Investigating the emergence of pathogen X

It has been predicted by WHO and other international experts that it is almost certain that a novel pathogen (collectively termed as 'pathogen X') will emerge and cause another global outbreaks again in the future. Therefore, it is important to learn from this COVID-19 pandemic to better prepare for the future pathogen. It was evident that pathogen genomic data and analyses are very useful to investigate the disease transmission, because the transmission leave footprints on the pathogen genome sequence. For example, phylogenetic tree analysis of the pathogen genome sequences can find the relatedness of the newly emerged pathogen in human to any existing known pathogens, shedding light on the animal source of human infections [1, 2, 3]. It can also be used to track the time when the disease emergence started [4, 5], particularly based on the principle that many pathogens especially fast-evolving RNA viruses mutate in certain rate over the time (different types of pathogens may have different mutation rates) [6], and thus the more mutations indicate the more distant evolutionary relationship between the two pathogen sequences. This approach is useful to model the evolution of the pathogen and back track its emergence history.

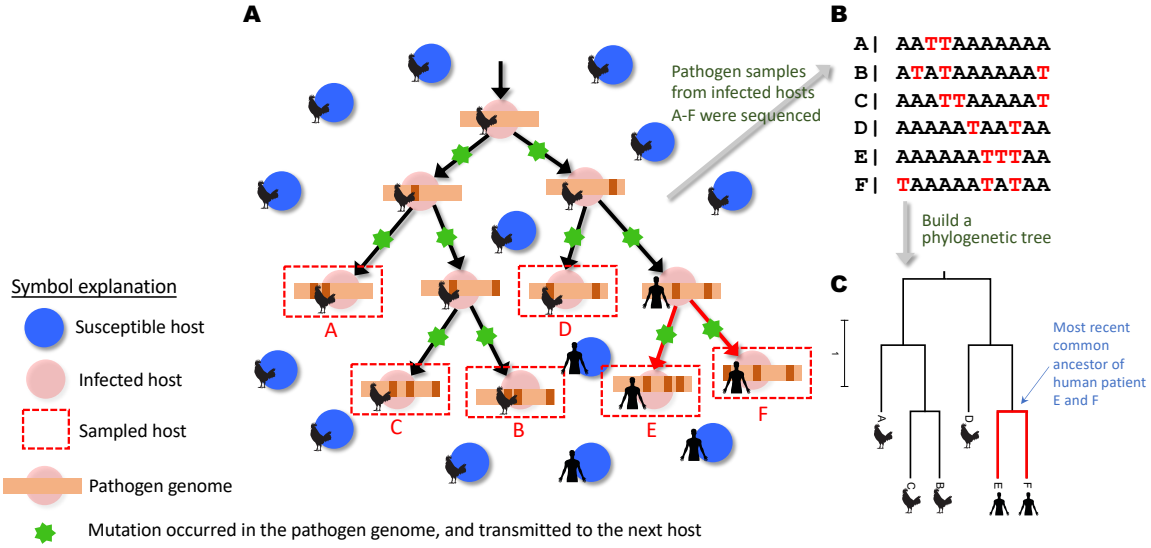

Figure 1. The concept of analysing disease transmission footprint in pathogen genomes.

(A) A scenario of disease transmission (indicated by black arrows) in chicken, which subsequently transmitted to three human patients. Mutations (green stars or dark bars in the rectangle) can occur in the pathogen genome when it infected a chicken or human host; the mutant pathogen might transmit to the next host. Pathogen samples from several hosts (four chicken and two humans in this example) were sequenced. (B) The pathogen genomic sequences (illustrative) obtained from the hosts A-F. (C) A phylogenetic tree can be built based on the genome sequences, where the tree branch lengths reflects the amount of mutations occurred in the genomes, and the grouping of branches refer to the hypothetic ancestor. For example, the node grouping E and F is called 'the

most recent common ancestor' of E and F, which is the patient zero (the first patient who has not been sequenced) in the human transmission cluster. The scale bar shown on the left indicates the branch length that reflects a single mutation in the sequence.

