Bayesian Tests of Topology Hypotheses with an Example from Diving Beetles

JOHANNES BERGSTEN1,*, ANDERS N. NILSSON2, AND FREDRIK RONQUIST3

1Department of Entomology, Swedish Museum of Natural History, Box 50007, SE-104 05 Stockholm, Sweden; 2Department of Ecology and Environmental Science, Umeå University, SE-90187 Umeå, Sweden; and 3Department of Biodiversity Informatics, Swedish Museum of Natural History, Box 50007, SE-104 05 Stockholm, Sweden

*Correspondence to be sent to: Department of Entomology, Swedish Museum of Natural History, Box 50007, SE-104 05 Stockholm, Sweden; E-mail: johannes.bergstesen@nrm.se.

Received 30 October 2012; reviews returned 16 January 2013; accepted 23 April 2013

Abstract.—We review Bayesian approaches to model testing in general and to the assessment of topological hypotheses in particular. We show that the standard way of setting up Bayes factor tests of the monophyly of a group, or the placement of a sample sequence in a known reference tree, can be misleading. The reason for this is related to the well-known dependency of Bayes factors on model-specific priors. Specifically, when testing tree hypotheses it is important that each hypothesis is associated with an appropriate tree space in the prior. This can be achieved by using appropriately constrained searches or by filtering trees in the posterior sample, but in a more elaborate way than typically implemented. If it is difficult to find the appropriate tree sets to be contrasted, then the posterior model odds may be more informative than the Bayes factor. We illustrate the recommended techniques using an empirical test case addressing the issue of whether two genera of diving beetles (Coleoptera: Dytiscidae), Suphrodytes and Hydroporus, should be synonymized. Our refined Bayes factor tests, in contrast to standard analyses, show that there is strong support for Suphrodytes nesting inside Hydroporus, and the genera are therefore synonymized. [Bayes factors; Coleoptera; Dytiscidae; marginal likelihood; model testing; posterior odds; reversible-jump MCMC, stepping-stone sampling.]

Testing hypotheses about the structure of trees is one of the most fundamental tasks in phylogenetic systematics. It helps answer questions such as: Does an unknown sample X belong to a known reference group A? Can two taxa A and B be maintained as separate or is one nested inside the other? Is clade A more closely related to clade B than to clade C? Is taxon A monophyletic?

Bayesian statistics includes several tools for comparing and testing models, and they have been applied to the testing of tree hypotheses in a number of recent papers (e.g., Lavoué et al. 2007; Marek and Bond 2007; Parker et al. 2007; Azuma et al. 2008; Rabeling et al. 2008; Yamanoue et al. 2008; Pavlicev et al. 2009; Tank and Olmstead 2009; Zakharov et al. 2009; Makowsky et al. 2010; Yang et al. 2010; Kelly et al. 2011; Knight et al. 2011; Schweizer et al. 2011; Drummond et al. 2012; Miller and Bergsten 2012). However, it is quite difficult to formulate relevant hypotheses about tree structure, and Bayes factor tests, as commonly applied in the literature, can be quite misleading. The purpose of this article is to discuss the Bayesian techniques for testing tree hypotheses, point out potential pitfalls, and provide general recommendations for empiricists. We illustrate both the pitfalls and the recommended techniques in a test of whether two genera of diving beetles (Coleoptera: Dytiscidae), Suphrodytes Gozis and Hydroporus Clairville, can be maintained as separate genera or should be synonymized.

THEORY

Bayesian Model Testing

We will assume that the reader is familiar with the basics of Bayesian phylogenetic inference (see, e.g., Holder and Lewis 2003; Yang 2006; Ronquist and D eas 2010). For an introduction to statistical testing of phylogenetic models in general, see Sullivan and Joyce (2005). The theory of Bayesian model testing was worked out primarily by Jeffreys (1935, 1961). An excellent summary of the early work on Bayesian model testing is provided by Kass and Raftery (1995), who also laid the foundation for much of the modern work in the field.

In Bayesian inference we are interested in the posterior probability of model parameters ($\theta$) given some observed data ($D$) and a model ($M$). The posterior probability $f(\theta|M)$ is obtained from the prior probability $f(\theta)$ of the parameter values, and the probability of the data given the parameter values and the model $f(D|\theta,M)$ (also called the “likelihood”), by Bayes’ theorem

$$f(\theta|M) = \frac{f(\theta|M)f(D|\theta,M)}{\int f(\theta|M)f(D|\theta,M)d\theta}.$$  

The normalizing constant in Bayes’ theorem, $f(D|M)$, is the marginal likelihood of the data. It is also the predictive probability of the data: the probability of seeing the observed data calculated from the model before the data are taken into account (Kass and Raftery 1995). Sometimes it is called the marginal [model] likelihood; we will refer to it simply as the model likelihood.

It is this model likelihood, which is used in Bayesian model comparison. Assume we have two models, $M_1$ and $M_2$. If we specified the prior probabilities of the two models, $f(M_1)$ and $f(M_2)$, we could calculate the posterior odds of the models as

$$\frac{f(M_1|D)}{f(M_2|D)} = \frac{f(M_1)}{f(M_2)} \times \frac{f(D|M_1)}{f(D|M_2)}.$$
is natural to focus on the marginal probabilities of the models in predicting the data (Kass and Raftery 1995; Xie et al. 2011) albeit at a significantly increased computational cost. More accurate estimates of the model likelihood can be obtained by thermodynamic integration (Lartillot and Philippe 2006), also known as path sampling (Baele et al. 2012), or the so-called stepping-stone method (Fan et al. 2011; Xie et al. 2011) albeit at a significantly increased computational cost.

Bayesian Tests of Tree Hypotheses

In Bayesian phylogenetic inference, the tree is often viewed as a discrete parameter, which can take on many different values, one for each tree topology. However, mapping parameters associated with nodes or branches, such as branch lengths, from one tree to another is not straightforward. In some sense, all nodes and branches of two distinct tree topologies are different regardless of

| Bayes factor | $2 \times \log_{e} BF$ | Interpretation |
|-------------|----------------------|----------------|
| 1–3         | 0–2                  | Not worth more than a bare mention |
| 3–20        | 2–6                  | Positive |
| 20–150      | 6–10                 | Strong |
| $> 150$     | $> 10$               | Very strong |

The last ratio is the ratio of the model likelihoods, also known as the Bayes factor. We could thus specify the same equation in words as

posterior model odds $= \text{prior model odds} \times \text{Bayes factor}.$

Both posterior model odds and Bayes factors are used to compare models in Bayesian inference. When two or more models are considered in the same analysis, it is natural to focus on the marginal probabilities of the models, $f(M_i | D)$, in the posterior of the supermodel analysis. This is how you would approach inference about any other discrete parameter in the analysis. The posterior model odds are simply the ratio between the marginal model probabilities.

A possible disadvantage of posterior model odds is that they depend on the prior on models. We can get rid of this dependency by focusing on the Bayes factor. Even the Bayes factor, however, is sensitive to the priors on model-specific parameters (as opposed to the prior probabilities of the models themselves). For instance, if one model is a special case of the other, then the Bayes factor is dependent on how focused the prior is on the larger parameter space (Kass and Raftery 1995 and references cited therein). It is an analogous effect that can cause problems with Bayes factor tests of topological hypotheses, as we will discover.

