



## 结合三代测序与二代测序技术揭示蜜蜂球囊菌孢子转录组的复杂性

孙明会<sup>1#</sup>, 刘佳美<sup>1#</sup>, 王思懿<sup>1</sup>, 朱乐冉<sup>1</sup>, 王紫馨<sup>1</sup>, 叶亚萍<sup>1</sup>, 钱加珺<sup>1</sup>, 顾小雨<sup>1</sup>, 徐细建<sup>3</sup>, 陈大福<sup>1,2</sup>, 郭睿<sup>1,2\*</sup>, 徐国钧<sup>1,2\*</sup>

- 1 福建农林大学动物科学学院(蜂学学院), 福建 福州 350002
- 2 福建农林大学蜂疗研究所, 福建 福州 350002
- 3 江西省养蜂研究所, 江西 南昌 330000

孙明会, 刘佳美, 王思懿, 朱乐冉, 王紫馨, 叶亚萍, 钱加珺, 顾小雨, 徐细建, 陈大福, 郭睿, 徐国钧. 结合三代测序与二代测序技术揭示蜜蜂球囊菌孢子转录组的复杂性. 微生物学报, 2022, 62(8): 2981–2994.

Sun Minghui, Liu Jiamei, Wang Siyi, Zhu Leran, Wang Zixin, Ye Yaping, Qian Jiajun, Gu Xiaoyu, Xu Xijian, Chen Dafu, Guo Rui, Xu Guojun. Unraveling the complexity of transcriptome in *Ascospaera apis* spore: based on third-generation and next-generation sequencing. *Acta Microbiologica Sinica*, 2022, 62(8): 2981–2994.

**摘要:**【目的】蜜蜂球囊菌(*Ascospaera apis*)是一种专性侵染蜜蜂幼虫的致死性真菌病原。本研究旨在利用 PacBio 单分子实时(single molecule real-time, SMRT)测序技术对蜜蜂球囊菌孢子(AaS)中基因的可变剪切(alternative splicing, AS)和可变多聚腺苷酸化(alternative polyadenylation, APA)以及长链非编码 RNA (long non-coding RNA, lncRNA)进行鉴定和分析, 进而揭示蜜蜂球囊菌孢子中转录组的复杂性。【方法】采用 Suppa 软件对蜜蜂球囊菌孢子中基因的 AS 事件进行鉴定。通过 RT-PCR 对不同类型的 AS 事件进行验证。采用 TAPIS pipeline 对蜜蜂球囊菌孢子基因的 APA 位点进行鉴定。利用 MEME 软件分析孢子全长转录本的 poly(A)剪接位点上游 50 bp 的序列特征并鉴定 motif。联用 CPC 和 CNCI 软件和比对 Swiss-prot 数据库的方法预测 lncRNA, 取三者的交

**基金项目:** 国家自然科学基金(31702190); 财政部和农业农村部: 国家现代农业产业技术体系(CARS-44-KXJ7); 福建农林大学硕士生导师团队项目(郭睿); 江西省蜜蜂生物学与饲养重点实验室开放基金(JXKLHBB-2020-04); 福建省病原真菌与真菌毒素重点实验室开放课题(郭睿)

Supported by the National Natural Science Foundation of China (31702190), by the Earmarked Fund for China Agriculture Research System (CARS-44-KXJ7); by the Master Supervisor Team Fund of Fujian Agriculture and Forestry University (Guo Rui); by the Open fund for Jiangxi Province Key Laboratory of Honeybee Biology and Beekeeping (JXKLHBB-2020-04) and by the Open Project of Key Laboratory of Pathogenic Fungus and Fungal Toxin of Fujian Province (Guo Rui)

<sup>#</sup>These authors contributed equally to this work.

\*Corresponding author. Tel/Fax: +86-591-87640197; E-mail: GUO Rui, ruiguo@fafu.edu.cn, XU Guojun, 000q060012@fafu.edu.cn

Received: 19 November 2021; Revised: 8 December 2021; Published online: 22 March 2022

集作为高可信度的 lncRNA 集合。进一步比较 lncRNA 和 mRNA 的转录本长度, 外显子数量与长度, 内含子长度, GC 含量, AS 事件数量。【结果】在蜜蜂球囊菌孢子中共鉴定到 2 609 次 AS 事件, 包括 1 227 次(47.03%)内含子保留(retained intron, RI), 842 次(32.27%)可变 3'剪切(alternative 3' splice sites, A3), 415 次(15.91%)可变 5'剪切(alternative 5' splice sites, A5), 85 次(3.26%)可变起始外显子(alternative first exons, AF), 35 次(1.34%)外显子跳跃(skipping exon, SE), 4 次(0.15%)可变末端外显子(alternative last exons, AL)和 1 次(0.04%)互斥外显子(mutually exclusive exons, MX)。通过 RT-PCR 证实了不同类型 AS 事件(RI、A5、SE 和 A3)的可靠性。共鉴定出 5 552 个基因含 APA 位点, 其中含有 >5 个 APA 位点的基因数量最多, 达到 2 197 个, 其次为含有 1 个 APA 位点的基因, 数量为 1 149 个; 此外, 含有 2、3、4 和 5 个 APA 位点的基因数量分别为 782、596、477 和 351 个。蜜蜂球囊菌孢子中全长转录本的上下游序列具有明显的碱基倾向性, U 和 A 分别在转录本的下游和上游富集。共鉴定到 953 条 lncRNA。与 mRNA 相比, 上述 lncRNA 的外显子数量更少且长度更短, 内含子长度更短, 转录本长度更短, GC 含量更低, AS 事件数量更少。【结论】本研究全面解析了 AaS 中的 AS、APA 和 lncRNA, 研究结果丰富了蜜蜂球囊菌的基础生物学信息, 深入揭示了蜜蜂球囊菌孢子中转录组的复杂性, 并为进一步探究不同剪接异构体在病原孢子和病原侵染中的分子功能打下了基础。

**关键词:** 蜜蜂球囊菌; PacBio 测序; 可变剪接; 可变多聚腺苷酸化; 长链非编码 RNA; 剪接异构体

## Unraveling the complexity of transcriptome in *Ascosphaera apis* spore: based on third-generation and next-generation sequencing

SUN Minghui<sup>1#</sup>, LIU Jiamei<sup>1#</sup>, WANG Siyi<sup>1</sup>, ZHU Leran<sup>1</sup>, WANG Zixin<sup>1</sup>, YE Yaping<sup>1</sup>, QIAN Jiajun<sup>1</sup>, GU Xiaoyu<sup>1</sup>, XU Xijian<sup>3</sup>, CHEN Dafu<sup>1,2</sup>, GUO Rui<sup>1,2\*</sup>, XU Guojun<sup>1,2\*</sup>

1 College of Animal Sciences (College of Bee Science), Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China

2 Apitherapy Research Institute, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China

3 Apicultural Research Institute of Jiangxi Province, Nanchang 330000, Jiangxi, China

**Abstract: [Objective]** *Ascosphaera apis* is a lethal fungal pathogen of honey bee larvae. This study aims to identify and analyze alternative splicing (AS) and alternative polyadenylation (APA) of genes and long non-coding RNAs (lncRNAs) in *A. apis* spore (AaS) by PacBio single molecular real-time (SMRT) sequencing, further unveiling the complexity of AaS's transcriptome. **[Methods]** AS events of genes in AaS were identified with Suppa, and AS events of various types were confirmed based on RT-PCR. TAPIS pipeline was utilized to explore APA sites of AaS genes. MEME was employed to investigate statistics of sequences at 50 bp upstream of poly(A) splicing sites and identify the motifs. The lncRNAs were predicted based on CPC, CNCI, and Swiss-prot database, and the intersection was regarded as lncRNA dataset. Then the transcript length, exon number and length, intron length, GC

content, and AS event number were compared between lncRNAs and mRNAs. **[Results]** A total of 2 609 AS events were identified in AaS, including 1 227 (47.03%) retained introns (RI), 842 (32.27%) alternative 3' splice sites (A3), 415(15.91%) alternative 5' splice sites (A5), 85 (3.26%) alternative first exons (AF), 35 (1.34%) skipping exons (SE), 4 (0.15%) alternative last exons (AL), and 1 (0.04%) mutually exclusive exon (MX). The reliability of various AS events such as RI, A5, SE, and A3 was validated via RT-PCR. Additionally, 5 552 genes containing APA sites were identified, among which genes with more than 5 APA sites were the most abundant (2 197), followed by genes with 1 APA site (1 149). The numbers of genes with 2, 3, 4, and 5 APA sites were 782, 596, 477, and 351, respectively. Moreover, upstream and downstream sequences of full-length transcripts in AaS had apparent base bias, with U and A respectively enriching at upstream and downstream. In total, 953 lncRNAs were identified, including 247 bidirectional lncRNAs, 171 long intergenic RNAs, 154 anti-sense lncRNAs, 141 sense lncRNAs and 9 intronic lncRNAs. Compared with mRNAs, these lncRNAs had fewer and shorter exons, shorter introns, shorter transcripts, lower GC content and fewer AS events. **[Conclusion]** In this study, we analyzed the AS, APA, and lncRNA in AaS. The findings enrich the basic biological information of *A. apis*, unravel the complexity of transcriptome in AaS, and lay a foundation for further exploration of the molecular function of various splicing isoforms in pathogenic spores and during pathogen infection.

