



RESOURCE ARTICLE

OneTwoTree: An online tool for phylogeny reconstruction

Michal Drori* | Anna Rice* | Moshe Einhorn | Ofer Chay | Lior Glick | Itay Mayrose

School of Plant Sciences and Food Security,
Tel Aviv University, Tel Aviv, Israel**Correspondence**Itay Mayrose, School of Plant Sciences and
Food Security, George S. Wise Faculty of
Life Sciences, Tel Aviv University, Tel Aviv
69978, Israel.
Email: itaymay@post.tau.ac.il**Funding information**United States-Israel Binational Science
Foundation, Grant/Award Number:
2016550; Israel Science Foundation, Grant/
Award Number: 961/17; Edmond J. Safra
Center for Bioinformatics**Abstract**

Phylogeny reconstruction is a key instrument in numerous biological analyses, ranging from evolutionary and ecology research, to conservation and systems biology. The increasing accumulation of genomic data makes it possible to reconstruct phylogenies with both high accuracy and at increasingly finer resolution. Yet, taking advantage of the enormous amount of sequence data available requires the use of computational tools for efficient data retrieval and processing, or else the process could quickly become an error-prone endeavour. Here, we present OneTwoTree (<http://onetwotree.tau.ac.il/>), a Web-based tool for tree reconstruction based on the supermatrix paradigm. Given a list of taxa names of interest as the sole input requirement, OneTwoTree retrieves all available sequence data from NCBI GenBank, clusters these into orthology groups, identifies the most informative set of markers, searches for an appropriate outgroup, and assembles a partitioned sequence matrix that is then used for the final phylogeny reconstruction step. OneTwoTree further allows users to control various steps of the process, such as the merging of sequences from similar clusters, or phylogeny reconstruction based on markers from a specific genome type. By comparing the performance of OneTwoTree to a manually reconstructed phylogeny of the Antirrhineae tribe, we show that the use of OneTwoTree resulted in substantially higher data coverage in terms of both taxon sampling and the number of informative markers assembled. OneTwoTree provides a flexible online tool for species-tree reconstruction, aimed to assist researchers ranging in their level of prior expertise in the task of phylogeny reconstruction.

KEYWORDS

phylogeny reconstruction, sequence clustering, supermatrix, systematics

1 | INTRODUCTION

Inferring the phylogeny of a group of organisms forms the basis of evolutionary, systematic and ecological studies by allowing the identification of key historical events, the comparison of evolving traits and the design of conservation strategies. Phylogenies further enable researchers from a wide range of disciplines to perform essential inference tasks, such as the detection of selective constraints acting on genes and proteins, functional sites prediction, identifying gene families and their dynamics, or for calculating various phylodiversity

metrics in community- and macro-ecology research (Felsenstein, 2004; Tucker et al., 2017).

Starting from the early years of molecular evolutionary research, a large number of tree reconstruction methods have been devised, ranging in computational complexity and statistical robustness (reviewed in Yang & Rannala, 2012). In general, different tree reconstruction strategies follow similar guidelines, which can be roughly divided into several distinct steps. Given a group of taxa of interest, the first step usually involves choosing an appropriate set of genomic markers and obtaining their sequences. These sequences are then aligned into a partitioned multiple sequence alignment (MSA), and based on the resulting MSA, the phylogeny is inferred. This

*These authors contributed equally to this work.

basic strategy involves multiple choices, including selecting the exact set of sequence markers and taxa to be analysed and choosing from among a large number of alternative methodologies and optional program parameters.

