

SARS 相关病毒基因组序列的研究

谷雪梅, 来茂德

(浙江大学医学院病理教研室, 浙江 杭州 310031)

[摘要] 严重急性呼吸道综合征(severe acute respiratory syndrome, SARS)是由 SARS 相关病毒(SARS-CoV)引起的一种新的烈性呼吸道传染病。SARS-CoV 是一种新的冠状病毒, 基因组结构符合典型的冠状病毒特征, 编码基因从 5'端起分别为: RNA 依赖的 RNA 聚合酶、突起蛋白、包膜蛋白、膜蛋白, 3'端为核衣壳蛋白。对 14 株 SARS-CoV 的 4 个变异点分析, 可将 14 株病毒分成两个基因型, 即第 9404、19084、22222、27827 位的 T : T : T : T/C : G : C : C 两型。对 SARS-CoV 的基因组的全面了解不仅为疾病的发病机理研究提供主要线索, 而且将为 SARS 的诊断、预防和治疗提供帮助。

[关键词] 严重急性呼吸道综合征; SARS; SARS 相关病毒; 基因组

[中图分类号] R 373; R 563.1 **[文献标识码]** A **[文章编号]** 1008-9292(2003)04-0362-07

严重急性呼吸道综合征(Severe Acute Respiratory Syndrome, SARS)已经在 31 个国家和地区被发现。世界卫生组织(World Health Organization, WHO)迅速建立起由全球 10 个国家和地区的 13 个实验室组成的协作监测网络, 围绕 SARS 的各方面, 全面展开研究工作(<http://www.who.int/csr/sarsarchive/en/>)。自香港大学率先宣布分离出引起 SARS 的一种未知冠状病毒后, 世界各地包括加拿大、美国、新加坡、中国的北京和台湾等的研究机构相继对该病毒进行了分析鉴定^[1~3]。2003 年 4 月 16 日, WHO 正式确认 SARS 相关冠状病毒(SARS-CoV)是 SARS 的主要致病病原体, 但也有可能有其他协同因素如 hMPV (human metapneumovirus)使病情恶化。自 4 月中旬加拿大 BCCA (British Columbia Cancer Center Agency Genome Sciences Centre, British Columbia Centre for Disease Control and National Microbiology Laboratory, Canada)和美国 CDC (National Centers for Disease Control and Prevention, USA)的实验室率先公布了 SARS 病毒的基因组序列以来, 世界各地都相继报告了各自的测序结果。截止 2003 年 5 月 28 日 NCBI 的 SARS 相关序列达到 23 条, 其中全基因组序列 14 条(表 1)。

1 SARS 相关冠状病毒(SARS-CoV)概述

SARS-CoV 为正链的单链 RNA 病毒, 复制不经过 DNA 中间体, 使用标准密码子。在分类学上的地位是: ssRNA positive-strand viruses, no DNA stage (正链的单链 RNA 病毒, 无 DNA 中间体); Nidovirales (嵌套病毒目); Coronaviridae (冠状病毒科); Coronavirus (冠状病毒属); SARS coronavirus (SARS 冠状病毒)^[4]。冠状病毒是一组具有包膜的线性 RNA 病毒, 其病毒基因组在 27 kb~30 kb 之间, 在已经发表的冠状病毒全基因组序列中, G 和 C 的比例占 37%~42%, 是所有 RNA 病毒中最大的一种。其病毒 RNA 还具有一个正链 RNA 病毒所特有的重要结构特征: 即 RNA 5'端有甲基化帽(methylated cap), 3'端有 Poly A 尾的结构。该结构是冠状病毒基因组 RNA 自身即可发挥翻译模板作用的重要结构基础。目前完成的 SARS-CoV 的全基因组序列大小在 29 700 bp~29 800 bp 之间, GC 含量

