

## · 新型冠状病毒肺炎 ·

DOI: 10.13602/j.cnki.jcls.2020.07.03

## 新型冠状病毒刺突糖蛋白氨基酸序列特征分析\*

杨春伟<sup>1a</sup>, 孙海燕<sup>2</sup>, 崔大伟<sup>1b</sup>, 丁伟<sup>1a</sup> (1. 浙江大学医学院附属第一医院 a. 病理科, b. 输血科, 杭州 310003; 2. 绍兴第二医院检验科, 浙江绍兴 312000)

**摘要:**目的 探讨我国大陆新型冠状病毒(SARS-CoV-2)刺突糖蛋白(S)的分子特征,为特效药物和疫苗的研发提供参考依据。**方法** 从全球共享禽流感数据库(GISAID)和 GenBank 数据库收集 2019 年 12 月至 2020 年 3 月 27 日我国大陆 255 株新型冠状病毒(SARS-CoV-2)的 S 蛋白基因序列,利用 MEGA6.0 软件分析 S 蛋白氨基酸序列的分子变异特征。**结果** 255 株 SARS-CoV-2 的 S 蛋白氨基酸序列处于系统进化树的同一分支,同源率为 96.8%~100%;bat-coronavirus-RaTG13 病毒的 S 蛋白序列与 255 株 SARS-CoV-2 的氨基酸同源率为 86.6%~93.4%。S 蛋白 6 个关键位点和受体结合域(RBD)的 3 个关键位点均未发生变异,但 4 株病毒 S 蛋白的 RBD 域发生个别氨基酸突变。**结论** 中国大陆 SARS-CoV-2 S 蛋白的氨基酸序列同源性较高,但仍需密切监测其分子特征。

**关键词:**新型冠状病毒;刺突糖蛋白;受体结合域;分子特征

**中图分类号:**R446.5

**文献标志码:**A

新型冠状病毒又称为严重急性呼吸综合征冠状病毒 2 (SARS-CoV-2),导致人类严重呼吸综合征(称为 COVID-19)<sup>[1-3]</sup>。根据 WHO 疫情通报显示:截止 2020 年 3 月 31 日,我国累计确诊 82 554 病例,死亡 3 314 例,全球 205 个国家发生确诊病例,累计确诊 750 890 例,死亡 57 610 例([https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200331-sitrep-71-covid-19.pdf?sfvrsn=4360e92b\\_8](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200331-sitrep-71-covid-19.pdf?sfvrsn=4360e92b_8))。SARS-CoV-2 与 SARS-CoV 的基因组同源性约为 80%,与蝙蝠冠状病毒(SARS-CoV-RaTG13)的同源性约为 96%<sup>[4-6]</sup>。冠状病毒表面的刺突糖蛋白(spike glycoprotein, S)通过其受体结构域(RBD)识别并结合宿主细胞表面受体血管紧张素转换酶 2 (ACE2),在介导病毒包膜与细胞膜融合导致细胞感染的过程中发挥关键作用,这使 S 蛋白成为中和抗体和疫苗的主要抗原靶点<sup>[6-8]</sup>。S 蛋白的变异严重影响 SARS-CoV-2 对宿主的感染,因此探讨 S 蛋白的变异情况将有助于 SARS-CoV-2 流行的控制。

## 1 资料与方法

**1.1 资料** 从全球共享禽流感数据库(GISAID)和 GenBank 数据库收集 2019 年 12 月至 2020 年 3 月 27 日中国大陆 255 株 SARS-CoV-2 的 S 蛋白基因序列,同时下载 SARS-CoV、MERS、蝙蝠冠状病毒和穿山甲感染的冠状病毒序列共 12 个,作为参考株:Human SARS coronavirus BJ01-2003 (AY278488.2)、Hu-

man SARS coronavirus GZ02-2003 (AY390556.1)、MERS coronavirus (NC-019843.3)-2012、bat-SL-CoV-ZC45-2018 (MG772933.1)、bat-SL-CoVZXC21-2018 (MG772934.1)、bat-coronavirus-RaTG13-2020 (MN996532.1)、hCoV-19/pangolin/Guangxi/P5L/2017/EPI-ISL-410540、hCoV-19/pangolin/Guangxi/P5E/2017/EPI-ISL-410541、hCoV-19/pangolin/Guangxi/P4L/2017/EPI-ISL-410538、hCoV-19/pangolin/Guangxi/P2V/2017/EPI-ISL-410542、hCoV-19/pangolin/Guangxi/P1E/2017/EPI-ISL-410539、hCoV-19/pangolin/Guangdong/1/2019/EPI-ISL-410721。

