

DOI: 10.3724/SP.J.1005.2011.01308

基因组选择及其应用

李恒德¹, 包振民², 孙效文^{1,3}

1. 中国水产科学研究院生物技术研究中心, 北京 100141;
2. 中国海洋大学海洋生命学院, 青岛 266003;
3. 中国水产科学研究院黑龙江水产研究所, 哈尔滨 150001

摘要: 品种选育在农业生产中占有十分重要的地位, 育种值估计是品种选育的核心。随着遗传标记的发展, 尤其是高通量的基因分型技术, 使得从基因组水平估计育种值成为可能, 即基因组选择。文章将基因组选择的方法分为两类: 一是基于估计等位基因效应来预测基因组估计育种值(GEBV), 如最小二乘法, 随机回归-最佳线性无偏预测(RR-BLUP)、Bayes、主成分分析等方法; 二是基于遗传关系矩阵来预测 GEBV, 通过采用高通量标记构建个体间的遗传关系矩阵, 然后用线性混合模型来预测育种值, 即 GBLUP 法, 并以这两种分类简要介绍了基因组选择各种方法的大致原理。影响基因组选择准确性的因素主要有标记类型和密度、单倍型长度、参考群体大小和标记-数量性状基因座(QTL)连锁不平衡(LD)大小等; 在基因组选择的各种方法中, 一般说来 Bayes 方法和 GBLUP 方法具有较高的准确性, 最小二乘法最差; GBLUP 计算速度快, 能够将标记和系谱结合起来, 因而比其他方法更具优势。尽管基因组选择取得了很大进展, 但在理论方面还面临着一些挑战, 如联合育种、长期选择的遗传进展及如何解析与性状有关和无关的标记等。基因组选择在一些动植物育种上已经开始应用, 在人类遗传倾向预测和进化动力学研究中也具有潜在的应用前景。基因组选择在个体间亲缘关系的量化上有了突破, 比传统方法更加精确, 因此, 基因组选择将会是动植物育种史上革命性的事件。

关键词: 基因组选择; 高通量遗传标记; 育种值估计

Genomic selection and its application

LI Heng-De¹, BAO Zhen-Min², SUN Xiao-Wen^{1,3}

1. The Centre for Applied Aquatic Genomics, Chinese Academy of Fishery Sciences, Beijing 100141, China;
2. College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China;
3. Heilongjiang River Fishery Research Institute, Chinese Academy of Fishery Sciences, Harbin 150001, China

Abstract: Selective breeding is very important in agricultural production and breeding value estimation is the core of selective breeding. With the development of genetic markers, especially high throughput genotyping technology, it becomes available to estimate breeding value at genome level, i.e. genomic selection (GS). In this review, the methods of GS was categorized into two groups: one is to predict genomic estimated breeding value (GEBV) based on the allele effect, such as

收稿日期: 2011-04-18; 修回日期: 2011-06-25

基金项目: 国家高技术研究发展计划(863 计划)项目(编号: SS2012AA100816)资助。

作者简介: 李恒德, 博士, 副研究员, 研究方向: 统计遗传学。E-mail: hengde.li@cafs.ac.cn

通讯作者: 孙效文, 研究员, 研究方向: 水产生物技术。E-mail: sunxw2002@163.com

网络出版时间: 2011-10-18 15:34:41

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20111018.1534.007.html>

least squares, random regression - best linear unbiased prediction (RR-BLUP), Bayes and principle component analysis, *etc.*; the other is to predict GEBV with genetic relationship matrix, which constructs genetic relationship matrix via high throughput genetic markers and then predicts GEBV through linear mixed model, *i.e.* GBLUP. The basic principles of these methods were also introduced according to the above two classifications. Factors affecting GS accuracy include markers of type and density, length of haplotype, the size of reference population, the extent between marker-QTL and so on. Among the methods of GS, Bayes and GBLUP are usually more accurate than the others and least squares is the worst. GBLUP is time-efficient and can combine pedigree with genotypic information, hence it is superior to other methods. Although progress was made in GS, there are still some challenges, for examples, united breeding, long-term genetic gain with GS, and disentangling markers with and without contribution to the traits. GS has been applied in animal and plant breeding practice and also has the potential to predict genetic predisposition in humans and study evolutionary dynamics. GS, which is more precise than the traditional method, is a breakthrough at measuring genetic relationship. Therefore, GS will be a revolutionary event in the history of animal and plant breeding.

