OneTwoTree

A step-by-step tutorial

This tutorial guides users in the process of phylogeny reconstruction using OneTwoTree.

OneTwoTree is an online tree reconstruction tool, designed to generate phylogenies in an unsupervised manner, aiming to a wide range of users' expertise.

Mainly, the tutorial will describe how to reconstruct different types of phylogenies using the tool. For a full description of each feature and its underlying pipeline please visit the online OVERVIEW page and the paper.

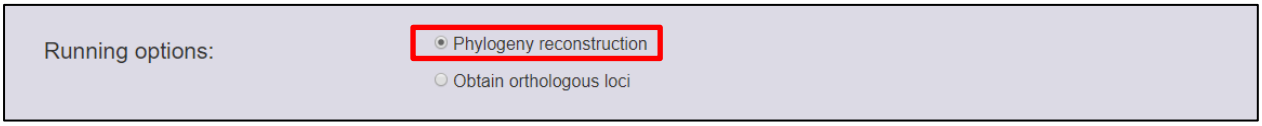
The topics covered by this tutorial are:

1) How to get a tree in a click.
2) How to handle your input.
3) How to root your tree
4) Phylogeny inference:
    a. How to add topological constraints.
    b. How to obtain time-calibrated phylogeny.
    c. Bootstrapping.
5) Navigating through the Results page.
6) Follow-up analyses.

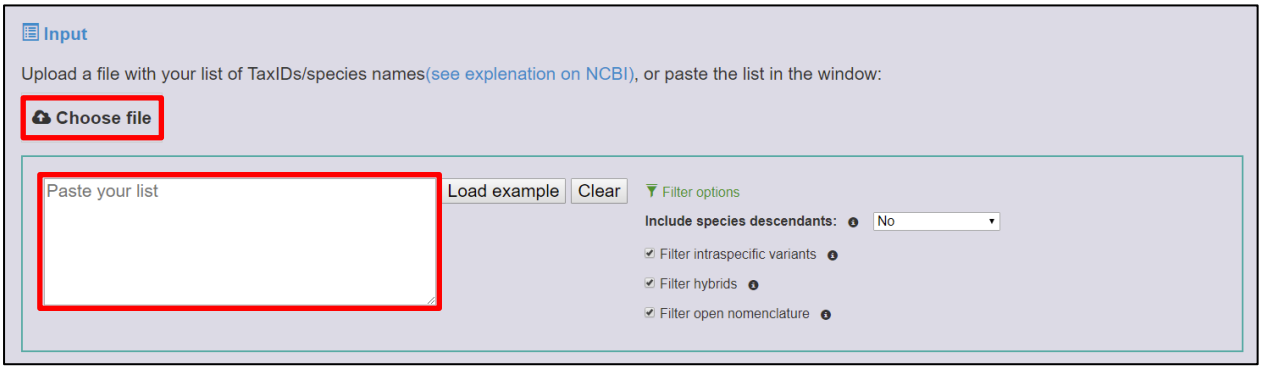
If you don't have your own data and wish to try out OneTwoTree, click on the Load example button to obtain sample data.

1) How to get a tree in a click

First, make sure that your operational mode is set to *Phylogeny reconstruction* under the *Running options* (this is the default):



Next, paste your input list or upload it as a file in *txt* or *csv* formats. This file should not contain any headers. To see an example, click on Load example.



The input can include both taxa names and TaxIDs, without any further specifications. You can also combine between species names and higher-ranked taxa.
Note: Currently, OneTwoTree supports species names separated by a *space* character between the genus and species names, and *not* other characters such as *underscore* or *dot*.

Fill in your Email address and give an informative job title for your own convenience.

Clicking the Submit button will produce a tree following your input list with all default parameters (namely, an unrooted, Maximum Likelihood phylogeny based on a supermatrix).

