Mycosphere 11(1): 2652–2676 (2020) www.mycosphere.org ISSN 2077 7019
Article
Doi 10.5943/mycosphere/11/1/18

Applied aspects of methods to infer phylogenetic relationships amongst fungi

Dissanayake AJ[1], Bhunjun CS[2], Maharachchikumbura SSN[1], Liu JK[1*]

[1]*School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China*
[2]*Center of Excellence in Fungal Research, Mae Fah Luang University, Chiang Rai 57100, Thailand*

Dissanayake AJ, Bhunjun CS, Maharachchikumbura SSN, Liu JK 2020 – Applied aspects of methods to infer phylogenetic relationships amongst fungi. Mycosphere 11(1), 2652–2676, Doi 10.5943/mycosphere/11/1/18

Abstract

There is a need to document the methodologies used for molecular phylogenetic analyses since the current fungal identification, classification and phylogeny are necessarily applied with DNA molecular sequence data. Hence this manuscript is mainly aimed to provide a basic reference or guideline for the mycologists venturing into the field of phylogenetic studies and to avoid unnecessary repetitions in related publications. This is not at all to elaborate the theoretical background but is intended as a compact guide of some useful software references and coding for the most commonly used phylogenetic methods applied in mycological research. This reference manuscript is originally conceived for usage with a set of programs necessary for inference of phylogenetic analyses in fungal research (taxonomy and phylogeny). For this purpose, principles, introductory texts and formulas have been omitted. Meanwhile, necessary steps from DNA extraction to DNA sequencing, sequences quality check, data alignment and general steps of phylogenetic tree reconstructions have been included.

Keywords – Fungal classification – Materials and methodology – Molecular data – Reference – Taxonomy

Introduction

A 'classification system' refers to placing organisms within hierarchical groups (e.g. species, genus) and aggregating them to form supergroups of upper rank (e.g. family, order) forming a taxonomic grading (Judd et al. 2007, Simpson & Michael 2010, Liu et al. 2017). Based on shared characteristics, 'taxonomy' defines clusters of biological organisms by naming those groups (Judd et al. 2007, Simpson & Michael 2010). Taxonomy is usually based on phylogenetic relationships, placing similar taxa together by organizing them into monophyletic groups. Formerly, organisms were classified based on morphological comparisons as advanced molecular techniques were not available to support traditional species classification (Guarro et al. 1999). This artificial classification scheme led to taxonomic disagreements.

'Mycology' is a comparatively young scientific field which developed with the innovation of the microscope in the 17th century. The inspiring work of the mycological history can be exemplified by the findings in Pier Antonio Micheli's book 'Nova plantarum genera' in 1729 (Alexopoulos 1963) which were based on the observation of fungal spores. Hendrik Persoon

Submitted 17 September 2020, Accepted 20 November 2020, Published 9 December 2020
Corresponding Author: Jian-Kui Liu – e-mail – liujiankui@uestc.edu.cn

(1761–1836) arranged the leading classification of fungi (mushrooms) and is considered an originator of modern mycology beginning with the '*Synopsis methodica fungorum* (1801)' which is the starting point for the nomenclature of the Uredinales, Ustilaginales and the Gasteromycetes. Based on spore colour and microscopic appearances, Elias Magnus Fries (1794–1878) auxiliary expanded the fungal classification. Owing to progress in biotechnology, biochemistry, genetics and molecular biology, the 20th century has witnessed a modernization of mycology. The incorporation of DNA sequencing expertise and phylogenetic investigation has resulted in a better knowledge of fungal relationships together with biodiversity and has challenged traditional morphology-based groupings in fungal taxonomy. Combination of both morphology and molecular data developed a more trustworthy and natural classification structure that reveals true phylogenetic associations (Jeewon et al. 2002, 2003, Hyde et al. 2013, Phillips et al. 2013, 2019, Wijayawardene et al. 2014, 2016, 2020, Maharachchikumbura et al. 2015, 2016). Hence some taxonomists and most phylogeneticists prefer applying phylogenetic analyses combining the morphological data for classification purposes. Addressing species concepts through phylogenetic analysis is more reliable to identify micro-fungi, especially when morphological characteristics are similar or overlap. Even in these cases, the phylogeneticists do not avoid studying or comparing the morphology. The practice of 'divergence times' was recently proposed as a comprehensively standardized criterion for classifying entities (Avise & Johns 1999, Avise & Mitchell 2007, Zhao et al. 2016, Hyde et al. 2017, Liu et al. 2017) and this has been applied to certain types of fungi (Pang et al. 2013, Zhao et al. 2016, Liu et al. 2019, Zhang et al. 2019).

Phylogeny could be simply defined as the '*evolution of a genetically related group of organisms as distinguished from the development of each individual organism'*. Molecular phylogenetic trees illustrate the evolutionary relationships among groups of organisms inferred from nucleic acid or protein sequences (Hall 2013, 2017). This approach discovers the evolutionary relationships between individuals by evaluating the variations arising in different organisms and understanding the connections between an ancestral sequence and its descendants. The phylogenetic analysis is conducted based on hereditary DNA data, which reveal more about where an organism belongs taxonomically. This has helped to confirm species identification and place fungi in more natural groupings. Phylogenetic trees are simply composed of nodes and branches. One branch connects to its adjacent two nodes, representing taxonomic entities (e.g. genus, species). The three main methods of building phylogenetic trees are: 1) distance matrix methods, 2) validation methods and 3) character-based methods (Hall 2013). Distance matrix methods accept a molecular clock, implying that all mutations are neutral and therefore, they ensue a random clock-like rate (Felsenstein 2004, Mount 2004). The most commonly used distance matrix method is 'neighbour-joining (NJ)'. Validation methods denote to 'bootstrapping' and 'Jackknife estimation' approaches. Most frequently used character-based methods are 'maximum parsimony (MP)' and 'maximum likelihood (ML)' analyses. Maximum parsimony is an optimality criterion, hence more strictly one or more trees will be chosen with the fewest number of changes necessary to explain the character distribution observed in the data. This method is applicable for similar sequences or a group of sequences and is sensitive to inadequate taxon sampling, letting 'long branch attraction' (Kück et al. 2012). The method ML practices a tree model for nucleotide substitutions, this method will choose the 'most likely to observe the data' out of all the trees of the particular dataset. Bayesian inference (BI) is also a character-based method which employs a likelihood function to get the posterior probability of trees. The Markov Chain Monte Carlo (MCMC) algorithm samples from a probability distribution to provide a measure of precision. It is not a measure of variation in the data but rather how strongly a topology or set of typologies are supported. This is, of course, related to the variation in the data. Nonetheless, it is not a simple correlation. All these methods can merely offer appraisals of what a phylogenetic tree might look like for a given set of data. Most virtuous methods also provide a sign of how much variation there is in these estimates.

The progress of the 'polymerase chain reaction' (PCR; Bartlett & Stirling 2003) has modernized numerous areas of biology, containing mycological systematics (Schesser et al. 1991, Gargas & Taylor 1992, Valones et al. 2009) and this endorsed the expansion of molecular

applications in the field of phylogenetic inference. During last two decades, classification system of the kingdom fungi have been incorporated with molecular phylogenetic investigations of one or more genes (Hibbett et al. 2007, Nilsson et al. 2012a, 2014, Wijayawardene et al. 2020). The standard of applying phylogenies become popular (Leebens-Mack et al. 2006, Hyde et al. 2013), as many command-line programmes are now available for successful development and implementation of such phylogenetic software (Hall 2017). The major objective of the present manuscript is to provide a basic reference of step by step guide for molecular phylogenetic analyses. A case study of 'introducing a new species in the genus *Botryosphaeria'* has been provided to follow each step in MP, ML and BI analysis.

Basics of fungal molecular characterization

DNA extraction

Fungal tissues are used for DNA extraction. For culturable fungi, cultures derived from 'single spore isolation' is commonly used and for unculturable fungi, fruiting bodies present on the hosts are carefully checked and removed to subject for 'direct DNA sequencing'. Isolates obtained from single spore isolation are grown on culture media (MEA, PDA etc.). To extract DNA, mycelia are scraped from pure cultures. If the attempt of culture isolation failed, the fruiting bodies on the host are carefully surface sterilized before they undergo molecular treatments. Fungal DNA extraction can be done in many ways. Most commonly and widely used method is the CTAB (cetyl trimethyl ammonium bromide) based extraction buffers (Doyle & Doyle 1990). However, nowadays Genomic DNA Extraction Kits (e.g. QIAGEN GmbH, QIAGEN Strasse 1, BioFlux®) are commonly used.

PCR amplification

The PCR is used to make millions of copies of a target piece of DNA. It is an indispensable tool in modern molecular biology and has transformed scientific research producing enough DNA copies to be analyzed. Some of the common types of PCR are; Real-Time PCR (quantitative PCR or qPCR), Reverse-Transcriptase (RT-PCR), Multiplex PCR, Nested PCR, High Fidelity PCR, Fast PCR, Hot Start PCR and GC-Rich PCR. PCR amplification is carried out using relevant primers. The PCR is frequently conducted in a thermal cycler. PCR mixtures are commonly composed of *Ex-Taq* DNA polymerase (a robust PCR enzyme with proofreading activity), *Ex-Taq* DNA polymerase buffer, dNTPs (deoxyribonucleotide triphosphate), genomic DNA, primers and ddH_2O (double distilled water). The PCR conditions of most gene regions, as well as the primers depend on the genus, family or the order of the particular study. Most common gene regions and the conditions of primer pairs used in fungal studies can be found at (http://ulab360.com/files/prod/manuals/201310/09/441335001.pdf, http://lutzonilab.org/aftol/primers/, https://www2.clarku.edu/faculty/dhibbett/Protocols_Folder/Primers/Primers.pdf and the supplementary material of https://pubmed.ncbi.nlm.nih.gov/30442909/). Agarose gel electrophoresis is conducted to check the positive amplicons from a PCR. Products are visualized on an agarose gel comprised of stains under UV light using a molecular imaging system. The commonly used, but mutagenic, staining agent is ethidium bromide, but it is being replaced by a range of safer but more expensive alternatives. These safe stains vary in cost, sensitivity and the impedance of DNA as it migrates through the gel. The other common gel stains are GelRed™, GelGreen™, SYBR™ safe, SafeView and EZ-Vision®In-Gel Solution. The positive PCR products are sent to sequencing companies for sequencing.

Basic steps of a phylogenetic analysis

1) Sequence quality assurance

The generated sequences are visualized using various types of sequence visualizing software.

There are more than fifty enrolled software for this purpose (https://en.wikipedia.org/wiki/List_of_alignment_visualization_software). Chromatograms of the generated sequences are mostly checked for their quality by analyzing the sharpness of the peaks using BioEdit v.5 (Fig. 1) (Hall 1999) or AliView (Larsson 2014) to assure the sequence quality. BioEdit should recognize it as an ABI Autosequencer Trace file and open it as a chromatogram. The size of the chromatogram is adjusted tracing with the horizontal scale and vertical scale bars to the top left of the image to see the peaks of the trace. For sequences where both forward and reverse reads are available, the consensus sequences are obtained using DNASTAR Laser-gene. The generated sequences are matched to those in the GenBank with the BLAST tool (Basic Local Alignment Search Tool, https//:blast.ncbi.nlm.nih.gov/Blast.cgi). The worldwide largest public DNA sequence database is maintained by National Center for Biotechnology Information (NCBI; Sayers et al. 2020a). BLAST (Altschul et al. 1997) is a tool available in the NCBI website to search for similar sequences to a given sequence in the GenBank sequence database (Sayers et al. 2020b). Results of BLAST searches have to be interpreted with caution (Nilsson et al. 2012b). It is not recommended to rely solely on BLAST to identify fungi, as NCBI does not curate all the sequences and species names.

