JH Bioinformatics Note – 1

Mycena Revisited

"It is not difficult to make a Phylogenetic Tree"

Jurrie Hubregtse

October 8, 2020

Cruentomycena viscidocruenta

Introduction

To start with, I must point out that I am not an expert in bioinformatics, nor am I a biologist; however, I am interested in fungi, and recently my interest has been directed towards the phylogenetic relationships between fungal species. I have been retired for the past 20 years. During my working career I carried out research into telecommunications networks and associated components, and in this capacity I have used and written computer models to assist me. I still have some computer skills, occasionally write some of my own software, and prefer to use a computer with a Linux operating system.

Late last year I offered to give a talk to the FNCV Fungi Group with the title "Making phylogenetic trees is not difficult". The aim of my presentation was to show that making a phylogentic tree is not rocket science and if someone really wanted to study phylogenetic analysis the skills required can readily be mastered. Due to the impact of COVID-19 I was not able to give that talk, so instead I decided to explore this statement a bit more and write it down as a Note. What prompted me to propose this talk was a number of totally inadequate phylogenetic trees I came across in some fungal taxonomy based journals: some trees did not illustrate what the author claimed, and some were no better than decorations. Although these poorly produced phylogenetic trees were largely the exception, this did make me wonder how difficult it would be to make a phylogenetic tree.

As a non-expert I will outline the steps I have taken to produce several phylogenetic trees which infer evolutionary relationships between a number of species (in this case fungi). At the outset the whole process may look daunting, but when it is broken up into its basic steps each step is relatively easy to perform.

It must also be emphasised that a phylogenetic tree is produced for a purpose, which is to illustrate inferred evolutionary relationships between members of a group of species so that a conclusion based on their inferred relationships can be drawn.

I tend to think that when I am reconstructing inferred phylogenetic relationships from a set of gene sequences I am conducting a scientific experiment for the purpose of illustrating their relationships over evolutionary time. As with all good science, scientific experiments should be repeatable. To illustrate that it is not difficult to produce a phylogenetic tree I propose to describe the process, and repeat the construction of a tree I have found in the literature, then add some variations to see if the result is still the same.

Basic Steps Involved in Estimating a Phylogenetic Tree

It is necessary to understand that the evolutionary relationship between two or more organisms implies that they share a common ancestry (i.e. they are homologous). To be able to infer this relationship it is necessary to obtain sequence data from the DNA of each individual; this data represents real measurements taken from real specimens. I will refer to these organisms as 'species', which in this instance are defined as a conceptual grouping of

organisms. The DNA obtained from specimens is sequenced and the result is a long string of text consisting of four letters, A, C, G and T, which symbolically represent the four different nucleotides present in the sequenced string of DNA. Examples of sequence data can be found in Appendix B.

Evolutionary relationships cannot be directly observed, but they can to be deduced from the collected sequence data. By applying statistical techniques to the data, it is possible to estimate an evolutionary relationship (an inferred phylogenetic tree); it is possible to estimate its accuracy; and it is also possible to improve the estimate by adding more data, but we will never be able to produce the one and only true phylogenetic tree. Fortunately, a statistically well supported estimate – an inferred phlyogenetic tree – can be a very useful tool in aiding our understanding of how various species (fungi in this case) are related. Also very fortunately, the statistical tools needed to carry out the various processes required to produce a phylogenetic tree are readily and freely available via the WEB.

Next I will outline the four basic steps that are required to make a phylogenetic tree. I will not try to describe each step in detail as there is abundant information available on the WEB. More details on the various software packages needed can be found in Appendix A. None of the procedural steps are particularly difficult.

Step 1: Identify and acquire sequences

The most common practice for selecting sequences for making phylogenetic estimates is to search the NCBI (National Center for Biotechnology Information) database by opening the BLAST (Basic Local Alignment Search Tool) search page on the NCBI website (https://www.ncbi.nlm.nih.gov/BLAST/). On the BLAST search page you can either enter a sequence or a GenBank ID as a "query" to search a global database for closely related sequences. Detailed information on how to use the NCBI website can be obtained from the help button on the top right-hand corner of the BLAST search page.

The other common practice for obtaining sequence IDs is from published phylogenetic reconstructions found in the literature. If you have some GenBank IDs but you do not require to do any BLAST searches, then using the NCBI website (https://www.ncbi.nlm.nih.gov/nucleotide) would be the quickest way to retrieve the sequences.

It is preferable to download sequences in FASTA file format, which is compatible with most alignment tools. The FASTA file format is a text file, which can be easily edited and concatenated using a text editor (not a word processor). See Appendix B: for an example of a sequence saved in FASTA file format.

I have just described the easy part, which is acquiring and manipulating sequence files. The hard part is to decide what problem you are trying to solve and which sequences you need to select in order to make a phylogenetic tree that will give you the evolutionary inference you need. I checked the Materials and Methods section of a number of articles containing phylogenetic trees to see which methods the authors employed to select the sequences they used, but unfortunately there was a void on the subject. I suspect that most sequences are either selected by using trial and error, or are taken from reference articles

covering a similar taxonomic topic. There does not seem to be an accepted protocol for sequence selection. There is nothing wrong in using trial and error – it is an accepted methodology. From my experience I found that using multiple sequences of the same species in critical areas of a phylogenetic inference to be useful in minimising errors.

A phylogenetic tree is a representation of the genealogical relationships between species, which is best achieved with a rooted tree. A rooted tree that shows the direction of evolution will allow inferences to be made about the order of descent. This is achieved by including some distantly related species, commonly called the outgroup, to act as the root of your tree.

Step 2: Multiple Sequence Alignment (MSA)

While Step 1 is considered to be the most difficult, Step 2 would be considered the most critical. The aim of this step is to obtain a signal inferring common evolutionary history between a matrix of homologous sequences. In this process the sequences are concatenated and then aligned in such a way that the homologous sites form columns. Obtaining the one and only true alignment is virtually an impossibility because the number of permutations and combinations that need to be tested even for a moderate number of sequences is huge, and the number of calculations required is astronomically large. To solve this problem a number of software packages have been developed that employ statistical algorithms to make a good estimate of the MSA. The resultant phylogenetic signal that is obtained from a matrix of aligned sequences can be used to estimate a phylogenetic tree. The stronger the signal the more robust the tree. Obviously any errors in the MSA estimation tend to produce errors in estimated trees and in any other downstream analysis.

Step 3: Estimate the phylogenetic tree

This step is possibly the least critical; this is where the phylogenetic signal is detected and analysed. There are four primary approaches used to analyse aligned sequences. These are :-

Maximum Parsimony Methods: Maximum Parsimony Methods seek the tree or trees (there may be more than one tree topology that fits the maximum parsimony condition) that require the least number of changes to explain the differences between the observed sequences. This method does not use any evolutionary models in weighting changes observed in the sequence data.

Distance-based Methods: Distance-based Methods seeks the tree in which the branch lengths correspond as closely as possible to the observed evolutionary distances between sequences in the alignment. The main distance-based methods include the neighbor-joining (NJ) and unweighted-pair group method with arithmetic mean (UPGMA). Distance-based methods of phylogeny are computationally fast, so they are particularly useful for analyses of a large number of sequences.

Maximum Likelihood Method: Maximum Likelihood (ML) seeks the tree that maximizes the likelihood of observing the given data (sequences).

One of the advantages of this method is that the likelihood of the resulting tree is known. This method is computationally intense, as a large number of trees need to be considered before converging on the maximum likelihood tree.

Bayesian Inference Method: Bayesian Inference (BI) is a variant of Maximum Likelihood (ML). Where ML seeks to find a tree that maximizes the likelihood of observing the data (sequences), BI seeks those trees with greatest likelihoods given the data (sequences). This method converges onto a set of trees with roughly equal likelihoods. The results of Bayesian analysis of these trees can be readily interpreted because the frequency of any clade in any set of trees is virtually equal to the probability of that clade existing. Therefore no bootstrapping is required to assess the confidence of the tree topology.

