



OFPT: a one-stop software for fungal phylogeny

Zeng XY^{1,2}, Tan TJ¹, Tian FH¹, Wang Y¹, and Wen TC^{3,4*}

¹Department of Plant Pathology, College of Agriculture, Guizhou University, Guiyang 550025, China

²Center of Excellence in Fungal Research, Mae Fah Luang University, Chiang Rai 57100, Thailand

³National Key Laboratory of Green Pesticide, Key Laboratory of Green Pesticide and Agricultural Bioengineering, Ministry of Education, Guizhou University, Guiyang 550025, China

⁴The Engineering Research Center of Southwest Bio-Pharmaceutical Resources, Ministry of Education, Guizhou University, Guiyang 550025, China

Zeng XY, Tan TJ, Tian FH, Wang Y, Wen TC 2023 – OFPT: a one-stop software for fungal phylogeny. Mycosphere 14(1), 1730–1741, Doi 10.5943/mycosphere/14/1/20

Abstract

DNA-based phylogeny is indispensable in modern fungal taxonomy. However, the procedure of conducting phylogenetic analysis is pretty complicated, and mistakes are commonplace leading to results among different researchers with low reliability and reproducibility. On top of that, obtaining robust phylogenetic trees is also very time-consuming, which may cost researchers lots of time performing mostly repetitive operations. There is software offering functions to perform part or complete procedure of the phylogenetic analysis, but they are either too multi-functional or have limited functions. Therefore, we developed a click-and-run software named ‘One-click Fungal Phylogenetic Tool’ (<https://ofpt.guhongxin.com>), which automates and standardizes the repetitive process of constructing phylogenetic trees and does not require any additional operations while running. This software offers users a step-by-step process and requires only an Excel file with DNA sequence information to automatically generate phylogenetic trees, which will improve the efficiency of phylogenetic study and the reliability of resulting phylogenetic trees.

Keywords – Bayesian inference – maximum likelihood – Python – sequence data – standardized process

Introduction

DNA-based phylogeny is an essential mean of current fungal taxonomic studies, especially for groups with limited morphological characteristics. However, conducting phylogenetic analysis comprises a complicated and time-consuming procedure, including retrieving sequence data from GenBank, sequence alignment and trimming, selecting the best-fit evolutionary model, sequence concatenation, converting sequence format with specific commands where necessary, and tree editing (Dissanayake et al. 2020, Zhang et al. 2020). Each individual who does phylogenetic analysis only focuses on a limit number of fungal groups, where parameters for relevant analysis are mostly the same. Therefore, many researchers have been spending a lot of time performing a large number of repetitive operations, which greatly reduces the efficiency of scientific researches. On the other hand, such analytical operations require researchers to have certain background knowledge of phylogeny, and be familiar with relevant websites and software. One common problem during downloading sequence is the duplication and misspelling of accession numbers, resulting in failure of downloading some DNA sequences. Another common mistake is on the model test, where researchers either do not test for the best-fit models, or they cannot apply them

correctly, especially for Bayesian inference. Hence, the reliability and reproducibility of the phylogenetic trees are always questionable among different researchers.

Software is the basis of phylogenetic studies, as conducting phylogenetic analysis requires lots of mathematical calculations. There are several specific software and websites developed for every step of phylogenetic analysis. To get start with, one will need to access the NCBI database and request the information by entering accession numbers to download reference sequence data. Afterwards, homologous regions should be aligned based on different algorithms using programs such as MAFFT (Kato et al. 2002), MUSLE (Edgar 2004), ClustalW (Thompson et al. 2003), whereas those poorly aligned regions are then needed to be excluded using TrimAl (Capella-Gutiérrez et al. 2009) or GBlocks (Talavera & Castresana 2007). For multi-locus phylogenetic analysis, SequenceMatrix is a good choice to concatenate different datasets into a single one (Vaidya et al. 2011). The best-fit nucleotide substitution models will be necessary if the phylogenetic tree is generated based on maximum likelihood (ML) and Bayesian inference (BI), and this can be tested and selected using ModelFinder (Kalyaanamoorthy et al. 2017), jModeltest (Posada 2008), or MrModeltest (Nylander 2008). Before constructing phylogenetic trees, the sequence alignment must be converted into the required format regarding to the software. The website ALTER (<https://sing.ei.uvigo.es/ALTER/>), as well as some software such as Aliview (Larsson 2014), MEGA (Kumar et al. 1994) and Bioedit (Hall 1999), provides the function to convert between different formats of the alignment. After the generation of phylogenetic trees, it can be then visualized using FigTree (Rambaut 2016) or iTOL website (Letunic & Bork 2019)

Since phylogenetic analysis is conducted based on computer programs, there must be a way to integrate and simplify the whole process into a single software rather than switching between different ones. Some software or websites, such as Phylogeny.fr (Dereeper et al. 2008), phyloGenerator (Pearse & Purvis 2013), Osiris (Oakley et al. 2014) and PhyloSuite (Zhang et al. 2020), can perform some or all of the above steps well, yet they are often designed to meet the requirements of researchers from various fields. That will often result in a multi-functional user interface, which still requires user interventions between steps. This is somehow less effective when someone wants to test different datasets with the same parameters, and wrong operations for both choosing files and setting parameters may also lead to erratic results. Moreover, there is currently no software or program specifically developed for fungal phylogeny.

