

# 利用 SNP 标记进行北京地区中国荷斯坦牛亲子推断的研究

郭刚<sup>2</sup>, 周磊<sup>1</sup>, 刘林<sup>3</sup>, 李东<sup>1</sup>, 张胜利<sup>2</sup>, 刘剑锋<sup>1</sup>, 丁向东<sup>1</sup>,  
张毅<sup>1</sup>, 王雅春<sup>1\*</sup>, 张勤<sup>1</sup>, 张沅<sup>1</sup>

(1. 中国农业大学动物科技学院, 农业部动物遗传育种试验室, 北京 100193; 2. 首农集团北京三元种业科技股份有限公司三元绿荷奶牛养殖中心, 北京 100076; 3. 北京奶牛中心, 北京 100085)

**摘要:** 本研究旨在利用 SNP 标记对北京地区中国荷斯坦牛群进行亲子推断, 并分析场、母牛出生年月、公牛家系对系谱错误率的影响, 以期为指导奶牛育种和生产管理提供依据。共选取了 255 个最小等位基因频率大于 0.45 的高多态 SNPs 标记, 利用似然法, 采用 Cervus3.0 软件对北京地区 84 头荷斯坦公牛和 1 927 头母牛进行亲子推断研究。结果显示, 试验群体平均系谱错误率为 20.9%, 不同的场、出生年份和月份的母牛的系谱错误率有显著差异 ( $P < 0.05$ ), 而各公牛家系间系谱错误率差异不显著 ( $P > 0.05$ )。结果说明, 错误系谱的发生主要是由于牛场本身记录不完善造成的。在我国亟需建立利用遗传标记监测、校正系谱准确性的制度, 采取措施提高系谱的准确性, 加快我国荷斯坦牛遗传改良进程。

**关键词:** 中国荷斯坦牛; 似然法; 亲子推断; SNP; 系谱错误

中图分类号: S823; Q348

文献标识码: A

文章编号: 0366-6964(2012)01-0044-06

## Parentage Inference with Single Nucleotide Polymorphism Markers in the Chinese Holstein in Beijing

GUO Gang<sup>2</sup>, ZHOU Lei<sup>1</sup>, LIU Lin<sup>3</sup>, LI Dong<sup>1</sup>, ZHANG Sheng-li<sup>2</sup>, LIU Jian-feng<sup>1</sup>,  
DING Xiang-dong<sup>1</sup>, ZHANG Yi<sup>1</sup>, WANG Ya-chun<sup>1\*</sup>, ZHANG Qin<sup>1</sup>, ZHANG Yuan<sup>1</sup>

(1. Key Laboratory of Animal Genetics and Breeding of Ministry of Agriculture, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China; 2. Beijing Sanyuan Luhe Dairy Cattle Center, Beijing Sanyuan Breeding Technology Co., Ltd, Capital Agribusiness Group, Beijing 100076, China; 3. Beijing Dairy Cattle Center, Beijing 100085, China)

**Abstract:** The aim of this study were to use single nucleotide polymorphisms (SNPs) markers for parentage inference in the Chinese Holstein, to discuss the effects of herd, birth year and birth month of the cow, sire family on the pedigree error rate, and to guide the cattle breeding and management. 255 highly polymorphic SNP markers, with the average minor allele frequency (MAF) over 0.45, were used for parentage inference among 84 Chinese Holstein bulls and 1 927 cows in Beijing. The analysis was carried out by the Cervus3.0 software adopting likelihood method. The average pedigree error rate was 20.9% in the experimental population. Moreover, the pedigree error rate among different herds, birth year and birth month of the cow were significantly different ( $P < 0.05$ ), while there was no significant difference among sires ( $P > 0.05$ ).

收稿日期: 2011-02-21

基金项目: “863”重大项目(2008AA101002); 国家奶牛产业技术体系专项资金; “十二五”国家科技支撑计划课题(2011BAD28B02); 国际合作项目(2008DFA31120); 948 项目(2010-C14)

作者简介: 郭刚(1980-), 男, 河北河间人, 博士, 主要从事动物遗传育种与繁殖研究, E-mail: guogang2180@gmail.com

\* 通讯作者: 王雅春(1968-), 博士, 副教授, 主要从事分子数量遗传学研究, Tel: 010-62732439, E-mail: wangyachun@cau.edu.cn

Therefore, the occurrence of pedigree error was mainly due to the poor management in the dairy herds. It is very necessary to establish pedigree supervision and correction system and take measures to improve the pedigree accuracy, which would speed up the process of dairy genetic improvement in China.

