

# An open-access database of infectious disease transmission trees to explore superspreader epidemiology

Juliana C. Taube<sup>1</sup>, Paige B. Miller<sup>2</sup>, and John M. Drake<sup>2,3</sup>

<sup>1</sup> Bowdoin College, Brunswick, ME;

<sup>2</sup> Odum School of Ecology, University of Georgia, Athens, GA;

<sup>3</sup> Center for the Ecology of Infectious Diseases, University of Georgia, Athens, GA

The database is available at <http://outbreaktrees.ecology.uga.edu/>

✉ [taubejc@gmail.com](mailto:taubejc@gmail.com)

🏠 [jtaube.github.io](https://github.com/jtaube)

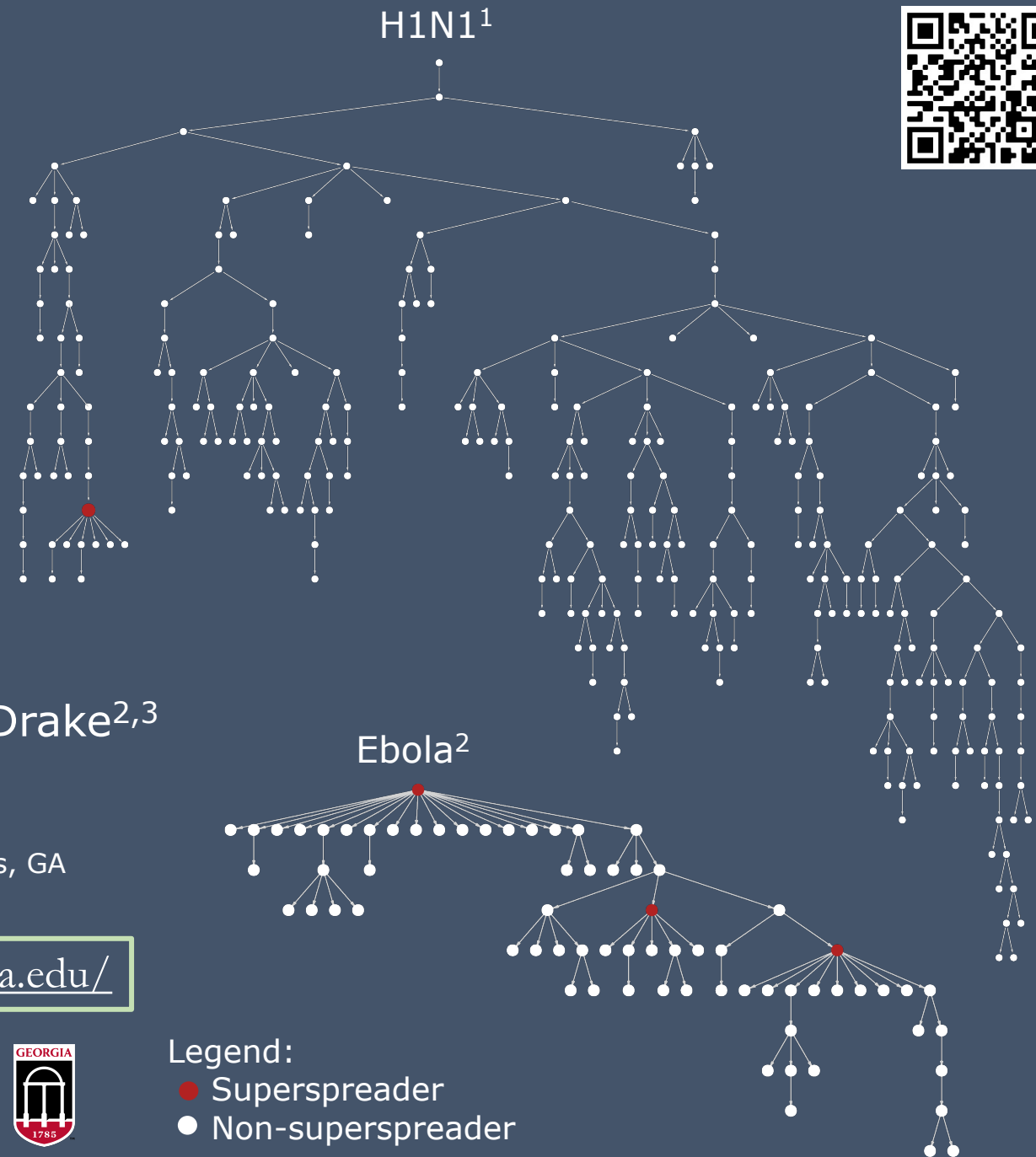
🐦 [@julianactaube](https://twitter.com/julianactaube)



