

Article

Not peer-reviewed version

Three-Dimensional Structural Heteromorphs of Mating-Type Proteins in *Hirsutella sinensis* and the Natural *Cordyceps sinensis* Insect-Fungal Complex

[Xiu-Zhang Li](#), [Yu-Ling Li](#), [Jia-Shi Zhu](#) *

Posted Date: 10 March 2025

doi: 10.20944/preprints202501.1307.v2

Keywords: Heteromorphic stereostructures of MAT1-1-1 and MAT1-2-1 proteins; Bayesian clustering; AlphaFold-predicted 3D protein structures; *Hirsutella sinensis*; reproduction of *Ophiocordyceps sinensis*; sexual life of the natural *Cordyceps sinensis* insect-fungal complex



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Three-Dimensional Structural Heteromorphs of Mating-Type Proteins in *Hirsutella sinensis* and the Natural *Cordyceps sinensis* Insect-Fungal Complex

Running title: Heteromorphic Mating Proteins in Natural *Cordyceps sinensis*

Xiu-Zhang Li, Yu-Ling Li and Jia-Shi Zhu *

State Key Laboratory of Plateau Ecology and Agriculture, Qinghai Academy of Animal and Veterinary Sciences, Qinghai University, Xining, Qinghai, China

* Correspondence: zhujosh@163.com

Abstract: The MAT1-1-1 and MAT1-2-1 proteins are essential for the sexual reproduction of *Ophiocordyceps sinensis*. Although *Hirsutella sinensis* has been postulated to be the sole-anamorph of *O. sinensis* and to undergo self-fertilization under homothallism or pseudohomothallism, little is known about the three-dimensional (3D) structures of the mating proteins in the natural *Cordyceps sinensis* insect-fungal complex, which is a valuable therapeutic agent in traditional Chinese medicine. However, the alternative splicing and differential occurrence and translation of the MAT1-1-1 and MAT1-2-1 genes have been revealed in *H. sinensis*, negating the self-fertilization hypothesis but rather suggesting the occurrence of self-sterility under heterothallic or hybrid outcrossing. In this study, the MAT1-1-1 and MAT1-2-1 proteins in 173 *H. sinensis* strains and wild-type *C. sinensis* isolates were clustered into 6 and 5 clades in the Bayesian clustering trees and belonged to 24 and 21 diverse AlphaFold-predicted 3D structural morphs, respectively. Over 3 quarters of the strains/isolates contained either MAT1-1-1 or MAT1-2-1 proteins but not both. The diversity of the heteromorphic 3D structures of the mating proteins suggested functional alterations of the proteins and provided additional evidence supporting the self-sterility hypothesis under heterothallism and hybridization for *H. sinensis*, Genotype #1 of the 17 genome-independent *O. sinensis* genotypes. The heteromorphic stereostructures and mutations of the MAT1-1-1 and MAT1-2-1 proteins in the wild-type *C. sinensis* isolates and natural *C. sinensis* insect-fungi complex suggest various sources of the mating proteins produced by two or more cooccurring heterospecific fungal species in natural *C. sinensis* that have been discovered in mycobiotic, molecular, metagenomic, and metatranscriptomic studies and may inspire future studies on the biochemistry of mating and pheromone-receptor proteins and the reproductive physiology of *O. sinensis*.

Keywords: heteromorphic stereostructures of MAT1-1-1 and MAT1-2-1 proteins; Bayesian clustering; AlphaFold-predicted 3D protein structures; *Hirsutella sinensis*; reproduction of *Ophiocordyceps sinensis*; sexual life of the natural *Cordyceps sinensis* insect-fungal complex

1. Introduction

The natural *Cordyceps sinensis* insect-fungal complex is one of most expensive therapeutic agents in traditional Chinese medicine and has a rich history of clinical applications for centuries in health maintenance, disease amelioration, post illness and postsurgery recovery, and antiaging therapy [Zhu *et al.* 1998a, 1998b, 2011]. As defined by the Chinese Pharmacopoeia, natural *C. sinensis* is an insect-fungal complex containing the *Ophiocordyceps sinensis* fruiting body and the remains of a *Hepialidae* moth larva (an intact, thick larval body wall with numerous bristles, an intact larval intestine and head tissues, and fragments of other larval tissues) [Ren *et al.* 2013; Zhang *et al.* 2014; Lu *et al.* 2016; Zhu & Li 2017; Li *et al.* 2022, 2023a]. Studies of natural *C. sinensis* have demonstrated its

multicellular heterokaryotic structures of hyphal and ascospore cells and genetic heterogeneity, including at least 17 genomically independent genotypes of *O. sinensis* fungi and >90 other fungal species spanning at least 37 fungal genera and larval genes [Jiang & Yao 2003; Zhang *et al.* 2010, 2018; Xia *et al.* 2015; Guo *et al.* 2017; Zhu & Li 2017; Li *et al.* 2016b, 2020, 2022, 2023a, 2023d; Zhong *et al.* 2018; Kang *et al.* 2024]. Among the numerous heterogeneous fungal species, *Hirsutella sinensis* was postulated by Wei *et al.* [2006] to be the sole-anamorph of *O. sinensis*; however, 10 years later, the key authors reported a species contradiction in an artificial cultivation project conducted in a product-oriented industrial setting between anamorphic inoculates of 3 GC-biased *H. sinensis* strains on *Hepialidae* moth larvae and the AT-biased sole-teleomorph (Genotype #4 of *O. sinensis*) in cultivated *C. sinensis* [Wei *et al.* 2016]. Notably, the Latin name *Cordyceps sinensis* has been used indiscriminately since the 1840s for both the teleomorph/holomorph of the fungus *C. sinensis* and the wild insect-fungal complex, and the fungus was renamed *Ophiocordyceps sinensis* in 2007 [Sung *et al.* 2007; Zhang *et al.* 2012; Ren *et al.* 2013; Zhu & Yao 2014; Zhu & Li 2017; Li *et al.* 2022]. Zhang *et al.* [2013] proposed improper implementation of the “One Fungus=One Name” nomenclature rule of the International Mycological Association [Hawksworth *et al.* 2011] while disregarding the presence of multiple genomically independent genotypes of *O. sinensis* fungi and inappropriately replacing the anamorphic name *H. sinensis* with the teleomorphic name *O. sinensis*. Thus, we continue using the anamorphic name *H. sinensis* for Genotype #1 of the 17 *O. sinensis* genotypes in this paper and refer to the genomically independent Genotypes #2–17 fungi as *O. sinensis* before their systematic positions are taxonomically determined, regardless of whether they are genetically GC- or AT-biased. We continue the customary use of the name *C. sinensis* to refer to the wild or cultivated insect-fungal complex because the renaming of *C. sinensis* to *O. sinensis* in 2007 did not involve the indiscriminate use of the Latin name for the natural insect-fungal complex, although this practice will likely be replaced in the future by the differential use of proprietary and exclusive Latin names for the multiple genome-independent *O. sinensis* genotypic fungi and the insect-fungi complex.

The sexual reproductive behavior of ascomycetes is controlled by transcription factors encoded at the mating-type (MAT) locus [Debuchy *et al.* 2006; Jones & Bennett 2011; Zheng & Wang 2013; Wilson *et al.* 2015]. Holliday *et al.* [2008], Stone *et al.* [2010], and Hu *et al.* [2013] reported failures to induce the development of *C. sinensis* fruiting bodies and ascospores *via* the use of pure *H. sinensis* cultures as inoculants. Zhang *et al.* [2013] summarized the failures over 40 years of academic experience in research-oriented academic settings. Hu *et al.* [2013] and Bushley *et al.* [2013] hypothesized that *H. sinensis* undergoes self-fertilization under homothallism or pseudohomothallism; however, Zhang *et al.* [2009, 2011] and Zhang and Zhang [2015] reported the differential occurrence of the MAT1-1-1 and MAT1-2-1 genes in numerous wild-type *C. sinensis* isolates and hypothesized that *O. sinensis* underwent facultative hybridization. Moreover, Li *et al.* [2023b, 2024b] reported the alternative splicing, differential occurrence, and differential transcription of mating-type and pheromone receptor genes in *H. sinensis* and natural *C. sinensis*, suggesting the occurrence of self-sterility in *H. sinensis* under heterothallism or hybridization and the demand of sexual partners during the sexual life of the natural *C. sinensis* insect-fungi complex.

Sequences of the MAT1-1-1 and MAT1-2-1 genes and proteins of *H. sinensis* are available in the GenBank database, but little is known about the polymorphic stereostructures of the proteins in *H. sinensis* strains and wild-type *C. sinensis* isolates, which are extremely crucial to the sexual reproduction of *O. sinensis* and for the maintenance of the natural ecological population volume of the Level II endangered authentic traditional Chinese medicinal “herb” [China Ministry of Agriculture and Rural Affairs 2021], which is a natural *C. sinensis* insect-fungi complex. In this work, we analyzed and correlated the statistical clustering of the primary structures and AlphaFold-predicted 3D structural models of the MAT1-1-1 and MAT1-2-1 proteins from 173 *H. sinensis* strains and wild-type *C. sinensis* isolates and correlated the heteromorphic structures of the protein sequences encoded by the genome, transcriptome and metatranscriptome assemblies of *H. sinensis* and natural *C. sinensis*.

2. Materials and Methods

2.1. *C. sinensis* Isolates and Accession Numbers of the MAT1-1-1 and MAT1-2-1 Proteins

The AlphaFold database lists the accession numbers of the MAT1-1-1 and MAT1-2-1 proteins and the 3D protein structures, which were derived from 173 *H. sinensis* strains and wild-type *C. sinensis* isolates that were collected from various production areas on the Qinghai-Tibet Plateau [Zhang *et al.* 2009, 2011; Hu *et al.* 2013; Zhang & Zhang 2015; Tunyasuvunakool *et al.* 2021].

2.2. Genome, Transcriptome, and Metatranscriptome Assemblies of *H. sinensis* Strains and the Natural *C. sinensis* Insect-Fungal Complex

The genome assemblies ANOV000000000, JAAVMX000000000, LKHE000000000, LWBQ000000000, and NGJJ000000000 of the *H. sinensis* strains Co18, IOZ07, 1229, ZJB12195, and CC1406-20395, respectively, were used for mating protein analysis [Hu *et al.* 2013; Li *et al.* 2016a; Jin *et al.* 2020; Liu *et al.* 2020; Shu *et al.* 2020].

The transcriptome assembly GCQL000000000 for the *H. sinensis* strain L0106 and the metatranscriptome assembly GAGW000000000 for the natural *C. sinensis* samples collected from Kangding County, Sichuan Province, China, were also used for mating protein analysis [Liu *et al.* 2015; Xiang *et al.* 2014].

Another metatranscriptome assembly was derived from mature natural *C. sinensis* samples collected from Deqin, Yunnan Province, China (*cf.* the Appendix of [Xia *et al.* 2017]). The metatranscriptome assembly sequences were uploaded to a repository database, www.plantkingdomgdb.com/Ophiocordyceps_sinensis/data/cds/Ophiocordyceps_sinensis_CDS.fasta, which is currently inaccessible, but a previously downloaded cDNA file (accessed from 5/18/2017 to 1/18/2018) was used for mating protein analysis.

2.3. Statistical Clustering Analysis for the MAT1-1-1 and MAT1-2-1 Protein Sequences

Multiple protein sequences of the *H. sinensis* strains and wild-type *C. sinensis* isolates were analyzed *via* the auto mode of MAFFT (v7.427). Bayesian clustering trees of the MAT1-1-1 and MAT1-2-1 protein sequences were then inferred *via* MrBayes v3.2.7 software (Markov chain Monte Carlo [MCMC] algorithm) with a sampling frequency of 100 iterations after discarding the initial 25% of the samples from a total of 1 million iterations [Huelsenbeck & Ronquist 2001; Ronquist *et al.* 2012; Li *et al.* 2022, 2023c, 2024b]. Clustering analysis was conducted at Nanjing Genepioneer Biotechnologies Co.

2.4. AlphaFold. -Based Prediction of 3D Structures of Mating Proteins

The 3D structures of the MAT1-1-1 and MAT1-2-1 proteins of the 173 *H. sinensis* strains and wild-type *C. sinensis* isolates were computationally predicted from their amino acid sequences *via* the artificial intelligence (AI)-based machine learning technology AlphaFold (<https://alphafold.com/>) and downloaded from the AlphaFold database (accessed from 10/18/2024 to 12/31/2024) for structural polymorphism analysis [Jumper *et al.* 2021; David *et al.* 2022; Abramson *et al.* 2024; Varadi *et al.* 2024]. The heteromorphic 3D structures of the MAT1-1-1 and MAT1-2-1 proteins were grouped based on the results of AlphaFold structural and Bayesian clustering analyses.

The AlphaFold database provides per-residue model confidence, the prediction of its score in the local distance difference test (pLDDT), between 0 and 100, a per-residue score that is assigned to each individual residue [Mariani *et al.* 2013; Jumper *et al.* 2021; David *et al.* 2022; Monzon *et al.* 2022; Xu *et al.* 2023; Abramson *et al.* 2024; Varadi *et al.* 2024]. Model confidence bands are used to color-code the residues in the 3D structure:  very high confidence (pLDDT > 90) residues are shown in dark blue,  high (90 > pLDDT > 70) in light blue,  low (70 > pLDDT > 50) in yellow, and  very low (pLDDT < 50) in orange [Mariani *et al.* 2013; Wroblewski & Kmiecik 2024]. Note that a protein region that is assigned a low pLDDT score does not necessarily indicate that this region is the most variable

region in the protein sequence; in contrast, a substantially variable region of a protein may be assigned a high pLDDT score. The AlphaFold database provides an average pLDDT score for each of the predicted 3D structure models of mating proteins, representing the overall model confidence in the predicted 3D structures.

2.5. Alignment Analysis of Protein Sequences

The amino acid sequences of the MAT1-1-1 and MAT1-2-1 proteins of *H. sinensis* and natural *C. sinensis* were aligned and compared via the GenBank Blastp program (<https://blast.ncbi.nlm.nih.gov/>).

2.6. Amino Acid Properties and Scale Analysis

The amino acid components of the mating proteins were scaled based on the general chemical characteristics of their side chains (cf. Supplementary Table S1) and plotted sequentially with a window size of 21 amino acid residues for the α -helices, β -sheets, β -turns, and coils of the MAT1-1-1 and MAT1-2-1 proteins via the linear weight variation model of the ExPASy ProtScale algorithm (<https://web.expasy.org/protscale/>) [Deleage & Roux 1987; Gasteiger *et al.* 2005; Peters & Elofsson 2014; Simm *et al.* 2016; Li *et al.* 2024b]. The plotting topologies and waveforms of the ProtScale plots for the proteins were compared to explore alterations in the 2D structures of the mating proteins.

3. Results

3.1. Diversity of the MAT1-1-1 and MAT1-2-1 Proteins in *H. sinensis* Strains and Wild-Type *C. sinensis* Isolates on the Basis of the AlphaFold-Predicted 3D Structures

A prior publication [Li *et al.* 2023b, 2024b] reported the differential occurrence, alternative splicing, and differential transcription of mating-type genes in *H. sinensis* and natural *C. sinensis*. The current paper focuses on the diverse stereostructures of the translation products, namely, MAT1-1-1 and MAT1-2-1 proteins of *O. sinensis*.

The AlphaFold database lists the accession numbers for 138 MAT1-1-1 proteins and 79 MAT1-2-1 proteins, which were derived from 173 *H. sinensis* strains and wild-type *C. sinensis* isolates [Zhang *et al.* 2009, 2011, 2013; Bushley *et al.* 2013; Hu *et al.* 2013; Zhang & Zhang 2015]. Among the 173 strains/isolates, 42 (24.3%) had records of AlphaFold-predicted 3D structures for both the MAT1-1-1 and MAT1-2-1 proteins. A majority (75.7%) of the strains/isolates presented 3D structure records for either the MAT1-1-1 or MAT1-2-1 protein, suggesting differential cooccurrences of the 2 mating proteins essential for the sexual reproduction of *O. sinensis*. In addition, strains CS68-2-1229 and CS2 have duplicated accession numbers for either the MAT1-1-1 or MAT1-2-1 protein, unlike the 171 other strains/isolates (*i.e.*, $173=138+79-42-2$).

Strain CS68-2-1229 has 2 accession numbers for the MAT1-1-1 protein, namely, AGW27560 and AGW27528, which share 100% sequence identity. However, AGW27560 is a full-length protein containing 372 amino acids, whereas AGW27528 is an N- and C-terminally truncated protein containing 301 amino acids with 80.9% query coverage.

Strain CS2 has 2 accession numbers for the full-length MAT1-2-1 protein, namely, AEH27625 and ACV60400, which contain 249 amino acids and share 100% sequence identity; however, the sequences were released by GenBank 6 years apart, on 03-JUN-2010 and 25-JUL-2016, respectively.

The 138 MAT1-1-1 proteins belong to diverse 3D structural models or morphs under 24 UniProt codes in the AlphaFold database, 118 of which are full-length proteins belonging to 3D structural morphs under 15 AlphaFold UniProt codes and are listed in Table 1. Among the 118 full-length proteins, 89 (75.4%) are under the UniProt code U3N942 and are considered “likely authentic” proteins. The remaining 20 of the 138 MAT1-1-1 proteins are truncated and belong to 3D structural morphs under 9 other UniProt codes.

Table 1. GenBank accession numbers (in green, in parentheses) for the full-length MAT1-1-1 proteins of the *H. sinensis* strains and wild-type *C. sinensis* isolates under the corresponding AlphaFold UniProt codes.

AlphaFold UniProt code (Bayesian cluster/branch*)	Strain/isolate number (GenBank accession number)
U3N942 (A1)	CS68-2-1229 (AGW27560)(AGW27528),GS09_111 (ALH24945),GS09_131 (ALH24947), ID10_1 (ALH24954), IOZ07 (KAF4512729), NP10_1 (ALH24955), NP10_2 (ALH24956), QH07_188 (ALH24957), QH07_197 (ALH24958), QH09_37 (ALH24968), QH09_46 (ALH24969), QH09_56 (ALH24970), QH09_66 (ALH24971), QH09_78 (ALH24972), QH09_93 (ALH24973), QH09_122 (ALH24959), QH09_131 (ALH24960), QH09_151 (ALH24961), QH09_20L (ALH24965), QH09_33L (ALH24967), QH10_1 (ALH24974), QH10_4 (ALH24975), QH10_7 (ALH24976), SC09_21 (ALH24987), SC09_36 (ALH24988), SC09_37 (ALH24989), SC09_47 (ALH24990), SC09_57 (ALH24991), SC09_77 (ALH24993), SC09_107 (ALH24978), SC09_117 (ALH24979), SC09_128 (ALH24980), SC09_147 (ALH24981), SC09_157 (ALH24982), SC09_167 (ALH24983), SC09_180 (ALH24984), SC09_190 (ALH24985), SC09_200 (ALH24986), SC10_18 (ALH24996), SC10_21 (ALH24997), SC10_4 (ALH24998), XZ05_3 (ALH25002), XZ05_7 (ALH25004), XZ05_12 (ALH25000), XZ06_124 (ALH25006), XZ06_152 (ALH25007), XZ07_108 (ALH25009), XZ07_133 (ALH25010), XZ07_154 (ALH25011), XZ07_166 (ALH25012), XZ07_176 (ALH25013), XZ07_180 (ALH25014), XZ08_4 (ALH25018), XZ08_10 (ALH25015), XZ08_24 (ALH25016), XZ08_26 (ALH25017), XZ08_56 (ALH25019), XZ08_59 (ALH25020), XZ08_A1 (ALH25021), XZ08_B1 (ALH25022), XZ09_4 (ALH25029), XZ09_46 (ALH25030), XZ09_48 (ALH25031), XZ09_59 (ALH25032), XZ09_71 (ALH25033), XZ09_80 (ALH25055), XZ09_106 (ALH25024), XZ09_113 (ALH25025), XZ09_118 (ALH25026), XZ09_15 (ALH25027), XZ09_32 (ALH25028), XZ10_7 (ALH25038), XZ10_15 (ALH25035), XZ10_17 (ALH25036), XZ10_23 (ALH25037), XZ12_1 (ALH25056), XZ12_33 (ALH25058), XZ12_43 (ALH25059), YN07_6 (ALH25039), YN07_8 (ALH25040), YN09_3 (ALH25044), YN09_72 (ALH25049), YN09_81 (ALH25050), YN09_85 (ALH25051), YN09_89 (ALH25052), YN09_96 (ALH25053),YN09_101 (ALH25041), YN09_140 (ALH25042)
A0A0N9QMM1 (A1)	GS09_121 (ALH24946), GS09_201 (ALH24949), GS09_225 (ALH24950), SC09_1 (ALH24977)
T5A511 (A1)	Co18 (EQK97643) (KE657544 410←1519 and ANOV01017390 410←1519)
A0A0N9R5B3 (A2)	SC09_65 (ALH24992)
A0A0N7G849 (A2)	SC09_97 (ALH24995)
A0A0N9QUF3 (A3)	GS09_143 (ALH24948)

A0A0N9R4V2 (A3)	YN09_61 (ALH25047)
A0A0N9QMS9 (B)	YN09_6 (ALH25046), YN09_22 (ALH25043), YN09_51 (ALH25045), YN09_64 (ALH25048)
A0A0N7G845 (C)	GS09_229 (ALH24951), GS09_281 (ALH24952), GS09_311 (ALH25054), GS10_1 (ALH24953), QH09_164 (ALH24962), QH09_173 (ALH24963), QH09_201 (ALH24964), QH09_210 (ALH24966), SC09_87 (ALH24994)
A0A0N9QUK2 (D1)	XZ05_8 (ALH25005)
A0A0N9QMT4 (D2)	XZ07_H2 (ALH24999), XZ12_16 (ALH25057)
A0A0N9QMR3 (E1)	XZ06_260 (ALH25008), XZ09_100 (ALH25023)
A0A0N9QMS4 (E2)	XZ09_95 (ALH25034)
A0A0N7G850 (E3)	XZ05_6 (ALH25003)
A0A0N9R4Q4 (E4)	XZ05_2 (ALH25001)

Note: The names of the pure *H. sinensis* strains are highlighted in bold and underlined, whereas those of the wild-type *C. sinensis* isolates are not. *, Branch 1 is shown in red, Branch 2 in pink, Branch 3 in purple, and Branch 4 in brown with the cluster codes (English letters) in parentheses determined *via* the Bayesian analysis shown in Figure 1 below. The “←” arrows indicate sequences in the antisense strands of the genome of the *H. sinensis* strain Co18.

The 79 MAT1-2-1 proteins belong to diverse 3D structural morphs under 21 UniProt codes in the AlphaFold database, 74 of which are full-length proteins containing 249 amino acids belonging to 3D structural morphs under 17 AlphaFold UniProt codes and are listed in Table 2. Among the 74 full-length MAT1-2-1 proteins, 38 (51.4%) are under the UniProt code D7F2E9 and are considered “likely authentic” proteins. The remaining 5 of the 79 MAT1-2-1 proteins are truncated and belong to 3D structural morphs under 4 other UniProt codes.

