# "FLEXIBILITY" AS A TRAIT AND METHODOLOGICAL ISSUES IN SPECIES DIVERSITY VARIATION AMONG ANGIOSPERM FAMILIES

Jonathan Silvertown,[1,2] Kevin J. McConway,[3] Mike E. Dodd,[1] and Mark W. Chase[4]

[1]*Ecology and Conservation Research Group, Department of Biological Sciences, Open University, Milton Keynes MK7 6AA, United Kingdom*
[2]*E-mail: j.silvertown@open.ac.uk*
[3]*Department of Statistics, Open University, Milton Keynes MK7 6AA, United Kingdom*
[4]*Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, United Kingdom*

## The Utility of Tree-Based Comparative Methods

Ricklefs and Renner (2000) and Ricklefs and Starck (1997) argue that comparative studies that have used phylogenetically independent contrasts (PIC) usually add little or nothing to the results that can be obtained using comparative methods that ignore phylogeny. Although it is true that the results of Dodd et al (1999) are remarkably similar to those of Ricklefs and Renner (1994), this demonstrates the robustness of the latter's results rather than the redundancy of the phylogenetic methods used by Dodd et al (1999) in their reanalysis. There are many instances where PIC and cross-species comparative analyses (TIP) produce similar results (Ricklefs and Starck 1997, Ackerly 2000), but there are also exceptions which caution against the assumption that phylogeny can be routinely ignored. For example, Tofts and Silvertown (2000) found quite different results from PIC and TIP analysis of twelve plant traits including a major effect of one trait affecting community assembly that was only revealed to be significant using PIC.

TIP analyses are prone to pseudoreplication and simulation studies have demonstrated the unacceptably high Type I error rate of the method (Martins and Garland 1991). Why, then, do so many empirical studies show that PIC and TIP results concur? One possible reason is that biologists choose to test patterns of trait correlation which they have good reason to believe already exist in their data. This was certainly true of Ricklefs and Renner (1994), whose work confirmed the earlier results of Eriksson and Bremer (1992) and others. Science proceeds by pursuing positive results in preference to negative ones (Silvertown and McConway 1997), but to forget this risks complacency. To test the idea that concurrence of PIC and TIP results is influenced by how hypotheses are selected, we have performed a Bayesian analysis of the two methods.

Consider a study where a correlation under investigation is significant ($P < 0.05$), using a TIP procedure. As is well known, the $P$-value derived from a statistical test is not the probability that the null hypothesis is true; it is the probability of getting results as extreme as those actually observed, under the assumption that the null hypothesis is true. Using a Bayesian approach to calculate the probability that the observed correlation really exists in nature and is not simply the result of random sampling variability, one needs to know the "pri-

or" probability, unconditional on the observed data, that the hypothesis of a real correlation in nature is actually true. If the correlation being investigated had simply been chosen at random from all the possible trait correlations, the prior probability that it is real would be likely to be very low—for illustration let us take it as 0.01. If, as is much more realistic in our view, the correlation had been selected for investigation because it looked potentially interesting on theoretical grounds and because of indicative positive results from previous studies, the prior probability of a real correlation would be much higher—0.5 is a plausible value. Indeed, if the research depends on funding, an even higher prior probability such as 90% might well be appropriate, because a study for which the prior evidence of a real correlation was only 50% would be unlikely to be funded.

The other necessary inputs to a crude Bayesian analysis are the probabilities of positive and negative results, which are conditional on the actual correlation present in nature. Information on plausible sizes for these probabilities can be obtained from simulation studies, where it is known what the true "state of nature" is. For example, the simulation study of Purvis et al. (1994, tables 1, 2) found that TIP correctly identified significant correlations ($P < 0.05$) in 58.6% of cases and PIC did so in 68.6% of cases. Where the true correlation was actually zero, TIP erroneously found a significant correlation in 31.4% of cases and PIC in 12.4%. Using these figures and Bayes Theorem, we compared the performance of PIC and TIP with the prior probability of a real correlation set at 0.01, 0.50 or 0.90, and obtained the results shown in Table 1. The results show the huge impact that the prior probability of a true correlation can have on the interpretation of a significant test result. With a prior probablity of 0.50, TIP performs only 15% better than guesswork (0.651–0.500, Table 1).

Similar calculations can be used to investigate the extent to which TIP and PIC results are likely to concur. To do so fully would require results from simulation studies on how often the results of the two methods agree, in terms of giving statistically significant correlations. Such information is not available in reports of simulation studies, because it is not directly relevant to what they are investigating. However, we can make some progress by noting that the observed perfor-