Keywords: genomic selection; high throughput genetic marker; breeding value estimation

在农业生物(作物、畜禽和水产生物)的产业链中, 品种培育占有十分重要的地位。20 世纪 70 年代, 美国前国务卿基辛格曾说过: “谁控制了石油, 谁就控制了所有国家; 谁控制了粮食, 谁就控制了所有的人”。世界上与农业生产有关的著名公司, 如泰国正大集团、美国孟山都公司、瑞士先正达公司等之所以能够影响全球某个产业就是因为这些公司有优秀的品种出售, 占有着全球市场的最大份额。所以毫不夸张地讲“农业生产中, 谁控制了品种, 谁就控制了产业”。

通过选择进行品种(系)培育是进行育种的重要手段, 选择育种的核心是育种值估计, 即选择依据。传统的育种值估计方法是从Hazel的选择指数法^[1]到Henderson提出的基于性状和系谱的最佳线性无偏预测(Best linear unbiased prediction, BLUP)^[2], 记为TP-BLUP, TP-BLUP法通过系谱构建分子亲缘相关矩阵(Numerator relationship matrix)(通常记为A)将个体间的遗传联系用矩阵形式表示出来, 利用线性混合模型(Linear mixed model)计算个体的估计育种值(Estimated breeding value, EBV)。这一方法已经得到广泛应用, 并为动植物的遗传改良做出了重大贡献。

随着微卫星(Microsatellite)、单核苷酸多态(Single nucleotide polymorphisms, SNP)等遗传标记的发现, 将遗传标记与表型信息关联起来, 通过连锁分析找到与性状有关的标记或数量性状基因座(Quantitative trait loci, QTL), 与TP-BLUP结合起来

进行遗传评定, 这样不仅可以提高育种值估计的精确性, 而且可以在能够获得DNA时进行早期选择, 缩短世代间隔, 提高遗传进展, 即标记辅助选择(Marker assisted selection, MAS)^[3,4]。标记辅助选择中育种值估计的准确性随所用标记数目的增加而增高。

农业动植物基因组信息的公布和高通量基因型分型技术的实用化, 使得我们可以得到高通量标记如SNP, 通过整个基因组SNP信息估计出单个SNP或者不同染色体片段的效应值, 然后再将这些效应值相加最终得到个体全基因组估计育种值(Genomic estimated breeding value, GEBV), 由于同时利用这些高通量的SNP标记来进行育种值估计, 大大提高了育种值估计的准确性, 即基因组选择(Genomic selection)^[5]。研究表明应用基因组选择可以节省大量成本, 成倍提高遗传进展^[6]。目前农业生物中已经开展了基因组选择研究^[7-9]。

基因组选择实质上是全基因组水平的标记辅助选择, 从育种值估计的思路大致分为两种: 一是通过参考群体(Reference population)估计出每个SNP等位基因或者不同染色体片段的效应值, 然后利用这些效应值来计算候选群体(Candidate population)的GEBV。可以看出以参考群体估计标记效应计算GEBV的方法只能够对具有基因型的个体估计GEBV, 目前在大的育种群体中往往很难对全部或大部个体做到基因型检测, 另外参考群体和育种群体的遗传结构也不尽相同, 随着选育进展差别会越来越大, 所以需经常构建参考群体并进行标记效应

测定加以校正,同时这种方法也不能利用长期育种过程中积累的丰富翔实的系谱和育种记录。二是通过已测定的基因型计算个体间的相关关系,利用已测定基因型的个体计算的 G 矩阵和根据系谱计算的 A 矩阵合并成一个矩阵(记为 H),采用类似 TP-BLUP 方法进行育种值估计的方法,以提高基因组选择的准确性,这种基于遗传关系矩阵估计 GEBV 的方法可以将系谱和基因型信息结合起来,充分利用育种过程中积累的资料。

1 基于估计等位基因效应的基因组选择方法

基于估计等位基因效应的主要方法有最小二乘法(Least squares)^[5]、随机回归BLUP法(Random regression BLUP, RR-BLUP)^[5]、主成分分析法(Principal component analysis)^[10,11]和Bayes法^[5,12~14]等。

高通量的遗传标记数目往往多于用于基因型测定的个体数目,难以直接采用多元回归的方法同时估计每个单倍型(Haplotype)或等位基因(Allele)的效应,因为误差自由度将耗尽,所以通常的做法是,先对单个染色体片段或标记进行回归分析,将对数似然值比(Log-likelihood ratio)依每个片段或标记进行作图,根据合适的阈值,确定哪些片段或标记对性状有影响,然后用多元回归估计这些片段或标记的效应,将个体每个基因型相应的效应相加就是其估计育种值,这种方法即最小二乘法。其实质上就是先通过全基因组关联分析(Genome-wide association studies, GWAS)^[15]筛选与性状有关的染色体片段或标记,然后同时估计这些片段或标记的效应,计算个体育种值。RR-BLUP方法则将染色体片段或标记的效应是作为随机效应处理,用线性混合模型估计其效应,每个个体相应染色体片段或标记效应的和即为其估计育种值^[5]。主成分分析法则将基因型编码,两种纯合基因型分别编码为 1 和 -1,杂合基因型编码为 0,然后对所有标记所构成的矩阵进行标准化处理;对矩阵进行分解获得特征根最大的若干主成分;将这些主成分得分与表型建立回归关系,求解每个基因型的效应;根据这些基因型效应估计个体的育种值^[10,16]。Bayes方法结合了标记(或染色体片段)效应方差的先验分布(Prior distribution)和数据资料两方面的信息,这些效应方差的先验分布和后验分布(Posterior distribution)均为逆卡方分布

