bioRxiv preprint doi: https://doi.org/10.1101/035360; this version posted December 25, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

# Do Phylogenetic Tree Viewers correctly display Support Values?

Lucas Czech* and Alexandros Stamatakis*†
*Heidelberg Institute for Theoretical Studies, Germany
Email: {alexandros.stamatakis, lucas.czech}@h-its.org
†Karlsruhe Institute of Technology, Institute for Theoretical Informatics, Germany

***Abstract*—Phylogenetic trees are routinely visualized to present and interpret the evolutionary relationships of the species that are being studied. Virtually *all* empirical evolutionary data studies contain a visualization of the inferred tree with support values using one of the popular and highly cited (e.g., TreeView, Dendroscope, FigTree, Archaeopteryx, etc.) tree viewing tools. As a consequence, programming errors or ambiguous semantics in tree file formats can lead to erroneous tree visualizations and consequently incorrect interpretations of phylogenetic analyses.**

**Here, we discuss the problems that can and *do* arise when displaying branch support values on trees. Presumably for historical reasons, branch support values (e.g., bootstrap support or Bayesian posterior probabilities) are typically stored as node labels in the widely-used Newick tree format. However, support values are attributes of branches (bipartitions) in unrooted phylogenetic trees. Therefore, storing support values as node labels can potentially lead to incorrect support-value-to-bipartition mappings when re-rooting trees in tree viewers. This depends on the mostly implicit semantics of tree viewers for interpreting node labels. To assess the potential impact of these ambiguous and predominantly implicit semantics of support values, we analyzed 10 distinct tree viewers. We find that, most of them exhibit some sort of incorrect or unexpected behavior when re-rooting trees with support values. We find that Dendroscope interprets Newick node labels as simply that, node labels in Newick trees. However, if they are meant to represent branch support values, the support value to branch mapping is incorrect when re-rooting trees with Dendroscope. We illustrate such an incorrect mapping by example of an empirical phylogenetic study.**

**As a solution, we suggest that (i) branch support values should exclusively be stored as meta-data associated to branches (and not nodes), and (ii) if this is not feasible, tree viewers should include a user dialogue that explicitly forces users to define if node labels shall be interpreted as node or branch labels, prior to tree visualization.**

(Keywords: phylogenetic trees, tree visualization, tree viewer, Newick format, branch support values, branch labels, software, bugs )

## I. Introduction

The Newick format is widely used to store and visualize phylogenies. Archie *et al.* introduced it in 1986 [1]. Since then, it has become the de-facto standard for storing, exchanging, and displaying phylogenies. It uses parenthesis and commas to specify the nesting structure of the tree and also allows for storing node labels and branch lengths.

In many cases, however, additional vital information needs to be associated with the branches of a tree. Published phylogenies usually display branch support values, such as bootstrap [2], Bayesian posterior probability [3], or aLRT test [4] values. These values are associated with branches (splits/bipartitions) of the tree and *not* nodes of the tree. In the original specification of the Newick format, the authors had not foreseen an option for specifying such branch labels.

Thus, as a work-around, branch support values are often stored as inner node labels in the output of phylogenetic inference codes. Node labels of tip nodes usually contain the species name of the extant organisms for which the phylogeny was built. Inner nodes, however, represent hypothetical common ancestors and do therefore generally not have a species name. Thus, these inner node labels can be used to store branch support information.

This convention or work-around of storing branch support values as node labels exhibits potential pitfalls. This is because, in an unrooted binary tree, it is not clear to which of the three outgoing branches of an inner node such a node label refers to.

However, for rooted trees, there is an unambiguous mapping of node labels to branches: The node label (support value) at an inner node can always be associated (mapped to) with the outgoing branch that points toward the root. Note that, unrooted trees often have a dedicated inner node that serves as a 'hook' for both, computing, and visualizing the tree. This so-called top-level trifurcation is not a root in the strict sense, but required for storing and parsing the tree because we need to start to recursively traverse the tree somewhere. We can chose the inner node that serves as top-level trifurcation arbitrarily, that is, the same underlying unrooted tree can be displayed or written to file in a plethora of distinct ways. For $n$ taxa, an unrooted binary tree has $n-2$ inner nodes, hence we can chose $n-2$ top-level trifurcations. For each of these possible top-level trifurcations, we can then also freely chose by which order we descend into the subtrees defined by the three outgoing branches to print out or visualize the tree. The chosen top-level trifurcation induces an artificial direction for branches in the tree, and can thus be used to unambiguously associate node labels with branches. Figure 1 shows an unrooted tree with this structure.