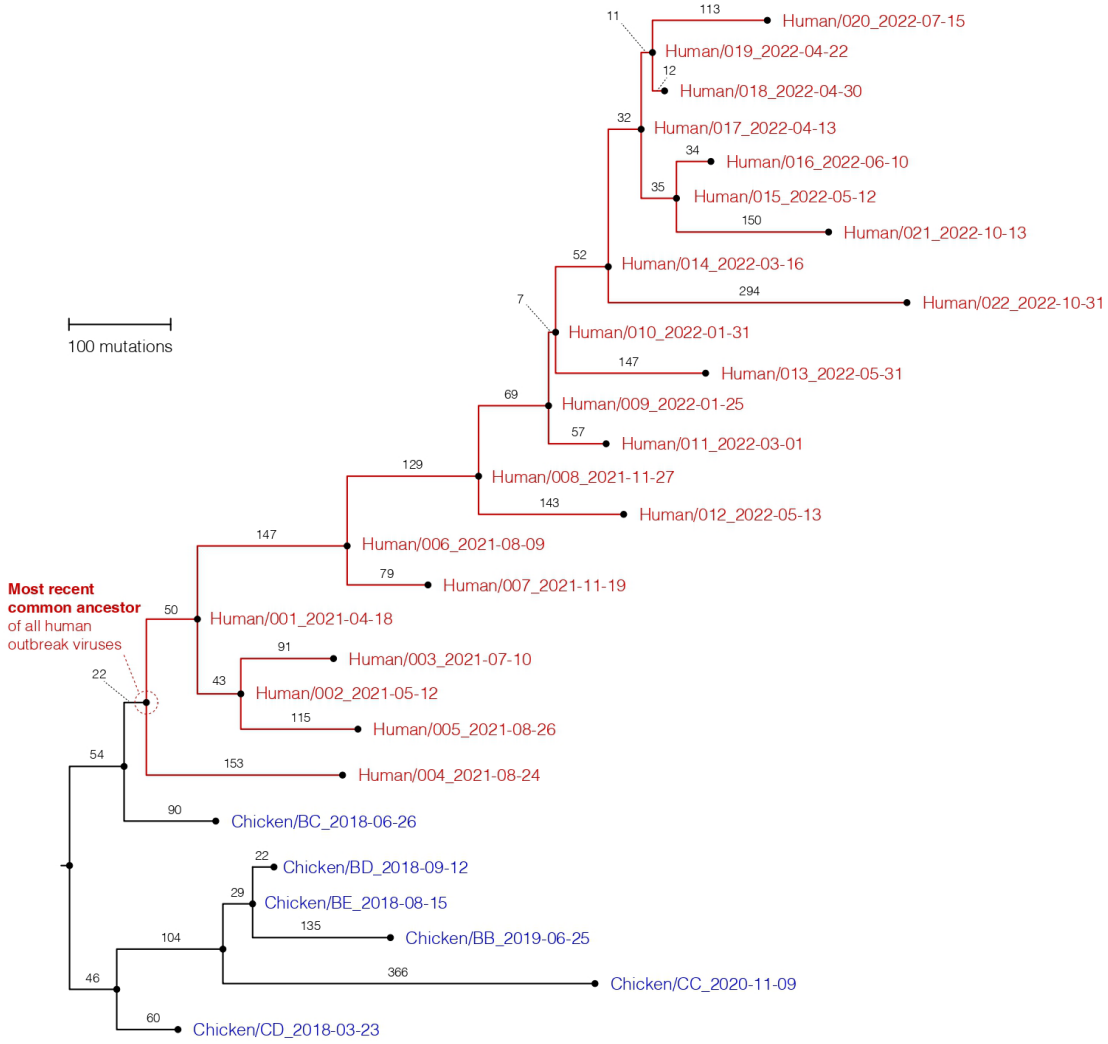

Figure 2. The phylogenetic tree built from the virus sequences from human patients (red) and infected chickens (blue).

