

印象初 印展 施鉴屏 编著

非典型肺炎(SARS) 冠状病毒基因全序列

SEQUENCES DATA OF SARS
CORONAVIRUS COMPLETE GENOME

山东科学技术出版社
www.lkj.com.cn

280363

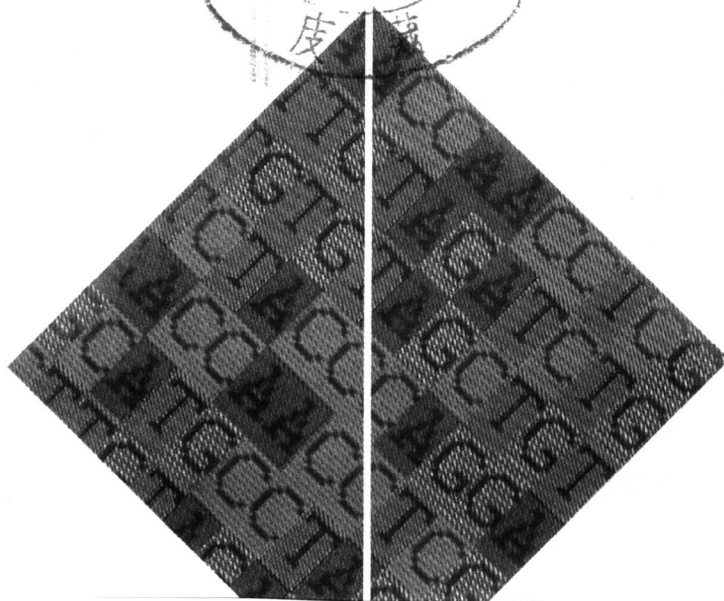
R373.1
Y451
2003
C.1

印象初 印 展 施鉴屏 编著

By Yin Xiangchu Yin Zhan and Shi Jianping

非典型肺炎(SARS) 冠状病毒基因全序列

SEQUENCES DATA OF SARS CORONAVIRUS COMPLETE GENOME



00125581

山东科学技术出版社

图书在版编目(CIP)数据

非典型肺炎(SARS)冠状病毒基因全序列/印象初, 印展, 施鉴屏编著. —济南: 山东科学技术出版社, 2003

ISBN 7-5331-3463-X

I. 非... II. ①印... ②印... ③施... III. 重症呼吸综合症—日冕形病毒—基因—序列 IV. R373.1

中国版本图书馆 CIP 数据核字 (2003) 第 038615 号

非典型肺炎(SARS)冠状病毒基因全序列

印象初 印 展 施鉴屏 编著

出版者: 山东科学技术出版社

地址: 济南市玉函路 16 号

邮编: 250002 电话: (0531) 2065109

网址: www.lkj.com.cn

电子邮件: sdkj@jn-public.sd.cninfo.net

发行者: 山东科学技术出版社

地址: 济南市玉函路 16 号

邮编: 250002 电话: (0531) 2020432

印刷者: 山东新华印刷厂

地址: 济南市胜利大街 56 号

邮编: 250001 电话: (0531) 2059512

开本: 787mm × 1092mm 1/16

印张: 16.5

字数: 365 千

版次: 2003 年 6 月第 1 版第 1 次印刷

ISBN 7-5331-3463-X

R·1061

定价: 40.00 元

内 容 提 要

本书依据 Genbank 的数据,将非典型肺炎 SARS 冠状病毒基因全序列对比后认为,目前全世界为 18 变种:即多伦多 2NC(Tor2NC) = 多伦多 2AY(Tor2AY),台湾 1(TW1),新加坡 2679(Sin2679),新加坡 2677(Sin2677),新加坡 2774(Sin2774),新加坡 2500(Sin2500),新加坡 2748(Sin2748),香港中大-su10(CUHK-su10),法兰克福(Frankfurt),乌尔巴尼(Urbani),香港大学(HKU),香港中大-w1(CUHK-w1),北京 1 号(BJ01),北京 2 号(BJ02),北京 3 号(BJ03),北京 4 号(BJ04),浙江 01(ZJ01)和广州 1 号(GZ01),并列出了它们的亲缘关系。非典型肺炎 SARS 冠状病毒基因全序列可能有 29787 bp,但各变种实有数为 29705-29757 bp,丢失数为 30-82 bp,丢失的位置也不同,其原因尚不明。各变种间差异数为:1-72 bp,说明变异明显有异,可能非同一起来源。书中 18 变种的基因全序列对比数据可用来鉴定已知种类和新变种。

