The Effects of Partitioning on Phylogenetic Inference

David Kainer*[,1] and Robert Lanfear[1,2,3]

[1]Division of Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra, ACT, Australia

[2]National Evolutionary Synthesis Center, Durham, NC

[3]Department of Biological Sciences, Macquarie University, Sydney, NSW, Australia

*Corresponding author: E-mail: david.kainer@anu.edu.au.

Associate editor: Koichiro Tamura

Abstract

Partitioning is a commonly used method in phylogenetics that aims to accommodate variation in substitution patterns among sites. Despite its popularity, there have been few systematic studies of its effects on phylogenetic inference, and there have been no studies that compare the effects of different approaches to partitioning across many empirical data sets. In this study, we applied four commonly used approaches to partitioning to each of 34 empirical data sets, and then compared the resulting tree topologies, branch-lengths, and bootstrap support estimated using each approach. We find that the choice of partitioning scheme often affects tree topology, particularly when partitioning is omitted. Most notably, we find occasional instances where the use of a suboptimal partitioning scheme produces highly supported but incorrect nodes in the tree. Branch-lengths and bootstrap support are also affected by the choice of partitioning scheme, sometimes dramatically so. We discuss the reasons for these effects and make some suggestions for best practice.

Key words: partitioning, tree topology, tree inference, molecular evolution.

Introduction

Statistical frameworks for phylogenetic tree inference, such as Bayesian and Maximum Likelihood (ML) approaches, use parametric models to account for changes in DNA sequences over time. An overly simplistic model may fail to adequately account for certain aspects of the evolutionary process, and can limit the accuracy of the inferences made from the data. A more complex model may overfit the data and result in similarly compromised phylogenic inferences (Sullivan and Joyce 2005). As the choice of a model of molecular evolution can affect the results of a phylogenetic analysis (Sullivan and Swofford 1997; Buckley and Cunningham 2002; Kelchner and Thomas 2007), the process of model selection can be important.

As phylogenetic data sets have increased in size, it has become increasingly important for models of molecular evolution to adequately account for variation in rates and patterns of substitution among sites (Buckley et al. 2001; Brown and Lemmon 2007). The most common approach to incorporating this heterogeneity is to partition a given sequence alignment into groups of sites that are assumed to have evolved under the same model (Nylander et al. 2004; Brandley et al. 2005; Blair and Murphy 2011). Traditionally, this is performed by using a priori knowledge of the data set, for example by assigning an independent model to each codon position in each protein-coding gene (Shapiro et al. 2006; Lanfear et al. 2012). Various other solutions have been proposed, including mixture models and some related Bayesian methods (Huelsenbeck and Nielsen 1999; Krajewski et al. 1999; Yang et al. 2000; Lartillot and Philippe 2004; Nylander et al. 2004; Pagel and Meade 2004;

Wu et al. 2013). In the general mixture model approach, each site in an alignment is given a probability of being assigned to each model. Traditional partitioning, therefore, can be viewed as a special case where the probability of assignment of a site to a model is simply reduced to either 0 or 1. A recent Bayesian approach to the partitioning problem (Wu et al. 2013) employed a Dirichlet process mixture model to jointly estimate both the number of partitions and the assignment of sites to those partitions. Of these methods, traditional partitioning remains by far the most widely used, and is currently the only method that is feasible for the large multilocus alignments that are increasingly common. In this article, we focus on assessing the effects of various approaches to traditional partitioning on phylogenetic inference.

Having defined groups of homogeneous sites, or data blocks, based on a priori knowledge of the data, the biggest challenge in traditional partitioned analysis is to choose the best partitioning scheme for the data at hand. Researchers typically take either an ad hoc or algorithmic approach to this problem. In the ad hoc approach, one may assess a handful of alternative ways in which data blocks can be further combined into homogenous groups (partitioning schemes) and select the best-fit partitioning scheme for subsequent phylogenetic analyses. In the algorithmic approach, a variety of automated methods can be used to heuristically evaluate the space of all possible partitioning schemes containing the initial data blocks (Lanfear et al. 2012, 2014). In both cases, statistical model evaluation metrics such as the corrected Akaike Information Criterion (AICc) or the Bayesian Information Criterion (BIC) may be used to select the

partitioning scheme with the best fit to the data (Schwarz 1978; Burnham and Anderson 2002).

For any given partitioning scheme, an appropriate model of molecular evolution needs to be chosen for each group of sites before a phylogenetic analysis can be performed. This aspect of model selection has received a lot of attention, particularly with respect to selecting one of the submodels of the general time-reversible (GTR) model for a specific group of sites (Posada and Crandall 2001; Posada and Buckley 2004; Sullivan and Joyce 2005). It is known that alternative model-evaluation criteria (e.g., AIC or BIC) can lead to the selection of alternative best-fit substitution models for the same data, and in some cases the use of alternative best-fit models for phylogenetic inference can change the resulting tree topology (Posada and Crandall 2001; Lemmon and Moriarty 2004; Ripplinger and Sullivan 2008). However, although partitioning is one of the most popular approaches to account for variation in patterns of substitution among sites, little is known about the way that alternative approaches to defining the partitioning scheme itself can affect phylogenetic inference. Just as the use of alternative best-fit nucleotide substitution models can change the results of a phylogenetic analysis (Pol 2004; Ripplinger and Sullivan 2008), so too could the use of alternative partitioning schemes.

Many studies have examined the effect of partitioning scheme choice for a particular data set. Brandley et al. (2005) investigated the effect of nine alternative ad hoc partitioning strategies on model fit, topology and node support in Bayesian framework and found variation in inferred tree topologies at weakly supported nodes. Other studies have noted that partitioning scheme choice affects node support (Ward et al. 2010; Rota and Wahlberg 2012; Powell et al. 2013), tree topology (Strugnell et al. 2005; Leavitt et al. 2013; Tao et al. 2013), and branch-lengths (Poux et al. 2008; Ho and Lanfear 2010). Occasionally the choice of partitioning scheme has little or no effect on tree characteristics at all (Cameron et al. 2012).

Although these studies demonstrate that partitioning can have important effects on phylogenetic inference, it is difficult to draw any general conclusions from them. The partitioning schemes compared within each study were selected on an ad hoc basis with very different approaches; some studies compared a small set of very similar partitioning schemes (e.g., Cameron et al. 2012), and others compared extensive sets of very different partitioning schemes (e.g., Rota and Wahlberg 2012; Leavitt et al. 2013). Furthermore, previous studies used a variety of different phylogenetic methods, software implementations, and tree inference methods. Simulations have shown that underpartitioning typically causes more severe error than overpartitioning, particularly with respect to node support (Brown and Lemmon 2007), and that the inappropriate modeling of among-partition rate variation can strongly affect branch-lengths (Marshall et al. 2006). These simulations give us a good picture of the effects of partitioning in specific cases when the truth is known. However, it is not clear whether a wide variety of empirical data sets are consistently affected in a similar way. To better understand the effects of partitioning scheme choice on phylogenetic inference, we need to compare a range of partitioning schemes across many empirical data sets within a unified statistical framework.

In this study, we test the effects of partitioning scheme choice on phylogenetic inference by analyzing a set of 34 published empirical data sets (table 1) under four different partitioning schemes that represent the full range of popular approaches to partitioning. The published alignments range from 13 to 2,872 taxa and 823 to 25,919 sites. They include data from the mitochondrial, nuclear, and chloroplast genomes from a diverse array of animals and plants (table 1). The four partitioning schemes we compared are 1) a scheme that leaves the data set unpartitioned; 2) a scheme with distinct groups of sites defined by genes, exons, introns, and codon positions; 3) an algorithmically estimated partitioning scheme selected with AICc; and 4) an algorithmically estimated partitioning scheme selected with BIC. For each data set, we estimated phylogenetic trees with bootstrap support under each partitioning scheme. We then used these data to assess the impact of partitioning scheme choice on phylogenetic tree topology, branch-lengths, and node support in an ML framework.

We find that the choice of partitioning scheme almost always affects tree topology, and the effects are particularly severe when data are left unpartitioned. Most notably, we find occasional instances where the use of a suboptimal partitioning scheme produces highly supported but incorrect nodes in the tree. Branch-lengths and node support are also affected by the choice of partitioning scheme, sometimes dramatically so. We discuss the reasons for these effects and make some suggestions for best practice.

Results

For each of the 34 data sets in our study, we estimated the ML tree using four different partitioning schemes: 1) $Adhoc_{one}$—an ad hoc partitioning scheme that groups all sites into a single subset; 2) $Adhoc_{feat}$—an ad hoc partitioning scheme that partitions the data set into the maximum possible number of data blocks based on genomic features (genes, codon positions, exons, and introns) and treats each data block as a separate subset of sites; 3) Opt_{AICc}—an algorithmic partitioning scheme in which the scheme and models of molecular evolution are estimated with the greedy algorithm using the AICc in PartitionFinder (Lanfear et al. 2012), using the $Adhoc_{feat}$ scheme as a starting point; and 4) Opt_{BIC}—an algorithmic partitioning scheme in which the scheme and models of molecular evolution are estimated with the greedy algorithm using the BIC in PartitionFinder, using the $Adhoc_{feat}$ scheme as a starting point. For the first two partitioning schemes, we chose the best-fit substitution model for each subset of sites using the BIC in PartitionFinder (see Materials and Methods). We compared the topologies, branch-lengths, and bootstrap support values inferred under each of these partitioning schemes.

Table 1. Details of the 34 Data Sets Used in This Study.

Data Set	Data Type[a]	Taxa	Sites	Data Blocks	Clade (Latin)	Clade (English)	Study Ref.	Data Set Ref.
Anderson_2013	M	145	3,037	4	Loliginidae	Pencil squids	Anderson et al. (2014)	Anderson et al. (2013)
Bergsten_2013	M,N	38	2,111	8	Dytiscidae	Diving beetles	Bergsten et al. (2013a)	Bergsten et al. (2013b)
Broughton_2013	M,N	61	19,997	61	Osteichthyes	Bony fishes	Broughton et al. (2013b)	Broughton et al. (2013a)
Brown_2012	N	41	1,665	7	Ptychozoon	Asian geckos	Brown et al. (2012b)	Brown et al. (2012a)
Caterino_2001	M,N	37	3,228	9	Papilionidae	Butterflies	Caterino et al. (2001)	Kuo et al. (2001)
Cognato_2001	M,N	44	1,896	7	Scolytinae	Bark beetles	Cognato and Vogler (2001b)	Cognato and Vogler (2001a)
Day_2013	M,N	152	3,586	11	Synodontis	African catfish	Day, Peart, Brown, Friel, et al. (2013)	Day, Peart, Brown, Bills, et al. (2013)
Devitt_2013	M	69	823	4	Ensatina	Salamander	Devitt et al. (2013b)	Devitt et al. (2013a)
Dornburg_2012	M,N	44	5,919	21	Holocentridae	Squirrel fishes	Dornburg et al. (2012b)	Dornburg et al. (2012a)
Dsouli_2011	M,N	39	1,635	7	Muscidae	Flies	Dsouli et al. (2011)	NA
Ekrem_2010	M,N	74	2,701	10	Chironomidae	Midges	Ekrem et al. (2010)	NA
Elias_2009	M,N	143	4,159	12	Nymphalidae	Butterflies	Elias, Joron, Willmott, Silva-Brandāo, et al. (2009)	Elias, Joron, Willmott, Kaiser, et al. (2009)
Fishbein_2001	N,C	40	9,005	11	Saxifragales	Core Eudicots	Fishbein et al. (2001b)	Fishbein et al. (2001a)
Fong_2012	N	110	25,919	168	Vertebrata	Vertebrates	Fong et al. (2012b)	Fong et al. (2012a)
Grande_2013	M,N	65	4,027	12	Paracanthopterygii	Fish	Grande et al. (2013a)	Grande et al. (2013b)
Guschanski_2013	M,C	110	17,092	63	Cercopithecini	Monkeys	Guschanski et al. (2013b)	Guschanski et al. (2013a)
Kaffenberger_2011	M,N	54	6,548	26	Gephyromantis	Malagasy frogs	Kaffenberger et al. (2012)	Kaffenberger et al. (2011)
Kang_2013a	N	28	7,276	15	Xiphophorus	Swordtail fish	Kang et al. (2013)	NA
Kang_2013b	M	28	1,239	6	Xiphophorus	Swordtail fish	Kang et al. (2013)	NA
Kawahara_2013	M,N	70	2,238	9	Hyposmocoma	Caterpillar	Kawahara and Rubinoff (2013a)	Kawahara and Rubinoff (2013b)
Lartillot_2012	N	78	15,117	51	Eutheria	Mammals	Lartillot and Delsuc (2012b)	Lartillot and Delsuc (2012a)
Leavitt_2013	M	34	15,404	87	Acridoidea	Grasshoppers	Leavitt et al. (2013)	NA
Li_2008	N	56	7,995	30	Actinopterygii	Fishes	Li et al. (2008)	NA
Murray_2013	M,N	237	3,111	9	Eucharitidae	Wasps	Murray et al. (2013a)	Murray et al. (2013b)
Rightmyer_2013	M,N	94	3,692	25	Hymenoptera	Bee	Rightmyer et al. (2013b)	Rightmyer et al. (2013a)
Sauquet_2011	N,C	51	5,444	10	Nothofagus	Beeches	Sauquet et al. (2012)	Sauquet et al. (2011)
Seago_2011	M	116	2,253	7	Coccinellidae	Ladybirds	Seago et al. (2011b)	Seago et al. (2011a)
Sharanowski_2011	N	139	3,982	11	Braconidae	Wasps	Sharanowski et al. (2011b)	Sharanowski et al. (2011a)
Siler_2013	M,N	61	2,697	7	Lycodon	Wolf snakes	Siler, Oliveros, et al. (2013)	Siler, Brown, et al. (2013)
Tolley_2013	M	203	5,054	16	Chamaeleonidae	Chameleons	Tolley et al. (2013b)	Tolley et al. (2013a)
Unmack_2013	M	139	6,827	25	Melanotaeniidae	Rainbowfish	Unmack et al. (2013b)	Unmack et al. (2013a)
Wainwright_2012	N	188	8,439	30	Acanthomorpha	Fishes	Wainwright, Smith, Price, Tang, Sparks, Ferry, Kuhn, Eytan, et al. (2012)	Wainwright et al. (2012)
Ward_2010	N	54	9,173	27	Dolichoderinae	Ants	Ward et al. (2010)	NA
Welton_2013	M,N	145	4,552	16	Varanus	Lizards	Welton et al. (2013b)	Welton et al. (2013a)

NOTE.—NA, not applicable.

