绵羊 *HMGA1* 基因的生物信息学分析

靳皓宇，武 顺，权 洁，韩真真，张小雪
（甘肃农业大学动物科学技术学院，甘肃 兰州 730070）

摘要： 高迁移率族蛋白 A1 基因（high mobility group protein A1，*HMGA1*）是一类对 DNA 转录起调控作用的基因，广泛存在于多种动物体内。为探究 *HMGA1* 基因在绵羊体内的功能，对该基因及其编码产物进行了生物信息学分析。结果显示，绵羊 *HMGA1* 基因编码 152 个氨基酸残基，其编码的蛋白质分子式为 $C_{443}H_{759}N_{151}O_{151}S_1$，分子质量为 10.648 35 kDa，等电点 pI 为 10.31。绵羊 HMGA1 蛋白为不稳定蛋白、亲水性蛋白、非分泌蛋白且无信号肽序列和跨膜结构；亚细胞定位主要存在于细胞核（95.7%）。绵羊与牛的 HMGA1 蛋白相似度为 100%。其二级结构和三级结构主要以无规卷曲组成。

关键词： 绵羊；*HMGA1* 基因；生物信息学分析
中图分类号： S826；Q57 **文献标志码：** A **文章编号：** 1001-1463(2022)02-0078-06
doi:10.3969/j.issn.1001-1463.2022.02.019

Bioinformatics Analysis of Sheep *HMGA1* Gene

JIN Haoyu，WU Shun，QUAN Jie，HAN Zhenzhen，ZHANG Xiaoxue
（College of Animal Science and Technology of Gansu Agriculture University，Lanzhou Gansu 730070，China）

Abstract： High mobility group protein A1（*HMGA1*）gene is a kind of gene that regulates DNA transcription and exists widely in many kinds of animals. In order to explore the function of *HMGA1* gene in sheep，bioinformatics analysis of *HMGA1* gene and its encoding product was conducted in this study. The results showed that sheep *HMGA1* gene encoded 152 amino acid residues，and its protein molecular formula was $C_{443}H_{759}N_{151}O_{151}S_1$，molecular weight was 10.64 835 Kda，isoelectric point pI was 10.31. Sheep HMGA1 protein was an unstable protein，hydrophilic protein，non-secretory protein and has no signal peptide sequence and transmembrane structure. The localization of HMGA1 at a subcellular level was mainly existed in the nucleus（95.7%）. The HMGA1 protein similarity between sheep and cattle was 100%. Its secondary and tertiary structures were mainly composed of random crimp.

Key words： Sheep；*HMGA1* gene；Bioinformatics analysis

高迁移率族蛋白（high mobility group-protein，HMG）是 Goodwin 等[1]在 1973 年于牛胸腺细胞中首次发现的一类染色质相关非组蛋白。Bianchi 等[2]将 HMG 蛋白分为 3 类：高迁移率族蛋白 B（high mobility group-proteinB，HMGB）、高迁移率族蛋白 A（high mobility group-proteinA，HMGA）和高迁移率族蛋白 H（high mobility group-proteinH，HMGH）。而HMGA 又可分为两大类：即高迁移率族蛋白 A1

收稿日期： 2021-10-21
基金项目： 甘肃农业大学学生科研训练计划项目（202104026）。
作者简介： 靳皓宇（2000—），男，山西晋城人，本科在读，研究方向为动物科学。联系电话：（0）18235626912。Email：1073959479@qq.com。
通信作者： 张小雪（1984—），女，湖北武汉人，副教授，主要从事动物遗传育种与繁殖研究及教学工作。联系电话：（0931）7631225。Email：zhangxx@gsau.edu.cn。

版），2008，37(4)：415-419.
[13] 王秀兰，包玉海. 土地利用动态变化研究方法探讨[J]. 地理科学进展，1999，18(1)：83.
[14] 史培军，官 鹏. 土地利用/覆盖变化研究的方法和实践[M]. 北京：科学出版社，2000.
[15] 陈述彭. 遥感信息机理研究[M]. 北京：科学出版社，1998.

