

网络出版时间:2023-09-04 14:29 DOI:10.13207/j.cnki.jnwafu.2024.03.001
网络出版地址:https://link.cnki.net/urlid/61.1390.S.20230831.1424.010

中华蜜蜂细胞吞噬与包囊作用相关基因 全长转录本鉴定及分析

郭思佳¹,张凯遥¹,荆欣¹,高旭泽¹,冯佩林¹,邹培缘¹,张浩宇¹,
陈大福^{1,2},郭睿^{1,2},付中民^{1,2}

(1 福建农林大学 动物科学学院(蜂学学院),福建 福州 350002;2 福建省蜂疗研究所,福建 福州 350002)

[摘要] 【目的】系统鉴定和分析中华蜜蜂(*Apis cerana cerana*)吞噬与包囊作用相关基因和全长转录本,为深入开展相关基因和剪接体的功能研究奠定基础。【方法】基于前期已获得的高质量中华蜜蜂纳米孔长读段测序数据,通过 Blast 工具将全长转录本比对 Nr 数据库筛选出吞噬与包囊作用相关基因和全长转录本。利用 gffcompare 软件将全长转录本与东方蜜蜂(*Apis cerana*)参考基因组上注释的转录本进行比较,鉴定未注释的新基因和新转录本。利用 TAPIS pipeline 预测和分析吞噬与包囊作用相关基因的可变多聚腺苷酸化(alternative polyadenylation,APA)位点,并通过 TBtools 软件鉴定 APA 位点上游的基序(motif)。使用 Astalavista 软件鉴定可变剪接(alternative splicing,AS)事件,并通过 IGV 浏览器进行结构可视化。通过 RT-PCR 验证 AS 事件的真实性。【结果】共鉴定到中华蜜蜂吞噬与包囊作用相关的基因 66 个和全长转录本 395 条,发掘出东方蜜蜂参考基因组未注释的 2 个新基因和 303 条新转录本。对参考基因组已注释的 34 个基因进行了结构优化,分别延伸了 18 个基因的 5'端和 12 个基因的 3'端,同时延长了 4 个基因的 5'端和 3'端。共鉴定到含有 1 个及以上 APA 位点的吞噬与包囊作用相关基因 47 个,其中多于 5 个 APA 位点的基因最多,为 32 个。在 APA 位点上游鉴定到多个基序,一致性序列为:GRBGCNKSDAACAAAYTRB-GCBMRNGGBYAYTAYWCNVWNGG。共鉴定到吞噬与包囊作用相关基因的 AS 事件 296 次,其中包括 131 次可变 3'端剪接(alternative 3' splice site,A3SS)、85 次内含子保留(intron retention,IR)、70 次可变 5'端剪接(alternative 5' splice site,A5SS)和 10 次外显子跳跃(exon skipping,ES)。RT-PCR 结果显示,扩增的目的片段大小符合预期,证实了随机选择的 2 次 AS 事件的真实性。【结论】系统鉴定了中华蜜蜂吞噬与包囊作用相关基因和全长转录本以及 AS 事件和 APA 位点,优化了东方蜜蜂参考基因组注释的吞噬与包囊作用相关基因的结构。

[关键词] 中华蜜蜂;吞噬作用;包囊作用;全长转录本;纳米孔测序;可变剪接;可变多聚腺苷酸化

[中图分类号] S895.137

[文献标志码] A

[文章编号] 1671-9387(2024)03-0001-10

Identification and analysis of full-length transcripts of genes relative to phagocytosis and capsulation in *Apis cerana cerana*

GUO Sijia¹,ZHANG Kaiyao¹,JING Xin¹,GAO Xuze¹,FENG Peilin¹,ZOU Peiyuan¹,
ZHANG Haoyu¹,CHEN Dafu^{1,2},GUO Rui^{1,2},FU Zhongmin^{1,2}

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

2 Apitherapy Research Institute of Fujian Province,Fuzhou,Fujian 350002,China)

Abstract: 【Objective】Systematic identification and investigation of genes and full-length transcripts

[收稿日期] 2022-12-14

[基金项目] 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建设专项资金项目(CARS-44-KXJ7);福建农林大学硕士生导师团队项目;福建农林大学动物科学学院(蜂学学院)科研扶持项目;福建省省级大学生创新创业训练计划项目(202310389027,X202310389084)

[作者简介] 郭思佳(1998—),女,四川什邡人,在读硕士,主要从事蜜蜂分子生物学研究。E-mail:guosijia1998@163.com

[通信作者] 付中民(1972—),男,河北唐山人,副教授,主要从事蜜蜂科学研究。E-mail:zmf@fafu.edu.cn

郭睿(1987—),男,安徽六安人,副教授,主要从事昆虫-病原互作研究。E-mail:ruiguo@fafu.edu.cn

associated with phagocytosis and capsulation in *Apis cerana cerana* were conducted, aiming to provide basis for further functional study on related genes and isoforms. 【Method】 Based on previously gained high-quality Nanopore long read sequencing data from *Apis cerana cerana*, the mapping of full-length transcripts to Nr database was conducted by Blast tool to identify full-length transcripts relative to phagocytosis and capsulation. The full-length transcripts were compared with those annotated on reference genome using the gffcompare software to identify unannotated novel genes and full-length transcripts. Prediction and analysis of alternative polyadenylation (APA) sites were conducted with TAPIS pipeline, followed by identification of motifs upstream of APA sites by TBtools software. Astalavista software was employed to identify alternative splicing (AS) events followed by structural visualization by IGV browser. RT-PCR was performed to validate the authenticity of AS events. 【Result】 A total of 66 genes and 395 full-length transcripts relevant to phagocytosis and capsulation in *Apis cerana cerana* were discovered, including 2 new genes and 303 new transcripts unannotated on the reference genome of *Apis cerana*. The structure of 34 annotated genes of the reference genome was optimized, among which 5' ends of 18 genes and 3' ends of 12 genes were extended, and 5' ends and 3' ends of 4 genes were prolonged. Additionally, 47 genes related to phagocytosis and capsulation were identified to contain one or more APA sites. The number of genes with more than 5 APA sites was 32. Multiple motifs were identified in upstream of APA sites, and the consistent sequence was GRBGCNKSDAACAAAYTRBGCBMRNGGBYAYTAYWCNVWNGG. In total, 296 AS events were identified, including 131 alternatives 3' splice sites (A3SS), 85 intron retention (IR), 70 alternative 5' splice sites (A5SS) and 10 exon skipping (ES). RT-PCR showed that the sizes of amplified fragments were consistent with expected sizes, confirming that the authenticity of the 2 randomly selected AS events. 【Conclusion】 Genes and full-length transcripts relative to phagocytosis and capsulation in *Apis cerana cerana* and AS events as well as APA sites were systematically identified, and structures of phagocytosis and capsulation-associated genes were optimized.

