

大豆花叶病毒密码子使用及其进化分析

袁 媛,黄赛花,王成坤,智海剑

(南京农业大学 大豆研究所/国家大豆改良中心,作物遗传与种质创新国家重点实验室,江苏 南京 210095)

摘要:对 57 个来自不同地区的大豆花叶病毒(SMV)分离物的基因组序列进行密码子使用偏性与进化分析。结果显示,SMV 基因组 GC 含量较低;GC3 < 50%,说明 SMV 偏好使用以 A/T 结尾的密码子;SMV 基因组密码子 3 个位置的 GC 含量均相互呈极显著相关($P < 0.01$),说明密码子 3 个位置的碱基组成相关;ENC 值为 49.42 ~ 51.78,密码子偏性存在但并不显著;对应性分析结果表明,第 1 轴显示 67.84% 的差异,ENC、GC3、GC12 之间的相关性均达极显著水平($P < 0.01$),说明密码子的偏性与碱基突变相关;中性绘图及 ENC 绘图分析均表明密码子使用模式主要受碱基突变影响;密码子使用频率比较结果显示 SMV 存在 19 个最优密码子;RSCU 的聚类分析证明 SMV 密码子偏性与病毒地理来源相关。

关键词:大豆花叶病毒;密码子偏性;多元分析;进化

中图分类号:S565.1

文献标识码:A

DOI:10.11861/j.issn.1000-9841.2014.06.0801

Analysis on Codon Usage and Evolution of Soybean Mosaic Virus

YUAN Yuan, HUANG Sai-hua, WANG Cheng-kun, ZHI Hai-jian

(Soybean Research Institute of Nanjing Agricultural University/National Center for Soybean Improvement, National Key Laboratory for Crop Genetics and Germplasm Enhancement, Nanjing 210095, China)

Abstract: The codon usage and evolution of 57 SMV isolates from different countries were analyzed in this study. The results indicated that GC content of codons was lower than AT in SMV isolates. GC3 content was less than 50%. SMV preferred to use those codons ending with A or T. GC content at three positions of codons (GC1, GC2 and GC3) had a significant correlation with each other ($P < 0.01$). ENC value was between 49.42 and 51.78, which meant that the codon bias existed while it was not significant. Corresponding analysis (COA) indicated that Axis 1 appeared 67.84% variation. There were significant correlation ($P < 0.01$) among ENC, GC3 and GC12. Above results with the analysis of Neutralityplot and ENC plot revealed that codon bias was influenced mainly by mutation pressure. And the comparative analysis of RSCU discovered 19 optimized codons in SMV. Cluster analysis based upon RSCU proved that SMV codon bias was related with their geographical origin.

Key words: Soybean Mosaic Virus; Codon usage bias; Multivariate analysis; Evolution

大豆花叶病毒(Soybean Mosaic Virus, SMV)是马铃薯 Y 病毒属(Potyvirus)的重要成员,存在于世界各大豆产区。SMV 基因组为单链正义 RNA,长约 9.6 kb,5'端有共价键结合金属蛋白 Vpg(Virus genome-linked protein, Vpg),3'端有一个 Poly(A)尾巴。基因组含有 2 个开放阅读框(ORF),翻译后产生一个多聚蛋白前体。该蛋白经自身编码的 3 个蛋白酶(P1, HC-Pro 和 NIa-Pro)切割加工后形成 11 个不同功能的成熟蛋白:P1、HC-Pro、P3、PIPO、6K1、CI、6K2、NIa-Vpg、NIa-Pro、NIb 和 CP。其中,CP 是唯一的结构蛋白^[1]。

不同物种或是同一物种的不同基因其同义密码子使用并不平均。在物种长期进化过程中,受其基因表达水平、GC 含量、长度、tRNA 等因素的影响,密码子使用具有不同偏好特点^[2-3]。密码子使用偏好性研究对物种密码子使用规律、目的基因的

高效表达等具有重要意义,可对分子水平阐明物种的起源及遗传分化等提供一定依据^[4]。但在植物病毒密码子偏性方面研究较少,目前,有关 SMV 密码子偏性的研究鲜有报道。

本文对来自不同地区的 57 个大豆花叶病毒(SMV)分离物基因组序列进行了密码子偏性分析,并对影响 SMV 基因组密码子使用偏性及其主要影响因素进行了探讨。

1 材料与方法

1.1 SMV 基因组序列

所使用的 57 个大豆花叶病毒分离物的基因组序列均来自 GenBank(表 1)。其中 4469-4、4547、6067-1、6202-2、Sc3、Sc6 分离物序列由本实验室提供。

收稿日期:2014-03-12

基金项目:转基因生物新品种培育科技重大专项(2008ZX08004-004);国家自然科学基金(31171574,31371646);现代农业产业技术体系(CARS-004)。

