

谈谈 SARS-CoV-2 的生物学及其变异毒株：对病毒进化的预期以及我们能做些什么

Ruibang Luo¹, Agnès Delaunay-Moisan², Kenneth Timmis³, Antoine Danchin^{4,5}

翻译者：Min Ou¹, Amy Wing-Sze Leung¹, Ruibang Luo¹

¹ Department of Computer Science, The University of Hong Kong, Pokfulam, Hong Kong, China

² Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette, 91198, France

³ Institute of Microbiology, Technical University of Braunschweig, Braunschweig, Germany

⁴ Kodikos Labs, Institut Cochin, 24 rue du Faubourg Saint-Jacques, Paris 75014, France

⁵ School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, 21 Sassoon Road, Hong Kong, China

通讯方法

电话: +331 4441 2551; 传真: +331 4441 2559

电子邮件: antoine.danchin@normalesup.org

摘要

SARS-CoV-2 的全球传播和越来越多变异毒株被识别（其中一些已取代原始进化枝成为优势毒株）突显了该病毒正在努力拓展其进化空间。病毒越是长时间高水平传播及繁衍，就越能适应人类宿主并找到继续流行的途径。第三波 COVID-19 的疫情正在多个国家和地区发生，凸显了当前遏制传播的措施的不足。制定这些措施所依据的其中一个理念是，接种疫苗将迅速阻碍病毒的传播并使社会回归常态。然而，接种疫苗本身将给疫苗带来选择压力并促使让疫苗失效的变异毒株的产生。因此，在大规模疫苗接种期间同时放任病毒继续高水平传播繁殖，并期待疫苗能终止这场大流行的政策，是不现实的。在缺乏有效的抗病毒药物的情况下，预防 SARS-CoV-2 的方式将予与预防流感相似，比如每年都需要针对优势毒株进行疫苗接种。能否战胜 COVID-19 或与其共存是 SARS-CoV-2 将如何进化

和变异的问题。因此，为了更好地预测未来的走向，理解 SARS-CoV-2 将如何进化及各种限制因素至关重要。迄今为止，研究的主要焦点还集中在负责与受体结合的突刺蛋白上。但该病毒总共编码 26 种与宿主以各种方式相互作用的蛋白，因此其进化有多种可能性，并且难以预知。但不管怎样，如果要对 COVID-19 进行最佳防御，就必须对其可能发生的变种进行防御。要做到这一点，我们必须要了解该病毒的各种进化可能性。除了该病毒与细胞间一般的相互作用外，人类中还存在着大量的多态性（例如 Lewis, HLA 等），有些存在于大多或所有人口中，有些则只存在于特定的种族。这些多态性也是推动或限制病毒进化因素。现在，我们拥有足够的资源和手段（包括病毒基因组测序，蛋白质结构测定/建模，蛋白质相互作用分析），对变异毒株进行功能测定，但获得全面的基因组数据并非易事。再者，如需了解病毒进化对传播和致病能力的影响，我们必须将基因组数据与传播（病毒流行病学）和疾病数据（患者临床数据），以及人群大小的信息联系起来。在这篇社论中，我们探讨了病毒生物学中的关键方面，以及人类多态性、人类行为、地理和气候的相关方面对病毒进化的影响，并在此基础上提出了一系列建议，以监测病毒的进化并预测哪些类型的变异将会出现。

引言

当前的 COVID-19 大流行像极了一个实时发生的分子进化实验。打头天起，它就一步步向被蒙在鼓里的群众展示了病毒的行为和能力。这个“实验”几乎涵盖了整个地球 78 亿具有丰富环境多样性和遗传多态性的智人。SARS-CoV-2 β 冠状病毒的起源仍然不明 – 除了我们知道一个在蝙蝠中找到的与 SARS-CoV-2 进化距离相近的病毒 (Ji *et al.*, 2020; Makarenkov *et al.*, 2021) – 但现在，我们可以通过它在不同国家和人群间的传播特征，较准确地追踪其进化过程了。截至 2021 年 3 月 2 日，通过 Global Initiative on Sharing All Influenza Data (GISAID, <https://www.gisaid.org>) 数据库获得 193,687 个 SARS-CoV-2 全基因组，并将它们与首个公开毒株“Wuhan-Hu-1” [INSDC AccNum MN908947.3, (H. Wang *et al.*, 2020; Wu *et al.*, 2020; Zhu *et al.*, 2020)] 的序列进行比较，我们找到了 19,794 个包含一个或以上单碱基突变 (SNP) 的基因组位置。

在我们使用这些突变作为研究 SARS-CoV-2 进化和 COVID-19 发展的重要特征的工具的同时，需要保持警惕。因为其中一些突变可能只是源于糟糕的数据质量，而非代表真正的

病毒多样性。另外，病毒序列采集的地理分布不均匀问题可能导致部分观测到的突变频率偏离实际。识别跟踪具有行为变化的变异毒株（行为包括毒力、趋向性、和传播）是病毒基因组监测的关键目标。我们在 SARS-CoV-2 中看到，随着时间的推移，某些积累下来的突变加强了病毒的传播能力，某些突变也改变了病毒的毒力 (Oulas *et al.*, 2021)。收集更多毒性有所加强或削弱的毒株的突变数据有助于得到更准确的分析结果。但鉴于病毒变异、地理条件及人类多态性这些因素的变化范围之广，我们在分析这些突变时需保持谨慎（参见补充图片 1 及补充表格 1 的注解）。对这些因素进行全面分析的难度，从质量参差不齐的关于 SARS-CoV-2 和 COVID-19 的著作的大量发表（截止 2021 年 3 月 3 日在 PubMed 上已查到 109,000 个相关文献），以及病毒序列数据库对病毒元数据采集的不足，就可见一斑。在本社论中，我们尽可能采用显而易见的分析思路（以避免追求分析热点和依赖于文献统计所产生的偏差）来预期病毒和大流行病演变的特征。我们将通过分析病原体的各种功能及其与宿主的相互作用的两个基础病毒生物学手段（基于其两个原始功能-传播和生存-随着时间的变化）来重新审视 SARS-CoV-2 将遇到的可能会影响流行病的发展的进化限制。这将突出一些我们虽然正在目睹但任未受到广泛关注特征，这些特征可以帮助我们预测这场大流行过程中可能发生的状况。

关于进化的几个重要方面

从定义上讲，进化学是基于已发生的事实所总结出来的。“时间”做为其变量意味着，对一个生物体分别进行短期和长期的演化分析，将得出非常不同的结果。此外，进行进化结果的预测是冒险的，因为进化是短视的。它不具备对未来进行宏伟设计的能力。但是，从长远来看，某种生物能存活下来一定归因于某些使其能够在那段时间内继续繁殖的能力。从短期看则有所不同。其存活能力只与少数与繁衍后代相关的能力有关，而非使其有能力在遥远的将来继续传播的能力。一个病毒从一个已经与之共存了很长时间的宿主转移到一个新的宿主，将使该病毒突然暴露在一个陌生的环境中。但短时间里它将仍然遵循和发挥使其能在先前宿主中存活的能力。同样的，新宿主也是长期演进的结果。新宿主尽管对于新入侵的病毒还很无知，但长期的自然选择也使新宿主拥有各种常见的应激反应来应付入侵。就以针对病毒入侵而言，先天免疫能力将通过特征识别新病毒的存在，进而预防或减慢病毒的发展 (Nan *et al.*, 2014; Chen *et al.*, 2017; Hur, 2019)。许多这些能力是动物甚

至植物所共有的 – 意外发现 Toll 受体在果蝇中的作用就是个例子 (Belvin and Anderson, 1996)。

进化无法预先确定一个病毒在较远的将来能否继续产生后代。所以从短期来看，这意味着一个“成功”的病毒将忽略其宿主的生死并极尽所能产生后代（除非宿主在病毒产生后代之前就已经死去）。而从长远来看，为了能持续传代，导致高死亡率的强病毒毒力是不可持续的。这项病毒的基本特性将启示我们，随着疫情的推移，病毒可能会发生什么功能变化。不幸的是，对进化多样性理解的努力导致了好些过度简化的假说。这些假说以人类为中心并带有一些经济学和道德的特点：一个生物体的行为可被视为“无私的”或“自私的”。突变可被分为“有利的”、“有害的”或“中性的”。这些假说是否适用于病毒的进化值得探讨。以一个病毒的进化过程为例，能提高病毒传代能力的方法多种多样。比方说，即便一个碱基变化并不导致一个基因功能的改变，它也有可能影响宿主的代谢系统、小幅度地改进病毒的复制酶、干扰免疫系统、影响体温、改变重要 tRNA 的可得性、与另一种病原体发生协同感染等。在一场大流行中，这些方法可同时发生而互相之间并不一定有一致性。这将直接影响病毒的繁衍，并受到包含如上所述的各个过程和系统的约束。没有任何的一个突变是真正“中性”的，中性或类似的温和解释或许只是用于描述我们不了解的过程并掩盖无知的一种手段。如果我们想要预测这场大流行的走向，哪怕只是一小部分，我们也要避免被“中立”的看法所误导，即便“黑天鹅事件”是生物进化中常有的事 (Taleb, 2008)。

病毒基因组中一个的变化将产生怎样的影响将随着时间的推移而显现。从短期、中期或长期来看，其结果会有所不同。对我们而言，重要的不是我们要应用的演化模型，而是要阐明分别在流行病的扩张、达到顶峰和散去时，有可能会或一定会出现的一系列特征。从这方面来说，当前的流行病是高度可预测的，且实际上在许多研究中都已做过预测 (Moya *et al.*, 2004; Turinici and Danchin, 2007; Horby *et al.*, 2013)。与其使用模型来进行预测，我们的研究将尝试找到在病毒感染期间将可能发挥作用的功能，并理解它们将如何帮助我们推断该场流行病的未来。

功能分析：从功能到序列，而不是相反

众所周知，生物学中功能的概念复杂无比 (Allen *et al.*, 1998)。让我们在这里使用工业设计作比喻。为了实现一个有意义的设计，在工业设备和流程设计中首先要了解是与该设计相关的功能，即功能分析 [Fantoni *et al.* in (Norell Bergendahl and Stanford University, 2009)。一种常用的分析方法是自底向上，列出各个必要组件并将它们组合为更加复杂的部分，直到完成。相比之下，从最终用户的角度出发，一种更好的分析方法是自上而下。我们首先确定设备的主功能（例如，打印机的主功能是打印），然后逐步识别主功能运行所需的辅助功能（例如，进纸机构，供墨机构，供电机构，控制机构等），最终得到构成了该设备的基本组件。话虽这么说，但了解病毒的功能，尤其是其主功能并非易事。在这里我们提出病毒有两个必须的，截然不同但相互影响的主功能：繁殖（产生后代）和探索（到达宿主）。这两个功能的实现方式基本上是无限制的。但这并不意味着它们不会受到外在约束：它们的组成部分，与生命过程、特定构造单元、大分子相关。幸运的是，这缩小了我们的研究范围。由于病毒早在动物出现之前就已存在，动物早已设计出旨在对付病毒入侵的各种功能。

探索

让我们来聊聊与主功能中的“探索”相关的病毒功能和宿主反应。任何与生物有互动的个体的主功能在不同的层面上都离不开“到达宿主”这个辅助功能。对我们来说病原性病毒的目标是人类。病毒首次接触的是一个人的皮肤等表层，这规范了病毒进入宿主的途径。以下是几个常见例子：

- 面部/肠道的趋向性，这被我们有意无意地使用双手深深影响着（只需要看看周围有多少人在碰自己的脸就可以知道了）。这类趋向性可以根据部位具体分为眼睛、鼻子、呼吸道、嘴巴以及肠道，而且可能受到常见的粪口传播影响，在 COVID-19 疫情爆发时都有发现与这类趋向性相关的例子 (Jin *et al.*, 2020)。在废水中检测到的病毒反映了经肠道传播的重要性 (Sharma *et al.*, 2021)。但暂时发现通过这种途径感染 SARS-CoV-2 个案还比较少 (Goldman,

2020)。然而，我们必须意识到在对病毒进行预测的时候肠道传播有可能会突然变成了一个主导的方式，这一点在下文会进行讨论。

- 通过气溶胶传播的病毒有着直接的呼吸道趋向性。当感染集中在鼻腔或是深入肺部有着不同的严重性，所以我们应该谨慎地监测着这类传播途径。这也指出了病毒可通过不同宿主细胞进入身体。
- 通过媒介传播的病毒：普遍的传播方式是通过皮肤和血液。虽然我们暂且不考虑新型感染方式的风险，但不排除有更加复杂的情况已发生，譬如当病原体本身是寄生虫的宿主 (Ng *et al.*, 2007)。

COVID-19 大流行主要与上述前两种传播途径有关，但是我们应该注意到取决于人类习惯的变化（譬如放宽手部和脸部卫生的规范），使很容易为病毒的进化打开的新窗口。

关于人类宿主

COVID-19 有着很多人类社会疾病的特征 (Danchin, 2003)，即便只是暂时的，但人们也不愿作出行为上的改变 (Cherif *et al.*, 2016)。由于社会中存在着各种政治原因，以及不同社会体系对于公益的重要性有着不同的看法，悬浮微粒的作用一直存在争议，尤其是因为戴口罩主要用于保护他人而不是佩戴者 (Wei *et al.*, 2021)

（<https://interactives.lowyinstitute.org/features/covid-performance/>）。而在不同社区中对所谓的共同利益则有着不同的看法。然而即使在户外悬浮微粒也可以将病毒长时间保持在空气中 (Zhang *et al.*, 2020)，因此这很可能会成为 SARS-CoV-2 主要传播途径 (Dumont-Leblond *et al.*, 2020; Lieber *et al.*, 2021)。相关的争议导致出现大量关于配戴口罩的不当建议，因此有着极大的破坏性 (Czypionkae *et al.*, 2020)。而忽略这个传播途径的重要性则成了病毒长久存在的主要风险，长期而言会产生危险的毒株。许多人类的习性，譬如喜欢在拥挤的环境中进行社交活动，也促进了疾病的传播。以下几个有关感染源头的例子说明了各种触发感染的可能性以及病毒的传播方式和持久性。

环境变化

呼吸系统疾病有着很大的季节性 (Audi *et al.*, 2020)。但是那确切的因素与传染病之间真正的因果关系可能会被其他的混杂因素掩盖了。季节性、污染、城市化、生物多样性或纬度都被认为与传染病有关 (Wood *et al.*, 2017)。但事实上，温度或雨水可能会改变人类在室内外的模式，这人为因素很少被考虑进去 (Bulfone *et al.*, 2020)。感染的可能性与空气中的病毒浓度有关，因此从感染者所呼出的病毒的浓度是传播的关键，在拥挤的环境（包括室外）中高度密密集的人群必定会有着高度传染性 (Derjany *et al.*, 2020)。此外，与农村地区的开放环境相比，城市环境中的防风屏障会导致空气的停滞，对稀释病毒非常不利。颗粒物污染也与疾病的严重程度有关 (Brauer *et al.*, 2021)。譬如，紫外线在室外的分布会根据纬度和海拔发生巨大的变化 (Karapiperis *et al.*, 2020)。因为不同感染和传播途径的选择压力会有所不同，病毒的进化模式会在不同的环境下有所区别。因此单单了解衣壳和突刺蛋白是不够的，了解病毒包膜的作用也非常重要。

肺部和肠道的趋向性

就如其他冠状病毒一样，肠道和呼吸道在预测当前大流行的未来时也都至关重要。譬如猪传染性胃肠炎病毒 (transmissible gastroenteritis alphacoronavirus, TGEV) 在新生小猪的小肠的肠绒毛细胞以及肺部细胞中都可以复制，该病毒早在 1946 年对美国养猪场造成近 100% 的死亡率，对养猪业做成重大的打击 (Doyle and Hutchings, 1946)。另一次爆发是在 1984 年，猪呼吸道冠状病毒 (porcine respiratory coronavirus, PRCV) 肆虐 (Wesley *et al.*, 1990)。人们发现 PRCV 带有一些缺失突变的 TGEV 突变体，这些缺失使病毒的趋向性从肠上皮转为呼吸道上皮 (Rasschaert *et al.*, 1990)。值得注意的是 PRCV 感染能保护猪免受 TGEV 的侵害，形成了天然的疫苗 (Bernardet *et al.*, 1989)。从这个例子我们可以知道趋向性的变化与病毒毒力的减弱有关。这些观察结果表明应该把改变肠道健康的因素纳入监测范围。这种乐观的进化趋势极少发生并且与趋向性无关，并取决于一系列的巧合：先是病毒毒力减弱，紧接着有效地传播减毒变体，最后针对主要病原体进行免疫交叉

保护。尽管减毒变体可能与病毒传播同时出现，但是这三个事件在短期内自然地发生的可能性很小。不管怎样，让人倍感意外的是猪 TGEV 突刺蛋白 N 末端的两个氨基酸变化足以导致病毒丧失对肠道的趋向性 (Ballesteros *et al.*, 1997)，因此我们应该记住这点并积极寻找减低病毒毒力的方式。然而不幸的是，冠状病毒经常与其他 RNA 病毒进行重组 (Zhanget *al.*, 2005; Chenet *al.*, 2019)，这过程可能会阻碍减毒变体的有效传播（见下文进一步讨论）。

人类的多态性

我们近年才慢慢地开始意识到人类群组的遗传结构的特征，但这也是应该被纳入考虑的因素之一 (Williams *et al.*, 2021)。虽然在 3 月 21 日那一天于 PubMed 中并没有搜索到“human polymorphism” AND (“Covid-19” OR “SARS-CoV-2”)的任何结果。去掉引号后也列出了 281 条参考文献，而其中只有少数相关的文献 [譬如 (Ovsyannikova *et al.*, 2020)]。一项探索冠状病毒决定多态性的基因组学研究表明，东亚人群的祖先已经习惯了与人类相伴数千年的冠状病毒的感染，因此其后代可能没有那么容易受到当前流行病的影响 (Souilmi *et al.*, 2020)。

在到达目标细胞之前，SARS-CoV-2 必须附着并穿过一层层粘蛋白和其他分泌物。根据 Lewis 系统中不同方式的抗原分泌，可以把人类分为几种类型 (Lemieux *et al.*, 1979; Nordgren and Svensson, 2019)。基于各种血型的感染特征也对了解病毒有所帮助 (Bloch *et al.*, 2021; Le Pendu *et al.*, 2021; Schetelig *et al.*, 2021)。病毒还必须在体液和细胞水平上通过获得性免疫的屏障，而当中涉及到的人类多态性非常多。如果基于在分隔开来的人群中收集到病毒基因组序列进行分析，可能会得到在该人群中关于疾病传播的重要观察结果，使我们能够预测疫情的走向。就环境因素影响的获得性免疫而言，过去的感染案例可能已经有了新 SARS-CoV-2 病毒呈现的抗原决定簇的特征 (Ng *et al.*, 2020)。定义人类多态性主要标志物的人类白细胞抗原 (human leucocyte antigens, HLA) 在此范畴上可能起重要作用。这需要进一步深入研究，因为可能发生由 HLA 标记的特定病毒表位与宿主因子一致的事件，从而引发严重的自身免疫反应的情况 [请参见与特定 HLA 和 H1N1 流感病

