

MOLECULAR BIOLOGY AND EVOLUTION

academic.oup.com/mbc



Downloaded from <https://academic.oup.com/mbc/issue/35/2> by guest on 28 January 2019

Society for Molecular Biology and Evolution

Print ISSN 0737-4038
Online ISSN 1537-1719

- Rates for mechanisms of reproductive isolation
- Rapid human colonization of South America
- Modular path from scales to feathers
- Tropical adaptation in domestic dogs
- Association between cold and cancer

MOLECULAR BIOLOGY AND EVOLUTION

(ISSN 0737-4038)

Founding Editors:
Editor in Chief:

Walter Fitch and Masatoshi Nei
Sudhir Kumar

Institute for Genomics and
Evolutionary Medicine,
Temple University,
Philadelphia, PA, USA
EIC.MBE@gmail.com

Editorial Assistant:
Editor's Advisors:

Marjorie Gillis
Barry Hall
Marcy Uyenoyama
Joseph Caspermeyer
Heather Rowe

EAssist.MBE@gmail.com
barryghall@gmail.com
marcy@duke.edu
joseph.caspermeyer@asu.edu
roweheat@gmail.com

News Writer:
Assistant Managing Editor:

Editors (Senior and Associate)

A list of senior and associate editors is available at
academic.oup.com/mbe/pages/Editorial_Board

Information for Authors

Guidelines for author submissions are available at
academic.oup.com/mbe/pages/General_Author_Guidelines

SOCIETY FOR MOLECULAR BIOLOGY AND EVOLUTION

www.smbe.org

COUNCIL

Bill Martin, *President (Editor-in-chief, Genome Biology and Evolution)*

Institut for Molecular Evolution, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany. Email: bill@hhu.de

Laura Landweber, *Past-President*

Departments of Biochemistry & Molecular Biophysics and Biological Sciences, Columbia University, New York, NY 10032.
Email: Laura.Landweber@columbia.edu

Aoife McLysaght, *President-Elect*

Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2 Ireland. Email: aoife.mclysaght@tcd.ie

David Pollock, *Secretary*

Department of Biochemistry and Molecular Genetics, University of Colorado, Aurora, CO 80045. Email: secretary.smbe@gmail.com

Jeffrey Thorne, *Treasurer*

Biological Sciences and Statistics, North Carolina State University, Raleigh, NC 27695. Email: treasurer.smbe@gmail.com

Belinda Chang, *Councillor (2018-2020)*

Ecology & Evolutionary/Cellular & Systems Biology, University of Toronto, Ontario M5S 3G5 Canada. Email: belinda.chang@utoronto.ca

Nicholas Galtier, *Councillor (2018-2020)*

Institut des Sciences de l'Evolution, Université Montpellier Montpellier, France 34095. Email: Nicolas.Galtier@univ-montp2.fr

Maud Tenaillon, *Councillor (2016-2018)*

Quantitative Genetics and Evolution - Le Moulon INRA, University of Paris-Sud, CNRS, AgroParisTech, Gif-sur-Yvette, France.
Email: tenaillon@moulon.inra.fr

Adam Eyre-Walker, *Councillor (2016-2018)*

School of Life Sciences, University of Sussex, Brighton, BN1 9QG, United Kingdom. Email: a.c.eyre-walker@sussex.ac.uk

Joanna Masel, *Councillor (2017-2019)*

Department of Ecology & Evolutionary Biology, University of Arizona, Tucson, AZ 85721. Email: masel@u.arizona.edu

Jay Storz, *Councillor (2017-2019)*

School of Biological Sciences, University of Nebraska, Lincoln, NE 68588. Email: jstorz2@unl.edu

Sudhir Kumar, *ex officio (Editor-in-chief, Molecular Biology and Evolution)*

Institute for Genomics & Evolutionary Medicine (iGEM), Temple University, Philadelphia, PA 19122. Email: EiC.MBE@gmail.com

Molecular Biology and Evolution (ISSN 0737-4038) is a peer-reviewed journal published twelve times a year by Oxford University Press for the Society for Molecular Biology and Evolution.

Molecular Biology and Evolution (MBE) publishes research at the interface between molecular and evolutionary biology. The journal publishes molecular evolutionary patterns, processes, and predictions at all levels of taxonomic, functional, genomic, and phenotypic organization. It also publishes new and improved methods, resources, and theories that are critical for advancing the molecular evolutionary research. MBE does not publish manuscripts containing largely descriptive work, confirmatory research, and discoveries with a limited scope or impact.

Membership in the Society for Molecular Biology and Evolution

Individuals interested in becoming a member of the Society for Molecular Biology and Evolution can obtain application forms from jnls.cust.serv@oup.com.

Subscriptions

A subscription to *Molecular Biology and Evolution* comprises 12 issues. All prices include postage; for subscribers outside the Americas, issues are sent air freight. *Molecular Biology and Evolution* Advance Access contains papers that have recently been accepted but have not yet been included within an issue. Advance Access is updated daily.

