Tree-for-all hackathon

Ann Arbor Michigan, September 15-19, 2014

Link to schedule, including webex connection info

Top-level document for notes from hackathon. Subgroups - add link to your notes and code repositories!

Subgroups

Integration of Trees and Traits

- People: Jeff Cavner/remote(skype Jeff Cavner, email jcavner@ku.edu), Luke Harmon, Zack Galbreath, Jorrit Poelen, Julienne Ng, Alex Harkess, Chodon Sass, Shan Kothari, Mark Westneat, (0.5*Curt)
- Link to notes:
- Code: https://github.com/lukejharmon/traitathon
- All final products are on the wiki

Library wrappers for OT APIs

- People: (0.5 * Joseph), mark holder(remote contact via IRC or G+ preferred. Skype possible), Jon Hill, Matt Yoder, François Michonneau, Jeet Sukumaran, David Winter, Karolis Ramanauskas
- Link to notes: https://docs.google.com/document/d/1OUS1-AtP6Ub6lGl87Ge_-06AdpnwLBF9Px2Che5l6aA
- Code:
  - R: https://github.com/fmichonneau/rotl
  - Python: https://github.com/OpenTreeOfLife/pyopentree
  - Ruby: https://github.com/SpeciesFileGroup/bark
  - Tests: https://github.com/OpenTreeOfLife/shared-api-tests

Phylogeny viz style-sheets

- People: Peter Midford (remote, IRC, G+, Skype OK), Jim Allman (remote), Pandurang (remote), Daisie, Gaurav, Julie, MIke Rosenberg (remote)
- Link to notes:
- Code: https://github.com/daisieh/ete [move to OpenTreeOfLife org?]

Sampling taxa

- People: Andréa, Arlin, Kayce, Dilrini, Jonathan (0.75) , Nicky
- Link to notes: Google doc Sampling taxa
- Code: https://github.com/arlin/sample_taxa
- Demo1: http://phylojive.acis.ufl.edu/PhyloJive/Otldigbio.jsp
- Demo2: https://arbor.kitware.com/
- draft of blog for OpenTree blog: sampling taxa blog

Branch lengths

- People: Lyndon (remote), Rachel Warnock, Josef Uyeda, Katie Lyons, Miranda Sinnott-Armstrong, Bob (remote), 0.5*Curt Lisle (remote)
- Link to notes Now including workflow!
- Code: https://github.com/ladykiyenz/branchlengths
- Workflow example: https://raw.githubusercontent.com/ladykiyenz/branchlengths/master/otTimetree_tutorial.md

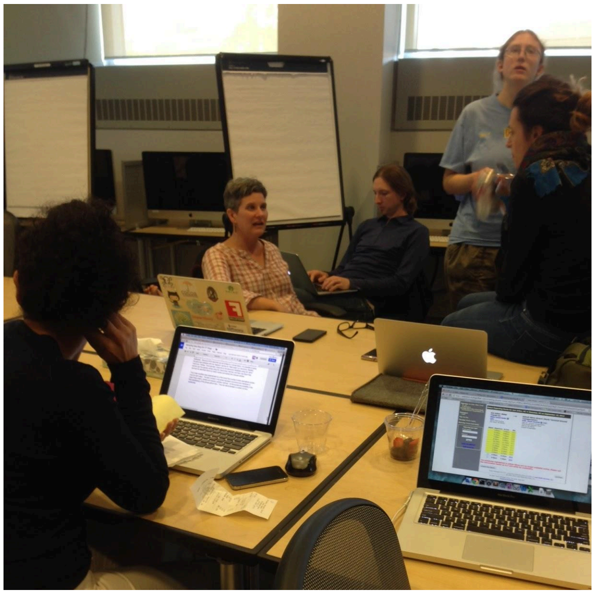
Annotation database

- People: Cody, Karen, Stephen, (0.5 * Joseph), Mark H. (remote), Hilmar (remote), Temi
- Link to notes

- Code: https://github.com/OpenTreeOfLife/muriqui

Wrap up notes

- branch lengths - see what trees have branch lengths, and units
- have branch lengths on synthetic tree
- for a given ott id, return taxonomic lineage [include lineage option with services]
    - add lineage to TNRS results, not just taxonomy
- query speed - can they be snappier?
- can the methods be GET instead of POST? especially the TNRS methods
- get citation for a study in a consistent format [note that about 10-20% of studies don't have DOIs]
- yay! for taxonomy bulk export; can we also get the lineage? could provide a script that does this



Day 1 notes

For IRC difficulties, try port: 7070, use SSL encryption

Arbor demonstration videos:
https://github.com/arborworkflows/arborworkflows.github.com/wiki/Arbor-Demonstration-Videos

potential bootcamps

- building the opentree taxonomy
- importing data into neo4j and doing synthesis

Pitches!

