

Article

Altered stereostructures of the DNA-binding domains in the mutant mating proteins of *Ophiocordyceps sinensis* and the wild insect–fungal complex

Xiu-Zhang Li ¹, Yu-Ling Li ¹, Wei Liu ² and Jia-Shi Zhu ^{1,*}

¹ State Key Laboratory of Plateau Ecology and Agriculture, Qinghai Academy of Animal and Veterinary Sciences, Qinghai University, Xining, Qinghai 810016, China; xiuzhang11@163.com, yulingli2000@163.com, zhujoah@163.com

² Institute of Immunology, Army Medical University, Chongqing, 400038, China; weiliu@tmmu.edu.cn

* Correspondence: zhujoah@163.com

Simple Summary

The natural *Cordyceps sinensis* insect–fungal complex, comprising the *Ophiocordyceps sinensis* fruiting body and the remains of a *Hepialidae* moth larva, is a highly valued therapeutic agent in traditional Chinese medicine. The sexual reproduction of *O. sinensis* is ultimately controlled by the tertiary structures of functional domains within the MAT1-1-1 and MAT1-2-1 proteins. This study reveals the primary structures of MAT1-1-1 and MAT1-2-1 protein variants derived from numerous wild-type *C. sinensis* isolates. The protein sequences exhibited various amino acid substitutions in the functional domains and clustered into several Bayesian clades correlated with the secondary and tertiary structures of the critical function domains. In combination with data on the alternative splicing and differential occurrence and transcription of the *MAT1-1-1* and *MAT1-2-1* genes, the altered 3D structures of the mating protein functional domains refute the hypothesis that *O. sinensis* can self-fertilize via homothallic mating and instead suggest that *O. sinensis* is self-sterile and requires a mating partner for sexual reproduction under heterothallism or hybridization during the sexual life of natural *C. sinensis* insect–fungal complex. This fundamental paradigm shift provides an accurate theoretical basis for understanding *O. sinensis* and the *C. sinensis* insect–fungal complex and will guide future academic and applied studies.

Abstract

The MAT α _HMGbox and HMG-box_ROX1-like domains of the MAT1-1-1 and MAT1-2-1 proteins, respectively, play essential roles in DNA binding and the regulation of gene transcription controlling *Ophiocordyceps sinensis* sexual reproduction. Alternative splicing, differential occurrence/transcription of the *MAT1-1-1* and *MAT1-2-1* genes have been revealed in *Hirsutella sinensis* (Genotype #1 of the 17 *O. sinensis* genotypes), suggesting self-sterility under heterothallic or hybrid outcrossing. In this study, the MAT α _HMGbox domains of MAT1-1-1 proteins in wild-type *Cordyceps sinensis* isolates cluster into 5 clades in the Bayesian clustering tree and belong to diverse stereostructure morphs under 19 AlphaFold codes. The HMG-box_ROX1-like domains of MAT1-2-1 proteins, on the other hand, cluster into 2 branched clades and belong to stereostructure morphs under 25 AlphaFold codes. Correlation analysis revealed that 1–4 amino acid substitutions in the DNA-binding domains of the mating proteins resulted in the altered hydrophobicity and secondary and tertiary structures of the DNA-binding domains of the proteins. Fungal

Academic Editor: Firstname Last-name

Received: date

Revised: date

Accepted: date

Published: date

Citation: To be added by editorial staff during production.

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

origin analysis revealed heterospecific fungal sources of mating proteins with
stereostructure variations in wild-type *C. sinensis* isolates that represented changed func-
tionalities involved in ensuring the accuracy and genetic diversity of the heterothallic and
hybrid reproduction of *O. sinensis* during the lifecycle of the *C. sinensis* insect–fungi com-
plex.

Keywords: MAT α _HMGbox domain of MAT1-1-1 protein; HMG-box_ROX1-like domain
of MAT1-2-1 protein; heteromorphic protein structures; wild-type *Cordyceps sinensis* iso-
lates; sexual reproduction of *Ophiocordyceps sinensis*; sexual life of the *Cordyceps sinensis*
insect–fungal complex

1. Introduction

Natural *Cordyceps sinensis* is one of most expensive therapeutic agents in traditional Chinese medicine and has a rich history of clinical use for several centuries for health maintenance, disease amelioration, post-illness and postoperative recovery, and antiaging therapy [1–3]. This natural therapeutic agent consists of the *Ophiocordyceps sinensis* fruiting body and the remains of a *Hepialidae* moth larva containing an intact, thick larval body wall with numerous bristles; an intact larval intestine; head tissues; and fragments of other larval tissues [4–10]. Studies of natural *C. sinensis* have demonstrated its multicellular heterokaryotic structures and genetic heterogeneity including 17 genomically independent genotypes of *O. sinensis* fungi, >90 other fungal species spanning at least 37 fungal genera, and larval genes [4,7,10–29]. Thus, the Chinese Pharmacopoeia defines natural *C. sinensis* as an insect–fungal complex classified as a LEVEL-II endangered natural species [30]. Among the numerous heterogeneous fungal species [15,21,26], *Hirsutella sinensis* has been postulated to be the sole anamorph of *O. sinensis* by Wei *et al.* [31]. However, ten years later, in an artificial cultivation study conducted in an industrial product-oriented setting, Wei *et al.* [32] reported a species contradiction between anamorphic inoculates of GC-biased Genotype #1 *H. sinensis* strains on *Hepialidae* moth larvae and the sole teleomorph (the genomically independent AT-biased Genotype #4 of *O. sinensis*) in the fruiting body of cultivated *C. sinensis*. Notably, the Latin name *Cordyceps sinensis* has been indiscriminately used since the 1840s to refer to both the teleomorph/holomorph of the fungus *C. sinensis* and the wild insect–fungal complex; the fungus was renamed *O. sinensis* in 2007, with the use of the *H. sinensis* strain EFCC 7287 as the nomenclature reference [4,7,33–35]. Although Zhang *et al.* [36] proposed improper implementation of the “One Fungus=One Name” nomenclature rule set by the International Mycological Association [37], 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*, in this work, we continue to use the anamorphic name *H. sinensis* for GC-biased Genotype #1 of the 17 *O. sinensis* genotypes that may share the same evolutionary ancestor [17] and refer to the genomically independent Genotypes #2–17 fungi as *O. sinensis* following the taxonomic descriptions that are used in several public depository databases, such as the GenBank and AlphaFold databases, before the systematic positions of these evolutionarily related genotypes are individually determined and differentiated, regardless of whether they are GC- or AT-biased genetically. In this study, 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 using the *H. sinensis* strain EFCC 7287 as the nomenclature reference did not involve the indiscriminately used Latin name for the natural insect–fungal complex [7,33]. However, this practice will likely be revised 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–fungal complex.

The sexual reproductive behavior of ascomycetes is strictly regulated by transcription factors encoded at the mating-type (*MAT*) locus. These transcription factors constitute the core mechanism that determines mating compatibility, regulates mating type recognition, and controls the development of fruiting bodies and the reproductive structures ascocarps and ascospores [38–46]. Like other fungi in Ascomycota, the reproductive behavior of *O. sinensis* relies on the synergistic interaction of 2 mating proteins, the *MAT1-1* and *MAT1-2* idiomorphs. These proteins contain 2 types of critical domains that specifically regulate gene expression related to sexual reproduction: (1) the mating-type alpha high mobility group box (*MAT* α _HMGbox) domain in the *MAT1-1-1* protein and (2) the high mobility group box ROX1-like (HMG-box_ROX1-like) domain in the *MAT1-2-1* protein [41,47–50]. However, alternative splicing, differential occurrence, and differential

transcription of mating-type and pheromone receptor genes and heteromorphic stereostructures of the entire MAT1-1-1 and MAT1-2-1 proteins in *H. sinensis* strains and wild-type *C. sinensis* isolates have been demonstrated, invalidating the self-fertilization hypothesis under homothallism and pseudohomothallism for *H. sinensis*, which was postulated as the sole anamorph of *O. sinensis*, and instead suggesting that self-sterility under heterothallism or hybridization requires sexual partners to accomplish sexual reproduction within the lifecycle of the *C. sinensis* insect–fungi complex [31,50–55].

Mutations of the 2 critical DNA-binding domains of mating proteins that result in a few amino acid substitutions may cause conformational changes in the core stereostructures of the domains, thereby affecting the synergistic interaction of the proteins, probably through the regulation of the expression of complementary downstream genes of the 2 idiomorphic mating proteins. Continuing our previous study [50], which demonstrated variations in the 3D structures of complete mating proteins, this study focuses on amino acid substitutions within the MAT α _HMGbox domain and the HMGbox_ROX1-like domain of the full-length MAT1-1-1 and MAT1-2-1 proteins, respectively, and correlates the variable primary structures of these domains with changes in the hydrophobic properties and the secondary and tertiary structures of the DNA-binding domains in wild-type *C. sinensis* isolates. Correlations between changes in hydrophobicity and the primary and secondary structures of the DNA-binding domains of mating proteins encoded by the genome, transcriptome and metatranscriptome assemblies of *H. sinensis* and *C. sinensis* insect–fungal complexes were also analyzed.

2. Materials and Methods

2.1. The MAT1-1-1 and MAT1-2-1 protein sequences of the wild-type *C. sinensis* isolates

The AlphaFold database lists the AlphaFold-predicted 3D structures for 138 MAT1-1-1 proteins and 74 MAT1-2-1 proteins [50]. Of these proteins, 118 MAT1-1-1 proteins and 69 MAT1-2-1 proteins are full-length proteins, as shown in Tables S1–S2. Three full-length MAT1-1-1 proteins (ALH24945, AGW27560, and EQK97643) and 2 full-length MAT1-2-1 proteins (AEH27625 and EQL04085) ~~are~~ derived from *H. sinensis* strains. All other full-length mating proteins ~~are~~ from wild-type *C. sinensis* isolates that were collected from various production areas on the Qinghai–Tibet Plateau [5,34,36,49–50,52,56–57]. The internal transcribed spacer (ITS) genotypic information is available in the GenBank database for the mutant full-length MAT1-1-1 and MAT1-2-1 proteins that ~~are~~ derived from 34 wild-type *C. sinensis* isolates and aligned with the reference ITS sequences of the 17 genotypes of *O. sinensis* (cf. Table S3).

The remaining 20 MAT1-1-1 proteins (AGW27517–AGW27536) and 5 MAT1-2-1 proteins (AGW27543, AGW27548, AGW27552, AGW27554, and AGW27555) ~~are~~ truncated and will be analyzed elsewhere [23,48].

2.2. Genome, transcriptome, and metatranscriptome assemblies of *H. sinensis* strains and *C. sinensis* insect–fungal complexes

The GenBank database lists 5 genome assemblies, LKHE000000000, NGJJ000000000, ANOV000000000, JAAVMX000000000, and LWBQ000000000, of the *H. sinensis* strains 1229, CC1406-20395, Co18, IOZ07, and ZJB12195, respectively, and a transcriptome assembly, GCQL000000000, of the *H. sinensis* strain L0106 [49,59–63].

The metatranscriptome assembly GAGW000000000 for the *C. sinensis* insect–fungal samples (unknown maturation stages) collected from Kangding County, Sichuan Province, China [64], is available in the GenBank database.

Another metatranscriptome assembly ~~was~~ derived from mature *C. sinensis* samples collected from Deqin, Yunnan Province, China, and was uploaded to a repository

database, www.plantkingdomgdb.com/Ophiocordyceps_sinensis/data/cds/Ophiocordyceps_sinensis_CDS.fas [65]. Although this database is currently inaccessible, it was accessed from 18 May 2017 to 18 January 2018, and a cDNA file was downloaded.

The genome, transcriptome, and metatranscriptome assemblies were used to analyze the MAT α _HMGbox and HMG-box_ROX1-like domains of the MAT1-1-1 and MAT1-2-1 proteins, respectively.

2.3. Alignment of the DNA-binding domain sequences of the mating proteins

The amino acid sequences of the MAT α _HMGbox domains of the MAT1-1-1 proteins and the HMG-box_ROX1-like domains of the MAT1-2-1 proteins of the wild-type *C. sinensis* isolates and *C. sinensis* insect–fungal complexes were aligned via the GenBank Blastp program (<https://blast.ncbi.nlm.nih.gov/> (Bethesda, MD, USA), accessed from 18 October 2024 to 10 July 2025).

2.4. Bayesian clustering analysis of the DNA-binding domains of the mating proteins

Multiple sequence alignment of the DNA-binding domains of the mating proteins of the wild-type *C. sinensis* isolates and *C. sinensis* insect–fungal complexes was performed via the auto mode of MAFFT (v7.427; Osaka, Japan). The intron regions were removed via trimAl (v1.4. rev15; Barcelona, Spain). Under the Bayesian information criterion, ProtTest v3.4.2 (University of Vigo and University of La Coruña, Spain) was used to find the optimal amino acid substitution model. Bayesian clustering trees of the DNA binding domain sequences of the MAT1-1-1 and MAT1-2-1 proteins were then inferred via MrBayes v3.2.7 software (Markov chain Monte Carlo [MCMC] algorithm; Uppsala, Sweden) at a sampling frequency of 100 iterations after discarding the initial 25% of the samples from a total of 1 million iterations. This yielded final majority rule consensus trees for the MAT α _HMGbox domains of the full-length MAT1-1-1 proteins and the HMG-box_ROX1-like domains of the full-length MAT1-2-1 proteins [7,50,54–55,66]. Clustering analysis was conducted at Nanjing Genepioneer Biotechnologies Co. (Nanjing, China).

2.5. Amino acid properties and scale analysis

The amino acid sequences of the DNA-binding domains of the mating proteins were scaled on the basis of the general chemical characteristics of their side chains (Table S4) as reported at <https://web.expasy.org/protscale/> (Basel, Switzerland), accessed from 18 October 2024 to 20 May 2025 [50,55,68–72]. The MAT α _HMGbox domain of MAT1-1 proteins (amino acids 51→225 of the ~~reference~~ sequence AGW27560 derived from the *H. sinensis* strain CS68-2-1229 [48]) and the HMG-box_ROX1-like domain of MAT1-2-1 proteins (amino acids 127→197 of the ~~reference~~ sequence AEH27625 derived from the *H. sinensis* strain CS2 [56]), with an additional 9 amino acid residues extending upstream and downstream of the domains, were plotted sequentially with a window size of 9 amino acid residues via the linear weight variation model of the ExPASy ProtScale algorithms [50,55,68–72] to generate ExPASy ProtScale plots to display and compare the topology and waveform changes in hydrophobicity and the 2D structures for α -helices, β -sheets, β -turns, and coils of the DNA-binding domains of the mating proteins.

H-6 AlphaFold-based prediction of the 3D structures of the mating proteins

For the heteromorphous stereostructure analysis in this study, the 3D structures of the MAT1-1-1 and MAT1-2-1 proteins of the wild-type *C. sinensis* isolates were computationally predicted from their amino acid sequences by using the artificial intelligence (AI)-based machine learning technology AlphaFold (<https://alphafold.com/>, Cambridgeshire, UK), which was downloaded from the AlphaFold database (accessed from 18 October 2024 to 10 May 2025) [50,73–82].

The AlphaFold database provides per-residue model confidence, predicts scores between 0 and 100 on the local distance difference test (pLDDT) and provides a per-residue score that is assigned to each individual residue [73–76,78–79]. The model confidence bands are used to color-code the residues in the 3D structures: residues with very high confidence (pLDDT>90) are shown in dark blue, those with high confidence (90>pLDDT>70) appear in light blue, residues with low confidence (70>pLDDT>50) are shown in yellow, and residues with very low confidence (pLDDT<50) are shown in orange [50,80,83]. The AlphaFold database provides an average pLDDT score for each of the AI-predicted 3D structural models.

2.7. Correlation of the primary, secondary, and tertiary structures of the DNA-binding domains of the MAT1-1-1 and MAT1-2-1 proteins

The AlphaFold-predicted 3D structures of the MAT α _HMGbox domains of the MAT1-1-1 proteins and the HMG-box_ROX1-like domains of the MAT1-2-1 proteins were amplified at the sites of amino acid substitutions. The locally magnified 3D structures at the mutation sites were correlated with the amino acid substitutions to obtain the primary structures and the changes in topology and waveform via ExPASy ProtScale plotting, which revealed hydrophathy, α -helices, β -sheets, β -turns, and coils for altered hydrophobicity and the 2D structures of the DNA binding domains of the MAT1-1-1 and MAT1-2-1 proteins.

3. Results

3.1. Primary structures of the MAT α _HMGbox domains of the full-length MAT1-1-1 proteins

Of the 138 MAT1-1-1 proteins in the AlphaFold database, 118 (85.5%) are full-length. Each of these proteins contains 372 amino acid residues. These proteins contribute to 15 heteromorphous 3D structure morphs belonging to 5 Bayesian clusters (Table S1) [50]. Twenty-nine (24.6%) of the 118 full-length MAT1-1-1 proteins derived from wild-type *C. sinensis* isolates contain variable amino acid substitutions.

The MAT1-1-1 proteins contain a MAT α _HMGbox domain that is located at amino acids 51→225 (shown in blue and underlined in Figure S1), as represented by the reference MAT1-1-1 protein AGW27560 derived from the *H. sinensis* strain CS68-2-1229 [48]. Other MAT1-1-1 proteins derived from wild-type *C. sinensis* isolates or encoded by the genome assemblies of the *H. sinensis* strains and the metatranscriptome assemblies of *C. sinensis* insect–fungal complexes contain various amino acid substitutions, as shown in red in Figure S1. The MAT α _HMGbox domains of 10 (34.5%) of the 29 mutant proteins are 100% identical to those of the query protein AGW27560. Each of the remaining 19 proteins (65.5%) contains a mutation or mutations that results in 1 or 2 substitutions in the amino acid residues at various sites within the MAT α _HMGbox domains.

Figure S1 also shows the MAT α _HMGbox domains of the MAT1-1-1 proteins encoded by the genome assemblies ANOV01017390 (794←1129 & 1280←1396), LKHE01001116 (4183←4620 & 4667←4759), and JAAVMX010000001 (6,699,061→6,699,153 & 6,699,203→6,699,637) of the *H. sinensis* strains Co18, 1229, and IOZ07, respectively [49,61,63]. (Note: the arrows “→” and “←” indicate sequences in the sense and antisense strands of the genomes, respectively; “&” refers to the removed intron portion.) These genome-encoded MAT1-1-1 proteins are truncated at their C-termini; the truncation segments are located far downstream of the MAT α _HMGbox domains.

The genome assemblies LWBQ00000000 and NGJJ00000000 and the transcriptome assembly GCQL00000000 of the *H. sinensis* strains ZJB12195, CC1406-20395, and L0106, respectively, do not contain the genes or transcripts that encode the MAT1-1-1 proteins [59–61].

The metatranscriptome assemblies GAGW01008880 (300←1127) and OSIN7648 (1→1065) of natural *C. sinensis* insect–fungal complexes encode the MAT1-1-1 proteins (cf. Figure S1) [64–65]. The MAT1-1-1 protein OSIN7648 is truncated in the middle sequence by 18 amino acid residues (SMQREYQAPRFFYDYSVS) between residues 863 and 864; the truncation segment is located downstream of the MAT α _HMGbox domain (151→675). The N-terminus of the MAT α _HMGbox domain (741←1127) of the MAT1-1-1 protein encoded by GAGW01008880 is truncated by 46 amino acid residues (AAASRATRQT-KEASCDRAKRPLNAFMAFRSYLKLFPDVQQKTASG).

3.2. Primary structures of the HMG-box_ROX1-like domains of the full-length MAT1-2-1 proteins

Of the 74 MAT1-2-1 proteins listed in the AlphaFold database, 69 (93.2%) are full-length proteins, each of which contains 249 amino acid residues; these proteins contribute to 17 heteromorphous AlphaFold tertiary structure morphs belonging to 5 Bayesian clusters (Table S2) [50]. Of the 69 full-length proteins, 39 (56.5%) are 100% identical to the representative MAT1-2-1 protein AEH27625, which was derived from the *H. sinensis* strain CS2 [56]. The remaining 30 full-length proteins (43.5%) contain amino acid substitutions at various sites, as shown in red in Figure S2. The HMG-box_ROX1-like domains of the MAT1-2-1 proteins are located at amino acid residues 127→197, as shown in blue and underlined in Figure S2, whereas the 9 external amino acid residues upstream and downstream of the domain are shown in blue but not underlined. The HMG-box_ROX1-like domain sequences of 3 of the 30 variable MAT1-2-1 proteins are 100% identical to the domain sequence of the query protein AEH27625, whereas the domains of the remaining 27 proteins contain 1–3 amino acid substitutions, as shown in red in Figure S2.

Figure S2 also shows the MAT1-2-1 proteins encoded by the genome assemblies ANOV01000063, LKHE01001605, LWBQ01000021, and NGJJ01000619 of the *H. sinensis* strains Co18, 1229, ZJB12195, and CC1406-20395, respectively [49,60–62]. An S-to-A substitution occurred in each of the HMG-box_ROX1-like domains of the genome-encoded MAT1-2-1 proteins ANOV01000063 (9759→9851 & 9907→10,026), LKHE01001605 (14,016←14,135 & 14,191←14,283), LWBQ01000021 (239,029←239,148 & 239,204←239,269), and NGJJ01000619 (23,186←23,305 & 23,361←23,453). An additional Y-to-H substitution occurred in the HMG-box_ROX1-like domains of LKHE01001605, LWBQ01000021 and NGJJ01000619 but not in the HMG-box_ROX1-like domain of ANOV01000063, resulting in 97.2–98.6% similarity to the sequence of the HMG-box_ROX1-like domain of the query protein AEH27625 [56].

The genome assembly JAAVMX000000000 of the *H. sinensis* strain IOZ07 and the metatranscriptome assembly GAGW000000000 of the *C. sinensis* insect–fungi complex do not contain genes or transcripts encoding MAT1-2-1 proteins [63–64].

Figure S2 also shows the protein sequences encoded by the MAT1-2-1 transcripts of the transcriptome assembly GCQL01020543 (397←1143) of the *H. sinensis* strain L0106 and the metatranscriptome assembly OSIN7649 (1→747) of the mature *C. sinensis* insect–fungal complex [59,65]. The HMG-box_ROX1-like domains (553←765 of GCQL01020543 and 379→591 of OSIN7649) of these proteins contain Y-to-H substitutions, resulting in 98.6% similarity to the sequence of that domain in the query protein AEH27625 [36].

3.3. Bayesian analysis of the MAT α _HMGbox domains of the full-length MAT1-1-1 proteins

The sequences of the MAT α _HMGbox domains of 19 full-length MAT1-1-1 proteins that display various amino acid substitutions were subjected to Bayesian clustering analysis and compared with the sequences of 6 authentic MAT1-1-1 proteins as references (Figure 1). The MAT α _HMGbox domains of the full-length MAT1-1-1 proteins were clustered into 5 Bayesian clusters (Figure 1), among which Clusters a and d were branched.

The domains of the 19 proteins with various amino acid substitutions were clustered into Clusters/Branches a2, b, d, and e1–e2 in the Bayesian clustering tree. Cluster d has a longer clustering distance than some of the other clusters do, and Branch e2 has the longest clustering distance. The MAT α _HMGbox domains encoded by the genome assemblies JAAVMX010000001 (6,699,061→6,699,153 & 6,699,203→6,699.637) and LKHE01001116 (4183←4620 & 4667←4759) of the *H. sinensis* strains 1229 and IOZ07 contain Y-to-M substitutions [61,63]. The domains encoded by the genome assembly ANOV01017390 (794←1129 & 1280←1396) of *H. sinensis* strain Co18 and the metatranscriptome assembly OSIN7648 (151→675) of the *C. sinensis* insect–fungal complex contain no changed amino acid residues [49,65]. The domain encoded by the metatranscriptome assembly GAGW01008880 (714←1127) of the *C. sinensis* insect–fungal complex is truncated at the N-terminus (Figure S1) [64]. The MAT α _HMGbox domains encoded by the genome assembly ANOV01017390 and metatranscriptome assemblies OSIN7648 and GAGW01008880 cluster into Branch a1 in the Bayesian clustering tree, together with the 6 authentic MAT1-1-1 proteins (Figure 1).

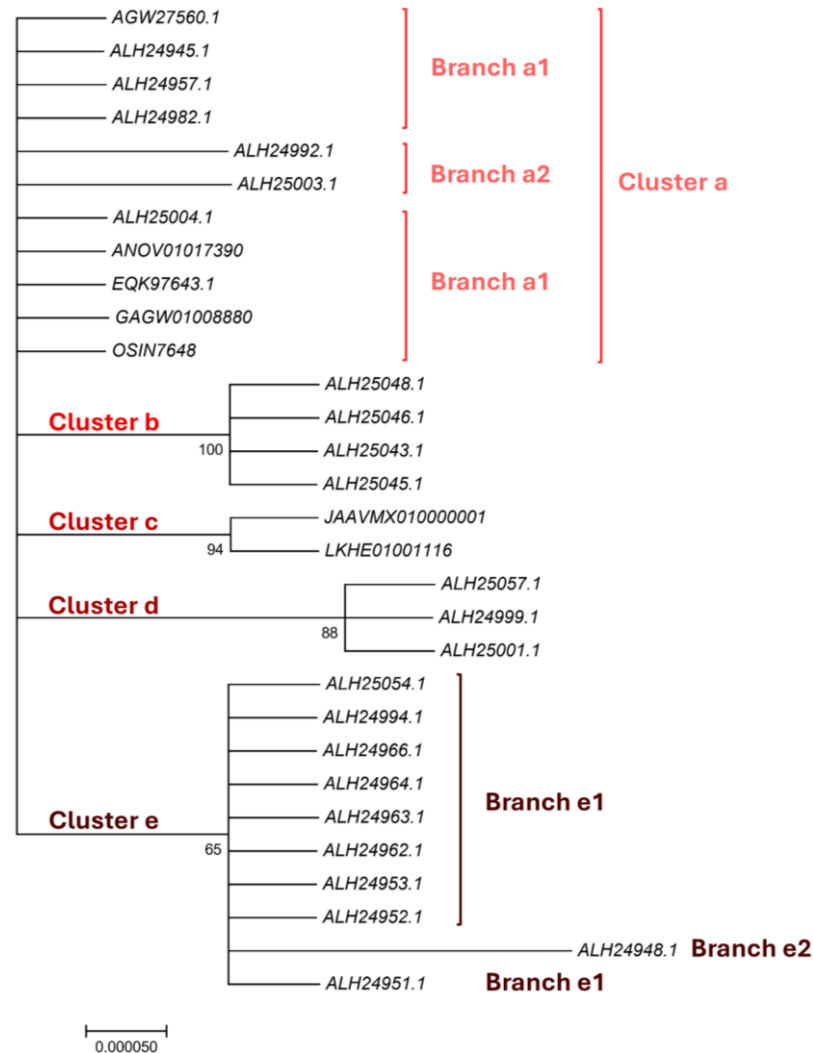


Figure 1. A Bayesian majority rule consensus clustering tree for the MAT α _HMGbox domain sequences of the 25 MAT1-1-1 proteins of the wild-type *C. sinensis* isolates and the corresponding domain sequences encoded by the genome and metatranscriptome assemblies of the *H. sinensis* strains and *C. sinensis* insect–fungal complexes was obtained via MrBayes v3.2.7 software.

