doi:10.5598/imafungus.2018.09.02.01 IMA FUNGUS · 9(2): 225–242 (2018)

ARTICLE

The MSDIN family in amanitin-producing mushrooms and evolution of the prolyl oligopeptidase genes

Hong Luo[1], Qing Cai[1], Yunjiao Lüli[1, 2], Xuan Li[3], Rohita Sinha[4], Heather E. Hallen-Adams[5], and Zhu L. Yang[1]

[1]Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China; corresponding author e-mail: luohong@mail.kib.ac.cn
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]Department of Environmental Science and Engineering, Kunming University of Science and Technology, Kunming 650091, Yunnan, China
[4]Viracor Eurofins, Lee's Summit, MO 64086, USA
[5]Department of Food Science and Technology, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

Abstract: The biosynthetic pathway for amanitins and related cyclic peptides in deadly *Amanita* (*Amanitaceae*) mushrooms represents the first known ribosomal cyclic peptide pathway in the *Fungi*. Amanitins are found outside of the genus in distantly related agarics *Galerina* (*Strophariaceae*) and *Lepiota* (*Agaricaceae*). A long-standing question in the field persists: why is this pathway present in these phylogenetically disjunct agarics? Two deadly mushrooms, *A. pallidorosea* and *A. subjunquillea*, were deep sequenced, and sequences of biosynthetic genes encoding MSDINs (cyclic peptide precursor) and prolyl oligopeptidases (*POPA* and *POPB*) were obtained. The two *Amanita* species yielded 29 and 18 MSDINs, respectively. In addition, two MSDIN sequences were cloned from *L. brunneoincarnata* basidiomes. The toxin MSDIN genes encoding amatoxins or phallotoxins from the three genera were compared, and a phylogenetic tree constructed. Prolyl oligopeptidase B (POPB), a key enzyme in the biosynthetic pathway, was used in phylogenetic reconstruction to infer the evolutionary history of the genes. Phylogenies of *POPB* and *POPA* based on both coding and amino acid sequences showed very different results: while *POPA* genes clearly reflected the phylogeny of the host species, *POPB* did not; strikingly, it formed a well-supported monophyletic clade, despite that the species belong to different genera in disjunct families. *POPA*, a known house-keeping gene, was shown to be restricted in a branch containing only *Amanita* species and the phylogeny resembled that of those *Amanita* species. Phylogenetic analyses of MSDIN and *POPB* genes showed tight coordination and disjunct distribution. A *POPB* gene tree was compared with a corresponding species tree, and distances and substitution rates were compared. The result suggested *POPB* genes have significant smaller distances and rates than the house-keeping *rpb2*, discounting massive gene loss. Under this assumption, the incongruency between the gene tree and species tree was shown with strong support. Additionally, k-mer analyses consistently cluster *Galerina* and *Amanita POPB* genes, while *Lepiota POPB* is distinct. Our result suggests that horizontal gene transfer (HGT), at least between *Amanita* and *Galerina,* was involved in the acquisition of *POPB* genes, which may shed light on the evolution of the α-amanitin biosynthetic pathway.

Key words:
Amanita
Galerina
Lepiota
amatoxin
phallotoxin
phylogeny
horizontal gene transfer

Article info: Submitted: 22 November 2017; Accepted: 24 July 2018; Published: 1 August 2018.

INTRODUCTION

Amatoxins and related cyclic peptides produced by deadly *Amanita* and *Galerina* mushrooms are biosynthesized through a ribosomal cyclic peptide pathway (Hallen *et al.* 2007, Luo *et al.* 2012). α-Amanitin, the major cyclic peptide toxin, is responsible for the vast majority (>90%) of deadly mushroom poisonings worldwide (Bresinsky & Besl 1990). The biosynthesis of this toxin and related cyclic peptides begins with activation of genes that encode a precursor peptide of 34–37 amino acids, named the MSDIN gene family after the highly conserved first five residues (Hallen *et al.* 2007). The precursor peptides are cleaved and macrocyclized into 7–10 amino acid cyclic peptides by a specialized prolyl oligopeptidase enzyme, POPB (Luo *et al.* 2009, 2014, Riley *et al.* 2014). POPB is the key enzyme of the cyclic peptide pathway, catalyzing both hydrolysis of the peptide bond and transpeptidation (Luo *et al.* 2014).

Besides certain species of *Amanita* and *Galerina*, a few species of *Lepiota* also produce α-amanitin (Haines *et al.* 1986, Mottram *et al.* 2010, Sgambelluri *et al.* 2014). Some *Conocybe* species are also reported to produce similar toxins, but we failed to detect any cyclic peptides in recently collected *C. apala*, reported to produce phallotoxins, and therefore we did not continue further with that study. Phylogenetically, the genera *Amanita*, *Galerina* and *Lepiota* are distantly related, belonging to three disjunct agaric families, *Amanitaceae*, *Strophariaceae*, and *Agaricaceae*, respectively. This non-

continuous distribution of amatoxins has raised major questions in the field: why does this pathway occur in these isolated lineages and not others? Did amatoxin biosynthesis evolve independently on multiple occasions, or did it originate from a common ancestor followed by gene loss or horizontal gene transfer (HGT)? Researchers had tried to infer the evolution of the pathway, but major hurdles existed and have prevented significant progress. One serious problem is the lack of suitable molecular markers, specifically for those genes related to the pathway, to infer essential and well resolved phylogenetic frameworks for the target mushroom groups, which are critical for providing evolutionary evidence on how this pathway evolved among the groups. A second problem is that phylogenetically important species have been difficult to obtain, especially fresh samples suitable for genetic and genomic analyses. Amatoxin-producing *Amanita* species are obligately mycorrhizal and grow slowly in culture, necessitating the use of wild-collected basidiomes. Thirdly, sequenced relevant agaric genomes were insufficient to support comparative studies among the target mushroom groups. This lack of data motivated more and deeper genome sequencing of deadly *Amanita* species by our group. Recent evidence showed that *POPB* is a key biosynthetic gene for the amatoxins and related cyclic peptides of lethal mushrooms (Luo *et al.* 2014), and these data have provided clues for rigorously elucidating the evolution of the pathway, even though genomic data on the mushrooms remained incomplete. Not only can the phylogeny of *POP* genes resolve the relationships among these genes, it can assist with HGT detection, as the most reliable method for HGT detection is based on phylogenetic inference (Ragan 2001, Fitzpatrick 2012). In the absence of experimental systems to track HGT, the standard method for identifying putative HGT events has relied on phylogenetic incongruence — a strongly supported disagreement between a well-supported gene phylogeny and the species phylogeny is often used to justify the acceptance of one or more putative HGT events as the cause of the phylogenetic conflict (Andersson 2005, Keeling & Palmer 2008).

Prolyl oligopeptidases (POPs; EC 3.4.21.26) are present in most phyla of life (Venalainen *et al.* 2004, Kaushik & Sowdhamini 2014). They play important but varied housekeeping functions. In mammals (including humans), POPs are apparently multifunctional enzymes involved in the maturation and degradation of peptide hormones and neuropeptides (Polgar 2002). As such, POPs play important roles in a number of physiological processes, including: learning and memory (Yoshimoto *et al.* 1987, Garcia-Horsman *et al.* 2007), cell signaling (Williams *et al.* 1999, Duan *et al.* 2014), sperm motility (Yoshida *et al.* 1999, Kimura *et al.* 2002), and cell proliferation and differentiation (Ohtsuki *et al.* 1994, Moreno-Baylach *et al.* 2008, Sakaguchi *et al.* 2011). Furthermore, abnormalities in POP activity are associated with diseases (Momeni *et al.* 2005). The majority of reported POPs are intracellular enzymes, while mushrooms produce some extracellular POPs (Chen *et al.* 2012) that can only be speculated to perform more specific peptide hydrolyzation roles compared to more general proteases.

While many POPs perform housekeeping functions, one POP has a specialized role as a biosynthetic enzyme for the MSDIN family of cyclic peptides. Deadly *Amanita* and *Galerina* species carry two copies of *POP* genes, *POPA* and *POPB*, in contrast to only a single copy in other basidiomycete genomes; most ascomycetes do not contain *POP*s. *POPB* is a specialized form involved in the toxin biosynthesis (Luo *et al.* 2014). By comparison, *POPA* is considered to carry out housekeeping roles as the homologs are present in all the mushrooms examined to date, poisonous or not (Luo *et al.* 2010, 2012, 2014). In *G. marginata*, evidence shows that *GmPOPA* does not catalyse cyclization of the precursor peptide for α-amanitin, and therefore is not involved in the biosynthesis of the cyclic peptides (Luo *et al.* 2014).

As for taxonomic distribution, to date, evidence indicates that *POPB* is strictly confined to mushrooms producing MSDIN-family cyclic peptides, a family which includes phallotoxins (not toxic to humans on ingestion) as well as the dangerous amatoxins. Our initial investigation indicated that, unlike many other single-copy genes, the phylogeny of *POPB* does not reflect that of the agaric species that harbor this gene. Rather, they tend to cluster according to functional chemical diversity, contradicting their species phylogeny. During our efforts to sequence more amanitin-producing mushrooms, we tried to gain insights into the evolution of the pathway using comparative genomics, but no clear conclusion has yet been reached. However, this has provided more putative *POPB* gene sequences over time. Recently Jonathan D. Walton's laboratory at Michigan State University started sequencing a deadly *Lepiota*, *L. subincarnata*, and kindly sent us the genomic sequences of two *POPB* sequences, one from each of the two strains sequenced (the species contains only one *POPB*, and no *POPA*). As a result, we now have *POP* gene sequences from all three taxonomic groups confirmed to produce MSDIN-family cyclic peptides, which makes it possible to reconstruct the evolutionary histories of these genes in a well-represented species composition, and to perhaps shed light on the history of the biosynthetic pathway.

In this research, two amanitin-producing mushrooms, *A. subjunquillea* and *A. pallidorosea*, were sequenced through the Beijing Genomics Institute (BGI) in Wuhan, China, and the genomes were surveyed for MSDIN genes. MSDIN sequences were also cloned from two *L. brunneoincarnata* strains, and toxin MSDINs (defined as MSDIN genes encoding amatoxins or phallotoxins) from all three genera were compared. Furthermore, DNA and amino acid sequences of *POP* genes were mined from the genome assemblies. Together with two *POPB* genes from *L. subincarnata*, predicted coding and amino acid sequences for *POP*s from genome mining and databases were used for phylogenetic analyses. A *POP* gene tree was compared with the species tree for incongruency analysis. Distances and substitution rates were compared among the three genera. A topology test was performed to determine the robustness of the *POPB* phylogeny. Gene structure was analyzed by examining intron placement in *POPB* and toxin MSDIN genes, and conducting k-mer analyses on di-, tri- and tetranucleotide frequencies on *POP* and *POPB* genes. Based on the results, we assessed the evolutionary history of *POPB*.

METHODS

Mushroom samples

Fresh wild basidiomes of *Amanita subjunquillea*, *A. pallidorosea* and *Lepiota brunneoincarnata* were harvested, stored at -80 °C, and lyophilized. Upon collection, all specimens were immediately put on dry ice after they were removed from soil. These species are a major cause of lethal mushroom poisonings in Eastern Asian countries (Chen *et al.* 2013, 2016).

