

Phylotastic! Hackathon, NESCent; 2012-06-04

**phylomatic:  
a simple tool for  
'phylogenetic empowerment'**

Cam Webb

Arnold Arboretum of Harvard University

Version 0.1-2-g7263348



# General question

(2)

- “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

(3)

- 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

(4)

- 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

(5)

megatree : Newick

```

      species                      section      genus
      |                            |            |
      v                            v            v
..., (banksia_grandis,banksia_solandri)grandes))phanerostomata)banksia)),
carnarvonian),grevillea)proteaceae))proteales, ((buxaceae...
      ^                ^
      |                |
      family          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

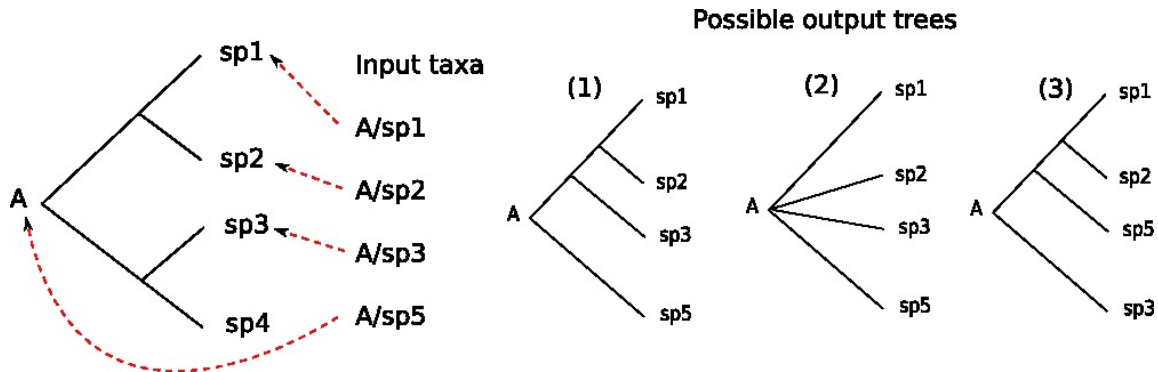
(6)

- 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

(8)

- 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](#))
- 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

(9)

- [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

(10)

- **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 L<sup>A</sup>T<sub>E</sub>X



## References

- APG. 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* **85**:531–553.
- Chase, M., M. Fay, J. Reveal, D. Soltis, P. Soltis, A. Anderberg, M. Moore, R. Olmstead, P. Rudall, et al. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161**:105–121.
- Webb, C. O., and M. J. Donoghue. 2005. Phylomatic: tree assembly for applied phylogenetics. *Molecular Ecology Notes* **5**:181–183. URL [http://www.phylodiversity.net/cwebb/pubs/webb2005\\_men.pdf](http://www.phylodiversity.net/cwebb/pubs/webb2005_men.pdf).