

# analysis

Juliana Taube

1/8/2021

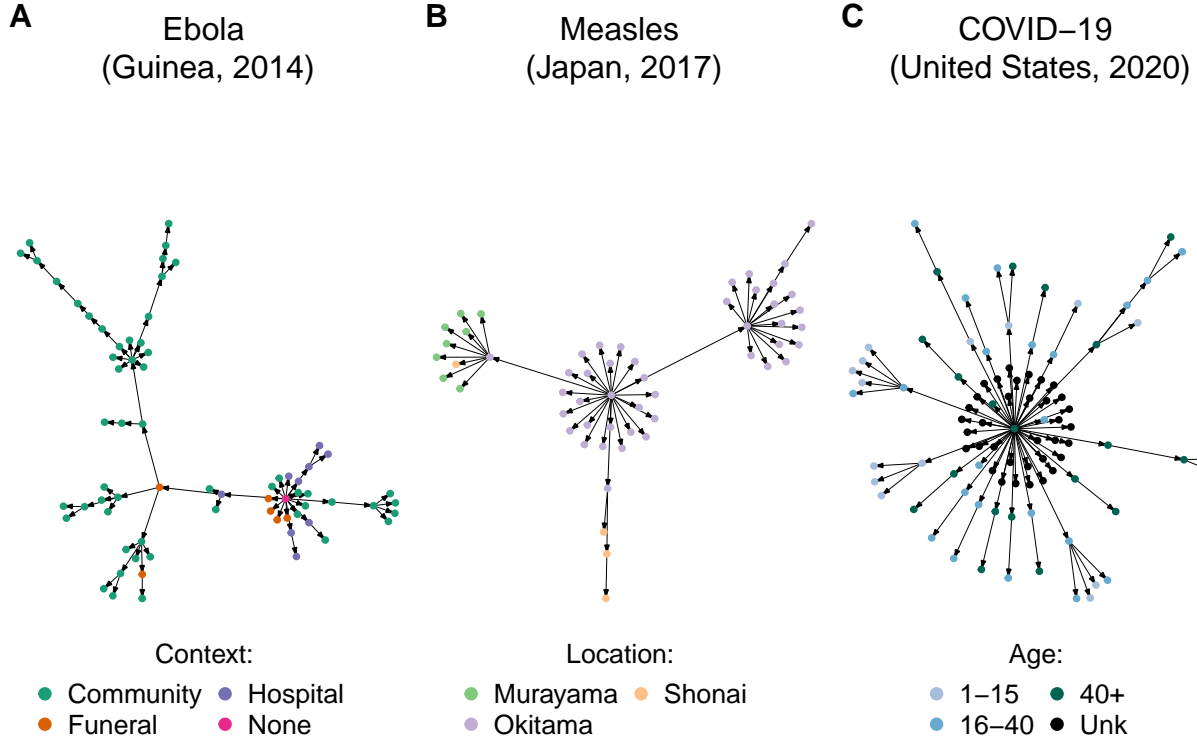


Figure 1: We compiled infectious disease transmission trees from the literature along with reported attribute information. Shown here are example trees in the database. (A) Ebola spread in different contexts [9]. (B) Measles spread in different locations [10]. (C) COVID-19 spread among age classes [11]. Primary sources for transmission trees are available in **OutbreakTrees** and listed in the Supplemental Material. **OutbreakTrees** may be accessed online at [OutbreakTrees.ecology.uga.edu](https://OutbreakTrees.ecology.uga.edu)

Table 1: List of most common attributes for individuals in trees.

