

水稻矮缩病毒第一号组份基因和编码蛋白的序列分析*

肖 锦¹ 李 毅^{1,2**} 张 净¹ 刘劲锋¹ 陈章良^{1,2}

(¹ 北京大学生命科学院蛋白质工程及植物基因工程重点实验室 北京 100871)

(² 德蒙特伏特大学 莱切斯特 英国)

摘 要 水稻矮缩病毒(Rice Dwarf Virus, 简称 RDV)是我国南方水稻病毒病的重要病原,属植物呼肠孤病毒。从中国福建分离物中克隆了基因组第一号片段(S1)的全长cDNA并对其进行全序列分析,结果表明 RDV 福建分离物 S1 克隆片段全长 4422bp,含有一个长 4332bp 的开放阅读框架,编码一个由 1444 个氨基酸组成的多肽(P1),分子量为 164kD。根据基因序列,对推测的 P1 氨基酸序列分析表明,序列中含有依赖于 RNA 的 RNA 聚合酶(RNA-dependent polymerase-RDRP)保守序列: motif I (DXXXXD)、motif II (SGXXXTXXXN)和 motif III (GDD),除此之外,在模式 III 后还存在一个很保守的区域 EXXKXY。由此说明 RDVS1 编码的蛋白 P1 可能是病毒的一种 RDRP。将 RDV 福建分离物 S1 核苷酸和编码蛋白氨基酸序列与日本流行株系相比,同源性分别为 95% 和 97%。RDV 福建分离物 S1 序列已被 GenBank 接受,号码为 U73201。

关键词 水稻矮缩病毒,第一号片段,序列分析,依赖于 RNA 的 RNA 聚合酶

分类号 S432.4

水稻矮缩病毒(Rice Dwarf Virus, RDV)属丁呼肠孤病毒科植物呼肠孤病毒属(phytoreoviruses)第一亚属,是水稻普通矮缩病的病原。在自然条件下, RDV 由二点黑尾叶蝉(*Nephotettix cincticeps*)和电光叶蝉(*Recilia dorsalis*)以持久方式传播,主要侵染少数禾本科植物,这种病毒病于 1883 年在日本首先发现,目前在日本、朝鲜、东南亚和我国广泛流行,造成水稻严重减产。

RDV 病毒粒子为无突起、无脂蛋白外膜的正二十面体,直径为 70nm,具有双层外壳蛋白^[1],内含由 12 条 dsRNA 组成的基因组和一些酶类,如转录酶(RDRP)负责从基因组 dsRNA 合成正负链 RNA。RDV 的转录酶活性象其他呼肠孤病毒一样在去除外层外壳蛋白后仍保留其活性。因此认为,转录酶活性与一种或几种核心蛋白有关。至今已有一些病毒编码的 RDRP 转录酶已经从受病毒感染的真核细胞中分离出来,通过计算机分析揭示出正链和双链 RNA 病毒的 RDRP 有一些保守的序列元件与功能有关,这些序列元件有助于识别 RDRP^[2]。

依照在聚丙烯酰胺凝胶电泳中的泳动速度的快慢, RDV 基因组 12 条 dsRNA 被依次

* 国家科委“863”水稻抗病毒工程资助课题(863-101)。

** 通讯作者

收稿日期: 1997-05-10

命名为 S2~S12, 除 RNA12(S12) 以外, 其余每一个片段只包含一个开放阅读框架, 其中 S2 和 S8 编码外层外壳蛋白, S3 编码内层外壳蛋白, S1、S5 和 S7 编码核心蛋白, S4、S6、S9、S10、S11 和 S12 编码非结构蛋白^[3~5]。

由于 RDV 造成水稻减产, 给农业造成很大危害, 因此引起东南亚及我国等许多国家科学家的重视。近年来, 随着分子生物学等技术的逐渐成熟, 对 RDV 整个基因组功能结构的基础研究取得了很大进展。我们实验室自 1988 年以来先后克隆了 S1~S12 所有基因 (S2 文章正在整理之中)^[6~15], 并完成了其全序列测定, 而且对各组份蛋白功能的研究及转化水稻的工作正在进行^[16, 17], 对于今后应用基因工程方法防治该病害有很重要的作用, 本文是 S1 基因组份工作的报道。

1 材料和方法

1.1 材料

RDV 感染的水稻植株来自福建省福州市郊感病稻田。cDNA 合成试剂盒、PCR 合成试剂盒、DNA 序列分析试剂盒、限制性内切酶及其他工具酶分别购自 Promega 公司、华美生物工程公司、Pharmacia 和 Bio-Lab 公司。

PCR 引物由美国 PE 公司合成。

(1) 5'-GGCAAATGATCGGCTGTGCCTGGAC-3'

(2) 5'-TTGATTGGATCTAGAATAGCCACCA-3'

XbaI

(3) 5'-TGGTGGCTATTCTAGATCCAATCAA-3'

XbaI

(4) 5'-CGCGTAGGGTACCAATCGCTCTTTT-3'

KpnI

(5) 5'-AAAAGAGCGATTGGTACCCTACGCG-3'

KpnI

(6) 5'-ATCATATATGATCGGCTGATTAACCTAT-3'

1.2 方法

1.2.1 病毒的分离纯化: 将 RDV 侵染的病株茎叶剪碎, 加液氮研磨。每克组织加入 3ml 0.1mol/L 磷酸缓冲液, 充分匀浆, 低速离心后取上清, 用氯仿抽提两次, 上清加入 7%PEG6000、3mol/LNaCl 沉淀, 最后再用两次差速离心和一次蔗糖密度梯度离心纯化病毒。

1.2.2 病毒 RNA 的抽提、回收、变性: 取病毒纯化液, 用蛋白酶 K 消化病毒外壳蛋白, 56℃ 保温 1h。然后分别用饱和酚、酚:氯仿 (1:1)、氯仿:异戊醇 (24:1) 各抽提一次, 乙醚去除残留氯仿后, 乙醇沉淀, 溶于 TE 缓冲液中, 将 RDVsRNAs 在 1.2% 低熔点琼脂糖凝胶上电泳分离, 将 S1 条带回收纯化, 然后经 DMSO 变性, 乙醇沉淀, 溶于无 RNAase 的重蒸水中。