毒有关的嗜睡症讨论组 (Schinkelshoek *et al.*, 2019)]。总的来说，获得性免疫反应贡献了大量的表型，这些表型与庞大数目的人类多态性所对应。只有通过收集大量病毒基因组序列，才能对此了解透彻（除了与炎症发生过程相关的常见特征），从而避免在预测中将获得性免疫的代谢后果与人类多态性的特定特征相关联。

血管紧张素转化酶 2（angiotensin-converting enzyme 2, ACE2）是病毒进入细胞过程中的受体 (Tortorici and Velesler, 2019)，这点将在下一部分中进行讨论。ACE2 基因在 X 染色体上，有着丰富的多态性。这个值得注意的特征暗示着性别二态性会影响感染。女性体内每个细胞中 X 染色体失活相关的随机变异会导致镶嵌多态性，这将可能导致一连串感染以及不同的疾病严重程度。在调查病毒的传播、严重程度和进化时，应该考虑到这种遗传特征 (Khayat *et al.*, 2020; Hamet *et al.*, 2021)。

到达细胞表面受体

病毒利用各种受体和宿主因子来结合靶标 (Baranowski *et al.*, 2001)。在冠状病毒中，与同源受体相关的趋向性类型的主要决定因素是突刺蛋白 (Hulswit *et al.*, 2016)。类似于前身 SARS-CoV-1 但又与使用另一种受体的 MERS-CoV 不同，SARS-CoV-2 的突刺蛋白 S 通过结合其主要靶标受体 ACE2（一种肾素-血管紧张素系统中必需的羧肽酶）介导病毒附着并进入宿主细胞 (Gross *et al.*, 2020)。ACE2 在心脏、肾脏、睾丸和胃肠系统中都有表达。在肺部，它在某些 2 型肺泡细胞中低水平表达，而该表达似乎是因人而异的 (Hikmet *et al.*, 2020; Zou *et al.*, 2020)。基于目前的状况，我们还注意到干扰素诱导了 ACE2 的表达 (Ziegler *et al.*, 2020)，揭示了感染过程中的前馈回路，这点或许解释了冠状病毒进化过程中使用 ACE2 作为受体的过程。

另一个有趣的特征是 ACE2 与跨膜丝氨酸蛋白酶 2（transmembrane serine protease 2, TMPRSS2）在鼻杯状分泌细胞、角膜、肺泡 2 型细胞、回肠吸收性肠上皮细胞、小肠上皮细胞和胆囊中共表达 (Lukassen *et al.*, 2020; Trypsteen *et al.*, 2020)。已经确定蛋白水解功能的参与对于病毒感染至关重要 (Laporte and Naesens, 2017)，而且冠状病毒突刺蛋

白 (S) 糖蛋白的蛋白水解裂解可激活糖蛋白进入宿主细胞 [参见下文 (Hoffmann, Kleine-Weber, Schroeder, *et al.*, 2020)]。在进行预测时可能会发现高度相似的病毒引起的疾病之间的差异。在 SARS-CoV-1 感染的情况下, ACE2 受体上的精氨酸和赖氨酸残基的裂解增强了病毒的感染性。这些残基对于通过 TMPRSS2 和人类气道胰蛋白酶样蛋白酶的裂解是必不可少的 [HAT, (Bertram *et al.*, 2011)]。而对于 SARS-CoV-1 而言 ACE2 裂解对于病毒 S 蛋白的激活是必需的 (Heurich *et al.*, 2014)。可能还有另一种蛋白酶作为受体: 二肽基肽酶 4 (dipeptidyl peptidase 4, DPP-4) 被认为是 SARS-CoV-2 的辅助受体, 但尚未得到进一步证实 (Badawi and Ali, 2021)。该病毒还使用细胞蛋白酶来促发突刺蛋白糖蛋白进入细胞。在翻译后进行修饰、蛋白激酶活性或各种类型的炎症细胞中它们的作用对于了解人类冠状病毒的传播至关重要 (Tharappel *et al.*, 2020)。无论如何, 随着流行病的发展必须仔细研究蛋白水解在感染初期的作用。它不仅参与了病毒与其宿主细胞的初始结合, 而且还介导了膜融合, 见下文进一步讨论。

病毒通过膜融合进入细胞

SARS-CoV-2 的感染有两个步骤: 当 S 蛋白与 ACE2 受体结合后, 病毒包膜与宿主细胞膜立即融合在一起。S 蛋白是同源三聚体, 每个单体由两个功能不同的亚基 S1 和 S2 组成。在冠状病毒中, 蛋白水解 (称为促发) 将 S1 亚基与 S2 亚基分离, 其中 S1 亚基包含受体结合域 (RBD), 促进与细胞受体的结合, 从而触发了膜融合 (Krueger *et al.*, 2001; Bosch *et al.*, 2003)。有效的融合需要 S2 进一步裂解成 S2' 片段 (Chambers *et al.*, 2020)。

令人感到惊讶的是, SARS-CoV-2 在 S 基因内部的 S1-S2 边界有一个编码多碱基位点基序 Arg-Arg-Ala-Arg (RRAR) 的插入片段, 这是病毒来源仍旧未知而引起的主要问题。这使病毒成为高尔基体枯草杆菌蛋白酶样 furin 蛋白酶的典型靶标 (Coutard *et al.*, 2020)。SARS-CoV-1 中并不存在此位点, 取而代之的是只包含单个 Arg 残基的片段, 这可能与多种蛋白酶的参与有关, 并直接影响病毒毒力和宿主选择性。这个插入片段也可以增强细胞之间的融合而不影响病毒进入细胞 (Andersen *et al.*, 2020)。为了证实该观点而进

行的功能分析发现宿主跨膜丝氨酸蛋白酶 2 (TMPRSS2) 和 furin 蛋白酶起着主要作用，它们共同促进细胞表面的病毒膜融合 (Hoffmann, Kleine-Weber, and Pöhlmann, 2020; Papa *et al.*, 2021)。S2 亚基的裂解引发 S 蛋白对 TMPRSS2 的 S2'位点进行加工，触发与邻近细胞的膜融合，并形成合胞体 (Buchrieser *et al.*, 2020; Papa *et al.*, 2021)。这突显了一个庞大的介导 SARS-CoV 病毒进入细胞的相互作用网络，为进化留下了很大空间。因此，监测这个区域中的突变并将其与临床数据相关联显得特别重要。可以通过该区域的变异来进行预测是因为该区域的变异很可能增强或相反地减弱病毒对人类细胞的侵袭能力。在不同的情况下，其他蛋白酶（包括组织蛋白酶和 furin 蛋白酶家族的成员）可能会帮助病毒通过内吞作用进入细胞。突刺蛋白的 S2 结构域会触发膜融合过程，这会使细胞的溶酶体触发内吞作用，从而使病毒进入细胞 (Ballout *et al.*, 2020)。和其他此处未谈及的病毒进入细胞的途径一样，这个过程会产生与溶酶体融合的胞内体，随后触发被包裹着的病毒基因组在细胞质中的释放 (Zhao *et al.*, 2021)并进入复制周期（在下文复制部分中讨论）。

有趣的是，“许多细胞融合成合胞体”是一种常见的的传染过程现象，例如在观察传染性支气管炎病毒的时候也能看到 (Yamada and Liu, 2009)。科学家还提出了一种有趣的生化活动来解释细胞中无膜隔室的形成 (Ditlev, 2021)。该过程由蛋白质中的所谓“无序”区域介导，该区域已证明可导致相位分离，这可能是无膜隔室的起源 (Shea *et al.*, 2021)。在冠状病毒中，该过程可能涉及 N 蛋白，特别是在脱壳步骤中，产生一个病毒特异性腔室，该腔室随着病毒的繁殖而逐渐变大，从而进一步利用了细胞机能及其代谢产物 (Dang *et al.*, 2021; Lu *et al.*, 2021)。除了涉及病毒基因组翻译的核糖体之外，与病毒蛋白相互作用的多种宿主蛋白也参与了隔室化过程，但是相关的文献记载仍然很少。

病毒特有的保护

自然选择为病毒提供了避免被物理化学或生物过程轻易识别和灭活的特征。冠状病毒被包裹着，这为它们提供了额外的保护层，但与宿主的相互作用又需要与脂质代谢相匹配。SARS-CoV-2 成熟病毒体的基因组编码着四种结构蛋白-突刺 (S)，包膜 (E)，膜

(M) 蛋白和基因组紧密核蛋白 (N)。我们已经讨论了蛋白 S 和 N 的作用。其他两种蛋白具有保护作用、调节作用和结构作用，用以塑造病毒体。此外，在形成病毒包膜时需要应对渗透压和/或电势的新功能。除了明显干扰宿主免疫反应以外，还解释了蛋白 E 以及可能与膜结合的辅助蛋白的离子通道活性 (DeDiego *et al.*, 2014)。

蛋白质对蛋白水解非常敏感，并且会自然地老化 (Truscott *et al.*, 2016)。翻译后的修饰有助于克服这些弱点。涉及细胞入侵的病毒蛋白被广泛糖基化有几个原因：协助蛋白质折叠、提高稳定性，最重要的是从宿主的免疫识别中屏蔽病毒 [这通常被描述为一个“聚糖屏障” (Walls *et al.*, 2016)]。糖基化水平是各种各样的，而在流感感染中非常重要的唾液酸化尚未针对 SARS-CoV-2 进行仔细研究 (Östbye *et al.*, 2020)。每个蛋白 S 单体有 22 个潜在的糖基化位点 (D. Wang *et al.*, 2020)。这些位点中大多数已被发现 *N*-糖基化，而假定的 *O*-糖基化位点的占有率则较低。因此，蛋白质中的相应氨基酸残基具有重要的屏蔽作用。非同义突变可能会影响突变菌株的后代，所以应优先监测涉及天冬酰胺的变化。

环境中病毒的成熟和释放

在探索的第二阶段，病毒从受感染的细胞、受感染的器官和有机体中释放出来，这一阶段的特征尚未被研究透彻。它可能在某些病毒变种的可传播性中起关键作用 (Lemey *et al.*, 2021)。宿主细胞的死亡释放了病毒体，但也可能套牢了它们。因此主动释放是病毒传播后代的更有效方法。长期以来这点已被病毒成熟后出芽所证实 (Garoff *et al.*, 1998)。冠状病毒出芽所涉及的宿主因子仍然有许多未知数，但已了解了胞质内的整体步骤顺序，既涉及内质网又涉及高尔基体 (Boson *et al.*, 2020)。与其他病毒一样 (Ortego *et al.*, 2007)，SARS-CoV-2 的 E 蛋白对于病毒成熟和释放途径也很重要。一旦复制完成，基因组就通过核蛋白 N 和蛋白 M 与内质网的脂质融合，驱动了病毒包装所必需的相分离过程 (Lu *et al.*, 2021)。组装完核衣壳 N/基因组复合物后，包膜将与蛋白 E、M 和 S 放在一起。病毒随后被运输至细胞表面，并通过非常规地劫持溶酶体的胞外功能以释放到环境中 (Ghosh *et al.*,

2020)。总的来说，我们再次强调蛋白酶作为病毒循环的重要特征是无处不在的，而我们必须仔细监控蛋白酶以预测未来的病毒进化。

最后，病毒蛋白与宿主蛋白的相互作用形成了各种稳定性的复合物。可以预料在成熟过程中，病毒包膜将包含病毒蛋白和在病毒体中的宿主蛋白。其中一些可能来自非特异性捕获，而其他则是由于它们与病毒蛋白的特异性相互作用，而更加系统地被捕获。虽然这步骤可能在准备下一轮感染的病毒中起到作用，但它通常被忽略。因为这些蛋白质，特别是病毒蛋白质，在感染后会立即注入受体宿主细胞中，所以这一点特别重要。这种不可避免的“污染”肯定是由进化形成的，是病毒操纵其宿主并加快启动病毒繁殖的一种方式。有着蛋白质结构的 CoV3D 数据库 (Gowthaman *et al.*, 2021) 是帮助我们通过这些病毒蛋白质介导的相互作用来预测潜在网络的重要资源。

进入细胞后，病毒基因组立即被翻译成两个大的多肽，它们被病毒蛋白酶迅速分解成有活性的非结构蛋白。将它们包含在病毒体中显然代表着一种选择性优势 (Haas *et al.*, 2021)。例如，SARS-CoV-2 木瓜蛋白酶样半胱氨酸蛋白酶（非结构蛋白 Nsp3 的结构域）对于病毒成熟、干扰宿主炎症和抗病毒免疫反应至关重要。已经通过使用质谱和蛋白激酶分析法研究了存在于 SARS-CoV-1 病毒体中的蛋白质补体 (Neuman *et al.*, 2008)。实验表明，除了蛋白质补体有着操纵宿主蛋白激酶的调节潜能以外 (Siddell *et al.*, 1981)，纯化的病毒体中确实存在两种病毒蛋白酶 Nsp3 和 Nsp5。找到多个大型跨膜蛋白 Nsp3 特别出乎意料，这促使我们去了解它是怎么整合到病毒体中的。另外，在成熟的病毒体中还已经发现了其他几种蛋白质，例如 Orf3a、Orf9b 和 Nsp2，目前对它们的功能了解仍然有限。病毒体成分的完整特征对于我们了解感染早期的动力学至关重要。

增殖

病毒结合到宿主细胞及通过细胞内吞作用进入细胞的过程中涉及使用 S 蛋白和其他刚刚提及的病毒蛋白，这情况仍未能完全被了解清楚。病毒进入细胞后必须拥有允许其自身复制的机制。由于 SARS-CoV-2 的基因组被病毒包膜包裹，它必须首先解开其包膜并

与几种蛋白结合（特别是衣壳蛋白 N），这样才能通过翻译产生所需要的酶供其基因组复制及形成新的包膜使用。

翻译机制

在进入特定类型的细胞后，病毒的 RNA 基因组需到达内质网中正确的位置并立即开始翻译。组成这活跃的起始复合物需要通过把长的 5'UTR 接合到核糖体 (Tidu *et al.*, 2020)。为使其进入细胞后能立即进行翻译，病毒基因组会形成 5'端以模仿标准的人类细胞的 mRNA (Yan *et al.*, 2021)。我们已经知道随后的翻译过程 (Hartenian *et al.*, 2020) 以及 SARS-CoV-2 病毒所产生的功能蛋白的数量 (Kim *et al.*, 2020)。但是，翻译过程中的步骤包括了许多未知功能，其中的一些功能可能对病毒未来的进化过程有着关键的影响 (Neches *et al.*, 2021)。除了核糖体以外，亦甚少有文献记载翻译过程中宿主蛋白与病毒蛋白的相互作用。

SARS-CoV-2 十分强大遗传的编码指导着 26 种蛋白质的合成 (Finkel *et al.*, 2021)。其中较为特别的是病毒 RNA 非对称的翻译过程：在脱离包膜后，前三分之二的基因组即从一个大的编码区中翻译成 Orf1a 和 Orf1ab 两个多肽，两者虽从同一 RNA 序列翻译出来但产生的数量并不平均。而 Orf1ab 的 RNA 是利用了假结结构和-1 移码翻译了其 C 端区域。由于冠状病毒的翻译无处不在，导致 Nsp 蛋白的非对称翻译而创造了一个特定的环境去精准地控制 Orf1a 和 Orf1ab 蛋白的比例或延迟 Orf1ab 的生产（其中包括 RNA 依赖性 RNA 复制酶，RdRp，Nsp12），这会一直发生直至 Orf1a 多肽（Nsp1-11）创造出合适 RNA 复制的环境为止 (Fehr and Perlman, 2015)。这些多肽随后被分成 16 种非结构蛋白 Nsp1-16（Orf1a 末端的 Nsp11 可能没有真实的功能），而其中一些则如我们所见存在于离开宿主后的病毒体中。转录组的远程部分用以下所述的机制被翻译成单独的蛋白质，包括病毒体主要蛋白质 E，M，N 和 S。

检测大量的病毒基因组序列有助鉴定在其编码中所有导致蛋白变异的突变。图 1 显示了以计算最高突变率和最低突变率之差后所得的每种病毒蛋白里所有突变的最高等位基因频率（AF）的分布。而补充表 1 提供了每种蛋白的突变数目。除了 S 蛋白预期具有高

度可变性以外，RdRP（Nsp12）也同样具有很大的可变性的事实告诉我们寻找针对病毒复制的抗病毒药物可能是一项艰巨的任务。这点同样适用于蛋白酶 Nsp3，尽管它有着多结构域可以提供了多个靶标，看上去作为抗病毒药物目标完全不是问题。虽然 Nsp2 的功能暂时仍然知之甚少，但它那些高度变化的范围让它特别受关注。而 Orf3 的变异则特别令人感兴趣，因为该蛋白可能通过维持一个对渗透压的正确反应来控制病毒包膜的稳定性。

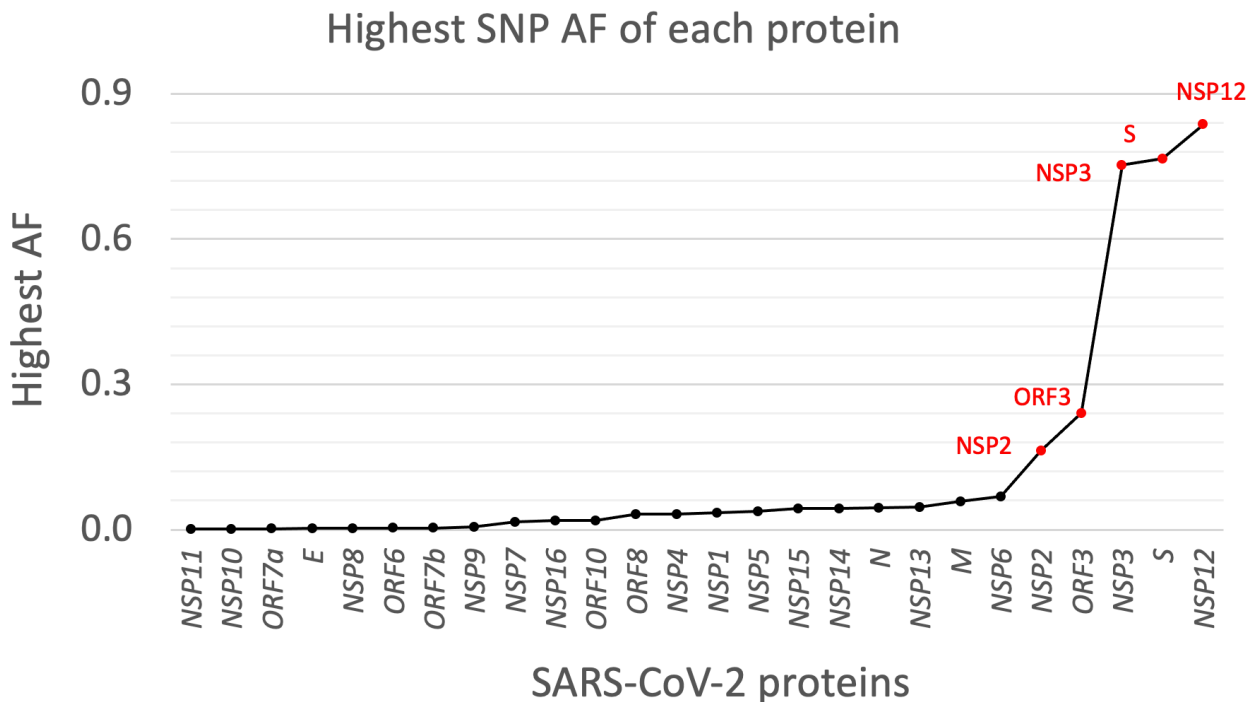


图 1. 由统计 193,687 个 SARS-CoV-2 毒株得到的每种蛋白质里所有突变（包括 SNP 和缺失突变）中的最高等位基因频率（AF）的分布。AF 大于 10% 的蛋白质由红色标签标记。y 轴显示在该蛋白质中观察到的最高的突变率（即最高的等位基因频率）。

