Dissertation Title	Phylogeny, Evolutionary History and Pan-genome Dynamics of *Armillaria*
Author	Jianwei Liu
Degree	Doctor of Philosophy (Biological Science)
Advisor	Chitrabhanu Sharma Bhunjun, Ph. D.
Co-Advisor	Professor Zhu-Liang Yang, Ph. D.

ABSTRACT

Recent studies have shown that phylogenetic discordance, caused by incomplete lineage sorting, hybridization, introgression, and horizontal gene transfer, is much more common across the tree of life than previously recognized. This widespread discordance makes reconstructing evolutionary relationships particularly challenging. *Armillaria* species are a compelling example, given their complex ecological roles. They function as destructive forest pathogens with significant economic impact, as saprotrophs contributing to wood decomposition and nutrient cycling, and as symbiotic partners of medicinal plants and fungi such as *Gastrodia elata* (Tianma) and *Polyporus umbellatus* (Fuling). These combined roles support an estimated annual market value of ¥14–21 billion. However, several fundamental questions remain unresolved, including the true extent of species diversity, boundaries between species, compatibility with *Gastrodia* under cultivation, and the evolutionary processes driving speciation.

Many studies have revealed that incomplete reproductive isolation between species is common in the genus *Armillaria*. This potential incomplete reproductive isolation may indicate ongoing gene flow, which could be the one factor underlying past difficulties in delineating species boundaries within this fungal genus. Here, focusing on China while including global specimens, we analyzed 600 biparentally inherited nuclear genomes and 605 maternally inherited mitochondrial genomes to establish a robust phylogenetic framework for the genus *Armillaria*. Through comprehensive comparisons of population genetic differentiation metrics, including Weir & Cockerham's F_{ST}, absolute differentiation (dxy), mitochondrial gene relative differentiation (ΦST), we delineated 21 distinct *Armillaria* lineages in China, comprising 6 cryptic novel taxa

(designated as *A. borealis* subsp. *sinoborealis*, *A. griseocapilla*, *A. linzhiensis*, *A. pseudoostoyae*, and *A. tabescens* subsp. *orientala*, *A. violacea* (CBS N) lineage 2-mtDNA), and a comprehensive review of all 73 previously described species was performed. Threshold analysis of population differentiation coefficients demonstrated that, from the nuclear-genomic perspective, an F_{ST}-nDNA value of ≥ 0.09 can serve as a preliminary threshold for delimiting species in *Armillaria*; however, the status of a lineage as a species or subspecies must still be confirmed with biological mating trial statistics and other supporting evidence. For those clades that receive strong support in the mitochondrial phylogeny, one must first verify that they also form independently supported clades in the nuclear phylogeny; only under that condition, when the corresponding mitochondrial divergence metrics reach F_{ST}-mtDNA = 0.37, dxy = 0.04, and ΦST = 0.45, be considered important criteria for species or subspecies delimitation. Meanwhile, a nuclear genomic divergence threshold of approximately 1.052 Ma may serve as a meaningful temporal criterion for species delimitation may serve as empirical criteria for species delimitation within the genus, facilitating the recognition of cryptic novel taxa.

Our results demonstrate strong concordance between phylogenetic and biological species concepts across all 14 recognized taxa. In light of this, we recommend retaining *A. pungentisquamosa* as a valid name, despite its potential synonymy with *A. xiaocaobanensis*, given the global consensus that species delimitation in *Armillaria* should prioritize reproductive compatibility and genomic distinctiveness over nomenclatural priority. Notably, analyses of widespread species complexes uncovered significant maternal-lineage phylogeographic patterns. For example, the Chinese biological species (hereafter abbreviated as CBS) J complex is resolved into five discrete geographic lineages: Xinjiang, China; Central China; the Eastern Himalayas; the Hengduan Mountains; and Southern Tibet based on mitochondrial genome data. Despite these clear mitochondrial breaks, nuclear-genome monophyly and high interfertility support retaining the current taxonomic circumscription without further subdivision. Although we did not delimit species according to these mitochondrial geographic lineages, their well-defined spatial structure provides an ideal framework for future fungal biogeographic (mycobiogeographic) investigations. Building on a clear species delimitation, our

systematic *Armillaria* and *Gastrodia* co-cultivation experiments suggest that, to date, only species in sect. *Gallicae* may be suitable for *Gastrodia elata* cultivation.

Our study also uncovers extensive phylogenetic discordance between nuclear and mitochondrial genomes, as well as between gene trees and the species tree, across the genus *Armillaria*. Phylogenetic-cause analyses indicate that incomplete lineage sorting is the primary driver of phylogenetic incongruence within *Armillaria*, although the roles of hybridization and introgression should not be overlooked. PhyloNet-based network analysis shows that the *Gallicae* section represents a hybrid lineage derived from the *Ostoyarum* section (weight 0.844) and the *Armillaria* section (weight 0.156), while the *A. tibetica* (CBS J) complex originates from a mixture of *A. bruneocystidia* (CBS H) (weight 0.51) and *A. sinapina* subsp. *alpina* (CBS A) + *A. luteopileata* (CBS C) + *A. sinensis* (CBS F) + *A. cepistipes* + *A. sinapina* subsp. *sinapina* (weight 0.49). Divergence time and ancestral range reconstruction analyses revealed that the most recent common ancestor (MRCA) of *Armillaria* originated in the Oligocene from Holarctic-Asia and subsequently underwent a complex history of global dispersal and diversification, with an estimated age of 27.7 Ma based on nuclear genome data and 24.40 Ma based on mitochondrial genome data.

To enable rapid and accurate species identification in *Armillaria*, we developed an integrated, multi-level identification framework combining genomic and computational approaches. First, we constructed two offline-searchable genomic databases (mitochondrial and nuclear) using *skani*, allowing users to perform efficient species classification with locally stored data. Second, we implemented a deep learning-based method that transforms raw genome sequences into visual DNA barcodes by encoding k-mer frequencies, enabling high-throughput pattern recognition for species discrimination. Third, through systematic evaluation of 31 primer pairs targeting 11 candidate genes, including internal primers for the longer loci, we found that almost every gene is suitable as a candidate barcode. Finally, we developed three SSR (simple sequence repeat) fingerprinting systems tailored for strain-level identification, providing breeders, industries, and researchers with robust tools for cultivar authentication, intellectual property protection, and population genetic analyses.

Keywords: DNA Barcording, Gene Flow, New Species, Phylogenetic Discordance, SSR