



Phyldynamics with structured population models

Denise Kühnert

Max-Planck Institute for the Science of Human History

Phylogenetic Inference

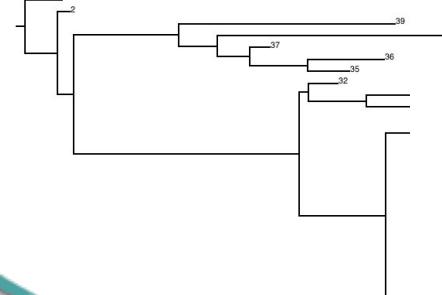


Pathogen genomes can track ecological processes and population dynamics

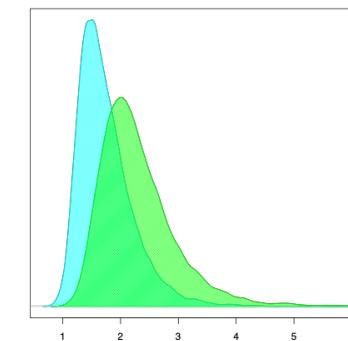
Sequence alignment

E.g. pol sequences obtained from HIV resistance testing

Reconstructed phylogeny



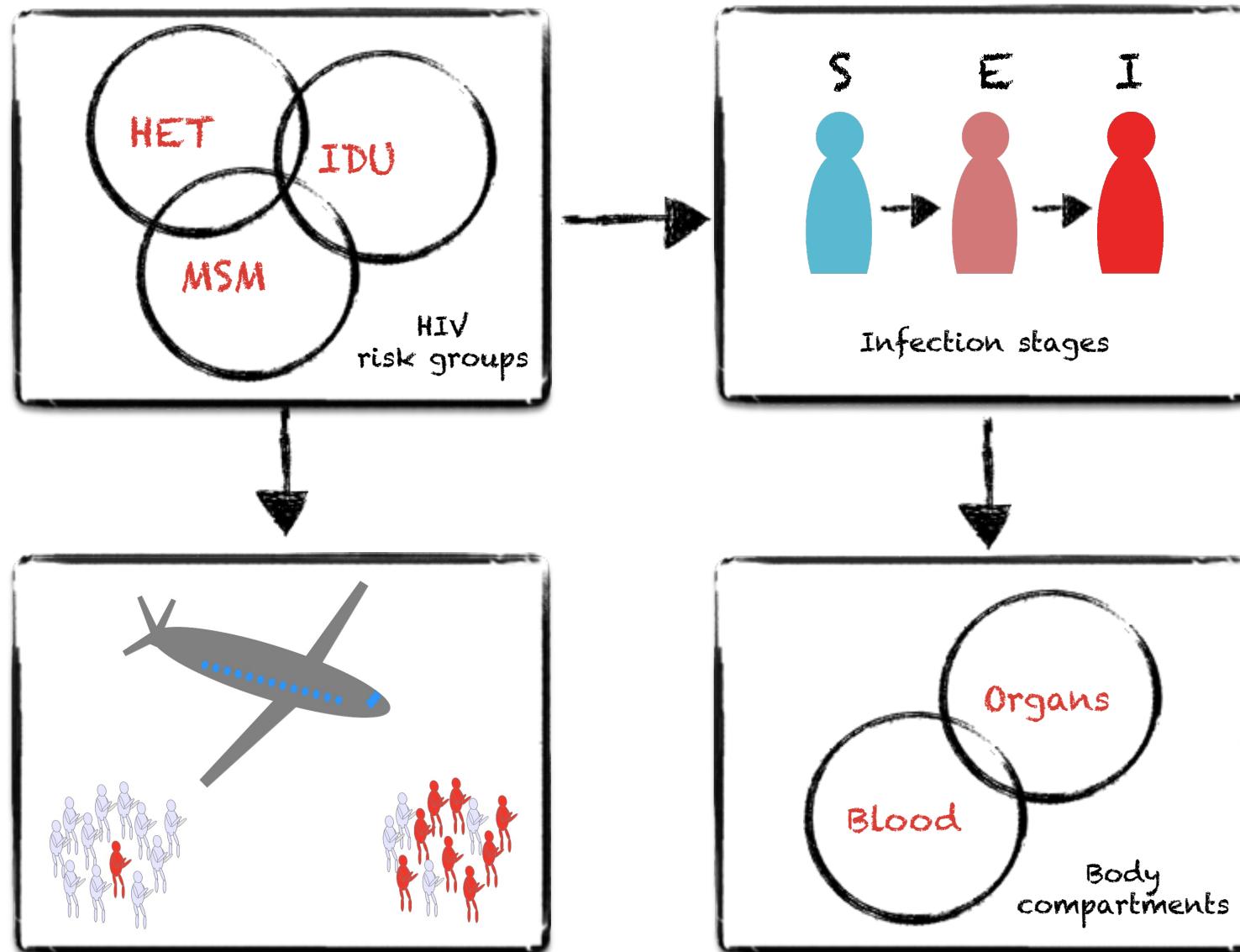
Epidemiological parameters



Joint reconstruction using Bayesian MCMC in BEAST



Infectious Disease Dynamics are Manifold





Structured population models – overview

- Migration process independent of tree prior
 - discrete trait analysis (“mugration”)
- Migration process part of tree prior
 - Structured coalescent
 - Structured coalescent approximations
 - BASTA
 - MASCOT
 - Structured birth-death model



"Mugration" – discrete trait analysis

Migration process inferred independent of the tree reconstruction





"Mugration" – discrete trait analysis

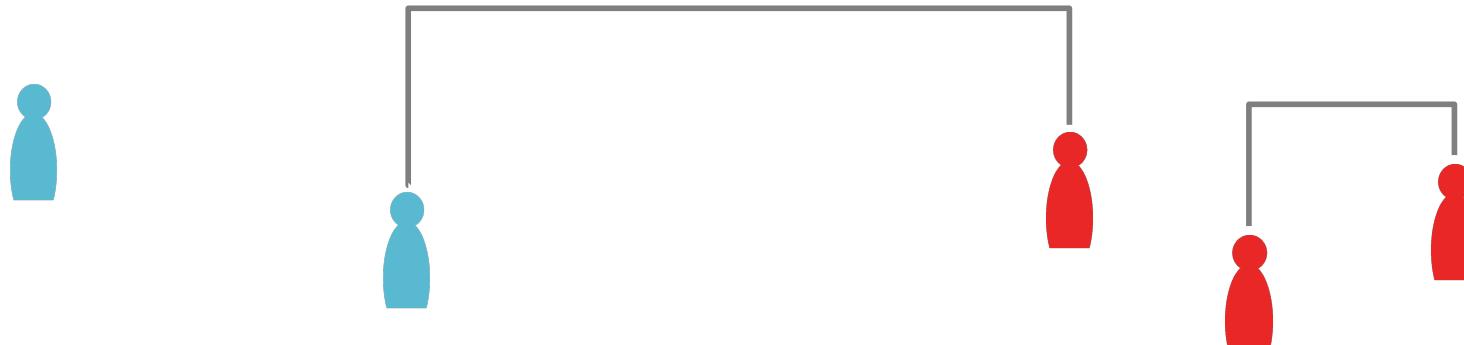
Step 1: Reconstruct tree with panmictic tree prior





"Mugration" – discrete trait analysis

Step 1: Reconstruct tree with panmictic tree prior





"Mugration" – discrete trait analysis

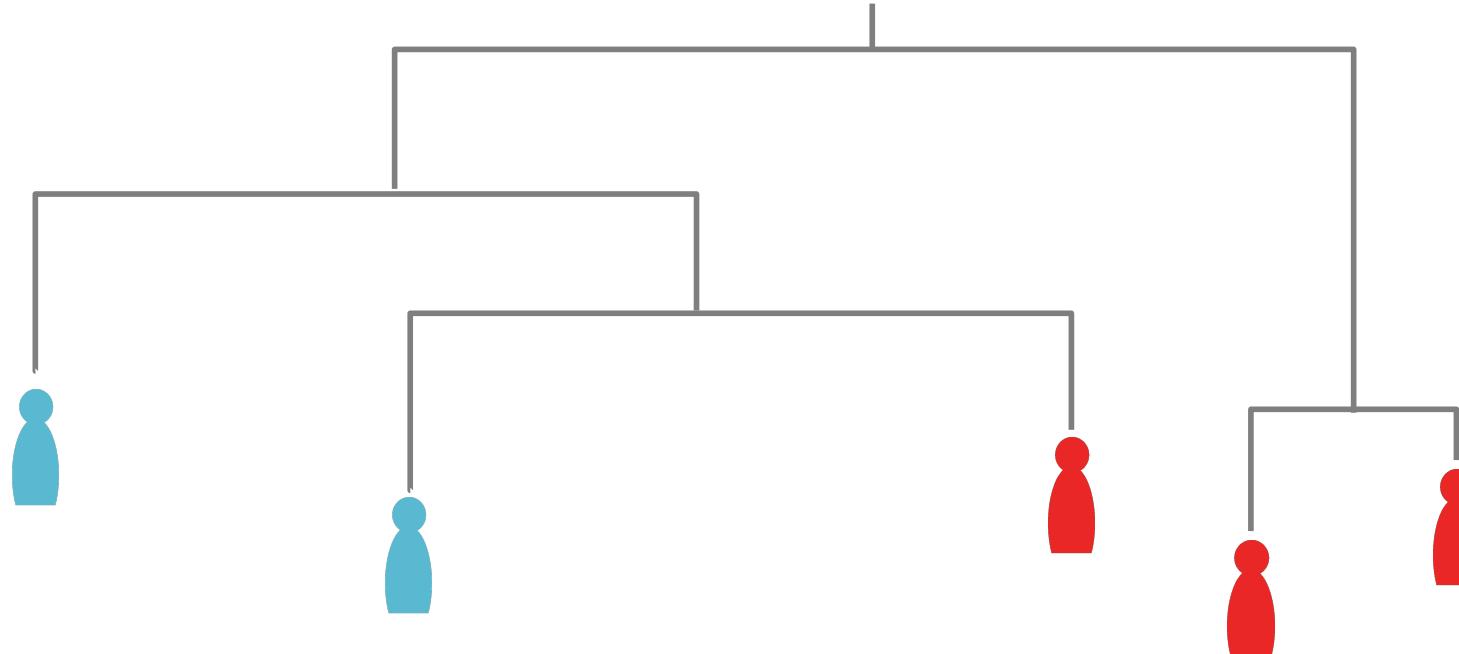
Step 1: Reconstruct tree with panmictic tree prior





"Mugration" – discrete trait analysis

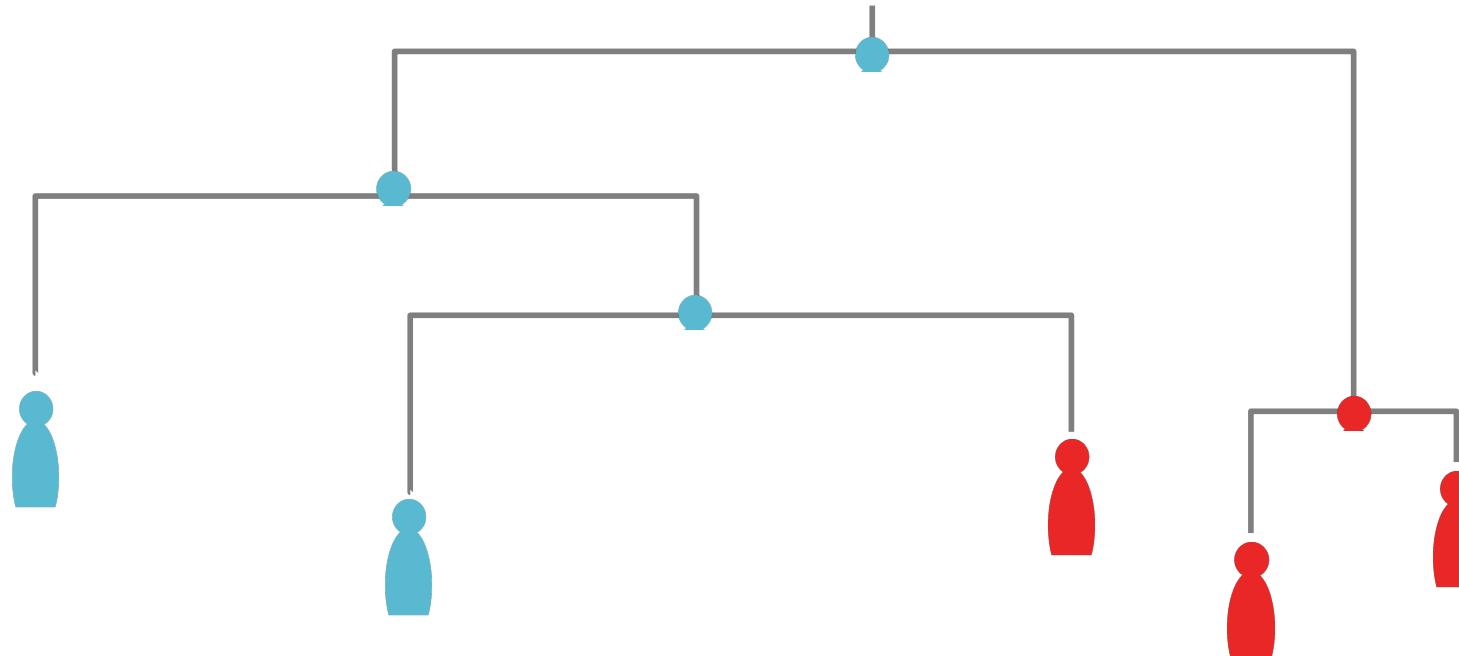
Step 1: Reconstruct tree with panmictic tree prior



