

冬虫夏草菌 3 个细胞核蛋白编码基因的分子进化

张姝¹ 张永杰^{2*}

(1. 山西大学 应用化学研究所 山西 太原 030006)

(2. 山西大学 生命科学学院 山西 太原 030006)

摘要:【目的】分析 3 个细胞核蛋白编码基因(*csp1*、*MAT1-1-1* 和 *MAT1-2-1*)在不同冬虫夏草菌株间的分子进化。【方法】从 125 个冬虫夏草样品中分别扩增 *csp1*、*MAT1-1-1* 和 *MAT1-2-1* 基因序列,比较外显子和内含子间以及 2 个交配型基因间的序列变异程度,比较基于不同基因或基因区域所构建的系统发育树拓扑结构的差异,分析 3 个基因承受的选择压力和 DNA 重组情况。【结果】3 个蛋白编码基因外显子区的长度在不同菌株间高度保守,具有 4.5%–5.7%的变异位点;内含子区的长度在不同菌株间相同或不同,具有 1.8%–22%的变异位点。对于 2 个交配型基因,*MAT1-1-1* 的碱基变异率低于 *MAT1-2-1*。基于外显子与内含子构建的系统发育树的拓扑结构,以及基于 2 个交配型基因外显子构建的系统发育树的拓扑结构都存在明显差异。3 个蛋白编码基因都经历着净化选择作用。基因内部的不同 DNA 位点间有重组,但 3 个基因片段之间没有明显的重组发生。【结论】由于冬虫夏草菌不同基因以及基因的不同区域表现出进化上的差异,所以在开展冬虫夏草菌进化相关的研究时,应该联合使用多个不同的基因片段。

关键词: 冬虫夏草菌, 丝氨酸蛋白酶基因, 交配型基因, 分子进化, 选择压力, 重组

Molecular evolution of three protein-coding genes in the Chinese caterpillar fungus *Ophiocordyceps sinensis*

ZHANG Shu¹ ZHANG Yong-Jie^{2*}

(1. Institute of Applied Chemistry, Shanxi University, Taiyuan, Shanxi 030006, China)

(2. College of Life Sciences, Shanxi University, Taiyuan, Shanxi 030006, China)

Abstract: [Objective] To analyze the molecular evolution of three protein-coding genes (*csp1*, *MAT1-1-1* and *MAT1-2-1*) among various *Ophiocordyceps sinensis* isolates. [Methods] We amplified partial *csp1* fragment and complete *MAT1-1-1* and *MAT1-2-1* gene sequences from 125 broadly sampled *O. sinensis* isolates, and compared the topological difference of phylogenetic trees constructed by exons and introns, and of those constructed by *MAT1-1-1* and *MAT1-2-1*. We also measured selection pressures and detected evidence of recombination for the three genes. [Results] The length of exons of the three protein-coding genes was conserved among different *O. sinensis* isolates with

基金项目: 国家自然科学基金项目(No. 81102759, 31140013); 教育部高等学校博士学科点专项科研基金项目(No. 20101401120007); 山西省青年科技基金项目(No. 2014021030-2)

*通讯作者: ✉: zhangyj2008@sxu.edu.cn

收稿日期: 2015-04-03; 接受日期: 2015-05-15; 优先数字出版日期(www.cnki.net): 2015-05-26

4.5%–5.7% of variable sites. The length of introns was either identical or different among different isolates with more than 1.8%–22% of variable sites. For the two mating-type genes, *MAT1-1-1* was less variable than *MAT1-2-1*. Phylogenetic topology differences were significant between exons and introns, and between *MAT1-1-1* and *MAT1-2-1* exon sequences. Each of the three genes was undergoing purified selection. Recombination events were detected between different DNA sites within each gene, but not obvious between different gene fragments. **[Conclusion]** Because different gene fragments and different regions of a gene in *O. sinensis* might be evolutionarily different, we should use multiple gene fragments in future evolution-related studies in *O. sinensis*.

Keywords: *Ophiocordyceps sinensis*, Serine protease gene, Mating-type gene, Molecular evolution, Selection pressure, Recombination

冬虫夏草是我国传统的名贵中药材,与人参、鹿茸并称“中药三大宝”。它是由冬虫夏草菌 *Ophiocordyceps sinensis* (Berk.) G.H. Sung, J.M. Sung, Hywel-Jones & Spatafora 侵染青藏高原高山草甸土中的蝙蝠蛾(主要为钩蝠蛾属 *Thitarodes Viette*)幼虫而形成的幼虫尸体与真菌子座的复合体。由于具有独特的药用价值,加上野生资源不断减少、市场价格飙升,冬虫夏草受到国内外各界人士的关注,也成为近年科学研究的热点。冬虫夏草菌丝体已实现规模化发酵生产,但接种蝙蝠蛾幼虫培育产生子实体仍很困难^[1]。冬虫夏草菌的基因组已经完成测序^[2],但冬虫夏草菌的基础研究工作整体上还比较薄弱。为了更好地保护冬虫夏草野生资源,我国有关部门已将冬虫夏草菌列为“国家二级保护物种”。为了加快冬虫夏草菌的基础研究,最近冬虫夏草菌被推荐为中国的“国菌”(The national fungus of China)^[3]。

冬虫夏草不同分布区地形和气候的差异性,以及寄主昆虫种类的多样性使得冬虫夏草菌表现出明显的遗传分化。目前已有许多关于冬虫夏草菌遗传多样性的研究工作,使用的分子标记包括随机扩增多态性 DNA (Random amplified polymorphic DNA, RAPD)、简单序列重复间区扩增多态性 (Inter-simple sequence repeat, ISSR) 和一些 DNA 片段(如核糖体 DNA 内转录间隔区、交配型基因 *MAT1-2-1*) 的序列分析等^[4-7]。多基因序列分析的结果除揭示冬虫夏草菌的遗传多样性外,还增加了我们对冬虫夏草菌的起源(时间和地点)、扩散、与寄

主昆虫的进化关系等问题的认识^[6,8-9]。

之前的研究已证实不同冬虫夏草菌株在 *csp1*、*MAT1-1-1* 和 *MAT1-2-1* 这 3 个蛋白编码基因序列上存在碱基变异^[6,9-10],但却没有对冬虫夏草菌蛋白编码基因的分子进化进行更深入的分析。*csp1* 基因全长 1 372 bp,由 4 个外显子和 3 个内含子组成,编码的丝氨酸蛋白酶(信号肽 15 aa,前肽 90 aa,成熟肽 285 aa)可能参与冬虫夏草菌对寄主昆虫体壁的侵染^[11]。*MAT1-1-1* 基因全长 1 219 bp,含有 2 个内含子,编码含有 372 个氨基酸的蛋白质^[10]。*MAT1-2-1* 基因全长 857 bp,含有 2 个内含子,编码含有 249 个氨基酸的蛋白质^[12]。*MAT1-1-1* 和 *MAT1-2-1* 为交配型基因,参与调控冬虫夏草菌的有性生殖。本研究从 125 个分布于几乎所有主要产地的冬虫夏草样品中扩增 *csp1*、*MAT1-1-1* 和 *MAT1-2-1* 基因的序列,比较这 3 个基因外显子与内含子,以及两个交配型基因之间的进化关系,分析这些基因所承受的选择压力,并寻找 DNA 重组的证据。我们发现冬虫夏草菌不同蛋白编码基因以及基因的不同区域表现出进化方面的差异,因此,开展冬虫夏草菌进化相关的研究时应该联合使用多个基因片段,以降低单基因分析造成的结果偏差。

