

# 水稻矮缩病毒最外层外壳蛋白基因( $S_2$ )cDNA 克隆、 序列分析及其在大肠杆菌中的表达\*

鲁瑞芳\*\* 李毅 杨崇林 颜华 陈章良

(北京大学生命科学院蛋白质工程及植物基因工程国家重点实验室 北京 100871)

**摘 要** 从水稻矮缩病毒(Rice dwarf virus, RDV)中国福建分离物中克隆分离了最外层外壳蛋白基因( $S_2$ )全长 cDNA, 并对其进行了序列分析, 结果表明 RDV  $S_2$  cDNA 全长 3512bp, 仅含一个 3348bp 的阅读框架, 编码一个含有 1116 个氨基酸的蛋白( $P_2$ )。与基因库中已知基因序列比较, 发现它与日本 RDV H 株系相应片段的核苷酸和氨基酸同源率分别为 94.6% 和 95.4%, 与轮状病毒 VP<sub>2</sub> 氨基酸序列有一定的同源性。 $S_2$  核苷酸序列二级结构预测结果表明, 5' 端 50 个核苷酸的二级结构为一个发夹结构和一个茎环结构。 $P_2$  有 4 个富含亮氨酸的区域, 位于 N 端亲水区域的 10 个氨基酸(AA 69~78)残基形成一个  $\alpha$ -螺旋, 这些特点均与轮状病毒 VP<sub>2</sub> 的结构特征相似。SDS-PAGE 和 Western 印迹分析表明在大肠杆菌中分段高效表达了  $S_2$  编码蛋白的 N 端和 C 端。

**关键词** 水稻矮缩病毒, 外壳蛋白基因, 序列分析, 基因表达, Western 印迹分析

**分类号** S432.4 **文献标识码** A **文章编号** 0001-6209(1999)04-0305-14

水稻矮缩病毒(Rice dwarf virus, RDV)主要分布于日本、朝鲜和我国南方各水稻产区, 是水稻病毒病的主要毒源之一。RDV 基因组由十二条双链 RNA(dsRNA)组成, 属于呼肠孤病毒科植物呼肠孤病毒组的成员, 在自然界该病毒通过叶蝉以持久方式传播。缺失 RDV 第二号片段( $S_2$ )编码的外壳蛋白( $P_2$ )时, RDV 不能被叶蝉传播并丧失侵染能力, 推测  $P_2$  在 RDV 与叶蝉的识别和入侵过程中起着重要作用<sup>[1-3]</sup>。我国已经完成了 RDV 中国福建分离物 11 个片段的克隆和序列分析工作<sup>[4-17]</sup>, 本文报道  $S_2$  cDNA 克隆和序列分析, 并构建了原核表达载体, 在大肠杆菌中分段表达了  $S_2$  编码的蛋白。

## 1 材料和方法

### 1.1 材料

RDV 感染的水稻叶片采自福建省福州市郊区。大肠杆菌菌株 DH5 $\alpha$ 、XL1-Blue、BL<sub>21</sub>(DE)<sub>3</sub>, 质粒 pBluescript(SK)、pTrcHisC 和 pET-11d 由本实验室提供。RDV  $S_2$  编码蛋白 C 端抗血清由日本 Suzuki N. 博士惠赠。

### 1.2 方法

**1.2.1** RDV  $S_2$  片段 RNA 的分离: 按高谦等人<sup>[12]</sup>的方法从感病水稻叶片中提取病毒。

\* 本研究为国家科委 863 水稻抗病毒基因工程资助课题(863-101)

本文报道的核苷酸序列已被 GenBank 收录, 接受号为  $S_2$  U73202

\*\* 现通讯地址: 中国科学院微生物研究所, 北京 100080

参加本工作的还有本实验室的安利忻同学和西北农业大学植保系的魏宁生教授

收稿日期: 1998-04-07, 修回日期: 1998-06-04

用蛋白酶 K 溶液消解病毒外壳蛋白,制备病毒核酸<sup>[4]</sup>。将 RDV dsRNAs 在 1.2% 的低熔点琼脂糖凝胶上电泳,分离回收 S<sub>2</sub> 片段。

**1.2.2 反转录和 PCR 扩增:**依据 Uyeda 等人<sup>[18]</sup>报道的 RDV H 株系 S<sub>2</sub> 核苷酸序列设计三对引物(表 1),以便分三段克隆。以经 DMSO 变性得到的单链 RNA 为模板,在引物 S<sub>22</sub>、S<sub>24</sub>、S<sub>26</sub> 和 AMV 反转录酶作用下分三段合成 cDNA 第一链。PCR 扩增条件为 94℃ 40s, 55℃ 或 58℃ 90s, 72℃ 120s, 35 个循环。

**1.2.3 目的片段的克隆和筛选:**PCR 产物经 T<sub>4</sub> DNA 聚合酶补平,与被 EcoRV 切开的 pBluescript(SK)连接,转化 *E. coli* DH5 $\alpha$ 。经含氨苄青霉素、X-gal 和 IPTG 的培养基筛选,煮沸法提取质粒,酶切鉴定重组子。依据重组子的物理图谱进行亚克隆。

表 1 引物序列及其在 RDV S<sub>2</sub> 中的位置

Table 1 Primer sequences and their locations in the RDV S<sub>2</sub> cDNA

人工合成的寡核苷酸引物 Synthetic primer	序列 Nucleotide sequence	在图 2 中的位置 Position in Figure 2
S <sub>21</sub>	5' GGCAAAACCTCGCCATGGCTTATCC3'	1~25nt
S <sub>22</sub>	5' AGCTGCAGAGTTGCGTTAATCCAGA3'	1516~1540nt 的互补序列
S <sub>23</sub>	5' AACGCAACTCTGCAGCTTAGTCTGT3'	1524~1548nt
S <sub>24</sub>	5' CCGTGTCTAGATAGTCAGTGCTTGA3'	2691~1715nt 的互补序列
S <sub>25</sub>	5' TCAAGCACTGACTATCTAGACACGG3'	2691~2715nt
S <sub>26</sub>	5' ATCATTTTAACTCAGAAGTACGCAT3'	3488~3512nt 的互补序列