Based on the above drawbacks, a click-and-run software named One-click Fungal Phylogenetic Tool (OFPT) is developed herewith, specifically for those who will conduct fungal phylogenetic analyses. It provides a step-by-step guide for users and does not requires any additional operations while running. After setting and preparing the required Excel file, users can just simply click and run the software repetitively without any extra manual process.

Materials & Methods

OFPT is written and packaged by Python 3.9. It integrates eight modules corresponding to the process of doing phylogenetic analysis into a pipelined process, viz. sequence downloading, sequence alignment, alignment trimming, alignment concatenating, model selection, format conversion, construction of ML tree, and construction of Bayesian tree (Fig. 1). These modules can be executed separately or in sequence with the above orders.

Sequence downloading

The program will first detect if only a single Excel file (.xls or .xlsx) is included in the working directory and then read the accession numbers from the 1st sheet of the Excel file. The 2nd sheet will be detected simultaneously. Request of downloading will then be sent to NCBI (<https://www.ncbi.nlm.nih.gov/>), and sequences in FASTA format will be downloaded and renamed according to the content in the first two columns. When finished, the program will verify if the number of entries in the Excel sheet is identical to the number of sequences downloaded from GenBank. An error message will appear if some of the listed accession numbers were not downloaded successfully. In that case, the user must correct the mistakes manually, or use the

function ‘check’ to allow the program to find the problematic entries. Afterwards, sequences (FASTA format) in the 2nd sheet, if any, will be matched based on their gene names (the header of each column) and added to the corresponding dataset accordingly. In addition, the program will detect the last two characters of the header, which represents the gene name, and check if they are digits matching the option of aligning strategy and trimming mode. If so, the program will treat these two digits as a customized option for aligning and trimming each genomic region. Finally, ‘*.fasta’ files for each gene region will be generated separately in the folder <in> to restore the sequence data.

