DOI:10.3969/j.issn.1005-202X.2023.11.014

医学生物物理

SARS-CoV-2与SARS病毒基因组中密码子使用的比较

王芳平[1],王志坚[1],孙咏萍[2],安战海[3]
1.天水师范学院电子信息与电气工程学院物理系，甘肃 天水 741001；2.内蒙古师范大学物理与电子信息学院，内蒙古 呼和浩特 010021；3.天水市田家炳中学，甘肃 天水 741000

【摘要】目的：分析冠状病毒基因中密码子的进化和变异。**方法：**运用生物信息学方法和生物统计学方法，对新型冠状病毒（SARS-CoV-2）和非典病毒（SARS）基因中密码子使用的偏好性和密码子的上下文关系进行比较分析。**结果：**两种病毒密码子的使用都存在显著的偏好性（RSCU值的变化为0.10~2.67），且偏好使用的密码子（偏爱密码子）和避免使用的密码子（稀有密码子）基本是一致的；通过相关性分析发现在两种病毒中密码子适应性指数（CAI）与密码子第三位点的GC含量（GC3）呈显著的负相关性，但SARS-CoV-2的负相关性更强（SARS-CoV-2：R^2=0.76，P<0.001；SARS：R^2=0.48，P<0.001），随着CAI的增大，GC3降低由于CAI是反映基因表达水平的参数，即高表达基因偏好低GC3的同义密码子。比较偏爱密码子和稀有密码子的上下文关系，发现偏爱密码子与稀有密码子各位点的核苷酸组分存在显著差异，这种差异性表现为在密码子不同位点GC和AT的显著区别，反映了GC含量对同义密码子使用偏爱性的约束。**结论：**SARS-CoV-2和SARS病毒基因组GC含量、密码子的上下文关系、密码子各位点核苷酸的组分、基因的表达水平是影响密码子偏爱性的重要因素，研究结果对SARS-CoV-2和SARS病毒基因组的进化和变异的研究具有参考意义。

【关键词】新型冠状病毒；非典病毒；二核苷酸；GC含量；密码子适应性指数
【中图分类号】R318；R373 **【文献标志码】**A **【文章编号】**1005-202X(2023)11-1402-06

Comparative analysis of codon usage in the genomes of SARS-CoV-2 and SARS virus

WANG Fangping[1], WANG Zhijian[1], SUN Yongping[2], AN Zhanhai[3]
1.Department of Physics, School of Electronic Information and Electrical Engineering, Tianshui Normal University, Tianshui 741001, China; 2. College of Physics and Electronic Information, Inner Mongolia Normal University, Hohhot 010021, China; 3. Tianshui Tianjiabing Middle School, Tianshui 741000, China

Abstract: Objective To analyze the genetic evolution and variation of coronavirus. **Methods** The codon usage bias and the codon context in the genes of SARS-CoV-2 and SARS virus were compared and analyzed using bioinformatics methods and biometric methods. **Results** The codon usage bias was observed in the two kinds of virus, with RSCU value varying from 0.10 to 2.67, and the codons preferred (preferred codons) and avoided (rare codons) were basically the same. The correlation analysis showed that the codon adaption index (CAI) was negatively correlated with the GC content at the third site of the codon (GC3) in the two viruses, and that the negative correlation in SARS-CoV-2 was higher than that in SARS (SARS-CoV-2: R^2=0.76, P<0.001; SARS: R^2=0.48, P<0.001). GC3 decreased with the increase of CAI which reflected the level of gene expression, suggesting highly expressed genes preferred the synonymous codons with low GC3. Though the analysis on the codon context between preferred codons and rare codons, it was found that there were significant differences in the nucleotide composition between preferred codons and rare codons, which were manifested as differences in GC and AT in different codon sites, reflecting the constraints of GC content on the synonymous codon usage bias. **Conclusion** GC content, codon context, nucleotide composition and gene expression level of SARS-CoV-2 and SARS virus are important factors affecting codon usage bias. The study has reference significance for the study on the genetic evolution and variation of SARS-CoV-2 and SARS virus.