1.2.3 cDNA 合成及 PCR 扩增: 取 2 μg 经变性处理的单链 RNA 样品与合成的两个特异引物各 2 μl (20 μmol/L) 在 70℃ 下退火 5min, 加入 dNTP、AMV 反转录酶合成 cDNA 第一条链。cDNA 第二条链的合成采用 PCR 技术扩增。扩增条件为 94℃ 变性 35s, 54℃ 退火 90s, 72℃ 延伸 120s, 反应进行 35 个循环。

GG	CAA	ATG	ATC	GGC	TGT	GCC	TGG	ACT	TGA	TTC	GTC	ATG	GAT	CTT	GCG	47
	M	I	G	C	A	W	T	*				M	D	L	A	14
AGG	TCG	GAT	ATC	ATT	CCG	CAT	TTG	CTA	TGC	CTC	TTC	CAA	GAG	ATA	ATC	95
R	S	D	I	I	P	H	L	L	C	L	F	Q	E	I	I	30
CAA	GCC	AAC	ATC	CAA	AAA	GTA	TCT	GAG	GCG	TAC	GAT	TTG	GAG	TTG	AAA	143
Q	A	N	I	Q	K	V	S	E	A	Y	D	L	E	L	K	46
ATA	ATG	AAC	ATT	TTG	ACG	TTG	TGG	CGA	AAT	GGA	TCT	GAC	AAT	CTG	ATT	191
I	M	N	I	L	T	L	W	R	N	G	S	D	N	L	I	62
TCA	GAC	AAC	GAT	GAT	TAC	ATG	AAA	AAG	GGA	CTA	TTC	TTC	AGT	AGG	AGT	239
S	D	N	D	D	Y	M	K	K	G	L	F	F	S	R	S	78
AAT	GAT	CCC	TTA	GCT	TTA	CAG	GCT	CGT	TAT	GCT	CGA	ATG	TAC	GAT	GAT	287
N	D	P	L	A	L	Q	A	R	Y	A	R	M	Y	D	D	94
CTG	TTT	AAG	CTA	AAC	AAC	TAT	GAA	GTG	CCA	GAT	GAT	GTT	GTG	CGT	AGG	335
L	F	K	L	N	N	Y	E	V	P	D	D	V	V	R	R	110
CAT	GAC	AAC	AAG	ATT	TTA	GAC	ATA	ATA	TTA	AAA	GAG	TCT	TCC	GTG	CCA	383
H	D	N	K	I	L	D	I	I	L	K	E	S	S	V	P	126
TTC	TGG	TAT	GAC	ATA	TCG	GAT	GAC	GAA	GCT	CAT	GAA	TCA	ATG	CTG	CCT	431
F	W	Y	D	I	S	D	D	E	A	H	E	S	M	L	P	142
GAA	TTC	AGA	CTG	CAA	GAC	ATT	CAC	GAG	TTC	AGA	TTG	AAT	TTG	AAG	CGC	479
E	F	R	L	Q	D	I	H	E	F	R	L	N	L	K	R	158
GTA	CGC	GTA	GTA	CCT	GAC	GAG	TCT	GAA	GAG	ATT	CAG	ATG	GAT	GAA	AGT	527
V	R	V	V	P	D	E	S	E	E	I	Q	M	D	E	S	174
CAG	AGT	GAT	AAG	CGG	CGT	CGC	AAG	AAG	AGG	ATG	GAG	AAG	AGT	CGT	CCT	575
Q	S	D	K	R	R	R	K	K	R	M	E	K	S	R	P	190
GTG	TGG	CTG	TCG	GGA	TCG	GAA	AGC	GAC	AGG	CGT	ATA	GAA	TTG	AAC	GAC	623
V	W	L	S	G	S	E	S	D	R	R	I	E	L	N	D	206
TCT	TTG	AAA	CCC	AGT	CAA	AAA	TTT	GAA	ACC	AAG	TTA	TCG	TCT	TAT	TTA	671
S	L	K	P	S	Q	K	F	E	T	K	L	S	S	Y	L	222
TTA	AAT	AGA	TTA	ATG	AAT	GAG	ATG	AAT	CCT	CAT	TAT	TGT	GGG	CAT	CCG	719
L	N	R	L	M	N	E	M	N	P	H	Y	C	G	H	P	238
TTA	CCA	GCA	TTA	TTT	GTC	ACA	CTC	ATA	ATG	CTA	AAG	GCT	TAT	TCT	ATT	767
L	P	A	L	F	V	T	L	I	M	L	K	A	Y	S	I	254
AAG	AAC	AAG	TTC	TTT	TCA	TAT	GGT	ATA	AGA	TAT	ATG	GAG	CTC	GTG	TGT	815
K	N	K	F	F	S	Y	G	I	R	Y	M	E	L	V	C	270
AAT	GAG	ATC	GCT	GGT	CCG	GAT	CTC	AAT	ACC	CGA	ACG	TTC	CCA	GTT	TTA	863
N	E	I	A	G	P	D	L	N	T	R	T	F	P	V	L	286
TTT	GGA	AGT	GAC	GGT	TCA	TTT	GTG	GGC	ACA	CGA	GTA	TAT	TCG	CAC	TAT	911
F	G	S	D	G	S	F	V	G	T	R	V	Y	S	H	Y	302
CCA	ATA	AAG	CTA	CGT	ATG	ATA	CTG	AAC	GAC	TTA	ACT	TAT	TTA	TTA	ACC	959
P	I	K	L	R	M	I	L	N	D	L	T	Y	L	L	T	318
TAC	TCA	GAT	CTT	CAT	AAG	TTC	CAA	GAA	TTC	GAG	CTA	GAT	GTA	AAT	GAT	1007
Y	S	D	L	H	K	F	Q	E	F	E	L	D	V	N	D	334
GAG	GTT	CTA	CTT	CAC	ATG	TTG	CGC	ACG	CCG	AAT	GAT	GGG	AGG	CAA	CTT	1055
E	V	L	L	H	M	L	R	T	P	N	D	G	R	Q	L	350
AAG	AAG	GCG	GTG	ACA	AGG	TTA	AAT	CAC	TAT	TAT	GGA	TTG	AAG	TTC	AAT	1103
K	K	A	V	T	R	L	N	H	Y	Y	G	L	K	F	N	366
CCG	AAG	ACC	ACC	GAC	TGC	GGT	GTG	GTC	AAT	GGA	ATG	GAT	TTC	ACG	CAT	1151
P	K	T	T	D	C	G	V	V	N	G	M	D	F	T	H	382
AAG	CAT	CCG	ATC	ACT	AAG	ACC	GCT	GAC	TTT	ACG	TCC	CCA	GTC	TTA	CCC	1199
K	H	P	I	T	K	T	A	D	F	T	S	P	V	L	P	398

