

意大利蜜蜂细胞内吞相关基因 与全长转录本鉴定*

范小雪^{1**} 王思懿^{1**} 朱乐冉¹ 荆欣¹ 郭思佳¹ 牛庆生³
蒋海宾³ 徐细建⁴ 陈大福^{1,2} 付中民^{1,2} 徐国钧^{1***} 郭睿^{1,2***}

(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. 吉林省养蜂科学研究所, 吉林 132000; 4. 江西省养蜂研究所, 南昌 330000)

摘要 【目的】利用已获得的纳米孔(Nanopore)长读段测序数据对意大利蜜蜂 *Apis mellifera ligustica* 的细胞内吞相关基因和全长转录本进行鉴定、分析和验证, 为深入开展相关研究提供参考和依据。【方法】通过 Blast 工具将意大利蜜蜂全长转录本比对 Nr 和 KEGG 数据库以筛选出细胞内吞相关基因和全长转录本。采用 gffcompare 软件将全长转录本与西方蜜蜂 *Apis mellifera* 参考基因组(Amel_HAv3.1)上注释的转录本进行比较以优化已注释基因的结构。使用 Astalavista 软件鉴定细胞内吞相关基因的可变剪接(Alternative splicing, AS)事件类型, 利用 IGV 浏览器对剪切体结构进行可视化, 通过 PCR 验证 AS 事件的真实性。利用 TAPIS pipeline 进行可变多聚腺苷酸化(Alternative polyadenylation, APA)位点预测和分析, 并通过 MEME 软件鉴定 APA 位点上游的基序(motif)。【结果】共鉴定到意大利蜜蜂细胞内吞相关的 170 个基因与 991 条转录本。共优化了 89 个细胞内吞相关基因的结构, 其中 5'端延长的基因有 36 个, 3'端延长的基因有 35 个, 5'端和 3'端同时延长的基因有 18 个。鉴定到细胞内吞相关的 28 个基因的 666 次 AS 事件, 其中最丰富的 AS 类型是 3'端可变剪接(Alternative 3' splice site, A3SS)。PCR 结果证实了涉及整合素 β -nu 样异构体基因、Ras 样 GTP 结合蛋白基因和 rab11 家族相互作用蛋白 4 基因的 2 种类型的 4 次 AS 事件真实性。共鉴定到 120 个细胞内吞相关的基因含有 1 个及以上的 APA 位点, 并在 APA 位点上游鉴定到多个 motif, 一致性序列为 NBMNHHHBMNSNNCBNVSNNNNNNNNNNVNNNN。【结论】鉴定到意大利蜜蜂细胞内吞相关的 170 个基因及 991 条全长转录本, 发现细胞内吞相关基因发生大量的 AS 事件并含有丰富的 APA 位点, 优化了西方蜜蜂参考基因组上注释的细胞内吞相关基因的结构。
关键词 意大利蜜蜂; 细胞内吞; 纳米孔测序; 全长转录本; 可变剪切

Identification of endocytosis-associated genes, and full-length transcripts, in *Apis mellifera ligustica*

FAN Xiao-Xue^{1**} WANG Si-Yi^{1**} ZHU Le-Ran¹ JING Xin¹ GUO Si-Jia¹ NIU Qing-Sheng³
JIANG Hai-Bin³ XU Xi-Jian⁴ CHEN Da-Fu^{1,2} FU Zhong-Min^{1,2} XU Guo-Jun^{1***} GUO Rui^{1,2***}

(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. Apiculture Science
Institute of Jilin Province, Jilin 132000; 4. Apicultural Research Institute of Jiangxi Province, Nanchang 330000, China)

Abstract [Objectives] To identify, analyze and confirm the genes and full-length transcripts associated with endocytosis in

*资助项目 Supported projects: 国家自然科学基金面上项目(32172792); 国家现代农业产业技术体系专项资金(CARS-44-KXJ7); 福建省自然科学基金面上项目(2022J01131334); 宁夏回族自治区重点研发项目(2022BBF02037); 福建农林大学硕士生导师团队项目(郭睿); 福建农林大学科技创新专项基金(郭睿); 福建农林大学动物科学学院(蜂学学院)科研扶持项目(郭睿)

**共同第一作者 Co-first authors, E-mail: imfanxx@163.com; siyiwang2021@163.com

***共同通讯作者 Co-corresponding authors, E-mail: ruiguo@fafu.edu.cn; 000q060012@fafu.edu.cn

收稿日期 Received: 2022-07-22; 接受日期 Accepted: 2023-04-13

Apis mellifera ligustica using previously obtained Nanopore long-read sequencing data, thereby providing a foundation for further functional study of these genes. **[Methods]** All full-length transcripts available for *Apis mellifera ligustica* were Blast-searched in the Nr and KEGG databases to screen out endocytosis-related genes and full-length transcripts. Structures of annotated genes were optimized by using gffcompare software to compare endocytosis-associated full-length transcripts with annotated transcripts in the *A. mellifera* reference genome (Amel_HAv3.1). Types of AS events in genes related to endocytosis were identified using Astalavista software. Visualization of isoforms' structure was performed with an IGV browser. The authenticity of AS events was verified using PCR. Prediction and investigation of APA sites were done using a TAPIS pipeline, followed by identification of motifs upstream of APA sites with MEME software. **[Results]** In total, 170 genes and 991 full-length transcripts relevant to endocytosis were identified. Structures of 89 annotated genes were optimized, among these, 5' ends of 36 genes were prolonged, 3' ends of 35 genes were prolonged, and both 5' ends and 3' ends of 15 genes were prolonged. In addition, 666 AS events in 28 genes were identified, the most common of which was an alternative 3' splice site. PCR results validated the authenticity of four randomly selected AS events involving an integrin beta-nu-like isoform, ras-like GTP-binding protein and Ras-like GTP-binding protein. Moreover, 120 genes were identified containing one or more APA sites in endocytosis, the most abundant of which were genes containing three APA sites (14, 11.67%). Several motifs upstream of APA sites were observed, with the consensus sequence: NBMNHHHBMNSNNCBNVSNNNNNNNNNNNVNNNN. **[Conclusion]** 170 genes and 991 full-length transcripts associated with endocytosis in *A. m. ligustica* were identified. A number of AS events in endocytosis-related genes that also contained abundant APA sites were detected, and structures of endocytosis-related genes annotated in the reference genome of *A. mellifera* were optimized.

