



# Comparative methods in R hackathon

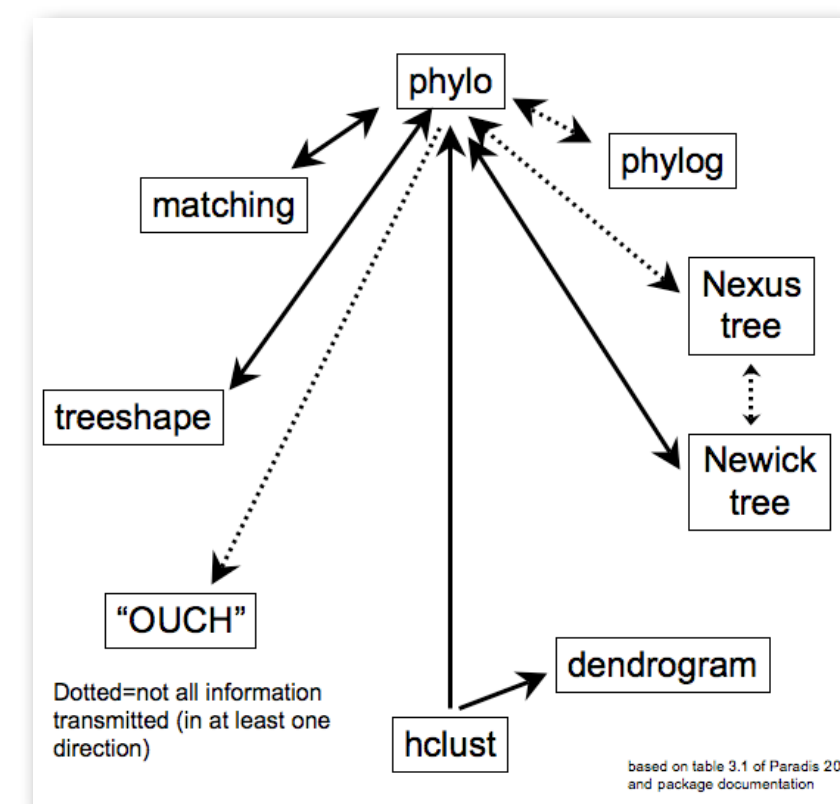
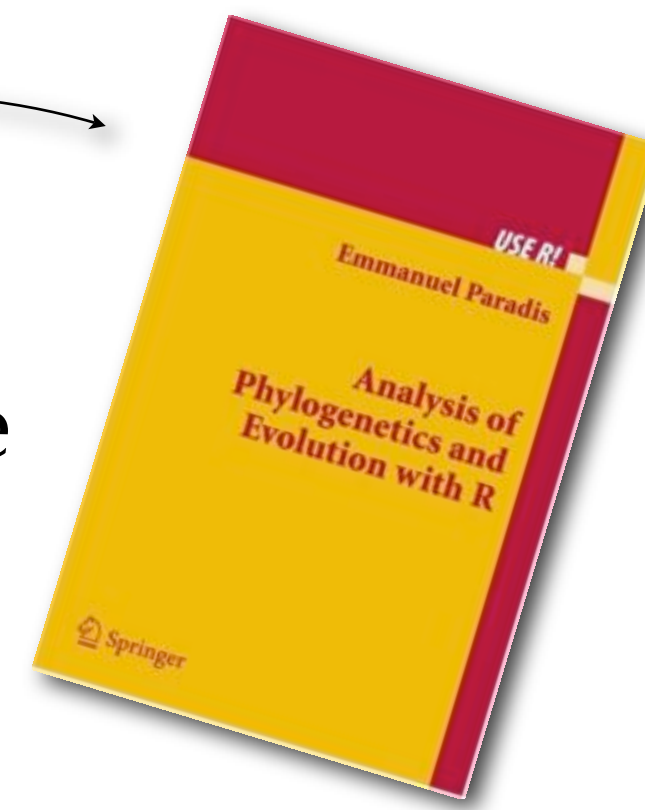
URL: [http://hackathon.nescent.org/R\\_Hackathon\\_1](http://hackathon.nescent.org/R_Hackathon_1) email: [hackathon2@nescent.org](mailto:hackathon2@nescent.org)

**Brian O'Meara**, Michael Alfaro, Charles Bell, Ben Bolker, Marguerite Butler, Peter Cowan, Damien de Vienne, Richard Desper, Joe Felsenstein, Luke Harmon, Christoph Heibl, Andrew Hipp, Gene Hunt, Thibaut Jombart, **Steve Kembel**, **Hilmar Lapp**, Scott Loarie, Wayne Maddison, Peter Midford, David Orme, Emmanuel Paradis, **Sam Price**, Dan Rabosky, Brian Sidlauskas, Stacey Smith, Dave Swofford, **Todd Vision**, Peter Waddell, **Amy Zanne**, Derrick Zwickl *[bold indicates organizer]*



## Rationale

The R statistical analysis package has emerged as a popular platform for implementation of powerful comparative phylogenetic methods to understand the evolution of organismal traits and diversification. It includes methods such as independent contrasts, ancestral state estimation, various models of continuous and discrete trait evolution, lineage through time plots, diversification tests, generalized estimating equations, tree plotting, and more. This event was designed to bring together active R developers as well as end-users working on the integration of comparative phylogenetic methods within R to actively address issues of data exchange standards, code interoperability, usability, documentation quality, and the breadth of functionality for comparative methods available within R. The idea originated from a whitepaper submitted by NESCent postdocs Amy Zanne and Sam Price.