Key words: *Apis cerana cerana*; phagocytosis; capsulation; full-length transcripts; Nanopore sequencing; alternative splicing; alternative polyadenylation

蜜蜂是自然界最重要的授粉昆虫,在生态平衡和粮食安全方面发挥举足轻重的作用。中华蜜蜂(*Apis cerana cerana*),简称中蜂,是东方蜜蜂(*Apis cerana*)的指名亚种,也是我国养蜂生产中使用的主要蜂种之一,经过长期的适应性进化已高度适应我国地理环境,具有重要的生态和经济价值^[1]。

近年来,以牛津纳米孔(Nanopore)长读段测序技术为代表的第三代测序技术不断革新、突飞猛进,在组装染色体级别基因组和揭示转录组复杂性等方面应用广泛。由于具有超长读长的显著优势,牛津纳米孔测序技术已成功应用于可变剪接体(isoform)、可变剪切(alternative splicing, AS)和可变多聚腺苷酸化(alternative polyadenylation, APA)位点等的精确鉴定和分析^[2-3]。目前,基于牛津纳米孔测序技术的全长转录组研究已见诸于橄榄果蝇(*Bactrocera oleae*)^[4]、亚洲飞蝗(*Locusta migratoria*)^[5]和椰蛀犀金龟(*Oryctes rhinoceros*)^[6]等昆虫中。东方蜜蜂的全长转录组研究报道迄今仅有 2

例, Liu 等^[7]对长白山东方蜜蜂越冬期抗寒性的全长转录本进行了分析,确定了 5 941 个完整的 ORF 序列;蔡宗兵等^[8]在中华蜜蜂中肠中鉴定到 265 个与细胞色素 P450 相关的全长转录本。此前,本课题组利用牛津纳米孔测序技术对蜜蜂的 2 种真菌病原蜜蜂球囊菌(*Ascosphaera apis*)和东方蜜蜂微孢子虫(*Nosema ceranae*)进行了较为系统的全长转录组研究^[9-14],但中蜂的全长转录组研究总体上仍较为滞后。

作为无脊椎动物,蜜蜂等昆虫缺乏获得性免疫,但拥有包括细胞免疫和体液免疫在内的天然免疫系统。当侵入昆虫血淋巴的病原体被宿主细胞识别后,宿主的细胞免疫随即被激活,进而合成与分泌抗菌肽,抵御病原侵染^[15]。吞噬和包裹作用是昆虫的 2 种重要细胞免疫反应,吞噬能直接杀死真菌孢子和外源小分子,而包裹能抵御如寄生虫等无法被吞噬的入侵者^[16]。Wang 等^[17]研究发现,棉铃虫(*Helicoverpa armigera*) 20E-HaEcR-HaUSP 复合物刺激血淋巴中 *HaCTL1* 基因的表达,进而提高

HaCTL1 蛋白的合成与分泌,并促进细胞吞噬和包裹作用,抵抗中华卵索线虫(*Ovomeris sinensis*)的侵袭。在果蝇(*Drosophila*)中, *TM9SF4* 基因突变可引起吞噬与包裹作用缺失,导致果蝇对革兰氏阴性菌更加敏感^[18]。Park 等^[19]利用二代测序技术组装和注释了东方蜜蜂的参考基因组,但由于二代测序产生的读段较短,故该版本参考基因组仅组装到重叠群(contig)水平,东方蜜蜂参考基因组仍需要进一步完善。目前,人们对中蜂吞噬与包裹作用的认识非常有限,对其相关基因和全长转录组缺乏深入的研究。

前期研究中,本课题组已利用牛津纳米孔测序技术对中华蜜蜂工蜂幼虫肠道组织进行了测序,获得了高质量的长读段数据^[8]。在此基础上,本研究对中华蜜蜂吞噬与包裹作用相关基因和全长转录本进行了系统鉴定,对东方蜜蜂参考基因组已注释的吞噬与包裹作用相关基因进行结构优化,进而发掘和分析吞噬与包裹作用相关基因的 APA 位点和 AS 事件,并对 AS 事件进行分子验证,以期丰富中华蜜蜂吞噬与包裹作用相关基因和全长转录本的相关信息,为深入开展相关基因和剪接体的功能研究奠定基础。

1 材料与方法

1.1 细胞吞噬与包裹作用相关基因与全长转录本的筛选与分析

本课题组前期已完成中蜂幼虫 4,5 和 6 日龄肠道样品(Ac4 组、Ac5 组和 Ac6 组)制备、RNA 提取、cDNA 建库及牛津纳米孔测序。Ac4、Ac5 和 Ac6 组测序结果分别获得了 7 338 627,7 003 419 和 7 434 233 条原始读段,其 N50 分别为 1 276,1 306 和 1 404 bp,平均读长分别为 1 126,1 126 和 1 166 bp,最大长度分别为 16 628,12 808 和 14 359 bp,质量值均达到 Q12^[8]。高质量的长读段测序数据可用于本研究中中蜂细胞吞噬与包裹作用相关基因和全长转录本的鉴定及分析。使用 Blast 工具(参数为默认设置)将鉴定到的所有全长转录本序列比对到 GO(<http://geneontology.org/>)、KEGG(<https://www.kegg.jp/>)和 Nr(<https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/>)数据库,再根据注释信息筛选细胞吞噬与包裹作用相关基因和全长转录本。

1.2 东方蜜蜂参考基因组中已注释细胞吞噬与包裹作用相关基因的结构优化

按照本课题组已建立的技术流程^[12,20],利用

gffcompare 软件(参数为默认设置)将本研究中鉴定到的细胞吞噬与包裹作用相关全长转录本与中蜂参考基因组(ACSNU-2.0)上已注释的转录本进行比较,以发现未注释的新基因和转录本,并对已注释基因进行结构优化。如果在原有基因边界之外的区域有比对读段(mapped reads)支持,则将基因的非翻译区向上、下游延伸,以修正基因的边界。

1.3 细胞吞噬与包裹作用相关基因的 APA 位点鉴定与分析

参照杜宇等^[11]报道的方法,利用 TAPIS pipeline^[21]鉴定细胞吞噬与包裹作用相关基因的 APA 位点,参数设置为:-norc、-meme-minw 6、-meme-maxw 6、-spamo-skip、-fimo-skip。通过 TBtools 软件^[22]对细胞吞噬与包裹作用相关基因 APA 位点上游 50 bp 的序列特征进行分析,以鉴定基序(motif)。

1.4 细胞吞噬与包裹作用相关基因的可变剪接分析

参照陈华枝等^[13]和杜宇等^[11]报道的方法,使用 Astalavista 软件^[23]鉴定中蜂细胞吞噬与包裹作用相关基因的 AS 事件类型,其中包括互斥外显子(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)。参数为默认设置。根据预测结果统计以上可变剪接类型数量。

