

利用 SNP 标记估计西门塔尔牛亲缘关系系数的准确性

张静静^{1,2}, 高会江², 吴洋², 朱波², 齐欣², 高雪², 张路培², 陈燕^{2*}

(1. 吉林农业大学动物科技学院, 长春 130118; 2. 中国农业科学院北京畜牧兽医研究所, 北京 100193)

摘要: 本研究旨在利用 SNP 标记估计西门塔尔牛亲缘关系系数, 以期准确确定估计个体间亲缘关系系数所需的 SNPs 数量。研究以 1 059 头出生于 2008—2012 年的西门塔尔牛为试验群体, 利用 Illumina bovineHD(770 k) 芯片, 根据最小等位基因频率(MAF)区间, 分别选择 100、500、1 000、1 500、2 000、2 500 和 3 000 个 SNPs 用于个体间亲缘关系系数的估计。结果显示, 随着标记数目的增多, 估计的亲缘关系系数准确性逐渐增加。且当 SNP 标记数目达到 2 500 时, 所估计的亲缘关系系数与利用所有标记估计的个体间亲缘关系系数差异不显著, 二者相关系数达到 0.89 以上。同时, 利用不同等位基因频率区间内标记估计的个体间亲缘关系系数差异不显著。由此可以看出, 当所选标记数目达到 2 500 以上时, 可以得到较高的亲缘关系系数估计准确性。本研究为基于 SNP 标记信息估计亲缘关系系数的进一步研究提供了理论基础, 同时为西门塔尔牛群体个体间亲缘关系的研究提供依据。

关键词: 西门塔尔牛; 亲缘关系系数; SNP; 最小等位基因频率

中图分类号: S823; S813.3

文献标志码: A

文章编号: 0366-6964(2016)02-0268-08

Estimation of the Kinship Coefficient in Simmental Cattle Based on SNP Markers

ZHANG Jing-jing^{1,2}, GAO Hui-jiang², WU Yang², ZHU Bo², QI Xin²,
GAO Xue², ZHANG Lu-pei², CHEN Yan^{2*}

(1. College of Animal Science and Technology, Jilin Agricultural University, Changchun 130118, China;
2. Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China)

Abstract: The objective of the study was to determine the number of SNPs which could efficiently estimate the kinship coefficient in Simmental cattle. 1 059 Simmental cattle born between 2008 and 2012 year were used as the reference population. Based on the interval of the minor allele frequency (MAF), 100, 500, 1 000, 1 500, 2 000, 2 500 and 3 000 SNPs located in Illumina bovineHD (770 k) chip were selected to estimate individual kinship coefficient. The results showed that with the increase of the SNPs' number, the estimation accuracy showed an increasing trend as well. Especially, when the SNPs' number reached to 2 500, there was no significant difference between relationship coefficients estimated using 2 500 SNPs and all SNPs, and the relationship coefficients were above 0.89 between them. Furthermore, the SNPs in the same interval with different allele frequency had no significant impact on the results. It was concluded that when the number of selected SNPs reached more than 2 500, a relatively higher estimation accuracy could be obtained. Our work has built a theoretical basis for further study of the kinship coefficient with high density

收稿日期: 2015-03-30

基金项目: 国家自然科学基金(31402039); 北京市自然科学基金(6154032); 中国农业科学院基本科研业务费(2014ywf-yb-4); 农业部物种资源保护(畜禽)项目(F557); 中国农业科学院科技创新工程经费(cxgc-ias-03); 科技支撑计划(2011BAD28B04); 863项目(2013AA102505-4); 中国农业科学院院本级增量业务费(2013ZL031)

作者简介: 张静静(1989-), 女, 山东莘县人, 硕士, 主要从事动物遗传育种与繁殖研究, E-mail: zhang_jingjing89@163.com

* 通信作者: 陈燕, E-mail: chenyan@caas.cn

SNP and provides a new clue for analyzing individual kinship relationship in Simmental cattle.

Key words: Simmental cattle; kinship coefficient; SNP; MAF

肉牛遗传评估工作的核心内容就是育种值估计。而在众多的育种值估计的方法中,最佳线性无偏预测(Best Linear Unbiased Prediction, BLUP)仍是最为有效和使用最为广泛的方法。其通过构建个体间的遗传关系矩阵来完成个体育种值的估计。BLUP 方法的优势在于能够充分利用个体之间的相关信息,即通过系谱信息构建的加性遗传关系矩阵(A matrix)来反映个体间的遗传相关关系并估计育种值。然而在实际生产过程中,因为往往存在着系谱错误、缺失和近交等问题,A 阵可能有较低的准确性^[1]。而且这种只通过系谱资料所获得的遗传相关关系只是亲缘关系的期望值,而真实的遗传相关关系会由于孟德尔抽样误差而与期望值有所偏差^[2]。