第一作者简介:袁媛(1989-),女,博士,主要从事抗大豆花叶病毒基因定位研究。E-mail:2013201006@njau.edu.cn。

通讯作者:智海剑(1957-),男,教授,博导,主要从事大豆遗传育种研究。E-mail:zhj@njau.edu.cn。

1.2 同义密码子使用偏好性分析和相关性检测

运用在线软件 Codon W (<http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::codonw>) 计算 SMV 基因组密码子使用偏性的有关参数:GC 总含量、同义密码子相对使用频率(relative synonymous codon usage, RSCU)、有效密码子数(effective number of codons, ENC)等^[5-6]。

利用 CUSP (<http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::cusp>) 软件分别计算 GC 在密码子第一、二和三位上的含量(GC1、GC2 和 GC3)。

对密码子各参数进行相关性检测。

1.3 中性绘图分析

以各密码子 GC1、GC2 的平均值计为 GC12。以 GC12 为纵坐标,GC3 为横坐标绘图,图中每个点代表独立的 SMV 分离物,回归曲线斜率则反映 GC12 和 GC3 的相关性,回归曲线反映中性程度^[7-8]。

1.4 ENC 绘图分析

以 ENC 为纵坐标,GC3 为横坐标绘图。并利用公式 $ENC = 2 + GC3 + 29/[GC3^2 + (1 - GC3)^2]$ 绘制标准曲线^[9]。

1.5 对应性分析

对应分析(corresponding analysis, COA)是最常用的多变量统计分析方法^[10]。各分离物按来源分类,利用 Codon W 软件,将所有分离物分布在一个

59 维(59 个密码子)的向量空间中,基于变量 RSCU 值的变化趋势,将分离物沿轴连续分布^[11]。

1.6 最优密码子

计算各序列的 ENC 比值(ENC 实际观察值和期望值之比),根据比值对序列进行排序,两极各选 10% 的分离物,分别列为高偏倚和低偏倚序列数据库。计算两数据库中对应密码子的 RSCU 值,若差异大于 0.08,则认为该密码子是最优密码子^[12]。

1.7 基于密码子使用偏性的聚类分析

利用多元分析软件 SPSS18.0,将 SMV 分离物作为对象,将 RSCU 值(仅用 59 个密码子的 RSCU 值,去除编码蛋氨酸的密码子 AUG 和编码色氨酸的密码子 UGG 以及 3 个终止密码子 UAA、UAG、UGA)作为变量进行聚类分析。并运用 MEGA5.05 软件对 57 条 SMV 全基因组序列进行系统发育分析;系统进化树的构建基于 Neighbor-joining (NJ) 算法^[13]。

2 结果与分析

2.1 密码子偏好性分析

表 1 显示,SMV 基因总 GC 含量较低,为 41% ~ 44%,且 GC3 含量均小于 50%,说明 SMV 偏好使用以 A/T 结尾的密码子。57 个 SMV 分离物中有 49 个

表 1 大豆花叶病毒来源及密码子不同位置的 GC 含量、ENC 值

Table 1 Content of GC codon at three positions and ENC value and sources of SMV

分离物 Isolate	总 GC Total GC/%	GC/%			ENC	GenBank 登录号 Accession number	来源 Country
		GC1	GC2	GC3			
413	41.5	47.87	35.98	40.52	51.08	GU015011	未知 Unknown
4469-4	41.9	48.30	36.06	41.13	51.32	HM590055	中国 China
4547	42.0	48.39	36.24	41.28	51.30	HQ396725	中国 China
6067-1	43.0	35.89	41.52	48.37	50.49	JF833015	中国 China
6202-2	41.7	48.15	36.05	40.71	51.02	JF833014	中国 China
Aa	42.8	35.76	41.46	48.09	49.63	AB100442	日本 Japan
Aa15-M2	42.9	35.83	41.55	48.12	49.68	AB100443	日本 Japan
G1	42.8	35.76	41.08	48.15	49.57	FJ640977	美国 USA
G2	43.1	36.08	41.8	48.12	50.23	S42280	美国 USA
G3	42.7	35.89	40.71	48.25	49.98	FJ640978	美国 USA
G4	43.1	35.93	41.97	48.14	50.30	FJ640979	美国 USA
G5	42.9	35.95	41.30	48.28	49.91	AY294044	美国 USA
G5H	42.8	35.79	41.65	47.97	49.58	FJ807701	韩国 Korea
G6	42.7	35.43	41.44	48.23	49.46	FJ640980	美国 USA
G6H	42.8	35.59	41.72	48.01	49.91	FJ640981	韩国 Korea
G7	42.8	35.64	41.77	47.65	49.74	AY216010	美国 USA
G7A	42.8	35.83	41.74	47.62	49.76	FJ640982	美国 USA
G7H	43.0	35.79	41.46	48.40	50.16	AY294045	韩国 Korea
G7H2	42.9	35.76	41.36	48.37	50.30	FJ807700	韩国 Korea