[a]Data Type refers to DNA source (M, mitochondrial; N, nuclear; C, chloroplast).

Topology Comparison

In what follows, we use the topology estimated under the Opt_{BIC} partitioning scheme as a reference against which to compare the other three topologies. This decision is arbitrary in the sense that we could use any of the four trees as a reference, although there is some evidence that the BIC is one of the best-performing metrics for model selection in phylogenetics (Minin et al. 2003; Abdo et al. 2005).

Tree topologies differed considerably depending on the partitioning scheme used to estimate them. When compared with the trees estimated using the Opt_{BIC} scheme, different tree topologies were estimated in 31/34 data sets using the $Adhoc_{one}$ scheme, 19/34 data sets using the $Adhoc_{feat}$ scheme, and 18/34 data sets using the Opt_{AICc} scheme. Below, we investigate these results further by comparing

the trees using a range of different metrics, and by analyzing the nodes that were present in some trees but not others.

For each data set, we calculated the distance between the tree estimated with the Opt_{BIC} scheme and the trees estimated with the other three schemes using three metrics: The Robison–Foulds metric (RF) (Robinson and Foulds 1981), the Matching Split distance metric (MS) (Bogdanowicz and Giaro 2012), and the Path Difference metric (PD) (Steel and Penny 1993). All three metrics were normalized relative to the mean distance between two randomly generated trees of the same size, such that a result of 0 implies that the two topologies are identical, whereas a result of 1 implies that the two topologies are at least as different as two random topologies (Bogdanowicz et al. 2012). This normalization allows the metrics to be meaningfully compared across different data sets.

Table 2. Mean Topological Distances between Trees Inferred with Different Partitioning Schemes.

Partitioning Scheme (vs. Opt_{BIC})	All comparisons[a]			Distance > 0[b]			
	$\overline{RF}$	$\overline{MS}$	$\overline{PD}$	# Data Sets (out of 34)	$\overline{RF}$	$\overline{MS}$	$\overline{PD}$
Opt_{AICc}	0.05	0.04	0.11	18	0.09	0.07	0.21
$Adhoc_{feat}$	0.05	0.04	0.11	19	0.09	0.07	0.20
$Adhoc_{one}$	0.12	0.10	0.25	31	0.13	0.10	0.28

Note.—All distances are calculated relative to the trees inferred with the Opt_{BIC} scheme. RF, Robinson-Foulds; MS, Matching Split; PD, Path Difference.
[a]Includes every data set, including those where trees inferred under two different schemes are topologically identical (i.e., distance = 0).
[b]Only data sets where trees inferred under two different schemes are topologically different.

Mean distances between topologies estimated under different partitioning schemes across all 34 data sets ($\overline{RF}$, $\overline{MS}$, $\overline{PD}$) are shown in table 2. These data show that when compared with topologies estimated using the Opt_{BIC} scheme, topologies from the Opt_{AICc} scheme were the most similar ($\overline{RF} = 0.048$, $\overline{MS} = 0.038$, $\overline{PD} = 0.0110$), followed by those from the $Adhoc_{feat}$ scheme ($\overline{RF} = 0.051$, $\overline{MS} = 0.038$, $\overline{PD} = 0.114$) and then those from the $Adhoc_{one}$ scheme ($\overline{RF} = 0.118$, $\overline{MS} = 0.099$, $\overline{PD} = 0.255$). When compared across all data sets, the distances of topologies estimated using the Opt_{AICc} and $Adhoc_{feat}$ schemes were not significantly different from one another (Mann–Whitney test, unpaired, two-sided), but the distances of the topologies estimated using the $Adhoc_{one}$ scheme were significantly larger than those estimated using the Opt_{AICc} and $Adhoc_{feat}$ schemes (Mann–Whitney test, unpaired, two-sided, $P < 0.01$ in all comparisons and across all three tree distance metrics).

To further investigate the differences between tree topologies, we calculated the proportion of conflicted nodes, which we define as the proportion of nodes that are present in one of the trees estimated using the Opt_{AICc}, $Adhoc_{feat}$, or $Adhoc_{one}$ partitioning schemes but not in the tree estimated from the same data set using the Opt_{BIC} scheme. Out of a total of 8,568 internal nodes examined across all data sets, 646 (7.54%) were conflicted. Trees estimated using the Opt_{AICc} scheme or $Adhoc_{feat}$ scheme typically had far fewer conflicted nodes than trees estimated with the $Adhoc_{one}$ scheme (Opt_{AICc}: 150; $Adhoc_{feat}$: 154; $Adhoc_{one}$: 342, across all data sets). The proportion of nodes in a tree that were conflicted ranged from zero to over 15% of all nodes in a single tree (fig. 1A), and tended to be highest in trees estimated with the $Adhoc_{one}$ scheme, for which it reached a value greater than 10% for 6 of 34 data sets (fig. 1A).

Effects on Branch-Lengths and Bootstrap Support

Branch-lengths remained very similar in the trees estimated using the Opt_{BIC}, Opt_{AICc}, or $Adhoc_{feat}$ schemes, and the same was true of bootstrap support values. In contrast, trees estimated using the $Adhoc_{one}$ scheme sometimes had very different branch-lengths and bootstrap values than trees

estimated using the other three partitioning schemes, and the effect was particularly pronounced for branch-length estimates.

Within each data set we calculated the mean branch-length ($\overline{BL}$) for each ML tree inferred with the four partitioning schemes. Although the trees in most data sets showed quite consistent $\overline{BL}$ irrespective of the partitioning scheme used to estimate them (fig. 2A), in a few data sets the trees showed large fluctuations, with the $\overline{BL}$ of some trees inferred under the Opt_{AICc}, $Adhoc_{feat}$, and Opt_{BIC} schemes being an order of magnitude greater than the $\overline{BL}$ of the tree inferred under the $Adhoc_{one}$ scheme. These long branch-lengths are implausible, often reaching values much greater than 1, and appear to be an unintended consequence of the ML inference software that we used. In what follows we treat the six data sets which contained trees with $\overline{BL} > 0.15$ (fig. 2A) as outliers, and repeat our analyses with and without these data sets. After excluding outliers, visual inspection of the data (fig. 2B) shows that partitioning a given data set with any of the Opt_{AICc}, $Adhoc_{feat}$ or Opt_{BIC} schemes produced trees with only very minor differences in mean branch-lengths, whereas partitioning the data set with the $Adhoc_{one}$ scheme produced much greater variation in mean branch-lengths (e.g., in one case, branches estimated with the $Adhoc_{one}$ scheme were less than half as long as those estimated with the Opt_{BIC} scheme; see the Rightmyer_2013 data set, table 3).

To test the significance of any differences in branch-lengths within each data set, after excluding the six outlier data sets described above we compared the set of branch-lengths of the tree estimated with the Opt_{BIC} scheme with the sets of branch-lengths of the trees estimated with the Opt_{AICc}, $Adhoc_{feat}$, and $Adhoc_{one}$ schemes (table 3). Only 1 data set out of 28 showed a significant change in branch-length distribution when either the Opt_{AICc} or $Adhoc_{feat}$ partitioning schemes were used instead of the Opt_{BIC} scheme (two-sided Mann–Whitney test, $P < 0.05$). However, 13 out of 28 data sets showed a significant change in branch-length distribution (two-sided Mann–Whitney test, $P < 0.05$) when the $Adhoc_{one}$ partitioning scheme was used instead of the Opt_{BIC} scheme (table 3).

To assess whether some partitioning schemes lead to the inference of longer or shorter branch-lengths across data sets, we calculated three differences in mean branch-length ($\Delta\overline{BL}$) for each of the 34 data sets, that is, including the six outlier data sets, in our study: $\Delta\overline{BL}$ (Opt_{BIC}, Opt_{AICc}), $\Delta\overline{BL}$ (Opt_{BIC}, $Adhoc_{feat}$), $\Delta\overline{BL}$ (Opt_{BIC}, $Adhoc_{one}$). A two-tailed sign test across all data sets (including the six outlier data sets) showed that the mean branch-length of trees inferred with the $Adhoc_{feat}$ scheme had a small but significant tendency to be longer than the mean branch-length of trees inferred with the Opt_{BIC} scheme (two-tailed sign test, $P = 0.035$, 23 of 33 data sets with nonzero branch-length differences had longer mean branch-lengths with the $Adhoc_{feat}$ scheme; table 3), whereas there was no directional trend when comparing the mean branch-lengths of trees estimated using the Opt_{BIC} scheme and either the Opt_{AICc} or $Adhoc_{one}$ schemes (two-tailed sign test, $P > 0.05$ in both cases). However, after removing the six outlier data sets we found that there was no

4
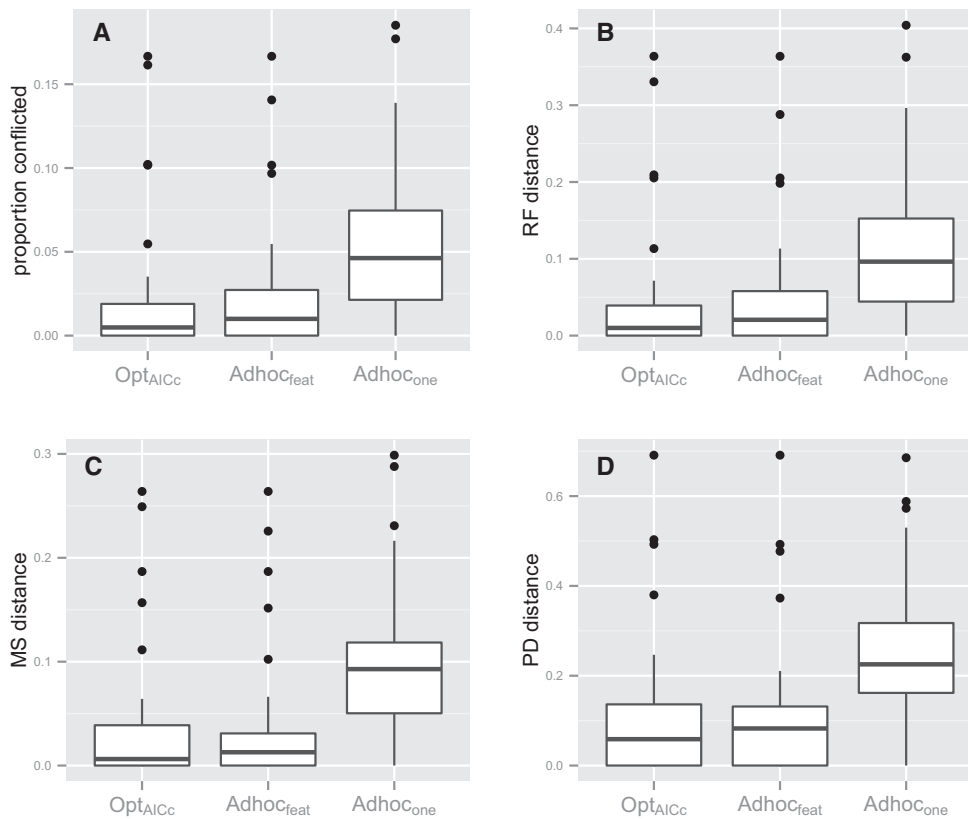
Fig. 1. The worse the partitioning scheme, the more different the tree topologies from topologies inferred with an optimal partitioning scheme. Each panel shows boxplots of a measure of topological distance between trees inferred with the Opt_{BIC} partitioning scheme and trees inferred with three other partitioning schemes. Each data point contributing to a boxplot represents one of the 34 data sets in our study. The three partitioning schemes are ordered from left to right in decreasing similarity to the Opt_{BIC} scheme. (A) Proportion of nodes that differ between the two trees, (B) normalized Robinson-Foulds distance, (C) normalized matching split distance, and (D) normalized Path Difference. In the three y axes in (B)–(D), a value of 0 corresponds to trees identical to the tree inferred using the Opt_{BIC} scheme, and a value of 1 represents the average difference between two randomly generated trees (see main text).

significant bias in the direction of the branch-length differences in any of the three comparisons (two-tailed sign test, $P > 0.05$ in all cases).