本书可供病毒分类工作者、医务工作者、医药研究工作者、大专院校师生参考。

Abstract

Based on the sequences data of SARS Coronavirus in Genbank, to compare those data showed that there are 18 varieties of SARS Coronavirus in the world, i. e. Tor2NC + Tor2AY, TW1, Sin2679, Sin2677, Sin2774, Sin2500, Sin2748, CUHK-su10, Urbani, Frankfurt, HKU, CUHK-w1, BJ01, BJ02, BJ03, BJ04, ZJ01 and GZ01. A phylogenetic tree of 18 varieties is given. SARS Coronavirus may has 29787 bp, but the real numbers were 29705-29757 bp, the lacking numbers were 30-82 bp, lacking places were very different, why? the reason is unknown. They have different number(1-72 bp) in 18 varieties. It showed that the degeneration is very distinct and it may be not one source. The comparison of sequences data of SARS Coronavirus may be to identify the known and new varieties.

This book is a reference for virus taxonomists, medical workers, medicine researchers, teachers and students of university or college.

前 言

2003年2月,非典型肺炎 SARS(Severe Acute Respiratory Syndrome)病在人间出现,这是一种新的非常严重的疾病,当时由于不知何种病原体,令世人束手无策和震惊。不久,科学家们找到了它的病原体为一种新的冠状病毒。4月18日,一种名为香港大学非典型肺炎冠状病毒 HKU SARS Coronavirus 基因全序列在国际互联网的基因银行 Genbank 中公布。到6月13日,共有19个 SARS 病毒基因序列公布,令人高兴。

当前首要问题是如何认识 SARS 病毒? 它们的变异有多大,有多少种? 各地为害的是一种,还是多种? 各个种发病情况有什么不同? 何故?

我们从 Genbank 下载了 SARS 基因序列数据,用 Clustal 和 Mega 软件进行分类研究。比较结果初步认为现有18个变种: 多伦多 2 NC(Tor2NC) = 多伦多 2AY(Tor2AY), 台湾 1(TW1), 新加坡 2679(Sin2679), 新加坡 2677(Sin2677), 新加坡 2774(Sin2774), 新加坡 2500(Sin2500), 新加坡 2748(Sin2748), 法兰克福(Frankfurt), 香港中大-su10(CUHK-su10), 乌尔巴尼(Urbani), 香港大学(HKU), 香港中大-w1(CUHK-w1), 北京1号(BJ01), 北京2号(BJ02), 北京3号(BJ03), 北京4号(BJ04), 浙江01(ZJ01)和广州1号(GZ01)。

本书建立一个虚拟种 Fiction, 基因序列总数应为 29787 碱基 ($a = 8496, c = 5949, g = 6189, t = 9153$), 可视为模式用来鉴定已知变种和新的变种。

我们向奋战在第一线的 SARS 的科学和医务工作者们,特别是从事基因序列研究的同行们致以崇高的敬意!

作者们非常感谢山东省人民政府和山东农业大学为我们提供经费和创造条件,进行本项目的研究;也感谢山东科学技术出版社以社会效益为重,国家需要为急,采用超常方式快速出版本书。

最后,我们相信在全人类的共同努力下,彻底战胜 SARS 为期不远。如果本书能在推动这一进程中起到一点作用,我们将感到非常高兴!

Preface

In February 2003, SARS (Severe Acute Respiratory Syndrome) is appeared in the human world, which is a new very severe disease, at same time, the pathogen of SARS is unknown, so it felt quite helpless and shocked the in the country and the world. In a short time, the scientists found that the new coronavirus is the pathogen of SARS. 18th April 2003, the sequences data of SARS Coronavirus HKU was reported in the Genbank. Up to the 13th June 2003, there are sequences data of 19 units of SARS Coronavirus were finished. When to read those results, we are very glad.

It is a first problem. How to know the SARS Coronavirus? How much is their degenerations? How many varieties are there in a locality? Every one has those characters in the symptom? Why? And so on.

