

口蹄疫病毒全基因组序列特征分析

杨鹏华, 倪凤娥

(长江大学 生命科学学院 长江中游湿地农业教育部工程研究中心, 湖北 荆州 434025)

摘要: 对口蹄疫病毒全基因组序列特征的分析有助于了解口蹄疫病毒复制、翻译等生命活动的相关机制。利用 DNASTar 和 DNAMAN 软件, 对 249 株口蹄疫病毒全基因组序列进行了比对分析。结果表明, 口蹄疫病毒 ORF 的长度为 5 982~7 020 bp, 平均长度为 6 975 bp, 编码 1 994~2 340 个氨基酸。各基因的核酸序列同源性均在 74% 以上, 最高达到 94.29%。在 5'-UTR 存在大量保守位点, 并发现 3 个长的保守序列。3'-UTR 区存在 4 段保守序列, 可以形成 2 个发卡结构。研究结果为进一步研究口蹄疫病毒基因组序列结构与病毒生存机制的关系提供了重要的生物信息, 为病毒防治提供了理论依据。

关键词: 口蹄疫病毒; 序列比对; 基因组

中图分类号: S852.65⁺9.6 文献标识码: A 文章编号: 1004-3268(2011)11-0140-05

Sequence Analysis of the Foot and Mouth Disease Virus Genome

YANG Peng-hua, NI Feng-e

(Engineering Research Center of Wetland Agriculture in the Middle Reaches of the Yangtze River,
Ministry of Education, College of Life Science, Yangtze University, Jingzhou 434025, China)

Abstract: There was no much information about the life circle of foot and mouth disease virus (FMDV), such as replication, translation as well as assembly of virus. Analysis of the FMDV genome may provide a chance to discover the mystery to some extent. Presently, the genomes of 249 FMDV isolates were aligned by the biological software, DNASTar and DNAMAN. The lengths of the open reading frame (ORF) in all isolates were between 5 982 to 7 020 bp, which code 1 994 to 2 340 amino acids. The homology of all genes of FMDV was above 74% and up to 94.29%. Three long conserved sequences and four conserved motifs were obtained in the 5'-UTRs and 3'-UTRs of FMDV, respectively. The result provided a new basis for understanding the relationship between the genome sequence and life mechanism of FMDV.

Key words: Foot and mouth disease virus; Sequence alignment; Genome

口蹄疫是一种由口蹄疫病毒引起的急性、热性和高度接触性人畜共患传染病, 主要感染各种偶蹄动物。口蹄疫病毒是正链单股 RNA 病毒, 属于小 RNA 病毒科口蹄疫病毒属, 共有 7 个血清型^[1]。基因组包括 2 个非翻译区 (UTR) 和 1 个开放阅读框 (ORF)。5'-UTR 包括 S 片段 (S fragment)、多聚 C (PolyC)、假结节 (PKs) 和内部核糖体进入位点 (IRES), 3'-端包括 3'-UTR 和多聚腺苷酸尾

(PolyA)^[2]。ORF 编码 1 个大的多聚蛋白, 翻译为 1 个前体蛋白, 最后水解为 12 种成熟蛋白, 分别为 L、1A (Vp4)、1B (Vp2)、1C (Vp3)、1D (Vp1)、2A、2B、2C、3A、3B (Vpg)、3C (Pro) 和 3D (Pol)^[3]。虽然对口蹄疫病毒的基因组已经有很多报道, 但是依然有很多区域的功能还不完全清楚^[4]。例如, 5'-UTR 中的 S 片段为什么是所有小 RNA 病毒科中最长的, 它具体参与 RNA 复制

收稿日期: 2011-06-21

基金项目: 转基因生物新品种培育科技重大专项 (2011ZX08009-003-006)

