Viruses causing acute respiratory infections transmit via tight population bottlenecks

Lei Zhao¹, Rodney S. Daniels², John W. McCauley², C. J. R. Illingworth^{3,4,*}

¹Section for Geogenetics, Globe Institute, University of Copenhagen, Øster Voldgade 5-7, 1350 København K, Denmark

²WHO Collaborating Centre for Reference and Research on Influenza, Crick Worldwide Influenza Centre, The Francis Crick Institute, United Kingdom

³MRC Biostatistics Unit, University of Cambridge, East Forvie Building, Forvie Site, Robinson Way, Cambridge, CB2 0SR, United Kingdom

⁴Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK

*Corresponding author: cjri2@cam.ac.uk

Abstract: Recent analyses of virus transmission have used genome sequence data to infer greatly contrasting transmission bottlenecks, with estimates differing by orders of magnitude. We here describe two simple mathematical models, the first considering transmission bottlenecks within an epidemiological framework, and the second evaluating the effect of particle diffusion on bottleneck sizes in a scenario where large numbers of particles could be transmitted. Our models suggest that, barring specific circumstances, tight population bottlenecks, involving few viral particles, are intrinsically likely in the spread of viruses causing respiratory infection. Our study provides a simple prior for studies of virus transmission, potentially grounding future research in this area.

The bottlenecks experienced by viruses during transmission between hosts are a key component of virus lifecycles. The number of particles transmitted between hosts strongly influences virus evolution; the tighter the bottleneck, the less genetic diversity is transmitted between individuals. To establish a new infection, a virus must overcome a series of bottlenecks; viruses must first be emitted by the infected host, then gain exposure to a novel host, then finally must get into a cell within that host so as to replicate and produce viable offspring (Figure 1A).

30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56

Studies of transmission bottlenecks for respiratory viruses have a somewhat mixed history. Household studies of influenza transmission in Hong Kong and Michigan inferred contrasting population bottlenecks of 200-250 and 1-2 virus genomes^{1,2}, this disparity being reconciled by the discovery of incorrectly assigned paired reads in the former dataset³. More recently, studies of SARS-CoV-2 transmission have presented similarly contrasting results, suggesting mean bottleneck sizes in the order of 10^3 and of between 1 and 8 virus genomes^{4,5}. Further studies of SARS-CoV-2⁶ and the context of previous studies of influenza transmission^{7,8} each favour the smaller of these values as being correct, while a reanalysis suggests that incorrect data processing may again have led to an inflated value⁹. However, the repeated publication of widely disparate bottleneck estimates suggests a lack of common understanding around this matter.

We here suggest that, independent of any genomic data, there are strong epidemiological and physical arguments suggesting that respiratory virus infections are likely to have a tight bottleneck at the point of transmission. While these arguments in no way preclude further study of respiratory viral transmission, we believe that they set a prior expectation for future research in this area.

In Figure 1B we show a simple diagram representing transmission. During the course of an infection, an infected individual (I) sheds viruses and makes contact with a series of susceptible individuals, which we denote S_1 to S_k . Adopting an epidemiological framework, a mean of R individuals will be infected, where R is the reproduction number of the virus. This fact alone tells us something about the population bottleneck at transmission. Where we denote by n_i the number of particles infecting the individual S_i , we have that, on average, n_i will be zero in $k-R$ cases, and greater than zero in the remaining R cases.

We next consider values of R for respiratory viruses. Estimated values of the basic reproductive number R_0 for influenza and SARS-CoV-2 are around 1.5 and 2.9 respectively^{10,11}, sufficient to cause rapid spread, but substantially lower than the equivalent value for measles¹². These sufficiently large but not extreme values of R suggest that, while these diseases are highly transmissible, it remains the case that on average most people in contact with an infected person do not themselves get infected. Putting this another way, among the contacts of a typical infected individual, the most common value of n_i is equal to zero.

