

談談 SARS-CoV-2 的生物學及其變異毒株：對病毒進化的預期以及我們能做些什麼

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

翻譯者：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

通訊方法

電話: +331 4441 2551; 傳真: +331 4441 2559

電子郵件: antoine.danchin@normalesup.org

摘要

SARS-CoV-2 的全球傳播和越來越多變異毒株被識別（其中一些已取代原始進化枝成為優勢毒株）突顯了該病毒正在努力拓展其進化空間。病毒越是長時間高水準傳播及繁衍，就越能適應人類宿主並找到繼續流行的途徑。第三波 COVID-19 的疫情正在多個國家和地區發生，凸顯了當前遏制傳播的措施的不足。制定這些措施所依據的其中一個理念是，接種疫苗將迅速阻礙病毒的傳播並使社會回歸常態。然而，接種疫苗本身將給疫苗帶來選擇壓力並促使讓疫苗失效的變異毒株的產生。因此，在大規模疫苗接種期間同時放任病毒繼續高水準傳播繁殖，並期待疫苗能終止這場大流行的政策，是不現實的。在缺乏有效的抗病毒藥物的情況下，預防 SARS-CoV-2 的方式將於與預防流感相似，比如每年都需要針對優勢毒株進行疫苗接種。能否戰勝 COVID-19 或與其共存是 SARS-CoV-2 將如何進化和變異的問題。因此，為了更好地預測未來的走向，理解 SARS-CoV-2 將如何進化及各種

限制因素至關重要。迄今為止，研究的主要焦點還集中在負責與受體結合的突刺蛋白上。但該病毒總共編碼 26 種與宿主以各種方式相互作用的蛋白，因此其進化有多種可能性，並且難以預知。但不管怎樣，如果要對 COVID-19 進行最佳防禦，就必須對其可能發生的變種進行防禦。要做到這一點，我們必須要瞭解該病毒的各種進化可能性。除了該病毒與細胞間一般的相互作用外，人類中還存在著大量的多態性（例如 Lewis，HLA 等），有些存在於大多或所有人口中，有些則只存在於特定的種族。這些多態性也是推動或限制病毒進化因素。現在，我們擁有足夠的資源和手段（包括病毒基因組測序，蛋白質結構測定/建模，蛋白質相互作用分析），對變異毒株進行功能測定，但獲得全面的基因組資料並非易事。再者，如需瞭解病毒進化對傳播和致病能力的影響，我們必須將基因組資料與傳播（病毒流行病學）和疾病資料（患者臨床資料），以及人群大小的資訊聯繫起來。在這篇社論中，我們探討了病毒生物學中的關鍵方面，以及人類多態性、人類行為、地理和氣候的相關方面對病毒進化的影響，並在此基礎上提出了一系列建議，以監測病毒的進化並預測哪些類型的變異將會出現。

引言

當前的 COVID-19 大流行像極了一個即時發生的分子進化實驗。打頭天起，它就一步步向被蒙在鼓裡的群眾展示了病毒的行為和能力。這個“實驗”幾乎涵蓋了整個地球 78 億具有豐富環境多樣性和遺傳多態性的智人。SARS-CoV-2 β 冠狀病毒的起源仍然不明 – 除了我們知道一個在蝙蝠中找到的與 SARS-CoV-2 進化距離相近的病毒 (Ji *et al.*, 2020; Makarenkov *et al.*, 2021) – 但現在，我們可以通過它在不同國家和人群間的傳播特徵，較準確地追蹤其進化過程了。截至 2021 年 3 月 2 日，通過 Global Initiative on Sharing All Influenza Data (GISAID, <https://www.gisaid.org>) 資料庫獲得 193,687 個 SARS-CoV-2 全基因組，並將它們與首個公開毒株“Wuhan-Hu-1” [INSDC AccNum MN908947.3, (H. Wang *et al.*, 2020; Wu *et al.*, 2020; Zhu *et al.*, 2020)] 的序列進行比較，我們找到了 19,794 個包含一個或以上單域基突變 (SNP) 的基因組位置。

在我們使用這些突變作為研究 SARS-CoV-2 進化和 COVID-19 發展的重要特徵的工具的同時，需要保持警惕。因為其中一些突變可能只是源於糟糕的資料品質，而非代表真正的病毒多樣性。另外，病毒序列採集的地理分佈不均勻問題可能導致部分觀測到的突變頻率

偏離實際。識別跟蹤具有行為變化的變異毒株（行為包括毒力、趨向性、和傳播）是病毒基因組監測的關鍵目標。我們在 SARS-CoV-2 中看到，隨著時間的推移，某些積累下來的突變加強了病毒的傳播能力，某些突變也改變了病毒的毒力 (Oulas *et al.*, 2021)。收集更多毒性有所加強或削弱的毒株的突變資料有助於得到更準確的分析結果。但鑒於病毒變異、地理條件及人類多態性這些因素的變化範圍之廣，我們在分析這些突變時需保持謹慎（參見補充圖片 1 及補充表格 1 的注解）。對這些因素進行全面分析的難度，從品質參差不齊的關於 SARS-CoV-2 和 COVID-19 的著作的大量發表（截止 2021 年 3 月 3 日在 PubMed 上已查到 109,000 個相關文獻），以及病毒序列資料庫對病毒中繼資料採集的不足，就可見一斑。在本社論中，我們盡可能採用顯而易見的分析思路（以避免追求分析熱點和依賴於文獻統計所產生的偏差）來預期病毒和大流行病演變的特徵。我們將通過分析病原體各種功能及其與宿主的相互作用的兩個基礎病毒生物學手段（基於其兩個原始功能-傳播和生存-隨著時間的變化）來重新審視 SARS-CoV-2 將遇到的可能會影響流行病的發展的進化限制。這將突出一些我們雖然正在目睹但任未受到廣泛關注特徵，這些特徵可以說明我們預測這場大流行過程中可能發生的狀況。

關於進化的幾個重要方面

從定義上講，進化學是基於已發生的事實所總結出來的。“時間”做為其變數意味著，對一個生物體分別進行短期和長期的演化分析，將得出非常不同的結果。此外，進行進化結果的預測是冒險的，因為進化是短視的。它不具備對未來進行宏偉設計的能力。但是，從長遠來看，某種生物能存活下來一定歸因於某些使其能夠在那段時間內繼續繁殖的能力。從短期看則有所不同。其存活能力只與少數與繁衍後代相關的能力有關，而非使其有能力在遙遠的將來繼續傳播的能力。一個病毒從一個已經與之共存了很長時間的宿主轉移到一個新的宿主，將使該病毒突然暴露在一個陌生的環境中。但短時間裡它將仍然遵循和發揮使其能在先前宿主中存活的能力。同樣的，新宿主也是長期演進的結果。新宿主儘管對於新入侵的病毒還很無知，但長期的自然選擇也使新宿主擁有各種常見的應激反應來應付入侵。就以針對病毒入侵而言，先天免疫能力將通過特徵識別新病毒的存在，進而預防或減慢病毒的發展 (Nan *et al.*, 2014; Chen *et al.*, 2017; Hur, 2019)。許多這些能力是動物甚

至植物所共有的 – 意外發現 Toll 受體在果蠅中的作用就是個好例子 (Belvin and Anderson, 1996)。

進化無法預先確定一個病毒在較遠的將來能否繼續產生後代。所以從短期來看，這意味著一個“成功”的病毒將忽略其宿主的生死並極盡所能產生後代（除非宿主在病毒產生後代之前就已經死去）。而從長遠來看，為了能持續傳代，導致高死亡率的強病毒毒力是不可持續的。這項病毒的基本特性將啟示我們，隨著疫情的推移，病毒可能會發生什麼功能變化。不幸的是，對進化多樣性理解的努力導致了好些過度簡化的假說。這些假說以人類為中心並帶有一些經濟學和道德的特點：一個生物體的行為可被視為“無私的”或“自私的”。突變可被分為“有利的”、“有害的”或“中性的”。這些假說是否適用於病毒的進化值得探討。以一個病毒的進化過程為例，能提高病毒傳代能力的方法多種多樣。比方說，即便一個域基變化並不導致一個基因功能的改變，它也有可能影響宿主的代謝系統、小幅度地改進病毒的複製酶、干擾免疫系統、影響體溫、改變重要 tRNA 的可得性、與另一種病原體發生協同感染等。在一場大流行中，這些方法可同時發生而互相之間並不一定有一致性。這將直接影響病毒的繁衍，並受到包含如上所述的各個過程和系統的約束。沒有任何的一個突變是真正“中性”的，中性或類似的溫和解釋或許只是用於描述我們不瞭解的過程並掩蓋無知的一種手段。如果我們想要預測這場大流行的走向，哪怕只是一小部分，我們也要避免被“中立”的看法所誤導，即便“黑天鵝事件”是生物進化中常有的事 (Taleb, 2008)。

病毒基因組中一個的變化將產生怎樣的影響將隨著時間的推移而顯現。從短期、中期或長期來看，其結果會有所不同。對我們而言，重要的不是我們要應用的演化模型，而是要闡明分別在流行病的擴張、達到頂峰和散去時，有可能會或一定會出現的一系列特徵。從這方面來說，當前的流行病是高度可預測的，且實際上在許多研究中都已做過預測 (Moya *et al.*, 2004; Turinici and Danchin, 2007; Horby *et al.*, 2013)。與其使用模型來進行預測，我們的研究將嘗試找到在病毒感染期間將可能發揮作用的功能，並理解它們將如何幫助我們推斷該場流行病的未來。

功能分析：從功能到序列，而不是相反

眾所周知，生物學中功能的概念複雜無比 (Allen *et al.*, 1998)。讓我們在這裡使用工業設計作比喻。為了實現一個有意義的設計，在工業設備和流程設計中首先要瞭解是與該設計相關的功能，即功能分析 [Fantoni *et al.* in (Norell Bergendahl and Stanford University, 2009)。一種常用的分析方法是自底向上，列出各個必要元件並將它們組合為更加複雜的部分，直到完成。相比之下，從最終用戶的角度出發，一種更好的分析方法是自上而下。我們首先確定設備的主功能（例如，印表機的主功能是列印），然後逐步識別主功能運行所需的協助工具（例如，進紙機構，供墨機構，供電機構，控制機構等），最終得到構成了該設備的基本元件。話雖這麼說，但瞭解病毒的功能，尤其是其主功能並非易事。在這裡我們提出病毒有兩個必須的，截然不同但相互影響的主功能：繁殖（產生後代）和探索（到達宿主）。這兩個功能的實現方式基本上是無限制的。但這並不意味著它們不會受到外在約束：它們的組成部分，與生命過程、特定構造單元、大分子相關。幸運的是，這縮小了我們的研究範圍。由於病毒早在動物出現之前就已存在，動物早已設計出旨在對付病毒入侵的各種功能。

探索

讓我們來聊聊與主功能中的“探索”相關的病毒功能和宿主反應。任何與生物有互動的個體的主功能在不同的層面上都離不開“到達宿主”這個協助工具。對我們來說病原性病毒的目標是人類。病毒首次接觸的是一個人的皮膚等表層，這規範了病毒進入宿主的途徑。以下是幾個常見例子：

- 面部/腸道的趨向性，這被我們有意無意地使用雙手深深影響著（只需要看看周圍有多少人在碰自己的臉就可以知道了）。這類趨向性可以根據部位具體分為眼睛、鼻子、呼吸道、嘴巴以及腸道，而且可能受到常見的糞口傳播影響，在 COVID-19 疫情爆發時都有發現與這類趨向性相關的例子 (Jin *et al.*, 2020)。在廢水中檢測到的病毒反映了經腸道傳播的重要性 (Sharma *et al.*, 2021)。但暫時發現通過這種途徑感染 SARS-CoV-2 個案還比較少 (Goldman,

2020)。然而，我們必須要意識到在對病毒進行預測的時候腸道傳播有可能會突然變成了一個主導的方式，這一點在下文會進行討論。

- 通過氣溶膠傳播的病毒有著直接的呼吸道趨向性。當感染集中在鼻腔或是深入肺部有著不同的嚴重性，所以我們應該謹慎地監測著這類傳播途徑。這也指出了病毒可通過不同宿主細胞進入身體。
- 通過媒介傳播的病毒：普遍的傳播方式是通過皮膚和血液。雖然我們暫且不考慮新型感染方式的風險，但不排除有更加複雜的情況已發生，譬如當病原體本身是寄生蟲的宿主 (Ng *et al.*, 2007)。

COVID-19 大流行主要與上述前兩種傳播途徑有關，但是我們應該注意到取決於人類習慣的變化（譬如放寬手部和臉部衛生的規範），使很容易為病毒的進化打開的新窗口。

關於人類宿主

COVID-19 有著很多人類社會疾病的特徵 (Danchin, 2003)，即便只是暫時的，但人們也不願作出行為上的改變 (Cherif *et al.*, 2016)。由於社會中存在著各種政治原因，以及不同社會體系對於公益的重要性有著不同的看法，懸浮微粒的作用一直存在爭議，尤其是因為戴口罩主要用於保護他人而不是佩戴者 (Wei *et al.*, 2021)

（<https://interactives.lowyinstitute.org/features/covid-performance/>）。而在不同社區中對所謂的共同利益則有著不同的看法。然而即使在戶外懸浮微粒也可以將病毒長時間保持在空氣中 (Zhang *et al.*, 2020)，因此這很可能會成為 SARS-CoV-2 主要傳播途徑 (Dumont-Leblond *et al.*, 2020; Lieber *et al.*, 2021)。相關的爭議導致出現大量關於配戴口罩的不當建議，因此有著極大的破壞性 (Czypionkae *et al.*, 2020)。而忽略這個傳播途徑的重要性則成了病毒長久存在的主要風險，長期而言會產生危險的毒株。許多人類的習性，譬如喜歡在擁擠的環境中進行社交活動，也促進了疾病的傳播。以下幾個有關感染源頭的例子說明了各種觸發感染的可能性以及病毒的傳播方式和持久性。

