

## 3 种不同利用微卫星标记进行个体识别方法的比较研究

毛永江,常 洪\*,杨章平,常国斌,刘桂琼  
(扬州大学动物科学与技术学院,扬州 225009)

**摘要:**以鲁西黄牛和渤海黑牛(各 30 头)12 个微卫星座位数据为基础,利用系统树分析法、最大似然法及 Bayes 方法,比较不同方法对个体识别效率的影响。结果表明:当利用 12 个微卫星座位和 6 个低杂合度位点数据时,3 种个体识别方法鉴定的准确性均为 100%。但利用 6 个高杂合度座位时,Bayes 方法的准确性最高(83.33%),最大似然法次之(76.67%),系统树分析法最低(71.67%)。结果表明,基于 Bayes 理论的个体识别方法最准确,是进行同类研究的首选工具,而系统树分析法和最大似然法可以作为其补充加以应用。

**关键词:**鲁西黄牛;渤海黑牛;微卫星;个体识别

中图分类号:S823.2

文献标识码:A

文章编号:0366-6964(2008)09-1171-05

### Comparison between 3 Individual Identification Methods Using Microsatellite Markers

MAO Yong-jiang, CHANG Hong\*, YANG Zhang-ping, CHANG Guo-bin, LIU Gui-qiong  
(College of Animal Science and Technology, Yangzhou University, Yangzhou 225009, China)

**Abstract:** 30 Luxi and 30 Bohai Black individuals were genotyped with 12 microsatellites. The efficiencies of individual identification among phylogenetic tree analysis, maximum likelihood and Bayes approach were compared. The results showed that the individual identification accuracy rates were 100% for 3 methods using 12 microsatellites and 6 low heterozygosity microsatellites. By using 6 high heterozygosity microsatellites, the accuracy rates were 83.33%, 76.67% and 71.67% for Bayes approach, maximum likelihood and phylogenetic tree analysis, respectively. All of these indicated that Bayes approach maybe the most accuracy and should be the first choice in individual identification research, and the maximum likelihood and phylogenetic tree analysis could be the complementary methods.

**Key words:** Luxi cattle; Bohai Black cattle; microsatellite; individual identification

动物个体或群体识别是根据个体或群体的某些特征判断其来源、归属及它们之间亲缘关系等方面的一种方法,在生物学研究中有着极其重要的意义。在家畜育种及遗传资源保护中,个体或群体识别对于判断动物个体或群体品种来源、品种间亲缘关系、品种的杂交育成史等方面都着重要意义。传统的动物个体或群体识别是根据其外貌特征、染色体特征及血液蛋白多态等方面来进行的。但由于这些方法

受到自身各方面特征的限制,阻碍了它们的进一步发展,特别是用外形相似、染色体特征相同、亲缘关系较近的动物进行动物个体或群体识别时效率很低。近年来,各种分子标记的发展为动物个体或群体识别提供了新的活力,其中微卫星标记由于其多态程度高、分布广泛、且在近缘物种间具有一定程度的同源性,受到广大学者的偏爱<sup>[1]</sup>。目前,微卫星标记已在家牛、水牛、马、狗、鸡等家畜禽个体或群体识