**Keywords:** *Ascosphaera apis*; PacBio sequencing; alternative splicing; alternative polyadenylation; long non-coding RNA; splicing isoform

蜜蜂球囊菌(*Ascosphaera apis*)是一种专性侵染蜜蜂幼虫的真菌病原,其导致的白垩病在养蜂生产中反复发生,可引起成年蜜蜂数量和蜂群群势的急剧下降,进而影响蜂产品的产量和质量,给养蜂业造成较大损失<sup>[1]</sup>。蜜蜂球囊菌的菌丝具有正负两种性别,通过异宗配合形成黑色孢子囊,外被坚韧的闭囊壳,内含一定数量的孢子球,孢子球由许多成熟孢子组成<sup>[2]</sup>。含蜜蜂球囊菌孢子的食物被蜜蜂宿主经口摄入后进入中肠,此时孢子低水平萌发、菌丝少量生长,至预蛹期后肠腔中大量生长的菌丝首先从尾部穿出,进而蔓延包裹整个虫体,形成白垩状虫尸<sup>[3]</sup>。真菌菌丝在增殖阶段大量生长,期间伴随着活跃的新陈代谢活动,而孢子是一种休眠态,仅维持低水平代谢和必要的生命活动<sup>[4]</sup>。此前,人们对真菌孢子内的生命活动(如基因转录和翻译)了解有限,技术手段是一个重

要的限制因素。以 Illumina 测序为代表的二代测序技术在过去十余年中已广泛应用于毛竹基腐病菌(*Arthrinium phaeospermum*)<sup>[5]</sup>、须发癣菌(*Trichophyton mentagrophytes*)<sup>[6]</sup>和灰花纹鹅膏菌(*Amanita fuligineoides*)<sup>[7]</sup>等真菌的转录组研究。相对于蜜蜂宿主,蜜蜂真菌病原的转录组研究较为滞后<sup>[8-9]</sup>。前人关于蜜蜂球囊菌的转录组研究仅有两例<sup>[10-11]</sup>。笔者团队前期通过分析蜜蜂球囊菌侵染意大利蜜蜂幼虫和中华蜜蜂幼虫过程的差异表达基因探讨了病原的侵染机制<sup>[12]</sup>;此外还通过 Illumina 测序对蜜蜂球囊菌菌丝和孢子中的 miRNA、circRNA 和长链非编码 RNA (long non-coding RNA, lncRNA)进行了系统的鉴定和比较分析<sup>[13-15]</sup>。

近年来,以 PacBio 单分子实时(single molecule real-time, SMRT)测序为代表的三代测序技术方兴未艾,因具有超长读长的特点而快速发展,

在全长转录本的鉴定及表达量和结构分析方面表现出巨大优势,已成功应用于动植物的全长转录组研究<sup>[16-17]</sup>。人们也利用 PacBio SMRT 测序技术对广东虫草(*Cordyceps guangdongensis*)<sup>[18]</sup>和哈氏疟原虫(*Plasmodium halstedii*)<sup>[19]</sup>等真菌的孢子进行了全长转录组研究,但总体上仍比较有限。对于蜜蜂真菌病原,相关研究更为匮乏。前期研究中,笔者团队运用纳米孔(Nanopore)长读段测序技术构建和注释了蜜蜂球囊菌的首个高质量全长转录组;基于高质量的测序数据完善了蜜蜂球囊菌参考基因组中的基因和转录本注释信息,并鉴定到较多新基因和新转录本<sup>[20-21]</sup>;此外,系统解析了蜜蜂球囊菌菌丝和孢子中基因的可变剪切(alternative splicing, AS)与可变腺苷酸化(alternative polyadenylation, APA)<sup>[22]</sup>。前期研究中,笔者团队利用 PacBio SMRT 测序技术对蜜蜂球囊菌的纯净菌丝样品进行测序,构建和注释了蜜蜂球囊菌的首个高质量全长转录组,并解析了基因的 AS 与 APA<sup>[23]</sup>。

鉴于目前对于真菌孢子内生命活动的了解有限及相关全长转录组研究较为缺乏,本研究在实验室条件下纯化得到蜜蜂球囊菌的纯净孢子样品,然后利用 PacBio SMRT 技术对构建的 cDNA 文库进行测序,基于高质量的测序数据鉴定全长转录本,并对基因的 AS 和 APA 以及 lncRNA 进行全面解析。研究结果可为蜜蜂球囊菌孢子中存在复杂生命活动提供直接证据,丰富蜜蜂球囊菌的基础生物学信息,深入揭示蜜蜂球囊菌孢子中转录组的复杂性,并为进一步探究不同剪接异构体(isoform)在病原孢子和病原侵染中的分子功能提供基础。

## 1 材料与方法

### 1.1 蜜蜂球囊菌孢子样品制备

蜜蜂球囊菌菌株由福建农林大学动物科学

学院(蜂学学院)蜜蜂保护实验室分离和保存<sup>[24]</sup>。按照笔者实验室已建立的技术流程<sup>[14]</sup>制备蜜蜂球囊菌的纯净孢子,将制备好的孢子样品(AaS)(约 100 mg)放入 RNA-free 的 EP 管,液氮速冻后保存于-80 °C 超低温冰箱。

### 1.2 cDNA 文库构建和 PacBio SMRT 测序

参照 Li 等<sup>[25]</sup>的方法进行 PacBio 文库构建和测序,利用 TRizol 试剂(Thermo Fisher 公司)提取 AaS 的总 RNA,通过 SMARTer™ PCR cDNA 合成试剂盒(TaKaRa 公司)合成高质量的 FL cDNAs。使用 BluePippin™ 系统(Select Science 公司)选择>4 kb 的 cDNA,进而通过 PCR 扩增得到双链 cDNA;再利用 AMPure PB 磁珠对 PCR 扩增产物进行纯化,然后进行 DNA 损伤修复、末端修复及测序接头连接;采用 DNA/Polymerase 结合试剂盒 P6 和 v2 引物将 SMRTbell 模板与聚合酶结合。委托广州基迪奥生物科技有限公司利用 PacBio Sequel 系统对构建好的 cDNA 文库进行测序。测序数据已上传到 NCBI SRA 数据库, BioProject 号: PRJNA557811。

### 1.3 Illumina 测序和数据质控

参照 Chen 等的方法<sup>[23]</sup>对 AaS 进行总 RNA 提取, cDNA 第 1 链合成, cDNA 第 2 链合成, 双链 cDNA 纯化, 目的片段纯化, PCR 扩增。委托广州基迪奥生物科技有限公司对建好的 cDNA 文库进行双端(paired-end)测序,测序平台为 Illumina HiSeq™ 4000。测序数据已上传到 NCBI SRA 数据库, BioProject 号: PRJ-NA560452。

