

DOI:10.13350/j.cjpb.240201

• 论著 •

猴痘病毒密码子偏爱性变迁分析^{*}

杨曼清,单静,李亚飞,罗春雨,刘奇^{*,*}

(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 了解猴痘病毒(mpox)密码子使用偏爱性特点,探讨影响其密码子使用偏爱性的主要因素。方法 从NCBI数据库中下载1260条mpox全基因编码序列,选择包括6种突变明显的蛋白编码序列(包括3个毒力基因和3个膜蛋白),运用软件CodonW、EMBOSS、IBM SPSS Statistics 26和SigmaPlot 14.0进行数据分析和绘图分析。结果 mpox 6种蛋白 ENC 取值范围为 40.65~52.24,CAI 均值为 0.64 ± 0.03 。ENC-plot、中性绘图及 PR2-plot 分析表明突变为影响 mpox 密码子偏爱性的主要因素。基于密码子使用偏爱性的聚类分析显示 6 种蛋白聚类相似,其中 2003-2016 毒株和标准株 Zaire-96-I-16 聚为一类,2017-2021 和 2022 至今的毒株聚为一类。**结论** mpox 6 种蛋白密码子主要受突变的影响,其偏爱性有增加趋势。2017 年以后的毒株密码子偏爱性发生较大改变,需要进一步加强监测。

【关键词】 猴痘病毒;密码子偏爱性;进化

【文献标识码】 A

【文章编号】 1673-5234(2024)02-0125-05

[*Journal of Pathogen Biology*. 2024 Feb;19(2):125-129.]

Analysis of codon preference changes in monkeypox virus

YANG Manqing, SHAN Jing, LI Yafei, LUO Chunyu, LIU Qi (Integrated Lab of Pathogenic Biology, Pre-clinical College, Dali University, Dali 671000, Yunnan, China)

【Abstract】 **Objective** To understand the codon usage preference of monkeypox virus (mpox) and explore the main factors affecting the codon usage preference. **Methods** A total of 1 260 full-length mpox gene coding sequences were downloaded from the NCBI database, including 6 protein coding sequences with obvious mutations (including 3 virulence genes and 3 membrane proteins). Software CodonW, EMBOSS, IBM SPSS Statistics 26 and SigmaPlot 14.0 were used for data analysis and plot analysis. **Results** The ENC values of the six mpox proteins ranged from 40.65 to 52.24, and the mean CAI value was 0.64 ± 0.03 . ENC-plot, neutral plot and PR2-plot analysis showed that mutation was the main factor affecting the codon preference of mpox. Clustering analysis based on codon usage preference showed that the six proteins clustered similar. The 2003-2016 strains clustered with standard strain Zaire-96-I-16, and the 2017-2021 and 2022-present strains clustered together. **Conclusion** The codons of the 6 mpox proteins were mainly affected by mutations, and their preference tended to increase. The codon preference of the strains after 2017 has changed greatly, and it is necessary to strengthen the monitoring.

【Key words】 mpox; codon preference; evolution

*** 猴痘病毒(mpox)感染引起人猴痘(human mpox, HMPX),该病自2022年5月以来在多个非地方性流行地区和国家暴发。截至2023年7月,已有113个国家和地区报告了超过8.8万例HMPX实验室确诊病例和150例死亡病例^[1]。国内相继报告了多例猴痘确诊病例,其中,仅2023年6月新增106例确诊病例。

mpox属于痘病毒科正痘病毒属,是一种包膜双链线性DNA病毒,其基因组长度约为197 kb,共编码200多种蛋白^[2-3]。在目前的疫情中,mpox可能主要通过密切接触传播^[4]。mpox目前包括分支I(刚果盆地进化分支)和分支II(西非进化分支)。分支I病毒感染患者临床症状相对较重,死亡率也更高^[5]。2022年疫情的演化支被认为是分支II^[6]。