1 材料与方法

1.1 研究材料及 DNA 提取

研究共使用 125 个来自不同产地(包括中国的西藏、青海、四川、云南和甘肃,以及印度和尼泊尔部分地区)的冬虫夏草样品,具体样品来源信息参

见 Zhang 等^[9]文献。野生冬虫夏草样品或纯培养物用 CTAB (十六烷基三甲基溴化铵)法提取基因组 DNA 后, 作为 DNA 模板扩增 3 个基因片段。

1.2 基因片段扩增

从冬虫夏草样品中分别扩增 3 个基因, 即 *csp1* (丝氨酸蛋白酶基因)、*MAT1-1-1* (交配型 MAT1-1 idiomorph 中的基因)和 *MAT1-2-1* (交配型 MAT1-2 idiomorph 中的基因)。*csp1* 的扩增引物 CSP1-ZF (5'-ATGAAGCTCTCCCTCCTCATCG-3') 和 TSPa2 (5'-CGATGACAGCGGAGATGGTG-3')位于 *csp1* 外显子序列上^[11], 所扩增区域为包括前 2 个内含子和前 3 个外显子的部分 *csp1* 基因序列, 约 800 bp。*MAT1-1-1* 的扩增引物(本研究新设计) MAT1-1F1 (5'-ATTGTGTTCTCTCGCCTT-3')和 MAT1-1R1 (5'-CCGTGCCTTGAAATGTGCT-3')位于 *MAT1-1-1* 基因上下游非编码区, 所扩增区域为包含 2 个内含子的完整 *MAT1-1-1* 基因序列, 约 1.4 kb。*MAT1-2-1* 的扩增引物 MAT1-2F1 (5'-CCACCGATCCAAGTC TCCT-3')和 MAT1-2R2 (5'-CAGTTTCAGTCGCTG TCGTG-3')位于 *MAT1-2-1* 基因上下游非编码区^[6], 所扩增区域为包含 2 个内含子的完整 *MAT1-2-1* 基因序列, 约 1.0 kb。PCR 扩增体系(50 μ L): ddH₂O 35.5 μ L, 10 \times PCR buffer 5 μ L, dNTPs (2.5 mmol/L) 5 μ L, 正、反向引物(10 μ mol/L)各 1.5 μ L, *Pfu* (5 U/ μ L) 0.5 μ L, DNA 1 μ L。扩增条件为: 94 $^{\circ}$ C 5 min; 94 $^{\circ}$ C 40 s, 53 $^{\circ}$ C (*csp1* 和 *MAT1-1-1*)或 55 $^{\circ}$ C (*MAT1-2-1*) 40 s, 72 $^{\circ}$ C 60 s (*csp1* 和 *MAT1-2-1*)或 90 s (*MAT1-1-1*), 35 个循环; 72 $^{\circ}$ C 10 min。扩增产物送北京诺赛基因组研究中心有限公司进行测序。*csp1* 和 *MAT1-2-1* 的序列登录号见 Zhang 等^[9]的附表 1。*MAT1-1-1* 的序列登录号为 KR003998-KR004112。

1.3 序列变异分析

将不同冬虫夏草菌株同一基因片段的序列整理在一起, 利用 MEGA 6.0^[13]软件中的 Muscle^[14]算法进行序列比对, 分别统计外显子区和内含子区变异位点的数目。按照通用遗传密码表将外显子序列

翻译成氨基酸, 比较不同冬虫夏草菌株间蛋白质序列的差异。利用 DnaSP 5.10^[15]软件鉴别各 DNA 片段等位基因的数目(等位基因指一种特定的 DNA 序列; 分析时将插入缺失作为序列差异处理), 并计算核苷酸多样性和单倍型多样性指数。

1.4 系统发育分析及拓扑结构比较

将比对好的序列, 利用 PAUP 4.0b10 软件^[16]中的最大简约法(MP 法)启发式搜索 MP 树, 搜索参数为: 将比对引入的空缺(Gap)作为第 5 种碱基, 1 000 次重复随机加入序列, TBR (Tree-bisection-reconnection)分支交换算法。在 PAUP 软件中分别通过 KH (Kishino-Hasegawa)检测、Templeton 检测和 SH (Shimodaira-Hasegawa)检测, 比较外显子与内含子系统发育树之间或两个交配型基因外显子系统发育树之间的拓扑结构是否存在显著差异。

1.5 选择压力分析

为了明确 3 个蛋白编码基因经历着中性选择、正向选择, 还是负向选择, 在 MEGA 6.0^[13]软件中通过 Z 测验(选择 Modified Nei-Gojobori method with Jukes-Cantor correction)判断各蛋白编码基因(使用外显子等位基因的序列)整体的非同义替代率(Non-synonymous substitution rate, d_N)与同义替代率(Synonymous substitution rate, d_S)之间是否存在显著差异。利用 MEGA 6.0^[13]内嵌的 HyPhy 方法计算各密码子位点的 d_N 与 d_S , 计算平均 d_N 与 d_S 值及其比值(ω)。 $\omega > 1$ 表示非同义突变被固定的速率大于同义突变, 说明基因受到了正选择(Positive selection); $\omega = 1$ 为中性选择(Neutral selection); $\omega < 1$ 代表负选择(Negative selection)。

1.6 重组分析

为了分析是否重组导致了蛋白编码基因较高的遗传变异, 在 DnaSP 5.10^[15]软件中采用 Hudson 和 Kaplan^[17]的四配子检验法分析不同 DNA 位点(使用包含外显子及内含子的基因序列)间的重组; 同时利用 MultiLocus 1.3b^[18]软件进行不同蛋白编码基因的等位基因之间的连锁不平衡(Linkage

disequilibrium)分析, 计算 I_A (Index of association), r_d (I_A 的变型, 消除了 I_A 对位点数的依赖, 利于不同研究之间的比较) 和 PcP (Percentage of phylogenetically compatible pairs of loci), 通过比较观察值与随机抽样 1 000 次获得的数据进行统计分析。

2 结果与分析

2.1 核苷酸变异分析

从所有 125 个冬虫夏草样品中获得了 *csp1* 和 *MATI-2-1* 基因的序列, 但仅从 115 个样品中获得了 *MATI-1-1* 基因的序列, 其他 10 个样品(包括 3 个甘肃样品、2 个青海样品和 5 个西藏样品)没有扩增成功。这 3 个基因扩增的区域各自包含 2 个内含子。就内含子插入的相位而言, *csp1* 中两个内含子都是 0 相位插入; *MATI-1-1* 的两个内含子都是相位 1, 分别插入丝氨酸和苏氨酸的密码子中; *MATI-2-1* 第 1 个内含子是相位 2, 插入蛋氨酸的密码子中, 而第 2 个内含子是相位 1, 插入丝氨酸的密码子中。

