

杨佳栋,刘月琴,张英杰. 不同哺乳动物 *GDF9* 密码子偏性及聚类分析[J]. 江苏农业科学,2017,45(15):41-45.
doi:10.15889/j.issn.1002-1302.2017.15.010

不同哺乳动物 *GDF9* 密码子偏性及聚类分析

杨佳栋,刘月琴,张英杰

(河北农业大学动物科技学院,河北保定 071000)

摘要:为更好地认识哺乳动物 *GDF9* 基因在长期进化过程中为适应环境形成的密码子使用模式,通过 CodonW 软件计算不同哺乳动物 *GDF9* 基因的密码子偏性指标,利用 SPSS 软件对密码子偏好性影响因素的相关性进行分析,结合密码子偏性和最小进化法分析其亲缘关系。GCT、TGT、TTT、ATT、CCT、AGA、ACT 为哺乳动物 *GDF9* 基因的最优密码子,该基因偏爱使用以 A 或 T 结尾的密码子;G、C 含量和蛋白质亲水性是影响 *GDF9* 基因密码子偏性的主要因素;聚类分析提示不同哺乳动物间的同源性越高,*GDF9* 基因密码子使用特征的相似性越高。

关键词:哺乳动物;*GDF9*;碱基替换;密码子偏性;聚类分析

中图分类号: Q755 **文献标志码:** A **文章编号:** 1002-1302(2017)15-0041-05

在雌性哺乳动物整个生殖周期过程中,卵母细胞只能来源于原始卵泡库^[1]。有研究表明,原始卵泡形成过程受激素、转录因子和相关通路的介导^[2]。生长分化因子 9 (growth differentiation factor 9, *GDF9*) 作为转化生长因子 β 超家族成员之一,在卵泡的发育过程中有关键的调节作用^[3]。密码子是 DNA 或 RNA 的碱基序列与其编码蛋白序列之间的对应关系^[4]。编码相同氨基酸的密码子为同义密码子,在蛋白合成过程中,同义密码子的使用频率存在差异,且物种和基因对某一种或几种密码子的使用具有偏好性^[5]。对基因密码子偏好参数进行分析,能更好地理解和研究基因水平转移和基因家族分化的发生^[6]。因此对密码子偏好性的研究可在分子的角度为 *GDF9* 基因序列特征、分类和遗传进化规律提供重要信息。前人相关研究表明,*GDF9* 基因具有刺激颗粒细胞减数分裂、抑制素的生成作用^[7],并影响卵泡发育和生殖功能^[8]。Wang 等发现通过 siRNA 敲除 *GDF9* 后能抑制仓鼠原始卵泡的形成,但添加 *GDF9* 纯品培养卵巢会加速原始卵泡的形成^[9]。在人、啮齿类、牛、绵羊和有袋类动物卵巢卵母细胞中,*GDF9* 基因特异性表达,但该基因在山羊卵母细胞和黄体中可同时表达^[10-11]。*GDF9* 基因还能通过多个信号通路

促进颗粒细胞增殖过程^[12]。马会明等通过 RNA 干扰使该基因表达沉默,也能抑制颗粒细胞增殖过程^[13]。这些研究通过探讨不同哺乳动物已被克隆的 *GDF9* 基因,并在卵巢卵母细胞中进行了相关表达分析,但未开展密码子使用偏性的研究,这不利于其异源表达和遗传转化等后续试验的进行。*GDF9* 基因密码子使用偏好性的研究能为该基因的分类和进化提供重要信息。本研究利用 CodonW 软件分析不同哺乳动物 *GDF9* 基因对密码子的使用情况,基于 *GDF9* 基因最小进化法和同义密码子相对使用度的欧式平方距离系数建立聚类关系,为 *GDF9* 基因功能的深入研究提供参考依据。

1 材料与方法

1.1 目的基因的来源

本研究所用基因资料都来源于 Nucleotide (<https://www.ncbi.nlm.nih.gov/nuccore/>) 数据库,12 种物种包括灵长目苏门答腊猩猩、普通狨猴和人,食肉目家猫和家犬,偶蹄目家牛、牦牛、绵羊、山羊和野猪,啮齿目小家鼠和褐家鼠。各哺乳动物 *GDF9* 都是完整的 cDNA 序列(表 1)。利用 MEGA 6.0 对上述核苷酸序列进行碱基分析。

1.2 目的基因的筛选

对候选的哺乳动物 *GDF9* 基因 cDNA 序列进行如下规则的筛选:(1) cDNA 序列为完整编码蛋白序列;(2) 该 cDNA 序列长度 > 300 bp;(3) 蛋白翻译部位位于细胞质;(4) 质粒、转座子及病毒上的基因不予考虑;(5) 多拷贝基因不重复计算。

收稿日期:2017-01-10

基金项目:国家现代肉羊产业技术体系项目(编号:CARs-39);河北农业大学青年科学基金(编号:QJ201218)。

作者简介:杨佳栋(1977—),男,河北保定人,硕士,讲师,主要从事反刍动物营养研究。E-mail: hbnys96@126.com。

[4] 朱琳,芦建国. 干旱胁迫对夏蜡梅幼苗生理特征的影响[J]. 南京林业大学学报(自然科学版),2015,39(2):179-182.

