Testing Evolutionary Processes with Phylogenetic Patterns: Test Power and Test Limitations

Philippe GRANDCOLAS *, Pierre DELEPORTE **
& Laure DESUTTER-GRANDCOLAS *

* E.P. 90 CNRS, Laboratoire d'Entomologie, Muséum national d'Histoire naturelle,
45, rue Buffon, 75005 Paris, France
** UMR 6552 CNRS, Laboratoire de Primatologie - Biologie évolutive, Station Biologique,
Université de Rennes I, 35380 Paimpont, France

ABSTRACT

Using parsimony, phylogenetic patterns may be inferred with cladistics, and may validate predictions issued from models of evolutionary processes. The use of parsimony is needed – whatever the evolutionary model implied – to minimize the number of unwarranted hypotheses, according to the elementary rules of comparative biology. Following this minimization, patterns are less hypothetical and more independent, and a higher number of evolutionary processes may be tested. One should be aware of possible biases in the comparison of the results provided by several tests in different clades, biases related to delineation of characters and ingroups.

RÉSUMÉ

Le test des processus évolutifs par les séquences phylogénétiques : puissance et limitations du test

La phylogénie cladistique permet d'établir par économie d'hypothèses des séquences d'évolution des caractères. Ces séquences peuvent valider les prédictions issues de modèles de processus évolutifs concernant ces mêmes caractères. L'usage de la parcimonie se justifie dans ce domaine, quelque soit le modèle évolutif qui y corresponde, par la nécessité de minimiser les hypothèses gratuites en biologie comparative. Il permet d'une part de ne pas rendre les résultats trop hypothétiques, et d'autre part de ne pas obérer le test d'hypothèses supplémentaires par manque d'indépendance. Il est recommandé de prendre en compte les biais possibles dans la comparaison de résultats de plusieurs tests dans des clades différents, biais pouvant découler de la définition des caractères et des groupes à l'étude.

INTRODUCTION

Phylogenetic tests of evolutionary scenarios formally existed since approximately twenty years (ANDERSEN, 1979). Following the development of cladistics, many people were interested in taking into account phylogenetic information for testing evolutionary hypotheses, as emphasized by several seminal papers (BROOKS, 1985; GREENE, 1986; CODDINGTON, 1988, 1990; CARPENTER, 1989). More recently, a large number of reviews dealt with this research field

(FUNK & BROOKS, 1990; WANNTORP *et al.*, 1990; BROOKS & MCLENNAN, 1991, 1993; BAUM & LARSON, 1991; CODDINGTON, 1994; EGGLETON & VANE-WRIGHT, 1994a; MADDISON, 1994; SPENCE & ANDERSEN, 1994; MILLER & WENZEL, 1995; DESUTTER-GRANDCOLAS, 1996). The goal of these studies in comparative biology is to use phylogenetic patterns either to infer an evolutionary history *per se* or to test previous hypotheses of evolutionary processes (ELDREDGE & CRACRAFT, 1980; GRANDCOLAS *et al.*, 1994).

The number of available methods using phylogenetic information in the study of processes has also greatly increased (*e.g.* HARVEY & PAGEL, 1991; MILES & DUNHAM, 1993; HARVEY *et al.*, 1995; MARTINS, 1996) generally without clear distinction of their respective pre-requisites or uses (CARPENTER, 1992; GRANDCOLAS *et al.*, 1994). Only some empirical modeling studies have been carried out to evaluate and to compare these methods, and they did not settle general issues in this respect (*e.g.* GITTLEMAN & HANG-KWANG, 1994; WESTNEAT, 1995; BJÖRKLUND, 1995). Several works have also criticized the reliability of phylogenetic tests. Regarding some specific evolutionary models, tests are supposed to be flawed either because parsimony is used or because adaptation is circumstantially detected (LEROI *et al.*, 1994; FRUMHOFF & REEVE, 1994; GRETHER, 1995; SCHLUTER, 1995).

The phylogeny user who compares taxa and builds phylogenies for inferring or testing evolutionary histories could now wonder which method is the most powerful and relevant in his case study, the more likely to provide him with robust and reliable results. He could also ask what are the limitations of these methods. We try to answer these questions, focusing mainly on the phylogenetic tests of evolutionary scenarios which seem to us of prime importance regarding the aim of comparative biology.

TEST POWER

A test results from the contrast of two independent sets of data: for instance, statistical tests compare an observed distribution and an expected distribution. The phylogenetic tests of evolutionary scenarios compare phylogenetic patterns and patterns implied by evolutionary processes (*i.e.* evolutionary scenarios), to infer sound hypotheses of evolution (ELDREDGE & CRACRAFT, 1980; CARPENTER, 1989; GRANDCOLAS *et al.*, 1994). As in any test, if expected and observed data sets are incongruent, the hypothesis under test (which has been obtained using unwarranted hypotheses) is rejected as unsatisfactory. Conversely, the congruence of the two data sets provides independent support (*i.e.* corroboration) for the unwarranted hypotheses used for obtaining one of the data sets. By unwarranted, we mean hypotheses which are not substantiated directly but made by extrapolation or by logical reasoning.

Phylogenetic tests may be ranked relative to other methods of extracting historical information, according to their respective testing power. This testing power may be estimated with respect to the range of different situations in which the tests can be performed, and with respect to the ratio and the reliability of refutations which they can produce. Estimating the testing power makes necessary to assess critically the kind of items to be compared in the test, the intrinsic properties of these items and thus the way to contrast them maximally. Both the phylogenetic patterns and the evolutionary scenarios should be examined in this perspective, in order to draw the guidelines for carrying out the tests.

Lessons from the phylogenetic patterns

Minimizing the burden of hypotheses. Evolution is a historical and unique phenomenon which occurred in the past and produced similarities and differences between taxa. The aim of comparative biology is to fill the gaps existing between the taxa to understand their differences, using the principle of descent with modification (Fig. 1). Consequently, comparative biology deals mainly with hypotheses, *i.e.* the basic hypotheses of descent patterns which link the respective characters' states in the different taxa (NELSON, 1970; FARRIS, 1983). These hypotheses will never be ascertained totally, because gaps in knowledge still remain (PATTERSON, 1994). Neither fossils nor additional taxa could provide anything other than hypotheses because these additional taxa could only insert themselves between other taxa without totally filling the gaps. Consequently, any methodological advance in comparative biology should consist in decreasing as much as possible the number of hypotheses. For reconstructing the past, one should not add any extra-hypothese (*e.g. ad hoc* hypotheses *sensu* FARRIS, 1983) to the basic and necessary descent hypotheses linking character states in taxa. Any additional *ad hoc* hypothesis will remain unwarranted (unsupported by the data) and thus decrease the reliability of
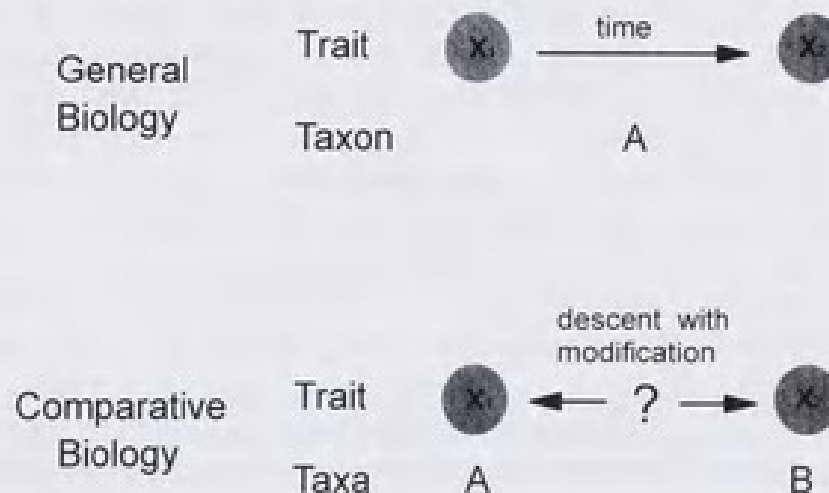

FIG. 1. — General biology deals with comparisons of different states of a trait "X" (X1 and X2) in a same taxon "A" at two different moments. Comparative biology deals with comparisons of different states of a trait "X" (X1 and X2) in two different taxa "A" and "B". In comparative biology, one relies on an assumption of descent, which will remains hypothetical ultimately (here quoted with a question mark).

the results. A usual argument for adding hypotheses that we called here "unwarranted" is to make analogy with previous case studies, in the way: "it is well-known that evolution proceeds in the way ...". For example, "it is well-known that transversions are more frequent than transitions". This kind of argument seems to us clearly inappropriate in science in the absence of directly supporting evidence.