We included two tests to further investigate the effects of partitioning on branch-length estimation. First, we assessed the tendency for underpartitioned models to underestimate longer branches in phylogenetic trees (Revell et al. 2005), by repeating the comparison of branch-length distributions between trees estimated with different partitioning schemes using only the longest quartile of the branches. When either the Opt_{AICc} or $Adhoc_{feat}$ partitioning schemes were used instead of the Opt_{BIC} scheme, the number of data sets with a significant change in the distribution of long branches (two-sided Mann–Whitney test, $P < 0.05$) was 1 and 2, respectively. When the $Adhoc_{one}$ partitioning scheme was used instead of the Opt_{BIC} scheme, 18 out of 28 data sets showed a significant change in the distribution of long branches (table 3 and supplementary table S1, Supplementary Material online). However there was still no significant directional bias for the change in branch-lengths across the 28 data sets (two-sided Mann–Whitney test based on changes in the mean of the

longest quartile of branches, $P > 0.05$ in all cases). Second, we investigated whether there was any tendency for under-specified models to systematically underestimate the length of deeper branches. To do this, we calculated the slope of the relationship between the depth of a branch and its length for the trees inferred with the Opt_{BIC} and $Adhoc_{one}$ schemes. Any tendency to systematically underestimate the length of deeper branches will cause the slope of this line to become more negative. Thus, if underspecified models tend to underestimate deeper branches of the tree, we would expect trees estimated with the $Adhoc_{one}$ scheme to have a more negative slope than trees estimated with the Opt_{BIC} scheme. We found that the slope for trees inferred with the $Adhoc_{one}$ scheme was more negative in 18 out of 34 data sets, and more positive in 16 out of 34 data sets, providing no evidence for a systematic underestimation of deeper branches across data sets (supplementary table S2, Supplementary Material online; two-tailed sign test, $P = 0.864$).

To test the effects of partitioning on bootstrap support within each data set, we compared the bootstrap support values of the tree estimated with the Opt_{BIC} to the bootstrap
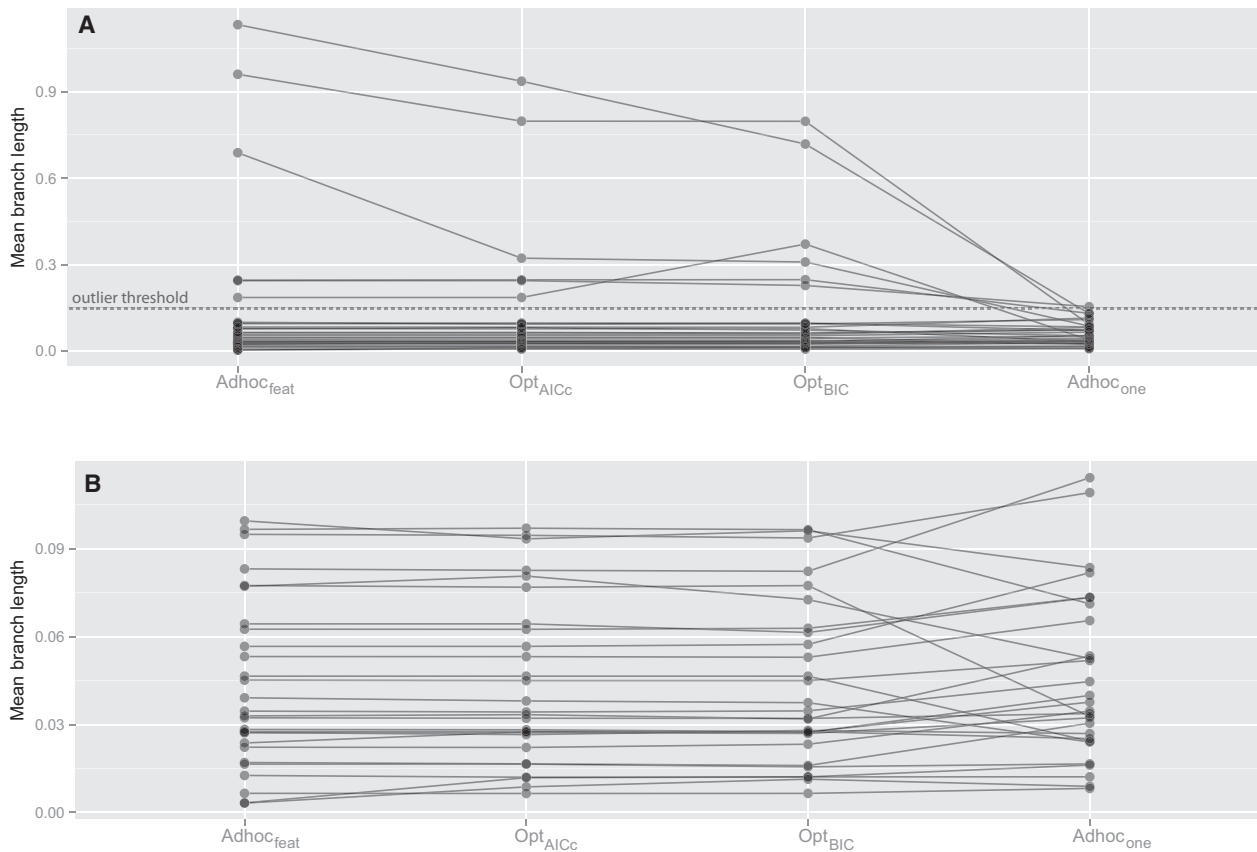
FIG. 2. Branch-lengths are consistent when trees are inferred with different partitioning schemes, except for trees inferred with no partitioning. Each panel shows the mean branch-length ($\overline{BL}$) for each data set when different partitioning schemes are used for tree inference. Horizontal lines connect the same data set across treatments. The dashed line shows the threshold for which any data set with higher $\overline{BL}$ from any of the partitioning schemes was excluded from further analysis. (A) includes all 34 data sets, (B) shows only the 28 data sets with $\overline{BL}$ less than the threshold of 0.15 for all partitioning schemes.

support values of the trees estimated with the $\mathrm{Opt_{AICc}}$, $\mathrm{Adhoc_{feat}}$, and $\mathrm{Adhoc_{one}}$ schemes (table 3). There were no significant differences in the distribution of bootstrap support values for any of the 34 data sets when comparing trees estimated using the $\mathrm{Opt_{BIC}}$ scheme to those estimated using the $\mathrm{Opt_{AICc}}$ or $\mathrm{Adhoc_{feat}}$ schemes (two-sided Mann–Whitney test, $P > 0.05$ in all cases). One data set, Fong_2012, showed a small but significant difference in bootstrap support when comparing the tree inferred with the $\mathrm{Adhoc_{one}}$ partitioning that inferred with the $\mathrm{Opt_{BIC}}$ scheme ($\overline{BS}$ increased from 74.9 to 79.4, two-sided Mann–Whitney test, $P = 0.011$).

Partitioning a given data set with any of the four partitioning schemes produced trees with only very minor differences in mean bootstrap support ($\overline{BS}$, fig. 3A). To assess whether some partitioning schemes have a general tendency to lead to higher or lower bootstrap support across data sets, we calculated three differences in mean bootstrap support ($\Delta\overline{BS}$) for each of the 34 data sets in our study: $\Delta\overline{BS}$ ($\mathrm{Opt_{BIC}}$, $\mathrm{Opt_{AICc}}$), $\Delta\overline{BS}$ ($\mathrm{Opt_{BIC}}$, $\mathrm{Adhoc_{feat}}$), and $\Delta\overline{BS}$ ($\mathrm{Opt_{BIC}}$, $\mathrm{Adhoc_{one}}$). Two-tailed sign tests across all data sets showed no significant bias in the direction of the bootstrap support difference in any of the three comparisons ($P > 0.05$ in all cases).

Properties of Conflicted and Congruent Nodes

To better understand the effects of partitioning on phylogenetic inference, we analyzed the properties of nodes that are the same (congruent) or different (conflicted) in the trees estimated using the four different partitioning schemes we employed in this study. As above, we use the topology estimated using the $\mathrm{Opt_{BIC}}$ partitioning scheme as a reference, and for each of trees estimated using the other three partitioning schemes we define conflicted nodes as those that are not present in the reference tree, and congruent nodes as those that are present in the reference tree.

Conflicted nodes tended to be associated with much shorter and more weakly supported branches than congruent nodes when the data were combined across all 34 data sets ($\overline{BL}_{conflicted} = 0.017$, $\overline{BL}_{congruent} = 0.058$; $\overline{BS}_{conflicted} = 25.6$, $\overline{BS}_{congruent} = 79.2$, Mann–Whitney test, unpaired, two-sided, $P < 0.01$ for the comparison of both branch-length and bootstrap values). This conclusion was not affected by removing the six long-branch outlier data sets ($\overline{BL}_{conflicted} = 0.007$, $\overline{BL}_{congruent} = 0.033$; $\overline{BS}_{conflicted} = 31.4$, $\overline{BS}_{congruent} = 80.3$, Mann–Whitney test, unpaired, two-sided, $P < 0.01$).

The branch-lengths associated with conflicted nodes did not differ significantly between partitioning schemes. Before

6

MBE

Table 3. Change in $\overline{BL}$ and $\overline{BS}$ due to Partitioning Scheme Choice.

