

DOI: 10.13718/j.cnki.xdzk.2018.09.006

蛋鸡资源群体的盲肠长度全基因组关联分析^①

王星果, 曲亮, 卢建, 窦套存, 马猛,
胡玉萍, 沈曼曼, 郭军, 李尚民, 王克华

江苏省家禽科学研究所, 江苏 扬州 225125

摘要:【目的】盲肠是鸡进一步消化和吸收养分,尤其是纤维的器官,对鸡的生长及后期产蛋都具有重要作用,因此鸡的盲肠长度是一个重要的生理指标。【方法】通过对白来航鸡与东乡绿壳蛋鸡杂交所得 F2 代个体的盲肠长度进行测量,并运用单核苷酸多态性(SNP)芯片检测其基因分型,根据 SNP 检测数据运用 SAS 进行遗传评估,并用全基因组混合模型关联算法(GEMMA)进行单变量全基因组关联分析(GWAS)。【结果】结果显示,盲肠长度表现出中等遗传力(0.39),GWAS 鉴定出 54 个 SNP 与盲肠的长度显著相关,且 1 号染色体 170 Mb 附近对于盲肠长度来说是一个重要区域。在这个区域,覆盖 26 个 SNP 位点的 18 个基因被定为候选基因,其中 2 个分别在编码序列(CDS)和 3'非翻译区(3'UTR),对应于 *NHLRC3* 和 *SIAH3*,它们可能是影响盲肠长度的重要 SNP 位点和基因。【结论】利用 GWAS 筛选并鉴定出和鸡盲肠长度相关联的 SNP 位点及基因,将为揭示蛋鸡盲肠发育的机制和分子育种提供基础。

关键词: 蛋鸡;盲肠长度;GWAS;遗传结构

中图分类号: S831.2

文献标志码: A

文章编号: 1673-9868(2018)09-0035-07

在未来的 20 年里人们对动物蛋白的需求将不断增长,而动物蛋白主要来源于牛奶和家禽产品^[1],这使得研究者们相信有必要提高畜禽产量,在发展中国家更有必要。饲养蛋鸡的主要成本是饲料,一般可以通过修改饮食配方,从而提高饲料效率的方法来减少用量。随着生物技术的发展,使用基因选择则是一种更有效的方法^[2]。

肠道既是重要的消化器官,又是维持内环境稳态的天然屏障。肠道发育是家禽饲料效率的一个重要因素,与消化酶活性和吸收营养的能力密切相关。肠道长度是消化道发育的一个重要指标,鸡肠道的长度与采食量有关,并影响鸡的生长^[3]。盲肠被认为是鸡消化粗纤维的主要场所,其长度与消化能力相关^[4]。虽然有一些文献报道了小肠长度的遗传背景^[5-8],然而很少有关于盲肠长度遗传背景的研究报道。因此,对鸡盲肠长度的遗传结构进行解析很有必要。

将整个基因组层面的序列变异(主要是单核苷酸多态性,SNPs)与表型和血统信息相结合,可用于

① 收稿日期: 2017-07-15

基金项目: 财政部、农业部国家蛋鸡产业技术体系项目(CARS-41-K02);江苏省农业科技自主创新项目(CX(16)1008);江苏省农业三新工程项目(SXGC[2016]290)。

作者简介: 王星果(1984-),男,博士,助理研究员,主要从事动物遗传育种和脂质代谢研究。

通信作者: 王克华,研究员。

进行全基因组关联分析(GWAS),并识别感兴趣性状特征的基因或调控元素^[9].与传统数量性状基因座(QTL)定位策略相比,GWAS的主要优势在于其检测变异效果的强大能力并将变异定义到较窄的基因组区域^[10].在过去的几年中,GWAS揭示了许多与鸡的形态特征、生产性状、抗病等相关的重要研究成果^[9].然而,还没有鸡盲肠长度的GWAS研究.因此,本研究主要使用600 K鸡SNP芯片对黑羽蛋鸡F2代资源群体(1 512 只母鸡)运用单变量GWAS检测与盲肠长度表型相关联的基因组区域或基因.

