

马敏彪, 武利恒, 丁琳, 等. 基于 SNP 分子标记的湖羊群体遗传结构分析 [J]. 畜牧与兽医, 2024, 56 (6): 1-8.

MA M B, WU L H, DING L, et al. Analysis of the genetic structure of Hu sheep population based on SNP molecular markers [J]. Animal Husbandry & Veterinary Medicine, 2024, 56 (6): 1-8.

基于 SNP 分子标记的湖羊群体遗传结构分析

马敏彪¹, 武利恒², 丁琳¹, 金慧佳³, 赵丽莉⁴, 曹丁壬¹, 王争光^{3*}

(1. 浙江省畜牧技术推广与种畜禽监测总站, 浙江 杭州 310021;

2. 浙江华欣牧业有限公司, 浙江 江山 324100;

3. 浙江大学动物科学学院, 浙江 杭州 310058;

4. 衢州市畜牧业发展中心, 浙江 衢州 324000)

摘要: 旨在运用现代生物信息学技术, 评估湖羊群体的遗传结构, 构建详细的系谱记录。通过对 40 只湖羊种公羊进行单核苷酸多态性 (SNP) 位点检测, 基于检测结果进行遗传结构分析、主成分分析和进化树构建。结果: 在 40 只种公羊中, 共检测到 952 214 个 SNPs; 平均核苷酸多样性在 0.212 9~0.242 5 之间, 平均观测杂合度大于平均期望杂合度, 遗传分化程度值在 0.370 2~0.452 4 之间; 状态同源 (IBS) 遗传距离在 0.173 9~0.228 0 之间, 平均为 0.220 9; 主成分分析和系统进化树的结果一致, 最终可以将 40 只种公羊分为 14 个家系, 湖羊群体中大部分个体间的亲缘关系较远, 具有较高的遗传多样性和选择潜力。本研究在全基因组水平上, 分析了湖羊种公羊的遗传结构和亲缘关系, 完善了系谱记录, 为湖羊场后续育种、配种等提供了科学依据。

关键词: 湖羊; 单核苷酸多态性; 遗传多样性; 遗传结构; 亲缘关系

中图分类号: S826 文献标志码: A 文章编号: 0529-5130(2024)06-0001-08

Analysis of the genetic structure of Hu sheep population based on SNP molecular markers

MA Minbiao¹, WU Liheng², DING Lin¹, JIN Huijia³, ZHAO Lili⁴, CAO Dingren¹, WANG Zhengguang^{3*}

(1. Zhejiang Province Animal Husbandry Technology Extension and Livestock and Poultry Monitoring Station, Hangzhou 310021, China;

2. Zhejiang Huaxin Animal Husbandry Co., Ltd., Jiangshan 324100, China;

3. College of Animal Sciences, Zhejiang University, Hangzhou 310058, China;

4. Quzhou Animal Husbandry Development Center, Quzhou 324000, China)

Abstract: Modern bioinformatics technology was used in this study to evaluate the genetic structure and to construct detailed pedigree records of Hu rams. Forty breeding rams were used and SNP site detection was performed on them. Then, genetic structure analysis, principal component analysis, and evolutionary tree construction were conducted in the Hu sheep. The results showed that a total of 952 214 SNPs were detected on 40 rams. The average nucleotide diversity of these sites was between 0.212 9 and 0.242 5. Their average observed heterozygosity (H_o) was greater than the average expected heterozygosity (H_e). The degree of their genetic differentiation was between 0.370 2 and 0.452 4. The genetic distance of their IBS was between 0.173 9 and 0.228 0, with an average of 0.220 9. The result of the principal component analysis was consistent with that of the phylogenetic tree. Taken together, the 40 rams could be divided into 14 families. The kinship between most of the individuals was distant. The present results indicated that Hu rams possessed high genetic diversity and selection potential, which contributed to a more complete pedigree record of Hu rams and provided scientific support for breeding and mating of Hu Sheep on farms.

Keywords: Hu sheep; SNP; genetic diversity; genetic structure; kinship

收稿日期: 2024-01-19; 修回日期: 2024-04-25

基金项目: 衢州市科技局项目 (2023T019); 浙江省农业重大技术协同推广计划项目 (2023ZDXT16); 浙江省科技厅项目 (2023C04004); 浙江省畜禽新品种选育重大科技专项 (2021C02068-6); 国家自然科学基金项目 (32172724)

第一作者: 马敏彪, 男, 畜牧师

* 通信作者: 王争光, 教授, 研究方向为畜禽育种与繁殖学, E-mail: wzhuang68@zju.edu.cn.

