

猪传染性胃肠炎病毒 SC-Y 株编码复制酶聚蛋白 ORF1 的克隆及结构特征分析

宋振辉^{1,2}, 郭万柱^{1*}, 殷华平¹, 曾智勇¹, 朱玲¹, 骆爱芳¹, 张鹦俊³

(1. 四川农业大学动物生物技术中心, 雅安 625014; 2. 西南大学荣昌校区动物医学系, 重庆 402460; 3. 西南大学荣昌校区图书馆, 重庆 402460)

摘要: 应用 RT-PCR 方法扩增了猪传染性胃肠炎病毒 SC-Y 株编码复制酶聚蛋白的 ORF1 序列, 将其进行克隆、序列测定和分析。确认 SC-Y 株 ORF1 全长 20 053 nt (GenBank 收录号 DQ390461)。该序列包含 2 个 ORF, 其中 ORF1a 由 12 053 个核苷酸构成, 可编码由 4 018 个氨基酸组成的多肽, ORF1b 由 8 036 个核苷酸构成, 可编码由 2 678 个氨基酸组成的多肽。对 SC-Y 与 TGEV 参考毒株及不同冠状病毒对应区进行序列比较, 结果显示, SC-Y 株与 PUR46-MAD 株的同源性最高, ORF1a 的核苷酸同源性为 99.5%, 推导氨基酸同源性为 99.2%, ORF1b 的核苷酸及推导氨基酸同源性均为 99.8%; 不同冠状病毒之间 ORF1b 比 ORF1a 具有更高的保守性。

关键词: 猪传染性胃肠炎病毒; 编码复制酶聚蛋白 ORF1

中图分类号: S852.65⁺9.6

文献标识码: A

文章编号: 0366-6964(2007)08-0822-05

Cloning and Analysis of Replicase Polyprotein-Encoding ORF1 of Transmissible Gastroenteritis Virus SC-Y Strain

SONG Zhen-hui^{1,2}, GUO Wan-zhu^{1*}, YIN Hua-ping¹, ZENG Zhi-yong¹,
ZHU Ling¹, LUO Ai-fang¹, ZHANG Ying-jun³

(1. Animal Biology Technology Center of Sichuan Agricultural University, Yaan 625014, China; 2. Department of Animal Medical Science, Southwest University, Chongqing 402460, China; 3. Library of Southwest University, Chongqing 402460, China)

Abstract: In this study, the complete sequence of the replicase polyprotein-encoding ORF1 of TGEV SC-Y strain was obtained by RT-PCR, the entire nucleotide sequence was determined (GenBank Accession No. DQ390461). Results showed that there were two partially overlapping open reading frames in replicase polyprotein-encoding ORF1 of SC-Y strain, the ORF1a contains 12 053 nucleotides encoding a polypeptide of 4 018 amino acids, the ORF1b contains 8 036 nucleotides encoding 2 678 amino acids. The sequences were compared among different TGEV strains and other coronary viruses. The results showed that the SC-Y has the highest homology with PUR46-MAD strain, the ORF1a and ORF1b nucleotide sequences had 99.5% and 99.8%, amino acid sequences had 99.2% and 99.8% respectively. The ORF1b is highly conserved among the different coronaviruses.

Key words: porcine transmissible gastroenteritis virus; replicase polyprotein-encoding ORF1

猪传染性胃肠炎是由猪传染性胃肠炎病毒 (Porcine transmissible gastroenteritis virus,

收稿日期: 2006-08-10

基金项目: 教育部长江学者和创新团队发展计划资助 (IRT0555)