Annual Subscription Rate (Volume 35, 12 issues, 2018)

Institutional

Print edition and site-wide online access: US\$1292.00/£736.00/€1149.00

Print edition only: US\$1191.00/£679.00/€1059.00

Site-wide online access only: US\$782.00/£447.00/€697.00

Please note: UK£ rate applies to UK and Rest of World, except US and Canada (\$) and Europe (€). Payment for orders to be delivered within Europe should be made in euros.

There are other subscription rates available; for a complete listing, please visit academic.oup.com/mbe/subscribe

Full prepayment in the correct currency is required for all orders. Payment should be in US dollars for orders being delivered to the USA or Canada; Euros for orders being delivered within Europe (excluding the UK); GBP sterling for orders being delivered elsewhere (i.e. not being delivered to USA, Canada, or Europe). All orders should be accompanied by full payment and sent to your nearest Oxford Journals office. Subscriptions are accepted for complete volumes only. Orders are regarded as firm, and payments are not refundable. Our prices include Standard Air as postage outside of the UK. Claims must be notified within four months of despatch/order date (whichever is later). Subscriptions in the EEC may be subject to European VAT. If registered, please supply details to avoid unnecessary charges. For subscriptions that include online versions, a proportion of the subscription price may be subject to UK VAT. Subscribers in Canada, please add GST to the prices quoted. Personal rate subscriptions are only available if payment is made by personal cheque or credit card, delivery is to a private address, and is for personal use only.

The current year and two previous years' issues are available from Oxford Journals. Previous volumes can be obtained from the Periodicals Service Company at <http://www.periodicals.com/oxford.html> or Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. E-mail: psc@periodicals.com. Tel: (518) 537-4700. Fax: (518) 537-5899.

Contact information

Journals Customer Service Department, Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, UK. E-mail: jnls.cust.serv@oup.com. Tel: +44 (0)1865 353907. Fax: +44 (0)1865 353485.

In the Americas, please contact: Journals Customer Service Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513, USA. E-mail: jnlorders@oup.com. Tel: (800) 852-7323 (toll-free in USA/Canada) or (919) 677-0977. Fax: (919) 677-1714. **In Japan, please contact:** Journals Customer Service Department, Oxford University Press, 4-5-10-8F Shiba, Minato-ku, Tokyo, 108-8386, Japan. E-mail: custserv.jp@oup.com. Tel: (81) 3 5444 5858. Fax: (81) 3 3454 2929.

Methods of payment

(i) Check (payable to Oxford University Press, to Oxford University Press, Cashiers Office, Great Clarendon Street, Oxford OX2 6DP, UK) in GB£ Sterling (drawn on a UK bank), US\$ Dollars (drawn on a US bank), or EU€ Euros. (ii) Bank transfer to Barclays Bank Plc, Oxford Group Office, Oxford (bank sort code 20-65-18) (UK), overseas only Swift code BARC GB 22 (GB£ Sterling to account no. 70299332, IBAN GB89BARC20651870299332; US\$ Dollars to account no. 66014600, IBAN GB27BARC20651866014600; EU€ Euros to account no. 78923655, IBAN GB16BARC20651878923655). (iii) Credit card (Mastercard, Visa, Switch or American Express).

Postal information

Molecular Biology and Evolution (ISSN 0737-4038) is published monthly by Oxford University Press, 2001 Evans Road, Cary, NC 27513-2009. Periodicals Postage Paid at Cary, NC, and additional mailing offices. POSTMASTER: Send address changes to *Molecular Biology and Evolution*, Mercury Media Processing, 187 E. Albert St., Rahway, NJ 07065, USA.

Environmental and ethical policies

Oxford Journals, a division of Oxford University Press, is committed to working with the global community to bring the highest quality research to the widest possible audience. Oxford Journals will protect the environment by implementing environmentally friendly policies and practices wherever possible. Please see academic.oup.com/journals/pages/authors_ethics for further information on environmental and ethical policies.

Digital Object Identifiers

For information on dois and to resolve them, please visit www.doi.org.

Permissions

For information on how to request permissions to reproduce articles or information from this journal, please visit academic.oup.com/journals/pages/authors

Advertising

Advertising, inserts, and artwork enquiries should be addressed to Advertising and Special Sales, Oxford Journals, Oxford University Press, Great Clarendon Street, Oxford, OX2 6DP, UK. Tel: +44 (0)1865 354767; Fax: +44 (0)1865 353774; E-mail: jnlsadvertising@oup.com.

Granting of License

It is a condition of publication in the journal that authors grant an exclusive license to the Society for Molecular Biology and Evolution. This ensures that requests from third parties to reproduce articles are handled efficiently and consistently and will also allow the article to be as widely disseminated as possible. As part of the license agreement, authors may use their own material in other publications provided that the journal is acknowledged as the original place of publication and Oxford University Press as the publisher.

Statements and opinions expressed in publications of the Society for Molecular Biology and Evolution or in presentations given during its regular meetings are those of the author(s) and do not necessarily reflect the official position of the Society for Molecular Biology and Evolution, the Editors, or the organizations with which the authors are affiliated. The Editor(s), publisher, and Society disclaim any responsibility or liability for such material and do not guarantee, warrant, or endorse any product or service mentioned, including those by advertisers, nor do they guarantee any claim made by the manufacturer of such product or service. Official positions of the Society are established only by its Council.

