

我国 2020 年下半年 SARS-CoV-2 病毒株的基因突变分析

孙慧敏^{1*}, 唐文彬², 张磊³, 李济伟⁴, 叶绍培²

(1. 厦门大学附属翔安医院中心实验室, 福建 厦门 361102; 2. 厦门大学附属翔安医院泌尿外科, 福建 厦门 361102; 3. 空军军医大学预防医学系, 陕西 西安 710032; 4. 厦门大学附属翔安医院呼吸和危重症及睡眠医学科, 福建 厦门 361102)

摘要: 基于 2020 年 6 月 1 日—12 月 1 日之间 GISAID 网站 (<http://platform.gisaid.org>) 上公布来自于中国的全部 21 条病毒基因组序列, 以 2019 年 12 月公布的新新型冠状病毒 (SARS-CoV-2) 参考株 NC_045512 和 MN996528 进行各基因组错义突变位点分析, 计算突变频率 (截止 2020 年 12 月), 寻找各突变最早出现的时间和地点, 网站内搜索相似病毒株。结果显示: 我国 2020 年下半年公布的北京市、浙江省、辽宁省大连市、云南省、山东省青岛市五地全部 21 个病毒株中共有 18 处错义突变, 集中分布于 *Spike (S)*、*ORF1AB* 和 *Nucleocapsid (N)* 基因。*S* 基因的 D614G 突变和 *NSP12 (RdRp)* 的 P323L 突变具有最高的突变频率; 除北京病毒株中发现一个新的突变位点外, 其他位点突变均主要出现于欧洲或北美洲。根据突变位点分类, 病毒株均属于欧洲的 G 支系, 其中云南省的病毒株属于 GH 支系, 与泰国等亚洲病毒株相似; 山东省青岛市、辽宁省大连市、浙江省、北京市的病毒株属于 GR 支系, 与来自于欧洲的病毒株相似。综上, 2020 年下半年我国新冠病毒毒株主要为外源输入病毒株, 在新型冠状病毒肺炎 (COVID-19) 防控策略上要在各运输环节严防外源输入; 病毒新突变在严格的防控措施下未发生扩散, 进一步说明我国采取的防控措施非常有效。

关键词: COVID-19; SARS-CoV-2; 基因突变

中图分类号: R 183; R 184 **文献标志码:** A

2020 年上半年中国一直处于严格防控新型冠状病毒肺炎 (COVID-19) 阶段, 在经历了一段时间的本土零病例以后, 社会各行各业开始有计划地慢慢恢复运行。2020 年 6 月 1 日—12 月 1 日半年内, 我国只有零星散在的疫情出现, 多数为输入病例, 也出现了几处小范围疫情, 其中北京市、浙江省、辽宁省大连市、云南省、山东省青岛市先后出现新型冠状病毒

收稿日期: 2020-10-13 **录用日期:** 2021-01-16

基金项目: 厦门大学新型冠状病毒防治应急科研攻关项目 (20720200047, 20720200032, 20720200017)

***通信作者:** hmsun@xah.xmu.edu.cn

毒 (SARS-CoV-2) 感染确诊病例, 或者在市场的环境以及进口冷冻食品中发现新冠病毒, 但由于各地均及时采取了严格的防控措施, 疫情均被迅速控制。

了解病毒基因组的位点突变及其突变频率, 研究突变位点在病毒进化及其与宿主相互作用的关系将对感染病例的治疗、疫苗和药物的研发以及疫情发展的判断有重要作用^[1]。此外病毒位点突变也可以用来追踪病毒的来源, 根据基因序列的相似性可以初步判定病毒来源及传播路径, 为流行病学调查过程中寻找传染源和切断传播途径提供重要依据^[2]。

1 资料与方法

1.1 资料来源

2020 年下半年 (2020 年 6 月 1 日—12 月 1 日) 中国五地 (云南($n=2$)、浙江($n=1$)、大连($n=4$)、青岛($n=11$)和北京($n=3$)) 提交给 GISAID (Global Initiative on Sharing All Influenza Data) 网站 (<http://platform.gisaid.org>) 的全部 21 条 SARS-CoV-2 病毒株序列以及网站中世界各地的全部 SARS-CoV-2 序列和相关数据。

1.2 分析方法

将 GISAID 网站上获得的病毒株各项信息, 以 NC_045512 和 hCoV-19/Wuhan/WIV04/2019 (EPI_ISL_402124, 2019 年 12 月 30 日采自湖北武汉金银潭医院感染者) 病毒株作为参考株^[3], 以 Staden/Spin 和 Lasergene 7 软件对病毒各基因进行注释, 以 CLUSTAL W 软件对病毒基因组进行比对, 对各基因突变位点进行统计, 计算突变频率 (截止 2020 年 12 月), 寻找各突变位点最早出现的时间和地点, 并在全部已知序列中寻找与各病毒株相似性最高的序列。

2 结果与分析

2.1 病毒株的基因组序列突变分析

病毒株的选择: GISAID 网站中获取 2020 年 6 月 1 日—12 月 1 日间采集的全部 SARS-CoV-2 病毒株共计 21 株, 提取各序列的相关信息如表 1 所示。

