Published online 15 September 2022

NAR Genomics and Bioinformatics, 2022, Vol. 4, No. 3 **1**
https://doi.org/10.1093/nargab/lqac064

Pitfalls of the site-concordance factor (sCF) as measure of phylogenetic branch support

Patrick Kück[1,*], Juliane Romahn[1,2,3] and Karen Meusemann[4]

[1]Centre for Molecular Biodiversity Research, Leibniz Institute for the Analysis of Biodiversity Change, Adenauerallee 160, 53113 Bonn, Germany, [2]LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Senckenberganlage 25, 60325 Frankfurt am Main, Germany, [3]Senckenberg Society for Nature Research, Senckenberganlage 25, 60325 Frankfurt am Main, Germany and [4]Directorate, Leibniz Institute for the Analysis of Biodiversity Change, Adenauerallee 160, 53113 Bonn, Germany

Received March 18, 2022; Revised August 10, 2022; Editorial Decision August 15, 2022; Accepted August 17, 2022

ABSTRACT

Confidence measures of branch reliability play an important role in phylogenetics as these measures allow to identify trees or parts of a tree that are well supported by the data and thus adequate to serve as basis for evolutionary inference of biological systems. Unreliable branch relationships in phylogenetic analyses are of concern because of their potential to represent incorrect relationships of interest among more reliable branch relationships. The site-concordance factor implemented in the IQ-TREE package is a recently introduced heuristic solution to the problem of identifying unreliable branch relationships on the basis of quartets. We test the performance of the site-concordance measure with simple examples based on simulated data and designed to study its behaviour in branch support estimates related to different degrees of branch length heterogeneities among a ten sequence tree. Our results show that in particular in cases of relationships with heterogeneous branch lengths site-concordance measures may be misleading. We therefore argue that the maximum parsimony optimality criterion currently used by the site-concordance measure may sometimes be poorly suited to evaluate branch support and that the scores reported by the site-concordance factor should not be considered as reliable.

INTRODUCTION

As a measure of confidence in phylogenetic trees, the computation and evaluation of branch support is an essential element of testing evolutionary hypotheses on the basis of trees. In molecular phylogenetics, various approaches exist for the assignment of support values to branches in trees to validate biological hypotheses (1–3). Well known measures for testing the statistical significance of a particular branching pattern are for example non-parametric bootstrap support (4,5), Bayesian posterior probabilities (6–8), or the interior branch test (9–11). Especially non-parametric bootstrap measures and Bayesian posterior probabilities have been established as methods of choice. Despite their popularity, both posterior probabilities and bootstrap measures are sensitive to model misspecification and can result in highly overestimated branch support (e.g. (12–17)). Given the sensitivity of both methods due to incorrect model assumptions and the linked problem of the unreliability of inflated branch support, it is beneficial to verify the robustness of phylogenetic branch events through additional support analyses. One recently published approach to complement Bayesian posterior probabilities and bootstrap support implies the calculation of a site concordance factor (sCF, (18)), which is part of the popular IQ-TREE package (18,19). Since the significance levels obtained by different methods for the same phylogenetic tree do not necessarily agree with each other, it is important to understand the statistical property of this method. The sCF is defined as the average percentage of maximum-parsimony (MP) informative site patterns in quartet-related sub-alignments, supporting a branch x of a given bifurcating and unrooted reference tree (18). Each quartet consists of one representative leaf sequence of each of the four branch x related subtrees with quartets drawn randomly from the overall pool of possible quartets if there are more than a user-specified number of quartets.

A site pattern is MP informative if it comprises at least two distinct characters, from which each character appears at least twice. Thus for a quartet of sequences, only three different MP informative site patterns exist (xxyy, xyxy and xyyx), each of them supporting one of three possible unrooted quartet phylogenies. An MP informative site pattern is concordant with branch x if it supports the reference tree branch x associated bipartition, e.g. pattern {yy}|{zz} is concordant for an internal branch ((A_y,B_y),(Cz, Dz))

*To whom correspondence should be addressed. Tel: +49 228 9122 404; Fax: +49 228 9122 212; Email: patrick_kueck@web.de

and discordant to an internal branch supporting either ((A$_y$,Cz),(B$_y$, Zz)) or ((A$_y$,Dz),(B$_y$, Cz)).

However, the usage of exclusively MP informative sites as measure of branch confidence in phylogenetic trees contains a number of statistical uncertainties. MP minimises the number of character substitutions on a quartet tree by assigning similar character states to the same interior node on the tree. This means that MP does not account for different branch lengths. Nevertheless, branch lengths have a significant impact on the tree inference, because they indicate the number of character substitutions or rather the evolutionary rate for a given branch. The longer a branch, the higher the probability of (single and multiple) state substitutions along that branch. Thus, degrees of length heterogeneities between short and long branches are an indication for evolutionary rate differences between simulated sequences. More distantly related sequence pairs of longer branches may have a higher probability of sharing the same character state because of parallel or convergent changes than closely related pairs of longer and shorter branches that have same character states due to a common ancestor. Omitting a correction for rate heterogeneity among lineages (20) and among sites (21) makes MP particularly susceptible to long-branch attraction (see also (22–31)).