Genome sequencing and assembly

High molecular weight DNA was extracted from lyophilized basidiomes using Genomic-tip 100/G (Qiagen 10243), following the manufacturer's protocols. The sequencing strategy for *A. subjunquillea* and *A. pallidorosea* used Illumina HiSeq 4000 and PacBio RSII at Beijing Genomics Institute (BGI) with 250 bp, 10 Kb and 20 Kb libraries constructed and sequenced using the company's standardized pipeline. In both cases, PacBio polymerase reads < 1000 bp, or with quality score less than 80 %, were removed. Subreads were extracted from polymerase reads, and adapter filtered. Subreads were corrected using Pbdagcon (https://github.com/PacificBiosciences/pbdagcon), Falcon (https://github.com/PacificBiosciences/FALCON-integrate) and Proovread (Hackl *et al.* 2014). Corrected reads were assembled with Celera Assembler (Myers *et al.* 2000) (v. 8.3, parameters: doTrim_initialQualityBased = 1, doTrim_finalEvidenceBased = 1, doRemoveSpurReads = 1, doRemoveChimericReads = 1, -d properties -U) or Falcon (v. 0.3.0, parameters: -v -dal8 -t32 -h60 -e.96 -l500 -s100 -H3000). Scaffolds were constructed through SSPACE Basic (v. 2.0) (Boetzer & Pirovano 2014) and gap closing with PBJelly2 (English *et al.* 2012) (15.8.24 with default settings). GATK (https://www.broadinstitute.org/gatk/) and SOAP tool packages (SOAP2, SOAPsnp, SOAPindel) (Li *et al.* 2009a, b) were applied for single-base corrections.

Cloning of MSDINs from *Lepiota brunneoincarnata*

Primers targeting conserved regions of MSDINs were designed based on genes from *G. marginata* and *A. bisporigera*. For PCR amplification, four primers out of 18 primers tried were used in four combinations. The forward primers were 5'-GGCTACCTCATGTCTGCTCTCG-3' and 5'-CAATCCGTCTGACTACCCACTC-3'. The reverse primers were 5'-ACCGAGCGTTGTATAGGGAGAA-3' and 5'-GCAAAGGCTAGCAGACAATACG-3'. PCR reactions were conducted under standard conditions, and products with predicted correct sizes directly sequenced.

Mining for MSDIN and prolyl oligopeptidase genes

Nucleotide sequences of MSDINs and/or *POP*s from the genomes of *A. subjunquillea* and *A. pallidorosea* (this study), and *A. phalloides* (Pulman *et al.* 2016), were obtained through standalone BLAST searches (NCBI BLAST+ 2.4.0) with corresponding query MSDIN and *POPB* sequences from *A. bisporigera* and *G. marginata*, which are well characterized by our molecular and biochemical approaches (Luo *et al.* 2010, 2012). In order to obtain reasonably reliable coding and amino acid sequences of the *POP* genes from the sequenced genomes, the genomic DNA sequences of the genes were compared to those of well characterized cDNA sequences from *A. bisporigera* and *G. marginata*. It quickly became apparent that the intron and exon structures are highly conserved among both *POPA* and *POPB* genes. Coding sequences were predicted using *POPB* cDNA from *A. bisporigera* as the reference. Similarly, *POPA* coding sequences were retrieved using *AbPOPA* cDNA for comparison. In above cases, GT-AG intron borders were predicted by aligning the gDNA sequences with the cDNAs, and the resulting amino acid sequences were further assessed by examining conservation among the amino acid sequences along the full length. In all cases, we obtained undisrupted ORFs after deleting the introns, and the amino acid sequences were conserved throughout the full length. With this method, final exon-intron structure was resolved without ambiguity. The same approach was applied to *L. subincarnata POPB* genomic DNA by comparing its only *POP*, *LsPOPB*, to the cDNAs of those in *G. marginata*. After the introns were predicted and removed, amino acid sequences were retrieved through translation with no ambiguity found. The resulting sequences were then used for phylogenetic analysis. The *Amanita POP* sequences can be found in Suppl. File 1. To generate a well-represented *POP* pool for macrofungi, *POP* coding (or cDNA) and protein sequences were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/) and JGI MycoCosm (http://genome.jgi.doe.gov/programs/fungi/index.jsf) (Table 1). For MSDIN gene comparison, additional sequences were obtained from previously published sources (Li *et al.* 2014, Pulman *et al.* 2016).

Sequence alignment and phylogenetic analysis

Three datasets, the coding sequences (CDSs) and amino acid sequences of the selected *POP* genes, and the CDSs of selected toxin MSDINs, were compiled. Sequences were aligned using Muscle 3.6 (Edgar 2004) with default settings, and then manually adjusted with BioEdit (Hall 1999, Suppl. Files 2–4). For the amino acid alignment, LG+ G was selected as the best-fitting empirical model by ProTest 3 (Darriba *et al*. 2011) under Akaike Information Criterion (AIC). For the nucleotide alignment, GTR + I + G and GTR + G were inferred as the best substitution models for the CDSs of *POP* and MSDIN genes by using MrModeltest v. 2.3 (Nylander 2004) under AIC, respectively. Maximum likelihood (ML) tree searching and bootstrapping (1 000 replicates) were done in RAxML v. 7 (Stamatakis 2006). Bayesian inference was carried out in MrBayes v. 3.2.6 (Ronquist & Huelsenbeck 2003) with two independent Markov chain Monte Carlo (MCMC) runs and four chains each. Runs were performed for 2 M generations, with trees sampled every 100 generations. Chain convergence was determined using Tracer v. 1.5 (http://tree.bio.ed.ac.uk/sofware/tracer/) to ensure convergence and sufficiently large effective sampling size values (>200). Subsequently, the sampled trees were summarized by discarding the first 25 % of trees as burn-in using the 'sump' and 'sumt' command implemented in MrBayes (Ronquist & Huelsenbeck 2003).

Table 1. Accession numbers of prolyl oligopeptidase gene and amino acid sequences included in the phylogenetic study.

Taxon	Strain	Prolyl Oligopeptidase	Source	Amino Acids	Coding Sequence
Agaricus bisporus var. *bisporus*	H97	POP	JGI	219134	
Agrocybe pediades	AH 40210	POP	JGI	736488	
Anomoporia bombycina	ATCC 64506	POP	JGI	1333562	
Antrodia sinuosa	LB1	POP	JGI	705706	
Artolenzites elegans	CIRM-BRFM 1663	POP	JGI	887929	
Auricularia subglabra	–	POP	JGI	1169635	
Auriculariopsis ampla	NL-1724	POP	JGI	534627	
Amanita bisporigera	–	POPA	NCBI	ADN19204	HQ225840
	–	POPB	NCBI	ADN19205	HQ225841
Amanita muscaria	Koide	POPA	JGI	74086	
Amanita pallidorosea	–	POPA	Genome	–	
	–	POPB	Genome	–	
Amanita phalloides	–	POPA	Genome	–	
	–	POPB	Genome	–	
Amanita rimosa	–	POPA	Genome		
	–	POPB	Genome		
Amanita subjunquillea	–	POPA	Genome	–	
	–	POPB	Genome	–	
Amanita thiersii	Skay4041	POPA	JGI	193040	
Beauveria bassiana	ARSEF 2860	POP	JGI	1657	
Bolbitius vitellinus	SZMC-NL-1974	POP	JGI	1302751	
Calocera cornea	–	POP	JGI	490308	
Ceraceosorus bombacis	MCA 4658	POP	JGI	333060	
Cerrena unicolor	–	POP	JGI	313626	
Clitocybe gibba	IJFM A 808	POP	JGI	1443961	
Colletotrichum nymphaeae	SA-01	POP	JGI	1020605	
Conocybe apala	–	POP	NCBI	ACQ65797	FJ906819
Coprinopsis cinerea	okayama7#130	POP	NCBI	XP_001841244	XM_001841192
Cortinarius glaucopus	AT 2004 276	POP	JGI	7259917	
Crepidotus variabilis	CBS 506.95	POP	JGI	870504	
Crucibulum laeve	CBS 166.37	POP	JGI	718333	
Cyathus striatus	AH40144	POP	JGI	1424557	
Cytidiella melzeri	FP 102339	POP	JGI	1412396	
Dichomitus squalens	LYAD-421 SS1	POP	JGI	160828	
Fibulorhizoctonia sp.	CBS 109695	POP	JGI	740187	
Fomitiporia mediterranea	–	POP	JGI	141658	
Galerina marginata	–	POPA	JGI	70906	
	–	POPB	JGI	146341	
Gloeophyllum trabeum	ATCC 11539	POP	NCBI	XP_007862595	XM_007864404
Gymnopilus chrysopellus	PR-1187 v1.0	POP	JGI	1688417	
Hebeloma cylindrosporum	h7	POP	JGI	444548	
Heliocybe sulcata	OMC 1185	POP	JGI	1641899	
Hydnomerulius pinastri	–	POP	JGI	174977	
Hydnopolyporus fimbriatus	CBS384.51	POP	JGI	983183	
Hypsizygus marmoreus	51987-8	POP	NCBI	KYQ30898	LUEZ01000114
Laccaria bicolor	–	POP	JGI	303722	
Lentinus tigrinus	–	POP	JGI	581405	
Leiotrametes sp.	BRFM 1775	POP	JGI	1337173	
Lepiota subincarnata	–	POPB	JDW*	–	
	–	POPB	JDW*	–	

Table 1. (Continued).

Taxon	Strain	Prolyl Oligopeptidase	Source	Amino Acids	Coding Sequence
Lepista nuda	CBS 247.69	POP	JGI	1271729	
Leucogyrophana mollusca	KUC20120723A-06	POP	JGI	1128529	
Macrolepiota fuliginosa	MF-IS2	POP	JGI	799170	
Malassezia pachydermatis	CBS 1879	POP	NCBI	KOS13970	LGAV01000004
Marssonina brunnea f. sp. '*multigermtubi*'	MB_m1	POP	JGI	1261	
Metarhizium robertsii	ARSEF 23	POP	JGI	9962	
Ophiocordyceps sinensis	CO18	POP	NCBI	EQL04271	KE652172
Panus rudis	PR-1116ss-1	POP	JGI	1587789	
Paxillus adelphus	Ve08.2h10	POP	JGI	30278	
Phanerochaete carnosa	HHB-10118-Sp	POP	JGI	191914	
Pisolithus tinctorius	Marx 270	POP	JGI	969495	
Pleurotus ostreatus	PC 15	POP	JGI	1092697	
Plicaturopsis crispa	–	POP	JGI	170617	
Pluteus cervinus	NL-1719	POP	JGI	759937	
Polyporus brumalis	BRFM 1820	POP	JGI	1399521	
Rhizoctonia solani	AG-1 IB	POP	JGI	7338	
Rhizopogon vinicolor	AM-OR11-026	POP	JGI	794021	
Schizophyllum commune	Tattone D	POP	JGI	421410	
Serpula himantioides	SHA21-2	POP	JGI	40692	
Suillus decipiens	EM49	POP	JGI	1139982	
Trametes cingulata	BRFM 1805	POP	JGI	1559098	
Trametes versicolor	–	POP	JGI	117177	
Xerocomus badius	84.06	POP	JGI	1422158	
Wolfiporia cocos	MD-104 SS10	POP	JGI	89370	
Stereum hirsutum	FP-91666 SS1	POP	JGI	59870	

* Provided by Jonathan D. Walton at Michigan State University.

