



MacroCAIC: revealing correlates of species richness by comparative analysis

PAUL-MICHAEL AGAPOW* and NICK J. B. ISAAC *Department of Biology, Imperial College at Silwood Park, Ascot, Berks SL5 7PY, U.K.*

Abstract. Studies of species richness have been hampered by the use of methods that fail to account for phylogenetic non-independence of character states. MacroCAIC is a computer program that extends the method of phylogenetically independent contrasts to encompass species-richness data. It examines user-selected characters

for correlation with species richness, thus allowing clearer identification of the factors driving large-scale patterns of diversity.

Key words. Species richness, phylogeny, comparative analysis, independent contrasts, macroevolution.

Clades show a wide distribution of species level diversity, even between closely related organisms. As only some of the variation can be explained by random factors (Purvis, 1996), identification of the factors that lead to species richness/paucity is essential for an understanding of large-scale patterns in biodiversity. Numerous hypotheses have been proposed and tested via taxonomic analyses of clades with different attributes (e.g. Slowinski & Guyer, 1989; Nee *et al.*, 1996). For example, a genus in one clade may represent an older (and thus perhaps more speciose) group than that in another. Also, a naive comparison of traits is thwarted by the non-independence of data across a phylogeny. Related clades may share characters via common descent rather than similar evolutionary pressures (Harvey & Pagel, 1991). This non-independence invalidates standard tests of correlation between characters, including species richness. For example, a naive examination of avian biodiversity might identify several characteristics of the passerines as promoting species richness, neglecting the possibility that these traits are seen in the group because they were present in their common ancestor.

Making comparisons solely between sister clades avoids this problem but sharply reduces the number of contrasts that can be drawn from a phylogeny.

MacroCAIC is a computer program that embodies a new method to correctly extracting correlates of species richness from a phylogeny. It uses a modification of comparative analysis by independent contrasts (CAIC) methodology (Purvis & Rambaut, 1995) to detect which comparisons within a phylogeny can attributed to independent evolutionary events, even in phylogenies that include polytomies. While there are several packages that calculate independent contrasts in inherited characters, the standard approach is inappropriate when dealing with species richness. For inherited traits, the character values used for calculating contrasts higher within the tree are essentially the average of more derived values. As this is clearly incorrect for calculating species richness, MacroCAIC instead calculates the richness of higher clades as the sum of their constituent clades. This general approach has been validated by an extensive set of simulations under a wide variety of evolutionary conditions (N. B. Isaac, P.-M. Agapow *et al.*, unpublished data). Note that MacroCAIC handles only continuous, not discrete characters. (See Purvis, 1996 for a review of methods handling categorical data.)

* Correspondence and present address: Paul-Michael Agapow, Department of Applied Statistics, University of Reading, Reading RG6 6FN, U.K., E-mail: p.m.agapow@reading.ac.uk

Table 1 Measures of species-richness correlates. X is the value of the contrasted independent character; N_1 is the species richness of the clade with the largest value of X , N_2 is the species-richness of the contrasted clade, and t is the time or distance separating the two. *RateDiff* is thus the difference in diversification rate of two sister-clades, *LnCladeRatio* and *PropnBigX-1/2* two simple time/distance independent contrasts of species-richness

Species-richness metric	Formula
<i>RateDiff</i>	$\frac{\ln N_1}{t} - \frac{\ln N_2}{t}$
<i>LnCladeRatio</i>	$\ln \frac{N_1}{N_2}$
<i>PropnBigX-1/2</i>	$\frac{N_1}{(N_1 + N_2)} - \frac{1}{2}$

The input data (i.e. the phylogeny and character values) consist of a series of plain text files in tab-delimited format similar to that used by other programs (Purvis & Rambaut, 1995), easily prepared by hand, or converted from other formats. The phylogeny need not have branch lengths (i.e. be dated) and in their absence dates can be extrapolated based on a punctuational or saltational model of evolution. The supplied phylogenies can also be higher than species level (e.g. genera), which is important where the lower relationships have not been clearly resolved. In this case, the species richness of every terminal taxon can be described by the user in an additional file. Character data are presented with character states as integers or floating-point numbers, and optional column names for easy identification of data. Not every character state of every terminal taxon need be known for analysis, with those of an uncertain value marked as missing.

