

山羊产羔数全基因组关联分析

兰蓉¹, 朱兰¹, 姚新荣², 王鹏², 邵庆勇¹, 洪琼花^{1*}

(1. 云南省畜牧兽医科学院, 昆明 650224; 2. 云南省种羊繁育推广中心, 寻甸 650000)

摘要: 本研究旨在通过对山羊产羔性状的全基因组关联分析(GWAS), 寻找和定位与山羊产羔性状密切相关的新基因。以云南黑山羊 6 个公羊家系的 302 只山羊为试验材料, 用 Illumina 公司 Iselect Goat60k 芯片技术进行 SNP 分型, 分型结果利用 plink V1.07 的线性回归模型对山羊产羔数性状进行全基因组关联分析。研究结果表明: 在 2 号染色体上有 2 个 SNPs 位点与山羊产羔数达到 5% 基因组水平显著相关 ($P < 1.48E-6$), 分别位于 *SLC4A10* 基因的下游和 *TBR1* 基因的上游; 5 个 SNPs 位点与山羊产羔数达潜在关联 ($P < 2.97E-5$), 分别位于 1 号染色体 *SEN7* 基因上游, 21 号染色体 Hypothetical Protein 基因的上游, 以及 28 号染色体 *WDFY4* 基因和 *TMEM26* 基因的上游、*BICC1* 基因的下游。这些基因可作为山羊产羔数性状的相关候选基因, 也可为山羊产羔性状的分子机制研究和今后标记辅助选择的开展提供理论基础及新的研究线索。

关键词: 云南黑山羊; 产羔数; 全基因组关联分析

中图分类号: S827; S813.3

文献标志码: A

文章编号: 0366-6964(2015)04-0549-06

Genome-wide Association Study of Lambing Number in Goat

LAN Rong¹, ZHU Lan¹, YAO Xin-rong², WANG Peng², SHAO Qing-yong¹, HONG Qiong-hua^{1*}

(1. Yunnan Institute of Animal Science and Veterinary, Kunming 650224, China;

2. Sheep Breeding Promotion Center in Yunnan Province, Xundian 650000, China)

Abstract: The aim of this study was to find and map novel genes related to lambing number in goat by genome-wide association study(GWAS). A total of 302 Yunnan Black goats from 6 families were genotyped by Illumina Goat60K iSelect array. The GWAS on goat lambing number was performed. The Results showed that 2 SNPs located at the downstream of *SLC4A10* and upstream of *TBR1* on chromosome 2 were significantly associated with lambing number ($P < 1.48E-6$) at 5% genome-wide level. Five SNPs were suggestive associated with lambing number ($P < 2.97E-5$). These 5 SNPs were located on chromosome 1, 21, 28, respectively, including *SEN7*, Hypothetical Protein, *WDFY4*, *TMEM26* and *BICC1* genes. These genes and SNPs offer essential information for understanding the molecular mechanisms of lambing number and provide foundation for the application of marker-assisted selection in breeding program of goat.

Key words: Yunnan Black goat; lambing number; genome-wide association study

山羊产羔数的不断提高一直是山羊生产的重要目标, 它关系到生产效率的提高和生产成本的降低。增加山羊的胎产羔数, 提高生产效益, 促进养羊业向持续环保发展道路的转变, 作为一个畜牧大国和人

口大国的中国显得尤为重要。但是山羊的产羔性状遗传力较低, 是一个受微效多基因控制的复杂数量性状, 同时也受一个或多个主基因控制, 常规育种手段的运用不能在较短时期内取得明显进展, 这在很

收稿日期: 2014-07-02

基金项目: 国家现代农业产业技术体系建设专项(CARS-39); 云南省科技计划项目(2014BB014)

作者简介: 兰蓉(1969-), 女, 云南人, 副研究员, 硕士, 主要从事家畜分子遗传育种研究, E-mail: rtlankitty@163.com

* 通信作者: 洪琼花, 研究员, E-mail: yxh7168@126.com

大程度上阻碍了山羊业的快速发展。准确筛选、鉴定产羔性状主效基因,进而进行分子育种,成为快速改良和提高山羊产羔数的前提条件。

目前已有很多学者对山羊产羔数相关基因做了研究,包括 *FSHR*^[1]、*gdf-9*^[2] 和 *NPY-Y1R* 等基因^[3],而利用目前较为先进的全基因组关联分析(GWAS)技术,开展山羊产羔性状的基因挖掘和定位的研究还未见报导。本研究拟采用 GWAS 技术对山羊产羔数的分子基础进行研究,以期挖掘和定位与山羊产羔数密切相关的新基因。