续表 1

分离物 Isolate	总 GC Total GC/%	GC/%			ENC	GenBank 登录号 Accession number	来源 Country
		GC1	GC2	GC3			
L	43.2	35.93	42.19	48.11	50.27	EU871724	加拿大 Canada
L-RB	43.3	35.99	42.41	48.11	50.21	EU871725	加拿大 Canada
N	43.1	35.92	41.83	48.12	50.53	D00507	美国 USA
Rsv4-RB3	43.0	35.82	42.34	47.86	50.31	JN416770	未知 Unknown
Sc3	41.6	48.03	36.05	40.61	51.02	JF833013	中国 China
Sc6	41.5	47.87	35.86	40.61	50.93	HM590054	中国 China
WS32	42.9	35.40	41.94	48.23	49.49	FJ640954	韩国 Korea
WS37	43.1	36.09	41.78	48.14	50.53	FJ640955	韩国 Korea
WS84	43.0	35.89	41.80	47.84	49.55	FJ640956	韩国 Korea
WS101	42.9	35.71	41.82	48.14	49.42	FJ640957	韩国 Korea
WS105	42.8	35.76	41.52	48.03	49.86	FJ640958	韩国 Korea
WS109	43.2	36.03	42.16	48.11	50.33	FJ640959	韩国 Korea
WS110	42.9	35.79	41.71	48.06	49.88	FJ640960	韩国 Korea
WS116	43.0	35.67	41.90	48.19	49.53	FJ640961	韩国 Korea
WS117	43.0	35.96	41.85	48.04	50.01	FJ640962	韩国 Korea
WS128	43.1	35.83	41.90	48.19	50.70	FJ640963	韩国 Korea
WS132	42.7	35.61	41.43	48.06	49.44	FJ640964	韩国 Korea
WS135	42.8	35.64	41.55	47.97	49.90	FJ640965	韩国 Korea
WS144	43.2	35.99	42.25	48.08	50.37	FJ640966	韩国 Korea
WS145	43.2	35.9	42.10	48.14	51.04	FJ640967	韩国 Korea
WS149	43.2	35.96	42.22	48.14	50.29	FJ640968	韩国 Korea
WS151	42.7	35.73	41.18	48.15	49.70	FJ640969	韩国 Korea
WS155	43.1	35.66	42.39	48.06	49.52	FJ640970	韩国 Korea
WS156	42.9	35.74	41.66	48.08	49.87	FJ640971	韩国 Korea
WS160	43.2	35.99	42.28	48.08	50.38	FJ640972	韩国 Korea
WS162	43.2	35.81	42.22	48.17	49.99	FJ640973	韩国 Korea
WS200	43.0	35.99	41.50	48.36	50.35	FJ548849	韩国 Korea
WS202	43.2	35.99	42.35	48.01	50.48	FJ640974	韩国 Korea
WS205	43.1	35.93	42.22	48.04	50.34	FJ640975	韩国 Korea
WS209	43.2	36.06	42.22	48.08	50.41	FJ640976	韩国 Korea
Am	42.5	48.18	36.32	42.95	51.78	KC845322	中国 China
Ar13	43.0	35.99	41.38	48.42	50.33	KF135488	伊朗 Iran
Ar33	43.0	36.07	41.58	48.09	49.96	KF135489	伊朗 Iran
Go11	42.9	35.53	41.67	47.61	51.37	KF135491	伊朗 Iran
Ky	41.4	47.87	36.08	40.33	50.76	HQ845736	美国 USA
Lo3	43.0	36.13	41.73	48.03	50.08	KF135490	伊朗 Iran
TNP	42.8	35.93	41.13	48.11	50.28	HQ845735	美国 USA
SX	42.8	36.03	41.44	48.14	50.36	KC845321	中国 China

表2 各基因相关参数的相关性分析

Table 2 Correlation analysis of each gene's related parameters

	GC	GC1	GC2	GC3	GC12
GC1	-0.894 **				
GC2	0.942 **	-0.982 **			
GC3	0.930 **	-0.989 **	0.974 **		
GC12	-0.825 **	0.986 **	-0.937 **	-0.972 **	
ENC	-0.438 **	0.692 **	-0.632 **	-0.649 **	0.723 **

** : $P < 0.01$ 极显著水平。

** : $P < 0.01$ extremely significant level.

的基因组密码子不同位点的 GC 含量表现为 $GC1 < GC2 < GC3$; 仅有 8 个 SMV 分离物的基因组碱基表现为 $GC2 < GC3 < GC1$, 其中有 6 个来源于中国。ENC 值为 49.42 ~ 51.78, 位于 20 (每个氨基酸只使用一个密码子的极端偏性情况) ~ 61 (各个密码子均被平均使用), 说明基因组中各密码子使用存在偏性但并不显著。由表 2 知, GC、GC1、GC2、GC3 的相关性均相互达到极显著水平 ($P < 0.01$), 且 GC1 对 GC2、GC3 呈负相关, GC2 对 GC3 则呈正相关, 说明密码子第一、二、三位的碱基组成相关。ENC 值与 3 个位置碱基的 GC 含量呈极显著相关 ($P < 0.01$), 表明碱基组成对密码子使用偏性有影响。