The constant rise in sequence data availability, along with the constant increase in the number of optional computational tools that can be used, makes the process of phylogeny reconstruction an increasingly daunting task, especially if performed over multiple clades that span dozens of species. This laid the foundation for the development of automated tools that aid researchers in various aspects involved in phylogeny inference (e.g., Agalma: Dunn, Howison, & Zapata, 2013; Hal: Robbertse, Yoder, Boyd, Reeves, & Spatafora, 2011; Phylogenie: Frickey & Lupas, 2004; phyloGenerator: Pearse & Purvis, 2013; SUMAC: Freyman, 2015). For example, *Phylogeny.fr* (Dereeper et al., 2008) is an online tree reconstruction tool that—given a provided set of sequences—allows the execution of distinct parts of the inference process, while offering a number of alternative programs that can be used for either the alignment or phylogeny reconstruction steps. Another useful tool, *PHLAWD* (Smith, Beaulieu, & Donoghue, 2009), was developed in the context of megaphylogeny reconstruction based on the supermatrix approach (de Queiroz & Gatesy, 2007). Once the program is installed on a local computer with Linux/MacOS operating systems, the user should provide the name of a clade of interest together with a representative sequence for each locus to be included in the analysis. The final output of this tool is an MSA for the requested marker. A supermatrix can then be assembled manually by the user, joining all loci of interest. *PhyLoTa* (Sanderson, Boss, Chen, Cranston, & Wehe, 2008) is designed to provide users with precomputed sets of clusters that were assembled through a single-linkage clustering approach and additionally provides precomputed gene trees that were reconstructed for each cluster. In particular, the results obtained by *PhyLoTa* are taxonomically constrained; that is, all sequences of the most recent common ancestor are collected even if one specifies only part of a clade. *SUPERSMART* (Antonelli et al., 2017) offers a customized framework for the inference of dated phylogenies, given a set of taxa of interest and a list of calibration points. Supporting both a virtual machine Linux-based platform and a Web server with simpler functionality, *SUPERSMART* relies on the data assembled by *PhyLoTa* for the clustering process. It then implements a novel technique for reconstructing species trees based on a three-step approach that combines the use of a sparse backbone tree with densely sampled phylogenies, which are created for smaller subclades using the supermatrix approach. Noteworthy, a simple-to-use online tool for either dated or undated species-tree reconstruction, given an unconstrained list of taxa of interest, while making use of all currently available GenBank sequence data, is, to our knowledge, missing.

Here, we present OneTwoTree (<http://onetwotree.tau.ac.il/>), a user-friendly online tool for species-tree reconstruction. This tool aims for a wide audience of researchers, ranging in their familiarity with the various steps involved in phylogeny reconstruction. OneTwoTree allows for both a simple and fully automated execution for the nonspecialist while providing skilled researchers with the ability to fine-tune various steps of the analysis. An overall summary of the

distinct features and limitations of OneTwoTree in comparison with several available tools is presented in Table 1. Among its unique features, the sole input requirement for OneTwoTree is a list of taxa names (in any taxonomic hierarchy) to be analysed. The computation then considers all sequence data available in NCBI GenBank (Benson et al., 2013) for the requested taxa, regardless of which specific loci are available. A clustering procedure is then performed to assign sequences to orthology groups that form the basis of the subsequent alignment and tree reconstruction steps. Another unique feature of OneTwoTree is its ability to automatically select an appropriate outgroup species to be included in the analysis. This outgroup can then be used to root the tree and to determine the directionality of evolution. OneTwoTree additionally provides users with the ability to perform follow-up analyses, including the possibility to adjust various steps of the computational pipeline or to modify the automatically obtained sequence clusters. The main steps of the computational pipeline are briefly outlined below and are illustrated schematically in Figure 1 (in-depth details of each step and updates are available in the OVERVIEW section of the Web server at <http://onetwotree.tau.ac.il/overview.html> and in the Supporting Information Appendix S1). We then demonstrate the use of OneTwoTree by reconstructing the phylogeny of the plant tribe Antirrhineae and comparing the results to a recently published phylogeny.

2 | MATERIALS AND METHODS

2.1 | Sequence generator

All sequences belonging to Eukarya (mammals, rodents, primates, plants, vertebrates and invertebrates) were downloaded from NCBI GenBank (Benson et al., 2013) and stored locally. Sequences were preprocessed and organized for rapid sequence retrieval. This step further enables rapid updating of the database upon new GenBank releases.

The only input requirement for OneTwoTree is a list of taxa names or TaxIDs of interest, of any taxonomical hierarchy. A name resolution process is applied for all names while accounting for synonymy, misspellings, and different naming conventions, following a user-specified reference naming database. The basic taxonomic unit for analysis is the species level but users can optionally include sequences from intraspecific variants or merge these with their recognized species. In case the user specifies a higher ranked taxon encompassing multiple species (e.g., a genus name), all circumscribed species will be included in the analysis as separate entities. By default, all available sequence data are retrieved for initial processing, while users can specify the exclusion of nuclear, mitochondrial, or chloroplast loci.