**1.2.4 cDNA 的序列分析:**用 Pharmacia T<sub>7</sub> DNA 聚合酶测序试剂盒进行双脱氧终止法测序。将得到的 RDV S<sub>2</sub> 全长核苷酸序列和推导的氨基酸序列进行 DNASIS 分析,并在 BLAST 数据库中查对。

**1.2.5 RDV S<sub>2</sub> 全长序列的克隆:**将回收的 pBRS<sub>2-2</sub> 和 pBRS<sub>2-3</sub> 中的目的片段同时与经 Pst I 和 BamH I 双酶切的质粒载体 pBluescript(SK)连接,转化 *E. coli* DH5 $\alpha$ , 筛选、酶切鉴定及 PCR 扩增。将插入片段大小合适的重组子命名为 pBRS<sub>2HL</sub>。然后将回收的 pBRS<sub>2HL</sub> 中的目的片段与具有 Pst I 和 BamH I 双酶切口的线性 pBRS<sub>2-1</sub> 连接。转化后,筛选、鉴定含全长序列的重组子,并将其命名为 pBRS<sub>2FL</sub>(图 1)。

**1.2.6 原核表达载体的构建:**pBRS<sub>2-1</sub> 和 pBRS<sub>2FL</sub> 均用 Nco I 和 BamH I 双酶切,回收目的片段,与经 Nco I 和 BamH I 双酶切的表达载体 pET-11d 连接,转化 *E. coli* BL<sub>21</sub>(DE)<sub>3</sub>。Apa I 切开 pBRS<sub>2-2</sub>, 补平,再用 Pst I 酶切;用 Pst I 和 Hind III 双酶切 pBRS<sub>2-3</sub>, 回收目的片段,再分别与表达载体 pTrcHisC 连接,转化 *E. coli* XLI-Blue。挑取单菌落,用煮沸法提取质粒 DNA,酶切和 PCR 扩增鉴定重组子。

**1.2.7 表达产物的 SDS-PAGE 及 Western 印迹分析<sup>[11]</sup>:**挑取单菌落接入 5mL LB 液体培养基中,37℃ 振荡培养过夜。按 1:100 稀释加入 LB 液体培养基中,37℃ 培养约 3h,经 IPTG 4h 后,离心收集菌体。菌体悬浮于 200 $\mu$ L 1 $\times$  SDS 加样缓冲液,沸水浴 10min,离心并进行 SDS-PAGE。RDV S<sub>2</sub> 3' 端编码产物制备的抗血清用作第一抗体<sup>[19]</sup>,碱性磷酸酶标羊抗兔 IgG(1:10000)为第二抗体。

**1.2.8 包含体的制备:**采用 Xu 等的方法提取包含体<sup>[15]</sup>。IPTG 诱导 4 h 后的菌液 1.5mL,经溶菌酶、DNase I、RNaseA 酶解,最后沉淀用 200 $\mu$ L 1 $\times$  SDS 加样缓冲液悬浮,

