Heteromorphic tertiary structures of MAT1-1-1 and MAT1-2-1 protein variants simultaneously produced in pairs by *Ophiocordyceps sinensis* strains

Xiu-Zhang Li [1], Yu-Ling Li [1], Wei Liu [2], Jian-Zhao Qi [3] and Jia-Shi Zhu [1]*

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

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

[3] Shaanxi Key Laboratory of Natural Products & Chemical Biology, College of Chemistry & Pharmacy, Northwest A&F University, Yangling, 712100, China; qjz@nwafu.edu.cn

* Corresponding author

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

Abstract

The MATα_HMGbox domain in the MAT1-1-1 protein and the HMG-box_ROX1-like domain in the MAT1-2-1 protein play essential roles in DNA binding and regulating the transcription of genes that control sexual reproduction in *Ophiocordyceps sinensis*. Previous studies have reported differential occurrences, differential transcription, and alternative splicing of the *MAT1-1-1*, *MAT1-2-1*, and pheromone receptor genes in *Hirsutella sinensis* (Genotype #1 among 17 genome-independent genotypes of *O. sinensis* fungi). In this study, we further demonstrated that the DNA-binding domains of the paired MAT1-1-1 and MAT1-2-1 proteins derived from each of the 20 purportedly "pure" *O. sinensis* strains exhibit heteromorphic tertiary structures, as predicted by AlphaFold 3D structural modeling. These differentially paired mating proteins are characterized by different truncations, diverse 1–4 amino acid substitutions at distinct sites, altered hydrophobicity and secondary structures, and heteromorphic tertiary structures, indicating divergency in the fungal origins of the mating proteins within the mycologically and genetically impure *O. sinensis* strains. Our findings support the self-sterility hypothesis for *O. sinensis* and likely ensure the fidelity and genetic diversity of heterothallic or hybrid reproduction during the lifecycle of the *Cordyceps sinensis* insect–fungal complex.

Keywords: MATα_HMGbox domain of the MAT1-1-1 protein; HMG-box_ROX1-like domain of the MAT1-2-1 protein; heteromorphic primary, secondary, and tertiary structures of proteins; *Ophiocordyceps sinensis* strain; *Cordyceps sinensis* insect−fungal complex; sexual reproduction; heterothallism; hybridization

1. Introduction

Natural *Cordyceps sinensis*, also known as Chinese cordyceps (non-Latin term) or DōngChóng XiàCǎo 冬虫夏草 in Chinese, exists only in high-altitude alpine meadows (3,000−5,000 m) on the Qinghai-Tibet Plateau, which contributes to its natural scarcity. It has a centuries-long history of clinical application in health maintenance, characterized by the dual reinforcement of "*YIN* 阴" and "*YANG* 阳", as well as in disease amelioration, postillness and postoperative recovery, and antiaging therapy [1−4]. Modern science is increasingly elucidating the complex biochemical mechanisms (*e.g.*, metabolic (glucose and lipid) regulation, organ protection, immunomodulation, anti-inflammatory activity, antioxidant activity, hormone-like activity supporting libido and fertility, and the enhancement of energy–vitality–stamina, etc.) that underpin these traditional applications, proving its status as a valuable

therapeutic resource, despite the complexity and rarity of this medicinal resource. This natural substance, with its unique therapeutic significance, embodies a fusion of nature's ingenuity and traditional medicinal wisdom; however, its high value has led to overharvesting, making it a Level-II protected natural substance [5].

Natural *C. sinensis* comprises the fruiting body of *Ophiocordyceps sinensis* and the remains of mummified *Hepialidae* moth larva, which retains a thick body wall with numerous bristles, an intact larval intestine, head tissues, and fragments of other larval tissues [6−13]. Dual-fluorescence microscopy has revealed multicellular heterokaryotic microstructures of *C. sinensis* ascospores and hyphae, containing mono-, bi-, tri-, and tetra-nucleate cells [14]. Natural *C. sinensis* exhibits apparent genetic heterogeneity, harboring >90 fungal species from at least 37 fungal genera, as well as 17 genomically independent genotypes of *O. sinensis* that differentially naturally co-occur, the abundance of which undergoes dynamic and reciprocal alterations in different compartments of *C. sinensis* during maturation, supporting their genome-independent natural characteristics [9,10,12,13,15−44]. Thus, the Chinese Pharmacopoeia defines this medicinal microecosystem as an insect–fungal complex.

Among the numerous intrinsic co-occurring fungal species [20,31−33,40,41], *Hirsutella sinensis* has been postulated to be the only anamorph of *O. sinensis* [45], although this hypothetical claim needs to be validated based on all 4 criteria of Koch's postulates. In terms of validation, Wei *et al.* [46] reported a successful artificial cultivation project conducted in an industrial product-oriented setting. Unfortunately, they reported a species contradiction that arose between the anamorphic inoculants of 3 *H. sinensis* strains (GC-biased Genotype #1 of *O. sinensis*) inoculated on *Hepialidae* moth larvae and the genome-independent, sole teleomorphic AT-biased Genotype #4 of *O. sinensis* identified in the fruiting body of cultivated *C. sinensis*.

Since the 1840s, the Latin name *Cordyceps sinensis* has been indiscriminately applied to both the teleomorph/holomorph of the fungus *C. sinensis* and the wild insect–fungal complex, disregarding Edward Doubleday's original identification in 1842 that the insect portion of the natural material belongs to Agrotis [47]. This imprecise use of the Latin name has blurred the distinction between the fungus, its host, and the insect−fungal complex, resulting in ambiguity in the scientific literature, commercial trade in the mass market, and governmental regulations worldwide. The fungus was renamed *O. sinensis* in 2007, and *H. sinensis* strain EFCC7287 (GC-biased Genotype #1 of *O. sinensis*) was used as the nomenclature reference [9,10,48−50]. Zhang *et al.* [51] improperly implemented the "One Fungus=One Name" nomenclature rule set by the International Mycological Association (IMA) and replaced the

anamorphic name *H. sinensis* with the teleomorphic name *O. sinensis* according to the international rule [52,53]. They interpreted *O. sinensis* as a single fungus to fulfill the prerequisite of "One Fungus" of the international rule but did not fully consider the 17 genomically independent genotypes of *O. sinensis* fungi, which differentially coexist in the different compartments of the *C. sinensis* insect–fungal complex. In the present paper, we retain the anamorphic name *H. sinensis* for GC-biased Genotype #1 among 17 genotypes of *O. sinensis* and refer to the genomically independent Genotypes #2–17 fungi as *O. sinensis* on the basis of the taxonomic annotations used in several public depositories, including GenBank and AlphaFold, pending the formal determination and differentiation of the systematic taxonomic positions of these GC- and AT-biased Genotypes #2–17. We also maintain the long-standing use of the traditional name *C. sinensis* for both wild and cultivated insect–fungal complexes because the 2007 renaming of *C. sinensis* to *O. sinensis*, which is based on *H. sinensis* strain EFCC7287 as the nomenclatural reference, did not address the long-standing, indiscriminate application of the Latin name to the insect–fungal complex [9,10,48,51,54]. Nevertheless, this imperfect and temporarily necessary practice will likely be revised in the future through the differential use of proprietary and exclusive Latin names for the multiple genome-independent genotypes of *O. sinensis* fungi and the insect–fungal complex as a unique medicinal entity.

The sexual reproductive behavior of ascomycetes is controlled by transcription factors encoded at the mating-type (*MAT*) locus, which constitute the key regulatory mechanism governing mating type recognition and compatibility and the development of fruiting bodies and sexual reproductive structures and cells such as ascocarps and ascospores [55−62]. The reproductive behavior of *O. sinensis* depends on the synergistic interplay of 2 core mating proteins, *MAT1-1* and *MAT1-2* idiomorphs. The MAT1-1-1 protein harbors a mating-type alpha high mobility group box (MATα_HMGbox) domain, whereas the MAT1-2-1 protein contains the high mobility group box ROX1-like (HMG-box_ROX1-like) domain [14,58,63−65]. These 2 functional domains of mating proteins play essential roles in regulating the expression of genes associated with sexual reproduction. Li *et al.* [65−68] documented the differential occurrence, differential transcription, and alternative splicing of mating-type and pheromone receptor genes in the genomes of *H. sinensis* strains, challenging the prior self-fertilization hypothesis for *H. sinensis via* homothallism or pseudohomothallism. Instead, these discoveries support self-sterility consistent with a heterothallic or hybrid lifestyle, indicating the necessity of compatible sexual partners to complete sexual reproduction within the *C. sinensis* insect–fungus complex. Moreover, Li *et al.* [65,69] reported the heteromorphic tertiary structures of the DNA-binding domains of the full-length MAT1-1-

110 1 and MAT1-2-1 proteins in wild-type *C. sinensis* isolates, which might be generated by either GC-
111 biased Genotype #1 (*H. sinensis*) or Genotype #3 of *O. sinensis*, further underscoring the complex
112 heterothallic or hybrid sexual reproduction of *O. sinensis*.

113 Variations in the DNA-binding domains of the 2 mating proteins, arising from diverse amino acid
114 substitutions, induce changes in their tertiary structures and synergistic interactions between the 2
115 idiomorphic mating proteins. These proteins are derived from numerous mycologically impure, wild-
116 type *C. sinensis* isolates [65], which were characterized by Prof. Zhang Y-J and his associates
117 [7,51,70−72]. The present study further analyzed differential N- and/or C-terminal truncations and
118 diverse amino acid substitutions of the mating proteins derived from 20 purportedly "pure" *O. sinensis*
119 strains reported by Prof. Yao Y-J and his collaborators [14,36] to clarify the protein structural basis of *O.*
120 *sinensis* sexual reproduction. The changes in the primary, secondary, and tertiary structures of the
121 MATα_HMGbox and the HMG-box_ROX1-like domains of the MAT1-1-1 and MAT1-2-1 proteins
122 were correlated. Critically, the present study focused on the differential pairings of MAT1-1-1 and
123 MAT1-2-1 protein variants generated simultaneously by 20 *O. sinensis* strains to explore their possible
124 heterospecific fungal origins within these heterogeneous but purportedly "pure" *O. sinensis* strains.

125 # 2. Materials and Methods

126 ## 2.1 MAT1-1-1 and MAT1-2-1 protein sequences from *O. sinensis* strains

127 The AlphaFold database provides AI-predicted 3D structural morphs for 20 truncated MAT1-1-1
128 proteins (AGW27517–AGW27536) and 5 truncated MAT1-2-1 proteins (AGW27543, AGW27548,
129 AGW27552, AGW27554, and AGW27555) derived from 20 *O. sinensis* strains [14]. Li *et al.* [36]
130 reported that 11 of the 20 strains were obtained from fresh caterpillar body samples collected from
131 various *C. sinensis* production regions on the Qinghai–Tibet Plateau, marked as "TS" in Table 1. The
132 remaining 9 *O. sinensis* strains were from the cultures of monoascospores of *C. sinensis* insect–fungal
133 complexes that were collected from a single production region (Maqên, Guoluo, Qinghai Provence of
134 China), which are marked as "SS" in Table 1.

135 The MAT1-1-1 protein sequences of different lengths were encoded by the *MAT1-1-1* cDNAs that
136 were amplified from total RNA extracted from the *O. sinensis* strains using the primer pair *m1F3/m1R3*,
137 while the *MAT1-2-1* cDNAs were amplified using the primer pair *Mat1-2F*/*Mat1-2R* (Figures S1–S2)
138 [14]. Li *et al.* [36] deposited the internal transcribed spacer (ITS) sequences into GenBank for 15 of the
139 20 *O. sinensis* strains, including 10 homogeneous strains shown in brown in Table 1 that reportedly

harbor only the GC-biased Group-A ITS sequences and 5 other strains shown in red in Table 1 that were characterized as heterogeneous, containing both GC-biased Group-A and AT-biased Group-B ITS sequences. The remaining 5 *O. sinensis* strains shown in black in Table 1 lack ITS sequence information in GenBank.

Table 1. *O. sinensis* strains and their tissue sources [36], GenBank accession numbers for the internal transcribed spacer (ITS) nucleic acid sequences and mating proteins, and AlphaFold UniProt codes for the truncated and full-length MAT1-1-1 and MAT1-2-1 proteins derived from the *O. sinensis* strains.

***O. sinensis* strain code**	**Source of strain**	**GenBank accession number**				**AlphaFold UniProt code for the mating protein**	
		ITS nucleotide sequence		**Protein sequence**			
		GC-biased Group-A	**AT-biased Group-B**	**MAT1-1-1**	**MAT1-2-1**	**MAT1-1-1**	**MAT1-2-1**
CS71-1219	SS	JQ900151	—	AGW27534	AGW27554	U3NE87 *	U3NEA9
CS91-1291	TS	JQ900166	—	AGW27523	AGW27543	U3N9T9 ⁋	U3N9W0
CS68-2-1229	SS	JQ900158	—	AGW27528	AGW27548	U3N9T9 ⁋	U3N9W5 ⁑
CS37-295	TS	JQ900169	—	AGW27519	AGW27539	U3N9T9 ⁋	D7F2E9 ♦
CS18-266	TS	JQ900168	—	AGW27520	AGW27540	U3N9T9 ⁋	—
CS560-961	TS	JQ900143	—	AGW27522	AGW27542	U3N6U0	D7F2E3
CS71-1218	SS	JQ900150	—	AGW27533	AGW27553	U3N6U8 †	U3N9X0
CS76-1284	TS	JQ900159	—	AGW27525	AGW27545	U3N919	—
CS561-964	TS	JQ900144	—	AGW27526	AGW27546	U3N7G5	—
CS25-273	TS	JQ900167	—	AGW27527	AGW27547	U3N6U4	—
CS71-1220	SS	JQ900152	JQ900175	AGW27535	AGW27555	U3N6U8 †	U3N9W5 ⁑
CS6-251	TS	JQ900163	JQ900180	AGW27517	AGW27537	U3N9T9 ⁋	U3N6V5
CS34-291	TS	JQ900162	JQ900179	AGW27524	AGW27544	U3N9T9 ⁋	—
CS70-1208	SS	JQ900146	JQ900172	AGW27529	AGW27549	U3NE79 ‡	—
CS68-2-1228	SS	JQ900157	JQ900178	AGW27536	AGW27556	U3N7H7	—
CS68-5-1216	SS	—	—	AGW27532	AGW27552	U3N6U8 †	U3N6W6
CS26-277	TS	—	—	AGW27521	AGW27541	U3N9T9 ⁋	D7F2E9 ♦
CS36-1294	TS	—	—	AGW27518	AGW27538	U3N9T9 ⁋	D7F2E9 ♦
CS70-1211	SS	—	—	AGW27530	AGW27550	U3NE79 ‡	—
CS70-1212	SS	—	—	AGW27531	AGW27551	U3NE87 *	—

Note: The *O. sinensis* strains shown in brown have ITS sequences of Group A; those shown in red have ITS sequences for both Groups A and B; and those shown in black do not have ITS sequences in GenBank [14,36]. "TS" indicates that the *O. sinensis* strains were isolated from the caterpillar body specimens. "SS" indicates that the *O. sinensis* strains were obtained from the cultures of monoascospores that were collected from mature *C. sinensis* insect–fungal complexes. The ITS sequences in dark blue belong to GC-biased Group A (Genotype #1) of *O. sinensis,* and those in light blue belong to AT-biased Group B (Genotype #17) of *O. sinensis* (Tables S1–S2) [9,10,36]. The GenBank protein accession numbers in green refer to the full-length MAT1-2-1 protein sequences, while the remaining protein accession numbers in pink refer to the truncated mating proteins. AlphaFold codes with the same suffix symbol (*, ⁋, †, ‡, ⁑, or ♦) belong to the same morphs of AlphaFold-predicted 3D structures. Those without a suffix symbol indicate unique 3D structural morphs. "—" indicates no recodes of ITS sequences or 3D structural models in the GenBank and AlphaFold databases.

2.2 Alignment of the DNA-binding domain sequences of the mating proteins

The amino acid sequences of the MATα_HMGbox domains of the MAT1-1-1 proteins and the HMG-box_ROX1-like domains of the MAT1-2-1 proteins from the *O. sinensis* strains were aligned using the GenBank Blastp program (https://blast.ncbi.nlm.nih.gov/ (Bethesda, MD, USA), accessed from 18 June 2025 to 31 January 2026).

2.3 Amino acid properties and scale analysis

The amino acid sequences of the DNA-binding domains of the mating proteins were scaled based on the general properties of their side chains (https://web.expasy.org/protscale/; Basel, Switzerland, accessed from 18 May 2025 to 20 October 2025) (Table 3) [65,68,69,73−77]. The MATα_HMGbox domain of MAT1-1 proteins (amino acids 51→225 of the reference sequence AGW27560 derived from the *H. sinensis* strain CS68-2-1229 [14]) and the HMG-box_ROX1-like domain of MAT1-2-1 proteins (amino acids 127→197 of the reference sequence AEH27625 [70]), with additional 9 amino acid residues extending upstream and downstream of the domains to ensure comprehensive coverage, were plotted sequentially with a window size of 9 amino acid residues using the linear weight variation model of the ExPASy ProtScale algorithms [65,68,69,73−75] to generate ExPASy ProtScale plots and predict the potential changes in the hydrophobic properties and secondary structures of proteins related to α-helices, β-sheets, β-turns, and coils. The resulting topological structures and waveforms shown in the ProtScale plots were compared to exploring changes in the hydrophobicity and 2D structural features of the DNA-binding domains of the mating proteins.

2.4 AlphaFold-based prediction of the tertiary structures of the mating proteins

To analyze the heteromorphic stereostructures in this study, the 3D structures of the MAT1-1-1 and MAT1-2-1 proteins from *O. sinensis* strains were predicted computationally from their amino acid sequences utilizing artificial intelligence (AI)-driven machine learning technology AlphaFold (https://alphafold.com/, Cambridgeshire, UK; accessed from 18 June 2025 to 20 October 2025) [78−88].

The AlphaFold database provides per-residue model confidence, predicts scores between 0 and 100 using the partial local distance difference test (pLDDT), and assigns a per-residue score to each individual residue [78−81,83−85]. A color-coding scheme based on the pLDDT score is used for the 3D structure to visually represent the confidence levels: residues with very high confidence (pLDDT>90) are shown in dark blue, those with high confidence (90>pLDDT>70) appear in light blue, residues with low confidence (70>pLDDT>50) are shown in yellow, and residues with very low confidence (pLDDT<50) are shown in orange [65,86,87,89].

The main reasons for the low confidence in the AlphaFold predictions include the following: (1) The supporting data are insufficient, as the AlphaFold-based prediction relies on the quality of the multiple sequence alignment (MSA). If few homologous sequences are available in a certain region of the MSA, the model will lack sufficient evolutionary information to infer the structure, resulting in low confidence. (2) Some protein regions are inherently flexible under physiological conditions, such as the activation domains of transcription factors, which do not have a fixed 3D structure, leading to low confidence in the predicted structure assigned by AlphaFold. (3) In terms of special structural regions, such as small-molecule binding sites and artificial linkers in fused proteins, the AlphaFold system may have limited predictive power, leading to insufficient confidence in the predictions of complex structures. Thus, the low-confidence regions in AlphaFold predictions are informative flags for highlighting the challenges of prediction and should not be interpreted with excessive confidence.

2.5 Correlations of the primary, secondary, and tertiary structures of the DNA-binding domains of the paired MAT1-1-1 and MAT1-2-1 proteins

Among diverse AlphaFold 3D structural heteromorphs, the changes in the tertiary structure around the mutation sites within the MATα_HMGbox domains of MAT1-1-1 proteins or the HMG-box_ROX1-like domains of MAT1-2-1 proteins were locally magnified to visualize the subtle 3D variations. Primary sequence variations in the DNA-binding domains of the mating proteins, such as peptide chain truncations and amino acid replacements, were correlated with topological structural and waveform changes in hydrophobicity, α-helices, β-sheets, β-turns, and coils, as presented in the ExPASy ProtScale plots, and with the locally magnified 3D structures around the variation sites. The variant MAT1-1-1 and MAT1-2-1 proteins produced by the *O. sinensis* strains were paired to explore their diverse fungal origins.

3. Results

3.1 Primary structures of the truncated MAT1-1-1 proteins in *O. sinensis* strains

In the alignment with the sequence of the representative authentic full-length MAT1-1-1 protein AGW27560 derived from the *H. sinensis* strain CS68-2-1229 [14], the 20 MAT1-1-1 proteins were differentially truncated at both the N- and C-termini (Figure 1). These truncated proteins of varying lengths were encoded by the cDNAs amplified using the same pair of primers (the forward primer *m1F3* and the reverse complementary primer *m1R3* shown in Figure S1) from the *MAT1-1-1* genes of *O. sinensis* strains. The MATα_HMGbox domains of the authentic full-length MAT1-1-1 proteins represented by the protein AGW27560 are located at amino acids 51→225 (shown in pink in Figures 1

and S1). The MATα_HMGbox domains of the 20 MAT1-1-1 proteins are truncated at the N-termini and correspond to 9 distinct AlphaFold 3D structural morphs (Figure 2, Table 1).

```
AGW27560 1   MTTRNEVMQRLSSVRADVLLNFLTDDAIFQLASRYHESTTEADVLTPVSTAAASRATRQTKE  62
AGW27528
AGW27535
AGW27530
AGW27534
AGW27522
AGW27525
AGW27526
AGW27527
AGW27536

AGW27560 63  ASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTTLWHKDPFRNKWALIAKVYSFVRD  123
AGW27528 1   ------------------------------------------------------------  60
AGW27535 1                                      --------------------------  26
AGW27530 1                                    V--------------------------  27
AGW27534 1                                                  -------------  13
AGW27522 1   ------------------------------------------------------------  60
AGW27525 1   ------------------------------------------------------------  60
AGW27526 1   -SSFT-------------------------------------------------------  60
AGW27527 1                  ---------------------------------------------  46
AGW27536 1                                  E-- --------------------------  28

AGW27560 124 QIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQDDDAGSQKLFQDESSANLDQSSL  183
AGW27528 61  ------------------------------------------------------------  120
AGW27535 27  ------------------------------------------------------------  86
AGW27530 28  ------------------------------------------------------------  87
AGW27534 14  ------------------------------------------------------------  73
AGW27522 61  ------------------------------------------------------------  120
AGW27525 61  ------------------------------------------------------------  120
AGW27526 61  ------------------------------------------------------------  120
AGW27527 47  ------------------------------------------------------------  106
AGW27536 29  ------------------------------------------------------------  88

AGW27560 184 LSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSHSGIMAPRAANCTPPVSYTKEKID  243
AGW27528 121 ----------------------------------------------------------  180
AGW27535 87  ----------------------------------------------------------  146
AGW27530 88  ----------------------------------------------------------  147
AGW27534 74  ----------------------------------------------------------  133
AGW27522 121 ----------------------------------------------------------  180
AGW27525 121 ----------------------------------------------------------  180
AGW27526 121 ----------------------------------------------------------  180
AGW27527 107 ----------------------------------------------------------  166
AGW27536 69  ----------------------------------------------------------  148

AGW27560 244 FINTIRSDPVQATKEILGDCYDETTIKLLGVKSHNVESVDSITHLSMQREYQAPRFFYDY  304
AGW27528 181 ------------------------------------------------------------  240
AGW27535 147 ------------------------------------------------------------  206
AGW27530 148 ------------------------------------------------------------  207
AGW27534 134 ------------------------------------------------------------  193
AGW27522 181 ------------------------------------------------------------  240
AGW27525 181 ------------------------------------------------------------  240
AGW27526 181 ------------------------------------------------------------  240
AGW27527 167 ------------------------------------------------------------  226
AGW27536 149 ------------------------------------------------------------  208

AGW27560 305 SVSYAGMDFGGSNEPVMNLNNLPENETFDIDSPFDLDKILGQSQSEGERTSHLPPSPPHN  364
AGW27528 241 ------------------------------------------------------------  300
AGW27535 207 ------------------------------------------------------------  266
AGW27530 208 ------------------------------------------------------------  267
AGW27534 194 ------------------------------------------------------------  254
AGW27522 241 ------------------------------------------------------------L--  300
AGW27525 241 ---------------------------------------------------------   299
AGW27526 241 -------------------------------------------------------     296
AGW27527 227 ---------------------------------------------------         277
AGW27536 209 ------------------------------------------------------------  268

AGW27560 365 PLDFYFAF                                                      372
AGW27528 301 -                                                             301
AGW27535 267 -                                                             267
AGW27530 268 -                                                             268
AGW27534 254
AGW27522 301 -                                                             301
AGW27525 299
AGW27526 296
AGW27527 277
AGW27536 269 -                                                             269
```

Figure 1. Alignment of the sequences of the query MAT1-1-1 protein AGW27560 derived from *O. sinensis* strain CS68-2-1229 and the 9 truncated MAT1-1-1 proteins (AGW27528, AGW27522, AGW27525, AGW27526, AGW27527, AGW27535, AGW27530, AGW27536, and AGW27534 [14]) that are truncated at both the N- and C-termini. These truncated proteins represent the 20 MAT1-1-1 proteins under 9 AlphaFold 3D structural models (U3N9T9, U3N6U0, U3N919, U3N7G5, U3N6U4, U3N6U8, U3NE79, U3N7H7, and U3NE87). The MATα_HMGbox domain (amino acids 51→225 in the AGW27560 sequence) is shown in pink, and those outside the domain are shown in blue. The amino acid substitutions are shown in green. The hyphens indicate identical amino acids, and the spaces denote unmatched sequence gaps.

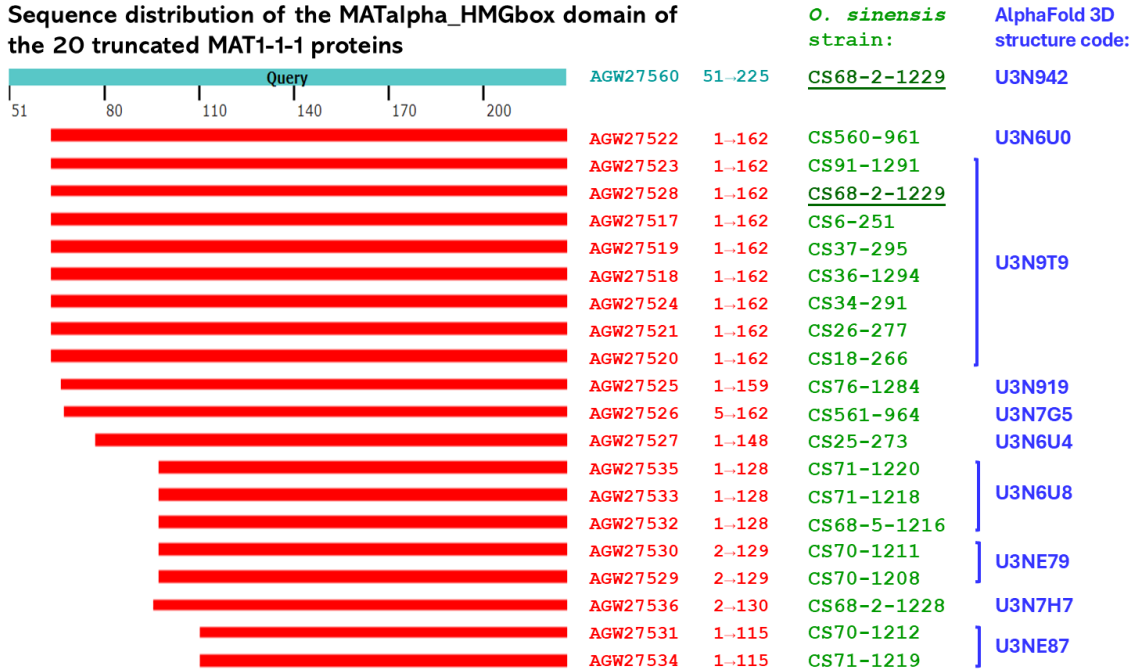


Figure 2. Sequence distribution of the N-terminally truncated MATα_HMGbox domains of 20 MAT1-1-1 proteins derived from *O. sinensis* strains (Table 1). The sequences are aligned with that of the full-length MATα_HMGbox domain (51→225) of the query MAT1-1-1 protein AGW27560 derived from the *H. sinensis*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

3.2 Primary structures of the truncated MAT1-2-1 proteins in *O. sinensis* strains

The AlphaFold database lists 5 MAT1-2-1 proteins produced by *O. sinensis* strains that are truncated differentially at their C-termini (Figure 3). These truncated MAT1-2-1 proteins of varying lengths are encoded by the *MAT1-2-1* cDNAs amplified using the same pair of primers (the forward primer *Mat1-2F* and the reverse complementary primer *Mat1-2R* shown in Figure S2 [14,70]) from the total RNA extracted from various *O. sinensis* strains. The full-length authentic MAT1-2-1 protein AEH27625 [70], under the AlphaFold code U3N942 derived from the *H. sinensis* strain CS2, is used as an alignment reference; the authentic MAT1-2-1 protein contains an HMG-box_ROX1-like domain at amino acids 127→197, as shown in pink in Figures 3 and S2. In alignment with the HMG-box_ROX1-like domain sequence of the reference protein AEH27625, the domain sequences of the MAT1-2-1 proteins derived from the 5 *O. sinensis* strains are truncated differentially at the C-termini and each contains a tyrosine-

to-histidine (Y-to-H) substitution, contributing to diverse 3D structures under 4 AlphaFold codes (Figure 4). The truncated MAT1-2-1 protein AGW27554 contains an additional glutamine-to-threonine (Q-to-T) substitution, as shown in Figure 3.

```
AEH27625 1   MANPINMIPNPQWNATDYEAIWKGLEAQVNPFSQILCLEGDFFRQLDDAAKLFIARKLME  60
AGW27543 1   ------------------------------------------------------------  60
AGW27554 1   ------------------------------------------------------------  60
AGW27555 1   ------------------------------------------------------------  60
AGW27548 1   ------------------------------------------------------------  60
AGW27552 1   ------------------------------------------------------------  60

AEH27625 61  HVQESVLYVNDGNGPDRVYLGAPRHFVVGGGMILQISGYAPYWIRRSVSKVVTATVLAPP  120
AGW27543 61  ------------------------------------------------------------  120
AGW27554 61  ------------------------------------------------------------  120
AGW27555 61  ------------------------------------------------------------  120
AGW27548 61  ------------------------------------------------------------  120
AGW27552 61  ------------------------------------------------------------  120

AEH27625 121 SPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYKDMS  180
AGW27543 121 -----------------------H------------------------------------  180
AGW27554 121 -----------------------H------------------------------------  180
AGW27555 121 -----------------------H-------------------------------     175
AGW27548 121 -----------------------H------------------------------      175
AGW27552 121 -----------------------H-----------------------------       174

AEH27625 181 QQVKQALLEKHPDYQYKPRRPCERRRRRRASPNQNPKQSTSRNAATRDAAISSEDTSTAT  240
AGW27543 181 -----                                                        185
AGW27554 181 ----T                                                         185
AGW27555 175
AGW27548 175
AGW27552 174

AEH27625 241 GDTNTANGF                                                     249
AGW27543 185
AGW27554 185
AGW27555 175
AGW27548 175
AGW27552 174
```

Figure 3. Alignment of the sequences of 5 truncated AMT1-2-1 proteins derived from the *O. sinensis* strains and the reference query of the full-length MAT1-2-1 protein AEH27625 from the *H. sinensis* strain CS2 under AlphaFold code D7F2E9. The amino acids in the HMG-box_ROX1-like domain (amino acids 127→197 in the AEH27625 sequence) are shown in pink, and those outside the domain are shown in blue. The amino acid substitutions are shown in green. The hyphens indicate identical amino acids, and the spaces denote unmatched sequence gaps.