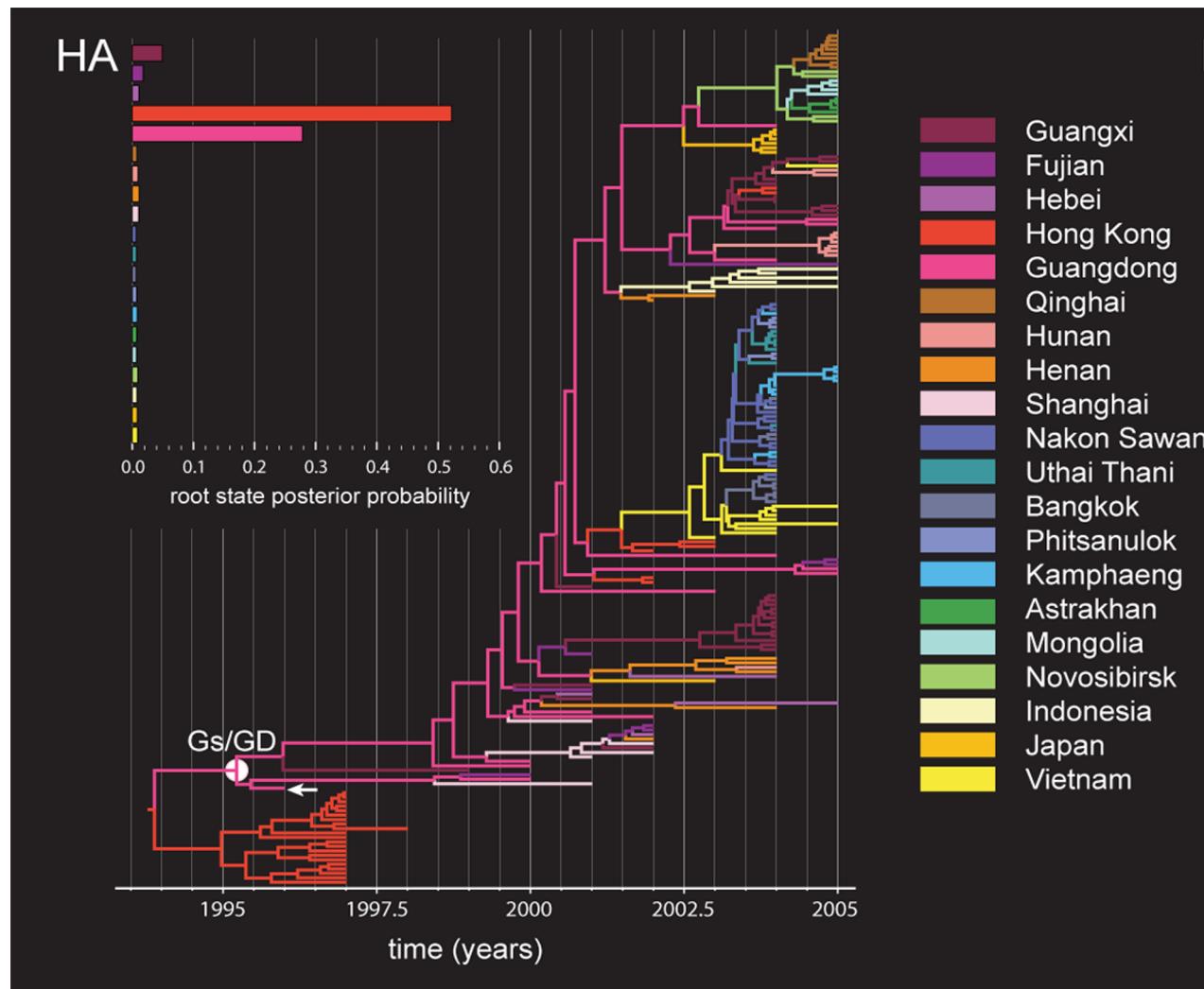


"Mugration" – discrete trait analysis

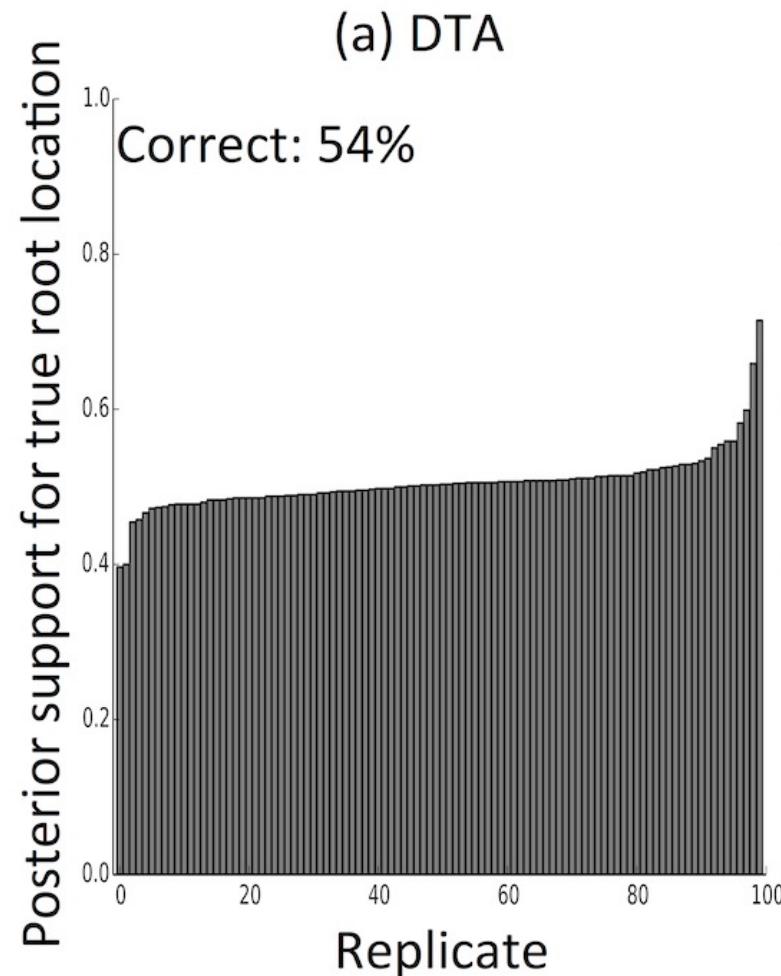
Step 2: Reconstruct migration process



The beginnings of Bayesian phylogeography



Ancestral reconstruction performance of discrete trait analysis





Structured coalescent

- Panmictic coalescent process within demes with migrations between demes



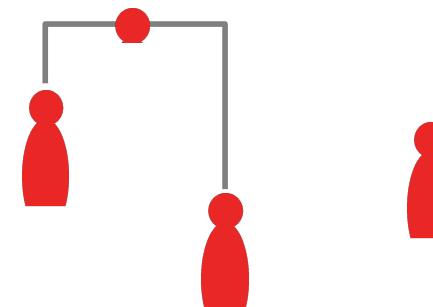
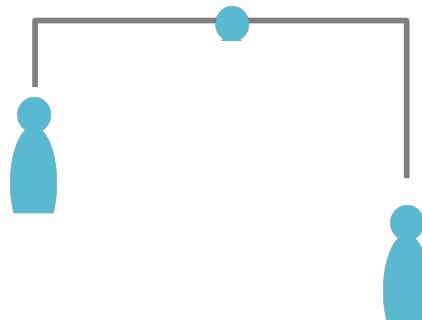
Structured coalescent