### 1.4 PacBio SMRT 测序数据质控和参考基因组比对

对于 PacBio Sequel 下机数据,使用 SMRT Link v5.0.1 对 cDNA 文库测序下机的原始数据进行分析和转录本聚类<sup>[26]</sup>,根据 full pass=1 和 minimum read score=0.65 的阈值从下机的

subreads 中提取 CCS (circular consensus sequence)序列;然后根据 CCS 序列是否含引物序列和 poly(A)尾巴信号将上述 CCS 序列聚类为短读段、嵌合体、非全长读段(non-full-length reads, nFL reads)、全长非嵌合读段(full-length non-chimeric reads, FLNC reads);进一步滤除短于 50 bp 的 reads, 并利用 ICE 软件对 FLNC reads 进行聚类, 从而得到一致性全长转录本。对于 Illumina HiSeq™ 4000 下机数据, 按照笔者团队前期的方法<sup>[27]</sup>数据质控。基于质控后 clean reads, 利用 LoRDEC tool (version 0.8)对低质量的 isoform 进行校正, 从而进一步提高一致性全长转录本的准确性。

### 1.5 AS 事件的鉴定和分析

采用 Suppa 软件(version: 2017-02-07)<sup>[28]</sup>对基因的 AS 事件进行鉴定。然后根据鉴定结果分类和统计 7 种类型的 AS 事件, 包括内含子保留(retention intron, RI)、外显子跳跃(exon skipping, ES)、可变 5'端剪切(alternative 5' splice sites, A5)和可变 3'端剪切(alternative 3' splice sites, A3)、互斥外显子(mutually exclusive exon, MX)、可变起始外显子(alternative first exons, AF)和可变末端外显子(alternative last exons, AL)。

### 1.6 AS 事件的 RT-PCR 验证

为验证本文中鉴定到的 AS 事件的可靠性, 通过 RT-PCR 对 4 种 AS 事件类型(RI、A5、SE 和 A3)进行验证, 选取包括发生 RI 的 Gene003631 (可通过 AS 形成 isoform009420 和 isoform009415)和 Gene000273 (可通过 AS 形成 isoform000723 和 isoform000725), 发生 A5 的 Gene000541 (可通过 AS 形成 isoform001360 和 isoform001361), 发生 SE 的 Gene001600 (可通过 AS 形成 isoform004216 和 isoform004217), 发生 A3 的 Gene000250 (可通过 AS 形成

isoform000642 和 isoform000644)产生的转录本。利用 Primer Premier 5.0 软件设计上述各转录本的特异性上下游引物(附表 1)。利用 FastPure® Cell/Tissue Total RNA Isolation Kit V2 试剂盒(诺唯赞)提取孢子样品的总 RNA。参照 HiScript® 1st Strand cDNA Synthesis Kit 试剂盒说明书, 利用 oligo(dT)引物进行反转录得到 cDNA, 作为模板进行 PCR 扩增。产物经 3.0% 的琼脂糖凝胶电泳检测, 再通过凝胶成像分析系统(培清)进行观察和拍照。

### 1.7 APA 位点的鉴定和分析

参照 Chen 等的方法<sup>[23]</sup>, 将能够比对上转录本且 3'末端位置为潜在 APA 位点的 isoform 用于 APA 位点分析, 采用 TAPIS 软件<sup>[29]</sup>对 APA 位点进行鉴定, 再根据鉴定结果统计含不同 APA 位点的基因数目。进一步采用 MEME 软件 (version 5.0.5)<sup>[30]</sup>对所有全长转录本的 APA 位点上游 50 bp 的序列特征进行分析以鉴定 motif。

### 1.8 lncRNA 的鉴定和分析

联用 CPC<sup>[31]</sup>和 CNCI<sup>[32]</sup>软件对 lncRNA 进行预测, 均采用默认参数。同时将所有全长转录本比对到 Swiss-prot 数据库(<https://www.uniprot.org>)以查询是否有相应的蛋白功能注释。进而将不具有编码蛋白潜力的转录本和在 Swiss-prot 数据库中无蛋白功能注释的转录本交集作为可靠的 lncRNA 集合。采用 Cuffcompare 软件统计 lncRNA 的类型, 包括正义链 lncRNA (sense lncRNA)、反义链 lncRNA (anti-sense lncRNA)、内含子 lncRNA (intronic lncRNA)、基因间区 lncRNA (long intergenic RNA, lincRNA)和双向 lncRNA (bidirectional lncRNA)。利用 StringTie (1.3.1)软件计算所有 lncRNA 和 mRNA 的 FPKM (fragment per kilobase of per million mapped reads)值, 进一步统计和比较 lncRNA 和 mRNA 的转录本长度, 外显子数量与长度, 内含子数量与长度, GC 含量, 表达量及 AS 事件数目。

## 2 结果与分析

### 2.1 PacBio SMRT 测序数据的质控和评估

AaS 的 PacBio SMRT 测序共产生 9 911 345 条 subreads, 包含 274 928 条 FLNC reads, 经严格校正得到 103 845 条高质量 reads, 平均长度和 N50 长度分别达到 2 502 bp 和 3 262 bp。以上结果表明 PacBio SMRT 测序数据质量良好, 可满足后续分析需要。

### 2.2 Illumina 测序数据的质控和评估

球囊菌孢子 (AaS) 的 RNA-seq 得到 105 675 408 条 raw reads, 经质控得到的 clean reads 为 104 621 402 条。AaS 两端 Q30 (99.9% 碱基正确率) 在 95.78% 及以上, AaS 的序列比对参考基因组的 mapping 率为 84.94%。以上结果表明获得的测序数据质量良好, 可满足后续分析。

### 2.3 蜜蜂球囊菌孢子中基因 AS 事件的鉴定和分析

在孢子中共鉴定到 2 609 次 AS 事件, 包括 1 227 次 (47.03%) RI, 842 次 (32.27%) A3, 415 次 (15.91%) A5, 85 次 (3.26%) AF, 35 次 (1.34%) SE, 4 次 (0.15%) AL 和 1 次 (0.04%) MX。蜜蜂球囊菌的参考基因组共包含 6 442 个基因, 每个基因均注释了 1 条转录本<sup>[25]</sup>。本研究鉴定到 10 693 个基因, 其中有 4 543 个基因对应 1 条转录本, 而多达 6 150 个基因含有 2 条及以上的转录本。KEGG 数据库注释结果显示, 包括发生可变剪接形成的转录本, 其可注释到氨基酸代谢 (546)、核苷酸代谢 (218)、脂质代谢 (364)、碳水化合物代谢 (695) 等多种物质代谢通路; 进一步分析发现, 有 242、81 条 isoform 分别在能量代谢通路和 MAPK 信号通路-酵母上富集。根据 Nr 数据库的蛋白功能注释信息, 筛选注释到丝氨酸/苏氨酸蛋白激酶上的转录本共有 21 条。括号内的数字代表注释到该或通路的 isoform

数。图 1 展示了部分注释到以上通路中的剪接异构体的 IGV 浏览器视图。

### 2.4 蜜蜂球囊菌孢子中 AS 事件的分子验证

利用 RT-PCR 对 RI、A5、SE 和 A3 等类型的 AS 事件进行验证, 琼脂糖凝胶电泳结果显示 Gene003631 可检测到 2 条 isoform 且大小符合预期; 类似地, Gene000273、Gene000541、Gene001600 和 Gene000250 均可检测到 2 条符合预期大小的 isoform (图 2)。以上结果与本研究中的预测结果一致, 证实了不同类型 AS 事件的可靠性。

### 2.5 蜜蜂球囊菌孢子中基因 APA 位点的鉴定和分析

在 AaS 中共鉴定到 5 552 个基因含有 1 个及以上的 APA 位点; 其中含 1 个 APA 位点的基因数目较多, 达到 1 149 个; 其次为 2 个 APA 位点的基因 (782 个); 含 5 个 APA 位点的基因数目较少 (351 个); 此外, 含 5 个以上 APA 位点的基因多达 2 197 个 (图 3A)。进一步的分析发现全长转录本的上下游序列具有明显的碱基倾向性, U 和 A 分别在转录本的下流和上游富集 (图 3B)。

另外, 如图 4 所示, 在全长转录本的 APA 位点上游鉴定到的 4 个 motif, 分别为 AUGAUAA、CAUGAAC、GUUCAAU 和 UCAUCAU。

### 2.6 蜜蜂球囊菌孢子中 lncRNA 的鉴定和分析

通过 CNCI 软件、CPC 软件和比对 SwissProt 数据库分别鉴定到 1 343、2 170 和 3 441 条 lncRNA, 三者的交集为 953 条 (图 5A)。在上述 953 条 lncRNA 中, lincRNA、bidirectional lncRNA、intronic lncRNA、anti-sense lncRNA 和 sense lncRNA 的数目分别为 171、247、9、

