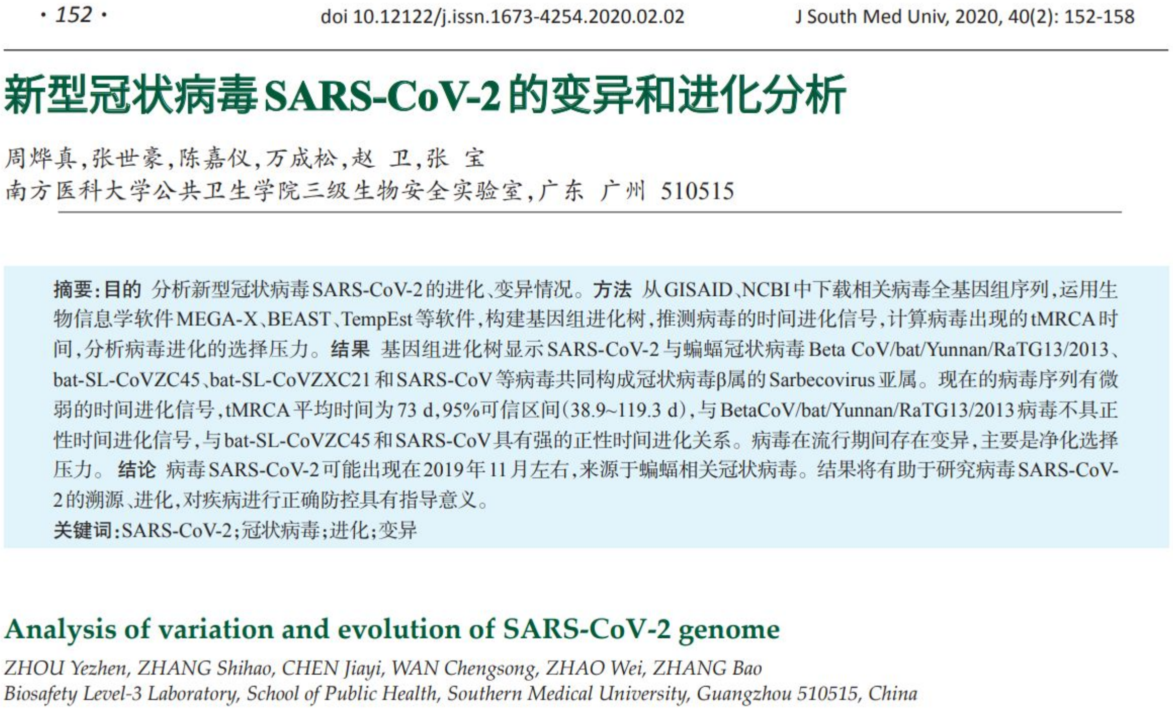

· 152 · doi 10.12122/j.issn.1673-4254.2020.02.02 J South Med Univ, 2020, 40(2): 152-158

新型冠状病毒SARS-CoV-2的变异和进化分析

周烨真,张世豪,陈嘉仪,万成松,赵 卫,张 宝
南方医科大学公共卫生学院三级生物安全实验室,广东 广州 510515

摘要:目的 分析新型冠状病毒SARS-CoV-2的进化、变异情况。**方法** 从GISAID、NCBI中下载相关病毒全基因组序列,运用生物信息学软件MEGA-X、BEAST、TempEst等软件,构建基因组进化树,推测病毒的时间进化信号,计算病毒出现的tMRCA时间,分析病毒进化的选择压力。**结果** 基因组进化树显示SARS-CoV-2与蝙蝠冠状病毒Beta CoV/bat/Yunnan/RaTG13/2013、bat-SL-CoVZC45、bat-SL-CoVZXC21和SARS-CoV等病毒共同构成冠状病毒β属的Sarbecovirus亚属。现在的病毒序列有微弱的时间进化信号,tMRCA平均时间为73 d,95%可信区间(38.9~119.3 d),与BetaCoV/bat/Yunnan/RaTG13/2013病毒不具正性时间进化信号,与bat-SL-CoVZC45和SARS-CoV具有强的正性时间进化关系。病毒在流行期间存在变异,主要是净化选择压力。**结论** 病毒SARS-CoV-2可能出现在2019年11月左右,来源于蝙蝠相关冠状病毒。结果将有助于研究病毒SARS-CoV-2的溯源、进化,对疾病进行正确防控具有指导意义。

关键词:SARS-CoV-2;冠状病毒;进化;变异

Analysis of variation and evolution of SARS-CoV-2 genome

ZHOU Yezhen, ZHANG Shihao, CHEN Jiayi, WAN Chengsong, ZHAO Wei, ZHANG Bao
Biosafety Level-3 Laboratory, School of Public Health, Southern Medical University, Guangzhou 510515, China

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7086142/pdf/nfykdxxb-40-202002152.pdf

Abstract: Objective To analyze the evolution and variation of SARS-CoV-2 during the epidemic starting at the end of 2019. Methods We downloaded the full-length genome sequence of SARS-CoV-2 from the databases of GISAID and NCBI. Using the software for bioinformatics including MEGA-X, BEAST, and TempEst, we constructed the genomic evolution tree, inferred the time evolution signal of the virus, calculated the tMRCA time of the virus and analyzed the selection pressure of the virus during evolution. Results The phylogenetic tree showed that SARS-CoV-2 belonged to the Sarbecovirus subgenus of β Coronavirus genus together with bat coronavirus BetaCoV/bat/Yunnan/RaTG13/2013, bat-SL-CoVZC45, bat-SL-CoVZXC21 and SARS-CoV. The genomic sequences of SARS-CoV-2 isolated from the ongoing epidemic showed a weak time evolution signal with an average tMRCA time of 73 days (95% CI: 38.9-119.3 days). No positive time evolution signal was found between SARS-CoV-2 and BetaCoV/bat/Yunnan/RaTG13/2013, but the former virus had a strong positive temporal evolution relationship with bat-SL-CoVZC45 and SARS-CoV. The major cause for mutations of SARS-CoV-2 was the pressure of purification selection during the epidemic. Conclusion SARS-CoV-2 may have emerged as early as November, 2019, originating most likely from bat-associated coronavirus. This finding may provide evidence for tracing the sources and evolution of the virus. Keywords: SARS-CoV-2; coronavirus; evolution; mutation

Since December 8, 2019, an outbreak of Breathing Dao is the main symptom of pneumonia.
After second-generation sequencing, virus isolation and identification, etc.
Paragraph, identified as a new type of coronavirus [1-2], The World Health Organization in 2020
It was temporarily named 2019 Novel Coronavirus (2019