作者简介: 宋振辉 (1976-), 内蒙古通辽人, 博士生, 从事传染病病原分子生物学研究

* 通讯作者: 郭万柱, 教授, 博士生导师, E-mail: wzguo@126.com

TGEV)引起的一种以严重腹泻、呕吐和脱水为临床症状的高度接触性传染病。TGEV 属冠状病毒科冠状病毒属成员,基因组为单股正链不分节段的 RNA,长约 28.5 kb。TGEV 在感染细胞内产生 7 种亚基因组 mRNA,这些 mRNA 分子在 3'-末端具有嵌套结构。其中基因组 mRNA 15'端约 20 kb 为一大开放的阅读框 ORF1,编码病毒复制酶聚蛋白,研究证明参与冠状病毒复制和转录的多种蛋白都由 ORF1 编码^[1],该序列又可分为相互重叠的 2 个开放阅读框(ORF),即 ORF1a 和 ORF1b,ORF1a 编码分子量约为 450 ku 的多聚蛋白 ppla,ORF1b 通过核糖体移码机制(Ribosomal frameshift)与 ORF1a 形成分子量约为 750 ku 的 pplab。ORF1a 含有 3 个蛋白酶区,即 2 个木瓜蛋白酶样区(Papain-like proteinase domain-1/2, PL-1/2)和 3C 样蛋白酶区(3C-like proteinase, 3CL^{pro})。研究表明,TGEV PL-1/2 和 3CL^{pro}可分别裂解 ppla/pplab 蛋白的 N 端和 C 端,而释放一系列非结构蛋白,它们与核衣壳蛋白和一些胞内蛋白共同构成复制酶-转录酶^[2]。尽管目前对 TGEV ORF1 中的主要功能区域有所鉴定,但仍对多种与病毒复制相关的蛋白缺乏可靠的研究方法,同时关于 TGEV ORF1 的材料较少,限制了对 TGEV 复制机理及生物特性的研究。

目前对 TGEV 的研究多集中在某些结构基因序列及编码产物功能分析上,非结构基因的研究较少,尤其是 5'端 ORF1 序列,由于该序列片段较长,结构复杂,编码产物较多,因此限制了对其深入研究。鉴于此,本研究对我国分离株 SC-Y 编码复制酶聚蛋白 ORF1 进行了基因克隆及结构特征分析,为进一步开展 TGEV 基因组分子特征研究以及 SC-Y 复制转录机制研究奠定了基础。

1 材料与方 法

1.1 毒株与试剂

TGEV SC-Y 株(由本室分离保存)、ST 细胞、DH5 α 由本实验室保存,pMD18-T 载体购自大连 TaKaRa 公司,反转录酶 superscript III、RNA 抽提试剂盒为 Invitrogen 产品,pGEM-T 购自 Promega。

1.2 引物设计

参考 GenBank 收录的 PUR46-MAD 株全基因序列(AJ271965)和 Purdue-115 株复制酶聚蛋白基

因序列(Z34093),设计 11 对引物对目的基因进行扩增,引物由宝生物工程(大连)有限公司合成,各引物核酸序列及在基因组中的位置见表 1。

1.3 病毒 RNA 提取

常规方法接种 TGEV SC-Y 病毒于 ST 细胞,出现明显的 CPE 时,收获细胞。按试剂盒说明操作抽提总 RNA。

1.4 RT-PCR 扩增

参照分子克隆实验指南^[3]进行。

1.5 PCR 产物的克隆与测序

将扩增的 11 个片段分别进行胶回收,与 pMD18-T 或 pGEM-T 载体连接,转化 DH5 α 感受态细胞后,经 Amp、IPTG、X-gal 筛选,PCR 鉴定阳性重组质粒送宝生物公司进行测序。

1.6 SC-Y 编码聚蛋白 ORF1 序列分析

利用 BioEdit、DNASTAR 生物信息学软件,对 SC-Y 与 GenBank 核酸数据库中的 TGEV 参考序列及冠状病毒相关序列进行比较分析,具体有 NC_002306(TGEV PUR46-MAD 株)、Z34093(TGEV Purdue-115 株)、DQ201447(TGEV TS 株)、HCVORFIAB(HcoV-229E 株)、AY700211(MHV A59 株)、NC-001451(IBV Beaudette 株、NC-004718(SARS Tor2 株)。

2 结 果

2.1 RT-PCR 扩增

用设计的 11 对引物对细胞毒 SC-Y 编码复制酶聚蛋白 ORF1 序列进行扩增,结果均获得了与预期相一致的片段,最长目的片段 3 842 bp,最短 733 bp(见图 1)。