Genome: obtained from *Amanita* genomes (Supplementary file 1).

Gene tree *vs*. species tree

A *POP* gene tree and a species tree were prepared as follows. For the *POP* gene tree, 6 species containing *POPB* and 12 other related agarics were selected (Suppl. File 5). For the species tree, 30 taxa encompassing *POPB*-possessing species were chosen, and the taxa covered those in the gene tree (Suppl. File 6). CDS sequences of *POP* and *rpb2* marker were applied for the two ML reconstructions. The *rpb2* sequences were obtained from our custom genomes, deposited in GenBank and JGI genomes (blastp) using *A. subpallidorosea rpb2* (KP691703) as the query. Alignments, model selection (GTR + I + G), and ML phylogeny were performed as described above. Comparison of distances and substitution rates of the resultant gene and species trees among the three amanitin-producing genera were performed using codeml in PMAL4.9 (Yang 2007) with REV model and runmode set to -2. Both trees were further analyzed by Notung 2.9 (Chen *et al.* 2000) with Divergence-Loss (DL) and Divergence-Transfer-Loss (DTL) models under default settings. The statistics were recorded and compared. The model with lower "Event Score" was chosen to show predicted evolutionary events.

Topology test

In order to test how congruent the topology of the POPB clade is with that expected based on the species tree, alternative topologies/hypotheses were generated using PAUP (4.0b10, the following CONSEL analysis does not support RAxML outputs) for comparison with the best tree generated by this program. These hypotheses are: (1) *POPB*s were monophyletic with *Amanita POPA*; (2) *POPB*s were monophyletic with *Amanita POPA* and *Galerina POP*s; and (3) *POPB*s were monophyletic with *Galerina POP*s. The alternative trees with site-wise log-likelihoods were input into CONSEL (V0.1i) to perform approximately unbiased tests (Shimodaira & Hasegawa 2001).

K-mer analyses

The sequences used to generate the *POP* gene tree were compared using the k-mer profiles for 2-, 3-, and 4-mers. K-mer profiles were used to calculate the pairwise distances (cosine distance), and the resulting distance matrices were used to generate neighbor-joining trees. Trees were drawn using the itol webserver (https://itol.embl.de).

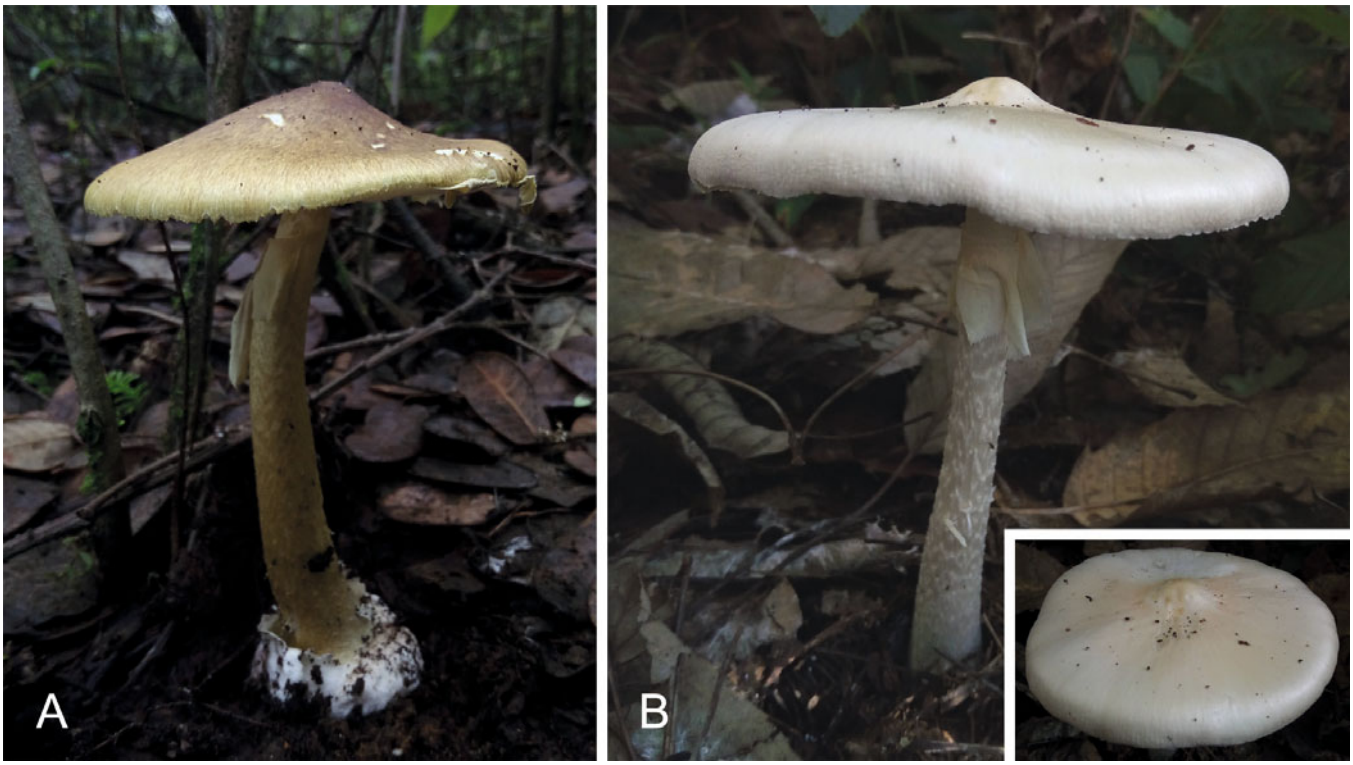

Fig. 1. Deadly *Amanita subjunquillea* and *A. pallidorosea* collected in China for genome sequencing: **A.** *A. subjunquillea* (HKAS 54509). **B.** *A. pallidorosea* (HKAS 82350); Note that the preserved specimens were different basidiomes from the same collection sites. The inset in B shows the characteristic light rose-tinged colour on the cap of the species.

RESULTS

Amanita genomes

Fresh tissues of *Amanita subjunquillea* and *A. pallidorosea* were obtained in excellent condition (Fig. 1). The resultant two draft genomes have the following statistics: *A. subjunquillea* N50 is 680 kb with the genome size at 53 Mb (GC content = 46.57 %) with the assembly containing 148 scaffolds; *A. pallidorosea* N50 is 450 kb, and the genome reached 56 Mb (GC content = 46.24 %) and the assembly comprised 251 scaffolds. Our assemblies had improved on N50s compared with previous genome sequencing of deadly *Amanita* mushrooms (Pulman *et al.* 2016). Because MSDIN genes are short and *POP* genes are intron-rich, reliable automated annotation cannot be achieved through standardized pipelines; these were therefore annotated manually. Further details of the genomes will be discussed elsewhere.

MSDINs and *POP* genes

Biosynthetic gene mining in the genomes revealed MSDIN and *POPB* genes in both of the newly sequenced *Amanita* species, in agreement with results from other amanitin-producing mushrooms in *Amanita* and *Galerina*. Tables 2 and 3 list the predicted MSDINs in both genomes, respectively. *Amanita subjunquillea* yielded 18 MSDINs while *A. pallidorosea* produced 29, each containing four toxin MSDIN genes that code for amatoxins or phallotoxins, namely the ones encoding α-amanitin, β-amanitin, phallacidin and phalloidin. Overall, the precursor genes have similar structures (leader peptide, core peptide, and recognition sequence; Arnison *et al.* 2013) with strong conservation in leader peptide and recognition sequence regions. Exon and intron structures are conserved among all MSDIN genes, and the alignments indicate that the genes span four exons and three introns, with the first intron disrupting the coding region.

In *A. subjunquillea* and *A. pallidorosea*, both *POPA* and *POPB* were unambiguously identified. The sequences share high similarity with those in *A. bisporigera*, which allowed the introns to be determined by aligning the cDNAs of *AbPOPA* and *AbPOPB* to their counterpart genomic DNAs in the newly sequenced *Amanita* genomes. All the genomic DNA sequences have highly conserved exon-intron structures, each having 19 predicted exons and 18 introns. *POP* genes from *A. phalloides* were obtained similarly, and the conservation is higher among *Amanita* species compared with *Galerina* and *Lepiota POP*s. *POPA* and *POPB* gene and aa sequences from the three *Amanita* species are given in Suppl. File 1. Surprisingly, only one *POP* gene has been identified in *L. subincarnata* (Jonathan D. Walton, pers. comm.). The gene has strongest homology to known *POPB* sequences and is therefore named *LsPOPB* for reference purpose. The alignment of *GmPOPB* cDNA in *Galerina* and *LsPOPB* genomic DNA predicted the gene's structure to have 17 exons and 16 introns.

MSDINs from *L. brunneoincarnata* were obtained by PCR using conserved sequences found in poisonous species of *Amanita* or *Galerina*. So far, only two MSDINs have been retrieved and they both code for α-amanitin.

Comparison of toxin MSDIN genes in three amanitin-producing agaric genera

In this study, toxin MSDINs are defined as the precursor genes in the MSDIN family that encode amatoxins or phallotoxins, the

Table 2. MSDIN peptide sequences from *Amanita subjunquillea* (*As*).

No.	Leader peptide	Core Peptide	Recognition Sequence	Scaffold	Coding and Notes
1	MSDINATCLP	**IWGIGCNP**	**IWGIGCNP**	1	α-amanitin
2	MSDINATRLP	**IWGIGCNP**	**IWGIGCNP**	1	β-amanitin
3	MSDINATRLP	**IWGIGCDP**	CIGDDVTALLTRGEALC	1	β-amanitin
4	MSDINATRLP	**AWLVDCP**	CVGDDVTALLTRGEALC	1	phallacidin
5	MSDINTARLP	**HFASFIPP**	CIGDDVTALLTRGEALC	1	
6	MSDINTARLP	**TFLPPLFVPP**	CVSDDVTALLTRGEALC	1	
7	MSDINATRLP	**AWLATCP**	CAGDDVTALLTRGEALC	2	phalloidin
8	MSDINATRLP	**LNILPFMLPP**	CVGDDVTALLTRGEALC	54	
9	MSDMNATRLP	**LIQRPFAP**	CVSDDVTALLTRGEALC	41	
10	MSDMNATRLP	**LIQRPYAP**	CVSDDVTALLTRGEALC	10	
11	MFDINITRLP	**IFWFIYFP**	CVGDDVTALLTRGEALC	8	
12	MSDINTARLP	**IGRPESIP**	CVGDDVTALLTRGEALC	13	
13	MSDINTARLP	**LRLPPFMIPP**	CVGDDVTALLTRGEALC	13	
14	MSDMNVARLP	**ISDPTAYP**	CVGGDVTALLTRGEALC	13	2 copies, in *Apa*
15	MSDINTVCLP	**LQKPWSRP**	CVGDDVTALLTRGEALC	13	
16	MSDVNATRLP	**FNFFRFPYP**	CIGDDVTALLTRGEALC	38	
17	MSDINATRLP	**SSVLPRP**	CVGDDVTALLTRGEALC	20	
18	MSDMNVARLP	**ISDPTAYP**	CVGGDVTALLTRGEALC	22	2 copies, in *Apa*

Core peptides (correspond to cyclic peptide toxins) are in bold.