Novel Coronavrius, 2019 nCoV) [3], On February 11, 2020
The formula is named SARS-CoV-2 virus, and the disease caused is called Coronavirus 2019 Viral disease (COVID-19) [4]. Until February 13, 2020, China's new crown
59,895 confirmed cases of pneumonia due to pneumonia, 16 067 suspected cases, and deaths 1367 cases; 492 cases outside China, spreading to 24 countries including Japan and Thailand [5] ;
With the adequate application of diagnostic reagents and the progress of the epidemic, the number of confirmed cases May increase further, seriously threatening people's lives and health Kang [6]. Therefore, on January 30, 2020, the World Health Organization
The document committee decided after discussion: the new coronavirus infection
Confirmed as "a public health emergency of international concern" [7]
To protect other Countries, especially to protect some countries with weaker medical systems, So that these countries can better prevent and control the pneumonia epidemic.
The disease became H1N1 in 2009, polio in 2014, Egypt in 2014
Bora, Zika in 2016, after the Ebola epidemic in 2019 [8], The sixth international
Public health emergencies of concern. It can be seen that the new type of coronavirus
The severity of the drug epidemic. As the epidemic progresses, Chinese scientists are
The characteristics of this article are well analyzed and described [9-10], For disease prevention and control Laid a solid foundation. The source of the virus is presumed to be bats as the virus The library spreads [11], As for whether there is an intermediate host, the possible intermediate host is

What has yet to be determined by further research. In the ongoing virus epidemic, there is
The important research content is: the origin time of the virus evolution process, the progress of the epidemic How is the mutation of the virus in the exhibition; this content is also useful for virus epidemic prevention and control It has important meaning. This article is based on the sequence submitted in GISAID (cut Until January 29, 2020) Analyze the relevant situation of virus evolution.

1 Materials and methods

1.1 Virus strain sequence

Download about SARS-CoV-2 from https://www.gisaid.org/
Sequence, a total of 45 full-length genome sequences. 39 of these sequences are used
Calculation of the time of the most recent common ancestor of the virus (tMRCA). Other coronary diseases
The poison sequence is downloaded from the NCBI sequence database.

1.2 Evolutionary tree construction, sequence similarity analysis software

MEGA-X is downloaded from https://www.megasoftware.net/;
tMRCA related analysis software combination BEAUti, BEAST,
TreeAnnotator is downloaded from http://www.beast2.org/, from http://
tree.bio.ed.ac.uk/software/Download evolutionary tree data display software
Tracer v1.6, evolutionary tree display software FigTree, evolutionary molecular clock detection Software TempEst.

1.3 Evolutionary tree construction method

Import the sequence into MEGA-X, the application software finds the best nucleoside

Acid substitute model parameters, and use this parameter to build evolutionary tree and tMRCA The calculation of the evolutionary tree uses the self-expanding method (Bootstrap=1000 repetitions); virus sampling time is imported into BEAUTi, with 2020 January 23 is regarded as 0 time, in days, with the virus strain BetaCoV/ Wuhan/IPBCAMS-WH-02/2019|EPI_ISL_403931 as Outgroup (outgroup, tree root), calculate tMRCA.

2 results

2.1 Basic information of genome sequence

Virus sequence is determined by second-generation sequencing or combined with third-generation sequencing.
The degree is between 29 688 and 29 899 bp [12]
, All covering the code of the virus
Area. Virus-encoded structural proteins S, E, M and N proteins, non-structural proteins White ORF1a, ORF1ab, ORF3, etc. are detailed in the literature and database Analysis of [13-14]
.

The nucleotide sequences of 39 SARS-CoV-2 full-length genomes were aligned Later, with SARS-CoV (SARS-CoV), Middle East Respiratory Syndrome (MERS CoV) sequence similarity averaged 78.7% and 48.7%, compared with the bat coronavirus strain bat-SL-CoVZC45 (abbreviated Write CoVZC45) and bat-SL-CoVZXC21 (abbreviated CoVZXC21) Closer, the similarity is 87.5% and 87.3%, which is similar to BetaCoV/bat/ Yunnan/RaTG13/2013 (abbreviated as RaTG13) is the closest, similarity Is 95.9%. There is a big difference between other coronaviruses, at 48.0%~ 87.4%, which also determines that they belong to different viruses in evolution Genus (Table 1).

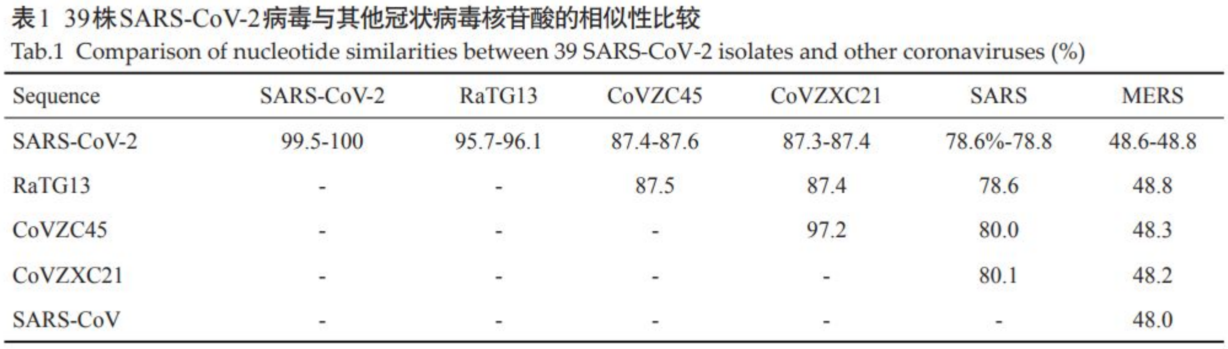

表1 39株SARS-CoV-2病毒与其他冠状病毒核苷酸的相似性比较

Tab.1 Comparison of nucleotide similarities between 39 SARS-CoV-2 isolates and other coronaviruses (%)

Sequence	SARS-CoV-2	RaTG13	CoVZC45	CoVZXC21	SARS	MERS
SARS-CoV-2	99.5-100	95.7-96.1	87.4-87.6	87.3-87.4	78.6%-78.8	48.6-48.8
RaTG13	-	-	87.5	87.4	78.6	48.8
CoVZC45	-	-	-	97.2	80.0	48.3
CoVZXC21	-	-	-	-	80.1	48.2
SARS-CoV	-	-	-	-	-	48.0

Table 1 Comparison of nucleotide similarities between 39 SARS-CoV-2 strains and other coronaviruses

Each coronavirus encodes the ammonia of the proteins ORF1ab, S, E, M and N The similarity of base acid is shown in Table 2~6. Select the sequence MN908947 as

For comparison with the reference strain, the result is the same as the nucleotide similarity of the genome
To. ORF1ab, S, M and N proteins and RaTG13, CoVZC45,
CoVZXC21 has the highest similarity, followed by SARS-CoV, and MERS
CoV has the lowest similarity, combined with the results of the subsequent evolutionary tree analysis, in the same subregion
There is a high similarity within the genera, and the similarity between different subgenus is low. value
It should be noted that E protein is highly conserved, SARS-CoV-2 and RaTG13,
CoVZC45 and CoVZXC21 are exactly the same, only 4 with SARS-CoV
The difference in amino acids.

