SCOTTI: Inferring transmission with the Structured Coalescent

Nicola De Maio, Chieh-Hsi Wu, Daniel Wilson
Complications: within-host coalescent

Transmission and evolution history

Transmission network

Phylogenetic tree
Complications

Within-host coalescent

Incomplete bottleneck

Non-sampled case

Multiple infections
Structured coalescent

Coalescence events only within demes
Migration moves single lineages between demes.

We use a recent efficient approximation to the structured coalescent: BASTA (De Maio et al 2015 PLOS Genetics).
Phylogeography with BASTA

(a) Migration

Calibration: 56%
Correlation: 0.58

Discrete trait
(Lemey et al. 2009)
fast but inaccurate.

(b) MultiTypeTree

Calibration: 87%
Correlation: 0.77

Structured coalescent
(Vaughan et al. 2014)
accurate but slow.

(c) BaStA

Calibration: 95%
Correlation: 0.83

BASTA
(De Maio et al. 2015)
accurate and fast.
SCOTTI: Efficient Reconstruction of Transmission within Outbreaks with the Structured Coalescent

- Hosts (demes) have same population size.
- Hosts have limited lifespan.
- No bottlenecks at transmission.
- Lineages do not migrate together at transmission.

Nicola De Maio¹,²*, Chieh-Hsi Wu², Daniel J Wilson¹,²,³
Simulated model

Extension of:

Didelot et al 2014 MBE.

Hall et al 2015
PLOS Comput Biol.

- Bottlenecks at transmission.

- Lineages migrate together at transmission.

- Only one transmission per host.
Benchmark model

Jombart et al 2014 PLOS Comput Biol

- No within-host population.
- No within-host evolution.
- Mutations accumulate at transmission.
- Only one sample per host.
- Generations of same lengths.
Models used

- Multispecies (simulated history)
- SCOTTI (ideal inference)
- Outbreaker (ideal inference)

Models used:

- Jombart et al 2014 PLOS Comput Biol
- De Maio et al 2016 PLOS Comput Biol
- Didelot et al 2014 MBE
- Hall et al 2015 PLOS Comput Biol
Simulations

Calibration 95% posterior sets

Accuracy of point estimate

Outbreaker
Simulations

Calibration 95\% posterior sets

Accuracy of point estimate

Calibration 95\% posterior sets

Outbreaker
Simulations

Within-host similarity

Mean accuracy of the point estimate

Calibration

Simulations

Within-host similarity

Mean accuracy of the point estimate

Calibration
# Simulations – Running time

<table>
<thead>
<tr>
<th>Generations</th>
<th>3</th>
<th>5</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
</tr>
<tr>
<td></td>
<td>4000</td>
<td>4000</td>
<td>4000</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>100s</td>
<td>118s</td>
<td>171s</td>
</tr>
<tr>
<td></td>
<td>124s</td>
<td>239s</td>
<td>346s</td>
</tr>
<tr>
<td>5</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
</tr>
<tr>
<td></td>
<td>4000</td>
<td>4000</td>
<td>4000</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>233s</td>
<td>314s</td>
<td>547s</td>
</tr>
<tr>
<td></td>
<td>385s</td>
<td>774s</td>
<td>1733s</td>
</tr>
<tr>
<td>7</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
</tr>
<tr>
<td></td>
<td>4000</td>
<td>4000</td>
<td>4000</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>414s</td>
<td>822s</td>
<td>1279s</td>
</tr>
<tr>
<td></td>
<td>819s</td>
<td>2021s</td>
<td>4039s</td>
</tr>
</tbody>
</table>
FMDV data

Outbreaker

Cottam et al. 2008 PLOS Pathogens
K. Pneumoniae data

Stoesser et al. 2014
Antimicrobial agents and chemotherapy
Different models result in different inferences.

New inference of transmission in BEAST2: SCOTTI.

Future work: transmission bottlenecks, introductions, epidemiological models.
Thanks for listening!

Daniel J Wilson

Chieh-Hsi Wu

Crook group (NDM Microbiology)