

京海黄鸡 *MyoG* 基因密码子特性分析

张 涛^{1,2}, 张跟喜^{1,2}, 韩昆鹏^{1,2}, 王金玉^{1,2}, 樊庆灿^{1,2},
李婷婷^{1,2}, 段 炼^{1,2}, 王永娟³

(1. 扬州大学 动物科学与技术学院, 江苏 扬州 225009; 2. 江苏省动物遗传繁育与分子设计重点实验室,
江苏 扬州 225009; 3. 江苏京海集团, 江苏 南通 226103)

摘要: 为了解鸡 *MyoG* 基因的密码子特性, 以便于为 *MyoG* 基因的表达选择合适的外源表达系统, 使用在线工具 CUSP 和 CHIPS 以及 codonW 软件对鸡 *MyoG* 基因进行分析, 并与鸡基因、模式动物基因组和其他动物 *MyoG* 基因进行比较。结果表明, *MyoG* 基因偏好于 G/C 结尾的密码子, CDS 序列中, GC 含量大于 AT 含量。*MyoG* 基因与鸡其他 31 个基因密码子偏好性相似, 均偏好于 G/C 结尾的密码子。与其他物种基因组密码子偏好性比较结果显示, 密码子偏好性与小鼠差异最小, 说明小鼠可以作为 *MyoG* 基因外源表达的宿主。与其他动物 *MyoG* 基因密码子偏好性比较显示, *MyoG* 基因在包括京海黄鸡在内的 32 个物种中均有较高表达, 几乎所有 *MyoG* 基因编码区对 GC 有较强烈的偏好性。基于 RSCU 值的聚类分析和基于 CDS 区系统发育树结果显示, 亲缘关系近的物种之间倾向于拥有相似的密码子偏好性。

关键词: 京海黄鸡; *MyoG*; 密码子; 偏好性; 聚类分析

中图分类号: Q78 文献标识码: A 文章编号: 1000-7091(2014)04-0071-09

Characterization of Codon Usage of *MyoG* Gene in Jinghai Yellow Chicken

ZHANG Tao^{1,2}, ZHANG Gen-xi^{1,2}, HAN Kun-peng^{1,2}, WANG Jin-yu^{1,2}, FAN Qing-can^{1,2},
LI Ting-ting^{1,2}, DUAN Lian^{1,2}, WANG Yong-juan³

(1. College of Animal Science and Technology, Yangzhou University, Yangzhou 225009, China; 2. Key Laboratory of Animal Genetics, Breeding, Reproduction and Molecular Design of Jiangsu Province, Yangzhou 225009, China; 3. Jiangsu Jinghai Poultry Group Co., Ltd., Nantong 226103, China)

Abstract: In order to understand the codon characteristics of *MyoG* gene in chicken and select the appropriate exogenous expression system, online tools CUSP, CHIPS and codonW software were used to analyze *MyoG* gene. *MyoG* gene was compared to other genes of chicken, model animal genomes and *MyoG* gene of other species. Result showed that *MyoG* of chicken was bias toward the synonymous codons with G and C at the third codon position. GC content was greater than AT content in CDS region. *MyoG* gene and other thirty-one genes of chicken shared similar codon bias toward the synonymous codons with G and C at the third codon position. Codon bias comparison to other species showed that differences of codon bias between *MyoG* gene and *Mus musculus* was the minimum, indicating that *Mus musculus* could be used as exogenous gene expression host of *MyoG* gene. Codon bias comparison to the *MyoG* gene of other species indicated that *MyoG* gene expressed high in all the thirty-two species, including Jinghai yellow chicken. Almost all the thirty-two *MyoG* genes were bias toward to GC. Result of clustering based on RSCU value and phylogenetic tree based on CDS region showed that species with close genetic relationship tend to have similar codon bias.

Key words: Jinghai yellow chicken; *MyoG*; Codon; Bias; Clustering analysis

蛋白质合成的模板是核酸, 核酸共可以组成 61 多的有 6 个) 共同编码一种氨基酸, 这些密码子称为同义密码子。以前研究表明, 基因对同义密码子

收稿日期: 2014-04-22

基金项目: 江苏省高校自然科学基金项目(12KJB230003); 国家肉鸡产业技术体系项目(nycytx-42-G1-05); 江苏高校优势学科建设工程项目; 江苏省动物遗传繁育与分子设计重点实验室项目

作者简介: 张 涛(1990-), 男, 山东临沂人, 在读硕士, 主要从事动物遗传育种与繁殖研究。

通讯作者: 王金玉(1952-), 男, 江苏射阳人, 教授, 博士生导师, 主要从事家禽育种研究。

的使用存在偏好性,即不是随机的使用密码子,而是优先使用其中的一些密码子。而对同一物种而言,不同基因也有不同的密码子偏好性,这与蛋白质二级结构、基因组的碱基组成、基因表达水平等有关^[1-3]。

密码子偏性现象存在中性理论和选择-突变-漂变学说2种理论^[4]。其中,中性理论认为是中性选择导致同义密码子的末位碱基突变,而中性选择不受自然选择压力的影响;选择-突变-漂变学说则认为同义密码子的偏好性使用主要是用于选择最优密码子。目前,在转基因研究中,密码子偏性分析已经得到成功运用,通过比较外源表达系统与某一特定基因之间的密码子使用偏性差异,分析是否会产生甲基化,进而导致基因沉默或者表达量下降^[5],因而,密码子偏好性分析可以用来预测目的基因的最佳宿主^[6]。

肌细胞生成素(Myogenic factor, MyoG)基因是成肌调节因子(Myogenic regulatory factors, MRFs)家族的成员之一^[7], MRFs 家族由生肌决定因子 D1(MyfD1)、肌细胞生成素(Myogenic factor, MyoG)、生肌因子5(Myf5)和生肌调节因子4(MRF4)4个基因组成^[8],是控制骨骼肌生成的关键因子^[9-10],它们编码的碱性螺旋-环-螺旋蛋白(Basic helix-loop-helix protein, bHLH)参与肌肉细胞的形成与分化过程^[11]。MyoG是生肌调节因子家族中最重要的一员,在中胚层细胞分化为成肌细胞,再由成肌细胞融合为肌纤维这一过程中发挥着重要的调控作用。MyoG基因是MRFs基因家族中唯一在所有骨骼肌肌细胞系中均可表达的基因^[12]。克隆MyoG基因,探讨该基因的特征特性,对提高鸡的生产性能具有一定意义。本研究分析了京海黄鸡MyoG基因的密码子特性,并将其与不同物种MyoG基因、同物种不同基因以及模式动物基因组的密码子偏好性进行了比较,对了解MyoG基因的密码子偏好性,并为其选择合适的外源表达系统,提高外源表达具有重要意义,并为MyoG基因转入模式动物进行功能验证奠定基础。