One way of understanding the Bayes factor is that it measures the strength of the data in changing the prior model odds. Alternatively, because the model likelihood is the predictive probability of the data, we can also view the Bayes factor as measuring the relative success of the models in predicting the data (Kass and Raftery 1995). The Bayes factor is closely related to the likelihood ratio statistic: change the integrated likelihoods to maximized likelihood and you essentially have the latter. Unlike the likelihood ratio statistic, however, Bayes factors can be used on any pair of models, regardless of whether or not they are nested. The former statistic is based on a chi-square distribution, taking the number of parameters into account through the degrees of freedom. Bayes factors, in contrast, have a natural way of penalizing overfitting, “a fully automatic Occam’s razor” (Kass and Raftery 1995).

Based on the similarity to the likelihood ratio statistic, general guidelines to the interpretation of Bayes factors were suggested by Kass and Raftery (1995). Specifically, they indicated that a Bayes factor larger than 3 should be interpreted as significant positive evidence in favor of the better model. A scale similar to the likelihood ratio statistic is obtained by taking twice the logarithm of the Bayes factor; on this scale, the critical value is 2 (Table 1).

### Computational Challenges

There are two main approaches to estimating posterior model odds and Bayes factors. The first one is to use a single Markov chain Monte Carlo (MCMC) analysis to estimate the posterior model probabilities (for phylogenetic examples, see Suchard et al. 2001, 2005; Huelsenbeck et al. 2004). If the models have a different number of dimensions, or include different parameters, then one has to implement reversible-jump MCMC, which is technically demanding. It can also be difficult to obtain good mixing across complex model spaces. Despite the difficulties, this is often the best approach for estimating moderate to large model probabilities. Small model probabilities, however, are difficult to estimate precisely. This means that more extreme posterior model odds are difficult to assess reliably. If the aim is to estimate Bayes factors, and the analysis uses prior model odds that are strongly biased, the results can be severely compromised as will be shown below. A possible solution is to modify the prior probabilities of models until the models of interest receive similar posterior probabilities. Bayes factors or posterior model odds can then be computed by taking the modified prior probabilities into account. Suchard et al. (2005) developed this approach further, referring to it as “Bayes factor titration”.

The other approach involves separate estimation of each of the model likelihoods. This means running a complete MCMC simulation on each model, so it can only be successfully applied to a small number of models. In addition, it is extremely difficult to estimate the model likelihood, which is typically a sum and integral over a large and complex parameter space. The most commonly used estimator is the harmonic mean estimator (HME) (Newton and Raftery 1994), which is simply the harmonic mean of the likelihoods of the MCMC output. The HME is sensitive to the inclusion of rare values with large influence on the estimate, and it tends to overestimate the model likelihood.

Bayesian Tests of Tree Hypotheses

In Bayesian phylogenetic inference, the tree is often viewed as a discrete parameter, which can take on many different values, one for each tree topology. However, mapping parameters associated with nodes or branches, such as branch lengths, from one tree to another is not straightforward. In some sense, all nodes and branches of two distinct tree topologies are different regardless of
how similar the topologies are or how many bipartitions they share.

Inspired by this, some workers consider each topology as a separate model (e.g., Suchard et al. 2001, 2005; Yang 2006) instead of treating the topology as an ordinary discrete model parameter. This leads to the view that a typical Bayesian phylogenetic analysis is a case of reversible-jump MCMC across topology model space. This is the view we will adopt here. If we only consider fully resolved trees, then the dimension of each model is the same but the node and branch length parameters are not identical. Thus, each jump between models amounts to discarding some parameters (dimensions) and adding in an equal number of other parameters (dimensions).

From an ordinary Bayesian phylogenetic analysis, then, we get the marginal posterior probabilities of the topology models. To get the posterior model odds for two distinct topologies, we simply divide the frequency of one tree with the frequency of the other in our sample of the posterior.

In inference of non-clock trees, it is standard procedure to associate each distinct topology with the same prior probability, in which case the posterior odds are the same as the Bayes factor. For clock trees, the standard models (birth−death, coalescent, uniform) typically put equal prior probability on unique labeled histories rather than on unique topologies. To calculate the Bayes factor for two distinct rooted topologies then, one has to take into account that they may be compatible with different numbers of labeled histories (Felsenstein 2004). For a simple example, consider a four-taxon clock tree. There are two histories compatible with a symmetric topology, depending on which speciation event happened first. In contrast, there is only one possible sequence of speciation events in an asymmetric, comb-like topology. If the prior puts equal probability on all speciation histories, then symmetric topologies are twice as likely to be present as asymmetric ones. To obtain Bayes factors from a posterior sample of trees under these conditions, simply divide the frequency of symmetric topologies with two to correct for the topology bias in the prior, and then use the ratio between the corrected frequencies. Except for this correction, topological hypothesis tests on clock trees behave essentially the same as tests on non-clock trees.

An important problem is that topology space is huge for even moderate numbers of taxa, and the posterior probabilities of an individual topology is likely to be small and difficult to estimate accurately. Even the best topologies may individually have low posterior probability in a large analysis. It is arguably easier to estimate the posterior probabilities of classes of trees than it is to assess the probability of individual topologies.

Focusing on classes of trees leads to other problems, however (for a discussion of similar problems from a different perspective, see Wheeler and Pickett 2008). When using Bayes factors, for instance, the hypotheses of interest are often associated with very different numbers of topologies (see below). In fact, the vastness of tree space often produces such extreme prior odds in standard analyses that it becomes hopeless in many cases to estimate Bayes factors from the posterior model odds. For instance, the prior odds might be so biased in favor of a particular clade being monophyletic that it would take an enormously large MCMC sample to show that the posterior odds are higher than the prior odds. One possibility is then to restructure the prior by introducing partition-associated probability factors and then do Bayes factor titration (Suchard et al. 2005; see also Ronquist et al. 2004).

A more serious problem, however, is to identify the relevant classes of trees to be contrasted. This is surprisingly difficult and can lead to counter-intuitive results, especially when using Bayes factors. In fact, we argue here that the Bayes factor test of topological hypotheses used today by most phylogeneticists (in the following referred to as "standard Bayes factor tests") should be abandoned in favor of more appropriate techniques. The problem is related to the sensitivity of Bayes factors to model-specific priors, and will be illustrated here with two hypothetical examples.

Consider first the problem of determining whether or not a sample sequence X belongs to a known group A, a problem analysed by Suchard et al. (2005). Assume for simplicity that the monophyly of A (without considering X) is overwhelmingly supported by the data. Further assume that there is no evidence allowing us to place X; all placements of X with different likelihood. Under these circumstances, the Bayes factor should be 1, that is, there should be no evidence for or against the hypothesis that X belongs to A compared to the hypothesis that X does not belong to A.

