



## Effects of parameter estimation on maximum-likelihood bootstrap analysis

Jennifer Ripplinger<sup>a,\*</sup>, Zaid Abdo<sup>a,b,c</sup>, Jack Sullivan<sup>a,d</sup>

<sup>a</sup> *Bioinformatics and Computational Biology, University of Idaho, Moscow, ID 83844-3017, USA*

<sup>b</sup> *Department of Mathematics, University of Idaho, Moscow, ID 83844-1103, USA*

<sup>c</sup> *Department of Statistics, University of Idaho, Moscow, ID 83844-1104, USA*

<sup>d</sup> *Department of Biological Sciences, University of Idaho, Moscow, ID 83844-3051, USA*

### ARTICLE INFO

#### Article history:

Received 1 November 2009

Revised 8 April 2010

Accepted 19 April 2010

Available online 29 April 2010

#### Keywords:

Bootstrap

Branch swapping

Heuristic search

Maximum-likelihood

Model selection

### ABSTRACT

Bipartition support in maximum-likelihood (ML) analysis is most commonly assessed using the nonparametric bootstrap. Although bootstrap replicates should theoretically be analyzed in the same manner as the original data, model selection is almost never conducted for bootstrap replicates, substitution-model parameters are often fixed to their maximum-likelihood estimates (MLEs) for the empirical data, and bootstrap replicates may be subjected to less rigorous heuristic search strategies than the original data set. Even though this approach may increase computational tractability, it may also lead to the recovery of suboptimal tree topologies and affect bootstrap values. However, since well-supported bipartitions are often recovered regardless of method, use of a less intensive bootstrap procedure may not significantly affect the results. In this study, we investigate the impact of parameter estimation (i.e., assessment of substitution-model parameters and tree topology) on ML bootstrap analysis. We find that while forgoing model selection and/or setting substitution-model parameters to their empirical MLEs may lead to significantly different bootstrap values, it probably would not change their biological interpretation. Similarly, even though the use of reduced search methods often results in significant differences among bootstrap values, only omitting branch swapping is likely to change any biological inferences drawn from the data.

© 2010 Elsevier Inc. All rights reserved.

### 1. Introduction

The nonparametric bootstrap (Efron, 1979) is a resampling procedure used to quantify the variability associated with a statistical estimate when the underlying distribution is unknown or difficult to calculate. If the data contains sufficient information to provide accurate parameter estimates, it can be resampled as a proxy for sampling from the true distribution. The bootstrap was first applied to phylogenetics by Felsenstein (1985) and has become one of the most common means of assessing support for tree bipartitions. The phylogenetic bootstrap involves resampling the characters of a character by taxon matrix (i.e., an alignment) with replacement to create a series of pseudoreplicates, which are subsequently analyzed in the same manner as the original data. Bootstrap values are equivalent to the frequency with which each node occurs in the resulting set of phylogenetic trees.

Sullivan (2005) noted that although maximum-likelihood (ML) bootstrap replicates should theoretically be analyzed in exactly the same manner as the original data, model selection is virtually never performed for bootstrap replicates and substitution-model parameters are often fixed to their maximum-likelihood estimates

(MLEs) for the empirical data (e.g., Ripplinger and Sullivan, 2008; Stamatakis et al., 2008). In addition, systematists often use less rigorous tree-search strategies to compensate for the computational burden associated with multiple ML analyses (e.g., Liao et al., 2007; Plam et al., 2008). Since inadequate parameter estimation (i.e., estimation of substitution-model parameters and tree topology) could lead to the recovery of suboptimal trees, the use of heuristics in ML analysis could artificially lower bootstrap values. Conversely, the use of a suboptimal substitution model could also lead to the consistent estimation of an inaccurate topology and erroneously increase bootstrap support for particular bipartitions. However, in advocating for the parsimony jackknife, Farris et al. (1996) suggested that there is no need to employ rigorous tree-search methods because well-supported nodes tend to be recovered regardless of the method of inference.

Investigation into the effects of parameter estimation on bootstrap values has been primarily limited to maximum-parsimony (MP) analysis. DeBry and Olmstead (2000) used simulated data to examine the effects of reduced tree search effort and found that the use of nearest-neighbor interchange (NNI) branch swapping and “no swap” methods (i.e., those that rely on a stepwise-addition tree) tended to decrease bootstrap support compared with intensive tree-bisection-reconnection (TBR) branch swapping. These results were especially pronounced for more poorly supported nodes. Similar conclusions for empirical data were reached by

\* Corresponding author. Address: Department of Biological Sciences, University of Idaho, P.O. Box 443051, Moscow, ID 83844-3051, USA.