(high mobility group-proteinA1，HMGA1)和高迁移率族蛋白 A2(high mobility group-proteinA2，HMGA2)，分别由 *HMGA1* 基因、*HMGA2* 基因编码。由于 *HMGA1* 基因转录后的可变剪接，HMGA1 又分为长度有所不同的高迁移率族蛋白 A1a(high mobility group-proteinA1a，HMGA1a)、高迁移率族蛋白 A1b(high mobility group-proteinA1b，HMGA1b)以及高迁移率族蛋白 A1c（high mobility group-proteinA1c，HMGA1c)3 种蛋白[3]。有关哺乳动物 *HMGA1* 基因在癌症细胞增殖和分化方面的研究较多，并且可能成为检测肿瘤标志物以及预测肿瘤预后的良好指标，针对 *HMGA1* 的治疗亦可能成为治疗肿瘤的新途径[4]。小鼠 *HMGA1* 基因对白色脂肪和棕色脂肪的生成起重要作用[5]，另有研究报道称猪 *HMGA1* 基因和黑皮质素受体 4(melanocortin receptor 4，MC4R)基因的 SNP 互作与其生长、瘦肉量以及脂肪含量均有一定关系[6-8]。有研究表明，猪 *HMGA1* 基因能提高软骨细胞增殖活性[9]，并通过调控透明软骨细胞增殖和分化来影响软骨组织的修复[10]，从而影响骨骼的结实度。目前针对 *HMGA1* 基因及其编码产物的研究大多集中于人类和猪，而针对绵羊的研究较少。我们通过调取生物基因组数据库中绵羊 *HMGA1* 的序列，利用生物信息学方法对绵羊 *HMGA1* 基因及其编码产物的序列、理化性质、蛋白质结构和生物学功能等进行了研究，以期为进一步研究该基因的结构和生物学功能提供参考。

1 材料与方法

1.1 序列来源

序列数据来源于 NCBI 网站 GeneBank 数据库，包括绵羊(XM_027958418.2、XP_027814219.1)、人(NM_001319077.2、NP_001306006.1)、家鼠(NM_001025427.3、NP_001020598.1)、鸡(NM_204369.1、NP_989700.1)、牛(NM_001076523.1、NP_001069991.1)、海龟(XM_037885374.1、XP_037741302.1)、黑猩猩（NM_001246496.1、NP_001233425.1)、猕猴（NM_001266352.1、NP_001253281.1)、猪(NM_001185154.1、NP_001172083.1)等 9 个物种的 mRNA 序列和氨基酸序列。括号内为 GeneBank 登录号。

1.2 方法

绵羊 *HMGA1* 基因开放阅读框（Open reading frame，ORF）分析采用 NCBI 的 ORF Finder 程序；理化性质分析采用 Bioedit 及 ExPASy 分析软件；绵羊 HMGA1 蛋白亚细胞定位采用 PSORTⅡ软件，蛋白潜在信号肽剪切位点预测采用 SignalP3.0 软件，蛋白跨膜螺旋区域预测采用 TMHMM 程序，蛋白保守结构域分析采用 Smart 软件，蛋白亲疏水性分析采用 Prot Scale 软件，二级结构预测采用 Jpred 软件，蛋白三级结构预测采用 Swiss-model 软件。多序列比对及同源性分析采用 DNAMAN 软件。

2 结果与分析

2.1 绵羊 HMGA1 的理化性质分析

蛋白质的基本性质包括相对分子质量、氨基酸组成、等电点基因编码产物半衰期和不稳定指数等[11]。利用 Prot Param 在线工具和 Bioedit 软件对绵羊 *HMGA1* 基因编码产物的理化性质进行分析，结果表明，其编码产物的分子式为 $C_{443}H_{759}N_{151}O_{151}S_1$，共含有 1 505 个原子。该编码产物氨基酸残基数为152 个，分子质量为 10.648 35 kDa，理论等电点 pI 为 10.31，提示绵羊 HMGA1 蛋白呈碱性。其氨基酸组成如图 1 所示，其中含量最多的是 Lys(赖氨酸)，所占比例为 16.67%。绵羊 *HMGA1* 基因编码产物中不含 Cys(半胱氨酸)、Phe(苯丙氨酸)、His(组氨酸)、Trp(色氨酸)、Tyr(酪氨酸)，负电荷残基总数(Asp + Glu)为 15，正电荷残基总数(Arg + Lys)为 27。该编码产物在哺乳动物体外的半衰期为 30 h，不稳定指数为 81.81，不稳定指数为 81.81>40，可确定该蛋白属于不稳定蛋白。

