

黄蔚虹, 陈永杰, 孙彦阔, 等. H9N2 禽流感病毒全基因组密码子使用偏好性及影响因素分析 [J]. 华南农业大学学报, 2020, 41(3): 15-22.  
HUANG Weihong, CHEN Yongjie, SUN Yankuo, et al. Codon usage bias of H9N2 avian influenza virus complete genomes and its influence factors[J].  
Journal of South China Agricultural University, 2020, 41(3): 15-22.

# H9N2 禽流感病毒全基因组密码子 使用偏好性及影响因素分析

黄蔚虹<sup>1</sup>, 陈永杰<sup>2</sup>, 孙彦阔<sup>2</sup>, 孙卫东<sup>1</sup>

(1 南京农业大学 动物医学院, 江苏 南京 210095;

2 广东省动物源性人兽共患病预防与控制重点实验室/华南农业大学 兽医学院, 广东 广州 510642)

**摘要:**【目的】探究 H9N2 禽流感病毒 (AIV) 全基因组的密码子使用偏好性及影响因素。【方法】选取 2010—2018 年国内 H9N2 AIV 流行毒株的全基因组为研究对象, 分析其碱基组成特性、最优密码子、密码子使用偏好性的影响因素以及病毒对宿主密码子使用模式的适应性。【结果】H9N2 AIV 的全基因组中 AU 含量高于 GC。大部分最优密码子以 A 或 U 结尾, 有效密码子数 (ENC) 平均值为 52.86, 提示存在密码子使用偏好性但偏好性较低。密码子使用偏好性主要受到突变压力和自然选择的共同作用, 其中自然选择 (所占比例为 61.79%~76.15%) 作用大于突变压力 (所占比例为 23.85%~38.21%)。H9N2 AIV 对人 *Homo sapiens* 的密码子适应指数平均值为 0.739~0.741, 提示 H9N2 AIV 禽流感病毒可能已适应人类的密码子使用模式。【结论】本研究为 H9N2 AIV 的基因进化分析、已有疫苗的密码子优化和新型疫苗 (密码子去优化疫苗) 研制提供了理论依据。

**关键词:** H9N2 禽流感病毒; 密码子使用偏好性; 自然选择; 突变压力

中图分类号: S852.659.5

文献标志码: A

文章编号: 1001-411X(2020)03-0015-08

## Codon usage bias of H9N2 avian influenza virus complete genomes and its influence factors

HUANG Weihong<sup>1</sup>, CHEN Yongjie<sup>2</sup>, SUN Yankuo<sup>2</sup>, SUN Weidong<sup>1</sup>

(1 College of Veterinary Medicine, Nanjing Agricultural University, Nanjing 210095, China;

2 Key Laboratory of Zoonosis Prevention and Control of Guangdong Province/College of Veterinary Medicine, South China Agricultural University, Guangzhou 510642, China)

**Abstract:** 【Objective】To study codon usage bias of H9N2 avian influenza virus (AIV) complete genomes and its influence factors. 【Method】The complete genomes of Chinese epidemic H9N2 AIV strains from 2010 to 2018 were selected. The characteristics of base composition, optimal codons, influence factors of codon usage bias and adaption to the codon usage patterns of the host were analyzed. 【Result】AU content was higher than GC content in the whole genomes of H9N2 AIV. Most of the optimal codons ended with A or U, and the average of effective number of codons (ENC) was 52.86, suggesting that codon usage bias existed, but the bias was low. The codon usage bias was mainly affected by mutation pressure and natural selection. Natural selection (accounting for 61.79%~76.15%) played a greater role than mutation pressure (accounting for 23.85%~38.21%). In addition, the average codon adaptation index of H9N2 AIV to *Homo sapiens* ranged from 0.739 to 0.741, suggesting that H9N2

收稿日期: 2019-08-14 网络首发时间: 2020-04-22 11:01:08

网络首发地址: <http://kns.cnki.net/kcms/detail/44.1110.s.20200421.1055.034.html>

作者简介: 黄蔚虹 (1995—), 女, 硕士研究生, E-mail: [mishelle-cat@163.com](mailto:mishelle-cat@163.com); 通信作者: 孙卫东 (1969—), 男, 教授, 博士, E-mail: [swd100@njau.edu.cn](mailto:swd100@njau.edu.cn)

基金项目: 国家重点研发计划 (2016YFD0500800-10); 江苏肉鸡生产全程关键技术集成示范应用 [TG(17)003]

AIV might have adapted to human codon usage patterns. 【Conclusion】 The study provides a theoretical basis for genetic evolution analysis of H9N2 AIV, codon optimization of existing vaccines and development of new vaccines (codon deoptimization vaccine).