Given a sequence quartet, this means: if branch lengths of non-directly related sequences are long enough, the number of MP informative site patterns for the correct tree may be smaller than the number of patterns which support an adjacent relationship of the two long branches, i.e. a wrongly inferred relationship caused by long branch attraction. Felsenstein (20) was the first who published a remarkable example of a quartet tree showing that MP can fail to be consistent even with an infinite amount of data if parallelism alongside longer non-adjacent related branches is expected to occur frequently. Several other studies found similar inconsistency results for parsimony in the case of quartets (e.g. (24,27,32–36)). Various studies of the last century have shown that even the existence of a molecular clock combined with a small amount of evolution does not guarantee the consistency of MP (37–39). MP inconsistency can even occur when the rate of nucleotide substitutions is constant (37,38,40), because unequal branch lengths can be caused by either unequal substitution rates, or as a consequence of an asymmetric topology (30). MP can be even inconsistent if all branches are of the same length, therefore the tree needs to be asymmetrical (26).

Furthermore, tree reconstruction methods from molecular sequences have been shown as unreliable if the base composition varies between sequences, a problem which was first identified in the late 1980s and early 1990s (e.g. (41–53)). There are only three MP informative patterns for a quartet of sequences: xxyy, xyxy and xyyx. Thus, for four sequences, the sCF measure uses only 36 (3 × 12) out of 256 possible site patterns. As stated by Yang (27), this is not an efficient use of the phylogenetic information content provided by the data. MP 'noninformative sites', like xxxy, xxyz and even 'invariable' (xxxx) or 'complete variable' (xywz) site patterns do contain phylogenetic information. For instance, differences of state frequencies in the data can provide information of substitution probabilities. Combining patterns like CCAA, CCGG, CCTT into a unique category (xxyy) implies that substitutions between different site characters (transitions and transversions), are expected to occur only with equal probability (27). It has been shown that the exclusion of base compositional differences among sequences, e.g. the case in which two unrelated taxa have independently acquired the same GC content, enhances assembling of sequences of similar composition rather than the sequences that last shared a common ancestor (42,44). Jermiin *et al.* (53) demonstrated on quartet simulations that the length of the internal branch is the decisive factor for finding the correct MP tree. The authors showed that for an internal branch length of 0.05, the frequency of correct MP reconstructions reached 0% when the difference in the GC content exceeded ≈15%. Reducing the internal branch length further to 0.01 produced similar results for much smaller compositional differences (≈10%).

Comparisons of different tree reconstruction methods based on different sequence simulations have shown that the tree success of MP in simulations is often worse than distance matrix methods (22–24,54–58).

The particular message widely adopted from those studies is that the MP method tends to infer trees incorrectly when there are non-directly related long branches in the tree. The reason is that long branches reconstructed next to each other gain more MP informative site patterns due to the accumulation of chance similarities than non-directly related long branches. Convergence due to a similar base composition in non-sister lineages is similarly misleading (compositional bias).

Since the sCF measure depends exclusively on quartet inferred MP informative site patterns, it can reasonably be assumed that both, branch length and base compositional heterogeneity in the data, have a negative impact on the outcome of the sCF analysis: reference trees with next related long-branch taxa or similar compositional state frequencies will be continuously overestimated, whereas internal branches of reference trees which represent the opposite case will be underestimated. Both scenarios undermine the usability of the sCF as a measure of confidence in phylogenetic trees.

We tested the efficiency of the IQ-TREE implemented sCF measure (18) as branch support estimation method under a wide range of nucleotide data simulations following different degrees of branch length combinations and among-site rate heterogeneity parameters of a ten-sequence tree. In total, our study comprises the analysis of 24,000 alignment simulations. We used maximum-likelihood (ML) for the calculation of alignment corresponding reference trees. Due to the comprehensive amount of simulations, we exclusively focused on the impact of rate heterogeneity among lineages and among sites and neglected tests on the impact of base compositional bias to the sCF measure.

MATERIALS AND METHODS

Simulation setup

For our test setting, we simulated nucleotide data of a ten-sequence tree using INDELible version 1.03 (59) with ten different combinations of fixed and pairwise elongated internal branches and a single root position with always the

