

## Testing evolutionary hypotheses in the classroom with MacClade software

Sylvio G Codella

Department of Biological Sciences, Kean University, New Jersey, USA

MacClade is a Macintosh-based software package that uses the techniques of cladistic analysis to explore evolutionary patterns. Although designed for the Macintosh platform, MacClade delivers excellent results on PCs when used with MacOS emulators. This paper describes a novel and effective exercise that allows undergraduate biology majors to test a hypothesis about behavioural evolution in insects. The exercise is derived from a real data set in the literature and provides students with the opportunity to learn firsthand how scientists evaluate hypotheses about patterns of adaptive change. The exercise can be modified to fit the needs of particular classes. In general, MacClade offers the life science teacher a wide variety of instructional possibilities.

**Key words:** Adaption, Cladistics, Computer, Evolution, Phylogenetic analysis, Software.

### Introduction

The concept of organic evolution unifies the breadth of modern biology, from the molecular through ecological levels of organisation. Yet natural science undergraduates are often sceptical when told that evolutionary hypotheses can be stated formally and tested with rigor. Their surprise is due in part to the unfortunate lack of attention paid to evolutionary concepts in many biology curricula. More practically, the problem stems from a lack of workable assignments that illustrate just how biologists go about evaluating hypotheses about patterns of adaptive change. As a result, thoughtful students sometimes see evolutionary studies as descriptive and speculative at best.

This regrettable state of affairs can be rectified through the implementation of class activities that allow students to experience evolutionary hypothesis testing first-hand. MacClade (Sinauer Associates, Sunderland, MA, USA) is a software package that uses the techniques of cladistic analysis to explore evolutionary patterns (Maddison and Maddison, 2000). Although MacClade was designed for the Macintosh platform, it also delivers excellent results on PCs when used with MacOS emulators such as Executor (Abacus Research and Development, Inc., Albuquerque, NM, USA). MacClade is a powerful research tool and an excellent instructional package as well. The software permits the easy input and manipulation of phylogenetic trees and provides statistical feedback so that results can be evaluated. MacClade can be used to introduce evolutionary hypothesis testing to undergraduates, at a greater or lesser level of detail as appropriate, in such courses as General Biology, Zoology, Evolutionary Biology and Animal Behaviour. This paper presents a novel and effective class exercise derived from a study that evaluates a hypothesis about correlated character evolution in insects (Ridley, 1990). The exercise allows

students, either individually or in teams, to readily apply sophisticated techniques to a real data set. The extent and complexity of the assignment can be adapted to the level and needs of a particular class. Students typically complete the exercise in about two hours, so it is appropriate as either an in-lab project or an out-of-class assignment. The mechanics of the software package are not explained in detail here; MacClade is very user-friendly, and students have little difficulty with it. Rather, the aim is to illustrate the value of MacClade as a teaching aid and to encourage its use in the classroom.

This exercise involves a question in behavioural ecology; instructors and students can find background information in such texts as Krebs and Davies (1993) and Alcock (2001). In order to benefit fully from the exercise as presented, students should be familiar with the concepts of natural selection and adaptation, as well as with the basic philosophy of cladistic classification (e.g. Hickman *et al.*, 2001 and many other undergraduate texts). The instructor, as well as advanced students, can find more sophisticated discussions of cladistics and phylogenetic analysis in Harvey and Pagel (1991), Wiley *et al.* (1991), Martins (1996) and Kitching *et al.* (1998), as well as in the MacClade documentation (Maddison and Maddison, 2000). Finally, knowledge of elementary statistical concepts (null and alternative hypotheses, *P*-values, etc.) plays an important role in this particular exercise. The assignment can be used as a working introduction to all of these ideas if the instructor provides some additional background beforehand.