Out of these four primary analytical approaches, all except those using parsimony methods require the use of an explicit evolutionary model. An evolutionary model is used in the calculation of the genetic change between an ancestor and its descendant, and it is the measure of this genetic change that determines the branch lengths of a phylogenetic tree.

There are a large number of models to choose from and it is preferable to choose an optimal model that best fits the aligned sequences. There are two criteria that can be used statistically to determine the relative fit of alternative competing models. These are the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The evolutionary model that best fits the aligned sequences is the one with the lowest AIC or BIC score. For this exercise I have chosen to use the BIC score.

Fortunately there are a number of software packages available to determine the optimum evolutionary model for aligned sequences. MEGA X has a model finder option; IQ-TREE, a phylogenetic tree inference software package, has a model finder built in; or you can use a dedicated model finder software such as PartitionFinder.

At present the two most commonly used approaches for producing phylogenetic trees are the Maximum Likelihood (ML) and Bayesian Inference (BI) methods. Both of these methodologies require an evolutionary model. There are a number of software packages that can be used to produce phylogenetic trees based on the ML method, but for the BI method the most commonly used software package is called MrBayes. All these software packages have comprehensive manuals and are not particularly difficult to implement.

When producing a phylogenetic tree using an ML method it is possible to assess the robustness of the phylogenetic inference produced (note that robustness is not the same as accuracy, which depends upon the included species). Robustness of a phylogenetic tree is most commonly tested by using the bootstrap method. In bootstrapping, a resampling of the alignment is used to build new trees, and the bootstrap value is calculated. This value represents the branch support as a percentage. The higher the percentages the more robust the tree. Most ML software packages have the bootstrap option; for best results set the resample number into the hundreds or thousands. The returned percentages should give a good indication of the tree's robustness.

Step 4: Draw the tree and assess the result

The output from all phylogenetic inference software is a text file that contains the description of the phylogenetic tree. It needs to be interpreted and then plotted. There is a plethora of software applications available to allow you to plot and visualise your phylogenetic tree. FigTree is the application I chose because it is convenient, intuitive to use, and compatible with the file formats I was using.

An approach to assessing the degree of confidence you can have in your phylogenetic inference would be to use several methodologies to produce it and, if the results are relatively consistent, this should boost your confidence in the relative accuracy of your phylogenetic inference.

Materials and Methods

The ITS gene (internal transcribed spacer regions ITS1, ITS2, with the 5.8S gene) used for the phylogenetic reconstructions is the same one used by Petersen *et al.* (2008). The species chosen were from published topologies in Petersen *et al.* (2008), Chew *et al.* (2013, 2014) and from select matches of BLAST searches at GenBank.

For more details on the software packages used see Appendix A.

Method used to produce the phylogenetic tree shown in Fig. 2 . To construct this phylogenetic tree, the MEGA X software package was used. The alignment tool Muscle with its default settings was used to perform the multiple sequences alignment (MSA). MEGA X's "MODELS-Find Best DNA–Protein Models" option was used to find the best-fit model of evolution, the HKY+I (Hasegawa-Kishino-Yano model + invariant sites) model was recommended because it had the lowest BIC score, implying it would best fit the aligned sequences. A Maximum Likelihood phylogenetic tree was produced, incorporating both the HKY+I evolutionary model and the bootstrap method (100 bootstrap replications), which were used to determine branch confidence scores (this gives the reliability of the inferred tree). The tree was viewed using MEGA X's tree editor and viewer, but the final version as seen in Fig. 2 was edited in FigTree.

Method used to produce the phylogenetic tree shown in Fig. 3 . To construct this phylogenetic tree, T-Coffee was used for the MSA. Then IQ-TREE was used to produce a maximum likelihood phylogenetic tree. This software package has an incorporated Evolutionary ModelFinder, which selects the best-fit model based on the lowest BIC score. The chosen model TN+F+G4 was used. Since IQ-TREE has an ultrafast bootstrap algorithm, a bootstrap of 1000 samples was used. The resultant tree was viewed and edited using FigTree.

Method used to produce the phylogenetic tree shown in Fig. 4 . To construct this phylogenetic tree, T-Coffee was used for the MSA and also to trim columns that are more than 80% blank. To determine the optimum evolutionary model the aligned sequences were analysed

using PartitionFinder2. The GTR+I+G (Generalised Time Reversible + Invariant sites + Gamma distributed rates) model had the lowest BIC score, making it the model of choice. IQ-TREE with its model option set to GTR+I+G, and with the bootstrap option set for 1000 replicates, was used to infer the phylogenetic tree. The resultant tree was viewed and edited using FigTree.

Method used to produce the phylogenetic tree shown in Fig. 5 .The previously used alignment and evolutionary model (GTR+I+G) was used used. RAxML (Randomized Accelerated Maximum Likelihood) version 8.2.12 was used to estimate the phylogenetic tree. The evolutionary model option was set to GTR+I+G and the bootstrap option was set to 1000 rapid bootstrap replicates, and thereafter a thorough ML search was used to determine the final phylogenetic tree. The resultant tree was viewed and edited using FigTree.

Method used to produce the phylogenetic tree shown in Fig. 6 . The previously used alignment and evolutionary model (GTR+I+G) was used. MrBayes version 3.2.7a was used to estimate the phylogenetic tree. It was set to run 100,000 generations with a burn-in fraction set to 25%. The tree produced was viewed and edited using FigTree.

Mycena Revisited – repeating an experiment

The very nature of a scientific experiment is that it can be repeated, and if all goes well the repeated experiment should produce similar results. The experiment that I proposed to repeat, using 3 different tree-building methods, is described in the paper by Petersen *et al.* (2008). The particular experiment is shown in Fig. 1, where *Mycena viscidocruenta* is shown not to be a *Mycena* and therefore should be placed in the newly created genus *Cruentomycena*.

Some observations concerning this phylogenetic tree: it has 3 specimens of *Mycena pura* as its outgroup plus 27 specimens of *Panellus stypticus* [sic]. This sort of arrangement is consistent for a phylogenetic tree constructed to infer that *Panellus stipticus* specimens found in different countries all belong to the same species. This it does convincingly. By including some species of *Cruentomycena* it also intends to fulfill the purpose of inferring that this genus is not directly related to species in the genus *Mycena*, which is represented by the outgroup *Mycena pura*.

When reconstructing this phylogenetic tree as shown in Fig. 1, some changes needed to be introduced. See Fig. 2 for the new configuration. Firstly, an outgroup of *Collybia cirrhata* specimens, marked in , was added; these would be the new root for the tree. By placing the root of the tree in a closely related family, Tricholomataceae, we will be able see the direction of evolution, and improve resolution between various genera within the Mycenaceae. Secondly, some extra species closely related to *Mycena pura* were added. These changes were made so that the phylogenetic tree would better portray the phylogenetic position of the genus *Cruentomycena* with respect to the genus *Mycena*. Most of the sequences used in Fig. 1 were reused. Genbank ID's for the extra sequences were obtained from articles by Chew *et al.* (2013, 2014).