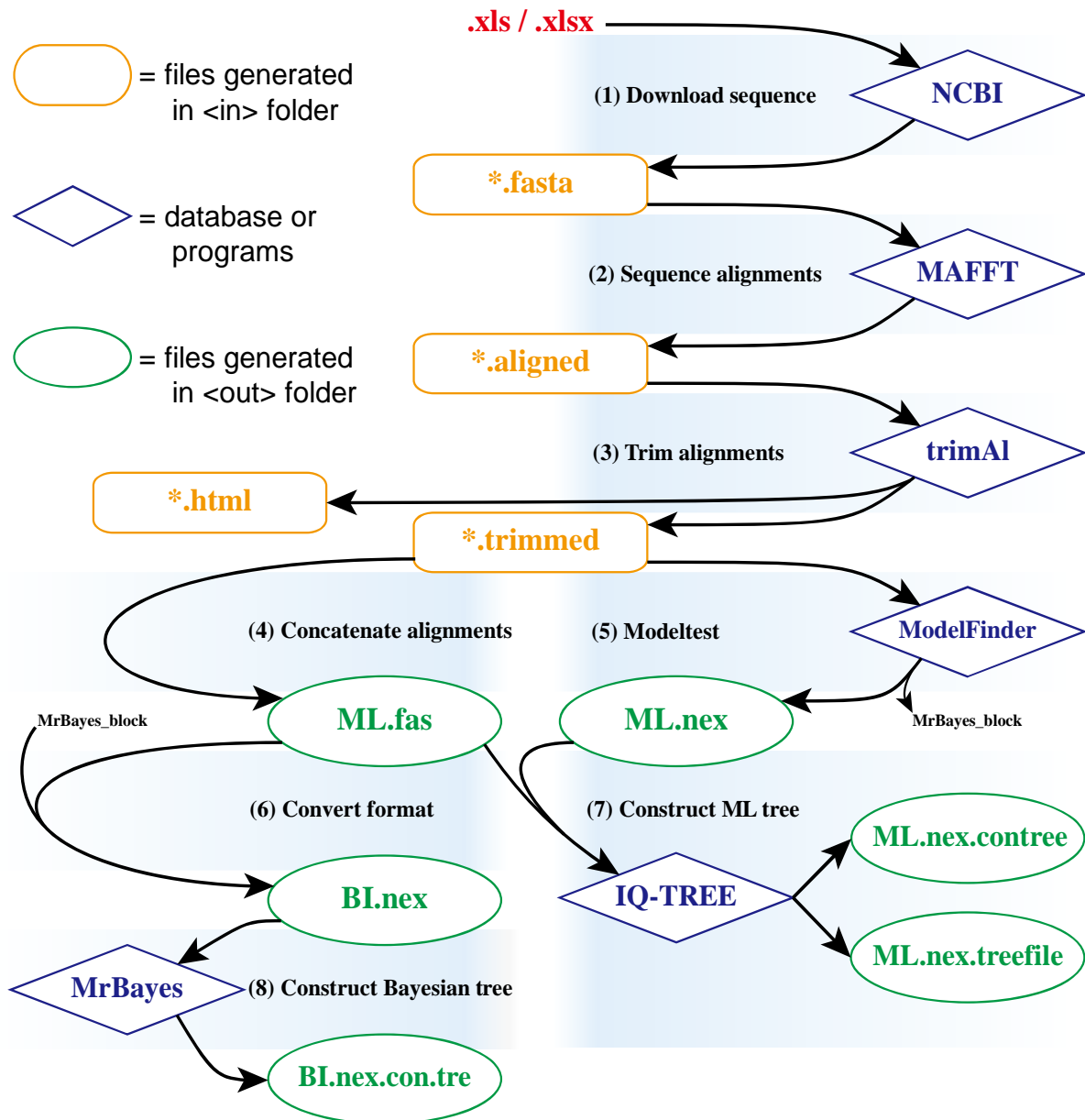


Figure 1 – The main workflow of OFPT.

Sequence alignments

Files with the ‘.fasta’ extension in the folder <in> will be detected and proceeded for alignment using MAFFT (Kato & Standley 2013). The global aligning strategy of commonly used ‘auto’, ‘FFT-NS-i’, ‘E-INS-i’, ‘L-INS-i’ and ‘G-INS-i’, can be selected when setting the configurations. If two digits at the end of the gene name have been detected in the downloading process, the aligning strategy will be set according to the first digit. Aligned sequence datasets will finally be written in ‘*.aligned’ file(s) separately in the folder <in>.

Alignment trimming

Files with the ‘.aligned’ extension in the folder <in> will be detected and proceeded for the alignment trimming using trimAl (Capella-Gutiérrez et al. 2009). The global trimming mode of commonly used ‘-gappyout’, ‘-strict’, ‘-strictplus’, ‘-gt 0.2–0.8’ can be selected when setting. If two digits at the end of the gene name have been detected in the downloading process, the trimming mode will be set according to the second digit. Being trimmed datasets will finally be written in ‘*.trimmed’ files separately in the folder <in>, and ‘*.html’ files demonstrating the excluded regions will be generated in the folder <in>.

Alignment concatenating

Files with the ‘.trimmed’ extension in the folder <in> will be detected and proceeded for concatenating. Sequences with the same ID will be concatenated in the order of file names. Sequences of each missing ID will be filled with question marks according to the length of corresponding gene regions. The concatenated dataset will finally be written into a single ‘ML.fas’ file in the folder <out>. The ‘ML.fas’ file can be directly used as the input file of IQ-TREE (Nguyen et al. 2015) and RAxML (Stamatakis 2014).

Model selection

Files with the ‘.trimmed’ extension in the folder <in> will be detected and proceeded for testing best-fit models using ModelFinder (Kalyaanamoorthy et al. 2017). Then the program will delete and recreate the <temp> folder before running this module to avoid errors caused by previously generated files, especially when users changed the data in the Excel file and rerun the program with the same working directory. Five files will be generated for each gene dataset in the folder <temp> including a ‘*.trimmed.iqtree’ file, which records the detailed results of the model test. The ‘ML.nex’ file will be generated in the folder <out> as one of the input file for IQ-TREE, which list the information of each gene partition and their corresponding model selected by the Bayesian information criterion (BIC). The command block for MrBayes will also be written in the ‘MrBayes_block’ file in the folder <temp>, including the commands for applying different models to different partitions. As some of the models are not supported in MrBayes, ‘nst=Mixed’ will be used when the best-fit model is not GTR/SYM, HKY/K80 or F81/JC (Huelsenbeck et al. 2004).

Format conversion

The ‘ML.fas’ file in the folder <out> will be detected and converted from FASTA to NEXUS format, and add the content of the ‘MrBayes_block’ file in the folder <temp> to the end. The converted dataset will finally be written into a ‘BI.nex’ file in the folder <out>, which can be directly used as the input file for MrBayes. The ‘BI.nex’ file can be directly used as the input file of MrBayes.

Construction of ML tree

The ‘ML.fas’ and ‘ML.nex’ file in the folder <out> will be detected and proceed for maximum likelihood (ML) tree construction using IQ-TREE (Nguyen et al. 2015), performing an ultra-fast bootstrap with 1000 replicates (Hoang et al. 2018). The program will read the partition and model information in the ‘ML.nex’ file to perform the analysis. The consensus ML tree with bootstrap supports will be generated in the folder <out> namely ‘ML.nex.contree’, and the ML tree with both bootstrap supports and SH-aLRT (SH-like approximate likelihood ratio test) result will be generated in the folder <out> namely ‘ML.nex.treefile’ file.