[5] 邵果园,蔡荣荣,王力超,等. 夏蜡梅组织培养试验初报[J]. 浙江林业科技,2006,26(5):28-30.

[6] 顾福根,万志刚,孙丙耀,等. 夏蜡梅的离体培养和快速繁殖[J]. 园艺学报,2007,34(4):324.

[7] Bergmann B A, Heungkyu M. *In vitro* adventitious shoot production in *Paulownia* [J]. Plant Cell Reports, 1997, 16(5):315-319.

[8] Fan G Q, Zhai X Q, Zhai C J, et al. Callus induction from leaves of

different *Paulownia* species and its plantlet regeneration [J]. Journal of Forestry Research, 2001, 12(4):209-214.

[9] 余茂云,殷桂香,赵佩,等. 植物组织培养过程中器官发生途径再生植株分子机制研究进展[J]. 科技导报,2015,33(2):91-98.

[10] 张变莉,王杨,刘荣宁,等. 同源四倍体台湾泡桐体外植株再生系统的建立[J]. 河南农业科学,2015,44(7):119-123.

[11] 翟晓巧,聂琳,张晓申. 灰楸体外植株再生体系建立[J]. 江西农业学报,2011,23(3):17-19.

1.3 相关密码子偏性分析指标

通过 CodonW 软件来分析 *GDF9* 基因密码子使用偏性,分析数值包括有效密码子数 (effective number of codons, ENc)、密码子偏爱指数 (condon bias index, CBI)、同义密码子相对使用度 (relative synonymous codon usage, RSCU);同时通过该软件来分析 *GDF9* 基因密码子偏性的影响因素,即 GC 含量 (GC)、密码子第 3 位的 GC 含量 (GC3s)、芳香族氨基酸比例 (aromatic amino acid ratio)、蛋白质亲水性 (protein hydrophilicity) 的相关性。

1.4 碱基替代饱和度检验

利用 BioEdit 软件对不同哺乳动物 *GDF9* 基因的 cDNA 序列进行比对,将比对结果保存为 fasta 格式,通过 DAMBE 软件对比对结果的碱基替代饱和度进行检验,若转换数和颠换数没有达到饱和,可进行下一步的系统发育分析,反之亦然。

1.5 系统发育树的构建

通过 CodonW 软件计算出不同哺乳动物 *GDF9* 基因的 RSCU 值,利用 SPSS 多元分析软件计算其相应的欧氏平方距离,并进行系统聚类。通过 MEGA 软件中最小进化法建立系统发育树。

2 结果与分析

2.1 不同物种 *GDF9* 基因序列碱基特征

由表 1 可知,在 12 种物种中,灵长目苏门答腊猩猩和人 cDNA 序列长度最短,但与普通狨猴差别较大,相差 250 bp。偶蹄目家牛、牦牛和绵羊 cDNA 序列最长,但野猪 cDNA 序列短 27 bp,这可能与单胃和反刍动物分类有关。在 ATCG 碱基分布中,除小家鼠 (A + T = 45.9%)、野猪 (A + T = 49.8%) 和褐家鼠 (A + T = 47.2%) 外,其余物种 A + T 碱基含量高于 G + C 碱基含量,表明 *GDF9* 基因更倾向于使用 A 或 T 碱基。

表 1 不同哺乳动物 *GDF9* 基因序列的相关信息

| 物种 | 登录号 | 碱基组成 (%) | | | | 总碱基数 (bp) | ENc | GC3s | CBI |
|--------|----------------|----------|------|------|------|-----------|-------|-------|-------|
| | | T | C | A | G | | | | |
| 苏门答腊猩猩 | XM_002815874.3 | 26.8 | 24.5 | 27.6 | 21.1 | 1 104 | 52.95 | 0.483 | 0.062 |
| 家猫 | NM_001165900.1 | 27.0 | 26.3 | 23.4 | 23.3 | 1 359 | 49.87 | 0.501 | 0.067 |
| 家牛 | NM_174681.2 | 27.4 | 25.1 | 25.4 | 22.1 | 1 362 | 50.60 | 0.473 | 0.024 |
| 家犬 | NM_001168013.1 | 26.2 | 26.0 | 23.4 | 24.4 | 1 344 | 53.26 | 0.516 | 0.095 |
| 牦牛 | XM_005890431.2 | 27.3 | 25.2 | 25.4 | 22.1 | 1 362 | 50.66 | 0.475 | 0.024 |
| 绵羊 | NM_001142888.2 | 26.8 | 25.7 | 24.3 | 23.2 | 1 362 | 51.51 | 0.503 | 0.068 |
| 普通狨猴 | XM_002744603.3 | 27.0 | 25.0 | 25.4 | 22.7 | 1 368 | 50.39 | 0.482 | 0.034 |
| 人 | NM_001288824.2 | 26.3 | 24.7 | 27.7 | 21.3 | 1 101 | 53.42 | 0.485 | 0.036 |
| 山羊 | NM_001285708.1 | 26.9 | 25.6 | 24.4 | 23.1 | 1 362 | 51.54 | 0.501 | 0.078 |
| 小家鼠 | NM_008110.2 | 23.0 | 29.3 | 22.9 | 24.8 | 1 326 | 50.51 | 0.621 | 0.107 |
| 野猪 | NM_001001909.1 | 25.3 | 27.3 | 24.5 | 22.9 | 1 335 | 51.84 | 0.543 | 0.075 |
| 褐家鼠 | NM_021672.1 | 23.2 | 29.1 | 24.0 | 23.7 | 1 323 | 50.81 | 0.597 | 0.041 |