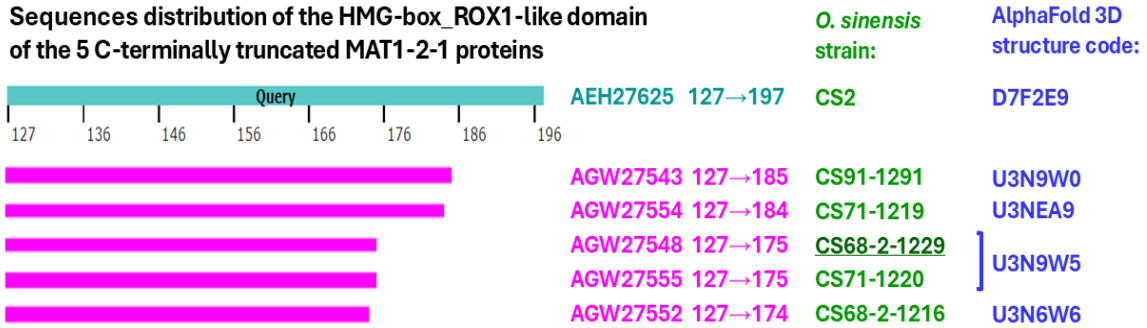


Figure 4. Sequence distribution of the C-terminally truncated HMG-box_ROX1-like domains of 5 MAT1-2-1 proteins derived from the *O. sinensis* strains (*cf.* Table 1). The sequences are aligned with the HMG-box_ROX1-like domain (127→197) of the full-length MAT1-2-1 protein AEH27625 from *H. sinensis* strain CS2.

3.3 Heteromorphic tertiary structures of the MATα_HMGbox domains of the truncated MAT1-1-1 proteins derived from *O. sinensis* strains

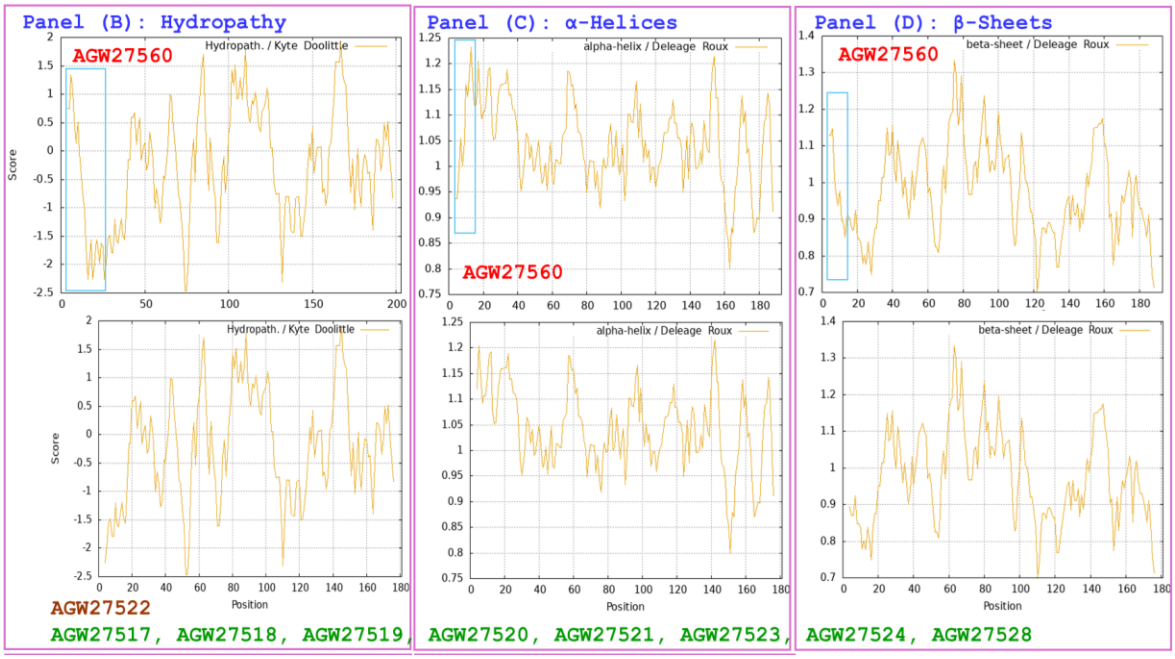
This section reports the results of the correlation analyses between changes in the hydrophobicity

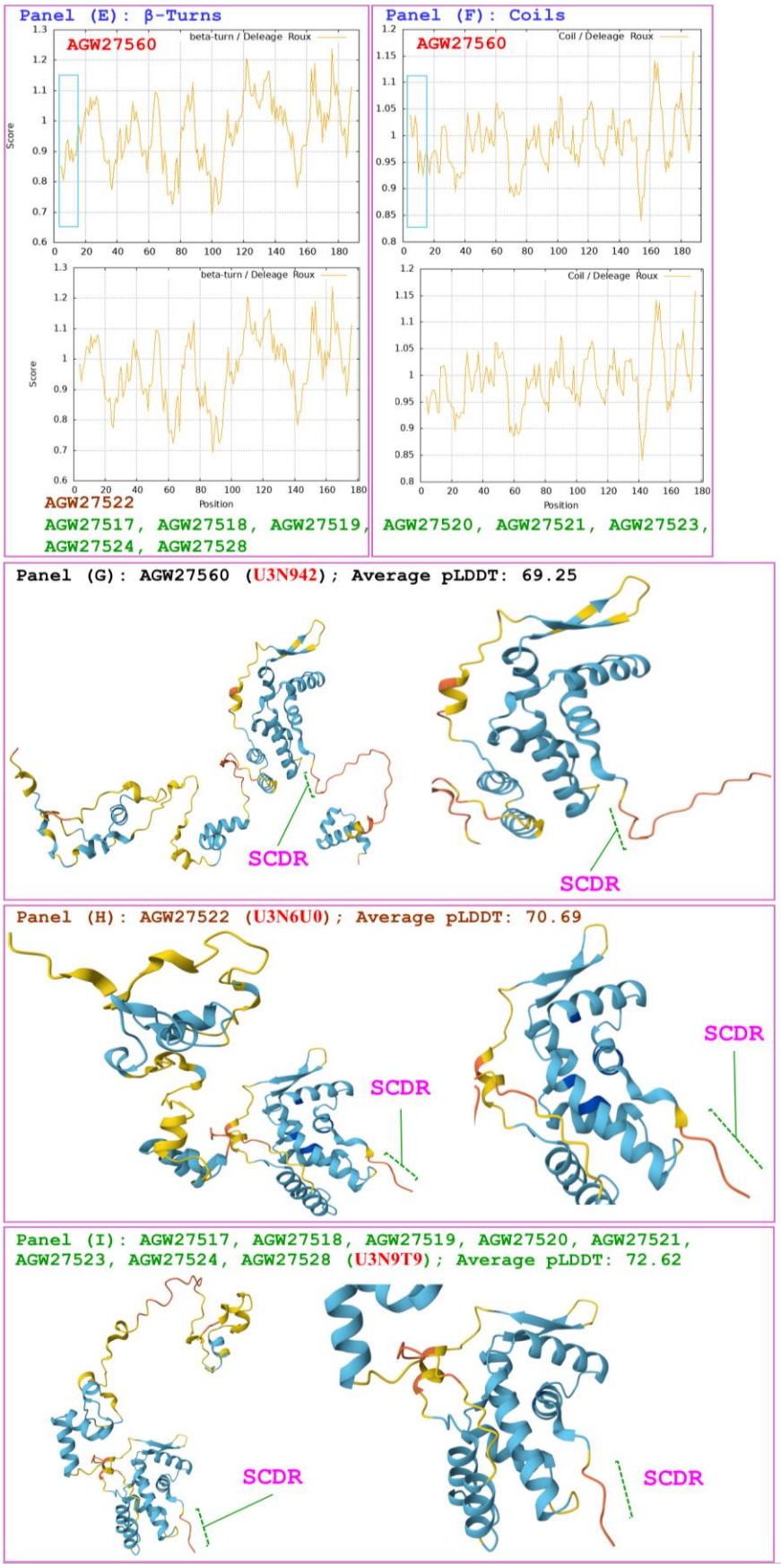
263 and structures of the MATα_HMGbox domains of truncated MAT1-1-1 proteins derived from various *O.*
264 *sinensis* strains, which correspond to 9 AlphaFold codes (Table 1, Figures 1−2), compared to the
265 reference MAT1-1-1 proteins. The reference protein, represented by AGW27560 (derived from the *O.*
266 *sinensis* strain CS68-2-1229; Table S4), is associated with the AlphaFold code U3N942 [14]. The
267 sequences of the MATα_HMGbox domain (amino acid residues 51→225) were extended by 9 amino
268 acid residues both upstream and downstream of the domains and were used in the sequential ExPASy
269 ProtScale plot to analyze hydropathy, α-helices, β-sheets, β-turns, and coils with a window size of 9
270 amino acid residues (*cf.* Section 2.3).

271 Figure 5 compares the hydrophobicity and structural characteristics of the MATα_HMGbox
272 domains between 2 groups of truncated MAT1-1-1 proteins: (1) the protein AGW27522 (under the
273 AlphaFold code U3N6U0) derived from the *O. sinensis* strain CS560-961; and (2) 8 proteins,
274 AGW27517, AGW27518, AGW27519, AGW27520, AGW27521, AGW27523, AGW27524, and
275 AGW27528 (all under the AlphaFold code U3N9T9), derived from *O. sinensis* strains CS6-251, CS36-
276 1294, CS37-295, CS18-266, CS26-277, CS91-1291, CS34-291, and CS68-2-1229, respectively (Table
277 1). The reference protein AGW27560 (under AlphaFold code U3N942) is derived from the *H. sinensis*
278 strain CS68-2-1229 [14]. Panel (A) of Figure 5 shows that the MATα_HMGbox domains of both groups
279 of subject proteins exhibit 13-residue truncations at N-termini of the DNA-binding domains. The
280 truncations result in changes in the topological structure and waveform of the ExPASy ProtScale plots,
281 as illustrated in Panels (B)–(F) for hydropathy, α-helices, β-sheets, β-turns, and coils, respectively. Panel
282 (G) shows the 3D structure of the full-length reference protein AGW27560, where the 3 α-helices within
283 the MATα_HMGbox domain form a core L-shaped stereostructure. Panels (H)–(I) illustrate that the N-
284 and C-terminal truncations (*cf.* Figure 1) significantly alter the 3D structures of the proteins under the
285 heteromorphic AlphaFold codes U3N6U0 and U3N9T9, respectively. The N-terminally truncated
286 regions of these variant proteins are located upstream of the core structure of 3 α-helices and may exert
287 a steric auxiliary effect on stabilizing the hydrophobic core structures of the MATα_HMGbox domains
288 [90,91].

Panel (A): Alignment of amino acid sequences of the MATα_HMGbox domains

```
AGW27560  42   ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTT  100
AGW27522  1                          ------------------------------------  37
AGW27517  1                          ------------------------------------  37
AGW27518  1                          ------------------------------------  37
AGW27519  1                          ------------------------------------  37
AGW27520  1                          ------------------------------------  37
AGW27521  1                          ------------------------------------  37
AGW27523  1                          ------------------------------------  37
AGW27524  1                          ------------------------------------  37
AGW27528  1                          ------------------------------------  37

AGW27560  101  LWHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQ  160
AGW27522  38   ------------------------------------------------------------  97
AGW27517  38   ------------------------------------------------------------  97
AGW27518  38   ------------------------------------------------------------  97
AGW27519  38   ------------------------------------------------------------  97
AGW27520  38   ------------------------------------------------------------  97
AGW27521  38   ------------------------------------------------------------  97
AGW27523  38   ------------------------------------------------------------  97
AGW27524  38   ------------------------------------------------------------  97
AGW27528  38   ------------------------------------------------------------  97

AGW27560  161  DDDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSH  220
AGW27522  98   ------------------------------------------------------------  157
AGW27517  98   ------------------------------------------------------------  157
AGW27518  98   ------------------------------------------------------------  157
AGW27519  98   ------------------------------------------------------------  157
AGW27520  98   ------------------------------------------------------------  157
AGW27521  98   ------------------------------------------------------------  157
AGW27523  98   ------------------------------------------------------------  157
AGW27524  98   ------------------------------------------------------------  157
AGW27528  98   ------------------------------------------------------------  157

AGW27560  221  SGIMAPRAANCTPP                                                234
AGW27522  158  --------------                                                171
AGW27517  158  --------------                                                171
AGW27518  158  --------------                                                171
AGW27519  158  --------------                                                171
AGW27520  158  --------------                                                171
AGW27521  158  --------------                                                171
AGW27523  158  --------------                                                171
AGW27524  158  --------------                                                171
AGW27528  158  --------------                                                171
```




Panel (E): β-Turns
AGW27560
beta-turn / Deleage Roux
Score
Position
AGW27522
AGW27517, AGW27518, AGW27519, AGW27524, AGW27528
Panel (F): Coils
AGW27560
Coil / Deleage Roux
Score
Position
AGW27520, AGW27521, AGW27523,
Panel (G): AGW27560 (U3N942); Average pLDDT: 69.25
SCDR
SCDR
Panel (H): AGW27522 (U3N6U0); Average pLDDT: 70.69
SCDR
SCDR
Panel (I): AGW27517, AGW27518, AGW27519, AGW27520, AGW27521, AGW27523, AGW27524, AGW27528 (U3N9T9); Average pLDDT: 72.62
SCDR
SCDR

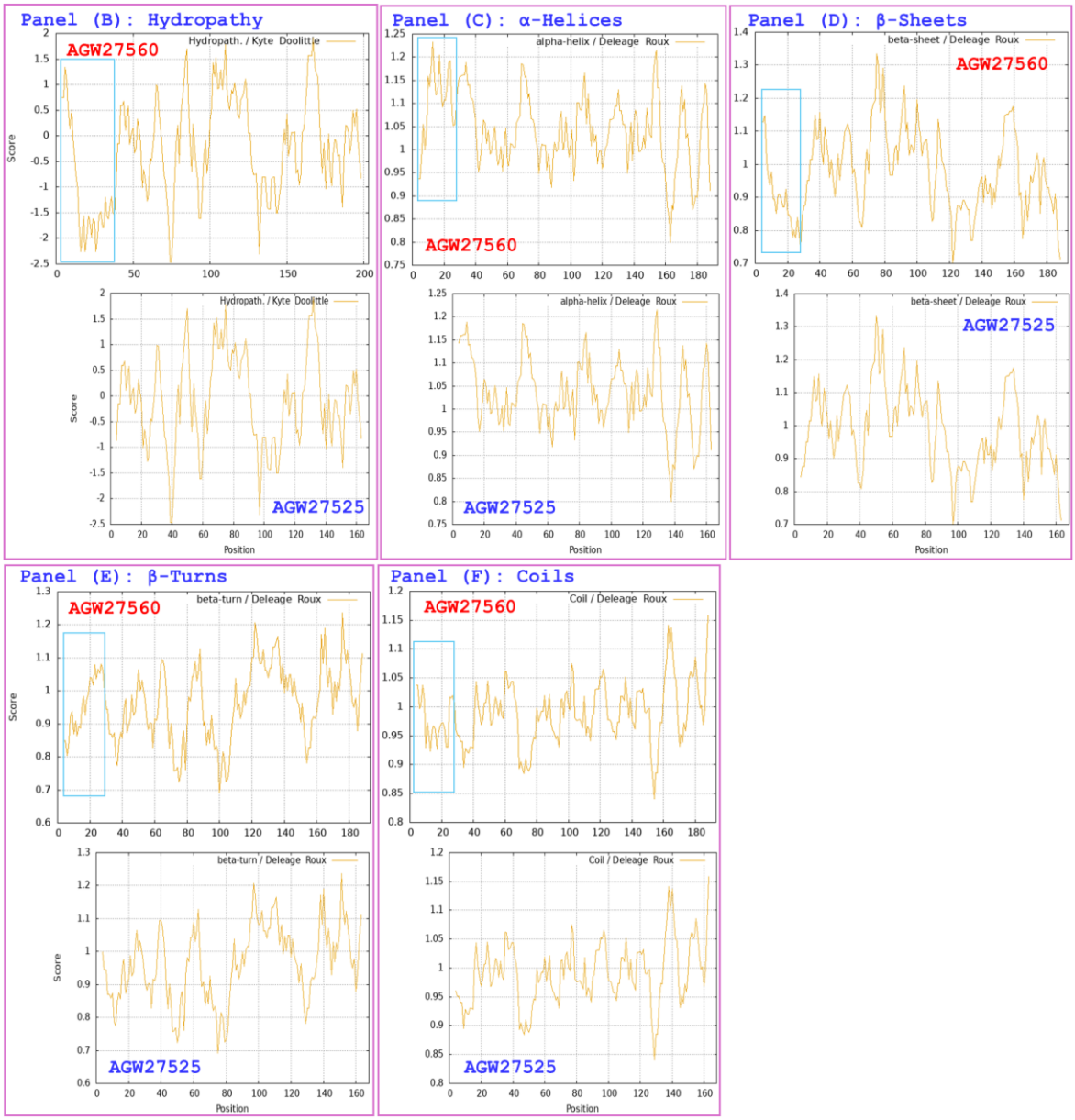
291

Figure 5. Correlations of the changes in the hydrophobicity and 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 *H. sinensis* strain CS68-2-1229 and the 2 groups of the truncated proteins: (1) AGW27522 shown in brown (under the AlphaFold code U3N6U0) derived from the *O. sinensis* strain CS560-961 and (2) the truncated proteins shown in green (under the AlphaFold code U3N9T9) derived from the *O. sinensis* strains CS6-251, CS36-1294, CS37-295, CS18-266, CS26-277, CS91-1291, CS34-291, and CS68-2-1229. Panel (A) shows an alignment of the amino acid sequences of the MATα_HMGbox domains of the MAT1-1-1 proteins, where the hyphens indicate identical amino acid residues and the spaces denote unmatched sequence gaps. The ExPASy ProtScale plots show the changes in hydrophobicity and the 2D structures in Panels (B)–(F) for hydropathy, α-helices, β-sheets, β-turns, and coils of the proteins, respectively; the open rectangles in blue highlight the N-terminally truncated region in the ExPASy plots. Panels (G)–(I) show the 3D structures of the full-length proteins on the left, and the locally magnified structures surrounding the truncated region are shown on the right. The model confidence for the AlphaFold-predicted 3D structures is as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

The hydrophobicity and structural characteristics of the MATα_HMGbox domain of the truncated MAT1-1-1 protein AGW27525 are compared with those of the reference full-length protein AGW27560 in Figure 6. The protein AGW27525 (under AlphaFold code U3N919) is derived from the *O. sinensis* strain CS76-1284 (Table 1), while the reference protein AGW27560 (under AlphaFold code U3N942) is derived from the *H. sinensis* strain CS68-2-1229 [14]. Panel (A) of Figure 6 shows a 16-residue truncation at the N-terminus of the MATα_HMGbox domain of the protein AGW27525. This truncation results in altered topological structures and waveforms in the ExPASy ProtScale plots, as illustrated in Panels (B)–(F) for hydropathy, α-helices, β-sheets, β-turns, and coils, respectively. Compared with the 3D structure of the reference protein AGW27560 shown in Panel (G), Panel (H) demonstrates that the truncated variant significantly alters the 3D structure of the protein AGW27525 under the AlphaFold code U3N919. Notably, the N-terminal amino acids 2–4 (AKR) shown in Panel (A) have hydropathy indices of 1.8, -3.9, and -4.5, respectively (Table S3) [73] and collectively exhibit hydrophilic characteristics. These amino acids initiate the formation of the β-sheet structure in the truncated protein AGW27525 shown in Panel (H); in contrast, the amino acids AKR in the full-length reference protein AGW27560 shown in Panel (G) are not part of the β-sheet structure or the hydrophobic core composed of the 3 critical α-helices within the MATα_HMGbox domain. The truncated region of the AGW27525 protein is located upstream of the 3 α-helix core within the MATα_HMGbox domain and may not directly participate in stabilizing the hydrophobic core of this domain [90,91].

Panel (A): Alignment of the amino acid sequences of the MATα_HMGbox domains

```
AGW27560  42   ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTT  100
AGW27525  1                                ---------------------------------  34

AGW27560  101  LWHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQ  160
AGW27525  35   ------------------------------------------------------------  94

AGW27560  161  DDDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSH  220
AGW27525  95   ------------------------------------------------------------  154

AGW27560  221  SGIMAPRAANCTPP                                               234
AGW27525  155  --------------                                               168
```



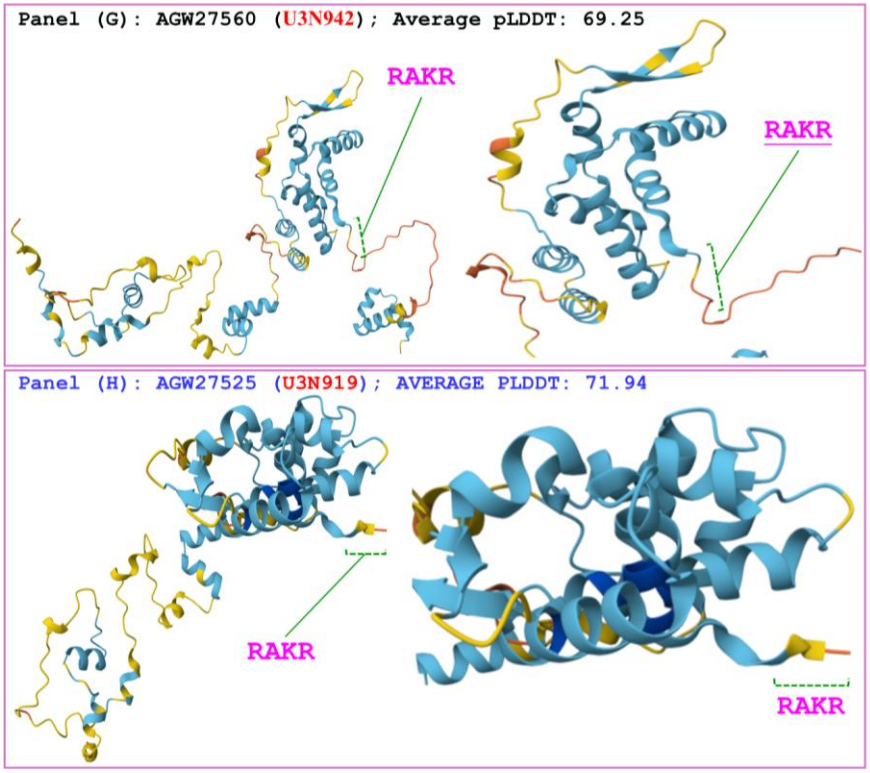


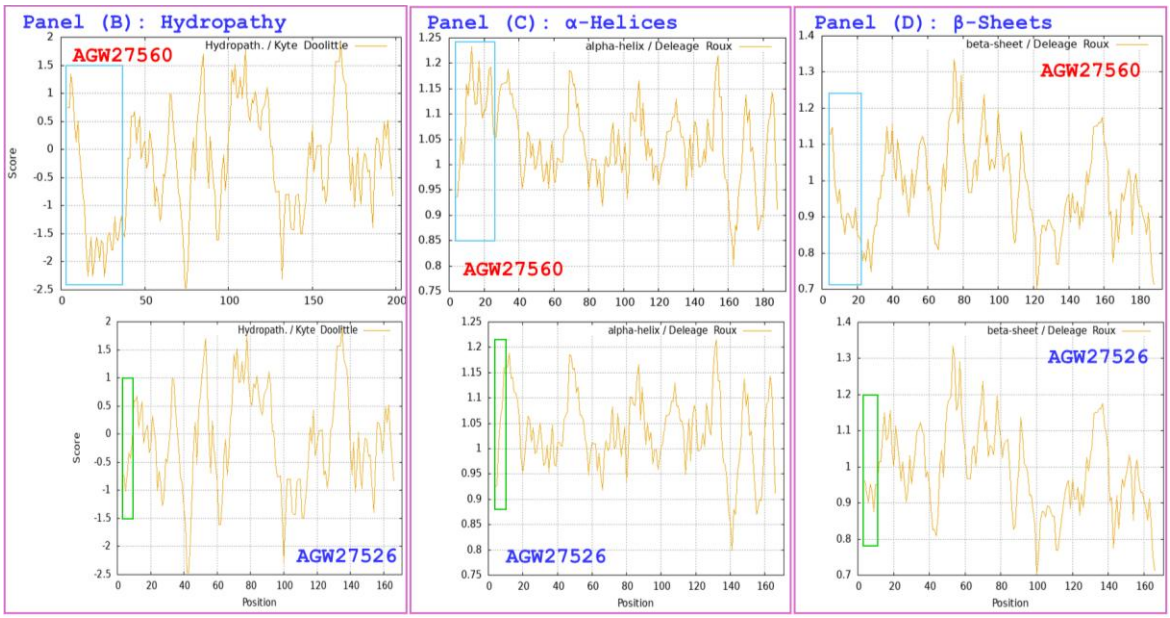
Figure 6. Correlations of the changes in hydrophobicity and the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the variant protein AGW27525 (under the AlphaFold code U3N919) derived from the *O. sinensis* strain CS76-1284. Panel (A) shows an alignment of the amino acid sequences of the MATα_HMGbox domains of the MAT1-1-1 proteins, where the hyphens indicate identical amino acid residues and the spaces denote unmatched sequence gaps. The ExPASy ProtScale plots show the changes in hydrophobicity and the 2D structure in Panels (B)–(F) for hydropathy, α-helices, β-sheets, β-turns, and coils of the protein, respectively; the open rectangles in blue highlight the N-terminally truncated region in the ExPASy plots. Panels (G)–(H) show the 3D structures of the full-length proteins on the left, and the locally magnified structures surrounding the variation region are shown on the right. The model confidence for the AlphaFold-predicted 3D structures is as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

The hydrophobicity and structural characteristics of the MATα_HMGbox domain of the variant MAT1-1-1 protein AGW27526 are compared with those of the full-length reference protein AGW27560 in Figure 7. The protein AGW27526 (under the AlphaFold code U3N7G5) is derived from the *O. sinensis* strain CS561-964 (Table 1), while the reference protein AGW27560 (under the AlphaFold code U3N942) is derived from the *H. sinensis* strain CS68-2-1229 [14]. Panel (A) of Figure 7 shows a 13-residue truncation at the N-terminus of the MATα_HMGbox domain of the protein AGW27526, along with 4 amino acid substitutions: C-to-S (cysteine to serine, both with polar neutral side chains; hydropathy indices changed from -3.5 to -0.8; *cf.* Table S3), D-to-S (acidic aspartic acid to serine with polar neutral side chains; hydropathy indices changed from 2.5 to -0.8), R-to-F (basic arginine to aromatic

phenylalanine; hydropathy indices changed from -4.5 to 2.8), and A-to-T (aliphatic alanine to threonine with polar neutral side chains; hydropathy indices changed from 1.8 to -0.7) (Table S3) [73]. The truncation and amino acid substitutions result in changes in the hydrophobicity and secondary structures of the variant protein, which are illustrated by the altered topological structure and waveform in the ExPASy ProtScale plots for hydropathy, α-helices, β-sheets, β-turns, and coils in Panels (B)–(F), respectively. Compared with the 3D structure of the reference protein AGW27560 shown in Panel (G), the truncation and replacement of amino acids in the variant protein AGW27526 shown in Panel (H) are located upstream of the 3 core α-helices of the MATα_HMGbox domain. These modifications caused apparent changes in the tertiary structure of the variant protein AGW27526 under the AlphaFold code U3N7G5.

Panel (A): Alignment of the amino acid sequences of the MATα_HMGbox domains

```
AGW27560  42   ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTT  100
AGW27526  1                            -SSFT--------------------------------  37

AGW27560  101  LWHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQ  160
AGW27526  38   ------------------------------------------------------------  97

AGW27560  161  DDDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSH  220
AGW27526  98   ------------------------------------------------------------  157

AGW27560  221  SGIMAPRAANCTPP                                               234
AGW27526  158  --------------                                               171
```



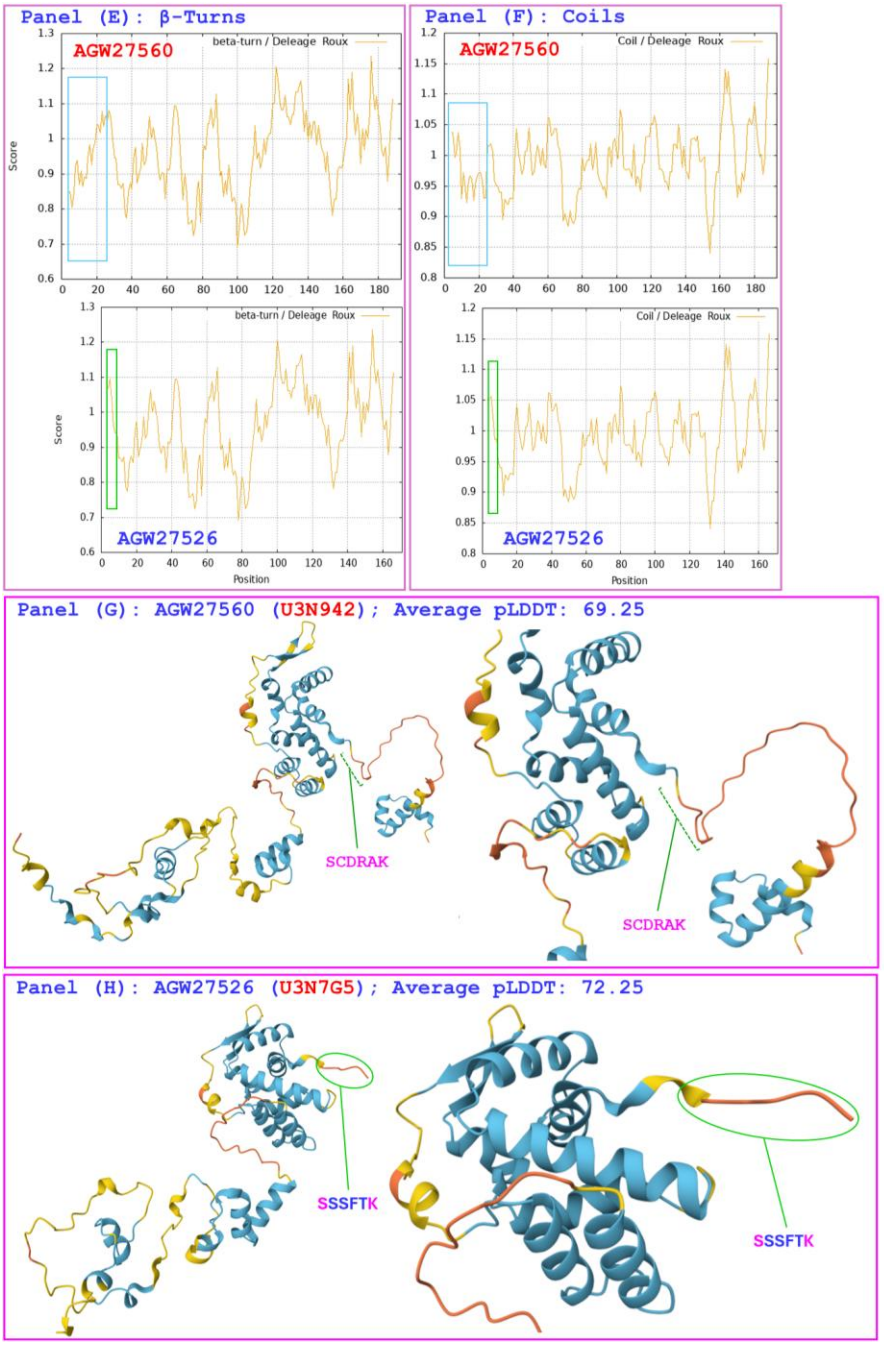


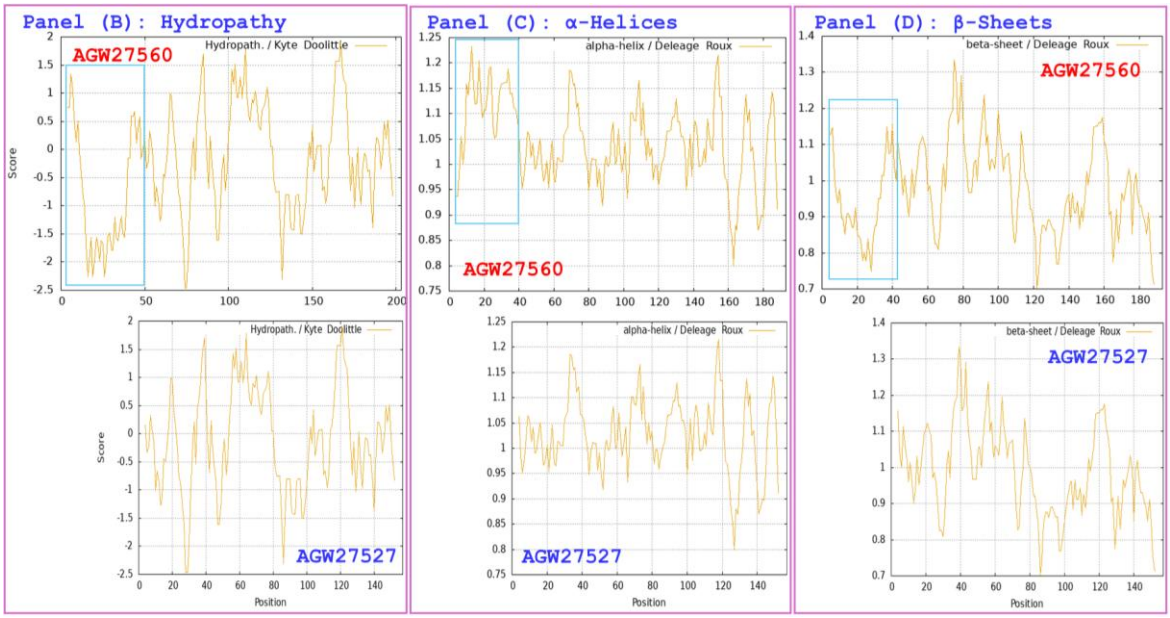
Figure 7. Correlations of the changes in the hydrophobicity and the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the variant protein AGW27526 (under the AlphaFold code U3N7G5) derived from the *O. sinensis* strain CS561-964. Panel (A) shows an alignment of the amino acid sequences of the MATα_HMGbox domains of the MAT1-1-1 proteins, where the hyphens indicate identical amino acid residues and the spaces denote unmatched sequence gaps. The ExPASy ProtScale plots show the changes in hydrophobicity and the 2D structure in Panels (B)–(F) for hydropathy, α-helices, β-sheets, β-turns, and coils of the protein, respectively; the open rectangles in blue highlight the N-terminally truncated regions, and those in green highlight the altered topological structure and waveform in the ExPASy plots. Panels (G)–(H) show the 3D structures of the full-length proteins on the left, and the locally magnified structures surrounding the regions of variation are shown on the right. The model confidence for the AlphaFold-predicted 3D structures is as follows:

very high (pLDDT>90); high (90>pLDDT>70); low (70>pLDDT>50); and very low (pLDDT<50).