Keywords: SARS-CoV-2; SARS virus; dinucleotide; GC content; codon adaption index

【收稿日期】2023-07-22
【基金项目】国家自然科学基金（11665019）；甘肃省科技计划（21JR7RE175）；甘肃省高等学校科研项目（2018B-41）；甘肃省教育科学“十四五”规划项目（GS[2023]GHB1387）
【作者简介】王芳平，博士，研究方向：生物物理，E-mail: wangfp@tsnu.edu.cn

前 言

新型冠状病毒（SARS-CoV-2）是目前已知的第7种可感染人类的冠状病毒，SARS-CoV-2与非典病毒（SARS）、中东呼吸综合征都属于Beta属冠状病毒[1-3]。2020年1月10日，SARS-CoV-2基因组序列公开发表[4]，基于病毒的基因组序列，相关的组学研究也逐渐开展起来，从基因组学研究的角度，解析其基因组结构[4]、预测蛋白的结构和功能、研究病毒序列的变异等[5]。SARS-CoV-2是具有包膜的单股正链RNA病毒[6]，结构不稳定，复制过程中相对容易发生变异，国内外科研工作者在病毒基因组的进化和变异[7-8]、疫苗和药物研发等方面进行广泛的研究[9-10]。如运用生物信息学方法，分析基因组进化[11]，推测病毒的时间进化信号[12]，基于比较基因组学研究冠状病毒位点的平均变异率、重组率[13]。

随着SARS-CoV-2的传播和不断变异，迄今为止变异的SARS-CoV-2主要更新有六大类，分别是Alpha变异毒株（编号为B.1.1.7），特性是攻击免疫；Beta变异毒株（编号为B.1.351），特性是规避疫苗；Gamma变异毒株（编号为P.1），特性是增长迅速，传染能力是原始SARS-CoV-2的两倍；Delta变异毒株（编号为B.1.617.2）和Lambda变异毒株（编号为C.37）；Omicron变异毒株（编号为B.1.1.529），最突出的特性是生长迅速，发生的变异最多，仅其表面刺突蛋白的变异就有32处[14]。Ji等[15]研究发现SARS-CoV-2基因组序列每个位点的平均变异率约为10^{-4}个，可以通过中间宿主发生变异和重组后，感染人类。在传播和感染过程中，SARS-CoV-2可能同样通过基因组变异和重组，发生适应性进化[4]。SARS-CoV-2感染和传播机制与SARS一样，都是刺突糖蛋白参与受体结合与宿主特异性识别的，有报道认为SARS-CoV-2的刺突糖蛋白是SARS与某种未知冠状病毒发生重组的结果，有助于病毒感染人类[15]。与SARS相比，SARS-CoV-2的刺突糖蛋白与人类细胞的血管紧张素转换酶II（Angiotensin Converting Enzyme II，ACE2）结合的亲和力更高[2,16]，这或许是SARS-CoV-2感染力更强的原因之一[17]。分析病毒的进化和变异是病毒防治和药物研发的首要任务，本研究基于国家生物技术信息中心（NCBI）中发表的SARS-CoV-2序列，用生物信息学方法，对病毒基因序列中密码子使用的偏好性进行分析，研究其基因的进化和变异规律，并讨论这些结果的进化意义，为病毒的分子实验研究提供理论参考，有助于SARS-CoV-2序列变异的诠释。

1 资料与方法

1.1 数据资料

本研究采用的SARS-CoV-2和SARS基因组数据取自NCBI序列库，SARS-CoV-2序列号为MN908947，含10个蛋白质编码序列；SARS序列号为DQ497008，含14个蛋白质编码序列。