2.2 克隆及测序

将获得的 11 个目的片段分别克隆到 pMD18-T 或 pGEM-T 载体,阳性质粒由宝生物公司测序,应用软件 BioEdit 对 11 个片段进行序列拼接,最后获得 SC-Y 株编码复制酶聚蛋白 ORF1 序列,全长共 20 053 bp(GenBank 登录号 DQ390461)。

2.3 SC-Y 编码复制酶聚蛋白 ORF1 序列分子特征

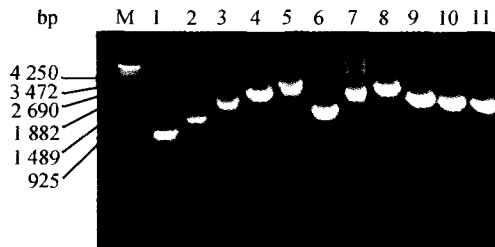
SC-Y 编码复制酶聚蛋白 ORF1 由 20 053 个核苷酸组成,包含 2 个 ORF,其中 ORF1a 由 12 053 个核苷酸构成,编码由 4 018 个氨基酸组成的多肽,ORF1b 由 8 036 个核苷酸构成,编码由 2 678 个氨基酸组成的多肽。这两个阅读框之间相互重叠 30

表 1 RT-PCR 扩增编码复制酶聚蛋白 ORF1 各引物序列及位置
Table 1 Sequence and position of primers in replicase polyprotein of ORF1 for RT-PCR

cDNA	引物 Primers	序列 Sequences	位置/bp Position	扩增产物/bp PCR products
1	F1	GTCGGCCTCTTAATACGACTCACTATAGACTTTTAAAGTAAAGTGAG	1-19	733
	R1	AATCGCCTTCAATGACTGGTTTTC	681-705	
2	F2	TGAGGACTTTGACCTTAAAATTGCT	601-629	1 100
	R2	AACATCAAATGCTTTTACAAGAACT	1 679-1 703	
3	F3	GCCAACATTTGTGGTTCCAGAC	1 622-1 643	2 021
	R3	TGTTGAAGAAATCAAAGGCCTG	3 621-3 642	
4	F4	CAGGCCTTTGATTTCTTCAAC	3 621-3 642	2 580
	R4	CTTGAGATGTTGAAAATCAGC	6 180-6 201	
5	F5	TAAAGTCTGCAGTCTGTGGC	6 115-6 134	3 842
	R5	TAATCCAAGTGAATGGTGTGTAC	9 935-9 957	
6	F6	ACACCATTCCTTGGATTAATCC	9 939-9 961	1 423
	R6	GTAGCCGATGCGGCTGGAATG	11 342-11 362	
7	F7	TCCAGCCGCATCGGCTACAAG	11 345-11 365	2 283
	R7	GTGCAAAGAAGAAGTGTTTAATG	13 605-13 628	
8	F8	AAAACACTTCTTCTTTGCACAGG	13 609-13 631	3 011
	R8	TGTGCCAAGAAGGCCTTGACAAC	16 580-16 620	
9	F9	CAAGGCCTTCTTGGCACATAATC	16 585-16 607	1 921
	R9	TTCTAGACCGGTGAAGGCACC	18 485-18 505	
10	F10	TTTACCGGTTTAAAAGGTGAC	18 491-18 511	1 616
	R10	TATGGTCCTAAGTAGTTAATACC	20 084-20 106	
11	F11	CTCGGACCATACTGTGACAAAGC	20 096-20 106	1 456
	R11	TGTGGTCCATCAGTTACGCCG	21 531-21 551	

F 为正向引物;R 为反向引物

F represent forward primer; R represent reverse primer



M. λ -EcoT14 I digest DNA marker; 1-11. cDNAs displayed in Table 1

图 1 TGEV (ORF1) 的 RT-PCR 扩增

Fig. 1 RT-PCR amplification of TGEV (ORF1)