The hydrophobicity and structural characteristics of the MATα_HMGbox domain of the variant MAT1-1-1 protein AGW27527 are compared with those of the reference protein AGW27560 in Figure 8. The protein AGW27527 (under the AlphaFold code U3N6U4) is derived from the *O. sinensis* strain CS25-273 (Table 1), while the reference protein AGW27560 (under the AlphaFold code U3N942) is derived from the *H. sinensis* strain CS68-2-1229 [14]. Panel (A) of Figure 8 shows a 27-residue truncation at the N-terminus of the MATα_HMGbox domain of the protein AGW27527. This truncation results in changes in the topological structure and waveform in the ExPASy ProtScale plots, which illustrate hydropathy, α-helices, β-sheets, β-turns, and coils in Panels (B)–(F), respectively. Compared with the 3D structure of the reference protein AGW27560 shown in Panel (G), Panel (H) shows that the truncation is located within the first of the 3 core α-helices of the MATα_HMGbox domain and that the tertiary structure of the truncated protein AGW27527 is apparently altered under the AlphaFold code U3N6U4.

Panel (A): Alignment of the amino acid sequences of the MATα_HMGbox domains

```
AGW27560  42   ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTT  100
AGW27527  1                                            ------------------------  23

AGW27560  101  LWHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQ  160
AGW27527  24   ------------------------------------------------------------  83

AGW27560  161  DDDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSH  220
AGW27527  84   ------------------------------------------------------------  143

AGW27560  221  SGIMAPRAANCTPP                                               234
AGW27527  144  --------------                                               157
```



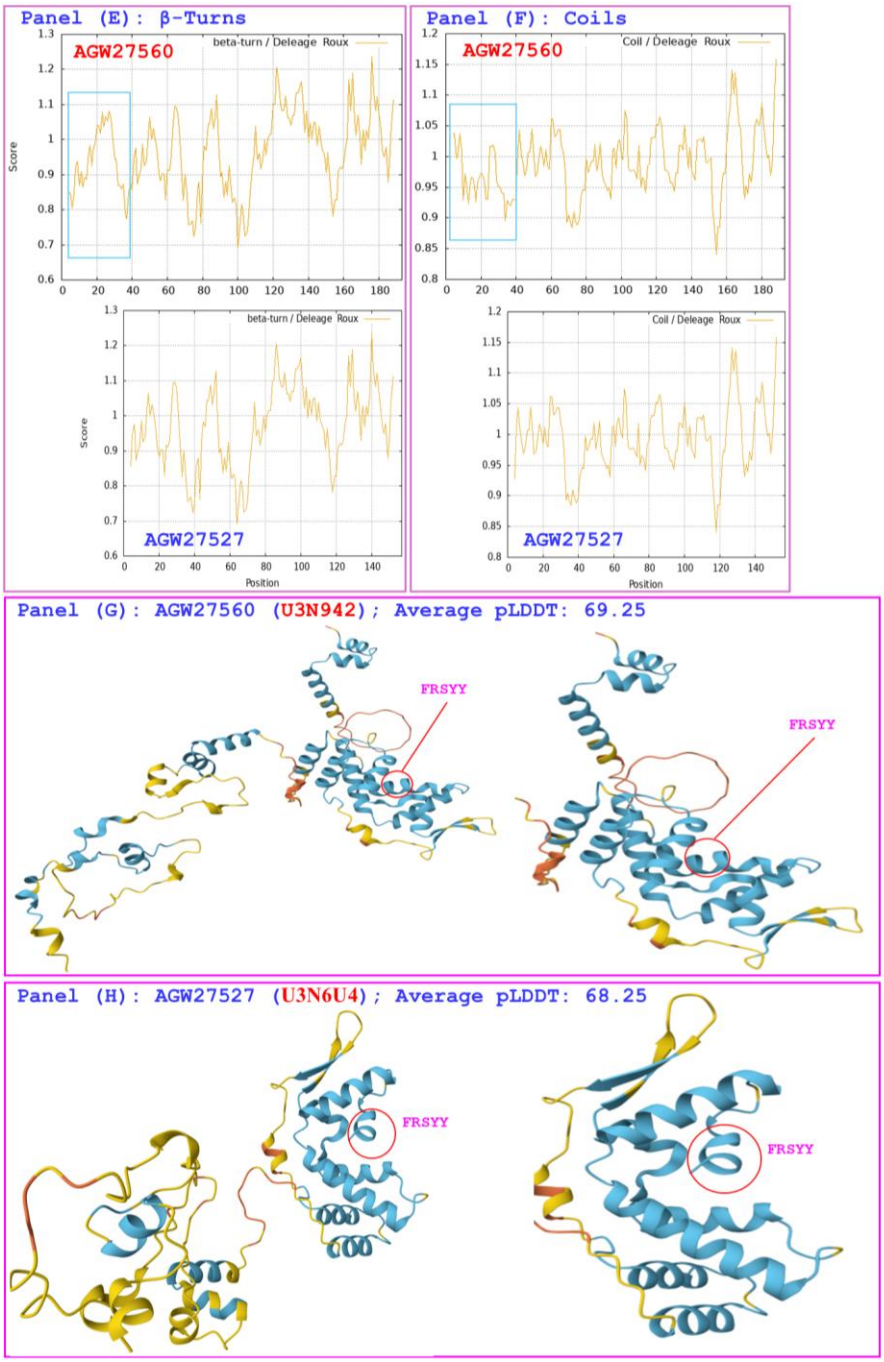


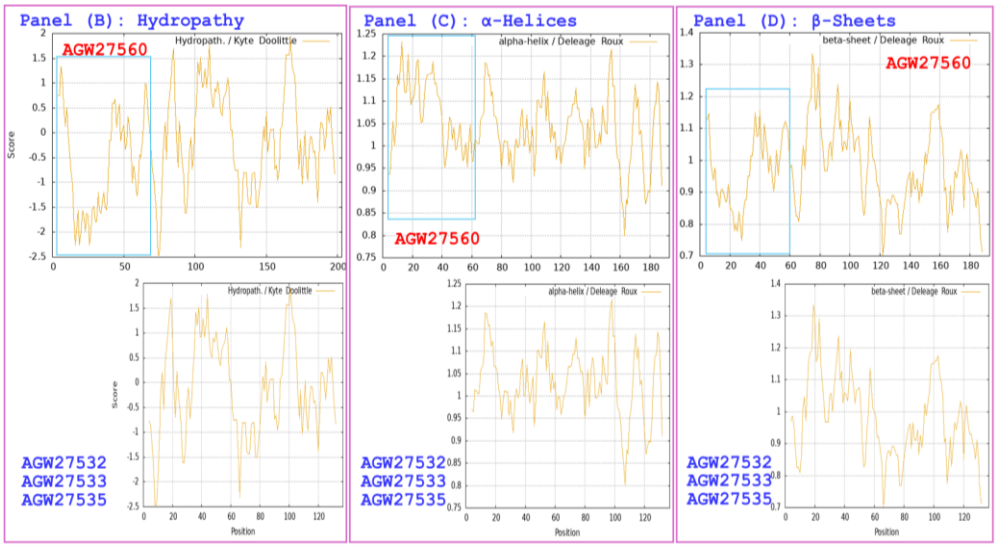
Figure 8. Correlations of the changes in the hydrophobicity and the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the variant protein AGW27527 (under the AlphaFold code U3N6U4) derived from the *O. sinensis* strain CS25-273. Panel (A) shows an alignment of the amino acid sequences of the MATα_HMGbox domains of the MAT1-1-1 proteins, where the hyphens indicate identical amino acid residues and the spaces denote unmatched sequence gaps. The ExPASy ProtScale plots) show the changes in hydrophobicity and the 2D structure in Panels (B)–(F for hydropathy, α-helices, β-sheets, β-turns, and coils of the protein, respectively; the open rectangles in blue highlight the N-terminal truncation regions in the ExPASy plots. Panels (G)–(H) show the 3D structures of the full-length proteins on the left, and the locally magnified structures surrounding the truncated region are shown on the right. The model confidence for the

AlphaFold-predicted 3D structures is as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

The hydrophobicity and structural characteristics of the MATα_HMGbox domain of the variant MAT1-1-1 proteins AGW27532, AGW27533, and AGW27535 are compared with those of the reference protein AGW27560 in Figure 9. These three variant proteins (all under the AlphaFold code U3N6U8) are derived from the *O. sinensis* strains CS68-5-1216, CS71-1218, and CS71-1220 (Table 1), whereas the reference protein AGW27560 (under the AlphaFold code U3N942) is derived from the *H. sinensis* strain CS68-2-1229 [14]. Panel (A) of Figure 9 shows 47-residue truncations at the N-termini of the MATα_HMGbox domains of the variant proteins. These truncations result in changes in the topological structure and waveform in the ExPASy ProtScale plots, which illustrate hydropathy, α-helices, β-sheets, β-turns, and coils in Panels (B)–(F), respectively. A comparison of the tertiary structures of the MAT1-1-1 proteins is illustrated in Panels (G)–(H). The truncations are located within the first of the 3 core α-helices of the MATα_HMGbox domains, which results in significant changes in the tertiary structures of the variant proteins AGW27532, AGW27533, and AGW27535 under the AlphaFold code U3N6U8.

Panel (A): Alignment of the amino acid sequences of the MATα_HMGbox domains

```
AGW27560  42   ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTT  100
AGW27532  1                                                              3
AGW27533  1                                                              3
AGW27535  1                                                              3

AGW27560  101  LWHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQ  160
AGW27532  4    ------------------------------------------------------------  63
AGW27533  4    ------------------------------------------------------------  63
AGW27535  4    ------------------------------------------------------------  63

AGW27560  161  DDDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSH  220
AGW27532  64   ------------------------------------------------------------  123
AGW27533  64   ------------------------------------------------------------  123
AGW27535  64   ------------------------------------------------------------  123

AGW27560  221  SGIMAPRAANCTPP                                               234
AGW27532  124  --------------                                               137
AGW27533  124  --------------                                               137
AGW27535  124  --------------                                               137
```



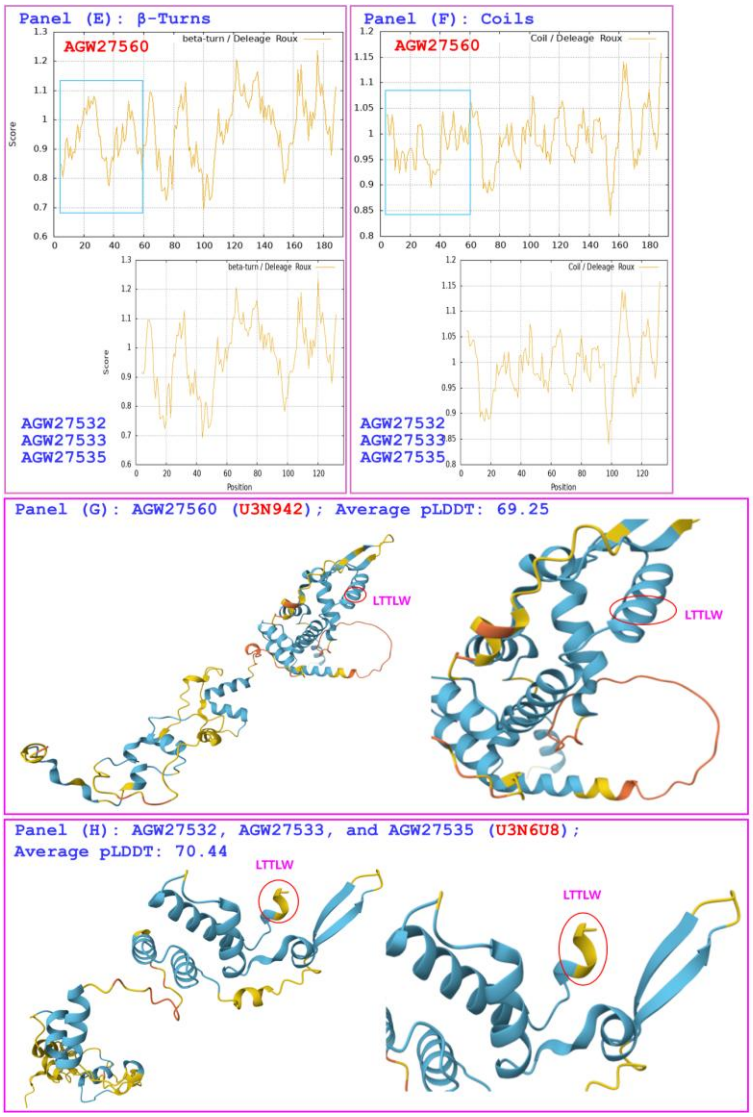


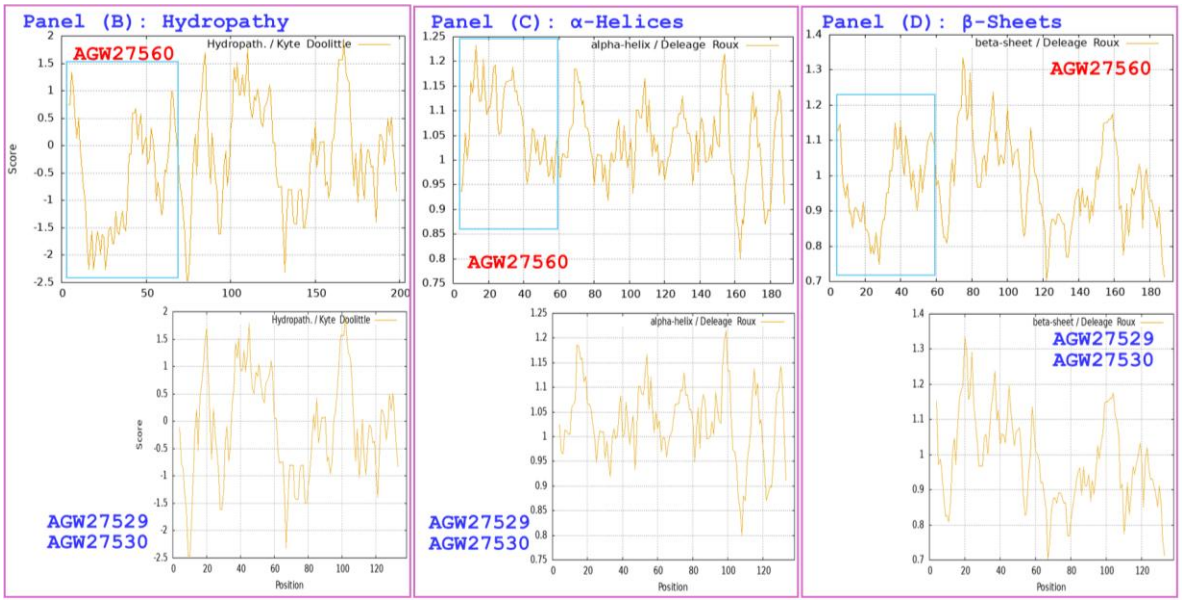
Figure 9. Correlations of the changes in the hydrophobicity and the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the variant proteins AGW27532, AGW27533, and AGW27535 (under the AlphaFold code U3N6U8) derived from the *O. sinensis* strains CS68-5-1216, CS71-1218, and CS71-1220, respectively. Panel (A) shows an alignment of the amino acid sequences of the MATα_HMGbox domains of the MAT1-1-1 proteins, where the hyphens indicate identical amino acid residues and the spaces denote unmatched sequence gaps. The ExPASy ProtScale plots show the changes in hydrophobicity and the 2D structure in Panels (B)–(F) for hydropathy, α-helices, β-sheets, β-turns, and coils of the protein, respectively; the open rectangles in blue highlight the N-terminally truncated regions in the ExPASy plots. Panels (G)–(I) show the 3D structures of the full-length proteins on the left and the locally magnified structures surrounding the truncation site on the right. The model confidence for the AlphaFold-predicted 3D structures is as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

The hydrophobicity and structural characteristics of the MATα_HMGbox domains of the variant MAT1-1-1 proteins AGW27529 and AGW27530 are compared with those of the reference protein AGW27560 in Figure 10. The two variant proteins (under the AlphaFold code U3NE79) are derived

from the *O. sinensis* strains CS70-1208 and CS70-1211 (Table 1), whereas the reference protein AGW27560 (under the AlphaFold code U3N942) is derived from the *H. sinensis* strain CS68-2-1229 [14]. Panel (A) of Figure 10 shows 46-residue truncations at the N-termini of the MATα_HMGbox domains of the proteins AGW27529 and AGW27530, along with F-to-V substitutions (aromatic phenylalanine to aliphatic valine; hydropathy indices changed from 2.8 to 4.2 (Table S3) [73] within the MATα_HMGbox domains, indicating increased hydrophobicity. The truncation and increased hydrophobicity result in an altered secondary structure, as illustrated by the changes in topological structure and waveform in the ExPASy ProtScale plots for hydropathy, α-helices, β-sheets, β-turns, and coils shown in Panels (B)–(F), respectively. Compared the 3D structures of the MAT1-1-1 proteins illustrated in Panels (G)–(H), the truncation and replacement of amino acids in the proteins AGW27529 and AGW27530 are located within the first of the 3 core α-helices of the MATα_HMGbox domains. These modifications result in the significantly altered tertiary structures of the proteins AGW27529 and AGW27530 corresponding to the AlphaFold code U3NE79.

Panel (A): Alignment of the amino acid sequences of the MATα_HMGbox domains

```
AGW27560  42   ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTT  100
AGW27529  1                                                          V---  4
AGW27530  1                                                          V---  4

AGW27560  101  LWHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQ  160
AGW27529  5    ------------------------------------------------------------  64
AGW27530  5    ------------------------------------------------------------  64

AGW27560  161  DDDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSH  220
AGW27529  65   ------------------------------------------------------------  124
AGW27530  65   ------------------------------------------------------------  124

AGW27560  221  SGIMAPRAANCTPP                                               234
AGW27529  125  --------------                                               138
AGW27530  125  --------------                                               138
```



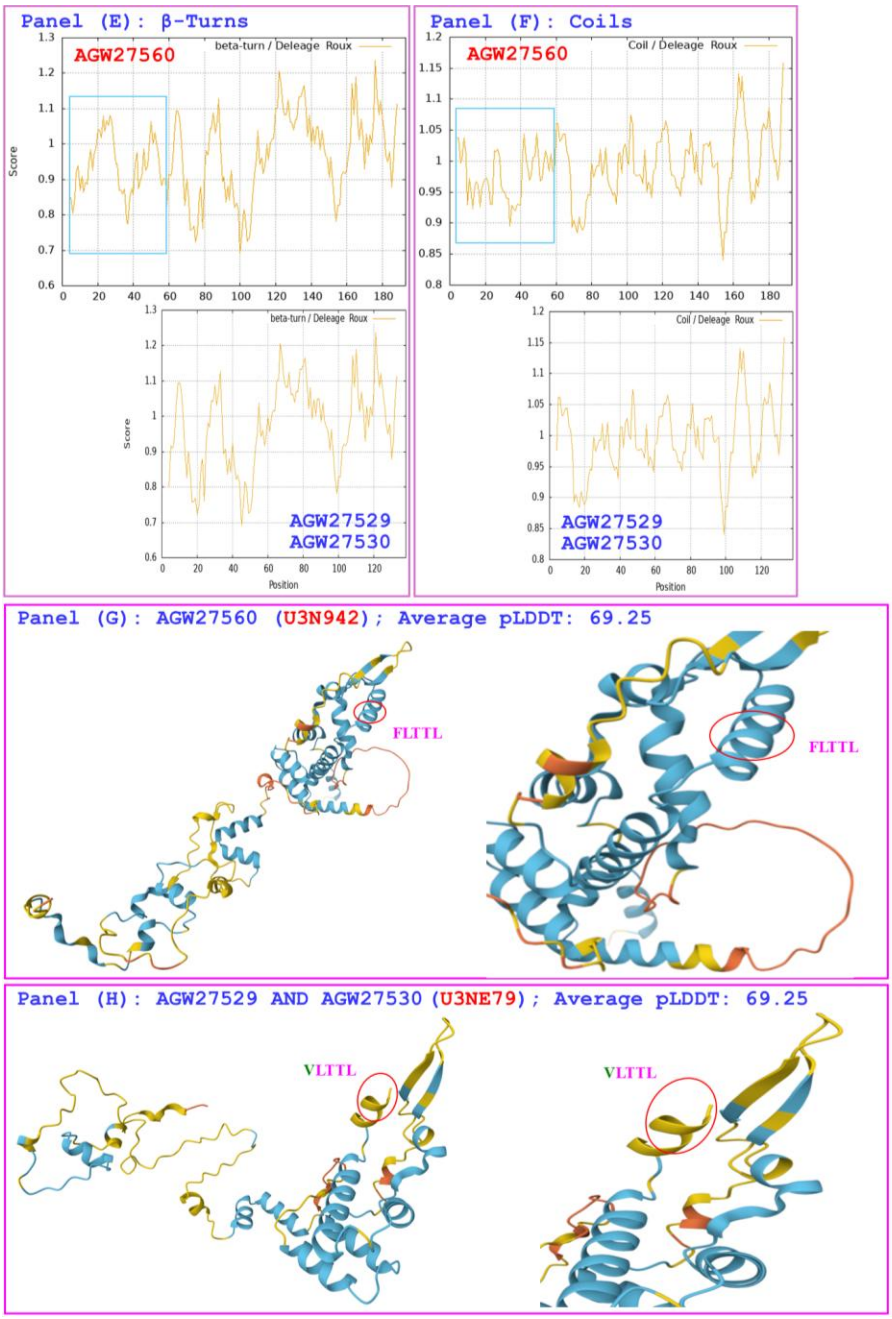


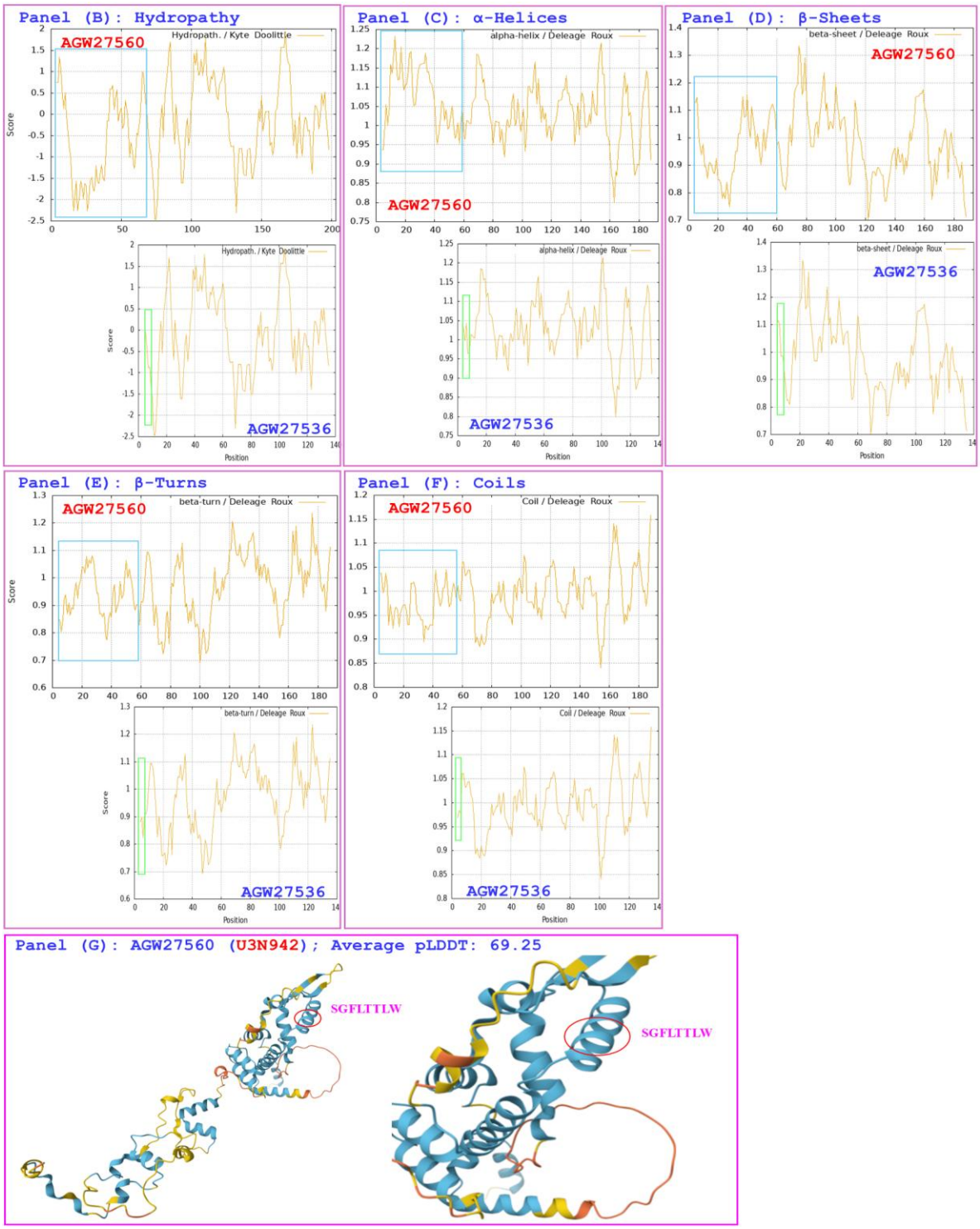
Figure 10. Correlations of the changes in the hydrophobicity and the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the variant proteins AGW27529 and AGW27530 (under the AlphaFold code U3NE79) derived from the *O. sinensis* strains CS70-1208 and CS70-1211. Panel (A) shows an alignment of the amino acid sequences of the MATα_HMGbox domains of the MAT1-1-1 proteins; amino acid substitutions are shown in green, the hyphens indicate identical amino acid residues, and the spaces denote unmatched sequence gaps. The ExPASy ProtScale plots show the changes in hydrophobicity and the 2D structure in Panels (B)–(F) for hydropathy, α-helices, β-sheets, β-turns, and coils of the protein, respectively; the open rectangles in blue highlight the N-terminally truncated regions in the ExPASy plots. Panels (G)–(I) show the 3D structures of the full-length proteins on the left, and the locally magnified structures surrounding the variation site are shown on the right. The model confidence for the AlphaFold-predicted 3D structures is as follows:

very high (pLDDT>90); high (90>pLDDT>70); low (70>pLDDT>50); and very low (pLDDT<50).

The hydrophobicity and structural characteristics of the MATα_HMGbox domain of the variant MAT1-1-1 protein AGW27536 are compared with those of the reference protein AGW27560 in Figure 11. The protein AGW27536 (under the AlphaFold code U3N7H7) is derived from the *O. sinensis* strain CS68-2-1228 (Table 1), whereas the reference protein AGW27560 (under the AlphaFold code U3N942) is derived from the *H. sinensis* strain CS68-2-1229 [14]. Panel (A) of Figure 11 shows a 44-residue truncation at the N-terminus of the MATα_HMGbox domain of the protein AGW27536, along with an S-to-E substitution and a deletion of leucine within the MATα_HMGbox domain. The S-to-E substitution involves a change from serine (with polar neutral side chains) to acidic glutamic acid, with the hydropathy index changing from -0.8 to -3.5 (Table S3) [73]; the deleted leucine has a hydropathy index of 3.8. The truncation and replacement and deletion of amino acids lead to reduced hydrophobicity and an altered secondary structure, as illustrated by the changes in topological structure and waveform in the ExPASy ProtScale plots for hydropathy, α-helices, β-sheets, β-turns, and coils in Panels (B)–(F), respectively. Compared the 3D structures of the MAT1-1-1 proteins illustrated in Panels (G)–(H), the truncation and replaced and deleted amino acids in the protein AGW27536 are located within the first of the 3 core α-helices of the MATα_HMGbox domain. These modifications result in significant changes in the 3D structure of the protein AGW27536, corresponding to the AlphaFold code U3N7H7.

