

绵羊 *RXRG* 基因的生物信息学分析

张司龙¹, 张小雪^{1,2}, 宋其志¹, 王维民¹

(1. 甘肃农业大学动物科学技术学院, 甘肃 兰州 730070; 2. 甘肃省肉羊繁育生物技术工程实验室, 甘肃 民勤 733300)

摘要: 以绵羊 *RXRG* 基因为目的基因, 利用生物信息学软件预测其结构和功能。结果表明, 绵羊 *RXRG* 基因编码 463 个氨基酸, 开放阅读框长度为 1 392 bp, 起始密码子位于 228 bp 处, 终止密码子位于 1 619 bp 处。*RXRG* 基因编码蛋白的相对分子质量为 50 845.19 Da, 等电点为 7.55, 在氨基酸组成中亮氨酸所占比率最高, 色氨酸占比最低。亚细胞定位主要位于细胞核中, 不属于分泌蛋白; 不存在信号肽序列; 存在两个保守结构域, 并且为疏水性蛋白, 二级结构主要以 α 螺旋和无规则卷曲为主, 三级结构主要由无规卷曲缠绕折叠形成。

关键词: 绵羊; *RXRG* 基因; 生物信息学分析

中图分类号: S826 **文献标志码:** A **文章编号:** 1001-1463(2020)02-0031-07

doi: 10.3969/j.issn.1001-1463.2020.02.008

Bioinformatics Analysis of *RXRG* Gene in Sheep

ZHANG Silong¹, ZHANG Xiaoxue^{1,2}, SONG Qizhi¹, WANG Weimin¹

(1. College of Animal Science and Technology, Gansu Agricultural University, Lanzhou Gansu 730070, China; 2. Engineering Laboratory of Mutton Sheep Breeding and Reproduction Biotechnology in Gansu Province, Minqin Gansu 733300, China)

Abstract: In this study, with the sheep *RXRG* gene as the target gene, the structure and function of sheep *RXRG* gene were predicted by bioinformatics software. The *RXRG* gene of sheep encodes 463 amino acids, the length of open reading frame is 1 392 bp, the initial codon is 228 bp, and the termination codon is 1 619 bp. The relative molecular weight and isoelectric point of *RXRG* protein were 50 845.19 Da and 7.55 respectively. Leucine accounted for the highest proportion of amino acid composition and tryptophan accounted for the lowest proportion. Subcellular localization is mainly located in the nucleus, not in the secretory protein; there is no signal peptide sequence; there are two conservative domains, and they are hydrophobic proteins. The secondary structure is mainly α helix and irregular curl, and the third structure is mainly formed by random curl winding and folding.

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

视黄酸是一种脂溶性的小分子物质, 在细胞分化、上皮细胞生长、视觉和组织维持、胎儿发育和繁殖等过程中发挥着重要作用^[1]。视黄酸受体 (Retinoic acid receptor,

RAR) 和视黄素 X 受体 (Retinoid X receptor, RXR) 是视黄酸受体家族的主要成员, RAR、RXR 由 3 种不同的基因 *RXRA*、*RXRB* 和 *RXRG* 编码, 形成了 *RXRA*、*RXRB*、*RXRG*

收稿日期: 2019-10-22

基金项目: 甘肃农业大学科技创新基金(GAU-XKJS-2018-023)。

作者简介: 张司龙(1997—), 男, 甘肃敦煌人, 本科在读。联系电话: (0)18809482521。Email: 1349989828@qq.com。

通信作者: 王维民(1984—), 男, 湖北武汉人, 副教授, 主要从事动物遗传育种与繁殖工作。联系电话: (0931)7631225。Email: wangwm@gsau.edu.cn。

等多种类型的受体^[2-3]。*RXRG* 基因作为配体激活的转录因子，结合到靶基因的特异性应答序列上，调节基因的转录表达^[2]，是细胞分化和组织形态发生的主要调节因子。*RXRG* 基因的研究多在黑猩猩、猕猴、狗、奶牛、大鼠、鸡、斑马鱼、青蛙等方面，而关于绵羊 *RXRG* 基因的研究较少。国内外研究表明，*RXRG* 基因不仅在动物妊娠的关键时期表达，还对具有高产性能猪种的窝产仔数具有明显的加性效应，对于高繁殖力的绵羊品种具有显著影响^[4]。因此，对 *RXRG* 基因进行深入研究，利用生物信息学相关软件和工具，对不同物种 *RXRG* 基因系统发育和绵羊 *RXRG* 基因 CDS 区序列蛋白质的理化性质、二级结构及多参数预测、蛋白质跨膜结构、信号肽预测、亚细胞定位和三级结构等进行分析，旨在寻找与双胎或者多胎性状有关的遗传标记位点，为培育高繁殖力绵羊品系提供理论依据，为绵羊 *RXRG* 基因结构与功能研究提供参考^[5]。