154 和 141 条(图 5B)。

结构特征比较分析结果显示, 与 mRNA 相比, 多数 lncRNA (670 条, 70.3%)含有 1 个外

显子(图 6A); lncRNA 的外显子和内含子长度更短, GC 含量更低, 转录本长度更短, 发生不同类型 AS 事件的数量更少(图 6B-F)。

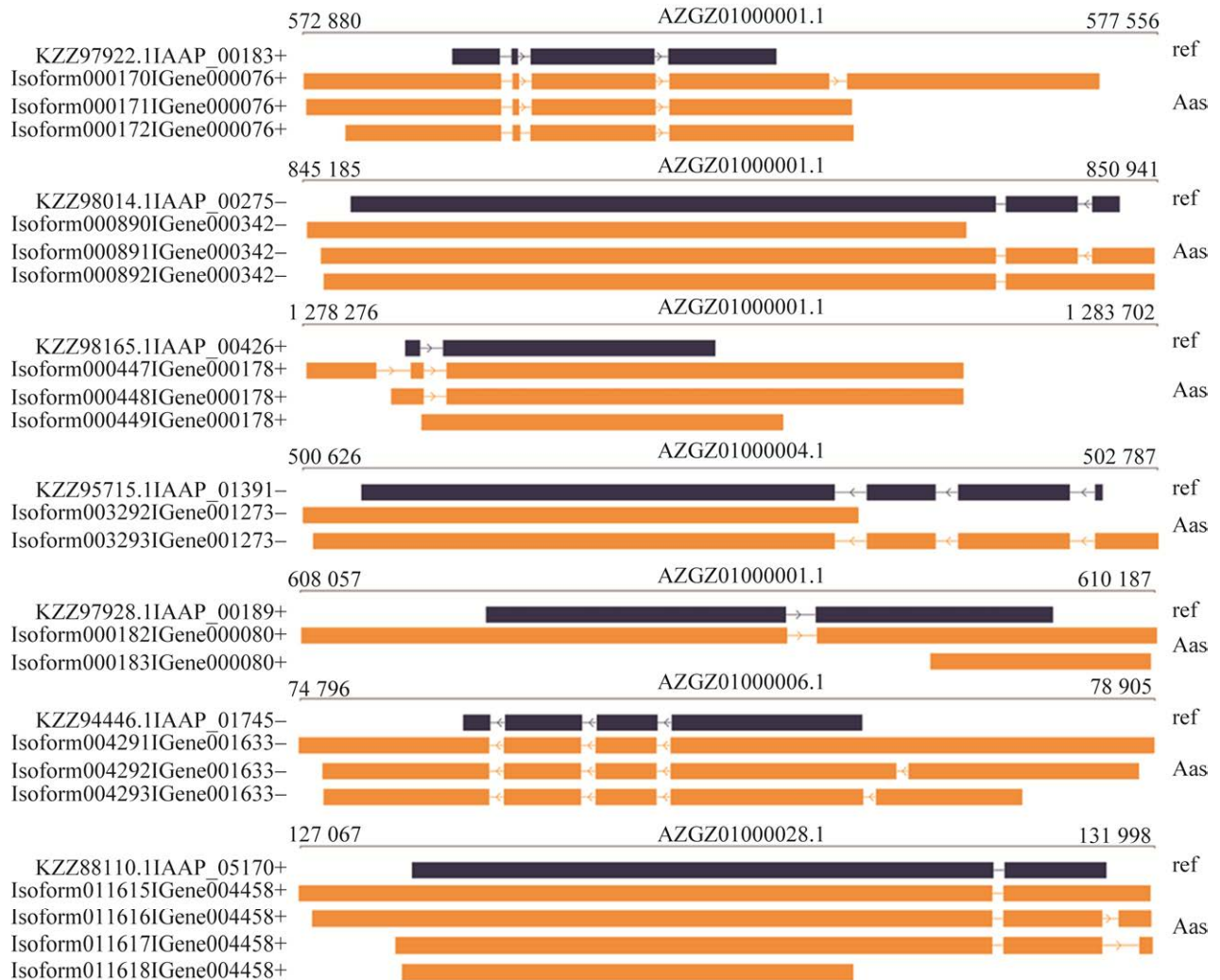


图 1 蜜蜂球囊菌孢子基因中剪接异构体的 IGV 浏览器视图

Figure 1 IGV browser view of gene isoforms in *A. apis* spores. Gene000076, Gene000342, Gene000178, Gene001273, Gene000080, Gene001633 and Gene004458 are the genes encoding cystathionine beta-synthase, DNA-directed RNA polymerase I, AB-hydrolase-associated lipase, alpha-1,6-mannosyltransferase, phosphoserine aminotransferase, MAP kinase kinase MKK1/SSP32 and Protein kinase-like domain protein, respectively. The gray horizontal line and the text above show the chromosome ID, start position and end position of the Genetic region. The rectangles represent exons and horizontal lines represent introns. Blue rectangles represent the annotated transcripts in the reference genome, texts on the left side indicate transcript ID, Gene ID and positive (negative) chain, respectively. Yellow rectangles represent the transcript identified in this study, texts on the left side are transcript ID, Gene ID and positive (negative) chain, respectively.

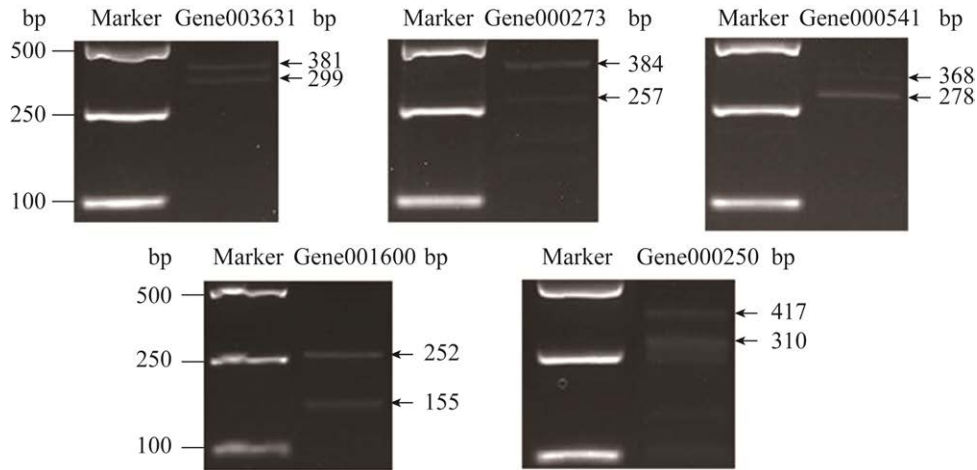


图 2 蜜蜂球囊菌孢子中 5 个基因的 AS 事件的 RT-PCR 验证

Figure 2 RT-PCR validation of AS events of five genes in *A. apis* spores.

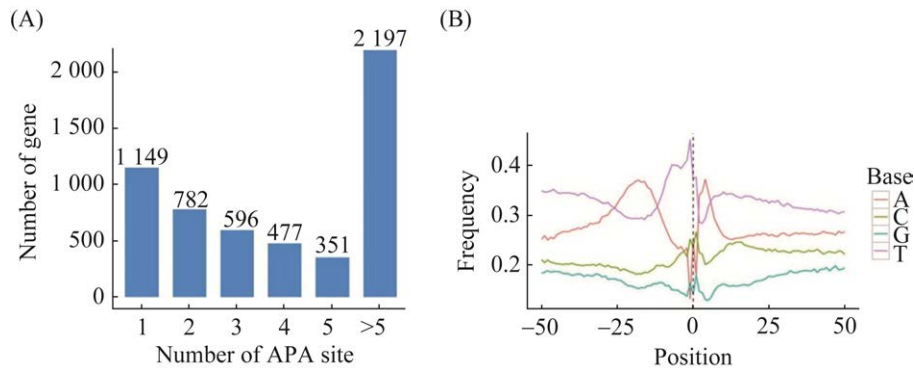


图 3 蜜蜂球囊菌孢子中基因的 APA 位点数量和上下游序列特征

Figure 3 Number (A) and characteristics of upstream and downstream sequences (B) of APA sites of genes in *A. apis* spores.