### The Hypothesis: male mating control favours high female mating frequency

Ridley (1990) was interested in whether female mating frequency in insects was influenced by which sex was "in

control" of mating. For example, in a mating system where females release a long-distance sex pheromone, males can only locate a mate if she is engaged in chemical signalling. Since most female insects can store sperm, a single mating may provide enough sperm for all of her eggs. She can then stop releasing pheromone and direct her energies elsewhere. Thus, she controls the male's ability to locate her.

In contrast, many dragonfly mating systems are under male control. Males defend territories containing resources that females need for reproduction — good egg-laying sites, for example (Alcock, 2001). Since a male's reproductive success is correlated with the number of females he inseminates, Ridley predicted an association between a high per capita mating frequency in females and male control of mating. This is expected because females must mate with such males in order to have access to the resources they defend, and may, therefore, mate more frequently than their sperm needs dictate. So, the question could be phrased: Does male control favour the evolution of multiple female mating? Or: Is female control a constraint on the evolution of multiple female mating? This hypothesis about behavioural adaptation is readily addressed with MacClade [Note: Ridley's (1990) actual analysis differs from mine; see also Harvey and Pagel, 1991].

### Preliminaries to the exercise

Students are provided with detailed instructions beforehand (Codella, 2000). Prior to the exercise, they complete the MacClade tutorial, which is itself a valuable learning experience. The tutorial uses vertebrate phylogeny to explain the mechanics and focus of the software, as well as to introduce some basic phylogenetic concepts. Through the tutorial, students learn that a MacClade file consists of a character matrix, wherein character states can be expressed either in words, or, more commonly, as numeric codes (e.g. 0 = wings absent; 1 = wings present; see Table 1). These data are then used to generate phylogenetic trees, or cladograms. Using several readily accessed descriptive statistics (e.g. treelength, consistency index), students see that some trees "fit" the matrix better than others do. The vertebrate tree included in the tutorial does not fit the data particularly well. By clicking on a tree branch, students can literally lift and place it at another location on the tree and observe the change, if any, in the statistics. This provides students with an opportunity to use real data, rather than their preconceived notions about vertebrate relationships, to construct a more robust cladogram.

In preliminary class discussions, two critical and related points are stressed. First, one must consider the evolutionary relationships within the Class Insecta in order to test the hypothesis in a meaningful way. Related taxa share many characteristics due to common ancestry. Consider Taxa I, J and K in Figure 1a. The simplest explanation for any similarities among the three is that the shared traits were inherited from a common ancestor. Therefore, the branching pattern of a particular tree affects our tally of trait occurrences, and one must count evolutionary *events*, not taxa, when evaluating hypotheses such as the one presented in this paper. Figure 1a, then, shows us that the character of interest has evolved four times. It also shows that the trait has been lost once; in the case of Taxon M, the branch colour has changed from black to white, rather than vice versa. This amounts to an evolutionary reversal.

