

水稻条叶枯病毒(RStV)基因组组分 4 的克隆与序列分析*

曲志才 沈大棱** 邓可京 吕英芝 李昌本

(复旦大学遗传学研究所 上海 200433)

Hull R

(约翰英纳斯研究中心病毒研究系 英国)

摘 要 利用 RT-PCR 技术合成并扩增了水稻条叶枯病毒(RStV)中国云南分离物基因组组分 4 的全长 cDNA,将 PCR 产物克隆在载体 pCRII 上,并进行全序列测定,所得核苷酸序列及推测的氨基酸序列与日本分离物 T 进行比较。结果表明,在核苷酸水平,两分离物的 vORF、vcORF 及基因间非编码区序列的同源性分别为 94.9%、94.1%、86.1%,5' 端非编码区序列相同,而 3' 非编码区同源性为 96.1%,仅有两个核苷酸不同;在氨基酸水平, vORF 及 vcORF 编码蛋白的同源性分别为 99.4% 和 98.3%。可见,编码区的大小及其氨基酸序列和两末端序列都是很保守的。因此,中国云南分离物 Y 与日本分离物 T 可能有很近的亲缘关系。

关键词 水稻条叶枯病毒,水稻条叶枯病,序列分析

分类号 Q74 文献标识码 A 文章编号 0001-6209(1999)01-0036-42

水稻条叶枯病毒(Rice Stripe Virus, RStV)是一种 RNA 病毒,隶属柔丝病毒组(Tenuivirus)。RStV 通过灰飞虱(*Laodelphax striatellus* Fallen)以持久方式传播^[1]。它广泛分布于我国、朝鲜、日本及前苏联东部各国^[2],是条叶枯病的病原,在流行年份会造成水稻严重减产。目前对水稻病毒病的防治尚缺乏有效的方法,主要通过更换水稻品种、改变耕作制度及喷洒杀虫剂防治传毒介体昆虫等间接手段控制病害的流行。因此,研究病毒-昆虫介体及病毒与植物之间的关系,寻找控制和防治水稻病毒病的有效手段与方法,具有重大的理论意义与实践价值。RStV 分子生物学及转基因植物的研究为植物抗病毒育种提供了一条可能的途径。

为研究 RStV 的中国云南分离物基因组组分 4 序列和基因产物功能,我们克隆得到了 RStV RNA4 的 cDNA,并进行了全序列分析。

1 材料和方法

1.1 材料

供试 RStV 毒源材料采自云南省楚雄地区感病稻田。大肠杆菌 INV α F' 及质粒 pCRII 购自美国 Invitrogen 公司,限制性内切酶、cDNA 合成、PCR 扩增以及序列测定试剂盒主要购自 Boehringer Mannheim、Gibco/BRL、Promega 和 US Biochemical。[α -³⁵P] 或 [α -³⁵P]dATP 购自 Amersham 公司。

* 本研究由美国 McKnight 基金资助。

** 联系作者

收稿日期:1997-06-10,修回日期:1997-09-26

1.2 方法

1.2.1 病毒的分离纯化和 RNA 的抽提 :参照文献 [3]。

1.2.2 cDNA 合成及 PCR 扩增 :根据已报道的 RStV 日本分离物 T 的全序列及柔丝病毒组所表现出的末端保守序列 ,设计 PCR 扩增反应的引物 :



以纯化的 RStV RNA 为模板 ,与合成的 PCR 引物在 65℃ 下退火 ,加入 dNTP、Superscript 反转录酶 ,反转录产生病毒 RNA 的 cDNA 第一链 ,再以此为模板进行 PCR 扩增。反应条件为 :94℃ 变性 1min ,40℃ 退火 1min ,72℃ 延伸 3min ,共 30 个循环 ,最后 72℃ 延伸 10min。PCR 产物经 1% 琼脂糖凝胶电泳检测。

1.2.3 基因克隆与序列测定 将纯化后的 PCR 产物直接克隆到 pCRII 载体上 ,电击穿法转化大肠杆菌 INVαF'。在含 IPTG 和 X-gal 的 LB 羧苄青霉素平板上筛选白色菌落 ,对筛选到的重组子作酶切和多克隆位点区两端的测序鉴定 ,得到多个重组阳性克隆 ,称之为 pCR2.2。采用双脱氧核苷酸链终止法 ,按试剂盒提供的程序进行双向序列测定。