2.2 | Clustering process

A clustering procedure is used to partition the entire sequence set into orthology groups. Sequence clustering over all non-ITS sequences is performed using a modified version of OrthoMCL (Li,

TABLE 1 Comparison between OneTwoTree and other phylogeny reconstruction tools

Name and main platform	Input types	Name resolution	Markers	Orthology assignment	Phylogeny inference	Possible outputs	Special Features	Limitations
OneTwoTree (web-server)	Taxa names	✓ Flexible ^a	All available data on GenBank 2017	Markov clustering	<ul style="list-style-type: none"> Maximum Likelihood (ML) or Bayesian Dated or undated 	<ul style="list-style-type: none"> MSA Cluster data Phylogeny 	<ul style="list-style-type: none"> Automatic outgroup selection Follow-up analyses: <ul style="list-style-type: none"> Modify parameters Handle markers manually Only tool that is based on all currently available GenBank data 	<ul style="list-style-type: none"> Running time depends on server load (it is advised to make large analyses using the stand-alone Virtual Box version)
Phylogeny.fr (web-server)	User-supplied sequence data	×	User-specified	User-specified	ML or Bayesian	<ul style="list-style-type: none"> MSA Phylogeny 	Dynamic tree editing interface	<ul style="list-style-type: none"> <200 sequences Jobs are restricted to 100 Mb
PhyLoTa (web-server)	Constraint taxa names	×	All available data on GenBank 2013	Single-linkage clustering	Maximum parsimony	<ul style="list-style-type: none"> Cluster data Phylogeny 	Short running time	<ul style="list-style-type: none"> Requires constant update (last available update – Feb. 2013) Incompatibility between input and output in case of paraphyly
PHLAWD (locally install on Linux-based OS ^b)	<ul style="list-style-type: none"> Taxa names Requested marker name & Seed sequence 	×	User-specified	User-specified	ML	Single-marker MSA	Taxonomy guided alignment	<ul style="list-style-type: none"> Only user-defined markers are retrieved Taxa and marker names should match NCBI naming Performed independently per marker
SUPERSMART (locally installed through a Virtual Machine ^c)	Taxa names	✓ NCBI Taxonomy	Available data from PhyLoTA	Single-linkage clustering	ML or Bayesian Dated	<ul style="list-style-type: none"> MSA Cluster data Phylogeny 	<ul style="list-style-type: none"> Incorporates NGS data Advanced dating options 	<ul style="list-style-type: none"> Depends on PhyLoTA updates

Note. ^aUser-specified reference naming database (Supporting Information Appendix S1). ^bOperating System; PHLAWD should be installed with its prerequisites. ^cProgram should be installed with its prerequisites. An online version with simpler functionality is also available.