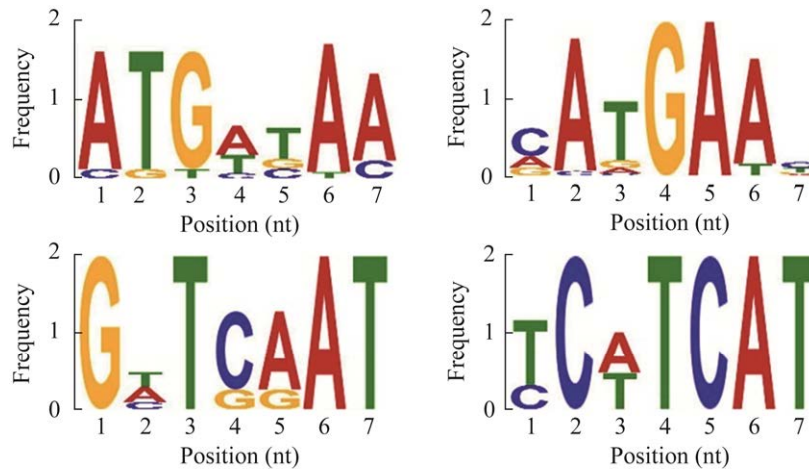


图 4 蜜蜂球囊菌孢子中全长转录本的 APA 位点上游 50 nt 位置鉴定到的 4 个 motif

Figure 4 Four motifs identified at 50nt upstream of APA sites of full-length transcripts in *A. apis* spores.



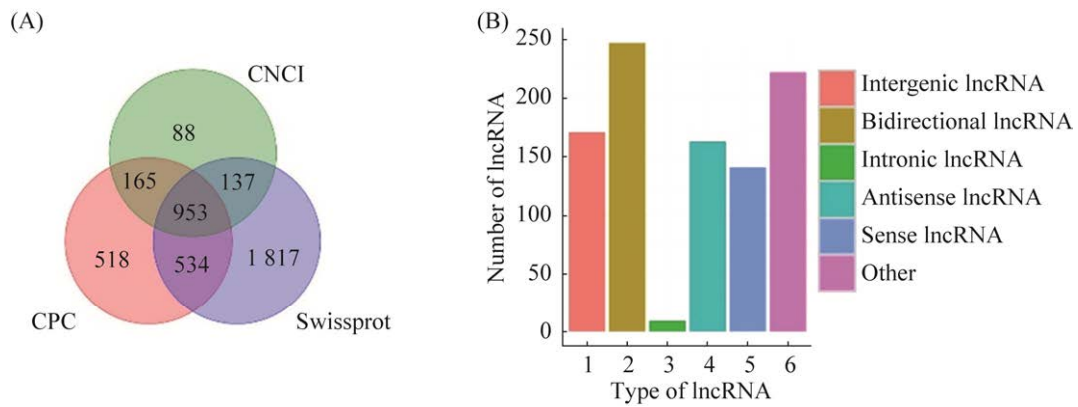


图 5 蜜蜂球囊菌孢子中 lncRNA 的数量和类型

Figure 5 Quantities and varieties of lncRNA in *A. apis* spores. A: Venn diagram of predicted lncRNAs; B: statistics of different types of lncRNA.

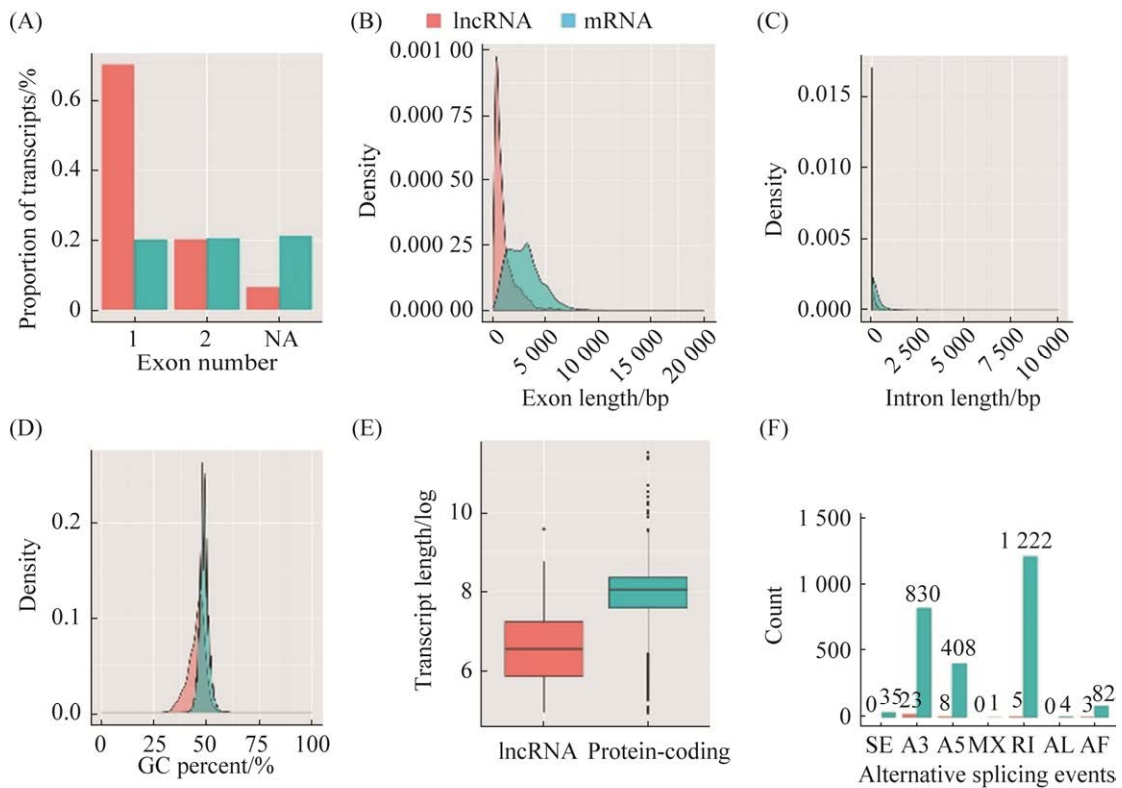


图 6 蜜蜂球囊菌孢子中 lncRNA 与 mRNA 的结构特征比较

Figure 6 Comparison of structural characteristics of lncRNA and mRNA in *A. apis* spores. A: comparison of exon number; B: comparison of exon length; C: comparison of intron length; D: comparison of GC content; E: comparison of transcript length; F: number of lncRNAs and mRNAs occurring various types of AS events.

### 3 讨论

真核生物的基因通常可通过 AS 产生不同的成熟 mRNA, 进而翻译出功能相似或不同的蛋白质, 从而极大地增加了的转录组和蛋白组的多样性<sup>[33]</sup>。蜜蜂球囊菌参考基因组(assembly AAP 1.0)<sup>[34]</sup>注释的基因总数为 6 442 个, 转录本总数为 6 442 条, 一个基因仅对应一条转录本, 这与真核生物的真实情况明显不符。蜜蜂球囊菌参考基因组的基于二代测序产生的短读段拼接组装而成的, 而由于读段较短的限制, 利用二代测序无法精确鉴定全长转录本和 isoform。本研究利用 PacBio SMRT 测序技术对蜜蜂球囊菌的纯净孢子样品进行测序, 共鉴定到 2 609 次 AS 事件, 包括 1 227 次(47.03%) RI, 842 次(32.27%) A3, 415 次(15.91%) A5, 85 次(3.26%) AF, 35 次(1.34%) SE, 4 次(0.15%) AL 和 1 次(0.04%) MX, 形成的 isoform 总数为 14 540 条, 说明蜜蜂球囊菌孢子中基因发生大量不同类型的 AS 事件, 实际上形成的转录本数量远远多于参考基因组上注释的转录本数量, 这为蜜蜂球囊菌参考基因组注释提供了重要补充。鉴于目前基于三代测序技术的真菌 AS 研究较为匮乏, 本研究的结果加深了对蜜蜂球囊菌基因 AS 的认识, 也为其他真菌的 AS 研究提供了有价值的参考信息。

