

4 株 SARS 冠状病毒基因组 5' 端序列的分析与比较

常国辉, 彭文明, 刘伯华, 范宝昌, 刘洪, 邓永强,

吕富双, 杨保安, 秦鄂德, 祝庆余**

(军事医学科学院微生物流行病学研究所, 北京, 100071)

Comparative Analysis of Genomic Organization of the 5' and of Four Strain SARS-associated Coronaviruses

CHANG Guo-hui, PENG Wen-ming, LIU Bo-hua, FAN Bao-chang, LIU Hong, DENG Yong-qiang,
LU Fu-shuang, YANG Bao-an, QIN E-de, ZHU Qing-yu**

(Institute of Microbiology and Epidemiology, Academy of Military Medical Sciences, Beijing 100071, China)

Abstract: The viral RNAs were isolated from the infected Vero-E6 cells, then the templates were produced by reverse transcription reaction. Two pairs of primers were used to amplify the 5'-ends of four strains of SARS-associated coronavirus (SARS-CoV) by RACE, cDNA fragments with the length of about 420 bp were amplified. The fragments were then purified and sequenced. The sequences of 5'-UTRs of SARS-CoV isolated in China were the same as those isolated in other countries and regions, such as Tor2 strain, Urbani strain, HKU-Su10 strain, CUHK-W1 strain, SIN2500 strain and SIN2677 strain. The secondary structures formed by 5'-UTRs of all known SARS-CoV were almost identical, but significantly different from other known non-SARS coronaviruses. Sequence analysis indicated that the conserved core sequence (5'-CUAAAC-3') of transcription regulating sequence in some coronaviruses was about -197 nt upstream from the start codon.

Key words: SARS-associated coronavirus (SARS-CoV); 5'-ends of genomic organization; Sequence analysis; Transcription regulating sequence

摘要: 利用 5' RACE 试剂盒对从中国不同地区、不同 SARS 患者体中分离的 SARS-CoV 基因组 5' 端序列进行 RT-PCR 扩增, 并将扩增产物克隆至 T easy vector。扩增片段的序列测定结果表明: 所分离的 4 株 SARS-CoV 基因组 5' 端非编码区的核苷酸序列和其他国家和地区报道的序列基本一致, 而且所形成二级结构也完全相同, 但与已知普通冠状病毒的差别较大。同时发现在依赖于 RNA 的 RNA 聚合酶起始密码子上游 -197 nt 处有冠状病毒典型的转录调控核心保守序列 5'-CUAAAC-3'。

关键词: SARS 冠状病毒 (SARS-CoV); 基因组 5' 端结构; 序列分析; 转录调控序列

中图分类号: R511; R373

文献标识码: A

文章编号: 1003-5125 (2003) 04-0330-05

我国广东省自 2002 年 10 月发现非典型肺炎以来, 目前该传染病已波及到世界许多国家和地区, 病例多达 8000 多例, 病死率高达 6% 以上。经过国际上多个实验室的通力研究, 已经确证引起该传染

病流行的病原体为一种新型的冠状病毒, WHO 将其命名为重症急性呼吸综合征冠状病毒 (Severe acute respiratory syndrome-associated coronavirus, SARS-CoV)^[1,2]。我们实验室继从广州和北京地区

收稿日期: 2003-05-19, 修回日期: 2003-07-05

* 基金项目: 国家 863 计划项目基金资助 (2003AA208201)

作者简介: 常国辉 (1969-), 男, 汉族, 河南籍, 博士, 从事分子病毒学的研究。

** 通讯作者。Correspondence author. Tel: 010-66948624; Fax: 010-63802193; E-mail: ZhuQY@nic.bmi.ac.cn

的 SARS 患者体内分离到 4 株 SARS-CoV, 在完成了初步的鉴定之后^[3], 又对这 4 株病毒基因组 5' 端非编码区的序列进行了测定和分析, 并与其他 SARS-CoV 和冠状病毒进行了比较。