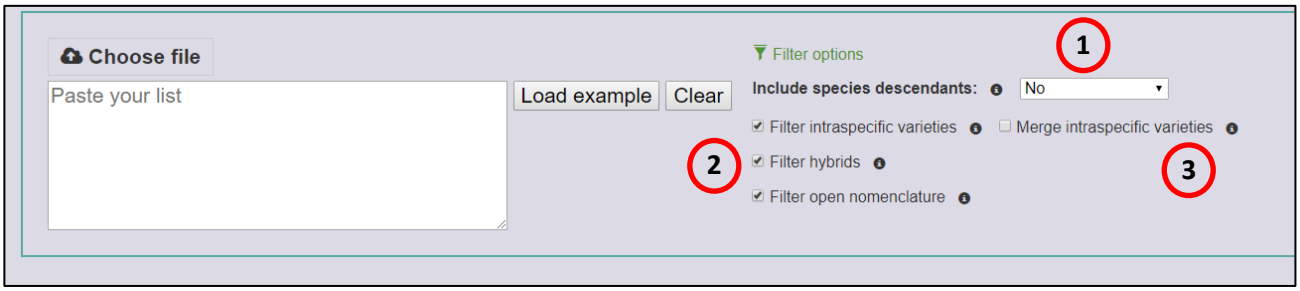
2) How to handle your input

There are several ways to retrieve sequences. The default parameters of the *Filter options* will fetch sequences only for the specified taxa. The filtering options include:

1. **Include species descendants:** The basic taxonomic unit of operation in OneTwoTree is the species-level. If you wish to retrieve sequences below the species-level (e.g., subspecies), change this parameter to "Yes".
   Note: This option is ignored in case the input name is a higher-ranked taxon (e.g., if the input name is a genus, all of its descendants, including intraspecific variants, are retrieved; for these to be omitted users should use the filters in #2).

2. **Filter intraspecific varieties/ Filter hybrids/ Filter open nomenclature:** Modify these filters to obtain intraspecific variants, hybrids and open nomenclature conventions. For example, in case the *Filter intraspecific variants* is unchecked, then intraspecific variants sequences will be included in the analysis.

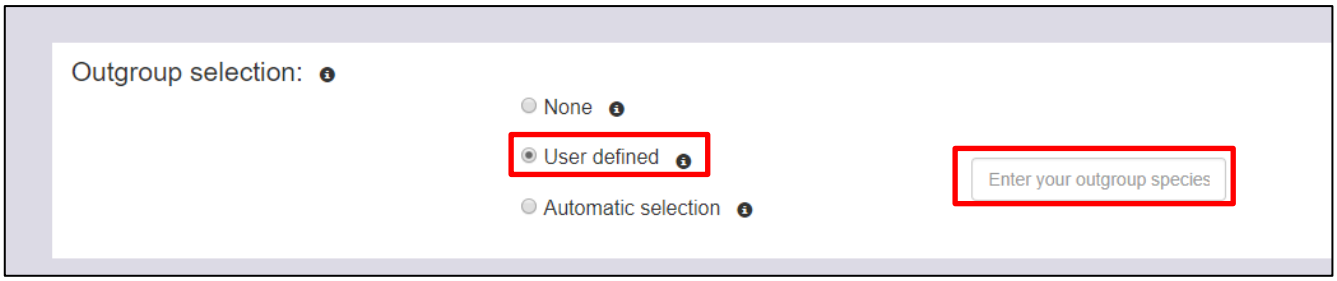
3. **Merge intraspecific varieties:** In case there are intraspecific varieties in the results, as well as their progenitor species (i.e., the binomial species name), their sequences can be unified into a single tip in the phylogeny.
   Note: The progenitor must be included in the original list (or under a higher-taxa) for its sequences to be included in the analysis. For example if the input includes *Iris caucaisca* (TaxID 292549) and *Iris caucasica subsp. turcica* (TaxID 995789), the merge option will retrieve all sequences in the database that belong to both TaxIDs, and will be placed under TaxID 292549. The phylogeny itself will include only *Iris caucaisca* and not *Iris caucasica subsp. turcica*.



