

# 昆虫质型多角体病毒同义密码子使用模式分析\*

石生林<sup>1,2,3</sup> 姜义仁<sup>2,3</sup> 杨瑞生<sup>2,3</sup> 刘彦群<sup>2,3</sup> 赵金莹<sup>2,3</sup> 秦利<sup>1,2,3\*\*</sup>

(1. 沈阳农业大学植物保护博士后流动站 沈阳 110866; 2. 沈阳农业大学生物科学技术学院  
沈阳 110866; 3. 辽宁省昆虫资源工程技术研究中心 沈阳 110866)

**摘要** 昆虫质型多角体病毒(cypovirus, CPV)是害虫种群重要调节因子,可用作生物防治剂。本研究采用多元统计分析方法对7种CPV进行密码子使用模式分析,结果表明:CPV密码子使用偏好性较弱,多数基因密码子使用模式受碱基组成影响,少数基因密码子使用模式除碱基组成外还有其它影响因素;中性绘图分析表明碱基组成主要受选择压力影响,受突变影响较小。同一电泳型CPV之间比同一宿主CPV之间共有的偏好性密码子多。CPV基因组内10个基因组片段之间密码子偏好性存在差异。CPV密码子偏好性与宿主昆虫密码子偏好性存在差异,所有CPV与其宿主昆虫共有的偏好性密码子均较少。对应分析进一步证明碱基组成是影响密码子使用的主要因素,不同电泳型CPV具有不同的密码子使用模式。聚类分析表明同一电泳型CPV密码子使用模式相似,同一宿主CPV密码子使用模式差异较大。

**关键词** 有效密码子数, ENc-plot 分析, 中性绘图分析, 对应分析, 偏好性密码子

## Analysis of synonymous codon usage patterns in insect cypoviruses

SHI Sheng-Lin<sup>1,2,3</sup> JIANG Yi-Ren<sup>2,3</sup> YANG Rui-Sheng<sup>2,3</sup> LIU Yan-Qun<sup>2,3</sup>  
ZHAO Jin-Ying<sup>2,3</sup> QIN Li<sup>1,2,3\*\*</sup>

(1. Postdoctoral Station of Plant Protection, Shenyang 110866, China; 2. College of Bioscience and Biotechnology, Shenyang Agricultural University, Shenyang 110866, China; 3. Insect Resource Engineering Research Center of Liaoning Province, Shenyang 110866, China)

**Abstract** Cypoviruses (CPVs) are key regulatory factors of pest insect populations and are considered to be very promising biological control agents. We detailed the synonymous codon usage patterns of seven CPVs through multivariate statistical methods. Our results show that CPVs possess a generally weak synonymous usage bias. The codon usage patterns of most genes were constrained by base composition, while those of a small minority of genes were influenced by both base composition and other factors. Neutrality plot analysis revealed that selective pressures dominated over mutation in shaping base composition. CPVs with the same electrophoretype shared more common preferred codons than CPVs from the same host. The 10 genome fragments within CPVs possessed statistically different codon biases. The codon biases of CPVs were significantly different from those of their insect host. All CPVs shared a few common preferred codons with their host. Correspondence analysis further confirmed that base composition was the main factor influencing synonymous codon usage and that different electropherotype CPVs possessed different codon usage patterns. Cluster analysis indicated that same-electropherotype CPVs possessed similar codon usage patterns, but same-host CPVs possessed dissimilar codon usage patterns.

**Key words** effective number of codons, ENc-plot analysis, neutrality plot analysis, correspondence analysis, preferred codons

呼肠病毒科(Reoviridae)、质型多角体病毒属( *Cypovirus* )的质型多角体病毒( *Cypoviruses*, CPV )只感染节肢动物并具有致病作用( King *et al.* , 2012)。质型多角体病毒基因组一般包括10个

\* 资助项目:现代农业产业技术体系专项(CARS-22)。  
\*\*通讯作者, E-mail: qinli1963@163.com  
收稿日期:2012-08-04, 接受日期:2012-12-22

dsRNA 基因组片段,可根据病毒电泳型和病毒原始宿主种类进行命名,其电泳型模式可分为 16 种[尚有一些未确定种](King *et al.*, 2012)。昆虫质型多角体病毒是昆虫重要病原,在家蚕 *Bombyx mori* L. 饲养中危害严重,对农林害虫具有生物防治潜力。

作为适应性进化结果(Goodarzi *et al.*, 2008),基因组内和基因组间同义密码子使用是非均衡的(Ingvarsson, 2008),称为同义密码子使用偏好性,也称密码子使用偏好性或密码子偏好性。密码子偏好性存在于从病毒到哺乳动物的广泛生物系统中(Eyre-Walker, 1991; Jiang *et al.*, 2008),一般认为是选择高效利用核糖体、遗传漂移和碱基突变的结果(Shah and Gilchrist, 2011)。不同病毒具有不同的密码子使用模式且与宿主密码子使用模式具有相关性,反映病毒的种属特异性以及对宿主的适应策略(Bahir *et al.*, 2009; Wang *et al.*, 2011b)。通过改变生物体密码子可在小至核苷酸(nucleotide)大至兆碱基(megabase)尺度上重建基因组(Isaacs *et al.*, 2011),研究密码子偏好性是研究分子进化和生物技术的核心(Plotkin and Kudla, 2011)。基于全基因组序列进行密码子使用模式比较分析是病毒分子进化研究的重要内容(李京生等, 2011; 石生林等, 2012)。目前 GenBank 中登录了 7 种昆虫质型多角体病毒全基因组序列:家蚕质型多角体病毒-1 (*Bombyx mori* cypovirus 1 [Suzhou strain], BmCPV-1)、马尾松毛虫质型多角体病毒-1 (*Dendrolimus punctatus* cypovirus 1, DpCPV-1)、舞毒蛾质型多角体病毒-1 (*Lymantria dispar* cypovirus 1, LdCPV-1)、棉铃虫质型多角体病毒-5 (*Heliothis armigera* cypovirus 5, HaCPV-5)、舞毒蛾质型多角体病毒-14 (*Lymantria dispar* cypovirus 14, LdCPV-14)、棉铃虫质型多角体病毒-14 (*Heliothis armigera* cypovirus 14, HaCPV-14)、粉纹夜蛾质型多角体病毒-15 (*Trichoplusia ni* cypovirus 15, TnCPV-15),为进行不同病毒基因组间及不同基因组片段间密码子使用模式比较分析奠定了基础。