密码子偏爱性是指生物在长期进化过程中对同义

密码子使用频率有所不同,从而形成的与其自身相适应的一套常用密码子^[7]。密码子偏爱性在外源蛋白的表达和分子进化研究中尤为重要。基于密码子偏爱性设计的基因表达载体可提高mpox基因的表达水平,这对探索物种进化和提高外源基因表达具有重要意义,也能为开发新型抗mpox疫苗提供一定参考。

本研究筛选的6个蛋白编码基因分别为mpox毒力候选基因及经分析有较高变异率的基因,其中B10R、B19R、D10L为西非和刚果盆地分支的毒力候

* 【基金项目】 国家自然科学基金项目(No. 81703573, 81660337);大理大学创新团队项目(No. ZKLX20191055)。

** 【通讯作者】 刘奇, E-mail: Qiliu@aliyun.com

【作者简介】 杨曼清(1998-),女,山西长治人,在读研究生。主要研究方向:感染与免疫。E-mail: 1196862382@qq.com

选基因^[8]; B2R、B21R、G10R 是包膜蛋白,且均在 2022 年 mpox 中出现新变体^[9]。分析各蛋白密码子使用偏爱性可以帮助了解其进化特点,为 mpox 重组疫苗和多表位疫苗的研发提供依据。

材料与方法

1 材料

1.1 基因序列 mpox 蛋白基因编码序列(coding DNA sequence, CDS)下载自 NCBI(<https://www.ncbi.nlm.nih.gov/labs/virus/>)。

1.2 分析工具 Codon W 软件以及 EMBOSS (<https://www.bioinformatics.nl/emboss-explorer>) 子程序 CAI(codon adaptation index) 和 CUSP(Create a codon usage table), 用于对目的基因密码子偏爱性计算; IBM SPSS Statistics 26 和 SigmaPlot 14.0, 用于数据分析和绘图分析。

2 方法

2.1 密码子相对适应度(CAI) CAI 用来衡量单个密码子的相对适应度值^[9], 取值在 0~1 之间, 若该值越接近于 1, 则表示其适应性越强; 反之则越弱。

2.2 有效密码子数(ENC) ENC 指密码子偏离随机选择的程度, 也是衡量同义密码子不均等使用偏好程度的关键性指标, 取值在 20~61 之间, 若该值等于 20, 则表示完全偏倚; 若等于 61, 则表示完全无偏倚^[10]。

2.3 相对同义密码子使用度(RSCU) RSCU 用于检测基因中所有同义密码子使用模式的变化。若该值小于 1, 表明该密码子的实际使用频率小于其他同义密码子; 若该值大于 1, 表示该密码子的实际使用频率大于其他同义密码子; 若该值等于 1, 表示该密码子没有偏爱性^[11]。

2.4 ENC-plot 优先密码数绘图分析 ENC-plot 用来分析 ENC 与 GC3 之间的分布关系。若基因位于标准曲线附近则表明该基因的密码子偏爱性主要受到突变的影响; 若基因位于标准曲线较远的位置, 则表明该基因密码子偏爱性主要受自然选择的影响^[12]。

2.5 中性绘图分析 中性绘图分析是通过对密码子 1、2 位和第 3 位碱基的相关性进行分析, 用于研究对密码子偏爱性产生直接影响的主要因素^[13]。以 GC12 为纵坐标, GC3 为横坐标绘图, 回归系数越接近于 0, 说明 GC12 与 GC3 相关性不显著, 1、2、3 位碱基组成不同, 密码子使用完全受自然选择的影响; 回归系数等于 1, 密码子使用完全受突变因素的影响。

2.6 PR2-plot 绘图分析 PR2 偏倚分析主要是为了

避免密码子的第 3 位碱基 A 与 T 以及 C 与 G 之间的线性突变不平衡。以 A3/(A3+T3) 为纵坐标, G3/(G3+C3) 为横坐标绘制散点图, 以 A=T 且 G=C 为中点, 各散点代表其基因方向和偏倚程度^[14]。

