**Microbial Source Attribution – Workshop**

**Introduction**

* to tree or not to tree? a central question
* phylogeny reconstruction vs. simple genetic distance

**Oliver Ratmann**

* Pangea HIV on figshare
* using information from high coverage, dense epidemics to much lower sampled epidemic with much larger prevalence (**incidence** in SA is 3x larger than **prevalence** in UK)

**Nicola De Maoi**

* Outbreak in a limited space (hospital) with high sampling coverage
* Genetic data
* Epidemiological data
  + admission and discharge date
* Complications
  + Phylogeny ≠ transmission tree
  + Incomplete bottleneck – multiple lines passed on
  + Un-sampled cases (asymptomatic)
  + High prevalence – multiple strains within host
* Structured coalescent (?)
  + accurate but slow (<4 patients feasible)
* Outbreaker has no within host evolution
* SCOTTI is their program
  + higher accuracy than outbreaker but uses more data (outbreaker only has date of sampling, SCOTTI uses ward admission/discharge)
* Their method includes within host evolution and then migration between hosts

**Thibaut Jombart**

* mutation at transmission
  + wrong, but still works
* average genetic signature <1
  + number genetic differences between two successive cases
* nice and slow
  + within-host evolution
  + using a phylogenetic likelihood (not just genetic distance)
* genetic data alone is insufficient in many cases
* contact tracing data
  + WHO have goData which uses contact tracing data to reconstruct outbreaks
* assume conditional independence between data and parameter
  + this allows you to add new terms to your likelihood by multiplying in different likelihood components
* modularity is the central idea of outbreaker2

**Samantha Lycett**

* Structured coalescent
  + is “correct” if we assume a structured population
* Joint inference
  + informed by genetic and spatial data
  + but only works if you have genetic variation, otherwise you just get clusters of locations
* Looking at avian influenza (H5)

**Philip O’Neill**

* Reconstructing transmission trees for communicable disease using densely sampled genetic data (Worby *et al.*)
* Compress WGS data in to a matrix of distances between individuals
  + similar to outbreaker
* Create a model to generate the observed data, not describing a likelihood around the data (?)
  + Allow multiple colonisations of different strains
  + Construct a model that implicitly includes “distances”
  + SI model (susceptible and colonised)
  + Allow import of colonised
  + Test sensitivity
  + Force of infection parameter β
* Modelling genetic distances
  + For each new colonisation; specify the distance between that colonisation and **every other patient**
  + Classify as **importation** or **acquisition** to calculate distance
  + Include the degrees of separation for acquisition (# generations)
  + They model transmission network and output data (distances)
* Use MCMC for inferring transmission model parameters
* Visualise transmission network with circle of individuals on outside and lines between points as transmission events
* A **generative modelling approach** (generates the data)
* Model differences between sequences rather sequences themselves
  + This is the same as simOutbreak
  + How do you model genetic differences to imports?
* Requires dense sampling

**Federica Giardina**

* Inferring unobserved contact networks from phylogenetic trees
* These contact networks are crucial to dynamics and control methods
* Try inferring network properties
* Contact structure leaves a footprint in the genetic structure of the pathogen population, given a fast enough evolutionary rate
* Coalescence ≠ transmission
* Include within-host diversity
* Generate network structures to capture features of MSM or needle sharing networks

**Sergei Pond**

* hiv-trace.org
* HIV transmission often occurs outside of “known partners” and contact tracing – the data is not reliable!
* Low per-contact transmission rates
* Defining transmission clusters; no standard definition

**Richard Goldstein**

* Mapping our knowledge of genealogical trees on to transmission trees

**Art Poon**

* ABC for phylogenies
* Using a kernel method

**Siu Yin Lau**

* Integrating genetic and epidemiological data
* partial data (infection time unknown, unobserved transmission)
* **No** imports
* Pseudo-likelihood as we only use observed sequences
* Accurate inferences requires explicit description of unobserved events
* Demonstrate effective imputation of transmitted sequences
* **Spatial** SEIR model using a kernel
* Evolutionary Kimura model
* Need to impute first sequence for imports
  + need a model to specify the first sequence
  + have a master sequence GM – the import is a variant of GM with a probability p
* In order to achieve efficient mixing
  + Narrow down parameter space for imputed sequence by inferring artificial evolution between two observed points
* Also describe partial genome sequencing
  + Most epidemiological parameters are equally well described
  + The **mutation rate** is strongly improved by considering all bases, but you can improve computational time by considering fewer bases if you are only interested in epidemiological parameters
* With subsampling (50%)
  + You cannot reconstruct exact transmission tree but cluster assignment is still successful

**Don Klinkenberg**

* Worked with Ypma
* Unobserved
  + infection to transmission
    - define a generation interval (w.dens)
  + infection to sampling
    - f.dens
    - shape of both is fixed so can be defined by one parameter
  + coalescence to transmission
    - coalescence within host (minitree)
    - Bottleneck of 1 at transmission
    - Define coalescence rates to describe different forms of coalescence
    - Defined by one parameter; in-host pathogen growth rate
  + mutation
    - fixed mutation rate
* Tree is updated not from phylogeny perspective but transmission
* Basic idea of 80% proposal steps:
  + Choose one host
  + propose next infection time
  + propose new infector
  + simulate phylogenetic minitrees of involved hosts (host & infector)
* For remaining 20% (with many SNPs)
  + phylogeny is essentially fixed (?)