

# 黄瓜花叶病毒山东株(CMV-SD)RNA2 全长 cDNA 克隆的 构建及全序列分析\*

张国华 许 燕 蔡文启 方荣祥

(中国科学院微生物研究所植物生物技术开放实验室 北京 100080)

**摘 要** 分别利用 5' RACE 和 3' RACE 确定了 CMV-SD RNA2 的 5' 和 3' 末端序列,在此基础上,利用 RT-PCR 得到了 RNA2 的 5' 端一半的 cDNA 克隆 pC25 和 3' 端一半的 cDNA 克隆 pC23,并通过拼接构建了 RNA2 全长 cDNA 克隆 pC2F。通过对 pC25 和 pC23 进行序列测定,得到了 RNA2 的全序列。序列分析结果表明 CMV-SD RNA2 由 3048nt 组成,其中存在 2 个部分重叠的阅读框 ORF1(79~2652nt)和 ORF2(2414~2746nt),分别编码 858aa 的 2a 蛋白和 111aa 的 2b 蛋白,并在 2a 蛋白的序列中发现了动植物病毒复制酶所特有的两个保守序列。该株系 RNA2 核苷酸序列与属 CMV I 亚组的 F<sub>ny</sub> 株系和 II 亚组的 Q 株系 RNA2 的核苷酸序列同源性分别为 91.7% 和 75.6%;2a 蛋白的氨基酸序列同源性分别为 93.8% 和 67.7% 2b 蛋白的氨基酸序列同源性分别为 83.0% 和 51.3%。同源性比较的结果表明 SD 株系属于 CMV I 亚组。

**关键词** 黄瓜花叶病毒山东株, RNA2 全长 cDNA 克隆, 序列分析

**分类号** S432.4 **文献标识码** A **文章编号** 0001-620X(1999)01-0008-14

黄瓜花叶病毒(CMV)是黄瓜花叶病毒属的典型成员,为三基因组分、正链 RNA 病毒<sup>[1,2]</sup>。研究表明,CMV RNA1、2 分别编码与复制有关的 1a、2a 蛋白<sup>[3]</sup>,RNA3 编码 3a 移动蛋白和外壳蛋白,外壳蛋白是通过由 RNA3 产生的亚基因组 RNA4 表达的<sup>[4]</sup>。CMV 是寄主最多、分布最广、对农作物危害最为严重的植物病毒之一。为了从分子水平上进一步深入研究 CMV 基因组的表达调控机制和致病机制、开辟新的抗病毒策略,我们合成并克隆了一株从山东大田感病烟草上分离得到的 CMV 分离物(CMV-SD)的全部基因组 cDNA,并进行了全序列测定。本文报道 CMV SD 株 RNA2 全长 cDNA 克隆的构建及其全序列分析结果。

## 1 材料和方法

### 1.1 病毒来源

CMV SD 株系由山东大田感病烟草中分离得到,并一直由我们实验室繁殖保存。

### 1.2 病毒及其 RNA 的提纯

用 CMV-SD 毒源接种三生烟(*N. tabacum var. Xanthi-nc*),14 天后采集带有花叶症状的顶部未接种叶片,参照 Franck<sup>[5]</sup>等综述的方法提纯病毒,即首先用含 0.1% 巯基乙醇的 0.5mol/L 柠檬酸钠(pH7.0)缓冲液进行匀浆、离心后得到的粗汁液经氯仿澄清、聚

\* 国家自然科学基金(No. 39400074)和 863 资助项目(No. 863-101-04-02-01)

本文报道的核苷酸序列被贮存于 GenBank/EMBL/DBJ 数据库,接受号 D86330

收稿日期:1997-09-11,修回日期:1997-11-29

乙二醇 6000 沉淀病毒后,再经 1.5% TritonX-100 处理及 10%~40% 蔗糖梯度离心得到提纯病毒。提纯病毒与等体积 2×RNA 提取液(50 mmol/L Tris-HCl pH7.6, 1 mmol/L EDTA, 50 mmol/L NaCl, 1% SDS)混合,先后用 TE 饱和酚及酚/氯仿(1:1)各抽提两次,经乙醇沉淀得到病毒 RNA,用焦碳酸二乙脂处理过的灭菌水溶解后置于 -70℃ 备用。