Thus, both rooted and unrooted trees in Newick format

bioRxiv preprint doi: https://doi.org/10.1101/035360; this version posted December 25, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

explicitly (root) or implicitly (top-level trifurcation) encode a direction for branches. Therefore, the mapping between branch support values and node labels in Newick files is well defined in principle: For restoring the correct association between node labels and branches, the direction towards the top-level node (root or top-level trifurcation) can be used. However, this entails an implicit semantic interpretation. When reading a Newick-formatted tree, the user or program needs to know if inner node labels need to be interpreted as branch labels. When this semantic distinction is not made, node labels need to be interpreted as being associated to the nodes. When node labels that should be interpreted as branch labels (e.g., support values) are erroneously interpreted as node labels, this can lead to incorrect visualizations as well as interpretations of phylogenies. While we focus on tree viewers here, the issues we discuss might also affect downstream analysis tools (e.g., tools for computing the weighted Robinson-Foulds distance [5] between phylogenies with support values) that parse phylogenies with node labels.

We show that, most common tree viewers do not offer an explicit option for specifying the semantics of inner node labels. A simple way to examine the behavior of viewers in this regard, is to (re-)root a given tree – a procedure that all of the viewers we tested offer. If node labels shall represent branch labels, a viewer behaves correctly (maintains the correct node label to branch mapping), if the association of some node labels with corresponding branches is changed during the re-rooting process. This is because the direction towards the root (or top-level trifurcation) changes.

Our unrooted bifurcating Newick test tree with inner node labels

$$T_N = \texttt{((C,D)1,(A,(B,X)3)2,E)R;}$$

has six leaf nodes (`A...E`), four inner nodes (labeled `1...3`, and the top-level trifurcation `R`). For the sake of simplicity, we do not use branch lengths. We use $T_N$ throughout this manuscript to test the behavior of tree viewers and to outline the potential problems that can arise due to the mostly implicit semantics of Newick node labels.

Note that, some programs can output support values as Newick comments in square brackets instead of using node labels. The tree

$$T_C = \texttt{((C,D)[1],(A,(B,X)[3])[2],E)[R];}$$

shows an example for this notation that contains the same information as tree $T_N$. For the semantics and the association of those comments with branches, the same rules apply as for the node label notation. Some of the tested tree viewers are able to correctly parse and display this format, but in general, the same semantic issues and mapping problems arise. Hence, we conduct our tests with $T_N$ instead of $T_C$.

In Figure 1 we show tree $T_N$ with nodes and branches that are colored by the correct node label to branch mapping.

If we now (re-)root $T_N$ at the branch that leads to tip `X`,

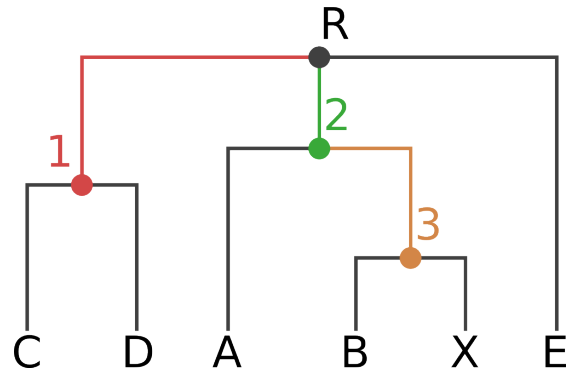



Figure 1: Original rooting (via top-level trifurcation) and visual representation of our Newick test tree $T_N$. Inner nodes and branches are colored according to the correct node label to branch mapping of $T_N$.

the mappings between all nodes and branches that lie on the path between the old and the new top-level node have to be altered. In our example, the nodes on the path between `R` and `X` are the inner nodes `2` and `3`. In Figure 2 we display the correct (2b) and incorrect (2a) way of mapping node labels (interpreted as branch support values) to branches after re-rooting. Note that, this rooted binary tree now contains one more node, which is the newly created root node `R'`. In both Figures, the node labels are correctly assigned to their corresponding nodes. However, the association of those labels to the corresponding branches is only correct in Figure 2b.