分别统计了外显子和内含子区域的碱基变异(表 1)。这 3 个基因外显子区的长度在不同菌株间高度保守, 但 4.5%–5.7% 的位点有碱基变化, 部分位点的碱基变化导致编码的氨基酸发生改变; 菌株间氨基酸突变率为 4.6%–5.9%。对于内含子, *MATI-1-1* 的 2 个内含子及 *MATI-2-1* 的第 2 个内含子在菌株间没有长度变化, 但 *csp1* 的两个内含子及 *MATI-2-1* 的第 1 个内含子在菌株间有长度变化。特别是 *csp1* 的第 1 个内含子, 由于一段 19–20 bp 的序列串联重复次数的不同(2–6 次), 导致不同菌株中 *csp1* 第 1 个内含子的长度有很大不同, 从 81 bp 到 161 bp。除长度差异外, 内含子区也存在一些碱基变异位点; 在除去比对产生的空缺位点后剩余的内含子位点中, 变异位点的比例占到 1.8% (*MATI-2-1* 第 2 个内含子)–22% (*csp1* 第 2 个内含子)(表 1)。从以上分析可以看出, 冬虫夏草菌具有较高水平的遗传分化, 但不同基因间, 以及不同基因区

域的变异水平有差异。

就等位基因的数目而言, 外显子表现出比内含子更多的等位基因(表 1)。具体来讲, *csp1* 外显子共有 15 个等位基因, 而其两个内含子分别有 14 和 10 个等位基因; *MATI-1-1* 外显子共有 19 个等位基因, 而其两个内含子分别只有 2 和 5 个等位基因; *MATI-2-1* 外显子共有 20 个等位基因, 而其两个内含子分别只有 5 和 2 个等位基因。*MATI-2-1* 外显子及其两个内含子在有 *MATI-1-1* 扩增的 115 个菌株中, 分别有 19、5 和 2 个等位基因。如果综合考虑外显子和内含子组成的基因序列, 本研究在 *csp1*、*MATI-1-1* 和 *MATI-2-1* 中分别检测到 25、19 和 21 个等位基因。

有趣的是, 我们发现冬虫夏草菌的 *MATI-2-1* 基因有两种不同的终止密码子, 即有的菌株使用的是 TAA, 有的菌株使用的是 TGA。对于 *MATI-1-1* 基因, 所有菌株都使用终止密码子 TAG。对于 *csp1*, 虽然本研究未扩增到终止密码子所处的位置, 但根据此前的报道, 终止密码子也是 TAG^[11]。

2.2 外显子与内含子系统发育关系的比较

由于外显子与内含子所具有的等位基因数不同, 当分别用外显子和内含子构建系统发育树时, 二者的拓扑结构也存在不同。无论是使用代表基因(即考虑包括外显子和内含子的完整序列)等位基因(图 1), 还是仅使用代表内含子等位基因(图 2), 我们都发现外显子与内含子拓扑结构的差异通常都达到显著或极显著水平(表 2)。

2.3 两个交配型基因碱基变异水平及系统发育关系的比较

MATI-1-1 含有比 *MATI-2-1* 更多的变异位点(60 vs. 52), 但由于 *MATI-1-1* 的序列较长, 其变异位点的比例低于 *MATI-2-1* (基因全长中 4.9% vs. 6.1%; 外显子区 4.7% vs. 5.7%), 核苷酸多样性和单倍型多样性指数也比 *MATI-2-1* 低(表 1)。*MATI-1-1* 和 *MATI-2-1* 含有的等位基因数虽然非常接近(表 1), 但等位基因中的样品构成并不完全一致。利用

表 1 3 个蛋白编码基因核苷酸及氨基酸序列的变异分析
Table 1 Nucleotide and amino acid variations in three protein-coding genes

DNA 或氨基酸 DNA or amino acid	项目 Item	<i>csp1</i>	<i>MATI-1-1</i>	<i>MATI-2-1</i>	<i>MATI-2-1</i>
	菌株数 No. isolates	125	115	125	115
基因(外显子+内含子) Gene (exon+intron)	长度 Length (bp)	661-740	1 219	857-862	857-862
	变异位点数 ^a Variable site ^a	129 (42, 4, 83)	60 (46, 14, 0)	52 (37, 10, 5)	51 (36, 10, 5)
	等位基因数(频率) ^b No. alleles (frequency) ^b	25 (61)	19 (74)	21 (76)	20 (70)
	核苷酸多样性 ^c Nucleotide diversity ^c	0.007 7±0.001 1	0.002 9±0.000 6	0.004 7±0.000 8	0.004 6±0.000 8
	单倍型多样性 ^c Haplotype diversity ^c	0.741±0.039	0.573±0.053	0.619±0.050	0.619±0.052
外显子 Exon	长度 Length (bp)	529	1119	750	750
	变异位点数 Variable site	24 (22, 2, 0)	53 (40, 13, 0)	43 (34, 9, 0)	42 (33, 9, 0)
	等位基因数(频率) No. alleles (frequency)	15 (62)	19 (74)	20 (76)	19 (70)
第 1 内含子 The 1 st intron	长度 Length (bp)	81-161	52	52-57	52-57
	变异位点数 Variable site	92 (9, 2, 81)	2 (2, 0, 0)	8 (3, 0, 5)	8 (3, 0, 5)
	等位基因数(频率) No. alleles (frequency)	14 (61)	2 (111)	5 (113)	5 (104)
第 2 内含子 The 2 nd intron	长度 Length (bp)	50-52	48	55	55
	变异位点数 Variable site	13 (11, 0, 2)	5 (4, 1, 0)	1 (0, 1, 0)	1 (0, 1, 0)
	等位基因数(频率) No. alleles (frequency)	10 (76)	5 (107)	2 (124)	2 (114)
氨基酸 Amino acid	长度 Length (aa)	175	372	249	249
	变异氨基酸数 Variable amino acid (%)	8 (4.6)	22 (5.9)	14 (5.6)	14 (5.6)

注: ^a: 对于变异位点, 括号外面为总的变异位点数, 括号里面分别为信息位点数、单碱基变异位点数和插入缺失位点数; ^b: 对于等位基因, 括号外面的数字代表等位基因数目, 括号里面的数字代表最优势等位基因含有的菌株数; ^c: 计算核苷酸多样性和单倍型多样性指数时, 不考虑比对产生的空缺, 因此, 实际值可能略微高一些。

Note: ^a: The total number of variable sites is given outside parentheses and that of phylogenetically informative sites, phylogenetically uninformative sites, and indel sites are given individually within parentheses; ^b: Numbers outside parentheses represent the total number of alleles, and those within parentheses are the number of individuals represented by the most dominant allele; ^c: When calculating nucleotide diversity and haplotype diversity indices, sites with gaps were excluded. Therefore, the actual value could be slightly higher.