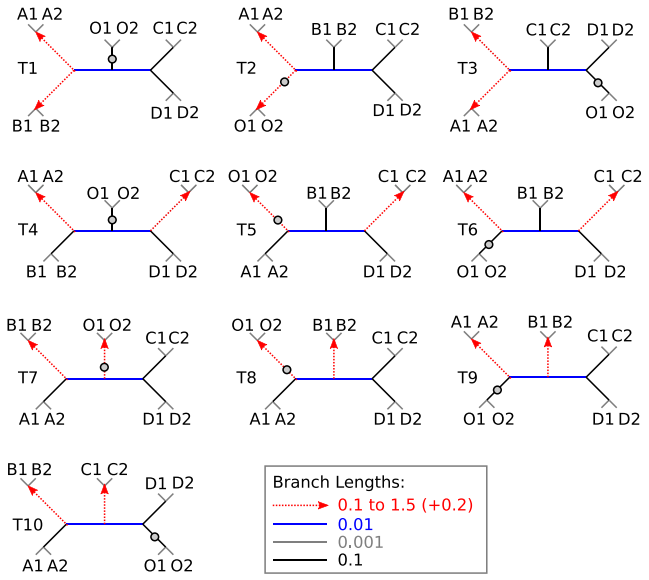

Figure 1. Simulation setup. Ten different combinations of pairwise internal branch elongation (red) of ten sequences in relation to the root position (dot marker) with sequence O1 and O2 defined as outgroup. Backbone-tree interlinked internal branches were simulated continuously short with a length of 0.01 (blue). Other internal branches were simulated with a length of 0.1 (black). To keep signal for pairwise adjacently related terminal branches comparatively strong, all terminal branches were set to 0.001 (grey). Overall, the simulation setup involved three symmetric trees with pairwise elongation either directly connected (T4), or separated by one (T7) or two backbone branches (T1), seven asymmetric trees with pairwise elongation either next related (T5, T6), or again separated by one (T8, T9, T10) or two backbone branches (T2, T3). The impact of outgroup affected branch elongation was thereby tested on four trees (T2, T5, T7, T8).

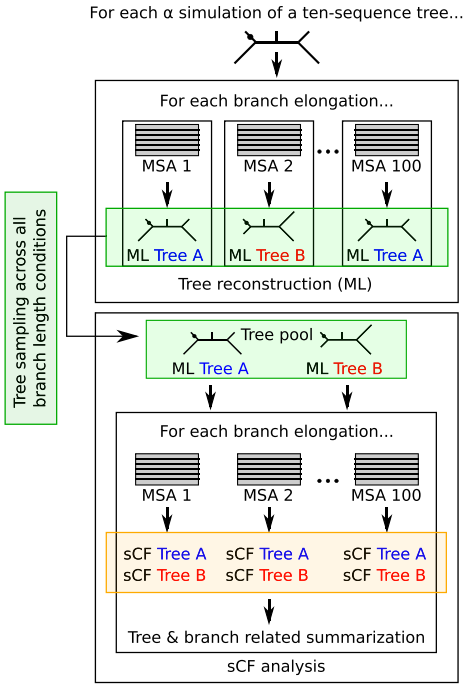

Figure 2. Overview of our analysis setup. For each rate heterogeneity parameter (α) of a ten-sequence tree simulation, ML trees were inferred for each of the 800 simulated alignments (100 alignment replicates per branch elongation). Different ML tree topologies were subsequently sampled (highlighted green). For the sCF test analysis, we evaluated all ML trees topologies of the sampled tree pool in combination with each of the 100 alignment replicates of a simulated branch condition and recorded for each step of branch elongation the average of single identified branch support.

same two sequences defined as outgroup (tree setup T1 to T10; Figure 1).

For each of the ten different tree settings, we simulated sequences of 250 kb. Internal branches which are not adjacent to a terminal branch were always set to 0.01. In each tree, we incrementally elongated (0.1 to 1.5 in steps of 0.2) all different combinations of two internal branches (Figure 1, highlighted in red) that are adjacent to the two innermost internal branches (Figure 1, highlighted in blue). Terminal branches were kept very small (0.001), all other branches were set to 0.1. Simulations used the GTR model of sequence evolution with an arbitrary relative rate matrix and set of state frequencies without indels or ambiguity states. State frequencies were simulated equally distributed with single frequencies of A, C, G and T set to 0.25 and with a relative rate matrix of C↔T: 0.2, A↔T: 0.4, G↔T: 0.6, A↔C: 0.8, C↔G: 1.2 and A↔G: 1.0. For each tree setup, among-site-rate-variation was modeled using a continuous Γ-rate distribution with three different rates of heterogeneity (α = 0.5, 1.0, 2.0) in combination with a fixed proportion of invariant sites (I = 0.3). Following the definition of Sullivan & Swofford (60), we chose the three different α scores to test rate-heterogeneity conditions from strong ($\alpha = 0.5$) to weak ($\alpha = 2.0$). For each branch length combination of elongated branch lengths, we generated 100 multiple sequence alignments, inferred an ML tree for each MSA and recorded the ML tree topologies (neglecting branch length differences).

Phylogenetic reconstruction

Maximum Likelihood (ML) trees were inferred from simulated data with IQ-TREE (19), using a mixed-distribution model (GTR+Γ+I) with estimated model parameters (α, I). The simulated continuous Γ-distribution was approximated from a discrete Γ-distribution with four relative substitution rate categories. The relative substitution rates and base compositions were estimated from the data.

Site concordance analysis

The efficiency of the site concordance factor (sCF) as measure of internal branch support was analysed separately for each α value of the ten tree settings. For each of the 800 alignment simulations (100 alignment replicates of each of the eight different combinations of fixed and pairwise elongated internal branches), we analysed the sCF branch support for all α value corresponding ML tree topologies. Finally, the identified sCF branch support of each ML tree was averaged individually for each alignment underlying branch length combination (Figure 2). All sCF analyses were executed using IQ-TREE version 2.1.3 (18) with default parameters, except for the number of single quartet analyses: to ensure all possible single quartet analyses along a given reference tree, we optionally set the maximum of single quartet analyses to 1,000.

