

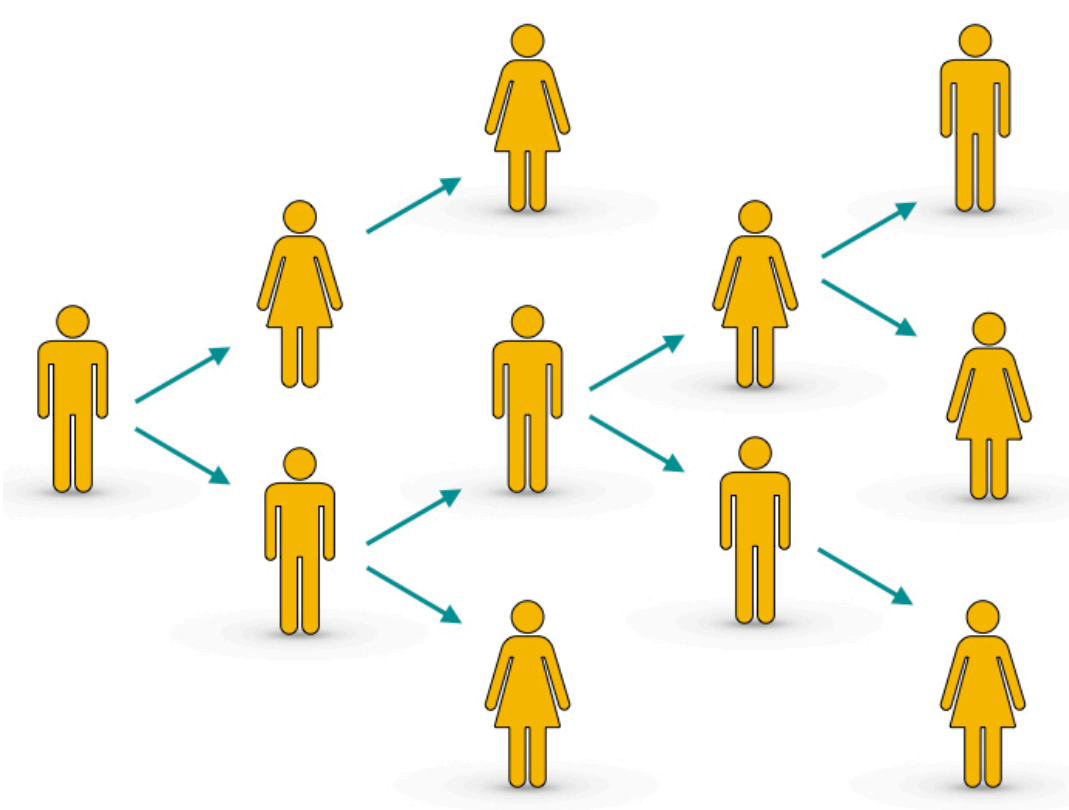
# Comparing and Summarising Outbreak Transmission Trees

Michelle Kendall<sup>1</sup>, Diepreye Ayabina<sup>2</sup>, Yuanwei Xu<sup>3</sup>, James Stimson<sup>3</sup>, Caroline Colijn<sup>3</sup>

Current affiliations: <sup>1</sup>Big Data Institute, University of Oxford, <sup>2</sup>Liverpool School of Tropical Medicine, <sup>3</sup>Imperial College London. Work conducted at Imperial College London, UK, and supported by the Engineering and Physical Sciences Research Council grants EP/ Ko26003/1 (CC), EP/No14529/1 (CC and YX). JS hosted by the Centre for Integrative Systems Biology and Bioinformatics at Imperial College London. DA funded by the Nigerian Universities Commission [under the presidential special scholarship scheme for innovation and development (PRESSID)].

