From phylogenetic trees to phylodynamic trees

David Welch work with Sasha Gavryushkina, Alexei Drummond, Tanja Stadler, Kylie Chen, Tim Vaughan, Alex Popinga

University of Auckland

10 September 2014 IdRec symposium Supported by Marsden Grant UOA1324.

Phylogenetics uses different models depending on data Genealogy: Coalescent



The coalescent relates node height to population size



Population size of N.

pictures from Beerli

The coalescent relates node height to population size



Population size of N. So choose same parent with prob 1/N pictures from Beerli

The coalescent relates node height to population size



Population size of N. So choose same parent with prob 1/Nk lineages coalesce at rate $\frac{k(k-1)}{2}\frac{1}{N}$

Coalescent based methods have been remarkably successful



from Stiller et al 2010

Coalescent based methods have been remarkably successful



from Bouckhaert et al 2012

Phylodynamics: applying phylogenetic methods to epidemiological data



from Hue et al 2005

Phylodynamics: applying phylogenetic methods to epidemiological data



from Hedge et al 2013

A hierarchical model of transmission trees



Trees with sampled ancestors



Trees with sampled ancestors



(only 3 out of 4 parameters are identifiable)

Sampled ancestors exist in data: HIV MSM cluster UK



Sampled ancestors exist in data: HIV MSM cluster UK



from Gavryushkina et al 2014

Parameter estimates are wrong if sampled ancestors are not included

Birth rate (λ)



Without sampled ancestors

from Gavryushkina et al 2014

Parameter estimates are wrong if sampled ancestors are not included

Diversification rate ($\lambda-\mu$): growth rate of epidemic $_{\rm from \ Gavryushkina \ et \ al \ 2014}$



Skyline methods are unreliable heuristics in phylodynamics



picture from Beerli

An exact model for reconstructing the transmission tree and prevalence



Points to remember

- Standard phylogenetic methods cannot be applied willy-nilly to genomic data from pathogens
- Models that capture the dynamics specific to epidemics result in more reliable inference
- We are currently working on a number of inference tools for models of sampled ancestor trees, oriented trees, exact inference of trees and prevalence curves, hierarchical models of viral and transmission trees.
- All tools will be made available as Beast 2 packages.
- ▶ We are always keen to collaborate with people who have data

References

- Withering Away 25,000 Years of Genetic Decline Preceded Cave Bear Extinction, Mathias Stiller, Gennady Baryshnikov, Herv Bocherens et al. Mol Biol Evol (2010) 27 (5): 975-978, doi:10.1093/molbev/msq083
- Mapping the Origins and Expansion of the Indo-European Language Family, Remco Bouckaert, Philippe Lemey, Michael Dunn et al. Science 24 August 2012: 337 (6097), 957-960. doi:10.1126/science.1219669
- Peter Beerli, coalescent pictures from talk slides at www.bieler-beerli.com/downloads/talks/madagascar-talk.pdf
- Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups, Stéphane Hue, Deenan Pillay, Jonathan P. Clewley, and Oliver G. Pybus, PNAS, 2005, vol. 102, no. 12, p44254429, doi:10.1073.pnas.0407534102
- Eduardo Castro-Nallar, Marcos Prez-Losada, Gregory F. Burton, Keith A. Crandall, The evolution of HIV: Inferences using phylogenetics, Molecular Phylogenetics and Evolution, Volume 62, Issue 2, February 2012, Pages 777-792, doi:10.1016/j.ympev.2011.11.019
- Real-time characterization of the molecular epidemiology of an influenza pandemic, J. Hedge, S. J. Lycett, and A. Rambaut. Biol. Lett. October 23, 2013. doi:10.1098/rsbl.2013.0331
- Recursive algorithms for phylogenetic tree counting, A Gavryushkina, D Welch, AJ Drummond Algorithms for Molecular Biology 8 (1), 26, 2013, doi:10.1186/1748-7188-8-26
- Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration, A Gavryushkina, D Welch, T Stadler, A Drummond, PLoS Computational Biology, 2014 to appear. arXiv preprint arXiv:1406.4573
- Complex population dynamics and the coalescent under neutrality, Erik M. Volz. 2012 Genetics, 190:1, 187-201. doi:10.1534/genetics.111.134627