E-mail address: [jiripplinger@gmail.com](mailto:jiripplinger@gmail.com) (J. Ripplinger).

Mort et al. (2000), who found that “no swap” methods lowered bootstrap values compared with TBR and NNI branch swapping, Sanderson and Wojciechowski (2000), who compared TBR and “no swap” methods, and Salamin et al. (2003), who found that bootstrap values tended to decline with decreased search effort (reviewed by Müller (2005)). However, although the use of less intensive search methods often led to statistically significant differences among bootstrap values, it rarely changed phylogenetic inferences made from the data. Based on the results for MP, it is likely that inadequate parameter estimation has some impact on ML bootstrap values. The purpose of this study is therefore to assess the performance of various model-selection and tree-search strategies on ML bootstrap analysis.

## 2. Materials and methods

### 2.1. Data collection

In order to assess the effects of parameter estimation on a range of phylogenetic hypotheses, we analyzed three empirical data sets whose ML tree contained a combination of long and short branches and whose bootstrap consensus tree contained both strongly and poorly supported bipartitions. All data sets were obtained from

TreeBASE (<http://www.treebase.org>) and included 12 hymenopteran cytochrome oxidase I sequences (Pedersen, 1996), 11 plant RNA polymerase II sequences (Geuten et al., 2007; Nickerson and Drouin, 2004), and 19 *Peromyscus* cytochrome *b* sequences (Sullivan et al., 1997). In addition to empirical data, we analyzed the effects of parameter estimation on a difficult-to-estimate eight-taxon Felsenstein zone tree (Felsenstein, 1978; Sullivan and Swofford, 2001). In order to obtain data for analysis, we used Seq-Gen (Rambaut and Grassly, 1997) to generate a 1000 base pair alignment based on the tree topology, branch lengths, and extreme rate-heterogeneity GTR + I +  $\Gamma$  parameter estimates outlined by Sullivan and Swofford (2001). PAUP\*4.0b10 (Swofford, 2002) was used to divide the resulting data set into three partitions (simulating codon positions) and obtain ML estimates of GTR + I +  $\Gamma$  model parameters for each partition. We then used Seq-Gen to simulate a new alignment for each partition and concatenated the resulting replicates to form a single simulated data set.

### 2.2. Maximum-likelihood analysis

For each data set, we used DT-ModSel (Minin et al., 2003) in conjunction with PAUP\* to select a substitution model with minimum posterior risk. PAUP\* was then used to estimate initial substi-