Pre-pitching ideas - link to hackathon issue tracker

1. **Integration of trees & traits (pt 1)** Jorrit: make a service invoked via R to query traits that can be associated with OT taxa, including but not limited to species interactions; use this in a workflow to enter an ancestral node for given tips, or go the other way (e.g., give me all the predators of bees in N. California, and retrieve the annotated tree from OT). (see http://www.globalbioticinteractions.org/ for predator-prey, host-parasite etc data for 30K species). Seeing this as an example of integration with trait databases (interaction as a trait) 4
    - https://github.com/OpenTreeOfLife/hackathon/issues/9
2. **library wrappers for OT API**Jeet: Python interoperability modules to query opentree API to generate (native) rich OO data objects (rather than the abbreviated information currently available by passing via Newick)

    - https://github.com/OpenTreeOfLife/hackathon/issues/28
    - Matt wants to do the same thing as a Ruby gem (https://github.com/SpeciesFileGroup/bark) [in the Ruby case this looks like goals of Peyotl steps 1-4]
    - François: library in R to query OT and get back tree to use in other R packages; also handle challenge of resolving names https://github.com/OpenTreeOfLife/hackathon/issues/14 8.5
3. **phylogeny viz style-sheets** Daisie: problem: published trees marked up with stars, support values, pics, etc, manually added in Illustrator (or Figtree), impossible to reproduce, hard to update; solution: separate out these parts as a rendering style analogous to CSS; could create a style for OT studies that includes some information available
    - idea: add generation of PDF
    - Jim has a use-case, tree viz tool
    - https://github.com/OpenTreeOfLife/hackathon/issues/17 4 [jimA makes 5?]
4. **Integration of trees & traits (pt 2)** Luke et al: use OT as the first step in a workflow that includes analysis of character evolution; user starts with character matrix; get tree from OT; then do ancestral state reconstruction
    - use-case (Julienne): floral trait evolution, has big char matrix and tree
    - use-case (Josef): do interesting things with trait evolution
    - Chodon: have rich trees with branch lengths; would like to upload trees for test cases (plants with rich data); test interaction / interoperability between opentree and arbor; idea is to test OT as a way to integrate, store, retrieve your study >10
5. Arlin: **taxonomic down-sampling**; user has a taxon such as Felidae or Mammalia, wants a tree of species; usually this will involve down-sampling, e.g., one species per family, most familiar species, species with genomes at NCBI; >10
    - https://github.com/OpenTreeOfLife/hackathon/issues/15
    - Cody: there are actually 2 distinct cases here: given me the species with property X in taxon T, and give me the top N species in taxon T according to metric X. The second one is much harder.
6. **Spec for reference taxonomy curation** Matt: opentree reference taxonomy, currently generated by automated alignment plus ad hoc patches suggested by curators and then vetted and implemented by one key staff member; develop spec for interface for multiple curators to suggest patches to OTT, so that the process can be more automated 4
    - may also be of interest to The Plant List
7. **Branch lengths for induced trees** Karen: some input trees come with branch lengths, but synthetic tree does not come with branch lengths: is there some way add branch length info to synthetic tree? (1) model-assigned lengths (2) go out and sample char data for implicated species (3) use info from source trees somehow (4) fossil calibrations (see DateLife: http://datelife.org/) 6
    - https://github.com/OpenTreeOfLife/hackathon/issues/25
    - Bob: can source trees be combined using STRAW as a web service? STRAW can generate tree with branch lengths.
    - Bob: if source trees are available in API, then Arbor can run NJst

    - Use the Fossil Calibration Database for this? *i'm glad to advise… (we're still in closed beta) -- jimA*
    - Google doc notes
8. **Pitch added: annotating the synthetic tree** Cody: a human- and machine-friendly way to designate clades in the synthetic tree for annotation purposes