

Chapter 14

Comparative analysis of sex ratio data

Peter Mayhew & Ido Pen

In: *The Sex Ratio Handbook* (I. Hardy, ed). Cambridge University Press. In press.

14.1 Summary

Comparative studies are those which use the characteristics of different taxa as a source of data, and such studies have made important contributions towards sex ratio theory. Comparative data require special methods for statistical analysis because not all the variance in taxon characteristics is evolutionary independent. Solving that problem requires explicit phylogenetic and evolutionary assumptions which create challenges at each stage of a comparative study. Here we review the essentials of a comparative approach: collecting data; choosing methods; analyzing data; and making conclusions. We include a worked example of a recent sex ratio study; that of New World non-pollinating fig wasps by West & Herre, which we analyze by independent contrasts and by simulation methods. Finally we review the relevant software for comparative analysis of sex ratios and how to obtain it.

14.2 Introduction

I find that in Great Britain there are 32 indigenous trees[:] of these 19 or more than half...have their sexes separated, - an enormous proportion compared with the remainder of the British flora: nor is this wholly owing to a chance coincidence in some one family having many trees and having a tendency to separated sexes: for the 32 trees belong to nine families and the trees with separate sexes to five families.

—Charles Darwin, from Staufer (1975).

In recent years an abundance of reviews have dealt with the theoretical problems of conducting a comparative study (Ridley, 1983; Burt, 1989; Ridley, 1989; Brooks & McLennan, 1991; Harvey & Pagel, 1991; Harvey & Purvis, 1991; Grafen & Ridley, 1996; Martins & Hansen, 1997; Price, 1997). Many biologists are today, like Darwin was, aware of the problem of non-independence of species characters and how a knowledge of phylogeny and evolutionary processes can in principle help to overcome it, as well as make full use of the data (Harvey, 1996). For researchers, this awareness does not in itself solve the problem for two reasons. First, the detail of how best to use phylogeny is still a matter of hot debate (see Harvey & Nee, 1997), and this makes comparative methods a potential minefield for the uninitiated. The uninitiated need to know succinctly what the different analysis options and their implications are. Second, for the empiricist the major problem remains how to get hands on experience with comparative methods. This chapter differs from all the above reviews, in trying to provide the practical as well as theoretical information sufficient to conduct a good comparative analysis in general, and a good comparative analysis of sex ratios in particular.

In the following sections we take the reader step by step through the stages which will most commonly characterize a comparative study of sex ratios. At each step we discuss whether and how comparative techniques can be useful, and the problems and challenges associated with them. At the end is a section containing information on the most relevant software, but we encourage readers to follow the intermediate sections too. Possession of software can be very useful, but to do good science you need to apply and interpret the software appropriately. If we have a single message it is that a good comparative study addresses that challenge, whereas a poor comparative study ignores it. If this chapter does nothing else, we hope it teaches the reader that care in choosing and interpreting statistics applies here as for any other empirical study.

14.3 Before starting

14.3.1 Why do a comparative analysis?

For the purposes of this chapter, comparative studies are defined as those which use the variation between different taxa as a source of data to help answer a biological question. Comparative studies are one of a number of empirical approaches to answering a question, and we distinguish studies in which the data consist of the properties of individuals within a species (observational, experimental, and “two-species comparative studies”, see T. *et al.*, 1994; Price, 1997), and those in which data consist of the results of a number of separate studies addressing an identical question in one or more species (meta-analyses, Arnqvist & Wooster, 1995).

Many readers will already be set on a comparative approach. Hypotheses may grow from informal comparative observations which then beg more formal investigation: “where the experimental biologist predicts the outcome of experiments, the evolutionary biologist retrodicts the experiment already performed by Nature; he teases science out of history” (Wilson, 1994, p.167). In other cases, scientists may begin with a question and then ask if a comparative approach can help answer it. We see four reasons for choosing the comparative approach. First, in addressing questions across different taxa, comparative analyses are likely to produce results of interest to more scientists than studies on a single species. Second, they do not suffer from the problem of extrapolating results from a single species to other species, which so often characterizes studies on individual “model organisms”: as soon as the problem of across taxon variation is addressed with real data, one is no longer extrapolating, one is doing a comparative study. Third, different taxa often show character variation which is difficult to obtain by experimental or other means: natural selection has performed manipulations and worked over timespans which are less possible for experimenters. As a result, and fourth, cross-species variation is often much larger than within species variation. In summary, variation across taxa is often large, easily obtained, widely interesting, and widely applicable: four good reasons for choosing a comparative approach to answering a question.

Although a comparative approach can be very illuminating (Harvey, 1996), it rarely tells a complete story. Comparative studies can be excellent ways to demonstrate correlation, but are much less adept at showing causation (although see Richman & Price, 1992). For the latter, experiments may provide an answer. We also stress that the answer to any question may depend on the taxonomic level within which the question is phrased (e.g. Mayhew & Hardy, 1998), so may depend on whether a comparative or other approach is taken. In sum, comparative and other kinds of analyses are not always strict alternatives, and can be complementary.

What specific comparative questions do sex ratio researchers ask? By

Table 14.1: Some comparative studies of the sex ratio. Abbreviations: CXA cross X analysis [X= populations (P), species (S), genera (G), family (F)], IC independent contrasts, GA graphical analysis (no statistics), MCCT Maddison's concentrated changes test.

Reference	Taxa studied	Methods
Dijkstra <i>et al.</i> (1998)	7 bird species	CSA
Fellowes <i>et al.</i> (2000)	44 species of Old World non-pollinating fig wasps	CSA, IC
Gowaty (1993)	40 bird species	CSA, CFA
Griffiths & Godfray (1988)	23 species of benthylid wasps	CSA
Hardy & Mayhew (1998)	19 species benthylid wasps	CSA, IC
Hardy <i>et al.</i> (1998)	16 populations/species parasitoid Hymenoptera	CSA, IC
Herre (1987)	13 fig wasp species	CSA
Herre <i>et al.</i> (1997)	22 fig wasp species	CSA
Johnson (1988)	15 primate genera	CGA
Mitani <i>et al.</i> (1996)	18 primate species	IC
Poulin (1997a)	193 populations/species nematodes/acanthocephalans	CPA, CSA IC
Poulin (1997b)	76 populations nematodes	CPA, IC
Read <i>et al.</i> (1995)	Protozoan bird parasites	GA
Schutler <i>et al.</i> (1995)	Haemoproteus from birds	CPA
Waage (1982)	31 species of scelionid wasps	CSA
Weatherhead & Montgomerie (1995)	16 bird species	MCCT
West & Herre (1998a)	17 fig wasp species	CSA, IC
West & Herre (1998b)	16 fig wasp species	CSA, IC
West <i>et al.</i> (1997)	55 fig wasp species	CSA, IC