1 材料与方法

1.1 序列来源

数据资料来源于 NCBI 网站的 GenBank 数据库，包括绵羊(XM_012185408.2)、瘤牛(XM_019986703.1)、黑猩猩(XM_513962.5)、人(NM_006917.5)、家鼠(NM_009107.3)、猪(NM_001130213.1)、马(XM_023640805.1)和鸡(NM_205294.1)8个物种的 mRNA 序列(括号内为Gen Bank登录号)。

1.2 方法

绵羊 *RXRG* 基因开放阅读框采用 NCBI

的开放阅读框查找器程序分析；采用 DNA 生物编辑器(Bioedit及DNA Star)分析软件预测 *RXRG* 基因编码产物的理化性质；多序列比对及同源性分析采用分子生物学综合应用软件(DNAMAN)；采用蛋白质亚细胞定位预测工具(PSORT)进行亚细胞定位。在蛋白质的分析预测方面采用信号肽预测工具(Signalp3.0软件)进行蛋白潜在信号肽剪切位点的预测，采用捆绑混合隐马尔可夫模型[tied-mixture hidden Markov models(TMhmm)]程序进行跨膜螺旋区域的预测，采用简单模块化架构研究工具(SMART)软件进行蛋白保守结构域分析；二级结构采用蛋白质二级结构预测服务器(Jpred)分析预测；采用瑞士模型(Swiss-model)软件分析蛋白三级结构；在蛋白亲疏水性分析方面采用疏水性分析(ProtScale)程序^[4-6]。

2 结果与分析

2.1 绵羊 *RXRG* 基因开放阅读框

开放阅读框的分析结果如图 1 所示，绵羊 *RXRG* 基因推测编码 463 个氨基酸残基，其基因序列长度为 1 392 bp，起始密码子位于 228 bp 处，终止密码子位于 1 619 bp 处。

2.2 绵羊 *RXRG* 编码产物的理化性质

蛋白质的理化性质分析包括对其相对分子质量、氨基酸组成以及等电点的理化性质分析^[7]。绵羊 *RXRG* 基因编码产物的理化性质结果(图 2)表明，该基因共编码 463 个氨基酸残基，相对分子质量为 50 845.19 Da，其氨基酸残基中亮氨酸所占比重最高，为 9.9%；其次为丝氨酸，所占比例为 9.5%，

