

---

# Phylogenetic Analysis Toolkits

---

Mark T. Holder  
Dept. Ecology and Evolutionary Biology  
University of Kansas  
Lawrence, KS, USA

## Session goal

---

- Brief presentations focussed on new developments in existing software tools:
  - Mesquite – David Maddison
  - CIPRES – Mark Miller
  - MorphoBank – Maureen O’Leary
  - brief reports on a wide variety of tools – me
- Open discussion of best practices, and current hurdles.

# RAXML

---

ML tree inference from DNA and AA data.

Alexandros Stamatakis: <http://icwww.epfl.ch/~stamatak>

A new version of RAxML-7.0.0 released in January, 2008:

- The rapid nonparametric bootstrapping algorithm - about 1 order of magnitude faster than the old RAxML, but yields qualitatively comparable results.
- improved parallelization based on PThreads.
- two web-servers available. In Switzerland:  
<http://phylobench.vital-it.ch/raxml-bb>  
and on the CIPRES portal:  
<http://8ball.sdsc.edu:8889/cipres-web/Bootstrap.do>
- Parallelized DNA analyses for the IBM BlueGene (tested on 2,000 seqs x 50kB and 250 seqs x 500kB).

# GARLI

---

ML tree inference from DNA data.

Derrick Zwickl (currently a NESCent postdoc)

<http://www.bio.utexas.edu/faculty/antisense/garli/Garli.html>

Garli0.96 is in beta testing (**not** available yet from the url above).

- amino acid and codon models
- backbone (in addition to monophyly) constraints
- fast, stepwise-addition starting tree under ML criterion (with the max. # of attachment branches per taxon.
- ability to read Nexus starting trees
- more rigorous reading of Nexus datasets using Paul Lewis' (with M. Holder's modifications) **NCL**
- runs multiple search replicates from a single config file
- outgroups can be specified for orientation of inferred trees
- parallelized (MPI) bootstrapping

# PhyML

---

ML tree inference from DNA data.

Stephane Guindon and Olivier Gascuel: <http://atgc.lirmm.fr/phyml>

An article describing a major upgrade PhyML 3.0 is in prep:

- 2 new tree search heuristics based on SPR moves (slower than the original NNI-based strategy, but much better at maximizing the likelihood – based on a large collection of real-world data sets)
- the fast branch test developed by Anisimova and Gascuel (2006) based on an approximate likelihood ratio test (aLRT). to evaluate the support of the data to each internal edge. A fast, complement to the (time-consuming) non-parametric bootstrap analysis.
- User-interface upgrade. PHYLIP-like text interface and standardized command-line syntax.
- A new server and a new web interface to this upgraded version of PhyML should be made available within the next couple of months. Binaries and sources will also be made available at the same time.

# Leaphy

---

ML tree inference from DNA and AA data.

Simon Whelan: <http://www.bioinf.manchester.ac.uk/leaphy>

- released version is still a beta
- it now contains most standard DNA and amino acid models (the paper only discussed amino acid models)
- Simon has test versions with covarion models (contact him if you want to try itout).

# IQPNNI - Important Quartet Puzzling and NNI

ML tree inference from DNA

Vinh, Quang, Schmidt, and von Haeseler:

<http://www.cibiv.at/software/iqpnni/>

- GTR and submodels
- $\Gamma + I$  or Meyer and von Haeseler site-specific rates
- $O(n^2)$  tree construction from important quartets

# POY

---

Simultaneous alignment and tree searching.

Ward Wheeler:

<http://research.amnh.org/scicomp/projects/poy.php>

Recent additions/improvements:

- Maximum likelihood approaches
- improvements to parallelization
- chromosomal character optimization

# TNT

---

Goloboff, Farris, and Nixon:

<http://www.zmuc.dk/public/phylogeny/tnt>

- Now freely available, for single user license – cross platform.
- very fast parsimony searching up to huge number of taxa (sectorial searches, ratcheting and tree-drifting, tree-fusing and tree-hybridization).
- wide range of options of cost matrices.

# MrBayes

---

Bayesian tree inference under a wide range of models from almost any type character data

- in the midst of a reimplementations in C++ (MrBayes 4) while maintaining and updating MrBayes 3
- moving toward treating:
  - character substitution model,
  - partition of characters into different models,
  - alignment as random variables.

# BEAST

---

Bayesian inference of trees and divergence times

- new local clocks (no need to specify the # of “clocks” or their position on the tree).
- Implementation of Thorne et al. correlated, relaxed clocks.
- Improvements to uncorrelated clock implementation
- Simultaneous tree inference and sequence alignment

# NESCent Efforts

---

Evolutionary informatics working group (Stoltzfus and Vos)

[https://www.nescent.org/wg\\_evoinfo](https://www.nescent.org/wg_evoinfo)

## Activities

- CDAO – an Ontology for comparative data analysis
- NeXML – a Future Data Exchange Standard (also supported by the CIPRES team)
- developing a “language” for expressing statistical models of character evolution (as opposed to software-specific syntax).
- developing validation and interconversion software to support legacy formats
- documenting Use Cases and terminology in evolutionary analysis

## NESCent Efforts

---

R-hackathon (December 2007).

- Diversification – BiSSE model; Ree’s key innovation test; updates to the laser package (e.g. functions to test for shifts in diversification rate using combined phylogenetic and taxonomic data)
- Reimplementation of divergence time estimation in ape.
- Analysis of continuous character evolution (improvements to geiger package)
- Internal data representation standards and I/O (NEXUS)
- Interoperability between R and Mesquite

# NESCent Efforts

---

Google Summer of Code 2008

Application being prepared (Hilmar Lapp is the organizer)

[http://hackathon.nescent.org/Phyloinformatics\\_Summer\\_of\\_Code\\_2008](http://hackathon.nescent.org/Phyloinformatics_Summer_of_Code_2008)

Project ideas:

- should be short – achievable in one summer,
- should be concrete – timeline of the steps toward goals should be written by the mentor and student in the initial week,
- must be written up very soon (today or tomorrow), but do not need rich detailed, descriptions.

## NESCent Efforts

---

- **PhyloWS** – phyloinformatics web services API – Contact Hilmar Lapp (NESCent) or Rutger Vos (CIPRES)
- **PhyloDB** extension to BioSQL (Hilmar Lapp and Bill Piel)
- **GeoPhyloBuilder** – “3D spatiotemporal ‘geophylogenetic’ GIS data model’ from a tree and locality data – Kidd and Liu

Hilmar Lapp has a blog: <http://blog.evolvingbits.org>

- HyPhy (Kosakovsky Pond, Muse, Frost, and Poon)
- Geneious (Biomatters)
- SplitsTree (Huson and Bryant) – version 4 is a Java reimplementation.
- “MANTiS: a phylogenetic framework for multi-species genome comparisons” by Tzika A.C., Helaers R., Van de Peer Y. & M.C. Milinkovitch *Bioinformatics* 2008 **24**(2):151-157  
[www.mantisdb.org](http://www.mantisdb.org)
- MetaPiga2 is under development
- ...