ATG	ACT	AAT	TCT	TTC	AAT	AAG	GCT	GAA	ATA	TGT	TAT	GGG	CAT	AAC	AGT	1247
M	T	N	S	F	N	K	A	E	I	C	Y	G	H	N	S	414
AAA	ATT	TTG	AAC	AGG	GCC	GTG	TTT	ACC	GAC	ACC	GTC	CGT	GGC	CAT	ATT	1295
K	I	L	N	R	A	V	F	T	D	T	V	R	G	H	I	430
CGC	GAG	GAC	CTT	AAG	AAT	GTG	GCT	GAC	TTA	GAC	TTG	CCA	AAA	CTG	TAT	1343
R	E	D	L	K	N	V	A	D	L	D	L	P	K	L	Y	446
GAG	CAT	GTT	TCT	AAA	CTA	GTG	GAT	ATG	CGT	GTT	AAT	TAC	ACA	ATT	ATC	1391
E	H	V	S	K	L	V	D	M	R	V	N	Y	T	I	I	462
TAT	GAT	CTC	ATG	TTT	CTA	CGA	GTC	ATG	TTG	AAT	CTT	GGT	GGC	TAT	TCT	1439
Y	D	L	M	F	L	R	V	M	L	N	L	G	G	Y	S	478
AGA	TCC	AAT	CAA	ATC	ACT	GAC	TTT	AGA	AAG	ACC	ATT	GAT	GAG	ATC	ACG	1487
R	S	N	Q	I	T	D	F	R	K	T	I	D	E	I	T	494
AAA	ATG	AAT	GAA	GGT	TTC	TTA	TCT	GGG	GCG	GAT	CCA	GAA	AAG	AAT	ATT	1535
K	M	N	E	G	F	L	S	G	A	D	P	E	K	N	I	510
GAC	ACA	TTG	AAT	GCT	TGG	ATG	GCG	CCT	ACA	ATG	GAG	GAT	TGT	GGT	TAT	1583
D	T	L	N	A	W	M	A	P	T	M	E	D	C	G	Y	526
CGA	CTT	ACT	AAG	TCT	ATC	TTG	TTC	GGT	AAA	TTT	AGG	AAA	GCC	AAA	TAC	1631
R	L	T	K	S	I	L	F	G	K	F	R	K	A	K	Y	542
CCA	TCT	GAT	TTG	GAA	GCT	AAG	TCC	AAT	ATT	GAT	TAT	TAC	GTA	ACT	GCT	1679
P	S	D	L	E	A	K	S	N	I	D	Y	Y	V	T	A	558
AGA	TCG	GCT	GGG	ATT	GGC	AAC	CTA	CGT	ATA	TCA	ATA	GAA	ACA	GAT	AAA	1727
R	S	A	G	I	G	N	L	R	I	S	I	E	T	D	K	574
AGA	AAG	TAC	AAA	GTA	CGG	ACT	ACA	TCG	AAG	TCC	GCC	TTT	GTT	AAC	GCC	1775
R	K	Y	K	V	R	T	T	S	K	S	A	F	V	N	A	590
ATG	GGT	TCT	GGA	ATA	CTT	GAT	GTT	AAT	CCC	GTC	AGT	AAT	GAA	CCT	ATG	1823
M	G	S	G	I	L	D	V	N	P	V	S	N	E	P	M	606
ATG	TTA	ACC	GAT	TAC	TTG	TTA	ACG	CAA	ACT	CCC	GAA	ACG	CGA	GCA	AAT	1871
M	L	T	D	Y	L	L	T	Q	T	P	E	T	R	A	N	622
TTA	GAG	GCG	GCT	ATA	GAT	AGC	GGA	TCT	AAA	TCC	GAC	TCT	GAA	TTA	ATG	1919
L	E	A	A	I	D	S	G	S	K	S	D	S	E	L	M	638
CGC	ATA	CTA	GGG	CAG	AAC	AGC	ATC	GGC	TCA	CGT	TCT	ACC	ACA	GCT	TGG	1967
R	I	L	G	Q	N	S	I	G	S	R	S	T	T	A	W	654
AGA	CCG	GTA	CGA	CCG	ATT	TAC	ATA	AAT	GTA	TTG	CAA	GCG	CAT	TTG	GCG	2015
R	P	V	R	P	I	Y	I	N	V	L	Q	A	H	L	A	670
CAG	GCT	TTC	ATT	ATC	GGG	CCT	CAC	ATA	AAT	GCA	ACT	GTT	AAT	CAG	CAT	2063
Q	A	F	I	I	G	P	H	I	N	A	T	V	N	Q	H	686
GAA	TAC	CAA	CCT	ACA	TCT	CTT	TGG	TTC	ACT	GGA	GAT	GAT	CTA	GGG	GTT	2111
E	Y	Q	P	T	S	L	W	F	T	G	D	D	L	G	V	702
GGG	TTT	GCT	ACA	CTG	TAT	CAA	AGC	GGA	ACC	GCG	GAT	ATA	ATC	GTT	CCA	2159
G	F	A	T	L	Y	Q	S	G	T	A	D	I	I	V	P	718
GCT	ATC	GAA	GCG	TCC	TCT	ACT	GGT	AAA	GCG	CTC	AGC	GTT	CTC	GCT	GAT	2207
A	I	E	A	S	S	T	G	K	A	L	S	V	L	A	D	734
TGC	TCA	TCC	TGG	GAT	CAG	ACG	TAT	CTG	ACG	GCG	ACG	ATG	ATT	CCG	TAC	2255
C	S	S	W	D	Q	T	Y	L	T	A	T	M	I	P	Y	750
TAT	AAT	GGA	ATA	AAA	AGA	GCT	TTG	CTG	GAA	TAC	CAA	CAA	GCG	GAT	ATG	2303
Y	N	G	I	K	R	A	L	L	E	Y	Q	Q	A	D	M	766
CGC	AAC	TTC	TAT	ATG	ATA	GAT	AGT	GGC	AGA	ACT	GGC	GTG	CCA	GGG	ATG	2351
R	N	F	Y	M	I	D	S	G	R	T	G	V	P	G	M	782
AAA	CTG	TCT	GAA	ATA	GTA	GAT	TGG	TTT	AAT	AGT	TTT	CAA	ACC	AAG	CGG	2399
K	L	S	E	I	V	D	W	F	N	S	F	Q	T	K	R	798