Altogether, our simulation study comprised 24,000 alignments and 24,000 ML tree reconstructions, and subsequently 996,000 single conducted sCF analyses with respect to the different ML tree topologies (see supplementary file S1). Due to this large amount of single data processings, we used 'Appetite', a comprehensive module based Perl pipeline for both, process execution and result evaluation. To enable transparency of the source code, all analysis steps of this study have been subsequently summarised in a single script (GreaterGlider.pl), which can be downloaded from https://github.com/PatrickKueck/GreaterGlider.

RESULTS

Site concordance support of correct ML trees

Independent from the simulated branch length condition, internal branches are consistently strongly supported if they are adjacent to continuously very small (0.001) simulated terminal branches (sCF $\geq$ 99; see supplementary file S1).

Given non-elongated (0.1) internal branch relationships of the two smaller simulated innermost branches (0.01), the sCF measure always resulted in a support range of 40–59 for correct internal branch relationships (see Figure 3 and supplementary file S1), whereas support of incorrect internal branch relationships was generally in a lower support range (20–39; see Figure 4 and supplementary file S1).

With increasing branch lengths, the branch support of adjacent elongated branches (setup T1 to T3) increases steadily (Figure 3A and supplementary file S1). Conversely, the sCF support of non-adjacently simulated long-branch relationships (T4 to T10) decreases continuously with increasing branch length differences (Figure 3B, C and supplementary file S1). The average sCF support of non-adjacently simulated long-branch relationships is thereby always < 40 when elongated branches reach a length of $\geq$ 0.3 (30 times the length of the two innermost branches). In all trees in which none of the internal branches is elongated (length of 0.1, equally with 10 times the length of the two innermost branches), the sCF is still < 50 (Figure 3A–C and supplementary file S1).

Site concordance support of incorrect ML trees

In our simulations, incorrectness of inferred ML trees always affects the position of pairwise elongated internal branches.

Branch support of long-branch repulsion of adjacently simulated long branches is always poorly supported with decreasing support for increasing branch lengths (Figure 4A and supplementary file S1).

The branch support of long-branch attracted relationships is stronger than for correctly inferred relationships if branch elongation is $\geq$ 0.3 (30 times the length of the two innermost branches). In contrast, support of correctly inferred long-branch relationships is steadily decreasing (Figure 4B and supplementary file S1).

When two longer branches $\geq$ 0.3 are incorrectly more closely related to each other, but not fully long-branch attracted and the outgroup is kept short (0.1), the sCF support is slightly better than for correct branch relationships (see supplementary file S1).

We found exceptions in scenarios where the outgroup was one of the two longer branches, with incomplete long-branch relationships slightly lower or equally supported as correctly inferred relationships. Nevertheless, the obtained sCF support was still low (see supplementary file S1).

Impact of the rate heterogeneity parameter

Our study shows that the overestimation of support for adjacent long-branch relationships is affected by the degree of data adopted site-rate heterogeneity. The α values used in our simulations determine the number of site positions allowed to substitute. The higher the α value, the higher the number of substitutional site positions and thus, the stronger the accumulation of chance similarities among long branches and the higher the frequency of MP informative site patterns for adjacent related long-branches. This means, that the strong branch support identified for adjacent long branches and the low support identified for non-adjacent related long branches are slightly more pronounced if the simulated datasets are less site-rate heterogeneous ($\alpha = 2.0$), especially in cases of stronger branch length conditions (Figures 3 and 4).

Impact of the outgroup position

The position of the outgroup does not make much difference with respect to the sCF support. Both, symmetric and asymmetric, tree topologies result in similar sCF support for correct and incorrect branch relationships (see supplementary file S1).

DISCUSSION AND CONCLUSION

The understanding of phylogenetic relationships is essential when studying evolutionary processes such as character evolution, speciation, adaptation, causes of evolutionary change, or relations between genotype and phenotype. Therefore, the critical assessment of phylogenetic relationships provides an important foundation for the interpretation of all comparative biological data; the reliability of an inferred phylogenetic tree is an important task, especially with a view of potential sources of error such as incomplete lineage sorting, stochastic errors and, especially, systematic errors. Considering the infamous long-branch attraction artifact as one potential main source of systematic bias, a branch evaluation method should be capable of identifying equally reliable support for both, adjacent and nonadjacent long branch relationships - otherwise, there is no credibility for any of the two cases. As expected from the results of previously conducted long-branch studies on the tree reconstruction success of MP (e.g. (20,23,26)), our study shows that an exclusive usage of MP informative site pattern frequencies as measure of branch support is not satisfying.