1 材料与方 法

1.1 材料

4 株 SARS-CoV (GZ01 株、BJ01 株、BJ02 株、BJ03 株) 分别从广州和北京 SARS 患者尸检组织和鼻咽拭子中分离得到, 详见文献^[3,4]。RNeasy Mini Kit、PCR Purification Kit 和 Gel Extraction Kit 均为 Qiagen 公司产品。5' RACE Kit 为 Roche 公司产品, T easy Vector Kit、限制性内切酶均为 Promega 公司产品。

1.2 引物设计与合成

利用 DNASTar 软件分别设计引物 51L: 5' -AGGCAGGACATGGCATTTCCTACTA-3'、52L: 5' -TGACCACCGGCTCCCTTATTACC-3' 和 53L: 5' -CTAGACCACAAGTGCCATTTTT-3'。引物由三博远志生物公司合成。

1.3 病毒核酸的提取和 RT-PCR 反应

分别取 4 株 SARS-CoV 感染的 Vero-E6 细胞, 待细胞病变后, 吸取细胞及其培养上清混合液 210 μ L, 参照试剂盒的说明书提取总 RNA。分别取 5 μ L 总 RNA, 组成 20 μ L 的反转录反应体系, 55 $^{\circ}$ C 反应 70min, 然后 65 $^{\circ}$ C 变性 10min。纯化后的 cDNA 在末端转移酶的作用下 37 $^{\circ}$ C 反应 30 min。分别取 5 μ L 加尾 cDNA, 利用 oligo dT-anchor primer 和引物 52L 进行第一次 PCR 扩增, 反应条件为: 94 $^{\circ}$ C 预变性 3min, 然后 94 $^{\circ}$ C 变性 15s, 54 $^{\circ}$ C 退火 30s, 72 $^{\circ}$ C 延伸 40s, 进行 10 个循环; 再于 94 $^{\circ}$ C 变性 15s, 55 $^{\circ}$ C 退火 30s, 72 $^{\circ}$ C 延伸 40s, 进行 4 个循环; 94 $^{\circ}$ C 变性 15s, 55 $^{\circ}$ C 退火 30s, 72 $^{\circ}$ C 延伸 60s, 进行 5 个循环; 94 $^{\circ}$ C 变性 15s, 55 $^{\circ}$ C 退火 30s, 72 $^{\circ}$ C 延伸 120s, 进行 6 个循环; 94 $^{\circ}$ C 变性 15s, 55 $^{\circ}$ C 退火 30s, 72 $^{\circ}$ C 延伸 140s, 进行 7 个循环; 最后于 72 $^{\circ}$ C 延伸 10 min, 4 $^{\circ}$ C 暂时保存。分别取第一次 PCR 扩增产物 1 μ L, 利用 anchor primer 和引物 53L 进行第二次 PCR 扩增, 反应条件除退火温度改为 61 $^{\circ}$ C 外, 其他条件和第一次 PCR 反应相同。

1.4 PCR 扩增片段的克隆和序列测定

按常规方法进行 PCR 扩增产物的分离纯化和连接克隆。序列测定应用 ABI PRISMTM3100 Genetic Analyzer, 反应试剂用 BigDyeTM Terminator Cycle Sequencing v2.0

1.5 核苷酸序列分析及比较

通过 DNASTar 等生物软件对 4 株病毒 5' 端 430 nt 左右的核苷酸序列进行分析, 与其他冠状病毒进行比较, 并利用 GeneQuest 软件, 结合计算机辅助分析系统, 绘制并分析非编码区二级结构。

2 结果

2.1 4 株 SARS-CoV 5' 末端序列测定与分析

由于 SARS-CoV 基因组的 5' 末端具有帽子结构以及非编码区常形成二级结构, 在 RT-PCR 过程中需要添加 cDNA 3' 末端的 PolyA, 使基因组 5' 端序列的测定更为复杂。通过 2 次 PCR 分别获得 4 株 SARS CoV 基因组的 5' 端 430 nt 左右核酸片段, 测序结果参见图 1 (270nt 后序列未显示 t)。分离的 4 株 SARS-CoV 基因组 5' 端非编码区除最前端的几个碱基不同外, 几乎具有一致的核苷酸一级结构, 说明该 4 株 SARS-CoV 基因组 5' 端非编码区比较稳定, 其长度大约为 270 nt 左右。在距离依赖于 RNA 的 RNA 多聚酶起始密码子上游 -197 nt 具有冠状病毒典型的转录调控核心保守序列 5' -CUAAAC-3'。