收稿日期: 2003-06-18 修回日期: 2003-07-04

作者简介: 谷雪梅(1978—), 女, 硕士生, 主要从事肿瘤分子病理学研究。

通讯作者: 来茂德(1960—), 男, 博士, 病理学教授, 主任医师, 博士生导师, 主要从事肿瘤分子病理学研究。

在 40％左右,其基因组结构具有典型的冠状病毒特征.作为一种新发现的冠状病毒,美、加、中三个国家的研究组在发表各自基因组分析结果的同时,对 SARS-CoV 进行了系统发生学的分析 (phylogenetic analyses)^[5~7],均得出了相同的结论,即 SARS-CoV 与以往任何已经确定的冠状病毒并不完全一致,在系统发生上与其他冠状病毒成员基本上是等距的,属于冠状病毒科中独立的第 4 分支(表 2)^[4]。

表 1 NCBI 上 SARS-CoV 基因组序列信息汇总(截止 2003 年 5 月 28 日)
Table 1 The genome sequences associated with SARS-CoV from NCBI(up to 2003-05-28)

Place	Accession No.	Name(length)
China,main land	AY297028	SARS coronavirus ZJ01,complete genome(29715 bp)
	AY278488	SARS coronavirus BJ01,complete genome(29725 bp)
	AY279354	SARS coronavirus BJ04,partial genome
	AY278490	SARS coronavirus BJ03,partial genome
	AY278487	SARS coronavirus BJ02,partial genome
	AY278489	SARS coronavirus GZ01,partial genome
China,Hong Kong	AY278554	SARS coronavirus CUHK-W1,complete genome(29736 bp)
	AY282752	SARS coronavirus CUHK-Su10,complete genome(29736 bp)
	AY278491	SARS coronavirus HKU-39849,complete genome(29742 bp)
	AY268070	SARS coronavirus Hong Kong/03/2003 RNA-directed RNA polymerase gene,complete
China,Tai Wan	AY286402	SARS coronavirus Taiwan JC-2003 RNA-directed RNA polymerase(pol) gene,partial cds
	AY291451	SARS coronavirus TW1,complete genome(29729 bp)
	AY268049	SARS coronavirus Taiwan RNA-directed RNA polymerase(pol) gene,partial cds
Singapore	AY283798	SARS coronavirus isolate SIN2774,complete genome(29711 bp)
	AY283797	SARS coronavirus isolate SIN2748,complete genome(29706 bp)
	AY283796	SARS coronavirus isolate SIN2679,complete genome(29711 bp)
	AY283795	SARS coronavirus isolate SIN2677,complete genome(29705 bp)
	AY293794	SARS coronavirus isolate SIN2500,complete genome(29711 bp)
Canada	NC—004718	SARS coronavirus,complete genome(29751 bp)
	AY274119	SARS coronavirus TOR2,complete genome(29751 bp)
	AY271716	SARS coronavirus Tor2 RNA polymerase 1b mRNA,partial cds
America	AY278741	SARS coronavirus Urbani,complete genome(29727 bp)
Vietnam	AY269391	SARS coronavirus Vietnam strain 200300592 polymerase gene,partial cds

表 2 SARS-CoV 与已知冠状病毒主要编码蛋白氨基酸序列相似性比较分析^[6]
Table 2 Pairwise identities of SARS-CoV proteins with other known coronaviruses^[6]

Groups	Virus	Pairwise amino acid identity(percent)						
		3CLPRO	POL	HEL	S	E	M	N
1	HCoV-229E	40.1	58.8	59.7	23.9	22.7	28.8	23.0
	PEDV	44.4	59.5	61.7	21.7	17.6	31.8	22.6
	TGEV	44.0	59.4	61.2	20.6	22.4	30.0	25.6
2	BCoV	48.8	66.3	68.3	27.1	20.0	39.7	31.9
	MHV	49.2	66.5	67.3	26.5	21.1	39.0	33.0
3	IBV	41.3	62.5	58.6	21.8	18.4	27.2	24.0