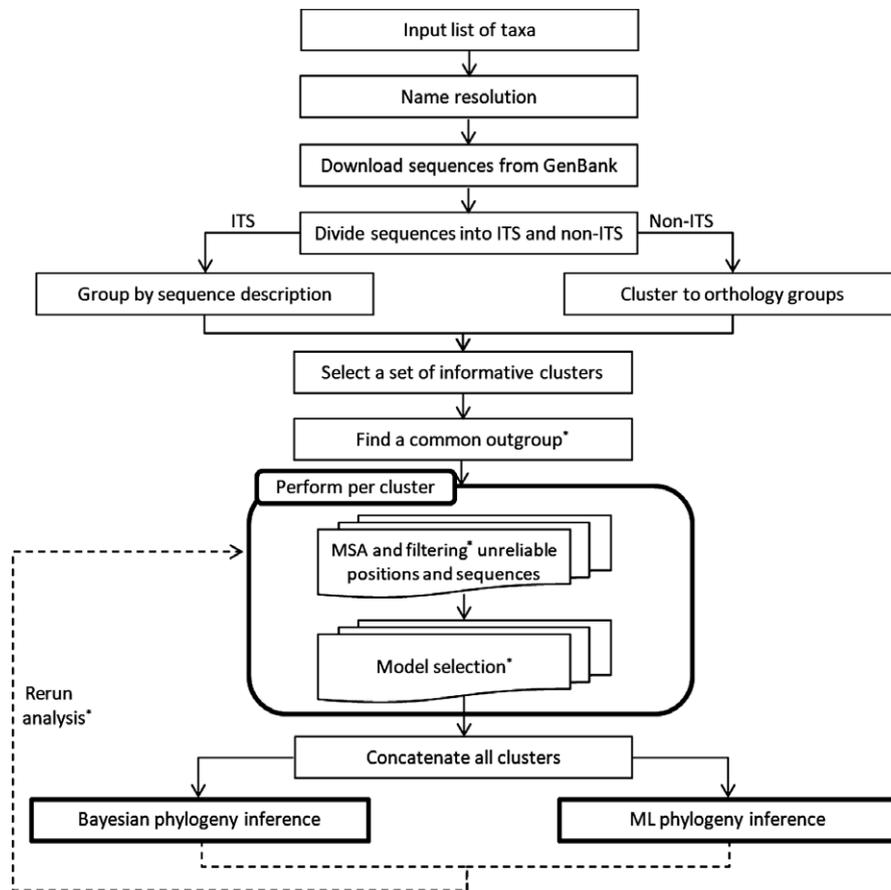


FIGURE 1 A schematic representation of the OneTwoTree computational pipeline. Asterisk denotes optional steps according to user preference. The full pipeline is described in the Supporting information Appendix S1

Stoekert, & Roos, 2003) or other alternative tools. In cases more than one accession per TaxID is found in a single cluster, a representative accession is selected. This selection process considers the similarity across TaxIDs, thus preferring the selection of orthologs over paralogs (see Supporting Information Appendix S1). The clustering (and subsequent alignment) of sequences belonging to the internal transcribed spacer (ITS) region is performed using a dedicated protocol to overcome probable errors that could occur if the two ITS sequences, ITS1 and ITS2, are clustered and then aligned together as suggested by Katoh & Standley, 2013. In the next version of OneTwoTree a more generalized approach will be available, allowing users to specify a set of markers to be identified using textual searches, independently of the sequence clustering process.

2.3 | Outgroup selection

One of the special features of OneTwoTree is its ability to automatically select an appropriate outgroup taxon for the clade of interest. Users can also provide as input a specified outgroup taxon or select to reconstruct the phylogeny without one. We note that the procedure developed in OneTwoTree is aimed at the detection of a species with sufficient sequence coverage that is taxonomically distant from all ingroup taxa (see Supporting Information Appendix S1). However, this automatic procedure cannot guarantee that the

selected taxon is not embedded within the ingroup taxa. It is thus recommended that users verify that the selected species is indeed a valid outgroup.

2.4 | Multiple sequence alignment

Aligning each cluster, along with the outgroup sequence (if selected), is performed using a number of possible programs, with MAFFT (Katoh & Standley, 2013) as default. Users can further filter poorly aligned positions or sequences using dedicated tools. An informative set of alignments is then assembled together, forming a concatenated MSA (i.e., a supermatrix) that serves as the input for subsequent phylogeny inference step.

2.5 | Phylogeny reconstruction

Given the MSA, users can choose to reconstruct a single tree, based on the supermatrix or obtain a *tree-per-cluster* based on individual sequence clusters, allowing the examination of incongruences among gene trees. Phylogeny reconstruction is performed using either RAxML (Stamatakis, 2014) or ExaML (Kozlov, Aberer, & Stamatakis, 2015) for maximum-likelihood (ML) inference, or MrBayes (Ronquist et al., 2012) for Bayesian inference. For each tool, various running parameters are available, including a range of clock and nonclock

models, resulting in phylogenies that are dated or undated and whose branch lengths are unconstrained or ultrametric. Users can also specify topological constraints or provide calibration points to enable divergence-time estimation. Inference reliability (i.e., node support) under the ML reconstruction can be obtained using a rapid bootstrap option (Stamatakis, Hoover, & Rougemont, 2008), while Bayesian posterior probabilities are an integral part of the Bayesian analysis.

2.6 | Rerun options

Upon completion, users may choose to rerun their analysis while modifying parameters and/or altering the sequence clusters that constitute the supermatrix (either by discarding specified clusters or by merging several clusters together).

3 | CASE STUDY: ANTIRRHINEAE PHYLOGENY

The plant clade Antirrhineae (*Plantaginaceae*) consists of ca. 30 genera that encompass over 320 species, including popular garden flowers, such as snapdragon (*Antirrhinum majus*) and toadflax (*Linaria vulgaris*). The evolutionary relationships within Antirrhineae were previously investigated by a number of studies, with the most recent one conducted by Ogutcen and Vamosi (2016). This study identified six well-supported clades within Antirrhineae and established the monophyly of the tribe. The inferred phylogeny included 146 species from Antirrhineae together with an additional 11 preselected outgroup species, one from each tribe of the family. Five markers were chosen for the phylogeny inference, including the nuclear *ITS* region, and four plastid regions: *ndhF*, *rbcl* large subunit, *rps16*, and *trnL-F*.