questioning the authorship of this volume, we compiled 19 published comparative sex ratio studies which consider taxon characteristics as data (Tab. 14.1). Comparative studies have been used to investigate the causes and effects of sex ratio evolution in a wide range of taxa; from nematodes and protozoa to birds and primates. Studies on Hymenoptera are particularly common. The Hymenoptera are species rich, often practical to study, and display particularly diverse sex ratios; they are ideal meat for comparative study. Studies on plants are notably absent from our list, though not from want of searching. Only a small proportion of plant species are dioecious and can therefore display actual sex ratios, but perhaps this review will stimulate some comparative work on them. If we had expanded our search to include sex allocation in monoecious taxa then plants would be better represented, not least by inclusion of Darwin's observations quoted at the

start of the chapter.

Within the taxa represented, comparative studies have provided some of the most important data contributing to sex ratio theory: questions about the effects of mating structure are very common, and the comparative data confirm the importance of local mate competition as an important factor in the evolution of female biased sex ratios (e.g. Fellowes *et al.*, 2000; Griffiths & Godfray, 1988; Hardy & Mayhew, 1998; Poulin, 1997a; Read *et al.*, 1995; Waage, 1982; West & Herre, 1998a, although see Schutler *et al.*, 1995). Three studies address the influence of local resource competition (Gowaty, 1993; Johnson, 1988; Weatherhead & Montgomerie, 1995). Two of them (Gowaty, 1993; Weatherhead & Montgomerie, 1995), study the same data set of brood sex ratios in birds but reach opposite conclusions due to the use of different comparative methods. A number of studies address inbreeding as an agent of sex ratio bias (Herre, 1987; Herre *et al.*, 1997; Schutler *et al.*, 1995), and one the influence of virginity under local mating (West *et al.*, 1997). Two studies address variation in sex ratio variances rather than just average sex ratios (Hardy *et al.*, 1998; West & Herre, 1998b). Comparative data relating to the effects of intensity of selection (e.g. Herre, 1987; West & Herre, 1998b) are especially valuable because such data are much more difficult to obtain by other means. One study (Dijkstra *et al.*, 1998) tests the causes of sex ratio bias in sexually dimorphic birds. Finally, sex ratio has sometimes been included in comparative studies as an explanatory variable, for example as a proxy for the intensity of sexual selection (Mitani *et al.*, 1996; Poulin, 1997b). Clearly, although comparative studies have so far made important contributions to sex ratio theory, there is also scope to widen the theoretical issues and which might be addressed with such techniques, and we hope this chapter will stimulate that development.

14.3.2 Comparative analysis and phylogeny

Comparative analyses, like other statistical analyses, face two important challenges: to not reject a null hypothesis if it is true and to reject it if it is false. Failure to meet these challenges is called Type I and Type II error, respectively. If a test avoids Type I error it has validity, and if it avoids Type II error it has power. Validity of statistical tests is what makes science different from random speculation: it stops us making the wrong turn at every step. Power is obviously important if we are to conclude all we might be concluding about the world, but if we never conclude anything, we are also not making wrong conclusions. So science gives validity priority over power: if a valid test can also be powerful, so much the better. The power and validity of any test can be demonstrated by using the test to analyze dummy data sets in which the true evolutionary relationships between characters are controlled, and observing how often they reject the

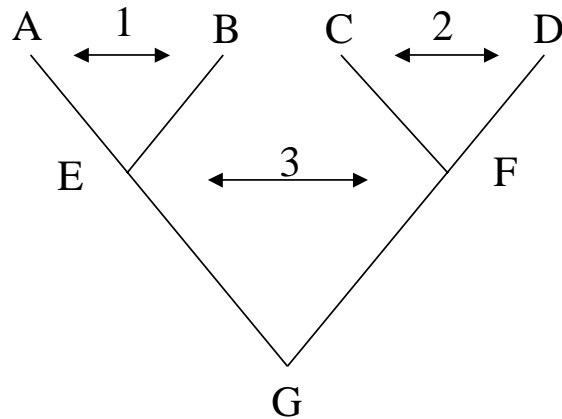


Figure 14.1: A schematic representation of some comparative methods. Using just raw species data (A,B,C, D) gives four data points but may be invalid due to the non-independence of species. Comparing extant sister taxa allows us to partition out variation in A and B that is independent of E (comparison 1) and C and D that is independent of F (comparison 2), but only leaves the two data points instead of four. Comparing ancestors also allows us to partition out the variation in E and F that is independent of G (comparison 3) if we are prepared to assume an evolutionary model to estimate E and F. That same evolutionary model also allows us to simulate null species data and hence use the raw species data (four points).

null hypothesis under consideration (e.g. Grafen, 1989; Purvis *et al.*, 1994; Grafen & Ridley, 1996).

The major obstacle to finding valid comparative tests arises from the non-independence of related taxa. Closely related species are likely to be similar because they will inherit characters from a shared common ancestor. This means that not all the variance in species characters is the result of independent evolutionary experiments: species can only have been independent of other species in the dataset since the last time they shared a common ancestor with one of them (Felsenstein, 1985). If, in addition, related taxa share important variables causing evolution to proceed in the same way on related parts of the evolutionary tree (so called third variables, Ridley, 1989), taxa can only be treated as evolutionarily independent if we know what the third variables are and can control for them statistically (Price, 1997). If we treat species as fully independent, for example by

taking raw species characters as our data for analysis, we may overestimate the amount of evolutionarily independent variation, and will artificially inflate the significance of any test used.

A valid statistical test can proceed in two ways: the first, and most common approach is to try to partition out the variation in the raw species data set that is evolutionarily independent, and use that for analysis instead (Fig. 14.1). The second possible approach is to continue using the raw species data but adjust the significance value of the test statistic to account for the non-independence of species (Fig. 14.1). To do both such tests we need to know something about the phylogeny and evolutionary processes that link the taxa in our data set together: "Tell us how to reconstruct the past, and we shall perform the comparative analysis with precision." (Harvey & Pagel, 1991).