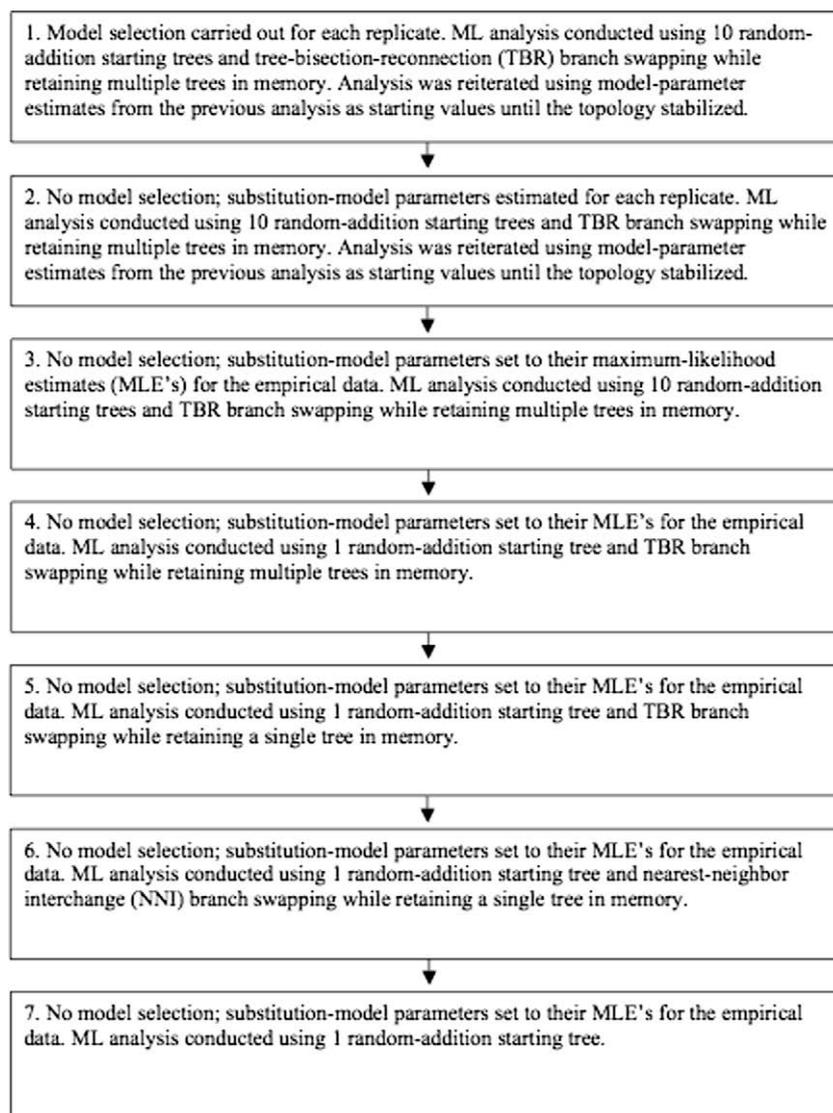


Fig. 1. Seven progressively simpler treatments used for ML bootstrap analysis.

**Table 1**

Results of model selection for 25,000 hymenopteran bootstrap replicates using the decision theory method implemented in DT-ModSel. Substitution models with frequency <0.1% are not shown.

Substitution model	Number of parameters	Frequency (%)
TIM + $\Gamma$	7	47.2
GTR + $\Gamma$	9	24.8
TIM + I + $\Gamma$	8	18.4
TIM + I	7	3.9
GTR + I	9	3.2
GTR + I + $\Gamma$	10	2.5

**Table 2**

Results of Dunnett's test conducted for hymenopteran bootstrap replicates. The most complex treatment, which involved conducting model selection and employing tree-bisection–reconnection (TBR) branch swapping with multiple starting trees for each bootstrap replicate, was designated as the control and compared with six progressively less intensive treatments. The mean bootstrap value obtained from the most intensive treatment is reported for each node. *P*-values less than 0.05 are indicated by an asterisk (\*).

Node	Mean	1 v 2	1 v 3	1 v 4	1 v 5	1 v 6	1 v 7
1	97	*	*	*	*	*	*
2	96						*
3	75	*	*	*	*	*	*
4	75	*	*	*	*	*	*
5	74	*	*	*	*	*	*
6	66	*	*	*	*	*	*
7	39	*	*	*	*	*	*
8	34	*	*	*	*	*	*
9	31	*	*	*	*	*	*

tution-model parameters from a neighbor-joining tree constructed with LogDet distance; these estimates were used as starting values for a ML heuristic search with 10 random-addition starting trees and TBR branch swapping. After the search was complete, we re-optimized substitution-model parameters on the ML tree and performed additional heuristic search iterations until the tree topology stabilized (Ripplinger and Sullivan, 2008; Sullivan et al., 2005). ML trees for each data set have been provided as supplementary material.

