

Selection of evolutionary models for phylogenetic hypothesis testing using parametric methods

B. C. EMERSON, K. M. IBRAHIM* & G. M. HEWITT

School of Biological Sciences, University of East Anglia, Norwich, UK

Keywords:

hypothesis;
Kishino–Hasegawa test;
parametric bootstrap;
phylogeny.

Abstract

Recent molecular studies have incorporated the parametric bootstrap method to test *a priori* hypotheses when the results of molecular based phylogenies are in conflict with these hypotheses. The parametric bootstrap requires the specification of a particular substitutional model, the parameters of which will be used to generate simulated, replicate DNA sequence data sets. It has been both suggested that, (a) the method appears robust to changes in the model of evolution, and alternatively that, (b) as realistic model of DNA substitution as possible should be used to avoid false rejection of a null hypothesis. Here we empirically evaluate the effect of suboptimal substitution models when testing hypotheses of monophyly with the parametric bootstrap using data sets of mtDNA cytochrome oxidase I and II (COI and COII) sequences for Macaronesian *Calathus* beetles, and mitochondrial 16S rDNA and nuclear ITS2 sequences for European *Timarcha* beetles. Whether a particular hypothesis of monophyly is rejected or accepted appears to be highly dependent on whether the nucleotide substitution model being used is optimal. It appears that a parameter rich model is either equally or less likely to reject a hypothesis of monophyly where the optimal model is unknown. A comparison of the performance of the Kishino–Hasegawa (KH) test shows it is not as severely affected by the use of suboptimal models, and overall it appears to be a less conservative method with a higher rate of failure to reject null hypotheses.

Introduction

Felsenstein (1985) proposed the use of the general statistical procedure of repeated sampling with replacement from a data set (bootstrapping) to estimate confidence limits of the internal branches in a phylogenetic tree. Although there have been debates on the interpretation of bootstrap results (e.g. Felsenstein & Kishino, 1993; Hillis & Bull, 1993) this is still the commonest method for assessing confidence in phylogenies. However, this nonparametric bootstrap method does not provide the

opportunity to test *a priori* hypotheses about the phylogeny of a group. The alternative, parametric bootstrap method does (Efron, 1982; Felsenstein, 1988; Bull *et al.*, 1993; Hillis *et al.*, 1996; Huelsenbeck *et al.*, 1996a, b). In contrast to the nonparametric method where replicate character matrices are generated by randomly sampling the original data, this method creates replicates using numerical simulation. A model of evolution is assumed, and its parameters are estimated from the empirical data. Then, using the same model and the estimated parameters, as many replicate data sets as needed of the same size as the original are simulated. Phylogenies constructed from these replicate character matrices are used to generate a distribution of difference measures between the maximum likelihood and *a priori* phylogenetic hypotheses. Huelsenbeck *et al.* (1996b) have advocated a likelihood based phylogenetic reconstruction for the generation of a distribution of difference measures but because of the excessive computation required a

Correspondence: Brent Emerson, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

Tel.: +44 1603 592 947; fax: +44 1603 592 250;

e-mail: b.emerson@uea.ac.uk

**Present address:* Department of Zoology, Southern Illinois University, Carbondale, IL 62901–6501, USA

parsimony based approach is an acceptable alternative with only a slightly lower level of discrimination (Hillis *et al.*, 1996).

Many recent molecular studies have used the parametric bootstrap method for testing *a priori* hypotheses against their molecular based phylogenies (Mallat & Sullivan, 1998; Ruedi *et al.*, 1998; Flook *et al.*, 1999; Jackman *et al.*, 1999; Oakley & Phillips, 1999). Other studies have utilized variations of the method to examine the possibility of long branch attraction (Felsenstein, 1978; Hendy & Penny, 1989; Huelsenbeck, 1997) causing erroneous topologies (Maddison *et al.*, 1999; Tang *et al.*, 1999), to estimate the sequence length required to resolve a particular phylogenetic tree (Halanych, 1998; Flook *et al.*, 1999), to evaluate the efficiency of different methods of phylogenetic reconstruction (Bull *et al.*, 1993; Hwang *et al.*, 1998), and to test the adequacy of models of DNA sequence evolution (Goldman, 1993; Yang *et al.*, 1994).

A critical first step of the parametric bootstrap is specifying a particular substitutional model, whose parameters will be estimated from the empirical data, and which will subsequently be used when generating the simulated, replicate DNA sequences. Hillis *et al.* (1996) stated that in limited studies the method appeared to be robust to changes in the model of evolution. However, Huelsenbeck *et al.* (1996a) suggest as realistic model of DNA substitution as possible should be used to reduce Type I error (false rejection of a null hypothesis) implying that parameter-rich models may perform better. Zhang (1999) has shown that if the substitution model is inadequate then the estimates of the substitution parameters will be biased. By using biased estimates of the parameters the simulated sequence data will deviate from the empirical data, such that the null distribution of tree score differences will be distorted.

In this study we evaluate empirically the effect of the choice of substitutional model when testing hypotheses of monophyly using DNA sequence data. We have chosen three gene regions (two mitochondrial and one nuclear) commonly used in phylogenetic analyses: cytochrome oxidase subunit I and II (COI and COII), the large mitochondrial ribosomal subunit gene (16S), and the nuclear ribosomal internal transcribed spacer 2 (ITS2). For COI and COII we have used published sequence data for Macaronesian *Calathus* beetles (Carabidae) (Emerson *et al.*, 1999, 2000) and for 16S and ITS2 we have used published sequence data for European *Timarcha* beetles (Chrysomelidae) (Gómez-Zurita *et al.*, 2000a, b). For each of these data sets we also examine the comparative performance of the nonparametric Kishino–Hasegawa (KH) test (Kishino & Hasegawa, 1989) which compares competing phylogenetic hypotheses under the maximum likelihood (ML) optimality criterion. This method uses likelihood ratio tests of the statistical significance of competing tree topologies. All three data sets offer clear

a priori hypotheses of monophyly for testing against empirical DNA sequence data.

Materials and methods

The following steps summarize the methodology used. For each DNA sequence data set (*Calathus* COI and COII, *Timarcha* 16S and *Timarcha* ITS2) we: (1) defined an appropriate model of nucleotide substitution for the purpose of parameterizing a maximum likelihood phylogenetic analysis of the data. For each of the three data sets the results of the phylogenetic analysis are in conflict with null hypotheses of monophyly at specified nodes. (2) To test the significance of the conflict in each case, we have carried out a parametric bootstrap analysis to evaluate the probability of such conflict arising by chance. (3) As a comparison of performance KH tests of the null hypotheses against conflicting maximum likelihood phylogenetic results were also performed. (4) For both the parametric bootstrap analyses and KH tests the effect of suboptimal parameter estimates for the model of nucleotide substitution were also evaluated.