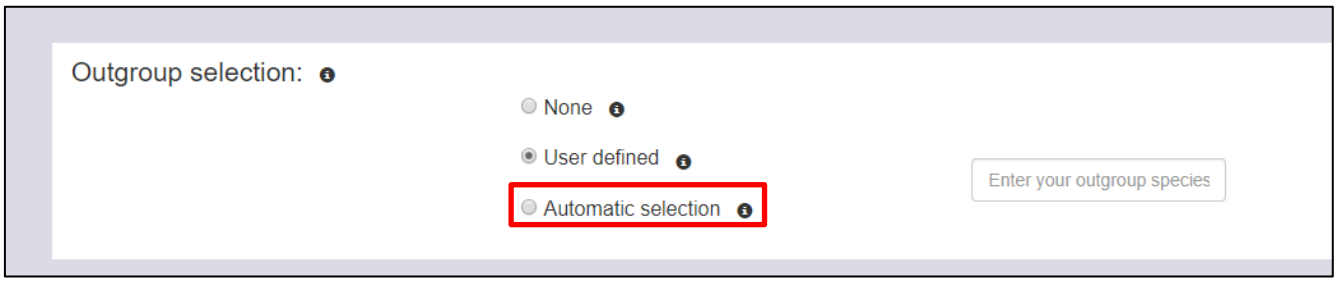
3) How to root your tree

Click on Advanced parameters.
The first parameter that can be modified is the *Outgroup selection* section. Once an outgroup is added to a phylogeny users can root the tree and determine the direction of evolution.

There are two ways to add an outgroup to the analysis. First, users can specify a taxon name or a TaxID in the designated input box. The specified name can be one of the names already part of the input list or not.



Second, users may choose to allow an automatic selection of the outgroup. Choosing this option will substantially increase the running time.

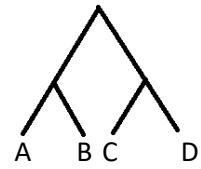


4) Phylogeny inference

Click on Advanced parameters.
Under the *Phylogeny inference* section you can alter the inference tool (Maximum Likelihood or Bayesian approach). Each inference tool has its own supported set of *Nucleotide models*.

a) How to add topological constraints

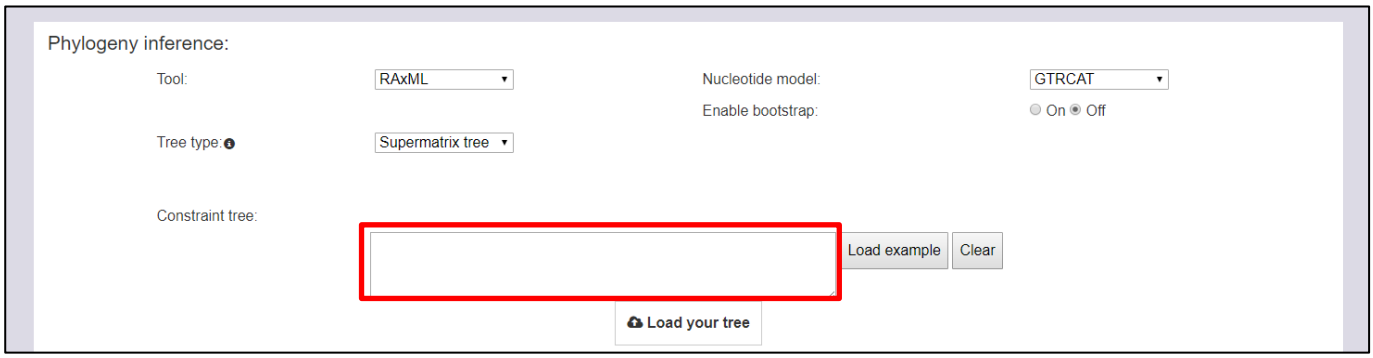
Adding a constraint means that there is a prior hypothesis or knowledge regarding several taxonomic groupings. OneTwoTree enables to take such knowledge under consideration. The constraint itself should be provided in a NEWICK format. For example, here is a tree plot with its NEWICK representation:



((A,B),(C,D));

In this case taxa A and B are constrained to be more closely related to each other than either one is to C or D. For example: Consider an input list that contains 10 species A-J. The phylogenetic relationship between four of these species (A-D) is known such that A and B form a monophyletic group and C + D form a monophyletic group, while there is no other knowledge regarding any other taxonomic groupings. Then, the constrain phylogeny shown above represents this constraint. In case species A-D are known to form a clade, and E+F are external to this clade but their relationship is unknown, the supplied NEWICK representation should be **(A,B,C,D);**.