Tables S5–S6 summarize the Bayesian clustering results and show the amino acid substitutions and deletions in the MAT α _HMGbox domains of the MAT1-1-1 proteins

derived from wild-type *C. sinensis* isolates and those encoded by the genome and metatranscriptome assemblies of *H. sinensis* strains and natural *C. sinensis* insect–fungal complexes, respectively.

3.4. Bayesian analysis of the HMG-box_ROX1-like domains of the full-length MAT1-2-1 protein sequences

Thirty full-length MAT1-2-1 proteins that are present in wild-type *C. sinensis* isolates or encoded by the genome and transcriptome assemblies of *H. sinensis* strains and by the metatranscriptome assembly of the *C. sinensis* insect–fungal complex and that have various amino acid substitutions in their HMG-box_ROX1-like domains were subjected to Bayesian clustering analysis and compared with the HMG-box_ROX1-like domains of 6 authentic MAT1-2-1 proteins as references. Figure 2 shows 2 branched clusters (a and b). Branches b2–b3 have longer clustering distances than some of the other branches do, and Branch b2β has the longest clustering distance.

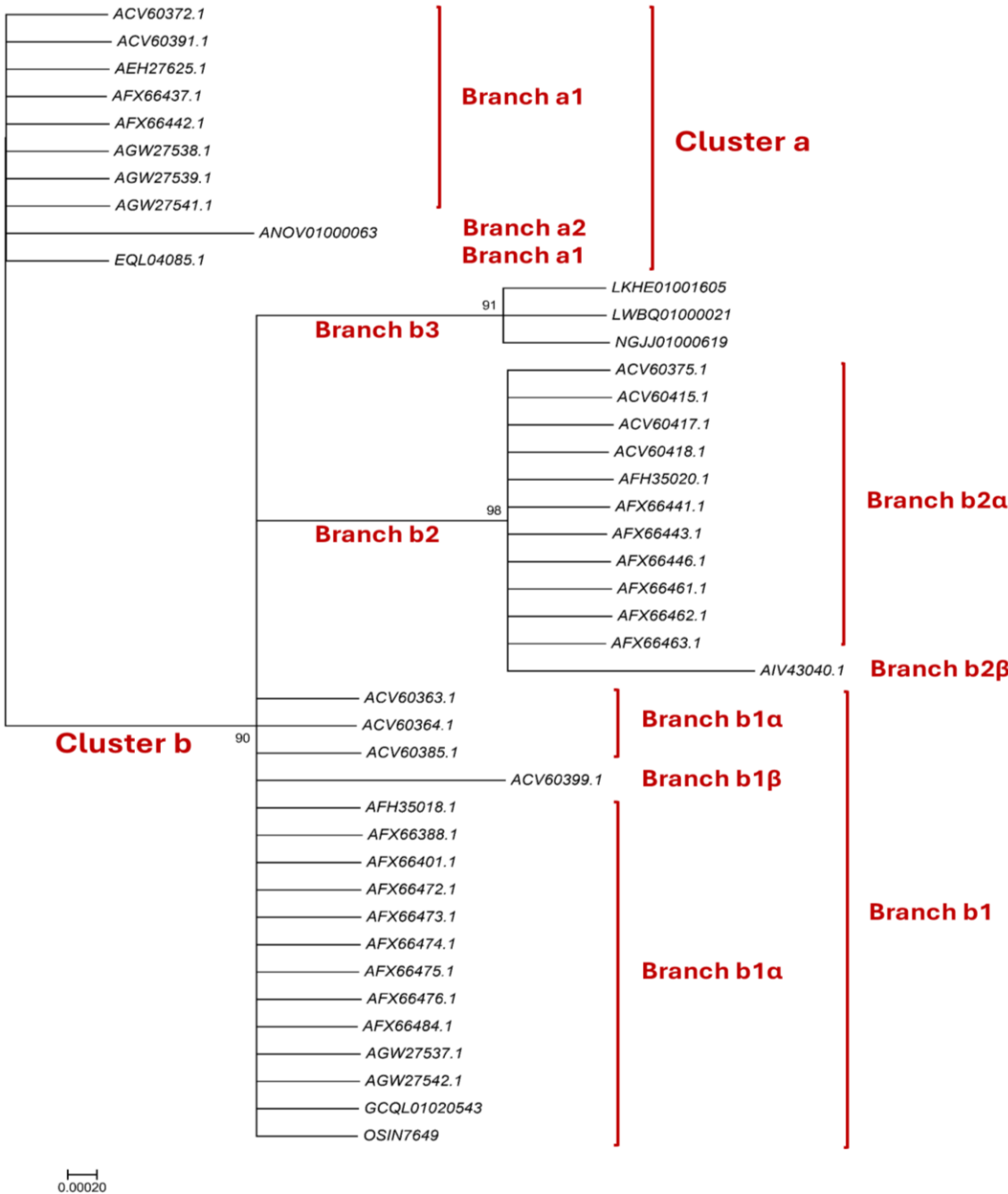


Figure 2. Bayesian majority rule consensus clustering tree inferred via MrBayes v3.2.7 software for the HMG-box_ROX1-like domain sequences of 35 full-length MAT1-2-1 proteins of the wild-type *C. sinensis* isolates and the corresponding domain segments of the translated genome, transcriptome, and metatranscriptome assemblies of the *H. sinensis* strains and *C. sinensis* insect–fungal complexes.

The HMG-box_ROX1-like domain encoded by the genome assembly ANOV01000063 (9759→9851 & 9907→10,026) of *H. sinensis* strain Co18 contains an S-to-A substitution and is clustered into Branch a2 with a slightly longer clustering distance (Figures 2 and S2) [49]. The domains encoded by other genome assemblies [LKHE01001605 (14,016←14,135 & 14,191←14,135 & 14,191←14,283), LWBQ01000021 (239,029←239,148 & 239,204←239,269), and NGJJ01000619 (23,186←23,305 & 23,361←23,453) of the *H. sinensis* strains 1229, ZJB12195, and CC1406-20395, respectively] [60–62] contain Y-to-H and S-to-A substitutions and are clustered into Branch b3 in the Bayesian clustering tree (Figures 2 and S2).

The HMG-box_ROX1-like domain of the MAT1-2-1 protein encoded by the transcriptome assembly GCQL01020543 (553←765) of the *H. sinensis* strain L0106 contains a Y-to-H substitution [59]. The domain of the MAT1-2-1 protein encoded by the metatranscriptome assembly OSIN7649 (379→591) of the *C. sinensis* insect–fungi complex is 100% identical to the reference MAT1-2-1 protein AEH27625 [56,65]. The domains of the MAT1-2-1 proteins encoded by the transcriptome and metatranscriptome assemblies were clustered into Branch b1α in the Bayesian clustering tree (Figures 2 and S2).

Tables S7–S8 summarize the Bayesian clustering results that reflect the amino acid substitutions and deletions in the HMG-box_ROX1-like domains of MAT1-2-1 proteins derived from wild-type *C. sinensis* isolates or encoded by the genome, transcriptome and metatranscriptome assemblies of *H. sinensis* strains and natural *C. sinensis* insect–fungal complexes, respectively.

3.5. Heteromorphic stereostructures of the MATα_HMGbox domains of full-length MAT1-1-1 proteins derived from wild-type *C. sinensis* isolates

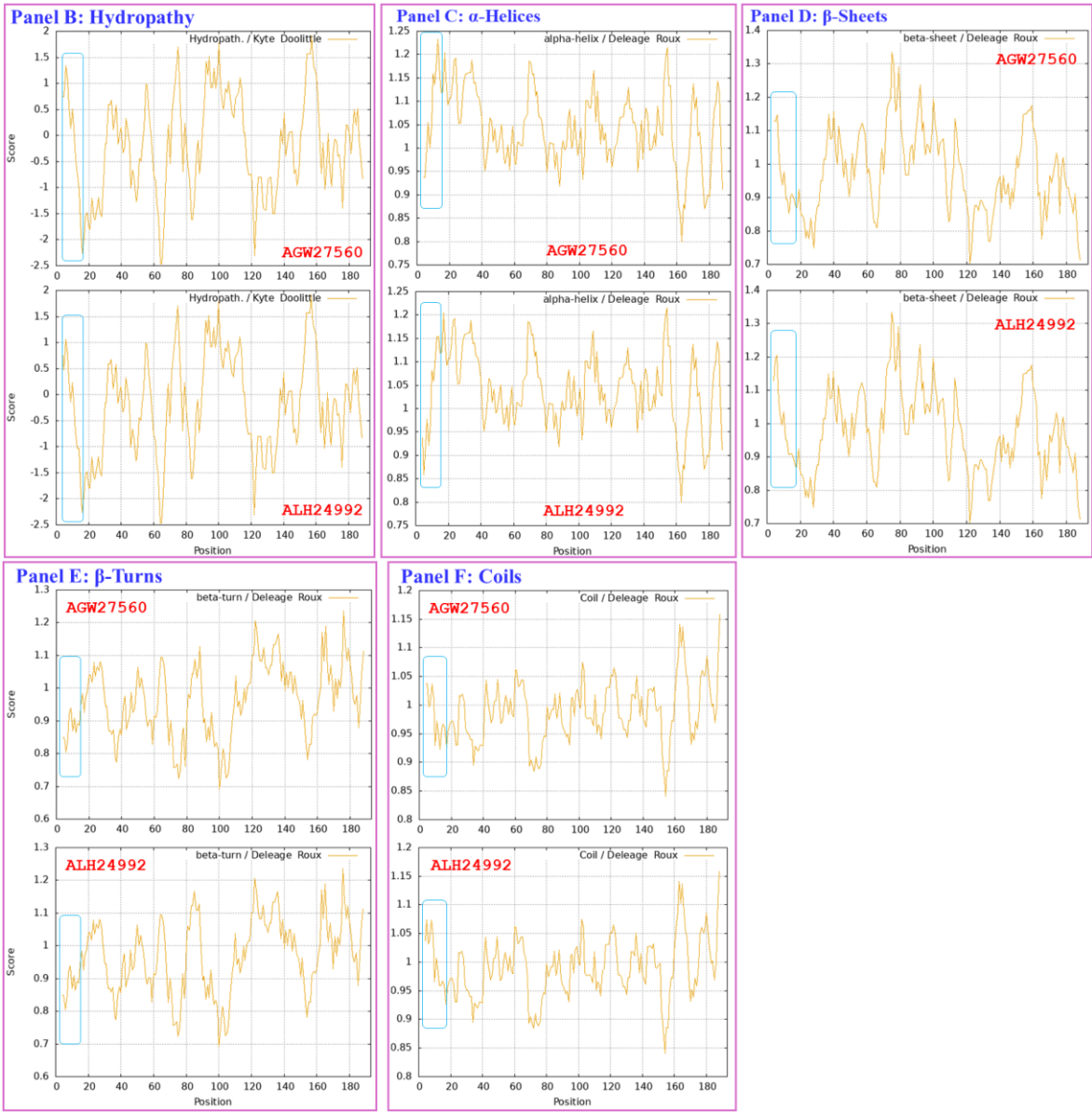
Section 3.5 presents analyses of the correlations between the changes in the structure and hydrophobicity of the MATα_HMGbox domains of full-length MAT1-1-1 proteins (under 7 AlphaFold 3D structure codes; cf. Table S5) derived from wild-type *C. sinensis* isolates compared with the reference MAT1-1-1 proteins represented by AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 (cf. Table S1) [48]. The MATα_HMGbox domain sequences (amino acid residues 51→225) were extended by 9 amino acid residues both upstream and downstream of the domains because of the setting of the sequential ExPASy ProtScale plotting for hydropathy, α-helices, β-sheets, β-turns, and coils at a window size of 9 amino acid residues (cf. Section 2.5 of Materials and methods).

Figure 3 compares the hydrophobicity and the structures of the MATα_HMGbox domain of the mutant MAT1-1-1 protein ALH24992 (under the AlphaFold code A0A0N9R5B3) derived from the wild-type *C. sinensis* isolate SC09_65 (cf. Tables S1 and S5) [52] with those of the MATα_HMGbox domain of the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [48]. Panel A shows an A-to-T substitution (aliphatic alanine to threonine with polar side chains; hydropathy indices changed from 1.8 to -0.7). This substitution causes reduced hydrophobicity (the larger the hydropathy value is, the stronger the hydrophobicity, whereas negative values indicate hydrophilicity [68]), as illustrated by the changes in topology and waveform in the hydropathy plot shown in Panel B. These changes result in an altered secondary structure of the protein in the area surrounding the mutation site in the MATα_HMGbox domain of the mutant MAT1-1-1 protein ALH24992, as illustrated

by the changes in topology and waveform in the ExPASy ProtScale plots for α -helices, β -sheets, β -turns, and coils shown in Panels C–F, respectively.

Panel A: Alignment of amino acid sequences

AGW27560	42	ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMFRSYYLKLPDVQOKTASGFLTTL	101
ALH24992	42	-----T-----	101
AGW27560	102	WHKDPFRNKWALIAKVYSFVRDQIGKDKVLSYFMSLACPTMTIIEPAAYLNALGWCVQD	161
ALH24992	102	-----	161
AGW27560	162	DDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSHS	221
ALH24992	162	-----	221
AGW27560	222	GIMAPRAANCTPP	234
ALH24992	222	-----	234



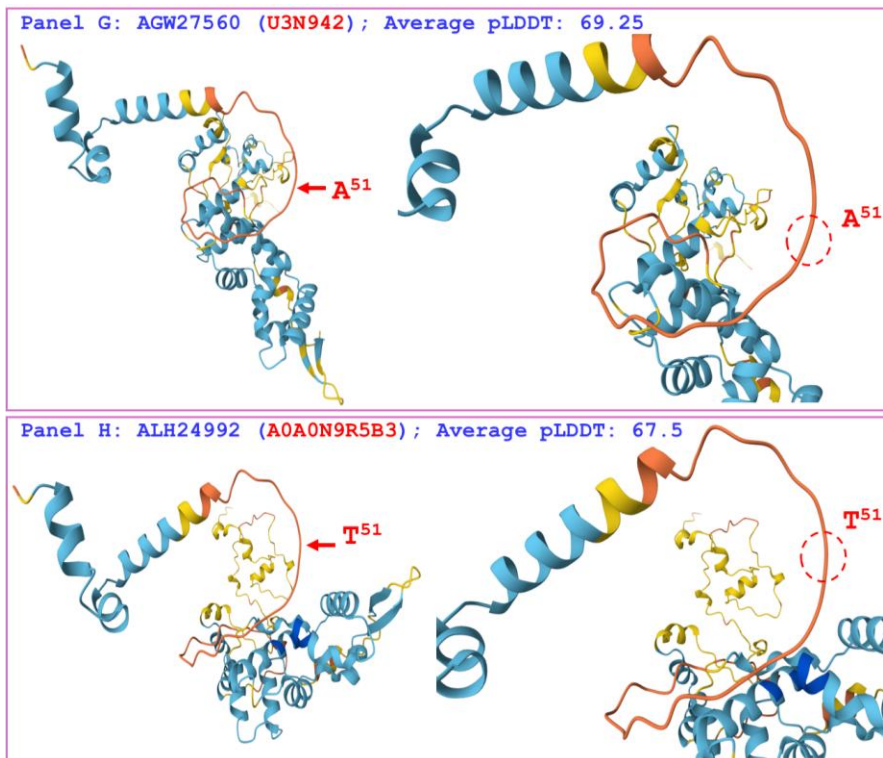


Figure 3. Correlation of the changes in hydrophobicity and the primary, secondary, and tertiary structures of the MAT α _HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the mutant protein ALH24992 (under the AlphaFold code A0A0N9R5B3) derived from the wild-type *C. sinensis* isolate SC09_65. Panel A shows an alignment of the amino acid sequences of the MAT α _HMGbox domains of the MAT1-1-1 proteins; the amino acid substitution is shown in red, whereas the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and in the 2D structure (Panels C–F for α -helices, β -sheets, β -turns, and coils, respectively) of the protein; the open rectangles in blue outline the changes in topology and waveform of the ExPASy plots. Panels G–H show the 3D structures of the full-length proteins on the left and the locally amplified structures surrounding the mutation site on the right. The model confidence values for the AlphaFold-predicted 3D structures are as follows: very high (pLDDT>90); high (90>pLDDT>70); low (70>pLDDT>50); and very low (pLDDT<50).

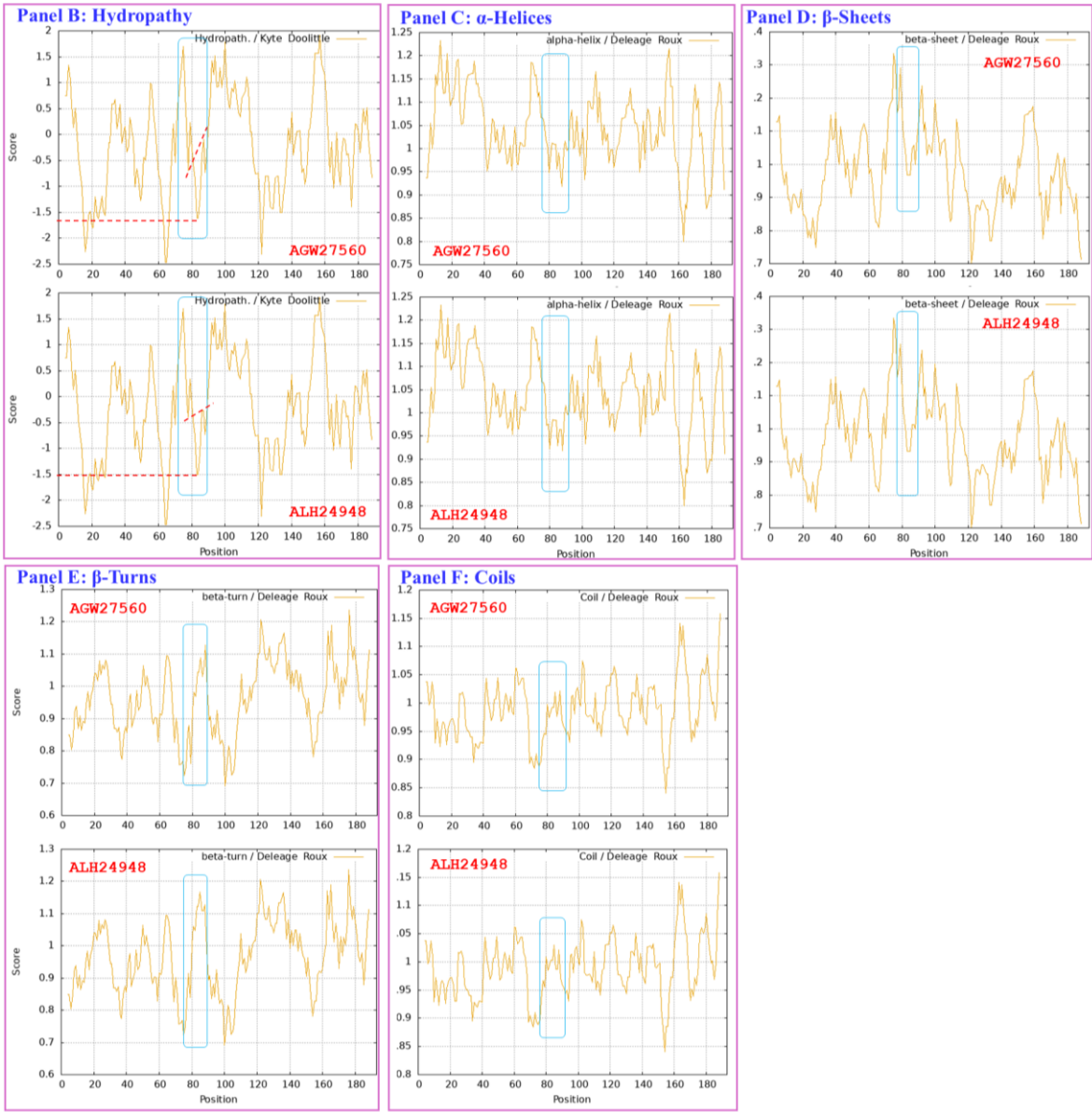
As shown in Panels G–H of Figure 3, which illustrate the 3D structures of the proteins, the substitution is located upstream of the core structure of 3 α -helices and may not directly participate in stabilizing the hydrophobic core structure of the DNA-binding domain [84–85]. However, the replacement of alanine with a threonine residue with reduced hydrophobicity altered the stereostructure of the MAT α _HMGbox domain of the MAT1-1-1 protein ALH24992 under the AlphaFold code A0A0N9R5B3, possibly through long-range spatial interactions. The mutant MAT α _HMGbox domain clustered into Branch a2 in the Bayesian clustering tree (*cf.* Figure 1, Table S5).

Figure 4 compares the hydrophobicity and structure of the MAT α _HMGbox domain of the MAT1-1-1 protein ALH24948 (under the AlphaFold code A0A0N9QUF3) derived from the wild-type *C. sinensis* isolate GS09_143 (*cf.* Table S1) [52] with those of the MAT α _HMGbox domain of the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [48]. Panel A shows substitutions of two adjacent residues: Q-to-S (glutamine to serine, both of which have polar side chains; hydropathy indices changed from -3.5 to -0.8 [68]) and I-to-F (aliphatic isoleucine to aromatic phenylalanine; hydropathy indices changed from 4.5 to 2.8), causing altered hydrophobicity, as illustrated by the changes in topology and waveform in the hydropathy plot

in Panel B. These changes alter the secondary structure surrounding the mutation sites in the MAT α _HMGbox domain, as illustrated by the changes in topology and waveform in the ExPASy ProtScale plots for α -helices, β -sheets, β -turns, and coils shown in Panels C–F. As shown in Panels G–H for the diagrams of the 3D structure, the adjacent amino acid substitutions are located at the end of one of the 3 α -helices that form the hydrophobic core stereostructure [84–85], and they cause the altered tertiary structure of the MAT α _HMGbox domain of the mutant protein ALH24948 under the AlphaFold code A0A0N9QUF3. The mutant MAT α _HMGbox domain clustered into Branch e2 in the Bayesian clustering tree (*cf.* Figure 1, Table S5).

