Phylogenetics

GEIGER: investigating evolutionary radiations

Luke J. Harmon1,2,*, Jason T. Weir3, Chad D. Brock4, Richard E. Glor5 and Wendell Challenger6

1Department of Biological Sciences, University of Idaho, Moscow, ID 83844, USA, 2Biodiversity Research Centre, 3Department of Zoology, University of British Columbia, Vancouver, BC, V6T1Z4, Canada 4IGERT Program in Evolutionary Modeling, Washington State University, Pullman, WA 99163, 5Department of Biology, University of Rochester, Rochester, NY 14627, USA and 6Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, V5A1S6, Canada

Received on May 31, 2007; revised on September 18, 2007; accepted on October 21, 2007
Advance Access publication November 15, 2007
Associate Editor: Keith Crandall

ABSTRACT

Summary: GEIGER is a new software package, written in the R language, to describe evolutionary radiations. GEIGER can carry out simulations, parameter estimation and statistical hypothesis testing. Additionally, GEIGER’s simulation algorithms can be used to analyze the statistical power of comparative approaches.

Availability: This open source software is written entirely in the R language and is freely available through the Comprehensive R Archive Network (CRAN) at http://cran.r-project.org/.

Contact: lukeh@uidaho.edu

1 INTRODUCTION

Phylogenetic trees are now available for a wide range of groups. Each of these trees contains a large amount of information about the history of diversification in the group of interest. Many tools have been developed to extract this information (Freckleton and Harvey, 2006; Nee, 2001; O’Meara et al., 2006). However, the required software is diffuse, working on different platforms and often requiring unique data formats.

Here, we describe a new software package that is dedicated to the analysis of phylogenetic comparative data. We call this package GEIGER, since its main purpose is to detect and describe evolutionary radiations. GEIGER is written in the cross-platform software language R and complements four other packages: APE (Paradis et al., 2004), ape (Bortolussi et al., 2006), OUCH (Butler and King, 2004) and LASER (Rabosky, 2006). In addition to permitting a range of tests that are not implemented by any of these packages, GEIGER’s simulation algorithms can be used to analyze the statistical power of comparative approaches implemented in other packages. Below, we describe the features of the software and illustrate a range of questions that it might be used to address. Full descriptions of all functions are found in GEIGER’s online help files (http://cran.r-project.org/src/contrib/Descriptions/geiger.html).

*To whom correspondence should be addressed.
GEIGER also allows simulation of both discrete and continuous character evolution on phylogenetic trees. For a discrete character, one must specify the character state at the root of the tree and a transition matrix that includes instantaneous rates of change among all possible states. The character can have any number of states, and the rate matrix need not be symmetrical. One can also simulate two or more correlated discrete traits by specifying a transition matrix that describes the transitions among all possible combinations of characters (Pagel, 1999a).

To simulate \( n \) continuous characters, the program requires an \( n \times n \) evolutionary variance–covariance matrix. This matrix describes the expected variances and covariances among characters per unit time; characters evolve under a multivariate Brownian motion model described by this matrix. Although numerous programs are available for the simulation of evolution via Brownian motion, GEIGER permits these types of simulations to be combined with tree simulations. For example, using GEIGER, one can grow birth-death trees and simulate the evolution of a set of characters on those trees.

### 2.2 Parameter estimation

GEIGER allows estimation of several key parameters associated with species diversification and character evolution. We focus on estimates of average rates of net diversification (speciation-extinction). There are currently three approaches to estimating this rate. One is based on waiting times between successive speciation events and is already implemented via the APE package; this method requires a complete phylogenetic tree with accurate ultrametric branch lengths. The two types of rate estimates provided by GEIGER require less certainty about the phylogenetic tree, but they are also less powerful and may result in estimates with larger sampling error. One estimate relies strictly on a clade’s age and species diversity (Magallon and Sanderson, 2001). The second uses the sum of all branch lengths in an ultrametric tree to obtain the Kendall–Moran estimate of diversification rate (Nee, 2001).

To estimate parameters associated with the evolution of multivariate continuous characters, we have implemented a new approach described in Revell et al. (2007) that obtains an unbiased estimate of the multivariate variance–covariance matrix for traits evolving under multivariate Brownian motion. For discrete characters, we provide a function to find the maximum-likelihood transition rate \( q \) for multistate characters with equal and symmetrical transition rates among character states. This function implements Felsenstein’s pruning algorithm (Felsenstein, 1981) in R, and provides a framework for future work using more complex models.

### 2.3 Hypothesis testing

Although numerous programs are available for the simulation of evolution, GEIGER provides a unified framework within which to implement several new types of tests as well as previously available tests that are also possible in other independent programs. First, GEIGER may be used to test whether particular clades are extraordinarily diverse or undiverse by calculating the probability of obtaining a number of extant species larger or smaller than the realized number given some specified rate and interval of time. For this calculation, we have corrected a code-error in Magallon and Sanderson’s (2001) equation 11a; the corrected version is available from the corresponding author.

GEIGER can also calculate the relative cladogenesis statistic, which examines the distribution of numbers of descendents for each branch in a phylogenetic tree existing at some specified time. Under a homogeneous model, the number of descendents should follow a known distribution (Schluter, 2000); lineages with more, or fewer, descendents than expected under this distribution may be hypothesized as lineages that were exceptionally successful or unsuccessful. This test, applied to the Galapagos finches, is shown in Figure 1A.

For character evolution, GEIGER provides a suite of functions designed to investigate the tempo and mode of evolutionary change on a phylogenetic tree. First, GEIGER includes functions to carry out tests of various likelihood models for both continuous and discrete data (Pagel, 1999b). These tests compare the fit of a constant rate Brownian model to various alternatives, including concentrations of character change early or late in the tree or at speciation events, and constrained models of evolution (Butler and King, 2004). Importantly, one can directly compare the likelihood fit of all of these models in an AIC framework. Second, GEIGER implements the approach for comparing levels of phenotypic disparity within and among clades that is described in Harmon et al. (2003). This approach provides a running average disparity for clades of a given age range through the history of the tree, and compares that to the expectation under a null model of Brownian motion (illustrated for the Galapagos finches in Fig. 1B). Finally, GEIGER can carry out phylogenetic ANOVA or MANOVA using simulation (Garland et al., 1993).

### 3 CONCLUSION

As the availability of phylogenetic trees increases, there is an increasing need to develop flexible software packages that permit a variety of comparative analyses. GEIGER expands existing phylogenetic and comparative utilities via the freely distributed, cross-platform R framework. Furthermore,
extensions to our software are facilitated by the flexibility of the R language.

ACKNOWLEDGEMENTS

We thank S. Rogers, A. Mooers, D. Schluter and W. Maddison for helpful comments. L.J.H. was funded by the University of British Columbia Biodiversity Research Centre. C.D.B’s work was supported by an NSF IGERT Fellowship (NSF BCS-0549425).

Conflict of Interest: none declared.

REFERENCES