環境變化

呼吸系統疾病有著很大的季節性 (Audi *et al.*, 2020)。但是那確切的因素與傳染病之間真正的因果關係可能會被其他的混雜因素掩蓋了。季節性、污染、城市化、生物多樣性或緯度都被認為與傳染病有關 (Wood *et al.*, 2017)。但事實上，溫度或雨水可能會改變人類在室內外的模式，這人為因素很少被考慮進去 (Bulfone *et al.*, 2020)。感染的可能性與空氣中的病毒濃度有關，因此從感染者所呼出的病毒的濃度是傳播的關鍵，在擁擠的環境（包括室外）中高度密密集的人群必定會有著高度傳染性 (Derjany *et al.*, 2020)。此外，與農村地區的開放環境相比，城市環境中的防風屏障會導致空氣的停滯，對稀釋病毒非常不利。顆粒物污染也與疾病的嚴重程度有關 (Brauer *et al.*, 2021)。譬如，紫外線在室外的分佈會根據緯度和海拔發生巨大的變化 (Karapiperis *et al.*, 2020)。因為不同感染和傳播途徑的選擇壓力會有所不同，病毒的進化模式會在不同的環境下有所區別。因此單單瞭解衣殼和突刺蛋白是不夠的，瞭解病毒包膜的作用也非常重要。

肺部和腸道的趨向性

就如其他冠狀病毒一樣，腸道和呼吸道在預測當前大流行的未來時也都至關重要。譬如豬傳染性胃腸炎病毒 (transmissible gastroenteritis alphacoronavirus, TGEV) 在新生小豬的小腸的腸絨毛細胞以及肺部細胞中都可以複製，該病毒早在 1946 年對美國養豬場造成近 100% 的死亡率，對養豬業做成重大的打擊 (Doyle and Hutchings, 1946)。另一次爆發是在 1984 年，豬呼吸道冠狀病毒 (porcine respiratory coronavirus, PRCV) 肆虐 (Wesley *et al.*, 1990)。人們發現 PRCV 帶有一些缺失突變的 TGEV 突變體，這些缺失使病毒的趨向性從腸上皮轉為呼吸道上皮 (Rasschaert *et al.*, 1990)。值得注意的是 PRCV 感染能保護豬免受 TGEV 的侵害，形成了天然的疫苗 (Bernardet *et al.*, 1989)。從這個例子我們可以知道趨向性的變化與病毒毒力的減弱有關。這些觀察結果表明應該把改變腸道健康的因素納入監測範圍。這種樂觀的進化趨勢極少發生並且與趨向性無關，並取決於一系列的巧合：先是病毒毒力減弱，緊接著有效地傳播減毒變體，最後針對主要病原體進行免疫交叉保護。儘管減毒變體可能與病毒傳播同時出現，但是這三個事件在短期內自然地發生的可

能性很小。不管怎樣，讓人倍感意外的是豬 TGEV 突刺蛋白 N 末端的兩個氨基酸變化足以導致病毒喪失對腸道的趨向性 (Ballesteros *et al.*, 1997)，因此我們應該記住這點並積極尋找減低病毒毒力的方式。然而不幸的是，冠狀病毒經常與其他 RNA 病毒進行重組 (Zhanget *al.*, 2005; Chenet *al.*, 2019)，這過程可能會阻礙減毒變體的有效傳播（見下文進一步討論）。

人類的多態性

我們近年才慢慢地開始意識到人類群組的遺傳結構的特徵，但這也是應該被納入考慮的因素之一 (Williams *et al.*, 2021)。雖然在 3 月 21 日那一天於 PubMed 中並沒有搜索到“human polymorphism” AND (“Covid-19” OR “SARS-CoV-2”)的任何結果。去掉引號後也列出了 281 條參考文獻，而其中只有少數相關的文獻 [譬如 (Ovsyannikova *et al.*, 2020)]。一項探索冠狀病毒決定多態性的基因組學研究表明，東亞人群的祖先已經習慣了與人類相伴數千年的冠狀病毒的感染，因此其後代可能沒有那麼容易受到當前流行病的影響 (Souilmi *et al.*, 2020)。

在到達目標細胞之前，SARS-CoV-2 必須附著並穿過一層層粘蛋白和其他分泌物。根據 Lewis 系統中不同方式的抗原分泌，可以把人類分為幾種類型 (Lemieux *et al.*, 1979; Nordgren and Svensson, 2019)。基於各種血型的感染特徵也對瞭解病毒有所幫助 (Bloch *et al.*, 2021; Le Pendu *et al.*, 2021; Schetelig *et al.*, 2021)。病毒還必須在體液和細胞水準上通過獲得性免疫的屏障，而當中涉及到的人類多態性非常多。如果基於在分隔開來的人群中收集到病毒基因組序列進行分析，可能會得到在該人群中關於疾病傳播的重要觀察結果，使我們能夠預測疫情的走向。就環境因素影響的獲得性免疫而言，過去感染案例可能已經有了新 SARS-CoV-2 病毒呈現的抗原決定簇的特徵 (Ng *et al.*, 2020)。定義人類多態性主要標誌物的人類白細胞抗原 (human leucocyte antigens, HLA) 在此範疇上可能起重要作用。這需要進一步深入研究，因為可能發生由 HLA 標記的特定病毒表位元與宿主因數一致的事件，從而引發嚴重的自身免疫反應的情況 [請參見與特定 HLA 和 H1N1 流感病毒有關的嗜睡症討論群組 (Schinkelshoek *et al.*, 2019)]。總的來說，獲得性免疫反應貢

獻了大量的表型，這些表型與龐大數目的人類多態性所對應。只有通過收集大量病毒基因組序列，才能對此瞭解透徹（除了與炎症發生過程相關的常見特徵），從而避免在預測中將獲得性免疫的代謝後果與人類多態性的特定特徵相關聯。

血管緊張素轉化酶 2（angiotensin-converting enzyme 2, ACE2）是病毒進入細胞過程中的受體 (Tortorici and Velesler, 2019)，這點將在下一部分中進行討論。ACE2 基因在 X 染色體上，有著豐富的多態性。這個值得注意的特徵暗示著性別二態性會影響感染。女性體內每個細胞中 X 染色體失活相關的隨機變異會導致鑲嵌多態性，這將可能導致一連串的感染以及不同的疾病嚴重程度。在調查病毒的傳播、嚴重程度和進化時，應該考慮到這種遺傳特徵 (Khayat *et al.*, 2020; Hamet *et al.*, 2021)。

到達細胞表面受體

病毒利用各種受體和宿主因數來結合靶標 (Baranowski *et al.*, 2001)。在冠狀病毒中，與同源受體相關的趨向性類型的主要決定因素是突刺蛋白 (Hulswit *et al.*, 2016)。類似於前身 SARS-CoV-1 但又與使用另一種受體的 MERS-CoV 不同，SARS-CoV-2 的突刺蛋白 S 通過結合其主要靶標受體 ACE2（一種腎素-血管緊張素系統中必需的羧肽酶）介導病毒附著並進入宿主細胞 (Gross *et al.*, 2020)。ACE2 在心臟、腎臟、睪丸和胃腸系統中都有表達。在肺部，它在某些 2 型肺泡細胞中低水準表達，而該表達似乎是因人而異的 (Hikmet *et al.*, 2020; Zou *et al.*, 2020)。基於目前的狀況，我們還注意到干擾素誘導了 ACE2 的表達 (Ziegler *et al.*, 2020)，揭示了感染過程中的前饋回路，這點或許解釋了冠狀病毒進化過程中使用 ACE2 作為受體的過程。

另一個有趣的特徵是 ACE2 與跨膜絲氨酸蛋白酶 2（transmembrane serine protease 2, TMPRSS2）在鼻杯狀分泌細胞、角膜、肺泡 2 型細胞、回腸吸收性腸上皮細胞、小腸上皮細胞和膽囊中共表達 (Lukassen *et al.*, 2020; Trypsteen *et al.*, 2020)。已經確定蛋白水解功能的參與對於病毒感染至關重要 (Laporte and Naesens, 2017)，而且冠狀病毒突刺蛋白（S）糖蛋白的蛋白水解裂解可啟動糖蛋白進入宿主細胞 [參見下文 (Hoffmann, Kleine-Weber, Schroeder, *et al.*, 2020)]。在進行預測時可能會發現高度相似的病毒引起的疾病之

間的差異。在 SARS-CoV-1 感染的情況下，ACE2 受體上的精氨酸和賴氨酸殘基的裂解增強了病毒的感染性。這些殘基對於通過 TMPRSS2 和人類氣道胰蛋白酶樣蛋白酶的裂解是不可避免的 [HAT, (Bertram *et al.*, 2011)]。而對於 SARS-CoV-1 而言 ACE2 裂解對於病毒 S 蛋白的啟動是必需的 (Heurich *et al.*, 2014)。可能還有另一種蛋白酶作為受體：二肽基肽酶 4 (dipeptidyl peptidase 4, DPP-4) 被認為是 SARS-CoV-2 的輔助受體，但尚未得到進一步證實 (Badawi and Ali, 2021)。該病毒還使用細胞蛋白酶來促發突刺蛋白糖蛋白進入細胞。在翻譯後進行修飾、蛋白激酶活性或各種類型的炎症細胞中它們的作用對於瞭解人類冠狀病毒的傳播至關重要 (Tharappel *et al.*, 2020)。無論如何，隨著流行病的發展必須仔細研究蛋白水解在感染初期的作用。它不僅參與了病毒與其宿主細胞的初始結合，而且還介導了膜融合，見下文進一步討論。

病毒通過膜融合進入細胞

SARS-CoV-2 的感染有兩個步驟：當 S 蛋白與 ACE2 受體結合後，病毒包膜與宿主細胞膜立即融合在一起。S 蛋白是同源三聚體，每個單體由兩個功能不同的亞基 S1 和 S2 組成。在冠狀病毒中，蛋白水解（稱為促發）將 S1 亞基與 S2 亞基分離，其中 S1 亞基包含受體結合域（RBD），促進與細胞受體的結合，從而觸發了膜融合 (Krueger *et al.*, 2001; Bosch *et al.*, 2003)。有效的融合需要 S2 進一步裂解成 S2' 片段 (Chambers *et al.*, 2020)。

令人感到驚訝的是，SARS-CoV-2 在 S 基因內部的 S1-S2 邊界有一個編碼多域基位元點基序 Arg-Arg-Ala-Arg (RRAR) 的插入片段，這是病毒來源仍舊未知而引起的主要問題。這使病毒成為高爾基體枯草桿菌蛋白酶樣 furin 蛋白酶的典型靶標 (Coutard *et al.*, 2020)。SARS-CoV-1 中並不存在此位點，取而代之的是只包含單個 Arg 殘基的片段，這可能與多種蛋白酶的參與有關，並直接影響病毒毒力和宿主選擇性。這個插入片段也可以增強細胞之間的融合而不影響病毒進入細胞 (Andersen *et al.*, 2020)。為了證實該觀點而進行的功能分析發現宿主跨膜絲氨酸蛋白酶 2 (TMPRSS2) 和 furin 蛋白酶起著主要作用，它們共同促進細胞表面的病毒膜融合 (Hoffmann, Kleine-Weber, and Pöhlmann, 2020; Papa *et al.*, 2021)。S2 亞基的裂解引發 S 蛋白對 TMPRSS2 的 S2' 位點進行加工，觸發與鄰

近細胞的膜融合，並形成合胞體 (Buchrieser *et al.*, 2020; Papa *et al.*, 2021)。這突顯了一個龐大的介導 SARS-CoV 病毒進入細胞的相互作用網路，為進化留下了很大空間。因此，監測這個區域中的突變並將其與臨床資料相關聯顯得特別重要。可以通過該區域的變異來進行預測是因為該區域的變異很可能增強或相反地減弱病毒對人類細胞的侵襲能力。在不同的情況下，其他蛋白酶（包括組織蛋白酶和 furin 蛋白酶家族的成員）可能會說明病毒通過內吞作用進入細胞。突刺蛋白的 S2 結構域會觸發膜融合過程，這會使細胞的溶酶體觸發內吞作用，從而使病毒進入細胞 (Ballout *et al.*, 2020)。和其他此處未談及的病毒進入細胞的途徑一樣，這過程會產生與溶酶體融合的胞內體，隨後觸發被包裹著的病毒基因組在細胞質中的釋放 (Zhao *et al.*, 2021)並進入複製週期（在下文複製部分中討論）。

有趣的是，“許多細胞融合成合胞體”是一種常見的傳染過程現象，例如在觀察傳染性支氣管炎病毒的時候也能看到 (Yamada and Liu, 2009)。科學家還提出了一種有趣的生化活動來解釋細胞中無膜隔室的形成 (Ditlev, 2021)。該過程由蛋白質中的所謂“無序”區域介導，該區域已證明可導致相位分離，這可能是無膜隔室的起源 (Shea *et al.*, 2021)。在冠狀病毒中，該過程可能涉及 N 蛋白，特別是在脫殼步驟中，產生一個病毒特異性腔室，該腔室隨著病毒的繁殖而逐漸變大，從而進一步利用了細胞機能及其代謝產物 (Dang *et al.*, 2021; Lu *et al.*, 2021)。除了涉及病毒基因組翻譯的核糖體之外，與病毒蛋白相互作用的多種宿主蛋白也參與了隔室化過程，但是相關的文獻記載仍然很少。

病毒特有的保護

自然選擇為病毒提供了避免被物理化學或生物過程輕易識別和滅活的特徵。冠狀病毒被包裹著，這為它們提供了額外的保護層，但與宿主的相互作用又需要與脂質代謝相匹配。SARS-CoV-2 成熟病毒體的基因組編碼著四種結構蛋白-突刺 (S)，包膜 (E)，膜 (M) 蛋白和基因組緊密核蛋白 (N)。我們已經討論了蛋白 S 和 N 的作用。其他兩種蛋白具有保護作用、調節作用和結構作用，用以塑造病毒體。此外，在形成病毒包膜時需要有應對滲透壓和/或電勢的新功能。除了明顯干擾宿主免疫反應以外，還解釋了蛋白 E 以及可能與膜結合的輔助蛋白的離子通道活性 (DeDiego *et al.*, 2014)。