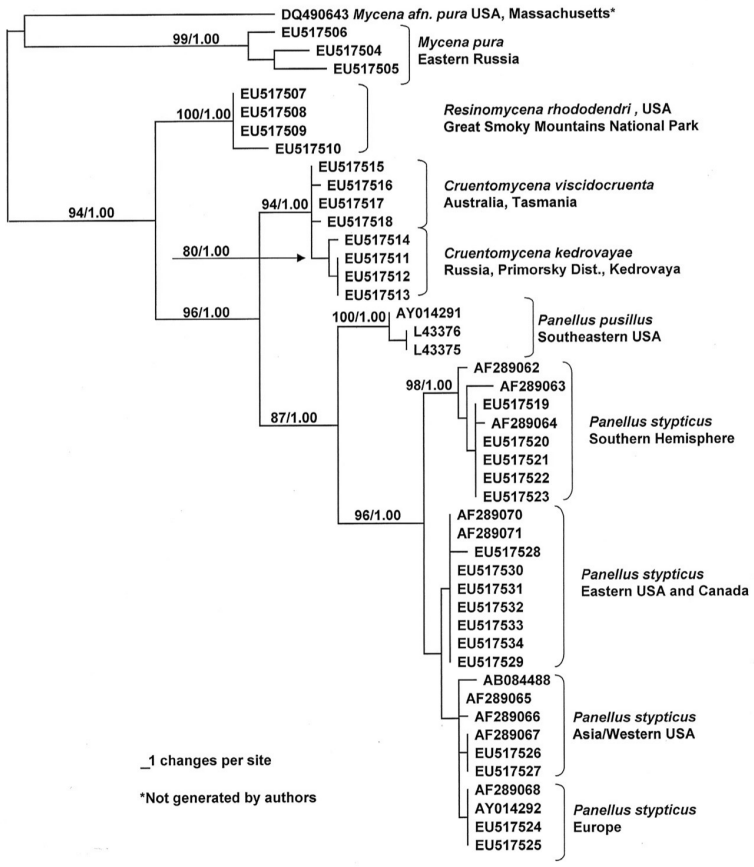

Figure 1: This phylogenetic tree is a direct copy of Fig. 5 from the paper by Petersen *et al.* (2008). Although the accompanying text says that the sequences are from the ribosomal large subunit (LSU) gene, they are actually all from the ITS gene.

The reconstructed phylogenetic tree (Fig. 2) is very similar to that in Fig. 1. The clade marked in is the *Mycena pura* group (*Mycena* sect. Calodontes), which is shown, with a high degree of confidence, to be a separate clade from the one marked in , with respect to the genus *Cruentomycena*. This is in full agreement with Petersen *et al.* (2008).

Normally this would settle the argument that *Mycena viscidocruenta* is not a *Mycena*. But since *Mycena viscidocruenta* initially satisfied the *Mycena* species concept (using the definition that a species is just a conceptual grouping of organisms) we need to be certain that *Mycena pura* and its closely related species are actually *Mycena*'s. To do this another group of 9 *Mycena* sequences have been included in the experiment. Most species within this group of *Mycena*'s exhibit bioluminescence.

The phylogenetic tree shown in Fig. 3 implies quite convincingly that *Mycena viscidocruenta* is not a species of *Mycena*. There are 2 *Mycena* clades, one shown in and the other shown in with a total of 18 *Mycena* species to which the *Cruentomycena* clade shown in is distinctly not related. Normally at this stage some would say Q.E.D ("quod erat demonstrandum", Latin for "that which was to be demonstrated") but for one small thing. *Mycena viscidocruenta*

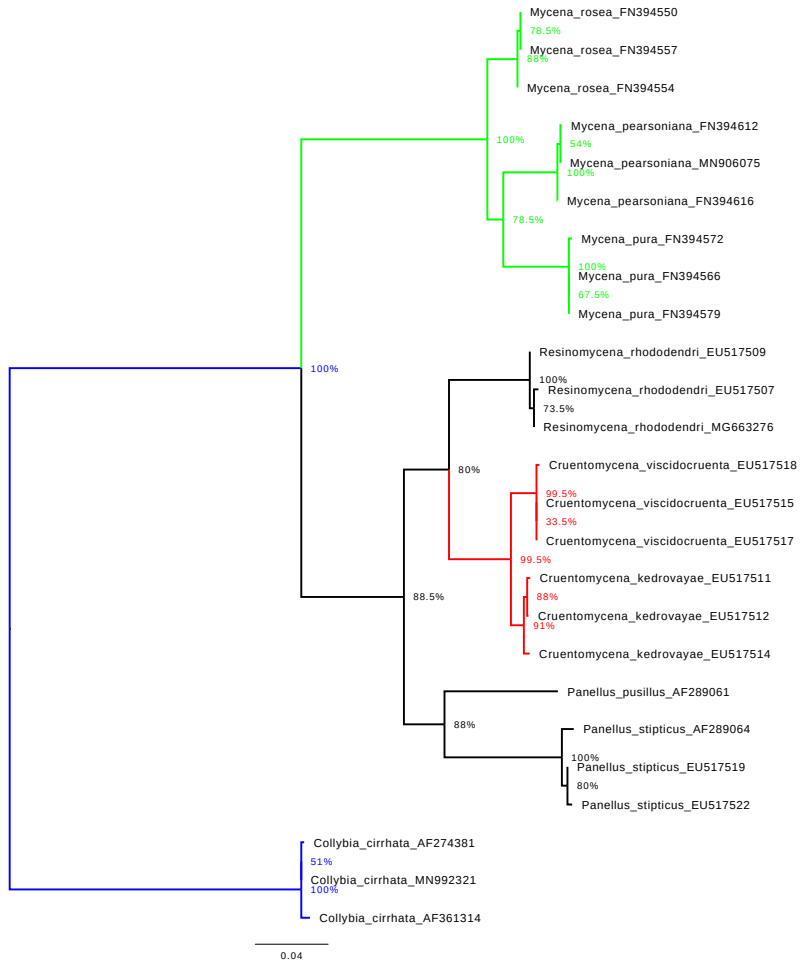

Figure 2: This phylogenetic tree was produced using the MEGA X software package. The sequences were aligned using Muscle and the tree was produced using the Maximum Likelihood method.

has not been compared with the genus *Mycena* type species *Mycena galericulata*, and it has not been convincingly shown that *Mycena viscidocruenta* actually needs its own genus, because it is still possible that it may be compatible with an already existing genus in the Mycenaceae. To address these issues I included some extra *Mycena*'s covering a broader spectrum of species including *Mycena galericulata* plus a number of species belonging to some closely related genera within the Mycenaceae. The next phylogenetic tree, shown in Fig. 4, is significantly different.

The phylogenetic tree in Fig. 4 clearly illustrates that *Cruentomycena* is a robust and a well defined clade, but at the same time it strongly infers that the genus *Mycena* is not a monophyletic group. That the genus *Mycena* is not monophyletic is well known, and was also illustrated by Chew *et al.* (2014). This clearly indicates that the species concept that describes the genus *Mycena* needs to be re-evaluated. If the genus *Mycena* is one day re-evaluated and the clade containing *Mycena galericulata* (the type species for *Mycena*) remains a *Mycena*, then the *Mycena pura* and *Mycena chlorophos* clades would most likely no longer be part of the genus *Mycena*.