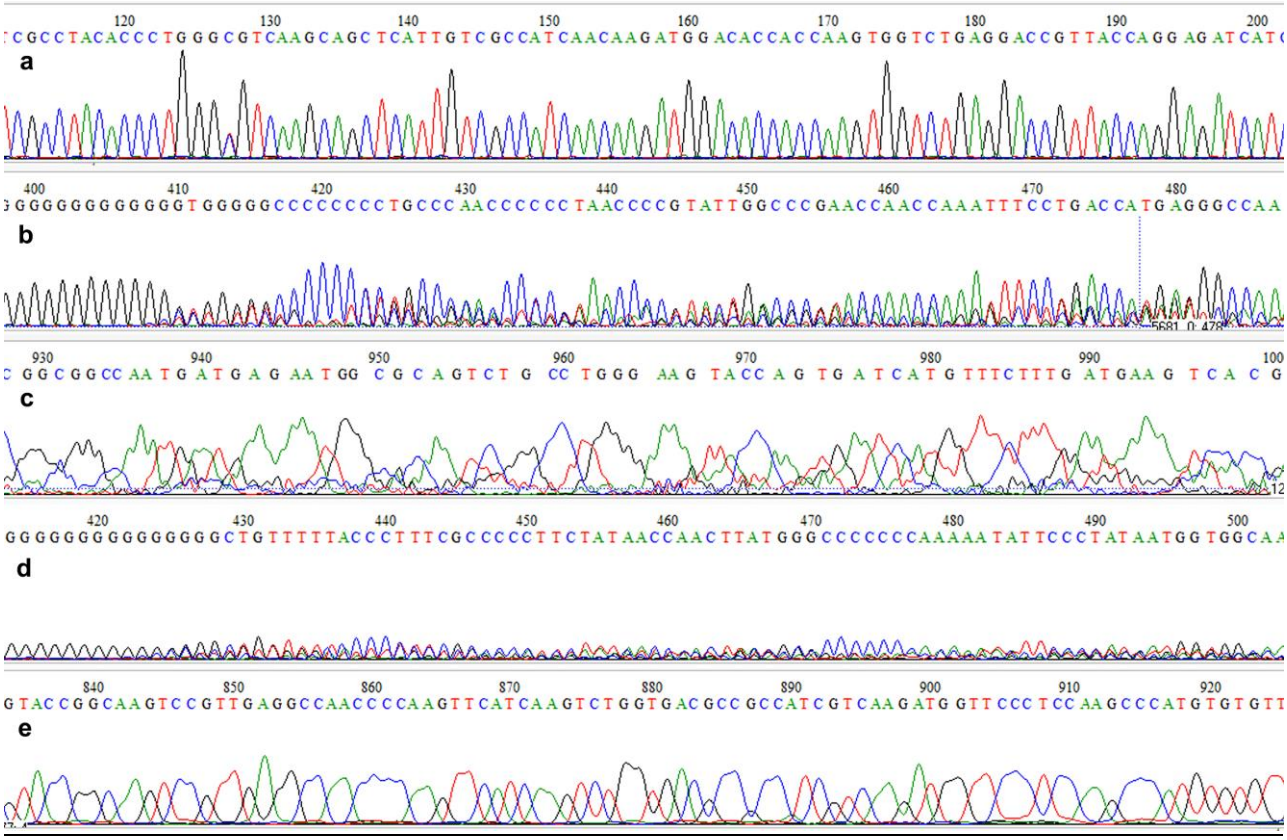

Figure 1 – Various types of chromatograms visualized in BioEdit. a Quality assured chromatogram. b–e Unqualified chromatograms.

BLAST search using mostly default parameters

(ftp://ftp.ncbi.nlm.nih.gov/pub/factsheets/HowTo_BLASTGuide.pdf).

The primary purpose of involving a BLAST search prior to a phylogenetic analysis is to rule out contamination as well as help identify members of gene families.

1. Navigate to the NCBI BLAST web server and click on “nucleotide blast”.

2. Click on ‘Browse’ and select the sequence file or paste the copied sequence directly into the Query box.

3. Enter a Job Title.

4. In the “Choose Search Set” section, change the database to ‘Nucleotide collection (nr/nt)’.
5. Under “Program Selection”, select ‘Highly similar sequences (blastn)’.

According to the BLAST guide (https://www.ncbi.nlm.nih.gov/blast/BLAST_guide.pdf), MEGABLAST is better at finding nucleotide sequences similar (but not identical) to your nucleotide query. The BLAST nucleotide algorithm finds similar sequences by breaking the query into short sub-sequences. If there is expected to be similar sequences available, it's better to go for the MEGABLAST option and set the maximum number of hits to 500, which increase the chances of finding out what it is have been sequenced.

6. Check the box “Show results in a new window” next to the “BLAST” button
7. Click “BLAST”

For initial species level confirmation, the internal transcribed spacer (ITS) region is commonly used. The ITS rDNA region is the formal barcode for fungi (Schoch et al. 2012) and it is capable of distinguishing species in most genera. Depending on the BLAST results of the ITS region, other relevant gene regions for the particular species/genus could be sequenced. The necessary gene regions (with proper primers) to be sequenced for each species/genus can be identified through relevant publications.

2) Determination of the extent of the trees to fit with the study

The particular study can be arranged after obtaining the BLAST results of the new sequences. Hence, the taxa range to be included in a phylogenetic study should be pre-organized to avoid the taxon sampling problem (Hillis et al. 2003).

i) Introducing a new species

For this, almost all required gene regions for the specific genus should be sequenced for the target isolates. The recommended primer pairs should be selected based on the literature. After obtaining the sequences, at least the top ten hits of BLAST results in NCBI database should be checked to have a proper idea about the isolates. However, it is recommended to include all type species and at least 2-3 strains from each species of the genus, when it comes to introducing a new species.

e.g. Introducing a novel species in the genus *Botryosphaeria*

At least ITS and TEF (translation elongation factor) regions should be sequenced (Phillips et al. 2013, Dissanayake et al. 2016). To date, there are fifteen type species (Chen et al. 2020), while Index Fungorum (http://www.indexfungorum.org/, accessed in October 2020) lists 283 names under the genus *Botryosphaeria*. The ‘type’ strains of each species (if they are available with molecular sequence data) should be included in the analysis when introducing a new species. For such a case, there’s no point in including sequence data from the particular family (Botryosphaeriaceae), the order (Botryosphaeriales) or the class (Dothideomycetes).

ii) Introducing a novel genus in a family

All type species of the genera in the particular family should be included in the phylogenetic analysis. Key species should be included to cover the phylogenetic breadth of each genus.

e.g. Introducing a new genus in Botryosphaeriaceae.

The LSU, ITS and TEF regions should be sequenced for the obtained isolates. All the genera in family Botryosphaeriaceae should be included for the phylogenetic analysis. If the new genus has shown phylogenetic exclusivity (monophyly) with respect to other genera, it should be introduced as a novel genus. Studies introducing new genera should limit the inclusion of taxa to the family level as it is not worth to include taxa at the order level.

iii) Revisiting a family/order/class

All available sequence data of the specific family/order/class should be incorporated in this

kind of study.

e.g. Reassessing the families in Botryosphaeriales

When assessing the higher taxonomic levels, the most popular molecular markers, such as ITS, LSU and SSU could be integrated to resolve the main lineages in the phylogenetic analysis. The sequences of representative genera and species in Botryosphaeriales could be retrieved from GenBank.

Estimation of divergence times emerged recently as universally accepted criterion for ranking higher taxa with additional evidence. Hence, maximum clade credibility (this has not been addressed in this study due to its particular parameters regarding the model, taxa, calibration points and divergence times) tree should be conducted in BEAST, so that each tree is weighted proportionally to its posterior probability.

3) Reference sequence, determination of outgroup and sequence alignment

Sequences can be searched and downloaded in several ways for phylogenetic analysis. The type or reference sequences can be downloaded using the search criteria of the NCBI, using the GenBank accession numbers (culture codes, species names could also use). Sequences could be saved in different formats in the available options of which the 'FASTA' format (Pearson & Lipman 1988) is generally favoured. The data downloaded from GenBank and the newly obtained, quality assured sequence data are added to the FASTA file together with the outgroup taxon/taxa (Fig. 2).

As a case study, we provide sequence data of *Botryosphaeria* species which were downloaded from GenBank together with sequence data of a new isolate (Fig. 2). This dataset will be processed for three major classes of phylogenetic analysis (MP, ML and BI) by providing details in each step (all the commands given here are compatible with the Windows versions).

Figure 2 – Inclusion of the sequence data in a FASTA file. a Outgroup taxon. b Species name. c Isolate code. d New isolate.

Selection of a proper outgroup

In phylogenetics, the outgroup taxon/taxa are the utmost indistinctly correlated entity or collection of entities that accommodate a reference cluster while defining the evolutionary connections within the ingroup (https://www.coursehero.com). The outgroup falls outside the clade being studied but is closely related to that clade (De Dreu et al. 2016) and acts as the comparison point of the ingroup and precisely permits for the phylogeny to be rooted. The choice of an outgroup is necessary as the direction of character change can be determined only on a rooted phylogeny, allowing to understand the evolution of traits along a phylogeny.

The evolutionary assumption of the outgroup entity has a shared progenitor with the ingroup, which is of age than the shared progenitor to the ingroup (Giribet & Ribera 1998). The root node can therefore be placed on the branch connecting the outgroup and ingroup. The selection of an outgroup might alter the topology of a phylogeny (De Dreu et al. 2016). Hence, the phylogeneticists generally add more outgroups in phylogenetic analysis. The inclusion of several outgroups is desirable as they offer an additional strong phylogeny, defending contrary to poor outgroup nominees and analysing the ingroup's postulated monophyly.

On the way to be an eligible outgroup, a taxon need content the below features https://en.wikipedia.org/wiki/Outgroup_(cladistics):

a) It need not to be an affiliate of the ingroup.
b) It needs to be associated to the ingroup, diligently sufficient for significant contrast (alignment) to the ingroup.
c) It should not partake taxonomic impediments (better to choose type strain) and it is essential to compose of all the gene regions used in the phylogeny.

Hence, an applicable outgroup needs to be explicitly free-standing to the clade of concern in the phylogenetic analysis (Maddison 1984). If an outgroup which is claded inside the ingroup is chosen to root the phylogeny, it effects in inappropriate assumptions of phylogenetic relationships and characteristic evolution (Maddison 1984). Yet, the finest level of affinity of the outgroup to the ingroup be subject to the distance of phylogenetic analysis. Selecting a strictly associated outgroup member to the ingroup is extra supportive once comparing the dissimilarities. Selecting an excessively distant outgroup may consequence in confounding convergent progression for a straight evolutionary connection owed to a mutual ancestor (Wilberg 2015, O'Brien et al. 2002). For trivial phylogenetics (e.g. resolving the evolutionary relationships of a clade within a genus) a fitting outgroup would be an associate of the sister clade. However, for deeper phylogenetic analysis, less closely associated taxa can be engaged (Jarvis 2014).

Rooted and unrooted trees

- A rooted tree is incorporated to create interpretations about the most common ancestor of the branches of the tree. Most frequently the root is denoted as 'Outgroup' (Taru & Shukla 2017).
- An unrooted tree is incorporated to create a graphic sketch of the branches, but not to make a postulation concerning a common ancestor (Taru & Shukla 2017) and evolutionary direction.

Check for indel (insertion and deletion)

Most phylogenetic approaches accept that respective placement of a sequence can adjust individually from the other positions. Gaps in alignments signify alterations in sequences such as insertion, deletion or genetic rearrangements (Rodriguez-Murillo & Salem 2013). Gaps are served in numerous ways by the phylogenetic methods (Rodriguez-Murillo & Salem 2013). Some alignment algorithms disregard gaps and treat them as missing data. Any vaguely aligned regions in the alignment should be omitted as they would then decline the precision of the phylogenetic analysis.

4) Sequence alignment and available software

Separate files are first prepared for single locus and are aligned using various software/programs. After obtaining the aligned file (from below mentioned software), sequences are manually adjusted if necessary, as manual aligning is essential to create high-quality reference alignments. BioEdit (Hall 1999) provide unique functionality which is analyzing sequence alignments as part of the manual annotation process. Though there are many different alignment tools available, 'AliView' is an alignment viewer and editor intended to encounter the necessities of next-generation sequencing era phylogenetic datasets, which offers the fastest and most intuitive to use (Larsson 2014). AliView handles alignments of limitless size in the formats most commonly used, i.e. FASTA, phylip, nexus and clustal.
The details of widely used software are given below.

* **MAFFT**

The alignment software MAFFT (Multiple alignments using iterative strategies, https://mafft.cbrc.jp/alignment/server/) (Katoh & Toh 2010, Rozewicki et al. 2019), allows effective and accurate alignment of DNA or amino acid sequences using various iterative alignment strategies. FASTA-formatted DNA or amino acid sequences are possible to submit to the MAFFT server (Fig. 3) and the sequences are aligned according to the user's settings.

As a case study, we show how to introduce a new species in the genus *Botryosphaeria*. The file which is composed of type sequences (obtained from GenBank), newly attained sequence data and the out-group could be uploaded to a sequence aligning software (e.g. MAFFT server, Fig. 3). The resulted FASTA file could be viewed using various formats (e.g. BioEdit).