Table 2. GenBank accession numbers (in green, in parentheses) for the full-length MAT1-2-1 proteins of the *H. sinensis* strains or wild-type *C. sinensis* isolates under the corresponding AlphaFold UniProt codes.

AlphaFold UniProt code (Bayesian cluster/branch**)	Strain/isolate number (GenBank accession number)
D7F2E9 (I-1)	CS2 (AEH27625)(ACV60400), CS26-277 (AGW27541), CS36-1294 (AGW27538), CS37-295 (AGW27539), SC-2 (ACV60395), SC-4 (ACV60396), SC-5 (ACV60398), SC-7 (ACV60397), SC09-37 (AFH35019), SC09_47 (AFX66423), SC09_57 (AFX66424), SC09_77 (AFX66426), SC09_97 (AFX66428), XZ05_7 (AFX66442), XZ05_12 (AFX66444), XZ06_152 (AFX66445),XZ07_11 (AFX66447), XZ07_46 (AFX66448), XZ09_106 (AFX66464), XZ09_15 (AFX66455), YN09_101 (AFX66482), YN09_72 (AFX66477), XZ09_113 (AFX66465), XZ-LZ06-1 (ACV60369), XZ-LZ06-7 (ACV60370), XZ-LZ06-21 (ACV60371), XZ-LZ06-108 (ACV60373), XZ-LZ07-30 (ACV60377), XZ-LZ07-108 (ACV60379), XZ-ML-191 (ACV60376), YN-1 (ACV60390), YN-5 (ACV60392), YN-6 (ACV60393),

	YN-8 (ACV60394), YN09_81 (AFX66478), YN09_85 (AFX66479), YN09_89 (AFX66480)
T5AF56 (I-1)	Co18 (EQL04085) (ANOV01000063 9329→10182)
V9LW10 (I-2)	SC09_200 (AFX66437)
D7F2H1 (I-2)	YN-4 (ACV60391)
D7F2F2 (I-2)	XZ-LZ06-61 (ACV60372)
A0A0A0RCF5 (II-1)	XZ12_16 (AIV43040)
D7F2J7 (II-2)	XZ05_8 (AFX66443), XZ06-124 (AFH35020), XZ-LZ07-H2 (ACV60418), XZ-LZ07-H1 (ACV60417)
D7F2F5 (III)	XZ05_2 (AFX66441), XZ06_260 (AFX66446), XZ09_80 (AFX66461), XZ09_95 (AFX66462), XZ09_100 (AFX66463), XZ-LZ05-6 (ACV60415), XZ-SN-44 (ACV60375),
V9LWC9 (IV-1)	YN09_64 (AFX66476)
V9LVS8 (IV-2)	YN09_6 (AFX66472), YN09_22 (AFX66473), YN09_51 (AFX66474)
D7F2E3 (V-1)	GS09_111 (AFX66388), CS560-961 (AGW275424), QH09-93 (AFH35018), XZ-NQ-154 (ACV60363), XZ-NQ-155 (ACV6036)
D7F2G5 (V-2)	QH-YS-199 (ACV60385)
D7F2H9 (V-2)	SC-3 (ACV60399)
V9LW71 (V-2)	QH09_11 (AFX66401)
V9LVU8 (V-2)	YN09_61 (AFX66475)
V9LWG5 (V-2)	ID10_1 (AFX66484)
U3N6V5 (V-2)	CS6-251 (AGW27537)
‡	NP10_1 (AFX66485), NP10_2 (AFX66486), YN09_3 (AFX66471), YN09_96 (AFX66481), YN09_140 (AFX66483)

Note: The names of the pure *H. sinensis* strains are highlighted in bold and underlined, whereas those of the wild-type *C. sinensis* isolates are not. **, Branch 1 is shown in red and Branch 2 in pink with the cluster codes (Roman numerals) in the parentheses determined *via* the Bayesian analysis shown in Figure 2 below. ‡, The 5 MAT1-2-1 protein sequences are included in the GenBank database but not in the AlphaFold database (*cf.* Supplementary Figure S1). The “→” arrow indicates the sequence in the sense strand of the genome of the *H. sinensis* strain Co18.

3.2. Bayesian Analysis of the MAT1-1-1 and MAT1-2-1 Proteins

Figure 1 shows the Bayesian clustering tree for 40 protein sequences covering the diverse structural morphs of MAT1-1-1 proteins under 24 UniProt codes. The sequences ALH24945, ALH24947, and AGW27560 represent a group of 89 sequences under UniProt code U3N942, which were clustered into Branch A1 of Cluster A, as shown in red alongside the tree in Figure 1. Branch A1 in Figure 1 also includes the full-length MAT1-1-1 proteins under UniProt codes A0A0N9QMM1 and T5A511. Cluster A includes other full-length MAT1-1-1 protein sequences with very similar 3D structures, which were clustered into Branch A2 in pink and Branch A3 in purple alongside the tree. The full-length MAT1-1-1 protein sequences with significantly variable 3D structures were clustered within Clusters B–E in Figure 1, either branched or unbranched, under various UniProt codes in red for Branch 1, in pink for Branch 2, in purple for Branch 3, or in brown for Branch 4.

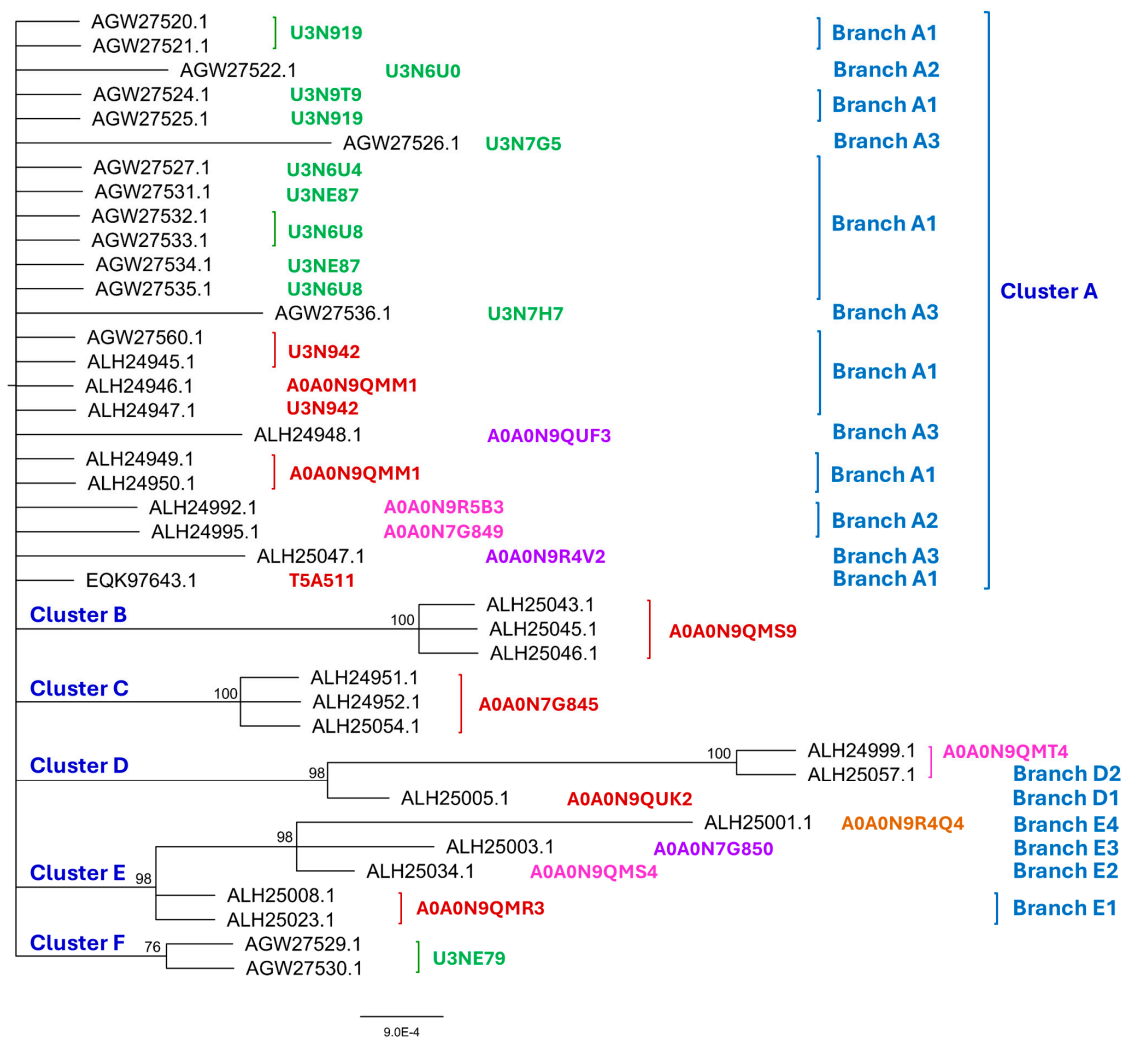


Figure 1. The Bayesian majority rule consensus clustering tree was inferred *via* MrBayes v3.2.7 software for the 40 full-length and truncated MAT1-1-1 proteins of the *H. sinensis* strains and wild-type *C. sinensis* isolates. The clusters and their branches (in blue) are shown alongside the tree. The AlphaFold UniProt codes for the 3D structures of the full-length proteins are shown in red alongside the tree for Branch 1 of the clusters, in pink for Branch 2, in purple for Branch 3, and in brown for Branch 4. The AlphaFold UniProt codes in green indicate the N-/C-terminally truncated MAT1-1-1 proteins.

Many truncated MAT1-1-1 proteins were found under UniProt codes U3N919, U3N6U0, U3N9T9, U3N7G5, U3N6U4, U3NE87, U3N6U8, and U3N7H7, which are shown in green alongside the tree in Figure 1 and were clustered into Branches A1–A3 of Cluster A. In addition, Cluster F contains the truncated MAT1-1-1 proteins under the UniProt code U3NE79 in green alongside the tree in Figure 1.

The 79 MAT1-2-1 proteins have various 3D structural morphs under 21 UniProt codes in the AlphaFold database (*cf.* Table 2), among which 32 representative sequences were subjected to Bayesian clustering analysis, as shown in Figure 2.

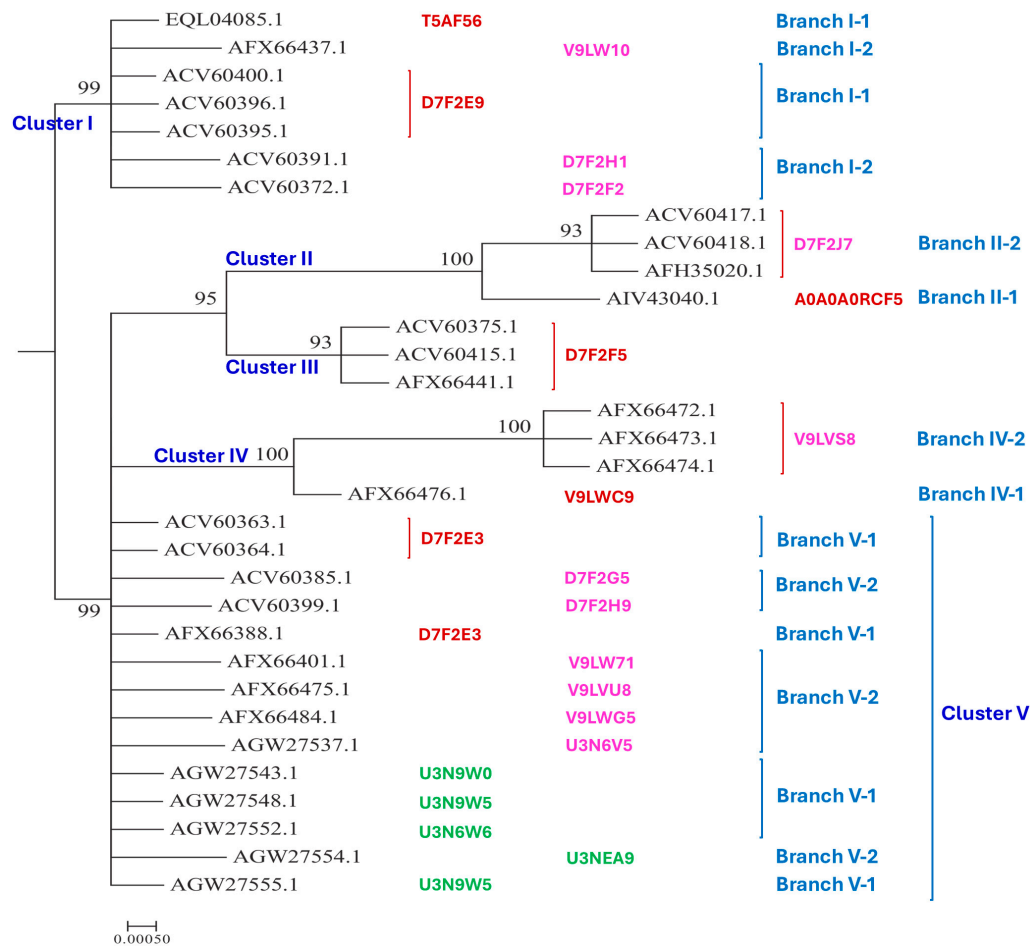


Figure 2. The Bayesian majority rule consensus clustering tree was inferred *via* MrBayes v3.2.7 software for the 32 full-length and truncated MAT1-2-1 proteins of the *H. sinensis* strains and wild-type *C. sinensis* isolates. The clusters and their branches (in blue) are shown alongside the tree. The AlphaFold UniProt codes for the 3D structures of the full-length proteins are shown in red alongside the tree for Branch 1 of the clusters and in pink for Branch 2 of the clusters. The AlphaFold UniProt codes in green indicate the C-terminally truncated MAT1-2-1 proteins.

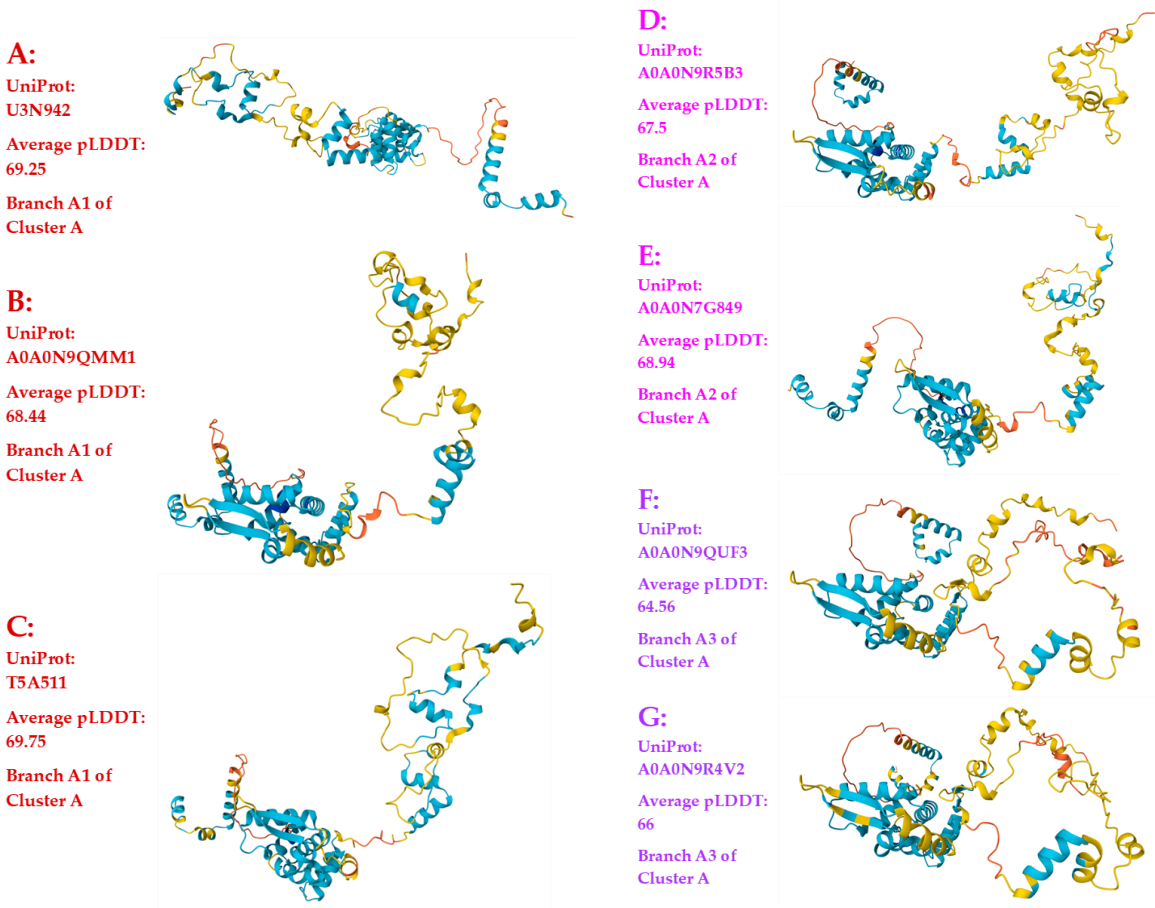
Among a total of 79 MAT1-2-1 proteins, 74 are full-length proteins, containing 249 amino acids and belonging to diverse 3D structural morphs under 17 AlphaFold UniProt codes. The remaining 5 MAT1-2-1 proteins are truncated and belong to 3D structural morphs under 4 other UniProt codes. Among the 74 full-length MAT1-2-1 proteins, 39 (52.7%) proteins under the UniProt codes D7F2E9 and T5AF56 were clustered into Branch I-1 of Cluster I of the Bayesian tree shown in Figure 2 and Table 2. Branch I-2 of Cluster I includes 3 MAT1-2-1 protein sequences with very similar 3D structures belonging to 3D structural morphs under the UniProt codes V9LW10, D7F2H1, and D7F2F2, as shown in Figure 2. Twenty-seven other full-length MAT1-2-1 proteins with significantly variable 3D structures were within Clusters II–V under various UniProt codes. Branches V-1 and V-2 of Cluster V also include 5 truncated MAT1-2-1 proteins under different UniProt codes, which are shown in green alongside the tree.

The GenBank database contains 5 other MAT1-2-1 protein sequences, namely, AFX66471, AFX66481, AFX66483, AFX66485, and AFX66486, which were derived from wild-type *C. sinensis* isolates YN09_3, YN09_96, YN09_140, NP10_1, and NP10_2, respectively, with predicted 3D structure records in the AlphaFold database for the MAT1-1-1 proteins but not for the MAT1-2-1 proteins (*cf.* Tables 1–2). The 5 MAT1-2-1 protein sequences are 100% identical to the reference sequence ACV60363 of Branch V-1 in the Bayesian clustering tree (*cf.* Figure 2 and Supplementary Figure S1), indicating that the 5 protein sequences likely belong to Branch V-1 of Cluster V, together

with 5 other Branch V-1 proteins, including the reference protein ACV60363, as listed in Table 1 and Figure 2.

3.3. Heteromorphic AlphaFold-Predicted 3D Structures of the MAT1-1-1 Proteins

Figure 3 shows the AlphaFold-predicted 3D structures of the 118 full-length MAT1-1-1 proteins under 15 structural morphs (Panels A–O), which are also listed in Table 1. Among the 118 full-length proteins, 89 (75.4%) are under the UniProt code U3N942 predicted by AlphaFold technology (Panel A of Figure 3). This 3D structure model most likely represents the authentic protein structure with full mating functionality.



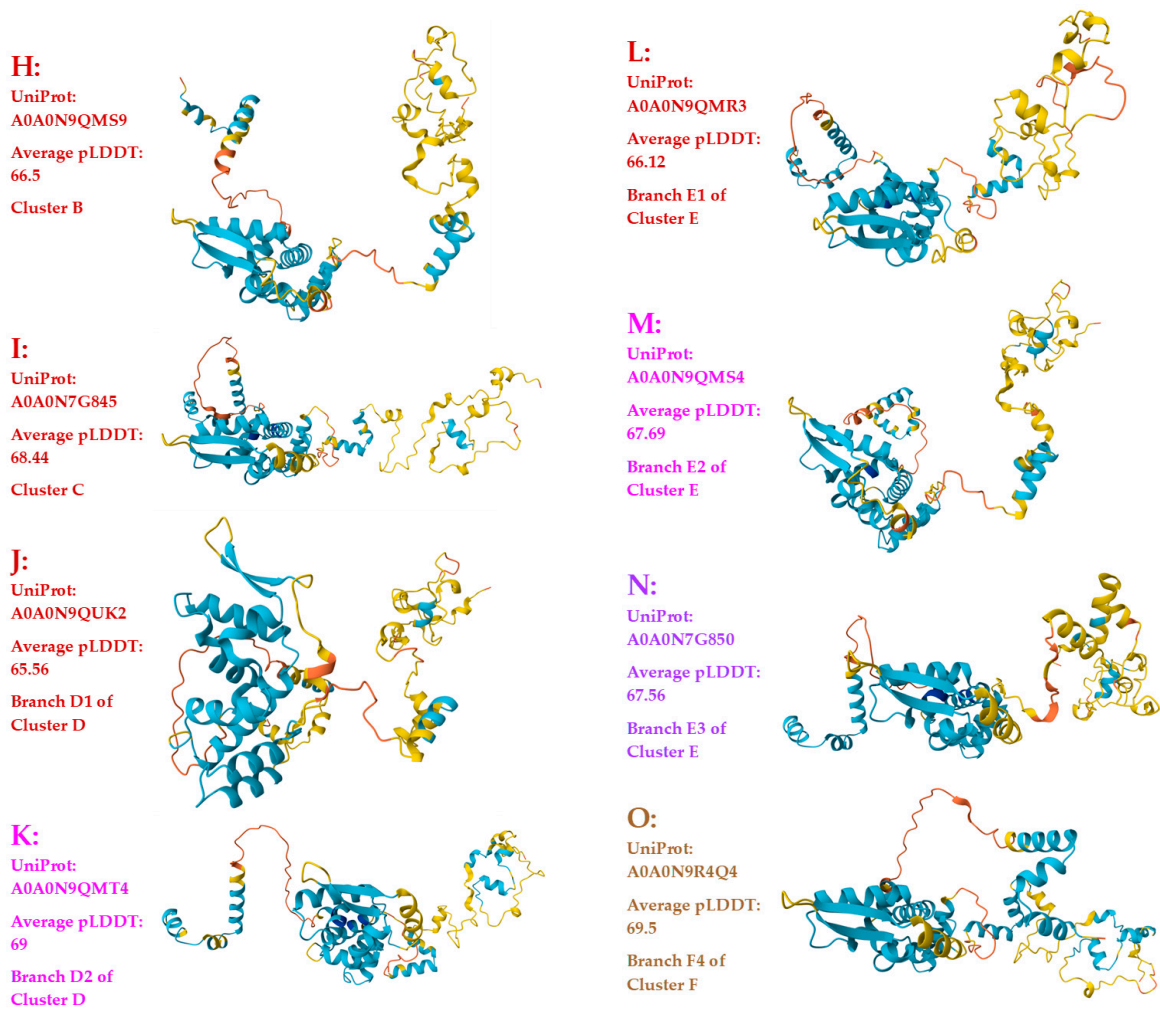
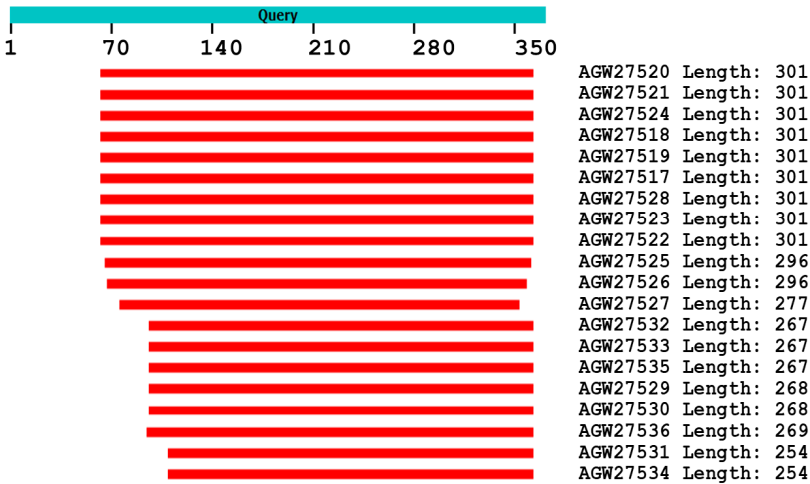


Figure 3. Fifteen 3D structural morphs for the 118 full-length MAT1-1-1 proteins of *H. sinensis* strains and wild-type *C. sinensis* isolates. The UniProt codes in red are for Branch 1 of each cluster shown alongside the Bayesian tree in Figure 1, those in pink are for Branch 2, those in purple are for Branch 3, and those in brown are for Branch 4. An average pLDDT score for each predicted 3D structural model was computed *via* AlphaFold technology and is shown in each of the 3D structural panels A–O, indicating model confidence (*cf.* Section 2.4): ■ very high (pLDDT > 90); ■ high (90 > pLDDT > 70); ■ low (70 > pLDDT > 50); and ■ very low (pLDDT < 50).