Incompatible tree formats are used in different R packages

Package	Function
geiger1.0-9.1	sim.char
ouch1.2-4	brown.dev
picante	evolve.brownian
ape2.01	evolve.phylo

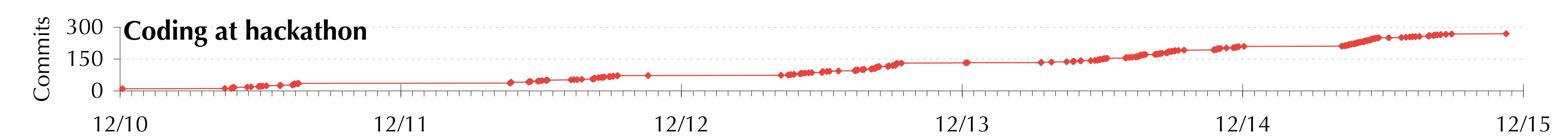
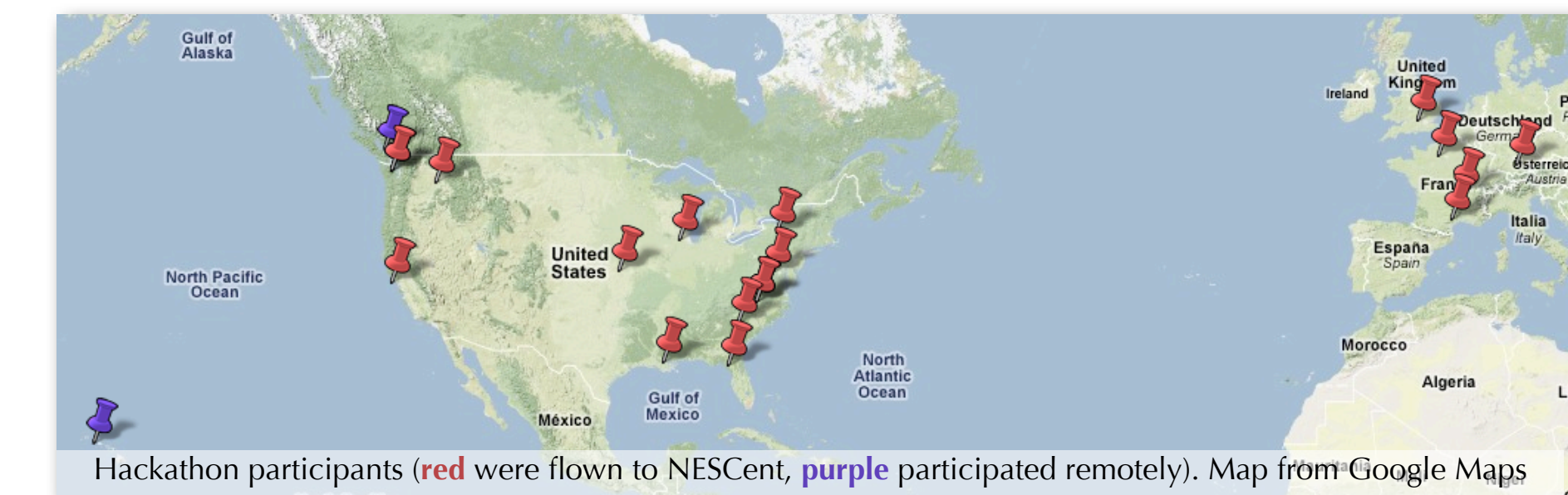
Redundancy (at least four functions to evolve traits up the tree using simple Brownian motion)

```
R Console
> data(bird.orders)
> x <- rep(0, 5)
> names(x) <- c("A", "B", "C", "D", "E")
> anc1 <- evolve.phylo(bird.orders, x, 1)
> anc2 <- evolve.phylo(bird.orders, x, 1)
> cor(anc1$tip.character, anc2$tip.character)
      A      B      C      D      E
A 0.31078191 -0.09568236 0.14090088 -0.06155229 -1.262620e-01
B -0.05132330 -0.02342825 -0.2522649 -0.18719187 7.461992e-02
C -0.09863444 0.05658986 -0.0637574 0.14725319 -9.226246e-05
D 0.03811630 0.11142563 -0.3654160 0.32579122 2.230738e-02
E 0.01063992 0.18845148 -0.2677129 0.02448515 1.700041e-02
```