ATA I	TTC F	AAT N	GCC A	TCC S	TAT Y	TTG L	AAG K	GAA E	AGG R	CAT H	TCA S	TTT F	GTG V	GTT V	AAG K	2447 814
TAT Y	ATG M	TGG W	TCG S	GGA G	CGA R	CTG L	GAC D	ACT T	TTC F	TTT F	ATG M	AAT N	TCC S	GTC V	CAG Q	2495 830
AAT N	GCA A	TTG L	ATC I	ACC T	AGG R	AGA R	ATA I	GCT A	GAA E	GAA E	GTA V	TCT S	TTG L	AGG R	GTG V	2543 846
TCA S	AAC N	ACT T	GGC G	TTA L	TCC S	TGG W	TTT F	CAG Q	GTT V	GCT A	GGC G	GAT D	GAC D	GCC A	ATC I	2591 862
ATG M	GTA V	TAC Y	GAT D	GGA G	TCA S	TCA S	ATC I	TCC S	ACA T	ACT T	GAG E	CAA Q	GTC V	ACT T	AGA R	2639 878
ATT I	AAT N	GAA E	ATC I	ACG T	GTT V	CGT R	AAT N	TAT Y	GAG E	GAG E	TCA S	AAT N	CAC H	ATA I	ATT I	2687 894
AAT N	CCG P	CAA Q	AAA K	ACG T	GTA V	ATT I	TCA S	CAT H	ATT I	TCT S	GGT G	GAG E	TAT Y	GCT A	AAA K	2735 910
ATA I	TAT Y	TAT Y	TAT Y	GCA A	GGA G	ATG M	CAT H	TTT F	CGC R	GAT D	CCT P	TCG S	ATT I	CAG Q	CTG L	2783 926
CAT H	GAA E	TCT S	GAA E	AAG K	GAT D	TCT S	GGA G	GCA A	AGT S	GAC D	GTG V	ACC T	GAG E	TCC S	CTA L	2831 942
AGG R	GAG E	TTT F	GGT G	CAA Q	GTC V	ATA I	TAT Y	GAA E	TAT Y	AAC N	AAA K	AGA R	GCG A	ATT I	GGT G	2879 958
ACC T	CTA L	CGC R	GTG V	AAC N	GCT A	TTA L	TAT Y	GGT G	AGG R	CTA L	ATA I	GCC A	GGG G	CTA L	GCA A	2927 974
TAT Y	AGT S	GTA V	AAC N	TGT C	CCG P	CAG Q	TAT Y	GAC D	GCT A	TCA S	AAA K	AGA R	ACA T	TAT Y	GCG A	2975 990
AAT N	ATG M	AAA K	TAT Y	TAT Y	CCG P	CCA P	CCA P	ACC T	TCA S	GTA V	ATT I	GCT A	CCA P	GCC A	GCA A	3023 1006
TTT F	AAG K	GGA G	GGA G	TTG L	GGA G	TTG L	TCA S	TTT F	ACT T	GGG G	TTA L	TCG S	TTG L	AAT N	GAG E	3071 1022
GTG V	CTA L	TTT F	ATA I	AAA K	TTG L	CAT H	CTT H	CAT H	GAA E	GCG A	GTA V	TCT S	CAG Q	GGT G	CTT H	3119 1038
CAC H	GTG V	ATT I	TCA S	ATG M	ATT I	TCA S	TTT F	GAA E	GCC A	AAT N	GAA E	GTG V	GTT V	TCA S	AAC N	3167 1054
TCT S	CTG L	TCC S	GCG A	TAT Y	TAT Y	CTT H	AAG K	GAC D	CAG Q	AAA K	GAT D	TTA L	CTG L	CGT R	GAC D	3215 1070
ATG M	AAG K	TTA L	GGC G	AAG K	CAT H	CTT H	GAA E	AAA K	GTT V	AAA K	GGT G	ATA I	TCC S	TTT F	AAA K	3263 1086
TCC S	TCT S	GAT D	TTA L	GCC A	TTT F	TCG S	GGT G	AGT S	GAT D	TTC F	TCG S	CAA Q	GGG G	CTA L	AAC N	3311 1102
CTG L	AAG K	AGG R	GAA E	TCT S	ATT I	GAT D	AAA K	GTG V	AAG K	CTT H	GAA E	GTG V	TCT S	CGC R	AAA K	3359 1118
TCT S	ATC I	CGT R	GAT D	TTG L	CGC R	TCT S	TCG S	GGT G	ATT I	AGT S	GTG V	CCA P	TCA S	ACG T	CAC H	3407 1134
GCA A	TAT Y	GAA E	AAT N	CTG L	CCA P	TAT Y	GCA A	TCA S	CTG L	CAT H	CAG Q	TCT S	TTC F	AAA K	AGT S	3455 1150
TTA L	AAA K	GTG V	GAT D	AGA R	GAC D	ACT T	TCG S	AAA K	TTT F	ACG T	AAT N	GAA E	CGG R	CTA L	CTG L	3503 1166
GTT V	TCA S	CTC L	CTT H	GAG E	TAT Y	AAA K	TCA S	GAC D	ATT I	CCT P	AGA R	GTA V	AGT S	GTT V	ACT T	3551 1182
TCG S	CAG Q	TAT Y	CCA P	GTT V	TAT Y	GAC D	TTG L	ATC I	AAC N	ATC I	TCA S	AAA K	GTT V	GAC D	GAA E	3599 1198