Panel A: Amino acid sequence alignment

AGW27560	42	ADVLTPVSTAAASRATRQ	KEASCDRAKRPLNAFM	AFRSYYLKLFPDVQ	QKTASGFLTTL	101
ALH24948	42	-----	-----	-----	-----	101
AGW27560	102	WHKDPFRNKWALIAKV	YSFVRDQIGKDKVSL	SYFMSLACPTMTI	IEPAAYLNALGWC	161
ALH24948	102	-----	-----	-----	-----	161
AGW27560	162	DDAGSQKLFQDESSAN	LQSSLLSAEYPSTE	IELLSALVNIGYFP	DHGADLVERMGSS	221
ALH24948	162	-----	-----	-----	-----	221
AGW27560	222	GIMAPRAANCTPP				234
ALH24948	222	-----				234



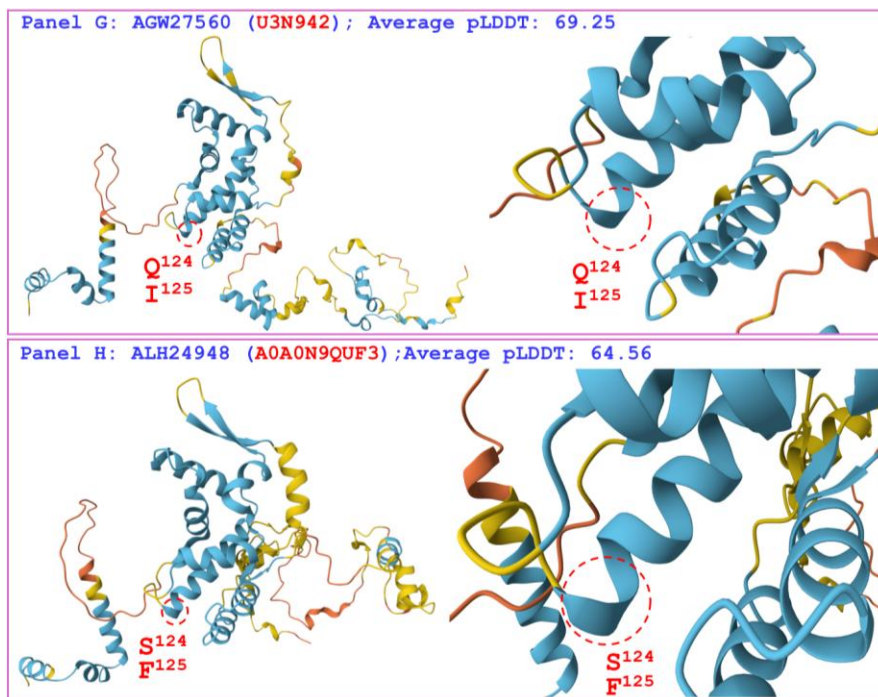


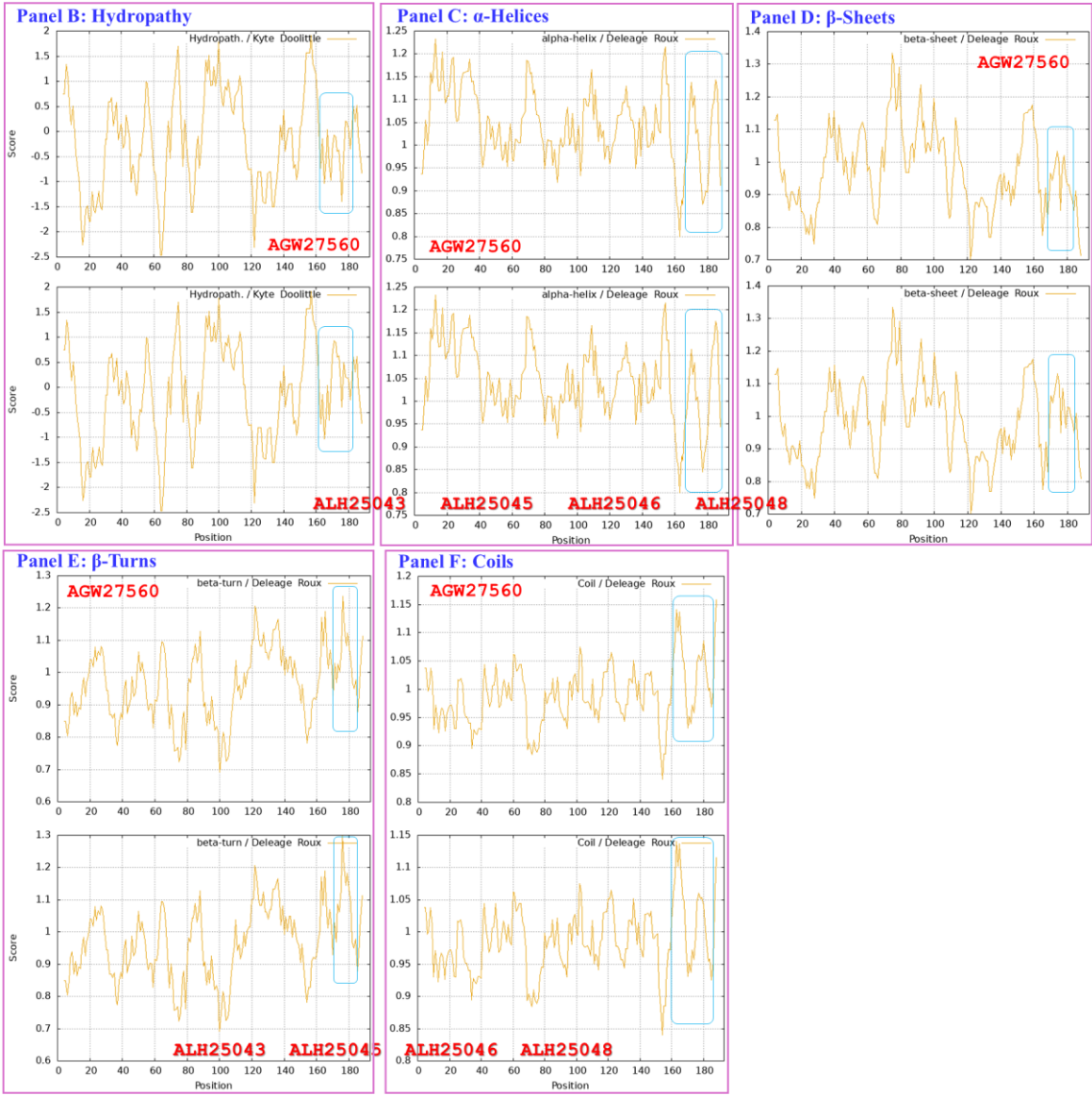
Figure 4. Correlations among the changes in the primary, secondary, and tertiary structures of the MAT α _HMGbox domains of the MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the mutant MAT1-1-1 protein ALH24948 (under the AlphaFold code A0A0N9QUF3) derived from the wild-type *C. sinensis* isolate GS09_143. Panel A shows an alignment of the MAT α _HMGbox domain sequences of the MAT1-1-1 proteins; amino acid substitutions are shown in red, whereas the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and in the 2D structure (Panels C–F for α -helices, β -sheets, β -turns, and coils, respectively) of the protein; the open rectangles in blue outline the changes in topology and waveform shown in the ExPASy plots. Panels G–H show the 3D structures of the full-length proteins on the left and the locally amplified structures at the sites of substitutions on the right. The model confidence values for the AlphaFold-predicted 3D structures are as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

Figure 5 compares the hydrophobicities and the structures of the MAT α _HMGbox domains of the MAT1-1-1 proteins ALH25043, ALH25045, ALH25046, and ALH25048 (under the AlphaFold code A0A0N9QMS9) derived from the wild-type *C. sinensis* isolates YN09_22, YN09_51, YN09_6, and YN09_64, respectively (*cf.* Table S1) [52], with those of the MAT α _HMGbox domain of the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [48]. Panel A shows substitutions of R-to-I (basic arginine to aliphatic isoleucine; hydropathy indices changed from -4.5 to 4.5 [68]) and P-to-T (proline to threonine, both of which possess polar neutral side chains; hydropathy indices changed from -1.6 to -0.7). These substitutions result in increased hydrophobicity, as illustrated by the changes in topology and waveform that are apparent in the hydropathy plot in Panel B. They also result in altered secondary structure surrounding the mutation sites in the MAT α _HMGbox domains of the MAT1-1-1 proteins ALH25043, ALH25045, ALH25046, and ALH25048, as illustrated by the changes in topology and waveform shown in the ExPASy ProtScale plots for α -helices, β -sheets, β -turns, and coils in Panels C–F. As shown in the diagrams of the 3D structure in Panels G–H, the sites of the R-to-I and P-to-T substitutions are inside and outside, respectively, the hydrophobic core formed by the 3 α -helices, and the substantial increases in hydrophobicity alter the tertiary structures of the MAT α _HMGbox domains of the MAT1-1-1 proteins

under the AlphaFold code A0A0N9QMS9. The mutant MAT α _HMGbox domain clustered into Cluster 2 in the Bayesian clustering tree (*cf.* Figure 1, Table S5).

Panel A: Amino acid sequence alignment

AGW27560	42	ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLPDVQQKTASGFLTTL	101
ALH25043	42	-----	101
ALH25045	42	-----	101
ALH25046	42	-----	101
ALH25048	42	-----	101
AGW27560	102	WHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTTIEPAAYLNALGWCVQD	161
ALH25043	102	-----	161
ALH25045	102	-----	161
ALH25046	102	-----	161
ALH25048	102	-----	161
AGW27560	162	DDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSHS	221
ALH25043	162	-----I-----	221
ALH25045	162	-----I-----	221
ALH25046	162	-----I-----	221
ALH25048	162	-----I-----	221
AGW27560	222	GIMAPRAANCTPP	234
ALH25043	222	---T-----	234
ALH25045	222	---T-----	234
ALH25046	222	---T-----	234
ALH25048	222	---T-----	234



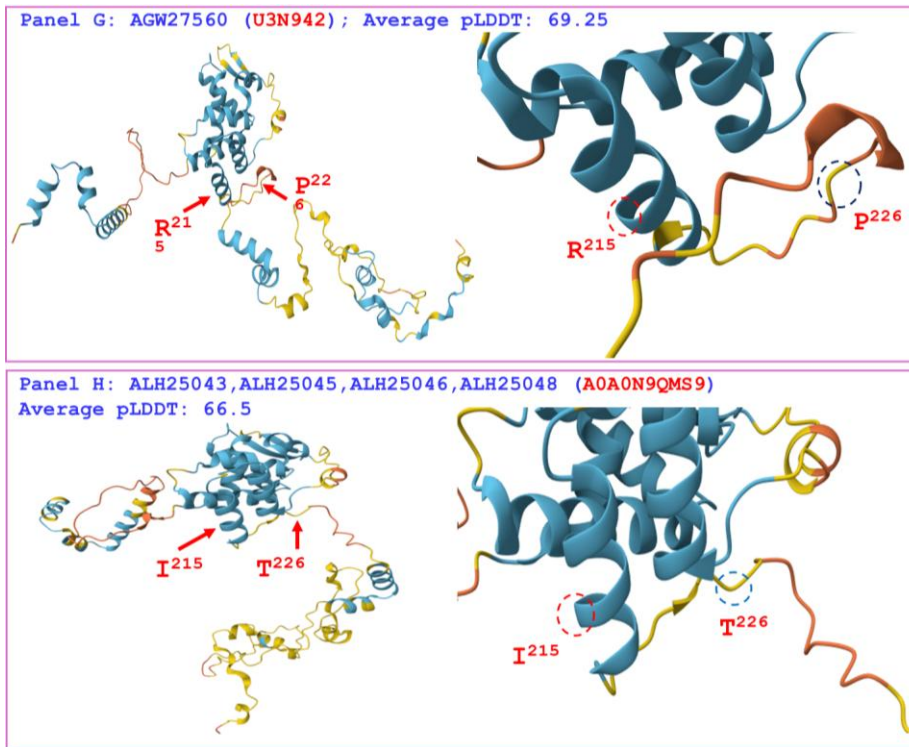
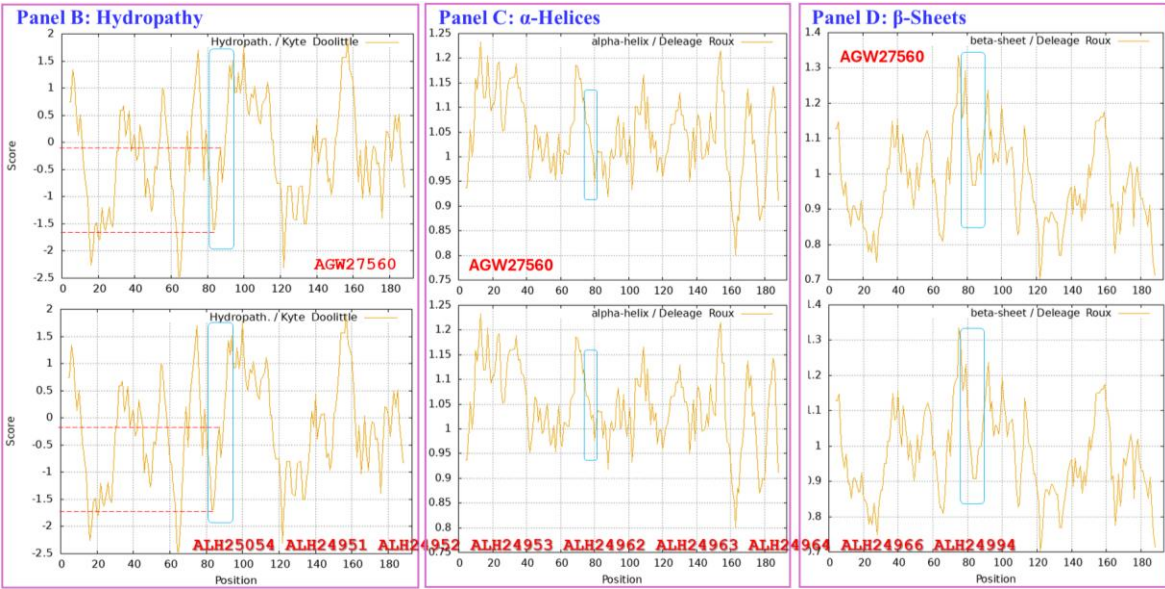


Figure 5. Correlation of changes in the primary, secondary, and tertiary structures of the MAT α _HMGbox domains of the MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) is derived from the *H. sinensis* strain CS68-2-1229, and the mutant MAT1-1-1 proteins ALH25043, ALH25045, ALH25046, and ALH25048 (under the AlphaFold code A0A0N9QMS9) are derived from the wild-type *C. sinensis* isolates YN09_22, YN09_51, YN09_6, and YN09_64, respectively. Panel A shows an alignment of the amino acid sequences of the MAT α _HMGbox domains of the MAT1-1-1 proteins; amino acid substitutions are shown in red, whereas the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and in the 2D structure (Panels C–F for α -helices, β -sheets, β -turns, and coils, respectively) of the proteins; the open rectangles in blue outline the changes in topology and waveform of the plots. Panels G–H show the 3D structures of the full-length proteins on the left and the locally amplified structures at the sites of substitutions on the right. The model confidence values for the AlphaFold-predicted 3D structures are as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

Figure 6 compares the hydrophobicities and structures of the MAT α _HMGbox domains of the MAT1-1-1 proteins ALH25054, ALH24951, ALH24952, ALH24953, ALH24962, ALH24963, ALH24964, ALH24966, and ALH24994 (under the AlphaFold code A0A0N7G845) derived from the wild-type *C. sinensis* isolates GS09_311, GS09_229, GS09_281, GS10_1, QH09_164, QH09_173, QH09_201, QH09_210, and SC09_87, respectively (*cf.* Table S1) [52], with those of the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [48]. Panel A shows I-to-L substitutions (aliphatic isoleucine to aliphatic leucine, hydrophathy indices changed from 4.5 to 3.8 [68]); these substitutions resulted in reduced hydrophobicity, as illustrated by the changes in topology and waveform shown in Panel B for the hydrophathy plot.

Panel A: Amino acid sequence alignment

AGW27560	42	ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLPDVQQKTASGFLTTL	101
ALH25054	42	-----	101
ALH24951	42	-----	101
ALH24952	42	-----	101
ALH24953	42	-----	101
ALH24962	42	-----	101
ALH24963	42	-----	101
ALH24964	42	-----	101
ALH24966	42	-----	101
ALH24994	42	-----	101
AGW27560	102	WHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTTIEPAAYLNALGWCVD	161
ALH25054	102	-----L-----	161
ALH24951	102	-----L-----	161
ALH24952	102	-----L-----	161
ALH24953	102	-----L-----	161
ALH24962	102	-----L-----	161
ALH24963	102	-----L-----	161
ALH24964	102	-----L-----	161
ALH24966	102	-----L-----	161
ALH24994	102	-----L-----	161
AGW27560	162	DDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSHS	221
ALH25054	162	-----	221
ALH24951	162	-----	221
ALH24952	162	-----	221
ALH24953	162	-----	221
ALH24962	162	-----	221
ALH24963	162	-----	221
ALH24964	162	-----	221
ALH24966	162	-----	221
ALH24994	162	-----	221
AGW27560	222	GIMAPRAANCTPP	234
ALH25054	222	-----	234
ALH24951	222	-----	234
ALH24952	222	-----	234
ALH24953	222	-----	234
ALH24962	222	-----	234
ALH24963	222	-----	234
ALH24964	222	-----	234
ALH24966	222	-----	234
ALH24994	222	-----	234



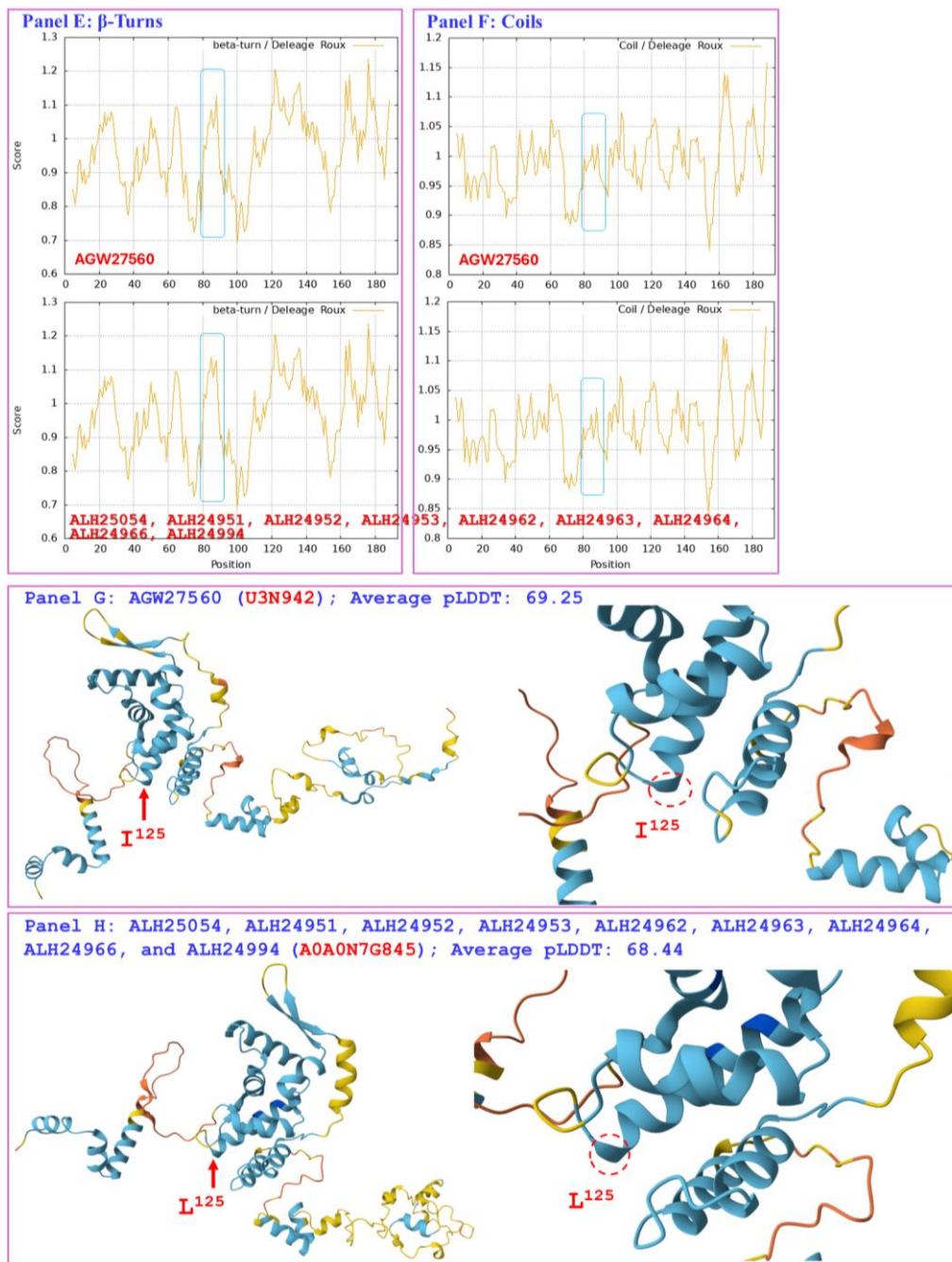


Figure 6. Correlation of changes in the primary, secondary, and tertiary structures of the MAT α _HMGbox domains in the MAT1-1-1 proteins. The reference protein AGW27560 (under the AlphaFold code U3N942) is derived from the *H. sinensis* strain CS68-2-1229, and the mutant MAT1-1-1 proteins ALH25054, ALH24951, ALH24952, ALH24953, ALH24962, ALH24963, ALH24964, ALH24966, and ALH24994 (under the AlphaFold code A0A0N7G845) are derived from the wild-type *C. sinensis* isolates GS09_311, GS09_229, GS09_281, GS10_1, QH09_164, QH09_173, QH09_201, QH09_210, and SC09_87, respectively. Panel A shows an alignment of the amino acid sequences of the MAT α _HMGbox domains of the MAT1-1-1 proteins; amino acid substitutions are shown in red, whereas the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and in the 2D structure (Panels C–F for α -helices, β -sheets, β -turns, and coils, respectively) of the proteins; the open rectangles in blue outline the changes in topology and waveform shown in the plots. Panels G–H show the 3D structures of the full-length proteins on the left and the locally amplified structures at the site of substitution on the right. The model confidence values for the AlphaFold-predicted 3D structures are as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

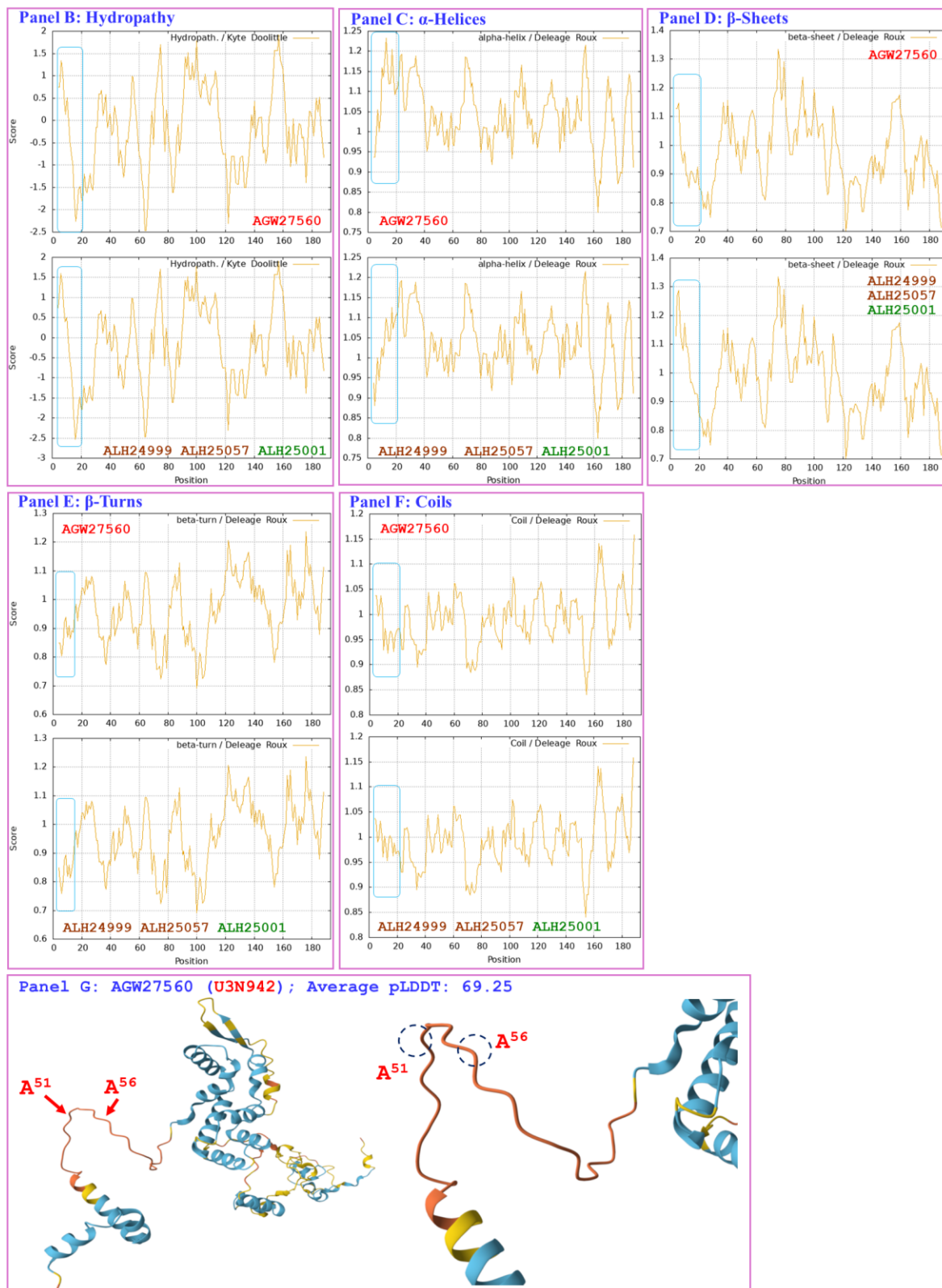
As shown in Panels C–F of Figure 6, the secondary structure of the proteins near the mutation sites in the MAT α _HMGbox domains of the MAT1-1-1 proteins ALH25054, ALH24951, ALH24952, ALH24953, ALH24962, ALH24963, ALH24964, ALH24966, and ALH24994 (*cf.* Table S1) was altered, as illustrated by the changes in topology and waveform in the ExPASy ProtScale plots for α -helices, β -sheets, β -turns, and coils, respectively. As shown in the representations of the 3D structures presented in Panels G–H, the substituted amino acid residues are located within the core structures of the α -helices, and they destabilize the hydrophobic core of the protein [84–85]; this alters the tertiary structures of the MAT α _HMGbox domains of the MAT1-1-1 proteins under the AlphaFold code A0A0N7G845, which cluster into Branch e1 in the Bayesian clustering tree (*cf.* Figure 1, Table S5).

Figure 7 shows the hydrophobicities and structures of the MAT α _HMGbox domains of the MAT1-1-1 proteins ALH24999 and ALH25057 (under the AlphaFold code A0A0N9QMT4) and the MAT1-1-1 protein ALH25001 (under the AlphaFold code A0A0N9R4Q4) derived from the wild-type *C. sinensis* isolates XZ07_H2, XZ12_16, and XZ05_2, respectively (*cf.* Table S1) [52], compared with the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [48]. Panel A shows substitutions of A-to-V (aliphatic alanine to aliphatic valine; hydrophathy indices changed from 1.8 to 4.2 [68]) and A-to-T (aliphatic isoleucine to threonine, an amino acid that contains polar neutral side chains; hydrophathy indices changed from 1.8 to -0.7). These substitutions altered the hydrophobicity of the protein, as illustrated by the changes in topology and waveform shown in the hydrophathy plot in Panel B. These changes altered the secondary structure surrounding the mutation sites in the MAT α _HMGbox domain of the MAT1-1-1 proteins ALH24999, ALH25057, and ALH25001, as illustrated by the changes in topology and waveform in the ExPASy ProtScale plots for α -helices, β -sheets, β -turns, and coils shown in Panels C–F, respectively. As shown in the 3D structures represented in Panels G–H, the mutation sites are located upstream of the hydrophobic core structure of 3 α -helices in the MAT α _HMGbox domains, and mutations at these sites alter the tertiary structures of the MAT α _HMGbox domains of MAT1-1-1 proteins under the AlphaFold codes A0A0N9QMT4 and A0A0N9R4Q4, which are clustered into Cluster d in the Bayesian clustering tree (*cf.* Figure 1, Table S5).