1.2 分析方法

1.2.1 密码子使用偏好性的分析参数 用DAMBE程序[18]分析SARS-CoV-2基因的密码子相对使用频率（RSCU）[19]。该指标可以直观地反映同义密码子使用偏好性而被广泛应用。RSCU的定义：某同义密码子使用频数的观察值除以该密码子所编码的氨基酸的所有简并密码子平均使用时的频数。对于一个给定的氨基酸i，其第j个密码子的RSCU值计算公式为：

$$\mathrm{RSCU}_{ij}=\frac{A_{ij}}{\overline{A}}=\frac{A_{ij}}{\frac{1}{n_i}\sum_{j=1}^{n_i}A_{ij}} \quad (1)$$

其中，A_{ij}是该密码子出现次数的实际观察值，n_i是编码此氨基酸的密码子简并度，其数值为1~6。RSCU值大于1表示密码子出现频率大于期望频数[20]，把RSCU值大于1的密码子定义为偏爱密码子，把RSCU值小于1的密码子定义为稀有密码子。

1.2.2 密码子适应性指数（CAI） CAI值是用来描述基因表达水平的一个精确的参数[21]，它表示一个基因相对于高表达基因，其密码子使用的相对适应性[19]。CAI值为0~1，较高的CAI值意味着相对于高表达的参考基因集，目标基因有较高的密码子使用偏好性和较高的密码子使用模式的相似性[21]。本研究用DAMBE软件计算SARS-CoV-2不同毒株的CAI值[18]。

1.2.3 密码子中二核苷酸的相对丰度r_{xy} 对密码子（1, 2）、（2, 3）、（3, 1）3种不同位点的组合，分析每个位点二核苷酸出现的频数的偏差r_{xy}：

$$r_{xy}=\frac{O_{xy}}{O_{xy}-E_{xy}} \quad (2)$$

其中，O_{xy}、E_{xy}分别为二核苷酸xy的观察频数和期望频数。r_{xy}>0表示二核苷酸xy是偏好使用的，反之亦然。

2 结果与讨论

2.1 密码子使用的偏好性

用DAMBE软件计算SARS-CoV-2和SARS的RSCU值见表1和图1。通过分析RSCU值可以发现：RSCU值变化为0.10~2.67，可以说明两种病毒密码子

使用都存在显著的偏好性，且两种病毒密码子的相对使用频率值的分布基本是一致的（表1），表明其密码子的偏爱性是相似的。其密码子的使用有如下特点。如对SARS-CoV-2的61个有义密码子，以下括号中为SARS-CoV-2密码子的RSCU值：（1）起始密码子AUG（1.00）和密码子UGG（1.00），终止密码子偏好UAA（2.40）；（2）在RSCU>1的27个偏爱的密码子中，密码子第三位点核苷酸为U或A，而RSCU值最大的一些密码子第三位点为U［AGA（2.67）和终止密码子例外］：GGU（2.34），GCU（2.18），UCU（1.96），GUU（1.95），CCU（1.93），ACU（1.78），CUU（1.74）；SARS-CoV-2和SARS最偏爱密码子第三位点都是核苷酸U或A；（3）在RSCU<1的32个稀有密码子中，有28个第三位点是核苷酸G或C。RSCU值最小的（稀有密码子）第三位点是核苷酸G，如：UCG（0.10），GGG（0.11），CCG（0.17），CGG（0.18），ACG（0.20）。当RSCU值增大趋向于1时，第三位点C居多，如UAC（0.77），CUC（0.60），UCC（0.46），AUC（0.55）。从表1可以看出，按照RSCU值由大到小，密码子第三位点核苷酸有一个基本变化趋势从U/A，到G/C的变化趋势，即偏爱密码子第三位点为U或A，而稀有密码子第三位点为G或C。