2.2 绵羊 *HMGA1* 基因开放阅读框分析

开放阅读框(Open Reading Frame，ORF)从起始密码子开始，是 DNA 序列中具有编码蛋白质潜能的序列，结束于终止密码子连续的碱基序列[12]。由 ORF 分析可知，该序列最大的开放阅读框长度

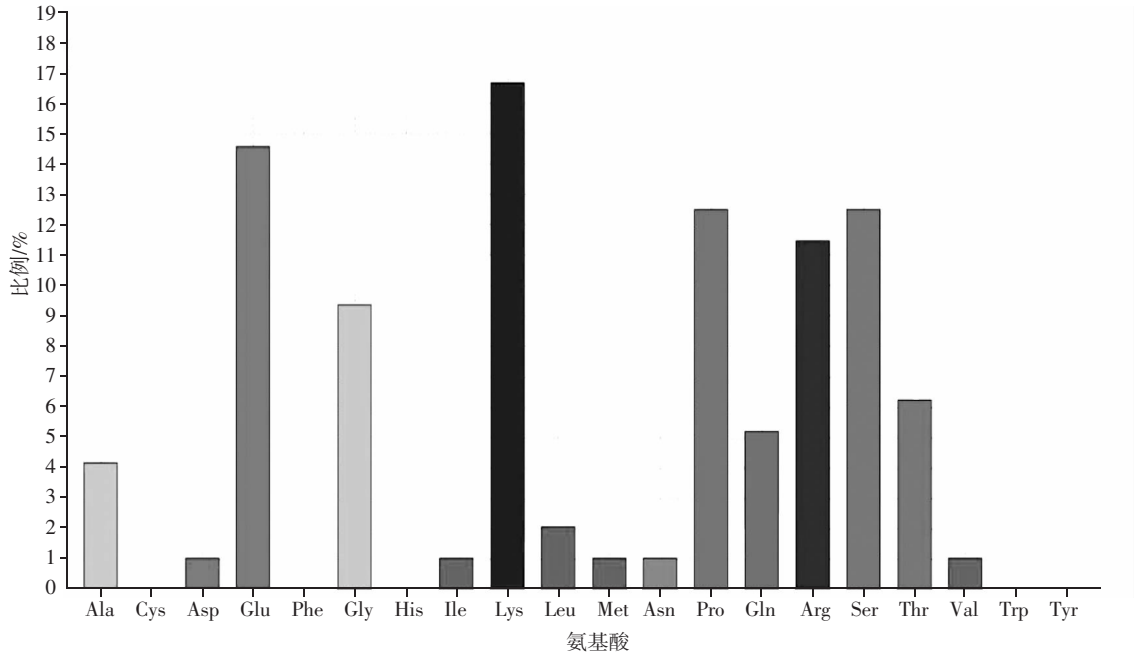

图 1 绵羊 *HMGA1* 基因编码的蛋白氨基酸组成

为 459 bp（起始密码子位于 101 bp 处，终止密码子位于 559 bp 处），编码了 152 个氨基酸残基(图2)。

2.3 绵羊 HMGA1 蛋白亚细胞定位

从亚细胞定位结果可知，绵羊 HMGA1 蛋白分布于细胞核的可能性为 95.7%，分布于细胞质的可能性为 4.3%。由此推断，绵羊 *HMGA1* 基因编码产物主要在细胞核中发挥生物学作用。

2.4 不同物种 HMGA1 蛋白的同源性分析

将绵羊 HMGA1 蛋白序列与另外一些已发表的动物如人、绵羊、家鼠、鸡、牛、海龟、黑猩猩、猕猴和猪 9 种动物的蛋白序列构建系统进化树，采用 DNAMAN 对 9 个物种的 HMGA1 蛋白的氨基酸序列进行多序列比对，结果如图 3 所示。可以看出，*HMGA1* 基因在这 9 个物种中均有表达。系统发育分析结果(图4)表明，人、黑猩猩、猕猴的