控制翻译的过程是病毒发展中十分关键的一环，这与是否有着精确的 tRNA 补体直接相关。值得注意的是单靠人类基因组并不能编码完全的 tRNA 分子，前体 tRNA 必须先要在细胞核中成熟，然后在其 3' 端添加 CCA 补体 (Augustin *et al.*, 2003)。为了让 tRNA 适应目标基因的密码子使用偏好还需要控制许多 tRNA 核苷酸的修饰，这导致了病毒进化的瓶颈

并可能通过病毒编码的蛋白反映出来。虽然相关功能尚未被研究，但是监测那些导致病毒蛋白使用密码子的偏好发生重大变化的突变将有助于我们预测该流行病的发展。

蛋白水解

正如上文中所强调的，利用多蛋白去进行蛋白水解是许多病毒的重要功能 (Yost and Marcotrigiano, 2013)。Nsp3 的木瓜蛋白酶样蛋白酶结构域在合成后会立即识别 LXGG 氨基酸序列和切割 Orf1a 及 Orf1ab (Barretto *et al.*, 2005; Gao *et al.*, 2021)。这蛋白的 3D 结构是已知的 (Gao *et al.*, 2021)。特别的是该切割位点序列也可以被细胞去泛素化酶所识别，使得 Nsp3 能够干扰其宿主的调节功能。该蛋白酶还会选择性地切割干扰素调节因子 3 (IRF3)。在 SARS-CoV-2 感染期间这种切割可能导致了小程度的 I 型 IFN 应答 (Moustaqil *et al.*, 2021)。由于多种重要功能都使用着一样的识别位点，这种蛋白酶进化的可能性极受限制。另一种病毒编码蛋白酶 Nsp5 通过识别 [AVPT][VTKRM][LF]Q[ASN] 序列，在 Nsp3 末端切割 Nsp4，并在 Nsp4 末端切割后释放自身。随后 Nsp5 切除了从 Nsp6 至 Nsp16 的所有 Nsps。与 Nsp3 的情况一样，Nsp5 调控宿主蛋白的分解，特别是有效的炎症缓解剂 NLRP12 (Normand *et al.*, 2018)，以及炎症反应相关的转化生长因子 β 激活激酶 1 结合蛋白 (TGFbeta activated kinase 1 binding protein 1, TAB1) (Xu and Lei, 2020)。这指出了增强细胞因子生产和炎症反应的分子机制 (Moustaqil *et al.*, 2021)。在这里，再一次显示了病毒进化影响疾病严重性的可能性仍然有限，尤其是在“long covid”的发展情况下。反之，这也突显了这些蛋白酶作为良好的药物靶标的可能性。

转录机制

通过使用从 Orf1a 和 Orf1ab 所翻译出来的 Nsp 蛋白，特别是 Nsp12 和 Nsp13 (Arya *et al.*, 2021)，病毒基因组末端的三分之一随后被表达为个别蛋白的转录本，而这些蛋白在促进病毒复制的过程中会减少宿主防御系统的活性。病毒的 3' 端序列被转录成的多个会被翻译成个别蛋白的转录本，上文中已讨论过当中的结构蛋白 E、M、N 和 S。N 蛋白的低

聚状态对于复制和转录都是必需的 (Ahamad *et al.*, 2020)。而转录本作为病毒的完整基因组会被加帽。mRNA 加帽复合物包括 Nsp14、Nsp16 和 Nsp10。Nsp16 与 Nsp10 关联并于 2' 位置甲基化形成 cap-0 型帽子结构，在病毒的 5' 端形成 cap-1 型的功能结构，通过这些过程它会被识别为真正的宿主功能性 RNA，因此并不会受到抗病毒反应的攻击 (Perveen *et al.*, 2021)。正如在其他病毒蛋白所观察到的，这些蛋白可能有着其他功能，进一步限制了它们的进化能力。Nsp9 是另一种 RNA 结合蛋白，但其功能仍未被完全了解（请参见下文“质量控制”的部分）。除了加帽外，转录起始也需要将其 RNA 折迭成具病毒特性的 3D 结构 (Madhugiri *et al.*, 2016)，当发生变异时，这结构不会形成有效的病毒基因组（下文中提及的 C241U 突变是一个例外）。重要的是，形成病毒形态涉及几个必须的转录步骤，这揭示了转录控制区中的序列进化受着很大的限制。于病毒复制或转录起始序列缺乏对突变的耐受性相反，蛋白质编码序列在积累突变后仍然允许合成有效的病毒基因组。但是这点取决于蛋白质，有些蛋白倾向接受突变，而另一些则相反 (Nagy *et al.*, 2021)。

复制机制

复制病毒基因组是病毒增殖的核心。对冠状病毒而言，它们是正单链 RNA 基因组病毒，它的复制过程应该是十分不对称的，而病毒基因组的互补拷贝的数量要比基因组本身少很多。这个不对称的病毒复制模式尚未研究透彻。这可能受到许多与 RNA 种类有关的 RNA 复制机制的因素影响，而这些复制机制与内质网有关。如上文所讨论有关病毒翻译的部分，Orf1a 和 Orf1ab 的蛋白质的表达是高度不对称的，这创造了一个合适 RNA 复制的环境 (Fehr and Perlman, 2015)。SARS-CoV-2 的复制程序的启动需要使用蛋白 Nsp7 和 Nsp8 才能从病毒的 3' 端结构起动。随后，再次在 Nsp7 和 Nsp8 的帮助下 RdRP 的主要亚基 Nsp12 会继续复制基因组 (Gao *et al.*, 2020)，而 Nsp13 则帮助 RNA 解链。由于与 RNA 种类有关的 RNA 聚合酶因为决定了复制过程的效率和准确性，所以有着独特的重要性，亦因此已被选作抗病毒分子的靶标。然而，如图 1 所见，Nsp12 中仍保留了许多突变，这表明了病毒在进化过程中得到了许多解决方案来躲避药物的抑制作用，说明 Nsp12 可能并不是抗病毒药物的良好靶标。另一个亚基 Nsp15 是在冠状病毒中十分保守的 U 特异

性核酸内切酶 (Pillon *et al.*, 2021)。它的功能尚未完全被了解，但似乎与躲避宿主的抗病毒反应有关 (Zhao *et al.*, 2020)。因此，它的进化用于预测病毒毒力的未来发展趋势十分有用。

质量控制

复制和转录过程缺乏足够的准确率来限制长达 30,000 个核苷酸的 SARS-CoV-2 基因组中出现错误 (Bradwell *et al.*, 2013)。在这种选择压力下，病毒衍生出来一些防止在每个复制周期中积累太多错误的校对复合体。这点十分重要，因为错误率过高会导致形成快速灭绝的缺陷病毒 (Pauly and Lauring, 2015)。而降低 RNA 病毒复制的准确率则是驱动某些核苷酸类似物（例如 favipiravir）的药物的设计理念，这种药物可导致突变，以至突变后的病毒后代不能存活 (Baranovich *et al.*, 2013)。Nsp14 既是产生 cap-0 型结构的鸟嘌呤-N7-甲基转移酶，又是可用作消除基因组复制过程中出现的错配的校对 3'至 5'核酸外切酶 (Ogando *et al.*, 2020)。该蛋白的许多突变会增加病毒基因组的突变负荷 (Eskier *et al.*, 2020)。研究这种双功能酶中突变的结果已经揭示了新型病毒谱系“绽放”般形成，其中一些可能会导致病毒毒力减弱 (Cluzel *et al.*, 2020)。

病毒的蛋白水解活性的质量控制可能在其长期存活中有着重要的作用。值得注意的是 Nsp9 二聚体在体外实验中会与肽 LEVL 结合，这与 Nsp5 蛋白酶切割位点相似 (Littler *et al.*, 2020)。可是，该蛋白的主要功能也可能涉及调节分子伴侣的活性。尤其是在压力下（例如在病毒感染期间），这些因素是关键的质量控制素。由此推论，SARS-CoV-2 可能会通过翻译后的修饰来调节这些功能。与此一致的是，谷氨酸残基的核苷酸化修饰了 Nsp9（在体外实验中略微偏好 UTP），这通过一个和 Nsp12（上文讨论过的取决于 RNA 种类的 RNA 聚合酶）的 NiRAN 域相关的锰依赖性活动进行。该残基属于保守的 N 端 NNE 三肽。而它是在冠状病毒 Nsp9 的同源中唯一不变的残基 (Slanina *et al.*, 2021)。有趣的是，对应 Nsp9 的鼠伤寒沙门氏菌的 SelO (YdiU) 蛋白质是在生命树的三个域中都十分保守的。该蛋白是线粒体中的一种硒蛋白，证实了它在控制氧化应激中的作用。它在

ATP 有限的情况下通过尿苷酸化去修饰分子伴侣 (Y. Yang *et al.*, 2020)。这种核苷酸化作用应该从 C>U 趋势的角度被优先探讨，以下一节将讨论其起源。

预测未来

谁都知道预测未来是困难的 (Sullivan, Gregory F, 2013)，尤其是考虑到人的行为的不规律性和地域性。但在许多情况下，我们必须做出我们最好的估计。好的估计有积极的作用，且即使是外行人也可以做出有趣的估计。举个例子，Andrew Wright 在 2020 年初 COVID-19 大流行之初，就在其小说 *The End of October* 里颇具几分真实地预测了大流行的发展。管理一场大流行并使其尽可能短而少害可以挽救数以百万计的生命和减轻其所导致的巨大的经济负担。在总结了大量的有关病毒的知识与文献，并筛除了好些错误信息后，我们在此着重讨论对于 SARS-CoV-2 的进化而言，它通过自然选择的方式所探索过的或多或少受到限制的进化空间。

在第一年大流行的普遍进化模式

尽管大量论文和在线出版物中都有讨论，但公众对 SARS-CoV-2 的反应却存在极大的偏差和局限性，更不用说那些本应该被紧急考虑的方面。公众讨论保持集中在病毒那些最明显特征上，例如包膜上的突刺蛋白（下文将进一步讨论它的部份作用），以及在缺乏信息的体外细胞模型中的假性治疗方法。这一点可以被简单地使用 Google 搜索“SARS-CoV-2”“spike”反映出来，在 3 月 3 日搜索的结果包含 6,000,000 页（2 月 15 日的结果为 9,000,000 页，这见证着热潮的力量，我们由此看到一个下降的趋势）。而搜索“SARS-CoV-2”“Orf8”的结果仅包含了 50,400 页。然而即便不是全部，但绝大多数病毒蛋白质仍在持续进化 (Jaroszewski *et al.*, 2020)。如图 2 和补充表 1 所展示，病毒基因中的突变数目要比在它的突刺蛋白中的多很多，而且蛋白之间突变的数目也存在相当大的差异。如本文第一部分所述，稳定这些变化的选择压力是由多种原因引起的，它们中的一部份是可以通通过人为干预的。那些影响多个病毒分离株的那些突变（直方图中的最后一列）可能是病毒传播中最重要的突变。

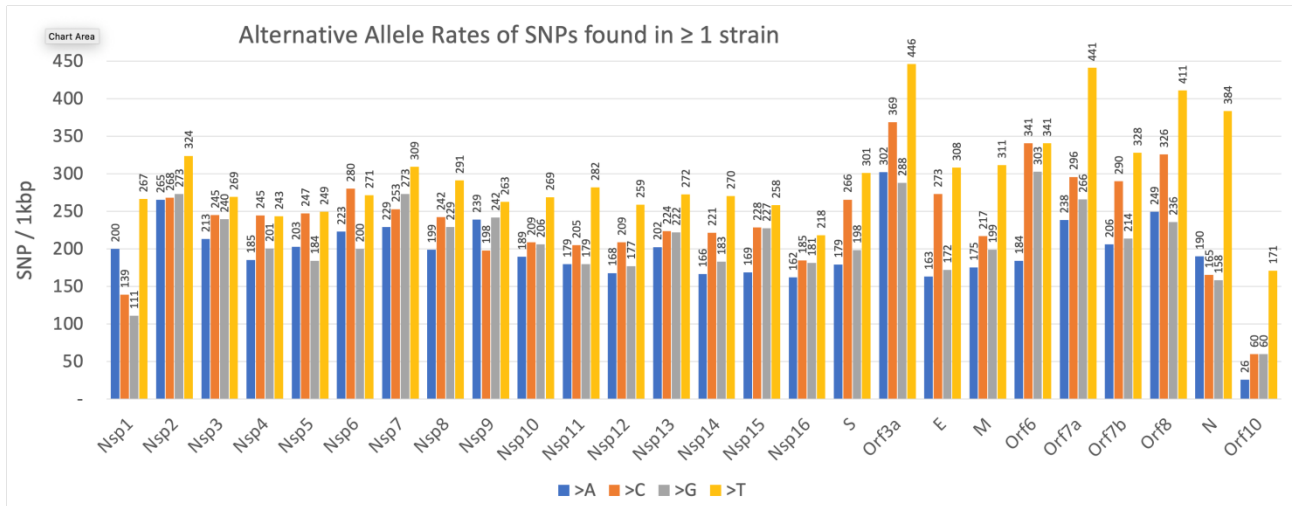
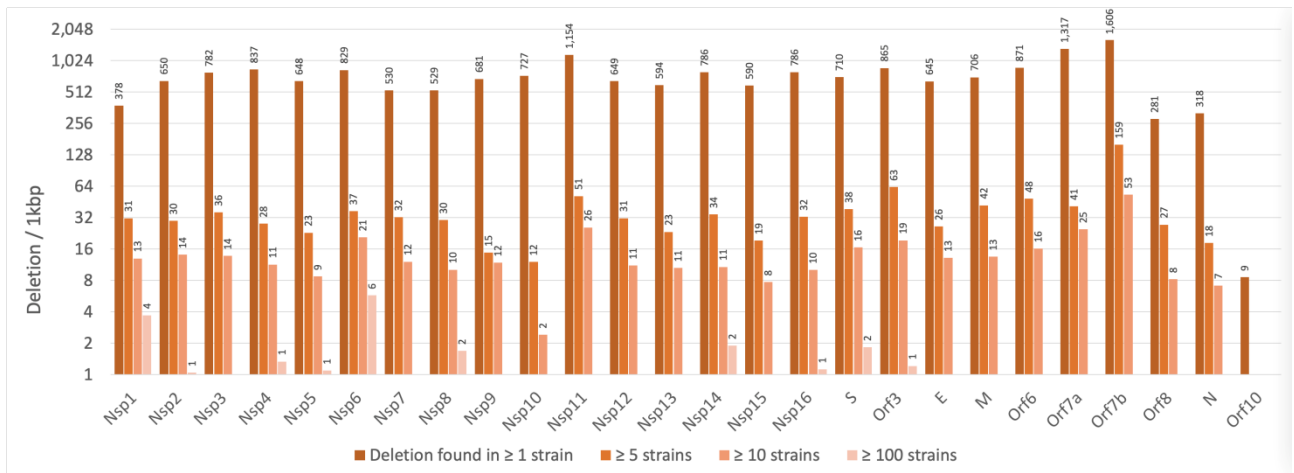
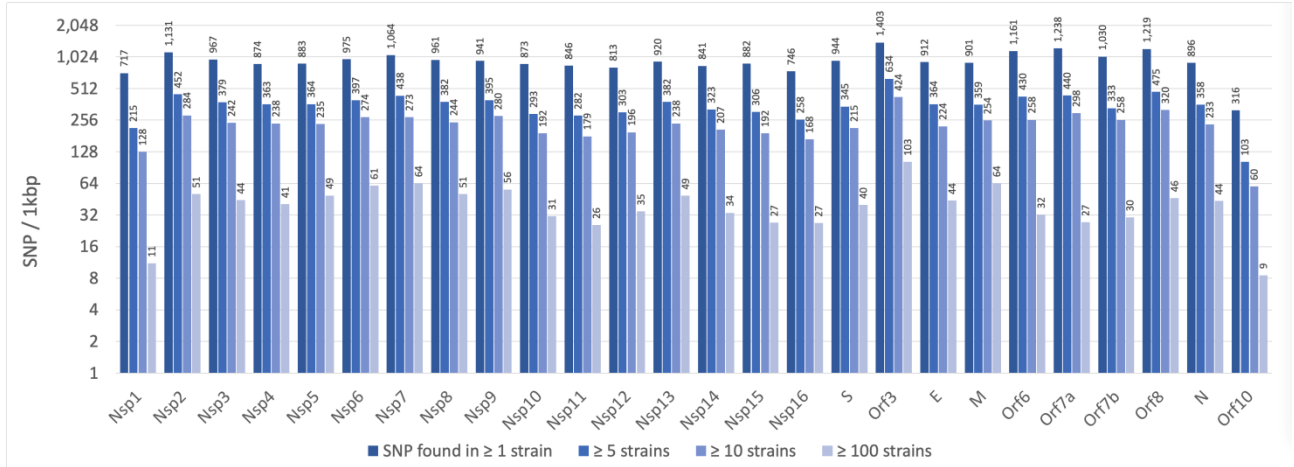


图 2. 总结了来自 193,687 个 SARS-CoV-2 毒株的每个基因的突变率。(a) SNP 率, (b) 缺失率, (c) 在 ≥1 个毒株中发现的 SNP 的突变碱基频率。‘>A’ 表示 SNP 从非 ‘A’ 突变到 ‘A’。y 轴已经根据每个基因的长度标准化为每 1kbp 的突变数。在 (a) 和

(b) 中，y 轴使用了以 2 为底的对数刻度，并且暂时了在 ≥ 1 株， ≥ 5 株， ≥ 10 株和 ≥ 100 株中的突变率。补充注解中记录了这些每个基因突变率是如何分析得到的。

以下是一个具体的例子去帮助我们预测病毒进化的未来变化。Nsp1 蛋白是从多蛋白 Orf1a 和 Orf1ab 中第一个分离出来的病毒蛋白。出乎意料的是，该蛋白质中的突变数量相对其他相同大小的 Nsp 蛋白都少得多，这表明了长期进化过程中有着很大的选择压力。Nsp1 对病毒基因组翻译的开始有着关键的作用。其翻译特征有别于该病毒的其他蛋白质的特征，涉及从大量的细胞 mRNA 中区分出病毒 mRNA 基因组 (Ou *et al.*, 2020)。因为它为了避免变异承受着相当程度的压力，因此应优先分析保留在其序列中的少数突变以及报告它们对毒力的间接影响。最后，正如前文所述，稳定这些变化的选择压力是由多种因子所引起的，其中一些可能可通过人为干预。收集相关数据支持基因组信息的充分利用的需求十分迫切。