**Key words:** Chinese Holstein; likelihood method; parentage inference; single nucleotide polymorphism (SNP); pedigree error

系谱信息在奶牛育种中有着重要的作用。错误的系谱会大大降低遗传评定准确性,影响选种选配的效果,降低群体的遗传进展,给奶牛生产带来巨大的经济影响。然而,在实际养殖生产中,系谱记录错误是难免的,它受多种因素影响,如配种记录、产犊记录、系谱录入及整理中的人为错误等等,尤其在饲养管理较粗放的散养式牧场,犊牛出生后很难确定父亲甚至双亲,往往会造成系谱错误。世界范围平均的奶牛系谱错误率约为 11%<sup>[1]</sup>。汪湛等<sup>[2]</sup>报道了天津部分奶牛场的系谱错误率为 11.83%,初芹等<sup>[3]</sup>利用微卫星标记对北京地区 21 个场的系谱错误率进行检测,平均为 16.3%。

进入 21 世纪后,随着 SNP 标记检测手段的不断成熟,利用 SNP 标记进行牛的亲子推断成为可能,Heaton 和 Werner<sup>[4-5]</sup> 等也分别报道了用 32~37 个 SNPs 标记在肉牛、奶牛的多品种群中进行亲子推断的结果。2009 年 Fisher 等人的研究表明利用一定数目的 SNPs 标记进行奶牛亲子鉴定可以达到或超过目前所常用的微卫星标记体系的效率<sup>[6]</sup>。

目前我国尚无利用 SNP 标记进行实际奶牛群体亲子推断的研究。本研究利用北京地区中国荷斯坦牛基因组选择资源群体的 SNP 标记基因型信息,研究其亲子推断的效率,并分析系谱错误率的影响因素,以期降低系谱错误,改善牛场生产管理,加快我国奶牛育种进程提供理论依据。

## 1 材料与amp;方法

### 1.1 资源群体

研究群体为北京地区的中国荷斯坦牛,包括 13 头种公牛及其 1 927 头女儿,其中 2 头种公牛为同卵双生且基因型完全相同,合并处理;另外随机选择了 71 头荷斯坦青年公牛同时作为候选公牛。公牛女儿分布于北京地区 15 个牛场,系谱信息由中国奶业协会数据处理中心提供。

### 1.2 标记基因型检测与筛选

使用 DP318 基因组提取试剂盒(天根,北京)从

母牛凝血和公牛冻精中提取基因组 DNA。检测 DNA 浓度后全部定量到  $10 \text{ ng} \cdot \mu\text{L}^{-1}$ 。使用 Illumina Bovine SNP50 芯片检测标记基因型,由 Illumina BovineSNP50 Beadchip 仪器完成,每个芯片含有 54 001 个 SNPs。

用于亲子推断的 SNP 标记的筛选标准参照美国农业部的研究(<http://cgemm.louisville.edu/USDA/index.html>):1) 最小等位基因频率(MAF)大于 0.45; 2) 每个标记的检出率(Call rate)大于 0.95; 3) 每条染色体上相邻 SNPs 间距大于 10 Mb。

美国农业部推荐 121 个 SNPs 标记进行亲子鉴定,其中有 112 个标记在本研究的芯片中,而本研究所选的 255 个标记与美国推荐的标记没有直接关系。

### 1.3 亲子推断

亲子推断采用 Cervus3.0 软件<sup>[7-9]</sup>,以 84 头公牛作为候选亲本,利用最大似然原理对母牛群体进行亲本推断。Cervus 软件针对每头母牛,对提供的所有候选亲本进行计算,并给出似然得分最高的亲本。

本研究中,当软件推断出的父亲与后代的似然得分大于零时,就认为该公牛是真实父亲,似然得分越高,可靠性越高;如果推断的母牛真实父亲与实际系谱记录不一致,则认为其为系谱错误。当软件推断出的父亲与后代的似然得分小于零时,认为它不是真实父亲,即该后代的真实父亲不在候选亲本中,也作为系谱错误处理。这两部分的系谱错误合并为总的系谱错误率。