1 材料和方法

1.1 构建资源群体

F2代资源群体由白来航(WL)和东乡绿壳蛋鸡(DX)2个纯系杂交产生.6只WL公鸡与133只DX母鸡交配、6只DX公鸡与80只WL母鸡交配形成F1代.然后,从WL/DX杂交产生的25只公鸡和407只母鸡与从DX/WL杂交产生的24只公鸡和235只母鸡选留下来生产F2代.共生产3 749只F2代个体(1 856只公鸡和1 893只母鸡).然后,从550个全同胞和49个半同胞家系中选择1 534只母鸡进行SNP基因分型,以确保足够的表型和背景信息.

1.2 表型测量

为描述盲肠长度的遗传结构,将72周龄F2代鸡屠宰,收集整个盲肠,挤压出内容物后测量其长度.使用所有可用的记录,运用SAS软件包的MEANS程序对数据进行描述性统计计算.对偏离正态分布的特征,使用SAS软件进行正态转换后再进行关联分析,然后这些转换值被用于下游分析,包括GWAS和遗传估计.

1.3 基因分型、质量控制

F2代鸡进行静脉采血,使用标准的苯酚/氯仿萃取法提取基因组DNA,使用600 K Affymetrix Axiom鸡基因芯片进行基因分型.删除7 883个未知基因位置的SNP和112个冗余基因坐标的SNP,使用Affymetrix工具v1.16.0软件的Axiom GT1算法收集SNP原始数据.只将质量控制(QC)在0.82或更高的样品用于下游分析,共得到1 512个样本和532 299个SNP.此外,考虑到当前统计方法检测表型和常染色体基因型之间的联系更有效,将性染色体的6 402个SNP剔除,使用PLINK v1.90程序包删除67 330个最小等位基因频率(MAF)小于5%和22 700个偏离哈迪温伯格平衡测试($p < 1 \times 10^{-6}$)的SNP,再使用BEAGLE v4.0程序归因一些零星的缺失基因型,只保留质量分数(R^2)大于0.5的SNP^[11].最后,共有1 512个样本和435 867个SNP可进行GWAS.

1.4 全基因组关联分析

GWAS之前先进行主成分分析,消除虚假相关.使用simpleM法更正测试数量,建立全基因组潜在和显著关联的合适阈值.建立有效的独立测试数值 $M_{\text{eff}}=59\,308$,全基因组显著和潜在的 p 值分别为 8.43×10^{-7} (0.05/59 308)和 1.69×10^{-5} (1.00/59 308).首先使用GEMMA v0.94软件的精确混合模型方法进行单变量分析,确定 $MAF \geq 0.05$ 的SNP的关联性^[12],再用独立SNP计算集中相似度矩阵,然后针对SNP和表型之间的显著水平计算Wald测试的 p 值.单变量线性混合模型为

$$y = W\alpha + x\beta + \mu + \epsilon$$

y 是 n 个个体的 $n \times 1$ 表型值向量, W 是 $n \times c$ 协变量矩阵(固定效应), α 是 $c \times 1$ 对应系数向量, x 是 $n \times 1$ 标记基因型向量, β 是标记的相应效应大小, μ 是协方差结构的 $n \times 1$ 随机多基因效应向量, ϵ 是 $n \times 1$ 随机残差向量.

曼哈顿和quantile-quantile(QQ)图使用R软件附带的GAP包绘制.用于确定假阳性信号程度的基因组膨胀因素 λ ,也使用R软件附带的GenABEL包中的estlambda函数计算^[13].

1.5 遗传和表型方差估计

运用 GCTA v1.24 程序中的单变量限制极大似然法估计遗传力^[14]. 针对各部分肠长度运用二元混合模型估计遗传和表型相关^[15]. 对于全基因组显著 SNP, 表型方差贡献使用下面的混合线性模型估计为

$$y = Xb + gG + e$$

y 是 n 个个体的 $n \times 1$ 表型值向量, b 是固定效应向量, X 是它的关联矩阵, gG 是 SNP 聚合效应的向量, e 是随机残余项.

