Guidelines for DNA-derived phylogenetic trees for Field Mycology articles

1. What is a phylogenetic tree?

A phylogenetic tree is a branching diagram that represents the evolutionary history and relationships between different taxa or groups showing how they've diverged from common ancestors over time. Each branch point indicates a divergence with the term "clade" used to define any group of taxa derived from a single common ancestor. Smaller clades can exist within larger ones.

A tree may include an outgroup which helps to determine the oldest common ancestor of the group of interest. The outgroup is selected to be closely related to, but outside, this group.

Trees usually show bootstrap (BS) values at each branch point. These are values derived by the tree-building program with repeated sampling of the dataset and are usually expressed as a percentage which provides a measure of confidence in the result of the analysis. Values of 70% or more are often regarded as acceptable with lower values excluded from the tree.

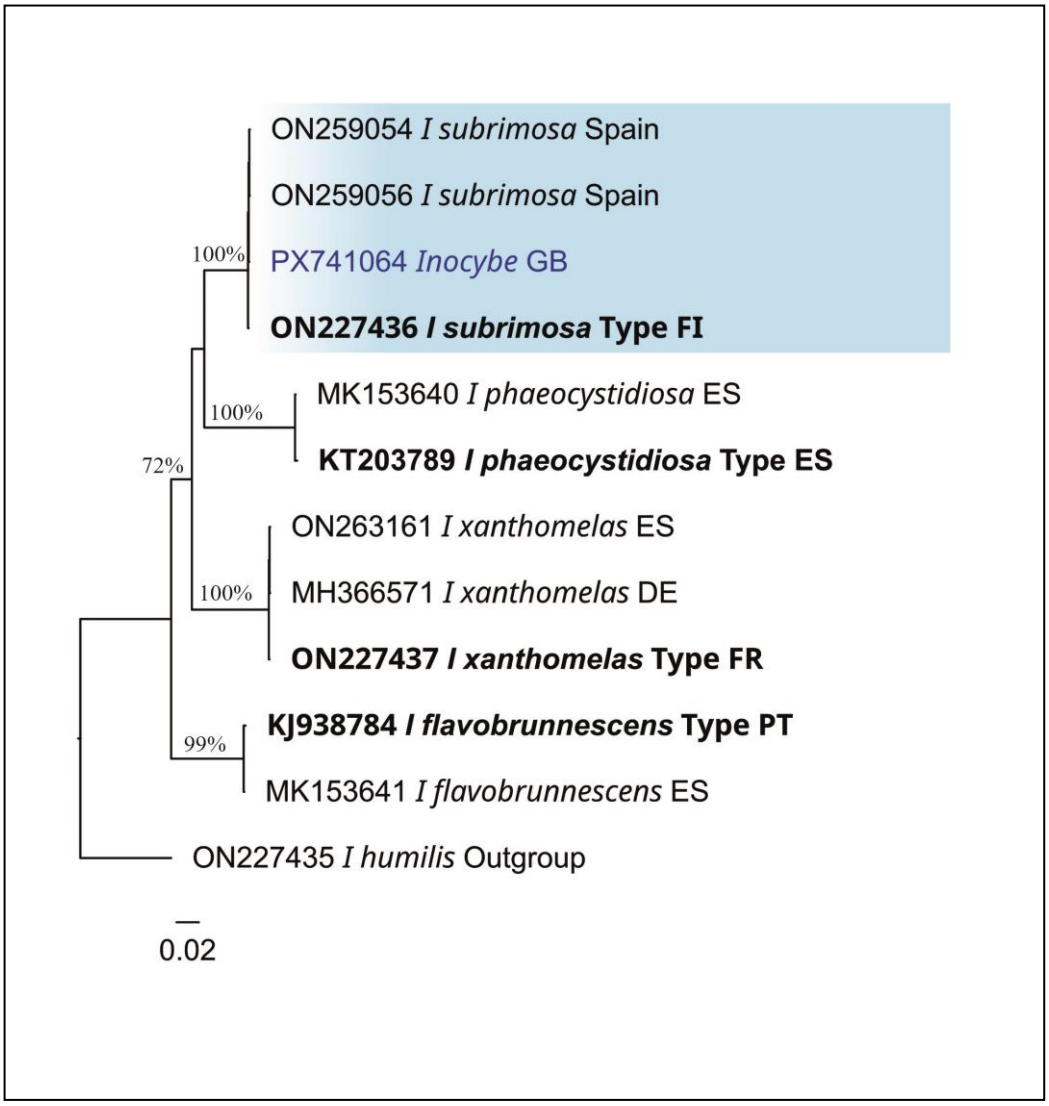

Figure 1. Example tree.