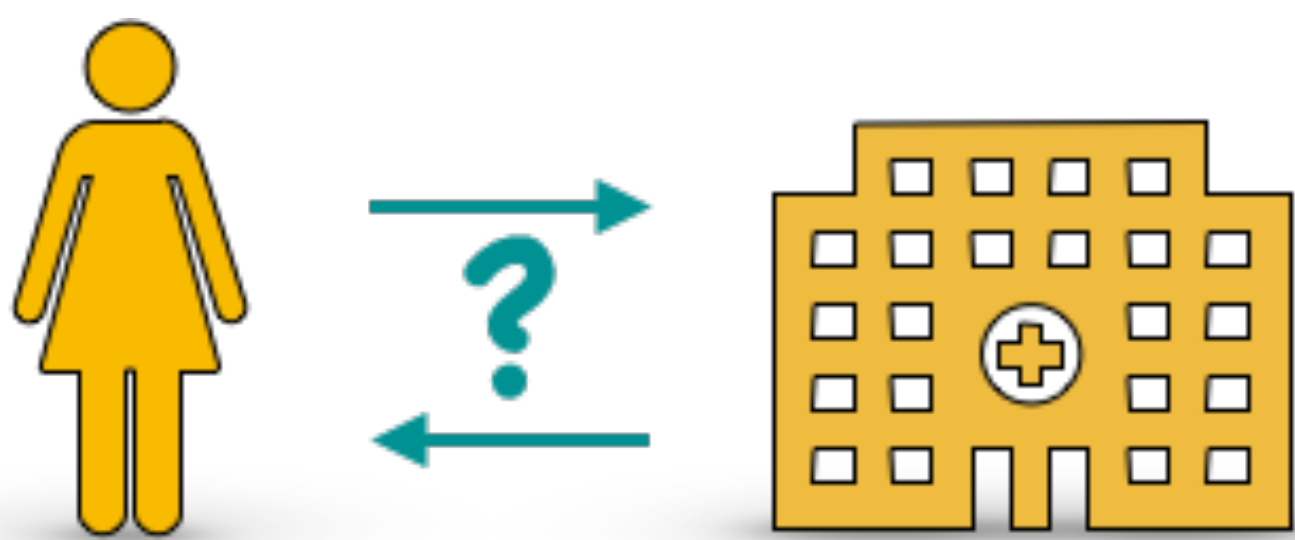
## Problem:

### Knowing who infected whom



When there is a single source case and there are no cases of multiple infection, the diagram of “who infected whom” is a tree: the *transmission tree*.