2.2 SARS-CoV 基因组 5' 端非编码区序列的比较分析

将分离的 4 株 SARS-CoV 基因组 5' 端非编码区序列与加拿大、美国、新加坡和我国香港等国家和地区所报道的 SARS-CoV 基因组序列进行比较, 结果如图 1。分离的 4 株 SARS-CoV 基因组 5' 端核苷酸一级结构与美国的 Urbani 株、加拿大的 TOR2 株和香港的 HKU-W1、CUHK-Su10 株以及新加坡的 SIN2500 株、SIN2677 株、SIN2679 株、SIN2774 株都高度一致, 而且 5' 端转录调控序列 (transcription regulating sequence, TRS) 均在起始密码子上游 -197 nt, 距离 5' 末端约 70 nt, 差别仅存在于最前端约 17 nt 的部分缺失。

与其他普通冠状病毒的 5' 端非编码区序列比较结果表明: SARS-CoV 基因组的一级结构与人冠状病毒 (Human coronavirus)、牛冠状病毒 (Bovine coronavirus)、鸟类传染性支气管炎病毒 (Avian infectious bronchitis virus)、猪传染性腹泻病毒 (Porcine epidemic diarrhea coronavirus)、鼠肝炎病毒 (Murine hepatitis virus) 和猪传染性肠胃炎病毒 (Porcine transmissible gastroenteritis virus) 存在显著的差别。同时各株冠状病毒的 TRS 核心保守序列相对位置统计结果显示见表 1, SARS-CoV 基因组的核 心保守序列相对位置与普通冠状病毒相比有一定差别。