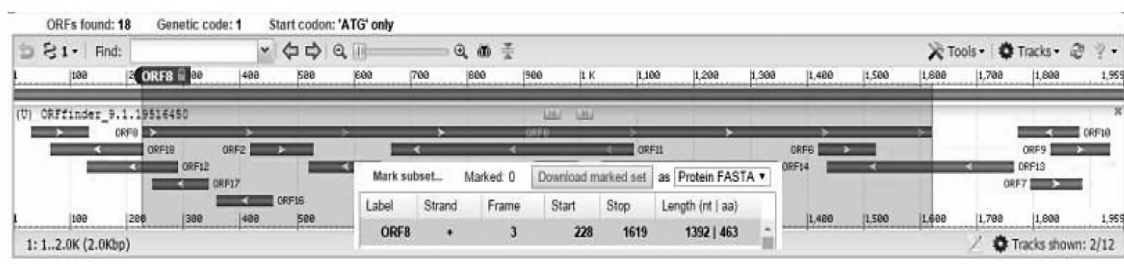


图 1 *RXRG* 基因开放阅读框

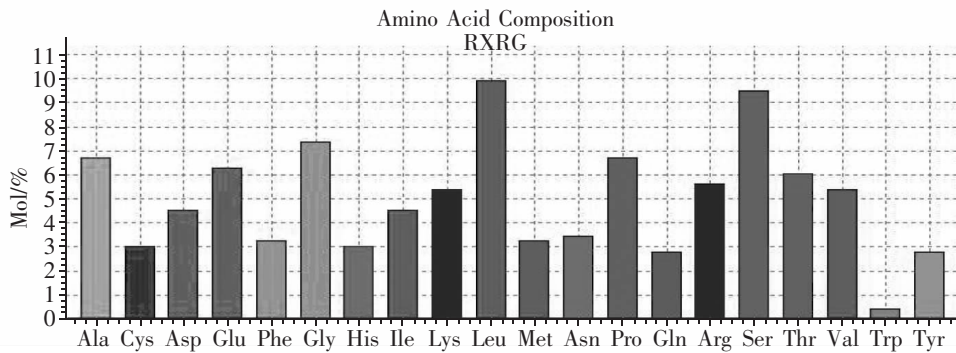


图 2 RXRG 基因编码产物的理化性质

甘氨酸所占比例为 7.3%；色氨酸占比最低，仅为 0.4%。

2.3 绵羊 RXRG 基因蛋白亚细胞定位

从蛋白质亚细胞定位预测工具程序预测得到的绵羊 RXRG 基因蛋白亚细胞定位结果(表1)可以看出，绵羊 RXRG 蛋白的亚细胞大多数分布在细胞核，其可能性为 78.3%。而在其他位置分布的可能性所占概率较小。其中，分布于细胞质的可能性为 8.7%，分布于线粒体的可能性为 4.3%，还有 4.3%分布在过氧化物酶体。由此推断，绵羊 RXRG 基因主要在细胞核中发挥生物学作用，其次在细胞质中发挥作用。

表 1 RXRG 编码产物亚细胞定位预测

亚细胞定位	概率 /%
细胞核	78.3
细胞质	8.7
线粒体	4.3
过氧化物酶体	4.3
细胞外,包括细胞壁	4.3

2.4 绵羊 RXRG 编码产物序列同源性

通过 NCBI 数据库找到绵羊、瘤牛、猪、马、黑猩猩、人、家鼠、鸡八种动物 DNA 序列，并对其进行同源性分析(图3)，可以看出，绵羊与他们之间的亲源性分别为 98%、93%、92%、91%、91%、86%、73%。由系统发育树可看出在这 8 种动物中绵羊 RXRG 基因与瘤牛亲缘关系较近，与鸡的亲

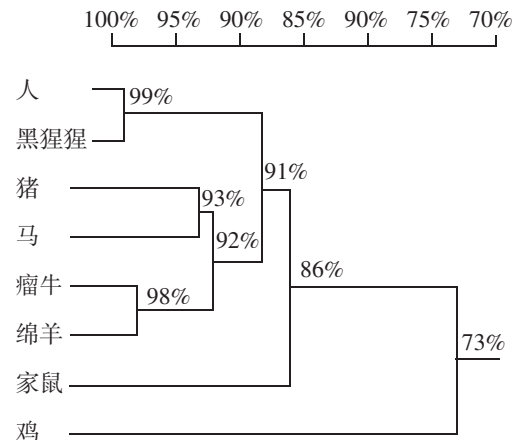


图 3 8 种动物 RXRG 编码产物发育树

缘关系最远，RXRG 基因在哺乳动物间的保守性较高(图4)。

2.5 绵羊 RXRG 蛋白潜在信号肽剪切位点预测

通过分析预测绵羊 RXRG 蛋白潜在信号肽剪切位点的情况可以得知基因编码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，信号肽序列是存在于分泌蛋白基因编码序列中起始密码子之后的一段富含疏水氨基酸多肽的序列，绵羊 RXRG 蛋白氨基酸在 23 的位置出现最大 C 值为 0.119，截止点为 0.32；在 1 的位置出现最大 S 值为 0.824，截止点为 0.87；在 1-19 的位置平均出现 S 值为 0.560，截止点为 0.48；在 20 的位置出现最大 Y 值为 0.178，截止点为 0.33；在 1-19 的位置出现 D 值为 0.369，截止点为 0.43 因此推断绵羊 RXRG 基因的编码产物属于分泌

