

doi: 10.13241/j.cnki.pmb.2018.21.006

青藏高原人群铁调素基因 hamp 密码子偏好性分析*

施 静¹ 雷 鸣² 张勇群^{1Δ} 孙舒遥¹ 李明霞¹ 孙曾梅¹ 郭云红¹

(1 西藏自治区人民政府驻成都办事处医院 四川 成都 610041; 2 西藏大学 西藏 拉萨 850000)

摘要 目的:通过对青藏高原人群及平原人群、恒河猴等其他 5 种物种的密码子使用进行分析,从而得出青藏高原人群铁调素基因(hamp)的密码子偏好性。**方法:**采用 PCR 技术获得青藏高原人群铁调素全基因序列,利用在线软件 CodonW 进行密码子偏好性分析,通过在线软件 PredictProtein 以及 Signal P 等软件进行青藏高原人群铁调素基因的结构分析,比较与 GenBank 数据库中选取的平原人群、恒河猴等其他 5 种物种的密码子偏好性的差异。**结果:**青藏高原人群的铁调素基因全长为 2681 bp,由 3 个外显子和 2 个内含子组成,编码 84 个氨基酸的铁调素前体肽,包含信号肽、中间肽与成熟肽,其密码子偏好性与平原人群、恒河猴等其他 5 个物种的密码子偏好性均有不同程度的差异。**结论:**青藏高原人群铁调素基因 hamp 密码子偏好性与其他物种的密码子偏好性均有不同程度的差异。

关键词:青藏高原人群;铁调素基因;密码子偏好性

中图分类号:R-33;Q593;R135.6 **文献标识码:**A **文章编号:**1673-6273(2018)21-4027-08

Analysis of the Codon Bias of Hamp in the Tibetan*

SHI Jing¹, LEI Ming², ZHANG Yong-qun^{1Δ}, SUN Shu-yao¹, LI Ming-xia¹, SUN Zeng-mei¹, WU Yun-hong¹

(1 Hospital of Chengdu Office of People's Government of Tibetan Autonomous Region (Hospital.C.T.), Chengdu, Sichuan, 610041, China;

2 Tibet University, Lasa, Tibet, 850000, China)

ABSTRACT Objective: The codon usage of hamp in the Tibetan and other species which include the plain and the rhesus monkey were analyzed to find the codon bias of hamp in the Tibetan. **Methods:** The PCR technology was used for the full genetic sequence of the Tibetan, using online software CodonW to analysis codon bias, other online software that PredictProtein and Signal P were used to analysis the structure of hamp of the Tibetan and comparing the codon bias of the Tibetan with other 5 species which were selected from the GenBank database. **Results:** The genetic sequence of hamp in Tibetan was 2681 bp, the protein structure with a signal peptide but no transmembrane domain, there was significant difference between the Tibetan and other species in the codon bias. **Conclusions:** There was significant difference between the Tibetan and other species in the codon bias.

Key words: Tibetan; Hamp; Codon bias

Chinese Library Classification(CLC): R-33; Q593; R135.6 **Document code:** A

Article ID: 1673-6273(2018)21-4027-08

前言

铁调素是一种由肝脏合成并分泌后释放入血浆的细胞因子诱导的富含半胱氨酸的抗菌多肽,不仅能抑制某些细菌和真菌的生长繁殖,也是一种控制小肠铁吸收及调节体内铁稳态的关键物质^[1]。铁调素在人体氧化代谢、细胞生长于增殖、氧的运输及铁代谢中扮演至关重要的角色,过多则会促进自由基的产生,引发神经退行性变及多种老年疾病,过少则会造造成贫血等多种疾病(缺铁性贫血、遗传性血色素沉着症、急性高山病、肿瘤等)^[2-4],给人类的健康造成了极大的影响^[5-7]。

目前我们在临床诊断中发现青藏高原人群的铁调素水平普遍高于平原地区体内的铁调素水平,而目前关于青藏高原人群铁调素基因的生物信息学分析较少^[8],又加上铁调素的表达

已经在青藏高原人群和其他平原地区人群中出现了差异,而关于这个差异是否会给青藏高原人群的健康造成影响目前还未见相关报道^[9]。在研究中我们也还未发现关于高原人群铁调素基因(hamp)密码子偏好性的相关研究。因此,对青藏高原人群铁调素基因的研究具有重要意义。

由此,本实验以青藏高原人群铁调素基因 hamp 全长序列为材料,对青藏高原人群 hamp 密码子偏好性进行探究,通过对采集的藏族高原人群外周静脉血进行 DNA 提取及 PCR 扩增,将结果通过在线软件进行铁调素的生物信息学及密码子偏好性分析。本次研究结果将为临床治疗铁代谢异常疾病提供实验基础与理论依据,能更好的为疾病的治疗提供依据。

1 材料与方法

* 基金项目:西藏自治区科技厅 2017 年重点研发与转化计划项目(ZH20170003);成都市卫生和计划生育委员会医学科研课题(2015031)

作者简介:施静(1989-),硕士研究生,主要研究方向:分子生物学,E-mail: 602705261@qq.com

Δ 通讯作者:张勇群(1976-),硕士生导师,副教授,主要研究方向:分子生物学,E-mail: yongqunzhang@yahoo.com

(收稿日期:2018-04-26 接受日期:2018-05-24)

1.1 试验材料

此研究采用从某院常规体检健康者中随机选取的西藏高原人群藏族 1 例,用乙二胺四乙酸三钾 (EDTA-K3) 抗凝采血管收集外周静脉血 2 mL。

1.2 试验方法

1.2.1 DNA 提取及 PCR 扩增 使用血液基因组 DNA 提取试剂盒提取 DNA(天根生化科技有限公司),-80℃ 保存备用。引物使用 Primer5 软件设计(表 1),由生工生物工程股份有限公司

合成,并采用 30 μL 的 PCR 反应扩增体系,其中包含 2× Mix (TsingKe)15 μL、F/R 引物各 2 μL、DNA 模板 2 μL 和 ddH₂O 9 μL。反应条件:95℃ 预变性 5 min,94℃ 变性 30 s,Tm℃ 退火 30 s,72℃ 延伸 30 s,重复 30 个循环。PCR 产物经纯化后用 AB 3730xL Genetic Analyzer 测序仪进行测序。拼接获得西藏高原人群全长序列,进而对西藏高原人群全长序列进行生物信息学分析。

表 1 6 对引物序列及基本信息

Table 1 6 pairs of primer sequences and basic information

Primer name	Primer sequences	Tm values (°C)	PCR product size(bp)
hamp1	5'-CAACCCTGTTCCCTGTCTCG-3'(F)	56.5	382
	5'-AGTGCTCGGGTGTCTCCA-3'(R)		
hamp2	5'-GTTCCCGCTTATCTCTCCCG-3'(F)	60	971
	5'-CTACTGTGTGGCGAATGGGT-3'(R)		
hamp3	5'-ATGAAGCACCGAGTGACAGG-3'(F)	60	692
	5'-GCTTATGGGGGCTTCTCTT-3'(R)		
hamp4	5'-GTGGGACTTGGGGATAAGGC-3'(F)	60	586
	5'-AGAGGTGGGGTCTGTTGTG-3'(R)		
hamp5	5'-AGAAAACCACAATCTCACTGGC-5'(F)	59.5	813
	5'-GGAAGGGAATGTGAGCAGGG-3'(R)		
hamp6	5'-AAGGCCCTCCTAAGAGTCC-3'(F)	59.2	602
	5'-CAACCTCCAGGACACACCAG-3'(R)		