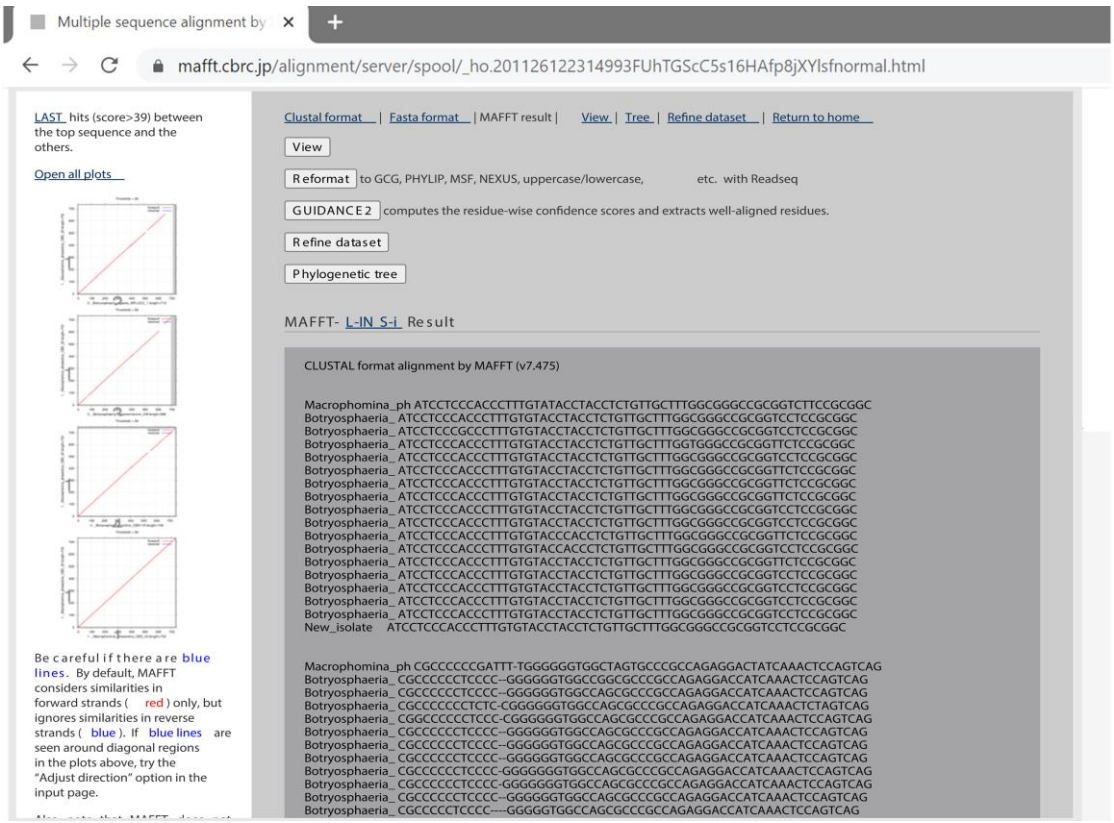

Figure 3 – Uploading a text file to MAFFT server.

* **BioEdit**

BioEdit (Biological sequence alignment Editor) (Hall 1999), is one of the most extensively used software to view and edit the sequence alignments (Fig. 4). The most recent version is 7.2 (from 2019). By overtyping, the sequence names can simply be renamed. BioEdit helps to convert

DNA into the complementary sequence or the reverse complement. The ambiguously aligned regions in the alignment can be manually excluded both from front and reverse as this increase the accuracy of phylogenetic analysis significantly (Talavera & Castresana 2007) and is thus highly recommended. Also, manual editing is done by many molecular phylogeneticists. Two or more gene regions can be combined in the same alignment (Fig. 4). Most of the phylogeneticists add a mark to differentiate one gene region from other (Here we marked with 4N) just for the easiness to spot each region. Hence this is not essentially beneficial.

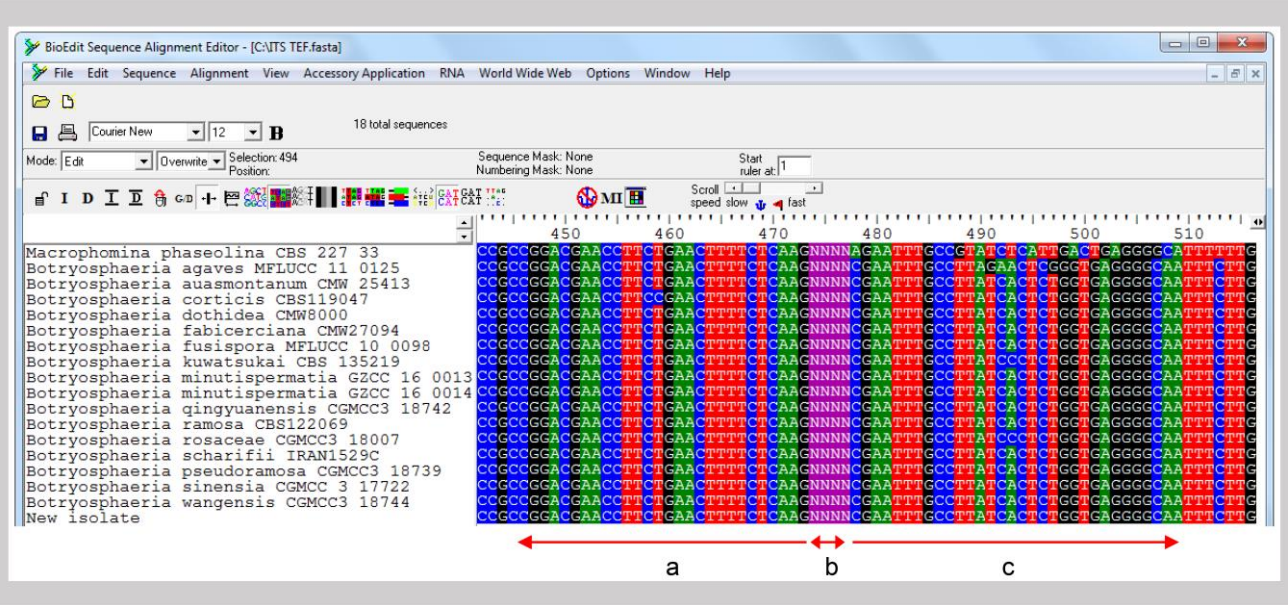

Figure 4 – Combining the gene regions of the 'final FASTA file' using BioEdit. a Gene region 1. b Combination of two loci by adding 4N. c Gene region 2.

* **ClustalX**

The clustalX program (http://bips.u-strasbg.fr/fr/Documentation/ClustalX/) (Larkin et al. 2007) is a still commonly used method for aligning sequences (Fig. 5). However, for larger or heterogeneous datasets, an iterative approach such as MAFFT software is recommended to use.

e.g. nexus format (Maddison et al. 1997)

Sequence/isolate names may exceed 10 characters. Standard isolate names could contain letters, numerals and underscore symbol but blanks are not allowed. Names composed of other symbols or blanks should be contained within quotes (but should ideally be avoided altogether). Since some legacy software requires unique sequence names, and truncate words longer than, e.g., 10 characters, it is a good habit to use sequence names of the form "KX099758_Mycorrhaphoides_stalpersii". Truncation would still leave the name unique in these cases.

Construct single locus trees and multiple loci tree

The 'final alignment' could be a single locus or a combination of multiple loci. The separately aligned single locus can be combined using BioEdit (Fig. 4) and should be done in such a way that the order of sequences is kept the same. According to the purpose of each study, single locus or multiple loci can be used to construct the phylogenetic analyses.

* **TrimAl**

TrimAl (Capella-Gutiérrez et al. 2009) is an automated tool for alignment trimming that can remove spurious sequences or poorly aligned regions (Fig. 6). TrimAl provides several automated options that can select the most appropriate thresholds for the alignment. TrimAl can be accessed using the Phylemon web server (http://phylemon.bioinfo.cipf.es/utilities.html) or using the

standalone version (http://trimal.cgenomics.org). The alignment of the ITS region after trimming using the automated method on TrimAl is shown in Fig. 7.

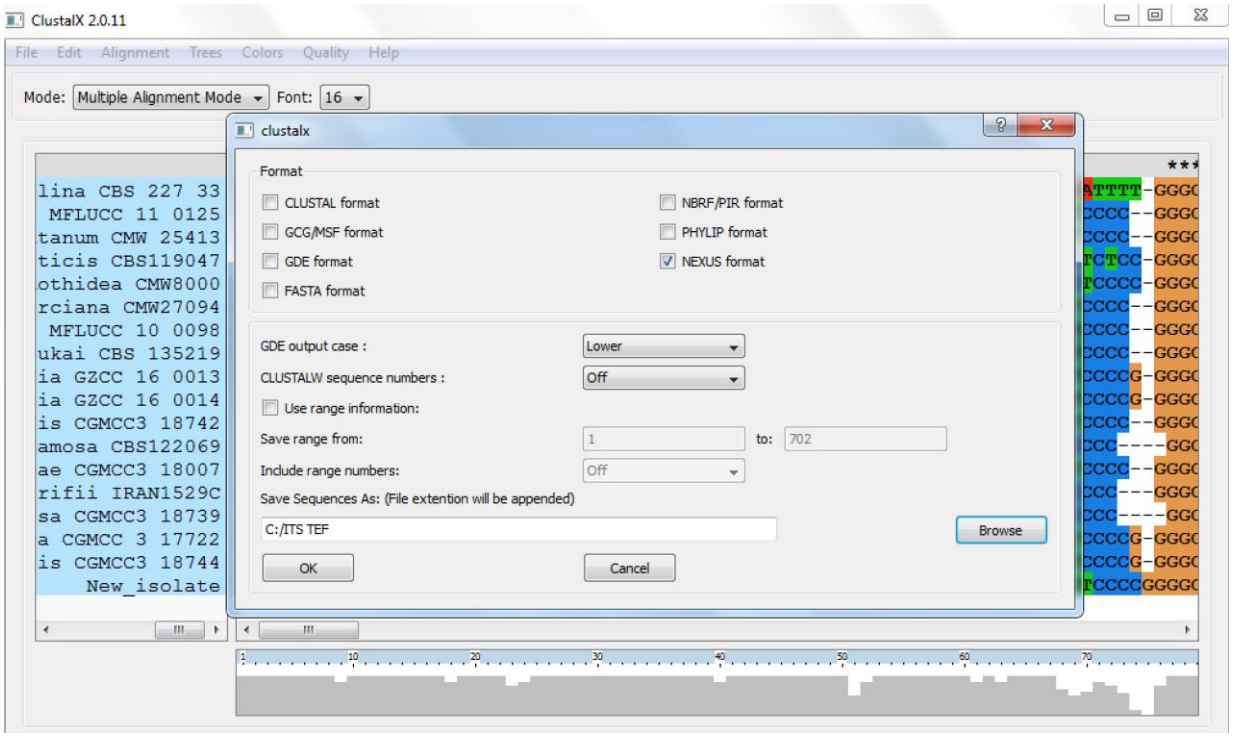

Figure 5 – ClustalX converts the FASTA formatted file to a nexus file

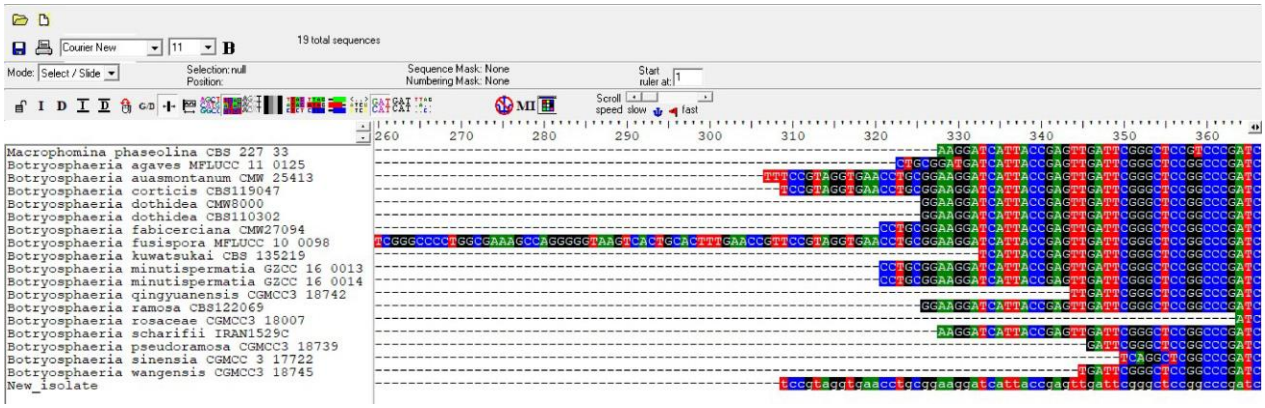

Figure 6 – The alignment of the untrimmed ITS region obtained from the MAFFT server