We download sequences data of 19 units SARS Coronavirus complete genome from Genbank. According to these data, the comparison of sequences of 18 varieties SARS Coronavirus complete genome is given by the Clustal Software. The result showed that there are 18 varieties of SARS Coronavirus with different numbers of bp (see Table 1) in the world as below: Tor2NC = Tor2AY, TW1, Sin2679, Sin2677, Sin2774, Sin2500, Sin2748, Frankfurt, CUHK-su10, Urbani, HKU, CUHK-wl, BJ01, BJ02, BJ03, BJ04, ZJ01 and GZ01.

A fiction of SARS Coronavirus (total 29787 bp, a = 8496, c = 5949, g = 6189, t = 9153) is established in this book. It may be as a type for to identify known varieties and new variety.

We pay sincere tribute to the scientists and medical workers in the first line of SARS, specially the colleagues on the work of sequences data.

The authors are grateful to The People's Government of Shandong Province and Shandong Agricultural University for providing grant and conditions to do this program. We are also particularly grateful to the Shandong Sciences and Technology Press for to attach importance to social effect and urgent to national need with above average speed to publish this book quickly.

Finely we believe that all humanity make great efforts altogether and can conquer the SARS entirely in the near future. If this book has some effect in the process, we are very delighted.

目 录

1. 18 个变种病毒基因全序列的数据来源	1
2. 方法和初步结果	2
3. 18 个变种病毒基因全序列比较	8
4. 参考文献和模拟变种基因全序列的数据	229

Contents

1. Source of sequences data of 18 SARS Coronavirus varieties complete genome	1
2. Method and preliminary result	2
3. Table on the comparison of sequences of 18 SARS Coronavirus variety complete genome	8
4. References and sequences data of fiction	229

1.18 个变种病毒基因全序列的数据来源

1. Source of sequences data of 18 SARS Coronavirus varieties complete genome

18 个变种病毒基因全序列的数据来自互联网的基因银行 (见表 1)。

Download sequences data of 18 varieties SARS Coronavirus complete genome from Genbank. (see Table 1).

表 1 SARS 冠状病毒基因全序列的数据

Table 1 Sequences data of SARS Coronavirus complete genome

编号 No.	Total	a	Bp c	g	t	others	产地名 Locality	作者 Authors	基因库序列号 No. In Genbank
1	29742	8475	5942	6183	9142		HKU	Leang, et al, 2003 - 04 - 18	AY278491
2	29727	8455	5940	6188	9144		Urbani	Mouroe, S., 2003 - 04 - 21	AY278741
3	29725	8465	5941	6185	9134		BJ01	Qin et al, 2003 - 05 - 01	AY278488
4	29736	8476	5938	6185	9137		CUHK-su10	Tsui, et al, 2003 - 05 - 07	AY282752
5	29711	8453	5936	6185	9137		Sin2774	Ruan et al, 2003 - 05 - 09	AY283798
6	29706	8451	5935	6184	9135	1	Sin2748	Ruan, et al 2003 - 05 - 09	AY283797
7	29711	8453	5936	6185	9137		Sin2679	Ruan, et al 2003 - 05 - 09	AY283796
8	29705	8447	5936	6187	9135		Sin2677	Ruan, et al 2003 - 05 - 09	AY283795
9	29711	8453	5937	6184	9137		Sin2500	Ruan, et al 2003 - 05 - 09	AY283794
10	29729	8457	5940	6188	9144		TW1	Yeh, et al 2003 - 05 - 14	AY291541
11	27951	8481	5940	6187	9143		Tor2AY	? 2003 - 05 - 16	AY274119
12	29715	8458	5934	6187	9135	1	ZJ01	Li, et al 2003 - 05 - 19	AY297028
13	29736	8476	5942	6185	9133		CUHK-wl	Tsui, et al, 2003 - 05 - 29	AY278554
14	29757	8481	5953	6189	9134		GZ01	Qin, et al, 2003 - 06 - 05	AY278489
15	29745	8472	5943	6184	9146		BJ02	Qin, et al 2003 - 06 - 05	AY278487
16	29740	8464	5951	6183	9142		BJ03	Qin, et al, 2003 - 06 - 05	AY278490
17	29732	8469	5942	6185	9136		BJ04	Bi, et al, 2003 - 06 - 05	AY279354
18	29727	8458	5935	6186	9148		Frankfurt	Thiel et al, 2003 - 06 - 11	AY291315
19	29751	8481	5940	6187	9143		Tor2NC	Marra, et al 2003 - 06 - 13	NC004718
20	29787	8496	5949	6189	9153		Fiction	Yin et al 2003 - 06 - ?	In this book
20*	+ 45	+ 21	+ 7	+ 6	+ 11		Fiction*	Yin et al 2003 - 06 - ?	In this book