1.2.4 Northern 印迹杂交 纯化的 RStV RNA 经变性琼脂糖凝胶电泳分离 ,然后转移至尼龙膜 ,用 $\alpha\text{-}^{32}\text{P}$ dCTP 随机标记法标记重组质粒作为探针 ,与膜杂交。

1.2.5 序列分析和同源性比较 采用 GCG 软件进行序列分析 [4] ,并与 RStV 日本分离物 T 基因组组分 4 作同源性比较。

2 结果

2.1 cDNA 的合成与扩增

根据已发表的 RStV 日本分离物 T 的 RNA4 全序列及末端序列的保守性 ,设计单一引物合成了覆盖 RNA4 全长的 cDNA 第一链 ,再按所设计的反应条件进行 PCR 扩增 ,扩增产物经 1% 琼脂糖凝胶电泳检测 ,其主要产物为 2.2kb 条带 ,大小与预期值一致(图 1)。

2.2 扩增产物的克隆与 Northern 印迹分析

利用 pCRII 载体 3'端 T-overhang ,将 PCR 产物直接插入到载体中。酶切鉴定和多克隆位点两端的序列分析表明 ,重组子的阳性率高达 60%~70%。

^{32}P 标记的 cDNA 探针的 Northern 印迹杂交只显示一条主带(图 2) ,其大小与预期的 RNA4 相符。此外 ,在组分 3 和 2 之间出现一些薄弱的杂交信号 ,这可能是由于它们具有共同或相似的末端序列所造

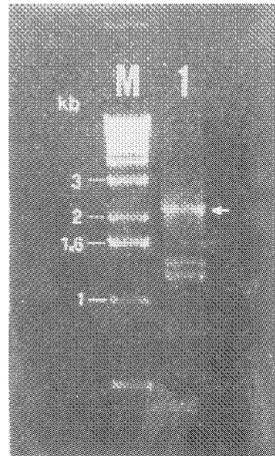


图 1 RT-PCR 产物电泳图

M ,1kb 标准 DNA ;1 扩增的 RStV 的 RNA4 全长 2.2kb 片段(箭头所示)

Fig. 1 Electrophoretic analysis of RT-PCR product
Lane M , 1 kb DNA Ladder ; Lane 1 , Amplified 2.2 kb (indicated by arrow) DNA fragment covering full-length RNA4 of RStV.

成的。

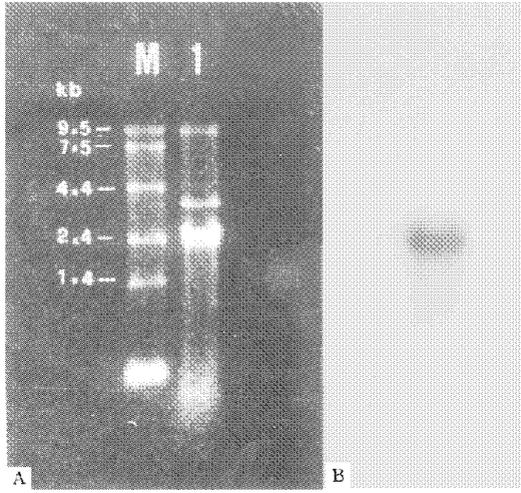


图 2 RNA4 的放射自显影图

A. 病毒总 RNA 电泳图谱。M ,RNA 标准分子 ;1 ,RStV 的四个组分。B. RNA4 的放射自显影。

Fig.2 Northern blot hybridization analysis of total denatured RStV RNAs.

A. Total RNAs of RStV extracted from the virus preparation. Lane M ,RNA ladder , Lane 1 ,Four fractionated components of RStV .B. Autoradiograph of RStV RNA4.