湖羊是列入《国家畜禽遗传资源保护名录》的多胎绵羊品种,在我国羊产业集约化发展中具有十分重要的地位。为了获得更大的养殖经济效益,部分羊场对湖羊进行了杂交改良,促进了湖羊肉用性能的提高,但也破坏了传统养殖区内湖羊的纯种优势,导致湖羊种质退化,系谱记录混乱^[1]。因此亟需通过多种有效手段对湖羊群体进行遗传结构分析和亲缘鉴定,完善种羊场的系谱记录,为湖羊种质资源的利用、保护和开发提供理论依据。

单核苷酸多态性(single nucleotide polymorphism, SNP)因具有数量多、分布广、遗传稳定性强、易大规模快速检测等特点^[2],被广泛应用于动物遗传育种研究,在标记辅助选择、动物遗传图谱构建、亲缘关系鉴定、品种溯源等领域具有重要的应用价值^[3-5]。Chessari等^[6]对48只Noticiana绵羊进行SNP检测,发现Noticiana绵羊与其他意大利绵羊的亲缘关系较远,推测与基因漂移、小种群规模和生殖隔离等因素有关。张任豹等^[7]对40只成年济宁青山羊进行了全基因组SNP检测和综合信息分析,成功将23只公羊划分为14个家系,在分子水平上完善了公羊的亲缘关系。史露露等^[8]对中国夏洛来羊进行了SNP检测,根据遗传距离的远近,将75只中国夏洛来羊划分为5个家系,并得出中国夏洛来羊和法国夏洛来羊之间已产生一定程度遗传分化的结论。

因此,为摸清湖羊场不同群体种公羊的遗传结构,提高种公羊的利用效率和养殖场的经济效益,本研究在全基因组水平上,利用SNP分子标记,对40只湖羊种公羊进行遗传多样性分析和家系划分,以完善湖羊场系谱记录,为湖羊场后续育种、配种等提供了科学依据。

1 材料与方法

1.1 样品采集

本研究采集了来自浙江华欣牧业有限公司湖羊场的40份种公羊血样,耳号分别为00806247、00417、M05098、M05096、00409、00670959、00422、00412、00440、00454、22A1013、M05569、00414、00467、22A1163、22A1115、00405、00451、00408、0403、0477、M05561、M05082、00402、M0555、22A1080、M05094、M03632、00452、M05081、M05084、00491、00447、00453、22A1096、M05563、00418、00424、22A1066、M05851。

1.2 测序

使用QIAGEN Blood Genome DNA Extraction Kit提取血样中的DNA。采用限制性核酸内切酶Dpn II和Bfa I对全基因组DNA进行完全酶切,回收插入片段

大小在200~600 bp之间的酶切片段,采用Double Digest Restriction Site-associated DNA Sequencing的模式建库,基于Illumina HiSeq测序平台,利用第二代测序技术,对文库进行双末端测序。

1.3 数据整理及质量控制

对样品进行上述一系列处理后,得到下机数据。下机数据以双端FASTQ格式保存,采用Illumina 1.8+版本编码,对每个样本的碱基识别准确率进行计算并统计。为了保证后续信息分析质量,采用FastQC对数据进行质量评估,采用fastp(v0.20.0)滑动窗口法对原始数据进行过滤,生成高质量序列^[9]。

1.4 比对分析

采用BWA(0.7.12-r1039)aln程序将过滤后的数据与参考基因组序列进行比对。为提高SNP预测的准确率,采用picard软件包中的“MarkDuplicates”命令,去除重复片段,即PCR Duplicates和Optical Duplicates,采用GATK程序中的“IndelRealigner”命令对所有InDel附近的reads进行重新比对。

1.5 SNP分析

采用stacks(version 2.55)程序包中的“ustacks”命令对单个样本中的reads按照序列相似性进行聚类,其次采用“cstacks”命令将所有样品的loci合并,得到每个loci共有序列,采用“sstacks”命令把各个样品的loci序列与共有序列进行比对,最后采用“populations”命令过滤并筛选出每个样品的SNP位点。采用ANNOVAR^[10]软件对SNP位点进行注释。

1.6 群体遗传多样性分析

基于SNP数据,采用stacks程序包中的“populations”命令,对来自多个群体的样本进行群体遗传多样性分析,计算群体所有位点的平均观测杂合度(observed heterozygosity, Ho)、平均期望杂合度(expected heterozygosity, He)以及群体内两两亚群间的遗传分化程度(用Fst值表示)。

1.7 群体亲缘关系分析

基于SNP数据,采用PLINK(v1.9)计算种公羊两两个体间的状态同源(IBS),得到IBS相似度矩阵,随后生成遗传距离矩阵,衡量样本间的相似性。采用GMatrix(Ver2)计算两两个体间的G值,构建各个样本间的全基因组关系G矩阵,进一步展现种公羊群体的亲缘关系。