作者简介: 杨鹏华 (1977-), 男, 河北保定人, 讲师, 博士, 主要从事转基因动物研究。E-mail: yphing@126.com

的机制也不清楚^[5-6]。而且人们对 2B 和 3A 的具体作用还不了解^[7]。虽然对口蹄疫病毒的复制过程进行了大量研究,但是仍然不清楚病毒确切的复制机制^[8-9]。分析全基因组序列特征可以帮助人们了解口蹄疫病毒复制、翻译等生命活动的相关机制。鉴此,对现有的 249 株口蹄疫病毒基因组进行了序列比对和特征分析,以期对口蹄疫病毒基因组序列与病毒复制、翻译和组装释放机制之间的关系有一个新的认识。

表 1 参考的口蹄疫病毒株基本情况

血清型	数量/个	登录号
A	56	AY593751—AY593794,AY593801—AY593803,EF117837,HM854021—HM854025,HQ632773,NC_011450,X74812
Asia1	40	AY304994,AY317098,AY390432,AY593795—AY593800,AY687333—AY687334,DQ533483,DQ989303—DQ989323,EF149009,EF149010,EF614458,FJ906802,GU931682,HQ632774,NC_004915
C	20	AF274010,AM409325,AY593804—AY593810,DQ409183—DQ409191,FJ824812,NC_002554
O	112	AB079061,AF026168,AF308157,AF377945,AF506822,AF511039,AH012984,AH012985,AJ320488,AJ539136—AJ539141,AJ633821,AY333431,AY359854,AY593811—AY593837,AY686687,DQ404158—DQ404180,DQ478936,DQ478937,EF175732,EF552688—EF552697,EF614457,EU214601,FJ175661—FJ175666,FJ542365—FJ542372,GU38482,GU384683,HM008917,HM229661,HQ009509,HQ412603,HQ632768—HQ632772,NC_004004,X00871
SAT1	10	AY593838—AY593846,NC_011451
SAT2	6	AF540910,AJ251473,AY5913847—AY593849,NC_003992
SAT3	5	AY593850—AY593853,NC_011452

1.2 口蹄疫病毒基因组序列比较

利用 DNASTar 软件对所有口蹄疫病毒全基因组序列进行比对,并拆分基因组为 5'-UTR 序列、单个结构基因、非结构基因及 3'-UTR 序列。用 DNAMAN 软件对每一个基因进行比对,分析保守位点。

2 结果与分析

2.1 口蹄疫病毒全基因组同源性分析

用 DNAMAN 软件对 249 个口蹄疫病毒全基因组序列进行同源性分析,结果表明,各基因的核酸序列同源性均在 74%以上,最高达到 94.29%,具体见表 2。其中 5'-UTR 和 3'-UTR 的核酸序列同源性分别是 78.48%和 82.76%。结构蛋白基因(1A、1B、1C 和 1D)的核酸序列和氨基酸序列同源性分别为 81.58%和 87.06%,非结构蛋白基因(除结构蛋白之外的其他蛋白)的核酸序列和氨基酸序列同源性分别为 91.89%和 95.55%。完整 ORF 的核酸序列和氨基酸序列同源性分别为 87.88%和 92.53%。

2.2 口蹄疫病毒 5'-UTR 区特征分析

对口蹄疫病毒基因组 5'-UTR 区进行比对,结果表明,此区同源性比较低,存在大量的突变位点,其中同源性最低的区域在 Poly(C)区,这一区域的 C 的数量变化很大,在图 1 中表示为一个大的凹陷(图 1)。另外,5'-UTR 的 3'端序列的保守性高于其

1 材料和方法

1.1 口蹄疫病毒基因组序列

口蹄疫病毒基因组序列收集自 GenBank 数据库中的已发表序列,共 249 个全基因组序列,具体见表 1。其中血清 A 型 56 个,亚洲 1 型 40 个,C 型 20 个,O 型 112 个,SAT1 型 10 个,SAT2 型 6 个,SAT3 型 5 个。