**1.2 方法** 利用 MEGA 6.0 分子进化遗传分析软件,用邻接法(Neighbour-Joining, N-J)对下载的冠状病毒 S 氨基酸序列构建氨基酸种系分子进化树,自展值(Bootstrap)设置为 1 000,同时对 S 氨基酸序列的分子变异特征进行分析。

## 2 结果

**2.1 S 糖蛋白同源性和种系进化分析** 中国大陆 SARS-CoV-2 的 S 蛋白氨基酸系统进化树分析显示:从 GISAID 和 GenBank 数据库收集的 255 株 SARS-CoV-2 的 S 蛋白氨基酸序列相互间差异较小,处于同一分支,其同源率为 96.8%~100%。其中,hCoV-19/Shanghai/SH0104/2020/EPI-ISL-416395 与其余 254 株 SARS-CoV-2 的 S 蛋白序列的氨基酸同源性仅为 89.9%~93.7%;12 株参考株 S 蛋白氨基酸序列

\* 基金项目:国家自然科学基金(81871709)。

作者简介:杨春伟,1987 年生,男,技师,大学本科,从事临床病理分子检验工作。

通信作者:丁伟,副主任技师,E-mail:1181010@zju.edu.cn;崔大伟,副主任技师,E-mail:daweicui@zju.edu.cn。

的差异较大, 同源性的为 34.1% ~ 99.9%, 其中 MERS coronavirus (NC-019843.3) 与其他 11 株的同源性最差, 为 34.1% ~ 39.1%; 12 株参考株中, bat-coronavirus-RaTG13 (MN996532.1) 的 S 蛋白序列与 255 株 SARS-CoV-2 的氨基酸同源性最近, 为 86.6% ~ 93.4%, 而 bat-SL-CoVZC45 (MG772933.1) 和 bat-SL-CoVZXC21 (MG772934.1) 与 255 株 SARS-CoV-2 的 S 蛋白同源性分别为 70.3% ~ 75.6% 和 69.6% ~

75.1%; MERS coronavirus (NC-019843.3) 与 255 株 SARS-CoV-2 的氨基酸同源性最远 (34.9% ~ 39.1%); 2 株 Human SARS coronavirus BJ01 和 Human SARS coronavirus GZ02, 与 255 株 SARS-CoV-2 的 S 蛋白同源性均为 66.2% ~ 71.8%; 6 株穿山甲感染的冠状病毒与 255 株 SARS-CoV-2 的 S 蛋白同源性为 76.9% ~ 84.8%。见图 1。