**Key words:** H9N2 avian influenza virus; codon usage bias; natural selection; mutation pressure

H9N2 禽流感在 1994 年首次暴发于我国广东省, 引起蛋鸡产蛋率降低 (14%~75%) 和肉鸡死亡 (10%~40%)<sup>[1]</sup>。在过去 20 多年中, H9N2 禽流感病毒 (Avian influenza virus, AIV) 迅速蔓延传播, 成为我国流行最为广泛的 AIV 亚型<sup>[2]</sup>。H9N2 AIV 与细菌、病毒等病原体共感染时, 可造成家禽死亡率大幅升高, 对家禽养殖业构成严重威胁。同时, H9N2 亚型禽流感已出现跨种传播的现象, 广东、香港、湖南和安徽均出现人感染 H9N2 AIV 的病例<sup>[2]</sup>。

同义密码子的使用具有某种倾向性, 并非是一个随机的过程。在蛋白质翻译过程中, 部分同义密码子会比其他密码子使用度更高, 这被称为密码子使用偏好性 (Codon usage bias)<sup>[3]</sup>。若病毒在同一宿主长时间循环, 则该病毒的密码子使用模式将趋同于宿主的密码子使用模式 (即病毒使用与宿主相同的同义密码子以翻译某一氨基酸), 以便更好地利用宿主细胞资源进行病毒复制、繁殖<sup>[4]</sup>。研究密码子使用模式有助于解释病毒在基因进化过程中所发生的适应性变化, 进一步认识病毒的基因表达调控, 有助于病毒疫苗的研制<sup>[5-6]</sup>。H9N2 AIV 对家禽养殖和人类健康的威胁日益加剧, 但目前对 H9N2 AIV 密码子使用模式的研究较少。

为了探讨近年来 H9N2 AIV 的密码子使用偏好性及其影响因素, 本研究选取 90 株 2010—2018 年 H9N2 AIV 的全基因组为研究对象, 分析其碱基组成特性、最优密码子、密码子使用偏好性的影响因素以及病毒对宿主密码子使用模式的适应性, 以期对 H9N2 AIV 基因进化、已有疫苗密码子的优化和新型疫苗 (密码子去优化疫苗) 研制提供理论依据。

## 1 材料与方法

### 1.1 序列来源

利用 NCBI 数据库 (<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>) 检索获取 2010—2018 年中国 H9N2 AIV 全基因组 (内含 8 个节段的完整编码序列), 根据遗传进化分析结果, 从中选取 90 个 H9N2 AIV 流行毒株的全基因组。每株病毒的编码序列按 PB2+PB1+PA+HA+NP+NA+M+NS 顺序进行拼接。

### 1.2 密码子使用偏好性参数分析

利用 CodonW 1.4.4 软件计算 H9N2 AIV 基因组的密码子使用偏好性参数, 包括同义密码子第 3 位各碱基 (A<sub>3S</sub>、U<sub>3S</sub>、C<sub>3S</sub> 和 G<sub>3S</sub>) 含量, 同义密码子第 3 位的 GC(GC<sub>3S</sub>) 含量, 有效密码子数 (Effective number of codons, ENC), 总平均亲水性 (General average hydrophobicity) 和芳香性 (Aromaticity)。用 Puigbò 等<sup>[7]</sup>的方法计算各碱基 (A、U、C 和 G) 含量, GC 含量, 密码子第 3 位各碱基 (A<sub>3</sub>、U<sub>3</sub>、C<sub>3</sub> 和 G<sub>3</sub>) 含量, 密码子第 3 位的 GC(GC<sub>3</sub>) 含量, 密码子第 1、2 位的 GC(GC<sub>12</sub>) 平均含量, 同义密码子相对使用度 (Relative synonymous codon usage) 和密码子适应指数 (Codon adaptation index)。宿主鸡 *Gallus gallus* 和人 *Homo sapiens* 的密码子使用模式获取自密码子使用数据库 (<http://www.kazusa.or.jp/codon/>)<sup>[8]</sup>。

### 1.3 有效密码子数绘图分析

ENC 为 20~60, 与密码子使用偏好性呈负相关, ENC ≤ 35 说明密码子使用偏好性高, ENC > 50 说明密码子使用偏好性低<sup>[9]</sup>。ENC 绘图是以 ENC 为纵坐标、GC<sub>3S</sub> 含量为横坐标绘制的二维散点图, 若基因组位于标准曲线 (即 ENC<sub>期望</sub> 与 GC<sub>3S</sub> 含量的函数关系) 上或其附近, 表明密码子使用偏好性仅受突变压力影响; 若基因组位于标准曲线下方较远的位置, 表明密码子使用偏好性受其他因素 (如自然选择) 影响。ENC<sub>期望</sub> 与 GC<sub>3S</sub> 含量的函数关系为:

$$ENC_{期望} = 2 + s + \frac{29}{s^2 + (1 - s)^2},$$

式中,  $s$  为 GC<sub>3S</sub> 含量<sup>[9]</sup>。

### 1.4 中性绘图分析

采用中性绘图 [ $w(GC_3)$  为横坐标、 $w(GC_{12})$  为纵坐标] 分析突变压力和自然选择对密码子使用偏好性的影响程度。若回归曲线斜率接近 1, 即基因组几乎沿对角线分布, 表示密码子使用偏好性仅受突变压力影响; 随着斜率逐渐减小, 甚至减小为 0, 自然选择对密码子使用偏好性的作用逐渐增大<sup>[5]</sup>。

