

## Phylogenetics

# UglyTrees: a browser-based multispecies coalescent tree visualizer

Jordan Douglas  \*

School of Computer Science, University of Auckland, Auckland 1010, New Zealand

\*To whom correspondence should be addressed.

Associate Editor: Yann Ponty

Received on December 19, 2019; revised on June 22, 2020; editorial decision on July 19, 2020; accepted on July 22, 2020

## Abstract

**Summary:** Visualization is a vital task in phylogenetics and yet there is a deficit in programs which visualize the multispecies coalescent (MSC) model. UglyTrees (UT) is an easy-to-use program for visualizing multiple gene trees embedded within a single species trees. The mapping between gene and species nodes is automatically detected allowing for ready access to the program. UT can scrape the contents of a website for MSC analyses, enabling the sharing of interactive MSC figures through optional parameters in the URL. If a posterior distribution is uploaded, the transitions between MSC states are animated allowing the visual tracking of trees throughout the sequence.

**Availability and implementation:** UT runs in all major web browsers including mobile devices, and is hosted at [www.uglytrees.nz](http://www.uglytrees.nz). The MIT-licensed code is available at <https://github.com/UglyTrees/uglytrees.github.io>.

**Contact:** [jordan.douglas@auckland.ac.nz](mailto:jordan.douglas@auckland.ac.nz)

## 1 Introduction

As biological sequence data become increasingly available, it becomes enticing to infer species phylogeny by concatenating genes sequences and inferring the phylogeny of the species as that of the gene tree. However, this approach makes for a biased estimator of species divergence times and substitution rates when incomplete lineage sorting is present (Arbogast *et al.*, 2002; Mendes and Hahn, 2016; Ogilvie *et al.*, 2016), and an inconsistent estimator of topology when divergence times are small (Pamilo and Nei, 1988). Bayesian multispecies coalescent (MSC) methods address these issues (Flouri *et al.*, 2018; Heled and Drummond, 2010; Höhna *et al.*, 2016; Jones, 2017; Ogilvie *et al.*, 2017; Ronquist *et al.*, 2012).

Visualization is an essential task in phylogenetics. Consequently, gene tree visualisation programs are ubiquitous [see Dendroscope—Huson *et al.* (2007); FigTree—Rambaut (2012); DensiTree—Bouckaert and Heled (2014); IcyTree—Vaughan (2017) and ape—Paradis and Schliep (2019)]. Unfortunately, MSC visualizers are far less common [the only program which we are aware of is a script used in Heled and Drummond (2010)].

In a conventional MSC depiction (Degnan and Rosenberg, 2009; Heled and Drummond, 2010; Rannala and Yang, 2003), one or more gene trees are embedded inside a species tree. Figure heights correspond to gene/species divergence times, while widths correspond to species' (effective) population sizes.

In an MSC analysis, an arbitrary number of gene trees could be used, sometimes even hundreds or thousands (Ogilvie *et al.*, 2017; Singhal *et al.*, 2018). There is no guarantee that branches will not overlap. Moreover, although continuous population models exist

(Heled and Drummond, 2010; Heled *et al.*, 2013), MSC analyses quite frequently invoke piecewise population size models where each species has its own freely determined population size (Gómez-Hernández *et al.*, 2019; Pinto *et al.*, 2019; Singhal *et al.*, 2018).

These two components of the MSC (embedded gene trees and piecewise population models) make its visualization an inherently inelegant task. This is compounded by the inverse relationship between the rate of coalescence and population size, which results in coalescent events tending to be clustered together in the narrowest of branches. This article presents UglyTrees (UT)—an easy-to-use browser-based program for visualizing MSC models. UT reads trees represented in Newick/NEXUS format and is therefore compatible with trees produced by \*BEAST, StarBEAST2, STACEY, MrBayes and RevBayes (Heled and Drummond, 2010; Höhna *et al.*, 2016; Jones, 2017; Ogilvie *et al.*, 2017; Ronquist *et al.*, 2012).

## 2 Visualization of the MSC

UT renders zero-or-more (rooted binary) gene trees embedded within a single (rooted binary) species tree using scalable vector graphics (SVG). The tree parser is built on top of that of IcyTree (Vaughan, 2017). The mapping between gene and species nodes is automatically detected allowing for ready access to the program.

The mapping algorithm attempts to map each gene to exactly one species, first by direct substring comparison, and if that fails, the labels are split using a range of delimiters ('\_', '-', and ':'). If a mapping cannot be found, the user is prompted to give one. Consider the following example:

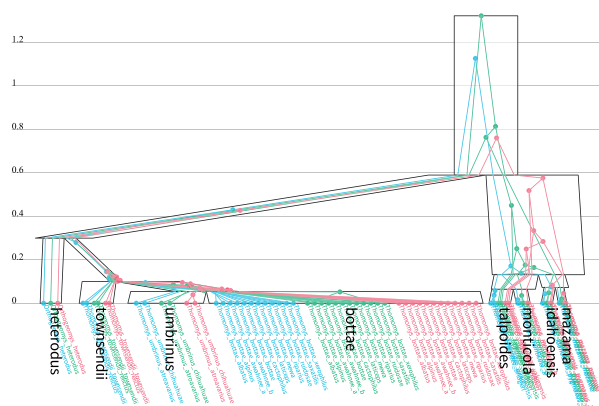


Fig. 1. MSC analysis performed on three genes from a Gopher dataset (Belfiore *et al.*, 2008) using a piecewise constant population size model

Genes: {horse\_1, horse\_2, seahorse\_1}, Species: {horse, seahorse}. horse\_1 and horse\_2 are mapped to horse and seahorse\_1 is mapped to seahorse.