### 1.4 系谱错误的影响因素分析

采用以下统计模型对影响系谱错误的因素进行统计分析:

$$y = \mu + sire + herd + year + month + e$$

其中, $y$  表示系谱是否错误,错误用 1 表示,否则用 0; $\mu$  是群体总平均值; $sire$  为公牛家系效应; $herd$  为场效应; $year$  为母牛出生年份效应; $month$  为母牛出生月份效应; $e$  为随机残差效应。统计分

析采用 SAS 9.2 GLM 软件包来完成。

## 2 结果

### 2.1 标记分布

本研究从 Illumina Bovine50 SNP 芯片数据结果中共筛选出 255 个高多态性 SNPs 标记, 这些标记分布于 29 条常染色体上, 平均间距为 10.62 Mb, 具体的分布情况如图 1 所示。其中, 1 号染色体上标记数最多, 共 16 个, 平均间隔为 10.53 Mb。筛选出的标记最小等位基因频率均高于 0.45, 检出率高于 0.95, 适合用于进行亲子推断。

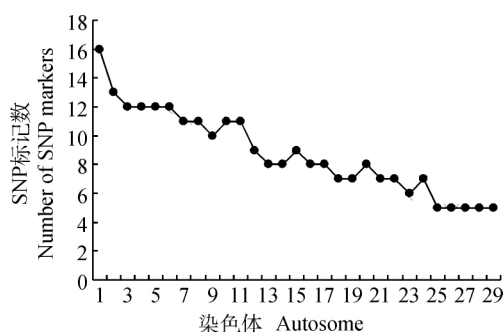


图 1 255 个 SNP 标记在 29 条常染色体上的分布情况  
Fig. 1 The distribution of 255 SNPs markers on 29 autosomes

### 2.2 亲子推断结果

使用 255 个 SNPs 标记, 将 84 头公牛作为候选亲本, 对 1 927 头母牛进行了亲子推断研究, 其中 1 847 头(95.8%)母牛被推断出的最可能父亲的似然得分大于零且推断父亲的置信度均达到 95% 以上; 有 402 头母牛存在系谱错误, 其中 80 头(4.1%)的 pair LOD 值小于 0, 可认为没有推断出父亲, 即真实父亲不在 84 头公牛内; 其余的 322 头推断出了父亲, 其中 196 头(占 322 头的 60.9%)的真实父亲在 13 头公牛中, 126 头(占 322 头的 39.1%)的真实父亲在其余 71 头公牛中。1 847 头母牛的最可能父亲的似然得分的分布情况如图 2。可见似然得分多在 30~50 之间, 即这些母牛推断的父亲可信度很高。

理论上, 真实亲本与其后代的基因型之间不应存在孟德尔矛盾, 但这 1 847 头母牛中仍有少数个体(4%比例)的基因型与其推断的最可能父亲基因型之间存在孟德尔矛盾(表 1)。然而, 这些亲子对中存在孟德尔矛盾的标记数均很少, 最多仅为 4 个。造成这种现象的最可能原因是标记判型错误。

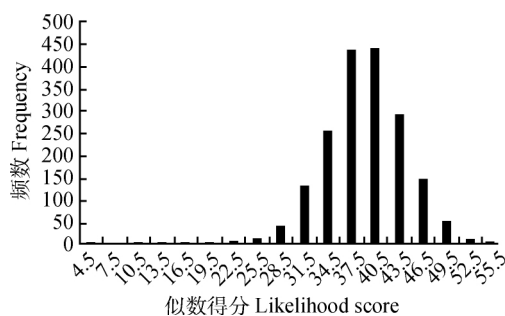


图 2 1 847 头母牛与其推断的最可能父亲的似然得分的频数分布

Fig. 2 The distribution of likelihood scores of 1 847 cows with their most likely sires referenced

表 1 1 847 头母牛与其推断的最可能父亲的孟德尔矛盾标记数

Table 1 The number of markers with Mendelian conflict of 1 847 cows with their most likely sire referenced