As shown in Table 1 and Figures 1 and 3, 94 (79.9%) of the 118 full-length MAT1-1-1 proteins are under the UniProt codes U3N942, A0A0N9QMM1, and T5A511 and clustered into Branch A1 of Cluster A in the Bayesian tree belonging to the 3D structure morphs A–C. The 94 full-length MAT1-1-1 proteins are most frequently detected and are likely authentic with full mating functionality.

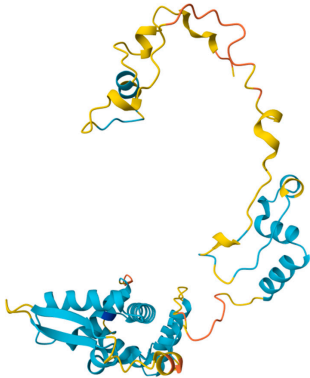
Figure 4 shows the sequence distributions (Panel A) and predicted 3D structures of 20 other MAT1-1-1 proteins, which are truncated at the N- and C-termini; these structures constitute the remaining 9 diverse morphs of 3D structures (Panels B–J).

A. Distribution of 20 N-/C-terminal truncated MAT1-1-1 protein



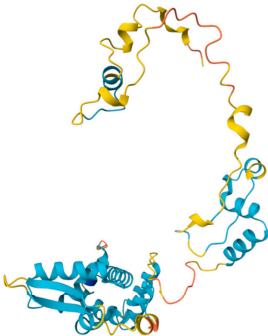
B:
UniProt: U3N9T9
Branch A1 of Cluster A

Strain	Accession
CS18_266	AGW27520
CS26_277	AGW27521
CS34_291	AGW27524
CS36_1294	AGW27518
CS_37-295	AGW27519
CS6-251	AGW27517
CS68-2-1229	AGW27528
CS91-1291	AGW27523



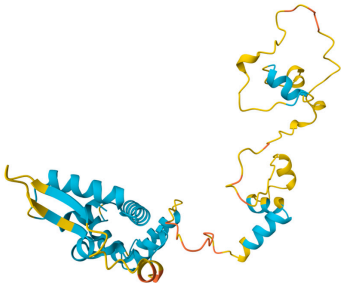
C:
UniProt: U3N919
Branch A1 of Cluster A

Strain	Accession
CS76-1284	AGW27525



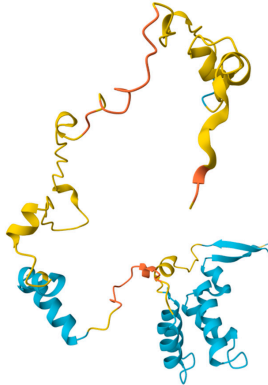
D:
UniProt: U3N6U4
Branch A1 of Cluster A

Strain	Accession
CS25-273	AGW27527



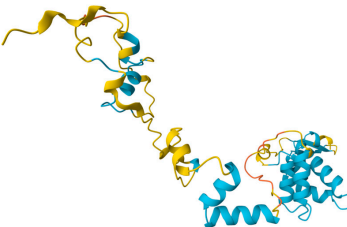
E:
UniProt: U3NE87
Branch A1 of Cluster A

Strain	Accession
CS70-1212	AGW27531
CS71-1219	AGW27534



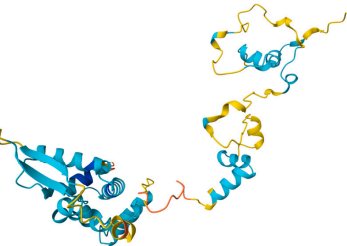
F:
UniProt: U3N6U8
Branch A1 of Cluster A

Strain	Accession
CS68-5-1216	AGW27532
CS71-1218	AGW27533
CS71-1220	AGW27535



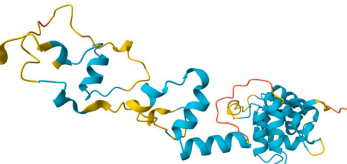
G:
UniProt: U3N6U0
Branch A2 of Cluster A

Strain	Accession
CS560-961	AGW27522



H:
UniProt: U3N7G5
Branch A3 of Cluster A

Strain	Accession
CS561-964	AGW27526



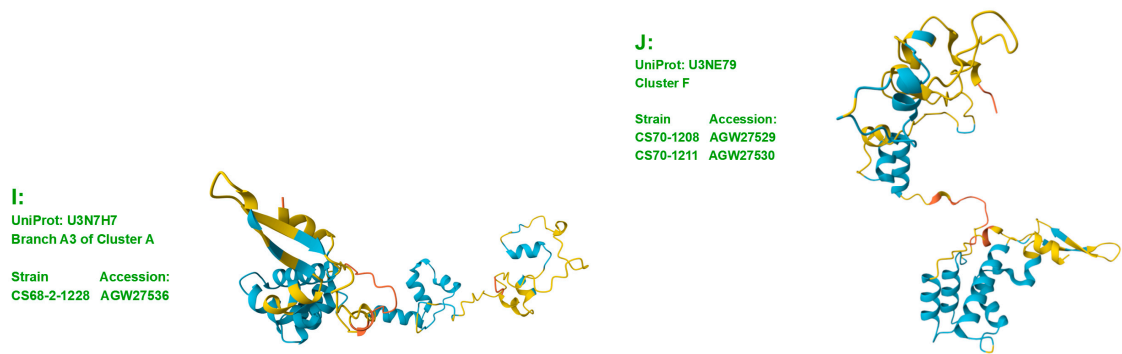
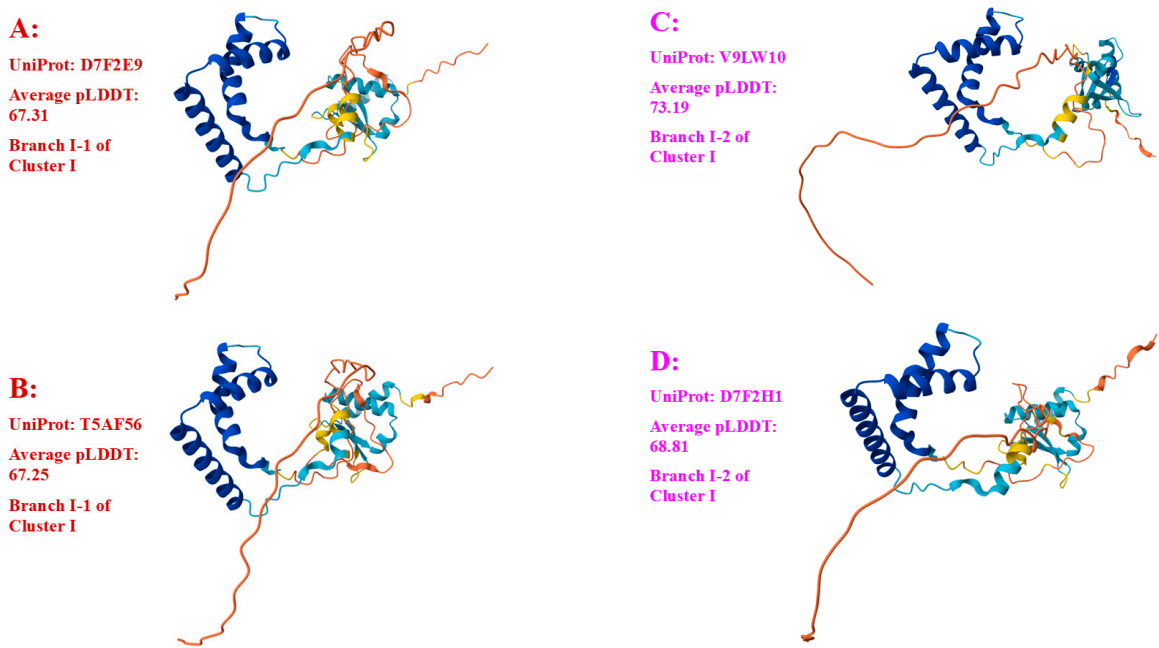


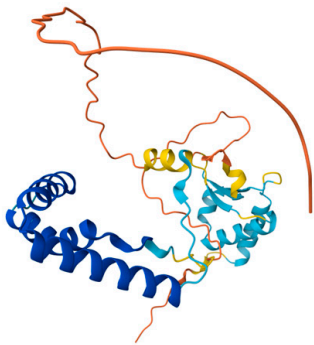
Figure 4. The sequence distribution (Panel A) and 9 diverse 3D structure morphs (Panels B–J) of the 20 N-/C-terminally truncated MAT1-1-1 proteins of *H. sinensis* strains and wild-type *C. sinensis* isolates. Model confidence: ■ very high (pLDDT > 90); ■ high (90 > pLDDT > 70); ■ low (70 > pLDDT > 50); and ■ very low (pLDDT < 50).

3.4. Heteromorphic AlphaFold-Predicted 3D Structures of the MAT1-2-1 Proteins

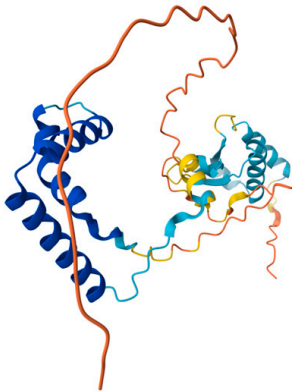
Among the 79 MAT1-2-1 protein sequences belonging to the 21 diverse 3D structural morphs, 74 are full-length proteins, 69 of which belong to diverse structural morphs under 17 UniProt codes and are shown in Panels A–Q of Figure 5.



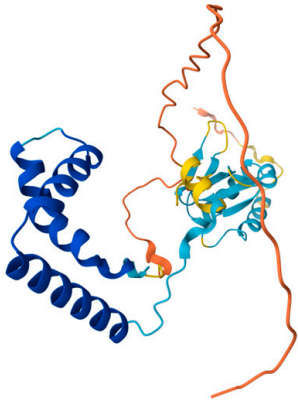
E:
UniProt: D7F2F2
Average pLDDT: 66.44
Branch I-2 of Cluster I



I:
UniProt: V9LWC9
Average pLDDT: 68.56
Branch IV-1 of Cluster IV



F:
UniProt: A0A0A0RCF5
Average pLDDT: 71.44
Branch II-1 of Cluster II



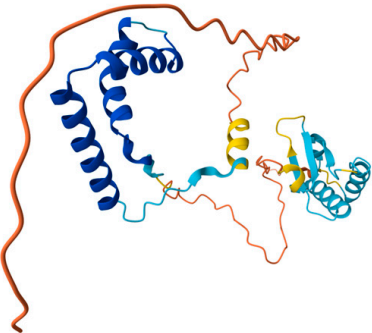
J:
UniProt: V9LVS8
Average pLDDT: 74.44
Branch IV-2 of Cluster IV



G:
UniProt: D7F2J7
Average pLDDT: 66.31
Branch II-2 of Cluster II



K:
UniProt: D7F2E3
Average pLDDT: 68.56
Branch V-1 of Cluster V



H:
UniProt: D7F2F5
Average pLDDT: 72.44
Cluster III



L:
UniProt: D7F2G5
Average pLDDT: 73.75
Branch V-2 of Cluster V



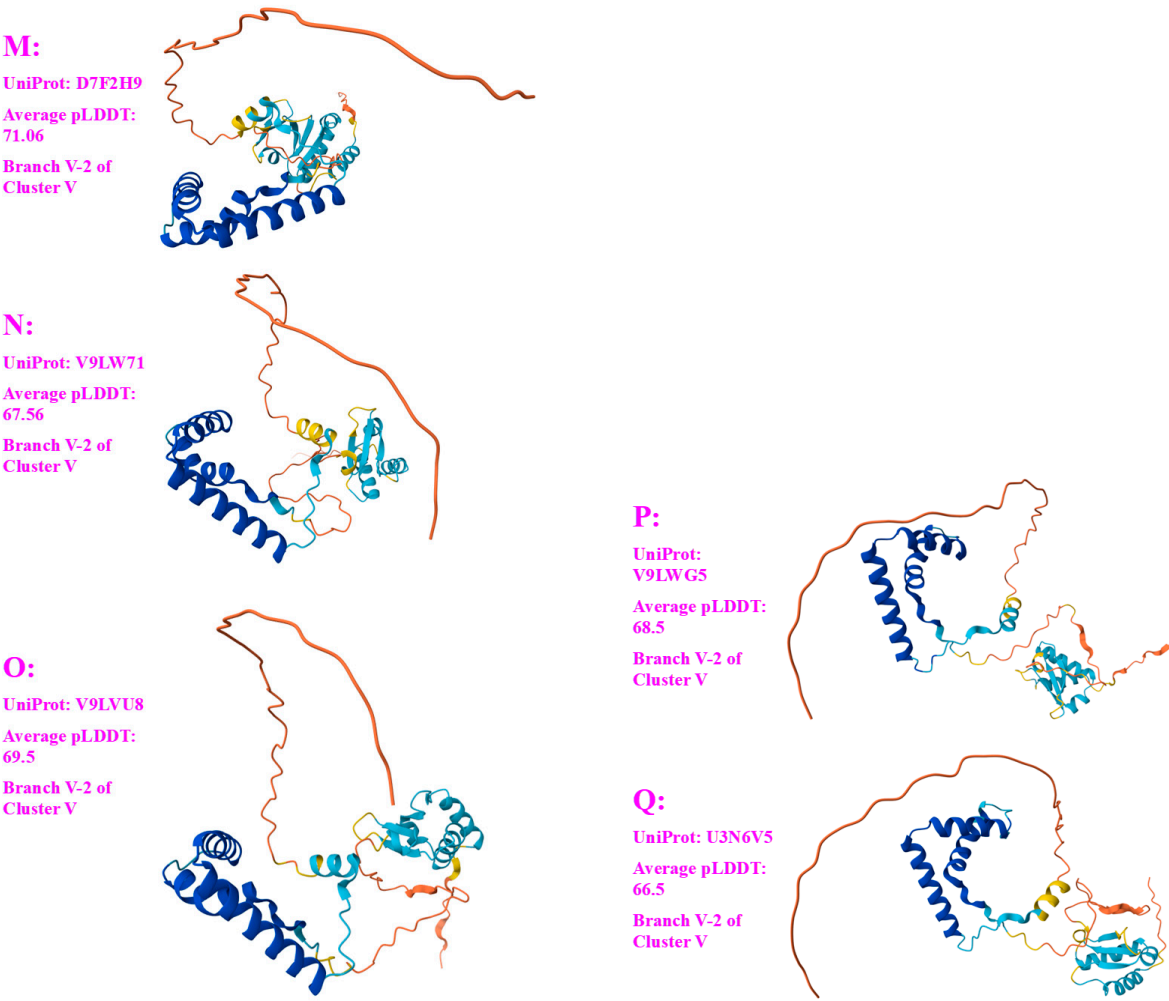


Figure 5. Seventeen 3D structural morphs for the full-length MAT1-2-1 proteins of *H. sinensis* strains and wild-type *C. sinensis* isolates. The UniProt codes in red are for Branch 1 of each cluster shown in the Bayesian tree in Figure 2, and those in pink are for Branch 2. Model confidence: ■ very high (pLDDT > 90); ■ high (90 > pLDDT > 70); ■ low (70 > pLDDT > 50); and ■ very low (pLDDT < 50).

As shown in Table 2 and Figures 2 and 5, 39 (52.7%) of the 74 full-length MAT1-2-1 proteins are under the UniProt codes D7F2E9 and T5AF56 and clustered into Branch I-1 of Cluster I in the Bayesian tree belonging to 3D structural morph A of the MAT1-2-1 proteins. The 39 full-length MAT1-2-1 proteins are frequently detected and are likely authentic with full mating functionality.

As shown in Supplementary Figure S1, the 5 other full-length MAT1-2-1 protein sequences (AFX66471, AFX66481, AFX66483, AFX66485, and AFX66486) exhibit 100% sequence identity with the Branch V-1 reference protein ACV60363. Thus, the 5 protein sequences without AlphaFold-predicted 3D structure records likely belong to the 3D structural morph K, together with the Branch V-1 reference protein ACV60363 (*cf.* Figures 2 and 5).

Figure 6 shows the sequence distribution (Panel A) and the AlphaFold-predicted 3D structures of the C-terminally truncated MAT1-2-1 proteins, which constitute the remaining 4 diverse morphs of 3D structures (Panels B–E).

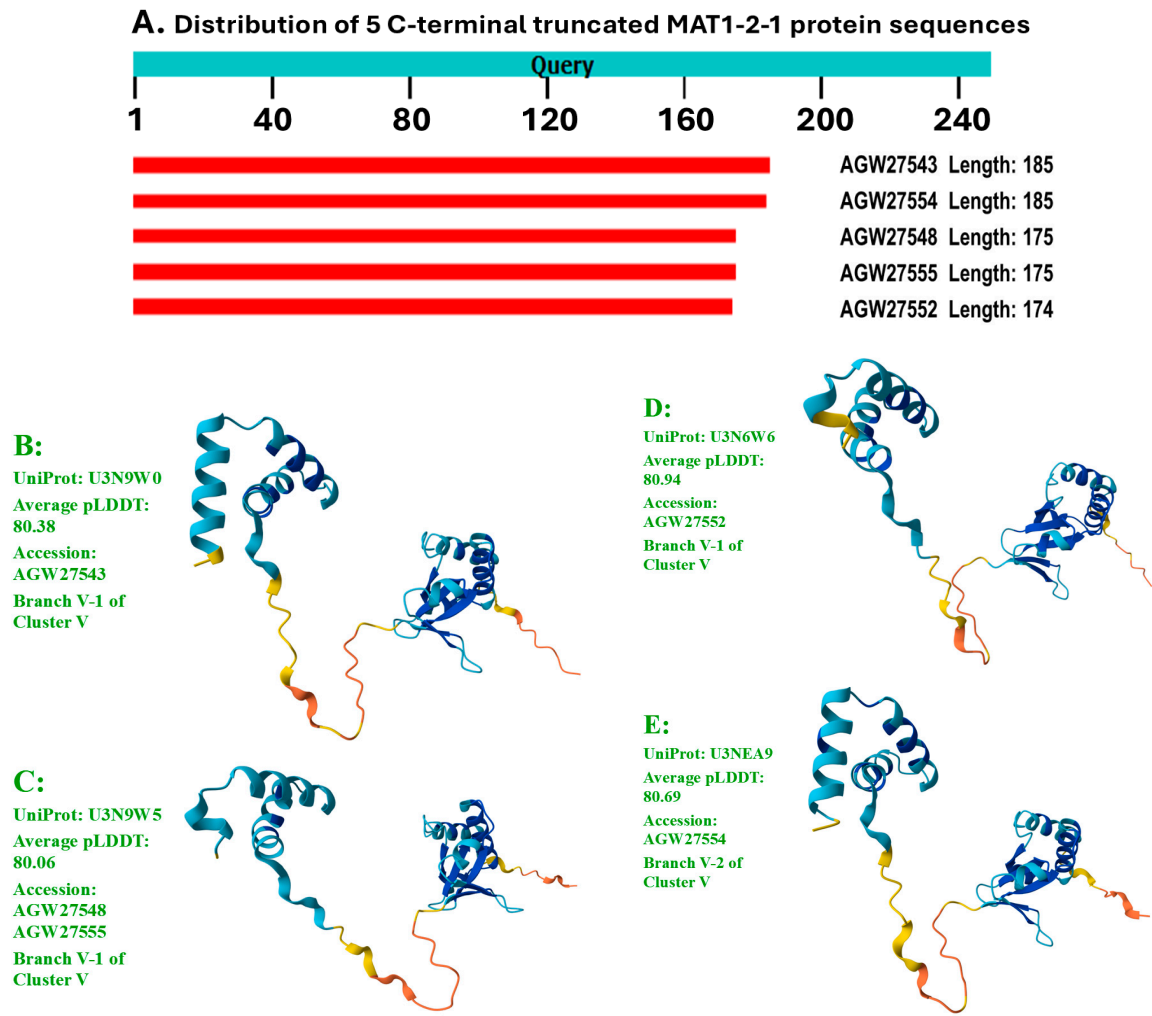


Figure 6. The sequence distribution (Panel A) and 3D structures (Panels B–E) of 5 C-terminally truncated MAT1-2-1 proteins of *H. sinensis* strains and wild-type *C. sinensis* isolates belonging to 4 diverse 3D structural morphs. Model confidence: very high (pLDDT > 90); high (90 > pLDDT > 70); low (70 > pLDDT > 50); and very low (pLDDT < 50).