Data Set	BL				ΔBL (Opt_BIC other)			ΔBL.25 (Opt_BIC other)			BS				ΔBS (Opt_BIC other)		
	Opt BIC	Opt AICc	Adhoc Feat	Adhoc One	Opt AICc	Adhoc Feat	Adhoc One	Opt AICc	Adhoc Feat	Adhoc One	Opt BIC	Opt AICc	Adhoc Feat	Adhoc One	Opt AICc	Adhoc Feat	Adhoc One
Anderson_2013	0.0465	0.0465	0.0465	0.0243	0.0000	0.0000	0.0222*	0.0000	0.0000	0.0875*	55.6	55.6	55.6	58.8	0.0	0.0	−3.2
Bergsten_2013	0.0965	0.0971	0.0967	0.0712	−0.0005	−0.0001	0.0253	−0.0016	−0.0009	0.0670*	76.5	76.9	77.1	75.4	−0.4	−0.6	1.1
Broughton_2013	0.0937	0.0946	0.0950	0.1092	−0.0009	−0.0012	−0.0155	−0.0019	−0.0015	−0.0345	96.8	96.8	97.0	96.3	<0.0	−0.3	0.5
Brown_2012[a]	0.2273	0.2442	0.2440	0.1537	−0.0169	−0.0167	0.0736*	−0.0407	−0.0401	0.1916*	78.0	77.6	77.6	75.9	0.4	0.4	2.1
Caterino_2001[a]	0.7970	0.7973	0.9601	0.0921	−0.0004	−0.1631	0.7048*	−0.0013	−0.3816	1.6599*	73.4	73.2	73.9	69.6	0.3	−0.4	3.9
Cognato_2001[a]	0.2474	0.2469	0.2459	0.1305	0.0006	0.0015	0.1170*	0.0023	0.0046	0.3031*	73.6	74.2	74.0	73.9	−0.6	−0.4	−0.3
Day_2013	0.0121	0.0120	0.0126	0.0162	0.0001	−0.0005	−0.0040*	0.0004	−0.0016	−0.0115*	80.5	80.5	80.4	80.9	0.0	0.1	−0.4
Devitt_2013	0.0114	0.0087	0.0032	0.0089	0.0026*	0.0082	0.0025	0.0094	0.0295*	0.0092	45.7	45.3	46.1	45.3	0.4	−0.4	0.5
Dornburg_2012	0.0160	0.0166	0.0170	0.0305	−0.0006	−0.0010	−0.0144*	−0.0015	−0.0029	−0.0426*	83.7	85.5	85.6	84.2	−1.8	−2.0	−0.5
Dsouli_2011	0.0727	0.0807	0.0772	0.0526	−0.0080	−0.0045	0.0201	−0.0230	−0.0134	0.0640*	85.8	86.5	86.4	86.0	−0.7	−0.6	−0.3
Ekrem_2010	0.0962	0.0934	0.0996	0.0836	0.0028	−0.0034	0.0126	0.0079	−0.0098	0.0444	83.7	83.8	84.1	81.4	−0.1	−0.4	2.3
Elias_2009	0.0065	0.0065	0.0065	0.0082	>0.0000	<0.0000	−0.0017*	0.0001	−0.0001	−0.0045	73.9	75.0	73.6	76.3	−1.1	0.3	−2.4
Fishbein_2001	0.0276	0.0282	0.0283	0.0252	−0.0006	−0.0007	0.0024	−0.0016	−0.0019	0.0056	88.6	88.1	87.7	87.3	0.5	0.8	1.3
Fong_2012	0.0273	0.0275	0.0237	0.0400	−0.0002	0.0036	−0.0127*	−0.0008	0.0103	−0.0357*	74.7	75.6	74.9	79.4	−0.9	−0.2	−4.7*
Grande_2013	0.0629	0.0625	0.0626	0.0735	0.0003	0.0003	−0.0106	0.0008	0.0007	−0.0215*	76.8	76.1	76.1	74.9	0.7	0.7	1.9
Guschanski_2013	0.0270	0.0272	0.0272	0.0323	−0.0002	−0.0002	−0.0053	−0.0008	−0.0008	−0.0137	92.5	92.4	92.6	92.9	>0.0	−0.1	−0.4
Kaffenberger_2011	0.0530	0.0532	0.0532	0.0655	−0.0002	−0.0002	−0.0125*	−0.0010	−0.0009	−0.0236*	84.5	84.1	84.0	83.0	0.4	0.5	1.5
Kang_2013a	0.0121	0.0119	0.0033	0.0122	0.0003	0.0089*	<0.0000	0.0005	0.0220*	<0.0000	78.6	77.3	78.8	79.2	1.2	−0.2	−0.6
Kang_2013b	0.0280	0.0266	0.0273	0.0269	0.0014	0.0007	0.0011	0.0035	0.0015	0.0045	49.8	52.4	52.0	53.3	−2.6	−2.2	−3.5
Kawahara_2013	0.0375	0.0381	0.0392	0.0241	−0.0006	−0.0017	0.0134	−0.0026	−0.0060	0.0468*	84.4	84.1	84.1	84.5	0.3	0.3	−0.1
Lartillot_2012	0.0450	0.0450	0.0452	0.0519	<0.0000	−0.0002	−0.0069	−0.0003	−0.0006	−0.0165*	94.0	94.1	94.3	91.6	−0.1	−0.3	2.4
Leavitt_2013[a]	0.7181	0.9359	1.1321	0.1298	−0.2178	−0.4140*	0.5883*	−0.6196	−1.1753*	1.6722*	81.0	81.9	82.7	79.4	−1.0	−1.7	1.6
Li_2008	0.0823	0.0827	0.0832	0.1143	−0.0003	−0.0008	−0.0319*	−0.0006	−0.0011	−0.0708*	82.7	82.6	82.4	84.5	0.1	0.2	−1.8
Murray_2013[a]	0.3709	0.1857	0.1857	0.0382	0.1852*	0.1852*	0.3326*	0.5452*	0.5452*	0.9934*	63.8	64.4	64.4	64.2	−0.5	−0.5	−0.4
Rightmyer_2013	0.0775	0.0769	0.0775	0.0328	0.0006	<0.0000	0.0447*	0.0013	−0.0003	0.1285*	53.7	54.1	54.4	53.2	−0.5	−0.8	0.5
Sauquet_2011	0.0156	0.0165	0.0165	0.0166	−0.0009	−0.0009	−0.0010	−0.0034	−0.0034	−0.0005	81.6	81.8	81.8	81.6	−0.3	−0.3	0.0
Seago_2011[a]	0.3088	0.3220	0.6879	0.0864	−0.0132	−0.3791*	0.2224*	−0.0272	−0.8398*	0.4933*	51.8	51.4	51.6	49.6	0.5	0.3	2.2
Sharanowski_2011	0.0574	0.0567	0.0567	0.0819	0.0007	0.0007	−0.0245*	0.0017	0.0017	−0.0571*	74.9	74.7	74.7	71.8	0.2	0.2	3.1
Siler_2013	0.0347	0.0343	0.0346	0.0447	0.0004	0.0001	−0.0100	0.0010	0.0001	−0.0262*	79.8	79.2	79.5	79.3	0.6	0.3	0.5
Tolley_2013	0.0615	0.0644	0.0644	0.0734	−0.0029	−0.0029	−0.0119*	−0.0088	−0.0088	−0.0260*	77.1	76.9	77.0	75.6	0.1	0.1	1.5
Unmack_2013	0.0275	0.0276	0.0276	0.0377	−0.0001	−0.0001	−0.0102*	−0.0006	−0.0005	−0.0305*	85.3	86.2	85.8	86.8	−0.9	−0.5	−1.5
Wainwright_2012	0.0319	0.0334	0.0330	0.0534	−0.0015	−0.0010	−0.0215*	−0.0036	−0.0026	−0.0530*	85.2	85.6	85.2	84.9	−0.4	0.0	0.3
Ward_2010	0.0233	0.0222	0.0223	0.0346	0.0011	0.0010	−0.0113*	0.0028	0.0026	−0.0274*	87.8	88.1	88.2	88.8	−0.2	−0.4	−1.0
Welton_2013	0.0321	0.0322	0.0323	0.0338	<0.0000	−0.0002	−0.0016	−0.0002	−0.0008	−0.0053	74.7	75.2	75.1	74.9	−0.5	−0.5	−0.3

NOTE.—ΔBL and ΔBS show the change in the mean branch-lengths and mean bootstrap values due to partitioning scheme choice, relative to the tree inferred with the Opt_BIC scheme. ΔBL.25 shows the change in the mean branch-lengths of the longest 25% of branches in each tree.

*Significant (P < 0.05) result of Mann–Whitney U test for change in the distribution of branch-lengths relative to the tree inferred with the Opt_BIC scheme. <0.0000 or >0.0000 indicates values were slightly less than or greater than zero, and therefore were not counted as zero in the sign tests for significant.

[a]Data sets defined as having mean branch-lengths that were outliers (>0.15) and were subsequently removed from analyses.
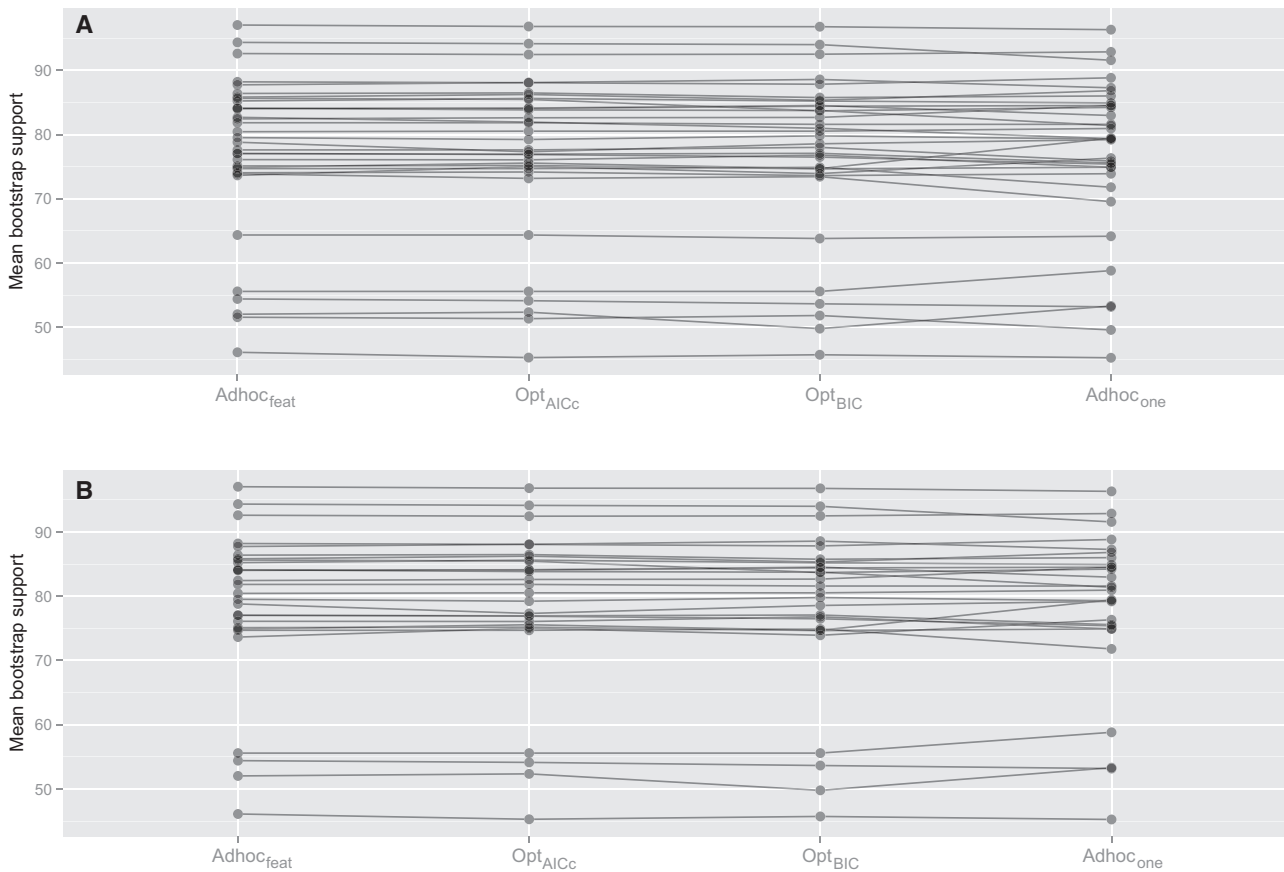
Fig. 3. Bootstrap support is usually consistent when trees are inferred with different partitioning schemes, but varies somewhat for trees inferred with no partitioning. Each panel shows the mean bootstrap support ($\overline{BS}$) for each data set when different partitioning schemes are used for tree inference. Horizontal lines connect the same data set across treatments. (A) includes all 34 data sets, (B) shows only the 28 data sets with $\overline{BL}$ less than 0.15 (a cutoff used to define outliers, see main text).

removing outliers, there was a tendency for conflicted nodes inferred with the $Adhoc_{one}$ scheme to be associated with shorter branches than were conflicted nodes inferred with the Opt_{AICc} or $Adhoc_{feat}$ schemes ($\overline{BL}_{conflicted} = 0.021$, 0.031, and 0.009 in the trees inferred with Opt_{AICc}, $Adhoc_{feat}$, and $Adhoc_{one}$ schemes respectively; Mann–Whitney test, unpaired, two-sided, $P < 0.01$; fig. 4A). However, once the six outlier data sets were removed (fig. 4B), the mean branch-lengths associated with conflicted nodes were more similar across partitioning schemes ($\overline{BL}_{conflicted} = 0.0053$, 0.0053, and 0.0076 in the trees inferred with Opt_{AICc}, $Adhoc_{feat}$, and $Adhoc_{one}$ schemes, respectively) and we found no significant difference in the distributions of branch-lengths associated with conflicted nodes inferred under each of the Opt_{AICc}, $Adhoc_{feat}$, and $Adhoc_{one}$ schemes (two-sided Mann–Whitney test, $P > 0.05$ in all cases).

The bootstrap support values of conflicted nodes inferred under the $Adhoc_{one}$ scheme were significantly higher than the those inferred under both the Opt_{AICc} and $Adhoc_{feat}$ schemes (two-sided Mann–Whitney test, $P < 0.01$ in both cases), whereas the difference between the bootstrap support values of conflicted nodes inferred using the Opt_{AICc} and $Adhoc_{feat}$ schemes was not significant (two-sided Mann–Whitney test, $P > 0.05$). This result remained the same

after removing the six data sets with outlying branch-lengths (fig. 4C and D).

Out of the total of 646 conflicted nodes found across all the data sets, only four were very highly supported (BS > 95) and two of those had maximum bootstrap values of 100. These four highly supported conflicted nodes were all found in topologies inferred with the $Adhoc_{one}$ scheme in two data sets: Ward_2010 and Fong_2012 (fig. 4D). Neither of these two data sets was among the six outliers that were removed in analyses of branch-lengths and bootstrap support.