The standard approach explored by Suchard et al. (2005) would be to test the hypothesis that X + A form a monophyletic group (H1) against the hypothesis that they do not (H0). This will likely produce misleading results, as we will see in the following. In the tree without X, let a be the number of taxa in A, and n the number of taxa in total. If A is a non-trivial group in the tree without X, then we have a ≤ n − 2 (group A excludes at least two taxa). Further, let B(n) be the number of unrooted, fully resolved trees for n taxa. If the prior associates each tree with the same prior probability, the prior odds for H1 against H0 would be

\[
\frac{B(n-a+1)B(a+2)/B(n+1)}{1-B(n-a+1)B(a+2)/B(n+1)}.
\]

For large n, the denominator is close to 1, so this ratio is very nearly equal to

\[
\frac{B(n-a+1)B(a+2)}{B(n+1)}.
\]

In the posterior, we expect all or almost all trees to have A as monophyletic. Given that this is true, and that X can be placed with equal probability on every branch in the tree, the posterior odds of X being placed inside A whereas it being placed outside is simply the ratio between the number of branches in A and the number...
of branches outside A,

\[
\frac{2a - 2}{2(n - a) - 1}
\]

Since the Bayes factor is the posterior odds divided by the prior odds, we obtain the Bayes factor as

\[
\frac{B(n + 1)(2a - 2)}{B(n - a + 1)B(a + 2)(2(n - a) - 1)}
\]

This is guaranteed to be a large number, especially for large \( n \), suggesting that there is significant support for \( H_0 \). The Bayes factor is misleading. The reason for this counter-intuitive result is that the Bayes factor reflects both the strength of evidence in favor of placing \( X \) in A, and the evidence in favor of monophyly of A regardless of the position of \( X \). Given the way the problem is formulated, we should take strong evidence for the monophyly of A for granted, and account for that in the hypothesis test.

Because such background information is often reflected in the posterior, the posterior odds can be more informative than Bayes factors calculated in the standard way. For instance, in this particular case, the posterior odds for \( X \) grouping with A would be

\[
\frac{2a - 2}{2(n - a) - 1}
\]

This will not deviate much from 1, so the posterior odds will suggest that there is little evidence for placing \( X \) inside A. However, the odds will be exactly equal to 1 only when there are as many tree tips inside as outside of \( A \) (\( a = n - a \), and assuming that placement as sister to A is indecisive).

To calculate a meaningful Bayes factor for this problem, we need to restrict our attention to trees that have A monophyletic after pruning of \( X \). In the best case, all the trees in the posterior sample from an unconstrained analysis will satisfy this constraint, in which case the posterior tree sample can be used directly for the calculation of Bayes factors. If only some trees in the posterior sample violate the constraint, we can filter them out, for instance using a backbone constraint in PAUP, and then use the remaining trees for hypothesis testing. One can also run an MCMC analysis using a backbone constraint forcing \( A \) to be monophyletic regardless of the position of \( X \), as permitted, for example, in the most recent version of MrBayes (Ronquist et al. 2012), and then use this tree sample for model testing.

An alternative, but likely less accurate method, is to estimate the model likelihood separately for the two hypotheses. One would then run an analysis with \( A \) constrained to be monophyletic without \( X \). The other analysis would have \( A+X \) constrained to be monophyletic, and \( A \) without \( X \) constrained not to be monophyletic. The latter constraint is important if \( X \) being the sister of \( A \) is not considered part of the tree set relevant for \( H_1 \).

If we restrict our attention to the backbone trees with \( A \) monophyletic, then the prior odds for \( X \) being placed inside \( A \) would be

\[
\frac{2a - 2}{2(n - a) - 1}
\]

This is the same as the posterior odds, and thus gives the expected Bayes factor of 1. Essentially, what we have done is to take the background information that \( A \) (with or without \( X \)) is likely to be monophyletic into account in the topology prior, making the Bayes factor test informative rather than misleading. This illustrates how sensitive Bayes factor tests of topology hypotheses are to the specification of the topology prior, unlike posterior model odds. Unfortunately, the standard flat prior on topologies (or histories) is rarely appropriate for Bayes factor tests of topological hypotheses.

For a second example of the problems with standard Bayes factor tests of topological hypotheses, consider the very common question of whether a group \( A \) is monophyletic. Also here it is easy to get contradictory results if one does not consider the relevant tree classes carefully. This can be illustrated with a simple example (Fig. 1). Assume that \( A \) consists of two subgroups, \( A_1 \) and \( A_2 \), which are both overwhelmingly supported as monophyletic, while there is no evidence concerning the relationships in the rest of the tree. The hypothesis test should then tell us that there is no evidence for or against the hypothesis that \( A \) is monophyletic.

We first consider the standard hypothesis test where we contrast \( A \) being monophyletic (\( H_1 \)) with \( A \) not being monophyletic (\( H_0 \)). As before, we have \( n \) taxa in \( A_1 + A_2 \), \( n \) taxa in the tree in total and \( a \leq n - 2 \). The prior odds for \( A \) being monophyletic are very nearly

\[
\frac{B(n - a + 1)B(a + 1)}{B(n)}
\]

which is a very small number if \( n \) is large. If we consider the tree outside of \( A_1 \) and \( A_2 \), there will be \( n-a+2 \) tips in it (two tips representing \( A_1 \) and \( A_2 \), respectively). There will be \( B(n-a+2) \) equally well-supported trees, only \( B(n-a+1) \) of which will have \( A \) monophyletic. The posterior odds for \( A \) being monophyletic will then be

\[
\frac{1}{2(n-a+1)-3}
\]

since we know in general that \( B(n+1)=(2n-3)B(n) \). This gives a Bayes factor of

\[
\frac{B(n)}{B(n-a+1)B(a+1)(2(n-a+1)-3)} = \frac{B(n-a+2)B(a+1)}{B(n-a+2)B(a+1)}
\]

which is guaranteed to be a large number for large \( n \). Again, the standard Bayes factor test is misleading, and just using the posterior odds as a guide is more informative. Even better is to take the monophyly of \( A_1 \) and \( A_2 \) into account in both \( H_1 \) and \( H_0 \), in which case we get the expected Bayes factor of 1 (Fig. 1).

What if there is also structure in the outgroup? Let the outgroup consist of two groups \( B_1 \) and \( B_2 \) with a total of \( b \) species. Furthermore, assume that there is no information about the structure of the tree except...
that $A_1$, $A_2$, $B_1$, and $B_2$ are each strongly supported as monophyletic. The prior odds for $A$ being monophyletic are the same but the posterior odds are going to be much higher (less extreme) because there are more constraints on outgroup relationships. It could be as high as $1/3$, if $n = a + b$. Specifically, the Bayes factor is going to be

$$B(n) = B(n-a+1)B(a+1)(2(n-a-b+3)-3).$$

Thus, the effect is to make the Bayes factor larger, resulting in the standard test being even more misleading.

We now turn our attention to a specific empirical case that illustrates the difficulties described above. In the discussion, we return to how these observations can be used to provide general recommendations for Bayesian tests of tree hypotheses.

