Differential pairings of MAT1-1-1 and MAT1-2-1 protein variants with heteromorphic tertiary structures produced by *Ophiocordyceps sinensis* strains

Xiu-Zhang Li [1], Yu-Ling Li [1], Wei Liu [2], 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

[2] Institute of Immunology, Army Medical University, Chongqing, 400038, China

* Corresponding author

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

Simple Summary

Cordyceps sinensis insect–fungal complex is a combination of the remains of a Hepialidae moth larva and the *Ophiocordyceps sinensis* fruiting body. For centuries, it has been used in traditional Chinese medicine for health maintenance, disease amelioration, help recover after being sick or having surgery, and even for slowing down the aging process. The sexual reproduction of *O. sinensis* is controlled by the 3D structures of specific functional parts of the paired MAT1-1-1 and MAT1-2-1 proteins. This study examined the structures of paired MAT1-1-1 and MAT1-2-1 protein variants from 20 *O. sinensis* strains. Differentially altered functional parts of some of the mating proteins are shorter than others with divergent replacement of amino acids, resulting in altered 3D structures of the proteins. Together with findings from previous and present studies, the paired mating proteins with divergent alterations of 3D structures prove that *O. sinensis* cannot reproduce on its own (self-fertilize). Instead, *O. sinensis* is "self-sterile" and its sexual reproduction requires mating partner(s). This new understanding of how *O. sinensis* reproduces sexually provides a new perspective to learn about *O. sinensis* and the *C. sinensis* insect–fungal complex and will guide future academic and applied studies.

Abstract

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

Keywords: MATα_HMGbox domain of MAT1-1-1 protein; HMG-box_ROX1-like domain of MAT1-

2-1 protein; heteromorphic primary, secondary, and tertiary structures of proteins; *Ophiocordyceps sinensis* strain; sexual reproduction; the *Cordyceps sinensis* insect−fungal complex

1. Introduction

Natural *Cordyceps sinensis*, also known as Chinese cordyceps (a non-Latin term) or DōngChóng XiàCǎo in Chinese, is endemic to high-altitude alpine meadows (3,000−5,000 m), 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*" and "*YANG*", as well as in disease amelioration, post-illness and postoperative recovery, and antiaging therapy [Zhu *et al*. 1998a, 1998b, 2011; Song *et al*. 2001]. Modern science is increasingly elucidating the complex biochemical mechanisms (*e.g*., immunomodulation, anti-inflammation, antioxidant activity, organ protection, enhancement of energy-vitality-stamina, etc.) that underpin these traditional applications, attesting to its status as a precious therapeutic resource, albeit the complexity and rarity of this medicinal resource. This natural substance with its unique therapeutic significance embodies a remarkable fusion of nature's ingenuity and traditional medicinal wisdom; however, its high value has led to overharvesting, rendering it a Level-II endangered natural substance [China Ministry of Agriculture and Rural Affairs 2021].

Natural *C. sinensis* comprises the fruiting body of *Ophiocordyceps sinensis* and the remains of a mummified *Hepialidae* moth larva, which retains a thick body wall with numerous bristles, intact larval intestine, head tissues, and fragments of other larval tissues [Ren *et al*. 2013; Zhang *et al*. 2014; Lu & St Leger 2016; Zhu & Li 2017; Li *et al*. 2022, 2023a, 2023d, 2023e]. Dual-fluorescence microscopy examinations have revealed multicellular heterokaryotic microstructures, containing mono-, bi-, tri-, and tetra-nucleate cells [Bushley *et al*. 2013]. Natural *C. sinensis* exhibits apparent genetic heterogeneity, harboring >90 fungal species from at least 37 fungal genera, as well as 17 genome−independent genotypes of *O. sinensis* that differentially co-occur naturally, the abundance of which undergoes dynamic alterations in different compartments during the maturation of natural *C. sinensis* [Li 1988; Dai *et al*. 1989, 2008; Liu *et al*. 1989; Kinjo & Zang 2001; Jiang & Yao 2003; Chen *et al*. 2004, 2011; Stensrud *et al*. 2005, 2007; Leung *et al*. 2006; Zhu *et al*. 2007, 2010; Yang *et al*. 2008, 2021; Xiao *et al*. 2009; Zhang *et al*. 2010a, 2010b, 2018; Li *et al*. 2013, 2016b, 2020b, 2022, 2023d, 2023e; Barseghyan *et al*. 2011; Gao *et al*. 2012; Ren *et al*. 2013; Meng *et al*. 2015; Xia *et al*. 2015, 2016; Guo *et al*. 2017; Zhu & Li 2017; Zhong *et al*. 2018; Kang *et al*. 2024]. Thus, the Chinese Pharmacopoeia defines this medicinal microecosystem as an insect–fungal complex.

Among the numerous intrinsic fungal species [Jiang & Yao 2003; Zhang *et al*. 2010a, 2010b; Xia *et al*. 2015, 2016; Zhang *et al*. 2018], *Hirsutella sinensis* has been postulated to be the sole anamorph of *O. sinensis* [Wei *et al*. 2006], although this hypothetical claim has not yet met all 4 criteria of Koch's postulates. Ten years later, however, Wei *et al*. [2016] reported a successful artificial cultivation project conducted in an industrial product-oriented setting and identified 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 as belonging to *Agrotis* [Lu 2014]. This imprecise usage of the Latin-name has blurred the distinction between the fungus, its host, and the insect−fungal complex, resulting in ambiguity in scientific literature, commercial trade, and governmental regulations worldwide. The fungus was renamed *O. sinensis* in 2007, using *H. sinensis* strain EFCC7287 (GC-biased Genotype #1 of *O. sinensis*) as the nomenclatural reference [Sung *et al*. 2007; Zhang *et al*. 2012; Ren *et al*. 2013; Zhu & Li 2017; Wang *et al*. 2018; Li *et al*. 2022, 2026]. Zhang *et al*. [2013] have improperly implemented the "One Fungus=One Name" nomenclature rule set by the International Mycological Association (IMA) and proposed replacing the anamorphic name *H. sinensis* with the teleomorphic name *O. sinensis* according to the international rule [Hawksworth *et al*. 2011; Taylor 2011]. They interpreted *O. sinensis* as a single fungus to fulfill the prerequisite of "One Fungus" of the international rule. However, their interpretation and implementation of the international rule did not fully consider the 17 genotypic *O. sinensis* fungi, which are genome-independent belonging to distinct fungi and differentially co-exist in the compartments of *C. sinensis* insect-fungal complex. In the present paper, we retain the anamorphic name *H. sinensis* for GC-biased Genotype #1 among the 17 genotypes of *O. sinensis* and refer to the genomically independent Genotypes #2–17 fungi as *O. sinensis* following 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 complex, because the 2007 renaming of *C. sinensis* to *O. sinensis*, based on *H. sinensis* strain EFCC7287 as the nomenclatural reference, did not address the long-standing indiscriminate application of the Latin name

108 to the insect–fungal complex [Sung *et al*. 2007; Yao & Zhu 2016; Zhu & Li 2017; Li *et al*. 2022, 2026].
109 Nevertheless, this imperfect and temporarily necessary practice will likely be revised taxonomically in
110 the future, through the differential use of proprietary and exclusive Latin names for the multiple genome–
111 independent *O. sinensis* genotypic fungi and the insect–fungal complex.

112 The sexual reproductive behavior of ascomycetes is controlled by transcription factors encoded at
113 the mating-type (*MAT*) locus, which constitute the core regulatory mechanism governing mating type
114 recognition and compatibility, and the development of fruiting bodies and sexual reproductive structures
115 such as ascocarps and ascospores [Turgeon & Yoder 2000; Debuchy & Turgeon 2006; Jones & Bennett
116 2011; Zheng & Wang 2013; Wilson *et al*. 2015; Sun *et al*. 2019; Ramšak *et al*. 2021; Ramšak & Kück
117 2022]. Similar to other ascomycetous fungi, the reproductive behavior of *O. sinensis* depends on the
118 synergistic interplay of 2 core mating proteins, the *MAT1-1* and *MAT1-2* idiomorphs. The MAT1-1-1
119 protein harbors a mating-type alpha high mobility group box (MATα_HMGbox) domain, whereas the
120 MAT1-2-1 protein contains the high mobility group box ROX1-like (HMG-box_ROX1-like) domain
121 [Metin *et al*. 2010; Bushley *et al*. 2013; Hu *et al*. 2013; Zheng & Wang 2013; Li *et al*. 2026]. These 2
122 conserved domains of mating proteins play essential roles in regulating the expression of genes
123 associated with sexual reproduction. Li *et al*. [2019, 2023c, 2024b] documented differential occurrence,
124 differential transcription, and alternative splicing of mating-type and pheromone receptor genes in the
125 genomes of *H. sinensis* strains, challenging the prior hypothesis that *H. sinensis* undergoes self-
126 fertilization *via* homothallism or pseudohomothallism. Instead, these discoveries support self-sterility
127 consistent with a heterothallic or hybrid lifestyle, indicating necessity of compatible sexual partners to
128 complete the sexual cycle within the *Cordyceps sinensis* insect-fungus complex. Moreover, Li *et al*.
129 [2025, 2026] demonstrated heteromorphic tertiary structures of the DNA-binding domains of full-length
130 MAT1-1-1 and MAT1-2-1 proteins in impure wild-type *C. sinensis* isolates, which might be generated
131 by either GC-biased Genotype #1 (*H. sinensis*) or Genotype #3 of *O. sinensis*. These observations further
132 underscore the complexity underlying heterothallic or hybrid sexual reproduction of *O. sinensis*.

133 Variations in the DNA-binding domains of the 2 mating proteins, arising from diverse amino acid
134 substitutions, induce conformational changes in their tertiary structures as well as synergistic effects of
135 the 2 idiomorphic mating proteins. These proteins are derived from numerous mycologically impure,
136 wild-type *C. sinensis* isolates [Li *et al*. 2026], which were collected and reported by Prof. Zhang Y-J's
137 group [Zhang *et al*. 2009, 2011, 2013, 2014; Zhang & Zhang 2015]. To clarify the protein structural basis
138 of *O. sinensis* sexual reproduction, the present study further analyzes the N- and/or C-terminal

truncations and amino acid substitutions of the reported by Prof. Yao Y-J's group and their collaborators [Bushley *et al*. 2013; Li *et al*. 2013] and correlates the changes in the primary, secondary, and tertiary structures and hydrophobic properties of the MATα_HMGbox and the HMG-box_ROX1-like domains of the MAT1-1-1 and MAT1-2-1 proteins, respectively. Critically, the present study investigates the differential pairings of MAT1-1-1 and MAT1-2-1 protein variants generated by the *O. sinensis* strains and the possible heterospecific fungal origins within these heterogeneous *O. sinensis* strains.

2. Materials and Methods

2.1 The MAT1-1-1 and MAT1-2-1 protein sequences of the *O. sinensis* strains

The AlphaFold database provides the AI-predicted 3D structural morphs for 20 truncated MAT1-1-1 proteins (AGW27517–AGW27536) and 5 MAT1-2-1 proteins (AGW27543, AGW27548, AGW27552, AGW27554, and AGW27555) [Bushley *et al*. 2013]. As reported by Li *et al*. [2013], 11 of the 20 *O. sinensis* strains were obtained from fresh caterpillar body specimens collected from various *C. sinensis* production regions on the Qinghai-Tibet Plateau, marked as "TS" in Table 1. The remaining 9 *O. sinensis* strains were from the cultures of monoascospores of *C. sinensis* insect–fungal complexes that were collected from a single production region (Maqên, Guoluo, Qinghai Provence of China), marked as "SS" in Table 1. The MAT1-1-1 protein sequences of different lengths were encoded by the *MAT1-1-1* cDNAs that were amplified from total RNA extracted from the *O. sinensis* strains using the primer pair of *m1cDNAF3/m1cDNAR5*, while the *MAT1-2-1* cDNAs were amplified using the primer pair of *Mat1-2F/Mat1-2R* (Figures S1–S2). Li *et al*. [2013] deposited the internal transcribed spacer (ITS) sequences to GenBank for 15 of the 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 [Li et al. 2013], 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 in brown have ITS sequences of Group A; those in red have ITS sequences for both Groups A and B; and those in black do not have ITS sequences in GenBank [Bushley *et al*. 2013; Li *et al*. 2013]. "TS" indicates that the *O. sinensis* strains were isolated from the caterpillar body specimens. "SS" indicates that the *O. sinensis* strains were 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) [Li *et al*. 2013, 2022]. 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. The AlphaFold codes with the same suffix symbol (*, ⁋, †, ‡, ⁑, or ♦) belong to the same morphs of AlphaFold-predicted 3D structures. Those without a suffix symbol indicate the unique 3D structural morphs.

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 of 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 on the basis of the general properties of their side chains (https://web.expasy.org/protscale/; Basel, Switzerland) (Table S3), accessed from 18 May 2025 to 20 October 2025 [Kyte & Doolittle 1982; Deleage & Roux 1987; Gasteiger *et al*. 2005; Peters & Elofsson 2014; Simm *et al*. 2016; Li *et al*. 2024b, 2025, 2026]. 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 [Bushley *et al*. 2013]) and the HMG-box_ROX1-like domain of MAT1-2-1 proteins (amino acids 127→197 of the reference sequence AEH27625 [Zhang *et al*. 2009]), 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 [Kyte & Doolittle 1982; Deleage & Roux 1987; Gasteiger *et al*. 2005; Li *et al*. 2024b, 2025, 2026] to generate ExPASy ProtScale plots to predict the potential changes in hydrophobicity property 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

For heteromorphic stereostructure analysis 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 the artificial intelligence (AI)-driven machine learning technology AlphaFold (https://alphafold.com/, Cambridgeshire, UK; accessed from 18 June 2025 to 20 October 2025) [Jumper *et al*. 2021; Tunyasuvunakool *et al*. 2021; David *et al*. 2022; Monzon *et al*. 2022; Rettie *et al*. 2023; Xu *et al*. 2023b; Abramson *et al*. 2024; Varadi *et al*. 2024; Wroblewski & Kmiecik 2024; Li *et al*. 2025; Yang *et al*. 2025].

The AlphaFold database provides per-residue model confidence, predicts scores between 0 and 100 in the partial local distance difference test (pLDDT), and assigns a per-residue score to each individual residue [Jumper *et al*. 2021; Tunyasuvunakool *et al*. 2021; David *et al*. 2022; Monzon *et al*. 2022; Xu *et al*. 2023b; Abramson *et al*. 2024; Varadi *et al*. 2024]. To visually represent confidence levels, the 3D structure uses a color-coding scheme based on the pLDDT score: 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 [Mariani *et al*. 2013; Wroblewski & Kmiecik 2024; Li *et al*. 2025, 2026].

The main reasons for the low-confidence in AlphaFold predictions include the following: (1) Insufficient supporting data, as the AlphaFold-based prediction rely on the quality of 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

218 protein regions are inherently flexible under physiological conditions, such as the activation domains of
219 transcription factors, which do not have a fixed 3D structure, leading to low prediction confidence
220 assigned by AlphaFold. (3) In terms of special structural regions, such as small-molecule binding sites
221 and artificial linkers in fused proteins, the AlphaFold system may have limited prediction power, leading
222 to insufficient confidence in the prediction of complex structures. Thus, the low-confidence regions in
223 AlphaFold predictions are informative flags for highlighting challenges of prediction and should not be
224 interpreted with excessive confidence.

225 2.5 Correlations of the primary, secondary, and tertiary structures of the DNA-
226 binding domains of the paired MAT1-1-1 and MAT1-2-1 proteins

227 Among various AlphaFold 3D structural heteromorphs, the tertiary structural changes around the
228 mutation sites within the MATα_HMGbox domains of MAT1-1-1 proteins or the HMG-box_ROX1-like
229 domains of MAT1-2-1 proteins are locally magnified to visualize the subtle 3D variations. Primary
230 sequence variations in the DNA-binding domains of the mating proteins, such as peptide chain
231 truncations and amino acid replacements, are correlated with topological structural and waveform
232 changes in hydrophobicity, α-helix, β-sheet, β-turn, and coil presented in the ExPASy ProtScale plots,
233 and the locally magnified 3D structures around the mutation sites. The variant MAT1-1-1 and MAT1-2-
234 1 proteins produced by the *O. sinensis* strains are paired to explore diverse fungal origins.

235 3. Results

236 3.1 Primary structures of the truncated MAT1-1-1 proteins in *O. sinensis* strains

237 As aligned with the representative full-length MAT1-1-1 protein AGW27560 derived from the *H.*
238 *sinensis* strain CS68-2-1229 [Bushley *et al*. 2013], the 20 MAT1-1-1 proteins are differentially truncated
239 at both N- and C-termini (Figure 1). These truncated proteins of varying lengths were encoded by the
240 cDNAs amplified using the same pair of primers (the forward primer *m1cDNAF3* and the reverse
241 complementary primer *m1cDNAR5* shown in Figure S1 [Bushley *et al*. 2013]) from the *MAT1-1-1* genes
242 of *O. sinensis* strains. As represented by the full-length reference protein AGW27560 [Bushley *et al.*
243 2013], the MATα_HMGbox domain of the MAT1-1-1 proteins locates at amino acids 51→225 (shown
244 in pink in Figures 1 and S1). The MATα_HMGbox domains of the 20 MAT1-1-1 proteins shown in pink
245 in Figure 1 are truncated at N-termini and contributed to 9 AlphaFold 3D structure morphs (Figure 2,
246 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 [Bushley et al. 2013] and the 9 truncated MAT1-1-1 proteins (AGW27528, AGW27522, AGW27525, AGW27526, AGW27527, AGW27535, AGW27530, AGW27536, and AGW27534 [Bushley et al. 2013]) 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) 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.

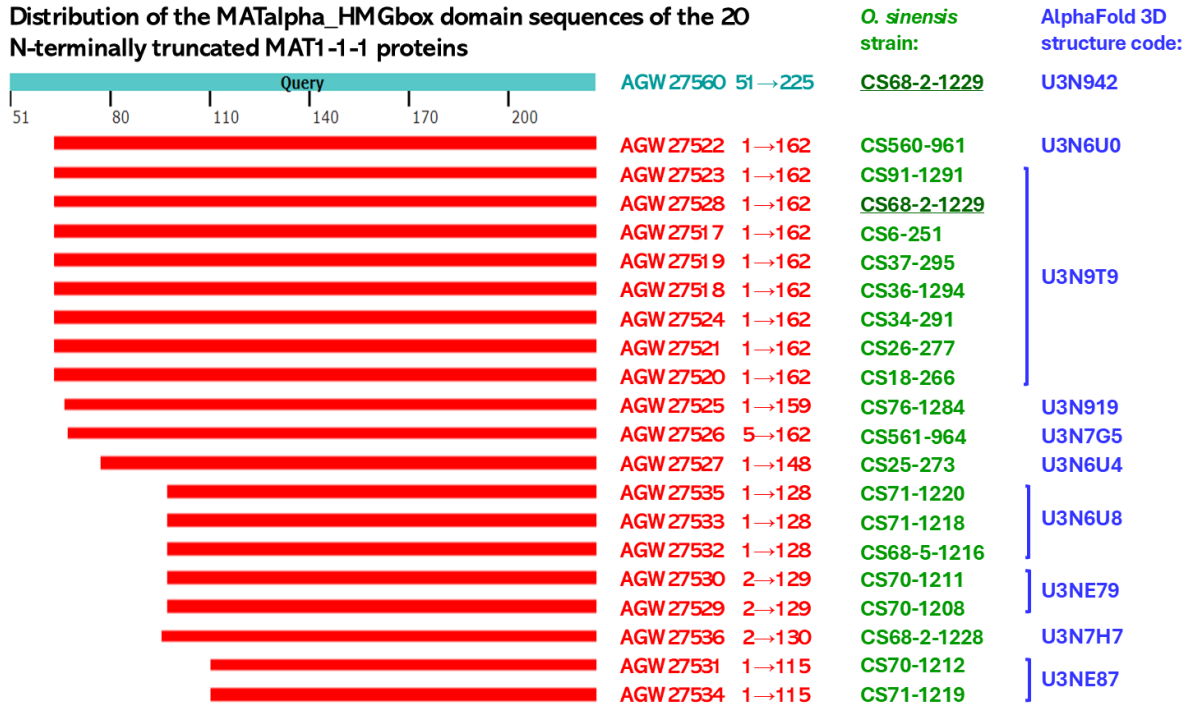


Figure 2. Sequence distribution of the N-terminally truncated MATα_HMGbox domains of 20 MAT1-1-1 proteins derived from *O. sinensis* strains (Table 1). The sequences are aligned with that of the full-length MATα_HMGbox domain (51→225) of the query MAT1-1-1 protein AGW27560 derived from the *H. sinensis* 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, which were produced by *O. sinensis* strains and are truncated at the C-termini (Figure 3). The truncated MAT1-2-1 proteins with 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 [Bushley *et al*. 2013]) from the total RNA extracted from various *O. sinensis* strains. The full-length authentic MAT1-2-1 protein AEH27625 under AlphaFold code U3N942derived from the *H. sinensis* strain CS2 is taken as the alignment reference, which contains an HMG-box_ROX1-like domain at amino acids 127→197 shown in pink in Figures 3 and S2. As aligned with the HMG-box_ROX1-like domain sequence of the query protein AEH27625 [Zhang *et al*. 2009], the domain sequences of the AMT1-2-1 proteins derived from the 5 *O. sinensis* strains are truncated at the C-termini and contain a Y-to-H substitution in each of the

truncated proteins, contributing to diverse 3D structures under 4 AlphaFold codes (Figure 4). The MAT1-2-1 protein AGW27554 contains an additional Q-to-T substitution 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 representing query full-length MAT1-2-1 protein AEH27625 of 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.

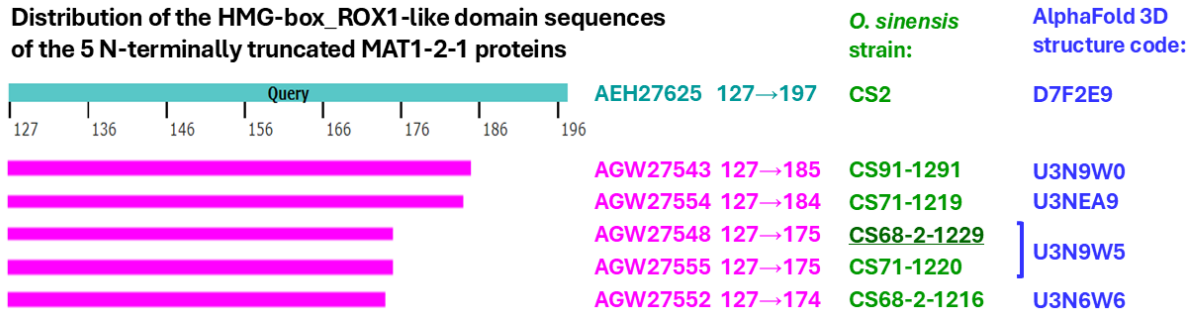


Figure 4. Sequence distribution of the C-terminally truncated HMG-box_ROX1-like domains of 5 truncated MAT1-2-1 proteins derived from the *O. sinensis* strains (*cf*. Table 1). The sequences are aligned with the HMG-box_ROX1-like domain (127→197) of the full-length MAT1-2-1 protein AEH27625 of *H. sinensis*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

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

Section 3.3 reports the correlation analyses between the changes in the hydrophobicity and structures of the MATα_HMGbox domains of truncated MAT1-1-1 proteins under 9 AlphaFold codes

287 (Table 1, Figures 1−2) derived from various *O. sinensis* strains, compared to the reference MAT1-1-1
288 proteins (under the AlphaFold code U3N942) represented by AGW27560 that was derived from the *O.*
289 *sinensis* strain CS68-2-1229 (Table S4) [Bushley et al. 2013]. The MATα_HMGbox domain sequences
290 (amino acid residues 51→225) were extended by 9 amino acids both upstream and downstream of the
291 domains because of the setting of the sequential ExPASy ProtScale plotting for hydropathy, α-helices,
292 β-sheets, β-turns, and coils at a window size of 9 amino acid residues (*cf*. Section 2.3).

293 Figure 5 compares the hydrophobicity property and the structures of the MATα_HMGbox domain
294 of 2 groups of the truncated MAT1-1-1 protein: (1) AGW27522 (under the AlphaFold code U3N6U0)
295 derived from the *O. sinensis* strain CS560-961 and (2) the truncated proteins AGW27517, AGW27518,
296 AGW27519, AGW27520, AGW27521, AGW27523, AGW27524, and AGW27528 (under the
297 AlphaFold code U3N9T9) derived from the *O. sinensis* strains CS6-251, CS36-1294, CS37-295, CS18-
298 266, CS26-277, CS91-1291, CS34-291, and CS68-2-1229, respectively (Table 1) with those of the
299 MATα_HMGbox domain of the query protein AGW27560 (under the AlphaFold code U3N942) derived
300 from the *H. sinensis* strain CS68-2-1229 [Bushley *et al*. 2013]. Panel (A) of Figure 5 shows 13-residue
301 truncations at the N-termini of MATα_HMGbox domains of the 2 groups of subject proteins, which
302 result in changes in topological structure and waveform in the ExPASy ProtScale plots for hydropathy,
303 α-helices, β-sheets, β-turns, and coils, as illustrated in Panels (B–F), respectively. Compared to the 3D
304 structure of the full-length protein AGW27560 as the comparison reference shown in Panel (G), where
305 the core structure of 3 α-helices within the MATα_HMGbox domain form a functional L-shape
306 stereostructure, Panels (H–I) illustrate that the truncations significantly altered the 3D structures of the
307 proteins under the heteromorphic AlphaFold codes U3N6U0 and U3N9T9, respectively. The truncation
308 regions of the variant proteins are located upstream of the core structure of 3 α-helices and may have
309 steric auxiliary influence on stabilizing the hydrophobic core structures of the MATα_HMGbox domains
310 [Baxevanis *et al*. 1995; Thapar 2015].