2.2 Evolutionary analysis

Using MEGA-X software to find the best model of nucleotide substitution
Block, the minimum value of BIC (Bayesian Information Criterion) is the most
Good nucleotide substitution model parameters show that the TN93 model is the best. To
TN93 replacement model, NJ (Neighbor joint) method to construct whole genome
The tree is transformed, and the result is shown in Figure 1A below. According to the International Virus Classification Committee
The classification of the current coronavirus is Alpha, Beta, Gamma and
Delta [15]
. From Figure 1, we can see that SARS-CoV-2 and
BetaCoV/bat/Yunnan/RaTG13/2013, bat-SL-CoVZC45
And bat-SL-CoVZXC21 form a member of the sarbecovirus genus
And SARS-CoV and other viruses constitute another part of the sarbecovirus genus
Branches.

2.3 tMRCA analysis

Use MEGA-X software to construct evolutionary tree and import evolutionary time information
No. analysis software TempEst, analyze whether there is evolutionary time information
interest. As shown in Figure 2A, there is a weak forward evolution signal, and the clock is roughly estimated

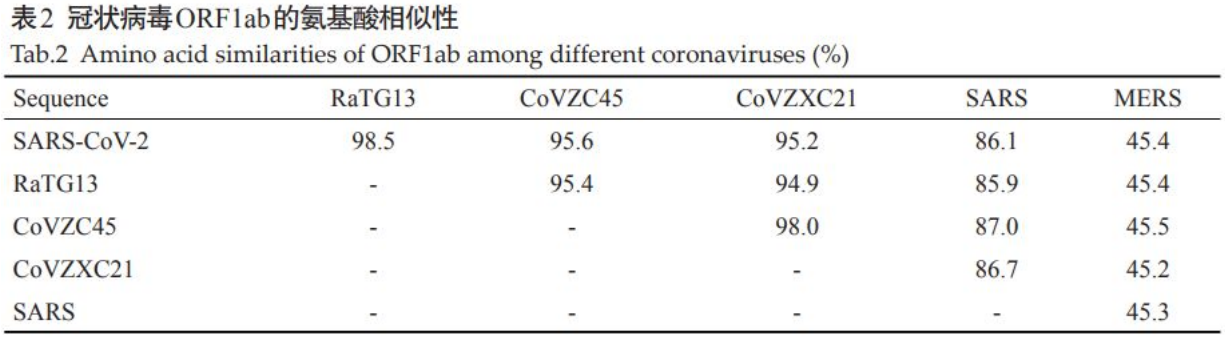

表2 冠状病毒ORF1ab的氨基酸相似性

Tab.2 Amino acid similarities of ORF1ab among different coronaviruses (%)

Sequence	RaTG13	CoVZC45	CoVZXC21	SARS	MERS
SARS-CoV-2	98.5	95.6	95.2	86.1	45.4
RaTG13	-	95.4	94.9	85.9	45.4
CoVZC45	-	-	98.0	87.0	45.5
CoVZXC21	-	-	-	86.7	45.2
SARS	-	-	-	-	45.3

Table 2 Amino acid similarity of coronavirus ORF1ab

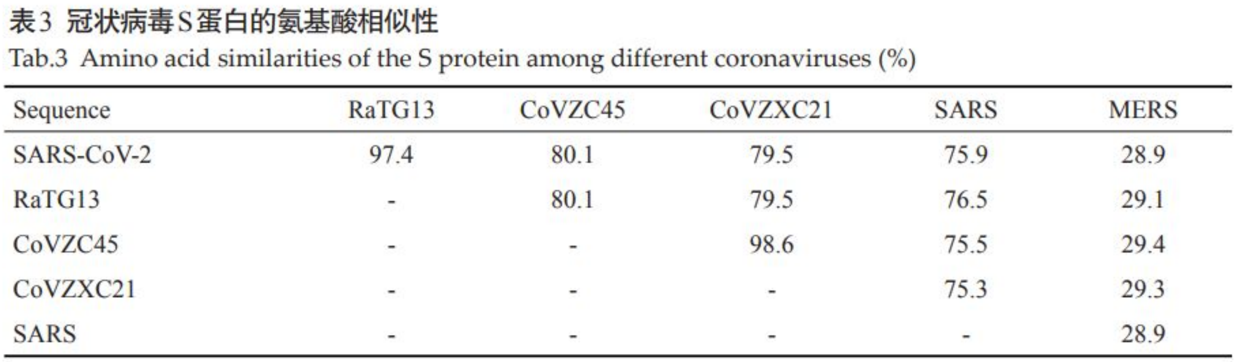

表3 冠状病毒S蛋白的氨基酸相似性

Tab.3 Amino acid similarities of the S protein among different coronaviruses (%)

Sequence	RaTG13	CoVZC45	CoVZXC21	SARS	MERS
SARS-CoV-2	97.4	80.1	79.5	75.9	28.9
RaTG13	-	80.1	79.5	76.5	29.1
CoVZC45	-	-	98.6	75.5	29.4
CoVZXC21	-	-	-	75.3	29.3
SARS	-	-	-	-	28.9