两个交配型基因的外显子序列分别构建系统发育树, 我们发现无论使用 115 个同时有 *MATI-1-1* 和 *MATI-2-1* 扩增的样品, 26 个代表 *MATI-1-1* 及 *MATI-2-1* 不同等位基因的样品(图 3), 19 个代表

MATI-1-1 不同等位基因的样品, 或者 19 个代表 *MATI-2-1* 不同等位基因的样品, 两个交配型基因外显子系统发育树之间的拓扑关系都存在极显著差异(表 2)。

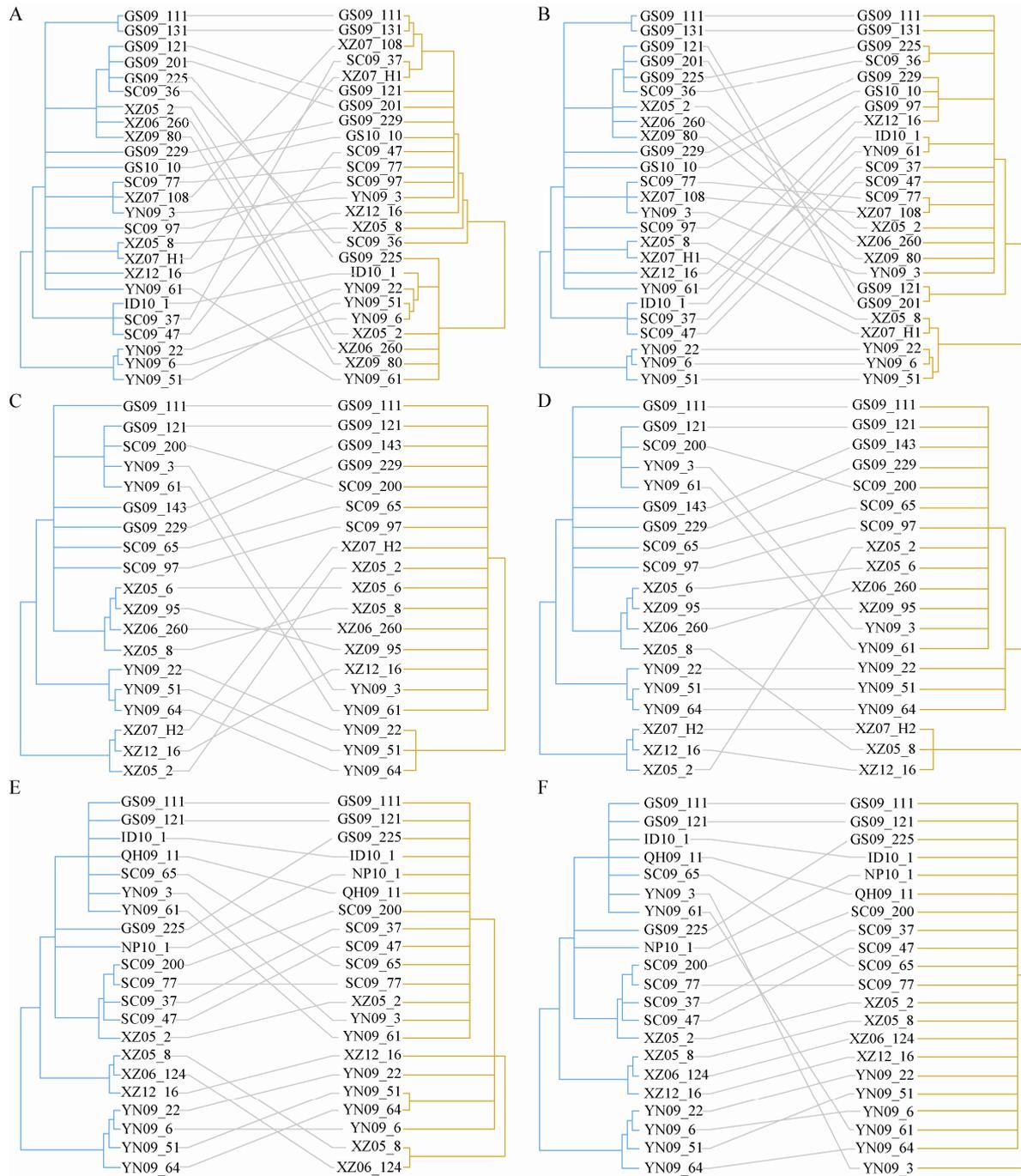


图1 利用代表基因各等位基因的样品比较外显子与内含子的系统发育树

Figure 1 Comparison between exon and intron phylogenies based on representative individuals of each gene allele

注: 基于基因(外显子+内含子)各等位基因, 选用了25、19和21个菌株分别对 *cspl* (A, B)、*MAT1-1-1* (C, D)和 *MAT1-2-1* (E, F) 进行分析。A、C、E 分别是比较 *cspl* (A)、*MAT1-1-1* (C)和 *MAT1-2-1* (E)外显子与各基因中对应的第1个内含子。B、D、F 分别是比较 *cspl* (B)、*MAT1-1-1* (D)和 *MAT1-2-1* (F)外显子与各基因中对应的第2个内含子。在每一图版中, 左侧为外显子系统发育树, 右侧为内含子系统发育树。

Note: In this figure, isolates representing *cspl* (25 isolates; A, B), *MAT1-1-1* (19 isolates; C, D), and *MAT1-2-1* (21 isolates; E, F) gene alleles (i.e., based on both exon and intron sequences) were employed. Panels A, C, and E are comparisons between exons and the 1st introns in *cspl* (A), *MAT1-1-1* (C), and *MAT1-2-1* (E). Panels B, D, and F are comparisons between exons and the 2nd introns in *cspl* (B), *MAT1-1-1* (D), and *MAT1-2-1* (F). Within each panel, the left tree is exon phylogeny, and the right one is intron phylogeny.

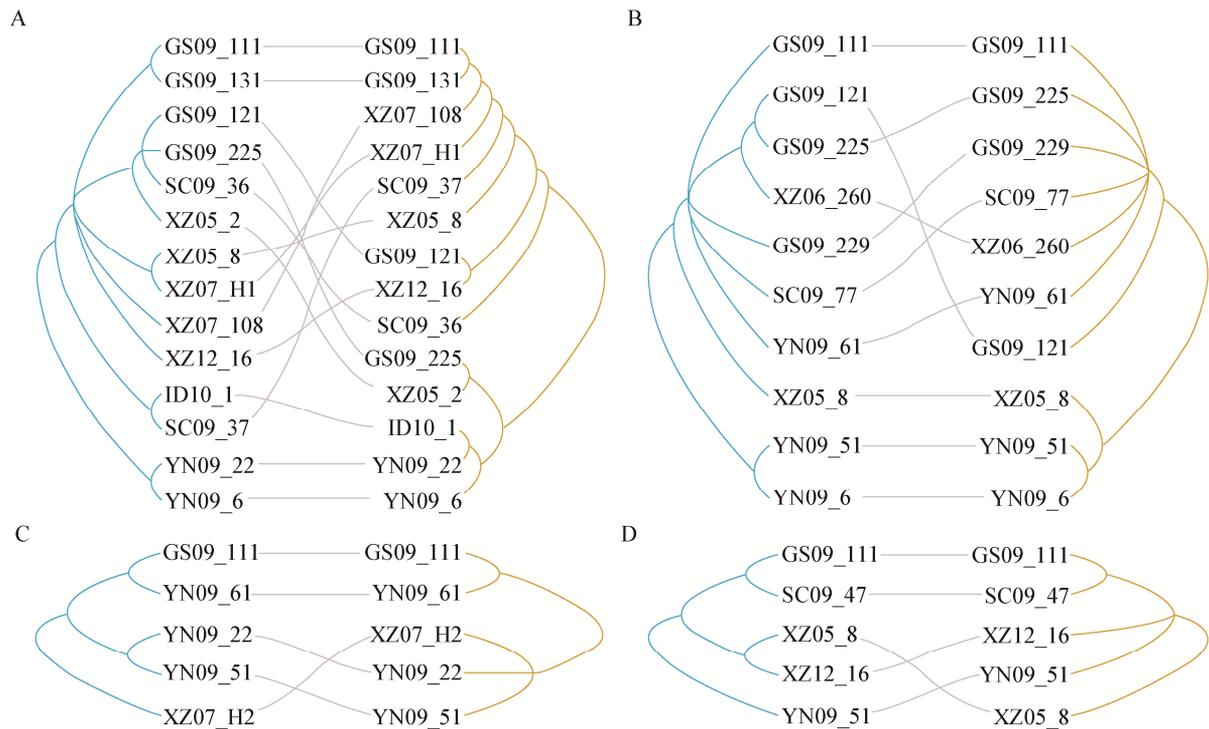


图 2 利用代表内含子各等位基因的样品比较外显子与内含子的系统发育树

Figure 2 Comparison between exon and intron phylogenies based on representative individuals of each intron allele