Key words *Apis mellifera ligustica*; endocytosis; Nanopore sequencing; full-length transcript; alternative splicing

蜜蜂是一种无脊椎动物,因具有真社会性而营群居性生活,其免疫防御系统可分为体液免疫和细胞免疫,后者主要由细胞介导的反应组成,如吞噬反应、成瘤、包囊和伤口闭合(梁勤和陈大福, 2009; Negri *et al.*, 2016)。细胞内吞可进一步分为网格蛋白依赖的内吞作用(Clathrin-dependent endocytosis)、细胞膜穴样凹陷(Caveolae)及细胞吞噬作用(Phagocytosis),其中吞噬体成熟是吞噬细胞功能的正常组成部分。细胞内吞在营养吸收、受体内化、细胞免疫和细胞信号转导等过程起到关键作用(Hinze and Boucrot, 2018; Mbengue *et al.*, 2016)。研究表明真核生物可通过细胞内吞作用抵御病原微生物入侵(Lin and Guttman, 2010)。在昆虫中,细胞内吞已被证实参与果蝇 *Drosophila* 和家蚕 *Bombyx mori* 等昆虫对病原微生物的免疫防御(Huang *et al.*, 2010)。Chen 等(2018)研究发现家蚕质型多角体病毒 BmCPV 经内吞作用介导进入家蚕卵巢 BmN 细胞,丹西卡维林、氯丙嗪、染料木素和 PP2 这 4 种内吞抑制剂能阻断 BmCPV 进入细胞的途径,并显著降低 BmCPV 的感染性。目前,蜜蜂细胞内吞方面的研究较为

滞后。最新的西方蜜蜂 *Apis mellifera* 参考基因组(Amel_HAv3.1)已经公布,基因组大小约 250 Mb,共有 12 374 个编码基因,其中包含细胞内吞相关的 168 个基因。

近年来,第三代测序技术逐渐兴起且发展迅速,纳米孔(Nanopore)长读段测序技术作为代表,因具有超长读长和直接读取核酸修饰的显著优势已成为动植物和微生物全长转录组研究的有力工具(Lu *et al.*, 2016)。AS 类型主要包括 5 种,分别为互斥外显子(Mutually exclusive exon, MEE)、内含子保留(Intron retention, IR)、外显子跳跃(Exon skipping, ES)、可变 3'端剪接(Alternative 3' splice site, A3SS)和可变 5'端剪接(Alternative 5' splice site, A5SS),Nanopore 测序技术在精确鉴定和分析结构变异方面表现出巨大优势,已成功应用于人(Jain *et al.*, 2018)、家山羊 *Capra hircus* (Bickhart *et al.*, 2017)、秀丽隐杆线虫 *Caenorhabditis elegans* (Tyson *et al.*, 2018)等物种的可变剪接(Alternative splicing, AS)与可变多聚腺苷酸化(Alternative polyadenylation, APA)研究。此前,笔者所在团队利用 Nanopore 测序技术对蜜蜂的 2 种真菌病原蜜蜂球囊菌

Ascosphaera apis 和东方蜜蜂微孢子虫 *Nosema ceranae* 进行了细致深入的全长转录组研究, 构建和注释了二者的全长转录组, 高质量完善了参考基因组注释(陈华枝等, 2020; 杜宇等, 2021b), 并系统鉴定了基因的可变剪接事件和可变多腺苷酸化位点(陈华枝等, 2021; 杜宇等, 2021a)。目前, 西方蜜蜂全长转录组滞后且有限。He 等(2022)基于 Nanopore 测序发现西方蜜蜂工蜂和蜂王之间差异异构体表达的程度, 转录剪接的灵活性和多腺苷酸化皆存在差异; 范小雪等(2023)通过 Nanopore 测序揭示意大利蜜蜂 *Apis mellifera ligustica* (简称意蜂) 丝氨酸/苏氨酸蛋白激酶相关基因和全长转录本进行鉴定和分析。

本研究基于已获得的高质量 Nanopore 测序数据对意蜂细胞内吞相关基因和全长转录本进行鉴定, 优化西方蜜蜂参考基因组已注释细胞内吞相关基因的结构, 分析和验证细胞内吞相关基因的 AS 事件和 APA 位点, 以期对意蜂细胞内吞相关基因及全长转录本信息提供有益补充, 并为细胞内吞相关基因和剪接体 (Isoform) 的功能研究提供数据资源和基础。

