Constructing a robust bioinformatic pipeline to investigate the phylogenetic relationships of species in the genus *Amanita*

Earth is experiencing significant losses of biodiversity1–3, but fungi are rarely included in conservation science4 due, in part to scientists historically assuming most fungi (especially mushrooms) were globally distributed. A local extinction of a cosmopolitan species is less concerning than the extinction of an endemic on. However, over the last few decades, DNA sequence data have changed our understanding of “cosmopolitan fungi”. In fact, most fungal lineages are made up of cryptic, endemic species found within restricted ranges5. Fungal species are often difficult to tell apart morphologically and, historically, more attention was paid to naming plants and animals. Only 5-10% of fungal species are formally described6. This discrepancy leaves us with little useful knowledge with which we can base conservation practices off. A specific example of this challenge is found the in species of the genus *Amanita* that are found in Wisconsin.

The *Amanita* genus encompasses species with a wide range of ecological niches and evolutionary histories, with some considered prize edibles like *A. caesarea*10, while others like *A. phalloides* are deadly11. Of the three subgenera of *Amanita*, two are considered ectomycorrhizal (ECM) and the other is considered saprotrophic12. Although it is a charismatic and famous genus, the biodiversity of the group remains cryptic and poorly understood. There is an abundance of species complexes in *Amanita* – a direct result of historical use of solely morphological data as the basis for taxonomic organization. As DNA sequencing technologies have become available and widespread, *Amanita* taxa going under a scientific name are repeatedly found to have enough genetic difference to be considered different lineages. For example, molecular evidence suggests *Amanita muscaria*, currently understood as a globally distributed species, is likely a complex composed of at least 8 lineages with endemic and relatively narrow ecological niches13.

The upshot: if we want to be effective in our efforts to conserve the biodiversity of fungi, we must have a firm grasp of the taxonomic organization of the kingdom. Increased efforts to build quality phylogenetic trees is needed across the entire kingdom. I am building a pipeline using a sample set of *Amanita* DNA sequences to have a robust tool with which I can apply to my own data once available. I present here an explanation of my pipeline and a description of my exploration of tools and models I can use to build robust phylogenies for *Amanita*.

**Methods:**

Data:

The data consist of 24 taxa each with sequences for three common loci that code for the nuclear large ribosomal subunit (nrLSU), trans-elongation factor 1 alpha (TEF1α), and RNA polymerase subunit 2 (RPB2) (Table 1). The data are a subset of taxa that were used to create the current accepted phylogeny for the genus of *Amanita* (Cui et al 2018). Taxon accession numbers for each of the 3 loci were used to create a fasta file with the tool Entrez Direct from NCBI (CITE). To investigate the phylogenetic space of these taxa, I implement two different alignment softwares (MAFFT and CLUSTAL-W), a maximum parsimony software (phangorn), a maximum likelihood software (iqtree2), a Bayesian inference software (MrBayes), and, finally, a coalescent software (ASTRAL).

Alignment:

Fasta files were aligned using two methods. The first software used was Clustal-W (CITE), a progressive multiple sequence alignment (PMSA) method that boosts sensitivity through dynamic use of weights. It creates an easily customizable alignment that more accurately represents mutation types and the probability of their occurrence in different areas of the genome. Further, it implements dynamic substitution matrices meaning it changes the substitution matrix used depending on sequence divergence – matrices become more relaxed as sequences become more diverged. Despite many strengths, Clustal-W still deals with local minimum issues as the alignment starts with the most similar sequences first. If there are initial errors in the beginning stages of alignment, they will proliferate as more sequences are added. For this project, I specified DNA sequences and conservative weights of 1 for both the gap opening and gap extension penalties. The alignments were then viewed with Aliview (CITE) and trimmed using trimAl (CITE) where gaps were removed if 50% or more of the sequences contained the gap.

The algorithm uses fast Fourier transform to identify homologous regions within the gene sequence of the taxa in question.

**Figures:**

Table 1: Data from Cui *et al.* 2018

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Species name | Location | nrLSU | TEF1α | RPB2 |
| *A.* aff. *brunneofuliginea* | Austria | MH486361 | MH508664 | MH485858 |
| *A. altipes* | Shangri-La, Yunnan, China | JN941158 | KR824810 | JQ031112 |
| *A. aspericeps* | Mount Wuyi, Fujian, China | MH486369 | MH508671 | MH485864 |
| *A. avellaneosquamosa* | Incheon's Central park, South Korea | KJ466482 | KJ481981 | KJ466647 |
| *A. battarrae* | Changdu, Tibet, China | MH486383 | MH508684 | MH485875 |
| *A. brunneofuliginea* | Changbai mountain, Jilin, China | MH486394 | MH508694 | MH485886 |
| *A. caojizong* | Yulong, Yunnan, China | MH486425 | MH508712 | MH485905 |
| *A. chiui* | Yanyuan, Sichuan, China | MH486447 | MH508727 | MH485930 |
| *A. cinereopannosa* | USA | MH486450 | MH508728 | MH485932 |
| *A. citrina* | Germany | MH486457 | MH508733 | MH485937 |
| *A. citrinoannulata* | Lianyungang, Jiangsu, China | MH486458 | MH508734 | MH485938 |
| *A. citrinoindusiata* | Shangri-La, Yunnan, China | MH486469 | MH508745 | MH485948 |
| *A. eijii* | Ninglang, Yunnan, China | MH486484 | MH508761 | MH485963 |
| *A. elata* | Xishuangbanna, Yunnan, China | MH486486 | MH508763 | MH485965 |
| *A. farinosa* | Baoshan, Yunnan, China | JN941154 | MH508772 | JQ031110 |
| *A. flavipes* | Shangri-La, Yunnan, China | MH486507 | MH508783 | MH485979 |
| *A. fulva* | France | MH486554 | MH508825 | MH486021 |
| *A. imazekii* | Tomakomai, Hokkaido, Japan | MH486591 | MH508859 | MH486052 |
| *A. ocreata* | CA, USA | MH486688 | KJ481947 | KJ466607 |
| *A. pantherina* | Czech Republic | KR824782 | KR824825 | KR824789 |
| *A. strobiliformis* | Germany | MH486895 | MH509117 | MH486298 |
| *A. suballiacea* | PA, USA | KJ466484 | KJ481940 | KJ466600 |
| *A. virosa* | Juva, Filand | JX998058 | JX998007 | KJ466664 |
| *L. delicata* | Austria | KT833807 | KT833835 | KT833822 |

**References:**