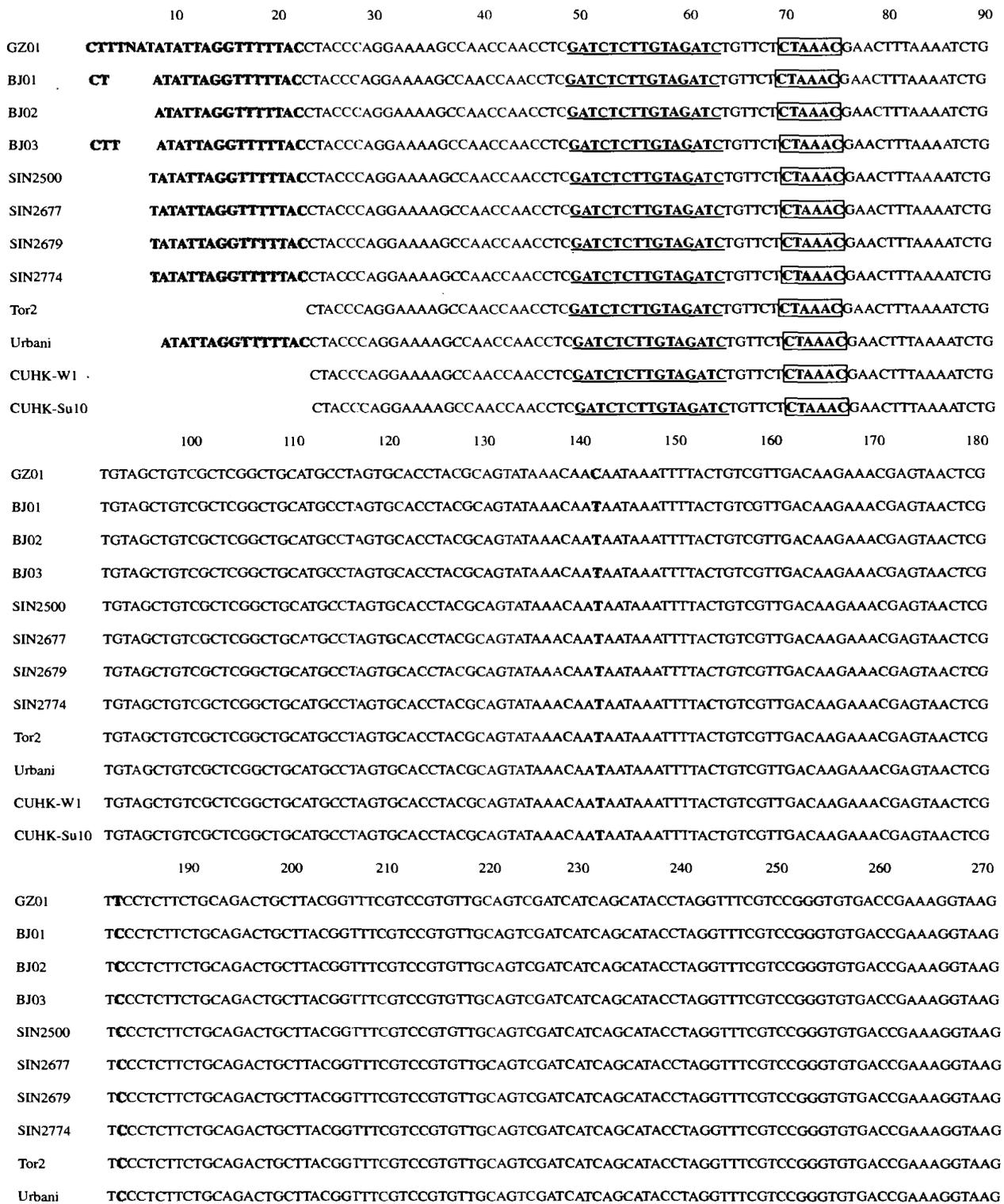


图 1 各株 SARS-CoV 基因组 5' 端非编码区序列比较

阴影表示碱基不同, 方框表示转录调控核心保守序列, 下划线表示保守序列上游的反向互补序列。

Fig.1 Comparison of 5' end UTR of SARS-CoV

Different bases are shaded; conserved core sequence of transcription regulating sequence are boxed; reverse complement sequence upstream of the conserved core sequence of transcription regulating sequence are underlined.

表 1 冠状病毒基因组 5' 端 TRS 核心保守区序列相对位置及反向互补序列碱基组成

Table 1 The position of TRS conserved core sequence and base composition of the reverse complement sequence upstream of the conserved core sequence in 5' end of coronavirus

Strain	The position of TRS conserved core sequence	Base composition of the reverse complement sequence upstream of the conserved core sequence
GZ01	- 197	5' -GAUCUCUUGUAGAUC-3'
BJ01	- 197	5' -GAUCUCUUGUAGAUC-3'
BJ02	- 197	5' -GAUCUCUUGUAGAUC-3'
BJ03	- 197	5' -GAUCUCUUGUAGAUC-3'
SIN2500	- 197	5' -GAUCUCUUGUAGAUC-3'
SIN2677	- 197	5' -GAUCUCUUGUAGAUC-3'
SIN2679	- 197	5' -GAUCUCUUGUAGAUC-3'
SIN2774	- 197	5' -GAUCUCUUGUAGAUC-3'
CUHK-W1	- 197	5' -GAUCUCUUGUAGAUC-3'
CUHK-Su10	- 197	5' -GAUCUCUUGUAGAUC-3'
Tor2	- 197	5' -GAUCUCUUGUAGAUC-3'
Urbani	- 197	5' -GAUCUCUUGUAGAUC-3'
HCV 229E	- 220	5' -UAGACUUUGUGUCUA-3'
BCV Mebus	- 141	5' -GAUCUCUUGUAGAUC-3'
PEDV CV-777	- 227	5' -UAGACCUUGUCUA-3'
MHV A-59	- 139	5' -AAACUCUUGUAGUUU-3'
PTGV Purdue	- 216	5' -CUAGCCUUGUGCUAG-3'

HCV, Human coronavirus; BCV, Bovine coronavirus; PEDV, Porcine epidemic diarrhea coronavirus; MHV: Murine hepatitis virus; PTGV, Porcine transmissible gastroenteritis virus. Reverse complement sequence are black.