1 材料与方法

1.1 供试生物材料

意蜂工蜂取自福建农林大学动物科学学院(蜂学学院)教学蜂场的实验蜂群。

1.2 细胞内吞相关基因和全长转录本的筛选与分析

前期研究中, 笔者所在团队已完成意蜂中肠样品制备、RNA 提取、cDNA 建库及 Nanopore 测序(范小雪等, 2023)。高质量测序数据可用于本研究中意蜂细胞内吞相关基因和全长转录本的鉴定、分析和验证。通过 Blast 工具将鉴定到的所有全长转录本序列比对到 Nr 数据库(<https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/>)及 KEGG (<https://www.omicsshare.com/tools/Home/Soft/pathtwaygsea>) 数据库, 然后根据数据库注释信息筛选细胞内吞相关基因和全长转录本。

1.3 西方蜜蜂参考基因组中已注释细胞内吞相关基因的结构优化

参照杜宇等(2020)的方法, 使用 gffcompare 软件将鉴定到的细胞内吞相关全长转录本与西方蜜蜂参考基因组 (Amel_HAv3.1) 中已注释的基因和转录本进行比较, 从而优化已注释基因的结构。如果在已注释基因边界之外的区域有 Mapped reads 支持, 将基因的非翻译区 (Untranslated region, UTR) 向上游或 (和) 下游延伸以修正基因边界。

1.4 细胞内吞相关基因的可变剪接事件鉴定与分析

参照陈华枝等(2021)的方法, 通过 Astalavista 软件鉴定细胞内吞相关基因的 AS 事件类型, 再根据预测结果统计不同类型 AS 事件的数目。

1.5 细胞内吞相关基因的 AS 事件的 PCR 验证

随机选择发生 ES 的 3 个基因 (LOC100576742, LOC409910 和 LOC409274) 和发生 IR 的 1 个基因 (LOC409274) 进行 PCR 验证。根据上述 4 基因的序列, 利用 Primer Premier 软件 (v.5.0) 设计跨剪接位点的特异性引物 (表 1), 委托生工生物 (上海) 公司进行合成。利用 RNA 抽提试剂盒 (Promega, 美国) 提取上述工蜂中肠样品的总 RNA, 再使用 Hifair® II 1st Strand cDNA Synthesis Kit (Yeast, 上海) 进行 cDNA 合成。以肌动蛋白基因 *actin* (GeneBank ID: LOC107999330) 作为内参。PCR 扩增在 T100 热循环仪上进行。反应体系和程序参照范小雪等(2023)进行设置。产物经 3% 琼脂糖凝胶电泳后用核酸凝胶成像仪 (上海培清, 中国) 进行检测和拍照。

1.6 细胞内吞相关基因的 APA 位点鉴定与分析

参照杜宇等(2021a)方法利用 TAPIS pipeline (Abdel-Ghany *et al.*, 2016) 鉴定细胞内吞相关基因的 APA 位点, 参数设置为 -norc、-meme-minw 6、-meme-maxw 6、-spamo-skip、-fimo-skip。通过 MEME 软件 (Bailey *et al.*, 2009) 对细胞内吞相关基因 APA 位点上游 50 bp 的序列特征进行分析以鉴定 motif。

表 1 本研究中使用的引物序列信息
Table 1 Information of primer sequences used in this study

引物名称 Primers	序列信息 (5'-3') Sequence information (5'-3')
ONT.593.23-ONT.593.27-F	TGCGTTACGAATCCAC
ONT.593.23-ONT.593.27-R	ACCATCACCCACAATAA
ONT.593.33-ONT.593.27-F	TGCGTTACGAATCCAC
ONT.593.33-ONT.593.27-R	TCCACTTGTTTCCCATC
rna14121-ONT.3771.4-F	ATTTATGTGCGGGTGT
rna14121-ONT.3771.4-R	GTGCTCGGTAAGATGG
ONT.5590.10-ONT.5590.12-F	ACATTTCGCTGGTGAT
ONT.5590.10-ONT.5590.12-R	ACTTCTTGGCTTCCTG
actin-F	TTATATGCCAACACTGTCCTT
actin-R	AGAATTGATCCACCAATCCA

2 结果与分析

2.1 意大利蜜蜂工蜂细胞内吞相关基因和全长转录本分析

共鉴定到细胞内吞相关的 170 个基因和 991 条全长转录本,其中吞噬体相关基因和全长转录本分别为 57 个和 397 条,内吞作用相关基因和全长转录本分别为 113 个和 594 条。LOC413409 产生的转录本数量最多 (46 条),其次为 LOC100576742 (28 条)。LOC724950、LOC410614、LOC551369、LOC410632、LOC725700 和

LOC551888 产生的转录本数量最少,仅为 1 条。表 2 展示了部分细胞内吞相关基因和全长转录本的详细信息。

2.2 参考基因组已注释细胞内吞相关基因的结构优化

对西方蜜蜂参考基因组已注释的 89 个细胞内吞相关基因进行结构进化,其中正链和负链结构优化的基因分别为 53 和 36 个。另外,5'端延长的基因有 36 个,3'端延长的基因有 35 个,5'端和 3'端同时延长的基因有 18 个。部分基因结构优化的详细结果如表 3 所示。

表 2 意大利蜜蜂细胞内吞相关基因和全长转录本详细信息 (仅展示 10 个)
Table 2 Detailed information about endocytosis-related genes symbol and full-length transcripts in *Apis mellifera ligustica* (only ten presented)