环境因素

在上文中已有强调，许多混杂因素会影响我们对呼吸系统疾病的理解。然而，我们已经充分了解到拥挤的环境，包括但又不局限于室内，对人与人之间的传播有着显著的贡献。其中通气系统的结构对此可能会产生很大的影响，这说明了对于建筑物的通气方式立刻进行深度研究的必要性，这包括使用空气过滤系统以及即便在户外环境中也要加强社交距离。为了控制将来呼吸道疾病的流行，在居住区、餐厅和办公楼以及公共交通的设施中应服从控制空气流动的强制性规定，就好比在手术室中一样。此外，气候也可能对病毒的进化有着重要的影响。B. Chen *et al.*, 2020 等研究并没有提供太多的资料，因为有相当数目的混杂因素影响流行病的发展。气候会影响人类的行为：比如我们倾向于在天气寒冷时留在室内。尽管大多数情况下仅仅是推论，但室内与室外感染传播率是最佳的记录参数。传播的参数与外界温度有关，当温度或湿度较高时，空调可能扮演着一个负面的角色。因此，封闭空间的通风系统 (Pease *et al.*, 2021) 可能会严重地影响着病毒的进化。这表明了特别监视其中几个功能的必要性。为了帮助进行预测，应对元数据作出全面调查以识别在封闭空间中的感染个案，这个调查应该与基于基因组测序的调查联系起来。基于基因组测

序的调查包括病毒包膜及其相关蛋白的结构、它们与脂质的结合、小的疏水性蛋白的进化，以及与离子回收利用和运输的关系。另一个或许应该探索的参数是紫外线辐射 (Karapiperis *et al.*, 2020)。在可以杀死病原体的同时，紫外线也可以以一种特定的方式导致诱变 (Wurtmann and Wolin, 2009)。那些更为标准的诱变核苷酸的修饰会生成尿嘧啶环丁烷二聚体。如果效果显著的话，这将令病毒的进化高度偏倚，这点值得研究，或许可以通过分析从较高 UV 的区域分离出的变异毒株的突变模式的偏差。然而，我们在文中只论及在病毒中一些鲜明的特征，这些特征可能会对预测病毒未来的进化做出最大的贡献而且适合通过实验方法研究。

普遍来说，正如大量有关该流行病的文献所指出，重点主要放在宿主对抗病毒的调控上。以这观点有助让我们把人类习性与感染联系起来。例如，由于吸烟者可能患有更多的呼吸道疾病，有机会因而与先前感染的病原体发生交叉反应，因此可能会产生一定程度的保护 (Saurabh *et al.*, 2021)，这就能解释了初期有关对吸烟可能产生保护作用这种矛盾的观察 (Landoni *et al.*, 2020)，并已确被定义为严重性风险因素之一 (Elliott *et al.*, 2021)。多种病毒同时感染也可能是病毒之间重组的源头，而下文讨论的特征可能会产生重要的结果。这个特定的行为在预测病毒在部分人口中进化时必须被考虑进去 – 应该再次利用整个基因组的序列来进行监测 – 通过联系先前的呼吸道和肠道感染纪录，而且应该再次地在利用整个基因组的序列来进行监测。

在一般情况下，患者临床数据可能会揭示病毒趋向性的进化，特别是与呼吸系统和消化功能相关的数据，这些数据作为与基因组联系起来的元数据而被纪录下来是很关键的。

取决于宿主的病毒发展的局限性

把病毒和它宿主的代谢系统联系起来对于了解它的短期和长期进化过程十分重要。病毒在进行增殖时必须将其宿主的代谢系统转向帮助自身的复制，换句话说病毒正操纵着宿主细胞的新陈代谢。因此病毒必须适应宿主的代谢系统的约束。在这方面，SARS-CoV-2 是一个非常有趣的生命体。基于它一些被忽略的生长特性，它增殖的方法体现了令人着迷的细胞代谢的普遍特性。细胞的生长过程中的大部代谢反应都在细胞质中发生，而细胞生

长所需的所有基本元素都会在细胞质中产生。这里引起了一个常被忽视的“非同向”增长的问题，类似的概念有常被套用在经济学上。细胞是在三维中生长的，而细胞膜则是在二维的表面上生长，更重要的是基因组是个一维增长的线性聚合物。这意味着，制造“过多”的膜会带来巨大的代谢压力，而制造“过多”的基因组会带来更大的压力。这些功能原则上对非包膜病毒及包膜病毒的增殖都有着莫大的益处。这些物理难以解释的现象形成了细胞演化中的和谐方式需要加以阐明。

实际上，通过利用整个代谢过程去合成三磷酸胞苷单分子[这对应着在基因组序列中的“C”]，上述一障碍已通过物竞天择的过程解决了 (Danchin and Marlière, 2020; Ou *et al.*, 2020)。嘧啶代谢中一个出乎意料的特征证实了这一观察的结果，这是通过一个不符合预期的酶活性的普遍缺失而证实的。虽然磷酸核糖基转移酶在细胞中无处不在，并可将其将嘌呤、嘧啶和其他杂环碱基拆成单核苷酸，但至今尚未在任何活生物体中找到胞嘧啶磷酸核糖基转移酶 (Ou *et al.*, 2020)。因为与它相关的酶无处不在，这使胞嘧啶磷酸核糖基转移酶的缺失显得十分重要。由于随机突变出现在现有同类酶中，导致该酶的缺失会持续出现，所以它的缺失一定与回收胞嘧啶的强烈的自然选择背道而驰。SARS-CoV-2 基因组中胞嘧啶有着下降趋势的结果是鸟嘌呤与之相似的缺失趋势，因为在复制过程中 G 和 C 是互补的，导致了病毒中 G 含量的相似下降。如下文所论述的，由于鸟嘌呤是涉及活性氧 (ROS) 的特定防御反应的靶标，这将衍生出有趣的结果。

因此，宿主在进化过程中保留了一些代谢功能，通过操纵 CTP 合成来干扰病毒的发展，这种现象也是不足为奇的。通过 viperin (干扰素可诱发的内质网病毒抑制蛋白) 合成类似 CTP 的抗病毒物 3'-deoxy-3',4'-didehydro-CTP (ddhCTP) 的过程就如化学家人工合成的抗病毒核苷酸一样引人入胜，类似的方式普遍存在于哺乳动物的先天抗病毒机制中 (Kang *et al.*, 2020)，但在牡蛎 (Green *et al.*, 2015) 甚至在细菌和古细菌中也能找到 (Bernheim *et al.*, 2020)。该抑制剂同时阻碍了 CTP 的四个关键功能：RNA 合成、在 tRNA 的 3'端添加 CCA、制造胞嘧啶核苷酸依赖性的膜脂以及制造作为蛋白质糖基化底物通用载体的合成磷酸盐 (Ou *et al.*, 2020)。这种代谢瓶颈的后果是 SARS-CoV-2 在低胞嘧啶含量的情况下进化 (见下文)，除非宿主中有专门的代谢设置以减轻其限制。

因此，无论是在 RNA 还是在氨基酸水平上能够保持不变的胞嘧啶残基对于病毒感染的永续性可能是十分重要的。这代表着它是 COVID-19 抗病毒药的重要靶标。因此，监测病毒的不同谱系在其短期的进化过程中如何应对这一局限性至为重要。这也可能包括胞嘧啶含量下降趋势的逆转。该逆转意味着改变 CTP 合成酶的控制和抑制 viperin。这样亦拓宽了病毒发展的可能性。这一点再次说明了尽量收集更多完整的病毒基因组序列的重要性。

当前对各种物种中的病毒的比较系统发生研究的大量增加，有助缓解对深入了解免疫和代谢系统之关系的迫切需要。在这方面，与不同代谢系统，尤其是对蝙蝠的代谢系统作比较可能十分有用，并有助于我们预测一些病毒进化过程中的限制。有趣的是这些动物会以非毒性的形式与病毒共存，以进化出飞行的蝙蝠为例，被选择出来的系统可能有着快速新陈代谢的特征 (Shen *et al.*, 2010)。

飞行是一种能量消耗极高的活动，其间会产生大量的 ROS。然而有趣的是，ROS 通过巨噬细胞和嗜中性粒细胞利用呼吸/氧化爆发来提供对病原体防御的积极作用 (Piacenza *et al.*, 2019)。尽管很少研究，但快速代谢的正面效果可能有助于在蝙蝠中病毒感染的无害化。然而，因感染引起的抗病毒反应可能导致恢复氧化还原的稳态失衡，引致 ROS 及促氧化剂细胞因子的释放不受控制，以及因过度炎症而引起失控病理。调节该过程与肺部感染中的多个事件特别相关，这是严重 COVID-19（当然还有“long covid”）的普遍特征。同时，过量的 ROS 也会对病毒基因组产生相当大的影响，其中过量的 ROS 会导致 8-氧鸟嘌呤的形成，这会在复制过程中触发 G>U 碱基颠换突变。因此，冠状病毒可能已发展出防御机制以减轻这种化学攻击造成的负担。由此可见，监测病毒中特定谱系中一致的碱基颠换突变十分重要 (Cluzel *et al.*, 2020)。

病毒的短期进化：提高传播性

病毒在新感染不同物种的短期内必须适应其增殖的速率以及感染新宿主的能力。在这期间有利于所有相关的基因的突变积累。流行病最可能的早期发展是该病毒开始提高其传播速度。这种发展方向源于两个基本的功能。最有可能的是提高其传播性，这跟许多社会

参数都有着密切的关系，包括拥挤的环境，或从受感染者所产生稳定而长期的感染源，及潜伏期的增加等。第二个关键特征是病毒的复制成功率增加，这高速的增殖率通常是与增加了面对突变负担的能力有关。

在 SARS-CoV-2 入侵人类的第一阶段，趋势是平均每年每个基因组约 22 个突变。与上述情况一样，除大多数蝙蝠病毒的基因组中有着尚未解释到的较高 C 核苷酸含量外，其他观察到的病毒基因组的 C 核苷酸含量普遍有下降 (Matyášek and Kovařík, 2020; Ou *et al.*, 2020; Simmonds, 2020)。这个网站 (<http://www.bio8.cs.hku.hk/sarscov2>) 一直有在跟踪新菌株中 C 的百分比 (Luo *et al.*, 2020)。允许不断出现突变也是因为其蛋白质序列的功能多样性和弹性。在大流行的头六个月期间，有研究监测该蛋白质组的进化 (Lubin *et al.*, 2020)。大量突变导致病毒蛋白中的氨基酸发生变化，这表示当中含有高度的正向选择 (Cluzel *et al.*, 2020)。尽管病毒能适应新的宿主细胞并不意味着它将导致严重疾病，但有些变化会影响 COVID-19 严重性。我们应该特别注意的是，去除 C 残基会使 ddhCTP 干扰效率降低，因此这可能会增加病毒的毒力。图 3 展示了该病毒的所有蛋白随着时间的变化。在大流行的头 8 个月（从 2019 年 12 月至 2020 年 7 月），每个毒株的平均突变率以每月约 1 个突变的速率增加，但随后便以相对较低的每月约 0.5 个突变速率增加（从 2020 年 8 月至 2021 年 1 月）。

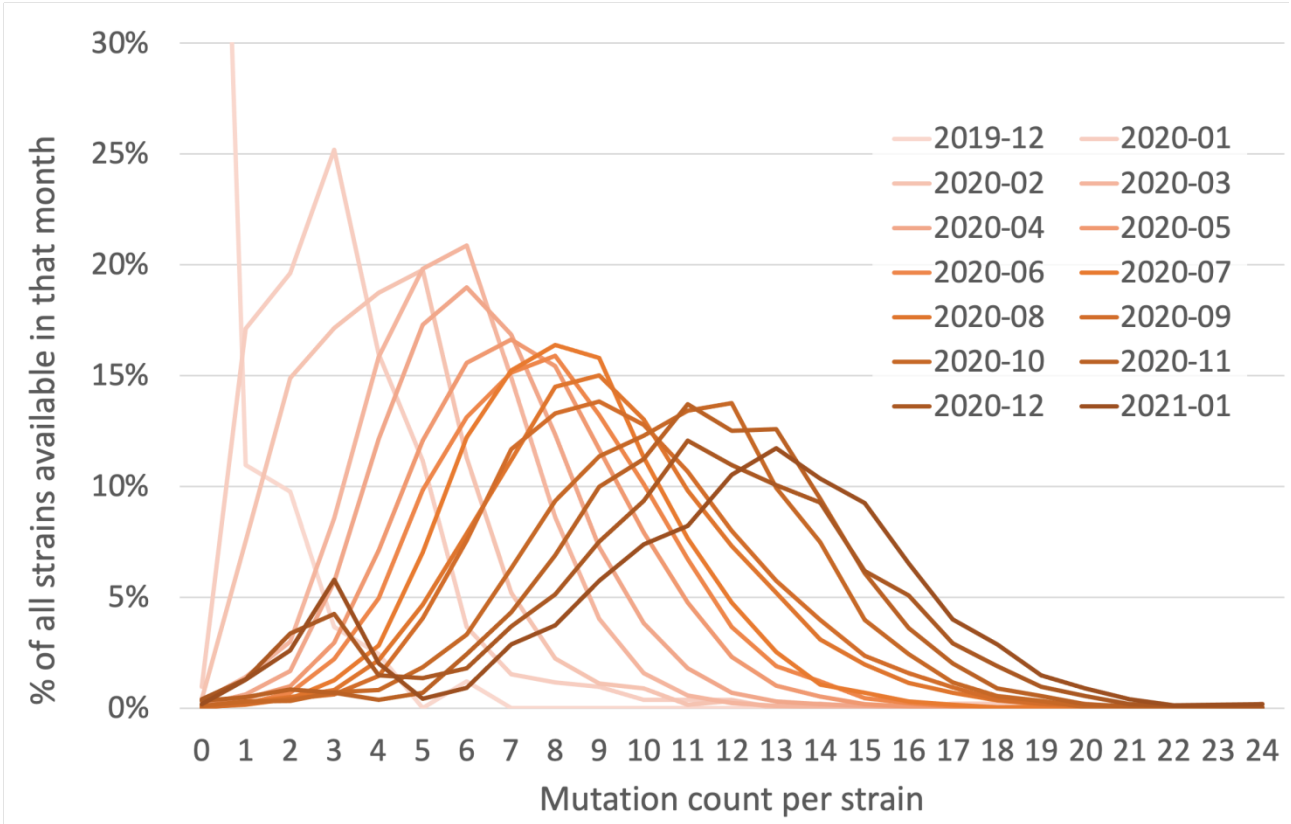
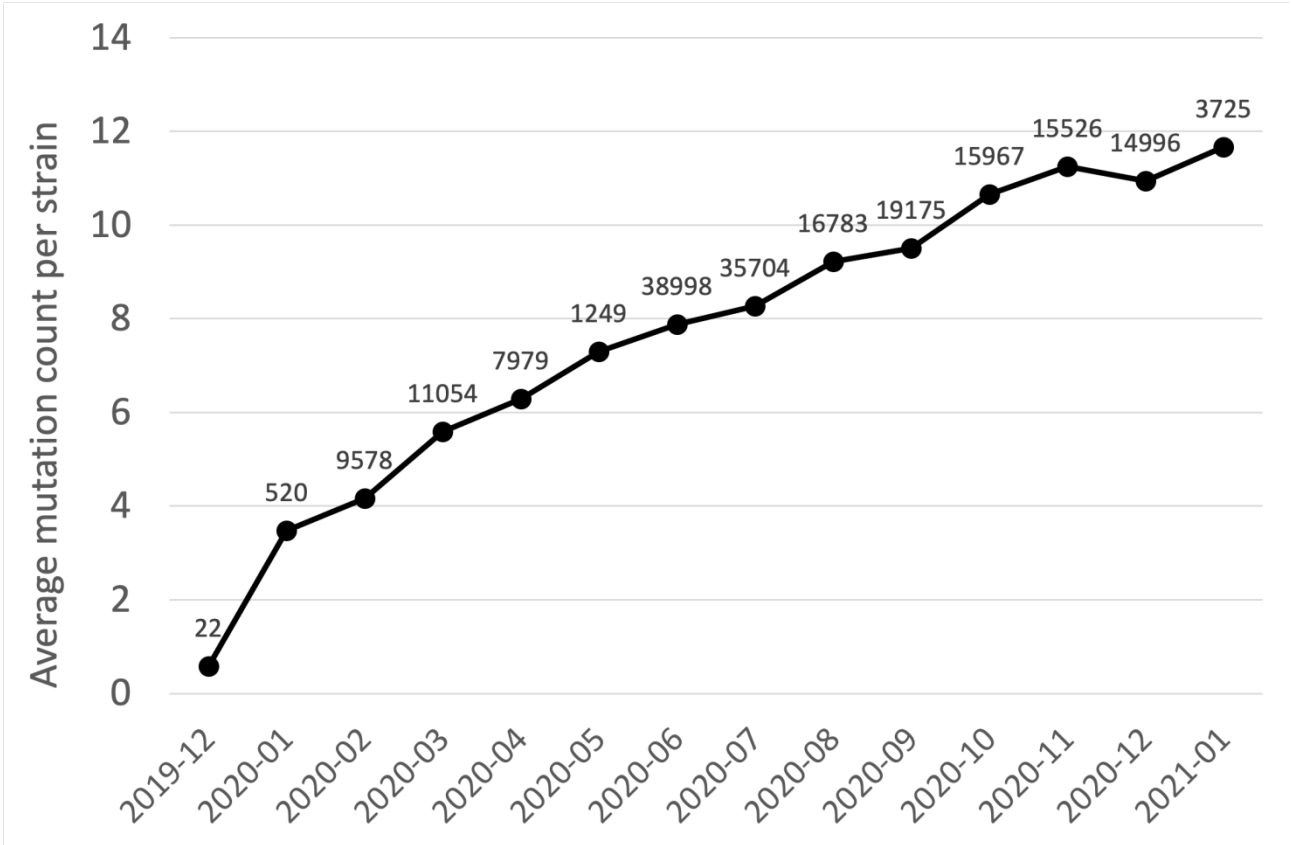


图 3. (a) 从 2019 年 12 月到 2021 年 1 月每个毒株中的平均突变计数。(b) 从 2019 年 12 月到 2021 年 1 月每个毒株的突变计数分布。a 中的数据标签显示了各个月用于分析的毒株数量。

基因组序列上各个位置自然发生的突变率不可能有很大差异，但是某些基因座进化迅速，而其他区域似乎受到极大限制。在迅速进化的区域，选择压力相对于病毒的整体增殖是宽松的，因此它可以迅速适应宿主的抗病毒反应。该观察结果应促使监测特定功能的进化是否与传播速率更高或更低的环境相关，同时还伴随疾病严重程度的变化。通常与轻度疾病相关的突变影响蛋白 Orf8、Nsp6、Orf3a、Nsp4 和核衣壳磷蛋白 N。相反，特定突变位于突刺糖蛋白、取决于 RNA 种类的 RNA 聚合酶中，有时还位于 Orf3a、Nsp3、Orf6 和 N 中，与严重的疾病后果有关。最后，与严重疾病后果相关的突变也位于 Orf3a 和 Nsp7 中。不幸的是，在疾病传播的背景下很少讨论这些蛋白质的作用。此外，除了蛋白 S 在触发免疫反应中的普遍作用外，在 22 个与疾病的临床结果发生重大变化（减轻或加重）相关的突变中，还有 4 个（3 个与严重疾病相关，1 个与轻度疾病相关）被定位到一个十个氨基酸长的磷酸化核衣壳 N 片段上。这指向病毒基因组中一个高度相关的位点 (Nagy et al., 2021)。