蛋白質對蛋白水解非常敏感，並且會自然地老化 (Truscott *et al.*, 2016)。翻譯後的修飾有助於克服這些弱點。涉及細胞入侵的病毒蛋白被廣泛糖基化有幾個原因：協助蛋白質折疊、提高穩定性，最重要的是從宿主的免疫識別中遮罩病毒 [這通常被描述為一個“聚糖屏障” (Walls *et al.*, 2016)]。糖基化水平是各種各樣的，而在流感感染中非常重要的唾液酸化尚未針對 SARS-CoV-2 進行仔細研究 (Östbye *et al.*, 2020)。每個蛋白 S 單體有 22 個潛在的糖基化位點 (D. Wang *et al.*, 2020)。這些位點中大多數已被發現 *N*-糖基化，而假定的 *O*-糖基化位點的佔有率則較低。因此，蛋白質中的相應氨基酸殘基具有重要的遮罩作用。非同義突變可能會影響突變菌株的後代，所以應優先監測涉及天冬醯胺酸的變化。

環境中病毒的成熟和釋放

在探索的第二階段，病毒從受感染的細胞、受感染的器官和有機體中釋放出來，這一階段的特徵尚未被研究透徹。它可能在某些病毒變種的可傳播性中起關鍵作用 (Lemey *et al.*, 2021)。宿主細胞的死亡釋放了病毒體，但也可能套牢了它們。因此主動釋放是病毒傳播後代的更有效方法。長期以來這點已被病毒成熟後出芽所證實 (Garoff *et al.*, 1998)。冠狀病毒出芽所涉及的宿主因數仍然有許多未知數，但已瞭解了胞質內的整體步驟順序，既涉及內質網又涉及高爾基體 (Boson *et al.*, 2020)。與其他病毒一樣 (Ortego *et al.*, 2007)，SARS-CoV-2 的 E 蛋白對於病毒成熟和釋放途徑也很重要。一旦複製完成，基因組就通過核蛋白 N 和蛋白 M 與內質網的脂質融合，驅動了病毒包裝所必需的相分離過程 (Lu *et al.*, 2021)。組裝完核衣殼 N/基因組複合物後，包膜將與蛋白 E、M 和 S 放在一起。病毒隨後被運輸至細胞表面，並通過非常規地劫持溶酶體的胞外功能以釋放到環境中 (Ghosh *et al.*, 2020)。總的來說，我們再次強調蛋白酶作為病毒迴圈的重要特徵是無處不在的，而我們必須仔細監控蛋白酶以預測未來的病毒進化。

最後，病毒蛋白與宿主蛋白的相互作用形成了各種穩定性的複合物。可以預料在成熟過程中，病毒包膜將包含病毒蛋白和在病毒體中的宿主蛋白。其中一些可能來自非特異性捕獲，而其他則是由於它們與病毒蛋白的特異性相互作用，而更加系統地被捕獲。雖然這步驟可能在準備下一輪感染的病毒中起到作用，但它通常被忽略。因為這些蛋白質，特別

是病毒蛋白質，在感染後會立即注入受體宿主細胞中，所以這一點特別重要。這種不可避免的“污染”肯定是由進化形成的，是病毒操縱其宿主並加快啟動病毒繁殖的一種方式。有著蛋白質結構的 CoV3D 資料庫 (Gowthaman *et al.*, 2021) 是說明我們通過這些病毒蛋白質介導的相互作用來預測潛在網路的重要資源。

進入細胞後，病毒基因組立即被翻譯成兩個大的多肽，它們被病毒蛋白酶迅速分解成有活性的非結構蛋白。將它們包含在病毒體中顯然代表著一種選擇性優勢 (Haas *et al.*, 2021)。例如，SARS-CoV-2 木瓜蛋白酶樣半胱氨酸蛋白酶（非結構蛋白 Nsp3 的結構域）對於病毒成熟、干擾宿主炎症和抗病毒免疫反應至關重要。已經通過使用質譜和蛋白激酶分析法研究了存在於 SARS-CoV-1 病毒體中的蛋白質補體 (Neuman *et al.*, 2008)。實驗表明，除了蛋白質補體有著操縱宿主蛋白激酶的調節潛能以外 (Siddell *et al.*, 1981)，純化的病毒體中確實存在兩種病毒蛋白酶 Nsp3 和 Nsp5。找到多個大型跨膜蛋白 Nsp3 特別出乎意料，這促使我們去瞭解它是怎麼整合到病毒體中的。另外，在成熟的病毒體中還已經發現了其他幾種蛋白質，例如 Orf3a、Orf9b 和 Nsp2，目前對它們的功能瞭解仍然有限。病毒體成分的完整特徵對於我們瞭解感染早期的動力學至關重要。

增殖

病毒結合到宿主細胞及通過細胞內吞作用進入細胞的過程中涉及使用 S 蛋白和其他剛剛提及的病毒蛋白，這情況仍未能完全被瞭解清楚。病毒進入細胞後必須擁有允許其自身複製的機制。由於 SARS-CoV-2 的基因組被病毒包膜包裹，它必須首先解開其包膜並與幾種蛋白結合（特別是衣殼蛋白 N），這樣才能通過翻譯產生所需要的酶供其基因組複製及形成新的包膜使用。

翻譯機制

在進入特定類型的細胞後，病毒的 RNA 基因組需到達內質網中正確的位置並立即開始翻譯。組成這活躍的起始復合物需要通過把長的 5'UTR 接合到核糖體 (Tidu *et al.*,

2020)。為使其進入細胞後能立即進行翻譯，病毒基因組會形成 5'端以模仿標準的人類細胞的 mRNA (Yan *et al.*, 2021)。我們已經知道隨後的翻譯過程 (Hartenian *et al.*, 2020) 以及 SARS-CoV-2 病毒所產生的功能蛋白的數量 (Kim *et al.*, 2020)。但是，翻譯過程中的步驟包括了許多未知功能，其中的一些功能可能對病毒未來的進化過程有著關鍵的影響 (Neches *et al.*, 2021)。除了核糖體以外，亦甚少有文獻記載翻譯過程中宿主蛋白與病毒蛋白的相互作用。

SARS-CoV-2 十分強大遺傳的編碼指導著 26 種蛋白質的合成 (Finkel *et al.*, 2021)。其中較為特別的是病毒 RNA 非對稱的翻譯過程：在脫離包膜後，前三分之二的基因組即從一個大的編碼區中翻譯成 Orf1a 和 Orf1ab 兩個多肽，兩者雖從同一 RNA 序列翻譯出來但產生的數量並不平均。而 Orf1ab 的 RNA 是利用了假結結構和 -1 移碼翻譯了其 C 端區域。由於冠狀病毒的翻譯無處不在，導致 Nsp 蛋白的非對稱翻譯而創造了一個特定的環境去精準地控制 Orf1a 和 Orf1ab 蛋白的比例或延遲 Orf1ab 的生產（其中包括 RNA 依賴性 RNA 複製酶，RdRp，Nsp12），這會一直發生直至 Orf1a 多肽（Nsp1-11）創造出合適 RNA 複製的環境為止 (Fehr and Perlman, 2015)。這些多肽隨後被分成 16 種非結構蛋白 Nsp1-16（Orf1a 末端的 Nsp11 可能沒有真實的功能），而其中一些則如我們所見存在於離開宿主後的病毒體中。轉錄組的遠端部分用以下所述的機制被翻譯成單獨的蛋白質，包括病毒體主要蛋白質 E，M，N 和 S。

檢測大量的病毒基因組序列有助鑑定在其編碼中所有導致蛋白變異的突變。圖 1 顯示了以計算最高突變率和最低突變率之差後所得的每種病毒蛋白裡所有突變的最高等位基因頻率 (AF) 的分佈。而補充表 1 提供了每種蛋白的突變數目。除了 S 蛋白預期具有高度可變性以外，RdRP (Nsp12) 也同樣具有很大的可變性的事實告訴我們尋找針對病毒複製的抗病毒藥物可能是一項艱鉅的任務。這點同樣適用於蛋白酶 Nsp3，儘管它有著多結構域可以提供了多個靶標，看上去作為抗病毒藥物目標完全不是問題。雖然 Nsp2 的功能暫時仍然知之甚少，但它那些高度變化的範圍讓它特別受關注。而 Orf3 的變異則特別令人感興趣，因為該蛋白可能通過維持一個對滲透壓的正確反應來控制病毒包膜的穩定性。

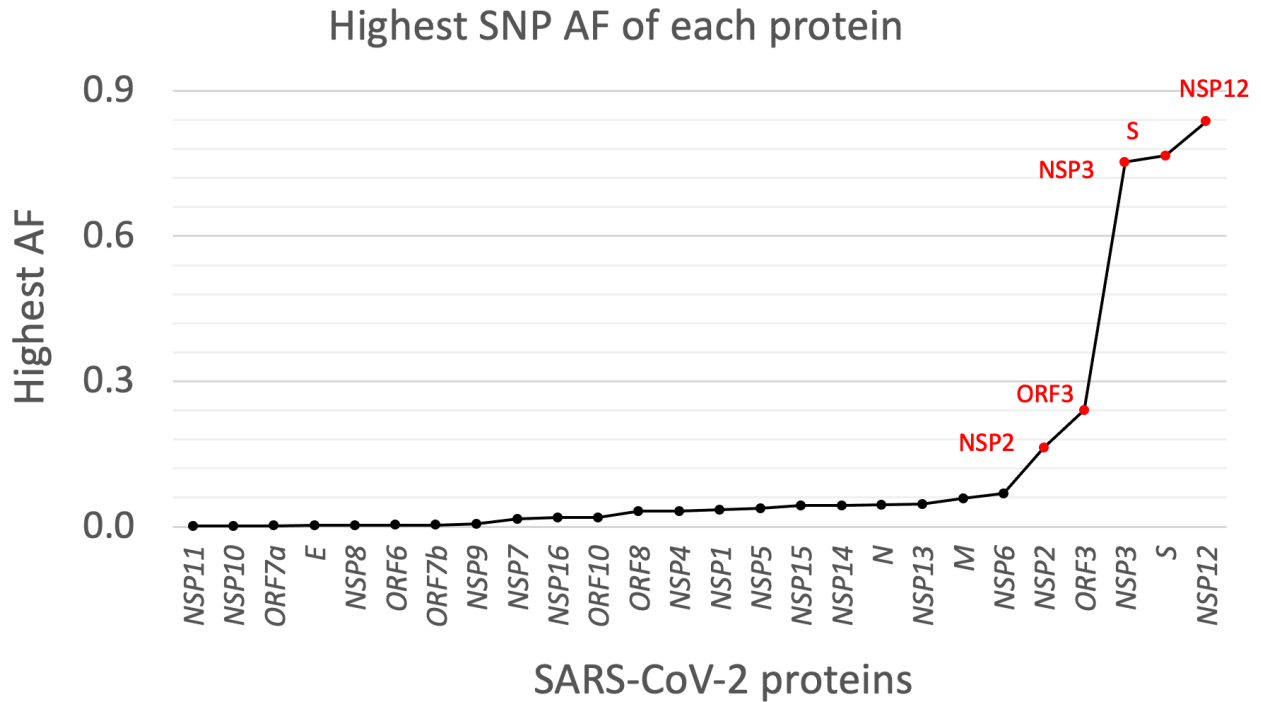


圖 1. 由統計 193,687 個 SARS-CoV-2 毒株得到的每種蛋白質裡所有突變（包括 SNP 和缺失突變）中的最高等位元基因頻率（AF）的分佈。AF 大於 10% 的蛋白質由紅色標籤標記。y 軸顯示在該蛋白質中觀察到的最高的突變率（即最高的等位元基因頻率）。

控制翻譯的過程是病毒發展中十分關鍵的一環，這與是否有著精確的 tRNA 補體直接相關。值得注意的是單靠人類基因組並不能編碼完全的 tRNA 分子，前體 tRNA 必須先在細胞核中成熟，然後在其 3' 端添加 CCA 補體 (Augustin *et al.*, 2003)。為了讓 tRNA 適應目標基因的密碼子使用偏好還需要控制許多 tRNA 核苷酸的修飾，這導致了病毒進化的瓶頸並可能通過病毒編碼的蛋白反映出來。雖然相關功能尚未被研究，但是監測那些導致病毒蛋白使用密碼子的偏好發生重大變化的突變將有助我們預測該流行病的發展。

蛋白水解

正如上文中所強調的，利用多蛋白去進行蛋白水解是許多病毒的重要功能 (Yost and Marcotrigiano, 2013)。Nsp3 的木瓜蛋白酶樣蛋白酶結構域在合成後會立即識別 LXGG 氨基酸序列和切割 Orf1a 及 Orf1ab (Barretto *et al.*, 2005; Gao *et al.*, 2021)。這蛋白的 3D 結

構是已知的 (Gao *et al.*, 2021)。特別的是該切割位點序列也可以被細胞去泛素化酶所識別，使得 Nsp3 能夠幹擾其宿主的調節功能。該蛋白酶還會選擇性地切割幹擾素調節因數 3 (IRF3)。在 SARS-CoV-2 感染期間這種切割可能導致了小程度的 I 型 IFN 應答 (Moustaqil *et al.*, 2021)。由於多種重要功能都使用著一樣的識別位點，這種蛋白酶進化的可能性極受限制。另一種病毒編碼蛋白酶 Nsp5 通過識別[AVPT][VTKRM][LF]Q[ASN]序列，在 Nsp3 末端切割 Nsp4，並在 Nsp4 末端切割後釋放自身。隨後 Nsp5 切除了從 Nsp6 至 Nsp16 的所有 Nsps。與 Nsp3 的情況一樣，Nsp5 調控宿主蛋白的分解，特別是有效的炎症緩解劑 NLRP12 (Normand *et al.*, 2018)，以及炎症反應相關的轉化生長因數 β 啟動激酶 1 結合蛋白 (TGFbeta activated kinase 1 binding protein 1, TAB1) (Xu and Lei, 2020)。這指出了增強細胞因數生產和炎症反應的分子機制 (Moustaqil *et al.*, 2021)。在這裡，再一次顯示了病毒進化影響疾病嚴重性的可能性仍然有限，尤其是在“long covid”的發展情況下。反之，這也突顯了這些蛋白酶作為良好的藥物靶標的可能性。