孟德尔矛盾标记数	个体数
0	1 771
1	52
2	21
3	1
4	2
总计	1 847

根据 255 个 SNPs 标记的亲子推断结果, 1 927 头母牛中有 80 头母牛未推断出真实父亲, 323 头母牛推断的真实父亲与系谱记录不一致, 即 403 头母牛的公牛系谱是错误的, 平均系谱错误率为 20.9%。

### 2.3 系谱错误影响因素

系谱错误的产生受多种因素影响。在本研究中, 考虑了母牛所在场、出生年份、出生月份及公牛家系对系谱错误的影响, 结果显示公牛间差异不显著( $P > 0.05$ ), 不同场间系谱错误率有极显著差异( $P < 0.001$ ), 母牛出生年份和出生月份间的系谱错误率差异显著( $P < 0.05$ ), 具体见表 2。

对各牛场的系谱错误情况分别进行了统计(表 3), 可见不同牛场间系谱错误率差别非常大, 系谱错误率最低的牛场其错误率仅为 6.9%, 而系谱错误率最高的牛场其错误率高达 46.3%。15 个牛场中有 8 个牛场的系谱错误率都在 20% 以上。

表 2 系谱错误率的各影响因素效应分析

Table 2 The effect of different factors on pedigree error

因素 Factor	场 Herd	出生年份 Birth year of daughter	出生月份 Birth month of daughter	父亲 Sire
F 值 F value	9.41	3.31	1.79	1.12
P 值 P value	<.000 1	0.005 6	0.040 5	0.266 3
显著性 Significance	***	**	**	NS

\*\*\*. 极显著 ( $P < 0.000 1$ ); \*. 显著 ( $P < 0.05$ ); NS. 不显著

\*\*\*. Extreme significance ( $P < 0.000 1$ ); \*. Significance; NS. No significance

表 3 各牛场系谱错误率

Table 3 Pedigree error rates of different herds

场 Herd	检测母牛数 Number of cows detected	系谱错误母牛数 Number of cows with rejected paternity	系谱错误率/% Pedigree error rate
01	80	26	32.5
02	82	38	46.3
03	83	27	32.5
04	107	15	14.0
05	110	34	30.9
06	110	10	9.1
07	114	16	14.0
08	118	13	11.0
09	135	14	10.4
10	139	25	18.0
11	146	30	20.5
12	158	40	25.3
13	158	46	29.1
14	188	13	6.9
15	199	56	28.1
总计 Total	1 927	403	20.9

系谱错误比例 = 系谱错误母牛数 / 检测母牛数

Pedigree error rate = Number of cows with rejected paternity / Number of cows detected

不同年份出生母牛的系谱错误率差异也非常大, 2002 年系谱错误率最低, 为 16.85%, 2006 年最高, 为 27.84%。2001-2006 年, 平均系谱错误率波动很大, 但总体呈上升趋势, 为 22.39%~27.8%(图 3)。

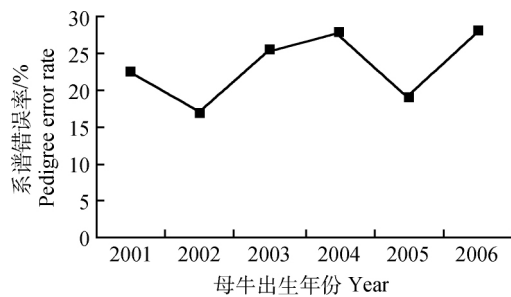


图 3 不同产犊年份系谱错误率

Fig. 3 Pedigree error rates of different years

从母牛出生月份角度来看, 1 月和 10~12 月份系谱错误率较低, 平均为 18.74%, 9 月份系谱错误率最高, 达到 28.82%, 而其它月份系谱错误率较平稳, 在 23.19%~26.89% 之间(图 4)。