Table 3 Amino acid similarity of coronavirus S protein

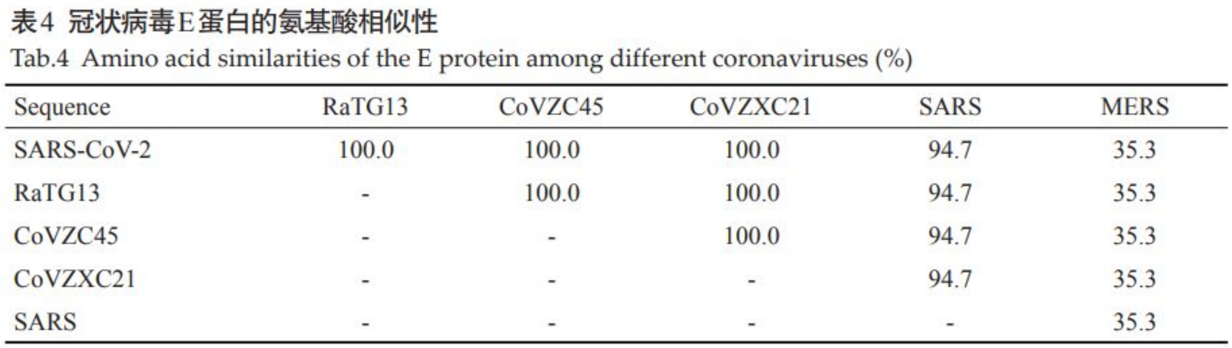

表4 冠状病毒E蛋白的氨基酸相似性

Tab.4 Amino acid similarities of the E protein among different coronaviruses (%)

Sequence	RaTG13	CoVZC45	CoVZXC21	SARS	MERS
SARS-CoV-2	100.0	100.0	100.0	94.7	35.3
RaTG13	-	100.0	100.0	94.7	35.3
CoVZC45	-	-	100.0	94.7	35.3
CoVZXC21	-	-	-	94.7	35.3
SARS	-	-	-	-	35.3

Table 4 Amino acid similarity of coronavirus E protein

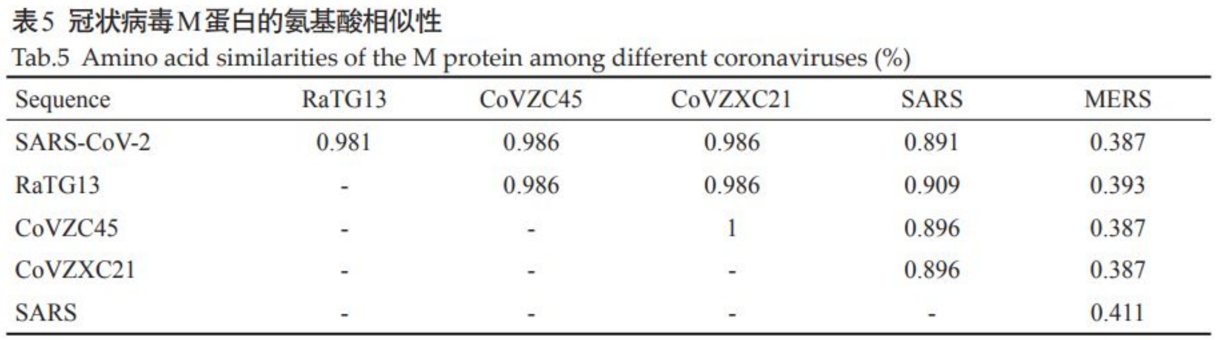

表5 冠状病毒M蛋白的氨基酸相似性

Tab.5 Amino acid similarities of the M protein among different coronaviruses (%)

Sequence	RaTG13	CoVZC45	CoVZXC21	SARS	MERS
SARS-CoV-2	0.981	0.986	0.986	0.891	0.387
RaTG13	-	0.986	0.986	0.909	0.393
CoVZC45	-	-	1	0.896	0.387
CoVZXC21	-	-	-	0.896	0.387
SARS	-	-	-	-	0.411

Table 5 Amino acid similarity of coronavirus M protein

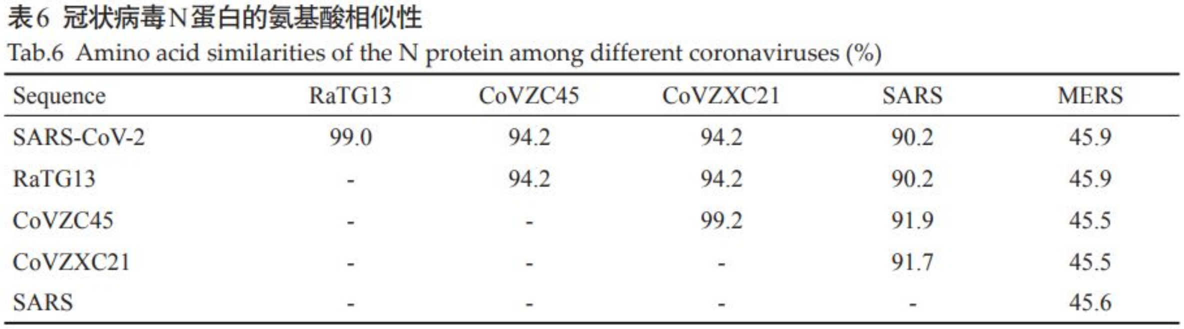

表6 冠状病毒N蛋白的氨基酸相似性

Tab.6 Amino acid similarities of the N protein among different coronaviruses (%)

Sequence	RaTG13	CoVZC45	CoVZXC21	SARS	MERS
SARS-CoV-2	99.0	94.2	94.2	90.2	45.9
RaTG13	-	94.2	94.2	90.2	45.9
CoVZC45	-	-	99.2	91.9	45.5
CoVZXC21	-	-	-	91.7	45.5
SARS	-	-	-	-	45.6

Table 6 Amino acid similarity of N protein of coronavirus

The rate is 2.24×10-6 substitution/site/day, tMRCA is -131 d, R2 =0.08. On this basis, we apply BEAST software for fine Assess tMRCA. As shown in Figure 2B, the average tMRCA time is 73.0 d (November 10, 2019), the 95% confidence interval is 38.9~119.3 d, after The test probability is 100%. The results suggest that the date of appearance of the virus is September 2019 Between 23rd and 15th December 2019. This is in line with the article reported in the literature The appearance of one case on December 1, 2019 is also consistent.