DNA sequences

For COI and COII we have analysed 71 sequences [924 base pairs (bp) of COI and 687 bp of COII] for 37 *Calathus* species, and an outgroup, *Calathidius accuminatus* (Emerson *et al.*, 1999, 2000). In the case of 16S a subset of the sequences presented by Gómez-Zurita *et al.* (2000a) from 19 species of the genus *Timarcha* were analysed. Ambiguous sites and sites with insertions or deletions were removed to obtain 34 sequences of length 495. Similarly, for ITS2 we analysed a subset of the sequences presented by Gómez-Zurita *et al.* (2000b) representing 19 *Timarcha* species. After removing ambiguous sites and sites with insertions or deletions, 30 sequences of length 532 were obtained. For both the 16S and ITS2 data sets *T. metallica* was used as an outgroup.

Model selection, phylogenetic analysis, and null hypotheses

For each of the three data sets the fit of sequence data to 56 models of base substitution, ranked in Table 1 roughly according to the complexity of their matrix of transition probability, was tested using Modeltest v.3.0 (Posada & Crandall, 1998). Modeltest uses the log likelihood ratio tests and the Akaike information criterion (AIC) (Akaike, 1973) to determine which of the models best describes the data. For each data set a maximum likelihood tree is constructed with the parameters defined as best describing the data set by Modeltest. We then explore null hypotheses of monophyly for each of these data sets in order to evaluate

Table 1 The fit of each of the 14 models of base substitution listed below to the sequence data was tested with a proportion of invariant sites (I) defined, a gamma correction (Γ) incorporated and both I + Γ included giving a total of 56 models.

Model	Complexity
Jukes Cantor	Equal base frequencies, one substitution rate
F81	Unequal base frequencies, one substitution rate
Kimura 2-parameter	Equal base frequencies, two substitution rates
HKY85	Unequal base frequencies, two substitution rates
Kimura 3-parameter	Equal base frequencies, three substitution rates
Kimura 3-parameter(uf)	Unequal base frequencies, three substitution rates
Tamura Nei(ef)	Equal base frequencies, three substitution rates
Tamura Nei	Unequal base frequencies, three substitution rates
TIM(ef)	Equal base frequencies, four substitution rates
TIM	Unequal base frequencies, four substitution rates
TVM(ef)	Equal base frequencies, five substitution rates
TVM	Unequal base frequencies, five substitution rates
SYM	Equal base frequencies, six substitution rates
GTR	Unequal base frequencies, six substitution rates

empirically the effect of the choice of substitutional model when testing hypotheses of monophyly using the parametric bootstrap.

Calathus COI and COII

For the *Calathus* COI and COII sequence data the general time reversible model of substitution (GTR), Rodríguez *et al.* (1990) fitted the data best with estimated substitution rates of: A–C: 6.278, A–G: 27.22, A–T: 10.51, C–G: 3.897, C–T: 99.55, G–T: 1, the proportion of invariant sites estimated to be 0.5647, a gamma shape parameter of $\Gamma = 0.994$, and, base frequencies A: 0.34, C: 0.10, G: 0.13, T: 0.43. Figure 1 is a ML tree of mtDNA sequence data for the 71 Macaronesian *Calathus* obtained using PAUP* (Sinauer Associates, Sunderland, MA, USA) (Swofford, 1998). Because computational limitations precluded the generation of nonparametric bootstrap values for branches using maximum likelihood, support was assessed with 1000 bootstrap replications using an unweighted maximum parsimony analysis. Maximum parsimony bootstrapping was also used to assess support for nodes for the 16S and ITS2 data sets.

The phylogeny in Fig. 1 is unresolved with regard to a number of the internal branches that are characterized by short length and lack of nonparametric bootstrap support. We consider five null hypotheses of monophyly with regard to island clades: (1) that with the exception of *C. subfuscus* which is clearly related to the continental *C. ambiguus*, the remaining Macaronesian *Calathus* are monophyletic (clades A, B, C, D); (2) that all Canary Island *Calathus* are monophyletic (clades B, C, D); (3) that excluding clade D the Macaronesian *Calathus* are monophyletic (clades A, B, C); that Madeira is either (4) monophyletic with the eastern Canary Islands (clades A, B) or (5) with the main Canary Island clade (clades A, C).

Timarcha 16S

For the *Timarcha* 16S sequence data the transition model of substitution (TIM), Rodríguez *et al.* (1990) fitted the data best with estimated substitution rates of: A–C: 1, A–G: 13.185, A–T: 6.077, C–G: 6.077, C–T: 45.935, G–T: 1, the proportion of invariant sites estimated to be 0.781, a gamma shape parameter of $\Gamma = 1.871$, and, base frequencies A: 0.404, C: 0.128, G: 0.089, T: 0.378. Figure 2a is a ML tree of mtDNA sequence data for the 34 Iberian *Timarcha* obtained using PAUP*. Within the phylogeny the *T. goettingensis* species complex, previously characterized as monophyletic based on cytogenetic studies (Petitpierre, 1970) is not monophyletic but includes the species *T. hispanica*, *T. granadensis* and *Timarcha* sp. Here we test the null hypothesis of monophyly for the *T. goettingensis* species complex.

Timarcha ITS2

For the *Timarcha* ITS2 sequence data the tranversion model of substitution with equal base frequencies (TVMef), Rodríguez *et al.* (1990) fitted the data best with estimated substitution rates of: A–C: 1.404, A–G: 3.193, A–T: 3.787, C–G: 1.067, C–T: 3.193, G–T: 1, the proportion of invariant sites estimated to be 0.361, and a gamma shape parameter of $\Gamma = 0.832$. Figure 2b is a ML tree of mtDNA sequence data for the 30 Iberian *Timarcha* obtained using PAUP*. Again the *T. goettingensis* species complex is not monophyletic but includes the species *T. hispanica*, *T. granadensis* and *Timarcha* sp. Similar to the 16S data set the null hypothesis we test is monophyly for the *T. goettingensis* species complex.

Parametric bootstrap

For a given hypothesis of monophyly, a constrained NJ tree (enforcing monophyly) was constructed from the sequence data in PAUP* using ML distances from the best fit model and the parameter estimates derived from the Modeltest analysis (Figs 3 and 4). Next, for each of the seven hypotheses, 500 replicate DNA sequence data sets

