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

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

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

* Corresponding author

E-mail: zhujosh@163.com (JSZ)

Abstract:

The MATα_HMGbox and HMG-box_ROX1-like domains of MAT1-1-1 and MAT1-2-1 proteins, respectively, play essential roles in DNA binding and downstream gene transcription regulation, controlling sexual reproduction of *Ophiocordyceps sinensis*. Alternative splicing, differential occurrence and transcription of the *MAT1-1-1* and *MAT1-2-1* genes have been unveiled in *Hirsutella sinensis*, suggesting the occurrence of self-sterility under heterothallic or hybrid outcrossing. This study further demonstrated that the MATα_HMGbox domains of MAT1-1-1 proteins in the wild-type *Cordyceps sinensis* isolates were clustered into 5 clades with or without branches in the Bayesian clustering tree, belonging to diverse 3D structural morphs under 19 AlphaFold codes. The HMG-box_ROX1-like domains of MAT1-2-1 proteins were clustered into 2 clades with branches in the Bayesian clustering tree, belonging to diverse 3D structural morphs under 25 AlphaFold codes. The correlation analysis reviewed that the 1–3 amino acid substitutions occurred in various sites of the DNA binding domains of the mutant mating proteins resulted in alteration of the hydrophobicity, secondary and tertiary structures of the DNA binding domains of the proteins. The fungal origin analysis demonstrated heterospecific fungal or genome-independent genotypical sources of the mating proteins with stereostructure variations in the wild-type *C. sinensis* isolates and *O. sinensis* strains, which represent the changed functionalities of the proteins in regulating and ensuring the accuracy and genetic diversity of the heterothallic or hybrid sexual reproductive process of *O. sinensis* in the lifecycle of *C. sinensis* insect-fungi complex.

Keywords: MATα_HMGbox domain of MAT1-1-1 protein; HMG-box_ROX1-like domain of MAT1-2-1 protein; heteromorphic primary, secondary, and tertiary structures of proteins; Bayesian clustering; *Ophiocordyceps sinensis* strain; wild-type *Cordyceps sinensis* isolates; reproduction of *O. sinensis*; sexual life of the *Cordyceps sinensis* insect-fungal complex

I. 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 in health maintenance, disease amelioration, post illness and postoperative recovery and antiaging therapy [Zhu *et al*. 1998a, 1998b, 2011]. This prestige 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, intact larval intestine and head tissues, and fragments of other larval tissues [Ren *et al*. 2013; Zhang *et al*. 2014; Lu *et al*. 2016; Zhu & Li 2017; Li *et al*. 2022, 2023a, 2023d, 2023e]. 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 [Li 1988; Dai et al. 1989; Kinjo & Zang 2001; Jiang & Yao 2003; Stensrud *et al*. 2005, 2007; Leung *et al*. 2006; Zhu *et al*. 2007, 2010; Yang *et al*. 2008, 2021; Zhang *et al*. 2010, 2018; Li *et al*. 2013, 2016b, 2020b, 2022, 2023e; Ren *et al*. 2013; Meng *et al*. 2015; Xia *et al*. 2015; Guo *et al*. 2017; Zhu & Li 2017; Zhong *et al*. 2018; Kang *et al*. 2024]. Thus, the Chinese Pharmacopoeia defines natural *C. sinensis* as an insect-fungal complex, which is classified as the LEVEL-II endangered natural species [China Ministry of Agriculture and Rural Affairs 2021]. Among the numerous heterogeneous fungal species, *Hirsutella sinensis* has been postulated to be the sole-anamorph of *O. sinensis* by Wei *et al*. [2006]. Ten years later, however, the key authors reported a species contradiction in an artificial cultivation project conducted in an industrial product-oriented setting between anamorphic inoculates of GC-biased Genotype #1 *H. sinensis* 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* [Wei *et al*. 2016]. Notably, the Latin name *Cordyceps sinensis* has been indiscriminately used since the 1840s for both the teleomorph/holomorph of the fungus *C. sinensis* and the wild insect-fungal complex, and the fungus was renamed *O. sinensis* in 2007 with using the *H. sinensis* strain EFCC 7287 as the nomenclature reference [Sung *et al*. 2007; Zhang *et al*. 2012; Ren *et al*. 2013; Zhu & Li 2017; Wang *et al*. 2018; Li *et al*. 2022]. Although Zhang *et al*. [2013] proposed improper implementation of the "One Fungus=One Name" nomenclature rule of the International Mycological Association [Hawksworth *et al*. 2011] while disregarding the presence of multiple genomically independent genotypes of *O. sinensis* fungi and inappropriately replacing the anamorphic name *H. sinensis* with the teleomorphic name *O. sinensis*, we continue using in this paper the anamorphic name *H. sinensis* for GC-biased Genotype #1 of the 17 *O. sinensis* genotypes and refer to the genomically independent Genotypes #2–17 fungi as *O. sinensis* following the taxonomic descriptions in several public depository databases, such as GenBank and

AlphaFold databases, before their systematic positions are individually determined, regardless of whether they are GC- or AT-biased genetically. 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-fungi complex [Sung *et al*. 2007; Yao & Zhu 2016; Zhu & Li 2017; Li *et al*. 2022], although 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-fungi complex.

The sexual reproductive behavior of ascomycetes is strictly regulated by transcription factors encoded at the mating-type (*MAT*) locus, which constitute the core mechanism that determine mating compatibility, regulating mating type recognition, and developing fruiting body, reproductive structure ascocarps and ascospores [Turgeon & Yoder 2006; Debuchy *et al*. 2006; Jones & Bennett 2011; Zheng & Wang 2013; Wilson *et al*. 2015; Sun *et al*. 2019; Rams *et al*. 2022]. Similar to other fungi in Ascomycota, reproductive behavior of *O. sinensis* relies on the synergistic interaction of 2 mating proteins, the *MAT1-1* and *MAT1-2* idiomorphs, which contain 2 types of critical domains specifically regulating gene expressions relating to sexual reproduction: (1) the Mating-type alpha High Mobility Group box (MATα_HMGbox) domain in MAT1-1-1 protein, and (2) the High Mobility Group box ROX1-like (HMG-box_ROX1-like) domain in MAT1-2-1 protein [Metin *et al*. 2010; Bushley *et al*. 2013; Hu *et al*. 2013; Zheng & Wang 2013; Li *et al*. 2025]. However, alternative splicing, differential occurrence, and differential transcription of mating-type and pheromone receptor genes and heteromorphic stereostructures of entire MAT1-1-1 and MAT1-2-1 proteins in *H. sinensis* strains and wild-type *C. sinensis* isolates were demonstrated, invalidating the self-fertilization hypothesis under homothallism and pseudohomothallism for *H. sinensis* that was postulated as the sole anamorph of *O. sinensis* but suggesting self-fertility under heterothallism or hybridization requiring sexual partners to accomplish sexual reproduction within the lifecycle of natural *C. sinensis* [Wei *et al*. 2006; Zhou *et al*. 2013; Zhang & Zhang 2015; Li *et al*. 2019, 2023c, 2024b, 2025].

Mutant sequences of the 2 critical DNA binding domains of mating proteins with a few amino acid substitutions may cause conformational changes in the core stereostructures of the domains, thereby affecting the synergistic effects probably through regulating complementary downstream genes of 2 idiomorphic mating proteins. Continuing our previous study [Li *et al*. 2025], which demonstrated the heteromorphic stereostructures of entire mating proteins, this study focuses on the amino acid substitutions within the MATα_HMGbox domain of the MAT1-1-1 protein and the HMG-box_ROX1-like domain of MAT1-2-1 protein and correlating the variable primary structures with the changes in hydrophobic properties and secondary and tertiary structures of the DNA binding domains in *O. sinensis* strains and wild-

type *C. sinensis* isolates. Correlations of the changes in hydrophobicity, 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-fungi complexes are also analyzed.

II. Materials and Methods

II-1 The MAT1-1-1 and MAT1-2-1 protein sequences of the *O. sinensis* strains and 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 [Li *et al*. 2025]. Among the proteins, 118 MAT1-1-1 proteins and 69 MAT1-2-1 proteins were full-length listed 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) were derived from *H. sinensis* strains. All other full-length mating proteins were from wild-type *C. sinensis* isolates that were collected from different production areas on the Qinghai–Tibet Plateau [Zhang *et al*. 2009, 2011, 2012, 2013, 2014; Hu *et al*. 2013; Zhang & Zhang 2015; Tunyasuvunakool *et al*. 2021; Li *et al*. 2025]. The mutant full-length MAT1-1-1 and MAT1-2-1 proteins were derived from 34 wild-type *C. sinensis* isolates listed in Table S3 with the genotypic information.

The remaining 20 MAT1-1-1 proteins (AGW27517–AGW27536) and 5 MAT1-2-1 proteins (AGW27543, AGW27548, AGW27552, AGW27554, and AGW27555) are truncated (Table S4) [Bushley *et al*. 2013; Li *et al*. 2013]. These protein sequences were derived from the *MAT1-1-1* and *MAT1-2-1* cDNA amplified from the total RNA that were extracted from the *O. sinensis* strains [Bushley *et al*. 2013], which were isolated from either the fresh caterpillar bodies (sclerotium) or the cultures of mono-ascospores of the *C. sinensis* insect-fungal complexes [Li *et al*. 2013]. The primer pairs of *m1cDNAF3/m1cDNAR5* and *Mat1-2F/Mat1-2R* were used to amplify the *MAT1-1-1* and *MAT1-2-1* cDNA, respectively [Bushley *et al*. 2013]. The sequences and locations of primers are shown in Figures S1–S2.

The internal transcribed spacer (ITS) molecular taxonomy/genotyping information for the wild-type *C. sinensis* isolates and *O. sinensis* strains in the GenBank database are listed in Tables S3–S6 with further alignment with the reference ITS sequences of genotypes of *O. sinensis*.

II-2 Genome, transcriptome, and metatranscriptome assemblies of *H. sinensis* strains and *C. sinensis* insect–fungal complexes

The GenBank database lists 5 genome assemblies LKHE00000000, NGJJ00000000, ANOV00000000, JAAVMX000000000, and LWBQ00000000 of the *H. sinensis* strains 1229, CC1406-20395, Co18, IOZ07, and ZJB12195, respectively, and a transcriptome assembly GCQL00000000 of the *H. sinensis* strain L0106

135 [Hu *et al*. 2013; Liu *et al*. 2015, 2020; Li *et al*. 2016a; Jin *et al*. 2020; Shu *et al*. 2020].

136 The metatranscriptome assembly GAGW00000000 is available in the GenBank database for the *C.*
137 *sinensis* insect-fungal specimens (unknown maturation stages) collected from Kangding County, Sichuan
138 Province of China [Xiang *et al*. 2014].

139 Another metatranscriptome assembly was derived from mature *C. sinensis* specimens collected from
140 Deqin, Yunnan Province of China and was uploaded to a repository database,
141 www.plantkingdomgdb.com/Ophiocordyceps_sinensis/data/cds/Ophiocordyceps_sinensis_CDS.fas [Xia
142 *et al*. 2017]. Although this database is currently inaccessible, it was accessed from 18 May 2017 to 18
143 January 2018 and a cDNA file was downloaded.

144 The genome, transcriptome, and metatranscriptome assemblies were used for analysis of the
145 MATα_HMGbox and HMG-box_ROX1-like domains of the MAT1-1-1 and MAT1-2-1 proteins,
146 respectively.

147 II-3 Alignment analysis of the DNA binding domain sequences of the mating proteins

148 The amino acid sequences of the MATα_HMGbox domains of the MAT1-1-1 proteins and the HMG-
149 box_ROX1-like domains of MAT1-2-1 proteins of the *O. sinensis* strains, wild-type *C. sinensis* isolates,
150 and *C. sinensis* insect-fungi complexes were aligned using the GenBank Blastp program
151 (https://blast.ncbi.nlm.nih.gov/ (Bethesda, MD, USA), accessed from 18 October 2024 to 10 June 2025).

152 II-4 Bayesian clustering analysis for the DNA binding domains of mating proteins

153 Multi sequence alignment of the DNA binding domains of mating proteins of the *O. sinensis* strains,
154 wild-type *C. sinensis* isolates, and *C. sinensis* insect-fungal complexes was performed *via* the auto mode of
155 MAFFT (v7.427; Osaka, Japan). The intron regions were removed using trimAl (v1.4. rev15; Barcelona,
156 Spain). Under the Bayesian information criterion, ProtTest v3.4.2 (University of Vigo and University of A
157 Coruña, Spain) was used to find the optimal amino acid substitution model. Bayesian clustering trees of the
158 DNA binding domain sequences of MAT1-1-1 and MAT1-2-1 proteins were then inferred using MrBayes
159 v3.2.7 software (Markov chain Monte Carlo [MCMC] algorithm; Uppsala, Sweden) with a sampling
160 frequency of 100 iterations after discarding the initial 25% of the samples from a total of 1 million iterations,
161 obtaining the final majority rule consensus trees for the MATα_HMGbox domains of the full-length MAT1-
162 1-1 proteins and the HMG-box_ROX1-like domains of the full-length MAT1-2-1 proteins [Huelsenbeck &
163 Ronquist 2001; Ronquist *et al*. 2012; Li *et al*. 2022, 2023c, 2024b, 2025]. Clustering analysis was conducted
164 at Nanjing Genepioneer Biotechnologies Co. (Nanjing, China).

165 II-5 Amino acid properties and scale analysis

166 The amino acid sequences of the DNA binding domains of mating proteins were scaled on the basis of

167 the general chemical characteristics of their side chains (Table S7) (https://web.expasy.org/protscale/ (Basel,
168 Switzerland), accessed from 18 October 2024 to 20 May 2025) [Kyte & Doolittle 1982; Deleage & Roux
169 1987; Gasteiger *et al*. 2005; Peters & Elofsson 2014; Simm *et al*. 2016; Li *et al*. 2024b, 2025]. The
170 MATα_HMGbox domain (51→225 of the reference sequence AGW27560 derived from the *H. sinensis*
171 strain CS68-2-1229 [Zhang *et al*. 2013]) of MAT1-1-1 proteins and the HMG-box_ROX1-like domain
172 (127→197 of the reference sequence AEH27625 [Zhang *et al*. 2009]) of MAT1-2-1 proteins with additional
173 9 amino acid residues extending both upstream and downstream of the domains were plotted sequentially
174 with a window size of 9 amino acid residues for the hydropathy, α-helices, β-sheets, β-turns, and coils of
175 the proteins using the linear weight variation model of the ExPASy ProtScale algorithms [Kyte & Doolittle
176 1982; Deleage & Roux 1987; Gasteiger *et al*. 2005; Li *et al*. 2024b, 2025]. The topologies and waveforms
177 of the ProtScale plots were compared to explore changes in the hydrophobicity and 2D structures of the
178 DNA binding domains of mating proteins.

179 II-6 AlphaFold-based prediction of 3D structures of mating proteins

180 The 3D structures of the MAT1-1-1 and MAT1-2-1 proteins of the *O. sinensis* strains and wild-type *C.*
181 *sinensis* isolates were computationally predicted from their amino acid sequences using the artificial
182 intelligence (AI)-based machine learning technology AlphaFold (https://alphafold.com/, Cambridgeshire,
183 UK) and downloaded from the AlphaFold database (accessed from 18 October 2024 to 10 May 2025) for
184 heteromorphic stereostructure analysis in this study [Jumper *et al*. 2021; David *et al*. 2022; Monzon *et al*.
185 2022; Rettie *et al*. 2023; Xu *et al*. 2023; Abramson *et al*. 2024; Varadi *et al*. 2024; Wroblewski & Kmiecik
186 2024; Li *et al*. 2025].

187 AlphaFold database provides a per-residue model confidence, the prediction of its score on the local
188 Distance Difference Test (pLDDT), between 0 and 100, a per-residue score that's assigned to each individual
189 residue [Jumper *et al*. 2021; David *et al*. 2022; Monzon *et al*. 2022; Xu *et al*. 2023; Abramson *et al*. 2024;
190 Varadi *et al*. 2024]. The model confidence bands are used to color-code the residues in the 3D structures:
191 very high (pLDDT>90) in dark blue, high (90>pLDDT>70) in light blue, low (70>pLDDT>50) in yellow,
192 and very low (pLDDT<50) in orange [Mariani *et al*. 2013; Wroblewski & Kmiecik 2024; Li *et al*. 2025].

193 II-7 Correlation of the primary, secondary, and tertiary structures of the DNA
194 binding domains of MAT1-1-1 and MAT1-2-1 proteins

195 The AlphaFold-predicted 3D structures of the mutant MATα_HMGbox domains of MAT1-1-1 proteins
196 and the mutant HMG-box_ROX1-like domains of MAT1-2-1 proteins were amplified localizing at the sites
197 of amino acid substitutions. The local magnification of 3D structures at the mutation sites were correlated
198 with the amino acid substitutions for the primary structures and the topology and waveform changes in

ExPASy ProtScale plots for hydropathy, α-helices, β-sheets, β-turns, and coils for altered hydrophobicity and secondary structures of the DNA binding domains of MAT1-1-1 and MAT1-2-1 proteins.

III. Results

III-1 Primary structures of the MATα_HMGbox domains of the full-length MAT1-1-1 proteins

Among the 138 MAT1-1-1 proteins in the AlphaFold database, 118 (85.5%) are full length consisting of 372 amino acids and contributing to 15 heteromorphic 3D structure models of 5 Bayesian clusters (Table S1) [Li *et al*. 2025]. Twenty-nine (24.6%) of the 118 full-length MAT1-1-1 proteins that were derived from wild-type *C. sinensis* isolates contain variable amino acid substitutions.

The MAT1-1-1 proteins contain an MATα_HMGbox domain at the amino acids 51→225 shown in blue and underlined in Figure S3, as represented by the reference MAT1-1-1 protein AGW27560 derived from the *H. sinensis* strain CS68-2-1229 [Zhang *et al*. 2013]. 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-fungi complexes contained various amino acid substitutions shown in red in Figure S3. The MATα_HMGbox domains of 10 (34.5%) of the 29 mutant proteins are 100% identical with the query protein AGW27560. The remaining 19 proteins (65.5%) contain 1–2 amino acid substitutions at various mutation sites within the MATα_HMGbox domains.

Figure S3 also lists the MATα_HMGbox domains of MAT1-1-1 proteins that are 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 [Hu *et al*. 2013; Li *et al*. 2016a; Shu *et al*. 2020]. (Note: "&" refers to the deleted intron section; the arrows "→" and "←" indicate sequences in the sense and antisense strands of the genomes, respectively.) The genome-encoded MAT1-1-1 proteins are truncated at the C-termini, which locate far downstream of the MATα_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 transcript encoding MAT1-1-1 proteins [Jin *et al*. 2020; Liu *et al*. 2015, 2020].

The metatranscriptome assemblies OSIN7648 (1→1065) and GAGW01008880 (300←1127) of natural *C. sinensis* insect-fungi complexes encode the MAT1-1-1 proteins (*cf*. Figure S3) [Xiang *et al*. 2014; Xia *et al*. 2017]. The MAT1-1-1 protein OSIN7648 is midsequence truncated by 18 aa between residues 863 and 864, locating downstream of the MATα_HMGbox domain (151→675). The N-terminus of the

230 MATα_HMGbox domain (741←1127) of MAT1-1-1 protein encoded by GAGW01008880 is truncated by
231 46 aa (AAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASG).

232 III-2 Primary structures of the HMG-box_ROX1-like domains of the full-length 233 MAT1-2-1 proteins

234 Among the 74 MAT1-2-1 proteins listed in the AlphaFold database, 69 (93.2%) are full length
235 containing 249 amino acids and contributed to 17 heteromorphic AlphaFold tertiary structure morphs of 5
236 Bayesian clusters (Table S2) [Li *et al*. 2025]. Among the 69 full-length proteins, 39 (56.5%) are 100%
237 identical to the representative MAT1-2-1 protein AEH27625 that was derived from the *H. sinensis* strain
238 CS2 [Zhang *et al*. 2009]. The remaining 30 full-length proteins (43.5%) contain amino acid substitutions at
239 various mutation sites shown in red in Figure S4. The HMG-box_ROX1-like domains of the MAT1-2-1
240 proteins are located at the amino acids 127→197 shown in blue and underlined in Figure S4 and the external
241 9 amino acids both upstream and downstream of the domain are in blue but not underlined. The HMG-
242 box_ROX1-like domain sequences of 3 of the 30 variable AMT1-2-1 proteins are 100% identical to the
243 domain sequence of the query protein AEH27625, while the domains of the remaining 27 proteins contain
244 1–3 amino acid substitutions shown in red in Figure S4.

245 Figure S4 also shows the MAT1-2-1 proteins encoded by the genome assemblies ANOV01000063,
246 LKHE01001605, LWBQ01000021, and NGJJ01000619 of the *H. sinensis* strains Co18, 1229, ZJB12195,
247 and CC1406-20395, respectively [Hu *et al*. 2013; Li *et al*. 2016a; Jin *et al*. 2020; Liu *et al*. 2020]. A
248 substitution of S-to-A occurred in each of the HMG-box_ROX1-like domains of genomically encoded
249 MAT1-2-1 proteins, ANOV01000063 (9759→9851 & 9907→10,026), LKHE01001605 (14,016←14,135
250 & 14,191←14,283), LWBQ01000021 (239,029←239,148 & 239,204←239,269), and NGJJ01000619
251 (23,186←23,305 & 23,361←23,453). An additional Y-to-H substitution occurred in the domains of
252 LKHE01001605, LWBQ01000021 and NGJJ01000619 but not in the domain of ANOV01000063, resulting
253 in 97.2– 98.6% similarities to the domain sequence of the query protein AEH27625 [Zhang *et al*. 2009].

254 The genome assembly JAAVMX000000000 of the *H. sinensis* strain IOZ07 and the metatranscriptome
255 assembly GAGW00000000 of the *C. sinensis* insect-fungi complex do not contain the gene or transcript
256 encoding MAT1-2-1 proteins [Xiang *et al*. 2014; Shu *et al*. 2020].

257 Figure S4 also shows a protein sequence encoded by the MAT1-2-1 transcript of the transcriptome
258 assembly GCQL01020543 (397←1143) of the *H. sinensis* strain L0106 [Liu *et al*. 2015]. The HMG-
259 box_ROX1-like domain (553←765) of this protein contains a Y-to-H substitution, resulting in 98.6%
260 similarity to the domain sequence of the query protein AEH27625 [Zhang *et al*. 2013].

261 The metatranscriptome assembly OSIN7649 (1→249) of the mature *C. sinensis* insect-fungi complex

262 represents a transcript encoding the MAT1-2-1 protein. The HMG-box_ROX1-like domain (379→591) of
263 the protein is 100% identical to the query sequence AEH27625 [Zhang *et al*. 2013; Xia *et al*. 2017].

264 III-3 Bayesian analysis of the MATα_HMGbox domains of the full-length MAT1-1-1 265 protein sequences

266 The MATα_HMGbox domain sequences of 19 full-length MAT1-1-1 proteins that have various amino
267 acid substitutions were subjected to the Bayesian clustering analysis, comparing with 6 MAT1-1-1 protein
268 sequences as references (Figure 1). The MATα_HMGbox domains of the full-length MAT1-1-1 proteins
269 were clustered into 5 Bayesian clusters (Figure 1), among which Clusters a and d are branched. The domains
270 of the 19 proteins with various amino acid substitutions were clustered into Clusters/Branches a2, b, d, and
271 e1−e2 in the Bayesian clustering tree. Cluster d has a longer clustering distance and Branch e2 has the
272 longest clustering distance. The MATα_HMGbox domains encoded by genome assemblies
273 JAAVMX010000001 (6,699,061→6,699,153 & 6,699,203→6,699.637) and LKHE01001116 (4183←4620
274 & 4667←4759) of *H. sinensis* strains 1229 and IOZ07 contain Y-to-M substitutions [Li *et al*. 2016; Shu *et*
275 *al*. 2020]. While the domains encoded by genome assembly ANOV01017390 (794←1129 & 1280←1396)
276 of *H. sinensis* strain Co18 and the metatranscriptome assembly OSIN7648 (151→675) of the *C. sinensis*
277 insect-fungi complex contain no mutant amino acids [Hu *et al*. 2013; Xia *et al*. 2017]. The domain encoded
278 by metatranscriptome assembly GAGW01008880 (714←1127) of the *C. sinensis* insect-fungi complex is
279 truncated at the N-terminus (Figure S3) [Xiang *et al*. 2014]. The MATα_HMGbox domains encoded by the
280 genome assembly ANOV01017390 and metatranscriptome assemblies OSIN7648 and GAGW01008880
281 were clustered into Branch a1 in the Bayesian clustering tree, together with the 6 authentic MAT1-1-1
282 proteins (Figure 1).

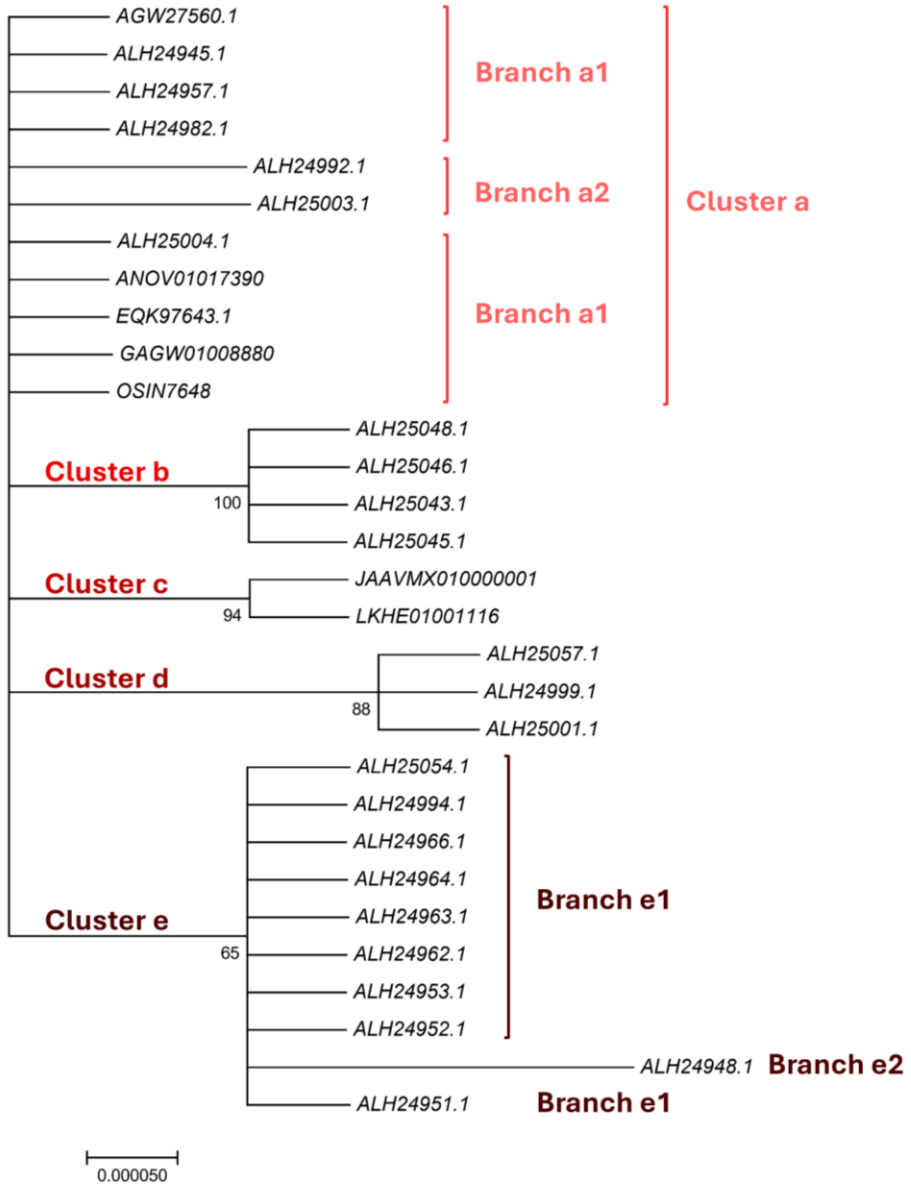


Figure 1. The Bayesian majority rule consensus clustering tree was inferred *via* MrBayes v3.2.7 software for the MATα_HMGbox domain sequences of the 25 MAT1-1-1 proteins of the *O. sinensis* strains and wild-type *C. sinensis* isolates and the corresponding domain sequences encoded by the genome and metatranscriptome assemblies of the *H. sinensis* strains and *C. sinensis* insect-fungi complexes.

