

DOI:10.13350/j.cjpb.220207

• 论著 •

尼帕病毒密码子偏爱性分析*

何振, 尚方建, 王聪, 刘奇, 孟余**

(大理大学基础医学院病原生物学综合实验室, 云南大理 671000)

【摘要】 **目的** 了解尼帕病毒(Nipah Virus, NiV)主要蛋白密码子使用偏爱性及其影响因素,寻找最佳病毒蛋白表达系统。**方法** 以人 NiV 的 352 条编码序列为材料,使用 EMBOSS 的子程序 CAI 和 CUSP 以及 CodonW、SPSS22.0 软件统计分析 NiV 6 个蛋白基因组密码子偏爱性,再用 SigmaPlot14.0 进行绘图。**结果** NiV 各蛋白基因组 ENC 值分布在 50.34~55.87 之间,CAI 值为 0.678(0.606~0.730)。RSCU>1 的密码子中以 A/U 结尾的约占 43.3%。AGA、AGG、ACA、CUU、GUU、UCA 为 6 个蛋白的多基因共有高频密码子。中性绘图分析、ENC-Plot 分析和 PR2 奇偶分析表明自然选择是影响 NiV 密码子偏爱性的主要因素,同时突变因素也具有一定的作用。将 NiV 基因密码子使用频率与大肠埃希菌、酵母、人和昆虫杆状病毒表达系统进行比较,显示 NiV 与酵母密码子使用频率相差最小。**结论** NiV 密码子在使用过程中偏爱 A/U 结尾,其偏爱性主要受自然选择作用的影响,酵母是最适于其蛋白表达的系统。

【关键词】 尼帕病毒;密码子偏爱性;自然选择

【中图分类号】 R373

【文献标识码】 A

【文章编号】 1673-5234(2022)02-0159-06

[*Journal of Pathogen Biology*. 2022 Feb;17(2):159-164, 169.]

Analysis of Nipah Virus codon usage bias

HE Zhen, SHANG Fang-jian, WANG Cong, LIU Qi, MENG Yu (*Integrated Lab of Pathology Biology, College of Basic Medical, Dali University, Dali 671000, Yunnan, China*)

【Abstract】 **Objective** To analysis codon usage preference and its influencing factors of Nipah Virus (Nipah Virus, NiV) main protein, and find the best expression system. **Methods** Use the EMBOSS programs CAI and CUSP, CodonW, SPSS22.0 statistically analyze the codon preference of 352 NiV coding sequences of human NiV genome. SigmaPlot 14.0 was used for mapping. Using the Codon Usage Database to compare and analyze the difference between the NiV genome and other expression systems. **Results** The average ENC values of NiV proteins ranged from 50.34 to 55.87, and the CAI value was 0.678(0.606 to 0.730), the GC3s value was 0.369(0.292~0.530). The codons with RSCU>1 and the codon ending in A/U accounted for about 43.3%. AGA, AGG, ACA, CUU, GUU, UCA are the multi-genes of six proteins that share high-frequency codons. Neutral analysis show that the slopes for the six proteins were -0.523, -0.256, -0.182, -0.277, 0.020, -0.249. PR2 parity analysis show 62.2% (331/532) probability of $y>0.5$ and 60.5% (332/532) probability of $x>0.5$. Neutral analysis, ENC-Plot analysis and PR2 parity analysis show that natural selection is the main factor affecting NiV codon preference, and mutation factors also have impact. Comparing the usage frequency of NiV gene codon with *E. coli*, yeast, human and insect baculovirus expression systems, it is shown that the difference between NiV and yeast codon usage frequency is the smallest. **Conclusion** The NiV codon prefers A/U endings, which mainly affected by natural selection. Yeast expression system will be the most suitable for expression of its protein.

【Key words】 Nipah virus; codon preference; natural selection

***尼帕病毒(Nipah virus, NiV)是一种高致病性的人畜共患病病毒,属于包膜 RNA 病毒,与亨德拉病毒(Hendra virus, HeV)密切相关,同属于副粘病毒科的亨德拉尼帕属(*Henipavirus*)^[1]。NiV 会在动物和人之间传播,同时也可以通过受污染的食物或直接在人与人之间传播。蝙蝠是 NiV 的主要宿主,同时 NiV 在猪群中也具有高度的传播特征^[2]。人感染 NiV 后会产一系列严重的疾病,包括从无症状状态感染(亚临床)到严重的急性呼吸道疾病和脑部炎症^[3]。NiV 于 1998-1999 年在马来西亚首次被报道^[4]。随后在许多亚洲国家均有爆发,包括印度尼西亚、印度、柬埔寨、巴

布亚新几内亚、孟加拉国、东帝汶、越南和泰国^[5-7]。我国南部毗邻东南亚国家等地区,虽然国内目前尚未有该病报道,但存在一定感染输入风险^[8]。对于 NiV 引起的感染,目前尚无有效的治疗方法,主要通过化学性药物利巴韦林(ribavirin)和法匹那韦(favipiravir)抑制

