Phylotastic! Hackathon, NESCent; 2012-06-04

**phylomatic:**
a simple tool for
‘phylogenetic empowerment’

Cam Webb

Arnold Arboretum of Harvard University
General question

- “What are the evolutionary relationships among my set of taxa?”... without collecting original molecular or morphological character data, or sequence re-analysis → re-use of existing phylogenies
Origins

- In ca. 1999, I wanted to estimate the phylogeny for 325 species trees in forest plots in Borneo to ask questions about the *phylogenetic community structure* of the forest.

- Didn’t have any genus-level phylogenies, but did have the first APG paper (*APG, 1998*) and a few ordinal phylogenies of families and started with hand-assembling a family tree including only those families in the Bornean forest.

- Simple list-to-phylogeny matcher originally coded in *gawk* (*gawk* still rocks!)
Evolution

• Put tools behind a web form ca. 2002, called it phylomatic (from beginning users could supply their own input tree URL)
• Basic ‘TNRS’ for plants: fuzzy (agrep) taxon lookup against megatree and APweb and Kew lists
• Started managing sub-trees as Newick text files in a Subversion repository (‘Tree-of-trees’ database)
• Mega-trees were archived and could be cited
• Ca. 2005, re-coded phylomatic in C, using tree I/O functions in phylocom, and released stand-alone phylomatic as part of phylocom distribution
• Added web service/API
I/O formats

megatree: Newick

\[
\text{species} \quad \text{section} \quad \text{genus} \\
| \quad | \quad | \\
\downarrow \quad \downarrow \quad \downarrow \\
\ldots, \text{(banksia\_grandis, banksia\_solandri)grandes})\text{phanerostomata)}banksia)\), \text{carnarvonia}, \text{grevillea})\text{proteaceae})\text{proteales, ((buxaceae...}
\]

\[ \hat{\quad} \quad \hat{\quad} \\
\quad \quad | \quad | \\
\text{family} \quad \text{order} \\
\]

taxa: slash-delimited node-paths

- proteaceae/banksia/banksia\_grandis
- proteaceae/banksia/banksia\_xxxx
- buxaceae/buxus/buxus\_aaaa
- buxaceae/buxus/myNode1/buxus\_bbbb <---- ‘node path’
- buxaceae/buxus/myNode1/buxus\_cccc
Basic algorithm

• Ascend each slash-delimited node path in taxa file, checking for the most terminal match in the megatree

• Add this node path to the megatree (allows user to add new phylogenetic structure), setting edge lengths to equidistantly space intermediate nodes

• Prune the megatree for the terminals in the taxa file

• Output new tree and list of non-matching taxa
Dealing with non-terminal matches (7)

- (1) Insert ‘blindly’ at first matching inner node (A)
- (2) Collapse all other matching taxa to polytomy at this inner node (A)
- (3) Insert randomly into megatree clade distal to A
Tree-of-trees database

- Backbone tree from APG3 (*Chase et al., 2009*)
- Families-in-orders from APweb (Peter Stevens et al.)
- Family trees: ad hoc selection hand coded from PDFs by me and volunteers
- with tools for assembly into megatrees
- Managed as text in CVS ([github](https://github.com))
- Also stores node age estimates (for `bladj`)
- Discussion topics: ‘Tree curation’ models (ToL, professional, etc); phylocode; community database for node age estimates (fossils, etc)
Users of phylomatic

- Webb and Donoghue (2005) paper has 160 citations
- Applications (in decreasing order)
  - Analysis of functional traits in ecological context
  - Phylogenetic community structure
  - Comparative analysis/independent contrasts
  - Phylogenetic diversity for conservation
- Majority of users just used default megatree, and did not augment with other trees
- I would like to present a single tree, with more information and caveats, with the tools to easily modify ‘to taste’
Acknowledgments

- **Ideas and feature suggestions** Michael Donoghue, David Ackerly, Rick Ree, Steven Kembel, and numerous Phylomatic users.

- **Funding** The Arnold Arboretum of Harvard University, Yale Institute for Biospheric Studies, the National Science and Engineering Research Council of Canada, and NSF grants (DEB-0212873, DEB-0515520, 1020868)

- **Software** GNU emacs, gcc, gdb, gawk, plus Subversion and LaTeX
References