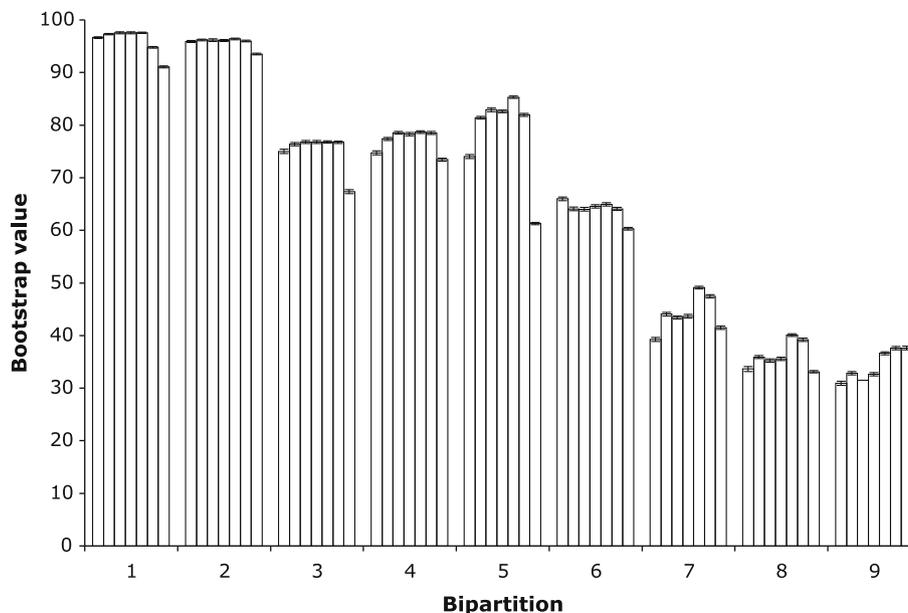
### 2.3. Bootstrap analysis

For each data set, we generated 1000 bootstrap replicates using the Seqboot program in the Phylip package (Felsenstein, 2005). The resulting output was converted to individual nexus files and analyzed with PAUP\* according to the following treatments: (1) replicates were analyzed in the same manner as the original data (i.e., DT-ModSel was used to identify an optimal substitution model for each replicate and replicates were analyzed using an iterated heuristic search strategy with 10 random-addition starting trees and TBR branch swapping), (2) same as the first treatment except replicates were analyzed using the substitution model selected for the original data, (3) same as the second treatment except replicates were analyzed with substitution-model parameters fixed to their MLEs for the original data, (4) same as the third treatment except each search was conducted with one random-addition starting tree, (5) same as the fourth treatment except a single optimal tree was retained during branch swapping, (6) same as the fifth treatment except each search was conducted with NNI branch swapping, and (7) same as the sixth treatment except branch swapping was excluded altogether (i.e., the replicates were analyzed based on a single random-addition tree; for further details, see Fig. 1). The results were summarized as 50% majority-rule consensus trees while retaining bipartitions with less than 50% support that were concordant with well-supported nodes. In order to account for the variance associated with bootstrap resampling, we repeated the bootstrap and subsequent ML analysis 25 times for each of the four data sets, resulting in a total of 100,000 model-selection and 700,000 ML analyses. All data collected as part of this study has been provided as supplementary material.

## 3. Results

### 3.1. Hymenopteran data

The seven-parameter transitional model with  $\Gamma$ -distributed rate variation (TIM +  $\Gamma$ ), which allows for unequal base frequencies, two transversion rates, and separate transition rates, had an optimal fit to the original data and was selected most frequently



**Fig. 2.** Mean bootstrap values and standard errors calculated for hymenopteran bootstrap replicates. Treatments are reported in order of decreasing intensity with the most rigorous method (which entailed conducting model selection for each replicate and employing TBR branch swapping with multiple starting trees) given first and the least intensive method (which involved holding the substitution model constant and forgoing branch swapping altogether) reported last.

**Table 3**

Results of model selection for 25,000 plant bootstrap replicates conducted using DT-ModSel. Substitution models with frequency <0.1% are not shown.

Substitution model	Number of parameters	Frequency (%)
GTR + I + $\Gamma$	10	82.1
TVM + I + $\Gamma$	9	17.8

**Table 4**

Results of Dunnett's test carried out for plant bootstrap replicates. *P*-values less than 0.05 are indicated by an asterisk (\*); for additional analytical details, see Table 2.

Node	Mean	1 v 2	1 v 3	1 v 4	1 v 5	1 v 6	1 v 7
1	100						
2	100						
3	100						*
4	94						*
5	90					*	*
6	88					*	*
7	85					*	*
8	54						*