TTA	AAT	GTT	AGA	TCT	GGT	GGA	CCA	GTC	AGG	TTC	ATC	TCT	GCT	CCA	ATA	3647
L	N	V	R	S	G	G	P	V	R	F	I	S	A	P	I	1214
GAA	GGT	AAG	CTA	TTG	GAA	GAA	AAT	ATT	GGT	ACC	CGT	CAA	GGA	GTT	CAG	3695
E	G	K	L	L	E	E	N	I	G	T	R	Q	G	V	Q	1230
TTT	AAA	AAC	CGA	GGG	TAC	GGA	GGA	TCC	CAA	GAA	GTA	TTA	CAC	TTC	ATA	3743
F	K	N	R	G	Y	G	G	S	Q	E	V	L	H	F	I	1246
AGA	TCC	AAC	GGC	TTG	GTG	ATA	ACT	GAA	CAG	GCT	CTG	ATT	GAC	TTA	ATC	3791
R	S	N	G	L	V	I	T	E	Q	A	L	I	D	L	I	1262
ATT	AAG	TCA	GGA	GTA	TG	CTA	ATG	ATA	AAT	CCA	CAA	CGT	GGG	TTG	ATT	3839
I	K	S	G	V	L	L	M	I	N	P	Q	R	G	L	I	1278
GAC	CTT	TTT	CAG	TCC	CTT	TCT	GGA	GAT	ACC	GCT	TCA	TCC	ATG	CAT	CTA	3887
D	L	F	Q	S	L	S	G	D	T	A	S	S	M	H	L	1294
GCG	AAC	TTC	TTC	ATG	GCT	GAA	AAG	CCG	CAC	TGG	GAA	GAT	AAT	GCC	ATA	3935
A	N	F	F	M	A	E	K	P	H	W	E	D	N	A	I	1310
TCT	CTC	ACC	ATT	GCG	GGG	TCA	TTG	CTT	GAG	AAC	TGC	GAT	AGT	AGA	ATT	3983
S	L	T	I	A	G	S	L	L	E	N	C	D	S	R	I	1326
GAG	AAC	GTA	AAG	AAT	TTT	GTC	AGC	GTC	CTT	GCT	ACT	GGC	ATG	CAA	AAG	4031
E	N	V	K	N	F	V	S	V	L	A	T	G	M	Q	K	1342
GAC	CTA	CAG	AGG	ATG	TTC	TAT	TAC	GTC	GGA	TTT	GTC	TAC	TAT	GCA	CAG	4079
D	L	Q	R	M	F	Y	Y	V	G	F	V	Y	Y	A	Q	1358
AGG	CTT	ATA	TGG	TCT	GGT	GGT	CAC	TCC	AGT	AAA	ATA	TTC	GTC	TCC	ATT	4127
R	L	I	W	S	G	G	H	S	S	K	I	F	V	S	I	1374
GAT	GAG	GAT	AAG	CTC	GCT	GAC	TTC	CTT	AGA	GGA	TCT	AAA	CCA	ATC	ACT	4175
D	E	D	K	L	A	D	F	L	R	G	S	K	P	I	T	1390
CGG	AGG	AGG	AAA	GCC	ATG	GCT	GGA	ACT	AAG	CGT	GAA	CCG	ATC	AAT	TTA	4223
R	R	R	K	A	M	A	G	T	K	R	E	P	I	N	L	1406
AGC	GCT	AAT	TTC	AGT	TAT	GAA	ATA	TCT	GAA	CCG	GAT	AGG	GAG	ATT	AGT	4271
S	A	N	F	S	Y	E	I	S	E	P	D	R	E	I	S	1422
GAG	TAC	GAC	CCA	CTA	ATC	TTA	TGT	CAT	CCC	CTA	TCA	ATG	CCA	TTT	TTT	4319
E	Y	D	P	L	I	L	C	H	P	L	S	M	P	F	F	1438
GGT	AAT	TGG	CAG	GAG	AAA	TAC	TCT	GTG	ATG	CAG	AGC	GAC	GAG	CAA	ATG	4367
G	N	W	Q	E	K	Y	S	V	M	Q	S	D	E	Q	M	1454
TGA	TTGATCGCCAGGCTATCCGTTGGGATAGGTTAATCAGCCGATCATATATGAT															4422

图1 RDV中国福建分离物S1片段核苷酸序列及编码蛋白的氨基酸序列

Fig.1 Nucleotide sequence and deduced amino acid sequence of RDV Fujian isolate segment S1

1.2.4 基因克隆及序列分析: 扩增产物经 T4 DNA polymerase 补平, 用氯仿抽提, 乙醇沉淀后, 溶于双蒸水中, 与 EcoRV 切开的 pBluescript(SK) 载体连接, 转化 *E.coli* DH5 α , 经选择培养基 Ampicilin(X-gal/IPTG) 筛选后, 煮沸法快速筛选重组子并酶切鉴定, 根据酶切图谱进行亚克隆并测序, DNA 序列采用 Sanger 双脱氧法, 质粒 DNA 制备用碱法和 PEG 纯化获得。

