

# 蜜蜂球囊菌菌丝和孢子中毒力因子 相关全长转录本的鉴定及分析\*

隆 琦<sup>1\*\*</sup> 余岢骏<sup>1\*\*</sup> 吴 鹰<sup>1</sup> 孙明会<sup>1</sup> 冯睿蓉<sup>1</sup> 赵 萧<sup>1</sup>  
胡 颖<sup>1</sup> 徐细建<sup>3</sup> 付中民<sup>1,2</sup> 陈大福<sup>1,2\*\*\*</sup> 郭 睿<sup>1,2\*\*\*</sup>

(1. 福建农林大学动物科学学院(蜂学学院), 福州 350002;

2. 福建农林大学蜂疗研究所, 福州 350002; 3. 江西省养蜂研究所, 南昌 330000)

**摘 要** 【目的】蜜蜂球囊菌 *Ascosphaera apis* 是一种专性感染蜜蜂幼虫而导致白垩病的致死性真菌病原。基于前期已获得的高质量纳米孔(Nanopore)长读段测序数据,对蜜蜂球囊菌菌丝(AaM)和孢子(AaS)中的毒力因子相关全长转录本进行鉴定和分析,为毒力因子相关剪接异构体的功能研究提供参考信息和基础。【方法】利用 BLAST 工具,将 AaM 和 AaS 中的所有全长转录本比对 Nr 数据库,以鉴定蜜蜂球囊菌的毒力因子(几丁质酶、脂肪酶、水解酶和蛋白酶)相关的全长转录本。使用 minimap2 软件,将 AaM 和 AaS 中的全长转录本序列与蜜蜂球囊菌参考基因组注释的已知转录本序列进行比对,将比对到参考基因组的全长转录本进行归一化处理,再通过每百万计数法(Counts per million, CPM)计算毒力因子相关全长转录本的表达量。利用百迈克云平台的相关工具绘制转录本的表达量聚类热图。通过 IGV 浏览器对部分毒力因子相关全长转录本结构进行可视化。【结果】在 AaM 鉴定到毒力因子相关的 367 个基因及 407 个全长转录本,包括 12 条几丁质酶相关全长转录本,48 条脂肪酶相关全长转录本,289 条水解酶相关全长转录本,58 条蛋白酶相关全长转录本。在 AaS 鉴定到毒力因子相关 367 个基因及 400 个全长转录本,包括 14 条几丁质酶相关全长转录本,63 条脂肪酶相关全长转录本,267 条水解酶相关全长转录本,56 条蛋白酶相关全长转录本。另外,AaM 和 AaS 特有的毒力因子(几丁质酶、脂肪酶)相关全长转录本分别有 0 条和 17 条,共有的毒力因子相关全长转录本有 60 条。进一步分析发现,蜜蜂球囊菌的部分毒力因子基因可通过可变剪接形成多条剪接异构体。【结论】共鉴定到蜜蜂球囊菌毒力因子(几丁质酶、脂肪酶、水解酶和蛋白酶)相关的 367 个基因和 486 条全长转录本;相比于蜜蜂球囊菌参考基因组注释的转录本,绝大多数毒力因子基因对应的全长转录本数量更多且结构更为复杂。研究结果丰富了蜜蜂球囊菌毒力因子相关基因和转录本的注释信息,为毒力因子相关剪接异构体的功能研究提供了基础,也为白垩病防控提供了潜在靶点。  
**关键词** 蜜蜂球囊菌; 菌丝; 孢子; 毒力因子; 全长转录本; 第三代测序

## Identification and analysis of virulence factor-related full-length transcripts in *Ascosphaera apis* mycelium and spores

LONG Qi<sup>1\*\*</sup> YU Ke-Jun<sup>1\*\*</sup> WU Ying<sup>1</sup> SUN Ming-Hui<sup>1</sup> FENG Rui-Rong<sup>1</sup>  
ZHAO Xiao<sup>1</sup> HU Ying<sup>1</sup> XU Xi-Jian<sup>3</sup> FU Zhong-Min<sup>1,2</sup> CHEN Da-Fu<sup>1,2\*\*\*</sup> GUO Rui<sup>1,2\*\*\*</sup>

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

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

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

**Abstract** [Objectives] To identify and analyze virulence factor-related full-length transcripts of *Ascosphaera apis*, a lethal

\*资助项目 Supported projects: 国家自然科学基金(31702190); 国家现代农业产业技术体系建设专项(CARS-44-KXJ7); 福建农林大学硕士生导师团队项目; 福建省病原真菌与真菌毒素重点实验室开放课题; 江西省应用研究培育计划(20181BBF68003); 福建省大学生创新创业训练计划项目(202010389016; 202010389162)

\*\*共同第一作者 Co-first authors, E-mail: 11070190695@126.com; kejunyu2021@163.com

\*\*\*共同通讯作者 Co-corresponding authors, E-mail: dfchen826@fafu.edu.cn; ruiguo@fafu.edu.cn

收稿日期 Received: 2021-06-15; 接受日期 Accepted: 2021-08-02

fungal pathogen that exclusively infects honeybee larvae causing chalkbrood disease, using previously gained high-quality long-read sequencing data, thereby providing reference information and a foundation for the functional study of virulence factor-associated isoforms of this pathogen. **[Methods]** All full-length transcripts of *A. apis* mycelium (AaM) and spores (AaS) were aligned to the Nr database using BLAST tool to identify full-length transcripts relative to virulence factors, such as chitinase, lipase, hydrolase and protease. Sequences of full-length transcripts in AaM and AaS were aligned to those of known transcripts annotated in the *A. apis* reference genome, followed by normalization of the numbers of full-length transcripts mapped to the reference genome. The expression levels of each virulence factor-related, full-length transcript were then calculated. A heatmap showing the spatial pattern of transcript expression was drawn using the appropriate tool in BMKCloud and the structures of partial virulence factor-related full-length transcripts were visualized in IGV browser. **[Results]** 367 genes and 407 full-length transcripts associated with virulence factors were identified in AaM, of which 12 were chitinase-related, 48 lipase-related, 289 hydrolase-related and 58 protease-related. 367 genes and 400 full-length transcripts associated with virulence factors were identified in AaS, of which 14 were chitinase-related, 63 lipase-related, 267 hydrolase-related and 56 protease-related. In addition, there were 0 and 17 unique virulence factor-related, full-length transcripts (lipase-related and chitinase-related transcripts) in AaM and AaS, respectively, and 60 that were common to mycelia and spores. Further analysis indicated that part of the virulence factor-related genes can generate several isoforms via alternative splicing. **[Conclusion]** A total of 367 genes and 486 full-length transcripts associated with virulence factors, such as chitinase, lipase, hydrolase and protease were identified. Most full-length transcripts corresponding to virulence factor-related genes were relatively abundant and had more complex structures. Our findings enrich the annotation of genes and transcripts relevant to *Ascosphaera apis* virulence factors, provide a foundation for the functional study of virulence factor-associated isoforms and offer potential targets for the prevention and control of chalkbrood in honey-bees.