1.5 细胞吞噬与包裹作用相关基因的 AS 事件验证

中蜂幼虫取自福建农林大学动物科学学院(蜂学学院)蜜蜂保护课题组的实验蜂群。参照本课题组已建立的技术流程^[24-26],从蜂群中提取巢脾,在实验室内将 2 日龄幼虫移至干净的 48 孔培养板,放入恒温恒湿箱,在温度为 35 ℃、相对湿度(RH)为 90%的条件下饲养,分别剖取 4,5 和 6 日龄幼虫肠道组织,液氮速冻后迅速转移到-80 ℃超低温冰箱保存,备用。

为验证细胞吞噬与包裹作用相关基因的 AS 事件准确性,以 β -actin 为内参基因,随机选择 2 种 AS 类型(A3SS、IR)的 1 个基因进行 RT-PCR 验证。根据上述基因的核酸序列设计跨剪接位点的特异性引物,交由生工生物(上海)公司合成(表 1)。以肌动蛋白基因 β -actin (LOC107999330)作为内参基因,将上述制备的 4,5 和 6 日龄幼虫肠道的 RNA 等物

质的量混合后进行反转录,获得 cDNA。以得到的 cDNA 作为模板进行 PCR 扩增,反应体系和程序参照

蔡宗兵等^[8]的报道进行。PCR 产物经 3% 琼脂糖凝胶电泳后使用核酸凝胶成像仪进行检测和拍照。

表 1 本研究中使用的引物序列

Table 1 Sequences of primers used in this study

AS 类型 Type of AS	基因或基因 ID Gene or gene ID	预测的转录本号 Predicted transcript number	序列 Sequence
IR	LOC107999286	ONT. 6127. 13(444 bp),	F:5'-CTCACGGA ACTACCACGG-3'
		ONT. 6127. 3(341 bp)	R:5'-GGGATATTCGCCACAACA-3'
A3SS	LOC107999764	ONT. 6364. 6(362 bp),	F:5'-AGAAAGAGCATTAGGAGA-3'
		ONT. 6364. 2(425 bp)	R:5'-GGAGTACGTTGAATAGGA-3'
—	<i>β-actin</i>	—	F:5'-TTATATGCCAACACTGTCTTT-3'
			R:5'-AGAATTGATCCACCAATCCA-3'

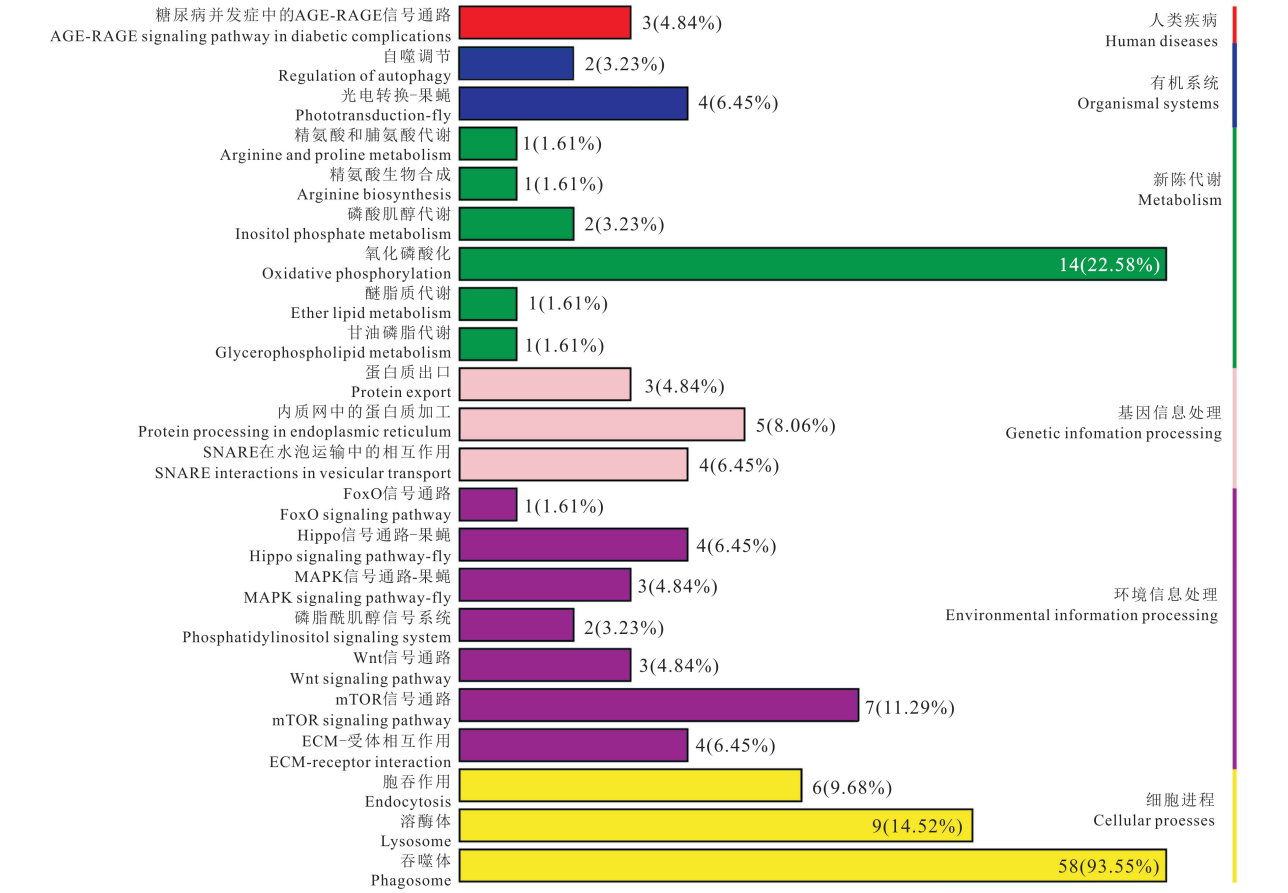
注:括号中的数据为片段长度。
Note:Data in the brackets is the length of transcript.

2 结果与分析

2.1 中蜂细胞吞噬与包裹作用相关基因和全长转录本的鉴定与分析

试验共鉴定到吞噬相关的基因 65 个和全长转录本 391 条,包裹作用相关的基因 1 个和全长转录

本 4 条,其中包括东方蜜蜂参考基因组上未注释的 2 个新基因和 303 条新转录本。对鉴定到的基因进行 KEGG 和 GO 注释,结果有 62 个基因注释到 22 条 KEGG 通路(图 1);有 65 个基因注释到 53 个 GO 条目,其中 *P* 值排序前 20 的条目见图 2。



图中数据为注释到的基因数及其占比。图 2 同
Digits indicate numbers of genes annotated and its ratio. Fig. 2 is the same

图 1 中蜂细胞吞噬与包裹作用相关基因的 KEGG 注释

Fig. 1 KEGG annotations of genes related to phagocytosis and capsulation in *Apis cerana cerana*