Construction of Bayesian tree

The ‘BI.nex’ file in the folder <out> will be detected and proceed for Bayesian inference (BI) using MrBayes (Ronquist et al. 2012). The default Markov chain Monte Carlo analysis will be performed with two parallel runs, each including four chains and 100 sampling frequencies with 1,000 diagnostic frequencies and 10,000 checkpoint frequencies. The runs will stop automatically

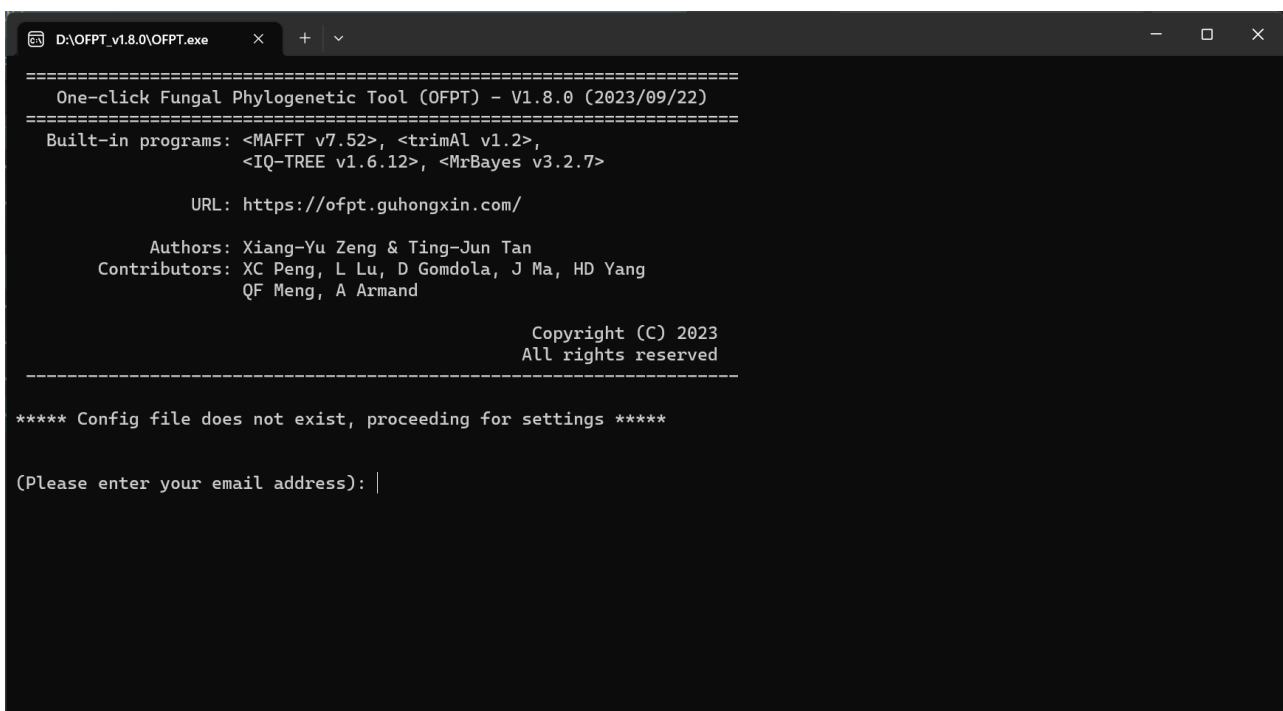
when the standard deviation falls below 0.01 or the generation number reaches 50,000,000 (can be changed in the MrBayes block manually). The first 25% sample fraction will be discarded when calculating convergence diagnostics. The consensus BI tree with posterior probabilities will be generated in the folder <out> namely 'BI.nex.con.tre'. This function is not recommended to use on PC (personal computer), as it takes a long time and will consume lots of CPU (Central Processing Unit) resources.

Results

OFPT is a streamlined software for conducting fungal phylogenetic analyses, with built-in programs of MAFFT, trimAl, IQ-TREE and MrBayes. It can be downloaded at the website <https://ofpt.guhongxin.com> and is currently supported for Windows operating system of version 10 or above. The software automates and standardizes the repetitive process of constructing phylogenetic trees, which will make relevant studies much easier and faster, and its resulting trees are identical to the manually generated ones. OFPT is simple and straightforward, offering users a step-by-step process, and its workflow is illustrated in Fig. 1.

Download and run

To use the software, you will first need to go to the url: <https://ofpt.guhongxin.com/>, click 'Download the newest version' button in the center, and a file named 'OFPT_v*.rar' (the asterisk represents the version number) will be downloaded to your computer. Then, you will need to extract the file, using 'WinRAR' or other compressing software, to a path without space or other illegal characters. Finally, you can double click the file 'OFPT.exe' to run the software.



```
D:\OFPT_v1.8.0\OFPT.exe x + v
=====
One-click Fungal Phylogenetic Tool (OFPT) - V1.8.0 (2023/09/22)
=====
Built-in programs: <MAFFT v7.52>, <trimAl v1.2>,
                  <IQ-TREE v1.6.12>, <MrBayes v3.2.7>

URL: https://ofpt.guhongxin.com/

Authors: Xiang-Yu Zeng & Ting-Jun Tan
Contributors: XC Peng, L Lu, D Gombola, J Ma, HD Yang
              QF Meng, A Armand

                          Copyright (C) 2023
                          All rights reserved
=====
**** Config file does not exist, proceeding for settings ****

(Please enter your email address): |
```

Figure 2 – The interface when running the software for the first time.