近年来,育种技术的不断发展使我国肉牛业取得了长足的进步。但是与发达国家相比,我国肉牛产业仍然处于初级阶段。公牛遗传评估技术的落后更是制约我国肉牛业发展的一个重要问题。错误的系谱信息将在很大程度上降低公牛遗传评估的准确性,进而降低群体的遗传进展。为降低系谱错误对育种带来的不利影响,提高育种值估计的准确性,进而加快我国肉牛的遗传进展,建立完整、准确的系谱信息显得尤为重要。而准确系谱信息的建立除严格的生产管理制度外,更有赖于个体间亲缘关系的估计。

在过去的十年里,对利用分子标记估计亲缘关系进行了很广泛的研究,出现了很多新方法以及软件^[3-5]。多项研究表明,分子标记如微卫星和 SNP 是检验和鉴定亲子关系的有效工具^[6]。SNP 标记信息也被用于估计群体中个体之间的亲缘关系^[7-8]。此外,研究还表明,有限的 SNPs 标记数目是限制亲缘关系估计准确性的主要瓶颈^[9]。随着高通量测序和高密度基因分型芯片的快速发展,数以千计的 SNPs 信息都能快速获得,进一步提高了分子标记在估计亲缘关系上的适用性。P. M. Vanraden 首次提出了基于高密度 SNP 芯片获得的 SNPs 信息估计亲缘关系矩阵(Genomic Relationship Matrix, 通常称为 G 阵),它可以代替 A 阵进行基因组育种值的估计^[10]。结果表明,系谱缺陷和孟德尔抽样误差等因素导致了 A 阵估计不准确,而利用高通量的 SNP 标记估计的 G 阵更接近于真实的个体间遗传相关关系^[11-12]。

另一方面,新的高通量基因分型技术的快速发展和高密度 SNP 芯片的出现产生了海量数据,也带来了高额的检测费用以及复杂的计算方法。为降低基因分型检测成本、节约计算时间,从高密度 SNP 标记中筛选高信息度的标记,用于估计个体亲缘关系受到众多研究者和育种家的关注。在猪的研究中,M. S. Lopes 等研究表明需要至少有 2 000 个 SNPs 标记用来估计遗传关系矩阵^[13]。在牛的研究中,M. M. Rolf 等研究发现利用 2 500~10 000 个标记时就能获得较高的亲缘关系估计准确性^[14]。

本研究以西门塔尔牛为研究群体,使用 Illumina bovineHD(770 k)牛高密度 SNP 芯片,依据不同的标记密度和最小等位基因频率(MAF),分别选择 100、500、1 000、1 500、2 000、2 500 和 3 000 个标记,估计个体间亲缘关系系数,并比较其准确性,以期优化标记选择方法和确定准确估计亲缘关系系数所需 SNPs 的数量。

1 材料与方法

1.1 试验材料

本研究的试验群体来自中国农业科学院北京畜牧兽医研究所牛遗传育种研究室在内蒙古乌拉盖地区构建的西门塔尔牛资源群体。试验动物选自出生于 2008~2012 年的 1 087 头西门塔尔公牛,详细情况见表 1。

表 1 西门塔尔牛出生年分布

Table 1 Birth year distribution for the genotyped Simmental cattle

出生年 Birth year	数量 Number	芯片类型 Panel
2008	17	770 k
2009	286	770 k
2010	243	770 k
2011	295	770 k
2012	246	770 k
合计 Total	1 087	770 k

1.2 基因型数据处理

试验群体使用了 Illumina bovineHD(770 k)芯片进行全基因组基因型测定,该芯片包含 777 962 个

SNPs 位点。在分析之前,所有 SNP 均进行质量控制,去除 SNP 检出率 $<95\%$ 、最小等位基因频率(Minor allele frequency,MAF) $<1\%$ 、极端不符合哈代-温伯格平衡检验 $P<10^{-6}$ 以及没有染色体位置信息的 SNP 位点和个体检出率 $<90\%$ 的个体。质量控制主要是利用 PLINK 软件^[15]完成。运行命令如下:plink-cow—file filename—maf0.01—geno0.01—mind0.1—hwe0.000001—recode12—out filename。

1.3 SNP 的不同筛选方法

本研究中采用在 3 种不同 MAF 区间(0.01~0.2,0.2~0.4 和 0.4~0.5)中等距离筛选多个标记密度形成分析数据集。标记数目方面分为 7 个密度梯度等级,分别为 100、500、1 000、1 500、2 000、2 500 和 3 000 个。然后,将不同 MAF 区域筛选的不同密度梯度的标记信息用于估计个体之间亲缘关系系数及它们的准确性,每组重复 100 次,并进行 t 检验评价准确性之间是否存在差异。以上工作,均是利用 R 语言自编程序计算完成。