Attribute	Database Code	Number of Trees
Transmission context	cont	301
Symptom Onset	onset	137
Sex	sex	86
Age	age	69
Location	loc	56
Quarantine Status	quar	36
Occupation	occp	34
Survival	surv	20

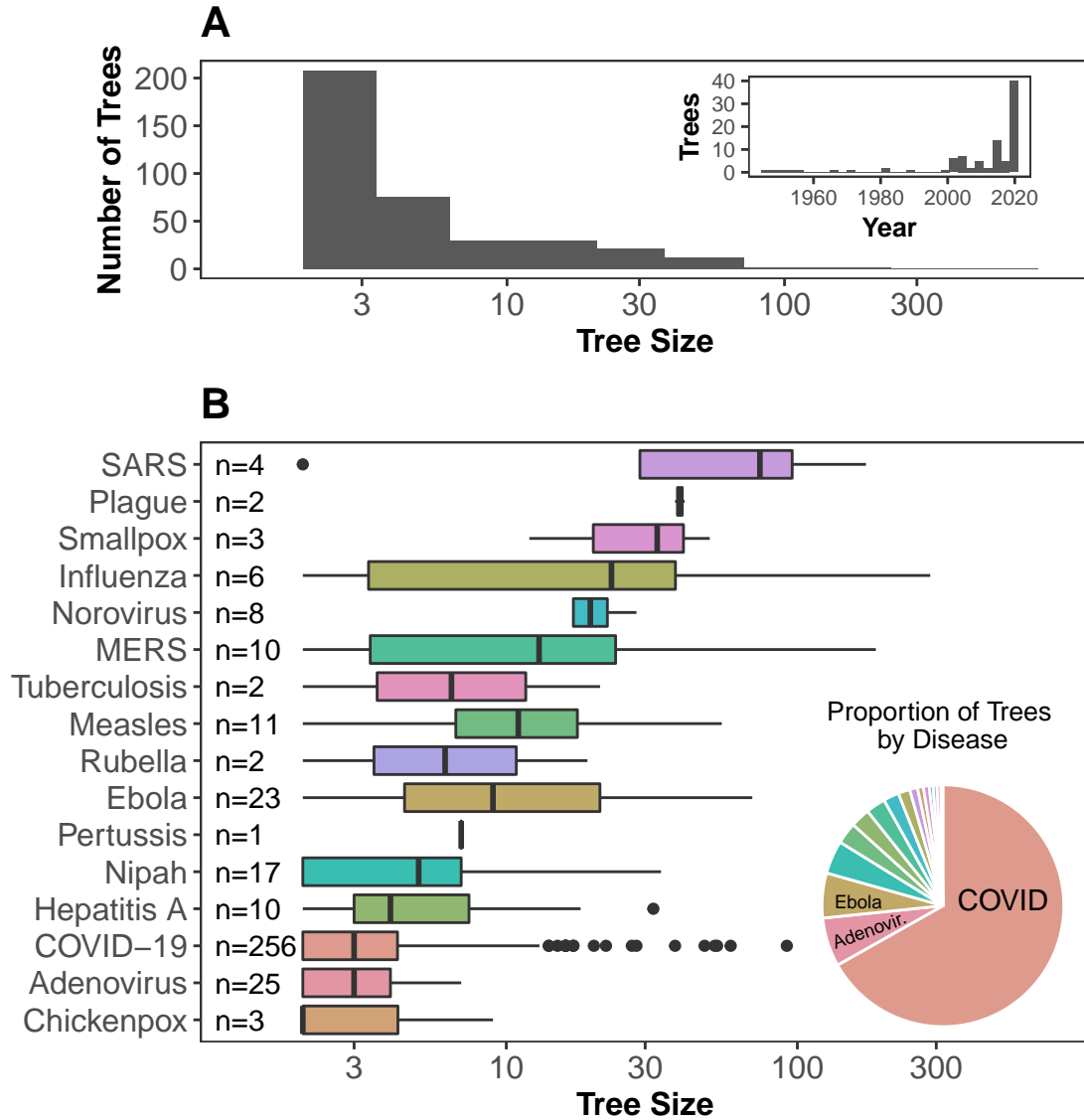


Figure 2: Characteristics of transmission trees in *OutbreakTrees*. (A) Tree size varies from 2 to 286 with a median of 3 and most trees represent outbreaks taking place in the past 20 years. (B) The largest trees are from H1N1 and SARS outbreaks while the highest proportion of trees in the database are from outbreaks of COVID-19, followed by adenovirus and Ebola. Tree size axes in both plots are shown on a log10 scale to better illustrate variation in medium-sized trees.