* **【基金项目】** 大理大学高层次人才启动费项目(No. KY2096107440)。

** **【通讯作者】** 孟余, E-mail: mengyu@dali.edu.cn

【作者简介】 何振(1997-),男,湖北孝感人,在读研究生。主要研究方向:感染与免疫。E-mail: 2273330137@qq.com

NiV 的复制^[9]。

密码子的使用是 DNA 序列的重要信息,密码子编码蛋白质具有简并性的特点,同一个氨基酸具有两个或多个密码子^[10]。不同的氨基酸所对应的密码子使用频率也不相同,这种现象被称为密码子的使用偏爱性(Codon usage bias, CUB)^[11]。分析这些基因密码子的使用模式,有助于理解 NiV 在体内的复制,逃避免疫反应并引起各种疾病的机制^[12]。NiV 基因组大约由 18.2 kb 的负链单链 RNA 组成,连续排列的 6 个基因编码 6 个蛋白。即核衣壳蛋白(nucleoprotein, N),基质蛋白(matrix protein, M),融合糖蛋白(fusion protein, F),附着糖蛋白(attachment glycoprotein, G)和 RNA 聚合酶蛋白或长蛋白(long polymerase, L),磷蛋白(phosphoprotein, P)^[13]。其中 P 基因组又可根据 RNA 编辑或开放阅读框编码 V/W/C 3 种蛋白。

本实验以 NiV 的基因编码序列为材料,综合分析 NiV 基因组密码子使用偏爱性及其影响因素,以便更好地了解 NiV 基因组密码子使用模式,为研究 NiV 的基因与蛋白的表达以及亚单位疫苗的研制提供依据。

材料与方法

1 材料

1.1 目的基因序列 通过 NCBI 的 GenBank 数据库(<https://www.ncbi.nlm.nih.gov/nuccore>)获取 352 条 NiV 基因编码序列(CDS)。

1.2 分析软件和程序 使用 EMBOSS(<https://www.bioinformatics.nl/emboss-explorer/>)中的 CAI (codon adaptation index), CUSP(Create a codon usage table)子程序和 CondonW 对基因组密码子进行偏爱性计算,再用 SPSS Statistics 22.0、SigmaPlot14.0 对上述数据进行统计分析并做图。

2 分析指标

2.1 密码子适应指数 密码子适应指数(codon adaptation index, CAI)是指编码同义密码子与最佳密码子之间的适应性水平,CAI 取值范围位于 0~1 之间^[10]。该值越接近于 1 表示适应性越强,反之适应性越弱^[14]。

2.2 有效密码子数 有效密码子数(Effective Number of Codon, ENC)被广泛用于评价密码子家族中同义密码子偏爱性的高低,以及测量 CUB 的偏爱程度。ENC 值的范围从 20(最高偏差)到 61(最低或无偏差)^[15]。即 ENC 的值为 20 时,属于极度偏爱,当 ENC 的值为 60 时,表明偏爱性较低^[16]。

2.3 相对同义密码子使用度 相对同义密码子使用度(Relative synonymous codon usage, RSCU)是指一个特定密码子的使用频率与相同氨基酸的所有预期同

义密码子的使用频率之比^[14]。如果计算得出 RSCU = 1,表示该密码子编码氨基酸时无使用偏爱性。如果 RSCU > 1,则表明该密码子作为翻译信息使用时具有使用偏爱性,属于偏爱密码子,当 RSCU ≥ 1.5 时,称这种密码子为高频偏爱密码子^[16]。

2.4 ENC-Plot 分析 以 GC3s、ENC 值分别为横、纵坐标并绘制散点图。ENC-Plot 分析用于研究 ENC 值与 GC3 分布的关系^[17],基点在标准曲线上或附近时,表明突变影响对密码子的使用偏性作用较大,基点偏离标准曲线则说明自然选择压力对密码子使用偏性作用较大。

2.5 中性分析 中性分析主要用于衡量突变以及自然选择对密码子使用偏爱性的影响。该方法以 GC3s、GC12 值为横、纵坐标并绘制二维散点图。如果散点的趋势线越靠近 Y = X 的对角线,则表明密码子的使用模式主要受到突变压力的影响。反之,则主要受到自然选择作用的影响^[18]。

2.6 奇偶校验规则分析 在正常情况下,密码子的使用模式是 A = T, G = C,但是在现实中会受到基因突变或者外界环境选择压力的影响导致 GC 使用度受到一定偏爱。本研究使用奇偶校验规则 2 (Parity Rule2, PR2)绘图分析来显示与密码子和四简并同义密码子氨基酸相关的 A3/(A3+T3)和 G3/(G3+C3) 值的关系^[19]。主要分析由 4 个同义密码子编码的氨基酸(丙氨酸,苏氨酸,精氨酸,甘氨酸,脯氨酸,丝氨酸,缬氨酸,和亮氨酸),将各氨基酸密码子计算得到的数据作散点图,各散点与中心点(A = T, G = C)的偏离程度与距离代表该基因的偏离程度与趋势。

2.7 对比分析 将 NiV 基因组密码子使用频率与大肠埃希菌、酵母、人和昆虫杆状病毒的密码子使用频率进行比较,如果比值位于 0.5~2.0 之间,说明两种表达系统密码子使用偏爱性基本相似,如果比值 < 0.5, > 2.0 说明两种表达系统的偏爱性相差较大。