1.4 连锁不平衡程度(LD)的计算

连锁不平衡程度(LD)的计算已有多种不同的方法,每一个都有不同的统计效用,其中以 D' ^[16] 和 r^2 ^[17] 最为常用。但是 r^2 对基因频率和有效群体大小的变化不敏感^[18],被广泛认为更加稳健并且是描述 LD 水平更好的参数。本研究采用 r^2 作为 LD 的衡量指标^[19]。

假定有两个位点,A 和 B,每个位点分别有两个等位基因, A_1, A_2 和 B_1, B_2 。其等位基因频率分别用 $P_{A1}, P_{A2}, P_{B1}, P_{B2}$ 表示。用 P_{11}, P_{12}, P_{21} 和 P_{22} 指代单倍型 A_1B_1, A_1B_2, A_2B_1 和 A_2B_2 的频率。计算这种连锁不平衡的方法:

$$r^2 = \frac{(P_{11}P_{22} - P_{12}P_{21})^2}{P_{A1}P_{A2}P_{B1}P_{B2}}$$

本研究运用 PLINK^[15] 软件 (<http://pngu.mgh.harvard.edu/~purcell/plink/>) 计算各子集中 SNP 之间的 r^2 。

1.5 亲缘关系系数的计算

把上下代个体间和同世代个体间的亲缘关系称为亲缘相关或血缘相关,有时也称为个体间的遗传相关。本研究使用 J. Yang 等^[20] 提出的利用个体间基因组亲缘关系系数的算法计算亲缘关系系数,公式:

$$A_{jk} = \frac{1}{N} \sum_i A_{ijk}$$

$$= \begin{cases} \frac{1}{N} \sum_i \frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i)}{2p_i(1 - p_i)} \\ 1 + \frac{1}{N} \sum_i \frac{x_{ij}^2 - (1 + 2p_i)x_{ij} + 2p_i^2}{2p_i(1 - p_i)} \end{cases}$$

其中, i 为第 i 个 SNP 位点, j 为第 j 个个体, k 为第 k 个个体, N 为 SNP 位点总数, P_i 为第 i 个位点等位基因频率。 X_{ij} 对应 aa、Aa、AA 基因型取 0、1、2。

本研究使用 770 k SNP 计算的亲缘系数和筛选 SNP 计算的亲缘系数的相关系数来表示准确性,公式:

$$r = \frac{\text{Cov}(Kin1, Kin2)}{\sqrt{\text{Var}(Kin1)\text{Var}(Kin2)}}$$

其中, r 为准确性, $Kin1$ 为使用所有标记计算的亲缘系数, $Kin2$ 为使用抽取的标记计算的亲缘系数。

2 结果

2.1 数据质控

数据详细的质控情况见表 2。经过质控后共得到 682 259 个 SNPs 位点和 1 059 个体用于下一步的分析。质量控制前后 SNP 位点在各染色体上的分布情况如图 1 所示,从图中可以看出不合格 SNP 位点数在各条染色体上所占比例相当。图 2 表示的是用于本研究群体中 SNPs 最小等位基因频率分布,从图中可以看出,全部 SNPs 的 MAF 分布均匀。此群体的平均 MAF 为 0.24。

表 2 SNPs 质量控制统计

Table 2 SNPs quality control statistic

质量控制标准 QC standards	SNP 标记数目 Number of SNPs
标记总数 Total number of SNPs	777 962
没有物理位置的标记 SNPs unknown position	2 078
Y 染色体上的标记 SNPs on chromosome Y	1 224
X 染色体上的标记 SNPs on chromosome X	39 367
最小等位基因频率 <0.05 的标记 SNPs with MAF <0.05	40 125
哈代-温伯格平衡检验 $P<10^{-6}$ 的标记 SNPs not in Hardy-Weinberg equilibrium ($P<10^{-6}$)	6 071
SNP 检出率 <0.95 SNPs with call rate <0.95	6 838
通过质量控制的 SNP 数 SNPs used after QC	682 259
SNP 缺失率 >0.1 的个体 Individuals with genotype missing rate >0.1	28