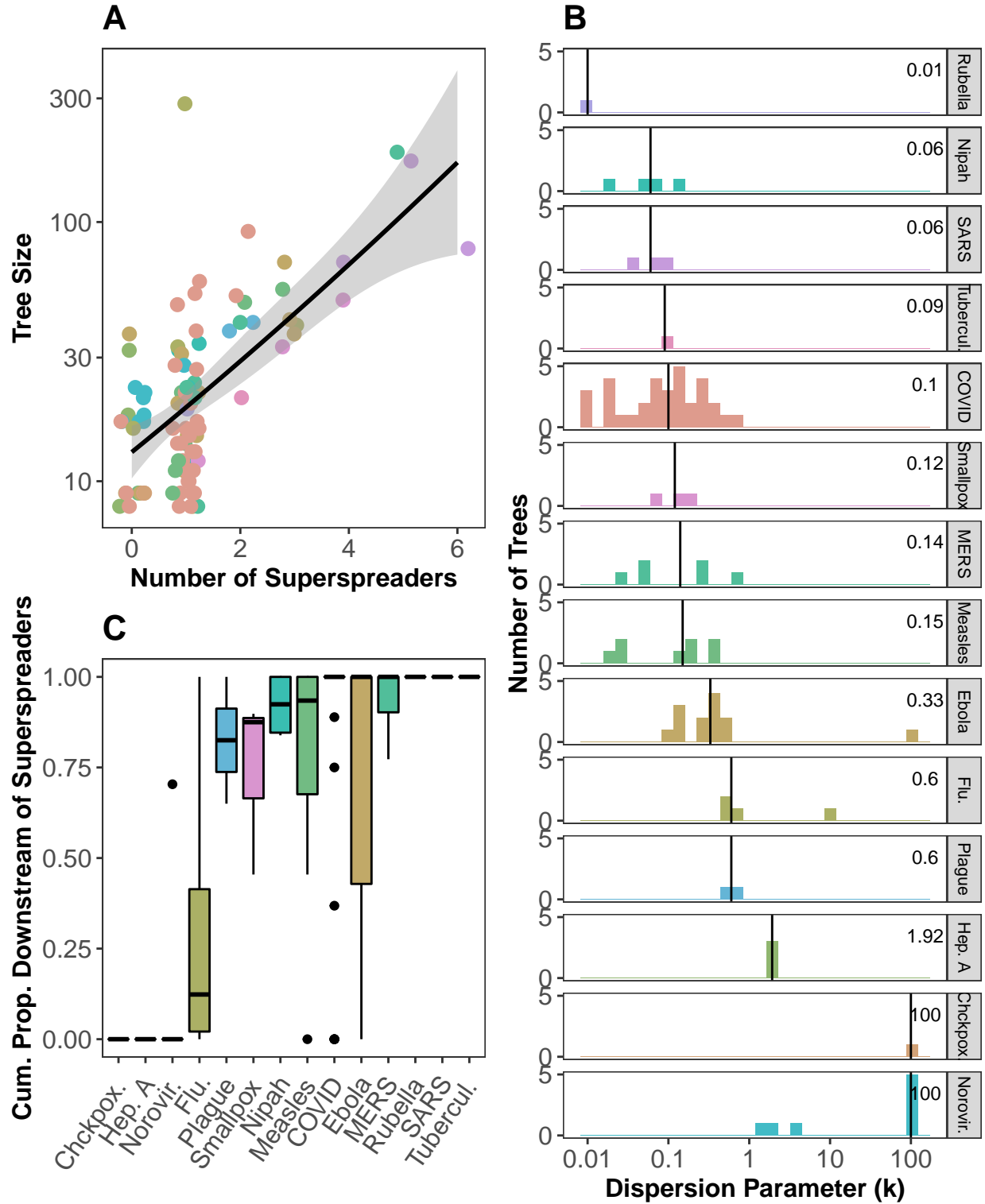


Figure 3: The importance of superspreading across diseases measured in three ways. (A) Larger outbreaks tend to have more superspreaders. Points are jittered vertically and y-axis is on a log10 scale for visual aid. (B) Dispersion parameter ( $k$ ) of a negative binomial distribution fit to the offspring distribution of trees by disease. Vertical line and value printed in each facet shows the median  $k$  for each disease. (C) Cumulative proportion of cases directly or indirectly attributed to superspreaders in a tree by disease. Only trees with 8 or more cases were used in these analyses with other cutoffs shown in S1 Fig and S2 Fig.