2.3 克隆片段的序列分析

RStV 中国分离物 Y 组分 4 的 cDNA 碱基序列分析(图 3)。

序列分析结果表明 ,cDNA 克隆 pCR2.2 含有全长的 RNA4 ,长度为 2235 个核苷酸 (图 3) ,各种核苷酸的组成是 32.21% A ,20.63% C ,18.84% G 和 28.32% U ,与日本分离物 T 的组成相似。序列分析得到两个开放阅读框(ORF) ,一个是位于 vRNA 上的 vORF ,另一个是 vcRNA 上的 vcORF(图 4) ,两个 ORF 之间是 732 个核苷酸的基因间非编码序列 ,比日本分离物 T 长 78 个核苷酸。

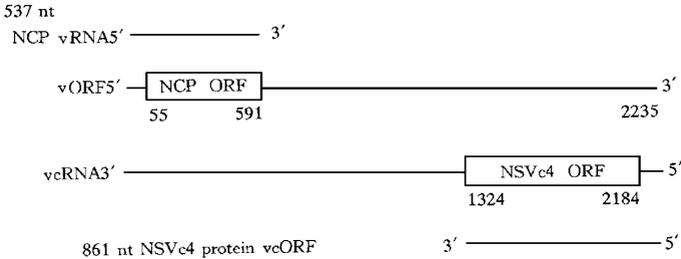


图 4 RStV 中国云南分离物 Y 基因组与表达策略

Fig.4 RStV RNA4 gene organization and expression strategy of Chinese isolate Y

非编码区 5'端非编码区 54 个核苷酸 ,不论大小还是组成与日本分离物 T 完全相同 (表 1) 。3'端非编码区 51 个核苷酸 ,同源性是 96.1% ,仅有两个核苷酸不同。

ACACAAAGTC	CAGGGCATT	GTACAACGAT	CCAGCAATTT	AATCAGAATC	GAAGATGCAA	60
					M Q	
GACGTACAAA	GGACAATAGA	AGTTTCTGTT	GGTCCTATTG	TAGGCCTAGA	TTACTACTCTG	120
D V Q R	T I E	V S V	G P I V	G L D	Y T L	
TTGTATGACA	CTCTGCCCGA	GACTGTGAGC	GATAATATAA	CTCTACCTGA	TTTGAAAGAT	180
L Y D T	L P E	T V S	D N I T	L P D	L K D	
CCAGAGAGAG	TCACGGAAGA	TACTAAGAAG	CTAATACTCA	AAGGCTGTGT	TTATATAGCA	240
P E R V	T E D	T K K	L I L K	G C V	Y I A	
TATCATCACC	CCTTGGAGAC	TGACACTCTT	TTCATTAAG	TTCACAAGCA	TATACCAGAG	300
Y H H P	L E T	D T L	F I K V	H K H	I P E	
TTTTGTCACT	CCTTCCTATC	ACACCTTCTA	GGAGGTGAAG	ATGATGACAA	TGCTCTGATA	360
F C H S	F L S	H L L	G G E D	D D N	A L I	
GACATTGGCC	TGTTCTTCAA	CATCTTGCAA	CCCTCTCTGG	GCGGTTGGAT	AACCAAGAAT	420
D I G L	F F N	I L Q	P S L G	G W I	T K N	
TTTCTTAGAC	ACCCAAACAG	GATGTCTAAG	GACCAAATTA	AAATGCTCCT	GGATCAGATT	480
F L R H	P N R	M S K	D Q I K	M L L	D Q I	
ATAAAGATGG	CTAAGGCAGA	GAGCTCAGAC	ACAGAAGAGT	ATGAGAAAGT	GTGGAAGAAG	540
I K M A	K A E	S S D	T E E Y	E K V	W K K	
ATGCCAACTT	ATTTGAGTC	AATCATCCAA	CCTCTTCTAC	ACAAAACATA	GTTTATTGCA	600
M P T Y	F E S	I I Q	P L L H	K T *		
TATCATATAG	TCCACCCTAC	GAGTGTCTGC	ATTCTAGGAG	TGAAGCCAGT	TTGGGTTGAC	660
CTGCTAGTGT	AAAAATATAA	GTATATTATA	GTTTAAAGAGA	TATACACGAT	GTTTGTTCAC	720
TTGAGCTTGG	CAGTTTGCAA	<u>GTAATTGTGT</u>	<u>GTAATTGTGC</u>	ATAAAATATAT	ATATGTCAAT	780
ATATAACTAT	ACATAGGTGT	GCAGATGTAA	ACACTTGAAA	CAAATATACA	CCTATCAACT	840
TATATACACA	GCTTGTAAGT	CTTGGTTCTT	CTTAACACAA	ACATAGAAAC	ATGGGAGTAT	900
<u>CAATTAGAAT</u>	<u>TGCCTTGGTG</u>	<u>TTATAAGCAC</u>	<u>AGATATAGCT</u>	<u>GTATGTAGAG</u>	<u>AACACTCAAG</u>	960
<u>GCTAAGAATT</u>	<u>CTCTAACATC</u>	GCTTTATACA	AAAACCATTA	TAAAATACT	AACAACTTTG	1020
TTTTACAAT	GTCAACGCTA	AAAAGTAAA	TGCCAAAAAC	ATGAGAAAAT	AGAAAATCAA	1080
AAACAATGAA	TGGTGCTAAG	CACCACATCC	GGATGTGGTG	CGTAGCACCA	TTTTCATAGT	1140
ATTGATCTAT	AACCTACTTG	TAGCATAAAC	ACCCACATTC	ATGAAGTGCA	TAGCACTCTC	1200
TTATGGATTG	CCCCTATGA	TTACATGAAT	GCATAGCATC	CTCATGGAAA	AAATCAGAGC	1260
AGAACAGACC	ACATCAGATG	CTTGGCACCA	GATGTTAGCA	TACCCTGGAG	CTAGTTATCT	1320
CACTTACATA	ATAACAGAAA	CTTCAGATTT	TGATTTCTTT	GTGATGGGTA	TCTTCTTTGG	1380
* M I	V S V	E S K	S K K T	I P I	K K P	
CTGCTTCACC	ACACCGAACT	CCTTTTCAGT	GTCTAGCTCA	GGGAAGCTCT	TGCTGGTGAA	1440
O K V V	G F E	K E T	D I F N	S F V	S T F	

