Synthetic Phylogeny of the Decapod Crustaceans

Keith A. Crandall - Director, Computational Biology Institute

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC
Data Availability

High archival rate of sequence data

~4% of all published phylogenetic trees
Karen Cranston, lead PI (Duke)
Gordon Burleigh (Florida)
Keith Crandall (GWU)
Karl Gude (MSU)
David Hibbett (Clark)
Mark Holder (Kansas)
Laura Katz (Smith)
Rick Ree (FMNH)
Stephen Smith (Michigan)
Doug Soltis (Florida)
Tiffani Williams (TAMU)

AVAToL: Assembling, Visualizing and Analysis of the Tree of Life
1. Build the **first complete draft tree** of life

2. Engage the community in **refinement and annotation**

3. Promote a **culture of data sharing** through software products

4. Develop **novel methods for phylogenetic synthesis**
Synthetic trees: description of the method

Input trees

Source tree + taxonomy (grey)

Synthesis

Smith et al. 2013
AVAToL: Open Tree of Life

- Phylogeny of all life
  - no new relationships
- Curated taxonomy
- Accessible to the public
- Add/Extract phylogenies
- Two additional AVAToL groups

Published Taxonomies + Published Phylogenies

www.opentreeoflife.org

Smith et al. 2013 PLOS Comp. Biol.
Fig. 1. Phylogenies representing the synthetic tree. The depicted tree is limited to lineages containing at least 500 descendants. (A) Colors represent proportion of lineages represented in NCBI databases. (B) Colors represent the amount of diversity measured by number of descendant tips. (C) Dark lineages have at least one representative in an input source tree.

- GenBank taxonomy has ~411,000 binomials (<5% of est.)
- AVAToL taxonomy = 2,227,481 terminals
- Tips represented by phylogenies = 49,487
- PNAS www.pnas.org/cgi/doi/10.1073/pnas.1423041112
Open Tree APIs

- command-line access to versions of the Open Tree of Life
- access the Graph of Life
- taxonomic name resolution services
- taxonomy
- studies containing source tree
- [https://github.com/OpenTreeOfLife/opentree/wiki/Open-Tree-of-Life-APIs](https://github.com/OpenTreeOfLife/opentree/wiki/Open-Tree-of-Life-APIs)
Impacts

- Benchmark for current state of phylogenetic knowledge
- Increasing rate of data archive
- Placing “dark taxa” in global informatics framework
- Phylogenies for any set of species easily available for conservation and ecological studies
Maximizing Outreach

What is the Tree of Life?
The 'tree' is an imaginary branching, treelike structure representing the evolutionary divergence of all living creatures, or species, on Earth. The Public Tree of Life is a small representation showing only 200 of the millions of species that exist on the planet.

Why these long shapes? With only 200 species shown here, many common ancestors are omitted from this tree, which results in these long shapes that go back to until they find an ancestor.

How to Read the Wheel
Each rectangle, regardless of size or shape, represents a species (also illustrated below as a white dot). All species evolved from another species that lived before it (their ancestor). All species share some common ancestor, the most recent ancestral form or species from which two or more different species evolved. For example, humans didn't evolve from chimps, but both humans and chimps evolved from a common ancestor long ago.

Click on any species to learn about a species
Common Sea Otter (Enhydra lutris)
Bald Eagle (Haliaeetus leucocephalus)
Stick Bug (Acanthocoris sordidus)
Human (Homo sapiens)
Peat Moss (Sphagnum sp.)
Common Cold (Rhinovirus)
Spicy Archaea (Crenarchaeota)

Mouse around to Explore, play, learn

Click here to watch a short video explaining this tree
Click here to watch a short video explaining the common ancestry.
Click here to watch a short video explaining the common ancestry.

Species A is the common ancestor to species B and F
Species B is the common ancestor to species C, D, and E
Species alive today (Extinct species in black)

Link here to the Encyclopedia of Life
Link here to the Public Tree Videos
Link here to the Open Tree of Life blog.
Future work: automatic updating

update trees with new sequence data

detect and incorporate newly published trees

NCBI

DRYAD

TreeBASE  A Database of Phylogenetic Knowledge
“OPEN” TREE OF LIFE

http://opentreeoflife.org