3.5. Primary Structures of the MAT1-1-1 Proteins

Because of the diversity of the 3D structures of the MAT1-1-1 proteins, the variations in their primary amino acid sequences were then analyzed. The 118 full-length MAT1-1-1 proteins (*cf.* Table 1) consisted of 372 amino acids and contributed to 15 diverse 3D structural morphs (*cf.* Figure 3). Among the 118 full-length proteins, 89 shared 100% sequence identity with the query protein sequence (AGW27560), whereas 20 other proteins shared 98.1–99.6% sequence similarity with the query sequence, as they contained various conservative and nonconservative substitutions of amino acid residues at isolated sites, which may have an impact on mating function. Figure 7 shows the alignment of the full-length MAT1-1-1 protein sequences covering 5 Bayesian clusters A–E (Branches A1, A2, A3, B, C, D1, D2, E1, E2, E3, and E4) shown in Figure 1 and Table 1 and 15 AlphaFold 3D structural morphs A–O (*cf.* Figure 3), as well as the MAT1-1-1 protein sequences encoded by the genome assemblies of *H. sinensis* and the metatranscriptome assemblies of the natural *C. sinensis* insect-fungal complex.

AGW27560	1	MTTRNEVMORLSSVRADVLLNFLTDDAIFQLASRYHESTTEADVLTVPVSTAAASRATROT	60
ALH24946	1		60
ALH24995	1	S	60
ALH24992	1		60
ALH25047	1		60
ALH24948	1		60
ALH25043	1		60
ALH25054	1		60
ALH25005	1		60
ALH24999	1		60
ALH25008	1	T I	60
ALH25034	1		60
ALH25003	1		60
ALH25001	1		60
EQK97643	1	T I	60
ANOV01017390	1519		1340
LKHE01001116	4909		4730
JAAVMX010000001	6698911		6699090
GAGW01008880			
OSIN7648	1		181
AGW27560	61	KEASCDRAKRPLNAFMAFRSYYLKLFDPVQOKTASGFLTTLWHKDPFRNKWALIAKVYSE	120
ALH24946	61		120
ALH24995	61		120
ALH24992	61		120
ALH25047	61		120
ALH24948	61		120
ALH25043	61		120
ALH25054	61		120
ALH25005	61		120
ALH24999	61		120
ALH25008	61		120
ALH25034	61		120
ALH25003	61		120
ALH25001	61		120
EQK97643	61		120
ANOV01017390	1339		1109
LKHE01001116	4729		4541
JAAVMX010000001	6699091		6699279
GAGW01008880	1127		1020
OSIN7648	182		360
AGW27560	121	VRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVDQDDAGSQKLFQDESSANLDQ	180
ALH24946	121		180
ALH24995	121		180
ALH24992	121		180
ALH25047	121		180
ALH24948	121	SF	180
ALH25043	121		180
ALH25054	121	L	180
ALH25005	121		180
ALH24999	121		180
ALH25008	121		180
ALH25034	121		180
ALH25003	121		180
ALH25001	121		180
EQK97643	121		180
ANOV01017390	1108		929
LKHE01001116	4540		4318
JAAVMX010000001	6699280		6699502
GAGW01008880	1019		840
OSIN7648	361		541
AGW27560	181	SSLLSAEYPSTEIELLSALVNIYGFDPDHGADLVERMGSSHSGIMAPRAANCTPPVSYTKE	240
ALH24946	181		240
ALH24995	181		240
ALH24992	181		240
ALH25047	181		240
ALH24948	181		240
ALH25043	181	I T	240
ALH25054	181		240
ALH25005	181		240
ALH24999	181		240
ALH25008	181		240
ALH25034	181		240
ALH25003	181	G	240
ALH25001	181		240
EQK97643	181		240
ANOV01017390	928		749
LKHE01001116	4317		4138
JAAVMX010000001	6699503		6699682
GAGW01008880	839		660
OSIN7648	542		719
AGW27560	241	KIDFINTIRSDPVQATKEILGDCYDETTIKLLGVKSHNVESVDSITHLSMQREYQAPRFF	300
ALH24946	241		300
ALH24995	241		300
ALH24992	241		300
ALH25047	241		300
ALH24948	241		300
ALH25043	241	I	300
ALH25054	241		300
ALH25005	241	V	300
ALH24999	241	V	300
ALH25008	241		300
ALH25034	241		300
ALH25003	241	K	300
ALH25001	241	K	300
EQK97643	241	K	300
ANOV01017390	748		569
LKHE01001116	4137		3958
JAAVMX010000001	6699683		6699862
GAGW01008880	659		480
OSIN7648	720	F	863

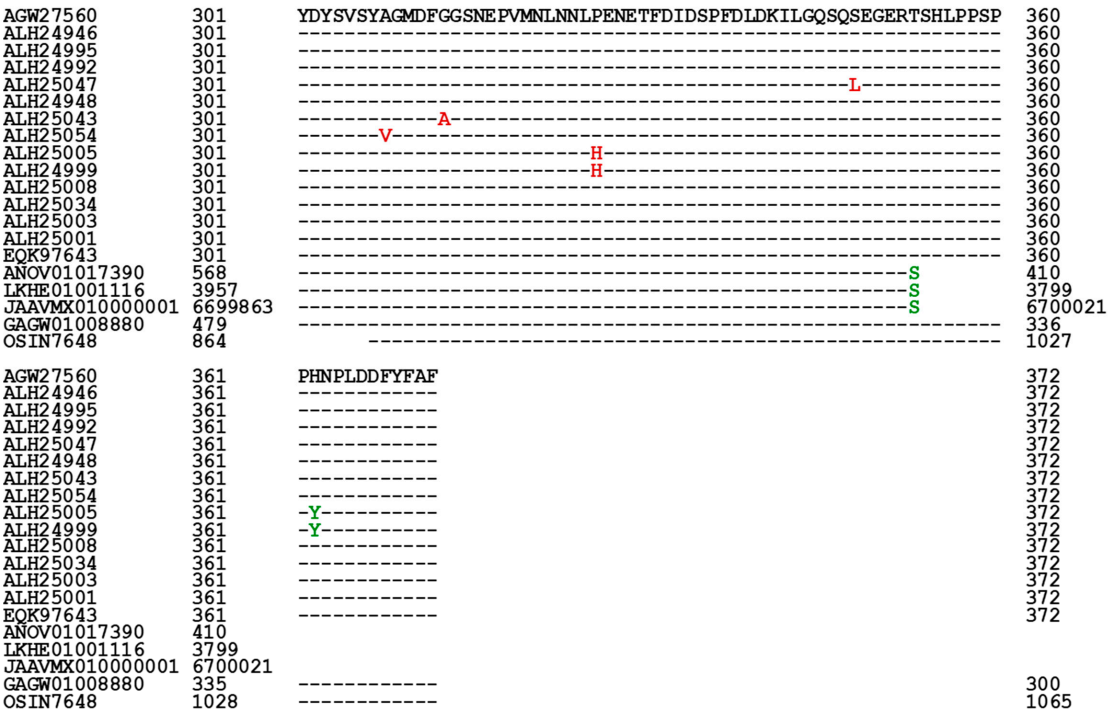


Figure 7. Alignment of the full-length sequences of representative MAT1-1-1 proteins of 15 structural morphs and corresponding translated sequence segments of the genome and metatranscriptome assemblies of *H. sinensis* strains and natural *C. sinensis*. The MATAlpha_HMGbox domain is highlighted in blue and underlined in the query protein sequence AGW27560 (51→225). The residues shown in green indicate conservative amino acid substitutions, and those in red indicate nonconservative amino acid substitutions. The hyphens indicate identical amino acid residues, and the spaces denote unmatched protein sequence gaps.

The MAT1-1-1 protein contains a MATAlpha_HMGbox domain, which is found in high-mobility group (HMG) proteins involved in DNA binding [Hu *et al.* 2013]. This domain is located in the segment 51→225 of the query sequence AGW27560, as highlighted in blue and underlined in Figure 7. Some nonconservative residue substitutions in other MAT1-1-1 proteins were found within this domain, as shown in red in Table 3.

The full-length MAT1-1-1 protein sequence EQK97643 (372 aa) was derived from *H. sinensis* strain Co18 under the AlphaFold UniProt T5A511 and published in GenBank on 22-MAR-2015 (*cf.* Table 1; Figure 3). A segment of the genome assembly ANOV01017390 (410←1519), which was also annotated as KE657544 (410←1519) in GenBank, was derived from the same *H. sinensis* strain but published in GenBank on 20-AUG-2013 and was found to be C-terminally truncated (352 aa; 95.1% query coverage *vs.* EQK97643).

Table 3 summarizes the protein sequence alignment results, including mutant amino acid residues and the percent similarity *vs.* the sequences of the “likely authentic” full-length MAT1-1-1 protein AGW27560, which are correlated with the statistical and structural analytical results obtained from Bayesian clustering and 3D structure prediction (including the stereostructure morphs and the associated AlphaFold UniProt codes) (*cf.* Table 1, Figures 1, 3, and 7). The correlations shown in Table 3 indicate that the minor sequence differences within or outside the MATAlpha_HMGbox domain of the MAT1-1-1 protein sequences may have had an impact on the diverse 3D structures.

Table 3. Summary of the full-length MAT1-1-1 protein sequence alignment results (mutant amino acids and the percent similarity *vs.* the likely authentic protein AGW27560), correlating with the Bayesian branches/clusters and the 3D structural models and associated with the AlphaFold UniProt codes.

Accession number	% similarity to AGW27560	Amino acid residue substitution		Bayesian cluster		3D structure model	AlphaFold UniProt code
		conservative	nonconservative	Branch	Cluster		
AGW27560	100%					A	U3N942
ALH24946	99.4%	Q-to-K, H-to-		A1	A	B	A0A0N9QMM1
EQK97643	100%	Y				C	T5A511
ALH24992	99.7%		A-to-T	A2	A	D	A0A0N9R5B3
ALH24995		T-to-S				E	A0A0N7G849
ALH25047	99.4%		S-to-L	A3	A	F	A0A0N9R4V2
ALH24948						G	A0A0N9QUF3
ALH25043	98.9%		R-to-I, P-to-T, T-to-I, G-to-A,		B	H	A0A0N9QMS9
ALH25054	99.4%	I-to-L	A-to-V		C	I	A0A0N7G845
ALH25005	99.2%	H-to-Y	P-to-H	D1	D	J	A0A0N9QUK2
ALH24999	98.1%	S-to-T, I-to-V, H-to-Y	A-to-V, A-to-T	D2	D	K	A0A0N9QMT4
ALH25008	99.7%	Y-to-H		E1	E	L	A0A0N9QMR3
ALH25034	99.4%	E-to-K, Y-to-H		E2	E	M	A0A0N9QMS4
ALH25003	99.2%	E-to-K, Y-to-H	S-to-G	E3	E	N	A0A0N7G850
ALH25001	98.4%	S-to-T, V-to-I, E-to-K, Y-to-H	A-to-V, A-to-T	E4	E	O	A0A0N9R4Q4

Note: The amino acid residue substitutions shown in red indicate nonconservative changes within the MATalpha_HMGbox domain of the MAT1-1-1 proteins. Other residue substitutions shown in black are located within or outside the MATalpha_HMGbox domain.

Among the 138 MAT1-1-1 protein sequences, 20 are truncated at both the N- and C-termini, showing 68–80% query coverage and belonging to 9 diverse 3D structural morphs (*cf.* Figure 4). Eighteen of the 20 truncated proteins presented 100% sequence identity with the representative full-length MAT1-1-1 protein AGW27560 under the UniProt code U3N942 (*cf.* Table 1). Among the 18 truncated proteins, 17 were clustered into Branch A1 of Cluster A in the Bayesian clustering tree (*cf.* Figure 1); however, the protein AGW27526 under the UniProt code U3N7G5 showed the longest Bayesian clustering distance *vs.* other truncated proteins and was clustered into Branch A3 of Cluster A (*cf.* Figure 1). Two other truncated proteins (AGW27522 and AGW27536) under the UniProt codes U3N6U0 and U3N7H7 shared 99.6–99.7% sequence similarity with the query sequence AGW27560 with either a nonconservative P-to-L substitution or an L residue deletion and were clustered into Branches A2 and A3 of Cluster A, respectively.

Figure 7 also shows the C-terminally truncated MAT1-1-1 proteins encoded by the genome assemblies ANOV01017390, LKHE01001116, and JAAVMX010000001 of *H. sinensis* strains Co18, 1229, and IOZ07, respectively [Hu *et al.* 2013; Li *et al.* 2016a, 2024b; Shu *et al.* 2020]. The truncated MAT1-1-1 proteins encoded by the genome assemblies had a deletion of 19 amino acid residues (SHLPPSPPHNPLDDFYFAF) at the C-termini and contained a nonconservative T-to-S substitution

(cf. Figure 7). The MAT1-1-1 protein encoded by the metatranscriptome assembly GAGW01008880 of natural *C. sinensis* is truncated by 96 amino acids at the N-terminus (MTTRNEVMQRLSSVRADVLLNFLTDDAIFQLA-SRHESTTEADVLTVPSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYLKLDPVQQ-QKTASG) partially with and outside the MATalpha_HMGbox domain [Hu *et al.* 2013; Xiang *et al.* 2014; Li *et al.* 2024b]. The MAT1-1-1 protein encoded by the metatranscriptome assembly OSIN7648 features midsequence truncation with a deletion of 18 amino acids (SMQREYQAPRFFYDYSVS) outside the MATalpha_HMGbox domain and a nonconservative L-to-F substitution within the exon II-encoding region of the MAT1-1-1 gene [Xia *et al.* 2017; Li *et al.* 2024b].

3.6. Primary Structures of the MAT1-2-1 Proteins

Among the 74 MAT1-2-1 proteins available in the AlphaFold database, 69 are full-length proteins containing 249 amino acids and are attributed to diverse 3D structural morphs under 17 UniProt codes (cf. Figure 5; Table 2). Among the 69 full-length proteins, 39 are 100% identical to the query sequence AEH27625 and clustered into Branch I-1 of Cluster I of the Bayesian tree (cf. Figure 2). The remaining 30 full-length proteins share 97.6–99.6% sequence similarity with the query protein sequence, containing various conservative and nonconservative substitutions of amino acid residues at isolated sites, which may have an impact on the mating function.

Figure 8 shows the alignment of the full-length MAT1-2-1 protein sequences covering 5 Bayesian clusters (Branches I-1, I-2, II-1, II-2, III, IV-1, IV-2, V-1, and V-2; cf. Figure 2) and 17 AlphaFold 3D structural morphs (cf. Figure 5 and Table 2), as well as the MAT1-2-1 protein sequences encoded by the genome and transcriptome assemblies of *H. sinensis* and the metatranscriptome assembly of the natural *C. sinensis* sample that was collected from Deqin, Yunnan Province, China. According to GenBank, both the MAT1-2-1 protein sequence EQL04085 and the genome assembly ANOV01000063 (9329→10182) were derived from *H. sinensis* strain Co18 and submitted to GenBank by the same group of authors. However, the genome assembly ANOV01000063 (9329→10182) contains a conservative S-to-A substitution, whereas EQL04085 does not contain this substitution. Note: the arrows “→” and “←” indicate sequences in the sense and antisense strands of the genomes, respectively.

AEH27625	1	MANPINMIPNPQWNATDYEAIWKGLEAQVNPFSQLCLEGDFFRQLDDAAKLFIAKRLME	60
AFX66437	1	-----N-----	60
ACV60391	1	-----	60
ACV60372	1	-----L-----	60
ATV43040	1	-----I-----	60
ACV60417	1	-----I-----	60
ACV60415	1	-----	60
AFX66476	1	-----G-----	60
AFX66472	1	-----G-----	60
ACV60363	1	-----	60
ACV60385	1	-----V-----	60
ACV60399	1	-----I-----	60
AFX66401	1	-----T-----	60
AFX66475	1	-----	60
AFX66484	1	-----	60
AGW27537	1	-----	60
EQL04085	1	-----	60
ANOV01000063	9329	-----	9508
LWBQ01000021	239726	-----	239547
NGJJ01000619	23883	-----	23704
LKHE01001605	14713	-----	14534
GCQL01020543	1143	-----	964
OSIN7649	1	-----	60
AEH27625	61	HVQESVLYVNDGNGPDRVYLGAPRHFVVGGMILQISGYAPYWIRRSVSKVVTATVLAPP	120
AFX66437	61	-----R-----	120
ACV60391	61	-----	120
ACV60372	61	-----	120
ATV43040	61	-----	120
ACV60417	61	-----	120
ACV60415	61	-----	120
AFX66476	61	-----A-----	120
AFX66472	61	-----A-----	120
ACV60363	61	-----	120
ACV60385	61	-----	120
ACV60399	61	-----	120
AFX66401	61	-----	120
AFX66475	61	-----	120
AFX66484	61	-----	120
AGW27537	61	-----T-----	120
EQL04085	61	-----	120
ANOV01000063	9558	-----	9740
LWBQ01000021	239506	-----	239315
NGJJ01000619	23666	-----	23472
LKHE01001605	14496	-----	14302
GCQL01020543	963	-----	784
OSIN7649	61	-----	120

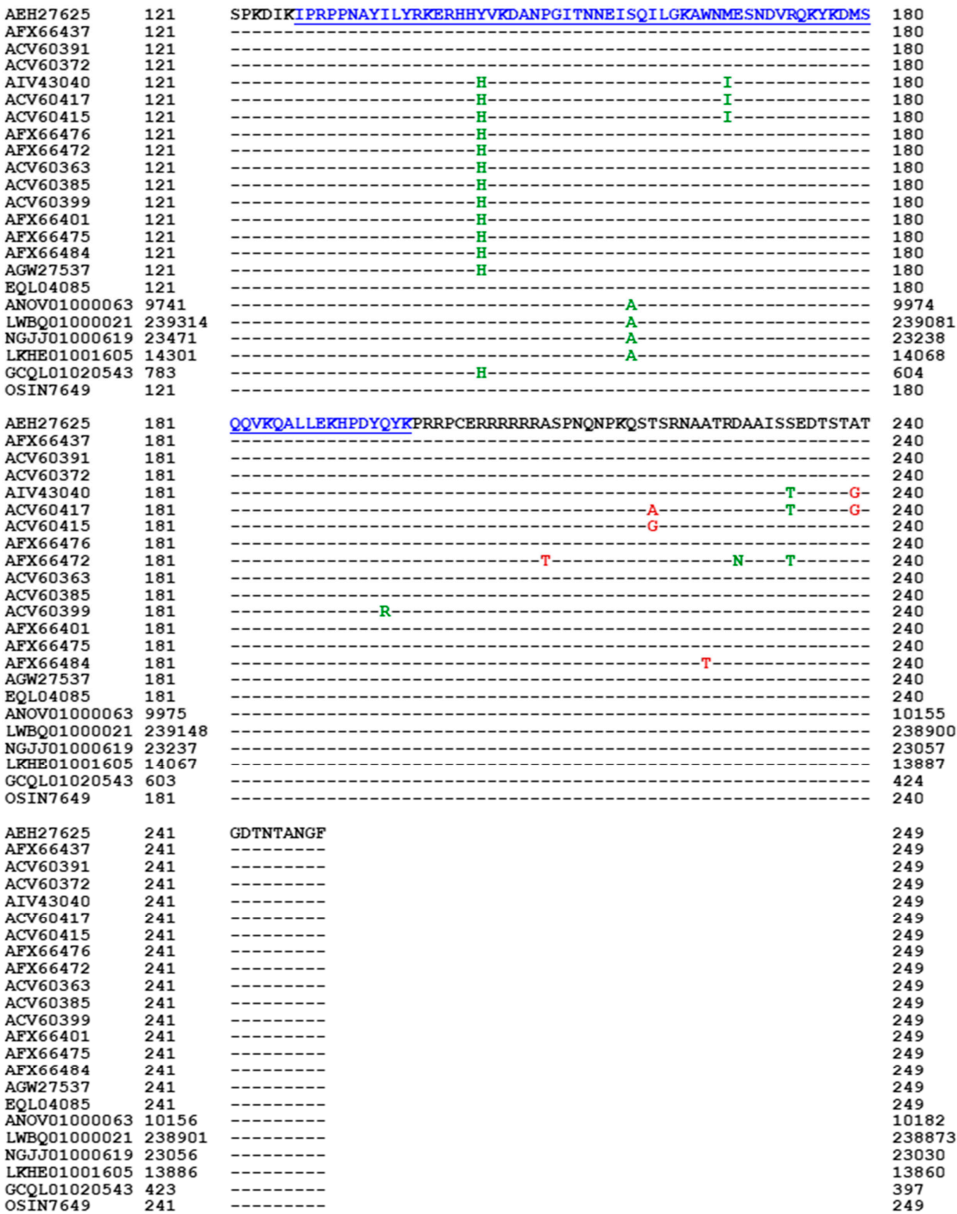


Figure 8. Alignment of the full-length sequences of representative MAT1-2-1 proteins of 17 diverse 3D structural morphs and the translated segments of the corresponding genome, transcriptome and metatranscriptome assemblies of *H. sinensis* strains and natural *C. sinensis*. An HMG-box_ROX1-like domain is highlighted in blue and underlined in the query protein sequence AEH27625 (127→197). The residues shown in green refer to conservative amino acid substitutions, and those in red indicate nonconservative amino acid substitutions. The hyphens indicate identical amino acid residues.

The MAT1-2-1 protein contains an HMG-box_ROX1-like domain, which binds the minor groove of DNA in a sequence-specific manner [Hu *et al.* 2013]. This domain is located in the segment 127→197 of the query sequence AEH27625 shown in blue and underlined in Figure 8. Some conservative residue substitutions in other MAT1-2-1 proteins were found within this domain, as shown in red in Table 4.

Table 4 summarizes the protein sequence alignment results, including mutant amino acid residues and percent similarities relative to the “likely authentic” full-length MAT1-2-1 protein AEH27625. The sequence alignment results are correlated with the statistical and structural analytical results obtained from the Bayesian clustering and 3D structure prediction (including the stereostructure models and the associated AlphaFold UniProt codes) (*cf.* Table 2, Figures 2, 5, and 8). The correlations shown in Table 4 indicate that the minor sequence differences within or outside the

HMG-box_ROX1-like domain of the MAT1-2-1 protein sequences may have an impact on the diverse 3D structures.

Table 4. Summary of the full-length MAT1-2-1 protein sequence alignment results (mutant amino acids and percent similarity *vs.* the likely authentic protein AEH27625), correlating with the Bayesian branches/clusters and the 3D structural models and associated with the AlphaFold UniProt codes.