1.8 群体家系结构分析

为探究湖羊种公羊的群体结构,直观展现其亲缘关系远近,采用FastTree软件中的“Maximum Likelihood”算法构建系统进化树。采用PLINK(v1.9)软件,利用SNP数据进行主成分分析,计算第一主

成分 (PC1) 和第二主成分 (PC2) 的变异比重。

2 结果

2.1 湖羊群体 SNP 分析

经过数据质控, 采用 stacks 软件, 挖掘 40 个样品的群体 SNP, 质检后一共得到 952 214 个 SNPs, 统

计结果见表 1。变异中群体的转换的碱基数量 (Ts) 多于颠换的碱基数量 (Tv), Ts/Tv 值大于 2, 符合属于同一品种的规律。其中, 以 300 kb 为窗口统计 SNP 的数目, 每条染色体上 SNP 的分布情况见图 1。每条染色体上的 SNP 数量具有较大的差异。

表 1 湖羊群体 SNP 结果统计

总基因型数目	杂合基因型数目	纯合基因型数目	发生转换的碱基数量	发生颠换的碱基数量	转换与颠换的比值
952 214	124 246	175 780	677 705	274 509	2.46

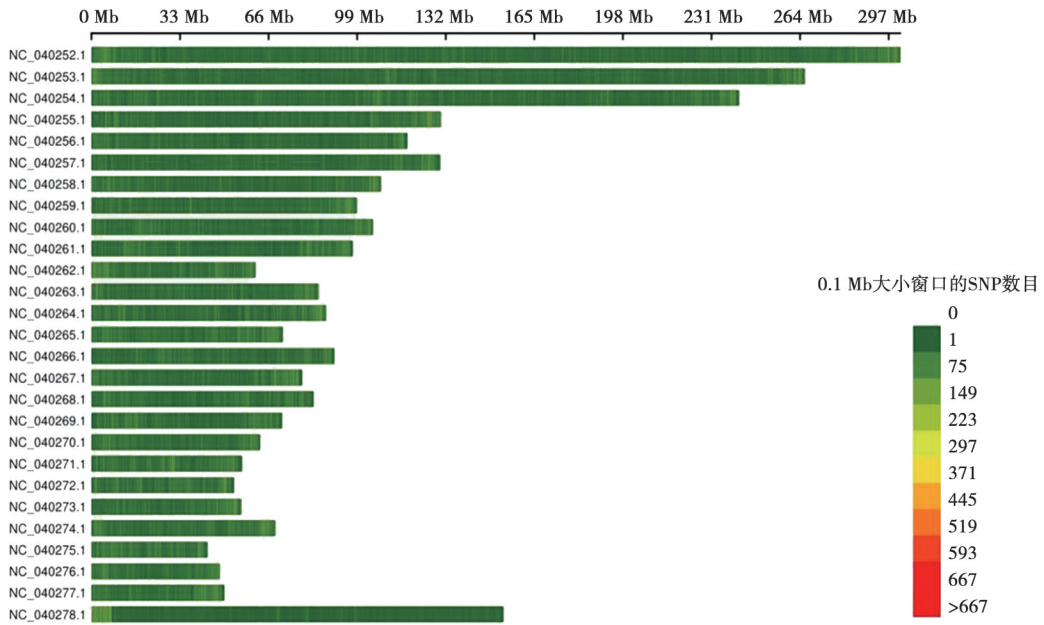


图 1 湖羊群体 SNP 的染色体分布图

2.2 湖羊群体的遗传多样性分析

采用 stacks 程序对来自不同群体的 40 个样本进行遗传多样性分析。各群体平均核苷酸多样性 (Pi) 在 0.212 9~0.242 5 之间, 说明群体内的遗传多样性相对比较丰富。根据哈代温伯格定律估算出各群体的

He, 结果见图 2, 每个群体所有位点的 Ho 均要远远大于 He, 说明各群体内杂合优势更明显。计算群体内两两亚群间的遗传分化程度, Fst 值在 0.370 2~0.452 4 之间, 说明群体间的遗传分化程度比较大。