Where transmission bottlenecks are considered in the literature, what is usually measured is the size of the bottleneck conditional on transmission having occurred; we now consider the distribution of the size of this statistic. If all exposure events were identical, then under the assumption that viruses behave independently of one another, we would expect a tight transmission bottleneck: With no viruses getting through in most cases, only a few viruses would be expected to establish infection where transmission occurred. However, exposure events are not likely to be identical. Under our assumption of independence, the number of viruses to which an individual S is exposed is likely to be proportional to the number of viruses emitted by I , while the number of viruses establishing an infection in S is likely to be proportional to the number to which S was exposed. Healthy human individuals differ substantially in the extent of aerosol that they exhale¹³, while symptoms such as coughing substantially increase aerosol emission¹⁴. Animal experiments show that different forms of exposure, such as physical contact or airborne transmission, can lead to differing numbers of viruses being transmitted¹⁵. We therefore consider the effect of differing emission and exposure upon virus transmission bottlenecks.

A simple mathematical model of exposure to airborne virus particles suggests that, with variable levels of emission and exposure, events involving the transmission of large numbers of viruses are theoretically possible. However, circumstances which facilitate such large

bottlenecks are likely to cause, on average, a much greater number of much tighter bottlenecks. Noting that the highest levels of exposure to virus particles are likely to arise in the most extreme events of viral shedding, we modelled the progress of particles shed from a sneezing event. Sneezing rapidly creates a cloud of particles just under one metre across¹⁶, which subsequently spread via diffusion, being affected by gravity and local patterns of air flow.

Figures 1C and 1D show a simple model of this process, considering diffusion in isolation. Beginning with a cloud of virus particles, we simulated the diffusion of these particles through three-dimensional space. In our model, increasing distance from the centre of the initial cloud led to rapidly decreasing exposure to virus particles; maximum levels of exposure at distances of one or two metres from the centre of the cloud were, respectively, only 5.1% and 0.6% of the maximum.

In a model of simulated virus transmission, we deliberately generated loose transmission bottlenecks, involving large numbers of viruses; we achieved this by assuming that the concentration of viruses at the centre of the cloud was sufficient to cause an expected transmission bottleneck of 100 virus particles. We next simulated transmission events at locations around the cloud centre, in which the expected transmission bottleneck was proportional to the level of exposure, transmission occurring if at least one virus was transmitted. Considering potential susceptible individuals distributed evenly around the centre of the initial cloud, more than 90% of cases of transmission involved a bottleneck of 10 viruses or fewer, with most infections being founded by a single virus (Figure 1E). This result held across a range of assumptions about virus emission (Figure 1S1). While our model is intentionally simplistic it suggests that, even where emission and exposure levels vary, and in circumstances where transmission has the potential to initiate infections with large numbers of viruses, an emission event is likely to establish a far greater number of infections with tight transmission bottlenecks.

We note potential circumstances in which our conclusion would not apply. Firstly, if viruses do not act independently during transmission, but in a co-operative manner, a bimodal distribution of transmission bottlenecks may be possible. Cooperativity between viruses has been observed in bacteriophage populations and in the abrogation of host cell mechanisms^{16,17}. If multiple viruses were to cooperate to invade a novel host, this would increase the proportion of cases in which multiple viruses founded infection. Multipartite viruses might need high MOIs to establish infection¹⁸. Secondly, if a virus was exceptionally transmissible, so that essentially all of the contacts of a host were infected, the mean transmission event could involve a large number of particles; this would require the virus to have a very high reproduction number (R). Thirdly, if the half-life of viruses in air was extremely short, preventing their spread via diffusion, direct exposure to the high concentrations of viruses emitted in a cough or a sneeze might be necessary for transmission, susceptible individuals receiving either a large exposure to the virus or no exposure at all. These circumstances do not, to the best of our knowledge, apply to most agents causing respiratory infections. Cooperativity at the point of infection is generally unlikely, with distinct viruses infecting separate subpopulations of cells within the airway¹⁹. Respiratory viruses may be highly infectious, yet it remains that most individuals in contact with an infected host do not get infected. Further, viruses causing respiratory diseases have some degree of stability in aerosol suspension, with for example numbers of viable SARS-CoV-2 virions in aerosol showing only a six-fold decrease over a period of three hours²⁰. We therefore conclude that most viruses causing acute respiratory infections are likely to transmit via narrow population bottlenecks, with few viruses being transmitted. Our result provides a prior model for future studies of bottlenecks occurring during in the transmission of viruses causing respiratory disease.