结果

1 CAI、ENC 分析

根据基因组数据分别计算 NiV 6 种蛋白的 ENC 值(图 1)。各蛋白的 ENC 值分布在 50.34~55.8 (52.03 ± 3.16)之间。可以看出 NiV 基因组有效密码子数均大于 50,表明 NiV 基因组密码子的偏爱性较低。CAI 值是 0.678 ± 0.027,取值范围为 0.606~0.730,接近于 0.7,表明 NiV 基因组表达水平相对较高。GC3s 值为 0.369 ± 0.046,取值范围为 0.292~0.530;GC 值为 0.433 ± 0.033,取值范围为 0.436~0.462。对以上各值进行相关性分析,CAI 与 GC3s 呈正相关($r = 0.849, P < 0.01$),说明 NiV 基因表达水平受组成限制影响的可能性较大。

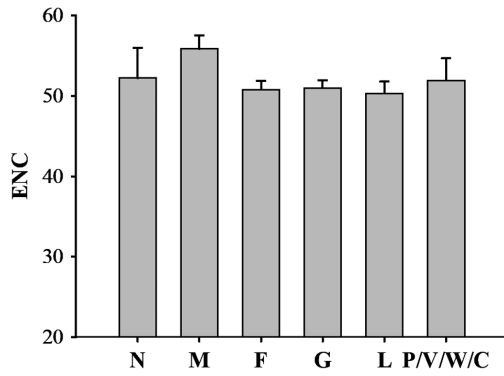


图 1 NiV 各蛋白基因组 ENC 值比较

Fig. 1 Comparison of the ENC value of each protein genome of NiV

2 相对密码子使用度分析

对 NiV 的 6 个蛋白基因组同义密码子进行 RSCU 值测定(表 1)。AUG, UGG 和 3 个终止密码子(UAG, UAA 和 UGA)因无同义密码子和 RSCU 值统计学意义,故排除分析 RSCU^[17]。6 个蛋白中 RSCU>1 即具有偏爱性的密码子个数分别是 24、31、28、29、29、26,其中 RSCU>1.5 的高频偏爱性密码子个数分别为 11、10、9、7、5、6。具有偏爱性的密码子多以 A、U 结尾,AGA、AGG、ACA、CUU、GUU、UCA 为 6 个蛋白共有高频密码子。还有一些密码子不具备代表性(RSCU<0.6),即 ACG、CGC、CGG、GCG、GGC、UCC、UCG。

3 ENC-Plot 分析

为了确定 NiV 基因组密码子使用模式是否受到突变和自然选择的影响,构建了 NiV6 个蛋白基因组的 ENC-Plot 分析图,并分析 GC3s 对密码子的影响(图 2)。结果表明 NiV 基因组 6 个蛋白的密码子使用模式大部分受突变和环境选择作用的双重影响,其中 P 基因的部分基因偏离曲线较大,可能与 P 基因共表达的 3 个蛋白差异有关。另外,N 基因的部分散点位于曲线偏上,说明 N 基因相对于其它蛋白基因在 NiV 基因中受到突变的影响作用较多。

4 中性分析

为了进一步探究自然选择和突变压力对 NiV 基因组密码子使用模式的影响,分析了 NiV6 个蛋白之间的 GC12 与 GC3s 值之间的线性回归关系。结果显示 6 个蛋白基因组 GC12 与 GC3s 值均有显著的相关性($P<0.05$),N、M、F、G、L、P/V/W/C 6 个蛋白的斜率分别是 -0.523、-0.256、-0.182、-0.277、0.020、-0.249(图 3)。因此自然选择的贡献率^[20]分别为 47.7%、74.4%、81.8%、72.3%、98.0%、75.1%,表明 NiV 大部分基因组密码子主要受到自然选择作用的影响。值得关注的是,N 蛋白受到突变压力和自然选择的双重影响但突变压力占主要作用。

表 1 NiV 各蛋白相对同义密码子使用度(RSCU)

