BIOINFORMATICS APPLICATIONS NOTE

Vol. 26 no. 10 2010, pages 1372–1373
doi:10.1093/bioinformatics/btq110

Phylogenetics

Advance Access publication March 12, 2010

DensiTree: making sense of sets of phylogenetic trees

Remco R. Bouckaert

Department of Computer Science, Auckland University

Associate Editor: Martin Bishop

ABSTRACT

Motivation: Bayesian analysis through programs like BEAST (Drummond and Rumbaut, 2007) and MrBayes (Huelsenbeck *et al*., 2001) provides a powerful method for reconstruction of evolutionary relationships. One of the benefits of Bayesian methods is that well-founded estimates of uncertainty in models can be made available. So, for example, not only the mean time of a most recent common ancestor (tMRCA) is estimated, but also the spread. This distribution over model space is represented by a set of trees, which can be rather large and difficult to interpret. DensiTree is a tool that helps navigating these sets of trees.

Results: The main idea behind DensiTree is to draw all trees in the set transparently. As a result, areas where a lot of the trees agree in topology and branch lengths show up as highly colored areas, while areas with little agreement show up as webs. This makes it possible to quickly get an impression of properties of the tree set such as well-supported clades, distribution of tMRCA and areas of topological uncertainty. Thus, DensiTree provides a quick method for qualitative analysis of tree sets.

Availability: DensiTree is freely available from http://compevol.auckland.ac.nz/software/DensiTree/. The program is licensed under GPL and source code is available.

Contact: remco@cs.auckland.ac.nz

Received on December 23, 2009; revised on March 5, 2010; accepted on March 8, 2010

1 INTRODUCTION

Sets of trees appear as the output of phylogenetic explorations through Bayesian and bootstrap analysis. The most common approaches to dealing with such sets of trees are to calculate a single summary tree, determine a set of most likely clades, draw neighbor networks (Huson and Bryant, 2006) or perform multidimensional scaling (MDS) (Hillis *et al*., 2005).

A popular method for analyzing tree sets is to find a single representative phylogeny and label the branches with uncertainty [for instance using the TreeAnnotator in BEAST (Drummond and Rumbaut, 2007)]. The benefit of this method is that it is easy to interpret the single hierarchy by visualizing it in a tree drawing program such as FigTree (available from http://tree.bio.ed.ac.uk/software/figtree/) and use error bars to indicate uncertainty in branch lengths. Unfortunately, it takes some skill to interpret situations where there is uncertainty in the topology. Such cases show in the tree as short branches with relatively large error bars. However, this is indistinguishable from the case where a single tree topology dominates but there is large uncertainty due to model and/or data.

Another method for interpreting tree sets is to find clades (i.e. subtrees) that occur with high frequency, for example, by using the TreeLogAnalyser in BEAST (Drummond and Rumbaut, 2007). The number of relevant clades may become very large, especially with large datasets since the number of possible trees grows exponentially in the number of labels. Furthermore, interpreting uncertainty within high-frequency clade may become cumbersome due to the large number of them.

Tree networks as in SplitsTree (Huson and Bryant, 2006) are graphs containing edges wherever such edges appear (possibly at some threshold frequency) in the tree set. Tree networks do not allow easy representation of uncertainty and can become unwieldy when large numbers of distinct topologies are present in the tree set.

MDS as implemented in (Hillis *et al*., 2005) is a technique that comes closest to our method in that it is qualitative as opposed to the more quantitative annotated summary tree and clade set methods. MDS allows identification of tree islands in a compelling way, but uncertainty of node heights is hard to interpret.

2 APPROACH

DensiTree draws all trees in the set simultaneously, but instead of using opaque lines, transparency is used when drawing the trees. As a result, in areas where a lot of the trees agree on the topology and branch length, there will be many lines drawn and the screen will show a densely colored area. Areas where there are a few competing topologies will be highlighted by a web of lines. Uncertainty in tMRCA and their distribution can be shown by smears around the mean MRCA. Where summary trees and clade sets are quantitative approaches to tree set analysis, DensiTree provides a qualitative approach. Figure 1 shows some examples that give an impression of the benefits of this approach. However, this being an inherently visual technique, the reader is invited to visit the gallery at the DensiTree website to view a larger variety of tree sets.

For each tree topology that occurs in the tree set, DensiTree calculates a so-called ‘consensus tree’. The branch lengths of a consensus tree are the average length of the branches for that particular topology. This set of consensus trees can be drawn independently from the rest of the tree set, and are drawn with intensity proportional to the frequency of the topology occurring in the set. The tree set can be navigated one topology at the time, so the most frequently occurring topology and its properties can be studied apart from the other topologies. This is especially useful for tree sets where there are a small number of topologies dominating the set.