1 材料和方法

1.1 目的基因的获取

本研究中,京海黄鸡MyoG基因序列由江苏省动物遗传繁育与分子设计重点实验室克隆获得,其CDS区全长684 bp,编码227个氨基酸,GenBank登录号为KJ130962。人(*Homo sapiens*)、小鼠(*Mus musculus*)等31种动物的MyoG基因序列以及筛选的31个鸡基因序列均来源于GenBank(表1)。

表1 鸡31条序列名称及GenBank登录号

Tab. 1 Name and GenBank accession number of 31 genes of chicken

| 基因 Gene name | 序列登录号 GenBank accession No. |
|---|-----------------------------------|
| Activin A receptor(<i>ACVRI</i>) | NM_204560 |
| Calmodulin 2(<i>CALM</i>) | NM_205005 |
| CaM kinase-like vesicle-associated(<i>CAMKV</i>) | XM_001234174 |
| Black-bone CAPN1(<i>CAPN1</i>) | FJ232589 |
| Calpastatin(<i>CAST</i>) | NM_001137650 |
| CD9 molecule(<i>CD9</i>) | NM_204762 |
| Corticotropin releasing hormone receptor 1(<i>CRHRI</i>) | NM_204321 |
| Fas(<i>FAS</i>) | NM_001199487 |
| POU class 1 homeobox 1(<i>POU1F1</i>) | NM_204319 |
| Follicle stimulating hormone beta(<i>FSHB</i>) | AB077362 |
| Follicle stimulating hormone receptor(<i>FSHR</i>) | NM_205079 |
| Growth hormone receptor(<i>GHR</i>) | NM_001001293 |
| Galactosidase, alpha(<i>GLA</i>) | XM_420183 |
| Gonadotropin-releasing hormone receptor(<i>GNRHR</i>) | NM_001012609 |
| Insulin-like growth factor 1(<i>IGF1</i>) | NM_001004384 |
| Insulin-like growth factor I receptor(<i>IGF-IR</i>) | KC242244 |
| LH beta subunit(<i>LHB</i>) | HQ872606 |
| Lipoprotein lipase(<i>LPL</i>) | NM_205282 |
| Melanocortin 1 Receptor(<i>MC1R</i>) | KF379749 |
| Melanocortin 4 receptor(<i>MC4R</i>) | NM_001031514 |
| Myostatin(<i>MSTN</i>) | NM_001001461 |
| Leptin precursor(<i>ob</i>) | AF082500 |
| Phospholipid transfer protein(<i>PLTP</i>) | NM_001162406 |
| Peroxisome proliferator-activated receptor alpha(<i>PPAR</i>) | EU049891 |
| Prolactin(<i>PRL</i>) | NM_205466 |
| Prolactin receptor(<i>PRLR</i>) | NM_204854 |
| Parathyroid hormone(<i>PTH</i>) | NM_205452 |
| Thyroid hormone responsive(<i>THRSP</i>) | NM_213577 |
| Thyrotropin-releasing hormone(<i>TRH</i>) | NM_001030383 |
| Growth hormone(<i>GH</i>) | NM_204359 |
| Insulin-like growth factor 2(<i>IGF2</i>) | NM_001030342 |

1.2 目的基因的筛选

筛选目的基因的原则为:编码序列长度大于300 bp,并且编码序列完整,编码蛋白质的基因;基因在细胞质中翻译,排除细胞器中翻译的基因,排除位于质粒、转座子、病毒上的基因,对多拷贝基因只统计一次^[13]。

1.3 方法

本研究使用在线工具 CUSP(<http://pro.genomics.purdue.edu/emboss/>)和 codonW 软件分析 GC 含量、GC3s、相对同义密码子使用度(Relatively synonymous codon usage, RSCU)以及密码子使用频率,运用软件 CHIPS(<http://pro.genomics.purdue.edu/emboss/>)对MyoG基因的有效密码子数(ENc)进行

统计,利用 SPSS 19 对 32 条 *MyoG* 基因的密码子使用偏好性进行聚类,使用 DNAMAN 软件对 32 条 *MyoG* 序列构建系统发育树。

2 结果与分析

2.1 *MyoG* 基因密码子偏好性分析

2.1.1 有效密码子数(ENc)、GC 含量、和 GC3s 的计算 ENc 是一个基因的密码子使用频率与同义密码子平均使用频率偏差的量化值,该值的范围为 20~61,越接近 20 说明密码子的偏好性越强。此值可以反映基因对同义密码子使用的偏好程度^[14]。GC3s 表示 G 和 C 碱基在第 3 位密码子碱基中出现的频率。本研究利用 CHIPS 和 codonW 计算的结果显示,京海黄鸡 *MyoG* 基因的 ENc 值、GC 含量和 GC3s 分别为 35.24、0.631 和 0.817,由此可见,京海黄鸡 *MyoG* 基因的 ENc 值较小,密码子存在偏好性;GC3s 大于 0.5,说明鸡 *MyoG* 基因偏好以 G/C 结尾的密码子;GC 含量大于 0.5,说明鸡 *MyoG* 基因 CDS 序列中,GC 含量大于 AT 含量。