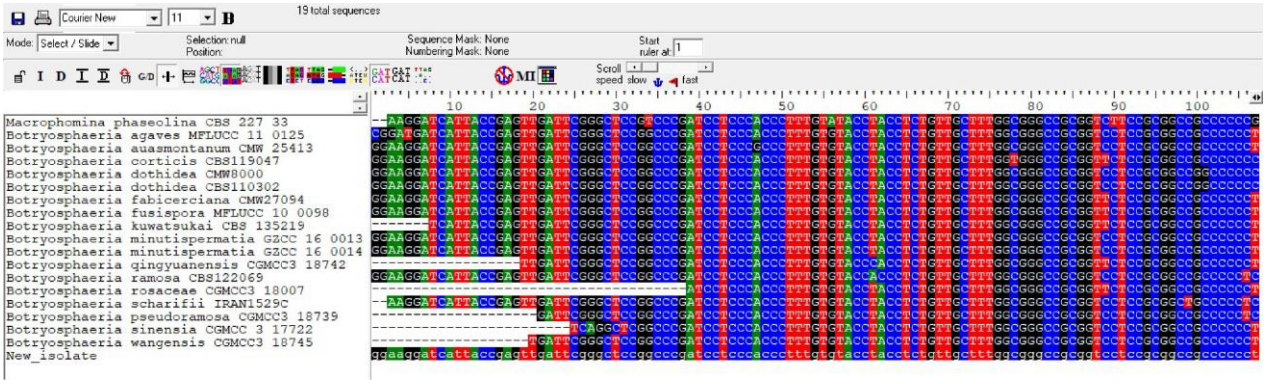

Figure 7 – The trimmed output from TrimAl using the automated method on the ITS region

* **GUIDANCE2**

GUIDANCE2 (Sela et al. 2015) is an important tool that can assign a confidence score for each sequence in the alignment. GUIDANCE2 identifies sequences that cannot be reliably aligned and enables their automatic removal. The confidence score from GUIDANCE2 is a reflection of the robustness of the generated alignment. An important feature of GUIDANCE2 is the color-coded multiple sequence alignment (Fig. 8) which is based on the confidence score of each residue in the alignment. Another output includes a text file of the generated alignment on which the results are based. GUIDANCE2 provides the choice of three algorithms for alignment including MAFFT, PRANK and clustalW.

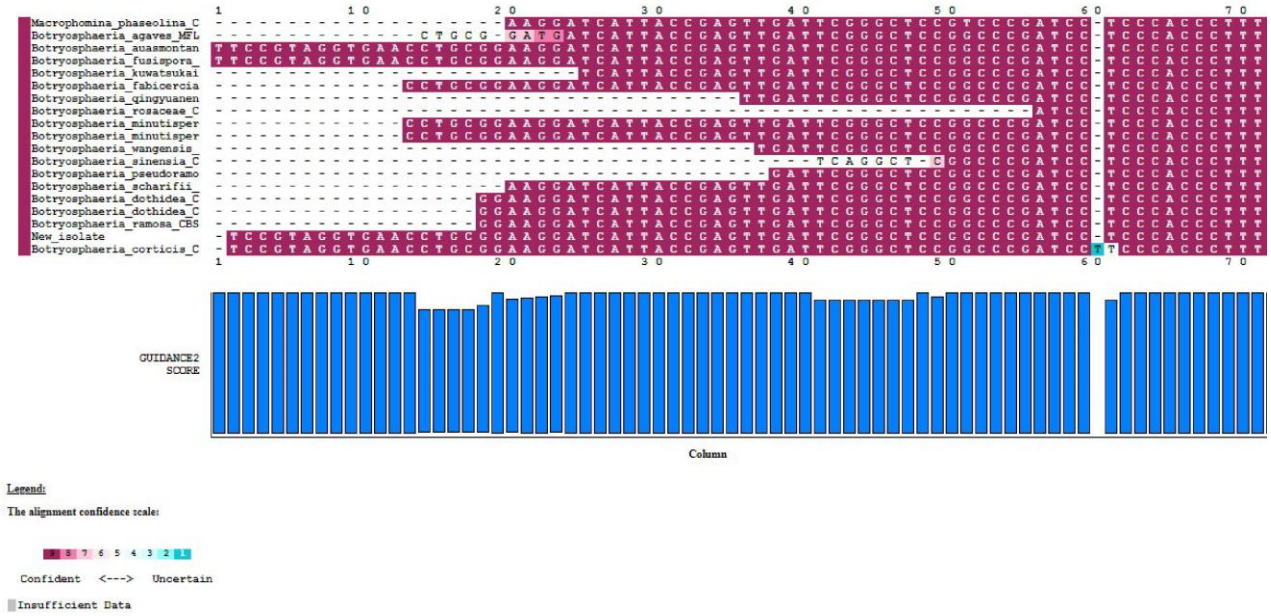

Figure 8 – The multiple sequence alignment colored according to the confidence score in GUIDANCE2 based on the ITS region

5) Conversion of final alignment to different formats

*** ALTER**

ALTER (ALignment Transformation EnviRonment, http://www.sing-group.org/ALTER/) is a web-based tool to convert between multiple sequence alignments formats.

e.g. phylip format

This is the standard input format for RAxML. The sequence names should only contain letters and numerals. A blank between sequence name and the sequence is not obligatory, but a newline character should be included (Fig. 9).

* **AliView**

AliView (Larsson 2014) is another option to change the formats of multiple sequence alignments. AliView is an alignment viewer and editor which can open files from a range of formats including FASTA, nexus and phylip. The sequence names could contain letters, numerals and underscore symbol but blanks should be removed before phylogenetic analyses (Fig. 10).

6) Software tools/websites/databases use for major phylogenetic analyses

Phylogenetic relationships of the isolated taxa are mainly identified using MP applied in PAUP (Swofford 2003), ML in RAxML (Silvestro & Michalak 2012) and Bayesian analysis in MrBayes (v32..7a) (Ronquist et al. 2012). In these three methods, ambiguous regions in the alignments are excluded, and gaps could be treated as missing data or as characters.

The CIPRES Science Gateway (Cyber Infrastructure for Phylogenetic RESearch, https://www.phylo.org/portal2/login!input.action) is a public resource for inference of bulky datasets. It is designed to provide all researchers with access to NSF XSEDE's large computational resources through a simple browser interface.

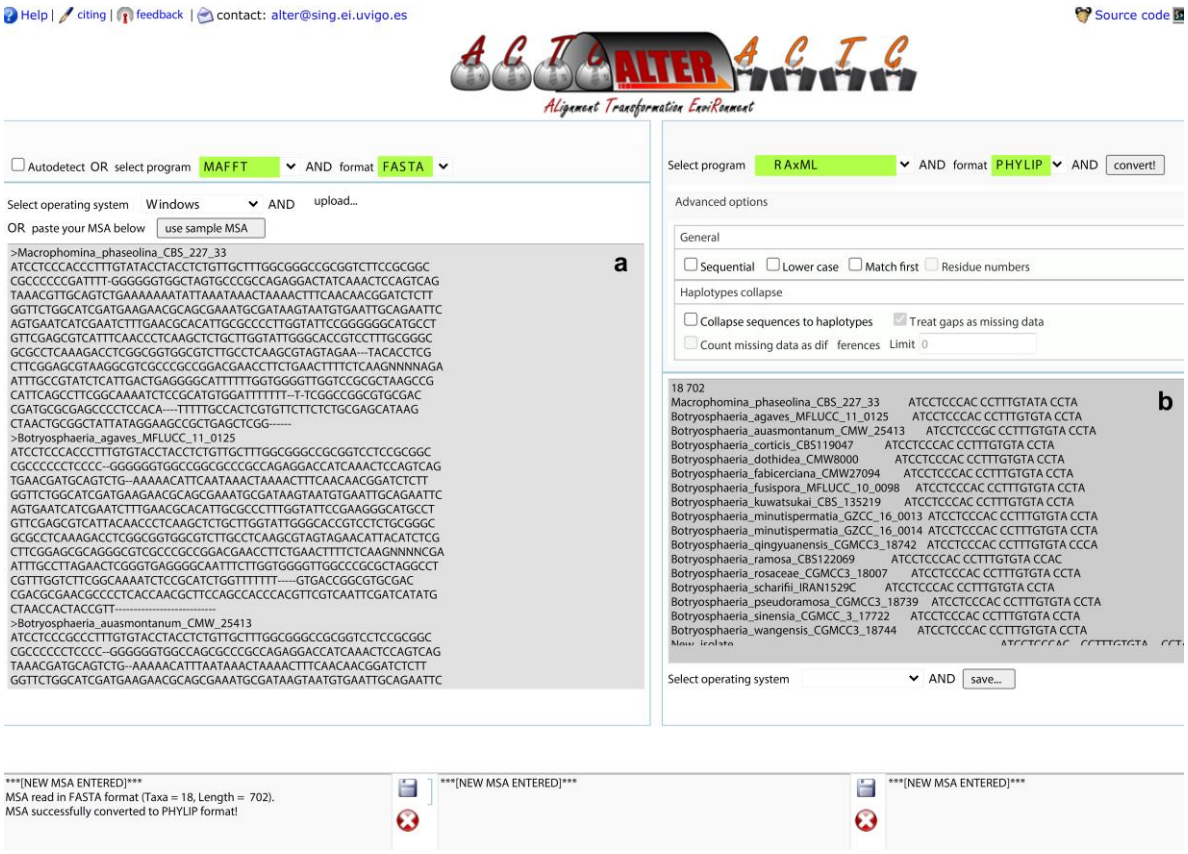

Figure 9 – Conversion of a FASTA file to a phylip file using ALTER. a Loaded the FASTA file. b The resulted phylip file.

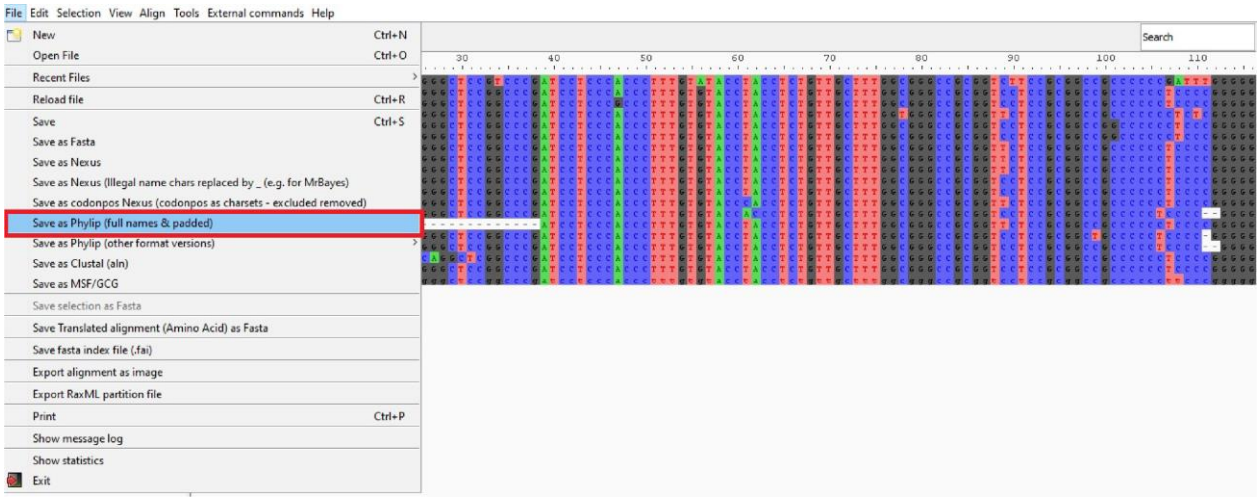

Figure 10 – Conversion of a FASTA file to a phylip file using AliView

i) Maximum parsimony analyses (MP)

Branch support is evaluated through bootstrap or jackknife replicates. For analysis of large matrices, parsimony jackknifing converges faster than extensive branch-swapping (Farris et al. 1996). The branches with zero-length are collapsed, and all multiple parsimonious trees are saved.

Tree-length (TL), consistency index (CI), retention index (RI), relative consistency index (RC) and homoplasy index (HI) are considered. The resulting trees are statistically assessed using Kishino-Hasegawa test (KHT) (Kishino & Hasegawa 1989).

e.g. 1 PAUP (Phylogenetic Analysis Using Parsimony, Swofford 2003)

This is the most commonly used computer program for phylogenetic analysis of fungi. This tool has the major advantage that complex procedures can be saved and easily be applied to new data sets (Weiß 2010).

e.g. 2 TNT (Tree analysis using New Technology)

A variety of methods for diagnosing trees and exploring character evolution is available in TNT (Goloboff et al. 2008).

Basic steps in an MP analysis

This could be changed depending on the user requirements.