**Table 1** Data matrix for insect correlated character evolution (after Ridley 1990).

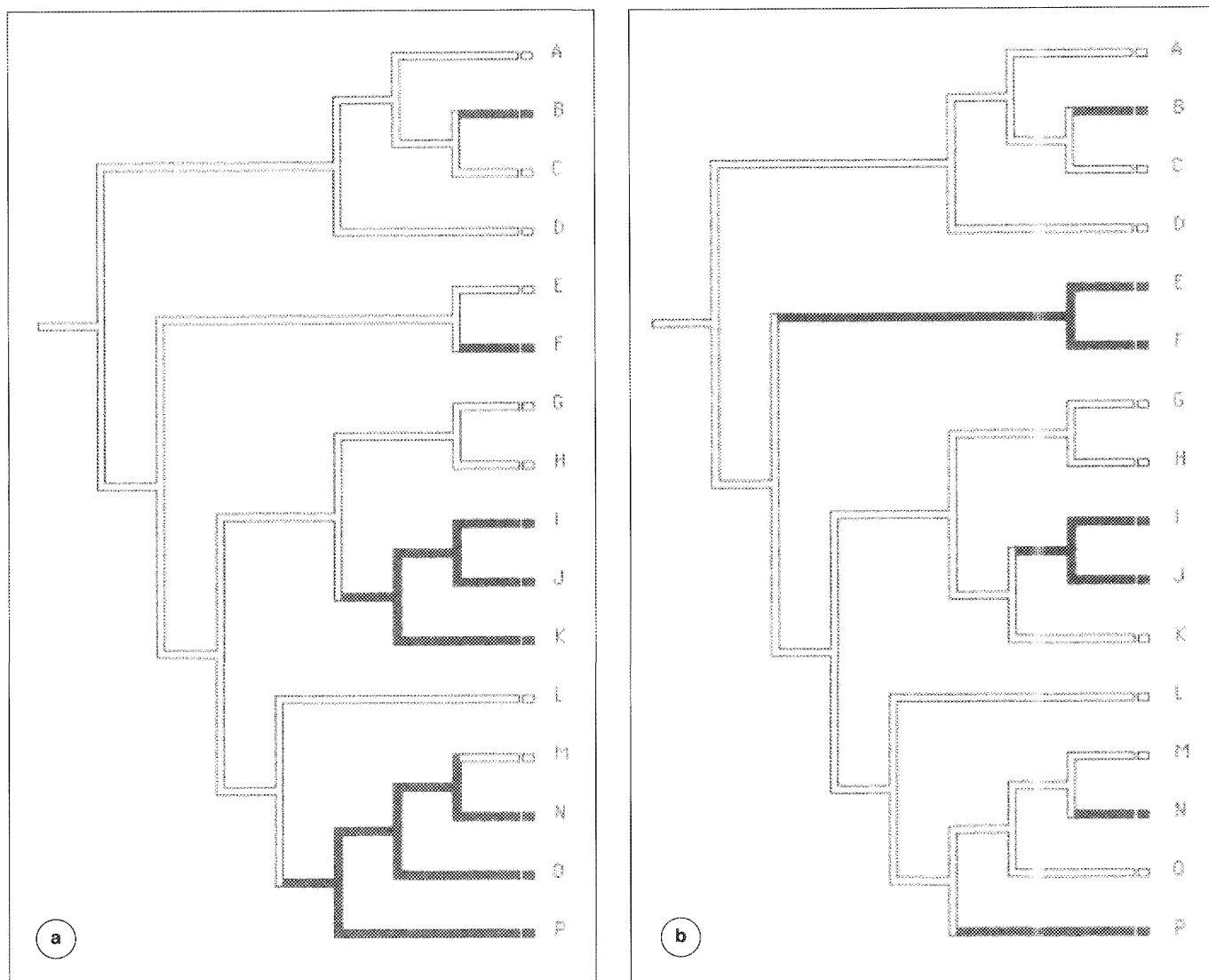
Insect Taxon	Mating Control	Mating Frequency
EPHEMEROPTERA	0	0
ODONATA	1	1
<i>Ischnura</i>	0	0
BLATTARIA	0	0
ORTHOPTERA	0	1
Acrididae	1	1
PSOCOPTERA	0	0
HETEROPTERA	1	1
HYMENOPTERA	0	0
Eumenidae	0	0
<i>Abispa ephippium</i>	1	1
<i>Paralastor</i>	1	0
Sphecidae	1	1
<i>Philanthus bicinctus</i>	1	0
<i>Andrena vaga</i>	0	0
Andrenidae	1	1
Halictidae	1	0
Megachilidae	1	1
<i>Centris pallida</i>	0	0
Anthophoridae	1	0
COLEOPTERA	1	1
Lampyridae	0	0
Dermestidae	0	0
Nematocera	0	0
Empididae	0	1
Syrphidae	1	1
Pulicophora	0	1
Scatophaga	1	1
Tephritidae	0	0
LEPIDOPTERA	0	0

Mating Control: 0 (female), 1 (male); Mating Frequency (female): 0 (1); 1 (> 1).