2.2 中性绘图分析

57 个 SMV 分离物基因组序列的中性绘图分析见图 1, 多数 SMV 分布在对角线下, 只有少部分在对角线上。GC3 分布于 0.404 4 ~ 0.484 2, GC12 分布于 0.383 0 ~ 0.423 2, 两者分布范围均较紧凑。GC12、GC3 的相关系数为 -0.972, 呈负相关, 且相关性达到显著水平 ($P < 0.01$), 回归曲线斜率为 -0.944, 均说明 GC12 与 GC3 高度相关, 且 3 个位点的碱基组成有明显的位置效应。此时基因组承

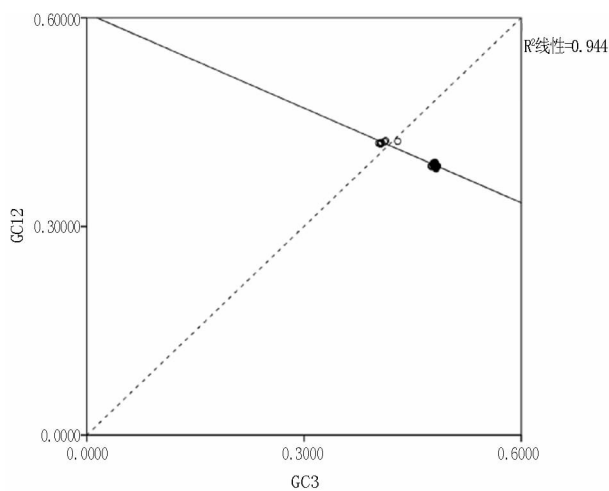


图1 中性绘图分析

Fig. 1 Neutrality plot analysis

受的进化约束力为突变, 即碱基的改变对密码子偏性有很大影响。

2.3 ENC 绘图分析

57 个 SMV 分离物基因组序列的 ENC 绘图分析如图 2 所示, 图中标准曲线表示在不存在选择压力时, 密码子偏好性完全取决于碱基组成偏好性, 即突变压^[9]。SMV 均分布在标准曲线下, GC3 分布于 0.404 4 ~ 0.484 2, ENC 分布于 49.42 ~ 59.78, ENC 实际观察值较接近标准曲线, 说明密码子偏好性主要反映在 GC 组成的偏好性上, 起主要决定因素是突变压。图 2 也显示, ENC 实际观察值并未紧贴标准曲线, 说明密码子偏性除受突变压作用外, 还受选择压的作用, 但影响较弱。

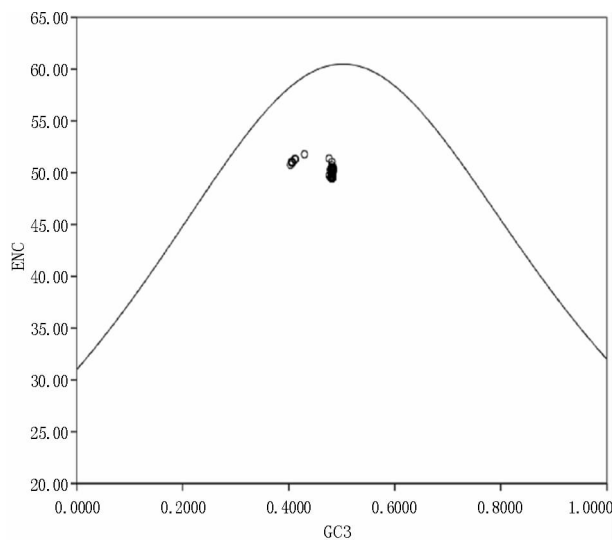


图2 ENC 和 GC3 的相关性分析

Fig. 2 Correlation analysis of ENC and GC3

2.4 对应性分析

对已有的 57 个 SMV 分离物基因组序列基于 RSCU 进行对应性分析, 并按其来源划分, 结果如图 3 所示。数据在第一轴显示密码子使用的差异为 67.84%, 另三轴显示的差异分别为 7%、5.58%、3.62%。第一轴的差异远远高于另外 3 轴, 表明第一轴是主要描述数据模式的轴向; 影响第一轴的因