Branch lengths supplied in the NEWICK format are ignored. Paste your constraint tree in the designated place or load a file.



Note: Adding a constraint tree when working with the ExaML tool for phylogeny inference, all taxa should appear in the user-constraint tree, otherwise an unconstraint tree will be produced.

b) How to obtain an ultrametric tree

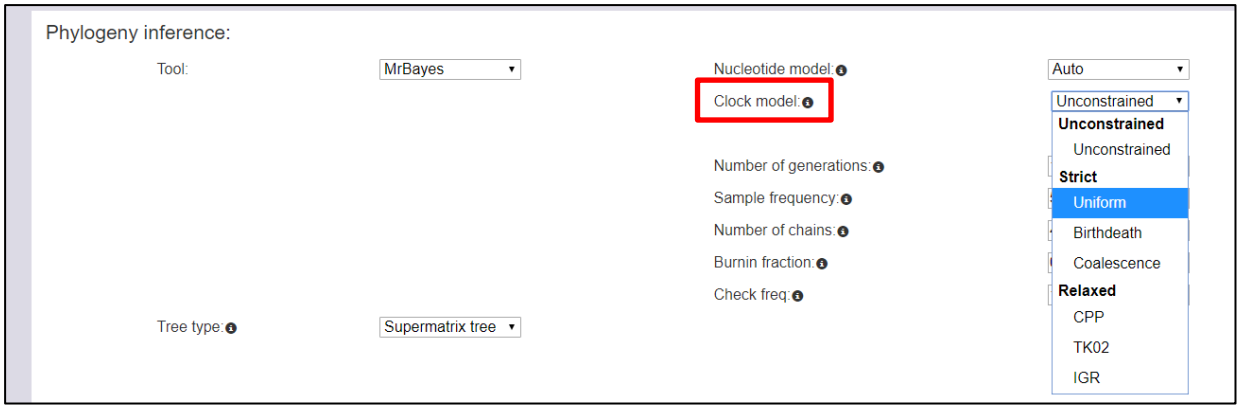
The divergence time estimation option will only be enabled if the tree is about to be rooted (see *Root your tree* section to find out how this can be done).

Using RAxML or ExaML as an inference tool

To get an ultrametric tree: First choose to root your tree, then the *Divergence time estimation* option will appear under the *Phylogeny inference* section. Set the *Divergence time estimation* to *Yes* for an ultrametric tree output.

Using MrBayes as an inference tool

To obtain an ultrametric tree using MrBayes the clock model should be any of the Strict or Relaxed clock models.



c) How to obtain time-calibrated phylogeny

Users can obtain a time-calibrated tree by adding calibration points. Constraining the divergence time of a certain ancestral node in the phylogeny helps to calibrate the branch lengths according to a time scale defined by the user. Usually, these time estimates are based on fossil data. To use this option, first set the *Divergence time* estimation parameter to *Yes.*

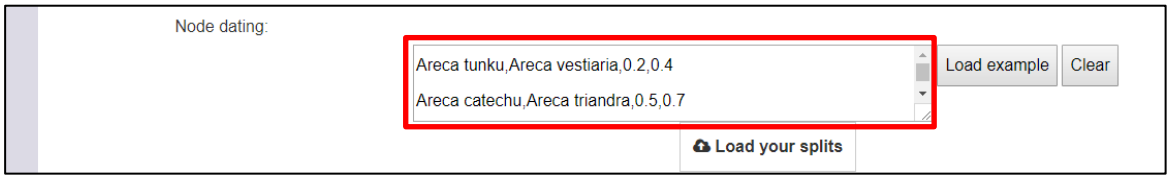
Using RAxML or ExaML as an inference tool

Two optimized tools are available: (1) treePL or (2) PLL-DPPDiv.

