Molecular Phylogenetics and Evolution 67 (2013) 429–435

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Assessing statistical reliability of phylogenetic trees via a speedy double bootstrap method

Aizhen Ren *, Takashi Ishida, Yutaka Akiyama

Graduate School of Information Science and Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan

ARTICLE INFO

Article history:
Received 27 August 2012
Revised 17 January 2013
Accepted 8 February 2013
Available online 26 February 2013

Keywords:
Phylogenetic tree
Maximum likelihood
Tree selection
Bootstrap probability
Double bootstrap
Rapid computation

ABSTRACT

Evaluating the reliability of estimated phylogenetic trees is of critical importance in the field of molecular phylogenetics, and for other endeavors that depend on accurate phylogenetic reconstruction. The bootstrap method is a well-known computational approach to phylogenetic tree assessment, and more generally for assessing the reliability of statistical models. However, it is known to be biased under certain circumstances, calling into question the accuracy of the method. Several advanced bootstrap methods have been developed to achieve higher accuracy, one of which is the double bootstrap approach, but the computational burden of this method has precluded its application to practical problems of phylogenetic tree selection. We address this issue by proposing a simple method called the speedy double bootstrap, which circumvents the second-tier resampling step in the regular double bootstrap approach. We also develop an implementation of the regular double bootstrap for comparison with our speedy method. The speedy double bootstrap suffers no significant loss of accuracy compared with the regular double bootstrap, while performing calculations significantly more rapidly (at minimum around 371 times faster, based on analysis of mammalian mitochondrial amino acid sequences and 12S and 16S rRNA genes). Our method thus enables, for the first time, the practical application of the double bootstrap technique in the context of molecular phylogenetics. The approach can also be used more generally for model selection problems wherever the maximum likelihood criterion is used.

1. Introduction

The analytical methods used in the field of molecular phylogenetics are important basic tools for reconstructing the evolutionary history (phylogenetic relationships) of molecules and organisms. Molecular phylogenetic methods are primarily used in the context of biological systematics, but they find applications in a wide variety of other fields in addition, as diverse as community ecology (Webb et al., 2002), biogeography (Wiley, 1981) and proteomics, including the inference of protein–protein interactions (Pazos and Valencia, 2001) similarity. Many methods of phylogenetic reconstruction have been developed and are in regular use (Felsenstein, 2004). However, those based on maximum likelihood estimation have proved most effective for reconstructing phylogenies using molecular sequence data (DNA, protein, etc.). Early work on this application of maximum likelihood was conducted by Felsenstein (1981), whose approach involved computing the maximum likelihood value for many topologies, and selecting the topology with the highest likelihood (the maximum likelihood (ML) tree) as the most probable candidate for the true topology.

It must be noted that the maximum likelihood values are dependent on the particular characteristics of a random variable: the molecular sequences that constitute the underlying data for phylogeny reconstruction. Thus, some analysis of the statistical reliability of the estimated ML tree or multiple alternative trees should be undertaken. Statistical hypothesis testing is commonly used for this purpose, and the bootstrapping technique is a well-known computational method for calculating reliability when a simple mathematical formula is difficult to derive. Bootstrapping is a resampling method that approximates a random sample by creating a bootstrap sample, generated by random sampling with replacement from the original single data set. In the context of phylogenetic tree selection, Felsenstein (1985) proposed the use of bootstrapping to place confidence intervals on phylogenies. He defined the *p*-value of a tree according to a frequency called the bootstrap probability (BP); the proportion of bootstrap pseudoreplicates of the original data set in which the tree is found to be optimal. However, it is known that under some circumstances the naive bootstrap probability can be biased (e.g., Hillis and Bull, 1993; Sanderson and Wojciechowski, 2000). Thus, some advanced bootstrap methods have been proposed, to achieve higher accuracy (Hall, 1992; Efron et al., 1996; Efron and Tibshirani, 1998; Shimodaira, 2002). Among these, the double bootstrap (Hall, 1992; Efron and Tibshirani, 1998) has been shown to be third order accurate

* Corresponding author. Fax: +81 3 5734 3646.
E-mail address: ren@bi.cs.titech.ac.jp (A. Ren).

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.02.011

and may hold great potential as a measure of phylogenetic tree support. However, the method imposes huge computation burdens and has yet to be applied in the context of molecular phylogenetics. To overcome this computational difficulty we propose a speedy double bootstrap method to compute the reliability of phylogenetic trees. For comparison, we also developed a procedure to implement the regular double bootstrap (Hall, 1992; Efron and Tibshirani, 1998), and we used these methods to analyze the mammalian mitochondrial protein sequences and genes for 12S and 16S rRNA. To illustrate the utility of our speedy double bootstrap, we compared results from this method with those from the regular double bootstrap, the traditional bootstrap proportion (BP), and the multiscale bootstrap technique (AU test) described by Shimodaira (2002).

2. Materials and methods

2.1. The double bootstrap method

In this study, homologous sites of aligned molecular sequence data are regarded as the units of sampling, and we use DNA data as the example for the following methodological descriptions. Suppose we have m homologous sequences, each with n nucleotide sites. These data can be represented as a $m \times n$ matrix $\mathbf{X} = \{x_{jh}\} = \{\mathbf{x_1}, \ldots, \mathbf{x_n}\}$, where $\mathbf{x}_h$ is the value of the h-th site and x_{jh} is one of the four deoxyribonucleotides (T, C, A, or G).

$$\begin{aligned} &Species1 : x_{11}\ x_{12}\ \cdots\ x_{1n} \\ &Species2 : x_{21}\ x_{22}\ \cdots\ x_{2n} \\ &\quad\vdots\ \ \vdots \\ &Speciesm : x_{m1}\ x_{m2}\ \cdots\ x_{mn} \end{aligned} \tag{1}$$