Table 1 Relative synonymous codon usage of each protein of NiV							
AA	Condon	F	G	L	M	N	P/V/W/C
F(Phe)	UUU	<u>1.59</u>	0.91	1.17	0.90	1.17	1.11
F	UUC	0.36	1.09	0.83	1.10	0.83	0.86
L(Leu)	UUA	1.24	0.90	1.44	0.23	0.55	0.66
L	UUG	1.33	1.07	1.26	1.27	1.02	1.32
L	CUU	0.91	0.88	1.14	1.25	0.88	0.94
L	CUC	0.76	0.17	0.56	1.36	<u>2.00</u>	0.55
L	CUA	0.84	1.33	0.92	0.85	1.17	1.01
L	CUG	0.92	<u>1.65</u>	0.68	1.03	0.38	<u>1.53</u>
I(Ile)	AUU	1.19	0.91	0.87	<u>1.50</u>	1.18	1.28
I	AUC	0.85	1.10	0.88	1.04	0.98	0.97
I	AUA	0.96	0.99	1.25	0.45	0.84	0.74
V(Val)	GUU	<u>1.59</u>	1.48	1.27	<u>1.81</u>	<u>1.77</u>	1.49
V	GUC	0.99	0.59	0.71	0.81	0.63	0.72
V	GUA	0.91	1.26	1.27	0.62	0.26	0.79
V	GUG	0.52	0.66	0.75	0.76	1.34	1.00
S(Ser)	UCU	1.04	1.27	<u>1.57</u>	<u>1.53</u>	0.99	1.26
S	UCC	0.19	0.51	0.71	1.16	0.17	0.58
S	UCA	<u>1.74</u>	<u>1.89</u>	<u>1.95</u>	1.06	<u>1.77</u>	<u>2.02</u>
S	UCG	0.37	0.11	0.32	0.00	0.06	0.41
P(Pro)	CCU	<u>1.89</u>	<u>1.90</u>	<u>1.91</u>	1.40	<u>1.51</u>	1.32
P	CCC	0.57	0.62	0.62	0.25	0.27	0.86
P	CCA	0.89	1.18	1.05	<u>1.86</u>	0.77	1.20
P	CCG	0.65	0.29	0.43	0.49	1.44	0.56
T(Thr)	ACU	1.13	1.04	1.45	0.87	<u>1.59</u>	<u>1.65</u>
T	ACC	0.47	1.10	0.57	1.68	1.23	0.77
T	ACA	<u>1.97</u>	<u>1.67</u>	<u>1.66</u>	1.45	1.18	1.36
T	ACG	0.42	0.19	0.32	0.00	0.00	0.23
A(Ala)	GCU	1.40	1.26	1.49	<u>1.74</u>	<u>1.73</u>	1.10
A	GCC	0.52	0.62	0.72	1.32	0.58	0.54
A	GCA	<u>2.02</u>	<u>1.70</u>	1.47	0.80	1.49	<u>2.12</u>
A	GCG	0.06	0.42	0.31	0.15	0.20	0.24
Y(Tyr)	UAU	1.26	0.89	1.23	0.82	0.68	0.97
Y	UAC	0.74	1.11	0.77	1.18	0.25	1.03
H(His)	CAU	<u>1.62</u>	1.05	1.27	0.76	0.87	1.20
H	CAC	0.38	0.95	0.73	1.24	0.87	0.80
Q(Gln)	CAA	1.19	1.31	1.40	0.77	1.15	1.19
Q	CAG	0.81	0.69	0.60	1.23	0.81	0.81
N(Asn)	AAU	1.06	1.43	1.34	0.95	1.10	1.14
N	AAC	0.94	0.57	0.66	1.05	0.90	0.86
K(Lys)	AAA	1.18	1.32	1.21	0.86	<u>1.50</u>	1.11
K	AAG	0.82	0.68	0.79	1.14	0.50	0.89
D(Asp)	GAU	1.09	0.89	1.33	<u>1.64</u>	1.21	1.30
D	GAC	0.91	1.11	0.67	0.36	0.79	0.70
E(Glu)	GAA	1.12	0.98	1.12	0.56	1.35	1.17
E	GAG	0.88	1.02	0.88	1.44	0.65	0.83
C(Cys)	UGU	0.98	1.19	1.31	<u>1.82</u>	0.00	1.07
C	UGC	1.02	0.81	0.69	0.18	0.00	0.90
R(Arg)	CGU	0.00	0.61	0.62	0.85	0.48	0.38
R	CGC	0.00	0.00	0.03	0.25	0.00	0.20
R	CGA	1.03	0.42	0.58	0.90	0.48	0.88
R	CGG	0.00	0.19	0.08	0.18	0.19	0.39
S(Ser)	AGU	1.17	1.06	1.01	1.22	<u>2.10</u>	0.99
S	AGC	1.49	1.16	0.44	1.03	0.92	0.65
R(Arg)	AGA	<u>3.12</u>	<u>3.85</u>	<u>3.18</u>	<u>1.89</u>	<u>3.17</u>	<u>2.80</u>
R	AGG	<u>1.85</u>	0.94	1.50	<u>1.93</u>	<u>1.67</u>	1.35
G(Gly)	GGU	1.23	0.87	1.10	1.07	0.59	0.61
G	GGC	0.32	0.39	0.41	0.81	0.92	0.56
G	GGA	1.22	<u>1.64</u>	1.23	1.25	<u>1.86</u>	1.80
G	GGG	1.22	1.10	1.25	0.86	0.63	1.03

注:终止密码子,Trp 及 Met(RSCU 值为 1)未列入表中;AA 表示氨基酸,加下划线表示各蛋白的高频密码子(RSCU>1.5)。

Notes: The stop codon and Trp and Met (RSCU value of 1) are not included in the table; AA represents an amino acid; Underlined indicates the high frequency codons of each protein (RSCU>1.5).