注: 基于内含子各等位基因, 我们分别选用了 14 或 10、5、5 个菌株对 *csp1* (A, B)、*MAT1-1-1* (C) 和 *MAT1-2-1* (D) 进行分析. 图版 A、B、C、D 分别是比较 *csp1* 外显子与其第 1 个内含子(A)、*csp1* 外显子与其第 2 个内含子(B)、*MAT1-1-1* 外显子与其第 2 个内含子(C) 和 *MAT1-2-1* 外显子与其第 1 个内含子(D). 在每一图版中, 左侧为外显子系统发育树, 右侧为内含子系统发育树. 我们没有测试 *MAT1-1-1* 中的第 1 个内含子和 *MAT1-2-1* 的第 2 个内含子, 因为它们各自只含有两个等位基因.

Note: In this figure, isolates representing *csp1* (14 or 10 isolates; A, B), *MAT1-1-1* (5 isolates; C), and *MAT1-2-1* (5 isolates; D) intron alleles (i.e., based on only one representative isolate of each intron allele) were employed. Panels A, B, C, and D are comparisons between *csp1* exon and the 1st intron (A), *csp1* exon and the 2nd intron (B), *MAT1-1-1* exon and the 2nd intron (C), and *MAT1-2-1* exon and the 1st intron (D), respectively. Within each panel, the left tree is exon phylogeny, and the right one is intron phylogeny. We did not test for the first *MAT1-1-1* intron and the second *MAT1-2-1* intron because they each had only two intron alleles.

2.4 3 个蛋白编码基因选择压力分析

通过 Z 测验对 3 个蛋白编码基因进行整体分析时, d_N 都显著小于 d_S , 3 种对立假设分析的结果支持这 3 个蛋白编码基因都经历过净化选择作用(表 3)。通过 HyPhy 对各密码子位点逐一分析的结果与此一致, 即平均 d_N/d_S 值小于 1 (表 3)。

2.5 重组分析

在 3 个蛋白编码基因中都检测到 DNA 位点间的重组事件(表 4)。其中, 在本研究所测定的部分 *csp1* 序列中至少检测到 6 次重组事件(S281 & S286、S286 & S457、S490 & S517、S517 & S538、S571 & S697、S697 & S710), 除 S538 位于内含子外, 其它

位点都位于外显子中。在完整的 *MAT1-1-1* 基因序列中至少检测到 1 次重组事件(S586 & S766), 这两个位点都位于外显子中。在完整的 *MAT1-2-1* 基因序列中至少检测到 6 次重组事件(S121 & S487、S513 & S595、S595 & S613、S745 & S770、S770 & S810、S811 & S861), 这些位点都位于外显子中。当对 2 个或 3 个基因片段综合分析时, 至少有一次重组事件发生在两个基因间, 即 *MAT1-1-1* 中的 1 个位点(S1195)与 *MAT1-2-1* 中的 1 个位点(S47)间连锁不平衡分析的结果不支持这 3 个基因之间有重组发生(表 4)。

表 2 不同系统发育树之间拓扑结构的比较
Table 2 Comparison on topological difference between different phylogenetic trees

待比较的项目 ^a Items in comparison ^a	使用的菌株 ^b Isolates used ^b	KH ^c	Templeton ^c	SH ^c
<i>csp1</i> 外显子与第 1 个内含子 <i>csp1</i> exon vs. 1 st intron	25 个基因等位基因菌株 25 gene allele isolates	0.001 5	0.001 8	0.002
	14 个第 1 内含子等位基因菌株 14 1 st intron allele isolates	0.000 1	0.000 1	<0.001
<i>csp1</i> 外显子与第 2 个内含子 <i>csp1</i> exon vs. 2 nd intron	25 个基因等位基因菌株 25 gene allele isolates	0.047 3	0.046 4	0.014
	10 个第 2 内含子等位基因菌株 10 2 nd intron allele isolates	0.180 0	0.179 7	0.068
<i>MATI-1-1</i> 外显子与第 1 个内含子 <i>MATI-1-1</i> exon vs. 1 st intron	19 个基因等位基因菌株 19 gene allele isolates	0.000 1	0.000 1	<0.001
<i>MATI-1-1</i> 外显子与第 2 个内含子 <i>MATI-1-1</i> exon vs. 2 nd intron	19 个基因等位基因菌株 19 gene allele isolates	0.000 3	0.000 2	<0.001
	5 个第 2 内含子等位基因菌株 5 2 nd intron allele isolates	0.000 9	0.000 9	0.002
<i>MATI-2-1</i> 外显子与第 1 个内含子 <i>MATI-2-1</i> exon vs. 1 st intron	21 个基因等位基因菌株 21 gene allele isolates	<0.000 1	<0.000 1	<0.001
	5 个第 1 内含子等位基因菌株 5 1 st intron allele isolates	0.002 6	0.002 7	0.007
<i>MATI-2-1</i> 外显子与第 2 个内含子 <i>MATI-2-1</i> exon vs. 2 nd intron	21 个基因等位基因菌株 21 gene allele isolates	<0.000 1	<0.000 1	<0.001
<i>MATI-1-1</i> 外显子与 <i>MATI-2-1</i> 外显子 <i>MATI-1-1</i> exon vs. <i>MATI-2-1</i> exon	115 个两个基因都有扩增的菌株 115 isolates with both gene amplifications	0.001 2	<0.000 1	0.001
	26 个代表 <i>MATI-1-1</i> 及 <i>MATI-2-1</i> 等位基因的菌株 26 either gene allele isolates	<0.000 1	<0.000 1	<0.001
	19 个代表 <i>MATI-1-1</i> 等位基因的菌株 19 <i>MATI-1-1</i> allele isolates	<0.000 1	<0.000 1	<0.001
	19 个代表 <i>MATI-2-1</i> 等位基因的菌株 19 <i>MATI-2-1</i> allele isolates	0.003 0	<0.000 1	<0.001