表1 SARS-CoV-2和SARS基因中密码子的RSCU值

Table 1 RSCU values of codons in the genes of SARS-CoV-2 and SARS virus

密码子	氨基酸	SARS-CoV-2	SARS	密码子	氨基酸	SARS-CoV-2	SARS
AGA	Arg	2.67	2.09	GUA	Val	0.90	0.86
UAA	Stop	2.40	2.14	GGA	Gly	0.82	0.91
GGU	Gly	2.34	1.91	AGG	Arg	0.80	0.99
GCU	Ala	2.18	2.05	UAC	Tyr	0.77	0.88
UCU	Ser	1.96	1.92	GAC	Asp	0.71	0.75
GUU	Val	1.95	1.67	GGC	Gly	0.71	1.00
CCU	Pro	1.93	1.67	AAG	Lys	0.69	0.94
ACU	Thr	1.78	1.72	CUA	Leu	0.66	0.70
CUU	Leu	1.74	1.70	AAC	Asn	0.65	0.75
UCA	Ser	1.65	1.76	GCG	Ala	0.64	0.53
ACA	Thr	1.64	1.77	CAG	Gln	0.61	0.78
UUA	Leu	1.63	1.55	CUC	Leu	0.60	0.85
CCA	Pro	1.59	1.04	UUC	Phe	0.59	0.76
UGU	Cys	1.55	1.69	CGC	Arg	0.58	0.77
AUU	Ile	1.51	1.23	GUC	Val	0.56	0.69
AGU	Ser	1.45	1.70	GAG	Glu	0.56	0.94
CGU	Arg	1.45	1.12	GCC	Ala	0.56	0.57
GAA	Glu	1.44	1.57	AUC	Ile	0.55	0.63
UUU	Phe	1.41	1.06	UCC	Ser	0.46	0.40
UUU	Phe	1.41	1.24	UGC	Cys	0.44	0.77
CAU	His	1.39	1.24	ACC	Thr	0.37	0.54
CAA	Gln	1.39	1.29	AGC	Ser	0.35	0.54
AAU	Asn	1.38	1.21	UAG	Stop	0.30	0.43
AAA	Lys	1.30	1.25	UGA	Stop	0.30	0.43
GAU	Asp	1.28	1.06	CUG	Leu	0.29	0.59
UAU	Tyr	1.22	1.25	CCC	Pro	0.29	0.42
GCA	Ala	1.09	1.12	CGA	Arg	0.29	0.47
UUG	Leu	1.07	1.11	ACG	Thr	0.20	0.20
AUG	Met	1.00	1.05	CGG	Arg	0.18	0.11
UGG	Trp	1.00	1.00	CCG	Pro	0.17	0.18
AUA	Ile	0.92	1.00	GGG	Gly	0.11	0.17
CAC	His	0.61	0.71	UCG	Ser	0.10	0.25

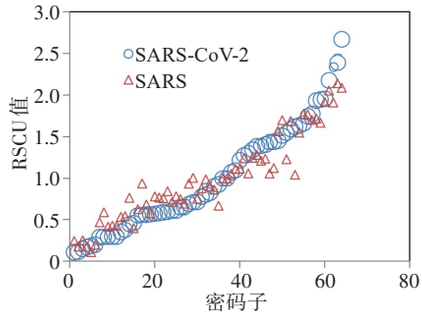

图1 SARS-CoV-2和SARS基因中64个密码子的RSCU值的分布

Figure 1 Distribution of RSCU of 64 codons in the genes of SARS-CoV-2 and SARS virus

2.2 密码子中及密码子上下文二核苷酸的相对丰度

为探讨密码子各个位点上二核苷酸频率分布规律，对SARS-CoV-2基因密码子中3种二核苷酸组合位点(1, 2)、(2, 3)、(3, 1)分别进行分析，计算各位点二核苷酸的观察频数与期望频数，并计算偏差r_{xy}，结果见表2。图2为(3, 1)位点二核苷酸的分布。其中对表2中密码子各个位点二核苷酸的观察值O_{xy}与期望值E_{xy}的分布关系作了线性拟合，拟合结果如下：(1, 2)位点：$E_{xy}=0.36O_{xy}+390$，$R^2=0.35$，$P<0.001$；(2, 3)位点：$E_{xy}=0.75O_{xy}+153$，$R^2=0.70$，$P<0.001$；(3, 1)位点：$E_{xy}=0.35O_{xy}+393$，$R^2=0.33$，$P<0.001$。