**Key words** *Ascosphaera apis*; mycelium; spore; virulence factor; full-length transcript; third-generation sequencing

蜜蜂球囊菌 *Ascosphaera apis* 专性感染蜜蜂幼虫, 导致的白垩病对蜜蜂健康和养蜂业造成较大危害。对于蜜蜂球囊菌的毒力因子, 相关研究不够全面和深入。早在 2006 年, Qin 等 (2006) 就利用 Sanger 测序技术组装了蜜蜂球囊菌的基因组, 但作者公布序列信息的同时并未公布功能信息, 导致该版本长期无法被利用。Shang 等 (2016) 通过二代测序技术重新组装和注释了蜜蜂球囊菌的参考基因组 (Assembly AAP 1.0), 为开展相关分子生物学及组学研究奠定了基础, 该版本虽组装到 scaffold 水平, 但基因序列和功能注释仍需完善, 所有注释基因均缺乏 5'UTR 和 3' UTR 信息。球孢白僵菌 *Beauveria bassiana* 是感染昆虫的重要病原真菌, Zhang 等 (2010) 研究发现 *Bbhog1* 和 *Bbmpk1* 与病原侵染过程中调节疏水性及穿透体壁相关, 并对球孢白僵菌的菌种毒力具有潜在影响。Carollo 等 (2010) 在对金龟子绿僵菌 *Metarhizium anisopliae* 的研究中发现磷脂和肉碱等脂代谢相关物质的增多可加速菌丝生长进而促进侵染进程。与球孢白僵菌

和金龟子绿僵菌等其他真菌病原不同, 蜜蜂球囊菌只能通过经口摄入的途径感染蜜蜂幼虫, 病原孢子在幼虫中肠内低水平萌发, 至预蛹期中肠和后肠肠间的隔膜消失, 病原孢子随食物涌入后肠, 在接触氧气后剧烈萌发, 同时菌丝大量生长并相继穿透肠壁和体壁, 进而杀死幼虫, 期间伴随着几丁质酶、蛋白酶、脂肪酶和水解酶等多种毒力因子的合成与释放 (Cornman *et al.*, 2012)。Cornman 等 (2012) 利用 454 焦磷酸测序技术对液体或固体培养基上培养的蜜蜂球囊菌菌丝及患白垩病的西方蜜蜂 *Apis mellifera* 幼虫体表长出的菌丝进行测序和比较分析, 鉴定到 1 个葡糖淀粉酶基因, 3 个几丁质酶基因, 16 个氨酰酶基因, 30 个酯酶基因, 42 个蛋白酶基因, 24 个脂肪酶基因以及若干其他毒力因子基因。笔者所在课题组前期利用 Illumina 测序得到的短读段组装和注释了蜜蜂球囊菌的第一个参考转录组 (张翌楠等, 2017), 并通过解析蜜蜂球囊菌感染意大利蜜蜂 *Apis mellifera ligustica* 幼虫和中华蜜蜂 *Apis cerana cerana* 幼虫过程的转录组动态初步

揭示了病原的侵染机制(陈大福等, 2017a; 郭睿等, 2017b)。

此前, 笔者所在团队曾对蜜蜂球囊菌的毒力因子进行了一些生化研究, 在蜜蜂球囊菌的体外培养实验中, 通过诱导蜜蜂球囊菌分泌胞外蛋白酶并利用相应底物进行酶活测定, 结果显示蜜蜂球囊菌分泌的胞外蛋白酶主要为酪蛋白酶、偶氮酪蛋白酶和胶原蛋白酶(李飞雄等, 2011); 通过 SDS-PAGE 和染色对健康蜜蜂幼虫和患白垩病幼虫的血淋巴进行分析和比较, 发现患病幼虫血淋巴中的蛋白组分发生降解, 蜜蜂球囊菌能分泌胞外酶降解幼虫的组织器官, 为自身的增殖提供营养物质(郑志阳等, 2011); 通过配制含不同金属离子的产酶培养基分析和探讨不同金属离子对蜜蜂球囊菌几丁质酶活力的影响(刘鹏虎等, 2008), 发现胞外酶在侵染过程中对于蜜蜂组织可进行有效降解从而促进球囊菌对于蜜蜂的侵染进程(李江红等, 2012)。此外, 也通过二代转录组分析揭示了蜜蜂球囊菌的毒力因子基因在病原侵染意大利蜜蜂幼虫和中华蜜蜂幼虫过程的潜在功能(陈大福等, 2017b; 郭睿等, 2017a)。

纳米孔(Nanopore)长读段测序技术是第三代测序技术的代表, 具有超长读长的显著优势, 无需拼接即可直接获取完整的转录本信息, 近年来已成功应用于动植物和微生物的全长转录组的相关研究。近期, 笔者所在团队利用 Nanopore 测序技术对蜜蜂球囊菌的菌丝样品(AaM)和孢子样品(AaS)分别进行测序, 获得了高质量的长读段测序数据, 构建和注释了蜜蜂球囊菌的首个全长转录组(杜宇等, 2021a; 2021b), 系统解析了 AaM 和 AaS 中基因的可变剪接(Alternative splicing, AS)与可变腺苷酸化(Alternative polyadenylation, APA)(杜宇等, 2021c), 并高质量完善了参考基因组的序列和功能注释。目前, 蜜蜂球囊菌的毒力因子研究较为滞后, 毒力因子相关的全长转录本信息仍然缺失。本研究基于前期已获得的牛津纳米孔(Nanopore)长读段测序数据对蜜蜂球囊菌菌丝和孢子中的毒力因子相关全长转录本进行鉴定、定量和结构分析, 研究结果可丰富蜜蜂球囊菌的毒力因子信息, 并