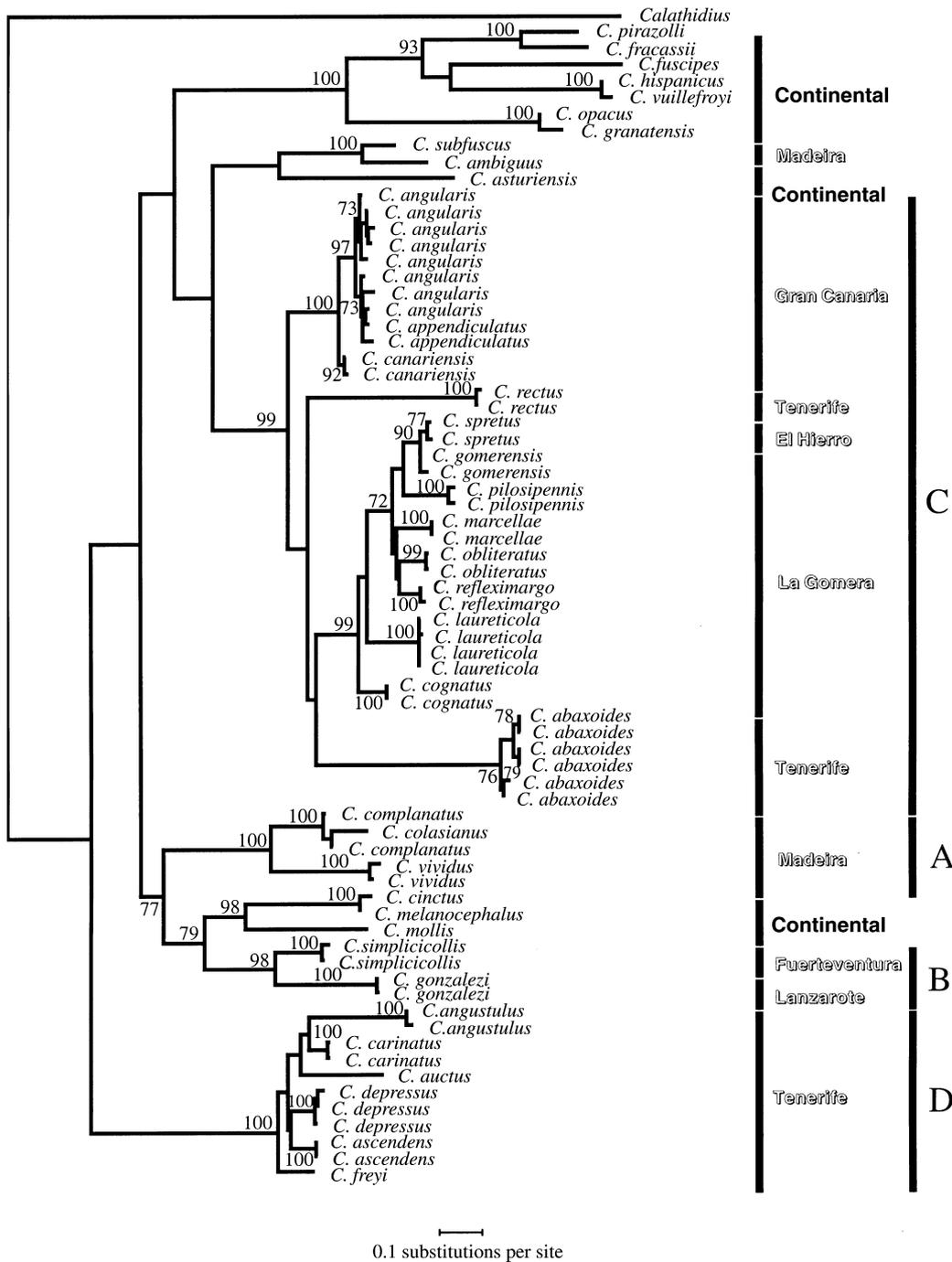


Fig. 1 Maximum likelihood tree for species of *Calathus* using mitochondrial COI and COII sequence data. Major clades of Macaronesian *Calathus* have been assigned the names A, B, C, and D. Bootstrap values are indicated for nodes gaining more than 70% support (1000 replications) from an unweighted maximum parsimony analysis.

were generated using Seq-Gen v.1.1 (Rambaut & Grassly, 1997) which simulates the evolution of DNA sequences along a defined phylogeny under a specified model of the substitution process. Again, the sequences were gener-

ated using the best-fit model of substitution and the parameter estimates derived from the empirical data. However, because of software limitations within Seq-Gen v.1.1 we were unable to generate sequences with a

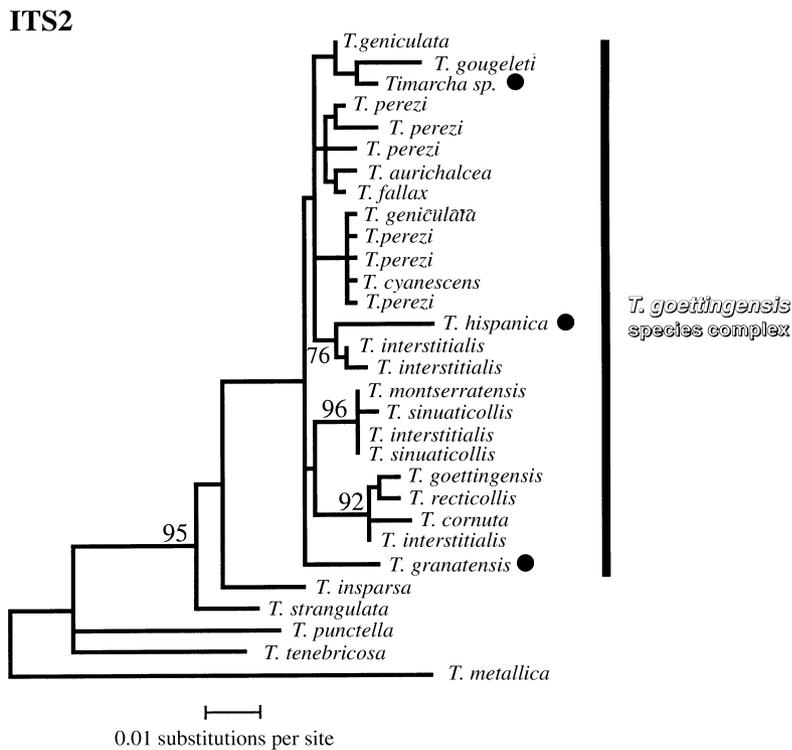
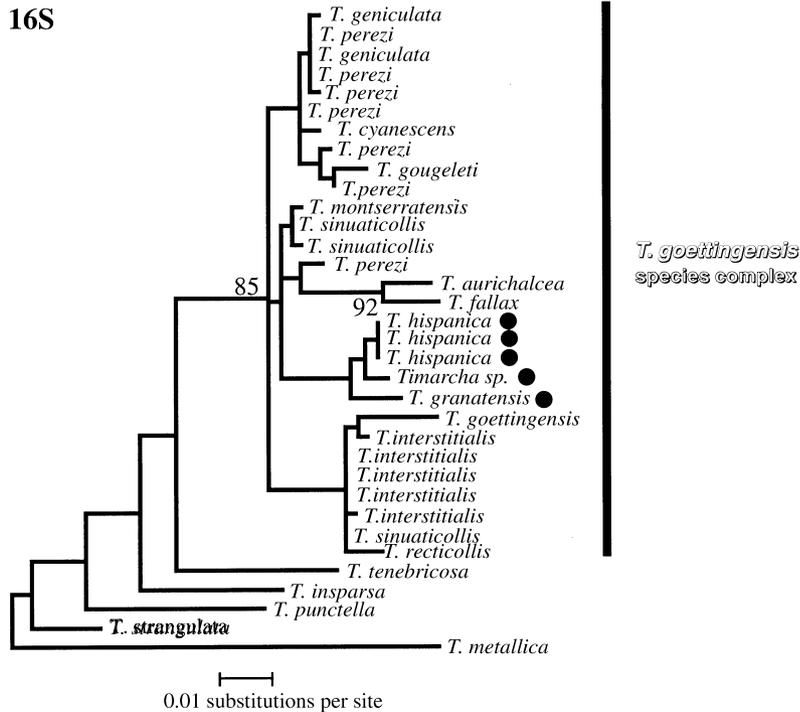