轉錄機制

通過使用從 Orf1a 和 Orf1ab 所翻譯出來的 Nsp 蛋白，特別是 Nsp12 和 Nsp13 (Arya *et al.*, 2021)，病毒基因組末端的三分之一隨後被表達為個別蛋白的轉錄本，而這些蛋白在促進病毒復制的過程中會減少宿主防禦系統的活性。病毒的 3'端序列被轉錄成的多個會被翻譯成個別蛋白的轉錄本，上文中已討論過當中的結構蛋白 E、M、N 和 S。N 蛋白的低聚狀態對於復制和轉錄都是必需的 (Ahamad *et al.*, 2020)。而轉錄本作為病毒的完整基因組會被加帽。mRNA 加帽複合物包括 Nsp14、Nsp16 和 Nsp10。Nsp16 與 Nsp10 關聯並於 2'位置甲基化形成 cap-0 型帽子結構，在病毒的 5'端形成 cap-1 型的功能結構，通過這些過程它會被識別為真正的宿主功能性 RNA，因此並不會受到抗病毒反應的攻擊 (Perveen *et al.*, 2021)。正如在其他病毒蛋白所觀察到的，這些蛋白可能有著其他功能，進一步限制了它們的進化能力。Nsp9 是另一種 RNA 結合蛋白，但其功能仍未被完全瞭解（請參見下文“質量控制”的部分）。除了加帽外，轉錄起始也需要將其 RNA 折疊成具病毒特性的 3D 結構 (Madhugiri *et al.*, 2016)，當發生變異時，這結構不會形成有效的病毒基因組（下文

中提及的 C241U 突變是一個例外)。重要的是，形成病毒形態涉及幾個必須的轉錄步驟，這揭示了轉錄控制區中的序列進化受著很大的限制。於病毒複製或轉錄起始序列缺乏對突變的耐受性相反，蛋白質編碼序列在積累突變後仍然允許合成有效的病毒基因組。但是這點取決於蛋白質，有些蛋白傾向接受突變，而另一些則相反 (Nagy *et al.*, 2021)。

複製機制

複製病毒基因組是病毒增殖的核心。對冠狀病毒而言，它們是正單鏈 RNA 基因組病毒，它的複製過程應該是十分不對稱的，而病毒基因組的互補拷貝的數量要比基因組本身少很多。這個不對稱的病毒複製模式尚未研究透徹。這可能受到許多與 RNA 種類有關的 RNA 複製機制的因素影響，而這些複製機製與內質網有關。如上文所討論有關病毒翻譯的部分，Orf1a 和 Orf1ab 的蛋白質的表達是高度不對稱的，這創造了一個合適 RNA 複製的環境 (Fehr and Perlman, 2015)。SARS-CoV-2 的複製程式的啟動需要使用蛋白 Nsp7 和 Nsp8 才能從病毒的 3'端結構起動。隨後，再次在 Nsp7 和 Nsp8 的幫助下 RdRP 的主要亞基 Nsp12 會繼續複製基因組(Gao *et al.*, 2020)，而 Nsp13 則幫助 RNA 解鏈。由於與 RNA 種類有關的 RNA 聚合酶因為決定了複製過程的效率和準確性，所以有著獨特的重要性，亦因此已被選作抗病毒分子的靶標。然而，如圖 1 所見，Nsp12 中仍保留了許多突變，這表明了病毒在進化過程中得到了許多解決方案來躲避藥物的抑製作用，說明 Nsp12 可能並不是抗病毒藥物的好靶標。另一個亞基 Nsp15 是在冠狀病毒中十分保守的 U 特異性核酸內切酶 (Pillon *et al.*, 2021)。它的功能尚未完全被瞭解，但似乎與躲避宿主的抗病毒反應有關 (Zhao *et al.*, 2020)。因此，它的進化用於預測病毒毒力的未來發展趨勢十分有用。

質量控制

複製和轉錄過程缺乏足夠的準確率來限制長達 30,000 個核苷酸的 SARS-CoV-2 基因組中出現錯誤 (Bradwell *et al.*, 2013)。在這種選擇壓力下，病毒衍生出來一些防止在每個

複製週期中積累太多錯誤的校對複合體。這點十分重要，因為錯誤率過高會導致形成快速滅絕的缺陷病毒 (Pauly and Lauring, 2015)。而降低 RNA 病毒複製的準確率則是驅動某些核苷酸類似物（例如 favipiravir）的藥物的設計理念，這種藥物可導致突變，以至突變後的病毒後代不能存活 (Baranovich *et al.*, 2013)。Nsp14 既是產生 cap-0 型結構的鳥嘌呤-N7-甲基轉移酶，又是可用作消除基因組複製過程中出現的錯配的校對 3'至 5'核酸外切酶 (Ogando *et al.*, 2020)。該蛋白的許多突變會增加病毒基因組的突變負荷 (Eskier *et al.*, 2020)。研究這種雙功能酶中突變的結果已經揭示了新型病毒譜系“綻放”般形成，其中一些可能會導致病毒毒力減弱 (Cluzel *et al.*, 2020)。

病毒的蛋白水解活性的質量控制可能在其長期存活中有著重要的作用。值得注意的是 Nsp9 二聚體在體外實驗中會與肽 LEVL 結合，這與 Nsp5 蛋白酶切割位點相似 (Littler *et al.*, 2020)。可是，該蛋白的主要功能也可能涉及調節分子伴侶的活性。尤其是在壓力下（例如在病毒感染期間），這些因素是關鍵的質量控制素。由此推論，SARS-CoV-2 可能會通過翻譯後的修飾來調節這些功能。與此一致的是，谷氨酸殘基的核苷酸化修飾了 Nsp9（在體外實驗中略微偏好 UTP），這通過一個和 Nsp12（上文討論過的取決於 RNA 種類的 RNA 聚合酶）的 NiRAN 域相關的錳依賴性活動進行。該殘基屬於保守的 N 端 NNE 三肽。而它是在冠狀病毒 Nsp9 的同源中唯一不變的殘基 (Slanina *et al.*, 2021)。有趣的是，對應 Nsp9 的鼠傷寒沙門氏菌的 SelO (YdiU) 蛋白質是在生命樹的三個域中都十分保守的。該蛋白是線粒體中的一種硒蛋白，證實了它在控制氧化應激中的作用。它在 ATP 有限的情況下通過尿苷酸化去修飾分子伴侶 (Y. Yang *et al.*, 2020)。這種核苷酸化作用應該從 C>U 趨勢的角度被優先探討，以下一節將討論其起源。

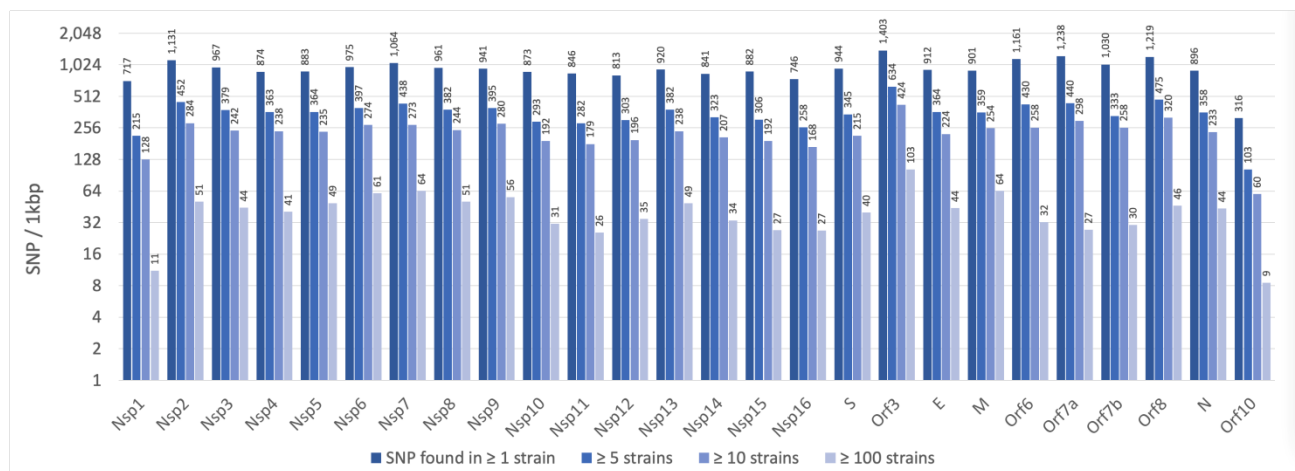
預測未來

誰都知道預測未來是困難的 (Sullivan, Gregory F, 2013)，尤其是考慮到人的行為的不規律性和地域性。但在許多情況下，我們必須做出我們最好的估計。好的估計有積極的作用，且即使是外行人也可以做出有趣的估計。舉個例子，Andrew Wright 在 2020 年初 COVID-19 大流行之初，就在其小說 *The End of October* 裡頗具幾分真實地預測了大流行的發展。管理一場大流行並使其盡可能短而少害可以挽救數以百萬計的生命和減輕其所導致

的巨大的經濟負擔。在總結了大量的有關病毒的知識與文獻，並篩除了好些錯誤資訊後，我們在此著重討論對於 SARS-CoV-2 的進化而言，它通過自然選擇的方式所探索過的或多或少受到限制的進化空間。

在第一年大流行的普遍進化模式

儘管大量論文和在線出版物中都有討論，但公眾對 SARS-CoV-2 的反應卻存在極大的偏差和局限性，更不用說那些本應該被緊急考慮的方面。公眾討論保持集中在病毒那些最明顯特徵上，例如包膜上的突刺蛋白（下文將進一步討論它的部份作用），以及在缺乏資訊的體外細胞模型中的假性治療方法。這一點可以被簡單地使用 Google 搜索“SARS-CoV-2” “spike”反映出來，在 3 月 3 日搜索的結果包含 6,000,000 頁（2 月 15 日的結果為 9,000,000 頁，這見證著熱潮的力量，我們由此看到一個下降的趨勢）。而搜索“SARS-CoV-2” “Orf8”的結果僅包含了 50,400 頁。然而即便不是全部，但絕大多數病毒蛋白質仍在持續進化 (Jaroszewski *et al.*,2020)。如圖 2 和補充表 1 所展示，病毒基因中的突變數目要比在它的突刺蛋白中的多很多，而且蛋白之間突變的數目也存在相當大的差異。如本文第一部分所述，穩定這些變化的選擇壓力是由多種原因引起的，它們中的一部份是可以通過人為幹預的。那些影響多個病毒分離株的那些突變（直方圖中的最後一列）可能是病毒傳播中最重要的突變。



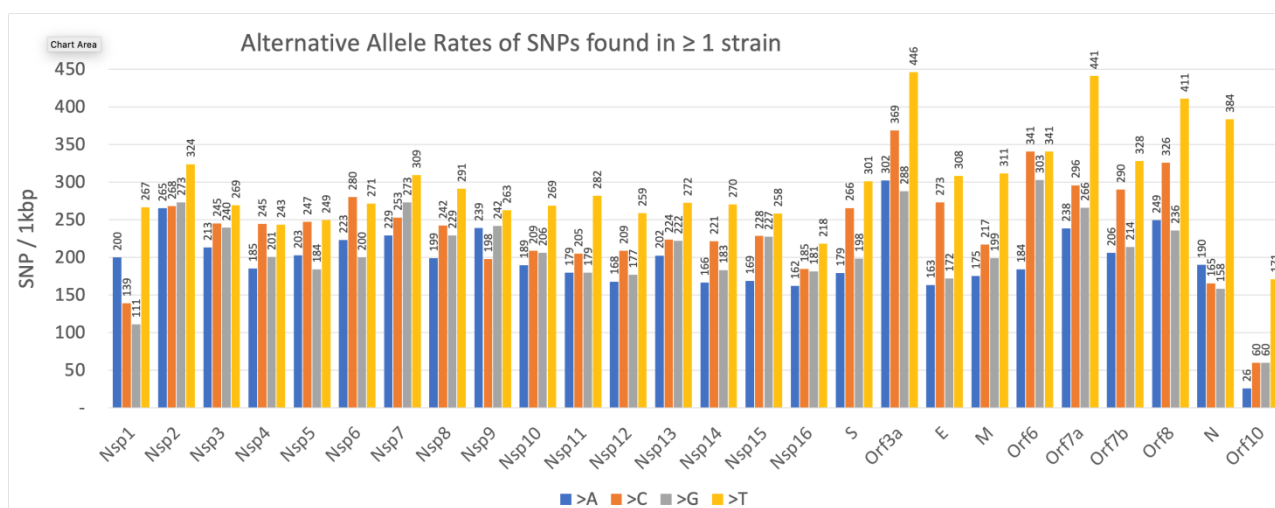
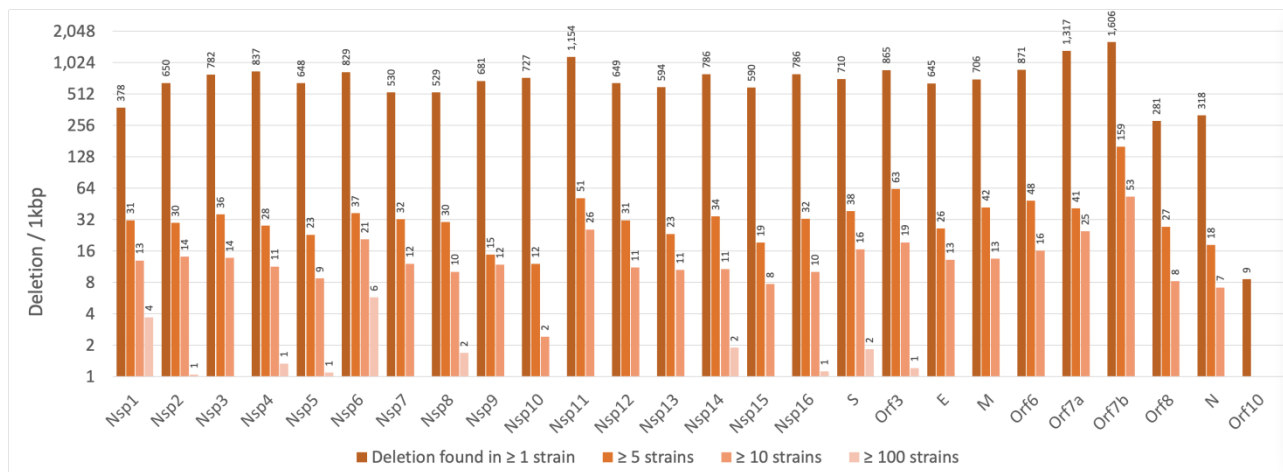