- Coalescence in red population



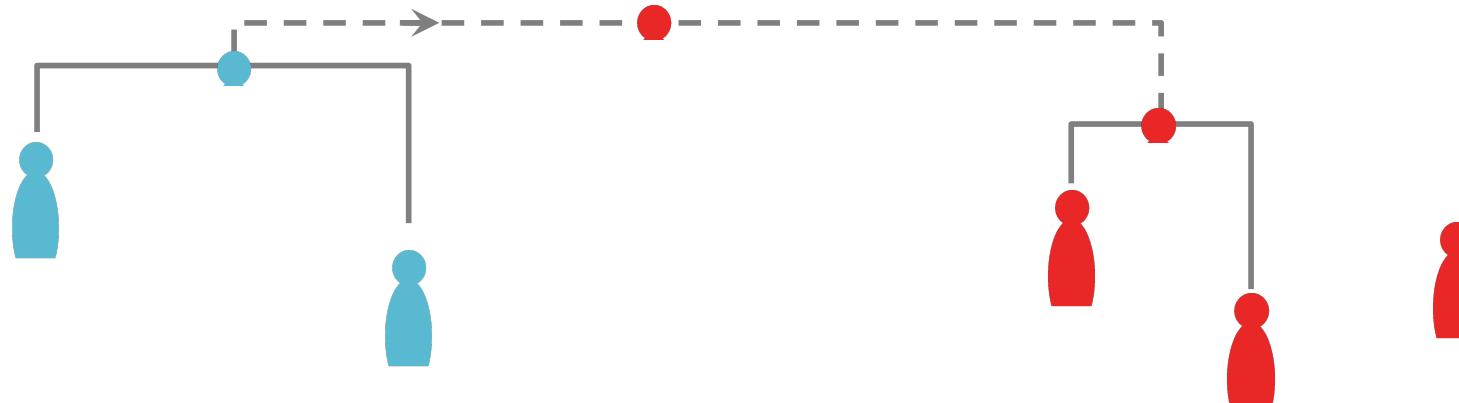
Structured coalescent

- Coalescence in blue population



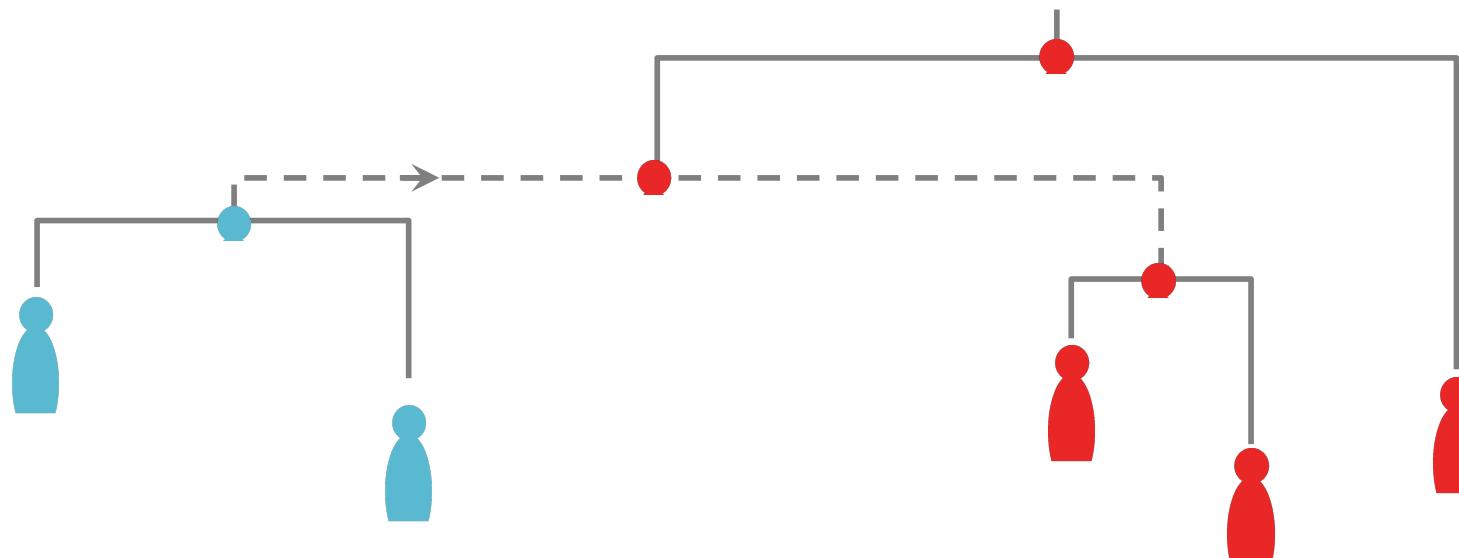
Structured coalescent

- Migration blue → red
 - followed by coalescence in red population

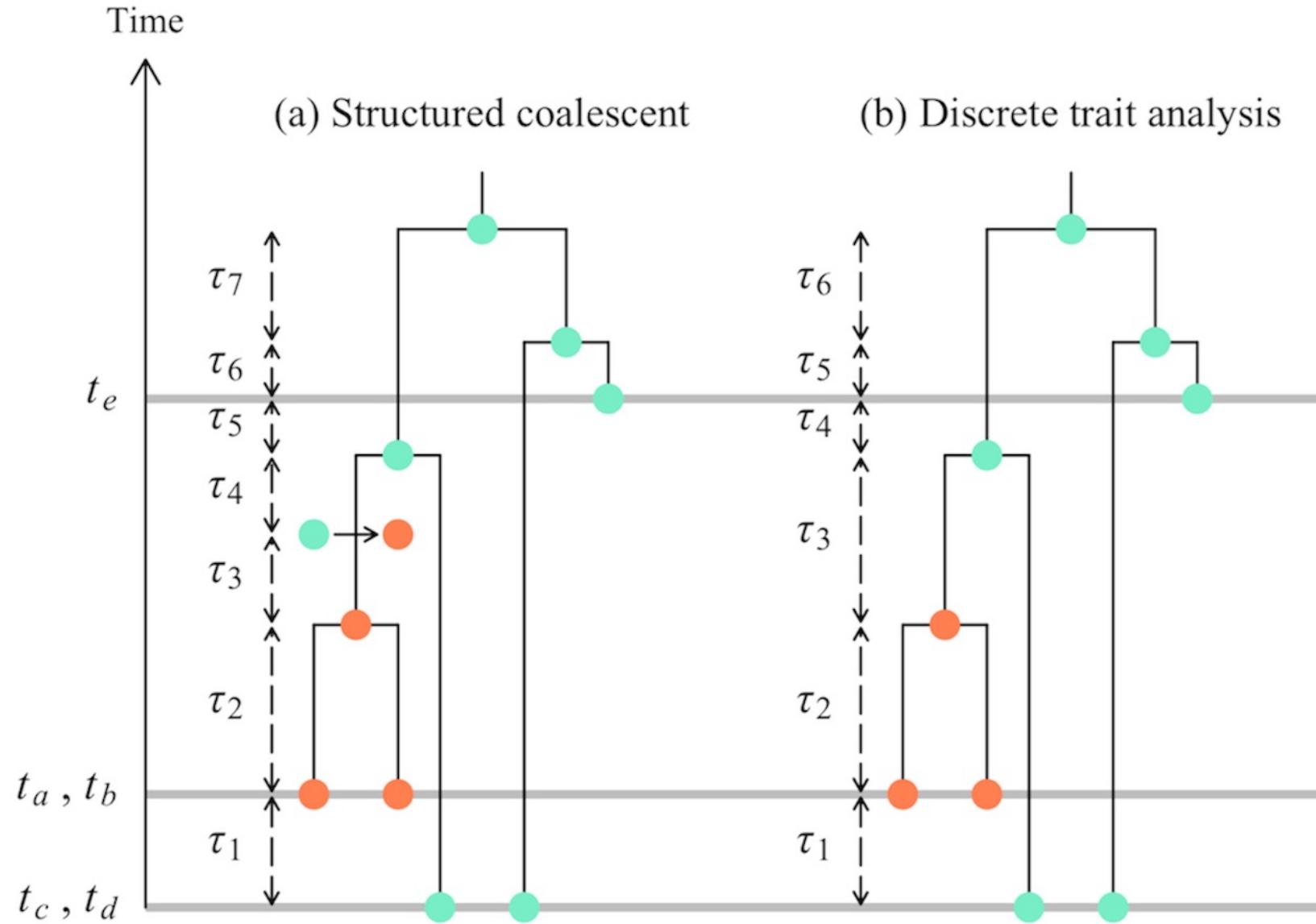


Structured coalescent

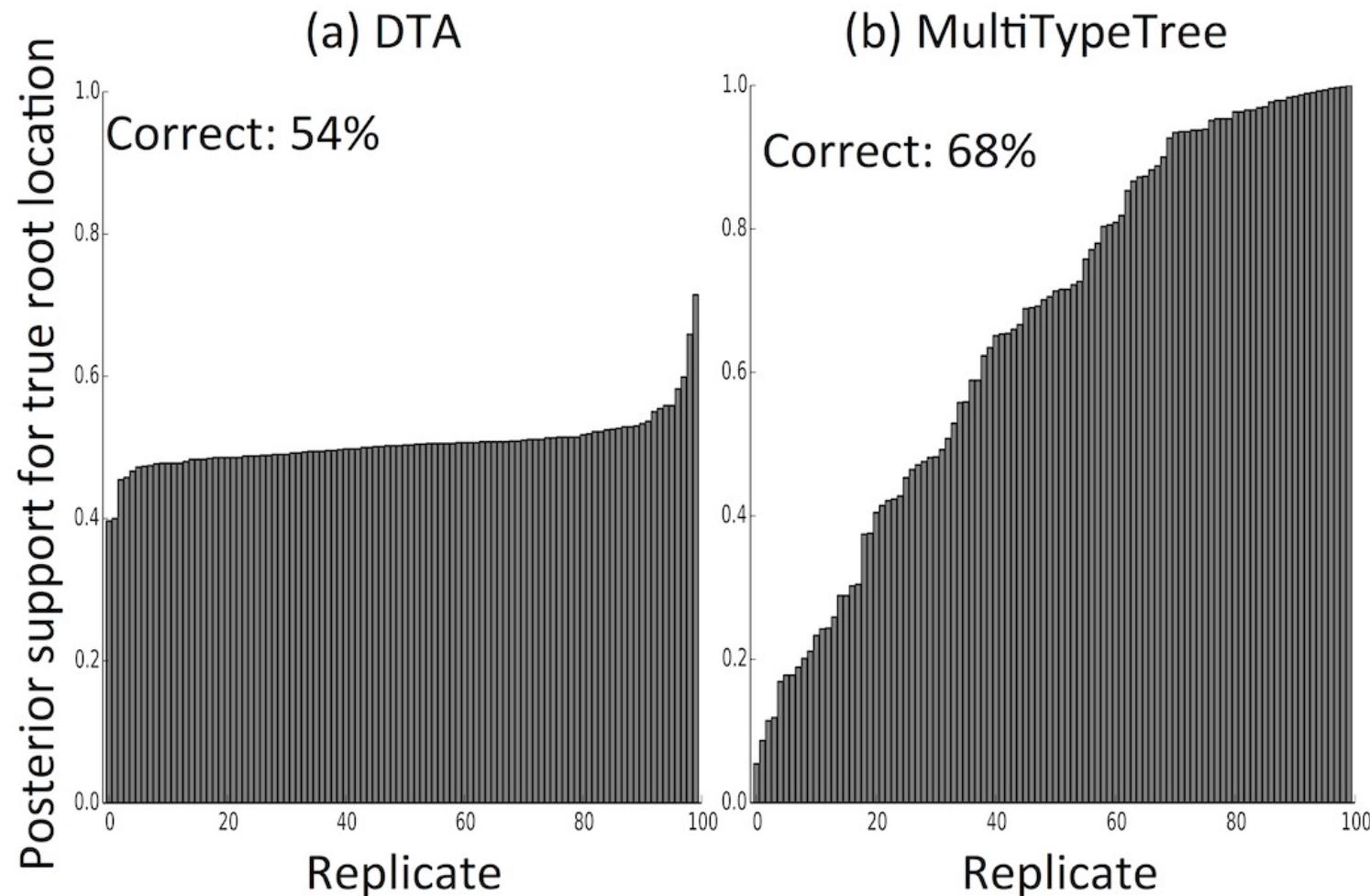
- Final coalescence event



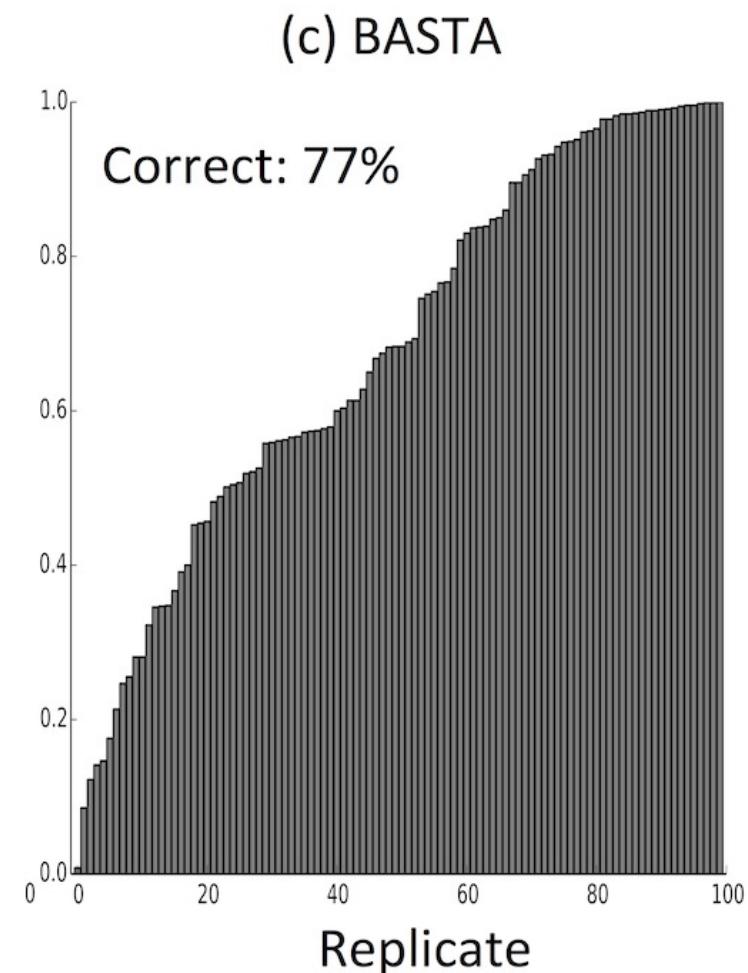
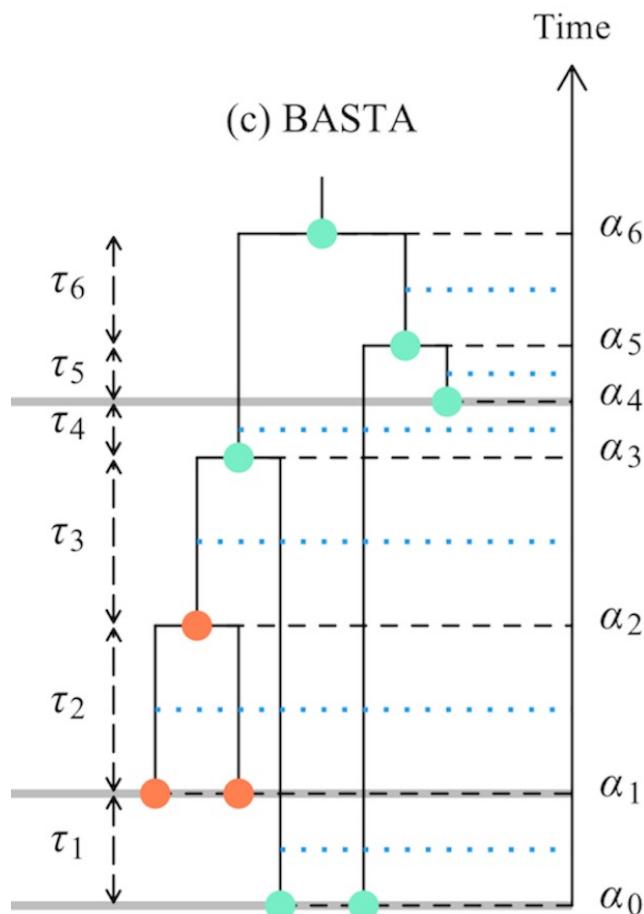
Structured coalescent versus “Mugration”