Panel A: Amino acid sequence alignment

Panel A: Amino acid sequence alignment

AGW27560	42	ADVLT	PVSTAAAS	RAT	RQTKEAS	CDRAK	RPLNA	FMAFR	SYYLK	LFDPV	QQKTAS	GFL	TTL	101	
ALH24999	42	-----	V-----	T-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	101	
ALH25057	42	-----	V-----	T-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	101	
ALH25001	42	-----	V-----	T-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	101	
AGW27560	102	WHKDP	FRNKW	ALIAK	VYSFVRD	QIGKDK	VSLSY	FMSLAC	PTMTT	IEPAAY	LNALG	WC	VQD	161	
ALH24999	102	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	161	
ALH25057	102	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	161	
ALH25001	102	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	161	
AGW27560	162	DDAGS	QKLFQ	DESSAN	LQSSLLS	AEYP	STEI	ELLSA	LVNIG	YFPD	HGAD	LIVER	MGS	SSH	221
ALH24999	162	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	221
ALH25057	162	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	221
ALH25001	162	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	221
AGW27560	222	GIMAP	RAANCT	PP	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	234
ALH24999	222	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	234
ALH25057	222	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	234
ALH25001	222	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	234



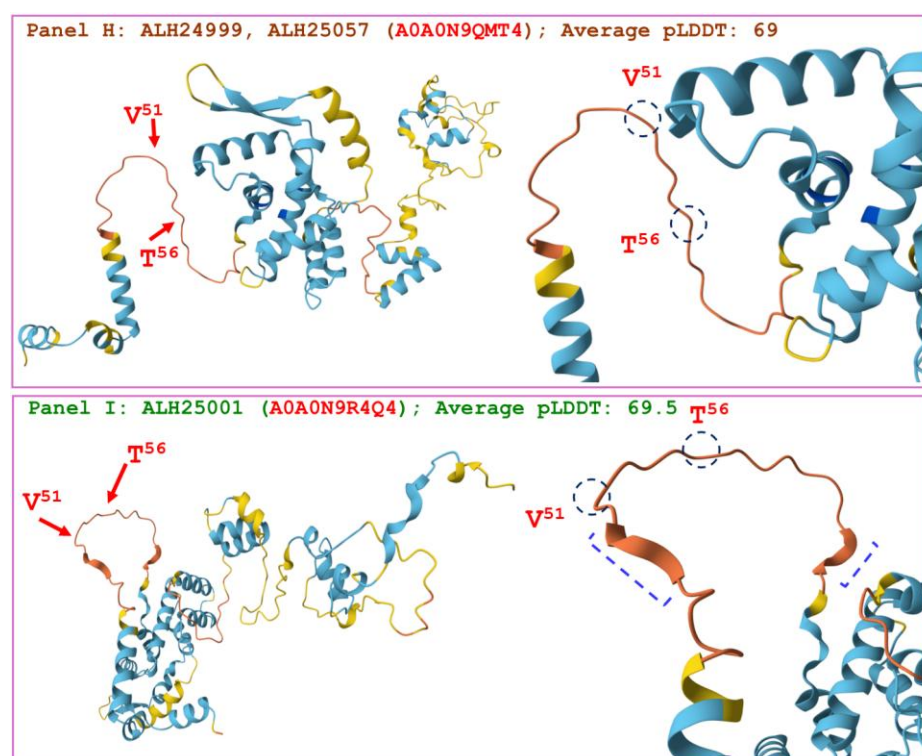


Figure 7. Correlation of the changes in the primary, secondary, and tertiary structures of the MAT α _HMGbox domains in the MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) is derived from the *H. sinensis* strain CS68-2-1229, and the mutant MAT1-1-1 proteins ALH24999 and ALH25057 (under the AlphaFold code A0A0N9QMT4) and the protein ALH25001 (under the AlphaFold code A0A0N9R4Q4) are derived from the wild-type *C. sinensis* isolates XZ07_H2, XZ12_16, and XZ05_2, respectively. Panel A shows an alignment of the amino acid sequences of the MAT α _HMGbox domains of the MAT1-1-1 proteins; amino acid substitutions are shown in red; the sequences displayed in brown and green represent the MAT1-1-1 proteins under the AlphaFold codes A0A0N9QMT4 and A0A0N9R4Q4, respectively; and the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and in the 2D structure (Panels C–F, which display the α -helices, β -sheets, β -turns, and coils, respectively) of the proteins. The open rectangles in blue outline the changes in topology and waveform shown in the plots. Panels G–H show the 3D structures of the full-length proteins on the left and the locally amplified structures at the substitution sites on the right. The model confidence values for the AlphaFold-predicted 3D structures are as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

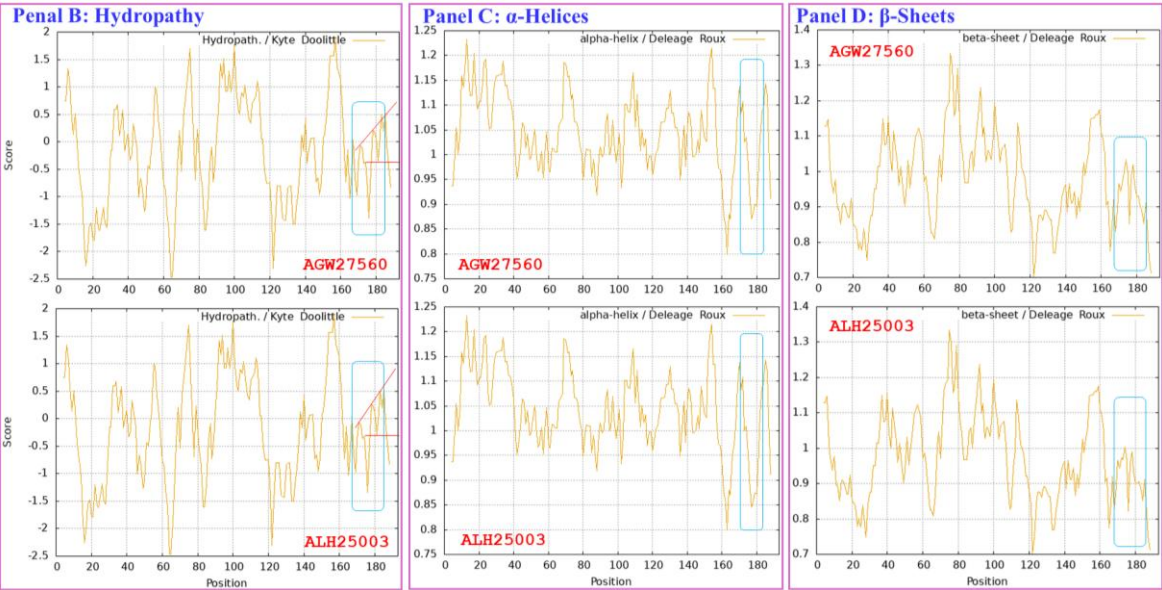
Although Panel A of Figure 7 shows the same substitutions of A-to-V and A-to-T in the MAT α _HMGbox domains of the mutant proteins ALH24999, ALH25057, and ALH25001 (under the AlphaFold codes A0A0N9QMT4 and A0A0N9R4Q4) [52], Panels H and I show different tertiary structures for the MAT α _HMGbox domains of the mutant MAT1-1-1 proteins under the 2 AlphaFold codes. Specifically, the protein ALH25001 contains 2 short β -sheet structures upstream and downstream of the mutation sites in the MAT α _HMGbox domain shown in Panel I; these structures are marked with dashed half brackets in blue. In contrast, the distinct tertiary structures of the mutant proteins ALH24999 and ALH25057 without β -sheet structures are shown in Panel H.

Figure 8 compares the hydrophobicity and structures of the MAT α _HMGbox domain of the MAT1-1-1 protein ALH25003 (under the AlphaFold code A0A0N7G850) derived from the wild-type *C. sinensis* isolate XZ05_6 (cf. Table S1) [52] with those of the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H.*

sinensis strain CS68-2-1229 [48]. Panel A shows an S-to-G substitution (serine, which has polar neutral side chains, is replaced by the unique amino acid glycine; hydropathy indices changed from -0.8 to -0.4 [68]). This substitution resulted in a slight increase in hydrophobicity, as illustrated by the changes in topology and waveform shown in the hydropathy plot in Panel B. These changes altered the secondary structure of the mutant MAT1-1-1 protein ALH25003 in the region surrounding the mutation site in the MAT α _HMGbox domain, as illustrated by the changes in topology and waveform in the ExPASy ProtScale plots for α -helices, β -sheets, β -turns, and coils shown in Panels C–F. As shown in the representations of the protein’s 3D structure in Panels G–H, the amino acid residue (serine), which is located at the end of one of the 3 α -helices that form the hydrophobic core structure of the MAT α _HMGbox domain [84–85], was substituted by glycine, which became the first residue after the α -helix in the domain. This single-residue substitution plus additional E-to-K and Y-to-H substitutions downstream of the MAT α _HMGbox domain (*cf.* Figure S1) significantly altered the tertiary structure of the protein ALH25003 under the unique AlphaFold code A0A0N7G850. The mutant MAT α _HMGbox domain clustered into Branch a2 in the Bayesian clustering tree (*cf.* Figure 1, Table S5).

Panel A: Amino acid sequence alignment

AGW27560	42	ADVLTPVSTAAASRATROTKEASCDRAKRPLNAFMFRSYLLKLPDVPQKTASGFLTTL	101
ALH25003	42	-----	101
AGW27560	102	WHKDPFRNKWALIAKVYSFVRDQIGKDKVLSYFMSLACPTMTTIEPAAYLNALGWCVQD	161
ALH25003	102	-----	161
AGW27560	162	DDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSHS	221
ALH25003	162	-----G-----	221
AGW27560	222	GIMAPRAANCTPP	234
ALH25003	222	-----	234



587

588

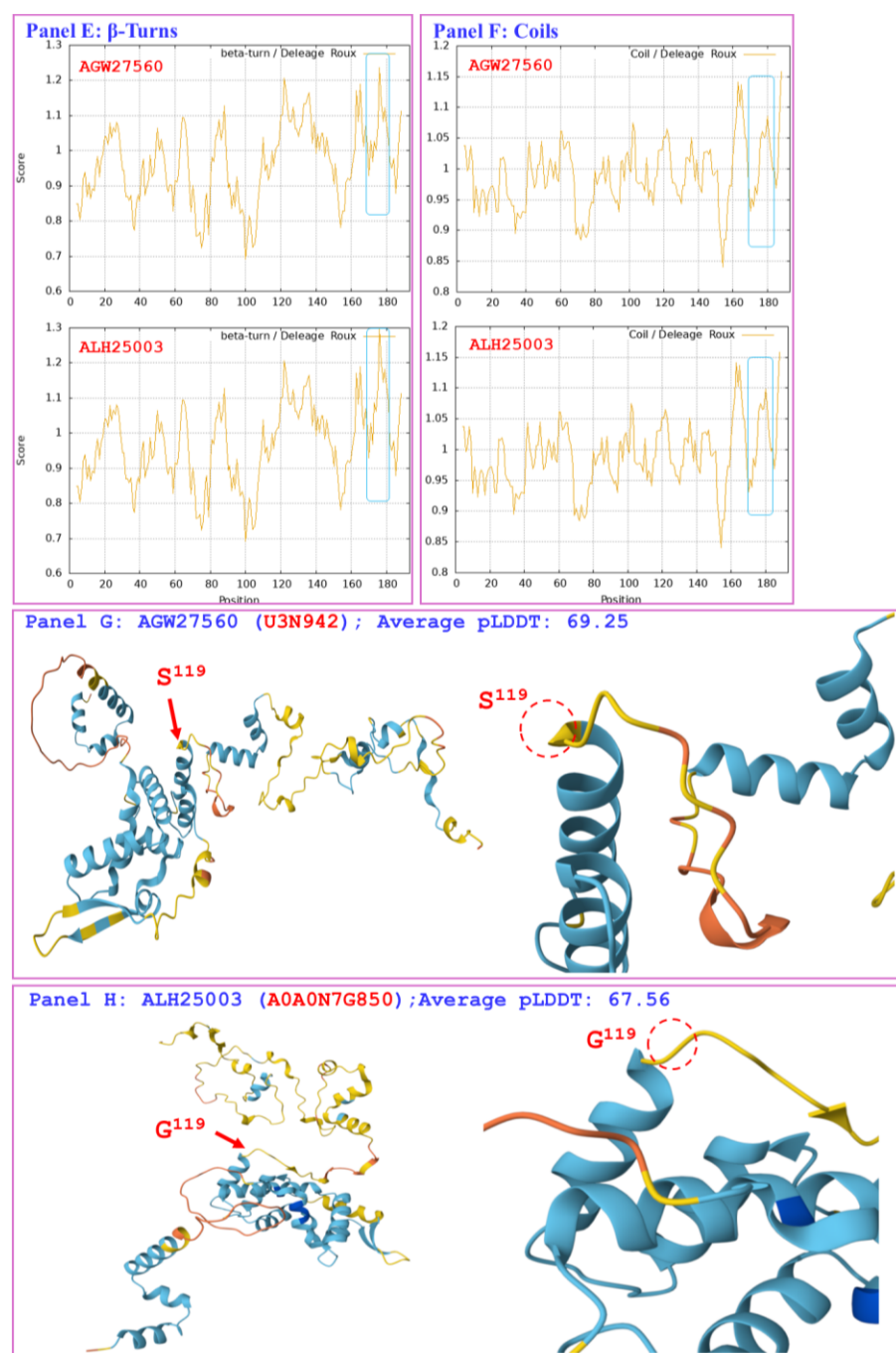


Figure 8. Correlation of the changes in the primary, secondary, and tertiary structures of the MAT α _HMGbox domains of the MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) is derived from the *H. sinensis* strain CS68-2-1229, and the mutant MAT1-1-1 protein ALH25003 (under the AlphaFold code A0A0N7G850) is derived from the wild-type *C. sinensis* isolate XZ05_6. Panel A shows an alignment of the amino acid sequences of the MAT α _HMGbox domains of the MAT1-1-1 proteins; the amino acid substitution is shown in red, whereas the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in the hydrophobicity (Panel B) and in the 2D structure of the protein (Panels C–F for the α -helices, β -sheets, β -turns, and coils, respectively) of the protein, and the open rectangles in blue outline the changes in topology and waveform shown in the plots. Panels G–H show the 3D structures of the full-length proteins on the left and the locally amplified structures at the sites of substitutions on the right. The model confidence values for the AlphaFold-predicted 3D structures are as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

Aggregating the data presented in Section 3.5, Panels A of Figures 3–8 demonstrate the 1–2 amino acid substitutions in the MAT α _HMGbox domains of MAT1-1-1 proteins under various AlphaFold 3D structure codes derived from wild-type *C. sinensis* isolates compared with the reference MAT1-1-1 protein AGW27560 under the AlphaFold code U3N942 [48]. The substitutions within the MAT α _HMGbox domains caused changes in the topologies and waveforms shown in the ExPASy ProtScale plots in Panels B–F of Figures 3–8. These changes indicate changes in the hydrophobicity and secondary structure (α -helices, β -sheets, β -turns, and coils) of the mutant full-length proteins. The changes in hydrophobicity and in the primary and secondary structures of the MAT α _HMGbox domains altered the AlphaFold 3D structures of the domains of the proteins under various AlphaFold codes (*cf.* Panels H of Figures 3–8 and Panel I of Figure 7). The mutant domain sequences containing 1 or 2 amino acid substitutions were clustered into different clades in the Bayesian clustering tree (*cf.* Figure 1, Table S5), and the altered 3D structures of the proteins ultimately changed the DNA binding affinities and the functionalities of gene transcriptional regulation related to the sexual reproduction of *O. sinensis*.

3.6. Diverse primary and secondary structures of the MAT α _HMGbox domains of MAT1-1-1 proteins encoded by the genome and metatranscriptome assemblies of *H. sinensis* and *C. sinensis* insect–fungal complexes

Figure 9 compares the hydrophobicities and secondary structures (α -helices, β -sheets, β -turns, and coils) of the MAT α _HMGbox domains of the MAT1-1-1 proteins encoded by the genome assemblies JAAVMX010000001 (6,699,061→6,699,153 & 6,699,206→6,699,637) and LKHE01001116 (4183←4620 & 4667←4759) of the *H. sinensis* strains IOZ07 and 1229 [61,63] with those of the reference MAT1-1-1 protein AGW27560 derived from the *H. sinensis* strain CS68-2-1229 [48]. Panel A shows a Y-to-M substitution (aromatic tyrosine to methionine, an amino acid that has polar neutral side chains; hydrophathy indices changed from -1.3 to 1.9 [68]). This substitution caused increased hydrophobicity in the MAT α _HMGbox domains of the MAT1-1-1 proteins encoded by the genome assemblies JAAVMX010000001 and LKHE01001116 [61,63], as illustrated by the changes in topology and waveform in the hydrophathy plot shown in Panel B. These changes resulted in altered secondary structures surrounding the mutation sites in the MAT α _HMGbox domains of the MAT1-1-1 proteins, as illustrated in the ExPASy ProtScale plots for α -helices, β -sheets, β -turns, and coils in Panels C–F, respectively. The MAT α _HMGbox domains of the MAT1-1-1 proteins were clustered into Branch c in the Bayesian clustering tree (*cf.* Figure 1, Table S6).

Panel A: Amino acid sequence alignment

AGW27560	42	ADVLT	FPVSTAAASRAT	QTKEASCDRAK	RPLNAFM	AFRSYYLKL	FPDVQ	QKTASG	FLTTL	101
JAAVMX010000001	6,699,034	-----	-----	-----	-----	M-----	-----	-----	-----	6,699,213
LKHE01001116	4786	-----	-----	-----	-----	M-----	-----	-----	-----	4555
AGW27560	85	WHKDP	FRNKWALIAK	VYSFVRDQ	IGKDKVSL	SFMSLACPT	MTTIEP	AAAYLNAL	GWC	VD 161
JAAVMX010000001	6,699,214	-----	-----	-----	-----	-----	-----	-----	-----	6,699,393
LKHE01001116	4554	-----	-----	-----	-----	-----	-----	-----	-----	4375
AGW27560	145	DDAGS	QKLFQDESSAN	LQSSLLSAE	YPSTEI	ELLSALVN	IGYFPD	HGADL	VERM	GSSHS 221
JAAVMX010000001	6,699,214	-----	-----	-----	-----	-----	-----	-----	-----	6,699,573
LKHE01001116	4374	-----	-----	-----	-----	-----	-----	-----	-----	4195
AGW27560	205	GIMAP	PRAANCTPP							234
JAAVMX010000001	6,699,574	-----	-----							6,699,664
LKHE01001116	4194	-----	-----							4156

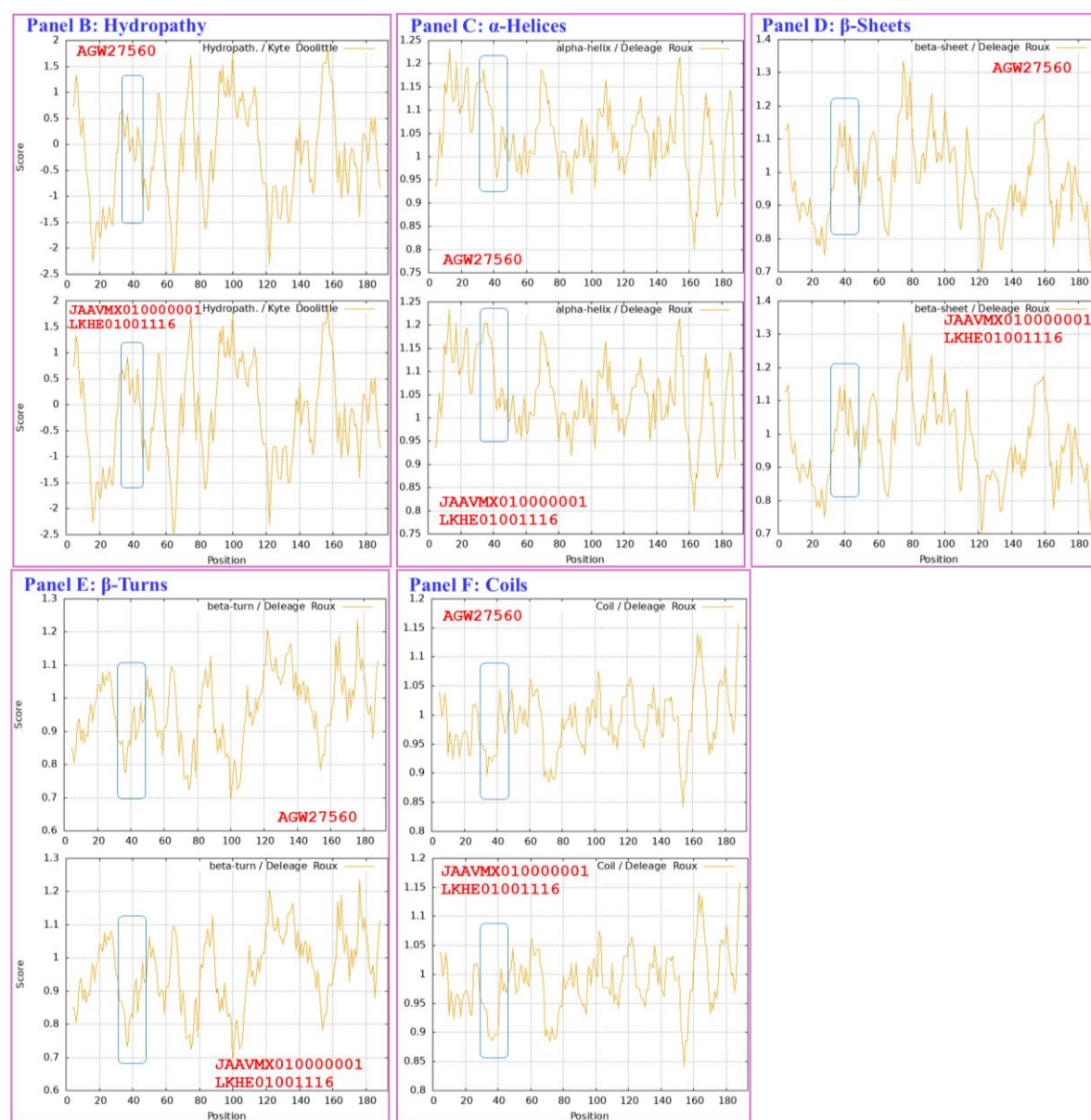


Figure 9. Correlation of the changes in the primary and secondary structures of the MAT α _HMGbox domains of the MAT1-1-1 proteins: the reference protein AGW27560 is derived from the *H. sinensis* strain CS68-2-1229, and the truncated MAT1-1-1 proteins are encoded by the genome assemblies JAAVMX010000001 and LKHE01001116 of the *H. sinensis* strains IOZ07 and 1229, respectively. Panel A shows an alignment of the amino acid sequences in the MAT α _HMGbox domains of the MAT1-1-1 proteins; amino acid substitutions are shown in red, whereas the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and in the 2D structures (Panels C–F show the α -helices, β -sheets, β -turns, and coils, respectively) of the proteins; the open rectangles in blue outline the changes in topology and waveform shown in the plots.

Compared with the reference MAT1-1-1 protein AGW27560 (*cf.* Figure S1, Table S6), no changes were detected in the amino acid sequences, hydrophobicities, or secondary structures (α -helices, β -sheets, β -turns, and coils) of the MAT α _HMGbox domains of the MAT1-1-1 proteins encoded by the genome assembly ANOV01017390 (794←1228 & 1280←1369) of the *H. sinensis* strain Co18 or the metatranscriptome assembly OSIN7648 (151→675) of the *C. sinensis* insect–fungal complex collected from Deqin, Yunnan, China [48–49,59,68]. The MAT α _HMGbox domains of the genome- and metatranscriptome-

encoded MAT1-1-1 proteins clustered into Branch a1 in the Bayesian clustering tree (*cf.* Figure 1, Table S6).

The MAT α _HMGbox domain of the MAT1-1-1 protein encoded by the metatranscriptome assembly GAGW01008880 (714←1127) of the *C. sinensis* insect-fungi complex is truncated at its N-terminus, as shown in Figure S1 and Panel A of Figure S3 [64]. The remaining portion of the MAT α _HMGbox domain of this protein is 100% identical to the query protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [48], and it showed no apparent changes in hydrophobicity or secondary structure, as illustrated by the identical topology and waveform that can be observed in the ExPASy ProtScale plots for hydropathy, α -helices, β -sheets, β -turns, and coils shown in Panels B–F, respectively, of Figure S3. Although the MAT α _HMGbox domain of the metatranscriptome-encoded protein clustered into Branch a1 in the Bayesian clustering tree together with the 6 reference authentic MAT1-1-1 proteins (*cf.* Figure 1, Table S6), the truncated MAT α _HMGbox domain, which lacks 50 and 48 amino acid residues encoded by exons I and II of the *MAT1-1-1* gene, respectively, may not be able to form a core stereostructure containing 3 critical α -helices that support high-affinity DNA binding and full functionality in the regulation of gene transcription related to the sexual reproduction of *O. sinensis* [50,84–85].

The genome assemblies LWBQ00000000 and NGJJ00000000 and the transcriptome assembly GCQL00000000 of the *H. sinensis* strains ZJB12195, CC1406-20395, and L0106, respectively, do not contain genes or transcripts that encode MAT1-1-1 proteins [59–60,62].