1.2.2 生物信息学分析 将所得序列在 NCBI 上进行 Blast 比对。用 Jellyfish 将所得序列翻译成氨基酸序列。并通过在线工具 SOPMA、Signal P 和 CodonW 等分析生物信息学。

1.2.3 密码子参数计算 采用 CodonW 中的 CUSP 程序对西藏高原人群铁调素基因的密码子组成及偏好性进行参数计算,并与其他物种进行比较,将所有物种的铁调素基因的密码子参数在 SPSS 中进行录入分析各物种铁调素基因密码子的使用模式及偏好性比值。主要计算指标有:有效密码子数(Effective Number of Codon, ENC),可以反映同义密码子非均衡使用的偏好程度,ENC 值范围在 20 到 61,越靠近 20 偏好性越强^[10];密码子第 3 位出现碱基 G 或 C 频率(GC 3 skew, GC3s),表示密码子第 3 位碱基中 G+C 出现的频率^[11];同义密码子相对使用度(Relative Synonymous Codon Usage, RSCU),是某一特定的密码子在编码对应氨基酸的同义密码子中的相对概率,可直观的反映出密码子使用的偏好性程度,RSCU 值等于 1,表明该密码子的使用没有偏好,当 RSCU 值大于 1,则表明该密码子的使用频率较高^[12]。将各物种的铁调素基因密码子 RSCU 值在 SPSS 中进行物种间的差异比较与分类分析。

1.2.4 聚类分析 根据西藏高原人群的铁调素基因特征,在 GenBank 数据库中选取若干物种进行系统进化树分析,利用 MEGA5 分析软件将各物种铁调素氨基酸序列进行归类比较分析。

2 结果

2.1 西藏高原人群铁调素 hamp 基因的结构和特点

编码人类 hamp 基因定位于 19 号染色体,本研究中西藏高原人群铁调素基因 hamp 全长有 2681bp (图 1),GC 含量为 0.603,与 NCBI 数据库中公布的人类全长序列对比同源性为 99.7%,有 7 个碱基有差异,其中 NCBI 数据库中公布的铁调素全长序列的位点 1385 的 T 碱基、1725 位点的碱基 T、2030 位点的碱基 G、2185 位点的碱基 G、2212 位点的碱基 G、2247 位点的碱基 G 和 2276 位点的碱基 G 在西藏人群铁调素全长序列中分别为碱基 G、C、A、A、A、A、A。且西藏高原人群铁调素基因全长序列有 3 个外显子和 2 个内含子,其中位点 44~205 是第一段外显子区,2327~2386 是第二外显子区,2476~2580 为第三外显子区。三个外显子拼接成一个 427bp 的开放阅读框,编码 84 个铁调素氨基酸前体肽。

信号肽是引导新合成的蛋白质向分泌通路转移的短肽链(长度 5-30 个氨基酸)^[13]。Signal P 是预测给定的氨基酸序列中是否存在潜在的信号肽剪切位点及其所在的在线软件。C-score 是信号肽的酶切位点值,一般剪切位点处的 C-score 值最高;S-score 则表示信号肽值,一般为信号肽区域的 S-score 最高;而 Y-score 是综合得出的剪切位点的分值。从图中可以看出经 Signal P 预测出该蛋白有一条信号肽,信号肽位置(如图 2)在 1-24 号氨基酸位点(MALSSQIWAACLLLLLLASLTSG),且 1-24 号氨基酸的 mean S-score 为 0.976,远高于分泌性蛋白标准 0.5。

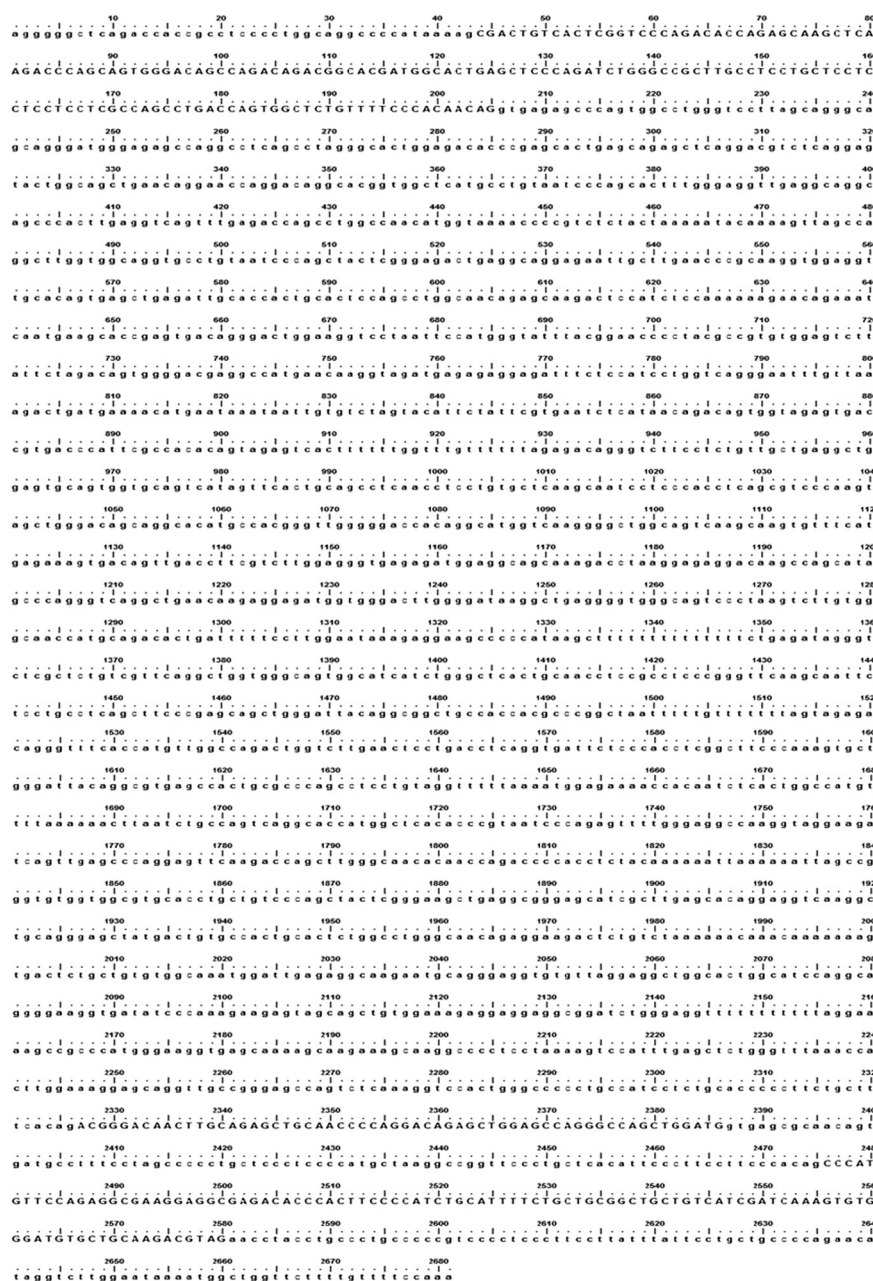


Fig.1 The complete sequence diagram of hamp gene in the Tibetan plateau population

Note: The capital area is the three exon regions.