为毒力因子相关剪接异构体(Isoform)的功能研究提供基础。

## 1 材料与方法

### 1.1 Nanopore 长读段测序数据来源

前期研究中, 笔者所在团队利用 Nanopore 测序技术对 AaM 和 AaS 进行测序, 分别测得 6 321 704 和 6 259 727 条原始读段(Raw reads), 经严格质控和过滤得到 5 669 436 和 6 233 159 条有效读段(Clean reads), 进一步鉴定到 9 859 条和 16 795 条非冗余全长转录本, 平均长度分别为 1 187 bp 和 1 303 bp, N50 分别达到 1 482 bp 和 1 658 bp(杜宇等, 2021b)。高质量的 Nanopore 长读段测序数据可用于本研究中蜜蜂球囊菌毒力因子相关全长转录本的鉴定及分析。

### 1.2 毒力因子相关全长转录本的鉴定

本研究综合参考笔者所在团队前期有关蜜蜂球囊菌的生物化学、分子生物学和组学方面的研究结果(刘鹏虎等, 2008; 郑志阳等, 2011; 李江红等, 2012; 郭睿等, 2017a, 2017b; 熊翠玲等, 2020)、前人(Cornman *et al.*, 2012; Svjetlana *et al.*, 2012; Shang *et al.*, 2016)关于蜜蜂球囊菌的研究报道以及真菌毒力因子的相关研究报道(Calderone and Fonzi, 2001; Fang *et al.*, 2009; 罗飞飞等, 2014), 并运用 BLAST 工具将 AaM 和 AaS 中的所有全长转录本比对 Nr(<https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/>)数据库以鉴定蜜蜂球囊菌的毒力因子(几丁质酶、脂肪酶、水解酶和蛋白酶)相关的全长转录本。

### 1.3 毒力因子相关全长转录本的定量与分析

参照杜宇等(2021b)的方法, 先采用 Minimap2 软件(Li, 2018)将 AaM 和 AaS 中的全长转录本序列与蜜蜂球囊菌参考基因组(Assembly AAP 1.0)注释的已知转录本序列进行比对, 以获取全长转录本与已知转录本的对应信息。因 Nanopore 全长转录组测序可以模拟成一个随机抽样的过程, 故先对各样品中的比对上蜜蜂球囊菌参考基因组的全长转录本序列数量

进行归一化处理以真实反映转录本的表达量,再通过每百万计数法 (Counts per million, CPM) (Zhou *et al.*, 2014) 计算每一条全长转录本的表达量, 公式为:  $CPM = \text{比对到转录本上的读段数 (Reads mapped to transcript)} / \text{样本中对齐的总读段数 (Total reads aligned in sample)} \times 1\,000\,000$ 。利用百迈克云平台 (<https://international.biocloud.net>) 的相关工具绘制转录本的表达量聚类热图。参照 Sylvain 和 Michael (2007) 的方法, 通过 IGV (Integrative genomics viewer) 浏览器对毒力因子相关全长转录本进行可视化。

## 2 结果与分析

### 2.1 蜜蜂球囊菌菌丝和孢子中毒力因子相关全长转录本的表达谱

在 AaM 中鉴定到 60 条毒力因子相关全长转

录本, 它们的表达量介于 0.17-244.91, 其中表达量最高的全长转录本分别为 ONT.5196.1 (溶血磷脂酶), 其次为 ONT.1648.1 (磷脂酶 D/转磷脂酶) 和 ONT.4754.1 (脂肪酶)。在 AaS 中鉴定到 77 条毒力因子相关全长转录本, 它们的表达量介于 0.16-931.93, 其中表达量最高的全长转录本为 ONT.5196.1 (溶血磷脂酶), 其次为 ONT.7545.1 (磷脂酶 PldA) 和 ONT.1436.2 (脂肪酶)。括号内文字为转录本对应蛋白在 Nr 数据库的注释。

进一步分析发现 AaM 和 AaS 特有的全长转录本分别有 0 条和 17 条, 共有的全长转录本有 60 条。表达量聚类分析结果显示上述共有全长转录本在 AaM 和 AaS 中呈现出不同的表达谱, 部分全长转录本在 AaS 中的表达量更高, 但更多的全长转录本在 AaM 中具有更高的表达水平 (图 1)。

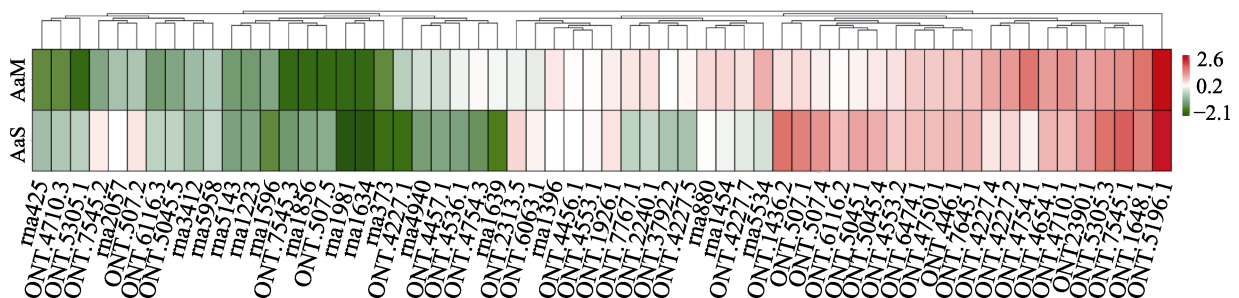


图 1 AaM 和 AaS 共有的毒力因子相关全长转录本的表达量聚类

Fig. 1 Expression clustering of virulence factor-associated full-length transcripts in AaM and AaS

### 2.2 蜜蜂球囊菌菌丝中毒力因子相关全长转录本的统计与分析

在 AaM 中鉴定到几丁质酶相关的 12 条全长转录本, 包括 ONT.4227.2 (内切几丁质酶 1)、ONT.1926.1 (几丁质酶 3)、ONT.4553.2 (几丁质结合, 结构域 3) 和 rna4940 (几丁质结合, 1 型) 等; 脂肪酶相关的 48 条全长转录本, 包括 ONT.5196.1 (溶血磷脂酶) 和 rna1981 (溶血磷脂酶) 等。几丁质酶相关全长转录本的详细信息如表 1 所示; 水解酶相关的 289 条全长转录本, 包括 rna4522 (糖基水解酶族 43, 五叶  $\beta$ -推进器结构域蛋白) 和 ONT.60.1 (糖基水解酶族 17) 等; 蛋白酶相关的 58 条全长转录本条, 包括