3.7. Heteromorphic stereostructures of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins derived from wild-type *C. sinensis* isolates

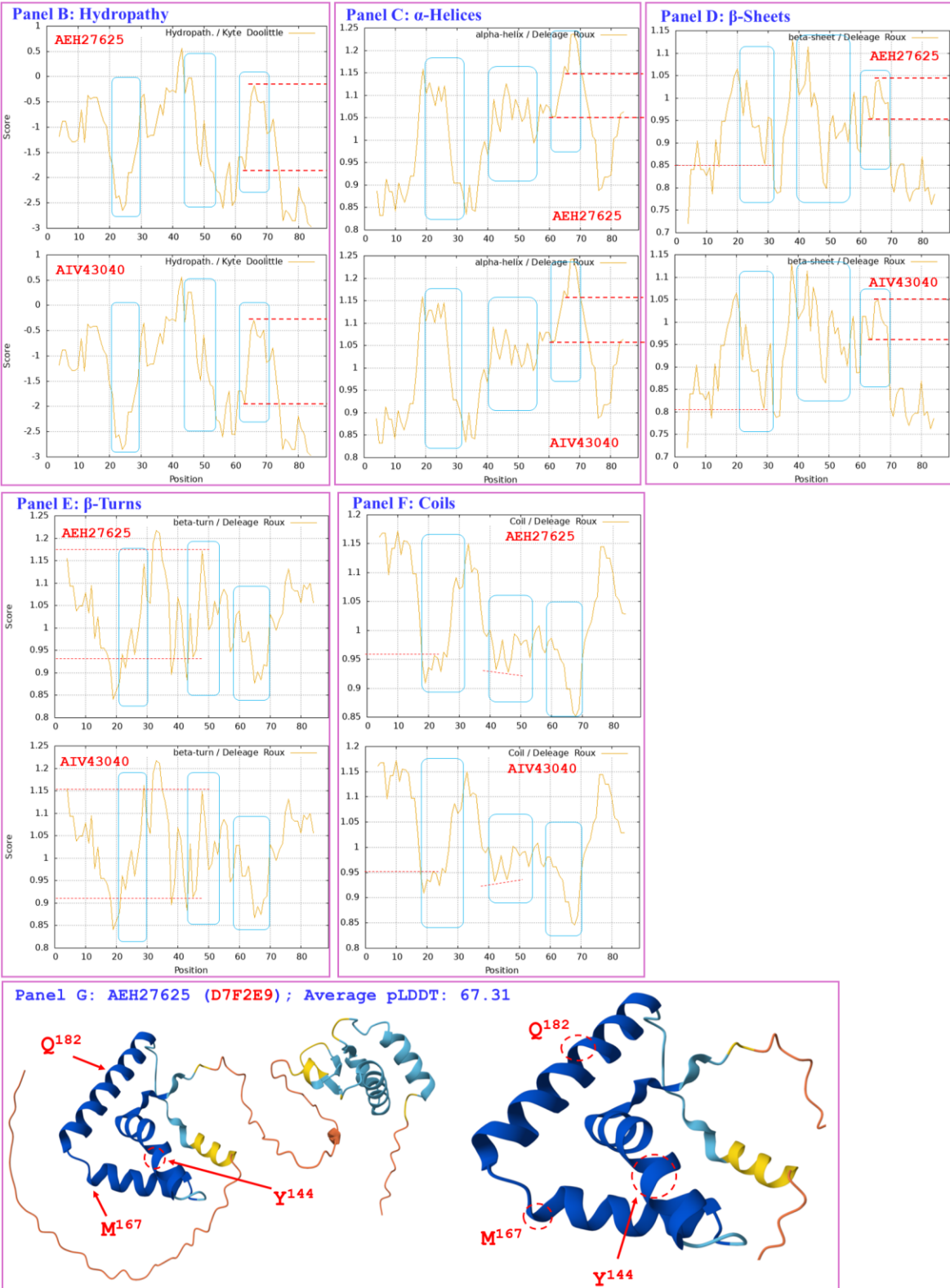
Section 3.7 presents analyses of the correlation between the changes in the hydrophobicity and structure of the HMG-box_ROX1-like domains of full-length MAT1-2-1 proteins (under 12 AlphaFold codes, *cf.* Table S7) derived from wild-type *C. sinensis* isolates compared with the reference MAT1-2-1 proteins represented by AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 (*cf.* Table S2) [56]. The HMG-box_ROX1-like domain sequences (amino acid residues 127→197) were extended both upstream and downstream by 9 external amino acid residues because the sequential ExPASy ProtScale plotting for hydropathy, α -helices, β -sheets, β -turns, and coils was set up at a window size of 9 amino acids (*cf.* Section 2.5 of Materials and Methods).

Figure 10 compares the hydrophobicity and structures of the HMG-box_ROX1-like domain of the MAT1-2-1 protein AIV43040 (under the AlphaFold code A0A0A0RCF5) derived from the *C. sinensis* isolate XZ12_16 (*cf.* Table S2) [5] with those of the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [56]. Panel A shows 3 amino acid substitutions: Y-to-H (aromatic tyrosine to basic histidine; hydropathy indices changed from -1.3 to -3.2 [68]), M-to-I (methionine, which possesses polar neutral side chains, to aliphatic isoleucine; hydropathy indices changed from 1.9 to 4.5) and Q-to-R (glutamine, which has polar neutral side chains, to basic arginine; hydropathy indices changed from -3.5 to -4.5). Individually, these substitutions caused decreased or increased hydrophobicity, as illustrated by the changes in topology and waveform in the hydropathy plot shown in Panel B. These changes resulted in altered secondary structures surrounding the mutation sites in the HMG-box_ROX1-like domain of the MAT1-2-1 protein AIV43040, as illustrated by the changes in topology and waveform in the ExPASy ProtScale plots for α -helices, β -sheets, β -turns, and coils shown in Panels C–F, respectively. As shown in the representations of 3D structures in Panels G–H, the replaced residues are located in each of the 3 core α -helices, and they alter the stability of the hydrophobic core of the HMG-box_ROX1-like domain [84–85] as well as the

stereostructure of the domain of the MAT1-2-1 protein AIV43040 under the AlphaFold code A0A0A0RCF5. The mutant HMG-box_ROX1-like domain clustered into Branch b2β, which displays the longest clustering distance in the Bayesian clustering tree (cf. Figure 2, Table S7).

Panel A: Amino acid sequence alignment

AEH27625	118	APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGAWNMESNDVRQKYK	177
AIV43040	118	-----H-----I-----	177
AEH27625	178	DMSQQVKQALLEKHPDYQYKPRRPCERRR	206
AIV43040	178	-----R-----	206



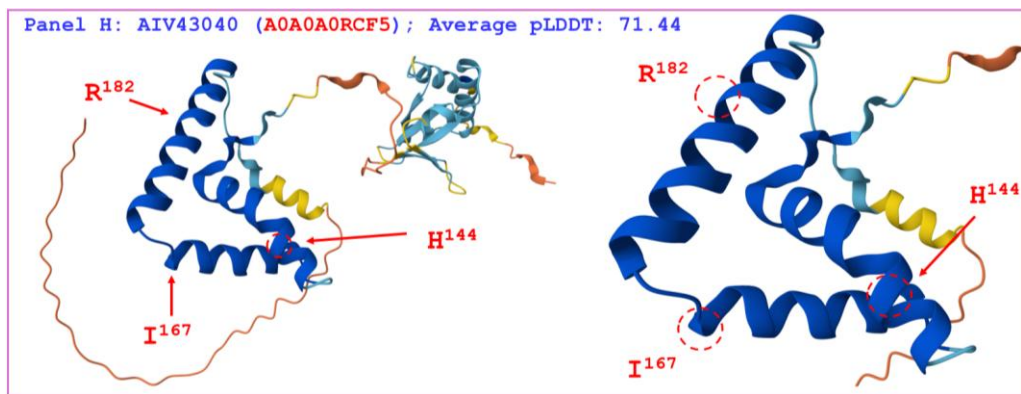


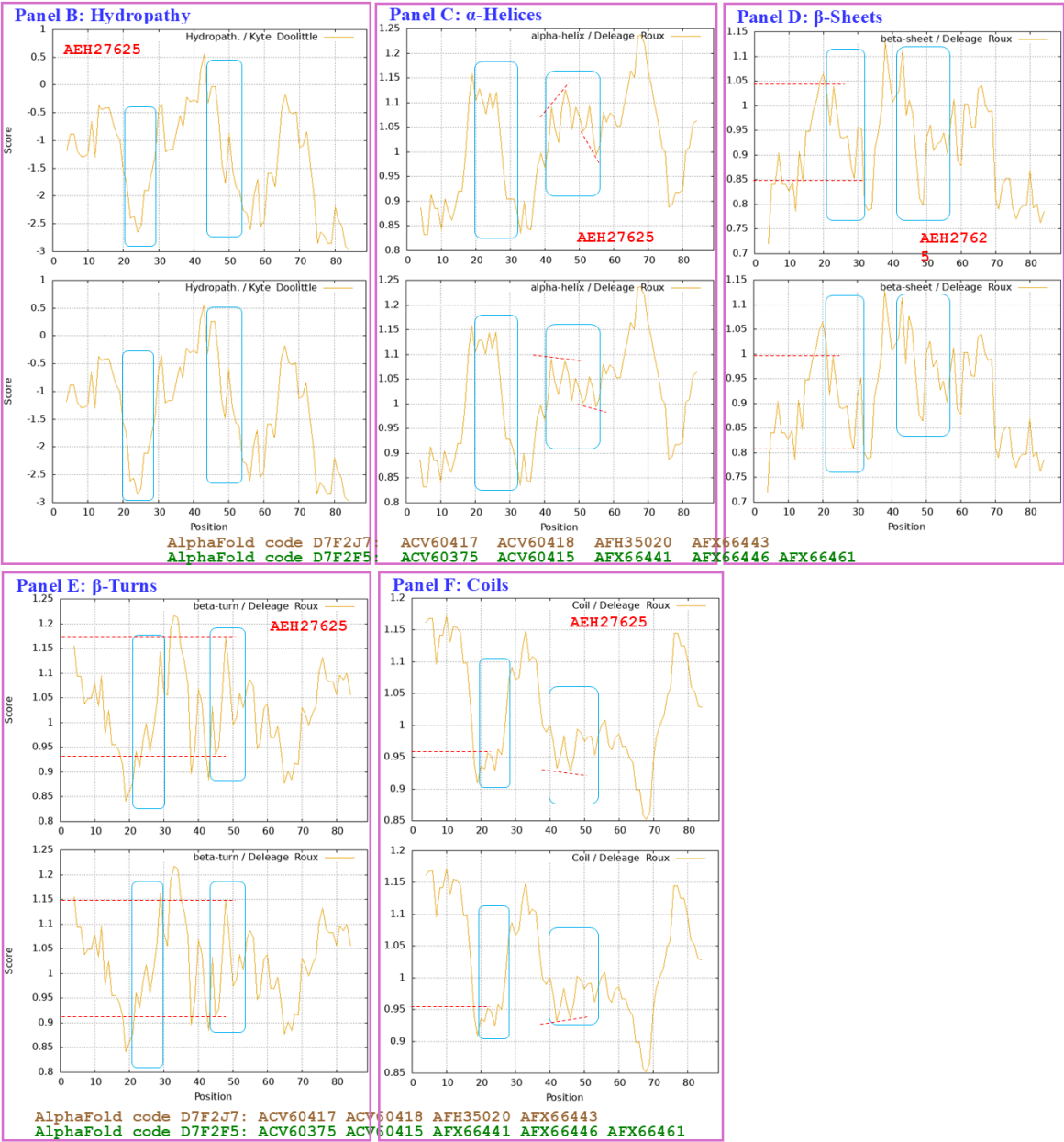
Figure 10. Correlation of the changes in the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins. The reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2, and the mutant MAT1-2-1 protein AIV43040 (under the AlphaFold code A0A0A0RCF5) is derived from the wild-type *C. sinensis* isolate XZ12_16. Panel A shows an alignment of the amino acid sequences of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins; amino acid substitutions are shown in red, whereas the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and in the 2D structure (Panels C–F show the α -helices, β -sheets, β -turns, and coils, respectively) of the protein; the open rectangles in blue outline the changes in topology and waveform seen in the plots. Panels G–H show representations of the 3D structures of the full-length proteins on the left; the locally amplified structures at the sites of substitutions are shown on the right. The model confidence values for the AlphaFold-predicted 3D structures are as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

Figure 11 compares the hydrophobicities and structures of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins under the AlphaFold code D7F2J7 (ACV60417, ACV60418, AFH35020, and AFX66443 shown in brown in Figure 11, derived from the wild-type *C. sinensis* isolates XZ-LZ07-H1, XZ-LZ07-H2, XZ06-124, and XZ05_8, respectively), and the hydrophobicities and structures of the proteins under the AlphaFold code D7F2F5 (ACV60375, ACV60415, AFX66441, AFX66446, and AFX66461 shown in green in Figure 11, derived from the wild-type *C. sinensis* isolates XZ-SN-44, XZ-LZ05-6, XZ05_2, XZ06_260, and XZ09_80, respectively) (cf. Table S2) [5,34,56] with those of the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [56]. Panel A shows substitutions of Y-to-H (aromatic tyrosine to basic histidine; hydropathy indices changed from -1.3 to -3.2 [68]) and M-to-I (methionine, which contains polar neutral side chains, to aliphatic isoleucine; hydropathy indices changed from 1.9 to 4.5). These substitutions caused changes in hydrophobicity, as illustrated by the changes in topology and waveform shown in the hydropathy plot in Panel B. They also resulted in altered secondary structures surrounding the mutation sites in the HMG-box_ROX1-like domains of the mutant MAT1-2-1 proteins ACV60417, ACV60418, AFH35020, and AFX66443 (under the AlphaFold code D7F2J7) and in those regions in the proteins ACV60375, ACV60415, AFX66441, AFX66446, and AFX66461 (under the AlphaFold code D7F2F5), as illustrated by the changes in topology and waveform shown in Panels C–F, which display the ExPASy ProtScale plots for α -helices, β -sheets, β -turns, and coils. As shown in the 3D structures presented in Panels G–I of Figure 11, the substituted residues are located within two of the 3 core α -helices, and they differentially alter the stability of the core structure of the HMG-box_ROX1-like domain [84–85] and the tertiary structure of this domain in the MAT1-2-1 proteins under the AlphaFold 3D structural codes D7F2J7 and D7F2F5. The sequences of the HMG-box_ROX1-like domain of the mutant proteins ACV60417, ACV60418, AFH35020, AFX66443, ACV60375, ACV60415, AFX66441,

AFX66446, and AFX66461 clustered into Branch b2α in the Bayesian clustering tree (cf. Figure 2, Table S7).

Panel A: Amino acid sequence alignment

AEH27625	118	APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK	177
ACV60417	118	-----H-----I-----	177
ACV60418	118	-----H-----I-----	177
AFH35020	118	-----H-----I-----	177
AFX66443	118	-----H-----I-----	177
ACV60375	118	-----H-----I-----	177
ACV60415	118	-----H-----I-----	177
AFX66441	118	-----H-----I-----	177
AFX66446	118	-----H-----I-----	177
AFX66461	118	-----H-----I-----	177
AEH27625	178	DMSQQVKQALLEKHPDYQYKPRRPCERRR	206
ACV60417	178	-----	206
ACV60418	178	-----	206
AFH35020	178	-----	206
AFX66443	178	-----	206
ACV60375	178	-----	206
ACV60415	178	-----	206
AFX66441	178	-----	206
AFX66446	178	-----	206
AFX66461	178	-----	206



756

757

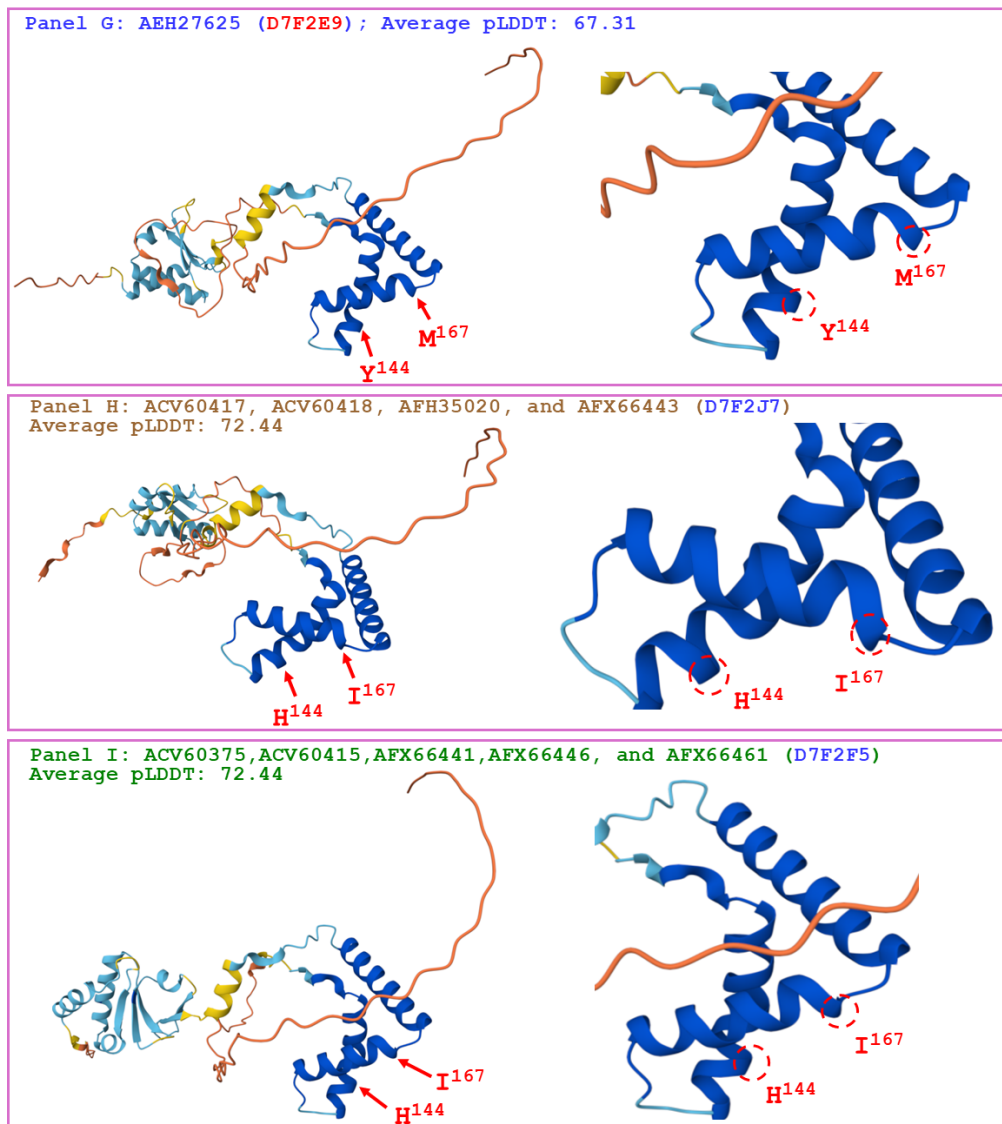


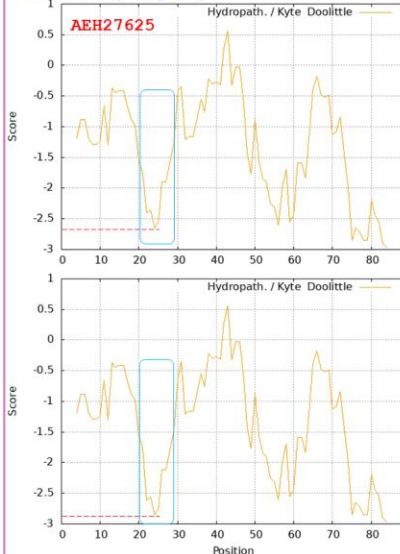
Figure 11. Correlation of the changes in the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins. The reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2; the mutant MAT1-2-1 proteins under the AlphaFold code D7F2J7 (ACV60417, ACV60418, AFH35020, and AFX66443 shown in brown) are derived from the wild-type *C. sinensis* isolate XZ-LZ07-H1, XZ-LZ07-H2, XZ06-124, and XZ05_8); and the mutant proteins under the AlphaFold code D7F2F5 (ACV60375, ACV60415, AFX66441, AFX66446, and AFX66461 shown in green) are derived from the wild-type *C. sinensis* isolates XZ-SN-44, XZ-LZ05-6, XZ05_2, XZ06_260, and XZ09_80. Panel A shows an alignment of the amino acid sequences of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins; amino acid substitutions are shown in red; the sequences displayed in brown and green represent the MAT1-2-1 proteins under the AlphaFold codes D7F2J7 and D7F2F5, respectively, and the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and in the 2D structures (Panels C–F for α -helices, β -sheets, β -turns, and coils, respectively) of the HMG-box_ROX1-like domains of the proteins; the open rectangles in blue outline the changes in topology and waveform seen in the plots. Panels G–H show representations of the 3D structures of the full-length proteins on the left and of the locally amplified structures at the sites of substitutions on the right. The model confidence values for the AlphaFold-predicted 3D structures are as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

Figure 12 compares the hydrophobicities and structures of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins under the AlphaFold codes **D7F2E3** (ACV60363, ACV60364, AFX66388, AFH35018, and AGW27542), **D7F2G5** (ACV60385), **V9LW71** (AFX66401), **V9LVS8** (AFX66472, AFX66473, and AFX66474), **V9LVU8** (AFX66475), **V9LWC9** (AFX66476), **V9LWG5** (AFX66484), and **U3N6V5** (AGW27537) derived from the wild-type *C. sinensis* isolates YN09_64, YN09_6, YN09_22, YN09_51, XZ-NQ-154, XZ-NQ-155, GS09_111, QH09-93, CS560-961, QH-YS-199, QH09_11, ID10_1, and CS6-251, respectively (cf. Table S2) [5,34,48,56], with the hydrophobicity and structure of the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [56]. Panel A shows Y-to-H substitutions (aromatic tyrosine to basic histidine; hydropathy indices changed from -1.3 to -3.2 [68]) in the HMG-box_ROX1-like domains of all the mutant MAT1-2-1 proteins mentioned above; these substitutions caused reduced hydrophobicity, as illustrated by the changes in topology and waveform shown in the hydropathy plots in Panel B. The substitutions resulted in alterations of the secondary structure surrounding the mutation sites in the HMG-box_ROX1-like domains of the MAT1-2-1 proteins, as illustrated by the changes in topology and waveform in the ExPASy ProtScale plots for α -helices, β -sheets, β -turns, and coils that can be seen in Panels C–F, respectively.