### 1.5 PR2 绘图分析

以  $w(A_3)/w(A_3+U_3)$  为纵坐标、 $w(G_3)/w(G_3+C_3)$

为横坐标绘制二维散点图,若  $w(A_3)=w(U_3)$  且  $w(C_3)=w(G_3)$ , 基因组均匀分布于图中心(横、纵坐标均为 0.5), 表明突变压力和选择率(替代率)对密码子使用的作用相等;反之,两者作用存在差异<sup>[5]</sup>。

### 1.6 密码子适应指数分析

密码子适应指数为 0~1, 值越高, 病毒对宿主的适应性越强<sup>[10]</sup>。密码子适应指数和 ENC 之间的相关性分析可确定突变压力和自然选择对密码子使用偏好性的相对影响。若自然选择优于突变压力, 则密码子适应指数和 ENC 之间的相关性高, 相关系数( $r$ )趋向于-1; 相反, 若突变压力更为重要, 则  $r$  应接近 0, 密码子适应指数和 ENC 之间无相关性<sup>[11]</sup>。

### 1.7 对应性分析

同义密码子相对使用度是其实际使用频率与期望频率(某一特定氨基酸的所有同义密码子能被均等使用的情况下该密码子的使用频率)的比率。同义密码子相对使用度等于 1 代表对应氨基酸无密码子使用偏好性, 大于 1 代表对应氨基酸对该密码子的使用较多, 小于 1 代表对该密码子的使用较少。对应性分析(Correspondence analysis)是常用于研究密码子使用模式的多元统计分析方法, 基于各基因组的同义密码子相对使用度, 将每个基因组依密码子使用频率在 1 个 59 维的向量空间内表示<sup>[12]</sup>。

### 1.8 最优密码子分析

各基因组的 ENC 从高到低依次排序, 从上限

区域和下限区域各取 10% 的基因组作为高、低表达组, 使用  $t$  检验分析两组之间密码子使用的差异显著性。比较两组的同义密码子相对使用度, 若差异  $\geq 0.08$ , 并且高表达组显著高于低表达组 ( $P<0.01$ ), 则该密码子被定义为最佳密码子<sup>[9, 13]</sup>。

### 1.9 统计学分析

使用 CodonW 1.4.4 软件对同义密码子相对使用度进行对应性分析, 通过 SPSS 23.0 软件完成  $t$  检验和 Spearman 相关性分析, 使用 SigmaPlot 12.5 软件绘图。

## 2 结果与分析

### 2.1 碱基组成特性

由 2010—2018 年 H9N2 AIV 基因组的碱基组成分析结果(表 1)可知, 各碱基含量为 A(32.92%~33.03%)>G(24.65%~24.72%)>U(22.88%~22.90%)>C(19.41%~19.49%)。2010—2018 年 GC 含量整体呈下降趋势, 且均少于 50%, 提示 H9N2 AIV 基因组中 AU 含量多于 GC 含量, 碱基组成更倾向于使用 AU, 随着时间推移, 该倾向性逐渐加大。同义密码子第 3 位碱基含量为  $A_{3S}$ (41.42%~41.80%)> $U_{3S}$ (30.76%~30.95%)> $G_{3S}$ (27.74%~28.11%)> $C_{3S}$ (26.51%~26.73%),  $GC_{3S}$  含量变化与 GC 相同, 说明以 A 或 U 作为末位碱基的同义密码子的使用度较大, 随着时间推移, 其使用度将逐渐增大。

表 1 H9N2 禽流感病毒基因组的碱基组成分析<sup>1)</sup>  
Table 1 Base composition analysis of H9N2 avian influenza virus genomes

碱基 <sup>2)</sup> Base	年份 Year		
	2010—2012	2013—2015	2016—2018
A	32.92±0.09	32.92±0.10	33.03±0.08
C	19.49±0.09	19.48±0.07	19.41±0.09
U	22.89±0.09	22.88±0.10	22.90±0.10
G	24.71±0.15	24.72±0.13	24.65±0.11
GC	44.19±0.16	44.20±0.15	44.07±0.14
$A_{3S}$	41.42±0.52	41.48±0.79	41.80±0.54
$C_{3S}$	26.57±0.86	26.51±1.05	26.73±0.70
$U_{3S}$	30.95±0.60	30.83±1.48	30.76±1.07
$G_{3S}$	28.08±0.49	28.11±0.98	27.74±1.24
$GC_{3S}$	42.06±0.58	42.07±0.69	41.93±0.73

1)表中数据为平均值±标准差; 2)下标“3S”代表同义密码子第3位

1)The datum in the table is average ± standard deviation; 2) The subscripted “3S” indicates the 3rd place of synonymous codon

### 2.2 最优密码子分析

以 ENC 为依据形成的高、低表达组数据如表 2 所示, 大部分同义密码子相对使用度差异较小且无显著性差异, 仅有 10 个密码子(UUU、GUU、GUA、

GCU、GCC、GAU、CGU、CGA、GGU 和 GGC) 满足相对使用度差异  $\geq 0.08$  且高表达组相对使用度显著高于低表达组 ( $P<0.01$ ) 的条件, 因此这 10 个密码子可判定为最优密码子。其中有 8 个最优密码子均

表 2 H9N2 禽流感病毒基因组的最优密码子分析  
Table 2 Optimal codon analysis of H9N2 avian influenza virus genomes