HCoV-229E:human coronavirus 229E(人冠状病毒),af304460; PEDV:porcine epidemic diarrhea virus(猪传染性腹泻病毒),af35351; TGEV:transmissible gastroenteritis virus(传染性胃肠炎病毒),aj271965; BCoV:bovine coronavirus(牛冠状病毒),af220295; MHV:murine hepatitis virus(鼠肝炎病毒),af201929; IBV:infectious bronchitis virus(感染性支气管炎病毒),m95169; 3CLPRO:3 chymotrypsin-like proteinase(3CL 样蛋白酶); POL:RNA-dependent RNA polymerase(RNA 依赖的 RNA 多聚酶); HEL:RNA helicase(RNA 螺旋酶); S:spike protein(突起蛋白); E:small envelop protein(包膜蛋白); M:membrane protein(膜蛋白); N:nucleocapsid(核衣壳蛋白)

2 SARS-CoV 的基因组结构特征

各地报道的 SARS 病毒基因组均为正链的单链 RNA 病毒,在 5'端和 3'端分别含有甲基化帽子和多聚腺苷酸尾巴。如前所述,其基因组结构符合典型的冠状病毒特征,基因顺序从 5'端起分别为 RNA 依赖的 RNA 聚合酶 [rep]、突起蛋白 [S]、包膜蛋白 [E]、膜蛋白 [M]、3'端为核衣壳蛋白 [N]。在 5'和 3'两端还有短的非编码序列区(UTR)。Rep 基因占基因组全长的 2/3,预计可以编码两种蛋白质(分别由 ORF1a 和 ORF1b 编码)。在 rep 基因下游有 4 个 ORF,编码病毒结构蛋白,按基因组上的排列顺序依次为 S、E、M 和 N,与其他已知的冠状病毒相似(图 1)。但是冠状病毒基因组中位于某些病毒 ORF1b 和 S 之间的血凝素酯酶

基因在 SARS-CoV 中并没有发现。冠状病毒基因组还编码一些位于 S 和 E、M 和 N 之间以及 N 蛋白下游的非结构蛋白。这些非结构蛋白在不同的冠状病毒中变异很大,在已有蛋白质序列数据库中未找到任何同源的蛋白质氨基酸序列,其功能还不明确,故称之为 PUP(putative uncharacterized protein)。但各个研究组报告的 SARS-CoV 基因组之间 PUP 存在着差异。中国对 BJ01 株及美国对 Urbani 株的分析报道,基因间的这些区域编码 5 个大于 50 个氨基酸的 PUP。2 个位于 S 和 E 之间,由两个重叠的 ORF 分别编码,其余 3 个 PUP 位于 M 和 N 之间;而加拿大对 Tor2 株的分析,则发现了 9 个可能的未知 PUPs 和 1 个位于 3'端(第 29590-29621 位)的 s2m motif^[5]。

在各株 SARS 病毒全基因组序列中研究

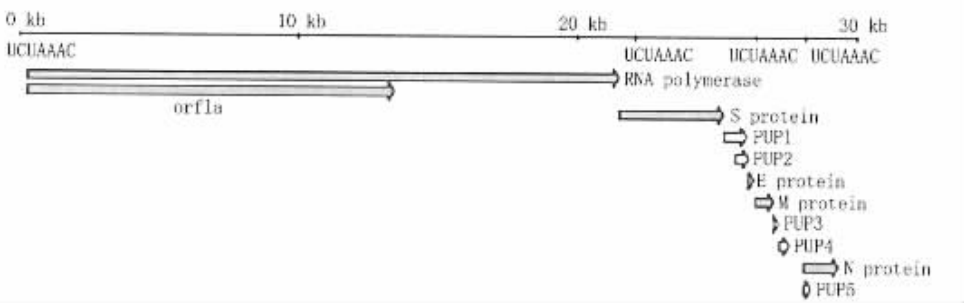


图 1 SARS 相关病毒的基因组结构

Fig. 1 Genome structure of SARS-associated virus^[7]

还发现,在编码 RNA 依赖的 RNA 聚合酶的起始端和编码 N 蛋白与 PUP 蛋白的部分重叠区都有明显的 GC 富集区。同时在基因组中还发现有典型的重复序列 AAACGAAC,位于 ORF1a, S 蛋白, PUP1(BJ01 ORF3), M 蛋白, PUP4(BJ01 ORF8), PUP5(Tor2 ORF9)和 N 蛋白上游,它的功能很可能是转录调节序列(TRS)的保守核心序列,其功能及作用有待于进一步的实验证实。该发现也进一步提示 SARS-CoV 的不连续转录模型,也为 ORF 预测的准确性提供了强有力的证据。