2.4 The relationship between SARS-CoV-2 and other coronaviruses

Use the evolution time signal analysis software TempEst to detect SARSCoV-2 and BetaCoV/bat/Yunnan/RaTG13/2013, bat-SL

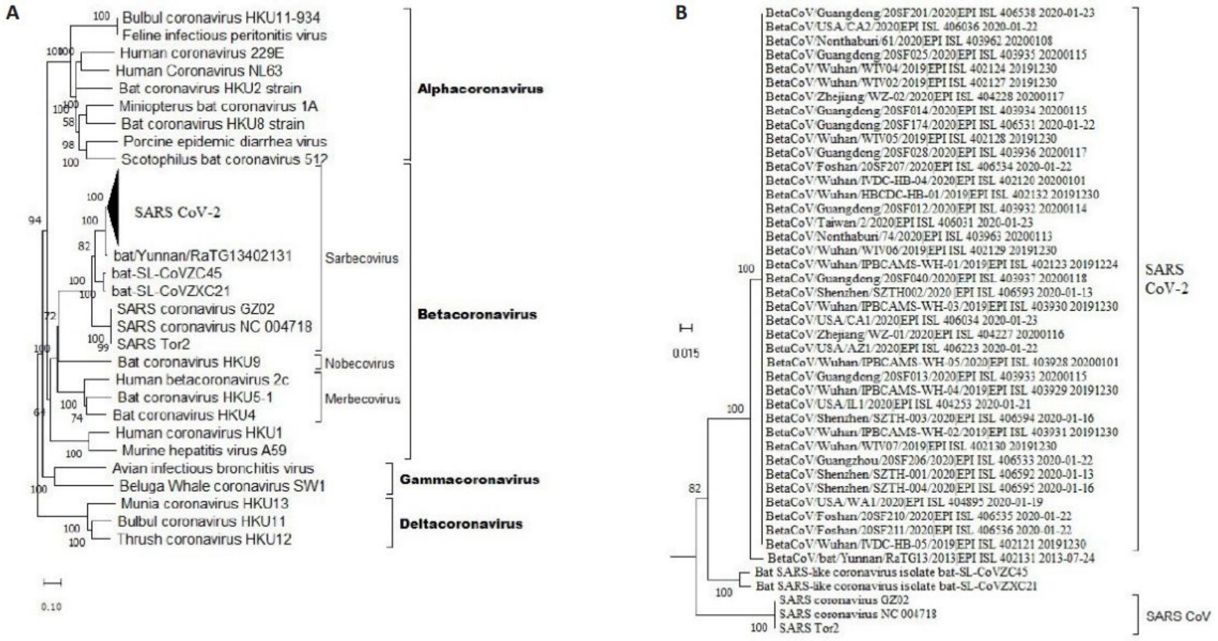
A
100
Bulbul coronavirus HKU11-934
Feline infectious peritonitis virus
100100
Human coronavirus 229E
Human Coronavirus NL63
Bat coronavirus HKU2 strain
100
100
Miniopterus bat coronavirus 1A
58
Bat coronavirus HKU8 strain
98
Porcine epidemic diarrhea virus
100
Scotophilus bat coronavirus 512
Alphacoronavirus
100
100
SARS CoV-2
82
94
bat/Yunnan/RaTG13402131
100
bat-SL-CoVZC45
100
bat-SL-CoVZXC21
Sarbecovirus
72
SARS coronavirus GZ02
SARS coronavirus NC 004718
Betacoronavirus
100
100
99
SARS Tor2
Bat coronavirus HKU9
Nobecovirus
Human betacoronavirus 2c
100
Bat coronavirus HKU5-1
Merbecovirus
74
Bat coronavirus HKU4
61
Human coronavirus HKU1
100
Murine hepatitis virus A59
Avian infectious bronchitis virus
100
Beluga Whale coronavirus SW1
Gammacoronavirus
Munia coronavirus HKU13
100
Bulbul coronavirus HKU11
100
Thrush coronavirus HKU12
Deltacoronavirus
0.10
B
BetaCoV/Guangdong/20SF201/2020|EPI ISL 406538 2020-01-23
BetaCoV/USA/CA2/2020|EPI ISL 406036 2020-01-22
BetaCoV/Nonthaburi/61/2020|EPI ISL 403962 20200108
BetaCoV/Guangdong/20SF025/2020|EPI ISL 403935 20200115
BetaCoV/Wuhan/WIV04/2019|EPI ISL 402124 20191230
BetaCoV/Wuhan/WIV02/2019|EPI ISL 402127 20191230
BetaCoV/Zhejiang/WZ-02/2020|EPI ISL 404228 20200117
BetaCoV/Guangdong/20SF014/2020|EPI ISL 403934 20200115
BetaCoV/Guangdong/20SF174/2020|EPI ISL 406531 2020-01-22
BetaCoV/Wuhan/WIV05/2019|EPI ISL 402128 20191230
BetaCoV/Guangdong/20SF028/2020|EPI ISL 403936 20200117
BetaCoV/Foshan/20SF207/2020|EPI ISL 406534 2020-01-22
BetaCoV/Wuhan/IVDC-HB-04/2020|EPI ISL 402120 20200101
BetaCoV/Wuhan/HBCDC-HB-01/2019|EPI ISL 402132 20191230
BetaCoV/Guangdong/20SF012/2020|EPI ISL 403932 20200114
BetaCoV/Taiwan/2/2020|EPI ISL 406031 2020-01-23
BetaCoV/Nonthaburi/74/2020|EPI ISL 403963 20200113
BetaCoV/Wuhan/WIV06/2019|EPI ISL 402129 20191230
100
BetaCoV/Wuhan/IPBCAMS-WH-01/2019|EPI ISL 402123 20191224
BetaCoV/Guangdong/20SF040/2020|EPI ISL 403937 20200118
BetaCoV/Shenzhen/SZTH002/2020 |EPI ISL 406593 2020-01-13
BetaCoV/Wuhan/IPBCAMS-WH-03/2019|EPI ISL 403930 20191230
BetaCoV/USA/CA1/2020|EPI ISL 406034 2020-01-23
BetaCoV/Zhejiang/WZ-01/2020|EPI ISL 404227 20200116
BetaCoV/USA/AZ1/2020|EPI ISL 406223 2020-01-22
BetaCoV/Wuhan/IPBCAMS-WH-05/2020|EPI ISL 403928 20200101
BetaCoV/Guangdong/20SF013/2020|EPI ISL 403933 20200115
BetaCoV/Wuhan/IPBCAMS-WH-04/2019|EPI ISL 403929 20191230
100
BetaCoV/USA/IL1/2020|EPI ISL 404253 2020-01-21
BetaCoV/Shenzhen/SZTH-003/2020|EPI ISL 406594 2020-01-16
BetaCoV/Wuhan/IPBCAMS-WH-02/2019|EPI ISL 403931 20191230
BetaCoV/Wuhan/WIV07/2019|EPI ISL 402130 20191230
BetaCoV/Guangzhou/20SF206/2020|EPI ISL 406533 2020-01-22
82
BetaCoV/Shenzhen/SZTH-001/2020|EPI ISL 406592 2020-01-13
BetaCoV/Shenzhen/SZTH-004/2020|EPI ISL 406595 2020-01-16
BetaCoV/USA/WA1/2020|EPI ISL 404895 2020-01-19
BetaCoV/Foshan/20SF210/2020|EPI ISL 406535 2020-01-22
BetaCoV/Foshan/20SF211/2020|EPI ISL 406536 2020-01-22
BetaCoV/Wuhan/IVDC-HB-05/2019|EPI ISL 402121 20191230
SARS CoV-2
0.015
BetaCoV/bat/Yunnan/RaTG13/2013|EPI ISL 402131 2013-07-24
Bat SARS-like coronavirus isolate bat-SL-CoVZC45
Bat SARS-like coronavirus isolate bat-SL-CoVZXC21
100
SARS coronavirus GZ02
SARS coronavirus NC 004718
100
SARS Tor2
SARS CoV

图1 SARS-CoV-2的全基因组进化分析

Fig.1 Phylogenetic analysis of the full-length genome of SRAS-CoV-2. **A**: Whole genome evolutionary tree of SARS-CoV-2 virus strains (compressed in triangle); **B**: Unfolded evolutionary tree of 38 isolates of SARS-CoV-2 virus.