Using the software for the first time

When first used, the user will be asked to enter an email address (Fig. 2), set the parameters for aligning and trimming (Fig. 3), and determine the working directory (the absolute path of a folder containing target Excel file). Afterwards, a file named 'Config.ini' will be generated to record the setting information. The parameters can be changed afterwards by pressing 's' in the main interface.

```

D:\OFPT_v1.8.0\OFPT.exe
-----XX-----XXXX-----XXXX-----
3. L-INS-i - is suitable for alignments like this:
ooooooooooooooooooooooooooooooooXXXXXXXXXX-XXXXXXXXXXXXXXXX-----
-----XX-XXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXoooooooooooooooo-----
-----ooooooooooooooooXXXXXXXX-XXXXXXXXXX-XXXXXXXXXXXXXXXX-----
-----ooooooooooooooooooooooooXXXXXXXX-XXXXXXXXXX-XXXXXXXXXXXXXXXX-----
-----XXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXX-----
4. G-INS-i - is suitable for alignments like this:
XXXXXXXXXX-XXXXXXXXXXXXXXXX
XX-XXXXXXXXXXXXXXXX-XXXXXXXX
XXXXX-XXXXXXX-XXXXXXX
XXXXX-XXXXXXXXXX-XXXXXXX
XXXXXXXXXXXXXXXX-XXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXX
=====
(Please select the global aligning strategy): 0
=====
0. gappyout - trim based on gaps' distribution (default)
1. strict - trim based on the fraction of gaps in a column and their similarity scores
2. strictplus - more stringent than strict
3. -gt 0.2 - allow 20% fraction of sequences to have gaps
4. -gt 0.3 - allow 30% fraction of sequences to have gaps
5. -gt 0.4 - allow 40% fraction of sequences to have gaps
6. -gt 0.5 - allow 50% fraction of sequences to have gaps
7. -gt 0.6 - allow 60% fraction of sequences to have gaps
8. -gt 0.7 - allow 70% fraction of sequences to have gaps
9. -gt 0.8 - allow 80% fraction of sequences to have gaps
=====
(Please select the global trimming method): 0

```

Figure 3 – The interface for choosing alignment strategy and trimming method.

Main interface

If the ‘Config.ini’ file exists, the main interface of the software will appear (Fig. 4). Users can run the 8 different modules individually or sequentially according to their demands. This can be either a single number, or a series of numbers, such as ‘1234’ or ‘1234567’. After pressing ‘ENTER’, the software will run the selected modules in sequence without human interventions. When all selected tasks are completed, the main interface will appear again with the message ‘All tasks finished’ showing above (Fig. 5). Information of each run will be recorded in the ‘Information.txt’ file in the <out> folder once all tasks are finished (Fig. 6).

```

D:\OFPT_v1.8.0\OFPT.exe
URL: https://ofpt.guhongxin.com/
Authors: Xiang-Yu Zeng & Ting-Jun Tan
Contributors: XC Peng, L Lu, D Gombdola, J Ma, HD Yang
              QF Meng, A Armand
Copyright (C) 2023
All rights reserved
-----
* * * * *
* (H) elp          /|||\  |||| /|||\  |||||  *
* (S) ettings     ||  ||  ||||  ||||/  ||  \ /  ||  ||_||  *
* (C) heck        \|||/  ||  ||  |||/  ||  \ /  ||  ||_||  *
* (1) Download sequence - requires a single '*.xls' or '*.xlsx' file *
*                   in the working directory *
* (2) Sequence alignments - requires '*.fasta' file(s) in <in> *
* (3) Trim alignments - requires '*.aligned' file(s) in <in> *
* (4) Concatenate alignments - requires '*.trimmed' file(s) in <in> *
* (5) Modeltest - requires '*.trimmed' file(s) in <in> *
* (6) Convert format - requires 'ML.fas' file in <out> and will *
*                   read 'MrBayes_block' file in <temp> *
* (7) Construct ML tree - requires 'ML.fas' & 'ML.nex' file in <out> *
* (8) Construct Bayesian tree - requires 'BI.nex' file in <out> *
* * * * *
<Current working directory>: D:\test
=====
Please select the function(s) you would like to proceed (e.g. 1234567 or 1234): 1234567