### 1.3 5' RACE

CMV RNA2 5' 末端区域利用 5' RACE(Rapid Amplification of cDNA Ends)法进行克隆,具体操作基本按 GIBCO BRL 5' RACE System 试剂盒说明进行。所用引物 SP1(5' TTCCCGGGTACATCGCGAGA3')、SP2(5' CGATATGCATGGGTTTGACC3')参照已发表的 CMV-Fny 株系(CMV-Fny)RNA2 的序列<sup>[6]</sup>设计合成。首先以纯化的 CMV 总 RNA 为模板,用引物 SP1 进行反转录反应,合成第一链 cDNA。经 RNase H 处理、Glass MAX DNA isolation spin cartridge 纯化后,用末端转移酶在合成的第一链 cDNA 3' 端加 poly(G) 尾。用引物 SP2 和 NS(5' TTGGTACCCCCCCCCCCCCCCC3')进行 PCR 扩增。反应程序为 94℃ 3min,94℃ 1min,50℃ 1min,72℃ 1min,30 个循环,72℃ 10min。PCR 产物经纯化后,克隆于 pGEM-T Vector(Promega)上,并进行序列分析。

### 1.4 3' RACE

CMV RNA2 3' 末端区域利用 3' RACE 法进行了克隆。首先在 CMV RNA 的 3' 端加 poly(A) 尾,反应条件参照 poly(A) polymerase(Pharmacia)的使用说明设定。50μL 的反应体系含有 50mmol/L Tris-HCl, 10mmol/L MgCl<sub>2</sub>, 2.5mmol/L MnCl<sub>2</sub>, 250mmol/L NaCl, 0.5ng/mL BSA, 0.3mmol/L ATP, 40U RNA 酶抑制剂, 5U poly(A) polymerase, 5μg CMV 总 RNA, 在 37℃ 反应 60min。反应产物用酚/氯仿抽提两次,乙醇沉淀,沉淀溶于 10μL TE 溶液中。取 5μL 加 poly(A) 尾 RNA 为模板, 1μL 引物 BT17(10pmol/μL), 加 4μL 5×反转录缓冲液, 1μL RNA 酶抑制剂(40U/μL), 2μL AMV 反转录酶(10U/μL), 42℃ 反应 60min, 合成第一链 cDNA。取 1/10 体积的反转录反应体系进行 PCR 扩增(条件同上), 所用引物为 BT17(5' GAGGATCCTTTTTTTTTTTTTTTTTTTT3') 和参照 CMV-Fny RNA2 的序列<sup>[6]</sup>设计合成的引物 C3T(5' CGTTCAGATCTCAGGCTGTT3')。PCR 产物经纯化后克隆于 pGEM-T Vector(Promega)上, 并进行序列分析。

### 1.5 RNA2 的全长 cDNA 合成和克隆

由于 CMV RNA2 比较长,很难一次得到全长 cDNA 克隆,我们采取了分两段进行 cDNA 合成、PCR 扩增和克隆的策略(图 2)。根据上述 5' RACE 和 3' RACE 的实验结果设计合成了位于 5'、3' 末端的引物 P4、P5, 位于内部的引物 P7(5' GACTTGTC-GAATTTGGAC3')、P8(5' AGAGATGCCTTCGAGAAC3') 根据我们已测定的 CMV-SD RNA2 的部分序列(未发表)并参照 CMV-Fny RNA2 的序列<sup>[6]</sup>设计合成。首先分别用引物 P5、P7 反转录合成 3' 端和 5' 端一半第一链 cDNA, 反应条件与上述 3' RACE 时的条件相同。分别以上述第一链 cDNA 为模板, 用引物 P5 和 P8, P4 和 P7 进行 PCR 扩增(延伸时间为 2min, 其余条件同上) 并克隆到已除去 Xho I 位点的 pBluescriptSK(+) 质粒上, 通过菌落 PCR 的方法筛选阳性克隆。