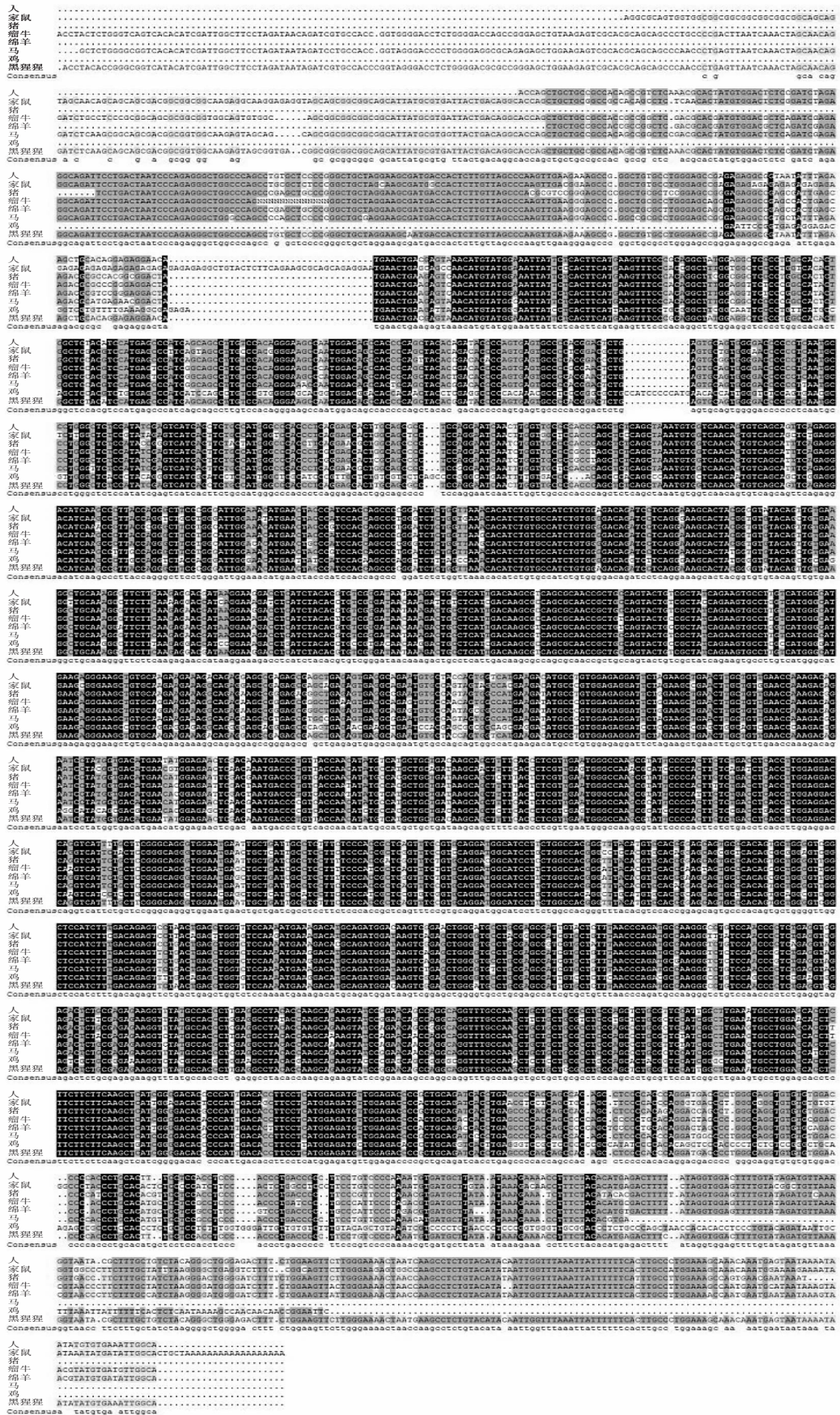


图4 8种动物RXRG基因同源性分析

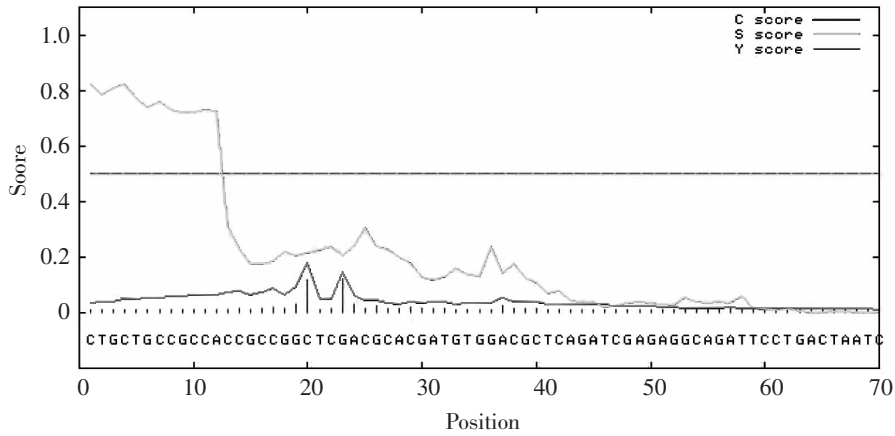


图5 绵羊RXRG基因蛋白潜在信号肽剪切位点分析结果

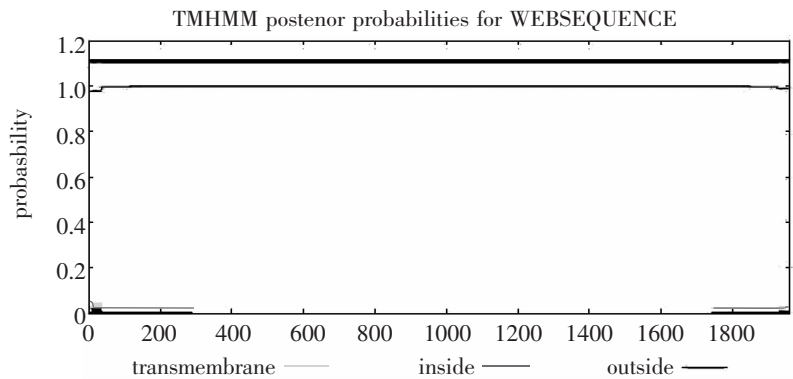


图6 绵羊RXRG基因蛋白跨膜螺旋结构分析结果

性蛋白。位置19和20之间最可能的裂解位为CGG-CT。

2.6 绵羊RXRG蛋白跨膜区域结构预测

从图6可以看出，通过使用TMHMM软件对绵羊RXRG蛋白跨膜区域结构进行预测分析，发现该基因所编码的蛋白没有跨膜区域结构。

2.7 绵羊RXRG蛋白保守结构域分析

由Smart软件分析结果得出，绵羊RXRG基因在第136~207位、第271~430位氨基酸残基之间存在跨膜区，在76~89的位置上出现了低复杂性区域(图7)。