```

Figure 4 – The main interface of the software.

```

Ultrafast bootstrap approximation results written to:
Split support values:      D:\test\temp\ML.nex.splits.nex
Consensus tree:           D:\test\temp\ML.nex.contree
Screen log file:          D:\test\temp\ML.nex.log

Date and Time: Thu Sep 21 12:34:01 2023
<ML tree complete>

||||||| All tasks finished |||||||

* * * * *
* (H) elp      /||| \  |||| /||| \  ||||
* (S) ettings  ||  ||  |||  |||/  ||  \ /  ||  ||_||
* (C) heck    \|||/  ||  ||  \  /  ||  ||_||
* (1) Download sequence - requires a single '*.xls' or '*.xlsx' file
*              in the working directory
* (2) Sequence alignments - requires '*.fasta' file(s) in <in>
* (3) Trim alignments - requires '*.aligned' file(s) in <in>
* (4) Concatenate alignments - requires '*.trimmed' file(s) in <in>
* (5) Modeltest - requires '*.trimmed' file(s) in <in>
* (6) Convert format - requires 'ML.fas' file in <out> and will
*              read 'MrBayes_block' file in <temp>
* (7) Construct ML tree - requires 'ML.fas' & 'ML.nex' file in <out>
* (8) Construct Bayesian tree - requires 'BI.nex' file in <out>
* * * * *
=====
<Current working directory>: D:\test
=====
Please select the function(s) you would like to proceed (e.g. 1234567 or 1234):

```

Figure 5 – The interface when finishing all tasks.

Data preparation

Before starting to use the software, users only need to put a single Excel file under the working directory, with the 1st sheet providing names and accession numbers of the datasets and the 2nd sheet with your own sequence data in FASTA format. In the 1st sheet, species name and strain code of each entry are required in the first two columns, which will then be treated as the sequence ID by the program. The rest columns should include accession numbers of each strain, with their corresponding gene name in the header of each column. If you would like to add reference after each entry, the corresponding column should be named as ‘Reference(s)’. In the 2nd sheet, the header of each columns should be identical to the gene name included in the 1st sheet, otherwise your own data will not be included to the corresponding dataset. The FASTA format of users’ sequence data should be placed into two cells, where sequence ID followed by the cell including DNA sequences. The template of the input Excel file is illustrated in Table 1, as well as in the ‘example.xlsx’ file under the directory <examples>.

Table 1 Template Excel file for using the software.

Species	Strain	Gene1	Gene2	Gene3	Gene2	Gene3
sp.1	st.1	ac. no. 1	ac. no. 2		>id1	>id2
sp.2	st.2	ac. no. 3		ac. no. 4	seq1	seq2
...		
Sheet1					Sheet2	

Requirement of input files

1. Downloading sequence requires a single ‘*.xls’ or ‘*.xlsx’ file in the working directory.
2. Sequence alignments requires ‘*.fasta’ file(s) in the folder <in>.
3. Trimming alignments requires ‘*.aligned’ file(s) in the folder <in>.
4. Concatenating alignments requires ‘*.trimmed’ file(s) in the folder <in>.
5. Modeltest requires ‘*.trimmed’ file(s) in the folder <in>.
6. Converting format requires ‘ML.fas’ file in the folder <out> and will read the ‘MrBayes_block’ file in the folder <temp>.

7. Constructing a ML tree requires ‘ML.fas’ and ‘ML.nex’ file in the folder <out>.
8. Constructing a Bayesian tree requires ‘BI.nex’ file in the folder <out>.

```

CPU: Intel64 Family 6 Model 186 Stepping 2, GenuineIntel
OS: Windows 10.0.22621
[Start: 2023-09-21 13:02:05]

=====
(1) Numbers of sequences
=====
ITS : 7
RPB2 : 8
TEF : 7

<Running time: 5.1105s>

=====
(2) Aligning strategies
=====
ITS : auto
RPB2 : FFT-NS-i
TEF : G-INS-i

<Running time: 5.1591s>

=====
(3) Trimming modes
=====
ITS : gaps < 20%
RPB2 : strictplus
TEF : gaps < 40%

<Running time: 0.1099s>

=====
(4) Concatenating information
=====
ITS : 1 - 627
RPB2 : 628 - 1690
TEF : 1691 - 2961

<Running time: 0.0147s>

=====
(5) Modeltest
=====
Twenty-two common DNA substitution models with rate heterogeneity were tested by ModelFinder
(Kalyaanamoorthy et al. 2017). The best-fit model for each gene selected by Bayesian information criterion
(BIC) is as follow:
ITS : K2P+I
RPB2 : TNe+G4
TEF : TN+F+I

```

Figure 6 – The log information generated in the ‘Information.txt’ file.

Production of output files

1. Downloading sequence will generate ‘*.fasta’ file(s) in the folder <in>, which comprises the dataset including downloaded sequences and users’ sequences (if any) of each gene region.
2. Sequence alignment will generate ‘*.aligned’ file(s) in the folder <in>, which comprises aligned sequence dataset of each gene region.
3. Trimming alignments will generate ‘*.trimmed’ and ‘*.html’ file(s) in the folder <in>, comprising optimized alignments and details of excluded regions of each gene region, respectively.
4. Concatenating alignments will generate ‘ML.fas’ file in the folder <out>, , which comprises the concatenated multi-gene dataset.