收稿日期:2007-09-05

基金项目:国家“863”计划(2002AA242011);国家自然科学基金(30571323);扬州大学高层次人才科研启动基金资助项目

作者简介:毛永江(1974-),男,贵州江口人,讲师,博士,主要从事动物遗传资源研究,E-mail: maoyongjiang@gmail.com

\* 通讯作者:常 洪,教授,博导,Tel:86-514-87997203, E-mail: hoch@yzcn.net

别中得到应用<sup>[2-8]</sup>。

根据分子标记进行动物个体或群体识别的方法,目前主要有系统树分析法、最大似然法及最近发展起来的 Bayes 方法<sup>[9]</sup>。系统树分析法是根据动物个体或群体间的遗传距离,采用某种方法构建系统树,再根据系统树的聚类情况进行判断和评价。最大似然法是根据动物个体特定标记的基因型,采用最大似然法计算每个个体属于各自采样群体的最大似然概率值,从而判断个体的归属。而 Bayes 方法是根据动物个体或群体的基因型组成,运用 Bayes 概率理论,从而判断个体或群体属于各自采样群体的概率。目前,这 3 种方法都有相应的理论基础和软件,在使用过程中极为方便,并且得到了广泛应用<sup>[10-12]</sup>。目前对利用这 3 种方法进行动物个体或群体识别的效率还没有进行系统研究,仅 Behl 等<sup>[5]</sup>对系统树分析法和最大似然法进行了简单比较,而其它更深层次的研究尚未见系统报道。

鲁西黄牛和渤海黑牛是我国黄河中下游流域黄牛的两个代表,同时也是列入国家级畜禽品种资源保护名录的两个品种<sup>[13]</sup>。鲁西黄牛被毛呈浅黄色,以肉质好、体大力强而著称,而渤海黑牛以其全身(被毛、蹄角及鼻镜)黑色,肉质细嫩,商品活牛在香港市场颇受欢迎。从外形上看,这两个品种差异明显,在采样时不存在品种混杂的问题,适合于动物个体或群体识别方法的比较研究。因此,本研究拟以鲁西黄牛和渤海黑牛 12 个微卫星座位数据为基础,对 3 种动物个体或群体识别方法的效率进行比较,同时探索微卫星座位数和杂合度水平对个体或群体识别的影响,以期对动物个体或群体识别提供参考依据。

## 1 材料与方法

### 1.1 采样及 DNA 提取

用中心产区典型群随机抽样于山东省鄆城县鲁西黄牛原种场和梁山县科龙畜牧有限公司采集鲁西黄牛共 30 头,同时于山东省滨州市渤海黑牛原种场采集渤海黑牛 30 头。从颈静脉或尾静脉采血 20 mL,ACD 抗凝。在实验室取血样 10 mL,按常规方法提取基因组 DNA,并用紫外分光光度计测定其浓度<sup>[14]</sup>。

### 1.2 微卫星选择及其 PCR 扩增

从联合国粮农组织(FAO)<sup>[15]</sup>推荐的用于牛亚科家畜遗传多样性研究的 30 对微卫星引物,结合

Ritz 等进行牛亚科家畜遗传多样性研究所使用的引物<sup>[16]</sup>,从中选择效果较好、杂合度在中度水平以上的 12 对:ETH3、ETH-225、BM-1824、AGLA-293、TGLA-73、TGLA-227、TGLA-122、IDVGA-11、IDVGA-2、IDVGA-55、CSSM033、CSSM043。PCR 反应及基因型判定详见参考文献<sup>[17]</sup>。

### 1.3 统计处理

群体遗传学一般统计学指标(如等位基因频率、基因型频率、期望杂合度、平均有效等位基因数等)采用 Popgene(Ver 1.32)完成。采用 3 种不同的方法进行个体识别,即:系统树分析、最大似然法和 Bayes 分析法。

1.3.1 系统树分析 根据 2 个黄牛群体 60 个个体 12 个微卫星座位的基因型,采用 DC 遗传距离和邻近给合法(Neighbor-joining, NJ)构建个体间的亲缘系统树。个体亲缘系统树构建由 Populations (1.2.28)和 Treeview 完成<sup>[18-19]</sup>。根据个体系统树的聚类情况,判定其正确性及其效率。

1.3.2 最大似然法 根据采样群体各个体 12 个微卫星座位的基因型,采用最大似然法计算每个个体属于各自采样群体的最大似然概率值(The Log of Odds, LOD)。为提高个体识别的准确性,本研究假定只有属于两个采样群体的最大似然概率值的比值大于等于 2 才视为正确。这样,即可在 99%的概率水平上认为该个体的判定是正确的。最大似然比值由软件 WHICHRUN 4.1 完成<sup>[12]</sup>。