Accession number	% similarity to AEH27625	Amino acid residue substitution		Bayesian cluster		3D structure model	AlphaFold UniProt code
		conservative	nonconservative	Branch	Cluster		
AEH27625	100%			I-1	I	A	D7F2E9
EQL04085	100%					B	T5AF56
AFX66437		D-to-N				C	V9LW10
ACV60391	99.6%		G-to-R	I-2	I	D	D7F2H1
ACV60372			Q-to-L			E	D7F2F2
AIV43040	97.6%	Y-to-H, M-to-I, Q-to-R, S-to-T	T-to-I, A-to-G	II-1	II	F	A0A0A0RCF5
ACV60417	97.6%	Y-to-H, M-to-I, S-to-T	T-to-I, T-to-A, A-to-G	II-2	II	G	D7F2J7
ACV60415	98.8%	Y-to-H, M-to-I	T-to-G		III	H	D7F2F5
AFX66476	98.8%	Y-to-H	D-to-G, V-to-A	IV-1	IV	I	V9LWC9
AFX66472	97.6%	Y-to-H, D-to-N, S-to-T	D-to-G, V-to-A, A-to-T	IV-2	IV	J	V9LVS8
ACV60363	99.6%	Y-to-H		V-1	V	K	D7F2E3
ACV60385	99.2%	I-to-V, Y-to-H		V-2	V	L	D7F2G5
ACV60399	99.2%	Y-to-H, Q-to-R		V-2	V	M	D7F2H9
AFX66401	99.2%	Y-to-H	D-to-I	V-2	V	N	V9LW71
AFX66475	99.2%	Y-to-H	I-to-T	V-2	V	O	V9LVU8
AFX66484	99.2%	Y-to-H	A-to-T	V-2	V	P	V9LWG5
AGW27537	99.2%	V-to-I, Y-to-H		V-2	V	Q	U3N6V5

Note: The amino acid residue substitutions shown in red indicate conservative changes within the HMG-box_ROX1-like domain of the MAT1-2-1 proteins. Other residue substitutions shown in black are located within or outside the HMG-box_ROX1-like domain.

In addition to the 74 full-length MAT1-2-1 proteins, 5 other protein sequences are C-terminally truncated and were clustered into Branches V-1 and V-2 of Cluster V in the Bayesian tree (shown in green in Figure 2), exhibiting 69–74% query coverage and 99.4–99.5% protein sequence similarity to the reference full-length MAT1-2-1 protein AEH27625 under the UniProt code D7F2E9. The 5 truncated MAT1-2-1 proteins contain a single conserved Y-to-H substitution with the HMG-box_ROX1-like domain, belonging to different 3D structural morphs under the 4 AlphaFold UniProt codes (*cf.* Figure 6).

Figure 8 also shows a conservative S-to-A substitution with the HMG-box_ROX1-like domain in the MAT1-2-1 proteins encoded by the genome assemblies ANOV01000063, LKHE01001605, LWBQ01000021, and NGJJ01000619 of *H. sinensis* strains Co18, 1229, ZJB12195, and CC1406-20395, respectively [Hu *et al.* 2013; Li *et al.* 2016a, 2023b, 2024b; Jin *et al.* 2020; Liu *et al.* 2020]. A conservative Y-to-H substitution with the HMG-box_ROX1-like domain was found in the transcriptome assembly GCQL01020543 of the *H. sinensis* strain L0106 [Liu *et al.* 2015; Li *et al.* 2024b]. No mutation was

detected in the MAT1-2-1 protein encoded by the metatranscriptome assembly OSIN7649 of natural *C. sinensis* [Xia *et al.* 2017; Li *et al.* 2024b], indicating no variation in the 3D structures and mating functionality of the MAT1-2-1 protein.

3.7. Differential Genomic Occurrence of the MAT1-1-1 and MAT1-2-1 Proteins in *H. sinensis*

Table 5 lists the differential occurrence of the MAT1-1-1 and MAT1-2-1 proteins encoded by the genome assemblies ANOV01017390/ANOV01000063, JAAVMX010000001, LKHE01001116/LKHE01001605, LWBQ01000021, and NGJJ01000619 of the *H. sinensis* strains Co18, IOZ07, 1229, ZJB12195, and CC1406-20395, respectively [Hu *et al.* 2013; Li *et al.* 2016a; Jin *et al.* 2020; Liu *et al.* 2020; Shu *et al.* 2020]. The genome assemblies LWBQ00000000 and NGJJ00000000 of the *H. sinensis* strains do not contain the genes encoding the MAT1-1-1 proteins, and the genome assembly JAAVMX000000000 does not contain the gene encoding the MAT1-2-1 protein.

Table 5. Percentage similarity between the sequences EQK97643 and AEH27625 for the MAT1-1-1 and MAT1-2-1 proteins, respectively, and the mating protein sequences encoded by the genome assemblies of *H. sinensis* strains.

<i>H. sinensis</i> strain	Genome assembly segment	Percentage similarity	
		MAT1-1-1 (<i>vs.</i> EQK97643)	MAT1-2-1 (<i>vs.</i> AEH27625)
Co18	ANOV01017390 (410←1519)	99.7%	
	ANOV01000063 (9329→10182)		99.6%
1229	LKHE01001116 (3799←4909)	99.7%	
	LKHE01001605 (13860←14713)		99.6%
IOZ07	JAAVMX010000001 (6698911→6700021)	99.7%	
	JAAVMX000000000		—
ZJB12195	LWBQ00000000	—	
	LWBQ01000021 (238873←239726)		99.6%
CC1406-20395	NGJJ00000000	—	
	NGJJ01000619 (23030←23883)		99.6%

Note: The “→” and “←” arrows indicate sequences in the sense and antisense strands of the genomes, respectively.

Although repetitive copies of many genes have been identified in the *H. sinensis* genome assemblies and the mutant repetitive genome sequences have normal transcriptional abilities and encode mutant proteins with altered functional specificities [Li *et al.* 2024a], no repetitive copies of the MAT1-1-1 or MAT1-2-1 genes were identified in the *H. sinensis* genome assemblies, proving that the mutant mating proteins with diverse stereostructures may be encoded by mutant mating-type genes of certain *H. sinensis* strains or heterospecific fungal species other than *H. sinensis*.

3.8. Differential Transcriptomic and Metatranscriptomic Occurrences of the MAT1-1-1 and MAT1-2-1 Proteins in *H. sinensis* and the Natural *C. sinensis* Insect-Fungal Complex

Table 6 shows the differential occurrence of the MAT1-1-1 and MAT1-2-1 proteins encoded by the transcriptome assembly of the *H. sinensis* strain L0106 and the metatranscriptome assemblies of natural *C. sinensis* specimens [Xiang *et al.* 2014; Liu *et al.* 2015; Xia *et al.* 2017]. The transcriptome assembly GCQL00000000 does not contain the gene encoding the MAT1-1-1 protein [Liu *et al.* 2015], and the metatranscriptome assembly GAGW00000000 of natural *C. sinensis* does not contain the gene encoding the MAT1-2-1 protein [Xiang *et al.* 2014]. Worth mentioning, Bushley *et al.* [2013] and Li *et*

al. [2023b, 2024b] reported disrupted translation of the MAT1-2-1 transcript due to alternative splicing of the gene with unspliced intron I, which contains 3 stop codons, in *H. sinensis* strain 1229. The alternative splicing of the MAT1-2-1 gene eventually produces a largely truncated protein lacking the C-terminal portions (including the entire HMG-box_ROX1-like domain) of the protein encoded by exons II and III of the gene and significantly alters the 3D structure of the protein with dysfunctionality.

Table 6. Percentage similarity between the sequences EQK97643 and AEH27625 for the MAT1-1-1 and MAT1-2-1 proteins, respectively, and the mating proteins encoded by the transcriptome assembly of *H. sinensis* strain L0106 and the metatranscriptome assemblies of natural *C. sinensis*.

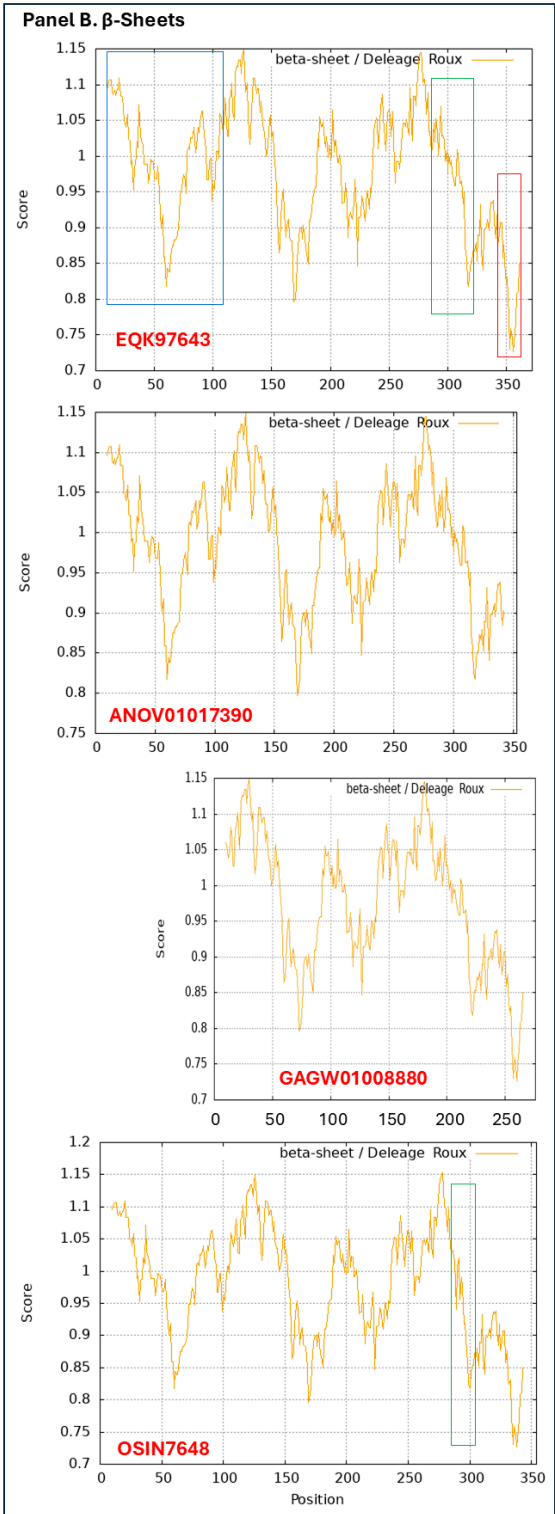
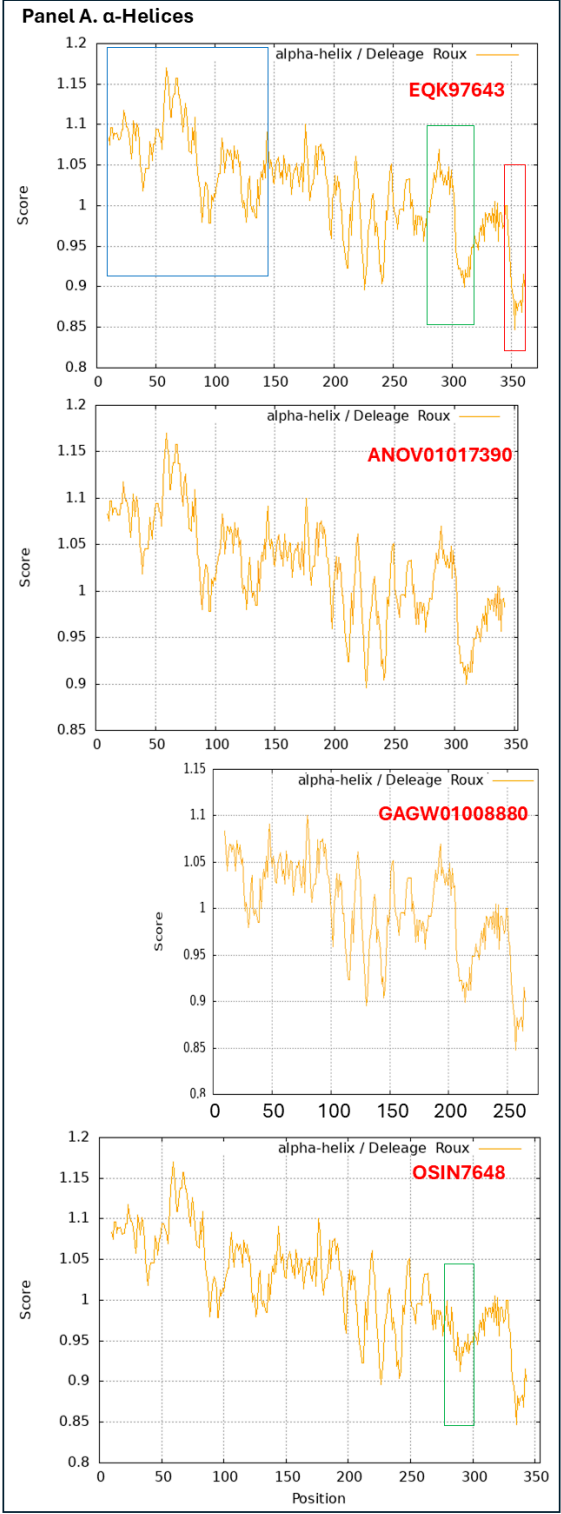
<i>H. sinensis</i> strain or natural <i>C. sinensis</i>	Transcriptome or metatranscriptome assembly segment	Percentage similarity	
		MAT1-1-1 (vs. EQK97643)	MAT1-2-1 (vs. AEH27625)
<i>H. sinensis</i> strain L0106	GCQL00000000	—	
	GCQL01020543 (397←1143)		99.6%
Mature natural <i>C. sinensis</i> (Collected at Deqin, Yunnan)	OSIN7648 (1→1065)	94.9%	
	OSIN7649 (1→397)		100%
Natural <i>C. sinensis</i> * (Collected at Kangding, Sichuan)	GAGW01008880 (300←1127)	100%	
	GAGW00000000		—

Note: *, Natural *C. sinensis* samples of unknown maturation stage. The “→” and “←” arrows indicate sequences in the sense and antisense strands of the genomes, respectively.

3.9. Diverse Secondary (2D) Structures of the MAT1-1-1 Proteins Encoded by the Genome of *H. sinensis* and the Metatranscriptome of Natural *C. sinensis* Insect-Fungal Complex

The predicted 3D structures of the truncated MAT1-1-1 proteins encoded by the genome and metatranscriptome assemblies are not available in the AlphaFold database. Figure 9 shows changes in the 2D structures, α -helices (Panel A), β -sheets (Panel B), β -turns (Panel C), and coils (Panel D) of the truncated MAT1-1-1 proteins that were analyzed *via* ExPASy ProtScale technology. EQK97643 (372 aa) of the *H. sinensis* strain GS09_111 was used as the reference for the authentic MAT1-1-1 proteins for the 2D analysis shown in the upper plots in all panels of Figure 9, which was clustered into Branch A1 of Cluster A (*cf.* Figure 1) and belongs to 3D structural morph A under the UniProt code U3N942 (*cf.* Figure 3).

The upper-middle plots in all panels show the 2D structures of the MAT1-1-1 protein encoded by the genome assembly ANOV01017390 of the *H. sinensis* strain Co18, which also represents 2 other MAT1-1-1 sequences within the genome assemblies LKHE01001116 and JAAVMX010000001 of the *H. sinensis* strains 1229 and IOZ07, respectively. The lower-middle and lowest plots in each panel of Figure 9 present the 2D structures of the MAT1-1-1 proteins encoded by the metatranscriptome assemblies GAGW01008880 and OSIN7648 of natural *C. sinensis*, respectively.



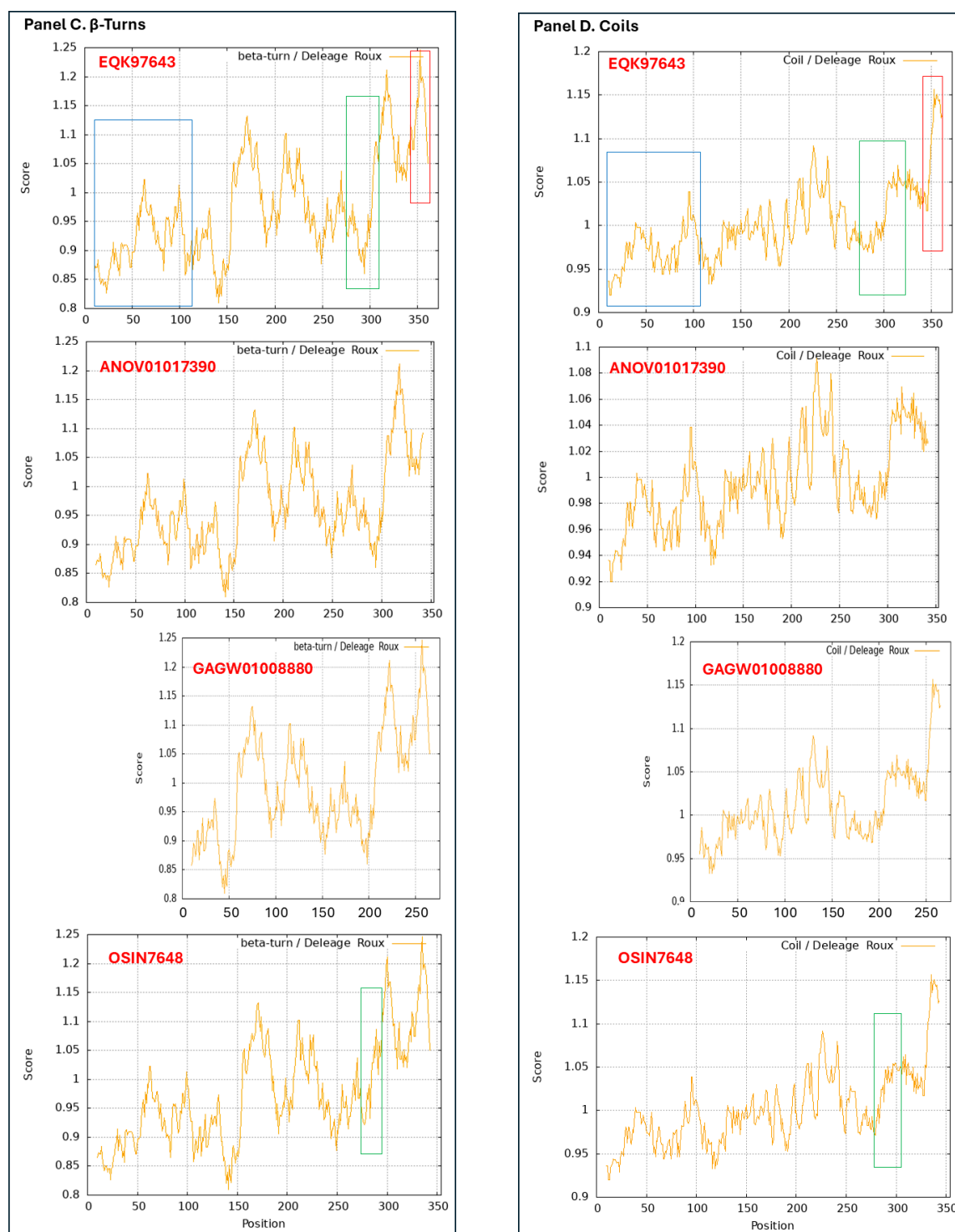


Figure 9. ExPASy ProtScale plots for the α -helices (Panel A), β -sheets (Panel B), β -turns (Panel C), and coils (Panel D) of the MAT1-1-1 proteins. Each panel contains 4 ProtScale plots for the 4 MAT1-1-1 proteins. The open boxes in blue in all EQK97643 plots indicate the N-terminal truncation region occurring in the MAT1-1-1 protein encoded by the metatranscriptome assembly GAGW01008880. The open boxes in red in all the EQK97643 plots indicate the C-terminal truncation region occurring in the genome assembly ANOV01017390. The open boxes in green in all the OSIN7648 plots, as well as in the corresponding region in all plots for the MAT1-1-1 protein EQK97643 for topology and waveform comparisons, indicate the midsequence truncation region occurring in the MAT1-1-1 protein encoded by the metatranscriptome assembly OSIN7648.

The open boxes shown in red in the EQK97643 plots indicate the C-terminal truncation region occurring in the genome assembly ANOV01017390 of *H. sinensis* strain Co18, which also represents

2 other genome assemblies LKHE01001116 and JAAVMX010000001 of *H. sinensis* strains 1229 and IOZ07, respectively.

The N-terminal truncation region of the MAT1-1-1 protein encoded by the metatranscriptome assembly GAGW01008880 of natural *C. sinensis* is indicated with open boxes shown in blue in the reference protein EQK97643 plots in Figure 9.

The open boxes shown in green in the OSIN7648 plots, as well as the corresponding region in the reference EQK97643 plots for the authentic MAT1-1-1 protein for structural comparison, outline the changes in the topology and waveform of the α -helix, β -sheet, β -turn, and coil in the midsequence truncation region in the MAT1-1-1 protein encoded by the metatranscriptome assembly OSIN7648. The topology and waveform changes appear to be more dramatic in the α -helix and β -turn plots in the midsequence truncation region of the protein OSIN7648 than in the β -sheet and coil plots. The midsequence truncation and apparent changes in the 2D structures imply significant alterations in the variable protein folding and 3D structures of the MAT1-1-1 proteins encoded by the genome and metatranscriptome assemblies and their mating functionalities.

3.10. Diverse 2D Structures of the MAT1-2-1 Proteins Encoded by the Genomes, Transcriptomes, and Metatranscriptomes of *H. sinensis* and the Natural *C. sinensis* Insect-Fungal Complex

The predicted 3D structures of the mutant MAT1-2-1 proteins encoded by these genome, transcriptome, and metatranscriptome assembly sequences are unavailable in the AlphaFold database. Figure 10 shows the 2D structures of the MAT1-2-1 proteins for the α -helices (Panel A), β -sheets (Panel B), β -turns (Panel C), and coils (Panel D). Each panel of Figure 10 contains 2 ProtScale plots for 2 MAT1-2-1 proteins. AEH27625 (249 aa) of the *H. sinensis* strain CS2 was used as the reference for the full-length MAT1-2-1 protein shown in the upper plots of all panels. The lower plots in all panels of Figure 10 represent the MAT1-2-1 protein encoded by the genome assembly ANOV01000063 (9,329→10,182), also representing 3 other genomic sequences, namely, NGJJ01000619 (23,030←23,883), LWBQ01000021 (238,873←239,726), and LKHE01001605 (13,860←14,713). Figure 10 shows slight changes in the topology and waveforms of the α -helices, β -turns, and coils in the MAT1-2-1 protein sequences encoded by the genome assembly ANOV01000063. The variation regions are outlined with the open boxes shown in red in the ANOV01000063 plots in Panels A, C, and D, respectively, as well as in the corresponding region in the reference MAT1-2-1 protein AEH27625 plots for topology and waveform comparisons. No apparent 2D changes were found in the topology or waveforms of the β -sheets of the MAT1-2-1 proteins encoded by the genome sequences.

In addition, no apparent changes were observed in the topology and waveforms in the ProtScale plots for the MAT1-2-1 proteins encoded by the transcriptome assembly GCQL01020543 (397←1143) of *H. sinensis* strain L0106 and the metatranscriptome assembly OSIN7649 (1→249) of natural *C. sinensis*, as compared with those for the reference protein AEH27625 [Liu *et al.* 2015; Xia *et al.* 2017; Li *et al.* 2024b]. Thus, these fully functional MAT1-2-1 proteins potentially belong to 3D structural morph A under the UniProt code U3N942 (*cf.* Figure 3) and are included within Branch I-1 of Cluster I in the Bayesian tree (*cf.* Figure 2).