Tasks

Imagine that recently there was an outbreak of respiratory diseases by unknown cause. The pathogen from the patients was soon sequenced and revealed to the genetically related to a virus previously found in chicken. A phylogenetic tree was built from the pathogen genetic sequences obtained from the 22 human patients (names start with 'Human/…') in the outbreak and infected chickens (names start with 'Chicken/…'). Your team is requested to analyse based on the phylogenetic tree to answer the following questions in your report:

a. How fast did this new pathogen evolve when it was transmitting in human population (i.e. what is the mutation rate in the unit of 'mutations per year' or 'mutation per day')?
b. When did this new pathogen start the outbreak in humans? How many days earlier than the first reported case?
c. What might this viral pathogen be? Why? (This is an open question.)
d. Based on all the data and your findings, what are your suggestions on preventing the future similar outbreaks from happening? (This is an open question.)

References:

[1] Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, & Pei Y-Y. (2020). A new coronavirus associated with human respiratory disease in China. Nature, 579(7798), 265-269. (https://www.nature.com/articles/s41586-020-2008-3)

[2] Lam TT-Y, Wang J, Shen Y, Zhou B, Duan L, Cheung C-L, Ma C, Lycett SJ, Leung CY-H, Chen X, Li L, et al. (2013). The genesis and source of the H7N9 influenza viruses causing human infections in China. Nature 502:241-244. (https://pubmed.ncbi.nlm.nih.gov/23965623/)

[3] Smith G J, Vijaykrishna D, Bahl J, Lycett S J, Worobey M, Pybus O G, Ma S K, Cheung C L, Raghwani J, Bhatt S, et al. (2009). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature, 459(7250), 1122-1125. (https://www.nature.com/articles/nature08182)

[4] Zeng F, Chow KY, Leung FC. (2003). Estimated timing of the last common ancestor of the SARS coronavirus. N Engl J Med 349:2469-70. (https://www.nejm.org/doi/full/10.1056/NEJM200312183492523)

[5] de Oliveira T, Pybus OG, Rambaut A, Salemi M, Cassol S, Ciccozzi M, Rezza G, Gattinara GC, D'Arrigo R, Amicosante M, Perrin L, Colizzi V, Perno CF, Benghazi Study G. (2006). Molecular epidemiology: HIV-1 and HCV sequences from Libyan outbreak. Nature 444:836-7. (https://www.nature.com/articles/444836a.pdf)

[6] Siebenga JJ, Beersma MF, Vennema H, van Biezen P, Hartwig NJ, Koopmans M. (2008). High prevalence of prolonged norovirus shedding and illness among hospitalized patients: a model for in vivo molecular evolution. J Infect Dis 198:994-1001.. (https://academic.oup.com/jid/article/198/7/994/2192243)

Submission

Your solution paper should include a 1-page Summary Sheet and a piece of short public science essay. The body cannot exceed 20 pages for a maximum of 23 pages with the Summary Sheet and short essay inclusive. The appendices and references should appear at the end of the paper and do not count towards the 23 pages limit.

调查新发病原体 X

根据世界卫生组织及其他国际专家的预测，未来几乎必定会出现一种新型病原体（统称为“病原体 X”），并再次引起全球性的疾病暴发。因此，必须从这次 COVID-19 大流行中吸取教训，以更好地应对未来的病原体传播。病原体的传播会在基因组序列上留下踪迹，因此分析病原体基因组数据对研究疾病传播发挥着重要作用。例如，基因组序列的系统发育树分析可以发现新兴病原体与任何已知病原体的关联，以揭示感染人类的动物来[参考资料 1,2,3]。系统发育树还可以用于追踪疾病出现的时间[参考资料 4,5]，其原理在于，许多病原体，尤其是快速进化的 RNA 病毒会随着时间以某种速度发生突变（不同类型的病原体可能有不同的突变率）[参考资料 6]，突变越多则说明两个病原体序列之间的进化关系越远。这种方法常用于建立病原体的进化模型，溯源其传播历史。