Fig. 2 Maximum likelihood trees for species of *Timarcha* using mitochondrial 16S and nuclear ITS2 sequence data. Bootstrap values are indicated for nodes gaining more than 70% support (1000 replications) from an unweighted maximum parsimony analysis. Species falling within the *T. goettingensis* species complex, but not recognized cytogenetically as belonging to this complex, are marked with black circles.

defined proportion of invariant sites as in the best-fit model identified using Modeltest. Instead, parameters were used from a similar model but without a proportion of invariant sites defined. For each of the null hypotheses,

each of the 500 replicate data sets were subject to heuristic parsimony searches first with and then without the specified constraint. The resulting distribution of differences between the step-length of constrained and

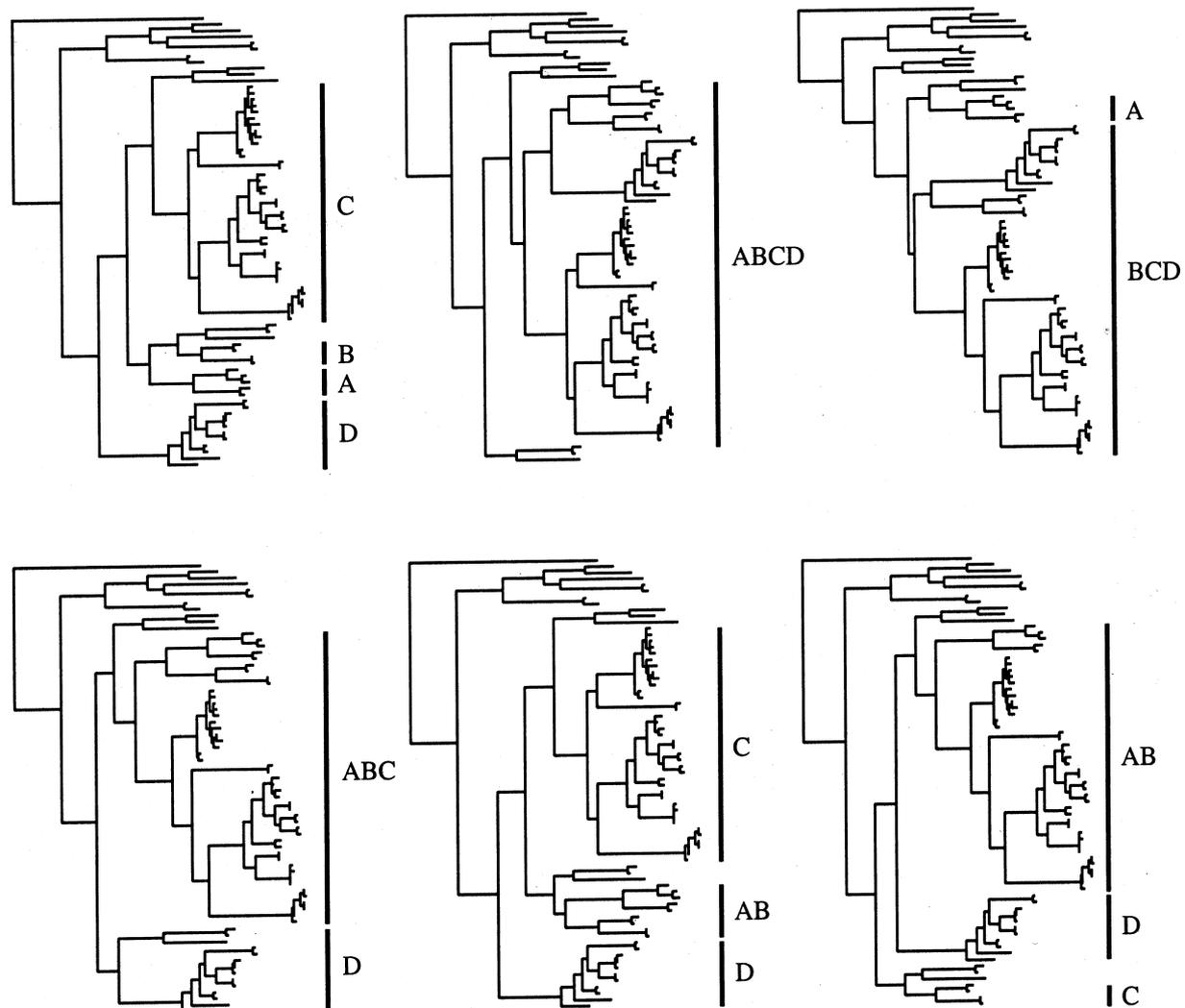


Fig. 3 Constrained trees constructed using neighbour joining with maximum likelihood distances under the GTR + I + Γ substitution model for the *Calathus* COI and COII sequence data. The first tree is the unconstrained tree and the other five represent null hypotheses about monophyly. Unconstrained and constrained trees were also generated under 5 other sub-optimal models (see text for details).

unconstrained trees was then compared with the tree length differences for the empirical constrained and nonconstrained trees. Alternatively we could have constructed ML trees and produced a distribution of likelihood differences but this was not feasible due to computational limitations.

This process was then repeated using the following less adequate substitution models for each of the three sequence datasets. For the *Calathus* COI and COII data: (1) GTR without Γ , (2) HKY85 (Hasegawa *et al.*, 1985), (3) Kimura 2-parameter (Kimura, 1980) with Γ , (4) Kimura 2-parameter without Γ , (5) Jukes-Cantor (Jukes & Cantor, 1969). For the *Timarcha* 16S data: (1) TIM without Γ , (2) HKY85, (3) Kimura 2-parameter with Γ , (4) Kimura 2-parameter without Γ , (5) Jukes-

Cantor. For the *Timarcha* ITS2 data: (1) TVM without Γ , (2) Kimura 3-parameter (Kimura, 1981), (3) Kimura 2-parameter with Γ , (4) Kimura 2-parameter without Γ , (5) Jukes-Cantor. Again, for each of these models, parameter estimates were derived from PAUP*, constrained NJ trees were constructed from the empirical data, sequence data was simulated under the same model, constrained and unconstrained parsimony trees were obtained from heuristic searches, and distributions of differences generated.