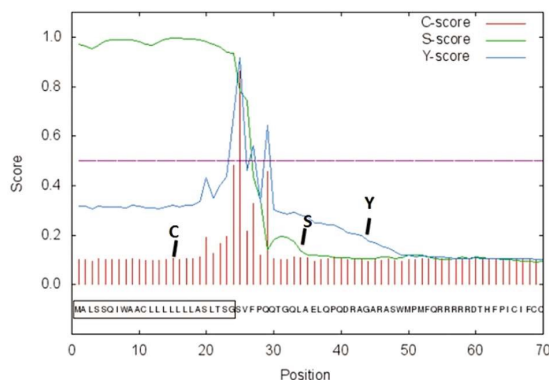


Fig.2 The prediction of peptide signal peptide in the Tibetan plateau population

由西藏高原人群铁调素前体肽信号肽的预测结果可以看出,西藏高原人群铁调素基因中有 74 bp 的开放阅读框编码着信号(MALSSSQIWAACLLLLLLASLTSG)。有文献报道,人类铁调素 hamp 基因编码的成熟肽为(DTHFPICIFCCGC-CHRSKCGMCKKT)^[14],此成熟肽序列与研究中西藏高原人群铁调素前体肽中的一段序列一致,表明西藏高原人群铁调素基因中有 136bp 的开放阅读框编码着成熟的氨基酸。其余阅读框编码的氨基酸则为中间肽(SVFPQQTGQLAELQPQDRAGARASWMPMFQRRRRR)(图 3),成熟氨基酸组分及通过软件 ProtParam 预测铁调素的理化性质,结果见表 2,理论等电点 8.22 为碱性蛋白,总平均疏水性大于 0 是一个疏水性蛋白。

gactgtcactcgggtccagacaccagagcaagctcaagaccagcagtgaggacagccagacagacggcacgATGGCACTGAGCTCC
M A L S S
CAGATCTGGGCCGCTTGCCTCCTGCTCCTCCTCCTCGCCAGCCTGACCACTGGCTCTGTTTTCCACAAC
Q I W A A C L L L L L L L A S L T S G S V F P Q
AGACGGGACAACCTTGCAGAGCTGCAACCCCAGGACAGAGCTGGAGCCAGGGCCAGCTGGATGCCCATGT
Q T G Q L A E L Q P Q D R A G A R A S W M P M
TCCAGAGGCGAAGGAGGCGAGACACCCACTTCCCCATCTGCATTTTCTGCTGCGGCTGCTGTCATCGATCAA
F Q R R R R R D T H F P I C I F C C G C C H R S
AGTGTGGGATGTGCTGCAAGACGTAgaacctacctgccctgccccgtccctccctcctatttattctgctgccccagaacat
K C G M C C K T
aggtcttgaataaaatggctggtcttttttccaaa

图3 铁调素基因拼接的外显子编码的氨基酸序列

Fig.3 The amino acid sequence of the exon coding of ferrostatin gene splicing

Note: The signal peptides are marked in black, intermediate peptides are underlined, and the mature peptides are unlabeled.

表2 铁调素理化性质分析表

Table 2 Analysis table of physical and chemical properties of ferritin

Amino acid numbe	Molecular weight(kDa)	Isoelectric point(pI)	Coefficient of instability	Total average hydrophobicity(GRAVY)	Acid amino acid(%)	Basic amino acid(%)	Polar neutral amino acids (%)	Nonpolar amino acid (%)
25	2797.41	8.22	37.76	0.388	3.57	14.28	40.49	41.66

利用在线预测软件 SOPMA 预测出高原人群铁调素前体肽的二级结构如图 4:α-螺旋是蛋白的主要二级结构, 所占比

例 50%, 其次为无规则卷曲, 29.76%, β-转角和延伸链分别为 14.29%, 5.95%。

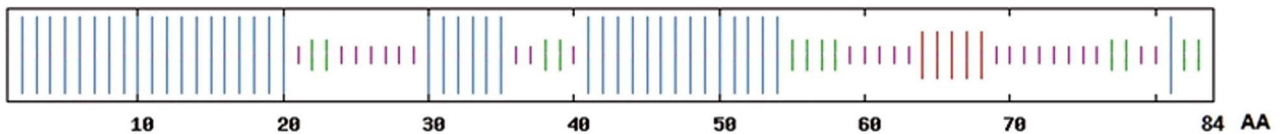


图4 高原人群铁调素前体肽的二级结构预测

Fig.4 Prediction of the secondary structure of precursors in the plateau population

Note: The blue is the alpha - helix, the purple is the unrulous curl, the green is the beta - corner, the red is the extension chain.

2.2 西藏高原人群铁调素 hamp 基因与其他物种的铁调素基因密码子偏好性分析

2.2.1 有效密码子数(ENC)和 GC 含量分析 此研究通过在线程序 codonW 计算高原人群铁调素基因 hamp (此全长序列已上传至 Genbank 数据库, 序列号为 MG679891) 的 ENC、GC 含量和 GC3s 分别为 41.94、0.603、0.705, 发现 ENC 值偏大, 说明西藏高原人群铁调素基因 hamp 中各密码子没有明显的偏好性; 西藏高原人群铁调素基因 hamp 编码区 GC>0.5, 表明西藏高原人群铁调素基因 hamp 在整个编码区序列中 G+C 含量大于 A+T; GC3s>0.5, 表明西藏高原人群铁调素基因 hamp 偏好使用以 G/C 结尾的密码子。通过在线软件分析从 GenBank

数据库中选取的其他物种的铁调素基因密码子偏好性。结果见表 3, 序列的 ENC 值介于 40.89 到 50.96 之间, 平均值为 46.09, 大部分密码子偏好性不强; ENC 值小于 30 或大于 55, 表明该基因可被预测为高表达或低表达基因。结果显示, 研究的 8 个物种的 ENC 值均在 30~55 范围内, 基因表达水平一般。在 8 个物种中除了斑马鱼和小鼠的 GC 含量分别为 0.495、0.5≤0.5, 剩余 6 个物种的 GC 含量均大于 0.5, 表明这些物种的铁调素基因 hamp 在整个编码区序列中 G+C 含量大于 A+T; 且除小鼠物种后剩余 7 个物种的 GC3s 均大于 0.5, 说明这 7 个物种的铁调素基因 hamp 基因偏好使用以 G/C 结尾的密码子。