Although validity has been the major motivation for incorporating phylogenetic knowledge into comparative tests, there are other benefits and costs involved. Another benefit is the satisfaction of making full use of the data (Harvey, 1996). Taking a phylogenetic perspective can allow one to ask whether evolutionary correlations apply equal well across different parts of the tree and that can allow one to say how the extant character states in your original data set have arisen (Mayhew & Hardy, 1998). Furthermore, statistics may be used not just to estimate significance but also coefficients of association and regression slopes. Again, if the assumed model of evolution is correct, phylogenetic techniques allow more accurate estimates of such parameters than cross-species comparisons (Harvey & Pagel, 1991). In particular, if correlated evolution varies slightly between different taxa, slopes calculated from cross-species analyses might be biased according to the relative representation of different taxa in the data set.

We identify two costs to phylogenetic awareness: time to do the test, and power if the phylogeny is not fully resolved. The time taken to perform a phylogenetically based test is often very small, because software is available to take the strain. Time may be required to familiarize oneself with each application, and to assemble phylogenetic knowledge on the group in question, but that is a worthwhile investment.

Although the power lost in a phylogenetically aware test might be great, it need not always be. If phylogeny and evolutionary processes are well known, more independent variation can be partitioned out, and very little power is lost. Greater power can be gained by comparing ancestors as well as extant taxa, or by simulation methods (Fig. 14.1), but both require a knowledge of internal phylogeny, and a model of evolution on which to simulate null data or to reconstruct ancestral states. If knowledge of those processes is not good or the phylogeny estimate is incorrect, the assumptions of the test might be inappropriate and introduce more Type I error (Price, 1997). A consensus on what approach to take when the model of

evolution is unknown (as it is in most cases) has yet to emerge from the academic community. One interim solution is to perform several analyses using a range of different assumptions, and observe if the results are robust, or to use methods, such as pairwise comparisons, which are themselves robust to a range of different evolutionary assumptions. Carrying out both cross-species analysis and analysis based on phylogeny can often be very informative (Price, 1997), and many of the studies in Tab. 14.1. do this. Unlike some authors (e.g. Ricklefs, 1998) we do not advocate abandoning phylogeny if evolutionary processes are poorly known. Our reason is that cross-species analyses assume that all species are equally related, which is probably wrong (Felsenstein, 1985). Nor do we advocate abandoning phylogeny when characters are thought to be phylogenetically labile (e.g. Westoby *et al.*, 1995). Our reason is that this is tantamount to making up the data: even if a species might have evolved to the same extant state from any previous ancestral state, in reality it attained its present state by diverging from an ancestor shared with its nearest relative in the data set, and the variance due to the latter amount of evolution is the variance we ought to be using in our tests. Cross-species analyses might sometimes produce the same results as those based on phylogeny (Ricklefs & Starck, 1996) but use incorrect assumptions about relatedness. That would be using the end to justify the means. It is possible to test for phylogenetic lability (e.g. Björklund, 1997) but that requires in itself a knowledge of phylogeny and evolutionary processes.

If phylogeny is poorly known, it may still be possible to explore the effect of different phylogeny estimates. For pairs of continuous characters, software is available to test the effect of different phylogenetic assumptions (Martins, 1996). Pairwise comparisons on extant sister taxa (see Moller & Birkhead, 1992, Fig. 14.1) can also be performed for both continuous and categorical variables and require very little knowledge of the internal phylogeny of a phylogenetic tree though they are less powerful than other tests.

We argue that the benefits of phylogenetic awareness - greater validity and making full use of the data - will normally outweigh the costs of time and power lost. But in extreme cases the costs may outweigh the benefits. What should one do then? One might perform a cross species analysis as a first step and point out the possible pitfalls (Herre *et al.*, 1997). Alternatively one can just ignore statistics and plot the data graphically (Read *et al.*, 1995). Such data can still be extremely valuable: they can point the way for further study when phylogenetic awareness is possible (see Hardy & Mayhew, 1998). Even if you have decided on a phylogenetically based test, it is a good idea to explore the cross-species data too. The former tells you about evolutionary correlations, but the latter tells you about trends in extant characters which are the product of those evolutionary correlations. Most biologists are interested in pattern as well as process. In short,

although phylogenetic awareness usually adds much at little cost, cross species comparisons have their place too if used carefully.

14.4 Collecting comparative data

14.4.1 Taxon characteristics

Data from comparative analyses generally come from two sources: direct collection by the same researchers using standardized methods (e.g. West & Herre, 1998a), or collection from several studies previously published by different authors (e.g. Griffiths & Godfray, 1988). Because data collection on several species tends to be time consuming and costly, the latter is more common. This leads to several challenges: one is that error, often quite substantial, is introduced into the estimates of species characteristics by different methods, and typographical errors. How should this error be dealt with? Second, there are several possible ways of summing up species properties with a single number. Which should be analyzed? Third, more than one independent estimate of species characters may exist. How should these be treated? Finally, how far should we go in our data collection, and which taxa should we be targeting?

One possible solution to the problem of measurement error is to set criteria for including a study in the analysis to try to reduce the noise in the data. Purvis & Harvey's comparative test of Charnov's mammal life-history model (Purvis & Harvey, 1995) illustrates how this might be done. Purvis & Harvey assembled a data set of 80 mammalian life tables after rejecting studies conforming to seven criteria, including those based on captive populations and those based on small samples (<30 individuals). Furthermore all data were examined to check for wholly exceptional values which might represent conversion or typographical errors, or data derived by chance from the tails of the real biological distributions.

Rejecting large quantities of data is a luxury affordable only with intensively studied taxa. More often studies will be forced by paucity of data to include most published information. Examining trends at different taxonomic levels can then suggest whether there was a large amount of measurement error in the species level data. When there is, relationships will probably hold better at deeper levels of the phylogeny, because the measurement error is eliminated by the averaging used to estimate ancestral states.

The problem of summarizing species properties with a single statistic is common to most comparative studies. Usually some kind of average is used for continuous variables; commonly the arithmetic mean, the median, or the mid-point of the range. What can be used depends on what is reported in the literature, but if raw data exist, it may pay to examine them

graphically to generate the best possible estimate. If the data are right skewed, as is common for ecological or life history characters, the median, or mean of the logged data, may be more appropriate than the arithmetic mean. If possible, it is worth trying to standardize the measure used within each study to reduce the error variance in the data.