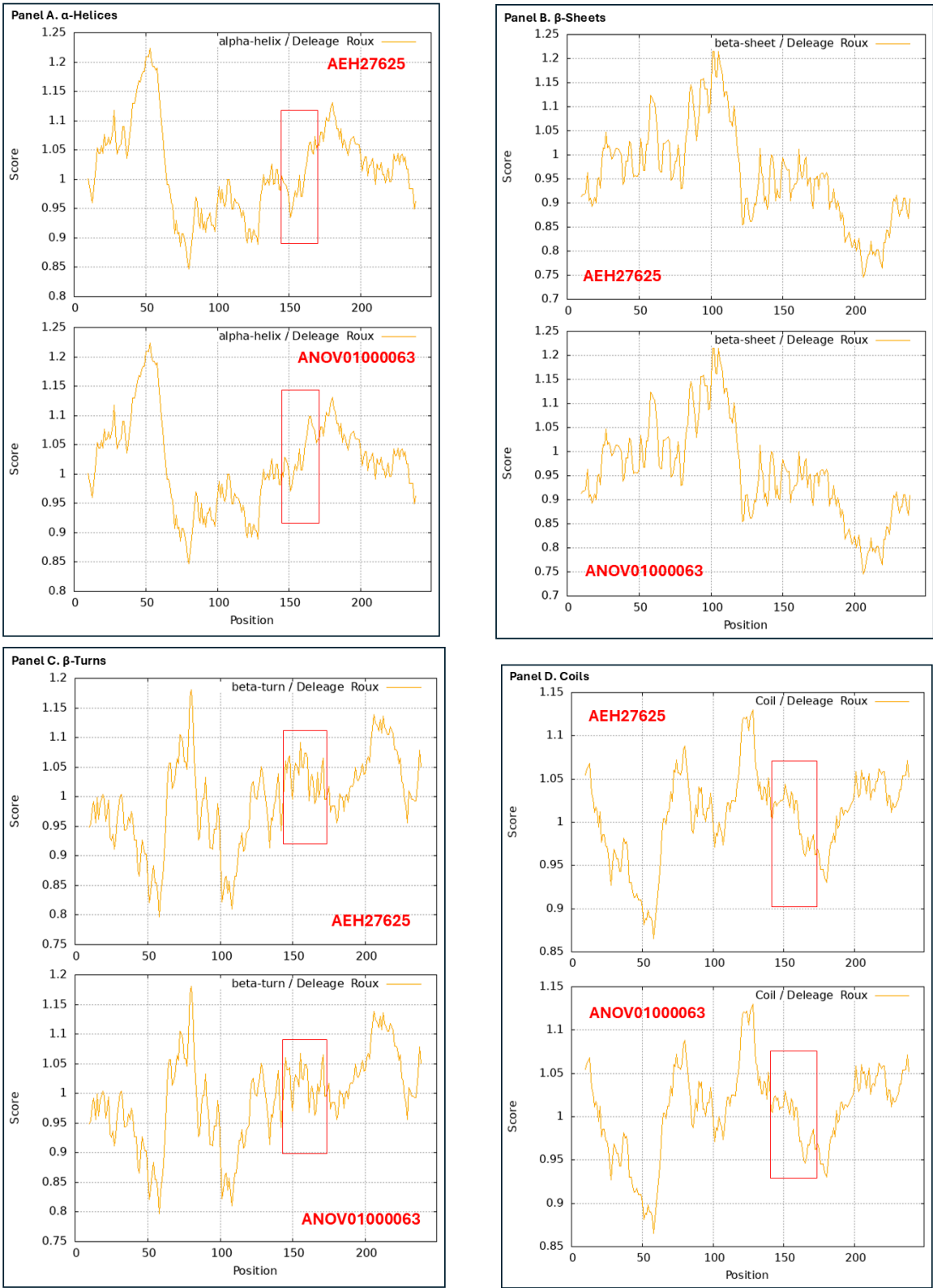


Figure 10. ExPASy ProtScale plots for the α -helices (Panel A), β -sheets (Panel B), β -turns (Panel C), and coils (Panel D) of the MAT1-2-1 proteins. Each panel contains 2 ProtScale plots. The open boxes in red in the ANOV01000063 plots in Panels A, C, and D for the α -helices, β -turns, and coils, as well as in the corresponding region in the AEH27625 plots for the authentic MAT1-2-1 protein for topology and waveform comparisons, indicate the variation region occurring in the genome assembly.

4. Discussion

4.1. Protein 3D Structure Analysis via the AI-Based AlphaFold Prediction System in Combination with Statistical Bayesian Clustering Technology to Stratify 3D Structure Models

The AI-based AlphaFold 3D structure prediction technology displayed extremely refined and powerful abilities in exploring and uncovering the subtle sequence differences of the proteins and in revealing the impact of these subtle sequence differences on 3D structure predictions. The AlphaFold database contains 15 diverse 3D structural models for the full-length MAT1-1-1 proteins (excluding the truncated proteins) and 17 diverse 3D structural models for the full-length MAT1-2-1 proteins (*cf.* Figures 3 and 5).

As shown in Figures 1–6 and Tables 1–2, the complexity of the 3D structure heteromorphs predicted using AlphaFold technology were clarified and organized after the combined use of statistical Bayesian clustering analysis. This combination of the 2 powerful technologies helped stratify the 3D structural heteromorph groups. As a result, the 15 and 17 stereostructure heteromorphs for the full-length MAT1-1-1 and MAT1-2-1 proteins, respectively, were stratified into 5 Bayesian clusters for each of the mating proteins, regardless of whether the clusters were branched or unbranched. Thus, different 3D structure models within a special Bayesian branch are closely structurally related, whereas other models may be structurally and statistically distant.

4.2. Heteromorphic 3D Structures of the MAT1-1-1 and MAT1-2-1 Proteins in *H. sinensis* Strains and Wild-Type *C. sinensis* Isolates

The study presented in this paper demonstrated the heteromorphic 3D structures of the MAT1-1-1 and MAT1-2-1 proteins in 173 *H. sinensis* strains and wild-type *C. sinensis* isolates. Appropriate interaction of the MAT1-1-1 and MAT1-2-1 proteins with functional stereostructures is essential for the sexual reproduction of *O. sinensis* during the lifecycle of the natural *C. sinensis* insect-fungi complex. However, 75.7% of the strains/isolates contained either MAT1-1-1 or MAT1-2-1 proteins but did not generate corresponding pairing mating proteins. These strains/isolates were harvested from scattered production locations on the Qinghai-Tibet Plateau. The harvesting locations and protein accession information are available in both the GenBank and AlphaFold databases.

A total of 6 Bayesian clusters with clustering branches and 24 AlphaFold-predicted 3D structural morphs were demonstrated for the heteromorphic stereostructures of 138 MAT1-1-1 proteins (*cf.* Figures 1 and 3–4). The full-length and truncated MAT1-1-1 proteins belonged to 15 and 9 diverse morphs of 3D structures, respectively. Most frequently detected “authentic” MAT1-1-1 proteins were clustered into Branch A1 of Cluster A in the Bayesian tree (*cf.* Figure 1 and Table 1), belonging to 3D structural morph A of the MAT1-1-1 proteins (*cf.* Figure 3). Figure 7 and Table 3 show that the scattered amino acid residue substitutions, conservative or nonconservative, are located within or outside the MATalpha_HMGbox domain of MAT1-1-1 protein sequences [Hu et al. 2013].

A total of 5 Bayesian clusters with clustering branches and 21 AlphaFold 3D structural morphs were demonstrated for the heteromorphic stereostructures of 79 MAT1-2-1 proteins (*cf.* Figures 2 and 5–6). The full-length and truncated MAT1-2-1 proteins belonged to 17 and 4 diverse morphs of 3D structures, respectively. Many (52.7%) of the MAT1-2-1 proteins were clustered into Branch I-1 of Cluster I in the Bayesian tree (*cf.* Figure 2 and Table 2), belonging to 3D structural morph A of the MAT1-2-1 proteins (*cf.* Figure 5). Figure 8 and Table 4 show that the dispersed amino acid residue substitutions, conservative or nonconservative, are located within or outside the HMG-box_ROX1-like domain of MAT1-2-1 protein sequences [Hu et al. 2013].

Zhang and Zhang [2015] reported 4.7% and 5.7% variations in the exon sequences of the MAT1-1-1 and MAT1-2-1 genes, respectively, in numerous wild-type *C. sinensis* isolates. Exon variations may disrupt the translation of coding sequences, in addition to alternative splicing and differential occurrence and transcription of mating genes, as reported by Li *et al.* [2023b, 2024b]. The mutation rates reported by Zhang and Zhang [2015] appeared to be much greater than those present in the GenBank database, which includes numerous variable sequences of the MAT1-1-1 and MAT1-2-1

proteins of *C. sinensis* isolates [Li *et al.* 2024b]. Presumably, Zhang and Zhang [2015] might not have uploaded all the variable sequences to the GenBank database to truthfully represent the natural diversity of the variable stereostructures of the mating proteins.

Although the AlphaFold database does not include the predicted 3D structures for the mating proteins encoded by the genome and transcriptome assemblies of *H. sinensis* strains and the metatranscriptome assemblies of the natural *C. sinensis* insect-fungi complexes, the 2D structures of the mutant MAT1-1-1 and MAT1-2-1 proteins (including those with truncation of large protein segments) that were translated from the nucleotide sequences were analyzed to explore the variations in the α -helices, β -sheets, β -turns, and coils *via* ExPASy ProtScale technology (*cf.* Figures 7–10). The apparent changes in the 2D structures of the mating proteins indicate altered 3D structures and subsequent dysfunction and even complete deactivation of the mating proteins.

Research has confirmed that the MAT1-1-1 and MAT1-2-1 proteins can interact with each other [Jacobsen *et al.* 2002; Rams & Kück 2022]. The interaction of mating proteins with functional stereostructures is essential for the recognition of compatible mating partners and plays a crucial role in regulating sexual reproduction by controlling gene expression related to mating compatibility within a fungal species. The MAT1-1-1 protein contains a MATalpha_HMGbox domain, and the MAT1-2-1 protein contains an HMG-box_ROX1-like domain [Hu *et al.* 2013]. Both domains are involved in DNA binding, binding to specific DNA sequences in the genome to regulate the transcription of genes involved in mating processes. We highlighted the domains in the sequences of the mating proteins in Figures 7–8 and summarized the conservative and nonconservative amino acid residue substitutions within or outside the domains in the sequences of the mating proteins of the wild-type isolates in Tables 3–4.

The heteromorphic stereostructures of the mating proteins might explain, at least partially, why efforts made in the past 40–50 years to cultivate pure *H. sinensis*, Genotype #1 of the 17 *O. sinensis* genotypes, in research-oriented academic settings to induce the production of fruiting bodies and ascospores have consistently failed, as reported and summarized previously [Holliday & Cleaver 2008; Stone 2010; Hu *et al.* 2013; Zhang *et al.* 2013]. Tables 5–6 of this paper confirmed the differential occurrence of the mating proteins encoded by the genome and transcriptome assemblies of the *H. sinensis* strains and by the metatranscriptome assemblies of the natural *C. sinensis* insect-fungi complex. Bushley *et al.* [2013] and Li *et al.* [2024b] reported alternative splicing of the MAT1-2-1 gene with unspliced intron I, which contains 3 stop codons, in *H. sinensis* strain 1229. Consequently, the C-terminally truncated MAT1-2-1 protein fragment lacked the major portion of the protein encoded by exons II and III of the gene.

4.3. Differential Occurrences of MAT1-1-1 and MAT1-2-1 Proteins with Heteromorphic Stereostructures in *H. sinensis* Strains and *C. sinensis* Isolates

Uploaded to the GenBank and AlphaFold databases by many researchers, 131 (75.7%) of the 173 *H. sinensis* strains and wild-type *C. sinensis* isolates generated either the MAT1-1-1 or MAT1-2-1 protein but not both. This phenomenon was confirmed by the differential occurrence of the mating proteins encoded by the genome, transcriptome, and metatranscriptome assemblies of the *H. sinensis* strains and the natural *C. sinensis* insect-fungi complex samples (*cf.* Tables 5–6).

However, 42 other strains/isolates (24.3%) produced both MAT1-1-1 and MAT1-2-1 proteins, the sequences of which were derived from genomic DNA isolated from *H. sinensis* strains or wild-type *C. sinensis* isolates but, unfortunately, not detected directly *via* gene transcription assays or biochemical examinations. In addition to the intracellular biological processes of the MAT1-1-1 and MAT1-2-1 proteins, such as differential transcription and alternative splicing of the genes, as reported by Li *et al.* [2024b], it needs to be considered whether the mating proteins with heteromorphic 3D structures are capable of expressing their mating functions to accomplish the sexual reproduction of *O. sinensis* during the lifecycle of the natural *C. sinensis* insect-fungi complex.

The MAT1-1-1 proteins from 35 of the 42 strains/isolates were clustered into Bayesian Cluster A (*cf.* Figure 1), accompanied by one of 20 MAT1-2-1 proteins that were clustered into Bayesian Cluster

I or one of 15 MAT1-2-1 proteins that were clustered into Bayesian Cluster V (cf. Figure 2). These results will be meaningful and useful for the future design of biochemical protein research and reproductive physiological research to explore the functionalities of mating proteins that are clustered into different Bayesian clusters and have diverse 3D structural morphs. The most challenging aspect of biochemical examinations is determining the stereostructures and molecular dynamics of fully functioning proteins under native conditions or after renaturation [Zhu & Gray 1994].

Figures 7–8 shows the amino acid residue substitutions in the mating protein sequences of the *H. sinensis* strains and wild-type *C. sinensis* isolates, which cause the alterations of 3D structures shown in Figures 1–6 and Tables 1–2. Table 3 summarizes the correlation between the altered primary structures and the diverse stereostructures. Some amino acid residue substitutions occur within the MATalpha_HMGbox domain and HMG-box_ROX1-like domain of the MAT1-1-1 and MAT1-2-1 protein sequences, respectively. In addition to the single residue substitutions, Figures 7–8 shows more pronounced mutations in the mating proteins encoded by the genome, transcriptome, and metatranscriptome assemblies of *H. sinensis* strains and natural *C. sinensis* insect-fungi complexes, which caused dramatic changes in the 2D structures of the mating proteins (cf. Figures 9–10). These findings will inspire further refined research to accurately locate the causal relationship between specific variations in amino acid sequences and protein 3D structural changes, as well as the potential impact on mating protein functionality.

4.4. Heteromorphic 3D Structures of the Mating Proteins and Sexual Reproductive Behavior of *H. sinensis*, Genotype #1 of *O. sinensis*

Sexual reproduction of *O. sinensis* is crucial for maintaining the natural population volume of the *C. sinensis* insect-fungi complex, which is endangered at Level 2 of National Key Protected Wild Plants [China Ministry of Agriculture and Rural Affairs 2021]. The following hypotheses have been previously proposed for *H. sinensis*, the sole anamorph of *O. sinensis* postulated by Wei *et al.* [2006]: (1) homothallism [Hu *et al.* 2013], (2) pseudohomothallism [Bushley *et al.* 2013], and (3) facultative hybridization [Zhang & Zhang 2015]. These hypotheses are based on nucleotide data derived from molecular, genome, and transcriptome studies of *H. sinensis*. In theory, self-fertilization under homothallism and pseudohomothallism in ascomycetes becomes a reality when interaction of MAT1-1-1 and MAT1-2-1 proteins with appropriate stereostructures exhibit full mating functions within a single fungal cell [Turgeon & Yoder 2000; Debuchy & Turgeo 2006; Jones & Bennett 2011; Zhang *et al.* 2011; Bushley *et al.* 2013; Hu *et al.* 2013; Zheng & Wang 2013; Wilson *et al.* 2015; Zhang & Zhang 2015]. However, after thoroughly analyzing genetic and transcriptional data for *H. sinensis* in the literature, Li *et al.* [2023b, 2024b] reported differential occurrences, alternative splicing, and differential transcription of the mating-type genes of the MAT1-1 and MAT1-2 idiomorphs and the pheromone receptor genes in 237 *H. sinensis* strains. Thus, the evidence indicated that *H. sinensis*, which is Genotype #1 of the 17 genomically independent *O. sinensis* genotypes, is self-sterilizing and incapable of completing self-fertilization but requires sexual partners to accomplish *O. sinensis* sexual reproduction under heterothallism or hybridization.

Because of the absence of repetitive copies of the MAT1-1-1 or MAT1-2-1 genes in the *H. sinensis* genome assemblies, the alternative splicing and differential occurrence and transcription of the mating-type genes and the diversity of heteromorphic 3D structures of the mating proteins with altered functionalities indicate that there may be two or more *H. sinensis* populations, either monoecious or dioecious. The different *H. sinensis* populations may participate as sexual partners capable of producing either the functioning MAT1-1-1 or the functioning MAT1-2-1 protein with proper stereostructures for reciprocal pairing and interaction during successful physiological heterothallism crossing. Thus, a- and α -pheromones and corresponding α - and a-pheromone receptor proteins play critical roles in the communication of sexual signals between sexual partners. If this assumption is correct, the sexual partners might possess indistinguishable *H. sinensis*-like morphological and growth characteristics, as elucidated previously [Kinjo & Zang 2001; Zhang *et al.*

2009; Chen *et al.* 2011; Li *et al.* 2013, 2016b; Mao *et al.* 2013]. For instance, the indistinguishable *H. sinensis* strains 1229 and L0106 produce complementary transcripts of the mating-type genes and proteins of the MAT1-1 and MAT1-2 idiomorphs [Bushley *et al.* 2013; Liu *et al.* 2015; Li *et al.* 2023b, 2024b].

Even if the physiological heterothallism hypothesis is incorrect for *O. sinensis*, one of the mating proteins might be produced by genome-independent, heterospecific fungal species, which would result in plasmogamy and the formation of heterokaryotic cells (*cf.* Figure 3 of [Bushley *et al.* 2013]) to ensure successful sexual hybridization or even parasexual reproduction if the heterospecific species are capable of breaking interspecific reproduction isolation, similar to many cases of fungal sexual hybridization and parasexual reproduction that promote adaptation to the extremely adverse ecological environment on the Qinghai-Tibet Plateau [Pfennig 2007; Seervai *et al.* 2013; Nakamura *et al.* 2019; Du *et al.* 2020; Hénault *et al.* 2020; Samarasinghe *et al.* 2020; Mishra *et al.* 2021; Steensels *et al.* 2021; Kück *et al.* 2022]. Alternatively, to complete physiological heterothallism or hybridization for reproduction, mating partners that produce the MAT1-1-1 and MAT1-2-1 proteins with functional stereostructures might exist in adjacent hyphal cells, which might determine their mating choices, and they may communicate with each other through a mating signal-based transduction system of pheromones and pheromone receptors and form “H”-shaped crossings of multicellular hyphae [Hu *et al.* 2013; Bushley *et al.* 2013; Mao *et al.* 2013]. In fact, Mao *et al.* [2013] illustrated the “H”-shaped morphology in *C. sinensis* hyphae that genetically contained either AT-biased Genotype #4 or #5 of *O. sinensis* without cooccurrence of the GC-biased Genotype #1 *H. sinensis* and that the genome-independent AT-biased *O. sinensis* genotypes shared indistinguishable *H. sinensis*-like morphological and growth characteristics. To date, no study has reported the identification of α - and α -pheromone genes in the genome or transcriptome assemblies of *H. sinensis* strains or in the metatranscriptome assemblies of natural *C. sinensis*; however, α - and α -pheromone receptor genes were found to differentially occur in the genome and transcriptome assemblies of *H. sinensis* and in metatranscriptome assemblies of the natural *C. sinensis* insect-fungi complex [Hu *et al.* 2013; Li *et al.* 2024b]. Thus, the mating signal transduction between sexual partners to ensure appropriate interaction between MAT1-1-1 and MAT1-2-1 proteins with functional stereostructures remains a mystery in *O. sinensis*, which may inspire further research.

4.5. Heteromorphic 3D Structures of the Mating Proteins and Sexual Reproduction Strategy During the Lifecycle of the Natural *C. sinensis*

In addition to the intensive attention given to *H. sinensis*, the N-terminally and midsequence-truncated MAT1-1-1 proteins and the variable MAT1-2-1 proteins encoded by the metatranscriptome assemblies of the natural *C. sinensis* insect-fungi complex exhibit alterations in the 2D structures of the proteins (*cf.* Figures 7–10). These results suggest heteromorphic 3D structures of the mating proteins and dysfunctional or anomalous fungal mating processes during the lifecycle of the natural *C. sinensis* insect-fungi complex.

Li *et al.* [2016b] reported the genetic heterogeneity of the wild-type *C. sinensis* isolates CH1 and CH2, which were isolated from the intestines of healthy living larvae of *Hepialus lagii* Yan, based on their *H. sinensis*-like morphology and growth characteristics. The *C. sinensis* isolates CH1 and CH2 contained GC-biased Genotype #1 (*H. sinensis*) and AT-biased Genotypes #4–5 of *O. sinensis*, as well as *Paecilomyces hepiali* [Dai *et al.* 1989], which was renamed *Samsoniella hepiali* in 2020 [Wang *et al.* 2020]. The impure wild-type *C. sinensis* isolates exhibited 15–39-fold greater inoculation potency on the larvae of *Hepialus armoricanus* than did pure *H. sinensis* ($n=100$ for each inoculant; $P<0.001$), indicating the symbiosis of multiple intrinsic fungi during the lifecycle of natural *C. sinensis*, at least in the larva infection stage [Li *et al.* 2016b].

We found that there are no repetitive copies of the MAT1-1-1 or MAT1-2-1 genes were identified in the *H. sinensis* genome assemblies. Thus, the genetic heterogeneity of the wild-type *C. sinensis* isolates suggests that the coexisting MAT1-1-1 and MAT1-2-1 proteins with varied amino acid sequences and diverse 3D structures detected in 24.3% of the wild-type *C. sinensis* isolates might have

been derived from genome-independent heterospecific fungi, which may pair complementarily and reciprocally to ensure proper interactions and accomplish mating processes during the heterothallic or hybrid reproduction of *O. sinensis*. The different fungal sources for the cooccurring MAT1-1-1 and MAT1-2-1 proteins are evidenced by the species contradiction between the inoculants (3 purified GC-biased *H. sinensis* strains) and the genome-independent teleomorphic AT-biased Genotype #4, which is reportedly the sole teleomorphic fungus present in the fruiting body of cultivated *C. sinensis* [Wei *et al.* 2016]. These authors reported that successful artificial inoculation-based cultivation projects under product-oriented industrial settings used a special cultivation strategy without pursuing strict fungal purification by adding soil that was transported from natural *C. sinensis* production areas on the Qinghai-Tibet Plateau into the industrial cultivation system [Wei *et al.* 2016].