氨基酸 Amino acid	密码子 <sup>1)</sup> Codon	同义密码子相对使用度 Relative synonymous codon usage		氨基酸 Amino acid	密码子 <sup>1)</sup> Codon	同义密码子相对使用度 Relative synonymous codon usage	
		高表达组 High expression group	低表达组 Low expression group			高表达组 High expression group	低表达组 Low expression group
		Ala	GCT*			1.130	0.957
	GCC*	0.884	0.730		TTC	1.040	1.184
	GCA	1.610	1.972	Leu	TTA	0.680	0.571
	GCG	0.376	0.340		TTG	1.330	1.164
Arg	AGA	2.386	2.875		CTT	1.162	1.081
	AGG	1.782	1.707		CTC	0.948	0.912
	CGT*	0.278	0.226		CTA	0.877	1.065
	CGC	0.290	0.277		CTG	1.002	1.208
	CGA*	0.834	0.462	Lys	AAA	1.076	1.129
	CGG	0.431	0.453		AAG	0.924	0.872
Asn	AAT	1.139	1.122	Pro	CCT	1.098	1.120
	AAC	0.862	0.878		CCC	0.880	0.703
Asp	GAT*	1.181	1.076		CCA	1.488	1.612
	GAC	0.819	0.924		CCG	0.535	0.564
Cys	TGT	0.938	0.885	Ser	AGT	1.189	1.141
	TGC	1.062	1.115		AGC	1.193	1.117
Gln	CAA	1.149	1.095		TCT	1.059	1.153
	CAG	0.851	0.905		TCC	0.782	0.802
Glu	GAA	1.206	1.206		TCA	1.340	1.428
	GAG	0.794	0.794		TCG	0.436	0.360
Gly	GGT*	0.711	0.580	Thr	ACT	1.040	0.908
	GGC*	0.735	0.559		ACC	0.902	0.863
	GGA	1.744	1.759		ACA	1.656	1.862
	GGG	0.810	1.103		ACG	0.402	0.367
His	CAT	1.243	1.259	Tyr	TAT	1.176	1.170
	CAC	0.757	0.741		TAC	0.824	0.830
Ile	ATT	1.162	1.129	Val	GTT*	1.037	0.915
	ATC	0.876	0.813		GTC	0.876	0.852
	ATA	0.962	1.058		GTA*	0.836	0.716
					GTG	1.250	1.517

1) “\*” 表示高表达组的同义密码子相对使用度显著高于低表达组且差异 $\geq 0.08$  ( $P < 0.01$ ,  $t$  检验)

1) “\*” indicates relative synonymous codon usage of high expression group is significantly higher than that of low expression group and the difference is equal or higher than 0.08 ( $P < 0.01$ ,  $t$  test)

以 A 或 U 结尾, 仅有 2 个以 C 结尾, 说明最优密码子的形成可能主要受到碱基组成限制, 可能存在突变压力影响密码子使用偏好性。

### 2.3 突变压力和自然选择共同作用于密码子使用偏好性

2010—2018 年 H9N2 AIV 的 ENC 为 51.90~55.01, 平均值 52.86, 标准差 0.64, 说明 H9N2 AIV 的密码子使用偏好性较低。在 ENC 绘图 (图 1)

分析中, 所有点均聚集于标准曲线的下方, 说明 H9N2 AIV 密码子使用偏好性的主导因素并非突变压力, 而是其他影响因素如自然选择。

PR2 绘图分析结果 (图 2) 显示, 2010—2018 年 H9N2 AIV 基因组主要聚集于图中心 (横、纵坐标均为 0.5) 的右上方, 基因组对 AU 的使用度大于 GC, 其中  $w(A_3) > w(U_3)$  且  $w(G_3) > w(C_3)$ , 这表明密码子使用偏好性可能由突变压力和其他因素 (如自然选

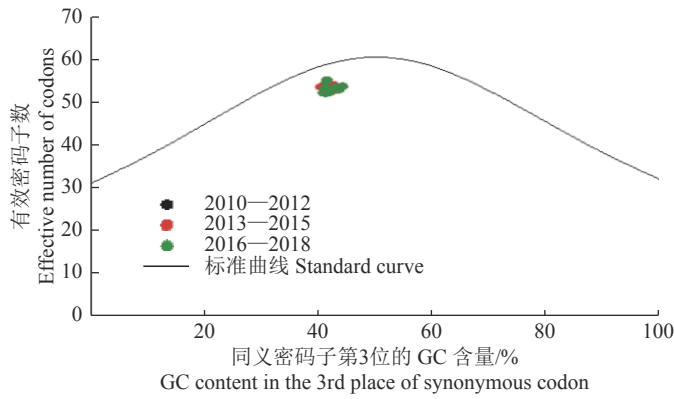
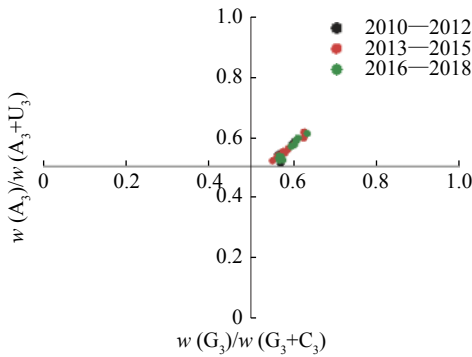