Legend:

● Superspreader

● Non-superspreader



# OVERVIEW



## Background

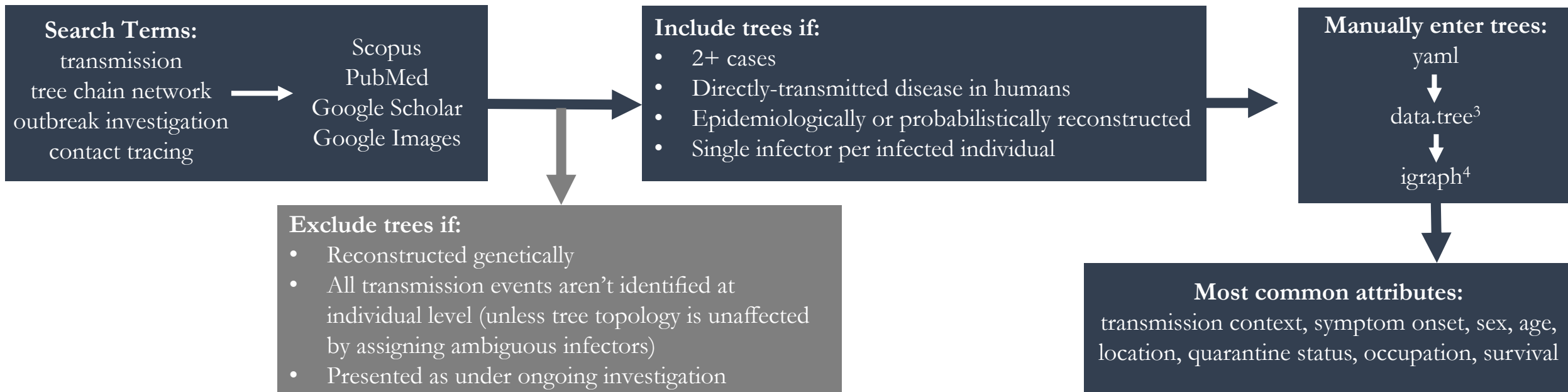
- Often outbreak investigations aim to identify **who infected whom** by reconstructing the outbreak **transmission tree**, which visualizes transmission between individuals as a network with nodes (individuals) and branches (transmission from person to person).
- These investigations are costly but highly valuable because transmission trees are information rich, including details about the settings of transmission and variation in number of secondary infections.
- Yet, when published, transmission trees are shown and described in a variety of formats that makes them difficult to compare across outbreaks.

## Summary

- We compiled a database of 383 published, standardized transmission trees consisting of 16 directly-transmitted diseases ranging in size from 2 to 286 cases.
- We demonstrated the potential utility of the database through short analyses addressing questions about superspreader epidemiology.
  - We compared the frequency and contribution of superspreaders to onward transmission across diseases.
  - We investigated patterns in how superspreaders are infected.

**Our goal: Create an accessible, easily analyzable, standardized database of transmission trees**

## DATABASE CONSTRUCTION



# DATABASE SUMMARY



Figure 1 (right). **Characteristics of transmission trees in OutbreakTrees.**

(A) Tree size varies from 2 to 286 with a median of 3 and most trees represent outbreaks taking place in the past 20 years (only trees with 10 or more cases shown in date plot). (B) The largest trees are from H1N1 and SARS outbreaks while the highest proportion of trees in the database are from outbreaks of COVID-19, followed by adenovirus and Ebola. Tree size axes in both plots are shown on a  $\log_{10}$  scale to better illustrate variation in medium-sized trees.