### 1.6 DNA 序列分析

DNA 手工测序按 Sequase Version 2.0(USB)试剂盒说明进行,自动测序使用 ABI 公

司自动测序仪 Model373A 及试剂。序列分析采用 DNASIS 软件。如没有特别说明,序列均为对 2 个独立的克隆进行测序的结果。

## 2 结果和讨论

### 2.1 CMV RNA2 末端序列的测定

CMV RNA2 5' 和 3' 末端的序列分别采用 5' RACE 和 3' RACE 方法进行了测定。通过 5' RACE 筛选到了多个含有相应 PCR 扩增片段插入的克隆,对其中的 4 个克隆进行了 DNA 序列测定。结果表明,除 1 个克隆 5' 端不完全外其余 3 个克隆均是全长的,其中 2 个克隆的 5' 端是由 GG 开始的(图 1a),第一个碱基 G 被认为是与 RNA 5' 端帽子结构中的 G 相对应。上述 4 个克隆相应位置的序列完全相同。根据所测定的序列设计合成了 PCR 扩增 CMV RNA2 5' 端片段的引物 P4(5' C CTGCAGTTTATTTACAAGAGC3'),在其 5' 端加入了一个 Pst I 位点(下划线处)。

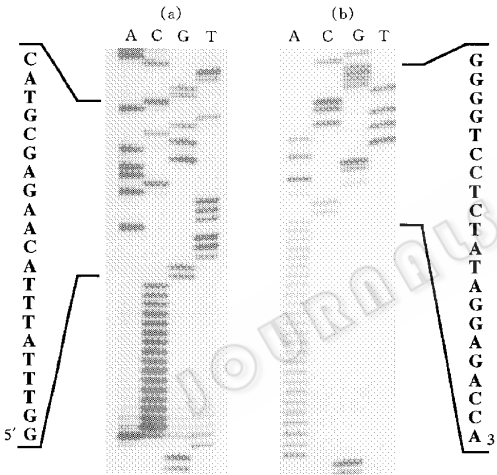


图 1 利用 5' RACE 和 3' RACE 确定 CMV-SD RNA2 末端序列  
a. 5' RACE b. 3' RACE

Fig. 1 Terminal nucleotide sequences of CMV-SD RNA2 determined by 5' RACE and 3' RACE

图 1b 显示的是 3' RACE 的测序结果,所测定的 3 个克隆的序列完全相同。由于植物 RNA 病毒的 3' 末端一般都是以 CCA 结尾的,因此图 1b 中的 poly(A)最上端的碱基 A 应该是 CMV-SD RNA2 3' 末端的第一个碱基。根据上述 3' RACE 的测序结果设计合成了 PCR 扩增 CMV RNA2 3' 端片段的引物 P5(5' C GGATCCTGGTCTCCTATA3'),在其 5' 端加入了一个 BamH I 位点(下划线处)。

### 2.2 CMV RNA2 全长 cDNA 克隆的构建

经菌落 PCR 筛选得到的 5' 端片段的 cDNA 克隆 pC25 和 3' 端片段的 cDNA 克隆 pC23。pC25 的 Pst I-Xho I 双酶切小片段与 pC23 的 Pst I-Xho I 双酶切大片段连接,转化后经菌落 PCR 筛选及酶切鉴定最终得到了含有 CMV RNA2 全长 cDNA 的克隆 pC2F。CMV RNA2 全长 cDNA 克隆的获得将为下一步建立 CMV RNA 体外转录体系,在体外

合成具有生物活性的 CMV RNA2,进一步深入研究 CMV RNA2 的功能奠定了基础。

### 2.3 CMV RNA2 的全序列测定和分析

在对 pC25 和 pC23 进行酶切图谱分析的基础上构建 5 个亚克隆,并进行了序列测定。测序结果表明(图 2),CMV-SD RNA2 全长 3048 nt,其基因结构与已测序的其他株系的 CMV RNA2 相同<sup>[6-8]</sup>。CMV RNA2 全序列测定策略如图 3 所示。通过用 DNASIS 软件对所得序列分析发现在 5' 端有一个大的阅读框 ORF1(79~2652nt),编码一个 858 aa 的 2a 蛋白,在其序列中也发现了动植物病毒复制酶的保守序列 TC<sub>3</sub>xxxTxxxNT 和