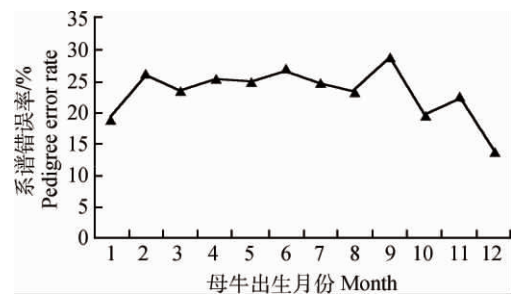


图 4 不同产犊月份的系谱错误率

Fig. 4 Pedigree error rates of different months

### 3 讨论

#### 3.1 系谱错误率及影响因素

本研究中,北京地区的荷斯坦母牛的系谱错误率平均为 20.9%,高于汪湛和初芹<sup>[2-3]</sup>等报道的错误率数据。一方面是由于本研究使用的 SNPs 标记数目多,分布广,具有更高的推断效率,如周磊<sup>[10]</sup>的模拟研究发现,利用 SNP 标记进行亲子鉴定,在最小等位基因频率为 0.35~0.50 时,以至少 2 个矛盾排除原则推断需要 54 个 SNPs,本研究共挑选了 255 个间距大、分布均匀、MAF 高的标记,达到了很高的推断水平。而另一方面也可能是由于研究群体本身之间系谱错误情况就存在差异造成的。

在影响系谱错误率的诸多因素中,牛场是主要因素,不同牛场间系谱错误率差异较大,最高牛场系谱错误率约是最低牛场系谱错误率的 7 倍,显示出不同牛场的管理水平及对系谱登记工作的重视程度和对分娩牛监督管理水平的差异。实际生产中有多种因素可造成系谱登记错误。例如,奶牛多在夜间分娩,如果接生人员没有及时监护,可能造成几个犊牛同晚出生而分不清母亲,进而导致犊牛父亲的记录错误;实际生产中牛耳标遗失后会根据谱系中犊牛的花片来确认牛只,往往也会造成一定的错误。此外,如果牛场的考核指标存在缺陷,也可能间接造成人为的系谱记录错误,如有的牛场将犊牛死淘率作为考核指标,可能用双胞胎个体中的一头去取代死胎而导致人为的系谱错误。

分析结果显示,不同母牛出生年份和月份间系谱错误率存在显著差异。2001-2006 年,北京地区牛场的系谱错误率有增加趋势,从 22.3% 升高到 27.8%,造成这种现象的原因可能是由于奶牛存栏数量逐年增加,牛场规模不断扩大增加了管理难度造成的。1 年的 12 个月份中,9 月份出生母牛系谱错误率最高,而 1 月份和 10~12 月份的系谱错误率较低,这与北京地区奶牛的产犊季节性有关。在实际生产中,9 月份是大多数牛场的产犊高峰期,产犊数量激增造成牛场值班人员工作量增加,管理难度加大,系谱错误率上升;而每年的 1~3 月份发情牛只较少,因此 10 月份到次年 1 月份的产犊数量较少,系谱错误率也相对较低。

此外,公牛对系谱错误率并没有显著影响,说明系谱错误主要是牛场本身的管理因素造成的,而非遗传因素。因此,在中国亟待采取一系列措施提高

牧场管理水平,加强繁殖监管,完善系谱记录制度,进而提高系谱准确性。

#### 3.2 系谱错误的监测与校正

Weller 等提出,如果保证每年牛群的遗传进展在 1% 以上,那么系谱错误率需控制在 8% 以下<sup>[11]</sup>。此外,据模拟研究结果,如果美国奶牛遗传评估中母牛的父亲记录错误率为 11%,群体遗传进展将会降低 11%~15%<sup>[11]</sup>。根据 Visscher 等<sup>[12]</sup>提出的系谱错误造成的选择进展损失公式,本研究中一半的牛场系谱错误率都在 20% 以上,如果假设选择性状遗传力为 0.25,每头公牛女儿数为 50,那么选择进展的损失率约为 12%。

为了降低系谱错误对生产和育种带来的不利影响,各奶业发达国家早已建立了系谱错误检测、校正的监管系统,利用遗传标记进行群体随机抽样的亲子鉴定。例如,澳大利亚荷斯坦协会,在全国每 500 头登记出生的犊牛中就抽检 1 头,要求农户将犊牛及其母亲的毛囊样品寄到协会,由协会出资进行亲子鉴定。

在我国奶牛业迅猛发展的现阶段,已成熟的标记检测方法,包括微卫星标记和 SNP 标记,建立系谱准确性的监管体系已经迫在眉睫,以监督后裔测定系统的正确性和准确性为出发点,进而加强记录管理意识和手段指导生产,可加速中国奶牛群体遗传进展。这也是今后我国奶牛遗传改良的重要基础工作之一。