The log-likelihood can be expressed as

$$l(\theta; \mathbf{X}) = \sum_{h=1}^{n} logf(\mathbf{x}_h; \theta) \tag{2}$$

where $f(\mathbf{x}_h;\ \theta) = f(x_{1h}, x_{2h}, \ldots, x_{mh};\ \theta)$ is the probability that at a particular homologous site, species 1 has base x_{1h}, species 2 has x_{2h} and species m has x_{mh}. The vector θ denotes unknown parameters such as the edge lengths (branch lengths) of a tree, and the base substitution rates along these branches. Here we assume that the base substitution rates have already been estimated, so θ denotes only the unknown edge lengths. For a given tree topology, θ is estimated by maximizing the log-likelihood, and the maximum log-likelihood of any tree topology i is given by

$$l_i(\hat{\theta}_i; \mathbf{X}) = \sum_{h=1}^{n} logf_i(\mathbf{x}_h; \hat{\theta}_i) \tag{3}$$

The topology with the highest value of $l(\hat{\theta};\ \mathbf{X})$ is the maximum likelihood phylogeny (T_{ML}) for data set $\mathbf{X}$, and is thus the most likely candidate for the true topology. To define null hypotheses for performing model comparisons, we must first recognize that molecular sequence data are discretely distributed; the true distribution for a random variable $\mathbf{x}$ can be expressed as

$$q(\cdot) = \{q(\mathbf{x}_1), q(\mathbf{x}_2), \ldots, q(\mathbf{x}_s)\} \tag{4}$$

where $s = 4^m$and is the expectation of $l_i(\hat{\theta}_i;\ \mathbf{X})$ with respect to $q(\cdot)$, $i = 1, \ldots, K$, i.e.

$$\mu_i = E_q[l_i(\hat{\theta}_i; \mathbf{X})] = \sum_{h=1}^{n} E_q[logf_i(\mathbf{x}_h; \hat{\theta}_i)] = nE_q[logf_i(\mathbf{x}; \hat{\theta}_i)] \tag{5}$$

where $E_q[logf_i(\mathbf{x}; \hat{\theta}_i)] = \sum_{t=1}^{s} q(\mathbf{x_t}) logf_i(\mathbf{x}_t; \hat{\theta}_i)$, for each $i = 1, \ldots, K$. So if we assume that tree T_1 is the best topology, the null and alternative hypotheses will be

$$H_1 : \mu_1 = max_{i=1,\ldots,K} \mu_i \ \ vs.\ H_1^A :\ others \tag{6}$$

and we must continue performing these comparisons as many times as is necessary, assuming in turn that tree T_i, $i = 2, \ldots, K$ is the best topology. Note that the null hypothesis H_1 involves multiple comparisons with the "best" topology (Hsu, 1981): as can be seen from (6), the null contains $k - 1$ hypotheses such that

$$H_{1j} : \mu_1 \geqslant \mu_j,\ j = 2, \ldots, K, \tag{7}$$

The null hypothesis H_1 is a polyhedral convex cone and ∂H_1, which is boundary of H_1 is nonsmooth at the vertex as well as on the faces of dimensions less than $K - 1$. Shimodaira and Hasegawa (1999) proposed a multiple comparisons procedure (the SH-test) to test H_1, but this was shown to be overly conservative and a different method was designed (the AU test), which uses a multiscale bootstrap technique to obtain third-order accurate p-values for testing the null hypothesis. Other authors (e.g., Hall, 1992; Efron and Tibshirani, 1998) had previously developed a double bootstrap method that was also able to provide third-order accurate p-values, but due to high computational requirements this method has not been adopted for phylogenetic applications.

At this juncture it is necessary to briefly review the double bootstrap method. The third-order accurate p-values was first proposed by Efron (1985) for the multivariate normal model, which can be represented as

$$\mathbf{Y} \overset{i.i.d.}{\sim} N_t(\eta, I_t) \tag{8}$$

This normal model is a simplification of reality. Let $\mathcal{H} \subset \mathbb{R}^t$ be an arbitrarily-shaped region with smooth boundaries denoted by $\partial\mathcal{H}$. We want to calculate a p-value $p(y)$ for testing the null hypothesis $\eta \in \mathcal{H}$. According to Efron (1985), when the true parameter η is on the boundary surface $\partial\mathcal{H}$, the third-order accurate p-value can be expressed as

$$p(y) = 1 - \Phi(d - c) \tag{9}$$

where d is the signed distance from y to $\hat{\eta}(y)$, with a positive or negative sign when y is, respectively, outside or inside $\mathcal{H}$. The point $\hat{\eta}(y)$ is the closest point to y (in Euclidean distance) on the surface $\partial\mathcal{H}$, and c in formula (9) is a quantity related to the curvature of $\partial\mathcal{H}$ at point $\hat{\eta}(y)$. The double bootstrap method of Hall (1992) and Efron and Tibshirani (1998) begins with a first tier of bootstrap resampling from the multivariate normal model with distribution

$$\mathbf{Y}^* \overset{i.i.d.}{\sim} N_t(\hat{\eta}(y), I_t) \tag{10}$$

A second tier of resampling is carried out for each of these vectors $\mathbf{Y}^*$, as well as for $\mathbf{Y}$, with the following distributions

$$\begin{aligned} &\mathbf{Y}^{**} \overset{i.i.d.}{\sim} N_t(\mathbf{Y}^*, I_t) \\ &\mathbf{Y}^{**} \overset{i.i.d.}{\sim} N_t(\mathbf{Y}, I_t) \end{aligned} \tag{11}$$

The second tier quantities in each case are as follows

$$\tilde{p}^* = P(y^{**} \in \mathcal{H}; y^*);\ \ \tilde{p} = P(y^{**} \in \mathcal{H}; y) \tag{12}$$

Then, according to Hall (1992) and Efron and Tibshirani (1998), the third-order accurate p-value (9) obtained by the double bootstrap method can be expressed as

$$1 - \Phi(d - c) = P(\tilde{p}^* < \tilde{p};\ \hat{\eta}(y)) + O(n^{-3/2}) \tag{13}$$