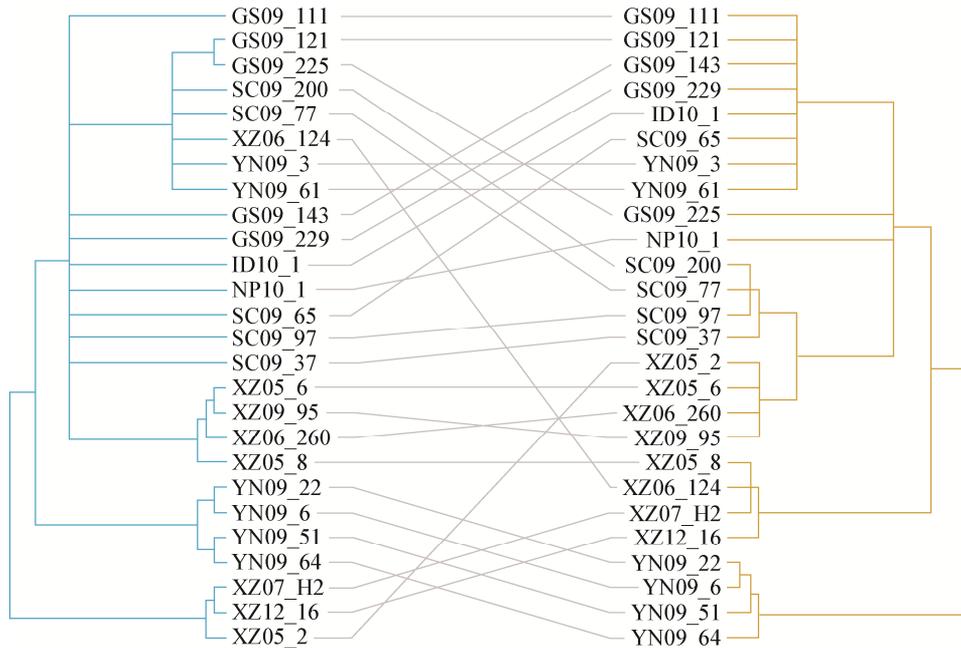
注: ^a: 未比较 *MATI-1-1* 第 1 个内含子和 *MATI-2-1* 第 2 个内含子, 因为它们各自分别含有 2 个等位基因; ^b: 对于基因(外显子+内含子)或内含子的各等位基因, 只选择一个代表菌株构建系统发育树; ^c: 在每一测试中, 我们都给出了拒绝零假设(即两棵树无差异)的概率值。

Note: ^a: We did not test for the 1st *MATI-1-1* intron and the 2nd *MATI-2-1* intron because they each had only two intron alleles; ^b: For each gene (exons + introns) or intron allele, one representative isolate was chosen for phylogeny construction; ^c: For each test, the probability to reject the null hypothesis of no difference between two trees is given.

3 讨论

之前的研究发现, 每一个被测试的冬虫夏草菌株都含有两种交配型“idiomorph” (即都能扩增出 *MATI-1-1* 和 *MATI-2-1*), 因此冬虫夏草菌可能为同宗配合真菌^[2,12]。Bushley 等根据冬虫夏草菌丝和子囊孢子细胞核染色观察到双核的结果, 认为冬虫夏草菌可能为假同宗配合真菌^[10], 然而在冬虫夏草菌

细胞内观察到的双核是否具有相同或不同的遗传物质尚不清楚。本研究从部分冬虫夏草样品中没有扩增到 *MATI-1-1* 基因, 但可以扩增出 *MATI-2-1* 基因; 作者在本研究未使用的其他冬虫夏草样品中还曾遇到过不能扩增出 *MATI-2-1*, 而能扩增出 *MATI-1-1* 的菌株。对于这些只能扩增到一种交配型基因的菌株, 需要通过分子杂交或基因组测序确认

图 3 *MATI-1-1* 和 *MATI-2-1* 外显子拓扑关系的比较Figure 3 Comparison between *MATI-1-1* and *MATI-2-1* exon phylogenies

注: 使用了 26 个代表 *MATI-1-1* 及 *MATI-2-1* 不同等位基因样品。

Note: In this figure, 26 isolates representing either *MATI-1-1* or *MATI-2-1* alleles were used.

表 3 3 个蛋白编码基因的选择压力分析

Table 3 Selection test of three protein-encoding genes

基因 Gene	基于密码子的 Z 测验 Codon-based Z test						平均 d_N/d_S 值 HyPhy Average d_N/d_S HyPhy
	中性假说 Neutrality		正选择 Positive selection		净化选择 Purifying selection		
	P-value	Stat	P-value	Stat	P-value	Stat	
<i>csp1</i>	0.005 7	-2.82	1.00	-2.79	0.002 7	2.83	0.08
<i>MATI-1-1</i>	0.000 5	-3.55	1.00	-3.49	0.000 3	3.58	0.26
<i>MATI-2-1</i>	0.000 2	-3.79	1.00	-3.83	0.000 1	4.00	0.19

注: 进行 Z 测验时, 拒绝零假设(即严格中性假说)的概率值取决于依赖的对立假设(HA)。因此, 分别测试了 3 种不同的对立假设, 即 $d_N \neq d_S$ (中性假说)、 $d_N > d_S$ (正选择)和 $d_N < d_S$ (净化选择)。在 Stat 栏下给出了 d_N 与 d_S 的差值, 但对于 HA=中性假说或正选择, 为 $d_N - d_S$; 对于 HA=净化选择, 为 $d_S - d_N$ 。

Note: When performing codon-based Z tests, the level of significance at which the null hypothesis of strict-neutrality ($d_N = d_S$) is rejected depends on the alternative hypothesis (HA). We tested three alternative hypotheses: $d_N \neq d_S$ (test of neutrality), $d_N > d_S$ (positive selection) and $d_N < d_S$ (purifying selection). The test statistic ($d_N - d_S$; for HA neutrality and positive selection) or ($d_S - d_N$; for HA purifying selection) is shown in the Stat column.

表 4 基因内部和基因间的重组分析

Table 4 Recombination analyses within individual gene and between different genes

Dataset ^a	R_M^b	I_A^c	r_d^c	PcP ^d
<i>csp1</i>	6	—	—	—
<i>MAT1-1-1</i>	1	—	—	—
<i>MAT1-2-1</i>	6	—	—	—
<i>csp1+MAT1-2-1</i>	12	0.554 (<0.001)	0.557 (<0.001)	1.00
<i>csp1+MAT1-1-1+MAT1-2-1</i>	13	1.137 (<0.001)	0.570 (<0.001)	1.00

注: ^a: 对于数据集 *csp1*、*MAT1-2-1* 和 *csp1+MAT1-2-1*, 使用了所有 125 个菌株; 对于数据集 *MAT1-1-1* 和 *csp1+MAT1-1-1+MAT1-2-1*, 只使用了 115 个 3 个基因都有扩增的菌株. ^b: R_M 代表最少重组事件数. 当对含有 115 个菌株的 *csp1+MAT1-1-1+MAT1-2-1* 数据集进行分析时, 之前从含 125 个菌株的 *MAT1-2-1* 数据集中找到的 S121 和 S487 位点间的重组消失了, 但出现另一对新的重组位点, 即在 *MAT1-1-1* S1195 和 *MAT1-2-1* S47 之间. ^c: I_A 代表位点间的关联指数. r_d 是 I_A 的变型, 经过调整可不受位点数目的影响, 从而利于不同研究间的比较. 二者的零假设都是随机重组. ^d: PcP 代表相容位点对的比例. 只要 PcP 值小于 1, 就代表有重组.