2.1.2 与鸡其他基因密码子偏好性比较 同义密码子相对使用度(Relative synonymous codon usage,

RSCU)是指对于某一特定的密码子,在编码对应氨基酸的同义密码子间的相对概率。它去除了氨基酸组成对密码子使用的影响,如果 RSCU 值等于 1,说明密码子的使用没有偏好;如果某一密码子的 RSCU 值大于 1,则表明该密码子具有相对较高的使用频率,反之亦然^[15]。本研究中,codonW 软件计算结果(表 2)表明,在京海黄鸡 *MyoG* 基因的密码子中,RSCU 值 ≥ 1.2 的密码子有 22 种,为 *MyoG* 基因的偏好密码子,其中 GGC、ATC、CGC、AGC、ACC、CTG 和 GTG(RSCU>2)偏好性较强,所有偏好密码子全部以 G/C 结尾。编码 Met 的密码子 ATG 和编码 Trp 的密码子的 RSCU 值等于 1,说明京海黄鸡 *MyoG* 基因中对 Met 和 Trp 使用没有偏好性。从 GenBank 中筛选了 31 种鸡基因,共包含 12 641 个密码子,其中,RSCU 值 ≥ 1.2 的密码子有 16 种,为鸡基因组偏好密码子,偏好性较强的为 CTG(RSCU>2);偏好密码子中,12 种为 G/C 结尾的密码子,占大多数,表明鸡 31 个基因总体偏好使用 G/C 结尾的密码子。由此可见,鸡基因组,密码子偏好性与 *MyoG* 基因的偏好性类似,表明鸡基因组可能具有较强的保守性。

表 2 condonW 程序分析 *MyoG* 基因与鸡 31 个基因的密码子偏好性

Tab.2 Analysis of codon bias of *MyoG* and thirty-one genes in chicken by codonW

| 密码子 Codon | 氨基酸 Amino acids | 个数(<i>MyoG</i> 基因) Number | 同义密码子相对使用度 (<i>MyoG</i> 基因)RSCU value | 个数(31 个 基因)Number | 同义密码子相对使用度 (31 个鸡基因)RSCU value |
|--------------|--------------------|-------------------------------|---|----------------------|-----------------------------------|
| GCA | A(Ala) | 4 | 0.89 | 234 | 1.12 |
| GCC | A | 6 | <u>1.33</u> | 295 | <u>1.41</u> |
| GCG | A | 3 | 0.67 | 80 | 0.38 |
| GCT | A | 5 | 1.11 | 227 | 1.09 |
| TGC | C(Cys) | 7 | <u>2.00</u> | 240 | <u>1.26</u> |
| TGT | C | 0 | 0.00 | 140 | 0.74 |
| GAC | D(Asp) | 5 | <u>1.25</u> | 298 | 0.95 |
| GAT | D | 3 | 0.75 | 327 | 1.05 |
| GAA | E(Glu) | 3 | 0.22 | 378 | 0.91 |
| GAG | E | 24 | <u>1.78</u> | 456 | 1.09 |
| TTC | F(Phe) | 4 | 0.89 | 310 | <u>1.23</u> |
| TTT | F | 5 | 1.11 | 196 | 0.77 |
| GGA | G(Gly) | 0 | 0.00 | 214 | 1.18 |
| GGC | G | 5 | <u>2.50</u> | 206 | 1.13 |
| GGG | G | 3 | <u>1.50</u> | 186 | 1.02 |
| GGT | G | 0 | 0.00 | 122 | 0.67 |
| CAC | H(His) | 3 | <u>2.00</u> | 173 | 1.17 |
| CAT | H | 0 | 0.00 | 122 | 0.83 |
| ATA | I(Ile) | 0 | 0.00 | 128 | 0.57 |
| ATC | I | 7 | <u>3.00</u> | 313 | <u>1.40</u> |
| ATT | I | 0 | 0.00 | 231 | 1.03 |
| AAA | K(Lys) | 3 | 0.67 | 334 | 0.92 |
| AAG | K | 6 | <u>1.33</u> | 395 | 1.08 |
| CTA | L(Leu) | 0 | 0.00 | 71 | 0.35 |

续表 2:

| 密码子 Codon | 氨基酸 Amino acids | 个数(<i>MyoG</i> 基因) Number | 同义密码子相对使用度 (<i>MyoG</i> 基因) RSCU value | 个数(31 个 基因) Number | 同义密码子相对使用度 (31 个鸡基因) RSCU value |
|--------------|--------------------|-------------------------------|--|-----------------------|------------------------------------|
| CTC | L | 7 | <u>1.75</u> | 239 | 1.18 |
| CTG | L | 14 | <u>3.50</u> | 527 | <u>2.60</u> |
| CTT | L | 0 | 0.00 | 141 | 0.70 |
| TTA | L | 0 | 0.00 | 88 | 0.43 |
| TTG | L | 3 | 0.75 | 148 | 0.73 |
| ATG | M(Met) | 1 | 1.00 | 322 | 1.00 |
| AAC | N(Asn) | 9 | <u>2.00</u> | 302 | 1.13 |
| AAT | N | 0 | 0.00 | 231 | 0.87 |
| CCA | P(Pro) | 3 | 0.71 | 192 | 1.18 |
| CCC | P | 8 | <u>1.88</u> | 177 | 1.09 |
| CCG | P | 1 | 0.24 | 84 | 0.52 |
| CCT | P | 5 | 1.18 | 197 | <u>1.21</u> |
| CAA | Q(Gln) | 3 | 0.46 | 144 | 0.60 |
| CAG | Q | 10 | <u>1.54</u> | 334 | <u>1.40</u> |
| AGA | R(Arg) | 1 | 0.29 | 140 | <u>1.38</u> |
| AGG | R | 3 | 0.86 | 128 | <u>1.26</u> |
| CGA | R | 0 | 0.00 | 54 | 0.53 |
| CGC | R | 9 | <u>2.57</u> | 116 | 1.14 |
| CGG | R | 6 | <u>1.71</u> | 114 | 1.12 |
| CGT | R | 2 | 0.57 | 58 | 0.57 |
| AGC | S(Ser) | 12 | <u>3.43</u> | 221 | <u>1.39</u> |
| AGT | S | 0 | 0.00 | 132 | 0.83 |
| TCA | S | 2 | 0.57 | 157 | 0.98 |
| TCC | S | 5 | <u>1.43</u> | 212 | <u>1.33</u> |
| TCG | S | 1 | 0.29 | 43 | 0.27 |
| TCT | S | 1 | 0.29 | 192 | <u>1.20</u> |
| ACA | T(Thr) | 0 | 0.00 | 229 | <u>1.26</u> |
| ACC | T | 5 | <u>2.22</u> | 232 | <u>1.27</u> |
| ACG | T | 3 | <u>1.33</u> | 97 | 0.53 |
| ACT | T | 1 | 0.44 | 170 | 0.93 |
| GTA | V(Val) | 0 | 0.00 | 94 | 0.49 |
| GTC | V | 1 | 0.44 | 191 | 0.99 |
| GTG | V | 8 | <u>3.56</u> | 329 | <u>1.71</u> |
| GTT | V | 0 | 0.00 | 157 | 0.81 |
| TGG | W(Trp) | 2 | 1.00 | 172 | 1.00 |
| TAC | Y(Tyr) | 5 | <u>2.00</u> | 238 | <u>1.22</u> |
| TAT | Y | 0 | 0.00 | 152 | 0.78 |

注: ≥ 1.2 的 RSCU 值用下划线表示,说明该密码子使用具有较强的偏好性。

Note: RSCU value with underline means $\text{RSCU} \geq 1.2$ and the codon have a strong codon bias.