5. Modeltest will generate 'ML.nex' file in the folder <out>, 'MrBayes_block' file in the folder <temp> and other intermediate files in the folder <temp>. The 'ML.nex' file is one of the input file for conducting ML analysis, which comprises partition and model information of the concatenated dataset.

6. Converting sequence will generate the 'BI.nex' file in the folder <out>, which is the input file for running MrBayes.

7. Constructing ML tree using IQ-TREE will generate 'ML.nex.conf' and 'ML.nex.treefile' file in the folder <out>, which are the output tree files from ML analysis.

8. Constructing a Bayesian tree using MrBayes will generate the 'BI.nex.con.tre' file in the folder <out>, which is the output tree file from BI.

Notice

1. If the anti-virus software reports a virus, please add the software to the trust list or temporarily disable the anti-virus software.

2. Paths of both the software and the working directory MUST NOT include SPACE or any other illegal characters.

3. When an error occurs, the software will prompt an error message and quit the current process.

4. Only a single Excel file should be included in the working directory.

5. Content in the first two columns of the Excel file will be treated as a sequence ID, information of accession numbers should be included in the other columns.

6. The missing accession number in the Excel file should be represented by an empty cell, a single hyphen, or a single en-dash.

7. Gene names (header of columns) in the Excel file should only include alphabets, numbers and underscores.

8. If users want to apply different strategies and modes for aligning and trimming to different genes, please directly add two digits representing aligning strategy and trimming mode together at the end of the headers (e.g. ITS10, TEF101).

9. The column header representing the same gene in sheets 1 and 2 should always be identical.

10. The sequence ID of the same strain the users provided should be identical, in order to avoid errors when concatenating multi-genes.

11. Please ensure all sequences are forward before aligning, as the function of adjusting sequence direction are not available in the software.

12. Running the model test will delete <temp> folder.

13. Details of the model test results can be found in '*.iqtree' files in <temp> folder.

14. Outgroups will not be assigned in the MrBayes command block.

15. Running 'convert sequence format' alone will not generate the MrBayes command block, unless a file named 'MrBayes_block' already exists in <temp> folder.

16. If an unexpected error occurs, the software window may automatically shut down while running. In this case, users can run the software from the Windows command prompt to check the exact errors during execution. First, press 'win + R' and type 'cmd' to open the command prompt. Then, type 'cd <path of OFPT>' to enter the path of the software. Finally, type the file name 'OFPT.exe' to run the software.

Discussion

There are useful software and platforms developed for conducting phylogenetic analysis. Phylogeny.fr is an online platform designed for researchers with little knowledge in phylogeny (Dereeper et al. 2008). It offers the 'one-click mode' covering steps from sequence aligning to tree rendering, where all parameters are set up to suit most studies. However, the platform lacks the module of downloading sequence from GenBank and testing best-fit models for analyses of ML or BI. phyloGenerator is a command-line application designed for ecologists, which allows

nonspecialists to produce robust and repeatable phylogenetic trees without requiring an extensive knowledge (Pearse & Purvis 2013). However, it lacks model testing and BI function and can only download the sequence data according to the name of organisms, which is error-prone for fungi. Osiris is multi-functional tool developed based on the workflow of Galaxy (Goecks et al. 2010), a web-based analysis platform (Oakley et al. 2014), which provide a set of functions covering all steps for phylogenetic analysis. However, it is too multi-functioned and is quite difficult to use. PhyloSuite is a powerful software recently developed for conducting phylogenetic analysis (Zhang et al. 2020). It possesses all necessary step for phylogenetic analysis and can manage the data itself. However, this multi-functional interface will lead to various operations by users, resulting in a non-standardized process, and the operation between each step is still time-consuming.

OFPT is developed to simplify and standardized the process for conducting phylogenetic analysis, specifically for fungal taxonomists, which integrates functions of sequence downloading, renaming, aligning, trimming, concatenating, model testing, format converting, Bayesian command block writing and tree construction in a single, streamlined software. It differs from the above software in its preset parameters, such as sequence ID, aligning strategy, trimming method, model selecting criteria, and parameters for running MrBayes, which does not require any intervention while running. Only an Excel sheet comprising sequence ID with corresponding accession number of each gene region, as well as users' own sequence data (if any), is needed to generate the phylogenetic trees.

The software has the following advantages:

1. It is a time-saving software that is easy to use. Users just need to click and run the software to finish the whole procedure, saving much time on handling this repeated work. It takes only 2 min, 4 min and 10 min to generate the ML tree from a 48×2391 (three partitions), 111×4660 (four partitions) and 122×5507 (five partitions) data matrix, respectively. Besides, those unfamiliar with phylogeny can also use it to generate reliable trees or datasets.