(Inverted chi-square distribution),采用Gibbs抽样方法进行上万次的循环,去掉burn in阶段的循环后,将其它循环的样本取均值作为这些效应的估计值。目前常用的Bayes方法主要是BayesA和BayesB^[5],主要的区别是BayesA允许不同标记有不同的方差且服从一定的分布,而BayesB允许一些标记的方差为零。

2 基于遗传关系矩阵的基因组选择方法

基因组选择的另一种思路是通过已测定的基因型计算个体间的相关关系,记为G矩阵,从而按照TP-BLUP的思路用G矩阵代替TP-BLUP中的A矩阵来估计育种值,即GBLUP法^[17]。

VanRaden^[17]提出了G矩阵的构建方法,将SNP标记进行编码,纯合基因型分别编码为 1 和 -1,杂合基因型为 0,令 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(Hardy-Weinberg Equilibrium)时的均值 $2(p_i - 0.5)$,得到Z矩阵,则可以计算得 $G = ZZ' / (2 \sum p_i(1 - p_i))$,这样G矩阵与分子血缘矩阵A类似。

可以看出GBLUP和这些以参考群体估计标记效应然后计算GEBV的方法只能够对具有基因型的个体估计GEBV。而大多数育种群体具有丰富翔实的系谱和育种记录,在大的育种群体中由于经费、人力、时间等因素的限制,难以对全部个体进行基因型测定,一些有育种记录但死亡的个体也无法再进行基因型测定。如果能将这些资料和基因型测定的个体等信息结合起来,将利用已测定基因型的个体计算的G矩阵和根据系谱计算的A矩阵合并成一个矩阵(记为H),采用类似TP-BLUP方法进行育种值估计^[18,19],无疑会提高基因组选择的准确性,在育种实践上也具有重要的应用价值。

利用H矩阵采用TP-BLUP方法进行育种值估计是GBLUP方法的推广。在一个大的育种资料中,可以根据系谱建立分子血缘矩阵A,这是进行TP-BLUP育种值估计的基础。对于已经测定基因型的个体,可以计算由这些个体组成的G矩阵,然后将G矩阵中的元素替换A矩阵中相对应的元素,从而形成矩阵 A_g , A_g 是比A更为准确的对个体间相关关系的衡量。对于一些未经基因型测定的后代,可以通过已经测定基因型的个体,根据系谱对这些后代

的相关关系进行修正, 进而得到矩阵 A_p , A_p 比 A_g 更为合理, 因为所有对公畜进行育种估计的绝大多数信息是包含在后代中的, 而不是祖先。 A_p 还可以根据整个系谱对祖先之间的相关关系进行修正, 进而得到矩阵 H , H 矩阵是由已经测定基因型的个体对整个 A 矩阵中的未经基因型测定的个体与其他个体之间相关关系的修正^[18]。传统的利用线性混合模型进行育种值估计需要 H 矩阵的逆阵, 由于 H 矩阵是 G 矩阵与系谱结合而成, 对大的系谱来说 H 矩阵的求逆将变的困难, Misztal等^[19]详尽描述了如何利用 H 矩阵进行遗传估计的计算过程。

3 影响基因组选择的因素

影响基因组选择的因素有很多, 如标记类型和密度、单倍型长度、参考群体大小和标记-QTL连锁不平衡(Linkage disequilibrium, LD)大小等。不同类型的标记具有不同的多态信息含量(Polymorphism information content, PIC), 一定长度基因组上标记的多少也直接影响遗传标记所能提供的总信息量, 同时标记数量越多, 越容易得到与QTL连锁不平衡的标记。基因组选择的准确性随标记密度的增加而增加, 微卫星标记密度从 $0.25Ne/morgan$ (Ne 为群体有效大小, effective population size)增加到 $2Ne/morgan$ 可使选择的准确性从 0.63 提高到 0.83; SNP标记密度从 $1Ne/morgan$ 增加到 $8Ne/morgan$, 可使选择的准确性从 0.69 提高到 0.86。要达到相似准确性所需的SNP标记数量大约是微卫星标记的 2~3 倍^[12]。

在一定的标记密度下, 由于标记间的距离和群体中LD的大小, 采用不同数目标记构成的单倍型进行基因组选择得到的准确性会有差异, Claus等^[21]模拟了标记密度约 $10Ne/morgan$ 时单倍型长度对基因组选择的准确性的影响, 发现当采用 10 个SNP标记作为单倍型时在各种标记密度下均可获得最高的准确性。Villumsen等^[14]在标记密度 $10Ne/morgan$ 时模拟了不同遗传力条件下单倍型长度对基因组选择准确性的影响, 同样得到采用 10 个SNP标记作为单倍型时育种估计的可靠性最高。