Tables S8–S9 summarize the Bayesian clustering results with the amino acid substitutions or deletions in the MATα_HMGbox domains of MAT1-1-1 proteins derived from wild-type *C. sinensis* isolates or encoded by the genome and metatranscriptome assemblies of *H. sinensis* strains and natural *C. sinensis* insect-fungal complexes, respectively.

III-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 have various amino acid substitutions in the HMG-

box_ROX1-like domains derived from the 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-fungi complex were subjected to the Bayesian clustering analysis, comparing with the HMG-box_ROX1-like domains of 6 authentic MAT1-2-1 proteins. Figure 2 shows 2 branched clusters (a and b). Branches b2–b3 have longer clustering distances and Branch b2β has the longest clustering distance.

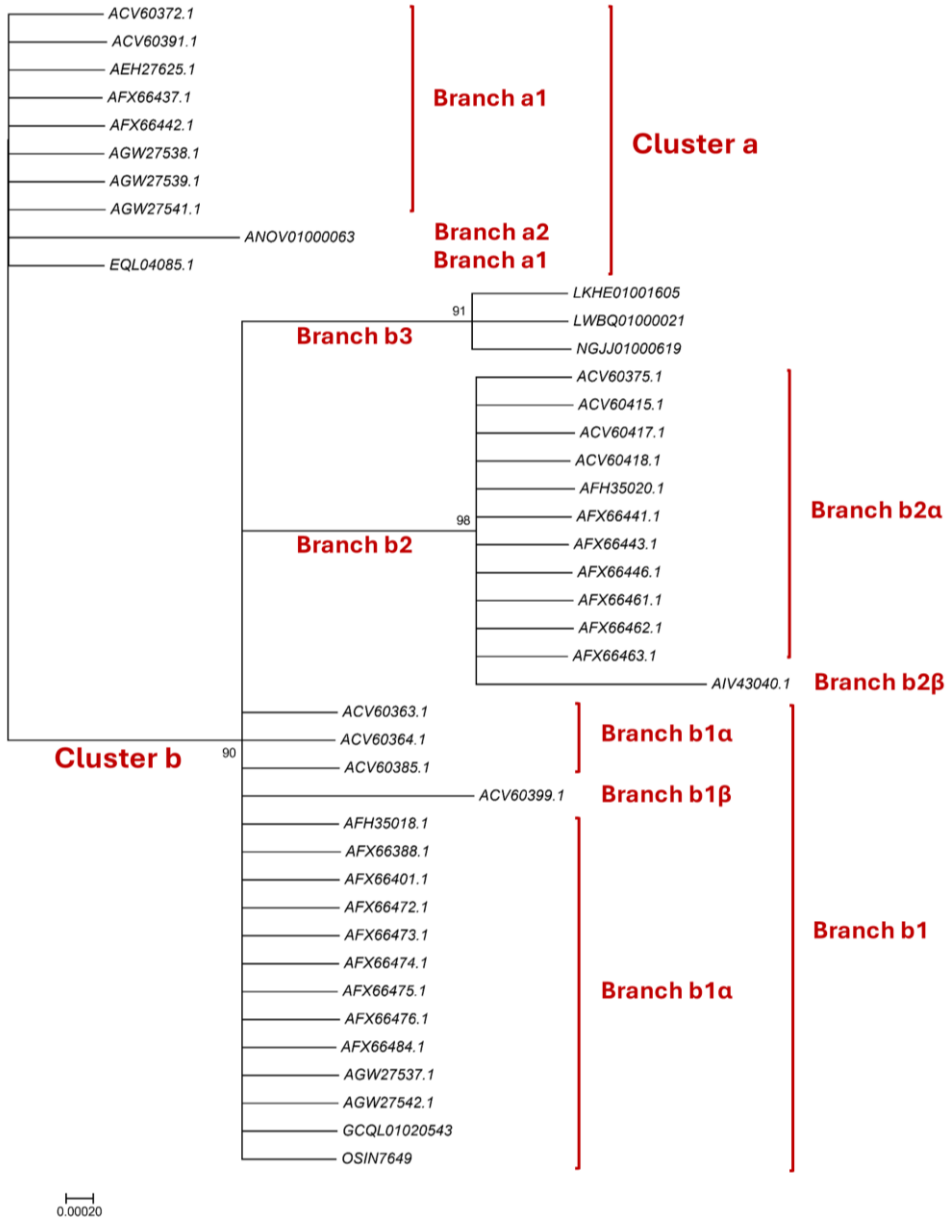


Figure 2. The Bayesian majority rule consensus clustering tree was inferred *via* MrBayes v3.2.7 software for the HMG-box_ROX1-like domain sequences of 35 full-length MAT1-2-1 proteins of the *O. sinensis* strains and wild-type *C. sinensis* isolates and the corresponding domain segments of translated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

The HMG-box_ROX1-like domain encoded by the genome assembly ANOV01000063 (9759→9851 & 9907→10,026) of *H. sinensis* strain Co18 contains a S-to-A substitution and was clustered into Branch a2 with a slightly longer clustering distance (Figures 2 and S4) [Hu *et al*. 2013]. The domains encoded by

308 other genome assemblies LKHE01001605 (14,016←14,135 & 14,191←14,283), LWBQ01000021
309 (239,029←239,148 & 239,204←239,269), and NGJJ01000619 (23,186←23,305 & 23,361←23,453) of *H.*
310 *sinensis* strains 1229, ZJB12195, and CC1406-20395 [Li *et al*. 2016; Jin *et al*. 2020; Liu *et al*. 2020],
311 respectively, contain Y-to-H and S-to-A substitutions and were clustered into Branch b3 in the Bayesian
312 clustering tree (Figures 2 and S4).

313 The HMG-box_ROX1-like domain of MAT1-2-1 protein encoded by the transcriptome assembly
314 GCQL01020543 (553←765) of the *H. sinensis* strain L0106 contains a Y-to-H substitution [Liu *et al*. 2015].
315 The domain of MAT1-2-1 protein encoded by the metatranscriptome assembly OSIN7649 (379→591) of
316 the *C. sinensis* insect-fungi complex is 100% identical to the reference MAT1-2-1 protein AEH27625
317 [Zhang *et al*. 2009; Xia *et al*. 2017]. The domains of the MAT1-2-1 proteins encoded by transcriptome and
318 metatranscriptome assemblies were clustered into Branch b1α in the Bayesian clustering tree (Figures 2,
319 S4).

320 Tables S10–S11 summarize the Bayesian clustering results with the amino acid substitutions or
321 deletions in the HMG-box_ROX1-like domains of MAT1-2-1 proteins derived from wild-type *C. sinensis*
322 isolates or encoded by genome, transcriptome and metatranscriptome assemblies of *H. sinensis* strains and
323 natural *C. sinensis* insect-fungal complexes, respectively.

324 III-5 Heteromorphic stereostructures of the MATα_HMGbox domains of full-length 325 MAT1-1-1 proteins derived from wild-type *C. sinensis* isolates

326 Section III-5 analyzes the correlations between the changes in the structures and hydrophobicity of the
327 MATα_HMGbox domains of full-length MAT1-1-1 proteins (under 7 AlphaFold 3D structure models; *cf*.
328 Table S8) derived from wild-type *C. sinensis* isolates, comparing with the reference MAT1-1-1 protein
329 AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 (*cf*.
330 Table S1) [Zhang *et al*. 2013]. The MATα_HMGbox domain sequences (amino acids 51→225) were
331 extended with external 9 amino acids both upstream and downstream of the domains because of the setting
332 of the ExPASy ProtScale plotting for the hydropathy, α-helices, β-sheets, β-turns, and coils with a window
333 size of 9 amino acid residues (*cf*. Section II-5 of Materials and Methods).

334 Figure 3 examined the structures of the MATα_HMGbox domain of MAT1-1-1 protein ALH24992
335 under the AlphaFold code A0A0N9R5B3 derived from the wild-type *C. sinensis* isolate SC09_65 (*cf*. Tables
336 S1and S8) [Zhang & Zhang 2015], comparing with the reference protein AGW27560 under the AlphaFold
337 code U3N942 derived from the *H. sinensis* strain CS68-2-1229 [Zhang *et al*. 2013]. Panel A shows an A-
338 to-T substitution (aliphatic alanine to threonine with polar side chains; hydropathy indexes of 1.8 to -0.7),
339 which caused reduced hydrophobicity, where the larger the hydropathy value, the stronger the

hydrophobicity; negative value indicates hydrophilicity [Kyte & Doolittle 1982]. The reduced hydrophobicity is illustrated by topology and waveform changes in Panel B for the hydropathy plot. These changes resulted in the altered secondary structures surrounding the mutation site in the MATα_HMGbox domain of MAT1-1-1 protein ALH24992 [Zhang & Zhang 2015], as illustrated by topology and waveform changes in Panels C–F for the ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils. As shown in Panels G–H for 3D structures, 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 DNA binding domain [Baxevanis *et al*. 1995; Thapar 2015]. However, the residue replacement with reduced hydrophobicity caused the altered stereostructure of the MATα_HMGbox domain of MAT1-1-1 protein ALH24992 under the AlphaFold code A0A0N9R5B3 probably through long-range spatial interactions.

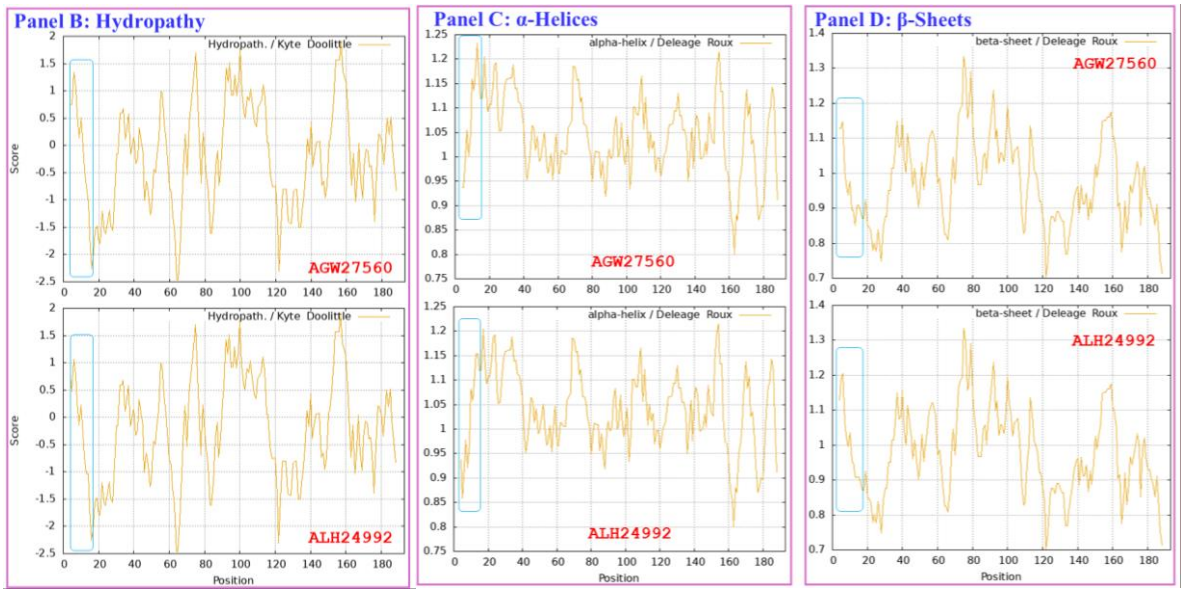
Panel A: Amino acid sequence alignment

```
AGW27560 42  ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTTL 101
ALH24992 42  ---------T-------------------------------------------------- 101

AGW27560 102 WHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQD 161
ALH24992 102 ------------------------------------------------------------ 161

AGW27560 162 DDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSHS 221
ALH24992 162 ------------------------------------------------------------ 221

AGW27560 222 GIMAPRAANCTPP                                                234
ALH24992 222 -------------                                                234
```



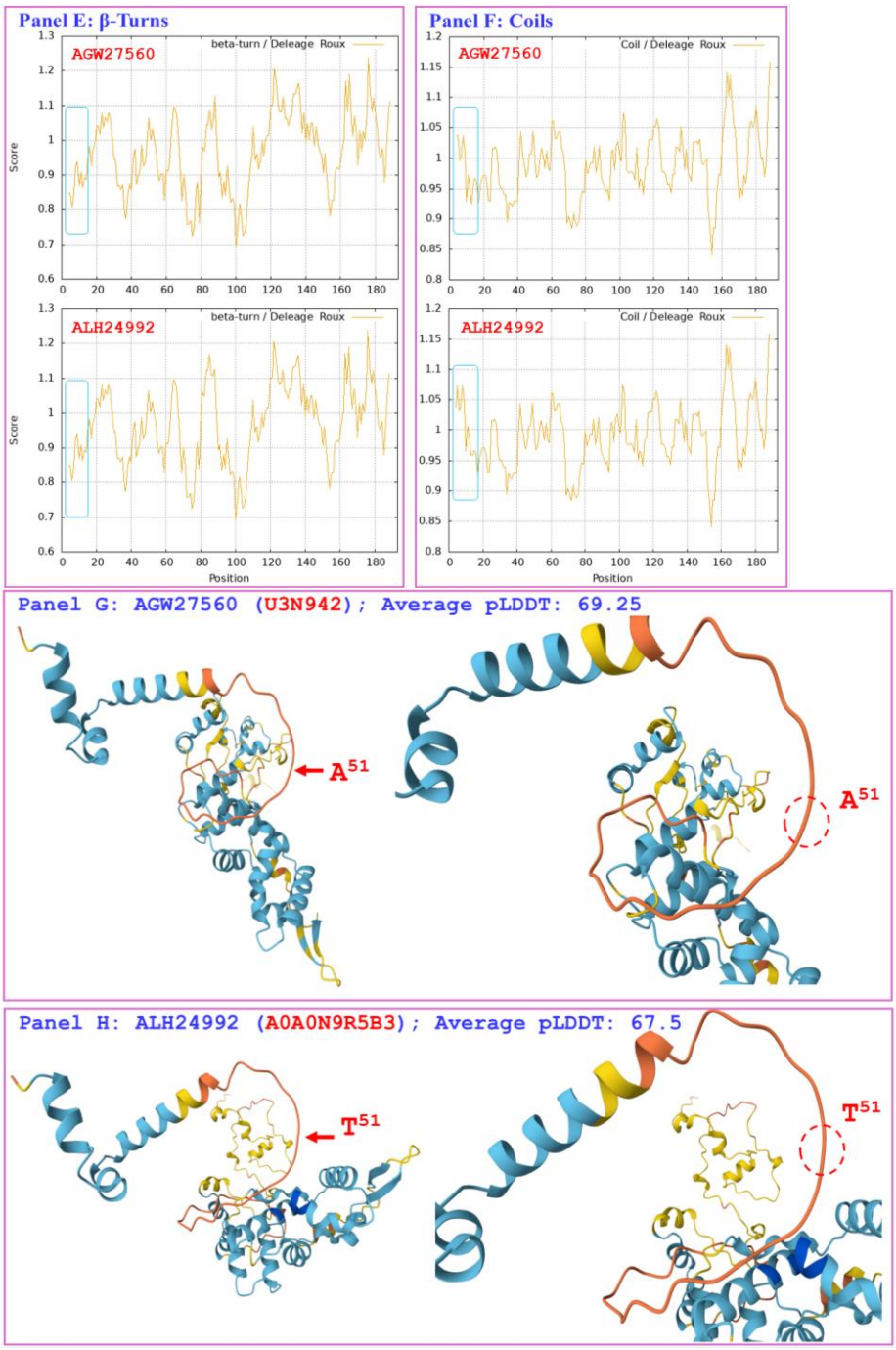


Figure 3. Correlation of the changes in the hydrophobicity, primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the mutant protein ALH24992 (under the AlphaFold code A0A0N9R5B3) derived from the wild-type *C. sinensis* isolate SC09_65. Panel A shows alignment of amino acid sequences of the MATα_HMGbox domain sequences of the MAT1-1-1 proteins with amino acid substitution in red; the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) with open rectangles in blue outlining the changes in topology and waveform of the plots. Each of 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. Model confidence for the AlphaFold-predicted 3D structures: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); ■ very low (pLDDT<50).

Figure 4 examined the structures of the MATα_HMGbox domain of MAT1-1-1 protein ALH24948 under the AlphaFold code A0A0N9QUF3 derived from the wild-type *C. sinensis* isolate GS09_143 (*cf.* Table S1) [Zhang & Zhang 2015], comparing with the reference protein AGW27560 under the AlphaFold code U3N942 derived from the *H. sinensis* strain CS68-2-1229 [Zhang *et al*. 2013]. Panel A shows substitutions of adjacent Q-to-S (glutamine to serine, both having polar side chains; hydropathy indexes of -3.5 to -0.8 [Kyte & Doolittle 1982]) and I-to-F (aliphatic isoleucine to aromatic phenylalanine; hydropathy indexes of 4.5 to 2.8), which caused the altered hydrophobicity, as illustrated in Panel B by topology and waveform changes for the hydropathy plot. These changes resulted in the altered secondary structures surrounding the mutation sites in the MATα_HMGbox domain, as illustrated by topology and waveform changes in Panels C–F for the ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils. As shown in Panels G–H for 3D structures, the adjacent amino acid substitutions are located at the end of one of the 3 α-helices that form the hydrophobic core stereostructure [Baxevanis *et al*. 1995; Thapar 2015] and caused the altered tertiary structure of the MATα_HMGbox domain of mutant protein ALH24948 under the AlphaFold code A0A0N9QUF3 [Zhang & Zhang 2015].

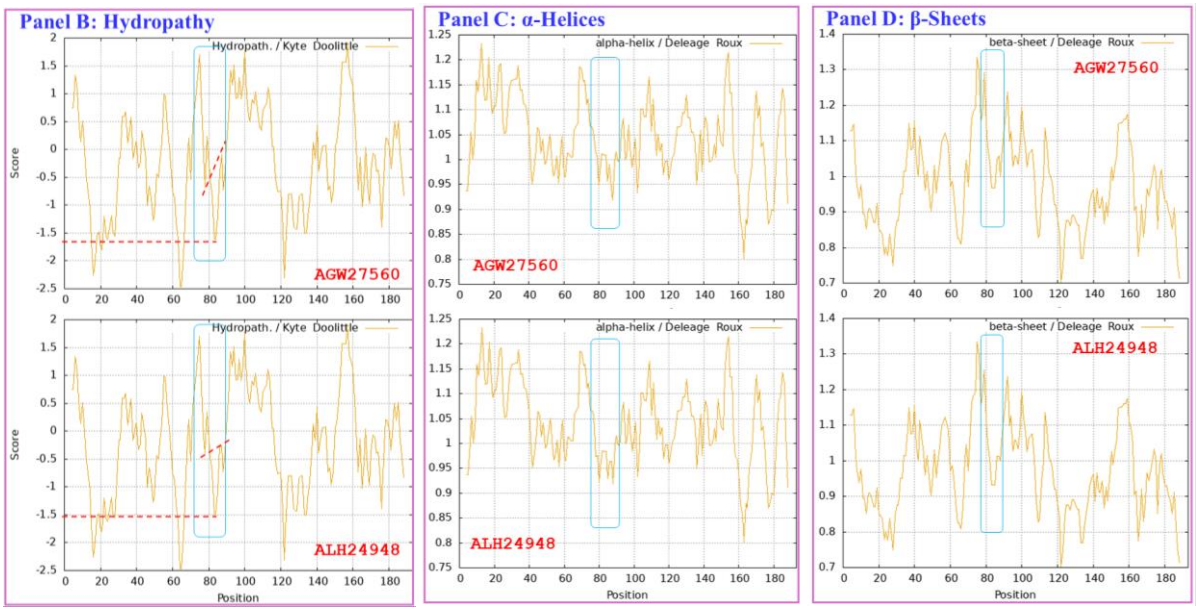
Panel A: Amino acid sequence alignment

```
AGW27560 42  ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTTL 101
ALH24948 42  ------------------------------------------------------------ 101

AGW27560 102 WHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQD 161
ALH24948 102 ----------------------SF------------------------------------ 161

AGW27560 162 DDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSHS 221
ALH24948 162 ------------------------------------------------------------ 221

AGW27560 222 GIMAPRAANCTPP                                                234
ALH24948 222 -------------                                                234
```



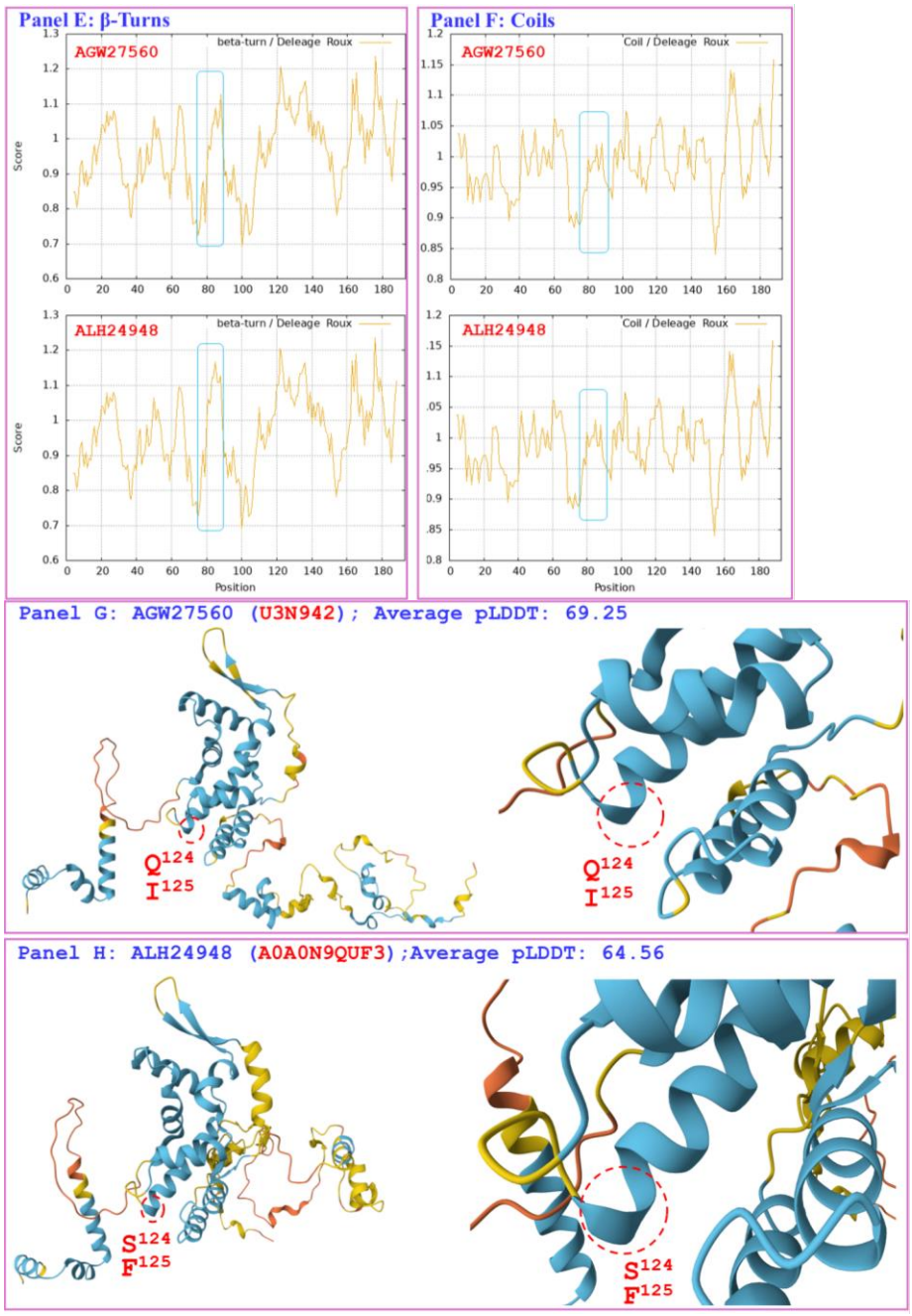


Figure 4. Correlation of the changes in the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the mutant MAT1-1-1 protein ALH24948 (under the AlphaFold code A0A0N9QUF3) derived from the wild-type *C. sinensis* isolate GS09_143. Panel A shows alignment of amino acid sequences of the MATα_HMGbox domain sequences of the MAT1-1-1 proteins with amino acid substitutions in red; the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) with open rectangles in blue outlining the changes in topology and waveform of the plots. Each of 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. Model confidence for the AlphaFold-predicted 3D structures: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); ■ very low (pLDDT<50).

Figure 5 examined the hydrophobicity and structures of the MATα_HMGbox domains of 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 (*cf*. Table S1) [Zhang & Zhang 2015], respectively, comparing with the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [Zhang *et al*. 2013]. Panel A shows substitutions of R-to-I (basic arginine to aliphatic isoleucine; hydropathy indexes of -4.5 to 4.5 [Kyte & Doolittle 1982]) and P-to-T (proline to threonine both with polar neutral side chains; hydropathy indexes of -1.6 to -0.7), which caused increases in hydrophobicity, as illustrated by topology and waveform changes in Panel B for the hydropathy plot. These changes resulted in the altered secondary structures surrounding the mutation sites in the MATα_HMGbox domains of MAT1-1-1 proteins ALH25043, ALH25045, ALH25046, and ALH25048 [Zhang & Zhang 2015], as illustrated by topology and waveform changes in Panels C–F for the ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils. As shown in Panels G–H for 3D structures, the R-to-I and P-to-T substitutions are located inside and outside the hydrophobic core of the 3 α-helices, respectively, with substantial increases in hydrophobicity altered the tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins under the AlphaFold code A0A0N9QMS9.