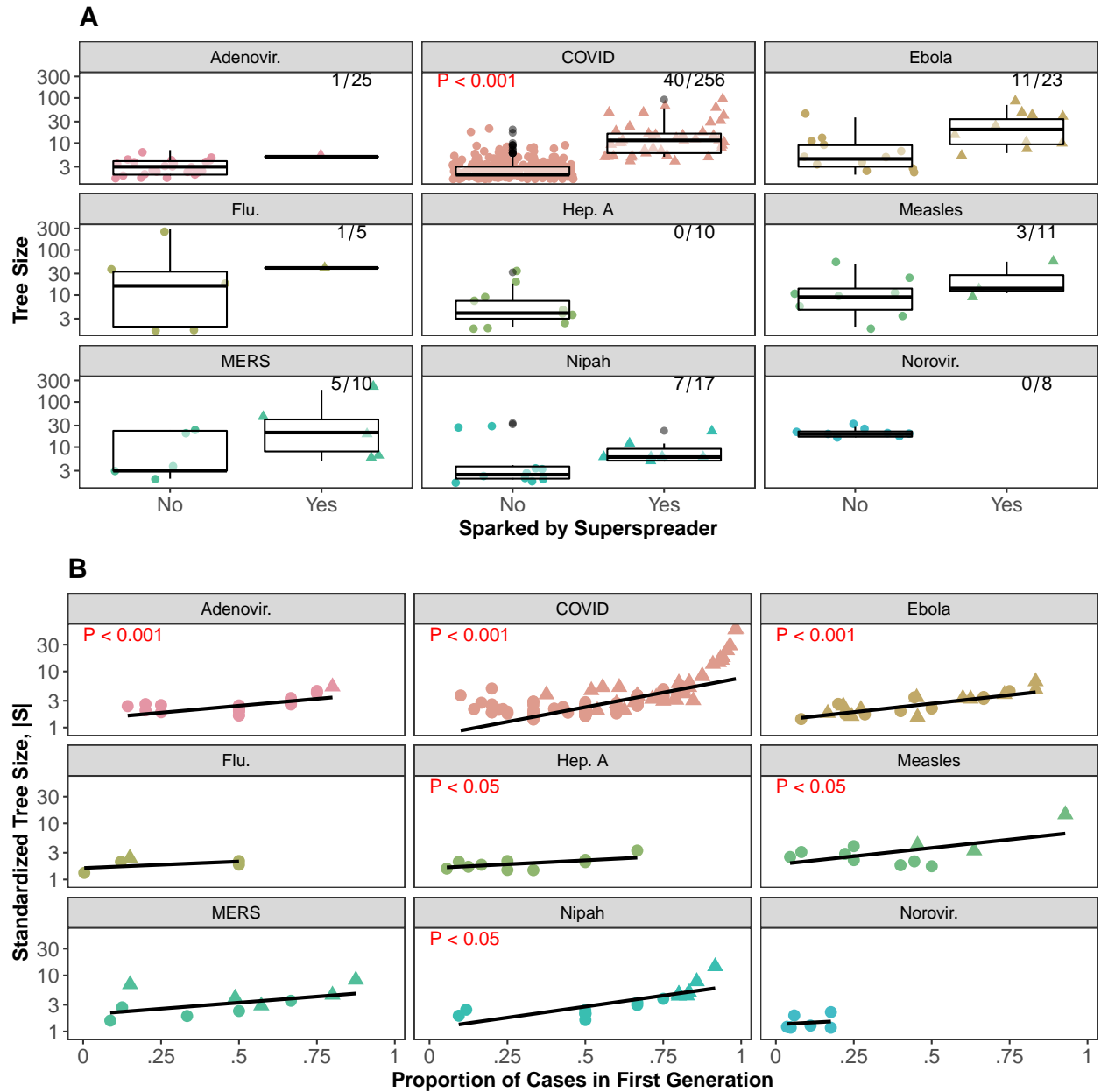


Figure 4: Importance of early spreading events to tree size across diseases. (A) The frequency (top right) and effect of trees sparked by superspreaders on tree size. Red text indicates results of Kruskal-Wallis test with Bonferroni-Holm correction of differences in tree size by superspreader sparking status (only P-values < 0.05 shown). (B) Across diseases there is a positive relationship between standardized tree size and proportion of cases in the first generation. Standardized tree size,  $|S|$ , is calculated by  $|S| = S^{1/(G-1)}$  where  $G$  is the number of generations in the tree and  $S$  is tree size. Lines show linear regression with results in red text (only coefficients with P-values < 0.05 shown). Note: y-axis on log10 scale. Trees sparked by superspreaders are represented by triangles, trees not sparked by superspreaders are represented by circles. Points have been jittered for clarity.