表 1 2020 年 6 月 1 日—12 月 1 日间中国采集并公布的全部 SARS-CoV-2 病毒株

Tab. 1 All SARS-CoV-2 strains collected in China from June 1 to December 1, 2020

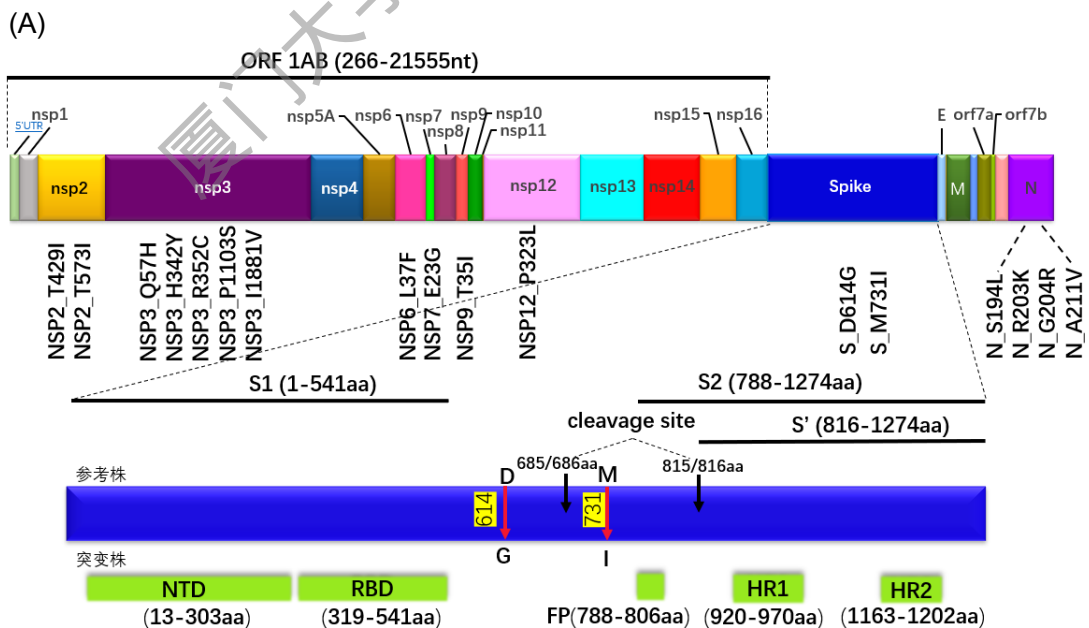
获取号	支系	亚支系	错义突变位点	采集地点	采集日期	样本来源
EPI_ISL_632935	GH	B.1.36	S D614G, N S194L, NS3 Q57H, NSP6 L37F,	云南	2020-09-13	患者