2.2 与其他基因组密码子偏好性比较

转基因研究常常需要进行基因的异源表达,虽然影响外源基因在宿主中表达的因素较多,但是选用合适的密码子对异源表达具有重要的影响。如果宿主表达系统与外源基因之间存在大量差异较大的罕见密码子,特别是连续分布的罕见密码子,就会降低表达量或导致翻译提前终止。而 *MyoG* 基因与宿主基因组的密码子差异会影响到表达水平,因此,研究鸡 *MyoG* 基因外源表达之前,必须考虑其与主要

宿主表达系统的密码子的偏好性差异。大肠杆菌和酵母菌作为原核和真核表达系统,小鼠作为模式动物常作为宿主用于外源表达研究,为预测 *MyoG* 基因能否在其中正常表达,有必要进行密码子偏好性比较。本研究将京海黄鸡 *MyoG* 基因与大肠杆菌、酵母菌和小鼠基因组密码子使用频率进行了比较(表 3),其中,C/E、C/Y 和 C/M 分别代表鸡 *MyoG* 基因与大肠杆菌、酵母菌和小鼠基因组密码子使用频率的比值。结果表明,与大肠杆菌和酵母菌的比

值中, 分别有 18, 25 种密码子比值 ≥ 2.0 或 ≤ 0.5 , 说明就 *MyoG* 基因而言, 大肠杆菌表达系统要优于酵母表达系统。与小鼠密码子使用频率比值中, 仅有 13 种密码子的比值 ≥ 2.0 或 ≤ 0.5 , 大大低于大肠杆菌和酵母菌, 说明小鼠为研究 *MyoG* 基因的理想模式生物。

表 3 *MyoG* 基因与 3 个物种基因组密码子使用频率的比值

Tab. 3 The ratio of codon usage frequencies between *MyoG* gene and genomes of three species

| 密码子 Codon | 氨基酸 Amino acids | 鸡 <i>MyoG</i> 基因 Chicken <i>MyoG</i> gene | 大肠杆菌基因组 <i>E. coli</i> genome | 酵母菌基因组 Yeast genome | 小鼠基因组 Mouse genome | C/E | C/Y | C/M |
|--------------|--------------------|--|----------------------------------|------------------------|-----------------------|-------------|--------------|-------------|
| GCA | A(Ala) | 17.54 | 23.00 | 16.20 | 15.80 | 0.76 | 1.08 | 1.11 |
| GCC | A | 26.32 | 21.60 | 12.60 | 26.00 | 1.22 | <u>2.09</u> | 1.01 |
| GCG | A | 13.16 | 21.10 | 6.20 | 6.40 | 0.62 | <u>2.12</u> | <u>2.06</u> |
| GCT | A | 21.93 | 18.90 | 21.20 | 20.00 | 1.16 | 1.03 | 1.10 |
| TGC | C(Cys) | 30.70 | 5.50 | 4.80 | 12.30 | <u>5.58</u> | <u>6.40</u> | <u>2.50</u> |
| TGT | C | 0.00 | 5.90 | 8.10 | 11.40 | 0.00 | 0.00 | 0.00 |
| GAC | D(Asp) | 21.93 | 17.90 | 20.20 | 26.00 | 1.23 | 1.09 | 0.84 |
| GAT | D | 13.16 | 33.70 | 37.60 | 21.00 | <u>0.39</u> | <u>0.35</u> | 0.63 |
| GAA | E(Glu) | 13.16 | 35.10 | 45.60 | 27.00 | <u>0.37</u> | <u>0.29</u> | <u>0.49</u> |
| GAG | E | 105.26 | 19.40 | 19.20 | 39.40 | <u>5.43</u> | <u>5.48</u> | <u>2.67</u> |
| TTC | F(Phe) | 17.54 | 13.90 | 18.40 | 21.80 | 1.26 | 0.95 | 0.80 |
| TTT | F | 21.93 | 24.40 | 26.10 | 17.20 | 0.90 | 0.84 | 1.28 |
| GGA | G(Gly) | 0.00 | 13.60 | 10.90 | 16.80 | 0.00 | 0.00 | 0.00 |
| GGC | G | 21.93 | 20.60 | 9.80 | 21.20 | 1.06 | <u>2.24</u> | 1.03 |
| GGG | G | 13.16 | 12.30 | 6.00 | 15.20 | 1.07 | <u>2.19</u> | 0.87 |
| GGT | G | 0.00 | 23.70 | 23.90 | 11.40 | 0.00 | 0.00 | 0.00 |
| CAC | H(His) | 13.16 | 7.30 | 7.80 | 15.30 | 1.80 | 1.69 | 0.86 |
| CAT | H | 0.00 | 12.40 | 13.60 | 10.60 | 0.00 | 0.00 | 0.00 |
| ATA | I(Ile) | 0.00 | 13.30 | 17.80 | 7.40 | 0.00 | 0.00 | 0.00 |
| ATC | I | 30.70 | 19.40 | 17.20 | 22.50 | 1.58 | 1.79 | 1.36 |
| ATT | I | 0.00 | 29.60 | 30.10 | 15.40 | 0.00 | 0.00 | 0.00 |
| AAA | K(Lys) | 13.16 | 37.20 | 41.90 | 21.90 | <u>0.35</u> | <u>0.31</u> | 0.60 |
| AAG | K | 26.32 | 15.30 | 30.80 | 33.60 | 1.72 | 0.85 | 0.78 |
| CTA | L(Leu) | 0.00 | 5.60 | 13.40 | 8.10 | 0.00 | 0.00 | 0.00 |
| CTC | L | 30.70 | 9.50 | 5.40 | 20.20 | <u>3.23</u> | <u>5.69</u> | 1.52 |
| CTG | L | 61.40 | 37.40 | 10.50 | 39.50 | 1.64 | <u>5.85</u> | 1.55 |
| CTT | L | 0.00 | 14.50 | 12.30 | 13.40 | 0.00 | 0.00 | 0.00 |
| TTA | L | 0.00 | 17.40 | 26.20 | 6.70 | 0.00 | 0.00 | 0.00 |
| TTG | L | 13.16 | 12.90 | 27.20 | 13.40 | 1.02 | <u>0.48</u> | 0.98 |
| ATG | M(Met) | 4.39 | 23.70 | 20.90 | 22.80 | <u>0.19</u> | <u>0.21</u> | <u>0.19</u> |
| AAC | N(Asn) | 39.47 | 20.30 | 24.80 | 20.30 | 1.94 | 1.59 | 1.94 |
| AAT | N | 0.00 | 29.30 | 35.70 | 15.60 | 0.00 | 0.00 | 0.00 |
| CCA | P(Pro) | 13.16 | 9.10 | 18.30 | 17.30 | 1.45 | 0.72 | 0.76 |
| CCC | P | 35.09 | 6.20 | 6.80 | 18.20 | <u>5.66</u> | <u>5.16</u> | 1.93 |
| CCG | P | 4.39 | 14.50 | 5.30 | 6.20 | <u>0.30</u> | 0.83 | 0.71 |
| CCT | P | 21.93 | 9.50 | 13.50 | 18.40 | <u>2.31</u> | 1.62 | 1.19 |
| CAA | Q(Gln) | 13.16 | 14.40 | 27.30 | 12.00 | 0.91 | <u>0.48</u> | 1.10 |
| CAG | Q | 43.86 | 26.70 | 12.10 | 34.10 | 1.64 | <u>3.62</u> | 1.29 |
| AGA | R(Arg) | 4.39 | 7.10 | 21.30 | 12.10 | 0.62 | <u>0.21</u> | <u>0.36</u> |
| AGG | R | 13.16 | 4.00 | 9.20 | 12.20 | <u>3.29</u> | 1.43 | 1.08 |
| CGA | R | 0.00 | 4.80 | 3.00 | 6.60 | 0.00 | 0.00 | 0.00 |
| CGC | R | 39.47 | 14.00 | 2.60 | 9.40 | <u>2.82</u> | <u>15.18</u> | <u>4.20</u> |
| CGG | R | 26.32 | 7.90 | 1.70 | 10.20 | <u>3.33</u> | <u>15.48</u> | <u>2.58</u> |
| CGT | R | 8.77 | 15.90 | 6.40 | 4.70 | 0.55 | 1.37 | 1.87 |
| AGC | S(Ser) | 52.63 | 14.30 | 9.80 | 19.70 | <u>3.68</u> | <u>5.37</u> | <u>2.67</u> |