Although Fig. 4 shows that Petersen *et al.* (2008) was correct in claiming

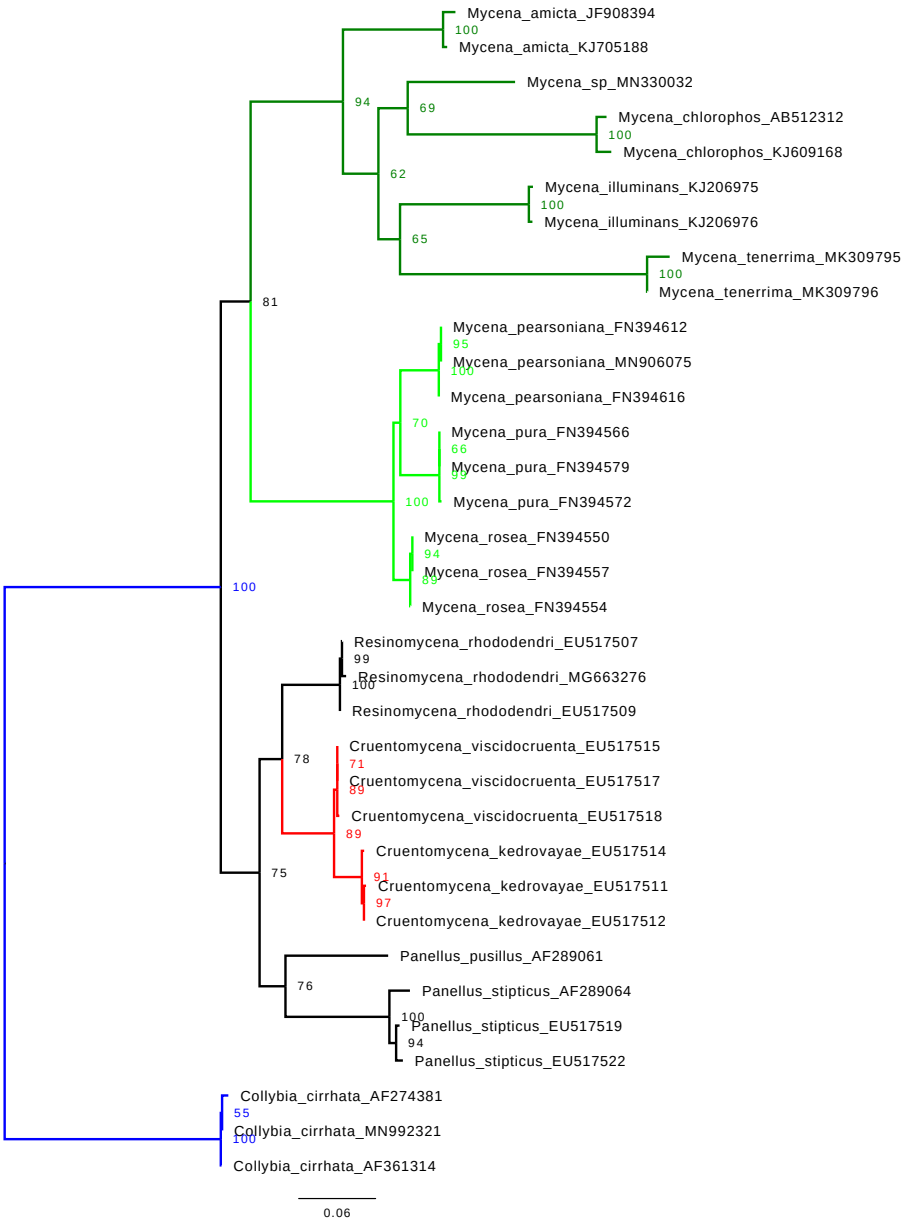

Figure 3: This phylogenetic tree was produced using the alignment software package T-Coffee and the IQ-TREE software package.

that *Mycena viscidocruenta* is not a *Mycena* and that it belongs in its own genus *Cruentomycena*, the phylogenetic inference that was supplied could not have supported their argument.

The phylogenetic tree in Fig. 4 was produced using the IQ-TREE software package, which uses a fast stochastic algorithm to reconstruct ML based inference phylogenetic trees. To test its robustness, a further two trees were produced using software that employed different ML algorithms. The next tree, in Fig. 5, was produced using RAxML. The final tree, in Fig. 6, was produced using MrBayes. It is quite evident that the three phylogenetic trees depicting the spread of species in the Mycenaceae all have similar topologies, which infers that these trees are robust interpretations of phylogenetic inference contained in the aligned sequence data.

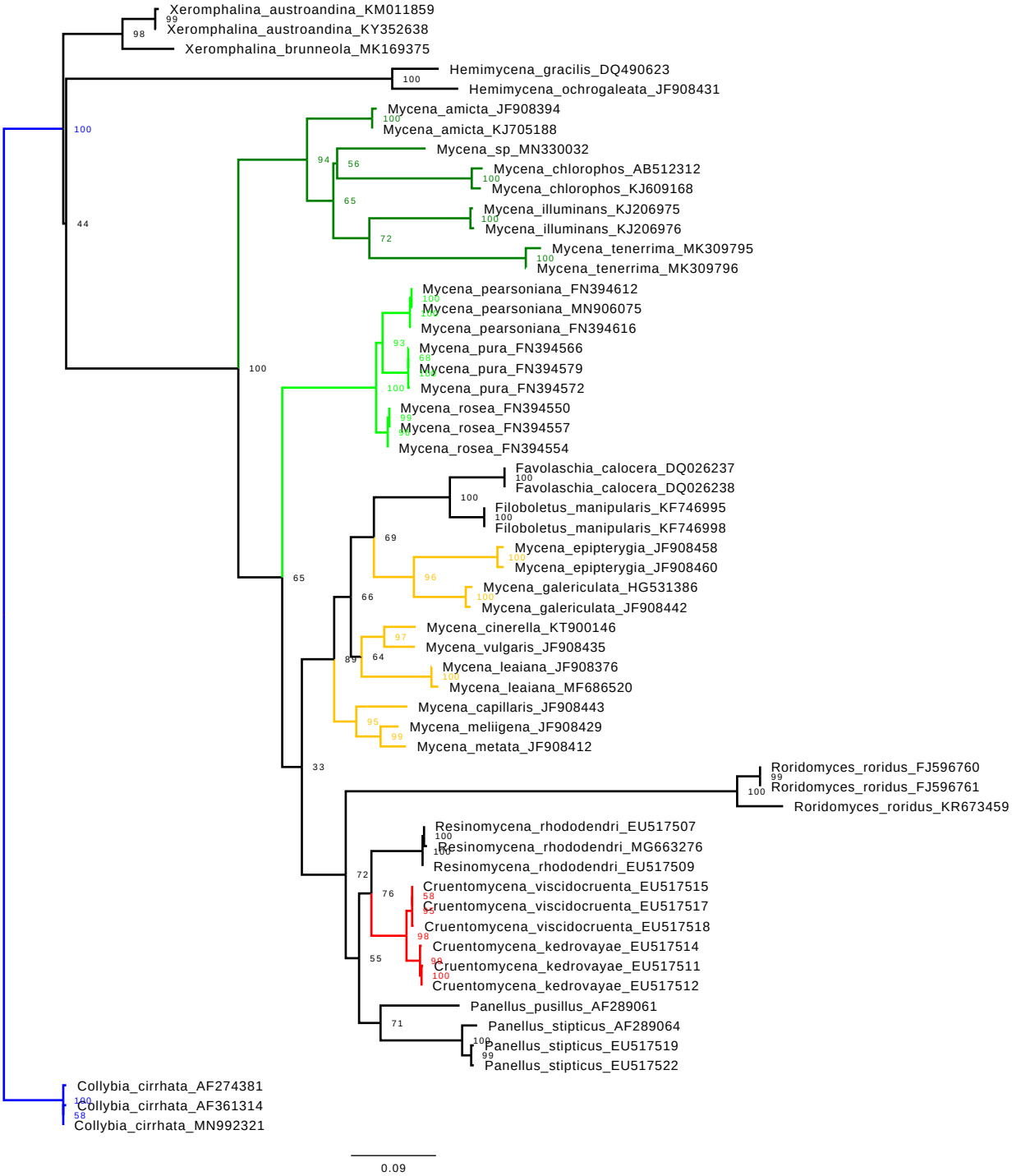

Figure 4: This phylogenetic tree was produced using the alignment software package T-Coffee and the IQ-TREE package.

Discussion

Figures 4, 5 and 6 clearly show that all species of *Cruentomycena* form a well supported clade in the family Mycenaceae and that they do not belong in the genus *Mycena*. This also reinforces the fact that *Mycena* is not monophyletic and needs to be re-assessed and possibly broken up into a number of genera.