尽管一些计算机仿真和体外实验的分析试图研究突变会否对该病毒造成伤害，但这些分析往往只是结构导向而不是功能导向的，这样会造成偏颇的结论。仅在持续时间短的谱系中才能了解突变是否有害。前期研究尚在不清楚某些快速传播的变异毒株是否因建立者效应而造成的超级传播事件，过了一段时间后，现在似乎已经确定某些突变确实在影响该疾病的传播 (Borges *et al.*, 2020)。实际上，这种观点只是事后的描述。而且很少会考虑到病毒是具有高度整合功能的。为了预测其进化方向，我们必须评估各种形成病毒的因素的功能性上位效应，即了解不同突变之间的联系如何促进病毒的适应性。在当前情况下，以同类型的分析确定了几个 SARS-CoV-2 基因座间的相互作用，包括 Orf3a 和 Nsp2、Nsp12、Nsp6 之间、Orf8 和 Nsp4 之间以及 Nsp2、Nsp13 和 Nsp14 之间 (Zeng *et al.*, 2020)。这些有待深入研究，因为部份相互作用尚未与疾病的严重程度关联（见上文）。因此，对那些尚未与疾病严重性相关连的蛋白序列的进化作进一步的监测有着极大的意

义。找出与轻度或重度疾病病例相关的病毒谱系是这类调查的重要结果，这结果能够促使与疾病传播行为相关的政策（通过限制聚集，戴口罩，限制旅游等）更精确地匹配在这过程中发现的关键病毒功能的进化。

当病毒长时间停留在人群中，病毒周期中的各个步骤所受的压力都会发生变化。冠状病毒与人类有着共同演化的过程，有迹象显示东亚人的基因网络可能起到了局部保护作用 (Souilmi *et al.*, 2020)。有趣的是，一些最基本的病毒蛋白并没有发生变化。例如，蛋白酶 Nsp5 的氨基酸序列是高度保守的，这远远超过 SARS-CoV-2 的变异毒株以及所有已知的冠状病毒。这些在不同的病毒家族中高度保守因而没有变化或变化很少的蛋白质是广谱抗病毒药物的潜在靶标。SARS-CoV-2 Nsp5 与 SARS-CoV-1 的氨基酸序列有 95% 相同。其三维结构可用于设计抑制剂，这种方法套用在控制 HIV 方面很成功 (Lubin *et al.*, 2020)。另一个关键特征是病毒的杀伤程度与特定突变之间的关系，这个特征必须与相关的元数据关联起来。在 692 个 SARS-CoV-2 基因组序列中，地理起源和 COVID-19 病例严重程度有着统计学上显著的相关性。研究发现地域差异与病例的严重性和等位基因变异有关，这在印度发现的病毒谱系中特别明显 (Goyal *et al.*, 2021)。而除了这些基本的观察外，也需要提倡对 SARS-CoV-2 全基因组序列进行系统地测序，同时追踪谱系在不同地区中的演变。应该要仔细检测病毒毒力的发展趋势，而地方性的变化则应立即触发相对应的遏制政策以控制病毒扩散。

在我们预测未来走向时收到的关于快速传播变异毒株的警告

正如前文描述，流行病最可能的早期发展是该病毒开始提高其传播速度。这很可能是由两个常见的功能引起的：病毒复制成功率以及传播的增加（此处涉及许多影响因素，包括拥挤的环境、或从受感染者所产生稳定而长期的感染源等）。这些在几个迅速传播并取代了已有病毒株的变异毒株中都可以看到 (<https://nextstrain.org/ncov/global>)。第一个记录下来的符合这类发展的例子是突刺蛋白中的 D614G 突变，该突变增强了 S1/S2 连接处的切割 (Gobeil *et al.*, 2021)。在那之后直至撰写本文时，全球至少发现了根据 Global Initiative on Sharing All Influenza Data (GISAID) 定义的八个主要进化枝：S、O、L、V、

G、GH、GR 和 GV (<https://www.gisaid.org/phylogenetics/global/nextstrain/>)。在亚洲有着一种特有的模式，包括变种 G、GH、GR、L、S、O (Sengupta *et al.*, 2021)，而在欧洲这病毒正在迅速演变为多个新谱系 (Hodcroft, Domman *et al.*, 2021)。

不幸的是变种命名法有些不稳定而且令人困惑：一个利用病原体基因组数据的科学和公共卫生的开源项目 Nexstrain (<https://nextstrain.org>)，提出了五个不同的进化枝：19A、19B、20A、20B 和 20C。Qingtian Guan 及其同事也提议将谱系分为五个进化枝，但名称不同：G614、S84、V251、I378 和 D392 (Guan *et al.*, 2020)，这与 Andrew Rambaut 和同事根据英国的疫情提出的 A、B、B.1、B.1.1 和 B.1.177 进化枝相关 (<https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563>)。

与上述这些命名法类似的是有着六个进化枝及作为其基础的标志性 SNP 的命名法，这些标志性 SNPs 已通过早期获得的变异序列进行了验证。例如，VI 型进化枝的特征是四个标志性的 SNP C241U (5'UTR)、C3037U (Nsp3 F924F)、C14408U (Nsp12 P4715L) 和 A23403G (Spike D614G)，它们有着强烈的等位基因关联。该变种由这些 C 到 U 的突变主导着，在很早的时候就已经成为优势毒株 (H.-C. Yang *et al.*, 2020)。为了说明主要进化枝的变化，在印度特兰甘纳邦最初的（武汉）进化枝 19A 被进化枝 20A 迅速取代，进化枝 20A 有着 5'UTR 的 C241U 突变和突刺蛋白 D614G 突变。紧接着的是已成为主流的进化枝 20B (Gupta *et al.*, 2021)。随后许多突刺蛋白中的其他变化被发现在迅速传播着 (Vilar and Isom, 2021)，表明该蛋白的大部分选择压力来自对宿主的适应。因此，我们可以预期，在疫苗的选择压力下，这种蛋白质以及较小程度的核衣壳蛋白将进化得最快。这反映在图 4 和补充图 1 中，它们展示了突变如何在世界范围内传播。但是由于数据收集极不均衡，因此这只是对目前状况的总体概述，强调了随着病毒在不同国家中传播应进行全面的病毒基因组测序以了解突变密度的需求。虽然突刺蛋白的全球突变密度在逐渐增加，但自 2020 年 9 月以来，在墨西哥和巴西等一些国家增加得非常快。

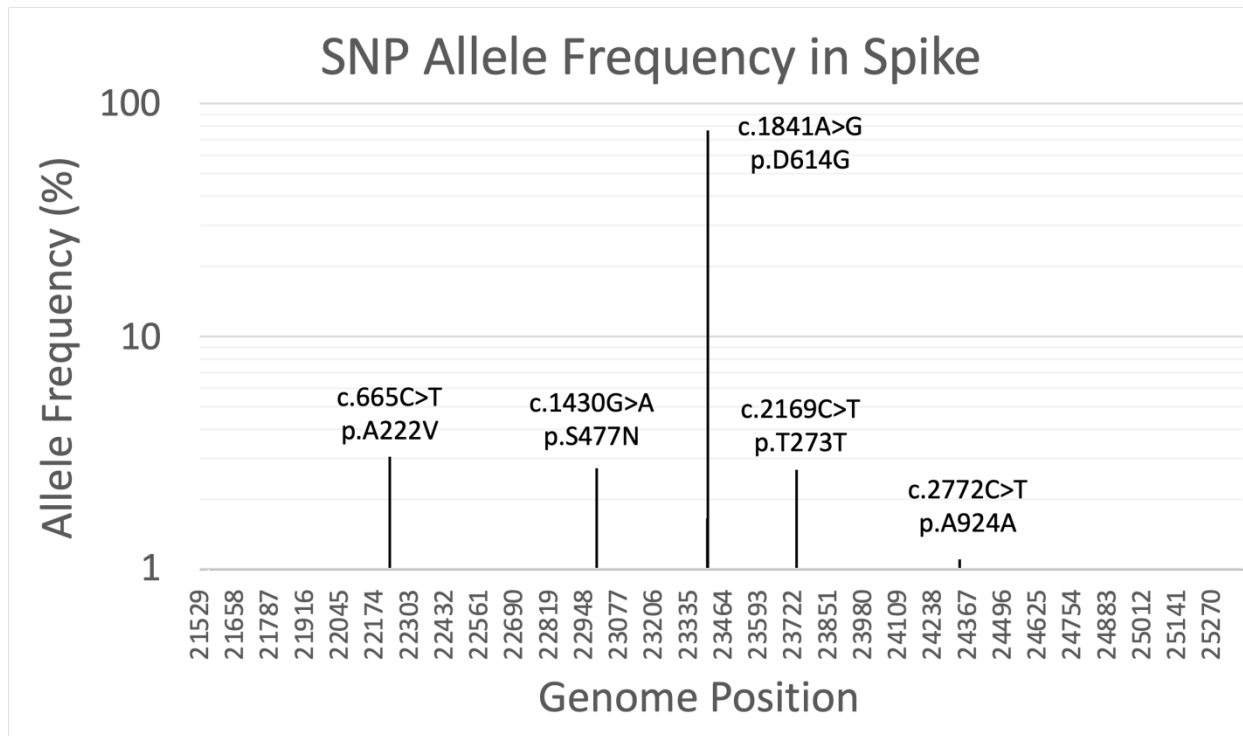


图 4. 193,687 个 SARS-CoV-2 全基因组中突刺蛋白突变的 SNP 等位基因频率 (Allele Frequency, AF)。y 轴是以 10 为底的对数刻度。我们仅展示了 AF ≥ 1% 的突变。

在发表这些研究后，其他重要变种的出现也有被报道。在撰写本文时，具有许多突变的几种变异毒株正在广泛传播 (<https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/scientific-brief-emerging-variants.html>)，英国变种 B.1.1.7、南非变种 B.1.351 和巴西变种 B.1.248 (Firestone *et al.*, 2021) 正在世界范围内推广。许多突变会同时发生变化 (<https://covariants.org>)。这可能是由于趋同进化导致的，这暗示着某些突变之间的上位相互作用。此类突变很可能会帮助指出病毒的功能特征，以预测其进一步的进化，以及可能的毒力减弱。这些突变的特征除了是一连串的突变之外，英国变种中还存在一个小缺失突变。通常来说在进化过程中插入/缺失突变是非常常见的 (请参见图 2)。但值得注意的是蝙蝠病毒基因组中的插入突变实际上是 SARS-CoV-2 感染人类的起源。谱系 B.1.525 展示了共同突变，谱系 B.1.525 是在 13 个国家/地区中发现的一个有趣新变异毒株，其突刺蛋白突变为 E484K、Q677H、F888L，并且有一系列与 B.1.1.7 相同的缺失突变 (https://cov-lineages.org/global_report_B.1.525.html)。

如果我们要预测疫情的未来，我们必须一直铭记突变的潜在重要性，而不仅是那些影响突刺蛋白的突变。

现在尚不清楚 D614G 突变是否会影响疾病的严重程度。但是在最近发现的三个主要变种中 (Zhou *et al.*, 2021)，已通过 B.1.1.7 变异毒株明确地证明了严重程度的提高 (Davies *et al.*, 2021)。许多变异毒株正在逐步替代早期病毒毒株，其中 B.1.525（丹麦、英国、尼日利亚）的严重程度未知，但它的传播潜力却不断在增加。

突刺蛋白以外一些重要变异的例子

大多数观察仅仅是描述性的，但是对于那些影响着进化的病毒功能改变的发现来说至关重要，特别是那些在可以预期的突刺蛋白变异之外的改变。譬如我们预测复制过程的改变将会产生两个截然不同的后果。一方面，长远地它将损害变异毒株的传播而导致病毒的灭绝，但另一方面，它又可能会在短期内拓展其进化空间。在这种情况下，应该把观察到的病毒产生进化“绽放”的趋势与急速扩张的谱系联系起来，例如对于病毒脱落胞嘧啶的自然趋势的可能的逆转 (Cluzel *et al.*, 2020)。这一过程的确对于预测在疫情前六个月内病毒进化的后果（产生了多个进化枝）非常有价值 (Koyama *et al.*, 2020)。

在重要的变异中，病毒基因组 5'UTR 中普遍存在的 C241U 突变尚未被发现与任何特定表型相关。它通常被认为是中性的突变，它使病毒对宿主的新陈代谢获得了小程度的适应性，并提高了对 viperin 的抵抗力。此外，它位于病毒基因组的 5'UTR 前导序列，这对于它控制翻译和宿主特异性而言特别重要 (Tidu *et al.*, 2020)。顺着这思路，有许多病毒进化“绽放”的例子。例如，从一个早期突变开始的一连串有趣的突变演替，G11083U（蛋白质 Nsp6，L37F）现在已广泛分布于世界各地，并且与印度的进化枝有关 (Banerjee *et al.*, 2020)。在变异 G2891A（A876T，蛋白 Nsp3 的泛素样结构域）之前，在多个国家发现了另一个突变 G1440A（G392D，蛋白质 Nsp2）(Liu *et al.*, 2020)，这随后引发了翻译 Orf7a 和 Orf7b 之间的冲突。这表明对于这些蛋白质来说任一表达都存在成本或收益难题。与这位点有关的病毒谱系值得追踪，因为它可能导致有趣的减毒形式 (Cluzel *et al.*, 2020)。同样，SARS 相关冠状病毒的 Orf8 区域也是高度变化的。它在疫情中不断变化，说明受到持

续的选择压力，有时会产生两种肽 Orf8a 和 Orf8b (S. Chen *et al.*, 2020)。在疫情的第一阶段来自四个不同国家的七个样本中，Orf8 区域有着一个开枝散叶般的突变产生过程，从第一次突变到最后一个突变跨越了六星期 (Cluzel *et al.*, 2020)。Orf8 蛋白在感染周期的末尾表达。监测它们促使病毒毒力进化的方式非常重要 (Neches *et al.*, 2021)。

最后，人们对监测病毒复制酶 Nsp12 的进化有着明显的兴趣。在疫情早期，从广泛的 5'端 C241U 突变开始的一连串突变，接着是 Nsp12 中锌指末端的突变 C14408U (P314L)。这种突变出现在病毒进化树的许多进化枝中。它以值得注意的方式改变了复制酶的活性，因为这种突变之后是病毒新谱系的“绽放”，这表明改变的复制过程是诱变的 (Cluzel *et al.*, 2020)。一个值得认真监测的例子是非常有趣的一连串演替：突刺蛋白中的 A23403G (D614G) 和 C3037U (同义) 突变、与 Orf3a 形成钾离子通道有关并可能对该蛋白的功能产生负面影响的突变 G25563U (Q57H) (Issa *et al.*, 2020)、蛋白质 Nsp2 中的 C1059U (T265I)、蛋白酶 Nsp3 的 SUD-N 区域中的三联体 G4181A (A1305T)，紧接着的是突变 G4285U (E1340D) 和 G28209U 导致翻译终止于 Orf8 蛋白质的 E106，它们的再次出现成为了病毒进化的重要标志 (Neches *et al.*, 2021)。自大流行开始已经过去了一年，应该使用新的解释来探究仍在不断出现的突变，特别是基于当前疫苗接种正在迅速进行的情况下。

疫苗接种及其预期后果

接种疫苗已成为控制甚至根除传染病的首选方法。这在多数情况下是正确的，但有些病毒却无法被研发出有效的疫苗，例如 HIV (Oyston and Robinson, 2012)。在减毒疫苗研发并使用成功之后 (Theiler and Smith, 1937)，许多新的疫苗种类在这之后被陆续开发 (Smith, 2012)。就冠状病毒而言，疫苗已成功应用于预防动物疾病 (Cruz *et al.*, 2010; Singh *et al.*, 2019)。因此，我们预期至少在这场大流行的早期，疫苗接种将有效控制疫情。但是，正如我们所看到的，该病毒的进化非常迅速，因此病毒的变异具有导致现有的 COVID-19 疫苗效力降低的持续风险。发展出多个毒株是常见的进化特征，并可能导致疫苗的减效或无效 (Zeng *et al.*, 2017)。在南非首次发现的 B.1.351 变异毒株似乎就是这种情况 (Diamond *et*

al., 2021)。此外，如果疫苗接种行动太慢，将为病毒提供充足的时间来进化出可以逃避疫苗免疫的变异毒株。还有，只基于病毒中单一蛋白的疫苗，或更糟糕的只基于某病毒蛋白结构域的疫苗，将会选择性地驱使突变在这些蛋白质或蛋白质结构域中快速富集 (*J. Yang et al.*, 2020)。

其他原因所导致的感染也可能影响病毒的进化。对于 SARS-CoV-1，某些人群似乎不受感染的这一事实表明，先前的某些感染可能已经诱导了交叉保护作用 (*Ng et al.*, 2003)。上面我们已经注意到了吸烟习惯和轻度感染之间出乎意料的关系。这意味着流感、流感疫苗接种与 COVID-19 疫情发展之间可能在未来会呈现关联。感染过流感病毒的人似乎更容易感染 SARS-CoV-2 (*Bai et al.*, 2021)。但是，回顾性研究并未发现接种了流感疫苗的人更容易得 COVID-19，且恰恰相反 (*Green et al.*, 2020)。总之，所有相关的观察结果都呼吁实施快速的疫苗接种计划，同时监视尚未达到有效接种比例的人群中的病毒基因组的进化相关变化。结合上述观点，我们认为，对这场大流行发展的预测，必须分别考虑已接种的人群和未完成接种的人群。

长期进化

1918-1919 年的流感疫情可能有助于我们预测 COVID-19 流行病可能会发生的情况。虽然无法完整重现当时发生的所有情景，至少我们知道这场疫情分三个发展阶段。第一阶段看上去像严重的流行性感冒。随后，它发展成一种非常严重的疾病。最终，它适应了人类宿主，并最终发展成目前的样子。在该疫情彻底消失之前，它还从人传到了猪 (*Shope*, 1936)。是基因重排 (reassortment) 事件导致了这种更强的流感病毒。流感病毒由独立的序列片段组成，当与相似的病毒共同感染时可能会发生基因重排，从而引入新蛋白质变体，帮助病毒繁殖。有趣的是，这些高毒性变体不涉及病毒感染并从宿主细胞中释放所需的血凝素或神经氨酸酶蛋白 (*Reid et al.*, 2004)。这再次强调，除了那些已经可被免疫系统快速识别的蛋白质，我们更要分析和考虑那些仍然被忽视的病毒特征。