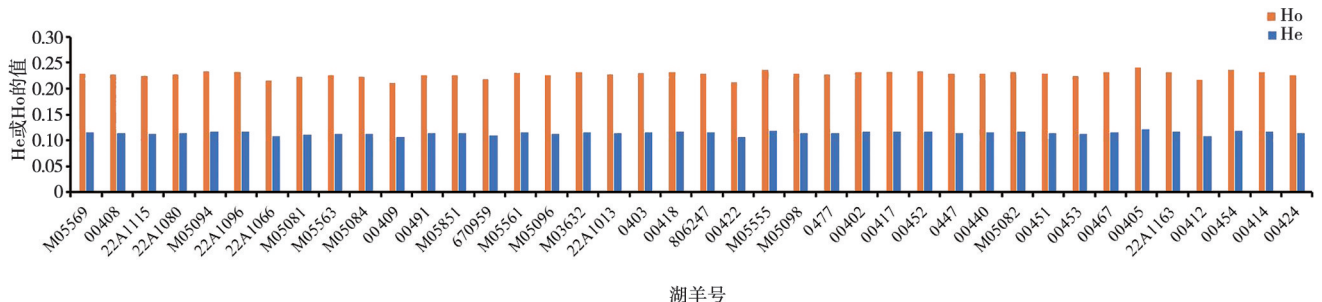


图 2 湖羊群体所有位点的 Ho 与 He

2.3 湖羊群体的状态同源 IBS 矩阵和全基因组关系 G 矩阵分析

根据 SNP 检测结果绘制状态同源 IBS 矩阵，如图 3 所示，每一个方块表示对应横纵坐标的两个样本间的遗传距离。方块的颜色越接近蓝色，说明 IBS 遗传距离越小，关系越近；颜色越红，说明 IBS 遗传距离越大，关系越远。40 只湖羊的 IBS 遗传距离在 0.173 9~0.228 0 之间，平均为 0.220 9，个体间 IBS

遗传距离差异较大，大部分个体间遗传距离较远，只有极少数个体间的遗传距离较近。对于全基因组关系 G 矩阵，如图 4 所示，每一个方块表示对应横纵坐标的两个样本间的 G 值大小。方块的颜色越接近红色，说明值越大，亲缘关系越近；颜色越接近蓝色，说明值越小，亲缘关系越远^[11]。与状态同源 IBS 矩阵显示的结果一致，极大部分个体间的亲缘关系较远，只有极少数个体间的亲缘关系较近。