GUUUUUUUACAAGAGCGUACGGUUUCAAUCCUGCCUCCCCUGUAAAAACUCCCUAGACUUU 60  
 CUAGUUUCUUUCUUUCCUAUGGCUUUUCCCGCCCCCGCUUUCACUAGCCAAUCUUUUG 120  
 ORF1 M A F P A P A F S L A N L L  
 AACGGUAGUUACGGGUGCGACACUCCCGAGGAAGUGGAACCGGUGCGAUCUGAGCAACGC 180  
 N G S Y G V D T P E E V E R V R S E Q R  
 GAAGAGGCUGCUGCGGCCUGUCGUAAUUACAGGCCUUACCCGUGUGGAUGCCAGCGAG 240  
 E E A A A A C R N Y R P L P A V D A S E  
 AGUGUCCAGAGGAUGUGCAUUCUCCGAACUCCAGACGGAGCUCUCCCGAAGAGGUG 300  
 S V P E E D V H S L R T P D G A P S E E V  
 UCUGUAGAGUUVUGUCACUUUAUGGUGCUGAAGAUUACCUUGAAAAAUCUGAUGAUGAGCUC 360  
 S V E F U V T Y G A E D Y L E K S D D E L  
 CUUGUCGCUUUUGAGACGAUGGUCAAACCCAUGGUAUCGGACAACUGUGGUCUCCGCG 420  
 L V A F E T M V K P M R I G Q L W C P A  
 UUUAAUAAAUGUUCUUUUUUUUCAGCAUUGCCAUGGCCAGGGCUUUGCUUUUAGCACCC 480  
 F N K C S F I S S I A M A R A L L L A P  
 AGAACAUCCACCGAACCAUGAAGUGUUUUGAGGACCUGGUCGCGGCUUUUUACACUAAA 540  
 R T S H R T M K C F E D L V A A I Y T K  
 UCCGAUUUCUACAACGAUGAUGAGUGUGAAAACCGACGCUUCAGAUAGAUUUCUGUCU 600  
 S D F Y N D D E C E T D D V Q I D I S  
 CGCGAUGUACCCGUAUUUCUUUCGAAACCGGUCUCCCGAAGCUGGAUUUGAACCACCA 660  
 R D V P G Y S F E P W S R T S G F E P P  
 CCUAUCUGUGAGGCGUGUGACAUGAUGAUGGACCAGUGCCCGUGUUUGAUUUCAAUGCU 720  
 P I C E A C D M I M D Q C P C F D F N A  
 UUAAGAAAUCGUGCGCUGAGAGGACAUUCGUGAUGAUUAUGUUAUCGAAGGUUUAGAU 780  
 L K K S C A E R T F A D D Y V I E G L D  
 GGUGUUGUAGACAAUGCGACUCUGUUGUCGAAUUUAGGCCCAAUUUUGGUGCCCGUGAAA 840  
 G V V D N A T L L S N L G P I L V P V K  
 UGUCAGUAGAGAAAUGTCAACCCACCCGUAUUGCGAAUUCUCCGAGUCUUAUUCUGUCU 900  
 C Q Y E K C P T P T V A N P P S L N R A  
 ACUGAUCGUGUUGAUUCAUUUAGUUCAAUCCAUUUGUGACUCGACUCUGCCCACUCAU 960  
 T D R V D I N L V Q S I C D S T L P T H  
 AGUAACUACGACGACUUUUUCAUCAAGUGUUCGUCGAGAGUGCUGACUAUUCUAUUGAU 1020  
 S N Y D D S F H Q V F V E S A D Y S I D  
 CUGGAUCUGUUAGACUUCGCCAGUCUGAUCUUAUUGCAAAAAUCCAGAUUCAGGGCAU 1080  
 L D H V R L R Q G S D L I A K I P D S G H  
 AUGAUACCGUUUUUGAACCCGGAGCGGUCACAAGAGUAGGUAACGACGAAGGAGGUC 1140  
 M I P V L N T G S G H K R V G T T K E V  
 CUUACAGCUAUCAGAAACGUAAUGCUGACGUUCCAGAGCUAGGUGAUUCCGUUAAUCUG 1200  
 L T A I K K R N A D V P E L G D S V N L  
 UCCAGAUUGAGUAAGGCUGUGGCUGAGAGAUUUUCAUUUCAUUAUUAUGGUGACUCU 1260  
 S R L S K A V A E R F F I S Y I N G D S  
 CUAGCAUCCAGCAACUUUGUUAAUGUCGUUAGCAACUCCACGAUUAUUAUGGAAAAAUGG 1320  
 L A S S N F V N V A V S N F H D Y M E K W  
 AAGUCCUAGGUCUUUCGUAUGAUGAUCUUCGGAUUCUUAUGCUGAAAAGUUGCAGUUU 1380  
 K S S G L S Y D D L P D L H A E K L Q F  
 UAUGAUCACAUGAUAAAAUCUGAUGUGAAAACCGUAGUGAGCGACACACUCAACAUUGAC 1440  
 Y D H M I K S D V K P V V S D T L N I D  
 AGACCGGUUCCGGCUACUAUACGUAUCAUAGAAAGGUUAUACCUCCAGUUCUCACCG 1500  
 R P V P A T I T Y H K K G I T S Q F S P  
 UUGUUUACUGCGCCAUUUGAGCGUUUCCAGAGAUGCCUUCGAGAACGCAUUUAUUCUCCU 1560  
 L F T A A P F E R F Q R C L R E R I I L P  
 GUUGGAAAGAUUCAUCCUUGAGAUUGGCAGGAUUUGAUGUUAAGAACAAGCACUGCUC 1620  
 V G K I S S L E M A G F D V K N K H C L  
 GAGAUUAAUUUGUCCAAAUUCGACAAGUCUCAAGGUGAAUUCUAUUAUUCAGGAA 1680  
 E I N L S K F D K S Q G E F H L M I Q E