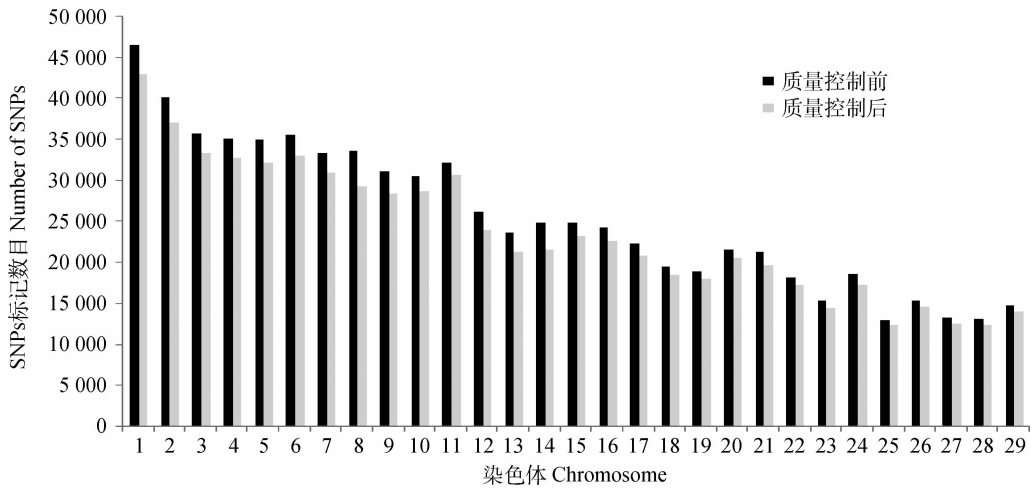


图 1 质量控制前后染色体上 SNP 位点数目分布

Fig. 1 Distribution of SNPs on each chromosome before and after quality control(QC)

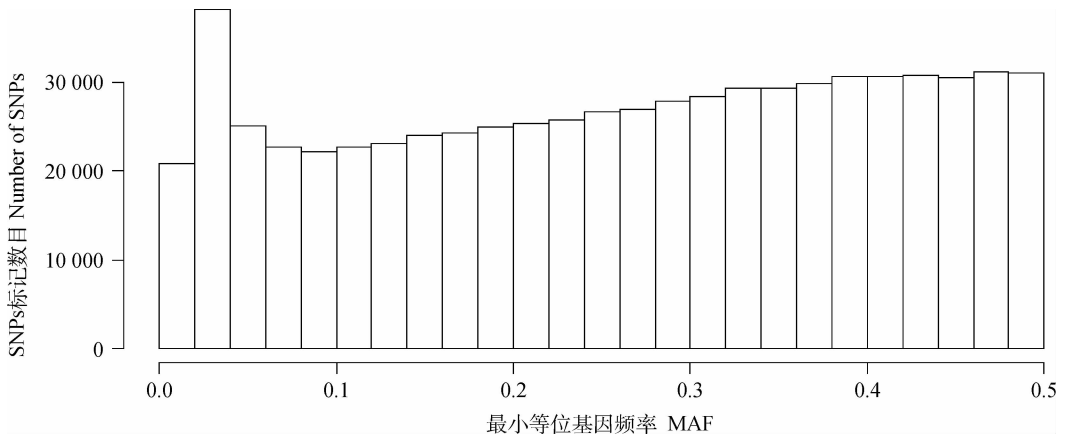


图 2 SNP 标记最小等位基因频率分布

Fig. 2 Distribution of minor allele frequency

2.2 亲缘关系系数分析

本研究在 3 个 MAF 区间(0.4~0.5, 0.2~0.4 和 0.01~0.2), 分别选取 100、500、1 000、1 500、2 000、2 500 和 3 000 个标记, 总计获得了 21 个 SNP 子集的数据。图 3 展示了在 3 个 MAF 区间内的标记数目分布。统计了选取的每一个 SNP 子集中相邻 SNP 间的连锁不平衡程度 r^2 值(表 3)。随着标记数目增多, SNP 之间的连锁程度加大。其中, 在选取 3 000 个来自于 MAF 区间(0.01~0.2) 的 SNP 有最大的 r^2 值, 连锁程度最高。在研究中将所筛选的不同密度梯度的标记信息用于估计个体之间的亲缘关系系数, 并和使用所有标记获得的亲缘关系系数计算相关性。

对质控后的 1 059 头西门塔尔公牛使用所有 682 259 个 SNPs 估计个体间的亲缘系数, 得到群体

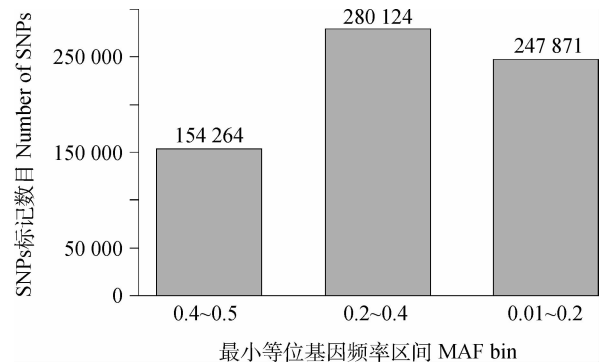


图 3 MAF 区间内标记数目

Fig. 3 Number of SNPs in each MAF bin

的亲缘系数平均值为 0.09, 标准差为 0.07。此外, 对 21 个 SNPs 子集进行了亲缘系数估计, 所得结果显示, 当所选标记数目为 $N=3 000$ 时, 通过 3 种 MAF 区间(0.4~0.5, 0.2~0.4 和 0.01~0.2) 筛选