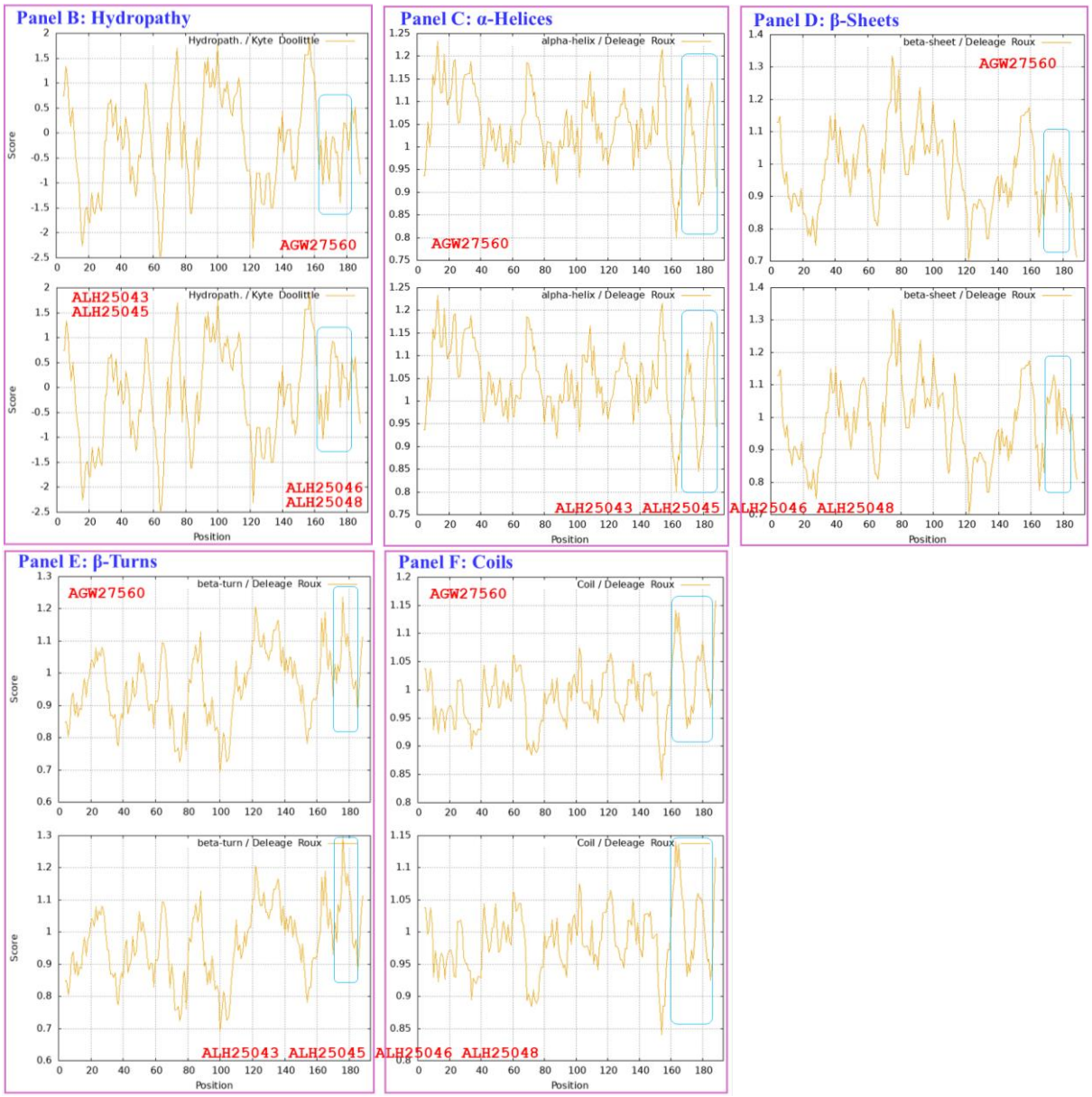
Panel A: Amino acid sequence alignment

```
AGW27560 42  ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTTL 101
ALH25043 42  ------------------------------------------------------------ 101
ALH25045 42  ------------------------------------------------------------ 101
ALH25046 42  ------------------------------------------------------------ 101
ALH25048 42  ------------------------------------------------------------ 101

AGW27560 102 WHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQD 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
```


Panel B: Hydropathy
Hydropath. / Kyte Doolittle
AGW27560
ALH25043
ALH25045
ALH25046
ALH25048
Score
Position
Panel C: α-Helices
alpha-helix / Deleage Roux
AGW27560
ALH25043 ALH25045 ALH25046 ALH25048
Position
Panel D: β-Sheets
beta-sheet / Deleage Roux
AGW27560
Position
Panel E: β-Turns
beta-turn / Deleage Roux
AGW27560
ALH25043 ALH25045 ALH25046 ALH25048
Score
Position
Panel F: Coils
Coil / Deleage Roux
AGW27560
Position

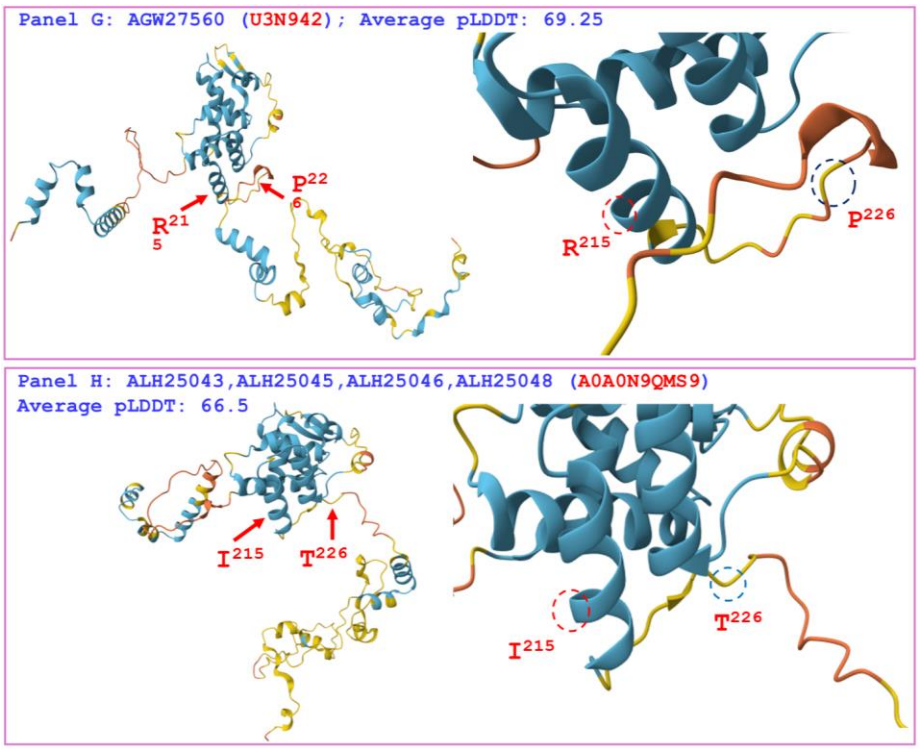


Figure 5. Correlation of the changes in the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the mutant MAT1-1-1 proteins ALH25043, ALH25045, ALH25046, and ALH25048 (under the AlphaFold code A0A0N9QMS9) derived from wild-type *C. sinensis* isolates YN09_22, YN09_51, YN09_6, and YN09_64, respectively. Panel A shows alignment of amino acid sequences of the MATα_HMGbox domain sequences of the MAT1-1-1 proteins with amino acid substitutions in red; the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) and hydropathy with open rectangles in blue outlining the changes in topology and waveform of the plots. Each of 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. Model confidence for the AlphaFold-predicted 3D structures: very high (pLDDT>90); high (90>pLDDT>70); low (70>pLDDT>50); very low (pLDDT<50).

Figure 6 examined the hydrophobicity and structures of the MATα_HMGbox domains of 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 (*cf.* Table S1) [Zhang & Zhang 2015], respectively, comparing with the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [Zhang *et al*. 2013].

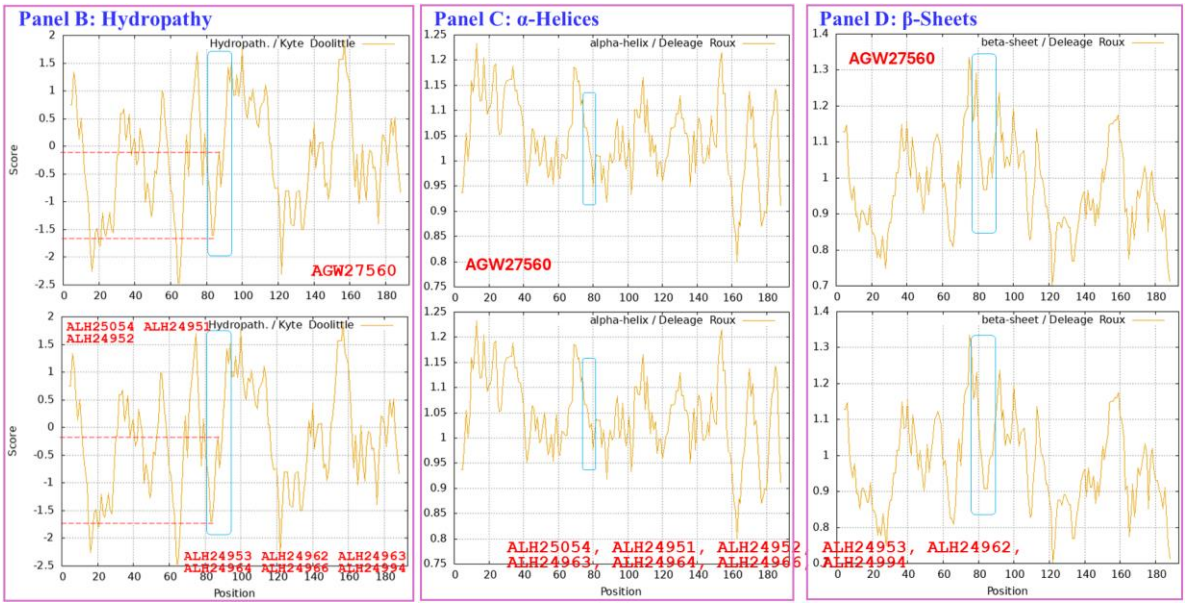
Panel A: Amino acid sequence alignment

```
AGW27560 42  ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTTL 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 WHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQD 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
```



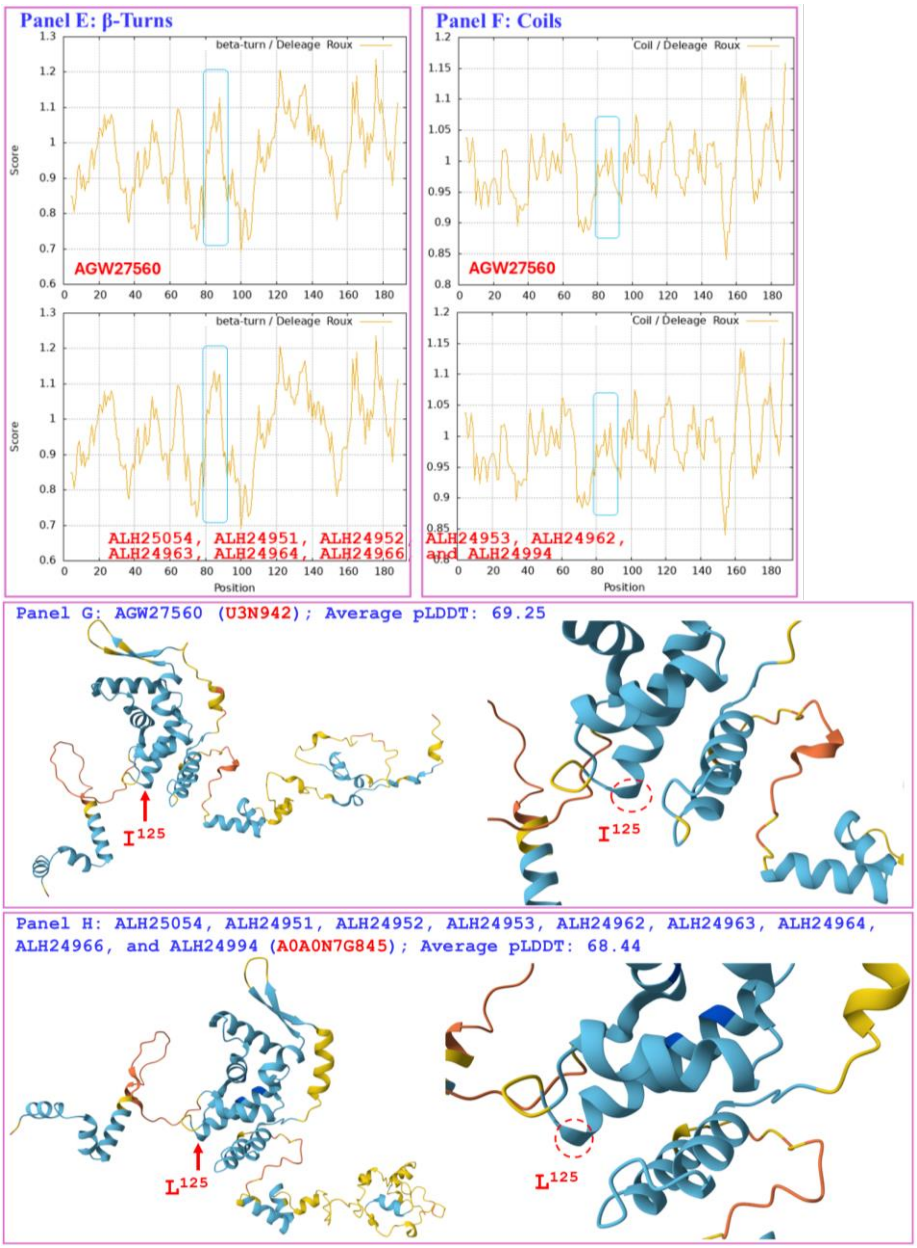


Figure 6. Correlation of the changes in the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the mutant MAT1-1-1 proteins ALH25054, ALH24951, ALH24952, ALH24953, ALH24962, ALH24963, ALH24964, ALH24966, and ALH24994 (under the AlphaFold code A0A0N7G845) derived from the wild-type *C. sinensis* isolate GS09_311, GS09_229, GS09_281, GS10_1, QH09_164, QH09_173, QH09_201, QH09_210, and SC09_87, respectively. Panel A shows alignment of amino acid sequences of the MATα_HMGbox domain sequences of the MAT1-1-1 proteins with amino acid substitutions in red; the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) with open rectangles in blue outlining the changes in topology and waveform of the plots. Each of 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. Model confidence for the AlphaFold-predicted 3D structures: very high (pLDDT>90); high (90>pLDDT>70); low (70>pLDDT>50); very low (pLDDT<50).

Panel A of Figure 6 shows I-to-L substitutions (aliphatic isoleucine to aliphatic leucine; hydropathy indexes of 4.5 to 3.8 [Kyte & Doolittle 1982]), which caused reduced hydrophobicity as illustrated by

topology and waveform changes in Panel B for the hydropathy plot. These changes resulted in the altered secondary structures surrounding the mutation sites in the MATα_HMGbox domains of MAT1-1-1 proteins ALH25054, ALH24951, ALH24952, ALH24953, ALH24962, ALH24963, ALH24964, ALH24966, and ALH24994 (*cf*. Table S1) [Zhang & Zhang 2015], as illustrated by topology and waveform changes in Panels C–F for the ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils. As shown in Panels G–H for the 3D structures, the amino acid substitutions with reduced hydrophobicity are located within the core structures of α-helices and destabilized the hydrophobicity core [Baxevanis *et al*. 1995; Thapar 2015], which caused the altered tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins under the AlphaFold code A0A0N7G845.

Figure 7 examined the hydrophobicity and structures of the MATα_HMGbox domains of MAT1-1-1 proteins ALH24999 and ALH25057 (under the AlphaFold code A0A0N9QMT4) derived from the wild-type *C. sinensis* isolates XZ07_H2 and XZ12_16 (*cf*. Table S1) [Zhang & Zhang 2015], respectively, comparing with the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [Zhang *et al*. 2013].

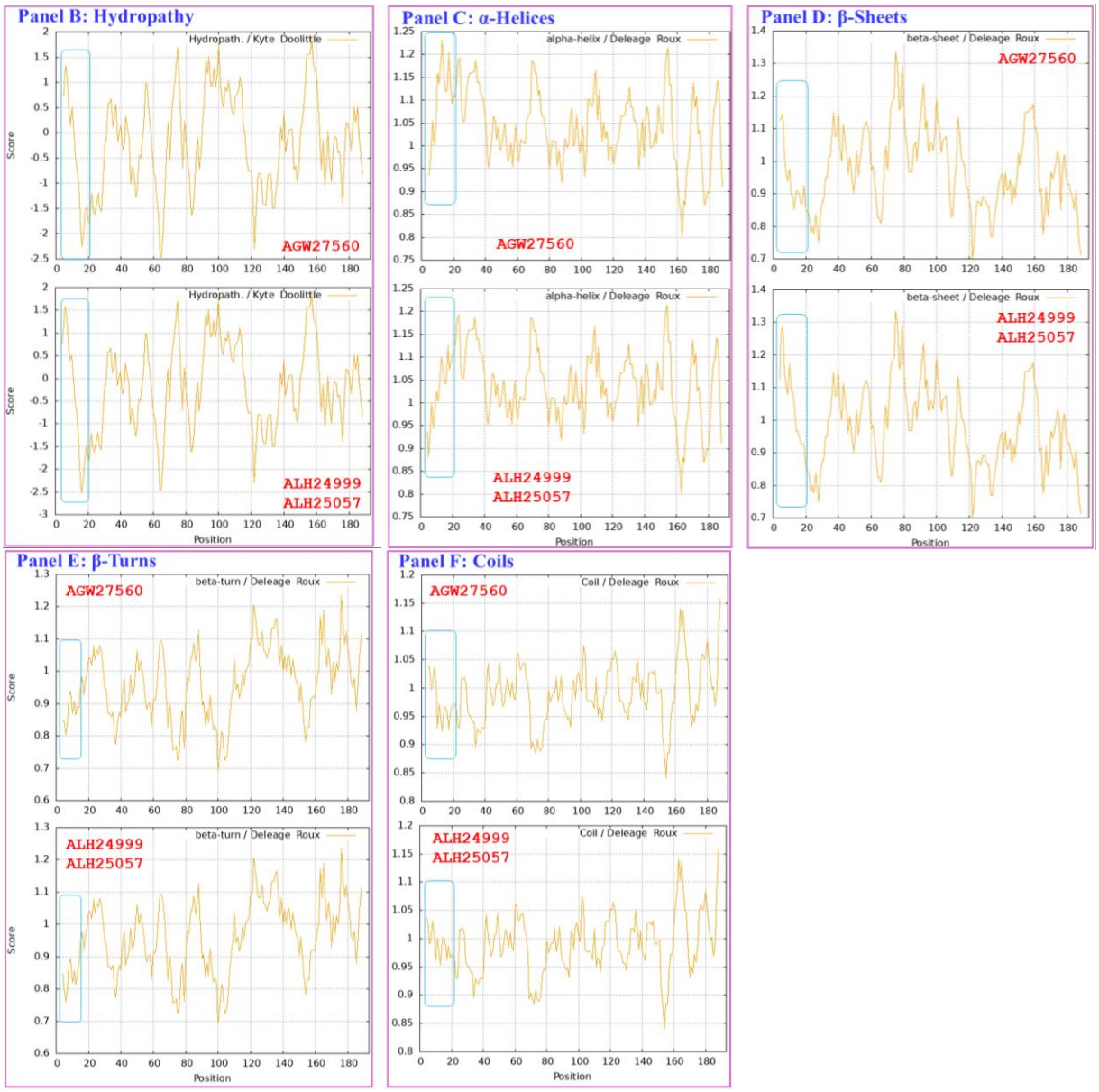
Panel A: Amino acid sequence alignment

```
AGW27560 42  ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTTL 101
ALH24999 42  ---------V----T---------------------------------------------- 101
ALH25057 42  ---------V----T---------------------------------------------- 101

AGW27560 102 WHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQD 161
ALH24999 102 ------------------------------------------------------------ 161
ALH25057 102 ------------------------------------------------------------ 161

AGW27560 162 DDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSHS 221
ALH24999 162 ------------------------------------------------------------ 221
ALH25057 162 ------------------------------------------------------------ 221

AGW27560 222 GIMAPRAANCTPP                                                234
ALH24999 222 -------------                                                234
ALH25057 222 -------------                                                234
```


Panel B: Hydropathy
Hydropath. / Kyte Doolittle
AGW27560
ALH24999
ALH25057
Score
Position
Panel C: α-Helices
alpha-helix / Deleage Roux
AGW27560
ALH24999
ALH25057
Position
Panel D: β-Sheets
beta-sheet / Deleage Roux
AGW27560
ALH24999
ALH25057
Position
Panel E: β-Turns
beta-turn / Deleage Roux
AGW27560
ALH24999
ALH25057
Score
Position
Panel F: Coils
Coil / Deleage Roux
AGW27560
ALH24999
ALH25057
Position

469

470

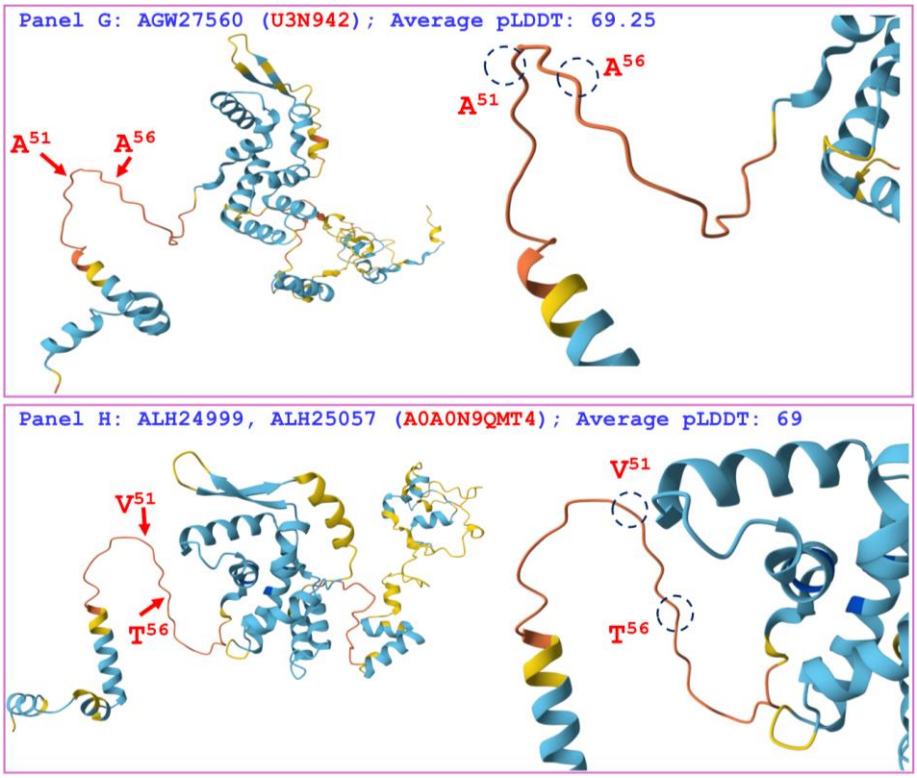


Figure 7. Correlation of the changes in the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the mutant MAT1-1-1 proteins ALH24999 and ALH25057 (under the AlphaFold code A0A0N9QMT4) derived from the wild-type *C. sinensis* isolates XZ07_H2 and XZ12_16, respectively. Panel A shows alignment of amino acid sequences of the MATα_HMGbox domain sequences of the MAT1-1-1 proteins with amino acid substitutions in red; the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) with open rectangles in blue outlining the changes in topology and waveform of the plots. Each of 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. Model confidence for the AlphaFold-predicted 3D structures: very high (pLDDT>90); high (90>pLDDT>70); low (70>pLDDT>50); very low (pLDDT<50).

Panel A of Figure 7 shows substitutions of A-to-V (aliphatic alanine to aliphatic valine; hydropathy indexes of 1.8 to 4.2 [Kyte & Doolittle 1982]) and A-to-T (aliphatic isoleucine to threonine with polar neutral side chains; hydropathy indexes of 1.8 to -0.7), which caused changes in hydrophobicity, as illustrated by topology and waveform changes in Panel B for the hydropathy plot. These changes resulted in the altered secondary structures surrounding the mutation sites in the MATα_HMGbox domain of MAT1-1-1 proteins ALH24999 and ALH25057 [Zhang & Zhang 2015], as illustrated by topology and waveform changes in Panels C–F for the ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils. As shown in Panels G–H for 3D structures, the mutation sites are located upstream the hydrophobic core structure of 3 α-helices in the MATα_HMGbox domains and altered the tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins under the AlphaFold code A0A0N9QMT4.

Figure 8 examined the hydrophobicity and structures of the MATα_HMGbox domain of MAT1-1-1

protein ALH25001 (under the AlphaFold code A0A0N9R4Q4) derived from the wild-type *C. sinensis* isolate XZ05_2 (*cf.* Table S1) [Zhang & Zhang 2015], comparing with the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [Zhang *et al*. 2013]. Panel A shows substitutions of A-to-V (aliphatic alanine to aliphatic valine; hydropathy indexes of 1.8 to 4.2 [Kyte & Doolittle 1982]) and A-to-T (aliphatic isoleucine to threonine with polar neutral side chains; hydropathy indexes of 1.8 to -0.7), which caused changes in hydrophobicity, as illustrated by topology and waveform changes in Panel B for the hydropathy plot. These changes resulted in the altered secondary structures surrounding the mutation sites in the MATα_HMGbox domain of MAT1-1-1 protein ALH25001, as illustrated by topology and waveform changes in Panels C–F for the ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils. As shown in Panels G–H for the 3D structures, the amino acid substitutions are located upstream the core structures of 3 α-helices of the domains and altered the tertiary structures of the domains of MAT1-1-1 proteins under the AlphaFold code A0A0N9R4Q4.

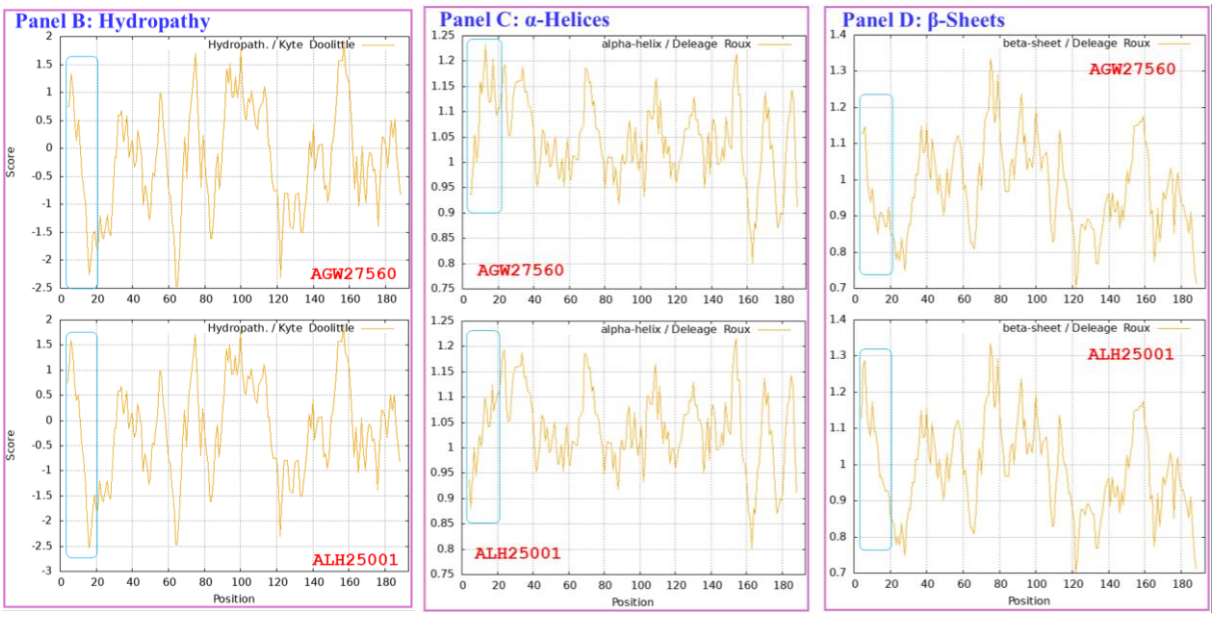
Panel A: Amino acid sequence alignment

```
AGW27560 42  ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTTL 101
ALH25001 42  ---------V----T------------------------------------------- 101

AGW27560 102 WHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQD 161
ALH25001 102 ------------------------------------------------------------ 161

AGW27560 162 DDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSHS 221
ALH25001 162 ------------------------------------------------------------ 221

AGW27560 222 GIMAPRAANCTPP                                                234
ALH25001 222 -------------                                                234
```



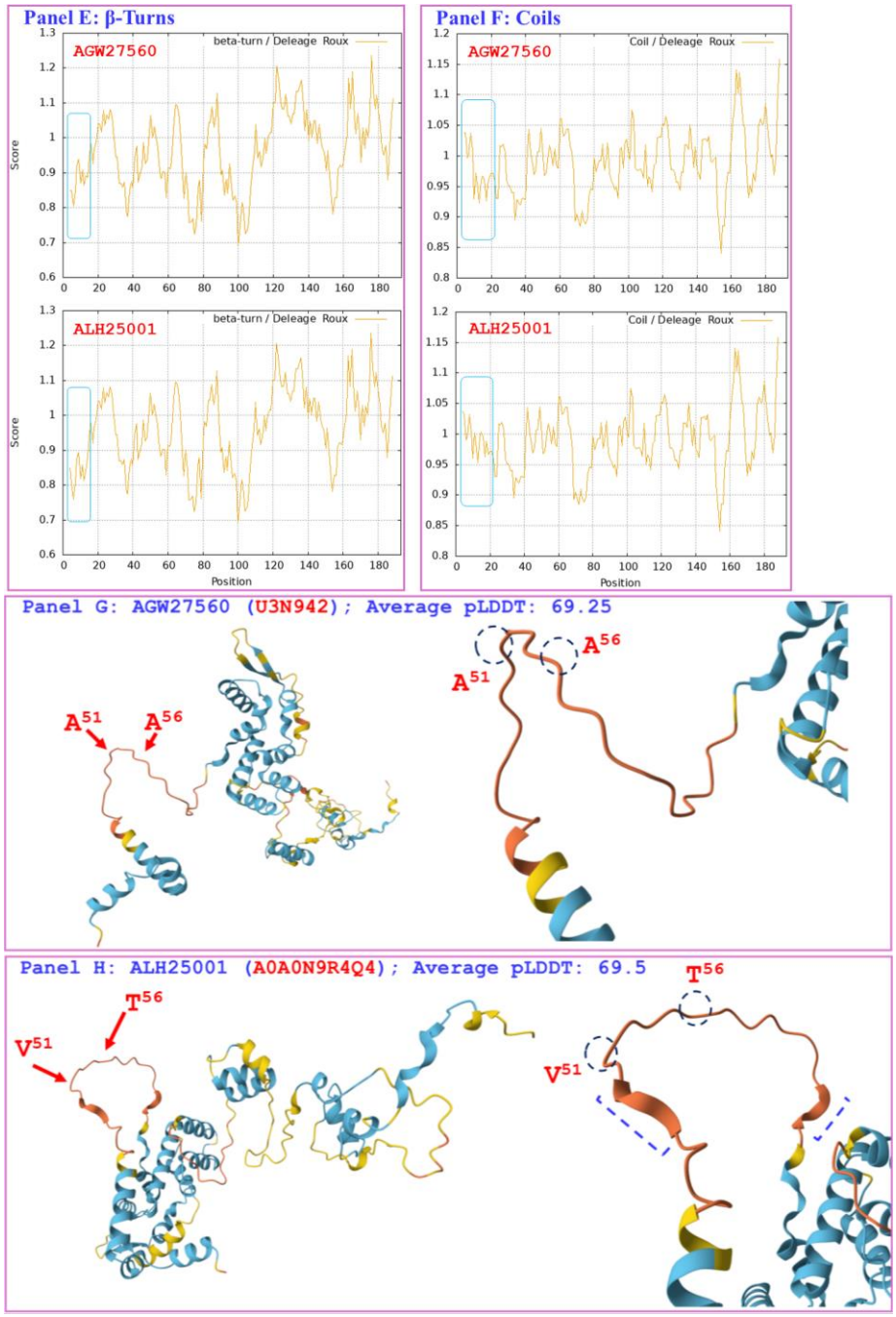


Figure 8. Correlation of the changes in the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 and the mutant MAT1-1-1 protein ALH25001 (under the AlphaFold code A0A0N9R4Q4) derived from the wild-type *C. sinensis* isolate XZ05_2. Panel A shows alignment of amino acid sequences of the MATα_HMGbox domain sequences of the MAT1-1-1 proteins with amino acid substitutions in red; the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) with open rectangles in blue outlining the changes in topology and waveform of the plots. Each of 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 with dashed brackets in blue indicating 2 extra β-sheets structures. Model confidence for the AlphaFold-predicted 3D structures: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); ■ very low (pLDDT<50).