(ONT.6150.2 (碱性丝氨酸蛋白酶) 和 ONT.5041.1 (自噬性丝氨酸蛋白酶 Alp2) 等。括号内文字为转录本对应蛋白在 Nr 数据库的注释。AaM 中几丁质酶相关全长转录本的详细信息如表 1 所示; 部分蛋白酶相关全长转录本的详细信息如表 2 所示。

### 2.3 蜜蜂球囊菌孢子中毒力因子相关全长转录本的统计与分析

在 AaS 中鉴定到几丁质酶相关的 14 条全长转录本, 包括 ONT.4227.2 (内切几丁质酶 1)、ONT.4227.6 (内切几丁质酶 1)、ONT.4553.2 (几丁质结合, 结构域 3) 和 rna1466 (几丁质结合, 结构域 3) 等; 脂肪酶相关的 63 条全长转录本,

表 1 AaM 中几丁质酶相关全长转录本的详细信息

Table 1 Detailed information of full-length transcripts related to chitinase in AaM

| 基因 ID<br>Gene ID | 转录本 ID<br>Transcript ID | AaM 中 CPM 值<br>CPM value in AaM | Nr 数据库注释<br>Nr database annotation |
|------------------|-------------------------|---------------------------------|------------------------------------|
| gene1223         | ONT.4227.2              | 22.248 960                      | Endochitinase 1                    |
| gene1223         | ONT.4227.4              | 18.251 100                      | Endochitinase 1                    |
| gene1466         | ONT.4553.2              | 7.474 260                       | Chitin-binding, domain 3           |
| gene1223         | ONT.4227.7              | 7.474 260                       | Endochitinase 1                    |
| gene4940         | ONT.7767.1              | 7.126 620                       | Chitin-binding, type 1             |
| gene5143         | ONT.1926.1              | 6.257 520                       | Chitinase 3                        |
| gene1223         | ONT.4227.5              | 5.736 060                       | Chitinase                          |
| gene1466         | ONT.4553.1              | 5.388 420                       | Chitin-binding, domain 3           |
| gene4940         | rna4940                 | 2.781 120                       | Chitin-binding, type 1             |
| gene1223         | ONT.4227.1              | 1.912 020                       | Endochitinase 1                    |
| gene1223         | rna1223                 | 0.521 460                       | Endochitinase 1                    |
| gene5143         | rna5143                 | 0.521 460                       | Chitinase 3                        |

表 2 AaM 中部分蛋白酶相关的部分全长转录本的详细信息

Table 2 Detailed information of partial full-length transcripts related to protease in AaM

| 基因 ID<br>Gene ID | 转录本 ID<br>Transcript ID | AaM 中 CPM 值<br>CPM value in AaM | Nr 数据库注释<br>Nr database annotation                   |
|------------------|-------------------------|---------------------------------|------------------------------------------------------|
| gene3164         | ONT.6150.2              | 955.140 883                     | Alkaline serine protease                             |
| gene1853         | ONT.5041.1              | 184.423 017                     | Autophagic serine protease Alp2                      |
| gene3605         | ONT.1206.1              | 51.276 899                      | ADAM family of metalloprotease ADM-B                 |
| gene6255         | ONT.3083.4              | 49.886 339                      | 26S protease subunit rpt4                            |
| gene3867         | ONT.6685.1              | 30.592 319                      | Acid protease                                        |
| gene92           | ONT.101.1               | 26.594 460                      | Aminopeptidase I zinc metalloprotease family protein |
| gene4877         | rna4877                 | 25.899 180                      | Ubiquitin-specific protease                          |
| gene3164         | rna3164                 | 21.727 500                      | Alkaline serine protease                             |
| gene3062         | ONT.6023.1              | 18.598 740                      | Ulp1 protease familyt                                |
| gene2554         | ONT.5387.1              | 15.643 800                      | ATP-dependent protease La                            |

包括 ONT.5196.1 (溶血磷脂酶) 和 ONT.4536.2 (溶血磷脂酶) 等; 水解酶相关的 267 条全长转录本, 包括 ONT.7198.1 (富马酰乙酰乙酸水解酶 FahA) 和 ONT.60.1 (糖基水解酶族 17) 等; 蛋白酶相关的全长转录本 56 条, 包括 ONT.6150.2 (碱性丝氨酸蛋白酶) 和 ONT.5041.1 (自噬性丝氨酸蛋白酶 Alp2) 等。AaS 中几丁质酶相关全长转录本的详细信息如表 3 所示; 部分蛋白酶相关全长转录本的详细信息如表 4 所示。

2.4 蜜蜂球囊菌毒力因子相关全长转录本的结构分析

进一步对鉴定到的毒力因子相关全长转录本进行结构分析, 结果显示与蜜蜂球囊菌参考基因组注释的转录本相比, 绝大多数毒力因子编码基因对应的全长转录本不仅数量更多, 结构也更为复杂。例如蜜蜂球囊菌的内切几丁质酶基因 (gene1223) 在参考基因组中仅注释了 1 条转录