续表 3:

| 密码子 Codon | 氨基酸 Amino acids | 鸡 <i>MyoG</i> 基因 Chicken <i>MyoG</i> gene | 大肠杆菌基因组 <i>E. coli</i> genome | 酵母菌基因组 Yeast genome | 小鼠基因组 Mouse genome | C/E | C/Y | C/M |
|--------------|--------------------|--|----------------------------------|------------------------|-----------------------|-------------|-------------|-------------|
| AGT | S | 0.00 | 13.20 | 14.20 | 12.70 | 0.00 | 0.00 | 0.00 |
| TCA | S | 8.77 | 13.10 | 18.70 | 11.80 | 0.67 | <u>0.47</u> | 0.74 |
| TCC | S | 21.93 | 9.70 | 14.20 | 18.10 | <u>2.26</u> | 1.54 | 1.21 |
| TCG | S | 4.39 | 8.20 | 8.60 | 4.20 | 0.53 | 0.51 | 1.04 |
| TCT | S | 4.39 | 13.10 | 23.50 | 16.20 | <u>0.33</u> | <u>0.19</u> | <u>0.27</u> |
| ACA | T(Thr) | 0.00 | 15.10 | 17.80 | 16.00 | 0.00 | 0.00 | 0.00 |
| ACC | T | 21.93 | 18.90 | 12.70 | 19.00 | 1.16 | 1.73 | 1.15 |
| ACG | T | 13.16 | 13.60 | 8.00 | 5.60 | 0.97 | 1.64 | <u>2.35</u> |
| ACT | T | 4.39 | 13.10 | 20.30 | 13.70 | <u>0.33</u> | <u>0.22</u> | <u>0.32</u> |
| GTA | V(Val) | 0.00 | 13.10 | 11.80 | 7.40 | 0.00 | 0.00 | 0.00 |
| GTC | V | 4.39 | 13.10 | 11.80 | 15.40 | <u>0.33</u> | <u>0.37</u> | <u>0.28</u> |
| GTG | V | 35.09 | 19.90 | 10.80 | 28.40 | 1.76 | <u>3.25</u> | 1.24 |
| GTT | V | 0.00 | 21.60 | 22.10 | 10.70 | 0.00 | 0.00 | 0.00 |
| TGG | W(Trp) | 8.77 | 13.40 | 10.40 | 12.50 | 0.65 | 0.84 | 0.70 |
| TAC | Y(Tyr) | 21.93 | 11.70 | 14.80 | 16.10 | 1.87 | 1.48 | 1.36 |
| TAT | Y | 0.00 | 21.60 | 18.80 | 12.20 | 0.00 | 0.00 | 0.00 |

注:下划线表示 2 个物种密码子比较具有明显偏差(≤ 0.5 , ≥ 2)的分值。

Note: Value with underline (≤ 0.5 , ≥ 2) means significant difference between two species.