Note: ^a: For datasets *csp1*, *MAT1-2-1*, and *csp1+MAT1-2-1*, all 125 isolates are included; for datasets *MAT1-1-1* and *csp1+MAT1-1-1+MAT1-2-1*, only the 115 isolates that had amplification of all the three DNA fragments are included. ^b: R_M , minimum number of recombination events. For the 115-isolate dataset *csp1+MAT1-1-1+MAT1-2-1*, the recombination between S121 and S487 within *MAT1-2-1* was missing (due to lower number of *MAT1-2-1* sequences than in the 125-isolate *MAT1-2-1* dataset), but we detected one new recombination between *MAT1-1-1* (at S1195) and *MAT1-2-1* (at S47). ^c: I_A , index of association. r_d is a modified measure of I_A , but adjusted for the number of loci. The adjustment helps facilitate comparison between populations and studies. The null hypothesis for both I_A and r_d is that there is random association among alleles from different genes (i.e., recombination). ^d: PcP, percentage of phylogenetically compatible pairs of loci. The null hypothesis for PcP is clonality, therefore, a PcP value lower than 1 would indicate recombination.

相反交配型的基因是否确实在基因组中缺失。如果冬虫夏草确实存在只有一种交配型的菌株, 也是可以理解的, 因为这种偶尔的异宗配合赋予冬虫夏草菌兼性杂交的能力, 有利于该菌的生存和进化。这种一个真菌物种同时表现出(假)同宗配合和异宗配合的现象在其他一些真菌中也有过报道, 如 *Neurospora tetrasperma*^[19]、*Cryptococcus neoformans*^[20]和 *Saccharomyces cerevisiae*^[21]。

通常, 交配型基因由于受净化选择压力的影响, 为避免有害突变产生而在真菌种内表现出较低的遗传变异。例如, *Cochliobolus heterostrophus* 和 *Phaeosphaeria nodorum* 交配型基因的种内变异率只有 0.15%–0.31%^[22-23]。然而, 我们在冬虫夏草菌中却发现交配型基因有很高的种内变异 (4.9%–6.1%), 而且 *MAT1-2-1* 基因的变异大于 *MAT1-1-1* 基因。Bushley 等也发现了类似的结果, 但其在 *MAT1-1-1* 和 *MAT1-2-1* 基因中分别只发现了 2 个和 9 个碱基变异位点^[10]。本研究通过对大量冬虫夏草菌株的分析从两个交配型基因中发现了非

常多的核苷酸变异位点(表 1), 用 *MAT1-1-1* 和 *MAT1-2-1* 外显子构建的系统发育树的拓扑结构存在显著差异(图 3), 然而造成 *MAT1-1-1* 和 *MAT1-2-1* 之间这些不同的原因仍有待分析。Stewart 等对 *Alternaria* 属真菌的研究发现交配型基因 *MAT1-1-1* 和 *MAT1-2-1* 在功能、选择压力和表达水平等方面的不同可能解释两个交配型基因变异水平的差异^[24]。除此之外, 冬虫夏草菌交配型基因中一些 DNA 位点的碱基变异改变了编码的氨基酸(表 1), 但这种氨基酸的改变是否影响交配型基因的功能尚不清楚。

本文发现 3 个蛋白编码基因(*csp1*、*MAT1-1-1* 和 *MAT1-2-1*)的外显子和内含子反映的系统发育关系不一致(图 1、2, 表 2), 这暗示着外显子与内含子承受的进化压力不同。由于外显子的长度在不同的冬虫夏草菌株中高度保守, 而内含子的长度在不同菌株间可能相同或不同, 因此, 内含子可能承受着比外显子更大的选择压力。外显子与内含子系统发育关系的不一致还暗示着这些内含子可能是通

过水平基因转移获得的。此外, 由于外显子序列中已经具有足够多的变异位点(4.5%–5.7%, 表 1), 今后用蛋白编码基因构建系统发育树时可以去掉内含子序列, 而只使用外显子序列。同时, 由于用不同基因的外显子序列构建的系统发育树间存在一定的拓扑结构差异(图 3, 表 2), 这又提示我们应该联合使用多个基因片段进行冬虫夏草菌进化相关的分析。

虽然冬虫夏草菌是同宗配合还是异宗配合还需更多的证据支持, 但可以肯定的是冬虫夏草菌能够通过有性生殖进行遗传重组。本研究在 3 个蛋白基因内部检测到多个 DNA 位点间的重组, 但这 3 个基因间并未检测到很强的重组信号(表 3), 这可能与 3 个基因在基因组中的相对遗传距离有关, 但对此我们尚无确切的数据。我们在冬虫夏草菌的其他一些 DNA 片段间曾检测到 DNA 重组。例如, 对于 Zhang 等使用的 7 个冬虫夏草菌 DNA 片段^[9], 我们在 1/3 的 DNA 片段组合中检测到重组发生, 包括 ITS 和 OSRC32、*csp1* 和 OSRC32、*MAT1-2-1* 和 OSRC32、OSRC17 和 OSRC32、OSRC27 和 OSRC32、ITS 和 OSRC17、OSRC17 和 OSRC27 (未发表数据)。

冬虫夏草菌的蛋白编码基因虽然经历着净化选择作用(表 2), 但却表现出较高的遗传分化(表 1), 这可能有多种原因。首先, 冬虫夏草菌能以有性生殖方式繁殖, 具备基因重组的前提条件; 其次, 青藏高原上不同分布区的地形和气候存在差异, 冬虫夏草菌为适应不同的环境而进行适应性进化; 再者, 冬虫夏草菌与寄主昆虫具有紧密的进化关系^[9], 而寄主昆虫种类非常丰富, 促进了冬虫夏草菌的分化。