The conserved MSDIN (except for the ones with variations) is underlined.

No. 14 and 18 have two copies and are also in *A. pallidorosea*.

Table 3. MSDIN peptide sequences from *Amanita pallidorosea* (*Apa*).

Name	Leader peptide	Core Peptide	Recognition Sequence	Scaffold	Coding
1	MSDINATRLP	**IWGIGCNP**	CVGDDVTALLTRGEALC	52	α-amanitin, 2 copies
2	MSDINATRLP	**IWGIGCNP**	CVGDDVTALLTRGEALC	52	α-amanitin, 2 copies
3	MSDINATRLP	**IWGIGCDP**	CVGDDVTALLTRGEALC	52	β-amanitin
4	MSDVNATRLP	**MAFPEFLA**	CVGDDVTALLTRGEALC	52	
5	MSDINATRLP	**AWLMTCP**	CVGDDVTALLTRGEALC	7	
6	MSDVNATRLP	**AWLVDCP**	CVGDDVTALLTRGEALC	62	phallacidin
7	MSDINASRLP	**FFPEVGFFP**	CVGDDVTALLTRGEALC	16	
8	MSDINAIRAP	**LPIFSLNP**	CVGDDVTALLTRGEALC	16	
9	MSDINATRLP	**NWHAGPTRPP**	CVADDVTALLTRGEALC	16	
10	MSDINAARLP	**NLFVWIPP**	CISDDVTALLTRGEALC	16	
11	MSDINASRLP	**AWLATCP**	CAGDDVTALLTRGEALC	16	phalloidin
12	MADINASRLP	**LNILPFHLPP**	CVSDDVTALLTRGEALC	16	
13	MSDINATRLP	**LGRPESLP**	CVGDDVTALLTRGEALC	3	
14	MSDINATRLP	**HPFPLGLQP**	CAGDDVTALLTRGEALC	73	2 copies
15	MSDINATRLP	**HPFPLGLQP**	CAGDDVTALLTRGEALC	101	2 copies
16	MSDMNVVRLP	**ISDPTAYP**	CVGDDVTALLTRGEALC	102	2 copies, in *As*
17	MSDINVIRLP	**IF*FIYFP**	CVGDNVTALLTRGEALC	60	
18	MSDINTARLP	**MHILAPPP**	CVSDDVTALLTRGEALC	10	
19	MSDMNVVRLP	**ISDPTAYP**	CVGDDVTALLTRGEALC	64	2 copies, in *As*
20	MSDTNATRLP	**SIFIVYPP**	CVSDDVTALLTRGEALC	86	
21	MSDINAARLP	**LVYMILFP**	SVGDDVTALLTRGEALC	111	
22	MSETNAARLP	**TIHLFSAP**	SVGDDVTALLTRGEALC	111	

Table 3. (Continued).

Name	Leader peptide	Core Peptide	Recognition Sequence	Scaffold	Coding
23	MSDINTARLP	**EFIVFGIFP**	CVGDDVTALLTRGEALC	78	
24	MSDINTTRLP	**YFFNDHPP**	CASDDVTALLTRGEALC	78	
25	MSDLNATRLP	**FNLFRFPYP**	CIGDDVTALLTRGEALC	9	
26	MSDINTARLP	**VFFMPPFIPP**	CVSDDVTALLTRGEALC	1	
27	MSNINTARLP	**FLVPSFPP**	CVSDDVTALLTRGEALC	1	
28	MYDINTTRLP	**HFFNLTPP**	CVRDDVTALLTRGEALC	39	
29	MSDINTARLP	**FVIIPPFIFP**	CVSDDVTALLTRGEALC	12	

Core peptides (correspond to cyclic peptide toxins) are in bold.

The conserved MSDIN (except for the ones with variations) is underlined.

No. 14 and 15 have two copies.

No. 16 and 19 have two copies and are also in *A. subjunquillea*.

*indicates a stop codon exists in the sequence.

major cyclic peptides in these agarics. The MSDIN sequences in *Amanita* are generally conserved (Fig. 2), with highlighted variations compared to the best represented consensus sequence from *Amanita* (not including the core peptides), although some variations are found in *A. phalloides* and two Asian *Amanita* species, *A. subjunquillea* and *A. pallidorosea*. This result also shows that Asian *Amanita* species share higher conservation with lowest amount of variation (red letters). In contrast, *A. phalloides* from Europe showed higher variation in its recognition sequences. Gene duplications are found in many species: in *A. bisporigera*, each copy for the listed two MSDIN genes is identical, indicating duplication happened recently without any accumulated variations. In *A. phalloides* and Asian amanitas, duplicates usually present some variations, indicating these duplications formed some time ago. In *Lepiota*, the leader peptides are 9 aa in length, while all others are 10 aa. In general, leader peptides are more conserved than the other sequences. In this case, the variations in this region reflect their generic position: MSDIN is specific to *Amanita*, MFDTN to *Galerina*, and MDAN to *Lepiota* (including *L. subincarnata* genomes). These sequences are highly conserved within the same genus, and true MSDIN sequences only exist in deadly *Amanita* species (underlined). Some sequences are highly conserved even across genera, including NATRLP in the leader peptide and LC (IC in *G. marginata*) at the very end of the recognition sequence. In addition, LTRG in the recognition sequence is conserved between the genera *Lepiota* and *Amanita*. Sequences of *Galerina* are closer to those of *Lepiota* than to those of *Amanita*: 10–11 variations *vs*. 15–16.

Phylogeny of *POP*s in macrofungi

The aligned coding sequences of *POP* genes comprised 75 taxa (Table 1) with 3489 bp in length (Suppl. File 2), and the dataset of amino acid sequences included sequences of the same 75 species with 1035 aligned sites (Suppl. File 3). The nucleotide alignment of MSDIN genes consisted of 21 sequences with 376 bp (Suppl. File 4). ML and BI analyses yielded identical tree topologies, and thus only the trees inferred from the ML analysis are shown (Figs 3–5). With regard to the *POP* genes, the lineage POPB from CDS (Fig. 4) had stronger statistical support than that from amino acid sequences (Fig. 3). This might be due to the higher conservation at the amino acid level due to degeneracy, quenching some of the phylogenetic signal that is present in coding sequences. The good support in the terminal clades including *POPA* and *POPB* allowed us to delineate boundaries of these genes. In the phylogenetic trees, *POPB* consistently formed a clade with strong support, while the remaining *POP* genes, including *POPA,* generated multiple strongly supported clades. *Galerina POPA* and *Amanita POPA* did not cluster together, but, rather, with *POP* genes from taxonomically related species, suggesting that "*POPA"* is simply the generic *POP* gene present in most basidiomycetes (Hibbett 2006, Matheny *et al.* 2006, 2007, Justo *et al.* 2011). Notable examples include one clade representing all *POP*s from the order *Boletales* (Binder & Hibbett 2006) and another from the order *Polyporales* (Justo *et al.* 2017). Strikingly, *POPB* displayed a very different pattern: all *POPB* genes, from three genera belonging to three disjunct families, clustered together forming a well-supported monophyletic clade (lineage POPB). Topology within *POPB* reflects species phylogeny; however, the apparent single origin of *POPB* within the *POP* tree (as opposed to derivation of *POPB* from each species' *POPA*) requires further explanation.

Phylogeny of MSDIN genes

The MSDIN phylogenetic tree was constructed without an outgroup. Fig. 5 shows that the MSDIN sequences from *Lepiota* and *Amanita* are separated and well supported. MSDINs are rather short and the hypervariable region (encoding cyclic peptides) interferes with phylogenetic analysis. As shown in the figure, *Amanita* MSDINs do not cluster according to species phylogeny, but do so based on chemical properties, in this case the toxins they encode. As a consequence, in *Amanita*, MSDINs encoding α-amanitin, β-amanitin, phalloidin, and phallacidin group together, respectively.

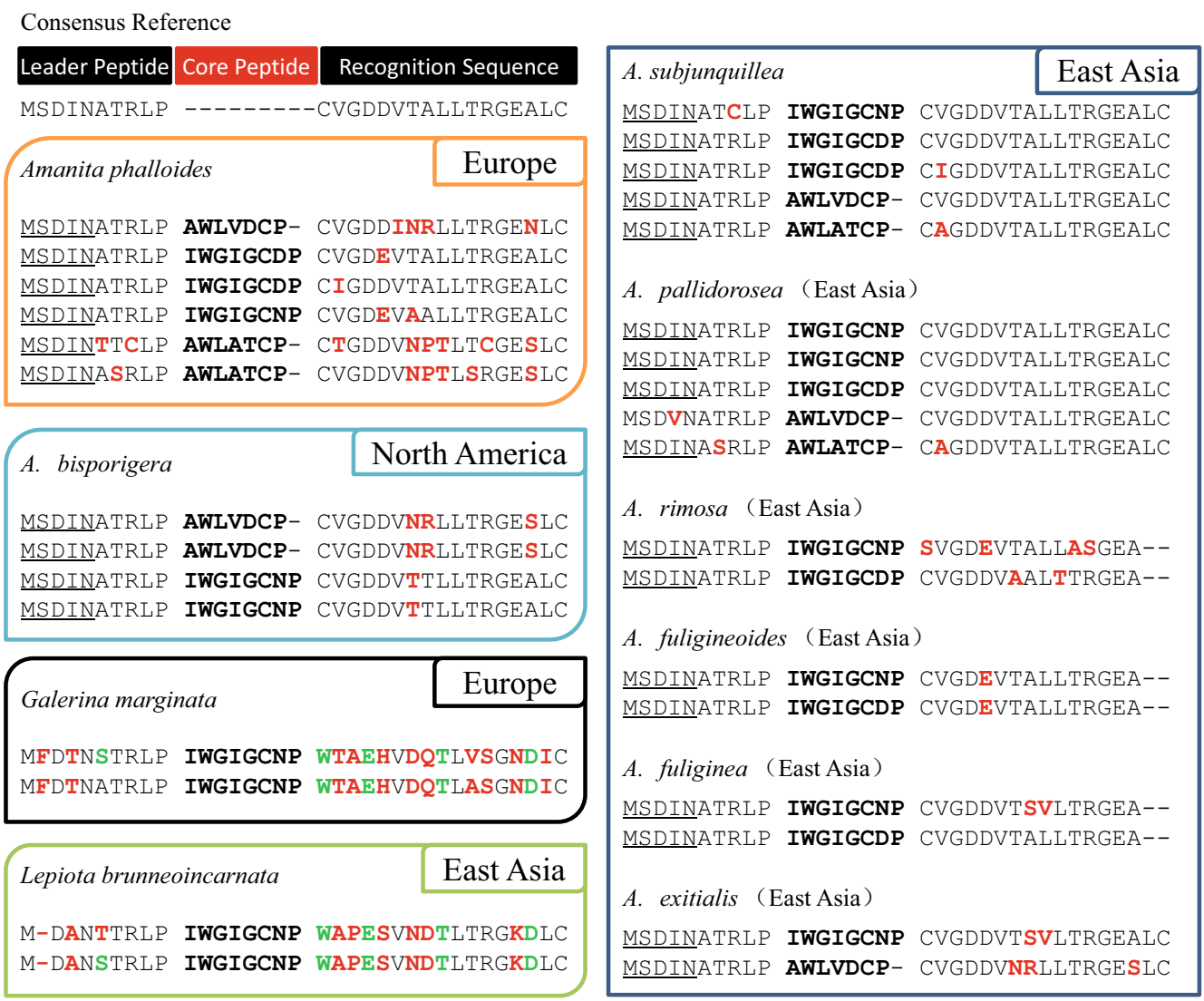

Fig. 2. Overview of MSDIN family genes in the three amanitin-producing agaric genera. Upper left indicates the best represented consensus sequence, with a general schematic structure of MSDIN genes with leader peptide, core peptide, and recognition sequence (not including the core peptide, the hypervariable region). Coloured boxes harbor species from specific geological locations. Coloured (red and green) letters indicate variations (differences) compared with the consensus sequence. In *G. marginata* and *L. brunneoincarnata*, green letters also designate conserved amino acids only between the two species. MSDIN sequences are underlined, and true MSDIN sequences are only found in *Amanita* species.