- Open the 'Final FASTA file' (Fig. 4) using 'clustalX' and save the file as 'nexus' format.
- Open 'PAUP' and open the nexus file.
- Copy 'single line' program to PAUP and run.

  e.g.
  Begin PAUP;
  export format = nexus
  interleaved = no
  file = filename.nex;
  end;
- Using the filename.nex file, run the PAUP for '1000 parsimony command (Fig. 11)

  e.g. 1000 parsimony command for *Botryosphaeria* dataset. The outgroup is Macrophomina_phaseolina_CBS_227_33

```
begin paup;
    exset * exclude =  ;
    delete   /only;
    log file= outputfile.txt;
    pset collapse=minbrlen;
    [ctype 1.5_1:all;]
    outgroup Macrophomina_phaseolina_CBS_227_33;
    set maxtrees=1000 increase=no;
    set criterion=parsimony;
    hsearch addseq=random  nreps=1000; roottrees outroot=monophyl;
    savetrees brlens=yes file=Mp_filename.tre;
    pscores ALL/ci=yes tl=yes hi=yes rc=yes ri=yes khtest=yes;
    bootstrap nreps=1000 Keepall=yes / AddSeq=random nreps=10;
    roottrees outroot=monophyl;
    savetrees file=BT_filename.tre from=1 to=1 savebootp=both maxdec=0;
end;
```

This could be modified according to the dataset

Some parts of the gene region can be excluded

The name of the log file could be changed

More than one outgroup could be included

- The resulting 'Maximum parsimony' (MP) tree (Fig. 12a) and 'Bootstrap support' (BT) tree (Fig. 12b) can be processed further by appending the support values from BT to MP.
- All other necessary details are given in the output file, herein named as 'outfile'.

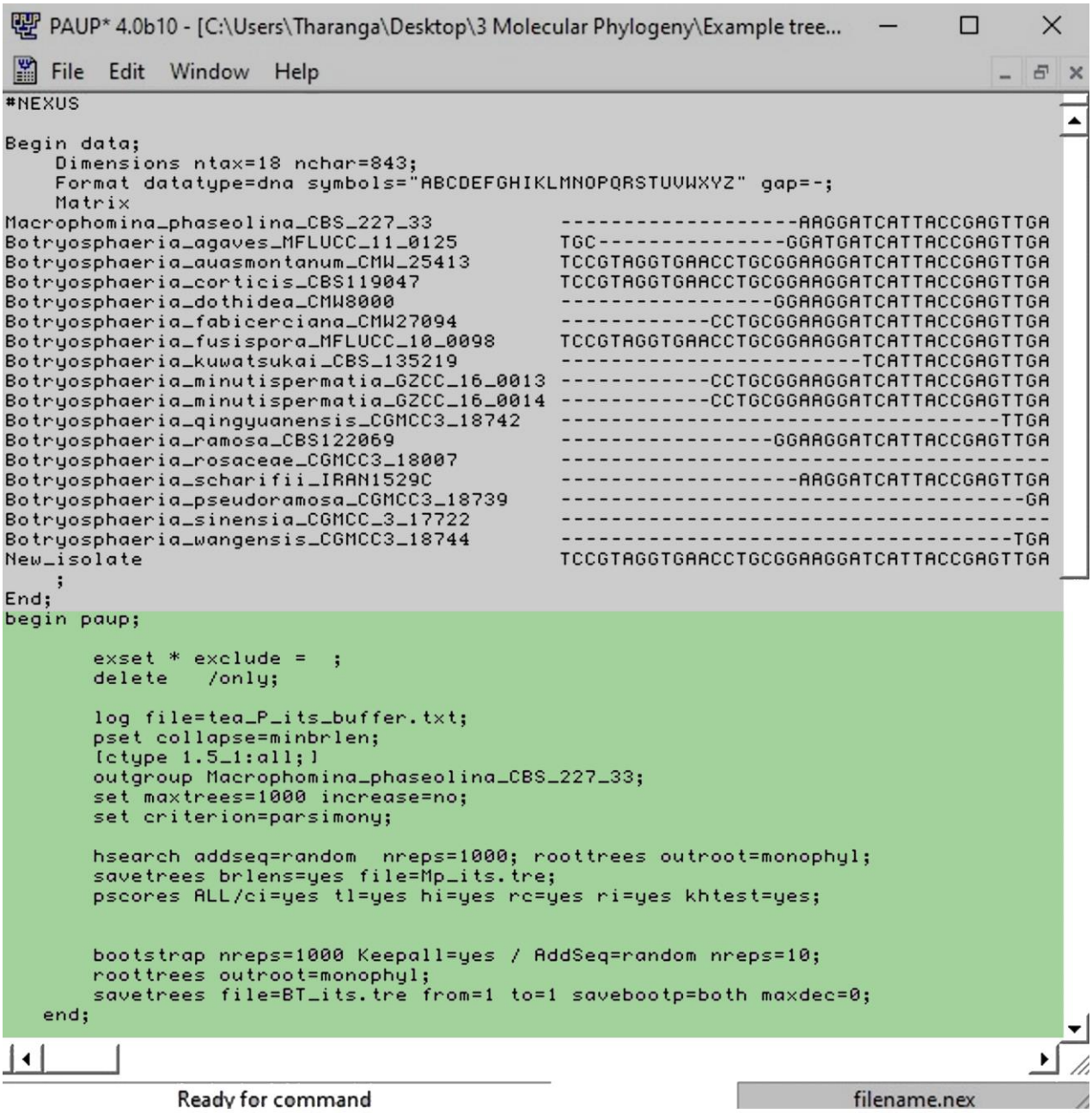

Figure 11 – Executing the PAUP for '1000 parsimony command'.

ii) Maximum likelihood analysis (ML)

The ML trees are conducted using RAxML-HPC2 on XSEDE (8.2.8) (Stamatakis et al. 2008, Stamatakis 2014) in the 'CIPRES Science Gateway platform' (Miller et al. 2010) incorporating various models of evolution with different non-parametric bootstrapping iterations.

e.g. RAxML (Randomized acc(x)elerated maximum likelihood**)**

RAxML is the fastest available software to run heuristic ML analysis on large datasets, which may include thousands of sequences (Stamatakis 2006).

Basic steps in a ML analysis

In this case study, we used the 'GUI' version of RAxML (raxmlGUI). The choice of substitution models for nucleotide sequences is limited to GTR models in RAxML.

- Navigate to ALTER link (Fig. 9). http://www.sing-group.org/ALTER/

- The file types can be changed according to the user's settings.
- Upload the 'Final FASTA file' (Fig. 4).
- The resulting file should be opened in 'raxmlGuI (python file)' program.
- Load the alignment and change the outgroup. Change the icons according to the user's settings (Fig. 13).
- Run RAxML.

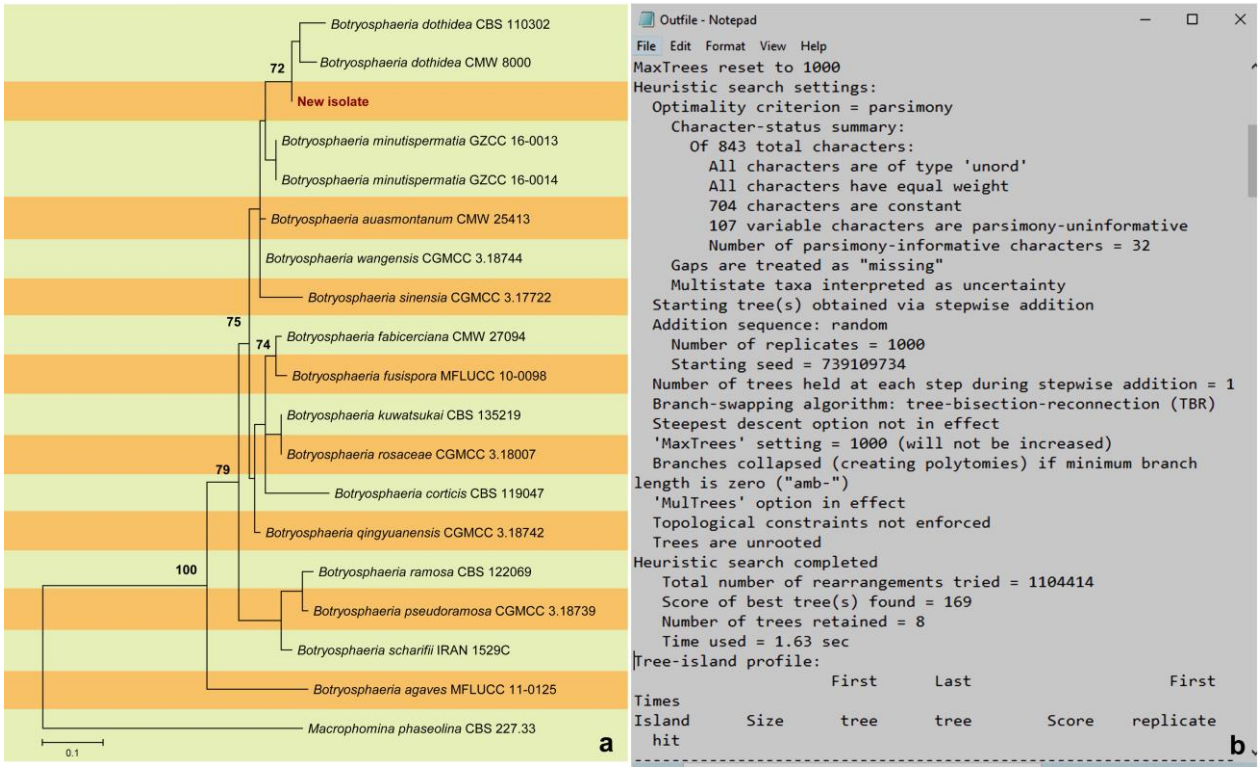

Figure 12 – The resulting files from the maximum parsimony analyses. a The Main tree file (bootstrap support values were imported from BT tree file). b The buffer file.

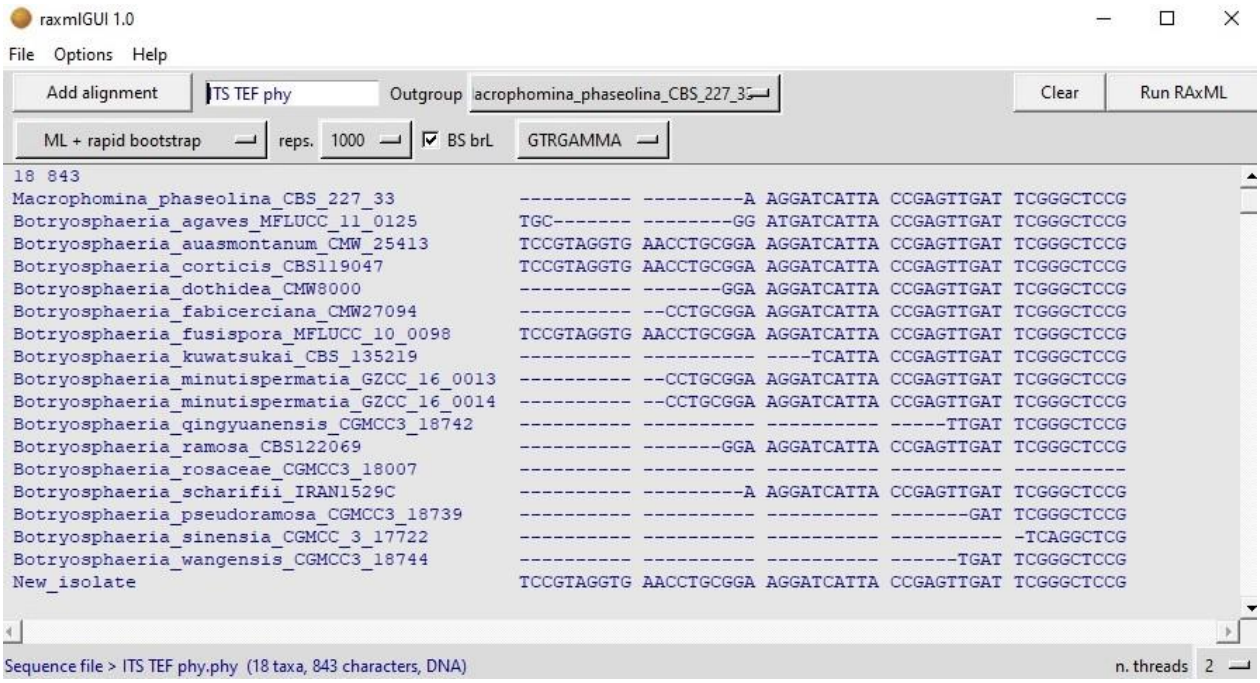

Figure 13 – Execute mode of RAxML program.

- Check the ‘TRE’ files. The resulted ‘Best tree’ and ‘Bipartitions tree’ can be processed further (Fig. 14a, b).
- All other necessary details are given in the ‘info’ file.