Independent of the degree of branch elongation and the correctness of internal branch relationships, the sCF measure solely results in strong support if longer branches are placed together while alternative, more distinct long-branch relationships are continuously lower supported (Figures 3 and 4, and supplementary file S1). Moreover, the stronger

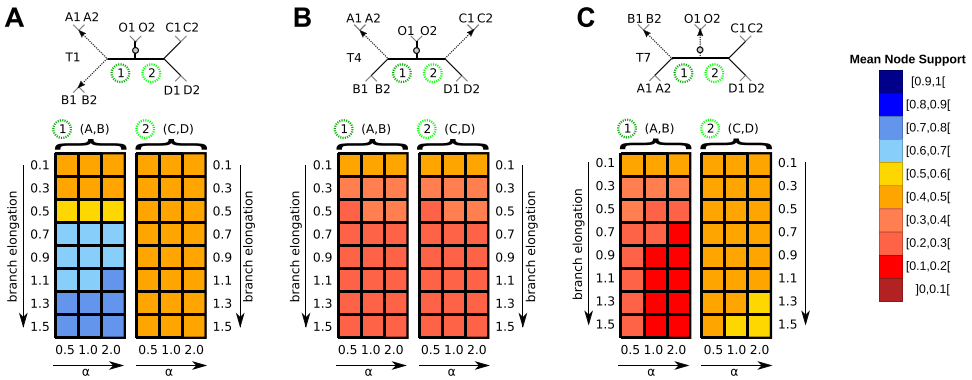

Figure 3. Exemplary examples of average sCF support ranges given correctly ML inferred internal branch relationships of three selected tree simulations (T1, T4, T7). As tree simulations with pairwise branch elongations divided by the same number of innermost branches resulted in the same range of sCF support (see supplementary file S1), we chose one representative tree for each of the three different branch elongation categories. (**A**) Individual sCF support of the two, very short (0.01) simulated innermost branches if the alignment underlying trees are based on adjacently pairwise elongated branches (T1 to T3, here T1). (**B**) Individual sCF support of the two innermost branches for alignment underlying trees with two non-adjacently elongated branches (T4 to T6, here T4), separated by two innermost branches. (**C**) Individual sCF support of the two innermost branches for scenario T7–T10 (here T7) where branch elongation takes place with one innermost branch in between.

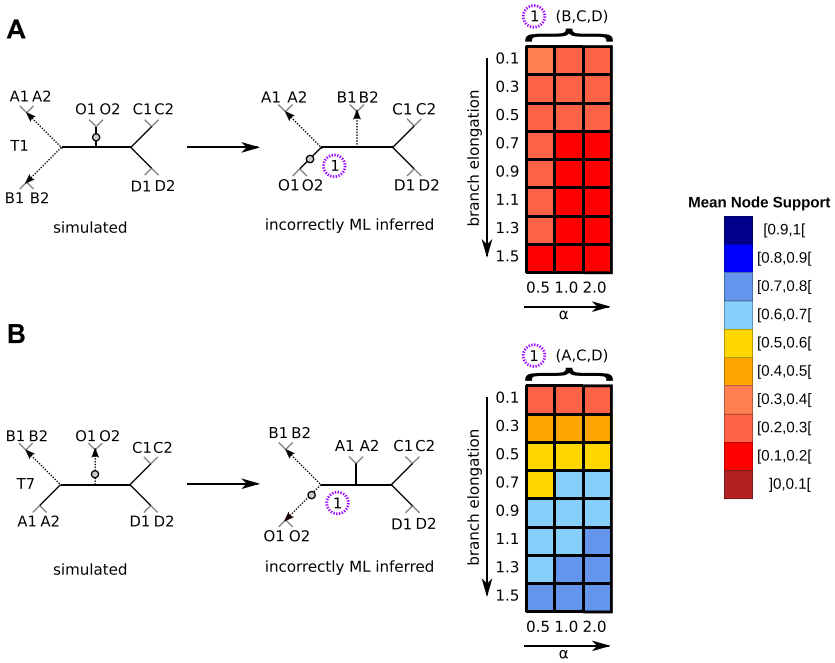

Figure 4. Exemplary examples of average sCF support ranges given incorrect long-branch relationships. For each step of branch elongation (0.1–1.5) and chosen α value (0.5, 1.0, 2.0) the average sCF support depends on 100 alignment replicates.(**A**) Branch support of incorrectly inferred long-branch repulsion based on ML with the original data based on adjacently simulated pairwise elongations in T1. (**B**) Increasing branch support of long-branch attracted elongated branches, inferred by ML based on simulated data given topology T7. In both examples, the support associated with the incorrectly resolved innermost branch is highlighted by a violet circled 1.

the branch length difference between adjacent long and short branches of a tree, the stronger the imbalance of MP informative site patterns in terms of strong branch support for adjacent and low support for non-adjacent long branch relationships.

Furthermore, different strengths of support for consistently correctly inferred relationships on ML basis of equally short branches cause an additional task. For example, continuously correct ML inferred internal branches that are ten times longer than the corresponding ancestral branch (0.1 versus 0.01) are moderately supported ($40 \geq$ sCF < 60), while always correctly ML inferred internal branch relationships next to terminal branches 100 times smaller than the internal branch of investigation (0.001 versus 0.1) gain continuously strong support (sCF ≥ 99). Thus, support differences of always correct ML reconstructed short branch relationships in respect of different branch length combinations raise the additional question of how and from when a sCF support value should be interpreted as strong or weak.