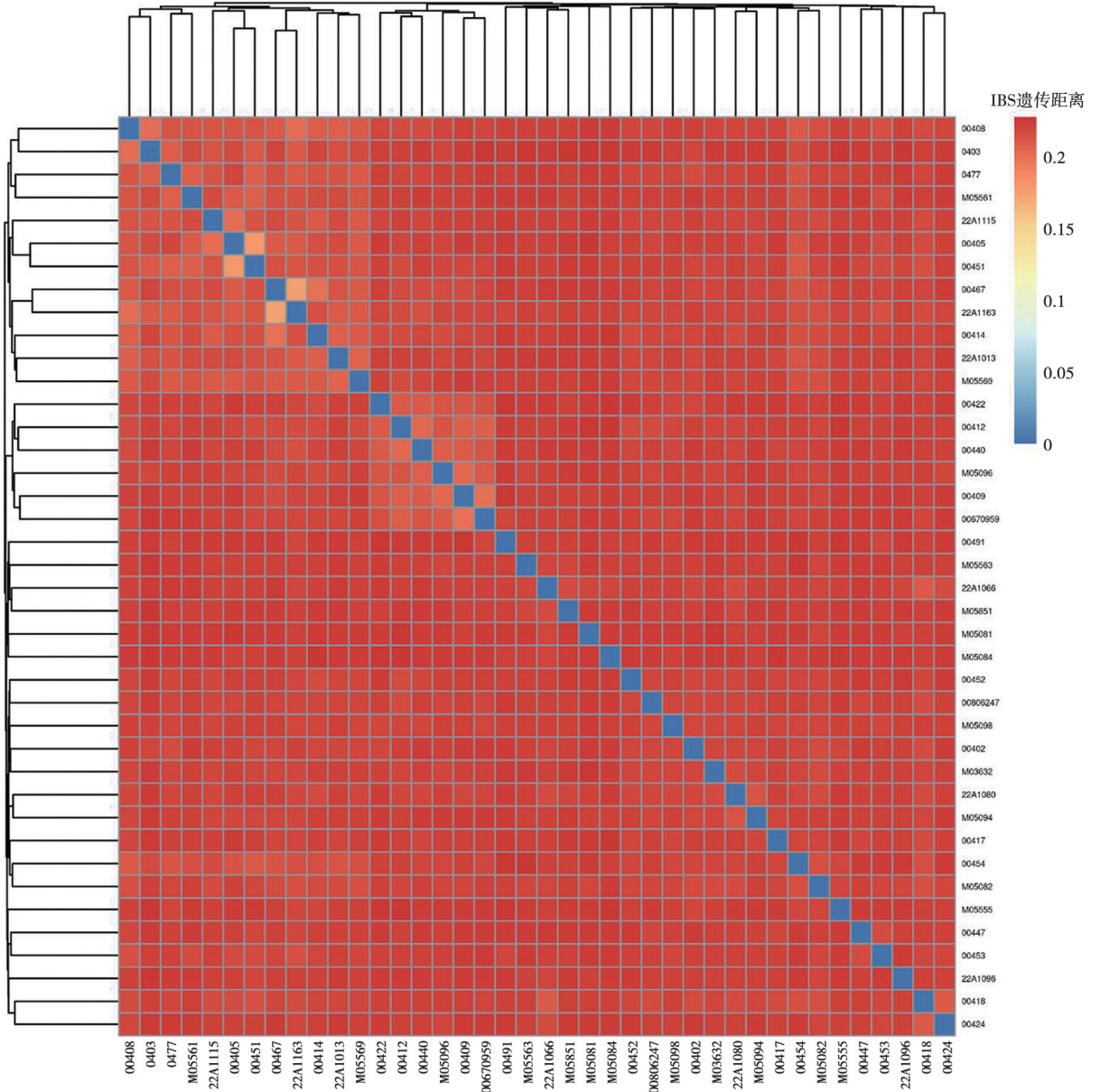


图 3 湖羊群体的 IBS 距离矩阵热图

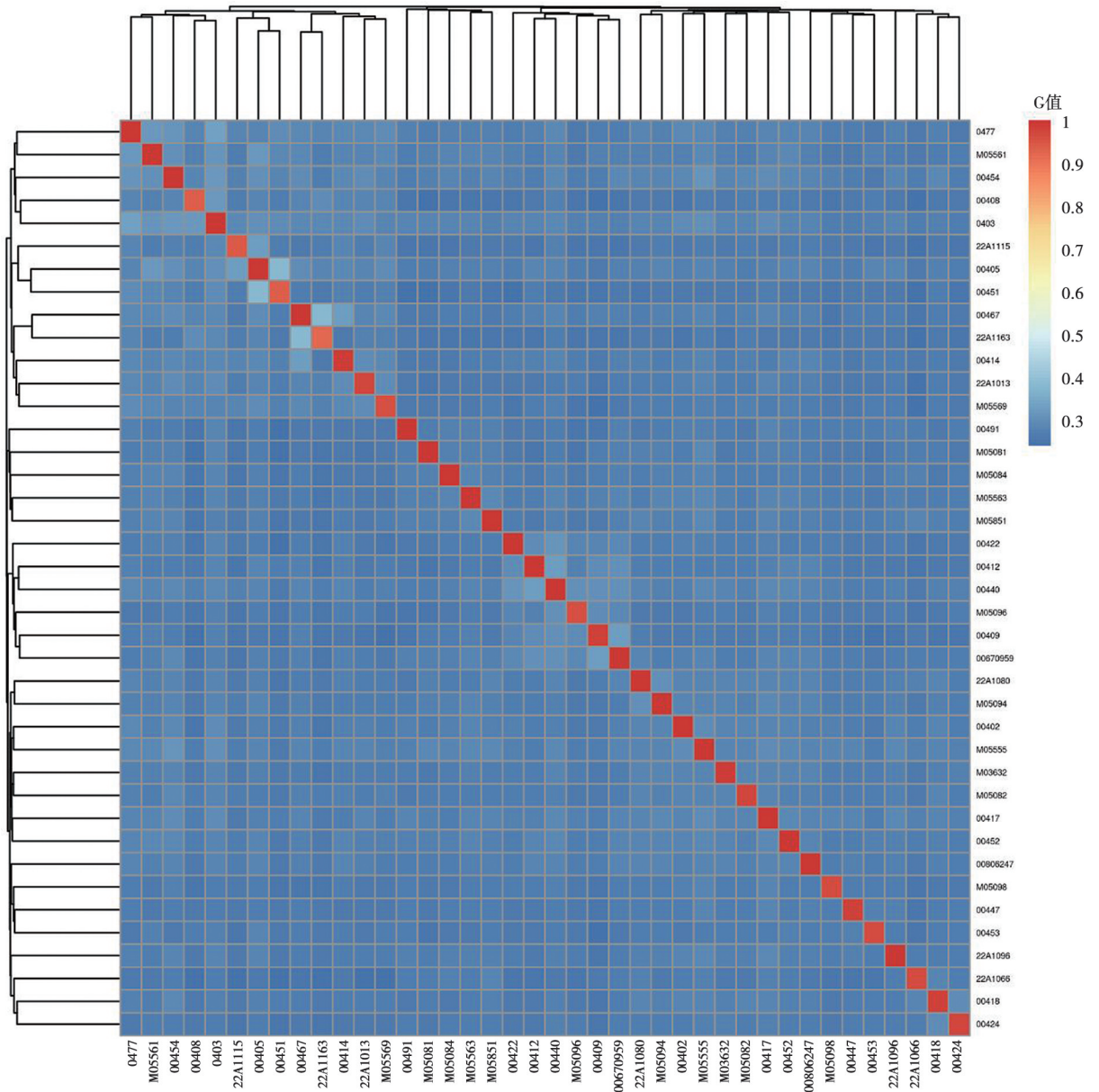


图 4 湖羊群体的 G 矩阵热图

2.4 群体主成分分析

采用 PLINK (v1.9) 软件进行主成分分析, PCA 的分析结果见图 5。PCA 显示, PC1 占变异的 6.5%, PC2 占变异的 6.1%, PC1 与 PC2 的区分度不大。40 只湖羊大致可以聚类成 3 大部分。其中, 00409、