CACAUUUUGAAUGGUCUAGGGUGCCCAGCUCGCCGAUAACUAAGUGGGUGUGAUUUCCAC 1740  
 H I L N G L G C P A P I T K W W C D F H  
 AGAUUCUCUUAUUAUAGAGACCCGUAGAGCUGGCGUUGGUAUGCCCAUCAGUUUCAGAGA 1800  
 R F S Y I R D R R A G V G M P I S F Q R  
 CGAACUGGUGAUGCAUUCACCUACUUUGGUAACACCAUCGUCACCAUGGCUGAAUUUCCA 1860  
 R T G D A F T Y F G N T I V T M A E F P  
 UGGUGUUACGACACCGACCAAUUCGAAAAAGCUUUUAUUCUCAGGGCGAUGAUUCUCUAGGA 1920  
 W C Y D T D Q F E K L L F S G D D S L G  
 UUUUCAGUACUCCUGUUGGUAUCGAGUACCGAGUAAGUUCACGACUUCUUCACAAUGGAA 1980  
 F S V L P P V G D P S K F T T L C F N M E  
 GCUAAGGUGAUGGAGCCCGCAGUACCAUAUAUUUGUUCGAAGUUCUUCUCUGACGAG 2040  
 A K V M E P A V P Y I C S K F L L S D E  
 UUCGGUACACUUUUUCCGUUCCAGACCCAUUGCGGAGGUUCAGCGGUUAGGCACGAAG 2100  
 F G N T F S V P D P L R E V Q R L G T K  
 AAAUCCCUAUUCAGACAACGAUGAAUUCUGUUGCUCACUUCUUCAGAGCUUUGUUGAU 2160  
 K I P Y S D N D E F L F A H F M S F V D  
 CGAUUGAAGUUUUUGGACCGAAUGACUCAGUCUUGUAUCGAUCAGCUUUCUCUUCUUCUU 2220  
 R L K F L D R M T Q S C I D Q L S L F F  
 GAGUUGAAAUCAGGAAGUCUGGGGUGAGGCUUUUAUGUUAAGGCGCCUUUAAGAAA 2280  
 E L K Y R K S G A E A A L M L G A F K K  
 UAUACCGCUAAUUUCAGUCCUAUAAAAGAACUCUACUUAUCAGAUUCGUCGUCAGUGCGAA 2340  
 Y T A N F Q S Y K E L Y Y S D R R Q C E  
 UUGAUCAAUUCGUUUAGUUGUGUAGAGUUAAGGAUUGAGCGUUCGAGUCCACUAAGCAA 2400  
 L I N S F S C V E L R I E R S S S T K Q  
 CGAAAGAAGAAUGGAAUUGAACGAAGGCGCAAUGACAAAACGUCGAACUCCAACUGGC 2460  
 R K K K D G I E R R R N D K R R T P T G  
 ORF2 M E L N E R G A M T N V E L Q L A  
 UCGUAUAGUGGAGGCGAAGAGGCAGAGACGAAGGUCUCACAAGCAGAAUCGACGGGAACG 2520  
 S Y S G G E E A E T K V S Q A E S T G T  
 R I V E A K R Q R R R S H K Q N R R E R  
 AGGUCACAAAAGUCCAGCGAGAGGGCGGUUCAAAUCUCAGGCUGUUCGCUUCCUACC 2580  
 R S Q K S Q R E G A F K S Q A V P L P T  
 G H K S P S E R A R S N L R L F R F L P  
 AUUCUAUCAAGUAGAUGGUUCGGAACUGAUAGAGAUGUACCACCAUGUGAACAUUGGUGG 2640  
 I L S S R W F G S T D R D V P P C E H G G  
 F Y Q V D G S E L I E M Y H H V N M V G  
 AUUGUCCGAGUCUGAGGCCCUUGUUUACGUUGCCAGCGGAAGAAGACCAUGAUUUCGA 2700  
 I V R V \* >2a  
 L S E S E A P C F T L P A E E D H D F D  
 CGAUACCGAUUGGUUCGUGGUAACGAGUGGGCGGAAGGUCGUUUCUGAAACCUCCCUU 2760  
 D T D W F A G N E W A E G R F \* >2b  
 CCUAUCUCCUCCGGUUUCUGAGGCGGGAGCUGAGUUGGCAGUACUGCUACAAACUGU 2820  
 CUGAAGUCACUAAAAGUUUAAAACGUGAACGGGUUGUCCAUUCAGCUAACGGCUAAAAU 2880  
 GGUCAGUCUGGAGAAAUCACGCGAAGGCUUAACAAGUUCUUCGAGACACCUUUGAAACC 2940  
 AUCUCCUAGGUUUCUUCGGAAGGACUUCGGUCCGUGUACUUCUAGCACAACGUGCUAGU 3000  
 UUAGGGUACGGGUGUCCUUCGCUUUGCGGGUCCUCUAUAGGAGACCA 3048