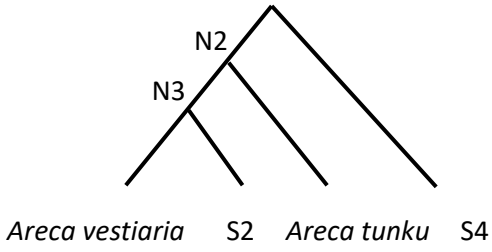
Using MrBayes as an inference tool

MrBayes has the option to calibrate nodes according to user-supplied data.
Note: When using MrBayes, only clock models that support the uniform branch-length prior can be applied together with user-provided calibration points.

For both phylogeny inference tools, users can add calibration points by defining a split (i.e., two taxa names that the desired calibration point is their most recent common ancestor) together with two time estimations – minimal and maximal ages of the relative split. In case no calibration points are provided, the tree will be calibrated relatively to a root age of 1.



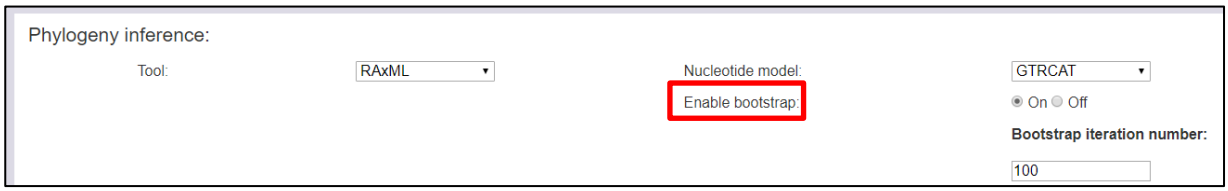
In the example shown above there are two splits that are calibrated, one in each line. The first number in each row represents the minimal age and the second one represents the maximal age. In this case, the first specifies that *Areca tunku* and *Areca vestiaria* diverged before 0.2 and 0.4 million years ago. Note that these two species need not form a monophyletic group. For example, if the inferred topology was:



The constraint will apply to N2.

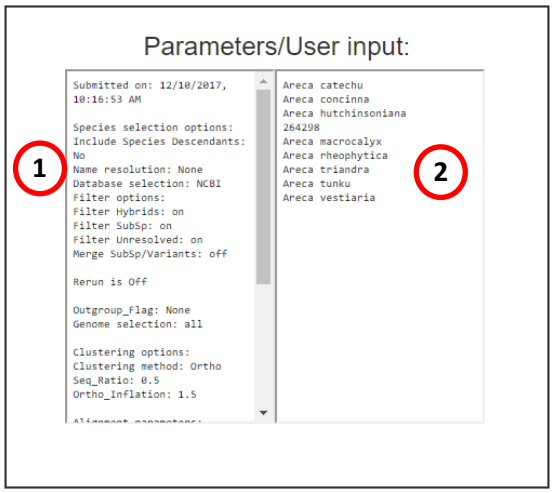
d) Bootstrapping

Under the ML approach one can add splits reliability support using a rapid bootstrap estimation. Click the *Enable bootstrap* option and specify the number of replicates.
Note: Choosing this option will substantially increase the running time.