Copyright © 2018 Society for Molecular Biology and Evolution

All rights reserved; no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without prior written permission of the publisher or a license permitting restricted copying issued in the UK by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1P 9HE.

Typeset by Cenveo Publisher Services, Bengaluru, India; Printed by Bell & Bain Ltd., Glasgow, UK.

Indexing and Abstracting

Molecular Biology and Evolution is abstracted or indexed in *Biological Abstracts*, *BIOSIS*, *Chemical Abstracts*, *Current Contents in Life Sciences*, *Excerpta Medica*, *Index Medicus*, *ISI/BIOMED*, *Science Citation Index*, and *Zoological Record*.

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide.

Contents

PERSPECTIVE

- 263 Multiple Innovations in Genetic and Epigenetic Mechanisms Cooperate to Underpin Human Brain Evolution
Mainá Bitar and Guy Barry

FAST TRACK

- 269 Interdependence, Reflexivity, Fidelity, Impedance Matching, and the Evolution of Genetic Coding
Charles W. Carter Jr and Peter R. Wills

DISCOVERIES

- 287 Whole-Genome Sequencing of African Dogs Provides Insights into Adaptations against Tropical Parasites
Yan-Hu Liu, Lu Wang, Tao Xu, Xiaomin Guo, Yang Li, Ting-Ting Yin, He-Chuan Yang, Yang Hu, Adeniyi C. Adeola, Oscar J. Sanke, Newton O. Otecko, Meng Wang, Yaping Ma, Olaogun S. Charles, Mikkel-Holger S. Sinding, Shyam Gopalakrishnan, José Alfredo Samaniego, Anders J. Hansen, Carlos Fernandes, Philippe Gaubert, Jane Budd, Philip M. Dawuda, Eli Knispel Rueness, Lubin Jiang, Weiwei Zhai, M. Thomas P. Gilbert, Min-Sheng Peng, Xiaopeng Qi, Guo-Dong Wang, and Ya-Ping Zhang
- 299 The Paleo-Indian Entry into South America According to Mitogenomes
Stefania Brandini, Paola Bergamaschi, Marco Fernando Cerna, Francesca Gandini, Francesca Bastaroli, Emilie Bertolini, Cristina Cereda, Luca Ferretti, Alberto Gómez-Carballa, Vincenza Battaglia, Antonio Salas, Ornella Semino, Alessandro Achilli, Anna Olivieri, and Antonio Torroni
- 312 The Rate of Evolution of Postmating-Prezygotic Reproductive Isolation in *Drosophila*
David A. Turissini, Joseph A. McGirr, Sonali S. Patel, Jean R. David, and Daniel R. Matute
- 335 Effects of Demographic History on the Detection of Recombination Hotspots from Linkage Disequilibrium
Amy L. Dapper and Bret A. Payseur
- 354 Horizontal Transfer of Non-LTR Retrotransposons from Arthropods to Flowering Plants
Dongying Gao, Ye Chu, Han Xia, Chunming Xu, Karolina Heyduk, Brian Abernathy, Peggy Ozias-Akins, James H. Leebens-Mack, and Scott A. Jackson
- 365 Evolutionary Epigenomics of Retrotransposon-Mediated Methylation Spreading in Rice
Jae Young Choi and Michael D. Purugganan
- 383 Mode and Rate of Evolution of Haemosporidian Mitochondrial Genomes: Timing the Radiation of Avian Parasites
M. Andreína Pacheco, Nubia E. Matta, Gediminas Valkiūnas, Patricia G. Parker, Beatriz Mello, Craig E. Stanley Jr, Miguel Lentino, Maria Alexandra Garcia-Amado, Michael Cranfield, Sergei L. Kosakovsky Pond, and Ananias A. Escalante
- 404 Reciprocal Nucleopeptides as the Ancestral Darwinian Self-Replicator
Eleanor F. Banwell, Bernard M.A.G. Piette, Anne Taormina, and Jonathan G. Heddle
- 417 Multiple Regulatory Modules Are Required for Scale-to-Feather Conversion
Ping Wu, Jie Yan, Yung-Chih Lai, Chen Siang Ng, Ang Li, Xueyuan Jiang, Ruth M. Elsey, Randall Wideltz, Ruchi Bajpai, Wen-Hsiung Li, and Cheng-Ming Chuong
- 431 Landscape of DNA Methylation on the Marsupial X
Shafagh A. Waters, Alexandra M. Livernois, Hardip Patel, Denis O'Meally, Jeff M. Craig, Jennifer A. Marshall Graves, Catherine M. Suter, and Paul D. Waters
- 440 Human-Specific Mutations and Positively Selected Sites in MARCO Confer Functional Changes
Kyle E. Novakowski, Nicholas V.L. Yap, Charles Yin, Kaori Sakamoto, Bryan Heit, G. Brian Golding, and Dawn M.E. Bowdish
- 451 Origin and Evolution of the *Bartonella* Gene Transfer Agent
Daniel Tamarit, Minna-Maria Neuvonen, Philipp Engel, Lionel Guy, and Siv G.E. Andersson