Let me re-emphasise that a species is a group of organisms based on a

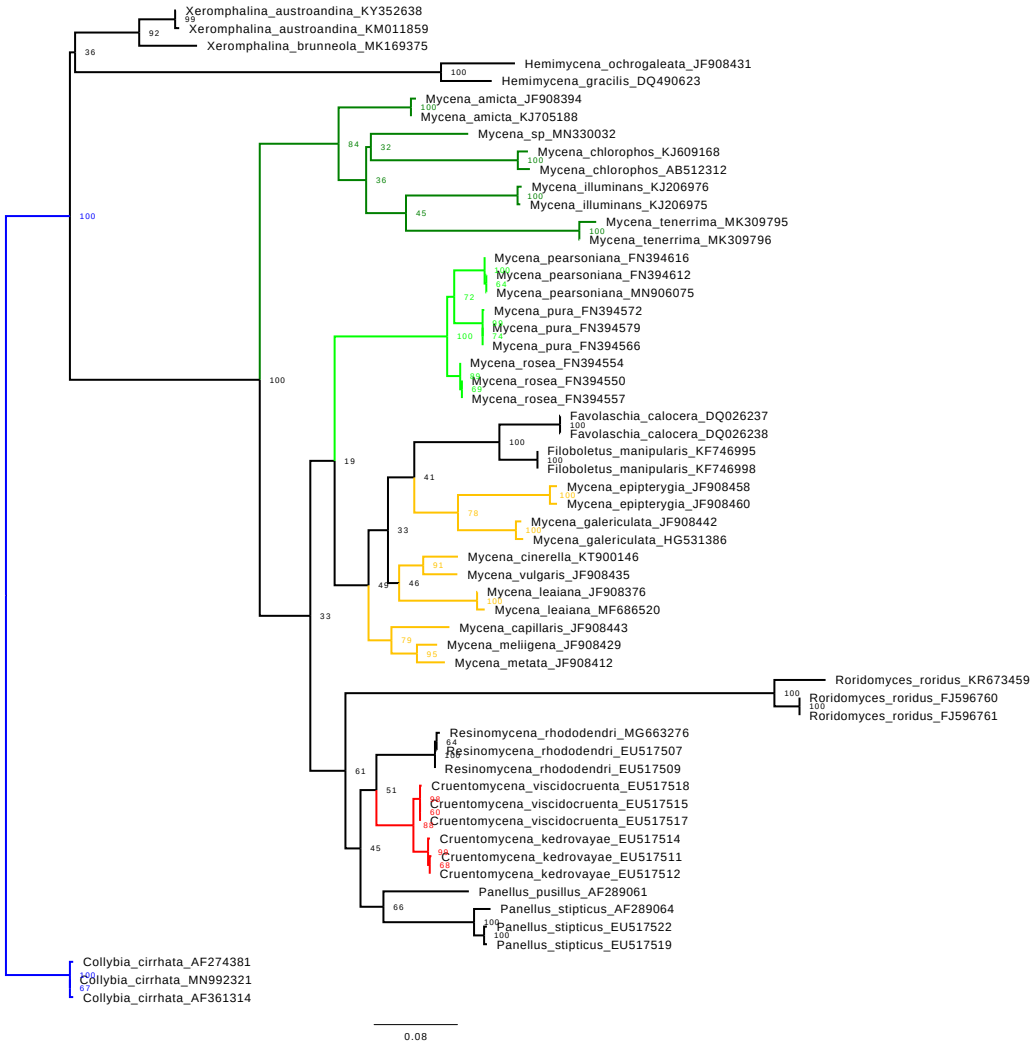

Figure 5: This phylogenetic tree was produced using the alignment software package T-Coffee and the RAxML software package.

species-concept. Many species-concepts use observable morphological similarities as a guide to homology, but for fungi this has been shown to be problematic. Similarity is the observation or measurement of resemblance or difference, independent of its cause. Homology means, specifically, that the sequences and the organisms in which they occur are descended from a common ancestor. By implication a phylogenetic tree shows only ancestral relationships (homology) and its topology is not influenced by observable similarities between the organisms.

Repeating an experiment in phylogenetics may teach us two things. Firstly, do not assume that any phylogenetic tree you see in the literature is correct, and secondly, recreating a phylogenetic tree to satisfy yourself that the inference is robust and supports the authors argument, is not such a difficult exercise.

In an attempt to make the phylogenetic trees in Figures 4, 5 and 6 more relevant to an Australian audience, I intended to include a number of Australian *Mycena* species in the mix, thinking that there should not be a problem in finding ITS sequences since they are the official barcoding markers for species-level identification. **Unfortunately** I was not able to find any ITS sequences of any

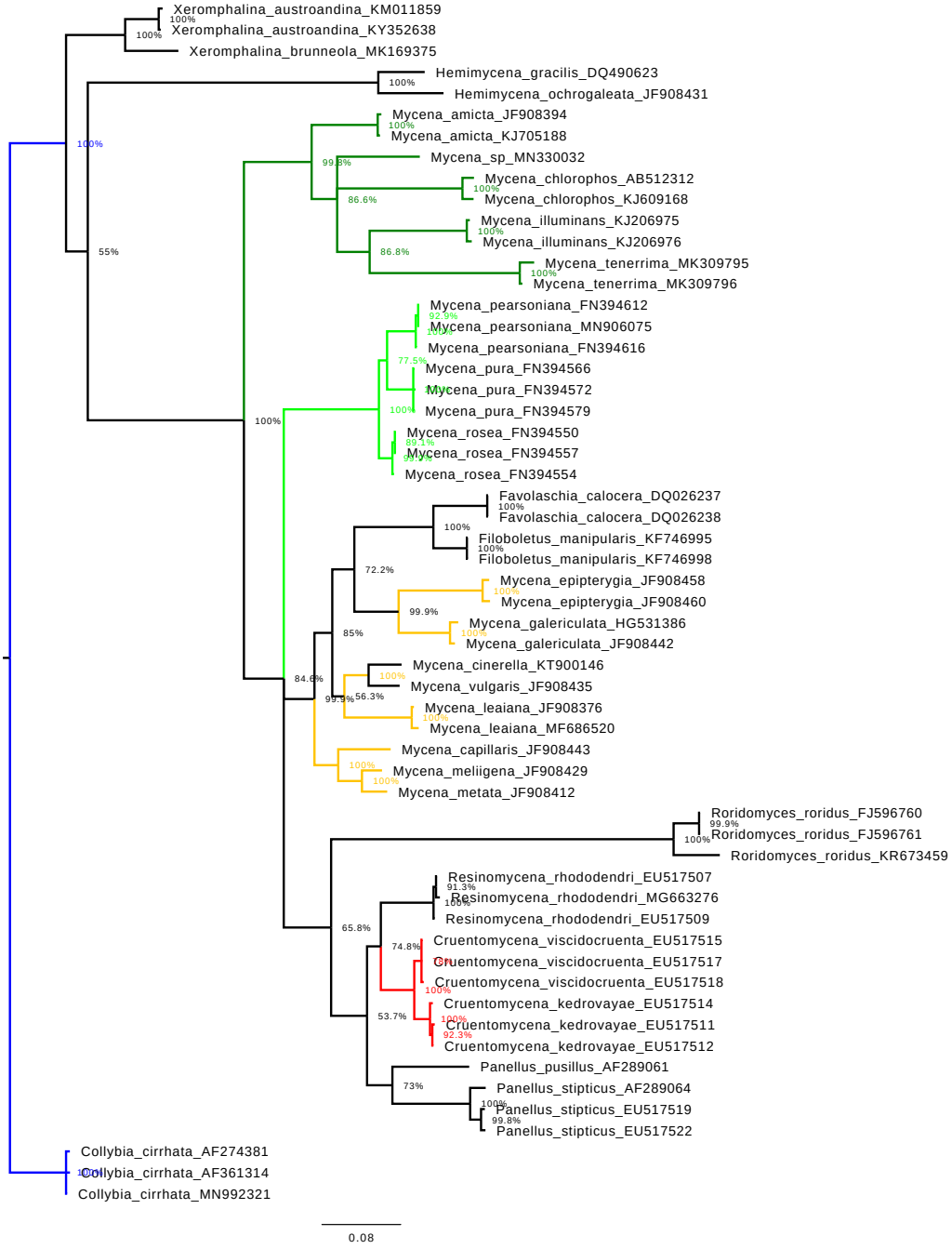

Figure 6: This phylogenetic tree was produced using the alignment software package T-Coffee and the Mr Bayes software package for making the phylogenetic tree.

described (holotype, or type) Australian *Mycena* species – none. I used the NCBI Taxonomy Browser (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Root) and searched for 52 *Mycena* species (see Appendix C for the list) and could not find a single match.