此外,中国 BJ01 的分析结果还发现,至少 4 个单位长度在 7 个碱基以上的回文结构(palindromes),140 多个可能形成发夹结构

(hairpin) 的片断,以及发现 BJ01 基因组上 155-211/861-920 这两个区域内的大约 60 个碱基出现重复现象^[7]。这些基因组现象的生物学意义,均有待于进一步研究来阐明其潜在的功能。

3 SARS-CoV 基因组编码的蛋白质分析及预测

目前对 SARS-CoV 蛋白质的研究,主要集中于基因组分析所得的核酸与氨基酸序列,利用生物信息学的方法和工具,即运用一系列的氨基酸序列分析和蛋白质结构预测工具,对其可能的编码蛋白的各种理化参数、生物学功能区(如:磷酸化位点、半胱氨酸富集区、糖基化位

点等)、二级结构、三级结构与功能域等各方面进行全面分析。主要的分析预测工具都可以在线得到,如表 3。北京大学医学院医学生物信息组借助网络平台——中国医学生物信息网

(<http://cmbi.bjmu.edu.cn/cmbidata/sars/>)还针对部分重要的蛋白质进行了细致的分析预测工作,并及时予以发布^[8~11]。

表 3 对 SARS-CoV 进行分析的主要软件
Table 3 Main analysis software to SARS-CoV

Software	Main function	On-line service or download website
ClustalW	general purpose multiple sequence alignment program for DNA or proteins	http://www.ebi.ac.uk/clustalw
Phylip	a free package of programs for inferring phylogenies	http://evolution.genetics.washington.edu/phylip.html
PredictProtein	an online service for sequence analysis, and structure prediction	http://www.cbi.pku.edu.cn/predictprotein/
Signal P	an online service for prediction of signal peptides of proteins	http://genome.cbs.dtu.dk/services/SignalP http://genome.cbs.dtu.dk/services/SignalP/mailserver.html
TMPRED	a transmembrane helix prediction program based on the statistical analysis of Tmbase databank.	http://www.ch.embnet.org/software/TMPRED—form.html
TMHMM	a membrane protein topology prediction method which is based on a hidden markov model	http://www.cbs.dtu.dk/services/TMHMM/
DNASar	a powerful program for most of sequence analysis	http://www.lmbe.seu.edu.cn/SARS/

全基因组分析表明,SARS-CoV 含有 11 个 ORFs,包括 6 个编码结构蛋白的 CDSs 和 5 个 PUPs 编码区,在蛋白氨基酸的组成中,亮氨酸出现频率最高,天冬氨酸出现频率最低。

编码 RNA 依赖的 RNA 聚合酶的 CDS 约占整个基因组的 2/3,与其它冠状病毒相比,该蛋白无论在长度和氨基酸序列上都是保守的,是 SARS 侵入宿主细胞之后,启动其它一切生命活动的关键蛋白质。目前认为 SARS 冠状病毒的 RNA 聚合酶是由两个完全不同的 ORF,即 ORF1a 和 ORF1b 分别编码,然后组装成成熟的聚合酶,而不是一个 ORF 编码之后,切割再组装而成的。这两个 ORF 的翻译及调控机制尚不明确,可能其翻译过程并不依赖于基因组 RNA 上 5'端的先导序列与非保守序列而独立地启动。这样的翻译机制可以极大提高病毒进入宿主细胞后合成其复制必需的 RNA 聚合酶的效率,缩短基因组 RNA 在细胞浆内“孤立存在”的时间,进而提高病毒生存的可能性,也

可以解释 SARS-CoV 的毒力为何远远高于同类其他冠状病毒,在短期内造成组织细胞严重损伤的情况。经蛋白质结构域分析表明,上述 ORF 编码的复合蛋白体中,除含有 RNA 聚合酶之外,还可能含有与鼠肝炎病毒类似的磷脂酶结构域(PLP^{PRO})、胰凝乳蛋白酶样结构域(3CL^{PRO})和一些疏水结构域^[6]。其中胰凝乳蛋白酶样结构域(3CL^{PRO})已经由找到的同源模型给予证实^[12]。