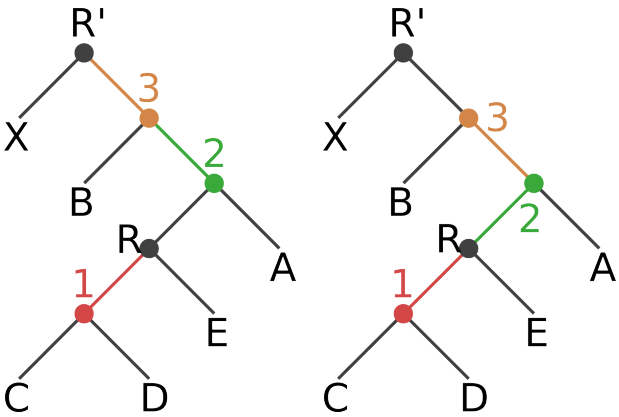
Evidently, an incorrect mapping of node labels to branches as presented in Figure 2a can yield incorrectly displayed branch support values in empirical phylogenetic studies. In addition, since a typically large fraction of the results and discussion sections of such papers is dedicated to interpreting the support values of the phylogeny, the conclusions of these studies might also be incorrect.

In the following, we examine different popular tree viewers to determine if they maintain the correct branch support mapping when we root our test tree $T_N$ at the branch leading to tip node `X`. Since Dendroscope [6] *does* yield an incorrect mapping, we also try to assess if there exist published empirical phylogenetic studies that contain incorrectly visualized trees with Dendroscope.

## II. Experimental Setup: Systematic Tests of Tree Viewers

While the original Newick format is well-defined, there is no official standard for it, including respective semantics. Hence, there is also no 'correct' way of using it – attributes of branches and nodes can be freely interpreted. Thus, users need to be aware of the semantics of such attributes. Their interpretation depends on the convention used when storing those values in Newick format. Yet, most viewers we

bioRxiv preprint doi: https://doi.org/10.1101/035360; this version posted December 25, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.




(a) Node labels are mapped incorrectly to branches, resulting in a tree with an erroneous node label to support value mapping.

(b) Node labels are correctly mapped to the branches of the tree.

Figure 2: Tree $T_N$ (and $T_C$) rooted on the branch leading to the tip node `X`. It contains an additional root node `R'`.

examined do not offer an option for explicitly defining these semantics. Users must therefore be aware and simply accept the implicit interpretation a particular viewer implements.

Given a Newick tree with inner node labels (e.g., tree $T_N$ with labels `1`, `2` and `3`), we distinguish between two possible interpretations for those labels. Those are:

1) The inner node labels are actual labels (e.g., ancestral species names). We call this the "node interpretation".
2) They represent branch labels (e.g., support values). We call this the "branch interpretation".

As already mentioned, the same applies to trees that use comments instead of node labels (e.g., tree $T_C$). In order for a program to support both interpretations, a reasonable solution would be to offer an option for choosing between the two, that is, to include an explicit semantic interpretation option.

We tested the tree viewers as follows:

- Load trees $T_N$ and $T_C$ from the corresponding Newick file.
- Check how the viewer interprets the values and if there is an option to specify their semantics.
- Re-root the tree at the branch leading to node `X`.
- Check whether the viewer works correctly based on its interpretation.

In Table I, we provide an overview of the viewers we tested.

There are of course many more phylogenetic tree viewers available. We excluded those that do not support re-rooting, since our test procedure can not be applied to them. We also chose some highly cited viewers, since the impact of potential errors in these on published phylogenies with support values is more pronounced.