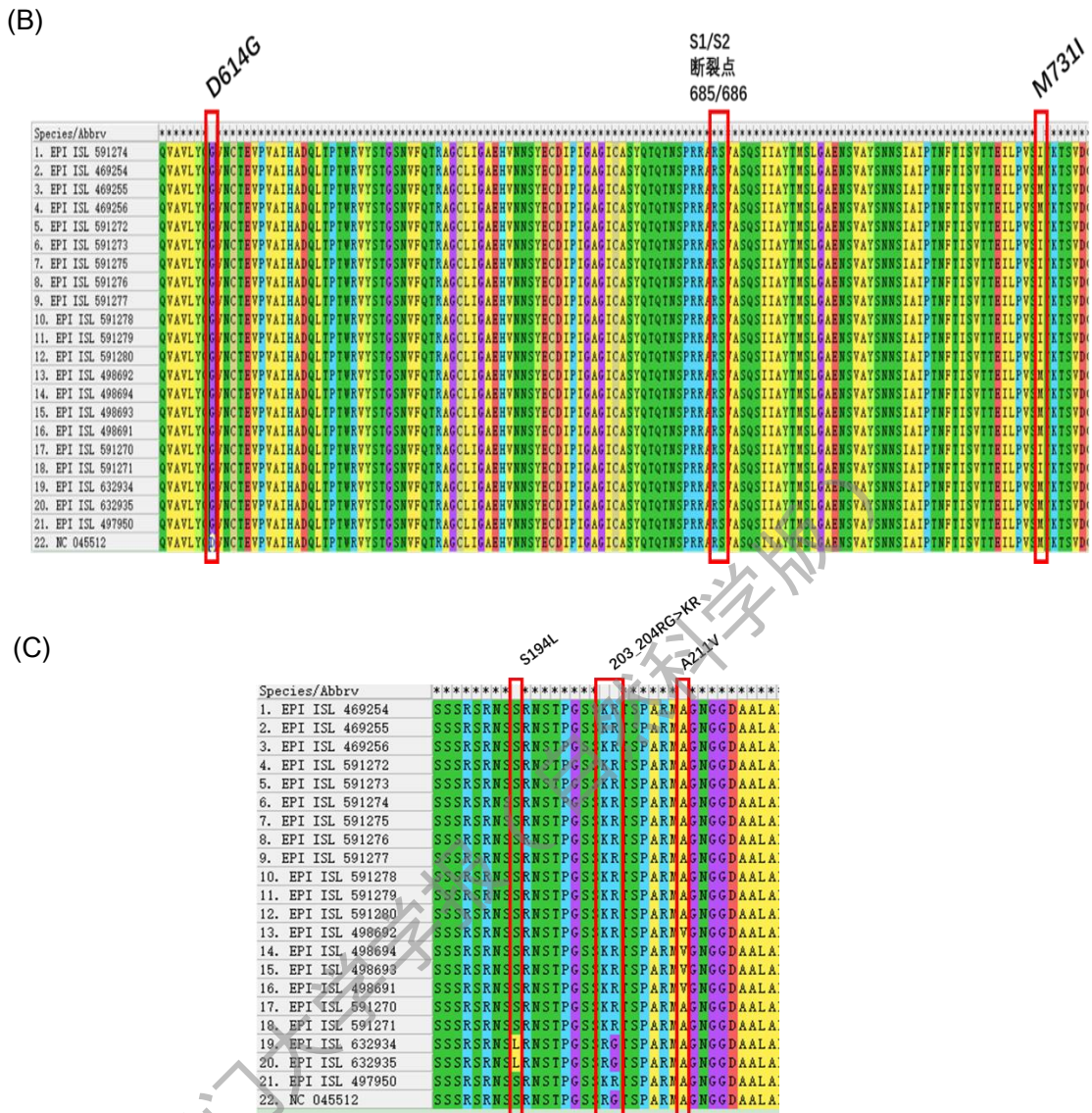
			NSP12 P323L				
EPI_ISL_632934	GH	B.1.36	S D614G, N S194L, NS3 Q57H, NSP6 L37F,	云南	2020-09-13	患者	
			NSP12 P323L				
EPI_ISL_591280	GR	B.1.1	S D614G, S M731I, N G204R, N R203K, NSP3	青岛	2020-10-07	环境	
			H342Y, NSP3 R352C, NSP12 P323L				
EPI_ISL_591279	GR	B.1.1	S D614G, S M731I, N G204R, N R203K, NSP3	青岛	2020-09-27	环境	
			H342Y, NSP3 R352C, NSP12 P323L				
EPI_ISL_591278	GR	B.1.1	S D614G, S M731I, N G204R, N R203K, NSP3	青岛	2020-09-24	环境	
			H342Y, NSP3 R352C, NSP12 P323L				
EPI_ISL_591277	GR	B.1.1	S D614G, S M731I, N G204R, N R203K, NSP3	青岛	2020-09-24	环境	
			H342Y, NSP3 R352C, NSP12 P323L				
EPI_ISL_591276	GR	B.1.1	S D614G, S M731I, N G204R, N R203K, NSP3	青岛	2020-09-24	环境	
			H342Y, NSP3 R352C, NSP12 P323L				
EPI_ISL_591275	GR	B.1.1	S D614G, S M731I, N G204R, N R203K, NSP3	青岛	2020-09-24	环境	
			H342Y, NSP3 R352C, NSP12 P323L				
EPI_ISL_591274	GR	B.1.1	S D614G, S M731I, N G204R, N R203K, NSP3	青岛	2020-09-24	环境	
			H342Y, NSP3 R352C, NSP12 P323L				
EPI_ISL_591273	GR	B.1.1	S D614G, S M731I, N G204R, N R203K, NSP3	青岛	2020-09-24	环境	
			H342Y, NSP3 R352C, NSP12 P323L				
EPI_ISL_591272	GR	B.1.1	S D614G, S M731I, N G204R, N R203K, NSP3	青岛	2020-09-24	环境	
			H342Y, NSP3 R352C, NSP12 P323L				
EPI_ISL_591271	GR	B.1.1	S D614G, S M731I, N G204R, N R203K, NSP2	青岛	2020-09-24	环境	
			T573I, NSP3 H342Y, NSP3 R352C, NSP12				
			P323L				
EPI_ISL_591270	GR	B.1.1	S D614G, S M731I, N G204R, N R203K, NSP3	青岛	2020-09-24	环境	
			H342Y, NSP3 R352C, NSP12 P323L				
EPI_ISL_498694	GR	B.1.1	S D614G, N A211V, N G204R, N R203K,	大连	2020-07-22	患者	
			NSP2 T429I, NSP3 I1881V, NSP12 P323L,				
			NSP15 V127F				
EPI_ISL_498693	GR	B.1.1	S D614G, N A211V, N G204R, N R203K,	大连	2020-07-22	患者	
			NSP2 T429I, NSP3 I1881V, NSP12 P323L,				
			NSP15 V127F				
EPI_ISL_498692	GR	B.1.1	S D614G, N A211V, N G204R, N R203K,	大连	2020-07-22	患者	
			NSP2 T429I, NSP3 I1881V, NSP12 P323L,				
			NSP15 V127F				
EPI_ISL_498691	GR	B.1.1	S D614G, N A211V, N G204R, N R203K,	大连	2020-07-22	患者	
			NSP2 T429I, NSP3 I1881V, NSP12 P323L,				
			NSP15 V127F				
EPI_ISL_497950	GR	B.1.1	S D614G, N G204R, N R203K, NSP9 T35I,	浙江	2020-07-15	患者	
			NSP12 P323L				
EPI_ISL_469256	GR	B.1.1	S D614G, N G204R, N R203K, NSP3 P1103S,	北京	2020-06-11	环境	
			NSP7 E23G, NSP12 P323L				
EPI_ISL_469255	GR	B.1.1	S D614G, N G204R, N R203K, NSP3 P1103S,	北京	2020-06-11	环境	
			NSP12 P323L				

21 株病毒株中，青岛和北京的病毒株来自于环境，云南、浙江和大连的病毒株来自于患者，根据病毒株突变情况进行分类，这些病毒株均具有 *Spike* 基因的 D614G 突变和 *RdRp* (*NSP12*) 的 P323L 双突变，按照 GISAID 分类均属于 G 支系^[3]。除云南病毒株外均具有 N 基因的 203_204RG>KR 突变，属于 GR, B.1.1 亚支系，云南病毒株具有 NSP3 Q57H 突变，属于 GH, B.1.36 亚支系^[4]。