图1 不同年份 H9N2 禽流感病毒的有效密码子数绘图分析

Fig. 1 Plot analysis of effective number of codons of H9N2 avian influenza virus in different years



下标“3”代表密码子第3位

The subscripted “3” indicates the 3rd place of codon

图2 不同年份 H9N2 禽流感病毒的 PR2 绘图分析

Fig. 2 PR2 plot analysis of H9N2 avian influenza virus in different years

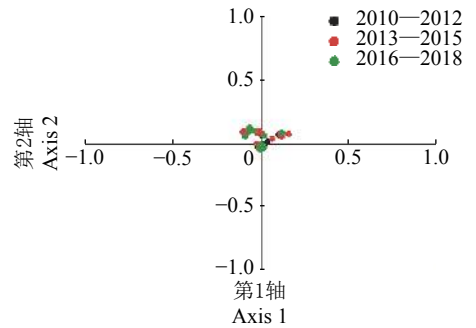


图3 不同年份 H9N2 禽流感病毒的对对应性分析

Fig. 3 Correspondence analysis of H9N2 avian influenza virus in different years

择) 共同决定, 但两者作用存在差异。

2.3.1 H9N2 禽流感病毒密码子使用偏好性受突变压力的影响 对 2010—2018 年 H9N2 AIV 进行对应性分析 (图 3), 结果显示第 1、2 轴分别解释 34.65% 和 21.34% 的变异, 大部分基因组都聚集在原点 (横、纵坐标均为 0) 周围, 彼此之间不存在太大差异, 这提示各基因组间的密码子使用偏好性差

异相对较小。第 1、2 轴与碱基组成指数进行相关性分析 (表 3), 结果显示第 1、2 轴与 ENC 均呈极显著正相关 ( $r=0.633, 0.216, P<0.001$ ); 第 1、2 轴与大部分碱基组成指数呈显著相关, 其中第 1 轴与 GC 含量呈极显著正相关 ( $r=0.510, P<0.001$ ), 第 2 轴与  $G_{3S}$  含量呈极显著负相关 ( $r=-0.717, P<0.001$ )。这提示碱基组成对密码子使用偏好性产生影响。

2.3.2 H9N2 禽流感病毒密码子使用偏好性受自然选择的影响 第 1、2 轴与总平均亲水性、芳香性进

表3 前两主轴对应性分析及与密码子使用偏好性主要参数的相关性分析

Table 3 Correspondence analyses of the first two principle axes and correlation analyses between the first two principle axes and main parameters of codon usage bias

前两主轴 First two principle axes	对应性分析差异比例 Variation proportion of correspondence analysis	相关性 <sup>1)</sup> Correlation									
		$w(U_{3S})$	$w(C_{3S})$	$w(A_{3S})$	$w(G_{3S})$	$w(GC_{12})$	$w(GC_{3S})$	$w(GC)$	ENC	总平均亲水性 General average hydropathicity	芳香性 Aromaticity
第1轴 Axis 1	34.65%	0.091	-0.348**	0.262*	-0.014	0.340**	-0.385**	0.510**	0.633**	-0.533**	-0.418**
第2轴 Axis 2	21.34%	0.471**	0.513**	-0.181	-0.717**	0.295**	-0.291**	-0.234*	0.216**	-0.233*	0.342**

1) “\*” 代表  $P<0.05$ , “\*\*” 代表  $P<0.01$ (Spearman 相关性分析); 下标“3S”代表同义密码子第3位, 下标“12”代表密码子第1、2位, “ENC”代表有效密码子数

1) “\*” represents  $P<0.05$ , “\*\*” represents  $P<0.01$ (Spearman’s rank correlation analysis); The subscripted “3S” indicates the 3rd place of synonymous codon, The subscripted “12” indicates the 1st and 2nd places of codon, “ENC” indicates effective number of codons

行相关性分析(表 3), 结果显示第 1、2 轴与总平均亲水性、芳香性均存在显著相关性, 这说明氨基酸的疏水性和芳香性均显著影响密码子使用偏好性。此外, 对 2010—2018 年 H9N2 AIV 进行密码子适应指数分析(表 4), 结果显示 H9N2 对宿主的密码子适应指数平均值随时间变化不大, 但 H9N2 AIV 对鸡的密码子适应指数平均值(0.766~0.768)

均高于对人的(0.739~0.741)。同时, 密码子适应指数与 ENC 相关性分析发现, 除了 2013—2015 年 H9N2 AIV 对人的密码子适应指数与 ENC 相关性不显著( $r=-0.223, P=0.237$ ), 其余密码子适应指数均与 ENC 呈显著负相关, 说明自然选择对 H9N2 AIV 的密码子使用偏好性产生重要影响, 但  $r \neq -1$ , 所以也存在突变压力产生影响。