AGCACCGCCA	ACAGCTTTCT	GAACGACATT	AGTCGTGAAG	TTACTGAAGC	TAGACAAATG	1500
A G G V	A K Q	V V N	T T F N	S F S	S L H	
ATTTATTGCC	TTATCTTCTA	GTGGTTCAAA	CTCAAACATC	ACGGTGTCCAC	TGGTCTTCAT	1560
N I A K	D E L	P E F	E F M V	T D S	T K M	
GGATACCGGC	AGATCAGTTC	TTTCAATCCC	CCAGAACCAA	AGAGTTCTAG	AGATGACACA	1620
S V P L	D T R	E I G	W F W L	T R S	I V C	
GTCTGTACT	GAGCTGTCAT	CAACAGACAC	AGACACCTGA	AGATTATGCT	TATCCTCTAA	1680
N Q V S	S D D	V S V	S V Q L	N H K	D E L	
AGCCAGAAAA	TTGCTAGAG	AACCCAAGAC	GGCAAAATTC	TTGCTAATTG	GGTAGCGAAC	1740
A L F N	A L S	G L V	A F N K	S I P	Y R V	
TTCTACCTCA	ACTTGATCAG	AGGGGTTCCAC	ATAGGACTTG	TCTATTATGC	GGAGAGTAGT	1800
E V E V	Q D S	P N V	Y S K D	I I R	L T T	
TACTCCACTA	GCTCTACCCT	TAATTCCAAT	CCAGACCAA	GCAAAAGTAG	CCACTCTGAA	1860
V G S A	R G K	I G I	W V L A	F T A	V R F	
GAATGGATAA	TGAGTGAGAG	GTTGATGAAA	CCAATATGGG	TCAACAAGTA	TGTTGAAAGT	1920
F P Y H	T L P	Q H F	W Y P D	V L I	N F T	
TGCCTTCTA	TGTGCCITCA	CAGAATAGTC	CTGTTTGGCT	ATGAACATAT	CATACTTGTT	1980
A K R H	A K V	S Y D	Q K A I	F M D	Y K N	
AACCTTGACA	TCTGAGAAGC	TGAAAGGCTC	CAATCCGAGC	ATGTAGCAG	CCTGGTCAAT	2040
V K V D	S F S	F P E	L G L M	T A A	Q D I	
TGTGACCCTA	CCCTGGITTA	GAGGCTCTT	GGAGAGAGCT	AGAGACTICC	TATTCTTATT	2100
T V R G	Q N L	P R K	S L A L	S K R	N K N	
ATCAACTCTC	TTTTGAGACT	CCTCACTAAG	GTCATCATAG	AGTACCTTAC	TTTTTGAAGT	2160
D V R K	Q S E	E S L	D D Y L	V K S	K S T	
GGACAAAAGT	CGAGACAAAG	CCATAATTAA	GTTATATATT	AGCTTAATCT	CAAAAGATAT	2220
S L L R	S L A	M				
GCCCTGACTT	TGTGT					2235