The second point is that the hypothesis presented in this exercise is *directional* (Harvey and Pagel, 1991): the prediction is that when male control evolves, high female mating frequency tends to follow. The interest is not merely in the association of male control and frequent female mating, but specifically in cases where male control arises *prior to* high female frequency on the evolutionary tree (Figure 1a, b). This will have practical implications when students perform the analysis.

### The Exercise: testing the hypothesis with MacClade

Following Ridley (1990), the class is provided with a cladogram of the major insect orders, which also includes some subordinate taxa (Figure 2). The tree has been simplified somewhat for the purposes of the exercise, but it is based on a reasonable phylogenetic hypothesis for the Insecta (Kristensen, 1981). Two variables are encoded in the data set (Table 1). Using the Trace function of MacClade, students view the distribution of these characters, in sequence, on the cladogram. The character states are superimposed on the tree (Figure 2) and can be viewed either in colour or in black-and-white patterns. The first variable is "mating control," the *independent* (or *x*) variable (Figure 2a). The distribution of the trait on the tree indicates that male



**Figure 1** Counting evolutionary events. In Tree (a), trait x (the independent variable, in black) has evolved four times (B, F, I+J+K, N+O+P). The trait has been lost in M. Tree (b) shows that trait y (dependent; again in black) has evolved five times (taxa B, E+F, I+J, N, P). These are the total gains and losses. In three of these five instances, the dependent character has been preceded by the independent character (taxa I+J, N, P). These are the designated gains and losses. Taxon B has acquired both traits simultaneously. For taxon F, the dependent character has evolved before the independent character.

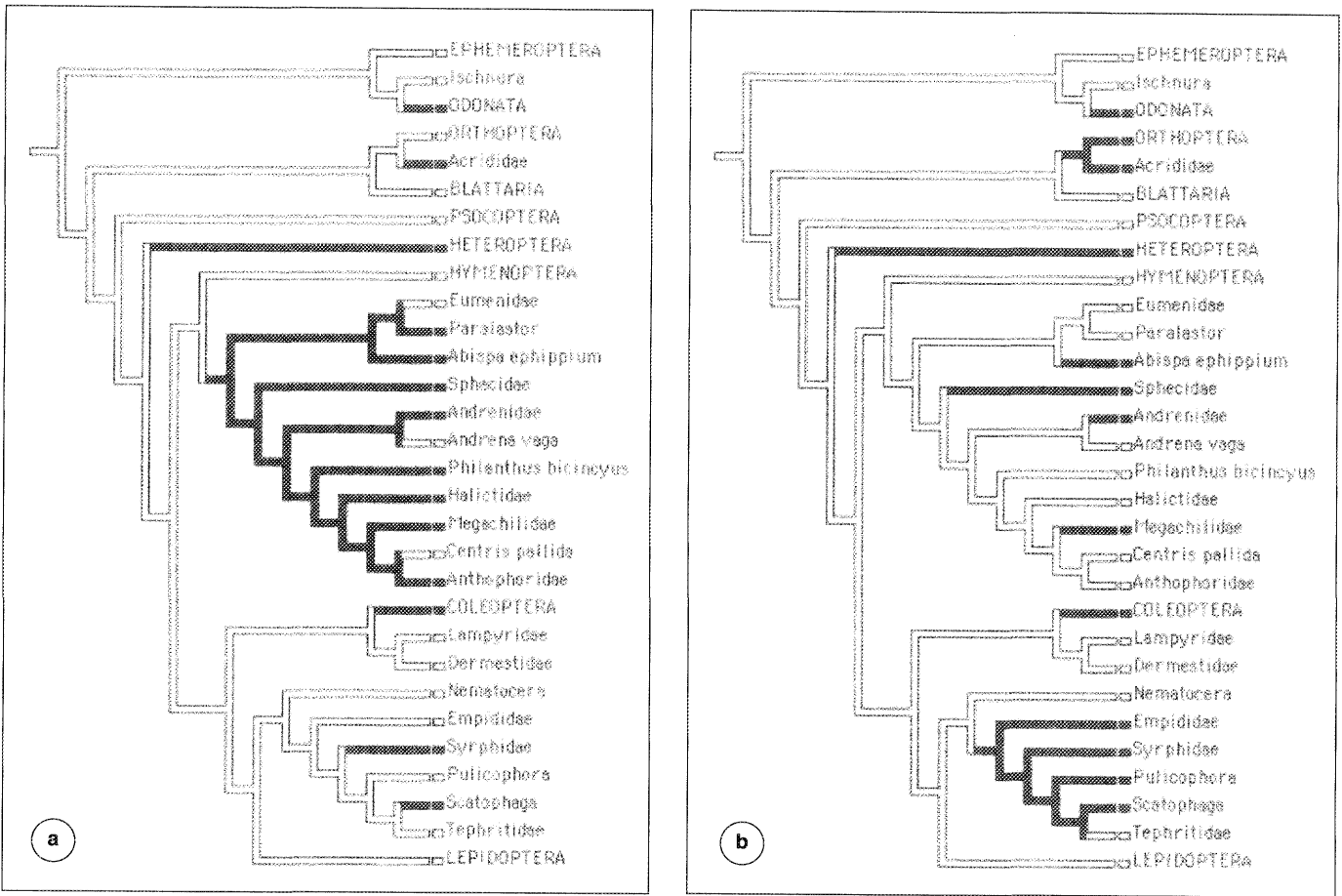
control has evolved seven times and been lost three times. The second (*dependent*, or *y*) variable is "mating frequency" (Figure 2b). Students count and find that a high frequency (defined here as mating more than once) has evolved nine times and been lost once; these values are defined as the *total* gains and losses. Again, some taxa share a high mating frequency due to common ancestry, so the instructor must repeatedly stress to students that they are to count the number of events, which may well be less than the number of taxa on any cladogram.