MacroCAIC allows the user to select groups of one or more characters for correlation with species richness. It outputs results in tab-delimited text files, identifying the node at which the contrast took place, contrasts for the examined characters (X), the extrapolated higher values, and three different measures of the species-richness contrast (δS). These are described in Table 1 and calculated such that each is positive when the clade with the larger value of the independent

character, X , contains more species than its sister clade. *RateDiff* (a measure of the difference in the rate at which the two clades have accumulated species) is the most obvious of the measures, but cannot be calculated when branch lengths are set equal due to nodal ages being unknown. Furthermore, simulation analysis suggests that *RateDiff* is particularly sensitive to variations in tree shape and departures from models of character evolution (N. B. Isaac, P.-M. Agapow *et al.*, unpublished data). The statistical performance of the other measures are virtually indistinguishable from one another. MacroCAIC also automatically performs analysis on the contrasts when a single independent variable is chosen. Association between contrasts in X and δS is assessed by regression through the origin, along with tests of the assumptions of Brownian motion and homogeneity of variance in both the contrasts and residuals of the regression model. Finally, an option exists for data exploration in which each character is run through the program and subjected to the tests described above.

MacroCAIC is currently available as a Macintosh binary at <http://www.bio.ic.ac.uk/evolve/software/macrocaic/>, but may easily be ported to other platforms. Even relatively large data sets may be explored within minutes, allowing easy searching for macroevolutionary correlates. It has already been used to study species richness over a range of phylogenies, including hoverflies (Katzourakis *et al.*, 2001) carnivores and primates (Gittleman & Purvis, 1998). It provides a robust way of testing possible correlates of species richness, giving an insight into the process that shape large-scale patterns of biodiversity.

We would like to thank Andy Purvis and Andrew Rambaut for the early development of MacroCAIC, and Aris Katzourakis, David Orme and Paul Harvey for advice. This work was supported by grant GR3/11526 from the Natural Environment Research Council (NERC).

REFERENCES

- Gittleman, J.L. & Purvis, A. (1998) Body size and species-richness in carnivores and primates. *Proceedings of the Royal Society of London B* **265**, 113–119.
- Harvey, P.H. & Pagel, M.D. (1991) *The comparative method in evolutionary biology*. Oxford University Press, Oxford.

- Katzourakis, A., Purvis, A., Azmeh, S., Rotheray, G. & Gilbert, F. (2001) Macroevolution of hoverflies (Diptera: Syrphidae): an analysis of phylogeny and adult characters. *Journal of Evolutionary Biology* **14**, 219–227.
- Nee, S., Barraclough, T.G. & Harvey, P.H. (1996) Temporal changes in biodiversity: detecting patterns and identifying causes. *Biodiversity: a biology of numbers and difference* (ed. by K.J. Gaston), pp. 230–252. Blackwell Science, Oxford.
- Purvis, A. (1996) Using interspecies phylogenies to test macroevolutionary hypotheses. *New uses for new phylogenies* (ed. by P.H. Harvey, A.J. Leigh Brown, J. Maynard Smith and S. Nee), pp. 153–168. Oxford University Press, Oxford.
- Purvis, A. & Rambaut, A. (1995) Comparative-analysis by independent contrasts (CAIC) — an Apple-Macintosh application for analyzing comparative data. *Computer Applications in the Biosciences* **11**(3), 247–251.
- Slowinski, J.B. & Guyer, C. (1989) Testing the stochasticity of patterns of organismal diversity: an improved null model. *American Naturalist* **134**, 907–921.