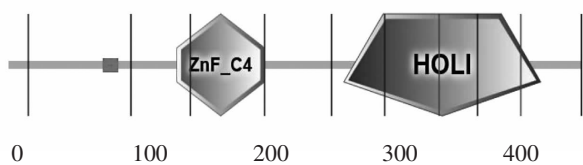


图7 绵羊RXRG蛋白保守结构域

2.8 绵羊RXRG蛋白亲疏水性分析

从图8看出，蛋白质亲疏水性研究发现该基因编码蛋白疏水性达到最大值的位点有很多，且最大值为4.000，其中Arg、Leu、Ser的单个值最高，达到了6.000。此外，亲

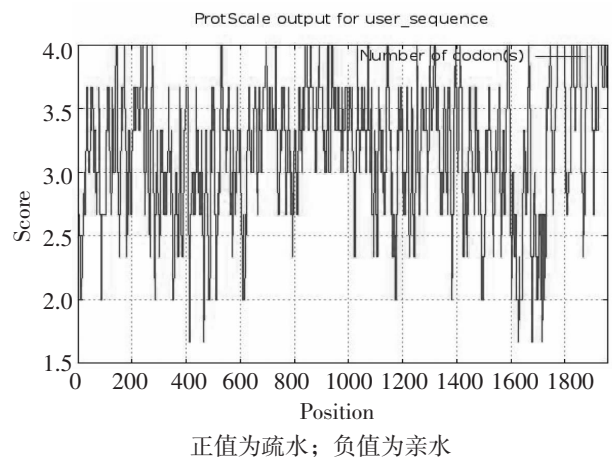


图8 绵羊RXRG基因编码蛋白质疏水性/亲水性预测分析

疏水性分析图中最小值为 1.667, 分析其基因得到的结果全部为正值, 可见其并无亲水性, 从而得出该基因编码的蛋白属于疏水性蛋白。

2.9 绵羊 RXRG 蛋白二级结构的预测

通过 Jpred 软件进行蛋白质二级结构的预测分析(图9)可知, 绵羊 RXRG 蛋白二级结构中有大部分结构为无规卷曲, 其次为 α 螺旋, β 折叠数量较少。其结果为 10 个 α 螺旋, 4 个 β 折叠, 其余全是无规卷曲结构。