素即是影响密码子使用模式的最主要因素^[12]。因此,对第一轴与 GC、GC3、ENC 进行相关性检测,发现呈极显著相关(r 值分别为 0.910, 0.995, -0.978, $P < 0.01$),说明碱基的组成突变与密码子使用的偏性直接关联。根据 SMV 分离物来源,将 SMV 分为 6 组。观察图 3 可知,轴 1 将 SMV 分为两类:0~0.1 区域中的 SMV 分离物分布较紧凑,说明各密码子碱基组成较相似;但位于 -0.3~0 区域的

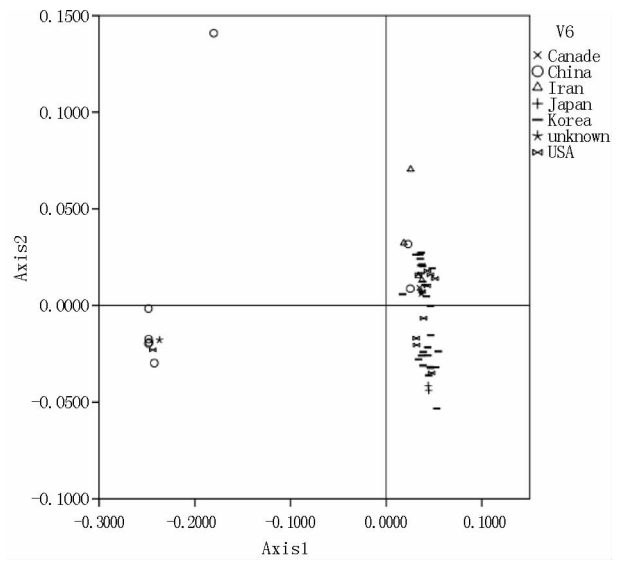


图 3 基于 RSCU 的对应性分析
Fig. 3 Corresponding analysis based on RSCU

SMV 较分散,且以来源于中国的分离物为主,说明来源于中国的 SMV 密码子碱基的组成与其他地区有较大差异。

2.5 最优密码子

通过对高偏倚和低偏倚两个数据库的比较,根据两个数据库中密码子 RSCU 差值,确定最优密码子,结果见表 3。这些密码子在高偏倚的数据库中相对使用频率明显高于其同义密码子以及其在低偏倚数据库中的频率,最优密码子的频繁出现可能与翻译水平相关。

