

---

# Reconstructing infectious disease outbreaks using genetic and epidemiological data

---

## RESEARCH QUESTION

Reconstructing the transmission tree of a disease outbreak is valuable for identifying a disease's dynamics, origin, and mechanisms of transmission. As genetic sequencing becomes increasingly affordable, the inference of transmission trees can be strengthened using sequence data from the pathogen. Therefore, there is a need for statistical methods that integrate this source of information with epidemiological data to derive the most likely transmission tree. My goal in this project is to explore methods for reconstructing the transmission tree of a disease, based on patients' geographical locations, the times at which they were tested, and genetic data of their pathogen samples.

The proposed method should be able to smoothly handle missing genetic or epidemiological data, and should be able to simultaneously estimate unknown parameters related to the disease, such as its rate of mutation and infectiousness.

## EXISTING WORK

Jombart et al [1] contrast the construction of transmission trees with the problem of phylogenetic tree reconstruction, noting that especially for newly emerging pathogens, ancestors and their descendents are likely to be present in the same population. They use genetic information to derive the most likely ancestor of each sequence, using epidemiological data only to break ties. Ypma et al [3] construct a likelihood function combining genetic, spatial and temporal data, assuming that these components are independent of one another. Morelli et al. [2] uses a Bayesian inference scheme which combines genetic and spatiotemporal data

using a dynamical model of disease progression, in which agents progress through infected, latent, and detected stages.

## REFERENCES

- [1] T. Jombart, R. Eggo, P. Dodd, and F. Balloux. Reconstructing disease outbreaks from genetic data: a graph approach. *Heredity*, 106(2):383–390, 2010.
- [2] M. J. Morelli, G. Thébaud, J. Chadœuf, D. P. King, D. T. Haydon, and S. Soubeyrand. A bayesian inference framework to reconstruct transmission trees using epidemiological and genetic data. *PLoS computational biology*, 8(11):e1002768, 2012.
- [3] R. Ypma, A. Bataille, A. Stegeman, G. Koch, J. Wallinga, and W. Van Ballegooijen. Unraveling transmission trees of infectious diseases by combining genetic and epidemiological data. *Proceedings of the Royal Society B: Biological Sciences*, 279(1728):444–450, 2012.