

新冠病毒Omicron变异株S蛋白变异位点及进化分析

吴秀兰¹, 李绮晴², 郑珊丽², 黄万信¹, 钟华斌¹, 唐文武^{2*}

¹肇庆学院食品与制药工程学院, 广东 肇庆

²肇庆学院生命科学学院, 广东 肇庆

收稿日期: 2022年9月26日; 录用日期: 2022年11月2日; 发布日期: 2022年11月9日

摘要

新冠病毒Omicron变异株是2021年11月首次在南非发现, 并已成为当前传染性最高的主流变异株, 其传染性和S蛋白位点变异存在相关性。本文通过生物信息学方法对Omicron变异株的S蛋白变异位点进行分析, 并与多种冠状病毒S蛋白进行多序列比对及构建系统进化树。结果显示Omicron变异株S蛋白序列保守性较高, 共发现26个变异位点, 其中K417N、Q493R、N501Y、Y505H变异位点可能增强病毒感染性。序列比对显示该变异株与Zeta变异株一致性最高, 达到98.11%, 其次与猫、猴的冠状病毒S蛋白序列一致性皆为98.04%, 与Lambda变异株相似性最低, 其一致性为97.01%。本研究有助于了解Omicron变异株S蛋白的功能, 并为后续筛选靶向药物提供理论依据。

关键词

新冠病毒, Omicron, S蛋白, 变异位点, 进化分析

Mutation Sites and Evolutionary Analysis of SARS-CoV-2 Omicron Spike Protein

Xiulan Wu¹, Qiqing Li², Shanli Zheng², Wanxin Huang¹, Huabin Zhong¹, Wenwu Tang^{2*}

¹College of Food and Pharmaceutical Engineering, Zhaoqing University, Zhaoqing Guangdong

²College of Life Science, Zhaoqing University, Zhaoqing Guangdong

Received: Sep. 26th, 2022; accepted: Nov. 2nd, 2022; published: Nov. 9th, 2022

Abstract

The omicron variant was first detected in South Africa in November 2021 and has become the most prevalent variant, with a correlation with variation in its Spike protein locus. Using bioinformatics methods and software, we analyzed the mutation sites of Spike protein with the Omicron variant.

*通讯作者。

文章引用: 吴秀兰, 李绮晴, 郑珊丽, 黄万信, 钟华斌, 唐文武. 新冠病毒 Omicron 变异株 S 蛋白变异位点及进化分析[J]. 计算生物学, 2022, 12(4): 59-65. DOI: 10.12677/hjcb.2022.124008

cron variant and compared it with other coronaviruses to construct the phylogenetic tree. The results showed that the sequence of Spike protein with the Omicron variant is highly conserved. A total of 26 mutation sites have been found in the Spike protein, among which K417N, Q493R, N501Y, and Y505H may enhance virus infectivity. Sequence alignment showed that the omicron variant had the highest identity with the Zeta variant, reaching 98.11%, followed by coronaviruses such as *Felinae* and *Primates*, with 98.04% identity. The lowest identity is with the Lambda variant, with an identity of 97.01%. This study will help to understand the function of the Spike protein of the Omicron variant and provide a theoretical basis for further research and screening of new antiviral drug targeting Spike protein.