基因符号 Gene symbol	转录本 ID Transcript ID	KEGG 数据库注释 KEGG database annotation	Nr 数据库注释 Nr database annotation
LOC100576742	ONT.5590.10	Phagosome	integrin beta-nu-like isoform X1
LOC413394	ONT.1542.2	Phagosome	vesicle-trafficking protein SEC22b-B-like
LOC724188	ONT.1755.2	Phagosome	protein transport protein Sec61 subunit beta
LOC409495	ONT.2145.2	Phagosome	lysosome-associated membrane glycoprotein 1-like isoform 2
LOC552528	ONT.3916.2	Phagosome	calnexin isoform X2
LOC409910	ONT.593.23	Endocytosis	ras-like GTP-binding protein Rho1 isoform X1
LOC107965281	ONT.5680.2	Endocytosis	GTP-binding ADP-ribosylation factor ARF6
LOC551100	rna24386	Endocytosis	endophilin-B1 isoform X4
LOC724991	ONT.658.6	Endocytosis	phosphatidylinositol 4-phosphate 5-kinase type-1 gamma isoform X14
LOC411857	ONT.6597.9	Endocytosis	charged multivesicular body protein 4b isoform X1

表 3 细胞内吞相关基因结构优化的详细信息 (仅展示 6 个)
Table 3 Detailed information about structure optimization of genes related to phagosome and endocytosis (only six presented)

基因符号 Gene symbol	优化后的起止位置 Start and end sites after optimization	正负链 Positive or negative strand	末端 End	原始位置 Original site	优化后的位置 Optimized site
LOC412886	NC_037652.1:7852048-7855293	+	3'	7854763	7855293
LOC551721	NC_037652.1:3786740-3789791	+	3'	3789477	7855293
LOC413200	NC_037652.1:1978802-2073463	—	5'	1980801	1978802
LOC410620	NC_037652.1:5231539-5237271	+	5'	5231544	5231539
LOC552410	NC_037652.1:7760742-7762128	+	5'	7760756	7760742
	NC_037652.1:7760742-7762128	+	3'	7761556	7762128
LOC552720	NC_037642.1:194671-197231	—	5'	194928	194671
	NC_037642.1:194671-197231	—	3'	197230	197231

2.3 细胞内吞相关基因的 AS 事件鉴定、分析及验证

共鉴定到细胞内吞相关的 28 个基因的 666 次 AS 事件, 涉及 ES、A5SS、A3SS、IR 和 MEE

5 种类型; 发生 AS 事件次数最多的类型为 A3SS (298 次), 其次为 ES (190 次)、A5SS (84 次) 和 IR (80 次), 最少的 AS 事件为 MEE (14 次)。部分细胞内吞相关基因的 AS 事件详细信息如表 4 所示。

表 4 细胞内吞相关基因的可变剪接事件详细信息 (仅展示 5 个)
Table 4 Detail information about alternative splicing events of endocytosis -related genes (only five presented)

基因符号 Gene symbol	转录本 ID Transcript ID	AS 事件类型 AS event type	区域 Region
LOC409481	ONT.472.2,ONT.472.1	Alternative 5' splice site	NC_037638.1:11885371-11888982W
LOC409910	ONT.593.37,ONT.593.38	Alternative 3' splice site	NC_037638.1:14864282-14868679C
LOC413683	ONT.4867.2,ONT.4867.1	Intron retention	NC_037647.1:5780037-5785078W
LOC413409	ONT.7405.10,ONT.7405.4	Exon skipping	NC_037653.1:6250113-6261700W
LOC727630	ONT.4448.1,ONT.4448.5	Mutually exclusive exon	NC_037646.1:2812469-2848141W

C 代表负义链, W 代表正义链。C represents nonsense strand, and W represents sense strand。

基于 Nanopore 测序数据的生物信息学预测结果显示, 细胞内吞相关基因整合素 β -nu 样异构体基因 (LOC100576742) 通过 ES 产生 ONT.5590.10 (约 371 bp) 和 ONT.5590.12 (目的片段大小约 842 bp) (图 1: C); Ras 样 GTP 结合蛋白基因 (LOC409910) 通过 ES 产生 ONT.593.6 (约 342 bp)、ONT.593.27 (约 386 bp) 和 ONT.593.33 (约 303 bp) (图 1: E), 通过 ES 产生 ONT.593.27 (约 268 bp) 和 ONT.593.23 (约 181 bp) (图 1: G), rab11 家族相互作用蛋白 4 基因 (LOC409274) 通过 IR 产生 rna14121 (约 439 bp)、ONT.3771.4 (目的片段大小为约 518 bp)

和 ONT.3771.7 (约 489 bp) (图 1: I)。上述结果说明 PCR 验证结果与生物信息学预测结果一致, 证明了意蜂细胞内吞相关基因的 AS 事件真实可靠。

2.4 细胞内吞相关基因的 APA 位点鉴定与分析

共鉴定到 120 个细胞内吞相关基因含有 1 046 个 APA 位点, 其中含有 3 个 APA 位点的基因数量最多, 为 14 个 (图 2: A); 含有 2 个 APA 位点的基因数量次之, 为 12 个 (图 2: A); 含有 1 个、4 个和 7 个 APA 位点的基因数量相同, 均为 10 个 (图 2: A)。此外, 在细胞内吞相关基