(Contents continued)

- 465 Sex Differences in 20-Hydroxyecdysone Hormone Levels Control Sexual Dimorphism in *Bicyclus anynana* Wing Patterns
Shivam Bhardwaj, Kathleen L. Prudic, Ashley Bear, Mainak Dasgupta, Bethany R. Wasik, Xiaoling Tong, Wei Fun Cheong, Markus R. Wenk, and Antónia Monteiro
- 473 Combination of 247 Genome-Wide Association Studies Reveals High Cancer Risk as a Result of Evolutionary Adaptation
Konstantinos Voskarides
- METHODS
- 486 Evaluating Fast Maximum Likelihood-Based Phylogenetic Programs Using Empirical Phylogenomic Data Sets
Xiaofan Zhou, Xing-Xing Shen, Chris Todd Hittinger, and Antonis Rokas
- 504 Bayesian Inference of Species Networks from Multilocus Sequence Data
Chi Zhang, Huw A. Ogilvie, Alexei J. Drummond, and Tanja Stadler
- RESOURCES
- 518 UFBoot2: Improving the Ultrafast Bootstrap Approximation
Diep Thi Hoang, Olga Chernomor, Arndt von Haeseler, Bui Quang Minh, and Le Sy Vinh
- NEWS
- Joseph Caspermeyer*
- 523 Scientists Find Evidence Our Best Friends, Dogs, Similarly Adapted to Malaria in Africa
- 523 Finding Their Inner Bird: Using Modern Genomics to Turn Alligator Scales into Birdlike Feathers

UFBoot2: Improving the Ultrafast Bootstrap Approximation

Diep Thi Hoang,^{†,1} Olga Chernomor,^{†,2} Arndt von Haeseler,^{2,3} Bui Quang Minh,^{*,2} and Le Sy Vinh^{*,1}

¹Faculty of Information Technology, University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam

²Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, Medical University Vienna, Vienna, Austria

³Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Vienna, Austria

[†]These authors contributed equally to this work.

***Corresponding authors:** E-mails: minh.bui@univie.ac.at; vinhls@vnu.edu.vn.

Associate editor: Michael S. Rosenberg

Abstract

The standard bootstrap (SBS), despite being computationally intensive, is widely used in maximum likelihood phylogenetic analyses. We recently proposed the ultrafast bootstrap approximation (UFBoot) to reduce computing time while achieving more unbiased branch supports than SBS under mild model violations. UFBoot has been steadily adopted as an efficient alternative to SBS and other bootstrap approaches. Here, we present UFBoot2, which substantially accelerates UFBoot and reduces the risk of overestimating branch supports due to polytomies or severe model violations. Additionally, UFBoot2 provides suitable bootstrap resampling strategies for phylogenomic data. UFBoot2 is 778 times (median) faster than SBS and 8.4 times (median) faster than RAXML rapid bootstrap on tested data sets. UFBoot2 is implemented in the IQ-TREE software package version 1.6 and freely available at <http://www.iqtree.org>.

Key words: phylogenetic inference, ultrafast bootstrap, maximum likelihood, model violation, polytomies.

Standard nonparametric bootstrap (SBS) (Efron 1979; Felsenstein 1985) is widely used in maximum likelihood (ML) phylogenetic analyses to estimate branch supports of a phylogenetic tree inferred from a multiple sequence alignment (MSA). To achieve this, SBS generates a large number of resampled MSAs and reconstructs an ML-tree for each bootstrapped MSA. The resulting bootstrap ML trees are then used either to compute branch supports for the ML-tree reconstructed from the original MSA or to build a consensus tree with support values.

Although fast ML-tree search algorithms exist for large data sets (Vinh and von Haeseler 2004; Stamatakis 2006; Guindon et al. 2010; Nguyen et al. 2015) SBS is still very computationally intensive. To improve computing time, rapid bootstrap (RBS; Stamatakis et al. 2008) and the ultrafast bootstrap (UFBoot; Minh et al. 2013) were developed. Although RBS resembles the conservative behavior of SBS (i.e., underestimating branch supports), UFBoot provides relatively unbiased bootstrap estimates under mild model misspecifications.

The key idea behind UFBoot is to keep trees encountered during the ML-tree search for the original MSA and to use them to evaluate the tree likelihoods for the bootstrap MSAs. To speed up likelihood computation even further for bootstrap MSAs, IQ-TREE employed the resampling estimated log-likelihood (RELL) strategy (Kishino et al. 1990). For each bootstrap MSA, the tree with the highest RELL score (RELL-tree) represents the ML-bootstrap tree. Contrary to SBS, UFBoot does not further ML optimize this tree. The discrepancy in branch supports between UFBoot and SBS emerges as bootstrap trees inferred by UFBoot and SBS might be different.