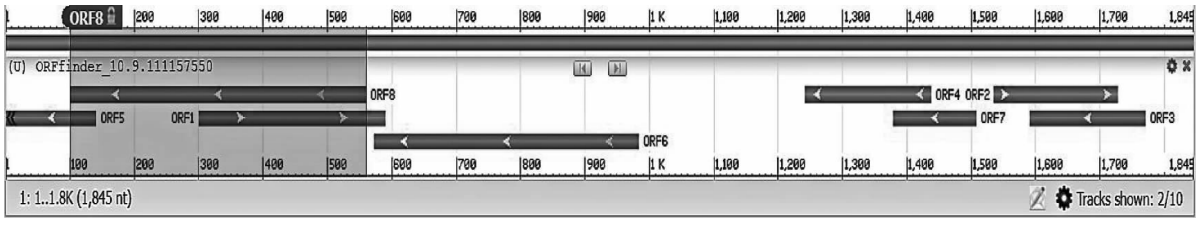

图 2 绵羊 *HMGA1* 基因序列的 ORF 分析

图 3 9 个物种的 *HMGA1* 基因编码产物序列的同源性分析

HMGA1 氨基酸序列同源性为 100%，绵羊与牛同源性高达 100%，说明人、黑猩猩、猕猴的亲缘关系较近，绵羊和牛的亲缘关系比较近，绵羊与鸡和海龟的同源性较低。

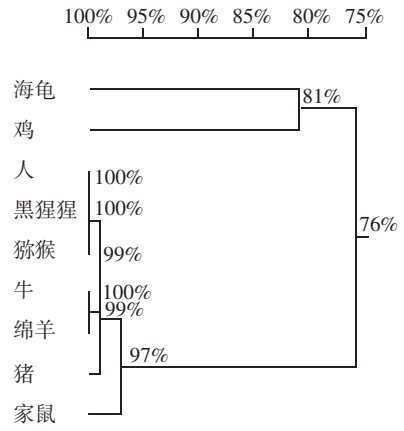

图 4　9 个物种的 *HMGA1* 基因编码产物序列的同源树

2.5　绵羊 HMGA1 蛋白潜在信号肽剪切位点预测

信号肽序列是存在于分泌蛋白基因编码序列中、在起始密码子之后的 1 段富含疏水氨基酸多肽的序列。蛋白质的合成在核糖体上进行，信号肽是蛋白质的一个片段，引导核糖体并定位于内质网上的一个通道上，使核糖体附着在内质网上，并将不断伸长的蛋白质链穿透通过通道，随后信号肽被切割下来，合成完成的蛋白质被释放进内质网腔，最后被转运到胞外[13]。通过检测绵羊 HMGA1 蛋白潜在信号肽的存在情况可以得知该基因编的产物是否是分泌蛋白和跨膜蛋白以及跨膜蛋白的基本信息。从图 5 可以看出，绵羊 *HMGA1* 基因编码产物的 C 值为 0.039、Y 值为 0.020、S 值为 0.064。因此推断 *HMGA1* 基因的编码产物不存在信号肽，且未发现跨膜区，所以该蛋白既为非分泌性蛋白，也为非跨膜蛋白，主要位于细胞核内。

2.6　绵羊 HMGA1 蛋白跨膜螺旋结构预测

采用 TMHMM 在线软件对于绵羊 HMGA1 蛋白跨膜螺旋结构进行分析预测，结果表明，绵羊 *HMGA1* 基因编码的蛋白质是非跨膜蛋白，即无跨膜结构(图6)。