1.3.3 Bayes 分析法 为验证采样的准确性,运用贝叶斯概率理论(Bayesian theory),使用 GENEALSS (V2.0)对所采样的每个个体所属群体进行判定。GENEALSS 分析采用 Rannala & Mountain 所用的方法<sup>[9]</sup>,判定准确性水平设定为 1%。

为验证微卫星座位数和杂合度水平对个体识别效率的影响,在利用 12 个微卫星座位的基础上,根据每个座位在总群体中(即把鲁西黄牛和渤海黑牛合并为一个群体)的期望杂合度水平(表 1),分为高杂合度组和低杂合度组,每组 6 个座位。再利用以上 3 种方法对所有个体进行判定。

## 2 结果

### 2.1 系统树分析法

2 个群体 12 个微卫星座位个体聚类图见图 1。由图 1 可看出,鲁西黄牛和渤海黑牛可分为两个明显的分支,且所有采样个体均聚到各自的群体中,这

表 1 不同组微卫星座位的期望杂合度  
Table 1 The expected heterozygosity of microsatellites for different groups

组别 Group	引物 Primer	鲁西黄牛 Luxi	渤海黑牛 Bohai Black	总群体 Total populations
6 对高杂合度引物 6 high <i>He</i> microsatellites	TGLA-227	0.765 0	0.910 7	0.873 8
	ETH-225	0.836 2	0.783 1	0.865 3
	IDVGA-2	0.843 5	0.850 3	0.858 7
	CSSM033	0.830 5	0.837 9	0.846 2
	AGLA-293	0.783 1	0.644 1	0.792 2
	TGLA-122	0.807 3	0.711 3	0.789 8
	Mean±SD	0.810 9±0.031 6	0.789 6±0.098 0	0.837 7±0.037 3
6 对低杂合度引物 6 low <i>He</i> microsatellites	CSSM043	0.589 3	0.444 6	0.738 5
	BM1824	0.269 5	0.587 6	0.716 7
	TGLA-73	0.157 6	0.235 0	0.594 3
	IDVGA-55	0.487 6	0.507 9	0.550 0
	ETH-3	0.494 4	0.453 7	0.488 0
	IDVGA-11	0.066 1	0.662 1	0.431 4
	Mean±SD	0.344 1±0.210 2	0.481 8±0.146 6	0.586 5±0.122 6
12 对引物的均值 Mean of 12 pairs microsatellites	Mean±SD	0.577 5±0.282 8	0.635 7±0.199 9	0.712 1±0.157 1

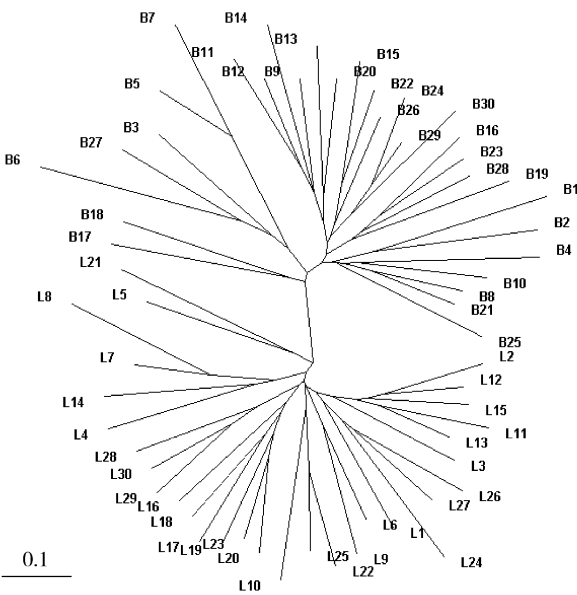


图 1 12 对微卫星引物 DC 遗传距离 NJ 个体聚类图  
Fig. 1 The radial individual NJ-tree based on DC distances from 12 microsatellites