表 3 AaS 中几丁质酶相关全长转录本的详细信息  
Table 3 Detailed information of full-length transcripts related to chitinase in AaS

| 基因 ID<br>Gene ID | 转录本 ID<br>Transcript ID | AaS 中 CPM 值<br>CPM value in AaS | Nr 数据库注释<br>Nr database annotation |
|------------------|-------------------------|---------------------------------|------------------------------------|
| gene1466         | ONT.4553.2              | 35.014 588                      | Chitin-binding, domain 3           |
| gene1223         | ONT.4227.2              | 28.360 232                      | Endochitinase 1                    |
| gene5143         | ONT.1926.1              | 23.131 809                      | Chitinase 3                        |
| gene1223         | ONT.4227.4              | 22.181 187                      | Endochitinase 1                    |
| gene1466         | ONT.4553.1              | 16.635 890                      | Chitin-binding, domain 3           |
| gene1223         | ONT.4227.7              | 9.823 097                       | Endochitinase 1                    |
| gene4940         | ONT.7767.1              | 5.386 860                       | Chitin-binding, type 1             |
| gene1223         | ONT.4227.5              | 3.327 178                       | Chitinase                          |
| gene1223         | rna1223                 | 1.425 933                       | Endochitinase 1                    |
| gene5143         | rna5143                 | 1.267 496                       | Chitinase 3                        |
| gene4940         | rna4940                 | 1.109 059                       | Chitin-binding, type 1             |
| gene1223         | ONT.4227.1              | 0.316 874                       | Endochitinase 1                    |
| gene1223         | ONT.4227.6              | 0.316 874                       | Endochitinase 1                    |
| gene1466         | rna1466                 | 0.158 437                       | Chitin-binding, domain 3           |

表 4 AaS 中蛋白酶相关的部分全长转录本的详细信息  
Table 4 Detailed information of partial full-length transcripts related to protease in AaS

| 基因 ID<br>Gene ID | 转录本 ID<br>Transcript ID | AaS 中 CPM 值<br>CPM value in AaS | Nr 数据库注释<br>Nr database annotation                   |
|------------------|-------------------------|---------------------------------|------------------------------------------------------|
| gene3164         | ONT.6150.2              | 955.140 883                     | Alkaline serine protease                             |
| gene1853         | ONT.5041.1              | 184.423 017                     | Autophagic serine protease Alp2                      |
| gene6255         | ONT.3083.4              | 51.276 899                      | 26S protease subunit rpt4                            |
| gene3867         | ONT.6685.1              | 49.886 339                      | Acid protease                                        |
| gene3062         | ONT.6023.4              | 30.592 319                      | Ulp1 protease familyt                                |
| gene3605         | ONT.1206.1              | 26.594 460                      | ADAM family of metalloprotease ADM-B                 |
| gene92           | ONT.101.1               | 25.899 180                      | Aminopeptidase I zinc metalloprotease family protein |
| gene3062         | ONT.6023.1              | 21.727 500                      | Ulp1 protease familyt                                |
| gene5608         | ONT.2323.1              | 18.598 740                      | 26S protease regulatory subunit 6B                   |
| gene1425         | ONT.4493.4              | 15.643 800                      | Cysteine protease                                    |

本,但本研究鉴定到该基因的 6 条转录本,它们与参考基因组注释的转录本结构均存在差异(图 2: A); 磷脂酶 D 基因 (gene2057) 在参考基因组中同样仅注释了 1 条转录本,但本研究鉴定到该基因的 4 条转录本,它们的结构与参考基因组注释转录本的结构也都具有差异 (图 2: B)。

3 讨论

蜜蜂球囊菌作为专性侵染蜜蜂幼虫的真菌病原,主要依靠菌丝的机械力穿透围食膜、肠壁和体壁杀死蜜蜂,侵染期间通过病原合成与分泌的毒力因子的降解作用损害蜜蜂机体( Aronstein

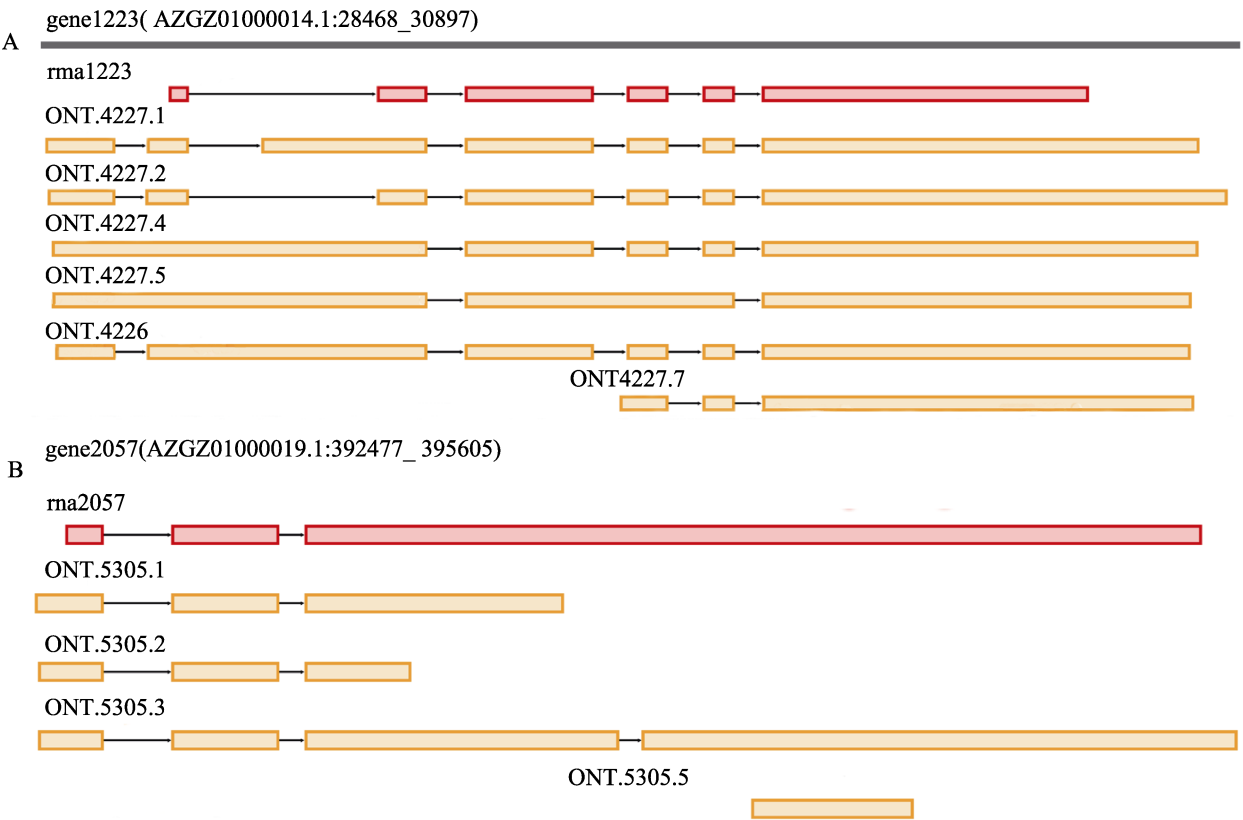


图 2 蜜蜂球囊菌几丁质酶基因 ( A ) 和磷脂酶 D 基因 ( B ) 对应转录本结构的 IGV 浏览器视图  
Fig. 2 IGV browser view of structures of transcripts corresponding to chitinase gene ( A )  
and phospholipase D gene ( B ) in *Ascosphaera apis*

