Literature Review

**Thibaut *et al.***

**Benefits**

* allows for multiple introductions (not in Ypma or Morelli)
* accounts for unobserved events (not in Ypma or Morelli)

**Drawbacks**

* Broader generation time distribution increases error on infection date estimates
* Generation estimates less effective with lower sampling density
* Mis-identifying imports increases error on mutation rate estimates
* Long generation times and subsequently high diversity between generations makes the detection of imports (identified by their high diversity) nearly impossible
* It is lacking denominator data (i.e. number of susceptibles) so the force of infection cannot be inferred
* Generation time distribution is not rigorously defined and, if unknown, involves manual testing of effective distributions (introducing arbitrariness)
* Currently not effective for pathogens with long/variable generation times; date of isolates (i.e. symptomatic) tells us less about the transmission route, because it may have happened over a very long time period
* Assumption of single lineage within patient; therefore not appropriate for multiple infections

**Questions**

* Outbreaker looks only at the number of mutations between pairs?
  + IE it doesn’t use accumulation of mutations along a chain to lend support to a given tree? As shown for sin2500 -> sin2679 upon manual inspection
* How does the convolution function work to incorporate several generations of transmission of the generation time distribution?
* Epidemiological pseudo-likelihood
  + Convolution operator?
  + Do this make any assumptions about generation times of the inferred generations?
* Defining the global influence GI to determine genetic outliers
  + You take 50 samples from the MCMC run and calculate the mean of the sum of genetic pseud-likelihoods (a higher score indicating strong genetic support for the proposed transmission network)
  + The GI of i is defined as the difference between the means when including and excluding i; it is a measure of difference of the pseudo-likelihood of i from the average pseudo-likelihood
  + Individuals with a GI 5x greater than the average GI is considered an import
* When analysing super-spreaders, do we subdivide the groups prior to analysis?
  + i.e. outbreaker does not figure out if individuals are super-spreaders or not
* Average reproductive number over time
  + Time 0 means the time of infection of the individual, and they average over all individuals in the given simulation?
  + Figure S.10 – black lines represent R(t) of the posterior trees?
* Reconstructing SARS
  + Figure S.14 – why exactly can transmission events not be readily inferred from the tree?
* How many accepted trees in the posterior?
* Has this been applied to the Oxfordshire dataset?
  + May struggle with a high influx of genetically diverse pathogens
  + However you have genetic information, date of infection and contact information
* What exactly are the difference between this and phylogenetic approaches?
  + This is looking only at mutation rate and number of genetic differences to compute the likelihood, instead of trying to reconstruct the phylogeny?
* I will be adding contact tracing as a component of the likelihood?
* Spelling mistakes
  + simOutbreak – “Genome length: 10000 nucleotids”
* How many days do you need to provide for the generation time distribution?

**Project**

* Contact tracing can allow the identification of imported cased of a closely related lineage, which are genetically indistinguishable from cases within a transmission chain
* Contact tracing allows for transmission routes of pathogens with long incubation periods to be inferred; the date of isolates tells us less about the time of transmission, but contact tracing allows restriction to times of contact with infected individuals

Look at convolution operator for generation times

Likelihood functions

Movement functions

outbreaker2 – simOutbreak

Try with fixing different parameters

Thinking theoretically about contact tracing

One way / two way

Will you fix all ancestries?

Look at how contact tracing is reported / used

How to incorporate it into the likelihood

**Ypma *et al.***

**Morelli *et al.***

**Paul Kellam** (lecture)

* Real time viral genetics
* Inform us of sustained introduction but low R0, compared to the inverse
* MERS
  + Used genetic information to determine that most cases are single transfers with little onward spread, in that putative cases were far too genetically distant to represent direct transmission events
  + Explained why MERS wasn’t spreading uncontrollably
  + Instead suggested a reservoir which was leaking into the general population
  + This completely changes infection control policy
  + Showed that camel viruses were dispersed across all cases, suggesting they were continuously seeding cases into humans with little human-human transmission

**Further literature**

* Chauchemez – Eurosurveillance (2013)