Gene tree *vs*. species tree

In order to investigate the relationship between *POP* gene tree and the corresponding species tree, a gene tree based on *POP*s (Fig. 6A) and a species tree based on *rpb2* marker (Fig. 6B), were constructed. The pairwise distances and substitution rates among three species representing the three disjunct genera were calculated (Table 4). Consistent with the general hypothesis that genes acquired *via* HGT would show significantly less divergence compared with house-keeping genes, the result indicated significantly smaller distances (to 1:6) and substitution rates (to 1:7) from the gene tree *vs*. the species tree. This result also allowed the discounting of the massive gene loss hypothesis, in which case the distances and substitution rates are expected be similar. In light of this, the non-POPB-containing taxa in the species cannot be removed in the topology comparison, and

Table 4. Comparison of distances and substitution rates of gene and species trees among the three amanitin-producing genera.

Species	Distance (Gt)	Distance (St)	dN (Gt)	dN (St)	dS (Gt)	dS (St)	dN/dS (Gt)	dN/dS (St)
A. rimosa vs. *G. marginata*	12.238	64.374	0.1572	0.0650	10.858	72.961	0.1448	0.0089
A. rimosa vs. *L. subincarnata*	17.565	43.247	0.1917	0.0916	16.058	52.356	0.1194	0.0175
G. marginata vs. *L. subincarnata*	14.226	34.662	0.1760	0.0836	12.357	49.097	0.1424	0.0170

Note: Gt = gene tree; St = species tree.

Fig. 3. Phylogeny of macrofungi inferred from maximum likelihood (ML) analysis based on amino acid sequences of *POP* gene. Maximum likelihood bootstraps over 50 % and Bayesian posterior probabilities over 0.90 are given at the internodes. Asterisks indicate orthologs.

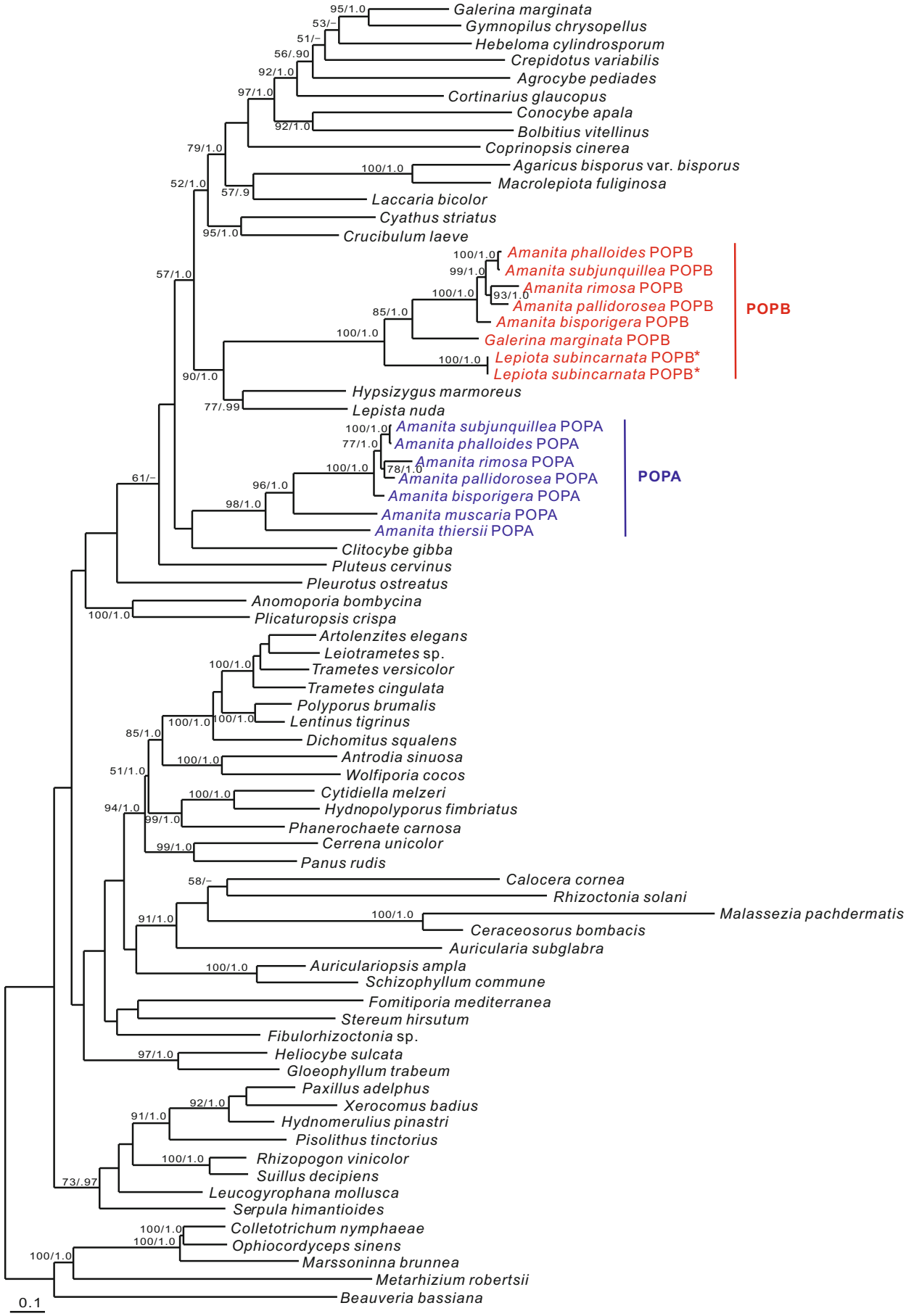

Fig. 4. Phylogeny of macrofungi inferred from maximum likelihood (ML) analysis based on coding sequences of *POP* gene. Maximum likelihood bootstraps over 50 % and Bayesian posterior probabilities over 0.90 are given at the internodes. Asterisks indicate orthologs.

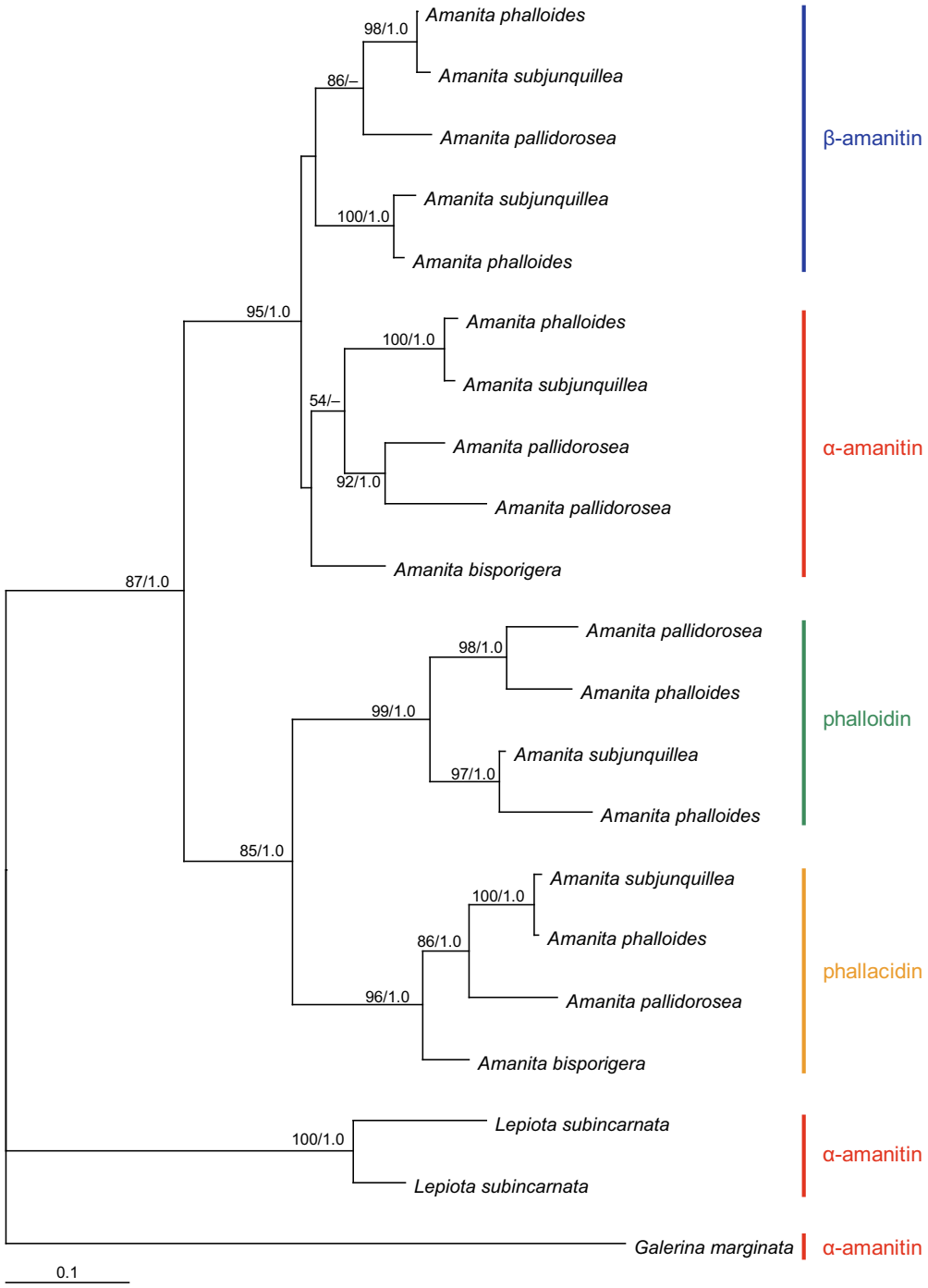

Fig. 5. Phylogeny of MSDIN genes in *Amanita, Galerina* and *Lepiota* inferred from maximum likelihood (ML). Maximum likelihood bootstraps over 50 % and Bayesian posterior probabilities over 0.90 are given at the internodes.

then incongruency was shown in 6A and 6B, with the POPB subclade marked in red. The strong statistical support ruled out the possibility of aligning conflicting clades under the settings. For example, *Galerina marginata* (POPB clade) will not cluster with *Gymnopilus chrysopellus* POP in the gene tree as the species do in the species tree. With Notung, the DL model returned the following general statistics: Event Score = 36.0, Dups = 4, Losses = 30, and Numbers of optimal solutions = 1. The DTL model produced: Event Score = 23.0, Dups = 0, Transfers = 5, Losses = 8, and Numbers of optimal solutions = 4. The DTL score was significantly smaller and therefore further analysis was based on DTL results. The DTL reconciled tree with one of the four optimal solutions is shown in Fig. 6C; their differences are at the gene transfer points (green circles). The illustrated transfer events in the POPB clade (T1, T2, and T3) indicate the possibility that the HGT happened from *L. subincarnata* to *G. marginata* (T1), then to an unknown species between *G. marginata* and *Amanita rimosa* (T2), and followed by another transfer within *Amanita* (T3). Three other optimal solutions have slightly different routes, but all indicated gene transfer.