Kishino-Hasegawa test

This parametric test evaluates alternative tree topologies by testing if the difference between their log

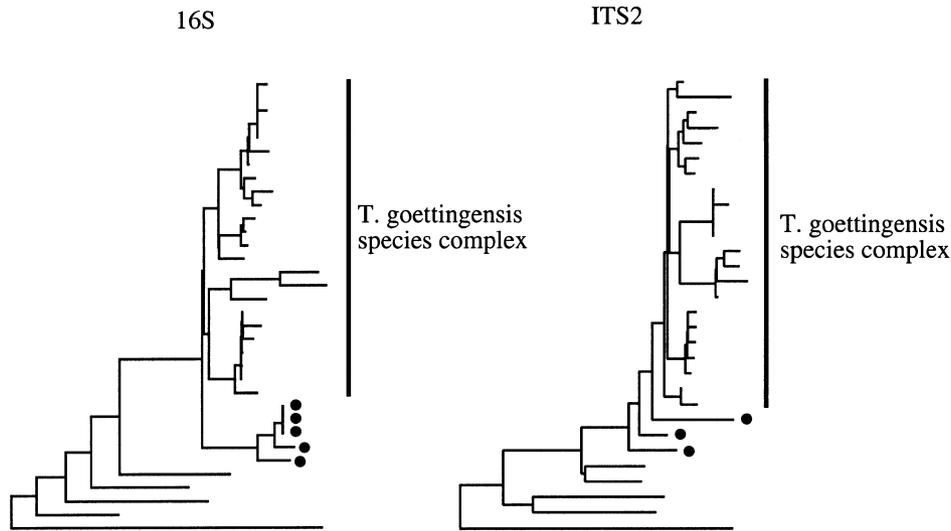


Fig. 4 Constrained trees constructed using neighbour joining with maximum likelihood distances under the TIM + I + Γ substitution model for the *Timarcha* 16S sequence data and the TVMef + I + Γ substitution model for the *Timarcha* ITS2 sequence data. For each of these data sets unconstrained and constrained trees were also generated under five other suboptimal models (see text for details).

likelihood values is statistically significant. This can be carried out either by bootstrap re-sampling which requires a high amount of computation (Hasegawa & Kishino, 1989) or using explicit estimates of the variance of the difference between log likelihoods as proposed in Kishino & Hasegawa (1989). The latter were carried out as implemented in PAUP*. Neighbour joining trees (Saitou & Nei, 1987) were constructed using ML distances for unconstrained and constrained trees using parameter estimates for each of the aforementioned models and each of the five *Calathus* null hypotheses evaluated. The smaller 16S and ITS2 data sets for *Timarcha* meant it was computationally feasible to construct trees under the optimality criterion or maximum likelihood.

Results and discussion

Parametric bootstrap

For the *Calathus* COI and COII data the distribution of differences in step length between hypothesis-constrained and unconstrained trees from the data sets simulated under the best-fit model (GTR + Γ) resulted in the rejection of hypotheses 1, 2 and 3 ($P \leq 0.01$) but hypotheses 4 and 5 could not be rejected at the 5% significance level (Fig. 5, Table 2). Thus, monophyly of Madeira with the eastern Canary Islands (clades A, B) or with the main Canary Island clade (A, C) cannot be excluded. For the *Timarcha* 16S data the hypothesis of monophyly for the *T. goettingensis* complex could not be rejected ($P > 0.25$) but was rejected for the ITS2 data ($P < 0.01$) (Fig. 5, Table 2).

Figure 6 shows the distribution of differences for each of the seven hypotheses when less than adequate substitution models were used both on the empirical data and for generating the replicate sequences using simulation. The significance of each of the null hypotheses under each of the models is summarized in Table 2. All null hypotheses are rejected with all the suboptimal models with the exception of the Kimura 2-parameter + Γ model. For this model of substitution the P -value for null hypothesis 5 (AC monophyly) is 0.14, and that for the null hypothesis of monophyly for the *T. goettingensis* complex using 16S sequences is 0.17.

The simplest of the models, the Jukes Cantor, for which base frequencies are assumed to be equal with only one substitution rate, produces the most left skewed distribution for all seven hypotheses (Fig. 6). The Kimura 2-parameter model is an improvement on the Jukes Cantor model by allowing one substitutional rate for transitions and one for transversions, but this has an apparently minor effect on the distribution of the step length differences. For the *Calathus* COI and COII data and the *Timarcha* 16S data, the HKY85 model further refines the substitutional model by allowing for unequal base frequencies, and both the TIM and GTR models improve on the HKY85 model by allowing for more substitutional rates (4 and 6, respectively). For the *Timarcha* ITS2 data the K3P model is an improvement upon the K2P model by incorporating an additional substitutional rate. The TVMef model further refines this by defining a total of five substitutional rates. These improvements impact on the distribution of expected differences in that the right tail is longer and the left skew is reduced, but again only marginally.