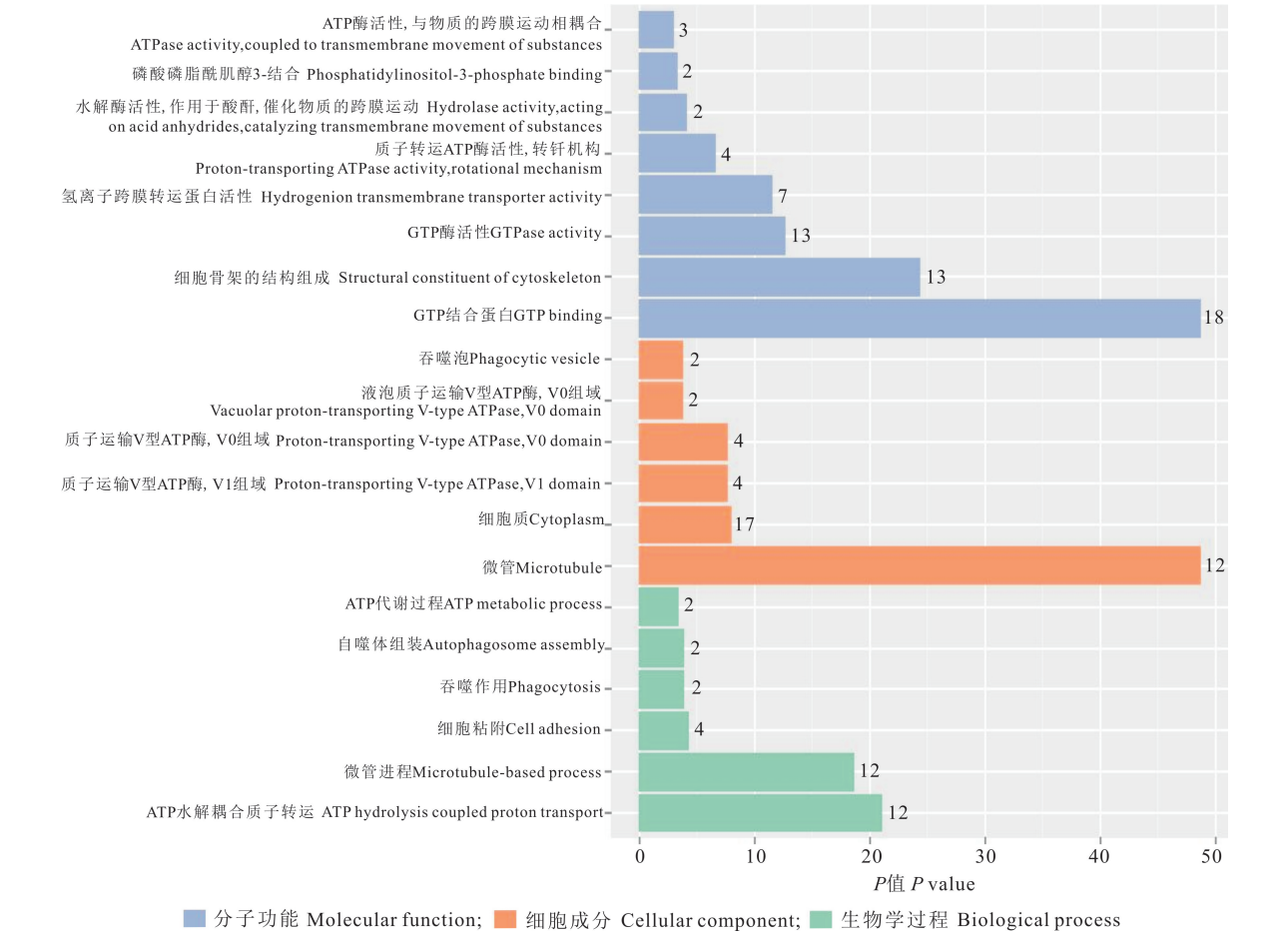


图 2 中蜂细胞吞噬与包裹作用相关基因的 GO 注释

Fig. 2 GO annotations of genes related to phagocytosis and capsule in *Apis cerana cerana*

2.2 东方蜜蜂参考基因组已注释细胞吞噬与包裹作用相关基因的结构优化

试验共对东方蜜蜂参考基因组上已注释的 34 个细胞吞噬与包裹作用相关基因的结构进行了优化,其中正链结构优化的基因为 17 个,负链结构优化的基因为 17 个;5'端延长的基因有 18 个,延长长度介于 4~5 512 bp;3'端延长的基因有 12 个,延长长度介于 1~3 218 bp;3'和 5'端均延长的基因有 4 个, 分 别 为 LOC107999172, LOC107999330, LOC108000752 和 LOC108003556(表 2)。

表 2 中蜂细胞吞噬与包裹作用相关基因结构优化的详细信息

Table 2 Detailed information of structural optimization of genes related to phagocytosis and capsulation in *Apis cerana cerana*

基因 ID Gene ID	参考序列 Reference sequence	优化前的起止位置 Start and end sites before optimization	优化后的起止位置 Start and end sites after optimization	正负链 Positive or negative strand	末端 End
LOC107998143	NW_016019219.1	1 777 187—1 778 417	1 771 675—1 778 417	+	5'
LOC108003298	NW_016019863.1	97 472—99 778	91 999—99 778	—	5'
LOC107995660	NW_016017456.1	1 776 644—1 806 926	1 774 455—1 806 926	+	5'
LOC107999172	NW_016019319.1	419 590—429 628	417 412—429 628	+	5'
LOC107997445	NW_016019153.1	1 089 494—1 103 782	1 087 610—1 103 782	—	5'
LOC107999159	NW_016019319.1	17 703—24 780	16 575—24 780	—	5'
LOC107994640	NW_016018544.1	5 265—7 101	4 277—7 101	—	5'
LOC107999790	NW_016019364.1	80 049—83 237	79 245—83 237	—	5'
LOC107994136	NW_016018344.1	1 816 913—1 822 666	1 816 182—1 822 666	—	5'
LOC107999764	NW_016019364.1	946 296—958 611	945 624—958 611	—	5'
LOC108000406	NW_016019441.1	709 872—713 105	709 245—713 105	—	5'
LOC107997632	NW_016019175.1	1 062 418—1 072 796	1 061 901—1 072 796	—	5'
LOC108000577	NW_016019442.1	960 650—965 211	960 161—965 211	—	5'

表 2(续) Table 2 (continued)