Taking into account the principle of independence. There is another reason to decrease the number of *ad hoc* hypotheses. To test evolutionary processes with phylogenetic patterns, it is
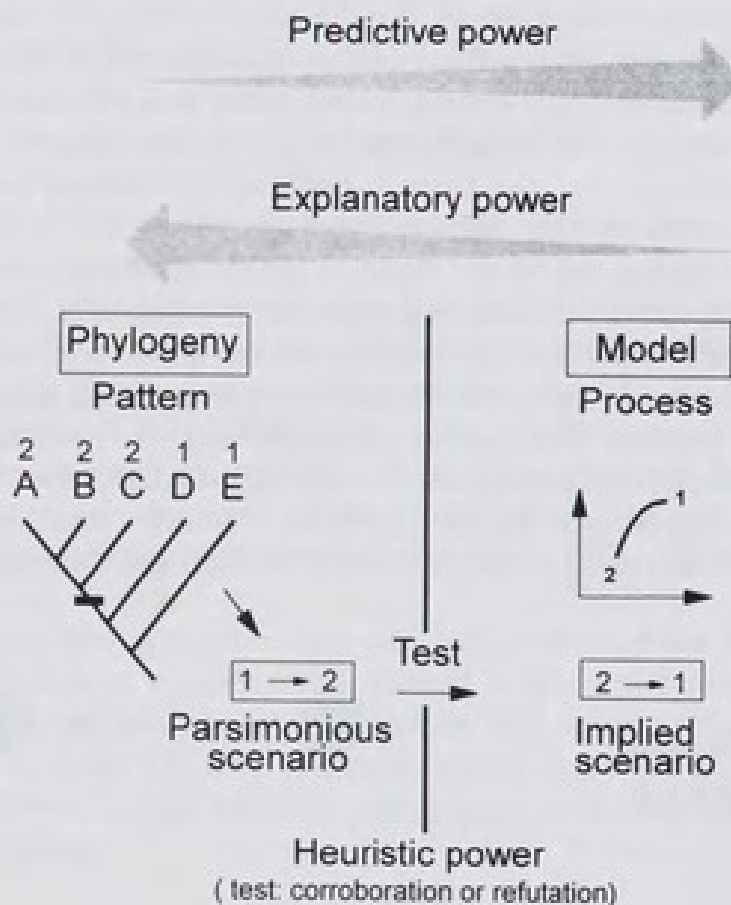
FIG. 2. — The phylogenetic test of evolutionary scenarios compares two independent issues: a pattern issued from a phylogenetic analysis (maximizing explanatory power) and a pattern issued from a model of evolutionary process (maximizing predictive power). The test itself has a maximal heuristic power, whether it provides a refutation or a corroboration as a result.

necessary to follow the principle of independence (DELEPORTE, 1993; GRANDCOLAS *et al.*, 1994). One should not test hypotheses of evolutionary processes with phylogenetic patterns which would have been inferred using these same hypotheses. The more *ad hoc* hypotheses used to infer phylogenetic patterns, the less validly evolutionary processes can be tested, *i.e.* tested with truly independent evidence.

The testing power of phylogenetic tests is inversely related to the number of *ad hoc* hypotheses made for reconstructing phylogenetic patterns. Using a lesser number of *ad hoc* hypotheses, one could test and refute a higher ratio of evolutionary processes with a higher reliability. This explicit principle is reminiscent of the earlier characterization of cladistics during the discussions among the different taxonomic schools. HENNIG (1950) himself already distinguished phylogenetic systematics from evolutionary systematics on the basis of the use of fewer *a priori* assumptions, as quoted by DUPUIS (1984).

Lessons from the evolutionary processes

The plea concerning this particular minimization of *ad hoc* hypotheses does not concern studies in general biology and especially in population biology. These kinds of biological studies mainly deal with processes rather than patterns, and they study them in a diachronic way but in the *same* taxa: the progress of a process can be observed along the time and the different states of a trait during a process may be put directly into relation without making too many hypotheses (Fig. 1). Along the time, several parameters can also be monitored to study their influence on the process. In this way, comparing a trait in the same species (or even in the same population of the same species) at different moments allows control of most influential parameters. The comparison of two different states of the trait under study at two different moments does not necessarily increase the number of uncontrolled parameters. This consequently does not decrease the number of degrees of freedom for these comparisons, as opposed to studies of comparative biology which compare different states of a trait in distinct taxa differing by many other characters.

Population biology can thus develop fairly directly testable models. Models formalize the relationships between several parameters on the basis of previous population studies. Models make predictions which can be validated by further observations on populations. The empirical validation of models is thus possible using complementary observations carried out at different moments on the same phenomenon (LEVINS, 1966; MICHALAKIS *et al.*, 1997, this volume). In general biology, predictions of models can be checked directly, while this is impossible for the same hypotheses in comparative biology. Many models in general biology are predictive regarding evolutionary processes in populations and are considered only secondarily as predictive in different situations, at a macroevolutionary level and in different taxa. These models acquire by extrapolation an heuristic value in comparative biology because their predictions can be addressed secondarily at a macroevolutionary level. The validity of models at this level can no longer be assessed empirically because the observations are no longer repeatable in the same taxa. It has been sometimes argued that validation may be possible however, using antagonistic models with opposite predictions (LEMEN & FREEMAN, 1989; MICHALAKIS *et al.*, 1997, this volume). But an identical prediction can be produced by several different models and thus cannot be validated solely by refutation of an opposite prediction generated by an antagonistic model (DUNBAR, 1989).

An evolutionary model at macroevolutionary level can only be validated by a comparison with the independent patterns which can be collected using phylogenetic analysis. This is an important methodological justification of the usefulness of phylogenetic tests of evolutionary scenarios.

Phylogenies versus models: explanatory power versus *predictive power*

Both approaches, phylogenetic analysis and process modeling, are obviously valuable for different reasons and they are complementary. There is an opportunity to compare the models of processes in general biology and the phylogenetic patterns in comparative biology. In this comparison, the patterns are testing the processes because patterns minimize *ad hoc* hypotheses at a macroevolutionary level while the models are *ad hoc* constructions at this level (Fig. 2). Analyses of patterns and processes have contrasting powers (Figs 2, 3). Phylogenetic patterns have a high explanatory power (FARRIS, 1979, 1983), because available data are explained by

themselves without any *ad hoc* additional hypothesis (Figs 2-3). Models of processes have a high predictive power, because they are designed to make predictions (Figs 2-3). The comparison of these two contrasted analyses has a higher heuristic power than each separate analysis (Fig. 2) because conclusions obtained when maximizing explanatory power are compared with conclusions obtained when maximizing predictive power.

	PHYLOGENY	EVOLUTIONARY MODEL
Concern	Pattern	Process
Object	Clade	Population / Clade
Power	Explanatory	Predictive
Level	Unique	Statistic
Reliability	Robustness	Validation
Pre-requisites	Descent with modification	Additional Hypotheses

FIG. 3. — Contrasted characteristics of phylogeny and model, including respective concern, object, power, level, reliability and pre-requisites.