5'端序列,并且发现大量非常保守的基序,这些基序主要集中在 IRES 区,与口蹄疫病毒的复制和翻译密切相关。用 BioEdit 软件对 5'-UTR 的 3'端序列进行一致性序列分析,结果发现了 3 个长的保守性非常高的序列,分别为 1、2 和 3(图 1)。其中保守序列 1 位于PKs 和 IRES 之间,保守序列 2 和 3 位于 IRES 内部,与蛋白质的翻译相关。

表 2 口蹄疫病毒基因组序列同源性分析

基因	同源性/%	
	核酸	氨基酸
5'-UTR	78.48	
3'-UTR	82.76	
L	85.66	88.61
1A	89.75	98.50
1B	81.62	87.87
1C	80.93	85.27
1D	74.01	76.59
2A	92.03	97.30
2B	93.21	96.88
2C	93.78	97.89
3A	91.08	93.65
3B	92.29	95.06
3C	92.81	97.46
3D	94.29	97.58
ORF	87.88	92.53

2.3 口蹄疫病毒 ORF 区特征分析

口蹄疫病毒 ORF 区的长度在 5982~7020 bp,平均长度为 6975 bp,编码 1994~2340 个氨基酸。通过分析发现,大部分病毒的 ORF 在 6960 bp 以

上,但是有 6 株病毒的 ORF 比较短,最短的只有 5 982 bp,为 DQ409184。进一步研究表明,这 6 株病毒缺失了大量基因序列,分别是 L、1BC 和 1CD(图 2)。口蹄疫病毒 ORF 区主要编码 12 种蛋白,包括

4 种结构蛋白和 8 种非结构蛋白,非结构蛋白的保守性明显大于结构蛋白。保守性较高的区域集中在 3 个部分,分别是 1A 和 1B 的 5' 序列、2ABC 和 3A 的 5' 端序列、3BCD(图 2 I, II, III)。