所有 21 株病毒株中的突变主要分布在 *ORF1AB* 基因、*Spike* (*S*) 基因以及 *N* 基因，*E*、*M* 基因则没有积累任何突变。在 *ORF1AB* 中的 NSP2、NSP3、NSP6、NSP7、NSP9、NSP12 和 NSP15 中均有数量不等的突变发生：其中 NSP3 突变数量最多，有 5 个位点；NSP6、NSP7、NSP9、NSP12 和 NSP15 则突变较少，分别只有 1 个突变位点。NSP1、NSP4、NSP5、NSP8、NSP10、NSP11、NSP13、NSP14 和 NSP16 则没有发生突变，为 *ORF1AB* 中保守区域。*N* 基因除了典型的 203_204 双突变外，在大连和云南病毒株中分别有一个罕见的 A211V 和 S194L 突变（表 1 和图 1(A)）。

S 基因上 2 个突变位点 D614G、M713I 均位于 *S1* 和 *S2* 亚基中间，在 2 个亚基断裂点附近（图 1(A)和(B)）。*N* 基因的突变位点集中于氨基酸序列 190~210 之间的 20 个氨基酸段（图 1(C)）。全部 18 个位点突变中 2/3 的发生于 *orf1ab* 的各个非结构蛋白的编码基因中，NSP2 的突变位点为 429 和 573，均由 T 突变为 I，分别只有一个突变（图 1(A)）。





nt. 核苷酸; aa. 氨基酸; NTD. N 末端结构域; RBD. 受体结合域; FP. 融合肽; HR. 七肽重复区。

图 1 突变位点在 SARS-CoV-2 基因组中的分布位置 (A) 以及 S 基因中 2 个突变位点 (B) 和 N 基因中 4 个突变位点 (C) 的位置

Fig. 1 Position and distribution of mutations in the SARS-CoV-2 genome (A), and locations of two mutations in S gene (B) and four mutations in N gene (C)

从图 2 中可以看出, Spike_D614G 和 NSP12_P323L 突变在所有地点的病毒株中 100% 出现。NSP7_E23G 突变仅出现于北京的病毒株中, NSP2_T573I 仅出现于青岛的一条序列中, 除北京和青岛病毒株外其他病毒株均具有相同的突变位点。N_A211V 突变仅出现在大连的病毒株中, 出现频率较低。同一地方的病毒株相似性接近 100%, 表明具有相同的来源并且