本研究采用多元统计分析方法,研究了 7 种昆虫质型多角体病毒全基因组密码子使用模式及其与宿主昆虫密码子使用模式的相关性,结果表明昆虫质型多角体病毒密码子偏好性较弱且与宿主密码子偏好性不同,为进一步研究昆虫质型多

角体病毒遗传进化及与宿主相互关系奠定了基础。

## 1 材料与方法

### 1.1 数据来源及处理

昆虫质型多角体病毒基因组序列: BmCPV-1 (GU323605、GQ924586、GQ924587、GU323606、GQ294468、GQ294469、GQ150538、GQ150539、GQ924588、GQ924589); DpCPV-1 (AY163247、AY147187、AY167578、AF542082、AY163248、AY163249、AY211091、AY211092、AY211093、AY211094); LdCPV-1 (NC\_003016、NC\_003017、NC\_003018、NC\_003019、NC\_003020、NC\_003021、NC\_003022、NC\_003023、NC\_003024、NC\_003025); HaCPV-5 (NC\_010670、NC\_010669、NC\_010668、NC\_010667、NC\_010666、NC\_010665、NC\_010664、NC\_010663、NC\_010662、NC\_010661); LdCPV-14 (NC\_003006、NC\_003007、NC\_003008、NC\_003009、NC\_003010、NC\_003011、NC\_003012、NC\_003013、NC\_003014、NC\_003015); HaCPV-14 (DQ242048、DQ388474、DQ388475、DQ388476、DQ388477、DQ017080、DQ010326、DQ010325、DQ010324、DQ010323); TnCPV-15 (NC\_002557、NC\_002558、NC\_002559、NC\_002567、NC\_002560、NC\_002561、NC\_002562、NC\_002563、NC\_002564、NC\_002565) 下载自 GenBank 核酸数据库和基因组数据库(截止至 2012 年 4 月 28 日)。GenBank 中 TnCPV-15 基因组片段 1~9 无编码序列(coding sequence, CDS)注释信息,本研究中通过 NCBI 在线 ORF Finder 程序(<http://www.ncbi.nlm.nih.gov/projects/gorf/>)鉴定并选取最大 ORF 作为 CDS 进行分析; LdCPV-14 基因组片段 10 注释了 2 个 CDS,为与其它病毒基因组片段 10 相对应,本研究中只选择较大的 1 个 CDS 进行分析; HaCPV-5 基因组片段 1 CDS 中有 6 个非 A/C/G/T 碱基,本研究将这 6 个碱基组成的密码子删除后再用于分析。

家蚕、马尾松毛虫 *Dendrolimus punctatus* Walker、舞毒蛾 *Lymantria dispar* L.、棉铃虫 *Heliothis armigera* Hübner、粉纹夜蛾 *Trichoplusia ni* Hübner 的密码子使用频率数据下载自密码子使用数据库(Codon Usage Database, Flat File Release 160.0 [June 15, 2007])(Nakamura *et al.*, 2000)。

## 1.2 密码子使用模式分析

有效密码子数 (effective number of codons, ENc) 常用于度量同义密码子使用偏好性,  $ENc \leq 35$  时认为密码子使用具有显著偏好性,  $ENc \geq 50$  时认为密码子使用偏好性较低 (Wright, 1990)。7 种昆虫质型多角体病毒基因组 ENc 及 GC 总含量 (GC)、同义密码子第 3 位 GC 含量 (GC3s)、相对同义密码子使用度 (relative synonymous codon usage, RSCU) 的计算、基于 RSCU 值的对应分析 (correspondence analysis, CA) 均利用 CodonW 软件完成。根据密码子使用数据库提供的密码子使用频率, 按照文献 (Wright, 1990) 提供的公式计算 5 种宿主昆虫 ENc 值, 利用 VisualGeneDeveloper 软件 (Jung and McDonald, 2011) 计算 5 种宿主昆虫 RSCU 值。每种氨基酸中 RSCU 值最高的密码子定义为偏好性密码子。

ENc-plot 分析中, 被分析的基因或基因组密码子使用模式如果由碱基组成决定, 则数据点会落在预期曲线上或沿曲线分布 (Wright, 1990); 中性绘图分析中,  $P_1$ 、 $P_2$  和  $P_3$  分别是密码子第 1 位、第 2 位和第 3 位的 GC 含量 ( $P_{12}$  是  $P_1$  和  $P_2$  均值, 计算时不包括密码子 ATG、TGG、ATA、TAA、TAG 和 TGA),  $P_{12}$  对  $P_3$  的回归系数既是突变-选择平衡系数 (Sueoka, 1988, 1999); 聚类分析采用离差平方和法 (Ward's method), 间距为欧氏距离平方 (Squared Euclidean distance)。相关分析采用 Spearman 等级相关系数, 2 个相关样本间差异显著性分析采用 Wilcoxon 符号秩法检验, 多个相关样本间差异显著性分析采用 Friedman 法, 各项分析均应用 SPSS 软件完成。