With respect to these principles, parsimony is not used as a particular model of evolution but as a logic for reasoning using as few *ad hoc* hypotheses as possible (FARRIS, 1983). This point has particularly been misunderstood (*e.g.* PAGEL & HARVEY, 1989; PAGEL, 1994) and has been a blind alley in discussions for several decades as noticed by RIEPPEL (1988) and EGGLETON & VANE-WRIGHT (1994b). Parsimony must be used as a logical principle and it has inevitable consequences concerning the reconstruction of evolution. However, any other method would be less valuable, because of the use of more *ad hoc* and unwarranted hypotheses. Parsimony in data analysis for phylogeny reconstruction is like democracy in the popular joke "the worst system, but nobody has ever found a better one". Assertions such as "in this case, parsimony does not work" are soundless because one does not know how evolution has proceeded in a given case and one cannot propose a model – to mitigate parsimony use – which is free of additional and costly assumptions.

It is sometimes asserted that phylogeny has also a predictive power (RIEPPEL, 1988; SYSTEMATICS AGENDA 2000, 1994), because it supplies parsimonious hypotheses of character states when one state is unknown within part of an ingroup. This assertion is misleading because it confounds the causation and the effect of parsimony use. Parsimony is used to provide hypotheses of phylogenetic patterns, even though some character states are unknown in some

taxa, because a phylogenetic explanation is needed even with incomplete data. But parsimony is primarily not used for predicting the value of missing data, such as unknown character states. Used in this exclusive way, parsimony would be nothing else than a model, and a poor one, of phylogenetic inertia through extrapolation of character states present in the sister taxa. The use of the term "predictive" should be restricted to modeling; it is misleading in the case of phylogenetic analysis and was probably mistaken for "heuristic", "informative", or better-conceived "explanatory".

TEST LIMITATIONS

Limitations can be intrinsic or extrinsic to the methodology of tests. Some intrinsic limitations have been emphasized in recent criticisms and are the product of unwarranted predictions by particular models of evolution. As these models cannot be validated, these hypotheses of limitations are not testable and are refuted in a first step. Other intrinsic limitations deal with the very nature of cladistic phylogenetic hypotheses and should be taken into account. A first limitation is related to the robustness of phylogenetic trees on which phylogenetic tests are based. Many authors have stressed that phylogenetic trees are not necessarily correct and that studies based on phylogenies should consider carefully this point (*e.g.* EGGLETON & VANE-WRIGHT, 1994c). Although this point must be obviously a matter of concern, it could not justify rejection of phylogenetic tests based on phylogenetic trees which have been *correctly* assessed even according to only one set of data (either morpho-anatomical, or behavioral, or molecular, etc.). As in any scientific study, a reasonable amount of evidence must be taken into consideration, even if additional evidence can possibly change the results in the future, provided that these results are refutable (QUIN & DUNHAM, 1983). It could be far less hazardous to use phylogenies even if they are young hypotheses still not much discussed in the literature than to use many *ad hoc* hypotheses to test evolutionary hypotheses. Cladistic phylogenies and related phylogenetic tests – even based on limited evidence – can be refuted contrary to *ad hoc* hypotheses of macroevolution. By the way, a further examination of the problem of tree robustness may be found in this volume (WENZEL, 1997).

A second intrinsic limitation deals with the absence of temporal scales when dealing with cladistics. Minimizing unwarranted hypotheses such as "evolutionary clocks" precludes any possible absolute dating in cladistics (except minimal age estimates using fossils, which is evidence independent of cladistics *per se*). This is particularly detrimental to the comparisons between clades for testing hypotheses of niche displacement, coevolution, etc. Conversely, studies which do not use this principle increase the burden of hypotheses. For instance, the validity of the conclusions of OWENS & BENNET (1995) relies on their hypothesis of an evolutionary clock in bird clades, a hypothesis less than reliable (CRACRAFT, 1992; MINDELL, 1992; O'HARA, 1991).

Most other limitations stay far beyond the tests and are related to the general and statistical significance of the addition of the results of several tests (Fig. 4). They are extrinsic to the tests but will undoubtedly become an important matter of concern when many phylogenetic tests are achieved in the future. The addition of their results will allow generalizations (GRANDE, 1994), provided that tests are carried out without sampling bias. These possible biases will be discussed in a second step.

Model-based criticisms

Recently, several authors have criticized phylogenetic tests, considering that parsimonious reconstructions do not work under the assumptions of particular evolutionary models (LEROI *et al.*, 1994; FRUMHOFF & REEVE, 1994; GRETHER, 1995; SCHLUTER, 1995).

A first criticism was based on a misunderstanding of phylogenetic tests. According to LEROI *et al.* (1994), pattern and process would be confused in phylogenetic tests and the pattern would not be sufficient in itself to prove the existence of a corresponding process (for example, polarity testing for the adaptive value of a trait). But, many phylogeneticists do not make the assumption of an obligatory and reciprocal relationship between a kind of pattern and a kind of process (CARPENTER, 1989; CODDINGTON, 1990; GRANDCOLAS *et al.*, 1994). This point has been clearly explained by CODDINGTON (1990) who showed that phylogenetic tests of evolutionary scenarios contrast two patterns, one from the phylogeny and one implied by evolutionary process (the scenario). In this way, the phylogenetic pattern is not taken as a direct indication of the presence of a process but tests for its lack versus its *possible* presence. The presence of this pattern in phylogeny is only a corroboration of the hypothesis of process. A corroboration is always weaker than a refutation (BERNARD, 1865; POPPER, 1959); it cannot be taken as a proof and thus it is necessary to substantiate the hypothesis of process by additional
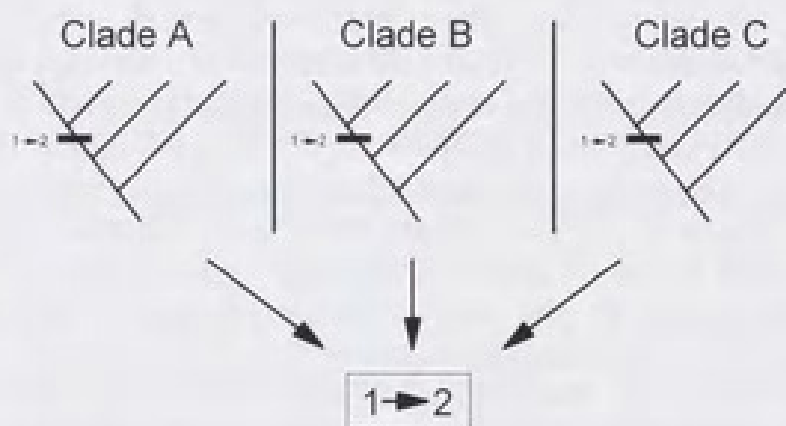

FIG. 4. — The generalization of a pattern (1 → 2) by the addition of phylogenetic analyses of three independent clades A, B and C. This generalized parsimonious pattern must be compared to the underlying scenario of an evolutionary model.

population studies. For example, character polarity may corroborate an hypothesis of adaptation but cannot prove directly the adaptive value of this character. The possible strong inference issuing from a phylogenetic test comes in fact from the observation of a phylogenetic pattern incompatible with the expected pattern, thus constituting a refutation of the tested process. More precisely, it constitutes a refutation of the idea that the process would have existed and played a major role in orienting macroevolution in the considered clade. The process is refuted by the phylogenetic pattern and not the contrary because it comprises much more unwarranted hypotheses at the macroevolutionary scale than the phylogenetic pattern. It is always possible to imagine that the process existed and left no traces behind, but this is not a testable and scientific proposition.

A second criticism deals with the possible genetic linkage between several traits (FRUMHOFF & REEVE, 1994; LEROI *et al.*, 1994; GRETHER, 1995). According to this criticism, a strong genetic link could better explain the changes of certain characters than their own adaptive value. This criticism is related to the misunderstanding commented upon above. Still, if the phylogenetic pattern of a trait is incompatible with the pattern implied by a hypothetical process concerning this trait, there is refutation of the process hypothesis, whatever any possible role of genetic linkage. As previously mentioned, if there is corroboration, there is still additional work to be achieved on populations before conclusion. This additional work should include genetic studies of linkage (see also MORAND, 1997, this volume).