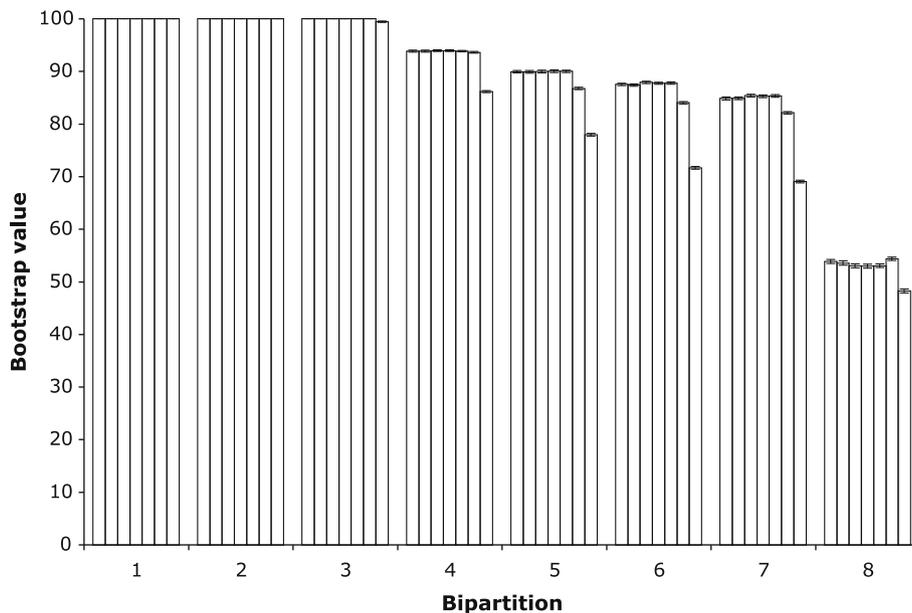
for the bootstrap replicates (~47%). Six additional substitution models with seven or more parameters were selected for the bootstrap replicates, all of which incorporated unequal base frequencies, independent transition rates (as well as additional rate classes), and among-site rate variation (ASRV) by including  $\Gamma$ -distributed rate variation, a proportion of invariable sites, or a combination of the two (Table 1). We compared the most complex treatment to each of the six less intensive treatments using Dunnett's test and found that forgoing model selection and using a reduced search strategy normally resulted in significantly different bootstrap values (evaluated at  $\alpha = 0.05$ ). This did not occur, however, for one strongly supported node ( $\bar{x} = 96\%$ ), for which only the "no swap" treatment led to a significant decrease in bootstrap support (Table 2). Nevertheless, forgoing model selection and/or setting substitution-model parameters to their MLEs for the original data resulted in less than a 3% average difference in bootstrap support ( $\bar{x} = 2.6\%$  and  $2.7\%$ , respectively). Use of less rigorous search strategies resulted in larger differences among bootstrap values ( $\bar{x} = 4.2\%$  for the NNI and "no swap" treatments), however, bootstrap support only declined more than 10% for a single node (Fig. 2).

### 3.2. Plant data

The 10-parameter GTR + I +  $\Gamma$  model, which incorporates unequal base frequencies and independent substitution rates, had minimum posterior risk when selected for the original data; this model was also selected much more frequently than other substitution models for the bootstrap replicates (~82%). Although six additional models were chosen for the bootstrap replicates, only the transversal model with  $\Gamma$ -distributed rate variation and invariable sites (TVM + I +  $\Gamma$ ), which allows for unequal base frequencies, one transition rate, and separate transversion rates, was selected for more than 1% of the remaining replicates (Table 3). Bipartitions that received 100% support using the most intensive treatment tended to be recovered for all bootstrap replicates using less rigorous methods, which is consistent with results obtained by DeBry and Olmstead (2000) for MP analysis. Only the use of the "no swap" and, in some instances, NNI branch swapping methods led to significantly different bootstrap values (Table 4); use of all other treatments changed the resulting bootstrap values by less than 1% ( $\bar{x} = 0.3\%$ ). However, while the use of NNI branch swapping led to less than a 2% average difference in bootstrap support, forgoing branch swapping resulted in an average difference of ~9% and a greater than 12% decrease for three nodes (Fig. 3).

### 3.3. Peromyscus data

The five-parameter HKY +  $\Gamma$  model, which incorporates unequal base frequencies and separate transition/transversion rates, had an optimal fit to the original data and was the most commonly selected model for the bootstrap replicates (~56%). Nineteen additional substitution models, almost all with five or more parameters, were selected for the bootstrap replicates, but many were selected with low frequency. All of the models selected for the bootstrap replicates incorporated ASRV and almost all included unequal base frequencies (Table 5). Nodes inferred with 100% support using the most rigorous treatment were again assigned 100% support using less intensive methods. Forgoing branch swapping resulted in significantly different bootstrap values for the majority of bipartitions (~69%) while the use of NNI branch swapping and other reduced search methods resulted in significantly different values for less than 50% of the nodes; setting substitution-model



**Fig. 3.** Mean bootstrap values and standard errors calculated for plant bootstrap replicates. For additional details, see Fig. 1.

**Table 5**

Results of model selection for 25,000 *Peromyscus* bootstrap replicates conducted using DT-ModSel. Substitution models with frequency <0.1% are not shown.

Substitution model	Number of parameters	Frequency (%)
HKY + $\Gamma$	5	55.6
K3Puf + $\Gamma$	6	19.4
HKY + I	5	12.3
K3Puf + I	6	3.6
TrN + $\Gamma$	6	2.4
HKY + I + $\Gamma$	6	1.8
TVM + $\Gamma$	8	1.1
TIM + $\Gamma$	7	1.1
TrN + I	6	1.0
K3Puf + I + $\Gamma$	7	0.7
TIM + I	7	0.3
GTR + $\Gamma$	9	0.3
TrN + I + $\Gamma$	7	0.2
TIM + I + $\Gamma$	8	0.2
GTR + I	9	0.1
TVM + I + $\Gamma$	9	0.1

**Table 6**

Results of Dunnett's test conducted for *Peromyscus* bootstrap replicates. *P*-values less than 0.05 are indicated by an asterisk (\*); see Table 2 for additional details.