To illustrate the applicability of OneTwoTree, we used it to reconstruct the phylogeny of the tribe while specifying the word “Antirrhineae” as the sole input, choosing to exclude all hybrids and intraspecific variants. The assembled supermatrix spanned 265 species and consisted of 1,054 sequences that were divided into 21 sequence clusters (Supporting Information Table S2). The five markers with the highest coverage were *ITS*, *rpl32*, *ndhF*, *trnG-S*, and a region comprising the *trnK* and *matK* genes (with 243, 150, 132, 101 and 55 species sequences per marker, respectively). The partial overlap between the set of markers with highest coverage and those used in the reference study demonstrates the usability of OneTwoTree in efficiently identifying the putatively most informative set of markers shared by a specified group. The resulting phylogeny (Figure 2) provides the most inclusive Antirrhineae phylogeny reconstructed to date in terms of both taxa coverage and data matrix size. The phylogeny included all ingroup taxa that were analysed in the reference study, except for four taxa that were treated as distinct species by Ogutcen and Vamosi (2016) but are ranked as subspecies according to NCBI Taxonomy, and an additional species that was not retrieved since it was not classified under Antirrhineae in NCBI Taxonomy. In particular, 125 species were included in the phylogeny

reconstructed by OneTwoTree but were missing from the reference analysis. Nearly all of these had sequences belonging to the *ITS* region, 37 had *ndhF* and 15 had *trnL-F*. This shows that OneTwoTree can be used to guide the selection and retrieval of the most informative markers while presenting data coverage statistics for all potential ones.

OneTwoTree was further used to reconstruct two additional phylogenies given the following inputs: (a) The list of 157 taxa that appear in the phylogeny of Ogutcen and Vamosi (2016) and (b) specifying the word “Antirrhineae” while including intraspecific variants (Figure S1 and Table S1 for input a, and Table S3 for input b). Again, these reconstructed phylogenies were based on substantially higher data coverage in terms of both number of markers and number of sequences per marker, compared to the reference study. A summary of these analyses is presented in Table 2, and a detailed comparison between the phylogenies can be found in the Supporting Information Appendix S1.

4 | DISCUSSION

A main consideration in developing OneTwoTree was to enable a fully unsupervised phylogeny inference for a group of taxa of interest, while at the same time allowing users to fine-tune most steps of the reconstruction pipeline. Therefore, the single input requirement for OneTwoTree is a list of taxa names to be included in the phylogeny and for which all available sequences from GenBank are retrieved. This further necessitates the incorporation of several features to the computational workflow, such as clustering sequences to orthologous groups and the option to automatically select an outgroup species. Another implementation consideration was to make OneTwoTree available as an online tool, alleviating the need of locally installing any program or updating the sequence database. We emphasize, however, that the phylogeny reconstructed using OneTwoTree is by no means a one-click solution and should be taken to represent an initial phylogenetic hypothesis that could then guide subsequent in-depth analyses or any sequencing efforts. As discussed below, there are a number of extensions that could be integrated in future releases of OneTwoTree that would provide researchers with enhanced capabilities to tweak various stages of the pipeline and to overcome some of the potential caveats that are inherent to automatic procedures.

The current implementation of OneTwoTree is grounded within the supermatrix paradigm, such that once sequence clusters are identified they are concatenated into a single partitioned alignment that forms the basis for phylogeny inference. Yet, clustering methods could occasionally partition a single marker into multiple subgroups that either evolved far enough from each other, or whose sequence overlap is not sufficient to be placed in a single cluster. It is thus advisable to carefully inspect the resulting clusters and, if necessary, to conduct a follow-up OneTwoTree analysis by modifying the parameters of the clustering step (e.g., the inflation index) or by merging several clusters together. In addition, a number of