What to do with independently collected estimates of variables depends on the phylogenetic independence of each study. If studies of distinct populations allow all the relevant variables to be estimated within each population, these can be included as separate taxa within the analysis. Moller & Birkhead (1992) did this for populations of barn swallows which nest solitarily or communally. More commonly it will not be clear if estimates are drawn from distinct populations, and not all relevant variables may be collected from a single study. Such cases may be treated as independent estimates of the same population and an average value calculated from both as above.

Taxon characteristics may require transformation prior to analysis. If sex ratio is the response variable and the data are assumed to be fully continuous, arcsine square root transformation is appropriate. Other continuous data are often right skewed and may need to be log transformed. A given change in a variable usually carries higher biological significance if it occurs when the absolute values are small than if they are large. Log transformation has the effect of transforming absolute into relative differences, so tell us more meaningful biology. Some tests, such as independent contrasts, assume a Brownian motion model of evolution, and logged data are more likely to conform to this model (Garland *et al.*, 1992).

How far should one go with data collection? One consideration is that more data can add power. The extent to which new data add power depends both on the intended method and the phylogeny used. Limitations may be reached well before then because of paucity of relevant studies. If there is ample data in the literature, one might profit from both collecting more estimates of species already covered, to reduce the error variance in the data set prior to analysis, and to include new taxa which bring to the data set novel information about correlated evolution. A datum about a different genus may be worth many data on genera already covered if it allows more evolutionarily independent variation to be partitioned out.

What taxa may reasonably be incorporated into a comparative analysis? Taxa must be sufficiently closely related, or have sufficiently similar biologies, for a single question to be relevant to them all, but sufficiently distantly related for evolution to have produced variation of interest. Sometimes this may include distant relatives, and sometimes populations of the same species. One situation to be wary of is when many of the included taxa are very close relatives, such as species in the same genus, but a few are more distantly related, such as species from different families (e.g. Hardy

et al., 1998). In such cases it is worth examining in detail whether comparisons among or between different families give very different results. If the analysis makes assumptions about the branch lengths of the phylogeny, checking the validity of and modifying the assumed branch lengths can be particularly illuminating (Garland *et al.*, 1992).

A final worthwhile step is to make the full dataset available via a published table, appendix, or journal website. Comparative datasets can take much effort to compile. Future researchers will be grateful not to have to start from scratch.

14.4.2 Phylogenies

In most cases, comparative analyses will require a two-pronged approach to data collection: collection of taxon character estimates, and collection of phylogenetic information. Well defined species level consensus trees exist for some groups such as primates (Purvis, 1995). More often, different species appear on different cladograms, some on several, and some on none. Then, estimates of phylogeny must be assembled according to the species in each study. Even where no cladograms are available, taxonomy may still allow some best guess at phylogenetic relationships (e.g. Mayhew & Blackburn, 2000), although one should be aware that taxonomies are sometimes not derived from cladistic principles and hence may not even attempt to estimate phylogeny. But only rarely can one afford the luxury of ignoring taxonomy. In exceptional cases, such as when species in the same genus are studied, there may be both no cladistic or taxonomic estimate available. In that case one might consider investigating the effect of different possible phylogenies, or restricting oneself to a cross-species analysis, with all its associated pitfalls.

Where different cladograms or taxonomies are available, one has several options. First, one might consider computing a composite tree, or "supertree" (Sanderson *et al.*, 1998). This can be done in two ways: reanalyzing the combined raw data sets of the different studies (which is impossible if taxonomies are to be included or if the combined data set is very large), or coding the trees themselves as data and analyzing them to find a consensus. A good example of the latter is Purvis' composite estimate of primate phylogeny (Purvis, 1995), which presents an estimate for all 203 extant primate species, resolved into 160 nodes of which 90 are dated, and derived from 112 source publications. This composite tree was calculated by recoding the nodes on the published trees as binary characters and then combining the data sets for different trees, appropriately weighted according to the quality of each study. The combined data were then analyzed by standard parsimony algorithms.

Such studies, though laudible, are time consuming and require famil-

ilarity with techniques for phylogeny reconstruction, so short cuts may be preferred. Luckily, in many cases trees can be combined by eye because there is little discrepancy between them, or little overlap in the taxa studied. A large drawback is that such estimates are not explicit about their assumptions, so different researchers may come to different solutions. Nonetheless, if this is the preferred option, we suggest at least consulting an expert taxonomist on the relevant group before attempting analysis. It is extremely easy to make big mistakes when combining trees by eye if one is not very familiar with the groups concerned.

An alternative, if two or more relatively complete but conflicting estimates are available, is to do a separate analysis on each estimate (e.g. Seehausen *et al.*, 2000). Discrepancies in the analysis outcome then may allow us to pinpoint the most important parts of the tree for further taxonomic work, as well how robust the outcome is under present phylogenetic uncertainties.

Whatever the final approach, the phylogeny used should be made available so that researchers can repeat the analysis if necessary, and the methods used to construct it made explicit, so that readers can judge its worth. No single estimate of phylogeny will ever be the last word, so facing up to its faults displays good scientific principles.

14.5 Choosing comparative methods

The analysis method chosen will depend on the way the data are coded, the computer system available, the extent of phylogenetic knowledge, and the evolutionary assumptions which one is prepared to make. This section is a quick guide to those requirements for the most common methods.

14.5.1 Cross-species comparisons

Because they may give biased or invalid estimates of the degree of correlated evolution, cross-species comparisons must be used with extreme care, or indeed “kept at the other end of a barge pole” (Ridley & Grafen, 1996). Cross-species comparisons are still the most common comparative method used (15 of the 19 sex ratio studies in Tab. 14.1). They do not require phylogenetic information, but that is exactly why they are often invalid. They can provide a useful way of illustrating trends in extant character states, but to conclude anything about correlated evolution is risky. We recommend that comparative researchers at the very least explore cross species trends graphically or with contingency tables. Both categorical and continuous variables and any combination of the two can be analyzed easily. If sex ratio is the response variable, logistic regression is to be recommended, or if not, then transformation (arcsine square root) of sex ratio prior to using

other parametric tests. Most statistics software packages will allow this, so one is not hardware restricted.