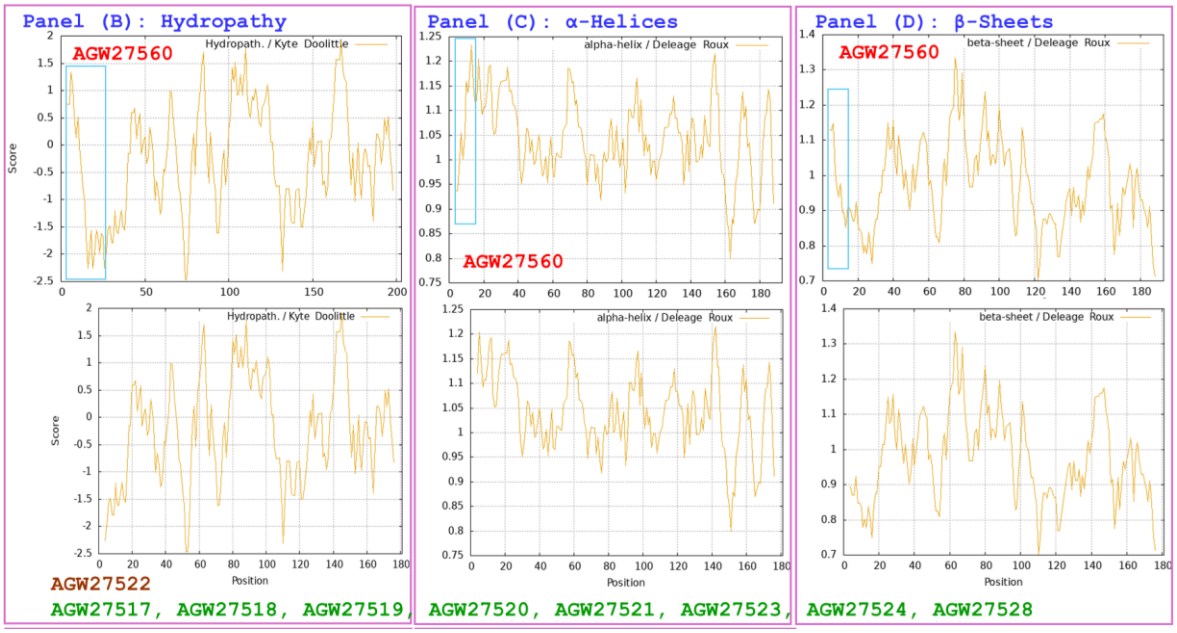
Panel (A): Alignment of amino acid sequences of the MATα_HMGbox domains

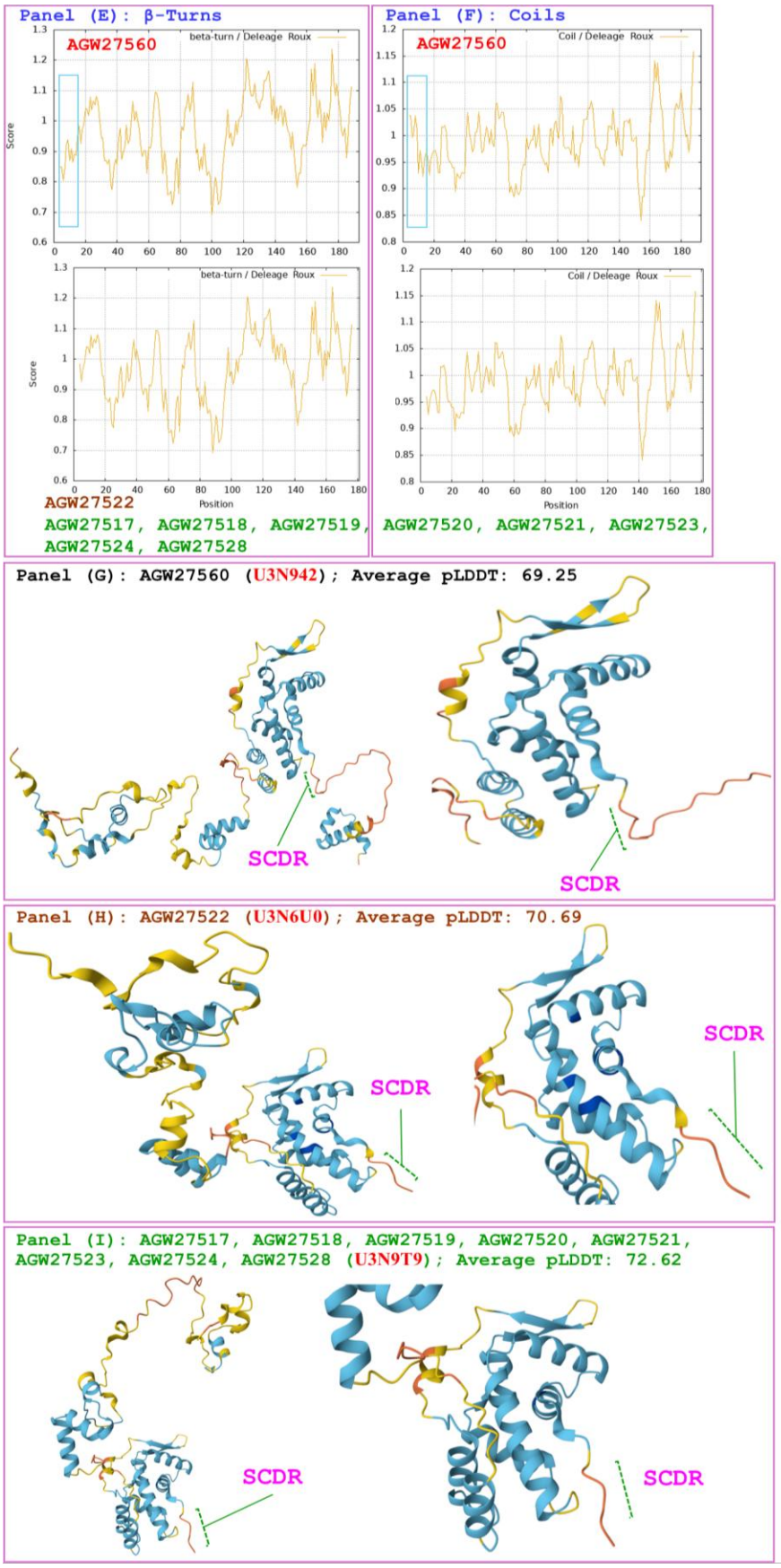
```
AGW27560  42   ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTT  100
AGW27522  1                                                                 37
AGW27517  1                                                                 37
AGW27518  1                                                                 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

313

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 the *H. sinensis* strain CS68-2-1229 and 2 groups of the truncated protein: (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, respectively. Panel (A) shows an alignment of the amino acid sequences of the MATα_HMGbox domains of the MAT1-1-1 proteins; whereas the hyphens indicate identical amino acid residues, 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 proteins, respectively; the open rectangles in blue highlight the N-terminal truncation region in the ExPASy plots. Panels (G–I) show the 3D structures of the entire proteins on the left and the locally magnified structures surrounding the truncation region 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).

Figure 6 compares the hydrophobicity property and structures of the MATα_HMGbox domain of the truncated subject MAT1-1-1 protein AGW27525 (under the AlphaFold code U3N919) derived from the *O. sinensis* strain CS76-1284 (Table 1) with those of the MATα_HMGbox domain of the reference full-length protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [Bushley *et al*. 2013]. Panel (A) of Figure 6 shows a 16-residue truncation at the N-terminus of the MATα_HMGbox domain of the proteins AGW27525, which results in altered topological structure and waveform in the ExPASy ProtScale plots for hydropathy, α-helices, β-sheets, β-turns, and coils, as illustrated in Panels (B–F), respectively. Compared to the 3D structure of the reference protein AGW27560 shown in Panel (G), Panel (H) illustrates the truncation variation and significantly altered 3D structure of the protein AGW27525 under the AlphaFold code U3N919. In particular, 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 [*cf*. Kyte & Doolittle 1982]), featuring hydrophilic feature in combination, and form the beginning of the β-sheet structure in the truncated protein AGW27525 in Panel (H); while the amino acids AKR in the full-length protein AGW27560 are not part of the β-sheet structure and the hydrophobic core of 3 critical α-helices within the MATα_HMGbox domain in Panel (G). The truncation region of the protein AGW27525 is located upstream of the core structure of 3 α-helices within the MATα_HMGbox domain and may not directly participate in stabilizing the hydrophobic core of the MATα_HMGbox domain [Baxevanis *et al*. 1995; Thapar 2015].

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

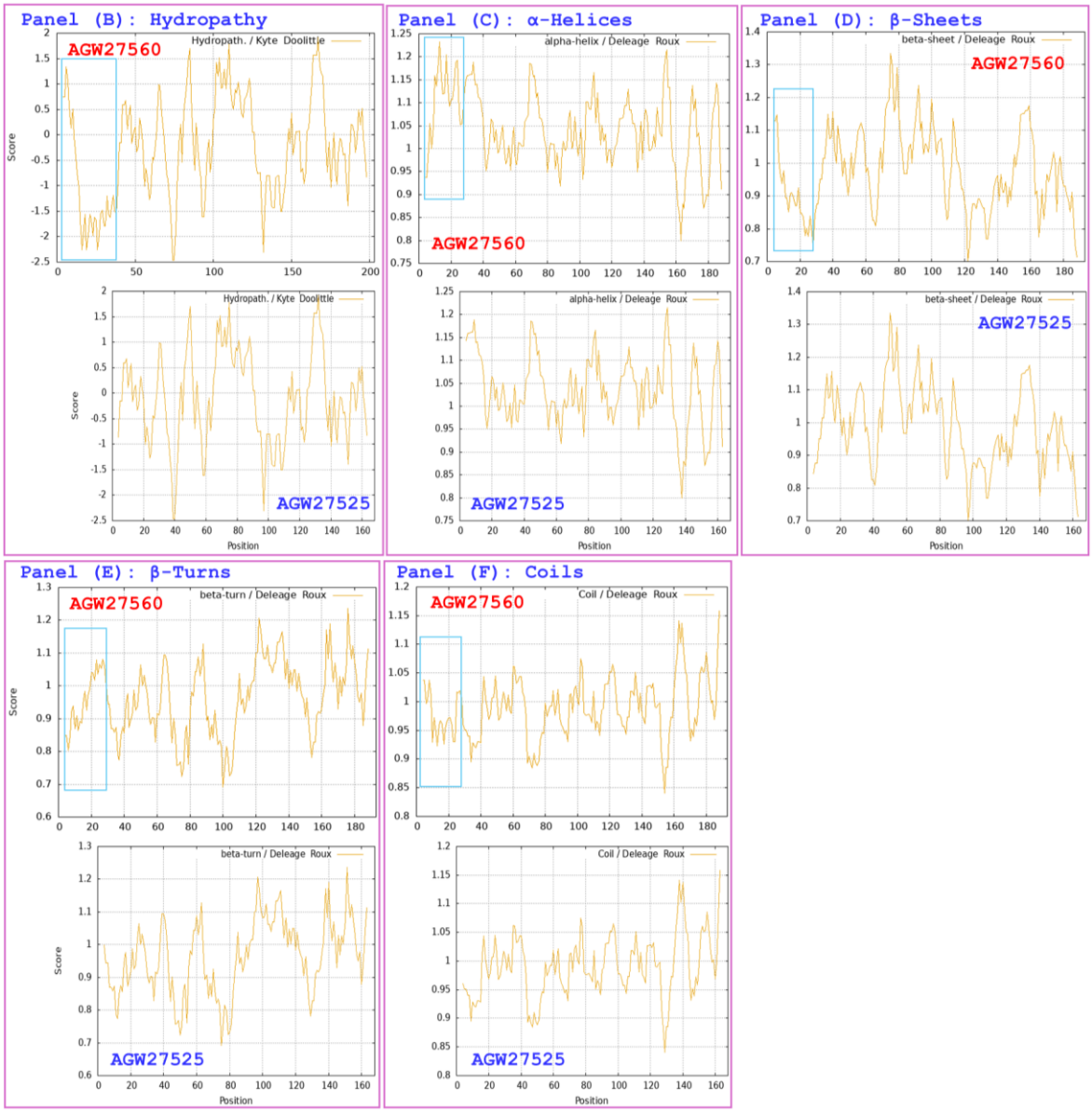
```
AGW27560  42   ADVLTPVSTAAASRATRQTKEASCDRAKRPLNAFMAFRSYYLKLFPDVQQKTASGFLTT  100
AGW27525  1                                ---------------------------------  34

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

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

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

346



347

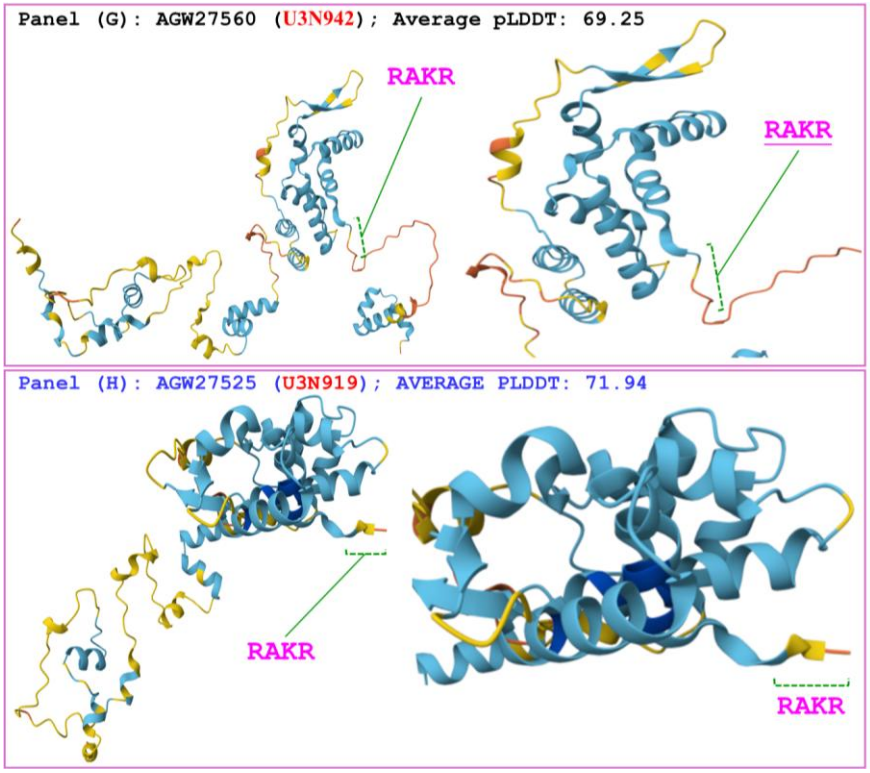


Figure 6. Correlations of the changes in hydrophobicity and the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the variant protein AGW27525 (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; whereas the hyphens indicate identical amino acid residues, 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 region in the ExPASy plots. Panels (G–H) show the 3D structures of the entire proteins on the left and the locally magnified structures surrounding the variation region 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).

Figure 7 compares the hydrophobicity property and the structures of the MATα_HMGbox domain of the variant MAT1-1-1 protein AGW27526 (under the AlphaFold code U3N7G5) derived from the *O. sinensis* strain CS561-964 (Table 1) with those of the MATα_HMGbox domain of the full-length protein AGW27560 as the comparison reference (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [Bushley *et al*. 2013]. Panel (A) of Figure 7 shows a 13-residue truncation at the N-terminus of the MATα_HMGbox domain of the proteins AGW27526 and amino acid substitutions of 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) [Kyte & Doolittle 1982]. The truncation and amino acid substitutions result in changes in hydrophobicity and secondary structures of the variant protein, as illustrated by the altered topological structure and waveform in the ExPASy ProtScale plots for hydropathy, α-helices, β-sheets, β-turns, and coils, as illustrated in Panels (B–F), respectively. Compared with the 3D structure of the reference protein AGW27560 shown in Panel (G), the truncation and replaced 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, which caused apparent changs in the tertiary structure of the variant protein under the AlphaFold code U3N919.

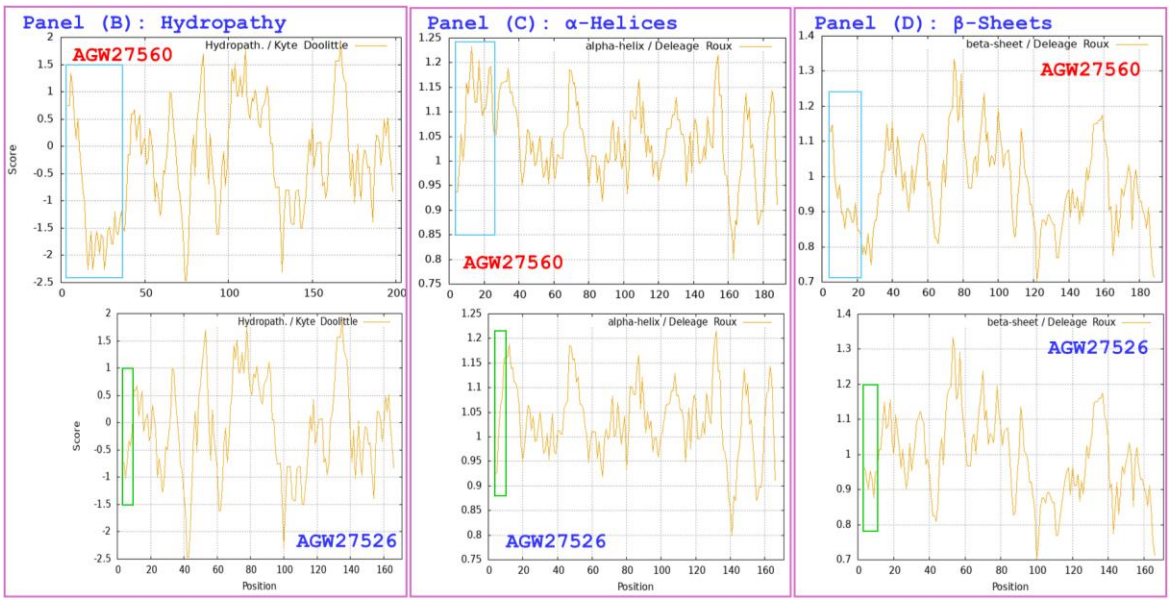
Panel (A): Alignment of the amino acid sequences of the MATα_HMGbox domains

```
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
```



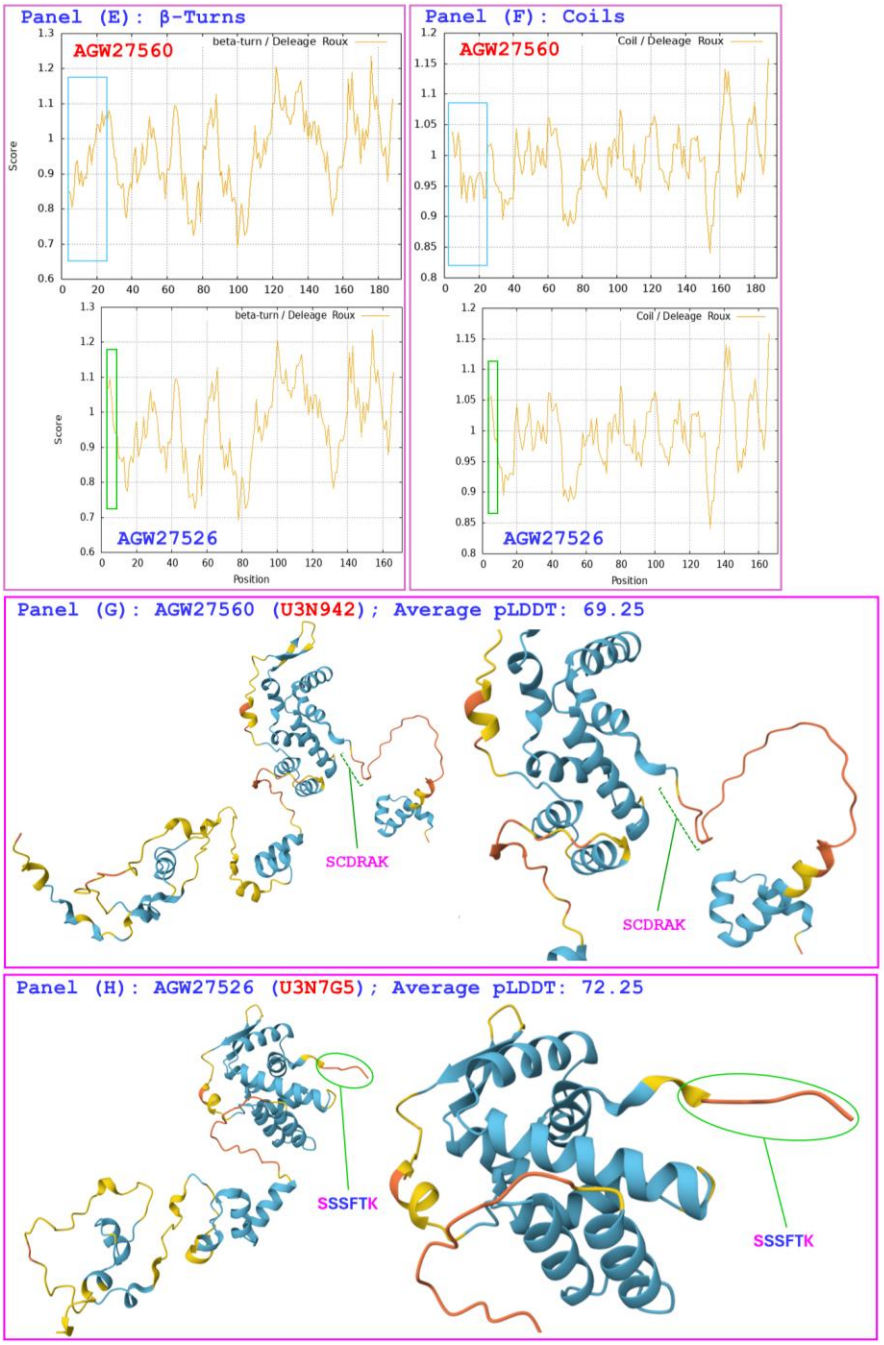


Figure 7. Correlations of the changes in the hydrophobicity and the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the variant protein AGW27526 (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; whereas the hyphens indicate identical amino acid residues, 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 and those in green highlight the altered topological structure and waveform in the ExPASy plots. Panels (G–H) show the 3D structures of the entire proteins on the left and the locally magnified structures surrounding the variation regions 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).

Figure 8 compares the hydrophobicity property and the structures of the MATα_HMGbox domain of the variant MAT1-1-1 protein AGW27527 (under the AlphaFold code U3N6U4) derived from the *O. sinensis* strain CS25-273 (Table 1) with those of the MATα_HMGbox domain of the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [Bushley *et al*. 2013]. Panel (A) of Figure 8 shows a 27-residue truncation at the N-terminus of the MATα_HMGbox domain of the protein AGW27527, which results in changes in topological structure and waveform in the ExPASy ProtScale plots in Panels (B–F) for hydropathy, α-helices, β-sheets, β-turns, and coils, respectively. Compared to the 3D structure of the reference protein AGW27560 shown in Panel (G), Panel (H) shows that the truncation that is located within the first of the 3 core α-helices and the apparent alteration of the tertiary structure of the MATα_HMGbox domain of the truncated protein AGW27527 under the AlphaFold code U3N6U4.

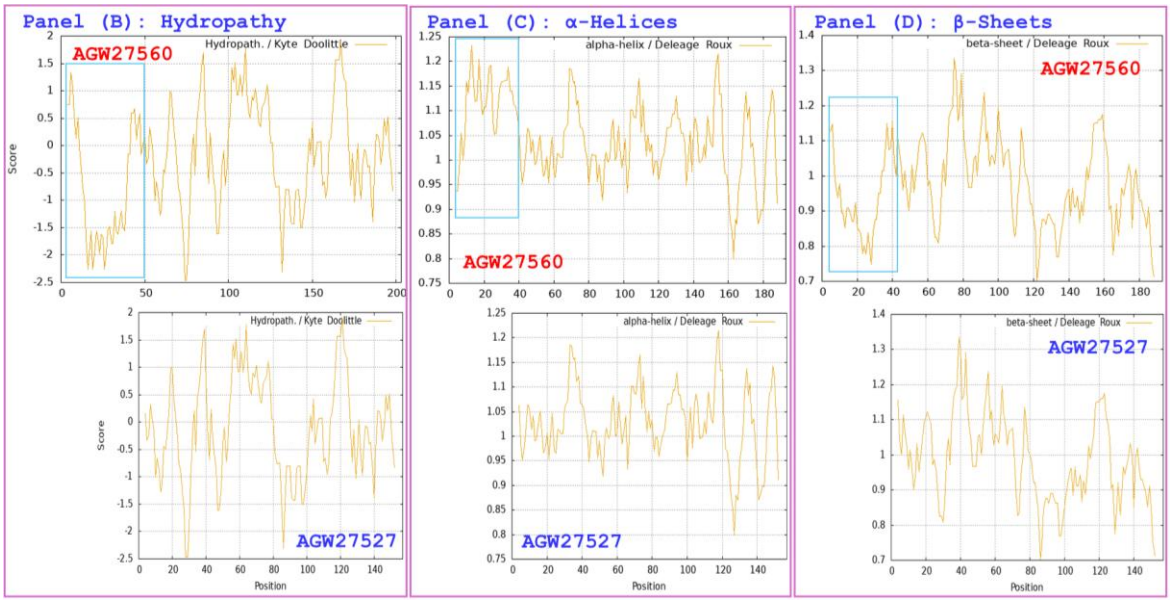
Panel (A): Alignment of the amino acid sequences of the MATα_HMGbox domains

```
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
```



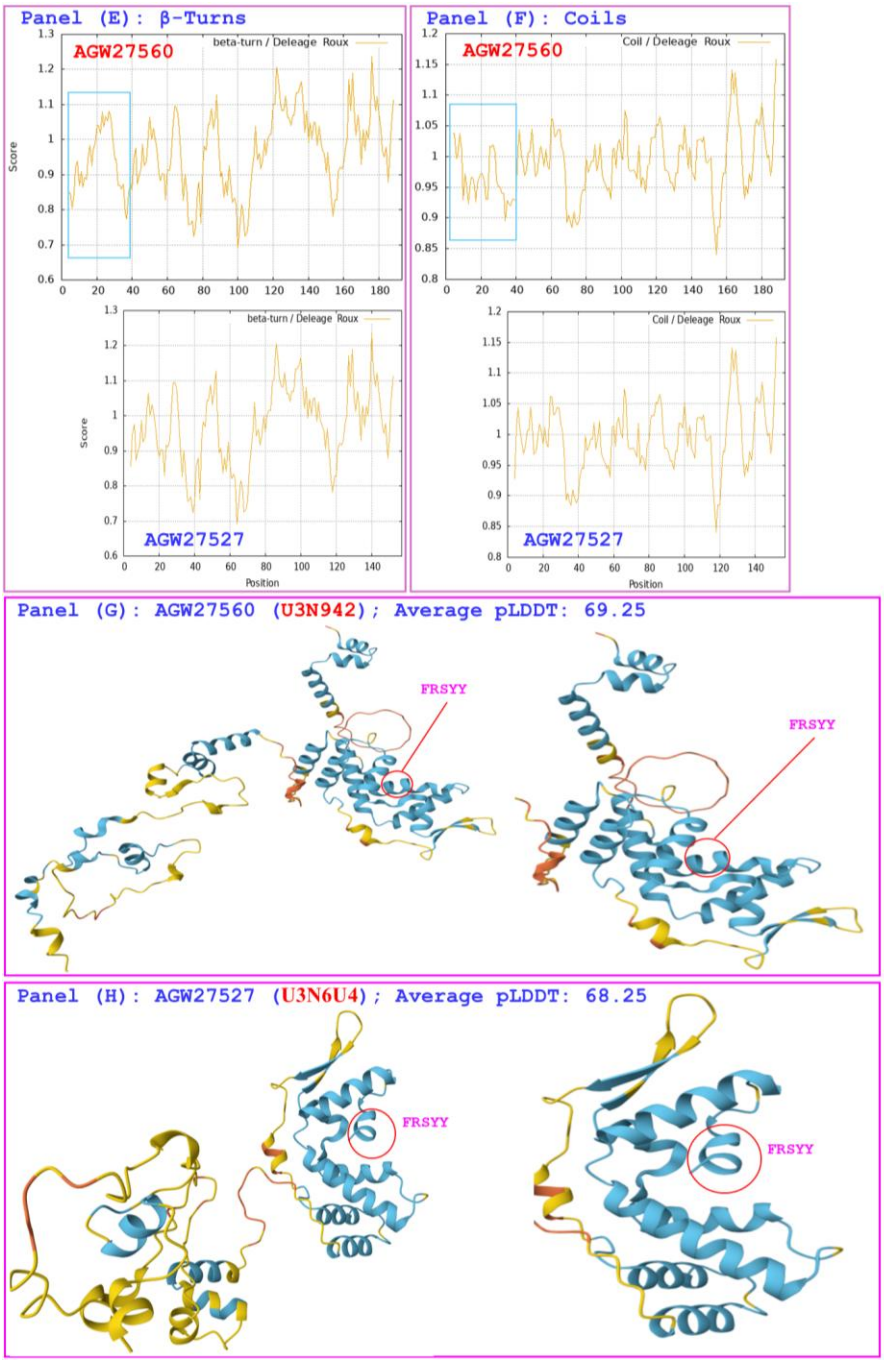


Figure 8. Correlations of the changes in the hydrophobicity and the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the variant protein AGW27527 (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; whereas the hyphens indicate identical amino acid residues, 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 entire proteins on the left and the locally magnified structures surrounding the truncation region 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).

