



食线虫真菌洛斯里被毛孢线粒体基因组的再分析

闫倩倩¹, 刘杏忠², 张永杰^{1*}

¹山西大学生命科学学院, 山西 太原 030006

²中国科学院微生物研究所真菌学国家重点实验室, 北京 100101

摘要:【目的】鉴定洛斯里被毛孢 OWVT-1 菌株的线粒体基因组, 验证公布的 USA-87-5 菌株线粒体基因组中的错误, 对洛斯里被毛孢正确的线粒体基因组序列进行注释并开展不同被毛孢物种间的比较线粒体基因组学分析。【方法】借助 DNA 高通量测序数据并通过必要的 Sanger 测序组装 OWVT-1 的线粒体基因组。通过 PCR 验证 OWVT-1 与公布的 USA-87-5 线粒体基因组序列差异的真实性。利用多种生物信息方法分析和注释洛斯里被毛孢的线粒体基因组。【结果】公布的洛斯里被毛孢 USA-87-5 菌株的线粒体基因组存在几处序列错误, 包括 3 处长片段的插入缺失和多处短片段的插入缺失。实际上, 洛斯里被毛孢 USA-87-5 与 OWVT-1 菌株的线粒体基因组序列完全相同。该菌的线粒体基因组全长 62949 bp, 在 7 个基因中共插入 13 个内含子, 部分内含子和基因间区显现出序列退化的特征。洛斯里被毛孢、明尼苏达被毛孢、线虫被毛孢的线粒体基因组具有较强的共线性关系。除一些独立的 ORF 外, 核心蛋白编码基因、rRNA 基因和 tRNA 基因的排列顺序非常保守。基因间区的长短是影响 3 种被毛孢线粒体基因组大小最主要的因素。【结论】公布的洛斯里被毛孢 USA-87-5 菌株线粒体基因组中存在序列错误。本文新报道了 OWVT-1 菌株的线粒体基因组, 并进行注释和比较线粒体基因组学分析。

关键词: 洛斯里被毛孢, 线粒体基因组, 校正, 注释, 比较线粒体基因组

线粒体(mitochondrion)是真核细胞中的一种半自主性细胞器, 被普遍认为起源于寄生在真核生物共同祖先细胞内的远古细菌, 具有自身的基因组, 独立于细胞核基因组而遗传(主要是单亲遗传), 不受减数分裂时核染色体重组的影响^[1]。线

粒体基因组虽然明显小于细胞核基因组, 但由于拷贝数高而成为细胞总 DNA 的重要组成部分。线粒体基因组特定的基因组成、遗传密码和复制方式, 记载着生物从单细胞到多细胞、从低等到高等进化过程中丰富的历史信息, 是研究物种起源

基金项目: 山西省自然科学基金(201601D011065); 山西省留学回国人员科技活动择优资助项目; 山西省回国留学人员科研资助项目(2017-015)

*通信作者。E-mail: zhangyj2008@sxu.edu.cn

收稿日期: 2018-03-29; 修回日期: 2018-04-25; 网络出版日期: 2018-05-12

与进化的有力工具^[2]。

近年来,随着 DNA 测序技术的发展和测序成本的降低,越来越多真菌的线粒体基因组被报道。截止 2018 年 3 月,NCBI 细胞器基因组参考数据库中已有约 260 种真菌的线粒体基因组([http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=4751 & opt=organelle](http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=4751&opt=organelle)),其中子囊菌 210 种(占 80%)。仅 2016 和 2017 两年就有 73 种真菌(其中 67 种为子囊菌)的线粒体基因组被测序。虽然基于高通量测序数据能够较容易地得到真菌的线粒体基因组,但在组装过程中需十分谨慎,必要时需通过 PCR 扩增和测序确认可疑区域的序列,否则组装出的结果容易出现序列错误。例如,肺囊康定产生菌 *Glarea lozoyensis* ATCC 74030 菌株由于 *cox2* 基因具有 2 个拷贝,当仅基于高通量测序数据组装线粒体基因组时,在这两个 *cox2* 基因处出现了组装错误^[3]。另外,我们发现一些发表的线粒体基因组中部分基因的注释也不够准确。

被毛孢属 *Hirsutella* Pat. 是一类寄主广泛、全球性分布的节肢动物和线虫病原真菌,在控制节肢动物种群、维持生态系统平衡、医药及生物活性物质开发等方面具有重要价值。自 1892 年 Patouillard 以嗜虫被毛孢 *Hirsutella entomophila* Pat. 为模式种建立以来,被毛孢属的成员不断增加,目前已有约 90 个有效名称^[4]。被毛孢属物种在现代分类系统中的分类地位尚未最终确定,但多数应归到子囊菌门 Ascomycota 肉座菌目 Hypocreales 的线虫草科 Ophiocordycipitaceae^[5]。根据国际命名法规“一菌一名”的要求, Quandt *et al.* (2014) 建议保留线虫草属 *Ophiocordyceps*, 而弃用被毛孢属(注:部分被毛孢是虫草的无性型),因为被毛孢属的模式材料已经无法得到,并且保留被毛孢属比保留线虫草属要做大量的物种名称

处理工作^[6]。被毛孢属多数真菌可能最终被组合到线虫草属,但目前大部分被毛孢属真菌缺乏分子数据,因此,科研人员尚未对被毛孢属的物种名称进行系统处理^[5]。被毛孢属真菌大部分以寄生昆虫为生,少数可寄生线虫、螨、蜘蛛等^[4]。已知能够寄生线虫的被毛孢只有 3 个种,即洛斯里被毛孢 *Hirsutella rhossiliensis* Minter & Brady 1980、明尼苏达被毛孢 *Hirsutella minnesotensis* Chen, Liu & Chen 2000 和线虫被毛孢 *Hirsutella vermicola* Xiang & Liu 2006; 前两种对植物寄生线虫具有较高的侵袭性和致死性,而第 3 种主要寄生食细菌线虫^[7]。虽然被毛孢属包含许多物种,但目前只有 3 个物种(即寄生线虫的 3 种被毛孢)有发表的线粒体基因组^[8-10]。