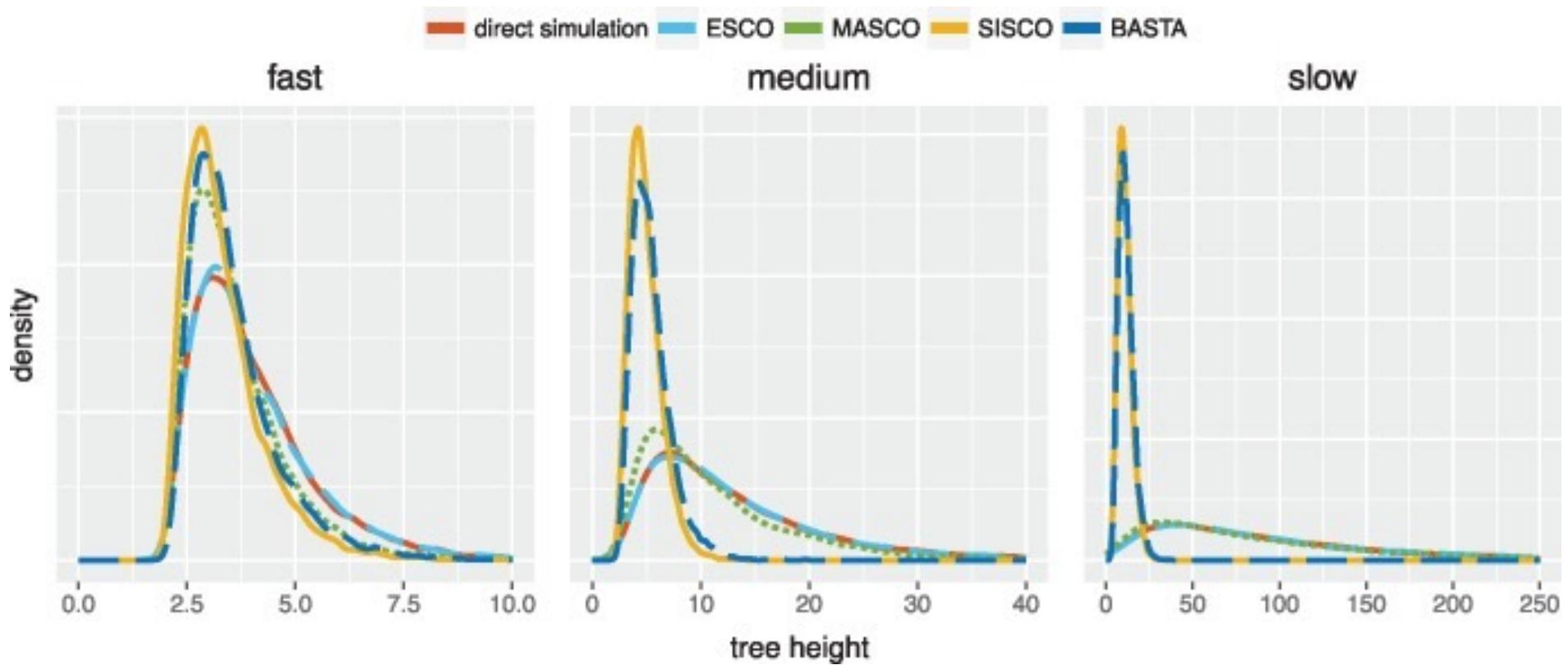
Ancestral reconstruction performance



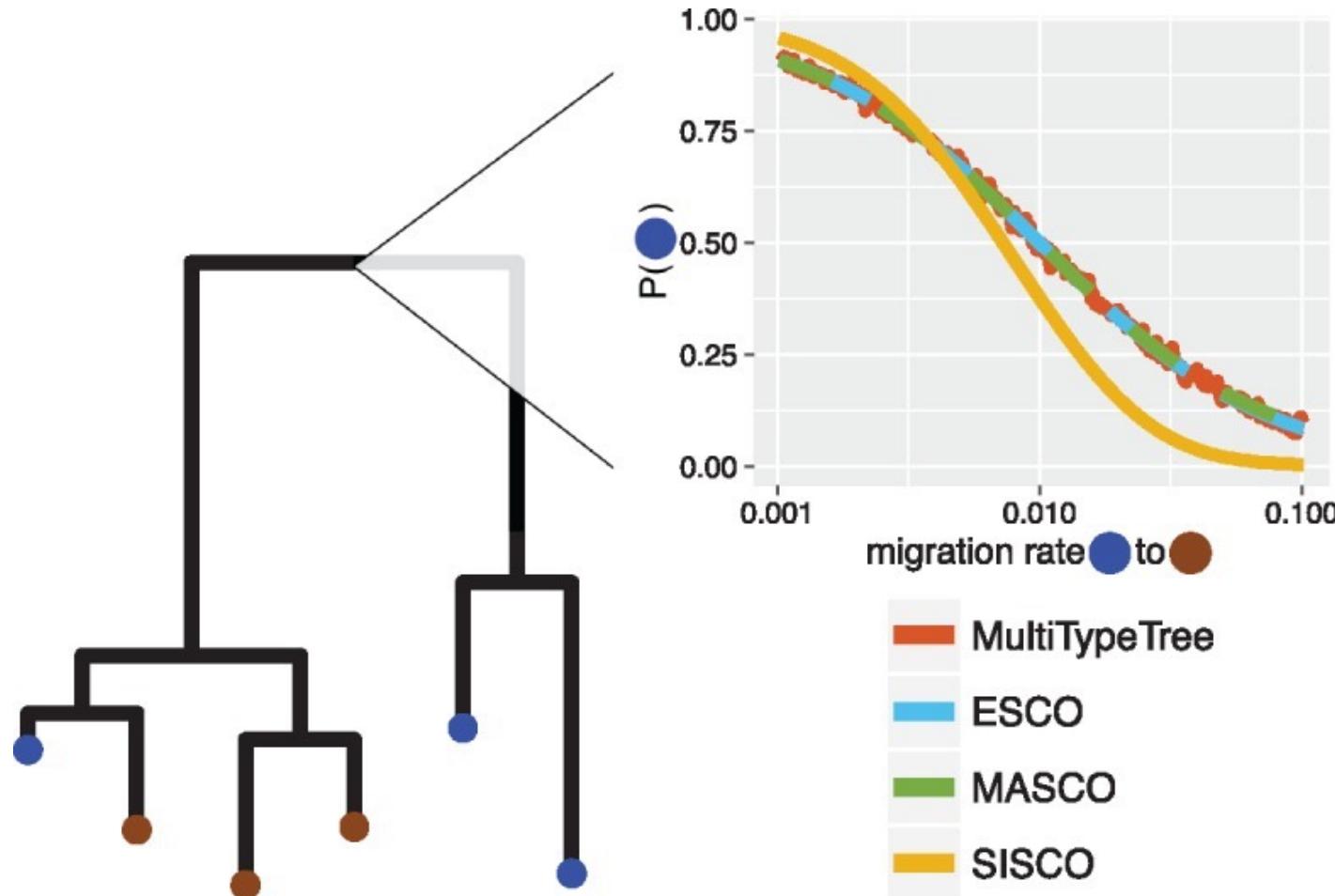
BASTA: Bayesian Structured coalescent Approximation



Biased tree height estimation when migration is slow



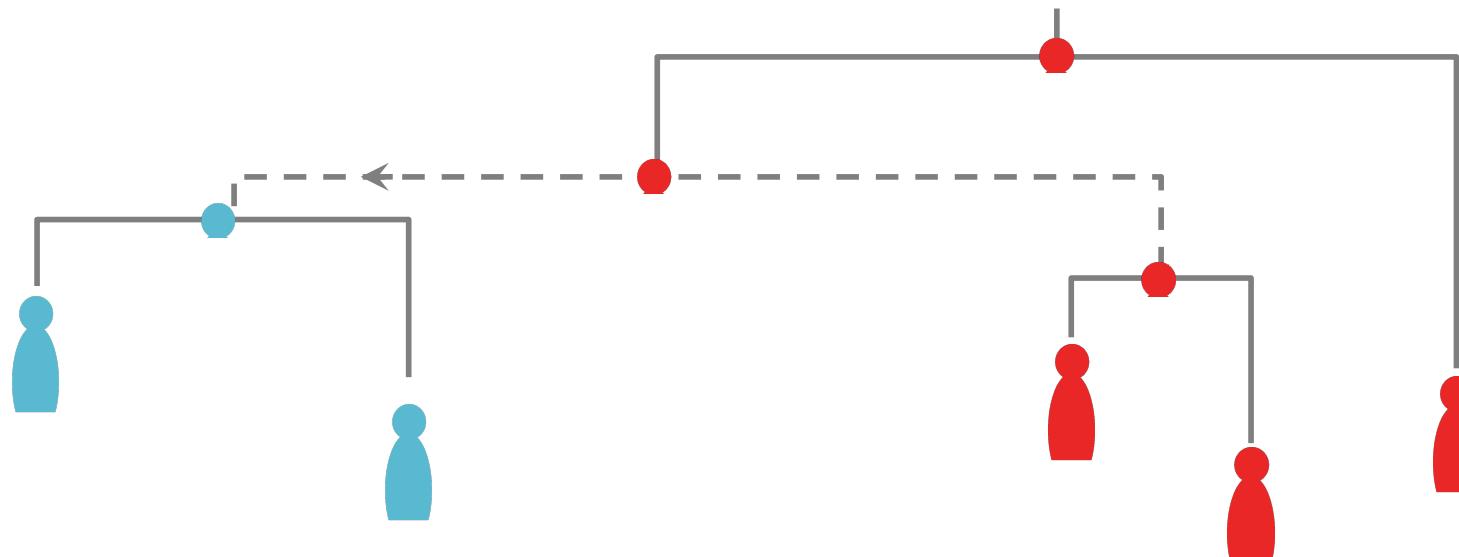
MASCOT: Marginal states approximation of the structured coalescent