2.2 密码子使用偏性分析

ENc 能反映密码子使用频率和同义密码子平均使用频率的偏差。ENc 越小,偏好性越强,即高表达基因易于使用一种或几种同义密码子,反之亦然,即低表达基因所包含的稀有密码子种类越多^[14-15]。由表 1 可知,除家猫之外,其余 11 种哺乳动物的有效密码子数均小于 55,表明 *GDF9* 基因对同义密码子使用存在一定的偏好性。但各哺乳动物的 GC3s 值与其 G + C 碱基含量有差异。当 CBI > 0 时,即在 0 ~ 1 范围之间;越接近 0,表示该基因对密码子使用为完全随机,不存在偏性,反之亦然。本研究所用物种的密码子偏爱指数均 > 0,其

中家牛和牦牛的 CBI 值最小,均为 0.024,小家鼠的 CBI 值最大,为 0.107,但该值也远离 1,表明不同哺乳动物 *GDF9* 基因对密码子的使用不存在绝对的偏好性。

RSCU 值的大小能反映同义密码子家族的密码子使用偏移情况,当 RSCU = 1 时,表明密码子使用没有偏性;当 RSCU > 1 时,表明此密码子使用频率较高^[16]。由表 2 可知, RSCU > 1.5 的密码子有 25 个,其中以 GCT、TGT、TTT、GGG、ATT、CTG、CCT、CAG、AGA、AGG、ACT、ACC 和 CTG 密码子的使用频率较高,同时可得出 *GDF9* 基因偏好使用 A 或 T 结尾的密码子。

表 2 不同物种 *GDF9* 基因的 RSCU 值

| 氨基酸 | 密码子 | RSCU 值 | | | | | | | | | | | |
|-----|-----|--------|------|------|------|------|------|------|------|------|------|------|------|
| | | 苏门答腊猩猩 | 家猫 | 家牛 | 家犬 | 牦牛 | 绵羊 | 普通狨猴 | 人 | 山羊 | 小家鼠 | 野猪 | 褐家鼠 |
| Ala | GCT | 2.82 | 2.06 | 2.00 | 2.07 | 2.00 | 1.88 | 2.53 | 2.59 | 1.94 | 0.92 | 1.12 | 1.83 |
| | GCC | 0.94 | 1.29 | 0.93 | 1.10 | 0.93 | 1.00 | 0.80 | 0.94 | 0.90 | 1.69 | 0.64 | 1.33 |
| | GCA | 0.24 | 0.39 | 0.53 | 0.41 | 0.53 | 0.25 | 0.53 | 0.47 | 0.26 | 1.08 | 0.64 | 0.67 |
| | GCG | 0.00 | 0.26 | 0.53 | 0.41 | 0.53 | 0.88 | 0.13 | 0.00 | 0.90 | 0.31 | 1.60 | 0.17 |
| Cys | TGT | 1.27 | 1.54 | 1.57 | 1.43 | 1.57 | 1.54 | 1.43 | 1.40 | 1.54 | 1.38 | 1.38 | 1.38 |
| | TGC | 0.77 | 0.64 | 0.43 | 0.57 | 0.43 | 0.46 | 0.57 | 0.60 | 0.46 | 0.62 | 0.63 | 0.62 |
| Asp | GAT | 0.77 | 0.00 | 0.95 | 1.00 | 0.95 | 0.84 | 0.47 | 0.77 | 0.84 | 0.33 | 0.56 | 0.67 |
| | GAC | 1.23 | 1.00 | 1.05 | 1.00 | 1.05 | 1.16 | 1.53 | 1.23 | 1.16 | 1.67 | 1.44 | 1.33 |
| Glu | GAA | 0.90 | 1.04 | 1.00 | 0.95 | 1.00 | 1.00 | 0.67 | 0.95 | 0.96 | 0.88 | 0.91 | 0.80 |
| | GAG | 1.10 | 0.96 | 1.00 | 1.05 | 1.00 | 1.00 | 1.33 | 1.05 | 1.04 | 1.12 | 1.09 | 1.20 |