表3 高原人群与其他物种铁调素基因的 hamp 的碱基构成和各偏好性指标结果

Table 3 The base composition of hamp and the results of the various preference indexes were obtained

Species	Sequence number	ENC	GC3s	GC	T3s	C3s	A3s	G3s
Tibetan(西藏高原)	MG679891	41.94	0.705	0.603	0.1471	0.5441	0.2131	0.3103
Plain(平原人群)	NM_021175	41.94	0.705	0.603	0.1471	0.5441	0.2131	0.3103
mouse(小鼠)	NM_032541.2	49.46	0.456	0.500	0.3214	0.2690	0.3310	0.2920
zebrafish(斑马鱼)	NM_205583.2	50.96	0.578	0.495	0.2958	0.4507	0.2698	0.3333

chimpanzee(黑猩猩)	NM_001109693.1	50.96	0.578	0.607	0.1471	0.5441	0.2097	0.3051
monkey(恒河猴)	XM_015123671.1	47.21	0.696	0.607	0.1571	0.5286	0.2063	0.3000
bat(蝙蝠)	XM_019624941.1	45.34	0.718	0.602	0.1600	0.5067	0.1765	0.3538
frog(青蛙)	NM_001097260.1	40.89	0.734	0.551	0.1667	0.6061	0.1724	0.3333

2.2.2 ENC 与 GC3s 的关联性分析 ENC 是衡量一个基因组偏爱性的重要指数,而 GC3s 是衡量一个基因组同义密码子组成的一个参数。标准曲线表示在非选择作用下,ENC 与 GC3s 的之间的关系,ENC-GC3s 关系图能直观有效的反应基因密码子使用模式,同时可以推断突变和选择作用在基因密码子使用模式中的相对重要性,其中 ENC 与 GC3s 的关系为: $Nc=2+s+29/[s^2+(1-s^2)]$ (注: Nc :有效密码子数; s :GC3s)^[15]。图 5 是铁调素基因 hamp 的 ENC 和 GC3s 的散点图,从图上我们可以看出铁调素基因 hamp 的 ENC 值分布较窄,表明大部分物种的铁调素基因对密码子的偏爱性差别不大,从中可以看出(图 5),只有非常少的基因位于曲线上,这说明突变对铁调素基因 hamp 密码子偏爱性的影响有限,自然选择对该铁调素基因 hamp 密码子偏爱性的影响更大。

2.3 青藏高原人群与其他物种铁调素基因 hamp 的密码子的使用特性分析

对青藏高原人群及其他物种铁调素基因 hamp 的密码子使用特性进行分析,结果见图 6-7,显示出了青藏高原人群中具有 19 个 $RSCU>1$ 的密码子,为高原人群铁调素基因 hamp 的偏好性密码子,其中以 U 结尾的密码子有 1 个,以 C 结尾的密码子有 10 个,以 G 结尾的有 5 个,以 A 结尾的有 3 个,说明青藏高原人群铁调素基因 hamp 的密码子偏好以 C 结尾的密码子。

平原人群、啮齿类小鼠、鱼类斑马鱼、灵长类黑猩猩、飞禽类蝙蝠、以及两栖类青蛙中 $RSCU>1$ 的分别有 19 个、21 个、24 个、19 个、24 个、23 个,都偏好使用以 C 结尾的密码子,可见不同物种的铁调素基因 hamp 大多是偏向使用以 C 结尾的密码子。将各物种的密码子使用特性进行对比分析(图 7)可以看出高原人群铁调素基因对密码子 GUU 的偏好性最强($RSCU=4$),与平原人群的密码子使用模式一致,而与其他 6 个物种的密码子偏好性均有不同程度的差异。

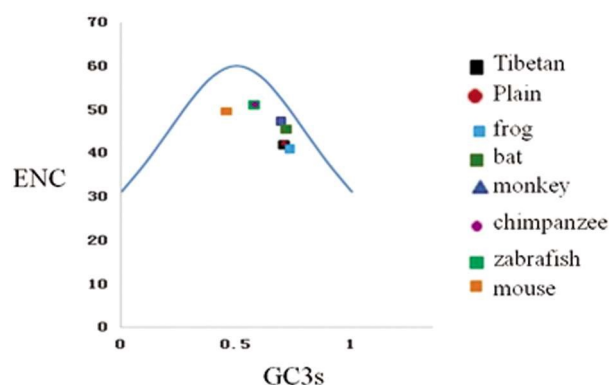


图 5 不同物种的铁调素基因的 ENC-GC3s 关联性分析图
Fig.5 ENC-GC3s correlation analysis diagram of ferrostatin gene in different species

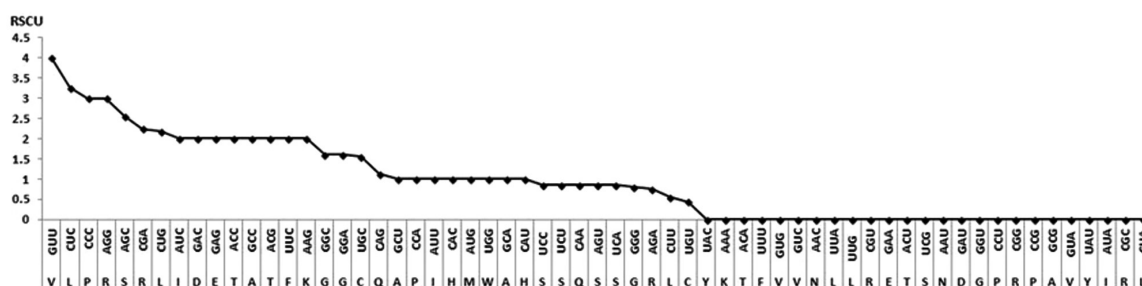


图 6 青藏高原人群铁调素基因 hamp 密码子使用模式
Fig.6 The hamp code of the Tibetan plateau population

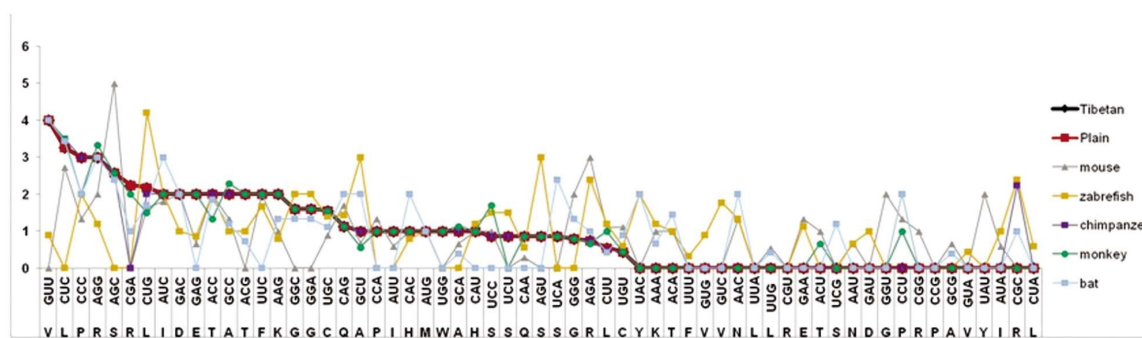


图 7 不同物种铁调素基因 hamp 的密码子使用模式
Fig.7 The code of the hamp gene of different species of ferritin gene was used