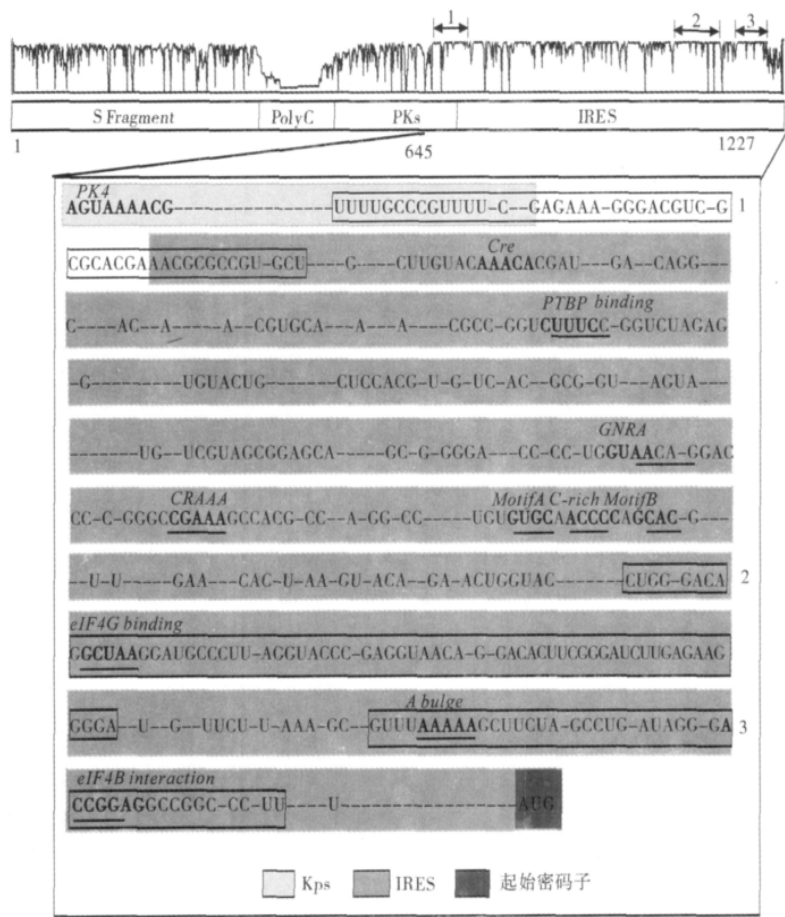


图 1 口蹄疫病毒 5'-UTR 区序列分析

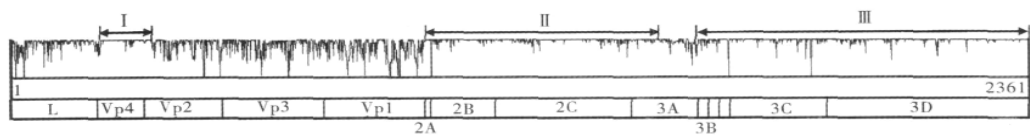


图 2 口蹄疫病毒 ORF 区蛋白序列分析

2.4 口蹄疫病毒 3'-UTR 区特征分析

口蹄疫病毒 3'-UTR 有约 90 个核苷酸,同源性为 82.76%,含有一些保守性非常强的序列,这些序列可以分为 4 部分,分别为 I、II、III 和 IV(图 3)。其中 I 和 II 区域能形成反向互补的茎,III 和 IV 能形成反向互补的茎,在 I 和 II 之间,III 和 IV 之间的序列形成环。这些保守序列与口蹄疫病毒的复制密切相关。

3 讨论

口蹄疫病毒基因组全长 8 083~8 287 bp,与文

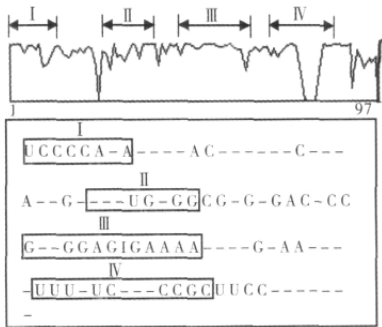


图 3 口蹄疫病毒 3'-UTR 区序列分析

献报道基本一致^[10]。但是本研究中发现,个别毒株的基因组序列发生了大规模丢失现象,最高的缺失了约 1000 bp,这样大量的核苷酸序列的减少并没有影响病毒在细胞上的复制。因为缺失只发生在了结构蛋白和 L 区,并没有影响有关复制的相关基因,并且随着病毒代数的增加,缺失的基因通过遗传重组又得到了修复^[11]。氨基酸序列同源性最高的是 1A 序列,同源性为 98.50%,同源性最低的是 1D,只有 76.59%。核酸序列同源性最高的是 3D,最低的是 1D。这可能是由于 1A 和 3D 蛋白主要参与病毒粒子的组装和复制,而病毒粒子的组装和复制是在宿主细胞的细胞质中发生的,对病毒的增殖至关重要,所以氨基酸序列的高度保守性保证了功能的完整性^[12]。而 1D 基因是最主要的抗原分子,是病毒入侵宿主细胞、刺激宿主免疫反应的主要因子,为了逃避宿主的免疫反应,1D 基因经常发生变异^[13]。非结构蛋白的保守性高于结构蛋白,因为非结构蛋白主要参与病毒的复制、前体蛋白的加工、病毒粒子的组装与释放,这些过程任何一步发生异常都将导致病毒的消亡,所以非结构蛋白的保守性较大^[2]。

各蛋白中同源性较低的是 L 和 3A 蛋白,L 蛋白 N 端和 3A 蛋白 C 端氨基酸序列的高变异率是导致它们同源性较低的原因^[14]。L 基因 2 个 AUG 之间序列的高度变异可能也导致了病毒更倾向于从第 2 个 AUG 位点起始翻译^[15-16]。而 3A 蛋白与病毒的毒力和感染宿主范围有关^[10]。同源性低的序列主要在结构蛋白 1BCD 区,因为这 3 个基因编码的蛋白构成衣壳蛋白,是口蹄疫病毒与宿主细胞直接接触的部分,所以面临着巨大的选择压力,这导致了这 3 个基因的变异程度非常高^[17]。其他的非结构蛋白和 1A 共同组成了相对稳定的区域,保守性相对较高^[18]。

