

Lecture 3.2

Analysing Populations

Simon Ho

Coalescent Theory

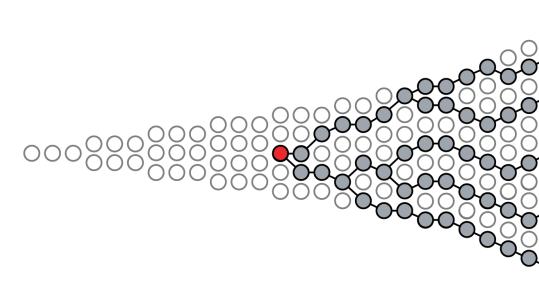
Levels of analysis

Interspecific	Intraspecific
One sequence per species	Sequences from one species
High sequence divergence	Low sequence divergence
Estimating species tree (phylogeny)	Estimating gene tree
Tree shape described by speciation process	Tree shape described by coalescent process
Gene trees often* congruent	Gene trees usually incongruent
Rates vary among branches	Rates constant among branches

2

Coalescent theory

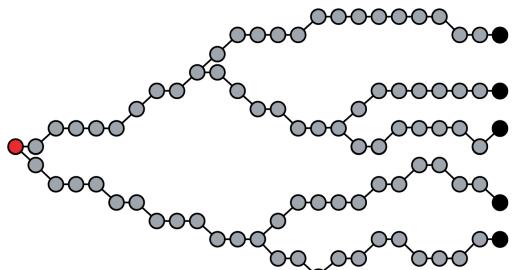
- Coalescent model used to put a prior on the tree
- Time between coalescent events depends on population size



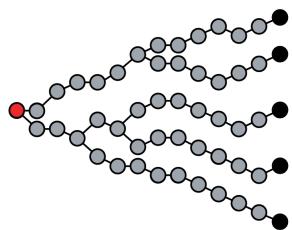
4

Coalescent theory

Constant size



Exponential growth



5

Coalescent theory

- Assumptions

- Random sampling

- No population subdivision

- No selection

- Non-overlapping generations

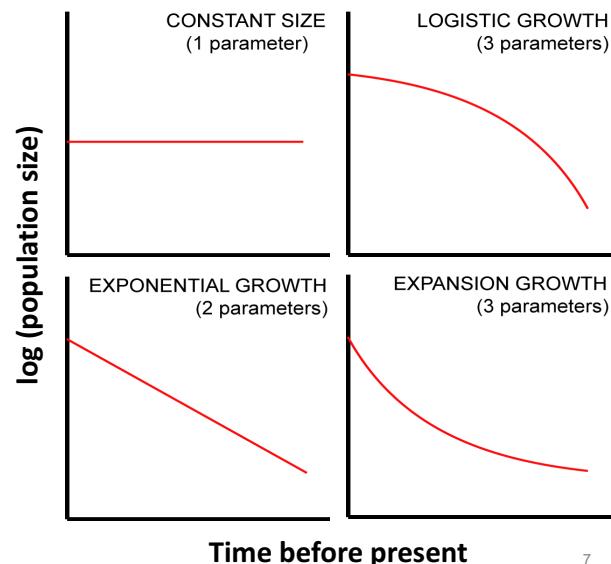
- No recombination

Panmictic population

6

Demographic models

- Several population models available in *BEAST*
- Can compare using Bayes factors



7

Demographic models

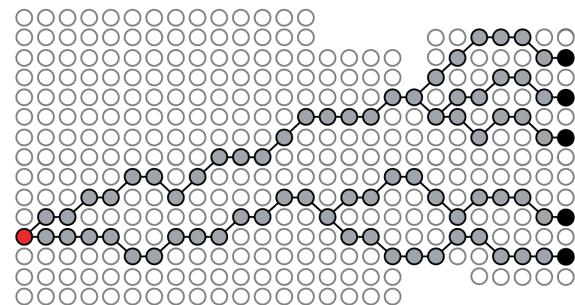
- The demographic history of a population leaves a signature in the DNA of its modern representatives
- Reconstructing this history might be of interest
 - Testing correlations with abiotic factors
 - Examine factors driving population dynamics
 - Tracing transmission and spread of viruses

8

Skyline Plots

Skyline-plot methods

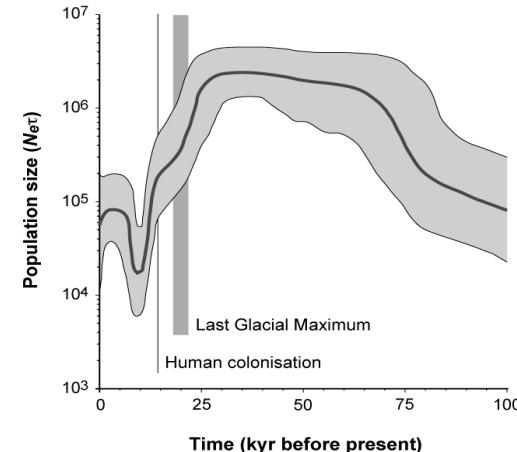
- Given a sequence alignment, demographic reconstruction comprises two separable steps:
 - Estimation of the genealogy from the alignment
 - Estimation of population history from the genealogy