The AMT1-1-1 and MAT1-2-1 proteins with diverse stereostructures may originate from heterogeneous fungal sources in single hyphal and ascosporic cells. Li *et al.* [2013] reported the genetic heterogeneity of 15 cultures of 2 groups derived from monoascospores, the reproductive cells of natural *C. sinensis*, after 25 days of *in vitro* incubation at 18 °C. The first group included 7 homogeneous clones (1207, 1218, 1219, 1221, 1222, 1225, and 1229), containing only *H. sinensis* (GC-biased Genotype #1 of *O. sinensis*). The second group included 8 heterogeneous clones (1206, 1208, 1209, 1214, 1220, 1224, 1227, and 1228), containing both GC-biased Genotype #1 and AT-biased Genotype #5 of *O. sinensis*. The sequences of the GC- and AT-biased *O. sinensis* genotypes reside in independent genomes and belong to independent fungi [Xiao *et al.* 2009; Zhu *et al.* 2010; Zhu & Li 2017; Li *et al.* 2022, 2024b]. Bushley *et al.* [2013], the collaborators of Li *et al.* [2013], observed multicellular heterokaryotic hyphae and ascospores of natural *C. sinensis* with mononucleated, binucleated, trinucleated, and tetranucleated structures and reported the PCR results for 22 clones, which included 7 additional clones (1210, 1211, 1212, 1213, 1216, 1223, and 1226) forming the third group of ascosporic clones in addition to the aforementioned first and second groups of clones (*cf.* Figures 2–3 of [Bushley *et al.* 2013]). However, neither Bushley *et al.* [2013] nor Li *et al.* [2013] reported the genetic features of the third group of ascosporic clones. However, there is no doubt from the 2 literature reports that the ascosporic clones from the third group were genetically distinct from those in the homogeneous first group or heterogeneous second group and apparently belonged to taxonomically different fungal species or fungal complexes.

Zhang and Zhang [2015] commented that the nuclei of binucleated hyphal and ascosporic cells (as well as mononucleated, trinucleated, and tetranucleated cells) of natural *C. sinensis* likely contained different genetic materials. These multicellular hyphal and ascosporic cells of natural *C. sinensis* might contain two or more sets of genomes of independent fungi, which might be responsible for the production of complementary mating proteins for sexual reproductive outcrossing. Thus, the monoascospores of natural *C. sinensis* might be characterized by more complex genetic heterogeneity, coexisting with more heterospecific fungal species than the cooccurring GC-biased Genotype #1 and AT-biased Genotype #5, which were reported by Li *et al.* [2013].

Unlike culture-dependent experiments, which are apparently unable to detect nonculturable fungal species, Li *et al.* [2023c, 2023d] reported culture-independent studies and demonstrated the genetic heterogeneity of *C. sinensis* ascospores and the stromal fertile portion (SFP), which is densely covered with numerous ascocarps, which are the reproductive cells and organs of natural *C. sinensis*. These authors observed semi-ejected and fully ejected ascospores of natural *C. sinensis* and reported the cooccurrence of the GC-biased *O. sinensis* Genotypes #1 and #13/14, AT-biased *O. sinensis* Genotypes #5–6 and #16 within AT-biased clade A in the Bayesian phylogenetic tree, *S. hepiali* ($\equiv$ *P. hepiali*), and an AB067719-type fungus. In addition, the *C. sinensis* SFP contained another fungal group, namely, AT-biased Genotypes #4 and #15 of *O. sinensis*, which were clustered into AT-biased clade B in the Bayesian phylogenetic tree [Li *et al.* 2023d]. Genotypes #4 and #15 were absent in the ascospores, which is consistent with the results of culture-dependent studies [Li *et al.* 2013]. The abundance of fungal components exhibited marked dynamic alterations in a disproportional and asynchronized manner in the *C. sinensis* SFP before and after ascospore ejection and in the 2 types of ascospores [Li *et al.* 2023c]. Thus, the coexistence of the MAT1-1-1 and MAT1-2-1 proteins detected

from the wild-type *C. sinensis* isolates might have been derived interindividually from fungi that might serve as mating partners of *O. sinensis* to accomplish heterothallic or hybrid reproduction.

Li *et al.* [2023b, 2024b] summarized prior scientific evidence regarding the sexual reproduction of *O. sinensis*. Based on genetic heterogeneity with multiple heterospecific fungal species in the natural *C. sinensis* insect-fungi complexes and multicellular heterokaryotic structures of ascospores and hyphae, the scientific evidence may be divided into 2 aspects: (1) the 17 cooccurring genome-independent genotypes of *O. sinensis* in different combinations and (2) the taxonomically heterospecific fungal species, which are based on the mycobiota of >90 co-colonizing fungi belonging to at least 37 fungal genera in the stromata and caterpillar bodies of natural *C. sinensis* insect-fungi complexes [Zhang *et al.* 2010, 2018; Xiang *et al.* 2014; Meng *et al.* 2015; Xia *et al.* 2015, 2017; Guo *et al.* 2017; Wang *et al.* 2018; Zhong *et al.* 2018; Li *et al.* 2019; Zhao *et al.* 2020; Yang *et al.* 2021; Kang *et al.* 2024].

(1) Evidence for the differential cooccurrence of multiple genotypes of *O. sinensis* in the compartments of the natural *C. sinensis* insect-fungi complex:

- (1-b). Differential occurrence of AT-biased Genotype #4 or #5 of *O. sinensis* without the cooccurrence of GC-biased *H. sinensis* in natural *C. sinensis* samples collected from different production areas in geographically remote locations [Engh 1999; Kinjo and Zang 2001; Stensrud *et al.* 2005, 2007; Mao *et al.* 2013].
- (1-c). Multiple cooccurring GC- and AT-biased genotypes of *O. sinensis* have been observed differentially in different combinations in the stroma, caterpillar body, ascocarps and ascospores of natural *C. sinensis* [Xiao *et al.* 2009; Zhu *et al.* 2010; Li *et al.* 2013, 2022, 2023c, 2023d; Mao *et al.* 2013]. The abundances of the *O. sinensis* genotypes underwent dynamic alterations in an asynchronous, disproportional manner in the caterpillar bodies and stromata of *C. sinensis* specimens during maturation, with a consistent predominance of the AT-biased genotypes of *O. sinensis*, not the GC-biased *H. sinensis*, in the stromata, indicating that the sequences of *O. sinensis* genotypes were present in independent genomes of different fungi [Xiao *et al.* 2009; Zhu *et al.* 2010; Gao *et al.* 2011, 2012; Hu *et al.* 2013; Li *et al.* 2013, 2016a, 2020, 2022, 2023c; Zhu & Li 2017; Jin *et al.* 2020; Liu *et al.* 2020; Shu *et al.* 2020].
- (1-d). The GC-biased Genotypes #1 and #2 of *O. sinensis* cooccur in the stromata of natural *C. sinensis*. The abundance of the GC-biased genotypes was dynamically altered during *C. sinensis* maturation [Zhu *et al.* 2010; Gao *et al.* 2012].
- (1-e). The cooccurrence of GC-biased genomically independent Genotypes #1 and #7 of *O. sinensis* was detected in the same specimen of natural *C. sinensis* [Chen *et al.* 2011].
- (1-f). A species contradiction between the anamorphic inoculants (GC-biased Genotype #1 *H. sinensis* strains) and the sole-teleomorph of AT-biased Genotype #4 of *O. sinensis* was detected in the fruiting body of cultivated *C. sinensis* in a product-oriented industrial setting [Wei *et al.* 2016].
- (1-g). Discovery of Genotypes #13 and #14 of *O. sinensis* in semi-ejected and fully ejected multicellular heterokaryotic ascospores, respectively, collected from the same *C. sinensis* samples [Li *et al.* 2023d].
- (1-h). The genetic heterogeneity of ascospores and SFP, the reproductive cells and organs of natural *C. sinensis*, involves multiple GC- and AT-biased *O. sinensis* genotypes in different combinations [Li *et al.* 2013, 2022, 2023c, 2023d].
- (2) Evidence for the differential cooccurrence of heterospecific fungal species in different compartments of the natural *C. sinensis* insect-fungi complex:
 - (2-a). Mycobiota findings for differential cooccurrence of >90 fungal species of at least 37 fungal genera in the caterpillar bodies and stromata of natural *C. sinensis* [Zhang *et al.* 2010, 2018; Xia *et al.* 2015; Guo *et al.* 2017; Zhong *et al.* 2018; Kang *et al.* 2024].
 - (2-b). A good number of *C. sinensis* isolates contained mutant MAT1-1-1 and MAT1-2-1 proteins, especially those proteins with C- and/or N-terminal truncations that belong to 9 and 4 diverse 3D structural morphs (*cf.* Figures 4 and 6), respectively. The mutant proteins were

either clustered into a separate Bayesian clade or clustered within the main clustering branches in the Bayesian trees (*cf.* Figures 1 and 2). The MAT1-1-1 and MAT1-2-1 proteins encoded by metatranscriptome assemblies of natural *C. sinensis* also exhibited either large-segment truncation or sequence variations similar to those observed in wild-type *C. sinensis* isolates (*cf.* Figures 7–10). Some of the mutant proteins might be produced by heterospecific fungi in impure wild-type *C. sinensis* isolates and in natural *C. sinensis* insect-fungi complexes.

- (2-c). Discoveries of the formation of the heterospecific *Cordyceps-Tolypocladium* complex in natural *C. sinensis* [Engh 1999; Stensrud *et al.* 2005, 2007] and the dual anamorphs of *O. sinensis*, involving psychrophilic *H. sinensis* and mesophilic *Tolypocladium sinensis* [Li 1988; Chen *et al.* 2004; Leung *et al.* 2006; Barseghyan *et al.* 2011].
- (2-d). A close association of psychrophilic *H. sinensis* and mesophilic *S. hepiali* ($\equiv$ *P. hepiali*) was found in the caterpillar body, stroma, ascospores, and stromal fertile portion, which was densely covered with numerous ascocarps of natural *C. sinensis*, and in the wild-type *C. sinensis* complexes, which appeared to be difficult to purify [Dai *et al.* 1989; Jiang & Yao 2003; Chen *et al.* 2004; Zhu *et al.* 2007, 2010; Yang *et al.* 2008; Li *et al.* 2016b, 2023d; Zhu & Li 2017].
- (2-e). Although Genotypes #13–14 are among the 17 genotypes of *O. sinensis*, these 2 GC-biased genotypes feature precise reciprocal cross substitutions of large DNA segments among 2 heterospecific parental fungi, namely, *H. sinensis* and an AB067719-type fungus. The taxonomic position of the AB067719-type fungus is undetermined to date, and more than 900 heterospecific fungal sequences, which are highly homologous to AB067719, have been uploaded to GenBank [Li *et al.* 2023d]. Chromosomal intertwining and genetic material recombination may occur after plasmogamy and karyogamy of the heterospecific parental fungi under sexual reproduction hybridization or parasexuality, which is characterized by the prevalence of heterokaryosis and results in concerted chromosome loss for transferring-substituting genetic materials without conventional meiosis [Bennett & Johnson 2003; Sherwood & Bennett 2009; Bushley *et al.* 2013; Seervai *et al.* 2013; Nakamura *et al.* 2019; Mishra *et al.* 2021; Kück *et al.* 2022; Li *et al.* 2023d].

The selection of different genotypes of *O. sinensis* or heterospecific fungal species as sexual partners depends on their mating choices for hybridization or parasexuality and their ability to break interspecific isolation barriers to adapt to extremely harsh ecological environments on the Qinghai-Tibet Plateau, and the seasonal climate changes from extremely cold winters, when *C. sinensis* is in its asexual growth phase, to warmer springs and early summers, when *C. sinensis* switches to the sexual reproduction phase [Pfennig 2007; Du *et al.* 2020; Hénault *et al.* 2020; Samarasinghe *et al.* 2020; Steensels & Gallone 2021].

5. Conclusions

The analysis of the MAT1-1-1 and MAT1-2-1 proteins in the 173 *H. sinensis* strains and wild-type *C. sinensis* isolates revealed heteromorphic stereostructures of the mating proteins, which were clustered into multiple Bayesian clustering clades and branches. In addition to the evidence of alternative splicing and differential occurrence and translation of the MAT1-1-1 and MAT1-2-1 genes in *H. sinensis* [Li *et al.* 2023b, 2024b], the diversity of heteromorphic mating proteins suggested stereostructure-related alterations in the mating functions of proteins and provided additional evidence supporting the self-sterility hypothesis under heterothallic and hybrid reproduction for *O. sinensis*, including *H. sinensis*, Genotype #1 of the 17 genome-independent *O. sinensis* genotypes. The heteromorphic stereostructures of the mutant MAT1-1-1 and MAT1-2-1 proteins discovered in wild-type *C. sinensis* isolates and the natural *C. sinensis* insect-fungi complexes may suggest diverse sources of the mating proteins produced by two or more cooccurring heterospecific fungal species in natural *C. sinensis* that have been discovered in mycobiota, molecular, metagenomic, and metatranscriptomic

studies, regardless of whether culture-dependent or culture-independent research strategies were used.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. TABLE S1 Amino acid scales based on the general chemical characteristics of their side chains for ProtScale analysis (<https://web.expasy.org/protscale/>) to predict secondary structures (α -helices, β -sheets, β -turns, and coils) of proteins; FIGURE S1 Alignment of the full-length sequences of reference MAT1-2-1 proteins AEH27625 of structural morph Branch I-1 of Cluster I, ACV60363 of structural morph Branch V-1, and 5 MAT1-2-1 proteins.

Acknowledgments: The authors are grateful to Prof. Mu Zang, Prof. Ru-Qin Dai, Prof. Ping Zhu, Prof. Wei Liu, and Prof. Zong-Qi Liang for their consultation.

Author Contributions: Conceptualization, XZL, YLL, and JSZ; methodology, JSZ; formal analysis, JSZ; data curation, XZL and JSZ; writing — original draft preparation, JSZ; writing — review and editing, XZL, YLL, and JSZ; project administration, YLL; funding acquisition, YLL. All the authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by (1) the Youth Fund (#2023-ZJ-958Q) awarded by the Department of Science & Technology of Qinghai Province; (2) the Joint Science Project of the Chinese Academy of Sciences, Qinghai Provincial Government, and Sanjiangyuan National Park (#LHZX-2022-01); and (3) the grant “The protective harvesting and utilization project for *Ophiocordyceps sinensis*” (#QHCY-2023-057) awarded by the Qinghai Province of China.

Institutional Review Board Statement: Not applicable because this paper is an *in silico* reanalysis of public data.

Informed Consent Statement: Not applicable because this paper is a public bioinformatic data reanalysis.

Data Availability Statement: All sequence and 3D structure data are available in the GenBank and AlphaFold databases, except for one set of metatranscriptome sequences from natural *C. sinensis* that was uploaded to the repository database www.plantkingdomgdb.com/Ophiocordyceps_sinensis/data/cds/Ophiocordyceps_sinensis_CDS.fas by Xia *et al.* [2017], which is currently inaccessible, but a previously downloaded cDNA file (accessed from 5/18/2017 to 1/18/2018) was used for mating protein analysis.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

- Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung CC, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E, Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Židek A, Bapst V, Kohli P, Jaderberg M, Hassabis D, Jumper JM. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. 2024; 630(8016): 493-500. doi: 10.1038/s41586-024-07487-w
- Barseghyan GS, Holliday JC, Price TC, Madison LM, Wasser SP. Growth and cultural-morphological characteristics of vegetative mycelia of medicinal caterpillar fungus *Ophiocordyceps sinensis* G.H. Sung *et al.* (Ascomycetes) Isolates from Tibetan Plateau (P. R. China). *Intl. J. Med. Mushrooms*. 2011; 13(6): 565–581. DOI: 10.1615/intjmedmushr.v13.i6.90
- Bennett RJ, Johnson AD. Completion of a parasexual cycle in *Candida albicans* by induced chromosome loss in tetraploid strains. *EMBO J*. 2003; 22(10): 2505–2515. DOI: 10.1093/emboj/cdg235

- Bushley KE, Li Y, Wang W-J, Wang X-L, Jiao L, Spatafora JW, Yao Y-J. Isolation of the MAT1-1 mating type idiomorph and evidence for selfing in the Chinese medicinal fungus *Ophiocordyceps sinensis*. Fungal Biol. 2013; 117(9): 599–610. DOI: 10.1016/j.funbio.2013.06.001
- Chen C-S, Hseu R-S, Huang C-T. Quality control of *Cordyceps sinensis* teleomorph, anamorph, and Its products. Chapter 12, in (Shoyama, Y., Ed.) Quality Control of Herbal Medicines and Related Areas. InTech, Rijeka, Croatia (2011). pp. 223–238. www.intechopen.com
- Chen Y-Q, Hu B, Xu F, Zhang W, Zhou H, Qu L-H. Genetic variation of *Cordyceps sinensis*, a fruit-body-producing entomopathogenic species from different geographical regions in China. FEMS Microbiol Lett. 2004; 230: 153–158. Doi: 10.1016/S0378-1097(03)00889-9
- China Ministry of Agriculture and Rural Affairs. Announcement (No. 15 of 2021) of National Forestry and Grassland Administration: List of National Key Protected Wild Plants. September 7, 2021. <https://www.forestry.gov.cn/c/www/lczc/10746.jhtm>
- Dai R-Q, Lan J-L, Chen W-H, Li X-M, Chen Q-T, Shen C-Y. Discovery of a new fungus *Paecilomyces hepiali* Chen & Dai. Acta Agricult. Univ. Pekin. 1989; 15(2): 221–224.
- David A, Islam S, Tankhilevich E, Sternberg MJE. The AlphaFold Database of Protein Structures: A Biologist's Guide. J Mol Biol. 2022; 434(2): 167336. <https://doi.org/10.1016/j.jmb.2021.167336>
- Debuchy R, Turgeo BG. Mating-Type Structure, Evolution, and Function in Euscomycetes. In (eds. Kües U. & Fischer, R.) Growth, Differentiation and Sexuality. Springer (2006). pp. 293–323.
- Deleage G, Roux B. An algorithm for protein secondary structure prediction based on class prediction. Protein Engineering, Design and Selection 1987; 1: 289–294. <https://doi.org/10.1093/protein/1.4.289>
- Du X-H, Wu D-M, Kang H, Wang H-C, Xu N, Li T-T, Chen K-L. Heterothallism and potential hybridization events inferred for twenty-two yellow morel species. IMA Fungus. 2020; 11: 4. <https://doi.org/10.1186/s43008-020-0027-1>
- Engh IB. Molecular phylogeny of the *Cordyceps-Tolypocladium* complex. Candidate scientific thesis, Department of Biology, University of Oslo, 1999.
- Gao L, Li X-H, Zhao J-Q, Lu J-H, Zhao J-G, Zhu J-S. Maturation of *Cordyceps sinensis* associates with alterations of fungal expressions of multiple *Ophiocordyceps sinensis* mutants with transition and transversion point mutations in stroma of *Cordyceps sinensis*. Beijing Da Xue Xue Bao. 2012; 44(3): 454–463. doi: 10.3969/j.issn.1671-167X.2012.03.025
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server. Chapter 52 (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005). pp. 571–607.
- Guo M-Y, Liu Y, Gao Y-H, Jin T, Zhang H-B, Zhou X-W. Identification and bioactive potential of endogenetic fungi isolated from medicinal caterpillar fungus *Ophiocordyceps sinensis* from Tibetan Plateau. Int J Agric Biol. 2017; 19: 307–313. DOI: 10.17957/IJAB/15.0281
- Hawksworth DL, Crous PW, Redhead SA, Reynolds DR, Samson RA, Seifert KA, and 82 other authors. The Amsterdam declaration on fungal nomenclature. IMA Fungus. 2011; 2: 105–112. DOI: 10.5598/ima fungus.2011.02.01.14
- Hénault M, Marsit S, Charron G, Landry CR. The effect of hybridization on transposable element accumulation in an undomesticated fungal species. eLife. 2020; 9: e60474. DOI: <https://doi.org/10.7554/eLife.60474>
- Holliday J, Cleaver M. Medicinal value of the caterpillar fungi species of the genus *Cordyceps* (Fr.) Link (Ascomycetes). A review. Int. J. Med. Mushrooms. 2008; 10: 219–234. DOI: 10.1615/IntJMedMushr.v10.i3.30

- Hu X, Zhang Y-J, Xiao G-H, Zheng P, Xia Y-L, Zhang X-Y, St Leger RJ, Liu X-Z, Wang C-S. Genome survey uncovers the secrets of sex and lifestyle in caterpillar fungus. *Chin. Sci. Bull.* 2013; 58: 2846–2854. doi: 10.1007/s11434-013-5929-5
- Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogeny. *Bioinformat.* 2001; 17: 754–755. DOI: 10.1093/bioinformatics/17.8.754
- Jacobsen S, Wittig M, Pöggeler S. Interaction Between Mating-Type Proteins From the Homothallic Fungus *Sordaria macrospora*. *Curr. Genet.* 2002; 41: 150–158. doi: 10.1007/s00294-002-0276-0
- Jiang Y, Yao Y-J. A review for the debating studies on the anamorph of *Cordyceps sinensis*. *Mycosistema.* 2003; 22(1): 161–176.
- Jin L-Q, Xu Z-W, Zhang B, Yi M, Weng C-Y, Lin S, Wu H, Qin X-T, Xu F, Teng Y, Yuan S-J, Liu Z-Q, Zheng Y-G. Genome sequencing and analysis of fungus *Hirsutella sinensis* isolated from *Ophiocordyceps sinensis*. *AMB Expr.* 2020; 10: 105. <https://doi.org/10.1186/s13568-020-01039-x>
- Jones SK, Bennett RJ. Fungal mating pheromones: choreographing the dating game. *Fungal Genet. Biol.* 2011; 48(7): 668–676. doi: 10.1016/j.fgb.2011.04.001
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021; 596: 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kang Q, Zhang J, Chen F, Dong C, Qin Q, Li X, Wang H, Zhang H and Meng Q. Unveiling mycoviral diversity in *Ophiocordyceps sinensis* through transcriptome analyses. *Front. Microbiol.* 2024; 15: 1493365. doi: 10.3389/fmicb.2024.1493365
- Kinjo N, Zang M. Morphological and phylogenetic studies on *Cordyceps sinensis* distributed in southwestern China. *Mycoscience.* 2001; 42: 567–574. DOI: 10.1007/BF02460956
- Kück U, Bennett RJ, Wang L and Dyer PS. Editorial: Sexual and Parasexual Reproduction of Human Fungal Pathogens. *Front. Cell. Infect. Microbiol.* 2022; 12: 934267. doi: 10.3389/fcimb.2022.934267
- Leung P-H, Zhang Q-X, Wu J-Y. Mycelium cultivation, chemical composition and antitumour activity of a *Tolypocladium* sp. fungus isolated from wild *Cordyceps sinensis*. *J. Appl. Microbiol.* 2006; 101: 275–283. DOI: 10.1111/j.1365-2672.2006.02930.x
- Li C-L. A study of *Tolypocladium sinense* C.L. Li. sp. nov. and cyclosporin production. *Acta Mycol. Sinica.* 1988; 7: 93–98.
- Li M-M, Zhang J-H, Qin Q-L, Zhang H, Li X, Wang H-T, Meng Q. Transcriptome and Metabolome Analyses of *Thitarodes xiaojinensis* in Response to *Ophiocordyceps sinensis* Infection. *Microorganisms* 2023a; 11: 2361. <https://doi.org/10.3390/microorganisms11092361>
- Li X, Wang F, Liu Q, Li Q-P, Qian Z-M, Zhang X-L, Li K, Li W-J, Dong C-H. Developmental transcriptomics of Chinese cordyceps reveals gene regulatory network and expression profiles of sexual development-related genes. *BMC Genomics.* 2019; 20: 337. <https://doi.org/10.1186/s12864-019-5708-z>
- Li X-Z, Li Y-L, Wang Y-N, Zhu J-S. Translations of mutant repetitive genomic sequences in *Hirsutella sinensis* and changes in secondary structures and functional specifications of the encoded proteins. *International J Mol Sci* 2024a; 25(20), 11178; <https://doi.org/10.3390/ijms252011178>
- Li X-Z, Li Y-L, Yao Y-S, Xie W-D, Zhu J-S. Further discussion with Li *et al.* (2013, 2019) regarding the “ITS pseudogene hypothesis” for *Ophiocordyceps sinensis*. *Mol. Phylogenet. Evol.* 2020; 146: 106728. <https://doi.org/10.1016/j.ympev.2019.106728>