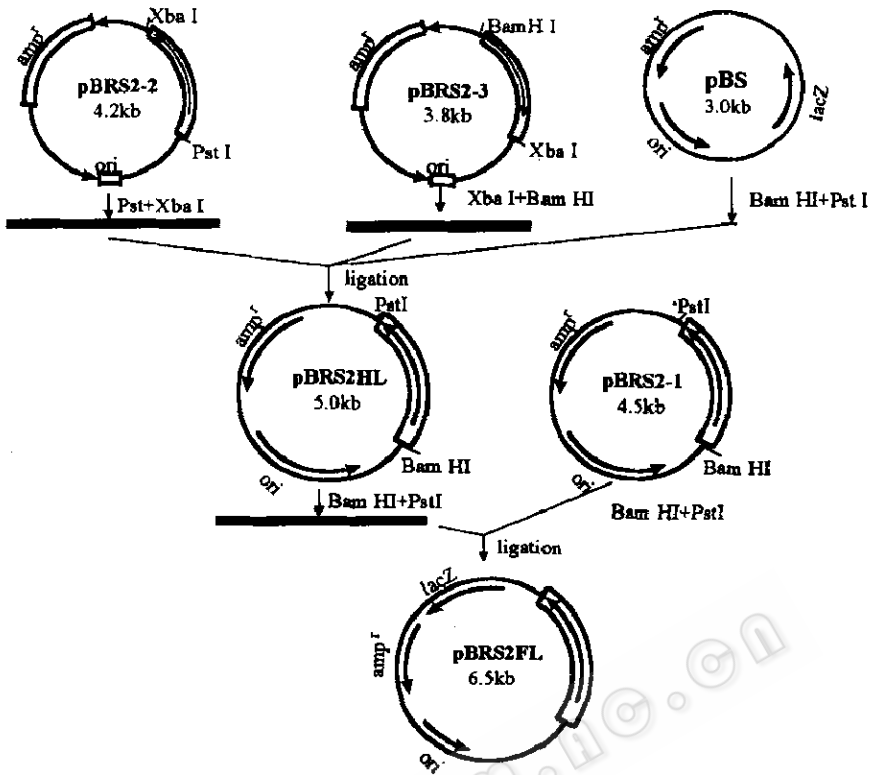


图 1 重组质粒 pBRS<sub>2FL</sub> 的构建

Fig. 1 Construction of pBRS<sub>2FL</sub> recombinant plasmid

沸水浴 10min, 取 15 $\mu$ L 进行 SDS-PAGE。

## 2 结果

### 2.1 RDV S<sub>2</sub> 基因的分离及其 cDNA 的 PCR 扩增

从纯化的病毒中提取了 RDV 的总 RNA, 并从低熔点琼脂糖凝胶中回收了 RDV S<sub>2</sub>, 其大小约 3.5kb, 浓度为 0.1 $\mu$ g/ $\mu$ L。以分段合成的 cDNA 第一链为模板进行 PCR 扩增, 得到了三条特异的扩增产物。它们的大小与预期的相当, 分别为 1.5kb(RDV S<sub>2-1</sub>)、1.2kb(RDV S<sub>2-2</sub>)、0.8kb(RDV S<sub>2-3</sub>)。

### 2.2 RDV S<sub>2</sub> cDNA 的分段克隆及其亚克隆策略

对上述三种 PCR 产物的转化产物进行筛选, 经酶切鉴定和 PCR 扩增, 得到三个重组子 pBRS<sub>2-1</sub>、pBRS<sub>2-2</sub> 和 pBRS<sub>2-3</sub>, 插入片段的大小分别为 1.5kb、1.2kb 和 0.8kb。其中 pBRS<sub>2-1</sub> 为反向插入, pBRS<sub>2-2</sub> 和 pBRS<sub>2-3</sub> 均为正向插入。用不同的限制性内切酶酶解, 得到它们的物理图谱, 并依此图谱制备了 9 个亚克隆。

### 2.3 RDV S<sub>2</sub> cDNA 的序列分析

pBRS<sub>2-1</sub>、pBRS<sub>2-2</sub>、pBRS<sub>2-3</sub> 及其 9 个亚克隆的测序结果表明 pBRS<sub>2-1</sub> 位于第 1 位碱基至第 1540 碱基, pBRS<sub>2-2</sub> 位于第 1524 碱基至 2715 碱基, pBRS<sub>2-3</sub> 位于第 2691 位碱基至第 3512 位碱基, 它们的大小分别为 1540bp、1192bp、822bp。RDV S<sub>2</sub> cDNA 的全长核苷

酸序列为 3512bp(图 2), 仅含有一个长为 3348bp 的开放阅读框架, 编码一个由 1116 个氨基酸组成、分子量为 123kD 的蛋白。经 DNASIS 软件对我们得到的 RDV S<sub>2</sub> cDNA 全长序列和推导的氨基酸序列进行分析, 发现它与日本 RDV H 株系的核苷酸和氨基酸序

	M	A	Y	P	N	H	V	R	N	V	W	
GG CAA AAC CTC GCC ATG GCT TAT CCC AAC CAC GTC AGA AAT GTT TGG												11
D V Y N L F R D V P N R E H L I												47
GAT GTG TAC AAC CTG TTC CGG GAC GTG CCA AAC CGT GAA CAT TTG ATT												27
R D I R N G L V T V R N L T N M												95
CGA GAT ATC CGC AAT GGG CTA GTT ACG GTC CGG AAT CTC ACA AAT ATG												43
L T N M E R D D Q L I I A Q L S												143
CTT ACT AAT ATG GAG CGC GAT GAT CAG TTG ATC ATT GCT CAA CTG TCC												59
N M M K S L S I G V E K A Q N E												191
AAT ATG ATG AAA TCG CTA TCT ATT GGG GTT GAG AAG GCT CAA AAT GAA												75
L S K L K T T D A D R A A V L A												239
CTT AGC AAA CTG AAG ACT ACG GAT GCC GAC CGT GCC GCA GTC TTA GCA												91
A Y Q T S V L N I E R N T M L L												287
GCT TAC CAA ACT TCG GTC CTT AAC ATA GAG CGA AAC ACT ATG TTA TTG												107
T G Y F K Q L V L D L T G Y V G												335
ACT GGG TAC TTC AAG CAG TTA GTC TTA GAC TTA ACT GGC TAT GTA GGT												123
A S V Y P I L P F M I T G D Q S												383
GCC AGT GTA TAT CCT ATC TTA CCT TTC ATG ATT ACG GGG GAT CAG TCT												139
M M V D S I K V N M K N V F D D												431
ATG ATG GTT GAC TCC ATT AAG GTT AAC ATG AAA AAT GTA TTT GAT GAT												155
K H E Q E I V L P I H P A C F V												479
AAG CAT GAG CAG GAG ATA GTC CTA CCT ATT CAT CCT GCT TGC TTT GTG												171
S T I T E D T S S V V Y A D G D												527
TCG ACC ATC ACT GAG GAC ACT TCT TCC GTA GTA TAC GCC GAT GGA GAT												187
E L Y S V H V R H A T M T M Y V												575
GAA CTG TAT TCA GTG CAC GTT AGG CAC GCA ACT ATG ACT ATG TAT GTC												203
N V L G E T V E T R Q L S M I G												623
AAT GTG CTT GGT GAG ACC GTC GAG ACC AGG CAG CTG TCT ATG ATA GGC												219
E S I V P D D F A P S L L I L R												671
GAA TCC ATC GTT CCT GAT GAT TTC GCG CCT TCC TTG CTG ATA TTG AGA												235
F S Q N S V G E V F Y L S H D N												719
TTT AGT CAA AAT TCA GTC GGA GAG GTT TTC TAT CTT AGT CAT GAC AAC												251
M K K F L G Y S L E Y T D K Y V												767
ATG AAA AAG TTT CTA GGC TAC AGT CTT GAA TAT ACT GAC AAG TAC GTT												267
I F D V A R R V S T T R N T I P												815
ATA TTC GAC GTA GCC AGA CGC GTG TCC ACG ACG CGT AAT ACG ATA CCT												283
N G F C S V D G V P Y L D G R F												863
AAT GGA TTC TGC TCT GTT GAT GGC GTG CCG TAT CTT GAT GGA CGG TTC												299
I Y Q P S G I S A D S N I C A I												911
ATT TAT CAA CCG AGC GGA ATT AGT GCG GAT AGT AAT ATC TGT GCT ATA												315
Y N S Y V L D T L R Y M T E C E												959
TAC AAT TCA TAT GTT TTA GAC ACT CTT AGA TAT ATG ACT GAG TGT GAA												331
V D T L R S V Y D Q T S S T A F												1007
GTT GAT ACG TTA AGA TCG GTG TAC GAT CAA ACT TCG TCA ACT GCA TTC												347
S K T D V L T S S L L T M Q S N												1055
TCA AAG ACC GAC GTA TTG ACA TCG AGT CTG CTA ACC ATG CAA AGT AAC												363
I S A L S A A T P Q L A N D V I												1103
ATT TCA GCT TTA TCT GCC GCT ACC CCC CAG TTA GCG AAT GAC GTA ATC												379
T F G S T D L L S L G T V L T V												1151
												395