A third criticism, addressed more widely, concerns some general assumptions of evolutionary models. Under specific evolutionary models dealing with rates or likelihoods of transitions and speciations, FRUMHOFF & REEVE (1994) and SCHLUTER (1995) imagined how phylogenetic tests could become inefficient in reconstituting past events. This sort of model-based assumptions are easily testable in populations but are unwarranted at a macroevolutionary scale, *a priori* to any phylogenetic reconstruction (see CARPENTER, 1997, this volume, and SCHULTZ *et al.*, 1996 for arguing against the model of FRUMHOFF & REEVE, 1994). Even if some patterns constructed with cladistics are biased because of some particular modes of evolution, there is *a priori* no other means to reconstruct them. The addition of the burden of any particular model would only make results less reliable because one can never substantiate this particular model concerning a past evolutionary phenomenon (analogy is not adequate in this respect to build a particular model).

These three kinds of criticisms either are based on a misunderstanding of the procedure of phylogenetic tests or do not follow a primary principle of comparative biology, that is to minimize unwarranted hypotheses.

Actual limitations: beyond the individual tests

Particular as well as general hypotheses can be tested using phylogenetic patterns. When dealing with general hypotheses, and to assess more strongly the conclusions, the phylogeny of several monophyletic groups may be studied to perform as many tests. Monophyletic groups may be considered as having evolved independently if they are not directly related (not sister-groups, or one group not included in another). This assumption is only statistical as even if only a few symplesiomorphic characters are shared, they can possibly determine evolutionary processes in two clades which were hypothesized to be independent. Consequently, if several tests bearing on different and independent groups provide the same results (refutation or corroboration of the hypothesis), the hypothesis is tested *by analogy* more strongly and generally. In this way, a kind of statistical significance may be assessed using the addition of several phylogenetic independent tests (Fig. 4). Such independent tests are not often possible today because of lack of available phylogenies. The opportunities of carrying out phylogenetic tests are still scarce. This should not preclude anticipating the future statistical pitfalls and the biases which could occur, but should incite to the realization of much more phylogenetic analyses.

Delineation of the trait under study. Depending on this delineation, the phylogenetic pattern may vary. Trait delineation comprises the definition of the trait itself, the definition of its

states and the establishment of primary homology. A trait may be used in phylogenetic tests either as a character for building the tree, or as an attribute optimized afterwards on the tree. Considering the trait either as a character or as an attribute depends on the primary homology of the trait (DE PINNA, 1991; GRANDCOLAS *et al.*, 1994), also named topographical correspondence by RIEPPEL (1988). The establishment of primary homology is often neglected although it is a

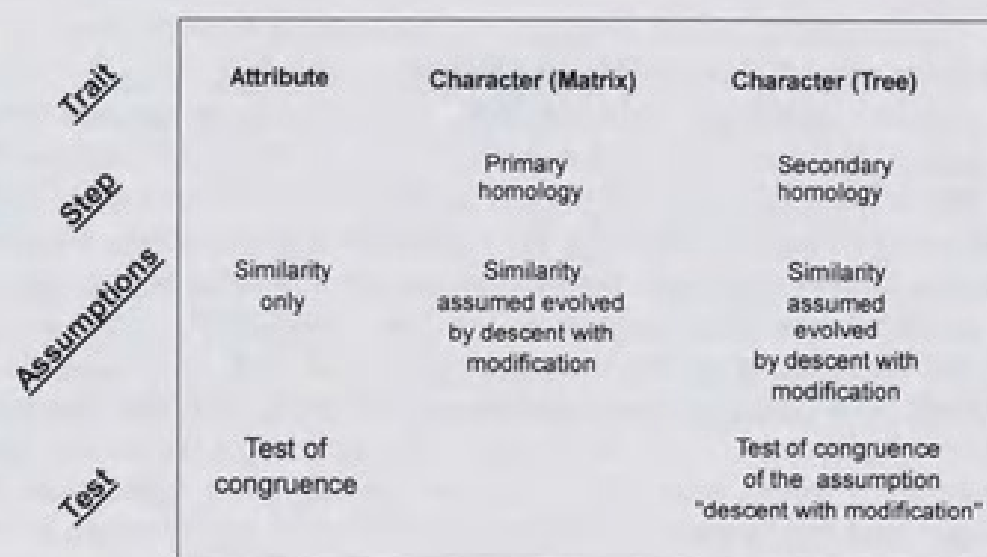| Trait	Attribute	Character (Matrix)	Character (Tree)
Step		Primary homology	Secondary homology
Assumptions	Similarity only	Similarity assumed evolved by descent with modification	Similarity assumed evolved by descent with modification
Test	Test of congruence		Test of congruence of the assumption "descent with modification"

FIG. 5. — The different operations applied during phylogenetic analysis to traits being attribute, or character in a matrix, or character in a tree. The attribute satisfies only to a statement of similarity, but not to a statement of homology; it is submitted to a test of congruence. The character is firstly assessed primary homologous on the basis of its similarity and on the basis of an assumption of descent with modification; it is secondly assessed secondarily homologous on the basis of a test of congruence of the assumption of primary homology.

critical step in phylogenetic analysis (GRANDCOLAS, 1993; GRANDCOLAS *et al.*, 1994). The primary homology of a trait is arbitrarily assessed by using statements of similarity which themselves rely mainly on the heritability and the delineation of this trait (Fig. 5). For example, traits such as geographical distributions may not be said to be strictly homologous because they are not heritable *sensu stricto* (DUPUIS, 1984). Also, macroecological traits such as "benthic" cannot be said homologous because they are defined at a too large scale (MICKEVICH & WELLER, 1991) and thus poorly defined. Most disagreements concerning primary homology come from the definition of primary homology itself. For example, all broadly similar traits could to be said to be primarily homologous (DELEPORTE, 1993), even if they are not used to build a tree, because they are similar and coded as such when mapped on the cladogram afterwards. This concept is however equivocal, in that it does not take into account the fact that these so-called homologous traits are not used as characters for building the tree, as all presumed *a priori* homologous traits should be with respect to the principle of total evidence (KLUGE, 1989). According to GRANDCOLAS *et al.* (1994), only similar traits which are used for building the tree should be said primarily homologous; they should be said to be only similar when optimized on the tree and when this mapping is the only way to assess their homology. In other words, primarily homologous traits – characters – are by definition similar traits which are postulated *a priori* to

be acquired by descent with modification and not to be homoplastic (Fig. 5). Conversely, attributes are similar but are not *a priori* postulated acquired by descent with modification (Fig. 5), and this is why one does not treat them as characters supporting phylogeny construction (but see in this volume: CARPENTER, 1997 for another distinction between characters and non-characters, or WENZEL, 1997 for arguing in favor of all traits taken as characters).
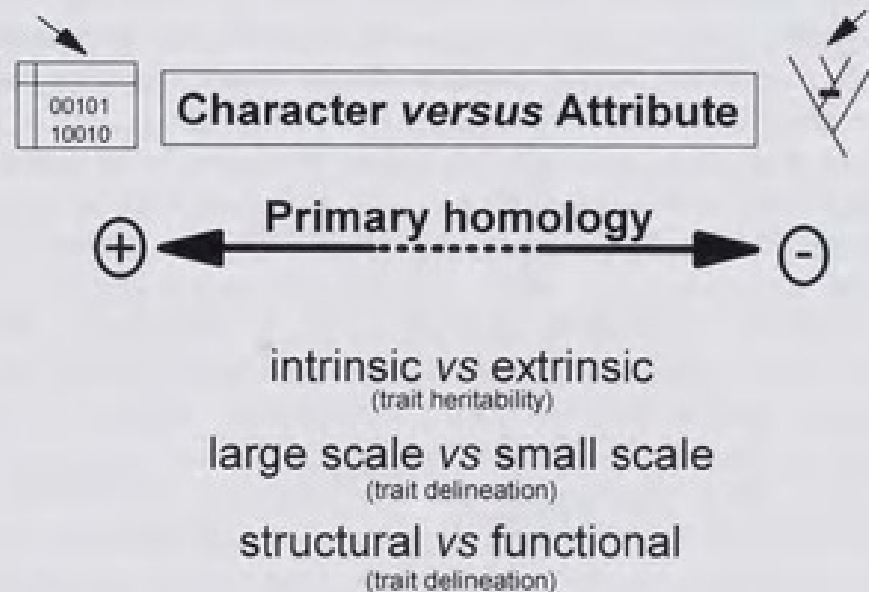

FIG. 6. — The distinction between character and attribute by the mean of a primary homology statement. This statement concerning a trait is based on the perception of its nature, intrinsic versus extrinsic (heritability), structural versus functional (delineation) and the scale large or small at which it has been defined previously (delineation).