Topology tests

The phylogenetic trees generated above shows that the POPB clade is not congruent with the species tree. The robustness of the POPB clade was assessed in this study.

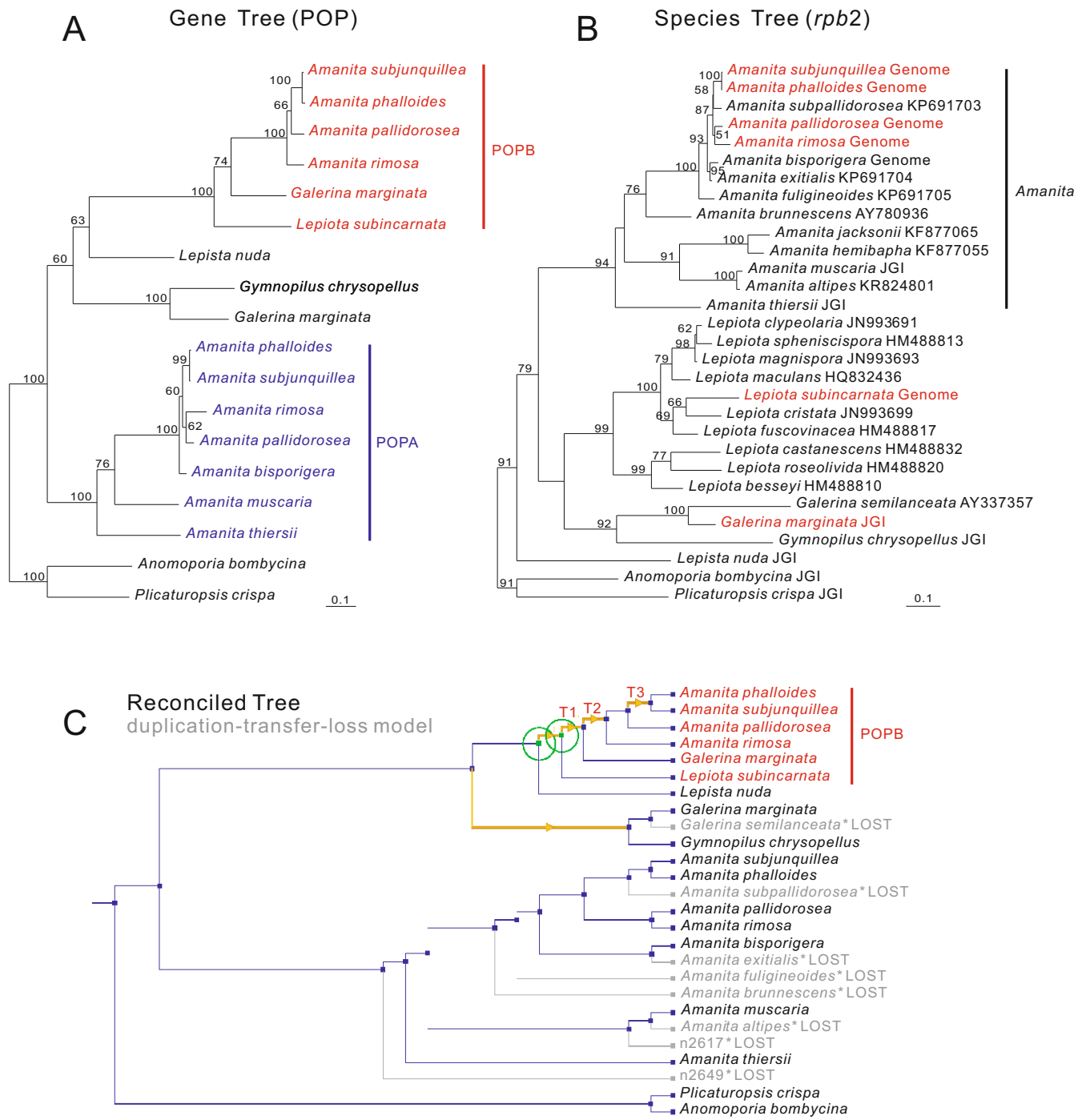

Fig. 6. Comparison of *POP* gene tree and species tree. **A.** *POP* gene tree. POPB lineage was highlighted in red; POPA lineage in blue. **B.** Species tree based on *rpb2*. Corresponding species for the POPB lineage were highlighted in red. *Amanita* lineage was indicated by a black bar. **C.** Reconciled tree by Notung. Yellow arrows indicated gene transfer. The green circles showed where the alternative transfer events occured. The transfer events in the strongly supported POPB lineage were marked as T1, T2 and T3.

The best tree generated by PAUP was consistent with those by RAxML (Figs 3–4, 6) and is not shown here. With the three alternative trees, Table 5 shows that the best tree by PAUP is highly supported over the alternative topologies for competing hypotheses, with both approximately unbiased *p*-values (AU) and bootstrap probability (NP) at 1. This result strongly suggested the monophyletic POPB clade is highly supported, rejecting a *de novo* origin of *POPB* from *POPA* within species.

Gene structure

As phylogenetic data cannot fully rule out an ancestral origin of *POPB* followed by multiple independent losses, we examined the intron structure of *POPB* and toxin MSDIN genes, and evaluated di-, tri-, and tetranucleotide frequencies of representative *POP* genes. Toxin MSDIN genes each contain three introns with a conserved size and placement, one near the 3' end of the coding region and two within the 3' UTR (Hallen *et al*. 2007, Luo *et al*. 2012), while *POPB*

Table 5. Approximately unbiased test on alternative *POP* trees.

Rank	Best tree vs. Hypothetical trees	obs	AU	NP
1	Best tree	-3346.3	1.000	1.000
2	*POPA* and *POPB* monophyletic	3346.3	100000000000,00E-7	3000000000000,00E-6
3	*POPA*, *POPB* and *Galerina POP* monophyletic	4866.1	2000000000000,00E-8	2000000000000,00E-7
4	*POPB* and *Galerina POP* monophyletic	5232.1	100000000000,00E-51	3000000000000,00E-18

obs: observed log-likelihood difference

AU: approximately unbiased *p*-values

NP: bootstrap probability

genes contain 17 introns of similar size and placement (one additional intron is present in *Amanita bisporigera;* Fig. 7). The k-mer profiles of *POP* genes consistently group *Galerina POPB* with *Amanita POPB* and *Amanita POPA*, and are distinct from *Galerina POPA* (Fig. 8). *Lepiota POPB* has distinct k-mer profiles from the other *POPB* genes.

DISCUSSION

All known amatoxin-producing *Amanita* species belong to section *Phalloideae*, which has at times been restricted to only lethal species. Recently the number of taxa in the section has undergone a minor expansion, and now includes at least four non-poisonous species in basal positions of the clade. Phylogenetic evidence also indicates a single origin of the cyclic peptide pathway within *Amanita* (Cai *et al.* 2014, Cui *et al*. 2018).

Diversity of MSDIN genes in three agaric genera

The genes discovered to date clearly share similar structures (including exon and intron structure), with leader peptide, core peptide and recognition sequence (Arnison *et al.* 2013), indicating they shared a common ancestor. The Asian *Amanita* species possess a similar pool of these genes compared to their European and North American relatives. In general, the toxin MSDINs (genes encoding amatoxins or phallotoxins, i.e. α-amanitin, β-amanitin, phallacidin, and phalloidin) are shared among amanitas, while the rest do not overlap significantly. Much less diversity of MSDIN genes was observed in *Galerina* and *Lepiota* species. *Galerina marginata* only possesses one *GmAMA1* gene in two copies; a rigorous PCR search only found two *AMA1* genes in two *L. brunneoincarnata* strains (our initial genome assembly now confirms this). Besides the structural similarities, other variations were evident. The actual "MSDIN" motif is restricted to *Amanita*, and in *Lepiota* and *Galerina*, the variations in this leader peptide region are distinctive, and likely genus-specific. In the recognition sequence region, there are conserved aa residues across genera but with a significant number of variations (Fig. 2).

Evolution of *POPB*

With limitations, phylogenetic reconstruction methods remain the only way to reliably infer historical events from gene sequences as they are the only methods that utilize large, comprehensive data sets (Eisen 2000). Non-tree-based ("surrogate") methods are increasingly used in identifying instances of lateral genetic transfer, but in many cases they lack reliability compared to rigorous phylogenetic analysis; further, phylogenetic methods are less dependent on subtle nucleotide-level signatures that could be unevenly distributed and subject to amelioration (Ragan *et al.* 2006). For these reasons, we took the phylogenetic route to assess the evolutionary history of the key biosynthetic gene *POPB*. In the distance and substitution rate analyses, significantly smaller numbers in both indicate the POPB lineage evolved at a much lower speed, incongruent with that of the three disjunct genera in the *rpb2* species tree, but consistent with the hypothesis of HGT. We therefore continued the topology comparison without the consideration of massive gene loss.