续表 2

| 氨基酸 | 密码子 | RSCU 值 | | | | | | | | | | | |
|-----|-----|--------|------|------|------|------|------|------|------|------|------|------|------|
| | | 苏门答腊猩猩 | 家猫 | 家牛 | 家犬 | 牦牛 | 绵羊 | 普通狢猴 | 人 | 山羊 | 小家鼠 | 野猪 | 褐家鼠 |
| Phe | TTT | 1.60 | 1.33 | 1.44 | 0.95 | 1.44 | 1.50 | 1.67 | 1.57 | 1.50 | 1.07 | 1.14 | 1.07 |
| | TTC | 0.40 | 0.67 | 0.56 | 1.05 | 0.56 | 0.50 | 0.33 | 0.43 | 0.50 | 0.93 | 0.86 | 0.93 |
| Gly | GGT | 0.92 | 0.84 | 0.57 | 1.04 | 0.57 | 0.94 | 1.20 | 0.80 | 1.00 | 0.47 | 0.75 | 0.00 |
| | GGC | 0.62 | 1.05 | 0.57 | 0.87 | 0.57 | 0.47 | 0.60 | 1.07 | 0.50 | 1.41 | 0.75 | 1.43 |
| | GGA | 1.54 | 0.63 | 1.14 | 1.22 | 1.14 | 0.94 | 1.20 | 1.33 | 1.00 | 0.71 | 0.25 | 1.43 |
| | GGG | 0.92 | 1.47 | 1.71 | 0.87 | 1.71 | 1.65 | 1.00 | 0.80 | 1.50 | 1.41 | 2.25 | 1.14 |
| His | CAT | 0.77 | 0.86 | 0.78 | 0.35 | 0.78 | 0.78 | 0.50 | 0.53 | 0.78 | 0.60 | 0.63 | 0.83 |
| | CAC | 1.23 | 1.14 | 1.22 | 1.65 | 1.22 | 1.22 | 1.50 | 1.47 | 1.22 | 1.40 | 1.38 | 1.17 |
| Ile | ATT | 1.59 | 1.15 | 1.71 | 1.09 | 1.85 | 1.38 | 1.67 | 1.50 | 1.38 | 1.25 | 1.62 | 1.50 |
| | ATC | 0.88 | 1.62 | 0.64 | 1.36 | 0.69 | 0.92 | 1.00 | 0.94 | 0.92 | 1.25 | 1.38 | 1.20 |
| | ATA | 0.53 | 0.23 | 0.64 | 0.55 | 0.46 | 0.69 | 0.33 | 0.56 | 0.69 | 0.50 | 0.00 | 0.30 |
| Lys | AAA | 0.92 | 0.83 | 0.97 | 0.96 | 0.97 | 0.89 | 0.83 | 0.75 | 0.93 | 1.14 | 1.07 | 1.05 |
| | AAG | 1.08 | 1.17 | 1.03 | 1.04 | 1.03 | 1.11 | 1.17 | 1.25 | 1.07 | 0.86 | 0.93 | 0.95 |
| Leu | TTA | 0.53 | 0.13 | 0.35 | 0.46 | 0.35 | 0.36 | 0.47 | 0.51 | 0.36 | 0.12 | 0.38 | 0.12 |
| | TTC | 1.24 | 0.88 | 0.94 | 1.15 | 0.94 | 0.96 | 1.06 | 1.20 | 0.84 | 0.86 | 1.25 | 1.10 |
| | CTT | 0.88 | 1.38 | 1.41 | 1.15 | 1.41 | 1.20 | 1.29 | 1.03 | 1.20 | 1.10 | 1.13 | 0.98 |
| | CTC | 1.24 | 1.13 | 1.18 | 0.92 | 1.18 | 1.20 | 1.41 | 1.20 | 1.20 | 1.71 | 1.13 | 1.96 |
| | CTA | 0.35 | 0.63 | 0.35 | 0.23 | 0.35 | 0.12 | 0.24 | 0.34 | 0.24 | 0.37 | 0.13 | 0.49 |
| | CTG | 1.76 | 1.88 | 1.76 | 2.08 | 1.76 | 2.16 | 1.53 | 1.71 | 2.16 | 1.84 | 2.00 | 1.35 |
| Met | ATG | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Asn | AAT | 0.67 | 0.63 | 0.44 | 0.67 | 0.42 | 0.44 | 0.75 | 0.89 | 0.44 | 0.57 | 0.25 | 1.07 |
| | AAC | 1.67 | 1.38 | 1.56 | 1.33 | 1.58 | 1.56 | 1.25 | 1.11 | 1.56 | 1.43 | 1.75 | 0.93 |
| Pro | CCT | 1.67 | 2.22 | 2.05 | 1.83 | 2.05 | 2.05 | 1.65 | 1.50 | 2.16 | 1.30 | 1.37 | 1.22 |
| | CCC | 0.83 | 1.00 | 0.97 | 0.80 | 0.97 | 0.97 | 1.06 | 1.00 | 0.86 | 1.73 | 1.26 | 1.89 |