表2 密码子中及两个密码子相邻位点的二核苷酸分布

Table 2 Dinucleotide distribution in codons and adjacent sites of two codons

二核苷酸	(1, 2)位点			(2, 3)位点			(3, 1)位点		
	O_{xy}	E_{xy}	r_{xy}	O_{xy}	E_{xy}	r_{xy}	O_{xy}	E_{xy}	r_{xy}
AA	1 093	918	0.19	944	861	0.09	739	824	-0.10
AC	717	588	0.22	587	557	0.05	661	453	0.46
AG	401	690	-0.42	430	411	0.05	867	611	0.43
AU	713	790	-0.10	1 087	1 063	0.02	459	941	-0.51
CA	551	588	-0.06	813	557	0.46	671	453	0.48
CC	394	377	0.05	241	360	-0.33	229	249	-0.08
CG	147	442	-0.67	90	266	-0.66	174	335	-0.48
CU	506	506	0.00	1 084	687	0.57	437	517	-0.15
GA	948	690	0.37	293	410	-0.29	325	610	-0.47
GC	656	442	0.48	242	266	-0.09	236	335	-0.29
GG	576	519	0.11	185	196	-0.06	309	452	-0.31
GU	780	593	0.31	809	507	0.60	360	697	-0.48
UA	456	790	-0.42	685	1 063	-0.36	1 179	942	0.25
UC	461	506	-0.09	441	687	-0.36	472	517	-0.08
UG	405	593	-0.32	526	507	0.03	1 610	697	1.31
UU	907	678	0.34	1 254	1 312	-0.04	973	1 075	-0.09

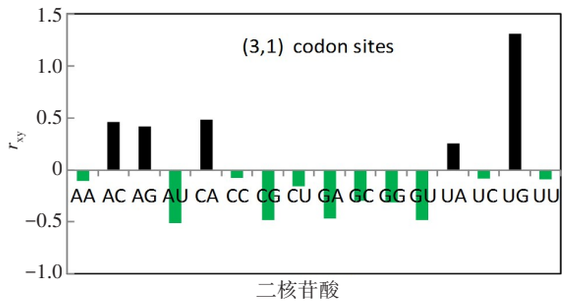

图2 密码子中(3, 1)位点二核苷酸相对丰度的分布

Figure 2 Distribution of the relative abundance of dinucleotide in (3, 1) codon sites

从拟合结果可以看出：(1)在密码子的3个二核苷酸位点中，位点(2, 3)相关性最强(决定系数最大，$R^2=0.70$)，即密码子的二核苷酸的期望值E_{xy}与观察值O_{xy}之间的线性相关性强，暗示着密码子对(2, 3)位点二核苷酸的偏好性相对弱；而对密码子(1, 2)和(3, 1)位点的期望值E_{xy}与观察值O_{xy}之间的线性相关性相对较小($R^2=0.35$和$R^2=0.33$)，核苷酸组分的偏好性明显高于(2, 3)位点。(2)在密码子位点(3, 1)，r值最大的二核苷酸是UG、AC、AG、CA，这些组合都含有A或U；反之，r值最小的，GU、GC、GA和GG，里面都含有G，可见在

密码子(3, 1)位点二核苷酸使用受AT或GC约束,特别是当第三位点为G时的4种组合(GA、GC、GG、GU)都是避免的,推测在(3, 1)位点,这4种二核苷酸组合的频率由上游密码子中的第三位点碱基类型决定。

密码子中二核苷酸的频率分布可以理解为密码子第二位点的碱基主要决定基因组的氨基酸,因此导致第二位点组合(2, 3)二核苷酸分布的保守,而第三位点决定同义密码子的选择,对同义密码子的偏爱性,显示出了对二核苷酸的偏爱。这是一个值得注意的密码子二核苷酸分布特点。这种表现的内在因素是值得进一步探讨的,首先,突变的压力或许对密码子碱基组分有重要的影响,如SARS-CoV-2的刺突糖蛋白基因序列发生插入突变,使该病毒的刺突糖蛋白与已报道的其他Beta冠状病毒的刺突糖蛋白序列间相似性较低[1, 22]。其次,基因氨基酸和肽链保守性可能是另一个约束密码子核苷酸组分的因素。最后,基因表达水平的要求对同义密码子的选择也影响密码子核苷酸组分[23],由于全基因组GC含量较低,密码子各位点的GC含量也是约束密码子选择的一个重要因素[24-25]。