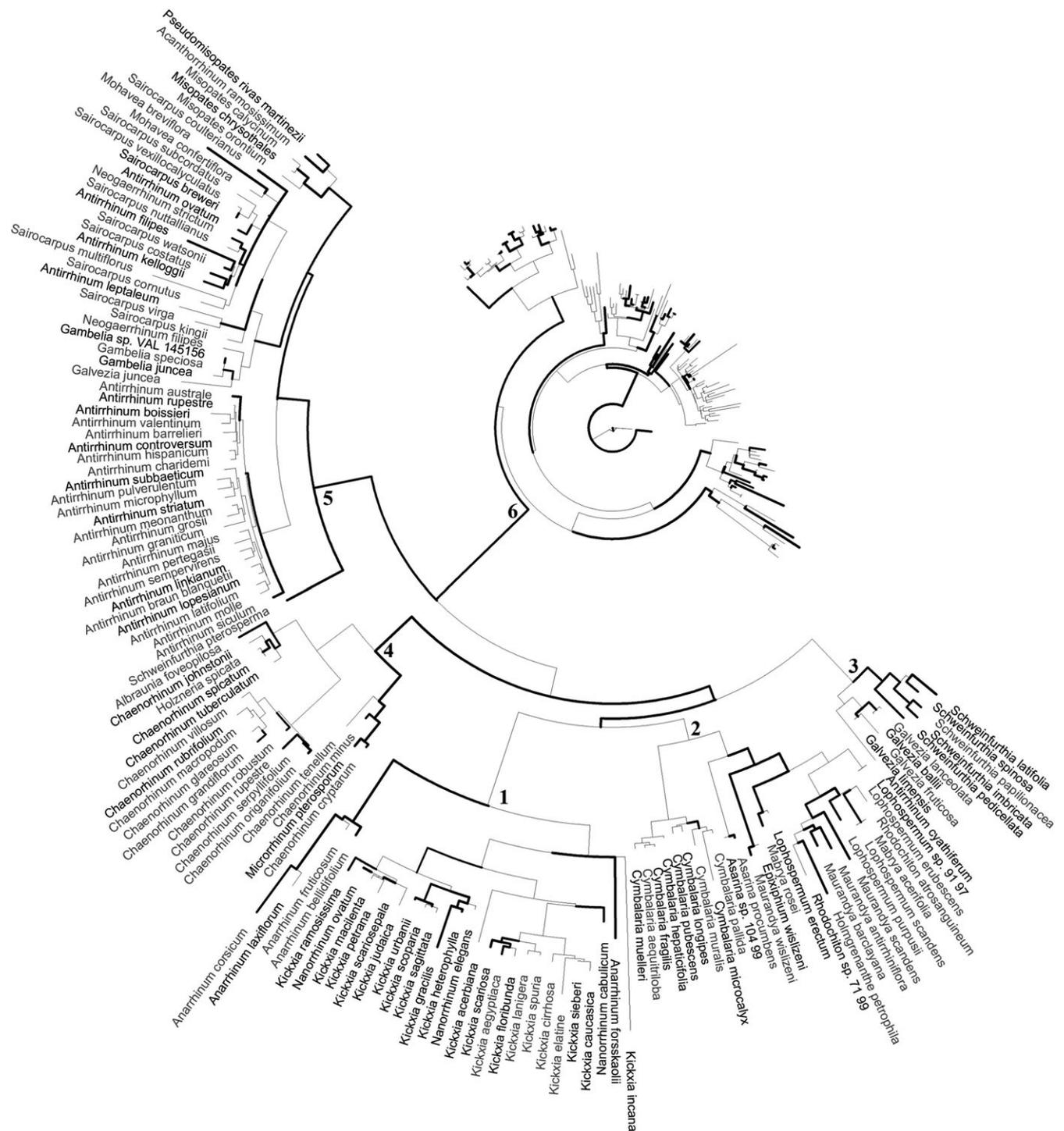


FIGURE 2 The Antirrhineae phylogeny produced by OneTwoTree using “Antirrhineae” as a sole input, excluding intraspecific variants. The most recent common ancestor of each clade of the tribe is assigned with a number— (1) *Anarrhinum*, (2) *Maurandya*, (3) *Gambelia*, (4) *Chaenorhinum*, (5) *Antirrhinum* and (6) *Linaria*; species in black are those that did not appear in the reference study, whereas the ones in gray did. A zoom-in on the *Linaria* clade is presented in Supporting Information Figure S2. The width of the branches is proportional to the bootstrap values, with thicker branch denoting higher support (ranging from 100, the thickest, to 85 or below, the thinnest)

processes, including hybridizations, lateral gene transfer and incomplete lineage sorting, may lead to incongruence among individual trees produced from distinct markers, resulting in inconsistent species-tree estimation (Degnan & Rosenberg, 2006). Several species-tree reconstruction methods have been designed to deal with

discordant data (e.g., Heled & Drummond, 2010; Mirarab & Warnow, 2015). These tools, however, are computationally demanding and it is thus advisable to apply them offline with powerful computational facilities, while OneTwoTree can be used to guide the data assembly step. Instead, users can test for congruence among the assembled