Leu 变为 Phe,第 268 位由 Leu 变为 His(图 2)。根据同源蛋白序列比较并结合 3CL^{pro}裂解氨基酸基序特征,推测多聚蛋白 ppla/pplab 中有 13 个 3CL^{pro}裂解位点,其中 9 个位于 ORF1a 中,4 个位于 ORF1b 中。

2.4 SC-Y 编码复制酶聚蛋白 ORF1 同源性比较

比较 SC-Y 株与 TGEV PUR46-MAD 株、Purdue-115 株、TS 株 ORF1 的同源性,结果显示,SC-Y 株与 PUR46-MAD 株的同源性最高,ORF1a 的核苷酸同源性为 99.5%,推导氨基酸同源性为 99.2%,ORF1b 的核苷酸及推导氨基酸同源性均为 99.8%,不同冠状病毒之间进行 ORF1a 和 ORF1b 序列比较,结果见表 2,SC-Y ORF1a 与 HcoV-229E ORF1a 同源性最高,为 38.9%,与 MHV、IBV 和 SARS 病毒的 ORF1a 同源性较低,分别为 21.7%、21.0% 和 22.0%。SC-Y ORF1b 与 HcoV-229E

个碱基。SC-Y 与 PUR46-MAD 株和 Purdue-115 株进行序列比较,三者具有相同的“滑脱位点序列”(Slippery site sequence),即 TTTAAC 和“假节”基本结构。SC-Y 株自身编码 3CL 蛋白酶的氨基酸序列中共有 2 个氨基酸发生变异,分别为第 3 位由

SC-Y	1	SGFRKMAQPSGLVEPCIVRVSYGNNVNLGLWLGDEVICPRHVIASDTTRVINYENEMSSV
Purdue-115	1	..L.....
PUR46-MAD	1	..L.....
SC-Y	61	RLHNEFSVSKNNVELGVVSARYKGVNLYLKVNVNPNTPEHKFKSIRKAGESENLACYEFC
Purdue-115	61
PUR46-MAD	61
SC-Y.pro	121	PGSVYGVNMRSQGTIKESFIAGTCGSVGYVLENGILYFVYMHLELGNCSHVGSNFECEM
Purdue-115	121
PUR46-MAD	121
SC-Y	181	YGGYEDQPSMQLGTFNVMSSDNVVAFLYAALINGEKWFVTNTSMSLESYNTWAKTNSFTE
Purdue-115	181
PUR46-MAD	181
SC-Y	241	LSSTDAFSMLRAKTGQSVKLLDSTVIMNKGFCCGRTILSYGSLCDEFTPTFRVIRQMYGVN
Purdue-115	241L.....
PUR46-MAD	241L.....
SC-Y	301	LQ
Purdue-115	301	..
PUR46-MAD	301	..

图 2 SC-Y 与 PUR46-MAD 和 Purdue-115 株的 3CL 蛋白酶同源性比较
 Fig 2 The sequence alignment of the SC-Y 3CL proteinase and 3CL^{pro} of PUR46-MAD and Purdue-115 strains

表 2 冠状病毒之间 ORF1a(右上方)和 ORF1b(左下方)的同源性

Table 2 Identity between ORF1a (top-right) and ORF1b (bottom-left) of Coronavirus %

	TGEV(SC-Y)	HcoV-229E	MHV A59	IBV Beaudette	SARS Tor2
TGEV(SC-Y)		38.9	21.7	21.0	22.0
HcoV-229E	72.0		22.0	20.1	22.1
MHV A59	54.5	52.6		22.0	28.2
IBV Beaudette	52.6	52.6	55.0		21.0
SARS Tor2	55.5	55.2	61.9	55.2	

ORF1b 同源性高达 72.0%，与 MHV、IBV 和 SARS 病毒的 ORF1b 同源性分别为 54.4%、52.6% 和 55.5%。由此可见，不同冠状病毒间 ORF1b 比 ORF1a 具有更高的保守性，说明在病毒进化过程中 ORF1b 的变异程度相对较小。