Discussion

Partitioning is one of the most popular methods used to model the heterogeneity of molecular evolution among sites in an alignment for phylogenetic inference. A number of studies have used simulations to investigate the effects of model misspecification (Lemmon and Moriarty 2004), partitioning scheme choice (Brown and Lemmon 2007), and of the inclusion or absence of rate variation among partitions (Marshall et al. 2006). Additionally, many studies have investigated the effects of partitioning on phylogenetic inference on individual empirical data sets (Brandley et al. 2005; Strugnell et al. 2005; McGuire et al. 2007; Li et al. 2008;
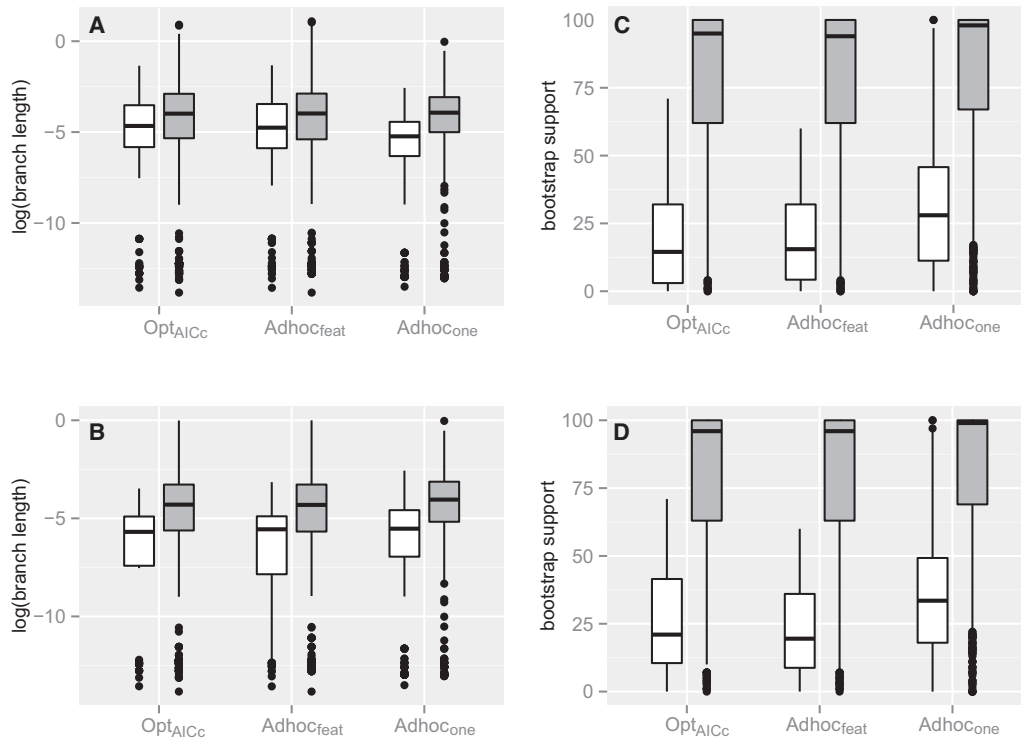
Fig. 4. Conflicts between trees inferred with Opt$_{BIC}$ partitioning and trees inferred with other partitioning schemes mostly occur at nodes with short branch-lengths and low bootstrap support. Plots (A) and (B) show the logarithm of the branch-lengths of conflicted (white) and congruent (gray) branches; plots (C) and (D) show the bootstrap support of conflicted (white) and congruent (gray) branches. Plots (A) and (C) include all branches from all 34 data sets in the study. Plots (B) and (D) show only the branches from the 28 data sets with mean branch-length ($\overline{BL}$) less than 0.15 (a cutoff used to define outliers, see main text). Conflicted and congruent branches are defined as those that disagree and agree with the topology inferred using the Opt$_{BIC}$ scheme, respectively.

Ward et al. 2010; Rota and Wahlberg 2012; Leavitt et al. 2013; Powell et al. 2013; Tao et al. 2013). However, to date there have been no systematic studies of the effects of partitioning on phylogenetic inference across empirical data sets, making it difficult to draw general conclusions. In this study, we compared the effects of four approaches to partitioning across 34 data sets: Using partitioning by genomic features (Adhoc$_{feat}$), using partitioning schemes that are optimized with algorithmic approaches such as those implemented in PartitionFinder (Opt$_{BIC}$ and Opt$_{AICc}$, above), and using a single subset of sites (i.e., no partitioning; Adhoc$_{one}$).

We quantified the effects of partitioning scheme choice on the inference of tree topology, branch-lengths, and bootstrap support in an ML framework. In general, we find that the range of partitioning schemes we compared often produced very similar trees for a given data set, with differences found mostly at nodes with very low support. Underpartitioning appears to produce more severe error than overpartitioning, but it is difficult to predict in advance which data sets will be more affected by partitioning scheme choice than others (supplementary text S1 and fig. S1, Supplementary Material online). We find a handful of cases in which different partitioning schemes lead to very strongly supported differences in tree topologies for the same data set, and so reiterate calls for care to be taken when choosing partitioning schemes for phylogenetic inference. In what follows, we begin by

comparing the results of phylogenetic inference under the three partitioned models we analyzed (Opt$_{BIC}$ Opt$_{AICc}$, Adhoc$_{feat}$) and then move on to discuss the effects of leaving the data unpartitioned (i.e., using Adhoc$_{one}$).

Surprisingly, despite the fact that the two algorithmic approaches (Opt$_{BIC}$ and Opt$_{AICc}$) often led to very large improvements in the information theoretic scores over the highly partitioned approach (e.g., a BIC score improvement of Opt$_{BIC}$ over Adhoc$_{feat}$ of 12,965 and 6,016 for the Fong_2012 and Leavitt_2013 data sets, respectively), the three approaches led to phylogenetic inferences with very few biologically meaningful differences. The inferred topologies were identical in about half of the data sets, while in those data sets where topological differences were found the conflicts were never highly supported and the measured distance between inferred topologies was relatively small. For example, the most strongly supported node that differed between the highly partitioned scheme and the algorithmically optimized scheme using the BIC had a bootstrap support of just 60%. Also, the choice of either a highly partitioned scheme or an algorithmically optimized scheme made little difference to the branch-lengths of the inferred trees in all but one data set (Kang_2013a, table 3), and made no significant difference to bootstrap support in any data sets (table 3). As the highly partitioned Adhoc$_{feat}$ scheme is most likely to be overpartitioned, these results are consistent with the view that

overpartitioning tends to produce less severe error in phylogenetic inference than underpartitioning (Brown and Lemmon 2007; Li et al. 2008).

In contrast to the comparison of partitioned approaches, leaving the data unpartitioned (i.e., using the Adhoc$_{one}$ partitioning scheme) often resulted in large differences in bootstrap values, branch-lengths, and topologies compared with using the algorithmically optimized scheme selected with BIC (i.e., Opt$_{BIC}$). In some data sets the mean bootstrap value changed by up to 4.7%, mean branch-length as much as doubled or halved, whereas the normalized distance between conflicted topologies was often greater than 0.10, indicating that the two trees were more than 10% as different as two randomly selected trees of the same size. Simulation studies have shown that underparameterization often precludes adequate modeling of the variation in rates and patterns of molecular evolution among sites (Lemmon and Moriarty 2004), and that underpartitioning can therefore induce large errors in phylogenetic inference (Brown and Lemmon 2007). In this study, we found that leaving the data unpartitioned led to large and occasionally significant differences in bootstrap support across the whole tree when compared with an algorithmically optimized partitioning scheme (table 3, fig. 3). However, these differences were not biased toward higher or lower bootstrap values when compared across data sets. This suggests, in line with simulation studies on the same topic (Brown and Lemmon 2007), that unpartitioned analyses increase error in bootstrap support values but do not cause predictable biases toward higher or lower values. Leaving the data unpartitioned also led to large and frequently significant differences in branch-lengths between data sets when compared with an algorithmically optimized partitioning scheme (table 2, fig. 2). As with bootstrap support values, we find no evidence of a tendency for unpartitioned analyses to have longer or shorter branch-lengths, suggesting that unpartitioned analyses are more prone to error but not necessarily to systematic bias in branch length estimation across the entire tree. These results differ from a previous simulation study (Brown and Lemmon 2007), which found no such error in branch-length estimates due to either under- or overpartitioning. However, that study simulated data in which evolutionary rates were identical across partitions, which is very different to the case in most empirical data sets. Many other studies, including some simulation studies, have shown that correctly accounting for variation in rates among sites (either through partitioning and/or the use of distributions of rates among sites) is often crucial to accurately estimating molecular branch-lengths (Lemmon and Moriarty 2004; Marshall et al. 2006; Ho and Lanfear 2010). For example, studies have shown that underspecified models tend to underestimate longer and deeper branches of phylogenetic trees (Yang et al. 1994; Lemmon and Moriarty 2004; Phillips 2009), particularly when the model fails to accommodate site rate heterogeneity through the use of a gamma parameter, although some of these effects can be reversed for some small data sets (Revell et al. 2005). We found no evidence that the underpartitioning led to systematic effects on the estimation of either longer or deeper

branches (see table 3, supplementary tables S1 and S2, Supplementary Material online). This might be because these biases are not constant across data sets (Revell et al. 2005). It could also be because all of the partitioning schemes we compared in our study used, at the very least, a model with gamma distributed rates across sites, which may be sufficient to account for systematic biases in estimating longer or deeper branches in some cases (see Lemmon and Moriarty 2004).

In addition to the increased variance in bootstrap values and branch-lengths, the trees inferred from unpartitioned data had the most topological differences when compared with the trees inferred using the Opt$_{BIC}$ scheme. For example, the Adhoc$_{one}$ scheme produced different topologies in 31 out of 34 data sets, and the measured distances between these alternate topologies were typically much greater than distances between topologies from the partitioned approaches (see table 2, fig. 1). Importantly, these differences were sometimes highly supported (fig. 4C and D). On the whole this suggests that severe underpartitioning can sometimes produce large errors in phylogenetic inference.

As alignments increase in length, we expect the degree of model misspecification to increase. This is reflected in the increasing difference in AICc and BIC scores between the four different partitioning schemes as the alignments increase in length (supplementary table S3, Supplementary Material online). Based on this, one might also expect the differences in tree topologies, branch-lengths, and node support to increase with alignment length. However, our data show the opposite pattern (supplementary fig. S2, Supplementary Material online)—the longer the alignment, the less the results depend on the partitioning scheme. We attribute this to two possible factors: 1) Phylogenetic signal presumably increases as data sets get larger, thus larger data sets may converge on the correct tree regardless of the partitioning scheme; and 2) systemic error is likely to increase as data sets get larger (e.g., due to the limitations of using stationary, reversible, and homogeneous models; Galtier and Gouy 1995; Rodríguez-Ezpeleta et al. 2007), thus larger data sets may converge on a single incorrect tree regardless of the partitioning scheme. Whatever the contribution of these two factors, it is clear from our data that their combined effect serves to reduce the effects of commonly used partitioning schemes on phylogenetic inference as data sets increase in size.

Our results suggest that the choice of partitioning scheme tends to predominantly affect nodes that have low bootstrap support and/or are associated with short branches (fig. 4), in concordance with previous studies (Brandley et al. 2005; Li et al. 2008; Ripplinger and Sullivan 2008; Miller et al. 2009; Leavitt et al. 2013). In one sense this is encouraging—if the effects of choosing a suboptimal partitioning scheme are limited to poorly supported nodes, then there is little to worry about as researchers typically do not make strong inferences based on nodes with low support. However, in a few data sets in our study, underpartitioning led to very highly supported but likely incorrect inferences. For example, one of the data sets we analyzed (Fong et al. 2012b) focused on the phylogenetic position of turtles within the amniotes, a subject of long-

term debate (Lee et al. 2004; Crawford et al. 2012). In the original analysis of this data set (see Fong_2012; table 1) turtles were placed as either the sister group to Archosauria or the sister group to Crocodilians (a clade within Archosauria), depending on the data set and partitioning scheme used. After rigorous analyses of many data sets, the authors concluded that the turtle–archosaur relationship was correct, and this conclusion has been corroborated by subsequent studies (Shen et al. 2011; Chiari et al. 2012; Crawford et al. 2012; Fong et al. 2012b). In our analysis of one of their data sets, when partitioned either algorithmically or with a highly partitioned ad hoc scheme, we recovered the turtle–archosaur relationship. However when using an underpartitioned scheme we recovered the incorrect turtle–crocodilian relationship with bootstrap support of 100%. Notably, the authors also recovered this incorrect relationship using the same data set but with an ad hoc partitioning scheme that ignored variation in patterns of molecular evolution between codon positions. Thus, in this case, the use of a poor partitioning scheme could lead to very high confidence in an almost certainly incorrect result. In a similar case, Tao et al. (2013) found that defining data blocks by gene alone for a four-locus data set resulted in a very highly supported but incorrect arrangement of a deep sister-group relationship in a complex Cyprinidae phylogeny. Ward et al. (2010) found in a phylogenetic study of ants that both unpartitioned and slightly partitioned analyses strongly supported incorrect topologies, and that "only partitioning the data both by gene and by codon position within gene reverses this effect" (Ward et al. 2010). Additionally, various studies have shown that failing to account for variation between codon positions can mislead phylogenetic inference (Li et al. 2008; Miller et al. 2009; Ward et al. 2010). These results, in combination with our own, highlight some cause for concern: Phylogenetic inference may be significantly misled when trees are inferred under very poor partitioning schemes.