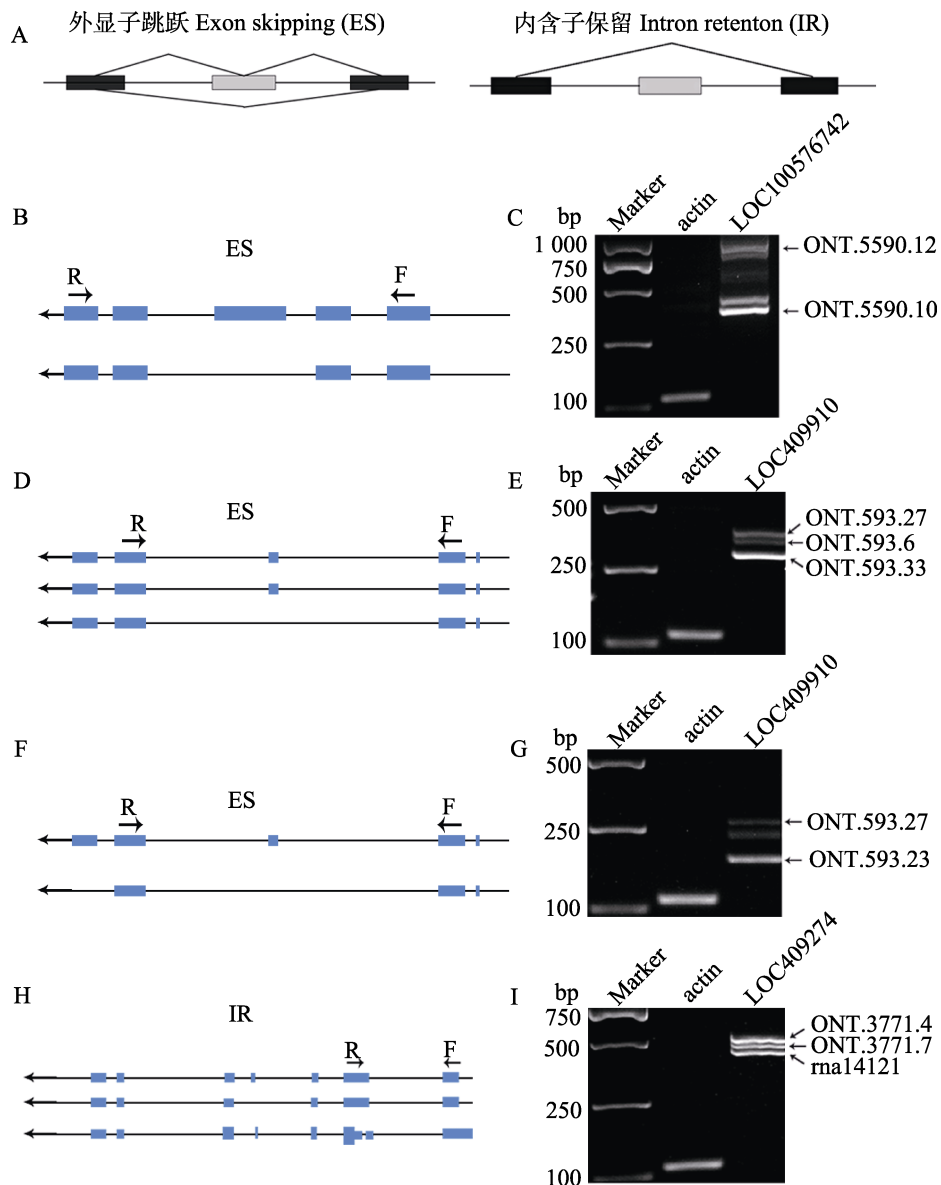


图 1 意大利蜜蜂细胞内吞相关的 3 个基因的 AS 事件验证

Fig. 1 Validation of AS events of three genes associated with endocytosis in *Apis mellifera ligustica*

A. ES, IR 示意图; B, D, F, H. 剪接体结构示意图; C, E, G, I. 剪接体扩增产物的琼脂糖凝胶电泳。

A. Schematic diagram of ES and IR; B, D, F, H. Diagrams of structures of isoforms; C, E, G, I. Agarose gel electrophoresis for amplified products from isoforms.

因的 APA 位点上游鉴定到多个 motif, 一致性序列为 NBMNHHHBMNSNNCBVSNNNNNNNNNNNNNNNNNNNNN (图 2: B)。

3 讨论

西方蜜蜂的第一个基因组版本 (Amel_4.0) 早在 2006 年就已完成测序并公布 (Honeybee Genome Sequencing Consortium, 2006)。为获得

更高质量的参考基因组, Wallberg 等 (2019) 运用 PacBio 测序技术、10×Chromium、BioNano 和 Hi-C 多种最新技术重新组装了接近染色体水平的西方蜜蜂参考基因组版本 (Amel_HAv3.1), 基因组大小约 250 Mb, 共有 12 374 个编码基因, 其中包含细胞内吞相关的 168 个基因。本研究利用 Nanopore 测序技术共鉴定到细胞内吞相关的 170 个基因和 991 条全长转录本, 其中吞噬体相