S 蛋白紧靠 RNA 多聚酶编码区,含有 1 255个氨基酸,是 SARS 结构蛋白中最大的一个蛋白质,相对分子量约为 139 kD。现已确定 S 蛋白含有一个疏水跨膜区和至少 9 个糖基化位点,由两个结构域组成,即靠近 N 端的球状结构和靠近 C 端的穿膜的棒状结构域,与冠状病毒侵入细胞过程密切相关。通过对其二级结构的预测发现存在着 3 个螺旋(helix)区域,可能和病毒外壳膜与宿主细胞膜的融合有关^[8]。另外,在 SignalP 信号肽预测还发现其 N 端 1~

13 个氨基酸可能为信号肽序列,具体功能未知^[5]。

E 蛋白是 SARS 病毒中最小的结构蛋白,编码一段 76 个氨基酸的蛋白,其疏水性最强,为 II 型包膜蛋白,定位在病毒颗粒表面。SignalP 和 TMpred 分析均显示其序列第 12~34 位存在跨膜区。另外,TMHMM (transmembrane hidden markov model) 分析显示其 C 末端有一个 46 个氨基酸的疏水区域位于病毒包膜外^[5]。

M 蛋白是一个跨膜糖蛋白,中等大小,由 221 个氨基酸构成,存在三个疏水跨膜螺旋区。它与 S 蛋白的结合对病毒包膜构成是必不可少的,并且在病毒装配位点上两种蛋白的集聚也是必要的。其 C 末端的 121 个氨基酸位于病毒颗粒内部,推测是与病毒核衣壳蛋白(N 蛋白)相互作用的部分^[5]。

N 蛋白(核衣壳蛋白)位于基因组 3'端,大小仅次于 S 蛋白,由 422 个氨基酸构成,为一碱性蛋白,等电点为 10.1,它的亲水性最强。N 蛋白处于病毒颗粒的核心部分,与基因组 RNA 结合形式存在,对于病毒基因组 RNA 特征性序列的识别及其与其他结构蛋白的相互作用,对于病毒的正确组装有重要意义^[13]。加拿大研究小组发现,SARS-CoV 病毒的 N 蛋白存在一独特的富含赖氨酸区域(KTFPPTEPKKDKKKKTDEAQ)。该特征性序列的功能尚不清楚,可能提示 SARS-CoV 的 N 蛋白含有一异常的核功能,参与 SARS 致病机制^[5]。

如前所述,SARS-CoV 基因组中还存在多个 ORF 编码的未知蛋白,在已知蛋白质库中找不到任何同源序列。对于这类蛋白,加拿大、美国、中国等地的研究小组利用已有的工具软件,对其特征进行了初步的分析和预测,还对其功能作出了初步的推断,但最终仍需进一步的实验室工作予以证实。