Although the double bootstrap has third-order accuracy, formula (13) suggests that it requires enormous numbers of bootstrap pseudoreplicates (many more than would be practically feasible in most cases), and in addition, computation of $\hat{\eta}(y)$ is known to be difficult. However, we propose a manipulation of the regular double bootstrap that will greatly speed its implementation and thus facilitate its application to real phylogenetic problems. Our method relies on use of formula (14) below (Efron and Tibshirani, 1996),

and on computation of $\hat{\eta}(y)$ using the PAVA method (Ayer et al., 1955; Zhao, 2007), for assessment of $\mathcal{H}$. To avoid of confusion, we call the proposed approach the speedy double bootstrap method. In the context of this section, Efron and Tibshirani (1996) showed that $1-\Phi(d-c)$ in formula (9) can be stated as follows

$$1-\Phi(d-c)=P(d^*>d;\ \hat{\eta}(y))+O(n^{-3/2}) \tag{14}$$

where d^* is the signed distance from $y^*\sim N_t(\hat{\eta}(y),I_t)$ to $\partial\mathcal{H}$. Formula (13) and (14) lead immediately to next resultant formula, when the true parameter η is on the boundary surface $\partial\mathcal{H}$.

$$P(\tilde{p}^*<\tilde{p};\ \hat{\eta}(y))=P(d^*>d;\ \hat{\eta}(y))+O(n^{-3/2}) \tag{15}$$

It is shown that double bootstrap probability also equals to third order accurate p-value $P(d^*>d;\ \hat{\eta}(y))$, the error being $O(n^{-3/2})$. The formula $P(d^*>d;\ \hat{\eta}(y))$ indicates that if we can calculate d^* and d using y^* and y respectively, then we do not need to resample from y^* and y. Now, we return to the problem of phylogenetic trees, as seen in H_1 and vector $(l_1,l_2,\ldots,l_K)$. Practically, in addition to the difficulty of computing $\hat{\eta}(y)$, calculation of d^* and d is also problematic. However, in the case of H_1, d can be analogous to $max_{j=2,\ldots,15}l_j-l_1$ (Shimodaira and Hasegawa, 2005) and as already mentioned, computation of $\hat{\eta}(y)$ can be achieved using the PAVA method. Furthermore, d^* can be analogous, d^*'s analogous will be considered in the following subsection.

2.2. The speedy double bootstrap procedure for assessing reliability of phylogenetic trees

We propose a simple double bootstrap method for assessing reliability of phylogenetic trees, that significantly mitigates the challenges of the regular double bootstrap. First we find a vector corresponding to $\hat{\eta}(y)$ in formula (10). According to Kishino et al. (1990), the vector

$$\mathbf{l}=(l_1(\hat{\theta}_1),\cdots,l_K(\hat{\theta}_K)) \tag{16}$$

asymptotically follows a multivariate normal distribution, the mean vector of which is

$$\boldsymbol{\mu}=(\mu_1,\ldots,\mu_K) \tag{17}$$

Note that, $\mathbf{l}$ of formula (16) is an unrestricted maximum likelihood estimate for $\boldsymbol{\mu}$. Assuming $\mu_1=max_{i=1,\ldots,K}\mu_i$ is the same as in H_1, under this restriction, the restricted estimator for $\boldsymbol{\mu}$ can be estimated using the PAVA method and expressed as

$$\hat{\boldsymbol{\mu}}=(\hat{\mu}_1,\ldots,\hat{\mu}_K) \tag{18}$$

Then we excise a subset W of the numerical set $\{1,\ldots,K\}$, including element 1, so that

$$\hat{\mu}_1=\frac{\sum_{j\in W}l_j(\hat{\theta}_j)}{\#W}$$
$$\hat{\mu}_j=min(\hat{\mu}_1,l_j(\hat{\theta}_j)),\ j\in\{2,\ldots,K\} \tag{19}$$

The symbol $\#W$ denotes the number of set W, and vector $\hat{\boldsymbol{\mu}}=(\hat{\mu}_1,\ldots,\hat{\mu}_K)$ corresponds to $\hat{\eta}(y)$ in formula (10). Also, the covariance matrix of vector $(l_1,l_2,\ldots,l_K)$ can be estimated by $\Sigma=(\sigma_{ij})$, with σ_{ij} given by

$$\frac{n}{n-1}\sum_{h=1}^{n}\left[logf_i(\mathbf{x_h};\hat{\theta}_i)-\frac{1}{n}\sum_{h=1}^{n}logf_i(\mathbf{x_h};\hat{\theta}_i)\right]\times\left[logf_j(\mathbf{x_h};\hat{\theta}_j)-\frac{1}{n}\sum_{h=1}^{n}logf_j(\mathbf{x_h};\hat{\theta}_j)\right] \tag{20}$$

Then we need to calculate another quantity, corresponding to d^* in formula (14). For this, we must generate $B1$ bootstrap pseudoreplicates of vector $(\hat{\mu}_1,\hat{\mu}_2,\ldots,\hat{\mu}_K)$ in formula (18). The pseudoreplicates $\left(\hat{\mu}_1^{*(b1)},\ldots,\hat{\mu}_K^{*(b1)}\right),\ b1=1,\ldots,B1$ are sampled from

$$\left(\hat{\mu}_1^{*(b1)},\ldots,\hat{\mu}_K^{*(b1)}\right)^T\overset{i.i.d.}{\sim}N_K((\hat{\mu}_1,\hat{\mu}_2,\ldots,\hat{\mu}_K)^T,\Sigma) \tag{21}$$

where T represents transpose, and Σ is used as above. Vectors $\left(\hat{\mu}_1^{*(b1)},\ldots,\hat{\mu}_K^{*(b1)}\right)$ constitute the first-order (first-tier) bootstrap pseudoreplicate. Now, d^* in formula (14) can be presented as

$$max_{j=2,\ldots,K}\hat{\mu}_j^{*(b1)}-\hat{\mu}_1^{*(b1)} \tag{22}$$

The following summary of the discussion above serves as a convenient step-by-step outline of our proposed procedure to test the null hypothesis H_1, dubbed the sDBP-test (see Fig. 1 for a specific example based on Tree-1 from Table 2).

sDBP – *test*

Step 1 Generate $B1$ bootstrap pseudoreplicates of the vector $(\hat{\mu}_1,\hat{\mu}_2,\ldots,\hat{\mu}_K)$ in (18). These pseudoreplicates

$$\left(\hat{\mu}_1^{*(b1)},\ldots,\hat{\mu}_K^{*(b1)}\right),\ b1=1,\ldots,B1 \tag{23}$$

are sampled from

$$\left(\hat{\mu}_1^{*(b1)},\ldots,\hat{\mu}_K^{*(b1)}\right)^T\overset{i.i.d.}{\sim}N_K((\hat{\mu}_1,\hat{\mu}_2,\ldots,\hat{\mu}_K)^T,\Sigma) \tag{24}$$