圖 2. 總結了來自 193,687 個 SARS-CoV-2 毒株的每個基因的突變率。(a) SNP 率，(b) 缺失率，(c) 在 ≥1 個毒株中發現的 SNP 的突變域基頻率。‘>A’ 表示 SNP 從非 ‘A’ 突變到 ‘A’。y 軸已經根據每個基因的長度標準化為每 1kbp 的突變數。在 (a) 和 (b) 中，y 軸使用了以 2 為底的對數刻度，並且暫時了在 ≥1 株，≥5 株，≥10 株和 ≥100 株中的突變率。補充注解中記錄了這些每個基因突變率是如何分析得到的。

以下是一個具體的例子去幫助我們預測病毒進化的未來變化。Nsp1 蛋白是從多蛋白 Orf1a 和 Orf1ab 中第一個分離出來的病毒蛋白。出乎意料的是，該蛋白質中的突變數量相對其他相同大小的 Nsp 蛋白都少得多，這表明了長期進化過程中有著很大的選擇壓力。Nsp1 對病毒基因組翻譯的開始有著關鍵的作用。其翻譯特徵有別於該病毒的其他蛋白質的特徵，涉及從大量的細胞 mRNA 中區分出病毒 mRNA 基因組 (Ou *et al.*, 2020)。因為它為了避免變異承受著相當程度的壓力，因此應優先分析保留在其序列中的少數突變以

及報告它們對毒力的間接影響。最後，正如前文所述，穩定這些變化的選擇壓力是由多種因數所引起的，其中一些可能可通過人為幹預。收集相關數據支援基因組資訊的充分利用的需求十分迫切。

環境因素

在上文中已有強調，許多混雜因素會影響我們對呼吸系統疾病的理解。然而，我們已經充分瞭解到擁擠的環境，包括但又不侷限於室內，對人與人之間的傳播有著顯著的貢獻。其中通氣系統的結構對此可能會產生很大的影響，這說明瞭對建築物的通氣方式立刻進行深入研究的必要性，這包括使用空氣過濾系統以及即便在戶外環境中也要加強社交距離。為了控制將來呼吸道疾病的流行，在居住區、餐廳和辦公樓以及公共交通的設施中應服從控制空氣流動的強制性規定，就好比在手術室中一樣。此外，氣候也可能對病毒的進化有著重要的影響。B. Chen *et al.*, 2020 等研究並沒有提供太多的資料，因為有相當數目的混雜因素影響著流行病的發展。氣候會影響人類的行為：比如我們傾向於在天氣寒冷時留在室內。儘管大多數情況下僅僅是推論，但室內與室外感染傳播率是最佳的記錄參數。傳播的參數與外界溫度有關，當溫度或濕度較高時，空調可能扮演著一個負面的角色。因此，封閉空間的通風系統 (Pease *et al.*, 2021) 可能會嚴重地影響著病毒的進化。這表明了特別監視其中幾個功能的必要性。為了幫助進行預測，應對元數據作出全面調查以識別在封閉空間中的感染個案，這個調查應該與基於基因組測序的調查聯繫起來。基於基因組測序的調查包括病毒包膜及其相關蛋白的結構、它們與脂質的結合、小的疏水性蛋白的進化，以及與離子回收利用和運輸的關係。另一個或許應該探索的參數是紫外線輻射 (Karapiperis *et al.*, 2020)。在可以殺死病原體的同時，紫外線也可以以一種特定的方式導致誘變 (Wurtmann and Wolin, 2009)。那些更為標準的誘變核苷酸的修飾會生成尿嘧啶環丁烷二聚體。如果效果顯著的話，這將令病毒的進化高度偏倚，這點值得研究，或許可以通過分析從較高 UV 的區域分離出的變異毒株的突變模式的偏差。然而，我們在文中只論及在病毒中一些鮮明的特徵，這些特徵可能會對預測病毒未來的進化做出最大的貢獻而且適合通過實驗方法研究。

普遍來說，正如大量有關該流行病的文獻所指出，重點主要放在宿主對抗病毒的調控上。以這觀點有助讓我們把人類習性與感染聯繫起來。例如，由於吸煙者可能患有更多的呼吸道疾病，有機會因而與先前感染的病原體發生交叉反應，因此可能會產生一定程度的保護 (Saurabh *et al.*, 2021)，這就能解釋了初期有關對吸煙可能產生保護作用這種矛盾的觀察 (Landoni *et al.*, 2020)，並已被定義為嚴重性風險因素之一 (Elliott *et al.*, 2021)。多種病毒同時感染也可能是病毒之間重組的源頭，而下文討論的特徵可能會產生重要的結果。這個特定的行為在預測病毒在部分人口中進化時必須被考慮進去 – 應該再次利用整個基因組的序列來進行監測 – 通過聯繫先前的呼吸道和腸道感染紀錄，而且應該再次地在利用整個基因組的序列來進行監測。

在一般情況下，患者臨床數據可能會揭示病毒趨向性的進化，特別是與呼吸系統和消化功能相關的數據，這些數據作為與基因組聯繫起來的元數據而被紀錄下來是很關鍵的。

取決於宿主的病毒發展的局限性

把病毒和它宿主的代謝系統聯繫起來對於瞭解它的短期和長期進化過程十分重要。病毒在進行增殖時必須將其宿主的代謝系統轉向幫助自身的複製，換句話說病毒正操縱著宿主細胞的新陳代謝。因此病毒必須適應宿主的代謝系統的約束。在這方面，SARS-CoV-2 是一個非常有趣的生命體。基於它一些被忽略的生長特性，它增殖的方法體現了令人著迷的細胞代謝的普遍特性。細胞的生長過程中的大部代謝反應都在細胞質中發生，而細胞生長所需的所有基本元素都會在細胞質中產生。這裏引起了一個常被忽視的“非同向”增長的問題，類似的概念有常被套用在經濟學上。細胞是在三維中生長的，而細胞膜則是在二維的表面上生長，更重要的是基因組是個一維增長的線性聚合物。這意味著，製造“過多”的膜會帶來巨大的代謝壓力，而製造“過多”的基因組會帶來更大的壓力。這些功能原則上對非包膜病毒及包膜病毒的增殖都有著莫大的益處。這些物理難以解釋的現象形成了細胞演化中的和諧方式需要加以闡明。

實際上，通過利用整個代謝過程去合成三磷酸胞苷單分子[這對應著在基因組序列中的“C”]，上述一障礙已通過物競天擇的過程解決了 (Danchin and Marlière, 2020; Ou *et al.*,

2020)。嘧啶代謝中一個出乎意料的特徵證實了這一觀察的結果，這是通過一個不符合預期的酶活性的普遍缺失而證實的。雖然磷酸核糖基轉移酶在細胞中無處不在，並可將嘌呤、嘧啶和其他雜環鹼基拆成單核苷酸，但至今尚未在任何活生物體中找到胞嘧啶磷酸核糖基轉移酶 (Ou *et al.*, 2020)。因為與它相關的酶無處不在，這使胞嘧啶磷酸核糖基轉移酶的缺失顯得十分重要。由於隨機突變出現在現有同類酶中，導致該酶的缺失會持續出現，所以它的缺失一定與回收胞嘧啶的強烈的自然選擇背道而馳。SARS-CoV-2 基因組中胞嘧啶有著下降趨勢的結果是鳥嘌呤與之相似的缺失趨勢，因為在複製過程中 G 和 C 是互補的，導致了病毒中 G 含量的相似下降。如下文所論述的，由於鳥嘌呤是涉及活性氧 (ROS) 的特定防禦反應的靶標，這將衍生出有趣的結果。

因此，宿主在進化過程中保留了一些代謝功能，通過操縱 CTP 合成來幹擾病毒的發展，這種現象也是不足為奇的。通過 viperin (幹擾素可誘發的內質網病毒抑制蛋白) 合成類似 CTP 的抗病毒物 3'-deoxy-3',4'-didehydro-CTP (ddhCTP) 的過程就如化學家人工合成的抗病毒核苷酸一樣引人入勝，類似的方式普遍存在於哺乳動物的先天抗病毒機制中 (Kang *et al.*, 2020)，但在牡蠣 (Green *et al.*, 2015) 甚至在細菌和古細菌中也能找到 (Bernheim *et al.*, 2020)。該抑制劑同時阻礙了 CTP 的四個關鍵功能：RNA 合成、在 tRNA 的 3'端添加 CCA、製造胞嘧啶核苷酸依賴性的膜脂以及製造作為蛋白質糖基化底物通用載體的合成磷酸鹽 (Ou *et al.*, 2020)。這種代謝瓶頸的後果是 SARS-CoV-2 在低胞嘧啶含量的情況下進化 (見下文)，除非宿主中有專門的代謝設置以減輕其限制。

因此，無論是在 RNA 還是在氨基酸水準上能夠保持不變的胞嘧啶殘基對於病毒感染的永續性可能是十分重要的。這代表著它是 COVID-19 抗病毒藥的重要靶標。因此，監測病毒的不同譜系在其短期的進化過程中如何應對這一局限性至為重要。這也可能包括胞嘧啶含量下降趨勢的逆轉。該逆轉意味著改變 CTP 合成酶的控制和抑制 viperin。這樣亦拓寬了病毒發展的可能性。這一點再次說明瞭盡量收集更多完整的病毒基因組序列的重要性。

當前對各種物種中的病毒的比較系統發生研究的大量增加，有助緩解對深入瞭解免疫和代謝系統之關係的迫切需要。在這方面，與不同代謝系統，尤其是對蝙蝠的代謝系統作比較可能十分有用，並有助於我們預測一些病毒進化過程中的限制。有趣的是這些動物會

以非毒性的形式與病毒共存，以進化出飛行的蝙蝠為例，被選擇出來的系統可能有著快速新陳代謝的特徵 (Shen *et al.*, 2010)。

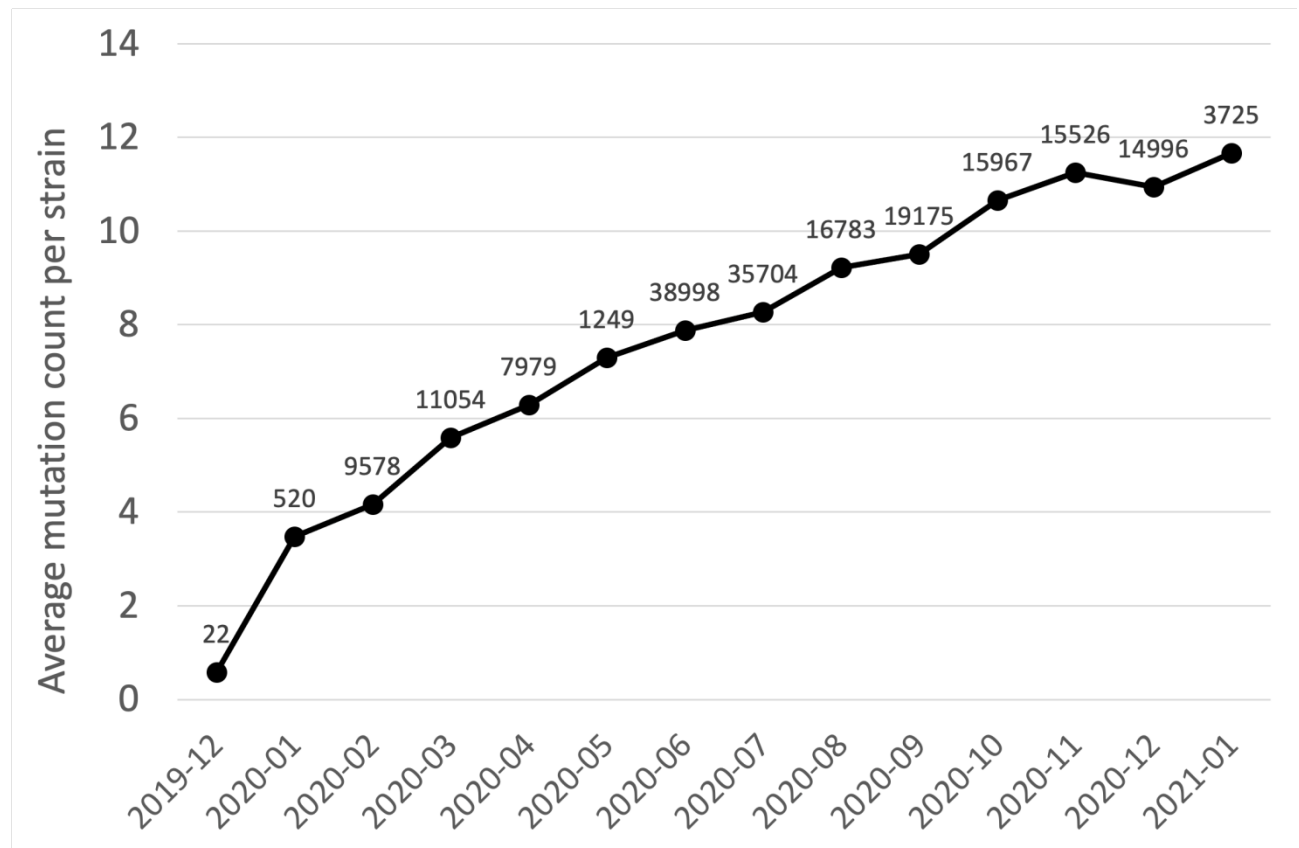
飛行是一種能量消耗極高的活動，其間會產生大量的 ROS。然而有趣的是，ROS 通過巨噬細胞和嗜中性粒細胞利用呼吸/氧化爆發來提供對病原體防禦的積極作用 (Piacenza *et al.*, 2019)。儘管很少研究，但快速代謝的正面效果可能有助於在蝙蝠中病毒感染的無害化。然而，因感染引起的抗病毒反應可能導致恢復氧化還原的穩態失衡，引致 ROS 及促氧化劑細胞因數的釋放不受控制，以及因過度炎症而引起失控病理。調節該過程與肺部感染中的多個事件特別相關，這是嚴重 COVID-19（當然還有“long covid”）的普遍特徵。同時，過量的 ROS 也會對病毒基因組產生相當大的影響，其中過量的 ROS 會導致 8-氧鳥嘌呤的形成，這會在複製過程中觸發 G>U 鹼基顛換突變。因此，冠狀病毒可能已發展出防禦機制以減輕這種化學攻擊造成的負擔。由此可見，監測病毒中特定譜系中一致的鹼基顛換突變十分重要 (Cluzel *et al.*, 2020)。