1 材料与方法

1.1 试验材料

本研究试验羊只来自云南省种羊繁育推广中心云南黑山羊核心群,由 6 个公羊家系的 302 只母羊组成,采集颈静脉 EDTA 抗凝血 2 mL, -20 °C 保存备用,记录试验羊只第二胎产羔数。

1.2 方法

1.2.1 基因组 DNA 提取及检测 按 blood genome DNA Extraction kit 试剂盒操作流程提取血样 DNA, 1×TE 缓冲液溶解,用 NANO 核酸蛋白分析仪检测 DAN 的浓度和质量。使 DNA 的吸光度 $OD_{260\text{ nm}}/OD_{280\text{ nm}}$ 为 1.7~2.0, 浓度大于 $50\text{ ng}\cdot\mu\text{L}^{-1}$ 、体积大于 20 μL 的样品作为基因分型样品,保存于 -20 °C 备用。

1.2.2 SNP 基因分型及质量控制 将质量达标的 DNA 样品送至北京斯克尔基因生物技术有限公司,用美国 Illumina 公司的山羊 Iselect Goat60k 芯片进行 SNP 分型。

1.2.3 芯片分型流程 包括 DNA 定量、扩增、片段化、片段化 DNA 沉淀及回收、芯片杂交、芯片清洗、单碱基延伸及染色、芯片扫描、数据分析、得出 SNP 分型结果。

1.2.4 质量控制 使用 Plink(1.07) 进行质量控制^[4],将 MAF 小于 0.05,检出频率小于 0.9,哈迪-温伯格平衡卡方检验将 $P<1\text{ E-}06$ 的 SNP 位点排除,同时还排除检出率 <0.9 的样本。

1.3 数据统计分析

1.3.1 表型性状统计分析 用 SAS(8.01)中的 CORR 过程对试验羊只第二胎产羔数进行描述性统计及羊只出生年度与产羔数的相关性分析。

1.3.2 群体结构分析 为了消除连锁不平衡对个体间遗传相关估计的偏差,选用常染色体上独立

的 SNPs,应用 Plink(1.07)全基因组 IBS(Identity-by-state)多维度分析^[4],进行 MDS 分析。SNPs 筛选时以 25 个位点为一个窗口,以 5 个位点为步长, r^2 设为 0.2,基于得到的独立 SNP 点计算个体之间的 IBS 距离,应用 R(V2.13.1)作图^[5],进而进行群体结构分析。

1.3.3 全基因组关联分析 运用 PLINK(V1.07)线性回归模型进行产羔性状的全基因组关联分析。为减少重复检验的假阳性,使用连锁不平衡修正的 Bonferroni 修正,以 25 个位点为一个窗口,5 个位点为步长, r^2 为 0.4,得到独立的 LD 块和 SNP,并据此进行全基因组关联分析值校正。

1.3.4 生物信息分析 应用软件 BLAST(2.28+)将潜在关联的 SNP 位点上下 60 bp 序列比对到参考基因组中^[6-7],得到 SNP 所在的染色体位置坐标,结合山羊参考基因组 Annotation-GFF 文件,查找 SNP 位置前后 500 K 的基因,获得 Gene-ID,根据 Gene-ID 获取基因的编码区(CDS)序列和 pep 序列,再用基因的 pep 序列进行 Nr 库注释,获取基因的注释信息。

2 结果

2.1 产羔数统计结果

云南黑山羊二胎产羔数平均为 2.04 只,变异系数为 0.75,通过相关统计分析,产羔数与母羊的出生年龄相关系数为 0.15,相关不显著($P>0.05$),说明母羊年龄对其产羔数无影响,因而在后面 GWAS 分析中采用了线性回归模型。