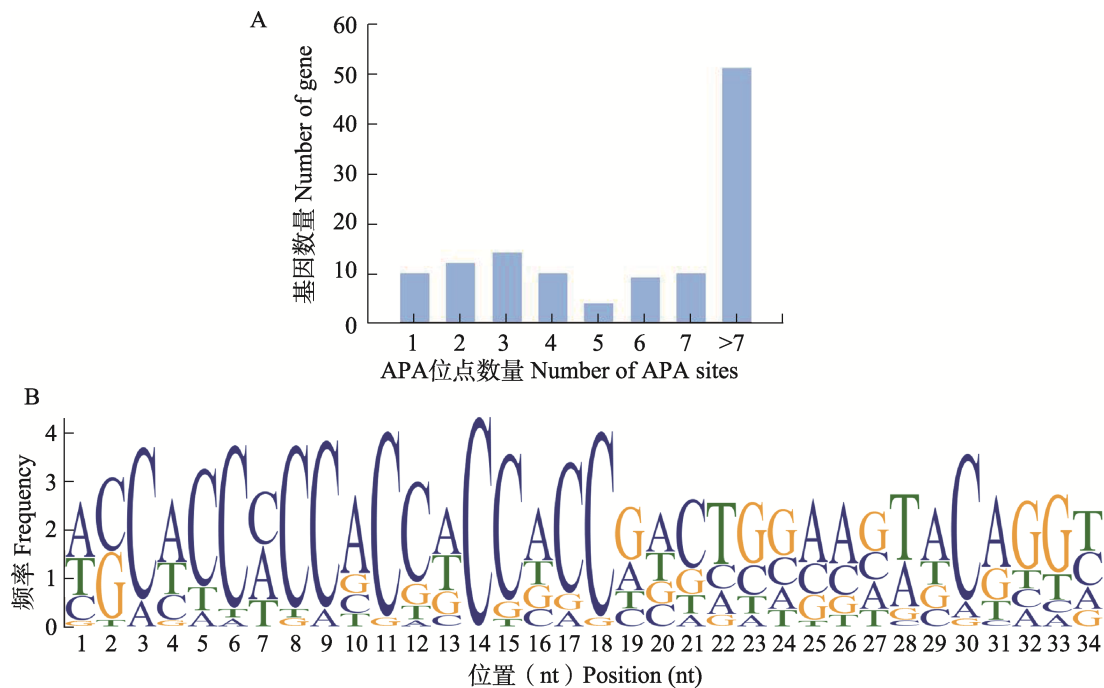


图 2 意大利蜜蜂细胞内吞相关基因的 APA 位点数量统计及 motif 鉴定
Fig. 2 Statistics and analysis of APA sites and motif identification of genes associated with endocytosis in *Apis mellifera ligustica*

A. 含有不同 APA 位点的基因数量统计; B. APA 位点上游 50 nt 位置鉴定到的 motif。
A. Number statistics of genes with APA sites; B. Motifs identified at 50 nt upstream of APA sites.

关基因和全长转录组数量分别为 57 和 397 条，内吞作用相关基因和全长转录组数量分别为 113 和 594 条，为相关基因和剪接体的克隆及功能研究提供了资源。

编码序列和 UTR 长度影响基因表达，miRNA 等因子通过结合 mRNA 的 3'UTR 和(或) 5'UTR 调节基因表达 (Barrett *et al.*, 2012)。利用 Nanopore 测序数据完善基因结构在蜜蜂球囊菌 (杜宇等, 2020; 隆琦等, 2021) 和东方蜜蜂微孢子虫 (陈华枝等, 2020) 等物种中已见诸报道。本研究共对 89 个细胞内吞相关基因结构进行了优化，延长了 36 个基因的 5'端，35 个基因的 3'端，18 个基因的 5'端和 3'端，说明即使对于染色体级别基因组，注释基因的序列仍有进一步完善空间，而 Nanopore 测序产生的长读段是一个很好的选项。考虑到 5'端和 3'端在基因表达调控方面的关键作用，上述基因结构优化对于进一步深入研究意蜂细胞内吞相关基因的表达调控具有重要意义。

真核生物基因普遍发生 AS，AS 参与调节 mRNA 稳定性和蛋白翻译，使基因在数量不增加的前提下大幅提高转录组和蛋白质的多样性与复杂性 (Venables *et al.*, 2012)。在拟南芥 *Arabidopsis thaliana* 中突变 2 个基因形成 2 个 A3SS，对其进行有效剪切，产生大量新的错误转录本，进而影响植物正常发育 (Cabezas-Fuster *et al.*, 2022)。本研究共鉴定到细胞内吞相关的 28 个基因的 666 次 AS 事件，类型包括 ES、A5SS、A3SS、IR 和 MEE。意蜂细胞内吞相关基因的 AS 事件中最丰富的类型也为 A3SS，占比达到 44.68%，在意蜂丝氨酸/苏氨酸蛋白激酶中最频繁的 AS 类型为 ES (范小雪等, 2023)，在不同基因及不同生物途径中甚至在不同物种之间发挥主要作用的 AS 类型不同，A3SS 在细胞内吞中的作用需要进一步探究。另外，本研究通过 PCR 证实整合素 β -nu 样异构体基因 (LOC100576742) 发生 1 次 ES 事件和 Ras 样 GTP 结合蛋白基因 (LOC409910) 发生的 2 次 ES 事件，以及 rab11

家族相互作用蛋白基因 (LOC409274) 发生 1 次 IR 事件, 证明了本研究中 Nanopore 测序数据的可靠性与鉴定到的 AS 事件的真实性。陈华枝等 (2020) 研究发现利用 Nanopore 测序技术鉴定果蝇 (Miller *et al.*, 2018)、拟南芥 (Michael *et al.*, 2018) 和东方蜜蜂微孢子虫等动植物和微生物的 AS 事件具有较高的准确性, 本研究的结果提供了又一例证。

本研究鉴定到 Rab 蛋白相关的 2 个基因: *ras* 相关蛋白编码基因 *Rab-5C* (LOC100577764) 和 *ras* 相关蛋白编码基因 *Rab-7a* (LOC724873), 并发现二者可通过 AS 分别产生 18 和 5 条转录本, 其中包括 12 和 4 条新转录本。RabGTP 酶类如 *rab5* 和 *rab7* 参与调节吞噬体融合与核周转运, 在吞噬体成熟进程中起到关键作用 (Gutierrez, 2013; 王芳等, 2019)。研究表明 *rab3* 基因参与调控棉铃虫 *Helicoverpa armigera* 吞噬作用 (李杰, 2014), *rab7* 和 *rab14* 通过增加果蝇对金黄色葡萄球菌 *Staphylococcus aureus* 的敏感性调节果蝇细胞的吞噬作用 (Garg and Wu, 2014)。鉴于目前蜜蜂 *Rab-5C* 和 *Rab-7a* 的研究仍然缺失, 下一步拟在本研究的基础上克隆上述 2 个基因及其剪接体的全长序列, 进而通过 RNAi 等技术探究其在意蜂生长发育及免疫防御中的功能。