Methods

141 Diffusion model

142

143 Calculation of our diffusion model was calculated using the Mathematica software package.

144 We defined a function $V(\mathbf{x}, t)$ describing the concentration of viruses in three dimensional

145 space $\mathbf{x}=\{x, y, z\}$, and calculated a solution to the equation:

146

$$147 \quad \frac{\partial V(x, t)}{\partial t} = \nabla^2 V(x, t)$$

148

149 with the initial condition

150

$$151 \quad V(x, 0) = e^{-4(x^2 + y^2 + z^2)}$$

152

153 This approximates the initial existence of a sphere of viruses centred on the origin at time

154 zero. We note that this equation has an analytical solution

155

$$156 \quad V(x, t) = (16t + 1)^{-3/2} e^{-\frac{4(x^2 + y^2 + z^2)}{16t + 1}},$$

157

158 and use this a simplified model of the diffusion of virus particles following an initial dispersal.

159

160 In order to calculate the maximum exposure at a given position vector $\mathbf{x}=\{x, y, z\}$, we

161 calculated the time of maximum exposure

162

$$163 \quad T_{max}(x) = \max \left\{ 0, \frac{8(x^2 + y^2 + z^2) - 3}{48} \right\}$$

164

165 In order to simulate transmission, we built a model encompassing variation in the extent of

166 viral emission, and differences in exposure arising from the diffusion model. The extent of

emission was calculated as a series of random draws from a gamma distribution with mean one, multiplied by 100:

$$e_i \sim \frac{100x^{-a-1}a^ae^{-ax}}{\Gamma(a)}$$

Here, varying the parameter a alters the variance of the distribution, allowing for the modelling of a greater or lesser amount of variance in particle emission; results shown in the main text were calculated with $a=1$, with a broader range of values of a being further considered.

We next defined a set of points S at radial distances r from the centre of the cloud of viruses, with r spaced at intervals of 1mm between 1mm and 10m, spaced at equal intervals of length 2π mm around concentric circles in the $z=0$ plane. At each point s in S we calculated the maximum exposure over time

$$x_i = e_i V(s, T_{max}(S))$$

We next simulated transmission bottlenecks using a simple Poisson model

$$n_i \sim \text{Poisson}(x_i)$$

Statistics of transmission bottlenecks were then calculated across cases in which n_i was non-zero.

Acknowledgements

CJRI acknowledges funding from the Medical Research Council (ref: MC_UU_00002/11).

194   **References**

- 195   1.   Poon, L. L. M. *et al.* Quantifying influenza virus diversity and transmission in humans.
196   *Nat. Genet.* **48**, 195–200 (2016).
- 197   2.   McCrone, J. T. *et al.* Stochastic processes constrain the within and between host
198   evolution of influenza virus. *Elife* **7**, (2018).
- 199   3.   Xue, K. S. & Bloom, J. D. Reconciling disparate estimates of viral genetic diversity
200   during human influenza infections. *Nature genetics* vol. 51 1298–1301 (2019).
- 201   4.   Popa, A. *et al.* Genomic epidemiology of superspreading events in Austria reveals
202   mutational dynamics and transmission properties of SARS-CoV-2. *Sci. Transl. Med.* **12**,
203   (2020).
- 204   5.   Lythgoe, K. A. *et al.* SARS-CoV-2 within-host diversity and transmission. *Science* (2021)
205   doi:10.1126/science.abg0821.
- 206   6.   Braun, K. M. *et al.* Transmission of SARS-CoV-2 in domestic cats imposes a narrow
207   bottleneck. *PLoS Pathog.* **17**, e1009373 (2021).
- 208   7.   Lumby, C. K., Nene, N. R. & Illingworth, C. J. R. A novel framework for inferring
209   parameters of transmission from viral sequence data. *PLoS Genet.* **14**, e1007718
210   (2018).
- 211   8.   Valesano, A. L. *et al.* Influenza B Viruses Exhibit Lower Within-Host Diversity than
212   Influenza A Viruses in Human Hosts. *J. Virol.* **94**, (2020).
- 213   9.   Martin, M. A. & Koelle, K. Reanalysis of deep-sequencing data from Austria points
214   towards a small SARS-COV-2 transmission bottleneck on the order of one to three
215   virions. doi:10.1101/2021.02.22.432096.
- 216   10. Biggerstaff, M., Cauchemez, S., Reed, C., Gambhir, M. & Finelli, L. Estimates of the
217   reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic
218   review of the literature. *BMC Infect. Dis.* **14**, 480 (2014).
- 219   11. Billah, M. A., Miah, M. M. & Khan, M. N. Reproductive number of coronavirus: A
220   systematic review and meta-analysis based on global level evidence. *PLoS One* **15**,