没有出现新的变异，大连和青岛的病毒株分别有 8 处突变，浙江和云南的病毒株中分别只有 5 个突变位点。

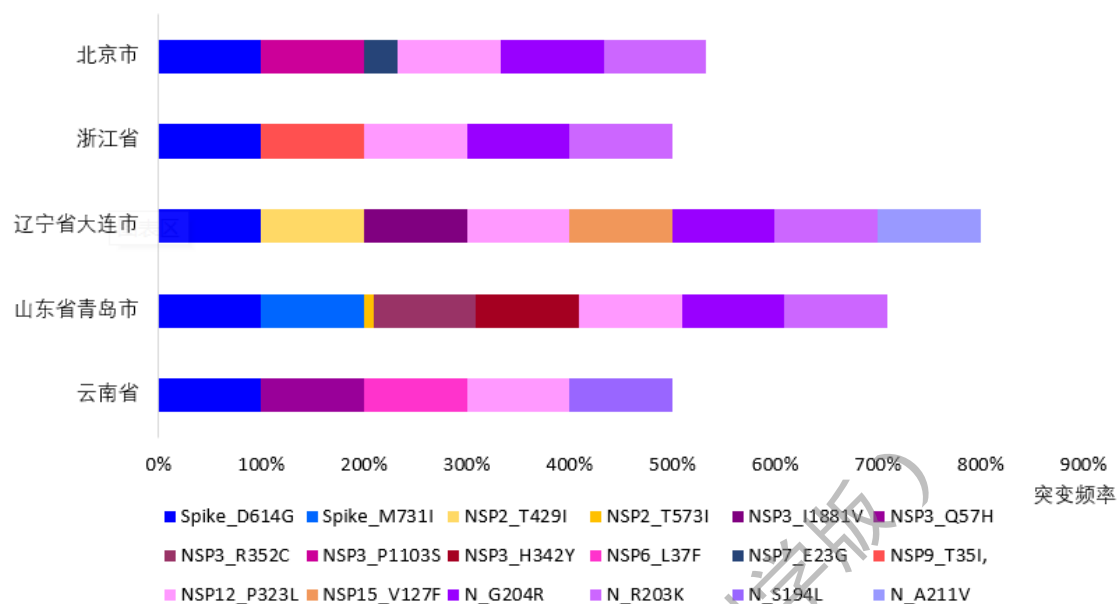


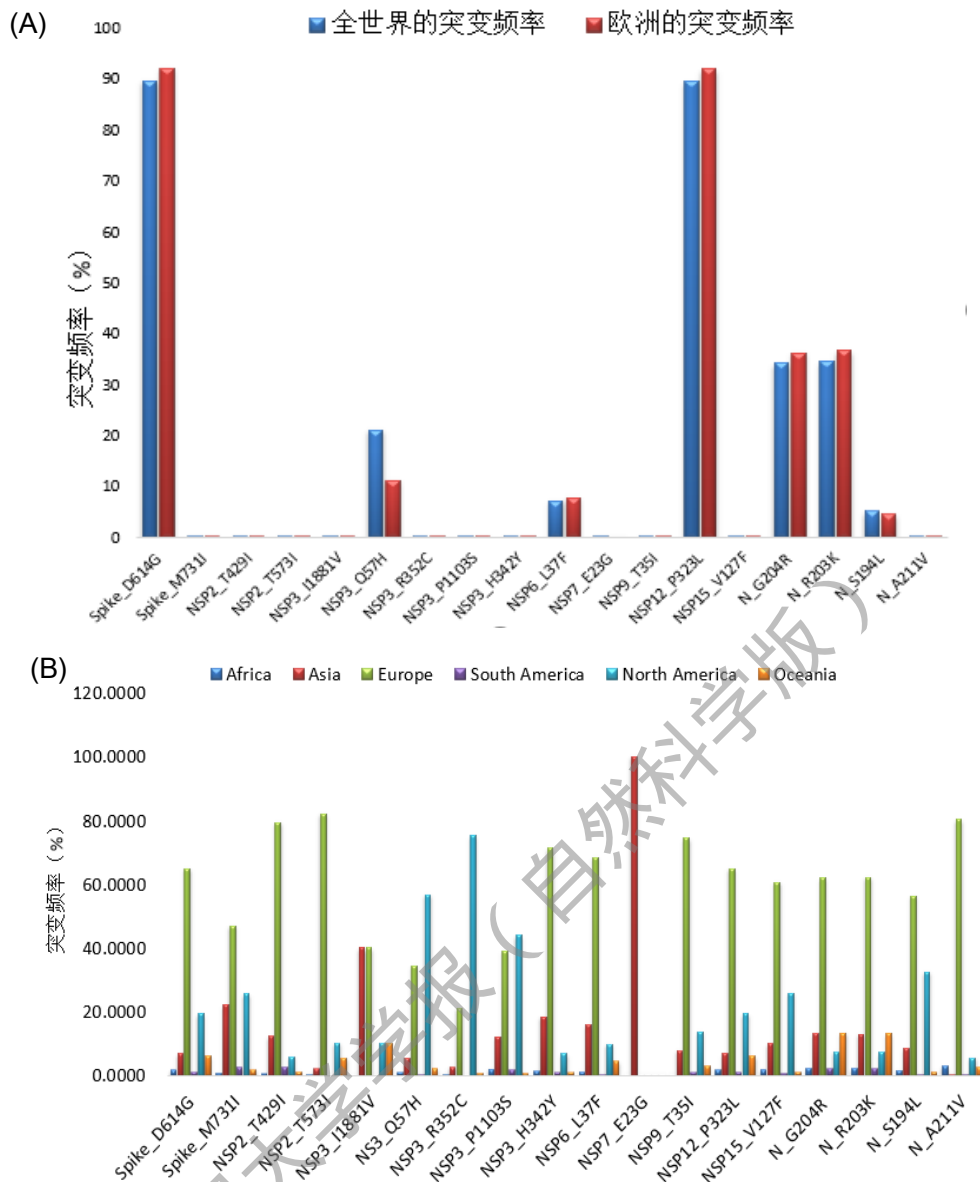
图 2. 2020 年下半年中国各地公布的病毒序列中错义突变的发生频率

Fig. 2 Proportion of missense mutations in each city of China in the second half of 2020

2.2 突变频率分析

突变频率是指一个位点突变在群体中出现的比率。分析上述各突变分别在亚、非、欧、南美、北美和澳洲的突变频率，结果显示：截止 2020 年 12 月，D614G 突变在全世界范围内已经达到 89%，而在欧洲更是高达 92%，该类型毒株因其具有较强的传播能力而成为优势毒株；NSP12_P323L 同样在欧洲的突变频率为 92%，在全世界的病毒株中突变频率是 89%（图 3(A)）。本研究中的 21 株病毒株均含有这两个突变位点。

除 NSP3 外，其他 *ORF1AB* 上的突变位点均主要分布于欧洲（图 3(B)），NSP3 的 3 个位点突变（Q57H、R352C、P1103S）主要发生于北美洲，Q57H 在欧洲的突变频率为 11%，而在世界范围内为 21%（图 3(A)）。NSP7_E23G 因为只有一株，故并无统计意义。*N* 基因上的 4 个突变主要分布于欧洲，203/204 位点形成 N GR>RK 突变，该突变的频率在欧洲为 36%，在全世界为 34%，统计发现这两个位点的突变频率也存在差异，203 位点突变较 204 位点突变略高。N_S194L 在欧洲的突变频率为 4.6%，而在世界范围内为 5.15%（图 2(A)），主要出现在欧洲和北美洲（图 2(B)）。A211V 突变非常少见，在欧洲病毒株中突变频率为 0.12%（181 株），而在全世界范围内更低至 0.09%（224 株）（图 2(A)），且主要（80.8%）出现在欧洲（图 2(B)），在 21 株中存于大连的 4 个病毒株中。



(A) 各突变位点在欧洲和全世界病毒株中的具有相似的突变频率； (B) 各突变位点在世界范围内（6个洲）的分布情况。

图3 中国2020年下半年（截至2020年12月1日）出现的SARS-CoV-2病毒株中的各突变位点在欧洲和全世界范围内的突变频率及地理分布

Fig. 3 The frequency of each amino acid mutation of SARS-CoV-2 virus strains, which emerging in China in the second half of 2020 (by the time of 2020-12-01), in Europe and the whole world as well as their geographical distribution

2.3 突变最早出现的时间和地点

在GISAID网站内通过筛选，寻找各突变在欧洲和全世界范围内最早出现的时间和地点（表2），发现2020年下半年出现在中国的病毒株中18个突变位点最早主要出现在亚洲（中