1.2.5 DNA 序列分析和同源性比较: 应用 HITACHI SOFTWARE ENGINEERING CO.LTD 1990 年 DNASIS 软件。

2 结果

2.1 RDV dsRNA 的提取和分离

从纯化病毒悬浮液中抽提病毒总 RNA, 在 1.2% 低熔点琼脂糖凝胶上进行电泳分离,

```
rdvcs1      EAKSNIDYYVTARSAGIGNLRISIETDKRKYK----RTTSKSAFVNAMG
rdvjs1      EAKSNIDYYVTARSAGIGNLRISIETDKRKYK----RTTSKSAFVNAMG
rota        IQDWLVDCSIEKFPLMAKIYSWSFHVGFRRQKMLDAALDQLKTEY--TED
reo         DTTSSIARYLY-RTWYLAARMAAQTRTWDP LF----QAIMRSQYVTARG
BTV         MILASFLLSIQTTITGYGRAWVKNVSTEFDKQ-----LKPNPSNLVR--D
rdvcs1      SGI-LDV-NPVSNEPMMMLTDYL-LTQTPETRANLEAAIDS-GSKSDSELM
rdvjs1      SGI-LDV-NPVSNEPMMMLTDYL-LTQTPETRANLEAAIDS-GSISDSELM
rota        VDD-EMYREYTMILIRDEVVKMLEEPVKHDDHLLQDSELAGLLSMSSAS-N
reo         GSG-AALRESLYAINVSLPDFKGLFVKAAATKIFQAAQLAN-LPFSHTSVA
BTV         VSD-LT-REFFKQAYVEAKERREEMVKPEDLYTSMRLRLARNTS-SGFSTE
rdvcs1      RILGQNSIGS-RSTTAWRPVRPIYI-----NVLOAHLAQAFIIGPH-INA
rdvjs1      RILGQNSIGS-RSTTAWRPVRPIYI-----NVLOAHLAQAFIIGPH-INA
rota        GESRQLKFG--RKTIFSTKKMMHVM-----DDMANGRYTPGIIPPVNVDK
reo         -ILADTSMGL-RNQVQRRPRSIMPL-----NVPQQQVSAPHTLTADYINY
BTV         IYVKKRFGPRLRDKDLVKINSRIKALVIFTKGHTVETDEELHKKYNSVEL
          ***
rdvcs1      TVNQHEYQPTSLWFTGDDLGVGFATLYQSGTAD--IIVPAI-----
rdvjs1      TVNQHESQPTSLWFTGDDLGVGFATLYQNGTAD--IIAPAI-----
rota        PIPLGRDVPGRRTRIIFILPYEYFIAQHAVVEKMLIYAK-----HTREYA
reo         HMNLSPTSGSAVIEKVIPLGVYASSPPNQSIINI-----
BTV         YQTGSRDVPKATRTIYSINLSVLVLPQLIVTLPLNE----YFSRVGGITR
          * * *
rdvcs1      EASSTGKAL-SVLA-----DCSSWDQT
rdvjs1      EASSTGKAL-SVLA-----DCSSWDQT
rota        EFYSQSNQLLSYGDVTRFLSNNSMVLYT-----DVSQWDS
reo         PDYKKIGGVIVGDLEATGSRVMDAADCFRNSADRDIFTIAIDYSEYDTH
BTV         -----YLATMIPYYNGIKRALLEYQQADMRFYIMDSGRT-GVPGMKL
rdvcs1      -----FLTATIIIPYYNGIKKALLEYQQADMRFYIMDSRT-GVPGMKL
rdvjs1      -----QHNTQ--PFRKGIIMGLDML--ANMTNDARVIQTLN----LYKQ
rota        ITWDFFLSVIM-AAIHEGVASSSIGK--PFMGYPASIVNDES----VVG
reo         LTRHMFRTGML-QGIREAMAPYR-AL--RYEGYTTLEQIIDFG----YGE
BTV         SEIV---DWFNSFQTKRIFNA-----
rdvcs1      SEIV---DWFNSFQTKRIFNA-----
rdvjs1      TQINL-MDS-YVQI-----
rota        VRAAR-PISGMOMNIOHLSKLYKRG-----FSYRVND
reo         GRVANTLWNGKRLFKTTFDAYIRLDESERDKGSFKVPKGVLPVSSVDVAN
BTV         -----
          * * *
rdvcs1      -----SYLKERHSFVVKYMWSGRLDTFFMNSVQN---ALIT
rdvjs1      -----SYLKERHSFVVKYMWSGRLDTFFMNSVQN---ALIT
rota        -----PDGNVIKKIQYGAVASGEKQTKAANSIANL---ALIK
reo         -----SFSFGNDFTHMTTFPSGSTATSTEHTANNS--TMET
BTV         RIAVDKGFDTLIAATDGSDL-ALIDTHLSENGSTLIANSMHMMAIGTLI-
          ***
rdvcs1      -RRIAEE-----VSLRVNSNTGLSWFQVAGDDAIMVYDGSSI-STTEQ
rdvjs1      -RRIAEE-----VSLRVNSNTGLSWFQVAGDDAIMVYDGSSI-STTEQ
rota        --TVLS-----RISKYSFATKI-IRVDGDDNYAVLQFNTE-VTEQM
reo         FLTVMVGPEHTDDPDVLRMLKSLTIQRNYVCQGGDGLMIDGTTA-GKVNS
BTV         QRAVGREQP-GILTFLS-----EQYVGGDDTLFYTKLHTDITVFD
          * * *
rdvcs1      VTRINEIT-VRNYEES--NHIINPQKTVISHIS-GEYAKIYYYAGMHFRD
rdvjs1      VTRVNEIT-VRNYEES--NHIINPQKTVISHIS-GEYAKIYYYAGMHFRD
rota        VQDVSM DV-RETYARM--NAKV---KALVSTVG-IEIAKRYIAGGKIIFR
reo         ETIONDLELISKYGEFGW-----KYDIA YDGTAEYKLKLYIFGCRIPN
BTV         KVAASIFD TVAKCGHE-----ASPSKTMMT-----
rdvcs1      PSIQ LH-ESEKDSGASDVTESLREFGQVIYEYNKRAIGTLR
rdvjs1      PSIQ LH-ESEKDSGASDVTESLREFGQVIYEYNKRAIGTLR
rota        AGINLL-NNEKRGQSTQWDQAAVLYSNYIVN-RLRGFETDR
reo         LSRHP IVGKERANSAAEFWPAILDQIMGVF-FNGVHD-GL
BTV         -----
```

图 2 呼吸孤病毒科成员间包含聚合酶保守模式区域的氨基酸序列比较

Fig. 2 Alignment of partial amino acid sequences which contain viral polymerase module region of four members of the family Reoviridae

RDVCS1. Partial amino acid sequence of RDV Fujian isolate S1; RDVJS1. Partial amino acid sequence of RDV Japanese isolate S1; Rota. Partial amino acid sequence of the bovine rotavirus RF stain RDRP; Reo. Partial amino acid sequence of reovirus serotype 3 RDRP; BTV. Partial amino acid sequence of the bluetongue virus serotype 10 RDRP.

12 条 RNA 可以分离成 9 条带,回收第一条带,即是 RDVS1 组份。

2.2 RDVS1cDNA 的合成及 PCR 扩增

根据已发表的日本流行株 S1 的全序列合成 6 个引物,分三段合成 cDNA。RDVS1 双链 RNA 经 DMSO 变性后,合成 cDNA 第一条链,然后经 PCR 扩增得到各段的 PCR 产物,各约为 1.4kb,与预计的片段大小相近。

2.3 扩增产物的克隆、鉴定和序列测定

PCR 产物经克隆筛选,根据已发表的 S1 序列的酶切图谱进行酶切鉴定,酶切结果表明,其中有些位点存在(如 54~59bp 的 EcoRV 和 1517~1522bp 的 BamHI),有些位点丢失(如 420~425bp 的 ClaI 和 2273~2278bp 的 HindIII),同时也突变出了新的位点(如 984~989bp 的 EcoRI 和 3675~3680bp 的 KpnI)。在酶切结果的基础上进行亚克隆,通过对所获得的三段克隆的序列分析得到 S1 全序列,见图 1。RDV 中国福建分离物 S1 由 4422 个碱基组成,与日本流行株系 S1 相比二者核苷酸同源率为 95%。该序列已被 GenBank 接受,接受号码为 U73201。