Keywords

SARS-CoV-2, Omicron, Spike Protein, Mutation Sites, Evolution Analysis

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

2019年12月出现的新型冠状病毒(Severe Acute Respiratory Syndrome Coronavirus 2, SARS-CoV-2, 简称新冠病毒)感染引发的新型冠状病毒肺炎(Coronavirus Disease 2019, COVID-19, 简称新冠肺炎), 已在全球范围内肆虐传播两年多, 对全球公共卫生及人类生命健康造成巨大威胁[1]。尽管全球各地已采取各种防疫措施, 但每日新增确诊病例数及死亡病例数仍在快速增加。据世界卫生组织(World Health Organization, 简称 WHO)实时统计数据, 截至北京时间 2022 年 5 月 30 日, 全球累计新冠肺炎确诊病例 252,467,084 例, 累计死亡病例 6,285,171 例[2]。新冠病毒在大规模传播的同时不断发生变异进化, 产生基本传播系数(Basic reproductive number, 简称 R_0)或致病性更强的变异株, 如 Alpha 变异株、Beta 变异株、Gamma 变异株、Delta 变异株, 对新冠防控和疫苗开发造成巨大影响[3]。2021 年 11 月 9 日于南非首次检测到的 Omicron (B.1.1.529)变异株, 在短短一周时间即成为南非豪登省主流新冠变异株, 并于同月 26 日被 WHO 定性为最高级别的“关切变异株” [4]。目前全球有 150 多个国家及地区检测出 Omicron 变异株[5], 该变异株已成为全球优势变异株[6]。

作为 Omicron 变异株发现前传染性最高的变异株, Delta 变异株的 R_0 值约为 5~8 [7], 而 Omicron 变异株的 R_0 数值在全球范围内可能高达 10 [8]。流行病学追踪显示, Omicron 变异株比 Delta 变异株更易于传播, 其传播速度亦高于 Delta 变异株[9]。病毒基因测序显示, 2022 年 2 月底开始爆发, 导致感染人数超过 52 万的上海市新冠疫情, 主要是 Omicron 变异株传播造成[10]。新冠病毒主要有四种结构蛋白, 分别为表面刺突蛋白(S 蛋白)、膜蛋白(M 蛋白)、核衣壳蛋白(N 蛋白)及包膜蛋白(E 蛋白) [11]。其中 S 蛋白具有识别人体细胞表面血管紧张素转化酶 II (简称 ACE2 蛋白)受体蛋白, 并与之特异性结合, 从而介导细胞膜融合的功能, 能够影响冠状病毒感染宿主细胞、特异性识别特定宿主的能力[12]。S 蛋白与 ACE2 受体蛋白的亲和力决定了新冠病毒感染能力, 因此 S 蛋白的氨基酸位点变异与新冠病毒传染性密切相关 [13]。Omicron 变异株是当前 R_0 数值最高的新冠病毒变异株, 其 S 蛋白的变异位点及其特征研究具有重要意义。本文对新冠病毒 Omicron 变异株的 S 蛋白变异位点进行分析, 与其他冠状病毒的 S 蛋白进行多序列比对及构建系统进化树, 以便加深人们对 Omicron 变异株 S 蛋白功能的理解, 并对后续研究、筛选靶向 S 蛋白的新型抗病毒药物提供理论依据。

2. 材料与方法

2.1. Omicron 变异株的 S 蛋白序列获取

从 NCBI (<https://www.ncbi.nlm.nih.gov>) 下载原始新冠病毒 Wuhan-Hu-1 (NC_045512.2) S 蛋白氨基酸序列及核苷酸序列。以来源于印度的 Omicron 变异单株(EPI_ISL_10204983)为例, 通过 GISAID (<https://www.gisaid.org/>) 数据库下载该测序株的基因组序列。通过线上网站 EMBOSS Needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/) 获取其 S 蛋白核苷酸序列, 通过线上网站 ExPasy (<https://web.expasy.org/translate/>) 将其转换为氨基酸序列。

2.2. Omicron 变异株 S 蛋白变异位点预测

通过线上网站 Nextclade (<https://clades.nextstrain.org>) 对 Omicron 变异株 S 蛋白进行变异位点分析, 并利用在线工具 Phyre2 对 S 蛋白三级结构进行模拟构建, 使用 PyMOL 对结果进行可视化操作, 标记主要突变位点的所在空间位置。