As the trees inferred with the Adhoc_feat scheme and the two algorithmically optimized schemes are generally very similar, our study suggests that the benefits of algorithmically optimizing partitioning schemes using programs such as PartitionFinder (Lanfear et al. 2012) may be limited, as long as the initial data blocks are sufficiently carefully defined (see also McGuire et al. 2007; Ward et al. 2010; Tao et al. 2013). But this highlights that the a priori definition of data blocks is a fundamental limitation of many current partitioning methods: Defining initial data blocks that are too small risks overpartitioning the data, and defining data blocks that are too large risks underpartitioning the data. In our data (supplementary fig. S1 and text S1, Supplementary Material online) the differences between trees inferred with a highly partitioned scheme and trees inferred with an algorithmically optimized scheme appear to decline as the number of sites per data block increases, and it may be possible to use such measures to help decide whether algorithmic optimization of partitioning schemes is necessary for any given data set, although such analyses may be more time consuming than the algorithmic optimization itself.

Various methods exist that make use of the data to automatically generate data blocks with increased accuracy. Partitioning by automatically grouping sites based on site rates (Kjer et al. 2001; Kjer and Honeycutt 2007) or through Bayesian methods (Wu et al. 2013) are two such approaches that warrant further investigation. We did not apply these methods in our study as many the data sets we analyzed are too large to be processed with many of these methods. Finally, experimental determination of evolutionary models (Bloom 2014) may eventually obviate the need for partitioning in data in certain very well-characterized cases.

Conclusions

Various approaches to partitioning have been established to improve the way evolutionary heterogeneity is accounted for in phylogenetic studies. We analyzed the effects of four common partitioning approaches on a large collection of empirical data sets by comparing the differences in resulting tree topology, branch-lengths, and node support. The most dramatic effect comes from leaving data unpartitioned, resulting in sometimes large and unpredictable variation in inferred branch-lengths, bootstrap support, and topology. Our results highlight that the use of very poor partitioning schemes can occasionally lead to strongly supported but incorrect inferences from phylogenetic analyses. On the other hand, a carefully defined ad hoc partitioning scheme will often perform similarly to algorithmically optimized partitioning schemes. As the algorithmic method for estimating an optimal scheme takes little extra effort, we recommend it over the ad hoc approach as a way to mitigate the risk of overpartitioning and poor parameter estimation.

Materials and Methods

Data Acquisition

We downloaded 34 published data sets used in previous phylogenetic studies from Dryad Digital Repository, Treebase, and other sources, each of which contains a multilocus DNA alignment (table 1). The alignments range from 13 to 2,872 taxa and 823 to 25,919 sites. They include data from the mitochondrial, nuclear, and chloroplast genomes from a diverse array of animals and plants. We avoided data sets where the corresponding published paper noted that the sequence alignment was problematic for phylogenetic inference. Additionally, the inclusion of a data set was blind to whether or not the original study had described any effects of partitioning on phylogenetic inference.

If an alignment contained a group of more than two identical sequences, we randomly removed sequences from this group until there were two remaining. This is because groups of three or more identical sequences have equal likelihoods under a range of different binary tree topologies, potentially confounding our comparisons of ML trees under different partitioning schemes. We also removed nonnucleotide (e.g., morphological) sites from those alignments that contained them. Original and curated alignments are available from Figshare at DOI: 10.6084/m9.figshare.991367.

Partitioning

We partitioned each alignment with PartitionFinder software v 1.1.0 (Lanfear et al. 2012). Partitioning involves two steps: Defining groups of sites that are assumed to have evolved in similar ways; and then choosing an appropriate model of molecular evolution for each group of sites. The first step in partitioning involves the assignment of each site in an alignment to a data block. Data blocks are user-defined sets of sites, typically encompassing distinct DNA features such as genes, intron, exons, and codon positions. Distinct data blocks may be further grouped together if they are deemed to evolve homogeneously. We used data blocks defined by the data sets' original authors where possible. In cases where the original study did not define data blocks in protein-coding genes based on codon positions, we defined these by finding the correct open-reading frame in Geneious version R6 (http://www.geneious.com, Kearse et al., 2012) using the published protein sequence for each gene.

In order to assess the effect of partitioning on tree inference, we defined four partitioning schemes that represent commonly used approaches:

1) $Adhoc_{one}$—this is a user-defined partitioning scheme that assigns every data block to one single group. This is the equivalent of leaving the data unpartitioned as only one substitution model is applied to all sites in the alignment. Substitution model selection and fit was assessed with BIC.

2) $Adhoc_{feat}$—this is a user-defined partitioning scheme that partitions the data set into the maximum possible number of data blocks based on genomic features (genes, codon positions, exons, and introns) and assigns every data block to its own group. In cases where data blocks are defined by genes and codon positions this scheme represents the commonly used gene-by-codon partitioning (Nylander et al. 2004; Shapiro et al. 2006). Substitution model selection and fit was assessed with BIC.

3) Opt_{AICc}—this is an algorithmically optimized partitioning scheme using the data blocks defined in the $Adhoc_{feat}$ scheme as a starting point, with substitution model selection and fit assessed with AICc. We chose to use AICc rather than AIC as it takes into account the size of the alignment when assessing model fit (Burnham and Anderson 2002).

4) Opt_{BIC}—this is an algorithmically optimized partitioning scheme using the data blocks defined in the $Adhoc_{feat}$ scheme as a starting point with substitution model selection and fit assessed with BIC. Using PartitionFinders greedy algorithm this is not guaranteed to be the absolute best scheme possible, but it has been shown to approach the optimal result in most cases.

We configured PartitionFinder for each data set and partitioning scheme. The subset definitions for the $Adhoc_{feat}$ and $Adhoc_{one}$ configurations were applied using the "search = user" setting. The subset definitions for the Opt_{BIC} and Opt_{AICc} configurations were calculated using the greedy algorithm in PartitionFinder by setting "search = greedy." We set "branchlengths = linked" for all analyses. All schemes except the Opt_{AICc} scheme used "model_selection = BIC." With these settings we ran two independent PartitionFinder analyses per partitioning scheme—one with "models = GTR+G" and one with "models = GTR+I+G"—and selected the analysis with the lowest information theoretic score as the best for that scheme. This approach is necessary because these are the only two models implemented in RAxML 7.5., and each predefined subset input to RAxML using the "-q" command must be assigned an independent version of the same model. Partition definitions for RAxML were copy-pasted from the PartitionFinder "best_scheme.txt" file.

Tree Inference

We estimated the ML tree for each of the four partitioning schemes for each data set using RAxML 7.5.3 MPI (Stamatakis 2006), resulting in four ML trees per data set.

We chose to use RAxML as it is the only ML tree inference software capable of analyzing all of the data sets included in this study. It is one of the most popular pieces of phylogenetics software and was the most commonly used software in the 34 studies whose data sets we analyze in this article.

For each data set we used RAxML to perform 96 ML tree searches per analysis, and a full ML bootstrap analysis with 1,024 replicates per analysis. In total, 13,440 ML trees were inferred with 143,360 bootstrap runs with the assistance of the computational facilities of the National Computational Infrastructure, Australia.

Data Analysis

In order to assess the effects of partitioning scheme choice on phylogenetic inference, we focused the analysis on three key metrics: 1) Topology, 2) branch-lengths, and 3) bootstrap support.

For each data set we used R and the "ape" package (Paradis et al. 2004) to read in the four best trees, one for each of the four partitioning schemes. The best tree inferred with the Opt_{BIC} scheme was used as the standard against which we compared the trees inferred with the Opt_{AICc}, $Adhoc_{feat}$, and $Adhoc_{one}$ schemes. Although it would have been possible to perform further pairwise analyses, such as comparisons between trees inferred with the $Adhoc_{one}$ and Opt_{AICc} schemes, we felt this would add unnecessary complexity to the study. We chose the tree inferred with the Opt_{BIC} scheme as the standard for reference. Any of the four trees could have been chosen as the reference, but since models selected with the BIC have been shown to perform equally well as those selected with AIC or AICc (Minin et al. 2003; Abdo et al. 2005), we elected to use those estimated with the algorithmically generated Opt_{BIC} scheme as this scheme is most representative of current best practice in model selection.

All scripts and data files used in the analysis are available for download from Figshare at DOI: 10.6084/m9.figshare.991367

Analysis of Topology

Topological distance metrics provide a measure of the dissimilarity of the branching structure of two or more trees. By comparing the distance between a tree inferred with the Opt_{BIC} partitioning scheme and each of the other three trees, we obtain an indication of how different the relationships among taxa would be under different partitioning schemes. A distance of zero indicates that two trees are identical in topology, though not necessarily with identical branch-lengths. For each data set, we measured the topological distance between the tree inferred with the Opt_{BIC} partitioning scheme and each of the trees inferred with the other three schemes using three metrics: Robinson-Foulds (Robinson and Foulds 1981), Path Difference (Steel and Penny 1993), and Matching Split (Bogdanowicz and Giaro 2012). Robinson-Foulds is a measure of symmetric difference between two trees that counts the absence of nodes in one tree but not the other. Matching Split is also a measure of symmetric difference, similar to Robinson-Foulds, but assigns greater importance to differences at deeper nodes in the tree as opposed to differences closer to the leaves. Path Difference is a measure of the difference in edge-based distances between pairs of taxa. We used Robinson-Foulds as it is the most commonly used metric, whereas Path Difference and Matching Split offer alternative approaches that have more attractive statistical properties, and have been shown to be more commonly in agreement with each other than Robinson-Foulds (Bogdanowicz and Giaro 2012).

Tree comparison values often scale with the number of taxa in the subject trees. As our data sets varied substantially in the number of taxa they contained, we normalized the tree distances to enable reliable comparison of the effects of partitioning on topological distances. We calculated normalized distances using TreeCmp software (Bogdanowicz et al. 2012). TreeCmp calculates a normalized distance for each metric by comparing the distance between two trees to the average distance between randomly generated trees with the same number of taxa. A value of 0 implies that two trees are identical whereas a value of 1 implies that two trees are as dissimilar as two randomly generated trees.

Analysis of Branch-Lengths

We analyzed branch-lengths in two ways: Within and between data sets. Within any given data set, we asked whether the total distribution of branch-lengths changed significantly depending on which of the four partitioning schemes was used to estimate the ML tree. To minimize the total number of tests, we chose the tree estimated using the Opt_{BIC} partitioning scheme as the reference, as above. As a result, we performed three comparisons of branch-length distributions for each of the 34 data sets in our study: Opt_{BIC} versus Opt_{AICc}, Opt_{BIC} versus $Adhoc_{feat}$, and Opt_{BIC} versus $Adhoc_{one}$. Each comparison involved comparing the distribution of branch-lengths from a pair of trees estimated from the same data using a two-sided Mann–Whitney test (R function wilcox.test). Although performing 102 statistical tests is likely to lead to many false positive results, we use the P values from

these tests only to indicate whether different partitioning schemes tend lead to important differences in branch-length distributions.

Our between-data set analysis asks whether, across all 34 data sets in our study, there is a tendency for some partitioning schemes to lead to longer or shorter branch-lengths than other partitioning schemes. To do this, we calculated three differences in mean branch-length for each of the 34 data sets in our study: $\Delta\overline{BL}$ (Opt_{BIC}, Opt_{AICc}), $\Delta\overline{BL}$ (Opt_{BIC}, $Adhoc_{feat}$), and $\Delta\overline{BL}$ (Opt_{BIC}, $Adhoc_{one}$). We then used two-tailed sign tests to ask whether the trees estimated using the Opt_{BIC} scheme tended to have longer or shorter branch-lengths than the trees estimated under any of the other partitioning schemes (i.e., a total of three sign tests).

We repeated the above analyses using only the longest quartile of branches per tree to assess the tendency for underpartitioned models to under- or overestimate longer branches. Finally, we tested for the tendency of underpartitioned models to systematically underestimate the length of branches deeper in the tree. For each data set, the ape package in R was used to extract the depth of each node in the rooted trees inferred with the Opt_{BIC} and $Adhoc_{one}$ schemes. We then fit a linear model to the relationship of node depth and log(BL) for the branches of each tree, and extracted the slopes of the models. We used a two-tailed sign test to determine whether a systematic difference existed between the slopes of the two models across all data sets. Such a tendency would indicate bias in the estimation of branch-lengths at varying node depth due to partitioning scheme choice.