基因 ID Gene ID	参考序列 Reference sequence	优化前的起止位置 Start and end sites before optimization	优化后的起止位置 Start and end sites after optimization	正负链 Positive or negative strand	末端 End
LOC107999330	NW_016019330.1	1 372 267—1 377 034	1 371 781—1 377 034	—	5′
LOC108000587	NW_016019442.1	569 024—569 997	568 540—569 997	—	5′
LOC108000752	NW_016017512.1	181 972—183 084	181 594—183 084	—	5′
LOC107998213	NW_016019219.1	2 093 759—2 100 474	2 093 552—2 100 474	—	5′
LOC107998677	NW_016019275.1	760 522—765 147	760 336—765 147	—	5′
LOC107999232	NW_016019330.1	492 630—495 090	492 481—495 090	—	5′
LOC108001305	NW_016019552.1	187 553—189 571	187 438—189 571	—	5′
LOC107995080	NW_016017456.1	516 106—518 939	516 101—518 939	+	5′
LOC108003556	NW_016017567.1	2 875 991—2 878 618	2 875 987—2 878 618	+	5′
LOC107999330	NW_016019330.1	1 371 781—1 377 033	1 371 781—1 377 034	—	3′
LOC107992565	NW_016017856.1	183 783—185 914	183 783—185 924	+	3′
LOC108001195	NW_016017512.1	514 991—517 027	514 991—517 058	+	3′
LOC108000752	NW_016017512.1	181 594—182 906	181 594—183 084	—	3′
LOC107998372	NW_016019231.1	1 386 200—1 388 380	1 386 200—1 388 790	+	3′
LOC107995099	NW_016018567.1	136 406—140 802	136 406—141 304	+	3′
LOC107993924	NW_016018256.1	253 448—258 337	253 448—258 848	+	3′
LOC107997332	NW_016017457.1	590 209—593 951	590 209—594 488	+	3′
LOC108003264	NW_016017556.1	192 600—196 182	192 600—196 785	+	3′
LOC107992702	NW_016017899.1	211—1 820	211—2 563	+	3′
LOC107999172	NW_016019319.1	417 412—428 815	417 412—429 628	+	3′
LOC108000484	NW_016019441.1	4 376 457—4 380 212	4 376 457—4 381 188	+	3′
LOC108003556	NW_016017567.1	2 875 987—2 877 296	2 875 987—2 878 618	+	3′
LOC108002524	NW_016017455.1	845 010—847 931	845 010—849 556	+	3′
LOC122719477	NW_016017767.1	310 646—314 815	310 646—318 033	+	3′
LOC107995086	NW_016018567.1	848 331—849 310	848 331—849 899	+	3′

2.3 中蜂细胞吞噬与包囊作用相关基因的 APA 位点鉴定与分析

试验共鉴定到含有 1 个及以上 APA 位点的细胞吞噬与包囊作用相关基因 47 个(图 3),其中多于 5 个 APA 位点的基因最多,为 32 个;含有 1 个 APA 位点的基因次之,为 6 个;含有 2 个和 3 个 APA 位点的基因均为 4 个;含有 4 个 APA 位点的基因为 1 个。此外,在细胞吞噬与包囊作用相关基因的 APA 位点上游鉴定到多个基序,一致性序列为:GRBGCNKSDAACAAAYTRBGCBMRNGGBY-AYTAYWCNVWNGG(图 4)。

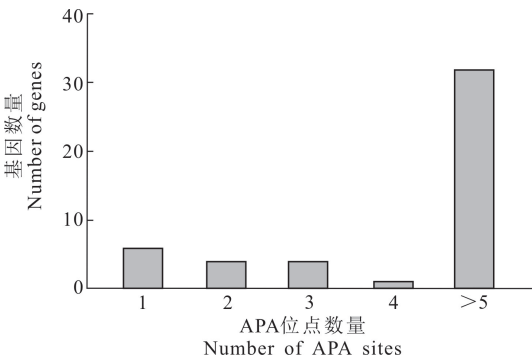


图 3 不同数量 APA 位点中蜂细胞吞噬与包囊作用相关基因的分布

Fig. 3 Distribution of different numbers of APA sites of genes related to phagocytosis and capsule in *Apis cerana cerana*



图 4 中蜂细胞吞噬与包囊作用相关基因的一致性基序

Fig. 4 Motifs of genes related to phagocytosis and capsulation in *Apis cerana cerana*

2.4 中蜂细胞吞噬与包裹作用相关基因的 AS 事件鉴定与分析

试验共鉴定到细胞吞噬与包裹作用相关基因的 296 次 AS 事件,其中数量最多的类型是可变 3'端剪

接(A3SS,131 次),其次为内含子保留(IR,85 次)和可变 5'端剪接(A5SS,70 次),数量最少的类型为外显子跳跃(ES,10 次)(图 5)。部分细胞吞噬与包裹作用相关基因的 AS 事件详细信息见表 3。

表 3 细胞吞噬与包裹作用相关基因的可变剪接事件(AS)详细信息(仅展示 6 个)

Table 3 Detail information of alternative splicing events occurred in phagocytosis and capsulation relevant genes (six displayed only)

基因 ID Gene ID	转录本 ID Transcript ID	AS 事件类型 AS event type	参考序列 Reference sequence	链属 Strand type	区域 Region
LOC107999286	ONT. 6127. 13,ONT. 6127. 3	IR	NW_016019330. 1	C	3 148 146—3 148 035
LOC107999764	ONT. 6364. 6,ONT. 6364. 2	A3SS	NW_016019364. 1	C	949 340—949 306
LOC107999286	ONT. 6127. 7,ONT. 6127. 2	ES	NW_016019330. 1	C	3 172 740—3 172 667
LOC107995660	ONT. 309. 1,ONT. 309. 2	A5SS	NW_016017456. 1	W	1 780 668—1 780 704
LOC107995099	ONT. 3315. 1,ONT. 3315. 3	A3SS	NW_016018567. 1	W	137 912—137 926
LOC107992565	ONT. 1750. 1,ONT. 1750. 2	IR	NW_016017856. 1	W	183 906—184 992

注:C 和 W 分别代表负义链和正义链。
Note:C and W represent nonsense strand and sense strand, respectively.

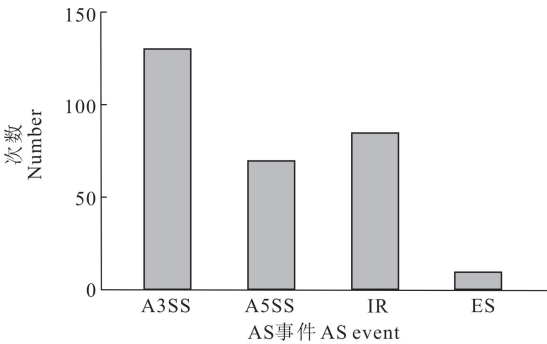
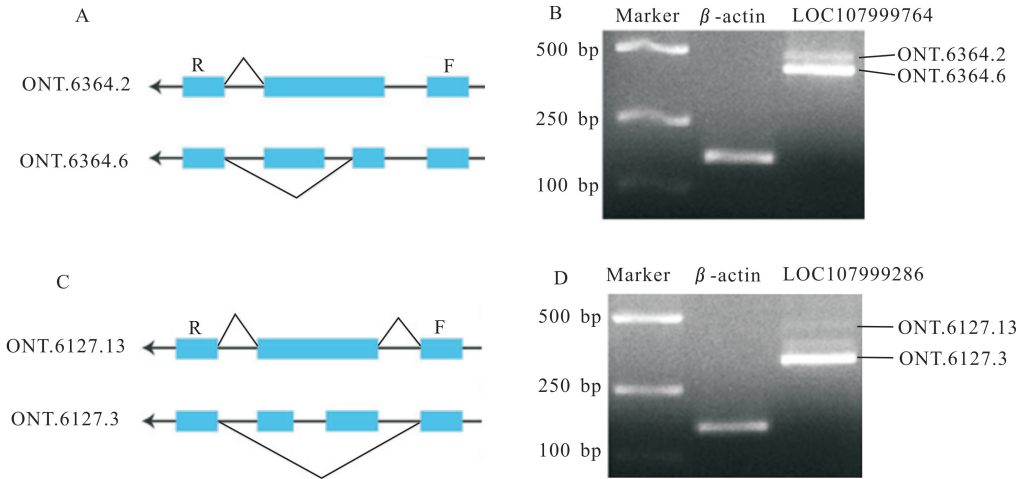