Node	Mean	1 v 2	1 v 3	1 v 4	1 v 5	1 v 6	1 v 7
1	100						
2	100						
3	99					*	*
4	97		*	*	*	*	*
5	80						*
6	72				*	*	*
7	70						*
8	69						*
9	61				*	*	*
10	60	*	*	*	*	*	*
11	60				*	*	*
12	59						*
13	58						*
14	55						*
15	52				*	*	*
16	1						*

parameters to their MLEs for the original data led to a significant difference in bootstrap support for one bipartition (Table 6). Use of all methods besides the NNI and “no swap” treatments resulted

**Table 7**

Results of model selection for 25,000 bootstrap replicates from simulated data conducted using DT-ModSel. Substitution models with frequency <0.1% are not shown.

Substitution model	Number of parameters	Frequency (%)
HKY + $\Gamma$	5	41.2
TVM + $\Gamma$	8	18.9
HKY + I + $\Gamma$	6	8.9
K3Puf + $\Gamma$	6	7.8
TrN + $\Gamma$	6	6.9
GTR + $\Gamma$	9	6.5
TVM + I + $\Gamma$	9	5.2
K3Puf + I + $\Gamma$	7	1.6
TIM + $\Gamma$	7	1.5
GTR + I + $\Gamma$	10	0.8
TrN + I + $\Gamma$	7	0.5
TIM + I + $\Gamma$	8	0.1

**Table 8**

Results of Dunnett's test conducted for bootstrap replicates from simulated data. *P*-values less than 0.05 are indicated by an asterisk (\*); for additional analytical details, see Table 2.

Node	Mean	1 v 2	1 v 3	1 v 4	1 v 5	1 v 6	1 v 7
1	100						*
2	98				*	*	*
3	86	*	*	*	*	*	*
4	72				*	*	*
5	61		*	*	*	*	*

in less than a 1% difference in average bootstrap support while use of the NNI method resulted in an  $\sim 1.5\%$  average difference in support. However, forgoing branch swapping led to an  $\sim 4\%$  average difference in bootstrap values with an  $\sim 12\%$  difference for two of the nodes (Fig. 4).

### 3.4. Simulation data

The five-parameter HKY +  $\Gamma$  model had minimum posterior risk when selected for the original data and was chosen most frequently for the bootstrap replicates ( $\sim 41\%$ ). Substitution models selected for the bootstrap replicates generated from the simulated

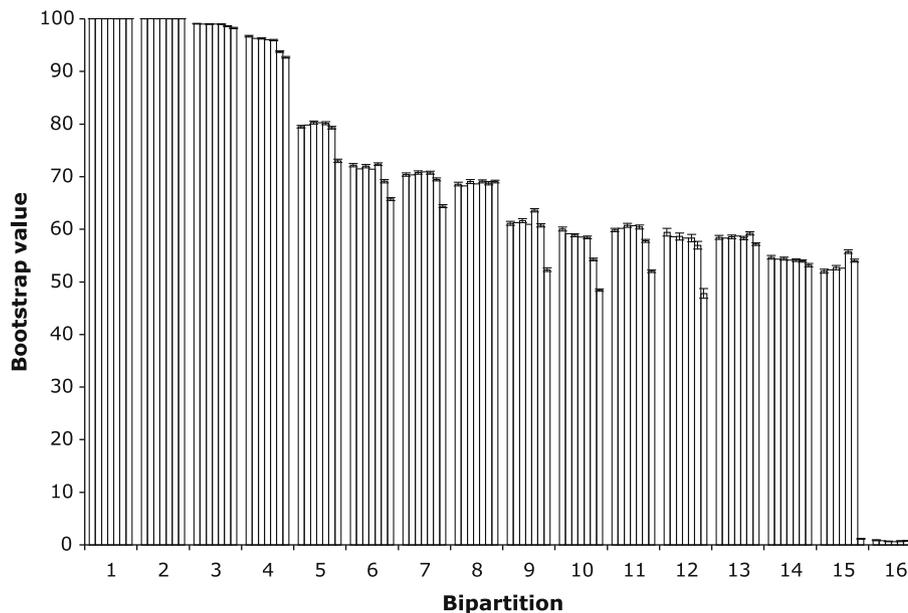


Fig. 4. Mean bootstrap values and standard errors calculated for *Peromyscus* bootstrap replicates. See Fig. 1 for further details.