病毒的短期進化：提高傳播性

病毒在新感染不同物種的短期內必須適應其增殖的速率以及感染新宿主的能力。在這期間有利於所有相關的基因的突變積累。流行病最可能的早期發展是該病毒開始提高其傳播速度。這種發展方向源於兩個基本的功能。最有可能的是提高其傳播性，這跟許多社會參數都有著密切的關係，包括擁擠的環境，或從受感染者所產生穩定而長期的感染源，及潛伏期的增加等。第二個關鍵特徵是病毒的複製成功率增加，這高速的增殖率通常是與增加了面對突變負擔的能力有關。

在 SARS-CoV-2 入侵人類的第一階段，趨勢是平均每年每個基因組約 22 個突變。與上述情況一樣，除大多數蝙蝠病毒的基因組中有著尚未解釋到的較高 C 核苷酸含量外，其他觀察到的病毒基因組的 C 核苷酸含量普遍有下降 (Matyášek and Kovařík, 2020; Ou *et al.*, 2020; Simmonds, 2020)。這個網站 (<http://www.bio8.cs.hku.hk/sarscov2>) 一直有在跟蹤新菌株中 C 的百分比 (Luo *et al.*, 2020)。允許不斷出現突變也是因為其蛋白質序列的功能多樣性和彈性。在大流行的頭六個月期間，有研究監測該蛋白質組的進化 (Lubin *et al.*,

2020)。大量突變導致病毒蛋白中的氨基酸發生變化，這表示當中含有高度的正向選擇 (Cluzel *et al.*, 2020)。儘管病毒能適應新的宿主細胞並不意味著它將導致嚴重疾病，但有些變化會影響 COVID-19 嚴重性。我們應該特別注意的是，去除 C 殘基會使 ddhCTP 幹擾效率降低，因此這可能會增加病毒的毒力。圖 3 展示了該病毒的所有蛋白隨著時間的變化。在大流行的頭 8 個月（從 2019 年 12 月至 2020 年 7 月），每個毒株的平均突變率以每月約 1 個突變的速率增加，但隨後便以相對較低的每月約 0.5 個突變速率增加（從 2020 年 8 月至 2021 年 1 月）。



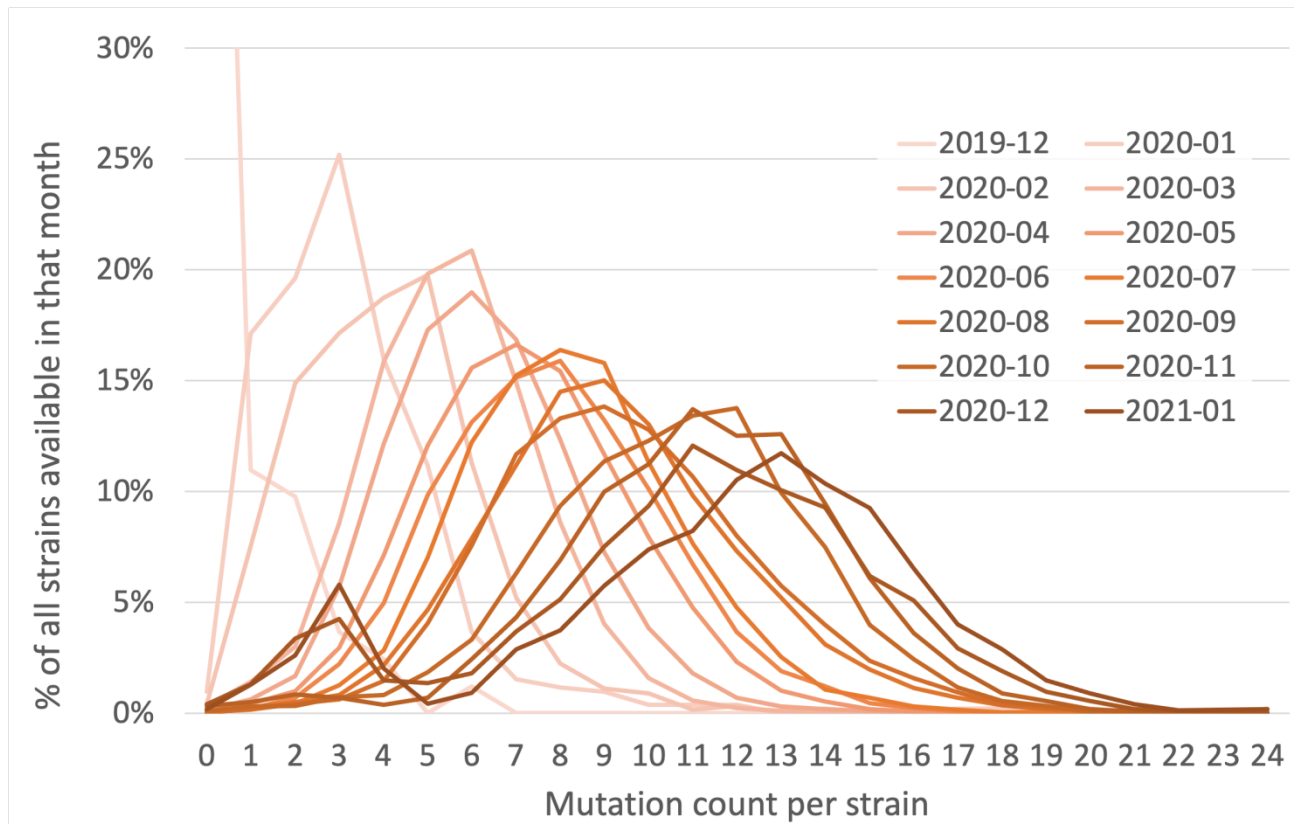


圖 3. (a) 從 2019 年 12 月到 2021 年 1 月每個毒株中的平均突變計數。(b) 從 2019 年 12 月到 2021 年 1 月每個毒株的突變計數分佈。a 中的資料標籤顯示了各個月用於分析的毒株數量。

基因組序列上各個位置自然發生的突變率不可能有很大差異，但是某些基因座進化迅速，而其他區域似乎受到極大限制。在迅速進化的區域，選擇壓力相對於病毒的整體增殖是寬鬆的，因此它可以迅速適應宿主的抗病毒反應。該觀察結果應促使監測特定功能的進化是否與傳播速率更高或更低的環境相關，同時還伴隨疾病嚴重程度的變化。通常與輕度疾病相關的突變影響蛋白 Orf8、Nsp6、Orf3a、Nsp4 和核衣殼磷蛋白 N。相反，特定突變位於突刺糖蛋白、取決於 RNA 種類的 RNA 聚合酶中，有時還位於 Orf3a、Nsp3、Orf6 和 N 中，與嚴重的疾病後果有關。最後，與嚴重疾病後果相關的突變也位於 Orf3a 和 Nsp7 中。不幸的是，在疾病傳播的背景下很少討論這些蛋白質的作用。此外，除了蛋白 S 在觸發免疫反應中的普遍作用外，在 22 個與疾病的臨床結果發生重大變化（減輕或加重）相關的突變中，還有 4 個（3 個與嚴重疾病相關，1 個與輕度疾病相關）被定位到一

個十個氨基酸長的磷酸化核衣殼 N 片段上。這指向病毒基因組中一個高度相關的位點 (Nagy et al, 2021)。

儘管一些電腦模擬和體外實驗的分析試圖研究突變會否對該病毒造成傷害，但這些分析往往只是結構導向而不是功能導向的，這樣會造成偏頗的結論。僅在持續時間短的譜系中才能瞭解突變是否有害。前期研究尚在不清楚某些快速傳播的變異毒株是否因建立者效應而造成的超級傳播事件，過了一段時間後，現在似乎已經確定某些突變確實在影響該疾病的傳播 (Borges *et al.*, 2020)。實際上，這種觀點只是事後的描述。而且很少會考慮到病毒是具有高度整合功能的。為了預測其進化方向，我們必須評估各種形成病毒的因素的功能性上位效應，即瞭解不同突變之間的聯系如何促進病毒的適應性。在當前情況下，以同類型的分析確定了幾個 SARS-CoV-2 基因座間的相互作用，包括 Orf3a 和 Nsp2、Nsp12、Nsp6 之間、Orf8 和 Nsp4 之間以及 Nsp2、Nsp13 和 Nsp14 之間 (Zeng *et al.*, 2020)。這些有待深入研究，因為部份相互作用尚未與疾病的嚴重程度關聯（見上文）。因此，對那些尚未與疾病嚴重性相關連的蛋白序列的進化作進一步的監測有著極大的意義。找出與輕度或重度疾病病例相關的病毒譜系是這類調查的重要結果，這結果能夠促使與疾病傳播行為相關的政策（通過限制聚集，戴口罩，限制旅遊等）更精確地匹配在這過程中發現的關鍵病毒功能的進化。

當病毒長時間停留在人群中，病毒週期中的各個步驟所受的壓力都會發生變化。冠狀病毒與人類有著共同演化的過程，有跡象顯示東亞人的基因網絡可能起到了局部保護作用 (Souilmi *et al.*, 2020)。有趣的是，一些最基本的病毒蛋白並沒有發生變化。例如，蛋白酶 Nsp5 的氨基酸序列是高度保守的，這遠遠超過 SARS-CoV-2 的變異毒株以及所有已知的冠狀病毒。這些在不同的病毒家族中高度保守因而沒有變化或變化很少的蛋白質是廣譜抗病毒藥物的潛在靶標。SARS-CoV-2 Nsp5 與 SARS-CoV-1 的氨基酸序列有 95% 相同。其三維結構可用於設計抑製劑，這種方法套用在控制 HIV 方面很成功 (Lubin *et al.*, 2020)。另一個關鍵特徵是病毒的殺傷程度與特定突變之間的關係，這個特徵必須與相關的元數據關聯起來。在 692 個 SARS-CoV-2 基因組序列中，地理起源和 COVID-19 病例嚴重程度有著統計學上顯著的相關性。研究發現地域差異與病例的嚴重性和等位基因變異有關，這在印度發現的病毒譜系中特別明顯 (Goyal *et al.*, 2021)。而除了這些基本的觀察外，也需

要提倡對 SARS-CoV-2 全基因組序列進行系統地測序，同時追蹤譜系在不同地區中的演變。應該要仔細檢測病毒毒力的發展趨勢，而地方性的變化則應立即觸發相對應的遏制政策以控制病毒擴散。

在我們預測未來走向時收到的關於快速傳播變異毒株的警告

正如前文描述，流行病最可能的早期發展是該病毒開始提高其傳播速度。這很可能是由兩個常見的功能引起的：病毒複製成功率以及傳播的增加（此處涉及許多影響因素，包括擁擠的環境、或從受感染者所產生穩定而長期的感染源等）。這些在幾個迅速傳播並取代了已有病毒株的變異毒株中都可以看到（<https://nextstrain.org/ncov/global>）。第一個記錄下來的符合這類發展的例子是突刺蛋白中的 D614G 突變，該突變增強了 S1/S2 連接處的切割 (Gobeil *et al.*, 2021)。在那之後直至撰寫本文時，全球至少發現了根據 Global Initiative on Sharing All Influenza Data (GISAID) 定義的八個主要進化枝：S、O、L、V、G、GH、GR 和 GV（<https://www.gisaid.org/phylogenetics/global/nextstrain/>）。在亞洲有著一種特有的模式，包括變種 G、GH、GR、L、S、O (Sengupta *et al.*, 2021)，而在歐洲這病毒正在迅速演變為多個新譜系 (Hodcroft, Domman *et al.*, 2021)。

不幸的是變種命名法有些不穩定而且令人困惑：一個利用病原體基因組資料的科學和公共衛生的開源專案 Nexstrain（<https://nextstrain.org>），提出了五個不同的進化枝：19A、19B、20A、20B 和 20C。Qingtian Guan 及其同事也提議將譜系分為五個進化枝，但名稱不同：G614、S84、V251、I378 和 D392 (Guan *et al.*, 2020)，這與 Andrew Rambaut 和同事根據英國的疫情提出的 A、B、B.1、B.1.1 和 B.1.177 進化枝相關（<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>）。

與上述這些命名法類似的是有著六個進化枝及作為其基礎的標誌性 SNP 的命名法，這些標誌性 SNPs 已通過早期獲得的變異序列進行了驗證。例如，VI 型進化枝的特徵是四個標誌性的 SNP C241U（5'UTR）、C3037U（Nsp3 F924F）、C14408U（Nsp12

P4715L) 和 A23403G (Spike D614G)，它們有著強烈的等位基因關聯。該變種由這些 C 到 U 的突變主導著，在很早的時候就已經成為優勢毒株 (H.-C. Yang *et al.*, 2020)。為了說明主要進化枝的變化，在印度特蘭甘納邦最初的 (武漢) 進化枝 19A 被進化枝 20A 迅速取代，進化枝 20A 有著 5'UTR 的 C241U 突變和突刺蛋白 D614G 突變。緊接著的是已成為主流的進化枝 20B (Gupta *et al.*, 2021)。隨後許多突刺蛋白中的其他變化被發現在迅速傳播著 (Vilar and Isom, 2021)，表明該蛋白的大部分選擇壓力來自對宿主的適應。因此，我們可以預期，在疫苗的選擇壓力下，這種蛋白質以及較小程度的核衣殼蛋白將進化得最快。這反映在圖 4 和補充圖 1 中，它們展示了突變如何在世界範圍內傳播。但是由於資料收集極不均衡，因此這只是對目前狀況的總體概述，強調了隨著病毒在不同國家中傳播應進行全面的病毒基因組測序以瞭解突變密度的需求。雖然突刺蛋白的全球突變密度在逐漸增加，但自 2020 年 9 月以來，在墨西哥和巴西等一些國家增加得非常快。