表 4 H9N2 禽流感病毒基因组的密码子适应指数分析<sup>1)</sup>

Table 4 Codon adaptation index analysis of H9N2 avian influenza virus genomes

年份 Year	有效密码子数 Effective number of codons	密码子适应指数 Codon adaptation index	
		鸡 <i>Gallus gallus</i>	人 <i>Homo sapiens</i>
2010—2012	52.86±0.64	0.768±0.005(-0.731**)	0.740±0.004(-0.583**)
2013—2015		0.766±0.006(-0.386*)	0.739±0.005(-0.223)
2016—2018		0.768±0.004(-0.635**)	0.741±0.003(-0.409*)

1)表中数据为平均值±标准差; 括号中数据为相关系数; “\*”表示 $P<0.05$ , “\*\*”表示 $P<0.01$ (Spearman相关性分析)

1) The datum in the table indicates average ± standard deviation; The datum in the bracket is correlation coefficient; “\*” represents  $P<0.05$ , “\*\*” represents  $P<0.01$ (Spearman’s rank correlation analysis)

2.3.3 突变压力和自然选择对密码子使用偏好性的相对作用比例 由中性绘图分析和 GC<sub>12</sub> 含量与 GC<sub>3</sub> 含量相关性分析(图 4)可知, 2010—2012、2013—2015 和 2016—2018 年的 GC<sub>12</sub> 含量与

GC<sub>3</sub> 含量均呈极显著负相关( $r=-0.720, -0.826, -0.595, P<0.001$ ), 提示密码子的这 3 个位置均受到突变压力的影响。对 2010—2012、2013—2015 和 2016—2018 年的 GC<sub>12</sub> 含量与 GC<sub>3</sub> 含量进行线性

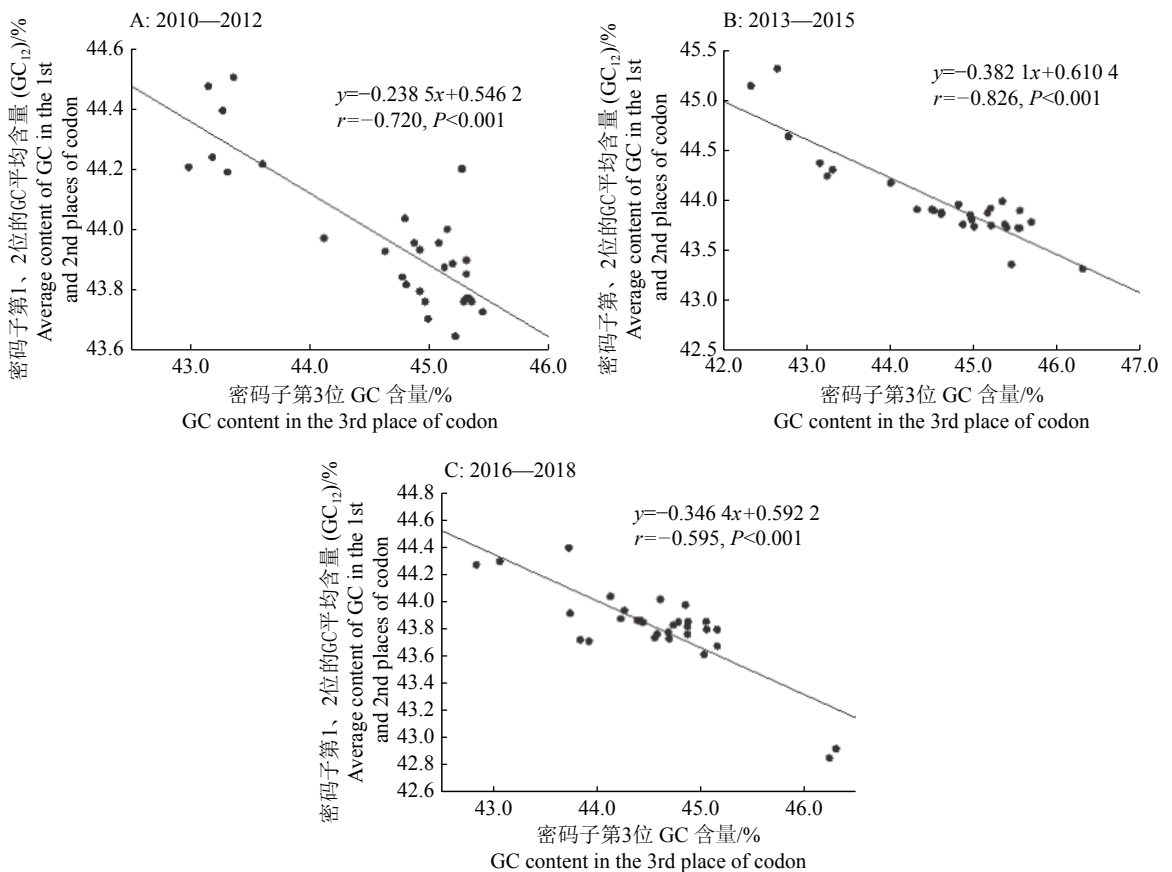


图 4 不同年份 H9N2 禽流感病毒的中性绘图分析

Fig. 4 Neutrality plot analysis of H9N2 avian influenza virus in different years

回归分析发现,斜率分别为-0.238 5、-0.382 1 和 -0.346 4,表明突变压力对 H9N2 AIV 密码子使用偏好性的影响分别为 23.85%(2010—2012 年)、38.21%(2013—2015 年) 和 34.64%(2016—2018 年),所以自然选择对密码子使用偏好性的影响较大,分别为 76.15%(2010—2012 年)、61.79%(2013—2015 年) 和 65.36%(2016—2018 年)。