图2 CMV-SD RNA2 的核苷酸全序列及推导的 ORF1 和 ORF2 的氨基酸序列  
下划线处为病毒复制酶的保守序列。

Fig.2 Complete nucleotide sequence of CMV-SD RNA2 and deduced amino acid sequences of ORF1 and ORF2

The consensus amino acid sequences of viral replicases are underlined.

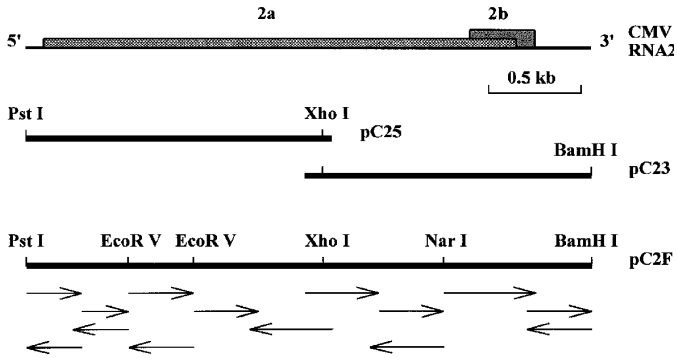


图3 CMV-SD RNA2 全长 cDNA 克隆的构建和序列测定策略

Fig.3 Construction of the full-length cDNA clone and sequencing strategy of CMV-SD RNA2

位于其后的 GDD<sup>[9]</sup>;在 3'端有一个小的与 ORF1 部分重叠的阅读框 ORF2(2414~2746nt),编码一个 111 aa 的 2b 蛋白,2b 基因的实际存在和表达由 Ding 等最初发现,并证明 2b 基因与病毒在植物体内的长距离移动和致病力密切相关<sup>[10-12]</sup>。

黄瓜花叶病毒株系根据其血清型、外壳蛋白肽图谱和核酸杂交分析可被划分为两个亚组,即 I 亚组和 II 亚组(或 Wt 亚组和 S 亚组)<sup>[13-15]</sup>。Fny 株系和 Q 株系分别是 I 亚组和 II 亚组研究得最多的代表成员。迄今,各株系病毒基因组 RNA 的测序结果显示,属于不同亚组的 CMV 株系的核苷酸序列和氨基酸序列同源性明显低于同属于一个亚组的 CMV 株系<sup>[16]</sup>。郭东川等曾对 CMV-SD 外壳蛋白基因进行了序列分析<sup>[17]</sup>,结果表明它与分属 I 亚组和 II 亚组的 D 株系和 Q 株系的外壳蛋白基因核苷酸序列同源性分别为 93.7% 和 76.0%。据此他们认为 SD 株系应属于 I 亚组。我们也将 SD 株系 RNA2 全序列分别与已发表的 Fny 株系、Q 株系 RNA2 全序列进行了比较,核苷酸序列同源性分别为 91.7% 和 75.6%。2a 蛋白的氨基酸序列同源性分别为 93.8% 和 67.7%,2b 蛋白的氨基酸序列同源性分别为 83.0% 和 51.3%。上述同源性比较结果进一步证明 SD 株系属于 I 亚组。