Panel A: Amino acid sequence alignment

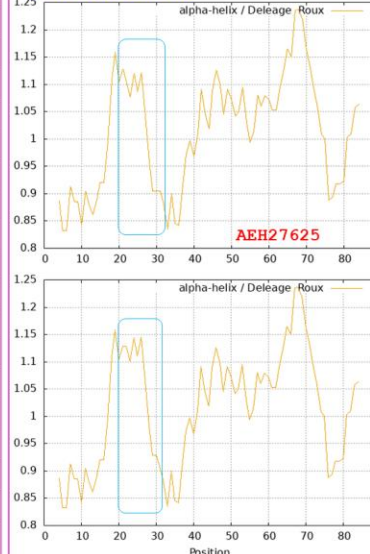
AEH27625	118	APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK	177
ACV60363	118	-----H-----	177
ACV60364	118	-----H-----	177
AFX66388	118	-----H-----	177
AFH35018	118	-----H-----	177
AGW27542	118	-----H-----	177
ACV60385	118	-----H-----	177
AFX66401	118	-----H-----	177
AFX66472	118	-----H-----	177
AFX66473	118	-----H-----	177
AFX66474	118	-----H-----	177
AFX66475	118	-----H-----	177
AFX66476	118	-----H-----	177
AFX66484	118	-----H-----	177
AGW27537	118	-----H-----	177
AEH27625	178	DMSQQVKQALLEKHPDYQYKPRRPCERRR	206
ACV60363	178	-----	206
ACV60364	178	-----	206
AFX66388	178	-----	206
AFH35018	178	-----	206
AGW27542	178	-----	206
ACV60385	178	-----	206
AFX66401	178	-----	206
AFX66472	178	-----	206
AFX66473	178	-----	206
AFX66474	178	-----	206
AFX66475	178	-----	206
AFX66476	178	-----	206
AFX66484	178	-----	206
AGW27537	178	-----	206

Panel B: Hydropathy

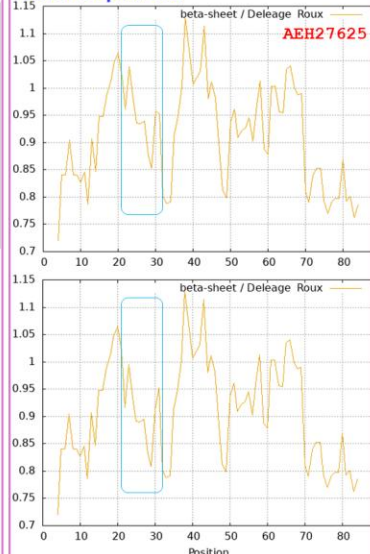


AlphaFold code D7F2E3: ACV60363 ACV60364 AFX66388 AFH35018 AGW27542
 AlphaFold code D7F2G5: ACV60385
 AlphaFold code V9LW71: AFX66401
 AlphaFold code V9LV88: AFX66472 AFX66473 AFX66474
 AlphaFold code V9LVU8: AFX66475
 AlphaFold code V9LVC9: AFX66476
 AlphaFold code V9LWG5: AFX66484
 AlphaFold code U3N6V5: AGW27537

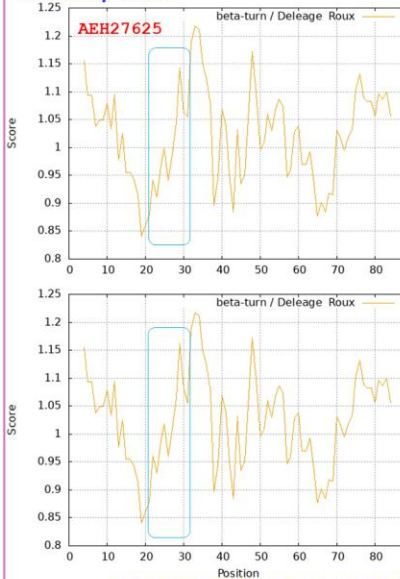
Panel C: α -Helices



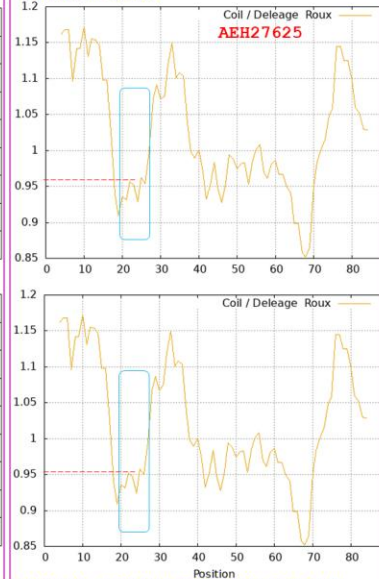
Panel D: β -Sheets



Panel E: β -Turns



Panel F: Coils



AlphaFold code D7F2E3: ACV60363 ACV60364 AFX66388 AFH35018 AGW27542
 AlphaFold code D7F2G5: ACV60385
 AlphaFold code V9LW71: AFX66401
 AlphaFold code V9LV88: AFX66472 AFX66473 AFX66474
 AlphaFold code V9LVU8: AFX66475
 AlphaFold code V9LVC9: AFX66476
 AlphaFold code V9LWG5: AFX66484
 AlphaFold code U3N6V5: AGW27537

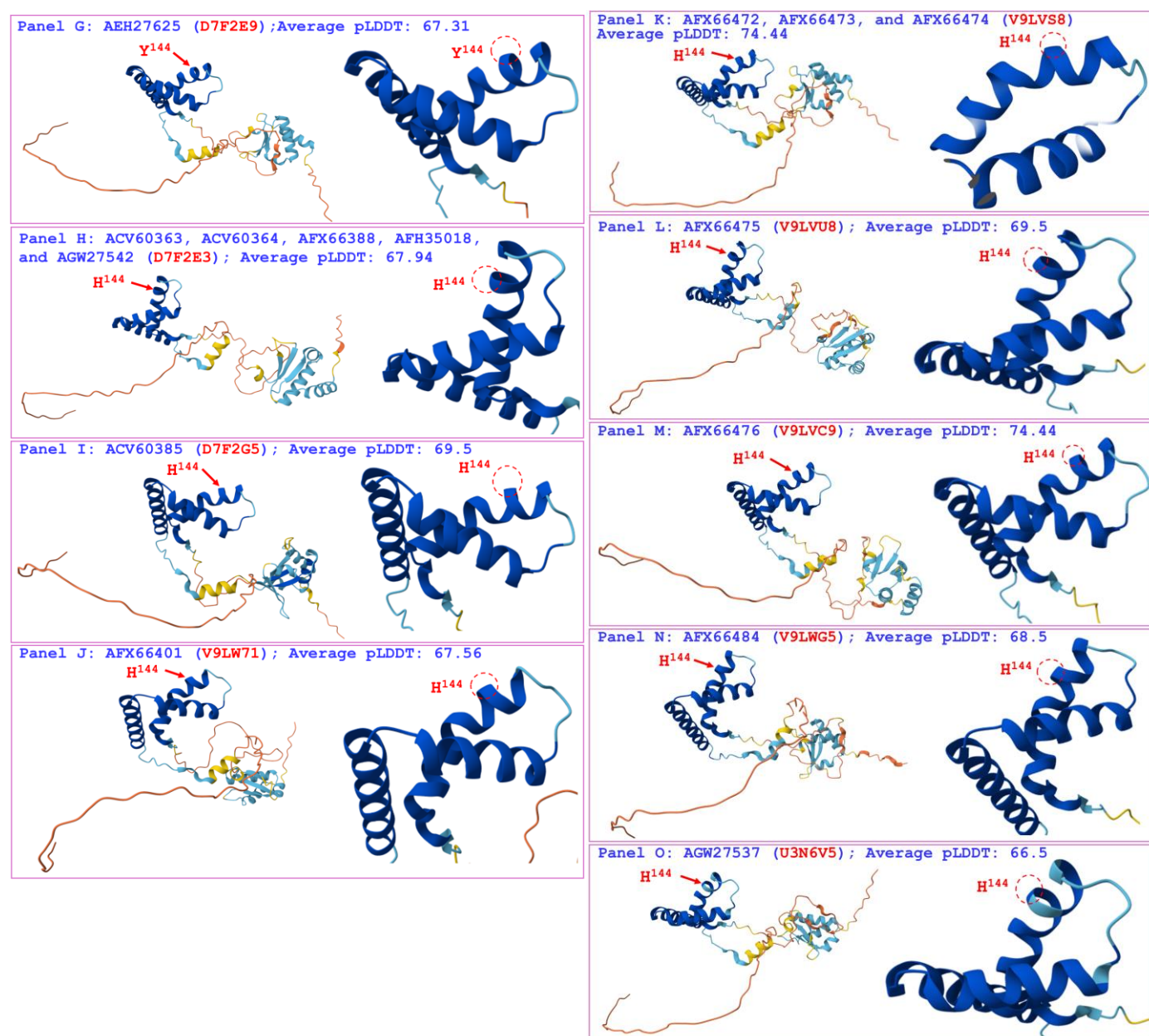


Figure 12. Correlation of the changes in the hydrophobicities and the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins: the reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2, and the mutant MAT1-2-1 proteins under the AlphaFold codes D7F2E3 (ACV60363, ACV60364, AFX66388, AFH35018, and AGW27542), D7F2G5 (ACV60385), V9LW71 (AFX66401), V9LVS8 (AFX66472, AFX66473, and AFX66474), V9LVU8 (AFX66475), V9LVC9 (AFX66476), V9LWG5 (AFX66484), and U3N6V5 (AGW27537) are derived from the wild-type *C. sinensis* isolates YN09_64, YN09_6, YN09_22, YN09_51, XZ-NQ-154, XZ-NQ-155, GS09_111, QH09-93, CS560-961, QH-YS-199, QH09_11, ID10_1, and CS6-251, respectively. Panel A shows an alignment of the amino acid sequences of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins; amino acid substitutions are shown in red, and the sequences depicted in various colors represent the MAT1-2-1 proteins under the AlphaFold codes V9LWC9, V9LVS8, D7F2E3, V9LVU8, D7F2G5, V9LW71, V9LWG5, and U3N6V5; the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and in the 2D structures (Panels C–F for α -helices, β -sheets, β -turns, and coils, respectively) of the proteins. The open rectangles in blue outline the changes in topology and waveform shown in the plots. Panels G–O show representations of the 3D structures of the full-length proteins on the left and the locally amplified structures at the mutation site on the

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

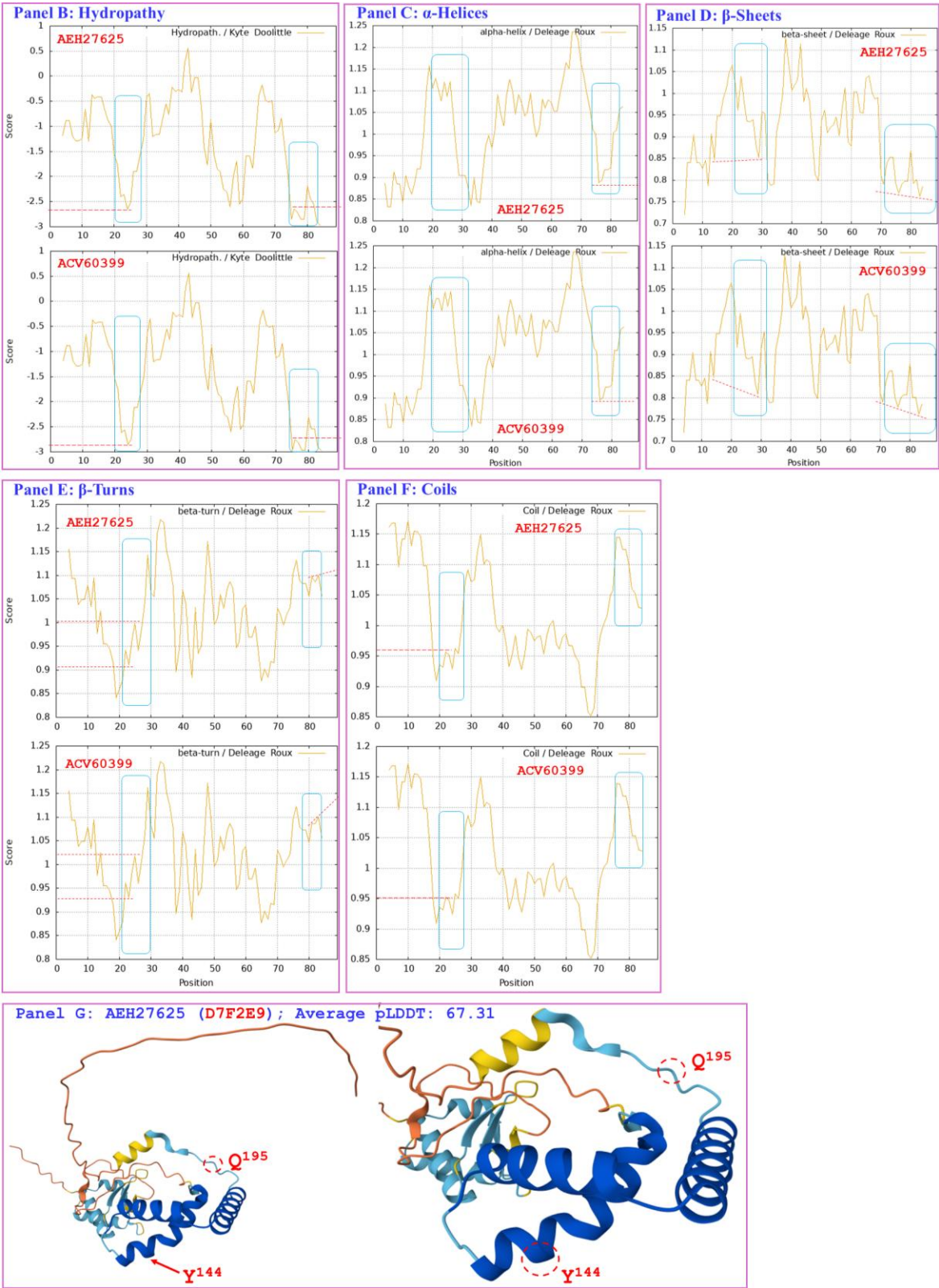
right. The model confidence values for the AlphaFold-predicted 3D structures are as follows:  very high (pLDDT>90);  high (90>pLDDT>70);  low (70>pLDDT>50); and  very low (pLDDT<50).

As shown in the representations of the tertiary structures presented in Panels G–O of Figure 12, the substituted residues with reduced hydrophobicity are located within the core α -helices in the HMG-box_ROX1-like domains, and they destabilize the hydrophobic core and alter the hydrophobic stereostructures of the domains [84–85]. Although Panel A of Figure 12 shows the same Y-to-H substitution that caused the changes in hydrophobicity and second-order structure (shown in the ExPASy ProtScale plots in Panels B–F) in the HMG-box_ROX1-like domains of the MAT1-2-1 proteins ACV60363–ACV60364, ACV60385, AFH35018, AFX66388, AFX66472–AFX66476, AFX66401, AFX66484, AGW27537, and AGW27542, additional amino acid substitutions outside the HMG-box_ROX1-like domains (*cf.* Figure S2) might play synergistic roles in altering the tertiary structures of the proteins through long-range spatial interactions with the mutations within the domains (Panels H–O of Figure 12) of the MAT1-2-1 proteins under the 8 AlphaFold codes summarized in Table S7. The mutant HMG-box_ROX1-like domains clustered into Branch b1 α in the Bayesian clustering tree (*cf.* Figure 2, Table S7).

Figure 13 compares the hydrophobicity and structure of the HMG-box_ROX1-like domain of the MAT1-2-1 protein ACV60399 (under the AlphaFold code D7F2H9) derived from the wild-type *C. sinensis* isolate SC-3 (*cf.* Table S2) [56] with those of the HMG-box_ROX1-like domain of the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [56]. Panel A shows substitutions of Y-to-H (aromatic tyrosine to basic histidine; hydropathy indices changed from -1.3 to -3.2 [68]) and Q-to-R (glutamine, which possesses polar neutral side chains, to basic arginine; hydropathy indices changed from -3.5 to -4.5); both of these substitutions caused reduced hydrophobicity, as illustrated by the changes in topology and waveform shown in the hydropathy plot in Panel B. These changes resulted in alterations in the secondary structure surrounding the mutation sites in the HMG-box_ROX1-like domain of the mutant MAT1-2-1 protein ACV60399, as illustrated by the changes in topology and waveform seen in the ExPASy ProtScale plots for α -helices, β -sheets, β -turns, and coils presented in Panels C–F. As shown in the representations of 3D structures in Panels G–H, the 2 replaced residues with reduced hydrophobicity are located inside and outside the hydrophobic core structure of the 3 α -helices, and they destabilize and alter the core stereostructure of the domain [84–85]. The HMG-box_ROX1-like domain of the mutant protein ACV60399 under the AlphaFold code D7F2H9 clustered into Branch b1 β in the Bayesian clustering tree (*cf.* Figure 2, Table S7).

Panel A: Amino acid sequence alignment

AEH27625	118	APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK	177
ACV60399	118	-----H-----	177
AEH27625	178	DMSQQVKQALLEKHPDYQYKPRRPCERRR	206
ACV60399	178	-----R-----	206



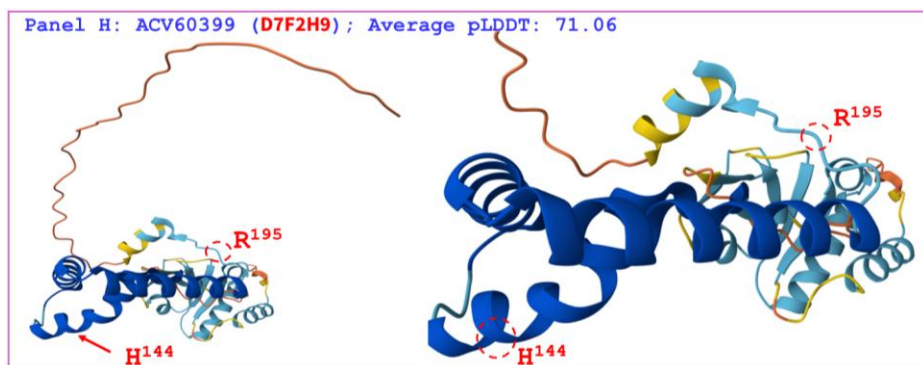


Figure 13. Correlation of the changes in the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins: the reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2, and the mutant MAT1-2-1 protein ACV60399 (under the AlphaFold code D7F2H9) is derived from the wild-type *C. sinensis* isolate SC-3. Panel A shows an alignment of the amino acid sequences of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins; amino acid substitutions are shown in red, whereas the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and in the 2D structure (Panels C–F for α -helices, β -sheets, β -turns, and coils) of the protein; the open rectangles in blue outline the changes in topology and waveform of the plots. Panels G–H show the 3D structures of the full-length proteins on the left and the locally amplified structures at the sites of substitutions on the right. The model confidence values for the AlphaFold-predicted 3D structures are as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

By aggregating the data presented in Section 3.7, Panels A in Figures 10–13 present the 1–3 amino acid substitutions detected in the HMG-box_ROX1-like domains that caused changes in the topologies and waveforms shown in the ExPASy ProtScale plots presented in Panels B–F of the figures. These changes indicate the presence of altered hydrophobicity and secondary structures (α -helices, β -sheets, β -turns, and coils) of the HMG-box_ROX1-like domains of the mutant MAT1-2-1 proteins. These changes resulted in alterations in the stereostructures of the domains of the mutant proteins under different AlphaFold codes, as illustrated in Panel H of Figures 10–13, Panel I of Figure 11, Panels I–O of Figure 12 and summarized in Table S7. The mutant HMG-box_ROX1-like domains clustered into various clades in the Bayesian clustering tree (*cf.* Figure 2; Table S7) and ultimately changed the DNA-binding specificities and affinities and the functionalities of gene transcriptional regulation, factors that play key roles in the sexual reproduction of *O. sinensis*.

3.8. Diverse primary and secondary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins encoded by the genome and transcriptome assemblies of *H. sinensis* and the metatranscriptome assembly of the *C. sinensis* insect–fungal complex

Figure 14 shows a comparison of the hydrophobicity and structure of the HMG-box_ROX1-like domain of the MAT1-2-1 protein encoded by the genome assembly ANOV01000063 (9759→9851 & 9907→10,026) derived from *H. sinensis* strain Co18 [49] with the hydrophobicity and structure of the reference MAT1-2-1 protein AEH27625 derived from *H. sinensis* strain CS2 (under the AlphaFold code D7F2E9) [56]. Panel A shows an S-to-A substitution (serine with polar neutral side chains to aliphatic alanine; hydrophathy indices changed from -0.8 to 1.8 [68]). This substitution caused increased hydrophobicity surrounding the mutation site in the domain of the MAT1-2-1 protein encoded by the genome assembly ANOV01000063, as illustrated by the changes in topology and waveform shown in the hydropathy plot in Panel B.

Panel A: Amino acid sequence alignment

AEH27625118

ANOV010000639732

APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK

-----A-----

AEH27625178

ANOV010000639968

DMSQQVKQALLEKHPDYQYKPRRPCERRR

206

10,164

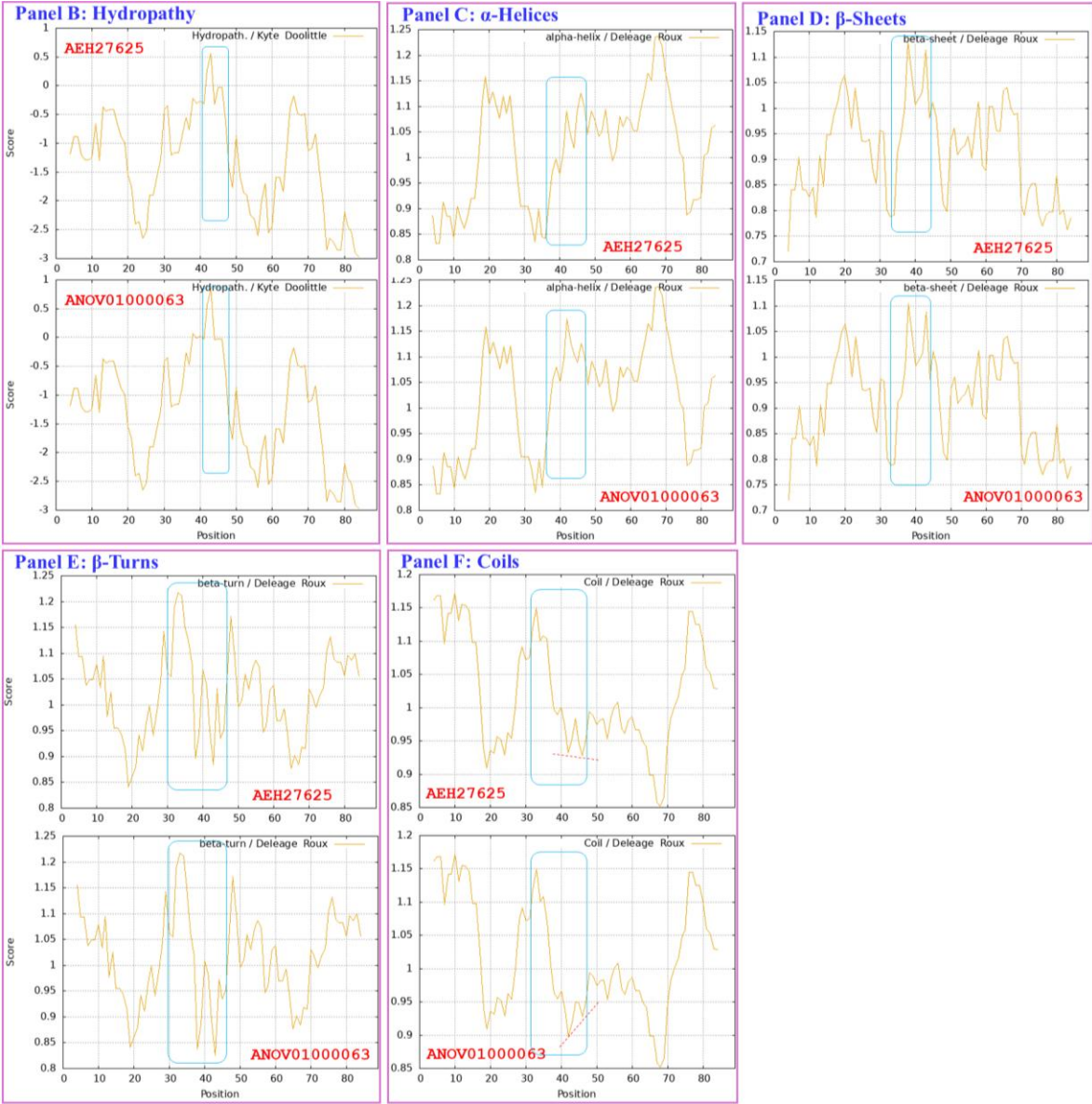


Figure 14. Correlation of the changes in the primary and secondary structures of the HMG-box_ROX1-like domains of the reference MAT1-2-1 protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 and the mutant MAT1-2-1 protein encoded by the genome assembly ANOV01000063 derived from the *H. sinensis* strain Co18. Panel A shows an alignment of the amino acid sequences of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins; amino acid substitutions are shown in red, whereas the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and in the 2D structure (Panels C–F for α -helices, β -sheets, β -turns, and coils, respectively) of the protein; the open rectangles in blue outline the changes in topology and waveform of the plots.

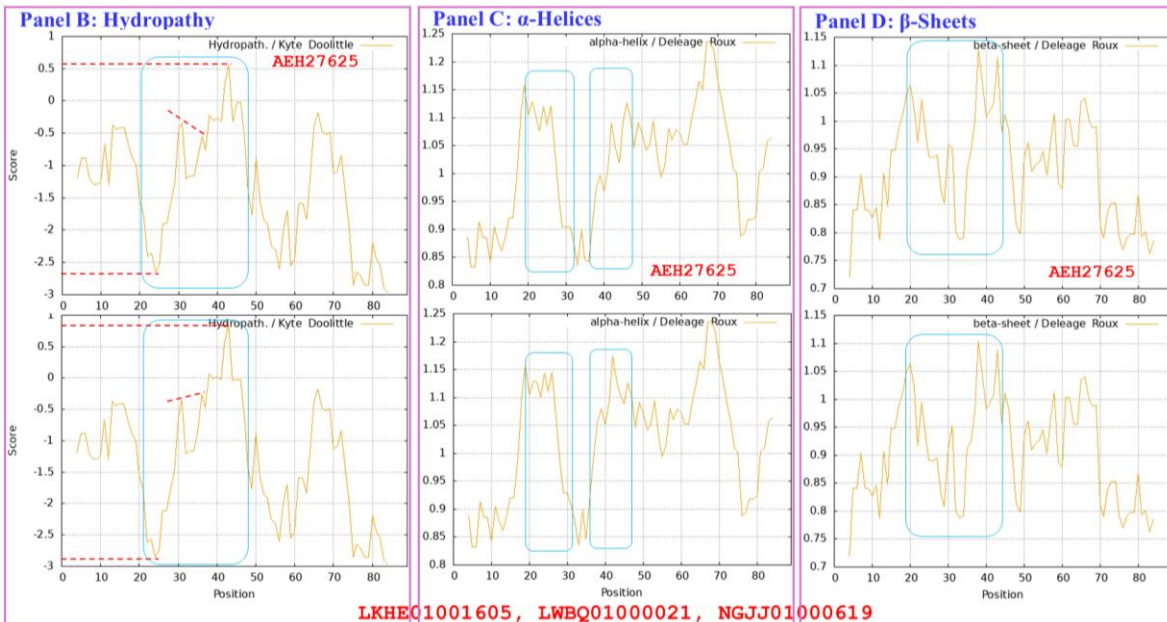
These changes in amino acid sequence and hydrophobicity resulted in altered secondary structure (α -helices, β -sheets, β -turns, and coils) within the HMG-box_ROX1-like domain of the mutant MAT1-2-1 protein encoded by the genome assembly ANOV01000063, as illustrated in the ExPASy ProtScale plots for α -helices, β -sheets, β -turns, and coils shown in Panels C–F, respectively, in Figure 14 [84–85]. These changes in

sequence and hydrophobicity may have altered the tertiary structure of the HMG-box_ROX1-like domain of the genome-encoded protein. The altered HMG-box_ROX1-like domain clustered into Branch a2 in the Bayesian clustering tree (cf. Figure 2; Table S8).

Figure 15 compares the hydrophobicities and structures of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins encoded by the genome assemblies LKHE01001605 (14,016←14,135 & 14,191←14,283), LWBQ01000021 (14,016←14,135 & 14,191←14,283), and NGJJ01000619 (23,186←23,305 & 23,361←23,453) of the *H. sinensis* strains 1229, ZJB12195, and CC1406-20395, respectively [60–62], with the hydrophobicity and structure of the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [56]. Panel A shows Y-to-H (aromatic tyrosine to basic histidine; hydropathy indices changed from -1.3 to -3.2 [68]) and S-to-A substitutions (serine, which has polar neutral side chains, to aliphatic alanine; hydropathy indices changed from -0.8 to 1.8), which resulted in altered hydrophobicity, as illustrated in the hydropathy plot in Panel B. These changes altered the secondary structures surrounding the mutation sites in the HMG-box_ROX1-like domains of the mutant MAT1-2-1 proteins encoded by the genome assemblies LKHE01001605, LWBQ01000021, and NGJJ01000619, as illustrated in the ExPASy ProtScale plots for α -helices, β -sheets, β -turns, and coils shown in Panels C–F, respectively [84–85], and subsequently altered the tertiary structures of the core stereostructures of the domains. The mutant HMG-box_ROX1-like domains of the genome-encoded proteins clustered into Branch b3 in the Bayesian clustering tree (cf. Figure 2, Table S8).