2.3 与其他动物 *MyoG* 基因密码子偏好性比较

各物种 *MyoG* 基因 ENc 值、GC3s 和 GC 含量结果见表 4。所选 31 种动物 *MyoG* 基因的 ENc 值为 32.46 ~ 53.99, 平均值为 40.36。大部分已知物种的 *MyoG* 基因 CDS 序列均一性较好, 偏好性一般。一般来说, ENc 值越小表明该基因的表达水平越高, ENc 值小于 30 或大于 55 的基因可被预测为高表达或低表达基因^[16]。数据分析表明, 野猪 (*Sus scrofa*)、东非狒狒 (*Papio anubis*)、猎隼 (*Falco cherrug*)

等 11 种动物 ENc 值在 35 左右或小于 35, 表达水平较高; 其余动物, 除非洲爪蟾外, ENc 值大都处于 40 ~ 50, 表达水平一般, 总体而言, *MyoG* 基因在包括京海黄鸡在内的 32 个物种中均有较高表达。32 个物种中, 除非洲爪蟾外, GC3s 含量均大于 0.664, 说明动物中, 几乎所有 *MyoG* 基因偏好以 G/C 结尾的密码子; 32 种动物 GC 含量, 也只有非洲爪蟾一种动物低于 0.5, 说明几乎所有 *MyoG* 基因编码区对 GC 有较强的偏好性。

表 4 32 个物种的 ENc 值、GC3s 和 GC 含量

Tab. 4 ENc value, GC3s and GC content of 32 species

| 物种 Species | 序列登录号 GenBank accession No. | ENc | GC3s | GC |
|---|--------------------------------|-------|-------|-------|
| 非洲爪蟾 <i>Xenopus (Silurana) tropicalis</i> | NM_001016725 | 53.99 | 0.472 | 0.462 |
| 斑马鱼 <i>Danio rerio</i> | NM_131006 | 49.33 | 0.664 | 0.548 |
| 草鱼 <i>Ctenopharyngodon idella</i> | JQ793897 | 48.13 | 0.673 | 0.560 |
| 大鼠 <i>Rattus norvegicus</i> | NM_017115 | 46.74 | 0.731 | 0.603 |
| 罗非鱼 <i>Oreochromis niloticus</i> | NM_001279526 | 45.82 | 0.697 | 0.585 |
| 小鼠 <i>Mus musculus</i> | NM_031189 | 44.08 | 0.748 | 0.604 |
| 金色大鼠 <i>Mesocricetus auratus</i> | XM_005073375 | 43.91 | 0.721 | 0.598 |
| 松江鲈 <i>Trachidermus fasciatus</i> | JQ905626 | 43.20 | 0.686 | 0.586 |
| 绵羊 <i>Ovis aries</i> | NM_001174109 | 41.87 | 0.769 | 0.631 |
| 牦牛 <i>Bos mutus</i> | XM_005895606 | 41.80 | 0.774 | 0.631 |
| 牛 <i>Bos taurus</i> | NM_001111325 | 41.80 | 0.774 | 0.631 |
| 瘤牛 <i>Bos indicus</i> | EU332914 | 41.78 | 0.774 | 0.631 |
| 赤鹿 <i>Cervus elaphus kansuensis</i> | FJ746497 | 41.73 | 0.763 | 0.628 |
| 藏羚羊 <i>Pantholops hodgsonii</i> | XM_005959599 | 41.01 | 0.774 | 0.632 |
| 水牛 <i>Bubalus bubalis</i> | KC107779 | 40.81 | 0.792 | 0.634 |
| 火鸡 <i>Meleagris gallopavo</i> | AY560111 | 40.70 | 0.776 | 0.614 |
| 甲鱼 <i>Pelodiscus sinensis</i> | AB480162 | 40.68 | 0.830 | 0.623 |

续表 4:

| 物种 Species | 序列登录号 GenBank accession No. | ENc | GC3s | GC |
|-----------------------------------|--------------------------------|-------|-------|-------|
| 山羊 <i>Capra hircus</i> | NM_001285733 | 40.53 | 0.783 | 0.635 |
| 原鸽 <i>Columba livia</i> | XM_005505337 | 39.90 | 0.830 | 0.642 |
| 抹香鲸 <i>Physeter catodon</i> | XM_007121541 | 38.89 | 0.786 | 0.629 |
| 猫 <i>Felis catus</i> | XM_003999381 | 37.94 | 0.818 | 0.641 |
| 野猪 <i>Sus scrofa</i> | NM_001012406 | 36.31 | 0.782 | 0.628 |
| 东非狒狒 <i>Papio anubis</i> | XM_003893278 | 35.76 | 0.809 | 0.637 |
| 猎隼 <i>Falco cherrug</i> | XM_005447228 | 35.61 | 0.866 | 0.646 |
| 游隼 <i>Falco peregrinus</i> | XM_005239599 | 35.61 | 0.866 | 0.646 |
| 苏门答腊猩猩 <i>Pongo abelii</i> | XM_003775454 | 35.49 | 0.805 | 0.637 |
| 人 <i>Homo sapiens</i> | NM_002479 | 35.43 | 0.809 | 0.638 |
| 京海黄鸡 Jinghai yellow Chicken | KJ130962 | 35.24 | 0.817 | 0.631 |
| 家兔 <i>Oryctolagus cuniculus</i> | FJ605116 | 34.03 | 0.886 | 0.661 |
| 绿头鸭 <i>Anas platyrhynchos</i> | GQ303573 | 33.27 | 0.897 | 0.656 |
| 白喉雀 <i>Zonotrichia albicollis</i> | XM_005492447 | 32.68 | 0.904 | 0.670 |
| 中地雀 <i>Geospiza fortis</i> | XM_005426784 | 32.46 | 0.904 | 0.676 |