洛斯里被毛孢是线虫的一种常见内寄生真菌,在自然界分布广泛,寄主范围也较广,已在 16 个属的 23 种土传线虫和 8 个国家被报道,包括美国、荷兰、法国、英国、意大利、新西兰、澳大利亚和中国等^[11]。洛斯里被毛孢显示出较好的生防潜力并被广泛研究^[12-13],但不同菌株在形态、致病性、遗传特征等方面存在差异^[14-16]。Liu & Chen (2001) 利用琼脂平板、室内土壤和温室试验等,从 93 个洛斯里被毛孢菌株中筛选到一个优良菌株 OWVT-1,该菌株对大豆胞囊线虫卵和二龄幼虫的种群密度抑制率分别达到 95% 和 98%^[13,15]。随后,研究人员围绕 OWVT-1 菌株生长、产孢和孢子萌发的营养条件^[17-18]、丝氨酸蛋白酶的分离合纯化和基因克隆^[19-20]、在土壤中的种群动态^[21-22]、对胞囊线虫的功能反应^[23]等做了大量的研究工作。

洛斯里被毛孢 OWVT-1 菌株的基因组测序已经完成,但尚未正式发表^[24]。Wang 等 (2016) 发表了洛斯里被毛孢 USA-87-5 菌株的线粒体基因

组^[9],但我们后来在使用该线粒体基因组序列的过程中发现可能存在序列错误。为了提供洛斯里被毛孢正确的线粒体基因组,本文组装了 OWVT-1 菌株完整的线粒体基因组,确认 USA-87-5 菌株发表的线粒体基因组存在序列错误,对洛斯里被毛孢正确的线粒体基因组序列进行注释并开展比较线粒体基因组学分析。

1 材料和方法

1.1 菌株培养与 DNA 提取

洛斯里被毛孢菌株 USA-87-5 和 OWVT-1 分别分离自美国明尼苏达州卡顿伍德县(Cottonwood)和沃西卡县(Waseca)大豆田中的胞囊线虫 *Heterodera glycines* 上。其中,OWVT-1 菌株分离自一块连作 27 年并出现线虫密度自然衰退的大豆田中^[25]。这两个菌株均由中国科学院微生物研究所刘杏忠研究员课题组提供。将这两个菌株分别接种到铺有玻璃纸的 PDA 培养基上,在 25 °C 黑暗培养 2 周,收集菌丝体后用月桂酸钠法提取总 DNA^[26]。

1.2 OWVT-1 线粒体基因组的组装

利用高通量 Roche 454 和 Illumina Hiseq 2000 结合的测序技术对 OWVT-1 菌株进行基因组测序,得到约 50.6 Mb 的基因组组装数据^[24]。为了得到该菌株的线粒体基因组序列,以明尼苏达被

毛孢的线粒体基因组(NC_027660)为 query (查询序列),通过本地 BLAST 搜索 OWVT-1 组装数据,找到 4 条代表线粒体 DNA 的序列。这 4 条序列均存在多处缺口,因此,根据缺口两端的已知序列设计引物,通过 PCR 扩增和 Sanger 测序填补序列内部存在的缺口。通过与明尼苏达被毛孢线粒体基因组序列比较,判断这 4 条序列的前后顺序关系,然后在相邻序列的端部设计引物,通过 PCR 扩增和 Sanger 测序拼接这 4 条序列,从而得到 OWVT-1 完整的线粒体基因组。

1.3 USA-87-5 和 OWVT-1 线粒体基因组的比较

利用 MAFFT version 7 软件(<http://mafft.cbrc.jp/alignment/server/>)对组装完成的 OWVT-1 的线粒体基因组与公布的 USA-87-5 的线粒体基因组(KU203675)进行比对,比对结果用 Mauve 2.3.1 软件进行展示^[27]。为便于序列比对,将公布的 USA-87-5 线粒体基因组序列的起点调整到与 OWVT-1 一致(即 *rnl* 基因前 15 bp 处)。对 OWVT-1 和 USA-87-5 的线粒体基因组比对存在差异的区域,根据两端的保守序列设计引物(表 1),然后通过 PCR 扩增和琼脂糖凝胶电泳确认两个菌株间是否确实存在差异。为了进一步确认 USA-87-5 的线粒体基因组序列,利用专门的线粒体基因组组装软件 MITObim 1.9^[28]从 USA-87-5 的高通量测序数据中重新组装线粒体基因组。

表 1. 验证 OWVT-1 与 USA-87-5 的序列使用的引物

Table 1. Primers used to verify sequences in USA-87-5 and OWVT-1

Locus	Primer name	Primer sequence (5'→3')	Expected size/kb
VG1	VG1-F1	GTATTTAGTATTGGGTGTTTCGAAT	1.1
	VG1-R1	AGACCGTAGATTTGTCCTTTGT	
VG2	VG2-F1	TTTTGAGCTTATTCCATAGTGC	0.7
	VG2-R1	TAGCATTAGAACCCGATCCATT	
VG3	VG3-F1	TCATTCCTTTCTGCTGTTCCAT	2.9
	VG3-R1	TACAGCCATTTAAACACCCCAT	

1.4 OWVT-1 线粒体基因组的注释

由于公布的 USA-87-5 菌株的线粒体基因组存在序列及注释方面的错误, 以 OWVT-1 菌株为代表重新注释洛斯里被毛孢的线粒体基因组。首先通过 MFannot 软件(<http://megason.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl>)进行自动注释(遗传密码表 4), 然后进行人工校对。通过与近缘真菌中无内含子基因的比较, 确定核心蛋白编码基因和核糖体 RNA 基因的边界、内含子的有无及其序列边界。使用 RNAweasel (<http://megason.bch.umontreal.ca/cgi-bin/RNAweasel/RNAweaselInterface.pl>)确定内含子的类型。通过 tRNAscan-SE 2.0^[29]注释 tRNA 基因。通过 ORF Finder (<http://www.ncbi.nlm.nih.gov>)寻找内含子及基因间区序列中存在的 ORF (开放阅读框), 只考虑长度大于 300 bp 的 ORF。通过在线 BLAST 搜索判断 ORF 的功能。