4 SARS-CoV 基因组的变异分析

SARS 是一种短期爆发的烈性传染性疾病。分析比较不同地区不同时间 SARS-CoV 的序列差异与突变状况,对于了解疾病的传播过

程,病毒的毒力变化有非常重要的意义。同时,研究病毒尤其是编码抗原性蛋白的基因的变异性是研制 SARS 相关病毒疫苗必不可少的。

北京大学医学部生物信息组的成员对 SARS-CoV 突变迁移做了动态的观察(<http://cmbi.bjmu.edu.cn/cmbidata/sars/SARS-GENOME/MutationMovement.htm>)。到 2003 年 4 月 21 日,GenBank 收录有 2 条 SARS-CoV 全基因组序列(TOR2,Urbani),通过比对这两条序列,一共发现 7 个核苷酸变异,有 4 个可以导致错义突变,即氨基酸序列的改变。到 2003 年 5 月 1 日,GenBank 已经收录了 5 条 SARS-CoV 全基因组序列(除外前面两条外,又增加 BJ01,HKU-39849,CUHK-W1),此时核苷酸替换已达 31 个,其中 21 个可以导致错义突变。到 2003 年 5 月 13 日,GenBank 已经收录 SARS-CoV 全基因组序列已达 12 条,核苷酸替换已达 42 个,28 个可导致错义突变。而截止 2003 年 5 月 28 日,GenBank 已经收录 SARS-CoV 全基因组序列已达 14 条,127 个单核苷酸序列突变,94 个可导致错义突变^[14]。从分析结果看出,随着提交序列的增加,更多的变异被发现。比较这些来自不同国家和地区的 SARS 相关病毒的全基因组序列时发现,其 5'端稳定区编码 RNA 多聚酶,序列变异性较低,在 3'可变区分别编码有 S、E、M、N 结构蛋白的 4 个 CDSs 和 5 个 PUPs,存在较高的碱基替换突变率。

新加坡研究人员通过对 14 株 SARS 冠状病毒全基因序列的比较分析^[14],总共得到 127 个变异单核苷酸序列(single nucleotide sequence variation)和导致 94 个相应变异的氨基酸。此外,SIN2677 及 SIN2748 分别发生 5 和 6 个核苷酸缺失,缺失位于不明开放阅读框(uncharacterized open reading frame)与核衣壳蛋白编码序列之间的非编码区(<http://image.thelancet.com/extras/03art4454webappendix2.pdf>)。然而各分离株之间存在的小量核苷酸变异,很可能是病毒体外培养的结果,即 SARS 相关病毒在非洲绿猴肾(Vero)细胞中繁殖的自适应结果,因此研究人员为了排除这种细胞

培养中突变所造成的干扰,又采用似然法重新分析实验数据。序列对比分析表明 127 个序列变异中,2 株或更多有 16 个变异位点(variant loci),3 株或更多有 8 个变异位点。采用更为严格的标准发现,4 个变异位点在 14 株 SARS 冠状病毒重复出现了 5 次以上。此 4 个变异位点可将 14 株病毒分为两种基因型,即第 9 404、19 084、22 222、27 827 位的 T : T : T : T/C : G : C : C 两型,因为从统计学角度分析 4 个位点同时变异的几率相当小。结合 SARS 传染源分析发现,所有 T : T : T : T 基因型的标本都与香港 M 旅馆有关,而所有 C : G : C : C 基因型的标本与香港 M 旅馆都无关系,故这一四位点变异可能是第一个不同株 SARS 冠状病毒的遗传标识。基于共同变异序列的进化树分析进一步证明,与 M 旅馆感染相关的病例与无关病例分属截然不同的两支。

除了此四位点基因型,另外四个共有变异序列(均在 14 株病毒中重现 3 至 4 次)似乎可以进一步确定地理位置上的亚群。第 9 854 位、19 838 位、27 243 位碱基是北京株特异性的,而第 19 084 位碱基序列可以区分新加坡与其它地区的 SARS 冠状病毒。

综上所述,在比较了不同来源和地区的 SARS 相关病毒的基因组后,虽然各病毒株之间存在着差别,但从整体上看,SARS-CoV 还是一个相对比较稳定的冠状病毒。发现一些共同序列的多态性非常有助于确定不同 SARS 的感染来源,但是这些共同多态性是否会引起生物学及临床症状的差异,仍有待进一步探讨。在 SARS-CoV 的所有功能和抗原决定簇蛋白被确定以前,这些基因组间序列的差别对于病毒的发病机理的意义还不能确定。

5 结 语

SARS 严重威胁着人类的健康,也给世界各国的卫生事业及经济带来了严重的压力,因此了解此病毒的本质以及研究控制方法尤为重要。随着各国及地区对 SARS 相关病毒全基因组测序的完成和公布,人们对 SARS 有了全新的认识,也更加确证了 SARS 相关病毒是一种全新的病原体,一种新的冠状病毒,而不是其他