1.6 基因鉴定

对显著性 SNP 进行功能注释, 使用变异效应预测软件 VEP 和 Ensembl 中的 Biomart 工具, 对在给定基因组区域中的候选基因的显著性基因座进行搜索^[16-17].

2 结 果

2.1 表型和遗传参数

屠宰 72 周龄 F2 代鸡, 得到盲肠长度 CL . 共测量 1 505 只母鸡, 获得 1 434 份有效数据. 对其进行描述统计可知, CL 平均为 (14.94 ± 2.10) mm, 最大值、最小值分别为 22.30 cm, 6.10 cm, 变异系数为 14.06%. 所有表型值在逆正态转换后符合正态分布, 转换后的值被用于所有后续分析.

通过合格的 GWAS 标记对 CL 及相关性状十二指肠长度 DL 、空肠长度 JL 、回肠长度 IL 的加性遗传方差进行计算, 并统计遗传参数, 结果如表 1 所示. 单变量 GCTA 分析显示 CL 遗传力在中等水平 ($h^2 = 0.39$). 双变量 GCTA 分析表明, CL 与 IL 的遗传相关较高, 而与 DL 和 JL 表现出中等遗传相关. 表型相关分析表明 CL 与其他几种小肠长度具有较低的表型相关性.

表 1 盲肠长度遗传参数

性状		DL	JL	IL
CL	遗传相关	0.54(0.09)	0.54(0.08)	0.62(0.07)
CL	表型相关	0.31**	0.28**	0.36**

注: 括号中为标准误; “**”表示差异极具有统计学意义 ($p < 0.01$).

2.2 候选位点 GWAS 鉴定

对 CL 进行单变量全基因组关联分析. 共鉴定出 54 个全基因组显著位点与 CL 关联. 显著的基因组区域从 1 号染色体上(GGA1)的 166.66 Mb 到 171.06 Mb. 分析发现有 209 个全基因组潜在位点与 CL 关联, 这些位点分别位于 GGA1 的 165.16~172.54 Mb, GGA4 的 688.54~731.56 Mb 以及 GGA5 的 312.90 Mb. 根据所有影响盲肠长度的 SNP 的 p 值制作曼哈顿图和 QQ 图, 如图 1 所示.

2.3 显著位点的基因注释

运用 VEP 和 Biomart 对显著 SNP 周围的区域进行扫描, 鉴定与盲肠长度相关的基因. 由于基因内部的突变比位于基因间区域的 SNP 更有意义, 我们对上述 54 个 SNP 进行了分析. 覆盖 26 个 SNP 的 18 个基因被确定为候选基因, 其中 5 个基因有不止 1 个 SNP 位点(表 2). 在这 26 个 SNP 位点中, 有 2 个分别位于 CDS 和 3'UTR, 均可被转录成成熟的 mRNA. 这 2 个 SNP 位点 rs316180141 和 rs14914978 及其对应的基因 $NHLRC3$ 和 $SIAH3$ 被视为最重要的 SNP 位点和基因.

2.4 表型方差贡献(CPV)的估计

使用 GCTA 工具对和盲肠长度关联的这 2 个 SNP 位点进行表型方差贡献的估计, 它们对于盲肠长度的 CPV 如表 3 所示. 位点 rs316180141 和 rs14914978 分别可以解释 CL 表型方差的 4.11%和 3.50%.