Here, we present UFBoot2 that substantially speeds up UFBoot and reduces the risk for overestimated branch support due to polytomies or severe model violations. We also discuss several resampling strategies for phylogenomics data recently implemented in UFBoot2. In the following, we will outline these improvements.

Accelerating UFBoot

The likelihood computation is the major runtime bottleneck of all ML software because it lies at the core of all analyses. The pruning algorithm (Felsenstein 1981) efficiently computes the likelihood of phylogenetic trees, but still does not scale well for large data sets. Therefore, we adopted a modification to Felsenstein's algorithm (see supplementary method, Supplementary Material online), first introduced in RAXML. The modification exploits the reversible property of models of sequence evolution typically used in phylogenetic analysis, which led to a theoretical speedup of 4 (for DNA) or 20 (for protein data) when estimating branch lengths. Moreover, we employed the SIMD (single instruction, multiple data) feature to concurrently compute the likelihood of two MSA sites with streaming SIMD extensions or four MSA sites with advanced vector extensions, thus leading to a theoretical speedup of two or four compared with a non-SIMD implementation. IQ-TREE code was further optimized to avoid redundant computations.

We benchmark the runtimes on 70 DNA and 45 protein MSAs (DOI 10.5281/zenodo.854445) from TreeBASE, previously analyzed in Nguyen et al. (2015). The command-lines used to perform bootstrap methods are provided in

supplementary table S1, Supplementary Material online. UFBoot2 achieved a median speedup of 2.4 times (maximum: 77.3) compared with UFBoot version 0.9.6 (released on October 20, 2013).

Correction for Polytomies

Polytomies refer to multifurcating nodes in the tree that cannot be resolved due to low phylogenetic signal in the data. However, phylogenetic reconstruction always assumes strictly bifurcating trees. When resolving polytomies, there might be multiple equivalently optimal bifurcating trees (Whelan and Money 2010). As UFBoot (and other bootstrap approaches) saves only a single optimal bifurcating tree for each bootstrap MSA, it might cause overoptimistic bootstrap supports for short branches (Simmons and Norton 2014).

To correct for this shortcoming, UFBoot2 implemented the following technique. Instead of assigning the bootstrap tree with the highest RELL for each bootstrap MSA, UFBoot2 will randomly select one of the trees encountered during tree search, whose RELL score is less than ε_{boot} (default: 0.5) away from the highest RELL. As a result, UFBoot2 will not give high supports for branches resolving the multifurcations.

It was shown with a star tree simulation that SBS and RBS sometimes led to false positives (bootstrap supports $\geq 95\%$ for nonexisting branches), whereas with this technique UFBoot never supported such branches (support values $\leq 88\%$) (Simmons and Norton 2014). We repeated the star tree simulation for UFBoot2 with the same setting as proposed in (Simmons and Norton 2014). We used Seq-Gen 1.3.2x (Rambaut and Grass 1997) to evolve 100 DNA MSAs, each of 15,000 sites, along a 4-taxon star tree with four terminal branch lengths of 0.05, under JC model. For each MSA, we performed UFBoot2 runs under JC and GTR+ Γ , each with 1,000 bootstrap replicates and up to 1,000 search iterations (invoked in IQ-TREE via “-bcor 1” option). The simulation results show that UFBoot2 resembles the original UFBoot in that it never supports nonexisting branches (support values $\leq 88\%$).

Reducing the Impact of Model Violations

Minh et al. (2013) showed that severe model violations inflate UFBoot support values. To resolve this issue, UFBoot2 provides an option to conduct an additional step once the tree search on the original MSA is completed. Here, the best RELL-trees are further optimized using a hill-climbing nearest-neighbor interchange (NNI) search based directly on the corresponding bootstrap MSA. Thus, this extra step operates like SBS, but with a quick tree search to save time. Bootstrap supports are then summarized from the resulting corrected bootstrap trees. In the following, we called this UFBoot2 + NNI, which can be invoked in IQ-TREE via “-bnni” option.

We repeated the PANDIT simulations (Minh et al. 2013) to compare the accuracy of UFBoot2 and UFBoot2 + NNI with SBS (1,000 replicates using IQ-TREE) and RBS (RAXML bootstopping criterion). The simulations include 5,690 DNA MSAs (DOI 10.5281/zenodo.854445) generated by Seq-Gen

(Rambaut and Grass 1997), where the model parameters and the tree (which we will call the true tree in the following) were inferred from the original MSAs downloaded from the PANDIT database (Whelan et al. 2006). The accuracy of a bootstrap method M is defined by $f_M(x)$, the percentage of branches with support value x (across all reconstructed trees) that occur in the true tree (Hillis and Bull 1993). Thus, $f_M(x)$ reflects the probability that a branch with support x is a true branch. Figure 1 shows the results [y -axis depicts $f_M(x)$]. If the sequence evolution model used to infer the ML-tree agrees with the model used for simulations, then SBS, RBS, and UFBoot2 + NNI underestimated branch supports, the latter to a lower degree (fig. 1A; curves above the diagonal). This conservative behavior of SBS and RBS corroborates previous studies (Hillis and Bull 1993; Minh et al. 2013). Whereas UFBoot2 obtained almost unbiased branch supports (fig. 1A; curve close to the diagonal), that is, closely matching the true probability of branches being correct. Thus, UFBoot2 resembles the behavior of the original UFBoot (Minh et al. 2013).