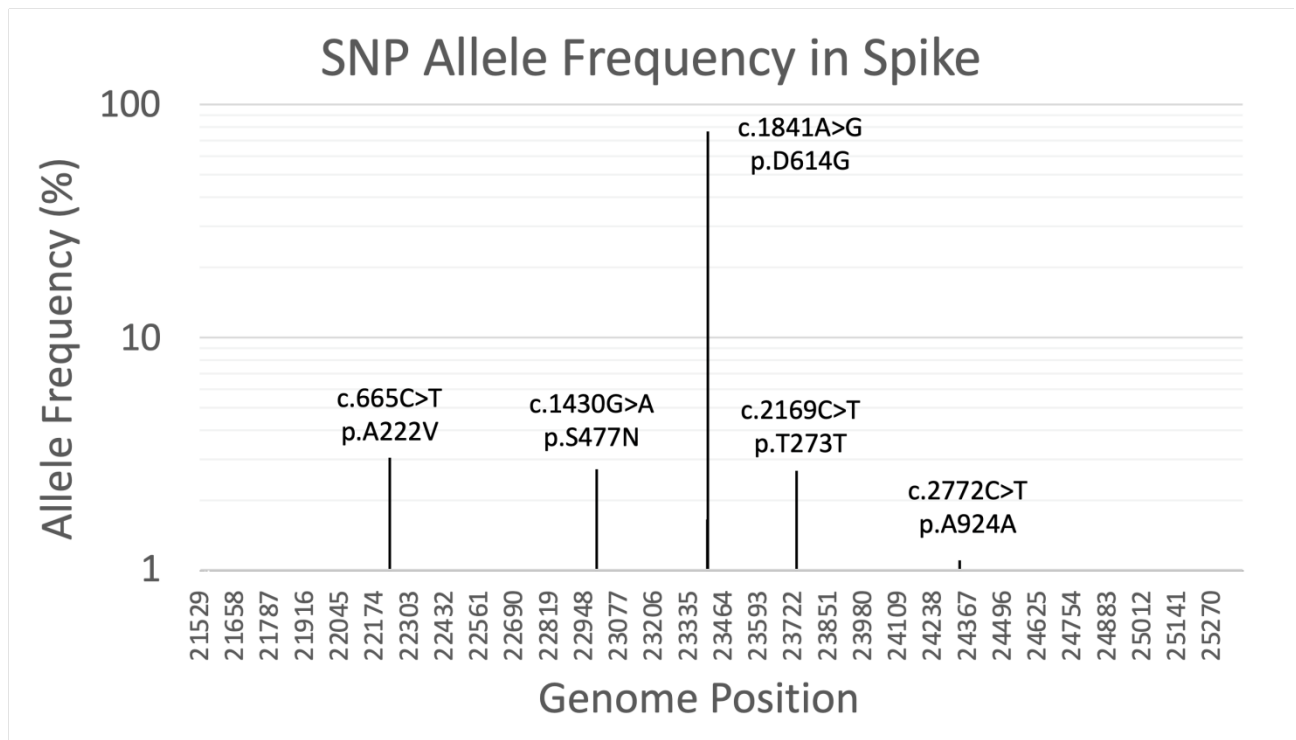


圖 4. 193,687 個 SARS-CoV-2 全基因組中突刺蛋白突變的 SNP 等位元基因頻率 (Allele Frequency, AF)。y 軸是以 10 為底的對數刻度。我們僅展示了 AF ≥ 1% 的突變。

在發表這些研究後，其他重要變種的出現也有被報導。在撰寫本文時，具有許多突變的幾種變異毒株正在廣泛傳播（<https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/scientific-brief-emerging-variants.html>），英國變種 B.1.1.7、南非變種 B.1.351 和巴西變種 B.1.248 (Firestone *et al.*, 2021) 正在世界範圍內推廣。許多突變會同時發生變化（<https://covariants.org>）。這可能是由於趨同進化導致的，這暗示著某些突變之間的上位相互作用。此類突變很可能會說明指出病毒的功能特徵，以預測其進一步的進化，以及可能的毒力減弱。這些突變的特徵除了是一連串的突變之外，英國變種中還存在一個小缺失突變。通常來說在進化過程中插入/缺失突變是非常常見的（請參見圖 2）。但值得注意的是蝙蝠病毒基因組中的插入突變實際上是 SARS-CoV-2 感染人類的起源。譜系 B.1.525 展示了共同突變，譜系 B.1.525 是在 13 個國家/地區中發現的一個有趣新變異毒株，其突刺蛋白突變為 E484K、Q677H、F888L，並且有一系列與 B.1.1.7 相同的缺失突變（https://cov-lineages.org/global_report_B.1.525.html）。如果我們要預測疫情的未來，我們必須一直銘記突變的潛在重要性，而不僅是那些影響突刺蛋白的突變。

現在尚不清楚 D614G 突變是否會影響疾病的嚴重程度。但是在最近發現的三個主要變種中 (Zhou *et al.*, 2021)，已通過 B.1.1.7 變異毒株明確地證明了嚴重程度的提高 (Davies *et al.*, 2021)。許多變異毒株正在逐步替代早期病毒毒株，其中 B.1.525（丹麥、英國、尼日利亞）的嚴重程度未知，但它的傳播潛力卻不斷在增加。

突刺蛋白以外一些重要變異的例子

大多數觀察僅僅是描述性的，但是對於那些影響著進化的病毒功能改變的發現來說至關重要，特別是那些在可以預期的突刺蛋白變異之外的改變。譬如我們預測複製過程的改變將會產生兩個截然不同的後果。一方面，長遠地它將損害變異毒株的傳播而導致病毒的滅絕，但另一方面，它又可能會在短期內拓展其進化空間。在這種情況下，應該把觀察到的病毒產生進化“綻放”的趨勢與急速擴張的譜系聯繫起來，例如對於病毒脫落胞嘧啶的自

然趨勢的可能的逆轉 (Cluzel *et al.*, 2020)。這一過程的確對於預測在疫情前六個月內病毒進化的後果（產生了多個進化枝）非常有價值 (Koyama *et al.*, 2020)。

在重要的變異中，病毒基因組 5'UTR 中普遍存在的 C241U 突變尚未被發現與任何特定表型相關。它通常被認為是中性的突變，它使病毒對宿主的新陳代謝獲得了小程度的適應性，並提高了對 viperin 的抵抗力。此外，它位於病毒基因組的 5'UTR 前導序列，這對於它控制翻譯和宿主特異性而言特別重要 (Tidu *et al.*, 2020)。順著這思路，有許多病毒進化“綻放”的例子。例如，從一個早期突變開始的一連串有趣的突變演替，G11083U（蛋白質 Nsp6，L37F）現在已廣泛分佈於世界各地，並且與印度的進化枝有關 (Banerjee *et al.*, 2020)。在變異 G2891A（A876T，蛋白 Nsp3 的泛素樣結構域）之前，在多個國家發現了另一個突變 G1440A（G392D，蛋白質 Nsp2）(Liu *et al.*, 2020)，這隨後引發了翻譯 Orf7a 和 Orf7b 之間的衝突。這表明對於這些蛋白質來說任一表達都存在成本或收益難題。與這位元點有關的病毒譜系值得追蹤，因為它可能導致有趣的減毒形式 (Cluzel *et al.*, 2020)。同樣，SARS 相關冠狀病毒的 Orf8 區域也是高度變化的。它在疫情中不斷變化，說明受到持續的選擇壓力，有時會產生兩種肽 Orf8a 和 Orf8b (S. Chen *et al.*, 2020)。在疫情的第一階段來自四個不同國家的七個樣本中，Orf8 區域有著一個開枝散葉般的突變產生過程，從第一次突變到最後一個突變跨越了六星期 (Cluzel *et al.*, 2020)。Orf8 蛋白在感染週期的末尾表達。監測它們促使病毒毒力進化的方式非常重要 (Neches *et al.*, 2021)。

最後，人們對監測病毒複製酶 Nsp12 的進化有著明顯的興趣。在疫情早期，從廣泛的 5'端 C241U 突變開始的一連串突變，接著是 Nsp12 中鋅指末端的突變 C14408U（P314L）。這種突變出現在病毒進化樹的許多進化枝中。它以值得注意的方式改變了複製酶的活性，因為這種突變之後是病毒新譜系的“綻放”，這表明改變的複製過程是誘變的 (Cluzel *et al.*, 2020)。一個值得認真監測的例子是非常有趣的一連串演替：突刺蛋白中的 A23403G（D614G）和 C3037U（同義）突變、與 Orf3a 形成鉀離子通道有關並可能對該蛋白的功能產生負面影響的突變 G25563U（Q57H）(Issa *et al.*, 2020)、蛋白質 Nsp2 中的 C1059U（T265I）、蛋白酶 Nsp3 的 SUD-N 區域中的三聯體 G4181A（A1305T），緊接著的是突變 G4285U（E1340D）和 G28209U 導致翻譯終止于 Orf8 蛋白質的 E106，它們

的再次出現成為了病毒進化的重要標誌 (Neches *et al.*, 2021)。自大流行開始已經過去了一年，應該使用新的解釋來探究仍在不斷出現的突變，特別是基於當前疫苗接種正在迅速進行的情況下。

疫苗接種及其預期後果

接種疫苗已成為控制甚至根除傳染病的首選方法。這在多數情況下是正確的，但有些病毒卻無法被研發出有效的疫苗，例如 HIV (Oyston and Robinson, 2012)。在減毒疫苗研發並使用成功之後 (Theiler and Smith, 1937)，許多新的疫苗種類在這之後被陸續開發 (Smith, 2012)。就冠狀病毒而言，疫苗已成功應用於預防動物疾病 (Cruz *et al.*, 2010; Singh *et al.*, 2019)。因此，我們預期至少在這場大流行的早期，疫苗接種將有效控制疫情。但是，正如我們所看到的，該病毒的進化非常迅速，因此病毒的變異具有導致現有的 COVID-19 疫苗效力降低的持續風險。發展出多個毒株是常見的進化特徵，並可能導致疫苗的減效或無效 (Zeng *et al.*, 2017)。在南非首次發現的 B.1.351 變異毒株似乎就是這種情況 (Diamond *et al.*, 2021)。此外，如果疫苗接種行動太慢，將為病毒提供充足的時間來進化出可以逃避疫苗免疫的變異毒株。還有，只基於病毒中單一蛋白的疫苗，或更糟糕的只基於某病毒蛋白結構域的疫苗，將會選擇性地驅使突變在這些蛋白質或蛋白質結構域中快速富集 (J. Yang *et al.*, 2020)。

其他原因所導致的感染也可能影響病毒的進化。對於 SARS-CoV-1，某些人群似乎不受感染的這一事實表明，先前的某些感染可能已經誘導了交叉保護作用 (Ng *et al.*, 2003)。上面我們已經注意到了吸煙習慣和輕度感染之間出乎意料的關係。這意味著流感、流感疫苗接種與 COVID-19 疫情發展之間可能在未來會呈現關聯。感染過流感病毒的人似乎更容易感染 SARS-CoV-2 (Bai *et al.*, 2021)。但是，回顧性研究並未發現接種了流感疫苗的人更容易得 COVID-19，且恰恰相反 (Green *et al.*, 2020)。總之，所有相關的觀察結果都呼籲實施快速的疫苗接種計畫，同時監視尚未達到有效接種比例的人群中的病毒基因組的進化相關變化。結合上述觀點，我們認為，對這場大流行發展的預測，必須分別考慮已接種的人群和未完成接種的人群。

長期進化

1918-1919 年的流感疫情可能有助於我們預測 COVID-19 流行病可能會發生的情況。雖然無法完整重現當時發生的所有情景，至少我們知道這場疫情分三個發展階段。第一階段看上去像嚴重的流行性感冒。隨後，它發展成一種非常嚴重的疾病。最終，它適應了人類宿主，並最終發展成目前的樣子。在該疫情徹底消失之前，它還從人傳到了豬 (Shope, 1936)。是基因重排 (reassortment) 事件導致了這種更強的流感病毒。流感病毒由獨立的序列片段組成，當與相似的病毒共同感染時可能會發生基因重排，從而引入新蛋白質變體，說明病毒繁殖。有趣的是，這些高毒性變體不涉及病毒感染並從宿主細胞中釋放所需的血凝素或神經氨酸酶蛋白 (Reid *et al.*, 2004)。這再次強調，除了那些已經可被免疫系統快速識別的蛋白質，我們更要分析和考慮那些仍然被忽視的病毒碼。

SARS-CoV-2 不會發生基因重排，因為它是由單個 RNA 元件組成的，但它易於與其他 RNA 重組 (recombination)，所以 SARS-CoV-2 在短期內最令人擔憂的進化特徵是進行大規模的重組，而且其遵循重組率與基因組長度的負相關關係 (Zhang *et al.*, 2005; Goldstein *et al.*, 2021)。這個過程似乎取決於 Nsp14 蛋白的活性 (Gribble *et al.*, 2021)。這個過程可能源於同時感染了不同的 SARS-CoV-2 毒株 – 在這方面這將與有性繁殖和基因水準轉移相似，導致穆勒棘輪 (Muller's ratchet) 的重啟。這過程可能會隨著旅遊限制的放寬而增強 – 但也可能源於同時感染了不同的病毒，甚至是人工構建的序列都可以被重組進來。所以說，在最近研發的 mRNA 類疫苗中用偽尿苷替代尿苷可以說是一項積極的創新以避免病毒把疫苗序列給重組進來，從而擴大病毒的進化潛力。我們還要關注 Nsp14 校對核酸外切酶所可能進化出的變體。我們應仔細監測在 Nsp14 中發生的突變，因為它們可能會擴大病毒的進化資源，在其突刺蛋白、核蛋白殼和 Orf8 蛋白中產生重組熱點。冠狀病毒普遍有基因重組的特徵，這在正鏈單鏈 RNA 病毒中普遍存在 (Zhang *et al.*, 2005; Patiño-Galindo *et al.*, 2021)。除了進化逆轉之外，重組還可能將不同環境下各自所產生的突變整合到一起。由於病毒重組發生在同時感染不同毒株的過程中，因此要特別關注人群密集的環境，尤其是封閉 (室內) 的環境，安裝有空氣內迴圈的環境，以及需要進行聲樂活動的環境。為了