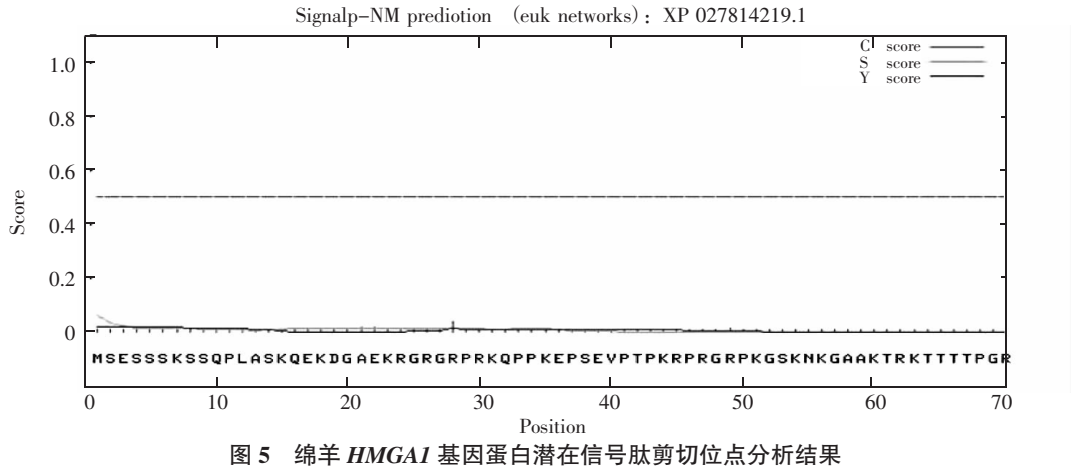

图 5　绵羊 *HMGA1* 基因蛋白潜在信号肽剪切位点分析结果

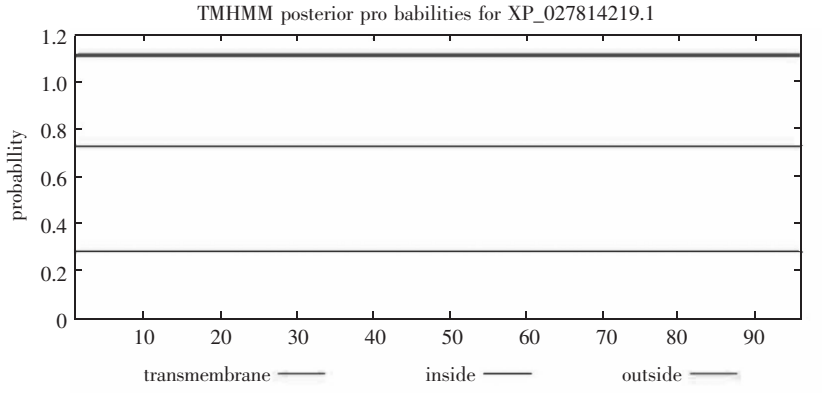

图 6　绵羊 *HMGA1* 基因蛋白跨膜螺旋结构分析结果

2.7 绵羊 HMGA1 蛋白保守结构域分析

由绵羊 HMGA1 蛋白潜在信号肽剪切位点预测已知绵羊 *HMGA1* 基因编码的蛋白是非跨膜蛋白，进而通过 Smart 软件分析，绵羊 HMGA1 蛋白没有跨膜结构。但是存在 3 个低复杂性区域，分别是 22~37、43~57、60~95(图 7；表 1)。

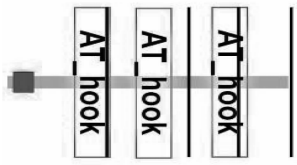

图 7 绵羊 HMGA1 蛋白保守结构域分析数据

表 1 绵羊 HMGA1 蛋白保守结构域

名称	起始位点	终止位点	E值
Low complexity	22	37	N/A
Low complexity	43	57	N/A
Low complexity	60	95	N/A

2.8 绵羊 HMGA1 蛋白亲疏水性分析

利用 Prot Scale 对绵羊 HMGA1 蛋白亲疏水性分析，结果表明，该基因编码蛋白疏水性的值均小于 0，最小值为 -3.200，总平均亲水性系数约为 -2.061，由此得出该基因编码的蛋白属于亲水蛋白(图8)。

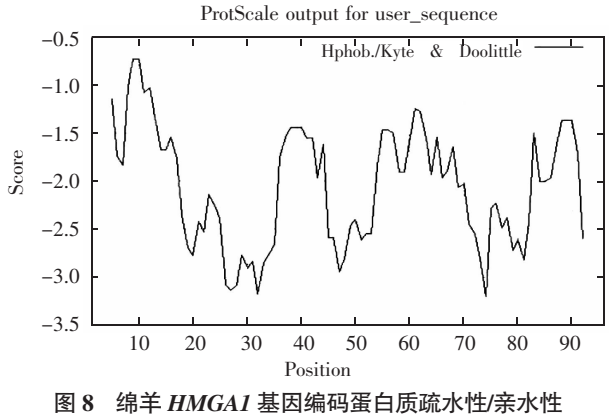

图 8 绵羊 *HMGA1* 基因编码蛋白质疏水性/亲水性预测分析

2.9 绵羊 HMGA1 蛋白二级结构的预测

蛋白质高级结构决定了它行使的生物功能，其二级结构是指在它的多肽链中具有规则重复的构象[14]。二级结构(secondary structure)是蛋白质第一个水平的折叠，即部分蛋白质链折叠形成一些通用结构，这些通用结构在所有蛋白质中都能找到[15]。通过 Jpred 软件分析可知（图9），绵羊 HMGA1 蛋白二级结构全部为无规卷曲结构。