00412、00422、00440、M05096、00670959 较近; 00454、22A1013、M05569、00414、00467、22A1163、22A1115、00405、00451、00408、0403、0477、M05561 较近; 剩余 21 只湖羊较近。

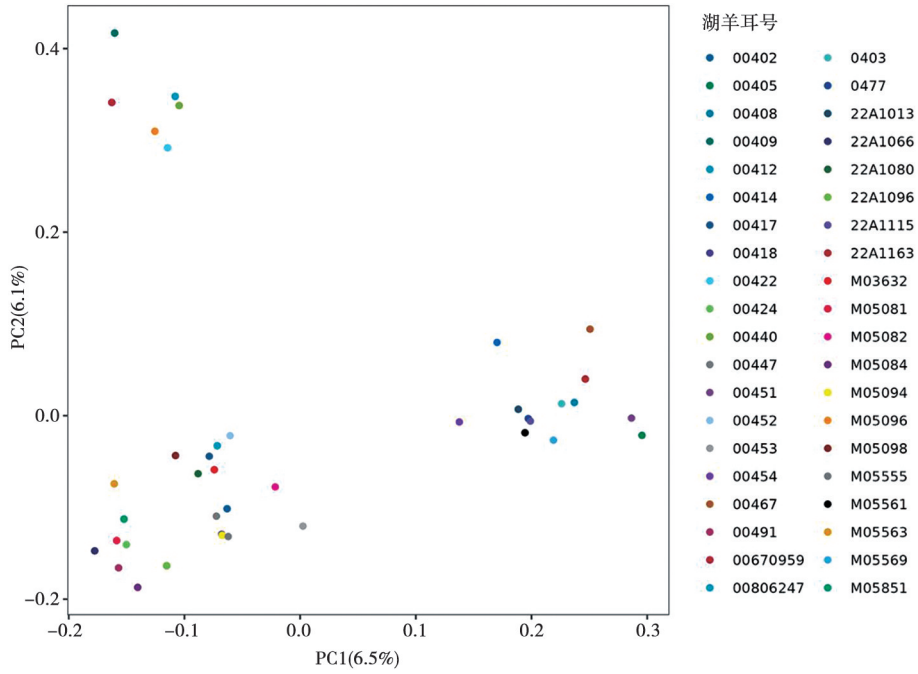


图5 湖羊群体的主成分分析

2.5 群体系统进化树分析

采用 FastTree 软件构建系统进化树，并对分支进行可靠性的验证，结果见图 6。由于样本均为单个一组，进化树样本聚类大体与 PCA 一致，可以分为 3

大部分，但具体个别样本分组与 PCA 有些差异，分别为 M05098、00417 和 00806247。结合状态同源 IBS 矩阵和全基因组关系 G 矩阵，这 40 只种公羊最多可以分为 14 个家系，详细的家系分类见表 2。