2.4 RDVS1 编码的蛋白

RDVS1 开放阅读框架有 4332 个碱基。根据 RDVS1 基因的核苷酸序列翻译成的氨基酸序列,发现 RDVS1 编码一个由 1444 个氨基酸组成的蛋白(P1)(图 1),分子量为 164kD,其氨基酸序列与 RDV 日本株 S1 相应 ORF 编码的蛋白在氨基酸水平上同源率为 97.7%。

对 P1 进行氨基酸序列分析表明,序列中(图 1)含有 RDRP 保守序列: motif I (DXXXXD)、motif II (SGXXXTXXXN) 和 motif III (GDD),这些 motif 集中位于 687-850aa 之间,除此以外,模式 III 后 897~904 位还存在一个很保守的区域 (EXXKXY)(图 2)。

3 讨论

我们对克隆到的 RDVS1 不同片断及不同亚克隆进行序列分析发现, RDVS1 由 4422nt 组成,与日本流行株系序列相比,同源率为 95%,总共有 217 处突变,其中同类碱基之间的转换($G \leftrightarrow A$ 或 $T \leftrightarrow C$)为 191 处,占 88%。在本实验室获得的 RDV 其他各片段的变异中也有这种特点^[11~20]。我们推测,这一现象的原因除了在 PCR 过程中可能发生错误配对外,主要是 RDRP 没有 3'→5' 的外切酶的功能,不能在复制过程中校正错配的碱基对,而在 RNA 分子中,除了 A-U、G-C 之间的正常配对外,G-U 之间也可以形成较弱的配对,这样在转录和复制过程中,发生错配的碱基对中以 G-U 配对为多数,从而导致了 $G \leftrightarrow A$ 、 $C \leftrightarrow U$ 的突变。这种突变对病毒进化的意义,目前不清楚。同时我们发现突变多是发现在密码子的第三位碱基上,点突变率的 78.3%,这种多数不引起氨基酸序列变化的突变占显著多数,说明病毒在进化过程中有保持相对稳定的机制。同时我们也看到,含有保守模式的蛋白质区域中的突变明显少于其他部分,保证了功能区的作用。

对 RDVS1 采用的密码子进行偏爱性分析,XYA 有 373 个,XYT 有 467 个,XYC 有 282 个,XYG 有 348 个, $(XYA+XYT)/(XYG+XYC)=1.34$,与碱基组成 $A+T/G+C=1.39$ 基本相符。在 RDV 的其他片段中,把以单子叶植物为宿主的 RDV 与以双子叶为宿主的 WTV(wound tumor virus)进行比较,人们发现 RDV 的 G+C 含量以及 XYG+XYC 的百

分比比 WTV 高^[3,7],而单子叶植物的 G+C 含量以及 XYG+XYC 的百分比要高于双子叶植物。RDVS1 编码蛋白是否有此规律,还要有待 WTVS1 的序列测定。这种病毒与其宿主植物碱基组成和密码子偏爱性的相似性,可以说明由于病毒借助宿主的翻译系统合成蛋白质,所以在进化过程中,一些宿主的翻译系统偏爱的密码子突变由于在生存竞争中占优势而保留下来,形成了这种规律。

RDVS1 有两个开放阅读框架,其中微型顺反子 (6nt-29nt) 编码七个氨基酸,而主要的 ORF 起始于 36nt,终止于 4367nt,编码 1444 个氨基酸,分子量 164kD,这种微型顺反子现象在动物细胞和病毒的 mRNA 有报告,但在植物呼肠孤病毒中没有发现。体外翻译研究结果表明^[2],用全长 cDNA 的转录产物翻译时,164kD 的蛋白翻译效率极低,而产生一些更小的蛋白产物;当通过点突变把微型顺反子的 ATG 突变成 AAG 时,翻译出 170kD 蛋白,同时伴有更小的一些蛋白产物。170kD 蛋白在 SDS-聚丙烯酰胺凝胶电泳中,迁移率与 RDV 最大的核心蛋白相同。微型顺反子功能不清楚,推测可能是一个衰减子可以在翻译水平上控制前导区转录的终止。根据体外翻译的 P1 蛋白的迁移率,推测 S1 编码 RDV 的一个核心蛋白,而出现一些更小的蛋白产物可能是由于在翻译过程中由于 S1 的 ORF 很长而出现提早终止,或基因内部 AUG 起始翻译或大的翻译产物的降解。在两种情况下,没有检测到微型顺反子的翻译产物。

在主要 ORF 的氨基酸序列中,有 RDRP 特异的保守模式,位于 687~850 氨基酸残基间,包括模式 I (DXXXXD)、模式 II (SGXXXTXXXN) 和 2 个模式 III (GDD),并且在模式 III 之后还存在一个很保守的区域 EXXKXY (897~904aa) 这一保守区域也在其他轮状病毒株 (蓝舌病毒除外) 和呼肠孤病毒血清类型的病毒的 RDRP 中均存在 (图 2)^[18~20],但其重要性尚不清楚。除了在这一段区域中集中分布的保守模式外,在 162~167 还有一个 motif (DXXXXD)。结合体外转录和翻译结果,由于 RDVS1 的翻译产物在 SDS-聚丙烯酰胺电泳中的迁移率与 RDV 最大的核心蛋白相同,用 RDVS1 编码区部分序列的体外表达产物制备得到的抗体与 RDV 最大的核心蛋白 (164kD) 有特异的反应,从而认为 RDVS1 编码 RDV 最大核心蛋白。在呼肠孤病毒基因中最大的 dsRNA 片段编码的核心蛋白为 RDRP,由此可推测 RDVS1 编码的蛋白是一种 RDRP,在与核心颗粒有关的转录酶活性中起关键酶的作用。目前,RDV 第三号片段编码的核心蛋白也发现有一个 RDRP 类似序列模式,因此有待进一步确定 RDVP1 和 P3 在复制和转录中的作用。