14.5.2 Independent contrasts

The most popular phylogenetically based method, used in 9 of the 19 studies in Tab. 14.1, is that of independent contrasts. There are several possible implementations, based on different assumptions. The method is due to Felsenstein (1985), who noted that differences between two daughter lineages across a node (contrasts) represent variation acquired since those taxa last shared a common ancestor, and so are phylogenetically independent in the absence of confounding third variables, which they may control for to some extent (Price, 1997). Contrasts can be calculated between ancestral taxa as well as extant taxa if a given model of evolution is assumed which allows those ancestral states to be calculated. Most commonly, Brownian motion is the assumed evolutionary model, although how suitable such a model is may, and perhaps should, be questioned. Contrasts may be analyzed parametrically by linear regression through the origin or non-parametrically by a sign test (Garland *et al.*, 1992). The original implementation is in Felsenstein's PHYLIP software package (for MS-DOS or Apple-Macintosh), but the phylogeny used must be fully resolved.

Two subsequent implementations allow analysis even if parts of the phylogeny are uncertain by making contrasts across soft polytomies (nodes representing uncertain relationships of more than two sister taxa). The most popular of these is CAIC (Purvis & Rambaut, 1995) for Apple-Macintosh, and outputs the contrasts which must then be analysed elsewhere. It does allow analysis of categorical characters if they are ordered explanatory variables, but otherwise variables must be coded as continuous. This means that bounded variables, such as sex ratios should be transformed prior to analysis, usually arcsine square root. In addition, it makes sense to log transform other continuous variables prior to analysis to conform to the Brownian motion model of evolution (Garland *et al.*, 1992).

Phylogenetic Regression (Grafen, 1989) is a macro run in the GLIM statistical package (hence is MS-DOS or UNIX compatible), and there are different versions corresponding to the different versions of GLIM. The output is perhaps less useful than CAIC - an F statistic, P value, and slope of the relationship. The raw contrasts are not part of the output, and neither are trustable values of R -squared. Phylogenetic Regression has been used to analyze categorical as well as continuous characters, both ordered and unordered, both as dependent and independent variables, and can simultaneously control for several independent variables, so is highly flexible (see Mayhew, 1998; Mayhew & Blackburn, 2000). Its major drawbacks are the output content and format, and the fact that it is user-unfriendly.

If the phylogeny is completely unknown, Martin's (1996) implementation of Felsenstein's method in COMPARE (for MS-DOS or UNIX) will allow contrasts to be calculated over a range of phylogenetic hypotheses to see if significant relationships are possible or likely.

14.5.3 Other methods for continuous data

A popular method which, to our knowledge, has yet to be applied to sex ratio studies is Phylogenetic Autocorrelation (Cheverud *et al.*, 1985; Gittleman & Kot, 1990). It can be implemented using the COMPARE software for PC, and a version for Apple-Macintosh is available elsewhere. Like independent contrasts, Phylogenetic Autocorrelation is designed for use with continuous data, and works by partitioning the data for each species into phylogenetic and specific components, then using the specific components to test for evolutionary correlations (Gittleman & Luh, 1992). Phylogenetic Autocorrelation can provide interesting comparative data on evolutionary lability (see Gittleman *et al.*, 1996). It is not always as valid as independent contrasts, although it can be more powerful (Purvis *et al.*, 1994).

If internal phylogeny is poorly known, it is often still possible to carry out phylogenetically aware tests by comparing pairs of sister taxa, so called pairwise comparisons (Moller & Birkhead, 1992). This might be preferred because conclusions are robust to many different models of evolution (Harvey & Nee, 1997), although they lack power precisely because they ignore internal phylogeny. No special software is required and simple non-parametric tests are often used.

14.5.4 Methods for categorical data

Although most comparative analyses of sex ratios will use some continuous measure of sex ratios as the response variable, it is possible that sex ratio could be coded categorically, such as male-biased versus female-biased (e.g. Weatherhead & Montgomerie, 1995), or that sex ratio is used as one of the explanatory variables in an analysis of other categorical variables such as habitat. For mixtures of categorical and continuous variables, some of the above methods, such as Phylogenetic Regression, are still useful, but an additional range of tests are open which especially consider correlated evolution of one categorical, usually binary, character on another (Ridley, 1983; Burt, 1989; Maddison, 1990; Harvey & Pagel, 1991; Moller & Birkhead, 1992; Pagel, 1994; Grafen & Ridley, 1996). Software is available to perform some of these (e.g. Maddison & Maddison, 1992), but if the data set is small such tests can be done by hand. One software package, MacClade, allows one to map the evolution of categorical characters across a tree and is extremely useful for identifying appropriate comparisons between variables.

Tests for correlated evolution of categorical characters are more problematic than those for continuous characters because change in categorical characters will tend to be rarer than change in continuous characters, leaving large areas of the tree in which no variation occurs. Different tests tend to treat these areas of the tree in different ways, and there is much controversy about how best to do it (see Harvey & Pagel, 1991; Ridley & Grafen, 1996).

14.5.5 Simulation methods

Instead of using analytical methods to partition the variance in the data into dependent and independent components, it is also possible to use phylogenetic information in computer (Monte Carlo) simulations to adjust the significance-levels of cross-species analyses (Garland, 1993). The simulation approach proceeds in two steps. First, a large number (typically 1000) of dummy data sets is generated by simulating evolution of characters across a tree using a null model of uncorrelated evolution. Second, a cross-species analysis is performed on each of the dummy data sets, yielding distributions of the relevant statistics. By comparing the statistics (e.g. *F* values) obtained in the cross-species analysis of the real data with the distributions obtained from simulations, one can judge the likelihood of the null-hypothesis being true.

The first step, simulation of character evolution, requires assumptions about a model of character evolution and an estimate of the ancestral character value at the root of the phylogenetic tree. Whereas analytical methods such as independent contrasts are nearly always based on the Brownian motion model of evolution, simulation methods allow much greater flexibility in the choice of an evolutionary model. For (arcsine square root transformed) sex ratio data, the assumption of Brownian motion may sometimes be a poor one, especially for taxa with chromosomal sex determination (e.g. mammals, birds), where it is likely there is some evolutionary "pull" towards a sex ratio of 0.5. In such a case, the Ornstein-Uhlenbeck process (Felsenstein, 1988) might be a more appropriate choice. In an Ornstein-Uhlenbeck process the expected change of a character is toward a fixed point (e.g. 0.5), at a rate proportional to the character's distance from this point.

Ancestral states in simulations can either be guessed (e.g. 0.5 for bird or mammal sex ratios), or estimated (e.g. with independent contrasts if Brownian motion is the assumed evolutionary model). By using a range of different ancestral states, the robustness of this assumption can be checked. Another advantage of the simulation approach is that unresolved nodes (at least three taxa originating from the same node) do not lead to a loss of degrees of freedom, as is the case in most analytical methods. Unresolved

nodes are simply kept as the best available information and it is assumed that at least three taxa have originated simultaneously from the same ancestor.