APA 是多聚腺苷化因子对多聚腺苷化位点信号选择性识别之后, 切割并加上多聚腺苷尾巴, 产生长短不同的转录本, 导致 RNA 分子或翻译出的蛋白质存在差异 (黄芝秀, 2018)。本研究共鉴定到 120 个细胞内吞相关基因含有 1 个及以上的 APA 位点, 其中含有 3 个 APA 位点的基因最为丰富, 其次为含有 2 个 APA 位点的基因 (图 2: A)。这表明意蜂细胞内吞相关基因广泛发生 APA。本研究在意蜂细胞内吞相关基因的 APA 位点上游鉴定到多个 motif, 一致性序列为 NBMNHHHBMNSNNCBNSNNNNNNNNNN VNNNN (图 1: B), 与已报道的人和小鼠 *Mus musculus*、高粱 *Sorghum bicolor* (L.) Moench 和东方蜜蜂微孢子虫 *Nosema ceranae* 均不相同 (Tian *et al.*, 2005; Abdel-Ghany *et al.*, 2016; 陈华枝等, 2021)。这表明不同生物体基因的 APA

位点上游 motif 具有物种特异性。上述鉴定到的 motif 是否及如何调节意大利蜜蜂 APA 位点形成及细胞内吞值得进一步研究。

综上, 本研究首次鉴定到意蜂细胞内吞相关的 170 个基因和 991 条转录本, 优化了西方蜜蜂参考基因组已注释的细胞内吞相关的 89 个基因结构, 发掘出细胞内吞相关基因的 666 次 AS 事件和 1 046 个 APA 位点。

参考文献 (References)

- Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy AS, 2016. A survey of the sorghum transcriptome using single-molecule long reads. *Nature Communications*, 7: 11706.
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS, 2009. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Research*, 37: W202–W208.
- Barrett LW, Fletcher S, Wilton SD, 2012. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences*, 69(21): 3613–3634.
- Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, Lee J, Lam ET, Liachko I, Sullivan ST, Burton JN, Huson HJ, Nystrom JC, Kelley CM, Hutchison JL, Zhou Y, Sun J, Crisà A, Ponce de León FA, Schwartz JC, Hammond JA, Waldbieser GC, Schroeder SG, Liu GE, Dunham MJ, Shendure J, Sonstegard TS, Phillippy AM, Van Tassell CP, Smith TP, 2017. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics*, 49(4): 643–650.
- Cabezas-Fuster A, Micol-Ponce R, Fontcuberta-Cervera S, Ponce MR, 2022. Missplicing suppressor alleles of Arabidopsis PRE-MRNA PROCESSING FACTOR 8 increase splicing fidelity by reducing the use of novel splice sites. *Nucleic Acids Research*, 50(10): 5513–5527.
- Chen F, Zhu L, Zhang Y, Kumar D, Cao G, Hu X, Liang Z, Kuang S, Xue R, Gong C, 2018. Clathrin-mediated endocytosis is a candidate entry sorting mechanism for *Bombyx mori* cypovirus. *Scientific Reports*, 8(1): 7268.
- Chen HZ, Du Y, Fan XX, Zhu ZW, Jiang HB, Wang J, Fan YC, Xiong CL, Zheng YZ, Fu ZM, Xu GJ, Chen DF, Guo R, 2020. Construction and annotation of the full-length transcriptome of