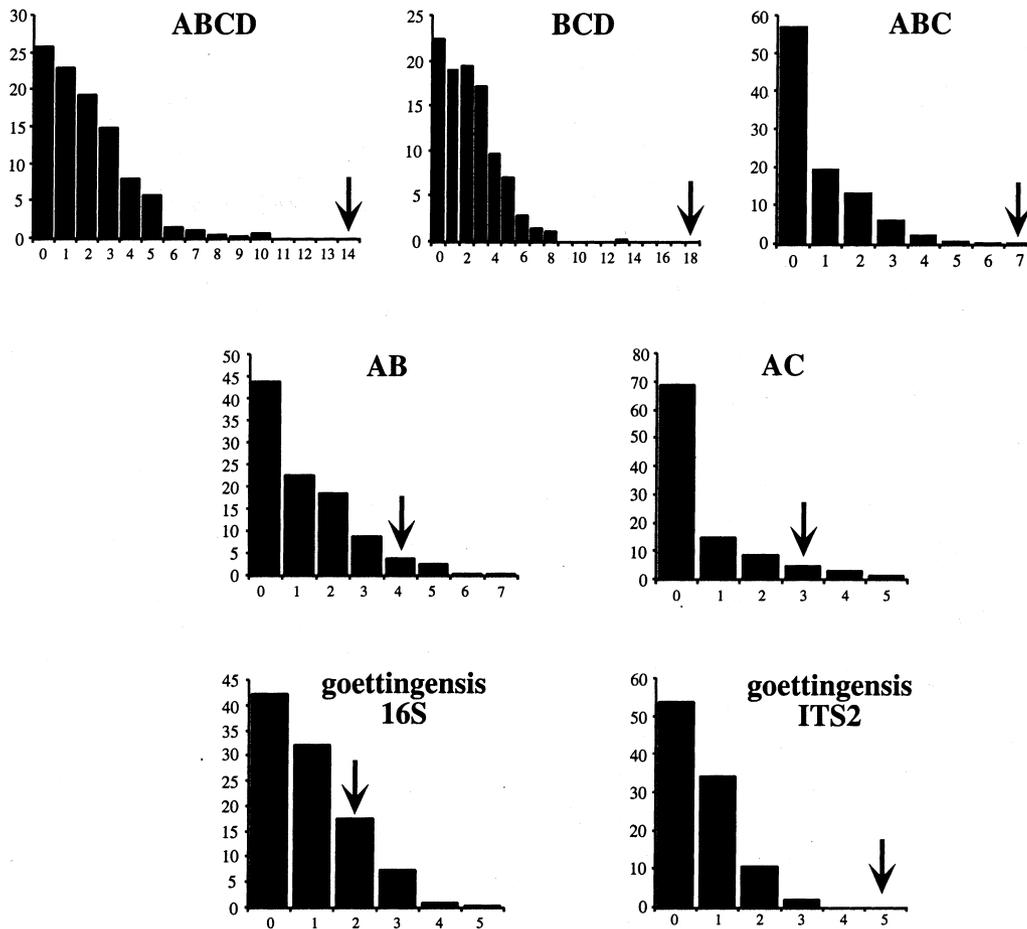


Fig. 5 Results from the parametric bootstrap tests for five hypotheses of monophyly for the Macaronesian *Calathus* clades A, B, C, and D (letters associated with each graph represent the hypothesis tested, see Fig. 1) for a dataset of COI and COII sequence data, and hypotheses of monophyly for the *Timarcha goettingensis* species complex for data sets of 16S and ITS2 sequence data. For each of the null hypotheses DNA sequence data were simulated on a topology constrained to the hypothesis 500 times under the GTR + Γ substitution model for COI and COII, the TIM + Γ substitution model for 16S, and the TVMef + Γ substitution model for ITS2. Arrows indicate the observed difference in steplength for the hypothesis of monophyly for the empirical data.

The addition of an estimate of the gamma shape parameter stands out as being an important factor in generating higher and more frequent step differences between the hypothesis constrained and unconstrained trees. This translates to a higher probability of accepting the null hypotheses, which we shall phrase in terms of failure to reject because the hypotheses were initially framed relative to an empirical tree which, although indicative, does not support them.

The addition of a gamma shape parameter to the GTR model meant the *Calathus* null hypotheses 4 (AB) and 5 (BC) could not be rejected. Adding a gamma shape parameter to the TIM model resulted in the failure to reject the *Timarcha* 16S null hypothesis. Similarly, the addition of a gamma shape parameter to the suboptimal Kimura 2-parameter model results in the inability to reject *Calathus* null hypothesis 4 and the *Timarcha* 16S

null hypothesis. All these null hypotheses are rejected when a gamma shape parameter is not included. Although the gamma shape parameter stands out as the singularly greatest factor contributing to the reduction of the left skew of the distribution of expected differences, it is also clear that additional parameters that apparently cause little effect individually can be important in combination with the gamma shape parameter.

Two conclusions can be drawn from the above observations. First, whether a particular hypothesis of monophyly is rejected or accepted appears to be highly dependent on whether the nucleotide substitution model being used is optimal. Second, as the available substitution models are essentially improvements over a basic model using additional parameters, it seems safe to conclude that a parameter rich model is either equally or

Table 2 Results of parametric bootstrap analyses for hypotheses of monophyly for Macaronesian *Calathus* beetles and for the *Timarcha goettingensis* species complex under different models of sequence evolution. COI and COII sequence data were used for the former, and 16S and ITS2 sequence data in the case of *T. goettingensis*. See Fig. 1 for the five hypotheses of monophyly represented by letter combinations. The *P*-values represent the proportion of the 500 simulated data sets which yielded a length difference equal to or greater than the empirical difference between the hypothesis constrained and unconstrained tree. See text for explanation.

Model	Hypothesis					<i>T. goettingensis</i>	
	ABCD	BCD	ABC	AB	AC	16S	ITS2
Jukes Cantor	$P < 0.01$	$P < 0.01$					
K2P	$P < 0.01$	$P < 0.01$					
K2P + Γ	$P < 0.01$	$P < 0.01$	$P < 0.01$	$P = 0.03$	$P = 0.14$	$P = 0.17$	$P < 0.01$
HKY85	$P < 0.01$	–					
GTR	$P < 0.01$	–	–				
GTR + Γ	$P < 0.01$	$P < 0.01$	$P = 0.01$	$P = 0.07$	$P = 0.08$	–	–
TIM	–	–	–	–	–	$P = 0.02$	–
TIM + Γ	–	–	–	–	–	$P = 0.26$	–
K3P	–	–	–	–	–	–	$P < 0.01$
TVMef	–	–	–	–	–	–	$P < 0.01$
TVMef + Γ	–	–	–	–	–	–	$P < 0.01$

less likely to reject a hypothesis of monophyly where the optimal model is unknown.

Kishino–Hasegawa test

In contrast to the parametric bootstrap, the KH tests did not reject hypothesis 3 (ABC) for the *Calathus* for the two substitution models with Γ correction, GTR + Γ and Kimura 2-parameter + Γ . Hypothesis 4 (AB), which was rejected by the bootstrap test in the case of all suboptimal models, could not be rejected under any of the suboptimal substitution models. Similarly, the hypothesis of monophyly for the *T. goettingensis* complex with the 16S sequence data which was rejected by the bootstrap test in the case of all models lacking the Γ parameter, could not be rejected by any of the suboptimal substitution models including those lacking Γ correction. For hypothesis 5 (AC), only the GTR + Γ and the Kimura 2-parameter + Γ models failed to reject the null hypothesis ($P \geq 0.17$). This and the fact that all substitutional models resulted in the rejection of hypotheses 1 (ABCD), 2 (BCD) and the hypothesis of *T. goettingensis* monophyly with the ITS2 sequence data are consistent with the results of the bootstrap test.

Given that the KH test (Table 3) produced three times as many acceptable model/hypotheses combinations compared with the parametric bootstrap method (Table 2), we conclude that its performance is not as severely affected as the latter by the use of suboptimal models. Furthermore, as all the model/hypothesis combinations rejected by the KH test have also been rejected by the parametric bootstrap but not the vice versa, we conclude that testing hypotheses of monophyly using the parametric bootstrap method is more conservative. By this we mean that the KH test fails to reject hypotheses of

monophyly more often than does the parametric bootstrap. It has recently been pointed out by Goldman *et al.* (2000) that the KH test will give inflated *P*-values unless both topologies being tested are defined *a priori*. However, they observed that a corrected test (Shimodaira & Hasegawa, 1999) also tends to fail to reject a null hypothesis more often than a parametric test.

It has been demonstrated that for a given data set, suboptimal models can often produce the same best supported tree as optimal models (Yang *et al.*, 1994). However, from our analyses it appears that the use of suboptimal models for hypothesis testing using parametric methods can lead to hypothesis rejection, when the use of an optimal model results in failure to reject the hypothesis. In particular the need to account for a gamma distribution of rates over sites appears to be a singularly important parameter to reduce this type I error. This is more apparent for the parametric bootstrap, but also true to some extent for the KH test. For our examples of the KH test, when there is failure to reject the null hypothesis with the optimal model, but with a comparatively low *P*-value (Table 3, hypotheses ABC and AC), suboptimal models not incorporating a Γ parameter lead to hypothesis rejection. However when there is failure to reject the null hypothesis with a comparatively high *P*-value [Table 3, hypotheses AB, and *T. goettingensis* monophyly (16S)] for the optimal model, suboptimal models also result in failure to reject. We have used a broad range of empirical data with a number of relevant hypotheses to address the issue of choice of evolutionary models for hypothesis testing using parametric methods. The general applicability of our conclusions will be enhanced by additional support from a numerical simulation study which we plan to undertake.