2.7 对应分析 对应分析(Correspondence analysis, CA) 用于分析同义密码子使用偏爱性产生的主要原因。运用 CodonW 进行基于 RSCU 值的对应分析, 将每个样本中所有基因区别分布在一个 59 维的向量空间中, 每个点代表一个同义密码子, 通过分析前两轴的相关性解释密码子使用的偏爱性^[15]。

2.8 聚类分析 基于密码子偏爱性的聚类分析将 mpox 各蛋白绘制树状图, 分析各蛋白特征相似性^[16-17]。

结 果

1 ENC、CAI 分析

mpox 6 个蛋白的 ENC 分析如图 1, 各蛋白 ENC 分布范围为 40.65~52.24(45.81±2.89), 均大于 35, 表明 mpox 6 个蛋白基因组密码子使用偏爱性均较弱。对各蛋白 ENC 值作进一步分析, 2022 年至今 B10R ($t = 33.269.178, P < 0.05$) 和 B19R ($t = 92.406.804, P < 0.05$) 2 种蛋白 ENC 值较标准株 Zaire-96-I-16 大, 说明其密码子偏性减弱; 而 B2R ($t = 30.794.288, P < 0.05$)、B21R ($t = 162.100.768, P < 0.05$)、G10R ($t = 38.650.573, P < 0.05$) 和 D10L ($t = 13.364.984, P < 0.05$) 4 种蛋白 ENC 值相较标准株 Zaire-96-I-16 减小, 表明其密码子偏性增强。

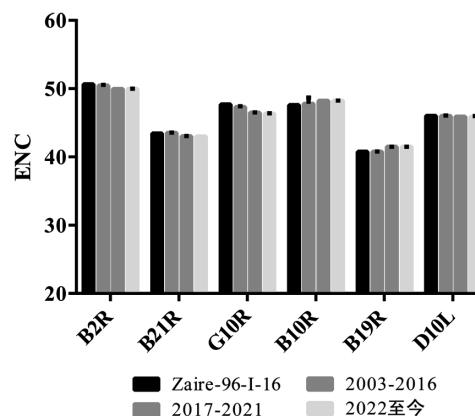
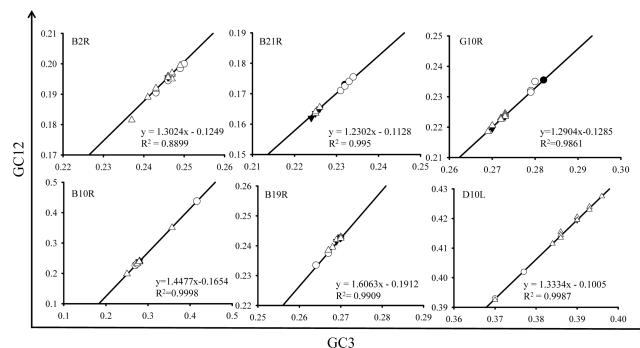


图 1 mpox 6 个蛋白基因组 ENC 值比较
Fig. 1 Comparison of ENC values for the six protein genomes of mpox

2 RSCU 分析

RSCU 分析如表 1。2022 年至今 B2R、B21R、B19R、D10L 分别有 11、29、7、6 个密码子较 Zaire-96-I-16 株 RSCU 值增大, 10、22、6、3 个密码子较 Zaire-96-I-16 株 RSCU 值小, 其密码子使用偏爱性整体上有增强的趋势, 与 ENC 分析结果一致。

明 mpox6 种蛋白密码子偏爱性主要受到突变因素的影响,与中性绘图分析结果一致。

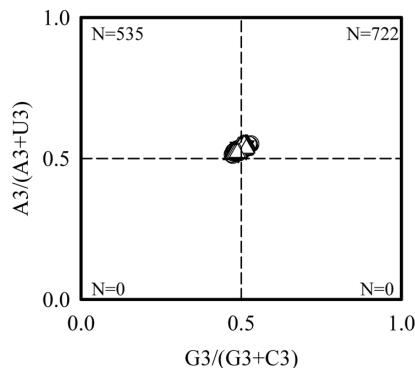


注:●表示 Zaire-96-I-16 株;○表示 2003-2016 毒株;▼表示 2017-2021 毒株;△表示 2022 年以后的毒株。

图 3 mpox 6 个蛋白中性绘图分析

Notes: ● denotes Zaire-96-I-16; ○ denotes the 2003-2016; ▼ Denotes 2017-2021; △ Indicates the strains from 2022 onwards.