- Li X-Z, Li Y-L, Zhu J-S. Differential transcription of mating-type genes during sexual reproduction of natural *Cordyceps sinensis*. Chin. J. Chin. Materia Medica. 2023b; 48(10): 2829–2840. DOI: 10.19540/j.cnki.cjcmm.20230213.102
- Li X-Z, Xiao M-J, Li Y-L, Gao L, Zhu J-S. Mutations and differential transcription of mating-type and pheromone receptor genes in *Hirsutella sinensis* and the natural *Cordyceps sinensis* insect–fungi complex. Biol. (Basel) 2024b; 13(8): 632; <https://doi.org/10.3390/biology13080632>
- Li Y, Hsiang T, Yang R-H, Hu X-D, Wang K, Wang W-J, Wang X-L, Jiao L, Yao Y-J. Comparison of different sequencing and assembly strategies for a repeat-rich fungal genome, *Ophiocordyceps sinensis*. J. Microbiol. Methods. 2016a; 128: 1–6. DOI: 10.1016/j.mimet.2016.06.025
- Li Y, Jiao L, Yao Y-J. Non-concerted ITS evolution in fungi, as revealed from the important medicinal fungus *Ophiocordyceps sinensis*. Mol. Phylogenet. Evol. 2013; 68: 373–379. doi: 10.1016/j.ympev.2013.04.010
- Li Y-L, Gao L, Yao Y-S, Wu Z-M, Lou Z-Q, Xie W-D, Wu J-Y, Zhu J-S. Altered GC- and AT-biased genotypes of *Ophiocordyceps sinensis* in the stromal fertile portions and ascospores of natural *Cordyceps sinensis*. PLoS ONE. 2023c; 18(6): e0286865. <https://doi.org/10.1371/journal.pone.0286865>
- Li Y-L, Li X-Z, Yao Y-S, Wu Z-M, Gao L, Tan N-Z, Lou Z-Q, Xie W-D, Wu J-Y, Zhu J-S. Differential cooccurrence of multiple genotypes of *Ophiocordyceps sinensis* in the stromata, stromal fertile portion (ascocarps) and ascospores of natural *Cordyceps sinensis*. PLoS ONE. 2023d; 18(3): e0270776. DOI: 10.1371/journal.pone.0270776
- Li Y-L, Li X-Z, Yao Y-S, Xie W-D, Zhu J-S. Molecular identification of *Ophiocordyceps sinensis* genotypes and the indiscriminate use of the Latin name for the multiple genotypes and the natural insect-fungi complex. Am J BioMed Sci. 2022b; 14(3): 115–135. doi: 10.5099/aj220300115
- Li Y-L, Yao Y-S, Zhang Z-H, Xu H-F, Liu X, Ma S-L, Wu Z-M, Zhu J-S. Synergy of fungal complexes isolated from the intestines of *Hepialus lagii* larvae in increasing infection potency. J. Fungal Res. 2016b; 14: 96–112.
- Liu J, Guo L-N, Li Z-W, Zhou Z, Li Z, Li Q, Bo X-C, Wang S-Q, Wang J-L, Ma S-C, Zheng J, Yang Y. Genomic analyses reveal evolutionary and geologic context for the plateau fungus *Ophiocordyceps sinensis*. Clin. Med. 2020; 15: 107–119. <https://doi.org/10.1186/s13020-020-00365-3>
- Liu Z-Q, Lin S, Baker PJ, Wu L-F, Wang X-R, Wu H, Xu F, Wang H-Y, Brathwaite ME, Zheng Y-G. Transcriptome sequencing and analysis of the entomopathogenic fungus *Hirsutella sinensis* isolated from *Ophiocordyceps sinensis*. BMC Genomics. 2015; 16: 106–123. DOI: 10.1186/s12864-015-1269-y
- Lu H-L, St Leger RJ. Chapter Seven - Insect Immunity to Entomopathogenic Fungi, Editors: Lovett B, St. Leger RJ. Advanc Genet, Academic Press, 2016. Vol. 94; pp. 251–85.
- Mariani V, Biasini M, Barbato A, Schwede T. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics. 2013; 29(21): 2722–2728. <http://dx.doi.org/10.1093/bioinformatics/btt473>
- Mao X-M, Zhao S-M, Cao L, Yan X, Han R-C. The morphology observation of *Ophiocordyceps sinensis* from different origins. J. Environ. Entomol. 2013; 35(3): 343–353.
- Meng Q, Yu H-Y, Zhang H, Zhu W, Wang M-L, Zhang J-H, Zhou G-L, Li X, Qin Q-L, Hu S-N, Zou Z. Transcriptomic insight into the immune defenses in the ghost moth, *Hepialus xiaojinensis*, during an *Ophiocordyceps sinensis* fungal infection. Insect Biochem. Mol. Biol. 2015; 64: 1–15. DOI: 10.1016/j.ibmb.2015.06.014
- Mishra A, Forche A and Anderson MZ. Parasexuality of Candida Species. Front. Cell. Infect. Microbiol. 2021; 11: 796929. doi: 10.3389/fcimb.2021.796929
- Monzon V, Haft DH, Bateman A. Folding the unfoldable: using AlphaFold to explore spurious proteins. Bioinformatics Advances, 2022; 1(2): vbab043. <https://doi.org/10.1093/bioadv/vbab043>

- Nakamura N, Tanaka C, Takeuchi-Kaneko Y. Transmission of antibiotic-resistance markers by hyphal fusion suggests partial presence of parasexuality in the root endophytic fungus *Glutinomyces brunneus*. *Mycol. Progress* 2019; 18: 453–462. <https://doi.org/10.1007/s11557-018-1455-9>
- Pfennig KS. Facultative Mate Choice Drives Adaptive Hybridization. *Sci.* 2007; 318: 965–967. DOI: 10.1126/science.1146035
- Peters C, Elofsson A. Why is the biological hydrophobicity scale more accurate than earlier experimental hydrophobicity scales? *Proteins*. 2014; 82(9): 2190–2198. doi: 10.1002/prot.24582
- Rams B, Kück U. The *Penicillium chrysogenum* tom1 gene a major target of transcription factor MAT1-1-1 encodes a nuclear protein involved in sporulation. *Front. Fungal Bio.* 2022; 3: 937023. doi: 10.3389/ffunb.2022.937023
- Ren Y, Wan D-G, Lu X-M, Guo J-L. The study of scientific name discussion for TCM Cordyceps. *LisShenzhen Med. Material Medica Res.* 2013; 24(9): 2211–2212.
- Rettie SA, Campbell KV, Bera AK, Kang A, Kozlov S, De La Cruz J, Adebomi V, Zhou G, DiMaio F, Ovchinnikov S, Bhardwaj G. Cyclic peptide structure prediction and design using AlphaFold. *bioRxiv [Preprint]*. 2023 Feb 26:2023.02.25.529956. doi: 10.1101/2023.02.25.529956
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space, *Systematic Biology*, 2012; 61(3): 539–542. <https://doi.org/10.1093/sysbio/sys029>
- Samarasinghe H, You M, Jenkinson TS, Xu J-P, James TY. Hybridization Facilitates Adaptive Evolution in Two Major Fungal Pathogens. *Genes*. 2020; 11: 101. doi:10.3390/genes11010101
- Seervai RNH, Jones SK, Hirakawa MP, Porman AM, Bennett RJ. Parasexuality and ploidy change in *Candida tropicalis*. *Eukaryotic Cell*. 2013; 12(12): 1629–1640. DOI: 10.1128/EC.00128-13
- Sherwood RK, Bennett RJ. Fungal meiosis and parasexual reproduction—lessons from pathogenic yeast. *Current Opinion Microbiol.* 2009; 12(6): 599–607. DOI: 10.1016/j.mib.2009.09.005
- Shu R-H, Zhang J-H, Meng Q, Zhang H, Zhou G-L, Li M-M, Wu P-P, Zhao Y-N, Chen C, Qin Q-L. A new high-quality draft genome assembly of the Chinese cordyceps *Ophiocordyceps sinensis*. *Genome Biol. Evol.* 2020; 12(7): 1074–1079. DOI: 10.1093/gbe/evaa112
- Simm S, Einloft J, Mirus O, Schleiff E. 50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification. *Biol. Res.* 2016; 49(1): 31. doi: 10.1186/s40659-016-0092-5
- Steensels J, Gallone B, Verstrepen KJ. Interspecific hybridization as a driver of fungal evolution and Adaptation. *Nat. Rev. Microbiol.* 2021; 19: 485–500.
- Stensrud Ø, Hywel-Jones NL, Schumacher T. Towards a phylogenetic classification of Cordyceps: ITS nrDNA sequence data confirm divergent lineages and paraphyly. *Mycol. Res.* 2005; 109: 41–56. DOI: 10.1017/s095375620400139x
- Stensrud Ø, Schumacher T, Shalchian-Tabrizi K, Svegardenib IB, Kauserud H. Accelerated nrDNA evolution and profound AT bias in the medicinal fungus *Cordyceps sinensis*. *Mycol. Res.* 2007; 111: 409–415. DOI: 10.1016/j.mycres.2007.01.015
- Stone R. Improbable partners aim to bring biotechnology to a Himalayan kingdom. *Sci.* 2010; 327: 940–941. DOI: 10.1126/science.327.5968.940
- Sung G-H, Hywel-Jones NL, Sung J-M, Luangsa-ard JJ, Shrestha B, Spatafora JW. Phylogenetic classification of *Cordyceps* and the clavicipitaceous fungi. *Stud. Mycol.* 2007; 57: 5–59. DOI: 10.3114/sim.2007.57.01

- Tunyasuvunakool, K., Adler, J., Wu, Z, Green T, Zielinski M, Židek A, Bridgland A, Cowie A, Meyer C, Laydon A, Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA, Potapenko A, Ballard AJ, Romera-Paredes B, Nikolov S, Jain R, Clancy E, Reiman D, Petersen S, Senior AW, Kavukcuoglu K, Birney E, Kohli P, Jumper J, Hassabis D. Highly accurate protein structure prediction for the human proteome. *Nature* 2021; 596: 590–596. <https://doi.org/10.1038/s41586-021-03828-1>
- Turgeon BG, Yoder OC. Proposed nomenclature for mating type genes of filamentous ascomycetes. *Fungal Genet. Biol.* 2000; 31: 1–5. DOI: 10.1006/fgbi.2000.1227
- Varadi M, Bertoni D, Magana P, Paramval U, Pidruchna I, Radhakrishnan M, Tsenkov M, Nair S, Mirdita M, Yeo J, Kovalevskiy O, Tunyasuvunakool K, Laydon A, Židek A, Tomlinson H, Hariharan D, Abrahamson J, Green T, Jumper J, Birney E, Steinegger M, Hassabis D, Velankar S. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* 2024; 52(D1): D368–D375. doi: 10.1093/nar/gkad1011
- Wang Y, Stata M, Wang W, Stajich JE, White MM, Moncalvo JM. Comparative genomics reveals the core gene toolbox for the fungus-insect symbiosis. *mBio.* 2018; 9: e00636-18. DOI: 10.1128/mBio.00636-18
- Wang Y-B, Wang Y, Fan Q, Duan D-E, Zhang G-D, Dai R-Q, Dai Y-D, Zeng W-B, Chen Z-H, Li D-D, Tang D-X, Xu Z-H, Sun T, Nguyen T-T, Tran N-L, Dao V-M, Zhang C-M, Huang L-D, Liu Y-J, Zhang X-M, Yang D-R, Sanjuan T, Liu X-Z, Yang Z-L, Yu H. Multigene phylogeny of the family Cordycipitaceae (Hypocreales): new taxa and the new systematic position of the Chinese cordycipitoid fungus *Paecilomyces hepiali*. *Fungal Divers* 2020; 103: 1. <https://doi.org/10.1007/s13225-020-00457-3>
- Wei J-C, Wei X-L, Zheng W-F, Guo W, Liu R-D. Species identification and component detection of *Ophiocordyceps sinensis* cultivated by modern industry. *Mycosystema.* 2016; 35(4): 404–410.
- Wei X-L, Yin X-C, Guo Y-L, Shen N-Y, Wei, J.-C. Analyses of molecular systematics on *Cordyceps sinensis* and its related taxa. *Mycosystema.* 2006; 25(2): 192–202.
- Wilson AM, Wilken PM, van der Nest MA, Steenkamp ET, Wingfield MJ, Wingfield BD. Homothallism: an umbrella term for describing diverse sexual behaviours. *IMA Fungus.* 2015; 6(1): 207–214. DOI: 10.5598/ima fungus.2015.06.01.13
- Wroblewski K, Kmiecik S. Integrating AlphaFold pLDDT Scores into CABS-flex for enhanced protein flexibility simulations. *Comput Struct Biotechnol J.* 2024; 30(23): 4350–4356. doi: 10.1016/j.csbj.2024.11.047
- Xia E-H, Yang D-R, Jiang J-J, Zhang Q-J, Liu Y, Liu Y-L, Zhang Y, Zhang H-B, Shi C, Tong Y, Kim C-H, Chen H, Peng Y-Q, Yu Y, Zhang W, Eichler EE, Gao L-Z. The caterpillar fungus, *Ophiocordyceps sinensis*, genome provides insights into highland adaptation of fungal pathogenicity. *Sci. Rep.* 2017; 7: 1806. DOI:10.1038/s41598-017-01869-z
- Xia F, Liu Y, Shen G-L, Guo L-X, Zhou X-W. Investigation and analysis of microbiological communities in natural *Ophiocordyceps sinensis*. *Can. J. Microbiol.* 2015; 61: 104–111. DOI: 10.1139/cjm-2014-0610
- Xiang L, Li Y, Zhu Y, Luo H, Li C, Xu X, Sun C, Song J-Y, Shi L-H, He L, Sun W, Chen S-L. Transcriptome analysis of the *Ophiocordyceps sinensis* fruiting body reveals putative genes involved in fruiting body development and cordycepin biosynthesis. *Genomics.* 2014; 103: 154–159. <https://doi.org/10.1016/j.ygeno.2014.01.002>
- Xiao W, Yang J-P, Zhu P, Cheng K-D, He H-X, Zhu H-X, Wang Q. Non-support of species complex hypothesis of *Cordyceps sinensis* by targeted rDNA-ITS sequence analysis. *Mycosystema.* 2009; 28(6): 724–730.
- Xu T, Xu Q, Li J-Y. Toward the appropriate interpretation of AlphaFold2. *Front Artif Intell.* 2023; 6: 1149748. doi: 10.3389/frai.2023.1149748
- Yang J-L, Xiao W, He H-X, Zhu H-X, Wang S-F, Cheng -KD, Zhu P. Molecular phylogenetic analysis of *Paecilomyces hepiali* and *Cordyceps sinensis*. *Acta Pharmaceut. Sinica* 2008; 43(4): 421–426.

- Yang J-Y, Tong X-X, He C-Y, Bai J, Wang F, Guo J-L. Comparison of endogenetic microbial community diversity between wild *Cordyceps sinensis*, artificial *C. sinensis* and habitat soil. *Chin. J. Chin. Materia Medica*. 2021; 46(12): 3106–3115.
- Yao Y-S, Zhu J-S. Indiscriminate use of the Latin name for natural *Cordyceps sinensis* and *Ophiocordyceps sinensis* fungi. *Chin. J. Chin. Mater. Med.* 2016; 41(7): 1316–1366.
- Zhang S; Zhang Y-J. Molecular evolution of three protein-coding genes in the Chinese caterpillar fungus *Ophiocordyceps sinensis*. *Microbiol. China*. 2015; 42(8): 1549–1560.
- Zhang S, Zhang Y-J, Liu X-Z, Wen H-A, Wang M, Liu D-S. Cloning and analysis of the MAT1-2-1 gene from the traditional Chinese medicinal fungus *Ophiocordyceps sinensis*. *Fungal Biol.* 2011; 115: 708–714.
- Zhang S, Zhang Y-J, Shrestha B, Xu J-P, Wang C-S, Liu X-Z. *Ophiocordyceps sinensis* and *Cordyceps militaris*: research advances, issues and perspectives. *Mycosystema*. 2013; 32: 577–597.
- Zhang S-W, Cen K, Liu Y, Zhou X-W, Wang C-S. Metatranscriptomics analysis of the fruiting caterpillar fungus collected from the Qinghai-Tibetan plateau. *Sci. Sinica Vitae*. 2018; 48(5): 562–570.
- Zhang Y-J, Li E-W, Wang C-S, Li Y-L, Liu X-Z. *Ophiocordyceps sinensis*, the flagship fungus of China: terminology, life strategy and ecology. *Mycol.* 2012; 3(1): 2–10. <https://www.tandfonline.com/doi/full/10.1080/21501203.2011.654354>
- Zhang Y-J, Sun B-D, Zhang S, Wàngmǔ, Liu X-Z, Gong W-F. Mycobiotal investigation of natural *Ophiocordyceps sinensis* based on culture-dependent investigation. *Mycosistema*. 2010; 29(4): 518–527.
- Zhang Y-J, Xu L-L, Zhang S, Liu X-Z, An Z-Q, Wàngmǔ, Guo Y-L. Genetic diversity of *Ophiocordyceps sinensis*, a medicinal fungus endemic to the Tibetan Plateau: implications for its evolution and conservation. *BMC Evol. Biol.* 2009; 9: 290. doi:10.1186/1471-2148-9-290
- Zhang Y-J, Zhang S, Li Y-L, Ma S-L, Wang C-S, Xiang M-C, Liu X, An Z-Q, Xu J-P, Liu X-Z. Phylogeography and evolution of a fungal–insect association on the Tibetan Plateau. *Mol. Ecol.* 2014; 23: 5337–5355. DOI: 10.1111/mec.12940
- Zhao Y-N, Zhang J-H, Meng Q, Zhang H, Zhou G-L, Li M-M, Wu P-P, Shu R-H, Gao X-X, Guo L, Tong Y, Cheng L-Q, Guo L, Chen C, Qin Q. Transcriptomic analysis of the orchestrated molecular mechanisms underlying fruiting body initiation in Chinese cordyceps. *Gene*. 2020; 763: 145061. <https://doi.org/10.1016/j.gene.2020.145061>
- Zheng, P.; Wang, C.-S. Sexuality Control and Sex Evolution in Fungi. *Sci. Sin. Vitae* 2013, 43, 1090–1097.
- Zhong X, Gu L, Wang H-Z, Lian D-H, Zheng Y-M, Zhou S, Zhou W, Gu J, Zhang G, Liu X. Profile of *Ophiocordyceps sinensis* transcriptome and differentially expressed genes in three different mycelia, sclerotium and fruiting body developmental stages. *Fungal Biol.* 2018; 122: 943–951. DOI: 10.1016/j.funbio.2018.05.011
- Zhu J-S, Gao L, Li X-H, Yao Y-S, Zhou Y-J, Zhao J-Q, Zhou Y-J. Maturation alterations of oppositely orientated rDNA and differential proliferations of CG:AT-biased genotypes of *Cordyceps sinensis* fungi and *Paecilomyces hepiali* in natural *C. sinensis*. *Am. J. Biomed. Sci.* 2010; 2(3): 217–238. doi: 10.5099/aj100300217
- Zhu J-S, Guo Y-L, Yao Y-S, Zhou Y-J, Lu J-H, Qi Y, Chen W, Zheng T-Y, Zhang L, Wu Z-M, Zhang L-J, Liu X-J, Yin W-T. Maturation of *Cordyceps sinensis* associates with co-existence of *Hirsutella sinensis* and *Paecilomyces hepiali* DNA and dynamic changes in fungal competitive proliferation predominance and chemical profiles. *J. Fungal Res.* 2007; 5(4): 214–224.
- Zhu J-S, Gray GM. Renaturative catalytic blotting of enzyme proteins. Chapter 17. in Dunbar BS (ed.), *Protein blotting: A practical approach (IRL series)*. Oxford University Press, Oxford, 1994, pp. 221–238. <https://doi.org/10.1093/oso/9780199634385.003.0017>

- Zhu J-S, Halpern GM, Jones K. The scientific rediscovery of a precious ancient Chinese herbal regimen: *Cordyceps sinensis*: Part I. J. Altern. Complem. Med. 1998a; 4(3): 289–303. DOI: 10.1089/acm.1998.4.3-289
- Zhu J-S, Halpern GM, Jones K. The scientific rediscovery of an ancient Chinese herbal medicine: *Cordyceps sinensis*: Part II. J. Altern. Complem. Med. 1998b; 4(4): 429–457. DOI: 10.1089/acm.1998.4.429
- Zhu J-S, Li C-L, Tan N-Z, Berger JL, Prolla TA. Combined use of whole-gene expression profiling technology and mouse lifespan test in anti-aging herbal product study. Proc. 2011 New TCM Products Innovation and Industrial Development Summit, Hangzhou, China (Nov 27, 2011). pp. 443–448. https://xueshu.baidu.com/usercenter/paper/show?paperid=08341c17fa58c8f85584b92572b90f75&site=xueshu_se
- Zhu J-S, Li Y-L. A Precious Transitional Chinese Medicine, *Cordyceps sinensis*: Multiple heterogeneous *Ophiocordyceps sinensis* in the insect-fungi complex. Lambert Academic Publishing, Saarbrücken. Germany, 2017.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.