此前, 笔者团队利用 PacBio SMRT 测序技术在蜜蜂球囊菌的纯净菌丝样品中鉴定到 3 617 次 AS 事件, 包括 1 834 次(50.70%) RI, 1 065 次(29.44%) A3, 538 次(14.87%) A5, 84 次(2.32%) AF, 85 次(2.35%) SE, 10 次(0.28%) AL 和 1 次(0.03%) MX。以上结果表明蜜蜂球囊菌菌丝中发生的 AS 事件总数多于孢子, 且多数类型的 AS 事件数量也多于孢子。这与蜜蜂球囊菌孢子是一种休眠态, 而菌丝是活跃的生

长状态的事实相符。在前期研究中, 我们利用 Nanopore 长读段测序技术对另一种常见的蜜蜂真菌病原东方蜜蜂微孢子虫的纯净孢子样品进行了测序, 共鉴定到 1 次 ADS 和 4 次 RI<sup>[35]</sup>。Burkhardt 等在研究古巴假霜霉菌(*Pseudoperonospora cubensis*)时发现 RI 是最主要的 AS 类型, 而可变 3' 剪切和可变 5' 剪切则以较低频率发生<sup>[36]</sup>。Jin 等研究发现在大丽轮枝菌(*Verticillium dahliae*)中 1 个基因能使用 4 个不同的 5' 剪接位点和 2 个不同的 3' 供体位点来产生多达 5 条成熟的 mRNA<sup>[37]</sup>。上述结果表明不同真菌中存在的 AS 事件类型和每一种 AS 事件发生数量存在差异, 真菌中存在活跃的基因转录和 AS 是一个普遍现象。

APA 是指同一基因转录出的不同转录本由于 poly(A) 尾巴发生 AS 而形成不同长度的 3' 非翻译区(3' untranslated region, UTR)。真核生物可通过 APA 产生编码序列或 3' UTR 长度不同的转录本, 从而调节 RNA 的定位、功能、稳定性和翻译效率, 并大幅增加转录组的多样性<sup>[38]</sup>。此前, 人们已在果蝇<sup>[39]</sup>(*Drosophila melanogaster*)、小鼠<sup>[40]</sup>(*Mus musculus*)、玉米<sup>[41]</sup>(*Zea mays*)和秀丽隐杆线虫<sup>[42]</sup>(*Caenorhabditis elegans*)等物种中发现 APA 现象。本研究在蜜蜂球囊菌孢子中共鉴定到 5 552 个基因含不同数量的 APA 位点, 其中含有 1 个 APA 位点的基因数量最多(1 149, 20.70%), 其次为含有 2 个 APA 位点的基因(782, 14.08%), 含 5 个及以上 APA 位点的基因数量达到 2 197 个(39.57%)。Chang 等利用 PacBio SMRT 测序技术对黄牛(*Bos taurus*)进行全长转录本研究, 在 2 643 个基因中共鉴定到 3 795 个 poly(A) 位点, 其中含 1 个 APA 位点的基因数量最多(1 929, 72.98%), 含有 5 个 APA 位点的基因数量最低(21, 0.79%)<sup>[43]</sup>。在高粱(*Sorghum bicolor* L. Moench)中, 超过 11 000 个基因包含

1个及以上APA位点,其中含1个APA位点的基因数最多,达到3289个(29.86%)<sup>[44]</sup>。前期研究中,笔者团队基于Nanopore长读段测序技术在东方蜜蜂微孢子虫的孢子中鉴定到233个基因含1个及以上APA位点,其中有多达61.37%的基因仅含1个APA位点<sup>[35]</sup>。以上结果共同表明无论是真菌菌丝和孢子还是动植物物种,APA都是普遍存在的现象,并且多数基因都只含1个APA位点。本研究中,在蜜蜂球囊菌孢子中全长转录本的poly(A)剪接位点上游鉴定到的4个motif,与前期基于Nanopore测序技术在蜜蜂球囊菌全长转录本的APA位点上游鉴定到4个的motif(UCUCCU、UCUUCU、CCCACC和CCCCU)不存在交集<sup>[22]</sup>,说明相比于Nanopore测序,PacBio SMRT测序能够鉴定到更多的motif,两种三代测序的主流技术在鉴定APA位点上游motif具有相互补充的潜力。AAUAAA作为一个在动植物中较为保守的motif,也曾在东方蜜蜂微孢子虫孢子中全长转录本APA位点上游鉴定到<sup>[35]</sup>,但本研究在蜜蜂球囊菌孢子中未鉴定到。这表明该motif同时具有保守性和种属特异性。

lncRNA是一类长度大于200 nt,包含2个以上外显子,缺少开放阅读框(ORF)的非编码RNA(ncRNA),可参与染色质修饰和转录调控等重要的生命活动<sup>[45]</sup>。目前,已在人类<sup>[46]</sup>(*Homo sapiens*)、拟南芥<sup>[47]</sup>(*Arabidopsis thaliana*)、家蚕<sup>[48]</sup>(*Bombyx mori*)和酿酒酵母<sup>[49]</sup>(*Saccharomyces cerevisiae*)等许多物种中鉴定到大量lncRNA。笔者团队前期基于Illumina测序技术在蜜蜂球囊菌孢子中鉴定到811条lncRNA<sup>[15]</sup>,最大长度和平均长度分别为4814 bp和740 bp。本研究在蜜蜂球囊菌孢子中共鉴定到953条lncRNA,最大长度和平均长度分别达到6327 bp和969 bp。这说明通过PacBio SMRT测序技术可以鉴定到

更多和更长的lncRNA,体现出较二代测序技术更为显著的优势。另外,本研究鉴定到的lncRNA与mRNA相比,具有更少的外显子,更短的外显子和内含子长度,更短的转录本长度,更低的表达水平和更少的AS事件。这些结构特征不仅与已报道的其他物种的lncRNA相似<sup>[50-51]</sup>,而且与我们前期基于二代测序鉴定到的蜜蜂球囊菌lncRNA一致<sup>[24]</sup>。以上结果体现了不同物种lncRNA的结构保守性。

孢子作为真菌的一种休眠态的形式存在,但对于真菌孢子中生命活动如转录、翻译和代谢的认识非常有限。本研究结合PacBio SMRT和Illumina测序技术鉴定到蜜蜂球囊菌孢子中的大量AS、APA和lncRNA,研究结果丰富了蜜蜂球囊菌的基础生物学信息,深入揭示了蜜蜂球囊菌孢子中转录组的复杂性,并为真菌孢子中存在复杂生命活动提供了新证据。

## 参考文献

- [1] Aronstein KA, Murray KD. Chalkbrood disease in honey bees. *Journal of Invertebrate Pathology*, 2010, 103: S20-S29.
- [2] Chen DF, Guo R, Xiong CL, Zheng YZ, Hou CS, Fu ZM. Morphological and molecular identification of chalkbrood disease pathogen *Ascosphaera apis* in *Apis cerana cerana*. *Journal of Apicultural Research*, 2018, 57(4): 516-521.
- [3] 梁勤, 陈大福. 蜜蜂保护学. 2版. 北京: 中国农业出版社, 2009.
- [4] Popham DL, Bernhards CB. Spore peptidoglycan. *Microbiology Spectrum*, 2015, 3(6). DOI: 10.1128/microbiolspec.TBS-0005-2012.
- [5] Li SJ, Tang YW, Fang XM, Qiao TM, Han S, Zhu TH. Whole-genome sequence of *Arthrrium phaeospermum*, a globally distributed pathogenic fungus. *Genomics*, 2020, 112(1): 919-929.
- [6] Zhang XK, Dai PX, Gao YP, Gong XW, Cui H, Jin YP, Zhang YH. Transcriptome sequencing and analysis of zinc-uptake-related genes in *Trichophyton mentagrophytes*. *BMC Genomics*, 2017, 18(1): 888.
- [7] He Y, Zhang CH, Deng WQ, Zhou XY, Li TH, Li CH.