HKU + Fiction* = Fiction

No.1 + No.20* = No.20

No.1 29742 8475 a 5942 c 6183 g 9142 t

No.20* +45 +21 a +7 c +6 g +11 t (来自 From ZJ01, GZ01, CUHK-su10 A C G T.)

No.20 29787 8496 a 5949 c 6189 g 9153 t

No.20* = No.12* + No.14* + No.4* = ZJ01* + GZ01* + CUHK-su10*

No.12* 7 3 a 0 c 1 g 3 t ZJ01*

No.14* 29 9 a 7 c 5 g 8 t GZ01*

No.4* 9 9 a 0 c 0 g 0 t CUHK-su10*

No.20* 45 21 a 7 c 6 g 11 t

2. 方法和初步结果

19 个病毒基因全序列的数据用 Clustal 软件进行比较,得出全序列比较结果。用 Mega 软件计算出 18 个变种病毒基因丢失序列的数量(见表 2)、不同序列的数量、百分率(见表 3)和分子系统树(图 1)。

2.Method and Preliminary Result

Comparison of sequences data of 19 units SARS Coronavirus complete genome with Clustal X (1.81), multiple sequence alignment is given. Percentage and numbers of different sequences(see Tab.3) and numbers of lacking sequence (see Tab.2) in 18 varieties SARS Coronavirus complete genome and phylogenetic tree of 18 varieties of SARS Coronavirus (see Fig.1) are given by Mega Software.

表 2 18 个变种 SARS 冠状病毒基因丢失序列的数量

Table 2 The numbers of lacking sequence
in 18 varieties SARS Coronavirus complete genome

	Total bp	a	c	g	t
Tor2NC	36	12	7	6	11
Tor2AY	36	12	7	6	11
TW1	58	34	7	6	11
Sin2679	76	40	9	8	19
Sin2677	82	43	10	8	21
Sin2774	76	40	9	8	19
Sin2500	76	40	9	8	19
Sin2748	81	41	11	8	21
CUHK-su10	51	16	8	8	19
Urbani	60	36	7	6	11
Frankfurt	60	36	7	6	11
HKU	45	21	7	6	11
BJ01	62	24	10	8	20
CUHK-w1	51	16	8	8	19
BJ04	55	19	9	8	19
BJ02	42	18	7	6	11
BJ03	47	21	7	6	13
ZJ01	71	39	8	8	16
GZ01	30	14	2	3	11

表 3 18 个变种 SARS 冠状病毒基因不同序列的数量和百分率

Table 3 Percentage and numbers of different sequences
in 18 varieties SARS Coronavirus complete genome