| | CCA | 1.00 | 0.67 | 0.97 | 1.14 | 0.97 | 0.86 | 1.06 | 1.17 | 0.76 | 0.43 | 1.16 | 0.67 |
| | CCG | 0.50 | 0.11 | 0.00 | 0.23 | 0.00 | 0.11 | 0.24 | 0.33 | 0.22 | 0.54 | 0.21 | 0.22 |
| Gln | CAA | 0.25 | 0.36 | 0.13 | 0.33 | 0.13 | 0.00 | 0.20 | 0.40 | 0.00 | 0.21 | 0.22 | 0.10 |
| | CAG | 1.75 | 1.64 | 1.88 | 1.67 | 1.80 | 2.00 | 1.80 | 1.60 | 2.00 | 1.79 | 1.78 | 1.90 |
| Arg | AGA | 2.29 | 2.69 | 2.36 | 2.52 | 2.36 | 2.28 | 2.77 | 2.45 | 2.20 | 1.29 | 2.36 | 1.86 |
| | AGG | 1.43 | 1.24 | 1.93 | 1.35 | 1.93 | 1.66 | 1.38 | 1.36 | 1.80 | 2.36 | 1.07 | 2.07 |
| | CGT | 0.57 | 1.03 | 0.64 | 0.58 | 0.64 | 0.62 | 0.46 | 0.55 | 0.60 | 0.64 | 0.64 | 0.62 |
| | CGC | 0.86 | 0.00 | 0.00 | 0.19 | 0.00 | 0.62 | 0.46 | 0.55 | 0.60 | 0.43 | 0.43 | 0.00 |
| | CGA | 0.29 | 0.41 | 0.21 | 0.58 | 0.21 | 0.21 | 0.46 | 0.27 | 0.20 | 0.43 | 0.21 | 0.21 |
| | CGG | 0.57 | 0.62 | 0.86 | 0.77 | 0.86 | 0.62 | 0.46 | 0.82 | 0.60 | 0.86 | 1.29 | 1.24 |
| Ser | AGT | 1.17 | 0.78 | 0.88 | 0.78 | 0.90 | 0.75 | 1.20 | 1.17 | 0.75 | 0.94 | 0.84 | 0.92 |
| | AGC | 0.73 | 0.78 | 1.02 | 0.91 | 0.90 | 0.90 | 0.93 | 0.73 | 0.90 | 1.29 | 0.84 | 1.38 |
| | TCT | 1.46 | 1.57 | 1.32 | 1.83 | 1.35 | 1.35 | 1.47 | 1.32 | 1.35 | 1.65 | 1.40 | 0.92 |
| | TCC | 1.17 | 1.70 | 1.32 | 0.91 | 1.35 | 1.35 | 1.20 | 1.32 | 1.35 | 1.53 | 1.67 | 1.73 |
| | TCA | 1.46 | 1.04 | 1.46 | 0.65 | 1.50 | 1.50 | 1.20 | 1.46 | 1.50 | 0.35 | 1.12 | 0.81 |
| | TCG | 0.00 | 0.13 | 0.00 | 0.91 | 0.00 | 0.15 | 0.00 | 0.00 | 0.15 | 0.24 | 0.14 | 0.23 |
| Thr | ACT | 1.20 | 1.68 | 1.78 | 1.00 | 1.68 | 1.78 | 1.40 | 1.26 | 1.56 | 0.83 | 1.14 | 0.80 |
| | ACC | 1.80 | 1.26 | 1.56 | 1.60 | 1.47 | 1.56 | 1.20 | 1.89 | 1.78 | 2.00 | 0.76 | 1.87 |
| | ACA | 1.00 | 0.63 | 0.67 | 0.40 | 0.63 | 0.44 | 1.20 | 0.84 | 0.44 | 0.83 | 1.33 | 0.53 |
| | ACG | 0.00 | 0.42 | 0.00 | 1.00 | 0.21 | 0.22 | 0.20 | 0.00 | 0.22 | 0.33 | 0.76 | 0.46 |
| Val | GTT | 1.14 | 1.04 | 1.33 | 1.60 | 1.33 | 1.28 | 1.33 | 1.20 | 1.38 | 0.41 | 1.12 | 0.46 |
| | GTC | 0.76 | 0.15 | 0.50 | 0.16 | 0.50 | 0.48 | 0.50 | 0.60 | 0.46 | 0.69 | 0.64 | 0.77 |
| | GTA | 0.57 | 0.44 | 0.67 | 0.48 | 0.67 | 0.80 | 0.50 | 0.60 | 0.77 | 0.28 | 0.64 | 0.46 |
| | GTG | 1.52 | 2.37 | 1.50 | 1.76 | 1.50 | 1.44 | 1.67 | 1.60 | 1.38 | 0.62 | 1.60 | 2.31 |
| Trp | TGG | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Tyr | TAT | 1.00 | 1.00 | 1.13 | 0.80 | 1.13 | 1.00 | 1.16 | 1.11 | 1.00 | 0.47 | 0.94 | 0.47 |
| | TAC | 1.00 | 1.00 | 0.88 | 1.20 | 0.88 | 1.00 | 0.84 | 0.89 | 1.00 | 1.53 | 1.06 | 1.53 |