2.2 基因分型数据质控结果

经质控后 MAF 小于 0.05 的位点有 6 162 个,检出频率 <0.9 的位点有 337 个,哈迪-温伯格平衡卡方 $P<0.000\ 001$ 的位点有 327 个,检出率 <0.9 的样本、排除染色体位置信息不确定点 474 个,最终获得 295 个样本在 45 608 个位点上的分型结果(表 1)。

2.3 群体结构分析结果

通过 MDS 分析,在常染色体可筛选出 15 309 个独立的 SNPs 位点,以计算的 MDS 组分 1 和组分 2 为坐标,应用 R(V2.13.1)绘制群体结构散点图(图 1),6 个公羊家系分别用一种颜色表示,由图 1 可见同一个家系个体几乎聚集在一起,并未均匀分布,表明存在群体分层,同一家系个体间的遗传相关程度较高。

表 1 每条染色体上 SNP 数及对应的 Bonferroni 校正的 5% 染色体水平显著阈值

Table 1 Bonferroni-corrected 5% chromosome-wise significance threshold after quality control

染色体 Chromosome	标记数 No. of SNPs	独立的标记数 No. of independent SNPs	5% 染色体水平显著阈值 5% chromosome-wise significant threshold
ch1	2 932	2 233	2.24E-05
ch2	2 550	1 946	2.57E-05
ch3	2 062	1 502	3.33E-05
ch4	2 167	1 553	3.22E-05
ch5	1 970	1 352	3.70E-05
ch6	2 010	1 433	3.49E-05
ch7	1 948	1 441	3.47E-05
ch8	2 070	1 529	3.27E-05
ch9	1 692	1 331	3.76E-05
ch10	1 893	1 462	3.42E-05
ch11	1 809	1 266	3.95E-05
ch12	1 573	1 168	4.28E-05
ch13	1 478	1 134	4.41E-05
ch14	1 665	1 259	3.97E-05
ch15	1 453	988	5.06E-05
ch16	1 438	1 064	4.70E-05
ch17	1 270	991	5.05E-05
ch18	1 096	853	5.86E-05
ch19	1 122	816	6.13E-05
ch20	1 295	998	5.01E-05
ch21	1 319	956	5.23E-05
ch22	1 043	799	6.26E-05
ch23	930	648	7.72E-05
ch24	1 191	855	5.85E-05
ch25	769	565	8.85E-05
ch26	948	722	6.93E-05
ch27	834	652	7.67E-05
ch28	801	646	7.74E-05
ch29	867	688	7.27E-05
ChX	1 413	826	6.05E-05
Total	45 608	33 674	

2.4 全基因组关联分析结果

使用连锁不平衡修正的 Bonferroni 进行修正,

得到独立的 LD 块和 33 674 个独立的 SNPs 用于 GWAS 分析, 据此得到 Bonferroni 修正的达 5% 基

基因组显著水平的 P 值为 $1.48E-6$ ($0.05/33674$), 基因组潜在关联阈值为 $2.97E-5$ ($1/33674$)。同时根据每条染色体上的独立 SNPs 数目, 得到每条染色体水平的显著性阈值(表 1), 通过对比, 发现达到潜在基因组水平显著差异的 SNPs 位点共 7 个(图 2, 表 2)。图 2 和表 2 显示: 2 个 SNPs 位点与山羊产

仔数达到 5% 基因组水平显著相关 ($P < 1.48E-6$), 分别位于 *SLC4A10* 基因的下游和 *TBR1* 基因的上游; 5 个位点达潜在关联 ($P < 2.97E-5$), 分别位于 1 号染色体 *SENP7* 基因上游, 21 号染色体 Hypothetical Protein 基因的上游, 以及 28 号染色体 *WDFY4* 和 *TMEM26* 基因的上游、*BICC1* 基因的下游。