Figure 9 compares the hydrophobicity property and the structures of the MATα_HMGbox domain of the variant MAT1-1-1 proteins AGW27532, AGW27533, and AGW27535 (under the AlphaFold code U3N6U8) derived from the *O. sinensis* strains CS68-5-1216, CS71-1218, and CS71-1220 (Table 1) with those of the MATα_HMGbox domain of the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [Bushley *et al*. 2013]. Panel (A) of Figure 9 shows 47-residue truncations at the N-termini of the MATα_HMGbox domains of the variant proteins; the truncations result in changes in topological structure and waveform in the ExPASy ProtScale plots in Panels (B–F) for hydropathy, α-helices, β-sheets, β-turns, and coils, respectively. Comparison of the tertiary structures of the MAT1-1-1 proteins shows the truncations of the variant proteins, as 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.

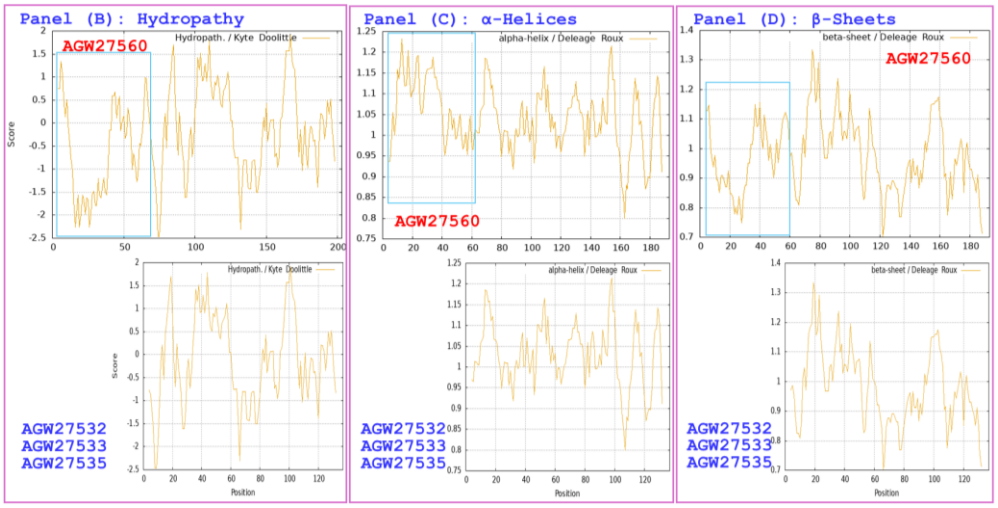
Panel (A): Alignment of the amino acid sequences of the MATα_HMGbox domains

```
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
```



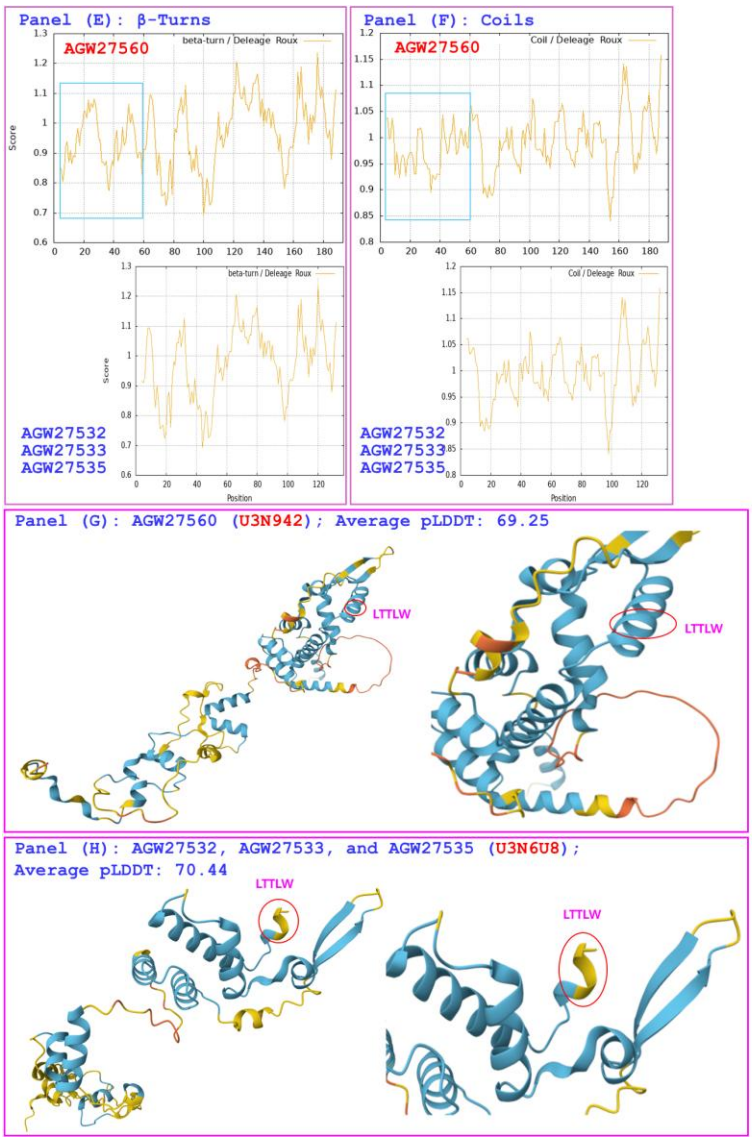


Figure 9. Correlations of the changes in the hydrophobicity and the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the variant protein AGW27532, 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; whereas the hyphens indicate identical amino acid residues, 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–I) show the 3D structures of the entire 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).

Figure 10 compares the hydrophobicity property and the structures of the MATα_HMGbox domains of the variant MAT1-1-1 proteins AGW27529 and AGW27530 (under the AlphaFold code U3NE79) derived from the *O. sinensis* strain CS70-1208 and CS70-1211, respectively (Table 1), with those of the

MATα_HMGbox domain of the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [Bushley *et al*. 2013]. Panel (A) of Figure 10 shows 46-residue truncations at the N-termini of the MATα_HMGbox domains of the proteins AGW27529 and AGW27530 and an F-to-V substitutions (aromatic phenylalanine to aliphatic valine; hydropathy indices changed from 2.8 to 4.2 (Table S3 [*cf.* Kyte & Doolittle 1982]) within the MATα_HMGbox domains, indicating increased hydrophobicity. The truncation and increased hydrophobicity result in altered secondary structure, as illustrated by the changes in topological structure and waveform in the ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils shown in Panels (B–F), respectively. Compared the 3D structures of the MAT1-1-1 proteins as illustrated in Panels (G–H), the truncation and replaced amino acids in the proteins AGW27529 and AGW27530 are located within the first of the 3 core α-helices of the MATα_HMGbox domains, resulting in significant change in the tertiary structures of the proteins AGW27529 and AGW27530 under the AlphaFold code U3NE79.

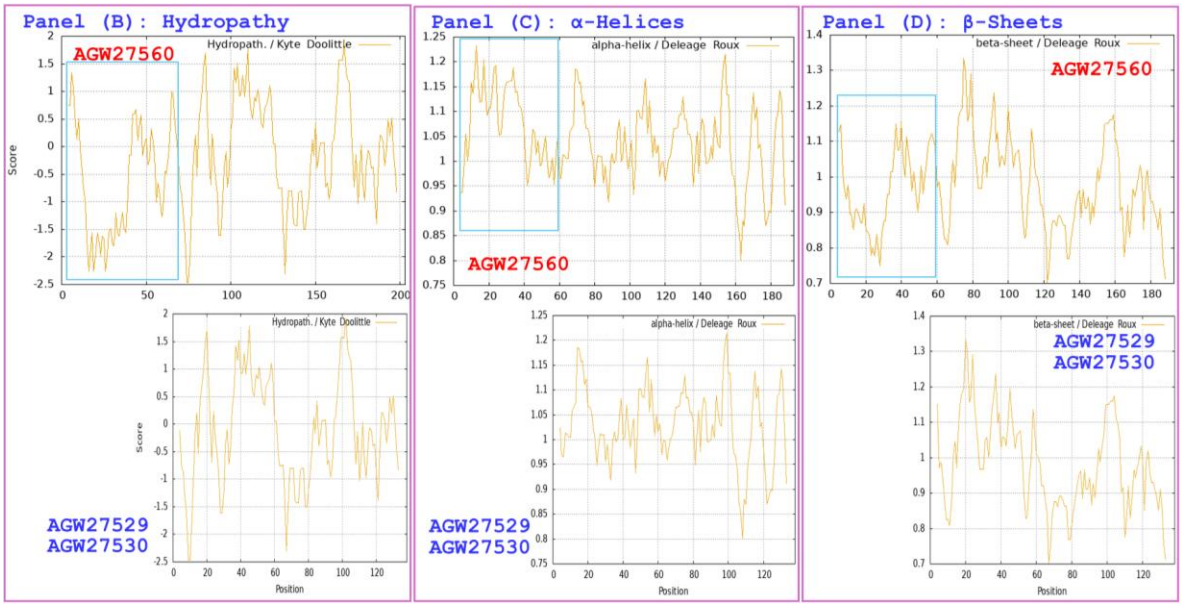
Panel (A): Alignment of the amino acid sequences of the MATα_HMGbox domains

```
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
```



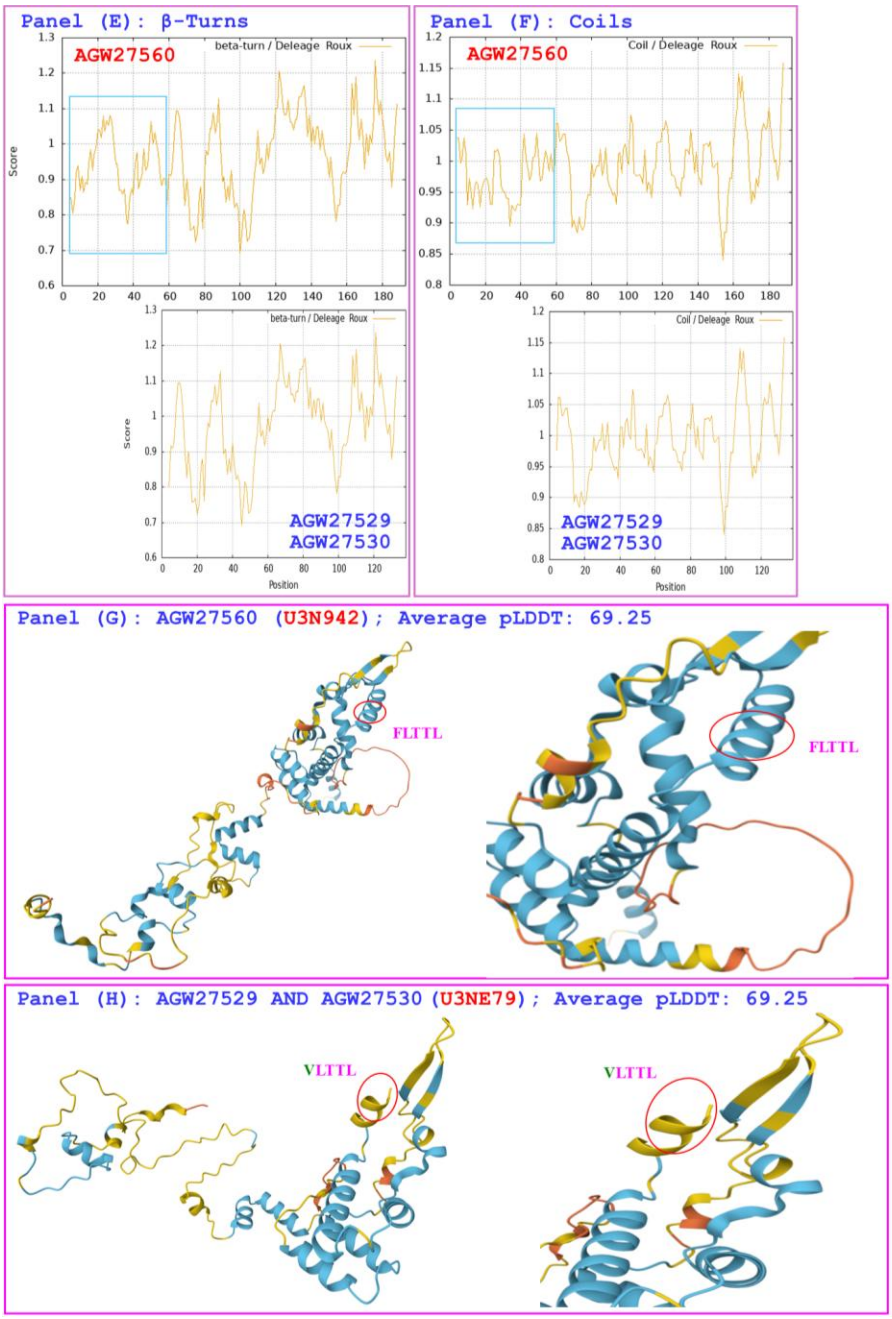


Figure 10. Correlations of the changes in the hydrophobicity and the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the variant proteins AGW27529 and AGW27530 (under the AlphaFold code U3NE79) derived from the *O. sinensis* strain 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, whereas the hyphens indicate identical amino acid residues, 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–I show the 3D structures of the entire proteins on the left and the locally magnified structures surrounding the variation 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).

Figure 11 compares the hydrophobicity property and the structures of the MATα_HMGbox domain of the variant MAT1-1-1 protein AGW27536 (under the AlphaFold code U3N7H7) derived from the *O. sinensis* strain CS68-2-1228 (Table 1) with those of the MATα_HMGbox domain of the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [Bushley *et al*. 2013]. Panel (A) of Figure 11 shows a 44-residue truncation at the N-terminus of the MATα_HMGbox domain of the protein AGW27536 and an S-to-E substitution (serine with polar neutral side chains to acidic glutamic acid; hydropathy indices changed from -0.8 to -3.5; *cf*. Table S3 [Kyte & Doolittle 1982]) and a deletion of leucine (hydropathy index 3.8) within the MATα_HMGbox domain, indicating largely reduced hydrophobicity. The truncation and reduced hydrophobicity result in altered secondary structure, as illustrated by the changes in topological structure and waveform in the ExPASy ProtScale plots in Panels (B–F) for hydropathy, α-helices, β-sheets, β-turns, and coils, respectively. Compared the 3D structures of the MAT1-1-1 proteins as illustrated in Panels (G–H), the truncation and replaced amino acids and the deletion of the protein AGW27536 are located within the first of the 3 core α-helices of the MATα_HMGbox domain, resulting in significant change in the tertiary structure of the proteins AGW27536 under the AlphaFold code U3N7H7.

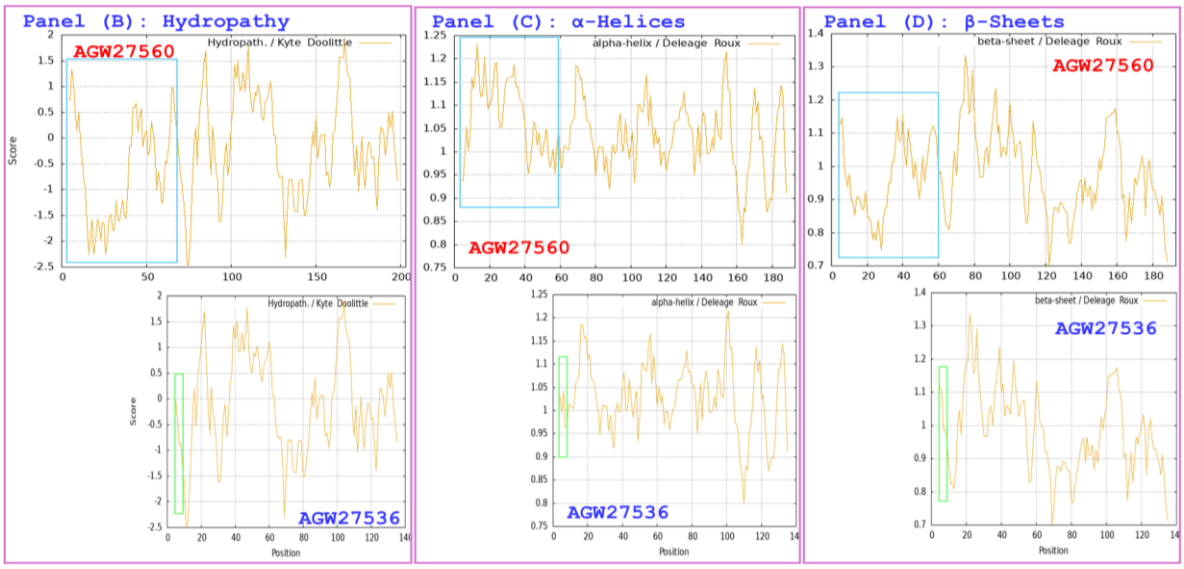
Panel (A): Alignment of the amino acid sequences of the MATα_HMGbox domains

```
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
```



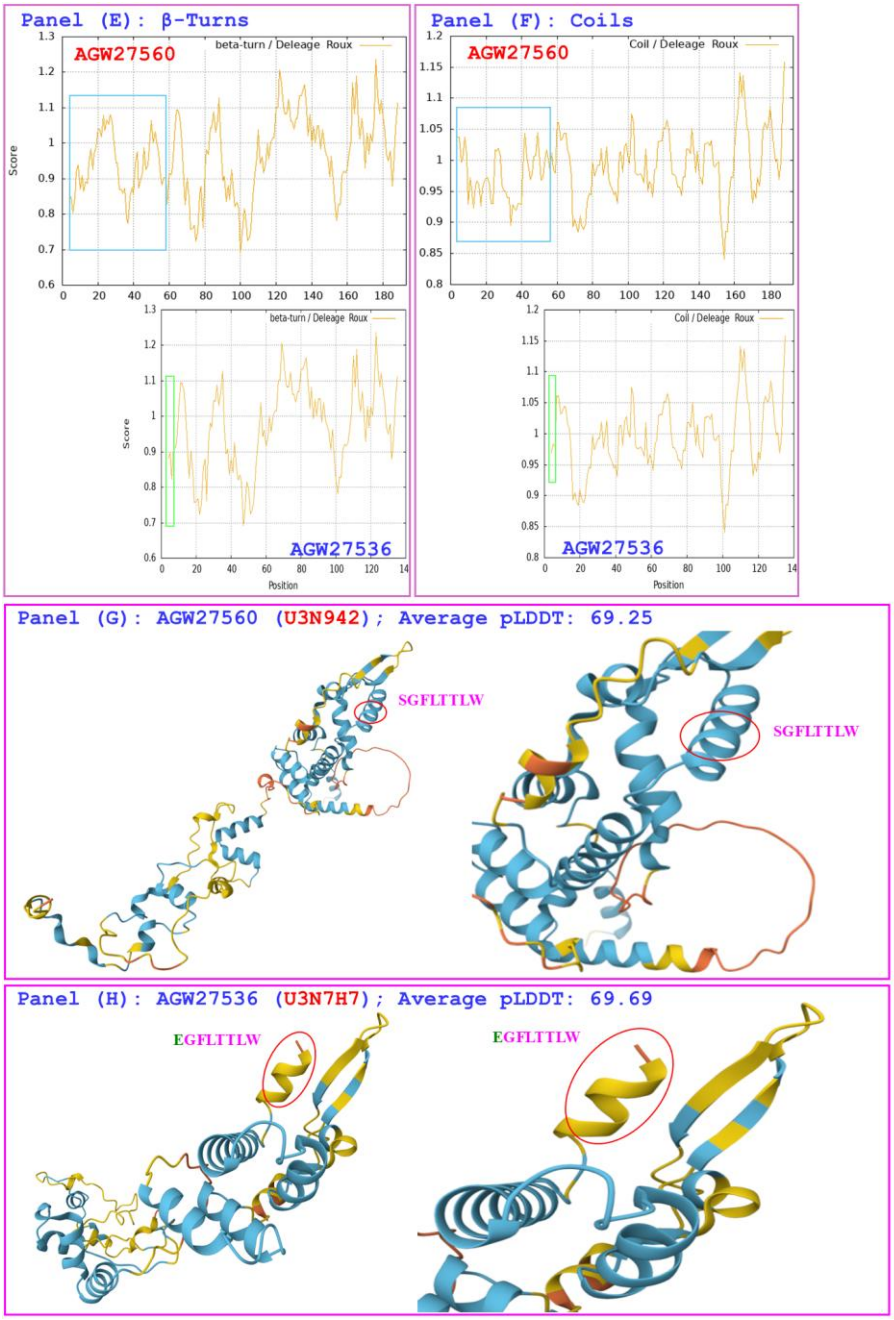


Figure 11. Correlations of the changes in the hydrophobicity and the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 and the variant protein 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; whereas the hyphens indicate identical amino acid residues, 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 and those in green highlight the altered topological structure and waveform in the ExPASy plots. Panels (G–H) show the 3D structures of the entire proteins on the left and the locally magnified structures surrounding the variation 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).

Figure 12 compares the hydrophobicity property and the structures of the MATα_HMGbox domains of the variant MAT1-1-1 proteins AGW27531 and AGW27534 (under the AlphaFold code U3NE87) derived from the *O. sinensis* strains CS70-1212 and CS71-1219, respectively (*cf.* Table 1) with those of the MATα_HMGbox domain of the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 [Bushley *et al*. 2013]. 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 in Panels (B–F) for hydropathy, α-helices, β-sheets, β-turns, and coils, respectively. Comparison of the 3D structures of the MAT1-1-1 proteins as illustrated in Panels (G–H) shows the truncations in the proteins AGW27531 and AGW27534, which are located within the 3 core α-helices of the MATα_HMGbox domains and result in the significantly altered tertiary structures of the variant proteins under the AlphaFold code U3NE87.

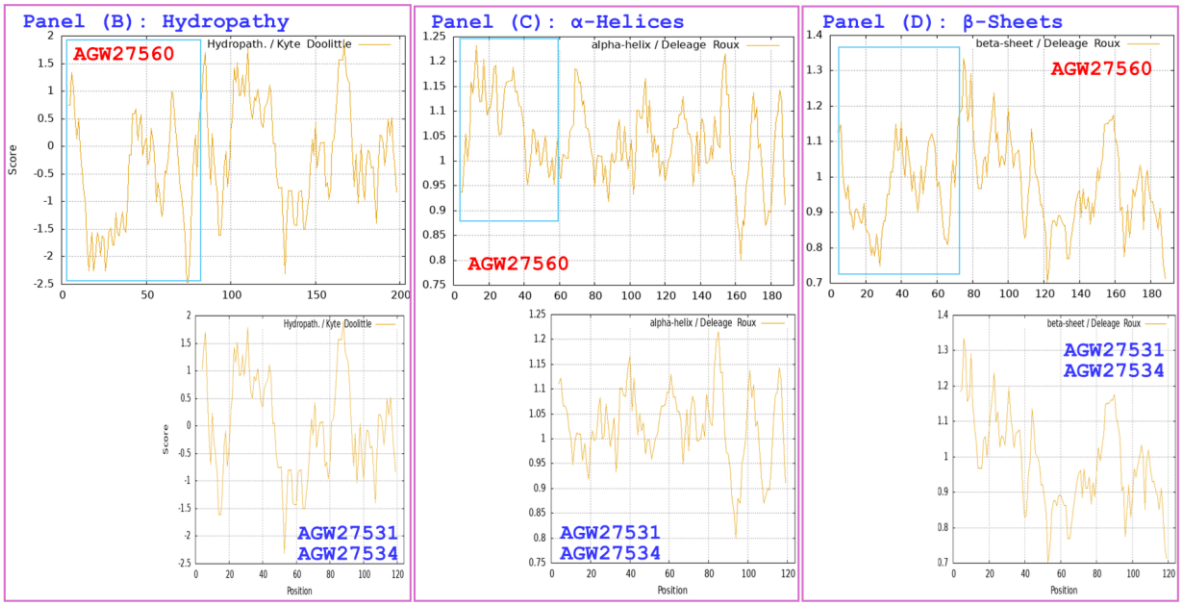
Panel (A): Alignment of the amino acid sequences of the MATα_HMGbox domains

```
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
```



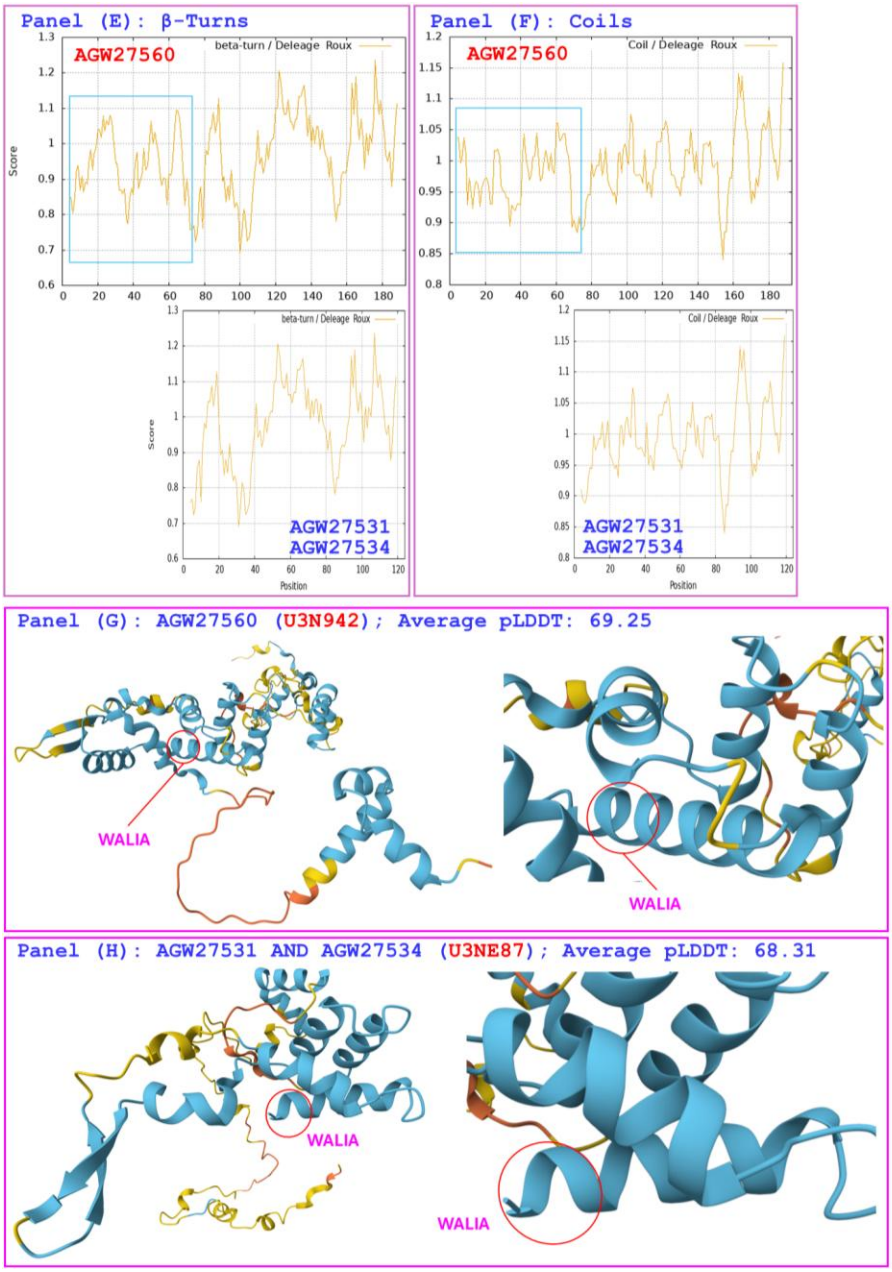


Figure 12. Correlations of the changes in the hydrophobicity and the primary, secondary, and tertiary structures of the MATα_HMGbox domains of MAT1-1-1 proteins: the reference protein AGW27560 (under the AlphaFold code U3N942) derived from the *H. sinensis strain* CS68-2-1229 and the variant proteins 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; whereas the hyphens indicate identical amino acid residues, 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 entire 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).