5'-UTR 5'端保守性非常高,在所选择的序列中除了 O/YM/YN/2000 株(序列号 HQ412603)外全都是 UUGAAA,5'端的 UU 对口蹄疫病毒复制起始至关重要,Vpg 蛋白也是通过酪蛋白的羟基与 5'端的 UU 相连接^[5,19]。经过分析,发现了几个保守性非常高的序列,有 121CCUG、156GCUGG、224CCGCC、253UGGGC、333UACUUGU、285UUUCA,其中 285UUUCA 序列与 5'端 UUGAAA 互补形成 S 片段长茎的根部,对维持 S 片段的稳定性非常重要。紧临 285UUUCA 序列就是 Poly(C),Poly(C)中 C 的数量在各毒株中变化较大,从 50 bp 至 250 bp 不等,而且在培养或增殖过程中还可以延长,可能与感染力或毒力有关,但

是有研究表明,只有 2 个 C 的毒株也具有感染性^[20]。PKs 中最后一个 PK 的保守性最高,保守序列为 AGUAAAA。分析发现,在 PKs 和 Cre 之间的一段较长序列相对保守,可能与维持 PKs 和 Cre 2 个茎结构的稳定有关。5'-UTR 区最保守的序列大多集中在 IRES 区,如 Cre 位点(AAACA)、PTBP 结合位点(UCUUUC)、GNRA、CRAAA、C-rich 位点(ACCC)、eIF4G 结合位点(GCUAA)、A bulge(AAAAA)、eIF4B 结合位点(ACCGGAG)等。另外,还发现了大量新的保守性较高的序列,在这些序列中大部分组成了 IRES 区,可能参与蛋白质翻译调控,但具体功能未知,有待进一步研究。

3'-UTR 区由 2 个茎环结构构成,第 1 个茎比较短,环比较大,第 2 个茎比较长,环比较小^[21]。在这一区域同样存在保守序列,如 UCCCCA、UGNGG、GGAGUGAAAA、UUNUC。前 2 个序列和后 2 个序列分别是反向互补序列,能形成 2 个茎结构,可能与病毒的复制相关^[22]。

通过对 249 株口蹄疫病毒的基因组进行序列比对,并分析它的基因组结构特征,发现在大量的保守序列中,存在一些新的功能未知的高度保守序列,这些序列在病毒生命活动中是否起到了某些关键性的作用,也许是防制病毒的潜在的靶位点。本研究结果为利用针对高度保守序列设计的小干扰 RNA(siRNA)来研究抗口蹄疫病毒奠定了基础,对口蹄疫病毒疫苗的研制、抗病毒药物的研发以及抗病转基因动物的研究提供了重要的生物信息。

参考文献:

- [1] 刘新生,王永录.口蹄疫病毒株 AF/72 结构蛋白 VP1 克隆及其 T 细胞表位预测[J].华北农学报,2010,35(5):25-32.
- [2] Mason P W,Grubman M J,Baxt B. Molecular basis of pathogenesis of FMDV[J]. Virus Res,2003,91:9-32.
- [3] Domingo E,Baranowski E,Escarmis C,et al. Foot-and-mouth disease virus[J]. Comp Immunol Microbiol Infect Dis,2002,25:297-308.
- [4] 张显,王永录,张永光,等.口蹄疫病毒结构蛋白 VP1 上 B 细胞表位的筛选鉴定[J].华北农学报,2009,34(5):81-85.
- [5] Lyons T,Murray K E,Roberts A W,et al. Poliovirus 5'-terminal cloverleaf RNA is required in cis for VPg uridylation and the initiation of negative-strand RNA synthesis[J]. J Virol,2001,75:10696-10708.
- [6] Barton D J,O'Donnell B J,Flanagan J B. 5' cloverleaf in poliovirus RNA is a cis-acting replication element required for negative-strand synthesis [J]. EMBO J, 1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