2.4 与其他物种铁调素基因 hamp 的密码子偏好性的比较

为了更直观的了解青藏高原人群铁调素基因密码子偏好

性与其他物种铁调素基因偏好性的区别,研究者将选取的 7 个物种的铁调素基因密码子频率与高原人群的铁调素基因密码

子频率进行了比较,结果见(表 4),比值在 0.5~2 之间的表明两物种间的密码子使用模式相近,比值小于 0.5 或大于 2,则表明两物种间的密码子使用模式较远(表 4 中用下划线表示)。在西藏高原人群与平原人群的铁调素基因密码子使用频率的比值结果均为 1,表明西藏高原人群的密码子偏好性的使用与平原人群的密码子偏好性使用一致;在西藏人群与啮齿类动物小鼠的密码子使用频率比值中 ≥ 2.0 或 ≤ 0.5 的密码子有 16 种;与灵长类动物黑猩猩和恒河猴的密码子使用频率比值中 ≥ 2.0 或 ≤ 0.5 的密码子均只有 1 种;与两栖类动物青蛙的密码子使用频率比值中 ≥ 2.0 或 ≤ 0.5 的密码子有 18 种;与飞行动物蝙蝠

的密码子使用频率比值中 ≥ 2.0 或 ≤ 0.5 的密码子有 17 种;与鲤形类斑马鱼的密码子使用频率比值中 ≥ 2.0 或 ≤ 0.5 的密码子有 19 种;表明西藏高原人群的铁调素基因 hamp 与这 7 个物种的铁调素基因密码子使用偏好性均存在不同程度的差异,但与平原人群、灵长类物种的差异普遍小于其他物种的差异。另有 24 个密码子 UUU、UUA、UUG、CUA、AUA、GUC、GUA、GUG、UCG、CCU、CCG、ACU、ACA、GCG、UAC、AAU、AAC、AAA、GAU、GAA、CGU、CGC、CGG、GGU 在西藏高原人群中使用频率为 0,被避免使用,未单独列在表中与其他物种一一进行比较。

表 4 西藏高原人群与其他物种铁调素基因的密码子频率的比值

Table 4 The ratio of the population of the Tibetan to the codon frequency of the other species of hamp gene

Codon	Amin acid	Mouse/ Tibetan	Zebrafish/ Tibetan	Chimpanzee/ Tibetan	Monkey/ Tibetan	Bat/Tibetan	Frog/Tibetan	Plain/ Tibetan
UUC	Phe	1.00	0.84	1.00	1.00	0.00	1.00	1.00
CUU	Leu	1.98	2.18	0.91	1.82	0.78	1.00	1.00
CUC	Leu	0.83	0.00	1.07	1.07	1.05	0.83	1.00
CUG	Leu	0.75	1.93	0.92	0.69	0.78	0.75	1.00
AUU	Ile	0.60	0.00	1.00	1.00	0.00	0.00	1.00
AUC	Ile	0.90	1.00	1.00	1.00	1.50	1.50	1.00
AUG	Met	1.00	1.00	1.00	1.00	1.00	1.00	1.00
GUU	Val	0.00	0.22	1.00	1.00	1.00	0.00	1.00
UCU	Ser	0.00	1.74	1.00	0.00	0.00	0.00	1.00
UCC	Ser	1.16	1.74	1.00	1.99	0.00	4.36	1.00
UCA	Ser	0.00	0.00	1.00	1.00	2.79	0.00	1.00
AGU	Ser	0.00	3.49	1.00	1.00	0.00	0.87	1.00
AGC	Ser	1.95	0.00	1.00	1.00	0.93	0.58	1.00
CCC	Pro	0.44	0.67	1.00	0.67	0.67	0.00	1.00
CCA	Pro	1.33	0.00	1.00	1.00	0.00	2.00	1.00
ACC	Thr	1.00	1.00	1.00	0.67	0.93	1.00	1.00
ACG	Thr	0.00	0.50	1.00	1.00	0.37	1.00	1.00
GCU	Ala	0.67	3.00	1.00	0.57	2.00	2.00	1.00
GCC	Ala	0.67	0.50	1.00	1.15	0.60	1.00	1.00
GCA	Ala	0.67	0.00	1.00	1.14	0.40	0.00	1.00
CAU	His	1.00	1.20	1.00	1.00	0.00	0.00	1.00
CAC	His	1.00	0.80	1.00	1.00	2.00	2.00	1.00
CAA	Gln	0.34	0.66	1.00	1.00	0.00	0.00	1.00
CAG	Gln	1.50	1.25	1.00	1.00	1.75	1.75	1.00
AAG	Lys	0.50	0.40	1.00	1.00	0.67	0.40	1.00
GAC	Asp	1.00	0.50	1.00	1.00	1.00	0.00	1.00
GAG	Glu	0.34	0.43	1.00	1.00	0.00	0.60	1.00
UGU	Cys	2.52	1.36	1.00	1.00	2.02	1.66	1.00
UGC	Cys	0.57	0.90	1.00	1.00	0.71	0.81	1.00
UGG	Trp	0.00	0.00	1.00	1.00	0.00	0.00	1.00
CGA	Arg	0.00	0.00	0.00	0.89	0.44	0.00	1.00
AGA	Arg	4.00	3.20	1.00	0.89	1.33	2.67	1.00
AGG	Arg	0.67	0.40	1.00	1.11	1.00	1.33	1.00
GGA	Gly	0.00	1.25	1.00	1.00	0.83	1.50	1.00
GGC	Gly	0.00	1.25	1.00	1.00	0.83	0.50	1.00
GGG	Gly	2.50	0.00	1.00	1.00	1.66	1.00	1.00

Note: The ratio is greater than or equal to 2 or less than 0.5, indicating that the codon is used with a strong preference.