Table I: Evaluated tree viewers with accumulated number of citations (https://scholar.google.com/, accessed on 2015-12-22).

| Viewer | Version | Citation | Citations |
|---|---|---|---|
| Archaeopteryx | 0.9901 beta (141014) | [7] | 220 |
| ATV | 4.00 alpha 13 | [8] | 270 |
| Dendroscope | 3.2.10 (built 19 Nov 2013) | [6] | 1,120 |
| ETE2 | 2.3.10 | [9] | 147 |
| EvolView | Accessed 2015-12-22 | [10] | 62 |
| FigTree | 1.4.2 | [11] | 79 |
| iTOL | N/A | [12] | 589 |
| PhyloWidget | Accessed 2015-12-22 | [13] | 96 |
| TreeView | 1.6.6 (Windows) | [14] | 10,113 |
| T-REX | N/A | [15] | 75 |
| Total | | | 12,771 |

## III. Results

In addition to testing the tree viewers, we inspected the output format for phylogenies with branch support values that is generated by three widely-used phylogenetic inference tools. PHYML [16] reports support values as node labels (see http://www.atgc-montpellier.fr/phyml/alrt/http://www.atgc-montpellier.fr/phyml/alrt/). RAxML [17] generates two tree files, one with comments and one with node labels. Finally, MrBayes [18] uses its own Nexus-based format, which internally uses a variation of Newick comments to report support values (posterior probabilities).

Those different output formats illustrate the difficulties associated to visualizing trees with branch support values. In the following, we summarize our test results for visualizing our test tree with the aforementioned tree viewers. In Table II we provide an overview of these results.

Table II: Evaluation of tree viewers. The columns "Nodes" and "Branches" indicate which of the two interpretations the viewer supports. The last column shows whether the behavior is correct according to the interpretation offered or implied by the viewer.

| Viewer | Nodes | Branches | Correct |
|---|---|---|---|
| Archaeopteryx | ✓ | ✓ | (✓) |
| ATV | | ✓ | ✓ |
| Dendroscope | ✓ | (✓) | ✓ |
| ETE2 | ✓ | ✓ | ✓ |
| EvolView | | ✓ | |
| FigTree | (✓) | ✓ | ✓ |
| iTOL | | ✓ | ✓ |
| PhyloWidget | ✓ | | |
| TreeView | | ✓ | ✓ |
| T-REX | | ✓ | (✓) |

Overall, most viewers face some issues related to re-rooting or displaying trees with support values. In the following we discuss our observations for each viewer.

*Archaeopteryx:* is one of the two tested viewers that is aware of the semantic issue (see https://sites.google.com/site/cmzmasek/

bioRxiv preprint doi: https://doi.org/10.1101/035360; this version posted December 25, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

home/software/archaeopteryx/documentation#TOC-Internal-Node-Names-are-Confidence-Vales). It offers an option to define the semantics of annotated values. The default option is to interpret nodes labels as node labels, thus the re-rooted tree is correctly displayed only for the node interpretation. When enabled (i.e., activating the node label as branch label interpretation), re-rooting correctly shifts the support values to the corresponding branches. However, it does not immediately display these values, which appears to be a bug. They only become visible when moving the mouse over the respective branch. Thus, in the overview table, we marked it as "almost" correct.

*ATV:* is the predecessor to Archaeopteryx. Different versions seem to alternate between the two possible interpretations of inner node labels. The one we tested uses the node labels as branch support values interpretation and thus correctly re-roots.

*Dendroscope:* only offers the node labels as node labels interpretation for our test trees. This leads to incorrect results when re-rooting trees with node labels that represent branch support values. If the tree also contains branch lengths, Dendroscope interprets the Newick comments as support values (e.g., tree $T_C$ plus branch lengths). The alternative notation using inner node labels (e.g., tree $T_N$) is not affected by this and always applies the interpretation as node labels. This behavior is not fully documented in the manual. We assess the impact of this behavior on published empirical phylogenetic studies in Section IV.

*ETE2:* is the other viewer that supports both interpretations. When reading a Newick formatted tree, it offers an option to choose the semantics of labels. The comment notation however is not supported.