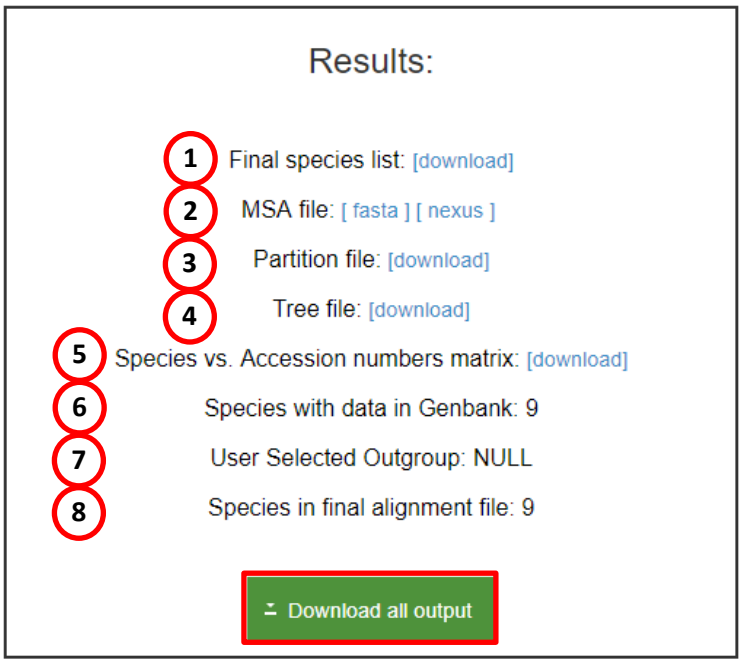


Navigating through the Results page

When the analysis is finished, the resulting phylogeny together with various other files are produced. All outputs are further available for download in a zip format, as well as a single-file download option. Users can also view their resulting phylogeny and a table representing the clusters that comprise the MSA.



1. Users can browse the parameters according to which the analysis was run.
2. In this window users will see the input as provided by them. These names will not necessarily represent the names that appear in the final phylogeny.



1. The *Final species list* contains the list of taxa that exist in the phylogeny. The will not necessarily match the set of names provided by the user (in case of name resolution conventions and/or data availability).

2. A *Multiple Sequence Alignment* file is available both in FASTA and in NEXUS formats.

3. The *Partition file* contains the positions of each marker in the final concatenated MSA.

4. Tree file is the resulting phylogeny. Apart from being ready for download, it is presented online (see below) using the WASABI platform (Veidenberg *et al.* 2016).

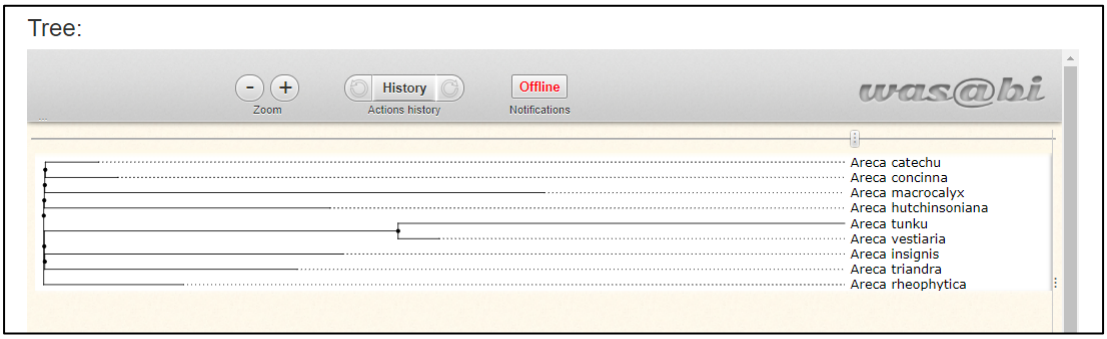
5. A matrix representing the taxa in the phylogeny, together with the accessions of the respective markers used in the MSA.

6. The number of taxa that had sequence data in GenBank, prior to any additional processing.

7. If user selected to add an outgroup, the selected (or user-chosen) one will be presented here.

8. The number of taxa in the final concatenated MSA (and therefore in the final phylogeny), after passing all filtering processes.

Tree:



Using the WASABI platform (Veidenberg *et al.* 2016) users can view, one the analysis is completed, their resulting phylogeny.