1.5 OWVT-1 线粒体基因组中的重复序列

通过本地 BLAST 对 OWVT-1 的线粒体基因组与其本身进行序列相似性搜索, 只考虑 E 值小于 10^{-5} 的比对结果。通过 Tandem Repeats Finder (<http://tandem.bu.edu/trf/trf.basic.submit.html>)分析 OWVT-1 线粒体基因组中的串联重复序列(使用默认参数)。

1.6 被毛孢属不同真菌线粒体基因组的比较

为分析被毛孢属真菌线粒体基因组的进化特征, 我们使用本研组装出的洛斯里被毛孢 OWVT-1 菌株的线粒体基因组、已公布的明尼苏达被毛孢 3608 菌株和线虫被毛孢 AS3.7877 菌株的线粒体基因组^[8,10], 比较它们线粒体基因组的大小、基因组成及排列顺序、内含子的插入位置等。使用 LAST 软件(<http://lastweb.cbrc.jp/>)对洛斯里被

毛孢与明尼苏达被毛孢或线虫被毛孢的线粒体基因组进行共线性分析。

1.7 序列提交

洛斯里被毛孢 OWVT-1 菌株的线粒体基因组已提交到 GenBank, 登录号为 MG979071。

2 结果和分析

2.1 洛斯里被毛孢 USA-87-5 和 OWVT-1 线粒体基因组序列的比较

借助 DNA 高通量测序数据并辅以 PCR 扩增和 Sanger 测序, 成功组装出 OWVT-1 的线粒体基因组, 全长 62949 bp, 与 USA-87-5 公布的线粒体基因组(62483 bp, KU203675)相差 466 bp。对 OWVT-1 的线粒体基因组与公布的 USA-87-5 线粒体基因组进行序列比对, 发现 3 处长片段的插入缺失与多处短片段的插入缺失(图 1)。然而, 通过对 3 处长片段插入缺失区域的 PCR 和凝胶电泳检测(图 2), 以及利用 MITObim 软件从高通量测序数据中重新组装 USA-87-5 的线粒体基因组, 都发现序列比对时出现的这些差异是由于公布的 USA-87-5 线粒体基因组存在序列错误造成的。也就是说, 菌株 OWVT-1 与 USA-87-5 的线粒体基因组实际均为 62949 bp, 并且无序列差异。

2.2 洛斯里被毛孢 OWVT-1 线粒体基因组的组成结构

OWVT-1 的线粒体基因组为环状分子, 包括 14 个核心蛋白编码基因(*atp6*、*atp8*、*atp9*、*cob*、*cox1-3*、*nad1-6* 和 *nad4L*)、2 个 rRNA 基因(*rnl* 和 *rns*)、26 个转运氨基酸的 tRNA 基因和另外 7 个独立的 ORF。除 *orf186* 外, 所有其他基因均在正义链上转录(图 3)。这 26 个 tRNA 基因可以转运所有

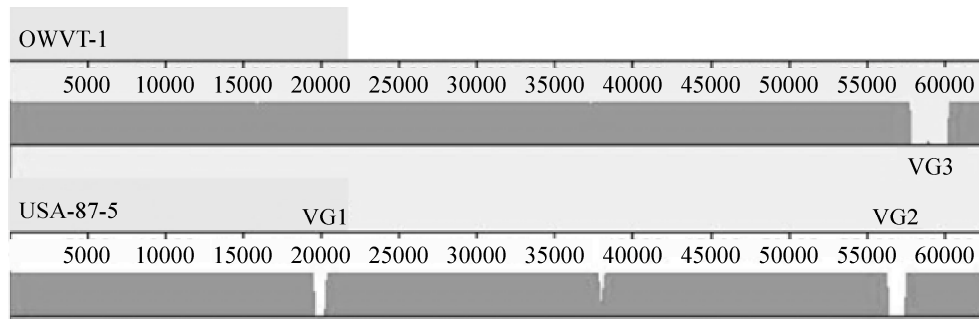


图 1. OWVT-1 与公布的 USA-87-5 线粒体基因组序列的比对

Figure 1. Alignment of the published USA-87-5 mitogenome and OWVT-1 mitogenome assembled in this study. Numbered scale bars indicate distances in base pairs. VG1, VG2, and VG3 indicate where three long insertion/deletion regions located.

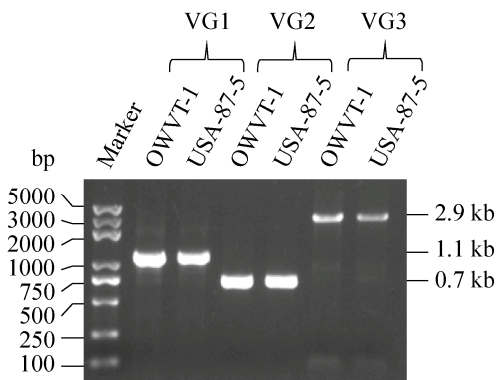


图 2. PCR 验证 OWVT-1 与公布的 USA-87-5 线粒体基因组比对存在差异的 3 处区域

Figure 2. PCR assays of the three long insertion/deletion regions resulting from alignment between the published USA-87-5 mitogenome and OWVT-1 mitogenome assembled in this study. Identical and expected amplicon sizes at each varying region between the two strains imply the presence of erroneous sequences in the published USA-87-5 mitogenome.