SARS-CoV-2 不会发生基因重排，因为它是由单个 RNA 组件组成的，但它易于与其他 RNA 重组 (recombination)，所以 SARS-CoV-2 在短期内最令人担忧的进化特征是进行大规模的重组，而且其遵循重组率与基因组长度的负相关关系 (*Zhang et al.*, 2005; *Goldstein et*

al., 2021)。这个过程似乎取决于 Nsp14 蛋白的活性 (Gribble *et al.*, 2021)。这个过程可能源于同时感染了不同的 SARS-CoV-2 毒株 – 在这方面这将与有性繁殖和基因水平转移相似, 导致穆勒棘轮 (Muller's ratchet) 的重启。这过程可能会随着旅游限制的放宽而增强 – 但也可能源于同时感染了不同的病毒, 甚至是人工构建的序列都可以被重组进来。所以说, 在最近研发的 mRNA 类疫苗中用伪尿苷替代尿苷可以说是一项积极的创新以避免病毒把疫苗序列给重组进来, 从而扩大病毒的进化潜力。我们还要关注 Nsp14 校对核酸外切酶所可能进化出的变体。我们应仔细监测在 Nsp14 中发生的突变, 因为它们可能会扩大病毒的进化资源, 在其突刺蛋白、核蛋白壳和 Orf8 蛋白中产生重组热点。冠状病毒普遍有基因重组的特征, 这在正链单链 RNA 病毒中普遍存在 (Zhang *et al.*, 2005; Patiño-Galindo *et al.*, 2021)。除了进化逆转之外, 重组还可能将不同环境下各自所产生的突变整合到一起。由于病毒重组发生在同时感染不同毒株的过程中, 因此要特别关注人群密集的环境, 尤其是封闭 (室内) 的环境, 安装有空气内循环的环境, 以及需要进行声乐活动的环境。为了抗疫, 在完成疫苗大范围覆盖或在有效的抗病毒药物可用之前, 我们必须尽量减少此类环境下的活动。

为了预测该场大流行的未来, 我们必须仔细分析病毒的趋向性 (tropism, 嗜性) 将如何演化。目前, 呼吸系统感染 (具有呼吸系统趋向性) 是主要途径, 但我们知道在其他冠状病毒中, 这种趋向可以非常迅速地改变。例如, 禽冠状病毒突刺蛋白中的三个氨基酸变化使该病毒能与肾细胞结合 (Bouwman, Parsons, *et al.*, 2020)。在小鼠中, 冠状病毒可能表现出神经趋向性 (Pasick *et al.*, 1994)。细胞进入和融合的自然选择过程与突刺蛋白的动态结构密切相关。我们已经看到, 在蝙蝠冠状病毒中插入的 furin 蛋白酶切割位点使其能够改变其宿主并适应人类的 ACE2 受体 (Coutard *et al.*, 2020)。现在, 该病毒已经找到了另一种进入受体, 即通过突刺蛋白与 CD147 蛋白的相互作用。至少在体外实验模型中 CD147 在多种细胞中均有表达, 包括上皮细胞和神经元细胞 (K. Wang *et al.*, 2020)。这点很重要, 因为利用多种受体可以使病毒通过不同的细胞进入人体, 也就是可以从不同的器官进入人体。另一方面, MERS-CoV 的受体, 二肽基肽酶 DPP4 的多态性会对病毒的进入产生负面影响 (Kleine-Weber *et al.*, 2020)。其他冠状病毒还会利用到其他一些受体: 猪 Δ 冠状病毒 (porcine delta coronavirus) 利用氨基肽酶 N (APN) 作为受体, 并通过突刺蛋白的结构域 S2 与 APN 相互作用 (Li *et al.*, 2018)。小鼠肝炎冠状病毒 (MHV) 是唯一已知的

利用其突刺蛋白的 N 末端结构域 (NTD, N-terminal domain) 识别另一种受体 CEACAM1a 的冠状病毒 (Shang *et al.*, 2020)。

另一个值得留意的病毒特征是保护病毒不被降解的蛋白质糖基化，它有可能被病毒用作进入靶细胞的一种方式。举个例子，鸡冠状病毒感染性支气管炎病毒 (IBV, chicken coronavirus infectious bronchitis virus) 通过病毒的“重度 N-糖基化的附着蛋白突刺”与宿主的“ α -2,3-连接唾液酸受体 Neu5Ac”结合而进入宿主细胞 (Bouwman, Habraeken, *et al.*, 2020)。人冠状病毒 OC43 显然是从牛冠状病毒 (BCoV) 而来。它通过其突刺蛋白与基于 9-O-乙酰化唾液酸聚糖 (sialoglycan-based receptors with 9-O-acetylated sialic acid) 的受体结合并以血凝素酯酶作为破坏受体的酶 (Lang *et al.*, 2020)。9-O-乙酰化唾液酸聚糖的受体结合位点在所有冠状病毒的突刺糖蛋白中都是保守的，其结构类似于冠状病毒血凝素酯酶和流感病毒 C/D 血凝素酯酶的融合糖蛋白的配体结合袋结构域 (Tortorici and Veessler, 2019)。因此，我们应该开发方法以监测针对病毒糖蛋白的抗体的变化，以及基因组突变导致的病毒蛋白糖基化改变。

展望

COVID-19 大流行开始了一年有多，我们正处于一个过渡时期，病毒正遭受着各种能将其推向长期存活的选择压力的考验。现在已经有三个主要因素促使病毒向可能的危险结果方向发展：毒株之间的重组、非常缓慢的疫苗接种运动以及抗病毒药物研究的不足。同时，受感染的人数非常多，因此在拥挤的环境中发生不同病毒毒株的共同感染已经不再稀奇。能够实时跟踪完整基因组序列的演变变得至关重要，这样我们才能查明病毒蛋白质中可能导致其减毒或相反导致更严重后果的目标位点。

我们提倡广泛收集病毒的完整基因组序列。而且，这些完整的序列只有在与相关的元数据相关联时才有意义。另外，病人的患病史是重要的元数据。元数据越多越好，而元数据的收集必须要合理地标准化。

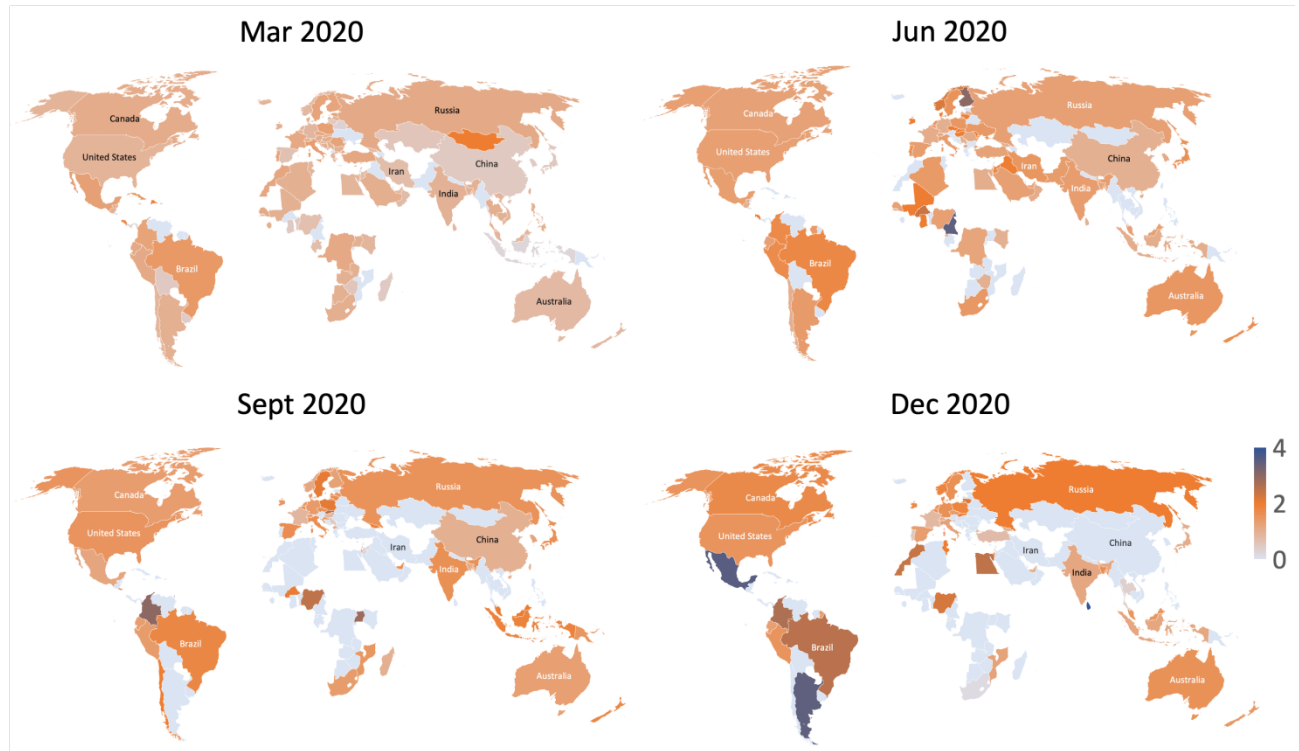
以下是迫切需要实施的措施，它们符合“认识你的敌人”的关键原则。

- 在任何地方都尽可能地对病毒的整个基因组（而不仅仅是突刺蛋白）进行测序。当测序技术不断改进时，这应该是切实可行的。例如，对 752 个临床样品进行纳米孔测序很快且很轻易地确定了该病毒的 3 个进化枝 (Bhoyar *et al.*, 2021)。

- 将重要的元数据与这些序列相关联（包括所有我们可以使用的关于感染者的信息以及临床数据），并将元数据与特定突变相关联，且不局限于只是突刺蛋白。
- 建立毒株谱系及其传播情况，并关联到各种参数，包括年龄、性别、种族（如果情况允许的话）、体重、人类遗传多态性特征（HLA）、Lewis 血型、营养习惯和平常行为（例如吸烟）等标准数据。另外，由于传播需要人与人之间的接触，我们应当确定感染的确切地点（国家、城市、建筑物），并将其与气象参数相关联。
- 专注于进化模式的变化，如毒株分支的爆发、核苷酸进化模式的改变（C 或 G 丢失趋势的逆转、颠换等情况）。并尝试将这些进化模式与特定蛋白中的突变联系起来。
- 识别正在衰减的毒株品系，并在受控下允许其传播。监控病毒从肺到肠的趋向性变化，反之亦然。
- 找出情况严重的毒株，并施加严格的局部防范措施。
- 追溯传播的上游，而不是帮助不大的下游。对感染者及其接触者的活动进行严格控制。
- 尽可能加快疫苗接种的速度，尤其是在高发病率和已存在多种变异毒株的人群中。紧急开发涵盖变异毒株第二代和第三代疫苗，以弥补第一代的不足。
- 基于较稳定的病毒蛋白，大量投资开发新的抗病毒剂，并进行广泛的筛查。

最后，似乎我们将不得不与 SARS-CoV-2 共存。这意味着，为了控制其负面后果，我们必须仔细跟踪其抗原决定簇的演变。COVID-19 可能会演变成类似季节性流感的情况，并且每年都需要使用不同的疫苗。其与其他呼吸系统疾病（尤其是流感）的共同进化必须要予以高度重视，因为忽略对后者保持稳定的群体免疫力可能会导致糟糕的后果。

补充图



补充图 1. 突刺蛋白在 2020 年四个不同月份的时间-地理突变密度。 当月可用基因组少于五个的国家没有被画出（灰色）。分析依赖于不同国家研究人员提交给 GISAID 的毒株，因为国家间基因组序列采样极为不均匀 (Hodcroft, De Maio, *et al.*, 2021)，我们要谨慎从此图中得出结论，该图仅是指示性的。例如，迄今为止，英国已对每 100,000 个案例中的近 5,000 个样本进行了排序，而美国为每 100,000 个案例中有 320 个样本，巴西为每 100,000 个案例中有 30 个样本（请参阅门户网站 <http://go.nature.com/3qutqwj>）。即使在英国，某些区域在测序数据方面的表现也要好于其他区域。

补充注解

我们筛选并从 GISAID 下载了从 2019 年 12 月 26 日算起至 2021 年 1 月 25 日采样的 193,687 株 SARS-CoV-2 的基因组。在我们的筛选中，我们要求毒株具有 1) 完整的基因组，2) 足够高的测序深度，和 3) 具体到月的样品收集日期。我们衷心感谢这 193,687 个基因组的，负责原始样品收集实验室，和负责进行测序并把序列提交给 GISAID 倡议进行共享的实验室。全部这些基因组的统一编号收录在这个 VCF 文件中，网址为 <http://www.bio8.cs.hku.hk/adanchin/variantsOf193687nCovAgainstWIV04.vcf.gz>，这其中已

包含了我们在这些毒株中所找到的突变。根据这些统一编号，我们可以直接通过 www.gisaid.org 与所有数据提交者取得联系。

我们使用 WIV04 (GISAID: EPI_ISL_402124, NCBI: MN996528) (Zhou *et al.*, 2020) 作为参考基因组，并将所有其他基因组与 WIV04 进行比对。WIV04 被用于开发国药集团的灭活 SARS-CoV-2 疫苗 (Xia *et al.*, 2020)。在最早发布的基因组中，另一个通常用作 SARS-CoV-2 参考的基因组是 Wuhan-Hu-1 (GISAID: EPI_ISL_402125, NCBI: NC_045512)。与 WIV04 相比，它在基因命名方面有细微差别（例如，Wuhan-Hu-1 中的 Orf8 与 WIV04 中的 NS8），但所有基因的起始和终止位置相同。因此，WIV04 和 Wuhan-Hu-1 之间的基因名称是可互换的。在这项研究中，我们将使用 Wuhan-Hu-1 中的基因命名（即以 'ORF' 代替 'NS' 开头）。

我们使用 MAFFT (Kato *et al.*, 2002) 进行多序列比对。多序列比对在以大量序列作为输入时非常的很慢。通过在 MAFFT 中设置使用参考基因组，我们得以在一个 24 核计算服务器上用了不到一个月的时间完成比对，但只能获得 SNP 和缺失突变 (deletion)。我们使用了 jvarkit-msa2vcf (<https://github.com/lindenb/jvarkit>) 将 MAFFT 输出中的突变提取到 VCF 文件中。我们使用 BCFtools (Li, 2011) 归一化所有样本中的突变表示方法（将不同样本中使用不同表示方法的相同突变合并）。我们删除了除了 'A', 'C', 'G', or 'T' 外，包含的未知碱基 'N' 和 IUPAC 核苷酸代码的突变。我们删除了长度超过 50bp 的缺失突变，因为它们更可能由不完整的基因组组装得来，而非一个真正的突变。第一段落中给出的 VCF 文件链接提供了所有 193,687 个基因组的突变。所有后续分析均基于此 VCF 文件。

补充表

补充表 1. 总结自 193,687 个 SARS-CoV-2 毒株的基因突变计数和发生率。 基于每个基因的长度，我们将突变发生率标准化为每 1kbp 的突变数。对于每个基因，我们展示了在 ≥ 1 个毒株， ≥ 5 个毒株， ≥ 10 个毒株， ≥ 100 个毒株和 ≥ 1000 个毒株中突变发生率。

引用文献

- Ahamad, S., Gupta, D., and Kumar, V. (2020) Targeting SARS-CoV-2 nucleocapsid oligomerization: Insights from molecular docking and molecular dynamics simulations. *J Biomol Struct Dyn* 1–14.
- Allen, C., Bekoff, M., and Lauder, G. (1998) *Nature's Purposes*, Cambridge, Massachusetts: MIT Press.
- Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R.F. (2020) The proximal origin of SARS-CoV-2. *Nat Med* 26: 450–452.

Arya, R., Kumari, S., Pandey, B., Mistry, H., Bihani, S.C., Das, A., *et al.* (2021) Structural insights into SARS-CoV-2 proteins. *J Mol Biol* **433**: 166725.

Audi, A., Allbrahim, M., Kaddoura, M., Hijazi, G., Yassine, H.M., and Zaraket, H. (2020) Seasonality of respiratory viral infections: will COVID-19 follow suit? *Front Public Health* **8**: 567184.

Augustin, M.A., Reichert, A.S., Betat, H., Huber, R., Mörl, M., and Steegborn, C. (2003) Crystal structure of the human CCA-adding enzyme: insights into template-independent polymerization. *J Mol Biol* **328**: 985–994.

Badawi, S. and Ali, B.R. (2021) ACE2 Nascence, trafficking, and SARS-CoV-2 pathogenesis: the saga continues. *Hum Genomics* **15**: 8.

Bai, L., Zhao, Y., Dong, J., Liang, S., Guo, M., Liu, X., *et al.* (2021) Coinfection with influenza A virus enhances SARS-CoV-2 infectivity. *Cell Res*.

Ballesteros, M.L., Sánchez, C.M., and Enjuanes, L. (1997) Two amino acid changes at the N-terminus of transmissible gastroenteritis coronavirus spike protein result in the loss of enteric tropism. *Virology* **227**: 378–388.

Ballout, R.A., Sviridov, D., Bukrinsky, M.I., and Remaley, A.T. (2020) The lysosome: A potential juncture between SARS - CoV - 2 infectivity and Niemann - Pick disease type C, with therapeutic implications. *FASEB j* **34**: 7253–7264.

Banerjee, A., Sarkar, R., Mitra, S., Lo, M., Dutta, S., and Chawla-Sarkar, M. (2020) The novel coronavirus enigma: phylogeny and analyses of coevolving mutations among the SARS-CoV-2 viruses circulating in India. *JMIR Bioinform Biotech* **1**: e20735.

Baranovich, T., Wong, S.-S., Armstrong, J., Marjuki, H., Webby, R.J., Webster, R.G., and Govorkova, E.A. (2013) T-705 (favipiravir) induces lethal mutagenesis in influenza A H1N1 viruses in vitro. *J Virol* **87**: 3741–3751.

Baranowski, E., Ruiz-Jarabo, C.M., and Domingo, E. (2001) Evolution of cell recognition by viruses. *Science* **292**: 1102–1105.

Barretto, N., Jukneliene, D., Ratia, K., Chen, Z., Mesecar, A.D., and Baker, S.C. (2005) The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. *J Virol* **79**: 15189–15198.

Belvin, M.P. and Anderson, K.V. (1996) A conserved signaling pathway: the Drosophila toll-dorsal pathway. *Annu Rev Cell Dev Biol* **12**: 393–416.

Bernard, S., Bottreau, E., Aynaud, J.M., Have, P., and Szymansky, J. (1989) Natural infection with the porcine respiratory coronavirus induces protective lactogenic immunity against transmissible gastroenteritis. *Vet Microbiol* **21**: 1–8.

Bernheim, A., Millman, A., Ofir, G., Meitav, G., Avraham, C., Shomar, H., *et al.* (2020) Prokaryotic viperins produce diverse antiviral molecules. *Nature*.

Bertram, S., Glowacka, I., Müller, M.A., Lavender, H., Gnirss, K., Nehlmeier, I., *et al.* (2011) Cleavage and activation of the severe acute respiratory syndrome coronavirus spike protein by human airway trypsin-like protease. *J Virol* **85**: 13363–13372.

Bhojar, R.C., Jain, A., Sehgal, P., Divakar, M.K., Sharma, D., Imran, M., *et al.* (2021) High throughput detection and genetic epidemiology of SARS-CoV-2 using COVIDSeq next-generation sequencing. *PLoS ONE* **16**: e0247115.