Based on the results of our study, the sCF of Minh *et al.* (18) as a measure of branch support is, at least, error-prone to different degrees of branch length heterogeneity, especially if the support associated branch is short. As long-branch attraction is one of the main topics when it comes to the necessity of branch reliability, the sCF is, at least on this problem, not a methodological alternative or reliable complement to non-parametric bootstrapping or Bayesian posterior probability estimations. Therefore we conclude not to rely on sCF measures as criterion to estimate the reliability of relationships affected by branch length heterogeneity.

In principle, it seems doubtful to estimate branch relationships inferred by a statistically robust and efficient tree method like ML or Bayesian tree inference with a less robust and statistically less understood MP approach (23,29,61–68).

An important scenario, exceptionally concerning a better statistical robustness of the MP metric in comparison to ML is given by Mendes & Hahn (69): The authors prove that for an asymmetric simulated four-taxon tree the parsimony metric of the concordance factor works well in the anomaly zone (AZ; (70)) of species trees if data are simulated in respect of incomplete-lineage sorting (ILS; (71–73)). However, the authors emphasise that only under the assumption of an infinite-sites mutation model and constant population sizes within and among species, parsimony-based methods should accurately identify the species tree in the AZ. MP correctly infers asymmetric species trees in the four-taxon case if (and only if) other phenomena capable of generating phylogenetic incongruence, such as introgression and homoplasy, are not present in the data. Since empirical real world phylogenetic data satisfies these conditions only very rarely, e.g. retrotransposon insertions as an example of homoplasy-free characters (69,74), the sCF can be only in exceptional cases another useful method for reconstructing asymmetric four-taxon species trees.

Although our simulations are based on a quite small number of sequences, our setup represents a recurrent range of short and long-branch relationships, which can be considered as either a subsection of branch relationships within a larger tree. Thus, we deem our results as quite representative in the sense of larger ("real world") data sets.

Since our study demonstrates that branch length heterogeneity has a negative impact on the sCF measure, we consider it as likely that other well known pitfalls of the MP approach, like base compositional heterogeneity or heterotachy (75,76), might also have a negative impact on the sCF measure. However, since both cases have not been tested here, further studies are necessary to achieve a sense of certainty.

Unfortunately, there is no perfect measure of branch reliability yet. As briefly mentioned in the introduction, non-parametric bootstrap and Bayesian methods have their own pitfalls.

Nevertheless, apart from the impact of inflated branch support due to systematic bias that affect both, ML and Bayesian, methods as well, we follow the statement of Douady *et al.* (15) that Bayesian inference and non-parametric bootstrapping can quite be used as potential upper and lower bounds of branch support to better explore the range of phylogenetic support estimates, especially when potential conflicts between data sets are explored.

SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

ACKNOWLEDGEMENTS

We thank the two anonymous reviewers for there inspiring comments.

FUNDING

This project was funded by the LIB.
Conflict of interest statement. None declared.

REFERENCES

1. Huelsenbeck,J., Rannala,B. and Masly,J.P. (2000) Accomodating phylogenetic uncertainty in evolutionary studies. *Science*, **288**, 2349–2350.
2. Lutzoni,F., Pagel,M. and Reeb,V. (2001) Major fungal lineages are derived from lichen symbiotic ancestors. *Nature*, **411**, 937–940.
3. Alfaro,M.E., Zoller,S. and Lutzoni,F. (2003) Bayes or bootstrap? A simulation study comparing the performance of bayesian markov chain monte carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol. Biol. Evol.*, **20**, 255–266.
4. Efron,B. (1979) Bootstrap methods: another look at the jackknife. *Ann. Stat.*, **7**, 1–26.
5. Felsenstein,J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.
6. Rannala,B. and Yang,Z. (1996) Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Mol. Evol.*, **43**, 304–311.
7. Larget,B. and Simon,D. (1999) Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.*, **16**, 750–759.
8. Mau,B., Newton,M.A. and Larget,B. (1999) Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics*, **55**, 1–12.
9. Nei,M., Stephens,J.C. and Saitou,N. (1985) Methods for computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes. *Mol. Biol. Evol.*, **2**, 66–85.
10. Li,W.H. (1989) A statistical test of phylogenies estimated from sequence data. *Mol. Biol. Evol.*, **6**, 424–435.
11. Rzhetsky,A. and Nei,M. (1992) A simple method for estimating and testing minimum-evolution trees. *Mol. Biol. Biol.*, **9**, 945–967.
12. Suzuki,Y., Glazko,G.V. and Nei,M. (2002) Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 16138–16143.
13. Huelsenbeck,J., Larget,B., Miller,R. and Ronquist,F. (2002) Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.*, **51**, 673–688.
14. Waddell,P.J., Kishino,H. and Ota,R. (2002) Very fast algorithms for evaluating the stability of ML and Bayesian phylogenetic trees from sequence data. *Genome Informatics*, **13**, 82–92.
15. Douady,C., Delsuc,F., Boucher,Y., Doolittle,W. and Douzery,E. (2003) Comparison of Bayesian and Maximum Likelihood bootstrap measures of phylogenetic reliability. *Mol. Biol. Evol.*, **20**, 248–254.
16. Erixon,P., Svennblad,B., Britton,T. and Oxelman,B. (2003) Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst. Biol.*, **52**, 665–673.
17. Huelsenbeck,J. and Rannala,B. (2004) Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst. Biol.*, **53**, 904–913.
18. Minh,P.Q., Hahn,M.W. and Lanfear,R. (2020) New methods to calculate concordance factors for phylogenomic datasets. *Mol. Biol. Evol.*, **37**, 2727–2733.