#### 3.3 SNP 和 SSR 标记的比较

随着芯片技术的不断发展,SNP 标记应用越来越广泛,SNP 标记具有自动化检测程度高,基因型判定错误率低,突变率低,基因组分布广泛等优点<sup>[5]</sup>,其缺点是单个标记多态性低,需要标记数多,但芯片技术给 SNP 的应用创造了有利条件。微卫星具有单个标记多态性高,需要标记数少的优点,但存在基因型判定错误率高的缺陷,而基因型判定错误对亲子鉴定的准确性影响很大。因此,SNP 在亲子推断上具有更大的应用前景。

### 4 结论

本研究利用 255 个高多态性 SNPs 标记实现了对中国荷斯坦牛亲子关系的高效推断。以分布于 15 个牛场的 1 927 头北京地区中国荷斯坦母牛为试验对象,平均系谱错误率估计值为 20.9%;场、母牛出生年份和月份间系谱错误率有显著差别,反映出

不同的牛场管理水平和对系谱登记工作重视程度的差异,说明必须在日常管理中采取鼓励、监督措施,尤其是产犊高峰期,应加大检查和监督力度,降低系谱错误。同时以遗传标记检测等手段,建立监督、校正系谱错误的检测系统,是我国奶牛生产和育种实践中亟需解决的问题之一,是加强监管、完善系谱登记制度和繁殖指标考核制度的技术保证,也是保证后裔测定公牛后代系谱准确的有力措施,可提高遗传评估的准确性,进而加快我国奶牛遗传改良进展。

#### 参考文献:

- [ 1 ] BANOS G, WIGGANS G R, POWELL R L. Impact of paternity errors in cow identification on genetic evaluations and international comparisons [J]. *J Dairy Sci*, 2001, 84: 2523-2529.
- [ 2 ] 汪 湛, 田雨泽, 刘和风. 应用血型分析技术对奶牛亲子关系正确率的调查初报 [J]. *中国畜牧兽医*, 2005, 3: 22-23.
- [ 3 ] 初 芹, 张 毅, 孙东晓, 等. 从 17 个微卫星标记错配分析影响中国荷斯坦牛系谱错误的因素 [J]. *畜牧兽医学报*, 2011, 42(2): 163-168.
- [ 4 ] HEATON M P, HARHAY G P, BENNETT G L, et al. Selection and use of SNP markers for animal identification and paternity analysis in US beef cattle [J]. *Mamm Genome*, 2002, 13: 272-281.
- [ 5 ] WERNER F A O, DURSTEWITZ G, HABERMANN F A, et al. Detection and characterization of SNPs useful for identity control and parentage testing in major European dairy breeds [J]. *Anim Genet*, 2004, 35: 44-49.
- [ 6 ] FISHER P J, MALTHUS B, WALKER M C, et al. The number of single nucleotide polymorphisms and on-farm data required for whole-herd parentage testing in dairy cattle herds [J]. *J Dairy Sci*, 2009, 92: 369-374.
- [ 7 ] MARSHALL T C, SLATE J, KRUIK L E B, et al. Statistical confidence for likelihood-based paternity inference in natural populations [J]. *Mol Ecol*, 1998, 7: 639-655.
- [ 8 ] KALINOWSKI S T, TAPER M L, MARSHALL T C. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment [J]. *Mol Ecol*, 2007, 16: 1099-1106.
- [ 9 ] ANDERSON E C, GARZA J C. The power of single-nucleotide polymorphisms for large-scale parentage inference [J]. *Genetics*, 2006, 172: 2567-2582.
- [ 10 ] 周 磊. 利用微卫星和 SNP 标记进行奶牛亲子鉴定模拟研究 [M]. 北京: 中国农业大学, 2010.
- [ 11 ] WELLER J I, FELDMESSER E, GOLIK M, et al. Factors affecting incorrect paternity assignment in the Israeli Holstein population [J]. *J Dairy Sci*, 2004, 87(8): 2627-2640.
- [ 12 ] VISSCHER M P, WOOLLIAMS J A, SMITH D, et al. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection [J]. *J Dairy Sci*, 2002, 85: 2368-2375.

(编辑 郭云雁)