## 2 结果与分析

### 2.1 ENc 值及 GC 含量分析

7 个 CPV 基因组共 70 个基因的 ENc 值变化范围为: 42.30 ~ 59.89, 均值为  $51.35 \pm 4.06$ ; 所有基因 ENc 值均大于 35, 46 个基因具有较高的 ENc 值 (50 ~ 61), 占基因总数的 65.71%, 说明质型多角体病毒同义密码子使用偏好性不显著。70 个基因的 GC 含量变化范围为: 34.4% ~ 49.8%, 均值为  $40.82 \pm 3.96$ ; GC3s 含量变化范围为: 23.3% ~ 57.7%, 均值为  $35.84 \pm 7.90$ , GC 含量与 GC3s 含量相关性有统计学意义 ( $\rho = 0.912, P < 0.001$ )。ENc 值与 GC 含量和 GC3s 含量相关性均有统计学

意义, 相关系数分别为  $\rho = 0.690 (P < 0.001)$  和  $\rho = 0.698 (P < 0.001)$ 。

对不同电泳型 CPV 之间 ENc 值和 GC 含量、GC3s 含量进行比较 (表 1), CPV-5 (HaCPV-5) 与 CPV-15 (TnCPV-15) 之间 3 项指标差异均无统计学意义 ( $P > 0.05$ ), 其它电泳型 CPV 两两之间差异均有统计学意义 ( $P < 0.05$ )。进一步分析表明, 3 种 CPV-1 之间除 BmCPV-1 与 LdCPV-1 间 ENc 值差异有统计学意义外 ( $P = 0.047$ ), 其它均无统计学意义; 2 种 CPV-14 之间 GC 含量差异无统计学意义 ( $P = 0.134$ ), ENc 值和 GC3s 含量差异有统计学意义 ( $P = 0.044, P = 0.009$ ); 宿主相同的 HaCPV-5 与 HaCPV-14 之间及 LdCPV-1 与 LdCPV-14 之间 ENc 值、GC 含量、GC3s 含量差异均有统计学意义 ( $P < 0.05$ )。这些结果说明不同电泳型 CPV 密码子使用模式不同。

CPV 基因组 10 个基因组片段分析结果表明, 10 个基因组片段之间 ENc 值、GC 含量和 GC3s 含量差异均有统计学意义, 显著水平分别为:  $P = 0.012, P = 0.001, P < 0.001$ , 说明 CPV 基因组内密码子使用模式存在差异, 且与编码基因功能有相关性。

以宿主 ENc 值为总体均数、CPV 基因组片段 ENc 值为样本进行单样本  $t$  检验的结果表明, BmCPV-1 与 *B. mori* (ENc = 58.97) [ $P < 0.001$ ]、DpCPV-1 与 *D. punctatus* (ENc = 51.79) [ $P = 0.002$ ]、LdCPV-14 与 *L. dispar* (ENc = 55.35) [ $P < 0.001$ ]、HaCPV-5 和 HaCPV-14 与 *H. armigera* (ENc = 55.58) [ $P < 0.001$ ]、TnCPV-15 与 *T. ni* (ENc = 54.96) [ $P < 0.001$ ] 之间的差异均有统计学意义, 而 LdCPV-1 与 *L. dispar* (ENc = 55.35) 间的差异无统计学意义 ( $P = 0.985$ ), 说明质型多角体病毒密码子偏好性与宿主密码子偏好性不同。

### 2.2 同义密码子使用分析

通过分析 7 种 CPV 同义密码子使用偏好性, 鉴定了除 Met、Trp 外 18 种氨基酸的偏好性密码子 (表 2)。7 种 CPV 之间有 8 种氨基酸的偏好性密码子相同: AGA (Arg)、AAU (Asn)、GAU (Asp)、UGU (Cys)、CAA (Gln)、CAU (His)、UUA (Leu)、UAU (Tyr), 这些密码子均以 A/U 结尾。3 种 CPV-1 之间除 Glu 外, 其它 17 种氨基酸的偏好性密码子均相同, 而偏好性不同的 Glu 的 2 个密码子

表 1 不同电泳型 CPV 之间 ENc 及 GC 含量的差异显著性

Table 1 Significance level of ENc and GC content between different electropherotype cypoviruses

| 基因组<br>Genome                       | 有效密码子数显著性<br>Significance level of ENc | GC 含量显著性<br>Significance level of GC( % ) | 同义密码子第 3 位 GC 含量显著性<br>Significance level of GC3s( % ) |
|-------------------------------------|----------------------------------------|-------------------------------------------|--------------------------------------------------------|
| CPV-1 与 CPV-5<br>CPV-1 vs CPV-5     | $P = 0.005$                            | $P = 0.007$                               | $P = 0.005$                                            |
| CPV-1 与 CPV-14<br>CPV-1 vs CPV-14   | $P = 0.007$                            | $P = 0.025$                               | $P = 0.012$                                            |
| CPV-1 与 CPV-15<br>CPV-1 vs CPV-15   | $P = 0.007$                            | $P = 0.005$                               | $P = 0.005$                                            |
| CPV-5 与 CPV-14<br>CPV-5 vs CPV-14   | $P = 0.007$                            | $P = 0.005$                               | $P = 0.005$                                            |
| CPV-5 与 CPV-15<br>CPV-5 vs CPV-15   | $P = 0.074$                            | $P = 0.838$                               | $P = 0.169$                                            |
| CPV-14 与 CPV-15<br>CPV-14 vs CPV-15 | $P = 0.047$                            | $P = 0.007$                               | $P = 0.005$                                            |

RSCU 值差异亦较小;2 种 CPV-14 之间 18 种氨基酸的偏好性密码子均相同;宿主相同的 HaCPV-5 和 HaCPV-14 之间有 15 种氨基酸的偏好性密码子相同,LdCPV-1 和 LdCPV-14 之间有 13 种氨基酸的偏好性密码子相同。