Panels A of Figures 9–10 show the same substitutions of A-to-V and A-to-T in the MATα_HMGbox

522 domains of the protein ALH25001 under AlphaFold code A0A0N9R4Q4 and the proteins ALH24999 and
523 ALH25057 under the AlphaFold code A0A0N9QMT4 [Zhang & Zhang 2015]. However, Panels H of
524 Figures 9–10 show different tertiary structures of the domains of the MAT1-1-1 proteins under 2 different
525 AlphaFold 3D morphs. In particular, the protein ALH25001 contains 2 extra short β-sheet structures both
526 upstream and downstream of the mutation sites in the MATα_HMGbox domain shown in Figure 8, which
527 are marked with dashed brackets in blue, as compared with the tertiary structures for the mutant proteins
528 ALH24999 and ALH25057 shown in Figure 7 [Zhang *et al*. 2013].

529 Figure 9 examined the hydrophobicity and structures of the MATα_HMGbox domain of MAT1-1-1
530 protein ALH25003 (under the AlphaFold code A0A0N7G850) derived from the wild-type *C. sinensis*
531 isolate XZ05_6 (*cf*. Table S1) [Zhang & Zhang 2015], comparing with the reference protein AGW27560
532 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [Zhang *et al*. 2013].
533 Panel A shows a S-to-G substitution (serine with polar neutral side chains to unique amino acid glycine;
534 hydropathy indexes of -0.8 to -0.4 [Kyte & Doolittle 1982], which caused a slight increase in hydrophobicity,
535 as illustrated by topology and waveform changes in Panel B for the hydropathy plot. These changes caused
536 the altered secondary structures surrounding the mutation site in the MATα_HMGbox domain of mutant
537 MAT1-1-1 protein ALH25003, as illustrated by topology and waveform changes in Panels C–F for the
538 ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils. As shown in Panels G–H for the 3D
539 structures, the last animo acid (serine) positioning at the end of one of the 3 α- helices that form the
540 hydrophobic core structure of the MATα_HMGbox domain [Baxevanis *et al*. 1995; Thapar 2015] was
541 substituted with glycine that became the first residue after the α-helix in the domain. The single residue
542 substitution plus other E-to-K and Y-to-H downstream the domain (*cf*. Figure S3) caused the significantly
543 altered tertiary structure of the protein ALH25003 under the unique AlphaFold code A0A0N7G850.

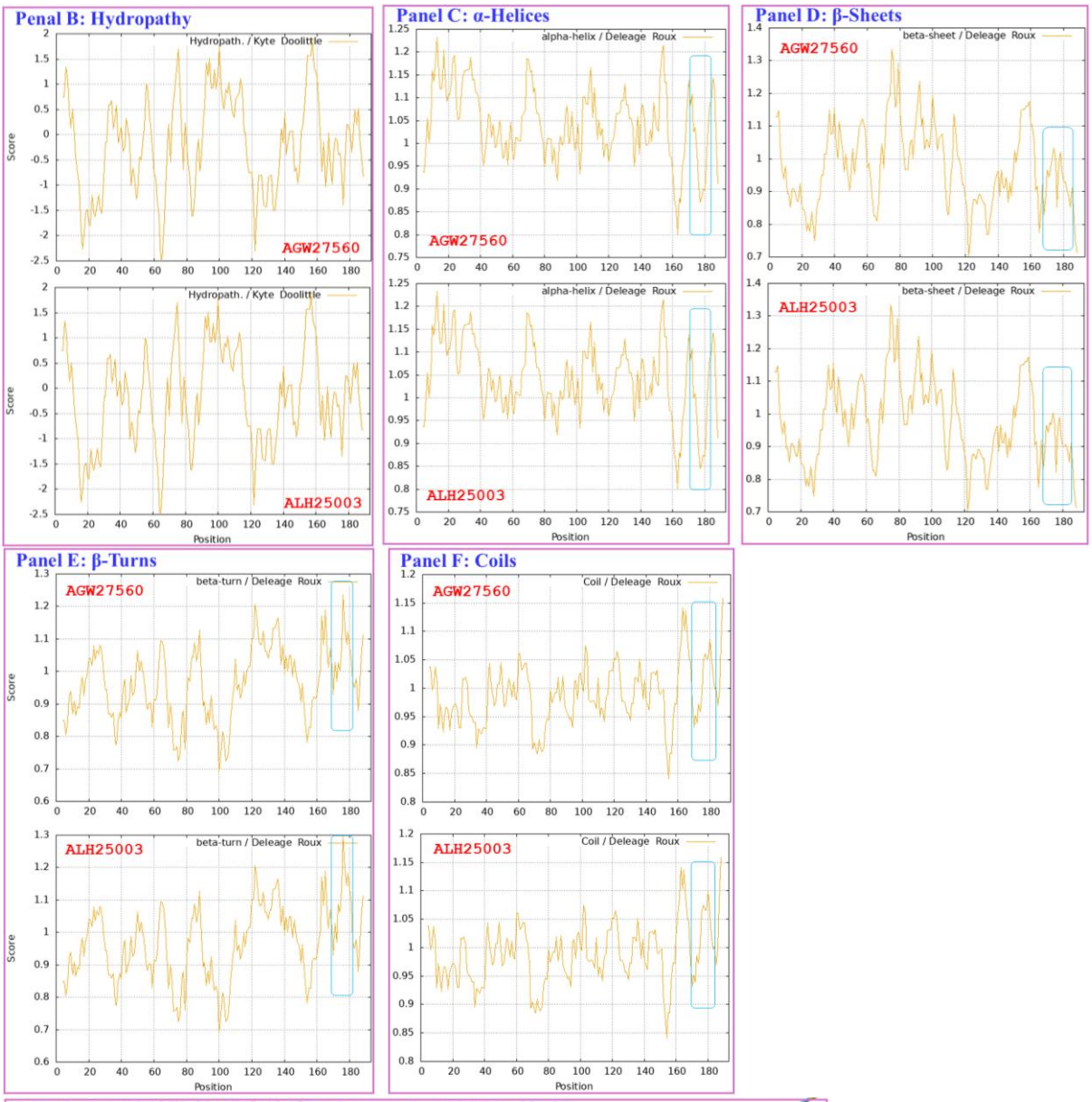
Panel A: Amino acid sequence alignment

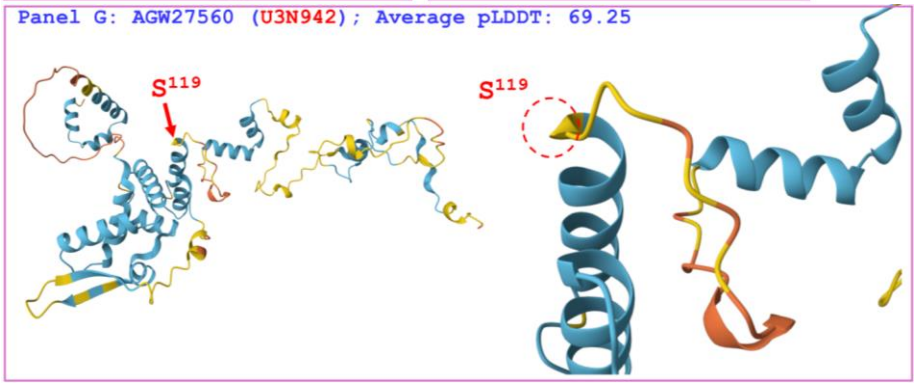
```
AGW27560 42  ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTTL 101
ALH25003 42  ------------------------------------------------------------ 101

AGW27560 102 WHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQD 161
ALH25003 102 ------------------------------------------------------------ 161

AGW27560 162 DDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSHS 221
ALH25003 162 ---------------------------------------------------------G-- 221

AGW27560 222 GIMAPRAANCTPP                                                234
ALH25003 222 -------------                                                234
```
544


Penal B: Hydropathy
Hydropath. / Kyte Doolittle
AGW27560
ALH25003
Score
Position
Panel C: α-Helices
alpha-helix / Deleage Roux
AGW27560
ALH25003
Position
Panel D: β-Sheets
beta-sheet / Deleage Roux
AGW27560
ALH25003
Position
Panel E: β-Turns
beta-turn / Deleage Roux
AGW27560
ALH25003
Score
Position
Panel F: Coils
Coil / Deleage Roux
AGW27560
ALH25003
Position


Panel G: AGW27560 (U3N942); Average pLDDT: 69.25
S119
S119

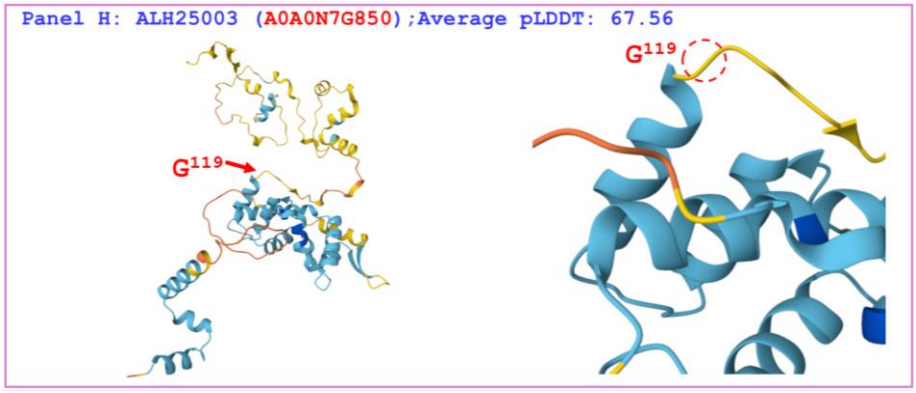


Figure 9. Correlation of the changes in the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the mutant MAT1-1-1 protein ALH25003 (under the AlphaFold code A0A0N7G850) derived from the wild-type *C. sinensis* isolate XZ05_6. Panel A shows alignment of amino acid sequences of the MATα_HMGbox domain sequences of the MAT1-1-1 proteins with amino acid substitutions in red; the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) with open rectangles in blue outlining the changes in topology and waveform of the plots. Each of 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. Model confidence for the AlphaFold-predicted 3D structures: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); ■ very low (pLDDT<50).

In aggregation, Panels A of Figures 5–11 demonstrated the 1–2 amino acid substitutions in the MATα_HMGbox domains of MAT1-1-1 proteins under various AlphaFold 3D structure codes derived from wild-type *C. sinensis* isolates, comparing with the reference MAT1-1-1 protein AGW27560 under the AlphaFold code U3N942 [Zhang *et al*. 2013]. The substitutions in the domains caused the changes in the topologies and waveforms of ExPASy ProtScale plots shown in Panels B–F of Figures 5–11, indicating the subsequently altered hydrophobicity and the secondary structures (α-helices, β-sheets, β-turns, and coils) of the mutant full-length proteins. The changes in hydrophobicity and primary and secondary structures of the domains then caused alterations of the AlphaFold-predicted 3D structures of the MATα_HMGbox domains of the proteins under various AlphaFold codes (*cf*. Panels H of Figures 5–11). The mutant domain sequences with 1 or 2 amino acid substitutions were clustered into different clades in the Bayesian clustering tree (*cf*. Figure 1, Table S8) and the altered 3D structures, ultimately changing the DNA binding affinities and the functionalities of gene transcription regulations relating to the sexual reproduction of *O. sinensis*.

III-6 Diverse stereostructures of the MATα_HMGbox domains of N-terminally truncated MAT1-1-1 proteins derived from *O. sinensis* strains

In addition to the 118 full-length proteins analyzed above, 20 other MAT1-1-1 proteins (AGW27517–AGW27536) are truncated at the N-termini of the MATα_HMGbox domains (Figure 10) [Bushley *et al*. 2013; Li *et al*. 2013, 2025]. These truncated proteins exhibited 66–93% query coverage

compared with the domain of the query sequence MAT1-1-1 protein AGW27560 under the AlphaFold UniProt code U3N942 [Zhang *et al*. 2013; Li *et al*. 2025]. The truncated MAT1-1-1 proteins derived from *O. sinensis* strains exhibited variable lengths shown in Figure 10, although the *MAT1-1-1* cDNA encoding the truncated proteins were amplified using the same pair of forward primer *m1cDNAF3* and revere compliment primer *m1cDNAR5* shown in Figure S1 [Bushley *et al*. 2013]. In addition to the variable truncation mutations, the protein AGW27526 contained an additional A-to-T substitution (aliphatic isoleucine to threonine with polar neutral side chains; hydropathy indexes of 1.8 to -0.7, *cf*. Table S7; indicating reduced hydrophobicity) and the protein AGW27536 had a deletion of leucine residue (hydropathy index of 3.8). The 20 truncated proteins exhibited diverse tertiary structures under 9 different AlphaFold 3D structure codes (Figure 10).

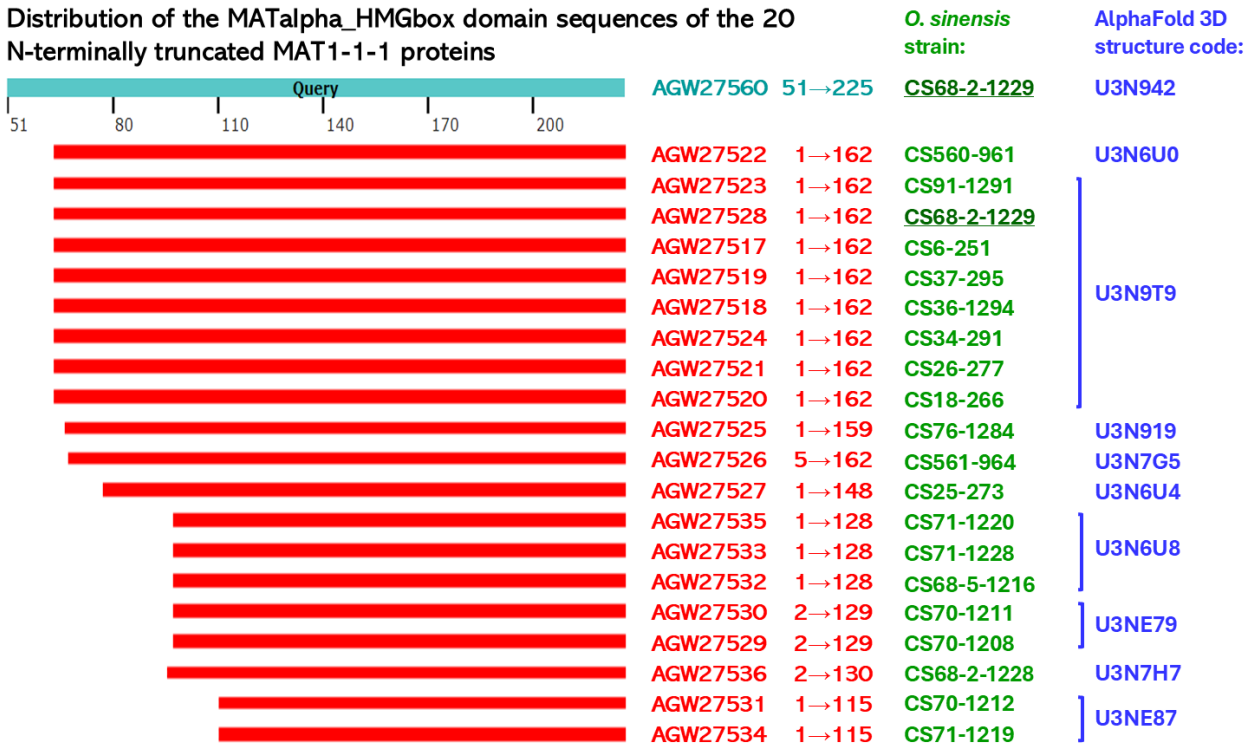


Figure 10. Sequence distribution of the N-terminally truncated MATα_HMGbox domains of 20 subject MAT1-1-1 proteins derived from the listed *O. sinensis* strains aligned with the full-length MATα_HMGbox domain (51→225) of the query MAT1-1-1 protein AGW27560 of the *H. sinensis* strain CS68-2-1229.

The 20 *O. sinensis* strains listed in Figure 10 were isolated either from the fresh caterpillar bodies or the cultures of mono-ascospores, marked as TS and SS in Table S4 reported by Li *et al*. [2013], respectively, of the *C. sinensis* insect-fungal complexes. Ten (CS18-266, CS25-273, CS37-295, CS68-2-1229, CS71-1219, CS71-1228, CS76-1284, CS91-1291, CS560-961, and CS561-964) of the 20 strains were genetically homogenous belonging to GC-biased Genotype #1 *H. sinensis* shown in blue in Table S4. However, 5 other strains in red in Table S4 (CS6-251, CS34-291, CS68-2-1228, CS70-1208, and CS71-1220) were

597 genetically heterogeneous, cooccurring GC-biased Genotype #1 and AT-biased Genotype #5 of *O. sinensis*
598 [Bushley *et al*. 2013; Li *et al*. 2013]. Worth noting, the 17 GC- and AT-biased genotypes of *O. sinensis* are
599 genome independent, belonging to different fungi [Kinjo & Zang 2001; Stensrud *et al*. 2007; Xiao *et al*.
600 2009; Yao & Zhu 2016; Zhu & Li 2017; Li *et al*. 2020b, 2022, 2023d, 2023e]. The remaining 5 strains
601 (CS26-277, CS36-1294, CS68-5-1216, CS70-1211, and CS70-1212) shown in black in Table S4 lacked ITS
602 molecular genotyping information in the GenBank database.

603 The variable lengths of MATα_HMGbox domains of the 20 truncated proteins shown in Figure 10
604 were derived from the *MAT1-1-1* cDNA that were amplified using the same pair of primers (*m1cDNAF3*
605 and *m1cDNAR5*) [Bushley *et al*. 2013]. The variable MAT1-1-1 proteins and the heterogeneous genotypic
606 backgrounds for the *O. sinensis* strains (CS6-251, CS34-291, CS68-2-1228, CS70-1208, and CS71-1220;
607 *cf*. Table S4) indicate different fungal sources of the truncated proteins, derived from either GC-biased
608 Genotype #1 or AT-biased Genotype #5 of *O. sinensis* fungi. While Strains CS26-277, CS36-1294, CS68-
609 5-1216, CS70-1211, and CS70-1212 that produced variable MAT1-1-1 proteins might also present fungal
610 heterogeneity but to date without ITS molecular genotyping evidence in the GenBank database, in addition
611 to the 4 other ascosporic strains (1210, 1213, 1223, and 1226) without ITS molecular taxonomic evidence
612 [Bushley *et al*. 2013; Li *et al*. 2013, 2025].

613 III-7 Diverse primary and secondary structures of the MATα_HMGbox domains of 614 MAT1-1-1 proteins encoded by genome and metatranscriptome assemblies of *H.* 615 *sinensis* and *C. sinensis* insect-fungi complex

616 Figure 11 examined the hydrophobicity and secondary structures (α-helices, β-sheets, β-turns, and
617 coils) of the MATα_HMGbox domain of MAT1-1-1 protein that were encoded by the genome assemblies
618 JAAVMX010000001 (6,699,061→6,699,153 & 6,699,203→6,699,637, removed the intron section) and
619 LKHE01001116 (4183←4614 & 4655←4759, removed the intron section) of the *H. sinensis* strains IOZ07
620 and 1229, comparing with the reference MAT1-1-1 protein AGW27560 derived from the *H. sinensis* strain
621 CS68-2-1229 [Zhang *et al*. 2009; Li *et al*. 2016a; Shu *et al*. 2020]. Panel A shows a Y-to-M substitution
622 (aromatic tyrosine to methionine with polar neutral side chains; hydropathy indexes of -1.3 to 1.9 [Kyte &
623 Doolittle 1982]), which caused increased hydrophobicity in the domains of MAT1-1-1 proteins that were
624 encoded by the genome assemblies JAAVMX010000001 and LKHE01001116 [Li *et al*. 2016a; Shu *et al*.
625 2020], illustrated by topology and waveform changes in Panel B for the hydropathy plot. These changes
626 resulted in the altered secondary structures surrounding the mutation sites in the domains of the MAT1-1-1
627 proteins, as illustrated in Panels C–F for the ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and
628 coils. The MATα_HMGbox domains of the MAT1-1-1 proteins were clustered into Branch c in the Bayesian

clustering tree (*cf*. Figure 1, Table S9).

Panel A: Amino acid sequence alignment

```
AGW27560        42        ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTTL 101
JAAVMX010000001 6,699,034 ----------------------------------------M-------------------- 6,699,213
LKHE01001116    4786      ----------------------------------------M-------------------- 4555

AGW27560        85        WHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQD 161
JAAVMX010000001 6,699,214 ------------------------------------------------------------ 6,699,393
LKHE01001116    4554      ------------------------------------------------------------ 4375

AGW27560        145       DDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSHS 221
JAAVMX010000001 6,699,214 ------------------------------------------------------------ 6,699,573
LKHE01001116    4374      ------------------------------------------------------------ 4195

AGW27560        205       GIMAPRAANCTPP                                                234
JAAVMX010000001 6,699,574 -------------                                                6,699,664
LKHE01001116    4194      -------------                                                4156
```

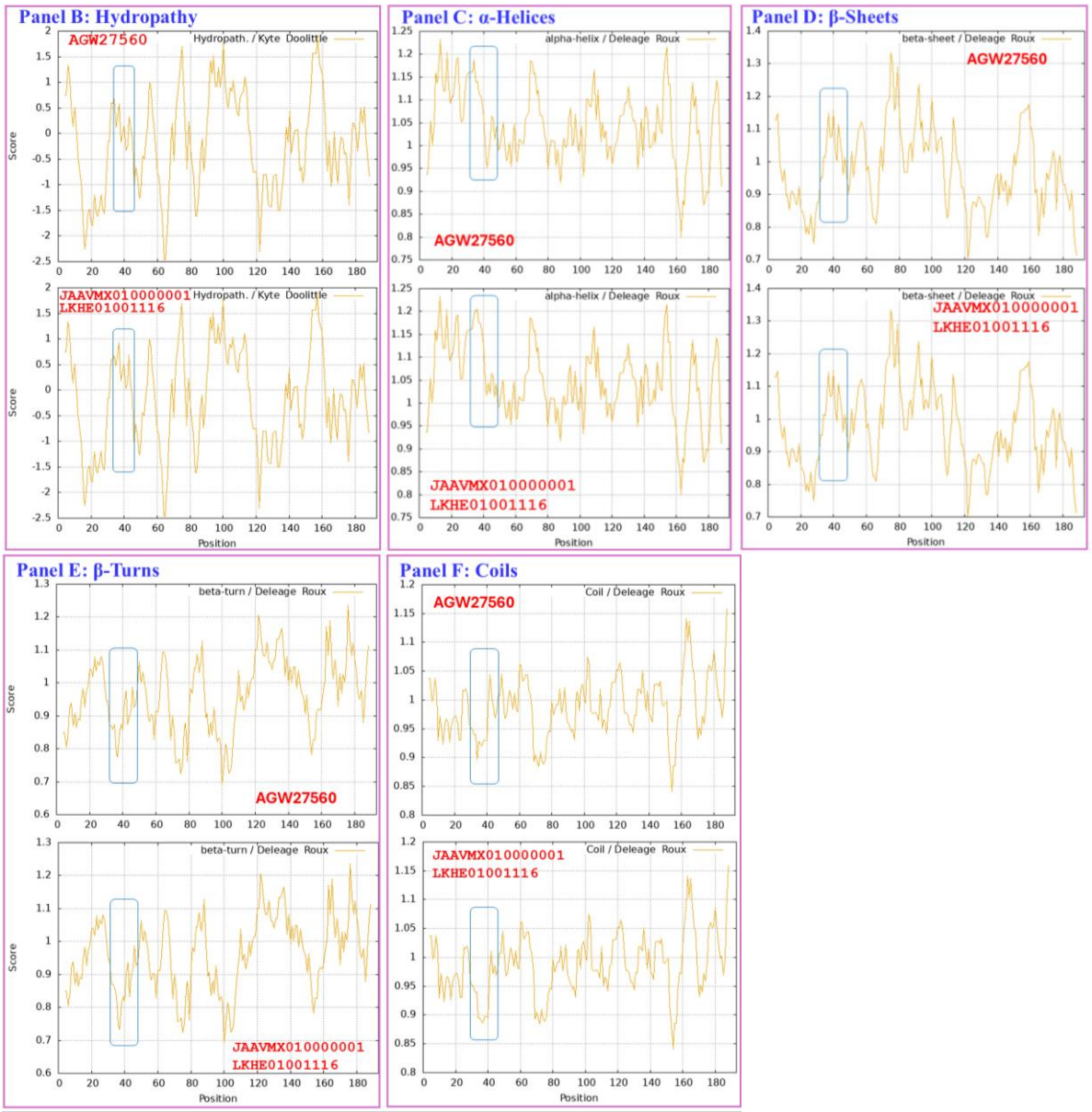


Figure 11. Correlation of the changes in the primary and secondary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 derived from the *H. sinensis* strain CS68-2-1229 and the truncated MAT1-1-1 proteins encoded by the genome assemblies JAAVMX010000001 and LKHE01001116 of the *H. sinensis* strains IOZ07 and 1229, respectively. Panel A shows alignment of amino acid sequences of the MATα_HMGbox domain sequences of the MAT1-1-1 proteins with amino acid substitutions in red; the hyphens

638 indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B)
639 and 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) with open rectangles in blue outlining the
640 changes in topology and waveform of the plots.

641 No changes were found in the amino acid sequences, hydrophobicity, and the secondary structures (α-
642 helices, β-sheets, β-turns, and coils) in the MATα_HMGbox domains of MAT1-1-1 proteins that were
643 encoded by the genome assembly ANOV01017390 (794←1229 & 1280←1369, removed the intron section)
644 of the *H. sinensis* strain Co18 and the metatranscriptome assembly OSIN7648 (1→1065) of the *C. sinensis*
645 insect-fungi complex collected from Deqin, Yunnan of China, comparing with the reference MAT1-1-1
646 protein AGW27560 (*cf*. Figure S3, Table S9) [Kyte & Doolittle 1982; Zhang *et al*. 2013; Hu *et al*. 2013;
647 Liu *et al*. 2015]. The MATα_HMGbox domains of the genome- and metatranscriptome-encoded MAT1-1-
648 1 proteins were clustered into Branch a1 in the Bayesian clustering tree (*cf*. Figure 1, Table S9).