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

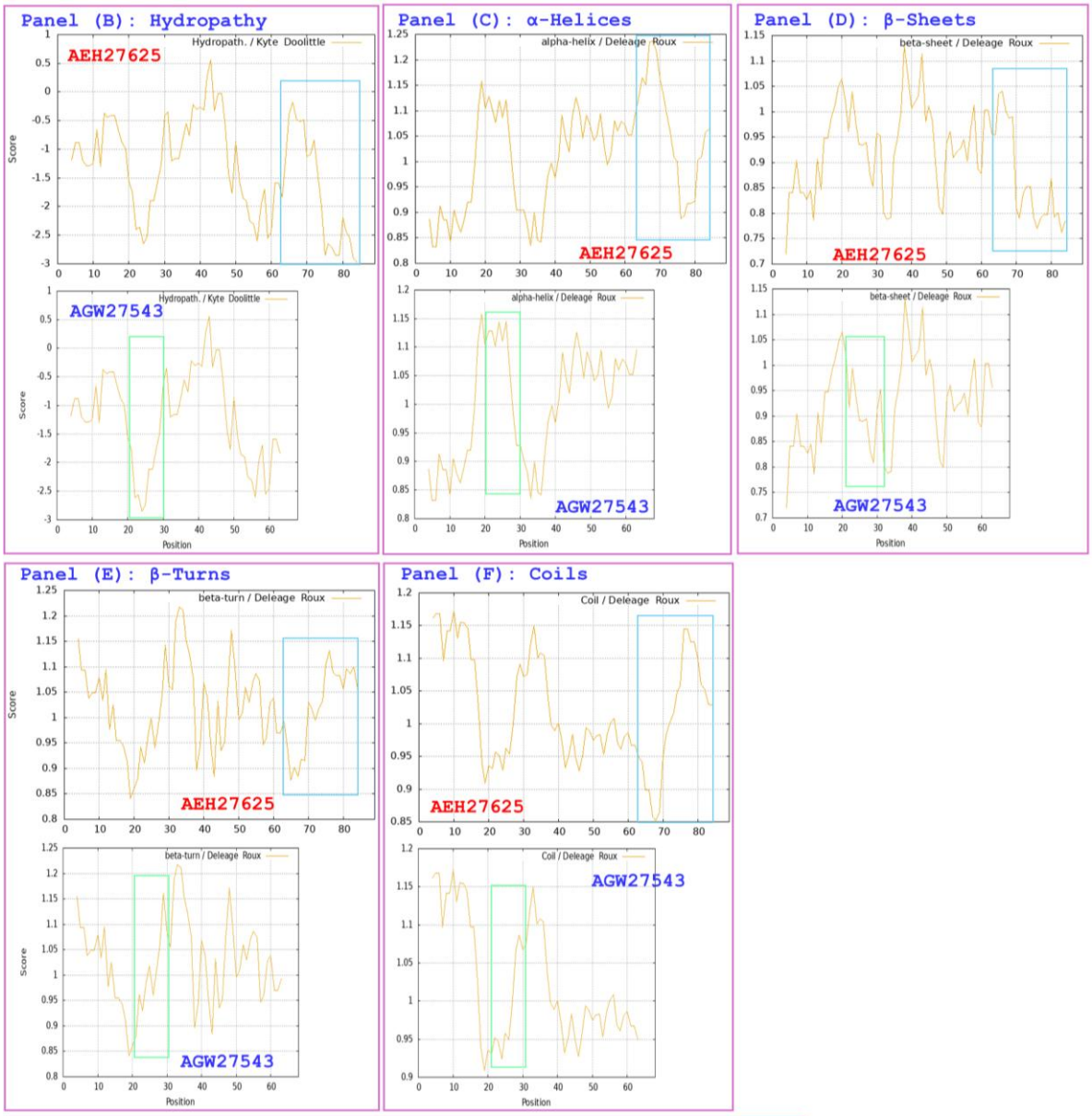
Section 3.4 presents the correlation analyses between the changes in the hydrophobicity property and structures of the HMG-box_ROX1-like domains of variant MAT1-2-1 proteins under 4 AlphaFold codes (*cf*. Table 1) derived from various *O. sinensis* strains, compared to the reference full-length MAT1-2-1 proteins (under the AlphaFold code D7F2E9) represented by AEH27625 derived from the *H. sinensis* strain CS2 (*cf*. Table S5) [Zhang et al. 2009]. 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 because the sequential ExPASy ProtScale plotting for hydropathy, α-helices, β-sheets, β-turns, and coils was set up at a window size of 9 amino acids (*cf*. Section 2.5).

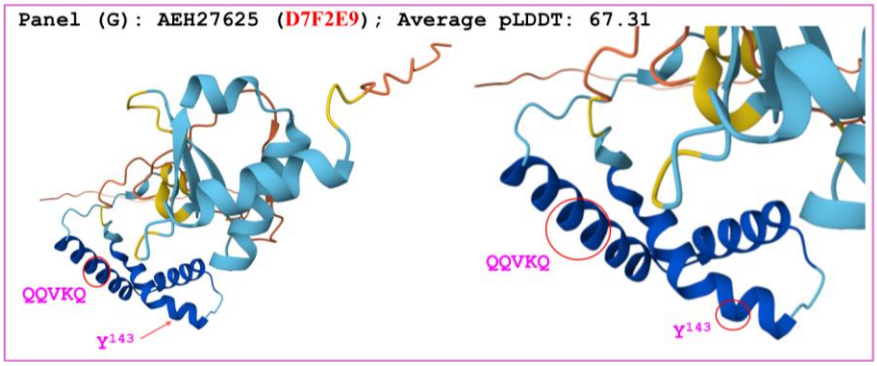
Figure 13 compares the hydrophobicity property and structures of the HMG-box_ROX1-like domain of the MAT1-2-1 protein AGW27543 (under the AlphaFold code U3N9W0) derived from the *O. sinensis* strain CS91-1291 (*cf*. Table 1) [Bushley *et al*. 2013] with those of the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [Zhang *et al*. 2009]. Panel (A) of Figure 13 shows a 12-residue truncation at the C-terminus of the HMG-box_ROX1-like domain of the protein AGW27543 and an amino acid substitution of Y-to-H (aromatic tyrosine to basic histidine; hydropathy indices changed from -1.3 to -3.2; *cf*. Table S3 [Kyte & Doolittle 1982]). The truncation and amino acid substitution with reduced hydrophobicity resulted in altered secondary structures surrounding the variation sites within the HMG-box_ROX1-like domain of the protein AGW27543, as illustrated by the changes in topological structure and waveform in the ExPASy ProtScale plots in Panels (B–F) for hydropathy, α-helices, β-sheets, β-turns, and coils, respectively. As shown in Panels (G–H) for 3D structures, the truncation and replaced amino acid residues are located in 2 of the 3 core α-helices and result in altered stereostructure 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

559

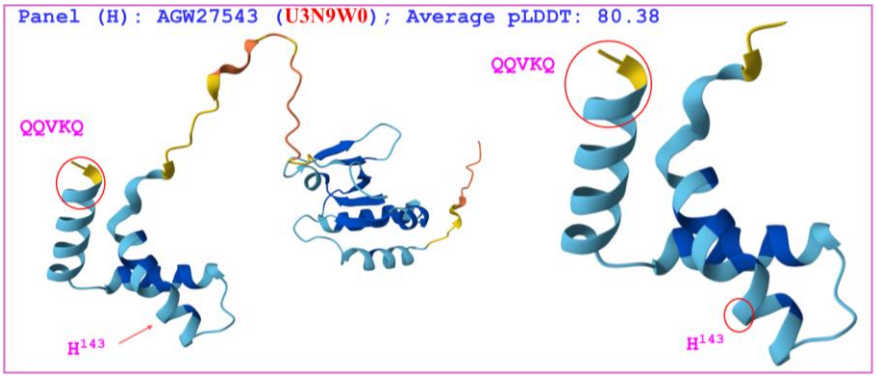


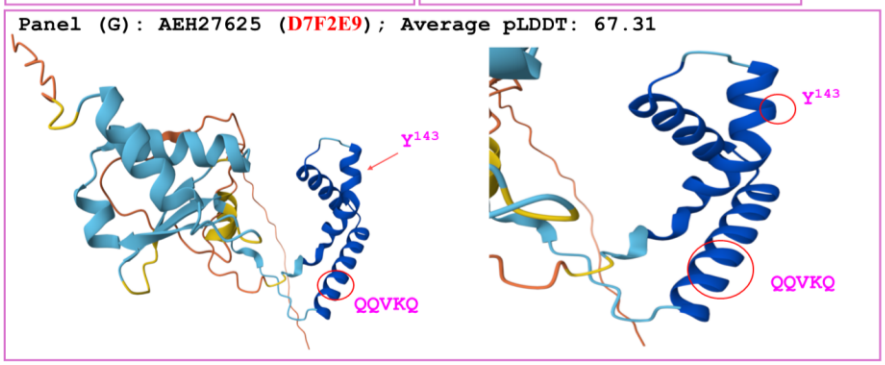
Figure 13. Correlations of the changes in hydrophobicity and the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins. The reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2, and the variant MAT1-2-1 protein 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, whereas the hyphens indicate identical amino acid residues, 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-terminal truncation regions in the ExPASy plots and those in green highlight the changes in topological structure and waveform seen in the plots. Panels (G–H) show representations of the 3D structures of the entire 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).

Figure 14 compares the hydrophobicity property and structures of the HMG-box_ROX1-like domain of the MAT1-2-1 protein AGW27554 (under the AlphaFold code U3NEA9) derived from the *O. sinensis* strain CS71-1219 [Bushley *et al*. 2013] with those of the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [Zhang *et al*. 2009]. 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 and amino acid substitutions: Y-to-H (aromatic tyrosine to basic histidine; hydropathy indices changed from -1.3 to -3.2) and Q-to-T (glutamine to threonine, both with polar neutral side chains; hydropathy indices changed from -3.5 to -0.7; *cf*. Table S3 [Kyte & Doolittle 1982]). The truncation mutation and amino acid substitutions with altered hydrophobicity result in changes in the secondary structure surrounding the mutation sites in the HMG-box_ROX1-like domain of the protein AGW27543, as illustrated by the changes in topological structure and waveform in the ExPASy ProtScale plots in Panels (B–F) for hydropathy, α-helices, β-sheets, β-turns, and coils, respectively. As illustrated in Panels (G–H) for 3D structures, the truncated and replaced residues are located in 2 of the 3 core α-helices and result in altered stereostructure 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
```



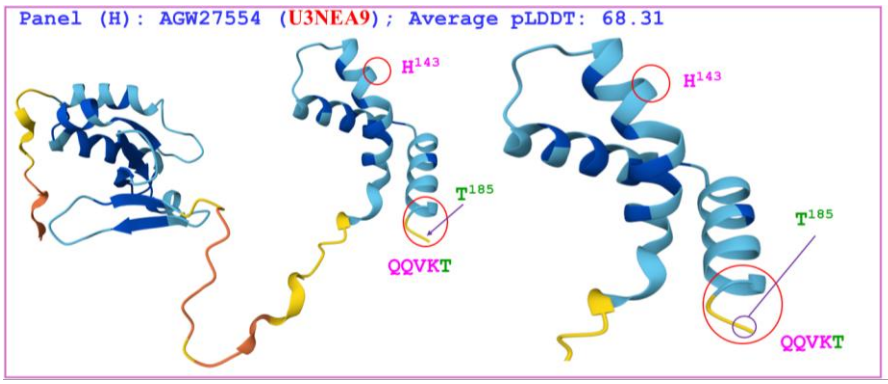


Figure 14. Correlations of the changes in hydrophobicity and the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins. The reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2, and the variant MAT1-2-1 protein 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, whereas the hyphens indicate identical amino acid residues, 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 topological structure and waveform. Panels (G–H) show representations of the 3D structures of the entire 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).

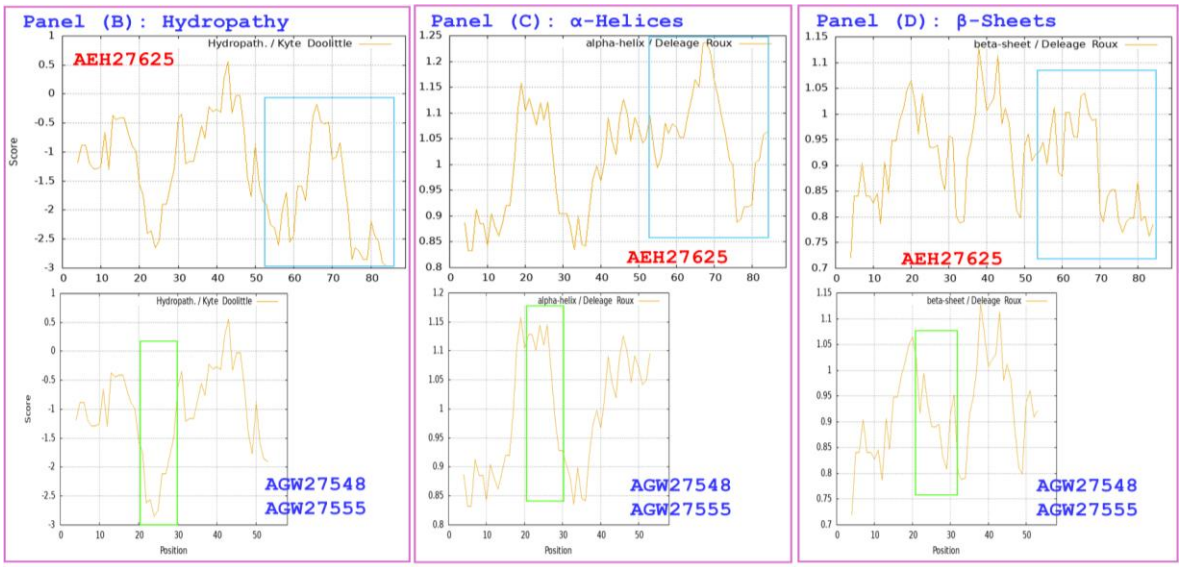
Figure 15 compares the hydrophobicity and structures of the HMG-box_ROX1-like domains of the MAT1-2-1 proteins AGW27548 and AGW27555 (under the AlphaFold code U3N9W5) derived from the *O. sinensis* strains CS68-2-1229 and CS71-1220 (*cf*. Table 1) [Bushley *et al*. 2013] with those of the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [Zhang *et al*. 2009]. Panel (A) of Figure 15 shows 22-residue truncations at the C-termini of HMG-box_ROX1-like domains of the MAT1-2-1 proteins AGW27548 and AGW27555 and amino acid substitutions in both variant proteins: Y-to-H (aromatic tyrosine to basic histidine; hydropathy indices changed from -1.3 to -3.2; *cf*. Table S3 [Kyte & Doolittle 1982]) within the HMG-box_ROX1-like domains. The truncation and amino acid substitutions caused decreased hydrophobicity and altered secondary structures surrounding the variation sites in the HMG-box_ROX1-like domains of the MAT1-2-1 proteins AGW27548 and AGW27555, as illustrated by the changes in topological structure and waveform in Panels (B–F) for the hydropathy, α-helices, β-sheets, β-turns, and coils, respectively. As shown in Panels (G–H) for 3D structures, the truncation and replaced residues are located in 2 of the 3 core α-helices, resulting 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.

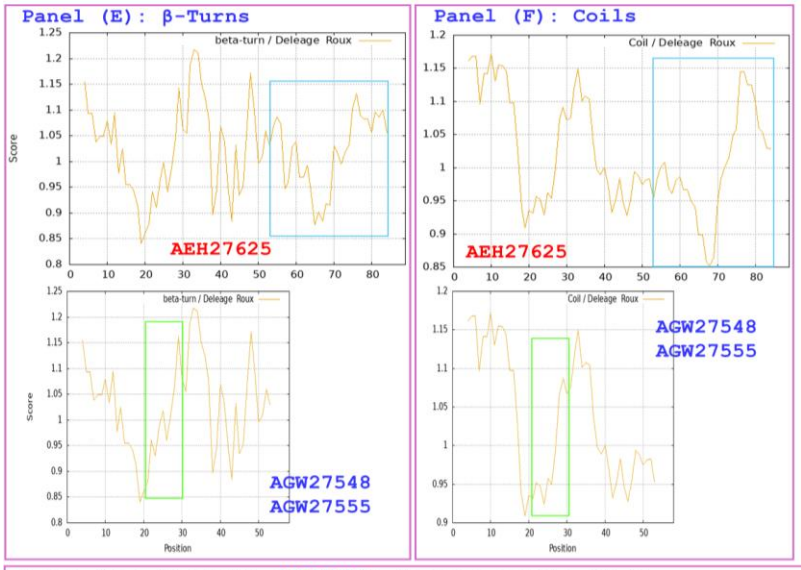
Panel (A): Alignment of amino acid sequences the HMG-box_ROX1-like domains

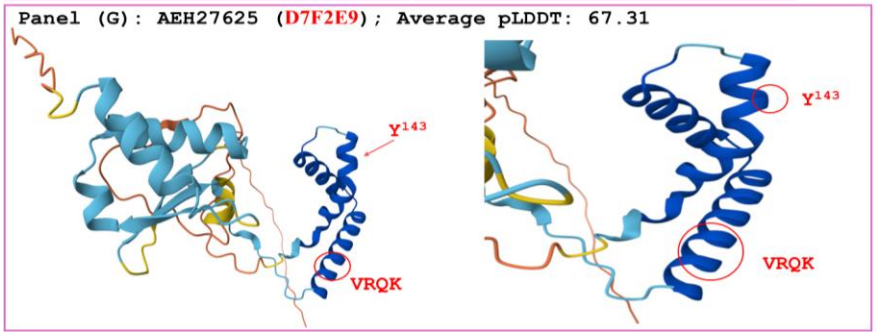
```
AEH27625 118 APPSPKDIKIPRPPNAYILYRKERHHYVKDANPGITNNEISQILGKAWNMESNDVRQKYK  177
AGW27548 118 --------------------------H------------------------------  175
AGW27555 118 --------------------------H------------------------------  175

AEH27625 178 DMSQQVKQALLEKHPDYQYKPRRPCERRR                               206
AGW27548 175
AGW27555 175
```

621







622

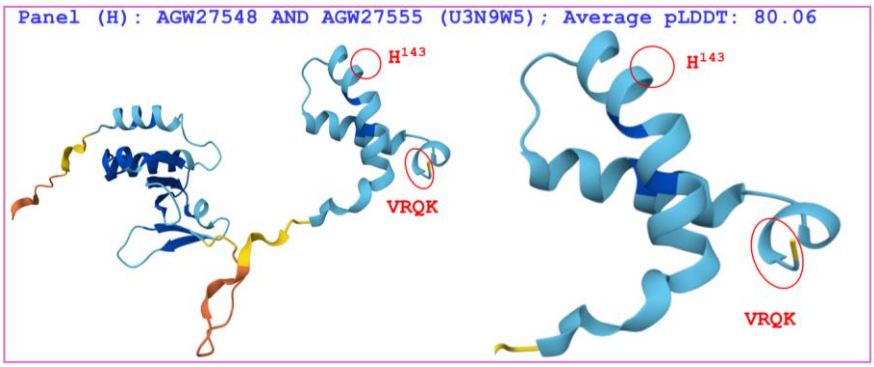


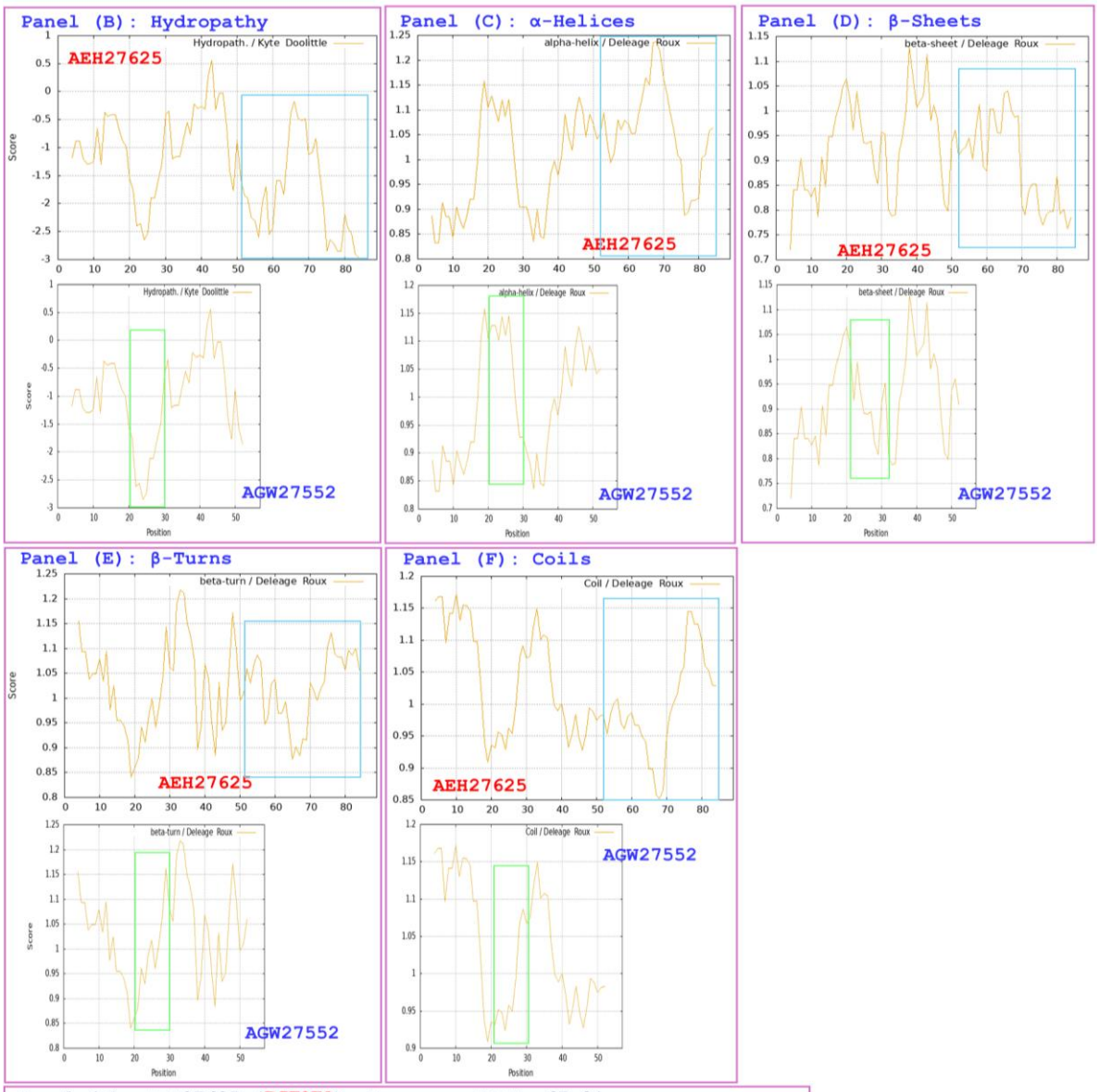
Figure 15. Correlations of the changes in hydrophobicity and the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins. The reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2, and the variant MAT1-2-1 proteins AGW27548 and AGW27555 (under the AlphaFold code U3N9W5) is 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, whereas the hyphens indicate identical amino acid residues, 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-terminal truncation regions in the ExPASy plots and in red highlight the changes in topological structure and waveform seen in the plots. Panels (G–H) show representations of the 3D structures of the entire 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).

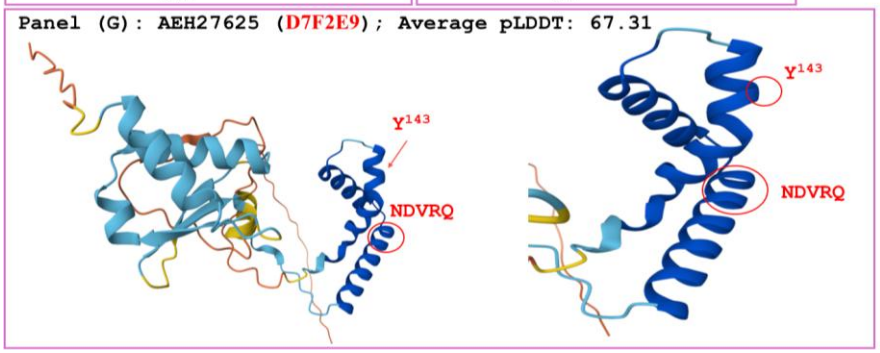
Figure 16 compares the hydrophobicity property and structures of the HMG-box_ROX1-like domain of the MAT1-2-1 protein AGW27552 (under the AlphaFold code U3N6W6) derived from the *O. sinensis* strain CS68-5-1216 [Bushley *et al*. 2013] with those of the reference protein AEH27625 (under the AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 [Zhang *et al*. 2009]. Panel (A) of Figure 16 shows a 23-residue truncation at the C-terminus of HMG-box_ROX1-like domain of the MAT1-2-1 protein AGW27552 and an amino acid substitution: Y-to-H (aromatic tyrosine to basic histidine; hydropathy indices changed from -1.3 to -3.2; *cf*. Table S3 [Kyte & Doolittle 1982]) within the HMG-box_ROX1-like domain. The replaced amino acid caused decreased hydrophobicity and altered secondary structures surrounding the mutation sites in the HMG-box_ROX1-like domain of the MAT1-2-1 protein AGW27552, as illustrated by the changes in topological structure and waveform in the ExPASy ProtScale plots for α-helices, β-sheets, β-turns, and coils, respectively. As shown in Panels (G–H) for 3D structures, the truncation and replaced residues are located in 2 of the 3 core α-helices and completely altered the stereostructure of the HMG-box_ROX1-like domain of the MAT1-2-1 protein AGW27552 under the 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
```





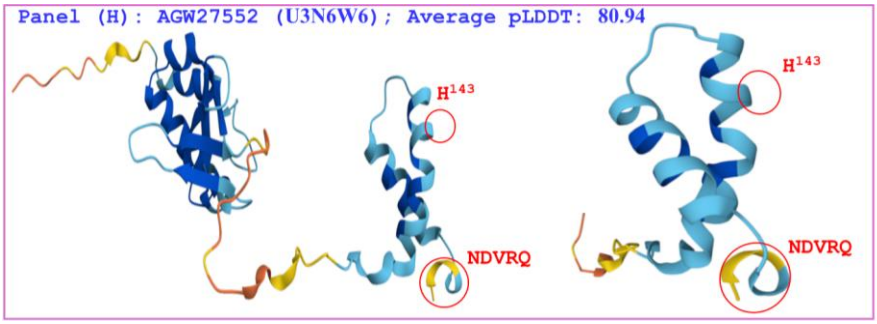


Figure 16. Correlations of the changes in hydrophobicity and the primary, secondary, and tertiary structures of the HMG-box_ROX1-like domains of MAT1-2-1 proteins. The reference protein AEH27625 (under the AlphaFold code D7F2E9) is derived from the *H. sinensis* strain CS2, and the variant MAT1-2-1 protein 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, whereas the hyphens indicate identical amino acid residues, 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-terminal truncation regions in the ExPASy plots and in red highlight the changes in topological structure and waveform seen in the plots. Panels (G–H) show representations of the 3D structures of the entire 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).