and Murray, 2010)。由于技术手段的限制，早期关于蜜蜂球囊菌毒力因子的研究主要依赖于酶活测定的方法，导致相关分子生物学的研究相对滞后。Cornman 等（2012）通过二代测序技术在蜜蜂球囊菌中鉴定到 3 个几丁质酶基因、30 个酯酶基因、42 个蛋白酶基因和 24 个脂肪酶基因。由于二代测序获得的读段较短，导致蜜蜂球囊菌参考基因组注释的 9 481 个基因均缺少 3'UTR 和 5'UTR（Shang *et al.*, 2016）。因此，基于已注释基因进行基因克隆难以获得全长序列，而 miRNA 等调控因子多通过结合 UTR 的部分序列发挥作用。前期研究中，笔者团队基于二代转录组数据分别延长了 51 个基因的 5'UTR 和 50 个基因的 3'UTR，并鉴定到 373 个新基因（郭睿等，2019）；近期又利用 Nanopore 长读段数据优化了蜜蜂球囊菌参考基因组上的已注释基因，分别延长了 4 744 和 4 737 个已注释基因 3'UTR 和 5'UTR，并鉴定到 1 558 个新基因（杜宇等，

2020）。ONT 测序技术能够直接测得转录本的全长序列，包括 CDS 区、5'UTR 和 3'UTR（Senol *et al.*, 2019）。基于已获得的高质量 ONT 测序数据，本研究鉴定到 AaM 中毒力因子相关的 367 个基因和 407 条全长转录本，AaS 中毒力因子相关的 367 个基因和 486 条全长转录本。上述鉴定到的毒力因子相关基因和转录本序列可为后续功能研究的深入开展提供数据基础。

现有的蜜蜂球囊菌参考基因组共注释了 6 442 个基因，每个基因对应 1 条转录本。而真核生物基因一般都会通过 AS 形成不同的剪接异构体，进而翻译出不同或相似结构和功能的蛋白质，发挥不同或相似的生物学功能（Ule *et al.*, 2019）。受制于二代测序产生的短读段，以往精准鉴定剪接异构体的难度很大。Nanopore 长读段测序技术在鉴定 AS 事件和剪接异构体方面展现出强大的能力，已广泛应用与人类 *Homo sapiens*、绿鳍马面鲉 *Thamnaconus septentrionalis*



和拟南芥 *Arabidopsis thaliana* 等动植物和微生物的相关研究( Bian *et al.*, 2020; Cui *et al.*, 2020; Tang *et al.*, 2020)。前期研究中, 笔者所在团队首次通过 Nanopore 测序技术对蜜蜂球囊菌和东方蜜蜂微孢子虫 *Nosema ceranae* 的基因和转录本结构进行了深入分析, 在蜜蜂球囊菌的菌丝和孢子中分别鉴定到 286 和 559 次 AS 事件( 杜宇等, 2021c), 在东方蜜蜂微孢子虫孢子中鉴定到 5 次 AS 事件( 陈华枝等, 2021), 体现出 Nanopore 测序技术在挖掘剪接异构体方面的优越性。本研究中发现, 与蜜蜂球囊菌参考基因组注释的转录本相比, 绝大多数毒力因子基因对应的全长转录本不仅数量更多, 而且结构也更为复杂; 其中内切几丁质酶基因( gene1223) 和磷脂酶 D 基因( gene2057) 可通过 AS 分别产生的 6 条和 4 条剪接异构体。这表明蜜蜂球囊菌可利用有限数量的毒力因子基因通过 AS 形成数量可观的剪接异构体, 进而翻译出具有特定生物学功能的蛋白质。同一毒力因子基因转录形成的不同剪接异构体在蜜蜂球囊菌的侵染过程中扮演的具体角色是什么? 不同的剪接异构体翻译出的蛋白质究竟发挥不同还是相似的功能? 这些有趣的科学问题是下一步的研究重点。

## 4 结论

在蜜蜂球囊菌菌丝中鉴定到毒力因子几丁质酶、脂肪酶、水解酶和蛋白酶相关的 12、48、289 和 58 条全长转录本; 在孢子中鉴定到几丁质酶、脂肪酶、水解酶和蛋白酶相关的 14、63、267 和 56 条全长转录本; 相比于蜜蜂球囊菌参考基因组注释的转录本, 绝大多数毒力因子基因对应的全长转录本数量更多且结构更为复杂。