2.3 冠状病毒基因组 5' 端非编码区二级结构分析
利用 GeneQuest 软件, 结合计算机辅助分析系统, 绘制出分离的 4 株 SARS-CoV 基因组 5' 端非编码区二级结构。图 2 结果显示: 4 株病毒基因组均可折叠成复杂的二级结构, 且形状完全一致, 此结构与其他国家和地区所报道的 SARS CoV 基因组 5' 端非编码区二级结构基本相同, 而与普通冠状病毒基因组相比有较大差别。但在 50 nt 左右, 所有冠状病毒基因组均可折叠成形状一致的、约由 15 nt 组成的“茎环结构”。

3 讨论

目前已知的普通冠状病毒分为 3 个族群, 主要包括人冠状病毒、牛冠状病毒、猪传染性腹泻病毒、鼠肝炎病毒和猪传染性肠胃炎病毒等, 主要导致人

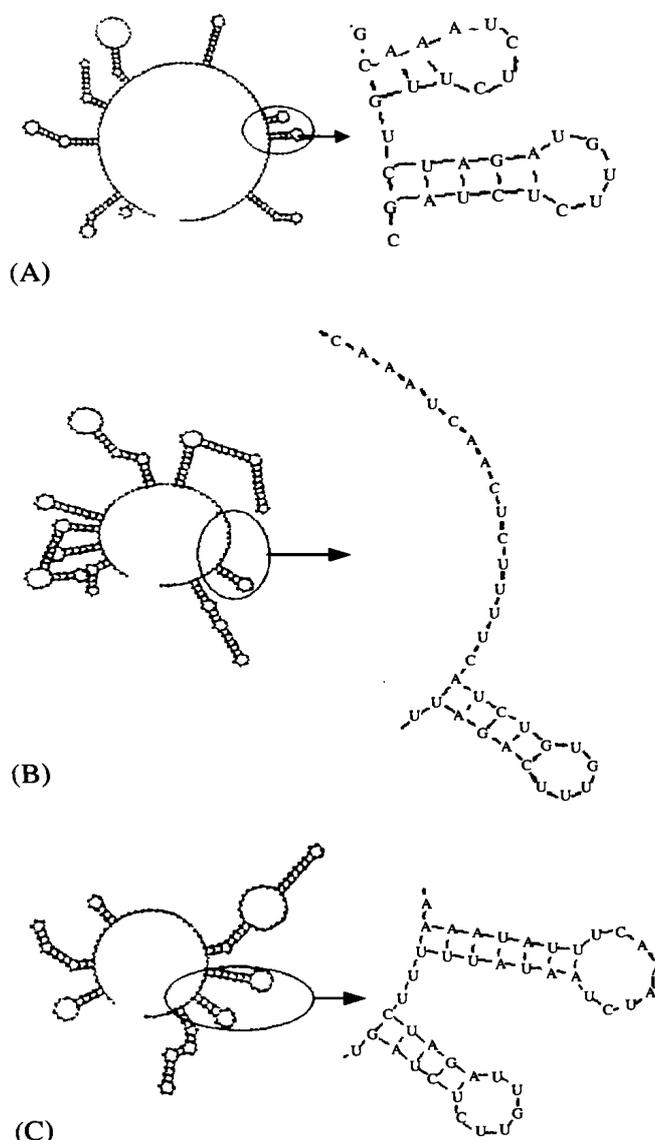


图 2 冠状病毒 5' 端非编码区二级结构及核心保守序列上游反向互补序列所形成的“茎环结构”

A, SARS CoV-GZ01 株; B, 人冠状病毒-229E 株; C, 牛冠状病毒-Mebus 株

Fig.2 Predicted secondary structure of 5' UTR of coronavirus and stem-loop structure formed by a reverse complement sequence upstream of the conserved core sequence

A, SARS-CoV GZ01 strain; B, Human coronavirus 229E strain; C, Bovine coronavirus -Mebus strain.