**Empirical Example**

The Northern Hemisphere genus *Hydroporus* is with its over 180 known valid species dominating the tribe Hydroporini. Most species of the genus inhabit boreal and arctic wetlands where they form an important part of the fauna as predators of zooplankton and benthic insect larvae, especially in the shallower parts of more temporary waters. The delimitation of the genus relative to a number of other smaller genera of the tribe remains problematic, as well as the recognition of subgenera within the genus, at present abandoned by Angus (1985), who stressed the isolated position of *dorsalis*, adding to characters from the genital tract, and changed the status of *Suphrodytes* to a monobasic genus. This view was to be accepted by most subsequent authors publishing on European Hydroporini (Nieuwerkerk 1992; Nilsson and Holmen 1995; Nilsson 2001, 2003).

More recently, the molecular phylogenetic analysis of Ribera et al. (2003, 2008) and Hernando et al. (2012) suggested that *Suphrodytes dorsalis* was nested within *Hydroporus*, providing evidence in favor of a synonymization of *Suphrodytes* with *Hydroporus*. However, Ribera et al. (2003) and Hernando et al. (2012) only used mitochondrial markers and support for the conclusion was weak or algorithm dependent and hence no formal change to the classification was made by either. (Ribera et al.’s, 2008) analysis used both nuclear and mitochondrial genes but focused on the higher-level relationships within the family Dytiscidae and the sample was too sparse at the species group level to make any definite conclusions regarding *Suphrodytes* and *Hydroporus*. Only four of the 29 species groups of

**FIGURE 1.** Schematic illustration of the Bayes factor test of the hypothesis that $A$ is a monophyletic group ($H_1$) against the hypothesis that it is not ($H_0$). We assume that $A$ consists of two strongly supported subclades, $A_1$ and $A_2$, and that the rest of the tree is unresolved. Although $H_0$ is a better explanation of the data, there is no evidence that $A_1$ and $A_2$ together form a monophyletic group. The Bayes factor compares the average height of the posterior over the prior tree space of each hypothesis. When using a standard $H_0$, the signal is spread over a large tree space, and the test suggests that $H_1$ is strongly supported. If we use an informed prior for $H_0$ by restricting the tree space to those trees that have $A_1$ and $A_2$ monophyletic, then the Bayes factor correctly identifies that there is no support for or against $H_1$ over $H_0$, as the average height is the same. A posterior odds test compares the total probability mass under each hypothesis (the area under the posterior distribution), and is therefore not as strongly influenced by the size of the tree space as the Bayes factor.
Hydroporus were represented in that study. As taxonomic sampling is imperative for accurate phylogenetic estimation (Zwickl and Hillis 2002; Heath et al. 2008) solving the Suphrodytes/Hydroporus controversy will require a study focused on the Hydroporus group of genera, a much broader sample of Hydroporus species groups, both nuclear and mitochondrial genes and explicit topological hypothesis testing.

**Approach Taken Here**

In this study we expand the four-gene dataset of Ribera et al. (2008) to include representatives of >75% of species groups in Hydroporus for the focused question of the phylogenetic placement of Suphrodytes in relation to Hydroporus. In a Bayesian hypothesis-testing framework we use a recently improved method of marginal likelihood estimation to explicitly test competing topologies using the Bayes factor. We show that the standard approach to calculating the Bayes factor results in the opposite conclusion compared to our preferred approach outlined above. We also calculate the posterior model odds for the alternative hypotheses by filtering the tree sample from the MCMC analysis, which supports the conclusion of our preferred Bayes factor approach. A common practice to estimate the marginal likelihood of the model (Lartillot and Philippe 2003). However it is known that the harmonic mean is a poor estimator and substantially overestimates the marginal likelihood of the model (Lartillot and Philippe 2003; Xie et al. 2011). Instead we used the stepping-stone sampling method recently described by Xie et al. (2011), which has been shown to be significantly more accurate than the harmonic mean method.

**MATERIALS AND METHODS**

**Taxon Sampling**

We restricted the taxon sampling to the Hydroporus group of genera as defined by Ribera et al. (2008) in their Figure 3 labeled as node number 26 of well-supported nodes. This included apart from Suphrodytes and Hydroporus also Neoporus Guignot, Heterosternula Strand, Hydrocolus, and Sanfilippodytes Francisco. Lo. From Ribera et al.’s analysis we consider it unambiguous that Suphrodytes belong nowhere else outside this group of genera and a test of its placement can be restricted to this subgroup within Hydroporinae. As outgroups to root the tree we used Anax insignis Sharp, Hyphydrus oratus (Linnaeus), Hexaephydrus minutissimus (Régimbart), Canthyporus hottomottos (Geeninger and Harold) and Laccornellus copelatoides (Sharp) which were part of the two closest clades to the Hydroporus group in Ribera et al. (2008). Within Suphrodytes we sampled the two existing species, *S. dorsalis* and *Suphrodytes figuratus* (Gyllenhal) after a recent treatment showing the genus is not monotypic as previously believed (Bergsten et al. 2012). Within Hydroporus we sampled representatives of as many additional species groups as possible to the ones included in Ribera et al. (2008), in particular some of the large-bodied species historically associated with *Suphrodytes*. This resulted in adding 19 new groups to Ribera et al.’s four to a total of 23 of the 29 presently recognized species groups being represented in the final data matrix (Appendix).

**DNA Extraction, PCR, and Sequencing**

For most species DNA was extracted from the head and prothorax of 96-well Wizard SV plates following the manufacturer’s instructions (Promega). Additional additional species were extracted from one hindleg with a GeneMole automated extraction robot (Mole Genetics AS). We targeted two nuclear (18S and H3) and two mitochondrial (CO1 and 16S) genes for the analysis using the primers listed in Supplementary Table S1 (doi:10.5061/dryad.s631d). For H3, 16 S, and 18S we used Ready-To-Go™PCR beads (Amersham Biosciences) in a 25-ul reaction volume with a 0.4 uM concentration of each primer and 2 ul of DNA (unknown concentration). Cycling conditions started with a 5 min denaturation step at 95°C followed by 40 cycles of 30 s at 95°C, 30 s at 50°C, and 60 s at 72°C, followed by a final extension step of 8 min at 72°C. Most CO1 sequences were amplified with Bioline Taq instead of beads but with the same primer and DNA concentration. Cycling condition for CO1 was 94°C for 2 min, 35–40 cycles of 94°C for 30 s, 51–53°C for 60 s and 70°C for 90–120 s, and a final extension of 70°C for 10 min. Most PCR products were purified with Exonuclease I and FastAP (Fermentas) in the proportion 1:4, and sequenced with a BigDye™Terminator ver. 3.1 Cycle Sequencing Kit (Applied Biosystems), cleaned with a DyeEx 96 kit (Qiagen) and run on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems). Some CO1 products were cleaned with a 96-well Millipore multiscreen plate and sequenced in both directions using a BigDye™Terminator ver. 2.1, and analysed on an ABI 3730 automated sequencer.

