

TreeBASE a bioinformatics tool for phylogenetic analysis: Submission guidelines made easy

Shunmugiah V Ramesh

ICAR-Indian Institute of Soybean Research (ICAR-IISR), Indore 452 001, Madhya Pradesh

ABSTRACT

TreeBASE is a data repository that accepts data related to sequence alignment and the consequential phylogenetic trees so that it could be made available, for wider scientific community. Further, increasing number of scientific journals requires raw data and supplementary information regarding phylogeny reconstruction and the resulting phylogenetic trees to be deposited in data repositories such as TreeBASE etc. Although the treeBASE website (<https://treebase.org/treebase-web/home.html>) provides a little information regarding submission, the present communication would like to supplement the same and provide a simple, easy to perform TreeBASE submission guide.

KEY WORDS: EVOLUTIONARY BIOLOGY, PHYLOGENY, REPOSITORY, TAXON

INTRODUCTION

TreeBASE is a database for phylogenetic information (Piel *et al*; 2002). It houses user submitted phylogenetic trees, linked data, including sequence alignment files. The repository admits all kinds of reconstructed phylogenies such as phylogeny of species, genes, and even whole population. The data submitted to TreeBASE will be made available for public viewing only when the data is published in a journal or in any other scientific publications. It would be made available only to the editors or referees of the peer-reviewed journal through a designated URL when user specifies that the submitted data is 'under review'. This way anonymous access to the data is provided to the peer-reviewers and editors of the jour-

nals well in advance. The status of data changes "Ready" immediately after the article has been published so that it is made available for the wider scientific community. TreeBASE is governed by The Phyloinformatics Research Foundation, Inc. (<http://www.phylorf.org/prf>).

The utility of such a repository is enormous as it provides easy access to the phylogenetic data for research community. This not only saves time but also enables researchers working in the specific area to avoid duplicity and to arrive at meaningful conclusions in a short span of time. Eventhough the TreeBASE help page provides a little information regarding how to submit the data, the information available is insufficient. Hence a step-by-step guide for making a treeBASE submission is provided herewith. Any researcher with a limited train-

ARTICLE INFORMATION:

*Corresponding Author: rameshsvbio@gmail.com

Received 28th May, 2016

Accepted after revision 20th June, 2016

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007

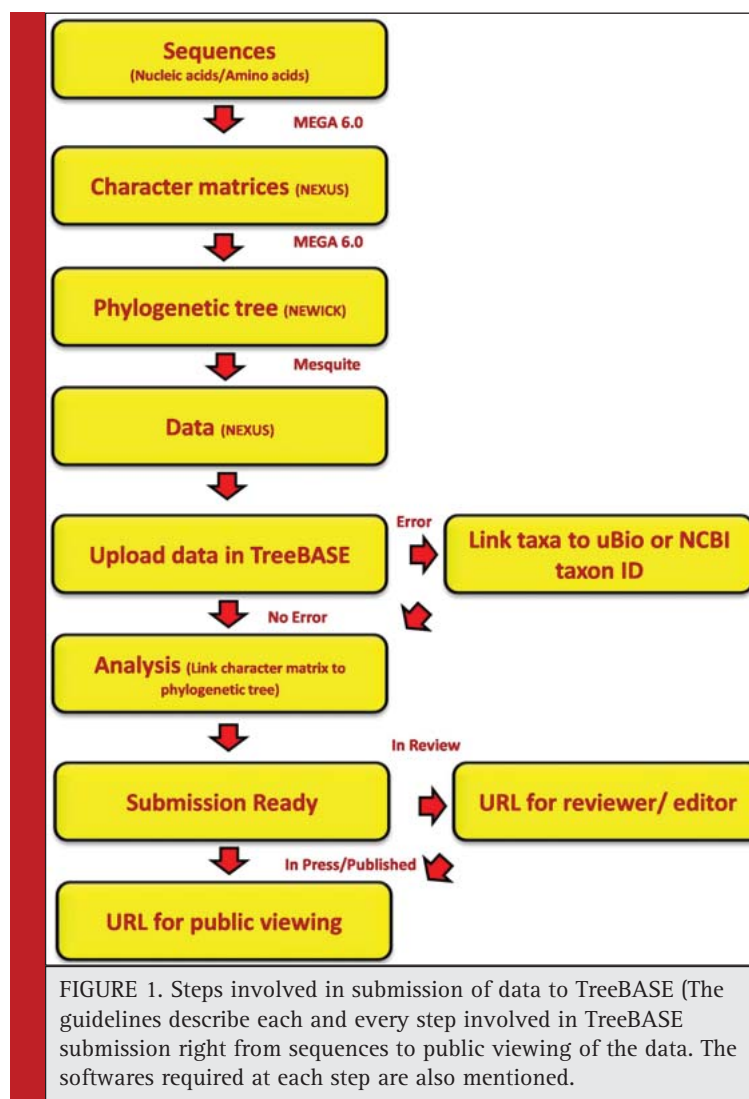


Thomson Reuters ISI SCI Indexed Journal

NAAS Journal Score : 3.48

© A Society of Science and Nature Publication, 2016. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>



ing on handling phylogenetic data would be able to perform this submission easily if he/she follows following steps (Fig.1):