结合 5 种宿主昆虫 RSCU 值,鉴定了各宿主昆虫与对应病毒间共有的偏好性密码子(表 2)。各病毒与其宿主共有的偏好性密码子均较少,进一步说明昆虫质型多角体病毒密码子偏好性与宿主昆虫密码子偏好性存在差异。

2.3 ENc-plot 及中性绘图分析

ENc-plot 分析中多数代表各个基因的数据点都落在期望曲线下方并沿曲线分布(图 1),表明多数基因密码子使用模式受碱基组成影响,少数基因密码子使用模式除受碱基组成影响外还受其它因素影响。中性绘图分析中(图 2), $P_{12}$ 对  $P_3$  的回归系数为  $0.164 \pm 0.039$  ( $P < 0.001$ ,  $R^2 = 0.209$ ),表明突变对  $P_3$  的影响为 16.4%,而选择约束对  $P_3$  的影响为 83.6%,即影响碱基组成的因素中选择压力的作用大于突变。

2.4 基于 RSCU 值的对应分析

基于 RSCU 值的对应分析中,第 1 轴和第 2 轴对变异的贡献率分别为 30.71% 和 11.20%,前 13 个轴对变异的贡献率为 85.66%,因此第 1 轴是描

述昆虫质型多角体病毒密码子使用差异的主要轴向,其代表的影响因素就是最能解释密码子偏好性的因素。各基因的 GC 含量与其在第 1 轴上位置的相关性有统计学意义 ( $\rho = -0.873$ ,  $P < 0.001$ ),GC3s 含量与其在第 1 轴上位置的相关性也有统计学意义( $\rho = -0.906$ ,  $P < 0.001$ ),说明编码序列碱基组成是影响密码子偏好性的主要因素。不同电泳型 CPV 之间各基因点在第 1 轴和第 2 轴上分布不同,相同电泳型 CPV 之间各基因点在第 1 轴和第 2 轴上分布类似(图 3),说明不同电泳型 CPV 具有不同的密码子使用模式。

同义密码子在两轴上的对应分布将多数 C/G 结尾密码子和 A/U 结尾密码子沿第 1 轴截然分开,少数 C/G 结尾密码子和 A/U 结尾密码子在第一轴上分布存在交叉(图 4),进一步说明多数基因同义密码子使用模式源于碱基组成,少数基因同义密码子使用模式除受碱基组成影响外,还有其它影响因素。

2.5 聚类分析

根据 RSCU 值对 7 种昆虫质型多角体病毒进行聚类分析(图 5),2 种 CPV-14 及 3 种 CPV-1 各自被聚为 1 类,同一电泳型 CPV 之间距离较近;不同电泳型 CPV 之间距离较远;同宿主不同电泳型 CPV 之间距离亦较远。聚类分析进一步表明,不同电泳型 CPV 的密码子使用模式不同。

表 2 昆虫质型多角体病毒偏好性密码子  
Table 2 Preferred codons in insect cypoviruses

| 氨基酸<br>Amino acid | 密码子<br>Codon | BmCPV-1     | DpCPV-1     | LdCPV-1     | HaCPV-5     | HaCPV-14    | LdCPV-14    | TnCPV-15    |
|-------------------|--------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Ala               | GCA          | <b>1.65</b> | <b>1.44</b> | <b>1.54</b> | <b>1.47</b> | <b>1.33</b> | <b>1.26</b> | 1.03        |
|                   | GCC          | 0.37        | 0.38        | 0.41        | 0.29        | 0.36        | 0.38        | 0.42        |
|                   | GCG          | 1.00        | 1.09        | 1.02        | 1.14        | 1.11        | 1.12        | 0.75        |
|                   | GCU          | 0.97        | 1.09        | 1.03        | 1.10        | 1.20        | 1.24        | <b>1.80</b> |
| Arg               | AGA          | <b>1.89</b> | <b>1.68</b> | <b>1.59</b> | <b>3.51</b> | <b>2.05</b> | <b>2.02</b> | <b>2.16</b> |
|                   | AGG          | 0.55        | 0.67        | 0.69        | 0.91        | 0.63        | 0.66        | 0.55        |
|                   | CGA          | 0.92        | 0.97        | 1.07        | 0.63        | 1.01        | 1.00        | 0.82        |
|                   | CGC          | 0.92        | 0.90        | 0.81        | 0.16        | 0.87        | 0.82        | 0.66        |
|                   | CGG          | 0.30        | 0.39        | 0.35        | 0.14        | 0.20        | 0.22        | 0.16        |
|                   | CGU          | 1.42        | 1.39        | 1.49        | 0.66        | 1.24        | 1.28        | 1.64        |
| Asn               | AAC          | 0.80        | 0.73        | 0.83        | 0.56        | 0.64        | 0.67        | 0.55        |
|                   | AAU          | <b>1.20</b> | <b>1.27</b> | <b>1.17</b> | <b>1.44</b> | <b>1.36</b> | <b>1.33</b> | <b>1.45</b> |
| Asp               | GAC          | 0.73        | 0.69        | 0.66        | 0.58        | 0.55        | 0.56        | 0.51        |
|                   | GAU          | <b>1.27</b> | <b>1.31</b> | <b>1.34</b> | <b>1.42</b> | <b>1.45</b> | <b>1.44</b> | <b>1.49</b> |
| Cys               | UGC          | 0.74        | 0.86        | 0.96        | 0.63        | 0.77        | 0.91        | 0.58        |
|                   | UGU          | <b>1.26</b> | <b>1.14</b> | <b>1.04</b> | <b>1.37</b> | <b>1.23</b> | <b>1.09</b> | <b>1.42</b> |
| Gln               | CAA          | <b>1.06</b> | <b>1.10</b> | <b>1.04</b> | <b>1.33</b> | <b>1.36</b> | <b>1.36</b> | <b>1.45</b> |
|                   | CAG          | 0.94        | 0.9         | 0.96        | 0.67        | 0.64        | 0.64        | 0.55        |
| Glu               | GAA          | <b>1.01</b> | 0.96        | 0.95        | <b>1.29</b> | <b>1.18</b> | <b>1.18</b> | <b>1.33</b> |
|                   | GAG          | 0.99        | <b>1.04</b> | <b>1.05</b> | 0.71        | 0.82        | 0.82        | 0.67        |
| Gly               | GGA          | 1.28        | 1.29        | 1.26        | <b>1.97</b> | 1.39        | 1.37        | 1.40        |
|                   | GGC          | 0.88        | 0.83        | 0.83        | 0.37        | 0.85        | 0.83        | 0.59        |
|                   | GGG          | 0.44        | 0.37        | 0.39        | 0.41        | 0.33        | 0.36        | 0.27        |
|                   | GGU          | <b>1.41</b> | <b>1.52</b> | <b>1.51</b> | 1.25        | <b>1.43</b> | <b>1.44</b> | <b>1.75</b> |
| His               | CAC          | 0.84        | 0.71        | 0.66        | 0.51        | 0.57        | 0.63        | 0.46        |
|                   | CAU          | <b>1.16</b> | <b>1.29</b> | <b>1.34</b> | <b>1.49</b> | <b>1.43</b> | <b>1.37</b> | <b>1.54</b> |
| Ile               | AUA          | <b>1.21</b> | <b>1.19</b> | <b>1.20</b> | <b>1.77</b> | 1.15        | 1.11        | 1.03        |
|                   | AUC          | 0.84        | 0.9         | 0.79        | 0.28        | 0.55        | 0.58        | 0.49        |
|                   | AUU          | 0.95        | 0.91        | 1.02        | 0.95        | <b>1.30</b> | <b>1.31</b> | <b>1.47</b> |
| Leu               | CUA          | 1.30        | 1.21        | 1.30        | 0.81        | 0.74        | 0.80        | 0.78        |