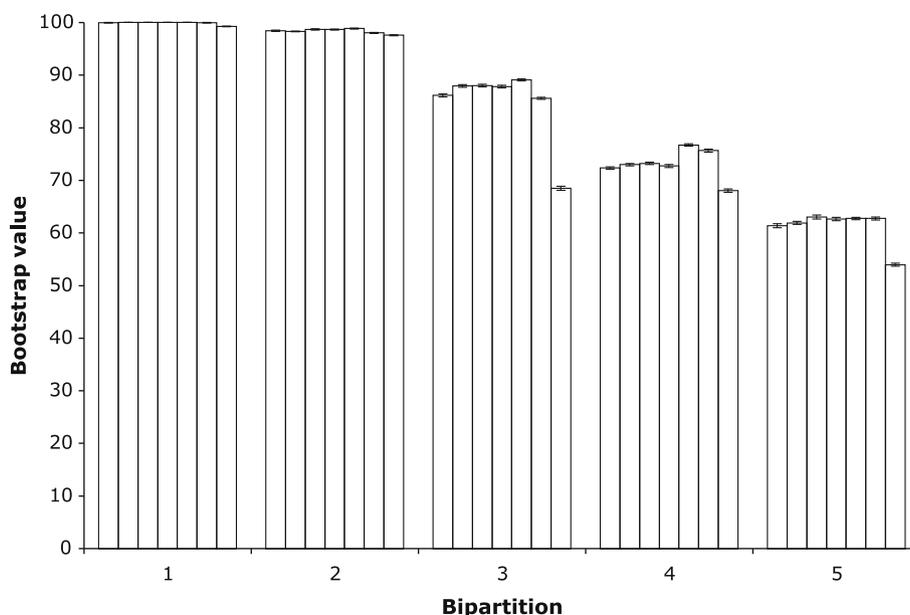


Fig. 5. Mean bootstrap values and standard errors calculated for bootstrap replicates generated from simulated data. For additional details, see Fig. 1.

data were more diverse than those chosen for the empirical data sets. Nineteen additional models were selected for the bootstrap replicates; most of these models incorporated five or more parameters and almost all incorporated unequal base frequencies and ASRV (Table 7). Forgoing branch swapping always led to significantly different bootstrap values while the use of reduced search methods and fixing substitution-model parameters to their MLEs for the original data led to significantly different values for approximately half of the nodes. Unlike the plant and *Peromyscus* data sets, using the substitution model selected for the empirical data and optimizing the substitution-model parameters on each bootstrap replicate resulted in a significantly different bootstrap value for one of the nodes (Table 8). Nonetheless, use of all methods besides the “no swap” treatment led to less than an ~2% average difference in bootstrap values. However, forgoing branch swapping resulted in an ~6% average difference in bootstrap support and a greater than 17% decrease for one bipartition (Fig. 5).

#### 4. Discussion

Even though forgoing model selection and setting substitution-model parameters to their MLEs based on the original data may lead to statistically different bootstrap values, these differences tend to be minor and probably would not change the biological interpretation of the results. The substitution model chosen for the original data tends to be selected most frequently for the bootstrap replicates and alternative models chosen for the replicates normally incorporate a similar number of parameters and almost always account for ASRV. Consequently, model selection appears unnecessary for adequate bootstrap analysis, which is reassuring since systematists rarely, if ever, conduct model selection for individual bootstrap replicates.

Although using less intensive bootstrap procedures may lead to significantly different bootstrap values, these differences are often small, especially for treatments that utilize some form of branch swapping. These results are similar to those obtained by Mort et al. (2000) for MP bootstrap analysis. Although the use of reduced search methods often changes the resulting bootstrap values, only the use of the “no swap” method leads to significantly different bootstrap values for more than 75% of the bipartitions and results

in larger differences among bootstrap values than more rigorous methods. As a result, it appears prudent to use some form of branch swapping when analyzing ML bootstrap replicates, even if computational constraints preclude using an intensive method.