A typical caption for a phylogenetic tree would be:

Figure 1. A maximum likelihood phylogram inferred from ITS sequences. Nodes are labelled to indicate branch support (shown as a bootstrap percentage) where this exceeds 70%. Sequences obtained in this study are shown in blue. The scale bar indicates the number of substitutions per site.

In this example, the term maximum likelihood (ML) relates to the analysis method used to generate the tree. There are several different methods, each with strengths and weaknesses, including assumptions made about the data and the speed of computation. ML is a commonly used method for rapidly generating trees, along with neighbour-joining (NJ) and maximum parsimony (MP).

Substitutions per site is shown as a horizontal scale bar at the bottom of the chart and provides a measure of evolutionary distance. The scale illustrates how many nucleotide changes have occurred per position in the sequence along a tree branch or between two taxa.

2. Purpose of a phylogenetic tree

Trees generated for articles in Field Mycology usually serve one of two purposes:

(a) To illustrate the position of one or more species of interest in relation to close relatives.

This can be useful when discussing previously described species as new to Britain or in other cases where a particular feature or distribution of a species is being covered.

(b) To support an argument which involves the description of a new species or changes our interpretation of a genus.

Although analyses based on a single barcode region can provide useful information to support fungal taxonomy studies, better information can be obtained by combining sequences from at least two or three genes to obtain a more robust phylogeny. Examples of different regions used for fungal DNA include:

Internal Transcribed Spacer region (ITS1, 5.8S, ITS2)
RNA polymerase II second largest subunit (rpb2)
nuclear ribosomal Large SubUnit (nrLSU-28S)
nuclear ribosomal Small SubUnit (nrSSU-18S)

Analyses of this type would usually be part of a collaborative effort involving specialists in the field and are outside the scope of this guidance document.

3. Sequences used in the analysis

The tree should be inferred from sequences drawn from public databases together with any query sequences generated in the current work. Where possible, sequences for type material should be included and labelled as such. The most commonly used sequence databases are **GenBank** (maintained by the National Center for Biotechnology Information, USA) and **Unite** (created by an international collaboration of research institutes, hosted by University of Tartu, Estonia). Note that a significant number of sequences in GenBank may be incorrectly labelled – ideally the sequences should be selected from a research paper which includes a phylogenetic tree with cross-references to source sequences. Unite can be particularly helpful as in some cases an entry will be designated

as either a "reference" or a "representative" sequence meaning that there some measure of confidence in the species identity of a sequence. Ensure that the downloaded sequences cover the same barcode region as the one used in your study.

All sequences should include the accession number from the source database (e.g., KJ938784 in Figure 1). It may be helpful to include a name (or a two or three letter country code from ISO3166) to identify the country of origin.

https://www.iso.org/obp/ui/#search select Country codes

Phylogenetic studies often contain a considerable number of sequences with trees sometimes extending over more than one page. For Field Mycology articles the aim should be to generate a tree with just sufficient sequences to support the article. As an approximate guide the tree could be based on 4 to 10 taxa, each supported by 1 to 4 sequences.

4. Alignment of sequences and tree generation

Prior to generating a tree, all sequences being used need to be aligned. Alignment is where the common blocks of DNA nucleotides are lined up irrespective of the start position of any sequence; these common blocks represent common ancestry and provide the framework for comparison of sequences and identification of differences that relate to determining speciation. Poor alignment of a query sequence can indicate that it is of poor quality and is of limited utility or that a public sequence in the selection used is of poor quality and should be removed from the analysis; the alignment visualisation helps assess this.

Various free resources exist for the alignment of sequences. **AliView** and **SeaView** are widely used programs that can be downloaded onto a computer. Two useful programs used for inferring a tree from aligned sequences are **FastTree** and **RaxmlGUI**: the former is fast and easy to use whilst the second is more widely used for specialist studies. **Phylogeny.fr** is an online resource that uses the MUSCLE alignment tool and provides a number of different methods for generating trees.

The **FigTree** program can be used to display and edit the tree. If required, the tree can be imported into presentation or desktop publishing programs for further enhancement. For more extensive work, the **Inkscape** open-source vector graphics editor can be very useful for handling files in SVG format and provides good editing features which, for example, facilitate removal of branch support values, replacement of underscores in sequence labels with spaces, and changes to fonts, line weight and colours to enhance the clarity of the displayed tree. On completion of any adjustments, the tree should be exported as a JPEG or PNG file with a resolution of not less than 300 pixels/inch.

5. Links

https://www.ncbi.nlm.nih.gov/genbank
https://unite.ut.ee
https://ormbunkar.se/aliview
https://doua.prabi.fr/software/seaview
https://github.com/morgannprice/fasttree
https://github.com/AntonelliLab/raxmlGUI
https://www.phylogeny.fr
https://github.com/rambaut/figtree/releases
https://inkscape.org