- Transcriptome sequencing analysis of the MSDIN gene family encoding cyclic peptides in lethal *Amanita fuligineoides*. *Toxicon*, 2020, 183: 61–68.
- [8] Huang Q, Evans JD. Identification of microRNA-like small RNAs from fungal parasite *Nosema ceranae*. *Journal of Invertebrate Pathology*, 2016, 133: 107–109.
- [9] Guo R, Chen DF, Xiong CL, Hou CS, Zheng YZ, Fu ZM, Liang Q, Diao QY, Zhang L, Wang HQ, Hou ZX, Kumar D. First identification of long non-coding RNAs in fungal parasite *Nosema ceranae*. *Apidologie*, 2018, 49(5): 660–670.
- [10] Getachew A, Abejew TA, Wu JL, Xu J, Yu HM, Tan J, Wu PJ, Tu YY, Kang WP, Wang Z, Xu SF. Transcriptome profiling reveals insertional mutagenesis suppressed the expression of candidate pathogenicity genes in honeybee fungal pathogen, *Ascosphaera apis*. *Scientific Reports*, 2020, 10: 7532.
- [11] Cornman RS, Bennett AK, Murray KD, Evans JD, Elsik CG, Aronstein K. Transcriptome analysis of the honey bee fungal pathogen, *Ascosphaera apis*: implications for host pathogenesis. *BMC Genomics*, 2012, 13: 285.
- [12] 熊翠玲, 杜宇, 王鸿权, 郑燕珍, 付中民, 王海朋, 张璐, 陈大福, 郭睿. 基于比较转录组学分析揭示中华蜜蜂及意大利蜜蜂幼虫的球囊菌抗性差异机制. *中国农业大学学报*, 2019, 24(5): 106–114.  
Xiong CL, Du Y, Wang HQ, Zheng YZ, Fu ZM, Wang HP, Zhang L, Chen DF, Guo R. Unraveling the mechanism regulating the *Ascosphaera apis*-resistance difference between *Apis cerana cerana* and *Apis mellifera ligustica* larvae based on comparative transcriptome analysis. *Journal of China Agricultural University*, 2019, 24(5): 106–114. (in Chinese)
- [13] 陈华枝, 蒋海宾, 祝智威, 范元婵, 许雅静, 孙明会, 刘佳美, 熊翠玲, 郑燕珍, 付中民, 徐国钧, 陈大福, 郭睿. 蜜蜂球囊菌菌丝和孢子中环状 RNA 的鉴定及比较分析. *微生物学报*, 2021, 61(5): 1299–1314.  
Chen HZ, Jiang HB, Zhu ZW, Fan YC, Xu YJ, Sun MH, Liu JM, Xiong CL, Zheng YZ, Fu ZM, Xu GJ, Chen DF, Guo R. Identification and comparison of circular RNAs in *Ascosphaera apis* mycelium and spore. *Acta Microbiologica Sinica*, 2021, 61(5): 1299–1314. (in Chinese)
- [14] 陈华枝, 祝智威, 蒋海宾, 王杰, 范元婵, 范小雪, 万洁琦, 卢家轩, 熊翠玲, 郑燕珍, 付中民, 陈大福, 郭睿. 蜜蜂球囊菌菌丝和孢子中微小 RNA 及其靶 mRNA 的比较分析. *中国农业科学*, 2020, 53(17): 3606–3619.  
Chen HZ, Zhu ZW, Jiang HB, Wang J, Fan YC, Fan XX, Wan JQ, Lu JX, Xiong CL, Zheng YZ, Fu ZM, Chen DF, Guo R. Comparative analysis of microRNAs and corresponding target mRNAs in *Ascosphaera apis* mycelium and spore. *Scientia Agricultura Sinica*, 2020, 53(17): 3606–3619. (in Chinese)
- [15] 陈华枝, 王杰, 祝智威, 蒋海宾, 范元婵, 范小雪, 万洁琦, 卢家轩, 郑燕珍, 付中民, 徐国钧, 陈大福, 郭睿. 蜜蜂球囊菌菌丝和孢子中链非编码 RNA 的比较及潜在功能分析. *中国农业科学*, 2021, 54(2): 435–448.  
Chen HZ, Wang J, Zhu ZW, Jiang HB, Fan YC, Fan XX, Wan JQ, Lu JX, Zheng YZ, Fu ZM, Xu GJ, Chen DF, Guo R. Comparison and potential functional analysis of long non-coding RNAs between *Ascosphaera apis* mycelium and spore. *Scientia Agricultura Sinica*, 2021, 54(2): 435–448. (in Chinese)
- [16] Zhang GQ, Sun M, Wang JF, Lei M, Li CJ, Zhao DJ, Huang J, Li WJ, Li SL, Li J, Yang J, Luo YF, Hu SN, Zhang B. PacBio full-length cDNA sequencing integrated with RNA-seq reads drastically improves the discovery of splicing transcripts in rice. *The Plant Journal*, 2019, 97(2): 296–305.
- [17] Hong F, Mo SH, Lin XY, Niu JZ, Yin J, Wei D. The PacBio full-length transcriptome of the tea aphid as a reference resource. *Frontiers in Genetics*, 2020, 11: 558394.
- [18] Zhang CH, Deng WQ, Yan WJ, Li TH. Whole genome sequence of an edible and potential medicinal fungus, *Cordyceps guangdongensis*. *G3: Bethesda, Md*, 2018, 8(6): 1863–1870.
- [19] Penouilh-Suzette C, Fourré S, Besnard G, Godiard L, Pecrix Y. A simple method for high molecular-weight genomic DNA extraction suitable for long-read sequencing from spores of an obligate biotrophic oomycete. *Journal of Microbiological Methods*, 2020, 178: 106054.
- [20] 杜宇, 祝智威, 王杰, 王秀娜, 蒋海宾, 范元婵, 范小雪, 陈华枝, 隆琦, 蔡宗兵, 熊翠玲, 郑燕珍, 付中民, 陈大福, 郭睿. 利用第三代纳米孔长读段测序技术构建和注释蜜蜂球囊菌的全长转录组. *中国农业科学*, 2021, 54(4): 864–876.  
Du Y, Zhu ZW, Wang J, Wang XN, Jiang HB, Fan YC, Fan XX, Chen HZ, Long Q, Cai ZB, Xiong CL, Zheng YZ, Fu ZM, Chen DF, Guo R. Construction and annotation of *Ascosphaera apis* full-length transcriptome utilizing nanopore third-generation