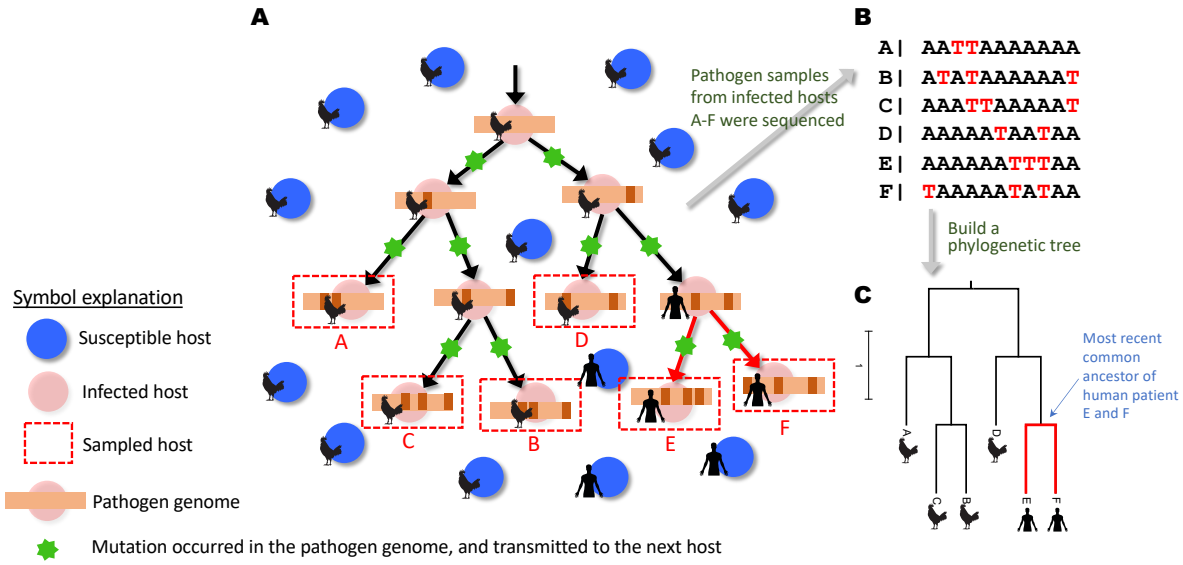

图 1. 分析病原体基因组中疾病传播踪迹的概念示意图

(A)疾病在鸡体内传播（黑色箭头），随后传播给 3 个人类患者。当病原体感染鸡或人类宿主时，其基因组可能发生突变（绿色星号和矩形中的深色位点），突变的病原体可能传播给下一个宿主。在这个例子中，我们对来自部分宿主（4 只鸡和 2 个人）的病原体样本进行测序。(B)从宿主 A-F（如图）获得的病原体基因组序列。(C)根据基因组序列建立的系统发育树，其中树的分支长度反映了发生突变的数量，分支共同归属的节点是假定祖先。例如，指向 E 和 F 的节点称为 E 和 F 的 “最近共同祖先”，它是人类传播集群中的零号病人（又名首例病人、索引病人，一般很难发现并测序）。左侧标尺表示序列中单个突变的支长。