2.3. 不同新冠病毒变异株的多序列比对

从 NGDC (<https://ngdc.cncb.ac.cn/ncov/>) 及 GISAID 数据库中, 下载 WHO 认可的 5 个关注 VOC 变种 (Alpha, Beta, Gamma, Delta, Omicron) 和 7 个需留意的 VOI 变种 (Epsilon, Zeta, Eta, Theta, Iota, Kappa, Lambda) 的全基因组序列, 按照上述 2.1 方法获取新冠病毒变异株的 S 蛋白序列。将 12 种新冠病毒变异株和 Wuhan-Hu-1、猫、狗、猴、马来菊头蝠等冠状病毒的 S 蛋白序列, 利用 MEGA11 软件进行多序列比对。

2.4. 不同冠状病毒的系统进化树构建

从 NCBI 网站中获取水貂 (*Neovison vison*)、孟加拉虎 (*Panthera tigris*)、亚洲狮 (*Panthera leo persica*)、白鼬 (*Mustela erminea*)、马来亚虎 (*Panthera tigris jacksoni*)、白尾鹿 (*Odocoileus virginianus*)、骆驼 (*Camelus*)、蝙蝠 (*Chiroptera*)、猪 (*Sus*)、牛 (*Bovini*)、麻雀 (*Passer*)、兔 (*Rabbit*) 等 13 种动物的新冠病毒 S 蛋白序列, 并与前文的 Alpha、Beta、Gamma、Delta、Omicron 5 种新冠病毒变异株, 以及猫 (*Felinae*)、狗 (*Canis lupus familiaris*)、猴 (*Primates*)、马来菊头蝠 (*Rhinolophus malayanus*) 等冠状病毒和 Wuhan-Hu-1 共 24 种冠状病毒的 S 蛋白序列, 利用 MEGA11 软件的邻接法构建系统进化树。

3. 结果与分析

3.1. Omicron 变异株 S 蛋白变异位点分析

由线上网站 Nextclade 比对分析显示, 所选 Omicron 变异株 S 蛋白与最早发现的 Wuhan-Hu-1 标准毒

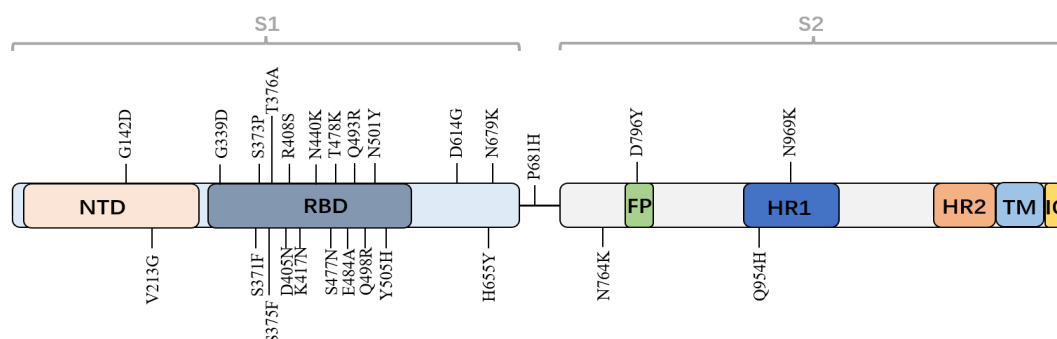


Figure 1. The S protein mutation site of the SARS-COV-2 Omicron variant

图 1. 新冠病毒 Omicron 变异株的 S 蛋白变异位点

株 S 蛋白进行比对, 共有 26 个位点发生突变, 分别为 G142D、V213G、G339D、S371F、S373P、S375F、T376A、D405N、R408、K417N、N440K、S477N、T478K、E484A、Q493R、Q498R、N501Y、Y505H、D614G、H655Y、N679K、P681H、N764K、D796Y、Q954H、N969K (见图 1)。其中与人 ACE2 受体结合特异性和亲和力相关的 RBD 区域, 共发现 16 个氨基酸突变位点, 分别为 G339D、S371F、S373P、S375F、T376A、D405N、R408S、K417N、N440K、S477N、T478K、E484A、Q493R、Q498R、N501Y、Y505H。