ACA TTC GGT TCT ACC GAT CTA CTT TCT CTC GGG ACA GTA TTG ACC GTA 1199  
 S N E F T A D D T I L S T S L A 411  
 TCG AAT GAG TTT ACA GCT GAT GAT ACA ATA TTA AGC ACT AGT CTT GCA 1247  
 G H C Q V D Y S E G S P Q D K S 427  
 GGT CAC TGC CAA GTT GAC TAC AGC GAG GGA TCA CCT CAA GAC AAG AGC 1295  
 M S I P V S C D S S Q F A S S T 443  
 ATG AGT ATA CCT GTA AGC TGT GAT TCG TCG CAG TTT GCG TCA TCT ACC 1343  
 V H S Y S A D I L G H G L K G D 459  
 GTC CAT TCT TAC TCA GCT GAC ATA CTG GGG CAT GGA CTC AAG GGT GAC 1391  
 R N M N L M I N V L G L M N P Q 475  
 CGA AAT ATG AAC TTA ATG ATA AAT GTA CTT GGA CTT ATG AAC CCC CAG 1439  
 K V T V D Y V Y S D G Y K L N F 491  
 AAA GTA ACG GTT GAC TAT GTC TAT TCG GAT GGT TAC AAA CTG AAC TTT 1487  
 A S V V A P G A P F W I N A T L 507  
 GCT TCA GTG GTT GCC CCC GGC GCG CCT TTC TGG ATT AAC GCA ACT CTG 1535  
 Q L S L S P S A H N M L S K L T 523  
 CAG CTT AGT CTG TCA CCT TCG GCA CAC AAT ATG CTG AGT AAG TTA ACA 1583  
 P L D N D A C P G L K A Q A N T 539  
 CCA CTG GAT AAT GAT GCA TGC CCT GGG CTT AAA GCG GCA AAC ACG 1631  
 P V L V S M T I N L D D A T P A 555  
 CCC GTA CTT GTA TCT ATG ACC ATC AAT CTT GAT GAC GCC ACT CCT GCT 1679  
 L G G E V I Q N C V F K I H H G 571  
 TTG GGA GGG GAG GTG ATT CAA AAC TGC GTG TTC AAG ATA CAC CAC GGC 1727  
 D D V Y S F V T D F D V V S Y T 587  
 GAT GAT GTG TAT AGC TTC GTC ACC GAC TTC GAT GTC GTT AGT TAT ACC 1775  
 S T S G T N C L K L I S S V D I 603  
 TCA ACG TCC GGG ACC AAT TGT TTG AAG CTT ATA TCA AGC GTG GAT ATC 1823  
 T S Q L P S D M V I Y V M N G S 619  
 ACC AGT CAG CTT CCC TCT GAC ATG GTG ATC TAC GTT ATG AAT GGG TCG 1871  
 P D A A F I S G D S V N M S S V 635  
 CCC GAT GCA GCT TTT ATA TCT GGT GAT TCC GTT AAT ATG TCA TCA GTG 1919  
 D W H Q S T S Q T V G N Y V Y T 651  
 GAC TGG CAC CAG TCC ACA AGC CAG ACT GTG GGG AAT TAT GTC TAT ACC 1967  
 T M K A Y W N V T S Y D V E A R 667  
 ACG ATG AAG GCT TAT TGG AAT GTA ACG TCG TAT GAT GTT GAG GCT CGT 2015  
 P Y T T Y V P G K V N F T A V E 683  
 CCT TAC ACC ACA TAC GTA CCT GGA AAG GTT AAC TTT ACA GCC GTA GAG 2063  
 H A D V F V D D Y N T G V N S Y 699  
 CAC GCT GAT GTT TTT GTA GAT GAC TAC AAC ACC GGC GTT AAC TCA TAT 2111  
 V I Y N S R I Y Y K G T P L Y I 715  
 GTG ATT GTA AAT AGT AGA ATA TAT TAT AAG GGA ACC CCT CTA TAT ATT 2159  
 E V P S G S F I K V S Y V T S P 731  
 GAG GTA CCA AGT GGC TCA TTC ATC AAG GTG AGC TAC GTC ACA AGC CCT 2207  
 L K N P T V D T Y N A E I S R N 747  
 TTG AAG AAC CCT ACC GTG GAC ACT TAT AAC GCT GAA ATC TCT CGC AAT 2255  
 S A Y L I K A N A S L D S V A A 763  
 TCA GCG TAC CTA ATT AAA GCA AAT GCT TCA TTA GAT TCA GTG GCC GCA 2303  
 M L N N I S N R I D A M E R L M 779  
 ATG CTC AAT AAT ATA TCG AAT CGA ATC GAT GCG ATG GAA CGC TTG ATG 2351  
 E P T R A Q Q I A G V V S S I G 795  
 GAG CCC ACA CGT GCG CAG CAG ATT GCA GGA GTA GTC TCA AGC ATA GGT 2399  
 G V I S L G M P L L G A I V V T 811  
 GGA GTC ATC TCA CTC GGC ATG CCA TTG CTC GGA GCG ATC GTG GTA ACC 2447  
 I G T I I S I A D P D K Q G I D 827