*EvolView:* is able to display numerical values at inner nodes. Re-rooting however misplaces those values to wrong nodes and sets some of them to zero. Re-rooting a given tree several times at different branches results in all inner node values becoming zero. Futhermore, re-rooting does not resolve the initial trifurcation properly, so that the resulting tree contains a multifurcation at node $\mathrm{R}$.

*FigTree:* is able to display multiple inner node values using both semantic interpretations. Of the tested viewers, it offers the best support for general tree annotation. When re-rooting the tree, however, there is no option to define the interpretation of node labels, since FigTree always assumes the branch interpretation. In addition, it can not parse certain Newick variants, such as trees that contain both, branch lengths, and support values stored as comments.

*iTOL:* works correctly, implicitly applying the node labels as branch support values interpretation.

*PhyloWidget:* interprets node labels as node labels. Thus, re-rooting a tree with branch support values yields errors. Also, re-rooting does not resolve the initial trifurcation, similar to EvolView. For this reason, it is marked as not correcet in Table II.

*TreeView:* correctly interprets node labels as branch support values. However, it displays the values next to the nodes instead of the branches, which may lead to potential confusion.

*T-REX:* also applies the branch interpretation and correctly re-roots. The values are however always displayed as percentages, which is not always the correct or desired way for displaying branch support values. Hence, we marked it as almost correct in the overview table. It also does not work with the comment notation.

In general, most tree viewers interpret node labels as branch support values and can correctly re-root the trees maintaining the correct branch support value mapping. Of the highly cited viewers, only Dendroscope implicitly interprets node labels as, simply that, node labels. Users who are not aware of this implicit semantic assumption might thus obtain tree visualizations with incorrectly mapped support values, because they expect Dendroscope to behave like the other viewers. We explore the potential impact of this implicit semantic assumption on empirical phylogenetic studies in the next section.

## IV. Impact on Empirical Phylogenetic Studies

The extent to which the described semantic issue in Dendroscope (see Section III) affects published phylogenies is hard to quantify. This is because, all phylogeny figures in all published papers citing Dendroscope (over $1000$ for the two Dendroscope papers based on Google scholar, accessed on 2015-12-23) would need to be checked and all original tree files would need to be available, which they should be, in principle. Hence, this is also an issue of the reproducibility of scientific results, even if in our case it simply boils down to making available a published Newick tree with support values for download. To at least get a feeling of the visualization and reproducibility issue, we contacted the authors of $14$ papers citing Dendroscope, published in journals such as *Nature, PLOS, BMC,* and *JBC*. Out of the contacted authors, $5$ replied, but only two were finally able to provide us with the trees that were used to generate the visualizations in their publications.

In the following, we analyze the trees visualized for these two papers with respect to the correctness of the branch support value mapping.

The first article [19] presents a phylogeny of 80 Arabidopsis accessions (see Fig. 4(b) of [19]) along with bootstrap values for some of the branches. The tree and bootstrap values were inferred with RAxML 7.3.5 [20], which writes a tree file that uses Newick comments for storing support values. Dendroscope [6] was used to re-root and visualize the tree. As already mentioned the tool *is* able to correctly handle this variant of storing support values. Thus, the error did not occur in this paper and the tree is correctly visualized.

bioRxiv preprint doi: https://doi.org/10.1101/035360; this version posted December 25, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