2.5 各物种的铁调素基因 hamp 密码子偏好性之间的比较

将各物种的铁调素基因 hamp 密码子偏好性 RSCU 值通过 SPSS20.0 软件聚类分析(图 8)可以看出,横坐标代表各物种间密码子相似性的距离。结果可以看出西藏高原人群(图中用 * 标记)与平原人群、恒河猴和黑猩猩的铁调素基因 hamp 密码子偏好性 RSCU 值聚类在同一分支,且分支的距离较近,表明分支的距离越近,物种间密码子的差异越小,表明这 4 个物种之间的铁调素基因密码子偏好性相似性较大;而另外 4 个物种之间的铁调素基因 hamp 密码子偏好性 RSCU 值聚类距离较大,表明他们的物种之间的密码子偏好性差异较大,各物种间的密码子相似性距离大部分不相同,表明不同的物种间的密码子偏好性均有不同程度的差异。

2.6 基于西藏高原人群铁调素基因 hamp 的序列比对与聚类分析

通过 BioEdit 软件和 MEGA5 软件将西藏高原人群铁调素前体肽序列与从 GenBank 中选取的 7 个物种的铁调素基因 hamp 的氨基酸序列进行序列比对(图 9)和聚类分析(图 10)(西藏高原人群用 * 标记)所示。通过铁调素基因 hamp 的氨基酸序列的比对与进化树分支可以看出西藏高原人群与平原人群的铁调素前体肽完全一致,且系统进化树分支西藏高原人群

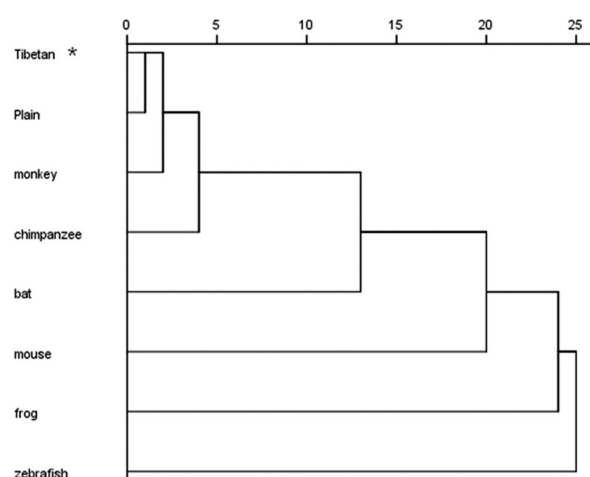


图 8 各物种铁调素基因 hamp 基因密码子聚类分析结果

Fig.8 The results of the hamp gene cluster analysis of gene hamp gene in each species

与平原人群、恒河猴、黑猩猩在同一分支上,表明与这四类物种基因信息相似性较大。

		identities
高原人群	MALSSQIWAACLLLLLLASLTSGSVFPQQTGQLAELQPQDRAGA-----RASWM-PMFQRRRRRDTHFPICFCCGCHRSKCGMCCKT	100%
平原人群	MALSSQIWAACLLLLLLASLTSGSVFPQQTGQLAELQPQDRAGA-----RASWM-PMFQRRRRRDTHFPICFCCGCHRSKCGMCCKT	100%
黑猩猩	MALSSQIWAACLLLLLLASLTSGSVFPQQTGQLAELQPQDRAGA-----RASWM-PMLQRRRRRDTHFPICFCCGCHRSKCGMCCKT	99%
恒河猴	MALSSQIWAACLLLLLLASLTSGSVFPQQTGQLAELQPQDRAGA-----RASWT-PMLQRRRRRDTHFPICFCCGCHRSKCGMCCKT	95%
小鼠	MALSTRQAAACLLLLLLASLTSTYLYHQMRQTTELQPLHGEES-----RADIAPM-QKRRKRDTHFPICFCCGCHRSKCGMCCKT	58%
蝙蝠	MAMNROIQATCLLLLLL-TSLTSAVLSROTTQLADLOTQDTAGAAA--GATAGLT-PGLQRLRRRDTHFPICFCCGCHRSKCGMCCKT	66%
青蛙	MKPVPICCLLLLLSFICHRGHSASLSGNEVTVTGNOIPETQMEESNAL-E-----PLL-----RSKR-QSHLSICVHCCNCKYKGCCKCLT	27%
斑马鱼	MKLSNVFLAAVILTCVCFQITAVPFIQVQVDEHHVESEELQENQHLTEAHERLTDPLVLFRTKQSHLSLRCFCKCCRNGKCGYCKCF	30%

图 9 不同物种的铁调素氨基酸序列比对图

Fig.9 The sequence of amino acid sequences of ferritin in different species

Note: The underlined part is the signal peptide region of each species, and the black font is cysteine.

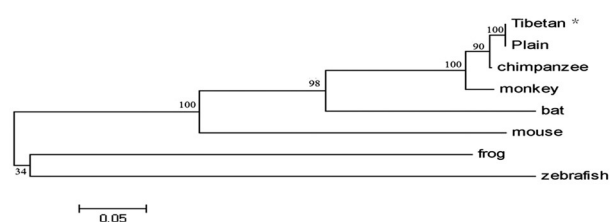


图 10 铁调素氨基酸序列系统进化树

Fig.10 The phylogenetic tree of the amino acid sequence system

3 讨论

贮存在 DNA 上的遗传信息通过 mRNA 传递到蛋白质上, mRNA 与蛋白质之间的联系是通过遗传密码的破译来实现的^[16]。mRNA 上每 3 个核苷酸翻译成蛋白质多肽链上的一个氨基酸,这 3 个核苷酸就称为密码子,也叫三联密码子^[17]。编码同一个氨基酸的密码子称为同义密码子。研究表明基因在同义密码子的使用上,普遍存在同义密码子非均衡使用的现象,即在同义密码子使用上优先使用其中的一些密码子,也叫最优密码子,这种现象就叫做密码子偏性^[18]。而密码子频率的偏性也是生物体中编码同一种氨基酸的同义密码子的非均衡使用形成的^[19]。在长期的进化过程中,每个物种都形成了自己的密码子

选择。人们对不同物种的密码子使用偏性已做了大量的研究,发现不同物种之间或同一物种不同基因家族之间的密码子使用特点可能存在差异,这种差异有助于了解物种间或者基因家族间的基因进化规律,进而了解转录和翻译过程中的调控机制,从而预测基因表达的关系^[20-21]。且密码子的使用除了受到自然选择压力的影响和突变或漂变的影响外,还有很多其他方面的因素影响,比如受物种本身碱基的组成、物种的全长基因长度、基因的表达情况、表达蛋白的二级结构、空间结构或者亲疏水性等等^[22-26]。因此,对物种进行密码子偏好性的分析对了解基因在外源表达系统的表达具有非常重要的意义。

在我们的研究中发现所获得的西藏高原人群铁调素基因 hamp 偏好使用以 C 结尾的密码子,对 19 种密码子具有偏好性,ENC 值为 41.94,与其他 7 个物种的偏好性(40.89~50.96)相接近。且西藏高原人群铁调素基因 hamp 避免使用密码子 UUU、UUA、UUG、CUA、AUA、GUC、GUA、GUG、UCG、CCU、CCG、ACU、ACA、GCG、UAC、AAU、AAC、AAA、GAU、GAA、CGU、CGC、CGG、GGU。在与 7 个物种的铁调素基因 hamp 密码子使用频率比值中发现 ≥ 2 或 ≤ 0.5 的种类为 0~19 种,表明西藏高原人群的铁调素基因 hamp 密码子偏好性与其他物种的铁调素基因 hamp 偏好性均有不同程度的差异。这可能与不