3.2. Omicron 变异株 S 蛋白三级结构分析

利用在线工具 Phyre2 对 Omicron 变异株 S 蛋白的三级结构进行模型构建, 在获得的结果中下载 pbd 文件, 使用 PyMOL 进行可视化分析并标记出 RBD 区域的主要突变位点, 相关结果见图 2。图 2(a) 显示出 S1 和 S2 蛋白的三维结构, 以及 S1 蛋白的 RBD 区域(灰色)和 NTD 区域(麦色)。图 2(b) 显示, 在 Omicron 变异株的 S 蛋白 RBD 区域的 16 个氨基酸变异位点中, K417N、N440K、T478K、E484A、Q493R、Q498R、N501Y、Y505H 共 8 个氨基酸位点与人受体蛋白 ACE2 结合域相互作用。因此上述位点变异可能增强新冠病毒与人 ACE2 的结合能力, 进而提高病毒感染力。

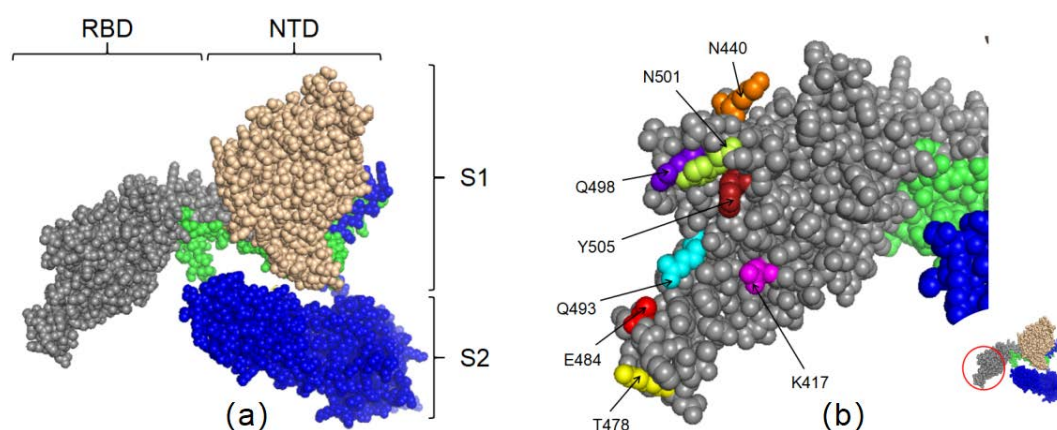


Figure 2. The tertiary structure of S protein and RBD mutation position with SARS-COV-2 Omicron variant; (a) The tertiary structure model of Omicron variant's S protein; (b) RBD mutation position of SARS-COV-2 Omicron variant

图 2. Omicron 变异株的 S 蛋白三级结构和 RBD 区域重要突变位点示意图; (a) Omicron 变异株的 S 蛋白三级结构; (b) S 蛋白 RBD 区域重要突变位点标记示意图

3.3. 不同新冠病毒变异株的多序列比对

对 5 个关注 VOC 变种, 7 个需留意 VOI 变种, 以及 4 种动物冠状病毒的 S 蛋白进行多序列比对, 部分比对集合图见图 3。结果显示, 17 种冠状病毒 S 蛋白中, 完全一致的氨基酸残基有 1187 个, 占 93.17%, 表明 S 蛋白的氨基酸序列保守性较高。其中与 Omicron 变异株 S 蛋白序列一致性最高的是 Zeta 变异株, 两者序列一致性高达 98.11%; 其次是猫、猴的新冠病毒 S 蛋白序列, 序列一致性皆为 98.04%; 与新冠病毒 Lambda 变异株相似性最低, 其序列一致性仅为 97.01%。