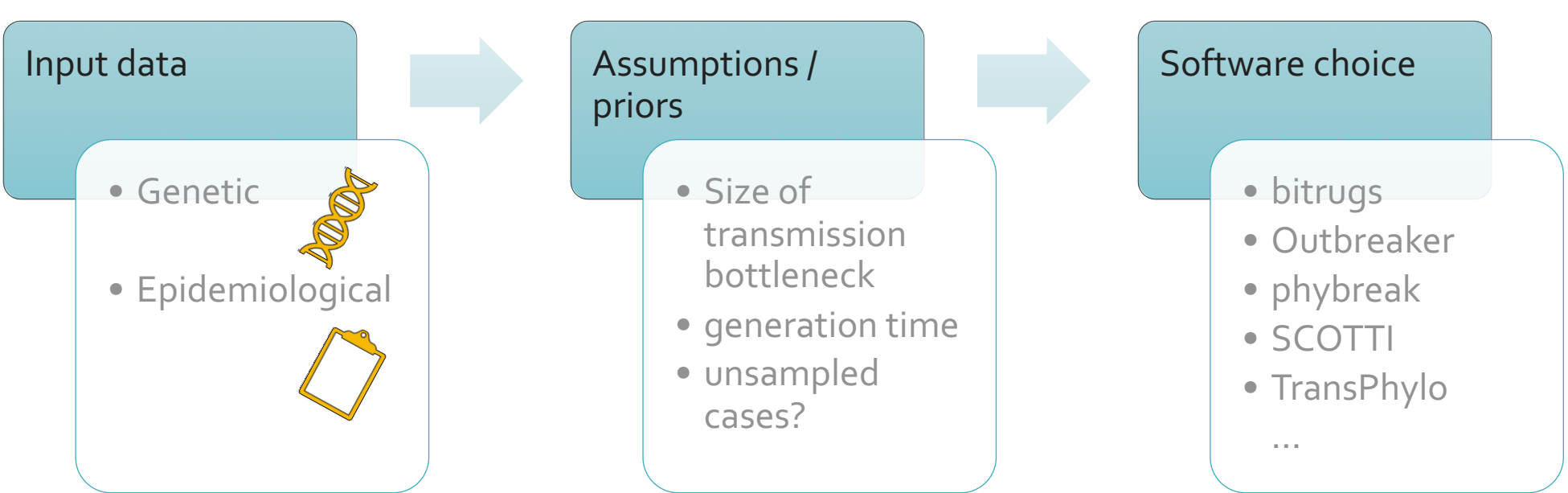
### ... is important



We use the transmission tree to

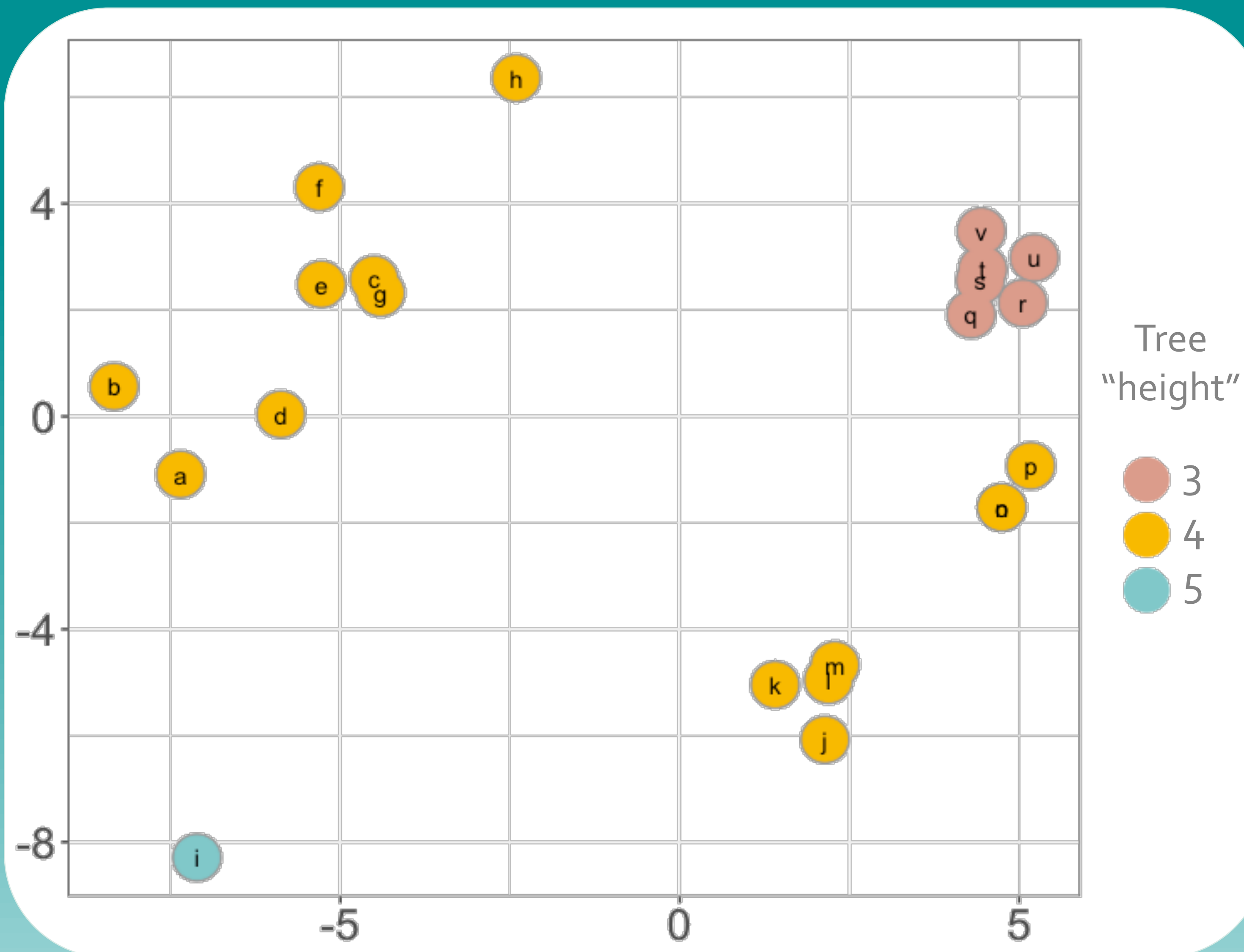
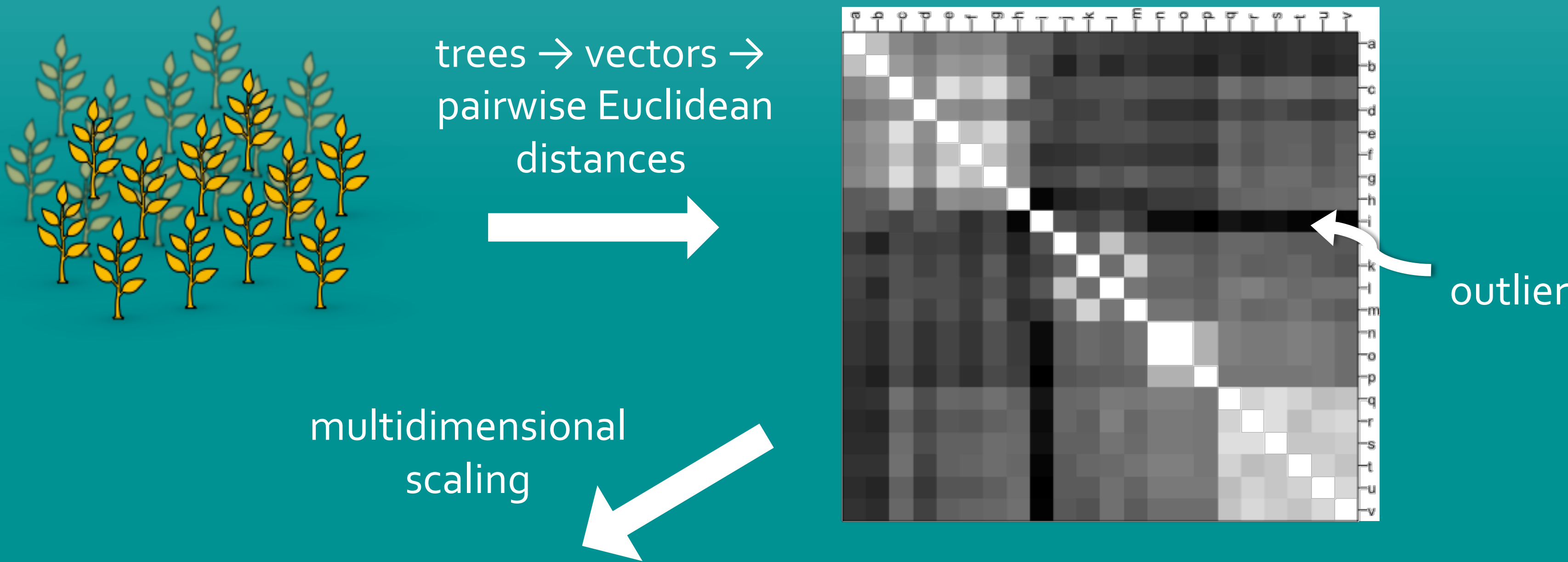
- understand transmission dynamics
- determine infectivity risk factors
- inform public health strategies