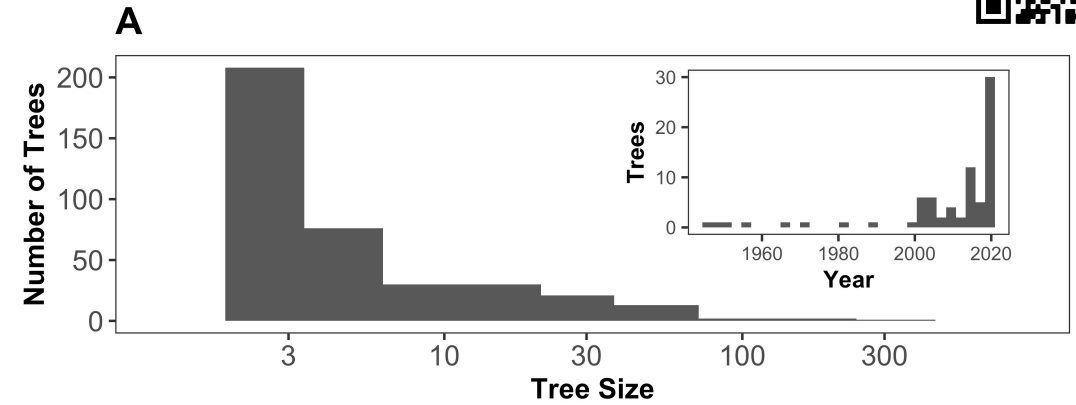
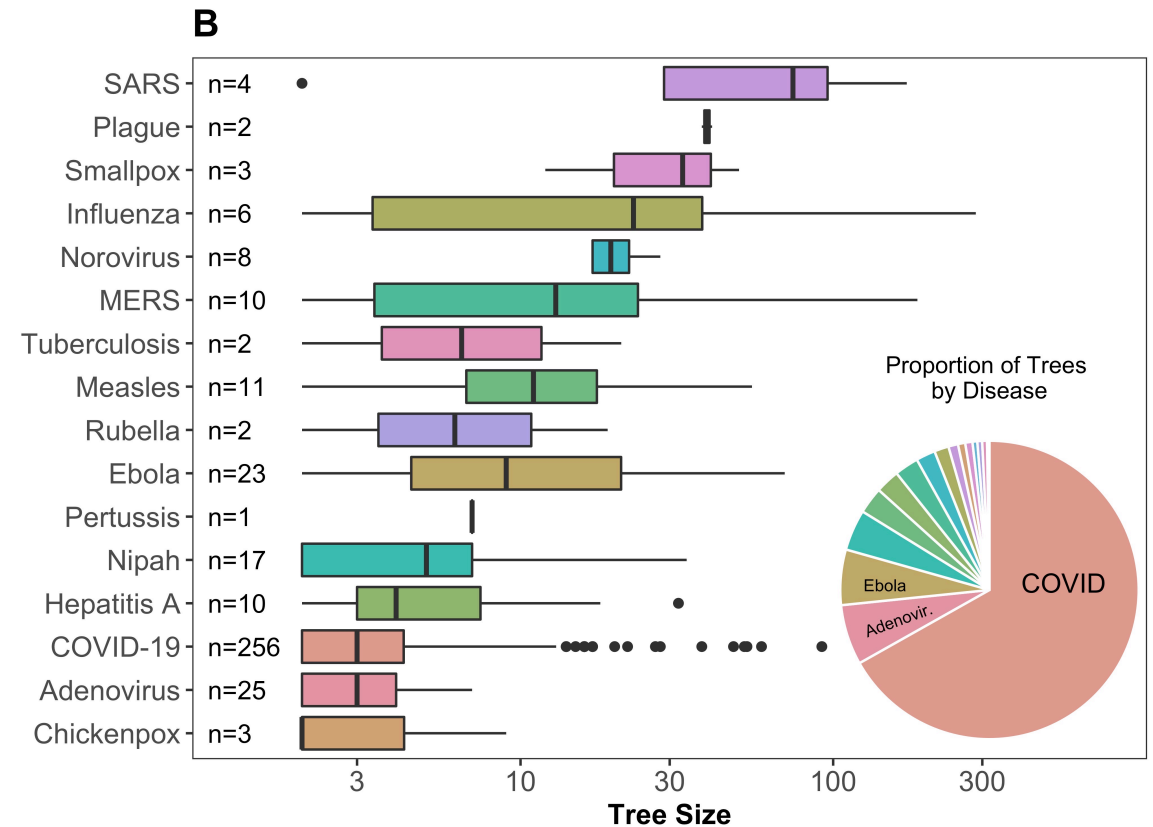