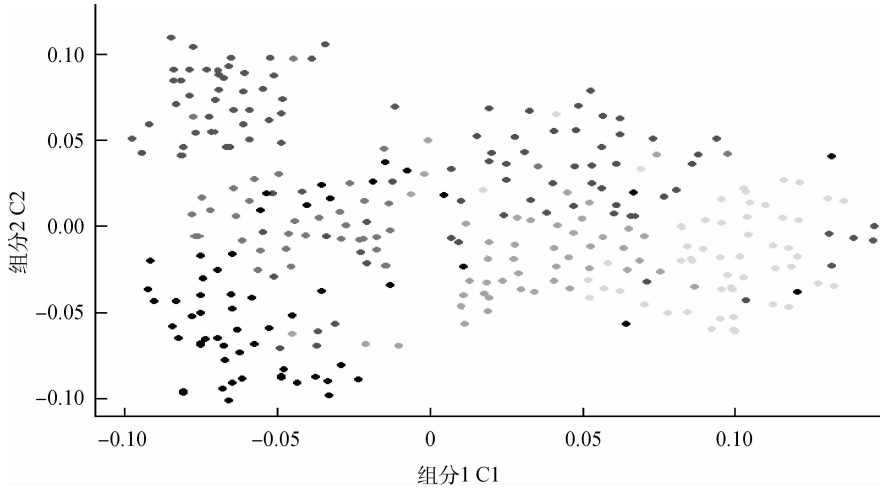
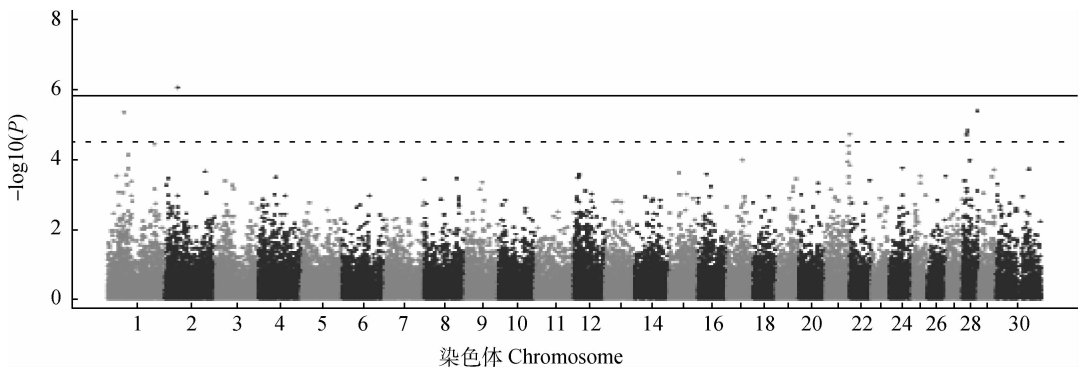


图 1 群体结构多维度分析图

Fig. 1 Population structure identification with multidimensional scaling analysis



横坐标为染色体号, 30 代表 X 染色体。纵坐标为关联分析 P 值的 $-\log_{10}$ 结果。黑色实线为达基因组水平 5% 显著阈值线, 黑色虚线为达基因组潜在关联阈值线

The scale on the X-axis represents ID of chromosomes. X chromosome is represented by 30. The scale on the Y-axis is the $-\log_{10}(P)$ score of association analysis. The black solid line is drawn at $-\log_{10}(1.48E-6)$ to show those significant at the genome-wide 5% threshold, and the black dashed line indicates genome-wide significance of suggestive association

图 2 产羔数全基因组关联分析曼哈顿图

Fig. 2 Manhattan plot of genome-wide association analysis for lambing number

3 讨论

全基因组关联分析 (Genome-wide association study, GWAS) 是一种利用遍布于生物体基因组中数以百万的分子标记 (主要是单核苷酸突变, 即

SNP), 并借助统计学工具对影响某些复杂性状如疾病易感性或重要经济性状的遗传变异进行鉴定和分析的一种新策略, 其基本思想是直接检测基因本身或基因附近微小区域的 SNP 标记与性状表型信息的关联来实现基因的精细定位。目前, GWAS 技术

表 2 产羔数 Bonferroni 校正的 5% 染色体水平显著的 SNPs 位点信息

Table 2 Bonferroni-corrected 5% chromosome-wide significant SNPs for lambing number