19. Nguyen,L.-T., Schmidt,H.A., Haeseler,A. and Minh,B.Q. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. *Mol. Biol. Evol.*, **32**, 268–274.
20. Felsenstein,J. (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Biol.*, **27**, 401–410.
21. Yang,Z. (1993) Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over time. *Mol. Biol. Evol.*, **10**, 1396–1401.
22. Hasegawa,M. and Yano,T. (1984) Maximum likelihood method of phylogenetic inference from DNA sequence data. *Bull. Biomet. Soc. Jpn.*, **5**, 1–7.
23. Huelsenbeck,J.P. and Hillis,D.M. (1993) Success of phylogenetic methods in the four-taxon case. *Syst. Zool.*, **42**, 247–264.
24. Tateno,Y., Takezaki,N. and Nei,M. (1994) Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. *Mol. Biol. Evol.*, **11**, 261–277.
25. Yang,Z., Goldman,N. and Friday,A.E. (1995) Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. *Syst. Biol.*, **44**, 384–399.
26. Kim,J.H. (1996) General inconsistency conditions for maximum parsimony: effects of branch lengths and increasing numbers of taxa. *Syst. Biol.*, **45**, 363–374.
27. Yang,Z. (1996) Among-site rate variation and its impact on phylogenetic analyses. *Tree*, **11**, 367–372.
28. Huelsenbeck,J.P. (1997) Is the Felsenstein zone a fly trap? *Syst. Biol.*, **46**, 69–74.
29. Swofford,D.L., Waddell,P.J., Huelsenbeck,J.P., Foster,P.G., Lewis,P.O. and Roger,J. (2001) Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Syst. Biol.*, **50**, 525–539.
30. Bergsten,J. (2005) A review of long-branch attraction. *Cladistics*, **21**, 163–193.
31. Yang,Z. and Rannala,B. (2012) Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.*, **13**, 303–314.
32. Hillis,D.M. and Huelsenbeck,J.P. (1993) Success of phylogenetic methods in the four-taxon case. *Syst. Biol.*, **42**, 247–264.
33. Huelsenbeck,J.P. (1995) Performance of phylogenetic methods in simulation. *Syst. Biol.*, **44**, 17–48.
34. Nei,M., Takezaki,N. and Sitnikova,T. (1995) Assessing molecular phylogenies. *Science*, **267**, 253–255.
35. Siddal,M.E. (1998) Success of parsimony in the four-taxon case: long branch repulsion by likelihood in the Farris zone. *Cladistics*, **14**, 209–220.
36. Jermiin,L.S., Ho,S. Y.W., Ababneh,F.A., Robinson,J. and Larkum,A. W.D. (2003) Hetero: a program to simulate the evolution of nucleotide sequences on a binary tree with four tips. *Appl. Bioinf.*, **2**, 159–163.
37. Hendy,M. and Penny,D. (1989) A framework for the quantitive study of evolutionary trees. *Syst. Zool.*, **38**, 297–309.
38. Zharkikh,A. and Li,W.H. (1993) Inconsistency of the maximum-parsimony method: the case of five taxa with a molecular clock. *Syst. Biol.*, **42**, 113–125.
39. Takezaki,N. and Nei,M. (1994) Inconsistency of the maximum parsimony method when the rate of nucleotide substitution is constant. *J. Mol. Evol.*, **39**, 210–218.
40. DeBry,R. (1992) The consistency of several phylogeny-inference methods under varying evolutionary rates. *Mol. Biol. Evol.*, **9**, 537–551.
41. Saccone,C. and Pesole,G. (1989) DNA microenvironments and the molecular clock. *J. Mol. Evol.*, **29**, 407–411.
42. Penny,D., Hendy,M.D., Zimmer,E.A. and Hamby,R. (1990) Trees from sequences: panacea or pandora's box? *Aust. Syst. Bot.*, **3**, 21–38.
43. Sidow,A. and Wilson,A.C. (1990) Compositional statistics: an improvement of evolutionary parsimony and its deep branches in the tree of life. *J. Mol. Evol.*, **31**, 51–68.
44. Lockhart,P.J., Penny,D., Hendy,M.D., Howe,C.J., Beanland,T.J. and Larkum,A.W.D. (1992) Controversy on chloroplast origins. *FEBS Lett.*, **301**, 127–131.
45. Lockhart,P.J., Howe,C.J., Bryant,D.A., Beanland,T.J. and Larkum,A. W.D. (1992) Substitutional bias confounds inference of cyanelle origins from sequence data. *J. Mol. Evol.*, **34**, 153–162.
46. Forterre,P., Benachenhou-Lafha,N. and Labedan,B. (1993) Universal tree of life. *Nature*, **362**, 795.
47. Hasegawa,M. and Hashimoto,T. (1993) Ribosomal RNA trees misleading? *Nature*, **361**, 23.
48. Steel,M.A., Lockhart,P.J. and Penny,D. (1993) Confidence in evolutionary trees from biological sequence data. *Nature*, **360**, 440–442.
49. Olsen,G.J. and Woese,C.R. (1993) Ribosomal RNA: a key to phylogeny. *FASEB J.*, **7**, 113–123.