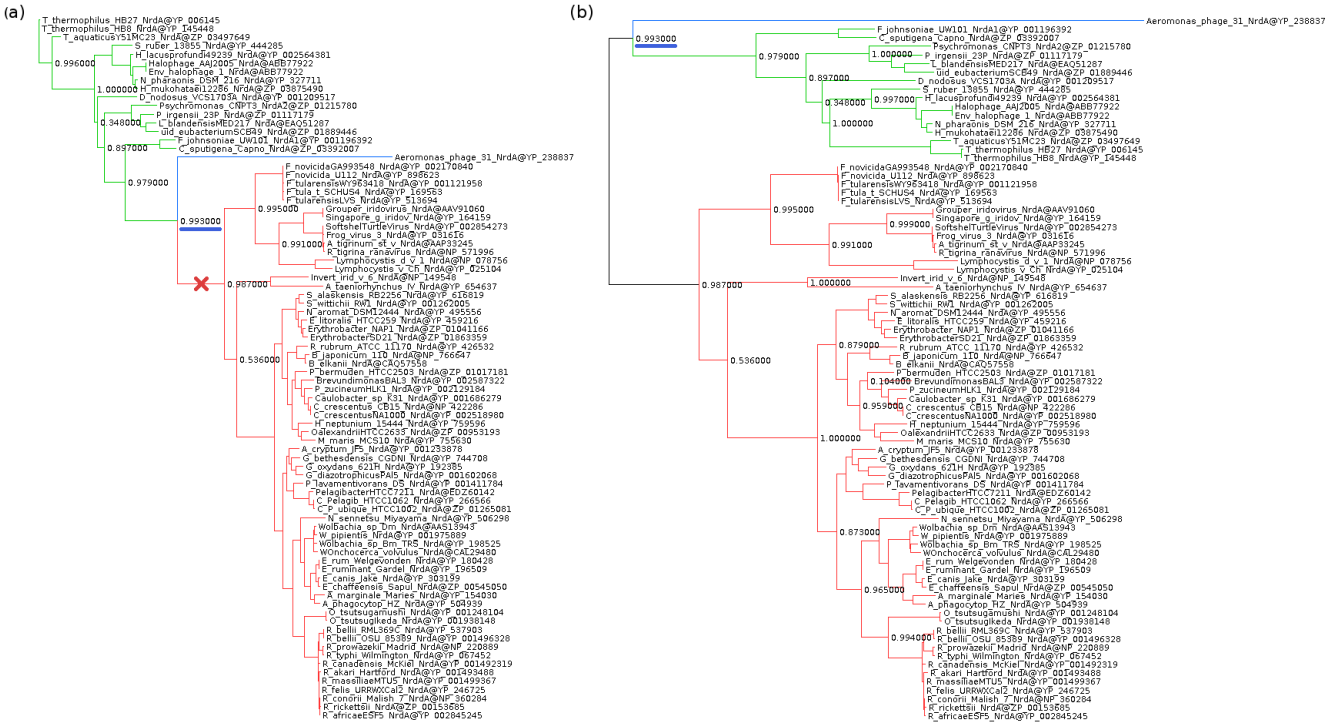



Figure 3: We used the original data from [21] to recreate Figure 2(a) of [21]. (a) The original tree with the branch used for re-rooting marked by a red cross. (b) The re-rooted tree with incorrectly placed branch support values (e.g., the underlined one). We colored the subtrees to highlight their positions after re-rooting.

The second article [21] presents several phylogenies for all three domains of life. The trees were inferred using RAxML v7.2.6 [20], [22], [23] and PHYML v3.0 [16], [24], [25]. Branch support values were estimated with PHYML using the SH-like likelihood ratio test [4], which reports support values as node labels. All trees in Figures 2 and 4-7 of [21] were re-rooted using Dendroscope such that they can be more easily compared to the comprehensive trees presented in Fig. 1 of the article. In all cases, branch support values were mapped incorrectly to the re-rooted trees in these Figures.

We illustrate this in Figure 3. Sub-figure (a) is the original Newick tree used to generate Figure 2(a) in [21]. We have marked the branch used for (re-)rooting the tree by a red cross. We colored the subtrees so that their corresponding position in the re-rooted tree is easily visible. Sub-figure (b) shows the re-rooted tree, which is identical to the one presented in [21]. The branch support values between the old and the new root node in our Figure 3 are not mapped to the same bipartition in sub-figure (a) and (b). For example, in sub-figure (a) the underlined support value refers to the bipartition `green taxa | blue taxon, red taxa` whereas in sub-figure (b) it refers to the bipartition `red taxa | green taxa, blue taxon`.

## V. Discussion & Conclusion

Our results indicate that an explicit convention and explicit semantics for interpreting node and branch labels in tree viewers is clearly missing. Both ways – interpretation as node labels and as branch values – are fairly common (8 and 4 of the tested viewers for the two interpretations, respectively). From the viewers we tested, only two (Archaeopteryx and ETE2) offer an explicit user option to define the semantics of node values. Dendroscope offers an implicit choice depending on the input format. Others can not read certain types of Newick variants. In summary, every tested tree viewer treats nodes labels and branch support values in its own, mostly undocumented, way.