- Nosema ceranae* based on the third-generation nanopore sequencing technology. *Acta Entomologica Sinica*, 63(12): 1461–1472. [陈华枝, 杜宇, 范小雪, 祝智威, 蒋海滨, 王杰, 范元婵, 熊翠玲, 郑燕珍, 付中民, 徐国钧, 陈大福, 郭睿, 2020. 基于第三代纳米孔测序技术的东方蜜蜂微孢子虫全长转录组构建及注释. *昆虫学报*, 63(12): 1461–1472.]
- 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, 2021. Analysis of alternative splicing and alternative polyadenylation of *Nosema ceranae* genes. *Mycosystema*, 40(1): 161–173. [陈华枝, 范小雪, 范元婵, 王杰, 祝智威, 蒋海滨, 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.]
- 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.]
- 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, 2021a. Analysis of alternative splicing and polyadenylation of *Ascosphaera apis* genes based on third-generation long-read sequencing dataset. *Acta Microbiologia 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, 2021a. 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*, 261(3): 667–682.]
- 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 Agricultural 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.]
- Fan XX, Zhang KY, Zhu LR, Wang ZX, Zhang KH, Niu QS, Xu XJ, Luo Q, Chen DF, Guo R, 2023. Identification and verification of genes and full-length transcripts of serine/threonine protein kinase in *Apis mellifera ligustica*. *Acta Entomologica Sinica*, 66(4): 478–485. [范小雪, 张凯遥, 朱乐冉, 王紫馨, 张奎昊, 牛庆生, 徐细建, 骆群, 陈大福, 郭睿, 2023. 意大利蜜蜂丝氨酸/苏氨酸蛋白激酶基因和全长转录本鉴定及验证. *昆虫学报*, 66(4): 478–485.]
- Garg A, Wu LP, 2014. *Drosophila* Rab14 mediates phagocytosis in the immune response to *Staphylococcus aureus*. *Cellular Microbiology*, 16(2): 296–310.
- Gutierrez MG, 2013. Functional role(s) of phagosomal Rab GTPases. *Small GTPases*, 4(3): 148–158.
- He XJ, Barron AB, Yang L, Chen H, He YZ, Zhang LZ, Huang Q, Wang ZL, Wu XB, Yan WY, Zeng ZJ, 2022. Extent and complexity of RNA processing in honey bee queen and worker caste development. *iScience*, 25(5): 104301.
- Hinze C, Boucrot E, 2018. Endocytosis in proliferating, quiescent and terminally differentiated cells. *Journal of Cell Science*, 131(23): jcs216804.
- Honeybee Genome Sequencing Consortium, 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature*, 443(7114): 931–949.
- Huang HR, Chen ZJ, Kunes S, Chang GD, Maniatis T, 2010. Endocytic pathway is required for *Drosophila* Toll innate immune signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 107(18): 8322–8327.
- Huang ZX, 2018. Dynamics of alternative polyadenylation and gene expression regulation by the absence of m6A on plant messenger RNA. Xiamen: Xiamen University. [黄芝秀, 2018. 植物信使 RNA 缺乏 m6A 甲基化时的选择性多聚腺苷酸化动态与基因表达调控研究. 硕士学位论文. 厦门: 厦门大学.]
- Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, Malla S, Marriott H, Nieto T, O'Grady J, Olsen HE, Pedersen BS, Rhie A, Richardson H, Quinlan AR, Snutch TP, Tee L, Paten B, Phillippy AM, Simpson JT, Loman NJ, Loose M, 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, 36(4): 338–345.
- Li J, 2014. Role of *Rab3* gene in cellular defense responses of the cotton bollworm, *Helicoverpa armigera*. Wuhan: Central China Normal University. [李杰, 2014. *Rab3* 基因在棉铃虫细胞免疫中的功能研究. 硕士学位论文. 武汉: 华中师范大学.]
- Liang Q, Chen DF, 2009. Honeybee Protection. Beijing: China Agriculture Press. 8–11. [梁勤, 陈大福, 2009. 蜜蜂保护学. 北

- 京: 中国农业出版社. 8–11.]
- Lin AE, Guttman JA, 2010. Hijacking the endocytic machinery by microbial pathogens. *Protoplasma*, 244(1/4): 75–90.
- Long Q, Yu KJ, Wu Y, Sun MH, Feng RR, Zhao X, Hu Y, Xu XJ, Fu ZM, Chen DF, Guo R, 2021. Identification and analysis of virulence factor-related full-length transcripts in *Ascosphaera apis* mycelium and spores. *Chinese Journal of Applied Entomology*, 58(5): 1073–1082. [隆琦, 余岢骏, 吴鹰, 孙明会, 冯睿蓉, 赵萧, 胡颖, 徐细建, 付中民, 陈大福, 郭睿, 2021. 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. *应用昆虫学报*, 58(5): 1073–1082.]
- Lu H, Giordano F, Ning Z, 2016. Oxford nanopore MinION sequencing and genome assembly. *Genomics, Proteomics & Bioinformatics*, 14(5): 265–279.
- Mbengue M, Bourdais G, Gervasi F, Beck M, Zhou J, Spallek T, Bartels S, Boller T, Ueda T, Kuhn H, Robatzek S, 2016. Clathrin-dependent endocytosis is required for immunity mediated by pattern recognition receptor kinases. *Proceedings of the National Academy of Sciences of the United States of America*, 113(39): 11034–11039.
- Michael TP, Jupe F, Bemm F, Motley ST, Sandoval JP, Lanz C, Loudet O, Weigel D, Ecker JR, 2015. High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nature Communications*, 9(1): 541.
- Miller DE, Staber C, Zeitlinger J, Hawley RS, 2018. Highly contiguous genome assemblies of 15 *Drosophila* species generated using nanopore sequencing. *G3 (Bethesda)*, 8(10): 3131–3141.
- Negri P, Maggi M, Ramirez L, Szawarski N, Feudis DL, Lamattina L, Eguaras M, 2016. Cellular immunity in *Apis mellifera*: Studying hemocytes brings light about bees skills to confront threats. *Apidologie*, 47(3): 379–388.
- Tian B, Hu J, Zhang H, Lutz CS, 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Research*, 33(1): 201–212.
- Tyson JR, O'Neil NJ, Jain M, Olsen HE, Hieter P, Snutch TP, 2018. MinION-based long-read sequencing and assembly extends the *Caenorhabditis elegans* reference genome. *Genome Research*, 28(2): 266–274.
- Venables JP, Tazi J, Juge F, 2012. Regulated functional alternative splicing in *Drosophila*. *Nucleic Acids Research*, 40(1): 1–10.
- Wallberg A, Bunikis I, Pettersson OV, Mosbech MB, Childers AK, Evans JD, Mikheyev AS, Robertson HM, Robinson GE, Webster MT, 2019. A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. *BMC Genomics*, 20(1): 275.
- Wang F, Yang JJ, Xu XX, 2019. Rab6 protein: Structure and function. *Medical Recapitulate*, 25(15): 2978–2982. [王芳, 杨俊俊, 徐兴祥, 2019. Rab6 蛋白的结构和功能. *医学综述*, 25(15): 2978–2982.]