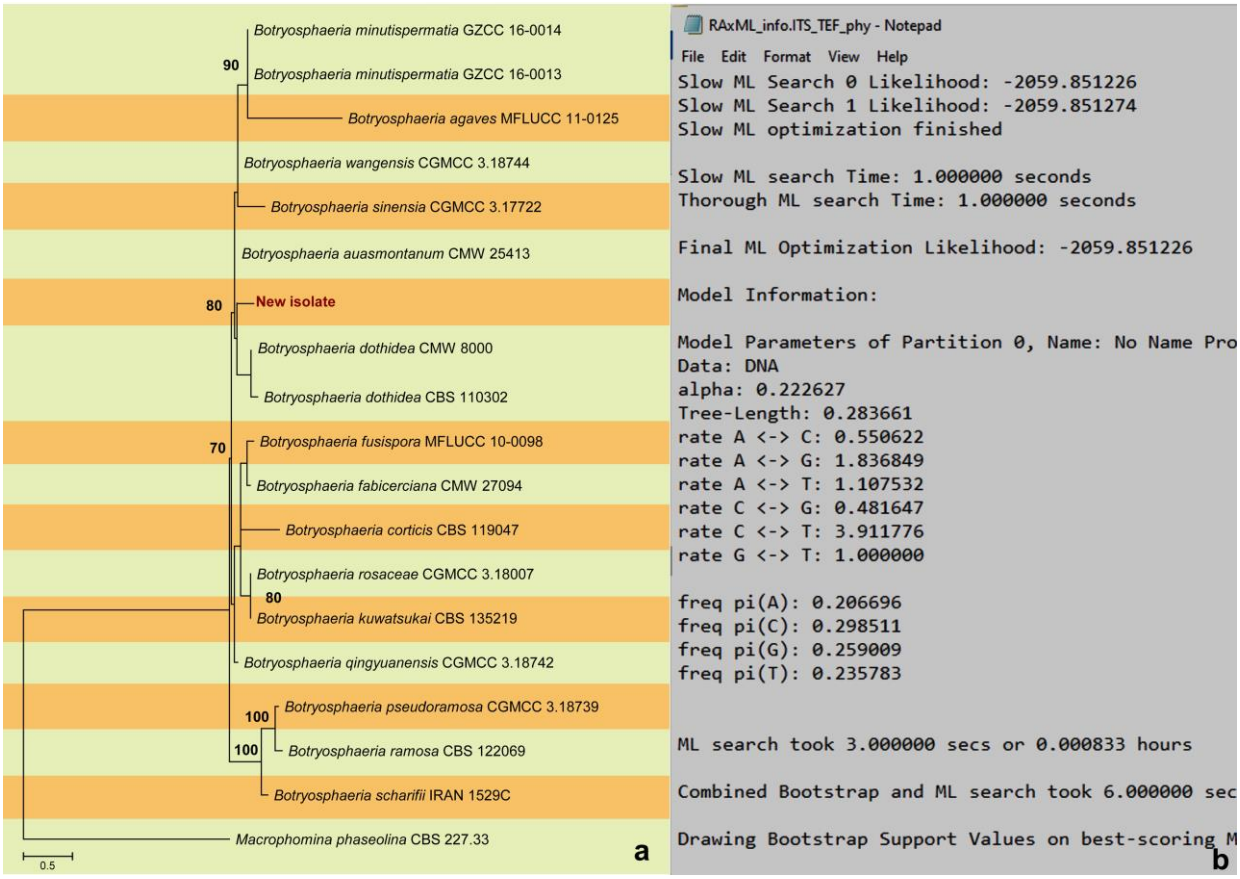

Figure 14 – The resulting files of maximum likelihood analyses. a The Main tree file (bootstrap support values were imported from BT tree file). b The out file.

PHYML

PHYML software allows effective and fast ML analysis (Guindon & Gascuel 2003). PHYML allows processing DNA as well as amino acid data with numerous substitution models. The output tree already includes the bootstrap percentages.

iii) Bayesian analysis

For the Bayesian analysis, the evolutionary models are designated by MrModeltest v.2.3 (Nylander 2004). Bayesian analysis is performed in MrBayes (v3.2.a) (Ronquist et al. 2012) and posterior probabilities (PPs) are selected by Markov chain Monte Carlo sampling (BMCMC). Synchronized Markov chains (four or six) are run for 50^6 to 10^7 generations, sampling the trees at each 100th generation. As of the 10,000 trees attained, 20% to 30% trees representing the burn-in phase are discarded. The remaining 80% to 70% trees are used to calculate PPs in a majority rule consensus tree.

e.g. MrBayes

MrBayes is the most widely used software to estimate phylogenetic relationships.

Basic steps in a Bayesian analysis

1) MrModelTest

- Prepare single gene FASTA files and convert them to nex files.

- Open nex file in PAUP (choose file type as ‘All files’ and file open mode as ‘Edit’).
- Remove the ‘Symbols’ line and execute the program.
- Open ‘MrModelblock’ (inside ‘MrModelTest’ folder) in PAUP file and execute the program.
- In MrModelTest’ folder ‘mrmodel.scores’ file and ‘mrmodelfit’ file can be visible. ‘mrmodeltest2’ file is available in the ‘bin’ folder.
- Open a command prompt and drag the files as follows.

“mrmodeltest2<mrmodel.scores> referencefile.txt

- The resulted ‘referencefile’ is available in the ‘user’ folder (Fig. 15). The Bayesian information can be obtained from this reference file. ‘GTR +I+G’ model and ‘SYM+I+G’ model are available.
- Repeat this process for each gene region to obtain the appropriate model.

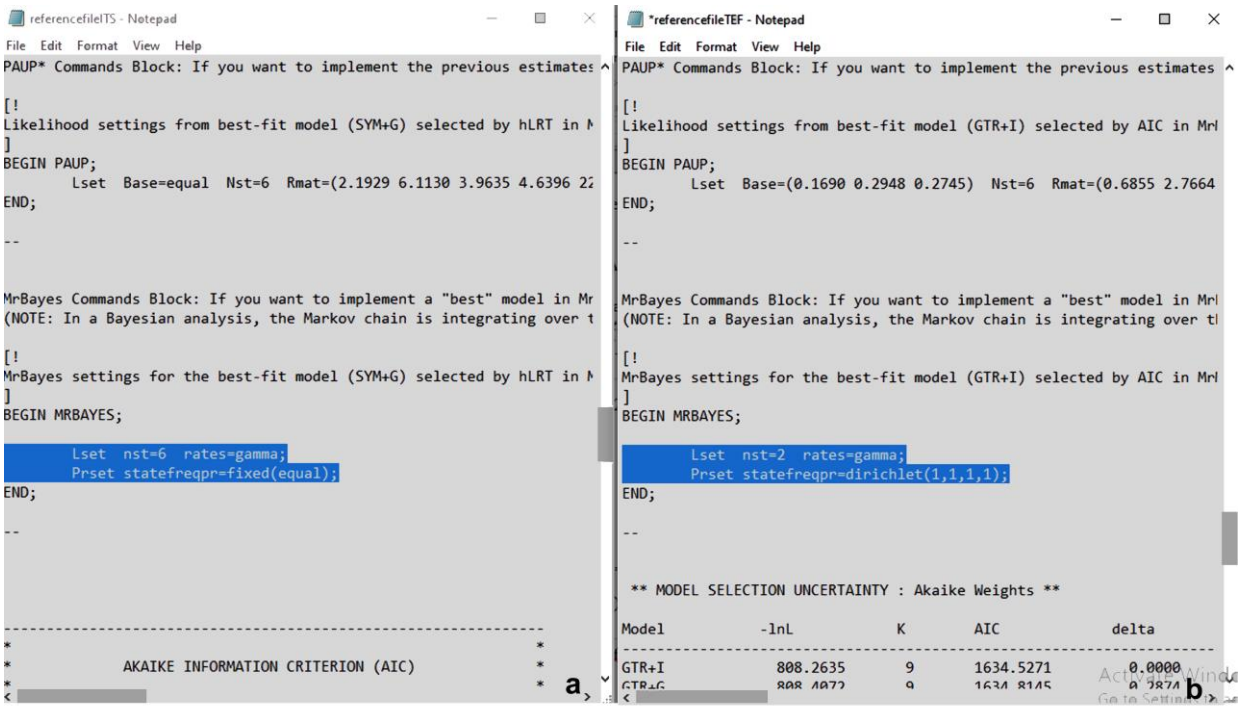

Figure 15 – Reference files showing model for each gene region. a ITS locus. b TEF locus.

2) Navigate to ALTER

- Upload the ‘Final FASTA file’ file to the ALTER server according to the user’s settings (Fig. 4).
- Open the resulted file, paste the Bayes command and replace the values and the outgroup.

e.g. Bayes command for *Botryosphaeria* dataset. The outgroup is Macrophomina_phaseolina_CBS_227_33

```
begin mrbayes;
outgroup Macrophomina_phaseolina_CBS_227_33;
charset ITS = 1-578;
charset TEF = 579-843;
partition matrices = 2: ITS, TEF;
set partition = matrices;
Lset applyto = (1) nst = 6 rates = gamma;
Lset applyto = (2) nst=2 rates = gamma;
prset applyto = (1) statefreqpr = fixed(equal);
prset applyto = (2) statefreqpr = dirichlet(1,1,1,1);
```

Rename the outgroup

Replace the values according to MrModelTest

unlink statefreq = (all) revmat = (all) shape = (all) pinvar = (all);
mcmc ngen = 5000000 printfreq = 1000 samplefreq = 1000 nchains = 6 savebrlens = yes;
mcmcp;
sumt burnin = 4000;
end;

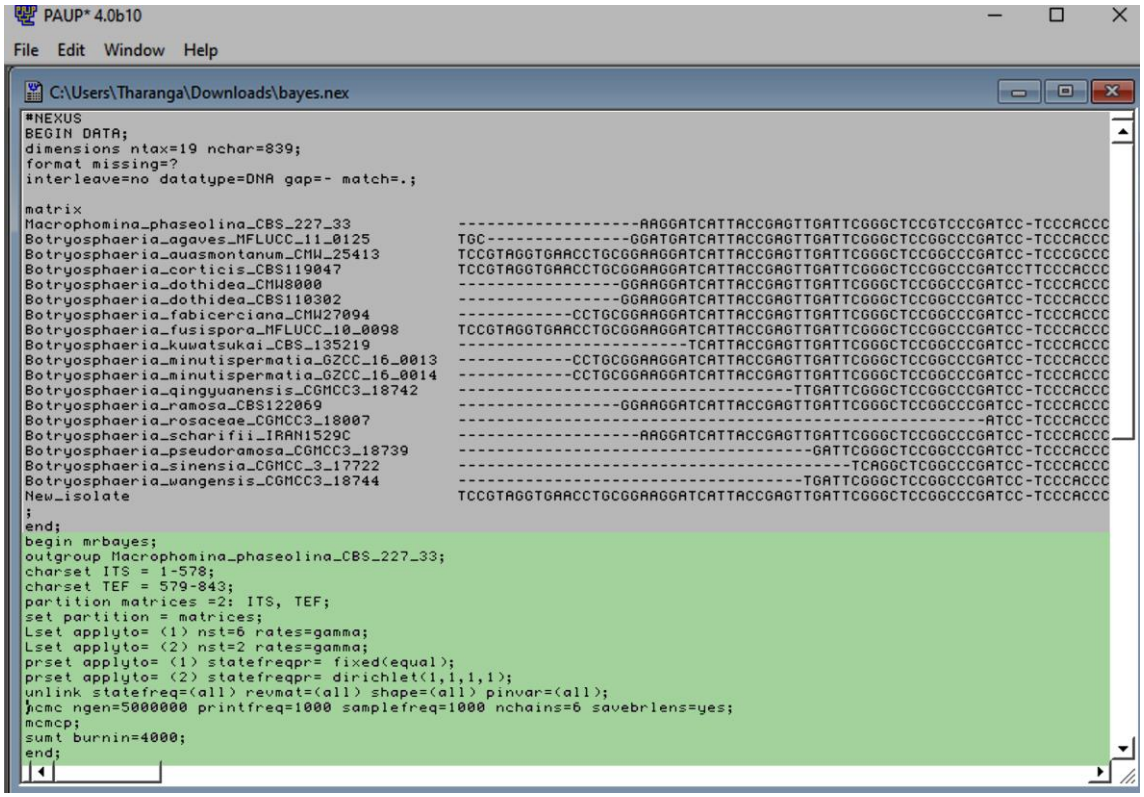

Figure 16 – Executing the PAUP for ‘Bayes command’.