Step 2 For each vector

$\left(\hat{\mu}_1^{*(b1)},\ldots,\hat{\mu}_K^{*(b1)}\right)$ of Step 1, calculate

$$max_{j=2,\ldots,K}\hat{\mu}_j^{*(b1)}-\hat{\mu}_1^{*(b1)},\ b1=1,\ldots,B1 \tag{25}$$

Step 3 For the vector $(l_1,l_2,\ldots,l_K)$ calculate

$$max_{j=2,\ldots,K}l_j-l_1 \tag{26}$$

Step 4 Calculate the p-value for H_1, defined below and denoted *sDBP*

$$sDBP=\frac{\#\left(max_{j=2,\ldots,K}\hat{\mu}_j^{*(b1)}-\hat{\mu}_1^{*(b1)}>max_{j=2,\ldots,K}l_j-l_1\right)}{B1} \tag{27}$$

In exactly the same way as shown for H_1, we can apply the sDBP-test to all other hypotheses H_k, $k=2,\ldots,K$.

2.3. The double bootstrap procedure for assessing reliability of phylogenetic trees

To properly assess the utility of our sDBP-test, it is necessary to compare our results with those generated using the standard double bootstrap procedure. To this end, we propose the following protocol for application of the regula double bootstrap to test the null hypothesis H_1, dubbed the DBP-test (see Fig. 2 for a specific example based on Tree-1 from Table 2).

DBP – *test*

Step 1 Generate $B1$ bootstrap pseudoreplicates of the vector $(\hat{\mu}_1,\hat{\mu}_2,\cdots,\hat{\mu}_K)$. These pseudoreplicates $\left(\hat{\mu}_1^{*(b1)},\ldots,\hat{\mu}_K^{*(b1)}\right),b1=1,\ldots,B1$ are sampled from

$$\left(\hat{\mu}_1^{*(b1)},\ldots,\hat{\mu}_K^{*(b1)}\right)^T\overset{i.i.d.}{\sim}N_K((\hat{\mu}_1,\hat{\mu}_2,\ldots,\hat{\mu}_K)^T,\Sigma) \tag{28}$$

This step is identical to Step 1 of the sDBP-test.

Step 2 Generate $B2$ bootstrap pseudoreplicates of each vector

$$\left(\hat{\mu}_1^{*(b1)},\hat{\mu}_2^{*(b1)},\ldots,\hat{\mu}_K^{*(b1)}\right) \tag{29}$$

from Step 1. These pseudoreplicates $\left(\hat{\mu}_1^{**(b2)},\ldots,\hat{\mu}_K^{**(b2)}\right),b2=1,\ldots,B2$ are sampled from

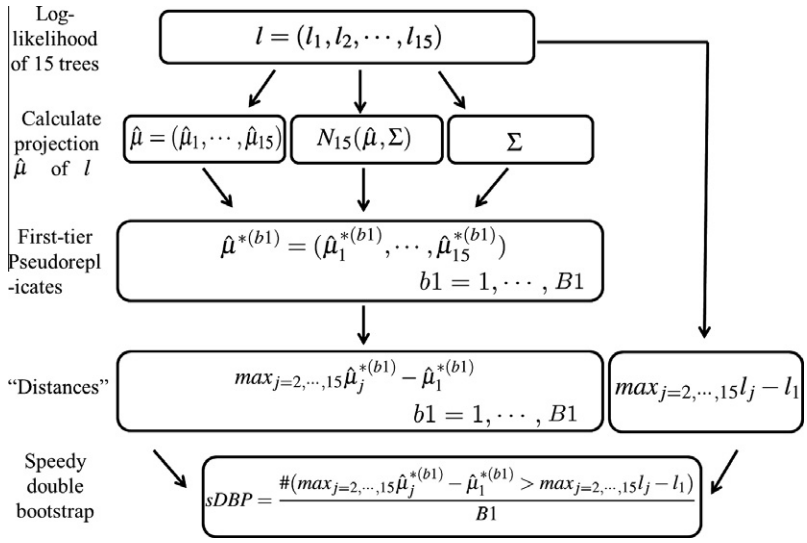

Fig. 1. The steps of the sDBP-test for Tree-1 Flow diagram to illustrate the steps in our speedy double bootstrap method (the sDBP-test of H_1), using Tree-1 (Table 2) as an example. The gDistancesh are anology distances d^* and d in formula (14), Σ is calculate from Eq. (20).

$$\left(\hat{\mu}_1^{**(b2)}, \ldots, \hat{\mu}_K^{**(b2)}\right)^T \overset{i.i.d.}{\sim} N_K\left(\left(\hat{\mu}_1^{*(b1)}, \ldots, \hat{\mu}_K^{*(b1)}\right)^T, \Sigma\right), b1 = 1, \ldots, B1 \tag{30}$$

Vectors $\left(\hat{\mu}_1^{**(b2)}, \ldots, \hat{\mu}_K^{**(b2)}\right)$ constitute the second-order (second-tier) bootstrap pseudoreplicate. Calculate the bootstrap probability $BP^{*(b1)}$ for each $\left(\hat{\mu}_1^{*(b1)}, \hat{\mu}_2^{*(b1)}, \ldots, \hat{\mu}_K^{*(b1)}\right)$, as follows

$$BP^{*(b1)} = \frac{\#\left(argmax\left(\hat{\mu}_1^{**(b2)}, \ldots, \hat{\mu}_K^{**(b2)}\right) = 1\right)}{B2}, b1 = 1, \ldots, B1 \tag{31}$$

Step 3 Generate $B2$ bootstrap pseudoreplicates of the vector $(l_1, l_2, \ldots, l_K)$. These pseudoreplicates $\left(l_1^{**(b2)}, \ldots, l_K^{**(b2)}\right)$, $b2 = 1, \ldots, B2$ are sampled from

$$\left(l_1^{**(b2)}, \cdots, l_K^{**(b2)}\right)^T \overset{i.i.d.}{\sim} N_K((l_1, l_2, \ldots, l_K)^T, \Sigma) \tag{32}$$

Now calculate the bootstrap probability BP, as follows

$$BP = \frac{\#\left(argmax\left(l_1^{**(b2)}, \ldots, l_K^{**(b2)}\right) = 1\right)}{B2} \tag{33}$$

Step 3 is the step for calculating traditional BP.