Panel (A): Alignment of the amino acid sequences of the MATα_HMGbox domains

```
AGW27560  42   ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTT  100
AGW27536  1                                                   E-- --  5

AGW27560  101 LWHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQ  160
AGW27536  6   ------------------------------------------------------------  65

AGW27560  161 DDDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSH  220
AGW27536  66  ------------------------------------------------------------  125

AGW27560  221 SGIMAPRAANCTPP                                               234
AGW27536  126 --------------                                               139
```


Panel (B): Hydropathy
AGW27560
Hydropath. / Kyte Doolittle
Score
AGW27536
Position
Panel (C): α-Helices
alpha-helix / Deleage Roux
AGW27560
AGW27536
Panel (D): β-Sheets
beta-sheet / Deleage Roux
AGW27560
AGW27536
Panel (E): β-Turns
beta-turn / Deleage Roux
AGW27560
AGW27536
Panel (F): Coils
Coil / Deleage Roux
AGW27560
AGW27536
Panel (G): AGW27560 (U3N942); Average pLDDT: 69.25
SGFLTTLW
SGFLTTLW

476

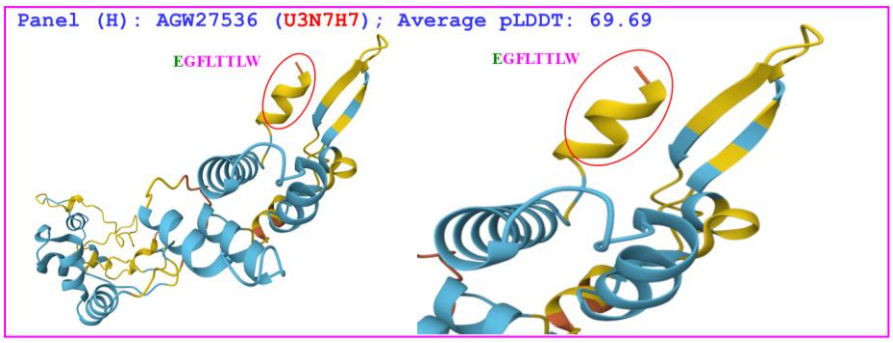


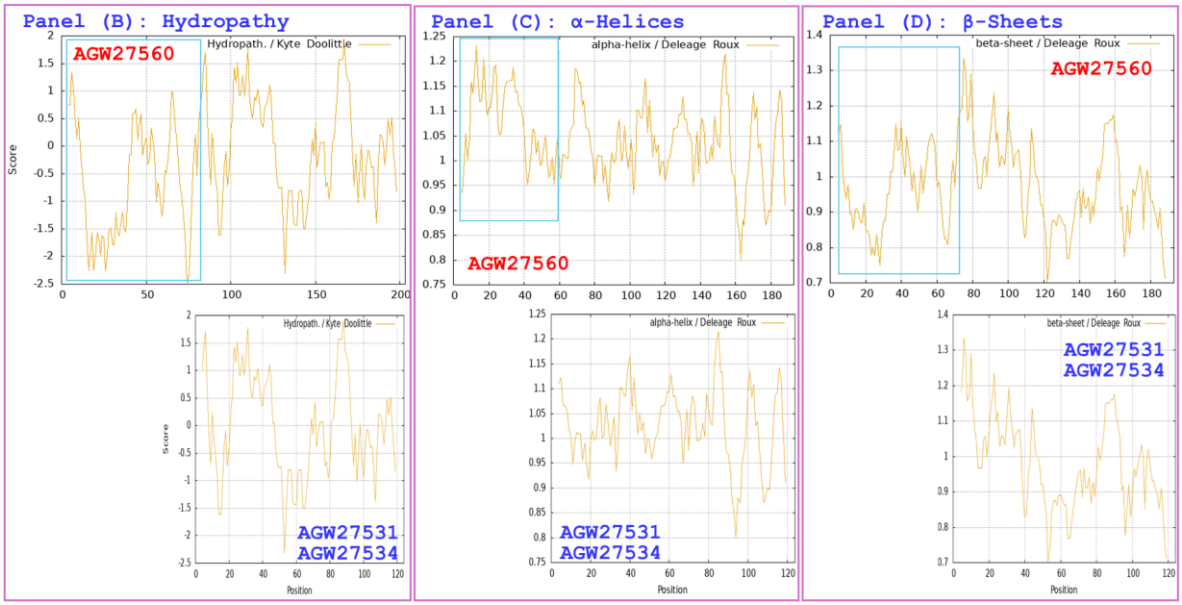
Figure 11. Correlations of the changes in the hydrophobicity and the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the variant protein AGW27536 (under the AlphaFold code U3N7H7) derived from the *O. sinensis* strain CS68-2-1228. Panel (A) shows an alignment of the amino acid sequences of the MATα_HMGbox domains of the MAT1-1-1 proteins, where the hyphens indicate identical amino acid residues and the spaces denote unmatched sequence gaps. The ExPASy ProtScale plots show the changes in hydrophobicity and the 2D structure in Panels (B)–(F) for hydropathy, α-helices, β-sheets, β-turns, and coils of the protein, respectively; the open rectangles in blue highlight the N-terminally truncated regions, and those in green highlight the altered topological structure and waveform in the ExPASy plots. Panels (G)–(H) show the 3D structures of the full-length proteins on the left, and the locally magnified structures surrounding the variation site are shown on the right. The model confidence for the AlphaFold-predicted 3D structures is as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

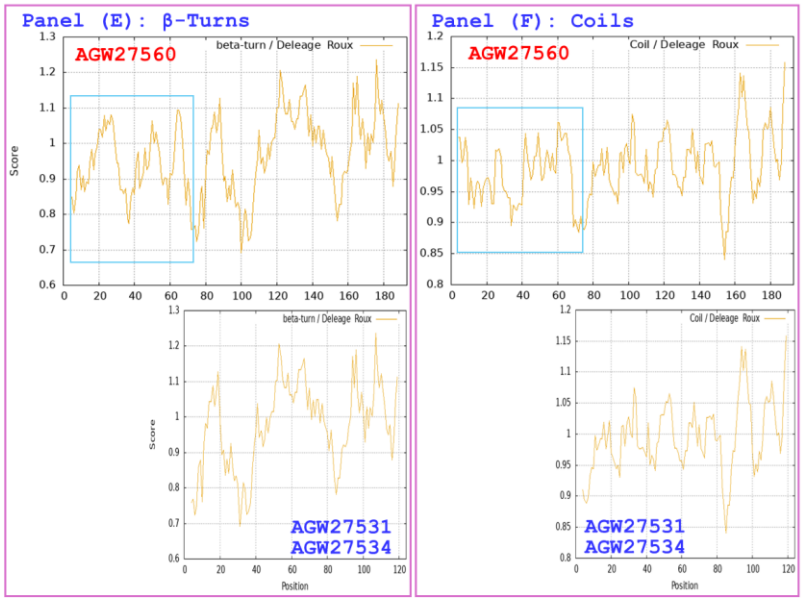
The hydrophobicity and structural characteristics of the MATα_HMGbox domains of the variant MAT1-1-1 proteins AGW27531 and AGW27534 are compared with those of the reference protein AGW27560 in Figure 12. These 2 variant proteins (both under the AlphaFold code U3NE87) are derived from the *O. sinensis* strains CS70-1212 and CS71-1219 (*cf.* Table 1), whereas the reference protein AGW27560 (under the AlphaFold code U3N942) is derived from the *H. sinensis* strain CS68-2-1229 [14]. Panel (A) of Figure 12 shows 60-residue truncations at the N-termini of the MATα_HMGbox domains of the proteins AGW27531 and AGW27534. The truncations result in altered hydrophobicity and secondary structures, as illustrated by the changes in topological structure and waveform in the ExPASy ProtScale plots for hydropathy, α-helices, β-sheets, β-turns, and coils in Panels (B)–(F), respectively. A comparison of the 3D structures of the MAT1-1-1 proteins, as illustrated in Panels (G)–(H), revealed that the truncations in the proteins AGW27531 and AGW27534 are located within the 3 core α-helices of the MATα_HMGbox domains and result in significantly altered tertiary structures of the variant proteins, both of which correspond to the same AlphaFold code U3NE87.

Panel (A): Alignment of the amino acid sequences of the MATα_HMGbox domains

```
AGW27560   42    ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTT   100
AGW27531
AGW27534

AGW27560   101   LWHKDPFRNKWALIAKVYSFVRDQIGKDKVSLSYFMSLACPTMTIIEPAAYLNALGWCVQ   160
AGW27531   1     ------------------------------------------------------------   50
AGW27534   1     ------------------------------------------------------------   50

AGW27560   161   DDDAGSQKLFQDESSANLDQSSLLSAEYPSTEIELLSALVNIGYFPDHGADLVERMGSSH   220
AGW27531   51    ------------------------------------------------------------   110
AGW27534   51    ------------------------------------------------------------   110

AGW27560   221   SGIMAPRAANCTPP                                                234
AGW27531   111   --------------                                                124
AGW27534   111   --------------                                                124
```

503





504

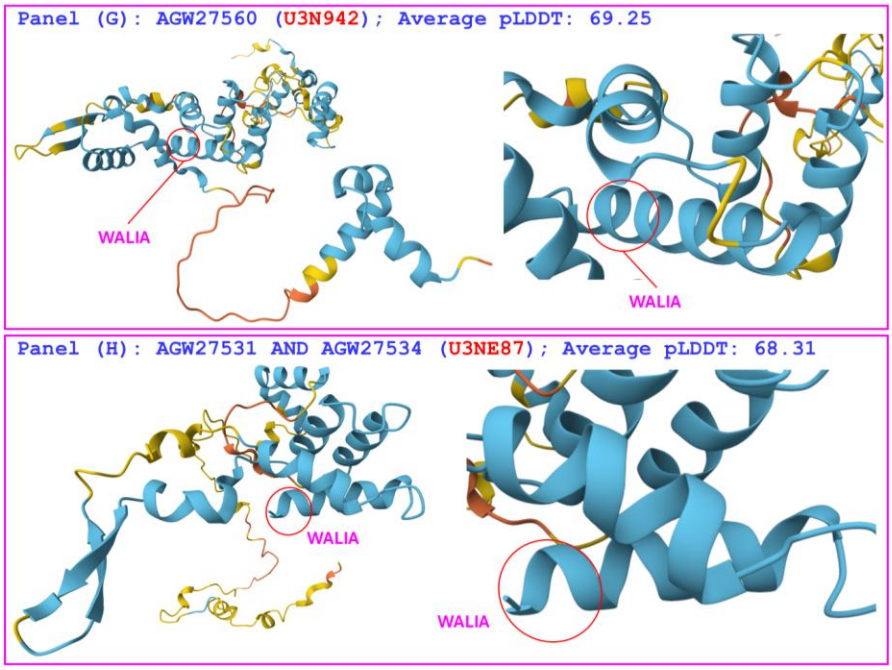


Figure 12. Correlations of the changes in the hydrophobicity and the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis strain* CS68-2-1229 and the variant proteins AGW27531 and AGW27534 (under the AlphaFold code U3NE87) derived from the *O. sinensis* strains CS70-1212 and CS71-1219, respectively. Panel (A) shows an alignment of the amino acid sequences of the MATα_HMGbox domains of the MAT1-1-1 proteins, where the hyphens indicate identical amino acid residues and the spaces denote unmatched sequence gaps. The ExPASy ProtScale plots show the changes in hydrophobicity and the 2D structure in Panels (B)–(F) for hydropathy, α-helices, β-sheets, β-turns, and coils of the protein; the open rectangles in blue highlight the N-terminally truncated regions in the ExPASy plots. Panels (G)–(H) show the 3D structures of the full-length proteins on the left, and the locally magnified structures surrounding the truncation site are shown on the right. The model confidence for the AlphaFold-predicted 3D structures is as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

Collectively, the data in Figures 5–12 in Section 3.3 demonstrate that the MAT1-1-1 protein variants derived from various *O. sinensis* strains exhibit N-terminal truncations (ranging from 13 to 60 residues) of their MATα_HMGbox domains, with different amino acid substitutions or deletions. These modifications significantly alter the hydrophobicity, secondary structures, and tertiary structures of the MATα_HMGbox domains, with truncation and substitution locations occurring either upstream of or within the three core α-helices of the domains. The variant MAT1-1-1 proteins correspond to one of nine distinct AlphaFold 3D structural conformations (*cf.* Table 1, Figure 2), highlighting the structural diversity of the MAT1-1-1 proteins.

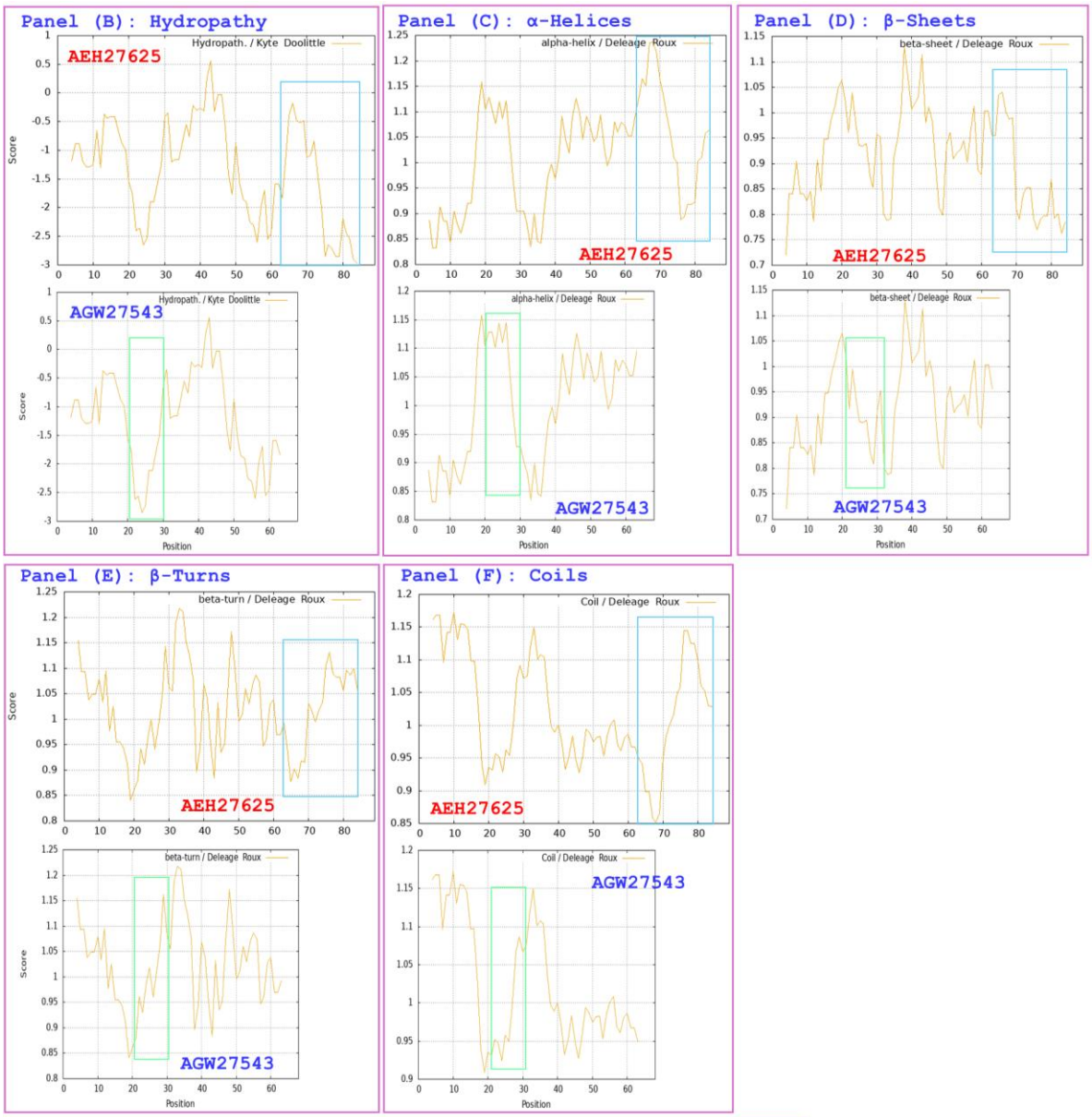
3.4 Heteromorphic tertiary structures of the HMG-box_ROX1-like domains of the truncated MAT1-2-1 proteins derived from *O. sinensis* strains

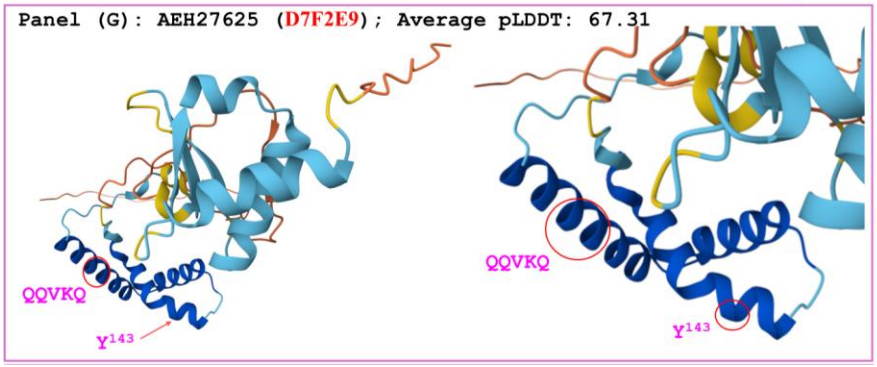
This section presents the results of the correlation analyses between the changes in the hydrophobic properties and structures of the HMG-box_ROX1-like domains of variant MAT1-2-1 proteins derived from various *O. sinensis* strains under 4 AlphaFold codes (*cf.* Table 1) compared with the reference full-length MAT1-2-1 proteins. This reference protein, represented by AEH27625 derived from the *H. sinensis* strain CS2 (*cf.* Table S5), is associated with the AlphaFold 3D structural code D7F2E9 [70]. The HMG-box_ROX1-like domain sequences (amino acid residues 127→197 of the authentic protein AEH27625) were extended both upstream and downstream by 9 external amino acids, which was used in the sequential ExPASy ProtScale plot for hydropathy, α-helices, β-sheets, β-turns, and coils with a window size of 9 amino acid residues (*cf.* Section 2.3).

The hydrophobicity and structural characteristics of the HMG-box_ROX1-like domain of the MAT1-2-1 protein AGW27543 are compared with those of the reference protein AEH27625 in Figure 13. The protein AGW27543 (under the AlphaFold code U3N9W0) is derived from the *O. sinensis* strain CS91-1291 (*cf.* Table 1) [14], while the reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2 [70]. Panel (A) of Figure 13 shows a 12-residue truncation at the C-terminus of the HMG-box_ROX1-like domain of the protein AGW27543, along with a Y-to-H substitution (tyrosine to histidine). This substitution changes the hydropathy index from -1.3 to -3.2 (Table S3) [73], resulting in reduced hydrophobicity. The C-terminal truncation and amino acid substitution, along with the accompanying reduction in hydrophobicity, resulted in altered secondary structures surrounding the variation sites within the HMG-box_ROX1-like domain of the protein AGW27543. This alteration is reflected by changes in the topological structure and waveform of the ExPASy ProtScale plots, which illustrate hydropathy, α-helices, β-sheets, β-turns, and coils in Panels (B)–(F), respectively. As shown in the 3D structures presented in Panels (G)–(H), the truncation and replaced amino acid residues are located in 2 of the 3 α-helices that form the core L-shaped stereostructure. These modifications result in an altered tertiary structure of the HMG-box_ROX1-like domain of the protein AGW27543 under the AlphaFold code U3N9W0.

Panel (A): Alignment of amino acid sequences the HMG-box_ROX1-like domains

```
AEH27625 118 APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK  177
AGW27543 118 --------------------------H--------------------------------  177

AEH27625 178 DMSQQVKQALLEKHPDYQYKPRRPCERRR                                206
AGW27543 178 --------                                                     185
```


Panel (B): Hydropathy
Hydropath. / Kyte Doolittle
AEH27625
AGW27543
Score
Position
Panel (C): α-Helices
alpha-helix / Deleage Roux
AEH27625
AGW27543
Position
Panel (D): β-Sheets
beta-sheet / Deleage Roux
AEH27625
AGW27543
Position
Panel (E): β-Turns
beta-turn / Deleage Roux
AEH27625
AGW27543
Score
Position
Panel (F): Coils
Coil / Deleage Roux
AEH27625
AGW27543
Position


Panel (G): AEH27625 (D7F2E9); Average pLDDT: 67.31
QQVKQ
Y143
QQVKQ
Y143

554

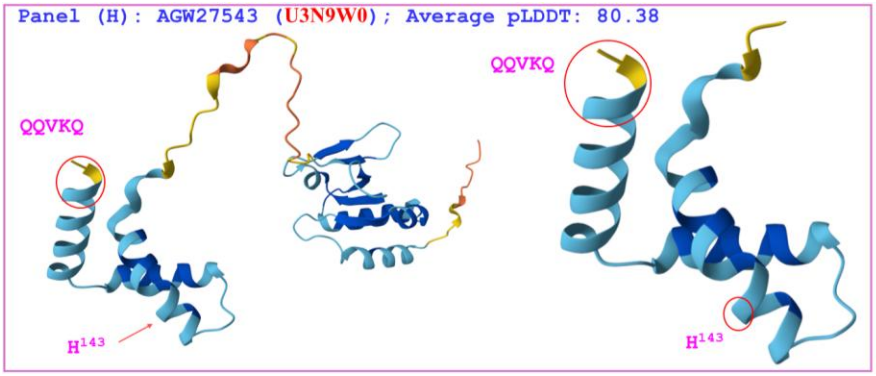


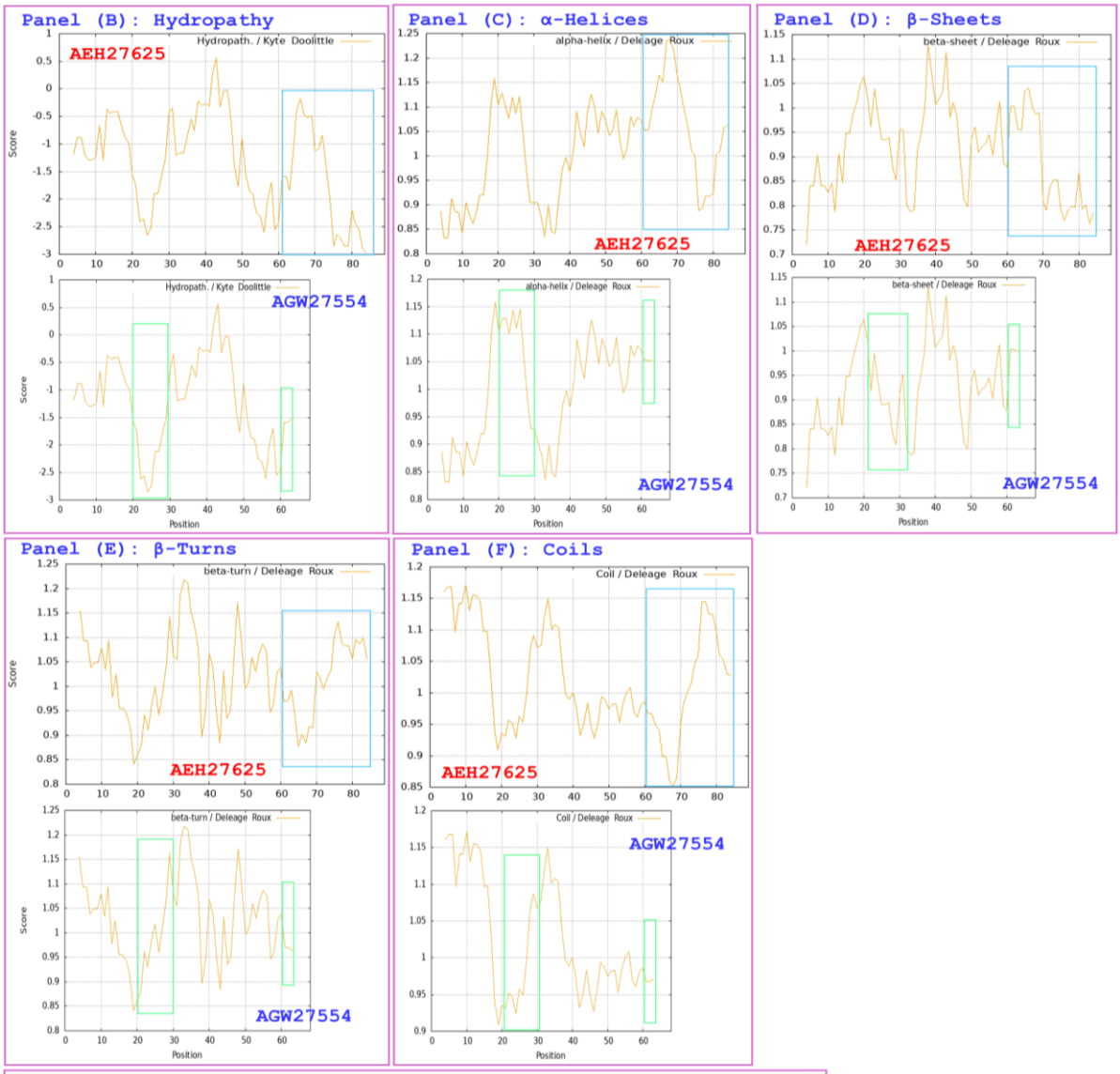
Figure 13. Correlations between the changes in hydrophobicity and the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins. The reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2, and the variant MAT1-2-1 protein AGW27543 (under the AlphaFold code U3N9W0) is derived from the *O. sinensis* strain CS91-1291. Panel (A) shows an alignment of the amino acid sequences of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins; the amino acid substitution is shown in green, the hyphens indicate identical amino acid residues, and the spaces denote unmatched sequence gaps. The ExPASy ProtScale plots show the changes in hydrophobicity and the 2D structure of the protein in Panels (B)–(F) for the hydropathy, α-helices, β-sheets, β-turns, and coils, respectively; the open rectangles in blue highlight the C-terminally truncated regions in the ExPASy plots, and those in green highlight the changes in topological structure and waveform observed in the plots. Panels (G)–(H) show representations of the 3D structures of the full-length proteins on the left; the locally magnified structures at the variation sites are shown on the right. The model confidence for the AlphaFold-predicted 3D structures is as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

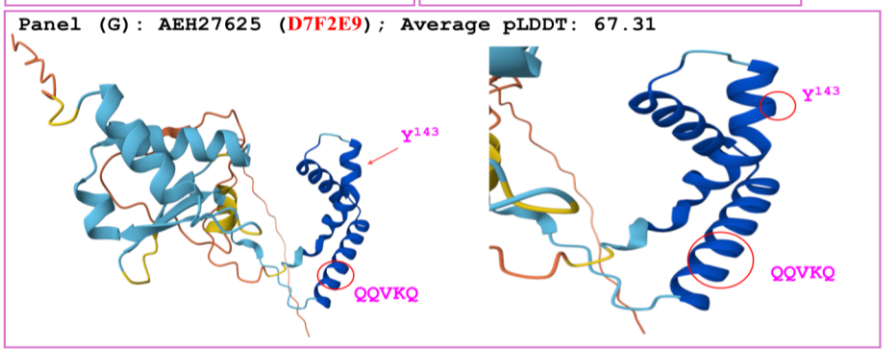
The hydrophobicity and structural characteristics of the HMG-box_ROX1-like domain of the MAT1-2-1 protein AGW27554 are compared with those of the reference protein AEH27625 in Figure 14. The protein AGW27554 (under the AlphaFold code U3NEA9) is derived from the *O. sinensis* strain CS71-1219 [14], whereas the reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2 [70]. Panel (A) of Figure 14 shows a 12-residue truncation at the C-terminus of the HMG-box_ROX1-like domain of the protein AGW27554, along with 2 amino acid substitutions: Y-to-H (tyrosine to histidine; hydropathy indices changed from -1.3 to -3.2) and Q-to-T (glutamine to threonine; hydropathy indices changed from -3.5 to -0.7 (Table S3) [73]. The C-terminal truncation and amino acid substitutions accompanied by altered hydrophobicity result in changes in the secondary structure surrounding the variation sites in the HMG-box_ROX1-like domain of the protein AGW27543. This alteration is reflected by the changes in the topological structure and waveform of the ExPASy ProtScale plots, which illustrate hydropathy, α-helices, β-sheets, β-turns, and coils, in Panels (B)–(F), respectively. As illustrated in the 3D structures in Panels (G)–(H), the truncated and replaced residues are located in 2 of the 3 core α-helices. These modifications result in an altered tertiary structure of the HMG-box_ROX1-like domain of the protein AGW27554 under the AlphaFold code U3NEA9.

Panel (A): Alignment of amino acid sequences the HMG-box_ROX1-like domains

```
AEH27625 118 APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK  177
AGW27554 118 --------------------------H-------------------------------  177

AEH27625 178 DMSQQVKQALLEKHPDYQYKPRRPCERRR                                206
AGW27554 178 -------T                                                     185
```

585





586

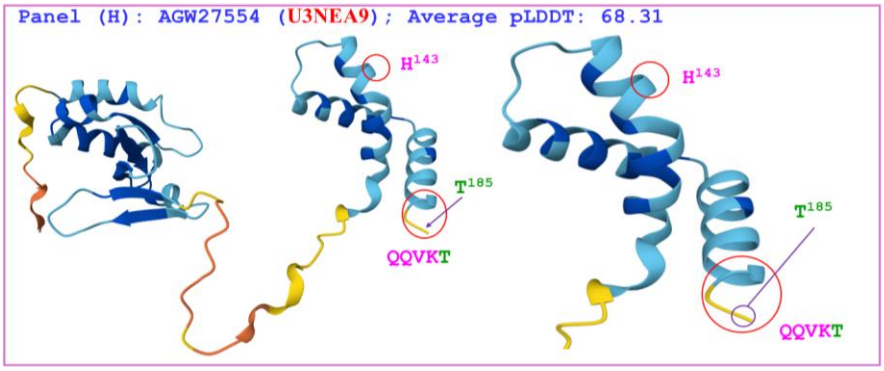


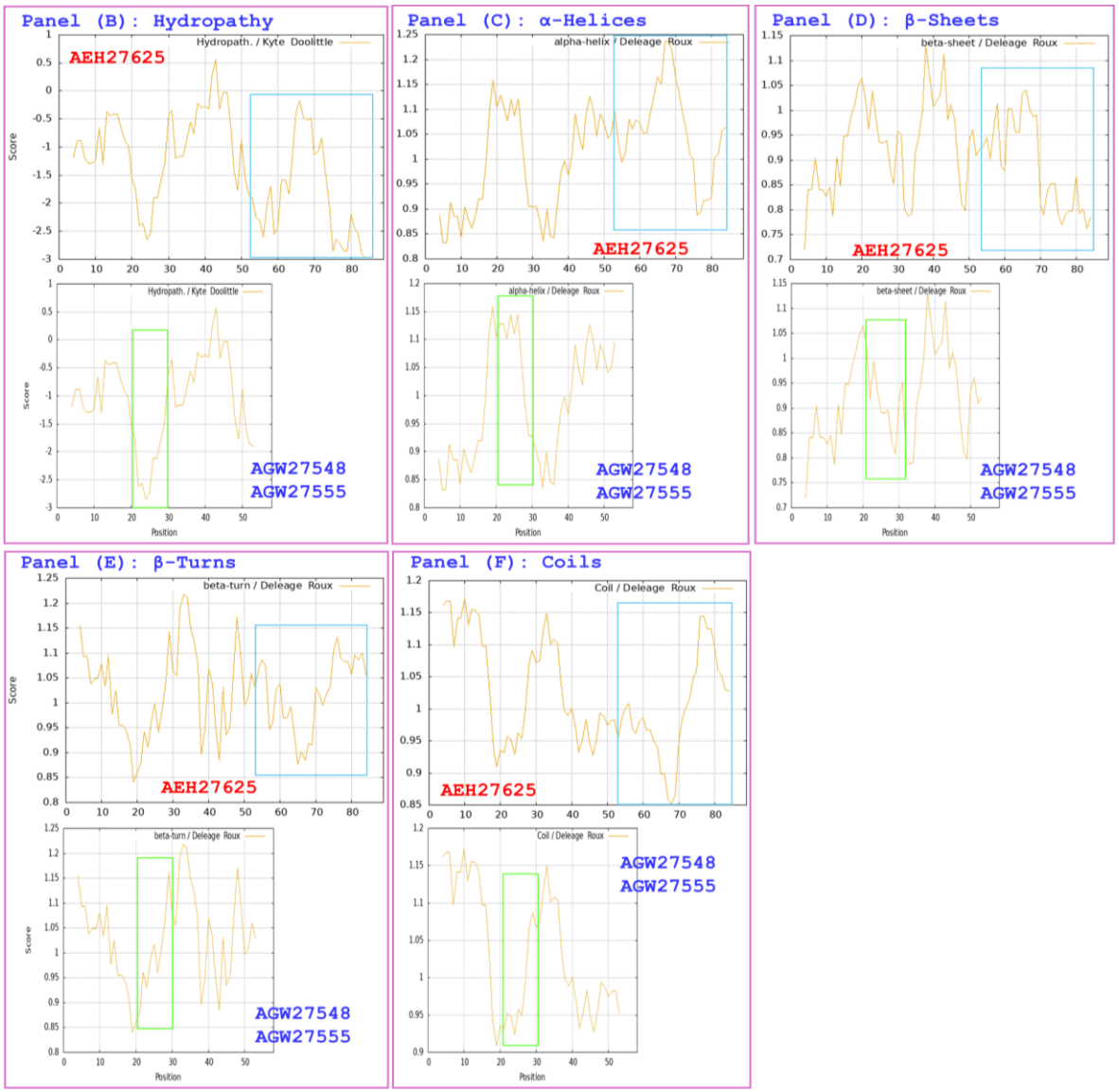
Figure 14. Correlations between changes in hydrophobicity and the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins. The reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2, and the variant MAT1-2-1 protein AGW27554 (under the AlphaFold code U3NEA9) is derived from the *O. sinensis* strain CS71-1219. Panel (A) shows an alignment of the amino acid sequences of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins; amino acid substitutions are shown in green, the hyphens indicate identical amino acid residues, and the spaces denote unmatched sequence gaps. The ExPASy ProtScale plots show the changes in hydrophobicity and the 2D structure of the protein in Panels (B)–(F) for hydropathy, α-helices, β-sheets, β-turns, and coils, respectively; the open rectangles in blue highlight the C-terminal truncation regions in the ExPASy plots, and those in green highlight the changes in the topological structure and waveform. Panels (G)–(H) show representations of the 3D structures of the full-length proteins on the left; the locally magnified structures at the variation sites are shown on the right. The model confidence for the AlphaFold-predicted 3D structures is as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

The hydrophobicity and structural characteristics of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins AGW27548 and AGW27555 are compared with those of the reference protein AEH27625, as shown in Figure 15. These 2 variant proteins (both under the AlphaFold code U3N9W5) are derived from the *O. sinensis* strains CS68-2-1229 and CS71-1220, respectively (*cf.* Table 1) [14], while the reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2 [70]. Panel (A) of Figure 15 shows 22-residue truncations at the C-termini of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins AGW27548 and AGW27555, along with a Y-to-H (tyrosine to histidine) amino acid substitution within the HMG-box_ROX1-like domains of both variant proteins. This substitution changes the hydropathy index from -1.3 to -3.2 (Table S3) [73]. The C-terminal truncations and amino acid substitutions caused decreased hydrophobicity and altered the secondary structures surrounding the variation sites in the HMG-box_ROX1-like domains of the MAT1-2-1 proteins AGW27548 and AGW27555. This alteration is reflected by changes in the topological structure and waveform of the ExPASy ProtScale plots, which illustrate hydropathy, α-helices, β-sheets, β-turns, and coils in Panels (B)–(F), respectively. As shown in the 3D structures in Panels (G)–(H), the truncations and replaced amino acid residues are located in 2 of the 3 core α-helices. These modifications result in altered tertiary structures of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins

AGW27548 and AGW27555 under the AlphaFold code U3N9W5.