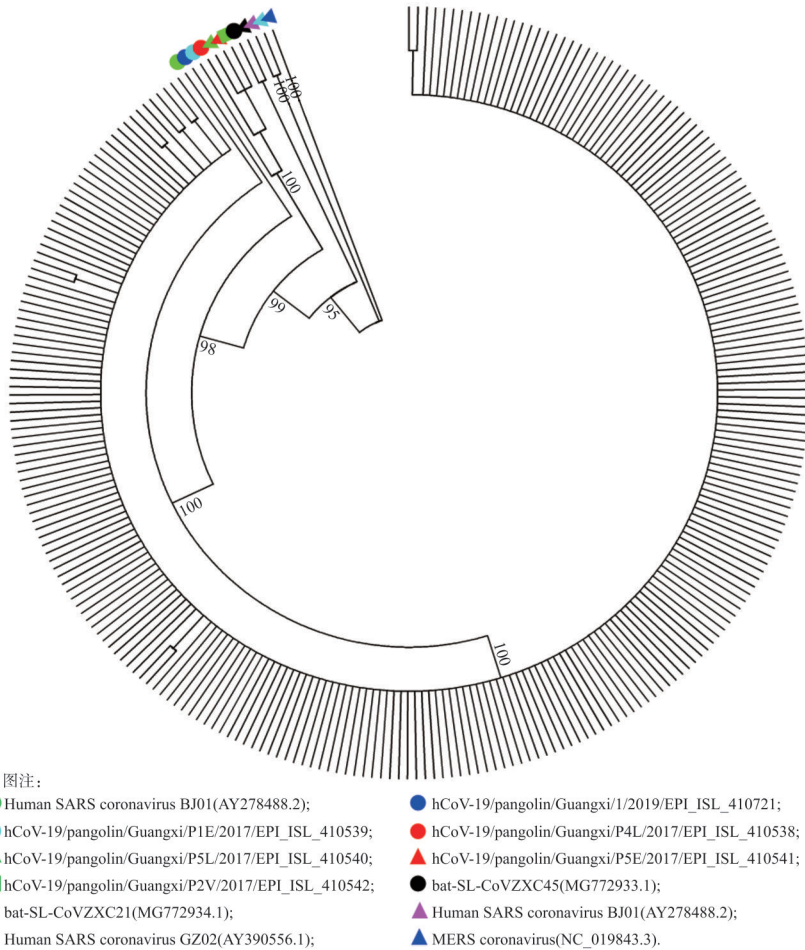


图 1 新型冠状病毒 (SARS-CoV-2) S 蛋白氨基酸种系进化树

**2.2 S 糖蛋白氨基酸位点变异分析** 255 株 SARS-CoV-2 病毒 S 蛋白的关键位点 L455、F486、Q493、S494、N501 和 Y505 均未发生变异, 且 S 蛋白 RBD 受体结合域关键位点 N439、V483 和 Q493 也均未发生变异; 255 株 SARS-CoV-2 病毒 S 蛋白的变异情况如下: 5 株发生 H49Y 突变, 3 株为 D614G 突变, 2 株为 M153T 突变, 2 株为 V1040F 突变, 2 株为 I1216T 突变, Y28D/N 突变各 1 株等。此外, 在 255 株病毒

的 RBD (306 ~ 543aa) 域中, 仅有 hCoV-19/Shenzhen/SZTH-004/2020/EPI\_ISL406595 发生 N354D 和 D364Y 突变, hCoV-19/Shanghai/SH0017/2020/EPI\_ISL416328 发生 S359N 突变, hCoV-19/Shanghai/SH0007/2020/EPI\_ISL416320 发生 K378R 突变, hCoV-19/Guangdong/GD2020087-P0008/2020/EPI\_ISL413863 发生 Q409E 突变。见表 1。

表 1 SARS-CoV-2 病毒 S 蛋白氨基酸变异分析

突变位点	突变氨基酸												
	D	N	I	Y	K	F	T	W	R	G	L	Q	E
Y28	1	1											
Y32			1										
H49				5									
H74					1								
S98						1							
S151							1						
M153								2					
S221									1				
N354	1												
S359		1											
D364				1									
K378									1				
Q409										1			
D614											3		
V615												1	
V860													1
K861												1	
E868				1									
V1040					2								
D1146													1
I1216							2						

注: D: 天冬氨酸; N: 天冬酰胺; I: 异亮氨酸; Y: 酪氨酸; K: 赖氨酸; F: 苯丙氨酸; T: 苏氨酸; W: 色氨酸; R: 精氨酸; G: 甘氨酸; L: 亮氨酸; Q: 谷氨酰胺; E: 谷氨酸。