另外,将获得的 RDVP1 氨基酸序列送入数据库中比较发现,它与人的 DNA 拓扑异构酶 II 有一定的同源性 (图 3)。DNA 拓扑异构酶在真核生物 DNA 复制过程中作用,解决解旋时产生的螺旋张力。在 RDV 中,可能没有单独的拓扑异构酶,但 RDV 是 dsRNA 病毒,复制过程也需要解开双螺旋结构,所以很可能 RDRP 具有一定的解旋功能以维持复制的正常进行,这种推测有待进一步证实。

由于 RDV 是双链 RNA 病毒,其外壳蛋白由多个基因编码,不象由单一蛋白外壳包装的 ssRNA 病毒那样容易获得 CP 介导的抗性。因此在目前复制酶介导的病毒抗性取得一些成功的背景下,通过对 RDRP 基因的克隆及其以后进一步的蛋白功能和水稻转化工作,不仅可以有利于揭示 RDV 这个具有复杂形态结构和基因组成的病毒的自身的复制组装机制和阐明病毒-昆虫介体-植物宿主三者之间的关系,而且对进一步研究复制酶介导的

```

A  LSGESDRRIELNDSLKPSQKFETKLSSYLLNRL-----NEGFLSGADPEKNIDTLNWMAPTMED
  ++GS D ++ LN + P + F + + YL ++L      ++ +S A +K ID W+ ME
B  IAGSTKDVKVFLNGNKL PVKGFRSYVDMYLLKDKL-----DDAAISLAFSKKQIDDRKEWLTNFMED

A  LGQNSIGSRSTTAWRPV-----SWDQTYLTATMIPYNGIKR-----SNHIINPQKTVISHISGE
  L QN + S + +P+ +W QTY + P NG ++ SN + POK+V+S + +
B  LAQNFVGSNNLNLQPI-----TWTQTYKEQVLEPMLNGTEK-----SNKELKPQKSVVSDLEAD

A  YAK-----RRRKAMAGTKREP
  K ++R A GTRK+P
B  DVK-----KKRAAPKGTGRDP

```

图3 RDV S1 中国福建分离物 ORF 编码蛋白 P1 与人 DNA 拓扑异构酶 II 氨基酸同源性比较

Fig.3 Amino acid sequence homology comparison of RDV Fujian isolate segment S1 ORF encoding protein with human DNA Topoisomerase II

A. Partial amino acid sequence of RDV Fujian isolate S1;

B. Partial amino acid sequence of human DNA Topoisomerase II.

抗病毒机制和获得复制酶介导的抗 RDV 水稻有很大帮助。

参 考 文 献

- [1] Inoue H, Timmins P A. *Virology*, 1985, **147**: 214~216.
- [2] Suzuki N, Tanimura M, Watanabe Y et al. *Virology*, 1992, **190**: 240~247.
- [3] Suzuki N, Watanabe Y, Kusano T et al. *Virology*, 1990, **179**: 446~454.
- [4] Suzuki N, Sugawara M, Kusano T. *Virology*, 1992, **191**: 992~995.
- [5] Uyeda I, Suda N, Yamada N et al. *Intervirology*, 1994, **37**: 6~11.
- [6] Zhang F J, Li Y, Liu Y F et al. *Acta Virologica*, 1997, **41**: 161~168.
- [7] 赵晓岚, 李 毅, 陈章良等. 微生物学报, 1996, **36**(2): 93~102.
- [8] 李 玮, 李 毅, 陈章良等. 应用基础及工程学报, 1994, **2**(2): 109~116.
- [9] 刘一飞, 李 毅, 陈章良等. 病毒学报, 1994, **10**: 246~250.
- [10] 曲 林, 李 毅, 陈章良等. 微生物学报, 1996, **36**(5): 335~343.
- [11] 李 玮, 李 毅, 陈章良等. 病毒学报, 1995, **11**(1): 56~62.
- [12] 曲 林, 李 毅, 朱玉贤等. 病毒学报, 1995, **11**(3): 271~275.
- [13] 储瑞银, 张 旭, 陈章良等. 病毒学报, 1993, **35**: 115~120.
- [14] 肖 锦, 李 毅, 陈章良等. 生物工程学报, 1996, **12**(3): 361~366.
- [15] 李 毅, 薛志宏, 陈章良等. 病毒学报, 1994, **10**: 339~345.
- [16] Xu H, Li Y, Chen Z L et al. *Virology*, 1997, **240**: 267~242.
- [17] Zheng H H, Li Y, Chen Z L et al. *Theor Appl Genet*, 1996, **94**: 522~527.
- [18] Cohen J, Charpilienne A, Chilmonczyk S et al. *Virology*, 1989, **171**: 131~140.
- [19] Wiener J R, Joklik W K. *Virology*, 1989, **169**: 194~203.
- [20] Roy P, Fukusho A, Ritter G D et al. *Nucleic Acids Res*, 1988, **16**: 11759~11767.

NUCLEOTIDE AND PROTEIN SEQUENCE ANALYSIS OF RICE DWARF VIRUS REPLICASE(SEGMENT S1)^{*}

Xiao Jin¹ Li Yi^{1,2**} Zhang Jing¹ Liu Jinfeng¹ Chen Zhangliang^{1,2}

(¹ *The National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871*)

(² *De Montfort University, The Gateway Leicester, LE19BH, UK*)

Abstract The full length cDNA of Rice Dwarf Virus (RDV) Fujian isolate genome segment S1 was cloned and full length sequence was determined. The results showed that S1 is 4422bp in length and contains a major open reading frame which encodes a polypeptide with 1444 amino acids. The major ORF contains the RNA-dependent RNA polymerase (RDRP) consensus sequences, such as motif I (DXXXXD) motif II (SGXXXTXXXN) and motif III (GDD). In addition to the three motifs, a well-conserved region (EXXXXY) is found following motif III. Therefore, we suggest that the protein encoded by S1 encodes the virus RDRP. Homology comparison of the nucleotide and amino acid sequences of the RDV Fujian isolate with the known sequences of the Japanese isolate showed that the both isolate have 95% and 97% on nucleotide and amino acid sequences, respectively. The sequence data have been deposited with GenBank database under the accession no. U73201.

Key words Rice dwarf virus (RDV), Segment S1, Sequence analysis, RNA-dependent RNA polymerase (RDRP)

^{*} The work was supported by 863 grant (863-101) from National Science and Technology Commission of China.

^{**} To whom correspondence should be addressed.