国、印度、越南)和欧洲各国。D614G 突变位点在 2020 年 1 月 24 日浙江的病毒株中已经出现。出现频率非常高的 NSP12_P323L 突变最早出现于意大利, NSP7_E23G 突变为所有突变中唯一新出现的突变, N 203_204 GR>RK 突变最早出现于德国和荷兰, NSP3_Q57H 最早出现于法国。18 个突变位点中的 8 个最早出现于欧洲, 8 个最早出现于亚洲, NSP2_T429I 和 NSP9_T35I 最早出现于北美。从突变出现的时间来看, 在全世界范围内 10 个突变出现于 2020 年 1 月和 2 月, 7 个突变最早出现于 3 月, 除 NSP7_E23G 外均最早出现于 2020 年 1—3 月。S 基因中的 M731I 突变虽然在 2020 年 2 月已经出现, 但是突变频率仅为 0.035%。

表 2 各突变位点最早和最近出现的时间地点

Tab. 2 The earliest and most recent occurrence time and collected location of each mutation

突变位点	世界范围内最早出现的时间	突变位点在世界范围内最早出现的地点	最近出现的时间	最近出现该突变的病毒株
Spike_D614G	2020-01-24	亚洲/中国/浙江	2020-12-20	hCoV-19/Australia/NSW1209/2020
Spike_M731I	2020-02-20	亚洲/中国	2020-12-20	hCoV-19/England/QEUH-B0 F376/2020
NSP2_T429I	2020-01-29	美国/加利福尼亚	2020-11-20	hCoV-19/New Zealand/20CV0587/2020
NSP2_T573I	2020-03-10	欧洲/英格兰	2020-11-20	hCoV-19/England/NOTT-11C EBB/2020
NSP3_I1881V	2020-03-30	欧洲/西班牙/安达卢西亚/格拉纳达	2020-11-20	hCoV-19/England/LOND-125 0F78/2020
NSP3_Q57H	2020-02-26	欧洲/法国/卢瓦尔/南特	2020-11-20	hCoV-19/England/NORW-F8 496/2020
NSP3_R352C	2020-03-29	亚洲/印度/泰米尔纳德邦	2020-11-20	hCoV-19/USA/WI-GMF-331 62/2020
NSP3_P1103S	2020-02-02	上海	2020-11-20	hCoV-19/England/ALDP-B5 C5FC/2020
NSP3_H342Y	2020-03-06	欧洲/瑞典/斯德哥尔摩	2020-11-20	hCoV-19/Belgium/ULG-1078 8/2020
NSP6_L37F	2020-01-17	云南/昆明	2020-11-20	hCoV-19/New Zealand/20CV0594/2020
NSP7_E23G	2020-06-11	中国/北京	/	/
NSP9_T35I	2020-03-19	美国/马萨诸塞	2020-11-20	hCoV-19/Belgium/ULG-1082 0/2020
NSP12_P323L	2020-02-20	欧洲/意大利/伦巴第	2020-12-20	hCoV-19/Australia/NSW1209 /2020
NSP15_V127F	2020-03-12	欧洲/荷兰/乌得勒支	2020-11-20	hCoV-19/England/QEUH-B8 19A7/2020
N_G204R	2020-02-25	欧洲/德国/巴登-符腾堡州	2020-11-20	hCoV-19/Ecuador/ZZL-4/2020

N_R203K	2020-02-24	欧洲/荷兰/柏林	2020-11-20	hCoV-19/Ecuador/ZZL-4/2020
N_S194L	2020-01-16	亚洲/中国/广东/深圳	2020-12-20	hCoV-19/Australia/NSW1209/2020
N_A211V	2020-03-27	亚洲/越南/河内	2020-11-20	hCoV-19/England/ALDP-B5C40E/2020

2.4 相似病毒株和来源分析

在 GISAID 网站中搜寻与目标病毒株最相近的病毒株,发现云南病毒株与来自我国台湾地区、泰国、沙特阿拉伯的病毒株具有最高的相似性。青岛病毒株与北美、南美的巴西和欧洲的瑞典的病毒株具有最高相似性,大连的病毒株与欧洲的俄罗斯,法国和瑞士的病毒株相似,浙江的病毒株与阿拉伯联合酋长国阿布扎比的两株病毒株相似性最高。北京的病毒株与美国宾夕法尼亚的一个病毒株以及沙特阿拉伯,越南和欧洲的瑞士及英国的一些病毒株非常相似。除云南外,其他四地的病毒株均与匹兹堡的一株病毒株(EPI_ISL_682000)和瑞典乌普萨拉的一株病毒株(EPI_ISL_676525)的病毒株非常相似(表3)。

表3 与各地病毒株相似度最高的 SARS-CoV-2 病毒株列表

Tab. 3 List of SARS-CoV-2 strains in the world that with the highest similarity to strains of each city