### 3 讨论

冠状病毒科 (CoV) 是已知的最大的单股正链 RNA 病毒, 根据血清学与基因组的结构特点, 冠状病毒通常被分为  $\alpha$ 、 $\beta$ 、 $\gamma$  和  $\delta$  4 个属, SARS、SARS-CoV-2 和 MERS 均属于  $\beta$ -CoV 属<sup>[4-5]</sup>。目前, 对于 SARS-CoV-2 感染人群的治疗, 除了有效隔离和常规的对症治疗以外, 特效药物和疫苗的研发仍在进行中。SARS-CoV-2 的基因特征特别是 S 蛋白的分子特征对于药物和疫苗的研发至关重要<sup>[6-7]</sup>。本研究从 GISAID 和 GenBank 数据库下载 255 条 SARS-CoV-2 的 S 蛋白序列, 其由 1 273 个氨基酸组成。通过系统发育树显示发现, 255 株 SARS-CoV-2 的 S 蛋白处于同一分支, 且同源率为 96.8%~100%, 表明中国大陆 SARS-CoV-2 S 蛋白的氨基酸序列高度同源, 相互间变异较小, 与近期的研究结果相似<sup>[9-11]</sup>。初步表明在我国流行的 SARS-CoV-2 S 蛋白序列并未发生明显变异。进一步分析显示, 255 株 SARS-CoV-2 的 S 蛋白序列与参考株之间存在一定差异, 与 bat-coronavirus-RaTG13 (MN996532.1) 的 S 蛋白序列的氨基酸同源性最近且高达 93.4%, 与近期的研究结果相一致<sup>[10-13]</sup>。

研究表明, SARS-CoV-2 通过 S 蛋白受体结构域

RBD 与宿主细胞表面受体结合, 介导病毒包膜与细胞膜融合, 促进病毒的侵入, 而 RBD 域是 SARS 及其相关冠状病毒基因组中最容易变异的区域, 特别是关键氨基酸对与人体 ACE2 受体结合及确定宿主类型至关重要<sup>[6-8]</sup>。本研究显示, S 蛋白的关键位点 L455、F486、Q493、S494、N501 和 Y505 均未发生变异, 且 S 蛋白 RBD 受体结合域中的关键位点 N439、V483 和 Q493 也均未发生变异, 提示 SARS-CoV-2 容易通过 RBD 域与人体 ACE2 结合, 导致病毒的感染与流行, 这与之前报道的结果相一致<sup>[10-11]</sup>。此外, 我们的研究还发现, S 蛋白中某些位点氨基酸发生变异, 如 5 株发生 H49Y 突变, 3 株为 D614G 突变等, 重要的是部分 SARS-CoV-2 的 S 蛋白区 RBD 域发生变异 (如 N354D、D364Y、S359N、K378R 和 Q409E), 但这些变异是否影响 S 蛋白及 RBD 域的结构和功能仍需进一步深入研究。

总之, 本研究初步探讨了在我国暴发流行的 SARS-CoV-2 病毒 S 蛋白区的分子特征, 显示该病毒的 S 蛋白并未发生明显的变异, 这为针对 SARS-CoV-2 病毒的药物研发、特异性抗体的制备及疫苗的设计提供参考依据, 且 S 蛋白区分子特征的持续监测有助于 SARS-CoV-2 病毒的预防和控制。

### 4 参考文献

- [1] Huang C, Wang Y, Li X, *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China [J]. *Lancet*, 2020, 395(10223): 497-506.
- [2] Zhu N, Zhang D, Wang W, *et al.* A novel coronavirus from patients with pneumonia in China, 2019 [J]. *N Engl J Med*, 2020, 382(8): 727-733.
- [3] Chan JF, Yuan S, Kok KH, *et al.* A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster [J]. *Lancet*, 2020, 395(10223): 514-523.
- [4] Lu R, Zhao X, Li J, *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding [J]. *Lancet*, 2020, 395(10224): 565-574.
- [5] Zhou P, Yang XL, Wang XG, *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin [J]. *Nature*, 2020, 579(7798): 270-273.
- [6] Yan R, Zhang Y, Li Y, *et al.* Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2 [J]. *Science*, 2020, 367(6485): 1444-1448.
- [7] Walls AC, Park YJ, Tortorici MA, *et al.* Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein [J]. *Cell*, 2020, 181(2): 281-292, e6.