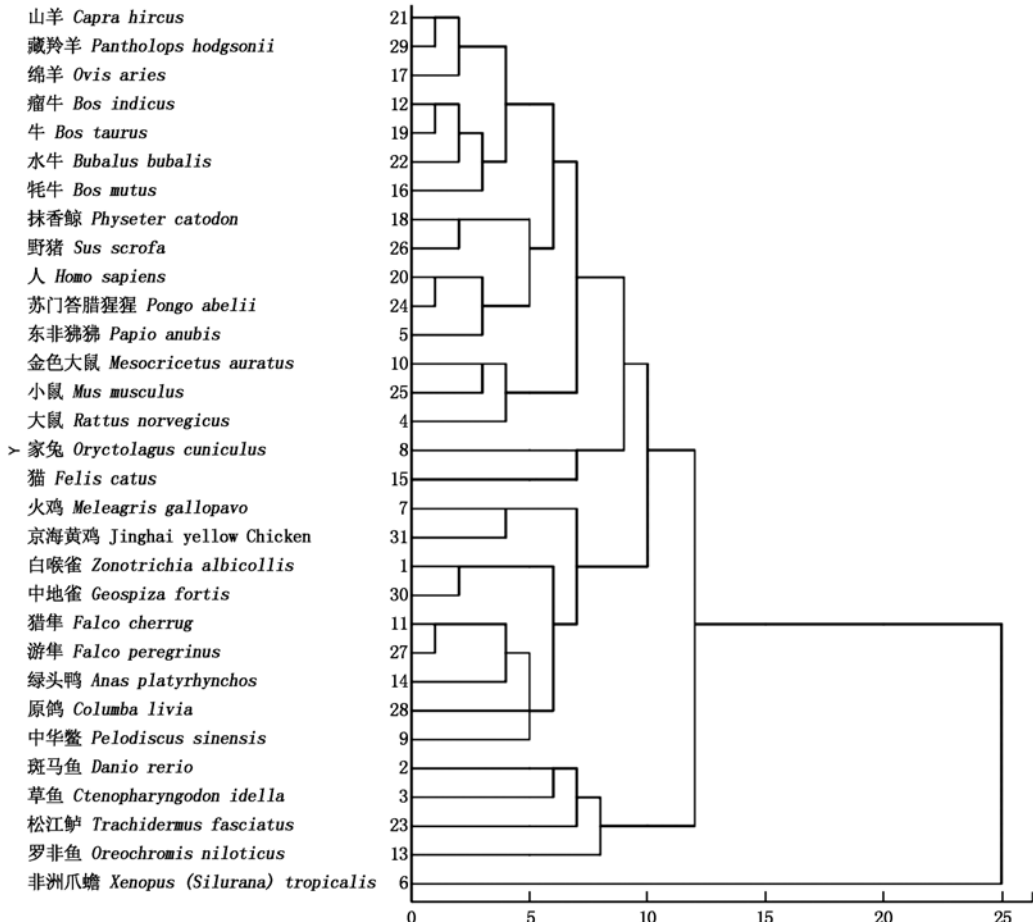


图 1 基于 RSCU 值的 *MyoG* 密码子偏好性聚类树

Fig. 1 Clustering tree of *MyoG* codon bias based on RSCU value

2.4 *MyoG* 基因密码子偏好性聚类分析

使用 SPSS 19 软件对 32 种动物 *MyoG* 基因的 RSCU 值进行聚类分析,采用系统聚类的方法,具体方法参照刘汉梅等^[17]的介绍。聚类结果见图 1,大部分同一纲的物种都聚在了一起,说明亲缘关系近

的物种之间倾向于拥有相同的密码子偏好性,但是,中华鳖却和鸟类聚为一支,同时也说明,亲缘关系较远的物种之间也可能具有相似的密码子偏好性。32 个物种中,非洲爪蟾作为两栖类模式动物,单独聚为一支,说明两栖类动物的密码子偏好性与其他动物

具有明显差异,而其他物种在密码子偏好性上,均有相似性。我们同时利用 DNAMAN 软件构建了 32 个物种 *MyoG* 基因 CDS 区的系统发育树(图 2),结果表明,系统发育大致分类与密码子偏好性的聚类结果类似,但并不是完全一致,说明该 32 个物种在亲缘关系远近和密码子偏好性上可能存在关联性。

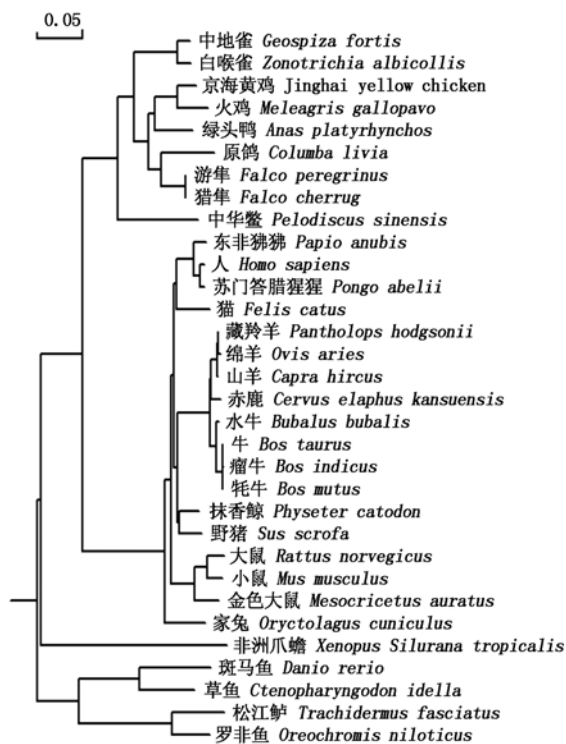


图 2 32 个物种 *MyoG* 基因 CDS 区系统发育树

Fig. 2 Phylogenetic tree of *MyoG* gene CDS region of thirty-two species

3 讨论

同一物种不同基因对密码子的偏好性都可能存在差异^[18-20]。本研究使用相关在线软件和工具分析了鸡 *MyoG* 基因的密码子特性,分析结果表明,鸡 *MyoG* 基因的 ENc 值、GC 含量和 GC3s 分别为 35.24, 0.631, 0.817, 在所有密码子中, GAG 使用的频率高(105.2)其次为 CTG(61.4)。鸡 *MyoG* 基因的 ENc 值较小,密码子存在偏好性; GC3s 大于 0.5, 说明鸡 *MyoG* 基因偏好以 G/C 结尾的密码子; GC 含量大于 0.5, 说明鸡 *MyoG* 基因 CDS 序列中, GC 含量大于 AT 含量。codonW 计算结果表明, 在鸡 *MyoG* 基因密码子中, 有 22 种密码子的 RSCU 值 ≥ 1.2 , 为 *MyoG* 基因的偏好密码子, 其中偏好性较强的是 GGC、ATC、CGC、AGC、ACC、CTG 和 GTG (RSCU > 2), 所有偏好密码子全部以 G/C 结尾, 这与在线工具 CUSP 和 CHIPS 的预测结果一致。与鸡 31 个基因比较显示, 鸡 *MyoG* 基因以及鸡的其他基