Panel (A): Alignment of amino acid sequences the HMG-box_ROX1-like domains

```
AEH27625 118 APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK  177
AGW27548 118 --------------------------H------------------------------  175
AGW27555 118 --------------------------H------------------------------  175

AEH27625 178 DMSQQVKQALLEKHPDYQYKPRRPCERRR                               206
AGW27548 175
AGW27555 175
```



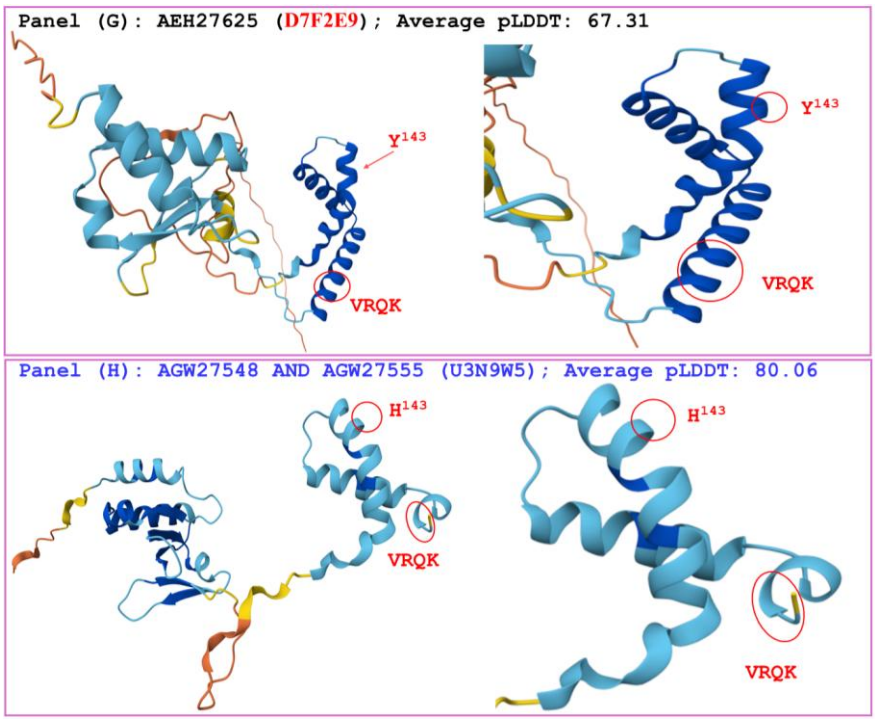


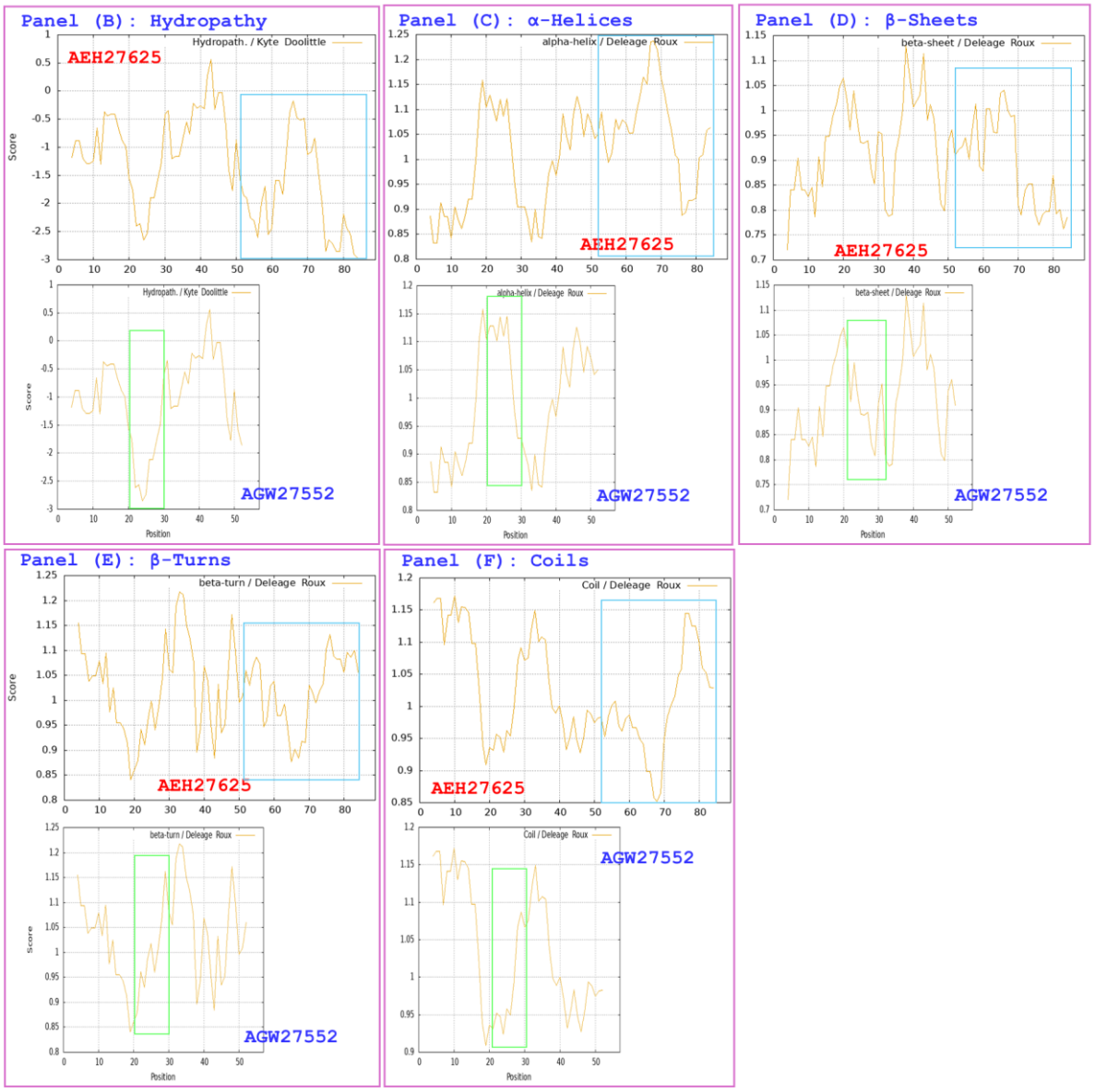
Figure 15. Correlations between changes in hydrophobicity and the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins. The reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2, and the variant MAT1-2-1 proteins AGW27548 and AGW27555 (under the AlphaFold code U3N9W5) are derived from the *O. sinensis* strains CS68-2-1229 and CS71-1220, respectively. Panel (A) shows an alignment of the amino acid sequences of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins; amino acid substitutions are shown in green, the hyphens indicate identical amino acid residues, and the spaces denote unmatched sequence gaps. The ExPASy ProtScale plots show the changes in hydrophobicity and the 2D structure in Panels (B)–(F) for the hydropathy, α-helices, β-sheets, β-turns, and coils of the proteins, respectively; the open rectangles in blue highlight the C-terminally truncated regions in the ExPASy plots, and the changes in topological structure and waveform observed in the plots are highlighted in red. Panels (G)–(H) show representations of the 3D structures of the full-length proteins on the left; the locally magnified structures at the variation sites are shown on the right. The model confidence for the AlphaFold-predicted 3D structures is as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

The hydrophobicity and structural characteristics of the HMG-box_ROX1-like domain of the MAT1-2-1 protein AGW27552 are compared with those of the reference protein AEH27625, as shown in Figure 16. The variant protein (under the AlphaFold code U3N6W6) is derived from the *O. sinensis* strain CS68-5-1216 [14], while the reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2 [70]. Panel (A) of Figure 16 shows a 23-residue truncation at the C-terminus of the HMG-box_ROX1-like domain in the MAT1-2-1 protein AGW27552, along with a Y-to-H substitution (tyrosine to histidine; hydropathy indices changed from -1.3 to -3.2) (Table S3) [73]) within the HMG-box_ROX1-like domain. This amino acid substitution caused decreased hydrophobicity and modified the secondary structures surrounding the sites of variation within the HMG-box_ROX1-

like domain of the MAT1-2-1 protein AGW27552. Changes in the topological structure and waveform are reflected in the ExPASy ProtScale plots corresponding to hydropathy, α-helices, β-sheets, β-turns, and coils in Panels (B)–(F), respectively. As shown in the comparison of the 3D structures in Panels (G)–(H) for a comparison of 3D structures, the truncation and replaced residue occur within 2 of the 3 core α-helices and completely remodel the overall stereostructure of the HMG-box_ROX1-like domain of the MAT1-2-1 protein AGW27552 (under AlphaFold code U3N6W6).

Panel (A): Alignment of amino acid sequences the HMG-box_ROX1-like domains

```
AEH27625 118 APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK  177
AGW27552 118 --------------------------H------------------------------  174

AEH27625 178 DMSQQVKQALLEKHPDYQYKPRRPCERRR                                206
AGW27552 174
```



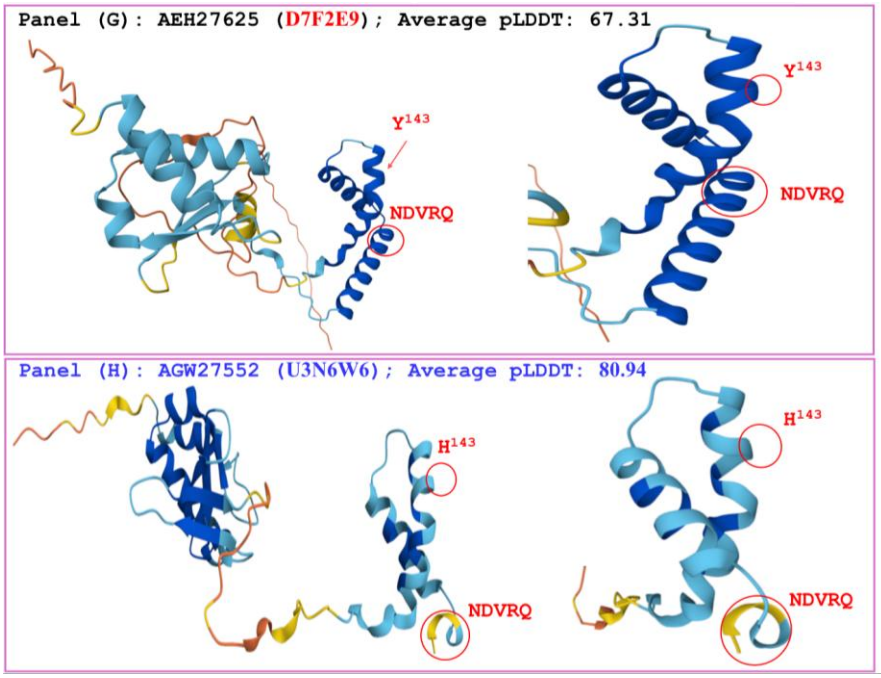


Figure 16. Correlations between changes in hydrophobicity and the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins. The reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2, and the variant MAT1-2-1 protein AGW27552 (under the AlphaFold code U3N6W6) is derived from the *O. sinensis* strain CS68-5-1216. Panel (A) shows an alignment of the amino acid sequences of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins; amino acid substitutions are shown in green, the hyphens indicate identical amino acid residues, and the spaces denote unmatched sequence gaps. The ExPASy ProtScale plots show the changes in hydrophobicity and the 2D structure of the protein in Panels (B)–(F) for the hydropathy, α-helices, β-sheets, β-turns, and coils, respectively; the open rectangles in blue highlight the C-terminally truncated regions in the ExPASy plots, and those in red highlight the changes in topological structure and waveform observed in the plots. Panels (G)–(H) show representations of the 3D structures of the full-length proteins on the left; the locally magnified structures at the sites of substitutions are shown on the right. The model confidence for the AlphaFold-predicted 3D structures is as follows: ■ very high (pLDDT>90); ■ high (90>pLDDT>70); ■ low (70>pLDDT>50); and ■ very low (pLDDT<50).

Collectively, the data in Figures 13–16 in Section 3.4 demonstrate that variant MAT1-2-1 proteins derived from various *O. sinensis* strains exhibit C-terminal truncations (ranging from 12 to 23 residues) within their HMG-box_ROX1-like domains, accompanied by tyrosine-to-histidine (Y-to-H) substitutions. These structural modifications alter the hydrophobicity, secondary structures, and tertiary structures of the HMG-box_ROX1-like domains. Both truncations and substitutions are localized within the 3 core α-helices of the domains. The variant MAT1-2-1 proteins correspond to one of four distinct AlphaFold 3D structural conformations (*cf.* Table 1, Figure 4), highlighting the structural diversity of this protein family among different *O. sinensis* strains.

3.5 Diverse fungal sources of the MAT1-1-1 and MAT1-2-1 protein variants

676 **produced simultaneously in pairs by the 20 *O. sinensis* strains**

677 **3.5a Complex genetic heterogeneity of the *O. sinensis* strains**

678 The ITS sequences used for phylogenetic analysis are available in GenBank for 15 of the 20 *O.*
679 *sinensis* strains obtained by Li *et al.* [36] but are lacking for the remaining 5 strains (*cf.* Table 1). The
680 design of the PCR experiment in that study included the use of a pair of fungal universal primers
681 (ITS5/ITS4 [92]) and 3 pairs of self-designed primers (GAF/GAR, GCF/GCR, and 5.8S-F/R) for ITS
682 sequence amplification, and the annealing temperatures were set to 53°C for primers ITS5/ITS4, 55°C
683 for primers GAF/GAR specific for the GC-biased sequences of Group A (Genotype #1 of *O. sinensis*),
684 and 60°C for primers GCF/GCR specific for the AT-biased sequences of group C (Genotype #5 of *O.*
685 *sinensis*). Table S6 lists the percentage identities and sequence coverages between the sequences of the
686 primers and the representative GC- and AT-biased genotypes of *O. sinensis*. Table S6 shows that many
687 genotypes exhibit no or significantly low sequence identities and coverages, indicating that they may not
688 be amplified if they truly co-occur in the test samples. Compared with other sequences, some genotypic
689 components of *O. sinensis* might be much more efficiently amplified, causing differential dominances
690 of the PCR amplicons. Furthermore, Li *et al.* [36] adopted a direct sequencing strategy for amplicons.
691 Obviously, this study did not employ rigorous amplicon cloning–sequencing techniques with the
692 selection of a sufficient number (30−50) of white colonies to achieve comprehensive, unbiased genetic
693 profiling, nor did it use the MassARRAY single-nucleotide polymorphism (SNP) MALDI–TOF mass
694 spectrum genotyping technique with multiple extension primers for comprehensive phylogenetic
695 examinations. Consequently, molecular examination after 25 days of shaker incubation of *C. sinensis*
696 monoascospores and tissues of caterpillar bodies (sclerotium) would have inevitably overlooked
697 nonculturable fungal sequences and some co-occurring amplicon sequences whose abundance in the
698 amplicon pools might have been low because of low amplification efficiency under the design used in
699 this study [36,37].

700 Under these experimental settings, Li *et al.* [36] reported that 10 of the 15 *O. sinensis* strains are
701 homogenous, harboring only GC-biased Genotype #1 *H. sinensis* (*cf.* Table 1); the remaining 5 strains
702 are heterogeneous, containing both GC-biased Genotype #1 (Group A) and AT-biased Genotype #17
703 (Group C) (*cf.* Tables S1−S2) [9,10,12,13,36,37]. However, because of the insufficiently refined research
704 design adopted by Li *et al.* [36], a high level of vigilance is warranted regarding the potential coexistence
705 of additional fungal species or *O. sinensis* genotypes in previously purportedly homogenous and
706 heterogeneous strains. Given the complex genetic background of the *O. sinensis* strains, the possibility

707 cannot be excluded that the detected MAT1-1-1 and/or MAT1-2-1 protein variants, which are
708 characterized by differential truncations and various amino acid substitutions at distinct variation sites,
709 originate from other unnoticed coexisting fungi.

710 Notably, the remaining 5 *O. sinensis* strains (CS26-277, CS36-1294, CS68-5-1216, CS70-1211, and
711 CS70-1212, shown in black in Table 1) lack ITS sequence records in the GenBank database, and their
712 homo/heterogeneous characteristics remain unknown. Researchers who possess these strains [14,36] are
713 expected to upload the missing ITS sequences to GenBank, ensuring the comprehensive and impartial
714 molecular documentation of these strains.

715 **3.5b Differential pairing of the MAT1-1-1 and MAT1-2-1 proteins simultaneously produced by**
716 **the 20 *O. sinensis* strains**

717 Sections 3.1−3.4 present various AlphaFold tertiary structural models of the mating protein variants
718 derived from 20 *O. sinensis* strains. These strains simultaneously produced paired MAT1-1-1 and MAT1-
719 2-1 proteins that harbor differential truncations at the N- and/or C-termini and diverse amino acid
720 substitutions. As shown in Table 2, 10 of the 20 strains are "homogenous" strains shown in brown, 5 are
721 heterogeneous strains shown in red, and the remaining 5 strains have unknown homo/heterogeneous
722 characteristics and are indicated in black. On the basis of the pairing patterns of the 2 mating proteins,
723 these strains can be categorized into 2 groups (Table 2). **Group I** consists of 5 strains, which
724 simultaneously produce truncated MAT1-1-1 and MAT1-2-1 proteins in pairs; **Group II** is composed of
725 the remaining 15 *O. sinensis* strains, which simultaneously produce the truncated MAT1-1-1 proteins
726 and the full-length MAT1-2-1 proteins in pairs (*cf.* Table S5). The MAT1-1-1 and MAT1-2-1 proteins
727 that share the same AlphaFold 3D structural morph are labeled with the same suffix symbol (⁋, †, *, ⁑,
728 ‡, or ♦) in Table 2.

729 **Table 2**. Two groups of the 20 *O. sinensis* strains based on the differential co-occurrence of pairs of MAT1-1-1
730 and MAT1-2-1 proteins with different truncations and amino acid substitutions

Grp.	*O. sinensis* strains	MAT1-1-1 protein					MAT1-2-1 protein			
		GenBank accession #	Number of residues truncated at		Amino acid substitution	AlphaFold code	GenBank accession #	Number of residue(s) truncated at C terminus	Amino acid substitution	AlphaFold code
			N termin.	C termin.						
I	CS71-1219	AGW27534	110	8		U3NE87 *	AGW27554	65	Y-to-H, Q-to-T	U3NEA9
	CS68-5-1216	AGW27532	97	8		U3N6U8 †	AGW27552	75	Y-to-H	U3N6W6
	CS71-1220	AGW27535	97	8		U3N6U8 †	AGW27555	74	Y-to-H	U3N9W5 ⁑
	CS68-2-1229	AGW27528	63	8		U3N9T9 ⁋	AGW27548	74	Y-to-H	U3N9W5 ⁑
	CS91-1291	AGW27523	63	8		U3N9T9 ⁋	AGW27543	64	Y-to-H	U3N9W0
II	CS37-295	AGW27519	63	8		U3N9T9 ⁋	AGW27539			D7F2E9 ♦
	CS36-1294	AGW27518	63	8		U3N9T9 ⁋	AGW27538			D7F2E9 ♦
	CS26-277	AGW27521	63	8		U3N9T9 ⁋	AGW27541			D7F2E9 ♦
	CS6-251	AGW27517	63	8		U3N9T9 ⁋	AGW27537		V-to-I, Y-to-H	U3N6V5
	CS18-266	AGW27520	63	8		U3N9T9 ⁋	AGW27540		Y-to-H	—
	CS34-291	AGW27524	63	8		U3N9T9 ⁋	AGW27544		Y-to-H	—
	CS71-1218	AGW27533	97	8		U3N6U8 †	AGW27553	1	Y-to-H, K-to-X	U3N9X0
	CS70-1211	AGW27530	97	8	F-to-V	U3NE79 ‡	AGW27550		Y-to-H	—
	CS70-1208	AGW27529	97	8	F-to-V	U3NE79 ‡	AGW27549		Y-to-H	—
	CS70-1212	AGW27531	110	8		U3NE87 *	AGW27551		Y-to-H	—
	CS560-961	AGW27522	63	8	P-to-L	U3N6U0	AGW27542		Y-to-H	D7F2E3
	CS76-1284	AGW27525	66	10		U3N919	AGW27545		Y-to-H	—
	CS561-964	AGW27526	63	13	CDRA-to-SSFT	U3N7G5	AGW27546		Y-to-H	—
	CS25-273	AGW27527	77	18		U3N6U4	AGW27547		Y-to-H	—
	CS68-2-1228	AGW27536	95	8	S-to-E, L-residue deletion	U3N7H7	AGW27556		Y-to-H	—

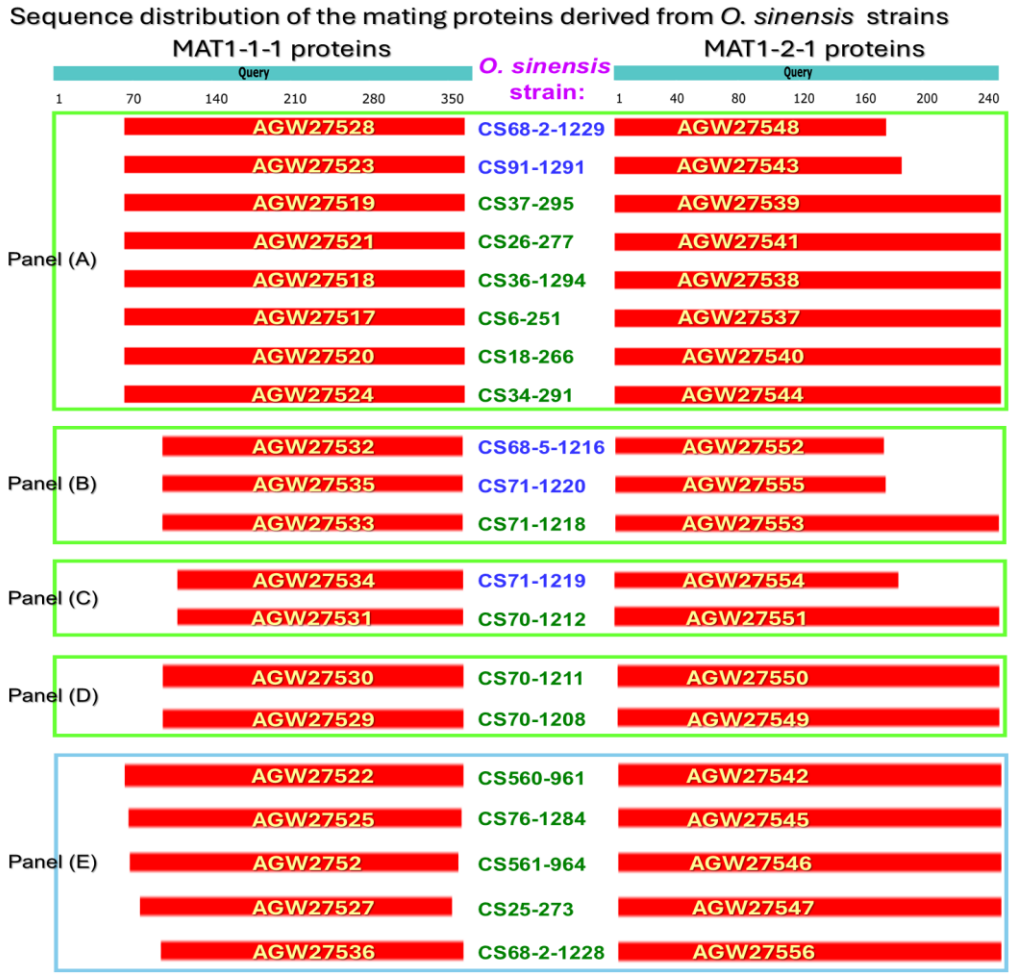
Note: The *O. sinensis* strains shown in brown are "homogenous" and were detected only GC-biased Genotype #1 of *O. sinensis*; those shown in red are heterogeneous, and both GC-biased Genotype #1 and AT-biased Genotype #17 of *O. sinensis* were detected; and those shown in black have unknown homo/heterogeneous characteristics (*cf.* Tables 1 and S1−S2) [9,10,14,36]. The GenBank protein accession numbers in pink refer to truncated MAT1-1-1 and MAT1-2-1 proteins, and those in green indicate full-length MAT1-2-1 proteins. **Group I** strains feature the production of paired truncated MAT1-1-1 and MAT1-2-1 proteins. **Group II** strains produce truncated MAT1-1-1 proteins that are paired with full-length MAT1-2-1 proteins. AlphaFold codes with the same suffix symbol (⁋, †, *, ⁑, ‡, or ♦) belong to the same AlphaFold-predicted 3D structural morphs. Those without a suffix symbol indicate unique 3D structural morphs.

The 15 **Group-II** *O. sinensis* strains produced the full-length MAT1-2-1 proteins shown in green in Table 2. Three (AGW27538, AGW27539, and AGW27541) of them are 100% identical to the authentic MAT1-2-1 proteins represented by protein AEH27625 and share the same tertiary structures under the AlphaFold code D7F2E9. The remaining 12 MAT1-2-1 proteins (AGW27537, AGW27540,

AGW27542, AGW27544−AGW27547, AGW27549−AGW27551, AGW27553, and AGW27556) contain 1−2 amino acid substitutions (Figure S3, Tables 2 and S5), all of which contain Y-to-H substitutions in the HMG-box_ROX1-like domains. The MAT1-2-1 protein AGW27537 derived from the *O. sinensis* strain CS6-251 contains an additional V-to-I (valine-to-leucine) substitution located upstream of the HMG-box_ROX1-like domain. The protein AGW27553 derived from the *O. sinensis* strain CS71-1218 contains an additional K-to-X (lysine-to-an unidentified residue) substitution located within the HMG-box_ROX1-like domain. Six of the 12 MAT1-2-1 proteins exhibit diverse tertiary structures under distinct AlphaFold 3D structural codes, whereas the remaining 6 proteins have not been assigned AlphaFold-predicted 3D structural morphs (Tables 2 and S5).

3.5c *O. sinensis* strains simultaneously produced MAT1-1-1 proteins with the same tertiary structures and MAT1-2-1 proteins with various tertiary structures

The sequence distributions of the paired MAT1-1-1 and MAT1-2-1 proteins that were simultaneously produced by each of the 20 *O. sinensis* strains are shown in Figure 17.



759 **Figure 17**. Alignment and sequence distributions of the paired MAT1-1-1 and MAT1-2-1 proteins derived from
760 the *O. sinensis* strains in **Group I** shown in blue, and those in **Group II** shown in green (*cf.* Table 2). The MAT1-
761 1-1 protein AGW27560 (AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 was used as
762 the query MAT1-1-1 sequence, and the MAT1-2-1 protein AEH27625 (AlphaFold code D7F2E9) derived from
763 the *H. sinensis* strain CS2 was used as the query MAT1-2-1 sequence in the alignment. The panels outlined with
764 open rectangles in green and blue are arranged according to the AlphaFold 3D structural models for the MAT1-1-
765 1 and MAT1-2-1 proteins, respectively.

766 Panel (A) of Figure 17 includes 8 *O. sinensis* strains: four purportedly homogenous strains are
767 shown in brown in Table 2, two heterogeneous strains are shown in red, and the remaining two strains
768 are shown in black with unknown homo/heterogeneous characteristics. All these strains produced MAT1-
769 1-1 proteins truncated by 63- and 8-residues at the N- and C-termini, respectively, which exhibit the
770 same tertiary structures under the AlphaFold 3D structural code U3N9T9 labeled with the suffix symbol
771 "¶" in Table 2 and analyzed above in Figure 5. Two purportedly homogenous strains, CS68-2-1229 and
772 CS91-1291, belong to **Group I**, as shown in blue in Figure 17; these strains simultaneously produce the
773 MAT1-2-1 proteins AGW27548 and AGW27543, which are differentially truncated by 74- and 64-
774 residues, respectively, at the C-termini and contain Y-to-H (tyrosine-to-histidine) substitutions within the
775 HMG-box_ROX1-like domains (Table 2; Figures 13, 15, and 17), resulting in altered 3D structures under
776 AlphaFold codes U3N9W5 and U3N9W0, respectively.

777 The remaining 6 *O. sinensis* strains in Panel (A) belong to **Group II**, as shown in green in Figure
778 17, including 2 purportedly homogenous strains, 2 heterogeneous strains, and 2 additional strains with
779 unknown homo/heterogeneous characteristics. In addition to the production of the same truncated
780 MAT1-1-1 proteins (under the AlphaFold code U3N9T9, labeled with the suffix symbol "¶" in Table 2),
781 these 6 strains simultaneously produced full-length MAT1-2-1 proteins with heteromorphic 3D
782 structures, which correspond to different AlphaFold codes or lack AlphaFold 3D structure codes and can
783 be subgrouped as described below (*cf.* Table 2 and Panel (A) of Figure 17):

784 (1) The coproduced full-length MAT1-2-1 proteins AGW27539, AGW27541, and AGW27538
785 exhibit the same 3D structure under the AlphaFold 3D structural code **D7F2E9**, which is marked
786 with the suffix symbol "♦" in Table 2. These proteins are 100% identical to the authentic MAT1-
787 2-1 protein AEH27625 (*cf.* Figure S3, Table S5) and were generated by the "homogenous" *O.*
788 *sinensis* strain CS37-295, as well as the strains CS26-277 and CS36-1294 that lack ITS sequence
789 records in GenBank (*cf.* Tables 1−2, Panel (A) of Figure 17).