Severe model violations do not influence SBS (fig. 1B; RBS not shown because RAXML does not support simpler models). However, UFBoot2 (like UFBoot) overestimated the branch supports (fig. 1B; curve below the diagonal), whereas UFBoot2 + NNI only slightly underestimated the bootstrap values (fig. 1B; curve closest to the diagonal). Thus, UFBoot2 + NNI helps to overcome the problem of unduly high supports by UFBoot2 in the presence of severe model violations.

In terms of computation times based on the analysis of 115 benchmark MSAs, UFBoot2, and UFBoot2 + NNI showed a median speedup of 778 (range: 200–1,848) and 424 (range: 233–749) compared with SBS, respectively. Compared with RBS, UFBoot2, and UFBoot2 + NNI are 8.4 (range: 1.5–51.2) and 5.0 (range: 0.8–32.6) times faster, respectively. Therefore, UFBoot2 + NNI is two times (median) slower than UFBoot2. Supplementary Figures S1–S3, Supplementary Material online, show the distributions of runtime ratios between SBS/RBS/UFBoot and UFBoot2/UFBoot2 + NNI.

We conclude that UFBoot2 and UFBoot2 + NNI are fast alternatives to other bootstrap approaches. Under no or mild model violations, UFBoot2 has the interpretation of unbiased bootstrap support as suggested for UFBoot (Minh et al. 2013). That is, one can trust branches with UFBoot2 support $\geq 95\%$. Users are advised to apply model violation detection methods (Goldman 1993; Weiss and von Haeseler 2003; Nguyen et al. 2011) before bootstrap analyses. UFBoot2 + NNI should be applied if severe model violations are present in the data set at hand.

Resampling Strategies for Phylogenomic Data

Recent phylogenetic analyses are typically based on multiple genes to infer the species tree, the so-called phylogenomics. To facilitate phylogenomic analysis, UFBoot2 implements several bootstrap resampling strategies: i) resampling MSA-sites within partitions (denoted as MSA-site resampling as the default option), ii) resampling genes instead of MSA-sites

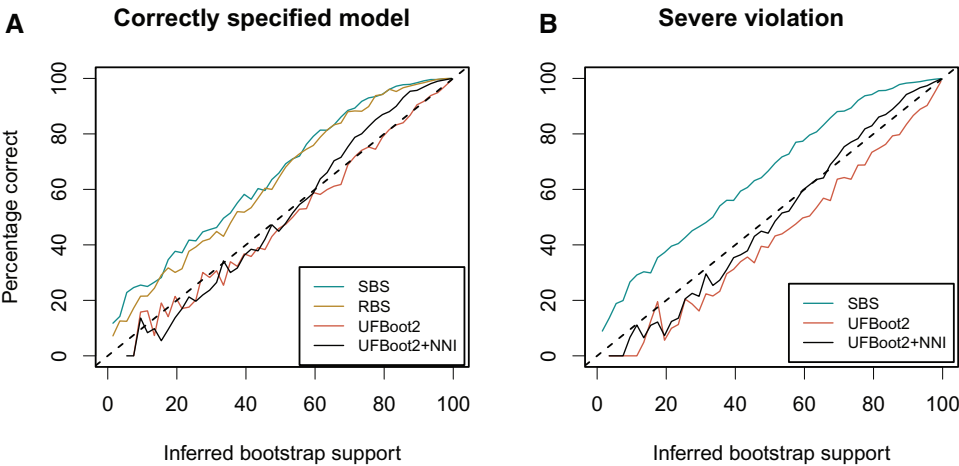


Fig. 1. Accuracy of the standard bootstrap (SBS), RAxML rapid bootstrap (RBS), ultrafast bootstrap (UFBoot2) and UFBoot2 with correction (UFBoot2 + NNI) for (A) correctly specified models and (B) severely misspecified models. The y-axis depicts the percentage of all branches with support value x (across all reconstructed trees) that occur in the true tree. Curves above the diagonal indicate underestimation of branch supports whereas curves below the diagonal indicate overestimation. For each point (x, y) in the curve representing the accuracy of bootstrap method M , x is an inferred bootstrap value by method M whereas y measures the probability of branches assigned by M with support value x to be true branches, that is, occurring on the true tree. Specifically, let X^+ (X^-) be the set of branches with support value x in all trees and present (absent) in the true tree. The y value is computed as the ratio between $|X^+|$ and $|X|$, where $|X| = |X^+| + |X^-|$.