2.3 密码子使用偏性的影响因素分析

由表 3 可知,GC、GC3s 和 Aromo 都与 CBI 呈正相关,其

中 GC 与 CBI 呈显著正相关,Gravy 与 CBI 呈负相关;GC、GC3s 和 Gravy 都与 ENc 呈负相关,其中 Gravy 与 ENc 呈显著

负相关,Aromo 与 ENc 呈正相关。上述结果表明,GC 含量和蛋白质亲水性对 *GDF9* 基因密码子偏性有影响。

表 3 偏性指标与影响因素的相关系数

| 指标 | GC | GC3s | Aromo | Gravy |
|-----|---------|--------|--------|----------|
| CBI | 0.596 * | 0.541 | -0.165 | 0.267 |
| ENc | -0.346 | -0.195 | 0.005 | -0.689 * |

注:“*”表示在 0.05 水平上显著相关, $P<0.05$ 。

2.4 不同物种 *GDF9* 基因的碱基替代饱和度检验

图 1 横坐标代表核苷酸转换数(s)和核苷酸颠换数(v),纵坐标代表基于 F84 模型建立的遗传距离。*GDF9* 基因的核苷酸转换数和颠换数均随着遗传距离的增加而变大,且两者呈线性增加关系,表明本研究所用的 12 种哺乳动物 *GDF9* 基因序列碱基替代未达到饱和,可进行下一步的聚类分析。

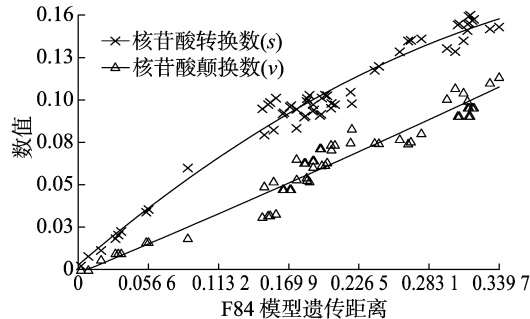


图 1 *GDF9* 基因序列碱基替代饱和度检验

2.5 *GDF9* 基因的聚类分析

由图 2 可知,12 种哺乳动物共分为三大类:一类是偶蹄目家牛、牦牛、绵羊、山羊和野猪;一类是灵长目苏门答腊猩猩、人、普通狢猴和食肉目家猫、家犬;另一类是啮齿目小鼠和褐家鼠,其中灵长目和食肉目的亲缘关系较近,而偶蹄目、灵长目、食肉目与啮齿目的亲缘关系较远,其中类间的最大距离标准化是 25。由图 3 可知,偶蹄目和食肉目亲缘关系较近,二者与灵长目亲缘关系较近,啮齿目与其亲缘关系最远。2 种不同方法构建的聚类结果不一致,但结合其密码子偏性指标结果发现,亲缘关系较近的物种,其密码子使用偏好性一致。