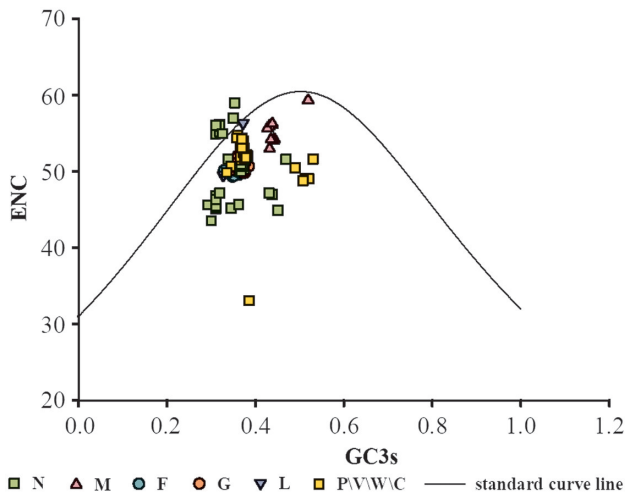


图 2 NiV 各蛋白基因 ENC-Plot 分析
Fig. 2 ENC-Plot analysis of NiV protein genes

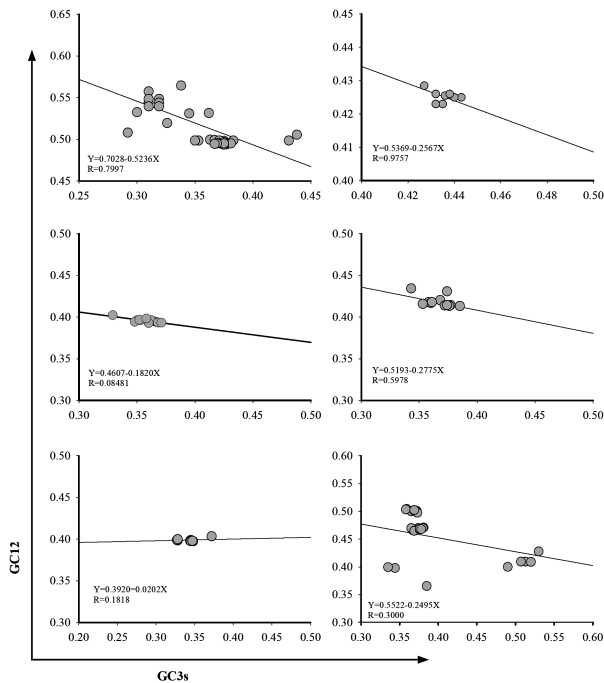


图 3 NiV 各蛋白中性绘图分析
Fig. 3 Neutrality analyses of each protein of the NiV

5 PR2 偏差图分析

为了分辨突变和自然选择作用对 NiV 基因组密码子使用模式影响的相对强度,对 NiV 基因组进行 PR2 分析(图 4)。结果显示 NiV 6 个蛋白的密码子偏爱性均有不同,6 个蛋白基因序列密码子第三位 A/U/C/G 的使用频率均有差异。平面图象限上方即 $y > 0.5$ 共有 331 个点,概率为 62.2% (331/532),象限右方即 $x > 0.5$ 共有 322 个点,概率为 60.5% (332/532)。说明在密码子的使用过程中,密码子第三位 A 的频率高于 U, G 的频率高于 C,即嘌呤的使用频率高于嘧啶。这种 A/G、C/U 使用的不对等性提示 NiV 密码子在使用过程中会受到突变压力和自然选择的双

重作用影响。

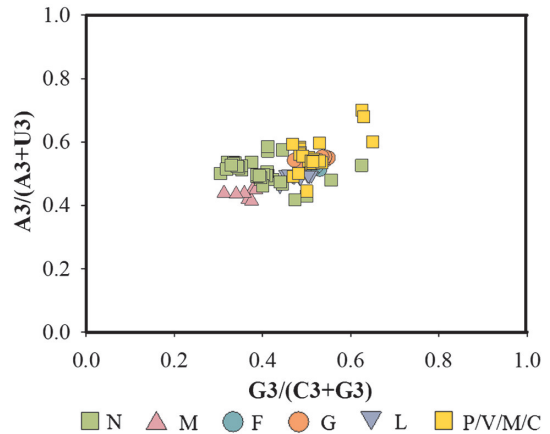


图 4 NiV 各蛋白基因组 PR2 分析
Fig. 4 Analysis of PR2 of each protein genome of NiV

6 对比分析

将 NiV 基因组密码子使用频率分别与大肠埃希菌、酵母、人类和昆虫杆状病毒表达系统进行比较(表 2),大肠埃希菌、酵母、人和昆虫杆状病毒表达系统密码子使用频率数据均来自 Codon Usage Database(<http://www.kazusa.or.jp/codon/>)。若密码子使用频率比值位于 0.5~2.0 之间,表明两种表达系统密码子使用偏爱性有一定相同。反之,数值大于 2.0 或小于 0.5 则表明偏爱性不相同。本研究中相对于 3 种不同表达系统的比值 < 0.5 , > 2.0 的数量分别为:大肠埃希菌 14 个(23.73%),酵母 6 个(10.17%),人类 20 个(33.90%),昆虫杆状病毒 18 个(30.50%)。表明酵母表达系统与 NiV 基因密码子偏爱性更为相似,更适合 NiV 外源基因的表达,人和昆虫杆状病毒表达系统与 NiV 基因密码子偏爱性存在一定差异,不适合 NiV 外源基因的表达。