Panel A: Amino acid sequence alignment

AEH27625	118	APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK	177
LKHE01001605	14,310	-----H-----A-----	14,076
LWBQ01000021	239,323	-----H-----A-----	239,089
NGJJ01000619	23,480	-----H-----A-----	23,246
AEH27625	178	DMSQQVKQALLEKHPDYQYKPRRPCERRR	206
LKHE01001605	14,075	-----	13,989
LWBQ01000021	239,088	-----	239,002
NGJJ01000619	23,245	-----	23,159



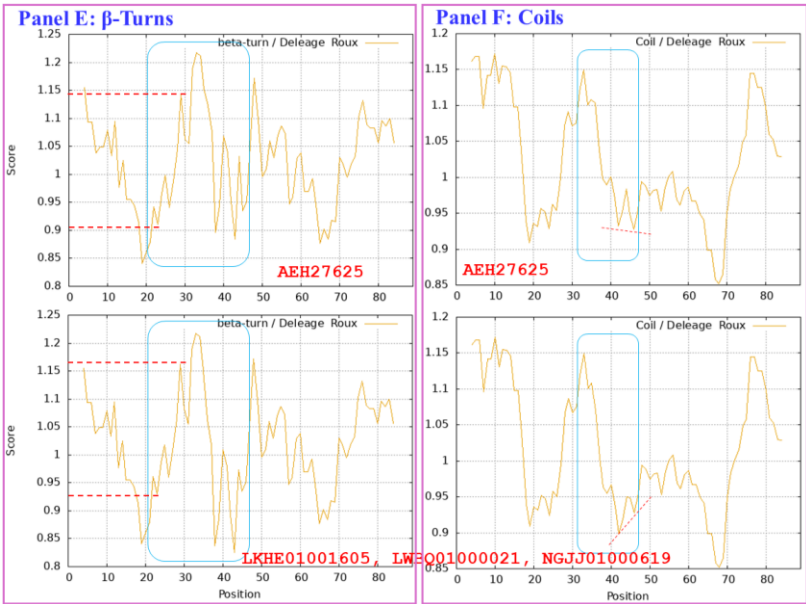


Figure 15. Correlation of the changes in the primary and secondary structures of the HMG-box_ROX1-like domain of the reference MAT1-2-1 protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 and the mutant MAT1-2-1 proteins encoded by the genome assemblies LKHE01001605, LWBQ01000021, and NGJJ01000619 derived from *H. sinensis* strains 1229, ZJB12195, and CC1406-20395, respectively. Panel A shows an alignment of the amino acid sequences of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins; amino acid substitutions are shown in red, whereas the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and in the 2D structures (Panels C–F for α -helices, β -sheets, β -turns, and coils, respectively) of the proteins; the open rectangles in blue outline the changes in topology and waveform seen in the plots.

The genome assembly JAAVMX000000000 of the *H. sinensis* strain IOZ07 does not contain a gene encoding the MAT1-2-1 protein [63].

Figure 16 compares the hydrophobicity and structure of the HMG-box_ROX1-like domain of the MAT1-2-1 proteins encoded by the transcriptome assembly GCQL01020543 (553–765) of the *H. sinensis* strain L0106 and by the metatranscriptome assembly OSIN7649 (379–591) of the mature *C. sinensis* insect–fungal complex [59,65] with the hydrophobicity and structure of the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [56]. Panel A shows Y-to-H substitutions (aromatic tyrosine to basic histidine; hydropathy indices changed from -1.3 to -3.2 [68]). The substitutions caused decreased hydrophobicity, as illustrated by the changes in topology and waveform shown in the hydropathy plot in Panel B. These changes resulted in alterations in the secondary structure surrounding the mutation sites in the HMG-box_ROX1-like domains of the MAT1-2-1 proteins encoded by the transcriptome assembly GCQL01020543 and the metatranscriptome assembly OSIN7649, as illustrated by the changes in topology and waveform shown in the ExPASy ProtScale plots for α -helices, β -sheets, β -turns, and coils in Panels C–F.

Panel A: Amino acid sequence alignment

AEH27625	118	APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK	177
GCQL01020543	792	-----H-----	613
OSIN7649	352	-----H-----	531
AEH27625	178	DMSQQVKQALLEKHPDYQYKPRRPCERRR	206
GCQL01020543	612	-----	526
OSIN7649	532	-----	618

932
933
934
935
936
937
938
939
940
941
942

943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958

959

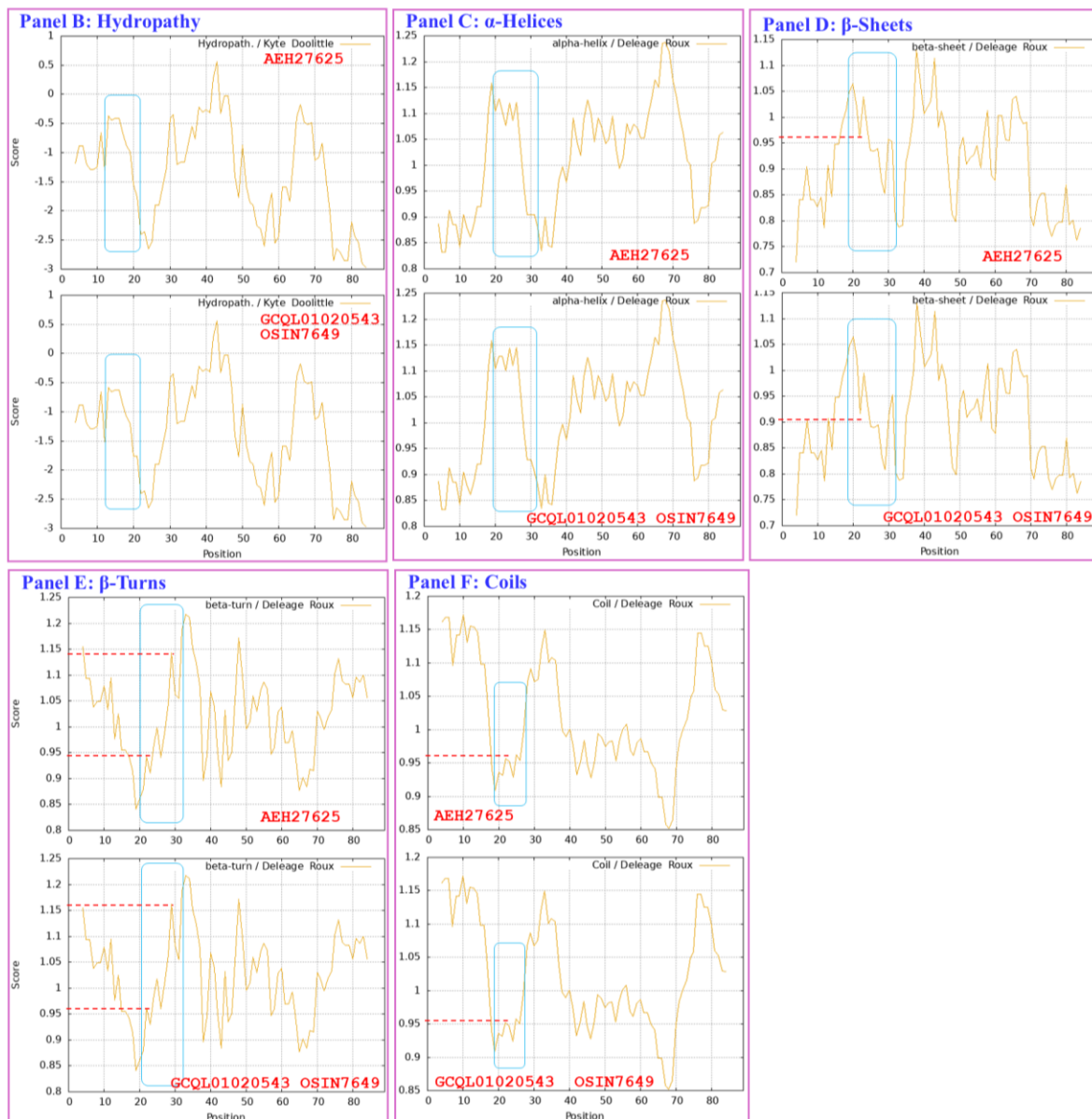


Figure 16. Correlation of the primary and secondary structures of the HMG-box_ROX1-like domain of the reference MAT1-2-1 protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 with the primary and secondary structures of the mutant MAT1-2-1 proteins encoded by the transcriptome assembly GCQL01020543 derived from the *H. sinensis* strain L0106 and the metatranscriptome assembly OSIN7649 (379→591) of the mature *C. sinensis* insect–fungal complex. Panel A shows an alignment of the amino acid sequences of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins; amino acid substitutions are shown in red, whereas the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and in the 2D structures (Panels C–F for α -helices, β -sheets, β -turns, and coils, respectively) of the proteins; the open rectangles in blue outline the changes in topology and waveform seen in the plots.

Similar to the Y-to-H substitutions in the HMG-box_ROX1-like domains of the mutant MAT1-2-1 proteins under various AlphaFold 3D structural models (such as V9LWC9, V9LVS8, D7F2E3, V9LVU8, D7F2G5, V9LW71, V9LWG5, and U3N6V5), as shown in Figure 12, the changes in hydrophobicities and primary and secondary structures shown in Figure 16 may subsequently alter the tertiary structures of the HMG-box_ROX1-like domains of the transcriptome-encoded and metatranscriptome-encoded mutant proteins. The mutant HMG-box_ROX1-like domains of the transcriptome- and metatranscriptome-

encoded proteins clustered into Branch b1 α in the Bayesian clustering tree (cf. Figure 2, Table S8).

3.9. Heterogenous fungal sources of the MAT1-1-1 and MAT1-2-1 proteins

Although wild-type *C. sinensis* isolates are often considered impure *H. sinensis*, analysis of mutant full-length mating proteins revealed their heterogeneous fungal sources, as shown in Table S3. Of the wild-type *C. sinensis* isolates whose full-length MAT1-1-1 and MAT1-2-1 proteins were analyzed (cf. Figures 3–8 and 10–13), 34 had ITS genotyping information in the GenBank database and belonged to GC-biased Genotypes #1 and #3 of genome-independent *O. sinensis* fungi (Table S3) [7]. Although the ITS sequences of 29 of the 34 wild-type *C. sinensis* isolates appeared to show high homology ($\geq 97\%$) to the ITS sequences of GC-biased Genotype #1 *H. sinensis*, the mutant MAT1-1-1 and MAT1-2-1 proteins might not necessarily be derived from GC-biased Genotype #1 *H. sinensis* within the impure fungal complex containing more than one fungal species [86], whereas the *H. sinensis* ITS sequences were most easily amplified via regular simple-step PCR under the experimental settings unless the PCR experiments used multiple pairs of primers combined with amplicon cloning techniques with selection and sequencing of sufficient (usually 30–50) white colonies. In other words, because of the impure nature of the wild-type *C. sinensis* isolates [7,86], the mutant proteins may be derived from cooccurring unvalidated heterospecific or genotypic fungi.

4. Discussion

4.1. Correlation of changes in the hydrophobic properties and the primary, secondary, and tertiary structures of the DNA-binding domains of MAT1-1-1 and MAT1-2-1 proteins

The MAT α _HMGbox domain of the MAT1-1-1 protein and the HMG-box_ROX1-like domain of the MAT1-2-1 protein are distinct DNA-binding domains that are critical for fungal mating-type regulation but differ in their evolutionary lineage, structural architecture, and functional roles [87–89]. The HMGbox domains in mating proteins are multifunctional motifs central to the transcriptional regulation of mating-related genes, *O. sinensis* sexual reproduction, and host adaptation in the *C. sinensis* insect–fungal complex. Our study demonstrates that various 1–4 amino acid substitutions within the MAT α _HMGbox and HMG-box_ROX1-like domains of the MAT1-1-1 and MAT1-2-1 proteins, respectively, of the wild-type *C. sinensis* isolates and *C. sinensis* insect–fungal complexes result in changes in the hydrophobic properties and in the secondary and tertiary structures of the DNA-binding domains of the proteins (cf. Figures 3–16). The variable sequences of the DNA-binding domains clustered into different clades in the Bayesian clustering trees (cf. Figures 1–2; Tables S5–S8).

The MAT α _HMGbox domain of the MAT1-1-1 protein contains a core structure of 3 α -helices that ensures high-affinity, sequence-dependent binding to AT-rich DNA motifs [44–46]. This binding facilitates the assembly of the transcriptional machinery in a way that involves mating-type-specific genes and activates gene transcription by inducing DNA conformational changes through physical bending or twisting of the target DNA segments [46,90–91]. This domain directly participates in the promoter recognition of specific genes related to the heterothallic reproductive process [92].

Table S4 shows the hydropathy values for amino acids that are defined by their side chains according to Kyte & Doolittle [68] (cf. <https://web.expasy.org/protscale/>). Larger hydropathy values indicate greater hydrophobicity, whereas negative values indicate hydrophilicity [68,84–85]. The core α -helical stereostructures of the MAT α _HMGbox domains are stabilized through a hydrophobic core that is composed of nonpolar amino acids such as isoleucine (Ile, I), valine (Val, V), leucine (Leu, L), phenylalanine (Phe, F),

cysteine (Cys, C), methionine (Met, M), and alanine (Ala, A); these amino acids have hydropathy indices of 4.5, 4.2, 3.8, 2.8, 2.5, 1.9, and 1.8, respectively (*cf.* Table S4) [68].

The HMG-box_ROX1-like domain of the MAT1-2-1 protein also consists of 3 asymmetrically arranged α -helices that are connected by flexible loops, forming a key hydrophobic core that is stabilized through the presence of the aforementioned nonpolar amino acids (*cf.* Table S4). The HMG-box_ROX1-like domain of the MAT1-2-1 protein regulates gene expression in a sequence-specific manner by binding to ATTAAT or ATTGTT motifs [87–89]. This domain has been implicated in chromatin remodeling and transcriptional repression through its interaction with other corepressors, similar to the repressor of oxygen-regulated genes 1 (ROX1) domain in yeast, through opposing pathways to the MAT α _HMGbox domain of the MAT1-1-1 protein [88,93–94]. ROX1 represses its own promoter, creating a feedback loop for hypoxic response regulation as a central hypoxia-responsive regulator that plays roles in chromatin remodeling following DNA binding and in the integration of metabolic signals [95].

To date, no studies have directly addressed heterodimerization between the MAT1-1-1 and MAT1-2-1 proteins of *O. sinensis*, although the formation of heterodimers between the MAT α _HMGbox domain of the MAT1-1-1 protein and the HMG-box_ROX1-like domain of the MAT1-2-1 protein is believed to be a critical mechanism in the regulation of fungal mating, particularly in heterothallic ascomycetes [96–98]. However, the MAT α _HMGbox and HMG-box_ROX1-like domains likely interact with each other, enabling them to cooperate synergistically in DNA binding through their complementary electrostatic surfaces [99]. The heteromorphous stereostructures of the HMGbox domains of the mutant mating proteins derived from wild-type *C. sinensis* isolates presented in this study likely alter the complementary cooperation of the mating proteins of the self-sterile *O. sinensis* under heterothallic or hybrid outcrossing and affect the subsequent functionalities of the activation or repression of mating-related gene transcription [50,54–55].

4.2. Heterogenous fungal sources of the MAT1-1-1 and MAT1-2-1 proteins

The analysis of mutant full-length mating proteins derived from wild-type *C. sinensis* isolates (Figures 3–8 and 10–13) revealed their heterogeneous fungal sources (Table S3) [5,34,36,52,56–57]. A total of 34 wild-type *C. sinensis* isolates were examined for molecular taxonomy ITS sequences that are available in GenBank and were found to belong to genomically independent GC-biased Genotypes #1 and #3 of *O. sinensis* fungi [7,10,34,52,56]. However, whether the impure wild-type *C. sinensis* isolates that cooccur with other heterospecific fungi remain to be determined through the use of culture-dependent techniques or culture-independent PCR protocols involving the use of multiple pairs of primers combined with amplicon cloning techniques with the selection and sequencing of sufficient (usually ≥ 30) white colonies.

Li *et al.* [86] conducted a similar *C. sinensis* isolate study and reported 2 other wild-type *C. sinensis* isolates, CH1 and CH2. With the use of multiple pairs of PCR and rigorous amplicon cloning techniques, these isolates were proven to coexist with GC-biased Genotype #1 and AT-biased Genotypes #4–5 of *O. sinensis* and heterospecific *Paecilomyces hepiali*. Li *et al.* [7,9–10,54–55] further confirmed that the ITS sequences of Genotypes #2–17 of *O. sinensis* with multiple transition and transversion point mutations are not repetitive copies in the genome of GC-biased Genotype #1 *H. sinensis* but belong to independent *O. sinensis* fungi, regardless of whether they are GC- or AT-biased. Although they display the key genetic features of fungal complexes, these wild-type isolates phenotypically exhibit *in vitro* growth and microscopic morphological characteristics of psychrophilic *H. sinensis*. Functionally, in the inoculation experiments on larvae of *Hepialus armoricanus* ($n=100$ larvae per study group), the infection rate significantly increased from nearly non-infectious values of 1–3% for *H. sinensis* to 55.2% for the wild-type isolates CH1 and CH2,

representing an up to 39-fold enhancement ($P < 0.001$), and the larval death latency largely decreased from 35–50 days for *H. sinensis*.

4.3. The MAT1-1-1 and MAT1-2-1 proteins are involved in the sexual reproduction of *O. sinensis*

The MAT1-1-1 and MAT1-2-1 proteins coregulate mating compatibility and sexual development during the sexual reproduction of *O. sinensis*, a process in which the MAT α _HMGbox and HMG-box_ROX1-like domains play key roles. The hypothesis of self-fertilization via a homothallic or pseudohomothallic strategy has been proposed for *O. sinensis* [48–49] on the basis of genetic research. However, Zhang and Zhang [52] argued against this hypothesis and suggested that *O. sinensis* uses facultative hybridization in its sexual reproduction on the basis of findings of differential occurrence of the MAT1-1-1 and MAT1-2-1 genes in more than 170 wild-type *C. sinensis* isolates derived from natural *C. sinensis* insect–fungal complex specimens collected from different production areas on the Qinghai–Tibet Plateau. Furthermore, Li *et al.* [50,54–55] reported differential occurrence, differential translation, and alternative splicing of MAT1-1-1 and MAT1-2-1 and pheromone receptor genes and heteromorphic stereostructures of the MAT1-1-1 and MAT1-2-1 proteins of wild-type *C. sinensis* isolates. In addition, the analysis of DNA-binding domains presented in this study further revealed diverse stereostructures of the MAT α _HMGbox domain of the MAT1-1-1 protein and the HMG-box_ROX1-like domain of the MAT1-2-1 protein. These findings demonstrated that *O. sinensis* experiences self-sterility and uses a heterothallic or hybrid strategy, even parasexuality, to accomplish sexual reproduction during the lifecycle of the *C. sinensis* insect–fungal complex [10,50,54–55,100–107].

Unlike homothallic reproduction, heterothallic or hybrid reproduction of *O. sinensis* requires mating partners. Among the 17 genotypes of *O. sinensis* [7,14,16–17,23,100,108–115], GC-biased Genotype #1 *H. sinensis* has been proposed as the sole anamorph of *O. sinensis* [31]. Wei *et al.* [32] reported successful industrial cultivation of *C. sinensis* insect–fungal complexes; however, the cultivated complexes presented a species contradiction between the anamorphic inoculants of 3 *H. sinensis* strains and the sole teleomorph of AT-biased Genotype #4 that was detected in the fruiting body of the cultivated *C. sinensis* insect–fungal complexes. The sequences of AT-biased Genotypes #4–6 and #15–17, as well as the sequences of GC-biased Genotypes #2–3 and #7–14 of *O. sinensis*, do not reside in the genome assemblies ANOV000000000, JAAVMX000000000, LKHE000000000, LWBQ000000000, or NGJJ000000000 of the *H. sinensis* strains Co18, IOZ07, 1229, ZJB12195, and CC1406-20395, respectively [7–9,17,23,49,60–63,113,116,]. Li *et al.* [116] further demonstrated that the sequences of Genotypes #2–17 of *O. sinensis* did not appear as repetitive copies in the genome of Genotype #1 *H. sinensis*, invalidating the hypothesis proposed by Li *et al.* [23–24,117] that “RIP mutation” induces or generates “ITS pseudogenes” and “rRNA pseudogenes” in the genome of *H. sinensis*. These data indicate that the 17 genotypes of *O. sinensis* are genomically independent and that they belong to different fungi. In addition, the data demonstrate that the sole anamorph hypothesis for *H. sinensis* proposed by Wei *et al.* [31] fails to pass examination via Koch’s postulates.

Li *et al.* [118–119] revised the inoculant information for the industrial cultivation project and reported that the cultivation project used cultures of *C. sinensis* ascospores rather than pure *H. sinensis* strains as inoculants, as previously reported by Wei *et al.* [32]. However, Li *et al.* [9–10] reported the detection of GC-biased Genotypes #1 and #14 and AT-biased Genotypes #5–6 and #16, as well as *P. hepiali*, in ascospores of natural *C. sinensis* insect–fungal complexes through a culture-independent approach using multiple pairs of primers and a PCR amplicon cloning technique with the selection and sequencing of ≥ 30 white colonies. In addition, Li *et al.* [23] used a culture-dependent protocol and reported

the detection of GC-biased Genotype #1 and AT-biased Genotype #5 in 8 cultures of mono-ascospores. Li *et al.* [86] demonstrated the nearly noninfectious feature of *H. sinensis* conidia and mycelia on 200 larvae of *H. armoricus* in inoculation experiments. Furthermore, Li *et al.* [7,9–10] reported that AT-biased Genotype #4 of *O. sinensis* occurred in the stromata throughout the entire course of *C. sinensis* maturation and that the stromal fertile portion (SFP) contained numerous ascocarps of the *C. sinensis* insect–fungal complex; the abundance of Genotype #4 was high in immature stromata, which are in asexual growth stages, and largely declined in mature stromata and SFP, which are in sexual production stages or in the transitional stages from asexual growth to sexual reproduction. Unfortunately, Genotype #4 of *O. sinensis* was absent from the *C. sinensis* ascospores. Thus, although Wei *et al.* [32] reported the detection of the sole teleomorphic AT-biased Genotype #4 in the cultivated *C. sinensis* insect–fungal complex and Li *et al.* [118–119] later revised the inoculant information, the fungal source of the sole teleomorphic, AT-biased Genotype #4 of *O. sinensis* remains scientifically uncertain in the artificial cultivation of the *C. sinensis* insect–fungal complex.

In addition to the formal reports of unsuccessful attempts at artificial cultivation of the fruiting bodies of *C. sinensis* [48,120–121], Zhang *et al.* [36] summarized the 40-year history of cultivation failures of insect–fungal complexes via a “pure” mycological strategy in academic research-oriented settings, and Qin *et al.* [122] summarized the obstacles to the cultivation of *O. sinensis* fruiting bodies and ascospores. In contrast, Wei *et al.* [32] reported success in such a cultivation effort in industrial product-oriented settings. This industrial success might be attributed to the application of a “mycologically impure” cultivation strategy on the basis of at least two facts:

(1) Cultures of mycologically impure ascospores of natural *C. sinensis* are used as inoculants [118–119]; these are most likely combined with cocultures of *C. sinensis* stroma and/or a stromal fertile portion (SFP) containing numerous ascocarps that, most importantly, contain AT-biased Genotype #4 of *O. sinensis*, which is consistent with the discovery of the sole teleomorphic *O. sinensis* genotype [32];

(2) Soil collected from the natural *C. sinensis* production areas on the Qinghai–Tibet Plateau was added to the industrial cultivation system as described by Wei *et al.* [32].

Unfortunately, the purification and genomic sequencing of GC- and AT-biased Genotypes #2–17 of *O. sinensis* have not been reported to date. Thus, genomic and transcriptomic information regarding mutations in the *MAT1-1-1* and *MAT1-2-1* genes in the genome-independent *O. sinensis* fungi of Genotypes #2–17, especially the DNA-binding domains of the *MAT1-1-1* and *MAT1-2-1* proteins, is lacking. Table S3 shows that the *MAT1-1-1* proteins ALH25057, ALH25005, and ALH25006 and the *MAT1-2-1* proteins AIV43040, AFX66443, ACV60417, AFH35020, and ACV60418 were most likely produced by Genotype #3 of *O. sinensis* if GC-biased Genotype #3 of *O. sinensis* does not coexist with other heterospecific fungi in the wild-type *C. sinensis* isolates XZ12_16, XZ05_8, XZ-LZ07-H1, XZ06-124, and XZ-LZ07-H2.

Regardless of whether the *MAT1-1-1* and *MAT1-2-1* proteins form heterodimers or interact via complementary electrostatic surfaces and thereby enable cooperative DNA binding, the synergy of the mating proteins of *O. sinensis* constitutes the core mechanism that regulates mating type recognition, triggering downstream nuclear fusion signaling pathways and fruiting body development [39–43]. The differential occurrence, differential transcription, and alternative splicing of mating-type and pheromone receptor genes in *H. sinensis* and wild-type *C. sinensis* isolates invalidates the self-fertilization hypothesis at the genomic and transcriptomic levels but suggests self-sterility of *O. sinensis* under heterothallic or hybrid reproduction [31,48–51,53–55]. Furthermore, at the protein level, many mutant *MAT1-1-1* and *MAT1-2-1* proteins reported by Li *et al.* [50] that contain mutant *MAT* α _HMGbox domains and HMG-box_ROX1-like domains, respectively, have

been discovered in wild-type *C. sinensis* isolates, and the insect–fungal complex demonstrated in this study may indicate that mating proteins from heterogeneous fungal sources may accomplish coordinated heterothallic or hybrid reproduction of *O. sinensis* during the lifecycle of *C. sinensis* insect–fungal complexes.