Increasing both the accuracy of the definition and the number of states improves primary homology because the criteria of homology may be more easily applied to the trait (Fig. 6). In this way, more available phylogenetic information existing in the traits is used. A trait the primary homology of which is assessed can be used to build the tree and is thus submitted to an internal test of congruence with other characters (Fig. 5). Increasing both the accuracy of the definition and the number of states optimizes in turn the secondary homology of the trait. When the primary homology of the trait has not been assessed, this trait can be optimized (as an attribute) on the tree to discover its phylogenetic pattern. This pattern can be more precise if the definition of both the trait and its states are accurate.

Concerning the problem of character delineation and especially the "character versus attribute" alternative, one should be aware that primary homologies should not be indirectly assessed. Unfortunately, homologies of behavioral or ecological traits are often based not really on direct examination of the criteria of homology but on indirect considerations. For instance, the homology of a behavioral trait is often assessed according to its neural or its anatomical correlates. If homology of the neural scheme or anatomical structures are assessed, we would better use neural schemes or anatomy as characters. Also, homology is often assessed using circular reasoning, especially in broadly similar traits: behavioral trait is observed in two taxa

known to be closely related, and so it is considered homologous, provided that they are related. This is obviously circular. Homology is not independently assessed for the ethological trait itself but by using a model of phylogenetic inertia. Determining the homology of behavioral traits is however possible using the classical criteria of homology, but actually applied to behavior itself. Most problems of plasticity and variability which are often said to prevent assessing behavior homology must be solved by appropriate ethological studies (WENZEL, 1992).

Selection of the ingroup. This term refers here to the selection of a group of taxa supposedly monophyletic, without any contingencies related to the sampling of taxa. The ingroups are generally studied for *a priori* reasons of suitability for specific phylogenetic tests of characters. Ingroups are often studied also according to some constraints of feasibility: are the taxa well known, have their phylogeny or at least their characters been preliminarily studied? A phylogenetic test deals with the evolution of one or several traits from an ancestral state toward derived state(s), possibly including reversals; this means that the group on which the test is carried out comprises taxa showing at least two states for each trait. Also, the groups under study are generally relatively small, still because of constraints of feasibility. Phylogenetic studies of larger groups are rarely carried out because many more character state occurrences must be documented according to the increased number of terminal taxa. Ingroups are consequently most often relatively *small* in size and *diverse* with respect to the trait under study. Consequently, patterns inferred from these phylogenies will be submitted statistically to scale effects. Comparing the results of several phylogenetic tests carried out on different clades could lead to a bias which, in turn, could prevent a statistical estimate of the general prevalence of a pattern and to assess the validity of the model corresponding to this pattern. For example, if someone wants to study the evolution of flying kinematics and behavior in insects, he would probably focus on Diptera, as this is the order which is currently very diverse and well-known in this respect. But he would not analyze the whole order of Diptera because to examine hundreds of taxa in this group will overwhelm his capacity to carry out phylogenetic studies within a few years. Thus, he would select a few groups which are *smaller*, which have been already partly studied, and which are *diverse* with respect to flying behavior. Selected groups should necessarily be diverse (character diversity), otherwise no comparative study may be carried out for want of different states of traits to be compared.

As they are statistically smaller and more diverse than if they were truly taken randomly in the tree of life, ingroups may present a non-random selection of patterns which are used to test evolutionary processes. In our example, our Dipterist would have certainly not selected very large taxa with very few variation in flying behavior (*e.g.* a monophyletic tribe comprising 500 species, of which 499 have a first kind of flight and only one another kind). These groups would be excluded from the analyses. Afterwards, generalizations based on these studies would not take into account patterns which could be more frequent in large and homogeneous groups. This non-random selection may be expected to be particularly biased. Indeed, the diversity of a given character should statistically increase with the size of a group. Thus, choosing small and diverse groups excludes most of groups present in a given part of the tree of life, those which are larger and moderately diverse, and those which are of the same size and which are not diverse.

The patterns and the relevant tested processes (Figs 7-8, see also GRANDCOLAS *et al.*, 1994) are listed below with the possible bias induced by the choice of the ingroup. The biases are mentioned below provided that all things are equal otherwise in the ingroup and in the tree of life, except the ingroup size and the diversity of the character under study in this ingroup. These biases may be expected statistically only (*i.e.* for a large number of ingroups); it is obvious that a unique and particular group may not conform to the statistical expectation.

– Polarity (testing for adaptation, Fig. 7): size and diversity of the ingroup may or may not have particular scale effects regarding this pattern/process. Polarity cannot be expected to have a particular value in a small and diverse group and only depends on the distribution of character' states on the taxa and on the structure of the phylogenetic tree.