表 3 相邻标记间的连锁不平衡程度统计

Table 3 Extent of linkage disequilibrium between 2 adjacent SNPs

筛选的 SNP 数量 No. of SNPs selected	最小等位基因频率区间 MAF bin		
	0.4~0.5	0.2~0.4	0.01~0.2
100	0.009±0.006	0.008±0.004	0.008±0.005
500	0.028±0.006	0.021±0.003	0.016±0.004
1 000	0.043±0.005	0.03±0.003	0.025±0.003
1 500	0.058±0.005	0.039±0.003	0.033±0.002
2 000	0.073±0.005	0.047±0.003	0.039±0.003
2 500	0.085±0.004	0.056±0.003	0.047±0.002
3 000	0.094±0.004	0.063±0.002	0.053±0.003

表中所列数值为 r^2 的平均值±标准误

Values listed in the table as the mean ± standard error of r^2

标记计算的西门塔尔公牛个体间亲缘关系系数与所有 SNP 数据计算的亲缘系数有最高的相关性,分别为 0.94 ± 0.001 ($P=0.792$)、 0.94 ± 0.001 ($P=0.339$)和 0.90 ± 0.004 ($P=0.693$),亲缘系数分别为 0.08、0.08 和 0.06。

图 4 列出了 3 种 MAF 区间(0.4~0.5, 0.2~0.4 和 0.01~0.2)下,不同 SNP 标记数目所得亲缘系数与所有 SNP 数据计算的亲缘系数相关性的盒图。从图中可以看出,当所选标记为 100 时,3 种筛选方法所得准确性均表现最低,分别为 $0.45 \pm$

0.007 ($P=0.834$), 0.44 ± 0.007 ($P=0.963$) 和 0.43 ± 0.023 ($P=0.680$)。当标记数目从 100 逐渐增加至 1 500 时,亲缘系数的相关性随着标记数目增加显著增加,之后增加的趋势减缓。且可以看出在标记数目为 100 时,亲缘系数的相关性有较大的波动。但是,选自 0.4~0.5 和 0.2~0.4 MAF 区间的标记,随着其数目增加,从图中可以看出亲缘系数的相关性非常稳定,而来自 0.01~0.2 区间标记,其相关性还是有明显的波动。