Next, students count the number of gains and losses of high mating frequency (dependent variable) that fall within the areas of male mating control (independent variable). This involves "togglng" back and forth between the two character tracings, and it is the most difficult part of the exercise for most students, both conceptually and practically (see Figure 1 for clarification). Students are again counting the gains and losses for mating frequency, but now only those on the black branches indicating male mating control (the designated branches). Thus, in Figure 2, "Sphecidae" is counted, since male control (*x*) evolved first.

However, "Acrididae" is not, as high female mating frequency (*y*) evolved first, and cases where *y* precedes *x* do not support the hypothesis. This count, the *designated* gains and losses, should therefore be less than or equal to the total gains and losses.

What about cases where branch colour changes simultaneously for both variables (e.g. "Heteroptera" in Figure 2a, b)? Should these be included in the designated gains and losses or not? The handling of such equivocal changes is controversial (Harvey and Pagel, 1991). In the initial part of this exercise, students are told to exclude these equivocal changes, which results in four designated gains and no losses. However, the effect of method choice on experimental outcome can be used to good instructional effect (see *Variations on the Exercise*, below), and the reader should now verify that if equivocal changes are included, the number of designated gains increases to seven (Figure 2a, b).

Students are now ready to perform the Concentrated Changes Test option of MacClade (Maddison, 1990; Maddison and Maddison, 2000) using their counts of gains and losses of high



**Figure 2** Phylogeny of the Insecta used in this exercise (after Kristensen, 1981). a) tracing of mating control, the independent variable in this exercise (white branches, female; black branches, male). b) tracing of female mating frequency, the dependent variable (white, 1; black, > 1).

mating frequency for both the whole tree and the designated areas. The test evaluates the null hypothesis that two discrete, dichotomous characters are randomly associated on a phylogenetic tree. Students input the required data in a sequence of two screens. In essence, MacClade runs a simulation and produces a sample of all possible ways that a specified number of total gains and losses can be distributed on a particular tree (again, the instructor should reinforce the idea that the branching pattern of a cladogram will affect this). The test then calculates the probability of having a specified number of designated gains and losses fall within the designated areas (independent variable) of the tree. Students use this *P*-value to reach a decision about the validity of the association between the two variables. A significant *P* (< 0.05) would support the alternative hypothesis that high female mating frequency follows the evolution of male control more consistently than expected by chance alone. This in turn would suggest that high female mating frequency is an adaptive response to male control that has been favoured by natural selection. However, the designated count obtained here (four gains, no losses; equivocal changes ignored), results in an insignificant *P* = 0.64. Thus, the present analysis has failed to refute the null hypothesis, which is in itself a good lesson for students preparing for research careers!