3.4. 不同冠状病毒的 S 蛋白系统进化分析

对 17 种不同动物冠状病毒、5 种新冠病毒变异株以及 SARS 病毒共 24 种冠状病毒的 S 蛋白进行多序列比对, 并采用邻接法构建系统进化树, 分析结果见图 4。结果显示, 骆驼、麻雀、猪、蝙蝠、牛、兔以及人类 SARS 等冠状病毒聚为一类, 自展值为 100, 说明它们在进化上相距较近, 可能来源于同一祖先。人类 Omicron 变异株(EPI_ISL_10204983)、Gamma 变异株(OU066503.1)以及 Beta 变异株(MZ314998.1)

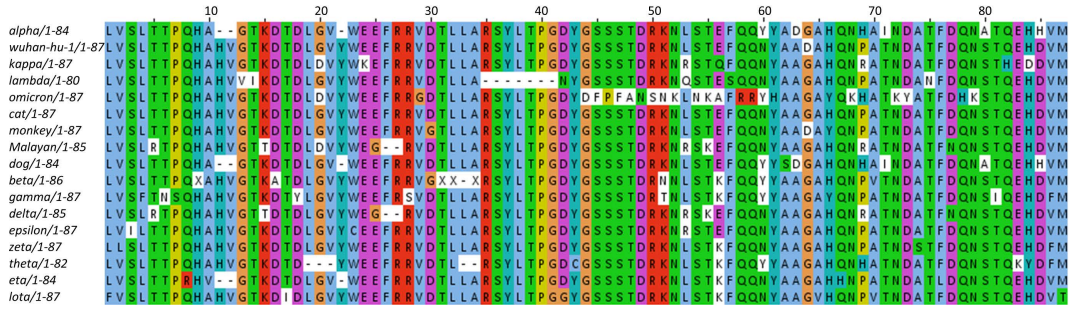
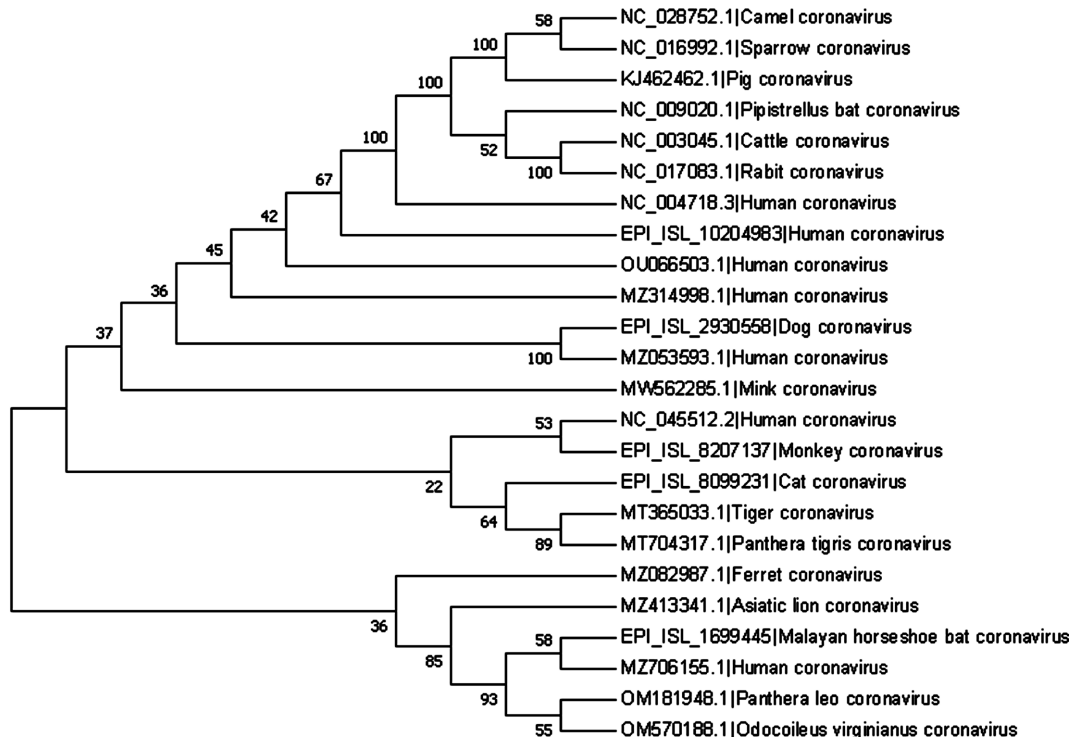