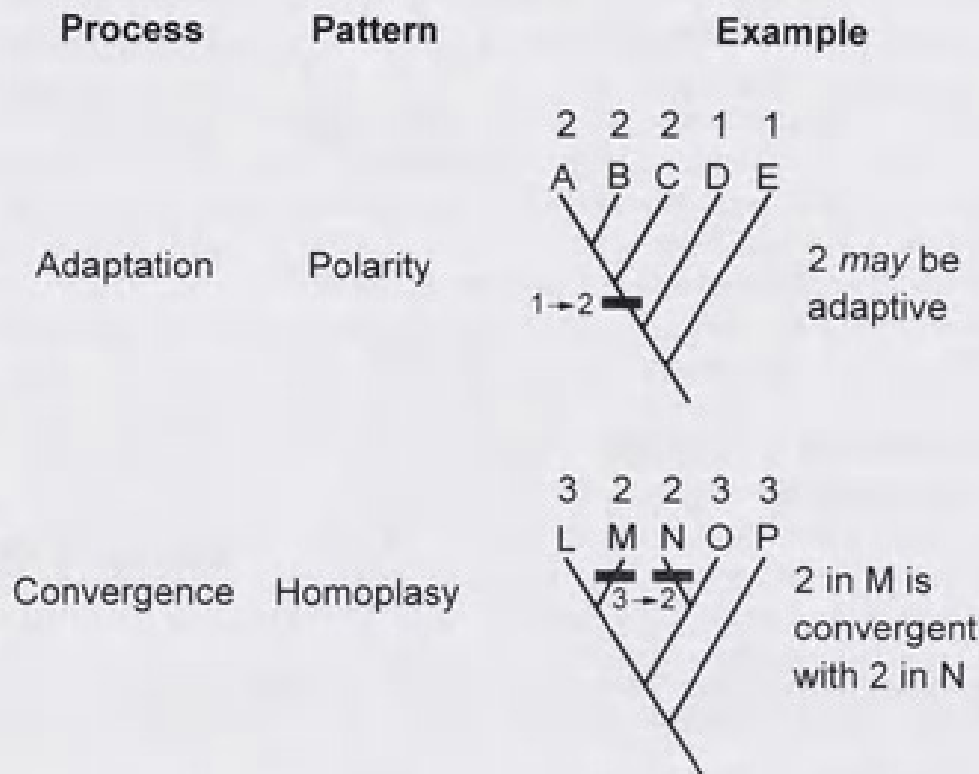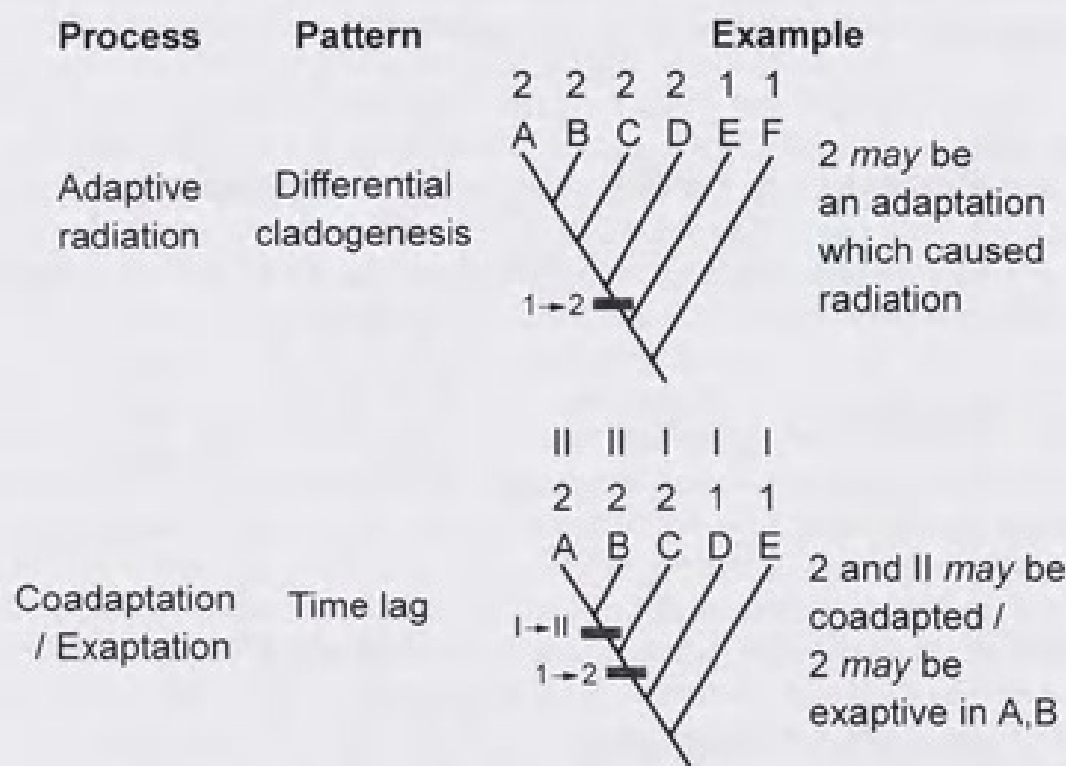| Process	Pattern	Example
Adaptation	Polarity	2 *may* be adaptive
Convergence	Homoplasy	2 in M is convergent with 2 in N

FIG. 7. — Two patterns relevant to the phylogenetic test of two processes (see GRANDCOLAS *et al.*, 1994 for more details). From left to right, the process to be tested, the pattern to be searched for testing, an example of phylogenetic test with its issue.

– Homoplasy (testing for convergence, Fig. 7): small and diverse ingroups may present statistically less homoplastic patterns because of the decrease of the number of subordinated nodes after a change in character state. The bias concerning this pattern is only related to the size of the ingroup: small ingroups do not allow to document as many reversals as could be expected because small ingroups have statistically fewer nodes. If there is a change of states of a character at a given node, there is simply more cases with no existing subordinated nodes which could permit to document another subsequent change of state such as a reversal.

– Time lag (testing for coadaptation-exaptation, Fig. 8): when testing for coadaptation or exaptation, (relative) time lag between the changes of two traits or between a trait and its function are searched for in phylogenies. Using smaller and diverse groups, there is a lower number of nodes where changes can take place. This can bias the correlation studies between two traits: after the change of a first trait, subsequent changes could take place in fewer places. Consequently, a smaller number of changes will necessarily be observed. This will bias the

FIG. 8. — Two other patterns relevant to the phylogenetic test of two other processes (see GRANDCOLAS *et al.*, 1994 for more details). From left to right, the process to be tested, the pattern to be searched for testing, an example of phylogenetic test with its issue.

frequency of observed time lags and will provide us with fewer corroborations of coadaptation-exaptation. This statement does not refer to a probabilistic approach for testing coadaptation-exaptation, such as that presented by MADDISON (1994) for challenging the views of SILLÉN-TULLBERG (1988). Probabilistic approaches deal with events occurring within the clades while our statement concerns the statistical meaning of (in)congruent results obtained from several clades.

– Differential cladogenesis (testing for radiation, Fig. 8): small ingroups with a high number of evolutionary changes cannot show relatively differential cladogenesis concerning the trait under study. Important differential cladogenesis can exist by definition only in very large ingroups because they imply a high number of taxa in the subgroup where occurred the most important cladogenesis. This can prevent to test for the importance of adaptive radiation which is

the process corresponding to the phylogenetic pattern of differential cladogenesis (GUYER & SLOWINSKI, 1991). This can prevent conversely testing for the role of evolutionary stasis, because the chosen small ingroups with a high number of evolutionary changes may not show evolutionary stasis.

Smaller ingroups are also *statistically* and *relatively* more recent groups, compared to larger ingroups, provided that both are taken in the same inclusive monophyletic group. Depending on the stability of evolutionary rates, this could lead to study only the relatively more recent evolutionary events. This is detrimental to the tests of evolutionary hypotheses which are linked to particular climatic or geological periods (but we can note that using too large a group could also lead to irrelevant correlations between a relatively old phylogenetic pattern and much more recent geological or climatic events). It should be kept in mind that the relation between ingroup size and age is not absolute but statistical. There also exist a few small and relatively old groups among all possible ingroups taken in the same inclusive monophyletic group (*e.g.* the so-called "relict taxa").

The last bias, but not the least, is related to the relevance of the ingroup for testing a particular evolutionary model. The phylogenetic test is designed to refute or to corroborate the prediction of an evolutionary model taking into account a number N of factors. The model could not be tested correctly when only $(N - 1)$ factors are considered in the phylogenetic test. This situation would occur if $(N - 1)$ factors are represented as apomorphies in the ingroup and if the N th factor is represented by a symplesiomorphy of the ingroup. This factor/plesiomorphy could make either trivial or extremely rare the pattern corroborating the model and could thus bias strongly the test toward corroboration or refutation. A recent example may be found in studies of Hymenoptera, where reversals of sociality were documented in Halictidae using phylogeny. PACKER *et al.* (1994) interestingly questioned why so many sociality reversals occur, while no appearances were documented. Together with other reasons, the phylogenetic inertia may have been quite important in biasing the tests. In Hymenoptera, most theories of social evolution put forward the role of brood care for favoring sociality. Higher-level phylogenetic analysis shows that brood care (the N th variable) is ancestral to Halictidae and this could bias the study toward a minimization of appearance events. Only studies at a much wider phylogenetic scale could adequately document appearances of sociality, for instance succeeding to the appearance of brood care and not preceding it. Another example deals with the origin of complex reproductive behaviors in cockroaches. These behaviors – ovoviviparity and viviparity – evolved following the appearance of "deposition of ootheca after sclerotization", which is apomorphic in cockroaches, relative to mantids and termites (GRANDCOLAS, 1996). If the females did not keep their ootheca after sclerotization, they could not have evolved toward subsequent retraction and nutrition of oothecae in a brood sac (ovoviviparity and viviparity). Anyone who would like to study subsequent evolution of reproductive behavior in a particular group of cockroaches should not forget that the character "deposition of ootheca after sclerotization", plesiomorphic at this level, is still influential (ROTH, 1989).