3 讨论

在 TGEV 基因组转录过程中，首先是以基因组为模板合成复制酶聚蛋白，即 ppla 和 pplab。其中 pplab 融合蛋白的形成是在核糖体移码机制下完成的。该过程至少需要 2 种顺式作用元件，一个是具有三级 RNA 结构的“假节”序列，另一个为核糖体“滑脱序列”^[4]。研究表明^[5]，“滑脱序列”和“假节”结构及二者之间的碱基数目都可影响核糖体移码的有效性。推测 TGEV 的“假节”结构由 3 个茎

(Stem)和 3 个环(Loop)组成，第 1 茎(Stem1)由 14 个碱基组成，位于“滑脱序列”下游 3 个碱基处，而滑脱位点(Slip site)可以通过改变阅读框，导致核糖体滑过 ORF1a 终止子而启动 ORF1b，从而形成融合蛋白。研究结果显示，SC-Y 株与 PUR46-MAD 和 Purdue-115 株的“滑脱位点”和“假节”序列完全一致，故推测 SC-Y 株在该区域可以表达融合蛋白 pplab。

SC-Y 株的 ORF1a 中 PL-1/2 主要参与 ppla 和 pplab 蛋白产物 N 端的裂解，3CL^{pro} 参与 C 端的裂解。序列比较表明，SC-Y 株 PL1 由 3 552—4 133 位核苷酸编码，其中共有 2 个碱基发生了同义密码子的转化，即核苷酸第 3 897 位由 G 变为 A，第 3 910 位由 A 变为 G。在 ppla/pplab 融合蛋白中^[6]，PL1 存在 2 个切割位点，分别为⁸⁷⁹Gly↓Gly⁸⁸⁰和¹¹⁰Gly↓Ala¹¹¹，同样在 SC-Y 对应区存在这 2 个裂解位点。

SC-Y 株 PL2 区域主要由 5 037—5 624 位核苷酸编码,与 PUR46-MAD 和 Purdue-115 株 PL2 的氨基酸同源性均为 99.5%。其中 5 099 位的碱基由 G 突变为 A,致使编码的第 1 700 位氨基酸由 Val 变成 Asp。

在 TGEV_{pp1a} 和 _{pp1ab} 多聚蛋白 C 端的加工过程中,3CL^{pro} 起非常重要的作用。SC-Y 株的该蛋白酶位于 _{pp1a} 的 C 端,在推测的二肽 Q^{2 878}-S^{2 879} 和 Q^{3 180}-A^{3 181} 处通过自身裂解作用而生成,研究表明^[7],冠状病毒 3CL^{pro} 优先裂解 L(V)-Q-S(AN) 氨基酸基序。根据此特征,在 SC-Y 复制酶融合蛋白中共有 13 个 3CL 蛋白酶裂解位点,其中位点 Q^{3 998}-S^{3 999} 和 Q^{4 928}-A^{4 929} 之间的氨基酸编码产物为 RNA 依赖性 RNA 聚合酶,位点 Q^{4 928}-A^{4 929} 和 Q^{5 527}-A^{5 528} 之间的产物包括金属离子结合区和解旋酶。它们在病毒的生命周期中发挥着许多重要的作用。