讨论

病毒的密码子使用偏好性是病毒在适应宿主的进化过程中逐渐形成的。各种人为和自然因素导致了 NiV 的传播和发展,目前针对于 NiV 的感染尚无有效治疗措施和疫苗^[2]。通过对 NiV 基因组密码子偏爱性的研究有助于阐明 NiV 密码子使用模式。通过了解 NiV 密码子使用模式有助于研究 NiV 基因与蛋白的功能,优化基因的表达,进而为研究针对病毒表面抗原特殊表位的亚单位疫苗,或者通过抑制 NiV 蛋白的复制,制作减毒疫苗提供参考。

本研究显示, NiV 的 ENC 值均大于 50,表明 NiV 基因组密码子在使用过程中偏爱性均较低, CDS 可能使用了所有的同义密码子^[14]。肝炎病毒^[21]、人博卡病毒^[22]的 ENC 结果也显示出低偏爱性。CUB 较低

表2 NiV与几种不同表达系统密码子使用频率比较
Table 2 Comparison of codon usage frequency between NiV and several different expression systems

氨基酸 Amino acid	密码子 Codon	尼帕 1/1000(N)	大肠埃希菌 1/1000(E)	酵母 1/1000(Y)	人 1/1000(H)	昆虫杆状病毒 1/1000(P)	N/E	N/Y	N/H	N/P
A(Ala)	GCA	23.1	23.0	16.2	15.8	12.8	1.00	1.43	1.46	1.80
A	GCC	9.1	21.6	12.6	27.7	20.8	0.42	0.72	0.33	0.44
A	GCG	3.5	21.1	6.2	7.4	12.4	0.16	0.56	0.47	0.28
A	GCT	21.9	18.9	21.2	18.4	24.6	1.16	1.03	1.19	0.89
C(Cys)	TGC	5.7	5.5	4.8	12.6	13.2	1.03	1.18	0.45	0.43
C	TGT	9.4	5.9	8.1	10.6	8.5	1.59	1.15	0.88	1.10
D(Asp)	GAC	22.2	17.9	20.2	25.1	32.4	1.24	1.10	0.89	0.69
D	GAT	36.1	33.7	37.6	21.8	21.9	1.07	0.96	1.66	1.65
E(Glu)	GAA	36.5	35.1	45.6	29.0	27.6	1.04	0.80	1.26	1.32
E	GAG	27.5	19.4	19.2	39.6	33.1	1.42	1.43	0.69	0.83
F(Phe)	TTC	14.4	13.9	18.4	20.3	27.5	1.03	0.78	0.71	0.52
F	TTT	19.0	24.4	26.1	17.6	10.1	0.78	0.73	1.08	1.88
G(Gly)	GGA	23.8	13.6	10.9	16.5	17.8	1.75	2.18	1.44	1.34
G	GGC	9.0	20.6	9.8	22.2	19.6	0.44	0.92	0.41	0.46
G	GGG	14.5	12.3	6.0	16.5	4.6	1.18	2.41	0.88	3.14
G	GGT	12.5	23.7	23.9	10.8	21.5	0.53	0.52	1.16	0.58
H(His)	CAC	6.8	7.3	7.8	15.1	15.6	0.93	0.87	0.45	0.44
H	CAT	9.9	12.4	13.6	10.9	8.7	0.80	0.73	0.91	1.14
I(Ile)	ATA	26.6	13.3	17.8	7.5	8.3	2.00	1.49	3.54	3.20
I	ATC	25.4	19.4	17.2	20.8	28.3	1.31	1.47	1.22	0.90
I	ATT	27.6	29.6	30.1	16.0	15.5	0.93	0.92	1.73	1.78
K(Lys)	AAA	36.0	37.2	41.9	24.4	26.8	0.97	0.86	1.47	1.34
K	AAG	25.6	15.3	30.8	31.9	49.2	1.67	0.83	0.80	0.52
L(Leu)	CTA	15.1	5.6	13.4	7.2	7	2.70	1.13	2.10	2.16
L	CTC	12.9	9.5	5.4	19.6	17	1.36	2.39	0.66	0.76
L	CTG	14.2	37.4	10.5	39.6	24.1	0.38	1.35	0.36	0.59
L	CTT	15.8	14.5	12.3	13.2	9.6	1.09	1.29	1.20	1.65
L	TTA	15.8	17.4	26.2	7.7	7.7	0.91	0.60	2.05	2.05
L	TTG	18.3	12.9	27.2	12.9	15.9	1.42	0.67	1.42	1.15
M(Met)	ATG	25.1	23.7	20.9	22.0	27	1.06	1.20	1.14	0.93
N(Asn)	AAC	21.7	20.3	24.8	19.1	28.8	1.07	0.87	1.13	0.75
N	AAT	35.6	29.3	35.7	17.0	13.4	1.22	1.00	2.10	2.66
P(Pro)	CCA	12.4	9.1	18.3	16.9	13.6	1.36	0.68	0.73	0.91
P	CCC	6.9	6.2	6.8	19.8	13.7	1.11	1.01	0.35	0.50
P	CCG	5.7	14.5	5.3	6.9	7.8	0.40	1.08	0.83	0.74
P	CCT	17.7	9.5	13.5	17.5	14.1	1.86	1.31	1.01	1.26
Q(Gln)	CAA	23.9	14.4	27.3	12.3	16.1	1.66	0.87	1.94	1.48
Q	CAG	14.4	26.7	12.1	34.2	21.6	0.54	1.19	0.42	0.67
R(Arg)	AGA	26.8	7.1	21.3	12.2	12.1	3.77	1.26	2.20	2.21
R	AGG	13.0	4.0	9.2	12.0	12.6	3.26	1.42	1.09	1.04
R	CGA	5.2	4.8	3.0	6.2	5.4	1.09	1.74	0.84	0.97
R	CGC	0.6	14.0	2.6	10.4	15.6	0.04	0.24	0.06	0.04
R	CGG	1.7	7.9	1.7	11.4	3.7	0.21	0.98	0.15	0.45
R	CGT	4.1	15.9	6.4	4.5	15.3	0.26	0.64	0.91	0.27
S(Ser)	AGC	12.7	14.3	9.8	19.5	11.1	0.89	1.29	0.65	1.14
S	AGT	19.1	13.2	14.2	12.1	8.4	1.45	1.35	1.58	2.27
S	TCA	27.3	13.1	18.7	12.2	10.6	2.09	1.46	2.24	2.58
S	TCC	7.8	9.7	14.2	17.7	12.6	0.81	0.55	0.44	0.62
S	TCG	3.7	8.2	8.6	4.4	7.7	0.46	0.43	0.85	0.49
S	TCT	20.2	13.1	23.5	15.2	10.7	1.54	0.86	1.33	1.88
T(Thr)	ACA	21.1	15.1	17.8	15.1	12.3	1.40	1.19	1.40	1.72
T	ACC	10.3	18.9	12.7	18.9	17.2	0.54	0.81	0.54	0.60
T	ACG	2.9	13.6	8.0	6.1	9.4	0.21	0.36	0.47	0.30
T	ACT	19.5	13.1	20.3	13.1	14.3	1.49	0.96	1.49	1.36
V(Val)	GTA	13.6	13.1	11.8	7.1	12	1.04	1.15	1.91	1.13
V	GTC	11.0	13.1	11.8	14.5	20.1	0.84	0.93	0.76	0.55
V	GTT	12.2	19.9	10.8	28.1	24.1	0.62	1.13	0.44	0.51
V	GTT	22.1	21.6	22.1	11.0	14.1	1.02	1.00	2.01	1.57
W(Trp)	TGG	9.1	13.4	10.4	13.2	13.3	0.68	0.87	0.69	0.68
Y(Tyr)	TAC	14.5	11.7	14.8	15.3	24.4	1.24	0.98	0.95	0.60
Y	TAT	18.7	21.6	18.8	12.2	10	0.86	0.99	1.53	1.87
*	TAA	0	2.0	1.1	1.0	2.3	0	0	0	0
*	TAG	0	0.3	0.5	0.8	0.7	0	0	0	0
*	TGA	0	1.1	0.7	1.6	0.7	0	0	0	0