**Analyses and Hypothesis Testing**

Sequence chromatograms were edited, trimmed of the primer ends and exported as fasta files in Sequencher v. 4.8 (Gene Codes Corporation). New DNA sequences are submitted to Genbank under the accession codes JX434757-JX434840 (Appendix). Sequences were aligned in Clustal X v. 2.0.12 (Larkin et al. 2007) with default gap opening and gap extension penalties (15 and 6.66, respectively). The alignments of the length-variable ribosomal genes were compared with using Mafft v. 6.850 (Katoh and Toh 2008) (E-INS-i, L-INS-I, and G-INS-i, all with 10 iterations) but differences were minimal (1 bp in length, 0.003 and 0.022 difference in
posterior probability branch support for the placement of *Suphrodytes* in single-gene analyses of 16s and 18s, respectively) and the clustal alignments were used for all four genes. The parallel mpi version of MrBayes version 3.2 (build 457) (Altekar et al. 2004; Ronquist et al. 2012) was used to infer a phylogenetic hypothesis from the combined four-gene dataset. We specified a partitioned model based on genes and codon positions. For each partition we specified a model a priori allowing for the estimation of base frequencies, the proportion of invariable sites and allowed for rate-variation across sites with a gamma distribution. However, the core of the model, the substitution rate matrix, was not specified a priori. Instead we used reversible-jump MCMC to integrate over the pool of all 203 possible reversible 4 × 4 nucleotide models. The method was first described by Huelserbeck et al. (2004) and has been implemented in the latest version of MrBayes 3.2. We specified two hypotheses to be tested against each other. Traditional classification treats *Hydroporus* each other. Traditional classification treats *Hydroporus* as monophyletic and the interference of other *Hydroporini*-taxa in relation to *Hydroporus*. For the Bayes factor test we contrast what we call the standard approach with our preferred approach and show that it changes the outcome of the test. In the standard test we calculate the marginal likelihood of *H*$_0$ by using an absolute monophyly constraint on *Hydroporus* as an informed topology prior, whereas the marginal likelihood of *H*$_1$ was calculated from an unconstrained analysis with an uninformative prior across topology space. This is the approach taken in many empirical studies (e.g., Lavoué et al. 2007; Marek and Bond 2007; Parker et al. 2007; Azuma et al. 2008; Yamanoue et al. 2010; Palickev et al. 2009; Tank and Olmstead 2009; Maksowsky et al. 2010; Yang et al. 2010; Knight et al. 2011). In our preferred Bayes factor test we calculate the marginal likelihood of the alternative hypotheses after specifying equally informed priors (constraints) on the topology (Fig. 2).

To calculate the marginal likelihood of models we used the stepping-stone sampling method of Xie et al. (2011), which is now implemented in MrBayes 3.2. (Ronquist et al. 2012). We used a value (0.4) for the $\alpha$-shape parameter of the beta distribution within the range Xie et al. (2011) found optimal and 204,000 MCMC steps were sampled every 100th generation for each of 48 $\beta$-values between 1 (posterior) and 0 (prior) after an initial 204,000 generations were discarded as burn-in. The contribution to the marginal likelihood from each step is estimated from a sample-size of 2040. The same setting was used for four independent runs for each model to be tested, each with one cold and one heated chain, and the arithmetic mean across runs of the estimated marginal likelihood for each model was used to calculate the Bayes factor. For all analyses, the chains and runs were distributed across eight cores of two 2.8 GHz Quad-Core Intel Xeon processors.

**RESULTS**

DNA extraction, PCR and sequencing was successful for all 21 taxa and four genes added to the dataset resulting in no complete gene gaps in the alignment. The combined data matrix consisted of 2111 aligned nucleotides of which 553 varied and 1558 were constant (Supplementary Table S2). The separate runs of the Bayesian analyses converged unproblematically to an average deviation of split frequencies of < 0.005, and the post burn-in, merged, runs resulted in the fully resolved topology shown in Figure 3. The tree could be rooted between ingroup (posterior probability, pp = 0.99) and outgroups as specified a priori. In the ingroup, *Hydrocolus* came out as sister to the remaining genera, but this must be seen as a tentative hypothesis as the support was just on the margin (pp = 0.51) to be recovered in the majority-rule consensus of the sampled trees. At this low level of support the recovered resolution is often sensitive to model specification. *Hydrocolus* apart, *Heterostenura-* *Neoporus-* *Sanfilippodytes* formed a strongly supported monophyletic group (pp = 0.99) sister to *Hydroporus*–*Suphrodytes*. *Hydroporus* and *Suphrodytes* were also highly supported as monophyletic (pp = 0.98). Within that clade, *Hydroporus* came out as paraphyletic due
to *Suphrodytes* being nested inside with a posterior probability of 0.87. *Hydroporus* was divided into two well-supported monophyletic groups, here labeled Clade I and II. Clade I (pp = 0.98) included the following species groups: *angustatus, fuscipennis, longulus, mennonius, neglectus, nigrita, tuberalus, striola, tessellatus and tristis*. Clade II (pp = 1.0) included the *appalachicus, axillaris, columbianus, erythrocephalus, lapporum, nigellus, niger, notabilis, obscurus, rufibrans, simniatipennis, subpubescens*, and *transpunctatus* species groups. *Suphrodytes* came out as sister group to Clade II with rather strong support (pp = 0.87). In the backbone-constrained analysis where *Hydroporus* was constrained to be monophyletic in relation to all taxa except “floating” *Suphrodytes*, the support for *Suphrodytes + Hydroporus* increased marginally to 1.0 and *Suphrodytes + Hydroporus* Clade II to 0.88.

Filtering the merged, post-burn-in, sample of trees for trees consistent with each hypothesis resulted in a posterior model odds of *H* 1 versus *H* 0 of 0.947/0.0334 for the unconstrained analysis (Table 2). According to the posterior model odds there was hence about 28 times higher probability for the hypothesis *H* 1 of a paraphyletic *Hydroporus* without *Suphrodytes* to the hypothesis *H* 0 of a monophyletic *Hydroporus*.

The stepping-stone MCMC sampling converged successfully among the four independent runs for all 48 β-values with an average deviation of split frequencies always < 0.03. The standard Bayes factor test estimated the marginal likelihood for *H* 1 from an unconstrained analysis and for *H* 0 under the topological constraint of *Hydroporus* monophyly. The estimation of the marginal log likelihood was −13.366.34 for the null hypothesis and −13.410.82 for the unconstrained *H* 1 hypothesis (Table 3). The test statistic 2 × log *BF* = 88.96 which, according to the scale of interpretation (Table 1), gives very strong support in favor of the constrained hypothesis (*H* 0) forcing *Hydroporus* to be monophyletic, despite the fact that this model has a much lower posterior probability (Table 2). In contrast, the preferred Bayes factor test where the two hypotheses are tested under equally informed priors on topology space, gives strong support (2 × log *BF* = 7.82) in favor of the alternative hypothesis (*H* 1) where *Suphrodytes* is nested within *Hydroporus*. Note that in a test with an ambiguous prior constraint on topology where *H* 0 is calculated only under constraint *H* 1 (= *Hydroporus* monophyletic) and *H* 1 is calculated only under constraint *H* 1 (= *Hydroporus* monophyletic), allowing a nested position of *Suphrodytes* but not excluding a sister-group relationship, the Bayes factor support for *H* 1 is reduced from strong to positive (Table 4; 2 × log *BF* = 4.42). Despite a large number of free parameters, given the combination of eight data partitions and reversible-jump MCMC across 203 substitution models for each, the consistent estimate of the marginal likelihood under various constraints across the four independent runs indicate that this was not a problem for the approximation (Table 4). Following the best supported...
hypothesis in our preferred Bayes factor test and from the posterior model odds, we reinstall Suphrodytes as a junior synonym of Hydroporus, following Zimmermann (1931), in order to make Hydroporus monophyletic, resulting in the reinstalled combinations H. dorsalis (Fabricius 1787) and Hydroporus figuratus (Gyllenhal 1826).