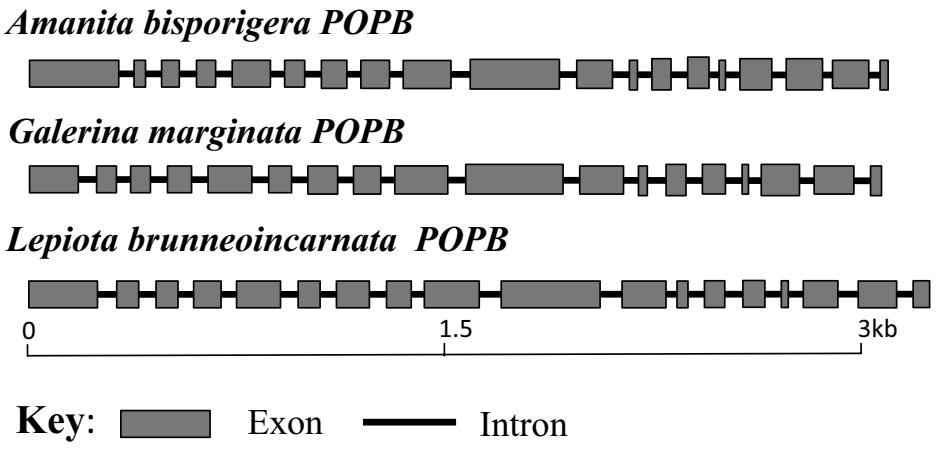

Fig. 7. *POPB* intron and exon structure for *Amanita, Galerina,* and *Lepiota.*

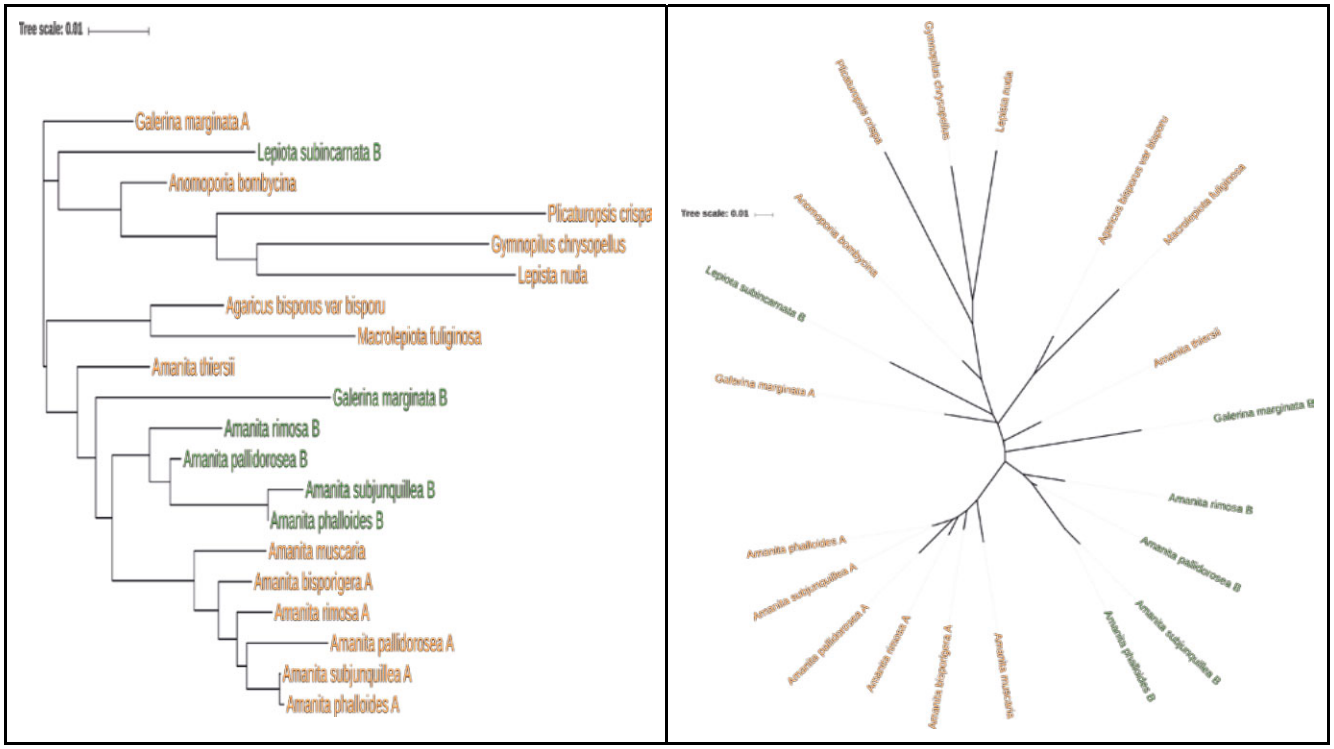

Fig. 8. Tetranucleotide frequency analysis of the POP sequences used in Fig. 6A. Tree given as rooted (left) and unrooted (right). *Galerina* POPB is placed within *Amanita* POPs, while *Lepiota* POPB is distinct. Di- and trinucleotide analyses give similar results (data not shown).

From the ML tree topologies, all *POPB*s reside exclusively in a well-supported monophyletic clade (lineage POPB), with the saprotrophic species *L. subincarnata* and *G. marginata* in basal positions. In contrast, all *POPB*s from *Amanita* species are terminal (Figs 3–4, 6). This scenario indicates *POPB*s from the saprobes are ancestral whereas *Amanita POPB*s are newer entities. The highly supported topologic incongruency between the *POP* gene tree and the species tree strongly suggest the acquisition of *POPB* in those lineages were likely the result of HGT. We also tested the incongruency using different markers, such as LSU, and the results were consistent, although some were with weak statistical support. The illustrated incongruency between *POP* gene trees and the species tree strongly suggested an HGT cause of the *POPB* distribution among *Lepiota*, *Galerina*, and *Amanita*. In addition, the topology test showed strong support for the POPB clade, rejecting all other three competing hypotheses. In Notung analysis, all predicted four best DTL solutions involve HGT with only minor variations, lending more support to the hypothesis.

Since *POPB* is a single-copy gene and is at the centre of the cyclic peptide biosynthesis, its phylogeny may reflect the evolutionary history of the α-amanitin biosynthetic pathway. In contrast, MSDIN sequences are less usable as they are short and have a highly variable region (core peptide) that interfers with the phylogenetic methods by pulling genes for same cyclic peptides together, although coding sequence analysis showed less of this problem than aa phylogeny (Fig. 5). The existence of the core peptides also partly causes low statistical support. Lacking a proper outgroup is another reason not enough information was obtained through the analysis.

Hypothesis for transferring the cyclic peptide pathway

At least three possible hypotheses for how the cyclic peptide pathway evolved in the three disjunct agaric genera can be envisioned: (1) the pathway arose independently in each of the genera; (2) the pathway originated in a common ancestor but was lost in most of the descendants except for the three genera; or (3) the pathway formed in a common ancestor and was then transferred through HGT to other recipients. If the pathway was due to independent origins as a result of convergent evolution, little resemblance among the pathways in the three genera would be expected. However, all pathways use MSDIN genes for the precursor peptides, and all the MSDINs share a conserved structure that features leader peptide, core peptide, and recognition sequence (Arnison *et al.* 2013). Furthermore, they all possess a specialized *POPB* gene that clusters into a single lineage, and they too, like the MSDINs, share exon and intron structures, indicating these genes are from a common ancestor (Fig. 7). Regarding the second hypothesis, our analyses on substitution rates and distances indicated that *POPB* genes evolve much slower than the house-keeping gene *rpb2*, consistent with the HGT hypothesis while conflicting with massive gene loss. The species trees only have a small subset of taxa, and the differences in distances and rates would only increase if more species are included. In addition, among the three families, we counted over 2000 species (or significantly more as the count was not complete). If there was a common ancestor in which the pathway originated, then thousands of agarics would have to lose at least two genes (MSDIN and *POPB*) to accommodate the toxin distribution. While this is not

entirely impossible, we consider the chance slim. Further, one would expect some pathway remnants detectable in the closely-related genomes, but in our BLAST searches and comparative genomic study, none has been found. As discussed above, our study now lends some support for the third hypothesis. Multiple phylogenetic reconstructions, comparison of substitution rates and distance, analyses of gene tree and species tree, predicted evolutionary events, and topology test, all suggest HGT was the underlining cause for the disjunct toxin distribution.

Other information pertaining to the HGT hypothesis

We have cloned one Class II transposon close to *GmPOPB* in *G. marginata* (unpubl.), and both Class I and Class II transposons (~ 60 within the 50 kb range) were detected in our initial assembly of *L. brunneosubincarnata* (unpubl.). Therefore, HGT of *POPB* could be assisted by transposons. It is known that at least some degree of gene clustering in this pathway is present in *A. bisporigera* and *G. marginata* (Luo *et al*. 2010, 2012). Clustering of genes is considered to be able to assist in HGT. Supernumerary chromosome transfers can also be explained by interspecific mating rather than HGT, but our comparative genomic study using Symap was negative on this assumption. K-mer analysis (Fig. 8) suggests strongly that *Galerina POPB* is the result of HGT from *Amanita*, while *Lepiota POPB* clusters neither with the other *POPB* genes, nor with the *POPs* from related species (*Macrolepiota fuliginosa* and *Agaricus bisporus*).

Fungal HGT

Most early reports on HGT cases involve bacterial donors, however recent studies show-cased an increasing number of HGT events between fungi and other eukaryotes (Slot 2017). Some good examples include depudecin biosynthesis in *Alternaria brassicicola* (Reynolds *et al.* 2017), ergotamine and loline biosyntheses in fungi belonging in *Clavicipitaceae* (Marcet-Houben & Gabaldon 2016), chaetoglobosin-like compound biosynthesis in *Mycosphaerella populorum* (Dhillon *et al.* 2015), fumonisin cluster in *Aspergillus niger* (Khaldi & Wolfe 2011), avirulence gene ACE1 cluster in *Magnaporthe grisea*, *Chaetomium globosum*, *Stagonospora nodorum,* and *A. clavatus* (Khaldi *et al.* 2008), four pathogenic chromosomes in *Fusarium* (Ma *et al.* 2010), eight chromosomes in the wheat pathogen *M. graminicola* (Goodwin *et al*. 2011), and a virulence gene called ToxA in *Pyrenophora tritici-repentis* (Friesen *et al.* 2006). Recent phylogenomic studies indicate that in a "typical" fungal genome, between 0.1–2.8 % of the genes may be the result of HGT (Wisecaver *et al.* 2014, Wisecaver & Rokas 2015), a much lower proportion than that in bacteria and archaea (Koonin *et al.* 2001). The structures of the MSDIN and *POPB* genes include multiple introns, strongly suggesting the fungal origin of these genes. Therefore, the evolution of *POPB* should be an example of fungal HGT. Based on the phylogenetic positions of *POPB*s from *Lepiota*, *Galerina*, and *Amanita*, one can conjecture the rough direction of the gene acquisition could also be in this direction. Substantial work is needed to provide more insights into this aspect.

Possible impact of the biosynthetic pathway

A long-observed phenomenon in ectomycorrhizal *Amanita* species is the rapid speciation within the lineage of lethal amanitas of the sect. *Phalloideae*, while the number of the non-amanitin-producing taxa in the section is significantly lower (Cai *et al*. 2014, Cui *et al*. 2018). A logical question regarding these opposing trends is: did the pathway drive the rapid speciation in deadly *Amanita*? At present, we have little clue regarding the target organisms, biological roles or selective advantages of amanitins, even less on many other cyclic peptides made by these agarics. Hopefully, the fast development in genome research combined with our transformation system will point us to the right directions.

ACKNOWLEDGEMENTS

This research was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant no. XDB31000000), Natural Science Foundation of China (Grant no. 31772377), Scientific Research Foundation of the Education Department of Human Resources and Social Security of Yunnan Province, China, and Scientific Research Foundation of Kunming Institute of Botany, Chinese Academy of Sciences. We are very grateful to Jonathan D. Walton for his very kind help with the *Lepiota POPB* sequences, and his long-term support.

REFERENCES

Andersson JO (2005) Lateral gene transfer in eukaryotes. *Cellular and Molecular Life Sciences* **62**: 1182–1197.

Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, *et al.* (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural Product Reports* **30**: 108–160.

Binder M, Hibbett DS (2006) Molecular systematics and biological diversification of Boletales. *Mycologia* **98**: 971–981.

Boetzer M, Pirovano W (2014) SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* **15**: 211.

Bresinsky A, Besl H (1990) *A Colour Atlas of Poisonous Fungi: a handbook for pharmacists, doctors and biologists*. Würzburg: Wolfe.

Cai Q, Tulloss RE, Tang LP, Tolgor B, Zhang P, *et al.* (2014) Multi-locus phylogeny of lethal amanitas: Implications for species diversity and historical biogeography. *BMC Evolutionary Biology* **14**: 143.

Chen JT, Chao ML, Wen CY, Chu WS (2012) Screening, purification, and characterization of an extracellular prolyl oligopeptidase from *Coprinopsis clastophylla*. *Journal of Microbiology* **50**: 652–659.