TABLE 2 Comparison of the Antirrhineae phylogenies produced by OneTwoTree and Ogutcen and Vamosi (2016)

Input and reconstruction method	Number of taxa	Number of sequences (top five markers) ^a	Number of markers	MSA length	Parsimony informative sites (%) ^b	Missing data (%) ^c	Running time (min)
Ogutcen and Vamosi (2016)	157	382 (382)	5	6,183	22.03	60	–
Taxa list identical to the reference study	157	839 (446)	24	28,331	12.70	83.21	52
“Antirrhineae” excluding intraspecific variants	265	1,054 (681)	21	25,788	12.62	86.58	84
“Antirrhineae” including intraspecific variants	337	1,297 (867)	20	23,689	13.51	86.41	133

Note. ^aThe number of sequences in the assembled supermatrix. In parentheses is the number of sequences in the five mostly covered markers. ^bPositions with at least two types of characters (e.g., nucleotides). ^cGaps and completely undetermined characters.

sequence clusters using a number of available approaches (e.g., *Concatenator*; Leigh, Susko, Baumgartner, & Roger, 2008) and then conduct a follow-up OneTwoTree analysis using one (or more) set of concordant clusters.

By default, OneTwoTree retrieves all sequences of the requested taxa. Yet, the inclusion of certain sequences could deteriorate the alignment. This could be the result of highly diverged sequences, partial sequences, as well as sequencing or annotation errors. The use of alignment filtering programs can be used to omit positions or sequences whose resolutions vary widely across alternative alignments. Likewise, the inclusion of species in the tree whose position is highly uncertain could distort the tree structure in such a way that would decrease the overall phylogenetic resolution. A possible solution would be to obtain a sample of possible phylogenies using a preliminary run and then to detect rogue taxa using tools such as RogueNaRok (Aberer, Krompass, & Stamatakis, 2013). In particular, this two-step approach requires extensive running time and could be offered as a follow-up option.

At last, the operational taxonomic units in a default OneTwoTree analysis correspond to terminal taxa in the hierarchy of NCBI taxonomy (usually species or subspecies). Following user preference, intraspecific variants can be excluded from analyses or included as distinct entities. Instead, this hierarchy could be collapsed, such that the tips of the phylogeny would include the sequences of all taxa circumscribed within a given rank. At present, OneTwoTree allows the merging of sequences from subspecies together with those of their recognized species into a single taxon. In the future, this could be performed above the species level, which would naturally enable the reconstruction of phylogenies whose terminal taxa are, for example, genera or families.

ACKNOWLEDGEMENTS

We thank Ezgi Ogutcen and Jana Vamosi for their insightful comments and data sharing, and Eli Levi Karin for the name suggestion. This study was supported by the Israel Science Foundation (961/17), the BSF-NSF (2016550), and by a fellowship to AR from the Edmond J. Safra Center for Bioinformatics.

AUTHOR CONTRIBUTIONS

I.M. conceived and supervised the study; M.D., M.E., A.R., O.C. and L.G. built the tool; A.R. analysed the case study; A.R. and I.M. wrote the manuscript.

DATA ACCESSIBILITY

OneTwoTree is freely available for online use at <http://onetree.tau.ac.il/>, together with an overview and a step-by-step tutorial. A stand-alone Virtual Box version is also provided under the *Download* tab. The source code, tutorial and example of results are freely available from the GitHub repository <https://github.com/MayroseLab/OneTwoTree>.