具有记录数的个体和进行基因型测定的个体数越多, 基因组选择的准确性越高^[5], 因为参考群体的扩大可以使单倍型或等位基因效应估计的准确性提高; 有时尽管有的个体没有基因型测定记录, 但

可以由系谱, 通过上下代之间的关系推断出一些个体的基因型, 进而还可以通过已推断出基因的个体推出其它一些个体的基因型, 将这些个体包括进参考群体也可以提高参考群体的数量, 进而提高基因组选择的准确性。将未经基因型测定的个体包含在参考群体内有助于提高低遗传力性状的育种值估计准确性, 提高程度与预测基因型的准确度有关^[22]。

标记-QTL间的LD可以增加基因组选择的准确性^[23], 随着世代的增加, 标记-QTL间的LD会逐渐降低, 进而引起基因组选择准确性的降低, 因此标记或单倍型效应经过若干世代就需要重新估计。Meuwissen等^[5]发现在基因型测定后的前两个世代基因组选择的准确性下降较快, 约为 5%, 其他世代则下降速度减慢。遗传力较高的性状随世代增加其基因组选择的准确性降低较慢^[14]。要想保证基因组选择较高的准确性, 每隔 2~3 个世代需要对标记或单倍型效应进行重新估计^[5]。不同交配设计也会显著影响标记-QTL之间的LD在每个世代的下降程度, 进而影响基因组选择的准确性^[24]。

基因组选择的准确性也受性状的影响, 不同性状采用同样的方法其准确性会有差异^[11], 这主要是因为不同性状的遗传力不同所造成的; 同一性状在不同环境条件下, 估计的标记效应不同, 显示基因型和环境的互作是遗传变异的重要组分^[11]。基因组选择很适合于世代间隔较长、繁殖力较低的物种, 如奶牛^[6], 由于不同物种的世代间隔和繁殖力不同, 因而基因组选择在不同物种中效果可能会有差异。

4 各种方法的比较

模拟研究发现在最小二乘法、RR-BLUP和Bayes方法中, 最小二乘法估计的准确性最低, 而Bayes方法估计的准确性最高, 其中BayesB估计的准确性要优于BayesA, 而RR-BLUP估计的准确性要略低于BayesA^[5,20]。从真实育种值对估计育种值的回归系数 b 来看, BayesB比BLUP和BayesA更接近 1, 尽管BayesA的准确性略高于RR-BLUP, 但RR-BLUP比BayesA的 b 值更接近 1^[5]; Lund等^[20]发现真实育种值对由RR-BLUP方法得到的GEBV的 b 值有时会比BayesB更接近 1。从计算的角度看, RR-BLUP法比Bayes方法更具吸引力, 而降低的效率却很少。

Zhong等^[24]用多个大麦近交系的基因型进行了

交配设计的模拟,发现基因组选择各种方法的准确性受不同模拟条件的影响,当QTL数目增加或性状的遗传力较高时,RR-BLUP法的准确性变得更高;大多数情况下,GBLUP比直接估计标记效应的方法更准确;当标记与大效应的QTL与标记存在强的连锁不平衡时,RR-BLUP方法的准确性最低。模型中包括标记信息要比TP-BLUP具有更高的准确性^[11]。

当标记-QTL间没有LD存在时,TP-BLUP具有最高的准确性,RR-BLUP法准确性接近于TP-BLUP,其次是BayesB,最小二乘法最差^[23];当标记-QTL间有LD存在时,BayesB和RR-BLUP的准确性最好,随着在世代数增加LD下降时,BayesB优于RR-BLUP和最小二乘法^[23]。尽管标记密度对基因组选择的准确性有影响,但对各种方法相对好坏影响不大^[24]。

从计算速度来讲,基于等位基因效应的育种值估计的方法有一些比较快,如主成分分析法,但比较常用的Bayes方法来讲则速度较慢。由于这些方法只能对已经基因型测定的个体进行育种值估计,无法利用过去的只有系谱而没有基因型测定的个体资料,而这一部分资料具有更多的记录信息。相比较而言,GBLUP方法与TP-BLUP方法类似,直观明了,更重要的是可以利用任何时候的系谱信息,然后将系谱资料与G矩阵结合起来,充分利用了育种资料所能提供的信息;由于GBLUP与TP-BLUP类似,只是提供给线性混合模型的矩阵元素上的差异,因而在计算速度上更具有优势。基因组选择的这两种思路在育种值估计的准确性上几乎没有差别^[17,20],因此,相对而言GBLUP更具优势。

由于GBLUP法相对于Bayes等其他方法,可以更好地充分利用系谱信息,不仅是系谱和基因型数据的完美融合,也是TP-BLUP方法在基因组时代的拓展,因而引起了育种者的极大兴趣,成为近年来的研究热点。

5 基因组选择面临的挑战

尽管基因组选择从提出到现在已经10年,而且实践中也发现比传统的遗传评估方式要具有更高的准确性,但目前仍然存在着一些挑战,如联合育种、品种间的基因组选择、长期进行基因组选择的准确性问题、如何区分与性状有关的标记和用于追踪个体间遗传关系的标记等问题。