同物种之间基因的表达水平、蛋白质得结构和功能都不同等因素有关^[27,28]。进而我们对不同物种的铁调素基因结构进行了对比分析。

结果发现,在本研究中斑马鱼与蝙蝠的铁调素基因 hamp 的密码子偏好程度普遍高于其他 6 个物种,针对此现象的出现我们对比了不同物种间的铁调素基因 hamp 的基因长度后发现,斑马鱼的铁调素基因 hamp 的基因长度普遍高于其他几个物种的铁调素基因 hamp 的基因长度,由此,我们推测物种间密码子偏好性使用程度可能与基因长度有关,基因长度越长的物种,密码子使用的偏好程度也会越大。另外,铁调素基因 hamp 的 ENC-GC3s 分布图结果显示各物种的铁调素基因 hamp 的密码子偏好性受突变偏好的影响较小,而受自然选择的影响较大,这与一研究报道中表述的一般高表达基因受到的自然选择压力更大的研究结果一致^[29],有报道还表明自然选择会让机体偏爱使用某一特性的密码子来提高翻译速率^[30]。在对西藏高原人群铁调素基因 hamp 的进行结构分析中我们发现,铁调素的前体肽主要为 α -螺旋结构,这种蛋白结构可能会影响蛋白表达水平及翻译速率,也可能导致密码子偏好性对自然选择的影响。同时,铁调素基因 hamp 有一信号肽,由信号肽预测出的成熟肽为碱性蛋白,且为疏水性蛋白,所含碱性氨基酸比例高于酸性氨基酸。由此,我们还推测物种间密码子偏好性的使用程度也可能与基因所表达的蛋白属性有关。

各物种的铁调素基因 hamp 的密码子 RSCU 值的聚类分析也表明了各物种间的铁调素基因 hamp 的密码子偏好性均有不同程度的差异,且各物种间的铁调素基因 hamp 的密码子的聚类分析结果与各物种的铁调素前体肽的进化树结果保持一致,均显示出高原人群与平原人群、黑猩猩、恒河猴聚类在同一分支,表明物种间的密码子分析结果在一定程度上是能反映物种间的进化树关系。在对不同物种间铁调素基因 hamp 的氨基酸序列进行对比分析,结果也能发现高原人群与平原人群、黑猩猩和恒河猴的氨基酸序列相似性较高。表明氨基酸序列的相似性能反映各物种间密码子偏好性的差异,推测密码子偏好性的使用程度还可能与氨基酸结构有关。

虽然在我们的研究中发现了高原人群与其他物种之间密码子偏好性均存在不同程度的差异,而未发现高原人群与平原人群之间密码子偏好性有明显差异,但我们发现西藏高原人群铁调素基因全长序列与数据库中公布的人群铁调素基因信息比对结果显示有 7 个碱基有差异,这些差异的碱基可能与高原人群与平原人群基因表达差异有关,这可能是与这些差异碱基造成了西藏高原人群体内铁调素基因 hamp 的单核苷酸多态性(SNP)有关,在以后的研究中我们将进一步探索。

综上所述,此研究者通过生物信息学及密码子偏好性对铁调素基因 hamp 进行了研究,为以后进一步研究西藏高原人群铁调素基因 hamp 的表达系统及密码子使用模式提供理论依据。

参考文献(References)

- [1] Liu J, Sun B, Yin H, et al. Hecpudin: A Promising Therapeutic Target for Iron Disorders: A Systematic Review [J]. *Medicine* (Baltimore), 2016, 95(14): e3150
- [2] den Elzen W P, de Craen A J, Wiegerinck E T, et al. Plasma hepcidin levels and anemia in old age. The Leiden 85-Plus Study[J]. *Haematologica*, 2013, 98(3): 448-454
- [3] Dostalíkova-Cimburova M, Kratka K, Balusikova K, et al. Duodenal expression of iron transport molecules in patients with hereditary hemochromatosis or iron deficiency[J]. *J Cell Mol Med*, 2012, 16(8): 1816-1826
- [4] Tesfay L, Clausen K A, Kim J W, et al. Hecpudin regulation in prostate and its disruption in prostate cancer [J]. *Cancer Res*, 2015, 75(11): 2254-2263
- [5] Daher R, Manceau H, Karim Z. Iron metabolism and the role of the iron-regulating hormone hepcidin in health and disease [J]. *Presse Med*, 2017
- [6] Agoro R, Benmerzoug S, Rose S, et al. An Iron-Rich Diet Decreases the Mycobacterial Burden and Correlates With Hecpudin Upregulation, Lower Levels of Proinflammatory Mediators, and Increased T-Cell Recruitment in a Model of Mycobacterium bovis Bacille Calmette-Guerin Infection[J]. *J Infect Dis*, 2017, 216(7): 907-918
- [7] Wang K, Wu J, Cheng F, et al. Acidic Polysaccharide from Angelica sinensis Reverses Anemia of Chronic Disease Involving the Suppression of Inflammatory Hecpudin and NF-kappaB Activation [J]. *Oxid Med Cell Longev*, 2017, 2017: 7601592
- [8] Chen Y, Jiang C, Luo Y, et al. Interaction of CARD14, SENP1 and VEGFA polymorphisms on susceptibility to high altitude polycythemia in the Han Chinese population at the Qinghai-Tibetan Plateau[J]. *Blood Cells Mol Dis*, 2016, 57: 13-22
- [9] Omena J, Cople-Rodrigues C, Cardoso J, et al. Serum Hecpudin Concentration in Individuals with Sick Cell Anemia: Basis for the Dietary Recommendation of Iron[J]. *Nutrients*, 2018, 10(4)[Epub ahead of print]
- [10] Choudhury M N, Uddin A, Chakraborty S. Gene expression, nucleotide composition and codon usage bias of genes associated with human Y chromosome[J]. *Genetica*, 2017, 145(3): 295-305
- [11] Chen Y, Li X, Chi X, et al. Comprehensive analysis of the codon usage patterns in the envelope glycoprotein E2 gene of the classical swine fever virus[J]. *PLoS One*, 2017, 12(9): e183646
- [12] Rodriguez A, Wright G, Emrich S, et al. %MinMax: A versatile tool for calculating and comparing synonymous codon usage and its impact on protein folding[J]. *Protein Sci*, 2017[Epub ahead of print]
- [13] Schaefer B, Moriishi K, Behrends S. Insights into the mechanism of isoenzyme-specific signal peptide peptidase-mediated translocation of heme oxygenase[J]. *PLoS One*, 2017, 12(11): e183344
- [14] Stasiak I M, Smith D A, Crawshaw G J, et al. Characterization of the hepcidin gene in eight species of bats [J]. *Res Vet Sci*, 2014, 96(1): 111-117
- [15] Bera B C, Virmani N, Kumar N, et al. Genetic and codon usage bias analyses of polymerase genes of equine influenza virus and its relation to evolution[J]. *BMC Genomics*, 2017, 18(1): 652
- [16] Seligmann H. Bijective codon transformations show genetic code symmetries centered on cytosine's coding properties [J]. *Theory Biosci*, 2017
- [17] Chevance F, Hughes K T. Case for the genetic code as a triplet of triplets[J]. *Proc Natl Acad Sci U S A*, 2017, 114(18): 4745-4750
- [18] Gajbhiye S, Patra P K, Yadav M K. New insights into the factors affecting synonymous codon usage in human infecting Plasmodium species[J]. *Acta Trop*, 2017, 176: 29-33