The second step in the simulation approach, a cross-species analysis of the dummy data sets (e.g. by means of logistic regression), offers the possibility to take account of the difference in sample size between studies. In this way, measurement error due to small sample sizes can be controlled for. This is particularly useful for sex ratio studies, because average sex ratios based on only a few individuals are more likely to be biased and may hide real biological trends in the data. In the simulated data sets, each species can be assigned the same sample size as in the real data set to observe if a biased sex ratio is to be expected by chance or not. Later on we give an example of how this can be done.

Moderately experienced users of spreadsheets with random-number generators should be able to implement their own simulations. Several of the computer packages listed below offer the possibility of generating simulated datasets based on a number of different evolutionary models (step 1). The output of these programs can be analyzed (step 2) by means of standard statistical software. One of us (I.P.) has produced a user-unfriendly program which performs both steps (simulation and logistic regression), available upon request in addition to assistance in operating the software.

14.6 Analyzing comparative data

Once you have your question, data set, phylogeny, and method, it is time to analyze. In this section we present a worked example to illustrate how that might be done. We have chosen the study of sex ratios in New World non-pollinating fig wasps by West & Herre (1998a) because the data are amenable to different forms of analysis, and illustrate well some of the challenges and problems involved.

The data consist of sex ratios of seventeen species of non-pollinating fig wasp collected by the authors from figs in Panama, together with the proportion of host fruit collected which contained each wasp species, and whether the wasp species has males which are winged or unwinged. The data are used to test two related hypotheses. First, that foundress number is positively correlated with proportion males. This is a prediction of Hamilton's theory of local mate competition (Hamilton, 1967). Second, that as the proportion of non-local mating increases, proportion males increases. That prediction is made by several models which extended Hamilton's theory (see Hardy, 1994). Both foundress number and the degree of non-local mating are measured by proxy variables because of the difficulty of observing them accurately in the field: non-pollinating fig wasps, unlike pollinating wasps, oviposit from the outside the fruit, so foundress

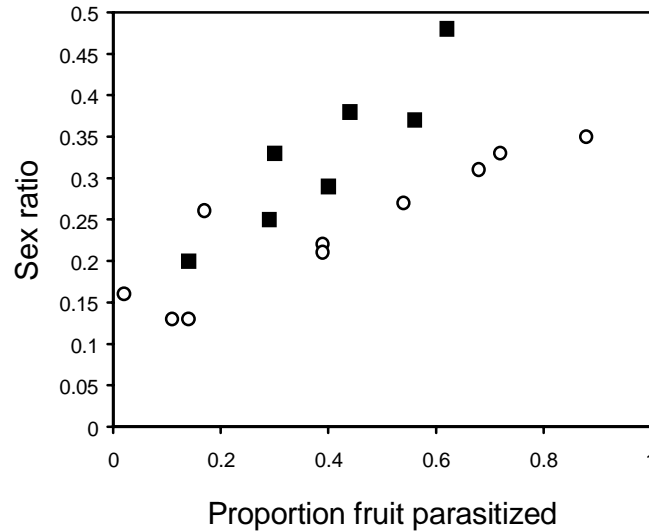


Figure 14.2: The raw species data on proportion fruit parasitized and sex ratio plotted for fig wasps with wingless males (closed squares) and winged males (open circles). Reproduced from West & Herre (1998a) with permission from Birkhäuser Verlag.

numbers cannot be observed directly without watching fruit continuously. Instead, West & Herre (1998a) used the proportion of fruit parasitized as a proxy for foundress number, because both are likely to be related to wasp density. They used a model to show that this relationship held under different assumptions about wasp distributions amongst fruit. The degree of non-local mating is also difficult to observe, but because some species have wingless males which never exit the fig in which they are born, and others are winged and can potentially fly to mate with females from other figs, male wingedness may be used a proxy for the degree of non-local mating.

West and Herre began by using the raw species data in a logistic regression to see if proportion fruit parasitized and male wingedness were correlated with sex ratio among extant taxa. Both relationships were as predicted from theory: species which parasitized a higher proportion of available fruit also had a greater proportion males as predicted, and although the intercept did not depend on male wingedness, the relationship was steeper for those species which had winged males, which therefore had

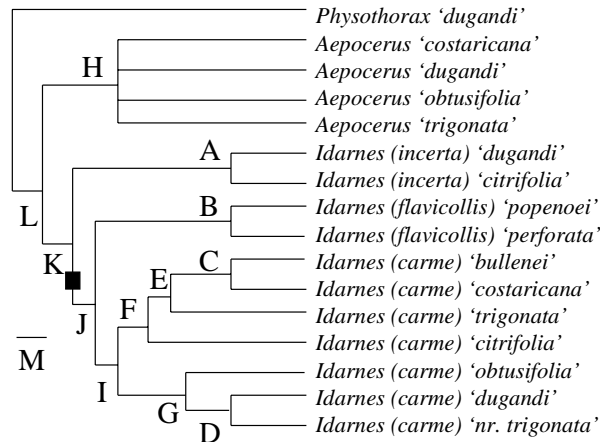


Figure 14.3: An estimate of the evolutionary relationships of the non-pollinating fig wasps in West & Herre's (1998a) study (Machado *et al.*, 1996; C.A. Machado pers. comm.). Provisional species names (in inverted commas) are assigned according to the *Ficus* species infested. Nodes are assigned letters referring to the contrasts in Fig. 14.4. The black square represents the likely moment when male wings were lost since only species in the *Idarnes flavicollis* and *Idarnes carme* species groups have wingless males.

a greater proportion males than unwinged species (Fig. 14.2.). Together proportion fruit parasitized and male wingedness accounted for 94.5% of the sex ratio deviance.

14.6.1 An analysis using independent contrasts

Is the cross-species relationship due to correlated evolution between the proportion fruit parasitized and sex ratio, and between male wingedness and sex ratio? One way to find out is to attempt to partition out the character variance which is evolutionarily independent. West & Herre used Purvis & Rambaut's (1995) implementation of Felsenstein's independent contrasts to achieve this (CAIC). They based their analysis on a preliminary estimate of phylogeny constructed using molecular data (for methods see Machado *et al.*, 1996), which was not fully resolved and led to just nine contrasts from the original 17 species. Subsequent work (C. Machado pers. comm) has resolved some parts of the tree, allowing us to do a more complete

analysis here. The topology is illustrated in Fig. 14.3 Like West & Herre, we use two different branch length assumptions: uniform branch lengths (Pagel, 1992) and assuming that taxon age is proportional to the number of species it contains (Grafen, 1989). The average sex ratio for each species was calculated by summing the total number of males collected and dividing by the total number of individuals of each species. Sex ratio was arcsine square root transformed prior to calculating the contrasts, but proportion fruit was left untransformed. Male wingedness was coded as a factor with possession of wings coded as 1, and winglessness coded as 0.