Figure 1 The whole genome evolution analysis of SARS-CoV-2

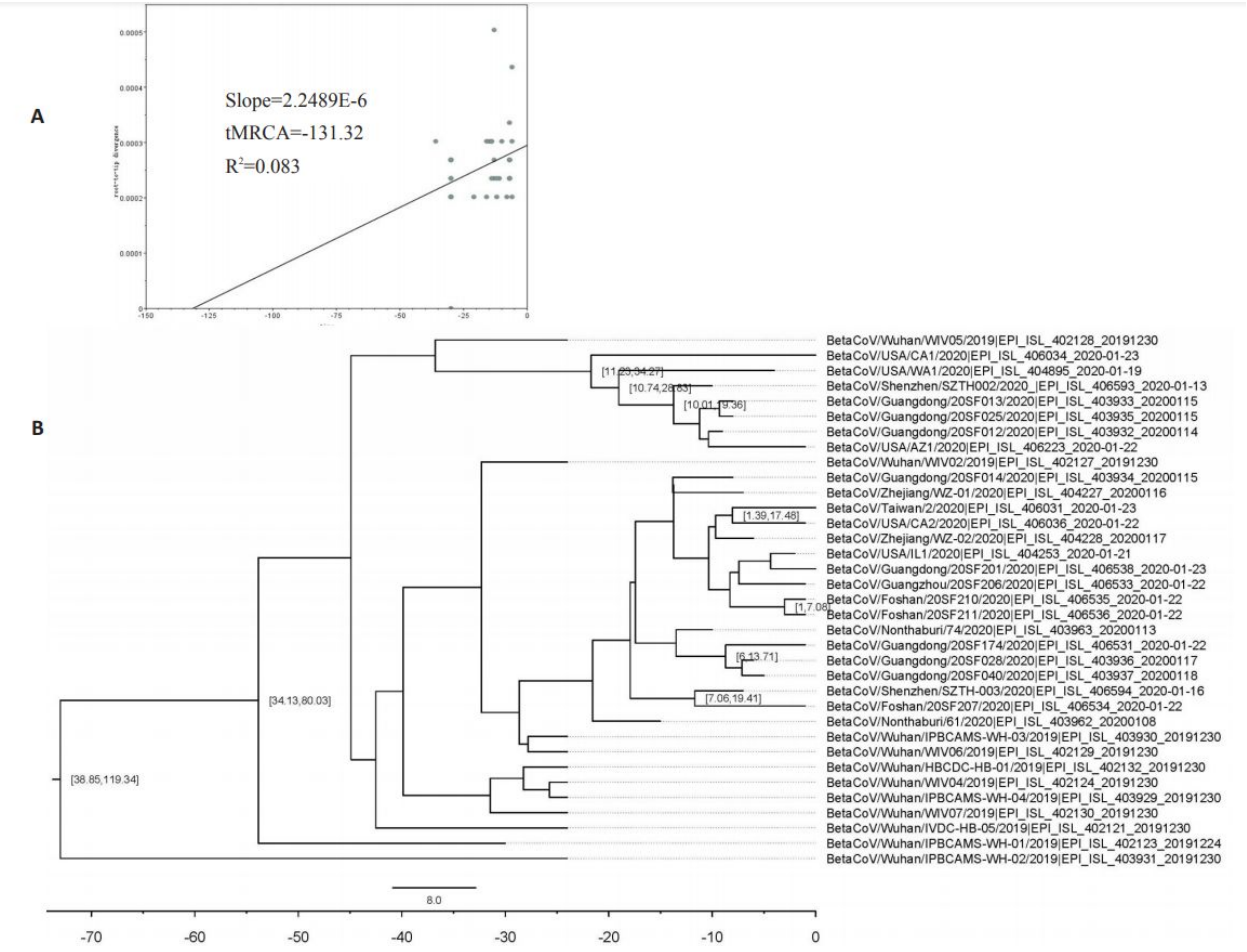

图2 SARS-CoV-2病毒的tMRCA计算

Fig.2 tMRCA of SARS-CoV-2. **A**: Calculation of evolutionary time signals of SARS-CoV-2 virus. The point indicates the genetic distance of each isolate of the virus from the reference strain (outgroup, tree root) on the time scale; **B**: Analysis of tMRCA of SARS-CoV-2 virus using BEAST.

Figure 2 Calculation of tMRCA of SARS-CoV-2 virus

Is there an evolutionary time gap between CoVZC45 and SARS-CoV?
system. The results show: SARS-CoV-2 and BetaCoV/bat/Yunnan/
RaTG13/2013 does not have a time evolution relationship, the slope is -1.8×10-5 substitution/site/day, R2=0.998 (Figure 3A), indicating that in nature Although it is difficult to achieve in the evolution, but with CoVZC45, SARS-CoV
There is indeed an obvious positive time evolution signal (Figure 3B, C).

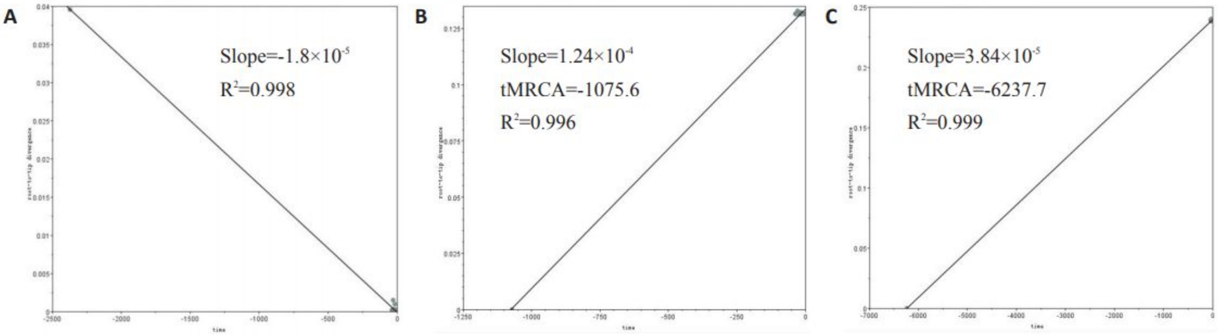

图3 SARS-CoV-2病毒与其他冠状病毒进化时间信号检测

Fig.3 Temporal signal test of molecular phylogenies of SARS-CoV-2 together with other coronaviruses. **A**, **B**, and **C**: Evolution time signal detection results of SARS-CoV-2 virus strain and RaTG13, CoVZC45, and SARS-CoV viruses, respectively.