产地名 Localities	Ur- Bani	TW1	HKU	Sin 2774	Fran- kfurt	Sin 2677	Sin 2500	Sin 2748	ZJ 01	Sin 2679	Tor2 NC	Tor2 AY	CU HK- su10	BJ 02	BJ 03	BJ 04	BJ 01	CU HK- wl	CZ 01
Urbani		0.020	0.047	0.027	0.040	0.027	0.024	0.020	0.094	0.024	0.024	0.024	0.027	0.091	0.091	0.067	0.057	0.040	0.182
TW1	6		0.034	0.013	0.027	0.013	0.020	0.007	0.081	0.010	0.010	0.010	0.013	0.077	0.077	0.054	0.044	0.034	0.168
HKU	14	10		0.040	0.047	0.040	0.037	0.034	0.108	0.037	0.037	0.037	0.040	0.104	0.104	0.081	0.071	0.060	0.195
Sin2774	8	4	12		0.020	0.013	0.010	0.007	0.088	0.017	0.017	0.017	0.020	0.084	0.084	0.061	0.050	0.040	0.175
Frankfurt	12	8	14	6		0.027	0.024	0.020	0.101	0.030	0.030	0.030	0.034	0.098	0.098	0.074	0.064	0.054	0.188
Sin2677	8	4	12	4	8		0.010	0.007	0.088	0.017	0.017	0.017	0.020	0.084	0.084	0.061	0.051	0.040	0.175
Sin2500	7	3	11	3	7	3		0.003	0.084	0.013	0.013	0.013	0.017	0.081	0.081	0.057	0.047	0.037	0.172
Sin2748	6	2	10	2	6	2	1		0.081	0.010	0.010	0.010	0.013	0.077	0.077	0.054	0.044	0.034	0.168
ZJ01	26	24	32	26	30	26	25	24		0.084	0.084	0.084	0.088	0.151	0.151	0.128	0.118	0.108	0.242
Sin2679	7	3	11	5	9	5	4	3	25		0.013	0.013	0.017	0.081	0.081	0.057	0.047	0.037	0.171
Tor2NC	7	3	11	5	9	5	4	3	25	4		0	0.017	0.081	0.081	0.057	0.047	0.037	0.171
Tor2AY	7	3	11	5	9	5	4	3	25	4	0		0.017	0.081	0.081	0.057	0.047	0.037	0.171
CUHK- su10	8	4	12	6	10	6	5	4	26	5	5	5		0.084	0.084	0.061	0.050	0.034	0.175
BJ02	27	23	31	25	29	25	24	23	45	24	24	24	25		0.087	0.084	0.054	0.071	0.178
BJ03	27	23	31	25	29	25	24	23	45	24	24	24	25	26		0.084	0.054	0.071	0.192
BJ04	20	16	24	18	22	18	17	16	38	17	17	17	18	25	25		0.050	0.061	0.182
BJ01	17	13	21	15	19	15	14	13	35	14	14	14	15	16	16	15		0.037	0.151
CUHK-wl	12	10	18	12	16	12	11	10	32	11	11	11	10	21	21	18	11		0.155
CZ01	54	50	58	52	56	52	51	50	72	51	51	51	52	53	57	54	45	46	

据表 1-3 和图 1 看出非典型肺炎 SARS 冠状病毒 18 个变种间的变异和亲缘关系为:

1. 多伦多 2NC(Tor2NC) = 多伦多 2AY(Tor2AY)基因全序列完全相同, 应

为同一变种。

2. 非典型肺炎 SARS 冠状病毒基因全序列数不等,从 29705-29757 bp(丢失序列的数量为 30-82 bp),数量不同,是否代表种类差异?

3. 非典型肺炎 SARS 冠状病毒基因全序列应为 29787bp(a 8496, c 5949, g 6189, t 9153). 在 18 变种中, 29549 bp 完全相同, 占 99.201%, 238 bp 不同, 占 0.799%, 变种间不同数为 1-72 bp, 说明变异明显有异, 可能非同一起来源。

4. 台湾 1, 新加坡 2679, 多伦多 2 NC(Tor 2NC) = 多伦多 2 AY(Tor 2AY), 新加坡 2500, 新加坡 2748, 新加坡 2774, 新加坡 2677, 香港中大-su10(CUHK-su10)和乌尔巴尼(Urbani)等 9 个变种间不同 bp 数为 1-8, 它们亲缘关系很近, 在系统树的上方。