### 3 讨论与结论

H9N2 AIV 全基因组的碱基组成特性和最优密码子结果显示,基因组中 AU 含量较多且偏好使用以 A 或 U 为末位碱基的同义密码子,这提示密码子使用偏好性可能对 H9N2 AIV 的进化产生影响。ENC 与密码子使用偏好性呈负相关,密码子使用偏好性大小可通过 ENC 值来进行评估,本研究发现 H9N2 AIV 的 ENC 平均值为 52.86,说明不同序列间的基因组组成保守且相对稳定,H9N2 AIV 的密码子使用偏好性较低,这与其他 A 型流感病毒和一些 RNA 病毒相似,如 H3N8(ENC=52.09)<sup>[14]</sup>、H5N1(ENC=50.91)<sup>[15]</sup>、H1N1pdm(ENC=52.5)<sup>[16]</sup>、H3N2(ENC=53.05)<sup>[3]</sup>、寨卡病毒(ENC=53.93)<sup>[17]</sup> 和马尔堡病毒(ENC=54.2)<sup>[5]</sup>。低密码子使用偏好性病毒与宿主的合成竞争较低,可在与其密码子使用模式不相同的宿主细胞内进行有效的复制,从而在宿主体内维持其存活周期<sup>[5, 16]</sup>。密码子使用偏好性可能会受到突变压力、自然选择、基因长度、GC 含量、基因表达水平和 RNA 结构等因素的影响,其中突变压力和自然选择为主要影响因素<sup>[18-19]</sup>。

为了确定突变压力对密码子使用偏好性的作用,对应性分析的第 1、2 轴与碱基组成指数的相关性分析结果显示第 1、2 轴与大部分碱基组成指数呈显著性相关,提示碱基组成对密码子使用偏好性产生影响,即突变压力为 H9N2 AIV 密码子使用偏好性的主要影响因素之一。但当病毒试图适应宿主细胞时,密码子使用偏好性往往会受到自然选择而降低<sup>[19]</sup>。对应性分析的第 1、2 轴与总平均亲水性、芳香性之间均呈现显著相关性,这说明氨基酸的疏水性和芳香性均显著影响着密码子使用偏好性,自然选择作用于 H9N2 AIV 的密码子使用偏好性。PR2 绘图分析发现基因组对 AU 的使用度大于 GC,揭示 H9N2 AIV 密码子使用偏好性可能由突变压力和其他因素(如自然选择)共同决定,但两者作用不相同。综合对应性分析、PR2 绘图和相关性分析结果,可得出结论:突变压力和自然选择为主要

影响因素,共同作用于 H9N2 AIV 密码子使用偏好性;两者作用的相对大小仍为未知。

为了衡量突变压力和自然选择对密码子使用偏好性作用的相对大小,ENC 绘图分析发现所有基因组均聚集于标准曲线的下方,提示突变压力并非 H9N2 AIV 密码子使用偏好性的主导因素,而是其他影响因素(如自然选择)起主导作用。但有报道<sup>[11]</sup>指出突变压力和自然选择可能影响 ENC,使其不能有效评估突变压力和自然选择对密码子使用偏好性的相对作用;此外,基因组为 AU 偏向的物种,其密码子使用偏好性将与基因组为 GC 偏向的存在差异,因此 ENC 绘图分析可能具有误导性。

密码子适应指数是研究自然选择影响密码子使用偏好性的不二选择,可用于评估病毒对宿主的适应性,所以利用 ENC 和密码子适应指数之间的相关性来研究突变压力和自然选择对密码子使用偏好性的影响<sup>[20]</sup>。H9N2 AIV 对鸡和人的密码子适应指数平均值偏高,提示 H9N2 AIV 已适应鸡和人的密码子使用模式,能够更好地利用宿主细胞资源进行病毒复制和繁殖,有利于病毒在宿主体内存活和进化;而且 H9N2 AIV 可能会发展为对人类有致病性的病毒,已有多例 H9N2 AIV 感染人类事件发生<sup>[2, 21]</sup>,随着 H9N2 AIV 逐渐适应人类密码子使用模式,对人类的致病性将逐渐增强,因此需要密切监测密码子适应指数变化。同时,密码子适应指数与 ENC 相关性分析发现大部分密码子适应指数与 ENC 均呈显著负相关且  $r \neq -1$ ,这说明自然选择对密码子使用偏好性的作用大于突变压力,但突变压力的作用也不能忽视。

进行中性绘图分析以揭示两者作用比例,从 GC<sub>12</sub> 含量与 GC<sub>3</sub> 含量的线性回归分析发现,突变压力对 H9N2 禽流感病毒密码子使用偏好性的作用分别为 23.85%(2010—2012 年)、38.21%(2013—2015 年) 和 34.64%(2016—2018 年),而自然选择对密码子使用偏好性的作用较大,分别为 76.15%(2010—2012 年)、61.79%(2013—2015 年) 和 65.36%(2016—2018 年)。中性绘图分析、ENC 绘图分析、密码子适应指数及其相关性结果揭示,在 H9N2 AIV 密码子使用偏好性中,自然选择为主导作用(所占比例为 61.79%~76.15%),而突变压力为次要作用(所占比例为 23.85%~38.21%)。与 H9N2 同为 A 型流感病毒的 H3N2<sup>[3]</sup> 和 H3N8<sup>[22]</sup> 的密码子使用偏好性也是以自然选择为主要影响因素。