649 The MATα_HMGbox domain of the MAT1-1-1 protein encoded by the metatranscriptome assembly
650 GAGW01008880 (714←1127) of the *C. sinensis* insect-fungi complex is largely truncated at the N-
651 terminus shown in Figure S3 and in Panel A of Figure S5 [Xiang *et al*. 2014]. The remaining portion of the
652 MATα_HMGbox domain of the metatranscriptome-encoded MAT1-1-1 protein was 100% identical to the
653 query protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-
654 2-1229 [Zhang *et al*. 2013] and showed no apparent changes in hydrophobicity and secondary structures,
655 as illustrated by the identical topology and waveform in Panels B–F of Figure S5 of the ExPASy ProtScale
656 plots for hydropathy, α-helices, β-sheets, β-turns, and coils. Although the MATα_HMGbox domain of the
657 metatranscriptome-encoded protein was clustered into Branch a1 in the Bayesian clustering tree together
658 with the 6 reference authentic MAT1-1-1 proteins (*cf*. Figure 1, Table S9), the remaining portion of the
659 truncated domain missing 50 and 48 amino acids that encoded by the exons I and II of the gene, respectively,
660 may not be able to form the core stereostructure of the 3 α-helices to support the high affinity of DNA
661 binding and fully functionality for gene transcription regulation relating to the sexual reproduction of *O.*
662 *sinensis* [Baxevanis *et al*. 1995; Thapar 2015].

663 The genome assemblies LWBQ00000000 and NGJJ00000000 and the transcriptome assembly
664 GCQL00000000 of the *H. sinensis* strains ZJB12195, CC1406-20395, and L0106, respectively, do not
665 contain the genes or transcripts that encode the MAT1-1-1 proteins [Liu *et al*. 2015, 2020; Jin *et al*. 2020].

666 III-8 Heteromorphic stereostructures of the HMG-box_ROX1-like domains of the
667 MAT1-2-1 proteins derived from wild-type *C. sinensis* isolates

668 This section analyzes the correlation between the hydrophobicity and structures of the HMG-
669 box_ROX1-like domains of MAT1-2-1 proteins derived from wild-type *C. sinensis* isolates under 12

AlphaFold codes (Table S10), comparing 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) [Zhang *et al*. 2009]. The HMG-box_ROX1-like domain sequences (amino acids 127→197) were extended both upstream and downstream with external 9 amino acids because the ExPASy ProtScale plotting for the hydropathy, α-helices, β-sheets, β-turns, and coils was set up with a window size of 9 amino acids (*cf*. Section II-5 of Materials and Methods).

Figure 12 examined the hydrophobicity and structures of the HMG-box_ROX1-like domain of MAT1-2-1 protein AIV43040 (under the AlphaFold code A0A0A0RCF5) derived from the *C. sinensis* isolate XZ12_16 (*cf*. Table S2) [Zhang *et al*. 2014], comparing with the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [Zhang *et al*. 2009]. Panel A shows 3 amino acid substitutions: Y-to-H (aromatic tyrosine to basic histidine; hydropathy index -1.3 to -3.2 [Kyte & Doolittle 1982]), M-to-I (methionine with polar neutral side chains to aliphatic isoleucine; hydropathy index 1.9 to 4.5), and Q-to-R (glutamine with polar neutral side chains to basic arginine; hydropathy index -3.5 to -4.5), which caused decreases and an increase in hydrophobicity, respectively, as illustrated by topology and waveform changes in Panel B for the hydropathy plot. These changes resulted in the altered secondary structures surrounding the mutation sites in the HMG-box_ROX1-like domain of MAT1-2-1 protein AIV43040 [Zhang *et al*. 2014], as illustrated by topology and waveform changes in Panels C–F for the ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils. As shown in Panels G–H for 3D structures, the replaced residues are located in each of the 3 core α-helices that altered the stability of hydrophobic core of the HMG-box_ROX1-like domain [Baxevanis *et al*. 1995; Thapar 2015] and the stereostructure of the domain of MAT1-2-1 protein AIV43040 under the AlphaFold code A0A0A0RCF5. The mutant HMG-box_ROX1-like domain was clustered into Branch b2β with the longest clustering distance in the Bayesian clustering tree (*cf*. Figure 2, Table S10).

Panel A: Amino acid sequence alignment

```
AEH27625 118 APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK 177
AIV43040 118 --------------------------H----------------------I---------- 177

AEH27625 178 DMSQQVKQALLEKHPDYQYKPRRPCERRR                                206
AIV43040 178 ----R------------------------                                206
```

Panel B: Hydropathy
Hydropath. / Kyte Doolittle
AEH27625
AIV43040
Score
Position
Panel C: α-Helices
alpha-helix / Deleage Roux
AEH27625
AIV43040
Position
Panel D: β-Sheets
beta-sheet / Deleage Roux
AEH27625
AIV43040
Position
Panel E: β-Turns
beta-turn / Deleage Roux
AEH27625
AIV43040
Score
Position
Panel F: Coils
Coil / Deleage Roux
AEH27625
AIV43040
Position

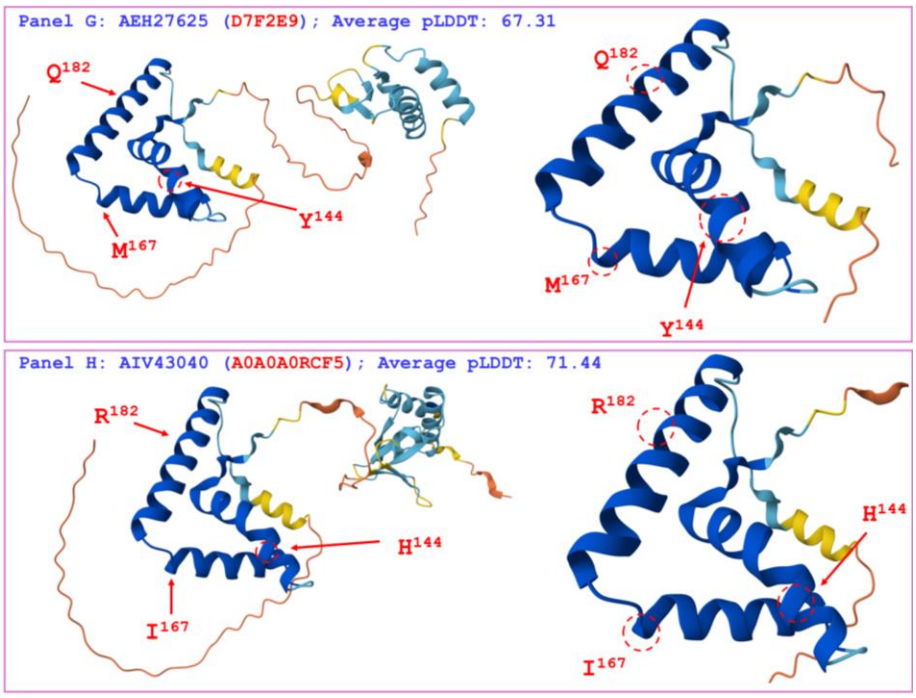


Figure 12. Correlation of the changes in the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins: the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 and the mutant MAT1-2-1 protein AIV43040 (under the AlphaFold code A0A0A0RCF5) derived from the wild-type *C. sinensis* isolate XZ12_16. Panel A shows alignment of amino acid sequences of the HMG-box_ROX1-like domain sequences of the MAT1-2-1 proteins with amino acid substitutions in red; the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) with open rectangles in blue outlining the changes in topology and waveform of the plots. Each of 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. Model confidence for the AlphaFold-predicted 3D structures: very high (pLDDT>90); high (90>pLDDT>70); low (70>pLDDT>50); very low (pLDDT<50).

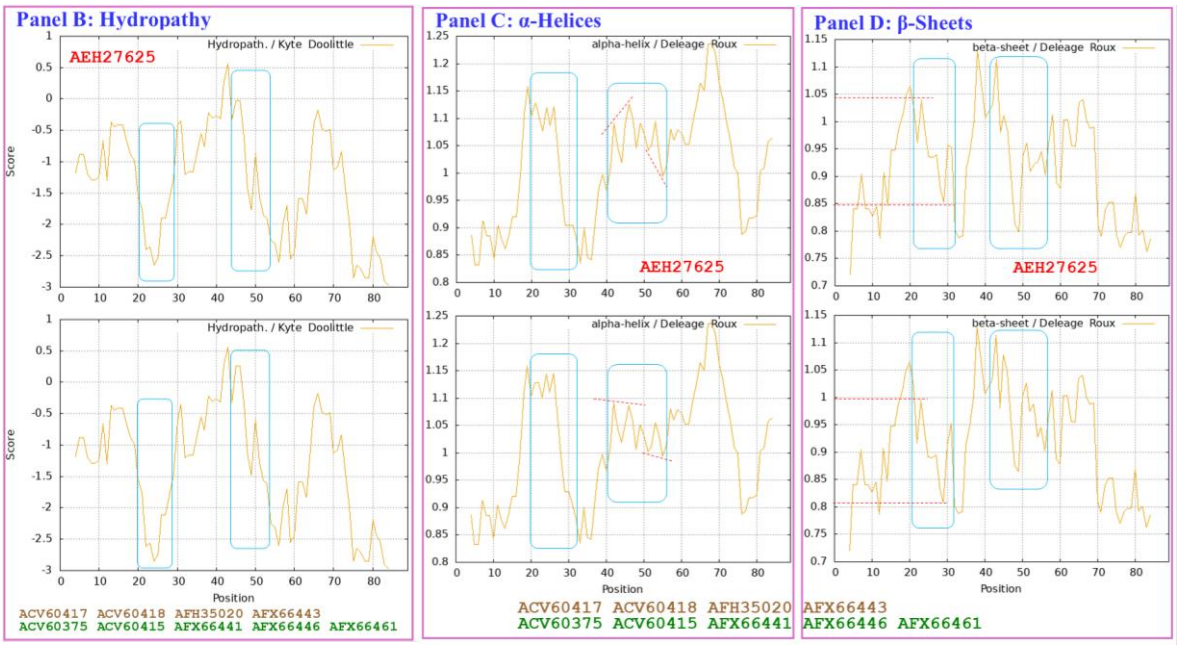
Figure 13 examined the hydrophobicity and structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins under the AlphaFold code D7F2J7 (ACV60417, ACV60418, AFH35020, and AFX66443 derived from the wild-type *C. sinensis* isolates XZ-LZ07-H1, XZ-LZ07-H2, XZ06-124, and XZ05_8) and of the proteins under the AlphaFold code D7F2F5 (ACV60375, ACV60415, AFX66441, AFX66446, and AFX66461 derived from the wild-type *C. sinensis* isolates XZ-SN-44, XZ-LZ05-6, XZ05_2, XZ06_260, and XZ09_80) (*cf*. Table S2) (*cf*. Table S2) [Zhang *et al*. 2009, 2012, 2014], respectively, comparing with the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [Zhang *et al*. 2009]. Panel A of Figure 13 shows substitutions of Y-to-H (aromatic tyrosine to basic histidine; hydropathy indexes of -1.3 to -3.2 [Kyte & Doolittle 1982]) and M-to-I (methionine with polar neutral side chains to aliphatic isoleucine; hydropathy indexes of 1.9 to of 4.5), which caused changes in hydrophobicity, respectively, as illustrated by topology and waveform changes in Panel B for the hydropathy plot. These changes resulted in the altered secondary structures surrounding the mutation sites in the HMG-box_ROX1-like domain of mutant MAT1-2-1 proteins ACV60417, ACV60418, AFH35020,

and AFX66443 under the AlphaFold code D7F2J7 and of the proteins ACV60375, ACV60415, AFX66441, AFX66446, and AFX66461 under the AlphaFold code D7F2F5 [Zhang *et al*. 2009, 2012, 2014], as illustrated by topology and waveform changes in Panels C–F for the ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils.

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
```



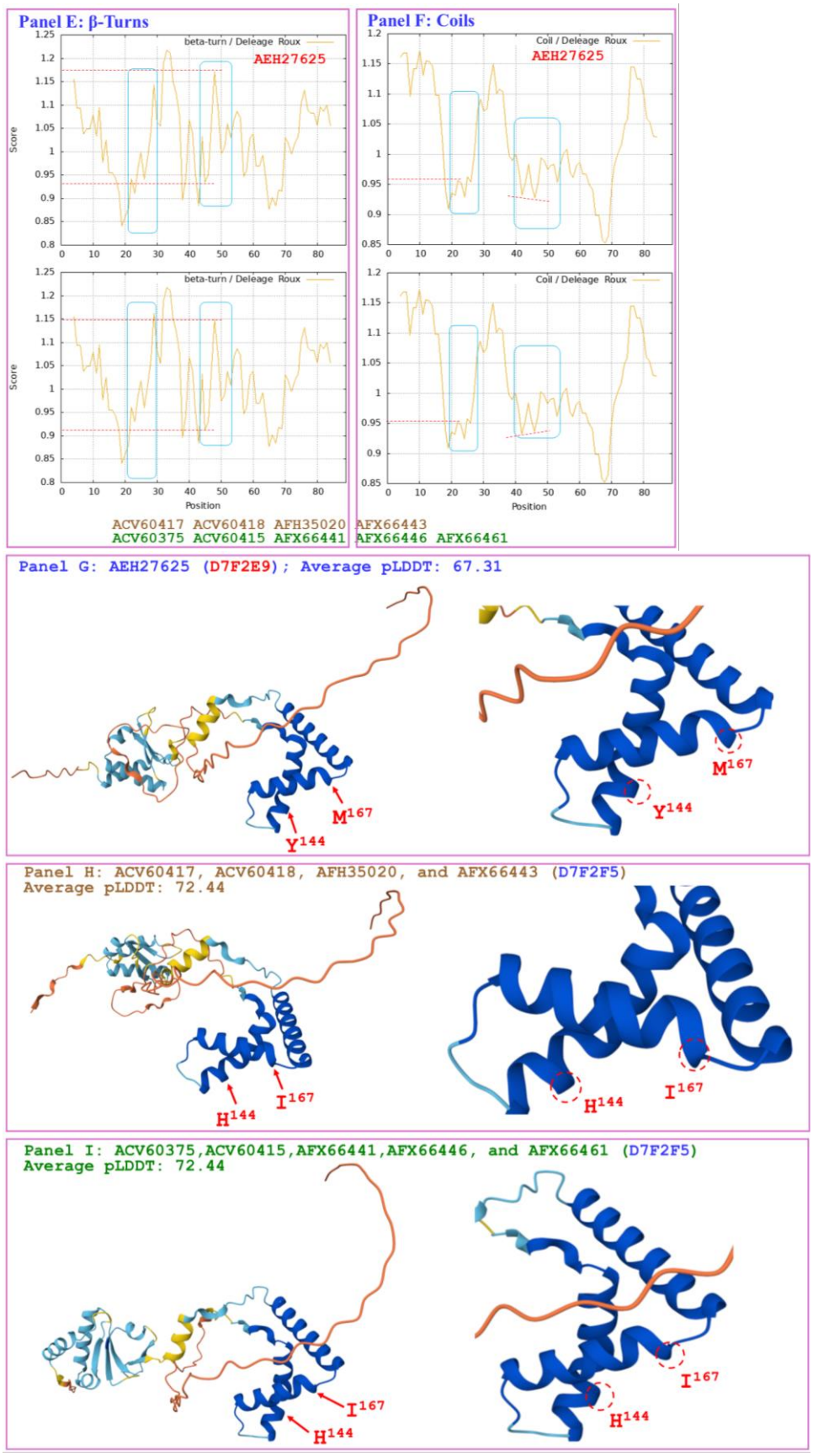


Figure 13. Correlation of the changes in the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins: the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from

the *H. sinensis* strain CS2, the mutant MAT1-2-1 proteins under the AlphaFold code D7F2J7 (ACV60417, ACV60418, AFH35020, and AFX66443 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 derived from the wild-type *C. sinensis* isolate XZ-SN-44, XZ-LZ05-6, XZ05_2, XZ06_260, and XZ09_80). Panel A shows alignment of amino acid sequences of the HMG-box_ROX1-like domain sequences of the MAT1-2-1 proteins with amino acid substitutions in red; the sequences in brown and green represent the MAT1-2-1 proteins under the AlphaFold codes D7F2J7 and D7F2F5, respectively; the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) with open rectangles in blue outlining the changes in topology and waveform of the plots. Each of 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. Model confidence for the AlphaFold-predicted 3D structures: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); ■ very low (pLDDT<50).

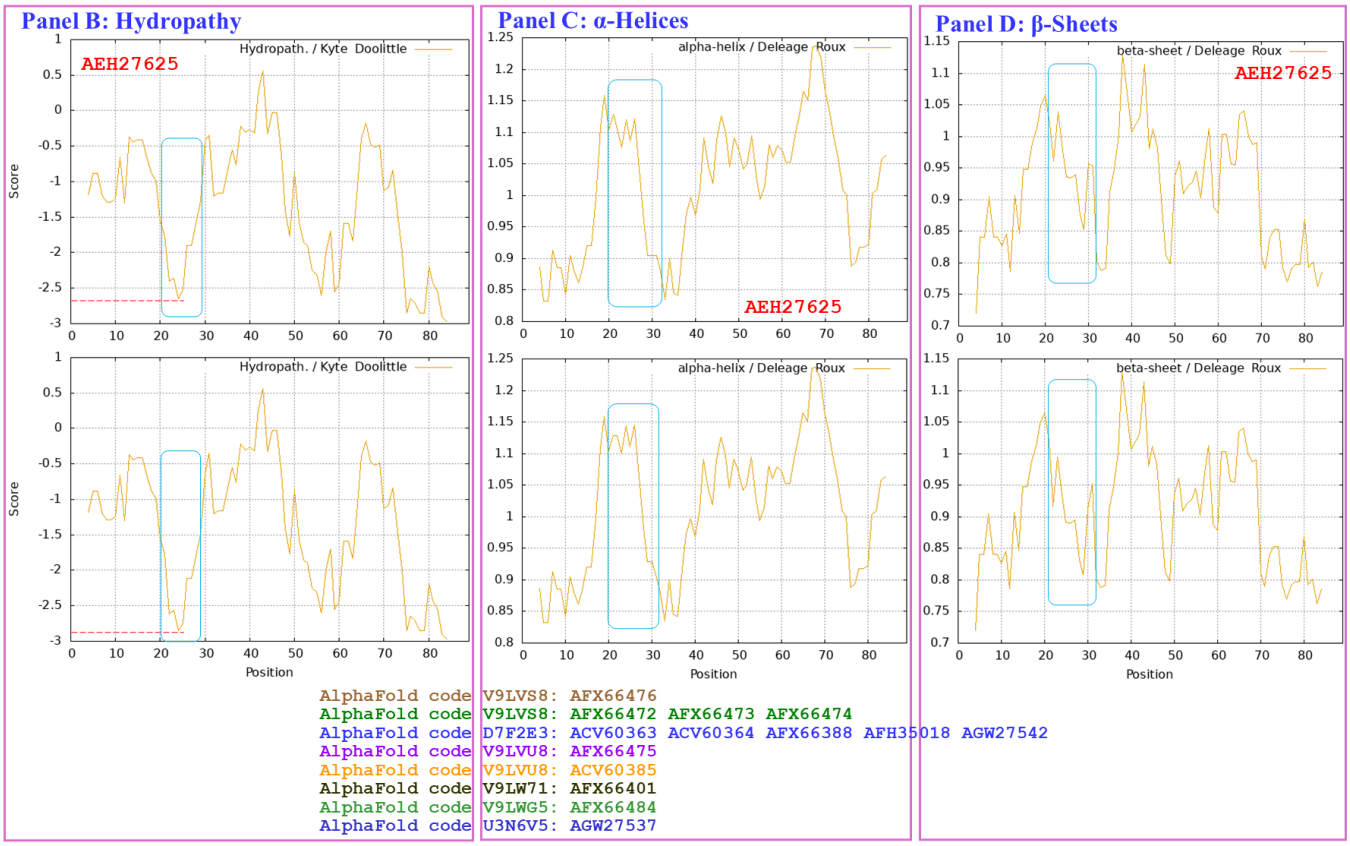
As shown in Panels G–I of Figure 13 for 3D structures, the replaced residues are located in two of the 3 core α-helices that differentially altered the stability of core structure of the HMG-box_ROX1-like domain [Baxevanis *et al*. 1995; Thapar 2015] and the tertiary structures of the domains of MAT1-2-1 proteins under the AlphaFold 3D structural codes D7F2J7 and D7F2F5. The variable domain sequences of the mutant proteins ACV60417, ACV60418, AFH35020, AFX66443, ACV60375, ACV60415, AFX66441, AFX66446, and AFX66461 were clustered into Branch b2α in the Bayesian clustering tree (*cf*. Figure 2, Table S10).

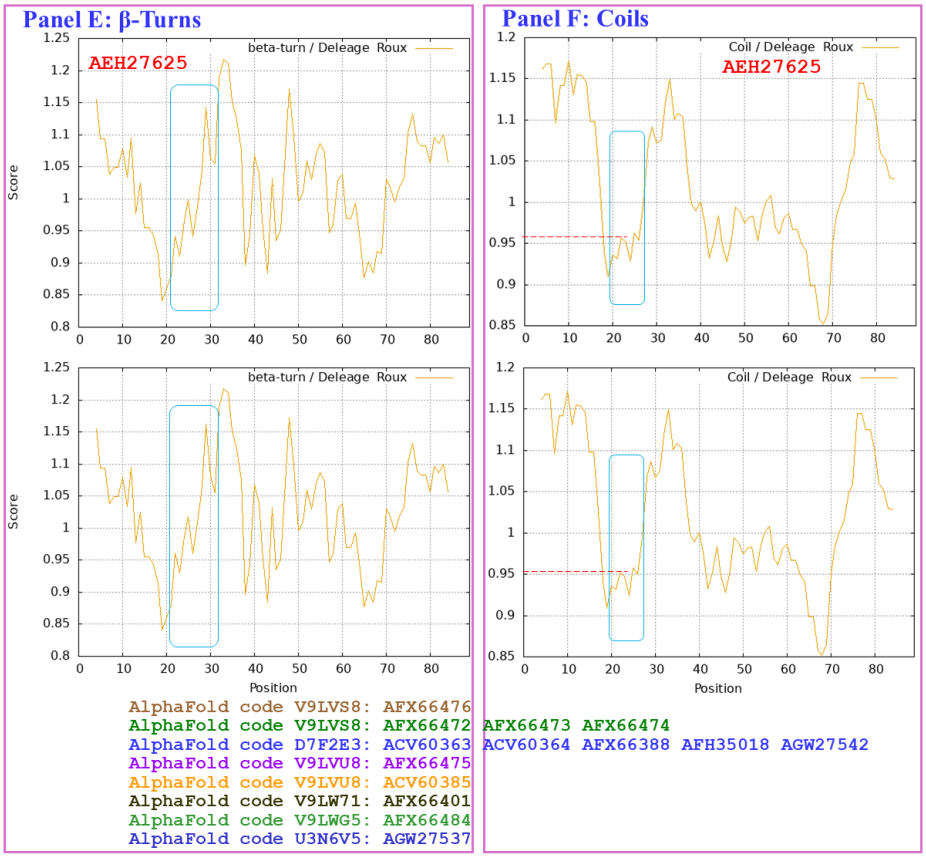
Figure 14 examined the hydrophobicity and structures of the HMG-box_ROX1-like domain of MAT1-2-1 protein under the AlphaFold codes V9LWC9 (AFX66476), V9LVS8 (AFX66472, AFX66473, and AFX66474), D7F2E3 (ACV60363, ACV60364, AFX66388, AFH35018, and AGW27542), V9LVU8 (AFX66475), D7F2G5 (ACV60385), V9LW71 (AFX66401), 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) [Zhang *et al*. 2014], comparing with the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [Zhang *et al*. 2009]. Panel A shows Y-to-H substitutions (aromatic tyrosine to basic histidine; hydropathy indexes of -1.3 to -3.2 [Kyte & Doolittle 1982]) in the HMG-box_ROX1-like domains of all the MAT1-2-1 proteins aforementioned, which caused reduced hydrophobicity, as illustrated by topology and waveform changes in Panel B for the hydropathy plots. These changes resulted in the altered secondary structures surrounding the mutation sites in the HMG-box_ROX1-like domains of MAT1-2-1 proteins, as illustrated by topology and waveform changes in Panels C–F for the ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils.

Panel A: Amino acid sequence alignment

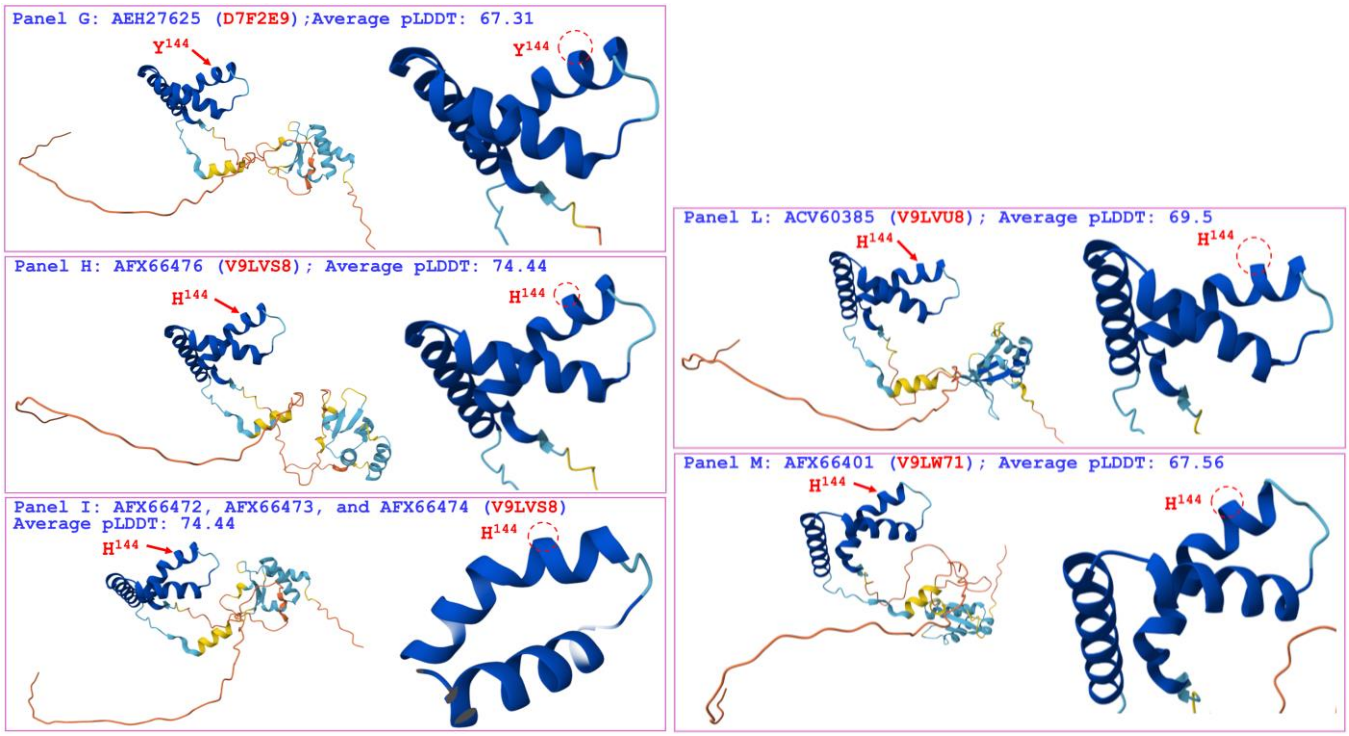
```
AEH27625 118 APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK 177
AFX66476 118 ---------------------------H---------------------------------- 177
AFX66472 118 ---------------------------H---------------------------------- 177
AFX66473 118 ---------------------------H----------------------------------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