11

Skyline-plot methods

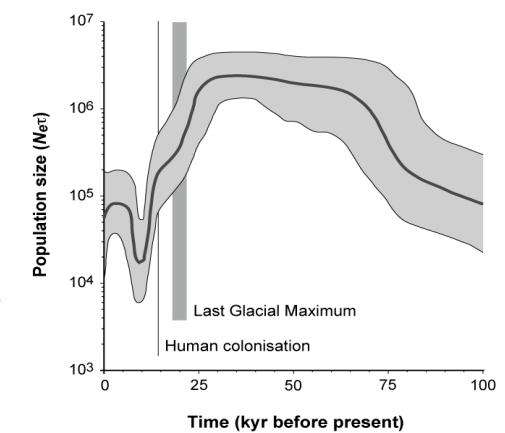
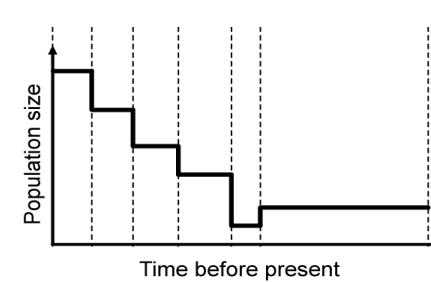
- In some cases it is inappropriate to limit our investigation to a small range of simple parametric models



Drummond et al. (2005) MBE

10

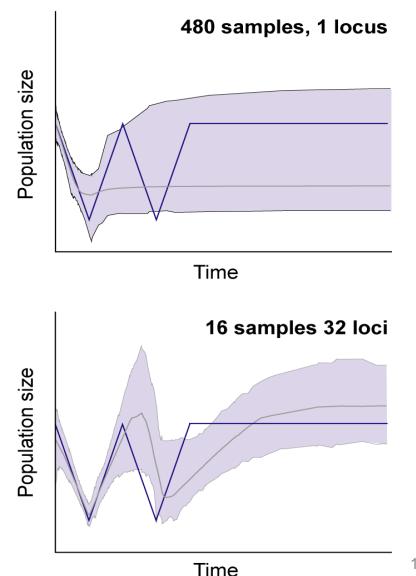
Skyline-plot methods



12

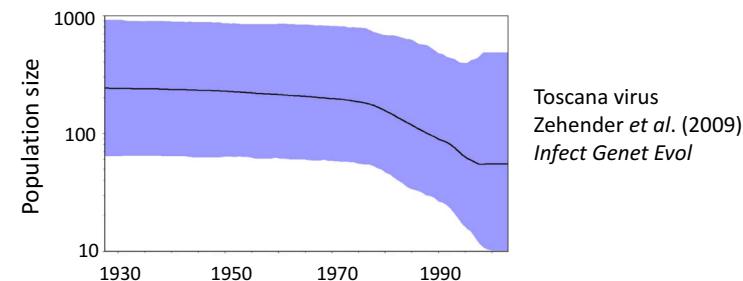
Extended Bayesian skyline

- Extended Bayesian skyline allows simultaneous analysis of multiple loci
 - Distinct substitution models
 - Distinct substitution rates
 - Distinct genealogies
- Advantages of multiple loci
 - Reduce coalescent error
 - Increase power to resolve bottleneck events



Evaluating support

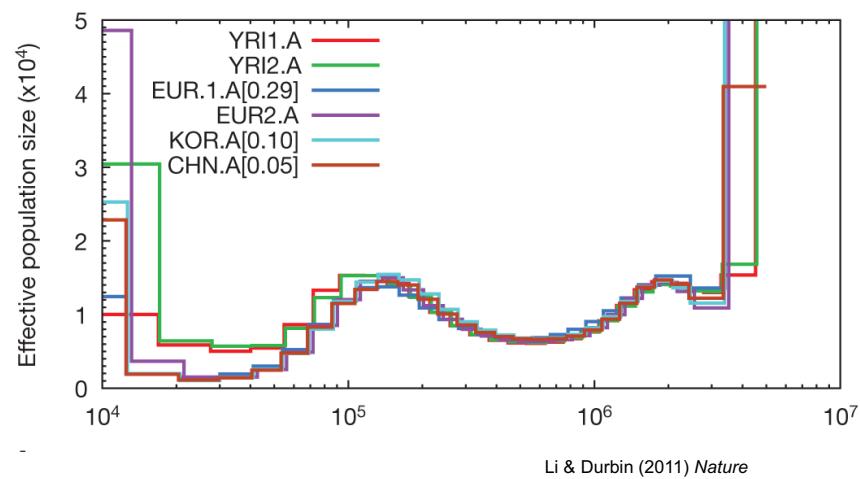
- Is an inferred demographic pattern meaningful?
 - Bayes factors
 - Visual inspection of skyline plot
 - Number of change points (eBSP)