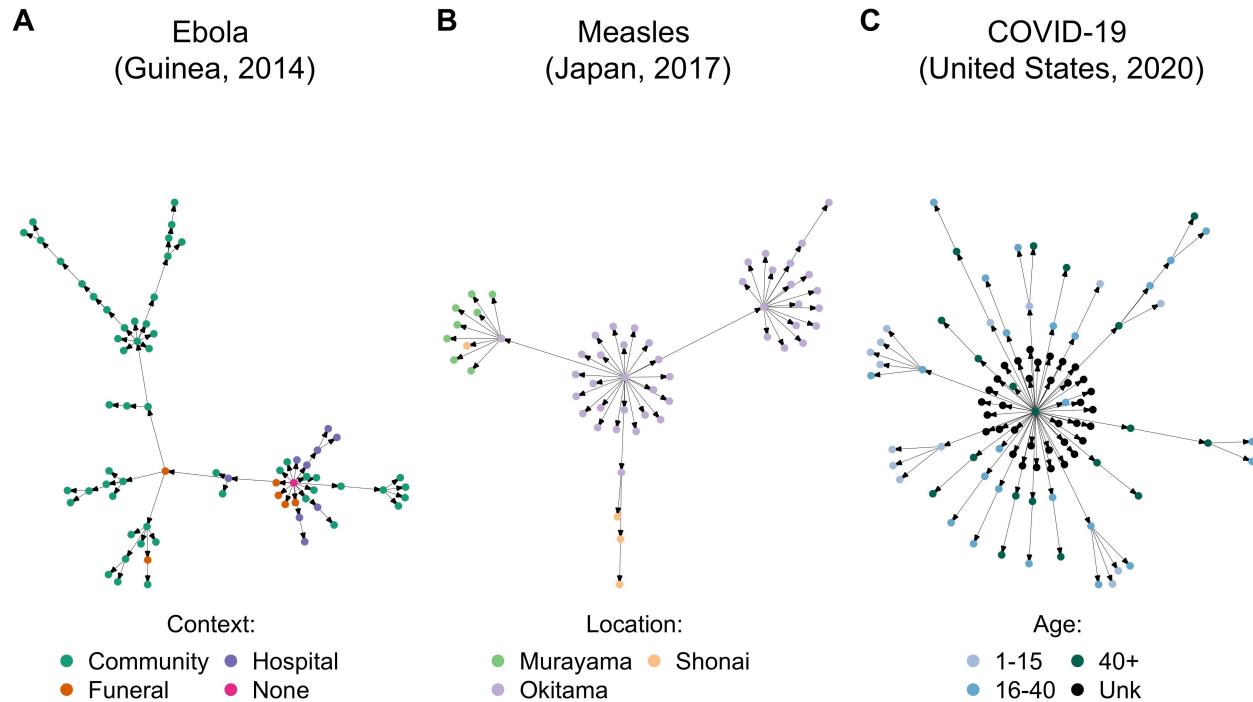