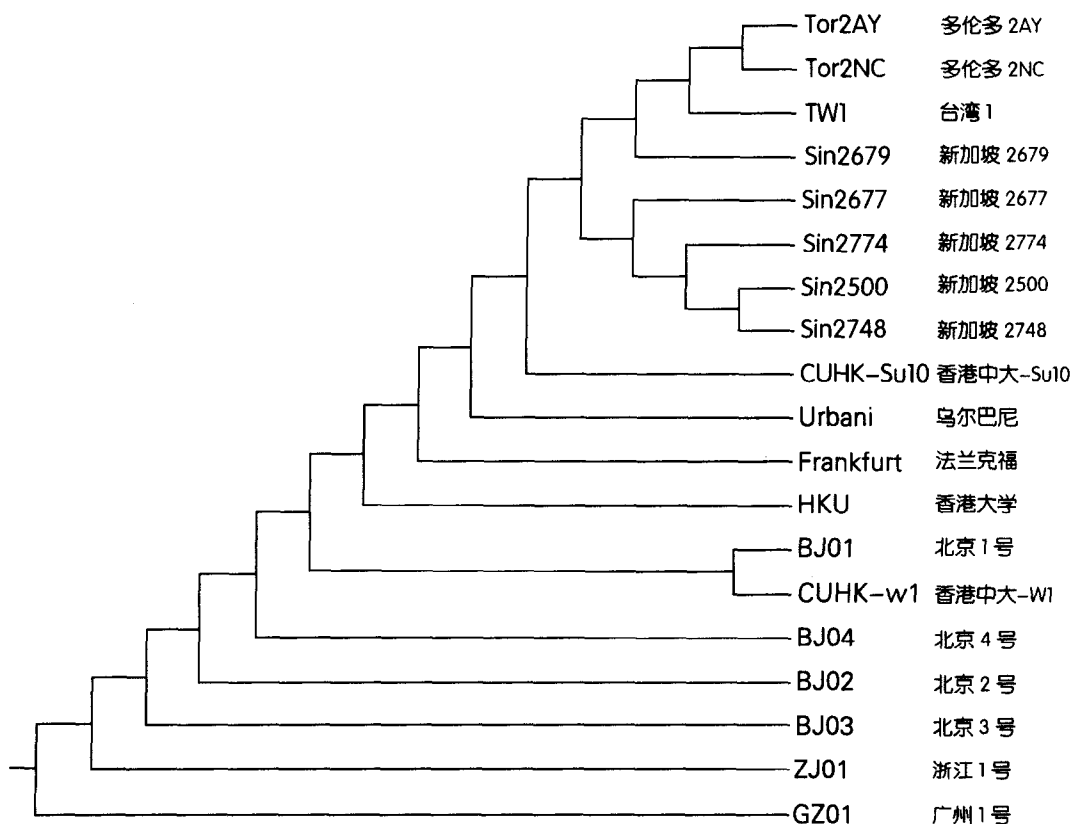


图 1 SARS 冠状病毒 18 个变种的分子系统树

Fig.1 Phylogenetic tree of 18 varieties of SARS Coronavirus

5. 法兰克福(Frankfurt)同乌尔巴尼(Urbani)间不同 bp 数为 12,同香港大学(HKU)间不同 bp 数为 14,故为独立一支。

6. 香港大学(HKU)同各变种间不同 bp 数为 10-58,差异较明显,居于系统树的中部。

7. 北京 1 号(BJ01)和香港中大-wl(CUHK-wl)间不同 bp 数为 11,而同其它变种间不同 bp 数为 13-45,相对较近,合为一支。

8. 北京 4 号(BJ04),北京 3 号(BJ03),北京 2 号(BJ02)和浙江 01(ZJ01)等 4 个变种同其它变种间不同 bp 数为 16-45,相对较远,均为独立一支。

9. 广州 1 号(GZ01)同其它变种间不同 bp 数为 45-72,亲缘关系最远,可视为最原始的变种,在系统树的底部。

Based on the Tab. 1-3 and Fig. 1, the different and relationships of 18 varieties of SARS are as below:

1. No.16 Tor 2 NC00 4718 = No.17 Tor 2 AY 274119 without different bp belong a variety.
2. SARS Coronavirus complete genome with 29705-29757 bp(lacking 30-82 bp), the different numbers behalf the characters of variety, or not?
3. SARS Coronavirus complete genome may has 29787 bp(a 8496, c 5949, g 6189, t 9153). In the 18 varieties, they are same 29549 bp, occupied 99.201%, different 238 bp, occupied 0.799%, and different 1-72 bp between varieties. It showed that the degeneration is very distinct and it may be not one source.

4. They are 9 varieties with different 1-8 bp and near relationships in the upper part of phylogenetic tree, including TW1, Sin2679, Tor2NC = Tor2AY, CUHK-sul0, Sin2500, Sin2748, Sin2774, Sin2677, CUHK-wl and Urbani.
5. Frankfurt has different 12 bp (with Urbani) and 14 bp (with HKU), so it is one dependent branch.
6. HKU has different 10-58 bp and differs others varieties distinctly, which placed in the middle part of phylogenetic tree.
7. BJ01 has different 13-45 bp (with others varieties) and 11 bp (with CUHK-wl), so BJ01 and CUHK-wl with comparatively near relationship grouped as one branch.
8. BJ04, BJ03, BJ02 and ZJ01 have different 16-45 bp (with others varieties) and far relationships, so they are being one dependent branch respectively.
9. GZ01 has different 45-72 bp (with others varieties) and most far relationships, so it may be one most original variety placed in the under part of phylogenetic tree.