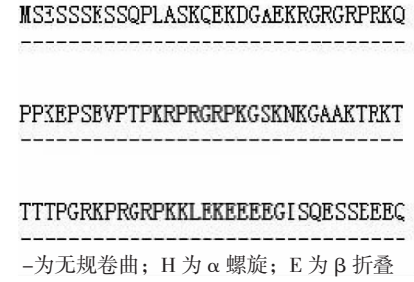

图 9 绵羊 *HMGA1* 基因编码蛋白质二级结构预测

2.10 绵羊 HMGA1 蛋白三级结构预测与分析

三级结构(tertiary structure)是指蛋白质在二级结构基础上的进一步折叠，通过将二级结构元素组装在一起形成每个蛋白质特有的三维构象[15-16]。由图 10 可知，绵羊 *HMGA1* 基因编码蛋白的三级结构主要由无规卷曲折叠缠绕形成。

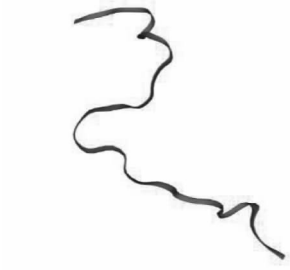

图 10 绵羊 *HMGA1* 基因编码蛋白的三级结构分析结果

3 小结与讨论

绵羊 *HMGA1* 基因的最长 ORF 长度为 459 bp，编码 152 个氨基酸残基；赖氨酸所占比例最多，为 16.67%，分子质量为 10.648 35 kDa，理论等电点（pI)为 10.31。*HMGA1* 基因编码的产物为不稳定性蛋白。其亚细胞定位的结果显示在细胞核的可能性最大，为 95.7%。*HMGA1* 在多种物种中氨基酸序列高度相似，绵羊与牛在系统发育树中距离最近。*HMGA1* 基因的编码产物中不存在信号肽，为非分泌性蛋白且没有跨膜区，故该蛋白是非跨膜蛋白。*HMGA1* 编码的蛋白为亲水蛋白，富含亲水氨基酸。绵羊 *HMGA1* 基因编码产物的二级结构主要以无规卷曲组成，三级结构主要由无规卷曲折叠缠绕形成。

在对猪全基因组序列数据对体尺性状进行全基因组关联研究中发现，*HMGA1* 基因是重要影响因子[17]。蒲蕾等[18]的研究表明，大白猪和民猪

HMGA1 基因的 g.543 T>C、g.1356 C>T 2 个突变位点与猪的生长、饲料利用性状相关。对猪群 *HMGA1* 基因突变位点检测试验的结果显示，T 等位基因型可以增加个体的体长[19]。在 *HMGA1* 基因与湖羊尾脂肪沉积的相关性分析中，得出了 *HMGA1* 基因可作为湖羊尾脂重量标记辅助选择的遗传标记的结论[20]。这些研究都表明，*HMGA1* 基因在提高生长性状上具有较大潜力，因此 *HMGA1* 基因可作为选种依据。此外，*HMGA1* 基因还可作为 RAD51 的新型调节器，在胆管癌中具有抗辐射作用[21]，也可作为一种潜在的胃癌治疗预后生物标志物[22]。

参考文献：

[1] GOODWIN G H，SANDERS C，JOHNS E W. A new group of chromatin-associated proteins with a high content of acidic and basic amino acids[J]. European Journal of Biochemistry，1973，38(1)：14-19.

[2] BIANCHI M E，BELTRAME M. Upwardly mobile proteins. Workshop：the role of HMG proteins in chromatin structure，gene expression and neoplasia[J]. EMBO Reports，2000，1(2)：109.

[3] 肖湘文，周建林，周 畅. 高迁移率族蛋白[J]. 细胞生物学杂志，2006(4)：501-506.