Clusters table:

1 Cluster	2 Description	3 Type	4 Median Length	5 Num of species	6 Included	7 Download
1	PRK	NUC	680	6	yes	Aligned
3	Areca hutchinsoniana gene for ITS1, 5.8S rRNA, ITS2, complete sequence, clone: 2.	NUC	723	6	yes	Aligned
2	RPB2	NUC	913	4	yes	Aligned

This table summarizes all clusters that were found in the data of the taxa in question. From this set of markers the final supermatrix was built:

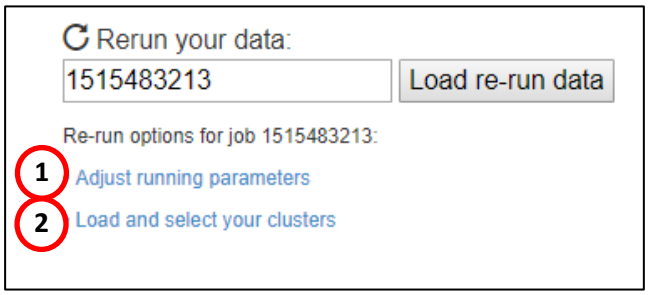
1. Cluster number.
2. The name of the cluster.
3. Genome type – NUC for nuclear, mtDNA for mitochondrial or cpDNA for chloroplast.
4. The median length of all available sequences belonging to that marker, from the taxa in question.
5. The number of species that had available data for this marker, from the taxa in question.
6. Was this marker included in the final supermatrix – Yes/No.
7. Users can choose to download each cluster in an unalighned, aligned, or both, manners.

Follow-up analyses

After inspecting the results, users can decide to rerun their analysis by clicking the

RERUN your data button. This will direct the user to OneTwoTree's homepage, with the

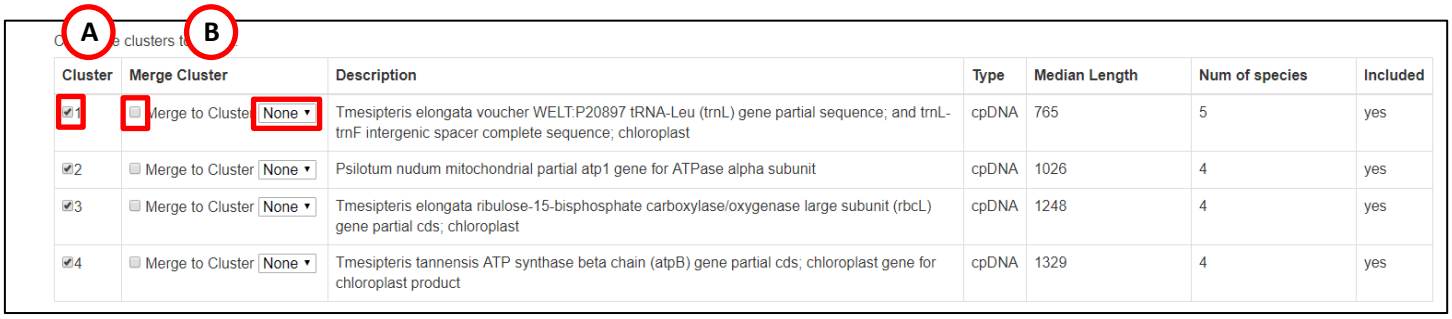
JobID copied to the *Rerun your data* section, and the results themselves loaded onto the page:



There are two options to rerun your results:

1. *Adjust running parameters* – Choosing this option will move the page to the *Advanced parameters* section where users can run the analysis once again with a different tree reconstruction method together with their respective parameters. Altering other parameters that are not related to the inference tool will take no effect and the rerun process will fail.

2. *Load and select your clusters* – This option allows the user to modify the clusters according to which the MSA was formed. The options are:



A. Choose which clusters to include by selecting and de-selecting the desired markers.

B. Merge two loci together. Tick the *Merge to cluster* option and select a cluster number. It is enough to merge one cluster with the other (no need to repeat this reciprocally).

After altering your rerun options, submit the job by clicking the Submit button.