因的密码子偏好性差异较小, 这可能与 *MyoG* 基因的保守性较强或者与蛋白结构和功能有关, 具体原因还需进一步研究。

了解基因与外源表达宿主的密码子差异有利于正确选择宿主, 从而顺利表达基因。本研究将京海黄鸡 *MyoG* 基因的密码子偏好性与大肠杆菌和酵母菌基因组密码子偏好性进行了比较, 结果显示, 大肠杆菌与鸡 *MyoG* 基因的密码子偏好性差异较小, 适合其进行外源表达。有研究表明, 转基因研究常常需要进行基因的异源表达, 而基因密码子用法在不同物种差异较大, 因而, 分析宿主系统与外源基因的密码子特性差异, 并改造一些密码子, 使其与宿主基因组密码子特性相符合, 对提高转录和翻译的效率, 进而提高表达量有重要意义。在本研究中, 将京海黄鸡 *MyoG* 基因密码子使用特性与模式动物小鼠基因组密码子使用特性进行比较分析, 发现具有明显偏差的密码子仅为 13 种, 大大低于大肠杆菌和酵母菌, 显示小鼠可能为 *MyoG* 异源表达的理想宿主。

通过比较不同物种 *MyoG* 基因可知, 除非洲爪蟾外, 大部分动物的 ENc 值均低于 50, 显示 *MyoG* 基因在大部分动物体内的表达水平较高, 这与 *MyoG* 蛋白的功能有关。GC3s 含量显示, 几乎所有 *MyoG* 基因偏好以 G/C 结尾的密码子; 32 种动物 GC 含量, 也只有非洲爪蟾一种动物低于 0.5, 说明几乎所有 *MyoG* 基因编码区对 GC 有较强的偏好性, 两栖类动物在密码子偏好性方面有自己独特的特性。基于基因 RSCU 值的系统聚类分析显示亲缘关系较近的物种之间倾向于拥有相同的密码子偏性, 系统发育与密码子偏好性的聚类结果类似, 说明该 32 个物种在亲缘关系远近和密码子偏好性上存在关联性。在聚类分析和系统发育分析中, 非洲爪蟾均单独聚为一类, 在此印证了两栖类动物在密码子偏好性方面有自己独特的特性。虽然偏好性聚类分析和系统发育分析结果相似, 但也存在诸多不同之处, 造成这一现象的主要原因可能是单基因在进化上有可能存在较大的突变, 密码子使用偏好性由于突变发生较大变化, 而分类地位与真实的分类地位存之间存在差异。Christianson 等^[21]认为真实的系统发育关系并不一定完全与密码子偏好性的聚类结果对应, 这一指标在较小的分类单元里比较可靠, 但是在分歧较大的分类单元里, 这一方法主要还是作为传统系统树构建方法的补充和参考。

参考文献:

[1] Urrutia A O, Hurst L D. Codon usage bias covaries with

- expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection[J]. *Genetics*, 2001, 159(3): 1191 – 1199.
- [2] Sueoka N, Kawanishi Y. DNA G + C content of the third codon position and codon usage biases of human genes [J]. *Gene*, 2000, 261(1): 53 – 62.
- [3] 李 平, 白云凤, 冯瑞云, 等. 籽粒苋苹果酸酶(NAD-ME)基因密码子偏好性分析[J]. *应用与环境生物学报*, 2011, 17(1): 12 – 17.
- [4] Bulmer M. The selection-mutation-drift theory of synonymous codon usage [J]. *Genetics*, 1991, 129(3): 897 – 907.
- [5] 张 乐, 金龙国, 罗 玲, 等. 大豆基因组和转录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.
- [6] 吴正常, 王 靖, 赵乔辉, 等. 猪脂多糖结合蛋白基因(*LBP*)的密码子偏好性分析[J]. *农业生物技术学报*, 2013, 21(10): 1135 – 1144.
- [7] Wright W E, Sassoon D A, Lin V K. Myogenin, a factor regulating myogenesis, has a domain homologous to *MyoD* [J]. *Cell*, 1989, 56(4): 607 – 617.
- [8] Rudnicki M A, Jaenisch R. The *MyoD* family of transcription factors and skeletal Myogenesis [J]. *Bioessays*, 1995, 17(3): 203 – 209.
- [9] Edmondson, Olson E. Helix-loop-helix proteins as regulators of muscle-specific transcription [J]. *J Biol Chem*, 1993, 268(2): 755 – 758.
- [10] Olson E N, Klein W H. bHLH factors in muscle development: deadlines and commitments, what to leave in and what to leave out [J]. *Genet Dev*, 1994, 8. pp: 1 – 8.
- [11] 高光宇, 王 杰, 王 永, 等. 成都麻羊 *MyoG* 基因的克隆与测序分析[J]. *西南民族大学学报: 自然科学版*, 2009, 35(1): 84 – 88.
- [12] Hastly P, Bradley A, Morris J H, *et al.* Muscle deficiency and neonatal death in mice with a targeted mutation in the *MyoGenin* gene [J]. *Nature*, 1993, 364: 501 – 506.
- [13] 郭秀丽, 王 玉, 杨路成, 等. 茶树 *CBFI* 基因密码子使用特性分析[J]. *遗传*, 2012, 34(12): 1614 – 1623.
- [14] 刘庆坡, 谭 军, 薛庆中. 籼稻品种 93-11 同义密码子的使用偏性[J]. *遗传学报*, 2003, 30(4): 335 – 340.
- [15] 赵 耀, 刘汉梅, 顾 勇, 等. 玉米 *waxy* 基因密码子偏好性分析[J]. *玉米科学*, 2008, 16(2): 16 – 21.
- [16] Sharp P M, Li W H. An evolutionary perspective on synonymous codon usage in unicellular organisms [J]. *Journal of Molecular Evolution*, 1986, 24(12): 28 – 38.
- [17] 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.
- [18] 马晓霞, 冯玉萍, 赵永清, 等. 口蹄疫病毒密码子使用偏性的位点差导[J]. *华北农学报*, 2013, 28(2): 32 – 37.
- [19] 晁岳恩, 常 阳, 王美芽, 等. 7 种作物叶绿体基因的密码子偏好性及聚类分析[J]. *华北农学报*, 2012, 27(4): 60 – 64.
- [20] 杨 平. 水稻 *HKT* 基因家族密码子使用特性分析[J]. *山西农业科学*, 2011, 39(11): 1137 – 1140, 1145.
- [21] Christianson M L. Codon usage patterns distort phylogenies from or of DNA sequences [J]. *American Journal of Botany*, 2005, 92(8): 1221 – 1233.