说明根据 12 个微卫星座位数据,可以对所采样的所有个体进行个体识别。当利用 6 个高杂合度座位时,所得个体聚类图也可分为明显的 2 个分支,但 2 群体中有 17 个个体出现了交叉的情况,也就是说,只有 71.67%的个体识别是正确的。当利用 6 个低

杂合度座位时,所得个体聚类图也可分为明显的 2 个分支(图略),但不如利用 12 个微卫星座位明显,且所有采样个体均聚到各自的群体中,说明其识别正确性为 100%。

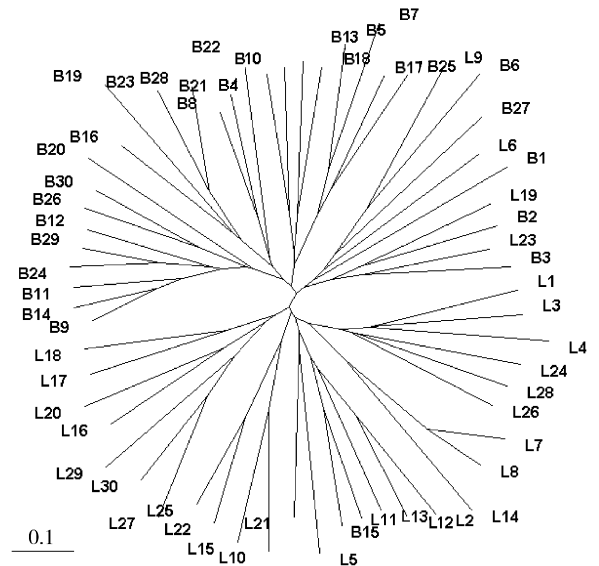


图 2 6 对高杂合度引物 DC 遗传距离 NJ 个体聚类图  
Fig. 2 The radial individual NJ-tree based on DC distances from 6 high *He* microsatellites

### 2.2 最大似然法

当利用 12 个微卫星座位数据时,各个体属于两

个采样群体的最大似然概率值的比值均大于 2, 最大值为 14。这说明在 99% 的概率水平上认为所有采样个体的归属判定是正确的, 均分属于各自的采样群体。当利用 6 个高杂合度座位时, 2 个群体共有 14 个个体(其中鲁西黄牛 8 个, 渤海黑牛 6 个)属于各自采样群体的最大似然概率值的比值均小于 2。当利用 6 个低杂合度座位时, 所有采样个体属于各自采样群体的最大似然概率值的比值均大于 2, 但其值稍低于利用 12 对微卫星引物数据时的比值。这说明利用 6 个低杂合度座位也能在 99% 的概率水平上正确判定所有采样个体的归属问题, 只是其准确性稍低于利用 12 对微卫星引物的数据。

### 2.3 Bayes 分析法

当利用 12 个微卫星座位数据时, 所有个体均正确地判定为各自的采样群体, 鉴定准确性为 100%。当利用 6 个高杂合度座位时, 2 群体有 50 个个体(83.33%)正确地判定为各自的采样群体, 仍有 10 个个体(占 16.67%, 其中鲁西黄牛 6 个, 渤海黑牛 4 个)判定概率低于 99% 或者判定不正确。当利用 6 个低杂合度座位时, 其判定结果与利用 12 个微卫星座位数据基本一致。同时, 利用 3 组不同微卫星组合数据在群体水平上进行判定, 其鉴定准确性均为 100%。

### 2.4 3 种个体识别方法的比较

比较 3 种个体识别方法, 可以发现: 当利用 12 个微卫星座位和 6 个低杂合度座位数据时, 3 种个体识别方法鉴定的准确性均为 100%。但利用 6 个高杂合度座位时, Bayes 方法的准确性最高(83.33%), 最大似然法次之(76.67%), 系统树分析法最低(71.67%) (图 3)。由此可见, 微卫星杂合度水平和座位数量对个体识别准确性有一定程度的影响。