CONCLUSION

Comparative biology is still a young and growing research field, as was phylogenetics when HENNIG (1965) published one of his last methodological accounts. Following the development of

phylogenetic methodology, it is now necessary to elaborate a cohesive methodology which can take into account the possible interrelations of phylogenetic patterns with evolutionary processes (and relevant models). This is generally made through the phylogenetic test of patterns which are expected under some process hypotheses.

As a contribution to this methodology, three rules are proposed which could improve phylogenetic analysis both intrinsically and extrinsically. These improvements should increase the phylogenetic test power and decrease the test limitations.

First, the burden of hypotheses in phylogenetic analysis should be reduced by decreasing the number of unwarranted hypotheses (with parsimony use). Comparative biology proceeds using hypotheses only. Adding unwarranted extra-hypotheses is detrimental to the reliability of the results.

Second, the independence of phylogenetic patterns relative to process hypotheses should be enhanced the same way, by decreasing the number of *ad hoc* hypotheses used to infer them. Particularly, to test an hypothesis of process, one should not use patterns inferred using this same process hypothesis.

Third, statistical bias during the generalization of the tests should be minimized. When several similar tests are carried out on different ingroups, their results may be compared to generalize them. The possible peculiarities of ingroups should be taken into account to minimize the possible bias in the generalization.

The first two rules deal with a general problem encountered in many research fields of evolutionary biology. Minimal hypotheses (sometimes named null hypotheses or null models, *e.g.* PATTERSON, 1994) are wanted in comparative studies as well as in population studies of adaptation (GOULD & LEWONTIN, 1979) or in studies of biotic interactions (QUINN & DUNHAM, 1983). These minimal hypotheses are needed to check the validity of the *ad hoc* hypotheses used to reconstruct the past. Both a lack of minimal hypotheses or an abuse of *ad hoc* hypotheses will make the results flawed or unreliable. It is stressed that comparative studies should take this principle into account, for consideration paid to previous methodological analyses in evolutionary biology. We must not reinvent the wheel in comparative biology, disregarding methodological advances in phylogenetics or in evolutionary biology.

ACKNOWLEDGEMENTS

REFERENCES

ANDERSEN, N. M., 1979. — Phylogenetic inference as applied to the study of evolutionary diversification of semiaquatic bugs (Hemiptera: Gerromorpha). *Systematic Zoology*, **28**: 554-578.

BAUM, D. A. & LARSON, A., 1991. — Adaptation reviewed: A phylogenetic methodology for studying character macroevolution. *Systematic Zoology*, **40**: 1-18.

BERNARD, C., 1865. — *Introduction à l'étude de la médecine expérimentale*. Paris, Baillère. [reprinted in 1966, Paris, Garnier Flammarion: 1-318].

BJORKLUND, M., 1994. — The independant contrast method in comparative biology. *Cladistics*, **10**: 425-433.

BROOKS , D. R., 1985. — Historical ecology: a new approach to studying the evolution of ecological associations. *Annals of the Missouri Botanical Garden*, **72**: 660-680.

BROOKS, D. R. & McLENNAN, D. A., 1991. — *Phylogeny, Ecology, and Behavior. A Research Program in Comparative Biology*. Chicago, The University of Chicago Press: 1-434.

BROOKS, D. R. & McLENNAN, D. A., 1993. — *Parascript. Parasites and the language of evolution*. Washington, Smithsonian Institution Press: 1-429.

CARPENTER, J. M., 1989. — Testing scenarios: Wasp social behavior. *Cladistics*, **5**: 131-144.

CARPENTER, J. M., 1992. — Comparing methods. Review of "The comparative method in evolutionary biology". Paul H. Harvey and Mark D. Pagel. 1991. Oxford University Press. *Cladistics*, **8**: 191-196.

CARPENTER, J. M., 1997. — Phylogenetic Relationships among European *Polistes* and the Evolution of Social Parasitism (Hymenoptera: Vespidae; Polistinae). *In*: P. GRANDCOLAS, *The Origin of Biodiversity in Insects: Phylogenetic Tests of Evolutionary Scenarios*. Mémoires du Muséum national d'Histoire naturelle, **173**: 135-161.

CODDINGTON, J. A., 1988. — Cladistic tests of adaptational hypotheses. *Cladistics*, **4**: 3-22.

CODDINGTON, J. A., 1990. — Bridges between evolutionary pattern and process. *Cladistics*, **6**: 379-386.

CODDINGTON, J. A., 1994. — The roles of homology and convergence in studies of adaptation. *In*: P. EGGLETON & R. VANE-WRIGHT, *Phylogenetics and Ecology. Linnean Society Symposium Series. n°17*. London, Academic Press: 53-78.

CRACRAFT, J., 1992. — Review of SIBLEY and AHLQUIST, Phylogeny and classification of birds. *Molecular Biology and Evolution*, **9**: 182-186.

DELEPORTE, P., 1993. — Characters, attributes and tests of evolutionary scenarios. *Cladistics*, **9**: 427-432.

DE PINNA, M. C. C., 1991. — Concepts and tests of homology in the cladistic paradigm. *Cladistics*, **7**: 367-394.

DESUTTER-GRANDCOLAS, L., 1996. — Studies in cave life evolution: a rationale for future theoretical developments using phylogenetic inference. *Journal of zoological systematics and evolutionary research*, **33**: in press.

DUNBAR, R. I. M., 1989. — Ecological modelling in an evolutionary context. *Folia Primatologica*, **53**: 235-246.

DUPUIS, C., 1984. — Willi Hennig's impact on taxonomic thought. *Annual Review of Ecology and Systematics*, **15**: 1-24.

EGGLETON, P. & VANE-WRIGHT, R., (eds), 1994a. *Phylogenetics and Ecology. Linnean Society Symposium Series. n°17*. London, Academic Press: 1-376.

EGGLETON, P. & VANE-WRIGHT, R. I., 1994b. — Introduction. *In*: P. EGGLETON & R. I. VANE-WRIGHT, *Phylogenetics and Ecology. Linnean Society Symposium Series. n°17*. London, Academic Press: IX-X.

EGGLETON, P. & VANE-WRIGHT, R. I., 1994c. — Some principles of phylogenetics and their implications for comparative biology. *In*: P. EGGLETON & R. I. VANE-WRIGHT, *Phylogenetics and Ecology. Linnean Society Symposium Series. n°17*. London, Academic Press: 347-366.

ELDREDGE, N. & CRACRAFT, J., 1980. — *Phylogenetic Patterns and the Evolutionary Process. Method and theory in Comparative Biology*. New York, Columbia University Press: 1-349.

FARRIS, J. S., 1979. — The information content of the phylogenetic system. *Systematic Zoology*, **28**: 483-519.

FARRIS, J. S., 1983. — The logical basis of the phylogenetic system. *In*: N. I. PLATNICK & V. A. FUNK, *Advances in Cladistics, Vol. 2: Proceedings of the Second Meeting of the Willi Hennig Society*. New York, Columbia University Press: 7-36.

FRUMHOFF, P. C. & REEVE, H. K., 1994. — Using phylogenies to test hypotheses of adaptation: A critique of some current proposals. *Evolution*, **48**: 172-180.

FUNK, V. A. & BROOKS, D. R., 1990. — *Phylogenetic Systematics as the Basis of Comparative Biology*. Washington, Smithsonian Institution Press: 1-45.

GITTLEMAN, J. L. & HANG-KWANG, L., 1994. — Phylogeny, evolutionary models and comparative methods: a simulation study. *In*: P. EGGLETON & R. I. VANE-WRIGHT, *Phylogenetics and Ecology. Linnean Society Symposium Series n°17*. London, Academic Press: 103-122.

GOULD, S. J. & LEWONTIN, R. C., 1979. — The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London, B*, **205**: 581-598.

GRANDCOLAS, P., 1993. — The origin of biological diversity in a tropical cockroach lineage: a phylogenetic analysis of habitat choice and biome occupancy. *Acta Oecologica*, **14**: 259-270.