标记 SNPs	染色体 Chromosome	物理位置 BP	基因型 Allele	最近基因 Nearest gene	P 值 P-value
snp20467-scaffold202-4458644	2	34235387	T/C	SLC4A10(D)	8.57E-07
snp20468-scaffold202-4490513	2	34267256	T/G	TBR1 (U)	8.57E-07
snp11435-scaffold1416-379578	28	40521920	T/C	WDFY4(U)	4.07E-06
snp45731-scaffold627-4874627	1	44576529	A/G	SENPF7(U)	4.49E-06
snp54840-scaffold838-3194600	28	14948614	T/C	TMEM26(U)	1.50E-05
snp25958-scaffold2688-260916	21	65919086	T/C	Hypothetical Protein(U)	1.86E-05
snp54781-scaffold838-728769	28	12482783	A/C	BICC1(D)	1.92E-05

D. SNP 位于基因的下游; U. SNP 位于基因的上游

D. SNP located in the downstream of the gene; U. SNP located in the upstream of the gene

已在奶牛、绵羊、猪、鸡等畜种中得以应用^[8-12],挖掘出许多与重要经济性状相关的新基因,包括繁殖、遗传抗性、产乳性状、生长性状等重要经济性状相关的新基因,而有关山羊产羔性状的 GWAS 研究则未见报导。

本研究 MDS 分析发现试验样本存在群体分层。已有的人类疾病的全基因组关联分析发现,由于祖先的遗传差异引起疾病和对照间等位基因频率的差异造成的群体分层现象,导致标记和性状间的伪关联^[13],利用全基因组标记的遗传信息对样本遗传相关进行估计,比系谱信息可更真实的反映个体间的相关度,因此本研究 GWAS 分析时使用了 MDS 分析产生的组分 1 作为协变量,以校正群体结构对关联分析结果的影响^[14],对应的全基因组膨胀系数由 1.14~2.15 减小为 1~1.05,可以认为修正后得到的结果没有受到群体分层等混杂因素的影响。

本研究在生物信息学分析中均是用差异位点与附近基因的编码区进行比对才确认 SNP 位点是位于该基因的上游或下游,因此本研究发现的与产羔数相关的基因均是位于功能基因上。

本研究发现,在 2 号染色体上的 SLC4A10 基因和 TBR1 基因显著影响山羊的产羔数。SLC4A10 基因编码钠离子推动的氯离子/碳酸氢根 ($\text{Cl}^-/\text{HCO}_3^-$) 交换蛋白,这个蛋白是各种细胞中重要的胞内 pH 调节剂^[15],对于产羔性状的作用可能是通过调节输卵管液中 HCO_3^- 的浓度,影响精子的获能^[16],从而影响到受精率,最终影响产羔率,已有研究表明,输卵管液中高浓度的 HCO_3^- 有助于精子

的获能^[17]。TBR1 基因编码 T-box 转录因子,这个蛋白参与机体发育过程中的转录调节,在发育时期的大脑皮质中有大量表达^[18],但它影响产羔性状的途径还不清楚,有待进一步研究。

本研究还发现,在 28 号染色体上有 3 个 SNPs 位点与产羔数存在潜在关联,分布在 12~40 Mb,这个区域内有 WDFY4、TMEM26 及 BICC 3 个候选基因,但对这 3 个基因的相关功能研究极少,在山羊中更是未见报导,所以本研究发现,它们与山羊的产羔数存在潜在的关联还需要做大样本、更为深入的研究,以确保得到的结论真实可靠。

本研究结果可作为山羊产羔性状的相关候选区域和候选基因,也可为山羊产羔性状的分子机制研究和今后的标记辅助选择的开展提供新的研究线索和理论基础。但本研究由于经费及时间所限,只采用了 302 只云南黑山羊作为试验样本,因此所获候选区域及基因仍然需要进行更大样本的重复验证,以确保今后在分子育种中应用的准确性和效率。

4 结 论

本研究鉴定出位于 2 号染色体上的 SLC4A10 和 TBR1 基因为影响山羊产羔性状的重要候选基因;28 号染色体 12~40 Mb 的区域内分布的 WDFY4、TMEM26 和 BICC1 基因对山羊产羔数有潜在的影响。

参考文献 (References):