进行联合育种时最主要的困难是测定基因型的个体数目相对于育种数据库中个体的数目要小很多,虽然对于已经测定基因型的个体间可以准确地估计出他们的遗传关系,但是这些信息在整个育种资料中所占的比重可能是有限的。Goddard和Hayes^[46]指出最现实的办法就是采用选择指数将用系谱资料估计得到的EBV和用遗传标记估计得到的GEBV进行加权。另一种办法是根据系谱和已经测定基因型的个体推断未经测定基因型的个体的基因型,这样可以避免不同育种群体使用不同遗传标记进行基因型测定所带来的问题,推断基因型也可以使得具有基因型个体的数目增加,更能准确衡量个体间的遗传关系,这个策略很诱人但在计算上是个很大的挑战^[46]。

长期进行基因组选择的遗传进展可能比表型选择和基于系谱选择的遗传进展小^[47,48,9],除非不断地加入新的遗传标记来预测育种值。可能的原因是GEBV是根据与QTL处于连锁不平衡的标记来估计的,如果不是完全连锁,固定标记并不能固定QTL,因此标记固定后,基因组选择将不能捕获部分QTL方差^[47];表型选择是同时利用所有的QTL,基因组选择只利用了能够检测到的QTL^[48]。Muir^[47]建议将多基因组分包含进GEBV中,Goddard^[48]和Jannink^[9]提出采用加权的基因组选择方法,对有利等位基因频率较低的标记给予较大的权重,防止低频率有利等位基因的丢失,以便维持遗传方差,使得长期的遗传进展最大化。

基因组选择在构建G矩阵时通常使用全部的遗传标记,其中有些标记与QTL处于连锁不平衡,因而这些标记可以用来估计GEBV;但是其他更多的标记可能与QTL并不处于连锁不平衡,由于这些标记可以描述个体间的部分遗传关系,因而可以估计到这些标记的效应。可能的解决方案是在模型中通过拟合多基因效应去除这些与QTL并不处于连锁不平衡的标记的影响,或者采用多个世代的个体做参考群体可以使GEBV的准确性维持更多世代^[47],也可以采用多个品种的个体做参考群体,因为如果一个标记在各个品种中均与性状有关,则这个标记与相应的QTL可能非常接近^[49]。Zhang等^[50]在构建G矩阵时采用方差对各个标记进行加权,避免了与性状无关的标记参与G矩阵的构建,但对于各个标记效应方差的估计仍然采用Bayes方法进行,在计算上仍

然不划算, 也没有彻底将与性状有关和无关的标记分离开来。

基因组选择依赖于候选群体与参考群体中标记与QTL处于相同的连锁相, 不同品种由于选择或进化的原因, 标记与QTL的连锁相可能不同^[51], 使用多个品种的个体作为参考群体可以显著提高GEBV估计的准确性^[49,51]。各个品种在参考群体中的组成仍然是需要研究的内容。在估计GEBV时只考虑了加性效应, 上位效应(Epistasis)对模型中每个方差组分有重要影响^[52], 如果模型中考虑非加性效应不仅可以增加GEBV估计的准确性, 也增加了表型值预测的准确性, 这对于制定选配方案有重要作用。Hu等^[53]利用 80 个标记对 126 个大豆重组近交系的研究表明, 模型中考虑上位效应预测表型值能力由单纯的加性模型的 0.33 提高到了 0.78。Gianola等^[54]提出了基因组选择的半参数方法, 可以估计数十万个标记间的互作。

目前高密度标记的基因型测定成本依然很高, 对一些物种来讲, 因其分布范围的局限或繁殖力低等原因, 开发高密度的遗传标记从经济上来说更不可行, 使用低密度标记比高密度标记可能更为现实, 因此采用低密度标记进行基因组选择可以降低基因型测定的成本。相对于高密度标记, 采用均匀分布的低密度标记引起基因组选择的准确性下降程度几乎不受QTL数目和方法的影响, 当选择与表型关联的低密度标记时, 则随QTL数目的增多, 基因组选择准确性下降越大^[55]。当选择相同数目的低密度标记时, 采用与表型关联的低密度标记并采用Bayes方法可以将准确性的损失降到最低^[55]。

此外, 目前的基因组选择研究仅局限于采用微卫星或SNP标记来估计GEBV, 对于其它一些可以引起遗传变异的因素还没有考虑, 如拷贝数变异(Copy number variation, CNV)和甲基化等, 考虑这些变异或许能够找回一些缺失的遗传力(Missing heritability)^[30]。高通量的SNP芯片检测技术可以用来估计CNV^[56], 因此, 采用高通量SNP检测技术, 将CNV等其它标记引入基因组选择模型不仅是可行的, 而且可能是必要的。