**DISCUSSION**

The main point of this article is that Bayes factor tests of topological hypotheses are extremely sensitive to the tree space associated with each hypothesis in the prior. One needs to carefully consider the relevant tree classes.
Table 2. Posterior probability of hypotheses based on an unconstrained analysis and a backbone-constrained analysis where *Hydroporus* is constrained monophyletic in relation to the outgroups but *Suphrodytes* is allowed to “float”.

| Constraints | H₀ | H₁ | H₂ | H₃ |
|-------------|----|----|----|----|
| Unconstrained | 0.03339 | 0.86825 | 0.08582 | 0.9662 |
| (30,004) | (1018) | (26,050) | (2601) | (28,411) |
| Backbone | 0.03376 | 0.88025 | 0.08582 | 0.9662 |
| (30,004) | (1013) | (26,410) | (2576) | (28,991) |

Notes: H₀ = *Hydroporus* monophyletic to the exclusion of *Suphrodytes*; H₁ = *Suphrodytes* + *Hydroporus* clade II monophyletic; H₂ = *Suphrodytes* + *Hydroporus* Clade I monophyletic; H₃ = *Suphrodytes* nested within *Hydroporus*. In parenthesis the total number of trees filtered out from the total (30,004) that supports the hypothesis.

Table 3. Estimations of the marginal likelihood using four independent stepping-stone sampling runs (SS) of three constrained hypotheses with an equally informed prior on topology, and of an unconstrained analysis (UC).

| SS runs | Lik. H₀ | Lik. H₁ | Lik. H₂ | Lik. UC | H₀ versus UC | H₁ versus UC |
|---------|---------|---------|---------|---------|-------------|-------------|
| 1 | −13.366 60 | −13.364 61 | −13.365 75 | −13.411 57 | 0.946 90 | 0.880 25 |
| 2 | −13.368 18 | −13.362 98 | −13.364 75 | −13.410 21 | 0.946 90 | 0.880 25 |
| 3 | −13.369 19 | −13.368 74 | −13.365 65 | −13.410 75 | 0.946 90 | 0.880 25 |
| 4 | −13.365 39 | −13.362 19 | −13.365 36 | −13.411 32 | 0.946 90 | 0.880 25 |
| Mean | −13.366 34 | −13.362 43 | −13.365 00 | −13.410 62 | 0.946 90 | 0.880 25 |

Table 4. Mean and standard deviation from four independent stepping-stone sampling estimations of the marginal likelihood under different constraints and unconstrained (UC).

| Constraints | Marg. Log. Lik. | St. Dev. (4 runs) |
|-------------|----------------|------------------|
| UC | −13.410 82 | 0.6078 |
| H | −13.391 51 | 0.3093 |
| HS | −13.389 30 | 1.9994 |
| HS, H₁, H₂, S | −13.363 21 | 1.0238 |
| HS, HS, H₁, H₂, S | −13.366 34 | 1.0845 |
| HS, HS, H₁, H₂, S | −13.362 43 | 1.2604 |
| HS, HS, H₁, H₂, S | −13.365 00 | 0.4941 |
| H, H₁, H₂ | −13.370 77 | 1.8202 |

Note: For explanation of constraints see Figure 2.

If the support is more diffuse, it may be better to focus on posterior model odds than to rely on Bayes factors calculated using the standard approach.

A special case concerns the rejection of hypotheses of monophony. Our examples suggest that the standard Bayes factor test is always biased toward acceptance of the monophony hypothesis. Thus, it appears that one can safely use a standard Bayes factor test to reject a hypothesis of monophony, but it is typically going to be an extremely conservative test. The conservative nature of the test can be seen in our empirical example (Fig. 2a,b): an attempt to reject the hypothesis of a monophyletic *Hydroporus* fails under the standard approach. In contrast, setting five equivalent topology constraints for the two hypotheses to be tested (Fig. 2c,d), yields an equally informed prior on tree space and the outcome of the Bayes factor test is reversed—the monophony of *Hydroporus* can be rejected. This internal resolution of *Hydroporus* and the position of *Suphrodytes* are also in agreement with previous studies (Ribera et al. 2003, 2008; Hernando et al. 2012). The sister clade of *Suphrodytes* (Clade II) includes the large-bodied northern species, for example *H. lapponum*, *H. notabilis* and *H. submuticus*, with which *Suphrodytes* historically have been associated (Zimmermann 1931, Guignot 1932, Zaitzev 1953).

Could the outcome of the Bayes factor test have been reversed if the posterior probability node support for a paraphyletic *Hydroporus* had been 1.0 instead of 0.87 in the unconstrained analysis? If, hypothetically, the posterior probability for one hypothesis is exactly 1, the probability of the other hypothesis would be exactly 0, and the posterior odds ratio would be infinite. However, in reality the support for a hypothesis is never exactly 1 or exactly 0. All topological hypotheses have at least an infinitesimal posterior probability, and the reason “1.0” is a common node value in empirical studies is only due to the finite MCMC sample-size, MCMC error and rounding of numbers. Since the Bayes factor is the ratio of the posterior odds to the prior odds, and the posterior odds is never infinite, there is always at least a theoretical possibility that a biased-enough prior odds can lead to the above situation.

Estimating Bayes factors or posterior model odds for tree hypothesis testing is not straightforward. The most accurate approach is usually to focus on the posterior sample of an unconstrained analysis. With appropriate filtering, such a sample can often be used directly to estimate posterior model odds or the more refined Bayes factors proposed here. There are essentially two cases where one needs to consider more elaborate methods, such as stepping-stone sampling (Xie et al. 2013) or thermodynamic integration (Lartillot and Philippe 2006). First, when the posterior probability of at least one topological hypothesis is so small that it cannot be estimated accurately from the posterior tree sample. Second, when the focus is on Bayes factors and the posterior odds are so strongly skewed that even accurately estimated model probabilities are not going to produce reliable Bayes factor estimates. In the latter case, Bayes factors...
factor titration may be a more attractive alternative to stepping-stone sampling or thermodynamic integration (Suchard et al. 2005). However, as shown both with our hypothetical and empirical examples, an accurate estimation of Bayes factor is no guarantee against misleading conclusions if the model-specific priors are not carefully considered. Restricting the tree space with an informed prior for one hypothesis but not the other, as is commonly done, will strongly bias the Bayes factor test in favor of the hypothesis under the constrained analysis. Although the improved methods for estimating the marginal likelihood with accuracy set the stage for more accurate model testing (Xie et al. 2001; Lartillot and Philippe 2006), the dependency of Bayes factor on model-specific priors remain the same. This was well explained by Kass and Raftery (1995), but phylogeneticists have so far failed to recognize the implications for tests of topology hypotheses.