综上所述,H9N2 AIV 全基因组为 AU 偏向且最优密码子多以 A 或 U 结尾,而且 H9N2 AIV 基因

组的密码子使用偏好性较低, 主要受到突变压力和自然选择的共同作用, 其中自然选择为主导作用(所占比例为 61.79%~76.15%), 而突变压力为次要作用(所占比例为 23.85%~38.21%)。H9N2 AIV 已适应人类的密码子使用模式, 可能对人类有致病性且致病性在逐渐增强。

#### 参考文献:

- [1] 陈伯伦, 张泽纪, 陈伟斌. 鸡 A 型禽流感病毒的分离与血清学初步鉴定[J]. 中国家禽, 1997(11): 4-6.
- [2] SUN Y, LIU J. H9N2 influenza virus in China: A cause of concern[J]. *Protein Cell*, 2015, 6(1): 18-25.
- [3] LI G, WANG R, ZHANG C, et al. Genetic and evolutionary analysis of emerging H3N2 canine influenza virus[J]. *Emerg Microbes Infect*, 2018: 7. doi: 10.1038/s41426-018-0079-0.
- [4] 胡明达. 禽流感病毒进化与传播规律研究[D]. 北京: 军事科学院, 2018.
- [5] NASRULLAH I, BUTT A M, TAHIR S, et al. Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution[J]. *BMC Evol Biol*, 2015: 15. doi: 10.1186/s12862-015-0456-4.
- [6] BAKER S F, NOGALES A, MARTÍNEZ-SOBRIDO L. Downregulating viral gene expression: Codon usage bias manipulation for the generation of novel influenza A virus vaccines[J]. *Future Virol*, 2015, 10(6): 715-730.
- [7] PUIGBÒ P, BRAVO I G, GARCIA-VALLVE S. CAIcal: A combined set of tools to assess codon usage adaptation[J]. *Biol Direct*, 2008: 3. doi: 10.1186/1745-6150-3-38.
- [8] NAKAMURA Y, GOJOBORI T, IKEMURA T. Codon usage tabulated from international DNA sequence databases: Status for the year 2000[J]. *Nucleic Acids Res*, 2000, 28(1): 292.
- [9] WANG L, XING H, YUAN Y, et al. Genome-wide analysis of codon usage bias in four sequenced cotton species[J]. *PLoS One*, 2018, 13(3): e0194372.
- [10] SHARP P M, LI W H. 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.
- [11] CHEN Y, CHEN Y F. Analysis of synonymous codon usage patterns in duck hepatitis A virus: A comparison on the roles of mutual pressure and natural selection[J]. *Virus Disease*, 2014, 25(3): 285-293.
- [12] CRISTINA J, FAJARDO A, SOÑORA M, et al. A detailed comparative analysis of codon usage bias in Zika virus[J]. *Virus Res*, 2016, 223: 147-152.
- [13] 金刚, 王丽萍, 龙凌云, 等. 普通野生稻线粒体蛋白质编码基因密码子使用偏好性的分析[J]. *植物科学学报*, 2019, 37(2): 188-197.
- [14] KUMAR N, BERA B C, GREENBAUM B D, et al. Revelation of influencing factors in overall codon usage bias of equine influenza viruses[J]. *PLoS One*, 2016, 11(4): e0154376.
- [15] ZHOU T, GU W J, MA J M, et al. Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses[J]. *Biosystems*, 2005, 81(1): 77-86.
- [16] ANHLAN D, GRUNDMANN N, MAKALOWSKI W, et al. Origin of the 1918 pandemic H1N1 influenza A virus as studied by codon usage patterns and phylogenetic analysis[J]. *RNA*, 2011, 17(1): 64-73.
- [17] 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: e107.
- [18] YANG X, LUO X, CAI X. Analysis of codon usage pattern in *Taenia saginata* based on a transcriptome dataset[J]. *Parasite Vector*, 2014: 7. doi: 10.1186/s13701-014-0527-1.
- [19] SHI S L, JIANG Y R, LIU Y Q, et al. Selective pressure dominates the synonymous codon usage in parvoviridae[J]. *Virus Genes*, 2013, 46(1): 10-19.
- [20] VICARIO S, MORIYAMA E N, POWELL J R. Codon usage in twelve species of *Drosophila*[J]. *BMC Evol Biol*, 2007: 7. doi: 10.1186/1471-2148-7-226.
- [21] HUANG Y W, LI X D, ZHANG H, et al. Human infection with an avian influenza A (H9N2) virus in the middle region of China[J]. *J Med Virol*, 2015, 87(10): 1641-1648.
- [22] 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. doi: 10.1186/s12864-017-4063-1.

【责任编辑 李庆玲】