



**HAL**  
open science

## Superspreading genomes

Samuel Alizon

► **To cite this version:**

Samuel Alizon. Superspreading genomes. *Science*, 2021, 371 (6529), pp.574-575. 10.1126/science.abg0100 . hal-03133445

**HAL Id: hal-03133445**

**<https://hal.science/hal-03133445>**

Submitted on 11 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Superspreading genomes

Samuel Alizon

Maladies Infectieuses et Vecteurs : Écologie, Génétique, Évolution Contrôlée  
(UMR CNRS 5290, UR IRD 224, UM), Montpellier, France

E-mail: samuel.alizon@cirs.fr

Alizon, S (2021) Superspreading genomes. *Science*, 371(6529):574-575. doi:10.1126/science.abg0100

## **Virus mutations help to unravel key epidemiological events**

Individual contributions to epidemic spread vary. Although some infections may not cause any secondary cases, others are associated with so-called "superspreading" events in which numerous infections result from the same case. These events can shape the course of an epidemic, but their detection remains challenging. On page 588 of this issue, Lemieux *et al.* (1) show that phylogenetic analyses of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genome sequence data help quantify the prevalence and impact of superspreading events on COVID-19 outbreaks.

Superspreading gained attention during the 2002–2004 SARS epidemic, and mathematical models highlighted its contrasting effects, by which the phenomenon is predicted to increase the probability that an outbreak will go extinct by chance but also to fuel the growth of outbreaks that evade extinction (2). Superspreading could also be involved in the adaptation of emerging infectious diseases to new hosts (3). Such events have since been identified in measles (2), Middle East respiratory syndrome (4), or ebola (5) outbreaks. Their presence can be detected from cluster sizes (4) or spatial incidence data (5), but they appear most strikingly in contact tracing data. However, studies to generate these data are expensive, invasive, and time consuming. Their quality is also limited because many people are unreliable respondents and list no or multiple potential sources of infection. Digital tracking could increase data quality but comes with substantial privacy risks. Most of these limitations are minimal for virus genome analyses. A common motivation to monitor the diversity of circulating viral strains is that some genetic variations may threaten treatment or long-term vaccine efficiencies (6). But genetic evolution can also be harnessed to track infections as they spread, informing public health decisions. Assuming a constant mutation rate, viruses originating from infections that are close in the transmission chain should be more alike, from a genetic standpoint, than viruses from infections that are temporally or geographically distant. Using sequence data, it is possible to infer phylogenetic trees, which bear many similarities with dated genealogies of infections.

Phylogenetic analyses readily provide insights regarding epidemic structure, including the presence of superspreading events, but also temporal aspects such as the date of origin of an epidemic wave (7) (see the figure). The Ebola virus (EBOV) epidemic in West Africa in 2013-2016 marked a shift in the production and sharing of virus genomic sequence data (8). This allowed monitoring of a potentially important role of superspreading events (9), which was later confirmed with spatial incidence data (5). The large number of EBOV genome sequences that were published during that time now seems limited. Owing to technological progress, rapid and affordable full genome sequencing protocols (10), and the involvement of many teams across the world, more than 300,000 open-licensed SARS-CoV-2 genomes were shared in 2020 (11). The impressive number of SARS-CoV-2 genomes should not mask the strong sampling heterogeneity. Half of the genomes are from the United Kingdom. Within the United States, more than half originate from only five states (California, Texas, Washington, Michigan, and New York). Timing also matters. For example, that the oldest SARS-CoV-2 genome dates from December 2019 limits our ability to accurately estimate the date of onset of the pandemic. The study by Lemieux *et al.* stands out because it uses dense and early sampling of the local epidemic in Boston, Massachusetts. This offers a detailed view of the structure and history of the epidemic in the city area. The study also confirms and analyzes two contrasting superspreading events. One of these occurred in a skilled nursing facility and had a substantial impact locally in terms of mortality. The second event was associated with a business conference. Thanks to their understanding of the local epidemics, the Global Initiative on Sharing All Influenza Data (GISAID) collaborative effort, and two key mutations in the virus genome, Lemieux *et al.* show that this event generated transmission chains all over the world that may account for hundreds of thousands of infections.