6 基因组选择的应用

基因组选择的最早应用实例是由荷兰 Euribrid 公司在 2006 年底开始的鸡的基因组选择(<http://www.worldpoultry.net/news/euribrid-makes-use-of-genomic-selection-883.html>), 采用 20000 个 SNP 标记的芯片对育种群进行基因型测定, 然后进行育种值估计。加拿大、德国、美国、新西兰、以色列、法国、爱尔兰、西班牙、澳大利亚、奥地利、波兰和丹麦等国也陆续开始了牛和猪的基因组选择研究。日本和墨西哥也开始了水稻和玉米的基因组选择研究。美国、新西兰和澳大利亚采用GBLUP方法估计的可靠性和较复杂的方法相比几乎一样高, 可靠性的增加使得一些育种公司已经开始商业运作, 这些公司的遗传进展至少提高了一倍^[7]。Su等^[25]对 3 330 头丹麦荷斯坦公牛 18 个性状的GEBV与EBV进行了比较, 发现 18 个性状的GEBV可靠性平均比EBV高出 0.13。Chen等^[26]对两个肉鸡品系共 348 030 只鸡采用GBLUP方法进行了育种值估计, 发现对高遗传力性状一步法GBLUP(one-step GBLUP)和BayesA方法的估计育种值准确性相似, 比传统的TP-BLUP方法提高 50%; 对低遗传力性状, GBLUP要比BayesA好 4~6 倍, 比TP-BLUP好 50%。Crossa等^[11]对玉米和小麦的研究表明, 由于标记的引入, 基因组选择比TP-BLUP预测能力提高 7.7~35.7%。Long等^[27]利用先筛选相关标记然后用Bayes方法对肉鸡死亡率的分类准确性可以提高至 90%。

利用全基因组标记在人类遗传倾向(Genetic predisposition)中也有潜在的应用前景^[28]。在复杂性状的GWAS中, 与性状相关的标记只能解释部分遗传方差^[29,30], 未能解释的部分称为缺失的遗传力(Missing heritability)。例如, 人类身高的遗传力高达约 0.8^[31,32], GWAS检测到的 50 个与身高有关的标记仅能解释总表型方差的约 5%^[33-36], 当借用GBLUP的思路, 同时采用全部标记时, 解释表型方差的比例可以达到约 45%, 通过对最小等位基因频率(Minor allele frequency)和LD的校正, 这一比例还会提高^[37]。同样的方法也被应用于阈值性状, 如Crohn病、躁郁症(Bipolar disorder)和I型糖尿病(type I diabetes)缺失遗传力的估计^[38]。

生物进化过程的衡量难度和其生态动力学

(Ecological dynamics)的遗传基础是研究基因如何影响表型、适应和群体动力学的主要困难,由动物育种所借鉴来的遗传方差-协方差矩阵(Genetic variance-covariance matrix)为进化学家研究宏观进化(Macroeolution)^[39~41]提供了方便。近年来,以混合线性模型和遗传方差-协方差矩阵为主的传统数量遗传学方法已经在研究现实生态条件下进化动力学的复杂性和限制性方面开始应用^[42~44],为进化生态学的研究带来了新的启示^[45]。由于GBLUP法相对于传统TP-BLUP法的优势,相信在不久的将来会被进化学家所借鉴,从而为进化生态学带来新的活力。

基因组选择的模拟研究结果十分诱人,但我们必须看到,基因组选择在理论上还有不完善的地方需要突破;无可否认,基因组选择在传统选择理论的基础上,结合了全基因组水平的信息,在量化个体间遗传关系方面有了突破,由高通量的遗传标记取代了传统的系谱,遗传关系的描述更加精确。因此,基因组选择将成为育种史上革命性的事件。

参考文献(References):