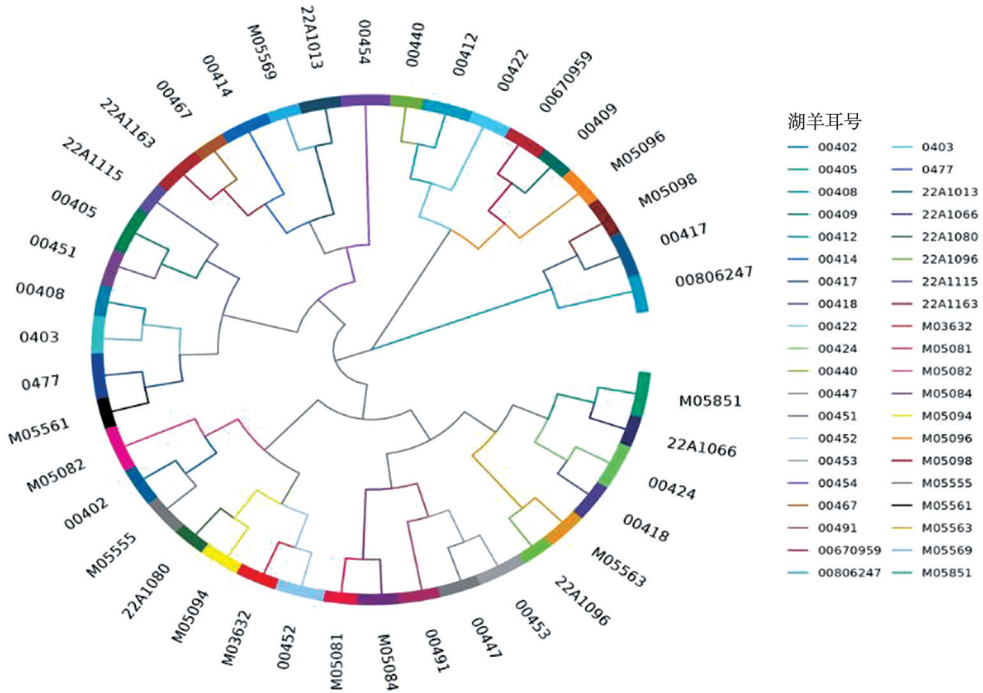


图6 湖羊群体的系统进化树

表2 湖羊种公羊家系分组

家系编号	湖羊耳号
1	00806247/00417/M05098
2	M05096/00409/00670959
3	00422/00412/00440
4	454
5	22A1013/M05569
6	00414/00467/22A1163
7	22A1115/00405/00451
8	00408/0403/0477/M05561
9	M05082/00402/M0555
10	22A1080/M05094/M03632/00452
11	M05081/M05084
12	00491/00447/00453
13	22A1096/M05563
14	00418/00424/22A1066/M05851

3 讨论

近年来, SNP 分子标记以结果稳定可靠等特点, 被广泛应用于生物遗传多样性分析。基于 SNP 分子标记的检测, 通过数据统计和数学模型构建, 借助生物信息技术可以准确计算出每个祖先品种对杂交后代个体基因组的遗传贡献比例, 从而判断亲缘关系的远近^[12]。由此衍生出的 SNP 芯片, 可以大大缩短检测时间, 使检测结果更加准确, 在猪、牛、羊、鸡等物种上已被成熟应用^[13-16]。本研究对 40 只种公羊在全基因组水平上进行 SNP 位点测定, 质检后得到 952 214 个 SNPs, 并进一步对湖羊进行遗传结构分析、主成分分析、进化树构建, 纠正并完善湖羊场系谱记录。

群体的遗传多样性可以用 P_i 、 H_e 、 H_o 、 F_{st} 等参数评估^[17]。本研究中, 各群体的 P_i 在 0.212 9 ~ 0.242 5 之间, 群体内的遗传多样性相对比较丰富。根据哈代温伯格定律得出各群体的 H_e , 均要远远小于对应的 H_o , 群体内更具有杂合的选择优势。群体内两两亚群间的 F_{st} 值均分布在 0.370 2 ~ 0.452 4 之间, 群体间的遗传分化程度比较大。与本研究的结果一致, 熊金珂等^[18]证实湖羊种公羊群体的 H_e (0.382 0) 低于 H_o (0.385 0), 一定程度上说明来自不同地区湖羊场的种公羊表现出较高的遗传多样性, 经过严格的人工选育。

状态同源序列 IBS 是个体间完全相同的 DNA 序列片段。IBS 相似度可以反映个体间相等等位基因的占比, 从而反映个体间的相似性。状态同源序列 IBS

和全基因组关系 G 矩阵是基于 SNP 位点检测的估计个体间亲缘关系的手段之一^[11]。本研究中, 两者的结果都表明大部分个体间的亲缘关系较远, 只有极少数个体间的亲缘关系较近, 说明湖羊场内可能存在近交情况。但是在一个有限的封闭群体中, 不同程度的近交又是不可避免的, 控制近交增量可以有效防止近交衰退^[19]。因此湖羊场可以考虑适当更新血统, 从其他湖羊场引入亲缘关系较远的优良种公羊。

在群体进化中, 进化树结果可以用来展示该物种不同品种之间的进化关系, 能够更加直观地展现亲缘关系远近。对于同一亚群中的不同样品, 在进化树上能很好合成一个分支^[20]。本研究根据 SNP 检测结果构建系统进化树, 40 只湖羊大致分为 3 个部分。此外, 在群体研究中, 通常利用个体基因组 SNP 的差异, 绘制散点图, 将不同个体按照主成分聚类成不同的亚群。本研究中的 40 只湖羊可以明显聚类成 3 个部分, 同一类群遗传距离较近, 这与进化树的结果基本一致, 可以相互印证。基于 SNP 分子标记, 综合多种分析方法, 可以完善种公羊的家系分类。本研究最终将 40 只湖羊分为了 14 个家系。熊金珂等^[18]利用 Illumina Ovine 50K SNP 芯片对湖羊种公羊进行基因分型, 将 248 只种公羊划分为 6 个家系。与之相比, 本研究对湖羊进行了全基因组水平上的 SNP 检测, 得到了更多的 SNP 位点, 获取的信息更加完整, 对湖羊的遗传结构分析更加精确, 湖羊种公羊家系分类也更加细致, 有利于维持湖羊的遗传多样性。