Figure 3 Multiple sequence alignment composite map of different coronavirus S proteins
图 3 不同冠状病毒的 S 蛋白多序列比对集合图

与上述冠状病毒遗传距离较近，进而可推测这些冠状病毒在进化上方向一致。

聚类分析显示，白鼬、亚洲狮、狮子、白尾鹿、马来菊头蝠的冠状病毒以及人类新冠病毒 Delta 变异株(Z706155.1)聚为一支，马来亚虎、虎、猫、猴的冠状病毒以及最早发现的新冠病毒 Wuhan-Hu-1 (NC_045512.2)聚为一支，上述结果说明冠状病毒进化可能与宿主物种无关，冠状病毒 S 基因的 RNA 序列变异具有随机性和多样化，导致其进化与宿主物种及发生时间等没有相关性。



注：SARS 冠状病毒(NC_004718.3)；水貂 *Neovison vison* (MW562285.1)；马来亚虎 *Panthera tigris jacksoni* (MT365033.1)；亚洲狮 *Panthera leo persica* (MZ413341.1)；狮子 *Panthera leo* (OM181948.1)；白鼬 *Mustela ermine* (MZ082987.1)；虎 *Panthera tigris* (MT704317.1)；白尾鹿 *Odocoileus virginianus* (OM570188.1)；牛 *Bovini* (NC 003045.1)；骆驼 *Camelus* (NC_028752.1)；蝙蝠 *Pipistrellus* (NC_009020.1)；猪 *Sus* (KJ462462.1)；麻雀 *Passer* (NC_016992.1)；兔 *Rabbit* (NC_017083.1)；Alpha 变异株(MZ053593.1)；Beta 变异株(MZ314998.1)；Gamma 变异株(OU066503.1)；Delta 变异株(Z706155.1)；Omicron 变异株(EPI_ISL_10204983)；猫 *Felinae* (EPI_ISL_8099231)；狗 *Canis lupus familiaris* (EPI_ISL_2930558)；猴 *Primates* (EPI_ISL_8207137)；马来菊头蝠 *Rhinolophus malayanus* (EPI_ISL_1699445)；Wuhan-Hu-1 (NC_045512.2)

Figure 4. Phylogenetic analysis of S proteins with different coronavirus
图 4. 各类冠状病毒 S 蛋白的系统进化树

4. 讨论

新冠肺炎是当前国际社会最受关注的公共卫生事件，SARS-CoV-2 的快速传播危及全世界人民的生命健康及全球经济的发展[14] [15]。2021 年 11 月首次发现的 Omicron 变异株，不仅传染速度高于其它新冠病毒变异株，而且容易让康复后的感染者再次感染[16]。S 蛋白对于冠状病毒入侵宿主细胞至关重要[17]，其 S1 亚基的 RBD 结合域能与人 ACE2 结合，协助病毒穿过细胞膜进入宿主细胞，进而感染人类，因此 RBD 区域变异位点与新冠病毒传染性强弱密切相关[18] [19] [20]。本研究显示，Omicron 变异株与原始新冠病毒 Wuhan-Hu-1 相比，S 蛋白中发现 26 个氨基酸突变位点，其中 16 个变异位点位于 RBD 区域。Shang 等报道新冠病毒 RBD 区域内的 417、449、487、489、493、500、501、502、505 共 9 个变异位点与 ACE2 结合力直接相关[21]，本研究发现 K417N、Q493R、N501Y、Y505H 4 个相同的变异位点，这些变异位点可能增强新冠病毒与人 ACE2 的结合能力，进而提高病毒传染性[22]。