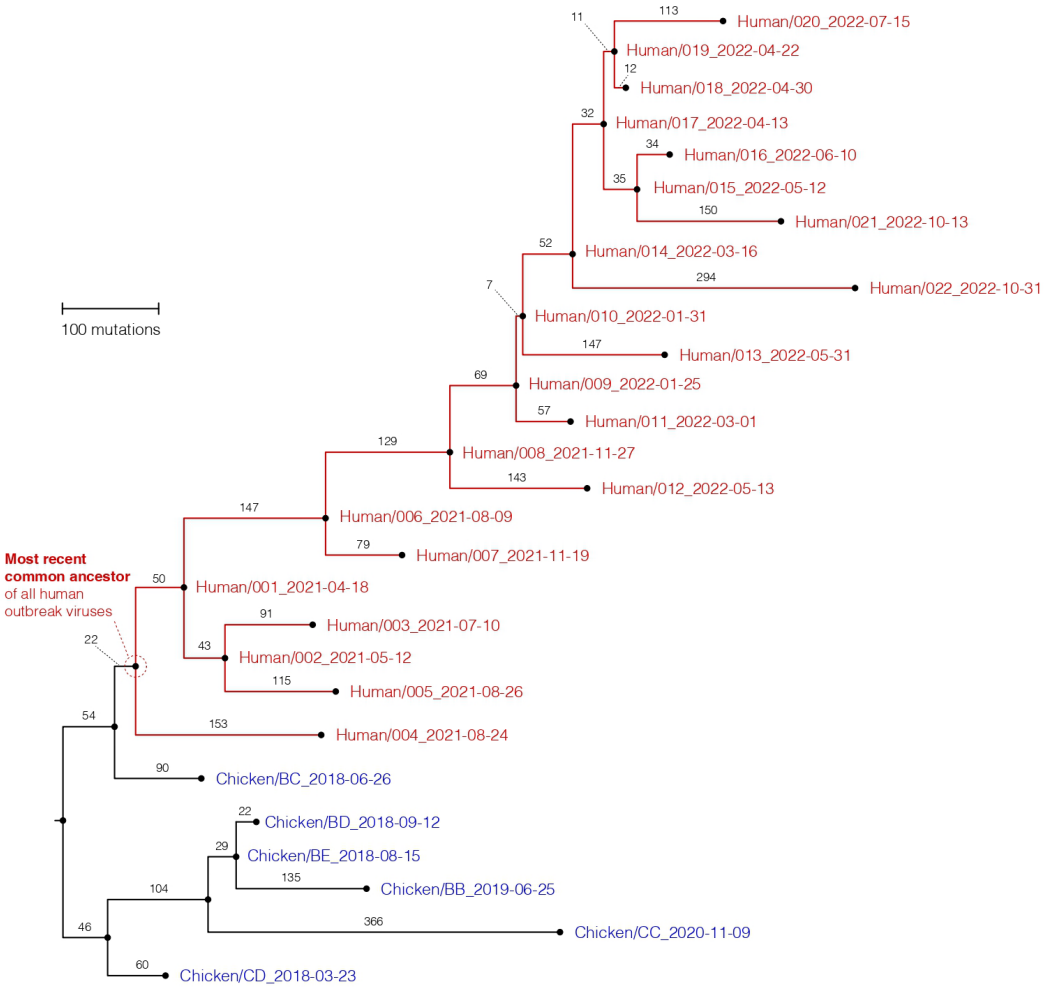

图 2. 由人类患者（红色）和受感染鸡（蓝色）的病毒序列建立的系统发育树

任务

假设近期爆发了一场原因不明的呼吸道疾病。来自病人的病原体很快被测序，并发现与以前在鸡身上发现的一种病毒基因相近。使用这次爆发的 22 名病人和受感染的鸡身上获得的病毒基因序列建立系统发育树（图 2；病人的病毒基因名字以‘Human/…’ 开头，受感染的鸡的病毒基因名字以‘Chicken/…’开头）。请你的团队根据系统发育树进行分析，并在报告中回答下列问题：

a. 该新病原体在人类群体中传播的进化速度是多少（即以每年的突变为单位的突变率是多少）？

b. 该新病原体何时开始在人类中爆发？比首个报告的病例早多少天？

c. 这个病毒病原体可能是什么？为什么？（开放式问题）

d. 根据所有数据和您的发现，您对预防未来类似疫情的发生有何建议？（开放式问题）

参考文献：

[1] Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, & Pei Y-Y. (2020). A new coronavirus associated with human respiratory disease in China. Nature, 579(7798), 265-269. (https://www.nature.com/articles/s41586-020-2008-3)

[2] Lam TT-Y, Wang J, Shen Y, Zhou B, Duan L, Cheung C-L, Ma C, Lycett SJ, Leung CY-H, Chen X, Li L, et al. (2013). The genesis and source of the H7N9 influenza viruses causing human infections in China. Nature 502:241-244. (https://pubmed.ncbi.nlm.nih.gov/23965623/)