ATC GGT ACT ATT ATC TCC ATT GCT GAC CCA GAC AAA CAG GGC ATT GAT	2495
Y H S V A N A F M S W C Q Y A A	843
TAC CAC TCG GTG GCT AAT GCC TTT ATG TCC TGG TGC CAG TAT GCG GCG	2543
V C R Y E Y G L L K R G D E K L	859
GTC TGT AGG TAT GAA TAC GGA CTT CTG AAG CGT GGG GAT GAG AAA CTA	2591
D V L S F M P K R V V S D F K N	875
GAT GTG TTG TCA TTT ATG CCG AAA CGT GTC GTG TCG GAT TTC AAG AAT	2639
K P D V I S L P E L G E S V L R	891
AAA CCC GAC GTT ATT AGT CTC CCA GAA TTA GGT GAA TCA GTA CTA CGT	2687
G S S T D Y L D T G I N I I Y N	907
GGA TCA AGC ACT GAC TAT CTA GAC ACG GGG ATT AAT ATC ATT TAT AAC	2735
D M Q L L G Q G K L S D W L N K	923
GAT ATG CAA CTG CTT GGA CAA GGC AAA CTC TCT GAC TGG CTC AAC AAA	2783
T V S R V E N N A A N F F E R N	939
ACT GTG AGC AGG GTG GAG AAT AAT GCT GCC AAC TTC TTT GAG AGG AAC	2831
L V K S L A N K E V L P M H A R	955
TTG GTT AAA AGT CTT GCG AAT AAG GAA GTC CTA CCA ATG CAT GCT CGA	2879
V E I T Q T E K I G D V Y R T T	971
GTT GAG ATT ACT CAA ACC GAG AAA ATT GGT GAT GTG TAT AGG ACT ACG	2927
I L Y T G I N E G S Y L G G D V	987
ATC CTA TAT ACA GGG ATA AAT GAA GGA TCA TAT TTG GGT GGG GAT GTC	2975
F A S R L G D K N I L R M N G F	1003
TTT GCT TCG CGG TTG GGG GAC AAA AAC ATC TTG CGT ATG AAT GGA TTT	3023
E S G P G R F K A I V E S T T E	1019
GAA AGC GGA CCA GGA AGG TTC AAA GCT ATC GTC GAA TCA ACT ACT GAA	3071
V G N F R V V D W T V S G M S R	1035
GTA GGC AAC TTT CGT GTA GTT GAT TGG ACG GTG TCT GGA ATG TCC AGG	3119
Y E I Y A A A G E I Y P S K D P	1051
TAC GAG ATT TAT GCT GCT GCC GGT GAA ATA TAT CCG AGC AAA GAC CCC	3167
S H A D V Q L L Y E S I V R D L	1067
TCT CAT GCT GAC GTA CAG CTG TTA TAC GAA AGC ATA GTT CGG GAT TTA	3215
T T R D G S F I L K H H D V L L	3263
L P G Q L D A F E E L I I R N A	1099
TTG CCT GGG CAA CTT GAC GCT TTT GAG GAG CTA ATC ATA AGA AAT GCC	3311
S N Y Q Y A F I G S N C S K L C	1115
TCG AAT TAC CAA TAT GCG TTT ATT GGT TCA AAC TGT TCA AAA TTA TGC	3359
A *	
GCA TGA TGT GGT TGA CAT CTT GAC CAA GTT CAA GCG ACC ACA AAG GTG	3407
GAT TAA AGA TGA TGA CTT CAA GTT GTA CAT CCA ATC TAT CTA TGA TCC	3455
ATT GTG ATA TGT ACC AGC CAG GGC TCA GAT GTA TGC GTA CTT CTG AGT	3503
TAA AAT GAT	3512

图 2 RDV S<sub>2</sub> cDNA 全长序列以及推导的氨基酸序列Fig. 2 The nucleotide sequence and its deduced amino acid sequence of RDV S<sub>2</sub> cDNA

列<sup>[1]</sup>的同源性分别为 94.6% 和 95.4%。在 192 处碱基突变中, 以同类碱基之间的转换为主, 占点突变的 85.8%。P<sub>2</sub> 氨基酸序列在 Blaster 数据库中比较, 发现它与轮状病毒的 VP<sub>2</sub> 蛋白有一定的同源性(图 3)。

```

VP2  ELSKLTDDADRAAVLADYQTSVLNIERNMMLLTGYFKQLVLD.....TNCLKLISSVDITSQLPSDMWI
      +L +L +T+A   + D Q  L I+ T  LTG  Q  D          TN + LIS + + + +P+DM+I
P2   DLKELVSTEAIQKMSQDLQLEALTIQSETQFLTGINSQAAND.....TNYMSLISGMWLLTVVPNDMFI

VP2  .....NFTAIDHADVFDDYNTGVNSYIVNSRI.....VRNLTNMLTNMERD
      N T I          YN VN +  N RI          +RN N + NM+R+
P2   .....NATVIPSPQTLFHYYNVNVNFHSNYNERI.....IRNDVNYILNMDRN
    
```