GRANDCOLAS, P., 1996. — The phylogeny of cockroach families: a cladistic appraisal of morpho-anatomical data. *Canadian Journal of Zoology*, **74**: 508-527.

GRANDCOLAS, P., DELEPORTE, P. & DESUTTER-GRANDCOLAS, L., 1994. — Why to use phylogeny in evolutionary ecology? *Acta Oecologica*, **15**: 661-673.

GRANDE, L., 1994. — Repeating patterns in nature, predictability, and "impact" in science. *In*: L. GRANDE & O. RIEPPEL, *Interpreting the Hierarchy of Nature. From Systematic Patterns to Evolutionary Process Theories*. San Diego, Academic Press: 61-84.

GREENE, H. W., 1986. — Diet and arboreality in the Emerald Monitor, *Varanus prasinus*, with comments on the study of adaptation. *Fieldiana, Zoology*, **31**: 1-12.

GRETHER, G. F., 1995. — Genetic correlation complicates the interpretation of phylogenetic transitions in sexually selected traits. *Behavioral Ecology*, **6**: 349.

GUYER, C. & SLOWINSKI, J. B., 1991. — Comparisons of observed phylogenetic topologies with null expectations among three monophyletic lineages. *Evolution*, **45**: 340-350.

HARVEY, P. H., LEIGH BROWN, A. J. & MAYNARD SMITH, J., (eds), 1995. — New uses for new phylogenies. *Philosophical Transactions of the Royal Society London, B*, **349**: 1-118.

HARVEY, P. H. & PAGEL, M. D., 1991. — *The Comparative Method in Evolutionary Biology*. Oxford, Oxford University Press: 1-248.

HENNIG, W., 1950. — *Grundzüge einer Theorie der phylogenetischen Systematik*. Berlin, Deutscher Zentralverlag: 1-370.

HENNIG, W., 1965. — Phylogenetic systematics. *Annual Review of Entomology*, **10**: 97-116.

KLUGE, A. G., 1989. — A concern for evidence and a phylogenetic hypothesis for relationships among *Epicrates* (Boidae, Serpentes). *Systematic Zoology*, **38**: 1-25.

LEMEN, C. A. & FREEMAN, P. W., 1989. — Testing macroevolutionary hypotheses with cladistic analyses: evidence against rectangular evolution. *Evolution*, **43**: 1538-1554.

LEROI, A. M., ROSE, M. R. & LAUDER, G. V., 1994. — What does the comparative method reveal about adaptation. *The American Naturalist*, **143**: 381-402.

LEVINS, R., 1966. — The strategy of model building in population biology. *The American Scientist*, **54**: 421-431.

MADDISON, D. R., 1994. — Phylogenetic methods for inferring the evolutionary history and processes of change in discretely valued characters. *Annual Review of Entomology*, **39**: 267-292.

MARTINS, E. P. (ed.), 1996. — *Phylogenies and the Comparative Method in Animal Behaviour*. New York, Oxford University Press: 1-415.

MICHALAKIS, Y., WAJNBERG, E. & BERNSTEIN, C., 1997. — On the utility of mathematical models and their use in evolutionary biology. *In*: P. GRANDCOLAS, *The Origin of Biodiversity in Insects: Phylogenetic Tests of Evolutionary Scenarios*. Mémoires du Muséum national d'Histoire naturelle, **173**: 47-52.

MICKEVICH, M. F. & WELLER, S. J., 1990. — Evolutionary character analysis: tracing character change on a cladogram. *Cladistics*, **6**: 137-170.

MILES, D. B. & DUNHAM, A. E., 1993. — Historical perspectives in ecology and evolutionary biology: the use of phylogenetic comparative analyses. *Annual Review of Ecology and Systematics*, **24**: 587-619.

MILLER, J. S. & WENZEL, J. W., 1995. — Ecological characters and phylogeny. *Annual Review of Entomology*, **40**: 389-415.

MINDELL, D. P., 1992. — DNA-DNA hybridization and avian phylogeny. *Systematic Biology*, **41**: 126-134.

MORAND, S., 1997. — Comparative analyses of continuous data: the need to be phylogenetically correct. *In*: P. GRANDCOLAS, *The Origin of Biodiversity in Insects: Phylogenetic Tests of Evolutionary Scenarios*. Mémoires du Muséum national d'Histoire naturelle, **174**: 73-90.

NELSON, G. F., 1970. — Outline of a theory of comparative biology. *Systematic Zoology*, **19**: 373-384.

O'HARA, R. J., 1991. — Phylogeny and classification of birds: a study in molecular evolution. Review. *Auk*, **108**: 990-994.

OWENS, I. P. F. & BENNETT, P. M., 1995. — Ancient geological diversification explains life-history variation among living birds. *Proceedings of the Royal Society of London, B*, **261**: 227-22.

PACKER, L., TAYLOR, J. & RICHARDS, M., 1994. — Social devolution in sweat bees. *In*: A. LENOIR, G. ARNOLD & M. LEPAGE , *Les Insectes sociaux. 12th Congress of the International Union for the Study of Social Insects IUSSI.*. Villetaneuse, Université Paris Nord: 358.

PAGEL, M. D., 1994. — The adaptationist wager. *In*: P. EGGLETON & R. I. VANE-WRIGHT, *Phylogenetics and Ecology. Linnean Society Symposium Series. n°17*. London, Academic Press: 29-51.

PAGEL, M. & HARVEY, P. H., 1989. — Comparative methods for examining adaptations depend on evolutionary models. *Folia Primatologica*, **53**: 203-220.

PATTERSON, C., 1994. — Null or minimal models. In: R. W. SCOTLAND, D. J. SIEBERT & D. M. WILLIAMS, *Models in Phylogeny Reconstruction. Systematics Association special volume n°52*, Oxford, Clarendon Press: 173-192.

QUINN, J. F. & DUNHAM, A. E., 1983. — On hypothesis testing in ecology and evolution. *The American Naturalist*, **122**: 602-617.

RIEPPEL, O., 1988. — *Fundamentals of Comparative Biology*. Basel, Birkhäuser Verlag: 1-202.

ROTH, L. M., 1989. — *Sliferia*, a new ovoviviparous cockroach genus (Blattellidae) and the evolution of ovoviviparity in Blattaria (Dictyoptera). *Proceedings of the Entomological Society of Washington*, **91**: 441-451.

SCHLUTER, D., 1995. — Uncertainty in ancient phylogenies. *Nature*, **377**: 108-109.

SCHULTZ, T. R., COCROFT, R. B. & CHURCHILL, G. A., 1996. — The reconstruction of ancestral character states. *Evolution*, **50**: 504-511.

SILLÉN-TULLBERG, B., 1988. — Evolution of gregariousness in aposematic butterfly larvae: a phylogenetic analysis. *Evolution*, **42**: 293-305.

SPENCE, J. R. & ANDERSEN, N. M., 1994. — Biology of water striders: interactions between systematics and ecology. *Annual Review of Entomology*, **39**: 101-128.

SYSTEMATICS AGENDA 2000, 1994. — *Systematics Agenda 2000. Charting the Biosphere. Technical Report*: 1-34.

WANNTORP, H. E., BROOKS, D. R., NILSSON, T. & NYLIN, S., 1990. — Phylogenetic approaches in ecology. *Oikos*, **57**: 119-132.

WENZEL J. W., 1992. — Behavioral homology and phylogeny. *Annual Review of Ecology and Systematics*, **23**: 361-381.

WENZEL J. W., 1997. — When is a phylogenetic test good enough? *In*: P. GRANDCOLAS, *The Origin of Biodiversity in Insects: Phylogenetic Tests of Evolutionary Scenarios*. Mémoires du Muséum national d'Histoire naturelle, **174**: 31-45.

WESTNEAT, M. W., 1995. — Feeding, function, and phylogeny: analysis of historical biomechanics in labrid fishes using comparative methods. *Systematic Biology*, **44**: 361-383.