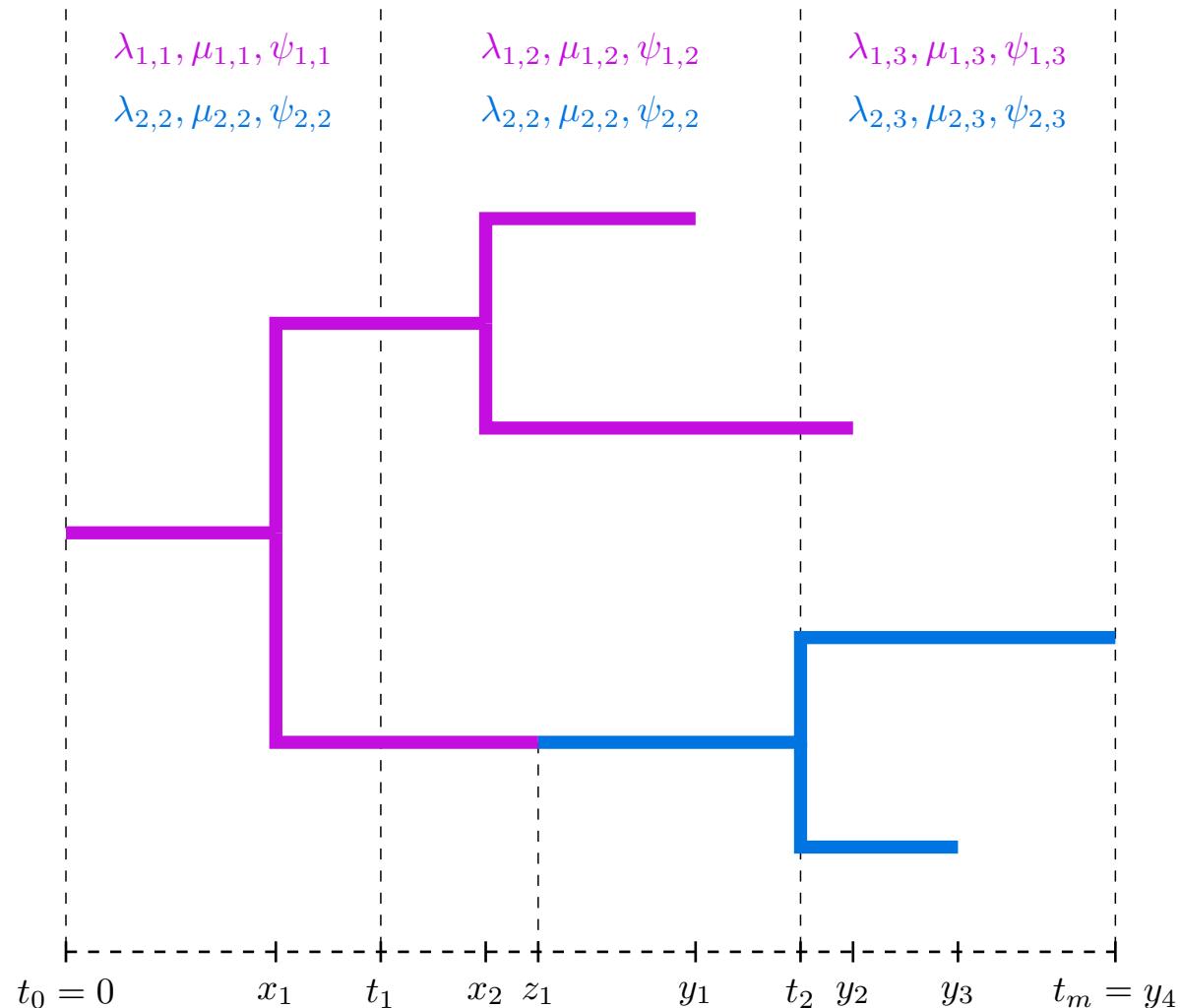
Structured birth-death model

Forward-time process



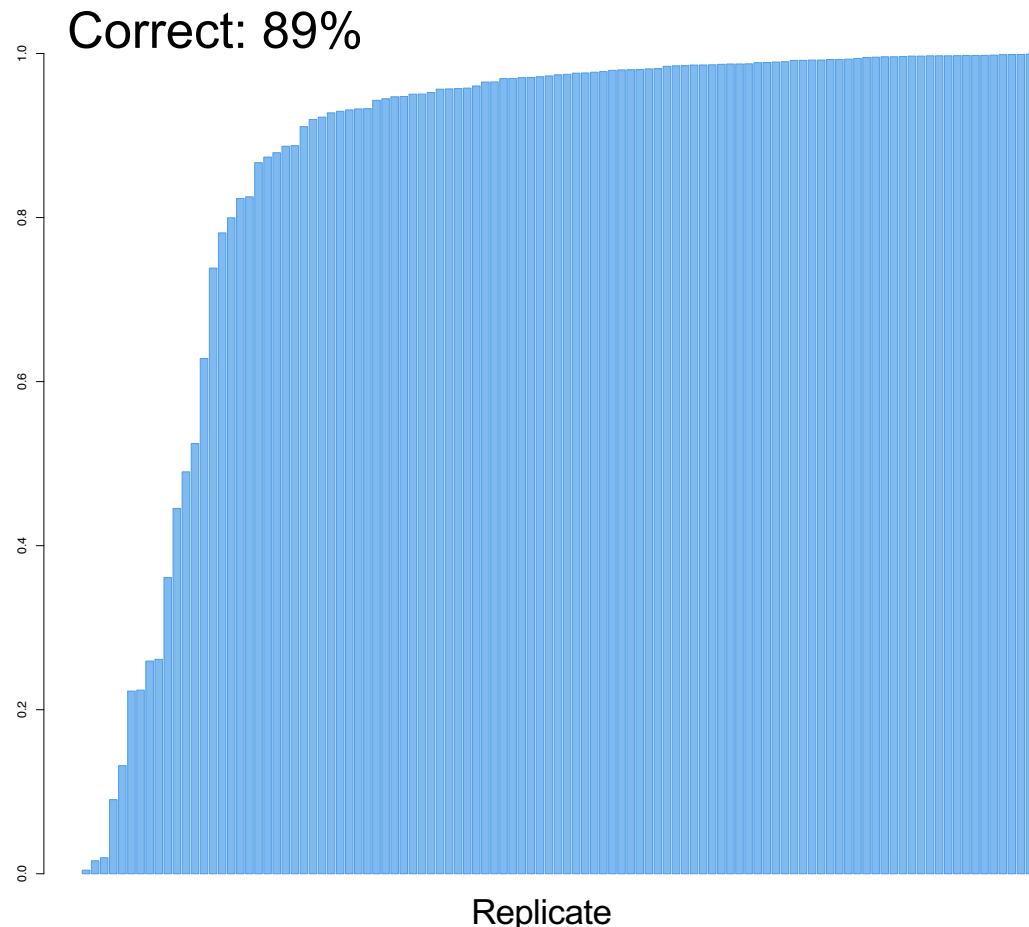


Structured birth-death model (bdmm)

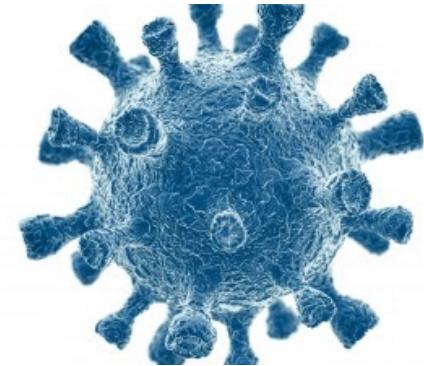




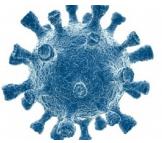
BDMM Ancestral reconstruction performance



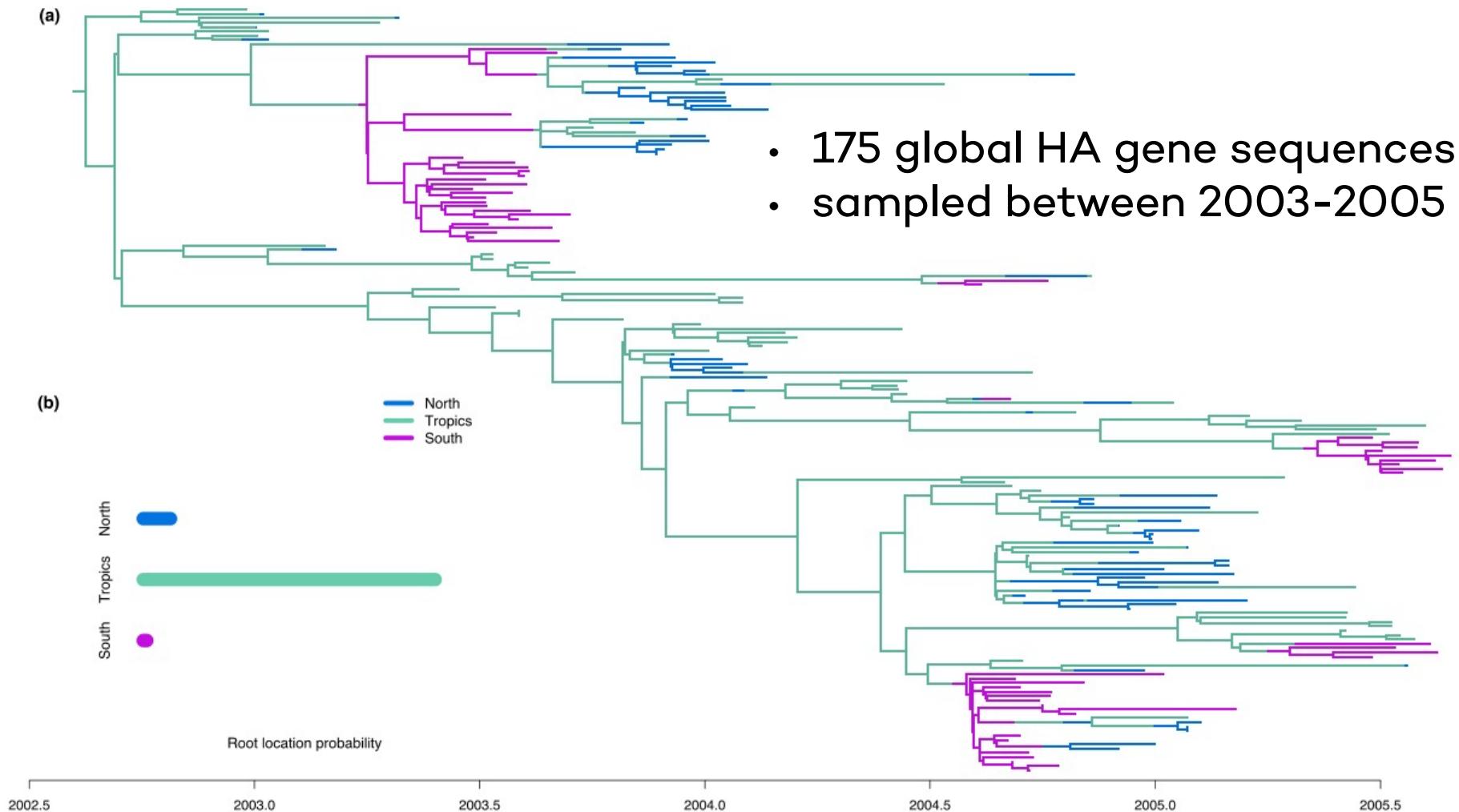
From 2-deme simulations in Kühnert et al. MBE. 2016

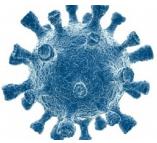


Global Influenza H3N2: Are the tropics the source?

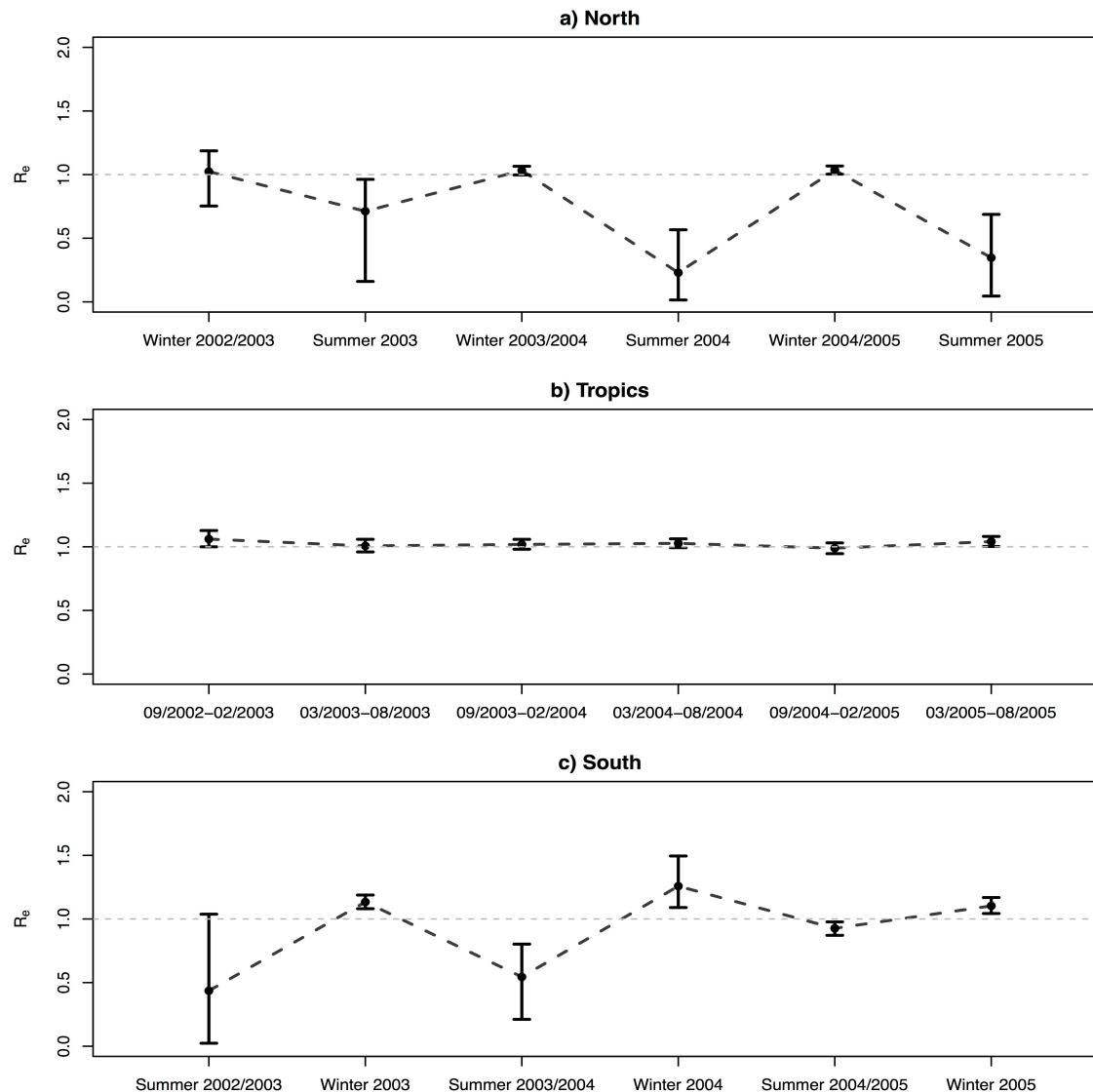


Global Influenza H3N2: Are the Tropics the Source?





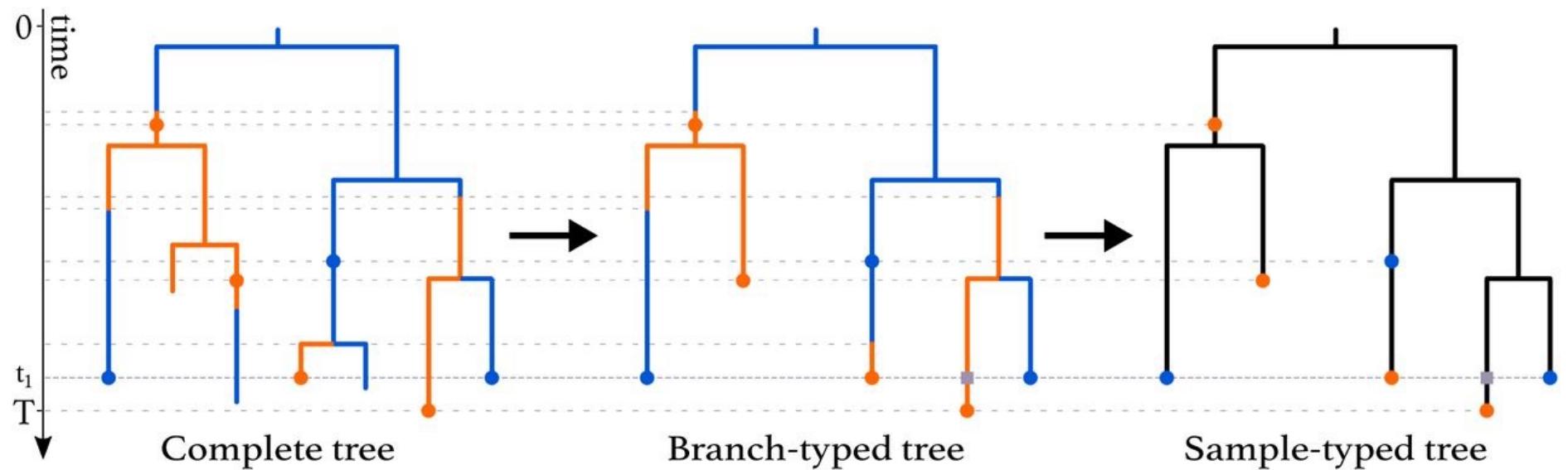
Global Influenza H3N2: Recovering Seasonal Dynamics





“BDMM 2020”

- Sampled-ancestor analysis
- Speed-up
- Bigger data sets ($>> 250$)





How to choose the right model?