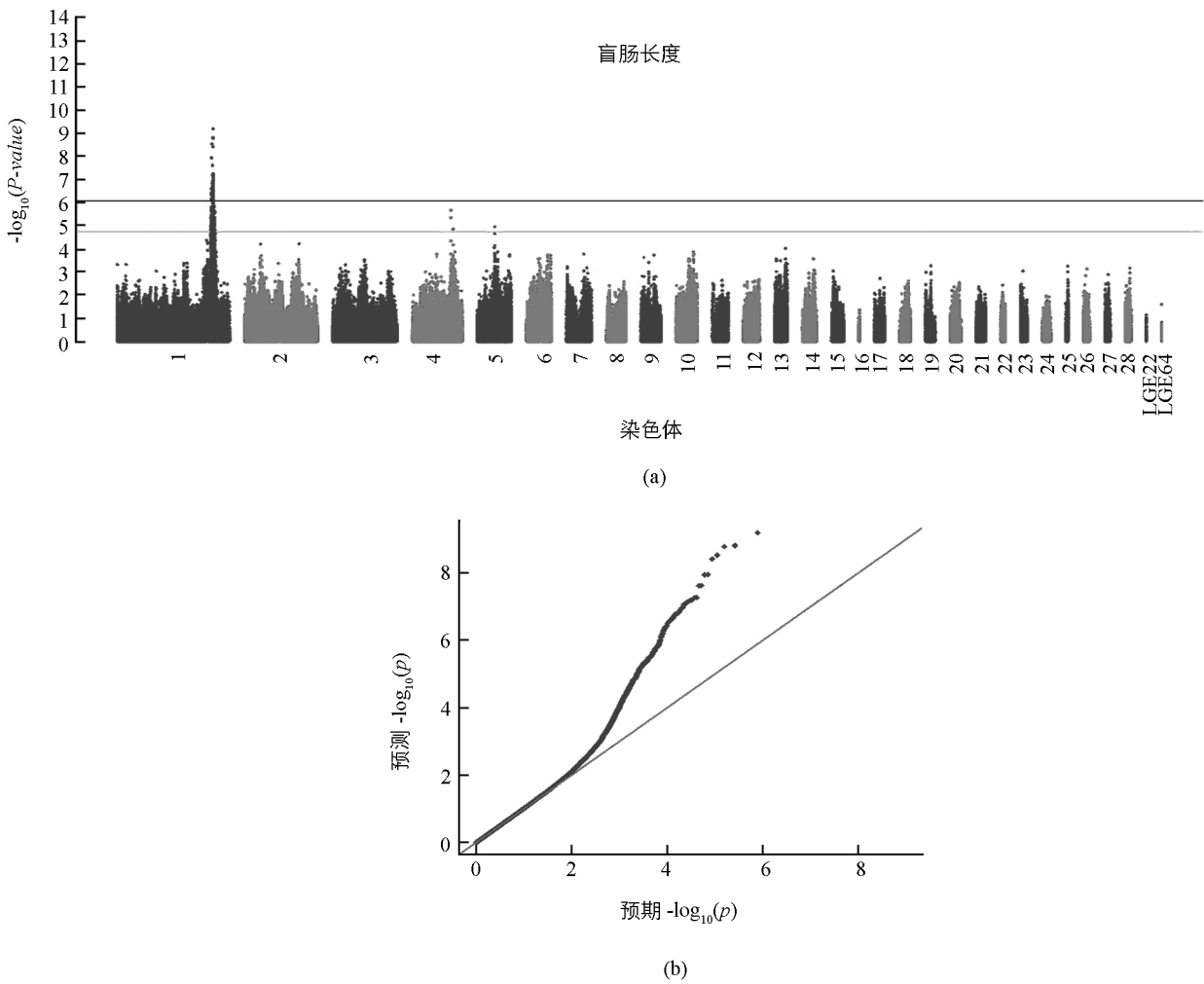


图 1 基于鸡盲肠长度 p 值的曼哈顿图(a)和 QQ 图(b)

表 2 潜在基因及其覆盖的 SNP 位点

基因	SNP	区域
CAB39L	rs14916642, rs14916609, rs313913043, rs313378297	内含子
VPS36	rs317477277, rs13973297, rs315690458	内含子
RB1	rs313207223	内含子
CKAP2	rs312737959	内含子
ENSGALT00000027531	rs13974362	内含子
ENSGALT00000027506	rs316815773	内含子
FREM2	rs314139974	内含子
NEK3	rs15500685	内含子
FOXO1	rs13553625, rs315124749	内含子
RNASEH2B	rs317812130	内含子
LOC428073	rs13973387, rs13973391	内含子
COG6	rs15502381, rs316214918	内含子
NHLRC3	rs316180141	CDS
COG3	rs317757047	内含子
GTF2F2	rs315041530	内含子
SIAH3	rs14914978	3'UTR
ENSGALT00000036488	rs316853653	内含子
THSD1	rs13553531	内含子