3.5 Diverse fungal sources of the MAT1-1-1 and MAT1-2-1 protein variants co-produced in pairs by the 20 *O. sinensis* strains

3.5a Genetic heterogeneity of the *O. sinensis* strains

The ITS sequences for phylogeny analysis are available in GenBank for 15 of the 20 *O. sinensis* strains but lack for the remaining 5 strains (*cf.* Table 1). Li *et al*. [2013] reported the design of PCR experiment including the use of 3 pairs of primers (GAF/GAR, GCF/GCR, and 5.8S-F/R) to amplify the ITS sequences and varying annealing temperatures at 53˚C, 55˚C, or 60℃. Percent identities and sequence coverages between the sequences of the primers and the representative 17 GC- and AT-biased genotypes of *O. sinensis* are listed in Table S6. Under this study setting, different component *O. sinensis* genotypes in the test samples might exhibit varying PCR amplification rates and the amplicons might be preferentially dominated by the most efficiently amplified fungal components. Li *et al*. [2013] also reported direct sequencing of the PCR amplicons. Apparently, this study did not apply strict amplicon cloning-sequencing techniques with a selection of sufficient 30−50 white colonies for comprehensive profiling without bias, nor use MassARRAY single nucleotide polymorphism (SNP) MALDI-TOF mass spectrum genotyping technique with use of multiple extension primers for phylogenetic analysis. Subsequently, the study would have inevitably overlooked some of cooccurring amplicon sequences,

which might be less-abundant within the amplicon pools due to relatively low-amplification efficiency under the study settings [Li *et al*. 2013, 2016b].

Li *et al*. [2013] reported that 10 of the 15 *O. sinensis* strains are homogenous, containing only GC-biased *H. sinensis*, which is Genotype #1 of *O. sinensis* (*cf*. Table 1); 5 other strains are heterogeneous, containing both GC-biased Genotype #1 and AT-biased Genotype #17 of *O. sinensis* strains (*cf*. Tables S1−S2) [Li *et al*. 2013, 2016b, 2022, 2023d, 2023e; Zhu & Li 2017]. However, because the research design adopted by Li *et al*. [2013] was not sufficiently refined, one should maintain a high level of vigilance whether additional fungal species or *O. sinensis* genotypes coexist in the purportedly homogenous and heterogeneous strains. Because of the complex genetic background of the *O. sinensis* strains, it cannot be excluded whether the MAT1-1-1 and/or MAT1-2-1 protein variants with differential truncations and various amino acid substitutions at varying variation sites originate from different coexisting fungi.

Notably, the remaining 5 *O. sinensis* strains (CS26-277, CS36-1294, CS68-5-1216, CS70-1211, and CS70-1212, shown in black in Table 1) lack the ITS sequence reports in the GenBank database with unknown homo/heterogeneity characteristics. Hopefully, the authors [Bushley *et al*. 2013; Li *et al*. 2013, 2025] who possess these strains will upload their ITS sequences to GenBank to ensure comprehensive and impartial molecular documentation of these strains.

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

The above sections present detailed tertiary structural modeling analyses of the 20 variant MAT1-1-1 and MAT1-2-1 proteins, which were assessed with various AlphaFold codes. These 20 strains simultaneously produced the paired mating proteins with differential truncations at the N- and/or C-termini and amino acid substitutions, as summarized in Tables 1−2, including the 10 "homogenous" strains shown in blown, 5 heterogeneous strains shown in red, and other 5 strains of unknown homo/heterogeneity characteristics shown in black. These strains can be divided into 2 groups on the basis of the pairing of 2 mating proteins (Table 2) [Bushley *et al*. 2013; Li *et al*. 2013]: **Group I** consists of 5 *O. sinensis* strains, which simultaneously produced both truncated MAT1-1-1 and MAT1-2-1 proteins; while **Group II** is composed of remaining 15 *O. sinensis* strains, which simultaneously produced the truncated MAT1-1-1 proteins and full-length MAT1-2-1 proteins listed in Table S5. The MAT1-1-1 and MAT1-2-1 proteins that share the same AlphaFold 3D structural morph are marked with the same suffix symbol (⁋, †, *, ⁑, ‡, or ♦) in Table 2.

716 **Table 2.** Two groups of the 20 *O. sinensis* strains on basis of differential cooccurrence of the MAT1-1-1 and
717 MAT1-2-1 protein in pairs with differential truncations and amino acid substitutions

Group	*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-terminus	C-terminus						
I	CS71-1219	AGW27534	110	8		U3NE87 *	AGW27554	65	Y-to-H, Q-to-T	U3NEA9
	CS68-5-1216	AGW27532	97	8		U3N6U8 †	AGW27552	74	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 ♦
	CS26-277	AGW27521	63	8		U3N9T9 ⁋	AGW27541			D7F2E9 ♦
	CS36-1294	AGW27518	63	8		U3N9T9 ⁋	AGW27538			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	--

718

Note: The *O. sinensis* strains in brown are "homogenous" and were detected only GC-biased Genotype #1 of *O. sinensis*; those in red are heterogeneous and were detected both GC-biased Genotype #1 and AT-biased Genotype #17 of *O. sinensis*; and those in black are unknown for their homo/heterogeneity characteristics (*cf*. Tables 1 and S1−S2) [Bushley *et al*. 2013; Li *et al*. 2013, 2022]. The GenBank protein accession numbers in pink fer 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 as producing paired truncated MAT1-1-1 and MAT1-2-1 proteins. **Group II** strains character as producing truncated MAT1-1-1 proteins that paired with the full-length MAT1-2-1 proteins. The AlphaFold codes with the same suffix symbol (⁋, †, *, ⁑, ‡, or ♦) belong to the same AlphaFold-predicted 3D structural morphs. Those without a suffix symbol indicate the unique 3D structural morphs.

In addition to the analyses of the truncated mating proteins in Sections 3.1−3.4 (*cf*. Figures 1−16), the 15 **Group-II** *O. sinensis* strains simultaneously produced the full length MAT1-2-1 proteins shown in Green in Table 2. Three (AGW27538, AGW27539, AGW27541) of the 15 MAT1-2-1 proteins are 100% identical to the authentic MAT1-2-1 proteins represented by protein AEH27625. The remaining 12 MAT1-2-1 proteins (AGW27537, AGW27540, AGW27542, AGW27544−AGW27547, AGW27549−AGW27551, AGW27553, AGW27556) contain 1−2 amino acid substitutions (Table 2, Figure S3). The substitutions are located in the HMG-box_ROX1-like domains of the MAT1-2-1 proteins, except the V-to-I substitution is located upstream of the HMG-box_ROX1-like domain of MAT1-2-1 protein AGW27537 originated from the *O. sinensis* strain CS6-251.

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

The sequence distribution of the MAT1-1-1 and MAT1-2-1 proteins simultaneously produced in pairs by each of the 20 *O. sinensis* strains are visualized in Figure 17, in which the panels are arranged according to the AlphaFold 3D structural models for the MAT1-1-1 proteins. Panel A of Figure 17 includes 8 *O. sinensis* strains, including 4 purportedly homogenous strains shown in blown in Table 2, 2 heterogeneous strains in red, and 2 strains in black with unknown homo/heterogeneity characteristics. These strains produced the 63-/8-residue truncated MAT1-1-1 proteins at the N-/C-termini, respectively, exhibiting the same tertiary structures under the AlphaFold code U3N9T9 labeled with a suffix symbol "⁋" in Table 2 and analyzed in Figure 5. Two purportedly homogenous strains CS68-2-1229 and CS91-1291 fall under **Group I**, which simultaneously produced the MAT1-2-1 proteins AGW27548 and AGW27543 that are 74- and 64-residue truncated at the C-termini (Table 2, Figures 13,15, and 17), resulting in differently altered tertiary structures under AlphaFold codes U3N9W5 and U3N9W0, respectively.

Figure 17. Alignment and sequence distributions of the paired MAT1-1-1 and MAT1-2-1 proteins derived from the *O. sinensis* strains in **Group I** in blue and **Group II** in green (*cf.* Table 2). The MAT1-1-1 protein AGW27560 (AlphaFold code U3N942) derived from the *H. sinensis* strain CS68-2-1229 was used as query MAT1-1-1 sequence and the MAT1-2-1 protein AEH27625 (AlphaFold code D7F2E9) derived from the *H. sinensis* strain CS2 was used as query MAT1-2-1 sequence in the alignment analysis.

The remaining 6 *O. sinensis* strains in Panel A fall under **Group II**, including 2 purportedly homogenous strains, 2 heterogeneous strains, and 2 other strains with unknown homo/heterogeneity characteristics. In addition to the productions of the same truncated MAT1-1-1 proteins under the AlphaFold code U3N9T9 (labeled with suffix symbol "‖" in Table 2), these 6 strains simultaneously produced the full-length MAT1-2-1 proteins with heteromorphic 3D structures under different AlphaFold codes or lacking AlphaFold 3D structure codes as follows (*cf.* Table 2 and Panel A of Figure 17):

(1) The co-produced full-length MAT1-2-1 proteins AGW27539, AGW27541, and AGW27538 exhibit

764 the same 3D structure under AlphaFold 3D structural code **D7F2E9,** with a suffix symbol "♦" in
765 Table 2. These MAT1-2-1 proteins are 100% identical to the authentic MAT1-2-1 protein AEH27625
766 (*cf*. Figure S3, Table S5). These MAT1-2-1 proteins were generated by the "homogenous" *O. sinensis*
767 strain CS37-295, and by the strains CS26-277 and CS36-1294 that lack the ITS sequence records in
768 GenBank (*cf*. Tables 1−2, Panel A of Figure 17).
769 (2) The co-produced full-length MAT1-2-1 protein AGW27537 has altered tertiary structure under
770 AlphaFold 3D structural code **U3N6V5** (*cf*. Figure 12 of [Li *et al*. 2026] and Table 5) and was
771 generated by the heterogeneous *O. sinensis* strain CS6-251 (*cf*. Tables 1−2, Panel A of Figure 17).
772 This protein shows 99.2% similarity to the authentic MAT1-2-1 protein AEH27625 and contains V-
773 to-I and Y-to-H substitutions, which are located upstream and within the HMG-box_ROX1-like
774 domain, respectively (*cf*. Figure S3 and 17, Table 2).
775 (3) The co-produced full-length MAT1-2-1 proteins AGW27540 and AGW27544 were generated by the
776 "homogenous" *O. sinensis* strain CS18-266 and the heterogeneous strain CS34-291 and do not have
777 predicted 3D structures in the AlphaFold database (*cf*. Tables 1−2, Panel A of Figure 17). These
778 MAT1-2-1 proteins contain Y-to-H substitutions within the HMG-box_ROX1-like domains, showing
779 99.6% identity to the authentic MAT1-2-1 protein AEH27625 (*cf*. Figure S3).
780 Panel B of Figure 17 and Tables 1−2 list the *O. sinensis* strains CS71-1220 and CS68-5-1216 of
781 **Group-I** and the strain CS71-1218 of **Group-II**. Strain CS71-1218 is "homogenous", Strain CS71-1220
782 is heterogenous, while Strain CS68-5-1216 is unknown about its homo/heterogeneity characteristics
783 [Bushley *et al*. 2013] (*cf*. Tables 1−2). These 3 strains produced the MAT1-1-1 proteins (AGW27532,
784 AGW27533, and AGW27535; *cf*. Figures 9 and 17) have 97-/8-residue truncations at the N-/C-termini,
785 respectively, under the same AlphaFold 3D structural model (U3N6U8 labeled with a suffix symbol "†"
786 in Table 2). The strains simultaneously generated the MAT1-2-1 proteins AGW27555 and AGW27552
787 74-residue truncated at their C-termini (*cf*. Figures 3 and 15−17) and the 1-residue truncated MAT1-2-1
788 protein AGW27553 at its C-terminus (*cf*. Figure S3) with distinct tertiary structures under AlphaFold 3D
789 structural codes U3N9W5, U3N6W6, and U3N9X0, respectively. These 3 MAT1-2-1 proteins contain
790 Y-to-H substitutions but the protein AGW27553 contains additional K-to-X substitution within the
791 HMG-box_ROX1-like domain, showing 99.6%, 99.6% and 99.2% similarities, respectively, to the
792 authentic MAT1-2-1 protein AEH27625 (*cf*. Figure S3).
793 Panel C of Figure 17 and Tables 1−2 list the "homogenous" *O. sinensis* strain CS71-1219 of **Group-**
794 **I** and the strain CS70-1212 of **Group-II** with unknown homo/heterogeneity characteristics. These 2

795 strains produced the MAT1-1-1 proteins AGW27534 and AGW27531 having 110-/8-residue truncations
796 at their N-/C-termini, respectively (*cf*. Figures 12 and 17), under the same AlphaFold 3D structural code
797 U3NE87 labeled with a suffix symbol "*" in Tables 1−2. Simultaneously, Strain CS71-1219 produced a
798 MAT1-2-1 protein AGW27554 having a 65-residue truncation at its C terminus and Y-to-H and Q-to-T
799 substitutions within the HMG-box_ROX1-like domain, resulting in the altered tertiary structure under
800 the AlphaFold 3D structural code U3NEA9 (*cf*. Figures 14 and 17). Whereas Strain CS70-1212 produced
801 the full-length MAT1-2-1 protein AGW27551 with a Y-to-H substitution within the HMG-box_ROX1-
802 like domain but without Q-to-T substitution (*cf*. Figure S3, Table 2). The MAT1-2-1 protein AGW27551
803 does not have an AlphaFold predicted 3D structure.

804 Panel D of Figure 17 and Tables 1−2 list 2 **Group-II** strain CS70-1211 and CS70-1208, the former
805 has unknown homo/heterogeneity characteristics and the latter is heterogeneous. These strains generated
806 the 97-/8-residue truncated MAT1-1-1 proteins AGW27530 and AGW27529 at their N-/C-termini and
807 have F-to-V substitutions within the MATα_HMGbox domains. These 2 MAT1-1-1 proteins are under
808 the same AlphaFold 3D model U3NE79 labeled with a suffix symbol "‡" in Tables 1−2 (*cf*. Figures 10,
809 and 17). Simultaneously, these strains co-produced the full-length MAT1-2-1 proteins AGW27550 and
810 AGW27549, which contain Y-to-H substitutions within the HMG-box_ROX1-like domains (*cf*. Tables
811 1−2; Figures 17 and S3), showing 99.6% sequence identity to the authentic MAT1-2-1 protein
812 AEH27625. The AlphaFold database does not contain AlphaFold predicted 3D structural model for these
813 2 MAT1-2-1 proteins.

814 **3.5e *O. sinensis* strains simultaneously produced the MAT1-2-1 proteins with the same tertiary**
815 **structures and the MAT1-1-1 proteins with various tertiary structures**

816 In **Group I** shown in Table 2, the "homogenous" strain CS68-2-1229 and the heterogeneous strain
817 CS71-1220 produced the 74-residue truncated MAT1-2-1 proteins AGW27548 and AGW27555 with Y-
818 to-H substitutions within the HMG-box_ROX1-like domains (*cf*. Figures 15 and 17), respectively. These
819 2 proteins share the same tertiary structures under the AlphaFold code U3N9W5 labeled with a suffix
820 symbol "⁑" in Tables 1−2. Simultaneously. these strains produced the MAT1-1-1 proteins AGW27528
821 and AGW27535, which are differentially truncated with deletions of 63/8 and 97/8 residues at the N-
822 termini of the proteins and have distinct 3D structures under AlphaFold codes U3N9T9 and U3N6U8
823 labeled with a suffix symbol "⁋"or "†" in Tables 1−2 that were fall differently under Panels (A) and (B)
824 of Figure 17, respectively (*cf*. Figures 5, 9, and 17).

825 The 5 **Group-II** *O. sinensis* strains (CS560-961, CS76-1284, CS561-964, CS25-273, and CS68-2-

826 1228; other than those fall under Panels A−D of Figure 17) generated the full-length MAT1-2-1 proteins
827 (AGW27542, AGW27545, AGW27546, AGW27547, and AGW27556; *cf*. Figures 17 and S3),
828 respectively. Although all these full-length MAT1-2-1 proteins contain the same Y-to-H substitutions
829 within the HMG-box_ROX1-like domains, only protein AGW27542 was assigned an AlphaFold 3D
830 structural code D7F2E3. Simultaneously, these strains generated the MAT1-1-1 protein variants
831 (AGW27522, AGW27525, AGW27526, AGW27527, and AGW27536; *cf*. Figures 9,17, Table 2) with
832 the differential truncations (63-/8-residue, 66-/10-residue, 63-/13-residue, 77-/18-residue, and 95-/8-
833 residue) at the N-/C-termini and the various amino acid substitutions under different AlphaFold 3D
834 structural codes (U3N6U0, U3N919, U3N7G5, U3N6U4, and U3N7H7), respectively (*cf*. Table 2,
835 Figures 5−8,11, and 17).

836 Aggregating the data presented in Section 3.5, the 20 *O. sinensis* strains simultaneously produced
837 the critical MAT1-1-1 and MAT1-2-1 proteins in pairs with differential truncations and amino acid
838 substitutions, indicating the genetic variations within the *MAT* loci in the genomes of impure *O. sinensis*
839 strains, regardless of whether the strains were previously reported as "homogenous" or heterogenous
840 based on the experimental results using less sophisticated study design by Li *et al.* [2013]. Because of
841 no repetitive copies of the *MAT1-1-1* and *MAT1-2-1* genes in the genome of *H. sinensis* [Li *et al.* 2025],
842 the various co-productions of the essential mating proteins in pairs indicate that they are not necessarily
843 the products of GC-biased Genotype #1 *H. sinensis*, instead the possible products of different
844 cooccurring heterospecific or genotypic fungi within the impure *O. sinensis* strains. Table S7 summarizes
845 differentially occurrence and co-occurrence of the MAT1-1-1 and MAT1-2-1 proteins in *O. sinensis*
846 strains, wild-type *C. sinensis* isolates, and *C. sinensis* insect–fungal complexes. Genomically, both the
847 *MAY1-1-1* and *MAT1-2-1* genes are present in the genome assemblies ANOV00000000 and
848 LKHE00000000 of *H. sinensis* strains Co18 and 1229, respectively [Hu *et al*. 2013; Li *et al*. 2016a].
849 However, the *MAY1-1-1* gene is absent in the genome assemblies LWBQ00000000 and NGJJ00000000
850 of *H. sinensis* strains ZJB12195 and CC1406-20395, respectively, and the *MAT1-2-1* gene is absent in
851 the genome assembly JAAVMX000000000 of *H. sinensis* strain IOZ07 [Jin *et al*. 2020; Liu *et al*. 2020;
852 Shu *et al*. 2020]. As shown in Figure S1 of [Li *et al*. 2026], the Y-to-M substitutions were found in the
853 MATα_HMGbox domains of MAT1-1-1 proteins encoded by the genome assemblies LKHE00000000
854 and JAAVMX000000000 of *H. sinensis* strains 1229 and IOZ07, respectively, but not by the genome
855 assembly ANOV00000000 of *H. sinensis* strain Co18 [Hu *et al*. 2013; Li *et al*. 2016a, 2026; Shu *et al*.
856 2020]. The Y-to-M substitution does not exist in the variant MAT1-1-1 proteins derived from the 20 *O.*

857 *sinensis* strains. The genomically encoded MAT1-2-1 proteins contain S-to-A substitutions within the
858 HMG-box_ROX1-like domains (*cf*. Figure S2 of [Li *et al*. 2026]), while the S-to-A substitution does not
859 occur in the MAT1-2-1 proteins derived from the **Group-II** *O. sinensis* strains listed in Table 2. It
860 indicates that the simultaneously detected MAT1-1-1 and MAT1-2-1 proteins in pairs with different
861 truncations and amino acid substitutions shown in Tables 1−2 and Figure 17 might be originated from
862 different cooccurring fungi within the impure *O. sinensis* strains.

863 # 4. Discussion

864 ## 4.1 The altered tertiary structures of the critical DNA-binding domains of mating
865 proteins produced by various *O. sinensis* strains

866 Previous studies [Li et al. 2023c, 2024b] reported the differential occurrence, differential
867 transcription, and alternative splicing of *MAT1-1-1* and *MAT1-2-1* and pheromone receptor genes in *H.*
868 *sinensis*, wild-type *C. sinensis* isolates, and *C. sinensis* insect−fungal complex. The full-length MAT1-
869 1-1 and MAT1-2-1 proteins contribute to 15 and 17 stereostructure heteromorphs, belonging to 5 and 5
870 Bayesian clusters, respectively [Li *et al*. 2025]. Whereas the DNA-binding domains of the MAT1-1-1
871 and MAT1-2-1 proteins derived from wild-type *C. sinensis* isolates clustered to 5 and 2 branched
872 Bayesian clusters, respectively (*cf*. Tables S6−S9 of [Li *et al*. 2026]). Evidence at genetic, transcriptional,
873 and protein structural levels supports that *O. sinensis* experiences self-sterility and employs a
874 heterothallic hybrid, or even parasexuality strategy to accomplish sexual reproduction during the
875 lifecycle of the *C. sinensis* insect–fungal complex [Li et al. 2023c, 2024b, 2025, 2026]. The research
876 presented in this paper extends the previous investigations by focusing on differential truncations and
877 diverse amino acid substitutions within the MATα_HMGbox domains of MAT1-1-1 proteins and HMG-
878 box_ROX1-like domains of MAT1-2-1 proteins, which are co-produced in pairs from *O. sinensis* strains.
879 The N- and/or C-terminal truncations and 1−4 amino acid substitutions within the DNA-binding domains
880 significantly influence hydrophobicity properties and modify the secondary and tertiary structures of the
881 functional domains of mating proteins—particularly the tertiary structures of hydrophobic cores formed
882 by 3 α-helices inside the DNA-binding domains. These structural variations ultimately modulate the
883 synergistic functions of the key functional domains of mating proteins during the sexual reproduction
884 process of *O. sinensis*.

885 The HMGbox domains in mating proteins are multifunctional motifs that are central to the
886 transcriptional regulation of mating-related genes, *O. sinensis* sexual reproduction, and host adaptation

887 within the *C. sinensis* insect–fungal complex [Zheng & Wang 2013; Li *et al*. 2023c, 2024b, 2025, 2026].
888 In particular, the MATα_HMGbox within the MAT1-1-1 protein represents is inferred to be of conserved
889 descent from the ancestral mating-type HMG lineage *via* conservative continuation model; whereas the
890 HMG-box_ROX1-like domain within the MAT1-2-1 protein arises from evolutionary cooption of a
891 different regulatory lineage (*e.g*., ROX1-like) through functional convergence—wherein it was
892 repurposed to control mating type specification [Martin *et al.* 2010; Czaja *et al.* 2011; Angelini *et al.*
893 2016; Poveda-Huertes *et al.* 2020; Ramšak & Kück 2021; Ramšak *et al.* 2022].

894 The MATα_HMGbox and HMG-box_ROX1-like domains of the MAT1-1-1 and MAT1-2-1
895 proteins, respectively, are structurally defined DNA-binding domains characterized by a conserved L-
896 shape conformation. This conformation is formed by a hydrophobic core containing 3 asymmetrically
897 arranged α-helices, whose stability is highly dependent on tightly packed, hydrophobic cores located at
898 the interface where these helices converge. The unique tertiary structures of these sequence-specific
899 DNA binders preferentially ensure high−affinity, sequence−dependent binding to short, AT-rich DNA
900 sequences within the promoters of their targets, mating-type-specific genes [Hansen *et al*. 2008;
901 Rajewska *et al*. 2012; Jackson *et al*. 2013; Zheng *et al*. 2013; Kim *et al*. 2015; Ramšak & Kück 2021].
902 The sequence-specific binding of the HMG-box domains to AT-rich promoter motifs serves as the initial
903 trigger, thereby playing pivotal roles in controlling fungal mating identity and gene expression
904 [Yamamoto *et al*. 1997; Martin et al. 2010; Ait Benkhali et al. 2013; Ramšak *et al.* 2022].