To begin, contrasts of arcsine square root sex ratio were calculated against proportion fruit parasitized using the CRUNCH algorithm in CAIC. From 17 species in the original data set, 13 independent contrasts can be calculated. We lose two potential contrasts among the *Aepocerus* genus which is unresolved, and one potential contrast because *Idarnes* "columbri-nae" has not been subjected to phylogenetic investigation and must be left out. Examination of the contrasts shows that differences between closely related species and those at the tips of the tree account for the majority of the evolutionarily independent variation; contrasts made deep in the phylogeny all cluster around the origin whereas those at the tips have greater magnitude (Fig. 14.4 a,b). This suggests little phylogenetic constraint in both variables, which is consistent with the flexibility and control over sex ratios normally attributed to Hymenoptera in experimental studies. The contrasts made under Grafen's branch length assumptions are slightly different; more variance is attributed to the tips of the tree than to the base because the branches originating from the base are assumed to be longer and therefore the variance in species characters must have evolved over a longer time than those at the tips. The most negative sex ratio contrast in both analyses refers to the comparison between the *Idarnes incerta* species group and the other *Idarnes* species groups, which differ in male wingedness.

Where possible we should check the validity of evolutionary and statistical assumptions, and CAIC allows us to do both by examining the output directly. Felsenstein's assumption of random-walk evolution of continuous characters will be contradicted if the absolute value of contrasts depends on the value of estimated ancestral states at nodes. However neither of these relationships is significant (uniform branches: $F_{1,11} = 0.003$, $P = 0.961$; Grafen: $F_{1,11} = 0.278$, $P = 0.608$), showing that the Brownian motion model has been correctly applied.

A second assumption, if we want to apply a parametric statistic like regression to our data, is that the variance in the contrasts must be homogeneous. That can be tested by plotting the absolute values of the standardized contrasts against the square root of the variance in the raw contrasts (Garland *et al.*, 1992), to observe if the relationship is significant. Both sets

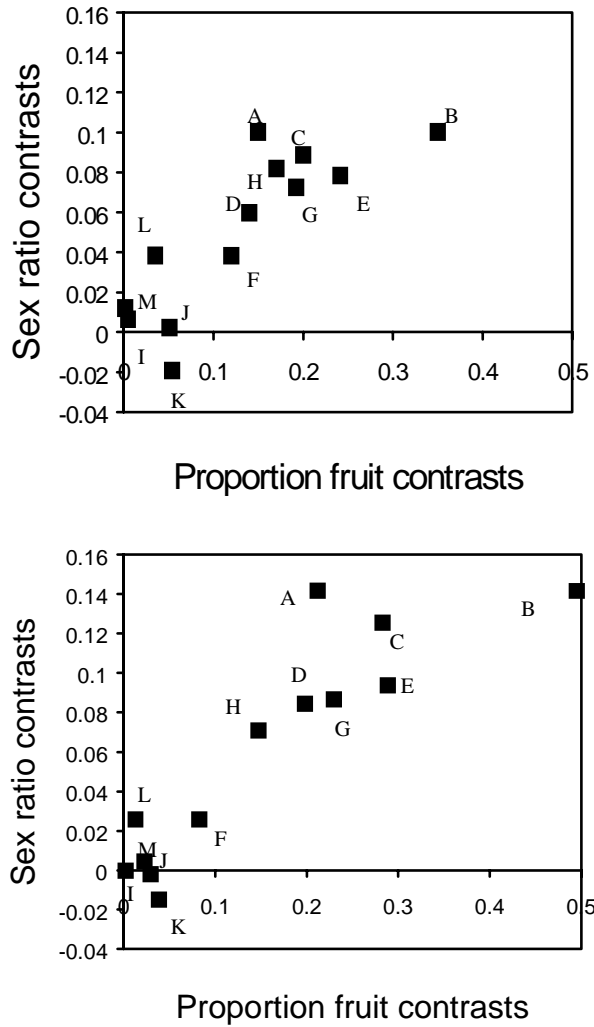


Figure 14.4: Independent contrasts of arcsine square root sex ratio on proportion fruit infested calculated in CAIC using a) uniform branch lengths and b) Grafen branch lengths. Letters refer to the contrasts across nodes in figure 14.2.

of branch length assumptions then show that simple regression is inappropriate (uniform branches: $F_{1,11} = 25.86$, $P = 0.004$, Grafen: $F_{1,11} = 28.06$, $P = 0.003$). One option would now be to apply a non-parametric test such as a sign test. Another would be to try transforming the data differently before analysis, or to try a different branch length assumption, but because we have already transformed the data and explored two branch length assumptions, a third alternative is preferable now: a weighted regression through the origin to take account of the relative uncertainty over the different data points. We use (1/square root contrast variance) as the weight variable because the magnitude of successfully standardized contrasts should be independent of the square root of their estimated variances. Regressions on contrasts must be forced through the origin because the average difference in the response variable to zero difference in explanatory variables must equal zero (Garland *et al.*, 1992).

The weighted regressions are significant for both branch length assumptions. For uniform branch lengths the estimated slope is 0.364, $F_{1,12} = 85.77$, $p < 0.0001$. For Grafen branch lengths the estimated slope is 0.359, $F_{1,12} = 75.53$, $P < 0.0001$. Thus, positive differences in proportion fruit parasitized are significantly associated with positive differences in sex ratio.

The only evolutionarily independent comparison which can be made between taxa which differ in male wingedness is between the Idarnes incerta species group and the other Idarnes species groups. Calculated using CAIC's BRUNCH algorithm, this contrast has the value 0.019 for uniform branch lengths and 0.015 for Grafen branch lengths. We note two points from this; first, that the direction of the contrast is as predicted from theory; presence compared to absence of wings in males is associated with an increase in sex ratio. Second, we cannot reject the null hypothesis that wingedness and sex ratios are evolutionarily correlated solely using contrasts between sister taxa which differ in state because this gives only one data point.