AEH27625 178 DMSQQVKQALLEKHPDYQYKPRRPCERRR                                 206
AFX66476 178 ------------------------------                                 206
AFX66472 178 ------------------------------                                 206
AFX66473 178 ------------------------------                                 206
AFX66474 178 ------------------------------                                 206
ACV60363 178 ------------------------------                                 206
ACV60364 178 ------------------------------                                 206
AFX66388 178 ------------------------------                                 206
AFH35018 178 ------------------------------                                 206
AGW27542 178 ------------------------------                                 206
AFX66475 178 ------------------------------                                 206
ACV60385 178 ------------------------------                                 206
AFX66401 178 ------------------------------                                 206
AFX66484 178 ------------------------------                                 206
AGW27537 178 ------------------------------                                 206
```




Panel E: β-Turns
beta-turn / Deleage Roux
AEH27625
Score
Position
Panel F: Coils
Coil / Deleage Roux
AEH27625
AlphaFold code V9LVS8: AFX66476
AlphaFold code V9LVS8: AFX66472 AFX66473 AFX66474
AlphaFold code D7F2E3: ACV60363 ACV60364 AFX66388 AFH35018 AGW27542
AlphaFold code V9LVU8: AFX66475
AlphaFold code V9LVU8: ACV60385
AlphaFold code V9LW71: AFX66401
AlphaFold code V9LWG5: AFX66484
AlphaFold code U3N6V5: AGW27537

767


Panel G: AEH27625 (D7F2E9);Average pLDDT: 67.31
Y144
Panel H: AFX66476 (V9LVS8); Average pLDDT: 74.44
H144
Panel I: AFX66472, AFX66473, and AFX66474 (V9LVS8)
Average pLDDT: 74.44
Panel L: ACV60385 (V9LVU8); Average pLDDT: 69.5
Panel M: AFX66401 (V9LW71); Average pLDDT: 67.56

768

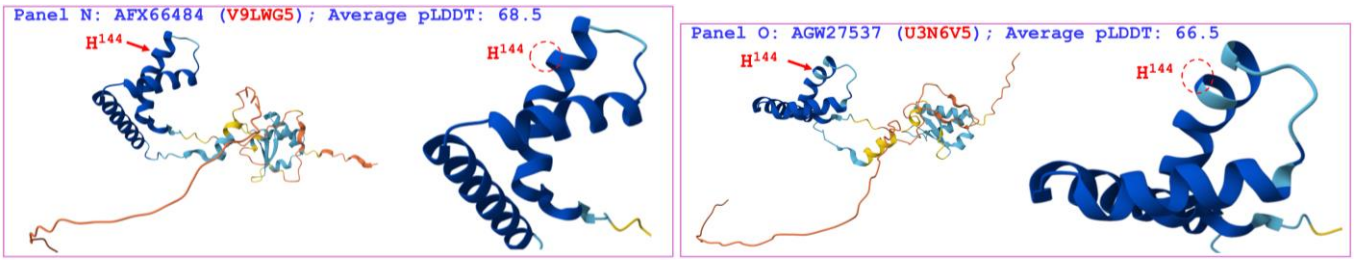


Figure 14. Correlation of the changes in the hydrophobicity, and primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins: the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 and the mutant MAT1-2-1 proteins under the AlphaFold codes **V9LWC9** (AFX66476), **V9LVS8** (AFX66472, AFX66473, and AFX66474), **D7F2E3** (ACV60363, ACV60364, AFX66388, AFH35018, and AGW27542), **V9LVU8** (AFX66475), **D7F2G5** (ACV60385), **V9LW71** (AFX66401), **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. Panel A shows alignment of amino acid sequences of the HMG-box_ROX1-like domain sequences of the MAT1-2-1 proteins with amino acid substitutions in red; the sequences 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 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) with open rectangles in blue outlining the changes in topology and waveform of the plots. Each of Panels G–O shows the 3D structures of the full-length proteins on the left and the locally amplified structures at the mutation site of substitution on the right. Model confidence for the AlphaFold-predicted 3D structures: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); ■ very low (pLDDT<50).

As shown in Panels G–O of Figure 14 for tertiary structures, the replaced residues with reduced hydrophobicity are located within the core α-helices that destabilized and altered the hydrophobic stereostructures of the HMG-box_ROX1-like domains [Baxevanis *et al*. 1995; Thapar 2015]. Although Panels A of Figure 14 shows the same Y-to-H substitutions in the HMG-box_ROX1-like domains of the mutant MAT1-2-1 proteins derived from various wild-type *C. sinensis* isolates, which caused the same changes in the hydrophobicity and second structures shown in Panels B–F for the ExPASy ProtScale plots, additional amino acid substitutions outside of the HMG-box_ROX1-like domains (*cf*. Figure S4) might play synergistic roles through long-range spatial interactions with the substituted residues within the domains on the alterations of the tertiary structures (Panels H–O of Figure 14; Table S10) under various AlphaFold codes. The domain sequences of the MAT1-2-1 proteins under variable AlphaFold codes were clustered into Branch b1α in the Bayesian clustering tree (*cf*. Figure 2, Tables S2 and S10).

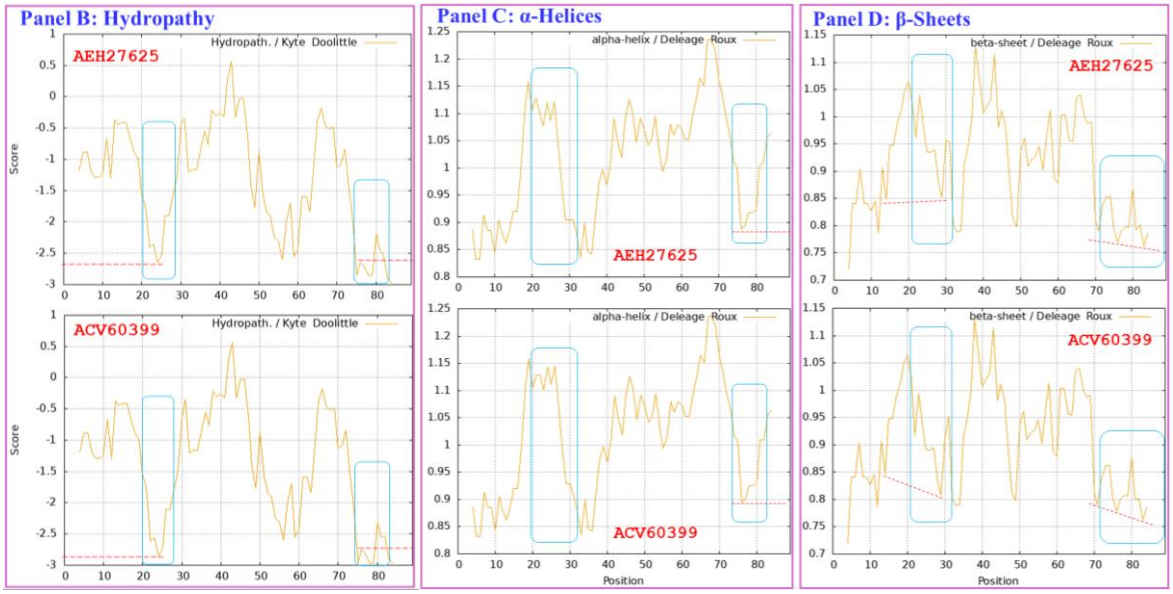
Figure 15 examined the hydrophobicity and structures of the domain of MAT1-2-1 protein ACV60399 (under the AlphaFold code D7F2H9) derived from the wild-type *C. sinensis* isolates SC-3 (*cf*. Table S2) [Zhang *et al*. 2009], comparing with the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [Zhang *et al*. 2009]. Panel A shows substitutions of Y-to-H (aromatic tyrosine to basic histidine; hydropathy indexes of -1.3 to -3.2 [Kyte & Doolittle 1982]) and Q-to-R (glutamine with polar neutral side chains to basic arginine; hydropathy indexes of -3.5 to -4.5), both of

which caused reduced hydrophobicity, as illustrated by topology and waveform changes in Panel B for the hydropathy plot. These changes resulted in the altered secondary structures surrounding the mutation sites in the HMG-box_ROX1-like domain of mutant MAT1-2-1 protein ACV60399 [Zhang *et al*. 2009], as illustrated by topology and waveform changes in Panels C–F for ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils. As shown in Panels G–H for 3D structures, the 2 replaced residues with reduced hydrophobicity are located inside as well as outside of the hydrophobic core structure of the 3 α-helices that destabilized and altered the core stereostructure of the domain [Baxevanis *et al*. 1995; Thapar 2015]. The HMG-box_ROX1-like domain of the mutant protein ACV60399 under the AlphaFold code D7F2H9 was clustered into Branch b1β in the Bayesian clustering tree (*cf*. Figure 2, Table S10).

Panel A: Amino acid sequence alignment

```
AEH27625 118 APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK 177
ACV60399 118 --------------------------H--------------------------------- 177

AEH27625 178 DMSQQVKQALLEKHPDYQYKPRRPCERRR                                206
ACV60399 178 -----------------R-----------                                206
```



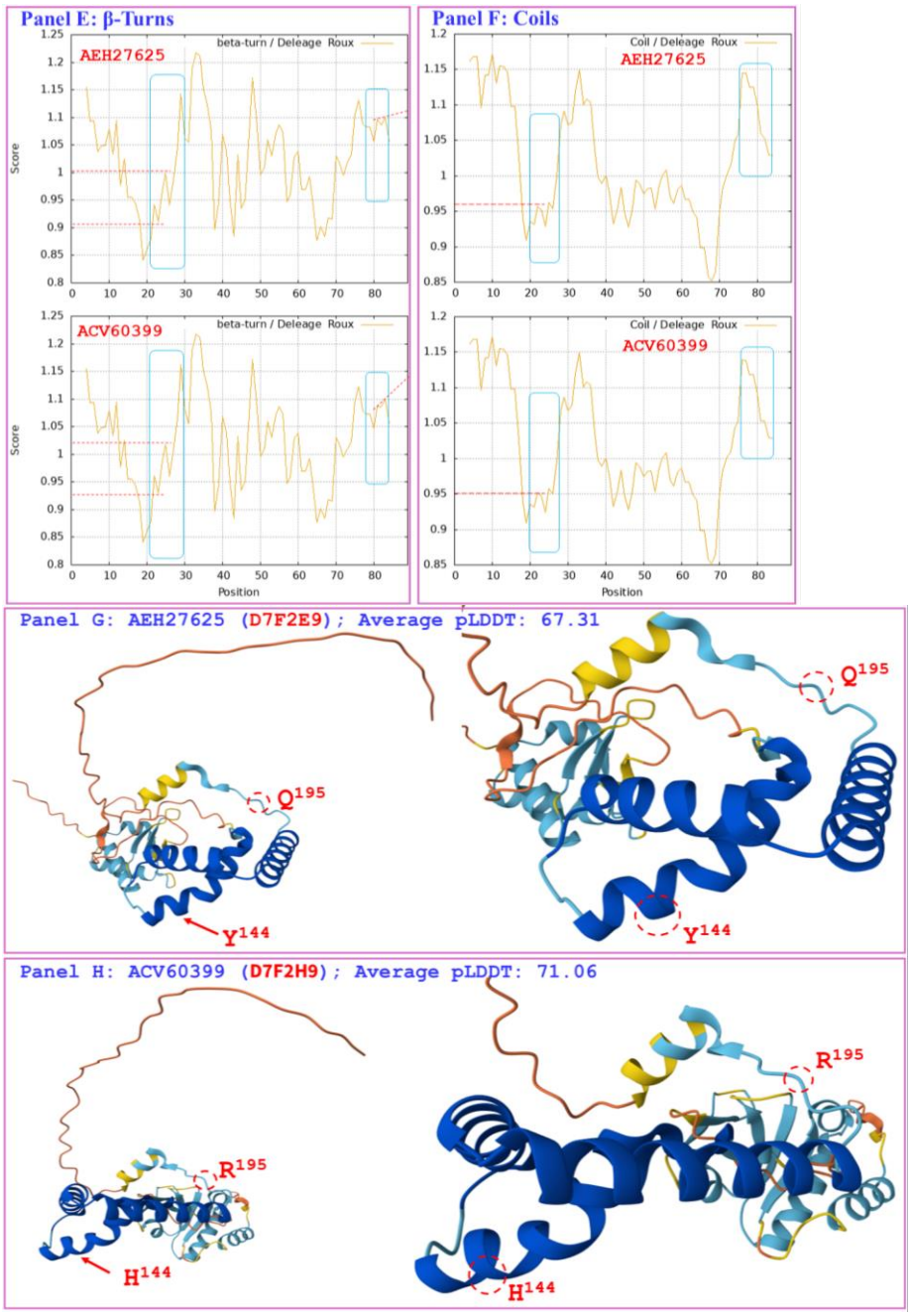


Figure 15. Correlation of the changes in the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins: the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 and the mutant MAT1-2-1 protein ACV60399 (under the AlphaFold code D7F2H9) derived from the wild-type *C. sinensis* isolate SC-3. Panel A shows alignment of amino acid sequences of the HMG-box_ROX1-like domain sequences of the MAT1-2-1 proteins with amino acid substitutions in red; the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show below the changes in hydrophobicity (Panel B) and 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) with open rectangles in blue outlining the changes in topology and waveform of the plots. Each of 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. Model confidence for the AlphaFold-predicted 3D structures: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); ■ very low (pLDDT<50).

Aggregating the data demonstrated in Section III-7, the 1–3 amino acid substitutions detected in the

HMG-box_ROX1-like domains (Panels A of Figures 12–15) subsequently caused the changes in topologies and waveforms of ExPASy ProtScale plots shown in Panels B–F of Figures 12–15, indicating altered hydrophobicity and secondary structures (α-helices, β-sheets, β-turns, and coils) of the domains of the mutant MAT1-2-1 proteins. These changes in the domains then resulted in the altered stereostructures of the domains of the mutant proteins under different AlphaFold codes as illustrated in Panels H–O of Figures 12–15 and summarized in Table S10.

The altered stereostructures of the domains ultimately change the DNA binding specificities and affinities and the functionalities of gene transcription regulations, which play key roles in the sexual reproduction of *O. sinensis*. As summarized in Table S10, the HMG-box_ROX1-like domains with 1-3 amino acid substitutions were clustered into various clades in the Bayesian clustering tree (*cf*. Figure 2).

III-9 Diverse stereostructures of the HMG-box_ROX1-like domains of C-terminally truncated MAT1-2-1 proteins derived from *O. sinensis* strains

In addition to the 69 full-length proteins, 5 other MAT1-2-1 proteins AGW27543, AGW27554, AGW27548, AGW27555, and AGW27552 derived from various *O. sinensis* strains are truncated at the C-termini of the HMG-box_ROX1-like domains. Although the same pair of primers (forward *Mat1-2F* and revere compliment *Mat1-2R* shown in Figure S2) were used to amplify MAT1-2-1 cDNA [Bushley *et al*. 2013], the truncated MAT1-2-1 proteins appeared to have variable lengths (Figure 16) with 69–74% query coverage compared with the query sequence AEH27625 of the full-length proteins under the AlphaFold UniProt code D7F2E9 [Zhang *et al*. 2013; Li *et al*. 2025].

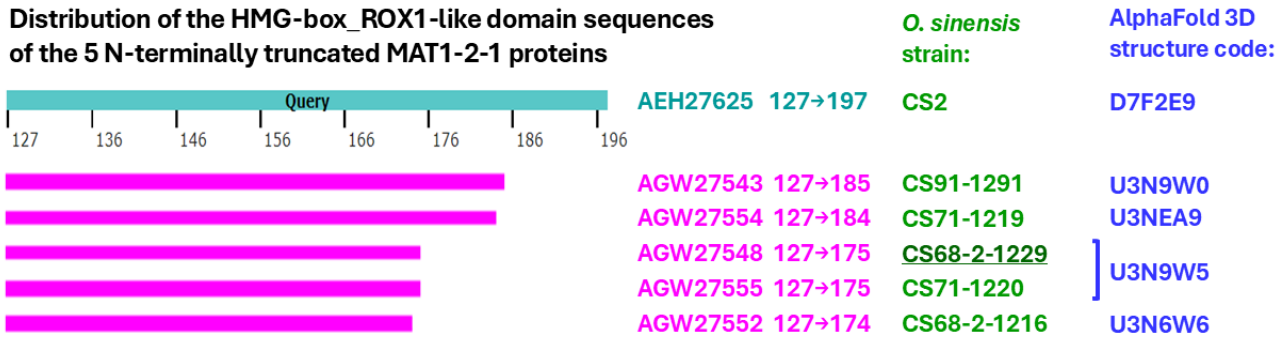


Figure 16. Sequence distribution of the C-terminally truncated HMG-box_ROX1-like domains of 5 subject MAT1-2-1 proteins derived from the listed *O. sinensis* strains aligned with the full-length HMG-box_ROX1-like domain (127→197) of the query MAT1-2-1 protein AEH27625 of *H. sinensis* strain CS2.

The C-terminal truncations of the HMG-box_ROX1-like domains of the 5 truncated proteins resulted in 97.9–98.3% similarities to the domain sequence of the query protein AEH27625. In addition to the truncation mutations, an additional Y-to-H substitution (aromatic tyrosine to basic histidine; hydropathy

854 indexes of -1.3 to -3.2 [Kyte & Doolittle 1982]) occurred in the HMG-box_ROX1-like domains of the 5
855 truncated MAT1-2-1 proteins. The truncation and point mutations caused reduced hydrophobicity and
856 destabilized of the hydrophobic core structures of the domains, resulting in diverse stereostructures under
857 4 AlphaFold 3D structure codes (Figure 16).

858 Three (CS91-1291, CS71-1219, and CS68-2-1229) of the 5 *O. sinensis* strains were genetically
859 homogenous shown in blue in Table S4 [Li *et al*. 2013], belonging to GC-biased Genotype #1 *H. sinensis*.
860 However, the strain CS71-1220 shown in red in Table S4 was genetically heterogeneous, containing both
861 GC-biased Genotype #1 and AT-biased Genotype #5 of *O. sinensis*. The GC- and AT-biased genotypes are
862 genome independent, belonging to different fungi [Kinjo & Zang 2001; Stensrud *et al*. 2007; Xiao *et al*.
863 2009; Li *et al*. 2013, 2020b, 2022, 2023d, 2023e; Yao & Zhu 2016; Zhu & Li 2017]. The remaining strain
864 CS68-5-1216 shown in black in Table S4 lacked ITS molecular genotyping evidence in the GenBank
865 database. The 5 *O. sinensis* strains were isolated either from the fresh caterpillar bodies or the cultures of
866 mono-ascospores (marked as TS or SS, respectively, in Table S4) of the *C. sinensis* insect-fungal complexes
867 [Li *et al*. 2013].

868 III-10 Diverse primary and secondary structures of the HMG-box_ROX1-like domain 869 of MAT1-2-1 proteins encoded by the genome and transcriptome assemblies of 870 *H. sinensis* and the metatranscriptome assembly of *C. sinensis* insect-fungi 871 complex

872 Figure 17 examined the hydrophobicity and structures of the HMG-box_ROX1-like domain of MAT1-
873 2-1 protein encoded by the genome assembly ANOV01000063 (9759→9851 & 9907→10,026, removed
874 the intron section) derived from *H. sinensis* strain Co18 [Hu *et al*. 2013], comparing with the reference
875 MAT1-2-1 protein AEH27625 derived from *H. sinensis* strain CS2 (under the AlphaFold code D7F2E9)
876 [Zhang *et al*. 2009]. Panel A shows a S-to-A substitution (serine with polar neutral side chains to aliphatic
877 alanine; hydropathy indexes of -0.8 to 1.8 [Kyte & Doolittle 1982]), which caused increased hydrophobicity
878 surrounding the mutation site in the domain of MAT1-2-1 protein encoded by the genome assembly
879 ANOV01000063 (9759→9851 & 9907→10,026) derived from *H. sinensis* strain Co18 [Hu *et al*. 2013],
880 comparing with the reference MAT1-2-1 protein AEH27625, as illustrated by topology and waveform
881 changes in Panel B for the hydropathy plot. These changes resulted in the altered secondary structures (α-
882 helices, β-sheets, β-turns, and coils) of the HMG-box_ROX1-like domain of the mutant MAT1-2-1 protein
883 encoded by the genome assembly ANOV01000063, as illustrated in Panels C–F for ExPASy ProtScale plots
884 for α-helices, β-sheets, β-turns, and coils [Baxevanis *et al*. 1995; Thapar 2015]. The sequence changes might
885 altered the tertiary structure of the domain of the genome-encoded protein and was clustered into Branch

a2 in the Bayesian clustering tree (*cf*. Figure 2; Table S11).

Panel A: Amino acid sequence alignment

```
AEH27625     118  APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK 177
ANOV01000063 9732 ----------------------------------------A------------------- 9967

AEH27625     178  DMSQQVKQALLEKHPDYQYKPRRPCERRR                                206
ANOV01000063 9968 ----------------------------                                  10,164
```

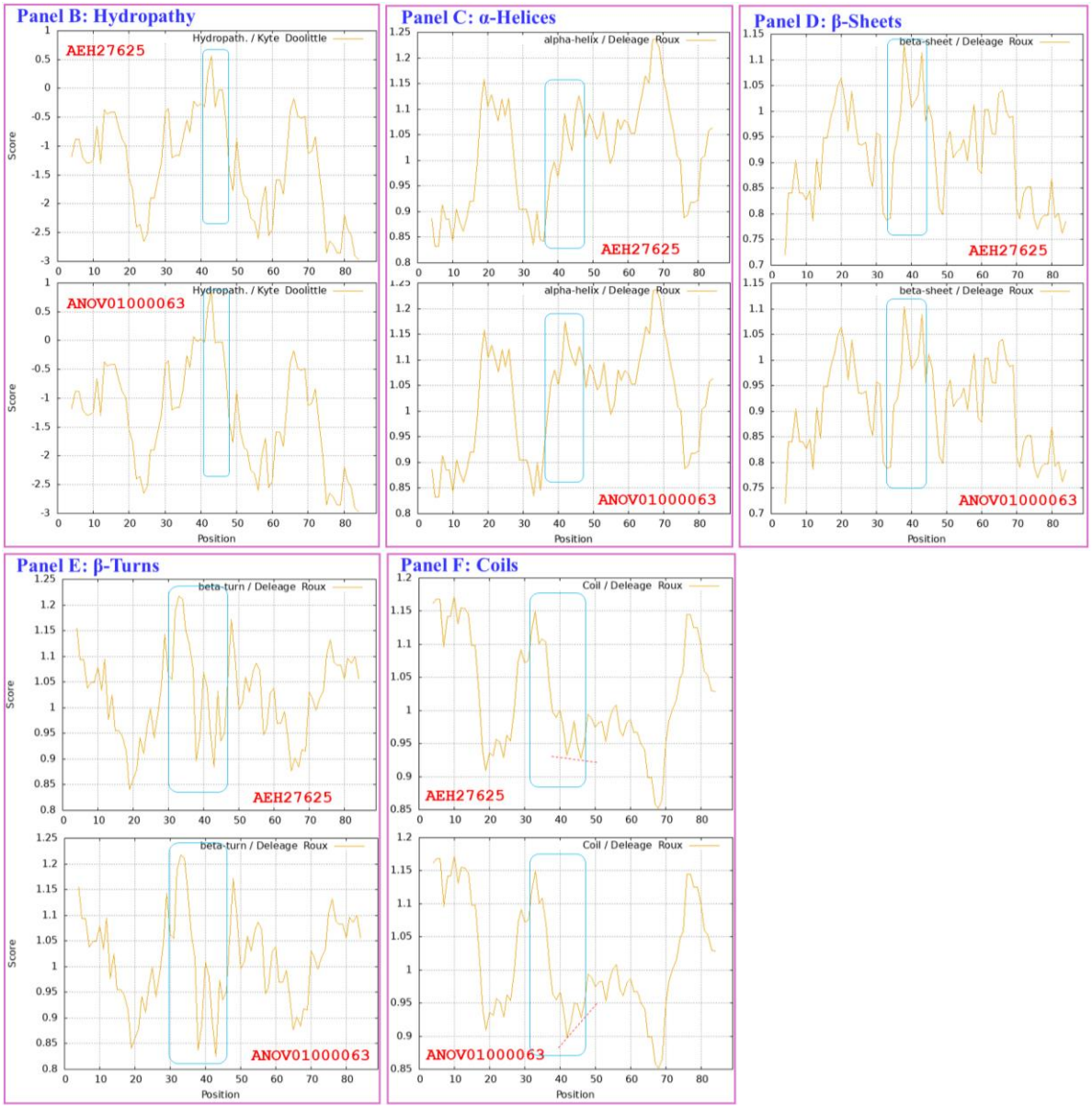


Figure 17. Correlation of the changes in the primary and secondary structures of the HMG-box_ROX1-like domains of the reference MAT1-2-1 protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 and the mutant MAT1-2-1 protein encoded by the genome assembly ANOV01000063 derived from the *H. sinensis* strain Co18. Panel A shows alignment of amino acid sequences of the HMG-box_ROX1-like domain sequences of the MAT1-2-1 proteins with an amino acid substitution in red; the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show below the changes in hydrophobicity (Panel B) and 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) with open rectangles in blue outlining the changes in topology and waveform of the plots.

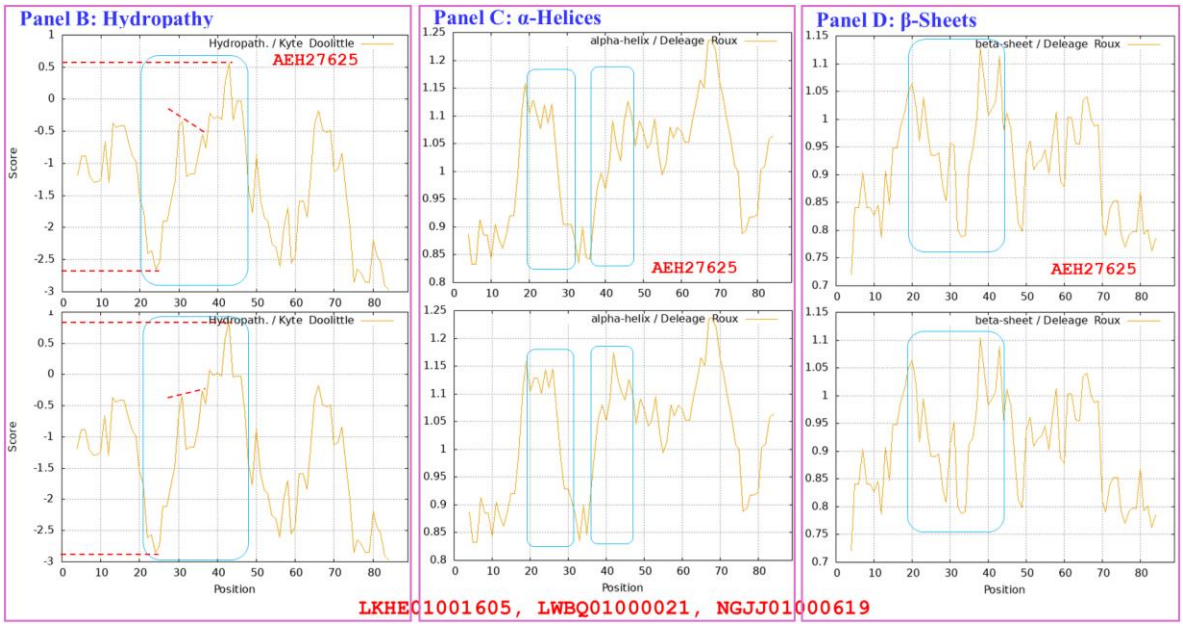
Figure 18 examines the hydrophobicity and structures of HMG-box_ROX1-like domains of MAT1-2-1 proteins encoded by the genome assemblies LKHE01001605 (14,016←14,135 & 14,191←14,283),

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, comparing with the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [Zhang *et al*. 2009]. Panel A shows Y-to-H (aromatic tyrosine to basic histidine; hydropathy indexes of -1.3 to -3.2 [Kyte & Doolittle 1982]) and S-to-A substitutions (serine with polar neutral side chains to aliphatic alanine; hydropathy indexes of -0.8 to 1.8) (Panel A), which caused changes in hydrophobicity, as illustrated in Panel B for hydropathy plot. The altered hydrophobicity altered the secondary structures surrounding mutation sites in the HMG-box_ROX1-like domain of mutant MAT1-2-1 proteins encoded by the genome assemblies LKHE01001605, LWBQ01000021, and NGJJ01000619, as illustrated in Panels C–F for ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils [Baxevanis *et al*. 1995; Thapar 2015], subsequently altering the tertiary structures of the core stereostructures of the domains. The HMG-box_ROX1-like domains of the genome-encoded mutant proteins were clustered into Branch b3 in the Bayesian clustering tree (*cf*. Figure 2, Table S11).

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
```



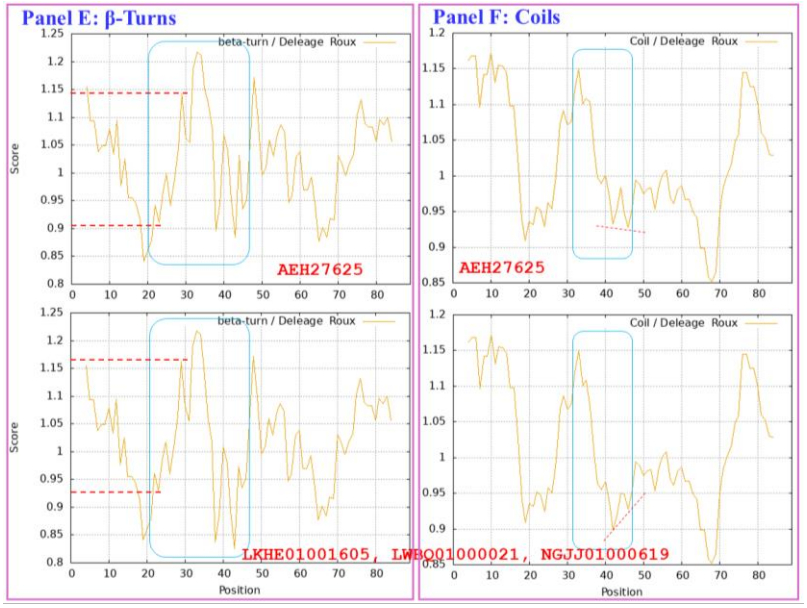


Figure 18. Correlation of the changes in the primary and secondary structures of the HMG-box_ROX1-like domains of the reference MAT1-2-1 protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 and the mutant MAT1-2-1 proteins encoded by the genome assemblies LKHE01001605, LWBQ01000021, and NGJJ01000619 derived from the *H. sinensis* strains 1229, ZJB12195, and CC1406-20395, respectively. Panel A shows alignment of amino acid sequences of the HMG-box_ROX1-like domain sequences of the MAT1-2-1 proteins with amino acid substitutions in red; the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show below the changes in hydrophobicity (Panel B) and 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) with open rectangles in blue outlining the changes in topology and waveform of the plots.