3. 18 个变种病毒基因全序列比较

3. Multiple sequence alignment of 18 varieties SARS

Coronavirus complete genome

- 表示碱基缺失 indicates deletion * 表示碱基相同 indicates base same

CLUSTAL X (1.81) multiple sequence alignment

1 x 60bp		60bp
Fiction	ATATTAGGTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
Urbani	ATATTAGGTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
TW1	ATATTAGGTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
HKU	ATATTAGGTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
Sin2774	-----TACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
Frankfurt	ATATTAGGTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
Sin2677	-----TACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
Sin2500	-----TACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
Sin2748	-----TACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
ZJ01	-----CCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
Sin2679	-----TACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
Tor2AY	ATATTAGGTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
Tor2NC	ATATTAGGTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
CUHK-Su10	-----CTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
BJ02	ATATTAGGTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
BJ03	----TAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
BJ04	-----TACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
BJ01	-----CCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
CUHK-w1	-----CTACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	
GZ01	-----TACCCAGGAAAAGCCAACCAACCTCGATCTCTTGTAGATCTGTT	

2 x 60bp		120bp
Fiction	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
Urbani	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
TW1	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
HKU	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
Sin2774	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
Frankfurt	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
Sin2677	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
Sin2500	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
Sin2748	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
ZJ01	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
Sin2679	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
Tor2AY	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
Tor2NC	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
CUHK-Su10	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
BJ02	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
BJ03	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
BJ04	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
BJ01	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	
CUHK-w1	CTCTAAACGAACCTTTAAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC	

CTCTAAACGAACTTTAAATCTGTGTAGCTGTCGCTCGGCTGCATGCCTAGTGCACCTAC

[illegible][illegible]

300bp
GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA

TW1	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
HKU	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
Sin2774	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
Frankfurt	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
Sin2677	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
Sin2500	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
Sin2748	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
ZJ01	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
Sin2679	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
Tor2AY	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
Tor2NC	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
CUHK-Su10	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
BJ02	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
BJ03	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
BJ04	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
BJ01	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
CUHK-w1	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA
GZ01	GTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTTCTTGGTGTCAACGAGAAAAACA

6 x 60bp	360bp
Fiction	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
Urbani	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
TW1	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
HKU	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
Sin2774	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
Frankfurt	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
Sin2677	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
Sin2500	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
Sin2748	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
ZJ01	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
Sin2679	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
Tor2AY	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
Tor2NC	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
CUHK-Su10	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
BJ02	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
BJ03	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
BJ04	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
BJ01	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
CUHK-w1	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG
GZ01	CACGTCCAACCTCAGTTTGCCTGTCCTTCAGGTTAGAGACGTGCTAGTGCGTGGCTTCGGG

7 x 60bp	420bp
Fiction	GACTCTGTGGAAGAGGCCCTATCGGAGGCACGTGAACACCTCAAAAATGGCACTTGTGGT
Urbani	GACTCTGTGGAAGAGGCCCTATCGGAGGCACGTGAACACCTCAAAAATGGCACTTGTGGT
TW1	GACTCTGTGGAAGAGGCCCTATCGGAGGCACGTGAACACCTCAAAAATGGCACTTGTGGT
HKU	GACTCTGTGGAAGAGGCCCTATCGGAGGCACGTGAACACCTCAAAAATGGCACTTGTGGT
Sin2774	GACTCTGTGGAAGAGGCCCTATCGGAGGCACGTGAACACCTCAAAAATGGCACTTGTGGT
Frankfurt	GACTCTGTGGAAGAGGCCCTATCGGAGGCACGTGAACACCTCAAAAATGGCACTTGTGGT
Sin2677	GACTCTGTGGAAGAGGCCCTATCGGAGGCACGTGAACACCTCAAAAATGGCACTTGTGGT
Sin2500	GACTCTGTGGAAGAGGCCCTATCGGAGGCACGTGAACACCTCAAAAATGGCACTTGTGGT