[3] Smith G J, Vijaykrishna D, Bahl J, Lycett S J, Worobey M, Pybus O G, Ma S K, Cheung C L, Raghwani J, Bhatt S, et al. (2009). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature, 459(7250), 1122-1125. (https://www.nature.com/articles/nature08182)

[4] Zeng F, Chow KY, Leung FC. (2003). Estimated timing of the last common ancestor of the SARS coronavirus. N Engl J Med 349:2469-70. (https://www.nejm.org/doi/full/10.1056/NEJM200312183492523)

[5] de Oliveira T, Pybus OG, Rambaut A, Salemi M, Cassol S, Ciccozzi M, Rezza G, Gattinara GC, D'Arrigo R, Amicosante M, Perrin L, Colizzi V, Perno CF, Benghazi Study G. (2006). Molecular epidemiology: HIV-1 and HCV sequences from Libyan outbreak. Nature 444:836-7. (https://www.nature.com/articles/444836a.pdf)

[6] Siebenga JJ, Beersma MF, Vennema H, van Biezen P, Hartwig NJ, Koopmans M. (2008). High prevalence of prolonged norovirus shedding and illness among hospitalized patients: a model for in vivo molecular evolution. J Infect Dis 198:994-1001.. (https://academic.oup.com/jid/article/198/7/994/2192243)

提交

你团队的解决方案论文应包括 1 页的摘要和 1 篇科普短文。正文不能超过 20 页，含摘要及短文最多 23 页。附录和参考资料应出现在正文之后，不算在 23 页的限制之内。

調查新發病原體 X

根據世界衛生組織及其他國際專家的預測，未來幾乎必定會出現一種新型病原體（統稱為“病原體 X”），並再次引起全球性的疾病暴發。因此，必須從這次 COVID-19 大流行中吸取教訓，以更好地應對未來的病原體傳播。病原體的傳播會在基因組序列上留下蹤跡，因此分析病原體基因組數據對研究疾病傳播發揮著重要作用。例如，基因組序列的系統發育樹分析可以發現新興病原體與任何已知病原體的關聯，以揭示感染人類的動物來源[參考資料 1，2，3]。系統發育樹還可以用於追蹤疾病出現的時間[參考資料 4，5]，其原理在於，許多病原體，尤其是快速進化的 RNA 病毒會隨著時間以某種速度發生突變（不同類型的病原體可能有不同的突變率）[參考資料 6]，突變越多則說明兩個病原體序列之間的進化關係越遠。這種方法常用於建立病原體的進化模型，溯源其傳播歷史。