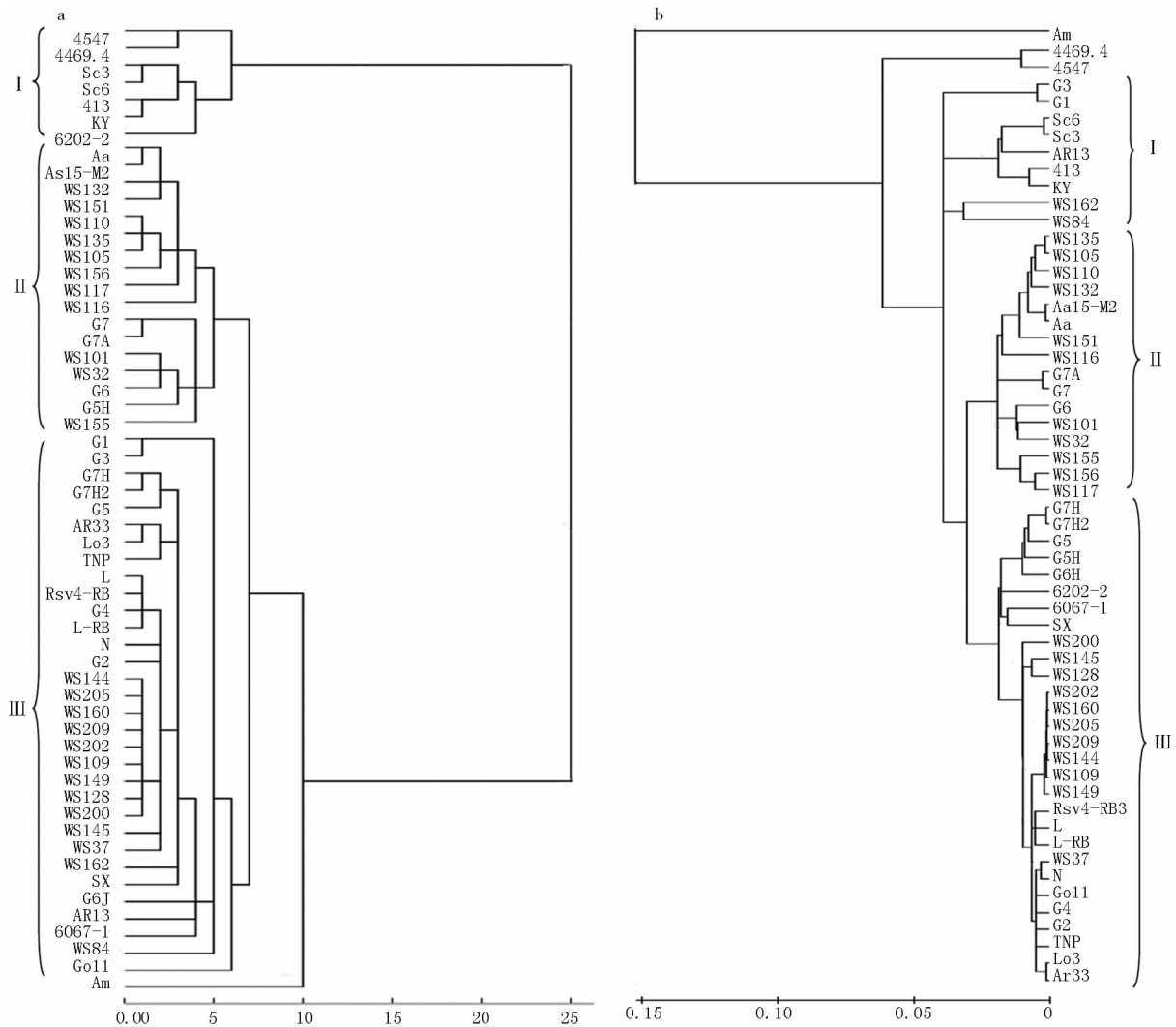
2.6 基于密码子偏性的聚类分析

对 57 个 SMV 分离物基因组序列的 RSCU 聚类分析。从图 4a 可知,SMV 被分为 3 类,第Ⅱ类和第Ⅲ类距离相对较近,与第Ⅰ类距离较远,且第Ⅰ类主要包含来自中国的 SMV 分离物。说明第Ⅱ类和第Ⅲ类 SMV 密码子使用较相似,但与来自中国的 SMV 分离物却有较大差异,结果与对应性分析一致,其原因可能是由碱基组成造成。按 57 个 SMV 分离物序列构建的系统进化树,它们明显聚为 3 类(图 4b),第Ⅱ、Ⅲ类与基于 RSCU 聚类构建的进化树较一致,但第Ⅰ类之间差异较大。由此可见,大豆花叶病毒的系统进化树与基于密码子偏性的聚类树总体趋势一致,但也存在一定差异,可能是单个基因在物种中的突变致使密码子偏性出现差异^[14]。

表 3 大豆花叶病毒基因组中的最优密码子
Table 3 Preferred codons in SMV genome

氨基酸 Amino acids	密码子 Codon	氨基酸 Amino acids	密码子 Codon	氨基酸 Amino acids	密码子 Codon	氨基酸 Amino acids	密码子 Codon
Phe	UUU	Tyr	UAU	Ser	UCU	Cys	UGU
	UUC		UAC *		UCC		UGC *
Leu	UUA	TER	UAA	Pro	UCA *	TER	UGA
	UUG *		UAG		UCG		UGG
	CUU	His	CAU		CCU	Arg	CGU
	CUC		CAC *	Thr	CCC *		CGC
Ile	CUA	Gln	CAA		CCA	Gly	CGA
	CUG *		CAG		CCG		CGG
	AUU	Asn	AAU	Ala	ACU	Ser	AGU
	AUC *		AAC *		ACC *		AGC *
Met	AUA	Lys	AAA		ACA	Arg	AGA *
	AUG		AAG *	Glu	ACG		AGG *
Val	GUU	Asp	GAU		GCU	Gly	GGU
	GUC		GAC *		GCC		GGC
	GUA *	Glu	GAA		GCA		GGA
	GUG *		GAG		GCG *		GGG *

* : ΔRSCU > 0.08.



a: 基于 SMV 相对密码子使用度的聚类图; b: SMV 基因序列分析进化分支图。

a: Cluster analysis dendrogram of RSCU for SMV genes; b: A dendrogram of SMV gene sequences.

图 4 SMV 基因构建的系统发育树

Fig. 4 Evolutional tree of SMV genes

3 讨论

物种在长期进化中主要受突变和自然选择的影响。生物体内密码子偏性被认为是受选择、突变和遗传漂变相互影响的结果^[15]。密码子的使用并非随机,不同物种、不同基因的密码子使用偏性具有不同特点^[14]。双生病毒偏爱使用以 A/T 结尾的密码子^[11],伪狂犬病毒偏爱使用以 G/C 结尾的密码子^[16]。本文发现 SMV 基因组 GC 含量较低,偏好使用以 A/T 结尾的密码子;ENC 值较高,密码子偏性存在但不显著。各密码子指标间相关性极显著,表明密码子碱基组成相关,其相关程度与相关原因有待研究。各绘图分析均指明 SMV 基因组的密码子在长期进化中主要受突变影响,与密码子碱基的组成密切相关,但受选择压力作用较小。