- Research question?
- High or low number of demes?
- Random sampling scheme? Unbiased?
- Prior knowledge?
- ...

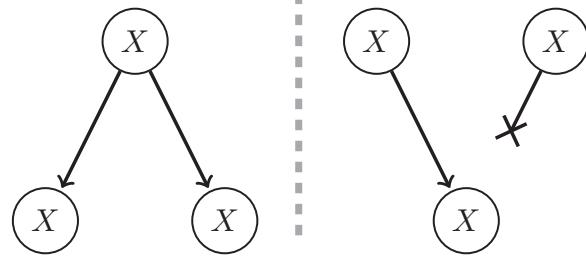
→ Test performance using simulations mimicking your own data



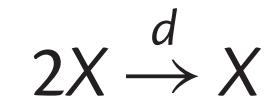
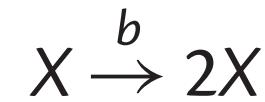
MASTER for phylodynamic simulations

(a)

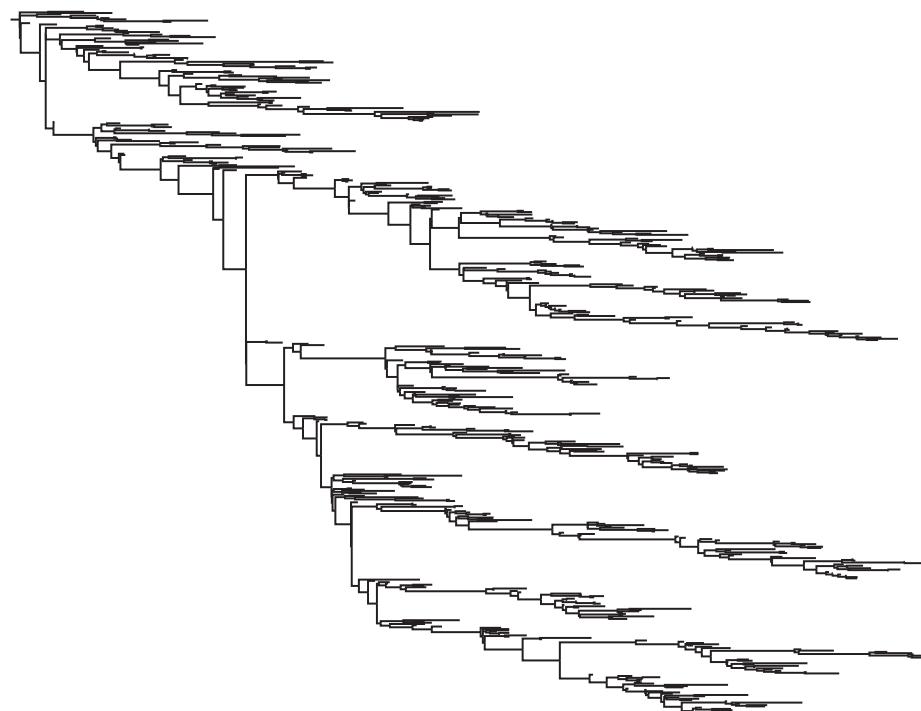
Birth



Death



(b)





MASTER simulation and BEAST inference possible in joint XML file

```
<reactionGroup spec='ReactionGroup' reactionGroupName='Birth'>
    <reaction spec='Reaction' rate="0.4">
        X[0] -> 2X[0]
    </reaction>
    <reaction spec='Reaction' rate="0.33">
        X[1] -> 2X[1]
    </reaction>
</reactionGroup>

<reactionGroup spec='ReactionGroup' reactionGroupName='Death'> <!-->
```

```
<reaction spec='Reaction' rate="0.33">
    X[1] -> X[0]
</reaction>
</reactionGroup>

<reactionGroup spec='ReactionGroup' reactionGroupName='Sampling'> <!-->
```

```
<reaction spec='Reaction' rate="0.001">
    X[0] -> X[0]
</reaction>
</reactionGroup>

<reactionGroup spec='ReactionGroup' reactionGroupName='Migration'> <!-->
```

```
<reaction spec='Reaction' rate="0.001">
    X[0] -> X[1]
</reaction>
</reactionGroup>

</model>

<initialState spec='InitState'>
    <lineageSeed spec='Individual'>
        <population spec='Population' type="@X" location="0"/>
    </lineageSeed>
</initialState>

<!-- Abort simulation if number of infected individuals drops to zero -->
<lineageEndCondition spec='LineageEndCondition' nLineages="0" isRejection="true"> <!-->
```

```
</lineageEndCondition>

<!-- Finish simulation when 100 samples have been generated. -->
<lineageEndCondition spec='LineageEndCondition' nLineages="100" isRejection="false"> <!-->
```

```
</lineageEndCondition>

<!-- Keep only sampled lineages. -->
<inheritancePostProcessor spec='LineageFilter' populationName="S"/>

<!-- Trim off post-sampling lineages. -->
<inheritancePostProcessor spec='LineageFilter' populationName="S"
    leavesOnly="true" discard="true"/>

<output spec='NewickOutput' fileName="BDMUC_treeSimAnalyzer_sim.newick"/>
<output spec='NexusOutput' fileName="BDMUC_treeSimAnalyzer_sim.nexus"/>
<!--><output spec='JsonOutput' fileName="BDMUC_treeSimAnalyzer_sim.json"/>
```

```
</tree>

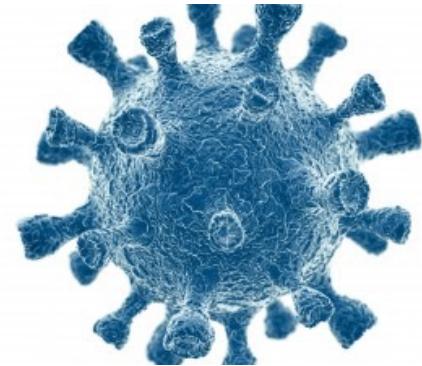
<data spec='beast.app.segen.SimulatedAlignment' id="alignment"
    data="@input_alignment"
    sequenceLength="2000"
    outputFileNames="BDMUC_SequenceSimAnalyzer_alignment.txt"
    trees="@masterTree">
    <sequence spec="Sequence" taxon="dummy" value="" />
    <siteModel spec="beast.evolution.sitemodel.SiteModel" id="SiteModel">
        <parameter name="lambda" value="0.001" />
        <substitutionModel spec="JukesCantor" />
    </siteModel>
</data>

<distribution spec="beast.evolution.speciation.BirthDeathMigrationModelUncoloured" id="treePrior" tree="@tree"
    typeLabel="type" stateNumber="2" conditionOnSurvival="true">
    <migrationMatrix spec="RealParameter" id="rateMatrix" value="0.01" dimension="2" lower="0" upper="100"/>
    <parameter name="frequencies" id="geo-frequencies" value=".5 .5" lower="0." upper="1." dimension="2"/>
    <parameter name="R0" id="R0" lower="0" dimension="2" value="3"/>
    <parameter name="becomeUninfectiousRate" id="becomeUninfectiousRate" value=".1" lower="0" dimension="2" />
    <parameter name="samplingProportion" id="samplingProportion" value="0.05 0.15" lower="0" dimension="2" upper="1."/>
</distribution>

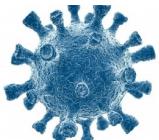
<RPNCalculator spec="beast.math.statistic.RPNCalculator" id="birth" expression="R0 becomeUninfectiousRate *" /> <!-- s/(d+s) -->
    <parameter idref="becomeUninfectiousRate" />
    <parameter idref="R0" />
</RPNCalculator>
<RPNCalculator spec="beast.math.statistic.RPNCalculator" id="sampling" expression="becomeUninfectiousRate samplingProportion *" />
    <parameter idref="becomeUninfectiousRate" />
    <parameter idref="samplingProportion" />
</RPNCalculator>
<RPNCalculator spec="beast.math.statistic.RPNCalculator" id="death" expression="becomeUninfectiousRate 1 samplingProportion - *" /> <!-- b=S0/(d+s) -->
    <parameter idref="becomeUninfectiousRate" />
    <parameter idref="samplingProportion" />
</RPNCalculator>

<run chainLength="5000000" id="mcmc" spec="MCMC" storeEvery="10000" numInitializationAttempts="100">

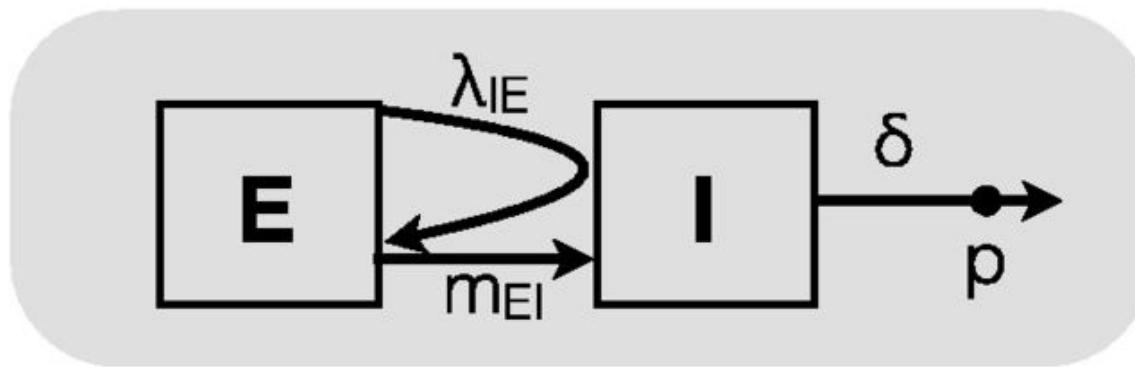
    <init spec="RandomTreeFromMaster" masterTree="@masterTree" taxa="@input_alignment">
        <tree spec="beast.evolution.tree.RandomTree" id="tree" taxa="@input_alignment">
            <populationModel spec="ConstantPopulation" />
                <parameter name="popSize" value="5.0" />
            </populationModel>
        </tree>
    </init>
    <state id="state" storeEvery="10000">
```



Modelling the exposed phase in SARS-CoV-2 outbreaks



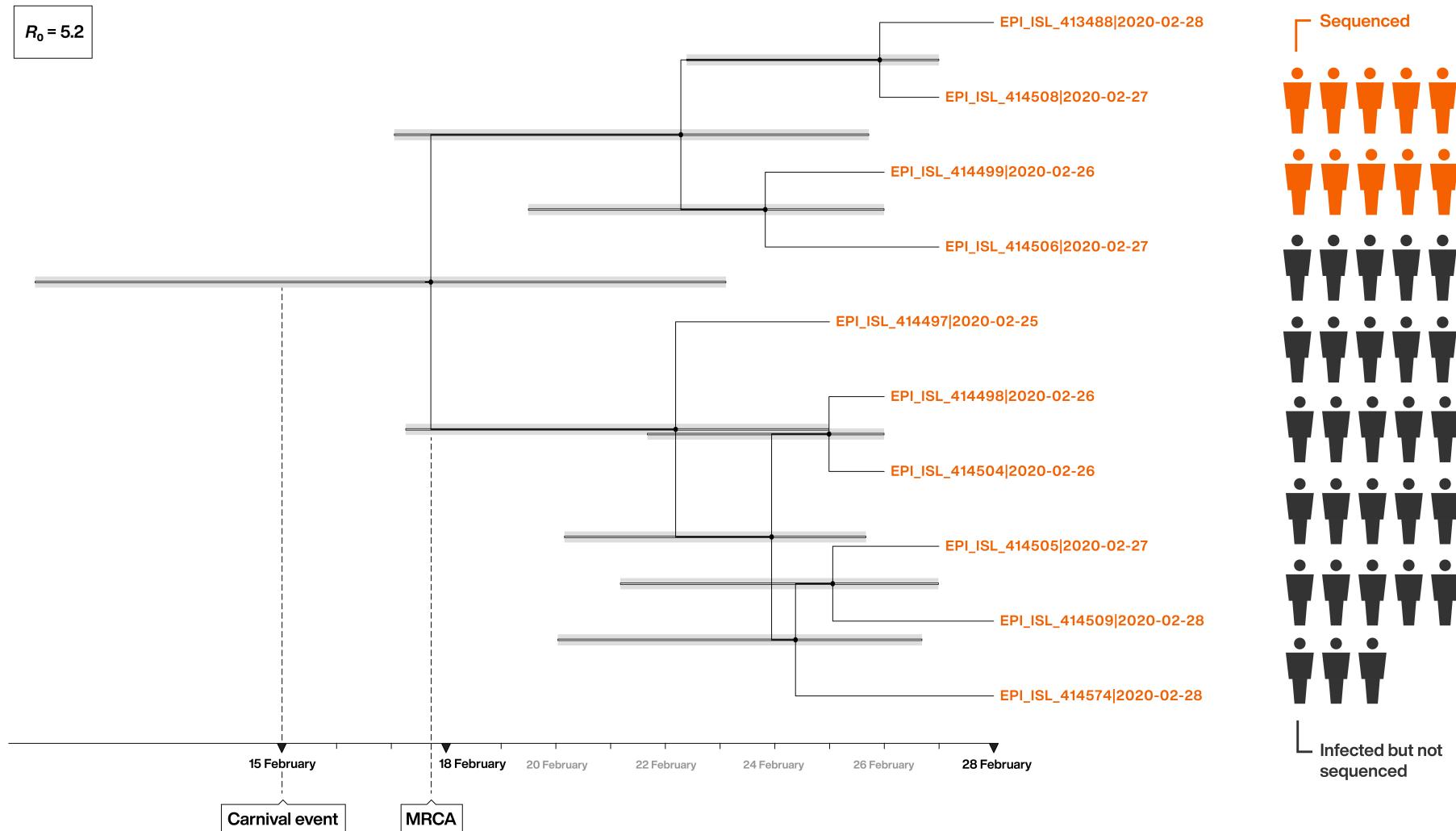
SARS-CoV-2 outbreak in Heinsberg modelled with exposed phase