221 e0242128 (2020).

222 12. Guerra, F. M. *et al.* The basic reproduction number (R) of measles: a systematic review.
 223 *Lancet Infect. Dis.* **17**, e420–e428 (2017).

224 13. Edwards, D. A. *et al.* Exhaled aerosol increases with COVID-19 infection, age, and
 225 obesity. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

226 14. Hamilton, F. *et al.* Aerosol emission from the respiratory tract: an analysis of relative
 227 risks from oxygen delivery systems. doi:10.1101/2021.01.29.21250552.

228 15. Varble, A. *et al.* Influenza A virus transmission bottlenecks are defined by infection route
 229 and recipient host. *Cell Host Microbe* **16**, 691–700 (2014).

230 16. Nishimura, H., Sakata, S. & Kaga, A. A new methodology for studying dynamics of
 231 aerosol particles in sneeze and cough using a digital high-vision, high-speed video
 232 system and vector analyses. *PLoS One* **8**, e80244 (2013).

233 17. Turner, P. E. & Chao, L. Prisoner's dilemma in an RNA virus. *Nature* vol. 398 441–443
 234 (1999).

235 18. Nee, S. The evolution of multicompartmental genomes in viruses. *J. Mol. Evol.* **25**, 277–
 236 281 (1987).

237 19. Sobel Leonard, A. *et al.* The effective rate of influenza reassortment is limited during
 238 human infection. *PLoS Pathog.* **13**, e1006203 (2017).

239 20. van Doremalen, N. *et al.* Aerosol and Surface Stability of SARS-CoV-2 as Compared
 240 with SARS-CoV-1. *N. Engl. J. Med.* **382**, 1564–1567 (2020).

242

243 **Figure 1: A model of viral transmission.** **A.** Stages in transmission. An infected individual

244 emits e viruses. A recipient is exposed to x viruses, of which n found infection. **B.** The

245 infected individual I exposes k individuals S_1 to S_k to the virus, infecting an average of R

246 individuals. In the individual S_i , the transmission bottleneck means that n_i viruses found

247 infection. For all but R individuals S_i , we thus have that $n_i=0$. **C.** Simplified model of the

248 diffusion of viruses following emission via a sneeze. The relative viral concentration drops

249 rapidly as the viral particles spread. Time here is shown in arbitrary units. **D.** Changes in

250 relative viral concentration at different distances from the centre of the initial cloud. An

251 initially high viral concentration rapidly falls to much lower, but more stable concentrations.

252 **E.** Transmission bottlenecks occurring in a simple model of transmission via particle diffusion

253 in which the mean peak exposure is sufficient to cause a transmission bottleneck of 100.

254

Supporting Figures

Figure 1S1: Effect of variance in viral emissions upon results obtained. A. Relative distributions of viral emissions by infected individuals were represented by a series of gamma distributions, here shown. **B.** Proportions of transmission events with $n_i \leq 10$, with $n_i \leq 5$, and with $n_i = 1$ were broadly consistent across a range of distributions of viral emissions.