图 5 中蜂细胞吞噬与包裹作用相关基因的 AS 事件类型
Fig. 5 Types of AS events in genes related to phagocytosis and capsulation in *Apis cerana cerana*

2.5 中蜂细胞吞噬与包裹作用相关基因的 AS 事件验证

随机选择 2 种 AS 事件类型进行 RT-PCR 验证,结果(图 6)显示,LOC107999764 基因可检测到 2 条转录本,大小分别约为 425 和 362 bp,与基于测序数据预测出的 ONT. 6364. 2 和 ONT6364. 6 的大小(表 1)一致;LOC107999286 基因可检测到 2 条转录本,大小分别约为 444 和 341 bp,与基于测序数据预测出的 ONT. 6127. 3 和 ONT. 6127. 13 的大小(表 1)一致。上述结果表明,本研究鉴定到的 AS 事件真实可靠。



A,C. A3SS 和 IR 的可变剪接事件类型示意图,方框表示外显子,直线表示内含子,箭头表示转录方向;

B,D. LOC107999764 和 LOC107999286 扩增产物的琼脂糖凝胶电泳

A and C show types of alternative splicing events occurred in A3SS and IR,rectangles represent exons,and black lines represent introns, arrow indicate direction of transcription;B and D show agarose gel electrophoresis for amplified products from LOC107999764 and LOC107999286

图 6 中蜂 2 个基因可变剪接事件的 PCR 验证

Fig. 6 PCR validation of alternative splicing events occurred in two genes in *Apis cerana cerana*

3 讨 论

此前, Park 等^[19]利用二代测序技术组装和注释了东方蜜蜂的参考基因组(ACSNU-2.0), 该版本的基因组大小约为 238 Mb, 共包含 10 651 个编码基因, 其中注释到吞噬作用的基因和转录本数目分别为 95 个和 182 条, 注释到包裹作用的基因和转录本数目分别为 1 个和 1 条。基于已获得的高质量牛津纳米孔测序数据, 本研究共鉴定到中蜂吞噬作用相关的基因 65 个和全长转录本 391 条, 包裹作用相关的基因 1 个和全长转录本 4 条, 其中包括参考基因组上未注释的新基因 2 个和新转录本 303 条。这为持续深入开展中蜂吞噬与包裹作用相关基因和 isoform 的研究提供了宝贵的材料。

UTR 主要通过顺式调节元件和 RNA 结合蛋白(或小 RNA 反式作用因子)之间的动态相互作用来调节特定的 mRNA 的代谢和翻译^[27]。3' UTR 参与 miRNA 或以蛋白质为基础的 mRNA 的降解机制, 而 5' UTR 影响 mRNA 的翻译效率^[28]。本研究共对东方蜜蜂参考基因组上已注释的 34 个吞噬与包裹作用相关基因结构进行优化, 延长了 18 个基因的 5' UTR(延长长度介于 4~5 512 bp)和 12 个基因的 3' UTR(延长长度介于 1~3 218 bp), 同时延长了 4 个基因的 5' UTR 和 3' UTR。这为进一步开展上述中蜂吞噬与包裹作用相关基因表达调控的深入研究提供了重要基础。

APA 通过产生编码序列或 3' UTR 不同的异构体, 从而调节目标 RNA 的功能、稳定性、定位和翻译效率, 促进转录组的复杂性^[29]。哺乳动物基因组中有 50%~80% 的基因发生 AS 和 APA, 进而产生数量众多的剪接异构体^[30]; 在人(*Homo sapiens*)中, 至少有 70% 的基因含有 APA 位点^[31-32]; 在小鼠(*Mus musculus*)中, 32% 的表达基因经历 APA^[33]; 在水稻(*Oryza sativa*)中, 约 48% 的基因含有 APA 位点^[34]; 在小麦(*Triticum aestivum*)中, 有 31% 的基因含有 APA 位点^[35]; 在拟南芥(*Arabidopsis thaliana*)中, 超过 60% 的基因含有 APA 位点^[36]。本研究鉴定到中蜂吞噬与包裹作用相关的 66 个基因, 其中有 47(约 71.2%) 个基因含有 1 个及以上的 APA 位点, 表明中蜂吞噬与包裹作用相关的基因多数发生 APA。其中, 多于 5 个 APA 位点的基因最多, 这与他人对热带爪蟾(*Xenopus tropicalis*)^[37]和蜜蜂球囊菌^[11]等物种的研究报道一致。但在高粱^[38]和草地贪夜蛾(*Spodoptera frugiperda*)^[39]等

物种中, 含 2 个 APA 位点的基因最为丰富。上述结果表明, 不同物种中, APA 位点数的分布规律存在差异。此外, 在吞噬与包裹作用相关基因的 APA 位点上游鉴定到多个基序, 一致性序列为: GRBGC-NKSDAACAAAYTRBGCMBMRNGGBYAYTAYW-CNVWNGG。人类基因的 APA 位点上游存在 1 个经典 motif (AAUAAA), 该基序同样存在于果蝇基因 APA 位点上游^[40]。Ma 等^[35]在小麦基因的 APA 位点上游鉴定到多个基序, 其中出现频率最高的是 AAUAAA、AUUAUUAU 和 UAUAUA。玉米(*Zea mays*)基因的 APA 位点上游出现频率前 3 位的基序为: AAUAAA、AUUAUUAU 和 UAUAUA^[41]。本课题组在东方蜜蜂微孢子虫基因的 APA 位点上游发现 3 个基序: AAUAAA、UGAUGC 和 GCGACG^[13], 在蜜蜂球囊菌基因的 APA 位点上游鉴定到 4 个基序: AUGAUAA、CAUGAAC、GUUCAAU 和 UCAUCAU^[42]。以上结果表明, APA 上游基序具有一定的种属特异性。Wang 等^[30]研究发现, 哺乳动物基因进化过程中聚腺苷酸化位点(polyadenylation sites, PAS)选择对 APA 施加进化压力的基因特征包括基因年龄、基因表达模式和基因功能。本研究鉴定到的基序与其他物种中 APA 位点上游的基序均不相同, 推测可能是由于高分化细胞一般倾向于使用远端 PAS, 导致具有较长 3' UTR 的转录本表达, 而生长较快的细胞则倾向于使用近端 PAS, 产生较短的 3' UTR, 从而导致产生的转录本具有潜在的高表达水平^[43]。AS 在基因的功能多样性中扮演着重要的角色, 由 AS 产生的蛋白质异构体在酶活性、定位或稳定性等方面存在差异, 从而影响细胞的活动^[44]。本研究共鉴定到吞噬与包裹作用相关基因的 296 次 AS 事件, 说明中蜂吞噬与包裹作用相关基因广泛发生 AS。其中, 最丰富的 AS 事件类型是 A3SS(131 次), 其次是 IR(85 次), 数量最少的为 ES(10 次)。在硝酸盐胁迫条件下, 大豆(*Glycine max* L.)根系中 RI 和 A3SS 是最常见的 AS 类型, 其中 A3SS 产生的剪接异构体的相对表达量在不同硝酸盐条件差异显著^[45]。Levesque 等^[46]报告了第一个关于癌症中 A3SS 使用的全基因组研究, 强调了乳腺癌中数百个不同剪接的事件, 表明 A3SS 在乳腺癌整体剪接网络中具有重要作用。IR 和 A3SS 主要涉及调节前体 mRNA 产生不同的成熟剪接体, 调节细胞中基因的表达, 然后影响编码蛋白的功能^[47]。本研究通过 RT-PCR 对随机选择 2 种 AS 事件类型进行验