Figure 2 (below). **Example trees in the database.** (A) Ebola spread in different contexts<sup>2</sup>. (B) Measles spread in different locations<sup>5</sup>. (C) COVID-19 spread among age classes<sup>6</sup>.

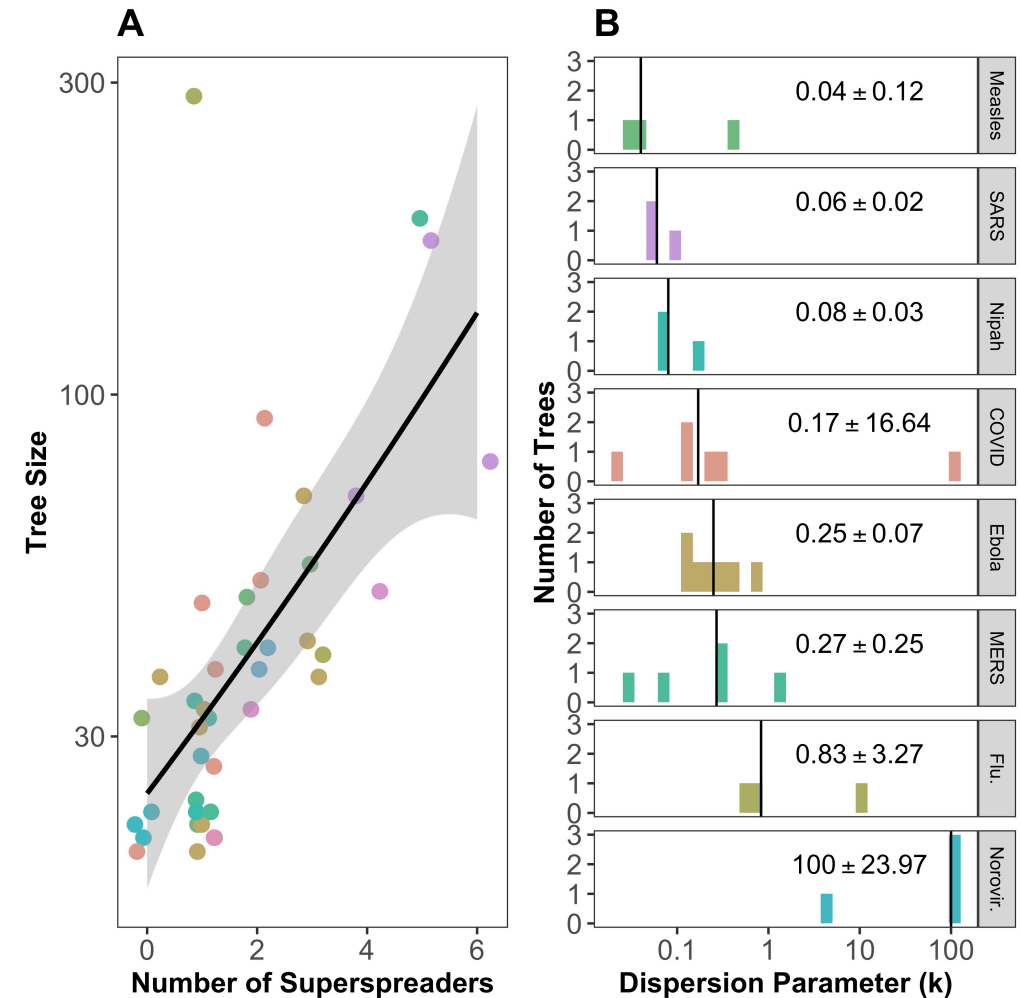




## Analyses Details

- Average individual reproductive number ( $R_0$ ): average number of secondary infections across all individuals in the tree excluding terminal nodes in the last generation
- Secondary infections were assumed to follow a negative binomial distribution ( $\mathbb{P}(X = k) = \binom{k+r-1}{k} p^k (1-p)^r$ ) and dispersion parameters ( $k$ ) were estimated using maximum likelihood methods (MASS package, fitdistr, no initial values given)
  - Large values of  $k$  denote little variation in number of secondary infections per case, while small values of  $k$  ( $k < 1$ ) correspond to high heterogeneity in the offspring distribution
- Superspreaders<sup>7</sup>: cases who transmitted to more individuals than the 99th percentile of a  $\text{Poisson}(R_0)$  ( $\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ )
- Superspreader-superspreader dyads occur when one superspreader infects another
  - We expect these dyads to occur at a frequency of  $s(s-1)/S$  where  $s$  denotes number of superspreaders and  $S$  denotes tree size