2.10 绵羊 RXRG 蛋白三级结构预测与分析

由图 10 可知, RXRG 基因编码蛋白的三级结构主要由无规卷曲折叠缠绕形成。

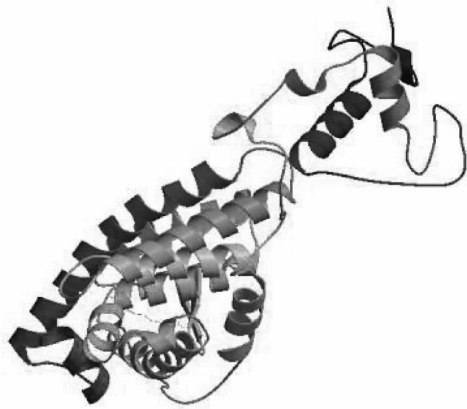


图 10 绵羊 RXRG 蛋白三级结构的分析结果

3 小结与讨论

绵羊 RXRG 基因的 ORF 长度为 1 392 bp, 编码 463 个氨基酸残基, 起始密码子位于 228 bp 处, 终止密码子位于 1 619 bp 处。

对其理化性质分析发现组成最多的氨基酸是亮氨酸, 所占比例为 9.9%, 其次, 丝氨酸占比为 9.5%, 色氨酸组成所占比例最少, 为 0.4%。在系统发育树中绵羊 RXRG 基因同源性与瘤牛关系最近, 为 98%, 其次为猪和马, 在所分析的动物中绵羊与鸡的亲缘关系最远。RXRG 基因编码蛋白潜在信号肽剪切位点的预测最大 C 值为 0.119, 截止点为 0.32; 最大 S 值为 0.824, 截止点为 0.87; 最大 Y 值为 0.178, 截止点为 0.33; 在 1-19 的位置平均出现 S 值为 0.560, 截止点为 0.48; 在 1-19 的位置出现 D 值为 0.369, 截止点为 0.43; 位置 19 和 20 之间最可能的裂解位点是 CGG-CT。绵羊 RXRG 基因编码蛋白质为疏水性蛋白, 其疏水性最大值为 4.000, 最小值为 1.667, 该基因编码的蛋白没有跨膜螺旋结构, α 螺旋为主, 三级结构主要为无规卷曲, 绵羊 RXRG 蛋白的亚细胞分布在细胞核的可能性最大, 为 78.3%。

RXRG 基因是调节细胞分化和组织形态发生的重要因子, 在提高羊生产力方面有重要影响, 李华振等^[4,8] 在绵羊繁殖轴相关组织 RXRs 基因的表达及 RXRA 基因多态性与季节性发情和绵羊基因多态性与产羔数方面进行了研究, Shaoze Cheng 等^[9]在 RXRG 调节鸡原始生殖细胞分化方面做过研究。此外, 黄萌等^[1-2,10]做过牛 RXRG 基因遗传变异与双胎性状的关联性分析方面已有研究。

```

MYGNYSHFMKFPTGFGGSPGHTGSTSMSPSAALSTGKPMDSHPSTDTTPVSAPRTLSAVGTPLNALG
SPYRVITSAMGPPSGALAAPPGLNLVAPPSSQLNVVNSVSI SEDIKPLPGLPGIGHMNCPTSPGSL
-----
VKHICAI CGDRSSGKHVYVYCEGCKGFFKRTIRKDLIYTCRDNDKCLIDKRQRNRCQYCRYQKCLV
-----EE-----HHHHH-----EE-----EEE-HHHHHH-HHHHHHHH
MGMKREAVQEERQRSRERA ESEAECANS GHEDMPVERILEAELAVEPKTESYGD MNTENSTNDPVTN
-----HHH
I CHAADKQLFTLVEWAKRIPHFSDLTLEDQVILLRAGWNELLIASFSHRVSVQD GILLATGLHVHR
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
SSAHSAGVGSIFDRVLT ELVSKMKDMQMDKSELGCLRAIVLFNPDAKGLSNPSEVETLREKVYATLE
-----HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AYTKQKYPEQPGRFAKLLLRPALRS IGLKCLEHLFFFKLIGDTPIDTFLMEMLETPLQTT
HHHHHHH-----HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