Analysis of Bootstrap Support

The same analysis that was performed for branch-lengths was performed for bootstrap support values, excluding the tests involving the longest quartile of branches and the depth of branches.

Analysis of Conflicted Nodes

A topological distance measure greater than zero indicates that, for a given data set, the choice of partitioning scheme has influenced the presence or absence of certain nodes in the inferred trees. We wished to investigate the characteristics of those nodes in order to understand where and why partitioning has an effect on phylogenetic inference.

We define a conflicted node as a node that is found in a tree estimated with either of the $Adhoc_{one}$, $Adhoc_{feat}$, or Opt_{AICc} schemes but that is not found in the tree estimated with the Opt_{BIC} scheme. For each data set, we used the "prop.clades" method of the ape package in R to extract sets of conflicted nodes and also the converse set of congruent nodes for each data set. To examine the properties of conflicted and congruent nodes, we used a Mann–Whitney test to compare, for each of the $Adhoc_{one}$, $Adhoc_{feat}$, or Opt_{AICc} partitioning schemes, the distributions of branch-lengths and bootstrap support of the conflicted nodes to the distributions of branch-lengths and bootstrap support of the congruent nodes across all data sets.

Finally, we investigated individual data sets that presented any conflicted nodes with very high bootstrap support (> 95) to determine the phylogenetic impact, if any, of the conflict. We used the "Distory" package in R (Chakerian and Holmes 2013) and Archaeopteryx 0.98 visual tree comparison tool (Han and Zmasek 2009) to highlight the nodes and branches with conflicts and to assess their phylogenetic impact with respect to the published phylogeny.

Supplementary Material

Supplementary text S1, tables S1–S3, and figures S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgment

References

Abdo Z, Minin VN, Joyce P, Sullivan J. 2005. Accounting for uncertainty in the tree topology has little effect on the decision-theoretic approach to model selection in phylogeny estimation. *Mol Biol Evol.* 22(3):691–703.

Anderson FE, Bergman A, Cheng SH, Pankey MS, Valinassab T. 2013. Data from: Lights out: the evolution of bacterial bioluminescence in Loliginidae. *Dryad Data Repository.* doi:10.5061/dryad.93s3n.

Anderson FE, Bergman A, Cheng SH, Pankey MS, Valinassab T. 2014. Lights out: the evolution of bacterial bioluminescence in Loliginidae. *Hydrobiologia* 725(1):189–203.

Bergsten J, Nilsson AN, Ronquist F. 2013a. Bayesian tests of topology hypotheses with an example from diving beetles. *Syst Biol.* 62(5):660–673.

Bergsten J, Nilsson AN, Ronquist F. 2013b. Data from: Bayesian tests of topology hypotheses with an example from diving beetles. *Dryad Data Repository.* doi:10.5061/dryad.s631d.

Blair C, Murphy RW. 2011. Recent trends in molecular phylogenetic analysis: where to next? *J Hered.* 102(1):130–138.

Bloom JD. 2014. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Mol Biol Evol.* 31(8):1956–1978.

Bogdanowicz D, Giaro K. 2012. Matching split distance for unrooted binary phylogenetic trees. *IEEE ACM Trans Comput Biol Bioinform.* 9(1):150–160.

Bogdanowicz D, Giaro K, Wróbel B. 2012. Treecmp: comparison of trees in polynomial time. *Evol Bioinform Online* 8:475.

Brandley MC, Schmitz A, Reeder TW. 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Syst Biol.* 54(3):373–390.

Broughton RE, Betancur RR, Li C, Arratia G, Orti G. 2013a. Data from: Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. *Dryad Data Repository.* doi:10.5061/dryad.f1t15.

Broughton RE, Betancur RR, Li C, Arratia G, Orti G. 2013b. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. *PLoS Curr TOL* 5(1).

Brown JM, Lemmon AR. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Syst Biol.* 56(4):643–655.

Brown RM, Siler CD, Das I, Min PY. 2012a. Data from: Testing the phylogenetic affinities of Southeast Asia's rarest geckos: Flap-legged geckos (Luperosaurus), Flying geckos (Ptychozoon), and their relationship to the pan-Asian genus *Gekko*. *Dryad Data Repository.* doi:10.5061/dryad.7bn0fr99.

Brown RM, Siler CD, Das I, Min Y. 2012b. Testing the phylogenetic affinities of Southeast Asia's rarest geckos: Flap-legged geckos (Luperosaurus), Flying geckos (Ptychozoon) and their relationship to the pan-Asian genus *Gekko*. *Mol Phylogenet Evol.* 63(3):915–921.

Buckley TR, Cunningham CW. 2002. The effects of nucleotide substitution model assumptions on estimates of nonparametric bootstrap support. *Mol Biol Evol.* 19(4):394–405.

Buckley TR, Simon C, Chambers GK. 2001. Exploring among-site rate variation models in a maximum likelihood framework using empirical data: effects of model assumptions on estimates of topology, branch lengths, and bootstrap support. *Syst Biol.* 50(1):67–86.

Burnham KP, Anderson DR. 2002. Model selection and multimodel inference : a practical information-theoretic approach. New York: Springer.

Cameron SL, Lo N, Bourguignon T, Svenson GJ, Evans TA. 2012. A mitochondrial genome phylogeny of termites (Blattodea: Termitoidae): robust support for interfamilial relationships and molecular synapomorphies define major clades. *Mol Phylogenet Evol.* 65(1):163–173.

Caterino MS, Reed RD, Kuo MM, Sperling FAH. 2001. A partitioned likelihood analysis of swallowtail butterfly phylogeny (Lepidoptera: Papilionidae). *Syst Biol.* 50(1):106–127.

Chakerian J, Holmes S. 2013. Distory: distance between phylogenetic histories. R package version 1.4.1. http://CRAN.R-project.org/package=distory.

Chiari Y, Cahais V, Galtier N, Delsuc F. 2012. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BCM Biol.* 10(1):65.

Cognato AI, Vogler AP. 2001a. Data from: Exploring data interaction and nucleotide alignment in a multiple gene analysis of Ips (Coleoptera: Scolytinae). *Dryad Data Repository.* doi:10.5061/dryad.678.

Cognato AI, Vogler AP. 2001b. Exploring data interaction and nucleotide alignment in a multiple gene analysis of Ips (Coleoptera: Scolytinae). *Syst Biol.* 50(6):758–780.

Crawford NG, Faircloth BC, McCormack JE, Brumfield RT, Winker K, Glenn TC. 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biol Lett.* 8(5):783–786.

Day JJ, Peart CR, Brown KJ, Bills R, Friel JP, Moritz T. 2013. Data from: Continental diversification of an African catfish radiation (Mochokidae: Synodontis). *Dryad Data Repository.* doi:10.5061/dryad.b6225.

Day JJ, Peart CR, Brown KJ, Friel JP, Bills R, Moritz T. 2013. Continental diversification of an African catfish radiation (Mochokidae: Synodontis). *Syst Biol.* 62(3):351–365.

Devitt TJ, Cameron Devitt SE, Hollingsworth BD, McGuire JA, Moritz C. 2013a. Data from: Montane refugia predict population genetic structure in the Large-blotched Ensatina salamander. *Dryad Data Repository.* doi:10.5061/dryad.k9g50.

Devitt TJ, Cameron Devitt SE, Hollingsworth BD, McGuire JA, Moritz C. 2013b. Montane refugia predict population genetic structure in the Large-blotched Ensatina salamander. *Mol Ecol.* 22(6):1650–1665.

Dornburg A, Moore JA, Webster R, Warren DL, Brandley MC, Iglesias TL, Wainwright PC, Near TJ. 2012a. Data from: Molecular phylogenetics of squirrelfishes and soldierfishes (Teleostei:Beryciformes: Holocentridae): reconciling more than 100 years of taxonomic confusion. *Dryad Data Repository.* doi:10.5061/dryad.3t19n.

Dornburg A, Moore JA, Webster R, Warren DL, Brandley MC, Iglesias TL, Wainwright PC, Near TJ. 2012b. Molecular phylogenetics of squirrelfishes and soldierfishes (Teleostei: Beryciformes: Holocentridae): reconciling more than 100 years of taxonomic confusion. *Mol Phylogenet Evol.* 65(2):727–738.

Dsouli N, Delsuc F, Michaux J, De Stordeur E, Couloux A, Veuille M, Duvallet G. 2011. Phylogenetic analyses of mitochondrial and nuclear data in haematophagous flies support the paraphyly of the genus *Stomoxys* (Diptera: Muscidae). *Infect Genet Evol.* 11(3):663–670.

Ekrem T, Willassen E, Stur E. 2010. Phylogenetic utility of five genes for dipteran phylogeny: a test case in the Chironomidae leads to generic synonymies. *Mol Phylogenet Evol.* 57(2):561–571.

Elias M, Joron M, Willmott K, Kaiser V, Silva-Brandão K, Arias C, Gomez Piñeres LM, Uribe S, Brower A, Freitas A, et al. 2009. Data from: Out of the Andes: patterns of diversification in clearwing butterflies. *Dryad Data Repository*. doi:10.5061/dryad.1705.

Elias M, Joron M, Willmott K, Silva-Brandão KL, Kaiser V, Arias CF, PiÑErez LMG, Uribe S, Brower AVZ, Freitas AVL, et al. 2009. Out of the Andes: patterns of diversification in clearwing butterflies. *Mol Ecol.* 18(8):1716–1729.

Fishbein M, Hibsch-Jetter C, Soltis DE, Hufford L. 2001a. Data from: Phylogeny of Saxifragales (angiosperms, eudicots): analysis of a rapid, ancient radiation. *Dryad Data Repository*. doi:10.5061/dryad.684.

Fishbein M, Hibsch-Jetter C, Soltis DE, Hufford L. 2001b. Phylogeny of Saxifragales (Angiosperms, Eudicots): analysis of a rapid, ancient radiation. *Syst Biol.* 50(6):817–847.

Fong JJ, Brown JM, Fujita MK, Boussau B. 2012a. Data from: A phylogenomic approach to vertebrate phylogeny supports a turtle-archosaur affinity and a possible paraphyletic Lissamphibia. *Dryad Data Repository*. doi:10.5061/dryad.25j6h.

Fong JJ, Brown JM, Fujita MK, Boussau B. 2012b. A phylogenomic approach to vertebrate phylogeny supports a turtle-archosaur affinity and a possible paraphyletic Lissamphibia. *PLoS ONE* 7(11):e48990.

Galtier N, Gouy M. 1995. Inferring phylogenies from DNA sequences of unequal base compositions. *Proc Natl Acad Sci U S A.* 92(24):11317–11321.

Grande T, Borden W, Smith W. 2013a. Limits and relationships of Paracanthopterygii: a molecular framework for evaluating past morphological hypotheses. *Mesozoic Fishes* 5:385–418.

Grande T, Borden WC, Smith WL. 2013b. Data from: Limits and relationships of Paracanthopterygii: a molecular framework for evaluating past morphological hypotheses. *Dryad Data Repository*. doi:10.5061/dryad.k4m8t.

Guschanski K, Krause J, Sawyer S, Valente LM, Bailey S, Finstermeier K, Sabin R, Gilissen E, Sonet G, Nagy ZT, et al. 2013a. Data from: Next-generation museomics disentangles one of the largest primate radiations. *Dryad Data Repository*. doi:10.5061/dryad.7k14q.

Guschanski K, Krause J, Sawyer S, Valente LM, Bailey S, Finstermeier K, Sabin R, Gilissen E, Sonet G, Nagy ZT, et al. 2013b. Next-generation museomics disentangles one of the largest primate radiations. *Syst Biol.* 62(4):539–554.

Han M, Zmasek C. 2009. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* 10(1):356.

Ho SY, Lanfear R. 2010. Improved characterisation of among-lineage rate variation in cetacean mitogenomes using codon-partitioned relaxed clocks. *Mitochondrial DNA* 21(3–4):138–146.

Huelsenbeck JP, Nielsen R. 1999. Variation in the pattern of nucleotide substitution across sites. *J Mol Evol.* 48:86–93.

Kaffenberger N, Wollenberg KC, Köhler J, Glaw F, Vieites DR, Vences M. 2011. Data from: Molecular phylogeny and biogeography of Malagasy frogs of the genus *Gephyromantis*. *Dryad Data Repository*. doi:10.5061/dryad.s791pg03.

Kaffenberger N, Wollenberg KC, Köhler J, Glaw F, Vieites DR, Vences M. 2012. Molecular phylogeny and biogeography of Malagasy frogs of the genus *Gephyromantis*. *Mol Phylogenet Evol.* 62(1):555–560.