对 SMV 全基因组序列的系统进化分析以及对 CP、P1、P3 基因序列进行的系统进化分析均未能发现基因核苷酸和氨基酸序列与 SMV 株系、致病力、样

品来源地间有明显的联系^[17]。而本研究发现 SMV 基因组的碱基组成(表 1)明显分为两类,一类是来源于中国的 SMV($GC2 < GC3 < GC1$),另一类是来源于中国以外地区的 SMV($GC1 < GC2 < GC3$)。对应性分析发现,第一轴显示 67.84% 的差异,显著高于其他轴。由图 3 可知,第一轴将 57 个 SMV 区分, $-0.3 \sim 0$ 的区域主要是来自中国的 SMV 分离物, $0 \sim 0.1$ 的区域主要为美国、韩国、日本等地的 SMV。而在聚类分析时,基于 RSCU 构建的系统进化树也将 SMV 分为三大类,第 I 类主要为中国地区的 SMV,且与其他地区 SMV 距离较远。来自韩国的 SMV 分离物也具有明显的地域性,第 II 类中的韩国分离物主要来源于韩国东部,第 III 类中的韩国分离物主要分布于韩国西部^[18]。表明 SMV 密码子偏性的使用与地域有一定的相关性。章红运等^[17]对 CI 基因进行系统进化分析时发现,CI 基因核苷酸序列与 SMV 致病力和病毒来源地之间也有着明显的联系。但基于 CI 基因构建的系统进化树与基于 RSCU 构建的系统进化树,以及基于 SMV 全基因组核苷酸构建的系统进化树间仍有差

异。因此,在研究 SMV 系统进化时,需进行多指标的全面研究,弥补指标间的不足。

大范围的寄主和载体、气候和环境的改变、新型农业的兴起等都会诱导植物病毒的进化^[19]。基因突变、基因重组等是植物病毒 RNA 变异或进化的主要动力^[20]。SMV 会因寄主、传播介体和环境的选择压力,以及病毒复制过程的误差而发生突变。其中,对抗宿主基因的抗性机制诱导的突变占主导地位^[21-22]。已有报道,基因重组在双生病毒^[23]和烟草脆裂病毒属^[24]的进化中发挥直接作用。本实验室经检测发现 4547、4469-4 等多个来自中国地区的 SMV 分离物为重组 SMV^[25]。由此推测不同地区 SMV 密码子使用差异可能是在进化过程中,受寄主、环境、基因重组等的影响造成。在病毒防治方面,由于 SMV 密码子不同位置 GC 含量不同,由其决定的病毒基因组不同,它们的致病性也会不同,因此应根据当地病毒类型开展有针对性的抗病育种。另外根据病毒基因的保守序列构建高效表达或沉默载体,获得广谱抗 SMV 转基因品种。