图 3 RDV P<sub>2</sub> 氨基酸序列与轮状病毒 VP2 的比较

Fig.3 Comparison amino acid sequence of RDV P<sub>2</sub> with that of rotivirus VP<sub>2</sub>

RDV S<sub>2</sub> 核苷酸序列二级结构预测结果表明,5'端 50 个核苷酸的二级结构为一个发夹结构和一个茎环结构(图 4),3'端的 50 个核苷酸也形成一个茎环结构。疏水性分析发现 P<sub>2</sub> N 端呈亲水,而 C 端为疏水(图略)。位于 N 端亲水区域的 10 个氨基酸残基(AA<sub>69-78</sub>)形成一个 α-螺旋。P<sub>2</sub> 蛋白还含有 4 个富含亮氨酸的区域(AA<sub>90-130</sub>, AA<sub>533-556</sub>, AA<sub>881-904</sub> 和 AA<sub>1059-1083</sub>),其中 AA<sub>533-566</sub> 推测可能形成一个亮氨酸拉链。

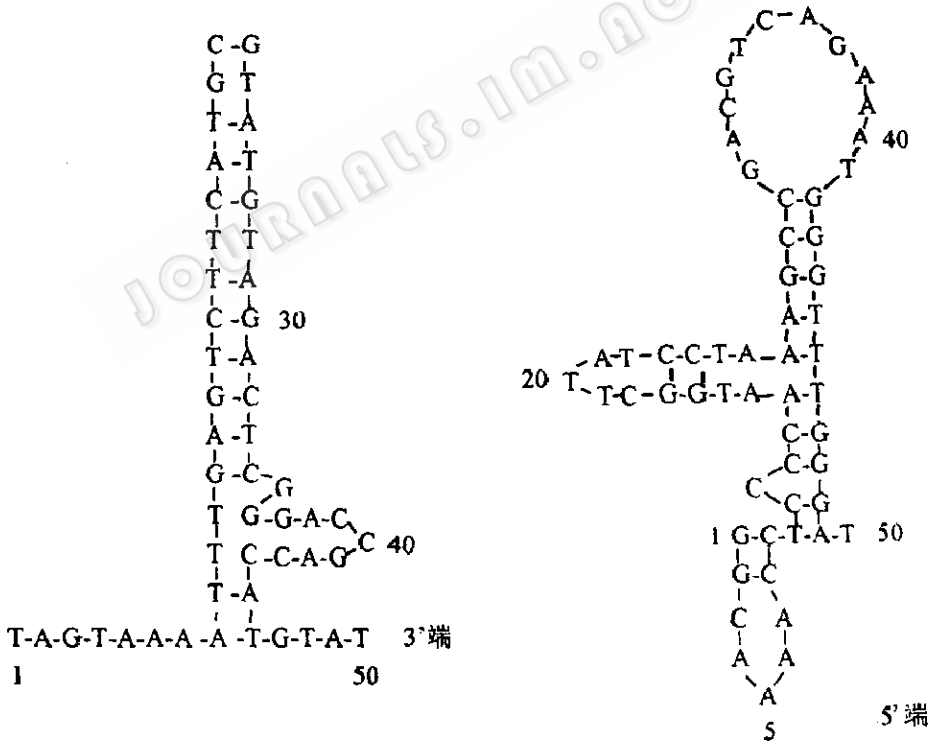


图 4 RDV S<sub>2</sub> cDNA 5'端和 3'端二级结构分析

Fig.4 The second structure of rice dwarf virus S<sub>2</sub> cDNA sequence in 5' and 3' termini

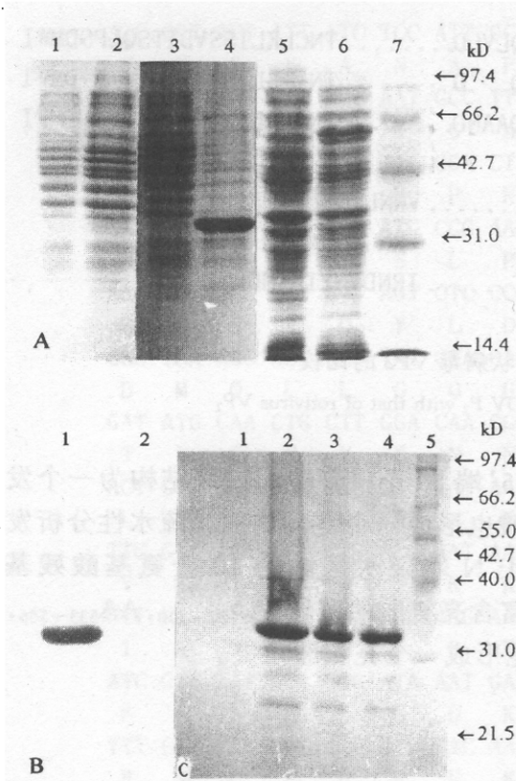


图5 DRV S<sub>2</sub> 表达产物的 SDS-PAGE 和 Western 印迹分析

A. SDS-PAGE: 1. pTCS<sub>2-2</sub>/XL-1Blue; 2. pTAS<sub>2fl</sub>/XL-1Blue; 3. pTrcHisC/XL-ue; 4. pTCS<sub>2-3</sub>/BL-1Blue; 5. pET-11d/BL<sub>21</sub>(DE)<sub>3</sub>; 6. pETS<sub>2-1</sub>/BL<sub>21</sub>(DE)<sub>3</sub>; 7. 标准蛋白分子量。

B. Western blot: 1. pTCS<sub>2-3</sub>/XL-1Blue; 2. pET-11d/BL<sub>21</sub>(DE)<sub>3</sub>。

C. pTCS<sub>2-3</sub> 包含体的 SDS-PAGE; 1. pET-11d/BL<sub>21</sub>(DE)<sub>3</sub>; 2~4. pTCS<sub>2-3</sub>/XL-1Blue 分别诱导 4、2、1 h; 5. 标准蛋白分子量。

Fig. 5 SDS-PAGE and Western blot analysis of the expression products of rice dwarf virus S<sub>2</sub> cDNA in *E. coli*.