注:下划线表示两个物种均具有明显偏差(>2.0, <0.5),大肠埃希菌、酵母、人和昆虫杆状病毒表达系统密码子使用频率数据均来自 Codon Usage Database(<http://www.kazusa.or.jp/codon/>)。

的病毒可能在入侵机体后更容易自我复制、转录和翻译蛋白^[23]。RSCU 结果显示 6 个蛋白的密码子以 A/U 结尾的较多, 鉴定出 6 个高频密码子和 7 个低频密码子, 与 SARS-CoV-2^[24]、Ebola virus^[25] 的 RSCU 分析结果相似。NiV 的 A/U 使用率较 C/G 大, A/U 含量又与热力稳定性密切相关, 提示 NiV 的碱基组成在复制过程中起到重要作用, 这可能与病毒蛋白的高效准确翻译和折叠有关, 有助于增强病毒对自然选择的适应。

ENC-Plot、中性绘图和 PR2 分析显示, NiV 基因组密码子使用模式受到突变压力和自然选择的双重影响, 但自然选择起主要作用。在新型冠状病毒(SARS-CoV-2)^[18]、新现 HIV-1L 亚型^[26] 的研究中也获得类似的结果。同样, 对 Zika virus^[27] 的研究显示 GC12 与 GC3s 存在显著的正相关性, 表明自然选择是影响 CUB 的主要因素。H3N2 犬流感病毒^[28] 研究结果显示, 突变压力对于密码子的偏爱性没有自然选择作用大。提示在影响密码子偏爱性方面, 自然选择的作用可能比突变压力和二核苷酸丰度等因素更为重要。本研究将 NiV 密码子使用频率与大肠埃希菌、酵母、人和昆虫杆状病毒做了比较分析, 结果表明 NiV 密码子使用频率与酵母更为接近, 提示真核表达系统可能是 NiV 最适合的体外表达系统。杜加亮等^[29] 运用酵母表达系统成功表达诺如病毒 VP1 蛋白的病毒样颗粒(VLP)。Lav T 等^[30] 运用毕赤酵母表达出登革热病毒 DENV-2 型的 VLP, 这些 VLP 引发小鼠机体产生病毒特异性中和抗体, 为开发廉价、安全和有效的亚单位疫苗奠定了基础。