If, as it seems, none of our *Mycena* species have been sequenced and phylogenetically compared with other *Mycena* species, then there is no scientific evidence to say that we have any *Mycena* species in Australia. There is expert

opinion that says we do, but this is not evidence. Likewise, we had expert opinion that there were many *Dermocybe* species in Australia, but phylogenetic evidence showed that they were all *Cortinarius* species.

It is now possible for clubs and individuals to purchase their own sequencers. Nanopore DNA sequencing devices are available for about the same price as a full frame DSLR camera. This is the good part – amateurs can now sequence the fungi they find. The bad part, in the case of *Mycena*, is that there are no official species reference sequences in GenBank to use for comparison. Eventually, sequences uploaded to GenBank by amateurs will become the defacto standard for our native species, whether properly identified or not. It is up to our national institutions responsible for documenting and cataloging our native biota to have sequence data lodged in GenBank or some other accessible database so that our native biota can be correctly identified. If this is not done, there is a significant probability that these institutions will become irrelevant.

Conclusion

I have been able to demonstrate that it is not difficult to reproduce a phylogenetic tree found in the literature. Now, having learned how to construct a phylogenetic tree, you will be able to ask the all-important question, does the tree really demonstrate what the author is implying? And the good part is you can go and test it for yourself. After all, this is the basis for good evidence-based science.

There are numerous ways by which the above phylogenetic trees could have been made. There are websites that would allow you to carry out all the relevant processes without the need to download any software packages. The method chosen above is more a reflection of my background. It is neither the optimum way nor the only way to produce phylogenetic trees.

I did not say or imply that you would be able to do it without the need to learn anything new. There will be a lot to learn, but it is all very interesting and straightforward. You will need to learn about evolutionary change, about the relatedness between species, about what homology implies, and a way of using your computer.

I like to emphasise that the ease with which a phylogenetic tree may be constructed should not be at the expense of understanding what is being done and the limitations of the computational tools. To assist in navigating your way through the various steps required in producing a phylogenetic tree I recommend using an excellent textbook such as “Bioinformatics and Functional Genomics” third edition by Jonathan Pevsner (it is possible to find a website where a PDF version of this book can be readily downloaded). This book contains many useful references to WEB resources, self-tests, mentions common pitfalls, and includes discussion questions, making it an excellent resource for independent study.

All the DNA sequence data, all of the software needed to make phylogenetic trees, and all of the educational material needed, is available for free on the WEB. How easy is that?

References

Chew ALC, Tan Y-S, Desjardin DE, Musa MY and Sabaratnam V (2013) "Taxonomic and phylogenetic re-evaluation of *Mycena illuminans*". *Mycologia* **Vol. 105(5), pp.1325–1335.**

Chew ALC, Desjardin DE, Tan Y-S, Musa MY and Sabaratnam V (2014) "Bioluminescent fungi from Peninsular Malaysia – a taxonomic and phylogenetic overview". *Fungal Diversity* **Vol. 70, pp.149–187.**

Petersen RH, Hughes KW, Lickey EB, Kovalenko AE, Olga V. Morozova OV and Psurtseva NV (2008) "A new genus, *Cruentomycena*, with *Mycena viscidocruenta* as type species". *Mycotaxon* **Vol. 105, pp.119–136.**

Appendices

Appendix A: Software Used

All of the software used in this exercise is in the public domain, and available on the WEB. The computer used to run the software was an HP Z210 workstation. The operating system was Linux, with Kubuntu (a flavour of the Ubuntu operating system) installed. The advantage of using Kubuntu is that it is compatible with most bioinformatic software and will run both GUI and command-line packages. An alternative to Linux is the Macintosh operating system (MacOS), which also offers a Unix/Linux-like command-line terminal.

Sequence file manipulation tools

There is no one file format that can be used between various phylogenetic packages. The most commonly used format is the FASTA format; other formats that will be required are Phylip, Clustal and Nexus. All these sequence files are plain text files and can be manipulated using a text editor (not a word processor), but this procedure is laborious and time consuming. It is best done using a dedicated software package or packages. There are packages on the WEB that can do this but I decided to write my own and include features that are consistent with the way I work. My package, written in Java, is called "FastaGBFileTool"; as the name suggests it's a tool that can be used to convert between various sequence file formats as well as concatenate sequence files. All the tools needed to write your own software are in the public domain and available on the WEB.

MEGA X – (Molecular Evolutionary Genetics Analysis)

The Molecular Evolutionary Genetics Analysis (MEGA) is a very popular GUI based software package that implements many analytical methods and tools that can be used for the production and testing of phylogenetic inference trees. All procedures, from the downloading of sequences all the way to viewing your phylogenetic tree, can be done with this software package. There are versions of MEGA X available for all major operating systems (Mac, Windows, Linux).

This is an excellent package to use if you are just starting. It comes with a 130 page manual, and there are numerous tutorials on its website.

MEGA X can be downloaded from :- https://megasoftware.net/

Multiple sequence alignment (MSA) tools

Muscle. The version of Muscle I used is included in the MEGA X software package.

T-Coffee, version 13.41.118.d921da1 T-Coffee is a very powerful command-line driven alignment package with a large range of options, and comes with a 159 page manual. This is a very versatile software package and well worth the learning curve.
T-Coffee can be downloaded from :- http://www.tcoffee.org/

Making Inference Trees

PartitionFinder version 2.1.1 is used for finding best-fit partitioning schemes and models of evolution for your aligned sequences. Partitioning schemes and models are required by ML and Bayesian methods when estimating phylogenetic trees.
PartitionFinder can be downloaded from :-
http://www.robertlanfear.com/partitionfinder/

IQ-TREE version 2.0.5 for Linux. IQ-TREE is a widely used software package for phylogenetic inference using the maximum likelihood (ML) method. IQ-TREE can find the best fit model of evolution via its inbuilt ModelFinder. It has a novel ultra-fast bootstrap approximation and exhibits good performance in terms of both computing times and likelihood maximization.
IQ-TREE can be downloaded from :- http://www.iqtree.org/

RAxML version 8.2.12. RAxML is a widely used software package for phylogenetic inference using the maximum likelihood (ML) method. It comes in two forms, the command-line and the GUI versions. **RaxmlGUI version 2.0** consists of a software wrapper that seamlessly integrates RAxML binaries so that they can be easily run from a GUI interface. This arrangement offers an automated pipeline for analyses that may require multiple successive calls of RAxML. For ease of use I chose **raxmlGUI version 2.0**, which implemented **RAxML version 8.2.12**. There are raxmlGUI software packages available for all major operating systems (Mac, Windows, Linux).
RaxmlGUI can be downloaded from :- https://antonellilab.github.io/raxmlGUI/
RAxML can be downloaded from :-
https://cme.h-its.org/exelixis/web/software/raxml/

MrBayes version 3.2.7a. MrBayes is a software package that performs Bayesian inference of phylogenetic and evolutionary models. The program has a command-line interface and should run on a variety of computer platforms. Depending on the settings, MrBayes analyses may demand a lot of your computer, both in terms of memory and processor speed. This package comes with an 152 page manual which explains how install and use it.
MrBayes can be downloaded from :- http://nbisweden.github.io/MrBayes/

Viewing and Editing FigTree version 1.4.4. FigTree is a graphical viewer of phylogenetic trees and can be used to produce publication-ready figures. It is written in Java and is compatible with Mac, Windows and Linux operating systems.
FigTree can be downloaded from :- http://tree.bio.ed.ac.uk/software/figtree/