Bloch, E.M., Patel, E.U., Marshall, C., Littlefield, K., Goel, R., Grossman, B.J., *et al.* (2021) ABO blood group and SARS-CoV-2 antibody response in a convalescent donor population. *Vox Sang*.

Borges, V., Isidro, J., Cortes-Martins, H., Duarte, S., Vieira, L., Leite, R., *et al.* (2020) Massive dissemination of a SARS-CoV-2 Spike Y839 variant in Portugal. *Emerg Microbes Infect* **9**: 2488–2496.

Bosch, B.J., van der Zee, R., de Haan, C.A.M., and Rottier, P.J.M. (2003) The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol* **77**: 8801–8811.

Boson, B., Legros, V., Zhou, B., Siret, E., Mathieu, C., Cosset, F.-L., *et al.* (2020) The SARS-CoV-2 envelope and membrane proteins modulate maturation and retention of the spike protein, allowing assembly of virus-like particles. *J Biol Chem* 100111.

Bouwman, K.M., Habraeken, N., Laconi, A., Berends, A.J., Groenewoud, L., Alders, M., *et al.* (2020) N-glycosylation of infectious bronchitis virus M41 spike determines receptor specificity. *J Gen Virol* **101**: 599–608.

Bouwman, K.M., Parsons, L.M., Berends, A.J., de Vries, R.P., Cipollo, J.F., and Verheije, M.H. (2020) Three amino acid changes in avian coronavirus spike protein allow binding to Kidney tissue. *J Virol* **94**.

Bradwell, K., Combe, M., Domingo-Calap, P., and Sanjuán, R. (2013) Correlation between mutation rate and genome size in riboviruses: mutation rate of bacteriophage Q β . *Genetics* **195**: 243–251.

Brauer, M., Casadei, B., Harrington, R.A., Kovacs, R., Sliwa, K., and WHF Air Pollution Expert Group (2021) Taking a stand against air pollution-The impact on cardiovascular disease: A joint opinion from the World Heart Federation, American College of Cardiology, American Heart Association, and the European Society of Cardiology. *Circulation* CIRCULATIONAHA120052666.

Buchrieser, J., Dufloo, J., Hubert, M., Monel, B., Planas, D., Rajah, M.M., *et al.* (2020) Syncytia formation by SARS-CoV-2-infected cells. *EMBO J* **39**: e106267.

Bulfone, T.C., Malekinejad, M., Rutherford, G.W., and Razani, N. (2020) Outdoor transmission of SARS-CoV-2 and other respiratory viruses: A systematic review. *The Journal of Infectious Diseases* jiaa742.

Chambers, J.P., Yu, J., Valdes, J.J., and Arulanandam, B.P. (2020) SARS-CoV-2, early entry events. *J Pathog* **2020**: 9238696.

Chen, B., Liang, H., Yuan, X., Hu, Y., Xu, M., Zhao, Y., *et al.* (2020) Predicting the local COVID-19 outbreak around the world with meteorological conditions: a model-based qualitative study. *BMJ Open* **10**: e041397.

Chen, F., Knutson, T.P., Rossow, S., Saif, L.J., and Marthaler, D.G. (2019) Decline of transmissible gastroenteritis virus and its complex evolutionary relationship with porcine respiratory coronavirus in the United States. *Sci Rep* **9**: 3953.

Chen, H., Ning, X., and Jiang, Z. (2017) Caspases control antiviral innate immunity. *Cell Mol Immunol* **14**: 736–747.

Chen, S., Zheng, X., Zhu, J., Ding, R., Jin, Y., Zhang, W., *et al.* (2020) Extended ORF8 Gene Region Is Valuable in the Epidemiological Investigation of Severe Acute Respiratory Syndrome-Similar Coronavirus. *J Infect Dis* **222**: 223–233.

Cherif, A., Barley, K., and Hurtado, M. (2016) Homo-psychologicus: Reactionary behavioural aspects of epidemics. *Epidemics* **14**: 45–53.

Cluzel, N., Lambert, A., Maday, Y., Turinici, G., and Danchin, A. (2020) Biochemical and statistical lessons from the evolution of the SARS-CoV-2 virus: paths for novel antiviral warfare. *C R Biol* **343**: 177–209.

Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N.G., and Decroly, E. (2020) The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Research* **176**: 104742.

Cruz, J.L.G., Zúñiga, S., Bécares, M., Sola, I., Ceriani, J.E., Juanola, S., *et al.* (2010) Vectored vaccines to protect against PRRSV. *Virus Res* **154**: 150–160.

Czypionka, T., Greenhalgh, T., Bassler, D., and Bryant, M.B. (2020) Masks and face coverings for the lay public : A narrative update. *Ann Intern Med*.

Danchin, A. (2003) Infection of society. As diseases have evolved to exploit the holes in our defences, including weaknesses in society, we have to reconsider our way of life, otherwise they will continue to haunt us. *EMBO Rep* **4**: 333–335.

Danchin, A. and Marlière, P. (2020) Cytosine drives evolution of SARS-CoV-2. *Environ Microbiol* **22**: 1977–1985.

Dang, M., Li, Y., and Song, J. (2021) ATP biphasically modulates LLPS of SARS-CoV-2 nucleocapsid protein and specifically binds its RNA-binding domain. *Biochem Biophys Res Commun* **541**: 50–55.

Davies, N.G., Abbott, S., Barnard, R.C., Jarvis, C.I., Kucharski, A.J., Munday, J.D., *et al.* (2021) Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* eabg3055.

DeDiego, M.L., Nieto-Torres, J.L., Jimenez-Guardeño, J.M., Regla-Nava, J.A., Castaño-Rodríguez, C., Fernandez-Delgado, R., *et al.* (2014) Coronavirus virulence genes with main focus on SARS-CoV envelope gene. *Virus Res* **194**: 124–137.

Derjany, P., Namilae, S., Liu, D., and Srinivasan, A. (2020) Multiscale model for the optimal design of pedestrian queues to mitigate infectious disease spread. *PLoS One* **15**: e0235891.

Diamond, M., Chen, R., Xie, X., Case, J., Zhang, X., VanBlargan, L., *et al.* (2021) SARS-CoV-2 variants show resistance to neutralization by many monoclonal and serum-derived polyclonal antibodies. *Res Sq*.

Ditlev, J.A. (2021) Membrane-associated phase separation: organization and function emerge from a two-dimensional milieu. *J Mol Cell Biol*.

Doyle, L.P. and Hutchings, L.M. (1946) A transmissible gastroenteritis in pigs. *J Am Vet Med Assoc* **108**: 257–259.

Dumont-Leblond, N., Veillette, M., Mubareka, S., Yip, L., Longtin, Y., Jouvét, P., *et al.* (2020) Low incidence of airborne SARS-CoV-2 in acute care hospital rooms with optimized ventilation. *Emerg Microbes Infect* **9**: 2597–2605.

Elliott, J., Bodinier, B., Whitaker, M., Delpierre, C., Vermeulen, R., Tzoulaki, I., *et al.* (2021) COVID-19 mortality in the UK Biobank cohort: revisiting and evaluating risk factors. *Eur J Epidemiol*.

Eskier, D., Suner, A., Oktay, Y., and Karakulah, G. (2020) Mutations of SARS-CoV-2 nsp14 exhibit strong association with increased genome-wide mutation load. *PeerJ* **8**: e10181.

Fehr, A.R. and Perlman, S. (2015) Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol* **1282**: 1–23.

Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., *et al.* (2021) The coding capacity of SARS-CoV-2. *Nature* **589**: 125–130.

Firestone, M.J., Lorentz, A.J., Wang, X., Como-Sabetti, K., Vetter, S., Smith, K., *et al.* (2021) First identified cases of SARS-CoV-2 variant B.1.1.7 in Minnesota — December 2020–January 2021. *MMWR Morb Mortal Wkly Rep* **70**.

Gao, X., Qin, B., Chen, P., Zhu, K., Hou, P., Wojdyla, J.A., *et al.* (2021) Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharm Sin B* **11**: 237–245.

Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., *et al.* (2020) Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **368**: 779–782.

Garoff, H., Hewson, R., and Opstelten, D.J. (1998) Virus maturation by budding. *Microbiol Mol Biol Rev* **62**: 1171–1190.

Ghosh, S., Dellibovi-Ragheb, T.A., Kerviel, A., Pak, E., Qiu, Q., Fisher, M., *et al.* (2020) β -Coronaviruses use lysosomes for egress instead of the biosynthetic secretory pathway. *Cell* **183**: 1520–1535.e14.

Gobeil, S.M.-C., Janowska, K., McDowell, S., Mansouri, K., Parks, R., Manne, K., *et al.* (2021) D614G mutation alters SARS-CoV-2 spike conformation and enhances protease cleavage at the S1/S2 junction. *Cell Rep* **34**: 108630.

Goldman, E. (2020) Exaggerated risk of transmission of COVID-19 by fomites. *The Lancet Infectious Diseases* **20**: 892–893.

Goldstein, S.A., Brown, J., Pedersen, B.S., Quinlan, A.R., and Elde, N.C. (2021) Extensive recombination-driven coronavirus diversification expands the pool of potential pandemic pathogens. *Evolutionary Biology*.

Gowthaman, R., Guest, J.D., Yin, R., Adolf-Bryfogle, J., Schief, W.R., and Pierce, B.G. (2021) CoV3D: a database of high resolution coronavirus protein structures. *Nucleic Acids Res* **49**: D282–D287.

Green, I., Ashkenazi, S., Merzon, E., Vinker, S., and Golan-Cohen, A. (2020) The association of previous influenza vaccination and coronavirus disease-2019. *Hum Vaccin Immunother* 1–7.

Green, T.J., Speck, P., Geng, L., Raftos, D., Beard, M.R., and Helbig, K.J. (2015) Oyster viperin retains direct antiviral activity and its transcription occurs via a signalling pathway involving a heat-stable haemolymph protein. *J Gen Virol* **96**: 3587–3597.

Gribble, J., Stevens, L.J., Agostini, M.L., Anderson-Daniels, J., Chappell, J.D., Lu, X., *et al.* (2021) The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS Pathog* **17**: e1009226.

Gross, L.Z.F., Sacerdoti, M., Piiper, A., Zeuzem, S., Leroux, A.E., and Biondi, R.M. (2020) ACE2, the receptor that enables infection by SARS-CoV-2: Biochemistry, structure, allostery and evaluation of the potential development of ACE2 modulators. *ChemMedChem* **15**: 1682–1690.

Guan, Q., Sadykov, M., Mfarrej, S., Hala, S., Naeem, R., Nugmanova, R., *et al.* (2020) A genetic barcode of SARS-CoV-2 for monitoring global distribution of different clades during the COVID-19 pandemic. *International Journal of Infectious Diseases* **100**: 216–223.

Gupta, A., Sabarinathan, R., Bala, P., Donipadi, V., Vashisht, D., Katika, M.R., *et al.* (2021) A comprehensive profile of genomic variations in the SARS-CoV-2 isolates from the state of Telangana, India. *Journal of General Virology*.

Haas, P., Muralidharan, M., Krogan, N.J., Kaake, R.M., and Hüttenhain, R. (2021) Proteomic approaches to study SARS-CoV-2 biology and COVID-19 pathology. *J Proteome Res* **20**: 1133–1152.

Hamet, P., Pausova, Z., Attaoua, R., Hishmih, C., Haloui, M., Shin, J., *et al.* (2021) SARS-CoV-2 receptor ACE2 gene is associated with hypertension and severity of COVID 19: interaction with sex, obesity and smoking. *Am J Hypertens*.

Hartenian, E., Nandakumar, D., Lari, A., Ly, M., Tucker, J.M., and Glaunsinger, B.A. (2020) The molecular virology of coronaviruses. *J Biol Chem* **295**: 12910–12934.

Heurich, A., Hofmann-Winkler, H., Gierer, S., Liepold, T., Jahn, O., and Pöhlmann, S. (2014) TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike protein. *J Virol* **88**: 1293–1307.

Hikmet, F., Méar, L., Edvinsson, Å., Micke, P., Uhlén, M., and Lindskog, C. (2020) The protein expression profile of ACE2 in human tissues. *Mol Syst Biol* **16**: e9610.

Hodcroft, E.B., De Maio, N., Lanfear, R., MacCannell, D.R., Minh, B.Q., Schmidt, H.A., *et al.* (2021) Want to track pandemic variants faster? Fix the bioinformatics bottleneck. *Nature* **591**: 30–33.

Hodcroft, E.B., Domman, D.B., Snyder, D.J., Oguntuyo, K., Van Diest, M., Densmore, K.H., *et al.* (2021) Emergence in late 2020 of multiple lineages of SARS-CoV-2 Spike protein variants affecting amino acid position 677. *medRxiv*.

Hoffmann, M., Kleine-Weber, H., and Pöhlmann, S. (2020) A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol Cell* **78**: 779–784.e5.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., *et al.* (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**: 271-280.e8.

Horby, P.W., Pfeiffer, D., and Oshitani, H. (2013) Prospects for emerging infections in East and southeast Asia 10 years after severe acute respiratory syndrome. *Emerg Infect Dis* **19**: 853–860.

Hulswit, R.J.G., de Haan, C. a. M., and Bosch, B.-J. (2016) Coronavirus spike protein and tropism changes. *Adv Virus Res* **96**: 29–57.

Hur, S. (2019) Double-stranded RNA sensors and modulators in innate immunity. *Annu Rev Immunol* **37**: 349–375.

Issa, E., Merhi, G., Panossian, B., Salloum, T., and Tokajian, S. (2020) SARS-CoV-2 and ORF3a: nonsynonymous mutations, functional domains, and viral pathogenesis. *mSystems* **5**: e00266-20.

Jaroszewski, L., Iyer, M., Alisoltani, A., Sedova, M., and Godzik, A. (2020) The interplay of SARS-CoV-2 evolution and constraints imposed by the structure and functionality of its proteins. *bioRxiv*.

Ji, W., Wang, W., Zhao, X., Zai, J., and Li, X. (2020) Cross-species transmission of the newly identified coronavirus 2019-nCoV. *J Med Virol* **92**: 433–440.

Jin, X., Lian, J.-S., Hu, J.-H., Gao, J., Zheng, L., Zhang, Y.-M., *et al.* (2020) Epidemiological, clinical and virological characteristics of 74 cases of coronavirus-infected disease 2019 (COVID-19) with gastrointestinal symptoms. *Gut* **69**: 1002–1009.

Kang, D., Gao, S., Tian, Z., Huang, D., Guan, G., Liu, G., *et al.* (2020) Ovine viperin inhibits bluetongue virus replication. *Molecular Immunology* **126**: 87–94.

Karapiperis, C., Kouklis, P., Papastratos, S., Chasapi, A., Danchin, A., and Ouzounis, C.A. (2020) Preliminary evidence for seasonality of Covid-19 due to ultraviolet radiation. *F1000Res* **9**: 658.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059–3066.

Khayat, A.S., de Assumpção, P.P., Meireles Khayat, B.C., Thomaz Araújo, T.M., Batista-Gomes, J.A., Imbiriba, L.C., *et al.* (2020) ACE2 polymorphisms as potential players in COVID-19 outcome. *PLoS One* **15**: e0243887.

Kim, D.-K., Knapp, J.J., Kuang, D., Chawla, A., Cassonnet, P., Lee, H., *et al.* (2020) A comprehensive, flexible collection of SARS-CoV-2 coding regions. *G3 (Bethesda, Md)* **10**: 3399–3402.

Kleine-Weber, H., Schroeder, S., Krüger, N., Prokscha, A., Naim, H.Y., Müller, M.A., *et al.* (2020) Polymorphisms in dipeptidyl peptidase 4 reduce host cell entry of Middle East respiratory syndrome coronavirus. *Emerg Microbes Infect* **9**: 155–168.

Koyama, T., Platt, D., and Parida, L. (2020) Variant analysis of SARS-CoV-2 genomes. *Bull World Health Organ* **98**: 495–504.

Krueger, D.K., Kelly, S.M., Lewicki, D.N., Ruffolo, R., and Gallagher, T.M. (2001) Variations in disparate regions of the murine coronavirus spike protein impact the initiation of membrane fusion. *J Virol* **75**: 2792–2802.

Landoni, G., Zangrillo, A., Romero García, C.S., Faustini, C., Di Piazza, M., Conte, F., *et al.* (2020) Nations with high smoking rate have low SARS-CoV-2 infection and low COVID-19 mortality rate. *Acta Biomed* **91**: e2020168.

Lang, Y., Li, W., Li, Z., Koerhuis, D., van den Burg, A.C.S., Rozemuller, E., *et al.* (2020) Coronavirus hemagglutinin-esterase and spike proteins coevolve for functional balance and optimal virion avidity. *Proc Natl Acad Sci U S A* **117**: 25759–25770.

Laporte, M. and Naesens, L. (2017) Airway proteases: an emerging drug target for influenza and other respiratory virus infections. *Curr Opin Virol* **24**: 16–24.

Le Pendu, J., Breiman, A., Rocher, J., Dion, M., and Ruvoën-Clouet, N. (2021) ABO blood Types and COVID-19: Spurious, anecdotal, or truly important relationships? A reasoned review of available data. *Viruses* **13**:

Lemey, P., Ruktanonchai, N., Hong, S., Colizza, V., Poletto, C., den Broeck, F.V., *et al.* (2021) SARS-CoV-2 European resurgence foretold: interplay of introductions and persistence by leveraging genomic and mobility data. *Res Sq*.

Lemieux, R.U., LePendu, J., and Hindsgaul, O. (1979) The Lewis antigens and secretor status. *Jpn J Antibiot* **32 Suppl**: S21-31.

Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993.

Li, W., Hulswit, R.J.G., Kenney, S.P., Widjaja, I., Jung, K., Alhamo, M.A., *et al.* (2018) Broad receptor engagement of an emerging global coronavirus may potentiate its diverse cross-species transmissibility. *Proc Natl Acad Sci U S A* **115**: E5135–E5143.

Lieber, C., Melekidis, S., Koch, R., and Bauer, H.-J. (2021) Insights into the evaporation characteristics of saliva droplets and aerosols: Levitation experiments and numerical modeling. *J Aerosol Sci* 105760.

Littler, D.R., Gully, B.S., Colson, R.N., and Rossjohn, J. (2020) Crystal structure of the SARS-CoV-2 non-structural protein 9, Nsp9. *iScience* **23**: 101258.

Liu, S., Shen, J., Yang, L., Hu, C.-D., and Wan, J. (2020) Distinct genetic spectrums and evolution patterns of SARS-CoV-2, Health Informatics.

Lu, S., Ye, Q., Singh, D., Cao, Y., Diedrich, J.K., Yates, J.R., *et al.* (2021) The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat Commun* **12**: 502.

Lubin, J.H., Zardecki, C., Dolan, E.M., Lu, C., Shen, Z., Dutta, S., *et al.* (2020) Evolution of the SARS-CoV-2 proteome in three dimensions (3D) during the first six months of the COVID-19 pandemic. *bioRxiv*.