## 3 讨论

### 3.1 3 种个体识别方法的比较

就 3 种方法的结果而言, 系统树分析法可利用聚类图形象地展示个体及群体的聚类情况及它们之间的亲缘关系, 但这种关系很难“量化”, 只能是一个定性的结果。相比之下, 最大似然法和 Bayes 方法能计算出每个采样个体在一定置信水平下属于各自采样群体的概率值, 基于 Bayes 方法的软件还能估计每个群体属于各自采样群体的概率值, 给出了量化的指标。3 种方法个体识别效率, 在相同条件下, Bayes 方法最高, 而系统树分析法最低。此外, 系统

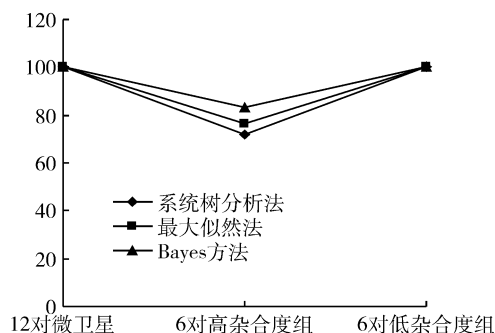


图 3 3 种个体识别方法效率的比较

Fig. 3 The comparison of efficiency for 3 individual identification methods

树分析法还有可能由于采用的遗传距离及聚类方法的不同而产生不同的聚类结果, 其结果在一定程度上带有“不确定性”。有关遗传距离及聚类方法对系统树构建的影响, Takezaki 等<sup>[20]</sup>对该问题进行了详细地研究, 在此不再详述。所以, 相比较而言, 作者认为 Bayes 方法在动物个体识别中最准确, 是进行同类研究的首选工具, 而系统树分析法和最大似然法可以作为其补充加以应用。

### 3.2 微卫星座位数和杂合度水平对个体识别准确性的影响

微卫星座位数对个体识别准确性有显著影响, 这已被类似的所有研究所证实<sup>[6]</sup>。在本研究所使用的 3 种方法中, 利用 12 个微卫星座位数据个体识别准确性均高于利用 6 个微卫星座位数据。可见, 座位数越多, 识别的准确性越高。在研究或解决实际问题时, 在特定置信水平和个体识别效率前提下, 应该采用的适宜的(或是最低的)微卫星座位数量。Bjornstad 等<sup>[6]</sup>在利用微卫星对 8 个马群体进行个体识别时, 采用 13 对微卫星引物, 其识别的准确性能达到 95% 以上。而利用微卫星在近缘间进行个体识别时, 使用不到 10 个微卫星座位也能使准确性达到 90%~95% 以上<sup>[21]</sup>。可见, 个体识别时适宜的(或是最低的)微卫星座位数量受到待识别个体及群体亲缘关系及遗传分化程度的影响。亲缘关系越近, 遗传分化程度越低, 在特定置信水平和个体识别效率前提下所需要的最低微卫星数量越多, 反之则越少。

微卫星座位杂合度水平对个体识别准确性也有一定程度的影响。在研究所使用的 3 种方法中, 6 个低杂合度座位个体识别的准确性均高于 6 个高杂合度座位, 但这并不能说明微卫星座位杂合度水平

越高,其个体识别的准确性越低。而是在特定置信水平和个体识别效率前提下,使用高杂合度座位可能比低杂合度座位需要更多的座位。Bjornstad 等<sup>[6]</sup>的研究表明:微卫星用于个体识别的误判率与观察等位基因数呈极显著负相关( $r = -0.47, P < 0.01$ ),而与座位杂合度水平没有显著相关关系( $r = -0.23, P = 0.26$ )。但 Blott 等<sup>[21]</sup>、Rosenberg 等<sup>[8]</sup>指出,拥有较多观察等位基因数和较高杂合度的微卫星座位,其进行个体及群体识别的效率也较高。可见,进行个体及群体识别时,微卫星座位的杂合度必须维持一定的水平。