The genome assemblies JAAVMX000000000 of the *H. sinensis* strain IOZ07 does not contain the gene encoding the MAT1-2-1 protein [Shu *et al*. 2020].

Figure 19 examines the hydrophobicity and structures of HMG-box_ROX1-like domain of MAT1-2-1 protein that was encoded by the transcriptome assembly GCQL01020543 (553←765) of the *H. sinensis* strain L0106 [Liu *et al*. 2015]. Panel A shows a Y-to-H substitution of (aromatic tyrosine to basic histidine; hydropathy indexes of -1.3 to -3.2 [Kyte & Doolittle 1982]), which caused decreased hydrophobicity as illustrated by topology and waveform changes in Panel B for the hydropathy plot. These changes resulted in the altered secondary structures surrounding the mutation site in the HMG-box_ROX1-like domain of MAT1-2-1 protein encoded by the transcriptome assembly GCQL01020543 [Liu *et al*. 2015], as illustrated by topology and waveform changes in Panels C–F for ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils. Similar to the mutant MAT1-2-1 proteins under various AlphaFold 3D structural models shown in Figure 14, the changes in the hydrophobicity, and primary and secondary structures would subsequently alter the tertiary structures of the HMG-box_ROX1-like domain of the transcriptome-encoded mutant protein. The domain of the transcriptome-encoded protein was clustered into Branch b1α in the Bayesian clustering tree (*cf*. Figure 2, Table S11).

Panel A: Amino acid sequence alignment

```
AEH27625     118 APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK 177
GCQL01020543 792 ----------------H------------------------------------------ 613

AEH27625     178 DMSQQVKQALLEKHPDYQYKPRRPCERRR                                206
GCQL01020543 612 ----------------------------                                 526
```

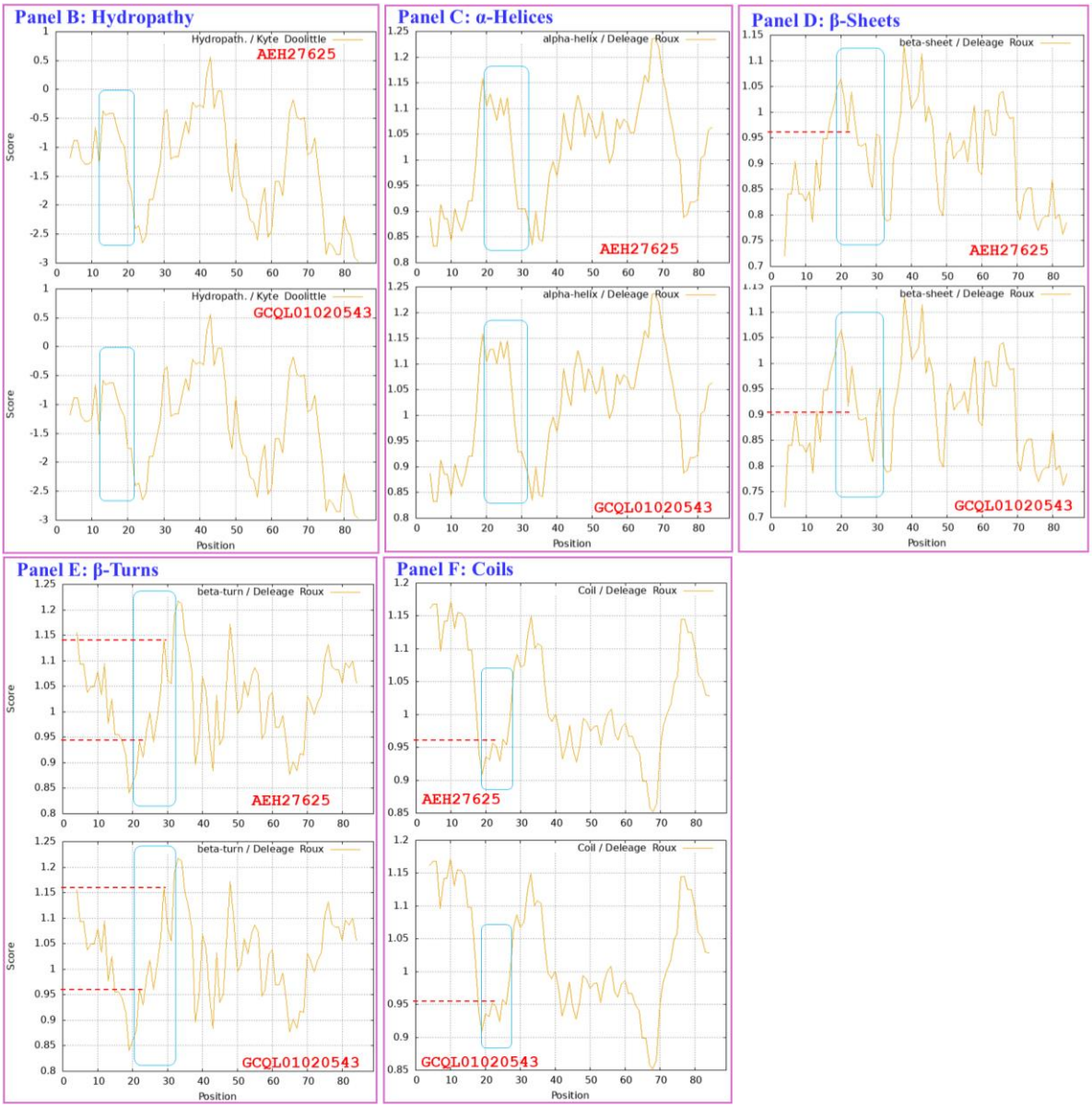


Figure 19. Correlation of the changes in the primary and secondary structures of the HMG-box_ROX1-like domains of the reference MAT1-2-1 protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 and the mutant MAT1-2-1 protein encoded by the transcriptome assembly GCQL01020543 derived from the *H. sinensis* strain L0106. Panel A shows alignment of amino acid sequences of the HMG-box_ROX1-like domain sequences of the MAT1-2-1 proteins with an amino acid substitution in red; the hyphens indicate identical amino acid residues. The ExPASy ProtScale plots show the changes in hydrophobicity (Panel B) and 2D structures (Panels C–F for α-helices, β-sheets, β-turns, and coils) with open rectangles in blue outlining the changes in topology and waveform of the plots.

No changes have been found in the amino acid sequence and the 2D structures (α-helices, β-sheets, β-turns, and coils) in the HMG-box_ROX1-like domain of MAT1-2-1 protein that was encoded by the

952 metatranscriptome assembly OSIN7649 (379→591) of the *C. sinensis* insect-fungi complex, comparing
953 with the reference MAT1-2-1 protein AEH27625 derived from the *H. sinensis* CS68-2-1229 (Figure S4,
954 Table S11) [Zhang *et al*. 2009; Xia *et al*. 2017]. The HMG-box_ROX1-like domain of the OSIN7649-
955 encoded MAT1-2-1 protein was clustered into Branch b1α in the Bayesian clustering tree (Figure 2, Table
956 S11).

957 **III-11 Heterogenous fungal sources of the MAT1-1-1 and MAT1-2-1 proteins**

958 Although wild-type *C. sinensis* isolates are often considered as impure *H. sinensis*, the analysis of
959 mutant full-length mating proteins disclosed their heterogenous fungal sources (Table S3). Among the wild-
960 type *C. sinensis* isolates that were analyzed for the full-length MAT1-1-1 and MAT1-2-1 proteins (*cf*.
961 Figures 3–9 and 12–15), 34 isolates have ITS genotyping information in the GenBank database, which
962 belonged to GC-biased Genotypes #1 and #3 of *O. sinensis* (Table S3). Although the ITS sequences in the
963 29 of 34 wild-type *C. sinensis* isolates appeared to show high homology (≥97%) to GC-biased Genotype
964 #1 *H. sinensis*, the mutant MAT1-1-1 and MAT1-2-1 proteins might not be necessarily derived from *H.*
965 *sinensis*, because the *H. sinensis* ITS sequences were most easily amplified during regular PCR under the
966 experimental settings, including the selection of primers, unless the PCR experiments included amplicon
967 cloning with selection and sequencing of sufficient (usually 30–50) white colonies. In other words, there
968 were still possibilities that mutant proteins were derived from unvalidated heterospecific or genotypic fungi,
969 because of the impure nature of the wild-type *C. sinensis* isolates.

970 Among the 20 *O. sinensis* strains analyzed for the truncated MAT1-1-1 and MAT1-2-1 proteins (*cf*.
971 Figures 10 and 16), 15 strains have ITS genotyping information in the GenBank database. Five (CS6-251,
972 CS34-291, CS70-1208, CS71-1220, and CS68-2-1228) of the 15 strains are heterogenous, containing both
973 GC-biased Genotype #1 and AT-biased Genotype #5 of *O. sinensis* (Tables S4–S6) [Bushley *et al*. 2013; Li
974 *et al*. 2013]. Actually, as shown in Table S6, the AT-biased *O. sinensis* may be AT-biased Genotype #17
975 (homology 98.1%–98.3%), rather than Genotype #5 (less similarities of 97.2%–97.6%) [Zhu & Li 2017;
976 Li *et al*. 2020b, 2022, 2023e]. In addition to the 15 *O. sinensis* strains, the taxonomy positions of the
977 remaining 5 strains (CS26-277, CS36-1294, CS68-5-1216, CS70-1211, and CS70-1212) listed in Table S4,
978 plus 4 other ascosporic strains (1210, 1213, 1223, and 1226) that were studied by Bushley *et al*. [2013],
979 lacked the validated ITS molecular taxonomic evidence, not being listed in Tables S5–S6.

980 As shown in Figure 16, the truncated proteins AGW27548 and AGW27555 that derived from *O.*
981 *sinensis* strains CS68-2-1229 and CS71-1220 are 9 and 10 amino acids shorter at the C-termini than the
982 proteins AGW27554 and AGW27543 that derived from the *O. sinensis* strains CS71-1219 and CS91-1291,
983 respectively. The *MAT1-2-1* genes (KC429550 and KC429557) in the *O. sinensis* strains CS68-2-1229 and

984 CS71-1220 are 31 nucleotides shorter at the 3' ends than the *MAT1-2-1* genes (KC429545 and KC429556)
985 in the strains CS71-1219 and CS91-1291. The 4 proteins were derived from cDNA that was amplified from
986 *MAT1-2-1* genes using the same pair of primers, *Mat1-2F*/*Mat1-2R* according to Bushley *et al*. [2013].
987 Bushley *et al*. [2013] and Li *et al*. [2013] reported that the *O. sinensis* strain CS68-2-1229 is homogenous
988 containing only GC-biased Genotype #1 and the strain CS71-1220 is heterogeneous containing the
989 independent genomes of both GC-biased Genotype #1 and AT-biased Genotype #17 of *O. sinensis* (Figures
990 S5–S6). The sequences of the truncated proteins AGW27548 and AGW27555 are 100% identical. It then
991 needs to be further verified whether the strain CS68-2-1229 was heterogeneous similar to the strain CS71-
992 1220 cooccurring with the genomes of both GC-biased Genotype #1 and AT-biased Genotype #17 of *O.*
993 *sinensis* (*cf*. Table S4–S6) and whether both the truncated proteins AGW27548 and AGW27555 were
994 derived from mutant *MAT1-2-1* genes, because there are no repetitive copies of *MAT1-1-1* and *MAT1-2-1*
995 genes in the genomes of GC-biased Genotype #1 *H. sinensis* [Li *et al*. 2025] and the sequences of all 17
996 genotypes of *O. sinensis*, regardless of whether they are GC-biased or AT-biased, reside in independent
997 genomes, belonging to different fungi [Kinjo & Zang 2001; Stensrud *et al*. 2007; Xiao *et al*. 2009; Yao &
998 Zhu 2016; Zhu & Li 2017; Li *et al*. 2020b, 2022, 2023d, 2023e].

999 # IV. Discussion

1000 ## IV-1 Correlation of the hydrophobicity, primary, secondary, and tertiary structures
1001 of the DNA binding domains in the MAT1-1-1 and MAT1-2-1 proteins

1002 The MATα_HMGbox domain of MAT1-1-1 protein and the HMG-box_ROX1-like domain of MAT1-
1003 2-1 protein are distinct DNA-binding domains critical for fungal mating-type regulation but differ in
1004 evolutionary lineage, structural architecture, and functional roles [Jackson *et al*. 2013; Zheng *et al*. 2013;
1005 Kim *et al*. 2015]. The HMG box domains in mating proteins are multifunctional motifs central to
1006 transcriptional regulation of mating-related genes, *O. sinensis* sexual reproduction, and host adaptation in
1007 *C. sinensis* insect-fungi complex. Our study presented in this paper demonstrated that the 1–3 amino acid
1008 substitutions within the MATα_HMGbox and HMG-box_ROX1-like domains of MAT1-1-1 and MAT1-2-
1009 1 proteins, respectively, of wild-type *C. sinensis* isolates and *C. sinensis* insect-fungi complexes resulted in
1010 changes in the hydrophobic properties and secondary and tertiary structures of the DNA binding domains
1011 of the proteins (*cf*. Figures 3–9, 11–15, 17–19, and S5). The variable sequences of the DNA binding
1012 domains were clustered into different clades in the Bayesian clustering trees (*cf*. Figures 1–2; Tables
1013 S8–S11).

1014 The MATα_HMGbox domain of MAT1-1-1 protein contains a core structure of 3 α helices, which

1015 ensures the high affinity of sequence-preferent binding to AT-rich DNA motifs (*e.g.*, ATTTA) to facilitate
1016 transcriptional machinery assembly involving mating-type-specific genes, and activates gene transcription
1017 by inducing DNA conformational changes through physical bending or twisting the target DNA segments
1018 [Martin *et al*. 2010; Ait Benkhali *et al*. 2013; Rams *et al*. 2022]. The domain directly participates in promoter
1019 recognition of specific genes relating to the heterothallic reproductive process [Yamamoto *et al*. 1997].

1020 Table S7 shows the hydropathy values for amino acids that are defined by their side chains according
1021 to Kyte & Doolittle [1982] provided by https://web.expasy.org/protscale/: the larger the value, the stronger
1022 the hydrophobicity; negative values indicate hydrophilicity [Kyte & Doolittle 1982; Baxevanis *et al*. 1995;
1023 Thapar 2015]. The core α-helical stereostructures of the domains are stabilized through a hydrophobic core
1024 that is composed of nonpolar amino acids, such as Isoleucine (Ile, I), Valine (Val, V), Leucine (Leu, L),
1025 Phenylalanine (Phe, F), Cysteine (Cys, C), Methionine (Met, M), and Alanine (Ala, A), which have
1026 hydropathy indexes of 4.5, 4.2, 3.8, 2.8, 2.5, 1.9, and 1.8 , respectively (*cf*. Table S7) [Kyte & Doolittle
1027 1982].

1028 The HMG-box_ROX1-like domain of MAT1-2-1 protein also consists of asymmetrically arranged 3 α
1029 helices connected by flexible loops, forming a key hydrophobic core that is stabilized through nonpolar
1030 amino acids aforementioned (*cf*. Table S7). The domain of MAT1-2-1 protein regulates gene expression in
1031 a sequence-specific manner by binding ATTAAT or ATTGTT motifs [Jackson *et al*. 2013; Zheng *et al*. 2013;
1032 Kim *et al*. 2015]. This domain is implicated in chromatin remodeling and transcriptional repression by
1033 interacting with other co-repressors, similar to the repressor of oxygen-regulated genes 1 (ROX1) domain
1034 in yeast, through opposing pathways to the MATα_HMGbox domain of MAT1-1-1 protein
1035 [Balasubramanian *et al*. 1993; Zitomer *et al*. 1997; Zheng *et al*. 2013]. ROX1 represses its own promoter,
1036 creating a feedback loop for hypoxic response regulation as a central hypoxia-responsive regulator playing
1037 roles in chromatin remodeling following DNA binding and integration of metabolic signals, etc.
1038 [Kastaniotis & Zitomer 2000].

1039 No published studies to date directly address heterodimerization between the MAT1-1-1 and MAT1-2-
1040 1 proteins of *O. sinensis*, although the heterodimer formation between the MATα_HMGbox domain of
1041 MAT1-1-1 protein and HMG-box_ROX1-like domain of MAT1-2-1 protein is believed as a critical
1042 mechanism in fungal mating regulation, particularly in heterothallic ascomycetes [Kües & Casselton 1993;
1043 Asante-Owusu *et al*. 1996; Jacobsen *et al*. 2002]. However, the MATα_HMGbox and HMG-box_ROX1-
1044 like domains likely interact each other enabling synergistic cooperation of DNA binding through
1045 complementary electrostatic surfaces [Hancock *et al*. 2019]. The heteromorphic stereostructures of the
1046 domains of the mutant mating proteins derived from wild-type *C. sinensis* isolates presented in this study
1047 likely alter the complementary cooperation of the mating proteins of the self-sterile *O. sinensis* under

1048 heterothallic or hybrid outcrossing and subsequent functionalities of activations or repressions of mating-
1049 related gene transcriptions [Li *et al*. 2023c, 2024b, 2025].

1050 **IV-2 Heterogenous fungal sources of the MAT1-1-1 and MAT1-2-1 proteins**

1051 The analysis of mutant full-length mating proteins derived from wild-type *C. sinensis* isolates (Figures
1052 3–9 and 12–15) disclosed their heterogenous fungal sources (Table S3) [Zhang *et al*. 2009, 2011, 2012,
1053 2013, 2014; Zhang & Zhang 2015]. A total of 34 wild-type *C. sinensis* isolates were examined for molecular
1054 taxonomy ITS sequences that are available in GenBank, which are genomically independent GC-biased
1055 Genotypes #1 and #3 of *O. sinensis* fungi. However, whether the impure wild-type *C. sinensis* isolates
1056 cooccur with other heterospecific fungi remains to be mycologically and molecularly addressed and
1057 outlined by using culture-dependent techniques or culture-independent PCR protocols using amplicon
1058 cloning techniques with selecting and sequencing sufficient amount (usually ≥30) of white colonies.

1059 The 5 *O. sinensis* strains (CS68-2-1229, CS68-5-1216, CS71-1219, CS71-1220, and CS91-1291; *cf*.
1060 Figures 10 and 16) contain both truncated MAT1-1-1 and MAT1-2-1 proteins shown in red in Table S4
1061 [Bushley *et al*. 2013; Li *et al*. 2013]. However, 15 other *O. sinensis* strains CS6-251, CS34-291, CS68-2-
1062 1228, CS70-1208, CS18-266, CS25-273, CS37-295, CS71-1218, CS76-1284, CS560-961, CS561-964,
1063 CS26-277, CS36-1294, CS70-1211, and CS70-1212 contain the truncated MAT1-1-1 proteins (*cf*. Figures
1064 10 and S1) but the full-length MAT1-2-1 proteins (*cf*. Figures 16 and S2; Table S2). The genotypic
1065 information for the *O. sinensis* strains indicates heterogeneous fungal sources of the truncated MAT1-1-1
1066 and MAT1-2-1 proteins, derived from GC-biased Genotype #1 and/or AT-biased Genotype #5 of *O. sinensis*,
1067 which are genomically independent belonging to different fungi [Stensrud *et al*. 2007; Xiao *et al*. 2008;
1068 Gao *et al*. 2012; Hu *et al*. 2013; Li *et al*. 2013, 2016a, 2020b, 2022, 2023d, 2023e; 2024a; Mao *et al*. 2013;
1069 Zhu & Li 2017; Jin *et al*. 2020; Liu *et al*. 2020; Shu *et al*. 2020]. Other strains CS26-277, CS36-1294,
1070 CS68-5-1216, CS70-1211, and CS70-1212 shown in black in Table S4, as well as 4 other strains (Clones
1071 1210, 1213, 1223, and 1226 derived from the cultures of mono-ascospores of *C. sinensis* [Li *et al*. 2025])
1072 previously reported by Bushley *et al*. [2013], lacked ITS molecular genotyping information and might be
1073 derived from other genomically independent, heterospecific or genotypic fungi, which need to be addressed
1074 publicly and updated in the GenBank database by researches of Bushley *et al*. [2013] and Li *et al*. [2013].

1075 A contradiction was noted regarding Strain CS68-2-1229 (in dark green and underlined in Figures 10
1076 and 16), which was isolated from a culture of mono-ascospores [Bushley *et al*. 2013; Li *et al*. 2013]. *C.*
1077 *sinensis* ascospores have been demonstrated as fungal complexes containing genomically independent GC-
1078 biased Genotypes #1 and #14, AT-biased Genotypes #5–6 and #16 of *O. sinensis*, and *Paecilomyces hepiali*
1079 [Li *et al*. 2023d, 2023e]. Two MAT1-1-1 protein sequences with different GenBank accession numbers,

1080 AGW27560 for a full-length protein and AGW27528 for an N-terminally truncated protein, were derived
1081 from the *O. sinensis* Strain CS68-2-1229. The C-terminally truncated MAT1-2-1 protein AGW27548 was
1082 also derived from this *O. sinensis* strain. All these 3 protein sequences were released by GenBank on the
1083 same date (28-SEP-2013). The truncated AGW27528 and AGW27548 protein sequences were translated
1084 from the cDNA sequences that were amplified using the pairs of primers, *i.e.*, *m1cDNAF3*/*m1cDNAR5* to
1085 amplify *MAT1-1-1* cDNA and *Mat1-2F*/*Mat1-2R* to amplify *MAT1-2-1* cDNA (*cf.* Figures S1–S2) [Bushley
1086 *et al*. 2013; Li *et al*. 2013]. Although the protein AGW27528 is among the 9 longest truncated MAT1-1-1
1087 proteins (162 amino acids) shown in Figure 10, the protein AGW27548 (49 amino acids) is 10 residues
1088 shorter than the longest truncated MAT1-2-1 protein AGW27543 (59 amino acids) shown in Figure 16. The
1089 complex results indicate that the protein AGW27548 might be derived from a mutant *MAT1-2-1* gene. Li *et*
1090 *al*. [2025] reported that no repetitive copies of *MAT1-1-1* and *MAT1-2-1* genes occurred in the genome
1091 assemblies of the 5 *H. sinensis* strains. Thus, it remains to be examined whether the strain CS68-2-1229
1092 was truly homogenous containing only Genotype #1 *H. sinensis* as reported by Bushley *et al*. [2013] and
1093 Li *et al*. [2013], or heterogeneous containing 2 or more independent genotypical or heterospecific genomes,
1094 similar to the heterogeneous strains (CS6-251, CS34-291, CS68-2-1228, CS70-1208, and CS71-1220) as
1095 shown in Tables S4–S6 [Bushley *et al*. 2013; Li *et al*. 2013].

1096 IV-3 The MAT1-1-1 and MAT1-2-1 proteins involving in sexual reproduction of *O.* 1097 *sinensis*

1098 The MAT1-1-1 and MAT1-2-1 proteins coregulate mating compatibility and sexual development in the
1099 sexual reproduction of *O. sinensis*, in which the MATα_HMGbox and HMG-box_ROX1-like domains play
1100 the key roles. The self-fertilization hypothesis using homothallism or psuedohomothalism strategy has been
1101 proposed for *O. sinensis* [Bushley *et al*. 2013; Hu *et al*. 2013] based on genetic research data. However,
1102 Zhang and Zhang [2015] argued against the hypotheses and believed that *O. sinensis* uses facultative
1103 hybridization for sexual reproduction on the basis of the findings of differential occurrence of *MAT1-1-1*
1104 and *MAT1-2-1* genes in over 170 wild-type *O. sinensis* isolates derived from natural *C. sinensis* insect-
1105 fungal complex specimens collected from different production areas on the Qinghai–Tibet Plateau.
1106 Furthermore, Li *et al*. [2023c, 2024b, 2025] reported alternative slicing, differential occurrence, and
1107 differential translation of *MAT1-1-1* and *MAT1-2-1* genes and the heteromorphic stereostructures of MAT1-
1108 1-1 and MAT1-2-1 proteins of *O. sinensis* strains and the wild-type *C. sinensis* isolates. In addition, the
1109 analysis of DNA binding domains presented in this study further demonstrated diverse stereostructures of
1110 the MATα_HMGbox domain of MAT1-1-1 proteins and the HMG-box_ROX1-like domain of MAT1-2-1
1111 proteins. It is clearly demonstrated that *O. sinensis* experiences self-sterility and uses a heterothallic or

1112 hybrid strategy, even parasexuality, to accomplish sexual reproduction during the lifecycle of *C. sinensis*
1113 insect-fungi complex [Kück 1993; Bennett & Johnson 2003; Sherwood & Bennett 2009; Seervai *et al*. 2013;
1114 Zhu & Li 2017; Nakamura *et al*. 2019; Samarasinghe *et al*. 2020; Steensels *et al*. 2021; Mishra *et al*. 2021;
1115 Kück *et al*. 2021; Li *et al*. 2023c, 2023e, 2024b, 2025].

1116 Distinct from the homothallic reproduction, the heterothallic or hybrid reproduction of *O. sinensis*
1117 requires mating partners. Among the 17 genotypes of *O. sinensis* [Kinjo & Zang 2001; Chen *et al*. 2004,
1118 2011; Stensrud *et al*. 2005, 2007; Xiao *et al*. 2008; Zhu *et al*. 2010; Barseghyan *et al*. 2011; Li *et al*. 2013,
1119 2022; Mao *et al*. 2013; Zhu & Li 2017; Du *et al*. 2020; Hėnault *et al*. 2020], GC-biased Genotype #1 *H.*
1120 *sinensis* has been proposed as the sole anamorph of *O. sinensis* [Wei *et al*. 2006]. Wei *et al*. [2016] reported
1121 the success of industrial cultivation of *C. sinensis* insect-fungi complexes, which, however, displayed a
1122 species contradiction between the anamorphic inoculants of 3 *H. sinensis* strains and the sole teleomorph
1123 of AT-biased Genotype #4 that was detected in the fruiting body of the cultivated *C. sinensis* insect-fungi
1124 complexes. The sequences of AT-biased Genotypes #4–6 and #15–17, as well as the sequences of GC-
1125 biased Genotypes #2–3 and #7–14 of *O. sinensis*, do not reside in the genome assemblies ANOV00000000,
1126 JAAVMX000000000, LKHE00000000, LWBQ00000000, and NGJJ00000000 of the *H. sinensis* strains
1127 Co18, IOZ07, 1229, ZJB12195, and CC1406-20395, respectively [Stensrud *et al*. 2007; Xiao *et al*. 2008;
1128 Gao *et al*. 2012; Hu *et al*. 2013; Li *et al*. 2013, 2016a, 2020b, 2022, 2023d, 2023e; 2024a; Mao *et al*. 2013;
1129 Zhu & Li 2017; Jin *et al*. 2020; Liu *et al*. 2020; Shu *et al*. 2020]. Li *et al*. [2024a] further demonstrated that
1130 the sequences of Genotypes #2–17 of *O. sinensis* were not the repetitive copies in the genome of Genotype
1131 #1 *H. sinensis*, invalidating the hypotheses of "RIP mutation" induced or generated "ITS pseudogene" and
1132 "rRNA pseudogenes" in the genome of *H. sinensis* raised by Li *et al*. [2013, 2017, 2020c]. These data
1133 indicate the 17 genotypes of *O. sinensis* are genome independent and belong to different fungi. In addition,
1134 the data demonstrated that the sole anamorph hypothesis for *H. sinensis* raised by Wei *et al*. [2006] has
1135 failed to pass the examination of Koch's postulates.