病毒的简单重组,其毒力远远强于其它已知的冠状病毒家族成员。对 SARS-CoV 的基因组知识的全面了解不仅为疾病的发病机理和病毒的毒力研究提供了重要的线索,而且将为 SARS 的诊断、预防和治疗提供帮助。对该病毒的多基因序列比对发现,不同的病毒株之间存在有一些共同序列的多态性,这些共同多态性是否会引起病毒的生物学及临床症状的差异仍值得研究;更多的 SARS 病毒的全基因组序列的完成,以及这些序列的多态性分析,有助于确定不同 SARS 感染的来源。

References:

- [1] Peiris J S M, Lai S L, Poon L L M, et al. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet*, Published on line on April 8, 2003. <http://image.thelancet.com/extras/03art3477web.pdf>
- [2] Drosten C, Gunther S, Preiser W, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome [J]. *N Engl J Med*, 2003, 348: 1967—1976.
- [3] Ksiazek T G, Erdman D, Goldsmith C S, et al. A novel coronavirus associated with severe acute respiratory syndrome [J]. *N Engl J Med*, 2003; 348: 1953—1966.
- [4] LAI Mao-de, ZHU Yi-min, GU Xue-mei (来茂德, 朱益明, 谷雪梅). Severe acute respiratory syndrome (SARS) [J]. *Journal of Zhejiang University (Medical Sciences)* [浙江大学学报(医学版)], 2003, 32(3): 167—170. (in Chinese)
- [5] Marco A M, Steven J M J, Caroline R A, et al. The genome sequence of the SARS-Associated Coronavirus [J]. *Science*, 2003, 300: 1399—1404.
- [6] Paul A R, Steven O M, Stephan S M et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome [J]. *Science*, published on line on May 1, 2003. <http://www.sciencexpress.org>
- [7] Qin E D, Zhu Q Y, Yu M, et al. A complete sequence and comparative analysis of a SARS-associated virus (Isolate BJ01) [J]. *Chinese Science Bulletin*, 2003, 48 (10): 941—948.
- [8] SHI Lei, ZHANG Qi-peng, LU Ming, et al (石磊, 张其鹏, 卢铭, 等). Preliminarily analyze the structure and function of the putative spike protein of SARS virus. CMBI, Published on line on May 1, 2003. <http://cmbi.bjmu.edu.cn/cmbidata/sars/sars—secstructure/spike.htm>. (in Chinese)
- [9] SHI Lei, ZHANG Qi-peng, RUI Wei, et al (石磊, 张其

鹏,芮伟,等). Preliminarily analyze the structure and function of the putative nucleocapsid protein. CMBI, Published on line on May 4, 2003. <http://cmbi.bjmu.edu.cn/cmbidata/sars/sars—secstructure/nuckeocapsid%20protein.htm>. (in Chinese)

[10] SHI Lei,ZHANG Qi-peng,RUI Wei,et al(石磊,张其鹏,芮伟,等). Searching the possible binding site of the SARS N protein and SARS RNA. CMBI, Published on line on May 9,2003. <http://cmbi.bjmu.edu.cn/cmbidata/sars/sars—secstructure/NBinding.htm>. (in Chinese)

[11] SHI Lei,ZHANG Qi-peng,LU Ming,et al(张其鹏,石磊,卢铭,等). Nuclear targeting sequence in SARS virus nucleocapsid protein. CMBI, Published on line on May 9, 2003. <http://cmbi.bjmu.edu.cn/cmbidata/sars/sars—secstructure/Nlocation.htm>. (in Chinese)

[12] RUI Wei,ZHANG Qi-peng,SHI Lei,et al(芮伟,张其鹏,石磊,等). Tertiary structure prediction of SARS coronavirus protein's partial sequence. CMBI, Published on line on May 9,2003. <http://cmbi.bjmu.edu.cn/cmbidata/sars/sars3d/sars3d.htm>. (in Chinese)

[13] Chen H,Wurm T,Britton P,et al. Interaction of the coronavirus nucleoprotein with nucleolar antigens and the host cell [J]. J Virol,2002,76(10):5233—5250.

[14] Ruan Y J,Chia L W,Ling A E,et al. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. Lancet, Published on line on May 9,2003. <http://image.thelancet.com/extras/03art4454web.pdf>

[责任编辑 张荣连]