2. The prepared datasets are reproducible. The whole process can be reproduced with the same parameters, and errors in copy-paste, parameter setting, and MrBayes command writing can be avoided. Therefore, different researchers can generate the same tree using the same dataset to compare and justify results from various publications.

3. Common mistakes can be avoided by using the software. The software can check the included accession numbers itself, which can make sure all of the included data are valid and can be successfully downloaded. It can also apply the best-fit model information to the input file of ML and BI analysis automatically, to avoid incorrect interpretation by the program of IQ-TREE and MrBayes.

4. The user-prepared Excel file can be directly added to their manuscript, as it shares the same format required by most journals. As the major procedures of phylogenetic analysis applied in the software is fixed, users can also simply cite the software in the material and methods part to describe the way they obtained the phylogenetic tree without the necessity to draft them in details. This helps those with less experience in conducting phylogenetic analysis.

5. It is the foundation of future standardized phylogenetic analyses. As long as a backbone dataset and relevant parameters are settled for a specific group, the phylogenetic relationship of newly collected species can be easily determined. This will avoid the inclusion of improper data and ensure phylogenetic trees being generated based on robust dataset.

The current version of the software is limited to the function of phylogenetic tree construction. However, in version 2, we will introduce standard sequence datasets for some typical fungal groups, so those who conduct phylogenetic analysis on these groups will no longer spend time on preparing such an Excel sheet for published data. In the future, standard morphological characteristics of specific groups will also be digitized, so both morphological and phylogenetic identifications will be much easier and faster. By that time, fungal taxonomic studies may step into a new digital era.

Acknowledgements

This work was supported by Guizhou Provincial Science and Technology Project QKH [2021] General 199; the National Natural Science Foundation of China (No. 32060012); Guizhou Provincial Science and Technology Projects ZK[2023] general 087; Natural Science Special Research Fund of Guizhou University, Special Post 2021(25); China Scholarship Council [2021]15 no. 202108520071. The authors would like to thank XC Peng, L Lu, D Gomdola, J Ma, HD Yang, QF Meng and A Armand for their feedback and suggestions.

References

- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009 – trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15), 1972–1973.
- Dereeper A, Guignon V, Blanc G, Audic S et al. 2008 – Phylogeny. fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research* 36(suppl_2), W465–W469.
- Dissanayake AJ, Bhunjun CS, Maharachchikumbura SSN, Liu JK. 2020 – Applied aspects of methods to infer phylogenetic relationships amongst fungi. *Mycosphere* 11(1), 2652–2676.
- Edgar RC. 2004 – MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5), 1792–1797.
- Goecks J, Nekrutenko A, Taylor J. 2010 – Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* 11, 1–13.
- Hall TA. 1999 – BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41, 95–98.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ et al. 2018 – UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35(2), 518–522.
- Huelsenbeck JP, Larget B, Alfaro ME. 2004 – Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Molecular Biology and Evolution* 21(6), 1123–1133.
- Kalyanamoorthy S, Minh BQ, Wong TKF, von Haeseler A et al. 2017 – ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14(6), 587–589.
- Katoh K, Misawa K, Kuma Ki, Miyata T. 2002 – MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30(14), 3059–3066.
- Katoh K, Standley DM. 2013 – MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30(4), 772–780.
- Kumar S, Tamura K, Nei M. 1994 – MEGA: molecular evolutionary genetics analysis software for microcomputers. *Bioinformatics* 10(2), 189–191.
- Larsson A. 2014 – AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30(22), 3276–3278.
- Letunic I, Bork P. 2019 – Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* 47(W1), W256–W259.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015 – IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32(1), 268–274.
- Nylander JAA. 2008 – MrModeltest v2.3. Department of Systematic Zoology, Uppsala University, Uppsala, Sweden.
- Oakley TH, Alexandrou MA, Ngo R, Pankey MS et al. 2014 – Osiris: accessible and reproducible phylogenetic and phylogenomic analyses within the Galaxy workflow management system. *BMC Bioinformatics* 15(1), 230.

- Pearse WD, Purvis A. 2013 – phyloGenerator: an automated phylogeny generation tool for ecologists. *Methods in Ecology and Evolution* 4(7), 692–698.
- Posada D. 2008 – jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 25(7), 1253–1256.
- Rambaut A – FigTree version 1.4.3. <https://github.com/rambaut/figtree/releases/tag/v1.4.3>
- Ronquist F, Teslenko M, van der Mark P, Ayres DL et al. 2012 – MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61(3), 539–542.
- Stamatakis A. 2014 – RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9), 1312–1313.
- Talavera G, Castresana J. 2007 – Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* 56(4), 564–577.
- Thompson JD, Gibson TJ, Higgins DG. 2003 – Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics* 00, 2.3.1–2.3.22.
- Vaidya G, Lohman DJ, Meier R. 2011 – SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27(2), 171–180.
- Zhang D, Gao F, Jakovli I, Zou H et al. 2020 – PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources* 20(1), 348–355.