Fig. 3 Analysis of neutral plots of six mpox proteins



注:●表示 Zaire-96-I-16 株;○表示 2003-2016 毒株;▼表示 2017-2021 毒株;△表示 2022 年以后的毒株。

图 4 mpox 基因 PR2-plot 绘图分析

Notes: ● denotes Zaire-96-I-16; ○ denotes the 2003-2016; ▼ Denotes 2017-2021; △ Indicates the strains from 2022 onwards.

Fig. 4 PR2-plot analysis of the mpox gene

6 对应分析

运用 CodonW 对不同年份 mpox 全基因组进行基于 RSCU 值的对应分析,结果如图 5,第一轴(Aixs 1)、第二轴(Aixs 2)分别解释了 44.10% 和 32.67% 的变异。各基因组分散分布,提示彼此存在较大差异,各基因组密码子使用偏爱性差异较大。参照吴炜倩等^[18]的方法将 Aixs 1 相应值与 CAI、CBI、Fop、ENC 和 GC3s 进行相关性分析(Pearson 相关分析),结果显示 Aixs 1 与 GC3s 和 ENC 相关(r 值分别为 0.556 和 0.590,均 $P < 0.01$),与 CAIP、CBI 和 Fop 相关(r 值分别为 0.952,0.961,0.937,均 $P < 0.01$),说明碱基组成是影响 mpox 密码子使用模式的主要因素。

7 聚类分析

基于密码子偏爱性对 mpox 6 种蛋白进行聚类分析,结果如图 6,6 种蛋白上的密码子偏爱性聚类分析

结果一致。其中 2003-2016 流行株和标准株 Zaire-96-I-16 聚为一类,2022 年以后的流行株与 2017-2021 流行株聚为一类。说明 2022 年疫情的 mpox 演化支与 2017-2021 年关系更为密切。

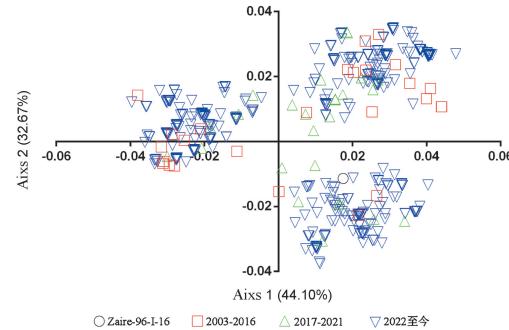


图 5 mpox 不同年份全基因组基于 RSCU 值的对应分析

Fig. 5 Correspondence analysis based on RSCU values for the whole genome of mpox in different years