查询病毒株	采集时间	相似病毒株获取号	采集地点	相似度
云南	2020-09-17	EPI_ISL_667809	我国台湾地区/桃园	99.99%
	2020-09-14	EPI_ISL_708812	泰国/曼谷	99.99%
	2020-06-25	EPI_ISL_678004	沙特阿拉伯	99.98%
	2020-04-24	EPI_ISL_513151	沙特阿拉伯	99.98%
青岛	2020-03-19	EPI_ISL_682000	北美/美国宾夕法尼亚州匹兹堡	99.97%
	2020-06-29	EPI_ISL_708529	南美/巴西/圣保罗	99.93%
	2020-06-30	EPI_ISL_677925	亚洲/沙特阿拉伯吉达	99.97%
	2020-04-30	EPI_ISL_676525	欧洲/瑞典/乌普萨拉	99.97%
	2020-05-25	EPI_ISL_660417	欧洲/瑞典/	99.97%
大连	2020-04-17	EPI_ISL_514750	俄罗斯/莫斯科	99.98%
	2020-05-20	EPI_ISL_447725	欧洲/法国	99.98%
	2020-03-19	EPI_ISL_682000	北美/美国宾夕法尼亚州匹兹堡	99.98%
	2020-04-30	EPI_ISL_676525	欧洲/瑞典/乌普萨拉	99.98%
浙江	2020-05-19	EPI_ISL_698983	亚洲/阿拉伯联合酋长国/阿布扎比	99.99%
	2020-05-18	EPI_ISL_698277	亚洲/阿拉伯联合酋长国/阿布扎比	99.99%
北京	2020-03-19	EPI_ISL_682000	北美/美国宾夕法尼亚州匹兹堡	99.99%
	2020-03-19	EPI_ISL_682000	北美/美国宾夕法尼亚州匹兹堡	99.99%
	2020-06-30	EPI_ISL_677925	亚洲/沙特阿拉伯吉达	99.99%
	2020-03-21	EPI_ISL_660542	亚洲/越南	99.99%
	2020-05-25	EPI_ISL_660417	欧洲/瑞典	99.99%
	2020-04-10	EPI_ISL_660413	欧洲/瑞典	99.99%

2020-04-30	EPI_ISL_676525	欧洲/瑞典/乌普萨拉	100%
2020-03-23	EPI_ISL_535007	欧洲/英格兰	100%
2020-03-29	EPI_ISL_534944	欧洲/英格兰	100%
2020-03-20	EPI_ISL_534902	欧洲/英格兰	99.99%

3 讨论

3.1 我国 2020 年下半年病毒株均具有典型欧洲病毒株支系突变位点

我国 2020 年下半年 21 株病毒株中所包含的突变位点均主要分布于欧洲和北美洲,最早出现的时间都是在 2020 年 1—3 月,最早出现的地点主要是亚洲、欧洲和北美洲。Takahiko Koyama 等在疫情早期将 D614G 支系病毒株细分为 D614G/Q57H、D614G/Q57H/T265I、D614G/203_204delinsKR/、D614G/203_204delinsKR/T175M^[5], 据此,我国下半年的病毒株中,云南病毒株属于 D614G/Q57H/支系,该支系最早在 2020 年 2 月 26 出现于法国 (EPI_ISL_418219)。其他四地的病毒株属于 D614G/203_204delinsKR/支系,最早于 2020 年 2 月 25 出现于德国 (EPI_ISL_412912)。我国 2020 年下半年公布的 21 株病毒株均具有 S 基因的 D614G 突变和 RdRp (NSP12)基因中的 P323L 突变,两者为共进化位点^[2,5]。这两个位点是欧洲 G 支系病毒株的主要特点^[3,5-6]。SARS-CoV-2 RdRp (也称为 NSP12)是复制/转录机制的关键组成部分,直接参与病毒复制,与 SARS-CoV 的 NSP12 具有较高的同源性,说明其功能和作用机制可能非常保守^[7-8]。SARS-CoV 中 RdRp 与 NSP 7 和 NSP8 形成了一个中空的圆柱形超复合体,复制超复合体与 NSP14 相互作用,NSP14 是一种具有巢病毒典型校对能力的外切酶,对于控制 RNA 复制中的保真和病毒突变非常重要^[9-10]。D614G 突变病毒株在南亚 1 月的时候还非常罕见,从 3 月 20 日后迅速增加,5 月已达到 81%,成为主要流行的病毒株,较祖先病毒株更有传播优势^[3],具有更强的适应性和传染性,感染者具有更 D614G 高的病毒载量;与 COVID-19 的感染严重程度无关;可能会降低对个别恢复期血清的中和敏感性,其中的机制很多研究者从分子、细胞及种群的层面分别进行了解释^[11-16]。

3.2 突变频率具有偏向性

根据突变位点在基因组上的位置,可以看出 21 株病毒株的主要突变分布于 ORF1AB、S 和 N 基因,但是同一基因上各突变位点发生的频率却存在很大差别,即使在位置上很靠近的两个位点,最早出现的时间也比较接近,如 D614G 和 M731I 都位于 S 基因的 S1 和 S2 两个亚基之间的断裂点附近,但是 D614G 的发生频率达到 90%,而 M731I 的突变频率只有 0.035%。ORF1AB 中,NSP3_Q57H、NSP6_L37F、NSP12_P323L 是突变频率较高的突变位点,

尤其是 NSP12_P323L 与 S_D614G 突变频率保持一致，其他突变位点的频率都非常低。N 基因上的 4 个突变位点集中于 190~210 之间的 20 个氨基酸范围内，出现的时间也很接近(表 2)，但是 203_204delinsKR 突变发生的频率（34%）较 N_S194L（4.6%）高很多。这种偏向性是病毒对宿主作用和适应的结果，与病毒进化的选择压力有关。

3.3 病毒株的突变分析对疫情防控措施的评价意义

2020 年入冬以来，世界各地的新冠疫情都有所加剧，尤其是一些国家出现了新的变异病毒株引起了全世界的恐慌。近一段时间，我国也在多地出现小范围疫情，在前期抗击疫情的工作中建立的防控体系的严密监控之下，采取隔离、限制人员流动等措施，使得小范围的疫情在短期内被控制并消除。突变分析也可以作为疫情防控的一种评价方法。北京的病毒株中出现了新发生的突变 NSP7_E23G，由于及时发现、隔离等有力措施，这类病毒株没有发生扩散，目前世界各地的病毒株中都还没有发现这个突变，这也说明我国采取的防控措施非常有效。

