Twitter thread

Excited to share our new preprint! We develop a new method to infer stochasticity in transmission dynamics from lineage frequency time series data, and we find consistently high levels of stochasticity in SARS-CoV-2 transmission in England.

Stochasticity in transmission, such as due to superspreading, affects disease transmission and evolution.

However, except in some specific scenarios, it has been challenging to quantify stochasticity in transmission more generally across time and space using traditional methods (contact tracing and models of outbreak trajectories).

Frequency time series data can tell us about the strength of genetic drift, which in this case is dominated by stochasticity in the transmission dynamics (because we think little within-host diversity for acute cases transmitted due to short times and small bottleneck sizes). We can leverage the large amounts of SARS-CoV-2 sequencing data to get lineage frequency time series (lineage defined here are genetically similar groups of sequences).

However, we have to deal with measurement noise, which arises from the process of how the sequences are collected. Who is sampled? How does sequencing capacity vary across locations? This also contributes noise the data that confounds that from genetic drift.

To deal with this, we develop a Hidden Markov Model that can jointly infer genetic drift and measurement noise, even when these two processes vary in strength over time. Genetic drift adds variance over time, whereas measurement noise does not. (make animation)

Diagram

Description automatically generated with medium confidence

Using this method, we were able to infer the strength of stochasticity in transmission dynamics in England across time, space, and across different variants.

We find much more stochasticity than we would expect from the number of people who are positive! It also is much more than expected from the extent of superspreading reported in the literature.

Host population community structure could lead to this increased stochasticity. More work needs to be done on these models to tease apart the cause of the increased stochasticity.

We think this method could be useful for inferring genetic drift and effective population size, compared to phylogenetic tree-based methods, when the mutation rate is low, there are a lot of sequences, and one needs to account for noise in the measurement process.

All of our code is available on github for others to explore. Please reach out if you think this method may be useful for your work! As a side note, we also develop a HMM that incorporates selection and drift in the manuscript.

We welcome any feedback!

This was a wonderful collaboration with Joao Ascensao, Takashi Okada, and Oskar Hallatschek from my PhD lab in Berkeley and Olivia and Erik from Imperial College London. Truly one of the most fun and rewarding collaborations!