抗疫，在完成疫苗大範圍覆蓋或在有效的抗病毒藥物可用之前，我們必須儘量減少此類環境下的活動。

為了預測該場大流行的未來，我們必須仔細分析病毒的趨向性（*tropism*，嗜性）將如何演化。目前，呼吸系統感染（具有呼吸系統趨向性）是主要途徑，但我們知道在其他冠狀病毒中，這種趨向可以非常迅速地改變。例如，禽冠狀病毒突刺蛋白中的三個氨基酸變化使該病毒能與腎細胞結合 (Bouwman, Parsons, *et al.*, 2020)。在小鼠中，冠狀病毒可能表現出神經趨向性 (Pasick *et al.*, 1994)。細胞進入和融合的自然選擇過程與突刺蛋白的動態結構密切相關。我們已經看到，在蝙蝠冠狀病毒中插入的 *furin* 蛋白酶切割位點使其能夠改變其宿主並適應人類的 ACE2 受體 (Coutard *et al.*, 2020)。現在，該病毒已經找到了另一種進入受體，即通過突刺蛋白與 CD147 蛋白的相互作用。至少在體外實驗模型中 CD147 在多種細胞中均有表達，包括上皮細胞和神經元細胞 (K. Wang *et al.*, 2020)。這點很重要，因為利用多種受體可以使病毒通過不同的細胞進入人體，也就是可以從不同的器官進入人體。另一方面，MERS-CoV 的受體，二肽基肽酶 DPP4 的多態性會對病毒的進入產生負面影響 (Kleine-Weber *et al.*, 2020)。其他冠狀病毒還會利用到其他一些受體：豬 Δ 冠狀病毒 (porcine delta coronavirus) 利用氨基肽酶 N (APN) 作為受體，並通過突刺蛋白的結構域 S2 與 APN 相互作用 (Li *et al.*, 2018)。小鼠肝炎冠狀病毒 (MHV) 是唯一已知的利用其突刺蛋白的 N 末端結構域 (NTD, N-terminal domain) 識別另一種受體 CEACAM1a 的冠狀病毒 (Shang *et al.*, 2020)。

另一個值得留意的病毒碼是保護病毒不被降解的蛋白質糖基化，它有可能被病毒用作進入靶細胞的一種方式。舉個例子，雞冠狀病毒感染性支氣管炎病毒 (IBV, chicken coronavirus infectious bronchitis virus) 通過病毒的“重度 N-糖基化的附著蛋白突刺”與宿主的“ α -2,3-連接唾液酸受體 Neu5Ac”結合而進入宿主細胞 (Bouwman, Habraeken, *et al.*, 2020)。人冠狀病毒 OC43 顯然是從牛冠狀病毒 (BCoV) 而來。它通過其突刺蛋白與基於 9-O-乙醯化唾液酸聚糖 (sialoglycan-based receptors with 9-O-acetylated sialic acid) 的受體結合並以血凝素酯酶作為破壞受體的酶 (Lang *et al.*, 2020)。9-O-乙醯化唾液酸聚糖的受體結合位點在所有冠狀病毒的突刺糖蛋白中都是保守的，其結構類似於冠狀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體結合袋結構域 (Tortorici and Veasler,

2019)。因此，我們應該開發方法以監測針對病毒糖蛋白的抗體的變化，以及基因組突變導致的病毒蛋白糖基化改變。

展望

COVID-19 大流行開始了一年有多，我們正處於一個過渡時期，病毒正遭受著各種能將其推向長期存活的選擇壓力的考驗。現在已經有三個主要因素促使病毒向可能的危險結果方向發展：毒株之間的重組、非常緩慢的疫苗接種運動以及抗病毒藥物研究的不足。同時，受感染的人數非常多，因此在擁擠的環境中發生不同病毒毒株的共同感染已經不再稀奇。能夠即時跟蹤完整基因組序列的演變變得至關重要，這樣我們才能查明病毒蛋白質中可能導致其減毒或相反導致更嚴重後果的目標位元點。

我們提倡廣泛收集病毒的完整基因組序列。而且，這些完整的序列只有在與相關的中繼資料相關聯時才有意義。另外，病人的患病史是重要的中繼資料。中繼資料越多越好，而中繼資料的收集必須要合理地標準化。

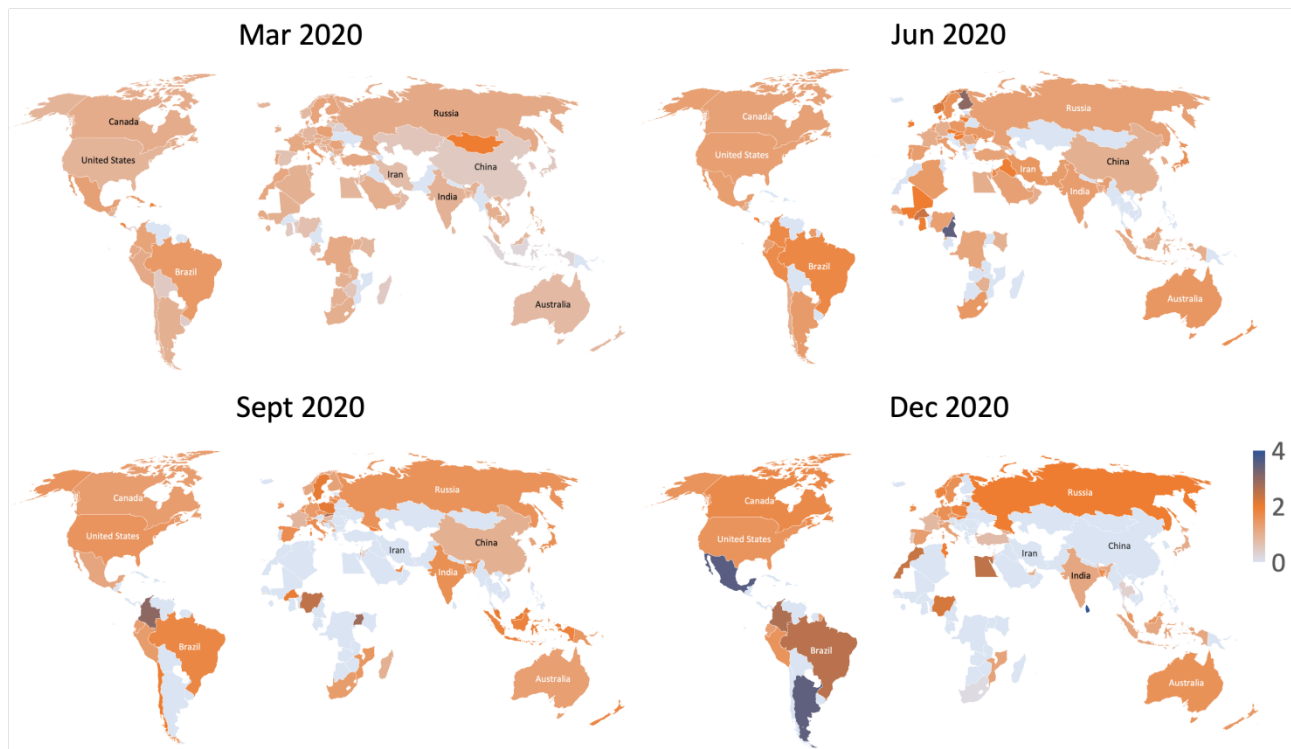
以下是迫切需要實施的措施，它們符合“認識你的敵人”的關鍵原則。

- 在任何地方都盡可能地對病毒的整個基因組（而不僅僅是突刺蛋白）進行測序。當測序技術不斷改進時，這應該是切實可行的。例如，對 752 個臨床樣品進行納米孔測序很快且很輕易地確定了該病毒的 3 個進化枝 (Bhoyar *et al.*, 2021)。
- 將重要的中繼資料與這些序列相關聯（包括所有我們可以使用的關於感染者的資訊以及臨床資料），並將中繼資料與特定突變相關聯，且不局限於只是突刺蛋白。
- 建立毒株譜系及其傳播情況，並關聯到各種參數，包括年齡、性別、種族（如果情況允許的話）、體重、人類遺傳多態性特徵（HLA）、Lewis 血型、營養習慣和平常行為（例如吸煙）等標準資料。另外，由於傳播需要人與人之間的接觸，我們應當確定感染的確切地點（國家、城市、建築物），並將其與氣象參數相關聯。
- 專注於進化模式的變化，如毒株分支的爆發、核苷酸進化模式的改變（C 或 G 丟失趨勢的逆轉、顛換等情況）。並嘗試將這些進化模式與特定蛋白中的突變聯繫起來。
- 識別正在衰減的毒株品系，並在受控下允許其傳播。監控病毒從肺到腸的趨向性變化，反之亦然。

- 找出情況嚴重的毒株，並施加嚴格的局部防範措施。
- 追溯傳播的上游，而不是幫助不大的下游。對感染者及其接觸者的活動進行嚴格控制。
- 盡可能加快疫苗接種的速度，尤其是在高發病率和已存在多種變異毒株的人群中。緊急開發涵蓋變異毒株第二代和第三代疫苗，以彌補第一代的不足。
- 基於較穩定的病毒蛋白，大量投資開發新的抗病毒劑，並進行廣泛的篩查。

最後，似乎我們將不得不與 SARS-CoV-2 共存。這意味著，為了控制其負面後果，我們必須仔細跟蹤其抗原決定簇的演變。COVID-19 可能會演變成類似季節性流感的情況，並且每年都需要使用不同的疫苗。其與其他呼吸系統疾病（尤其是流感）的共同進化必須要予以高度重視，因為忽略對後者保持穩定的群體免疫力可能會導致糟糕的後果。

補充圖



補充圖 1. 突刺蛋白在 2020 年四個不同月份的時間-地理突變密度。當月可用基因組少於五個的國家沒有被畫出（灰色）。分析依賴于不同國家研究人員提交給 GISAID 的毒株，因

為國家間基因組序列採樣極為不均勻 (Hodcroft, De Maio, *et al.*, 2021)，我們要謹慎從此圖中得出結論，該圖僅是指示性的。例如，迄今為止，英國已對每 100,000 個案例中的近 5,000 個樣本進行了排序，而美國為每 100,000 個案例中有 320 個樣本，巴西為每 100,000 個案例中有 30 個樣本（請參閱門戶網站 <http://go.nature.com/3qutqwj>）。即使在英國，某些區域在測序數據方面的表現也要好於其他區域。

補充注解

我們篩選並從 GISAID 下載了從 2019 年 12 月 26 日算起至 2021 年 1 月 25 日採樣的 193,687 株 SARS-CoV-2 的基因組。在我們的篩選中，我們要求毒株具有 1) 完整的基因組，2) 足夠高的測序深度，和 3) 具體到月的樣品收集日期。我們衷心感謝這 193,687 個基因組的，負責原始樣品收集實驗室，和負責進行測序並把序列提交給 GISAID 倡議進行共用的實驗室。全部這些基因組的統一編號收錄在這個 VCF 檔中，網址為 <http://www.bio8.cs.hku.hk/adanchin/variantsOf193687nCovAgainstWIV04.vcf.gz>，這其中已包含了我們在這些毒株中所找到的突變。根據這些統一編號，我們可以直接通過 www.gisaid.org 與所有資料提交者取得聯繫。

我們使用 WIV04 (GISAID: EPI_ISL_402124, NCBI: MN996528) (Zhou *et al.*, 2020) 作為參考基因組，並將所有其他基因組與 WIV04 進行比對。WIV04 被用於開發國藥集團的滅活 SARS-CoV-2 疫苗 (Xia *et al.*, 2020)。在最早發佈的基因組中，另一個通常用作 SARS-CoV-2 參考的基因組是 Wuhan-Hu-1 (GISAID: EPI_ISL_402125, NCBI: NC_045512)。與 WIV04 相比，它在基因命名方面有細微差別（例如，Wuhan-Hu-1 中的 Orf8 與 WIV04 中的 NS8），但所有基因的起始和終止位置相同。因此，WIV04 和 Wuhan-Hu-1 之間的基因名稱是可互換的。在這項研究中，我們將使用 Wuhan-Hu-1 中的基因命名（即以 'ORF' 代替 'NS' 開頭）。

我們使用 MAFFT (Katoh *et al.*, 2002) 進行多序列比對。多序列比對在以大量序列作為輸入時非常的很慢。通過在 MAFFT 中設置使用參考基因組，我們得以在一個 24 核計算伺服器上用了不到一個月的時間完成比對，但只能獲得 SNP 和缺失突變 (deletion)。我們使用了 jvarkit-msa2vcf (<https://github.com/lindenb/jvarkit>) 將 MAFFT 輸出中的突變提取到 VCF 檔中。我們使用 BCFtools (Li, 2011) 歸一化所有樣本中的突變表示方法（將不同樣本中使用不同表示方法的相同突變合併）。我們刪除了除了 'A', 'C', 'G', or 'T' 外，包含的未知域

基 'N' 和 IUPAC 核苷酸代碼的突變。我們刪除了長度超過 50bp 的缺失突變，因為它們更可能由不完整的基因組組裝得來，而非一個真正的突變。第一段落中給出的 VCF 檔連結提供了所有 193,687 個基因組的突變。所有後續分析均基於此 VCF 檔。

補充表

補充表 1. 總結自 193,687 個 SARS-CoV-2 毒株的基因突變計數和發生率。基於每個基因的長度，我們將突變發生率標準化為每 1kbp 的突變數。對於每個基因，我們展示了在 ≥ 1 個毒株， ≥ 5 個毒株， ≥ 10 個毒株， ≥ 100 個毒株和 ≥ 1000 個毒株中突變發生率。

引用文獻

- 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 Karakülah, 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., LePendou, 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 Györffy, 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., Kyrpidis, 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. *Viol 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.