905 The hydrophobic core, composed of tightly packed amino acid residues, is the foundation of the
906 structures [Kyte & Doolittle 1982; Baxevanis *et al*. 1995; Thapar 2015]. It provides precise
907 conformational stability and constraints that enable the 3 α-helices to form the characteristic L-shaped
908 fold, which creates stable resistance to environmental perturbations that might otherwise unfold or distort
909 the domains. Disruption of this hydrophobic core, whether by amino acid substitutions introducing
910 incompatible residues or by peptide chain truncations removing parts of it, directly impairs structural
911 integrity and abolishes the crucial functions of the domains in DNA binding and the transcription
912 regulation of mating-type-specific genes.

913 Apparently, the differential truncations and 1–4 amino acid substitutions identified in this study at
914 various mutation sites within the functional domains of paired mating proteins in *O. sinensis* strains
915 directly disrupt key structural folding elements (helices, angles, hydrophobic core) (*cf*. Figures 5–16),
916 and fundamentally impair the DNA-binding specificity in recognizing the target DNA segments. These
917 alterations have significant consequences for the regulation of *O. sinensis* sexual reproduction during

918 the sexual life of *C. sinensis* insect−fungal complex. Relative to the authentic MAT1-1-1 and MAT1-2-
919 1 proteins of *O. sinensis* analyzed previously by Li et al. [2025, 2026], the differentially varied DNA-
920 binding domains clustered into diverse clades/branches in the Bayesian clustering trees [Li *et al*. 2025,
921 2026].

922 Although the MATα_HMGbox domain within MAT1-1-1 protein and the HMG-box_ROX1-like
923 domain within MAT1-2-1 protein both belong structurally to the HMG-box superfamily and share a
924 conserved L-shaped fold that enables DNA binding and bending, they are functionally distinct in the
925 context of mating-type regulation in *O. sinensis via* regulating different gene sets, resulting in opposite
926 regulatory outcomes. The MATα_HMGbox domain of MAT1-1-1 protein activates transcription of genes
927 encoding MAT1-1-specific receptors or other regulators. Although the HMG-box_ROX1-like domain of
928 MAT1-2-1 protein binds to specific DNA sequences in the MAT1-2 specific gene promoter encoding the
929 MAT1-2 specific pheromone receptor, it may overlap differently with the MATα_HMGbox targets and
930 act through an opposite pathway to counterbalance the activity of MAT1-1-1 protein [Balasubramanian
931 *et al*. 1993; Zitomer *et al*. 1997; Zheng *et al*. 2013]. Functionally, the MATα_HMGbox domain primarily
932 functions as an activator of its cognate mating-type genes (*i.e.,* MAT1-1 genes). In contrast, the HMG-
933 box_ROX1-like domain acts as a repressor of genes specific to the opposite mating type (MAT1-1 or "a-
934 specific" genes) and may also serve as an activator for MAT1-2-specific genes. This repressive capacity
935 represents a key functional hallmark derived from its homology to ROX1 [Kastaniotis & Zitomer 2000].
936 Thus, despite their structural similarity as DNA-binding and DNA-bending HMG-box domains, the
937 MATα_HMGbox and HMG-box_ROX1-like domains are functionally specialized to control distinct,
938 complementary sets of target genes, thereby defining the two different mating types. Their divergent
939 regulatory outputs (activation *vs*. repression of key target gene sets) constitute the molecular and protein
940 structural basis of mating-type specification during the sexual reproduction of *O. sinensis* during the
941 lifecycle of *C. sinensis* insect-fugal complex.

942 The putative heterodimerization between the MATα_HMGbox domain of MAT1-1-1 protein and
943 the HMG-box_ROX1-like domain of MAT1-2-1 protein has been proposed as a potential regulatory
944 mechanism to enforce outbreeding and initiate the sexual cycle in heterothallic fungi [Kües & Casselton
945 1993; Asante-Owusu *et al*. 1996; Jacobsen *et al*. 2002]. However, because of the lack of direct evidence
946 in *O. sinensis*, it remains a compelling hypothesis for future research to confirm the occurrence of this
947 heterodimerization, characterize the interaction interface, and define its precise regulatory targets and
948 functional impact on mating and fruiting body development in this economically and ecologically

important insect−fungal complex. Nevertheless, the MATα_HMGbox and HMG-box_ROX1-like domains may interact electrostatically and complementarily, providing a plausible physicochemical basis for cooperative synergy in DNA binding. This interaction could coordinate the transition from asexual growth to sexual development distorting target DNA to repress haploid programs and activate mating/meiosis genes of *O. sinensis* during the sexual life of *C. sinensis* insect−fungal complex [Hancock *et al*. 2019]. Despite this, direct experimental validation using protein biochemistry and reproductive physiology approaches is still required to confirm this hypothetical model, clarify its species-specific characteristics, and determine the impact of the protein-level variations on the complementary interactions and mating functional alteration of *O. sinensis* [Li *et al*. 2023c, 2024b, 2025, 2026]—even as the genome-independent, multi-fungal, and multi-genotypical nature of *O. sinensis* remains to be fully elucidated [Kinjo & Zang 2001; Stensrud et al. 2007; Xiao et al. 2009; Zhu et al. 2010; Gao *et al*. 2012; Mao *et al*. 2013; Yao & Zhu 2016; Li et al. 2016b, 2020b, 2022, 2023d, 2023e; Zhu & Li 2017].

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

4.2a Phylogenetic heterogeneity of *O. sinensis* strains

Bushley *et al*. [2013] employed a dual-fluorescence microscopy technique to characterize the multicellular heterokaryotic microstructures of *C. sinensis* ascospores and hyphae, which contain mono-, bi-, tri-, and tetra-nucleate cellular configurations. Zhang & Zhang [2015] subsequently commented on these heterokaryotic microstructures and expressed doubts regarding whether binucleate (as well as trinucleate and tetra-nucleate) cells might inconsistently harbor homogenous or heterogeneous hereditary substances. Molecular biological observations by Li *et al*. [2013, 2016b, 2023d, 2023e], which identified multiple fungal components in the ascospores and ascocarps of *C. sinensis* insect-fungal complex and in “pure” *H. sinensis* strains, provided indisputable evidence supporting the hypothesis proposed by Zhang & Zhang [2015] that the multicellular heterokaryotic cells contain heterogeneous hereditary substances.

Important culture-dependent evidence was reported by Li *et al*. [2013], who deposited the ITS sequences of 15 of the 20 *O. sinensis* strains in GenBank. Among these 15 *O. sinensis* strains, 10 were identified as "homogenous" *O. sinensis* strains, reportedly harboring only the GC-biased Genotype #1. The remaining 5 strains were characterized as heterogeneous, containing both GC-biased Genotype #1,

979 and AT-biased Genotype #17. For this heterogeneity phenomenon, Li *et al*. [2013, 2017, 2020c] proposed
980 the hypothesis that the AT-biased sequences represent "ITS pseudogenes" or "rRNA pseudogenes"
981 within the genome of GC-biased *H. sinensis*, which are generated through Repeat-Induced Point (RIP)
982 mutation in a single genome. However, subsequent studies have demonstrated that the sequences of
983 Genotypes #2−17 of *O. sinensis* do not reside in the genome of *H. sinensis* but are genomically
984 independent and belong to distinct fungal lineages. These findings directly invalidate the hypotheses that
985 the "RIP" mutation induced the AT-biased "ITS pseudogene" or "rRNA pseudogenes" in the genome of
986 GC-biased *H. sinensis* [Li *et al*. 2016a, 2022, 2023d, 2023e, 2024a; Zhu & Li 2017]. As analyzed in
987 Section 3.5a, unfortunately, the research methodology employed by Li et al. (2013) was less
988 sophisticated. This technical limitation led to an imbalance in the amplification efficiency of coexisting
989 fungal sequences, with only the sequence(s) exhibiting the highest PCR amplification efficiency being
990 captured. which precluded the comprehensive detection of all coexisting fungal taxa in the samples.
991 Notably, the different abundance of amplicons (high *vs*. low)primarily reflects the level of PCR
992 amplification efficiency and velocity under specific research settings and should not be equated with the
993 natural occurrence ratio of coexisting fungi in the tested samples. The natural occurrence ratio of
994 coexisting fungal taxa can be quantified using specialized techniques such as Southern blotting [Zhu *et*
995 *al*. 2010].
996 Following the research clue of the inconsistent co-occurrence of GC-biased Genotype #1 and AT-
997 biased Genotype #17 in the cultures of monoascospores identified in the culture-independent study
998 (Tables 1 and S1−S2) [Li *et al*. 2013], Li *et al*. [2016b, 2023d, 2023e] further adopted an exhaustive
999 culture-independent approach, including multi-primers, "touch-down" PCR protocols and selection and
1000 sequencing of multiple white colonies, as well as MALDI-TOF SNP mass spectrum genotyping
1001 approaches with use of multiple extension primers, to exhaustively investigate the comprehensive fungal
1002 composition within *C. sinensis* ascospores and the "pure" *H. sinensis* strains Hs-Ma-4, Hs7.7, Hs7.8,
1003 and Hs-Qing2 (gifts from renowned mycologist Prof. Gou YL). The authors reported that *C. sinensis*
1004 ascospores and "pure" *H. sinensis* strains harbor cooccurrence of genomically independent fungal taxa,
1005 including genome–independent, psychrophilic GC-biased Genotypes #1 and #14, AT-biased Genotypes
1006 #5–6 and #16 of *O. sinensis*, as well as the mesophilic *Paecilomyces hepiali* (recently reclassified as
1007 *Samsoneilla hepiali* [Wang *et al*. 2020]). These findings clearly indicate that the study by Li *et al*. [2013]
1008 overlooked non-culturable fungal components under the particular experimental conditions and
1009 coexisting fungal sequences with low PCR amplification efficiency. The origins of paired MAT1-1-1

1010 and MAT1-2-1 proteins with differential truncations and various amino acid substitutions at different
1011 mutation sites require further clarification *via* a rigorously designed mycological and molecular
1012 protocols by the custodian(s) of the original *O. sinensis* strains.

1013 The paired heteromorphic MAT1-1-1 and MAT1-2-1 proteins are assumed to have distinct fungal
1014 origins, derived from the cooccurring GC-biased Genotypes #1 and #3 and AT-biased Genotype #17 of
1015 *O. sinensis* in wild-type *C. sinensis* isolates [Li *et al*. 2026] and the *O. sinensis* strains analyzed in this
1016 study. Given the limited available information on the diverse coexisting genotypes of *O. sinensis*, the
1017 different pairings of the MAT1-1-1 and MAT1-2-1 protein variants that exhibit differential truncations
1018 and diverse amino acid substitutions at distinct mutation sites and were co-produced by the differentially
1019 co-existing genomically independent fungi in *O. sinensis* strains [Li 1988; Dai *et al*. 1989, 2008; Liu *et*
1020 *al*. 1989; Kinjo & Zang 2001; Jiang & Yao 2003; Chen *et al*. 2004, 2011; Stensrud *et al*. 2005, 2007;
1021 Leung *et al*. 2006; Zhu *et al*. 2007, 2010; Yang *et al*. 2008, 2021; Xiao *et al*. 2009; Zhang *et al*. 2010a,
1022 2010b, 2018; Barseghyan *et al*. 2011; Gao *et al*. 2012; Li *et al*. 2013, 2016b, 2020b, 2022, 2023d, 2023e;
1023 Ren *et al*. 2013; Meng *et al*. 2015; Xia *et al*. 2015, 2016; Guo *et al*. 2017; Zhu & Li 2017; Zhong *et al*.
1024 2018; Kang *et al*. 2024].

1025 **4.2b Differential pairings of the MAT1-1-1 and MAT1-2-1 proteins simultaneously produced by**
1026 ***O. sinensis* strains**

1027 As summarized in Table S7, both authentic and variant MAT1-1-1 and MAT1-2-1 proteins display
1028 differential occurrence or co-occurrence in the various *O. sinensis* strains (including GC-biased
1029 Genotype #1 *H. sinensis* strains), wild-type *C. sinensis* isolates, and *C. sinensis* insect–fungal complexes.
1030 The 20 *O. sinensis* strains analyzed in this study are categorized into two distinct groups based on co-
1031 productions of their paired mating proteins: Group I produces the truncated MAT1-1-1 and truncated
1032 MAT1-2-1 proteins in pairs, while Group II harbors truncated MAT1-1-1 proteins paired with full-length
1033 MAT1-2-1 proteins (*cf*. Tables 2 and S5, Figures 5−17 and S3) [Bushley *et al*. 2013; Li *et al*. 2013, 2025].
1034 Most full-length or truncated MAT1-1-1 and MAT1-2-1 proteins contain diverse amino acid substitutions
1035 (*cf*. Figures 1−16 and S3).

1036 When addressing the issue of overlooked, low-abundance co-existing fungal sequences that likely
1037 arise from inefficient amplification and sequencing bias associated with suboptimal experimental design,
1038 the "homogenous" and heterogeneous cultures of *C. sinensis* monoascospores previously reported by Li
1039 et al. [2013] presumably contain a greater diversity of uncharacterized fungal sequences. The actual
1040 fungal cooccurrence within the tested strains may be far more complex than what reported by Li *et al*.

1041 [2013]. As analyzed in Sections 3.5 and 4.2, the distinct pairings of the highly divergent MAT1-1-1 and
1042 MAT1-2-1 protein variants that were simultaneously and differentially produced by the 20 *O. sinensis*
1043 strains present a complex, non-uniform pairing pattern, indicating divergent fungal origins from
1044 cooccurring fungal taxa in the samples. Further investigations are therefore warranted, employing
1045 unbiased and rigorous culture-independent methodologies to comprehensively characterize coexisting
1046 genomically independent genotypic *O. sinensis* and heterospecific fungal taxa [Li *et al*. 2016b, 2023d,
1047 2023e].

1048 **4.2c Paradoxical dilemmas regarding highly possible heterogeneity of the prior reported "pure"**
1049 ***O. sinensis* strains**

1050 The first dilemma concerns the purportedly homogenous strain CS68-2-1229 of **Group I** (dark
1051 green and underlined in Figures 2 and 4). Only the GC-biased Genotype #1 was detected in this "pure"
1052 culture of *Cordyceps sinensis* monoascospores [Bushley *et al*. 2013; Li *et al*. 2013], although the study
1053 design adopted by Li *et al*. [2013] was insufficiently rigorous to comprehensively detect all cooccurring
1054 fungal taxa (*cf*. Section 3.5a). Two MAT1-1-1 protein sequences from this strain have been deposited in
1055 GenBank: a full-length MAT1-1-1 protein AGW27560 containing 372 amino acids and an N-/C-
1056 terminally truncated MAT1-1-1 protein AGW27528 containing 301 amino acids (*cf*. Figure 5). These 2
1057 MAT1-1-1 proteins share 100% sequence identity, yet AGW27528 exhibits only 81% query coverage
1058 with 63- and 8-residue truncations at the N- and C-termini, respectively, (*cf*. Figures 17 and S3). This
1059 truncation pattern was likely attributable to the use of different primers for amplifying the *MAT1-1-1*
1060 gene or its transcript from the total RNA pool in the experiments. This strain also produced the truncated
1061 MAT1-2-1 protein AGW27548 (*cf*. Figure 15) containing 175 amino acids, which shares 99.4% sequence
1062 identity with the authentic full-length MAT1-2-1 protein and carries a Y-to-H substitution within the
1063 HMG-box_ROX1-like domain; notably, it shows only 70% sequence coverage due to a 74-residue
1064 truncation at the C-terminus relative to the authentic full-length MAT1-2-1 proteins containing 249
1065 amino acids. These 3 mating protein sequences derived from Strain CS68-2-1229 were all released by
1066 GenBank on the same date (28-SEP-2013). The truncated MAT1-1-1 AGW27528 and MAT1-2-1 protein
1067 AGW27548 were derived from cDNAs that were amplified from the same batch of total RNA pool
1068 extracted from the strain CS68-2-1229 using a primer pair *m1cDNAF3*/*m1cDNAR5* to amplify *MAT1-1-*
1069 *1* cDNA and a primer pair *Mat1-2F*/*Mat1-2R* to amplify *MAT1-2-1* cDNA (*cf*. Figures S1−S2) [Bushley
1070 *et al*. 2013; Li *et al*. 2013].

1071 As shown in Figures 2 and 5, the MATα_HMGbox domain (162 amino acids) of the MAT1-1-1

1072 protein AGW27528 is among the longest domains across all 9 truncated MAT1-1-1 proteins. In contrast,
1073 the HMG-box_ROX1-like domain (49 amino acids) of the MAT1-2-1 protein AGW27548 (*cf*. Figure 15)
1074 is 10 residues shorter than the longest truncated HMG-box_ROX1-like domains (59 amino acids) of the
1075 MAT1-2-1 protein AGW27543 (*cf*. Figure 4), and 22 residues shorter than the 71-residue HMG-
1076 box_ROX1-like domain of the authentic MAT1-2-1 protein AEH27625 [Li *et al*. 2026]. The truncated
1077 MAT1-2-1 protein AGW27548, which harbors the Y-to-H substitution, exhibits altered secondary and
1078 tertiary structures (*cf*. Figure 15). Li *et al*. [2025] reported no repetitive copies of the *MAT1-1-1* and
1079 *MAT1-2-1* genes in the genome of *H. sinensis* strain. These findings suggest that the MAT1-1-1 protein
1080 AGW27528 is most likely encoded by the authentic *MAT1-1-1* gene in the *H. sinensis* genome. However,
1081 the MAT1-2-1 protein AGW27548 is presumably encoded by a variated *MAT1-2-1* gene within the *H.*
1082 *sinensis* genome or by an independent cooccurring fungal taxon in the actually "impure" *O. sinensis*
1083 strain CS68-2-1229. Thus, the true genetic identity of strain CS68-2-1229 requires re-evaluation: it
1084 remains to be determined whether this strain is genuinely homogenous, containing only Genotype #1 *H.*
1085 *sinensis*, or it is indeed heterogeneous, harboring 2 or more fungal taxa like the other heterogeneous
1086 strains, as shown in Tables 1–2 and S1–S2.

1087 Another paradoxical purportedly homogenous *O. sinensis* strain CS561-964 (**Group II**) was
1088 reported to harbor only GC-biased Genotype #1 *H. sinensis* [Bushley *et al*. 2013; Li *et al*. 2013]. In
1089 contrast to the "homogenous" strain CS68-2-1229, Strain CS561-964 produced the MAT1-1-1 protein
1090 AGW27526, which features 63-/13-residue truncated at the N-/C-termini alongside 4 amino acid
1091 substitutions (CDRA-to-SSEF) within the MATα_HMGbox domains (*cf*. Figures 7 and 17, Tables 1−2).
1092 The markedly altered MAT1-1-1 protein AGW27526 is clearly encoded by a variated *MAT1-1-1* gene,
1093 either within the single genome of *H. sinensis* or, more plausibly, from a genome-independent,
1094 cooccurring fungal taxon. Such a fungal taxon is likely overlooked in the study by Li *et al*. [2013] due
1095 to the use of suboptimal experimental protocols (*cf*. Section 3.5a),

1096 Analyses of the purportedly homogenous *O. sinensis* strains raise fundamental questions regarding
1097 fungal purification and comprehensive detection of all cooccurring fungal taxa in such cultures. Two key
1098 questions remain unresolved: first, whether the purportedly homogenous strains listed in Tables 1−2 and
1099 S1−S2 truly harbor only GC-biased Genotype #1, or are indeed heterogeneous and contai multiple fungi?
1100 Second, whether the characterized heterogeneous strains only contain the GC-biased Genotype #1 and
1101 AT-biased Genotype #17, or an even greater diversity of undetected coexisting fungi? It is reasonable to
1102 hypothesize that the differentially truncated MAT1-1-1 and MAT1-2-1 protein variants, which form

1103 paired combinations and exhibit diverse amino acid substitutions at distinct mutation sites, are not
1104 necessarily from GC-biased Genotype #1 *H. sinensis*. Instead, these protein variants likely originate from
1105 AT-biased Genotype #17 (*cf*. Tables S1−S2), or from other uncharacterized cooccurring fungal taxa that
1106 were not detected and reported in the study by Li *et al*. [2013].

1107 4.3 The self-sterility behavior of *O. sinensis* and requirement of sexual partners to 1108 accomplish heterothallic or hybrid reproduction

1109 Wei *et al*. [2006] proposed that *H. sinensis* (GC-biased Genotype #1 of *O. sinensis*) represents the
1110 sole anamorph of *O. sinensis* teleomorph. Zhang *et al*. [2013] proposed implementation of "One
1111 Fungus=One Name" nomenclature principle established by the International Mycological Association
1112 [Hawksworth *et al*. 2011; Taylor 2011], yet inappropriately replaced the anamorphic name *H. sinensis*
1113 with the teleomorphic name *O. sinensis*. This designation was applied uniformly to all 17 genomically
1114 independent genotypes, irrespective of their genetic GC- and AT-biased characteristics. These 17
1115 genomically independent genotypes share a common hereditary ancestor and naturally co-exist in
1116 divergent combinational patterns. To support the sole anamorph hypothesis, Bushley *et al*. [2013] and
1117 Hu *et al*. [2013] proposed the self-fertilization hypothesis for *O. sinensis* to accomplish homothallic or
1118 pseudohomothallic reproduction. However, multiple research groups [Holliday & Cleaver 2008; Stone
1119 2010; Hu *et al*. 2013] have reported unsuccessful attempts to artificially cultivate the fruiting bodies and
1120 ascospores of *C. sinensis* insect−fungal complex using pure fungal cultures. Zhang *et al*. [2013]
1121 summarized the 40-year history of failed attempts, both published and unpublished , to cultivate the
1122 insect−fungal complex in academic research-oriented settings. Qin *et al*. [2018] further elaborated on
1123 the key obstacles hindering the artificial cultivation of *C. sinensis* fruiting bodies and ascospores.
1124 Evidently, the contentious sole anamorph hypothesis and the concomitant self-fertilization hypothesis
1125 have not yet been scientifically validated in academia to meet all 4 criteria of Koch's postulations.

1126 Contradictory to the hypothesis that *H. sinensis* represents the sole anamorph of *O. sinensis* [Wei *et*
1127 *al*. 2006], molecular studies have revealed considerable genetic complexity of *O. sinensis*. This Latin
1128 name *O. sinensis* encompasses at least 17 distinct genotypes [Stensrud *et al*. 2007; Xiao *et al*. 2009; Zhu
1129 *et al*. 2010; Gao *et al*. 2012; Li *et al*. 2022]. Although these *O. sinensis* genotypes may share a common
1130 evolutionary ancestor [Stensrud *et al*. 2007], the sequences of Genotypes #2−17 of *O. sinensis* do not
1131 reside in the genome of Genotype #1 *H. sinensis*, indicating that they represent divergent, modern fungal
1132 species with independent genomes [Xiao *et al*. 2009; Zhu and Li 2017; Li *et al*. 2022]. A medicinal
1133 mycology research group led by Prof. S.P. Wasser reported that *H. sinensis* and *Tolypocladium sinensis*

1134 serve as co-anamorphs of *O. sinensis* [Barseghyan *et al*. 2011], supporting the prior detections of *T.*
1135 *sinensis* from natural *C. sinensis* insect-fungal complexes [Li 1988; Jiang & Yao 2003; Chen *et al*. 2004;
1136 Leung *et al*. 2006]. Quandt *et al*. [2014] subsequently proposed taxonomic revisions within the genus
1137 *Tolypocladium*, subsequently proposed taxonomic revisions within the genus *Tolypocladium*, reassigning
1138 several species to the family Ophiocordycipitaceae. To date, the taxonomic positions of 17 evolutionarily
1139 related genotypes as independent fungi, as well as the precise anamorph–teleomorph relationship of *O.*
1140 *sinensis*, remains incompletely resolved. Regarding reproductive modes, early hypotheses proposed
1141 homothallic or pseudohomothallic reproduction for *O. sinensis*. However, Zhang and Zhang [2015]
1142 reported the differential presence of *MAT1-1-1* and *MAT1-2-1* genes across numerous wild-type *C.*
1143 *sinensis* isolates and advanced a facultative hybridization hypothesis for *O. sinensis*. Li *et al*. [2023c,
1144 2024b] further documented differential occurrence, alternative splicing, and differential translation of
1145 the *MAT1-1-1* and *MAT1-2-1* genes and the pheromone receptor genes. More recently, Li *et al*. [2025,
1146 2026] and the present study revealed heteromorphic tertiary structures of the MAT1-1-1 and MAT1-2-1
1147 protein variants, in particular variations within the DNA-binding domains of the mating proteins. These
1148 structural observations provide direct protein-level evidence for heterogenous fungal origins of mating-
1149 type loci in numerous wild-type *C. sinensis* isolates and *O. sinensis* strains. Notably, Li *et al*. [2025]
1150 reported that both *MAT1-1-1* and *MAT1-2-1* genes exist as single-copy genes in the *H. sinensis* genome
1151 with no repetitive genomic copies. Collectively, evidence from genetics, transcription, and protein
1152 structure invalidates the self-fertilization hypothesis for *O. sinensis* and instead support self-sterility,
1153 implying that *O. sinensis* relies on heterothallic mating, hybridization, or even parasexual outcrossing to
1154 accomplish sexual reproduction [Bennett & Johnson 2003; Sherwood & Bennett 2009; Seervai *et al*.
1155 2013; Nakamura *et al*. 2019; Samarasinghe *et al*. 2020; Mishra *et al*. 2021; Steensels *et al*. 2021; Li *et*
1156 *al*. 2023c, 2023e, 2024b, 2025, 2026]. Similar self-sterile reproductive strategies have been documented
1157 in many other *Ophiocordyceps* species, which require compatible mating partners to undergo sexual
1158 reproduction [Yokoyama *et al*. 2005; Lee *et al*. 2010; Zheng *et al*. 2013; Araújo *et al*. 2018; Xu *et al*.
1159 2023a; Zou *et al*. 2024].