### 参考文献:

- [1] BRUFORD M W, WAYNE P K. Microsatellite and their applications to population genetic studies [J]. *Current Opinion in Genetics and Development*, 1993, 3:939-943.
- [2] FREEMAN A R, MEGHEN C M, MACHUGH D E, et al. Admixture and diversity in West African cattle populations [J]. *Molecular Ecology*, 2004, 13: 3 477-3 487.
- [3] KUMAR P, FREEMAN A R, LOFTUS R T, et al. Admixture analysis of south Asian cattle [J]. *Heredity*, 2003, 91: 43-50.
- [4] KRUGER K, GAILLARD C, STRANZINGER G, et al. phylogenetic analysis and species allocation of individual equids using microsatellite data [J]. *Journal of Animal Breeding and Genetics*, 2005, 122 (Suppl. 1):78-86.
- [5] BEHL R, BEHL J, GUPTA N, et al. Individual identification and breed allocation with microsatellite markers: an evaluation in Indian horses [J]. *Asian Aust J Anim Sci*, 2007, 20 (1):25-30.
- [6] BJORNSTAD G, ROED K H. Evaluation of factors affecting individual assignment precision using microsatellite data from horse breeds and simulated breed crosses [J]. *Animal Genetics*, 2002, 33 (5): 262-270.
- [7] KOSKINEN M T. Individual assignment using microsatellite DNA reveals unambiguous breed identification in the domestic dog [J]. *Animal Genetics*, 2003, 34:297-301.
- [8] ROSENBERG N A, BARKE T, ELO K, et al. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds [J]. *Genetics*, 2001, 159: 699-713.
- [9] RANNALA B, MOUNTAIN J L. Detecting immigration by using multilocus genotypes [J]. *Proceeding of the National Academy of Sciences of the USA*, 1997, 94: 9 197-9 221.
- [10] CAVALLI-SFORZA L L, EDWARDS A W F. Phylogenetic analysis: Models and estimation procedures [J]. *American Journal of Human Genetics*, 1967, 19: 233-257.
- [11] NEI M, KUMAR S. *Molecular evolution and phylogenetics* [M]. New York: Oxford University Press, 2000.
- [12] BANKS M A, EICHERT W. WHICHRN, Version 4.1, a computer program for population assignment based on multilocus genotype data. Bodega Marine Laboratory, University of California at Davis, Bodega Bay, CA, USA. 2001.
- [13] 中国畜禽遗传资源状况编委会. 中国畜禽遗传资源状况 [M]. 北京:中国农业出版社, 2004.
- [14] 萨姆布鲁克 J, 费里奇 E F, 曼尼阿蒂斯 T. 分子克隆实验指南 [M]. 金冬雁, 黎孟枫, 等译. 第 2 版. 北京:科学出版社, 1999:483-486.
- [15] FAO. Measurement of domestic animal diversity (MoDAD): recommended microsatellite markers, new microsatellite sets-recommendations of joint JSAG/FAO Standing Committee [M]. Rome, 2004.
- [16] RITZ L R, GLOWATZKI M L, MACHUGH D E, et al. Phylogenetic analysis of the tribe Bovini using microsatellite [J]. *Animal Genetics*, 2000, 31:178-185.
- [17] 毛永江, 常洪, 杨章平, 等. 3 个中国地方黄牛品种遗传结构及其遗传分化的研究 [J]. *畜牧兽医学报*, 2007, 38(2):125-132.
- [18] OLIVIER L. Populations(1.2.28). URL: www.cnrs-gif.fr/pge/bioinfo/populations.
- [19] PAGE R D M. TREEVIEW: An application to display phylogenetic trees on personal computer [J]. *Computer Applications in the Biosciences*, 1996, 12: 357-358.
- [20] TAKEZAKI N, NEI M. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA [J]. *Genetics*, 1996, 144: 389-399.
- [21] BLOTT S C, MIDTHJELL L, HALEY C S. Discriminating among cattle breeds using genetic markers [J]. *Heredity*, 1999, 82:613-619.