20 种标准的氨基酸；除甲硫氨酸(M)、苯丙氨酸(F)、亮氨酸(L)、精氨酸(R)、丝氨酸(S)各有 2-3 个具有相同或不同反义密码子的 tRNA 基因外，其他氨基酸均只有一个 tRNA 基因。tRNA 基因主要聚集在 *nad6/rnl* (5 个)、*rnl/nad2* (13 个) 和 *rns/cox3* (4 个) 这 3 个区域(图 3)。

在洛斯里被毛孢的线粒体基因组中共有 13 个内含子,分布在 *rnl* (2)、*nad2* (1)、*nad5* (2)、*cob* (3)、*cox1* (3)、*nad1* (1) 和 *cox3* (1) 这 7 个基因中。除 *nad5-i1* 的内含子类型不能确定外,其他内含子均为 I 型,但又可细分成若干亚型(表 2)。除 *cox1-i1* 外(上游外显子以 C 结尾,内含子本身以 G 结尾),其他内含子都符合 I 型内含子典型的序列特征(即上游外显子以 T 结尾,内含子本身以 G 结尾)。除两个内含子 *nad5-i2* 和 *cox3-i1* 因序列较短而无 ORF 外,其他内含子中均有至少一个 ORF (编码核糖体蛋白 RPS3, GIY-YIG 或 LAGLIDADG 归巢内切酶),并且 4 个内含子 *rnl-i1*、*nad2-i1*、*cox1-i3* 和 *nad1-i1* 由于移码突变和/或终止密码子突变而显现出序列退化的特征。此外,在 *rnl/nad2* 基因间区还存在多处退化的 GIY-YIG 和 LAGLIDADG 归巢内切酶的编码序列,已不能鉴定出完整的读码框。

通过本地 BLAST 对 OWVT-1 的线粒体基因组与其自身进行相似性搜索,找到 42 个高相似性的比对结果,但没有长片段的序列重复,只在 29-101 bp (平均 58 bp) 的比对长度上有 79.2%-100% (平均 87.8%) 的相似性;最大的 101 bp 的比对结果对应的是 *trnL_2/trnQ* 基因间区的序列与 *trnH/trnM_3*

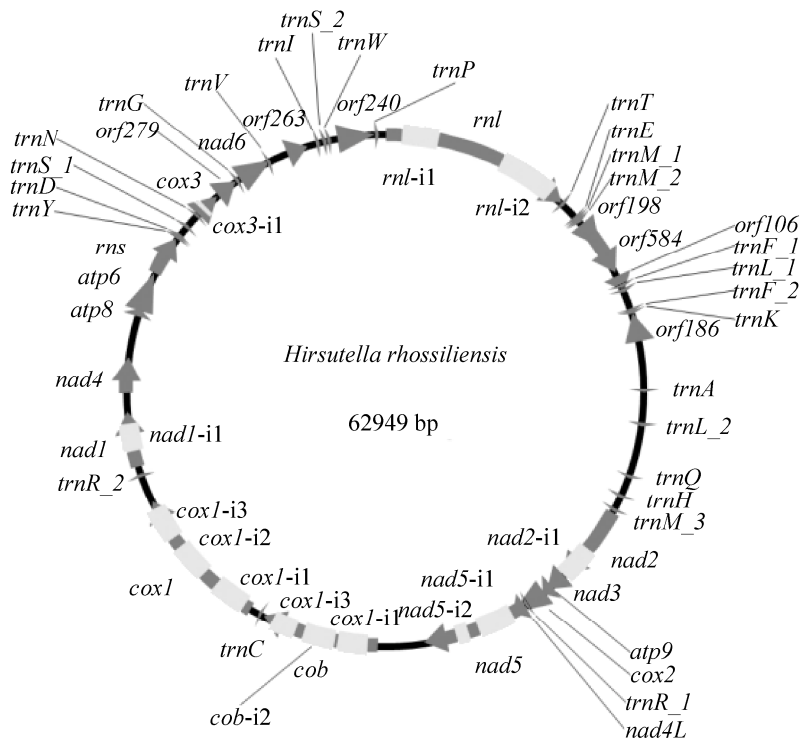


图 3. 洛斯里被毛孢线粒体基因组环状图

Figure 3. Circular map of the *H. rhossiliensis* OWVT-1 mitogenome. Arrows indicate transcription directions of mitochondrial genes. Introns are shown in white boxes.

基因间区的序列。Tandem Repeats Finder 找到 13 个串联重复序列，总长度约 550 bp (占线粒体基因组总长的 0.87%)，6–25 bp 的序列可重复 2–11 次；最长的串联重复序列是在 *nad5/cob* 基因间区，其中 CTGCTGA 以平均 95% 的匹配度串联重复了约 11 次。

2.3 3 种被毛孢属真菌线粒体基因组的比较

洛斯里被毛孢与线虫被毛孢、明尼苏达被毛孢的线粒体基因组具有较强的共线性关系(图 4)。它们的线粒体基因组大小存在差异，受到基因间区的长度、独立 ORF 的数量和内含子数目的影响(表 3)，相关性系数(*r*)分别为 0.98、0.93 和 0.38。在 3 种被毛孢的线粒体基因组中都有 14 个核心蛋白编码基因、2 个 rRNA 基因、25–26 个 tRNA 基因和 3–7 个独立的 ORF (图 5)。根据 BLASTN 的

分析结果，这些独立的 ORF 仅在各自所属真菌的线粒体基因组中存在，没有在其他两种真菌中发现有高相似性的序列。除 *trnA* 基因的位置在不同真菌中发生变化及洛斯里被毛孢的 *trnF* 基因多出 1 个拷贝外，其他所有核心蛋白编码基因、rRNA 基因和 tRNA 基因的排列顺序在 3 种被毛孢中完全相同(图 5)。相邻基因 *nad2* 和 *nad3* 在 3 种被毛孢中都是紧密连接在一起的，没有任何间隔碱基；*nad4L* 和 *nad5* 在 3 种被毛孢中都是重叠 1 个碱基(即 *nad4L* 的终止密码子 TAA 的最后一个碱基 A 充当 *nad5* 起始密码子 ATG 的第一个碱基 A)。