Step 4 Calculate the p-value for H_1, defined below and denoted DBP

$$DBP = \frac{\#\left(BP^{**(b1)} < BP\right)}{B1} \tag{34}$$

Similarly, we can apply the DBP-test to all other hypotheses H_k, $k = 2, \ldots, K$.

2.4. Analysis of mammalian mitochondrial amino acid sequences (protein-coding genes) and the 12S and 16S rRNA genes

To apply our methods to a real molecular dataset, we analyzed the amino acid sequences of the mammalian mitochondrial protein-coding genes and the DNA sequences of the 12S and 16S rRNA genes using the sDBP-test and the DBP-test. We included 20 mammalian species in these analyses, belonging to eight major clades: Primates, Lagomorpha, Rodentia, Fereuungulata, Chiroptera, Soricomorpha, Marsupialia and Monotremata (see Table 1). We used the representatives from Marsupialia and Monotremata as outgroups. In cases where a major clade was represented by more than two species, the following relationships were applied, based on the results of Cao et al. (2000): in the Fereuungulata, ((domestic cat, harbor seal), (horse, (Indian rhinoceros, white rhinoceros)), (cow, blue whale)); in the Primates, (((human, chimpanzee), western gorilla), Sumatran orangutan); and in the outgroup, ((American opossum, wallaroo), platypus). We also assumed the relationship (Fereuungulata, (Chiroptera, Soricomorpha)), based once again on strong support for these groupings in Cao et al. (2000). We can thus divide these taxa into five monophyletic groups: Primates (group I), Lagomorpha (group II), Rodentia (group III), (Fereuungulata, (Chiroptera, Soricomorpha)) (group IV), (Marsupialia, Monotremata) (group V). Finally all of the 15 unrooted trees (see Table 1 footnote *c*) compatible with these five groups were considered in our comparisons. These 15 unrooted trees of 20 species are shown in Table 2, respectively.

The 12 proteins coded for in the mammalian mitochondrial genome are ND1, ND2, COX1, COX2, ATP8, ATP6, COX3, ND3, ND4L, ND4, ND5, and CYTB. Amino acid alignments for these proteins were constructed using ClustalW version 1.83 (Thompson et al., 1994), and all positions with gaps were excluded from analyses. This resulted in a final total of 3593 amino acids in the alignment. DNA sequences for the small (12S) and large (16S) mitochondrial rRNA genes were also aligned using ClustalW. As with the amino acid sequences, alignment positions with gaps were excluded from the analyses leaving a total of 870 and 1416 sites for the 12S and 16S genes, respectively. Analysis of the amino acid alignment was conducted using the CodeML program of the PAML package (Yang, 1997), applying the Empirical + F model and the mtREV24.dat rate matrix (Adachi and Hasegawa, 1996), and modeling rate heterogeneity among sites with the discrete gamma distribution (Yang, 1996). The CodeML program provides site-wise log-likelihoods for each of the 3593 amino acids in the alignment. The 12S and 16S rRNA gene sequences were analyzed using the BaseML program of the PAML package (Yang, 1997), applying the REV model and modeling rate heterogeneity among

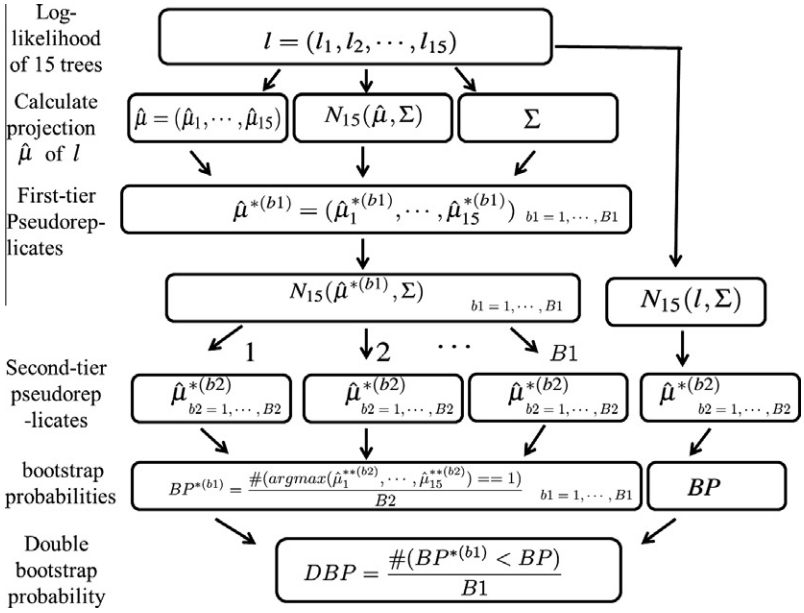

Fig. 2. The steps of the DBP-test for Tree-1 Flow diagram to illustrate the steps in our implementation of the regular double bootstrap method (the DBP-test of H_1), using Tree-1 (Table 2) as an example. Σ is calculate from Eq. (20), BP is traditional bootstrap proportion.

sites with the discrete gamma distribution. The BaseML program provides site-wise log-likelihoods for each of the bases in the 12S and 16S alignments. These site-wise log-likelihood scores for the amino acids and for the two rRNA gene sequences were summed to determine the best ML tree and to calculate the sDBP, DBP, AU, and BP for the 15 candidate trees. We compared our sDBP and DBP results with each other, and with the BP and AU results (see Table 2). First, mean square errors were used to assess whether sDBP is a reasonable approximation of DBP, and to compare its accuracy with the other two methods (BP and AU). Second, to determine whether there was any significant difference between the sDBP and DBP results, we used the availability of paired data for each phylogenetic tree, and applied the paired *t*-test (which is actually a *t*-test on a sample of differences). These analyses were performed using the software R 2.9.0 (R Core Team, 2012) on a personal computer with the following specifications: 2.50 GHz CPU (Core (TM) i5-2520 M CPU), 8.00 GB RAM.