- long-read sequencing technology. *Scientia Agricultura Sinica*, 2021, 54(4): 864–876. (in Chinese)
- [21] 杜宇, 蒋海宾, 王杰, 范小雪, 王秀娜, 冯睿蓉, 张文德, 隆琦, 熊翠玲, 郑燕珍, 陈大福, 郭睿. 蜜蜂球囊菌菌丝和孢子中全长转录本的差异表达分析. *昆虫学报*, 2021, 64(3): 363–373.  
Du Y, Jiang HB, Wang J, Fan XX, Wang XN, Feng RR, Zhang WD, Long Q, Xiong CL, Zheng YZ, Chen DF, Guo R. Differential expression analysis of full-length transcripts in mycelia and spores of *Ascosphaera apis*. *Acta Entomologica Sinica*, 2021, 64(3): 363–373. (in Chinese)
- [22] 杜宇, 王杰, 蒋海宾, 王秀娜, 范元婵, 范小雪, 祝智威, 隆琦, 张文德, 熊翠玲, 郑燕珍, 付中民, 陈大福, 郭睿. 基于第三代长读段测序数据解析蜜蜂球囊菌基因的可变剪切与可变腺苷酸化. *微生物学报*, 2021, 61(3): 667–682.  
Du Y, Wang J, Jiang HB, Wang XN, Fan YC, Fan XX, Zhu ZW, Long Q, Zhang WD, Xiong CL, Zheng YZ, Fu ZM, Chen DF, Guo R. Analysis of alternative splicing and polyadenylation of *Ascosphaera apis* genes based on third-generation long-read sequencing dataset. *Acta Microbiologica Sinica*, 2021, 61(3): 667–682. (in Chinese)
- [23] Chen DF, Du Y, Fan XX, Zhu ZW, Jiang HB, Wang J, Fan YC, Chen HZ, Zhou DD, Xiong CL, Zheng YZ, Xu XJ, Luo Q, Guo R. Reconstruction and functional annotation of *Ascosphaera apis* full-length transcriptome utilizing PacBio long reads combined with Illumina short reads. *Journal of Invertebrate Pathology*, 2020, 176: 107475.
- [24] Guo R, Chen DF, Xiong CL, Hou CS, Zheng YZ, Fu ZM, Diao QY, Zhang L, Wang HQ, Hou ZX, Li WD, Kumar D, Liang Q. Identification of long non-coding RNAs in the chalkbrood disease pathogen *Ascosphaera apis*. *Journal of Invertebrate Pathology*, 2018, 156: 1–5.
- [25] Li Y, Fang CC, Fu YH, Hu A, Li CC, Zou C, Li XY, Zhao SH, Zhang CJ, Li CC. A survey of transcriptome complexity in *Sus scrofa* using single-molecule long-read sequencing. *DNA Research*, 2018, 25(4): 421–437.
- [26] Gordon SP, Tseng E, Salamov A, Zhang JW, Meng XD, Zhao ZY, Kang DW, Underwood J, Grigoriev IV, Figueroa M, Schilling JS, Chen F, Wang Z. Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing. *PLoS One*, 2015, 10(7): e0132628.
- [27] 蒋海宾, 杜宇, 范小雪, 王杰, 祝智威, 范元婵, 熊翠玲, 付中民, 徐国钧, 陈大福, 郭睿. 蜜蜂球囊菌菌丝和孢子的比较转录组分析. *四川大学学报: 自然科学版*, 2020, 57(6): 1177–1185.  
Jiang HB, Du Y, Fan XX, Wang J, Zhu ZW, Fan YC, Xiong CL, Fu ZM, Xu GJ, Chen DF, Guo R. Comparative transcriptome analysis of *Ascosphaera apis* mycelium and spore. *Journal of Sichuan University: Natural Science Edition*, 2020, 57(6): 1177–1185. (in Chinese)
- [28] Alamancos GP, Pagès A, Trincado JL, Bellora N, Eyra E. Leveraging transcript quantification for fast computation of alternative splicing profiles. *RNA: New York, NY*, 2015, 21(9): 1521–1531.
- [29] Chao YH, Yuan JB, Li SF, Jia SQ, Han LB, Xu LX. Analysis of transcripts and splice isoforms in red clover (*Trifolium pratense* L.) by single-molecule long-read sequencing. *BMC Plant Biology*, 2018, 18(1): 300.
- [30] Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings International Conference on Intelligent Systems for Molecular Biology*, 1994, 2: 28–36.
- [31] Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei LP, Gao G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*, 2007, 35(suppl\_2): W345–W349.
- [32] Sun L, Luo HT, Bu DC, Zhao GG, Yu KT, Zhang CH, Liu YN, Chen RS, Zhao Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Research*, 2013, 41(17): e166.
- [33] Shenasa H, Hertel KJ. Combinatorial regulation of alternative splicing. *Biochimica et Biophysica Acta: BBA - Gene Regulatory Mechanisms*, 2019, 1862(11/12): 194392.
- [34] Shang YF, Xiao GH, Zheng P, Cen K, Zhan S, Wang CS. Divergent and convergent evolution of fungal pathogenicity. *Genome Biology and Evolution*, 2016, 8(5): 1374–1387.
- [35] 陈华枝, 范小雪, 范元婵, 王杰, 祝智威, 蒋海宾, 张文德, 隆琦, 熊翠玲, 郑燕珍, 付中民, 徐国钧, 陈大福, 郭睿. 东方蜜蜂微孢子虫基因的可变剪接及可变腺苷酸化解析. *菌物学报*, 2021, 40(1): 161–173.  
Chen HZ, Fan XX, Fan YC, Wang J, Zhu ZW, Jiang HB, Zhang WD, Long Q, Xiong CL, Zheng YZ, Fu ZM,

- Xu GJ, Chen DF, Guo R. Analysis of alternative splicing and alternative polyadenylation of *Nosema ceranae* genes. *Mycosystema*, 2021, 40(1): 161–173. (in Chinese)
- [36] Burkhardt A, Buchanan A, Cumbie JS, Savory EA, Chang JH, Day B. Alternative splicing in the obligate biotrophic oomycete pathogen *Pseudoperonospora cubensis*. *Molecular Plant-Microbe Interactions: MPMI*, 2015, 28(3): 298–309.
- [37] Jin LR, Li GL, Yu DZ, Huang W, Cheng C, Liao SJ, Wu QJ, Zhang Y. Transcriptome analysis reveals the complexity of alternative splicing regulation in the fungus *Verticillium dahliae*. *BMC Genomics*, 2017, 18(1): 130.
- [38] Elkon R, Ugalde AP, Agami R. Alternative cleavage and polyadenylation: extent, regulation and function. *Nature Reviews Genetics*, 2013, 14(7): 496–506.
- [39] Smibert P, Miura P, Westholm JO, Shenker S, May G, Duff MO, Zhang DY, Eads BD, Carlson J, Brown JB, Eisman RC, Andrews J, Kaufman T, Cherbas P, Celniker SE, Graveley BR, Lai EC. Global patterns of tissue-specific alternative polyadenylation in *Drosophila*. *Cell Reports*, 2012, 1(3): 277–289.
- [40] Kasowitz SD, Ma J, Anderson SJ, Leu NA, Xu Y, Gregory BD, Schultz RM, Wang PJ. Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. *PLoS Genetics*, 2018, 14(5): e1007412.
- [41] Jafar Z, Tariq S, Sadiq I, Nawaz T, Akhtar MN. Genome-wide profiling of polyadenylation events in maize using high-throughput transcriptomic sequences. *G3 Genes Genomes Genetics*, 2019, 9(8): 2749–2760.
- [42] Blazie SM, Babb C, Wilky H, Rawls A, Park JG, Mangone M. Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in *Caenorhabditis elegans* intestine and muscles. *BMC Biology*, 2015, 13: 4.
- [43] Chang TP, An BX, Liang M, Duan XH, Du LL, Cai WT, Zhu B, Gao X, Chen Y, Xu LY, Zhang LP, Gao HJ, Li JY. PacBio single-molecule long-read sequencing provides new light on the complexity of full-length transcripts in cattle. *Frontiers in Genetics*, 2021, 12: 664974.
- [44] Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy ASN. A survey of the *Sorghum* transcriptome using single-molecule long reads. *Nature Communications*, 2016, 7: 11706.
- [45] 杜宇, 冯睿蓉, 王杰, 祝智威, 张文德, 余岢骏, 隆琦, 蔡宗兵, 解彦玲, 熊翠玲, 郑燕珍, 陈大福, 郭睿. 中华蜜蜂 6 日龄幼虫响应蜜蜂球囊菌感染的长链非编码 RNA 应答研究. *微生物学报*, 2021, 61(5): 1338–1358.
- Du Y, Feng RR, Wang J, Zhu ZW, Zhang WD, Yu KJ, Long Q, Cai ZB, Xie YL, Xiong CL, Zheng YZ, Chen DF, Guo R. Long non-coding RNA response of 6-day-old *Apis cerana cerana* larvae to *Ascospaera apis* infection. *Acta Microbiologica Sinica*, 2021, 61(5): 1338–1358. (in Chinese)
- [46] Kazimierczyk M, Kasprowicz MK, Kasprzyk ME, Wrzesinski J. Human long noncoding RNA interactome: detection, characterization and function. *International Journal of Molecular Sciences*, 2020, 21(3): 1027.
- [47] Zhao X, Li J, Lian B, Gu H, Li Y, Qi Y. Global identification of *Arabidopsis* lncRNAs reveals the regulation of MAF<sub>4</sub> by a natural antisense RNA. *Nature Communications*, 2018, 9: 5056.
- [48] Wu YQ, Cheng TC, Liu C, Liu DL, Zhang Q, Long RW, Zhao P, Xia QY. Systematic identification and characterization of long non-coding RNAs in the silkworm, *Bombyx mori*. *PLoS One*, 2016, 11(1): e0147147.
- [49] Niederer RO, Hass EP, Zappulla DC. Long noncoding RNAs in the yeast *S. cerevisiae*. *Advances in Experimental Medicine and Biology*, 2017: 119–132.
- [50] Pauli A, Valen E, Lin MF, Garber M, Vastenhout NL, Levin JZ, Fan L, Sandelin A, Rinn JL, Regev A, Schier AF. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Research*, 2012, 22(3): 577–591.
- [51] Guo R, Chen DF, Xiong CL, Hou CS, Zheng YZ, Fu ZM, Liang Q, Diao QY, Zhang L, Wang HQ, Hou ZX, Kumar D. First identification of long non-coding RNAs in fungal parasite *Nosema ceranae*. *Apidologie*, 2018, 49(5): 660–670.

(本文责编 李磊)

#### 附表 1 RT-PCR 引物信息

本文补充材料见网络版 <http://journals.im.ac.cn/actamicrocn>。补充材料为作者提供的原始数据，作者对其学术质量和内容负责。