续表 2( Table 2 continued )

| 氨基酸<br>Amino acid | 密码子<br>Codon | BmCPV-1     | DpCPV-1     | LdCPV-1     | HaCPV-5     | HaCPV-14    | LdCPV-14    | TnCPV-15    |
|-------------------|--------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                   | CUC          | 0.55        | 0.49        | 0.45        | 0.14        | 0.18        | 0.21        | 0.37        |
|                   | CUG          | 0.62        | 0.63        | 0.65        | 0.43        | 0.49        | 0.44        | 0.25        |
|                   | CUU          | 0.76        | 0.8         | 0.85        | 0.49        | 0.84        | 0.82        | 1.03        |
|                   | UUA          | <b>1.85</b> | <b>1.86</b> | <b>1.79</b> | <b>2.99</b> | <b>2.16</b> | <b>2.18</b> | <b>2.54</b> |
|                   | UUG          | 0.92        | 1.02        | 0.96        | 1.14        | 1.58        | 1.55        | 1.03        |
| Lys               | AAA          | 0.96        | 1.00        | 0.99        | <b>1.39</b> | <b>1.17</b> | <u>1.17</u> | <b>1.31</b> |
|                   | AAG          | <b>1.04</b> | <u>1.00</u> | <b>1.01</b> | 0.61        | 0.83        | 0.83        | 0.69        |
| Phe               | UUC          | <u>1.04</u> | <u>1.09</u> | <u>1.12</u> | 0.64        | 0.79        | 0.80        | 0.56        |
|                   | UUU          | 0.96        | 0.91        | 0.88        | <b>1.36</b> | <b>1.21</b> | <b>1.20</b> | <b>1.44</b> |
| Pro               | CCA          | <b>2.00</b> | <b>1.82</b> | <b>1.68</b> | <b>2.01</b> | <b>2.01</b> | <b>1.96</b> | 1.36        |
|                   | CCC          | 0.48        | 0.62        | 0.52        | 0.33        | 0.28        | 0.32        | 0.40        |
|                   | CCG          | 0.51        | 0.6         | 0.74        | 0.76        | 0.77        | 0.81        | 0.50        |
|                   | CCU          | 1.01        | 0.95        | 1.06        | 0.91        | 0.95        | 0.91        | <b>1.73</b> |
| Ser               | AGC          | 0.77        | 0.86        | 0.84        | 0.78        | 0.79        | 0.76        | 0.62        |
|                   | AGU          | 1.26        | 1.12        | 1.11        | 1.67        | 1.45        | 1.43        | 1.61        |
|                   | UCA          | <b>2.00</b> | <b>1.82</b> | <b>1.85</b> | <b>1.98</b> | <b>1.58</b> | <b>1.61</b> | 1.20        |
|                   | UCC          | 0.71        | 0.51        | 0.5         | 0.29        | 0.45        | 0.44        | 0.46        |
|                   | UCG          | 0.61        | 0.66        | 0.66        | 0.61        | 0.80        | 0.78        | 0.35        |
| Thr               | UCU          | 0.66        | 1.02        | 1.03        | 0.67        | 0.93        | 0.98        | <u>1.76</u> |
|                   | ACA          | <u>1.59</u> | <b>1.43</b> | <b>1.46</b> | <b>1.73</b> | <b>1.30</b> | <b>1.28</b> | 1.10        |
|                   | ACC          | 0.48        | 0.58        | 0.55        | 0.43        | 0.54        | 0.56        | 0.44        |
|                   | ACG          | 0.92        | 1.03        | 0.99        | 0.92        | 0.94        | 0.94        | 0.59        |
|                   | ACU          | 1.01        | 0.95        | 1.00        | 0.92        | 1.21        | 1.22        | <u>1.88</u> |
| Tyr               | UAC          | 0.86        | 0.9         | 0.83        | 0.65        | 0.75        | 0.74        | 0.49        |
|                   | UAU          | <b>1.14</b> | <b>1.1</b>  | <b>1.17</b> | <b>1.35</b> | <b>1.25</b> | <b>1.26</b> | <b>1.51</b> |
| Val               | GUA          | <b>1.40</b> | <b>1.19</b> | <u>1.21</u> | <b>1.72</b> | 0.99        | 0.99        | 1.23        |
|                   | GUC          | 0.53        | 0.69        | 0.66        | 0.33        | 0.47        | 0.50        | 0.49        |
|                   | GUG          | 1.05        | 1.11        | 1.12        | 0.91        | 1.10        | 1.09        | 0.71        |
|                   | GUU          | 1.02        | 1.01        | 1.00        | 1.04        | <b>1.44</b> | <b>1.41</b> | <b>1.57</b> |