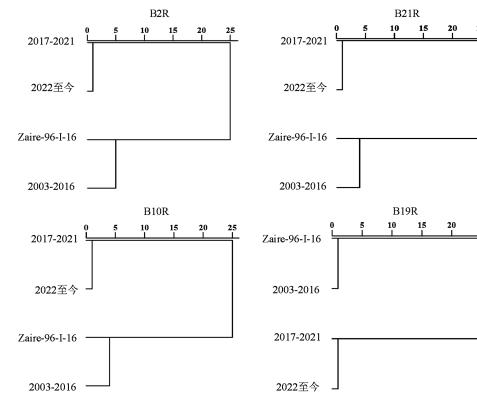


图 6 不同年份 mpox 6 种蛋白聚类分析

Fig. 6 Cluster analysis of mpox6 proteins in different years

讨 论

mpox 为包膜双链线性 DNA 病毒,其基因组长度巨大,约为 197 kb,共编码 200 多种蛋白^[2-3]。病毒基因组序列长度和基因含量与广谱寄主范围呈正相关,与致病性呈负相关^[18]。研究认为,mpox 刚果盆地进化支较西非进化支具有更大的基因组和含量,这可能是刚果盆地进化支毒力较低的原因^[19]。与 mpox 同为正痘病毒属的天花病毒(Variola Virus, VARV)其每年每个基因组仅出现 1~2 个核苷酸突变^[20],而 2022 年暴发的 mpox,该数字则增加到每个基因组有 46 个单核苷酸多态性位点(single-nucleotide polymorphisms, SNPs)^[21],为 2018-2019 流行株的 6~12 倍,这些变化可能与突变和病毒适应能力变化相关^[22]。研究认为,mpox 的突变现象不仅会加快了其进化速度,而且使病毒的传播能力增强,也增加了猴痘疫苗的研发难度^[23]。

本研究筛选了 mpox 变异较大的 6 个基因,包括 3

个毒力候选基因(B10R、B19R、D10L)和3个包膜蛋白(B2R、B21R、G10R)。相关分析显示,mpox6种蛋白ENC值均大于35,说明mpox密码子使用偏爱性较低。RSCU分析显示,新发(2022年以后的毒株)mpox6个蛋白总体密码子偏爱性较标准株Zaire-96-I-16增加,这可能进一步减弱mpox在宿主体内的复制能力。

另外,自然选择和碱基突变等因素共同作用导致同义密码子使用偏爱性产生,突变作用和正向选择作用越强,产生的高频密码子越多;反之,选择作用越强,高频密码子越少^[24]。本研究中ENC-plot、中性绘图及PR2-plot分析结果均提示mpox6种蛋白密码子使用模式主要受突变选择作用影响,与mpox ITR(反向末端区域)的同义密码子使用偏爱性分析^[24]相一致。

基于6种蛋白密码子偏爱性的聚类分析显示,2003-2016流行株和标准株Zaire-96-I-16聚为一类,提示早期流行株与标准株接近;而2022年以后的流行株与2017-2021流行株聚为一类,说明2022年疫情的mpox演化支与2017-2021流行株关系更为密切。同时也提示2017年以后mpox发生了较大改变。鉴于牛痘病毒前三代疫苗(Dryvax、ACAM2000、MVA-BN)的安全性、产生的副作用及对进化支的交叉免疫效果都不尽如人意^[25-28],因此仍需进一步加强猴痘疫苗的研发。