表 3 2 个显著 SNP 位点对盲肠长度的表型方差贡献

SNP	染色体	位置	基因	区域	等位基因	MAF	效应(s. e. m.)	CPV/%	<i>p</i> 值
rs316180141	1	170539428	<i>NHLRC3</i>	CDS	T/C	0.488	−0.288(0.056)	4.11	2.61E-07
rs14914978	1	167193066	<i>SIAH3</i>	3′UTR	A/T	0.429	−0.262(0.052)	3.50	4.43E-07

3 讨 论

盲肠是显著影响鸡消化吸收粗纤维的重要器官. 大部分关于鸡盲肠长度的研究集中在饲料营养方面^[18–19], 然而对鸡盲肠长度的基因组结构目前尚不清楚. GWAS 是分析动物重要遗传结构特点的一个强大工具. 目前, 在多种农业动物如牛^[20–21]、猪^[22–23]和家禽^[24–25]中已确定了许多 QTL 及 SNP. 因此, 有必要使用 GWAS 研究鸡盲肠长度. 本研究中包括 1 512 只蛋鸡, 使得特征差异最大化, 提高了确定性状 QTL 的概率, 同时高密度(600 K)染色体 SNP 芯片覆盖了整个鸡染色体, 以提高准确性和可靠性.

从盲肠长度的数据可以看出, 尽管最大值和最小值之间的差异很大, 但大多数测量的数值在正常范围内. 遗传评估的结果表明, *CL* 具有中等遗传力. 盲肠长度与小肠长度的 3 个特征相关分析结果表明, 其与回肠长度具有高遗传相关性, 但表型相关性并不高, 表明这些特征受环境因素影响较大. 这个结果与我们之前关于小肠长度和质量的研究结果类似. 十二指肠、空肠和回肠长度(质量)有较高的遗传相关性, 但却是中等表型相关, 表明小肠长度和质量也有相似的遗传特征.

GWAS 的结果显示, 对于盲肠长度大部分的显著性 SNP 位点位于 GGA1 的 170 Mb 附近, 表明该区域对盲肠长度很重要. 在我们之前对小肠长度和质量的研究中, 最重要的 SNP 位点也落在 GGA1 的 170 Mb 附近, 表明这个区域对于小肠长度和质量也很重要. 由此可见, 这段区域是研究肠道发育的重要区域, 值得做进一步研究.

在与盲肠长度相关联的 26 个候选 SNP 和相对应的 18 个基因中, 我们推测 2 个 SNP 位点 rs316180141 和 rs14914978 及其对应的基因 *NHLRC3* 和 *SIAH3* 是最重要的基因, 因为这 2 个 SNP 位于相应基因的转录区域. *NHLRC3*(NHL repeat containing 3)是 NHL 重复包含蛋白 3, 它可能参与了一些酶过程, 比如泛素化等, 我们预测它在鸡盲肠中通过各种酶对消化起到促进作用. *SIAH3* (siah E3 ubiquitin protein ligase family member 3)是 siah E3 泛素蛋白连接酶家族成员 3, 它参与泛素依赖的蛋白水解过程^[26], 对多器官发育起重要作用. 因此, 我们认为这 2 个基因与盲肠长度高度相关, 值得进一步研究. 而对应的 2 个 SNP 的 CPV 均较高, 进一步提示这 2 个基因是影响鸡盲肠长度的候选基因.

4 结 论

通过 GWAS 鉴定出 54 个与盲肠长度显著相关的 SNP 位点, 且大部分集中在 1 号染色体 170 Mb 附近. 共将覆盖 26 个 SNP 位点的 18 个基因定为候选基因, 其中 2 个分别在 CDS 和 3′UTR, 对应于 *NHLRC3* 和 *SIAH3*. 这些位点和基因可能是影响盲肠长度的重要 SNP 位点和基因.