## 参考文献 (References)

- Aronstein KA, Murray KD, 2010. Chalkbrood disease in honey bees. *Journal of Invertebrate Pathology*, 103(Suppl. 1): S20–S29.
- Bian L, Li F, Ge J, Wang P, Chang Q, Zhang S, Li J, Liu C, Liu K, Liu X, Li X, Chen H, Chen S, Shao C, Lin Z, 2020. Chromosome-level genome assembly of the greenfin horse-faced filefish (*Thamnaconus septentrionalis*) using Oxford Nanopore PromethION sequencing and Hi-C technology. *Molecular Ecology Resources*, 20(4): 1069–1079.
- Calderone RA, Fonzi WA, 2001. Virulence factors of *Candida albicans*. *Trends Microbiol.*, 9(7): 327–335.
- Carollo CA, Calil AL, Schiave LA, Guaratini T, Roberts DW, Lopes NP, Braga GU, 2010. Fungal tyrosine betaine, a novel secondary metabolite from conidia of entomopathogenic *Metarhizium* spp. fungi. *Fungal Biol.*, 114(5/6): 473–480.
- Chen DF, Guo R, Xiong CL, Liang Q, Zheng YZ, Xu XJ, Huang ZJ, Zhang ZN, Zhang L, Li WD, Tong XY, Xi WJ, 2017a. Transcriptomic analysis of *Ascosphaera apis* stressing larval gut of *Apis mellifera ligustica*. *Acta Entomologica Sinica*, 60(4): 401–411. [陈大福, 郭睿, 熊翠玲, 梁勤, 郑燕珍, 徐细建, 黄枳隼, 张墨楠, 张璐, 李汶东, 童新宇, 席伟军, 2017a. 胁迫意大利蜜蜂幼虫肠道的球囊菌的转录组分析. 昆虫学报, 60(4): 401–411.]
- Chen DF, Guo R, Xiong CL, Liang Q, Zheng YZ, Xu XJ, Zhang ZN, Huang ZJ, Zhang L, Wang H Q, Xie YL, Tong XY, 2017b. Transcriptome of *Apis cerana cerana* larval gut under the stress of *Ascosphaera apis*. *Scientia Agricultura Sinica*, 50(13): 2614–2623. [陈大福, 郭睿, 熊翠玲, 梁勤, 郑燕珍, 徐细建, 张墨楠, 黄枳隼, 张璐, 王鸿权, 解彦玲, 童新宇, 2017b. 中华蜜蜂幼虫肠道响应球囊菌早期胁迫的转录组学. 中国农业科学, 50(13): 2614–2623.]
- Chen HZ, Fan XX, Fan YC, Wang J, Zhu ZW, Jiang HB, Zhang WE, Long Q, Xiong C L, Zhen Y Z, Fu ZM, Chen DF, Guo R, 2021. Analysis of alternative splicing and alternative polyadenylation of *Nosema ceranae* genes based on Nanopore full-length transcriptome dataset. *Mycosystema*, 40(1): 161–173. [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, 2021. 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, 40(1): 161–173.]
- Cornman RS, Bennett AK, Murray KD, Evans JD, Elisk CG, Aronstein K, 2012. Transcriptome analysis of the honey bee fungal pathogen, *Ascosphaera apis*: Implications for host pathogenesis. *BMC Genomics*, 13(1): 285.
- Cui JW, Shen N, Lu ZG, Xu GL, Wang YY, Jin B, 2020. Analysis and comprehensive comparison of PacBio and nanopore-based RNA sequencing of the *Arabidopsis* transcriptome. *Plant Methods*, 16(11): 85.
- Du Y, Jiang HB, Wang J, Fan XX, Wang XN, Feng RR, Zhang WD, Long Q, Xiong CL, Zheng YZ, Chen DF, Guo R, 2021a. Differential expression analysis of full-length transcripts in mycelia and spores of *Ascosphaera apis*. *Acta Entomologica Sinica*, 64(3): 363–373. [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,



- 2021a. 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, 64(3): 363–373.]
- 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, 2021b. Construction and annotation of *ascosphaera apis* full-length transcriptome utilizing nanopore third-generation long-read sequencing technology. *Scientia Agricultura Sinica*, 54(4): 864–876. [杜宇, 祝智威, 王杰, 王秀娜, 蒋海滨, 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, 2021b. 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, 54(4): 864–876.]
- Du Y, Wang J, Jiang HB, Wang XN, Fan Y C, Fan XX, Zhu ZW, Long Q, Zhang WD, Xiong CL, Zheng YZ, Fu ZM, Chen DF, Guo R, 2021c. Analysis of alternative splicing and polyadenylation of *Ascosphaera apis* genes based on third-generation long-read sequencing dataset. *Acta Microbiologica Sinica*, 61(3): 667–682. [杜宇, 王杰, 蒋海滨, 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, 2021c. 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, 61(3): 667–682.]
- Du Y, Fu ZM, Zhu ZW, Wang J, Feng RR, Wang XN, Jiang HB, Fan YC, Fan XX, Xiong CL, Zheng YZ, Xu GJ, Chen DF, Guo R, 2020. Elongation of genic untranslated regions, exploration of SSR loci and identification of unannotated genes and transcripts based on the nanopore sequencing dataset of *Ascosphaera apis*. *Acta Entomologica Sinica*, 63(11): 1345–1357. [杜宇, 付中民, 祝智威, 王杰, 冯睿蓉, 王秀娜, 蒋海滨, 范元婵, 范小雪, 熊翠玲, 郑燕珍, 徐国钧, 陈大福, 郭睿, 2020. 基于蜜蜂球囊菌纳米孔测序数据的基因非翻译区延长、SSR 位点发掘及未注释基因和转录本鉴定. 昆虫学报, 63(11): 1345–1357.]
- Fang WG, Feng J, Fan YH, Zhang YJ, Bidochka MJ, Leger RJ, Pei Y, 2009. Expressing a fusion protein with protease and chitinase activities increases the virulence of the insect pathogen *Beauveria bassiana*. *Journal of Invertebrate Pathology*, 102(2): 155–159.
- Guo R, Xiong CL, Zheng YZ, Zhang L, Tong XY, Liang Q, Chen DF, 2017a. Transcriptome analysis of *Apis mellifera ligustica* larval gut during the early stage of stress induced by *Ascosphaera apis*. *Chinese Journal of Applied Entomology*, 54(4): 553–560. [郭睿, 熊翠玲, 郑燕珍, 张璐, 童新宇, 梁勤, 陈大福, 2017a. 意大利蜜蜂幼虫肠道响应球囊菌早期胁迫的转录组学分析. 应用昆虫学报, 54(4): 553–560.]
- Guo R, Chen DF, Huang ZJ, Liang Q, Xiong CL, Xu XJ, Zheng YZ, Zhang ZN, Tong XY, Hou ZX, Jiang LL, Dao C, 2017b. Transcriptome analysis of *Ascosphaera apis* stressing larval gut of *Apis cerana cerana*. *Acta Microbiologica Sinica*, 57(12): 1865–1878. [郭睿, 陈大福, 黄积隼, 梁勤, 熊翠玲, 徐细建, 郑燕珍, 张翌楠, 解彦玲, 童新宇, 侯志贤, 江亮亮, 刀晨, 2017b. 球囊菌胁迫中华蜜蜂幼虫肠道过程中病原的转录组学研究. 微生物学报, 57(12): 1865–1878.]
- Guo R, Chen HZ, Tong XY, Xiong CL, Zheng YZ, Fu ZM, Xie YL, Wang HP, Zhao HX, Chen DF, 2019. Structural optimization of annotated genes and identification of novel genes in *Ascosphaera apis*. *Journal of China Agricultural University*, 24(1): 61–68. [郭睿, 陈华枝, 童新宇, 熊翠玲, 郑燕珍, 付中民, 解彦玲, 王海朋, 赵红霞, 陈大福, 2019. 蜜蜂球囊菌基因结构优化及新基因鉴定. 中国农业大学学报, 24(1): 61–68.]
- Li H, 2018. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18): 3094–3100.
- Li FX, Li JH, Liang Q, Chen DF, 2011. Characteristics of *Ascosphaera apis* extracellular proteinase. *Scientia Agricultura Sinica*, 44(6): 1247–1254. [李飞雄, 李江红, 梁勤, 陈大福, 2011. 蜜蜂球囊菌几种胞外蛋白酶的特性. 中国农业科学, 44(6): 1247–1254.]
- Li JH, Zheng ZY, Chen DF, Liang Q, 2012. Factors influencing *Ascosphaera apis* infection on honeybee larvae and observation on the infection process. *Acta Entomologica Sinica*, 55(7): 790–797. [李江红, 郑志阳, 陈大福, 梁勤, 2012. 影响蜜蜂球囊菌侵染蜜蜂幼虫的因素及侵染过程观察. 昆虫学报, 55(7): 790–797.]
- Liu PJ, Zhou G, Chen DF, Liang Q, 2008. Effects of metal ions on the activity of the chitinase from *Ascosphaera apis*. *Journal of Fujian Agriculture and Forestry University (Natural Science Edition)*, 37(1): 96–98. [刘鹏虎, 周刚, 陈大福, 梁勤, 2008. 金属离子对蜜蜂球囊菌几丁质酶活力的影响. 福建农林大学学报(自然科学版), 37(1): 96–98.]
- Luo FF, Li SL, Chen LY, Zhang WC, Wang B, Huang B, Hu FL, 2014. Identification of spore germination and virulence related biomarkers from *Beauveria bassiana* using an LC-MS-based metabolomic technique. *Acta Microbiologica Sinica*, 54(1): 33–41. [罗飞飞, 李淑林, 陈龙云, 张万存, 王滨, 黄勃, 胡丰林, 2014. 代谢组学方法鉴定球孢白僵菌孢子萌发和杀虫毒力相关的标记物. 微生物学报, 54(1): 33–41.]
- Qin X, Evans JD, Aronstein KA, Murray KD, Weinstock GM, 2006. Genome sequences of the honey bee pathogens *Paenibacillus larvae* and *Ascosphaera apis*. *Insect Molecular Biology*, 15(5): 715–718.
- Senol CD, Kim JS, Ghose S, Alkan C, Mutlu O, 2019. Nanopore sequencing technology and tools for genome assembly: computational analysis of the current state, bottlenecks and future directions. *Brief. Bioinform.*, 20(4): 1542–1559.
- Shang YF, Xiao GH, Zheng P, Cen K, Zhan S, Wang CS, 2016. Divergent and convergent evolution of fungal pathogenicity. *Genome Biology and Evolution*, 8(5): 1374–1387.