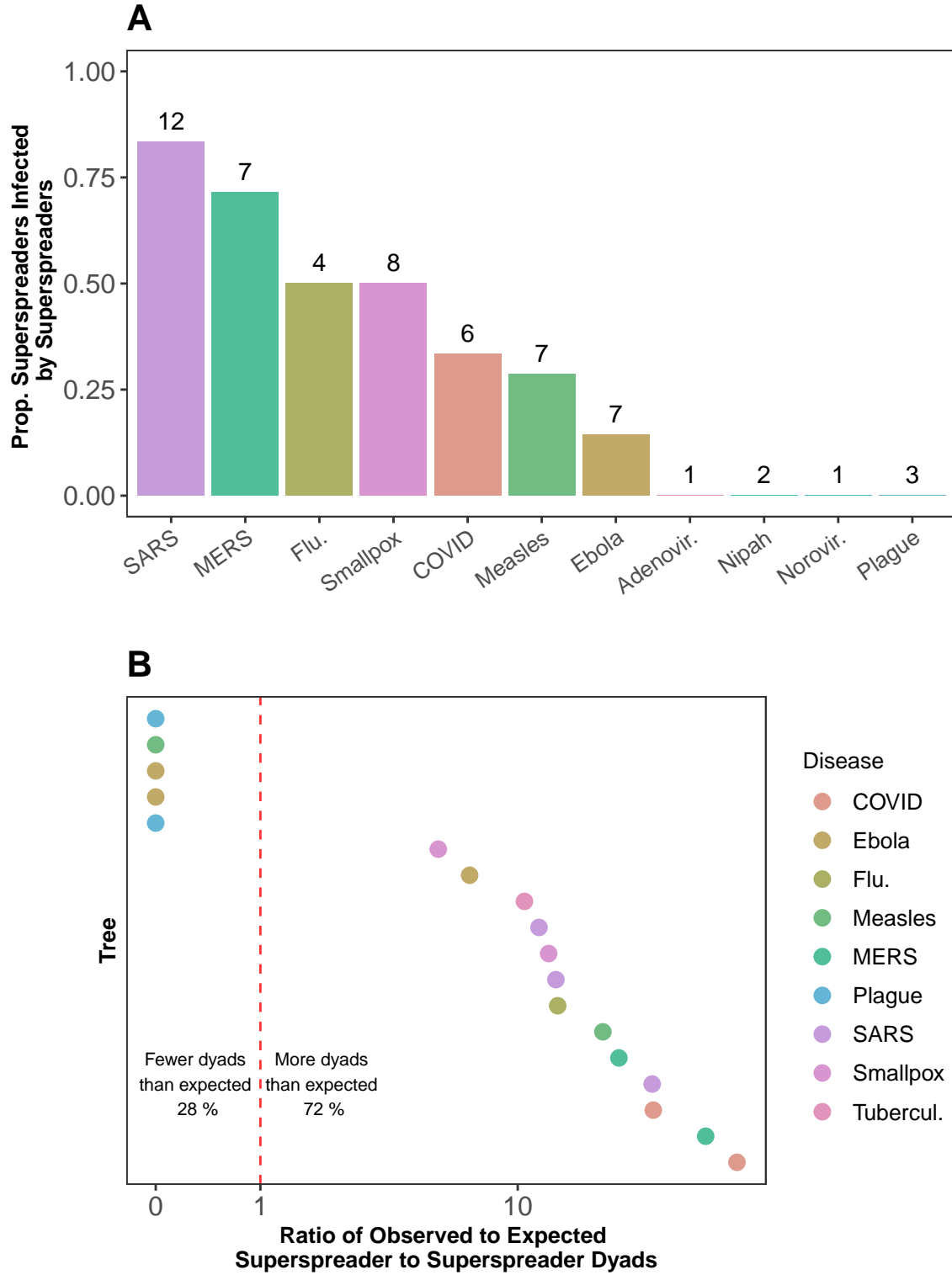


Figure 5: Characteristics of individuals infecting superspreaders. (A) The proportion of superspreaders infected by other superspreaders. Numbers above bars indicate the number of superspreaders for which there was sufficient information about their infector to calculate the