14

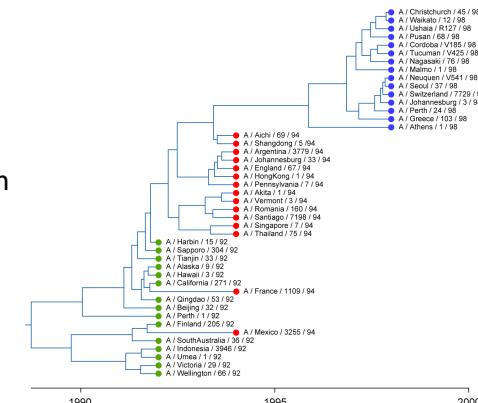
PSMC

- Pairwise sequentially Markovian coalescent (PSMC) model



Intraspecific molecular dating

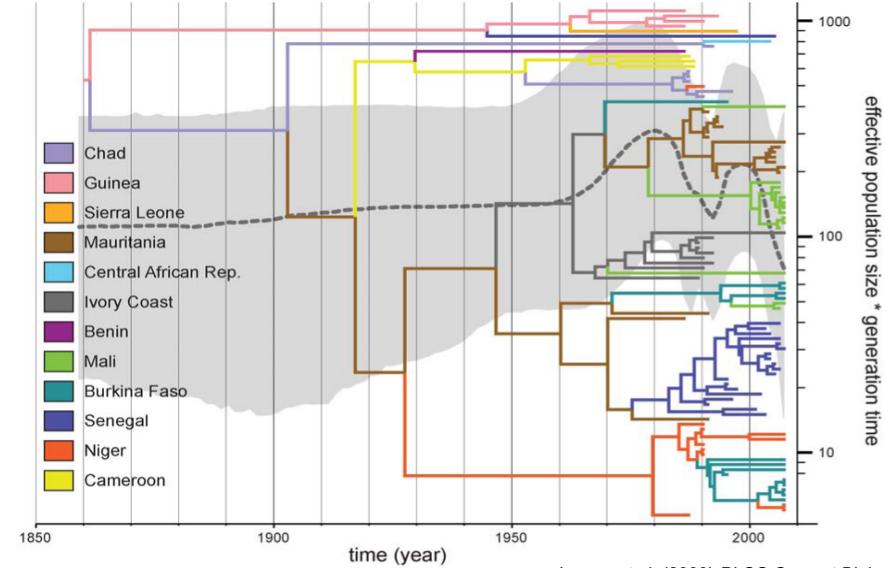
- Difficult to identify reliable calibrations
 - Biogeographic events
 - Sampling times (tip dates)
- Low sequence divergence
 - Risk of overparameterisation
 - Sensitivity to priors
- Various sources of bias



16

Phylogeography

Discrete phylogeographic model



Lemey et al. (2009) *PLOS Comput Biol*

18

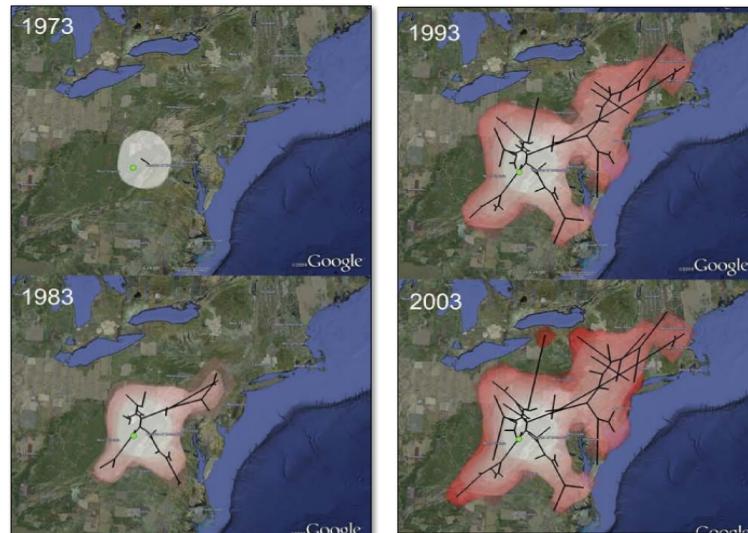
Discrete phylogeographic model



Lemey et al. (2009) *PLOS Comput Biol*

19

Continuous phylogeographic model



Lemey et al. (2010) *Mol Biol Evol*

20