参考文献:

[1] HUME D, WHITELAW C B A, ARCHIBALD A. The Future of Animal Production: Improving Productivity and Sustainability [J]. The Journal of Agricultural Science, 2011, 149(S1): 9–16.

[2] DE VERDAL H, NARCY A, BASTIANELLI D, et al. Improving the Efficiency of Feed Utilization in Poultry by Selection. 2. Genetic Parameters of Excretion Traits and Correlations with Anatomy of the Gastro-Intestinal Tract and Digestive Efficiency [J]. BMC Genetics, 2011, 12(1): 1–10.

- [3] MALJAARS P W, PETERS H P, KODDE A, et al. Length and Site of the Small Intestine Exposed to Fat Influences Hunger and Food Intake [J]. *British Journal of Nutrition*, 2011, 106(10): 1609—1615.
- [4] 李 利, 臧素敏, 王 鹏, 等. 太行鸡和海兰灰蛋雏鸡盲肠主要消化生理特性的比较 [J]. *畜牧与兽医*, 2010, 42(11): 37—40.
- [5] GAO J, REN J, ZHOU L H, et al. A Genome Scan for Quantitative Trait Loci Affecting the Length of Small Intestine in a White Duroc x Chinese Erhualian Intercross Resource Population [J]. *Journal of Animal Breeding and Genetics*, 2010, 127(2): 119—124.
- [6] GAO Y, DU Z Q, WEI W H, et al. Mapping Quantitative Trait Loci Regulating Chicken Body Composition Traits [J]. *Animal Genetics*, 2009, 40(6): 952—954.
- [7] AMBO M, MOURA A S, LEDUR M C, et al. Quantitative Trait Loci for Performance Traits in a Broiler x Layer Cross [J]. *Animal Genetics*, 2009, 40(2): 200—208.
- [8] MIGNON-GRASTEAU S, RIDEAU N, GABRIEL I, et al. Detection of QTL Controlling Feed Efficiency and Excretion in Chickens Fed a Wheat-Based Diet [J]. *Genetics Selection Evolution*, 2015, 47(1): 74.
- [9] HUI Z, WANG Z, WANG S, et al. Progress of Genome Wide Association Study in Domestic Animals [J]. *Journal of Animal Science & Biotechnology*, 2012, 4(1): 291—298.
- [10] HIRSCHHORN J N, DALY M J. Genome-Wide Association Studies for Common Diseases and Complex Traits [J]. *Nature Reviews Genetics*, 2005, 6(2): 95—108.
- [11] PURCELL S, NEALE B, TODD-BROWN K, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses [J]. *American Journal of Human Genetics*, 2007, 81(3): 559—575.
- [12] ZHOU X, STEPHENS M. Genome-Wide Efficient Mixed Model Analysis for Association Studies [J]. *Nature Genetics*, 2012, 44(7): 821—824.
- [13] ZHAO J H. Gap: Genetic Analysis Package [J]. *Journal of Statistical Software*, 2007, 23(23): 313—336.
- [14] YANG J, LEE S H, GODDARD M E, et al. GCTA: A Tool for Genome-Wide Complex Trait Analysis [J]. *American Journal of Human Genetics*, 2011, 88(1): 76—82.
- [15] LEE S H, DECANDIA T R, RIPKE S, et al. Estimating the Proportion of Variation in Susceptibility to Schizophrenia Captured by Common SNPs [J]. *Nature Genetics*, 2012, 44(3): 247—250.
- [16] MCLAREN W, PRITCHARD B, RIOS D, et al. Deriving the Consequences of Genomic Variants with the Ensembl API and SNP Effect Predictor [J]. *Bioinformatics*, 2010, 26(16): 2069—2070.
- [17] KINSELLA R J, KAHARI A, HAIDER S, et al. Ensembl BioMart: a Hub for Data Retrieval Across Taxonomic Space [M]. Oxford: Oxford University Press, 2011.
- [18] 李 燕, 康相涛, 孙桂荣, 等. 木寡糖对矮脚绿壳蛋鸡肠道长度及形态结构的影响 [J]. *饲料研究*, 2007(12): 67—69.
- [19] 占今舜, 詹 康, 霍永久, 等. 苜蓿草颗粒饲料对鹅生长性能、肠道长度和血液生化指标的影响 [J]. *中国农业大学学报(自然科学版)*, 2015, 20(3): 133—138.
- [20] SHERMAN E L, NKURUMAH J D, MOORE S S. Whole Genome Single Nucleotide Polymorphism Associations with Feed Intake and Feed Efficiency in Beef Cattle [J]. *Journal of Animal Science*, 2010, 88(1): 16—22.
- [21] BOLORMAA S, PRYCE J E, HAYES B J, et al. Multivariate Analysis of a Genome-Wide Association Study in Dairy Cattle [J]. *Journal of Dairy Science*, 2010, 93(8): 3818—3833.
- [22] DO D N, OSTERSEN T, STRATHE A B, et al. Genome-Wide Association and Systems Genetic Analyses of Residual Feed Intake, Daily Feed Consumption, Backfat and Weight Gain in Pigs [J]. *BMC Genetics*, 2014, 15(1): 27.
- [23] ZHANG W, YANG B, ZHANG J, et al. Genome-Wide Association Studies for Fatty Acid Metabolic Traits in Five Divergent Pig Populations [J]. *Scientific Reports*, 2016, 6: 24718.
- [24] YUAN J, WANG K, YI G, et al. Genome-Wide Association Studies for Feed Intake and Efficiency in Two Laying Peri-