Furthermore, programs that can infer branch support values use a plethora of distinct output formats. As alternative, developers of phylogenetic inference codes should consider using the PhyloXML format to store trees. PhyloXML explicitly supports storing information that is associated with branches. PhyloXML trees are, however, more difficult to parse and yield substantially larger tree files. For instance, our test tree $T_N$ requires 24 bytes in Newick, but 856 bytes in PhyloXML format. A 512 taxon tree with branch lengths requires $40,303$ bytes in Newick and $239,795$ bytes in PhyloXML.

To address the general problem, we suggest that all tree

bioRxiv preprint doi: https://doi.org/10.1101/035360; this version posted December 25, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

viewers shall offer an explicit option to choose between the two possible interpretations of node labels. Ideally, users should be forced to define the semantics of their node labels before the tree is displayed by the respective tool.

Finally, all published phylogenies with support values that used Dendroscope for re-rooting and tree visualization need to be re-assessed if support values were stored as node labels in the original Newick files.

We conclude with some practical suggestions for users of phylogenetic tree viewing tools.

- Pay attention to the options a viewer offers for interpreting node labels in Newick files.
- If available, use the option to set the desired interpretation (e.g., Archaeopteryx, ETE2).
- Double-check your results, maybe try other tools, or conduct a visual inspection, particularly if you re-rooted the tree.
- The behavior of viewers can easily be tested with our example trees $T_N$ and $T_C$ that are available for download at https://github.com/stamatak/tree-viz-issues.

## Acknowledgment

This work was financially supported by the Klaus Tschira Foundation.

We wish to thank the authors of the papers described in Section IV for sending us their original tree files:

- Anthony Poole and Daniel Lundin [21].
- Artem Pankin and Franziska Turck [19].

## References

[1] J. Archie, W. H. Day, W. Maddison, C. Meacham, F. J. Rohlf, D. Swofford, and J. Felsenstein, "The newick tree format," Online, accessed 2015-07-26., June 1986, http://evolution.genetics.washington.edu/phylip/newicktree.html.

[2] J. Felsenstein, "Confidence Limits on Phylogenies: An Approach Using the Bootstrap," *Evolution*, vol. 39, no. 4, pp. 783–791, Jul. 1985. [Online]. Available: http://www.jstor.org/stable/2408678

[3] J. P. Huelsenbeck, F. Ronquist, R. Nielsen, and J. P. Bollback, "Bayesian inference of phylogeny and its impact on evolutionary biology." *Science (New York, N.Y.)*, vol. 294, no. 5550, pp. 2310–4, Dec. 2001. [Online]. Available: http://www.sciencemag.org/content/294/5550/2310.abstracthttp://www.sciencemag.org/content/294/5550/2310.short

[4] M. Anisimova and O. Gascuel, "Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative." *Systematic biology*, vol. 55, no. 4, pp. 539–52, Aug. 2006. [Online]. Available: http://sysbio.oxfordjournals.org/content/55/4/539.short

[5] D. Robinson and L. R. Foulds, "Comparison of phylogenetic trees," *Mathematical Biosciences*, vol. 53, no. 1, pp. 131–147, 1981.

[6] D. H. Huson and C. Scornavacca, "Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks," *Systematic biology*, p. sys062, 2012.

[7] M. V. Han and C. M. Zmasek, "phyloXML: XML for evolutionary biology and comparative genomics." *BMC bioinformatics*, vol. 10, p. 356, Jan. 2009. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2774328\&tool=pmcentrez\&rendertype=abstract

[8] C. M. Zmasek and S. R. Eddy, "ATV: display and manipulation of annotated phylogenetic trees." *Bioinformatics (Oxford, England)*, vol. 17, no. 4, pp. 383–384, Apr. 2001. [Online]. Available: http://bioinformatics.oxfordjournals.org/content/17/4/383.short