It is possible to calculate contrasts in wingedness whilst controlling for proportion fruit parasitized by using residual values of arcsine square root sex ratio for each species. Because the contrast regression slopes are less biased than the species regression slopes, residuals are calculated as the species arcsine square root sex ratio value minus the product of the contrast slope and the species value of proportion fruit parasitized. Contrasts on these residuals can then be calculated against contrasts in wingedness using the BRUNCH algorithm of CAIC. Such contrasts confirm the direction of the trend: under equal branch lengths the contrast on wingedness has the value 0.078, and under Grafen branch lengths 0.046. These are both quite substantial positive differences in arcsine square root sex ratio (see Fig. 14.4).

Table 14.2: Logistic regression analysis of fig wasp sex ratio data from West & Herre (1998a). P values are based on regular F distributions without taking phylogeny into account and on F distributions obtained from simulated sex ratio evolution across the phylogenetic tree (Fig. 14.3). ^aUniform branch lengths. ^bBranch lengths according to Grafen's (1989) method.

Variable	Deviance	DF	F	P	P^a	P^b
Constant	216.1	16				
Prop. parasitized	102.7	15	16.57	0.001	0.009	0.01
Winged	17.81	14	66.72	<0.0001	0.001	0.002

14.6.2 An analysis by comparison of real with simulated null data

Tab. 14.2 contains statistics and test results based on logistic regression analysis applied to the raw data of West & Herre. In addition to the standard significance levels not taking into account the underlying phylogeny, the table contains P values based on F distributions generated by simulations. Results are shown for the assumption of uniform branch lengths and for branch lengths computed according to Grafen's (1989) method. Simulations were carried out using Brownian motion of arcsine square root transformed sex ratios. For the ancestral sex ratio (proportion males) at the root of the tree, an estimate of 0.31 was obtained using Felsenstein's (1985) approach. The variance of the Brownian motion process was scaled by trial and error in such a way that the variance in simulated sex ratios at the tips of the tree matched the observed variance in the real data. For each of the 1000 simulated sex ratio data sets, a random number of males was drawn for each species from a binomial distribution $B(n, p)$ where p is the expected proportion of males from the simulation and n the sample size of the real study. Logistic regression was carried out and F statistics were stored. The resulting F distribution was used to compute a P value for the F statistic obtained from the analysis of the real data. Although the P values based on simulations are about tenfold higher than those based on standard F distributions, they still suggest sex ratios have not evolved independently of foundress number and the degree of local mating.

14.7 Making conclusions about comparative data

The previous section illustrates that care is needed in drawing conclusions from comparative tests because slightly different tests can give different results. One can only make conclusions within the confines of test assumptions, so we have to know what those assumptions are and be explicit about

them. If we want to know how robust our conclusions are to uncertainties in our statistical assumptions, we need to do several tests using different assumptions. Sometimes, and perhaps often, this will lead to no simple conclusion. Fair enough. Again, this kind of uncertainty is not exclusive to comparative studies: it occurs in experimental studies too where assumptions have to be made which cannot immediately be verified. It is better to be informed about our lack of knowledge and withhold judgement than to ignore it and make erroneous conclusions.

In the above example, both logistic regression on the raw species values, CAIC, and simulations based on Brownian motion suggest that increasing foundress number leads to less female-biased sex ratios. However, we have to be more careful over our conclusions about the degree of local mating. The raw species values are significantly different from those expected if sex ratio were independent of male wingedness, but because only a single change in male wingedness is probably involved, any character which differs between the two clades involved could also be responsible. Clearly phylogenetic awareness is especially valuable in this study.

14.8 A comparative biologist's toolbox

Whilst we hope our text will help readers to make effective use of their comparative data, there is no substitute for getting your hands on some comparative software and having a go. This section provides information on websites relevant to obtaining the most useful software for comparative analyses. Most include the relevant documentation.

14.8.1 Software for analyzing comparative data:

A website which provides useful information on comparative software is <http://evolution.genetics.washington.edu/phylip/software.html>

The following paragraphs detail some of the most useful implementations.

CAIC. An Apple Macintosh implementation of independent contrasts; very practical and user-friendly. Can be downloaded free of charge from <http://evolve.zoo.ox.ac.uk/CAIC/CAIC.html>

COMPARE. A range of applications for analyzing comparative data and phylogenetic trees for MS-DOS, and UNIX. Includes two independent contrast methods (the tree must be completely resolved for Felsenstein's method, and for Martin's method the tree must be completely unknown), and Phylogenetic Autocorrelation. May be downloaded free of charge from <http://work.uoregon.edu/~COMPARE/>

MacClade. A very user friendly package for exploring phylogenetic trees for Apple-Macintosh. Main strength is in reconstructing ancestral states of

categorical variables, which is very useful for pairwise comparisons. Also performs Maddison's (1990) concentrated range test for comparative analysis of two-state variables. Published by Sinauer

<http://phylogeny.arizona.edu/macclade/macclade.html>

<http://www.sinauer.com/Titles/frmaddison.htm>

Phylogenetic autocorrelation

An Apple-Macintosh application of Gittleman & Kot's method (an MS-DOS version is also available in COMPARE, above), available by anonymous file transfer from <ftp.math.utk.edu> under the directory `pub/luh`.

Phylogenetic regression. A macro written for the GLIM statistics package, distributed free by the author:

Dr Alan Grafen, St John's College, Oxford, U.K.

alan.grafen@st-johns.oxford.ac.uk

For information on GLIM see <http://www.nag.co.uk/stats/GDGE.html>

14.8.2 Obtaining phylogenetic information

A useful starter website is the tree of life website

<http://phylogeny.arizona.edu/tree/>

This links to many other sites on phylogenies.

14.8.3 Other software for exploring phylogenies:

Useful lists of phylogenetics software are available at:

<http://evolution.genetics.washington.edu/phylip/software.html>

<http://www.vims.edu/~mes/hennig/hennig.html>

<http://evolve.zoo.ox.ac.uk/>

Two of the most widely used packages for reconstructing phylogenies are:

PHYLIP. For MS-DOS and Apple-Macintosh. May be downloaded free of charge from:

<http://evolution.genetics.washington.edu/phylip.htm>

PAUP. For Apple-Macintosh. Published by Sinauer

<http://www.sinauer.com/Titles/frswofford.htm>

Acknowledgments

We are grateful to Andrea Gillmeister, Ian Hardy, Allen Herre, Scott Johnson, Carlos Machado and Stuart West.