1160 Wei *et al*. [2016] reported successful industrial cultivation of *C. sinensis* insect−fungal complexes.
1161 A puzzling species paradox was observed between the input and output of the cultivated insect–fungal
1162 complexes: 3 GC-biased *H. sinensis* strains were used as inoculants, yet only the teleomorph of AT-
1163 biased Genotype #4 was detected in the fruiting bodies of the cultivated *C. sinensis* insect–fungal
1164 complexes. The sole teleomorphic AT-biased Genotype #4 identified in the cultivated fruiting bodies by

1165 Wei et al. [2016] had apparently replaced the inoculated GC-biased Genotype #1 strains, resulting in a
1166 fundamental species contradiction. The sequences of Genotypes #2–17 of *O. sinensis* are not present in
1167 the genome assemblies ANOV00000000, JAAVMX000000000, LKHE00000000, LWBQ00000000, or
1168 NGJJ00000000 of the pure *H. sinensis* strains Co18, IOZ07, 1229, ZJB12195, and CC1406-20395,
1169 respectively [Stensrud *et al*. 2007; Xiao *et al*. 2009; Gao *et al*. 2012; Hu *et al*. 2013; Li *et al*. 2013,
1170 2016a, 2020b, 2022, 2023d, 2023e; 2024a; Mao *et al*. 2013; Zhu & Li 2017; Jin *et al*. 2020; Liu *et al*.
1171 2020; Shu *et al*. 2020]. Instead, these genotypes belong to distinct fungi with independent genomes.
1172 These results invalidate the illusory assumptions proposed by Li et al. [2013, 2017, 2020c] that the so
1173 called "ITS pseudogene" or "rRNA pseudogenes" within the genome of *H. sinensis* were generated
1174 through "RIP" mutation. Furthermore, genomic or transcriptomic analysis provided no supporting
1175 evidence for the unfounded speculations [Li *et al*. 2020b. 2024a]. In contrast to the significantly variable
1176 maturation-related dynamics of AT-biased genotypes of *O. sinensis*, GC-biased *O. sinensis* genotypes
1177 maintain a relatively constant low abundance in the stromata throughout the entire maturation course of
1178 *C. sinensis* insect−fungal complex [Zhu *et al*. 2010; Gao *et al*. 2012]. These findings indicate that the 17
1179 genome−independent genotypes of *O. sinensis*, regardless of being GC-biased or AT-biased, represent
1180 distinct fungal species [Xiao *et al*. 2009; Li *et al*. 2022, 2023d, 2023e], further reinforcing the
1181 invalidation of the RIP mutation/pseudogenes hypotheses proposed by Li *et al*. [2013, 2017, 2020c]. The
1182 successful industrial cultivation of *C. sinensis* insect–fungal complexes, coupled with the apparent
1183 species paradox, indicates that the sole anamorph hypothesis for *H. sinensis* proposed by Wei *et al*.
1184 [2006] failed to satisfy all 4 criteria of Koch's postulates.

1185 Wei *et al*. [2016] identified AT-biased Genotype #4 of *O. sinensis* as the sole teleomorph in the
1186 fruiting body of cultivated insect−fungal complexes. Early studies reported that AT-biased Genotype #4
1187 occurred in the stromata throughout all maturation stages of *C. sinensis* insect−fungal complex, including
1188 the stromal fertile portion (SFP). However, its abundance exhibited a marked ontogenetic shift: while
1189 Genotype #4 was highly abundant in immature stromata during the asexual growth phase, its levels
1190 declined substantially in mature stromata and SFP during the sexual reproduction stage or the transition
1191 from asexual growth to sexual reproduction [Zhu *et al*. 2010; Gao *et al*. 2012; Li *et al*. 2020b, 2022,
1192 2023d, 2023e]. Li *et al*. [2020a] subsequently clarified that the inoculant used in the industrial
1193 inoculation-cultivation project, which is the cultures of the *C. sinensis* ascospores, rather than pure *H.*
1194 *sinensis* strains initially described by Wei *et al*. [2016].

1195 Regrettably, Li *et al*. [2013] failed to detect AT-biased Genotype #4 from cultures of *C. sinensis*

1196 monoascospores using culture–dependent methods but instead identified GC-biased Genotype #1 and
1197 AT-biased Genotype #5 (actually Genotype #17; *cf*. Table S2). Used culture–independent approaches, Li
1198 *et al*. [2016b, 2023d, 2023e] employed a comprehensive PCR amplicon cloning-sequencing strategy
1199 (with the selection of ≥30 white colonies) and SNP MALDI-TOF mass spectrum genotyping using
1200 multiple extension primers. These analyses identified GC-biased Genotypes #1 and #14 and AT-biased
1201 Genotypes #5–6 and #16, and *S. hepiali* in *C. sinensis* ascospores, yet again failed to detect AT-biased
1202 Genotype #4. Collectively, these findings confirm the absence of Genotype #4 of *O. sinensis* in the *C.*
1203 *sinensis* ascospores [Li *et al*. 2013, 2023d, 2023e]. Consequently, the fungal origin of the sole
1204 teleomorphic AT-biased Genotype #4 of *O. sinensis* detected in the fruiting body of cultivated *C. sinensis*
1205 insect−fungal complex [Wei *et al*. 2016] remains an unresolved mycological enigma.

1206 Li *et al*. [2026] summarized the evidence related to the industrial success of cultivating the
1207 insect−fungal complex and postulated that the technical design for the cultivation project adopts a
1208 "fungal impurity" strategy, with the synergistic effect of coexisting fungi potentially being the key to
1209 success. The supporting evidence includes: (1) Li *et al*. [2020a] reported that the cultivation inoculant
1210 was cultures of *C. sinensis* ascospores, which harbor multiple genotypic and heterospecific fungi [Li *et*
1211 *al*. 2013, 2023d, 2023e]; (2) the *C. sinensis* ascospores do not contain AT-biased Genotype #4 of *O.*
1212 *sinensis* [Li *et al*. 2013, 2020b, 2022, 2023d, 2023e]. It is therefore plausible that the origin of the
1213 teleomorphic Genotype #4 that was detected by Wei *et al*. [2016] originated from co-cultures of *C.*
1214 *sinensis* ascospores with stroma and/or SFP containing numerous ascocarps, which indeed contain AT-
1215 biased Genotype #4 of *O. sinensis*, along with other genotypic and heterospecific fungi; (3) the industrial
1216 cultivation system was supplemented with soil collected from the natural *C. sinensis* production areas
1217 on the Qinghai–Tibet Plateau [Wei *et al*. 2016].

1218 To date, the purification and genomic sequencing of GC- and AT-biased Genotypes #2–17 of *O.*
1219 *sinensis* have not been reported. Consequently, genomic and transcriptomic information regarding the
1220 significantly divergent *MAT1-1-1* and *MAT1-2-1* genes in the genome−independent Genotypes #2–17 of
1221 *O. sinensis*, especially the critical DNA-binding domains of the MAT1-1-1 and MAT1-2-1 proteins,
1222 remains lacking. Irrespective of the precise biophysical details of their interaction, whether the paired
1223 MAT1-1-1 and MAT1-2-1 proteins form heterodimers or interact *via* complementary electrostatic
1224 surfaces, the synergy of paired mating proteins serves as the central and indispensable molecular trigger
1225 for mating compatibility and the initiation of sexual reproduction in *O. sinensis*, [Debuchy & Turgeo
1226 2006; Jones & Bennett 2011; Zheng & Wang 2013; Wilson *et al*. 2015; Sun *et al*. 2019]. Molecular

1227 evidence regarding differential occurrence, alternative splicing, and differential transcription of mating-
1228 type and pheromone receptor genes in *H. sinensis* unequivocally indicates that *O. sinensis* is self-sterile,
1229 relying on heterothallism or hybrid outcrossing to complete sexual reproduction, thereby effectively
1230 ruling out homothallism (self-fertilization) as a viable mechanism in this species. [Hu *et al*. 2013;
1231 Bushley *et al*. 2013; Zhou *et al*. 2013; Wei *et al*. 2006; Li *et al*. 2019, 2023c, 2024b, 2025]. Furthermore,
1232 the discovery of heteromorphic tertiary structures of variant MATα_HMGbox domains (in MAT1-1-1
1233 proteins) and HMG-box_ROX1-like domains (in MAT1-2-1 proteins) provides the protein level
1234 validation that (1) self-fertilization is biochemically unfeasible and (2) coordinated heterothallic or
1235 hybrid reproduction is enforced through structural complementarity between the paired mating proteins,
1236 which are likely generated by divergent heterogeneous *O. sinensis* strains. This self-sterility hypothesis
1237 resolves longstanding uncertainty in *O. sinensis* cultivation—namely, the failure of single-species
1238 cultures to fruit—and underscores the role of multi-fungal symbiotic synergism in its ecological success
1239 throughout the lifecycle of the naturally occurring or artificially cultivated *C. sinensis* insect−fungal
1240 complex.

1241 # Conclusions

1242 This study presents the protein structural-level evidence for self-sterility in *O. sinensis* and further
1243 verifies the heterothallic or hybrid mating pattern of this fungus. For the 20 *O. sinensis* strains examined
1244 in this study, the diverse pairings of MAT1-1-1 and MAT1-2-1 protein variants exhibit differential
1245 truncations and 1−4 amino acid substitutions in their DNA-binding domains. The primary structural
1246 variations caused dramatic changes in the hydrophobicity property of the mating proteins and altered the
1247 secondary and tertiary structures of the MATα_HMGbox and HMG-box_ROX1-like domains of the
1248 differentially paired MAT1-1-1 and MAT1-2-1 proteins, respectively. The structural heterogeneity of
1249 differentially paired mating proteins not only indicates the co-occurrence of heterogeneous fungal origins
1250 in impure *O. sinensis* strains but also invalidates the hypothesis of homothallic or pseudohomothallic
1251 self-fertilization in *O. sinensis*. Complementing the evidence of differential occurrence, differential
1252 transcription, and alternative splicing of the *MAT1-1-1* and *MAT1-2-1* genes and pheromone receptor
1253 genes, the protein structural variations of these DNA-binding domains impair the homologous
1254 interaction required for homothallic or pseudohomothallic elf-fertilization. Collectively, our findings of
1255 differential pairings of the mating protein variants in *O. sinensis* strains clarify the sexual reproduction
1256 mechanism of *O. sinensis* and confirm that the selection of suitable heterologous mating partners is an

essential natural strategy for self-sterile *O. sinensis* to complete sexual reproduction during the lifecycle of *Cordyceps sinensis* insect-fungal complexes on the Qinghai–Tibet Plateau.

Acknowledgments

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 [Bushley *et al*. 2013]. 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 m1cDNAF3/*MAT1-1-1* (AGCGAAGCGACCTCTCAAT) and the reverse complementary primer m1cDNAR5/*MAT1-1-1* (GGAGGACTCGGAGGAAG) reported by [Bushley *et al*. 2013]. 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 MAT1-2-1 protein AEH27625 (1→249) by derived from *Ophiocordyceps sinensis* strain CS2 [Zhang *et al*. 2009]. 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 by [Bushley *et al*. 2013]. 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 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 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 [Li *et al*. 2025].

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 [Li *et al*. 2025].

Table S6. Similarity of the 3 pairs of primers that were used by Li *et al*. [2013] 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 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. Cooccurrence 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 that are listed in the GenBank database.

Author Contributions: Conceptualization, XZL, YLL, and JSZ; methodology, JSZ; formal analysis, JSZ; investigation, XZL and JSZ; data curation, XZL and JSZ; writing–original draft preparation, JSZ; writing—review and editing, XZL, YLL, and JSZ; project administration, YLL; funding acquisition, YLL. All 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) Wild *Ophiocordyceps sinensis* Identification and Application Project (QHRD-2025-004); (3) 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. Ait Benkhali, J.; Coppin, E.; Brun, S.; Peraza-Reyes, L.; Martin, T.; Dixelius, C.; Lazar, N.; van Tilbeurgh, H.; Debuchy, R. A Network of HMG-box Transcription Factors Regulates Sexual Cycle in the Fungus Podospora anserina. *PLoS Genet.* **2013**, *9*, e1003642. https://doi.org/10.1371/journal.pgen.1003642.
2. Angelini, R.M.D.M.; Rotolo, C.; Pollastro, S.; Faretra, F. Molecular analysis of the mating type (*MAT1*) locus in strains of the heterothallic ascomycete *Botrytis cinerea*. *Plant Pathol*. **2016**, *65*, 1221−1400. https://doi.org/10.1111/ppa.12509
3. Araújo JPM, Evans HC, Kepler R, Hughes DP. 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
4. Asante-Owusu, R.N.; Banham, A.H.; Böhnert, H.U.; Mellor, E.J.C.; Casselton, L.A. Heterodimerization between two classes of homeodomain proteins in the mushroom *Coprinus cinereus* brings together potential DNA-binding and activation domains. *Gene* **1996**, *172*, 25–31. https://doi.org/10.1016/0378-1119(96)00177-1.
5. Balasubramanian, B.; Lowry, C.V.; Zitomer, R.S. The Rox1 repressor of the *Saccharomyces cerevisiae* hypoxic genes is a specific DNA-binding protein with a high-mobility-group motif. *Mol. Cell Biol.* **1993**, *13*, 6071–6078. https://doi.org/10.1128/mcb.13.10.6071-6078.1993.
6. Barseghyan, G.S.; Holliday, J.C.; Price, T.C.; Madison, L.M.; Wasser, S.P. Growth and cultural-morphological characteristics of vegetative mycelia of medicinal caterpillar fungus *Ophiocordyceps sinensis* G.H. Sung et al. (Ascomycetes) Isolates from Tibetan Plateau (P. R. China). *Intl. J. Med. Mushrooms* **2011**, *13*, 565–581. https://doi.org/10.1615/intjmedmushr.v13.i6.90.
7. Baxevanis, A.D.; Bryant, S.H.; Landsman, D. Homology model building of the HMG-1 box structural domain. *Nucleic Acids Res.* **1995**, *23*, 1019–1029. https://doi.org/10.1093/nar/23.6.1019.
8. 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.

1335 9. 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
1336 idiomorph and evidence for selfing in the Chinese medicinal fungus *Ophiocordyceps sinensis*. *Fungal Biol.* **2013**, *117*, 599–
1337 610. https://doi.org/10.1016/j.funbio.2013.06.001.
1338 10. Chen, C.-S.; Hseu, R.-S.; Huang, C.-T. Quality Control of *Cordyceps sinensis* Teleomorph, Anamorph, and Its Products,
1339 Chapter 12. In *Quality Control of Herbal Medicines and Related Areas*; Shoyama, Y., Ed.; InTech: Rijeka, Croatia, 2011; pp.
1340 223–238. Available online: www.intechopen.com (accessed on 3 May 2024).
1341 11. Chen, Y.-Q.; Hu, B.; Xu, F.; Zhang, W.; Zhou, H.; Qu, L.-H. Genetic variation of *Cordyceps sinensis*, a fruit-body-producing
1342 entomopathogenic species from different geographical regions in China. *FEMS Microbiol. Lett.* **2004**, *230*, 153–158.
1343 https://doi.org/10.1016/S0378-1097(03)00889-9.
1344 12. China Ministry of Agriculture and Rural Affairs. Announcement (No. 15 of 2021) of National Forestry and Grassland
1345 Administration: List of National Key Protected Wild Plants. 7 September 2021. Available online:
1346 https://m.163.com/dy/article/HHCVOJPU055360T7.html (accessed on 3 May 2025).
1347 13. Czaja, W.; Miller, K.Y.; Miller, B.L. Complex mechanisms regulate developmental expression of the matA (HMG) mating
1348 type gene in homothallic Aspergillus nidulans. *Genetics*. **2011**, *189*, 795–808. doi: 10.1534/genetics.111.131458
1349 14. 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
1350 & Dai. *Acta Agricult. Univ. Pekin.* **1989**, *15*, 221–224.
1351 15. 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*
1352 *hepiali*. *Mycosystema* **2008**, *27*, 641–644.
1353 16. David, A.; Islam, S.; Tankhilevich, E.; Sternberg, M.J.E. The AlphaFold Database of Protein Structures, A Biologist's Guide.
1354 *J. Mol. Biol.* **2022**, *434*, 167336. https://doi.org/10.1016/j.jmb.2021.167336.
1355 17. Debuchy, R.; Turgeo, B.G. Mating-Type Structure, Evolution, and Function in Euascomycetes. In *Growth, Differentiation*
1356 *and Sexuality*; Kües, U., Fischer, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 293–323.
1357 18. Deleage, G.; Roux, B. An algorithm for protein secondary structure prediction based on class prediction. *Protein Eng. Des.*
1358 *Sel.* **1987**, *1*, 289–294. https://doi.org/10.1093/protein/1.4.289.
1359 19. Du, X.-H.; Wu, D.-M.; Kang, H.; Wang, H.-C.; Xu, N.; Li, T.-T.; Chen, K.-L. Heterothallism and potential hybridization
1360 events inferred for twenty-two yellow morel species. *IMA Fungus* **2020**, *11*, 4. https://doi.org/10.1186/s43008-020-0027-1.
1361 20. Gao L, Li X-H, Zhao J-Q, Lu J-H, Zhao J-G, Zhu J-S. Maturation of *Cordyceps sinensis* associates with alterations of fungal
1362 expressions of multiple *Ophiocordyceps sinensis* mutants with transition and transversion point mutations in stroma of
1363 *Cordyceps sinensis*. *Beijing Da Xue Xue Bao*. **2012**, *44*, 454–463. doi: 10.3969/j.issn.1671-167X.2012.03.025
1364 21. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein Identification and
1365 Analysis Tools on the ExPASy Server, Chapter 52. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Humana Press:
1366 Totowa, NJ, USA, 2005; pp. 571–607.
1367 22. Guo, M.-Y.; Liu, Y.; Gao, Y.-H.; Jin, T.; Zhang, H.-B.; Zhou, X.-W. Identification and bioactive potential of endogenetic
1368 fungi isolated from medicinal caterpillar fungus *Ophiocordyceps sinensis* from Tibetan Plateau. *Int. J. Agric. Biol.* **2017**, *19*,
1369 307–313. https://doi.org/10.17957/IJAB/15.0281.
1370 23. Hancock, S.P.; Cascio, D.; Johnson, R.C. Cooperative DNA binding by proteins through DNA shape complementarity.
1371 *Nucleic Acids Res.* **2019**, *47*, 8874–8887. https://doi.org/10.1093/nar/gkz642.
1372 24. Hansen FT, Madsen CK, Nordland AM, Grasser M, Merkle T, Grasser KD. A novel family of plant DNA-binding proteins
1373 containing both HMG-box and AT-rich interaction domains. *Biochemistry*. **2008**, *47*, 13207–13214. doi: 10.1021/bi801772k
1374 25. Hawksworth, D.L.; Crous, P.W.; Redhead, S.A.; Reynolds, D.R.; Samson, R.A.; Seifert, K.A.; Taylor, J.W.; Wingfield, M.J.;
1375 Abaci, Ö.; Aime, C.; et al. The Amsterdam declaration on fungal nomenclature. *IMA Fungus* **2011**, *2*, 105–112.
1376 https://doi.org/10.5598/imafungus.2011.02.01.14.
1377 26. Hénault, M.; Marsit, S.; Charron, G.; Landry, C.R. The effect of hybridization on transposable element accumulation in an
1378 undomesticated fungal species. *eLife* **2020**, *9*, e60474. https://doi.org/10.7554/eLife.60474.

27. 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.
28. 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 uncovers the secrets of sex and lifestyle in caterpillar fungus. *Chin. Sci. Bull.* **2013**, *58*, 2846–2854. https://doi.org/10.1007/s11434-013-5929-5.
29. Jackson, D.; Lawson, T.; Villafane, R.; Gary, L. Modeling the structure of yeast MATα1, An HMG-Box motif with a C-terminal helical extension. *Open J. Biophys.* **2013**, *3*, 1–12. https://doi.org/10.4236/ojbiphy.2013.31001.
30. Jacobsen, S.; Wittig, M.; Pöggeler, S. Interaction Between Mating-Type Proteins from the Homothallic Fungus Sordaria macrospora. *Curr. Genet.* **2002**, *41*, 150–158. https://doi.org/10.1007/s00294-002-0276-0.
31. Jiang, Y.; Yao, Y.-J. A review for the debating studies on the anamorph of *Cordyceps sinensis*. *Mycosistema* **2003**, *22*, 161–176.
32. Jin, L.-.Q.; Xu, Z.-.W.; Zhang, B.; Yi, M.; Weng, C.-.Y.; Lin, S.; Wu, H.; Qin, X.-T.; Xu, F.; Teng, Y.; Yuan, S.-J.; Liu, Z.-Q.; Zheng, Y.-G. Genome sequencing and analysis of fungus *Hirsutella sinensis* isolated from *Ophiocordyceps sinensis*. *AMB Expr.* **2020**, *10*, 105. https://doi.org/10.1186/s13568-020-01039-x.
33. Jones, S.K.; Bennett, R.J. Fungal mating pheromones: Choreographing the dating game. *Fungal Genet. Biol.* **2011**, *48*, 668–676. https://doi.org/10.1016/j.fgb.2011.04.001.
34. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.: Kohl, S,A,A,; Ballard, A.J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein; S.; Silver, D.; Vinyals, O.; Senior, A.W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
35. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. *Nature*. **2021**, *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2
36. Kang, Q.; Zhang, J.; Chen, F.; Dong, C.; Qin, Q.; Li, X.; Wang, H.; Zhang, H.; Meng, Q. Unveiling mycoviral diversity in *Ophiocordyceps sinensis* through transcriptome analyses. *Front. Microbiol.* **2024**, *15*, 1493365. https://doi.org/10.3389/fmicb.2024.1493365.
37. Kastaniotis, A.J.; Zitomer, R.S. Oxygen Dependent Repression in Yeast. In *Rox1 Mediated Repression*; Advances in Experimental Medicine and Biology; Springer Nature: Cham, Switzerland, 2000; Volume 475, pp. 185–195. https://doi.org/10.1007/0-306-46825-5_18.
38. Kim, H.-K.; Jo, S.-M.; Kim, G.-Y.; Kim, D.-W.; Kim, Y.-K.; Yun, S.-H. A large-scale functional analysis of putative target genes of mating-type loci provides insight into the regulation of sexual development of the cereal pathogen *Fusarium graminearum*. *PLoS Genet.* **2015**, *11*, e1005486. https://doi.org/10.1371/journal.pgen.1005486.
39. 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.
40. 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.
41. Kües, U.; Casselton, L.A. The origin of multiple mating types in mushrooms. *J. Cell Sci.* **1993**, *104*, 227–230. https://doi.org/10.1242/jcs.104.2.227.
42. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132. https://doi.org/10.1016/0022-2836(82)90515-0.
43. 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*, 298–340. doi: 10.1128/MMBR.00005-10

1424 44. Leung, P.-H.; Zhang, Q.-X.; Wu, J.-Y. Mycelium cultivation, chemical composition and antitumour activity of a
1425 *Tolypocladium* sp. fungus isolated from wild *Cordyceps sinensis*. *J. Appl. Microbiol.* **2006**, *101*, 275–283.
1426 https://doi.org/10.1111/j.1365-2672.2006.02930.x.
1427 45. Li, C.-L. A study of *Tolypocladium sinense* C.L. Li. sp. nov. and cyclosporin production. *Acta Mycol. Sin.* **1988**, *7*, 93–98.
1428 46. Li, M.-M.; Zhang, J.-H.; Qin, Q.-L.; Zhang, H.; Li, X.; Wang, H.-T.; Meng, Q. Transcriptome and Metabolome Analyses of
1429 Thitarodes xiaojinensis in Response to *Ophiocordyceps sinensis* Infection. *Microorganisms* **2023a**, *11*, 2361.
1430 https://doi.org/10.3390/microorganisms11092361.
1431 47. 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
1432 the ghost-moth larvae after ex situ infection by *Ophiocordyceps sinensis*. *Sci. China Life Sci.* **2020a**, *63*, 1093–1095.
1433 https://doi.org/10.1007/s11427-020-1686-7. Erratum in Sci. China Life Sci. **2023b**, 66, 892. https://doi.org/10.1007/s11427-
1434 022-2232-5.
1435 48. 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
1436 of Chinese cordyceps reveals gene regulatory network and expression profiles of sexual development-related genes. *BMC*
1437 *Genom.* **2019**, *20*, 337. https://doi.org/10.1186/s12864-019-5708-z.
1438 49. Li, X.-Z.; Li, Y.-L.; Wang, Y.-N.; Zhu, J.-S. Translations of mutant repetitive genomic sequences in *Hirsutella sinensis* and
1439 changes in secondary structures and functional specifications of the encoded proteins. *Int. J. Mol. Sci.* **2024a**, *25*, 11178.
1440 https://doi.org/10.3390/ijms252011178.
1441 50. 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
1442 hypothesis" for *Ophiocordyceps sinensis*. *Mol. Phylogenet. Evol*. **2020b**, *146*, 106728.
1443 https://doi.org/10.1016/j.ympev.2019.106728
1444 51. Li, X.-Z.; Li, Y.-L.; Zhu, J.-S. Differential transcription of mating-type genes during sexual reproduction of natural
1445 *Cordyceps sinensis*. *Chin. J. Chin. Mater. Medica* **2023c**, *48*, 2829–2840. https://doi.org/10.19540/j.cnki.cjcmm.20230213.102.
1446 52. Li, X.-Z.; Li, Y.-L.; Liu, W.; Zhu, J.-S. Altered stereostructures of the DNA-binding domains of variant mating proteins of
1447 *Ophiocordyceps sinensis* and the wild insect–fungal complex. *Biol. (Basel).* **2026**, *15*, 186.
1448 https://doi.org/10.3390/biology15020186
1449 53. Li, X.-Z.; Li, Y.-L.; Zhu, J.-S. Three-dimensional structural heteromorphs of mating-type proteins in *Hirsutella sinensis* and
1450 the natural *Cordyceps sinensis* insect–fungal complex. *J. Fungi.* **2025**, *11*, 244. https://doi.org/10.3390/jof11040244.
1451 54. Li, X.-Z.; Xiao, M.-J.; Li, Y.-L.; Gao, L.; Zhu, J.-S. Mutations and differential transcription of mating-type and pheromone
1452 receptor genes in *Hirsutella sinensis* and the natural *Cordyceps sinensis* insect–fungi complex. *Biology* **2024b**, *13*, 632.
1453 https://doi.org/10.3390/biology13080632.
1454 55. 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
1455 sequencing and assembly strategies for a repeat-rich fungal genome, *Ophiocordyceps sinensis*. *J. Microbiol. Methods* **2016a**,
1456 *128*, 1–6. https://doi.org/10.1016/j.mimet.2016.06.025.
1457 56. 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
1458 mutated ITS genes in populations of *Ophiocordyceps sinensis* and their implications for molecular systematics. *IMA Fungus*
1459 **2020c**, *11*, 18.
1460 57. Li, Y.; Jiao, L.; Yao, Y.-J. Non-concerted ITS evolution in fungi, as revealed from the important medicinal fungus
1461 *Ophiocordyceps sinensis*. *Mol. Phylogenet. Evol.* **2013**, *68*, 373–379. https://doi.org/10.1016/j.ympev.2013.04.010.
1462 58. Li, Y.; Yang, R.-H.; Jiang, L.; Hu, X.-D.; Wu, Z.-J.; Yao, Y.-J. rRNA Pseudogenes in Filamentous Ascomycetes as Revealed
1463 by Genome Data. *G3-Genes Genom. Genet.* **2017**, *7*, 2695–2703. https://doi.org/10.1534/g3.117.044016.
1464 59. 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
1465 of *Ophiocordyceps sinensis* in the stromal fertile portions and ascospores of natural *Cordyceps sinensis*. *PLoS ONE.* **2023d**, *18*,
1466 e0286865. https://doi.org/10.1371/journal.pone.0286865.