证,结果显示 LOC107999764 和 LOC107999286 均能扩增出 2 条转录本,大小与基于测序数据的预测结果一致,表明中蜂吞噬与包裹作用相关基因的 AS 事件真实可靠。

[参考文献]

[1] 曾志将. 养蜂学 [M]. 北京:中国农业出版社,2017.
Zeng Z J. Apiculture [M]. Beijing: China Agriculture Press, 2017.

[2] Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing [J]. Nature Biotechnology, 2016, 34(5): 518-524.

[3] Wang Y, Zhao Y, Bollas A, et al. Nanopore sequencing technology, bioinformatics and applications [J]. Nature Biotechnology, 2021, 39(11): 1348-1365.

[4] Bayega A, Oikonomopoulos S, Gregoriou M E, et al. Nanopore long-read RNA-seq and absolute quantification delineate transcription dynamics in early embryo development of an insect pest [J]. Scientific Reports, 2021, 11(1): 7878.

[5] Jiang F, Zhang J, Liu Q, et al. Long-read direct RNA sequencing by 5'-Cap capturing reveals the impact of Piwi on the widespread exonization of transposable elements in locusts [J]. RNA Biology, 2019, 16(7): 950-959.

[6] Filipovic I, Hereward J P, Rašić G, et al. The complete mitochondrial genome sequence of *Oryctes rhinoceros* (Coleoptera: Scarabaeidae) based on long-read nanopore sequencing [J]. Peer J, 2021, 9: e10552.

[7] Liu N N, Ren Z Y, Ren Q D, et al. Full length transcriptomes analysis of cold-resistance of *Apis cerana* in Changbai Mountain during overwintering period [J]. Gene, 2022, 830: 146503.

[8] 蔡宗兵, 王紫馨, 吴 鹰, 等. 中华蜜蜂细胞色素 P450 相关基因和全长转录本的鉴定及分析 [J]. 昆虫学报, 2023, 66(1): 11-18.
Cai Z B, Wang Z X, Wu Y, et al. Identification and analysis of genes and full-length transcripts related to cytochrome P450 in *Apis cerana cerana* [J]. Acta Entomologica Sinica, 2023, 66(1): 11-18.

[9] 隆 琦, 余尚骏, 吴 鹰, 等. 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 [J]. 应用昆虫学报, 2021, 58(5): 1073-1082.
Long Q, Yu K J, Wu Y, et al. Identification and analysis of virulence factor-related full-length transcripts in *Ascosphaera apis* mycelium and spores [J]. Chinese Journal of Applied Entomology, 2021, 58(5): 1073-1082.

[10] 杜 宇, 祝智威, 王 杰, 等. 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 [J]. 中国农业科学, 2021, 54(4): 864-876.
Du Y, Zhu Z W, Wang J, et al. Construction and annotation of *Ascosphaera apis* full-length transcriptome utilizing nanopore third-generation long-read sequencing technology [J]. Scientia Agricultura Sinica, 2021, 54(4): 864-876.

[11] 杜 宇, 王 杰, 蒋海滨, 等. 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 [J]. 微生物学报, 2021, 61(3): 667-682.

Du Y, Wang J, Jiang H B, et al. Analysis of alternative splicing and polyadenylation of *Ascosphaera apis* genes based on third-generation long-read sequencing dataset [J]. Acta Microbiologica Sinica, 2021, 61(3): 667-682.

[12] 陈华枝, 范元婵, 蒋海滨, 等. 基于纳米孔全长转录组数据完善东方蜜蜂微孢子虫的基因组注释 [J]. 中国农业科学, 2021, 54(6): 1288-1300.
Chen H Z, Fan Y C, Jiang H B, et al. Improvement of *Nosema ceranae* genome annotation based on nanopore full-length transcriptome data [J]. Scientia Agricultura Sinica, 2021, 54(6): 1288-1300.

[13] 陈华枝, 范小雪, 范元婵, 等. 东方蜜蜂微孢子虫基因的可变剪接及可变腺苷酸化解析 [J]. 菌物学报, 2021, 40(1): 161-173.
Chen H Z, Fan X X, Fan Y C, et al. Analysis of alternative splicing and alternative polyadenylation of *Nosema ceranae* genes [J]. Mycosystema, 2021, 40(1): 161-173.

[14] 陈华枝, 杜 宇, 范小雪, 等. 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 [J]. 昆虫学报, 2020, 63(12): 1461-1472.
Chen H Z, Du Y, Fan X X, et al. Construction and annotation of the full-length transcriptome of *Nosema ceranae* based on the third-generation nanopore sequencing technology [J]. Acta Entomologica Sinica, 2020, 63(12): 1461-1472.

[15] Strand M R. The insect cellular immune response [J]. Insect Science, 2008, 15: 1-14.

[16] 吴 珊, 凌尔军. 昆虫细胞免疫反应中的吞噬、集结和包裹作用 [J]. 昆虫学报, 2009, 52(7): 791-798.
Wu S, Ling E J. Phagocytosis, nodulation and encapsulation in cellular immune responses in insects [J]. Acta Entomologica Sinica, 2009, 52(7): 791-798.

[17] Wang G J, Wang W W, Liu Y, et al. Steroid hormone 20-hydroxyecdysone promotes CTL1-mediated cellular immunity in *Helicoverpa armigera* [J]. Insect Science, 2021, 28(5): 1399-1413.