Can be intimidating to beginners

## Work at hackathon (Dec. 10-14, 2007)

- 30 developers and users worked on programming & writing documentation
- Split into subgroups on diversification, divergence times, documentation, class design, Mesquite-R interaction, input/output, and trait evolution
- Package source code stored on shared repository hosted at R-forge ("PhyloConductor")



- Designed and began implementing a new S4 class for data and trees
- Ran "bootcamps" for developers on numerical optimization and S4 coding
- Used the Nexus Class Library (Lewis & Holder) and RCpp (Samperi) for reading and interpreting Nexus tree and data files
- Began work on R tutorials
- Tested existing methods in R, identifying errors
- Developed ways for R to call Mesquite and Mesquite to call R

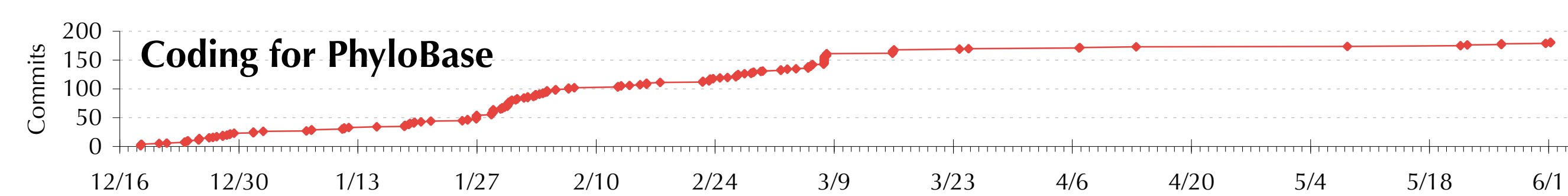
## Outcomes: Community

- **R-Phylo Wiki** (<http://www.r-phylo.org>): Tutorials and overview of available analyses and packages from the hackathon have been placed on a public website for all to use and improve. It's had >7,000 page visits from >30 countries and >600 edits since it went live in March 2008.
- **R-sig-phylo mailing list** (<https://stat.ethz.ch/mailman/listinfo/r-sig-phylo>): A mailing list for users of R for comparative methods and phylogenetics. Over 100 messages in its first four months.
- Comparative methods in R user tutorials planned for 2009 Society for Integrative and Comparative Biology and Evolution meetings.
- Addition of **R track** to NESCent summer course in phyloinformatics, featuring software developed at hackathon and taught by hackathon participant Marguerite Butler.
- Proposal to NSF for summer course in R for phyloinformatics.
- Ongoing collaborations between hackathon participants.
- Two Google Summer of Code projects to sponsor student developers:
  - Peter Cowan: Tree and data plotting in the **phylobase** project (see right)
  - Matthew Helmus: Enhancing the representation of ecophylogenetic tools in R in the **picante** project



## Outcomes: Software

- **Phylobase** (<http://r-forge.r-project.org/projects/phylobase/>): New package for phylogenetic trees and data. Can load trees and data from Nexus files, output to other tree formats, coordinate pruning of taxa from data and tree, traverse tree, handle DNA, morphological, and continuous data types. Work is ongoing (below) to enhance tree plotting and other functions. As with all hackathon products, new developers are welcome to join to further improve the code (one already has).



- Movement of existing packages to source code repositories allowing more collaborative development (i.e., Picante package has new Google Summer of Code 2008 developer Matthew Helmus)
- **R-Mesquite** interaction: Code written to allow Mesquite (Maddison & Maddison, 2007) to call R packages (such as OUCH (Butler & King 2004) and APE (Paradis et al. 2004)), and for R to call headless Mesquite, although easier installation needs to be created.
- Continuing improvement and release of packages by hackathon participants (**GEIGER**, **LASER**, **ape**).
- See [http://hackathon.nescent.org/R\\_Hackathon\\_1](http://hackathon.nescent.org/R_Hackathon_1) for more info.

## NESCent informatics

The US National Evolutionary Synthesis Center (<http://www.nescent.org>) encourages synthetic, interdisciplinary, and transformative research in evolutionary biology. NESCent, a collaborative effort of Duke, NC State University and UNC Chapel Hill, is located in Durham NC and is supported by the National Science Foundation (EF-0423641).



A major goal of NESCent's Informatics branch is to promote community-driven, collaborative open-source software development. This is achieved through hackathons, internships (such as the Google Summer of Code), summer courses, conference workshops, and by externally funded collaborations for the development and support of important cyberinfrastructure resources. NESCent accepts whitepapers that provide suggestions for future informatics activities from anyone at any time. See the website or the NESCent booth in the exhibit hall for more information.