50. Collins,T.M., Wimberger,P.H. and Naylor,G.J.P. (1994) Compositional bias, character-state bias, and character-state reconstruction using parsimony. *Syst. Biol.*, **43**, 482–496.
51. Foster,P.G. and Hickey,D.A. (1999) Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J. Mol. Evol.*, **48**, 284–290.
52. Tario,R., Rodriguez-Trelles,F. and Ayala,F.J. (2001) Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. *Mol. Biol. Evol.*, **18**, 1464–1473.
53. Jermiin,L.S., Ho,S.Y.W., Ababneh,F.A., Robinson,J. and Larkum,A.W.D. (2004) The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Syst. Biol.*, **53**, 638–643.
54. Saitou,N. and Imanishi,T. (1989) Relative efficiencies of the Fitch-Margoliash, maximum parsimony, maximum likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Mol. Biol. Evol.*, **6**, 514–525.
55. Jin,L. and Nei,M. (1990) Limitations of the evolutionary parsimony method of phylogenetic analysis. *Mol. Biol. Evol.*, **7**, 82–102.
56. Hasegawa,M. and Fujiwara,M. (1993) Relative efficiencies of the maximum likelihood, maximum parsimony, and neighbor-joining methods. *Mol. Phylogenet. Evol.*, **2**, 1–5.
57. Kuhner,M.K. and Felsenstein,J. (1994) A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.*, **11**, 459–468.
58. Takahashi,K. and Nei,M. (2000) Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Mol. Biol. Evol.*, **17**, 1251–1258.
59. Fletcher,W. and Yang,Z. (2009) INDELible: A flexible simulator of biological sequence evolution. *Mol. Biol. Evol.*, **26**, 1879–1888.
60. Sullivan,J. and 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.
61. Felsenstein,J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, **17**, 368–376.
62. Fukami-Kobayashi,K. and Tateno,Y. (1991) Robustness of maximum likelihood tree estimation against different patterns of base substitutions. *J. Mol. Evol.*, **32**, 79–91.
63. Gaut,S. and Lewis,P.O. (1995) Success of maximum likelihood phylogeny inference in the four-taxon case. *Mol. Biol. Evol.*, **12**, 152–162.
64. Gaucher,E.A. and Miyamoto,M.M. (2005) A call for likelihood phylogenetics even when the process of sequence evolution is heterogeneous. *Mol. Phylogenet. Evol.*, **37**, 928–931.
65. Sinsheimer,J.S., Lake,J.A. and Little,R.J.A. (1996) Bayesian hypothesis testing of four-taxon topologies using molecular sequence data. *Biometrics*, **52**, 193–210.
66. Simmons,M.P. and Masaki,M. (2004) Efficiently resolving the basal clades of a phylogenetic tree using Bayesian and parsimony approaches: a case study using mitogenomic data from 100 higher teleost fishes. *Mol. Phylogenet. Evol.*, **31**, 351–362.
67. Simmons,M.P., Zhang,L.B., Webb,C.T., Reeves,A. and Miller,J.A. (2006) The relative performance of Bayesian and parsimony approaches when sampling characters evolving under homogeneous and heterogeneous sets of parameters. *Cladistics*, **22**, 171–185.
68. Wolsan,M. and Sato,J.J. (2010) Effects of data incompleteness on the relative performance of parsimony and Bayesian approaches in a supermatrix phylogenetic reconstruction of Mustelidae and Procyonidae (Carnivora). *Cladistics*, **26**, 168–194.
69. Mendes,F.K. and Hahn,M.W. (2018) Why concatenation fails near the anomaly zone. *Syst. Biol.*, **67**, 158–169.
70. Degnan,J. and Rosenberg,N.A. (2006) Discordance of species trees with their most likely gene trees. *PLoS Genet.*, **2**, e68.

71. Hudson,R.R. (1983) Testing the constant-rate neutral allele model with protein sequence data. *Evolution*, **37**, 203–217.
72. Tajima,F. (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.
73. Pamilo,P. and Nei,M. (1988) Relationships between gene trees and species trees. *Mol. Biol. Evol.*, **5**, 568–583.
74. Suh,A. and Smeds,I. H.E. (2015) The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biol.*, **13**, e1002224.
75. Philippe,H., Zhou,Y., Brinkmann,H., Rodrigue,N. and Delsuc,F. (2005) Heterotachy and long-branch attraction in phylogenetics. *BMC Evol. Biol.*, **5**, 50.
76. Lockhart,P.J., Novis,P., Milligan,B.G., Riden,J., Rambaut,A. and Larkum,T. (2005) Heterotachy and tree building: a case study with plastids and Eubacteria. *Mol. Biol. Evol.*, **23**, 40–45.