2.5. Comparisons of computational speed

For the sDBP-test, the DBP-test and the BP-test, we measured the time taken to calculate a *p*-value for Tree-1 (see Table 2), based on the site-wise log-likelihood data. We used the RELL approximation method (Kishino et al., 1990) with the BP-test, and conducted

Table 1
The 20 mammalian species used in this study, including GenBank accession numbers for sequence data, and the major clade membership and group membership (Groups I–V) of each species.

GenBank accession number[a]	Binomial scientific name	Common name	Major clade[b]	Group[c]
NC_001807	*Homo sapiens*	Human	Primates	I
NC_001643	*Pan troglodytes*	Chimpanzee	Primates	I
NC_001645	*Gorilla gorilla*	Western gorilla	Primates	I
NC_002083	*Pongo abelii*	Sumatran orangutan	Primates	I
NC_001913	*Oryctolagus cuniculus*	Rabbit	Lagomorpha	II
NC_002658	*Thryonomys swinderianus*	Greater cane rat	Rodentia	III
NC_005089	*Mus musculus*	House mouse	Rodentia	III
NC_001700	*Felis catus*	Domestic cat	Fereuungulata	IV
NC_001325	*Phoca vitulina*	Harbor seal	Fereuungulata	IV
NC_001640	*Equus caballus*	Horse	Fereuungulata	IV
X97336	*Rhinoceros unicornis*	Indian rhinoceros	Fereuungulata	IV
NC_001808	*Ceratotherium simum*	White rhinoceros	Fereuungulata	IV
NC_001601	*Balaenoptera musculus*	Blue whale	Fereuungulata	IV
NC_006853	*Bos taurus*	Cow	Fereuungulata	IV
AF061340	*Artibeus jamaicensis*	Jamaican fruit-eating bat	Chiroptera	IV
AB042770	*Pteropus dasymallus*	Ryukyu flying fox	Chiroptera	IV
Y19192	*Talpa europaea*	European mole	Soricomorpha	IV
NC_001610[d]	*Didelphis virginiana*	North American opossum	Marsupialia	V
Y10524[d]	*Macropus robustus*	Wallaroo	Marsupialia	V
NC_000891[d]	*Ornithorhynchus anatinus*	Platypus	Monotremata	V

[a] NCBI (GenBank) accession number.
[b] The major clade that the species belongs to.
[c] The group (I–V) that the species belongs to. The 15 unrooted trees compatible with these five groups are t1: (((I,IV), II), III, V), t2: ((I,IV), (II,III), V), t3: (((I,II), IV), III, V), t4: ((I, (II,III)), IV, V), t5: ((I, (IV,II)), III, V), t6: (I, (IV, (II,III)), V), t7: (((I,IV), III), II, V), t8: (((I,II), III), IV, V), t9: (((I,III), II), IV, V), t10: (I, ((IV,II), III), V), t11: ((I,III), (IV,II), V), t12: ((I,II), (IV,III), V), t13: (I, ((IV,III), II), V), t14: (((I,III), IV), II, V), t15: ((I, (IV,III)), II, V).
[d] Outgroup species.

Table 2
sDBP, DBP, AU and BP results based on final amino acid and DNA sequence alignments, for each of the 15 candidate trees of 20 mammalian species. The p-values that are NOT significant at $\alpha = 0.05$ are emphasized in bold type.

Tree form[a]	$\triangle l_i$	BP_i[b]	DBP_i[c]	$sDBP_i$[d]	AU_i[e]
1: ((((((1,2), 3), 4), (((8,9), (10, (11,12)), (13,14)), ((15,16), 17))), 5), (6,7), ((18,19), 20))	−5.5	**0.474**	**0.688**	**0.748**	**0.881**
2: (((((1,2), 3), 4), (((8,9), (10, (11,12)), (13,14)), ((15,16), 17))), (5, (6,7)), ((18,19), 20))	5.5	**0.189**	**0.458**	**0.436**	**0.524**
3: ((((((1,2), 3), 4), 5), (((8,9), (10, (11,12)), (13,14)), ((15,16), 17))), (6,7), ((18,19), 20))	8.0	**0.115**	**0.332**	**0.384**	**0.375**
4: (((((1,2), 3), 4), (5, (6,7))), (((8,9), (10, (11,12)), (13,14)), ((15,16), 17)), ((18,19), 20))	11.6	**0.094**	**0.319**	**0.327**	**0.429**
5: (((((1,2), 3), 4), ((((8,9), (10, (11,12)), (13,14)), ((15,16), 17)), 5)), (6,7), ((18,19), 20))	11.9	0.035	**0.197**	**0.307**	**0.193**
6: ((((1,2), 3), 4), ((((8,9), (10, (11,12)), (13,14)), ((15,16), 17)), (5, (6,7))), ((18,19), 20))	14.5	0.042	**0.224**	**0.263**	**0.271**
7: ((((((1,2), 3), 4), (((8,9), (10, (11,12)), (13,14)), ((15,16), 17))), (6,7)),5, ((18,19), 20))	17.9	0.001	0.022	**0.164**	0.014
8: ((((((1,2), 3), 4), 5), (6,7)), (((8,9), (10, (11,12)), (13,14)), ((15,16), 17)), ((18,19), 20))	18.0	0.028	**0.201**	**0.223**	**0.332**
9: ((((((1,2), 3), 4), (6,7)), 5), (((8,9), (10, (11,12)), (13,14)), ((15,16), 17)), ((18,19), 20))	20.5	0.021	**0.142**	**0.181**	**0.205**
10: ((((1,2), 3), 4), (((((8,9), (10, (11,12)), (13,14)), ((15,16), 17)), 5), (6,7)), ((18,19), 20))	30.0	0.001	0.019	0.047	0.022
11: (((((1,2), 3), 4), (6,7)), ((((8,9), (10, (11,12)), (13,14)), ((15,16), 17)), 5), ((18,19), 20))	32.1	0.000	0.022	0.033	0.014
12: (((((1,2), 3), 4), 5), ((((8,9), (10, (11,12)), (13,14)), ((15,16), 17)), (6,7)), ((18,19), 20))	36.1	0.000	0.032	0.011	0.000
13: ((((1,2), 3), 4), (((((8,9), (10, (11,12)), (13,14)), ((15,16), 17)), (6,7)), 5), ((18,19), 20))	36.7	0.000	0.032	0.011	0.000
14: ((((((1,2),3),4), (6,7)), (((8,9), (10, (11,12)), (13,14)), ((15,16), 17))), 5, ((18,19), 20))	37.0	0.000	0.035	0.012	0.000
15: (((((1,2), 3), 4), ((((8,9), (10, (11,12)), (13,14)), ((15,16), 17)), (6,7))), 5, ((18,19), 20))	44.0	0.000	0.045	0.000	0.000