Omicron 变异株的 4 个关键变异位点中，N501Y 变异位点广泛存在于 Alpha、Beta、Gamma 等高传染性变异株中[23]。Frampton 等人研究证明，N501 位氨基酸是冠状病毒受体结合域的关键接触残基之一，直接参与 RBD 和 ACE2 的结合作用，而 N501Y 突变增强了 RBD 和 ACE2 的结合亲和力[24]。刘洋等研究表明，N501Y 变异位点能增加新冠病毒 S 蛋白对上呼吸道细胞受体的亲和力，在巴西、南非等国的趋同进化表明，N501Y 是最主要的关注适应性突变位点，是上呼吸道感染和冠状病毒传播性增强的关键因素[25]。Q493 能与 S 蛋白中和抗体 LY-CoV555 相结合从而形成三个稳定力，当 S 蛋白发生 Q493R 突变时，会导致该位点与抗体无法形成相互作用力，表明 Q493R 突变体可能是 LY-CoV555 逃逸突变体。K417N 突变不直接与 ACE2 相互作用，但其有助于病毒更有效地与 hACE2 结合[3]。

5. 结论

综上所述，本研究对 Omicron 变异株 S 蛋白变异位点进行分析，与其他新型冠状病毒变异株的 S 蛋白进行多序列比对及构建系统进化树，有助于深入了解 Omicron 变异株 S 蛋白的功能，为后续研究、筛选靶向 S 蛋白的新型抗病毒药物提供理论依据。

基金项目

广东省大学生创新训练计划项目(S202010580047)。

参考文献

- [1] 任彪, 廖彬佑, 王建楠, 等. 奥密克戎(Omicron)变异株及其对口腔诊疗的潜在影响[J]. 四川大学学报(医学版), 2022, 53(2): 175-180.
- [2] WHO (2022) Coronavirus (COVID-19) Dashboard. <https://covid19.who.int>
- [3] 吴俣, 刘珏, 刘民, 梁万年. 新型冠状病毒 Omicron 变异株的流行病学特征及防控研究[J]. 中国全科医学, 2022, 25(1): 14-19.
- [4] WHO (2021) Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern. [https://www.who.int/news/item/26-11-2021-classification-of-omicron-\(b.1.1.529\)-sars-cov-2-variant-of-concern](https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern)
- [5] 今冬疫情又起, 但黎明的曙光已不再遥远[EB/OL]. <http://www.CRNTT.com>, 2022-01-21.
- [6] 美国专家反驳“疫情终结论”: 奥密克戎传越广, 下一个变种越早来[EB/OL]. <https://new.qq.com/omn/20220118/20220118A0BLND00.html>, 2022-01-19.
- [7] 14 秒就能人传人, 钟南山口中“传染性极强”的德尔塔毒株是怎么回事?[EB/OL]. [http://finance.sina.com.cn/tech/csj/2021-07-29/doc-ikqciyzk8293182.shtml#~:text=](http://finance.sina.com.cn/tech/csj/2021-07-29/doc-ikqciyzk8293182.shtml#~:text=,2021-12-09), 2021-12-09.
- [8] Burki, T.K. (2022) Omicron Variant and Booster COVID-19 Vaccines. *The Lancet Respiratory Medicine*, **10**, E17. [https://www.thelancet.com/journals/lanres/article/PIIS2213-2600\(21\)00559-2](https://www.thelancet.com/journals/lanres/article/PIIS2213-2600(21)00559-2)