注:加粗显示的密码子为偏好性密码子,下划线密码子为 CPV 与宿主共有的偏好性密码子。  
The preferred codons are in bold, the preferred codons shared by viruses and their host are underlined.

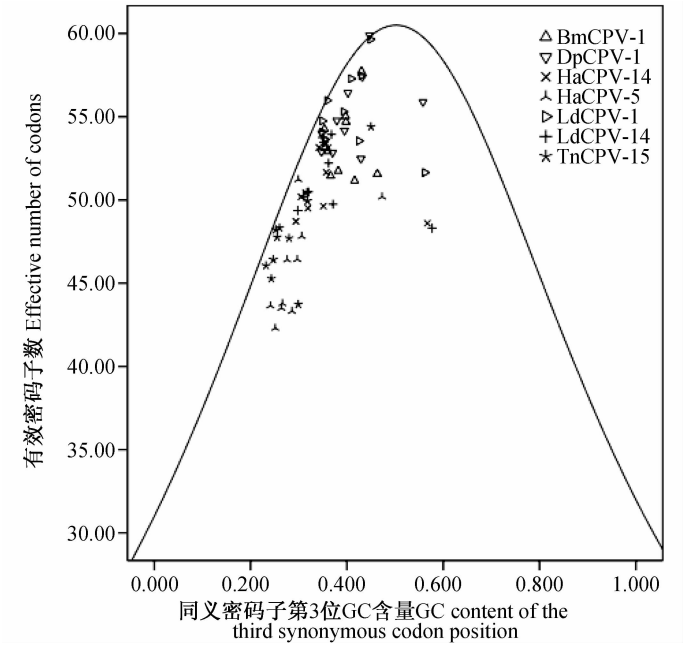


图1 昆虫质型多角体病毒密码子使用模式 ENc-plot 分析

**Fig.1 ENc-plot of codon usage patterns of insect cypoviruses**

实线代表 ENc 对同义密码子第 3 位 GC 含量的期望值。

The expected ENc values from GC3s are shown as a solid line.

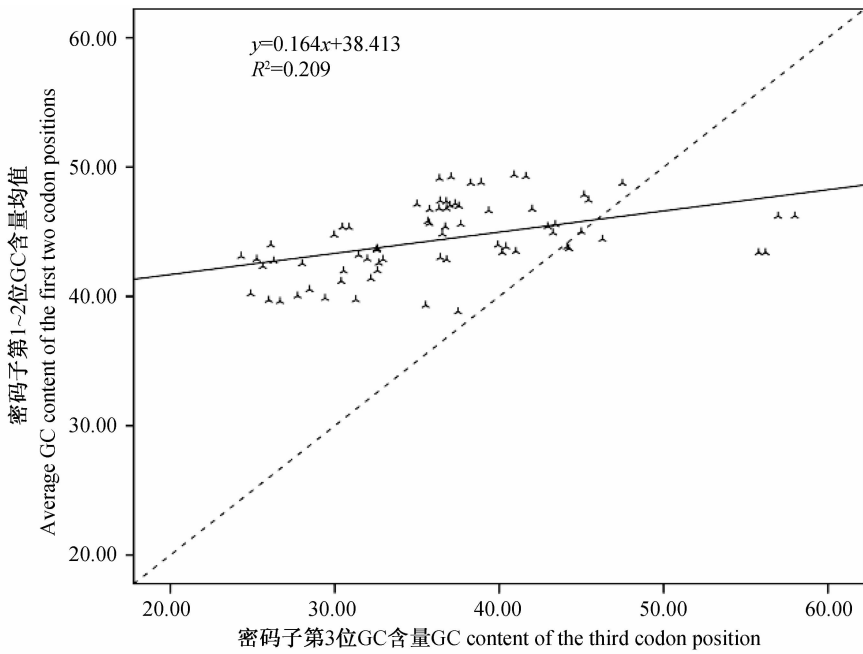


图2 昆虫质型多角体病毒密码子使用模式中中性绘图分析

**Fig.2 Neutrality-plot of codon usage patterns of insect cypoviruses**

实线代表  $P_{12}$  对  $P_3$  的直线回归,  $R^2 = 0.209$ ,  $P < 0.001$ 。

The solid line is the linear regression of  $P_{12}$  against  $P_3$ ,  $R^2 = 0.209$ ,  $P < 0.001$ .