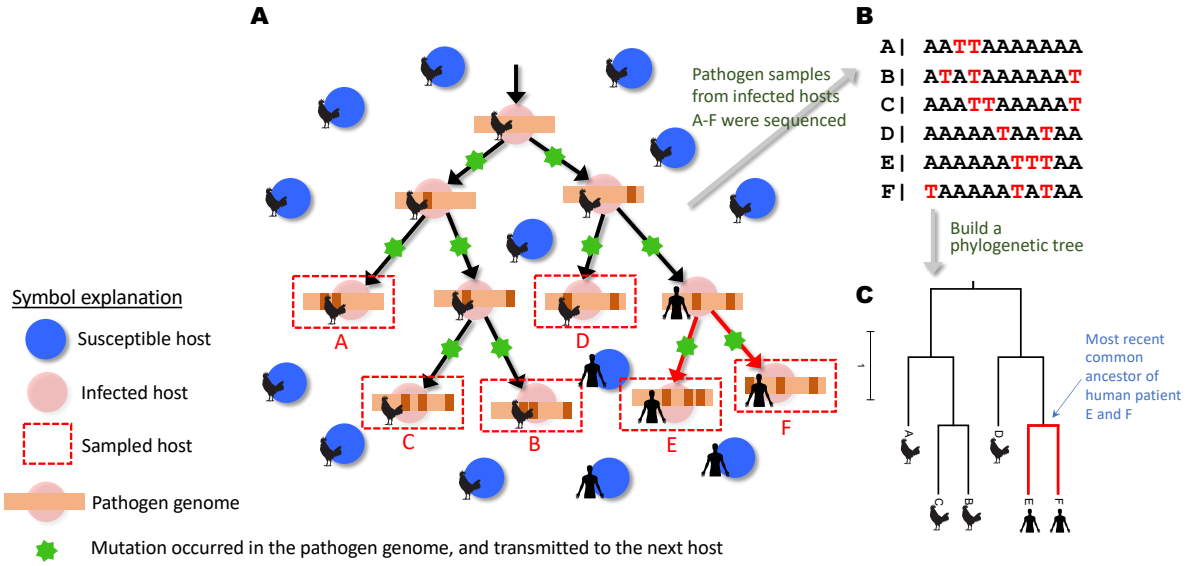

圖1. 分析病原體基因組中疾病傳播蹤跡的概念示意圖

（A）疾病在雞體內傳播（黑色箭頭），隨後傳播給 3 個人類患者。當病原體感染雞或人類宿主時，其基因組可能發生突變（綠色星號和矩形中的深色位點），突變的病原體可能傳播給下一個宿主。在這個例子中，我們對來自部分宿主（4 隻雞和 2 個人）的病原體樣本進行測序。（B）從宿主 A-F（如圖）獲得的病原體基因組序列。（C）根據基因組序列建立的系統發育樹，其中樹的分支長度反映了發生突變的數量，分支共同歸屬的節點是假定祖先。例如，指向 E 和 F 的節點稱為 E 和 F 的「最近共同祖先」。它是人類傳播集群中的零號病人（又名首例病人、索引病人，一般很困難發現並測序）。左側標尺表示序列中單個突變的支長。