- [9] 新冠病毒又双叒变异! 加息进程受阻?[EB/OL]. <https://finance.sina.com.cn/money/forex/forexinfo/20211129/docikyakumx0681628.shtm>, 2021-11-28.
- [10] 上海疫情已超 52 万人感染, 多为奥密克戎 BA.2 和 BA.2.2 [EB/OL]. https://cj.sina.com.cn/articles/view/1643971635/61fd043300100xa47?finpagefr=p_104, 2022-4-29.
- [11] 徐本锦, 宣焱, 杜淼, 等. 新冠病毒刺突糖蛋白结构与功能的生物信息学分析及原核表达[J]. 病毒学报, 2022, 38(1): 41-56.
- [12] 何金娇, 刘洋洋, 毛雪飞, 等. SARS-CoV-2 S 蛋白结构与功能的生物信息学分析[J]. 基因组学与应用生物学, 2021, 40(Z1): 2346-2355.
- [13] 王嘉君. 浅谈 Omicron 新型冠状病毒变异株的特点和防控措施[J]. 福建轻纺, 2022(2): 33-37.
- [14] Li, Q., Guan, X., Wu, P., *et al.* (2020) Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *New England Journal of Medicine*, **382**, 1199-1207.
- [15] Huang, C., Wang, Y., Li, X., *et al.* (2020) Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. *The Lancet*, **395**, 497-506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
- [16] 南非卫生部: 奥密克戎毒株或致新冠二次感染风险增加[EB/OL]. http://www.news.cn/world/2021-12/11/c_1128153618.htm, 2021-12-12.
- [17] Wrapp, D., Wang, N., Corbett, K.S., *et al.* (2020) Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. *Science*, **367**, 1260-1263. <https://doi.org/10.1126/science.abb2507>
- [18] Astuti, I. (2020) Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An Overview of Viral Structure and Host Response. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, **14**, 407-412. <https://doi.org/10.1126/science.abb2507>
- [19] Kumar, S. and Saxena, S.K. (2021) Structural and Molecular Perspectives of SARS-CoV-2. *Methods*, **195**, 23-28. <https://doi.org/10.1016/j.ymeth.2021.03.007>
- [20] Rodríguez-Maldonado, A.P., Vázquez-Pérez, J.A., Cedro-Tanda, A., *et al.* (2021) Emergence and Spread of the Potential Variant of Interest (VOI) B.1.1.519 of SARS-CoV-2 Predominantly Present in Mexico. *Archives of Virology*, **166**, 3173-3177. <https://doi.org/10.1007/s00705-021-05208-6>
- [21] Shang, J., Ye, G., Shi, K., *et al.* (2020) Structural Basis of Receptor Recognition by SARS-CoV-2. *Nature*, **581**, 221-224. <https://doi.org/10.1038/s41586-020-2179-y>
- [22] Qin, S., Cui, M., Sun, S., *et al.* (2021) Genome Characterization and Potential Risk Assessment of the Novel SARS-CoV-2 Variant Omicron (B.1.1.529). *Zoonoses*, **1**, 1-5. <https://doi.org/10.15212/ZOONOSES-2021-0024>
- [23] 穆雪纯, 李丽, 王凌航. 新型冠状病毒流行变异株的分型及研究进展[J]. 国际病毒学杂志, 2021, 28(4): 336-340.
- [24] Frampton, D., Rampling, T., Cross, A., *et al.* (2021) Genomic Characteristics and Clinical Effect of the Emergent SARS-CoV-2 B.1.1.7 Lineage in London, UK: A Whole-Genome Sequencing and Hospital-Based Cohort Study. *The Lancet Infectious Diseases*, **21**, 1246-1256. [https://doi.org/10.1016/S1473-3099\(21\)00170-5](https://doi.org/10.1016/S1473-3099(21)00170-5)
- [25] Liu, Y., Liu, J., Plante, K.S., *et al.* (2022) The N501Y Spike Substitution Enhances SARS-CoV-2 Transmission. *Nature*, **602**, 294-299. <https://doi.org/10.1038/s41586-021-04245-0>