Chen K, Durand D, Farach-Colton M (2000) NOTUNG: a program for dating gene duplications and optimizing gene family trees. *Journal of Computational Biology* **7**: 429–447.

Chen Z, Zhang P, Zhang Z (2013) Investigation and analysis of 102 mushroom poisoning cases in Southern China from 1994 to 2012. *Fungal Diversity* **64**: 123–131.

Chen ZH, Yang ZL, Bau T, Li TH (2016) *Poisonous Mushrooms: recognition and poisoning treatment.* Beijng: Science Press.

Cui YY, Cai Q, Tang LP, Liu JW, Yang ZL (2018) The family *Amanitaceae*: molecular phylogeny, higher-rank taxonomy and the species in China. Fungal Diversity, doi.org/10.1007/s13225-

018-0405-9

Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**: 1164–1165.

Dhillon B, Feau N, Aerts AL, Beauseigle S, Bernier L, *et al.* (2015) Horizontal gene transfer and gene dosage drives adaptation to wood colonization in a tree pathogen. *Proceedings of the National Academy of Sciences, USA* **112**: 3451–3456.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**: 1969–1973.

Duan L, Ying G, Danzer B, Perez RE, Shariat-Madar Z, *et al.* (2014) The prolyl peptidases PRCP/PREP regulate IRS-1 stability critical for rapamycin-induced feedback activation of PI3K and AKT. *Journal of Biological Chemistry* **289**: 21694–21705.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Eisen JA (2000) Assessing evolutionary relationships among microbes from whole-genome analysis. *Current Opinion in Microbiology* **3**: 475–480.

English AC, Richards S, Han Y, Wang M, Vee V, *et al.* (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* **7**: e47768.

Fitzpatrick DA (2012) Horizontal gene transfer in fungi. *FEMS Microbiology Letters* **329**: 1–8.

Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, *et al.* (2006) Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature Genetics* **38**: 953–956.

Garcia-Horsman JA, Mannisto PT, Venalainen JI (2007) On the role of prolyl oligopeptidase in health and disease. *Neuropeptides* **41**: 1–24.

Goodwin SB, M'Barek SB, Dhillon B, Wittenberg AH, Crane CF, *et al.* (2011) Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genetics* **7**: e1002070.

Hackl T, Hedrich R, Schultz J, Forster F (2014) proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* **30**: 3004–3011.

Haines JH, Lichstein E, Glickerman D (1986) A fatal poisoning from an amatoxin containing *Lepiota*. *Mycopathologia* **93**: 15–17.

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.

Hallen HE, Luo H, Scott-Craig JS, Walton JD (2007) Gene family encoding the major toxins of lethal *Amanita* mushrooms. *Proceedings of the National Academy of Sciences, USA* **104**: 19097–19101.

Hibbett DS (2006) A phylogenetic overview of the *Agaricomycotina*. *Mycologia* **98**: 917–925.

Justo A, Miettinen O, Floudas D, Ortiz-Santana B, Sjokvist E, *et al.* (2017) A revised family-level classification of the *Polyporales* (*Basidiomycota*). *Fungal Biology* **121**: 798–824.

Justo A, Vizzini A, Minnis AM, Menolli N jr, Capelari M, *et al.* (2011) Phylogeny of the *Pluteaceae* (*Agaricales*, *Basidiomycota*): taxonomy and character evolution. *Fungal Biology* **115**: 1–20.

Kaushik S, Sowdhamini R (2014) Distribution, classification, domain architectures and evolution of prolyl oligopeptidases in prokaryotic lineages. *BMC Genomics* **15**: 985.

Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics* **9**: 605–618.

Khaldi N, Collemare J, Lebrun MH, Wolfe KH (2008) Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. *Genome Biology* **9**: R18.

Khaldi N, Wolfe KH (2011) Evolutionary origins of the fumonisin secondary metabolite gene cluster in *Fusarium verticillioides* and *Aspergillus niger*. *International Journal of Evolutionary Biology* **2011**: 423821.

Kimura A, Matsui H, Takahashi T (2002) Expression and localization of prolyl oligopeptidase in mouse testis and its possible involvement in sperm motility. *Zoological Science* **19**: 93–102.

Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annual Review of Microbiology* **55**: 709–742.

Li P, Deng WQ, Li TH (2014) The molecular diversity of toxin gene families in lethal *Amanita* mushrooms. *Toxicon* **83**: 59–68.

Li R, Li Y, Fang X, Yang H, Wang J, *et al.* (2009a) SNP detection for massively parallel whole-genome resequencing. *Genome Research* **19**: 1124–1132.

Li R, Yu C, Li Y, Lam TW, Yiu SM, *et al.* (2009b) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**: 1966–1967.

Luo H, Hallen-Adams HE, Scott-Craig JS, Walton JD (2010) Colocalization of amanitin and a candidate toxin-processing prolyl oligopeptidase in *Amanita* basidiocarps. *Eukaryotic Cell* **9**: 1891–1900.

Luo H, Hallen-Adams HE, Scott-Craig JS, Walton JD (2012) Ribosomal biosynthesis of alpha-amanitin in *Galerina marginata*. *Fungal Genetics and Biology* **49**: 123–129.

Luo H, Hallen-Adams HE, Walton JD (2009) Processing of the phalloidin proprotein by prolyl oligopeptidase from the mushroom *Conocybe albipes*. *Journal of Biological Chemistry* **284**: 18070–18077.

Luo H, Hong SY, Sgambelluri RM, Angelos E, Li X, *et al.* (2014) Peptide macrocyclization catalyzed by a prolyl oligopeptidase involved in alpha-amanitin biosynthesis. *Chemistry and Biology* **21**: 1610–1617.

Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, *et al.* (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367–373.

Marcet-Houben M, Gabaldon T (2016) Horizontal acquisition of toxic alkaloid synthesis in a clade of plant associated fungi. *Fungal Genetics and Biology* **86**: 71–80.

Matheny PB, Curtis JM, Hofstetter V, Aime MC, Moncalvo JM, *et al.* (2006) Major clades of *Agaricales*: a multilocus phylogenetic overview. *Mycologia* **98**: 982–995.

Matheny PB, Wang Z, Binder M, Curtis JM, Lim YW, *et al.* (2007) Contributions of rpb2 and tef1 to the phylogeny of mushrooms and allies (*Basidiomycota*, *Fungi*). *Molecular Phylogenetics and Evolution* **43**: 430–451.

Momeni N, Nordstrom BM, Horstmann V, Avarseji H, Sivberg BV (2005) Alterations of prolyl endopeptidase activity in the plasma of children with autistic spectrum disorders. *BMC Psychiatry* **5**: 27.

Moreno-Baylach MJ, Felipo V, Mannisto PT, Garcia-Horsman JA (2008) Expression and traffic of cellular prolyl oligopeptidase are regulated during cerebellar granule cell differentiation, maturation, and aging. *Neuroscience* **156**: 580–585.

Mottram AR, Lazio MP, Bryant SM (2010) *Lepiota subincarnata* J.E. Lange induced fulminant hepatic failure presenting with

pancreatitis. *Journal of Medical Toxicology* **6**: 155–157.
Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP*, et al.* (2000) A whole-genome assembly of *Drosophila*. *Science* **287**: 2196–2204.
Nylander JA (2004) MrModeltest v. 2.2. Uppsala: Evolutionary Biology Centre, Uppsala University.
Ohtsuki S, Homma K, Kurata S, Komano H, Natori S (1994) A prolyl endopeptidase of Sarcophaga peregrina (flesh fly): its purification and suggestion for its participation in the differentiation of the imaginal discs. *Journal of Biochemistry* **115**: 449–453.
Polgar L (2002) The prolyl oligopeptidase family. *Cellular and Molecular Life Sciences* **59**: 349–362.
Pulman JA, Childs KL, Sgambelluri RM, Walton JD (2016) Expansion and diversification of the MSDIN family of cyclic peptide genes in the poisonous agarics *Amanita phalloides* and *A. bisporigera*. *BMC Genomics* **17**: 1038.
Ragan MA (2001) On surrogate methods for detecting lateral gene transfer. *FEMS Microbiology Letters* **201**: 187–191.
Ragan MA, Harlow TJ, Beiko RG (2006) Do different surrogate methods detect lateral genetic transfer events of different relative ages? *Trends in Microbiology* **14**: 4–8.
Reynolds HT, Slot JC, Divon HH, Lysoe E, Proctor RH*, et al.* (2017) Differential retention of gene functions in a secondary metabolite cluster. *Molecular Biology and Evolution* **34**: 2002–2015.
Riley R, Salamov AA, Brown DW, Nagy LG, Floudas D*, et al.* (2014) Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proceedings of the National Academy of Sciences, USA* **111**: 9923–9928.
Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.
Sakaguchi M, Matsuda T, Matsumura E, Yoshimoto T, Takaoka M (2011) Prolyl oligopeptidase participates in cell cycle progression in a human neuroblastoma cell line. *Biochemical and Biophysical Research Communications* **409**: 693–698.
Sgambelluri RM, Epis S, Sassera D, Luo H, Angelos ER*, et al.* (2014) Profiling of amatoxins and phallotoxins in the genus *Lepiota* by liquid chromatography combined with UV absorbance and mass spectrometry. *Toxins* **6**: 2336–2347.
Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**: 1246–1247.
Slot JC (2017) Fungal gene cluster diversity and evolution. *Advances in Genetics* **100**: 141–178.
Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
Venalainen JI, Juvonen RO, Mannisto PT (2004) Evolutionary relationships of the prolyl oligopeptidase family enzymes. *European Journal of Biochemistry* **271**: 2705–2715.
Williams RS, Eames M, Ryves WJ, Viggars J, Harwood AJ (1999) Loss of a prolyl oligopeptidase confers resistance to lithium by elevation of inositol (1,4,5) trisphosphate. *EMBO Journal* **18**: 2734–2745.
Wisecaver JH, Rokas A (2015) Fungal metabolic gene clusters-caravans traveling across genomes and environments. *Frontiers in Microbiology* **6**: 161.
Wisecaver JH, Slot JC, Rokas A (2014) The evolution of fungal metabolic pathways. *PLoS Genetics* **10**: e1004816.
Yoshida K, Inaba K, Ohtake H, Morisawa M (1999) Purification and characterization of prolyl endopeptidase from the Pacific herring, *Clupea pallasi*, and its role in the activation of sperm motility. *Development Growth and Differentiation* **41**: 217–225.
Yoshimoto T, Kado K, Matsubara F, Koriyama N, Kaneto H*, et al.* (1987) Specific inhibitors for prolyl endopeptidase and their anti-amnesic effect. *Journal of Pharmacobio-Dynamics* **10**: 730–735.
Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.

Supplementary files can be found on the IMA Fungus website, http://www.imafungus.org/:

Supplementary File 1. *POPA* and *POPB* coding and aa sequences from the genomes of *Amanita subjunquillea*, *A. pallidorosea* and *A. phalloides*.
Supplementary File 2. Alignment of *POP* coding sequences from 75 fungi.
Supplementary File 3. Alignment of POP aa sequences from the same 75 fungi as in Supplementary File 2.
Supplementary File 4. Alignment of coding sequences from 21 MSDIN genes.
Supplementary File 5. Alignment of coding sequences from 18 *POP* genes used in the gene tree.
Supplementary File 6. Alignment of coding sequences from 30 MSDIN genes used in the species tree.