In conclusion, we argue that phylogeneticists should abandon the standard Bayes factor tests that are commonly used today to test topological hypotheses. Instead, they should take the irrelevant background signal about topological structure into account in the Bayes factor test. If this is not possible, then it is better to focus on posterior model odds than on Bayes factors calculated in the standard way, which are almost surely misleading.

**Supplementary Material**

Supplementary material, including data files and online-only Appendices, can be found in the Dryad data repository (doi:10.5061/dryad.s631d). Datamatrix and tree files are available in TreeBase (http://purl.org/phylo/treebase/phylows/study/TB2: S13927).

**Funding**

This work was supported by the Swedish Research Council (Grant No: 2009-3744 to J.B. and Grant No: 2011-5622 to FR).

**Acknowledgments**

We are grateful to Elisabeth Weingartner for assistance in the molecular lab and to Maxim Teslenko for support on the stepping-stone sampling programming and debugging.

**Appendix**

Data on specimens used in the analyses and Genbank accession numbers. Accession numbers in bold are new sequences, remaining sequences are from Ribera et al. (2008) or Bergsten et al. (2012).

| ID     | Cat. No. | Genus       | Species         | Species gr. | 16S         | COI         | 18S         | H3         |
|--------|----------|-------------|-----------------|-------------|-------------|-------------|-------------|------------|
| BMNH 74409 |          | Suphrodytes | doralis         |             | JX221577    | JX221617    | JX221991    | JX221993   |
| BMNH 74410 |          | Suphrodytes | figuratus       |             | JX221576    | JX221618    | JX221992    | JX221994   |
| BMNH 722681 |         | Hydroporus  | longulus        |             | JX434757    | JX434799    | JX443778    | JX443820   |
| BMNH 722685 |         | Hydroporus  | raartzi         |             | JX434757    | JX434799    | JX443778    | JX443820   |
| BMNH 729585 |         | Hydroporus  | rufifrons       |             | JX434758    | JX434800    | JX443779    | JX443821   |
| BMNH 739567 |         | Hydroporus  | tristis         |             | JX434759    | JX434801    | JX443780    | JX443822   |
| BMNH 800416 |         | Hydroporus  | nigter          |             | JX434760    | JX434802    | JX443781    | JX443823   |
| BMNH 800417 |         | Hydroporus  | nigter          |             | JX434760    | JX434802    | JX443781    | JX443823   |
| BMNH 800432 |         | Hydroporus  | assilariis      |             | JX434761    | JX434803    | JX443782    | JX443824   |
| BMNH 800437 |         | Hydroporus  | longiusculus    |             | JX434762    | JX434804    | JX443783    | JX443825   |
| BMNH 800458 |         | Hydroporus  | notabilis       |             | JX434763    | JX434805    | JX443784    | JX443826   |
| BMNH 800761 |         | Hydroporus  | carri           |             | JX434764    | JX434806    | JX443785    | JX443827   |
| BMNH 800760 |         | Hydroporus  | appalachius     |             | JX434765    | JX434807    | JX443786    | JX443828   |
| BMNH 800765 |         | Hydroporus  | appalachius     |             | JX434766    | JX434808    | JX443787    | JX443829   |
| BMNH 801066 |         | Hydroporus  | mannerheimi     |             | JX434767    | JX434809    | JX443788    | JX443830   |
| BMNH 800464 |         | Hydroporus  | lapparatus      |             | JX434768    | JX434910    | JX443790    | JX443831   |
| BMNH 801288 |         | Hydroporus  | nigritta        |             | JX434769    | JX434911    | JX443790    | JX443832   |
| BMNH 801293 |         | Hydroporus  | tesselatus      |             | JX434770    | JX434912    | JX443791    | JX443833   |
| BMNH 801298 |         | Hydroporus  | membroni       |             | JX434771    | JX434913    | JX443792    | JX443834   |
| BMNH 801303 |         | Hydroporus  | obscurus        |             | JX434772    | JX434914    | JX443793    | JX443835   |
| BMNH 8024791 |        | Hydroporus  | forius          |             | JX434773    | JX434915    | JX443794    | JX443836   |
| BMNH 8024796 |        | Hydroporus  | forius          |             | JX434774    | JX434916    | JX443795    | JX443837   |
| BMNH 8024800 |        | Hydroporus  | erthropeathedus  |             | JX434775    | JX434917    | JX443796    | JX443838   |
| NHRS-JLBK00000528 | | Hydroporus | appalachius     |             | JX434776    | JX434918    | JX443797    | JX443839    |
| NHRS-JLBK00000678 | | Hydroporus | subnaticus      |             | JX434777    | JX434919    | JX443798    | JX443840    |
| BMNH 818779 |         | Hydroporus  | nigellas         |             | JX36327    | JX363311    | JX850515    | E0f74019   |
| BMNH 818248 |         | Hydroporus  | nigellas         |             | JX36327    | JX363311    | JX850515    | E0f74019   |
| BMNH 818250 |         | Hydroporus  | pubescens       |             | EJ49327    | EJ38733    | E0f79096    |
| BMNH 818249 |         | Hydroporus  | pubescens       |             | EJ49327    | EJ38733    | E0f79096    |
| BMNH 818239 |         | Hydroporus  | scalestains     |             | EJ858278   | EJ858309    | E0f79097    |
| BMNH 818238 |         | Hydroporus  | nagrippus       |             | EJ858281   | EJ858312    | E0f79098    |
| ID Cat. No. | Genus        | Species       | Species gr. | 16S     | COI      | 18S     | H3      |
|------------|--------------|---------------|-------------|---------|----------|---------|---------|
| BMNH 693521| Hydrocorbus  | salmborgyi    | AJ850379    | AJ850629| AJ850518 | E670199 |
| BMNH 681287| Herpoderanea | pulcher       | AF518252    | AF518282 | AJ318732 | E670194 |
| BMNH 681544| Neoporus     | arizonicus    | AJ850380    | AJ850630 | AJ850519 | E670193 |
| BMNH 681288| Neoporus     | undulatus     | AJ850381    | AJ850631 | AJ318741 | E670210 |
| BMNH 681625| Sandlyraulus | terminalis    | AJ850426    | AJ850673 | AJ850552 | E670212 |
| MNCN A9   | Anax         | insigilis     | EF056665    | EF059870 | EF05633  | E670550 |
| MNCN A1126| Heteranthrus | minussonianis | EF056677    | EF056606 | EF05642  | E670563 |
| BMNH 681625| Hyphydrus    | onus           | EF056688    | EF056618 | EF05651  | E670576 |
| BMNH 681625| Hyphydrus    | onus           | EF056688    | EF056618 | EF05651  | E670576 |
| BMNH 681331| Laccornella  | copeloides    | AY334131    | AY334247 | AJ385758 | E670578 |