ods of Chickens [J]. Genetics Selection Evolution, 2015, 47(1): 82.

[25] YI G, SHEN M, YUAN J, et al. Genome-Wide Association Study Dissects Genetic Architecture Underlying Longitudinal Egg Weights in Chickens [J]. BMC Genomics, 2015, 16(1): 746.

[26] HASSON S A, KANE L A, YAMANO K, et al. High-Content Genome-Wide RNAi Screens Identify Regulators of Parkin Upstream of Mitophagy [J]. Nature, 2013, 504(7479): 291–295.

Genome-Wide Association Study (GWAS) for Cecum Length in Laying Hens

WANG Xing-guo, QU Liang, LU Jian,
DOU Tao-cun, MA Meng, HU Yu-ping,
SHEN Man-man, GUO Jun, LI Shang-min, WANG Ke-hua

Jiangsu Institute of Poultry Science, Yangzhou Jiangsu 225125, China

Abstract: **【Objective】** Cecum is a chicken organ for digestion and absorption of nutrients, especially fiber, which plays an important role in the growth and egg production of chicken. So chicken cecum length is an important physiological index. Because studies on cecum length are limited, it is necessary to carry out a study of its genetic structure. **【Methods】** In the current study, the cecum length of the F2s of reciprocal crosses between white leghorn and Dongxiang was measured, the genotypes were detected using the SNP (single-nucleotide polymorphism) chip, the genetic evaluation was detected using the SAS according to the SNP data, and univariate analysis of genome-wide association (GWAS) was conducted with genome-wide efficient mixed model (GEMMA) correlation algorithm. **【Results】** Results showed that the cecum length exhibited a medium heritability (0.39). By GWAS we identified 54 SNPs significantly associated with cecum length, and about 170 Mb on Chromosome 1 was an important area for cecum length. In this region, 18 genes covering 26 SNPs were classified as candidate genes, two of which were in the coding sequence (CDS) and 3' untranslated region (3'UTR), corresponding to the *NHLRC3* and *SIAH3*, respectively. They might be important SNPs and genes which influenced the cecum length in chicken. **【Conclusion】** The SNPs and genes associated with cecum length in chicken were identified by GWAS. It will help to provide the basis for understanding the mechanism of cecum development and molecular breeding in laying hens.

Key words: laying hen; cecum length; GWAS; genetic architecture

责任编辑 夏 娟