线虫被毛孢、明尼苏达被毛孢和洛斯里被毛孢分别在 5、7 和 7 个基因中含有 7、13 和 13 个内含子；明尼苏达被毛孢和洛斯里被毛孢虽然都在 7 个基因中插入了 13 个内含子，但含有内含子

表 2. 洛斯里被毛孢线粒体基因组中的内含子及其 ORF

Table 2. Introns and intronic ORFs in the *H. rhossiliensis* OWVT-1 mitogenome

Intron	Intronic ORF	Length/bp	Intron type	Notes
<i>rnl</i> -i1		1443	IB(3')	Degeneration
	<i>orf165</i>	498		GIY-YIG endonuclease
<i>rnl</i> -i2		2240	IA	
	<i>orf544</i>	1635		RPS3
<i>nad2</i> -i1		1370	IA	Degeneration
	<i>orf102</i>	309		LAGLIDADG endonuclease
<i>nad5</i> -i1		1403	Unknown	
	<i>orf428</i>	1287		LAGLIDADG endonuclease
<i>nad5</i> -i2		492	ID	
<i>cob</i> -i1		1179	ID	
	<i>orf287</i>	864		GIY-YIG endonuclease
<i>cob</i> -i2		1244	IA	
	<i>orf298</i>	897		LAGLIDAG endonuclease
<i>cob</i> -i3		987	IB	
	<i>orf237</i>	714		LAGLIDADG endonuclease
<i>cox1</i> -i1		1372	IB	
	<i>orf345</i>	1038		LAGLIDADG endonuclease
<i>cox1</i> -i2		1402	IB(5')	
	<i>orf318</i>	957		LAGLIDADG endonuclease
<i>cox1</i> -i3		1364	IB	Degeneration
	<i>orf277</i>	834		GIY-YIG endonuclease
	<i>orf103</i>	312		GIY-YIG endonuclease
<i>nad1</i> -i1		1080	IB	Degeneration
	<i>orf100</i>	303		GIY-YIG endonuclease
<i>cox3</i> -i1		190	IB	

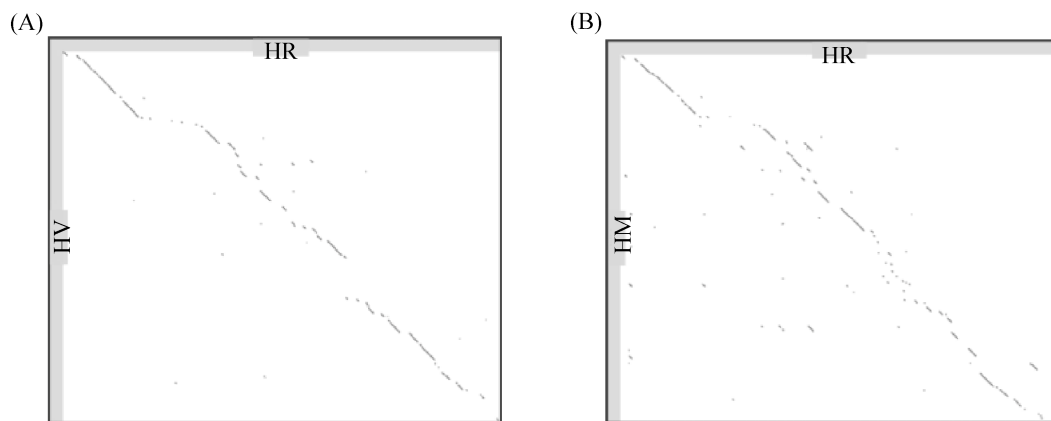


图 4. 3 种被毛孢线粒体基因组序列的共线性关系

Figure 4. Dot plots between different *Hirsutella* mitogenomes. A: between *H. rhossiliensis* (HR) and *H. vermicola* (HV) mitogenomes; B: between *H. rhossiliensis* (HR) and *H. minnesotensis* (HM) mitogenomes.

的不是相同的 7 个基因(图 5)。3 种真菌在 *rnl*、*nad5* 和 *cob* 中都有内含子插入, 并且在这 3 个基因中都有一共同的内含子插入位点。在 *rnl* 共同的内含子插入位点中都有编码核糖体蛋白 RPS3 的 ORF;

在 *nad5* 共同的内含子插入位点中无 ORF 或有未知功能的 ORF; 在 *cob* 共同的内含子插入位点中有未知功能的 ORF 或编码 GIY-YIG 归巢内切酶的 ORF。

表 3. 3 种被毛孢线粒体基因组的比较

Table 3. Comparison among three *Hirsutella* mitogenomes

Item	HV	HM	HR
Accession No.	KY465721	KR139916	MG979071
Strain	AS3.7877	3608	OWVT-1
Size (bp)	53793	52245	62949
No. tRNAs	25	25	26
No. introns	7	13	13
No. ORFs	3	4	7
Length of exons	24.6 kb	24.9 kb	26.0 kb
Length of introns	15.5 kb	17.1 kb	15.8 kb
Length of intergenic regions	13.7 kb	10.2 kb	21.1 kb

HR, *H. rhossiliensis*; HV, *H. vermicola*; HM, *H. minnesotensis*. Intergenic regions are those remaining regions after excluding exons and introns in core protein-encoding genes, rRNA genes, tRNA genes, and free-standing ORFs.