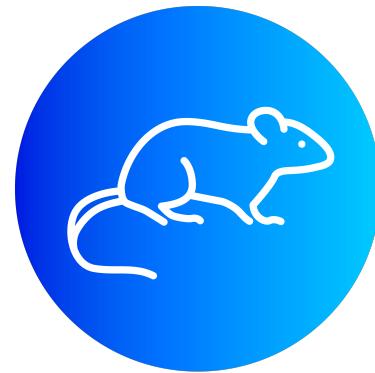


- Fixed clock rate: 8E-4
- Fixed incubation period: 2.5 days
- Narrow prior around infectious period: 11 days



Heinsberg outbreak – exemplary results from (too) small data set

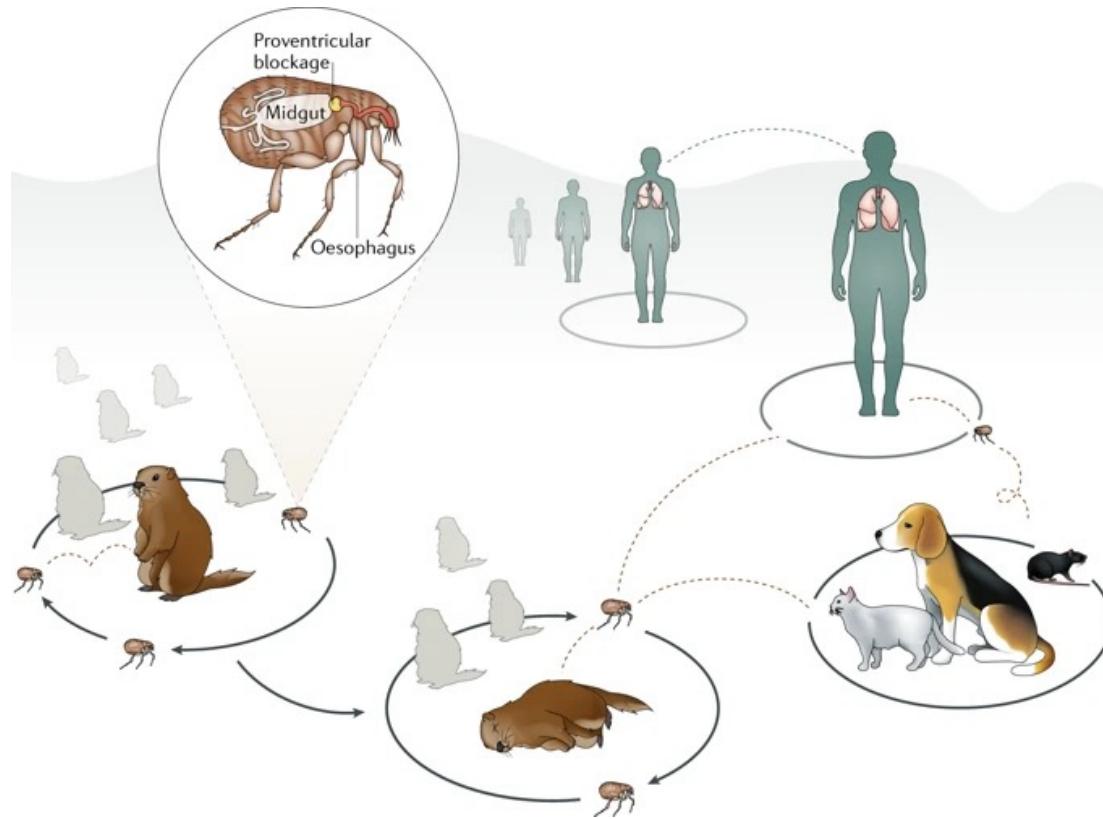




Phyloodynamics of *Yersinia pestis*



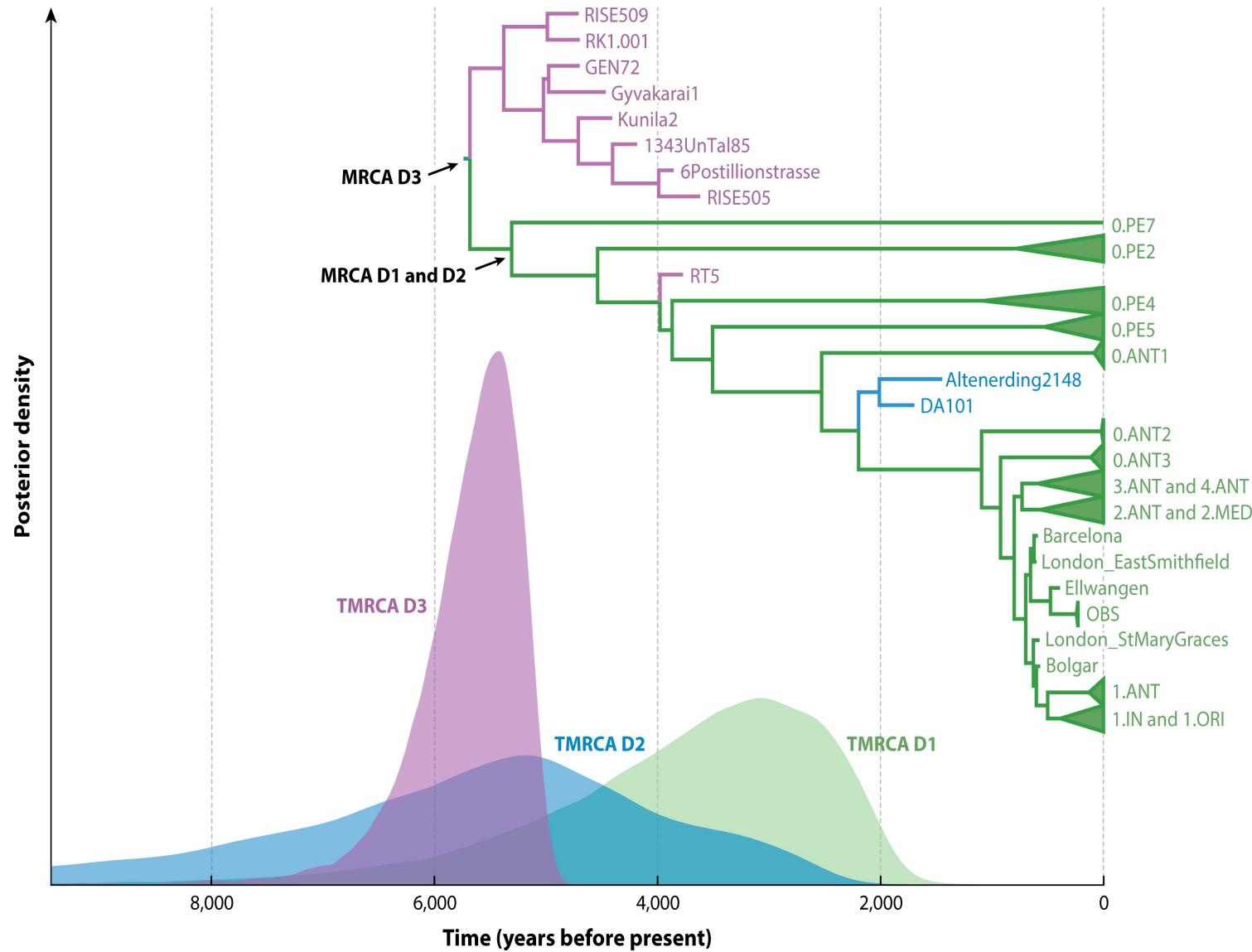
Plague transmission worldwide



Currently plague is considered endemic in regions of Asia, the Americas and Africa, with over 90% of the cases occurring on the African continent



Dating the origin of *Yersinia pestis*



Plague in Madagascar



- ~400 cases annually since 2010
- Disease introduced from India in 1898
- Reached the capital city Antananarivo in 1921

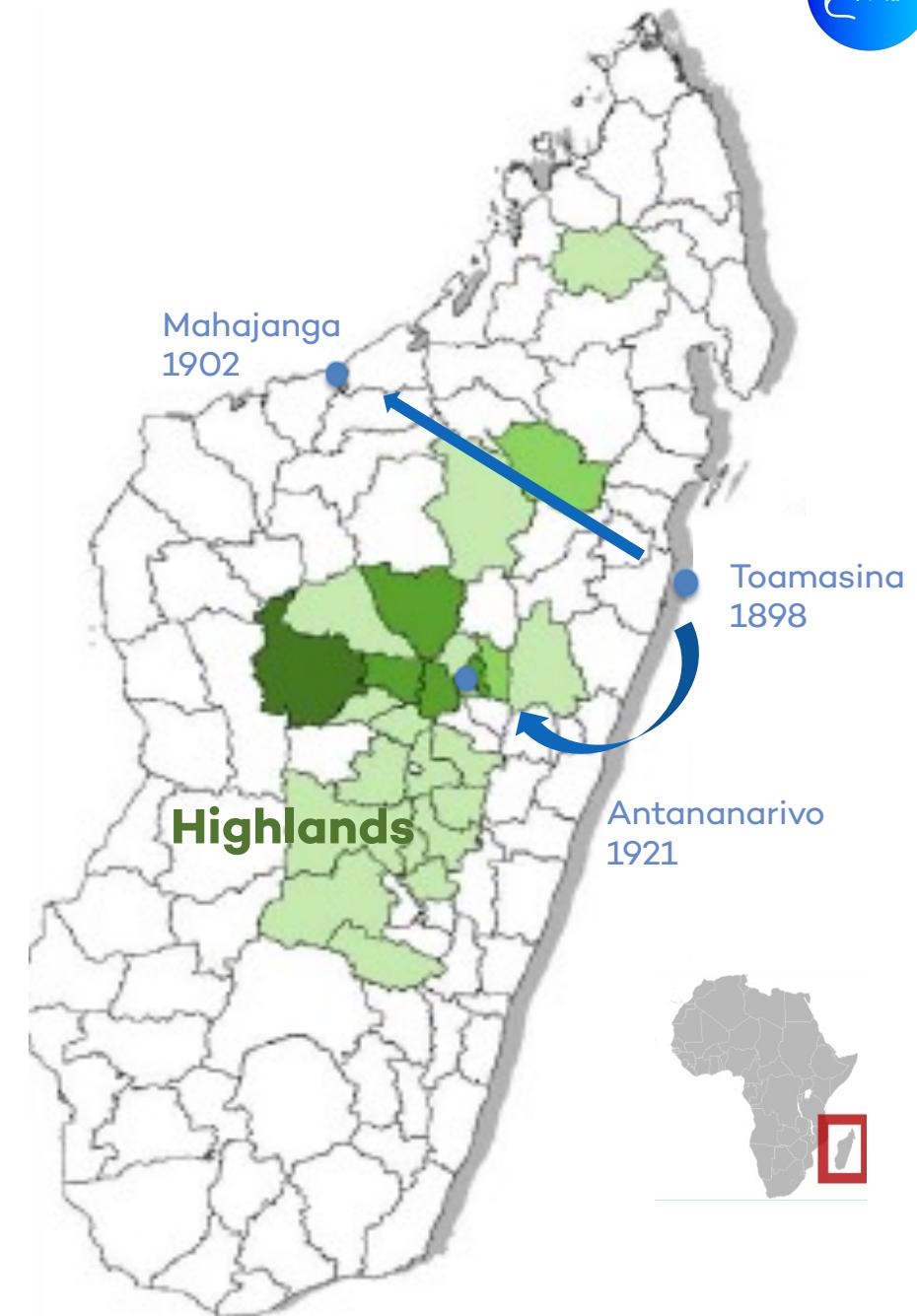


Figure adapted from Andrianaivoarimanana et al. 2013



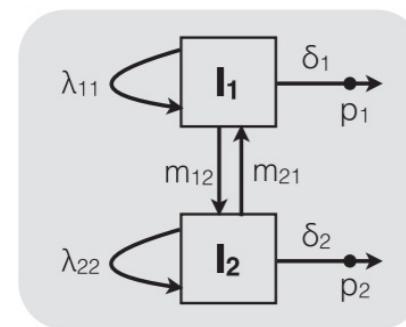
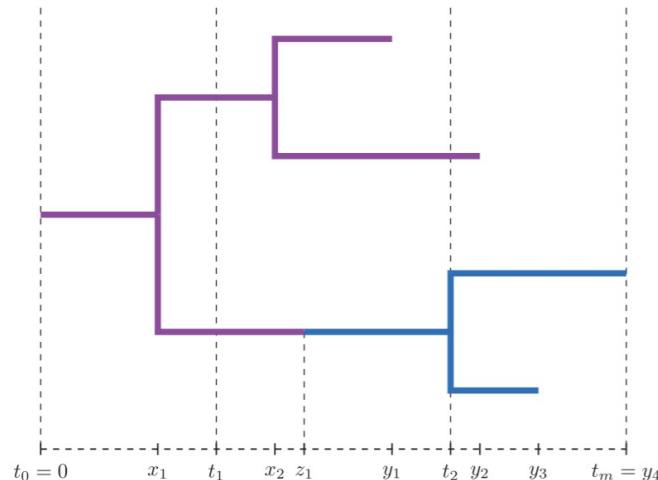
Re-introduction of plague in Mahajanga during the 90s?