- Save this file (Fig. 16) as ‘bayes.nex’ and copy it to the MrBayes folder.
- Open the ‘MrBayes image’ and Type ‘bayes.nex’ and hit the ‘Enter’ button (Fig. 17).

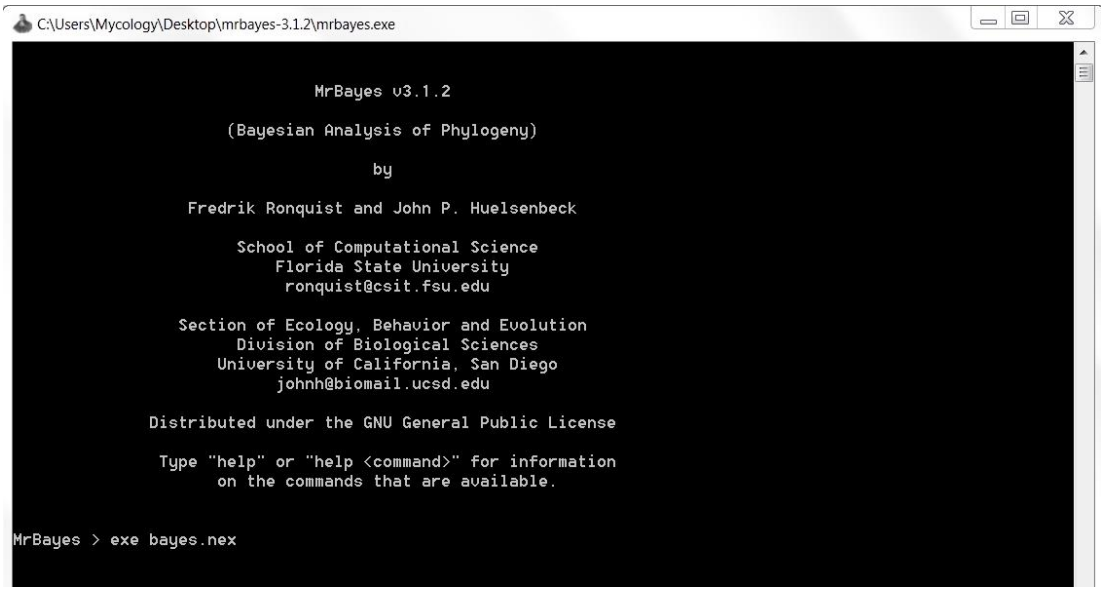

Figure 17 – Executing the file in MrBayes.

This depends on the data and should check with 'Tracer' (Rambaut et al. 2018)

- After the program is run, type 'Sumt burnin = 2000' to discard the burnin phase.
- The final result files appear in the MrBayes folder.
- The resulting 'Con' file can be displayed in, e.g. iToL (https://itol.embl.de/) or in FigTree (Fig. 18)

Tree reading software

There are a lot of tools existing to customize and gloss the phylogenetic trees. The commonly incorporated software/tools are as follows.

- iTOL

Interactive Tree Of Life (iTOL) is an internet-based utensil for the display, operation and modification of phylogenetic trees. The datasets can be directly drag and drop onto the tree, with comprehensive control of respective picturing choice. Branch and label colors, styles and typefaces can be modify interactively or by organized data records. Trees can be interactively trimmed and re-rooted. Several sorts of records such as genome sizes or protein field collections can be plotted onto the tree (https://itol.embl.de/) (Letunic & Bork 2016).

- TreeView

A tree view is a sketching component that presents a graded outlook of the evidence. Inferred phylograms can be imported, displayed and plotted using Tree view (https://treeview.software.informer.com/).

- MEGA

MEGA is an advantageous software in building phylogenies and envisioning them, and also for recording transformation (Tamura et al. 2013, Kumar et al. 2016). It also transforms alignment records to additional layouts. The MEGA tree explorer is supportive in modifying trees very simply, subtrees can also be designated and modified distinctly. Specific tree image exporting possibilities are also offered. The input arrangements are newick, phylip, mega, and nexus (https://www.megasoftware.net/).

- Dendroscope

This supports envision of bulky trees and offers numerous choices to trace the sketching through a command line (Huson & Scornavacca 2012). A number of diverse outlooks are also accessible, trees can be simply re-rooted and node labels and branches can be easily structured. It can transfer trees in newick and nexus format. The consumers will have to register themselves first to procedure this tool (http://dendroscope.org/).

- FigTree

FigTree is a graphical viewer for phylogenetic trees and a program for producing publication-ready figures. Trees can be graphically manipulated and annotated in various ways. FigTree is aimed to envision the trees that are created by BEAST. Tip labels and node labels can be simply modified. This will straightforwardly export trees in nexus, newick, and JSON format through certain graphics export possibilities such as emf, pdf, sg, png, etc. (http://tree.bio.ed.ac.uk/software/figtree/).

Formatting the trees

Adobe illustrator is the most commonly used tool to format phylogenetic trees. Phylograms are plotted in treeview and edited in Adobe Illustrator CS6 v.16.0.0

(https://www.adobe.com/cn/products/illustrator.html). Microsoft PowerPoint, InkSpace (https://inkscape.org/) is also commonly used to edit and format the trees.

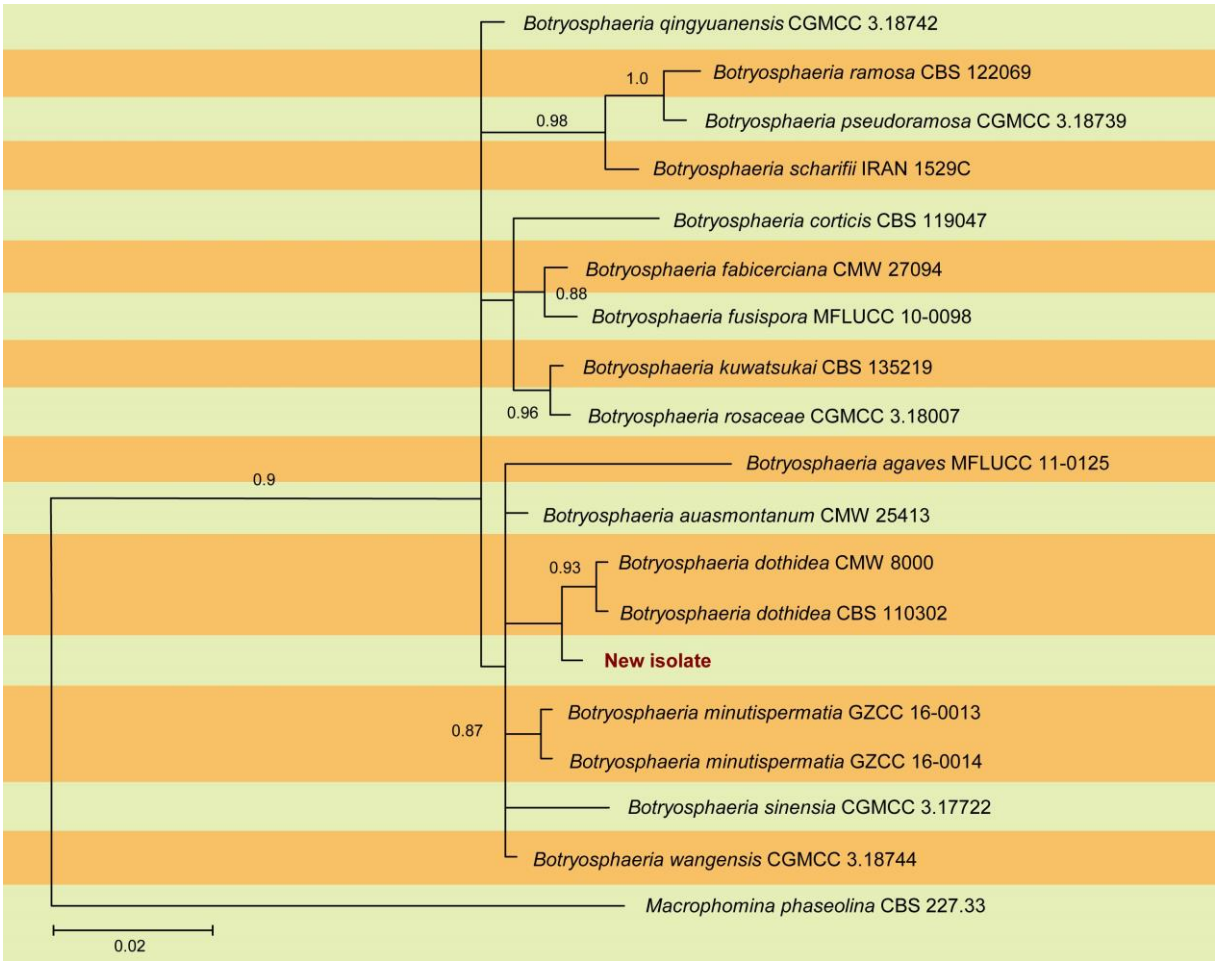

Figure 18 – The resulting Bayes tree.

Writing tree legends

The tree legends briefly describe the major aspects of the tree such as the mode used for the analysis, gene regions used in the analysis, node probabilities, the outgroup taxon/taxa and how the taxa are present/treat. Tree scale bar indicates the extent of the evolutionary changes. Often, the number on the scale should be the percentage of estimated genetic variation. For phylogenetic studies, the length of the scale bar would usually be much shorter (with a scale bar at about 2-5%). However, this is not always the case, and that is why scale bars should always be explained under the figures.

Submit the sequences to GenBank (https://submit.ncbi.nlm.nih.gov/subs/genbank/)

GenBank accepts sequence data directly determined by the submitter. The submission must include information about the source organism and annotation provided by the submitter. Novel generated sequences should be deposited in GenBank with required information, as well as the proper annotation for those protein gene regions. The utmost significant cause of new data for GenBank is direct acquiesce from researchers. GenBank is contingent on its contributors to assist in retaining the database as broad, modern, and precise as probable. NCBI offers timely and accurate handling and natural evaluation of original entries and updates to prevailing entries and is equipped to support authors who have novel data to submit. The GenBank offers numerous options (BankIt, Sequin, direct submission) to submit the sequences. The GenBank accession numbers should be published/available online for the use of future studies.

Submit the alignment to TreeBASE

An author normally starts a TreeBASE submission (https://treebase.org/treebase-web/submitTutorial.html), before submitting a paper to a journal for evaluation. This step is categorized as "in progress." The submitting author receives a specified URL and a submission ID. This URL can be sent to the journal editor so that referees or commentators can review (but cannot do modification) the provided data. This is convenient for manuscript reviewers to track the alignment and the phylogenetic trees.

Additional Software

In this section, we highlight two approaches the Automatic Barcode Gap Discovery (ABGD) method and objective clustering, which have been applied for species delineation in several organisms including fungi. A recent application of these approaches for species delineation was in *Bipolaris* (Bhunjun et al. 2020). The Automatic Barcode Gap Discovery (ABGD) method (Puillandre et al. 2012) can be used as additional evidence for species delineation. ABGD is an analytical tool that sorts sequences into putative species. ABGD delineates species based on a barcode gap between intraspecific and interspecific diversity. ABGD can be accessed using the web server (https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html) or by using the command line version.

Objective clustering in TaxonDNA (Meier et al. 2006) is another approach that can be used to determine the number of hypothetical species in a dataset. This method delineates species based on intra and interspecific genetic distances. TaxonDNA can also be used to calculate the *p*-distance or K2P distance of the taxa in the dataset.

General Conclusion

This manuscript aims to provide the basic guidelines of MP, ML and BI analysis for the beginners who are willing to engage with mycological researches based on taxonomic and phylogenetic studies. A newcomer should start by reading up on the relevant topics of phylogenetic inference and not just try to produce a tree as quickly as possible. Since the manuscript is providing the details of most needed software/links/databases, the beginners would be able to easily access or follow the steps to construct the phylogenetic analysis. A thorough phylogenetic tree requires more practice and should be tailored to the specifics of the dataset and study.

Acknowledgment

We are grateful to Prof. Dr. Kevin D. Hyde for his valuable suggestions to shape the paper and Dr. Martin Ryberg is thanked for providing comments to improve it. Dhanushka N. Wanasinghe, Chada Norphanphoun and Milan C. Samarakoon are thanked for their assistance on figure editing.

References

Alexopoulos CJ. 1963 – The Myxomycetes. II Botanical Review 29, 1–78.

Altschul SF, Madden TL, Schäffer AA, Zhang J et al. 1997 – Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25, 3389–3402.

Avise JC, Johns GC. 1999 – Proposal for a standardized temporal scheme of biological classification for extant species. Proceedings of the National Academy of Sciences USA. 96, 7358–7363.

Avise JC, Mitchell D. 2007 – Time to standardize taxonomies. Systematic Biology. 56, 130–133.