790 (2) The coproduced full-length MAT1-2-1 protein AGW27537 has an altered tertiary structure under
791 the AlphaFold 3D structural code **U3N6V5** (*cf.* Figure 12 of [65] and Table S5) and was produced
792 by the heterogeneous *O. sinensis* strain CS6-251 (*cf.* Tables 1−2, Panel (A) of Figure 17). This

793 protein is 99.2% similar to the authentic MAT1-2-1 protein AEH27625 and contains V-to-I and Y-
794 to-H substitutions, which are located upstream of and within the HMG-box_ROX1-like domain,
795 respectively (*cf.* Figures S3 and 17, Tables 2 and S3).
796 (3) The coproduced full-length MAT1-2-1 proteins AGW27540 and AGW27544 were generated by
797 the "homogenous" *O. sinensis* strain CS18-266 and the heterogeneous strain CS34-291. The
798 predicted 3D structures of these proteins are not available in the AlphaFold database (*cf.* Tables
799 1−2, Panel (A) of Figure 17). These proteins contain Y-to-H substitutions within the HMG-
800 box_ROX1-like domains and show 99.6% identity to the authentic MAT1-2-1 protein AEH27625
801 (*cf.* Figure S3).
802 Panel (B) of Figure 17 includes 3 *O. sinensis* strains: CS68-5-1216 and CS71-1220 from **Group I**
803 and CS71-1218 from **Group II** (Table 2). Strain CS71-1218 is "homogenous", strain CS71-1220 is
804 heterogeneous, and strain CS68-5-1216 has unknown homo/heterogeneous characteristics [14,36] (*cf.*
805 Table 1). All 3 strains produced the MAT1-1-1 proteins (AGW27532, AGW27535, and AGW27533; *cf.*
806 Figures 9 and 17) with 97- and 8-residue truncations at the N- and C-termini, respectively, which
807 conform to the same AlphaFold 3D structural model (U3N6U8 marked with the suffix symbol "†" in
808 Table 2). Simultaneously, strains CS68-5-1216 and CS71-1220 produced the MAT1-2-1 proteins
809 AGW27552 and AGW27555 with 75- and 74-residue truncations, respectively, at the C-termini, whereas
810 strain CS71-1218 produced the MAT1-2-1 protein AGW27553 with a single residue deletion at the C-
811 terminus (*cf.* Figures 3,15−17, and S3). These 3 MAT1-2-1 proteins contain Y-to-H substitutions within
812 the HMG-box_ROX1-like domains, whereas the full-length protein AGW27553 contains an additional
813 K-to-X (lysine-to-an unidentified residue) substitution in the HMG-box_ROX1-like domain. They show
814 99.6%, 99.6% and 99.2% similarity to the authentic MAT1-2-1 protein AEH27625 (*cf.* Figure S3) and
815 exhibit distinct 3D structures under the AlphaFold codes U3N6W6, U3N9W5, and U3N9X0,
816 respectively.
817 Panel (C) of Figure 17 lists the **Group-I** "homogenous" *O. sinensis* strain CS71-1219 and the
818 **Group-II** strain CS70-1212, which have unknown homo/heterogeneous characteristics (Tables 1−2).
819 These 2 strains produced the MAT1-1-1 proteins AGW27534 and AGW27531 with 110- and 8-residue
820 truncations at the N- and C-termini, respectively (*cf.* Figures 12 and 17), which conform to the same
821 AlphaFold 3D structural code U3NE87 (labeled with a suffix symbol "*" in Tables 1−2). Simultaneously,
822 strain CS71-1219 produced the MAT1-2-1 protein AGW27554, which is 65-residue truncated at the C-
823 terminus and contains Y-to-H and Q-to-T (glutamine-to-threonine) substitutions within the HMG-

824 box_ROX1-like domain. These variations resulted in an altered tertiary structure under the AlphaFold
825 3D structural code U3NEA9 (*cf.* Figures 14 and 17). In contrast, strain CS70-1212 simultaneously
826 produced the full-length MAT1-2-1 protein AGW27551, which contains only a Y-to-H substitution
827 within the HMG-box_ROX1-like domain (*cf.* Figure S3, Table 2). No AlphaFold-predicted 3D structural
828 model has been assigned to the MAT1-2-1 protein AGW27551.

829 Panel (D) of Figure 17 lists 2 **Group-II** strains, CS70-1211 and CS70-1208; the former has
830 unknown homo/heterogeneous characteristics, and the latter is heterogeneous (Tables 1−2). These strains
831 produced the MAT1-1-1 proteins AGW27530 and AGW27529, whose N- and C-termini were truncated
832 by 97 and 8 residues, respectively. Both proteins contain F-to-V (phenylalanine-to-valine) substitutions
833 within the MATα_HMGbox domains, which conform to the same AlphaFold 3D model U3NE79 labeled
834 with a suffix symbol "‡" in Tables 1−2 (*cf.* Figures 10 and 17). These strains simultaneously produced
835 the full-length MAT1-2-1 proteins AGW27550 and AGW27549, which contain Y-to-H (tyrosine-to-
836 histidine) substitutions within the HMG-box_ROX1-like domains (*cf.* Tables 1−2; Figures 17 and S3)
837 and show 99.6% sequence identity to the authentic MAT1-2-1 protein AEH27625. No AlphaFold-
838 predicted 3D structural models have been assigned to the 2 MAT1-2-1 proteins.

839 **3.5d *O. sinensis* strains simultaneously produced MAT1-2-1 proteins with the same tertiary**
840 **structures and MAT1-1-1 proteins with various tertiary structures**

841 As shown in Table 2, the **Group-I** "homogenous" strain CS68-2-1229 and the heterogeneous strain
842 CS71-1220 produced the 74-residue truncated MAT1-2-1 proteins AGW27548 and AGW27555,
843 respectively, with Y-to-H substitutions within the HMG-box_ROX1-like domains (*cf.* Figures 15 and
844 17). These 2 proteins share the same tertiary structures under the AlphaFold code U3N9W5 labeled with
845 the suffix symbol "⁑" in Tables 1−2. These strains simultaneously produced the MAT1-1-1 proteins
846 AGW27528 and AGW27535, which have 63- and 8-residue truncations and 97- and 8-residue
847 truncations at the N- and C-termini of the proteins, respectively, and form distinct 3D structures under
848 AlphaFold codes U3N9T9 and U3N6U8 labeled with the suffix symbols "⁋" and "†" in Tables 1−2,
849 which were fall differently under Panels (A) and (B) of Figure 17, respectively (*cf.* Figures 5 and 9).

850 Panel (E) in Figure 17 (highlighted by the open rectangle in blue) includes 5 **Group-II** *O. sinensis*
851 strains (CS560-961, CS76-1284, CS561-964, CS25-273, and CS68-2-1228), which produced the full-
852 length MAT1-2-1 proteins (AGW27542, AGW27545, AGW27546, AGW27547, and AGW27556; *cf.*
853 Figure S3), respectively. These MAT1-2-1 proteins contain the same Y-to-H substitutions within the
854 HMG-box_ROX1-like domains. Unfortunately, only the protein AGW27542 was given an AlphaFold

855 3D structural code D7F2E3; the remaining 4 proteins exhibited the same structures as the protein
856 AGW27542 but were not assigned an AlphaFold 3D structural code. These strains simultaneously
857 generated truncated MAT1-1-1 protein variants (AGW27522, AGW27525, AGW27526, AGW27527,
858 and AGW27536; *cf.* Figures 9 and 17; Table 2) with different truncations (63-/8-residue, 66-/10-residue,
859 63-/13-residue, 77-/18-residue, and 95-/8-residue truncations) at the N-/C-termini, respectively, and
860 contained various amino acid substitutions, as shown in Table 2 and Figures 5−8, 11, and 17. The diverse
861 variations result in altered tertiary structures of the MAT1-1-1 proteins under distinct AlphaFold 3D
862 structural codes (U3N6U0, U3N919, U3N7G5, U3N6U4, and U3N7H7) (*cf.* Table 2, Figures 5−8 and
863 11).

864 **3.5e Distinct patterns of variation between the mating-type genes in the genome of pure *H.***
865 ***sinensis* and the mating proteins derived from the *O. sinensis* strains**

866 Table S7 summarizes the co-occurrence and differential occurrence of the MAT1-1-1 and MAT1-
867 2-1 proteins in *C. sinensis* insect–fungal complexes, wild-type *C. sinensis* isolates, and *O. sinensis* strains,
868 including various *O. sinensis* genotypes. Among the 183 total samples, the MAT1-1-1 and MAT1-2-1
869 proteins co-occur in 25.1% of the samples but differentially occur in 52.5% and 22.4% of the samples,
870 respectively.

871 Genomically, both the *MAT1-1-1* and *MAT1-2-1* genes are present in the genome assemblies
872 ANOV00000000 and LKHE00000000 of *H. sinensis* strains Co18 and 1229 (GC-biased Genotype #1 of
873 *O. sinensis*), respectively [64,93]. However, the *MAT1-1-1* gene is absent in the genome assemblies
874 LWBQ00000000 and NGJJ00000000 of the *H. sinensis* strains ZJB12195 and CC1406-20395,
875 respectively, and the *MAT1-2-1* gene is absent in the genome assembly JAAVMX000000000 of the *H.*
876 *sinensis* strain IOZ07 [94−96]. However, this pattern of the differential absence of the *MAT1-1-1* and
877 *MAT1-2-1* genes in the genome of *H. sinensis* was not observed at the protein level, *i.e.*, all the 20 *O.*
878 *sinensis* strains simultaneously produced the MAT1-1-1 and MAT1-2-1 proteins in pairs without
879 omissions. The inconsistency of the absence patterns at the genetic and protein levels indicates divergent
880 fungal origins of the MAT1-1-1 and MAT1-2-1 proteins coproduced by the 20 *O. sinensis* strains.

881 Figure S1 of [65] revealed Y-to-M (tyrosine-to-methionine) substitutions in the MATα_HMGbox
882 domains of MAT1-1-1 proteins encoded by the genome assemblies LKHE00000000 and
883 JAAVMX000000000 of the *H. sinensis* strains 1229 and IOZ07, respectively, but not by the genome
884 assembly ANOV00000000 of the *H. sinensis* strain Co18 [64,65,93,96]. However, Y-to-M substitution
885 was not detected in the MATα_HMGbox domains of variant MAT1-1-1 proteins derived from the 20 *O.*

886 *sinensis* strains.

887 Figure S2 of [65] revealed S-to-A (serine-to-alanine) substitutions in the HMG-box_ROX1-like
888 domains of the MAT1-2-1 proteins encoded by the genome assemblies of ANOV00000000,
889 LKHE00000000, LWBQ00000000, and NGJJ00000000 of *H. sinensis* strains Co18, 1229, ZJB12195,
890 and CC1406-20395, respectively. However, S-to-A substitution does not occur in the HMG-box_ROX1-
891 like domains of MAT1-2-1 proteins, regardless of whether they are truncated or full-length proteins
892 derived from the 20 *O. sinensis* strains.

893 Aggregating the observations presented in Section 3.5, the 20 *O. sinensis* strains simultaneously
894 produced the MAT1-1-1 and MAT1-2-1 proteins in pairs, with differential truncations and diverse amino
895 acid substitutions. These findings highlight the genetic variations within the *MAT* loci in the fungal
896 genomes of impure *O. sinensis* strains, regardless of whether the strains were previously reported as
897 "homogenous" or heterogeneous by Li *et al.* [36]. The apparent differences in the presence/absence of
898 the *MAT1-1-1* and *MAT1-2-1* genes in the genome assemblies of the 5 *H. sinensis* strains were not
899 replicated at the protein level in the coproduction of MAT1-1-1 and MAT1-2-1 proteins in pairs from the
900 20 *O. sinensis* strains. Because repetitive copies of the *MAT1-1-1* or *MAT1-2-1* genes are not present in
901 the *H. sinensis* genome [69], these observations suggest that the paired MAT1-1-1 and MAT1-2-1
902 proteins are not necessarily the products of GC-biased Genotype #1 *H. sinensis* but are instead most
903 likely the products of different co-occurring heterospecific or genotypic fungi within the impure *O.*
904 *sinensis* strains.

905 # 4. Discussion

906 ## 4.1 Altered tertiary structures of the DNA-binding domains of mating proteins
907 produced by various *O. sinensis* strains

908 Li *et al*. [67,68] previously reported the differential occurrence, differential transcription, and
909 alternative splicing of *MAT1-1-1* and *MAT1-2-1* and pheromone receptor genes in *H. sinensis*, wild-type
910 *C. sinensis* isolates, and the *C. sinensis* insect−fungal complex. The full-length MAT1-1-1 and MAT1-
911 2-1 proteins contribute to 15 and 17 heteromorphic stereostructures, belonging to 5 and 5 Bayesian
912 clusters, respectively [69], whereas the DNA-binding domains of the full-length mating proteins derived
913 from wild-type *C. sinensis* isolates clustered into 5 and 2 branched Bayesian clusters, respectively, as
914 summarized in Tables S6−S9 of [65]. Thus, the evidence at the genetic, transcriptional, and protein
915 structural levels suggests that *O. sinensis* is self-sterile and employs a heterothallic, hybrid, or even

916 parasexual strategy to accomplish sexual reproduction during the lifecycle of the *C. sinensis* insect–
917 fungal complex [65,67−69]. The data analyzed in this paper extend the previous observations by focusing
918 on differential truncations and diverse amino acid substitutions within the MATα_HMGbox and HMG-
919 box_ROX1-like domains of the MAT1-1-1 and MAT1-2-1 proteins, respectively, which are coproduced
920 simultaneously in pairs by the 20 *O. sinensis* strains (Figure 17; Tables 1−2). N- and/or C-terminal
921 truncations and 1−4 amino acid substitutions within the DNA-binding domains significantly alter the
922 hydrophobic properties and modify the structures of the functional domains of mating proteins,
923 particularly the tertiary structures of the hydrophobic cores formed by 3 α-helices within the DNA-
924 binding domains. These structural variations ultimately modulate the synergistic interactions between
925 the key functional domains of mating proteins during the sexual reproduction of *O. sinensis*.

926 The HMGbox domains in mating proteins are multifunctional motifs that are central to the
927 transcriptional regulation of mating-related genes, *O. sinensis* sexual reproduction, and host adaptation
928 within the *C. sinensis* insect–fungal complex [59,65,67−69]. It has been inferred that the MATα-
929 HMGBox in the MAT1-1-1 protein has a conserved ancestral mating-type HMG lineage *via* conservative
930 continuation model, whereas the HMG-box_ROX1-like domain within the MAT1-2-1 protein represents
931 evolutionary cooperation of a different regulatory lineage (*e.g.*, ROX1-like) through functional
932 convergence, where it was repurposed to control mating-type specification [62,63,97−100].

933 The MATα_HMGbox and HMG-box_ROX1-like domains of the MAT1-1-1 and MAT1-2-1
934 proteins, respectively, are structurally defined DNA-binding domains characterized by a hydrophobic
935 core containing 3 asymmetrically arranged α-helices. The hydrophobic core, which is composed of
936 tightly packed amino acid residues, is the foundation of the structures and provides a precise
937 conformational stability and constraints that enable the 3 α-helices to form the characteristic L-shaped
938 fold [73,90,91]. These α-helices converge and create stable resistance to environmental perturbations
939 that might unfold or distort the domains. The unique tertiary structures of these sequence-specific DNA
940 binders preferentially ensure high-affinity, sequence-dependent binding to short AT-rich DNA sequences
941 within the promoters of their target mating-type-specific genes [61,101−106]. The sequence-specific
942 binding of the HMG-box domains to AT-rich promoter motifs serves as the initial trigger, thereby playing
943 pivotal roles in controlling fungal mating identity and gene expression [62,97,107,108].

944 Disruption of this hydrophobic core, whether by 1–4 amino acid substitutions at distinct sites that
945 introduce incompatible residues or by differential peptide chain truncations that remove parts of it, as
946 identified in this study, directly impairs structural integrity and abolishes the crucial synergistic functions

947 of the domains in DNA binding specificity in recognizing the target DNA segments and the
948 transcriptional regulation of mating-type-specific genes. These variations have significant consequences
949 for the regulation of *O. sinensis* sexual reproduction during the sexual life of the *C. sinensis* insect−fungal
950 complex. Compared with the authentic MAT1-1-1 and MAT1-2-1 proteins of *O. sinensis* analyzed
951 previously by Li *et al.* [65,69], the differentially altered DNA-binding domains clustered into diverse
952 clades/branches in the Bayesian clustering trees [65,69].

953 Although the MATα_HMGbox domain within the MAT1-1-1 protein and the HMG-box_ROX1-like
954 domain within the MAT1-2-1 protein both belong structurally to the HMG-box superfamily and share a
955 conserved L-shaped fold that enables DNA binding and bending, they are functionally distinct in the
956 context of mating-type regulation in *O. sinensis* because they regulate different gene sets, resulting in
957 opposite regulatory outcomes. The MATα_HMGbox domain of the MAT1-1-1 protein activates the
958 transcription of genes encoding MAT1-1-specific receptors or other regulators. Although the HMG-
959 box_ROX1-like domain of the MAT1-2-1 protein binds to specific DNA sequences in the *MAT1-2*-
960 specific gene promoter encoding the *MAT1-2*-specific pheromone receptor, it may overlap differently
961 with the MATα_HMGbox targets and act through an opposite pathway to counterbalance the activity of
962 the MAT1-1-1 protein [109−111]. Functionally, the MATα_HMGbox domain primarily functions as an
963 activator of its cognate mating-type genes (*i.e., MAT1-1* genes). In contrast, the HMG-box_ROX1-like
964 domain acts as a repressor of genes specific to the opposite mating type (*MAT1-1* or "a-specific" genes)
965 and may also serve as an activator for *MAT1*-2-specific genes. This repressive capacity represents a key
966 functional hallmark derived from its homology to ROX1 [112]. Thus, despite their structural similarity
967 as DNA-binding and DNA-bending HMG-box domains, the MATα_HMGbox and HMG-box_ROX1-
968 like domains are functionally specialized to control distinct, complementary sets of target genes, thereby
969 defining the two different mating types. Their divergent regulatory outputs (activation *vs*. repression of
970 key target gene sets) constitute the molecular and protein structural basis of mating-type specification
971 during the sexual reproduction of *O. sinensis* during the lifecycle of the *C. sinensis* insect–fugal complex.

972 The putative heterodimerization between the MATα_HMGbox domain of the MAT1-1-1 protein and
973 the HMG-box_ROX1-like domain of the MAT1-2-1 protein has been proposed as a potential regulatory
974 mechanism to enforce outbreeding and initiate the sexual cycle in heterothallic fungi [113−115].
975 However, because of the lack of direct evidence in *O. sinensis*, it remains a compelling hypothesis and
976 future research must confirm the occurrence of this heterodimerization, characterize the interaction
977 interface, and define its precise regulatory targets and functional effects on mating and fruiting body

978 development in this economically and ecologically important insect–fungal complex. Nevertheless, the
979 MATα_HMGbox and HMG-box_ROX1-like domains may interact electrostatically and
980 complementarily, providing a plausible physicochemical basis for cooperative synergy in DNA binding.
981 This interaction could coordinate the transition from asexual growth to sexual development, thereby
982 distorting target DNA to repress haploid programming and activating the mating/meiosis genes of *O.*
983 *sinensis* during the sexual life of the *C. sinensis* insect–fungal complex [116]. Nevertheless, direct
984 experimental validation using protein biochemistry and reproductive physiological approaches is
985 warranted to confirm this hypothetical model, clarify its species-specific characteristics, and determine
986 the effects of protein-level variations on the complementary interactions and alterations in the mating
987 function of *O. sinensis* [65,67−69], even as the genome-independent, multifungal, and multigenotypic
988 nature of *O. sinensis* remains to be fully elucidated [9,10,12,13,19,24,27,30,35,37,38,54,117]. Such
989 protein-level variations may provide critical evidence to support or refute the hypotheses of the
990 anamorph–teleomorph linkage and sexual reproduction strategy of *O. sinensis* [14,20,36,45,64,65,
991 67−69,72].

992 4.2 Heterogenous fungal sources of the MAT1-1-1 and MAT1-2-1 proteins 993 simultaneously produced in pairs by the impure *O. sinensis* strains

994 4.2a Phylogenetic heterogeneity of *O. sinensis* strains

995 By employing a dual-fluorescence microscopy technique, Bushley *et al.* [14] observed the
996 multicellular heterokaryotic microstructures of *C. sinensis* ascospores and hyphae, which contain mono-,
997 bi-, tri-, and tetra-nucleate cellular structures. Zhang & Zhang [72] subsequently commented on these
998 heterokaryotic microstructures and expressed doubts regarding whether binucleate (as well as tri- and
999 tetra-nucleate) cells might inconsistently harbor homogenous or heterogeneous hereditary substances.
1000 Molecular biological observations reported by Li *et al.* [12,13,36−38] revealed multiple fungal
1001 components in the ascospores and ascocarps of *C. sinensis* insect–fungal complex and in purportedly
1002 "pure" *H. sinensis* strains. These findings provide indisputable evidence supporting the hypothesis
1003 proposed by Zhang & Zhang [72] that multicellular heterokaryotic cells indeed contain heterogeneous
1004 hereditary substances.

1005 Evidence of culture dependence was reported by Li *et al.* [36], who deposited the ITS sequences of
1006 15 of the 20 *O. sinensis* strains in GenBank. Among the 15 *O. sinensis* strains, 10 were reported as
1007 "homogenous" strains, purportedly harboring only the GC-biased Genotype #1. The remaining 5 strains

1008 were characterized as heterogeneous, containing the ITS sequences of both GC-biased Genotype #1
1009 (Group A) and AT-biased Genotype #17 (Group C). Li *et al*. [36,118,119] hypothesized that AT-biased
1010 genotypic sequences are nonfunctional repetitive copies within the single genome of GC-biased *H.*
1011 *sinensis* and are referred to as "ITS pseudogenes" and "rRNA pseudogenes", which are generated
1012 through repeat-induced point (RIP) mutation. However, subsequent studies have demonstrated that the
1013 sequences of Genotypes #2−17 of *O. sinensis*, including both GC- and AT-biased genotypes, do not
1014 reside in the genomic assemblies of 5 pure *H. sinensis* strains [9,10,12,13,38,120]. Thus, the 17 *O.*
1015 *sinensis* genotypes are actually genomically independent and belong to distinct fungal lineages, although
1016 they may share a common ancestral heredity [24]. These findings directly invalidate the hypotheses that
1017 "RIP" mutation induces AT-biased "ITS pseudogenes" and "rRNA pseudogenes" as nonfunctional
1018 repetitive copies in the genome of GC-biased Genotype #1 *H. sinensis* [36,118,119].

1019 The less sophisticated research methodology employed by Li *et al*. [36] was analyzed in Section
1020 3.5a. The technical limitation led to an imbalance in the amplification efficiency of coexisting fungal
1021 sequences. Consequently, Li *et al*. [36] detected the sequence(s) that exhibited the highest PCR
1022 amplification efficiency, likely precluding the comprehensive detection of all the coexisting fungal taxa
1023 in the samples. Notably, the different abundances (high *vs*. low) of amplicons primarily reflect the level
1024 of PCR amplification efficiency and velocity under specific research settings and, obviously, should not
1025 be equated with the natural ratio of coexisting fungi in the tested samples. The natural ratio of coexisting
1026 fungal taxa can be quantified using other specialized non–PCR techniques, such as Southern blotting
1027 with the use of specific probes [27]. Notably, naturally occurring, less abundant fungal taxa may still
1028 play synergistic roles in the lifecycle of the *C. sinensis* insect–fungal complex.

1029 Furthermore, Li *et al*. [12,13,37] adopted an exhaustive culture-independent approach, including the
1030 use of multiple primers, "touch-down" PCR protocols, and the selection and sequencing of multiple
1031 (>30) white colonies, as well as MALDI-TOF SNP mass spectrum genotyping approaches using multiple
1032 extension primers, to exhaustively investigate the comprehensive fungal composition within *C. sinensis*
1033 ascospores and the "pure" *H. sinensis* strains Hs-Ma-4, Hs7.7, Hs7.8, and Hs-Qing2 (gifts from the
1034 renowned mycologist Prof. Gou YL). These studies demonstrated that *C. sinensis* ascospores and "pure"
1035 *H. sinensis* strains co-occur with genomically independent fungal taxa, including the psychrophilic GC-
1036 biased Genotypes #1 and #14 and AT-biased Genotypes #5−6 and #16 of *O. sinensis*, as well as the
1037 mesophilic *Paecilomyces hepiali* (recently reclassified as *Samsoniella hepiali* [121]). These findings
1038 clearly indicate that the study by Li *et al*. [36] overlooked some coexisting fungal sequences with low

1039 PCR amplification efficiency and nonculturable fungal components under particular experimental
1040 conditions. Thus, the origins of paired MAT1-1-1 and MAT1-2-1 proteins with differential truncations
1041 and various amino acid substitutions at distinct sites require further clarification using rigorously
1042 designed mycological and molecular protocols by the custodians of the original *O. sinensis* strains.

1043 The paired heteromorphic MAT1-1-1 and MAT1-2-1 proteins are likely to have distinct fungal
1044 origins and were possibly derived from the co-occurring GC-biased Genotypes #1 and #3 and AT-biased
1045 Genotype #17 of *O. sinensis* in wild-type *C. sinensis* isolates [65] and the *O. sinensis* strains analyzed in
1046 this study. Given the limited available information on the diverse coexisting genotypes of *O. sinensis*,
1047 the different pairings of the MAT1-1-1 and MAT1-2-1 protein variants that exhibit differential
1048 truncations and diverse amino acid substitutions at distinct sites were coproduced simultaneously by the
1049 differentially coexisting genomically independent fungi in *O. sinensis* strains [6,9−13,15−44].

1050 **4.2b Differential pairings of the MAT1-1-1 and MAT1-2-1 proteins simultaneously produced by**
1051 ***O. sinensis* strains**

1052 Table S7 summarizes the differential occurrence or co-occurrence of both authentic and variant
1053 MAT1-1-1 and MAT1-2-1 proteins in *C. sinensis* insect–fungal complexes, wild-type *C. sinensis* isolates,
1054 and various *O. sinensis* strains (including GC-biased Genotype #1 *H. sinensis* strains). The differential
1055 occurrence pattern of the *MAT1-1-1* and *MAT1-2-1* genes in the genome of the Genotype #1 *H. sinensis*
1056 protein is inconsistent with the genetic expression pattern at the protein level; namely, the MAT1-1-1
1057 and MAT1-2-1 proteins were produced simultaneously by the 20 *O. sinensis* strains without any omission.
1058 The paired MAT1-1-1 and MAT1-2-1 proteins exhibit differential truncations and diverse amino acid
1059 substitutions (*cf.* Tables 2 and S5, Figures 1−17 and S3) [14,36,69]. The distinct pairings of the highly
1060 divergent MAT1-1-1 and MAT1-2-1 protein variants that were simultaneously and differentially
1061 produced by the 20 *O. sinensis* strains present a complex, nonuniform pairing pattern (*cf.* Table 2 and
1062 Figure 17), indicating diverse fungal origins of the mating proteins from co-occurring fungal taxa in the
1063 samples. In addressing the issue of overlooked co-occurring fungal taxa due to the unsophisticated study
1064 designed by Li *et al.* [36] (*cf.* Section 3.5a), namely, the nonculturable fungal taxa and low-abundance
1065 coexisting fungal sequences that arise from less efficient PCR amplification and sequencing bias
1066 associated with the suboptimal experimental design, the 20 *O. sinensis* strains presumably exhibit a much
1067 greater diversity of detected and undetected fungal sequences than the "homogenous" and heterogeneous
1068 cultures of *C. sinensis* monoascospores and tissues of caterpillar bodies previously reported by Li *et al.*
1069 [36]. Further clarification is warranted, employing unbiased and rigorous culture-independent methods

1070 to comprehensively profile coexisting genomically independent genotypes of *O. sinensis* and
1071 heterospecific fungal taxa [12,13,27,35,37].

1072 **4.2c Paradoxical dilemmas regarding the highly possible heterogeneity of purportedly "pure" *O.***
1073 ***sinensis* strains**

1074 The first dilemma concerns the purportedly homogenous strain CS68-2-1229 of **Group I**
1075 (highlighted in dark green and underlined in Figures 2 and 4). Only the GC-biased Genotype #1 was
1076 previously detected in this "pure" culture of *C. sinensis* monoascospores under the experimental settings
1077 adopted by Li *et al.* [36], which was incapable of completely covering all co-occurring fungal taxa (*cf.*
1078 Section 3.5a). Two MAT1-1-1 protein sequences from this strain have been deposited in GenBank: a
1079 full-length MAT1-1-1 protein, AGW27560, containing 372 amino acids, and a truncated MAT1-1-1
1080 protein, AGW27528, containing 301 amino acids (*cf.* Figure 5). These 2 proteins share 100% sequence
1081 identity, yet AGW27528 exhibits only 81% query coverage with 63- and 8-residue truncations at the N-
1082 and C-termini, respectively (*cf.* Figure 17, Table 2). The truncation pattern was likely attributable to the
1083 use of different primers for amplifying the *MAT1-1-1* gene or its transcript from the total RNA pool in
1084 the experiments. This strain simultaneously produced the truncated MAT1-2-1 protein AGW27548 (175
1085 amino acids; *cf.* Figure 15), which shares 99.4% sequence identity with the authentic full-length MAT1-
1086 2-1 protein and has a Y-to-H substitution within the HMG-box_ROX1-like domain; notably, it shows
1087 only 70% sequence coverage because of a 74-residue truncation at the C-terminus relative to that of the
1088 authentic full-length MAT1-2-1 proteins, which contain 249 amino acids. All 3 mating protein sequences
1089 derived from strain CS68-2-1229 were released in GenBank on the same date (28-SEP-2013). The
1090 truncated MAT1-1-1 AGW27528 and MAT1-2-1 protein AGW27548 were derived from cDNAs that
1091 were amplified using the primer pair *m1F3*/*m1R3* to amplify *MAT1-1-1* cDNA and the primer pair *Mat1-*
1092 *2F*/*Mat1-2R* to amplify *MAT1-2-1* cDNA from the same batch of total RNA pool extracted from the
1093 strain CS68-2-1229 (*cf.* Figures S1−S2) [14,36,70]. Figure 2 shows that the MATα_HMGbox domain
1094 (162 amino acids) of the MAT1-1-1 protein AGW27528 is among the longest domains of all 9 truncated
1095 MAT1-1-1 proteins. In contrast, Figure 4 shows that the HMG-box_ROX1-like domain (49 amino acids)
1096 of the MAT1-2-1 protein AGW27548 (*cf.* Figures 4 and 15) is 10 residues shorter than the longest
1097 truncated HMG-box_ROX1-like domain (59 amino acids) of the MAT1-2-1 protein AGW27543 [65,70].
1098 The truncated MAT1-2-1 protein AGW27548, which harbors a Y-to-H substitution, exhibits altered
1099 secondary and tertiary structures under the AlphaFold 3D structural code U3N9W5 (*cf.* Figure 15, Table
1100 2). Li *et al.* [69] reported no repetitive copies of the *MAT1-1-1* and *MAT1-2-1* genes in the genome of

1101 the *H. sinensis* strain. These findings suggest that the MAT1-1-1 protein AGW27528 is likely encoded
1102 by the authentic *MAT1-1-1* gene in the *H. sinensis* genome. However, the MAT1-2-1 protein AGW27548
1103 is presumably either encoded by a variant *MAT1-2-1* gene within the *H. sinensis* genome or by an
1104 independent co-occurring fungal taxon in the actual "impure" *O. sinensis* strain CS68-2-1229. Thus, the
1105 true genetic characteristics of strain CS68-2-1229 require re-evaluation: whether this strain is genuinely
1106 homogenous, containing only Genotype #1 *H. sinensis*, or is indeed heterogeneous, harboring 2 or more
1107 fungal taxa, remains to be determined.

1108 Another paradoxical "homogenous" *O. sinensis* strain, CS561-964 (**Group II**), was also reported
1109 to harbor only GC-biased Genotype #1 *H. sinensis* [14,36]. This strain produced the MAT1-1-1 protein
1110 AGW27526, featuring 63- and 13-residue truncations at the N- and C-termini, respectively, along with
1111 4 amino acid substitutions (CDRA-to-SSEF) within the MATα_HMGbox domains (*cf.* Figures 7 and 17,
1112 Tables 1−2). Clearly, the altered protein AGW27526 is encoded by a variant *MAT1-1-1* gene, plausibly
1113 from a genome-independent, co-occurring fungal taxon in the *O. sinensis* strain. The ITS sequence of
1114 this fungal taxon was most likely overlooked in the study by Li *et al.* [36] because of the use of
1115 suboptimal culture-dependent experimental protocols (*cf.* Section 3.5a).