#### Acknowledgments

This research is part of the University of Idaho Initiative in Bioinformatics and Evolutionary Studies (IBEST); funding for the IBEST Bioinformatics Core is provided by NIH/NCRR Grants P2ORR16448 and P2ORR016454. We would like to thank C. Brown and an anonymous reviewer for their comments that helped improve this manuscript and our systems administrators for their assistance in using the IBEST Beowulf clusters. We would also like to thank J. Evans for providing us with a script to convert the output from Phylip's Seqboot program to individual nexus files.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ymp.2010.04.035.

#### References

- DeBry, R.W., Olmstead, R.G., 2000. A simulation study of reduced tree-search effort in bootstrap resampling analysis. *Syst. Biol.* 49, 171–179.
- Efron, B., 1979. Bootstrap methods: another look at the jackknife. *Ann. Stat.* 7, 1–26.
- Farris, J.S., Albert, V.A., Källersjö, M., Lipscomb, D., Kluge, A.G., 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12, 99–124.
- Felsenstein, J., 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* 27, 401–410.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- Felsenstein, J., 2005. PHYLIP (Phylogeny Inference Package). Version 3.6. Distributed by the Author. Department of Genome Sciences, University of Washington, Seattle, Washington.
- Geuten, K., Massingham, T., Darius, P., Smets, E., Goldman, N., 2007. Experimental design criteria in phylogenetics: where to add taxa. *Syst. Biol.* 56, 609–622.
- Liao, W.-J., Yuan, Y.-M., Zhang, D.-Y., 2007. Biogeography and evolution of flower color in *Veratrum* (Melanthiaceae) through inference of a phylogeny based on multiple DNA markers. *Plant Syst. Evol.* 267, 177–190.
- Minin, V., Abdo, Z., Joyce, P., Sullivan, J., 2003. Performance-based selection of likelihood models for phylogeny estimation. *Syst. Biol.* 52, 1–10.
- Mort, M.E., Soltis, P.S., Soltis, D.E., Mabry, M.L., 2000. Comparison of three methods for estimating internal support on phylogenetic trees. *Syst. Biol.* 49, 160–171.

- Müller, K.F., 2005. The efficiency of different search strategies in estimating parsimony jackknife, bootstrap, and Bremer support. *BMC Evol. Biol.* 5, 58.
- Nickerson, J., Drouin, G., 2004. The sequence of the largest subunit of RNA polymerase II is a useful marker for inferring seed plant phylogeny. *Mol. Phylogenet. Evol.* 31, 403–415.
- Pedersen, B.V., 1996. A phylogenetic analysis of cuckoo bumblebees (*Psithyrus*, Lepeletier) and bumblebees (*Bombus*, Latreille) inferred from sequences of the mitochondrial gene cytochrome oxidase I. *Mol. Phylogenet. Evol.* 5, 289–297.
- Plam, M., Jørgensen, A., Kristensen, T.K., Madsen, H., 2008. Sympatric *Biomphalaria* species (Gastropoda: Planorbidae) in Lake Albert, Uganda, show homoplasies in shell morphology. *Afr. Zool.* 43, 34–44.
- Rambaut, A., Grassly, N.C., 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* 13, 235–238.
- Ripplinger, J., Sullivan, J., 2008. Does choice in model selection affect maximum likelihood analysis? *Syst. Biol.* 57, 76–85.
- Salamín, N., Chase, M.W., Hodkinson, T.R., Savolainen, V., 2003. Assessing internal support with large phylogenetic DNA matrices. *Mol. Phylogenet. Evol.* 27, 528–539.
- Sanderson, M.J., Wojciechowski, M.F., 2000. Improved bootstrap confidence limits in large-scale phylogenies, with an example from Neo-Astragalus (Leguminosae). *Syst. Biol.* 49, 671–685.
- Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* 57, 758–771.
- Sullivan, J., 2005. Maximum-likelihood methods for phylogeny estimation. *Methods Enzymol.* 395, 757–779.
- Sullivan, J., Swofford, D.L., 2001. Should we use model-based methods for phylogenetic inference when we know that assumptions about among-site rate variation and nucleotide substitution pattern are violated? *Syst. Biol.* 50, 723–729.
- Sullivan, J., Markert, J.A., Kilpatrick, C.W., 1997. Phylogeography and molecular systematics of the *Peromyscus aztecus* species group (Rodentia: Muridae) inferred using parsimony and likelihood. *Syst. Biol.* 46, 426–440.
- Sullivan, J., Abdo, Z., Joyce, P., Swofford, D.L., 2005. Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation. *Mol. Biol. Evol.* 22, 1386–1392.
- Swofford, D.L., 2002. PAUP\*. Phylogenetic Analysis Using Parsimony (\* and Other Methods). Version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.