- [21] 王桂梅. 蓝光联合茵栀黄颗粒对新生儿病理性黄疸治疗效果及不良反应的影响[J]. 中国医药导刊, 2015, 17(09): 922-923+925
Wang Gui-mei. Effect of treatment and adverse reaction on neonatal pathological jaundice of blue light and Yinzhihuang granules[J]. Chinese Journal of Medicinal Guide, 2015, 17(09): 922-923+925
- [22] 李婕, 胡燕. 蓝光联合茵栀黄颗粒治疗新生儿病理性黄疸的临床研究[J]. 中国临床药理学杂志, 2015, 31(18): 1809-1811
Li Jie, Hu Yan. Blue light combined with Yinzhihuang granules in the treatment of neonatal pathological jaundice clinical research Chinese [J]. Journal of Clinical Pharmacology, 2015, 31(18): 1809-1811
- [23] 李婕, 胡燕. 蓝光联合茵栀黄颗粒治疗新生儿病理性黄疸的临床研究[J]. 中国临床药理学杂志, 2015, 31(18): 1809-1811
Li Jie, Hu Yan. Clinical study of blue light combined with Yinzhihuang granules in the treatment of neonatal pathological jaundice[J]. Chin J Clin Pharmacol, 2015, 31(18): 1809-1811
- [24] 钟晓波. 茵栀黄颗粒联合间隙蓝光照射佐以双歧杆菌三联活菌治疗新生儿黄疸临床分析[J]. 中药材, 2015, 38(03): 643-644
Zhong Xiao-bo. Clinical analysis of Yinzhihuang granules combined with blue light and bifidobacterium triad in the treatment of neonatal jaundice [J]. Journal of Chinese Medicinal Materials, 2015, 38 (03): 643-644
- [25] 林艾. 茵栀黄颗粒、蓝光照射联合双歧三联活菌胶囊治疗新生儿黄疸的机制研究[J]. 海南医学院学报, 2017, 23(03): 397-400
Lin Ai. Study on the mechanism of Yinzhihuang granules, blue light irradiation combined with bifidobacteria in the treatment of neonatal jaundice [J]. Journal of Hainan Medical University, 2017, 23 (03): 397-400
- [26] 潘军平, 王水云, 温晓红. 茵栀黄颗粒联合人免疫球蛋白治疗新生儿 ABO 溶血性黄疸的疗效观察 [J]. 现代药物与临床, 2017, 32(05): 871-874
Pan Jun-ping, Wang Shui-yun, Wen Xiao-hong. To observe the curative effect of Yinzhihuang Granules Combined with Human Immunoglobulin treatment of neonatal ABO hemolytic jaundice [J]. Drugs & Clinic, 2017, 32(05): 871-874
- [27] 马晓丽, 李莲英. 茵栀黄颗粒联合蓝光照射辨治湿热熏蒸型新生儿黄疸的临床研究[J]. 辽宁中医杂志, 2016, 43(02): 328-330
Ma Xiao-li, Li Lian-ying. Clinical study of Yinzhihuang granules combined with blue light irradiation treatment of damp heat type fumigation of neonatal jaundice [J]. Liaoning Journal of Chinese medicine, 2016, 43(02): 328-330
- [28] 谭桢, 刘爱明, 罗敏, 等. 茵栀黄注射液抗胆汁淤积药效成分的筛选及其作用机制研究[J]. 中国中药杂志, 2016, 41(6): 1113-1118
Tan Zhen, Liu Ai-ming, Luo Min, et al. Yinzhihuang Injection anti cholestatic drug ingredient of [J]. China Journal of Chinese Medicine Medics, 2016, 41(6): 1113-1118
- [29] 吴海滨, 余世锋, 兰绍阳. 基于 FXR 探讨茵栀黄注射液利胆退黄的机制研究[J]. 辽宁中医杂志, 2016, 43(4): 845-848
Wu Hai-bin, She Shi-feng, Lan Shao-yang. To investigate the effect of removing Yinzhihuang injection mechanism of yellow based on FXR [J]. Liaoning Journal of Chinese medicine, 2016, 43 (4): 845-848
- [30] 李煊琼, 钟斌才. 高胆红素血症新生儿联合检测血清心肌酶谱与肌钙蛋白的临床意义[J]. 海南医学, 2015, 26(12): 1775-1777
Li Huan-qiong, Zhong Bin-cai. Clinical significance of combined detection of serum myocardial enzymes and troponin in neonates with hyperbilirubinemia[J]. Hainan medicine, 2015, 26(12): 1775-1777

(上接第 4034 页)

- [19] Yannai A, Katz S, Hersberg R. The codon usage of lowly expressed genes is subject to natural selection[J]. Genome Biol Evol, 2018[Epub ahead of print]
- [20] Song H, Liu J, Song Q, et al. Comprehensive Analysis of Codon Usage Bias in Seven Epichloe Species and Their Peramine-Coding Genes[J]. Front Microbiol, 2017, 8: 1419
- [21] Zhou Z, Dang Y, Zhou M, et al. Codon usage biases co-evolve with transcription termination machinery to suppress premature cleavage and polyadenylation[J]. Elife, 2018, 7[Epub ahead of print]
- [22] Jeacock L, Faria J, Horn D. Codon usage bias controls mRNA and protein abundance in trypanosomatids[J]. Elife, 2018, 7[Epub ahead of print]
- [23] Goffena J, Lefcort F, Zhang Y, et al. Elongator and codon bias regulate protein levels in mammalian peripheral neurons[J]. Nat Commun, 2018, 9(1): 889
- [24] Wang Y N, Ji W H, Li X R, et al. Unique features of nucleotide and codon usage patterns in mycoplasmas revealed by information entropy[J]. Biosystems, 2018, 165: 1-7
- [25] Wang Y, Jia P, Sharif R, et al. High-Level Production of DNA-Specific Endonuclease AsEndI with Synonymous Codon and its Potential Utilization for Removing DNA Contamination [J]. Appl Biochem Biotechnol, 2017
- [26] Chen Y, Xu Q, Tan C, et al. Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Senecavirus A evolution[J]. Microb Pathog, 2017, 112: 313-319
- [27] Deb B, Uddin A, Mazumder G A, et al. Analysis of codon usage pattern of mitochondrial protein-coding genes in different hookworms[J]. Mol Biochem Parasitol, 2018, 219: 24-32
- [28] Shi S L, Jiang Y R, Yang R S, et al. Codon usage in Alphabaculovirus and Betabaculovirus hosted by the same insect species is weak, selection dominated and exhibits no more similar patterns than expected[J]. Infect Genet Evol, 2016, 44: 412-417
- [29] Dilucca M, Cimini G, Giansanti A. Essentiality, conservation, evolutionary pressure and codon bias in bacterial genomes [J]. Gene, 2018 [Epub ahead of print]
- [30] Galtier N, Roux C, Rousselle M, et al. Codon Usage Bias in Animals: Disentangling the Effects of Natural Selection, Effective Population Size, and GC-Biased Gene Conversion [J]. Mol Biol Evol, 2018, 35 (5): 1092-1103