REFERENCES

- Aberer, A. J., Krompass, D., & Stamatakis, A. (2013). Pruning rogue taxa improves phylogenetic accuracy: An efficient algorithm and web service. *Systematic Biology*, *62*, 162–166. <https://doi.org/10.1093/sysbio/sys078>
- Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M. J., ... Vos, R. A. (2017). Toward a self-updating platform for estimating rates of speciation and migration, ages, and relationships of taxa. *Systematic Biology*, *66*, 152–166.
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, *41*, 36–42.
- de Queiroz, A., & Gatesy, J. (2007). The supermatrix approach to systematics. *Trends in Ecology and Evolution*, *22*, 34–41. <https://doi.org/10.1016/j.tree.2006.10.002>
- Degnan, J. H., & Rosenberg, N. A. (2006). Discordance of species trees with their most likely gene trees. *PLoS Genetics*, *2*, e68. <https://doi.org/10.1371/journal.pgen.0020068>
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., ... Gascuel, O. (2008). Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, *36*, W465–W469. <https://doi.org/10.1093/nar/gkn180>
- Dunn, C. W., Howison, M., & Zapata, F. (2013). Agalma: An automated phylogenomics workflow. *BMC Bioinformatics*, *14*, 330. <https://doi.org/10.1186/1471-2105-14-330>
- Felsenstein, J. (2004). *Inferring phylogenies*. Sunderland, MA: Sinauer Associates.
- Freyman, W. A. (2015). SUMAC: Constructing phylogenetic supermatrices and assessing partially decisive taxon coverage. *Evolutionary Bioinformatics Online*, *11*, 263–266.
- Frickey, T., & Lupas, A. N. (2004). PhyloGenie: Automated phylome generation and analysis. *Nucleic Acids Research*, *32*, 5231–5238. <https://doi.org/10.1093/nar/gkh867>
- Heled, J., & Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, *27*, 570–580. <https://doi.org/10.1093/molbev/msp274>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*, 772–780. <https://doi.org/10.1093/molbev/mst010>

- Kozlov, A. M., Aberer, A. J., & Stamatakis, A. (2015). ExaML version 3: A tool for phylogenomic analyses on supercomputers. *Bioinformatics*, *31*, 2577–2579. <https://doi.org/10.1093/bioinformatics/btv184>
- Leigh, J. W., Susko, E., Baumgartner, M., & Roger, A. J. (2008). Testing congruence in phylogenomic analysis. *Systematic Biology*, *57*, 104–115. <https://doi.org/10.1080/10635150801910436>
- Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*, 2178–2189. <https://doi.org/10.1101/gr.1224503>
- Mirarab, S., & Warnow, T. (2015). ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, *31*, i44–i52. <https://doi.org/10.1093/bioinformatics/btv234>
- Ogutcen, E., & Vamosi, J. C. (2016). A phylogenetic study of the tribe Antirrhineae: Genome duplications and long-distance dispersals from the Old World to the New World. *American Journal of Botany*, *103*, 1071–1081. <https://doi.org/10.3732/ajb.1500464>
- Pearse, W. D., & Purvis, A. (2013). phyloGenerator: An automated phylogeny generation tool for ecologists. *Methods in Ecology and Evolution*, *4*, 692–698. <https://doi.org/10.1111/2041-210X.12055>
- Robbertse, B., Yoder, R. J., Boyd, A., Reeves, J., & Spatafora, J. W. (2011). Hal: An automated pipeline for phylogenetic analyses of genomic data. *PLoS Currents*, *3*, RRN1213.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*, 539–542. <https://doi.org/10.1093/sysbio/sys029>
- Sanderson, M. J., Boss, D., Chen, D., Cranston, K. A., & Wehe, A. (2008). The PhyLoTA browser: Processing GenBank for molecular phylogenetics research. *Systematic Biology*, *57*, 335–346. <https://doi.org/10.1080/10635150802158688>
- Smith, S. A., Beaulieu, J. M., & Donoghue, M. J. (2009). Mega-phylogeny approach for comparative biology: An alternative to supertree and supermatrix approaches. *BMC Evolutionary Biology*, *9*, 37. <https://doi.org/10.1186/1471-2148-9-37>
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stamatakis, A., Hoover, P., & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology*, *57*, 758–771. <https://doi.org/10.1080/10635150802429642>
- Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., ... Mazel, F. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews*, *92*, 698–715. <https://doi.org/10.1111/brv.12252>
- Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: Principles and practice. *Nature Reviews Genetics*, *13*, 303–314. <https://doi.org/10.1038/nrg3186>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Drori M, Rice A, Einhorn M, Chay O, Glick L, Mayrose I. OneTwoTree: An online tool for phylogeny reconstruction. *Mol Ecol Resour.* 2018;18:1492–1499. <https://doi.org/10.1111/1755-0998.12927>