图3 RStV中国分离物Y组分4的cDNA碱基序列及氨基酸序列。

5'端55~591处为vORF,592~1323处为基因间非编码区,3'端1324~2184处为vcORF。

下划直线示插入序列,波线示正向重复序列,*示终止密码子。

Fig. 3 Complete nucleotide and amino acid sequence of ssRNA 4 of RStV isolate Y.

Nucleotides are numbered from the 5' terminus of vRNA. The nucleotides 55 to 591 is vORF.

The vcORF encoded within nucleotides 1324 to 2184. In the intergenic region directly repeating sequence is underlined, — Indicates position of insertion mentioned in the text.

The asterisk (*) indicates the (UAG) stop codon on vORF or (UAA) on vcORF.

基因间非编码区富含核苷酸A和U,其中35%A和29%U,与日本分离物T的同源性是86.1%。在897~980位有84个核苷酸的插入,其中包含一个EcoRI切点。为证实该插入序列的存在,当用EcoRI消化RNA4 cDNA时,只有插入了RNA4 cDNA片段的重组子,琼脂糖凝胶电泳可检测到1.3和0.9kb两条带(图5泳道6)。没有插入片段的则检测不到这两条带(图5泳道1~5)。可见,该插入序列确系自然发生而并非矫作物。计算机分析表明,在1088~1133位,可形成一茎环结构。类似的结构也见于日本分离物T。

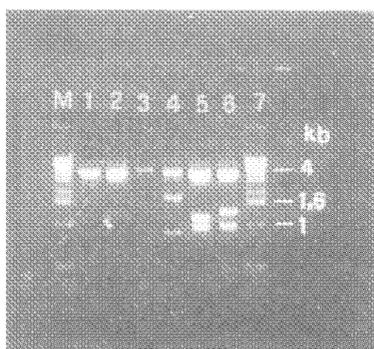


图 5 含有 RNA4 cDNA 的重组质粒 pCR2.2 的限制酶切图谱。

M 和 7.1 kb 标准 DNA ;1~6 不同重组子和 6 得到的 1260 和 970bp 二个片段

Fig.5 Electrophoretic analysis of pCR2.2 recombinants digested with EcoRI enzyme.

Lanes M and 7.1 kb DNA ladder ;Lanes 1 to 6 Different recombinants and Lane 6 gives rise to two fragments , 1260 and 970 bp respectively.

编码区 :位于 vRNA4 上的 vORF 全长 537 个核苷酸 ,其大小与分离物 T 相同。同源性分析表明 ,两分离物 RNA4 片段具极高的序列同源性。核苷酸序列的同源性为 94.9% ,由此推导的氨基酸序列同源性为 99.4%(表 1)。在 27 个核苷酸变异中 ,其中 26 个是沉默突变 ,即无义突变 ,唯一一个是高保守变异 ,由分离物 T 中的密码子 AUG 变成 AUC 不过 ,AUC 编码的氨基酸 Ile 与 AUG 编码的 Met 仍属于同型氨基酸。

位于 vcRNA4 上的 vcORF 全长 861 个核苷酸 ,其大小与分离物 T 也相同。核苷酸序列的同源性为 94.1% ,推测编码蛋白的氨基酸同源率为 98.3%。在 51 个核苷酸变异中 ,其中 46 个是沉默突变 ,4 个是保守变异 ,只有一个可引起氨基酸的变异。