- Outbreaks in 1902, 1907, between 1924 and 1928
- No human cases reported from 1929-1990
- Outbreaks in 1991/2 and 1995-1999
- No reports of rat epizootics (die-offs) in the coastal regions after plague “disappeared”
- Rats from the coastal regions show no development of plague resistance



Phylogenetics of *Yersinia pestis* in Madagascar

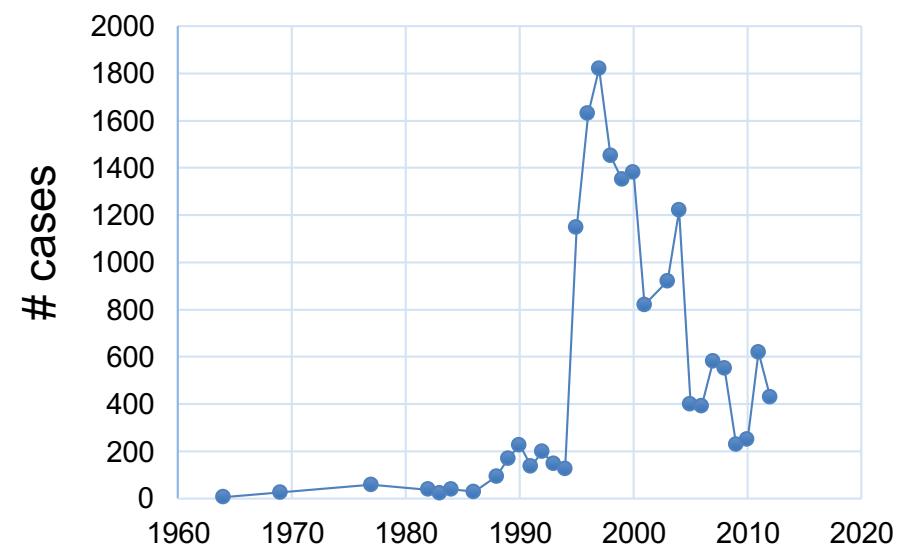
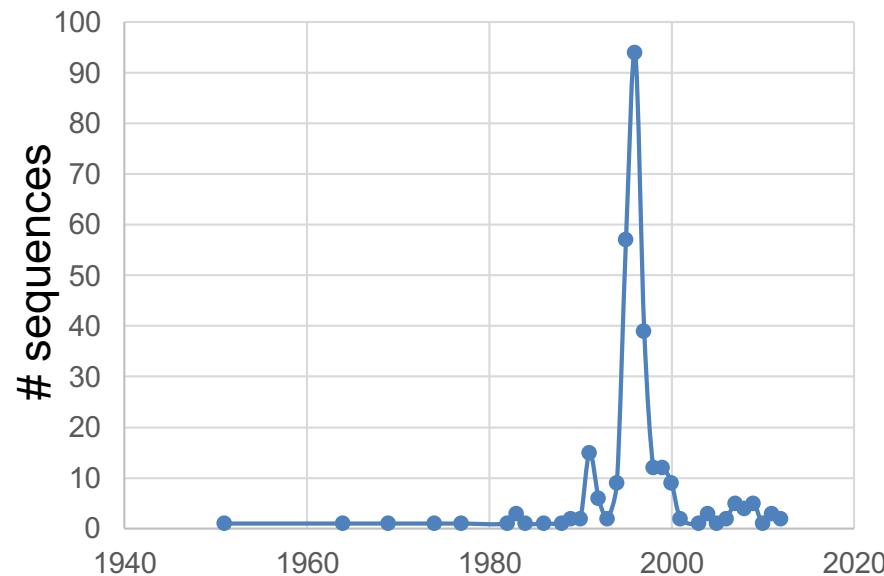
- 301 genomes sampled from Humans between 1951 and 2012
 - 184 from the Central Highlands
 - 117 from Mahajanga
- Temporal signal (BETS) ✓





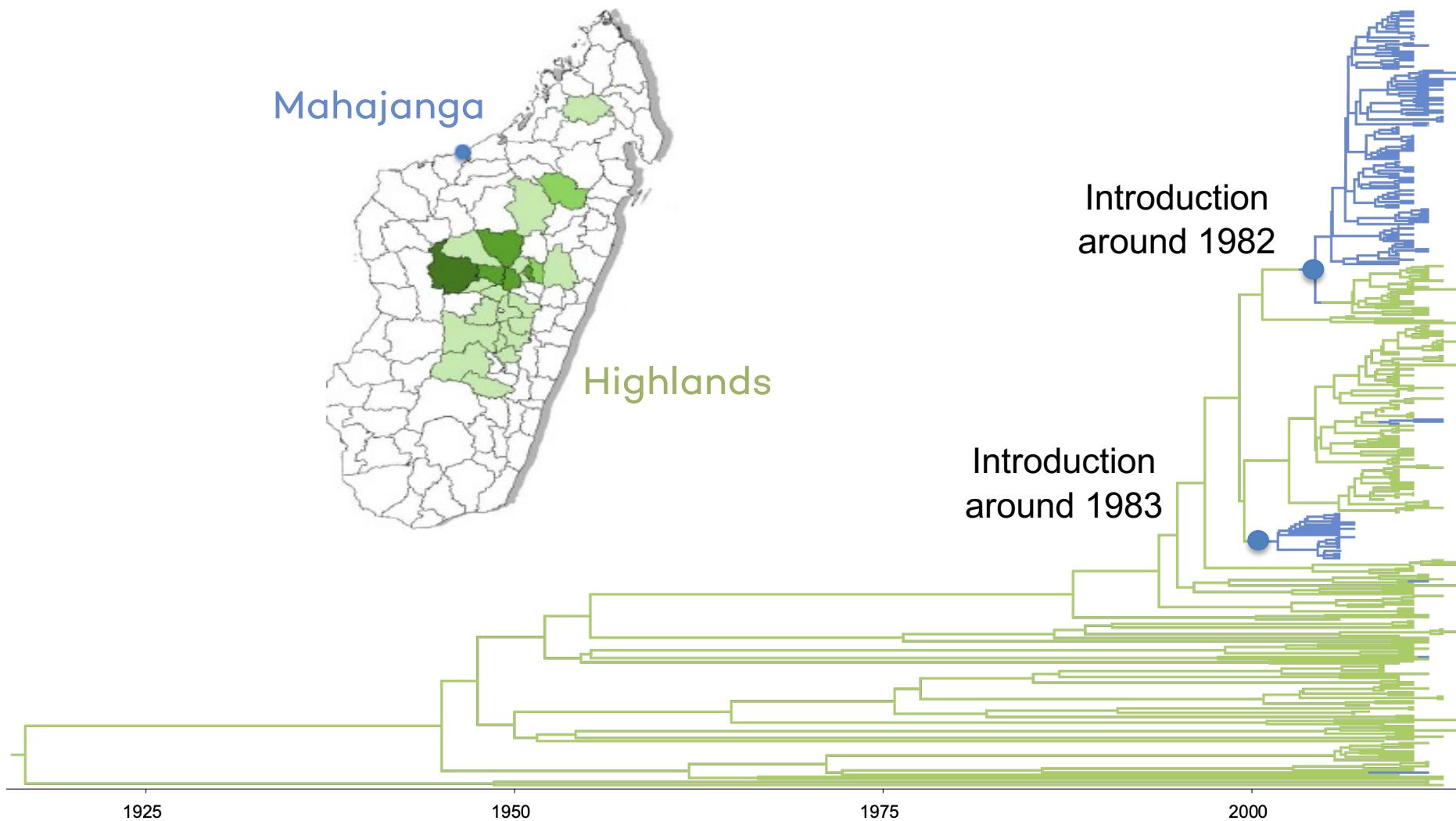
Using incidence data as prior information

- Sampling probability per (sampling) year based on observed incidence data





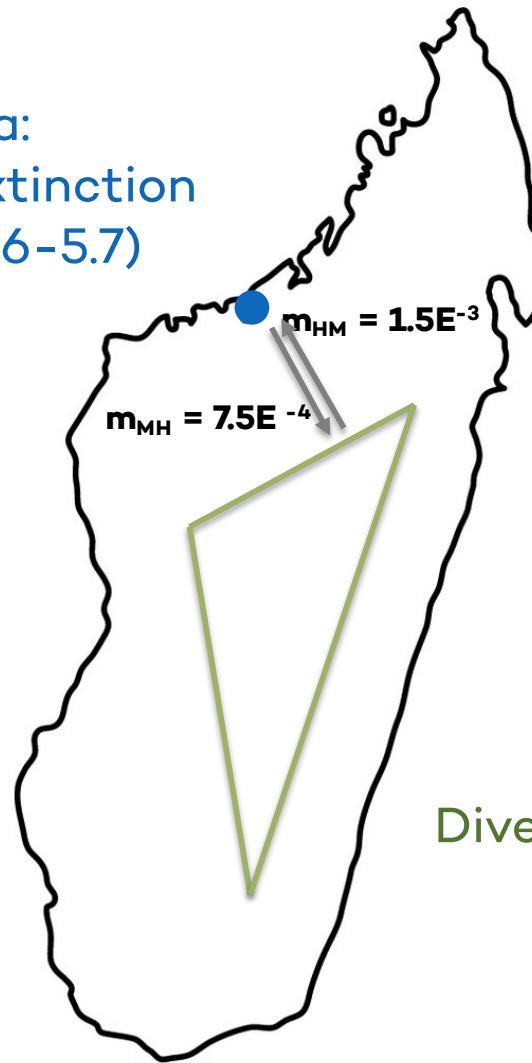
Two independent introductions from the Highlands into Mahajanga





More diversification in Mahajanga

Mahajanga:
Diversification/extinction
ratio **d = 5.2 (4.6-5.7)**



Highlands:
Diversification/extinction
ratio **d = 2.9**
(2.6 – 3.2)



Yersinia Pestis Summary

- Introduction of *Y.pestis* to Madagascar during late 19th - early 20th century confirmed
- Faster “transmission” of *Y. pestis* in Mahajanga
- More migration from Highlands to Mahajanga than vice versa
- The 1991-92 and the 1995-99 outbreaks represent independent introductions from the Highlands, with early seeding (local reservoir?)

Related tutorials



Species Trees with Relaxed Molecular Clocks

Estimating per-species substitution rates using StarBEAST2

Huw A. Oggilvie

Updated 21 Jun 2019 by tgvaughan

Structured birth death model

Population structure using the multi-type birth-death model

Denise Kühnert and Jūlija Pečerska

Updated 19 Nov 2020 by tgvaughan

Calibrated Species Trees

Estimating species divergence times using StarBEAST2

Huw A. Oggilvie

Updated 3 Feb 2017 by Unknown

MASCOT v2.1.2 Tutorial

Parameter and State inference using the approximate structured coalescent

Nicola F. Müller

Updated 24 Jul 2020 by nicfel

Structured coalescent

Population structure using MultiTypeTree

Nicola F. Müller and Tim Vaughan

Updated 12 Jul 2017 by laduplessis

The MSBD package

Inferring lineage-dependent birth and death rates

Joëlle Barido-Sottani

Updated 3 Aug 2020 by bjoelle

Questions?

Acknowledgements

tide

Luis Esquivel Gomez
Arthur Kocher
Ariane Weber

MPI archaeogenetics dept.

Johannes Krause
Alexander Herbig
Kirsten Bos

Warwick
Mark Achtman
Xavier Didelot

ETH Zürich
Tim Vaughan
Jeremie Scire
Tanja Stadler



Transmission, Infection, Diversification & Evolution Group