| ID Cat. No. | Legit        | Country       | Locality                                              |
|------------|--------------|---------------|-------------------------------------------------------|
| BMNH 74409 | J. Geijer    | Sweden        | Öland, Kalmar Lan, Mörbylånga, Ålgustrum, Jordtorp grustag. 27 August 2005 |
| BMNH 74403 | J. Geijer    | Sweden        | Öland, Kalmar Lan, Borgom, Hogorum, Viterkärret. 19 July 2005 |
| BMNH 722881| G. N. Foster | France        | Isère, Ruisseau des Fontinettes. 17 July 2005 |
| BMNH 729985| AN. Nilsson  | Sweden        | Ångermanland, Torbole, 23 May 2005 |
| BMNH 800446| J. Bergsten  | United States | New York, Richford, SE of Ithaca. 11 September 2002 |
| BMNH 800432| J. Bergsten  | United States | California, Eldorado Co., American river, campground by Silver lake. 20 September 2002 |
| BMNH 800437| J. Bergsten  | United States | California, Alpine Co., Hope Valley, Blue Lakes road, West Fork, Carson River. 21 September 2002 |
| BMNH 800458| J. Bergsten  | Canada        | Alberta, Meanoak biological station, W4mer. Twp65 Rge23 Sec12NW. 3 September 2002 |
| BMNH 800761| J. Bergsten  | Canada        | Alberta, Hwy.11 approx. 5 km E. of border to Banff National Park. 7 September 2002 |
| BMNH 800780| J. Bergsten  | Canada        | Alberta, W4mer. Twp67 Rge24 Sec34SE. 2 September 2002 |
| BMNH 800785| J. Bergsten  | Canada        | Alberta, W4mer. Twp65 Rge23 Sec13E. 3 September 2002 |
| BMNH 800806| J. Bergsten  | Canada        | Alberta, Hwy. 11, just E. of Big Horn. 7 September 2002 |
| BMNH 801288| D. Bilhon    | United Kingdom | Cornwall, The Lizard Pool 1 at Kynance Cove. 1 July 2005 |
| BMNH 801293| D. Bilhon    | United Kingdom | Cornwall, stream beside B3300 above bridge, 20 June 2005 |
| BMNH 801298| D. Bilhon    | United Kingdom | Cornwall, The Lizard Pool 2 at Kynance Cove. 1 July 2005 |
| BMNH 801303| J. Bergsten  | Sweden        | Vasterbotten, Umeå, Sirorfs, Umeahven. 26 August 2007 |
| BMNH 824791| J. Bergsten  | United States | California Sierra Co., Sierra Valley, Hwy 89 & 49. 23 September 2002 |
| BMNH 824798| J. Bergsten  | United States | California, Modoc Co., Hwy 299 approx. 5 km E. of Cedarville. 22 September 2002 |
| BMNH 824800| D. Bilhon    | United Kingdom | Cornwall, The Lizard, Hayle Kimbro pool. 1 July 2005 |
| NHRS-JLKB000000528| AN. Nilsson | Sweden | Torne Lappmark, Abisko, myre E. of Bjerkajakare. 28 June 2010 |
| NHRS-JLKB000000678| AN. Nilsson | Sweden | Vasterbotten, Vindeln, Strycksede, 29 May 2010 |
| BMNH 681779| B. Andreu    | Sweden        | S. Ha. Osuala, Kustgal, 1100 m SSV Rovrak, 3 November 2000 |
| BMNH 681248| D.T. Bilhon  | Tenerife      | Anaga, Roque Chinobre, December 1997 |
| BMNH 681250| I. Ribera    | Spain         | Burgos |
| BMNH 681249| I. Ribera    | England       | Dorest, Wareham, Morden bog, 5 July 1998 |
| BMNH 681239| I. Ribera    | Portugal      | Sa. Da Estrela, Torre, Lagoa 25 July 1998 |
| BMNH 693521| A.N. Nilsson | Sweden        | Prov. Vasterbotten, Åmåle, 3 August 1999 |
| BMNH 681257| Y. Alarie    | Canada        | Ontario |
| BMNH 681544| Y. Alarie    | United States | New Mexico, September 2000 |
| BMNH 681286| Y. Alarie    | Canada        | Ontario |
| BMNH 681625| I. Ribera & A. Cieslak | United States | 16 US California Mendocino co. / Rd. 1 Manchester / pond S City / 23 June 2000 |
| Outgroups  | MNCN A9      | South Africa  | North Cape, Stream at top of Studer Pass, 22 August 2004 |
| MNCN A1126 | M. Balke     | Madagascar    | Andasibe, Station Forestiere, orchid garden, 979 m, xi./xii.2004, MD 037 |
| BMNH 681225| I. Ribera    | United Kingdom | Sommerset Levels, Catcott Heath, 4 July 1998 |
| BMNH 681639| I. Ribera & A. Cieslak | South Africa | 17, W Cape, Limiet Berg, Tributary r. Wit, rd. R301 24 km NE Wellington. 25 March 2001 |
| BMNH 681331| I. Ribera    | Chile         | 16, X Reg. 6 km W La Unión, Pond in rd. to Huceicola, 29 January 1999 |
Suchard M. A., Weiss R. E., Sinsheimer J. S. 2005. Models for estimating Bayes factors with applications to phylogeny and tests of monophyly. Biometrics 61:665–673.

Sullivan J., Joyce P. 2009. Model selection in phylogenetics. Ann. Rev. Ecol. Evol. Syst. 36:445–466.

Swofford D. L. 2002. PAUP*: phylogenetic analysis using parsimony (and other methods). Version 4. Sunderland (MA): Sinauer Associates.

Tank D. C., Olmstead R. G. 2009. The evolutionary origin of a second radiation of annual Castilleja (Orobanchaceae) species in South America: the role of long distance dispersal and allopolyploidy. Am. J. Bot. 96:1907–1921.

Wheeler W. C., Pickett, K. M. 2008. Topology-Bayes versus Clade-Bayes in phylogenetic analysis. Mol. Biol. Evol. 25:447–453.

Xie W., Lewin P. O., Fan Y., Kuo L., Chen M.-H. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst. Biol. 60:150–160.

Yamanoue Y., Miya M., Matsuura K., Katoh M., Sakai H., Nishida M. 2008. A new perspective on phylogeny and evolution of tetraodontiform fishes (Pisces: Acanthopterygii) based on whole mitochondrial genome sequences: Basal ecological diversification? BMC Evol. Biol. 8:212.

Yang Z. 2006. Computational molecular evolution. Oxford: Oxford University Press.

Yang B. B., Guo X. G., Hu X. S., Zhang J. G., Liao L., Chen D. L., Chen J. F. 2010. Species discrimination and phylogenetic inference of 17 Chinese Leishmania isolates based on internal transcribed spacer 1 (ITS1) sequences. Parasitol. Res. 107:1049–1055.

Zaitzev F. A. 1953. Nasekomye zhestkokrylye. Plavuntsovye i vertyachki. Fauna SSSR 4(N. S. 58), 1–376.

Zakharev E. V., Lobov N. F., Nowak C., Hellmann J. 2009. Intrgression as a likely cause of mtDNA paraphyly in two allopatric skippers (Lepidoptera: Hesperiidae). Heredity 102:590–599.

Zimmermann A. 1931. Monographie der palaarktischen Dytisciden, II. Hydroporinae (Z. Teil). Koleopterol. Rundsch. 17:97–159.

Zwickl D. J., Hillis D. M. 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51:588–598.