表 1 中国分离物 RStV 组分 4 的基因组分析及其与日本分离物 T 的比较

Table 1 Genome analysis of RNA4 of the Chinese isolate and its comparison with the corresponding regions of Japanese isolate T

区段 Region	产物 Product	起始 Start	终止 Stop	分子量 Mol. mass /kD	氨基酸数 Length amino acid	核苷酸同源性 Identity of nucleotide /%	氨基酸同源性 Identity of amino acid /%
5'UTR ^a	-	1	54	-	-	100	-
vORF	NCP	55	591	19.8	178	94.9	99.4
IR ^b	-	592	1323	-	-	86.1	-
vcORF	NSvc4	1324	2184	31.8	286	94.1	98.3
3'UTR	-	2185	2235	-	-	96.1	-

^a 非翻译区

^b 基因间非编码区

3 讨论

Taq 聚合酶的不依赖模板活性 ,使 PCR 循环之后的最后延伸得以在扩增产物的 3'端加上 A-overhang ,利用 pCRII 载体的 3'端 T-overhang ,可以快速、有效的将 PCR 产物克隆进 pCRII 载体中。结果表明 T-overhang 克隆策略是行之有效的和高重复性的克隆技术。

两分离物序列同源性比较的一个显著特点是序列变异的不对称性。尤以 5'和 3'端非编码区保守性最高 ,基因间非编码序列变异性最大 ,而编码区变异性介于二者之间。这

意味着低保守区对其功能的影响是不重要的,而高保守区则不同,尤其是两末端的保守序列对病毒 RNA 的转录、复制以及蛋白质的翻译是至关重要的^[1]。两分离物序列同源性的比较也显示位于基因间非编码区的 84 个核苷酸的插入并未改变其稳定的发夹结构。类似的插入也见于具有双义编码策略的番茄斑萎病毒(TSWV)^[5]。可能这些基因间非编码区具有复杂的高级结构,对突变敏感,因而其二级结构并非核苷酸序列是病毒活性所必需的。需要指出的是两分离物之间的序列多样性是否与它们的生物学特性(包括条叶枯病症)有关,尚需进一步研究。

参 考 文 献

- [1] Ramirez B C, Haenni A L. *Journal Gen Virol*, 1994, **75**: 467-475.
 [2] Toriyama S. *Microbiological Sciences*, 1986, **3**: 347-351.
 [3] Qu Z C, Harper G, Hull R. *Virus Genes*, 1997, **15**(2): 99-103.
 [4] Devereux J, Haerberli P, Smith O. *Nucleic Acids Research*, 1984, **12**: 387-395.
 [5] de Haan P, Wagemakers L, Peters D *et al.* *Journal Gen Virol*, 1990, **71**: 1001-1007.

CLONING AND SEQUENCING ANALYSIS OF RICE STRIPE VIRUS GENOME SEGMENT 4 OF CHINESE ISOLATE Y*

Qu Zhicai Shen Daleng** Deng Kejing Lu Yingzhi Li Changben

(*Institute of Genetics, Fudan University, Shanghai 200433*)

Hull R

(*Department of Virus Research, John Innes Centre, Norwich, UK*)

Abstract The cDNA fragment covering full-length sequence of RStV RNA4 of Yunnan isolate in China was obtained by RT-PCR. The PCR-derived fragment was then cloned into vector pCRII. The cloned cDNA was sequenced. Comparison of the nucleotide and deduced amino acid sequences with those of the Japanese isolate T was made. The results showed that at the nucleotide level, vORF, vcORF and the intergenic region had 94.9%, 94.1% and 86.1% identity respectively, the 5'-untranslational region was exactly the same as that of Japanese isolate T, while the 3'-terminal sequence had 96.1% identity, differing by two nucleotides; at the amino acid level, vORF and vcORF had 99.4% and 98.3% identity respectively. Therefore, as well as being exactly the same size for the two isolates, the amino acid sequences of the coding regions and the 5'- and 3'-terminal sequences were well conserved. Our results indicated that the Chinese isolate is closely related to the Japanese isolate T.

Key words Rice stripe virus, Rice stripe disease, Sequencing analysis

* This work was supported by McKnight Foundation.

** To whom correspondence should be addressed.