5. Conclusions

This study demonstrated the variations in the tertiary structures of the mating proteins in numerous wild-type *C. sinensis* isolates, *H. sinensis* strains, and *C. sinensis* insect–fungal complex specimens, focusing on the full-length DNA-binding domains, *i.e.*, the MAT α _HMGbox domain of the MAT1-1-1 protein and the HMG-box_ROX1-like domain of the MAT1-2-1 protein. These DNA-binding domains are critical for specifically controlling the expression of genes related to the sexual reproduction of *O. sinensis*. The sequences of DNA-binding domains in the mating proteins exhibit various amino acid substitutions at different mutation sites and are clustered into 5 and 2 major Bayesian clades, respectively, each encompassing several Bayesian branches, and AlphaFold stereostructure predictions indicate that they belong to different tertiary protein structural morphs. The findings confirmed the hypothesis of *O. sinensis* self-sterility under heterothallic mating or hybrid mating at the protein structure level, further complementing the comprehensive evidence from the genetic and transcriptional levels, refuting the self-fertilization hypothesis for *O. sinensis* under homothallic or pseudohomothallic reproduction. Thus, finding an appropriate mating partner is the only means by which self-sterile *O. sinensis* can accomplish sexual reproduction during the lifecycle of the natural *C. sinensis* insect–fungal complex on the Qinghai–Tibet Plateau.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/doi/s1>, **Figure S1.** Alignment of the reference MAT1-1-1 protein AGW27560 sequence derived from the *Hirsutella sinensis* strain CS68-2-1229 [23] and the mutant protein sequences derived from the wild-type *Cordyceps sinensis* isolates with various amino acid residue substitutions derived from the wild-type *C. sinensis* isolates [52] and from the genome and metatranscriptome assemblies of *H. sinensis* strains or the *C. sinensis* insect–fungi complexes [49,61,63–65]. The underlined segment in blue refers to the MAT α _HMGbox domain (amino acids 51→225) of the reference MAT1-1-1 protein AGW27560, and the 9 external amino acid residues upstream and downstream of the domain are shown in blue but not underlined. The amino acid substitution is shown in red, whereas the hyphens indicate identical amino acid residues, and the spaces denote unmatched protein sequence gaps; **Figure S2.** Alignment of the reference MAT1-2-1 protein sequence AEH27625 derived from the *Hirsutella sinensis* strain CS2 [56] and the mutant MAT1-2-1 proteins derived from the wild-type *Cordyceps sinensis* isolates with various amino acid residue substitutions [52] and from the genome and metatranscriptome assemblies of *H. sinensis* strains or the *C. sinensis* insect–fungi complexes [49,59–62,65]. The underlined segment in blue refers to the HMG-box_ROX1-like domain (127→197 of the reference sequence AEH27625), and the 9 external amino acid residues upstream and downstream of the domain are shown in blue but not underlined. The amino acid substitution is shown in red, whereas the hyphens indicate identical amino acid residues, and the spaces denote unmatched protein sequence gaps; **Figure S3.** Correlation of the changes in the primary and secondary structures of the MAT α _HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 is derived from the *H. sinensis* strain CS68-2-1229 [23], and the truncated MAT1-1-1 protein is encoded by the metatranscriptome assembly GAGW01008880 derived from the *C. sinensis* insect–fungi complex [64]. Panel A shows an alignment of the amino acid sequences in the MAT α _HMGbox domains of the MAT1-1-1 proteins; amino acid substitutions are shown in red, whereas the hyphens indicate identical amino acid residues, and the spaces denote unmatched protein sequence gaps. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and the 2D structure (Panels C–F show the α -helices, β -sheets, β -turns,

and coils, respectively); the open rectangles in blue outline the truncation region in the plots; **Table S1**. GenBank accession numbers (in red in parentheses) for the full-length MAT1-1-1 proteins in the AlphaFold database under the corresponding AlphaFold UniProt codes [50]; **Table S2**. GenBank accession numbers (in red) for the full-length MAT1-2-1 proteins of 69 *H. sinensis* strains or *C. sinensis* isolates under the corresponding AlphaFold UniProt codes [50]; **Table S3**. Wild-type *C. sinensis* isolates, GenBank accession numbers for the ITS nucleic acid sequences and mating protein sequences, and percentage similarities *vs.* GC-biased *O. sinensis* Genotypes #1–3 and #7–9; **Table S4**. Amino acids are scaled on the basis of the general chemical characteristics of their side chains for ProtScale analysis (<https://web.expasy.org/protscale/>) to predict the hydrophobicity and secondary structures (α -helices, β -sheets, β -turns, and coils) of proteins; **Table S5**. Summary of the Bayesian clustering results in Figure S1 with the amino acid substitutions in the MAT α _HMGbox domains of the 19 full-length MAT1-1-1 proteins of the wild-type *C. sinensis* isolates under the AlphaFold UniProt codes and GenBank accession numbers; **Table S6**. Summary of the Bayesian clustering results in Figure S1 with the amino acid substitutions in the MAT α _HMGbox domains of the MAT1-1-1 proteins encoded by the genome assemblies of *H. sinensis* strains and the metatranscriptome assemblies of natural *C. sinensis* under the GenBank accession numbers; **Table S7**. Summary of the Bayesian clustering results in Figure 4 with the amino acid substitutions in the HMG-box_ROX1-like domains of the 25 full-length MAT1-2-1 proteins of wild-type *C. sinensis* isolates under the AlphaFold UniProt codes and GenBank accession numbers; **Table S8**. Summary of the Bayesian clustering results in Figure 4 with the amino acid substitutions and deletions in the HMG-box_ROX1-like domains of the MAT1-2-1 proteins encoded by the genome and transcriptome assemblies of *H. sinensis* strains and the metatranscriptome assembly of natural *C. sinensis* insect–fungi complexes found under the GenBank accession numbers;

Author Contributions: Conceptualization, X-Z.L., Y-L.L., W.L., and J-S.Z.; methodology, W.L. and J-S.Z.; formal analysis, J-S.Z.; investigation, X-Z.L. and J-S.Z.; data curation, X-Z.L. and J-S.Z.; writing—original draft preparation, J-S.Z.; writing—review and editing, X-Z.L., Y-L.L., W.L., and J-S.Z.; supervision, Y-L.L. and J-S.Z.; project administration, Y-L.L.; funding acquisition, Y-L.L. All the authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by (1) the Major Science and Technology Projects of Qinghai Province, China (#2021-SF-A4); (2) Joint Science Project of the Chinese Academy of Sciences, Qinghai Provincial Government, and Sanjiangyuan National Park (#LHZX-2022-01); and (3) Wild *Ophiocordyceps sinensis* Identification and Application Project (QHRD-2025-004).

Institutional Review Board Statement: Not applicable. This paper is an *in silico* reanalysis of publicly available data. Thus, this study does not require ethical approval from the Institutional Review Board or signed patient informed consent.

Informed Consent Statement: Not applicable.

Data Availability Statement: All sequence and 3D structure data are available in public depository databases: GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) and AlphaFold (<https://alphafold.com/>).

Acknowledgments: The authors are grateful to Prof. Mu Zang, Prof. Ru-Qin Dai, Prof. Zong-Qi Liang and Prof. Yong-Jie Zhang for their consultation.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhu J-S, Halpern GM, Jones K. The scientific rediscovery of a precious ancient Chinese herbal regimen: *Cordyceps sinensis*: Part I. *J. Altern. Complen. Med.* **1998**, 4(3): 289–303. DOI: 10.1089/acm.1998.4.3-289
2. Zhu J-S, Halpern GM, Jones K. The scientific rediscovery of an ancient Chinese herbal medicine: *Cordyceps sinensis*: Part II. *J. Altern. Complen. Med.* **1998**, 4(4): 429–457. DOI: 10.1089/acm.1998.4.429

3. 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. Available online: https://xueshu.baidu.com/usercenter/paper/show?paperid=08341c17fa58c8f85584b92572b90f75&site=xueshu_se (accessed on 30 January 2025)
4. 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.
5. 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](https://doi.org/10.1111/mec.12940)
6. 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.
7. 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.* **2022**, 14(3): 115–135. doi: [10.5099/aj220300115](https://doi.org/10.5099/aj220300115)
8. 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.* **2023**, 11: 2361. <https://doi.org/10.3390/microorganisms11092361>
9. 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.* **2023**, 18(6): e0286865. <https://doi.org/10.1371/journal.pone.0286865>
10. 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.* **2023**, 18(3): e0270776. DOI: [10.1371/journal.pone.0270776](https://doi.org/10.1371/journal.pone.0270776)
11. Li C-L. A study of *Tolypocladium sinense* C.L. Li. sp. nov. and cyclosporin production. *Acta Mycol. Sinica.* **1988**, 7: 93–98.
12. 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.
13. Dai R-Q, Li X-M, Shao A-J, Lin S-F, Lan J-L, Chen W-H, Shen C-Y. Nomenclatural validation of *Paecilomyces hepiali*. *Mycosystema.* **2008**, 27(5):641–644
14. Kinjo N, Zang M. Morphological and phylogenetic studies on *Cordyceps sinensis* distributed in southwestern China. *Mycoscience.* **2001**, 42: 567–574. DOI: [10.1007/BF02460956](https://doi.org/10.1007/BF02460956)
15. Jiang Y, Yao Y-J. A review for the debating studies on the anamorph of *Cordyceps sinensis*. *Mycosistema.* **2003**, 22(1): 161–176.
16. Stensrud Ø, Hywel-Jones NL, Schumacher T. Towards a phylogenetic classification of Cordyceps: ITS nrDNA sequence data confirm divergent lineages and paraphyly. *Micol. Res.* **2005**, 109: 41–56. DOI: [10.1017/s095375620400139x](https://doi.org/10.1017/s095375620400139x)
17. Stensrud Ø, Schumacher T, Shalchian-Tabrizi K, Svegardenib IB, Kauserud H. Accelerated nrDNA evolution and profound AT bias in the medicinal fungus *Cordyceps sinensis*. *Micol. Res.* **2007**, 111: 409–415. DOI: [10.1016/j.mycres.2007.01.015](https://doi.org/10.1016/j.mycres.2007.01.015)
18. 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](https://doi.org/10.1111/j.1365-2672.2006.02930.x)
19. Yang J-L, Xiao W, He H-X, Zhu H-X, Wang S-F, Cheng K-D, Zhu P. Molecular phylogenetic analysis of *Paecilomyces hepiali* and *Cordyceps sinensis*. *Acta Pharmaceut. Sinica.* **2008**, 43(4): 421–426.
20. 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. DOI: [10.19540/j.cnki.cjcmm.20210226.103](https://doi.org/10.19540/j.cnki.cjcmm.20210226.103)
21. 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.
22. 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.
23. 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](https://doi.org/10.1016/j.ympev.2013.04.010)
24. Li Y, Jiang L, Wang K, Wu H-J, Yang R-H, Yan Y-J, Bushley KE, Hawksworth DL, Wu Z-J, Yao Y-J. RIP mutated ITS genes in populations of *Ophiocordyceps sinensis* and their implications for molecular systematics. *IMA Fungus.* **2020**, 11: 18. DOI: [10.1016/j.mimet.2016.06.025](https://doi.org/10.1016/j.mimet.2016.06.025)
25. 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](https://doi.org/10.1016/j.ibmb.2015.06.014)
26. 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](https://doi.org/10.1139/cjm-2014-0610)
 27. Guo M-Y, Liu Y, Gao Y-H, Jin T, Zhang H-B, Zhou X-W. Identification and bioactive potential of endogenous 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](https://doi.org/10.17957/IJAB/15.0281)
 28. 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](https://doi.org/10.1016/j.funbio.2018.05.011)
 29. Kang Q, Zhang J, Chen F, Dong C, Qin Q, Li X, Wang H, Zhang H, Meng Q. Unveiling mycoviral diversity in *Ophiocordyceps sinensis* through transcriptome analyses. *Front. Microbiol.* **2024**, 15: 1493365. doi: [10.3389/fmicb.2024.1493365](https://doi.org/10.3389/fmicb.2024.1493365)
 30. 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. Available online: <https://m.163.com/dy/article/HHCV0IPU055360T7.html> (accessed on 3 May 2025).
 31. 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.
 32. 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.
 33. 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](https://doi.org/10.3114/sim.2007.57.01)
 34. 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>
 35. 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](https://doi.org/10.1128/mBio.00636-18)
 36. 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.
 37. 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](https://doi.org/10.5598/ima fungus.2011.02.01.14)
 38. 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](https://doi.org/10.1006/fgbi.2000.1227)
 39. Debuchy R, Turgeon BG. Mating-Type Structure, Evolution, and Function in Euascomycetes. In (eds. Kües U. & Fischer, R.) Growth, Differentiation and Sexuality. Springer (2006). pp. 293–323.
 40. 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](https://doi.org/10.1016/j.fgb.2011.04.001)
 41. Zheng, P.; Wang, C.-S. Sexuality Control and Sex Evolution in Fungi. *Sci. Sin. Vitae* **2013**, 43, 1090–1097.
 42. 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](https://doi.org/10.5598/ima fungus.2015.06.01.13)
 43. Sun S, Coelho MA, David-Palma M, Priest SJ, Heitman J. The evolution of sexual reproduction and the mating-type locus: Links to pathogenesis of *Cryptococcus* human pathogenic fungi. *Annu. Rev. Genet.* **2019**, 53: 417–444. doi: [10.1146/annurev-genet-120116-024755](https://doi.org/10.1146/annurev-genet-120116-024755)
 44. Ramšak B, Markau J, Pazen T, Dahlmann TA, Krappmann S, Kück U. The master regulator MAT1-1-1 of fungal mating binds to its targets via a conserved motif in the human pathogen *Aspergillus fumigatus*. *G3-Genes / Genomics / Genetics.* **2020**, 11(2): jkaa012/6026963. <https://doi.org/10.1093/g3journal/jkaa012>
 45. Ramšak B, Kück U, Hofmann E. The mating type transcription factor MAT1-1-1 from the fungal human pathogen *Aspergillus fumigatus*: synthesis, purification, and crystallization of the DNA binding domain. *BioRxiv.* **2021**. doi: <https://doi.org/10.1101/2021.12.13.472399>
 46. Ramšak 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](https://doi.org/10.3389/ffunb.2022.937023)
 47. Metin B, Findley K, Heitman J. The mating type locus (MAT) and sexual reproduction of *Cryptococcus heveanensis*: insights into the evolution of sex and sex-determining chromosomal regions in fungi. *PLoS Genet.* **2010**, 6(5): e1000961.

- [doi:10.1371/journal.pgen.1000961](https://doi.org/10.1371/journal.pgen.1000961) 1376
48. 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](https://doi.org/10.1016/j.funbio.2013.06.001) 1377
 49. 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](https://doi.org/10.1007/s11434-013-5929-5) 1378
 50. Li X-Z, Li Y-L, Zhu J-S. Three-dimensional structural heteromorphs of mating-type proteins in *Hirsutella sinensis* and the natural *Cordyceps sinensis* insect–fungal complex. *J. Fungi.* **2025**, 11(4): 244. <https://doi.org/10.3390/jof11040244> 1379
 51. Zhou XW, Li LJ, Tian EW. Advances in research of the artificial cultivation of *Ophiocordyceps sinensis* in China. *Critical Rev Biotechnol* **2013**, 34(3), 233–243. <https://doi.org/10.3109/07388551.2013.791245> 1380
 52. 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. 1381
 53. 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> 1382
 54. 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.* **2023**, 48(10): 2829–2840. DOI: [10.19540/j.cnki.cjcmm.20230213.102](https://doi.org/10.19540/j.cnki.cjcmm.20230213.102) 1383
 55. 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)* **2024**, 13(8): 632. <https://doi.org/10.3390/biology13080632> 1384
 56. 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 1385
 57. 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. 1386
 58. 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> 1387
 59. 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](https://doi.org/10.1186/s12864-015-1269-y) 1388
 60. 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> 1389
 61. 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.* **2016**, 128: 1–6. DOI: [10.1016/j.mimet.2016.06.025](https://doi.org/10.1016/j.mimet.2016.06.025) 1390
 62. 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> 1391
 63. 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](https://doi.org/10.1093/gbe/evaa112) 1392
 64. 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> 1393
 65. 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 1394

66. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogeny. *Bioinformat.* **2001**, 17: 754–755. DOI: [10.1093/bioinformatics/17.8.754](https://doi.org/10.1093/bioinformatics/17.8.754) 1427
67. 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 Biol.* **2012**, 61(3): 539–542. <https://doi.org/10.1093/sysbio/sys029> 1429
68. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, 157(1): 105–132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0) 1430
69. 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> 1431
70. 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. 1432
71. 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](https://doi.org/10.1002/prot.24582) 1433
72. 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](https://doi.org/10.1186/s40659-016-0092-5) 1434
73. 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](https://doi.org/10.1038/s41586-024-07487-w) 1435
74. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek 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> 1436
75. 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> 1437
76. Monzon V, Haft DH, Bateman A. Folding the unfoldable: using AlphaFold to explore spurious proteins. *Bioinform Adv.* **2022**, 1(2): vbab043. <https://doi.org/10.1093/bioadv/vbab043> 1438
77. 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.* **2023**; 26:2023.02.25.529956. doi: [10.1101/2023.02.25.529956](https://doi.org/10.1101/2023.02.25.529956) 1439
78. 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](https://doi.org/10.3389/frai.2023.1149748) 1440
79. 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](https://doi.org/10.1093/nar/gkad1011) 1441
80. 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](https://doi.org/10.1016/j.csbj.2024.11.047) 1442
81. Ellison TJ, Ellison CK, Improved DNA binding to a type IV minor pilin increases natural transformation. *Nucleic Acids Res.* **2025**, 53(10): gkaf467, <https://doi.org/10.1093/nar/gkaf467> 1443
82. Yang X, Zhu H-Q, Shi L-X, Song T-G, Gong W-B, He S-M, Shan S, Xu C-F, Zhou Z, AlphaFold-guided structural analyses of nucleosome binding proteins. *Nucleic Acids Res.* **2025**, 53(14): gkaf735, <https://doi.org/10.1093/nar/gkaf735> 1444
83. 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> 1445
84. Baxeavanis AD, Bryant SH, Landsman D. Homology model building of the HMG-1 box structural domain. *Nucleic Acids Res.* **1995**, 23(6): 1019–1029. DOI: [10.1093/nar/23.6.1019](https://doi.org/10.1093/nar/23.6.1019) 1446
85. Thapar R. Structure-specific nucleic acid recognition by L-motifs and their diverse roles in expression and regulation of the 1447

- genome. *Biochim Biophys Acta*. **2015**, 1849(6): 677–687. doi: 10.1016/j.bbagra.2015.02.006 1478
86. 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.* **2016**, 14: 96–112. 1479
 87. Jackson D, Lawson T, Villafane R, Gary L. Modeling the structure of yeast MAT α 1: An HMG-Box motif with a C-terminal helical extension. *Open J. Biophysics*. **2013**, 3: 1–12. <http://dx.doi.org/10.4236/ojbiphy.2013.31001> 1480
 88. Zheng Q, Hou R, Zhang J-Y, Ma J, Ma J-W, Wu Z-S, Wang G-H, Wang C-F, Xu J-R. The MAT locus genes play different roles in sexual reproduction and pathogenesis in *Fusarium graminearum*. *PLoS ONE*. **2013**; 8(6): e66980. doi:10.1371/journal.pone.0066980 1481
 89. Kim H-K, Jo S-M, Kim G-Y, Kim D-W, Kim Y-K, Yun S-H. A large-scale functional analysis of putative target genes of mating-type loci provides insight into the regulation of sexual development of the cereal pathogen *Fusarium graminearum*. *PLoS Genet.* **2015**; 11(9): e1005486. doi:10.1371/journal.pgen.1005486 1482
 90. Martin T, Lu S-W, van Tilbeurgh H, Ripoll DR, Dixelius C, Dixelius C, Turgeon BG, Debuchy R. Tracing the Origin of the Fungal a1 Domain Places Its Ancestor in the HMG-Box Superfamily: Implication for Fungal Mating-Type Evolution. *PLoS ONE* **2010**; 5(12): e15199. doi:10.1371/journal.pone.0015199 1483
 91. Ait Benkhali J, Coppin E, Brun S, Peraza-Reyes L, Martin T, Dixelius C, Lazar N, van Tilbeurgh H, Debuchy R. A Network of HMG-box Transcription Factors Regulates Sexual Cycle in the Fungus *Podospora anserina*. *PLoS Genet.* **2013**; 9(7): e1003642. doi:10.1371/journal.pgen.1003642 1484
 92. Yamamoto A, Ando Y, Yoshioka K, Saito K, Tanabe T, Shirakawa H, Yoshida M. Difference in affinity for DNA between HMG proteins 1 and 2 determined by surface plasmon resonance measurements. *J. Biochem.* **1997**; 122(3): 586–594. doi: 10.1093/oxfordjournals.jbchem.a021793. PMID: 9348088 1485
 93. Balasubramanian B, Lowry CV, Zitomer RS. The Rox1 repressor of the *Saccharomyces cerevisiae* hypoxic genes is a specific DNA-binding protein with a high-mobility-group motif. *Mol. Cell Biol.* **1993**; 13(10): 6071–6078. doi: 10.1128/mcb.13.10.6071-6078.1993 1486
 94. Zitomer RS, Limbach MP, Rodriguez-Torres AM, Balasubramanian B, Deckert J, Snow PM. Approaches to the study of Rox1 repression of the hypoxic genes in the yeast *Saccharomyces cerevisiae*. *Methods.* **1997**; 11(3): 279–288. doi: 10.1006/meth.1996.0422 1487
 95. Kastaniotis AJ, Zitomer RS. Rox1 mediated repression. Oxygen dependent repression in yeast. *Adv. Exp. Med. Biol.* **2000**; 475: 185–195. <https://pubmed.ncbi.nlm.nih.gov/10849660/> 1488
 96. Kües U, Casselton LA. The origin of multiple mating types in mushrooms. *J. Cell Sci.* **1993**; 104 (2): 227–230. <https://doi.org/10.1242/jcs.104.2.227> 1489
 97. Asante-Owusu RN, Banham AH, Böhnert HU, Mellor EJC, Casselton LA. Heterodimerization between two classes of homeo-domain proteins in the mushroom *Coprinus cinereus* brings together potential DNA-binding and activation domains. *Gene.* **1996**; 172(1): 25–31. [https://doi.org/10.1016/0378-1119\(96\)00177-1](https://doi.org/10.1016/0378-1119(96)00177-1) 1490
 98. 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 1491
 99. Hancock SP, Cascio D, Johnson RC. Cooperative DNA binding by proteins through DNA shape complementarity. *Nucleic Acids Res.* **2019**; 47(16):8874–8887. doi: 10.1093/nar/gkz642 1492
 100. Kück U, Bennett RJ, Wang L, Dyer PS. Editorial: Sexual and Parasexual Reproduction of Human Fungal Pathogens. *Front. Cell Infect. Microbiol.* **2022**; 12: 934267. doi: 10.3389/fcimb.2022.934267 1493
 101. 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 1494
 102. 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 1495
 103. 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 1496
 104. 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> 1497
 105. 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 1498
 106. Steensels J, Gallone B, Verstrepen KJ. Interspecific hybridization as a driver of fungal evolution and Adaptation. *Nat. Rev. Microbiol.* **2021**; 19: 485–500. 1499
 107. Mishra A, Forche A, Anderson MZ. Parasexuality of *Candida* Species. *Front. Cell. Infect. Microbiol.* **2021**; 11: 796929. doi: 1500

- [10.3389/fcimb.2021.796929](https://doi.org/10.3389/fcimb.2021.796929) 1529
108. 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](https://doi.org/10.1016/S0378-1097(03)00889-9) (accessed on 30 January 2025). 1530
 109. 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 1531
 110. 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. 1532
 111. Zhu J-S, Gao L, Li X-H, Yao Y-S, Zhou Y-J, Zhao J-Q, Zhou Y-J. Maturational 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](https://doi.org/10.5099/aj100300217) 1533
 112. 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](https://doi.org/10.1615/intjmedmushr.v13.i6.90) 1534
 113. 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. 1535
 114. 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> 1536
 115. 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](https://doi.org/10.7554/eLife.60474) 1537
 116. 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. *Int. J. Mol. Sci.* 2024, 25(20), 11178; <https://doi.org/10.3390/ijms252011178> 1538
 117. Li Y, Yang R-H, Jiang L, Hu X-D, Wu Z-J, Yao Y-J. rRNA Pseudogenes in Filamentous Ascomycetes as Revealed by Genome Data. *G3-Genes / Genomics / Genetics*. 2017, 7(8): 2695–2703. [doi: 10.1534/g3.117.044016](https://doi.org/10.1534/g3.117.044016) 1539
 118. Li W-J, Xia J-M, Li Q-P, Zhang Z-Y, Zhang W-W, Dong C-H, Wei J-C, Liu X-Z. Developmental recording of the ghost-moth larvae after *ex situ* infection by *Ophiocordyceps sinensis*. *Sci. China Life Sci.* 2020, 63, 1093–1095. <https://doi.org/10.1007/s11427-020-1686-7> 1540
 119. Li W-J, Xia J-M, Li Q-P, Zhang Z-Y, Zhang W-W, Dong C-H, Wei J-C, Liu X-Z. **Erratum** to: Developmental recording of the ghost-moth larvae after *ex situ* infection by *Ophiocordyceps sinensis*. *Sci. China Life Sci.* 2023, 66, 892. <https://doi.org/10.1007/s11427-022-2232-5> 1541
 120. 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](https://doi.org/10.1615/IntJMedMushr.v10.i3.30) 1542
 121. Stone R. Improbable partners aim to bring biotechnology to a Himalayan kingdom. *Sci.* 2010, 327: 940–941. [DOI: 10.1126/science.327.5968.940](https://doi.org/10.1126/science.327.5968.940) 1543
 122. Qin Q-L, Zhou G-L, Zhang H, Meng Q, Zhang J-H, Wang H-T, Miao L, Li X. Obstacles and approaches in artificial cultivation of Chinese cordyceps. *Mycol.* 2018, 9(1): 7–9. <https://doi.org/10.1080/21501203.2018.1442132> 1544

1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567