- [1] Hazel LN. The genetic basis for constructing selection indexes. *Genetics*, 1943, 28(6): 476–490. [DOI](#)
- [2] Hendersen CR. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 1975, 31(2): 423–447. [DOI](#)
- [3] Fernando RL, Grossman M. Marker assisted selection using best linear unbiased prediction. *Genet Sel Evol*, 1989, 21(4): 467–477. [DOI](#)
- [4] Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 1989, 121(1): 185–199. [DOI](#)
- [5] Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 2001, 157(4): 1819–1829. [DOI](#)
- [6] Schaeffer LR. Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed Genet*, 2006, 123(4): 218–223. [DOI](#)
- [7] Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. *Invited review: Genomic selection in dairy cattle: progress and challenges*. *J Dairy Sci*, 2009, 92(2): 433–443. [DOI](#)
- [8] Sonesson AK, Meuwissen TH. Testing strategies for genomic selection in aquaculture breeding programs. *Genet Sel Evol*, 2009, 41: 37. [DOI](#)
- [9] Jannink JL. Dynamics of long-term genomic selection. *Genet Sel Evol*, 2010, 42(1): 35. [DOI](#)
- [10] Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE. Reducing dimensionality for prediction of genome-wide breeding values. *Genet Sel Evol*, 2009, 41(1): 29. [DOI](#)
- [11] Crossa J, de los Campos G, Pérez P, Gianola D, Burgueno J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan JB, Arief V, Banziger M, Braun HJ. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, 2010, 186(2): 713–724. [DOI](#)
- [12] Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE. Genomic selection using different marker types and densities. *J Anim Sci*, 2008, 86(10): 2447–2454. [DOI](#)
- [13] Meuwissen THE, 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](#)
- [14] Villumsen TM, Janss L, Lund MS. The importance of haplotype length and heritability using genomic selection in dairy cattle. *J Anim Breed Genet*, 2009, 126(1): 3–13. [DOI](#)
- [15] Wang WYS, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet*, 2005, 6(2): 109–118. [DOI](#)
- [16] Long N, Gianola D, Rosa GJM, Weigel KA. Dimension reduction and variable selection for genomic selection: application to predicting milk yield in Holsteins. *J Anim Breed Genet*, 2011, 128(4): 247–257. [DOI](#)
- [17] VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*, 2008, 91(11): 4414–4423. [DOI](#)
- [18] Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. *J Dairy Sci*, 2009, 92(9): 4656–4663. [DOI](#)
- [19] 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](#)
- [20] Lund MS, Sahana G, de Koning DJ, Su G S, Carlborg Ö. Comparison of analyses of the QTLMAS XII common dataset. I: Genomic selection. *BMC Proc*, 2009, 3(Suppl. 1): S1. [DOI](#)
- [21] Calus MPL, Meuwissen THE, de Roos APW, Veerkamp RF. Accuracy of genomic selection using different methods to define haplotypes. *Genetics*, 2008, 178(1): 553–561. [DOI](#)
- [22] Pszczola MJ, Mulder HA, Calus MPL. The Accuracy of genomic selection using(un)genotyped animals to enlarge the reference population. paper presented at 9th world congress on genetics applied to livestock production. Leipzig, Germany, 2010 August. [DOI](#)

- [23] Habier D, Fernando RL, Dekkers JCM. The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 2007, 177(4): 2389–2397. [DOI](#)
- [24] Zhong SQ, Dekkers JCM, Fernando RL, Jannink JL. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics*, 2009, 182(1): 355–364. [DOI](#)
- [25] Su G, Guldbrandtsen B, Gregersen VR, Lund MS. Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. *J Dairy Sci*, 2010, 93(3): 1175–1183. [DOI](#)
- [26] Chen CY, Misztal I, Aguilar I, Tsuruta S, Meuwissen THE, Aggrey SE, Wing T, Muir WM. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: an example using broiler chickens. *J Anim Sci*, 2011, 89(1): 23–28. [DOI](#)
- [27] 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](#)
- [28] de los Campos G, Gianola D, Allison DB. Predicting genetic predisposition in humans: the promise of whole-genome markers. *Nat Rev Genet*, 11(12): 880–886. [DOI](#)
- [29] Maher B. Personal genomes: The case of the missing heritability. *Nature*, 2008, 456(7218): 18–21. [DOI](#)
- [30] Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TFC, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature*, 2009, 461(7265): 747–753. [DOI](#)
- [31] Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C, Dunkel L, De Lange M, Harris JR, Hjelmborg JVB, Luciano M, Martin NG, Mortensen J, Nisticò L, Pedersen NL, Skytthe A, Spector TD, Stazi MA, Willemsen G, Kaprio J. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res*, 2003, 6(5): 399–408. [DOI](#)
- [32] Macgregor S, Cornes BK, Martin NG, Visscher PM. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Hum Genet*, 2006, 120(4): 571–580. [DOI](#)
- [33] Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, Sulem P, Thorlacius S, Gylfason A, Steinberg S, Helgadóttir A, Ingason A, Steinthorsdóttir V, Olafsdóttir EJ, Olafsdóttir GH, Jonsson T, Borch-Johnsen K, Hansen T, Andersen G, Jorgensen T, Pedersen O, Aben KK, Witjes JA, Swinkels DW, den Heijer M, Franke B, Verbeek AL, Becker DM, Yanek LR, Becker LC, Tryggvadóttir L, Rafnar T, Gulcher J, Kiemeny LA, Kong A, Thorsteinsdóttir U, Stefansson K. Many sequence variants affecting diversity of adult human height. *Nat Genet*, 2008, 40(5): 609–615. [DOI](#)
- [34] Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, Eyheramendy S, Voight BF, Butler JL, Guiducci C, Illig T, Hackett R, Heid IM, Jacobs KB, Lyssenko V, Uda M, Diabetes Genetics Initiative; FUSION; KORA; Prostate, Lung Colorectal and Ovarian Cancer Screening Trial; Nurses' Health Study; SardiNIA, Boehnke M, Chanock SJ, Groop LC, Hu FB, Isomaa B, Kraft P, Peltonen L, Salomaa V, Schlessinger D, Hunter DJ, Hayes RB, Abecasis GR, Wichmann HE, Mohlke KL, Hirschhorn JN. Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet*, 2008, 40(5): 584–591. [DOI](#)
- [35] Melzer D, Perry JR, Hernandez D, Corsi AM, Stevens K, Rafferty I, Lauretani F, Murray A, Gibbs JR, Paolisso G, Rafiq S, Simon-Sanchez J, Lango H, Scholz S, Weedon MN, Arepalli S, Rice N, Washecka N, Hurst A, Britton A, Henley W, van de Leemput J, Li R, Newman AB, Tranah G, Harris T, Panicker V, Dayan C, Bennett A, McCarthy MI, Ruokonen A, Jarvelin MR, Guralnik J, Bandinelli S, Frayling TM, Singleton A, Ferrucci L. A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet*, 2008, 4(5): e1000072. [DOI](#)
- [36] Visscher PM. Sizing up human height variation. *Nat Genet*, 2008, 40(5): 489–490. [DOI](#)
- [37] Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*, 2010, 42(7): 565–569. [DOI](#)
- [38] Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet*, 2011, 88(3): 294–305. [DOI](#)
- [39] Shaw FH, Shaw RG, Wilkinson GS, Turelli M. Changes in genetic variances and covariances: G whiz! *Evolution*, 1995, 49(6): 1260–1267. [DOI](#)
- [40] Paulsen SM. Quantitative genetics of the wing color pattern in the buckeye butterfly (*Precis coenia* and *Precis evarete*): evidence against the constancy of G. *Evolution*, 1996, 50(4): 1585–1597. [DOI](#)
- [41] Steppan SJ, Phillips PC, Houle D. Comparative quantitative genetics: evolution of the G matrix. *Trends Ecol Evol*, 2002, 17(7): 320–327. [DOI](#)