**Variations on the exercise**

The idea that male control of mating in insects favours the evolution of a high female mating frequency did not hold up in this particular analysis. Of course, things are usually not quite so

simple in real science, and the subtleties of phylogenetic analysis can be incorporated into this exercise at the instructor's discretion. As mentioned above, the inclusion of equivocal branch changes can vary the outcome of an analysis. In this data set, inclusion of such branches results in seven designated gains and no losses. This produces a significant result in the concentrated changes test (*P* = 0.012). An instructor wishing for a "cleaner" result to the exercise can limit the assignment to this configuration. For a more advanced assignment, students can perform the analysis both ways. By comparing the two sets of results, students can discuss the effect of method choice on experimental outcome and gain insights into its particular effect on phylogenetic analysis.

Students can learn for themselves about the effect of branching pattern on the outcome of evolutionary hypothesis testing. This is accomplished by moving branches into different arrangements (easy and fun with MacClade!), and repeating the procedure. This often generates surprisingly sophisticated discussion.

Also, students can explore the outcome when phylogeny is ignored, which was warned about earlier in this paper. By counting the number of taxa with each of the four possible character distributions — female control, high frequency; female control, low frequency; male control, high frequency; male control, low frequency — they can cast their results in a 2x2 contingency table and perform a chi-square analysis (Zar, 1996). For students familiar with elementary statistics, this may well be what their intuition suggested in the first place. Such an approach yields significant

results ( $\chi^2 = 6.4$ ,  $df = 1$ ;  $0.025 > P > 0.01$ ). If compared with the initial, more conservative run of the concentrated changes test (which ignored simultaneous changes and thereby gives insignificant results), students can learn first-hand how the choice of analysis can have marked effects on the outcome of a study.

### MacClade in introductory settings

This exercise can be simplified for introductory college courses. Students can learn a great deal simply by studying the distribution of characters on the insect cladograms in lieu of the detailed analyses. An interactive component can be added by having students move particular branches and observe the effect on various statistics. This approach allows key evolutionary concepts, such as monophyly, homology and convergence, to emerge intuitively, thereby preparing students for a more analytical presentation of these ideas in advanced courses.

MacClade can even be used profitably at the pre-university level. A simplified experience with the tutorial alone provides sufficient material for a lesson in evolution. Students can evaluate the effect of branch movements on the vertebrate cladogram's robustness by focusing on changes in an intuitively straightforward statistic, such as treelength. The fact that students' results may often be at odds with what they expected can generate very valuable discussion. For an even simpler approach, teacher and class can work together to construct a basic character matrix for a few common organisms (e.g. bird, lizard, butterfly). Students can suggest characters about natural history, behaviour, etc., based on their personal experiences with these animals. By using the Trace function, they can then rearrange cladogram branches so as to "pair like with like", which amounts to a very simple introduction to the idea that cladograms attempt to minimise convergent characters. Elementary school students exposed even to these very simple activities will be much better prepared to meet the challenges of evolutionary study at the university level.

### Conclusions and educational implications

Although evolutionary theory is the centrepiece of modern biology, effective presentation of its methodologies can pose a special challenge to the life science teacher. The MacClade software package offers an array of solutions that is only limited by the instructor's imagination — the exercise presented here is merely one possible approach. Use of MacClade can help restore evolution to the centre stage it should occupy in Biology curricula.