proportion. (B) Ratio of 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.

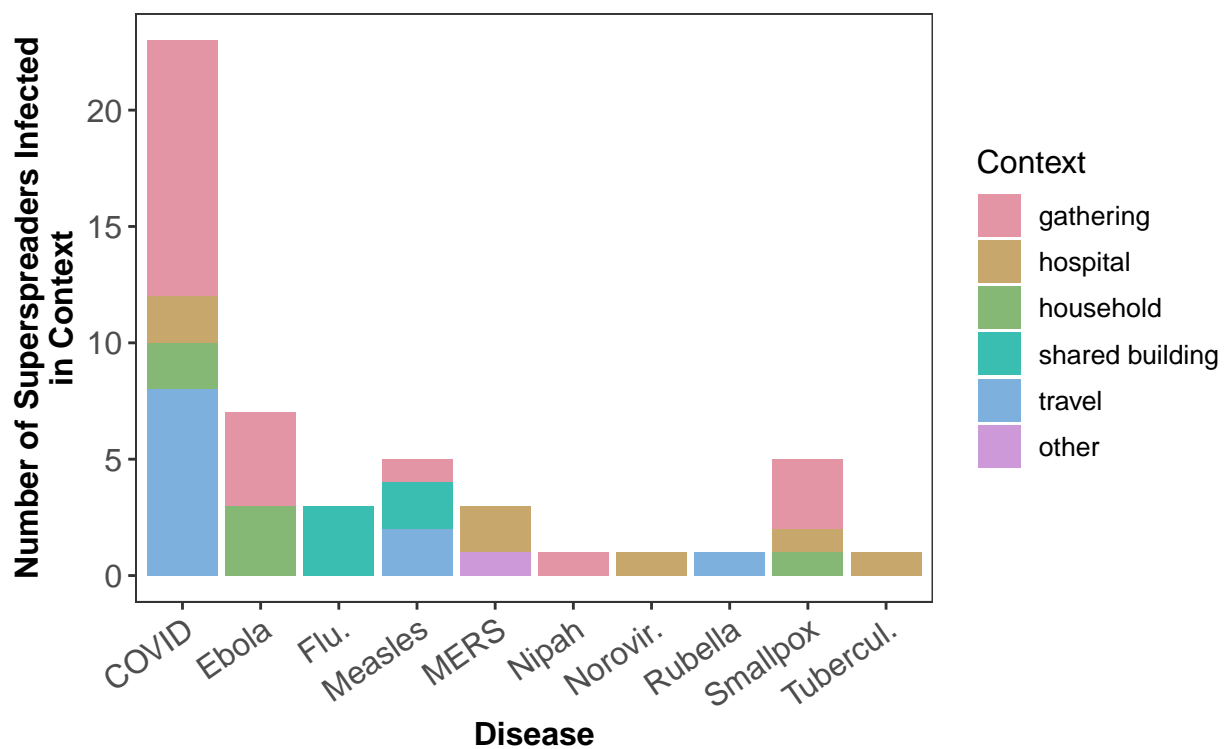


Figure 6: Number of superspreaders infected in various transmission contexts. Shared buildings included worker dormitories and hotels. The category “other” includes a spillover event of MERS to a camel owner.