3 讨论

本文报道洛斯里被毛孢 OWVT-1 菌株的线粒体基因组, 并验证公布的 USA-87-5 菌株线粒体基因组(KU203675)中存在的错误。这两个菌株的分离地点均位于美国明尼苏达州, 但属于两个不同的县, 相距约 144 公里。我们发现两个菌株的线粒体基因组序列实际上是完全相同的。公布的 USA-87-5 菌株的线粒体基因组(KU203675)中存在几处序列错误并引起一些注释错误, 包括在 *nad4L* 终止处多出一个碱基 A, 因而改变了其与 *nad5* 的结合方式(二者本应重叠 1 个碱基, 但多出的碱基使二者直接相连); 在 *trnQ* 和 *trnM_3* 间一段 0.7 kb 的序列串联重复了 2 次(即 VG1 处), 使得 *trnH* 基因多出 1 个拷贝; 在 *trnS_1* 和 *cox3-E2* 间一段 1.0 kb 的序列重复了 2 次(即 VG2 处), 使得无法准确注释出 *cox3* 基因的起始位置; 在 *nad6* 和 *trnI* 间丢失了一段 2.4 kb 的序列(即 VG3 处, 也是公布序列的结尾处), 因此漏掉了 *trnV* 和 *orf263* 两个基因; 在 *cox1* 基因中也存在序列错误等。在发表的 USA-87-5 菌株线粒体基因组序列正确的地方也存在一些注释错误, 如 *nad2* 基因中的内含子没有被鉴定出来, *rnl* 基因中内含子的鉴定不准确等^[9]。

我们对洛斯里被毛孢正确的线粒体基因组序列进行注释, 并与明尼苏达被毛孢和线虫被毛孢的线粒体基因组进行比较。3 种被毛孢线粒体基因组的共线性较好, 而且常见基因的顺序也比较保守, 但洛斯里被毛孢比明尼苏达被毛孢和线虫被毛孢的线粒体基因组大(相差约 10 kb)。我们发现基因间区的长度、独立 ORF 的数量和内含子数目均会影响被毛孢线粒体基因组的大小, 但基因间区长度和独立 ORF 数量的影响程度较大(表 3)。例如, *rnl/nad2* 基因之间的序列(除去其中的 tRNA 和独立 ORF 本身的序列)在线虫被毛孢、明尼苏达被毛孢和洛斯里被毛孢中分别为 2.4、2.2 和 9.2 kb。Xiao *et al.* (2017)通过对酵母菌种间和种内多个线粒体基因组的分析, 发现引起物种间和物种内线粒体基因组大小变化的主要因素是不同的, 物种间差异主要是由基因间区序列引起的^[30]。这与本文的结果相一致。

洛斯里被毛孢在 7 个基因中共插入 13 个内含子, 其中 4 个内含子 *rnl-i1*、*nad2-i1*、*cox1-i3* 和 *nad1-i1* 由于移码突变和/或终止密码子突变而显现出序列退化的特征; 2 个内含子 *nad5-i2* 和 *cox3-i1* 序列较短, 分别为 492 bp 和 190 bp, 小于明尼苏达被毛孢和线虫被毛孢中线粒体内含子的

HV	HM	HR
<i>rnl</i> (1)	<i>rnl</i> (1)	<i>rnl</i> (2)
<i>trnT</i>	<i>trnT</i>	<i>trnT</i>
<i>trnE</i>	<i>trnE</i>	<i>trnE</i>
<i>trnM_1</i>	<i>trnM_1</i>	<i>trnM_1</i>
<i>trnM_2</i>	<i>trnM_2</i>	<i>trnM_2</i>
		<i>orf198</i>
		<i>orf584</i>
		<i>orf106</i>
		<i>trnF_1</i>
<i>trnL_1</i>	<i>trnL_1</i>	<i>trnL_1</i>
<i>orf827</i>		
	<i>orf798</i>	
	<i>trnA</i>	
<i>trnF</i>	<i>trnF</i>	<i>trnF_2</i>
<i>trnK</i>	<i>trnK</i>	<i>trnK</i>
		<i>orf186</i>
<i>trnA</i>		<i>trnA</i>
<i>trnL_2</i>	<i>trnL_2</i>	<i>trnL_2</i>
<i>trnQ</i>	<i>trnQ</i>	<i>trnQ</i>
<i>trnH</i>	<i>trnH</i>	<i>trnH</i>
<i>trnM_3</i>	<i>trnM_3</i>	<i>trnM_3</i>
<i>nad2</i>	<i>nad2</i>	<i>nad2</i> (1)
<i>nad3</i>	<i>nad3</i>	<i>nad3</i>
<i>atp9</i>	<i>atp9</i>	<i>atp9</i>
<i>cox2</i> (1)	<i>cox2</i>	<i>cox2</i>
<i>trnR_1</i>	<i>trnR_1</i>	<i>trnR_1</i>
<i>nad4L</i>	<i>nad4L</i> (1)	<i>nad4L</i>
<i>nad5</i> (1)	<i>nad5</i> (1)	<i>nad5</i> (2)
<i>cob</i> (1)	<i>cob</i> (2)	<i>cob</i> (3)
<i>trnC</i>	<i>trnC</i>	<i>trnC</i>
<i>cox1</i> (3)	<i>cox1</i> (6)	<i>cox1</i> (3)
<i>trnR_2</i>	<i>trnR_2</i>	<i>trnR_2</i>
<i>nad1</i>	<i>nad1</i>	<i>nad1</i> (1)
<i>nad4</i>	<i>nad4</i> (1)	<i>nad4</i>
	<i>orf121</i>	
<i>atp8</i>	<i>atp8</i>	<i>atp8</i>
<i>atp6</i>	<i>atp6</i>	<i>atp6</i>
	<i>orf263</i>	
	<i>orf276</i>	
<i>rns</i>	<i>rns</i>	<i>rns</i>
<i>trnY</i>	<i>trnY</i>	<i>trnY</i>
<i>trnD</i>	<i>trnD</i>	<i>trnD</i>
<i>trnS_1</i>	<i>trnS_1</i>	<i>trnS_1</i>
<i>trnN</i>	<i>trnN</i>	<i>trnN</i>
<i>cox3</i>	<i>cox3</i> (1)	<i>cox3</i> (1)
		<i>orf279</i>
<i>trnG</i>	<i>trnG</i>	<i>trnG</i>
<i>nad6</i>	<i>nad6</i>	<i>nad6</i>
<i>trnV</i>	<i>trnV</i>	<i>trnV</i>
		<i>orf263</i>
<i>trnI</i>	<i>trnI</i>	<i>trnI</i>
<i>trnS_2</i>	<i>trnS_2</i>	<i>trnS_2</i>
<i>trnW</i>	<i>trnW</i>	<i>trnW</i>
		<i>orf240</i>
	<i>orf266</i>	
	<i>orf300</i>	
<i>trnP</i>	<i>trnP</i>	<i>trnP</i>

图 5. 3 种被毛孢线粒体基因排列顺序的比较

Figure 5. Comparison of gene orders among three *Hirsutella* mitogenomes. The number of introns is given in parentheses after gene names if present. HR, *H. rhossiliensis*; HV, *H. vermicola*; HM, *H. minnesotensis*.