### Acknowledgements

I thank J Franko, F Glazer, R Kliman, A Porta, K Raffa and X Yu for thoughtful comments on earlier drafts of this paper. I am also grateful to the many undergraduates at Northland College (Ashland,

Wisconsin, USA) and Kean University (Union, New Jersey, USA) who have offered suggestions about this exercise over the years, especially J Franko, E Rhynes, G Struble and K Tischler. Completion of this paper was supported in part by an Untenured Faculty Research Initiative award from Kean University.

### References

- Alcock J (2001) *Animal behavior: an evolutionary approach*. 7th edition. Sunderland, MA, USA: Sinauer Associates.
- Codella S G (2000) *Zoology: pattern and process*. Dubuque, IA, USA: Kendall-Hunt.
- Harvey P H and Pagel M D (1991) *The comparative method in evolutionary biology*. Oxford, UK: Oxford University Press.
- Hickman C P, Roberts L S and Larson A (2001) *Integrated principles of zoology*. 11th edition. New York, USA: McGraw-Hill.
- Kitching I J, Forey P L, Humphries C J and Williams D M (1998) *Cladistics: the theory and practice of parsimony analysis*, 2nd edition. (Systematics Association Publication No. 11). Oxford, UK: Oxford University Press.
- Krebs J R and Davies N B (1993) *An introduction to behavioural ecology*, 3rd edition. Cambridge, MA, USA: Blackwell Scientific.
- Kristensen N P (1981) Phylogeny of insect orders. *Annual Review of Entomology*, **26**, 135 – 157.
- Maddison W P (1990) A method for testing the correlated evolution of two binary characters: Are gains or losses concentrated on certain branches of a phylogenetic tree? *Evolution*, **44**, 539 – 557.
- Maddison W P and Maddison D R (2000) *MacClade: analysis of phylogeny and character evolution*. Version 4. Sunderland, MA, USA: Sinauer Associates.
- Marins E P (Ed.) (1996) *Phylogenies and the comparative method in animal behavior*. New York, USA: Oxford University Press.
- Ridley M (1990) The control and frequency of mating in insects. *Functional Ecology*, **4**, 75 – 84.
- Wiley E O, Siegel-Causey D, Brooks D R and Funk V A (1991) The complete cladist: a primer of phylogenetic procedures. *University of Kansas Museum of Natural History Special Publications*, No. 19.
- Zar J H (1996) *Biostatistical analysis*. 3rd edition. Englewood Cliffs, NJ, USA: Prentice-Hall.

### Appendix

#### MacClade 4

Sinauer Associates, Inc., 23 Plumtree Road, Sunderland, MA 01375 USA, Tel: 1-413-549-4300. [www.sinauer.com](http://www.sinauer.com), accessed 28 February 2002.

MacClade Home Page, <http://phylogeny.arizona.edu/macclade/macclade.html>, accessed 28 February 2002.

#### Executor 2.1 (MacOs emulator software)

Abacus Research and Development, Suite 4-101, 1650 University Blvd., NE, Albuquerque, NM 87102 USA, Tel: 1-505-766-9115. [www.ardi.com](http://www.ardi.com), accessed 28 February 2002.

*Professor Sylvio Codella works in the Department of Biological Sciences, Kean University, 1000 Morris Avenue, Union, NJ 07083 USA. Tel. (908) 527-2464; Fax. (908) 659-1976; Email: [scodella@kean.edu](mailto:scodella@kean.edu)*

A vertical bar on the left side of the page, consisting of a series of horizontal segments in shades of yellow and orange, with a small red diamond at the top.

COPYRIGHT INFORMATION

TITLE: Testing evolutionary hypotheses in the classroom with  
MacClade software

SOURCE: Journal of Biological Education 36 no2 Spr 2002

WN: 0210502544008

The magazine publisher is the copyright holder of this article and it is reproduced with permission. Further reproduction of this article in violation of the copyright is prohibited. To contact the publisher:  
<http://www.iob.org/>.

Copyright 1982-2002 The H.W. Wilson Company. All rights reserved.