[4] 杨 霞，李国华. *HMGA1* 与肿瘤的研究进展[J]. 江西医药，2009，44(1)：79-82.

[5] ARCE-CEREZOALTAMIRA，GARCÍAMIQUEL，RODRÍGUEZ-NUEVO AIDA，et al. HMGA1 overexpression in adipose tissue impairs adipogenesis and prevents diet-induced obesity and insulin resistance[J]. Scientific Reports，2015，5：14487.

[6] 巩建飞，岳静伟，赵玲玲，等. 猪 *HMGA1* 基因的组织表达及其多态性与背膘厚的关联分析[J]. 畜牧兽医学报，2019，50(7)：1340-1346.

[7] KIM K S，LEE J J，SHIN H Y，et al. Association of melanocortin 4 receptor (MC4R) and high mobility group AT-hook 1 (HMGA1) polymorphisms with pig growth and fat deposition traits[J]. Animal Genetics，2006，37(4)：419-421.

[8] HONG JOONKI，KIMDUWAN，CHOKYUHO，et al. Effects of genetic variants for the swine FABP3，HMGA1，MC4R，IGF2，and FABP4 genes on fatty acid composition[J]. Meat Science，2015，110：46-51.

[9] CLEYNENISABELLE，VAN DE VENWIM J M. The HMGA proteins： a myriad of functions (Review)[J]. Int J Oncology，2008，32(2)：289-305.

[10] RICHTER ANDREAS，HAUSCHILDGREGOR，MURUA ESCOBAR HUGO，et al. Application of high-mobility-group-A proteins increases the proliferative activity of chondrocytes in vitro[J]. Tissue engineering. Part A，2009，15(3)：473-7.

[11] 张小雪，李发弟，王维民. 绵羊 *ANXA10* 基因生物信息学分析[J]. 甘肃农业科技，2016(6)：1-4.

[12] SLIGHTOM J L，DURAND-TARDIF M，JOUANIN L，et al. Nucleotide sequence analysis of TL-DNA of Agrobacterium rhizogenesagropine type plasmid. Identification of open reading frames[J]. The Journal of Biological Chemistry，1986，261(1)：108-121.

[13] 叶方寅. 信号肽假说的提出及证实[J]. 国外医学分子生物学分册，1999，21(6)：377-379.

[14] 陈 珂，张 君，刘嘉斐，等. 绿豆 AP2/ERF 转录因子家族的生物信息学鉴定与特征分析 [J]. 分子植物育种，2020，18(20)：6605-6617.

[15] 孔繁良. 基于二级结构的蛋白质三级结构预测[D]. 济南：济南大学，2016.

[16] 刘 佳，王 祯，代友超，等. 绵羊 *HTR4* 基因的生物信息学分析[J]. 甘肃农业科技，2020(10)：35-40.

[17] LIU HUATAO，SONGHAILIANG，JIANGYIFAN，et al. A single-step genome wide association study on body size traits using imputation-based whole-genome sequence data in yorkshire pigs[J]. Frontiers in Genetics，2021，12：629049.

[18] 蒲 蕾，刘 欣，岳静伟，等. 杜洛克猪 *HMGA1* 基因多态位点与生长、饲料利用性状的关联分析[J]. 中国畜牧兽医，2016，43(12)：3268-3274.

[19] 杨广礼，章 焕，田慧月，等. 野猪、松辽黑母猪及其杂交一代多性状功能基因分子遗传标记研究[J]. 中国畜牧兽医，2021，48(3)：946-953.

[20] WANG JIANGHUI，ZHANGXIAOXUE，WANGXIAOJUAN et al. Polymorphism and expression of the HMGA1 gene and association with tail fat deposition in Hu sheep[J]. AnimBiotechnol，2021，undefined：1-9.

[21] SONG JIANPING，CUIDONGHAI，WANG JING et al. Overexpression of HMGA1 confers radioresistance by transactivating RAD51 in cholangiocarcinoma [J]. Cell Death Discov，2021，7：322.

[22] YANG QIONG，WANGYUSI，LIMENGSHU et al. HMGA1 promotes gastric cancer growth and metastasis by transactivating SUZ12 and CCDC43 expression [J]. Aging(Albany NY)，2021，13：16043-16061.