长度, 可能也是序列退化的证据。实际上, 线虫被毛孢 7 个内含子的平均长度为 2.2 kb (范围 617–5789 bp)^[8]; 明尼苏达被毛孢 13 个内含子的平均长度为 1.3 kb (范围 893–1818 bp)^[10]; 洛斯里被毛孢 13 个内含子的平均长度为 1.2 kb (范围 190–2240 bp; 表 2)。此外, 我们在洛斯里被毛孢 *rnl/nad2* 基因之间的序列中发现多处退化的 GIY-YIG 和 LAGLIDADG 归巢内切酶的编码序列, 但已不能鉴定出完整的读码框。这些序列退化的证据意味着, 洛斯里被毛孢祖先线粒体基因组可能比目前已知的 62949 bp 大。为了详细了解该菌的进化过程, 我们目前正在利用不同地理来源的洛斯里被毛孢菌株开展种内比较线粒体基因组学研究。

致谢

中国科学院上海生命科学研究院植物生理生态研究所赖屹玲和中国科学院动物研究所王牛牛分别参与完成 OWVT-1 和 USA-87-5 菌株的基因组测序工作, 在此表示感谢。

参考文献

- [1] Bullerwell CE. Organelle genetics: evolution of organelle genomes and gene expression. Berlin, Heidelberg: Springer, 2012.
- [2] Desmond E, Brochier-Armanet C, Forterre P, Gribaldo S. On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes. *Research in Microbiology*, 2011, 162(1): 53–70.
- [3] Zhang YJ, Zhao YX, Zhang S, Chen L, Liu XZ. Reanalysis of the mitochondrial genome of the pneumocandin-producing fungus *Glarea lozoyensis*. *Acta Microbiologica Sinica*, 2017, 57(5): 724–736.
- [4] Ciancio A, Colagiero M, Rosso LC, Gutierrez SNM, Grasso G. Phylogeny and morphology of *Hirsutella tunicata* sp. nov.

- (Ophiocordycipitaceae), a novel mite parasite from Peru. *Mycoscience*, 2013, 54(5): 378–386.
- [5] Simmons DR, Kepler RM, Rehner SA, Groden E. Phylogeny of *Hirsutella* species (Ophiocordycipitaceae) from the USA: remedying the paucity of *Hirsutella* sequence data. *IMA Fungus*, 2015, 6(2): 345–356.
- [6] Quandt CA, Kepler RM, Gams W, Araújo JPM, Ban S, Evans HC, Hughes D, Humber R, Hywel-Jones N, Li ZZ, Luangsa-Ard JJ, Rehner SA, Sanjuan T, Sato H, Shrestha B, Sung GH, Yao YJ, Zare R, Spatafora JW. Phylogenetic-based nomenclatural proposals for Ophiocordycipitaceae (Hypocreales) with new combinations in *Tolypocladium*. *IMA Fungus*, 2014, 5(1): 121–134.
- [7] Xiang MC, Yang EC, Xiao QM, Liu XZ, Shen SY. *Hirsutella vermicola* sp. nov., a new species parasitizing bacteria-feeding nematodes. *Fungal Diversity*, 2006, 22: 255–265.
- [8] Zhang YJ, Zhang HY, Liu XZ, Zhang S. Mitochondrial genome of the nematode endoparasitic fungus *Hirsutella vermicola* reveals a high level of synteny in the family Ophiocordycipitaceae. *Applied Microbiology and Biotechnology*, 2017, 101(8): 3295–3304.
- [9] Wang NN, Zhang YJ, Hussain M, Li K, Xiang MC, Liu XZ. The mitochondrial genome of the nematode endoparasitic fungus *Hirsutella rhossiliensis*. *Mitochondrial DNA Part B, Resources*, 2016, 1(1): 114–115.
- [10] Zhang YJ, Zhang S, Liu XZ. The complete mitochondrial genome of the nematode endoparasitic fungus *Hirsutella minnesotensis*. *Mitochondrial DNA Part A, DNA Mapping, Sequencing, and Analysis*, 2016, 27(4): 2693–2694.
- [11] Zhao XH, Xu YL. Advance on *Hirsutella* sp. of biological control. *System Sciences and Comprehensive Studies in Agriculture*, 2011, 27(3): 376–381. (in Chinese)
赵晓晖, 许艳丽. 生防真菌被毛孢的研究进展. *农业系统科学与综合研究*, 2011, 27(3): 376–381.
- [12] Liu SF, Chen SY. Efficacy of the fungi *Hirsutella minnesotensis* and *H. rhossiliensis* from liquid culture for control of the soybean cyst nematode *Heterodera glycines*. *Nematology*, 2005, 7(1): 149–157.
- [13] Chen SY, Liu XZ. Control of the soybean cyst nematode by the fungi *Hirsutella rhossiliensis* and *Hirsutella minnesotensis* in greenhouse studies. *Biological Control*, 2005, 32(2): 208–219.
- [14] Wang NN, Zhang YJ, Jiang XZ, Shu C, Hamid MI, Hussain M, Chen SY, Xu JP, Xiang MC, Liu XZ. Population genetics of *Hirsutella rhossiliensis*, a dominant parasite of cyst nematode juveniles on a continental scale. *Applied and Environmental Microbiology*, 2016, 82(21): 6317–6325.
- [15] Liu XZ, Chen SY. Screening isolates of *Hirsutella* species for biocontrol of *Heterodera glycines*. *Biocontrol Science and Technology*, 2001, 11(1): 151–160.
- [16] Xiang MC, Yang XH, Wang ZX, Liu XZ, Chen SY, Xiao QM. Variability of morphology, parasitism, and nucleotide sequences among isolates and species of nematophagous *Hirsutella*. *Biological Control*, 2007, 41(1): 110–119.
- [17] Liu XZ, Chen SY. Nutritional requirements of the nematophagous fungus *Hirsutella rhossiliensis*. *Biocontrol Science and Technology*, 2002, 12(3): 381–393.
- [18] Sun MH, Liu XZ. Carbon requirements of some nematophagous, entomopathogenic and mycoparasitic Hyphomycetes as fungal biocontrol agents. *Mycopathologia*, 2006, 161(5): 295–305.
- [19] Wang B, Liu XY, Wu WP, Liu XZ, Li SD. Purification, characterization, and gene cloning of an alkaline serine protease from a highly virulent strain of the nematode-endoparasitic fungus *Hirsutella rhossiliensis*. *Microbiological Research*, 2009, 164(6): 665–673.
- [20] Wang B, Wu WP, Liu XZ. Purification and characterization of a neutral serine protease with nematocidal activity from *Hirsutella rhossiliensis*. *Mycopathologia*, 2007, 163(3): 169–176.
- [21] Zhang LM, Yang EC, Xiang MC, Liu XZ, Chen SY. Population dynamics and biocontrol efficacy of the nematophagous fungus *Hirsutella rhossiliensis* as affected by stage of the soybean cyst nematode. *Biological Control*, 2008, 47(2): 244–249.
- [22] Zhang LM, Liu XZ, Zhu SF, Chen SY. Detection of the nematophagous fungus *Hirsutella rhossiliensis* in soil by real-time PCR and parasitism bioassay. *Biological Control*, 2006, 36(3): 316–323.
- [23] Shu C, Lai YL, Yang EC, Chen SY, Xiang MC, Liu XZ. Functional response of the fungus *Hirsutella rhossiliensis* to the nematode, *Heterodera glycines*. *Science China Life Sciences*, 2015, 58(7): 704–712.
- [24] 赖屹玲. 被毛孢寄生大豆孢囊线虫的分子机制以及杀线虫化合物高通量筛选方法的建立. 北京: 中国科学院大学博士学位论文, 2014.
- [25] Chen SY. Infection of *Heterodera glycines* by *Hirsutella rhossiliensis* in a Minnesota soybean field. *Journal of Nematology*, 1997, 29: 573.
- [26] Cai WJ, Xu DB, Lan X, Xie HH, Wei JG. A new method for the extraction of fungal genomic DNA. *Agricultural Research and Application*, 2014, (3): 1–5. (in Chinese)