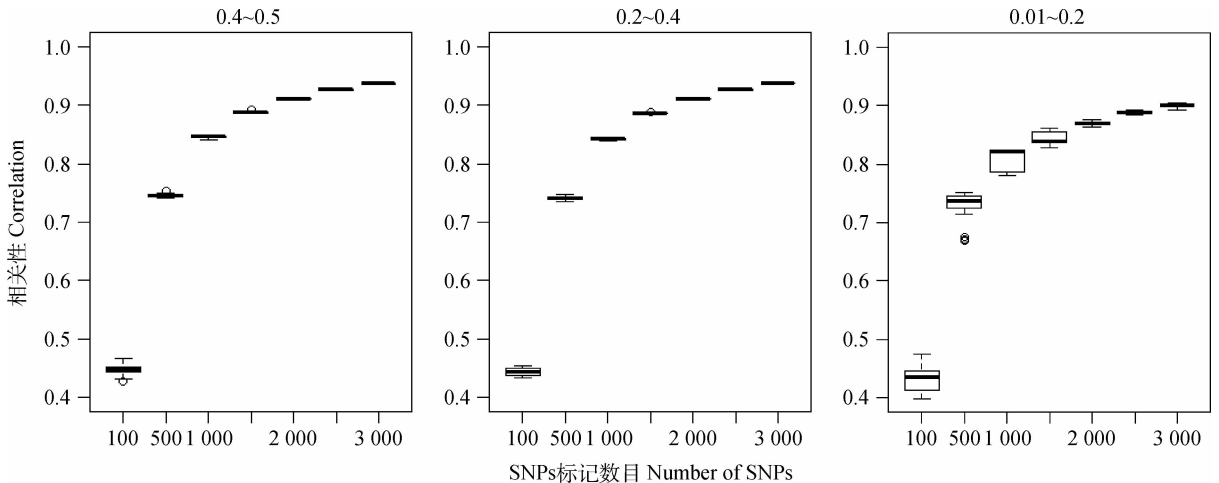


图 4 选择的 SNPs 和所有标记获得亲缘关系相关性盒形图

Fig. 4 Correlation between selected SNPs and all SNPs for kinship estimation

图 5 表示的是不同标记数目下,不同 MAF 区间所得亲缘系数的准确性。从图中可以得到,SNP 选自 MAF 0.4~0.5 和 0.2~0.4,亲缘关系相关系

数增长趋势一致,且基本吻合。SNP 选自 MAF 0.01~0.2,亲缘系数准确性也是保持增长趋势,但是增长速度相对较低。

综上所述,我们可知选择最小等位基因频率大于 0.2 且不连锁的 SNP,在数目达到 2 500 个时亲缘系数的估计值有高的准确性(相关系数大于 0.89)。

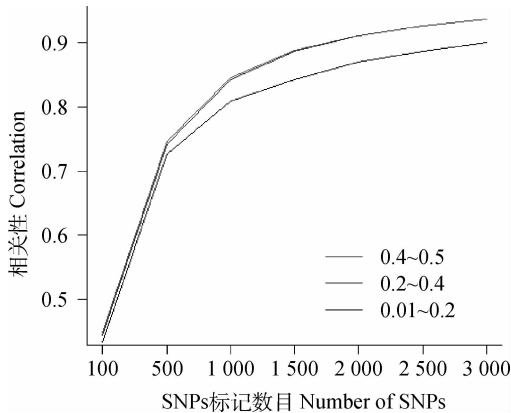


图 5 选择的 SNPs 亲缘关系系数增长趋势

Fig. 5 The growth trend of kinship coefficient between selected SNPs and all SNPs

3 讨论

随着高通量测序和基因分型技术的出现,利用 SNP 标记构建基因组亲缘关系矩阵(G 阵)来替代传统的加性遗传矩阵已经展开了大量的研究^[10,12,21-22]。本研究在不同最小等位基因频率区间内随机抽取 7 个标记密度(N = 100、500、1 000、1 500、2 000、2 500 和 3 000)的 SNPs,用于群体内个体间亲缘系数的计算。研究结果表明,标记位点的最小等位基因频率对个体间亲缘关系系数的计算有一定的影响,且在标记数目达到 2 500 时估计的亲缘系数有高的准确性。

3.1 SNP 的筛选

本研究中,将标记位点的最小等位基因分为 3 个区间,区间内位点随机抽取,并没有考虑位点之间是否关联。但是,当标记密度足够大的时候标记之间存在连锁不平衡状态,它们并不能完全独立分离的传递给下一代。在遗传信息传递过程中,每个子代总是精确的遗传来自双亲常染色体遗传物质的一半。但对位于 X 染色体上的遗传信息而言,它的传递往往与性别有关,这就使得在计算亲缘关系系数时,位于 X 染色体上的 SNP 标记信息将能更有效的估计出雌性个体间的亲缘相关关系,对雄性间的亲缘关系系数估计能力较差。而同性别后代的孟德尔抽样误差要比父本低^[23]。此次研究群体以雌性

个体为主,为此,该研究过程中,选择了剔除 X 染色体上的标记位点。

最小等位基因频率和标记密度是影响亲缘相关系数准确性的重要因素。C. Y. Chen 等报道了在荷兰斯坦牛群体中使用相等的等位基因组频率能够获得非常准确的亲缘相关系数和较小的偏差^[24]。S. Ogawa 等在日本黑牛群体中研究发现,随着标记数目增加亲缘相关系数估计准确性增加,当标记数目达到 10 000 时达到稳定^[25]。因此,本研究在 3 个 MAF 区间(0.4~0.5,0.2~0.4 和 0.01~0.2),分别随机选择 7 组 SNP(数目分别为 100、500、1 000、1 500、2 000、2 500、3 000)形成子集,来估计亲缘系数。

3.2 亲缘关系系数的估计

本研究选择了西门塔尔牛资源群体,通过选择 3 个 MAF 区间的 SNP 标记来分析最小等位基因频率对亲缘关系系数估计的影响。结果表明,使用来自 MAF 0.2~0.4 和 0.4~0.5 两个区间的 SNP 估计的群体的亲缘系数的平均值和标准差是相近的,高于使用 MAF 0.01~0.2 区间内的 SNP。根据前人研究报道,增加标记的数据量可能不会明显增高亲缘相关系数估计的准确性,且群体亲缘系数矩阵(G 阵)的估计最好使用完全不连锁的标记^[14]。另一项研究^[26]认为,随着标记密度增加,标记的连锁程度加大,导致了大量的标记信息的浪费。本研究使用了 7 个梯度的 SNPs 数目来估计亲缘系数,准确性是用和所有标记估计的亲缘系数相关性来表示,结果也验证了降低标记数目,没有导致亲缘系数准确性的显著下降。因此,使用低密度的 SNP 标记能够准确估计群体间亲缘相关系数,这就使降低基因分型成本成为了可能。