Figure 3 Detection of the evolutionary time signal of SARS-CoV-2 virus and other coronaviruses

2.5 Variations in the evolution of SARS-CoV-2

We compare the sequences of 38 strains, the differences between nucleotides
Very small, in the sequence BetaCoV/Wuhan/
IPBCAMS- WH- 01/2019|EPI_ISL_402123 is the reference order
Column, there are 117 mutations in total, as shown in Figure 4, the mutation sites are not clustered, relatively evenly distributed throughout the genome. It can also be seen that there are 3 Variations at each locus (occurred on ORF1ab), stably expressed in other
Among the 36 strains of viruses, multiple sites are clustered in the sequence of the virus strain at the same time.

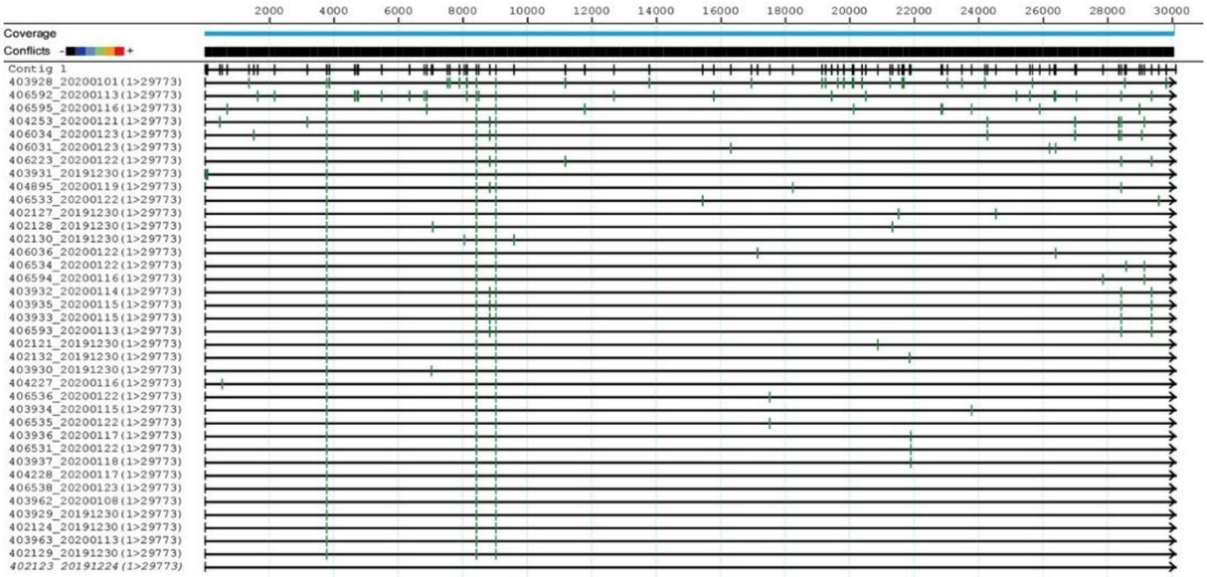

图4 38株病毒序列变异

Fig.4 Variation of 38 SARS-CoV-2 strains. Each green vertical line represents the the location of variation in the genome, using the earliest BetaCoV/Wuhan/IPBCAMS-WH-01/2019| EPI_ISL_402123 20191224 as the reference sequence.

Figure 4 Sequence variation of 38 strains of virus

We import the mutation sequence into http://www.datamonkey.org/
, Analyze whether there is selection pressure in its variation. As shown in Table 7, participation
The virus replication-related protein has 48 mutation sites, and the fixed effect is like
The Ranby Test Model (FEL) found that there are 4 purification options, and the
The relatively conservative method Single Likelihood Ancestor Count (SLAC) method can only detect
A site. Only FEL method can be used in S protein, M protein and N protein
Detect statistically significant 2~3 purification options, and detect in N protein
One kind of positive selection evolution.

3 Discussion

Since the beginning of December 2019, the emerging coronavirus epidemic in Wuhan
Poison [8]
, As of February 2020, it has triggered a total of 60,000 people in China so far