致谢： 特别感谢厦门大学生命科学学院吴勤刚博士在病毒基因组突变分析方面的悉心指导。

参考文献：

- [1] LI Q Q, WU J J, NIE J H, et al. The impact of mutations in SARS-CoV-2 Spike on viral infectivity and antigenicity[J]. Cell, 2020, 182(5):1284-1294.
- [2] MARIA P, BRUNA M, FRANCESCA B, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant[J]. J Transl Med, 2020,18(1):179.
- [3] ISLAM O K, AL-EMRAN H M, HASAN M S, et al. Emergence of European and North American mutant variants of SARS-CoV-2 in South-East Asia [J/OL]. Transbound Emerg Dis, 2020. [2020-12-30]. doi: 10.1111/tbed.13748.
- [4] MARCELO H S P, DUSCHINKA R D G, CASSIA D, et al. Multiple introductions followed by ongoing community spread of SARS-CoV-2 at one of the largest metropolitan areas of northeast Brazil[J] Viruses, 2020, 12:1414.
- [5] KOYAMA T, PLATT D, PARIDA L. Variant analysis of SARS-CoV-2 genomes[J]. Bulletin of the World Health Organization, 2020, 98(7): 495-504.
- [6] FORSTER P, FORSTER L, RENFREW C, et al, Phylogenetic network analysis of SARS-CoV-2 genomes[J]. Proceedings of the National Academy of Sciences of the United States of America, 2020,117(17): 9241-9243.

- [7] ZEBUHRL J. The coronavirus replicase[J]. *Curr Top Microbiol Immunol*, 2005, 287:57-94.
- [8] WU F, ZHAO S, YU B, et al. A new coronavirus associated with human respiratory disease in China[J]. *Nature*, 2020, 579(7798):265-269.
- [9] KIRCHDOERFER R N, WARD A B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors[J]. *Nat Commun*, 2019, 10:2342.
- [10] ZHAI Y, SUN F, LI X, et al. Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer [J]. *Nat Struct Mol Biol*, 2005, 12(11):980-986.
- [11] BULUT C, KATO Y. Epidemiology of COVID-19[J]. *Turk J Med Sci*, 2020, 50 (SI-1):563-570.
- [12] WANG Q, ZHANG Y, WU L, et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2[J]. *Cell*, 2020,181(4):894-904.e9
- [13] BELOUZARD S, MILLET J K, LICITRA B N, et al, Mechanisms of coronavirus cell entry mediated by the viral spike protein [J]. *Viruses*, 2012, 4(6):1011-1033.
- [14] ZHANG L, JACKSON C B, MOU H, et al. The D614G mutation in the SARS-CoV-2 Spike protein reduces S1 shedding and increases infectivity [J]. *Nat Commun*. 2020,11(1):6013.
- [15] DANILOSKI Z, GUO X, SANJANA N E. The D614G mutation in SARS-CoV-2 Spike increases transduction of multiple human cell types[EB/OL]. *bioRxiv*. [2020-12-30]. doi: 10.1101/2020.06.14.151357.
- [16] OMOTUYI I O, NASH O, AJIBOYE O B, et al. Atomistic simulation reveals structural mechanisms underlying D614G spike glycoprotein-enhanced fitness in SARS-CoV-2[J]. *Journal of Computational Chemistry*, 2020, 41(24): 2158-2161.

Genetic mutation analysis of SARS-CoV-2 strains in the second half of 2020 in China

SUN Huimin¹, TANG Wenbin², ZHANG Lei³, LI Jiwei⁴, YE Shaopei²

(1. Central Laboratory of Affiliated Xiang'an Hospital of Xiamen University, Xiamen 361102, China; 2. Department of Urology, affiliated Xiang 'an Hospital of Xiamen University, Xiamen 361102, China; 3. Department of Preventive Medicine, Air Force Medical University, Xi'an 710032, China; 4. Department of Respiratory and Critical Illness and Sleep Medicine, affiliated Xiang'an Hospital of Xiamen University, Xiamen 361102, China)

Abstract: The present study based on all the 21 SARS-CoV-2 genome sequences on GISAID (<http://platform.gisaid.org>) that collected from June 1, to December 1, 2020 in China, the NC_045512 and MN996528 strains were used as reference strain, we found the missense mutation sites of every strain, calculated the mutation frequency, and searched for similar virus strains in the GISAID. There were 18 missense mutations in all 21 strains, mainly located in *Spike (S)* gene, *ORF1AB* gene and *N* gene. D614G mutation in *S* gene and P323L mutation in *NSP12 (RdRp)* gene had the highest mutation frequency. Except for a new missense mutation in Beijing strain, the other mutations were mainly distributed in the strains from Europe or North America. According to the classification of mutation sites, all the virus strains belong to the G clade, among which the Yunnan virus strains belong to the GH clade. The strains from Qingdao, Dalian, Zhejiang and Beijing belong to the GR clade and are similar to those strains from Europe. Taken together, the SARS-CoV-2 strains collected from China in the second half of 2020 are mainly imported from other countries. In terms of prevention and control strategy, we should prevent the importation of foreign sources virus strains at all transportation links. The strain contains novel mutation did not spread to other places further indicates the efficiency of prevention and control measures of COVID-19 adopted in China.

Keywords: COVID-19; SARS-CoV-2; gene mutation