对于群体亲缘系数的估计,使用的 SNP 标记来自于 MAF 0.2~0.4 和 0.4~0.5,这两个区间估计准确性非常相近,且都是显著高于来自于 MAF 0.01~0.2 区间 SNP 估计的亲缘系数。这个可能是很多低频的 SNP 由于基因分型技术导致基因型错误,进而影响了亲缘系数估计准确性。因此在估计群体中个体间的亲缘系数时,应选择高频的 SNP 标记(MAF > 0.2)。研究中还统计了每个子集中相邻 SNP 间的连锁程度(表 3), r^2 值都是小于 0.1 的。随着标记数目的增多,连锁程度增加,亲缘系数估计的准确性也增加,这个研究结果是和前人报道相一致的^[25]。

当前研究已经显示,在西门塔尔牛群体中利用大于等于 2 500 个不连锁且最小等位基因频率大于 0.2 的 SNP 标记就能获得准确的亲缘相关系数。这个研究结果是和 M. M. Rolf 等在安格斯牛群体估计 SNP 数目在 2 500~10 000 个时能够获得一个稳健的亲缘相关系数^[14]相一致。M. S. Lopes 等在猪中发现,获得一个高精度的亲缘相关系数,至少需要 2 000 个 SNPs^[13]。在猪群体中可以用比较少的 SNPs 来构建 G 阵,这可能是因为它们基因组大小和单倍型块长度的差异引起的;J. A. Arias 等观测到牛的基因组遗传长度是 3 249 cM^[27];而猪的基因组遗传长度则在 1 797 和 2 149 cM 之间^[28]。此外,R. Veroneze 等发现,猪群体中平均的单倍型块大小是 395 kb,远大于荷斯坦奶牛的平均单倍型块大小(164 kb)^[29-30]。

4 结 论

高密度的 SNP 芯片已经在动物育种中广泛应用。但是这种芯片的价格较贵,并需要花费成倍时间运算。因此利用少量 SNP 标记进行基因组选择成为研究热点。本研究探讨了标记最小等位基因频率和密度对估计亲缘系数准确性的影响,结果表明,选择最小等位基因频率大于 0.2 且不连锁的 SNP,在数目达到 2 500 个时亲缘系数的估计值有高的准确性。

参考文献 (References):

- [1] BAUMUNG R, SÖLKNER J. Pedigree and marker information requirements to monitor genetic variability[J]. *Genet Sel Evol*, 2003, 35(4): 369-383.
- [2] VISSCHER P M, MEDLAND S E, FERREIRA M A, et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings[J]. *PLoS Genet*, 2006, 2(3): e41.
- [3] LI C C, HORVITZ D G. Some methods of estimating the inbreeding coefficient [J]. *Am J Hum Genet*, 1953, 5(2): 107-117.
- [4] AULCHENKO Y S, RIPKE S, ISAACS A, et al. GenABEL: an R library for genome-wide association analysis[J]. *Bioinformatics*, 2007, 23(10): 1294-1296.
- [5] QUELLER D C, GOODNIGHT K F. Estimating relatedness using genetic markers[J]. *Evolution*, 1989, 43(2): 258-275.
- [6] 周 磊, 初 芹, 刘 林, 等. 利用微卫星和 SNP 标记信息进行奶牛亲子鉴定的模拟研究[J]. *畜牧兽医学报*, 2011, 42(2): 169-176.
- [7] ZHOU L, CHU Q, LIU L, et al. Simulation study on paternity identification in dairy cattle with microsatellite and SNP markers [J]. *Acta Veterinaria et Zootechnica Sinica*, 2011, 42(2): 169-176. (in Chinese)
- [7] 张 哲, 罗元宇, 李晴晴, 等. 一种基于高密度遗传标记的亲缘鉴定方法及其应用[J]. *遗传*, 2014, 36(8): 835-841.
- [8] ZHANG Z, LUO Y Y, LI Q Q, et al. Developing and applying of a parentage identification approach based on high density genetic markers[J]. *Hereditas (Beijing)*, 2014, 36(8): 835-841. (in Chinese)
- [8] 郭 刚, 周 磊, 刘 林, 等. 利用 SNP 标记进行北京地区中国荷斯坦牛亲子推断的研究[J]. *畜牧兽医学报*, 2012, 43(1): 44-49.
- [9] GUO G, ZHOU L, LIU L, et al. Parentage inference with single nucleotide polymorphism markers in the Chinese holstein in Beijing [J]. *Acta Veterinaria et Zootechnica Sinica*, 2012, 43(1): 44-49. (in Chinese)
- [9] ZHANG Z, TODHUNTER R J, BUCKLER E S, et al. Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood[J]. *J Anim Sci*, 2007, 85(4): 881-885.
- [10] VANRADEN P M. Efficient methods to compute genomic predictions[J]. *J Dairy Sci*, 2008, 91(11): 4414-4423.
- [11] HILL W G, SALISBURY B A, WEBB A J. Parentage identification using single nucleotide polymorphism genotypes: application to product tracing[J]. *J Anim Sci*, 2008, 86(10): 2508-2517.
- [12] SANTURE A W, STAPLEY J, BALL A D, et al. On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs [J]. *Mol Ecol*, 2010, 19(7): 1439-1451.
- [13] LOPES M S, SILVA F F, HARLIZIUS B, et al. Improved estimation of inbreeding and kinship in pigs using optimized SNP panels[J]. *BMC Genet*, 2013, 14: 92.
- [14] ROLF M M, TAYLOR J F, SCHNABEL R D, et al. Impact of reduced marker set estimation of genomic relationship matrices on genomic selection for feed efficiency in Angus cattle[J]. *BMC Genet*, 2010, 11: 24.
- [15] PURCELL S, NEALE B, TODD-BROWN K, et al. PLINK: a tool set for whole-genome association and