2.3 基因组密码子第三位点的GC含量(GC3)与CAI的相关性分析

SARS-CoV-2病毒和SARS病毒编码序列的CAI与GC3含量结果见表3和表4。可以看出,两种病毒基因的GC和GC3含量都较低,而基因表达水平相对较高,为了分析这种相关性,对病毒基因密码子的GC3和CAI做了线性拟合,拟合结果如下:SARS-CoV-2:y=-0.65x+0.83,R^2=0.76,P<0.001;SARS:y=-0.31x+0.72,R^2=0.48,P<0.001(图3)。结果显示两种病毒基因的CAI和GC3都呈显著的负相关性,即随着GC3的增大,基因表达水平降低,即高表达基因更偏好使用低GC3的同义密码子。GC3含量与基因表达水平之间的这种关联,反映出基因表达水平影响同义密码子的选择,而密码子GC3含量是影响同义密码子使用的一个内在因素。两种病毒表现出了共同的特征,值得注意的是SARS-CoV-2的CAI和GC3负相关性更强。这或许是冠状病毒进化过程自然选择的约束,根据SARS-CoV-2和SARS基因的GC3和CAI之间的相关性,推测基因表达的调控是影响密码子使用进化的一个因素。

3 结 论

随着密码子RSCU值从最大增加到最小,密码子第三位点核苷酸从U到A,再到G、C的变化(U~A~G~C),可以看出,在这两种病毒基因中,密码子的单核苷酸频率分布按照密码子位点呈一定的结构,这个结果显示了偏爱密码子与稀有密码子中核苷酸的使用有显著区别,偏爱密码子第三位点为U或A,而稀有密码子第三位点是G或C。结果表明,密码子第三位点的核苷酸组分与密码子使用偏爱性有相关性。而进一步比较发现,密码子第三位点核苷酸的频数差异主要表现在GC含量的约束,推测密码子第三位点核苷酸是决定密码子偏爱性的重要因素。

有研究显示在细菌中,GC3的结构在原核生物中有一个共同的特征,基因的碱基含量在染色体上的位置呈一定的结构化[26],本研究通过分析SARS-CoV-2同义密码子使用的偏性,密码子GC3和CAI之间的关联,得到以下结论:密码子的偏好性和密码子不同位点的

表3 COVID-19基因名称和长度、基因序列的GC含量、GC3和CAI值
Table 3 Gene name and length, GC content of gene sequence, GC3 and codon adaptive index in SARS-CoV-2

基因名称	GC/%	序列长度	GC3/%	CAI
QHN73809.1	0.37	21 291	0.27	0.67
QHN73810.1	0.37	3 822	0.26	0.67
QHN73811.1	0.39	828	0.35	0.62
QHN73812.1	0.38	228	0.33	0.60
QHN73813.1	0.42	669	0.39	0.57
QHN73814.1	0.27	186	0.29	0.60
QHN73815.1	0.38	366	0.39	0.61
QHN73816.1	0.36	366	0.26	0.70
QHN73817.1	0.47	1 260	0.53	0.50

表4 SARS基因名称和长度、基因序列的GC含量、GC3和CAI值
Table 4 Gene name and length, GC content of gene sequence, GC3 and codon adaptive index in SARS