Appendix B: Examples of GenBank & FASTA File Formats

Cruentomycena viscidocruenta EU517515 GenBank File

```
LOCUS       EU517515                632 bp    DNA     linear   PLN 21-NOV-2008
DEFINITION  Cruentomycena viscidocruenta voucher K(M)117458 clone C3 18S
            ribosomal RNA gene, partial sequence; internal transcribed spacer
            1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2,
            complete sequence; and large subunit ribosomal RNA gene, partial
            sequence.
ACCESSION   EU517515
VERSION     EU517515.1
KEYWORDS    .
SOURCE      Cruentomycena viscidocruenta
  ORGANISM  Cruentomycena viscidocruenta
            Eukaryota; Fungi; Dikarya; Basidiomycota; Agaricomycotina;
            Agaricomycetes; Agaricomycetidae; Agaricales; Mycenaceae;
            Cruentomycena.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Petersen,R.H., Hughes,K.W., Lickey,E.B., Kovalenko,A.E.,
            Morozova,O.V., Psurtseva,N.V. and Morosova,O.
  TITLE     A new genus, Cruentomycena, with Mycena viscidocruenta as type
            species
  JOURNAL   Mycotaxon 105, 119-136 (2008)
REFERENCE   2  (bases 1 to 632)
  AUTHORS   Hughes,K.W., Petersen,R.H., Lickey,E., Psurtseva,N.V., Kovalenko,A.
            and Morosova,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2008) Ecology and Evolutionary Biology,
            University of Tennessee, Knoxville, TN 37996, USA
FEATURES             Location/Qualifiers
     source          1..632
                     /organism="Cruentomycena viscidocruenta"
                     /mol_type="genomic DNA"
                     /specimen_voucher="K(M)117458"
                     /db_xref="taxon:182006"
                     /clone="C3"
                     /country="Australia: Tasmania"
                     /identified_by="G. Gates, D. Ratkowski"
     rRNA            <1..13
                     /product="18S ribosomal RNA"
     misc_RNA        14..245
                     /product="internal transcribed spacer 1"
     rRNA            246..405
                     /product="5.8S ribosomal RNA"
     misc_RNA        406..621
                     /product="internal transcribed spacer 2"
     rRNA            622..>632
                     /product="large subunit ribosomal RNA"
ORIGIN
        1 ggaaggatca ttattgaata cgctttgagt gttgatgctg gccctttcgg gggcatgtgc
       61 tcgcattcaa aactatttat cttctcttgt gcacctttg tagtctttga atgtcacctt
      121 tccgctgcaa acgcgggttt tggagaggct tcgtgctttt tctctgggct tcaaagacta
      181 tgtcttcata tacgttaaac agtttagaat gtctttttaa cggcctaaca gccattaaac
      241 ctaatacaac tttcaacaac ggatctcttg gctctcccat cgatgaagaa cgcagcgaaa
      301 tgcgataagt aatgtgaatt gcagaattca gtgaatcatc gaatctttga acgcaccttg
      361 cgccctttgg tattccgaag ggcatgcctg tttgagtgtc attaaattat caaccttgaa
```

```
      421 agcttttgtg gctcttcttg gcttggatgt gagggctttg ctggcttcct tcagtggatt
      481 ggtctgctcc ctttaaatgc attagtggga tctcttgtgg accgtcactt ggtgtgataa
      541 ttatctacgc cgcttgactt tgaagcaaat cttatgggaa cctgcttata accgtccctc
      601 gcttggacaa ctttctttaa tttgacctca aa
//
```

Cruentomycena viscidocruenta EU517515 FASTA File

```
>Cruentomycena_viscidocruenta_EU517515
GGAAGGATCATTATTGAATACGCTTTGAGTGTTGATGCTGGCCCTTTCGGGGGCATGTGCTCGCATTCAAAACTATTTAT
CTTCTCTTGTGCACCTTTTGTAGTCTTTGAATGTCACCTTTCCGCTGCAAACGCGGGTTTTGGAGAGGCTTCGTGCTTTT
TCTCTGGGCTTCAAAGACTATGTCTTCATATACGTTAAACAGTTTAGAATGTCTTTTTAACGGCCTAACAGCCATTAAAC
CTAATACAACTTTCAACAACGGATCTCTTGGCTCTCCCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATT
GCAGAATTCAGTGAATCATCGAATCTTTGAACGCACCTTGCGCCCTTTGGTATTCCGAAGGGCATGCCTGTTTGAGTGTC
ATTAAATTATCAACCTTGAAAGCTTTTGTGGCTCTTCTTGGCTTGGATGTGAGGGCTTTGCTGGCTTCCTTCAGTGGATT
GGTCTGCTCCCTTTAAATGCATTAGTGGGATCTCTTGTGGACCGTCACTTGGTGTGATAATTATCTACGCCGCTTGACTT
TGAAGCAAATCTTATGGGAACCTGCTTATAACCGTCCCTCGCTTGGACAACTTTCTTTAATTTGACCTCAAA
```

Mycena pura EU517506 GenBank File

```
LOCUS       EU517506                623 bp    DNA     linear   PLN 21-NOV-2008
DEFINITION  Mycena pura voucher TENN60747 18S ribosomal RNA gene, partial
            sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene,
            and internal transcribed spacer 2, complete sequence; and large
            subunit ribosomal RNA gene, partial sequence.
ACCESSION   EU517506
VERSION     EU517506.1
KEYWORDS    .
SOURCE      Mycena pura
  ORGANISM  Mycena pura
            Eukaryota; Fungi; Dikarya; Basidiomycota; Agaricomycotina;
            Agaricomycetes; Agaricomycetidae; Agaricales; Mycenaceae; Mycena.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Petersen,R.H., Hughes,K.W., Lickey,E.B., Kovalenko,A.E.,
            Morozova,O.V., Psurtseva,N.V. and Morosova,O.
  TITLE     A new genus, Cruentomycena, with Mycena viscidocruenta as type
            species
  JOURNAL   Mycotaxon 105, 119-136 (2008)
REFERENCE   2  (bases 1 to 623)
  AUTHORS   Hughes,K.W., Petersen,R.H., Lickey,E., Psurtseva,N.V., Kovalenko,A.
            and Morosova,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2008) Ecology and Evolutionary Biology,
            University of Tennessee, Knoxville, TN 37996, USA
FEATURES             Location/Qualifiers
     source          1..623
                     /organism="Mycena pura"
                     /mol_type="genomic DNA"
                     /specimen_voucher="TENN60747"
                     /db_xref="taxon:153505"
                     /country="Russia: Primorsky Region, Hasansky District"
                     /lat_lon="43.10123 N 131.55208 E"
                     /identified_by="R.H. Petersen"
     rRNA            <1..13
                     /product="18S ribosomal RNA"
     misc_RNA        14..239
                     /product="internal transcribed spacer 1"
     rRNA            240..399
```

```
                     /product="5.8S ribosomal RNA"
     misc_RNA        400..612
                     /product="internal transcribed spacer 2"
     rRNA            613..>623
                     /product="large subunit ribosomal RNA"
ORIGIN
        1 ggaaggatca ttattgaata acttggtgtt gagctggccc cagtcgggca tgtgctcgca
       61 tcatattatt tatctatctc ttgtgcacct tttgtagtct ttgaagcgtt cgcagtcnat
      121 gcggttttgg gtcttgggct ttgccgccct tttccctgtt tgcttcaaag gctatgtttt
      181 atacacacta tttgaagtca cagaatgtct cttattgact ttcaagtcag taaatctata
      241 caactttcag caacggatct cttggctctc ccatcgatga agaacgcagc gaaatgcgat
      301 aagtaatgtg aattgcagaa ttcagtgaat catcgaatct ttgaacgcac cttgcgccct
      361 ttggtattcc gaagggcatg cctgtttgag tgtcattaaa ttctcaatct tgcagacttt
      421 tgtttgcgag gcttggatgt gagggctttg ctggcttcca ttcagttgga tggtctgctc
      481 cctttaaatt cattagtggg atcctttgtg gatggtcact tggtgtgata attatctacg
      541 ccgcctgact ctgaaacaag acttgtggga acctgcttat aaccgtctct tcagagacta
      601 tcttttgaca atttgacctc aaa
//
```