A. SDS-PAGE: 1. pTCS<sub>2-2</sub>/XL-1Blue; 2. pTAS<sub>2fl</sub>/XL-1Blue; 3. pTrcHisC/XL-ue; 4. pTCS<sub>2-3</sub>/XL-1Blue; 5. pET-11d/BL<sub>21</sub>(DE)<sub>3</sub>; 6. pETS<sub>2-1</sub>/BL<sub>21</sub>(DE)<sub>3</sub>; 7. Protein molecular weight marker.

B. Western blot: 1. pTCS<sub>2-3</sub>/XL-1Blue; 2. pET-11d/BL<sub>21</sub>(DE)<sub>3</sub>。

C. SDS-PAGE of pTCS<sub>2-3</sub> inclusion; 1. pET-11d/BL<sub>21</sub>(DE)<sub>3</sub>; 2~4. pTCS<sub>2-3</sub>/XL-1Blue induced for 4, 2, 1 h by IPTG, respectively; 5. Protein molecular weight marker.

## 2.4 RDV S<sub>2</sub> 基因在大肠杆菌中的表达

按 1.2.5 中所述的方法, 经过筛选和酶切鉴定以及 PCR 扩增得到了插入片段为 3.5kb 的重组子 pBRS<sub>2FL</sub>。按 1.2.6 所述方法, 将 pBRS<sub>2-1</sub> 和 pBRS<sub>2FL</sub> 中的外源片段克隆到表达载体 pET-11d 的 *Nco* I 和 *Bam* H I 位点, 得到表达质粒 pETS<sub>2-1</sub> 和 pETS<sub>2FL</sub>。pBRS<sub>2-2</sub> 和 pBRS<sub>2-3</sub> 中的目的片段克隆到 pTrcHisC 中的 *Pst* I 位点, 获得表达质粒 pTCS<sub>2-2</sub> 和 pTCS<sub>2-3</sub>。

SDS-PAGE 结果发现 pTCS<sub>2-2</sub> 和 pTAS<sub>2FL</sub> 没有出现所要表达蛋白的特异条带; pETS<sub>2-1</sub>、pTCS<sub>2-3</sub> 分别在大约 55kD、32kD 处出现一条明显的特异条带(图 5A), 其大小与预期的相当。对 SDS-PAGE 凝胶进行薄层层析扫描, pTCS<sub>2-3</sub> 在相应位置上比 pTrcHisC 多出一个峰, 诱导 4h 的表达产物占菌体总蛋白含量的 46.8%。用 S<sub>2</sub> C 端 cDNA 编码蛋白的抗体 AcRS<sub>2</sub>T<sub>1</sub> 对上述表达产物进行 Western 印迹检测, 结果仅 pTCS<sub>2-3</sub> 的特异表达产物与该抗体发生特异反应, 在相应位置上出现了特异的蛋白条带(图 5B)。

采用 Xu 等的方法<sup>[15]</sup> 提取 pTCS<sub>2-3</sub> 和 pTrcHisC 中的包含体。SDS-PAGE 结果发现 pTCS<sub>2-3</sub> 在 32kD 处有一条很明显的蛋白带, 而 pTrcHisC 则无此条带(图 5C), 说明 pTCS<sub>2-3</sub> 在大肠杆菌中的表达产物以包含体的形式存在。

## 3 讨论

将我们获得的 RDV S<sub>2</sub> 序列与日本 RDV H 株系的 S<sub>2</sub> 序列<sup>[18]</sup> 比较, 发现在 190 处点突变中, 同类碱基之间的转换为 163 处, 并且转换没有方向性。本实验室获得的 RDV S<sub>1</sub>、S<sub>3-12</sub> 各片段也有同样的变异, 这与 dsRNA 复制有关<sup>[13]</sup>。此外, 这种变异特点还与病毒不同株系的地域环境如寄主种类、昆虫介体种类有关。Uyeda 等在 PAGE 电泳中发现不同的 RDV 分离物基因组各片段之间的泳动率不同<sup>[20]</sup>。日本分离物与朝鲜分离物之间差异不如日本分离物与尼泊尔、日本分离物与菲律宾分离物间的大, 这除



了地域环境不同外,主要是由于它们的传播介体的种类不同<sup>[21]</sup>。在日本 RDV 主要是由二点黑尾叶蝉(*Nephotettix cincticeps*)传播,而菲律宾则由 *N. nigropictus* 传播。

利用 DNASIS 软件对 RDV 中国福建分离物 S<sub>2</sub> 核苷酸序列进行二级结构预测,发现 5' 端 50 个苷酸形成的二级结构与轮状病毒 5' 端 60 个核苷酸形成的二级结构相似<sup>[22]</sup>。在轮状病毒中,5' 端非编码区核苷酸形成的发夹结构和茎环结构对轮状病毒的复制、翻译均起作用<sup>[22]</sup>。RDV S<sub>2</sub> 5' 端的非编码区的九个核苷酸形成的发夹结构是否对 RDV S<sub>2</sub> 复制、翻译起调控作用有待研究。RDV P2 氨基酸序列分析结果表明 P2 蛋白与轮状病毒 VP2 有一定的同源性,其疏水性及 N 端、C 端的二级结构又与 VP2 的相同。在轮状病毒中,VP<sub>2</sub> 是一种核心蛋白,具有结合 RNA 的活性,参与病毒复制中间体的形成,并能够形成二聚体。P<sub>2</sub> 是一种外层衣壳蛋白,参与介体传播,决定病毒的侵染性<sup>[1~3]</sup>,而且具有 4 个富含亮氨酸区域,那么 P<sub>2</sub> 是否具有结合 RNA 的活性,能否在寄主体内形成二聚体有待进一步探索。