4 结论

该羊场的 40 个湖羊种公羊群体中大部分个体间的亲缘关系较远, 具有较高的遗传多样性和选择潜力, 可以分为 14 个家系。极少数个体间的亲缘关系较近, 可以考虑适当引种。后续, 应做好系谱记录, 进一步完善种公羊的各种表型和基因型数据, 为湖羊场规模化选配提供基础, 实现利益最大化。

参考文献:

- [1] 惠富平, 邵文挥, 孙雁冰. 湖羊繁育史及其当代保种考述 [J]. 畜牧与兽医, 2022, 54 (12): 141-145.
- [2] 苏睿, 林峻, 陈鲤群, 等. 高通量自动化 SNP 检测技术研究进展 [J]. 中国细胞生物学学报, 2019, 41 (7): 1412-1422.
- [3] 王晨阳, 王璐, 张锐虎, 等. SNP 标记在动物遗传育种及人类疾病动物模型研究中的应用 [J]. 中国比较医学杂志, 2019, 29 (4): 120-125.
- [4] 赵杰, 游新勇, 徐贞贞, 等. SNP 检测方法在动物研究中的应用 [J]. 农业工程学报, 2018, 34 (4): 299-305.
- [5] DEMENTIEVA N V, SHCHERBAKOV Y S, TYSHCHENKO V I, et al. Comparative analysis of molecular RFLP and SNP markers in

- assessing and understanding the genetic diversity of various chicken breeds [J]. *Genes (Basel)*, 2022, 13 (10): 1876.
- [6] CHESSARI G, CRISCIONE A, TOLONE M, et al. High-density SNP markers elucidate the genetic divergence and population structure of noticiana sheep breed in the mediterranean context [J]. *Front Vet Sci*, 2023, 10: 1127354.
- [7] 张任豹, 周东辉, 周李生, 等. 基于 70 K SNP 芯片分析济宁青山羊保种群体的遗传结构 [J]. *畜牧兽医学报*, 2023, 54 (7): 2836-2847.
- [8] 史露露, 胡明月, 赖伟宁, 等. 基于 50K SNP 芯片的夏洛来羊遗传结构及选择信号分析 [J]. *黑龙江畜牧兽医*, 2023 (6): 49-53.
- [9] SCHUBERT M, LINDGREEN S, ORLANDO L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging [J]. *BMC Res Notes*, 2016, 9: 88.
- [10] YANG H, WANG K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR [J]. *Nat Protoc*, 2015, 10 (10): 1556-1566.
- [11] 邓俊, 刘艺端, 许文坤, 等. 基于 SNP 芯片撒坝猪保种群体的遗传结构分析 [J]. *中国饲料*, 2022 (17): 7-11.
- [12] 何俊, 钱长嵩, TAIT R G, 等. SNP 芯片数据估计动物个体基因组品种构成的方法及应用 [J]. *遗传*, 2018, 40 (4): 305-314.
- [13] 陈鑫, 孙晓梅, 鲁慧文, 等. 基于 SNP 芯片的猪基因组选择效果分析 [J]. *畜牧与兽医*, 2023, 55 (2): 5-10.
- [14] 戎艳花, 贾雪纯, 李鹏飞, 等. 晋南牛遗传结构特征及选择信号分析 [J]. *中国畜牧兽医*, 2024, 51 (1): 160-171.
- [15] 章晓炜, 刘莉君, 马敏彪, 等. 全基因组水平湖羊不同保种方式的可行性探究 [J]. *黑龙江畜牧兽医*, 2021 (13): 54-57.
- [16] 顾丽红, 刘志勇, 张细权, 等. 180 日龄文昌鸡和隐性白羽克洛鸡基因结构及 SNP 分析 [J]. *畜牧与兽医*, 2019, 51 (3): 7-15.
- [17] 王亦秋, 姜自琴, 王宇哲, 等. 全基因组 SNP 分型策略及基因组选择在畜禽中的应用 [J]. *广东畜牧兽医科技*, 2023, 48 (5): 1-6.
- [18] 熊金珂, 鲍晶晶, 胡文萍, 等. 基于 SNP 芯片的湖羊群体遗传结构分析 [J]. *中国畜牧杂志*, 2023, 59 (11): 89-94.
- [19] 周隽, 覃星文, 吴细波, 等. 基于 SNP 芯片对陆川猪进行遗传多样性与遗传结构分析的研究 [J]. *中国畜牧杂志*, 2023, 59 (8): 257-263.
- [20] 黄树文, 张哲, 陈赞谋, 等. 广东省现有 5 个地方猪种基于 SNP 芯片的遗传多样性分析 [J]. *中国畜牧杂志*, 2018, 54 (6): 33-37.

证书

畜牧与兽医

2023年度江苏省高校优秀期刊

江苏省高等学校学报研究会

2023年12月20日