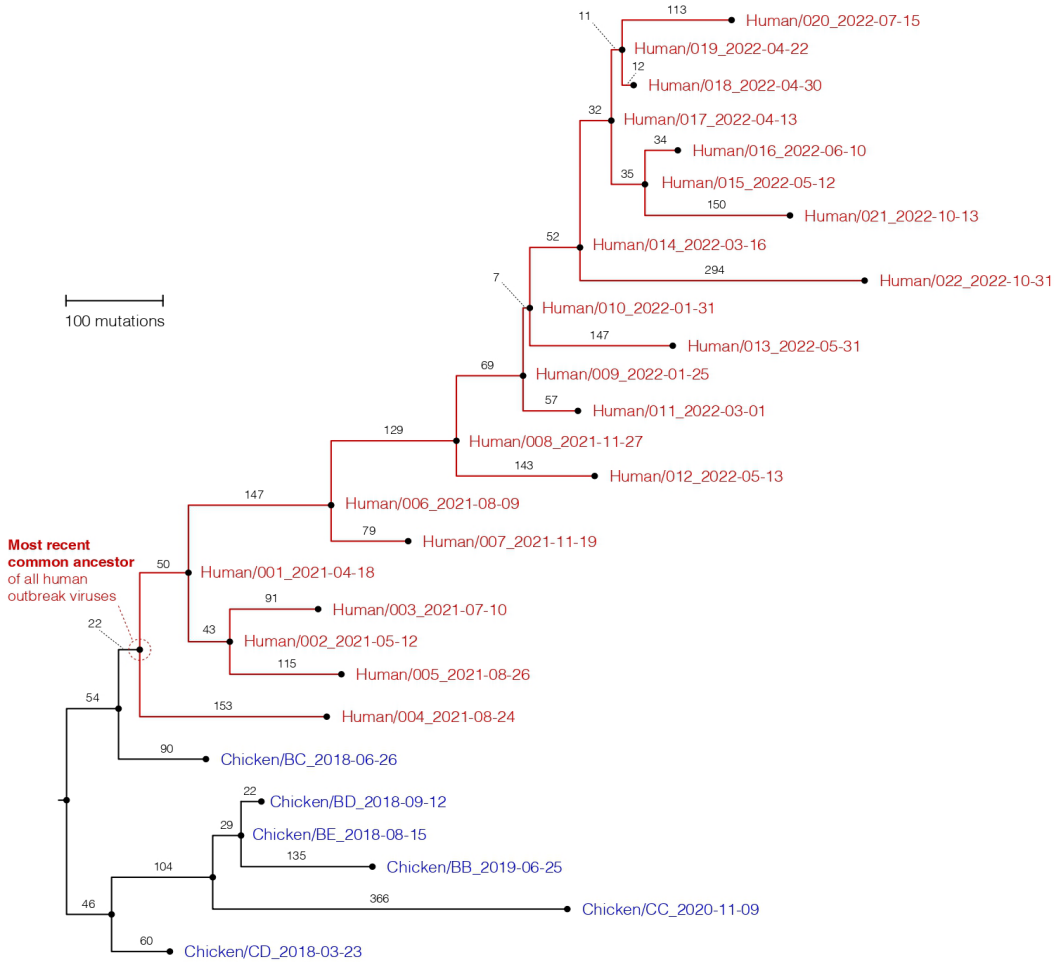

圖 2. 由人類患者（紅色）和受感染雞（藍色）的病毒序列建立的系統發育樹

任務

假設近期爆發了一場原因不明的呼吸道疾病。來自病人的病原體很快被測序，並發現與以前在雞身上發現的一種病毒基因相近。使用這次爆發的 22 名病人和受感染的雞身上獲得的病毒基因序列建立系統發育樹（圖 2 病人的病毒基因名字以『Human/...』 開頭，受感染的雞的病毒基因名字以 'Chicken/...'）開頭。請你的團隊根據系統發育樹進行分析，並在報告中回答下列問題：

a. 該新病原體在人類群體中傳播的進化速度是多少（即以每年的突變為單位的突變率是多少）？

b. 該新病原體何時開始在人類中爆發？比首個報告的病例早多少天？

c. 這個病毒病原體可能是什麼？為什麼？ （開放式問題）

d. 根據所有數據和您的發現，您對預防未來類似疫情的發生有何建議？（開放式問題）

參考文獻：

[1] Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, & Pei Y-Y. (2020). A new coronavirus associated with human respiratory disease in China. Nature, 579(7798), 265-269. (https://www.nature.com/articles/s41586-020-2008-3)

[2] Lam TT-Y, Wang J, Shen Y, Zhou B, Duan L, Cheung C-L, Ma C, Lycett SJ, Leung CY-H, Chen X, Li L, et al. (2013). The genesis and source of the H7N9 influenza viruses causing human infections in China. Nature 502:241-244. (https://pubmed.ncbi.nlm.nih.gov/23965623/)

[3] Smith G J, Vijaykrishna D, Bahl J, Lycett S J, Worobey M, Pybus O G, Ma S K, Cheung C L, Raghwani J, Bhatt S, et al. (2009). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature, 459(7250), 1122-1125. (https://www.nature.com/articles/nature08182)

[4] Zeng F, Chow KY, Leung FC. (2003). Estimated timing of the last common ancestor of the SARS coronavirus. N Engl J Med 349:2469-70. (https://www.nejm.org/doi/full/10.1056/NEJM200312183492523)

[5] de Oliveira T, Pybus OG, Rambaut A, Salemi M, Cassol S, Ciccozzi M, Rezza G, Gattinara GC, D'Arrigo R, Amicosante M, Perrin L, Colizzi V, Perno CF, Benghazi Study G. (2006). Molecular epidemiology: HIV-1 and HCV sequences from Libyan outbreak. Nature 444:836-7. (https://www.nature.com/articles/444836a.pdf)

[6] Siebenga JJ, Beersma MF, Vennema H, van Biezen P, Hartwig NJ, Koopmans M. (2008). High prevalence of prolonged norovirus shedding and illness among hospitalized patients: a model for in vivo molecular evolution. J Infect Dis 198:994-1001.. (https://academic.oup.com/jid/article/198/7/994/2192243)

提交

你團隊的解決方案論文應包括 1 頁的摘要和 1 篇科普短文。正文不能超過 20 頁，含摘要及短文最多 23 頁。附錄和參考資料應出現在正文之後，不算在 23 頁的限制之內。