- 蔡文娇, 徐大彬, 蓝霞, 谢红辉, 韦继光. 一种提取真菌基因组 DNA 的新方法. *农业研究与应用*, 2014, (3): 1–5.
- [27] Darling AE, Mau B, Perna NT. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*, 2010, 5(6): e11147.
- [28] Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research*, 2013, 41(13): e129.
- [29] Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*, 2016, 44(W1): W54–W57.
- [30] Xiao SJ, Nguyen DT, Wu BJ, Hao WL. Genetic drift and indel mutation in the evolution of yeast mitochondrial genome size. *Genome Biology and Evolution*, 2017, 9(11): 3088–3099.

Reanalysis of the mitochondrial genome of the nematophagous fungus *Hirsutella rhossiliensis*

Qianqian Yan¹, Xingzhong Liu², Yongjie Zhang^{1*}

¹ School of Life Sciences, Shanxi University, Taiyuan 030006, Shanxi Province, China

² State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

Abstract: [Objective] To identify the mitochondrial genome (mitogenome) of *Hirsutella rhossiliensis* OWVT-1 strain, verify erroneous sequences present in the published USA-87-5 mitogenome, and perform comparative mitogenomic analyses of three *Hirsutella* species. [Methods] The OWVT-1 mitogenome was assembled based on high-throughput DNA sequencing data and Sanger sequencing. PCR assays were performed to verify 3 long varying regions observed during alignment between the published USA-87-5 mitogenome and OWVT-1 mitogenome assembled in this study. Bioinformatics analyses were used to annotate the mitogenome sequence of OWVT-1. [Results] We detected several sequence errors in the published USA-87-5 mitogenome, including long sequence insertions/deletions and small indels. The mitogenomes of OWVT-1 and USA-87-5 were identical without any nucleotide difference. The complete mitogenome of *H. rhossiliensis* was 62949 bp in length, with 13 introns in seven genes. Several introns and intergenic regions seemed to have degenerated. The mitogenomes of *H. rhossiliensis*, *H. vermicola*, and *H. minnesotensis* showed a high synteny. Except free-standing ORFs, the gene order of core protein-encoding genes, rRNA genes, and tRNA genes were highly conserved among the three *Hirsutella* mitogenomes. The length at intergenic regions was a main factor affecting mitogenome sizes of different *Hirsutella* species. [Conclusion] There were erroneous sequences in the published USA-87-5 mitogenome. We reported the authentic mitogenome sequence of *H. rhossiliensis* using the highly virulent strain OWVT-1.

Keywords: *Hirsutella rhossiliensis*, mitochondrial genome, correction, annotation, comparative mitogenomics

(本文责编: 张晓丽)

Supported by the Natural Science Foundation of Shanxi Province (201601D011065), by the Fund Program for the Scientific Activities of Selected Returned Overseas Professionals in Shanxi Province and by the Research Project Supported by Shanxi Scholarship Council of China (2017-015)

*Corresponding authors. E-mail: zhangyj2008@sxu.edu.cn

Received: 29 March 2018; Revised: 25 April 2018; Published online: 12 May 2018