Mycena pura EU517506 FASTA File

```
>Mycena_pura_EU517506
GGAAGGATCATTATTGAATAACTTGGTGTTGAGCTGGCCCCAGTCGGGCATGTGCTCGCATCATATTATTTATCTATCTC
TTGTGCACCTTTTGTAGTCTTTGAAGCGTTCGCAGTCNATGCGGTTTTGGGTCTTGGGCTTTGCCGCCCTTTTCCCTGTT
TGCTTCAAAGGCTATGTTTTATACACACTATTTGAAGTCACAGAATGTCTCTTATTGACTTTCAAGTCAGTAAATCTATA
CAACTTTCAGCAACGGATCTCTTGGCTCTCCCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGAA
TTCAGTGAATCATCGAATCTTTGAACGCACCTTGCGCCCTTTGGTATTCCGAAGGGCATGCCTGTTTGAGTGTCATTAAA
TTCTCAATCTTGCAGACTTTTGTTTGCGAGGCTTGGATGTGAGGGCTTTGCTGGCTTCCATTCAGTTGGATGGTCTGCTC
CCTTTAAATTCATTAGTGGGATCCTTTGTGGATGGTCACTTGGTGTGATAATTATCTACGCCGCCTGACTCTGAAACAAG
ACTTGTGGGAACCTGCTTATAACCGTCTCTTCAGAGACTATCTTTTGACAATTTGACCTCAAA
```

Panellus stipticus EU517522 GenBank File

```
LOCUS       EU517522                517 bp    DNA     linear   PLN 21-NOV-2008
DEFINITION  Panellus stypticus voucher TENN61074 haplotype 2 18S ribosomal RNA
            gene, partial sequence; internal transcribed spacer 1, 5.8S
            ribosomal RNA gene, and internal transcribed spacer 2, complete
            sequence; and large subunit ribosomal RNA gene, partial sequence.
ACCESSION   EU517522
VERSION     EU517522.1
KEYWORDS    .
SOURCE      Panellus stipticus
  ORGANISM  Panellus stipticus
            Eukaryota; Fungi; Dikarya; Basidiomycota; Agaricomycotina;
            Agaricomycetes; Agaricomycetidae; Agaricales; Mycenaceae; Panellus.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   Petersen,R.H., Hughes,K.W., Lickey,E.B., Kovalenko,A.E.,
            Morozova,O.V., Psurtseva,N.V. and Morosova,O.
  TITLE     A new genus, Cruentomycena, with Mycena viscidocruenta as type
            species
  JOURNAL   Mycotaxon 105, 119-136 (2008)
REFERENCE   2  (bases 1 to 517)
  AUTHORS   Hughes,K.W., Petersen,R.H., Lickey,E., Psurtseva,N.V., Kovalenko,A.
            and Morosova,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2008) Ecology and Evolutionary Biology,
            University of Tennessee, Knoxville, TN 37996, USA
FEATURES             Location/Qualifiers
```

```
     source          1..517
                     /organism="Panellus stipticus"
                     /mol_type="genomic DNA"
                     /specimen_voucher="TENN61074"
                     /db_xref="taxon:5636"
                     /haplotype="2"
                     /country="New Zealand: Nelson Dist., Charming Creek Track"
                     /lat_lon="41.58357 S 171.96107 E"
                     /identified_by="K. Hughes"
     rRNA            <1..13
                     /product="18S ribosomal RNA"
     misc_RNA        14..175
                     /product="internal transcribed spacer 1"
     rRNA            176..335
                     /product="5.8S ribosomal RNA"
     misc_RNA        336..506
                     /product="internal transcribed spacer 2"
     rRNA            507..>517
                     /product="large subunit ribosomal RNA"
ORIGIN
        1 ggaaggatca ttattgaata cgcttttggg tgttgacgct ggcctttcga ggcatgtgct
       61 cgcattcaaa ctgtttaaat cttcacttgt gcacctttg tagtcttggg aggagactat
      121 gttttcatat acattgaaaa gttacagaat gtcttgaggc ctttataaag ttaatacaac
      181 tttcaacaac ggatctcttg gctctcccat cgatgaagaa cgcagcgaaa tgcgataagt
      241 aatgtgaatt gcagaattca gtgaatcatc gaatctttga acgcaccttg cgccctttgg
      301 tattccgaag ggcatgcctg tttgagtgtc attaaattat caaccttgga ggctttccta
      361 ggcttggatg tgagggcttt tgctggcttc cttcagtgga ttggtctgct ccctttaaat
      421 gcattagtgg gttccacaag tcgctcggtg tgataaatta tctacaccgt attgacttgc
      481 acctgcttat aaccgtagtt tgatatttga cctcaaa
//
```

Panellus stipticus EU517522 FASTA File

```
>Panellus_stipticus_EU517522
GGAAGGATCATTATTGAATACGCTTTTGGGTGTTGACGCTGGCCTTTCGAGGCATGTGCTCGCATTCAAACTGTTTAAAT
CTTCACTTGTGCACCTTTTGTAGTCTTGGGAGGAGACTATGTTTTCATATACATTGAAAAGTTACAGAATGTCTTGAGGC
CTTTATAAAGTTAATACAACTTTCAACAACGGATCTCTTGGCTCTCCCATCGATGAAGAACGCAGCGAAATGCGATAAGT
AATGTGAATTGCAGAATTCAGTGAATCATCGAATCTTTGAACGCACCTTGCGCCCTTTGGTATTCCGAAGGGCATGCCTG
TTTGAGTGTCATTAAATTATCAACCTTGGAGGCTTTCCTAGGCTTGGATGTGAGGGCTTTTGCTGGCTTCCTTCAGTGGA
TTGGTCTGCTCCCTTTAAATGCATTAGTGGGTTCCACAAGTCGCTCGGTGTGATAAATTATCTACACCGTATTGACTTGC
ACCTGCTTATAACCGTAGTTTGATATTTGACCTCAAA
```

Appendix C: Some Australian Mycena Species

The table contains all of the Australian *Mycena* species used in a search for ITS sequences in the NCBI database. As of 07/08/2020 there were no matches and hence I was not able to include any described Australian *Mycena* species.

No.	Species	ITS ID.	No.	Species	ITS ID.
1	Mycena albidocapillaris	—	27	Mycena mijoi	—
2	Mycena albidofusca	—	28	Mycena mulawaestris	—
3	Mycena atrata	—	29	Mycena nargan	—
4	Mycena atroavellanea	—	30	Mycena neerimensis	—
5	Mycena australiana	—	31	Mycena nivalis	—
6	Mycena austrofilopes	—	32	Mycena nullawarrensis	—
7	Mycena carmeliana	—	33	Mycena nyula	—
8	Mycena sp. (cf-lazulina)*	MN330032	34	Mycena piringa	—
9	Mycena clarkeana	—	35	Mycena roseoflava	—
10	Mycena cunninghamiana	—	36	Mycena simpsonii	—
11	Mycena cystidiosa	—	37	Mycena subalbida	—
12	Mycena epipterygia	—	38	Mycena subcorticalis	—
13	Mycena eucalyptorum	—	39	Mycena subgalericulata	—
14	Mycena fuhreri	—	40	Mycena subnigra	—
15	Mycena fumosa	—	41	Mycena subvulgaris	—
16	Mycena fusca	—	42	Mycena tallangattensis	—
17	Mycena insueta	—	43	Mycena tasmaniensis	—
18	Mycena interrupta	—	44	Mycena thunderboltensis	—
19	Mycena kurramulla	—	45	Mycena toyerlaricola	—
20	Mycena kuurkacea	—	46	Mycena trachycephala	—
21	Mycena kyeema	—	47	Mycena tuvara	—
22	Mycena lageniformis	—	48	Mycena vinacea	—
23	Mycena leaiana	—	49	Mycena waralya	—
24	Mycena lilliria	—	50	Mycena wubabulna	—
25	Mycena maldea	—	51	Mycena yirukensis	—
26	Mycena marangania	—	52	Mycena yuulongicola	—

*For this species I already had the sequence.