### ... but difficult.



Each methodological choice typically changes the final tree(s). Bayesian methods produce thousands of trees. Unless they are all the same it is hard to know how to interpret them.

## Tree comparison solution:



each tree is uniquely characterised by a vector  $v$ :  
for each pair  $(x, y)$  of individuals sampled,  
 $V_{(x,y)}$  = depth of most recent common infector of  $x$  and  $y$




Our tree comparison method enables quick and easy identification of:

- **tree agreement**: similarity? outliers? Often, the thousands of trees from a Bayesian analysis seem to disagree but can be grouped into just a few (equally likely) alternatives of the overall transmission history.
- **input / methods / software agreement**: do trees differ if we change our methods? (Typically yes.) By how much?
- **representative trees**: previous consensus trees have serious drawbacks. From a collection of trees we produce a credible summary tree with associated likelihood.
- **convergence**: has the MCMC converged on an area of tree space or is it still “moving”?

### For more information:

Kendall, M., Ayabina, D., Xu, Y., Stimson, J. and Colijn, C. (2018) **Estimating transmission from genetic and epidemiological data: a metric to compare transmission trees**, *Statistical Science*, 33(1):70-85

Full text available at:  
<https://michellekendall.github.io/publications/>

Functions and explanatory vignette available in R package  **treespace** from [CRAN](#) or [GitHub](#).

[michelle.kendall@bdi.ox.ac.uk](mailto:michelle.kendall@bdi.ox.ac.uk)

Follow @mishkendall