[a] Trees are numbered by increasing order of $\triangle l_i = max_{j \neq i} l_j - l_i$, the difference between the log-likelihood value of a given tree and the largest value among all other trees. Species labels: 1 = human, 2 = chimpanzee, 3 = western gorilla, 4 = Sumatran orangutan, 5 = rabbit, 6 = greater cane rat, 7 = house mouse, 8 = domestic cat, 9 = harbor seal, 10 = horse, 11 = Indian rhinoceros, 12 = white rhinoceros,13 = blue whale, 14 = cow, 15 = Jamaican fruit-eating bat, 16 = Ryukyu flying fox, 17 = European mole, 18 = North American opossum, 19 = wallaroo, 20 = platypus.
[b] Bootstrap probability, calculated from B1 = 10000 pseudoreplicates.
[c] Double bootstrap probability, calculated from 1 million pseudoreplicates (B1 = 1000, B2 = 1000).
[d] Speedy double bootstrap probability, calculated from B1 = 10,000 pseudoreplicates.
[e] Multiscale bootstrap probability, calculated from B1 = 10,000 pseudoreplicates.

Table 3
Comparison of the BP, DBP and sDBP methods, regarding their speed to compute a p-value for tree-1.

	DBP	sDBP	BP
Time (secs)[a]	900.02	2.42	0.448
Time (secs)[b]	23164	5.95	2.20
Speed increase (DBP/sDBP)	371 times[a]	3893 times[b]	

[a] ($B1 = 10^3, B2 = 10^3$ pseudoreplicates).
[b] ($B1 = 5 \times 10^3, B2 = 5 \times 10^3$ pseudoreplicates).

two separate sets of analyses. In the first set, we applied the sDBP-test with $B1 = 10^3$ pseudoreplicates, the DBP-test with $B1 = 10^3$ and $B2 = 10^3$ pseudoreplicates, and the BP-test with 10^3 pseudoreplicates. In the second set, we applied the sDBP-test with $B1 = 5 \times 10^3$ pseudoreplicates, the DBP-test with $B1 = 5 \times 10^3$ and $B2 = 5 \times 10^3$ pseudoreplicates, and the BP-test with 5×10^3 pseudoreplicates.

3. Results

3.1. Analysis of mammalian mitochondrial data

Table 2 presents results of our sDBP and DBP calculations for the 15 phylogenetic trees analyzed in this study, alongside values for BP and AU. The confidence sets of trees obtained, respectively, by the sDBP-test and the DBP-test at $\alpha = 0.05$ were {1,2,3,4,5,6,8,9} and {1,2,3,4,5,6,7,8,9} (Table 2). The sDBP tree set was thus slightly larger than the set selected by DBP. Tree-4 is most strongly supported as the T_{ML} by previous studies that have included a more comprehensive set of clades and species than we have used here (e.g., Cao et al., 2000; Madsen et al., 2001; Murphy et al., 2001). Our results for this tree indicate that sDBP = 0.327 > 0.05 and DBP = 0.319 > 0.05, and our conclusions are thus not in contradiction with the latest data.

Based on the values in Table 2, mean square errors between DBP and the other three methods (sDBP, BP and AU) were, respectively, 0.003, 0.022 and 0.006. Thus, the sDBP apparently provides a good approximation of the regular DBP, with the sDBP–DBP comparison having the lowest error (0.003). In addition, comparison of sDBP and DBP results using the paired t-test returned a p-value of 0.079, providing no evidence of a significant difference between these methods.

3.2. Comparisons of computational speed

Results of the two sets of analyses conducted to compare computational speed between the sDBP, DBP and BP methods are shown in Table 3. In both sets the BP-test was the fastest, followed by the sDBP-test, then the DBP-test. In the first set of calculations (lower numbers of pseudoreplicates) the sDBP-test was 371 times faster than DBP-test, and this advantage improved substantially in the second set (higher pseudoreplication), in which the sDBP-test was 3893 times faster than DBP-test.

4. Discussion

The maximum likelihood inference $(l_1, l_2, \ldots, l_K)$ for phylogenetic trees is also expressed approximately as (8), but the covariance matrix is not identity. This reduces again to the identity matrix case by applying a linear transformation to $(l_1, l_2, \ldots, l_K)$ (Shimodaira and Hasegawa, 2005). Thus, we can use the sDBP-test and the DBP-test for the general phylogenetic tree selection problem.

Let us define G as the region occupied by H_1 in formula (6), and ∂H_1 is then the boundary of the region G. From $\triangle l_i = max_{j \neq i} - l_i$, $i = 1, \ldots, 15$ of Table 2, for example, the following can be determined: If $\triangle\ l_1 = (max_{j=2,\ldots,K} l_j - l_1) > 0$, then we can establish that $\mathbf{l} \notin G$. If $\triangle l_1 = (max_{j=2,\ldots,K} l_j - l_1) < 0$, then we can establish that $\mathbf{l} \in G$. If $\triangle l_1 = (max_{j=2,\ldots,K} l_j - l_1) = 0$, then we can establish that $\mathbf{l} \in \partial G$. Therefore, there are analogs in nature for d^*, d, and formulae (25) and (26). However, in case of hypothesis H_1, the shape of the null is a polyhedral convex cone and ∂H_1 is nonsmooth at the vertex as well as on the faces of dimensions less than $K - 1$. As already mentioned, the double bootstrap method (Hall, 1992; Efron and Tibshirani, 1998) assumes that the boundary of the region is a smooth surface. Regions with nonsmooth boundaries, in particular, may lead to serious difficulties as discussed by Perlman and Wu (1999), Perlman and Wu (2003), and further study is needed in this regard.