【参考文献】

- [1] Singh R K, Dhama K, Chakraborty S, et al. Nipah virus: epidemiology, pathology, immunobiology and advances in diagnosis, vaccine designing and control strategies - a comprehensive review[J]. Vet Q, 2019, 39(1): 26-55.
- [2] Ang B, Lim T, Wang L. Nipah virus infection[J]. J Clin Microbiol, 2018, 56(6): e01875-17.
- [3] Bellini W J, Harcourt B H, Bowden N, et al. Nipah virus: an emergent paramyxovirus causing severe encephalitis in humans[J]. J Neurovirol, 2005, 11(5): 481-487.
- [4] Chua K B, Bellini W J, Rota P A, et al. Nipah virus: a recently emergent deadly paramyxovirus[J]. Science, 2000, 288(5470): 1432-1435.
- [5] Sendow I, Field H E, Adjid A, et al. Screening for Nipah virus infection in West Kalimantan province, Indonesia[J]. Zoonoses Public Health, 2010, 57(7-8): 499-503.
- [6] Wacharapluesadee S, Boongird K, Wanghongsa S, et al. A longitudinal study of the prevalence of Nipah virus in Pteropus lylei bats in Thailand; evidence for seasonal preference in disease transmission[J]. Vector Borne Zoonotic Dis, 2010, 10(2): 183-190.
- [7] Yadav P D, Raut C G, Shete A M, et al. Detection of Nipah virus RNA in fruit bat (Pteropus giganteus) from India[J]. Am J Trop Med Hyg, 2012, 87(3): 576-578.
- [8] 林志龙, 严延生, 张智芳, 等. 尼帕病毒病[J]. 中国人兽共患病学报, 2021, 37(3): 246-251.
- [9] Dawes B E, Kalveram B, Ikegami T, et al. Favipiravir (T-705) protects against Nipah virus infection in the hamster model[J]. Sci Rep, 2018, 8(1): 7604.
- [10] Hanson G, Collier J. Codon optimality, bias and usage in translation and mRNA decay[J]. Nat Rev Mol Cell Biol, 2018, 19(1): 20-30.
- [11] Behura S K, Severson D W. Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes[J]. Biol Rev Camb Philos Soc, 2013, 88(1): 49-61.
- [12] Alonso A M, Diambra L. SARS-CoV-2 codon usage bias down-regulates host expressed genes With similar codon usage[J]. Front Cell Dev Biol, 2020(8): 831.
- [13] Sharma V, Kaushik S, Kumar R, et al. Emerging trends of Nipah virus: A review[J]. Rev Med Virol, 2019, 29(1): e2010.
- [14] Chen Y, Shi Y, Deng H, et al. Characterization of the porcine epidemic diarrhea virus codon usage bias[J]. Infect Genet Evol, 2014(28): 95-100.
- [15] Chakraborty S, Deb B, Barbhuiya PA, et al. Analysis of codon usage patterns and influencing factors in Nipah virus[J]. Virus Res, 2019(263): 129-138.
- [16] Butt A M, Nasrullah I, Qamar R, et al. Evolution of codon usage in Zika virus genomes is host and vector specific[J]. Emerg Microbes Infect, 2016, 5(10): e107.
- [17] Wright F. The 'effective number of codons' used in a gene[J]. Gene, 1990, 87(1): 23-29.
- [18] 尚方建, 石哲芳, 王聪, 等. 新型冠状病毒(SARS-CoV-2)的密码子偏爱性分析[J]. 中国人兽共患病学报, 2021, 37(1): 15-21.
- [19] Duan H, Zhang Q, Wang C, et al. Analysis of codon usage patterns of the chloroplast genome in *Delphinium grandiflorum* L. reveals a preference for AT-ending codons as a result of major selection constraints[J]. PeerJ, 2021(9): e10787.
- [20] Yu X, Liu J, Li H, et al. Comprehensive analysis of synonymous codon usage patterns and influencing factors of porcine epidemic diarrhea virus[J]. Arch Virol, 2021, 166(1): 157-165.
- [21] Deb B, Uddin A, Chakraborty S. Codon usage pattern and its influencing factors in different genomes of hepadnaviruses [J]. Arch Virol, 2020, 165(3): 557-570.
- [22] Zhao S, Zhang Q, Liu X, et al. Analysis of synonymous codon usage in 11 Human Bocavirus isolates[J]. BioSystems, 2008, 92(3): 207-214.
- [23] Tao J, Yao H. Comprehensive analysis of the codon usage patterns of polyprotein of Zika virus[J]. Prog Biophys Mol Biol, 2020(150): 43-49.
- [24] Hou W. Characterization of codon usage pattern in SARS-CoV-2 [J]. Virol J, 2020, 17(1): 138.

(下转 169 页)