- [42] Kruuk LEB. Estimating genetic parameters in natural populations using the "animal model". *Philos Trans R Soc Lond B Biol Sci*, 2004, 359(1446): 873–890. [DOI](#)
- [43] Ellegren H, Sheldon BC. Genetic basis of fitness differences in natural populations. *Nature*, 2008, 452(7184): 169–175. [DOI](#)
- [44] Kruuk LEB, Slate J, Wilson AJ. New answers for old questions: the evolutionary quantitative genetics of wild animal populations. *Annu Rev Ecol Evol Syst*, 2008, 39(1): 525–548. [DOI](#)
- [45] Wilson AJ, Réale D, Clements MN, Morrissey MM, Postma E, Walling CA, Kruuk LEB, Nussey DH. An ecologist's guide to the animal model. *J Anim Ecol*, 79(1): 13–26. [DOI](#)
- [46] Goddard ME, Hayes BJ. Genomic selection. *J Anim Breed Genet*, 2007, 124(6): 323–330. [DOI](#)
- [47] Muir WM. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J Anim Breed Genet*, 2007, 124(6): 342–355. [DOI](#)
- [48] Goddard M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*, 2009, 136(2): 245–257. [DOI](#)
- [49] De Roos APW, Hayes BJ, Spelman RJ, Goddard ME. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics*, 2008, 179(3): 1503–1512. [DOI](#)
- [50] Zhang Z, Liu J, Ding X, Bijma P, de Koning DJ, Zhang Q. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS One*, 2010, 5(9): e12648 [DOI](#)
- [51] Harris BL, Johnson DL, Spelman RJ. Genomic selection in New Zealand and the implications for national genetic evaluation. paper presented at Proc Interbull meeting. Niagara falls, Canada, 2008. [DOI](#)
- [52] Cheverud JM, Routman EJ. Epistasis and its contribution to genetic variance components. *Genetics*, 1995, 139(3): 1455–1461. [DOI](#)
- [53] Hu Z, Li Y, Song X, Han Y, Cai X, Xu S, Li W. Genomic value prediction for quantitative traits under the epistatic model. *BMC Genetics*, 2011, 12(1): 15. [DOI](#)
- [54] Gianola D, Fernando RL, Stella A. Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*, 2006, 173(3): 1761–1776. [DOI](#)
- [55] Habier D, Fernando RL, Dekkers JCM. Genomic selection using low-density marker panels. *Genetics*, 2009, 182(1): 343–353. [DOI](#)
- [56] Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*, 2007, 17(11): 1665–1674. [DOI](#)

•综合信息•

《遗传》被评选为 2011 年度中国精品科技期刊

2011 年 12 月 2 日, 中国科学技术信息研究所在北京召开“中国科技论文统计结果发布会”, 300 种期刊被评选为 2011 年度中国精品科技期刊, 《遗传》杂志榜上有名。



这次获得 2011 年度中国精品科技期刊证书的生命科学类期刊有：

Molecular Plant, 昆虫学报, 生理学报, 生态学报, 生态学杂志, 生物多样性, 生物化学与生物物理进展, 水生生物学报, 西北植物学报, 遗传, 应用生态学报, 植物病理学报, 植物生态学报, 植物学报, 中国科学生命科学等。