1467 60. 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
1468 cooccurrence of multiple genotypes of *Ophiocordyceps sinensis* in the stromata, stromal fertile portion (ascocarps) and
1469 ascospores of natural *Cordyceps sinensis*. *PLoS ONE* **2023e**, *18*, e0270776. https://doi.org/10.1371/journal.pone.0270776.
1470 61. Li, Y.-L.; Li, X.-Z.; Yao, Y.-S.; Xie, W.-D.; Zhu, J.-S. Molecular identification of *Ophiocordyceps sinensis* genotypes and the
1471 indiscriminate use of the Latin name for the multiple genotypes and the natural insect–fungi complex. *Am. J. BioMed. Sci.*
1472 **2022**, *14*, 115–135. https://doi.org/10.5099/aj220300115.
1473 62. 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
1474 from the intestines of *Hepialus lagii* larvae in increasing infection potency. *J. Fungal Res.* **2016b**, *14*, 96–112.
1475 63. 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
1476 analyses reveal evolutionary and geologic context for the plateau fungus *Ophiocordyceps sinensis*. *Clin. Med.* **2020**, *15*, 107–
1477 119. https://doi.org/10.1186/s13020-020-00365-3.
1478 64. Liu, X.-J.; Guo, Y.-L.; Yu, Y.-X.; Zeng, W. Isolation and identification of the anamorph of *Cordyceps sinensis* fungus. *Acta*
1479 *Mycol. Sin.* **1989**, *8*, 35–40.
1480 65. Liu, Z.-Q.; Lin, S.; Baker, P.J.; Wu, L.-F.; Wang, X.-R.; Wu, H.; Xu, F.; Wang, H.-Y.; Brathwaite, M.E.; Zheng, Y.-G.
1481 Transcriptome sequencing and analysis of the entomopathogenic fungus *Hirsutella sinensis* isolated from *Ophiocordyceps*
1482 *sinensis*. *BMC Genom.* **2015**, *16*, 106–123. https://doi.org/10.1186/s12864-015-1269-y.
1483 66. Lu, D. Western records and studies of the Chinese caterpillar fungus to the beginning of the 20th century. *J. Fungal. Res.*
1484 **2014**, *12*, 233–244. https://doi.org/10.13341/j.jfr.2014.0043.
1485 67. Lu, H.-L.; St. Leger, R.J. Chapter Seven—Insect Immunity to Entomopathogenic Fungi. In *Advances in Genetics*; Lovett, B.,
1486 St. Leger, R.J., Eds.; Academic Press: Cambridge, MA, USA, 2016; Volume 94, pp. 251–285.
1487 68. Mao, X.-M.; Zhao, S.-M.; Cao, L.; Yan, X.; Han, R.-C. The morphology observation of *Ophiocordyceps sinensis* from different
1488 origins. *J. Environ. Entomol.* **2013**, *35*, 343–353.
1489 69. Mariani, V.; Biasini, M.; Barbato, A.; Schwede, T. lDDT, a local superposition-free score for comparing protein structures
1490 and models using distance difference tests. *Bioinformatics* **2013**, *29*, 2722–2728.
1491 https://doi.org/10.1093/bioinformatics/btt473.
1492 70. Martin, T.; Lu, S.-W.; van Tilbeurgh, H.; Ripoll, D.R.; Dixelius, C.; Dixelius, C.; Turgeon, B.G.; Debuchy, R. Tracing the
1493 Origin of the Fungal a1 Domain Places Its Ancestor in the HMG-Box Superfamily, Implication for Fungal Mating-Type
1494 Evolution. *PLoS ONE* **2010**, *5*, e15199. https://doi.org/10.1371/journal.pone.0015199.
1495 71. 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.
1496 Transcriptomic insight into the immune defenses in the ghost moth, Hepialus xiaojinensis, during an *Ophiocordyceps*
1497 *sinensis* fungal infection. Insect Biochem. *Mol. Biol.* **2015**, *64*, 1–15. https://doi.org/10.1016/j.ibmb.2015.06.014.
1498 72. Metin, B.; Findley, K.; Heitman, J. The mating type locus (MAT) and sexual reproduction of *Cryptococcus heveanensis*:
1499 Insights into the evolution of sex and sex-determining chromosomal regions in fungi. *PLoS Genet.* **2010**, *6*, e1000961.
1500 https://doi.org/10.1371/journal.pgen.1000961.
1501 73. Mishra, A.; Forche, A.; Anderson, M.Z. Parasexuality of Candida Species. *Front. Cell. Infect. Microbiol.* **2021**, *11*, 796929.
1502 https://doi.org/10.3389/fcimb.2021.796929.
1503 74. Monzon, V.; Haft, D.H.; Bateman, A. Folding the unfoldable, using AlphaFold to explore spurious proteins. *Bioinform.*
1504 *Adv.* **2022**, *1*, vbab043. https://doi.org/10.1093/bioadv/vbab043.
1505 75. Nakamura, N.; Tanaka, C.; Takeuchi-Kaneko, Y. Transmission of antibiotic-resistance markers by hyphal fusion suggests
1506 partial presence of parasexuality in the root endophytic fungus Glutinomyces brunneus. *Mycol. Progress.* **2019**, *18*, 453–
1507 462. https://doi.org/10.1007/s11557-018-1455-9.
1508 76. Peters, C.; Elofsson, A. Why is the biological hydrophobicity scale more accurate than earlier experimental hydrophobicity
1509 scales? *Proteins* **2014**, *82*, 2190–2198. https://doi.org/10.1002/prot.24582.
1510 77. Poveda-Huertes D, Matic S, Marada A, Habernig L, Licheva M, Myketin L, Gilsbach R, Tosal-Castano S, Papinski D,
1511 Mulica P, Kretz O, Kücükköse C, Taskin AA, Hein L, Kraft C, Büttner S, Meisinger C, Vögtle FN. An Early mtUPR:
1512 Redistribution of the Nuclear Transcription Factor Rox1 to Mitochondria Protects against Intramitochondrial Proteotoxic

1513 Aggregates. *Mol Cell*. **2020**, *77*, 180–188. doi: 10.1016/j.molcel.2019.09.026
1514 78. Qin, Q.-L.; Zhou, G.-L.; Zhang, H.; Meng, Q.; Zhang, J.-H.; Wang, H.-T.; Miao, L.; Li, X. Obstacles and approaches in
1515 artificial cultivation of Chinese cordyceps. *Mycology* **2018**, *9*, 7–9. https://doi.org/10.1080/21501203.2018.1442132.
1516 79. 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,
1517 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.
1518 Phylogenetic-based nomenclatural proposals for Ophiocordycipitaceae (Hypocreales) with new combinations in
1519 Tolypocladium. *IMA Fungus*. **5**, *1*, 121–134. doi:10.5598/imafungus.2014.05.01.12.
1520 80. Rajewska M, Wegrzyn K, Konieczny I. AT-rich region and repeated sequences - the essential elements of replication
1521 origins of bacterial replicons. *FEMS Microbiol Rev*. **2012**, *36*, 408–434. doi: 10.1111/j.1574-6976.2011.00300.x
1522 81. Ramšak, B.; Kück, U. The *Penicillium chrysogenum* tom1 gene a major target of transcription factor MAT1-1-1 encodes a
1523 nuclear protein involved in sporulation. *Front. Fungal Bio.* **2022**, *3*, 937023. https://doi.org/10.3389/ffunb.2022.937023.
1524 82. Ramšak, B.; Kűck, U.; Hofmann, E. The mating type transcription factor MAT1-1-1 from the fungal human pathogen
1525 *Aspergillus fumigatus*: Synthesis, purification, and crystallization of the DNA binding domain. *bioRxiv* **2021**.
1526 https://doi.org/10.1101/2021.12.13.472399.
1527 83. Ramšak, B.; Markau, J.; Pazen, T.; Dahlmann, T.A.; Krappmann, S.; Kűck, U. The master regulator MAT1-1-1 of fungal
1528 mating binds to its targets via a conserved motif in the human pathogen *Aspergillus fumigatus*. *G3 Genes Genom. Genet.*
1529 **2020**, *11*, jkaa012. https://doi.org/10.1093/g3journal/jkaa012.
1530 84. Ren, Y.; Wan, D.-G.; Lu, X.-M.; Guo, J.-L. The study of scientific name discussion for TCM Cordyceps. *LisShenzhen Med.*
1531 *Mater. Medica Res.* **2013**, *24*, 2211–2212.
1532 85. Rettie, S.A.; Campbell, K.V.; Bera, A.K.; Kang, A.; Kozlov, S.; De La Cruz, J.; Adebomi, V.; Zhou, G.; DiMaio, F.;
1533 Ovchinnikov, S.; et al. Cyclic peptide structure prediction and design using AlphaFold. *bioRxiv* **2023**, 26:2023.02.25.529956.
1534 https://doi.org/10.1101/2023.02.25.529956.
1535 86. Samarasinghe, H.; You, M.; Jenkinson, T.S.; Xu, J.-P.; James, T.Y. Hybridization Facilitates Adaptive Evolution in Two
1536 Major Fungal Pathogens. *Genes* **2020**, *11*, 101. https://doi.org/10.3390/genes11010101.
1537 87. Seervai, R.N.H.; Jones, S.K.; Hirakawa, M.P.; Porman, A.M.; Bennett, R.J. Parasexuality and ploidy change in *Candida*
1538 *tropicalis*. *Eukaryot. Cell.* **2013**, *12*, 1629–1640. https://doi.org/10.1128/EC.00128-13.
1539 88. Sherwood, R.K.; Bennett, R.J. Fungal meiosis and parasexual reproduction--lessons from pathogenic yeast. *Curr. Opin.*
1540 *Microbiol.* **2009**, *12*, 599–607. https://doi.org/10.1016/j.mib.2009.09.005.
1541 89. 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
1542 high-quality draft genome assembly of the Chinese cordyceps *Ophiocordyceps sinensis*. *Genome Biol. Evol.* **2020**, *12*, 1074–
1543 1079. https://doi.org/10.1093/gbe/evaa112.
1544 90. Simm, S.; Einloft, J.; Mirus, O.; Schleiff, E. 50 years of amino acid hydrophobicity scales, revisiting the capacity for peptide
1545 classification. *Biol. Res.* **2016**, *49*, 31. https://doi.org/10.1186/s40659-016-0092-5.
1546 91. Song, L.-R.; Hong, X.; Ding, X.-L.; Zang, Z.-Y. A comprehensive dictionary of modern pharmacy of Chinese traditional
1547 medicine. People's Medical Press, Beijing. 2001; pp 733–737.
1548 92. Steensels, J.; Gallone, B.; Verstrepen, K.J. Interspecific hybridization as a driver of fungal evolution and Adaptation. *Nat.*
1549 *Rev. Microbiol.* **2021**, *19*, 485–500.
1550 93. Stensrud, Ø.; Hywel-Jones, N.L.; Schumacher, T. Towards a phylogenetic classification of Cordyceps: ITS nrDNA
1551 sequence data confirm divergent lineages and paraphyly. *Mycol. Res.* **2005**, *109*, 41–56.
1552 https://doi.org/10.1017/s095375620400139x.
1553 94. Stensrud, Ø.; Schumacher, T.; Shalchian-Tabrizi, K.; Svegardenib, I.B.; Kauserud, H. Accelerated nrDNA evolution and
1554 profound AT bias in the medicinal fungus *Cordyceps sinensis*. *Mycol. Res.* **2007**, *111*, 409–415.
1555 https://doi.org/10.1016/j.mycres.2007.01.015.
1556 95. Stone, R. Improbable partners aim to bring biotechnology to a Himalayan kingdom. *Science* **2010**, *327*, 940–941.
1557 https://doi.org/10.1126/science.327.5968.940.

1558 96. Sun, S.; Coelho, M.A.; David-Palma, M.; Priest, S.J.; Heitman, J. The evolution of sexual reproduction and the mating-type
1559 locus: Links to pathogenesis of *Cryptococcus* human pathogenic fungi. *Annu. Rev. Genet.* **2019**, *53*, 417–444.
1560 https://doi.org/10.1146/annurev-genet-120116-024755.
1561 97. Sung, G.-H.; Hywel-Jones, N.L.; Sung, J.-M.; Luangsa-ard, J.J.; Shrestha, B.; Spatafora, J.W. Phylogenetic classification of
1562 *Cordyceps* and the clavicipitaceous fungi. *Stud. Mycol.* **2007**, *57*, 5–59. https://doi.org/10.3114/sim.2007.57.01.
1563 98. Taylor JW. One Fungus = One Name: DNA and fungal nomenclature twenty years after PCR. *IMA Fungus*. **2011**, *2*, 113–
1564 120. doi:10.5598/imafungus.2011.02.02.01
1565 99. Thapar, R. Structure-specific nucleic acid recognition by L-motifs and their diverse roles in expression and regulation of
1566 the genome. *Biochim. Biophys. Acta* **2015**, *1849*, 677–687. https://doi.org/10.1016/j.bbagrm.2015.02.006.
1567 100. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Žídek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon,
1568 A.; Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA, Potapenko
1569 A, Ballard AJ, Romera-Paredes B, Nikolov S, Jain R, Clancy E, Reiman D, Petersen S, Senior AW, Kavukcuoglu K, Birney
1570 E, Kohli P, Jumper J, Hassabis D. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*,
1571 590–596. https://doi.org/10.1038/s41586-021-03828-1.
1572 101. Turgeon, B.G.; Yoder, O.C. Proposed nomenclature for mating type genes of filamentous ascomycetes. *Fungal Genet. Biol.*
1573 **2000**, *31*, 1–5. https://doi.org/10.1006/fgbi.2000.1227.
1574 102. Varadi, M.; Bertoni, D.; Magana, P.; Paramval, U.; Pidruchna, I.; Radhakrishnan, M.; Tsenkov, M.; Nair, S.; Mirdita, M.;
1575 Yeo, J.; et al. AlphaFold Protein Structure Database in 2024, providing structure coverage for over 214 million protein
1576 sequences. *Nucleic Acids Res.* **2024**, *52*, D368–D375. https://doi.org/10.1093/nar/gkad1011.
1577 103. Wang, Y.; Stata, M.; Wang, W.; Stajich, J.E.; White, M.M.; Moncalvo, J.M. Comparative genomics reveals the core gene
1578 toolbox for the fungus-insect symbiosis. *mBio* **2018**, *9*, e00636-e18. https://doi.org/10.1128/mBio.00636-18.
1579 104. 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,
1580 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.,
1581 Yang, D.-R., Sanjuan, T., Liu, X.-Z., Yang, Z.-L., Yu, H. Multigene phylogeny of the family Cordycipitaceae (Hypocreales):
1582 new taxa and the new systematic position of the Chinese cordycipitoid fungus *Paecilomyces hepiali*. *Fungal Diversity* **2020**,
1583 *103*, 1–46. https://doi.org/10.1007/s13225-020-00457-3.
1584 105. Wei, J.-C.; Wei, X.-L.; Zheng, W.-F.; Guo, W.; Liu, R.-D. Species identification and component detection of *Ophiocordyceps*
1585 *sinensis* cultivated by modern industry. *Mycosystema* **2016**, *35*, 404–410.
1586 106. Wei, X.-L.; Yin, X.-C.; Guo, Y.-L.; Shen, N.-Y.; Wei, J.-C. Analyses of molecular systematics on *Cordyceps sinensis* and its
1587 related taxa. *Mycosystema* **2006**, *25*, 192–202.
1588 107. Wilson, A.M.; Wilken, P.M.; van der Nest, M.A.; Steenkamp, E.T.; Wingfield, M.J.; Wingfield, B.D. Homothallism: An
1589 umbrella term for describing diverse sexual behaviours. *IMA Fungus* **2015**, *6*, 207–214.
1590 https://doi.org/10.5598/imafungus.2015.06.01.13.
1591 108. Wroblewski, K.; Kmiecik, S. Integrating AlphaFold pLDDT Scores into CABS-flex for enhanced protein flexibility
1592 simulations. *Comput. Struct. Biotechnol. J.* **2024**, *30*, 4350–4356. https://doi.org/10.1016/j.csbj.2024.11.047.
1593 109. Xia, E.-H.; Yang, D.-R.; Jiang, J.-J.; Zhang, Q.-J.; Liu, Y.; Liu, Y.-L.; Zhang, Y.; Zhang, H.-B.; Shi, C.; Tong, Y.; Kim C-H,
1594 Chen H, Peng Y-Q, Yu Y, Zhang W, Eichler EE, Gao L-Z. The caterpillar fungus, *Ophiocordyceps sinensis*, genome provides
1595 insights into highland adaptation of fungal pathogenicity. *Sci. Rep.* **2017**, *7*, 1806. https://doi.org/10.1038/s41598-017-01869-
1596 z.
1597 110. Xia, F.; Liu, Y.; Guo, L.-X.; Shen, G.-L.; Lin, J.; Zhou, X.-W. Pyrosequencing analysis revealed complex endogenetic
1598 microorganism community from natural DongChong XiaCao and its microhabitat. *BMC Microbiol*. **2016**, *16*, 196. DOI
1599 10.1186/s12866-016-0813-5
1600 111. Xia, F.; Liu, Y.; Shen, G.-L.; Guo, L.-X.; Zhou, X.-W. Investigation and analysis of microbiological communities in natural
1601 *Ophiocordyceps sinensis*. *Can. J. Microbiol.* **2015**, *61*, 104–111. https://doi.org/10.1139/cjm-2014-0610.

1602 112. Xiang, L.; Li, Y.; Zhu, Y.; Luo, H.; Li, C.; Xu, X.; Sun, C.; Song, J.-Y.; Shi, L.-H.; He, L.; Sun W, Chen S-L. Transcriptome
1603 analysis of the *Ophiocordyceps sinensis* fruiting body reveals putative genes involved in fruiting body development and
1604 cordycepin biosynthesis. *Genomics* **2014**, *103*, 154–159. https://doi.org/10.1016/j.ygeno.2014.01.002.
1605 113. Xiao, W.; Yang, J.-P.; Zhu, P.; Cheng, K.-D.; He, H.-X.; Zhu, H.-X.; Wang, Q. Non-support of species complex hypothesis
1606 of *Cordyceps sinensis* by targeted rDNA-ITS sequence analysis. *Mycosystema* **2009**, *28*, 724–730.
1607 114. Xu M, Ashley NA, Vaghefi N, Wilkinson I, Idnurm A. Isolation of strains and their genome sequencing to analyze the
1608 mating system of *Ophiocordyceps robertsii*. *PLoS One*. **2023a**, *18*, e0284978. doi: 10.1371/journal.pone.0284978.
1609 115. Xu, T.; Xu, Q.; Li, J.-Y. Toward the appropriate interpretation of Alphafold2. *Front. Artif. Intell.* **2023a**, *6*, 1149748.
1610 https://doi.org/10.3389/frai.2023.1149748.
1611 116. Yamamoto, A.; Ando, Y.; Yoshioka, K.; Saito, K.; Tanabe, T.; Shirakawa, H.; Yoshida, M. Difference in affinity for DNA
1612 between HMG proteins 1 and 2 determined by surface plasmon resonance measurements. *J. Biochem.* **1997**, *122*, 586–594.
1613 https://doi.org/10.1093/oxfordjournals.jbchem.a021793. PMID 9348088.
1614 117. Yang, J.-L.; Xiao, W.; He, H.-X.; Zhu, H.-X.; Wang, S.-F.; Cheng, K.-D.; Zhu, P. Molecular phylogenetic analysis of
1615 *Paecilomyces hepiali* and *Cordyceps sinensis*. *Acta Pharmaceut. Sinica.* **2008**, *43*, 421–426.
1616 118. Yang, J.-Y.; Tong, X.-X.; He, C.-Y.; Bai, J.; Wang, F.; Guo, J.-L. Comparison of endogenetic microbial community diversity
1617 between wild *Cordyceps sinensis*, artificial *C. sinensis* and habitat soil. *Chin. J. Chin. Mater. Medica* **2021**, *46*, 3106–3115.
1618 119. 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
1619 structural analyses of nucleosome binding proteins. *Nucleic Acids Res.* **2025**, *53*, gkaf735.
1620 https://doi.org/10.1093/nar/gkaf735.
1621 120. Yao, Y.-S.; Zhu, J.-S. Indiscriminate use of the Latin name for natural *Cordyceps sinensis* and *Ophiocordyceps sinensis* fungi.
1622 *Chin. J. Chin. Mater. Med*. **2016**, *41*, 1316–1366.
1623 121. Yokoyama E, Yamagishi K, Hara A, Heterothallism in *Cordyceps takaomontana*. *FEMS Microbiol Lett*. **2005**, *250*, 145–150.
1624 https://doi.org/10.1016/j.femsle.2005.07.004.
1625 122. Zhang, S.; Zhang, Y.-J. Molecular evolution of three protein-coding genes in the Chinese caterpillar fungus *Ophiocordyceps*
1626 *sinensis*. *Microbiol. China.* **2015**, *42*, 1549–1560.
1627 123. 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
1628 traditional Chinese medicinal fungus *Ophiocordyceps sinensis*. *Fungal Biol.* **2011**, *115*, 708–714.
1629 124. Zhang, S.; Zhang, Y.-J.; Shrestha, B.; Xu, J.-P.; Wang, C.-S.; Liu, X.-Z. *Ophiocordyceps sinensis* and *Cordyceps militaris*:
1630 Research advances, issues and perspectives. *Mycosystema* **2013**, *32*, 577–597.
1631 125. Zhang, S.-W.; Cen, K.; Liu, Y.; Zhou, X.-W.; Wang, C.-S. Metatranscriptomics analysis of the fruiting caterpillar fungus
1632 collected from the Qinghai-Tibetan plateau. *Sci. Sinica Vitae* **2018**, *48*, 562–570.
1633 126. Zhang, Y.-J.; Li, E.-W.; Wang, C.-S.; Li, Y.-L.; Liu, X.-Z. *Ophiocordyceps sinensis*, the flagship fungus of China: Terminology,
1634 life strategy and ecology. *Mycology* **2012**, *3*, 2–10. https://doi.org/10.1080/21501203.2011.654354.
1635 127. Zhang, Y.-J.; Sun, B.-D.; Zhang, S.; Wàngmǔ Liu, X.-Z.; Gong, W.-F. Mycobiotal investigation of natural *Ophiocordyceps*
1636 *sinensis* based on culture-dependent investigation. *Mycosistema* **2010a**, *29*, 518–527.
1637 128. Zhang, Y.-J.; Xu, L.-L.; Zhang, S.; Liu, X.-Z.; An, Z.-Q.; Wàngmǔ Guo, Y.-L. Genetic diversity of *Ophiocordyceps sinensis*, a
1638 medicinal fungus endemic to the Tibetan Plateau: Implications for its evolution and conservation. *BMC Evol. Biol.* **2009**, *9*,
1639 290. doi:10.1186/1471-2148-9-290
1640 129. 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.
1641 Phylogeography and evolution of a fungal–insect association on the Tibetan Plateau. *Mol. Ecol.* **2014**, *23*, 5337–5355.
1642 https://doi.org/10.1111/mec.12940.
1643 130. Zhang, Y.-J.; Zhang, S.; Wàngmǔ; Bai, F.-Y.; Liu, X.-Z. High Diversity of the Fungal Community Structure in Naturally-
1644 Occurring *Ophiocordyceps sinensis*. *PLoS ONE* **2010b**, 5(12), e15570. doi:10.1371/journal.pone.0015570.
1645 131. Zheng, P.; Xia, Y.-L.; Zhang, S.-W.; Wang, C.-S. Genetics of Cordyceps and related fungi. *Appl Microbiol Biotechnol*, **2013**,
1646 *97*, 2797–2804. doi: 10.1007/s00253-013-4771-7.

1647 132. Zheng, P.; Wang, C.-S. Sexuality Control and Sex Evolution in Fungi. *Sci. Sin. Vitae* **2013**, *43*, 1090–1097.
1648 133. 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
1649 play different roles in sexual reproduction and pathogenesis in *Fusarium graminearum*. *PLoS ONE.* **2013**, *8*, e66980.
1650 https://doi.org/10.1371/journal.pone.0066980.
1651 134. 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
1652 *Ophiocordyceps sinensis* transcriptome and differentially expressed genes in three different mycelia, sclerotium and fruiting
1653 body developmental stages. *Fungal Biol.* **2018**, *122*, 943–951. https://doi.org/10.1016/j.funbio.2018.05.011.
1654 135. Zhou, X.W.; Li, L.J.; Tian, E.W. Advances in research of the artificial cultivation of *Ophiocordyceps sinensis* in China. *Crit.*
1655 *Rev. Biotechnol.* **2013**, *34*, 233–243. https://doi.org/10.3109/07388551.2013.791245.
1656 136. 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
1657 rDNA and differential proliferations of CG:AT-biased genotypes of *Cordyceps sinensis* fungi and *Paecilomyces hepiali* in
1658 natural *C.* sinensis. *Am. J. Biomed. Sci.* **2010**, *2*, 217–238. https://doi.org/10.5099/aj100300217.
1659 137. 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.
1660 Maturation of *Cordyceps sinensis* associates with co-existence of *Hirsutella sinensis* and *Paecilomyces hepiali* DNA and
1661 dynamic changes in fungal competitive proliferation predominance and chemical profiles. *J. Fungal Res*. **2007**, 5, 214–224.
1662 138. Zhu, J.-S.; Halpern, G.M.; Jones, K. The scientific rediscovery of a precious ancient Chinese herbal regimen: *Cordyceps*
1663 *sinensis*: Part I. *J. Altern. Complem. Med.* **1998a**, *4*, 289–303. https://doi.org/10.1089/acm.1998.4.3-289.
1664 139. Zhu, J.-S.; Halpern, G.M.; Jones, K. The scientific rediscovery of an ancient Chinese herbal medicine: *Cordyceps sinensis*:
1665 Part II. *J. Altern. Complem. Med.* **1998b**, *4*, 429–457. https://doi.org/10.1089/acm.1998.4.429.
1666 140. Zhu, J.-S.; Li, C.-L.; Tan, N.-Z.; Berger, J.L.; Prolla, T.A. Combined use of whole-gene expression profiling technology and
1667 mouse lifespan test in anti-aging herbal product study. In Proceedings of the 2011 New TCM Products Innovation and
1668 Industrial Development Summit, Hangzhou, China, 27 November 2011; pp. 443–448. Available online:
1669 https://xueshu.baidu.com/usercenter/paper/show?paperid=08341c17fa58c8f85584b92572b90f75&site=xueshu_se
1670 (accessed on 30 January 2025).
1671 141. Zhu, J.-S.; Li, Y.-L. A Precious Transitional Chinese Medicine, *Cordyceps sinensis*: Multiple heterogeneous *Ophiocordyceps*
1672 *sinensis* in the insect–fungi complex. Lambert Academic Publishing, Saarbrüchen. Germany, 2017.
1673 142. Zitomer, R.S.; Limbach, M.P.; Rodriguez-Torres, A.M.; Balasubramanian, B.; Deckert, J.; Snow, P.M. Approaches to the
1674 study of Rox1 repression of the hypoxic genes in the yeast Saccharomyces cerevisiae. *Methods* **1997**, *11*, 279–288.
1675 https://doi.org/10.1006/meth.1996.0422.
1676 143. Zou, J.; Zhang, Y.-T.; Luo, Y.; Fu, M.-H.; Sun, B.-L.; Liu, S.-G. MAT1-1 and MAT1-2 *Ophiocordyceps xuefengensis* and
1677 comparison of their chemical composition. methods. *Biol. (Basel)* **2024**, *13*, 686; https://doi.org/10.3390/biology13090686.