GUIDELINES FOR TREEBASE SUBMISSION

In general phylogenetic trees are generated using nucleic acid or amino acids sequence data. This input data requires to be formatted to suit the submission standards. Original sequence data (nucleic acid/aminoacid), has to be prepared as a notepad file. It is essential to match the names of the sequences (referred to as Taxa) with NCBI Taxon ID and/or uBio ID. This ensures that while uploading the in TreeBASE, all the taxon names are automatically validated in treeBASE. To illustrate this see the following example:

eg) KP827649 WAUSA- original taxon name generally read as GenBank

accession no followed by isolate/strain/place etc. It is imperative to change the same to Tomato spotted wilt virus. KP827649WAUSA

This not only ensures easy file upload but also original phylogenetic tree in the publication and TreeBASE submissions are easily related. The identifiers such as isolate/strain name, accession no.s, etc placed after taxon's scientific name with a period in between helps in relating the original tree to TreeBASE entry (Sanderson *et al.*, 1994).

Similarly, sequence alignment files, (called as character matrices), used in the preparation of phylogenetic trees are also required to be formatted. Prepare the character matrix using any sequence alignment editor programme such as MEGA (Tamura *et al*; 2013). Save these results in FASTA format and name the files explicitly, eg *Glycine max*.TF alignment. Using the same character matrices, generate the phylogenetic trees in the softwares such as MEGA ver. 6.0 (Tamura *et al*; 2013). It is important

to save the tree session by exporting the resulting trees in newick format. Now two important requirements for submission that are character matrix and phylogenetic trees are ready.

TreeBASE suggests Mesquite software to convert the data files (sequence alignment and phylogenetic trees) into nexus format (Maddison and Maddison 2015) (<http://mesquiteproject.org/>). Now open all the alignment files in Mesquite and save it in nexus format. Open the tree files in Mesquite and make certain that the tree branches and taxon names match exactly the original publication. Generally flipping off the branches and editing the taxon names will suffice to match the phylogenetic tree with the published data/or data under review. With this the files ready for upload are ready.

Create an account in TreeBASE with a preferred user name and password. It is important to keep information regarding the study such as title of the manuscript, abstract, key words etc. Enter title of the manuscript in “Name of the study” space. Under the section ‘notes for the study’, mention salient details of the study in few lines so as to ensure easy communication between you and TreeBASE staff once your submission is ready for public viewing.

In the ‘citation section’ enter your publication details (for articles in press or published work) or else set the status as “In Review”. Now that the metadata is entered, files are ready for uploading. Use the ‘upload’ link available in the toolbox section and perform data upload. Once files are uploaded, verify that taxa, matrices and trees subsection of tool box are not highlighted red as it implies error in uploaded taxa, or matrices or trees.

If error is shown with the uploaded file, click taxa and link it to uBio or NCBI taxon ID. Generally linking your taxa to uBio or NCBI taxon ID is an issue. If no error in the files uploaded go to analysis section. In the analysis section, incorporate steps so as to link your uploaded data matrix to the phylogenetic tree. This step essentially involves matching your sequence alignment data to the output phylogenetic tree by placing a phylogeny reconstruction analysis step in between. Analysis step ends

when all the uploaded sequence alignment and tree data is linked ie) none of the data submitted is “orphaned”.

Now the submission is ready for public viewing, one could provide the accession no, and or URL linked to your submission to reviewer/editor so that your data is accessible for them. Once accepted for publication you could make your data available for public viewing through a designated URL.

CONCLUSION

In an “open access” era that fuels free flow of scientific data, repositories such as TreeBASE play an indispensable role in dissemination of knowledge and saves time to arrive at meaningful conclusions. The easy to perform step-by-step guidelines for TreeBASE submission presented here would enable swift submission of evolutionary biology data into this data repository.

REFERENCES

- <http://mesquiteproject.org/>
<http://www.phylorlf.org/prf>
<https://treebase.org/treebase-web/home.html>
 Maddison WP and Maddison DR (2015) Mesquite: A modular system evolutionary analysis. Version 3.04 <http://mesquiteproject.org>
 Piel WH, Donoghue MJ and Sanderson MJ (2002) TreeBASE: a database of phylogenetic knowledge. Pp. 41-47. In: Shimura J, Wilson KL, and Gordon D (Eds.). To the interoperable Catalog of Life with partners Species 2000 Asia Oceania. Research Report from the National Institute for Environmental Studies No. 171, Tsukuba, Japan
 Sanderson M J, Donoghue M J, Piel W, and Eriksson T (1994) TreeBASE: a prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. *American Journal of Botany*. Vol 81 No. 6: Page 183.
 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*; Vol 30: pages 2725-2729