1136 Li *et al*. [2020a, 2023b] revised the inoculant information for the industrial cultivation project and
1137 reported that the cultivation project used the cultures of the *C. sinensis* ascospores as the inoculants, rather
1138 than pure *H. sinensis* strains as previously disclosed by Wei *et al*. [2016]. However, Li *et al*. [2023d, 2023e]
1139 reported detections of GC-biased Genotypes #1 and #14 and AT-biased Genotypes #5–6 and #16, as well
1140 as *Paecilomyces hepiali* in the ascospores of the natural *C. sinensis* insect-fungi complexes through a
1141 culture-independent approach using a PCR amplicon cloning technique with selecting and sequencing of
1142 ≥30 white colonies. Slightly differently, Li *et al*. [2013] used a culture-dependent protocol and reported
1143 detection of GC-biased Genotype #1 and AT-biased Genotype #5 in the 8 cultures of mono-ascospores.

1144 Furthermore, Li *et al*. [2020b, 2022, 2023d, 2023e] and Gao *et al*. [2012] reported that AT-biased Genotype
1145 #4 of *O. sinensis* occurred in the stromata during the entire course of *C. sinensis* maturation and stromal
1146 fertile portion (SFP) containing numerous ascocarps of *C. sinensis* insect-fungi complex; the abundance of
1147 Genotype #4 was high in immature stromata and was largely declined in mature stromata and SFP but,
1148 unfortunately, was absent from the *C. sinensis* ascospores. Thus, although Wei *et al*. [2016] reported the
1149 detection of the sole teleomorphic AT-biased Genotype #4 in the cultivated *C. sinensis* insect-fungi complex
1150 and Li *et al*. [2020a, 2023b] revised the inoculant information, the fungal source for the sole teleomorph,
1151 AT-biased Genotype #4 of *O. sinensis* remains scientifically uncertain in artificial cultivation of *C. sinensis*
1152 insect-fungi complex.

1153 In addition to the formal reports of unsuccessful studies in artificial cultivation of the fruiting bodies of
1154 *C. sinensis* [Holliday *et al*. 2008; Stone 2010; Hu *et al*. 2013], Zhang *et al*. [2013] summarized the 40-year
1155 history of cultivation failures of insect-fungal complex using "pure" mycological strategy in the academic
1156 research-oriented settings. In contrast, Wei *et al*. [2016] reported the cultivation success in the industrial
1157 product-oriented settings, which was likely attributed to the application of a "mycologically impure"
1158 cultivation strategy. This cultivation strategy has become the key factor to industrial success, based on the
1159 following reports:

1160 (1) using the cultures of mycologically impure ascospores of natural *C. sinensis* as the inoculants [Li *et al*.
1161 2020a, 2023b], which most likely combined with the cocultures of *C. sinensis* stroma and/or stromal fertile
1162 portion (SFP) containing numerous ascocarps that, most importantly, contain AT-biased Genotype #4 of *O.*
1163 *sinensis*, consistent with the discovery of the sole teleomorphic *O. sinensis* genotype [Wei *et al*. 2016]; and
1164 (2) adding the soil that was collected from the natural *C. sinensis* production areas on the Qinghai–Tibet
1165 Plateau into the industrial cultivation system as disclosed by Wei *et al*. [2016].

1166 Unfortunately, purification and genome sequencing of GC- and AT-biased Genotypes #2–17 of *O. sinensis*
1167 have not been reported to date. Thus, genomic and transcriptomic information is lacking regarding
1168 mutations of the *MAT1-1-1* and *MAT1-2-1* genes, especially the DNA binding domains of MAT1-1-1 and
1169 MAT1-2-1 proteins.

1170 Regardless of whether MAT1-1-1 and MAT1-2-1 proteins form heterodimer or interactively involve
1171 complementary electrostatic surfaces and enabling cooperative DNA binding, the synergy of the mating
1172 proteins of *O. sinensis* constitutes the core mechanism regulating mating type recognition, triggering the
1173 downstream nuclear fusion signaling pathways, and fruiting body development [Debuchy *et al*. 2006; Jones
1174 & Bennett 2011; Zheng & Wang 2013; Wilson *et al*. 2015; Sun *et al*. 2019]. The alternative splicing,
1175 differential occurrence, and differential transcription of mating-type and pheromone receptor genes in

Hirsutella sinensis and *C. sinensis* isolates invalidated the self-fertilization hypothesis at the genomic and transcriptomic levels but suggested self-sterility for *O. sinensis* under heterothallic or hybrid reproduction [Hu *et al*. 2013; Bushley *et al*. 2013; Zhou *et al*. 2013; Wei *et al*. 2006; Li *et al*. 2019, 2023c, 2024b, 2025]. Furthermore at the protein level, the discoveries of a good number of mutant mating proteins reported by Li *et al*. [2025], which have mutant MATα_HMGbox and HMG-box_ROX1-like domains in MAT1-1-1 and MAT1-2-1 proteins, respectively, detected in *O. sinensis* strains, wild-type *C. sinensis* isolates, and insect-fungal complex demonstrated in this study, may indicate heterogeneous fungal sources of the mating proteins to accomplish coordinated heterothallic or hybrid reproduction of *O. sinensis* in the lifecycle of *C. sinensis* insect-fungi complexes.

Conclusions

The analysis of the DNA binding domains of mating proteins in the 173 *O. sinensis* strains and wild-type *C. sinensis* isolates revealed heteromorphic stereostructures of the MATα_HMGbox and HMG-box_ROX1-like domains of MAT1-1-1 and MAT1-2-1 proteins, respectively. A few amino acid substitutions at various mutation sites in the DNA binding domains were demonstrated in numerous mutant mating proteins derived from wild-type *C. sinensis* isolates and correlated with the changes in hydrophobic properties and secondary and tertiary structures under different AlphaFold 3D structural cords, which were clustered into various Bayesian clustering clades. The diversity of heteromorphic MATα_HMGbox and HMG-box_ROX1-like domains of the mating proteins discovered in wild-type *C. sinensis* isolates may lead to stereostructure-related functional alterations of the mating process. The stereostructure variants of the DNA binding domains of mutant MAT1-1-1 and MAT1-2-1 proteins suggest diverse sources of the mating proteins produced by cooccurring *O. sinensis* genotypes or heterospecific fungi coexisting in natural *C. sinensis* that have been identified in mycobiota, molecular, metagenomic, and metatranscriptomic studies, regardless of whether culture-dependent or independent research strategies were used.

Acknowledgments

[REFERENCES]

1. 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, Žídek 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

2. Asante-Owusu RN, Banham AH, Böhnert HU, Mellor EJC, Casselton LA. Heterodimerization between two classes of homeodomain 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
3. 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
4. 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
5. Baxevanis AD, Bryant SH, Landsman D. Homology model building of the HMG-1 box structural domain. Nucleic Acids Research 1995; 23(6): 1019−1029. DOI: 10.1093/nar/23.6.1019
6. 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
7. 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
8. 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
9. 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
10. 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 (accessed on 30 January 2025).
11. 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/HHCVOJPU055360T7.html (accessed on 3 May 2025).
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. 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
14. Debuchy R, Turgeo BG. Mating-Type Structure, Evolution, and Function in Euascomycetes. In (eds. Kües U. & Fischer, R.) Growth, Differentiation and Sexuality. Springer (2006). pp. 293–323.
15. 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
16. 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
17. Gao L, Li X-H, Zhao J-Q, Lu J-H, Zhao J-G, Zhu J-S. Maturation of *Cordyceps sinensis* associates with alterations of fungal expressions of multiple *Ophiocordyceps sinensis* mutants with transition and transversion point mutations in stroma of *Cordyceps sinensis*. Beijing Da Xue Xue Bao. 2012; 44(3): 454–463. doi:

1251 10.3969/j.issn.1671-167X.2012.03.025
1252 18. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein Identification and
1253 Analysis Tools on the ExPASy Server. Chapter 52 (In) John M. Walker (ed): The Proteomics Protocols Handbook,
1254 Humana Press (2005). pp. 571–607.
1255 19. Guo M-Y, Liu Y, Gao Y-H, Jin T, Zhang H-B, Zhou X-W. Identification and bioactive potential of endogenetic
1256 fungi isolated from medicinal caterpillar fungus *Ophiocordyceps sinensis* from Tibetan Plateau. Int J Agric Biol.
1257 2017; 19: 307–313. DOI: 10.17957/IJAB/15.0281
1258 20. Hancock SP, Cascio D, Johnson RC. Cooperative DNA binding by proteins through DNA shape complementarity.
1259 Nucleic Acids Res. 2019; 47(16):8874–8887. doi: 10.1093/nar/gkz642
1260 21. Hawksworth DL, Crous PW, Redhead SA, Reynolds DR, Samson RA, Seifert KA, and 82 other authors. The
1261 Amsterdam declaration on fungal nomenclature. IMA Fungus. 2011; 2: 105−112. DOI:
1262 10.5598/imafungus.2011.02.01.14
1263 22. Hénault M, Marsit S, Charron G, Landry CR. The effect of hybridization on transposable element accumulation
1264 in an undomesticated fungal species. eLife. 2020; 9: e60474. DOI: https://doi.org/10.7554/eLife.60474
1265 23. Holliday J, Cleaver M. Medicinal value of the caterpillar fungi species of the genus Cordyceps (Fr.) Link
1266 (Ascomycetes). A review. Int. J. Med. Mushrooms. 2008; 10: 219–234. DOI: 10.1615/IntJMedMushr.v10.i3.30
1267 24. 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
1268 uncovers the secrets of sex and lifestyle in caterpillar fungus. Chin. Sci. Bull. 2013; 58: 2846–2854. doi:
1269 10.1007/s11434-013-5929-5
1270 25. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogeny. Bioinformat. 2001; 17: 754−755.
1271 DOI: 10.1093/bioinformatics/17.8.754
1272 26. Jackson D, Lawson T, Villafane R, Gary L. Modeling the structure of yeast MATα1: An HMG-Box motif with
1273 a C-terminal helical extension. Open J Biophysics, 2013, 3: 1−12. http://dx.doi.org/10.4236/ojbiphy.2013.31001
1274 27. Jacobsen S, Wittig M, Pöggeler S. Interaction Between Mating-Type Proteins From the Homothallic Fungus
1275 Sordaria macrospora. Curr. Genet. 2002; 41: 150–158. doi: 10.1007/s00294-002-0276-0
1276 28. Jiang Y, Yao Y-J. A review for the debating studies on the anamorph of *Cordyceps sinensis*. Mycosistema. 2003;
1277 22(1): 161–176.
1278 29. 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
1279 Y-G. Genome sequencing and analysis of fungus *Hirsutella sinensis* isolated from *Ophiocordyceps sinensis*.
1280 AMB Expr. 2020; 10: 105. https://doi.org/10.1186/s13568-020-01039-x
1281 30. Jones SK, Bennett RJ. Fungal mating pheromones: choreographing the dating game. Fungal Genet. Biol. 2011;
1282 48(7): 668–676. doi: 10.1016/j.fgb.2011.04.001
1283 31. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A,
1284 Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R,
1285 Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T,
1286 Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein
1287 structure prediction with AlphaFold. Nature 2021; 596: 583–589. https://doi.org/10.1038/s41586-021-03819-2
1288 32. Kang Q, Zhang J, Chen F, Dong C, Qin Q, Li X, Wang H, Zhang H and Meng Q. Unveiling mycoviral diversity
1289 in *Ophiocordyceps sinensis* through transcriptome analyses. Front. Microbiol. 2024; 15: 1493365. doi:
1290 10.3389/fmicb.2024.1493365
1291 33. Kastaniotis AJ, Zitomer RS. Rox1 mediated repression. Oxygen dependent repression in yeast. Adv Exp Med
1292 Biol. 2000; 475: 185–195. https://pubmed.ncbi.nlm.nih.gov/10849660/
1293 34. 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

35. Kinjo N, Zang M. Morphological and phylogenetic studies on *Cordyceps sinensis* distributed in southwestern China. Mycoscience. 2001; 42: 567–574. DOI: 10.1007/BF02460956
36. Kück U, Bennett RJ, Wang L and Dyer PS. Editorial: Sexual and Parasexual Reproduction of Human Fungal Pathogens. Front. Cell. Infect. Microbiol. 2022; 12: 934267. doi: 10.3389/fcimb.2022.934267
37. 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
38. 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
39. 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
40. Li C-L. A study of *Tolypocladium sinense* C.L. Li. sp. nov. and cyclosporin production. Acta Mycol. Sinica. 1988; 7: 93−98.
41. Li M-M, Zhang J-H, Qin Q-L, Zhang H, Li X, Wang H-T, Meng Q. Transcriptome and Metabolome Analyses of Thitarodes xiaojinensis in Response to *Ophiocordyceps sinensis* Infection. Microorganisms 2023a; 11: 2361. https://doi.org/10.3390/microorganisms11092361
42. 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 2020a; 63, 1093–1095. https://doi.org/10.1007/s11427-020-1686-7
43. 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 2023b; 66, 892. https://doi.org/10.1007/s11427-022-2232-5
44. 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
45. Li X-Z, Li Y-L, Wang Y-N, Zhu J-S. Translations of mutant repetitive genomic sequences in *Hirsutella sinensis* and changes in secondary structures and functional specifications of the encoded proteins. International J Mol Sci 2024a; 25(20), 11178; https://doi.org/10.3390/ijms252011178
46. Li X-Z, Li Y-L, Yao Y-S, Xie W-D, Zhu J-S. Further discussion with Li *et al*. (2013, 2019) regarding the "ITS pseudogene hypothesis" for *Ophiocordyceps sinensis*. Mol. Phylogenet. Evol. 2020b; 146: 106728. https://doi.org/10.1016/j.ympev.2019.106728
47. 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. 2023c; 48(10): 2829–2840. DOI: 10.19540/j.cnki.cjcmm.20230213.102
48. 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
49. Li X-Z, Xiao M-J, Li Y-L, Gao L, Zhu J-S. Mutations and differential transcription of mating-type and pheromone receptor genes in *Hirsutella sinensis* and the natural *Cordyceps sinensis* insect–fungi complex. Biol. (Basel) 2024b; 13(8): 632. https://doi.org/10.3390/biology13080632
50. 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

1337 sequencing and assembly strategies for a repeat-rich fungal genome, *Ophiocordyceps sinensis*. J. Microbiol.
1338 Methods. 2016a; 128: 1−6. DOI: 10.1016/j.mimet.2016.06.025
1339 51. 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
1340 ITS genes in populations of *Ophiocordyceps sinensis* and their implications for molecular systematics. IMA
1341 Fungus. 2020c; 11: 18. DOI:10.1016/j.mimet.2016.06.025
1342 52. Li Y, Jiao L, Yao Y-J. Non-concerted ITS evolution in fungi, as revealed from the important medicinal fungus
1343 *Ophiocordyceps sinensis*. Mol. Phylogenet. Evol. 2013; 68: 373−379. doi: 10.1016/j.ympev.2013.04.010
1344 53. Li Y, Yang R-H, Jiang L, Hu X-D, Wu Z-J, Yao Y-J. rRNA Pseudogenes in Filamentous Ascomycetes as Revealed
1345 by Genome Data. G3 (Bethesda) 2017; 7(8): 2695−2703. doi: 10.1534/g3.117.044016
1346 54. 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
1347 of *Ophiocordyceps sinensis* in the stromal fertile portions and ascospores of natural *Cordyceps sinensis*. PLoS
1348 ONE. 2023d; 18(6): e0286865. https://doi.org/10.1371/journal.pone.0286865
1349 55. 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
1350 cooccurrence of multiple genotypes of *Ophiocordyceps sinensis* in the stromata, stromal fertile portion
1351 (ascocarps) and ascospores of natural *Cordyceps sinensis*. PLoS ONE. 2023e; 18(3): e0270776. DOI:
1352 10.1371/journal.pone.0270776
1353 56. Li Y-L, Li X-Z, Yao Y-S, Xie W-D, Zhu J-S. Molecular identification of *Ophiocordyceps sinensis* genotypes and
1354 the indiscriminate use of the Latin name for the multiple genotypes and the natural insect-fungi complex. Am J
1355 BioMed Sci. 2022; 14(3): 115–135. doi: 10.5099/aj220300115
1356 57. 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
1357 from the intestines of *Hepialus lagii* larvae in increasing infection potency. J. Fungal Res. 2016b; 14: 96−112.
1358 58. 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
1359 analyses reveal evolutionary and geologic context for the plateau fungus *Ophiocordyceps sinensis*. Clin. Med.
1360 2020; 15: 107–119. https://doi.org/10.1186/s13020-020-00365-3
1361 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.
1362 Transcriptome sequencing and analysis of the entomopathogenic fungus *Hirsutella sinensis* isolated from
1363 *Ophiocordyceps sinensis*. BMC Genomics. 2015; 16: 106–123. DOI: 10.1186/s12864-015-1269-y
1364 60. Lu H-L, St Leger RJ. Chapter Seven - Insect Immunity to Entomopathogenic Fungi, Editors: Lovett B, St. Leger
1365 RJ. Advanc Genet, Academic Press, 2016. Vol. 94; pp. 251–85.
1366 61. Mao X-M, Zhao S-M, Cao L, Yan X, Han R-C. The morphology observation of *Ophiocordyceps sinensis* from
1367 different origins. J. Environ. Entomol. 2013; 35(3): 343–353.
1368 62. Mariani V, Biasini M, Barbato A, Schwede T. lDDT: a local superposition-free score for comparing protein
1369 structures and models using distance difference tests. Bioinformatics. 2013; 29(21): 2722–2728.
1370 http://dx.doi.org/10.1093/bioinformatics/btt473
1371 63. Martin T, Lu S-W, van Tilbeurgh H, Ripoll DR, Dixelius C, Dixelius C, Turgeon BG, Debuchy R. Tracing the
1372 Origin of the Fungal a1 Domain Places Its Ancestor in the HMG-Box Superfamily: Implication for Fungal
1373 Mating-Type Evolution. PLoS ONE 2010; 5(12): e15199. doi:10.1371/journal.pone.0015199
1374 64. 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.
1375 Transcriptomic insight into the immune defenses in the ghost moth, *Hepialus xiaojinensis*, during an
1376 *Ophiocordyceps sinensis* fungal infection. Insect Biochem. Mol. Biol. 2015; 64: 1−15. DOI:
1377 10.1016/j.ibmb.2015.06.014
1378 65. Metin B, Findley K, Heitman J. The mating type locus (MAT) and sexual reproduction of *Cryptococcus*
1379 *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

66. Mishra A, Forche A and Anderson MZ. Parasexuality of Candida Species. Front. Cell. Infect. Microbiol. 2021; 11: 796929. doi: 10.3389/fcimb.2021.796929

67. Monzon V, Haft DH, Bateman A. Folding the unfoldable: using AlphaFold to explore spurious proteins. Bioinformatics Advances, 2022; 1(2): vbab043. https://doi.org/10.1093/bioadv/vbab043

68. 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

69. 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

70. Rams B, Kück U. The *Penicillium chrysogenum* tom1 gene a major target of transcription factor MAT1-1-1 encodes a nuclear protein involved in sporulation. Front. Fungal Bio. 2022; 3: 937023. doi: 10.3389/ffunb.2022.937023

71. 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.

72. Rettie SA, Campbell KV, Bera AK, Kang A, Kozlov S, De La Cruz J, Adebomi V, Zhou G, DiMaio F, Ovchinnikov S, Bhardwaj G. Cyclic peptide structure prediction and design using AlphaFold. bioRxiv [Preprint]. 2023; 26:2023.02.25.529956. doi: 10.1101/2023.02.25.529956

73. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space, Systematic Biology, 2012; 61(3): 539–542. https://doi.org/10.1093/sysbio/sys029

74. 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

75. 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

76. 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

77. 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

78. 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

79. Steensels J, Gallone B, Verstrepen KJ. Interspecific hybridization as a driver of fungal evolution and Adaptation. Nat. Rev. Microbiol. 2021; 19: 485–500.

80. Stensrud Ø, Hywel-Jones NL, Schumacher T. Towards a phylogenetic classification of Cordyceps: ITS nrDNA sequence data confirm divergent lineages and paraphyly. Mycol. Res. 2005; 109: 41−56. DOI: 10.1017/s095375620400139x

81. Stensrud Ø, Schumacher T, Shalchian-Tabrizi K, Svegardenib IB, Kauserud H. Accelerated nrDNA evolution and profound AT bias in the medicinal fungus *Cordyceps sinensis*. Mycol. Res. 2007; 111: 409–415. DOI: 10.1016/j.mycres.2007.01.015

82. Stone R. Improbable partners aim to bring biotechnology to a Himalayan kingdom. Sci. 2010; 327: 940–941. DOI: 10.1126/science.327.5968.940

1422 83. Sun S, Coelho MA, David-Palma M, Priest SJ, Heitman J. The evolution of sexual reproduction and the
1423 mating-type locus: Links to pathogenesis of *Cryptococcus* human pathogenic fungi. Annu Rev Genet. 2019;
1424 53:4 17-444. doi: 10.1146/annurev-genet-120116-024755
1425 84. Sung G-H, Hywel-Jones NL, Sung J-M, Luangsa-ard JJ, Shrestha B, Spatafora JW. Phylogenetic classification
1426 of *Cordyceps* and the clavicipitaceous fungi. Stud. Mycol. 2007; 57: 5−59. DOI: 10.3114/sim.2007.57.01
1427 85. Thapar R. Structure-specific nucleic acid recognition by L-motifs and their diverse roles in expression and
1428 regulation of the genome. Biochim Biophys Acta 2015; 1849(6): 677–687. doi: 10.1016/j.bbagrm.2015.02.006
1429 86. Tunyasuvunakool, K., Adler, J., Wu, Z, Green T, Zielinski M, Žídek A, Bridgland A, Cowie A, Meyer C, Laydon
1430 A, Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA,
1431 Potapenko A, Ballard AJ, Romera-Paredes B, Nikolov S, Jain R, Clancy E, Reiman D, Petersen S, Senior AW,
1432 Kavukcuoglu K, Birney E, Kohli P, Jumper J, Hassabis D. Highly accurate protein structure prediction for the
1433 human proteome. Nature 2021; 596: 590–596. https://doi.org/10.1038/s41586-021-03828-1
1434 87. Turgeon BG, Yoder OC. Proposed nomenclature for mating type genes of filamentous ascomycetes. Fungal
1435 Genet. Biol. 2000; 31: 1–5. DOI: 10.1006/fgbi.2000.1227
1436 88. Varadi M, Bertoni D, Magana P, Paramval U, Pidruchna I, Radhakrishnan M, Tsenkov M, Nair S, Mirdita M,
1437 Yeo J, Kovalevskiy O, Tunyasuvunakool K, Laydon A, Žídek A, Tomlinson H, Hariharan D, Abrahamson J,
1438 Green T, Jumper J, Birney E, Steinegger M, Hassabis D, Velankar S. AlphaFold Protein Structure Database in
1439 2024: providing structure coverage for over 214 million protein sequences. Nucleic Acids Res. 2024; 52(D1):
1440 D368–D375. doi: 10.1093/nar/gkad1011
1441 89. Wang Y, Stata M, Wang W, Stajich JE, White MM, Moncalvo JM. Comparative genomics reveals the core gene
1442 toolbox for the fungus-insect symbiosis. mBio. 2018; 9: e00636-18. DOI: 10.1128/mBio.00636-18
1443 90. Wei J-C, Wei X-L, Zheng W-F, Guo W, Liu R-D. Species identification and component detection of
1444 *Ophiocordyceps sinensis* cultivated by modern industry. Mycosystema. 2016; 35(4): 404–410.
1445 91. Wei X-L, Yin X-C, Guo Y-L, Shen N-Y, Wei, J.-C. Analyses of molecular systematics on *Cordyceps sinensis* and
1446 its related taxa. Mycosystema. 2006; 25(2): 192–202.
1447 92. Wilson AM, Wilken PM, van der Nest MA, Steenkamp ET, Wingfield MJ, Wingfield BD. Homothallism: an
1448 umbrella term for describing diverse sexual behaviors. IMA Fungus. 2015; 6(1): 207–214. DOI:
1449 10.5598/imafungus.2015.06.01.13
1450 93. Wroblewski K, Kmiecik S. Integrating AlphaFold pLDDT Scores into CABS-flex for enhanced protein
1451 flexibility simulations. Comput Struct Biotechnol J. 2024; 30(23): 4350–4356. doi: 10.1016/j.csbj.2024.11.047
1452 94. 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
1453 H, Peng Y-Q, Yu Y, Zhang W, Eichler EE, Gao L-Z. The caterpillar fungus, *Ophiocordyceps sinensis*, genome
1454 provides insights into highland adaptation of fungal pathogenicity. Sci. Rep. 2017; 7: 1806.
1455 DOI:10.1038/s41598-017-01869-z
1456 95. Xia F, Liu Y, Shen G-L, Guo L-X, Zhou X-W. Investigation and analysis of microbiological communities in
1457 natural *Ophiocordyceps sinensis*. Can. J. Microbiol. 2015; 61: 104–111. DOI: 10.1139/cjm-2014-0610
1458 96. 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
1459 analysis of the *Ophiocordyceps sinensis* fruiting body reveals putative genes involved in fruiting body
1460 development and cordycepin biosynthesis. Genomics. 2014; 103: 154−159.
1461 https://doi.org/10.1016/j.ygeno.2014.01.002
1462 97. Xiao W, Yang J-P, Zhu P, Cheng K-D, He H-X, Zhu H-X, Wang Q. Non-support of species complex hypothesis
1463 of *Cordyceps sinensis* by targeted rDNA-ITS sequence analysis. Mycosystema. 2009; 28(6): 724–730.
1464 98. 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

99. 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

100. Yang J-L, Xiao W, He H-X, Zhu H-X, Wang S-F, Cheng -KD, Zhu P. Molecular phylogenetic analysis of *Paecilomyces hepiali* and *Cordyceps sinensis*. Acta Pharmaceut. Sinica 2008; 43(4): 421–426.

101. 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.

102. Yao Y-S, Zhu J-S. Indiscriminate use of the Latin name for natural *Cordyceps sinensis* and *Ophiocordyceps sinensis* fungi. Chin. J. Chin. Mater. Med. 2016; 41(7): 1316–1366.

103. 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.

104. 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.

105. 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.

106. 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.

107. 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

108. 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.

109. 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

110. 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

111. Zheng, P.; Wang, C.-S. Sexuality Control and Sex Evolution in Fungi. Sci. Sin. Vitae 2013, 43, 1090–1097.

112. 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

113. 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

114. 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

115. 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

116. Zhu J-S, Guo Y-L, Yao Y-S, Zhou Y-J, Lu J-H, Qi Y, Chen W, Zheng T-Y, Zhang L, Wu Z-M, Zhang L-J, Liu X-

J, Yin W-T. Maturation of *Cordyceps sinensis* associates with co-existence of *Hirsutella sinensis* and *Paecilomyces hepiali* DNA and dynamic changes in fungal competitive proliferation predominance and chemical profiles. J. Fungal Res. 2007; 5(4): 214–224.

117. Zhu J-S, Halpern GM, Jones K. The scientific rediscovery of a precious ancient Chinese herbal regimen: *Cordyceps sinensis*: Part I. J. Altern. Complem. Med. 1998a; 4(3): 289–303. DOI: 10.1089/acm.1998.4.3-289
118. Zhu J-S, Halpern GM, Jones K. The scientific rediscovery of an ancient Chinese herbal medicine: *Cordyceps sinensis*: Part II. J. Altern. Complem. Med. 1998b; 4(4): 429–457. DOI: 10.1089/acm.1998.4.429
119. 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)
120. Zhu J-S, Li Y-L. A Precious Transitional Chinese Medicine, *Cordyceps sinensis*: Multiple heterogeneous *Ophiocordyceps sinensis* in the insect-fungi complex. Lambert Academic Publishing, Saarbrüchen. Germany, 2017.
121. 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