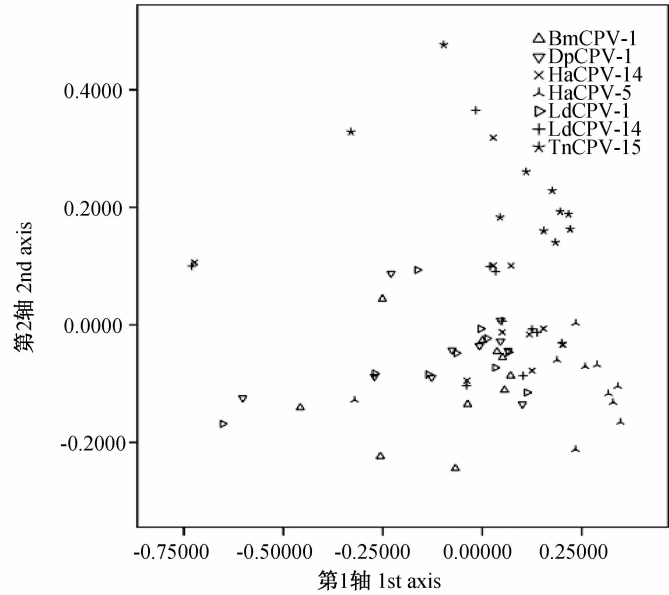


图 3 基因在对应分析第 1 轴和第 2 轴上的位置

Fig. 3 Positions of the genes in the plot of the first two major axes in correspondence analysis

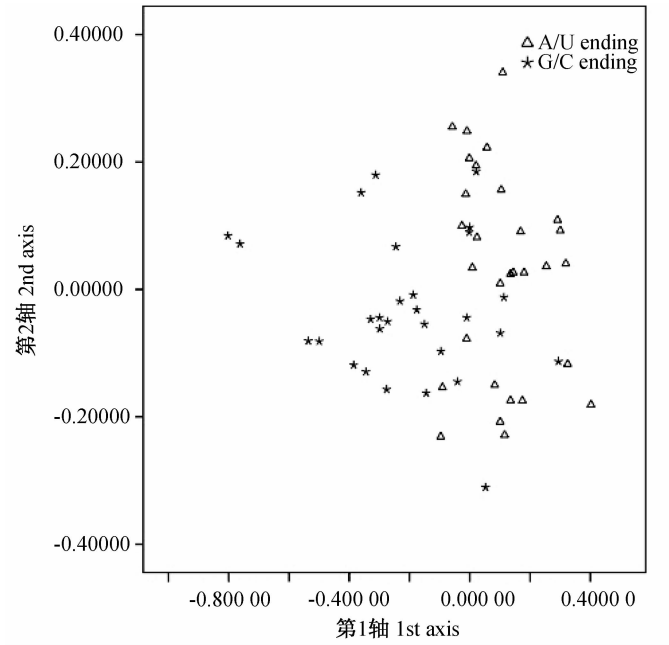


图 4 密码子在对应分析第 1 轴和第 2 轴上的位置

Fig. 4 Positions of the codons in the plot of the first two major axes in correspondence analysis

3 讨论

生物体在选择编码同一氨基酸的不同密码子时具有偏好性,这种偏好性与基因组进化有关。分析密码子偏好性及其影响因素对研究病毒进化

及病毒与宿主间的相互关系至关重要(Shackelton *et al.*, 2006),一些病毒中密码子偏好性显著,如甲型肝炎病毒(Andrea *et al.*, 2011);而另一些病毒中密码子偏好性则较弱,如脊髓灰质炎病毒(Zhang *et al.*, 2011)和新城疫病毒(Wang *et al.*,

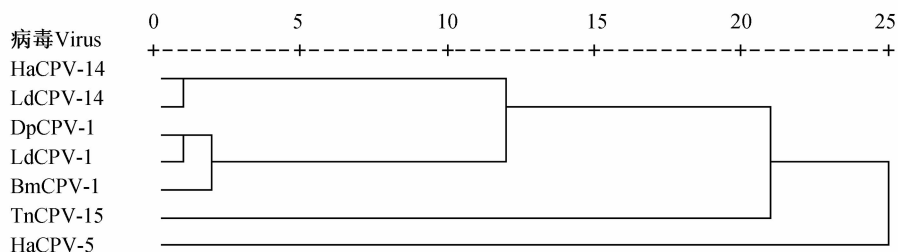


图5 昆虫质型多角体病毒基于密码子偏好性的聚类分析

Fig.5 Cluster analysis of insect cypoviruses based on codon bias

2011a)。

本研究首次对7种CPV密码子使用模式进行全面分析,结果表明CPV密码子偏好性较弱,与昆虫杆状病毒密码子使用模式的研究结果相一致(Jiang *et al.*, 2008)。RNA病毒同义密码子偏好性较弱的原因之一,可能是病毒选择与宿主不同的较弱的密码子偏好性有利于其在宿主体内高效复制(Zhang *et al.*, 2011)。作为适应性进化结果,病毒密码子使用与宿主有相关性(Chantawannakul and Cutler, 2008; Bahir *et al.*, 2009),并受宿主影响(Liu *et al.*, 2011; Zhou *et al.*, 2012),CPV宿主特异性较低、宿主域较广(贺倩等, 2010),因此来自宿主的选择可能也是较弱密码子偏好性的原因。本研究比较了CPV与宿主昆虫的密码子使用偏好性,发现CPV与宿主昆虫的密码子偏好性存在差异,但未能阐明宿主昆虫对CPV密码子使用模式的影响方式。CPV不同基因组片段密码子使用偏好性存在统计学差异,这一结果与已有研究密码子使用与蛋白质三级结构有相关性这一结论相符(顾万君等, 2002)。