Based on our comparison of the speedy double bootstrap method with other approaches for estimating the reliability of phylogenetic trees (regular double bootstrap, multiscale bootstrap (AU test) and traditional bootstrap probability) we recommend the sDBP-test for general tree selection problems. This method is computationally less burdensome than the AU test or the regular DBP, and has several other advantages over competing methods. The estimator $\hat{\mu}$ estimates the particular parameter configuration of a given molecular dataset when we assume that the true parameter lies on the boundary ∂H_1. This allows calculation of higher-order accurate p-values by the sDBP method, an advantageous feature shared with the double bootstrap method but lacking in the multiscale bootstrap (AU test) and the traditional bootstrap probability (BP-test). Furthermore, because the sDBP is not dependent on the BP-test, it is not susceptible to the potential biases associated with the BP. This is a unique advantage of the sDBP-test, and an additional argument for its superiority. However, the sDBP is impractical when $\hat{\mu}$ or the signed distance are difficult to estimate. In these cases, other methods of tree selection should be used instead.

We plan to make our R modules available via CRAN, and to develop a computer program to perform the sDBP and DBP tests described in this paper. In addition, implementation of the speedy double bootstrap does not involve difficult calculations such as the optimization of non-linear functions necessary for the AU test, so this method could be easily incorporated into any of the general phylogenetic analysis packages that calculate site-wise log-likelihoods from the dataset.

5. Conclusions

We have presented the speedy double bootstrap procedure (sDBP-test) for assessing confidence levels of phylogenetic trees, and for comparison we have also developed a double bootstrap procedure (DBP-test) for the same purposes. Our sDBP-test provides improvements in accuracy over the traditional bootstrap probability (BP-test), and substantial improvements in speed over the regular double bootstrap (DBP-test), for the first time enabling the double bootstrap technique to be practically applied in the context of molecular phylogenetics.

Our calculations show that the application of the sDBP-test is not confined to general tree selection problems; rather, it is appropriate for general model selection problems when the maximum likelihood criterion is used.

Acknowledgments

We thank Hidetoshi Shimodaira for helpful discussions. We also thank Osamu Watanabe. We also thank the anonymous reviewers for helpful comments. Our work was supported by the Global COE program gComputationism as a Foundation for the Sciencesh (CompView) of the Tokyo Institute of Technology.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2013.02.011.

References

Adachi, J., Hasegawa, M., 1996. Model of amino acid substitution in proteins encoded by mitochondrial dna. Journal of Molecular Evolution 42, 459–468.
Ayer, M., Brunk, H., Ewing, G., Reid, W., Silverman, E., 1955. An empirical distribution function for sampling with incomplete information. Annals of Mathematical Statistics, 641–647.
Cao, Y., Fujiwara, M., Nikaido, M., Okada, N., Hasegawa, M., 2000. Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data. Gene 259, 149–158.
Efron, B., 1985. Bootstrap confidence intervals for a class of parametric problems. Biometrika 72, 45–58.
Efron, B., Halloran, E., Holmes, S., 1996. Bootstrap confidence levels for phylogenetic trees. Proceedings of the National Academy of Sciences USA 93, 13429–13434.
Efron, B., Tibshirani, R., 1996. The Problem of Regions. Stanford Technical Report 192.
Efron, B., Tibshirani, R., 1998. The problem of regions. Annals of Statistics 26, 1687–1718.
Felsenstein, J., 1981. Evolutionary trees from dna sequences: a maximum likelihood approach. Journal of Molecular Evolution 17, 368–376.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution, 783–791.
Felsenstein, J., 2004. Inferring Phylogenies. Sinauer Associates, Sunderland, Massachusetts.
Hall, P., 1992. The bootstrap and Edgeworth expansion. Springer Verlag, New York.
Hillis, D., Bull, J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology 42, 182–192.
Hsu, J., 1981. Simultaneous confidence intervals for all distances from the gbesth. Annals of Statistics, 1026–1034.
Kishino, H., Miyata, T., Hasegawa, M., 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. Journal of Molecular Evolution 31, 151–160.
Madsen, O., Scally, M., Douady, C., Kao, D., DeBry, R., Adkins, R., Amrine, H., Stanhope, M., de Jong, W., Springer, M., 2001. Parallel adaptive radiations in two major clades of placental mammals. Nature 409, 610–614.
Murphy, W., Eizirik, E., Johnson, W., Zhang, Y., Ryder, O., O'Brien, S., 2001. Molecular phylogenetics and the origins of placental mammals. Nature 409, 614–618.
Pazos, F., Valencia, A., 2001. Similarity of phylogenetic trees as indicator of protein–protein interaction. Protein Engineering 14, 609–614.
Perlman, M., Wu, L., 1999. The emperorfs new tests. Statistical Science 14, 355–369.
Perlman, M., Wu, L., 2003. On the validity of the likelihood ratio and maximum likelihood methods. Journal of Statistical Planning and Inference 117, 59–81.
R Core Team, 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Sanderson, M., Wojciechowski, M., 2000. Improved bootstrap confidence limits in large-scale phylogenies, with an example from neo-astragalus (leguminosae). Systematic Biology 49, 671–685.
Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Systematic Biology 51, 492–508.
Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Molecular Biology and Evolution 16, 1114–1116.
Shimodaira, H., Hasegawa, M., 2005. Assessing the uncertainty in phylogenetic inference. In: Nielsen, R. (Ed.), Statistical Methods In Molecular Evolution: Statistics for Biology and Health. Springer, pp. 463–493 (chapter 17).
Thompson, J., Higgins, D., Gibson, T., 1994. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673–4680.
Webb, C., Ackerly, D., McPeek, M., Donoghue, M., 2002. Phylogenies and community ecology. Annual Review of Ecology and Systematics, 475–505.
Wiley, E., 1981. Phylogenetics: The Theory and Practice of Phylogenetic Systematics. Wiley-Interscience, New York.
Yang, Z., 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends in Ecology & Evolution 11, 367–372.
Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences 13, 555–556.
Zhao, H., 2007. Comparing several treatments with a control. Journal of Statistical Planning and Inference 137, 2996–3006.