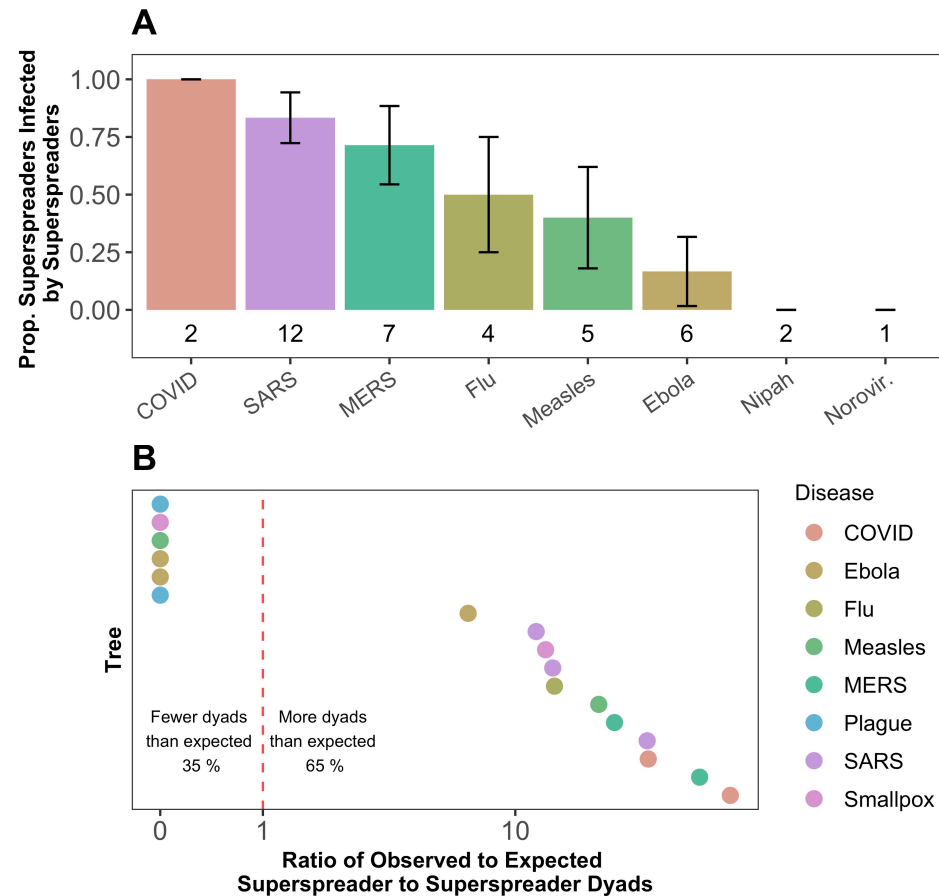
Figure 3. **The importance and expected frequency of superspreading across diseases.** (A) Larger trees tend to have more superspreaders. Points are jittered vertically, and y-axis is on a  $\log_{10}$  scale for visual aid. (B) Dispersion parameter ( $k$ ) of a negative binomial distribution fit to the offspring distribution of trees by disease (for diseases with 3+ trees). Vertical line and value printed in each facet shows the median  $k$  and standard error for each disease. Only trees with 20+ cases and 2+ generations of spread were used in these analyses.



# ILLUSTRATIVE ANALYSES



Figure 4. **Characteristics of individuals infecting superspreaders.** (A) The proportion of superspreaders infected by other superspreaders. Standard error bars shown. Numbers below bars indicate the number of superspreaders for which there was sufficient information about their infector to calculate the proportion. (B) Ratio of observed to expected superspreader-superspreader dyads in trees with more than one superspreader. The expected number of dyads is calculated by  $s(s - 1)/S$ , where  $s$  is the number of superspreaders in the tree and  $S$  is tree size. Both analyses are limited to trees with 20+ cases and 2+ generations. (A) is additionally limited to diseases with 3+ trees.



## Generation of superspreaders

Superspreaders may be more likely to infect other superspreaders due to:

- Biological patterns: Individuals with high viral shedding may cause those they infect to also have high viral shedding<sup>8</sup>
- Behavioral patterns: Individuals engaging in riskier behavior (e.g., attending large gatherings, not following precautionary measures) may infect others with similar behavior<sup>9</sup>

## CONCLUSIONS

- Transmission trees contain valuable information about specific outbreaks, which is costly to collect. Our database **OutbreakTrees** standardizes tree format, allowing for greater comparative analyses.
- We illustrated how this database can be used to explore questions surrounding superspreader epidemiology, including estimating the dispersion parameter of COVID-19 to be between that of SARS and MERS and providing support for the theory that superspreaders generate other superspreaders

## REFERENCES & ACKNOWLEDGMENTS

1. Cauchemez et al. (2011) *PNAS*, **108**(7).
2. Faye et al. (2015) *Lancet ID*, **15**(3).
3. Glur (2018) <https://CRAN.R-project.org/package=data.tree>
4. Csardi & Nepusz (2006) *InterJournal, Complex Systems*.
5. Komabayashi et al. (2018) *JJID*.
6. Ohio Department of Health. (2020) <https://coronavirus.ohio.gov/wps/portal/gov/covid-19/resources/news-releases-news-you-can-use/covid-19-update-08-04-20>.
7. Lloyd-Smith et al. (2005) *Nature*, **438**(7066).
8. Beldomenico (2020) *International Journal of Infectious Diseases*.
9. Susswein & Bansal (2020) *medRxiv*.
10. All sources for trees can be found at <http://outbreaktrees.ecology.uga.edu>

Population Biology of Infectious Diseases REU Site (NIH U54 GM111274) provided funding and the Drake Lab provided helpful comments and assistance. We thank E. Marty for assistance developing the online interface to **OutbreakTrees**.