Suzuki 等用杆状病毒载体在家蚕(*Spodoptera frugiperda*)细胞中表达了 RDV 的 S<sub>1</sub> 和 S<sub>11</sub> cDNA 编码的 11 个蛋白,并认为大肠杆菌表达系统不适于 RDV 蛋白的大量表达<sup>[19]</sup>。我们在大肠杆菌中成功地表达了 RDV S<sub>2</sub> 编码蛋白的 N 端和 C 端,而且 pTCS<sub>2-3</sub> 属于高效表达(占菌体蛋白的 46.8%),但其中间区域却没有表达。我们曾在原核表达载体 pTrcHisA 未能表达 RDV S<sub>2-1</sub>,但在 pET-11d 中得到了表达,说明欲表达 RDV S<sub>2-2</sub> 需要选用其它的原核表达载体。

## 参 考 文 献

- [1] Omura T, Takahashi A, Iino H *et al.* *Annals of the Phytopathological Society of Japan*, 1994, **60**:384.
- [2] Yan J, Tomaru M, Takahashi A *et al.* *Virology*, 1996, **224**:539~541.
- [3] Tomaru M, Maruyama W, Kikuchi A *et al.* *Journal Virology*, 1997, **71**:8019~8023.
- [4] 李毅,薛志安,刘一飞等. *病毒学报*, 1994, **10**:339~345.
- [5] 李玮,李毅,陈章良等. *应用基础与工程科学学报*, 1994, **2**:109~115.
- [6] 刘一飞,李毅,潘乃隧等. *病毒学报*, 1994, **10**:246~250.
- [7] 曲林,李毅,朱玉贤等. *病毒学报*, 1995, **11**:270~274.
- [8] 李玮,李毅,张旭等. *病毒学报*, 1995, **11**:56~62.
- [9] 肖锦,李毅,陈章良等. *生物工程学报*, 1996, **12**:361~366.
- [10] 赵晓岚,李毅,陈章良等. *微生物学报*, 1996, **36**:1~11.
- [11] 曲林,李毅,陈章良. *微生物学报*, 1996, **36**:335~343.
- [12] 高谦,狄阳新,刘玮等. *植物学报*, 1990, **32**:13~18.
- [13] 肖锦,李毅,陈章良. *微生物学报*, 1998, **38**(5):348~358.
- [14] Mao Z J, Li Y, Chen Z L. *Archives of Virology*, 1998, (in press)
- [15] Xu H, Li Y, Chen Z L. *Virology*, 1998, **240**:267~272.
- [16] Zhang F J, Li Y, Chen Z L. *Acta Virologica*, 1997, **41**:161~168.
- [17] Zheng H H, Li Y, Chen Z L *et al.* *Theor Appl Genet*, 1997, **94**:522~527.
- [18] Uyeda I, Suda N, Yamada N *et al.* *Intervirology*, 1994, **37**:6~11.
- [19] Suzuki N, Sugawara M, Kusano T *et al.* *Virology*, 1994, **202**:41~48
- [20] Fukumoto F. *Annals of the Phytopathological Society of Japan*, 1992, **58**:83~86.
- [21] Uyeda I. *Seminars in Virology*, 1995, **6**:117~121.

[22] Patton J T. *Journal General Virology*, 1995, **76**:2633~2644.

## MOLECULAR CLONING AND SEQUENCING OF OUTER CAPSID PROTEIN GENE OF RICE DWARF VIRUS AND ITS EXPRESSION IN *ESCHERICHIA COLI*

Lu Ruifang Li Yi Yang Chonglin Yan Hua Chen Zhangliang

(*National Laboratory of Protein Engineering and Plant Genetic Engineering,*

*College of Life Sciences, Peking University, Beijing 100871*)

**Abstract** The  $S_2$  full-length cDNA of rice dwarf phyto-reovirus which encodes the viral outer capsid protein was cloned and its complete nucleotide sequence was determined. The results showed that  $S_2$  is 3512bp long with a large open reading frame which encodes a protein of 1116 amino acids. It shares 94.6% and 95.4% identity with RDV of Japanese H isolate in terms of nucleotide and amino acid sequences, respectively, and it also shows some homology with VP2 of rotavirus at the level of amino acid sequence. The search of deduced RDV  $S_2$  amino acid sequence in Blast network found that there were 4 leucine-rich motifs in  $P_2$  protein, and ten amino acids within the hydrophobic region at amino-terminus could form an  $\alpha$ -helix. Predicted secondary structure of  $S_2$  cDNA indicated that a hairpin and a stem loop are present in the 5'-end within 50bp, and a stem loop in the 3'-end within 50bp. RDV  $S_2$  partial and full-length sequences were then cloned into expression vector pET-11d & pTrcHisC. SDS-PAGE and Western blot proved that amino- and carboxyl-termini of  $P_2$  were successfully expressed in *E. coli*.

**Key words** Rice dwarf phyto-reovirus, Outer capsid protein gene, Sequence analysis, Gene expression, Western blot

\* This work was supported by grants (863-101) from the Chinese Sciences and Technology Committee