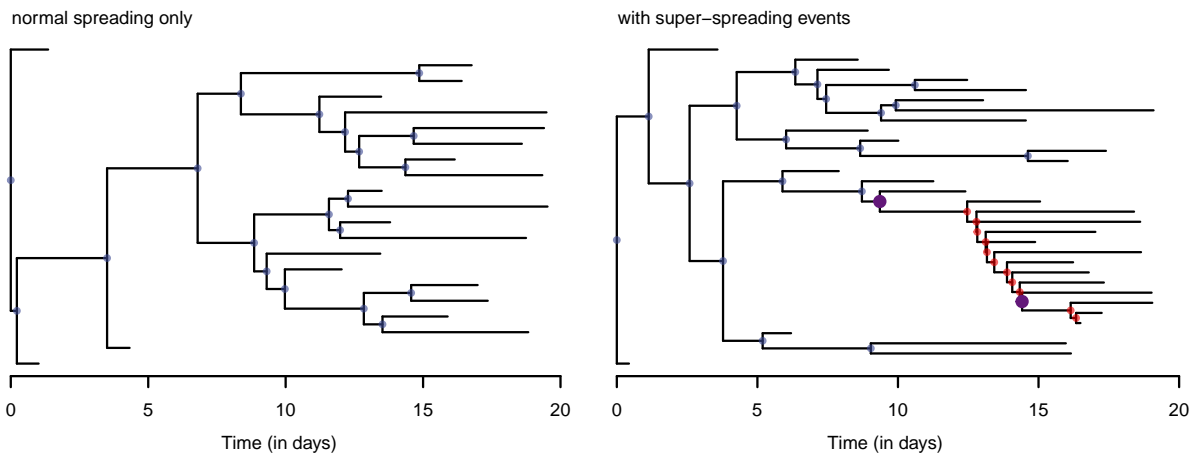
Phylogenetic analyses may interpret any type of heterogeneity not included in the underlying model, especially uneven sampling or spatial structure, as superspreading (7). A promising research avenue to address this issue is called "data integration." It hypothesizes that combining different types of data—for example, genome sequences and incidence data (12)—should contain more information. This could improve the detection of individual heterogeneity in transmission (13). It could also help answer a longstanding question: Are superspreading events due to individual biological (such as virus load) or behavioral (such as contact rate) properties, or to the environment (such as poorly ventilated meeting rooms)? This is illustrated in two ways in the study by Lemieux *et al.* They use their knowledge of the sampling location to distinguish superspreading events from clusters generated by uneven sampling. Also, they show that two virus introductions in the skilled nursing facility had opposite trajectories, with one leading to a handful of secondary cases and the other to a massive outbreak.

The unfolding of an epidemic depends on the underlying causes of superspreading events. If transmissibility is correlated with susceptibility and if natural immunity after recovery is strong, the effect of superspreaders should decrease as the epidemic unfolds, decreasing herd immunity thresholds (14). In terms of control, identifying risk factors associated with superspreading events—whether they relate to individuals, activities, or locations—opens the possibility for targeted interventions that can disproportionately hamper epidemic spread (2, 15).

The evolutionary rate of SARS-CoV-2 has, so far, been relatively slow, making multiple sources of data particularly complementary. The ease with which virus genomes can now be generated and the importance of monitoring virus evolution are opportunities for phylodynamics to become a routine tool in outbreak management.

## References and Notes

1. J. E. Lemieux, *et al.*, *Science* (2020).
2. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, *Nature* **438**, 355 (2005).
3. A. Yates, R. Antia, R. R. Regoes, *Proc. R. Soc. Lond. B* **273**, 3075 (2006).
4. A. J. Kucharski, C. L. Althaus, *Eurosurveillance* **20**, 21167 (2015).
5. M. S. Y. Lau, *et al.*, *Proc. Nat. Acad. Sci. USA* **114**, 2337 (2017).
6. D. A. Kennedy, A. F. Read, *PLoS Biol.* **18**, e3001000 (2020).
7. E. M. Volz, K. Koelle, T. Bedford, *PLoS Comput Biol* **9**, e1002947 (2013).
8. S. K. Gire, *et al.*, *Science* **345**, 1369 (2014).
9. T. Stadler, D. Kühnert, D. A. Rasmussen, L. d. Plessis, *PLoS Curr.* (2014).
10. N. E. Freed, M. Vlkov, M. B. Faisal, O. K. Silander, *Biol. Methods Protoc.* **5** (2020).
11. GISAID - Initiative. URL: <https://www.gisaid.org/>.
12. X. Du, A. A. King, R. J. Woods, M. Pascual, *Sci. Transl. Med.* **9** (2017).
13. L. M. Li, N. C. Grassly, C. Fraser, *Mol. Biol. Evol.* **34**, 2982 (2017).
14. T. Britton, F. Ball, P. Trapman, *Science* (2020).
15. M. E. J. Woolhouse, *et al.*, *Proc. Natl. Acad. Sci. USA* **94**, 338 (1997).
16. I thank CNRS, Institut pour la Recherche et le Dveloppement, and Rgion Occitanie for financial support and T. Kamiya, M. T. Sofonea, M. van Baalen, and the Experimental and Theoretical Evolution team for discussion.



**Figure 1: The effect of superspreading events.** Virus genome sequences are arranged in a phylogeny to track transmission over time. Each node corresponds to a transmission event ("birth"), and each tip corresponds to the end of an infection ("death"). Superspreading generates many secondary infections that cluster in the phylogenetic tree. In epidemics with superspreading, many infections do not cause any secondary case. Red dots are secondary infections caused superspreading events (in purple).