Kang JH, Schartl M, Walter R, Meyer A. 2013. Comprehensive phylogenetic analysis of all species of swordtails and platies (Pisces: Genus *Xiphophorus*) uncovers a hybrid origin of a swordtail fish, *Xiphophorus monticolus*, and demonstrates that the sexually selected sword originated in the ancestral lineage of the genus, but was lost again secondarily. *BMC Evol Biol.* 13(1):25.

Kawahara AY, Rubinoff D. 2013a. Convergent evolution of morphology and habitat use in the explosive Hawaiian fancy case caterpillar radiation. *J Evol Biol.* 26(8):1763–1773.

Kawahara AY, Rubinoff D. 2013b. Data from: Convergent evolution in the explosive Hawaiian Fancy Cased caterpillar radiation. *Dryad Data Repository*. doi:10.5061/dryad.gh895.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649.

Kelchner SA, Thomas MA. 2007. Model use in phylogenetics: nine key questions. *Trends Ecol Evol.* 22(2):87–94.

Kjer K, Honeycutt R. 2007. Site specific rates of mitochondrial genomes and the phylogeny of eutheria. *BMC Evol Biol.* 7(1):8.

Kjer KM, Blahnik RJ, Holzenthal RW. 2001. Phylogeny of Trichoptera (Caddisflies): characterization of signal and noise within multiple datasets. *Syst Biol.* 50(6):781–816.

Krajewski C, Fain MG, Buckley L, King DG. 1999. Dynamically heterogenous partitions and phylogenetic inference: an evaluation of analytical strategies with *cytochrome b* and *ND6* gene sequences in cranes. *Mol Phylogenet Evol.* 13(2):302–313.

Kuo MM, Sperling FAH, Caterino MS, Reed RD. 2001. Data from: A partitioned likelihood analysis of swallowtail butterfly phylogeny (Lepidoptera: Papilionidae). *Dryad Data Repository*. doi:10.5061/dryad.615.

Lanfear R, Calcott B, Ho SY, Guindon S. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol.* 29(6):1695–1701.

Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol Biol.* 14(1):82.

Lartillot N, Delsuc F. 2012a. Data from: Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. *Dryad Data Repository*. doi:10.5061/dryad.tt28qk6f.

Lartillot N, Delsuc F. 2012b. Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. *Evolution* 66(6):1773–1787.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol.* 21(6):1095–1109.

Leavitt JR, Hiatt KD, Whiting MF, Song H. 2013. Searching for the optimal data partitioning strategy in mitochondrial phylogenomics: a phylogeny of Acridoidea (Insecta: Orthoptera: Caelifera) as a case study. *Mol Phylogenet Evol.* 67(2):494–508.

Lee MS, Reeder TW, Slowinski JB, Lawson R. 2004. Resolving reptile relationships Assembling the Tree of Life. Oxford: Oxford University Press. p. 451–467.

Lemmon AR, Moriarty EC. 2004. The importance of proper model assumption in Bayesian phylogenetics. *Syst Biol.* 53(2):265–277.

Li C, Lu G, Ortí G. 2008. Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. *Syst Biol.* 57(4):519–539.

Marshall DC, Simon C, Buckley TR. 2006. Accurate branch length estimation in partitioned Bayesian analyses requires accommodation of among-partition rate variation and attention to branch length priors. *Syst Biol.* 55(6):993–1003.

McGuire JA, Witt CC, Altshuler DL, Remsen JV Jr. 2007. Phylogenetic systematics and biogeography of Hummingbirds: Bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy. *Syst Biol.* 56(5):19.

Miller KB, Bergsten J, Whiting MF. 2009. Phylogeny and classification of the tribe Hydaticini (Coleoptera: Dytiscidae): partition choice for Bayesian analysis with multiple nuclear and mitochondrial protein-coding genes. *Zool Scr.* 38(6):591–615.

Minin V, Abdo Z, Joyce P, Sullivan J. 2003. Performance-based selection of likelihood models for phylogeny estimation. *Syst Biol.* 52(5):674–683.

Murray EA, Carmichael AE, Heraty JM. 2013a. Ancient host shifts followed by host conservatism in a group of ant parasitoids. *Proc R Soc Lond B Biol Sci.* 280(1759)

Murray EA, Carmichael AE, Heraty JM. 2013b. Data from: Ancient host shifts followed by host conservatism in a group of ant parasitoids. *Dryad Data Repository.* doi:10.5061/dryad.qn57t.

Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey J. 2004. Bayesian phylogenetic analysis of combined data. *Syst Biol.* 53(1):47–67.

Pagel M, Meade A. 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst Biol.* 53(4):571–581.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2):289–290.

Phillips MJ. 2009. Branch-length estimation bias misleads molecular dating for a vertebrate mitochondrial phylogeny. *Gene* 441(1–2):132–140.

Pol D. 2004. Empirical problems of the hierarchical Likelihood Ratio Test for model selection. *Syst Biol.* 53(6):949–962.

Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst Biol.* 53(5):793–808.

Posada DC, Crandall KA. 2001. Selecting the best-fit model of nucleotide substitution. *Syst Biol.* 50:580–601.

Poux C, Madsen O, Glos J, de Jong WW, Vences M. 2008. Molecular phylogeny and divergence times of Malagasy tenrecs: influence of data partitioning and taxon sampling on dating analyses. *BMC Evol Biol.* 8:102.

Powell AF, Barker FK, Lanyon SM. 2013. Empirical evaluation of partitioning schemes for phylogenetic analyses of mitogenomic data: an avian case study. *Mol Phylogenet Evol.* 66(1):69–79.

Revell LJ, Harmon LJ, Glor RE. 2005. Under-parameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. *Syst Biol.* 54(6):973–983.

Rightmyer MG, Griswold T, Brady SG. 2013a. Data from: Phylogeny and systematics of the bee genus *Osmia* (Hymenoptera: Megachilidae) with emphasis on North American Melanosmia: subgenera, synonymies, and nesting biology revisited. *Dryad Data Repository.* doi:10.5061/dryad.jd5ff.

Rightmyer MG, Griswold T, Brady SG. 2013b. Phylogeny and systematics of the bee genus *Osmia* (Hymenoptera: Megachilidae) with emphasis on North American Melanosmia: subgenera, synonymies and nesting biology revisited. *Syst Entomol.* 38(3):561–576.

Ripplinger J, Sullivan J. 2008. Does choice in model selection affect maximum likelihood analysis? *Syst Biol.* 57(1):76–85.

Robinson D, Foulds LR. 1981. Comparison of phylogenetic trees. *Math Biosci.* 53(1):131–147.

Rodríguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst Biol.* 56(3):389–399.

Rota J, Wahlberg N. 2012. Exploration of data partitioning in an eight-gene data set: phylogeny of metalmark moths (Lepidoptera, Choreutidae). *Zool Scr.* 41(5):536–546.

Sauquet H, Ho SYW, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ, Bayly MJ, Bromham L, Brown GK, Carpenter RJ, et al. 2011. Data from: Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). *Dryad Data Repository.* doi:10.5061/dryad.qq106tm4.

Sauquet H, Ho SYW, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ, Bayly MJ, Bromham L, Brown GK, Carpenter RJ, et al. 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). *Syst Biol.* 61(2):289–313.

Schwarz G. 1978. Estimating the dimension of a model. *Ann Stat.* 6(2):461–464.

Seago AE, Giorgi JA, Li J, Ślipiński A. 2011a. Data from: Phylogeny, classification and evolution of ladybird beetles (Coleoptera: Coccinellidae) based on simultaneous analysis of molecular and morphological data. *Dryad Data Repository.* doi:10.5061/dryad.dc1r2.

Seago AE, Giorgi JA, Li J, Ślipiński A. 2011b. Phylogeny, classification and evolution of ladybird beetles (Coleoptera: Coccinellidae) based on simultaneous analysis of molecular and morphological data. *Mol Phylogenet Evol.* 60(1):137–151.

Shapiro B, Rambaut A, Drummond AJ. 2006. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol.* 23(1):7–9.

Sharanowski BJ, Dowling APG, Sharkey MJ. 2011a. Data from: Molecular phylogenetics of Braconidae (Hymenoptera: Ichneumonoidea)

based on multiple nuclear genes and implications for classification. *Dryad Data Repository.* doi:10.5061/dryad.1688p.

Sharanowski BJ, Dowling APG, Sharkey MJ. 2011b. Molecular phylogenetics of Braconidae (Hymenoptera: Ichneumonoidea), based on multiple nuclear genes, and implications for classification. *Syst Entomol.* 36(3):549–572.

Shen X-X, Liang D, Wen J-Z, Zhang P. 2011. Multiple genome alignments facilitate development of NPCL markers: a case study of tetrapod phylogeny focusing on the position of turtles. *Mol Biol Evol.* 28(12):3237–3252.

Siler C, Brown RM, Oliveros CH, Santanen A. 2013. Data from: Multilocus phylogeny reveals unexpected diversification patterns in Asian Wolf Snakes (genus *Lycodon*). *Dryad Data Repository.* doi:10.5061/dryad.cp6gg.

Siler CD, Oliveros CH, Santanen A, Brown RM. 2013. Multilocus phylogeny reveals unexpected diversification patterns in Asian wolf snakes (genus *Lycodon*). *Zool Scr.* 42(3):262–277.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.

Steel MA, Penny D. 1993. Distributions of tree comparison metrics—some new results. *Syst Biol.* 42(2):126–141.

Strugnell J, Norman M, Jackson J, Drummond AJ, Cooper A. 2005. Molecular phylogeny of coleoid cephalopods (Mollusca: Cephalopoda) using a multigene approach; the effect of data partitioning on resolving phylogenies in a Bayesian framework. *Mol Phylogenet Evol.* 37(2):426–441.

Sullivan J, Joyce P. 2005. Model selection in phylogenetics. *Annu Rev Ecol Evol Syst.* 36(1):445–466.

Sullivan J, Swofford D. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. *J Mamm Evol.* 4(2):77–86.

Tao W, Mayden RL, He S. 2013. Remarkable phylogenetic resolution of the most complex clade of Cyprinidae (Teleostei: Cypriniformes): a proof of concept of homology assessment and partitioning sequence data integrated with mixed model Bayesian analyses. *Mol Phylogenet Evol.* 66(3):603–616.

Tolley KA, Townsend TM, Vences M. 2013a. Data from: Large-scale phylogeny of chameleons suggests African origins and rapid Eocene radiation. *Dryad Data Repository.* doi:10.5061/dryad.11350.

Tolley KA, Townsend TM, Vences M. 2013b. Large-scale phylogeny of chameleons suggests African origins and Eocene diversification. *Proc R Soc Lond B Biol Sci.* 280(1759)

Unmack PJ, Allen GR, Johnson JB. 2013a. Data from: Phylogeny and biogeography of rainbowfishes (Melanotaeniidae) from Australia and New Guinea. *Dryad Data Repository.* doi:10.5061/dryad.qq846.

Unmack PJ, Allen GR, Johnson JB. 2013b. Phylogeny and biogeography of rainbowfishes (Melanotaeniidae) from Australia and New Guinea. *Mol Phylogenet Evol.* 67(1):15–27.

Wainwright PC, Smith WL, Price SA, Tang KL, Sparks JS, Ferry LA, Kuhn KL, Eytan RI, Near TJ. 2012. The evolution of pharyngognathy: a phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. *Syst Biol.* 61(6):1001–1027.

Wainwright PC, Smith WL, Price SA, Tang KL, Sparks JS, Ferry LA, Kuhn KL, Near TJ. 2012. Data from: The evolution of pharyngognathy: a phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. *Dryad Data Repository.* doi:10.5061/dryad.5h951h04.

Ward PS, Brady SG, Fisher BL, Schultz TR. 2010. Phylogeny and biogeography of dolichoderine ants: effects of data partitioning and relict taxa on historical inference. *Syst Biol.* 59(3):342–362.

Welton LJ, Siler CD, Oaks JR, Diesmos AC, Brown RM. 2013a. Data from: Multilocus phylogeny and Bayesian estimates of species boundaries reveal hidden evolutionary relationships and cryptic diversity in Southeast Asian monitor lizards. *Dryad Data Repository.* doi:10.5061/dryad.m0n61.

Welton LJ, Siler CD, Oaks JR, Diesmos AC, Brown RM. 2013b. Multilocus phylogeny and Bayesian estimates of species

boundaries reveal hidden evolutionary relationships and cryptic diversity in Southeast Asian monitor lizards. *Mol Ecol.* 22(13):3495–3510.

Wu C-H, Suchard MA, Drummond AJ. 2013. Bayesian selection of nucleotide substitution models and their site assignments. *Mol Biol Evol.* 30(3):669–688.

Yang Z, Goldman N, Friday A. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Mol Biol Evol.* 11(2):316–324.

Yang Z, Nielsen R, Goldman N, Pedersen A-MK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155(1):431–449.