[9] J. Huerta-Cepas, J. Dopazo, and T. Gabaldón, "ETE: a python Environment for Tree Exploration." *BMC bioinformatics*, vol. 11, no. 1, p. 24, Jan. 2010. [Online]. Available: http://www.biomedcentral.com/1471-2105/11/24

[10] H. Zhang, S. Gao, M. J. Lercher, S. Hu, and W.-H. H. Chen, "EvolView, an online tool for visualizing, annotating and managing phylogenetic trees," *Nucleic Acids Research*, vol. 40, no. W1, pp. 569–572, Jul. 2012. [Online]. Available: http://nar.oxfordjournals.org/content/40/W1/W569.short

[11] A. Rambaut, "FigTree, a graphical viewer of phylogenetic trees." Online, accessed 2015-07-27., 2007, http://tree.bio.ed.ac.uk/software/figtree/.

[12] I. Letunic and P. Bork, "Interactive Tree of Life v2: Online annotation and display of phylogenetic trees made easy," *Nucleic Acids Research*, vol. 39, no. SUPPL. 2, pp. 1–4, Jul. 2011. [Online]. Available: http://nar.oxfordjournals.org/content/early/2011/04/04/nar.gkr201.abstract

[13] G. E. Jordan and W. H. Piel, "PhyloWidget: web-based visualizations for the tree of life." *Bioinformatics (Oxford, England)*, vol. 24, no. 14, pp. 1641–2, Jul. 2008. [Online]. Available: http://bioinformatics.oxfordjournals.org/content/24/14/1641.short

[14] Roderic D. M. Page, "TreeView: an application to display phylogenetic trees on personal computers." *Computer applications in the biosciences : CABIOS*, vol. 12, no. 4, pp. 357–358, 1996.

[15] B. Alix, D. A. Boubacar, M. Vladimir, A. Boc, A. B. Diallo, and V. Makarenkov, "T-REX: A web server for inferring, validating and visualizing phylogenetic trees and networks," *Nucleic Acids Research*, vol. 40, no. W1, pp. 573–579, Jul. 2012. [Online]. Available: http://nar.oxfordjournals.org/content/40/W1/W573.short

[16] S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel, "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of phyml 3.0," *Systematic biology*, vol. 59, no. 3, pp. 307–321, 2010.

[17] A. Stamatakis, "Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies," *Bioinformatics*, vol. 30, no. 9, pp. 1312–1313, 2014.

bioRxiv preprint doi: https://doi.org/10.1101/035360; this version posted December 25, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[18] F. Ronquist, M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck, "Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space," *Systematic biology*, vol. 61, no. 3, pp. 539–542, 2012.

[19] L. Liu, J. Adrian, A. Pankin, J. Hu, X. Dong, M. von Korff, and F. Turck, "Induced and natural variation of promoter length modulates the photoperiodic response of flowering locus t," *Nature communications*, vol. 5, 2014.

[20] A. Stamatakis, P. Hoover, and J. Rougemont, "A rapid bootstrap algorithm for the raxml web servers," *Systematic biology*, vol. 57, no. 5, pp. 758–771, 2008.

[21] D. Lundin, S. Gribaldo, E. Torrents, B.-M. Sjoberg, and A. Poole, "Ribonucleotide reduction - horizontal transfer of a required function spans all three domains," *BMC Evolutionary Biology*, vol. 10, no. 1, p. 383, 2010. [Online]. Available: http://www.biomedcentral.com/1471-2148/10/383

[22] A. Stamatakis, "Raxml-vi-hpc: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models," *Bioinformatics*, vol. 22, no. 21, pp. 2688–2690, 2006.

[23] ——, "Phylogenetic models of rate heterogeneity: a high performance computing perspective," in *Parallel and Distributed Processing Symposium, 2006. IPDPS 2006. 20th International*. IEEE, 2006, pp. 8–pp.

[24] S. Guindon and O. Gascuel, "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood," *Systematic biology*, vol. 52, no. 5, pp. 696–704, 2003.

[25] S. Q. Le, N. Lartillot, and O. Gascuel, "Phylogenetic mixture models for proteins," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 363, no. 1512, pp. 3965–3976, 2008.