基因名称	GC/%	序列长度	GC3/%	CAI
AAP41036.1	0.41	21 222	0.60	0.65
AAP41037.1	0.38	3 768	0.56	0.66
AAP41038.1	0.40	825	0.33	0.63
AAP41039.1	0.40	465	0.46	0.59
AAP41040.1	0.40	231	0.39	0.61
AAP41041.1	0.45	666	0.31	0.57
AAP41042.1	0.32	192	0.48	0.54
AAP41043.1	0.40	369	0.28	0.59
AAP41044.1	0.32	135	0.39	0.63
AAP41045.1	0.38	120	0.38	0.64
AAP41046.1	0.40	255	0.37	0.60
AAP41047.1	0.48	1 269	0.30	0.58
AAP41048.1	0.52	297	0.34	0.50
AAP41049.1	0.54	213	0.30	0.54

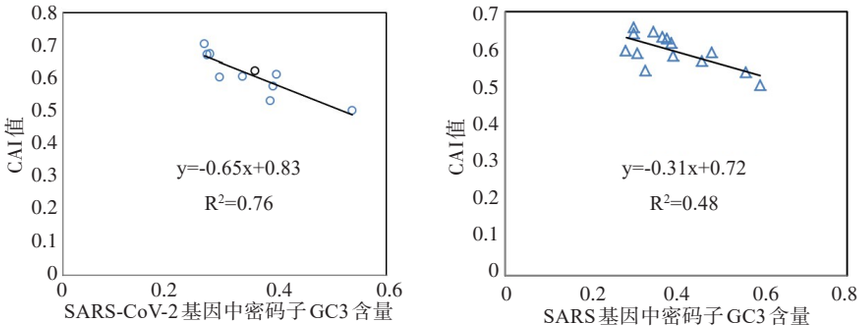

图3 密码子GC3和CAI的关系

Figure 3 Relationship between the GC3 of codons and CAI

核苷酸使用有关，在密码子同一位点，偏好密码子和稀有密码子偏爱的单核苷酸和二核苷酸模式是截然不同的，这个结果表现为同一位点对GC或AT的偏好与避免，揭示GC含量的约束。其次，密码子的GC3和CAI之间呈显著的负相关性，可以推测，基因表达水平影响着同义密码子的使用，早在1981年，Ikemura[27]发现在大肠杆菌中基因表达水平与同义密码子使用存在关联，这里SARS-CoV-2的密码子GC3和CAI之间的负相关性说明，作为RNA病毒的SARS-CoV-2，其密码子使用与细菌在某些方面也存在相似的机理。本研究认为SARS-CoV-2基因中密码子偏好性的压力，表现为密码子中的不同位点对单核苷酸和二核苷酸模式的选择，这与作用在SARS-CoV-2密码子上的核苷酸组成结构的约束是不可忽视的[11]，病毒基因使用密码子的这种结构也可能与基因组重新排列或进化速率的增加有关[28]。本研究从密码子使用的角度研究SARS-CoV-2的基因序列，为病毒的序列进化和变异提供理论参考，有助于SARS-CoV-2变异的诠释。

【参考文献】

[1] Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin[J]. Nature, 2020, 579 (7798): 270-273.

[2] Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding[J]. Lancet, 2020, 395(10224): 565-574.

[3] Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019[J]. N Engl J Med, 2020, 382(8): 727-733.

[4] Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China[J]. Nature, 2020, 579(7798): 265-268.

[5] Zhang J, Wu Y, Wang R, et al. Bioinformatic analysis reveals that the reproductive system is potentially at risk from SARS-CoV-2[J]. Ann Transl Med, 2021, 9(8): 678.

[6] Xu D, Zhang H, Gong HY, et al. Identification of a potential mechanism of acute kidney injury during the COVID-19 outbreak: a study based on single-cell tranome analysis[J]. Intensive Care Med, 2020, 46(6): 1114-1116.

[7] Chan JF, Kok KH, Zhu Z, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan[J]. Emerg Microbes Infect, 2020, 9(1): 221-236.

[8] Dimonte S, Muhammed BM, Taib HS, et al. Genetic variation and evolution ofthe 2019 novel coronavirus [J]. Public Health Genom, 2021, 24(1-2): 54-66.