参 考 文 献

- Zhang YJ. Biology of the Chinese Caterpillar Fungus *Ophiocordyceps sinensis*[M]. Beijing: Science Press, 2012 (in Chinese)
张永杰. 冬虫夏草菌的生物学研究[M]. 北京: 科学出版社, 2012
- Hu X, Zhang YJ, Xiao GH, et al. Genome survey uncovers the secrets of sex and lifestyle in caterpillar fungus[J]. Chinese Science Bulletin, 2013, 58(23): 2846-2854
- Zhang YJ, Li EW, Wang CS, et al. *Ophiocordyceps sinensis*, the flagship fungus of China: terminology, life strategy and ecology[J]. Mycology, 2012, 3(1): 2-10
- Chen YJ, Zhang YP, Yang YX, et al. Genetic diversity and taxonomic implication of *Cordyceps sinensis* as revealed by RAPD markers[J]. Biochemical Genetics, 1999, 37(5/6): 201-213
- Liang HH, Cheng Z, Yang XL, et al. Genetic diversity and structure of *Cordyceps sinensis* populations from extensive geographical regions in China as revealed by inter-simple sequence repeat markers[J]. Journal of Microbiology, 2008, 46(5): 549-556
- Zhang YJ, Xu LL, Zhang S, et al. Genetic diversity of *Ophiocordyceps sinensis*, a medicinal fungus endemic to the Tibetan Plateau: Implications for its evolution and conservation[J]. BMC Evolutionary Biology, 2009, 9: 290
- Zhang S, Zhang YJ, Bhushan S, et al. *Ophiocordyceps sinensis* and *Cordyceps militaris*: research advances, issues and perspectives[J]. Mycosystema, 2013, 32(4): 577-597 (in Chinese)
张姝, 张永杰, Bhushan S, 等. 冬虫夏草菌和蛹虫草的研究现状、问题及展望[J]. 菌物学报, 2013, 32(4): 577-597
- Quan QM, Wang QX, Zhou XL, et al. Comparative phylogenetic relationships and genetic structure of the Caterpillar Fungus *Ophiocordyceps sinensis* and its host insects inferred from multiple gene sequences[J]. Journal of Microbiology, 2014, 52(2): 99-105
- Zhang YJ, Zhang S, Li YL, et al. Phylogeography and evolution of a fungal-insect association on the Tibetan Plateau[J]. Molecular Ecology, 2014, 23(21): 5337-5355
- Bushley KE, Li Y, Wang WJ, et al. Isolation of the *MAT1-1* mating type idiomorph and evidence for selfing in the Chinese medicinal fungus *Ophiocordyceps sinensis*[J]. Fungal Biology, 2013, 117(9): 599-610
- Zhang YJ, Liu XZ, Wang M. Cloning, expression, and characterization of two novel cuticle-degrading serine proteases from the entomopathogenic fungus *Cordyceps sinensis*[J]. Research in Microbiology, 2008, 159(6): 462-469
- Zhang S, Zhang YJ, Liu XZ, et al. Cloning and analysis of the *MAT1-2-1* gene from the traditional Chinese medicinal fungus *Ophiocordyceps sinensis*[J]. Fungal Biology, 2011, 115(8): 708-714
- Tamura K, Stecher G, Peterson D, et al. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0[J]. Molecular Biology and Evolution, 2013, 30(12): 2725-2729
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput[J]. Nucleic Acids Research, 2004, 32(5): 1792-1797
- Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data[J]. Bioinformatics, 2009, 25(11): 1451-1452
- Swofford DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)[Z]. Version 4.0 b10. Sinauer Associates, Sunderland. Accessed 2003. <http://paup.csit.fsu.edu/index.html>
- Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences[J]. Genetics, 1985, 111(1): 147-164
- Agapow PM, Burt A. Indices of multilocus linkage disequilibrium[J]. Molecular Ecology Notes, 2001, 1(1/2): 101-102
- Merino ST, Nelson MA, Jacobson DJ, et al. Pseudohomothallism and evolution of the mating-type chromosome in *Neurospora tetrasperma*[J]. Genetics, 1996, 143(2): 789-799
- Lin X, Hull CM, Heitman J. Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*[J]. Nature, 2005, 434(7036): 1017-1021
- Katz Ezov T, Chang SL, Frenkel Z, et al. Heterothallism in *Saccharomyces cerevisiae* isolates from nature: effect of HO locus on the mode of reproduction[J]. Molecular Ecology, 2010,

- 19(1): 121-131
- [22] Bennett R, Yun S, Lee T, et al. Identity and conservation of mating type genes in geographically diverse isolates of *Phaeosphaeria nodorum*[J]. Fungal Genetics and Biology, 2003, 40(1): 25-37
- [23] Turgeon B. Application of mating type gene technology to problems in fungal biology[J]. Annual Review of Phytopathology, 1998, 36(1): 115-137
- [24] Stewart JE, Kawabe M, Abdo Z, et al. Contrasting codon usage patterns and purifying selection at the mating locus in putatively asexual *Alternaria* fungal species[J]. PLoS One, 2011, 6(5): e20083

2015 年中国微生物学会及各专业委员会学术活动计划表

序号	会议名称	主办/协办单位	时间	人数	地点	联系方式
1	第二届国际重症休克与脓毒症高峰论坛	中国微生物学会微生物毒素专业委员会	3月底	1500人	广东广州	张庆红 010-66867382 z_qinghong@aliyun.com
2	全国“发酵工程课程研讨会”	中国微生物学会生化过程模型化与控制专业委员会	4月	120	上海	夏建业 jyxia@ecust.edu.cn
3	兽医微生物教学研讨	中国微生物学会兽医微生物学专业委员会	5月	30	山东泰安	13683505108
4	《海洋生物高技术丛书》分册 5: 海洋微生物资源开发利用审稿会	中国微生物学会海洋微生物学专业委员会	5月	30	山东青岛	焦炳华
5	第六届传染病防控基础研究与应用技术论坛	中国微生物学会分析微生物专业委员会	6月	300	待定	吕相征 lvxz@cma.org.cn
6	第十五届微生物学教学和科研及成果产业化研讨会	中国微生物学会农业微生物学专业委员会和普通微生物学专业委员会联合主办	7月	200	新疆乌鲁木齐	努尔古丽·热合曼 nurgulum@163.com
7	第三届全国昆虫-微生物联合转化有机废弃物机制及资源化利用研讨会	中国微生物学会农业微生物学专业委员会	7月	150	山东泰安	刘玉升 ysl8877@163.com
8	全国酶工程学术研讨会	酶工程专业委员会	7-8月	200	待定	
9	工业企业微生物安全控制技术与实践研讨会	中国微生物学会工业微生物学专业委员会	8月	150	北京	010-53218310
10	第12届全国海洋药物论坛	中国微生物学会海洋微生物学专业委员会	8月	200	浙江舟山	林文瀚 13701285168
11	第7届全国微生物资源学术暨国际微生物系统与分类学研讨会	中国微生物学会微生物资源专业委员会	8月25-30日	400	浙江杭州	阮志勇 010-82108651-620 许学伟 0571-81963208
12	第六届中国临床微生物学大会暨生物学与免疫学论坛	中国微生物学会临床微生物学专业委员会	9月11-13日	400	待定	0574-87035856
13	生物安全培训会议	中国微生物学会微生物生物安全专业委员会	9月	80	武汉	18600189362
14	医学微生物学与免疫学专委会青年学组成立并学组研讨会	中国微生物学会医学微生物学与免疫学专业委员会	9月下旬	100	重庆	hoofuquan@aliyun.com
15	第十一届全国病毒学学术会议	中国微生物学会病毒学专业委员会	10月	600	湖北武汉	吴莹 wuying@im.ac.cn
16	全国发酵过程优化与控制高级技术培训班	中国微生物学会生化过程模型化与控制专业委员会	10月	80-100	上海	刘健 jliu@nc-bio.com
17	2015年医学真菌学新进展学术研讨会暨中美真菌班举办三十周年纪念会	中国微生物学会真菌学专业委员会	10月	200	江苏南京	刘维达
18	2015年中国微生物学会学术年会	中国微生物学会	10月23-26日	600	湖北宜昌	杨海花, 王旭 010-64807200
19	第十八次全国环境微生物学学术研讨会	中国微生物学会环境微生物学专业委员会	11月13-16日	500	江苏镇江	蒋建东 025-84399726
20	第十届全国芽胞杆菌青年工作者学术研讨会	中国微生物学会农业微生物学专业委员会	11月底	100	湖北武汉	胡晓敏 huxm@wh.iov.cn
21	中国微生物与白酒酿造技术研讨会	中国微生物学会工业微生物学专业委员会	12月	150	四川宜宾	010-53218310
22	生物安全研讨会	中国微生物学会微生物生物安全专业委员会	12月	60	北京	18600189362
23	微生物的全基因组测序及生物信息学分析	中国微生物学会生物制品专业委员会	待定	150	待定	67095437/67095601