1116 Analyses of the purportedly homogenous *O. sinensis* strains raise fundamental questions regarding
1117 fungal purification, *in vitro* culture characteristics, and comprehensive profiling of all co-occurring
1118 fungal taxa in such cultures. Two predicaments remain unresolved: First, do the purportedly homogenous
1119 strains listed in Tables 1−2 truly harbor only GC-biased Genotype #1, or are they indeed heterogeneous
1120 and contain 2 or more fungal taxa? Second, do the characterized heterogeneous strains contain only the
1121 GC-biased Genotype #1 and AT-biased Genotype #17 or an even greater number of undetected coexisting
1122 fungal taxa? A reasonable hypothesis is that the differentially paired truncated and full-length MAT1-1-
1123 1 and MAT1-2-1 protein variants, which exhibit diverse amino acid substitutions at distinct sites of
1124 variation, are not necessarily produced by GC-biased Genotype #1 *H. sinensis*. Instead, these protein
1125 variants likely originate from AT-biased Genotype #17 (*cf.* Tables S1−S2) or from other uncharacterized
1126 co-occurring fungal taxa that were not detected in the culture-dependent, suboptimal study by Li *et al.*
1127 [36].

1128 4.3 Self-sterility of *O. sinensis* and the requirement for sexual partners to 1129 accomplish heterothallic or hybrid reproduction

1130 Wei *et al.* [45] proposed that *H. sinensis* (GC-biased Genotype #1 of *O. sinensis*) represents the sole
1131 anamorph of *O. sinensis* teleomorph. Zhang *et al.* [51] proposed the implementation of the "One

1132 Fungus=One Name" nomenclature principle established by the International Mycological Association
1133 [52,53] and replaced the anamorphic name *H. sinensis* with the teleomorphic name *O. sinensis* [45]. This
1134 renaming implementation, however, was not only applied to GC-biased Genotype #1 *H. sinensis*, which
1135 was used as the nomenclature reference, but was also uniformly applied to all 17 genomically
1136 independent genotypes of *O. sinensis*, although they may share a common hereditary ancestor
1137 [9,10,24,54]. Eleven of the 17 genotypes are GC biased, whereas the remaining 6 genotypes are AT
1138 biased. The genotypes in various combinations coexist in different compartments of the *C. sinensis*
1139 insect–fungal complex, and their natural abundance changes dynamically and reciprocally during *C.*
1140 *sinensis* maturation, indicating a genomically independent nature [9,10,12,13,27,35,37]. In support of
1141 the sole anamorph hypothesis for *H. sinensis* [45], Bushley *et al.* [14] and Hu *et al.* [64] studied the
1142 mating-type genes in the genome of GC-biased Genotype #1 *H. sinensis* and proposed the self-
1143 fertilization hypothesis for *O. sinensis* to accomplish homothallic or pseudohomothallic reproduction.
1144 However, several research groups [64,122,123] have reported unsuccessful attempts to artificially
1145 cultivate the fruiting bodies and ascospores of *C. sinensis* insect–fungal complexes using pure fungal
1146 cultures. Zhang *et al.* [51] then summarized the 40-year history of unsuccessful attempts to cultivate
1147 insect–fungal complexes in academic research-oriented settings. Qin *et al.* [124] elaborated on the key
1148 obstacles hindering the artificial cultivation of *C. sinensis* fruiting bodies and ascospores. Li *et al.* [125]
1149 further attributed the decades-long history of failures in the cultivation of *C. sinensis* insect–fungal
1150 complexes to "the low rate of the fungal invasion to host insects and primordium induction after
1151 infection". Compared with the nearly noninfectious characteristics of pure *H. sinensis*, Li *et al.* [37]
1152 advanced an inoculation technique using cocultures of natural fungal clusters with a natural abundance
1153 ratio of multiple genotypic and heterospecific fungal taxa isolated from the *C. sinensis* insect–fungal
1154 complex. This advanced technique resulted in up to 39-fold greater infectivity in the larvae of *Hepialus*
1155 *armoricanus* and significantly shortened the latency (from 35−50 days to 5−8 days) from inoculation to
1156 mummification/death. Evidently, the hypotheses of contentious sole anamorph and concomitant self-
1157 fertilization have not yet been scientifically validated in academia to meet all 4 criteria of Koch's
1158 postulations.

1159 In contrast to the hypothesis that *H. sinensis* represents the sole anamorph of *O. sinensis* [45],
1160 molecular studies have revealed complex genetic diversity of *O. sinensis,* and this Latin name
1161 encompasses at least 17 distinct genotypes [9,10,12,13,27,30,35,54]. The sequences of Genotypes #2−17
1162 of *O. sinensis* are not present in the genome assemblies of 5 Genotype #1 *H. sinensis* strains, indicating

1163 that they represent diverse genome-independent fungal taxa [9,10,24,30]. A medicinal mycology group
1164 led by Prof. S.P. Wasser reported that *H. sinensis* and *Tolypocladium sinensis* serve as conjoint
1165 anamorphs of *O. sinensis* [34], supporting the prior detection of *T. sinensis* and *T. sinense* from natural
1166 *C. sinensis* insect–fungal complexes [15,20,21,25]. Quandt *et al.* [126] subsequently proposed
1167 taxonomic revisions within the genus *Tolypocladium*, reassigning several species to the family
1168 Ophiocordycipitaceae. To date, the taxonomic positions of 17 evolutionarily related genotypes as
1169 independent fungi, as well as the precise anamorph–teleomorph relationship of *O. sinensis*, remain
1170 unresolved. With respect to reproductive modes, although early hypotheses propose homothallic or
1171 pseudohomothallic reproduction for *O. sinensis* [10,58], Zhang and Zhang [72] reported the differential
1172 presence of *MAT1-1-1* and *MAT1-2-1* genes across numerous wild-type *C. sinensis* isolates and proposed
1173 a facultative hybridization hypothesis for *O. sinensis*. Li *et al.* [67,68] further documented the differential
1174 occurrence, alternative splicing, and differential translation of the *MAT1-1-1* and *MAT1-2-1* genes and
1175 the pheromone receptor genes. More recently, the studies by Li *et al.* [65,69] and the present study
1176 revealed heteromorphic tertiary structures of the MAT1-1-1 and MAT1-2-1 protein variants, particularly
1177 variations within the DNA-binding domains of the mating proteins. These structural observations
1178 provide protein-level evidence for heterogeneous fungal origins of mating-type loci in numerous wild-
1179 type *C. sinensis* isolates and *O. sinensis* strains. Notably, Li *et al.* [69] reported that both the *MAT1-1-1*
1180 and *MAT1-2-1* genes exist as single copies in the *H. sinensis* genome with no repetitive genomic copies.
1181 Collectively, the evidence from genetics, transcription, and protein structure is inconsistent with the self-
1182 fertilization hypothesis for *O. sinensis* and instead supports self-sterility, implying that *O. sinensis* relies
1183 on heterothallic mating, hybridization, or even parasexual outcrossing to accomplish sexual reproduction
1184 [13,65,67−69,127−136]. Similar self-sterile reproductive strategies have been documented in many other
1185 *Ophiocordyceps* species, which require compatible mating partners to undergo sexual reproduction
1186 [104,137−141].
1187 Wei *et al.* [46] reported the successful industrial cultivation of *C. sinensis* insect–fungal complexes.
1188 However, this study revealed a puzzling species paradox between the input and output of the artificial
1189 cultivation project: 3 GC-biased *H. sinensis* strains were used as inoculants, yet only the teleomorph of
1190 AT-biased Genotype #4 was detected in the fruiting bodies of the cultivated *C. sinensis* insect–fungal
1191 complexes. Hence, the sole teleomorphic AT-biased Genotype #4 identified in the cultivated fruiting
1192 bodies replaced the inoculant GC-biased Genotype #1 strains, resulting in a species contradiction. The
1193 sequences of Genotypes #2−17 of *O. sinensis* are not present in the genome assemblies ANOV00000000,

1194 JAAVMX000000000, LKHE00000000, LWBQ00000000, or NGJJ00000000 of the pure *H. sinensis*
1195 strains Co18, IOZ07, 1229, ZJB12195, and CC1406-20395, respectively [9,10,12,13,24,30,35,36,38,
1196 64,93−96,117,120]. Instead, these genotypic sequences are associated with distinct genome-independent
1197 fungal taxa. These results invalidate the unfounded assumptions proposed by Li *et al.* [36,118,119] that
1198 AT-biased genotypes are nonfunctional "ITS pseudogenes" or "rRNA pseudogenes" that were generated
1199 through "RIP" mutation in the *H. sinensis genome*. Supporting evidence at the genomic and
1200 transcriptomic levels is not available to validate these unfounded speculations [38,120]. In contrast to
1201 the significantly variable maturation-related dynamics of AT-biased genotypes of *O. sinensis*, GC-biased
1202 *O. sinensis* genotypes maintain a relatively constant low abundance in the stromata throughout the entire
1203 maturation course of the *C. sinensis* insect–fungal complex [27,35]. These observations suggest that the
1204 17 genome-independent genotypes of *O. sinensis*, regardless of whether they are GC or AT biased,
1205 represent distinct fungal species [10,12,13,30], further reinforcing the invalidation of the RIP
1206 mutation/pseudogene hypotheses proposed by Li *et al.* [36,38,118−120]. The successful industrial
1207 cultivation of *C. sinensis* insect–fungal complexes, coupled with the apparent species paradox, indicates
1208 that the sole anamorph hypothesis for *H. sinensis* proposed by Wei *et al.* [46] failed to satisfy all 4 criteria
1209 of Koch's postulates.

1210 Li *et al.* [125] subsequently clarified that the inoculant used in the industrial cultivation project was
1211 an *O. sinensis* conidial suspension obtained after 3 months of culture of *C. sinensis* ascospores, rather
1212 than the 3 pure *H. sinensis* strains initially described by Wei *et al.* [46]. However, Li *et al.* [125] did not
1213 disclose the molecular components of the inoculant to the public or redress the previously reported
1214 teleomorphic AT-biased Genotype #4 of *O. sinensis* in the fruiting bodies of cultivated *C. sinensis* insect–
1215 fungal complexes. On the other hand, several other studies reported that AT-biased Genotype #4 was
1216 detected in the stromata throughout all maturation stages of the *C. sinensis* insect–fungal complexes
1217 [10,27,35,36,38]. The abundance of Genotype #4 markedly altered ontogenetic dynamics; *i.e.*, it was
1218 highly abundant in immature stromata during the asexual growth phase and was substantially reduced in
1219 the mature stromata and the stromal fertile portion of the *C. sinensis* insect–fungal complex during the
1220 sexual reproduction stage or the transition from asexual growth to sexual reproduction
1221 [9,10,12,13,35,38]. Regrettably, several studies [12,13,36,37] failed to detect AT-biased Genotype #4
1222 from *C. sinensis* ascospores using both culture-dependent and culture-independent approaches. Thus, the
1223 dynamic maturation pattern for AT-biased Genotype #4 of *O. sinensis* and its absence in *C. sinensis*
1224 ascospores suggest that Genotype #4 may not participate in the sexual reproduction of *O. sinensis*.

1225 Consequently, the puzzling species paradox between the inoculant and the sole teleomorphic AT-biased
1226 Genotype #4 of *O. sinensis* detected in the fruiting body of the cultivated *C. sinensis* insect–fungal
1227 complex [46] remains unresolved. Unfortunately, Li *et al.* [125] did not address the genetic complexity
1228 of the ascospores of cultivated insect–fungal complexes.

1229 Li *et al.* [65] discussed the evidence related to the industrial success of cultivating the insect–fungal
1230 complex and postulated that the success of the industrial cultivation project may be attributed to the
1231 adaptation of a "mycological impurity" strategy, facilitating the synergistic interplay of coexisting fungi.
1232 The supporting evidence includes the following: (1) Li *et al.* [125] reported that the inoculant was a
1233 culture of *C. sinensis* ascospores to address the weak infectivity of GC-biased Genotype #1 *H. sinensis*.
1234 Li *et al.* [12,13,36] reported that *C. sinensis* ascospores harbor multiple genotypic and heterospecific
1235 fungi. (2) *C. sinensis* ascospores do not contain AT-biased Genotype #4 of *O. sinensis* [10,12,13,36,38].
1236 Therefore, the origin of teleomorphic Genotype #4 detected in the fruiting body of the cultivated *C.*
1237 *sinensis* insect–fungal complex by Wei *et al.* [46] might be derived from cocultures of *C. sinensis*
1238 ascospores with the stroma and/or stromal fertile portion of the *C. sinensis* insect–fungal complex that
1239 indeed contain AT-biased Genotype #4 of *O. sinensis*, along with other genotypic and heterospecific
1240 fungi. Li *et al.* [37] demonstrated that cocultures of multiple genotypic and heterospecific fungal taxa in
1241 wild-type *C. sinensis* isolates exhibit up to 39-fold greater infectivity in the larvae of *Hepialus*
1242 *armoricanus* than those of pure *H. sinensis.* (3) The industrial cultivation system was supplemented with
1243 soil collected from natural *C. sinensis* production areas on the Qinghai–Tibet Plateau [46].

1244 To date, the purification and genomic analysis of GC- and AT-biased Genotypes #2–17 of *O. sinensis*
1245 have not been documented. Thus, genomic and transcriptomic information for the divergent *MAT1-1-1*
1246 and *MAT1-2-1* genes in Genotypes #2–17 of *O. sinensis* is lacking, especially concerning the critical
1247 DNA-binding domains of the MAT1-1-1 and MAT1-2-1 proteins. Regardless of the precise biophysical
1248 details of their interaction, whether the paired MAT1-1-1 and MAT1-2-1 proteins form heterodimers or
1249 interact *via* complementary electrostatic surfaces, the synergy of paired mating proteins serves as the
1250 central and indispensable molecular trigger for mating compatibility and the initiation of sexual
1251 reproduction in *O. sinensis* [56–60]. Molecular evidence regarding the differential occurrence,
1252 alternative splicing, and differential transcription of mating-type and pheromone receptor genes in *H.*
1253 *sinensis* unequivocally indicates that *O. sinensis* is self-sterile and relies on heterothallism or hybrid
1254 outcrossing to complete sexual reproduction, thereby ruling out homothallism (self-fertilization) as a
1255 viable mechanism in this species [14,45,63,66–69,142]. Furthermore, the discovery of heteromorphic

1256 tertiary structures of variant MATα_HMGbox domains (in MAT1-1-1 proteins) and HMG-box_ROX1-
1257 like domains (in MAT1-2-1 proteins) provides protein-level validation that (1) self-fertilization is
1258 biochemically unfeasible and (2) coordinated heterothallic or hybrid reproduction (even parasexuality)
1259 is enforced through structural complementarity between the paired mating proteins, which are likely
1260 generated by divergent heterogeneous *O. sinensis* strains. This self-sterility hypothesis resolves the
1261 longstanding uncertainty in the cultivation of the *C. sinensis* insect–fungal complex, namely, the failure
1262 of single-species cultures to produce fruit, and underscores the role of the synergistic effect of multiple
1263 fungal symbionts on its ecological success throughout the lifecycle of the naturally occurring or
1264 artificially cultivated *C. sinensis* insect–fungal complex.

1265 # Conclusions

1266 This study presents evidence at the protein structure level for self-sterile *O. sinensis* and further
1267 verifies the heterothallic or hybrid reproduction of *O. sinensis*. The diverse pairings of MAT1-1-1 and
1268 MAT1-2-1 protein variants coproduced simultaneously by the 20 *O. sinensis* strains examined in this
1269 study exhibit differential truncations and 1−4 amino acid substitutions in their DNA-binding domains.
1270 The primary structural variations caused dramatic changes in the hydrophobic properties and secondary
1271 and tertiary structures of the MATα_HMGbox and HMG-box_ROX1-like domains of the differentially
1272 paired MAT1-1-1 and MAT1-2-1 proteins, respectively. The structural heterogeneity of differentially
1273 paired mating proteins not only indicates the co-occurrence of heterogeneous fungal origins in impure
1274 *O. sinensis* strains but also invalidates the hypothesis of homothallic or pseudohomothallic self-
1275 fertilization in *O. sinensis*. Complementing the genetic and transcriptional evidence of differential
1276 occurrence, differential transcription, and alternative splicing of the *MAT1-1-1* and *MAT1-2-1* genes and
1277 pheromone receptor genes, the variations in the protein structures of these DNA-binding domains of the
1278 mating proteins impair the homologous interaction required for homothallic or pseudohomothallic self-
1279 fertilization. Collectively, our findings of differential pairings of the mating protein variants in *O.*
1280 *sinensis* strains clarify the sexual reproduction mechanism of *O. sinensis* and confirm that the selection
1281 of suitable heterologous mating partners is an essential natural strategy for self-sterile *O. sinensis* to
1282 complete sexual reproduction during the lifecycle of *Cordyceps sinensis* insect–fungal complexes on the
1283 Qinghai–Tibet Plateau.

1284 # Acknowledgments

1285 The authors are grateful to Prof. Mu Zang, Prof. Ru-Qin Dai, and Prof. Zong-Qi Liang for their consultation.

Supplementary Materials: The following supporting information can be downloaded at www.xxxxxx.

Figure S1. Alignment of the DNA sequence of the *MAT1-1-1* gene KC437356 (6530→7748) and the encoded sequence of the MAT1-1-1 protein AGW27560 (1→ 372) derived from *Ophiocordyceps sinensis* strain CS68-2-1229 [14]. Residues in pink indicate the MATα_HMGbox domain of the *MAT1-1-1* gene and MAT1-1-1 protein; those in green indicate introns I and II. The underlined residues in red indicate the forward primer m1F3/*MAT1-1-1* (CCACTAGGCAGACCAAAGAAG) and the reverse complementary primer m1R3/*MAT1-1-1* (CGCAAAGTAAAAGTCGTCCAGA) reported by [14]. The triplet residues (TAG) in red represent the stop codon.

Figure S2. Alignment of the *MAT1-2-1* gene HM212637 (3241→4097) and the encoded sequence of the MAT1-2-1 protein AEH27625 (1→249) derived from *Ophiocordyceps sinensis* strain CS2 [70]. Residues in pink indicate the HMG-box_ROX1-like domain of the *MAT1-2-1* gene and MAT1-2-1 proteins; those in green indicate introns I and II. The underlined residues in red indicate the forward primer *Mat1-2F*/*MAT1-2-1* (TGGAATGCGACTGACTACGA) and the reverse complementary primer *Mat1-2R*/*MAT1-2-1* (CCAGGAGAGCTTGCTTGACT) reported previously [14,70]. The triplet residues (TAA) in red represent the stop codon.

Figure S3. Alignment of the sequences of full-length MAT1-2-1 proteins derived from **Group II** *O. sinensis* strains listed in Table 2 with that of the authentic MAT1-2-1 protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2. The residues in pink indicate the HMG-box_ROX1-like domains of the proteins, those in green indicate the variant amino acids, the hyphens indicate identical amino acid residues, and the space denotes the unmatched sequence gap.

Table S1. The *O. sinensis* strains, GenBank accession numbers for the ITS nucleic acid sequences corresponding to the GenBank accession numbers for the MAT1-1-1 and MAT1-2-1 proteins, and percent similarities *vs*. GC-biased Genotypes #1–3, #7–10, and #12 of *O. sinensis*.

Table S2. The *O. sinensis* strains, GenBank accession numbers for the ITS nucleic acid sequences corresponding to the GenBank accession numbers for the MAT1-1-1 and MAT1-2-1 proteins, and percent similarities *vs*. AT-biased Genotypes #4–6 and #15–17 of *O. sinensis*.

Table S3. Amino acid scales based on the general chemical characteristics of their side chains for ProtScale analysis (https://web.expasy.org/protscale/) to predict the hydrophobicity and secondary structures (α-helices, β-sheets, β-turns, and coils) of proteins.

Table S4. GenBank accession numbers (in red in parentheses) for the full-length MAT1-1-1 proteins in the AlphaFold database under the corresponding AlphaFold UniProt codes [87].

Table S5. GenBank accession numbers (in red) for the full-length MAT1-2-1 proteins of 69 *H. sinensis* strains or wild-type *C. sinensis* isolates under the corresponding AlphaFold UniProt codes [87].

Table S6. Similarity of the 3 pairs of primers that were used by Li *et al*. [36] to amplify the ITS sequences of the *O. sinensis* strains listed in Table 1 that were isolated either from the caterpillar body specimens that were collected from various production areas on the Qinghai–Tibet Plateau or from the cultures of mono-ascospores that were collected from mature *C. sinensis* insect–fungal complexes obtained from a single production region (Maqên, Guoluo, Qinghai Provence of China).

Table S7. Co-occurrence or differential occurrence of the MAT1-1-1 and MAT1-2-1 proteins in the *O. sinensis* strains, wild-type *C. sinensis* isolates, and *C. sinensis* insect–fungal complex listed in the GenBank database.

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

Funding: This research was funded by (1) the Chinese Academy of Sciences−People's Government of Qinghai Province on Sanjiangyuan National Park (#LHZX-2022-01); (2) Forestry and Grassland Bureau of Qinghai Province: "Wild *Ophiocordyceps sinensis* Identification and Application Project" (QHRD-2025-004); (3) Department of Science and Technology of Qinghai Province: "Process Optimization and Application of Antioxidant Performance of Yushu *Cordyceps sinensis* Extract" (2025-NK-P45); and (4) the Shaanxi Key Laboratory of Natural Product & Chemical Biology Open Foundation (SXNPCB 2024003).

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

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

Data Availability Statement: All sequence and 3D structure data are available in the GenBank and AlphaFold databases.

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

[REFERENCES]

1. Zhu, J.-S.; Halpern, G.M.; Jones, K. The scientific rediscovery of a precious ancient Chinese herbal regimen: *Cordyceps sinensis*: Part I. *J. Altern. Complem. Med.* **1998a**, *4*, 289–303. https://doi.org/10.1089/acm.1998.4.3-289.
2. Zhu, J.-S.; Halpern, G.M.; Jones, K. The scientific rediscovery of an ancient Chinese herbal medicine: *Cordyceps sinensis*: Part II. *J. Altern. Complem. Med.* **1998b**, *4*, 429–457. https://doi.org/10.1089/acm.1998.4.429.
3. Zhu, J.-S.; Li, C.-L.; Tan, N.-Z.; Berger, J.L.; Prolla, T.A. Combined use of whole-gene expression profiling technology and mouse lifespan test in anti-aging herbal product study. In Proceedings of the 2011 New TCM Products Innovation and Industrial Development Summit, Hangzhou, China, 27 November 2011; pp. 443–448. Available online: https://xueshu.baidu.com/usercenter/paper/show?paperid=08341c17fa58c8f85584b92572b90f75&site=xueshu_se (accessed on 30 January 2025).
4. Song, L.-R.; Hong, X.; Ding, X.-L.; Zang, Z.-Y. A comprehensive dictionary of modern pharmacy of Chinese traditional medicine. People's Medical Press, Beijing. 2001; pp 733-737.
5. 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. 7 September 2021. Available online: https://m.163.com/dy/article/HHCVOJPU055360T7.html (accessed on 3 May 2025).
6. Ren, Y.; Wan, D.-G.; Lu, X.-M.; Guo, J.-L. The study of scientific name discussion for TCM Cordyceps. *Lis Shenzhen Med. Mater. Medica Res.* **2013**, *24*, 2211−2212.
7. 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. https://doi.org/10.1111/mec.12940.
8. Lu, H.-L.; St. Leger, R.J. Chapter Seven—Insect Immunity to Entomopathogenic Fungi. In *Advances in Genetics*; Lovett, B., St. Leger, R.J., Eds.; Academic Press: Cambridge, MA, USA, 2016; Volume 94, pp. 251–285.
9. 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.
10. Li, Y.-L.; Li, X.-Z.; Yao, Y.-S.; Xie, W.-D.; Zhu, J.-S. Molecular identification of *Ophiocordyceps sinensis* genotypes and the indiscriminate use of the Latin name for the multiple genotypes and the natural insect–fungi complex. *Am. J. BioMed. Sci.* **2022**, *14*, 115–135. https://doi.org/10.5099/aj220300115.
11. 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.

12. Li, Y.-L.; Gao, L.; Yao, Y.-S.; Wu, Z.-M.; Lou, Z.-Q.; Xie, W.-D.; Wu, J.-Y.; Zhu, J.-S. Altered GC- and AT-biased genotypes of *Ophiocordyceps sinensis* in the stromal fertile portions and ascospores of natural *Cordyceps sinensis*. *PLoS ONE.* **2023d**, *18*, e0286865. https://doi.org/10.1371/journal.pone.0286865.
13. Li, Y.-L.; Li, X.-Z.; Yao, Y.-S.; Wu, Z.-M.; Gao, L.; Tan, N.-Z.; Lou, Z.-Q.; Xie, W.-D.; Wu, J.-Y.; Zhu, J.-S. Differential cooccurrence of multiple genotypes of *Ophiocordyceps sinensis* in the stromata, stromal fertile portion (ascocarps) and ascospores of natural *Cordyceps sinensis*. *PLoS ONE* **2023e**, *18*, e0270776. https://doi.org/10.1371/journal.pone.0270776.
14. Bushley, K.E.; Li, Y.; Wang, W.-J.; Wang, X.-L.; Jiao, L.; Spatafora, J.W.; 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*, 599–610. https://doi.org/10.1016/j.funbio.2013.06.001.
15. Li, C.-L. A study of *Tolypocladium sinense* C.L. Li. sp. nov. and cyclosporin production. *Acta Mycol. Sin.* **1988**, *7*, 93–98.
16. 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*, 221–224.
17. Dai, R.-Q.; Li, X.-M.; Shao, A.-J.; Lin, S.-F.; Lan, J.-L.; Chen, W.-H.; Shen, C.-Y. Nomenclatural validation of *Paecilomyces hepiali*. *Mycosystema* **2008**, *27*, 641–644.
18. Liu, X.-J.; Guo, Y.-L.; Yu, Y.-X.; Zeng, W. Isolation and identification of the anamorph of *Cordyceps sinensis* fungus. *Acta Mycol. Sin.* **1989**, *8*, 35–40.
19. Kinjo, N.; Zang, M. Morphological and phylogenetic studies on *Cordyceps sinensis* distributed in southwestern China. *Mycoscience* **2001**, *42*, 567–574. https://doi.org/10.1007/BF02460956.
20. Jiang, Y.; Yao, Y.-J. A review for the debating studies on the anamorph of *Cordyceps sinensis*. *Mycosistema* **2003**, *22*, 161–176.
21. 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. https://doi.org/10.1016/S0378-1097(03)00889-9.
22. Chen, C.-S.; Hseu, R.-S.; Huang, C.-T. Quality Control of *Cordyceps sinensis* Teleomorph, Anamorph, and Its Products, Chapter 12. In *Quality Control of Herbal Medicines and Related Areas*; Shoyama, Y., Ed.; InTech: Rijeka, Croatia, **2011**; pp. 223–238. Available online: www.intechopen.com (accessed on 3 May 2024).
23. Stensrud, Ø.; Hywel-Jones, N.L.; Schumacher, T. Towards a phylogenetic classification of Cordyceps: ITS nrDNA sequence data confirm divergent lineages and paraphyly. *Mycol. Res.* **2005**, *109*, 41–56. https://doi.org/10.1017/s095375620400139x.
24. Stensrud, Ø.; Schumacher, T.; Shalchian-Tabrizi, K.; Svegardenib, I.B.; Kauserud, H. Accelerated nrDNA evolution and profound AT bias in the medicinal fungus *Cordyceps sinensis*. *Mycol. Res.* **2007**, *111*, 409–415. https://doi.org/10.1016/j.mycres.2007.01.015.
25. 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. https://doi.org/10.1111/j.1365-2672.2006.02930.x.
26. 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.
27. 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*, 217–238. https://doi.org/10.5099/aj100300217.
28. Yang, J.-L.; Xiao, W.; He, H.-X.; Zhu, H.-X.; Wang, S.-F.; Cheng, K.-D.; Zhu, P. Molecular phylogenetic analysis of *Paecilomyces hepiali* and *Cordyceps sinensis*. *Acta Pharmaceut. Sinica.* **2008**, *43*, 421–426.

1412 29. Yang, J.-Y.; Tong, X.-X.; He, C.-Y.; Bai, J.; Wang, F.; Guo, J.-L. Comparison of endogenetic microbial community diversity
1413 between wild *Cordyceps sinensis*, artificial *C. sinensis* and habitat soil. *Chin. J. Chin. Mater. Medica* **2021**, *46*, 3106–3115.
1414 30. Xiao, W.; Yang, J.-P.; Zhu, P.; Cheng, K.-D.; He, H.-X.; Zhu, H.-X.; Wang, Q. Non-support of species complex hypothesis
1415 of *Cordyceps sinensis* by targeted rDNA-ITS sequence analysis. *Mycosystema* **2009**, *28*, 724–730.
1416 31. Zhang, Y.-J.; Sun, B.-D.; Zhang, S.; Wàngmǔ Liu, X.-Z.; Gong, W.-F. Mycobiotal investigation of natural *Ophiocordyceps*
1417 *sinensis* based on culture-dependent investigation. *Mycosistema* **2010a**, *29*, 518–527.
1418 32. Zhang, Y.-J.; Zhang, S.; Wàngmǔ; Bai, F.-Y.; Liu, X.-Z. High Diversity of the Fungal Community Structure in Naturally-
1419 Occurring *Ophiocordyceps sinensis*. *PLoS ONE* **2010b**, 5(12), e15570. doi:10.1371/journal.pone.0015570
1420 33. Zhang, S.-W.; Cen, K.; Liu, Y.; Zhou, X.-W.; Wang, C.-S. Metatranscriptomics analysis of the fruiting caterpillar fungus
1421 collected from the Qinghai-Tibetan plateau. *Sci. Sinica Vitae* **2018**, *48*, 562–570.
1422 34. Barseghyan, G.S.; Holliday, J.C.; Price, T.C.; Madison, L.M.; Wasser, S.P. Growth and cultural-morphological
1423 characteristics of vegetative mycelia of medicinal caterpillar fungus *Ophiocordyceps sinensis* G.H. Sung et al. (Ascomycetes)
1424 Isolates from Tibetan Plateau (P. R. China). *Intl. J. Med. Mushrooms* **2011**, *13*, 565–581.
1425 https://doi.org/10.1615/intjmedmushr.v13.i6.90.
1426 35. Gao, L., Li, X.-H., Zhao, J.-Q., Lu, J.-H., Zhao, J.-G., Zhu, J.-S. Maturation of *Cordyceps sinensis* associates with alterations
1427 of fungal expressions of multiple *Ophiocordyceps sinensis* mutants with transition and transversion point mutations in
1428 stroma of *Cordyceps sinensis*. *Beijing Da Xue Xue Bao*. **2012**, *44*, 454–463. doi: 10.3969/j.issn.1671-167X.2012.03.025
1429 36. Li, Y.; Jiao, L.; Yao, Y.-J. Non-concerted ITS evolution in fungi, as revealed from the important medicinal fungus
1430 *Ophiocordyceps sinensis*. *Mol. Phylogenet. Evol.* **2013**, *68*, 373–379. https://doi.org/10.1016/j.ympev.2013.04.010.
1431 37. 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
1432 from the intestines of *Hepialus lagii* larvae in increasing infection potency. *J. Fungal Res.* **2016b**, *14*, 96–112.
1433 38. 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
1434 hypothesis" for *Ophiocordyceps sinensis*. *Mol. Phylogenet. Evol*. **2020b**, *146*, 106728.
1435 https://doi.org/10.1016/j.ympev.2019.106728
1436 39. 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.; et al.
1437 Transcriptomic insight into the immune defenses in the ghost moth, Hepialus xiaojinensis, during an *Ophiocordyceps*
1438 *sinensis* fungal infection. Insect Biochem. *Mol. Biol.* **2015**, *64*, 1–15. https://doi.org/10.1016/j.ibmb.2015.06.014.
1439 40. Xia, F.; Liu, Y.; Shen, G.-L.; Guo, L.-X.; Zhou, X.-W. Investigation and analysis of microbiological communities in natural
1440 *Ophiocordyceps sinensis*. *Can. J. Microbiol.* **2015**, *61*, 104–111. https://doi.org/10.1139/cjm-2014-0610.
1441 41. Xia, F.; Liu, Y.; Guo, L.-X.; Shen, G.-L.; Lin, J.; Zhou, X.-W. Pyrosequencing analysis revealed complex endogenetic
1442 microorganism community from natural DongChong XiaCao and its microhabitat. *BMC Microbiol.* **2016**, *16*, 196. DOI
1443 10.1186/s12866-016-0813-5
1444 42. Guo, M.-Y.; Liu, Y.; Gao, Y.-H.; Jin, T.; Zhang, H.-B.; Zhou, X.-W. Identification and bioactive potential of endogenetic
1445 fungi isolated from medicinal caterpillar fungus *Ophiocordyceps sinensis* from Tibetan Plateau. *Int. J. Agric. Biol.* **2017**, *19*,
1446 307–313. https://doi.org/10.17957/IJAB/15.0281.
1447 43. 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
1448 *Ophiocordyceps sinensis* transcriptome and differentially expressed genes in three different mycelia, sclerotium and fruiting
1449 body developmental stages. *Fungal Biol.* **2018**, *122*, 943–951. https://doi.org/10.1016/j.funbio.2018.05.011.
1450 44. Kang, Q.; Zhang, J.; Chen, F.; Dong, C.; Qin, Q.; Li, X.; Wang, H.; Zhang, H.; Meng, Q. Unveiling mycoviral diversity in
1451 *Ophiocordyceps sinensis* through transcriptome analyses. *Front. Microbiol.* **2024**, *15*, 1493365.
1452 https://doi.org/10.3389/fmicb.2024.1493365.
1453 45. Wei, X.-L.; Yin, X.-C.; Guo, Y.-L.; Shen, N.-Y.; Wei, J.-C. Analyses of molecular systematics on *Cordyceps sinensis* and its
1454 related taxa. *Mycosystema* **2006**, *25*, 192–202.
1455 46. Wei, J.-C.; Wei, X.-L.; Zheng, W.-F.; Guo, W.; Liu, R.-D. Species identification and component detection of *Ophiocordyceps*
1456 *sinensis* cultivated by modern industry. *Mycosystema* **2016**, *35*, 404–410.