[18] Bergeret E, Perrin J, Williams M, et al. TM9SF4 is required for *Drosophila cellular* immunity via cell adhesion and phagocytosis [J]. Journal of Cell Science, 2008, 121(Pt 20): 3325-3334.

[19] Park D, Jung J W, Choi B S, et al. Uncovering the novel characteristics of Asian honey bee, *Apis cerana*, by whole genome sequencing [J]. BMC Genomics, 2015, 16(1): 1.

[20] 杜 宇, 付中民, 祝智威, 等. 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 [J]. 昆虫学报, 2020, 63(11): 1345-1357.
Du Y, Fu Z M, Zhu Z W, et al. 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* [J]. Acta Entomologica Sinica, 2020, 63(11): 1345-1357.

[21] Abdel-Ghany S E, Hamilton M, Jacobi J L, et al. A survey of

- the sorghum transcriptome using single-molecule long reads [J]. Nature Communications, 2016, 7: 11706.
- [22] Bailey T L, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers [J]. Proceedings International Conference on Intelligent Systems for Molecular Biology, 1994, 2: 28-36.
- [23] Foissac S, Sammeth M. ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets [J]. Nucleic acids research, 2007, 35: W297-W299.
- [24] 陈华枝, 付中民, 王杰, 等. 中华蜜蜂 6 日龄幼虫响应蜜蜂球囊菌胁迫的环状 RNA 应答 [J]. 微生物学报, 2020, 60(10): 2292-2310.
- Chen H Z, Fu Z M, Wang J, et al. Circular RNA response of *Apis cerana cerana* 6-day-old larvae to *Ascosphaera apis* stress [J]. Acta Microbiologica Sinica, 2020, 60(10): 2292-2310.
- [25] 杜宇, 冯睿蓉, 王杰, 等. 中华蜜蜂 6 日龄幼虫响应蜜蜂球囊菌侵染的长链非编码 RNA 应答研究 [J]. 微生物学报, 2021, 61(5): 1338-1358.
- Du Y, Feng R R, Wang J, et al. Long non-coding RNA response of 6-day-old *Apis cerana cerana* larvae to *Ascosphaera apis* infection [J]. Acta Microbiologica Sinica, 2021, 61(5): 1338-1358.
- [26] 冯睿蓉, 付中民, 杜宇, 等. 中华蜜蜂幼虫肠道中微小 RNA 的鉴定及分析 [J]. 中国农业科学, 2022, 55(1): 208-218.
- Feng R R, Fu Z M, Du Y, et al. Identification and analysis of micro RNAs in the larval gut of *Apis cerana cerana* [J]. Scientia Agricultura Sinica, 2022, 55(1): 208-218.
- [27] Schuster S L, Hsieh A C. The untranslated regions of mRNAs in cancer [J]. Trends in Cancer, 2019, 5(4): 245-262.
- [28] Navarro E, Mallén A, Hueso M. Dynamic variations of 3'UTR length reprogram the mRNA regulatory landscape [J]. Bio-medicines, 2021, 9(11): 1560.
- [29] Elkon R, Ugalde A P, Agami R. Alternative cleavage and polyadenylation: extent, regulation and function [J]. Nature Reviews Genetics, 2013, 14(7): 496-506.
- [30] Wang R J, Zheng D H, Yehia G, et al. A compendium of conserved cleavage and polyadenylation events in mammalian genes [J]. Genome Research, 2018, 28(10): 1427-1441.
- [31] Hoque M, Ji Z, Zheng D H, et al. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing [J]. Nature Methods, 2013, 10(2): 133-139.
- [32] Tian B, Manley J L. Alternative polyadenylation of mRNA precursors [J]. Molecular Cell Biology, 2017, 18(1): 18-30.
- [33] Tian B, Hu J, Zhang H B, et al. A large-scale analysis of mRNA polyadenylation of human and mouse genes [J]. Nucleic Acids Research, 2005, 33(1): 201-212.
- [34] Fu H H, Yang D W, Su W Y, et al. Genome-wide dynamics of alternative polyadenylation in rice [J]. Genome Research, 2016, 26(12): 1753-1760.
- [35] Ma H, Lin J C, Mei F M, et al. Differential alternative polyadenylation of homoeologous genes of allohexaploid wheat ABD subgenomes during drought stress response [J]. The Plant Journal: for Cell and Molecular Biology, 2023, 114(3): 499-518.
- [36] Yu Z B, Hong L, Li Q Q. Signatures of mRNA alternative polyadenylation in arabidopsis leaf development [J]. Frontiers in Genetics, 2022, 13: 863253.
- [37] Zhu S, Wu X H, Fu H J, et al. Modeling of genome-wide polyadenylation signals in *Xenopus tropicalis* [J]. Frontiers in Genetics, 2019, 10: 647.
- [38] Chakrabarti M, de Lorenzo L, Abdel-Ghany S E, et al. Wide-ranging transcriptome remodelling mediated by alternative polyadenylation in response to abiotic stresses in sorghum [J]. The Plant Journal: for Cell and Molecular Biology, 2020, 102(5): 916-930.
- [39] Fang L Y, Guo L N, Zhang M, et al. Analysis of polyadenylation signal usage with full-length transcriptome in *Spodoptera frugiperda* (Lepidoptera: Noctuidae) [J]. Insects, 2022, 13(9): 803.
- [40] Retelska D, Iseli C, Bucher P, et al. Similarities and differences of polyadenylation signals in human and fly [J]. BMC Genomics, 2006, 7: 176.
- [41] Jafar Z, Tariq S, Sadiq I, et al. Genome-wide profiling of polyadenylation events in maize using high-throughput transcriptomic sequences [J]. G3 (Bethesda, Md.), 2019, 9(8): 2749-2760.
- [42] 孙明会, 刘佳美, 王思懿, 等. 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 [J]. 微生物学报, 2022, 62(8): 2981-2994.
- Sun M H, Liu J M, Wang S Y, et al. Unraveling the complexity of transcriptome in *Ascosphaera apis* spore: based on third-generation and next-generation sequencing [J]. Acta Microbiologica Sinica, 2022, 62(8): 2981-2994.
- [43] Yuan F W, Hankey W, Wagner E J, et al. Alternative polyadenylation of mRNA and its role in cancer [J]. Genes & Diseases, 2019, 8(1): 61-72.
- [44] Blake D, Lynch K W. The three as: alternative splicing, alternative polyadenylation and their impact on apoptosis in immune function [J]. Immunological reviews, 2021, 304(1): 30-50.
- [45] Guo B H, Dai Y, Chen L, et al. Genome-wide analysis of the soybean root transcriptome reveals the impact of nitrate on alternative splicing [J]. G3, 2021, 11(7): 162.
- [46] Levesque L A, Roy S, Salazar N. CXCR3 expression and genome-wide 3' splice site selection in the TCGA breast cancer cohort [J]. Life, 2021, 11(8): 746.
- [47] Laloum T, Martin G, Duque P. Alternative splicing control of abiotic stress responses [J]. Trends in Plant Science, 2018, 23(2): 140-150.