[9] Ceraolo C, Giorgi FM. Genomic variance of the 2019-nCoV coronavirus[J]. J Med Virol, 2020, 92(5): 522-528.

[10] Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for COVID-19 and other lineage B betacoronaviruses[J]. Nat Microbiol, 2020, 5(4): 562-569.

[11] Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation [J]. Science, 2020, 367(6483): 1260-1263.

[12] 周烨真, 张世豪, 陈嘉仪, 等. 新型冠状病毒COVID-19的变异和进化分析[J]. 南方医科大学学报, 2020, 40(2): 152-158.
Zhou YZ, Zhang SH, Chen JY, et al. Analysis of variation and evolution of COVID-19 genome[J]. Journal of Southern Medical University, 2020, 40(2): 152-158.

[13] Su S, Wong G, Shi W, et al. Epidemiology, geneticrecombination, and pathogenesis of coronaviruses[J]. Trends Microbiol, 2016, 24(6): 490-502.

[14] Ji W, Wang W, Zhao X, et al. Cross-species transmission of the newly identified coronavirus 2019-nCoV[J]. J Med Virol, 2020, 92(4): 433-440.

[15] Ji W, Li X. Response to comments on "Cross-species transmission of the newly identified coronavirus 2019-nCoV" and "Codon bias analysis may be insufficient for identifying host(s) of a novel virus" [J]. J Med Virol, 2020, 92(9): 1440.

[16] Chen ZM, Fu JF, Shu Q, et al. Diagnosis and treatment recommendations for pediatric respiratory infection caused by the 2019 novel coronavirus[J]. World J Pediatr, 2020, 16(3): 240-246.

[17] Miyairi I, Ziebarth J, Laxton JD, et al. Host genetics and Chlamydia disease: prediction and validation of disease severity mechanisms[J]. PLoS One, 2012, 7(3): e33781.

[18] Xia X. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution[J]. Mol Biol Evol, 2013, 30(7): 1720-1728.

[19] Sharp PM, Li WH. The codon adaptation index: a measure of directional synonymous codon usage bias, and its potential applications [J]. Nucleic Acids Res, 1987, 15(3): 1281-1295.

[20] Dos RM, Wernisch L, Savva R. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome[J]. Nucleic Acids Res, 2003, 31(23): 6976-6985.

[21] Roy SW. The Plasmodium gaboni genome illuminates allelic dimorphism of immunologically important surface antigens in P. falciparum[J]. Infect Genet Evol, 2015, 36(2): 441-449.

[22] Habibzadeh P, Stoneman EK. The novel coronavirus: a bird's eye view[J]. Int J Occup Environ Med, 2020,11(2): 65-71.

[23] Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms[J]. Mol Biol Evol, 1985, 2(1): 13-35.

[24] 王芳平, 王志坚, 李永香. 蝇基因组中内含子数目随其长度的分布研究[J]. 基因组学与应用生物学, 2020, 39(3): 1062-1066.
Wang FP, Wang ZJ, Li YX. Research on the distribution of intron number with its length in drosophila melanogaster genome [J]. Genomics and Applied Biology, 2020, 39(3): 1062-1066.

[25] 王芳平, 王志坚, 李永香. 三种模式生物基因组中GC含量的比较[J]. 基因组学与应用生物学, 2019, 38(5): 2215-2220.
Wang FP, Wang ZJ, Li YX. Comparison of GC contents in genomes of the three model microbes[J]. Genomics and Applied Biology, 2019, 38(5): 2215-2220.

[26] Daubin V, Perriere G. G+C3 structuring along the genome: a common feature in prokaryotes[J]. Mol Biol Evol, 2003, 20(4): 471-483.

[27] Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes[J]. J Mol Biol, 1981, 146(1): 1-21.

[28] Benvenuto D, Giovanetti M, Ciccozzi A, et al. The 2019 - new coronavirus epidemic: evidence for virus evolution[J]. J Med Virol, 2020, 92(4): 455-459.

（编辑：陈丽霞）