致谢 本研究得到了刘思泽同志技术上的大力协助,在此表示感谢。

### 参 考 文 献

- [1] Peden K W C, Symons R H. *Virology*, 1973 **53**: 487~492.
- [2] Lot H, Marchoux G, Marrow J M *et al.* *J Gen Viral*, 1974 **22**: 81~83.
- [3] Hayes R J, Buck K W. *Cell*, 1990 **63**: 363~368.
- [4] Schwinghamer M W, Symons R H. *Virology*, 1977 **79**: 88~108.
- [5] Francki R J B, Mossop D W, Hatta T *et al.* CMII/AAB Descriptions of plant viruses, 1979, No. 213.
- [6] Rizzo T M, Palukaitis P. *J Gen Viral*, 1988 **69**: 1777~1787.
- [7] Rezaian M A, Williams R H V, Gordon K H *et al.* *Eur J Biochem*, 1984 **143**: 277~284.
- [8] Kataoka J, Masuta C, Takanami Y *et al.* *Ann Phytopath Soc Japan*, 1990 **56**: 495~500.

- [ 9 ] Morch M D , Boyer J C , Haenni A L. *Nucleic Acids Res* ,1988 ,**16** :6157~6172.
- [ 10 ] Ding S W , Anderson B J , Haase H R *et al.* *Virology* ,1994 ,**198** :593~601.
- [ 11 ] Ding S W , Li W X , Symons R H *et al.* *EMBO J* ,1995 ,**14** :5762~5772.
- [ 12 ] Ding S W , Shi B J , Li W X *et al.* *Proc Natl Acad Sci USA* ,**93** :7470~7474.
- [ 13 ] Dvergne J C , Cardin L. *Ann Phytopathol* ,1973 ,**5** :409~430.
- [ 14 ] Edwards M C , Gonsalves D. *Phytopathology* ,1983 ,**73** :269~273.
- [ 15 ] Owen J , Palukaitis P. *Virology* ,1988 ,**166** :485~502.
- [ 16 ] Hayakawa T , Mizukami M , Nakamura I *et al.* *J Gen Virol* ,1989 ,**70** :499~504.
- [ 17 ] 郭东川 , 乔利亚 , 方荣祥等. *微生物学报* ,1993 ,**33**(3) :233~235.

## CONSTRUCTION OF FULL-LENGTH cDNA CLONE AND COMPLETE NUCLEOTIDE SEQUENCE OF RNA2 OF CUCUMBER MOSAIC VIRUS SD STRAIN\*

Zhang Guohua Xu Yan Cai Wenqi Fang Rongxiang

( *Laboratory of Plant Biotechnology , Institute of Microbiology , Academia Sinica , Beijing 100080* )

**Abstract** The full-length cDNA clone of CMV-SD RNA2 was constructed from two cDNA clones , which were obtained by RT-PCR with primers synthesized based on terminal sequences determined using 5' RACE and 3' RACE , respectively ; and the complete sequence was determined. It consists of 3048nt and there are two partial overlapped ORFs. ORF1 , which is located in the 5' portion ( 79 ~ 2652nt ) , codes for 2a protein of 858aa , and GDD consensus sequence reported to be " polymerase site " was found in deduced 2a protein sequence. ORF2 , which is located in the 3' portion ( 2414 ~ 2746nt ) , codes for 2b protein of 111aa. The overall nucleotide sequence homologies with RNA2 of CMV Fny and Q strain , which are representatives of subgroups I and II , respectively , are 91.7% and 75.6% ; the amino acid sequence homologies are 93.8% and 67.7% for 2a protein , and 83.0% and 51.3% for 2b protein , respectively. These results demonstrated that CMV-SD belongs to subgroup I.

**Key words** Cucumber mosaic virus SD strain , RNA2 , Full-length cDNA , Nucleotide sequence

\* Project of Chinese National Programs for Science and Technology Development ( No. 863-101-04-02-01 )

Project of Chinese National Natural Science Fund for Free Application ( No. 39400074 )

\* The nucleotide sequence data reported in this paper have been deposited with GenBank/EMBL/DDBJ databases under the accession number D86330