- population-based linkage analyses [J]. *Am J Hum Genet*, 2007, 81(3): 559-575.
- [16] LEWONTIN R C. The interaction of selection and linkage. I. General considerations; heterotic models [J]. *Genetics*, 1964, 49(1): 49-67.
- [17] HILL W G. Estimation of linkage disequilibrium in randomly mating populations [J]. *Heredity (Edinb)*, 1974, 33(2): 229-239.
- [18] ZHAO H, NETTLETON D, DEKKERS J C. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms [J]. *Genet Res*, 2007, 89(1): 1-6.
- [19] TERWILLIGER J D, HAGHIGHI F, HIEKKALINNA T S, et al. A bias-ed assessment of the use of SNPs in human complex traits [J]. *Curr Opin Genet Dev*, 2002, 12(6): 726-734.
- [20] YANG J, BENYAMIN B, MCEVOY B P, et al. Common SNPs explain a large proportion of the heritability for human height [J]. *Nat Genet*, 2010, 42(7): 565-569.
- [21] GUO S W. Variation in genetic identity among relatives [J]. *Hum Hered*, 1996, 46(2): 61-70.
- [22] GARANT D, KRUK L E. How to use molecular marker data to measure evolutionary parameters in wild populations [J]. *Mol Ecol*, 2005, 14(7): 1843-1859.
- [23] PINTO N, GUSMÃO L, AMORIM A. X-chromosome markers in kinship testing; a generalisation of the IBD approach identifying situations where their contribution is crucial [J]. *Forensic Sci Int Genet*, 2011, 5(1): 27-32.
- [24] CHEN C Y, MISZTAL I, AGUILAR I, et al. Effect of different genomic relationship matrices on accuracy and scale [J]. *J Anim Sci*, 2011, 89(9): 2673-2679.
- [25] OGAWA S, MATSUDA H, TANIGUCHI Y, et al. Effects of single nucleotide polymorphism marker density on degree of genetic variance explained and genomic evaluation for carcass traits in Japanese Black beef cattle [J]. *BMC Genet*, 2014, 15: 15.
- [26] WEIR B S, ANDERSON A D, HEPLER A B. Genetic relatedness analysis: modern data and new challenges [J]. *Nat Rev Genet*, 2006, 7(10): 771-780.
- [27] ARIAS J A, KEEHAN M, FISHER P, et al. A high density linkage map of the bovine genome [J]. *BMC Genet*, 2009, 10: 18.
- [28] TORTEREAU F, SERVIN B, FRANTZ L, et al. A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content [J]. *BMC Genomics*, 2012, 13: 586.
- [29] VERONEZE R, LOPES P S, GUIMARÃES S E, et al. Linkage disequilibrium and haplotype block structure in six commercial pig lines [J]. *J Anim Sci*, 2013, 91(8): 3493-3501.
- [30] QANBARI S, PIMENTEL E C, TETENS J, et al. The pattern of linkage disequilibrium in German Holstein cattle [J]. *Anim Genet*, 2010, 41(4): 346-356.

(编辑 郭云雁)