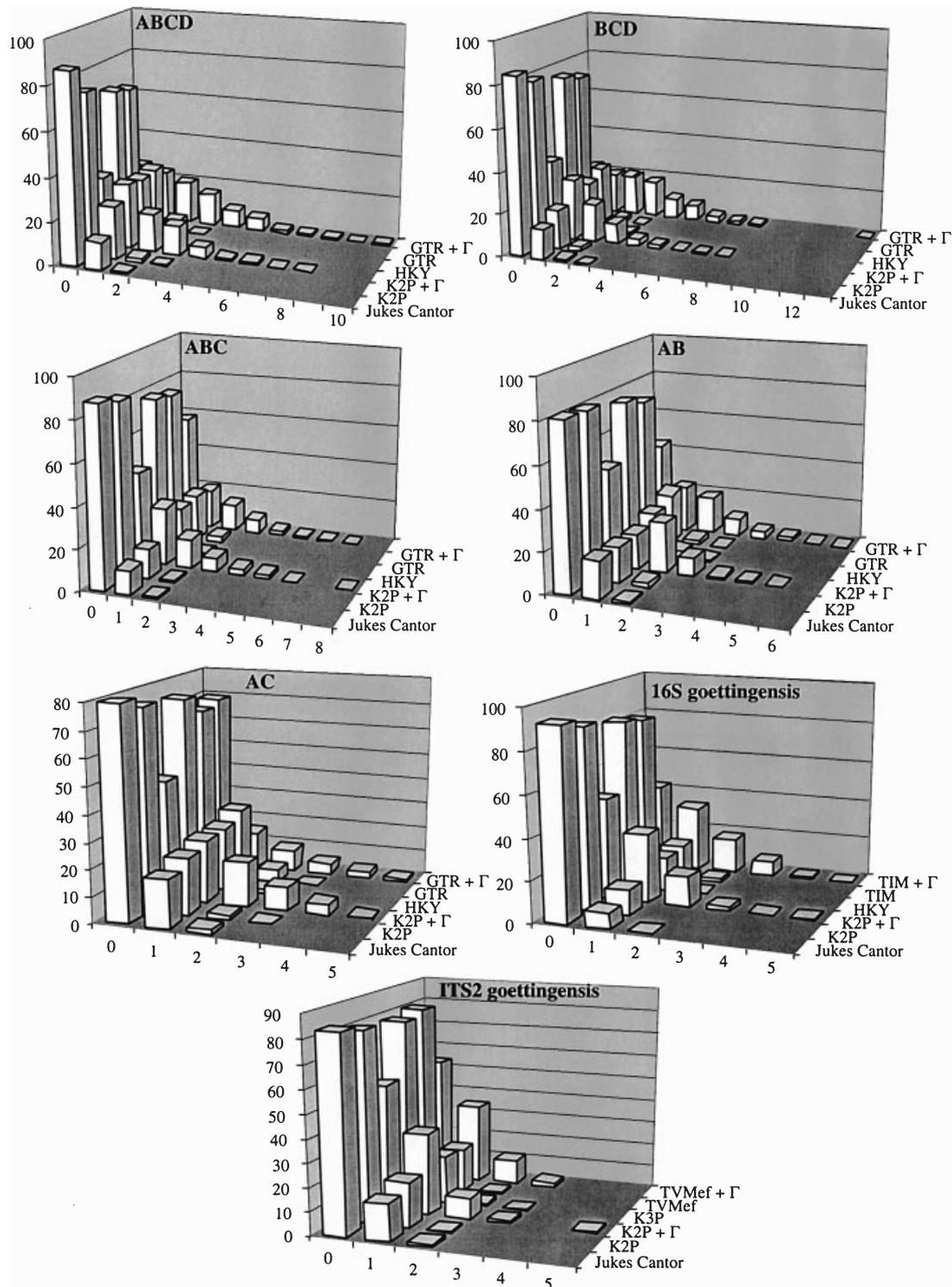


Fig. 6 Distribution of steplength differences from parametric bootstrap analysis of five hypotheses of monophyly for *Calathus* beetles from Macaronesia for a dataset of COI and COII sequence data, and hypotheses of monophyly for the *Timarcha goettingensis* species complex for data sets of 16S and ITS2 sequence data. Hypotheses are tested under six different substitutional models. *x* axis = steplength differences between constrained and unconstrained maximum parsimony analyses for each simulated data set. *y* axis = the frequency of simulated data sets giving a particular steplength difference.

Table 3 Results of the Kishino–Hasegawa test for hypotheses of monophyly for Macaronesian *Calathus* beetles and for the *Timarcha goettingensis* species complex under different models of sequence evolution. COI and COII sequence data were used for the former, and 16S and ITS2 sequence data in the case of *T. goettingensis*. See Fig. 1 for the five hypotheses of monophyly represented by letter combinations. The *P*-values represent the proportion of the 500 simulated data sets which yielded a length difference equal to or greater than the empirical difference between the hypothesis constrained and unconstrained tree. See text for explanation.

Model	Hypothesis					<i>T. goettingensis</i>	
	ABCD	BCD	ABC	AB	AC	16S	ITS2
Jukes Cantor	$P < 0.01$	$P < 0.01$	$P < 0.01$	$P = 0.24$	$P = 0.02$	$P = 0.50$	$P = 0.02$
K2P	$P < 0.01$	$P < 0.01$	$P = 0.01$	$P = 0.31$	$P = 0.03$	$P = 0.84$	$P = 0.03$
K2P + Γ	$P = 0.01$	$P < 0.01$	$P = 0.09$	$P = 0.36$	$P = 0.17$	$P = 0.89$	$P = 0.03$
HKY85	$P < 0.01$	$P < 0.01$	$P = 0.02$	$P = 0.33$	$P = 0.03$	$P = 0.70$	–
GTR	$P < 0.01$	$P < 0.01$	$P = 0.01$	$P = 0.17$	$P = 0.03$	–	–
GTR + Γ	$P = 0.01$	$P = 0.02$	$P = 0.10$	$P = 0.43$	$P = 0.18$	–	–
TIM	–	–	–	–	–	$P = 0.78$	–
TIM + Γ	–	–	–	–	–	$P = 0.57$	–
K3P	–	–	–	–	–	–	$P = 0.03$
TVMef	–	–	–	–	–	–	$P = 0.03$
TVMef + Γ	–	–	–	–	–	–	$P = 0.03$

Acknowledgments

We would like to thank Jesús Gómez-Zurita for kindly providing us with 16S and ITS2 sequence data for *Timarcha*, and two anonymous referees for helpful comments on an earlier draft.