1457 47. Lu, D. Western records and studies of the Chinese caterpillar fungus to the beginning of the 20th century. *J. Fungal. Res.*
1458 **2014**, *12*, 233–244. https://doi.org/10.13341/j.jfr.2014.0043.
1459 48. Sung, G.-H.; Hywel-Jones, N.L.; Sung, J.-M.; Luangsa-ard, J.J.; Shrestha, B.; Spatafora, J.W. Phylogenetic classification of
1460 *Cordyceps* and the clavicipitaceous fungi. *Stud. Mycol.* **2007**, *57*, 5–59. https://doi.org/10.3114/sim.2007.57.01.
1461 49. Zhang, Y.-J.; Li, E.-W.; Wang, C.-S.; Li, Y.-L.; Liu, X.-Z. *Ophiocordyceps sinensis*, the flagship fungus of China: Terminology,
1462 life strategy and ecology. *Mycology* **2012**, *3*, 2–10. https://doi.org/10.1080/21501203.2011.654354.
1463 50. Wang, Y.; Stata, M.; Wang, W.; Stajich, J.E.; White, M.M.; Moncalvo, J.M. Comparative genomics reveals the core gene
1464 toolbox for the fungus-insect symbiosis. *mBio* **2018**, *9*, e00636-e18. https://doi.org/10.1128/mBio.00636-18.
1465 51. Zhang, S.; Zhang, Y.-J.; Shrestha, B.; Xu, J.-P.; Wang, C.-S.; Liu, X.-Z. *Ophiocordyceps sinensis* and *Cordyceps militaris*:
1466 Research advances, issues and perspectives. *Mycosystema* **2013**, *32*, 577–597.
1467 52. Hawksworth, D.L.; Crous, P.W.; Redhead, S.A.; Reynolds, D.R.; Samson, R.A.; Seifert, K.A.; Taylor, J.W.; Wingfield, M.J.;
1468 Abaci, Ö.; Aime, C.; et al. The Amsterdam declaration on fungal nomenclature. *IMA Fungus* **2011**, *2*, 105–112.
1469 https://doi.org/10.5598/imafungus.2011.02.01.14.
1470 53. Taylor JW. One Fungus = One Name: DNA and fungal nomenclature twenty years after PCR. *IMA Fungus*. **2011**, *2(2),*
1471 113–120. doi:10.5598/imafungus.2011.02.02.01
1472 54. Yao, Y.-S.; Zhu, J.-S. Indiscriminate use of the Latin name for natural *Cordyceps sinensis* and *Ophiocordyceps sinensis* fungi.
1473 *Chin. J. Chin. Mater. Med*. **2016**, *41*(7), 1316–1366.
1474 55. Turgeon, B.G.; Yoder, O.C. Proposed nomenclature for mating type genes of filamentous ascomycetes. *Fungal Genet. Biol.*
1475 **2000**, *31*, 1–5. https://doi.org/10.1006/fgbi.2000.1227.
1476 56. Debuchy, R.; Turgeo, B.G. Mating-Type Structure, Evolution, and Function in Euascomycetes. In *Growth, Differentiation*
1477 *and Sexuality*; Kües, U., Fischer, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 293–323.
1478 57. Jones, S.K.; Bennett, R.J. Fungal mating pheromones: Choreographing the dating game. *Fungal Genet. Biol.* **2011**, *48*, 668–
1479 676. https://doi.org/10.1016/j.fgb.2011.04.001.
1480 58. Zheng, P.; Wang, C.-S. Sexuality Control and Sex Evolution in Fungi. *Sci. Sin. Vitae* **2013**, *43*, 1090–1097.
1481 59. Wilson, A.M.; Wilken, P.M.; van der Nest, M.A.; Steenkamp, E.T.; Wingfield, M.J.; Wingfield, B.D. Homothallism: An
1482 umbrella term for describing diverse sexual behaviours. *IMA Fungus* **2015**, *6*, 207–214.
1483 https://doi.org/10.5598/imafungus.2015.06.01.13.
1484 60. Sun, S.; Coelho, M.A.; David-Palma, M.; Priest, S.J.; Heitman, J. The evolution of sexual reproduction and the mating-type
1485 locus: Links to pathogenesis of *Cryptococcus* human pathogenic fungi. *Annu. Rev. Genet.* **2019**, *53*, 417–444.
1486 https://doi.org/10.1146/annurev-genet-120116-024755.
1487 61. Ramšak, B.; Kűck, U.; Hofmann, E. The mating type transcription factor MAT1-1-1 from the fungal human pathogen
1488 *Aspergillus fumigatus*: Synthesis, purification, and crystallization of the DNA binding domain. *bioRxiv* **2021**.
1489 https://doi.org/10.1101/2021.12.13.472399.
1490 62. Ramšak, B.; Kück, U. The *Penicillium chrysogenum* tom1 gene a major target of transcription factor MAT1-1-1 encodes a
1491 nuclear protein involved in sporulation. *Front. Fungal Bio.* **2022**, *3*, 937023. https://doi.org/10.3389/ffunb.2022.937023.
1492 63. Metin, B.; Findley, K.; Heitman, J. The mating type locus (MAT) and sexual reproduction of *Cryptococcus heveanensis*:
1493 Insights into the evolution of sex and sex-determining chromosomal regions in fungi. *PLoS Genet.* **2010**, *6*, e1000961.
1494 https://doi.org/10.1371/journal.pgen.1000961.
1495 64. Hu, X.; Zhang, Y.-J.; Xiao, G.-H.; Zheng, P.; Xia, Y.-L.; Zhang, X.-Y.; St Leger, R.J.; Liu, X.-Z.; Wang, C.-S. Genome survey
1496 uncovers the secrets of sex and lifestyle in caterpillar fungus. *Chin. Sci. Bull.* **2013**, *58*, 2846–2854.
1497 https://doi.org/10.1007/s11434-013-5929-5.
1498 65. Li, X.-Z.; Li, Y.-L.; Liu, W.; Zhu, J.-S. Altered stereostructures of the DNA-binding domains of variant mating proteins of
1499 *Ophiocordyceps sinensis* and the wild insect–fungal complex. *Biol. (Basel).* **2026**, *15(2)*, 186.
1500 https://doi.org/10.3390/biology15020186

1501 66. 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
1502 of Chinese cordyceps reveals gene regulatory network and expression profiles of sexual development-related genes. *BMC*
1503 *Genom.* **2019**, *20*, 337. https://doi.org/10.1186/s12864-019-5708-z.
1504 67. Li, X.-Z.; Li, Y.-L.; Zhu, J.-S. Differential transcription of mating-type genes during sexual reproduction of natural
1505 *Cordyceps sinensis*. *Chin. J. Chin. Mater. Medica* **2023c**, *48*, 2829–2840. https://doi.org/10.19540/j.cnki.cjcmm.20230213.102.
1506 68. Li, X.-Z.; Xiao, M.-J.; Li, Y.-L.; Gao, L.; Zhu, J.-S. Mutations and differential transcription of mating-type and pheromone
1507 receptor genes in *Hirsutella sinensis* and the natural *Cordyceps sinensis* insect–fungi complex. *Biology* **2024b**, *13*, 632.
1508 https://doi.org/10.3390/biology13080632.
1509 69. Li, X.-Z.; Li, Y.-L.; Zhu, J.-S. Three-dimensional structural heteromorphs of mating-type proteins in *Hirsutella sinensis* and
1510 the natural *Cordyceps sinensis* insect–fungal complex. *J. Fungi.* **2025**, *11*, 244. https://doi.org/10.3390/jof11040244.
1511 70. Zhang, Y.-J.; Xu, L.-L.; Zhang, S.; Liu, X.-Z.; An, Z.-Q.; Wàngmǔ Guo, Y.-L. Genetic diversity of *Ophiocordyceps sinensis*, a
1512 medicinal fungus endemic to the Tibetan Plateau: Implications for its evolution and conservation. *BMC Evol. Biol.* **2009**, *9*,
1513 290. doi:10.1186/1471-2148-9-290
1514 71. 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
1515 traditional Chinese medicinal fungus *Ophiocordyceps sinensis*. *Fungal Biol.* **2011**, *115*, 708–714.
1516 72. Zhang, S.; Zhang, Y.-J. Molecular evolution of three protein-coding genes in the Chinese caterpillar fungus *Ophiocordyceps*
1517 *sinensis*. *Microbiol. China.* **2015**, *42*, 1549–1560.
1518 73. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–
1519 132. https://doi.org/10.1016/0022-2836(82)90515-0.
1520 74. Deleage, G.; Roux, B. An algorithm for protein secondary structure prediction based on class prediction. *Protein Eng. Des.*
1521 *Sel.* **1987**, *1*, 289–294. https://doi.org/10.1093/protein/1.4.289.
1522 75. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein Identification and
1523 Analysis Tools on the ExPASy Server, Chapter 52. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Humana Press:
1524 Totowa, NJ, USA, **2005**; pp. 571–607.
1525 76. Peters, C.; Elofsson, A. Why is the biological hydrophobicity scale more accurate than earlier experimental hydrophobicity
1526 scales? *Proteins* **2014**, *82*, 2190–2198. https://doi.org/10.1002/prot.24582.
1527 77. Simm, S.; Einloft, J.; Mirus, O.; Schleiff, E. 50 years of amino acid hydrophobicity scales, revisiting the capacity for peptide
1528 classification. *Biol. Res.* **2016**, *49*, 31. https://doi.org/10.1186/s40659-016-0092-5.
1529 78. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.;
1530 Potapenko, A.; Bridgland, A.; Meyer, C.: Kohl, S,A,A,; Ballard, A.J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.;
1531 Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.;
1532 Bodenstein; S.; Silver, D.; Vinyals, O.; Senior, A.W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein
1533 structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
1534 79. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Žídek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon,
1535 A.; Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA, Potapenko
1536 A, Ballard AJ, Romera-Paredes B, Nikolov S, Jain R, Clancy E, Reiman D, Petersen S, Senior AW, Kavukcuoglu K, Birney
1537 E, Kohli P, Jumper J, Hassabis D. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*,
1538 590–596. https://doi.org/10.1038/s41586-021-03828-1.
1539 80. David, A.; Islam, S.; Tankhilevich, E.; Sternberg, M.J.E. The AlphaFold Database of Protein Structures, A Biologist's Guide.
1540 *J. Mol. Biol.* **2022**, *434*, 167336. https://doi.org/10.1016/j.jmb.2021.167336.
1541 81. Monzon, V.; Haft, D.H.; Bateman, A. Folding the unfoldable, using AlphaFold to explore spurious proteins. *Bioinform.*
1542 *Adv.* **2022**, *1*, vbab043. https://doi.org/10.1093/bioadv/vbab043.
1543 82. Rettie, S.A.; Campbell, K.V.; Bera, A.K.; Kang, A.; Kozlov, S.; De La Cruz, J.; Adebomi, V.; Zhou, G.; DiMaio, F.;
1544 Ovchinnikov, S.; et al. Cyclic peptide structure prediction and design using AlphaFold. *bioRxiv* **2023**, 26:2023.02.25.529956.
1545 https://doi.org/10.1101/2023.02.25.529956.

1546 83. Xu, T.; Xu, Q.; Li, J.-Y. Toward the appropriate interpretation of Alphafold2. *Front. Artif. Intell.* **2023b**, *6*, 1149748.
1547 https://doi.org/10.3389/frai.2023.1149748.
1548 84. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick,
1549 J.; Bodenstein SW, Evans DA, Hung CC, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E,
1550 Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB,
1551 Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C,
1552 Yakneen S, Zhong ED, Zielinski M, Žídek A, Bapst V, Kohli P, Jaderberg M, Hassabis D, Jumper JM. Accurate structure
1553 prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500. https://doi.org/10.1038/s41586-024-
1554 07487-w.
1555 85. Varadi, M.; Bertoni, D.; Magana, P.; Paramval, U.; Pidruchna, I.; Radhakrishnan, M.; Tsenkov, M.; Nair, S.; Mirdita, M.;
1556 Yeo, J.; et al. AlphaFold Protein Structure Database in 2024, providing structure coverage for over 214 million protein
1557 sequences. *Nucleic Acids Res.* **2024**, *52*, D368–D375. https://doi.org/10.1093/nar/gkad1011.
1558 86. Wroblewski, K.; Kmiecik, S. Integrating AlphaFold pLDDT Scores into CABS-flex for enhanced protein flexibility
1559 simulations. *Comput. Struct. Biotechnol. J.* **2024**, *30*, 4350–4356. https://doi.org/10.1016/j.csbj.2024.11.047.
1560 87. Li, X.-Z.; Li, Y.-L.; Zhu, J.-S. Three-dimensional structural heteromorphs of mating-type proteins in *Hirsutella sinensis* and
1561 the natural *Cordyceps sinensis* insect–fungal complex. *J. Fungi.* **2025**, *11*, 244. https://doi.org/10.3390/jof11040244.
1562 88. Yang, X.; Zhu, H.-Q.; Shi, L.-X.; Song, T.-G.; Gong, W.-B.; He, S.-M.; Shan, S.; Xu, C.-F.; Zhou, Z. AlphaFold-guided
1563 structural analyses of nucleosome binding proteins. *Nucleic Acids Res.* **2025**, *53*, gkaf735.
1564 https://doi.org/10.1093/nar/gkaf735.
1565 89. Mariani, V.; Biasini, M.; Barbato, A.; Schwede, T. lDDT, a local superposition-free score for comparing protein structures
1566 and models using distance difference tests. *Bioinformatics* **2013**, *29*, 2722–2728.
1567 https://doi.org/10.1093/bioinformatics/btt473.
1568 90. Baxevanis, A.D.; Bryant, S.H.; Landsman, D. Homology model building of the HMG-1 box structural domain. *Nucleic*
1569 *Acids Res.* **1995**, *23*, 1019–1029. https://doi.org/10.1093/nar/23.6.1019.
1570 91. Thapar, R. Structure-specific nucleic acid recognition by L-motifs and their diverse roles in expression and regulation of
1571 the genome. *Biochim. Biophys. Acta* **2015**, *1849*, 677–687. https://doi.org/10.1016/j.bbagrm.2015.02.006.
1572 92. White, T.J., Bruns, T., Lee, S., Taylor, J.W. Amplification and direct sequencing of fungal ribosomal RNA genes for
1573 phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.), PCR Protocols, A Guide to Methods and
1574 Applications. Academic Press, London, 1990; pp. 315–322.
1575 93. 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
1576 sequencing and assembly strategies for a repeat-rich fungal genome, *Ophiocordyceps sinensis*. *J. Microbiol. Methods* **2016a**,
1577 *128*, 1–6. https://doi.org/10.1016/j.mimet.2016.06.025.
1578 94. 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.;
1579 Zheng, Y.-G. Genome sequencing and analysis of fungus *Hirsutella sinensis* isolated from *Ophiocordyceps sinensis*. *AMB*
1580 *Expr.* **2020**, *10*, 105. https://doi.org/10.1186/s13568-020-01039-x.
1581 95. 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
1582 analyses reveal evolutionary and geologic context for the plateau fungus *Ophiocordyceps sinensis*. *Clin. Med.* **2020**, *15*, 107–
1583 119. https://doi.org/10.1186/s13020-020-00365-3.
1584 96. 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
1585 high-quality draft genome assembly of the Chinese cordyceps *Ophiocordyceps sinensis*. *Genome Biol. Evol.* **2020**, *12*, 1074–
1586 1079. https://doi.org/10.1093/gbe/evaa112.
1587 97. Martin, T.; Lu, S.-W.; van Tilbeurgh, H.; Ripoll, D.R.; Dixelius, C.; Dixelius, C.; Turgeon, B.G.; Debuchy, R. Tracing the
1588 Origin of the Fungal a1 Domain Places Its Ancestor in the HMG-Box Superfamily, Implication for Fungal Mating-Type
1589 Evolution. *PLoS ONE* **2010**, *5*, e15199. https://doi.org/10.1371/journal.pone.0015199.
1590 98. Czaja, W.; Miller, K.Y.; Miller, B.L. Complex mechanisms regulate developmental expression of the matA (HMG) mating
1591 type gene in homothallic *Aspergillus* nidulans. *Genetics*. **2011**, *189*, 795-808. doi: 10.1534/genetics.111.131458

1592 99. Angelini, R.M.D.M., Rotolo, C., Pollastro, S., Faretra, F. Molecular analysis of the mating type (*MAT1*) locus in strains of
1593 the heterothallic ascomycete *Botrytis cinerea*. *Plant Pathol*. **2016**, *65*, 1221-1400. https://doi.org/10.1111/ppa.12509.
1594 100. Poveda-Huertes, D., Matic, S., Marada, A., Habernig, L., Licheva, M., Myketin, L., Gilsbach, R., Tosal-Castano S, Papinski
1595 D, Mulica P, Kretz O, Kücükköse C, Taskin AA, Hein L, Kraft C, Büttner S, Meisinger C, Vögtle FN. An Early mtUPR:
1596 Redistribution of the Nuclear Transcription Factor Rox1 to Mitochondria Protects against Intramitochondrial Proteotoxic
1597 Aggregates. *Mol Cell*. **2020**, *77(1),* 180-188. doi: 10.1016/j.molcel.2019.09.026
1598 101. Hansen, F.T., Madsen, C.K., Nordland, A.M., Grasser, M., Merkle, T., Grasser, K.D. A novel family of plant DNA-binding
1599 proteins containing both HMG-box and AT-rich interaction domains. *Biochemistry*. **2008**; *47*, 13207–13214. doi:
1600 10.1021/bi801772k
1601 102. Rajewska, M., Wegrzyn, K., Konieczny, I. AT-rich region and repeated sequences - the essential elements of replication
1602 origins of bacterial replicons. *FEMS Microbiol Rev*. **2012**, *36(2),* 408–434. doi: 10.1111/j.1574-6976.2011.00300.x
1603 103. Jackson, D.; Lawson, T.; Villafane, R.; Gary, L. Modeling the structure of yeast MATα1, An HMG-Box motif with a C-
1604 terminal helical extension. *Open J. Biophys.* **2013**, *3*, 1–12. https://doi.org/10.4236/ojbiphy.2013.31001.
1605 104. Zheng, P.; Xia, Y.-L.; Zhang, S.-W.; Wang, C.-S. Genetics of Cordyceps and related fungi. *Appl Microbiol Biotechnol*, **2013a**,
1606 *97*, 2797–2804. doi: 10.1007/s00253-013-4771-7
1607 105. 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
1608 genes of mating-type loci provides insight into the regulation of sexual development of the cereal pathogen *Fusarium*
1609 *graminearum*. *PLoS Genet.* **2015**, *11*, e1005486. https://doi.org/10.1371/journal.pgen.1005486.
1610 106. Ramšak, B.; Markau, J.; Pazen, T.; Dahlmann, T.A.; Krappmann, S.; Kűck, U. The master regulator MAT1-1-1 of fungal
1611 mating binds to its targets via a conserved motif in the human pathogen *Aspergillus fumigatus*. *G3 Genes Genom. Genet.*
1612 **2020**, *11*, jkaa012. https://doi.org/10.1093/g3journal/jkaa012.
1613 107. Yamamoto, A.; Ando, Y.; Yoshioka, K.; Saito, K.; Tanabe, T.; Shirakawa, H.; Yoshida, M. Difference in affinity for DNA
1614 between HMG proteins 1 and 2 determined by surface plasmon resonance measurements. *J. Biochem.* **1997**, *122*, 586–594.
1615 https://doi.org/10.1093/oxfordjournals.jbchem.a021793. PMID 9348088.
1616 108. Ait Benkhali, J.; Coppin, E.; Brun, S.; Peraza-Reyes, L.; Martin, T.; Dixelius, C.; Lazar, N.; van Tilbeurgh, H.; Debuchy, R.
1617 A Network of HMG-box Transcription Factors Regulates Sexual Cycle in the Fungus Podospora anserina. *PLoS Genet.*
1618 **2013**, *9*, e1003642. https://doi.org/10.1371/journal.pgen.1003642.
1619 109. Balasubramanian, B.; Lowry, C.V.; Zitomer, R.S. The Rox1 repressor of the *Saccharomyces cerevisiae* hypoxic genes is a
1620 specific DNA-binding protein with a high-mobility-group motif. *Mol. Cell Biol.* **1993**, *13*, 6071–6078.
1621 https://doi.org/10.1128/mcb.13.10.6071-6078.1993.
1622 110. Zitomer, R.S.; Limbach, M.P.; Rodriguez-Torres, A.M.; Balasubramanian, B.; Deckert, J.; Snow, P.M. Approaches to the
1623 study of Rox1 repression of the hypoxic genes in the yeast Saccharomyces cerevisiae. *Methods* **1997**, *11*, 279–288.
1624 https://doi.org/10.1006/meth.1996.0422.
1625 111. 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
1626 play different roles in sexual reproduction and pathogenesis in *Fusarium graminearum*. *PLoS ONE.* **2013b**, *8*, e66980.
1627 https://doi.org/10.1371/journal.pone.0066980.
1628 112. Kastaniotis, A.J.; Zitomer, R.S. Oxygen Dependent Repression in Yeast. In *Rox1 Mediated Repression*; Advances in
1629 Experimental Medicine and Biology; Springer Nature: Cham, Switzerland, 2000; Volume 475, pp. 185–195.
1630 https://doi.org/10.1007/0-306-46825-5_18.
1631 113. Kües, U.; Casselton, L.A. The origin of multiple mating types in mushrooms. *J. Cell Sci.* **1993**, *104*, 227–230.
1632 https://doi.org/10.1242/jcs.104.2.227.
1633 114. Asante-Owusu, R.N.; Banham, A.H.; Böhnert, H.U.; Mellor, E.J.C.; Casselton, L.A. Heterodimerization between two
1634 classes of homeodomain proteins in the mushroom *Coprinus cinereus* brings together potential DNA-binding and
1635 activation domains. *Gene* **1996**, *172*, 25–31. https://doi.org/10.1016/0378-1119(96)00177-1.
1636 115. Jacobsen, S.; Wittig, M.; Pöggeler, S. Interaction Between Mating-Type Proteins from the Homothallic Fungus Sordaria
1637 macrospora. *Curr. Genet.* **2002**, *41*, 150–158. https://doi.org/10.1007/s00294-002-0276-0.

116. Hancock, S.P.; Cascio, D.; Johnson, R.C. Cooperative DNA binding by proteins through DNA shape complementarity. *Nucleic Acids Res.* **2019**, *47*, 8874–8887. https://doi.org/10.1093/nar/gkz642.
117. Mao, X.-M.; Zhao, S.-M.; Cao, L.; Yan, X.; Han, R.-C. The morphology observation of *Ophiocordyceps sinensis* from different origins. *J. Environ. Entomol.* **2013**, *35*, 343–353.
118. Li, Y.; Yang, R.-H.; Jiang, L.; Hu, X.-D.; Wu, Z.-J.; Yao, Y.-J. rRNA Pseudogenes in Filamentous Ascomycetes as Revealed by Genome Data. *G3-Genes Genom. Genet.* **2017**, *7*, 2695–2703. https://doi.org/10.1534/g3.117.044016.
119. Li, Y.; Jiang, L.; Wang, K.; Wu, H.-J.; Yang, R.-H.; Yan, Y.-J.; Bushley, K.E.; Hawksworth, D.L.; Wu, Z.-J.; Yao, Y.-J. RIP mutated ITS genes in populations of *Ophiocordyceps sinensis* and their implications for molecular systematics. *IMA Fungus* **2020c**, *11*, 18.
120. Li, X.-Z.; Li, Y.-L.; Wang, Y.-N.; Zhu, J.-S. Translations of mutant repetitive genomic sequences in *Hirsutella sinensis* and changes in secondary structures and functional specifications of the encoded proteins. *Int. J. Mol. Sci.* **2024a**, *25*, 11178. https://doi.org/10.3390/ijms252011178.
121. Wang, Y.-B., Wang, Y., Fan, Q., Duan, D-E., Zhang, G.-D., Dai6, R.-Q., Dai, Y.-D., Zeng, W.-B. Chen, Z.-H., Li, D.-D., Tang, D.-X., Xu, Z.-H., Sun, T., Nguyen, T.-T., Tran, N.-L., Dao, V.-M., Zhang, C.-M., Huang, L.-D., Liu, Y.-J., Zhang, X.-M., Yang, D.-R., Sanjuan, T., Liu, X.-Z., Yang, Z.-L., Yu, H. Multigene phylogeny of the family Cordycipitaceae (Hypocreales): new taxa and the new systematic position of the Chinese cordycipitoid fungus *Paecilomyces hepiali*. *Fungal Diversity* **2020**, *103*, 1–46. https://doi.org/10.1007/s13225-020-00457-3.
122. Holliday, J.; Cleaver, M. Medicinal value of the caterpillar fungi species of the genus *Cordyceps* (Fr.) Link (Ascomycetes). A review. *Int. J. Med. Mushrooms* **2008**, *10*, 219–234. https://doi.org/10.1615/IntJMedMushr.v10.i3.30.
123. Stone, R. Improbable partners aim to bring biotechnology to a Himalayan kingdom. *Science* **2010**, *327*, 940–941. https://doi.org/10.1126/science.327.5968.940.
124. Qin, Q.-L.; Zhou, G.-L.; Zhang, H.; Meng, Q.; Zhang, J.-H.; Wang, H.-T.; Miao, L.; Li, X. Obstacles and approaches in artificial cultivation of Chinese cordyceps. *Mycology* **2018**, *9*, 7–9. https://doi.org/10.1080/21501203.2018.1442132.
125. 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. Erratum in Sci. China Life Sci. **2023b**, 66, 892. https://doi.org/10.1007/s11427-022-2232-5.
126. Quandt, C.A., Kepler, R.M., Gams, W., Araújo, J.P.M., Ban, S., Evans, H.C., Hughes, D., Humber, R., Hywel-Jones, N., Li, Z.-Z., Luangsa-ard, J.J., Rehner, S.A., Sanjuan, T., Sato, H., Shrestha, B., Sung, G.-H., Yao, Y.-J., Zare, R., Spatafora, J.W. Phylogenetic-based nomenclatural proposals for Ophiocordycipitaceae (Hypocreales) with new combinations in Tolypocladium. *IMA Fungus.* 2014, **5(***1*), 121–134. doi:10.5598/imafungus.2014.05.01.12.
127. Bennett, R.J.; Johnson, A.D. Completion of a parasexual cycle in *Candida albicans* by induced chromosome loss in tetraploid strains. *EMBO J.* **2003**, *22*, 2505–2515. https://doi.org/10.1093/emboj/cdg235.
128. Sherwood, R.K.; Bennett, R.J. Fungal meiosis and parasexual reproduction--lessons from pathogenic yeast. *Curr. Opin. Microbiol.* **2009**, *12*, 599–607. https://doi.org/10.1016/j.mib.2009.09.005.
129. Seervai, R.N.H.; Jones, S.K.; Hirakawa, M.P.; Porman, A.M.; Bennett, R.J. Parasexuality and ploidy change in *Candida tropicalis*. *Eukaryot. Cell.* **2013**, *12*, 1629–1640. https://doi.org/10.1128/EC.00128-13.
130. 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.
131. 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.
132. Hėnault, M.; Marsit, S.; Charron, G.; Landry, C.R. The effect of hybridization on transposable element accumulation in an undomesticated fungal species. *eLife* **2020**, *9*, e60474. https://doi.org/10.7554/eLife.60474.
133. Samarasinghe, H.; You, M.; Jenkinson, T.S.; Xu, J.-P.; James, T.Y. Hybridization Facilitates Adaptive Evolution in Two Major Fungal Pathogens. *Genes* **2020**, *11*, 101. https://doi.org/10.3390/genes11010101.

134. Mishra, A.; Forche, A.; Anderson, M.Z. Parasexuality of Candida Species. *Front. Cell. Infect. Microbiol.* **2021**, *11*, 796929. https://doi.org/10.3389/fcimb.2021.796929.
135. Steensels, J.; Gallone, B.; Verstrepen, K.J. Interspecific hybridization as a driver of fungal evolution and Adaptation. *Nat. Rev. Microbiol.* **2021**, *19*, 485–500.
136. Kück, U.; Bennett, R.J.; Wang, L.; Dyer, P.S. Editorial, Sexual and Parasexual Reproduction of Human Fungal Pathogens. *Front. Cell Infect. Microbiol.* **2022**, *12*, 934267. https://doi.org/10.3389/fcimb.2022.934267.
137. Yokoyama, E., Yamagishi, K., Hara, A, Heterothallism in *Cordyceps takaomontana*. *FEMS Microbiol Lett*. **2005**, *250(1)*, 145–150. https://doi.org/10.1016/j.femsle.2005.07.004
138. Lee SC, Ni M, Li W, Shertz C, Heitman J. The Evolution of Sex: a Perspective from the Fungal Kingdom. Microbiol *Mol Biol Rev*. **2010**, *74(2)*, 298−340. doi: 10.1128/MMBR.00005-10
139. Araújo, J.P.M., Evans, H.C., Kepler, R., Hughes, D.P. Zombie-ant fungi across continents: 15 new species and new combinations within *Ophiocordyceps*. I. Myrmecophilous hirsutelloid species. *Stud Mycol*. **2018**, *90*, 119−160. doi: 10.1016/j.simyco.2017.12.002.
140. Xu M, Ashley NA, Vaghefi N, Wilkinson I, Idnurm A. Isolation of strains and their genome sequencing to analyze the mating system of *Ophiocordyceps robertsii*. *PLoS One*. **2023a**, *18(5)*, e0284978. doi: 10.1371/journal.pone.0284978
141. Zou, J.; Zhang, Y.-T.; Luo, Y.; Fu, M.-H.; Sun, B.-L.; Liu, S.-G. MAT1-1 and MAT1-2 *Ophiocordyceps xuefengensis* and comparison of their chemical composition. methods. *Biol. (Basel)* **2024**, *13(9)*, 686; https://doi.org/10.3390/biology13090686.
142. Zhou, X.W.; Li, L.J.; Tian, E.W. Advances in research of the artificial cultivation of *Ophiocordyceps sinensis* in China. *Crit. Rev. Biotechnol.* **2013**, *34*, 233–243. https://doi.org/10.3109/07388551.2013.791245.