- Svjetlana V, Jacobus JB, Jørgen E, Annette BJ, 2012. Virulence of mixed fungal infections in honey bee brood. *Bio. Med. Central*, 9(1): 5.
- Sylvain F, Michael S, 2007. Astalavista: Dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Research*, 35(Web server issue): W297–W299.
- Tang AD, Soulette CM, Van Baren MJ, Hart K, Hrabeta-Robinson E, Wu CJ, Brooks AN, 2020. Full-length transcript characterization of *SF3B1* mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nature Communication*, 11(1): 1438.
- Ule J, Blencowe BJ, 2019. Alternative splicing regulatory networks: Functions, mechanisms, and evolution. *Molecular Cell*, 76(2): 329–345.
- Xiong CL, Du Y, Feng RR, Jiang HB, Shi XY, Wang HP, Fan XX, Wang J, Zhu ZW, Fan YC, Chen HZ, Zhou DD, Zheng YZ, Chen DF, Guo R, 2020. Differential expression pattern and regulation network of microRNAs in *Ascosphaera apis* invading *Apis cerana cerana* 6-day-old larvae. *Acta Microbiologica Sinica*, 60(5): 992–1009. [熊翠玲, 杜宇, 冯睿蓉, 蒋海宾, 史小玉, 王海朋, 范小雪, 王杰, 祝智威, 范元婵, 陈华枝, 周丁丁, 郑燕珍, 陈大福, 郭睿, 2020. 侵染中华蜜蜂 6 日龄幼虫的蜜蜂球囊菌的微小 RNA 差异表达谱及调控网络. 微生物学报, 60(5): 992–1009.]
- Zhang YJ, Zhang JQ, Jiang XD, Wang GJ, Luo ZB, FAN YH, WU ZQ, Yan P, 2010. Requirement of a mitogen-activated protein kinase for appressorium formation and penetration of insect cuticle by the entomopathogenic fungus *Beauveria bassiana*. *Applied and Environmental Microbiology*, 76(7): 2262–2270.
- Zhang ZN, Xiong CL, Xu XJ, Huang ZJ, Zheng YZ, Luo Q, Liu M, Li WD, Tong XY, Zhang Q, Liang Q, Guo R, Chen DF, 2017. De novo assembly of a reference transcriptome and development of SSR markers for *Ascosphaera apis*. *Acta Entomologica Sinica*, 60(1): 34–44. [张墨楠, 熊翠玲, 徐细建, 黄枳隼, 郑燕珍, 骆群, 刘敏, 李汶东, 童新宇, 张琦, 梁勤, 郭睿, 陈大福, 2017. 蜜蜂球囊菌的参考转录组 de novo 组装及 SSR 分子标记开发. 昆虫学报, 60(1): 34–44.]
- Zheng ZY, Li JH, Liang Q, Chen DF, 2011. *Ascosphaera apis* secretes multiple extracellular enzymes to infect honeybee larvae. *Journal of Fujian Agriculture and Forestry University (Natural Science Edition)*, 40(3): 280–284. [郑志阳, 李江红, 梁勤, 陈大福, 2011. 蜜蜂球囊菌分泌多种胞外酶侵染蜜蜂幼虫. 福建农林大学学报(自然科学版), 40(3): 280–284.]
- Zhou X, Lindsay H, Robinson MD, 2014. Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic Acids Research*, 42(11): e91.