自然选择和突变压力被认为是不同生物体中导致密码子使用差异的主要因素(Hershberg and Petrov, 2008; Sharp *et al.*, 2010),本研究中密码子偏好性与碱基组成的相关分析及ENc-plot和中性绘图分析表明,选择对CPV同义密码子使用模式的影响大于突变。聚类分析中相同电泳型CPV聚类距离较近,不同电泳型CPV聚类距离较远,提示基于密码子使用模式的聚类分析可为CPV电泳型分类提供依据。然而,由于可供分析的基因组数量较少及密码子使用数据库更新滞后,本研究所提供的密码子使用模式结论可能具有局限性。随着CPV基因组测序数量及宿主密码子使用数据增

加,有必要进行更广泛深入的病毒和宿主间密码子使用模式分析,为进一步研究病毒和宿主间互作关系提供依据。

## 参考文献 (References)

- Andrea LD, Pintó RM, Bosch A, Musto H, Cristin J, 2011. A detailed comparative analysis on the overall codon usage patterns in hepatitis A virus. *Virus Res.*, 157(1):19–24.
- Bahir I, Fromer M, Prat Y, Linial M, 2009. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol. Syst. Biol.*, 5:311.
- Chantawannakul P, Cutler RW, 2008. Convergent host-parasite codon usage between honeybee and bee associated viral genomes. *J. Invertebr. Pathol.*, 98(2):206–210.
- Eyre-Walker AC, 1991. An analysis of codon usage in mammals; selection or mutation bias? *J. Mol. Evol.*, 33(5):442–449.
- Goodarzi H, Torabi N, Najafabadi HS, Archetti M, 2008. Amino acid and codon usage profiles; adaptive changes in the frequency of amino acids and codons. *Gene*, 407(1/2):30–41.
- Hershberg R, Petrov DA, 2008. Selection on codon bias. *Annu. Rev. Genet.*, 42:287–299.
- Ingvarsson PK, 2008. Molecular evolution of synonymous codon usage in Populus. *BMC Evol. Biol.*, 8:307.
- Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM, 2011. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science*, 333(6040):348–353.
- Jiang Y, Deng F, Wang H, Hu Z, 2008. An extensive analysis on the global codon usage pattern of baculoviruses. *Arch. Virol.*, 153(12):2273–2282.
- Jung SK, McDonald K, 2011. Visual gene developer; a fully

- programmable bioinformatics software for synthetic gene optimization. *BMC Bioinformatics*, 12:340.
- King AM, Adams MJ, Carstens EB, Lefkowitz E, 2012. Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses. San Diego: Elsevier Academic Press. 572 – 580.
- Liu YS, Zhou JH, Chen HT, Ma LN, Pejsak Z, Ding YZ, Zhang J, 2011. The characteristics of the synonymous codon usage in enterovirus 71 virus and the effects of host on the virus in codon usage pattern. *Infect. Genet. Evol.*, 11(5):1168 – 1173.
- Nakamura Y, Gojobori T, Ikemura T, 2000. Codon usage tabulated from international DNA sequence databases; status for the year 2000. *Nucleic Acids Res.*, 28(1):292.
- Plotkin JB, Kudla G, 2011. Synonymous but not the same; the causes and consequences of codon bias. *Nat. Rev. Genet.*, 12(1):32 – 42.
- Shackelton LA, Parrish CR, Holmes EC, 2006. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J. Mol. Evol.*, 62(5):551 – 563.
- Shah P, Gilchrist MA, 2011. Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *PNAS*, 108(25):10231 – 10236.
- Sharp PM, Emery LR, Zeng K, 2010. Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 365(1544):1203 – 1212.
- Sueoka N, 1988. Directional mutation pressure and neutral molecular evolution. *PNAS*, 85(8):2653 – 2657.
- Sueoka N, 1999. Two aspects of DNA base composition: G + C content and translation-coupled deviation from intra-strand rule of A = T and G = C. *J. Mol. Evol.*, 49(1):49 – 62.
- Wang M, Liu YS, Zhou JH, Chen HT, Ma LN, Ding YZ, Liu WQ, Gu YX, Zhang J, 2011a. Analysis of codon usage in Newcastle disease virus. *Virus Genes*, 42(2):245 – 253.
- Wang M, Zhang J, Zhou JH, Chen HT, Ma LN, Ding YZ, Liu WQ, Liu YS, 2011b. Analysis of codon usage in bovine viral diarrhea virus. *Arch. Virol.*, 156(1):153 – 160.
- Wright F, 1990. The effective number of codons' used in a gene. *Gene*, 87(1):23 – 29.
- Zhang J, Wang M, Liu WQ, Zhou JH, Chen HT, Ma LN, Ding YZ, Gu YX, Liu YS, 2011. Analysis of codon usage and nucleotide composition bias in polioviruses. *Virol. J.*, 8:146.
- Zhou JH, Gao ZL, Zhang J, Chen HT, Pejsak Z, Ma LN, Ding YZ, Liu YS, 2012. Comparative the codon usage between the three main viruses in pestivirus genus and their natural susceptible livestock. *Virus Genes*, 44(3):475 – 481.
- 顾万君, 马建民, 周童, 孙啸, 陆祖宏, 2002. 不同结构的蛋白编码基因的密码子偏性研究. *生物物理学报*, 18(1):81 – 86.
- 贺倩, 刘小侠, 张青文, 2010. 昆虫质型多角体病毒的研究进展. *昆虫知识*, 47(5):834 – 840.
- 李京生, 杨瑞生, 贾萍, 石生林, 刘彦群, 潘敏慧, 秦利, 2011. 柞蚕核型多角体病毒和家蚕核型多角体病毒的同义密码子使用模式分析. *蚕业科学*, 37(2):266 – 271.
- 石生林, 李苗苗, 姜义仁, 杨瑞生, 秦利, 2012. 5 种家蚕病毒的密码子及密码对使用模式分析. *蚕业科学*, 38(2):273 – 280.