【参考文献】

- [1] 2022-23 Mpoxy (Monkeypox) Outbreak: Global Trends [EB/OL]. [2023-04-29]. https://worldhealth.org.shinyapps.io/mpx_global/.
- [2] Alakunle E, Moens U, Nchinda G, et al. Monkeypox virus in Nigeria: infection biology, epidemiology, and evolution [J]. Viruses, 2020, 12(11):1257.
- [3] 宁玲,李磊. 人类猴痘的研究进展[J]. 安徽医学,2022,43(7):851-855.
- [4] Al-Mandhari A, Kodama C, Abubakar A, et al. Monkeypox outbreak and response efforts in the Eastern Mediterranean Region[J]. East Med Health J, 2022, 28(7):465-468.
- [5] Adler H, Gould S, Hine P, et al. Clinical features and management of human monkeypox: a retrospective observational study in the UK[J]. Lancet Infect Dis, 2022, 22(8):1153-1162.
- [6] Saxena SK, Ansari S, Maurya VK, et al. Re-emerging human monkeypox: a major public - health debacle[J]. J Med Virol, 2023, 95(1):e27902.
- [7] 吴炜倩,胡源,胡接力,等. 乙型肝炎病毒密码子偏好性及影响因素分析[J]. 中国病原生物学杂志,2020,15(12):1404-1410.
- [8] Chen N, Li G, Liszewski MK, et al. Virulence differences between monkeypox virus isolates from West Africa and the Congo basin [J]. Virology, 2005, 340(1):46-63.
- [9] GISAID Initiative [EB/OL]. [2023-05-06]. <https://www.epicov.org/epi3/frontend/#16f71b>.
- [10] Butt AM, Nasrullah I, Qamar R, et al. Evolution of codon usage in Zika virus genomes is host and vector specific [J]. Environ Microbes Infect, 2016, 5(1):1-14.
- [11] Karumathil S, Raveendran NT, Ganesh D, et al. Evolution of synonymous codon usage bias in West African and Central African strains of monkeypox virus [J]. Evolution Bioinformatics, 2018, 14:1610507672.
- [12] Zhang Y, Nie X, Jia X, et al. Analysis of codon usage patterns of the chloroplast genomes in the Poaceae family [J]. Australian J Botany, 2012, 60(5):461-470.
- [13] Wei L, He J, Jia X, et al. Analysis of codon usage bias of mitochondrial genome in Bombyx mori and its relation to evolution [J]. BMC Evolution Biol, 2014, 14:1-12.
- [14] He W, Wang N, Tan J, et al. Comprehensive codon usage analysis of porcine deltacoronavirus [J]. Molecular Phylogenetic Evolution, 2019, 141:106618.
- [15] 冉然,张晓明,宛涛,等. 密码子偏好的相关研究及方法分析[J]. 草原与草业,2022,34(3):5-10.
- [16] 尚方建,刘熙,田明明,等. 人与骆驼来源中东呼吸综合征冠状病毒密码子偏爱性分析[J]. 中国病原生物学杂志,2020,15(11):1246-1251.
- [17] 李亚飞,罗春雨,石哲芳,等. 新型冠状病毒Omicron变异株密码子偏爱性及其进化分析[J]. 中国人兽共患病学报,2022,38(7):559-565.
- [18] Hendrickson RC, Wang C, Hatcher EL, et al. Orthopoxvirus genome evolution: the role of gene loss [J]. Viruses, 2010, 2(9):1933-1967.
- [19] Meyer H, Sutter G, Mayr A. Mapping of deletions in the genome of the highly attenuated vaccinia virus MVA and their influence on virulence [J]. J General Virol, 1991, 72(5):1031-1038.
- [20] Firth C, Kitchen A, Shapiro B, et al. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses [J]. Molecular Biol Evolution, 2010, 27(9):2038-2051.
- [21] Hatmal MMM, Al-Hatamleh MAI, Olaimat AN, et al. Comprehensive literature review of monkeypox [J]. Emerging microbes & infections, 2022, 11(1).
- [22] Bailey SF, Alonso Morales LA, Kassen R. Effects of synonymous mutations beyond codon bias: the evidence for adaptive synonymous substitutions from microbial evolution experiments [J]. Genome Biol Evolution, 2021, 13(9):b141.
- [23] 苏琳佳,许方婧伟,张云涛. 猴痘病毒疫苗的研究进展[J]. 微生物学免疫学进展,2022,50(5):1-5.
- [24] Tang D, Wei F, Cai Z, et al. Analysis of codon usage bias and evolution in the chloroplast genome of Mesona chinensis Benth [J]. Development Gene Evolution, 2021, 231:1-9.
- [25] Karumathil S, Raveendran NT, Ganesh D, et al. Evolution of synonymous codon usage bias in West African and Central African strains of monkeypox virus [J]. Evolutionary Bioinformatics, 2018, 14:1610507672.
- [26] Fenner F, Henderson DA, Arita I, et al. Smallpox and its eradication [M]. World Health Organization Geneva, 1988.
- [27] Petersen E, Zumla A, Hui DS, et al. Vaccination for monkeypox prevention in persons with high-risk sexual behaviours to control on-going outbreak of monkeypox virus clade 3 [J]. Internat J Infect Dis, 2022, 122:569-571.
- [28] Ahmed SF, Sohail MS, Quadeer AA, et al. Vaccinia-virus-based vaccines are expected to elicit highly cross-reactive immunity to the 2022 monkeypox virus [J]. Viruses, 2022, 14(9):1960.