- 2001,20:1439-1448.
- [7] Mason P W, Pacheco J M, Zhao Q Z, *et al.* Comparisons of the complete genomes of Asian, African and European isolates of a recent foot-and-mouth disease virus type O pandemic strain (PanAsia) [J]. J Gen Virol, 2003, 84: 1583-1593.
- [8] Murray K E, Steil B P, Roberts A W, *et al.* Replication of poliovirus RNA with complete internal ribosome entry site deletions[J]. J Virol, 2004, 78: 1393-1402.
- [9] Steil B P, Barton D J. Cis-active RNA elements(CREs) and picornavirus RNA replication[J]. Virus Res, 2009, 139: 240-252.
- [10] Carrillo C, Tulman E R, Delhon G, *et al.* Comparative genomics of foot-and-mouth disease virus[J]. J Virol, 2005, 79: 6487-6504.
- [11] Garcia-Arriaza J, Ojosnegros S, Davila M, *et al.* Dynamics of mutation and recombination in a replicating population of complementing, defective viral genomes [J]. J Mol Biol, 2006, 360: 558-572.
- [12] George M, Venkataramanan R, Pattnaik B, *et al.* Sequence analysis of the RNA polymerase gene of foot-and-mouth disease virus serotype Asia1 [J]. Virus Genes, 2001, 22: 21-26.
- [13] Berryman S, Clark S, Monaghan P, *et al.* Early events in integrin alphavbeta6-mediated cell entry of foot-and-mouth disease virus[J]. J Virol, 2005, 79: 8519-8534.
- [14] Thomas A A, Rijnbrand R, Voorma H O. Recognition of the initiation codon for protein synthesis in foot-and-mouth disease virus RNA[J]. J Gen Virol, 1996, 77: 265-272.
- [15] Belsham G J. Dual initiation sites of protein synthesis on foot-and-mouth disease virus RNA are selected following internal entry and scanning of ribosomes *in vivo* [J]. EMBO J, 1992, 11: 1105-1110.
- [16] Cao X, Bergmann I E, Fullkrug R, *et al.* Functional analysis of the two alternative translation initiation sites of foot-and-mouth disease virus[J]. J Virol, 1995, 69: 560-563.
- [17] Domingo E, Escarmis C, Baranowski E, *et al.* Evolution of foot-and-mouth disease virus[J]. Virus Res, 2003, 91: 47-63.
- [18] Lewis-Rogers N, McClellan D A, Crandall K A. The evolution of foot-and-mouth disease virus: impacts of recombination and selection[J]. Infect Genet Evol, 2008, 8: 786-798.
- [19] Liu Y, Franco D, Paul A V, *et al.* Tyrosine 3 of poliovirus terminal peptide VPg(3B) has an essential function in RNA replication in the context of its precursor protein, 3AB[J]. J Virol, 2007, 81: 5669-5684.
- [20] Rieder E, Bunch T, Brown F, *et al.* Genetically engineered foot-and-mouth disease viruses with poly(C) tracts of two nucleotides are virulent in mice[J]. J Virol, 1993, 67: 5139-5145.
- [21] Saiz M, Gomez S, Martinez-Salas E, *et al.* Deletion or substitution of the aphthovirus 3' NCR abrogates infectivity and virus replication[J]. J Gen Virol, 2001, 82: 93-101.
- [22] Sharma N, O'Donnell B J, Flanagan J B. 3'-Terminal sequence in poliovirus negative-strand templates is the primary cis-acting element required for VPgUpU-primed positive-strand initiation[J]. J Virol, 2005, 79: 3565-3577.
- [23] Rodriguez Pulido M, Sobrino F, Borrego B. Attenuated foot-and-mouth disease virus RNA carrying a deletion in the 3' noncoding region can elicit immunity in swine[J]. J Virol, 2009, 83: 3475-3485.