Lukassen, S., Chua, R.L., Trefzer, T., Kahn, N.C., Schneider, M.A., Muley, T., *et al.* (2020) SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. *EMBO J* **39**: e105114.

Luo, R., Wong, Y.-S., and Lam, T.-W. (2020) Tracking cytosine depletion in SARS-CoV-2, *Bioinformatics*.

Madhugiri, R., Fricke, M., Marz, M., and Ziebuhr, J. (2016) Coronavirus cis-acting RNA elements. *Adv Virus Res* **96**: 127–163.

Makarenkov, V., Mazouze, B., Rabusseau, G., and Legendre, P. (2021) Horizontal gene transfer and recombination analysis of SARS-CoV-2 genes helps discover its close relatives and shed light on its origin. *BMC Ecol Evol* **21**: 5.

Matyášek, R. and Kovařík, A. (2020) Mutation patterns of human SARS-CoV-2 and bat RaTG13 coronavirus genomes are strongly biased towards C>U transitions, indicating rapid evolution in their hosts. *Genes (Basel)* **11**:

Moustaqil, M., Ollivier, E., Chiu, H.-P., Van Tol, S., Rudolffi-Soto, P., Stevens, C., *et al.* (2021) SARS-CoV-2 proteases PLpro and 3CLpro cleave IRF3 and critical modulators of inflammatory pathways (NLRP12 and TAB1): implications for disease presentation across species. *Emerg Microbes Infect* **10**: 178–195.

Moya, A., Holmes, E.C., and González-Candelas, F. (2004) The population genetics and evolutionary epidemiology of RNA viruses. *Nat Rev Microbiol* **2**: 279–288.

Nagy, Á., Pongor, S., and Gyórfy, B. (2021) Different mutations in SARS-CoV-2 associate with severe and mild outcome. *Int J Antimicrob Agents* **57**: 106272.

Nan, Y., Nan, G., and Zhang, Y.-J. (2014) Interferon induction by RNA viruses and antagonism by viral pathogens. *Viruses* **6**: 4999–5027.

Neches, R.Y., Kyripides, N.C., and Ouzounis, C.A. (2021) Atypical divergence of SARS-CoV-2 Orf8 from Orf7a within the coronavirus lineage suggests potential stealthy viral strategies in Immune evasion. *mBio* **12**:

Neuman, B.W., Joseph, J.S., Saikatendu, K.S., Serrano, P., Chatterjee, A., Johnson, M.A., *et al.* (2008) Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. *J Virol* **82**: 5279–5294.

Ng, K.W., Faulkner, N., Cornish, G.H., Rosa, A., Harvey, R., Hussain, S., *et al.* (2020) Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. *Science* **370**: 1339–1343.

Ng, T.-W., Turinici, G., Ching, W.-K., Chung, S.-K., and Danchin, A. (2007) A parasite vector-host epidemic model for TSE propagation. *Med Sci Monit* **13**: BR59-66.

Ng, T.W., Turinici, G., and Danchin, A. (2003) A double epidemic model for the SARS propagation. *BMC Infect Dis* **3**: 19.

Nordgren, J. and Svensson, L. (2019) Genetic susceptibility to human norovirus infection: An update. *Viruses* **11**:

Norell Bergendahl, M. and Stanford University eds. (2009) Design theory and research methodology, Glasgow: Design Society.

Normand, S., Waldschmitt, N., Neerincx, A., Martinez-Torres, R.J., Chauvin, C., Couturier-Maillard, A., *et al.* (2018) Proteasomal degradation of NOD2 by NLRP12 in monocytes promotes bacterial tolerance and colonization by enteropathogens. *Nat Commun* **9**: 5338.

Ogando, N.S., Zevenhoven-Dobbe, J.C., van der Meer, Y., Bredenbeek, P.J., Posthuma, C.C., and Snijder, E.J. (2020) The enzymatic activity of the nsp14 exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2. *J Virol* **94**:

Ortego, J., Ceriani, J.E., Patiño, C., Plana, J., and Enjuanes, L. (2007) Absence of E protein arrests transmissible gastroenteritis coronavirus maturation in the secretory pathway. *Virology* **368**: 296–308.

Östbye, H., Gao, J., Martinez, M.R., Wang, H., de Gier, J.-W., and Daniels, R. (2020) N-linked glycan sites on the influenza A virus neuraminidase head domain are required for efficient viral incorporation and replication. *J Virol* **94**:

Ou, Z., Ouzounis, C., Wang, D., Sun, W., Li, J., Chen, W., *et al.* (2020) A path towards SARS-CoV-2 attenuation: metabolic pressure on CTP synthesis rules the virus evolution.

Oulas, A., Zanti, M., Tomazou, M., Zachariou, M., Minadakis, G., Bourdakou, M.M., *et al.* (2021) Generalized linear models provide a measure of virulence for specific mutations in SARS-CoV-2 strains. *PLoS One* **16**: e0238665.

Ovsyannikova, I.G., Haralambieva, I.H., Crooke, S.N., Poland, G.A., and Kennedy, R.B. (2020) The role of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility and severity. *Immunol Rev* **296**: 205–219.

Oyston, P. and Robinson, K. (2012) The current challenges for vaccine development. *Journal of Medical Microbiology* **61**: 889–894.

Papa, G., Mallery, D.L., Albecka, A., Welch, L.G., Cattin-Ortolá, J., Luptak, J., *et al.* (2021) Furin cleavage of SARS-CoV-2 Spike promotes but is not essential for infection and cell-cell fusion. *PLoS Pathog* **17**: e1009246.

Pasick, J.M., Kalicharran, K., and Dales, S. (1994) Distribution and trafficking of JHM coronavirus structural proteins and virions in primary neurons and the OBL-21 neuronal cell line. *J Virol* **68**: 2915–2928.

Patiño-Galindo, J.Á., Filip, I., and Rabadan, R. (2021) Global patterns of recombination across human viruses. *Mol Biol Evol.*

Pauly, M.D. and Luring, A.S. (2015) Effective lethal mutagenesis of influenza virus by three nucleoside analogs. *J Virol* **89**: 3584–3597.

Pease, L.F., Wang, N., Salsbury, T.I., Underhill, R.M., Flaherty, J.E., Vlachokostas, A., *et al.* (2021) Investigation of potential aerosol transmission and infectivity of SARS-CoV-2 through central ventilation systems. *Build Environ* 107633.

Perveen, S., Khalili Yazdi, A., Devkota, K., Li, F., Ghiabi, P., Hajian, T., *et al.* (2021) A high-throughput RNA displacement assay for screening SARS-CoV-2 nsp10-nsp16 complex toward developing therapeutics for COVID-19. *SLAS Discov* 2472555220985040.

Piacenza, L., Trujillo, M., and Radi, R. (2019) Reactive species and pathogen antioxidant networks during phagocytosis. *J Exp Med* **216**: 501–516.

Pillon, M.C., Frazier, M.N., Dillard, L.B., Williams, J.G., Kocaman, S., Krahn, J.M., *et al.* (2021) Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics. *Nat Commun* **12**: 636.

Rasschaert, D., Duarte, M., and Laude, H. (1990) Porcine respiratory coronavirus differs from transmissible gastroenteritis virus by a few genomic deletions. *Journal of General Virology* **71**: 2599–2607.

Reid, A.H., Fanning, T.G., Janczewski, T.A., Lourens, R.M., and Taubenberger, J.K. (2004) Novel origin of the 1918 pandemic influenza virus nucleoprotein gene. *J Virol* **78**: 12462–12470.

Saurabh, S., Verma, M.K., Gautam, V., Kumar, N., Jain, V., Goel, A.D., *et al.* (2021) Tobacco, alcohol use and other risk factors for developing symptomatic COVID-19 vs asymptomatic SARS-CoV-2 infection: a case-control study from western Rajasthan, India. *Trans R Soc Trop Med Hyg*.

Schetelig, J., Baldauf, H., Wendler, S., Heidenreich, F., Real, R., Kolditz, M., *et al.* (2021) Blood group A epitopes do not facilitate entry of SARS-CoV-2. *J Intern Med*.

Schinkelshoek, M.S., Fronczek, R., Kooy-Winkelaar, E.M.C., Petersen, J., Reid, H.H., van der Heide, A., *et al.* (2019) H1N1 hemagglutinin-specific HLA-DQ6-restricted CD4+ T cells can be readily detected in narcolepsy type 1 patients and healthy controls. *J Neuroimmunol* **332**: 167–175.

Sengupta, A., Hassan, S.S., and Choudhury, P.P. (2021) Clade GR and clade GH isolates of SARS-CoV-2 in Asia show highest amount of SNPs. *Infect Genet Evol* **89**: 104724.

Shang, J., Wan, Y., Liu, C., Yount, B., Gully, K., Yang, Y., *et al.* (2020) Structure of mouse coronavirus spike protein complexed with receptor reveals mechanism for viral entry. *PLoS Pathog* **16**: e1008392.

Sharma, V.K., Jinadatha, C., Lichtfouse, E., Decroly, E., van Helden, J., Choi, H., and Chatterjee, P. (2021) COVID-19 epidemiologic surveillance using wastewater. *Environ Chem Lett* s10311-021-01188-w.

Shea, J.-E., Best, R.B., and Mittal, J. (2021) Physics-based computational and theoretical approaches to intrinsically disordered proteins. *Curr Opin Struct Biol* **67**: 219–225.

Shen, Y.-Y., Liang, L., Zhu, Z.-H., Zhou, W.-P., Irwin, D.M., and Zhang, Y.-P. (2010) Adaptive evolution of energy metabolism genes and the origin of flight in bats. *Proceedings of the National Academy of Sciences* **107**: 8666–8671.

Shope, R.E. (1936) The incidence of neutralizing antibodies for swine influenza virus in the sera of human beings of different ages. *J Exp Med* **63**: 669–684.

Siddell, S.G., Barthel, A., and ter Meulen, V. (1981) Coronavirus JHM: a virion-associated protein kinase. *J Gen Virol* **52**: 235–243.

Simmonds, P. (2020) Rampant C→U hypermutation in the genomes of SARS-CoV-2 and other coronaviruses: causes and consequences for their short- and long-term evolutionary trajectories. *mSphere* **5**: e00408-20.

Singh, G., Singh, P., Pillatzki, A., Nelson, E., Webb, B., Dillberger-Lawson, S., and Ramamoorthy, S. (2019) A minimally replicative vaccine protects vaccinated piglets against challenge with the porcine epidemic diarrhea virus. *Front Vet Sci* **6**: 347.

Slanina, H., Madhugiri, R., Bylapudi, G., Schultheiß, K., Karl, N., Gulyaeva, A., *et al.* (2021) Coronavirus replication-transcription complex: Vital and selective NMPylation of a conserved site in nsp9 by the NiRAN-RdRp subunit. *Proc Natl Acad Sci U S A* **118**:

Smith, K.A. (2012) Louis Pasteur, the father of immunology? *Front Immunol* **3**: 68.

Souilmi, Y., Lauterbur, M.E., Tobler, R., Huber, C.D., Johar, A.S., and Enard, D. (2020) An ancient viral epidemic involving host coronavirus interacting genes more than 20,000 years ago in East Asia, *Evolutionary Biology*.

Sullivan, Gregory F, aka O., Garson (2013) It's difficult to amke predictions, especially about the future. *Quote Investigator*.

Taleb, N. (2008) *The black swan: the impact of the highly improbable*, New York: Random House.

Tharappel, A.M., Samrat, S.K., Li, Z., and Li, H. (2020) Targeting crucial host factors of SARS-CoV-2. *ACS Infect Dis* **6**: 2844–2865.

Theiler, M. and Smith, H.H. (1937) The use of yellow fever virus modified by in vitro cultivation for human immunization. *The Journal of Experimental Medicine* **65**: 787–800.

Tidu, A., Janvier, A., Schaeffer, L., Sosnowski, P., Kuhn, L., Hammann, P., *et al.* (2020) The viral protein NSP1 acts as a ribosome gatekeeper for shutting down host translation and fostering SARS-CoV-2 translation. *RNA*.

Tortorici, M.A. and Veesler, D. (2019) Structural insights into coronavirus entry. *Adv Virus Res* **105**: 93–116.

Truscott, R.J.W., Schey, K.L., and Friedrich, M.G. (2016) Old proteins in man: A field in its infancy. *Trends Biochem Sci* **41**: 654–664.

Trypsteen, W., Van Cleemput, J., Snippenberg, W. van, Gerlo, S., and Vandekerckhove, L. (2020) On the whereabouts of SARS-CoV-2 in the human body: A systematic review. *PLoS Pathog* **16**: e1009037.

Turinici, G. and Danchin, A. (2007) The SARS case study: An alarm clock? In *Encyclopedia of Infectious Diseases*. Tibayrenc, M. (ed). Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 151–162.

Vilar, S. and Isom, D.G. (2021) One year of SARS-CoV-2: How much has the virus changed? *Biology (Basel)* **10**:

Walls, A.C., Tortorici, M.A., Frenz, B., Snijder, J., Li, W., Rey, F.A., *et al.* (2016) Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat Struct Mol Biol* **23**: 899–905.

Wang, D., Baudys, J., Bundy, J.L., Solano, M., Keppel, T., and Barr, J.R. (2020) Comprehensive analysis of the glycan complement of SARS-CoV-2 spike proteins using signature ions-triggered electron-transfer/higher-energy collisional dissociation (EThcD) mass spectrometry. *Anal Chem* **92**: 14730–14739.

Wang, H., Li, X., Li, T., Zhang, S., Wang, L., Wu, X., and Liu, J. (2020) The genetic sequence, origin, and diagnosis of SARS-CoV-2. *Eur J Clin Microbiol Infect Dis* **39**: 1629–1635.

Wang, K., Chen, W., Zhang, Z., Deng, Y., Lian, J.-Q., Du, P., *et al.* (2020) CD147-spike protein is a novel route for SARS-CoV-2 infection to host cells. *Signal Transduct Target Ther* **5**: 283.

Wei, J., Guo, S., Long, E., Zhang, L., Shu, B., and Guo, L. (2021) Why does the spread of COVID-19 vary greatly in different countries? Revealing the efficacy of face masks in epidemic prevention. *Epidemiol Infect* **149**: e24.

Wesley, R.D., Woods, R.D., Hill, H.T., and Biwer, J.D. (1990) Evidence for a porcine respiratory coronavirus, antigenically similar to transmissible gastroenteritis virus, in the United States. *J VET Diagn Invest* **2**: 312–317.

Williams, F.M.K., Freidin, M.B., Mangino, M., Couvreur, S., Visconti, A., Bowyer, R.C.E., *et al.* (2021) Self-reported symptoms of COVID-19, including symptoms most predictive of SARS-CoV-2 infection, are heritable. *Twin Res Hum Genet* 1–6.

Wood, C.L., McInturff, A., Young, H.S., Kim, D., and Lafferty, K.D. (2017) Human infectious disease burdens decrease with urbanization but not with biodiversity. *Phil Trans R Soc B* **372**: 20160122.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., *et al.* (2020) A new coronavirus associated with human respiratory disease in China. *Nature* **579**: 265–269.

Wurtmann, E.J. and Wolin, S.L. (2009) RNA under attack: cellular handling of RNA damage. *Crit Rev Biochem Mol Biol* **44**: 34–49.

Xia, S., Duan, K., Zhang, Yuntao, Zhao, D., Zhang, H., Xie, Z., *et al.* (2020) Effect of an inactivated vaccine against SARS-CoV-2 on safety and immunogenicity outcomes: interim analysis of 2 randomized clinical trials. *JAMA* **324**: 951–960.

Xu, Y.-R. and Lei, C.-Q. (2020) TAK1-TABs complex: A central semiosome in inflammatory responses. *Front Immunol* **11**: 608976.

Yamada, Y. and Liu, D.X. (2009) Proteolytic activation of the spike protein at a novel RRRR/S motif is implicated in furin-dependent entry, syncytium formation, and infectivity of coronavirus infectious bronchitis virus in cultured cells. *J Virol* **83**: 8744–8758.

Yan, L., Ge, J., Zheng, L., Zhang, Y., Gao, Y., Wang, T., *et al.* (2021) Cryo-EM structure of an extended SARS-CoV-2 replication and transcription complex reveals an intermediate state in cap synthesis. *Cell* **184**: 184-193.e10.

Yang, H.-C., Chen, C.-H., Wang, J.-H., Liao, H.-C., Yang, C.-T., Chen, C.-W., *et al.* (2020) Analysis of genomic distributions of SARS-CoV-2 reveals a dominant strain type with strong allelic associations. *Proc Natl Acad Sci U S A* **117**: 30679–30686.

Yang, J., Wang, W., Chen, Z., Lu, S., Yang, F., Bi, Z., *et al.* (2020) A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity. *Nature* **586**: 572–577.

Yang, Y., Yue, Y., Song, N., Li, C., Yuan, Z., Wang, Y., *et al.* (2020) The YdiU domain modulates bacterial stress signaling through Mn²⁺-dependent UMPylation. *Cell Reports* **32**: 108161.

Yost, S.A. and Marcotrigiano, J. (2013) Viral precursor polyproteins: keys of regulation from replication to maturation. *Curr Opin Virol* **3**: 137–142.

Zeng, H.-L., Dichio, V., Rodríguez Horta, E., Thorell, K., and Aurell, E. (2020) Global analysis of more than 50,000 SARS-CoV-2 genomes reveals epistasis between eight viral genes. *Proc Natl Acad Sci U S A* **117**: 31519–31526.

Zeng, Z., Li, T.-T., Jin, X., Peng, F.-H., Song, N.-H., Peng, G.-Q., and Ge, X.-Y. (2017) Coexistence of multiple genotypes of porcine epidemic diarrhea virus with novel mutant S genes in the Hubei Province of China in 2016. *Virol Sin* **32**: 298–306.

Zhang, R., Li, Y., Zhang, A.L., Wang, Y., and Molina, M.J. (2020) Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proc Natl Acad Sci USA* **117**: 14857–14863.

Zhang, X.W., Yap, Y.L., and Danchin, A. (2005) Testing the hypothesis of a recombinant origin of the SARS-associated coronavirus. *Arch Virol* **150**: 1–20.

Zhao, J., Sun, L., Zhao, Y., Feng, D., Cheng, J., and Zhang, G. (2020) Coronavirus endoribonuclease ensures efficient viral replication and prevents protein kinase R activation. *J Virol*.

Zhao, Z., Qin, P., and Huang, Y.-W. (2021) Lysosomal ion channels involved in cellular entry and uncoating of enveloped viruses: Implications for therapeutic strategies against SARS-CoV-2. *Cell Calcium* **94**: 102360.

Zhou, D., Dejnirattisai, W., Supasa, P., Liu, C., Mentzer, A.J., Ginn, H.M., *et al.* (2021) Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine induced sera. *Cell* S0092867421002269.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., *et al.* (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**: 270–273.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., *et al.* (2020) A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* **382**: 727–733.

Ziegler, C.G.K., Allon, S.J., Nyquist, S.K., Mbanjo, I.M., Miao, V.N., Tzouanas, C.N., *et al.* (2020) SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell* **181**: 1016-1035.e19.

Zou, X., Chen, K., Zou, J., Han, P., Hao, J., and Han, Z. (2020) Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. *Front Med* **14**: 185–192.