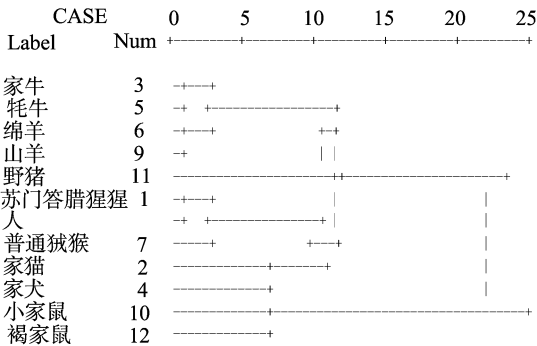


图 2 基于密码子使用偏性的聚类结果

3 讨论

密码子偏好性的研究对了解物种起源于进化、目的基因的改造、推测未知基因表达水平及预测目的基因最佳宿主真

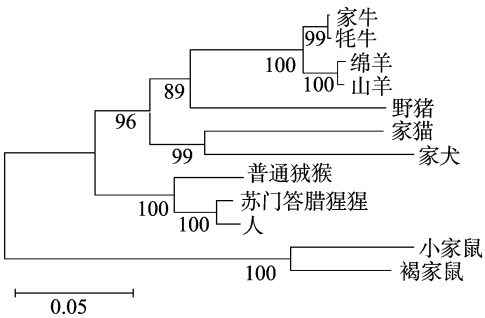


图 3 基于最小进化法的聚类结果

定了理论基础^[17-18],同时为转基因研究中的甲基化能否引起基因表达量下降或基因沉默提供参考^[19]。本研究所选 12 种哺乳动物 *GDF9* 基因的有效密码子数值在 49.57~53.42 区间波动,密码子使用上存在一定的偏好性;在碱基序列组成成分上,9 种哺乳动物的 AT 含量高于 GC 含量,但密码子第 3 位的 GC 含量(GC3s)与 GC 含量存在差异;12 种哺乳动物有 25 个密码子的同义密码子相对使用度>1.5,其中以 GCT、TGT、TTT、ATT、CCT、AGA、ACT 为主,说明 *GDF9* 基因偏好使用 A 或 T 结尾的密码子。因此推断碱基组成是造成密码子偏好性的原因之一,这与姜艳等的研究结果^[20]一致。

目前的研究认为,碱基差异(突变)、自然选择、cDNA 序列长度和 tRNA 丰度对密码子偏好性均有影响^[21]。Chen 等发现自然选择造成的密码子变化可能提高基因表达效率和多肽折叠动力,但这只发生在特定的基因中,甚至在特定基因的特定位置中^[22]。本研究选取不同哺乳动物相同基因,可忽略蛋白自然选择和 tRNA 丰度等影响,通过分析密码子使用偏性的影响因素,发现 GC 含量和蛋白质亲水性是影响 *GDF9* 基因密码子偏性的主要因素。已有研究表明,GC 含量对同义密码子使用偏好性有重要影响,密码子偏好性越强的基因越趋向于使用以 G 或 C 结尾的密码子,且密码子偏好性的变化往往由第 3 位密码子决定^[23]。因子在不存在自然选择压力的情况下,一定方向的突变压力对第 3 位碱基种类会有影响^[24]。

本研究基于密码子使用偏性的聚类与基于最小进化法形成的聚类相比,发现同一目下的物种聚为一类,即亲缘关系较近。但 2 种聚类结果存在差异,主要集中在前者将灵长目和食肉目聚为一类,后者将偶蹄目和食肉目聚为一类。这种聚类结果不一致的现象在植物和其他动物^[25-28]中均出现过,提示我们在今后的研究中,对物种的分离不能仅局限于对碱基组成的分析。可结合密码子使用偏性和编码区序列总和的结果进行分析,这样可为提高分类结果准确度提供重要参考。*GDF9* 基因偏好使用的密码子的确定有利于其转基因表达系统的建立,并对其目的基因的改造和最优密码子的优化等提供理论基础。

4 结论

本研究通过分析 12 种哺乳动物 *GDF9* 基因密码子使用模式,发现不同哺乳动物间密码子选择偏性具有差异性,确定该基因的最优密码子为 GCT、TGT、TTT、ATT、CCT、AGA、ACT,且偏爱使用 A 或 T 结尾的密码子。GC 含量和蛋白质亲

水性是影响 *GDF9* 基因密码子偏性的主要因素。基于密码子使用偏性的聚类与基于最小进化法形成的聚类均表明亲缘关系相近的物种间的密码子使用特征较为一致,但 2 种方法构建的聚类结果存在差异。

参考文献:

- [1] Zhang H, Risal S, Gorre N, et al. Somatic cells initiate primordial follicle activation and govern the development of dormant oocytes in mice[J]. *Current Biology*, 2014, 24(21): 2501–2508.
- [2] 赵立华. *Rac1* 在小鼠原始卵泡形成中的功能与机制[D]. 北京: 中国农业大学, 2016.
- [3] Tang K Q, Yang W C, Zhang X X, et al. Effects of polymorphisms in the bovine growth differentiation factor 9 gene on sperm quality in Holstein bulls[J]. *Genetics and Molecular Research*, 2013, 12(3): 2189–2195.
- [4] 刘庆坡, 薛庆中. 遗传密码子及其应用[J]. *中国生物化学与分子生物学报*, 2006, 22(11): 851–855.
- [5] 朱孝轩, 朱英杰, 宋经元, 等. 基于全基因组和转录组分析的赤芝密码子使用偏好性比较研究[J]. *药学报*, 2014, 49(9): 1340–1345.
- [6] Wei L, He J, Jia X, et al. Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution[J]. *BMC Evolutionary Biology*, 2014, 14: 262.
- [7] Vitt U A, Hayashi M, Klein C, et al. Growth differentiation factor – 9 stimulates proliferation but suppresses the follicle – stimulating hormone – induced differentiation of cultured granulosa cells from small antral and preovulatory rat follicles [J]. *Biology of Reproduction*, 2000, 62(2): 370–377.
- [8] 陶萨茹拉, 王 峰, 刘永斌. 生长分化因子 *GDF9* 基因的研究进展[J]. *畜牧与饲料科学*, 2009, 30(1): 54–55.
- [9] Wang J, Roy S K. Growth differentiation factor – 9 and stem cell factor promote primordial follicle formation in the hamster; modulation by follicle – stimulating hormone[J]. *Biology of Reproduction*, 2004, 70(3): 577–585.
- [10] Crawford J L, McNatty K P. The ratio of growth differentiation factor 9; bone morphogenetic protein 15 mRNA expression is tightly co – regulated and differs between species over a wide range of ovulation rates[J]. *Molecular and Cellular Endocrinology*, 2012, 348(1): 339–343.
- [11] Hosoe M, Kaneyama K, Ushizawa K, et al. Quantitative analysis of bone morphogenetic protein 15 (BMP15) and growth differentiation factor 9 (*GDF9*) gene expression in calf and adult bovine ovaries [J]. *Reproductive Biology and Endocrinology*, 2011, 9: 33.
- [12] 常秀峰, 张 敏, 吴小华, 等. 卵母细胞分泌的生长分化因子 – 9 和骨形成蛋白 – 15 与卵母细胞发育潜能的关系[J]. *生殖与避孕*, 2014, 34(3): 200–204.
- [13] 马会明, 张永芳, 王蒙蒙, 等. 生长分化因子 – 9 下调对小鼠卵巢颗粒细胞增殖能力的影响[J]. *安徽医科大学学报*, 2016, 51(1): 51–55.
- [14] Xu Y, Jia R, Zhang Z, et al. Analysis of synonymous codon usage pattern in duck circovirus[J]. *Gene*, 2015, 557(2): 138–145.
- [15] Zhang Z, Dai W, Wang Y, et al. Analysis of synonymous codon usage patterns in torque teno sus virus 1 (TTSuV1) [J]. *Archives of Virology*, 2013, 158(1): 145–154.
- [16] 张之一, 陈 蕾, 薛慧良, 等. 13 种哺乳动物 OTR 密码子使用偏性及其聚类分析[J]. *北华大学学报(自然科学版)*, 2012, 13(1): 44–48.
- [17] 吴宪明, 吴松峰, 任大明, 等. 密码子偏性的分析方法及相关研究进展[J]. *遗传*, 2007, 29(4): 420–426.
- [18] 宗秋芳, 戴开宇, 刘 颖, 等. 猪氨肽酶 N (APN) 基因密码子偏好性分析[J]. *扬州大学学报(农业与生命科学版)*, 2016, 37(2): 45–50.
- [19] 张 乐, 金龙国, 罗 玲, 等. 大豆基因组和转基因组的核基因密码子使用偏好性分析[J]. *作物学报*, 2011, 37(6): 965–974.
- [20] 姜 艳, 张全星, 管德龙, 等. 小鲵科动物 *Cytb* 基因密码子偏好性分析[J]. *陕西师范大学学报(自然科学版)*, 2016, 44(4): 77–82.
- [21] Liu G, Wu J, Yang H, et al. Codon usage patterns in *corynebacterium glutamicum*; mutational bias, natural selection and amino acid conservation [J]. *Comparative and Functional Genomics*, 2010, 2010: 343569.
- [22] Chen H, Sun S, Norenburg J L, et al. Mutation and selection cause codon usage and bias in mitochondrial genomes of ribbon worms (Nemertea) [J]. *PLoS One*, 2014, 9(1): e85631.
- [23] Carlini D B, Chen Y, Stephan W. The relationship between third – codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes *Adh* and *Adhr* [J]. *Genetics*, 2001, 159(2): 623–633.
- [24] 时 慧, 王 玉, 杨路成, 等. 茶树抗寒调控转录因子 ICE1 密码子偏性分析[J]. *园艺学报*, 2012, 39(7): 1341–1352.
- [25] 刘汉梅, 赵 耀, 顾 勇, 等. 几种植物 *waxy* 基因的密码子用法特性分析[J]. *核农学报*, 2010, 24(3): 476–481.
- [26] 晁岳愚, 吴政卿, 杨会民, 等. 11 种植物 *psbA* 基因的密码子偏好性及聚类分析[J]. *核农学报*, 2011, 25(5): 927–932.
- [27] 郭秀丽, 王 玉, 杨路成, 等. 茶树 *CBF1* 基因密码子使用特性分析[J]. *遗传*, 2012, 34(12): 1614–1623.
- [28] 李丽莎, 李祥龙, 周荣艳, 等. 山羊酪氨酸相关蛋白 1 (TYRP1) 基因密码子偏好性分析[J]. *贵州农业科学*, 2016, 44(3): 113–119.