[1] 朱 吉,杨仕柳,欧阳叙向,等. 山羊 FSHR 基因 5'端

- 多态性与产仔数性状的关系 [J]. 草食家畜, 2007, 134(1):12-16.
- ZHU J, YANG S L, OUYANG X X, et al. Relationship between polymorphism of follicle stimulating hormone receptor gene and litter size traits in goats. [J]. *Grass-Feeding Livestock*, 2007, 134(1):12-16. (in Chinese)
- [2] 任丽娜, 刘思国, 胡大伟, 崇明白山羊 gdf-9 基因多态性与产仔率的关联分析 [J]. 畜牧与兽医, 2012, 44(10):25-29.
- REN L N, LIU S G, HU D W, et al. Polymorphism of growth differentiation factor 9 gene and its relationship with litter size of Chong-ming white goats. [J]. *Animal Husbandry & Veterinary Medicine*, 2012, 44(10):25-29. (in Chinese)
- [3] 储明星, 冯涛, 赵有璋, 等. 神经肽 Y-Y1 受体基因 (NPY-Y1R) 多态性及其与山羊繁殖性能关系 [J]. 农业生物技术学报, 2009, 17(6):960-966.
- CHU M X, FENG T, ZHAO Y Z, et al. Polymorphism of neuropeptide Y-Y1 receptor gene (NPY-Y1R) and its relationship with reproductive performance in goats [J]. *Journal of Agriculture Biotechnology*, 2009, 17(6):960-966. (in Chinese)
- [4] 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.
- [5] DEVELOPMENT R, CORE T R. A language and environment for statistical computing R Foundation for Statistical Computing [M]. Vienna, Austria, 2011.
- [6] <http://goat.kiz.ac.cn/GGD/download9.htm>.
- [7] DONG Y, XIE M, JIANG Y, et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*) [J]. *Nat Biotechnol*, 2013, 31(2):135-141.
- [8] LI C, SUN D, ZHANG S, et al. Genome wide association study identifies 20 novel promising genes associated with milk fatty acid traits in chinese Holstein [J]. *PLoS ONE*, 2014, 9(5):e96186.
- [9] KIJJAS J W, SERRANO M, MCCULLOCH R, et al. Genome wide association for a dominant pigmentation gene in sheep [J]. *J Anim Breed Genet*, 2013, 130(6):468-475.
- [10] GARCÍA-GÁMEZ E, GUTIÉRREZ-GIL B, SAHANA G. GWA analysis for milk production traits in dairy sheep and genetic support for a QTN influencing milk protein percentage in the LALBA gene [J]. *PLoS ONE*, 2012, 7(10):e47782.
- [11] JUNG E J, PARK H B, LEE J B, et al. Genome-wide association study identifies quantitative trait loci affecting hematological traits in an F2 intercross between Landrace and Korean native pigs [J]. *Anim Genet*, 2014, 45(4):534-541.
- [12] SUN Y F, ZHAO G P, LIU R R, et al. The identification of 14 new genes for meat quality traits in chicken using a genome-wide association study [J]. *BMC Genomics*, 2013, 14:458-468.
- [13] PRICE A L, PATTERSON N J, PLENGE R M, et al. Principal components analysis corrects for stratification in genome-wide association studies [J]. *Nat Genet*, 2006, 38(8):904-909.
- [14] 顾晓荣. 影响鸡体重性状 QTL 的全基因组连锁和关联研究[D]. 北京: 中国农业大学, 2011.
- GU X R. Whole genome linkage and association studies on QTLs effecting body weight in chicken[D]. Beijing: China Agricultural University, 2011. (in Chinese)
- [15] DAMKIER H H, AALKJAER C, PRAETORIUS J. Na⁺-dependent HCO₃⁻ import by the slc4a10 gene product involves Cl⁻ export [J]. *J Biol Chem*, 2010, 285(35):26998-27007.
- [16] DEMARCO L A, ESPINOSA F, EDWARDS J, et al. Involvement of a Na⁺/HCO₃⁻ cotransporter in mouse sperm capacitation [J]. *J Biol Chem*, 2003, 278(9):7001-7009.
- [17] TIENHAI P, JOHANNISSON A, RODRIGUEZ-MARTINEZ H. Sperm capacitation in the porcine oviduct [J]. *Anim Reprod Sci*, 2004, 80(1-2):131-146.
- [18] BEDOGNI F, HODGE R D, ELSEN G E, et al. Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex [J]. *Proc Natl Acad Sci USA*, 2010, 107(29):13129-13134.