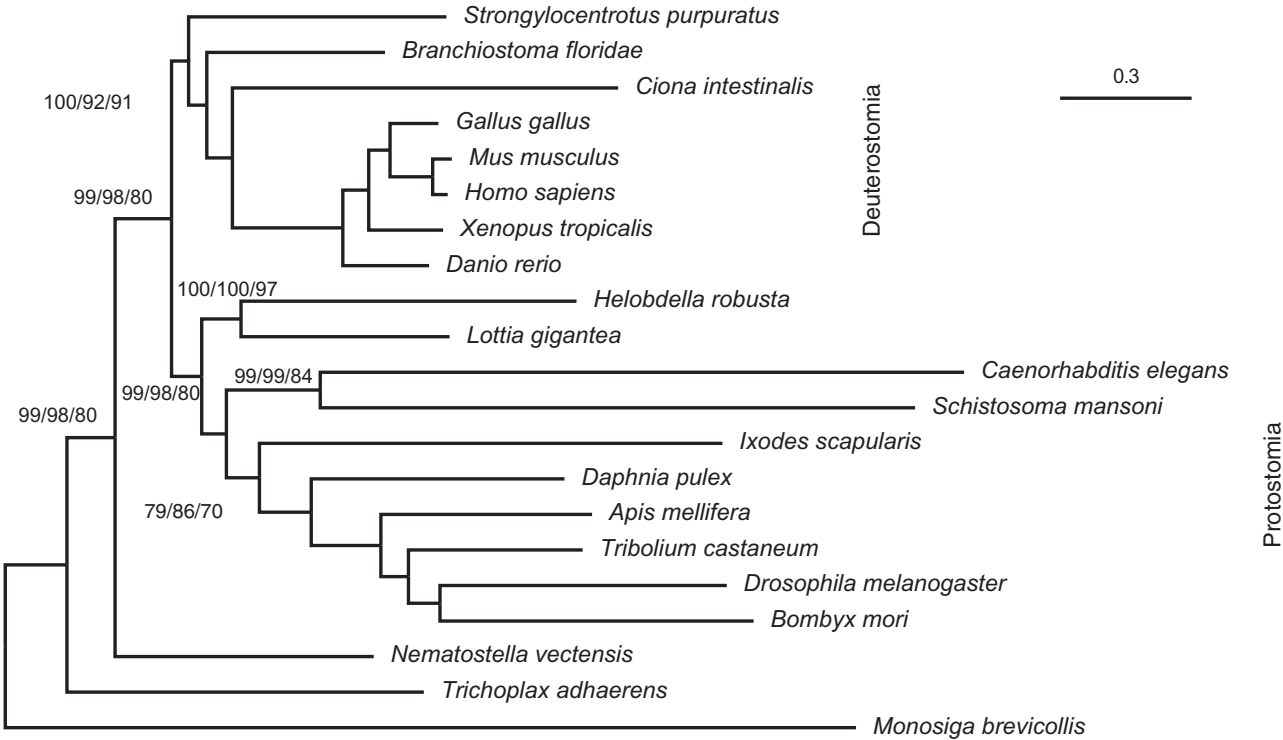


Fig. 2. Maximum-likelihood tree inferred under the edge-unlinked partition model. Numbers attached to the branches show the UFBoot2 bootstrap supports using MSA-site, gene, and gene-site resampling strategies (omitted when all three supports are 100%).

(gene-resampling invoked via “-bsam GENE” option), and iii) resampling genes and subsequently resamples MSA-sites within each gene (gene-site resampling, invoked via “-bsam GENESITE” option) (Gadagkar et al. 2005). Strategy (i) preserves the number of MSA-sites for all genes in the bootstrap MSAs, whereas strategies (ii) and (iii) will lead to different number of sites in the bootstrap MSAs.

To investigate the impact of the three resampling strategies, we reanalyzed the metazoan data with 21 species, 225 genes, and a total of 171,077 amino-acid sites (Salichos and Rokas 2013). Figure 2 shows the ML tree inferred with IQ-TREE under edge-unlinked partition model (Chernomor et al. 2016), which allows separate sets of branch lengths across partitions. The tree replicates previous results (Salichos and

Rokas 2013) and shows the Protostomia clade (Telford et al. 2015). However, discrepancies between resampling strategies are observed: while MSA-site and gene-resamplings obtained high supports (>95%) for branches along the backbone of the tree (fig. 2; bold lines), lower supports (80%) were estimated by gene-site resampling.

By further examining 14 other empirical data sets (Bouchenak-Khelladi et al. 2008; Fabre et al. 2009; Stamatakis and Alachiotis 2010; van der Linde et al. 2010; Pyron et al. 2011; Nyakatura and Bininda-Emonds 2012; Springer et al. 2012; Hinchliff and Roalson 2013; Salichos and Rokas 2013; Dell'Ampio et al. 2014), we observed more discrepancies between resampling strategies (data not shown). Exceptionally, for some data sets, a number of branches showed almost no support ($\leq 10\%$) for one resampling but high supports ($\geq 95\%$) for the other two resampling strategies. However, there is no tendency toward systematically lower supports obtained by one resampling strategy.

Taking into account the above findings, we recommend to apply all alternative resampling strategies. If similar bootstrap supports are obtained, then one can be more confident about the results.