SC-Y3CL^{pro} 两侧为疏水性氨基酸集中区,分别位于氨基酸 2 640-2 960 位和 3 184-3 438 位。TGEV 的 3CL^{pro} 由 2 个不同的折叠组成,包括结构域 I 和 II 的 N 端胰凝乳蛋白酶折叠,和 C 端额外的螺旋区域 III。其中 N 端的胰凝乳蛋白酶样的 2 个 β 折叠,即结构域 I (8-100) 和 II (101-183),通过 1 个 16 氨基酸残基的环(184-199)与由 5 个 α 螺旋组成的结构域 III (200-302) 连接,N 端残基(1-5)组成的多肽折叠在结构域 III 的上面。研究表明 N 端残基(1-5)在保持结构域 II、III 及 16 个氨基酸环正确排列方位上具有重要作用,且所有冠状病毒 3CL^{pro} 在第 3 位氨基酸都呈疏水性,与结构域 III 上的 Phe(206)、Ala(209)、Phe(287)、Val(292) 和 Met(296) 共同组成一段疏水区,缺失 N 端残基(1-5)的 3CL^{pro} 蛋白水解活性将有所下降^[7]。由此推测 N 端(1-5)的残基组成及构型可能对 3CL^{pro} 水解活性产生影响,SC-Y 与 PUR46-MAD 和 Purdue-115 株的 3CL^{pro} 序列比较,结果显示有 2 个氨基酸发生了变异,一个为第 3 位氨基酸由 Leu 变为 Phe,与 IBV 和 SARS 病毒 3CL^{pro} 相应位置残基相同,另一个为第 268 位氨基酸由 Leu 变为 His。其中第 3 位氨基酸 Phe 具有疏水性,但这一变异氨基酸能否影响 TGEV 3CL^{pro} 水解活性还难以预测。

比较 SC-Y、PUR46-MAD、Purdue-115、TS、HCV229E、IBV、MHV、SARS 毒株之间 ORF1a 和 ORF1b 同源关系,发现在不同冠状病毒之间 ORF1b 的同源性要高于 ORF1a。由此推断 ORF1a 比 ORF1b 更易发生变异。分析 SC-Y 株与 PUR46-

MAD 株序列,共有 36 个氨基酸发生变异,其中 31 个发生在 ORF1a 中,其余 5 个发生在 ORF1b 中。而在变异的碱基中,主要是 T-C 和 A-G 之间的变化,这一变化规律有何遗传学意义,值得探讨。

目前,国外学者已从猪传染性胃肠炎病毒基因组全长 cDNA 拯救了具有感染性的病毒粒子^[8,9],然后在体外对病毒基因组进行人工操作,研究病毒复制、致病机制及病毒与宿主细胞之间的关系。本文对 TGEV SC-Y 株基因组 5' 端 20 053 bp 的编码复制酶聚蛋白 ORF1 进行了克隆和测序,这为今后构建基因组全长 cDNA 提供了材料,也为进一步研究病毒的基因功能打下了良好基础。

参考文献:

- [1] Annette H, John Z. Conservation of substrate specificities among coronavirus main proteases [J]. *J Gen Vir*, 2002, 83: 595~599.
- [2] Ziebuhr J, Snijder E J, Gorbalenya A E. Virus-encoded proteinases and proteolytic processing in the *Nidovirales* [J]. *J Gen Virol*, 2000, 81: 853~879.
- [3] 萨姆布鲁克 J, 弗里奇 E F, 曼尼阿蒂斯 T. 分子克隆实验指南 [M]. 第 2 版. 北京: 科学出版社, 1998.
- [4] Eleouet J, Rasschaert D, Lambert P, *et al*. Complete sequence (20 kilobases) of the polyprotein-encoding gene 1 of transmissible gastroenteritis virus [J]. *Virology*, 1995, 206: 817~822.
- [5] Sawsan N, Jan L, Alison B, *et al*. The role of RNA pseudoknot stem 1 length in the promotion of efficient-1 ribosomal frameshifting [J]. *J Mol Biol*, 1999, 288: 305~320.
- [6] Putics A, Gorbalenya A E, Ziebuhr J. Identification of protease and ADP-ribose 1"-monophosphatase activities associated with transmissible gastroenteritis virus non-structural protein 3 [J]. *J Gen Vir*, 2006, 87: 651~656.
- [7] Kanchan A, Gottfried J P, Jeroen R M, *et al*. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α -helical domain [J]. *EMBO J*, 2002, 21: 3 213~3 224.
- [8] Yount B, Curtis K M, Baric R S. Strategy for systematic assembly of large RNA and DNA genome-s: Transmissible gastroenteritis virus model [J]. *J Vir*, 2000, 74: 10 600~10 611.
- [9] Almazan F, Gonzalez J M, Penzes Z, *et al*. Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome [J]. *Proc Natl Acad Sci U S A*, 2000, 97(10): 5 516~5 521.