参考文献

- [1] Chung B Y W, Miller W A, Atkins J F, et al. An overlapping essential gene in the *Potyviridae* [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105 (15): 5897-5902.
- [2] 杨春亮,王良,武斌,等. 大豆 *GmRAV* 基因的密码子偏好性分析[J]. *东北农业大学学报*, 2012, 43 (7): 34-41. (Yang C L, Wang L, Wu B, et al. Codon bias analysis of soybean *GmRAV* gene [J]. *Journal of Northeast Agricultural University*, 2012, 43 (7): 34-41.)
- [3] 武志娟,钟金城. 密码子偏性及其应用[J]. *生物学通报*, 2012, 47 (4): 9-11. (Wu Z J, Zhong J C. Codon bias and application [J]. *Bulletin of Biology*, 2012, 47 (4): 9-11.)
- [4] 续晨,蔡小宁,钱保俐,等. 葡萄基因组密码子使用偏好模式研究[J]. *西北植物学报*, 2012, 32 (2): 409-415. (Xu C, Cai X N, Qian B L, et al. Codon usage bias in *Vitisvinifera* [J]. *Acta Botanica Boreali-Occidentalia Sinica*, 2012, 32 (2): 409-415.)
- [5] Sharp P M, Li W H. An evolutionary perspective on synonymous codon usage in unicellular organisms [J]. *Journal of Molecular Evolution*, 1986, 24: 28-38.
- [6] Wright F. The 'effective number of codons' used in a gene [J]. *Gene*, 1990, 87: 23-29.
- [7] Sueoka N. Directional mutation pressure and neutral molecular evolution [J]. *Proceedings of the National Academy of Sciences*, 1988, 85 (8): 2653-2657.
- [8] Sueoka N. Two aspects of DNA base composition: G + C content and translation-coupled deviation from intra-strand rule of A = T and G = C [J]. *Journal of Molecular Evolution*, 1999, 49 (1): 49-62.
- [9] 邢朝斌,吴鹏,修乐山,等. 刺五加功能基因密码子偏好性的分析[J]. *中草药*, 2013, 44 (5): 598-603. (Xin C B, Wu P, Xiu L S, et al. Analysis on bias of functional gene codon of *Eleutherococcus senticosus* [J]. *Chinese Traditional and Herbal Drugs*, 2013, 44 (5): 598-603.)
- [10] Sharp M, Li W H. The codon adaptation index a measure of directional synonymous codon usage bias, and its potential applications [J]. *Nucleic Acids Research*, 1987, 15 (3): 1281-1295.
- [11] 徐晓忠,刘庆坡,樊龙江,等. 双生病毒的同义密码子用法及其进化分析[J]. *植物病理学报*, 2008, 38 (5): 478-488. (Xu X Z, Liu Q P, Fan L J, et al. Analysis on synonymous codon usage and evolution in geminiviruses [J]. *Acta Phytopathologica Sinica*, 2008, 38 (5): 478-488.)
- [12] 张文娟. 基于密码子水平的生物信息学分析及进化研究[D]. 上海:复旦大学, 2006: 37-47. (Zhang W J. Codon analysis and its application in bioinformatics and evolutionary studies [D]. Shanghai: Fudan University, 2006: 37-47.)
- [13] 周伟,陈洪俊,李明福,等. 基于相对密码子频率的芜菁花叶病毒基因组序列的进化分析[J]. *生物信息学*, 2010, 8 (2): 142-149. (Zhou Y, Chen H J, Li M F, et al. Evolutionary analysis of turnip mosaic virus genomic sequences based on synonymous codon usage [J]. *China Journal of Bioinformatics*, 2010, 8 (2): 142-149.)
- [14] 何闪,张淑杰,修乐山,等. 鲨烯合酶基因的密码子偏性分析[J]. *基因组学与应用生物学*, 2013, 32 (2): 232-239. (He S, Zhang S J, Xiu L S, et al. Codon usage analysis in squalene synthase gene [J]. *Genomics and Applied Biology*, 2013, 32 (2): 232-239.)
- [15] 晁岳恩,常阳,王美芳,等. 7 种作物叶绿体基因的密码子偏好性及聚类分析[J]. *华北农学报*, 2012, 27 (4): 60-64. (Chao Y E, Chang Y, Wang M F, et al. Codon usage bias and cluster analysis on chloroplastic genes from seven crop species [J]. *Acta Agriculturae Boreali-Sinica*, 2012, 27 (4): 60-64.)
- [16] 杨丽,郭万柱,殷华平,等. 伪狂犬病病毒基因组密码子用法特点分析[J]. *中国预防兽医学报*, 2007, 29 (2): 103-106. (Yang L, Guo W Z, Yin H P, et al. Analysis of characteristic codon usage in pseudorabies virus genome [J]. *Chinese Journal of Preventive Veterinary Medicine*, 2007, 29 (2): 103-106.)
- [17] 章红运. 大豆花叶病毒基因组全序列测定及大豆 *eIF4E* 和 *eIFiso4E* 基因与病毒 *VPg* 基因的互作分析[D]. 南京:南京农业大学, 2012: 28-32. (Zhang H Y. Determination of complete genomic of soybean mosaic virus and interaction analysis between soybean *eIF4E/eIFiso4E* gene and SMV *VPg* gene [D]. Nanjing: Nanjing Agricultural University, 2012: 28-32.)
- [18] Seo J K, Ohshima K, Lee H G, et al. Molecular variability and genetic structure of the population of soybean mosaic virus based on the analysis of complete genome sequences [J]. *Virology*, 2009, 393 (1): 91-103.
- [19] Chareil E R, Holmes E C. A phylogenetic survey of recombination frequency in plant RNA viruses [J]. *Archives of Virology*, 2006, 151: 933-946.
- [20] Valli A, Martinn-Hernandez A M, Lopez-Moya J J, et al. RNA silencing suppression by a second copy of the P1 serine protease of Cucumber vein yellowing ipomovirus (CVYV), a member of the family *Potyviridae* that lacks the cysteine protease HC-Pro [J]. *Journal of Virology*, 2006, 80: 10055-10063.
- [21] Choi B K, Koo J M, Ahn H J, et al. Emergence of Rsv-resistance breaking soybean mosaic virus isolates from Korean soybean cultivars [J]. *Virus Research*, 2005, 112: 42-51.
- [22] Acosta-Leal R, Bryan B K, Rush C M. Host effect on the genetic diversification of beet necrotic yellow vein virus single-plant populations [J]. *Phytopathology*, 2010, 100: 1204-1212.
- [23] Padidam M, Sawyer S, Fauquet C M. Possible emergence of new geminiviruses by frequent recombination [J]. *Virology*, 1999, 265: 218-255.
- [24] MacFarlane S. Natural recombination among plant virus genomes: evidence from tobamoviruses [J]. *Seminars in Virology*, 1997, 8: 25-31.
- [25] Yang Y Q, Lin J, Zheng G J, et al. Recombinant soybean mosaic virus is prevalent in Chinese soybean fields [J]. *Archives of Virology*, 2014, 159: 1793-1796.