Conclusions

UFBoot2 significantly improves speed and accuracy of bootstrap values compared with UFBoot. It also offers new functionalities in the presence of model violations and in its applicability to phylogenomic data. In general, since SBS, RBS, and UFBoot2 + NNI share a disadvantage of being conservative, more research is necessary to understand the different biases introduced by the available phylogenetic bootstrap estimation methods.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by Vietnam National Foundation for Science and Technology Development (102.01-2013.04). A.V.H., B.Q.M., and O.C. were supported by the Austrian Science Fund—FWF (grant nos. I-2805-B29 and I-1824-B22). The authors thank Stephen Crotty and two anonymous reviewers for helpful comments on the manuscript.

References

- Bouchenak-Khelladi Y, Salamin N, Savolainen V, Forest F, van der Bank M, Chase MW, Hodkinson TR. 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Mol Phylogenet Evol.* 47(2):488–505.
- Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol.* 65(6):997–1008.
- Dell'Ampio E, Meusemann K, Szucsich NU, Peters RS, Meyer B, Borner J, Petersen M, Aberer AJ, Stamatakis A, Walz MG, et al. 2014. Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects. *Mol Biol Evol.* 31(1):239–249.
- Efron B. 1979. Bootstrap methods: another look at the jackknife. *Ann Stat.* 7(1):1–26.
- Fabre P-H, Rodrigues A, Douzery EJP. 2009. Patterns of macroevolution among Primates inferred from a supermatrix of mitochondrial and nuclear DNA. *Mol Phylogenet Evol.* 53(3):808–825.
- Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol.* 17(6):368–376.
- Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution (NY)* 39(4):783–791.
- Gadagkar SR, Rosenberg MS, Kumar S. 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *J Exp Zool B Mol Dev Evol.* 304B(1):64–74.
- Goldman N. 1993. Statistical tests of models of DNA substitution. *J Mol Evol.* 36(2):182–198.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59(3):307–321.
- Hillis DM, Bull JJ. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol.* 42(2):182–192.
- Hinchliff CE, Roalson EH. 2013. Using supermatrices for phylogenetic inquiry: an example using the sedges. *Syst Biol.* 62(2):205–219.
- Kishino H, Miyata T, Hasegawa M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J Mol Evol.* 31(2):151–160.
- Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol.* 30(5):1188–1195.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
- Nguyen MAT, Klaere S, von Haeseler A. 2011. MISFITS: evaluating the goodness of fit between a phylogenetic model and an alignment. *Mol Biol Evol.* 28(1):143–152.
- Nyakatura K, Bininda-Emonds ORP. 2012. Updating the evolutionary history of Carnivora (Mammalia): a new species-level super-tree complete with divergence time estimates. *BMC Biol.* 10(1):12.
- Pyron RA, Burbrink FT, Colli GR, de Oca ANM, Vitt LJ, Kuczyński CA, Wiens JJ. 2011. The phylogeny of advanced snakes (Colubroidea), with discovery of a new subfamily and comparison of support methods for likelihood trees. *Mol Phylogenet Evol.* 58(2):329–342.
- Rambaut A, Grass NC. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics* 13(3):235–238.
- Salichos L, Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497(7449):327–331.
- Simmons MP, Norton AP. 2014. Divergent maximum-likelihood-branch-support values for polytomies. *Mol Phylogenet Evol.* 73:87–96.
- Springer MS, Meredith RW, Gatesy J, Emerling CA, Park J, Rabosky DL, Stadler T, Steiner C, Ryder OA, Janečka JE, et al. 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. Stanýon R, editor. *PLoS One* 7:e49521.
- Stamatakis A. 2006. RAXML-VI-HP: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.
- Stamatakis A, Alachiotis N. 2010. Time and memory efficient likelihood-based tree searches on phylogenomic alignments with missing data. *Bioinformatics* 26(12):i132–i139.
- Stamatakis A, Hoover P, Rougemont J, Renner S. 2008. A rapid bootstrap algorithm for the RAXML web servers. *Syst Biol.* 57(5):758–771.
- Telford MJ, Budd GE, Philippe HH. 2015. Phylogenomic insights into animal evolution. *Curr Biol.* 25(19):R876–R887.

- van der Linde K, Houle D, Spicer GS, Steppan SJ. 2010. A supermatrix-based molecular phylogeny of the family Drosophilidae. *Genet Res (Camb)* 92(1):25–38.
- Vinh LS, von Haeseler A. 2004. IQPNNI: moving fast through tree space and stopping in time. *Mol Biol Evol.* 21(8):1565–1571.
- Weiss G, von Haeseler A. 2003. Testing substitution models within a phylogenetic tree. *Mol Biol Evol.* 20(4):572–578.
- Whelan S, de Bakker PIW, Quevillon E, Rodriguez N, Goldman N. 2006. PANDIT: an evolution-centric database of protein and associated nucleotide domains with inferred trees. *Nucleic Acids Res.* 34:D327–D331.
- Whelan S, Money D. 2010. The prevalence of multifurcations in tree-space and their implications for tree-search. *Mol Biol Evol.* 27(12):2674–2677.