Another option is to use animation, where one topology is drawn per frame and each frame is drawn either on top of the old one or on a clear screen. The former shows how growing the tree set

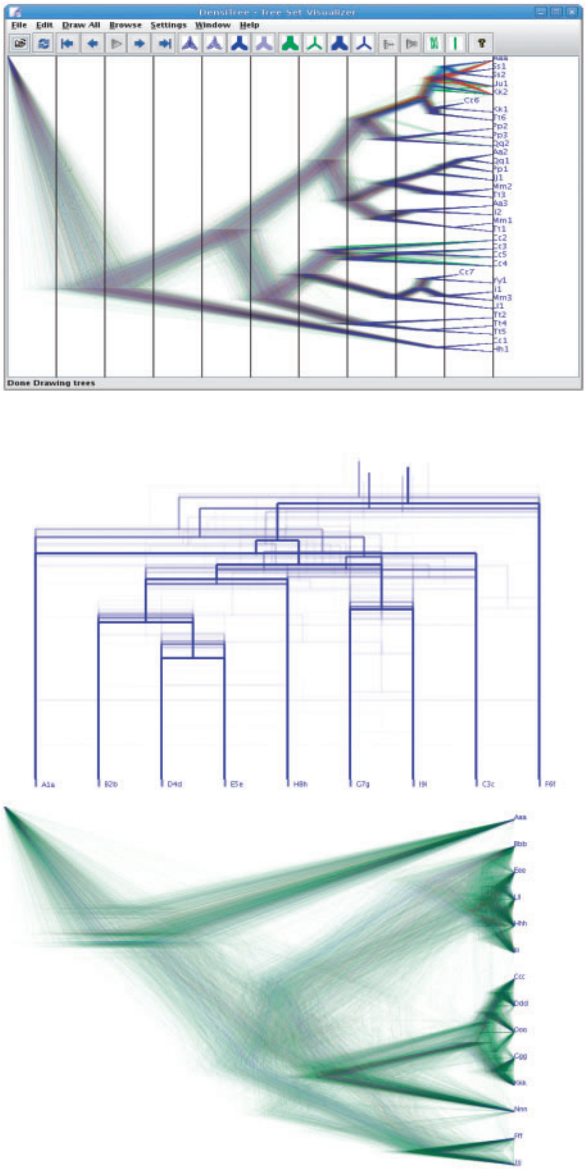
DensiTree - Tree Set Visualizer
File Edit Draw All Browse Settings Window Help
Done Drawing trees
A1a
B2b
D4d
E5e
H8h
G7g
I9i
C3c
F6f

Fig. 1. Screen shot (top) of DensiTree and some example outputs. Top: a tree set with a dominant topology in most of the tree, except for taxa at the top where there are three alternative interpretations (blue for most frequently occurring topolgy, red for second and green for third). Note the increase in uncertainty in tMRCA going up higher in the tree. Middle: block tree where only consensus trees are shown. There is a reasonable certainty of most of the tree, except for the area around the root. Bottom: example of a set with very distinct clades, where the two clades over five nodes have very high uncertainty.More examples are available in the gallery on the DensiTree web page.

increases the uncertainty in the topology, while the latter allows closer inspection of each of the different topologies in the set.

To explore distribution of tMRCA in trees, trees can be drawn in triangular (pyramidal) form or as a block tree. Both forms can be useful depending on the tree set.

While the tMRCAs determine one axis of the layout of the tree, the other axis is determined by the order of the tips, which make a dramatic impact on the interpretability of an image. Various heuristics are implemented in DensiTree to reorder the tips. A method that appears to perform well overall is to calculate the average distance of tips according to the number of branches separating leaves in a tree. The leaf ordering is started by selecting the closest two leaves, then extending the ordering left or right with the closest node to the left or right from the already ordered nodes till all nodes end up in the ordering. Other ordering heuristics are based on performing classical clustering using the distance as described above and use an ordering that lays out the thus obtained hierarchy. This is an area for ongoing research. A known issue it that the approach works best on clock like trees so that the leafs all are fixed at the same location.

3 TECHNICAL DETAILS

DensiTree can read tree sets in NEXUS format (Maddison *et al.*, 1997), such as those produced by programs like MrBayes and BEAST, and lists of Newick trees as produced by PHYLIP. Most aspects of the tree drawing can be configured, including line width, line color, intensity, font, background color, etc. and can be passed as command line options. Images can be exported in BMP, JPG and PNG bitmap formats. DensiTree is written in Java, so any computer that runs a Java runtime version 1.6 or later should be able to use DensiTree. The drawing of trees is performed with multiple threads, so that modern multicore machines are fully utilized. Large tree sets with many taxa may take a few minutes to draw, but drawing only consensus trees can speed up the process a bit. DensiTree is licensed under the GNU public license. A manual is available via `http://compevol.auckland.ac.nz/DensiTree/` and contains further details of the user interface.

ACKNOWLEDGEMENT

This work was greatly helped by discussions with Alexei Drummond and other members of the evolutionary computation group in Auckland.

Funding: Marsden grant 08-UOA-170.

Conflict of Interest: none declared.

REFERENCES

Drummond,A. and Rambaut,A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Hillis,D.M. *et al.* (2005) Analysis and Visualization of Tree Space. *Syst. Biol.*, **54**, 471–82.

Huelsenbeck,J.P. *et. al.* (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, **294**, 2310–2314.

Huson,D.H. and Bryant,D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.*, **23**, 254–267.

Maddison,D.R. *et al.* (1997) NEXUS: an extendible file format for systematic information. *Syst. Biol.*, **46**, 590–621.