The widths at the top and bottom of each species branch can be set independently (using tree meta-annotations) and the width in between is linearly interpolated (Fig. 1). This facilitates the visualization of two population size models commonly invoked in the literature: (i) piecewise *constant* models, for which each species branch has freely a determined population size (i.e. top and bottom are the same), and (ii) continuous *linear* models (Heled and Drummond, 2010), for which the population size at the bottom of each branch is equal to the sum of its two children's population sizes at the tops of their respective branches.

If multiple MSC states are uploaded (a posterior distribution for instance), they can be iterated through with smooth animated transitions. This enables the visual tracking of trees through the posterior distribution. UT's zooming feature makes it suitable for large datasets, however, performance depends on the number of SVG elements—with complexity dependent on the number of genes  $G$ , the number of species  $S$  and the taxon count  $N$ . When there is a large number of SVG elements, UT by default renders one gene tree at a time (Fig. 1).

### 3 Web scraping

Any changes made to the visual settings can be downloaded as a template in XML format. Display settings are restored upon subsequently uploading the template. By adding parameters to the URL, the simple backend of UT fetches a template file—and any tree files the template is pointing to—from the web. A customized message is optionally displayed to the user upon page load. This enables the sharing of MSC interactive visualizations with just one click. For example: <http://uglytrees.nz/?w=http://uglytrees.nz/examples/gopher/session.xml>

### Funding

This work was supported by Marsden grant [18-UOA-096], from the Royal Society of New Zealand.

*Conflict of Interest:* none declared.

### References

- Arbogast, B.S. *et al.* (2002) Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annu. Rev. Ecol. Syst.*, **33**, 707–740.
- Belfiore, N.M. *et al.* (2008) Multilocus phylogenetics of a rapid radiation in the genus *Thomomys* (rodentia: Geomyidae). *Syst. Biol.*, **57**, 294–310.
- Bouckaert, R. and Heled, J. (2014) Densitree 2: Seeing trees through the forest. *BioRxiv*, page 012401.
- Degnan, J.H. and Rosenberg, N.A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.*, **24**, 332–340.
- Flouri, T. *et al.* (2018) Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Mol. Biol. Evol.*, **35**, 2585–2593.
- Gómez-Hernández, C. *et al.* (2019) Evaluation of the multispecies coalescent method to explore intra-*Trypanosoma cruzi* I relationships and genetic diversity. *Parasitology*, **146**, 1063–1012.
- Heled, J. and Drummond, A.J. (2010) Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.*, **27**, 570–580.
- Heled, J. *et al.* (2013) Simulating gene trees under the multispecies coalescent and time-dependent migration. *BMC Evol. Biol.*, **13**, 44.
- Höhna, S. *et al.* (2016) RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.*, **65**, 726–736.
- Huson, D.H. *et al.* (2007) Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics*, **8**, 460.
- Jones, G. (2017) Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *J. Math. Biol.*, **74**, 447–467.
- Mendes, F.K. and Hahn, M.W. (2016) Gene tree discordance causes apparent substitution rate variation. *Syst. Biol.*, **65**, 711–721.
- Ogilvie, H.A. *et al.* (2016) Computational performance and statistical accuracy of \*BEAST and comparisons with other methods. *Syst. Biol.*, **65**, 381–396.
- Ogilvie, H.A. *et al.* (2017) Starbeast2 brings faster species tree inference and accurate estimates of substitution rates. *Mol. Biol. Evol.*, **34**, 2101–2114.
- Pamilo, P. and Nei, M. (1988) Relationships between gene trees and species trees. *Mol. Biol. Evol.*, **5**, 568–583.
- Paradis, E. and Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, **35**, 526–528.
- Pinto, B.J. *et al.* (2019) Population genetic structure and species delimitation of a widespread, Neotropical dwarf gecko. *Mol. Phylogenet. Evol.*, **133**, 54–66.
- Rambaut, A. (2012) *Figtree v1.4*. <http://tree.bio.ed.ac.uk/software/figtree/>
- Rannala, B. and Yang, Z. (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, **164**, 1645–1656.
- Ronquist, F. *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.*, **61**, 539–542.
- Singhal, S. *et al.* (2018) A framework for resolving cryptic species: a case study from the lizards of the Australian wet tropics. *Syst. Biol.*, **67**, 1061–1075.
- Vaughan, T.G. (2017) IcyTree: rapid browser-based visualization for phylogenetic trees and networks. *Bioinformatics*, **33**, 2392–2394.