References

- Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In: *Proceedings of the 2nd International Symposium on Information Theory* (B. N. Petrov & F. Csaki, eds), pp. 267–281. Akadémia Kiado, Budapest.
- Bull, J.J., Cunningham, C.W., Molineaux, L.J., Badgett, M.R. & Hillis, D.M. 1993. Experimental molecular evolution of bacteriophage T7. *Evolution* **47**: 993–1007.
- Efron, B. 1982. The jackknife, the bootstrap, and other resampling plans. *CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38*. Society Indust. Appl. Math., Philadelphia.
- Emerson, B.C., Oromí, P. & Hewitt, G.M. 1999. MtDNA phylogeography and recent intra-island diversification of Canary Island *Calathus* beetles (Carabidae). *Mol. Phylogenet. Evol.* **13**: 149–158.
- Emerson, B.C., Oromí, P. & Hewitt, G.M. 2000. Interpreting colonisation of the *Calathus* (Coleoptera: Carabidae) on the Canary Islands and Madeira through the application of the parametric bootstrap. *Evolution* **54**: 2081–2090.
- Felsenstein, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* **27**: 401–410.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.
- Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.* **22**: 521–565.
- Felsenstein, J. & Kishino, H. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. *Syst. Biol.* **42**: 193–200.

- Flook, P.K., Klee, S. & Rowell, C.H.F. 1999. Combined molecular phylogenetic analysis of the Orthoptera (Arthropoda, Insecta) and implications for their higher systematics. *Syst. Biol.* **48**: 233–253.
- Goldman, N. 1993. Statistical tests of models of DNA substitution. *J. Mol. Evol.* **36**: 182–198.
- Goldman, N., Anerson, J.P. & Rodrigo, A.P. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* **49**: 652–670.
- Gómez-Zurita, J., Juan, C. & Pettitpierre, E. 2000a. Sequence, secondary structure and phylogenetic analyses of the ribosomal internal transcribed spacer 2 (ITS2) in the leaf beetle genus *Timarcha* (Coleoptera, Chrysomelidae). *Insect Mol. Biol.* **9**: 591–604.
- Gómez-Zurita, J., Juan, C. & Pettitpierre, E. 2000b. The evolutionary history of the genus *Timarcha* (Coleoptera, Chrysomelidae) inferred from mitochondrial COII gene and partial 16S rDNA sequences. *Mol. Phylogenet. Evol.* **14**: 304–317.
- Halanych, K.M. 1998. Considerations for reconstructing metazoan history: signal, resolution, and hypothesis testing. *Am. Zool.* **38**: 929–941.
- Hasegawa, M. & Kishino, H. 1989. Confidence limits on the maximum-likelihood estimate of the hominoid tree from mitochondrial-DNA sequences. *Evolution* **43**: 672–677.
- Hasegawa, M., Kishino, H. & Yano, T. 1985. Dating the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **21**: 160–174.
- Hendy, M.D. & Penny, D. 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* **38**: 297–309.
- Hillis, D.M. & Bull., J.J. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* **42**: 182–192.
- Hillis, D.M., Mable, B.K. & Moritz, C. 1996. Applications of molecular systematics: the state of the field and a look to the future. In: *Molecular Systematics* (D. M. Hillis, C. Moritz & B. K. Mable, eds), pp. 515–543. Sinauer, Sunderland, MA.
- Huelsbeck, J.P. 1997. Is the Felsenstein zone a fly trap? *Syst. Biol.* **46**: 69–74.

- Huelsenbeck, J.P., Hillis, D.M. & Jones, R. 1996a. Parametric bootstrapping in molecular phylogenetics: applications and performance. In: *Molecular Zoology: Advances Strategies and Protocols* (J. D. Ferraris & S. R. Palumbi, eds), pp. 19–45. Wiley-Liss, New York, NY.
- Huelsenbeck, J.P., Hillis, D.M. & Nielsen, R. 1996b. A likelihood-ratio test of monophyly. *Syst. Biol.* **45**: 546–558.
- Hwang, U.W., Kim, W., Tautz, D. & Freidrich, M. 1998. Molecular phylogenetics at the Felsenstein zone: approaching the Strepsiptera problem using 5.8s and 28s rDNA sequences. *Mol. Phylogenet. Evol.* **9**: 470–480.
- Jackman, T.R., Larson, A., De Queiroz, K. & Losos, J.B. 1999. Phylogenetic relationships and tempo of early diversification in *Anolis* lizards. *Syst. Biol.* **4**: 254–285.
- Jukes, T.H. & Cantor, C.R. 1969. Evolution of protein molecules. In: *Mammalian Protein Metabolism* (H. M. Munro, ed.), pp. 21–132. Academic Press, New York, NY.
- Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**: 111–120.
- Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. USA* **78**: 454–458.
- Kishino, H. & Hasegawa, M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order of Homonoidea. *J. Mol. Evol.* **29**: 170–179.
- Maddison, D.R., Baker, M.D. & Ober, K.A. 1999. Phylogeny of carabid beetles as inferred from 18s ribosomal DNA (Coleoptera: Carabidae). *Syst. Entomol.* **24**: 103–138.
- Mallat, J. & Sullivan, J. 1998. 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. *Mol. Biol. Evol.* **15**: 1706–1718.
- Oakley, T.H. & Phillips, R.B. 1999. Phylogeny of salmonine fishes based on growth hormone introns: Atlantic (*Salmo*) and Pacific (*Oncorhynchus*) salmon are not sister taxa. *Mol. Phylogenet. Evol.* **11**: 381–393.
- Petitpierre, E. 1970. Variaciones morfológicas y de la genitalia en las *Timarcha* Lat. (Col. Chrysomelidae). *P. Inst. Biol. Apl.* **48**: 5–16.
- Posada, D. & Crandall, K.A. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.
- Rambaut, A. & Grassly, N.C. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comp. Appl. Biosci.* **13**: 235–238.
- Rodríguez, F., Oliver, J.L.A., Marín, A. & Medina, J.R. 1990. The general stochastic model of nucleotide substitution. *J. Theoret. Biol.* **142**: 485–501.
- Ruedi, M., Auberson, M. & Savolainen, V. 1998. Biogeography of Sulawesi shrews: testing for their origin with a parametric bootstrap on molecular data. *Mol. Phylogenet. Evol.* **9**: 567–571.
- Saitou, N. & Nei, M. 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- Shimodaira, H. & Hasegawa, M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**: 1114–1116.
- Swofford, D. 1998. *Paup*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sinauer Associates, Sunderland, MA.
- Tang, K.L., Berendzen, P.B., Wiley, E.O., Morrissey, J.F., Winterbottom, R. & Johnson, G.D. 1999. The phylogenetic relationships of the suborder Acanthuroidei (Teleostei: Perciformes) based on molecular and morphological evidence. *Mol. Phylogenet. Evol.* **11**: 415–425.
- Yang, Z., Goldman, N. & Friday, A. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Mol. Biol. Evol.* **11**: 316–324.
- Zhang, J. 1999. Performance of likelihood ratio tests of evolutionary hypotheses under inadequate substitution models. *Mol. Biol. Evol.* **16**: 868–875.

Received 11 April 2001; accepted 17 April 2001