Bartlett J, Stirling D. 2003 – A Short History of the Polymerase Chain Reaction. Methods in Molecular Biology 226, 3–6.

Bhunjun CS, Dong Y, Jayawardena RS, Jeewon R et al. 2020 – A polyphasic approach to delineate species in *Bipolaris*. Fungal Diversity 102, 225–256.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009 – trimAl, a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 2515, 1972–1973.
Chen YY, Dissanayake AJ, Liu ZY, Liu JK. 2020 – Additions to Karst Fungi 4, *Botryosphaeria* spp. associated with woody hosts in Guizhou province, China including *B. guttulata* sp. nov. Phytotaxa 454, 186–202.
De Dreu CKW, Gross J, Méder Z, Giffin M et al. 2016 – In-group defense, out-group aggression, and coordination failures in intergroup conflict. Proceedings of the National Academy of Sciences of the United States of America 113, 10524–10529.
Dissanayake AJ, Phillips AJL, Hyde KD, Li XH. 2016 – Botryosphaeriaceae, Current status of genera and species. Mycosphere 7, 1001–1073.
Doyle JJ, Doyle JL. 1990 – Isolation of plant DNA from fresh tissue. Focus 12, 13–15.
Farris JS, Albert VA, Källersjö M, Lipscomb D, Kluge AG. 1996 – Parsimony jackknifing outperforms neighbor-joining. Cladistics 12, 99–124.
Felsenstein J. 2004 – Inferring Phylogenies Sinauer Associates, Sunderland, MA.
Gargas A, Taylor JW. 1992 – Polymerase Chain Reaction PCR - Primers for Amplifying and Sequencing 18S rDNA from Lichenized Fungi. Mycologia 84, 589–592.
Giribet G, Ribera C. 1998 – The position of arthropods in the animal kingdom: a search for a reliable outgroup for internal arthropod phylogeny". Molecular Phylogenetics and Evolution 9, 481–488.
Goloboff PA, Farris JS, Nixon KC. 2008 – TNT, a free program for phylogenetic analysis. Cladistics 24, 774–786.
Guarro J, Gené J, Stchigel AM. 1999 – Developments in fungal taxonomy. Clinical Microbiology Reviews 12, 454–500.
Guindon S, Gascuel O. 2003 – A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Systematic Biology 52, 696–704.
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.
Hall BG. 2013 – Building phylogenetic trees from molecular data. Molecular Biology and Evolution 30, 1229–1235.
Hall BG. 2017 – Phylogenetic Trees Made Easy. A How-To Manual. Sinauer Associates. Fifth Edition.
Hibbett DS, Binder M, Bischoff JF, Blackwell M et al. 2007 –A higher-level phylogenetic classification of the fungi. Mycological Research 111, 509–547.
Hillis DM, Pollock DD, McGuire JA, Zwickl DJ. 2003 – Is Sparse Taxon Sampling a Problem for Phylogenetic Inference?. Systematic Biology 52, 124–126.
Huson DH, Scornavacca C. 2012 – Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Systematic Biology 61, 1061–1067.
Hyde KD, Udayanga D, Manamgoda DS, Tedersoo L et al. 2013 – Incorporating molecular data in fungal systematics: a guide for aspiring researchers. Current Research in Environmental & Applied Mycology 3, 1–32.
Hyde KD, Maharachchikumbura SSN, Hongsanan S, Samarakoon MC et al. 2017 – The ranking of fungi, a tribute to David L. Hawksworth on his 70th birthday. Fungal Diversity 84, 1–23.
Jarvis E. 2014 – Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346, 1320–331.
Jeewon R, Liew ECY, Hyde KD. 2002 – Phylogenetic relationships of *Pestalotiopsis* and allied genera inferred from ribosomal DNA sequences and morphological characters. Molecular Phylogenetics and Evolution 25, 378–392.
Jeewon R, Cai L, Liew ECY, Zhang KQ, Hyde KD. 2003 – *Dyrithiopsis lakefuxianensis* gen. et sp. nov. from Fuxian lake, Yunnan, China, and notes on the taxonomic confusion surrounding *Dyrithium*. Mycologia 95, 911–920.
Judd WS, Campbell CS, Kellogg EA, Stevens PF, Donoghue MJ. 2007 – Taxonomy in Plant Systematics – A phylogenetic approach, 3rd ed., Sinauer Associates, Sunderland.

Katoh K, Toh H. 2010 – Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics 26, 1899–1900.
Kishino H, Hasegawa M. 1989 – Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data. Journal of Molecular Evolution 29, 170–179.
Kück P, Mayer C, Wägele JW, Misof B. 2012 – Long Branch Effects Distort Maximum Likelihood Phylogenies in Simulations Despite Selection of the Correct Model. PLOS ONE 7, e36593.
Kumar S, Stecher G, Tamura K. 2016 – MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33, 1870–1874.
Larkin MA, Blackshields G, Brown NP, Chenna R et al. 2007 – Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.
Larsson A. 2014 – AliView, a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30, 3276–3278.
Leebens-Mack J, Vision T, Brenner E, Bowers JE et al. 2006 – Taking the first steps towards a standard for reporting on phylogenies: Minimum Information about a Phylogenetic Analysis (MIAPA). Omics 10, 231–7.
Letunic I, Bork P. 2016 – Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Research 44, W242–245. Doi: 10.1093/nar/gkw290
Liu JK, Hyde KD, Jeewon R, Phillips AJL et al. 2017 – Ranking higher taxa using divergence times, a case study in Dothideomycetes. Fungal Diversity 84, 75–99.
Liu NG, Hyde KD, Bhat DJ, Jumpathong J, Liu JK. 2019 – Morphological and phylogenetic studies of *Pleopunctum* gen. nov. Phaeoseptaceae, Pleosporales - from China. Mycosphere 10, 757–775.
Maddison W. 1984 – Outgroup Analysis and Parsimony. Systematic Zoology 33, 83–103.
Maddison DR, Swofford DL, Maddison WP. 1997 – Nexus: An Extensible File Format for Systematic Information. Systematic Biology 46, 590–621.
Maharachchikumbura SSN, Hyde KD, Jones EBG, McKenzie EHC et al. 2015 – Towards a natural classification and backbone tree for Sordariomycetes. Fungal Diversity 72, 199–301
Maharachchikumbura SSN, Hyde KD, Jones EBG, McKenzie EHC et al. 2016 – Families of Sordariomycetes. Fungal Diversity 79, 1–317.
Meier R, Shiyang K, Vaidya G, Ng PK. 2006 – DNA barcoding and taxonomy in *Diptera*, a tale of high intraspecific variability and low identification success. Systematic Biology 55, 715–728.
Miller MA, Pfeiffer W, Schwartz T. 2010 – Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In, Proceedings of the Gateway Computing Environments Workshop GCE, 14 Nov. 2010, New Orleans, Louisiana. Minitab Inc., Boston, MA, USA minitab release 15.1.1.0.
Mount DM. 2004 – Bioinformatics, Sequence and Genome Analysis 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Nilsson RH, Tedersoo L, Abarenkov K. 2012a – Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. MycoKeys 4, 37–63.
Nilsson RH, Tedersoo L, Abarenkov K, Ryberg M et al. 2012b – Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. MycoKeys 4, 37–63.
Nilsson RH, Hyde KD, Paw1owska J, Ryberg M et al. 2014 – Improving ITS sequence data for identification of plant pathogenic fungi. Fungal Diversity 67, 11–19.
Nylander JAA. 2004 – MrModeltest v2 Program distributed by the author. Evolutionary Biology Centre, Uppsala University, Sweden.
O'Brien, Michael J, Lyman RL, Saab Y et al. 2002 – Two issues in archaeological phylogenetics: taxon construction and outgroup selection. Journal of Theoretical Biology 215, 133–150.

Pang KL, Vrijmoed LLP, Jones EBG. 2013 – Genetic variation within the cosmopolitan aquatic fungus *Lignincola laevis* Microascales, Ascomycota. Organisms Diversity and Evolution 13, 301–309.
Pearson WR, Lipman DJ. 1988 – Improved tools for biological sequence comparison. Proceedings of the National Academy of Sciences of the United States of America 85, 2444–2448.
Phillips AJL, Alves A, Abdollahzadeh J, Slippers B et al. 2013 – The Botryosphaeriaceae, genera and species known from culture. Studies in Mycology 76, 51–167.
Phillips AJL, Hyde KD, Alves A, Liu JK. 2019 – Families in Botryosphaeriales, a phylogenetic, morphological and evolutionary perspective. Fungal Diversity 94, 1–22.
Puillandre N, Lambert A, Brouillet S, Achaz G. 2012 – ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Molecular Ecology 21, 1864–1877.
Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018 – Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology 67, 901.
Rodriguez-Murillo L, Salem RM. 2013 – Insertion/Deletion Polymorphism. In: Gellman M.D., Turner J.R. (eds) Encyclopedia of Behavioral Medicine. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-1005-9_706
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, 539–542.
Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. 2019 – MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Research 47, W5–W10.
Sayers EW, Beck J, Brister JR, Bolton EE et al. 2020a – Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 48, 9–16.
Sayers EW, Cavanaugh M, Clark K, Ostell J et al. 2020b – GenBank. Nucleic Acids Research 48, 84–86.
Schesser K, Luder A, Henson JM 1991 – Use of polymerase chain reaction to detect the take-all fungus, *Gaeumannomyces graminis*, in infected wheat plants. Applied and Environmental Microbiology 572, 553–556.
Schoch CL, Seifert KA, Huhndorf S, Robert V et al. 2012 – Nuclear ribosomal internal transcribed spacer ITS region as a universal DNA barcode marker for Fungi. Proceedings of the National Academy of Sciences of the United States of America 109, 6241–6246.
Sela I, Ashkenazy H, Katoh K, Pupko T. 2015 – GUIDANCE2, accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Research 43, W7-W14.
Silvestro D, Michalak I. 2012 – RaxmlGUI, a graphical front-end for RAxML. Organisms Diversity & Evolution 12, 335–337.
Simpson MG, Michael G. 2010 – Chapter 1 Plant Systematics, an Overview. Plant Systematics, 2nd ed., Academic Press.
Stamatakis A. 2006 – Phylogenetic models of rate heterogeneity, a high performance computing perspective. In Proceedings of IPDPS2006, HICOMB Workshop, Proceedings on CD, Rhodos, Greece.
Stamatakis A. 2014 – RAxML version 8, a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 301, 312–1313.
Stamatakis A, Hoover P, Rougemont J. 2008 – A rapid bootstrap algorithm for the raxml web servers. Systematic Biology 57, 758–771.
Swofford DL. 2003 – PAUP*, Phylogenetic Analysis Using Parsimony, * and Other Methods, Version 4.0b10, Sinauer Associates, Sunderland.
Talavera G, Castresana J. 2007 – Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology 56, 564–577.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013 – MEGA6, molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30, 2725–2729.

Taru S, Shukla D. 2017 – What is a phylogenetic tree. (http://iammdelhi.com/wp-content/uploads/2017/09/What-is-a-phylogenetic-tree-Dr-Taru-and-Dr-Shukla-1.pdf)
Valones MAA, Guimarães RL, Brandão LAC, de Souza PRE et al. 2009 – Principles and applications of polymerase chain reaction in medical diagnostic fields, a review. Brazilian Journal of Microbiology 40, 1–11
Weiß M. 2010 – Molecular Phylogenetic Reconstruction. Institut für Evolution und Ökologie Organismische Botanik.
Wijayawardene NN, Crous PW, Kirk PM, Hawksworth DL et al. 2014 – Naming and outline of Dothideomycetes-2014 including proposals for the protection or suppression of generic names. Fungal Diversity 69, 1–55.
Wijayawardene NN, Hyde KD, Wanasinghe DN, Papizadeh M et al. 2016 – Taxonomy and phylogeny of dematiaceous coelomycetes. Fungal Diversity 77, 1–316.
Wijayawardene NN, Hyde KD, Al-Ani LKT, Tedersoo L et al. 2020 – Outline of Fungi and fungus-like taxa. Mycosphere 11, 1060–1456.
Wilberg EW. 2015 – What's in an outgroup? The impact of outgroup choice on the phylogenetic position of *Thalattosuchia* (*Crocodylomorpha*) and the origin of *Crocodyliformes*. Systematic Biology 64, 621–637.
Zhang SN, Hyde KD, Jones EG, Jeewon R et al. 2019 – *Striatiguttulaceae*, a new pleosporalean family to accommodate *Longicorpus* and *Striatiguttula* gen. nov. from palms. MycoKeys 49, 99–129.
Zhao RL, Zhou JL, Chen J, Margaritescu S et al. 2016 – Towards standardizing taxonomic ranks using divergence times – a case study for reconstruction of the *Agaricus* taxonomic system. Fungal Diversity 84, 43–74.