类普通感冒和引起动物、家禽类死亡; 显然, 目前在世界范围内广泛流行的 SARS-CoV 与之有明显差异。基于对 SARS-CoV 部分基因序列的系统进化分析证实: SARS-CoV 为一新型冠状病毒, 在系统分类上不属于上述 3 个族群^[4,5]。

为进一步了解此新病毒基因组结构和深入阐

明其致病机制,在完成分离的4株 SARS-CoV 的初步鉴定后,我们又对其非编码区的序列进行了测定和比较分析。结果显示:4株 SARS-CoV 基因组5'端非编码区的序列与已知的其他普通冠状病毒有很大差异,但与国际上所报道的 SARS-CoV 基因组高度一致,提示目前在世界范围内广泛传播的 SARS 病原体之间关系极为密切。至于造成各株 SARS-CoV 基因组之间5'末端有10多个碱基差别的原因,我们认为可能与不同实验室所采用的 RACE 法 RT-PCR 反应条件有关,但也不能排除是病毒本身的自然差别,有关结果有待进一步研究加以证实。

冠状病毒为正链单股 RNA 病毒,其增殖过程是在先导-引物转录合成机制的调控下进行,5'端非编码区的转录调控序列和先导序列,以及所形成高度保守的二级结构在病毒 mRNA 的转录合成过程中具有重要作用^[5]。4株 SARS-CoV 基因组5'端非编码区序列分析表明,在其起始密码子上游-197nt 处均发现转录调控核心保守序列 5'-CUAAAC-3',同时,在距此序列上游6nt 处都存在一个有15nt 组成的反向互补序列,该特点与其他国家和地区所报道的 SARS-CoV 高度一致,而且与其他普通冠状病毒的差别也不是很大。说明该序列特点对于冠状病毒完成其正常复制非常重要。

4株 SARS-CoV 基因组5'端非编码区折叠所形成的二级结构和目前其他国家和地区所报道的 SARS-CoV 基本一致,但与其他普通冠状病毒有很大差异。值得注意的是,虽然非编码区二级结构存在很大差别,但是由15nt 反向互补序列所形成的“茎环结构”在所有冠状病毒中却高度保守,进一步的分析表明,该反向互补序列的碱基组成并非高

度保守,说明对于调控蛋白质的识别来讲,这种高度保守的空间二级结构比核苷酸的种类更为重要。我们也注意到,在该高度保守的二级结构的下游2-3 nt, SARS-CoV 基因组的核心保守序列也参与形成一个“茎环结构”,该结构特点与牛冠状病毒 Mebus 株非常相似,但与核心保守序列不参与形成二级结构的人冠状病毒相差很远。可以推测 SARS-CoV 基因组5'端非编码区二级结构与普通冠状病毒的显著不同,对于其感染、复制等生理生化过程肯定有很大的影响,但这些差别的存在对其毒力有何影响以及 SARS-CoV 与其他病毒的亲源关系等都有待进一步的深入研究。

致谢: 本研究得到了于曼,刘洪,邓永强,谭刚,姜涛,康小平,杨银辉,吴劲松,熊伟,韩伟国,李钟铎等同事的大力帮助,在此向他们表示衷心的感谢!

参考文献

- [1] CDC. Outbreak of severe acute respiratory syndrome-worldwide, 2003[J]. *Morbidity and Mortality Weekly Report*, 2003, 52: 241.
- [2] WHO. Cumulative number of reported cases (SARS)[J/OL]. <http://www.who.int/csr/sarscountry/2003-04-03>.
- [3] 祝庆余,秦鄂德,王翠娥,等.非典型肺炎病例标本中新型冠状病毒的分离与鉴定[J]. *中国生物工程杂志*, 2003, 4(23): 106-112.
- [4] 秦鄂德,祝庆余,彭文明,等.重症急性呼吸综合征病例尸解肺组织中新型冠状病毒部分聚合酶基因的序列测定[J]. *军事医学科学院院刊*, 2003, 27(2): 81-83.
- [5] Marco A M, Steven J M, Caroline R A. The genome sequence of the SARS-associated coronavirus[J/OL]. <http://www.scienceexpress.org/1May2003/page/1/10.1126/science.1085953>