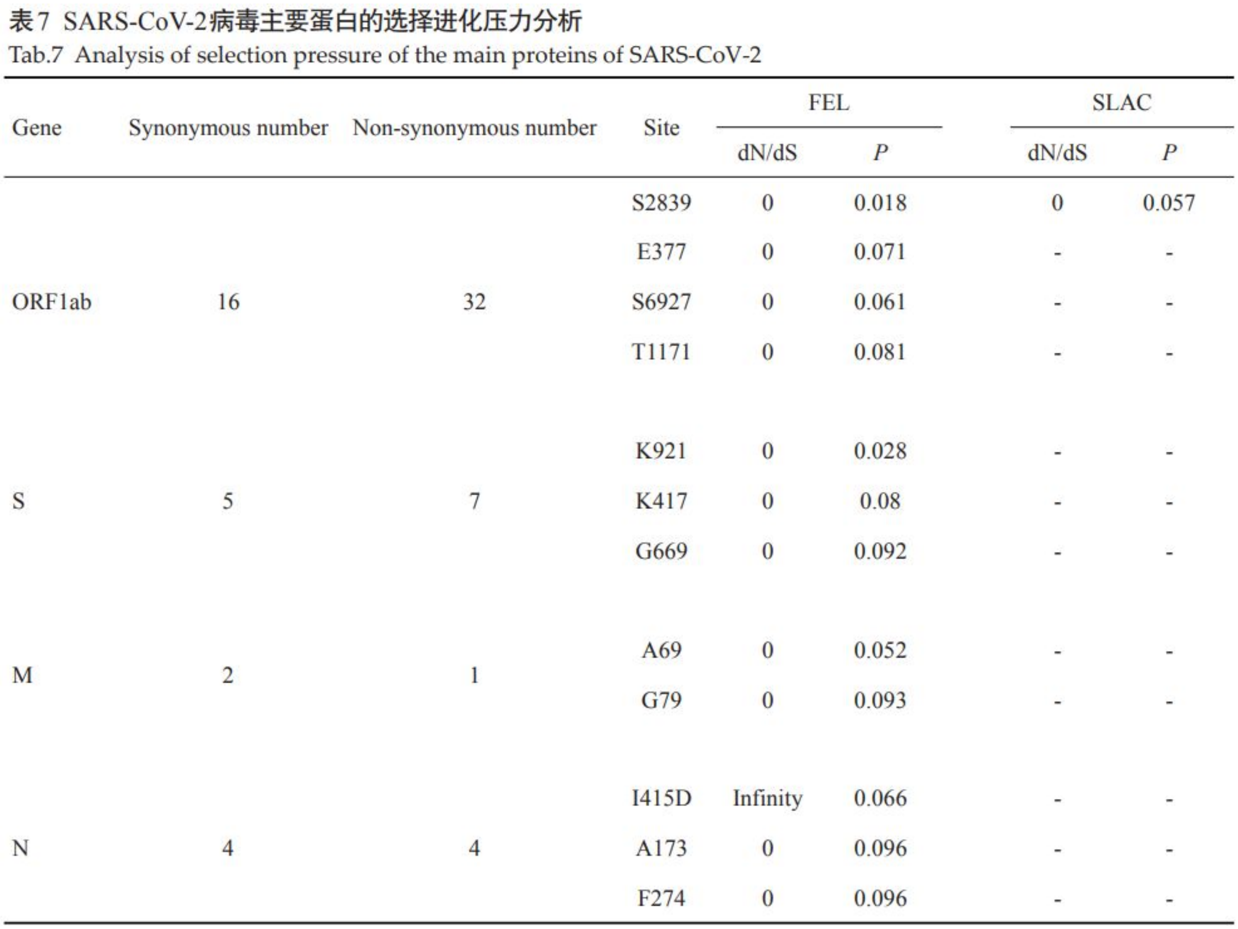

表7 SARS-CoV-2病毒主要蛋白的选择进化压力分析
Tab.7 Analysis of selection pressure of the main proteins of SARS-CoV-2

Gene	Synonymous number	Non-synonymous number	Site	FEL		SLAC	
				dN/dS	*P*	dN/dS	*P*
ORF1ab	16	32	S2839	0	0.018	0	0.057
			E377	0	0.071	-	-
			S6927	0	0.061	-	-
			T1171	0	0.081	-	-
S	5	7	K921	0	0.028	-	-
			K417	0	0.08	-	-
			G669	0	0.092	-	-
M	2	1	A69	0	0.052	-	-
			G79	0	0.093	-	-
N	4	4	I415D	Infinity	0.066	-	-
			A173	0	0.096	-	-
			F274	0	0.096	-	-

Table 7 Selection evolutionary pressure analysis of the main proteins of SARS-CoV-2 virus

Infection, more than 1,100 people died, and the harm to people’s health has far exceeded
The SARS epidemic around 2003［16-17］

. This virus has high similarity with SARSCoV virus, and belongs to β-corona like SARS-CoV. Sarbecovirus subgenus of the genus Sarbecovirus (Figure 1) and its impact on people's health A serious impact, the International Virus Nomenclature Committee named it SARSCoV-2, which is a sister virus to SARS-CoV [18-19]

. World Health Organization

The disease caused by the virus is called COVID-19. This is also currently issued

There is now the seventh coronavirus that can cause illness [20]

.

The virus has a stronger spreading power and longer duration than SARS-CoV

The incubation period is the basis for the virus to infect people extensively [21-22]

. Despite the virus

The genome was quickly resolved during the epidemic period [2]

, There are many literature reports

The characteristics of its genome [23-24]

, But these characteristics are still difficult to explain the strong

Communication power [8]

. Some other important factors affecting the spread of the virus are viruses

Traceability, including: the time when the virus appeared, whether there was an intermediate host, and

What is the source and whether the source of infection has always existed. According to the existing

The virus sequence information in the database, and the current traceability research, presumably

From the chrysanthemum-headed bat [25]

.

Regarding the possible time of the virus appearing, we use evolutionary analysis software

TempEst analyzes the current genomic data, and the results indicate that the score has appeared

Sub-clock event signal, which is the premise for tMRCA calculation later. here

Based on this, we use the BEAST software to estimate when the virus may appear

The period is 38.9~119.3 d (from January 23 forward), which is in 2019

Between September 23, 2019 and December 15, 2019. This also reminds the current Wu

One of the causes of high infection in Han City. We note that in this article about

The calculation of tMRCA may end because the time span is too short (only 30 days).

The result is not very accurate, but it also provides certain information.

Regarding the source of the virus, the current similarity to SARS-CoV-2 is compared

Gao's virus isolated from bats is RaTG13 (Yunnan, 2013),

bat- SL-CoVZC45 and bat- SL-CoVZXC21 (2017, Zhejiang

Jiang), the most similar is the RaTG13 strain. in case

SARS-CoV-2 is from the RaTG13 strain, so RaTG13

The virus should have a time evolution signal with SARS-CoV-2. We check

Measured the information on the time evolution of the two viruses, and found that there is a negative correlation (evolution speed

Rate is negative) (Figure 3A); if SARS-CoV-2 evolves from RaTG13
The virus should have a positive evolution rate, so we speculate: SARS-CoV-2
It is unlikely to be derived from the RaTG13 virus, although there are some differences between them
High similarity. We then calculated the evolutionary relationship between SARS-CoV-2 and bat-SLCoVZC45 and SARS-CoV and found that they have
There is an obvious positive correlation (Figure 3B, C). Prompt SARS-CoV-2
The production of bat-SL-CoVZC45 and other coronaviruses have a certain relationship system. This point still needs more experiments to support.
Does the virus adapt to external pressure during the epidemic?
Power is also an important factor affecting the ability of the virus to spread [26]
, From the clinical
Data show that the spread of the virus is indeed increasing [27]
. We pass the whole base
Due to the selection and evolutionary pressure analysis of the set of data, the results show that the virus exhibits a progressive
Chemical selection, multiple synonymous substitutions occur in highly conserved genes,
This substitution is much higher than the non-synonymous substitution, and the position of these substitutions may be
The vitality of poison has an important influence. Continue to strengthen these sites
Monitoring will help to analyze the strong transmission ability of SARS-CoV-2,
The prevention and control has a certain guiding role.