```

-为无规卷曲; H 为 α 螺旋; E 为 β 折叠

图 9 绵羊 RXRG 基因编码蛋白质二级结构预测

在国外还有关于 *RXRG* 基因在黑猩猩、斑马鱼以及其他不同功能方面的研究(全外显子鉴定家族分离性催乳素瘤中的 *RXRG* 和 TH 种系变异)^[11], 以及生物信息学的处理理论研究方法等^[12-13], 这些研究都表明视黄素受体基因在提高繁殖力方面具有很大潜力。本研究显示, *RXRG* 基因为 1 392 bp 长度的基因序列, 编码 463 个氨基酸残基, 其主要的功能区域是细胞核, 其理化性质预测中氨基酸组成中亮氨酸所占比例最大, 占比最少的氨基酸是色氨酸。通过系统发育树可以知道绵羊、瘤牛、马、猪在 *RXRG* 基因这一方面有比较近的亲缘关系, 和鸡的亲缘关系最远, 说明 *RXRG* 基因在哺乳动物间保守性较高。绵羊 *RXRG* 基因编码产物属于分泌蛋白, 位置 19 和 20 之间最可能的裂解位点是: CGG-CT, 该基因编码的蛋白为疏水性蛋白, 其保守结构域有两个功能区域, 其二级结构多为无规卷曲和 α 螺旋, 其三级结构主要以无规卷曲为主, 并且没有发现 *RXRG* 基因编码蛋白质产物的跨膜区域结构。

参考文献:

- [1] 黄 萌, 许尚忠, 答林森, 等. 牛 *RXRG* 基因遗传变异与双胎性状的关联分析[J]. 遗传, 2008(2): 190-194.
- [2] 黄 萌, 许尚忠, 答林森, 等. 牛 *RXRG* 基因 cDNA 的克隆及生物信息学分析[J]. 西北农林科技大学学报(自然科学版), 2008(11): 1-5; 10.
- [3] ALEXANDRA TALLAFUSS, LAURA A. HALE, YI-LIN YAN, et al. Postlethwait. Characterization of retinoid-X receptor genes *rxra*, *rxrba*, *rxrbb* and *rxrg* during zebrafish development[J]. Gene Expression Patterns, 2005, 6(5): 556-565.
- [4] 李华振, 狄 冉, 郭晓飞, 等. 绵羊繁殖轴相关组织 RXRs 基因的表达及 *RXRA* 基因多态性与季节性发情之间的关系[J]. 农业生物技术学报 2019, 27(8): 1401-1409.
- [5] 张小雪, 李发弟, 王维民. 绵羊 *STMN2* 基因生物信息学分析[J]. 甘肃农业科技, 2016(7): 58-61.
- [6] 王 伟, 杨巧丽, 谢开会, 等. 合作猪 *HMOX1* 基因克隆及生物信息学分析[JOL]. 基因组学与应用生物学, 2019[2019-04-29]. <http://kns.cnki.net/kcms/detail/45.1369.Q.20190426.1608.006.html>.
- [7] 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.
- [8] 王 辉. 中国美利奴羊 *RXRG* 基因多态性与产羔数的关联性分析[C]//中国畜牧兽医学会畜禽遗传标记分会. 第十二次全国畜禽遗传标记研讨会论文集. 北京: 中国畜牧兽医学会, 2010.
- [9] SHAOZE CHENG, MAN WANG, YINGJIE WANG, et al. *RXRG* associated in PPAR signal regulated the differentiation of primordial germ cell[J]. Journal of Cellular Biochemistry, 2018, 119(8): 6926-6934.
- [10] 刘 宇, 何向东, 夏 忆, 等. 牦牛 *MSMB* 基因的克隆及其生物信息学分析[J]. 中国草食动物科学, 2019, 39(5): 1-5; 22.
- [11] FLAVIA M. MELO, PATRÍCIA P. COUTO, ALLEN E. BALE, et al. Whole-exome identifies *RXRG* and TH germline variants in familial isolated prolactinoma[J]. Cancer Genetics, 2016, 209(6): 251-257.
- [12] ZHENG GAO, JONATHAN TERHORST, CRISTOPHER V. VAN HOUT, et al. U-PASS: unified power analysis and forensics for qualitative traits in genetic association studies[J]. Bioinformatics, 2020, 36(3): 974-975.
- [13] GOH KELVIN G K, MORIEL DANILO G, HANCOCK STEVEN J, et al. Bioinformatic and molecular analysis of inverse autotransporters from *Escherichia coli* [J]. mSphere, 2019, 4(4): e00572-e00591.

(本文责编: 陈 伟)