研究论文

意大利蜜蜂 MAPK 信号通路相关基因和全长转录本鉴定及分析*

范小雪[1, 2, 3**] 荆 欣[1**] 康 婧[1] 高旭泽[1] 张艺琼[1] 姚雨彤[1]
毛予立[1] 牛庆生[4] 陈大福[1, 2, 3] 付中民[1, 2, 3***] 郭 睿[1, 2, 3***]

（1. 福建农林大学蜂学与生物医药学院，福州 350002；2. 天然生物毒素国家地方联合工程实验室，福州 350002；
3. 福建省蜂疗研究所，福州 350002；4. 吉林省养蜂科学研究所，吉林 132000）

摘 要 **【目的】** 利用纳米孔（Nanopore）测序数据鉴定意大利蜜蜂 *Apis mellifera ligustica* 丝裂原活化蛋白激酶（Mitogen activated protein kinase，MAPK）信号通路相关基因和全长转录本，并发掘 MAPK 信号通路相关基因的可变多聚腺苷酸化（Alternative polyadenylation，APA）位点及可变剪接（Alternative splicing，AS）事件，以期加深对意大利蜜蜂 MAPK 信号通路的认识，为持续深入开展相关基因和剪接体（Isoform）的功能研究提供数据资源和基础。**【方法】** 采用 Blast 工具将所有全长转录本比对到 Nr 数据库和 KEGG 数据库，再根据注释信息筛选出 MAPK 信号通路相关基因和全长转录本。利用 gffcompare 软件将 MAPK 信号通路相关全长转录本与西方蜜蜂参考基因组（Amel_HAv3.1）上注释的转录本进行比较以优化已注释基因结构。采用 TAPIS pipeline 预测 APA 位点，并通过 MEME 软件鉴定所有 APA 位点上游 50 bp 的基序（motif）。使用 Astalavista 软件鉴定 AS 事件并统计不同类型的 AS 事件数目。通过 PCR 验证 MAPK 信号通路相关基因的 AS 事件。**【结果】** 共鉴定到意大利蜜蜂 MAPK 信号通路相关的 83 个基因与 443 条全长转录本。同时延长了 1 个基因的 5′ UTR 和 3′ UTR，延长了 12 个基因的 5′ UTR，延长了 10 个基因的 3′ UTR。共鉴定到 52 个 MAPK 信号通路相关基因含有 1 个及以上的 APA 位点，并在 APA 位点上游鉴定到 2 个 motif。共鉴定到 MAPK 信号通路相关的 13 个基因的 21 次 AS 事件，其中最丰富的类型为外显子跳跃（Exon skipping，ES）。PCR 结果证实了骨形态发生蛋白受体 1b 基因的 4 次 ES 事件和 Ets DNA 结合蛋白 pokkuri 基因的 1 次 ES 事件的真实性。**【结论】** 首次鉴定到意大利蜜蜂 MAPK 信号通路相关的 83 个基因和 446 条全长转录本，优化了该信号通路中西方蜜蜂参考基因组已注释的 21 个基因的结构，发掘出 MAPK 信号通路相关基因的 406 个 APA 位点和 52 次 AS 事件，证实了 MAPK 信号通路相关基因 AS 事件的真实性。

关键词 意大利蜜蜂；MAPK 信号通路；全长转录本；可变剪接；可变多聚腺苷酸化

Identification and analysis of MAPK signaling pathway-related genes and full-length transcripts in *Apis mellifera ligustica*

FAN Xiao-Xue[1, 2, 3**] JING Xin[1**] KANG Jing[1] GAO Xu-Ze[1] ZHANG Yi-Qiong[1]
YAO Yu-Tong[1] MAO Yu-Li[1] NIU Qing-Sheng[4] CHEN Da-Fu[1, 2, 3]
FU Zhong-Min[1, 2, 3***] GUO Rui[1, 2, 3***]

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

*资助项目 Supported projects：国家自然科学基金面上项目（32372943）；国家现代农业产业技术体系专项资金项目（CARS-44-KXJ7）；福建省自然科学基金面上项目（2022J01131334，2023J01133656）；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（KFb22060XA）；福建农林大学动物科学学院（蜂学学院）科研扶持项目（郭睿，付中民）

**共同第一作者 Co-first authors，E-mail：imfanxx@163.com；jingxin6662022@163.com

***共同通讯作者 Co-corresponding authors，E-mail：ruiguo@fafu.edu.cn；zmfchina@163.com

收稿日期 Received：2022-09-01；接受日期 Accepted：2023-11-25

2. National & Local United Engineering Laboratory of Natural Biotoxin, Fuzhou 350002, China; 3. Apitherapy Research Institute of Fujian Province, Fuzhou 350002, China; 4. Apiculture Science Institute of Jilin Province, Jilin 132000, China)

Abstract **[Aim]** To identify genes and full-length transcripts associated with the mitogen activated protein kinase (MAPK) signaling pathway in *Apis mellifera ligustica*, and explore alternative polyadenylation (APA) sites and alternative splicing (AS) events in MAPK signaling pathway-related genes. **[Methods]** All full-length transcripts were aligned to the Nr and KEGG databases, after which MAPK signaling pathway-associated genes and full-length transcripts were screened out. MAPK signaling pathway-associated full-length transcripts were compared with those annotated in the reference genome of *A. mellifera* (Amel_HAv3.1) to optimize the structures of annotated genes. APA sites were predicted using the TAPIS pipeline, followed by identification of motifs 50 bp upstream of all APA sites. AS events were identified using Astalavista software followed by statistics on the number of various kinds of AS events. PCR was used to validate AS events in genes related to the MAPK signaling pathway. **[Results]** In total, 83 genes and 443 full-length transcripts relevant to the MAPK signaling pathway were identified. 5′ UTR and 3′ UTR of one gene were prolonged, and 5′ UTR of 12 genes and 3′ UTR of 10 genes were prolonged, respectively. 52 genes associated with the MAPK signaling pathway contained one or more APA sites and there were two motifs upstream of the APA sites. Twenty-one AS events in 13 genes related to the MAPK signaling pathway were identified, of which the most abundant was exon skipping (ES). PCR confirmed the authenticity of four ES events in the Bone morphogenetic protein receptor type-1B gene and one ES event in the Ets DNA-binding protein pokkuri gene. **[Conclusion]** Eighty-three genes and 446 full-length transcripts associated with the MAPK signaling pathway in *A. m. ligustica* were identified for the first time, structures of 21 annotated genes in the *A. mellifera* reference genome were optimized, 406 APA sites and 52 AS events of genes related to MAPK signaling pathway were discovered, and the authenticity of AS events in MAPK signaling pathway-associated genes was verified.

Key words *Apis mellifera ligustica*; MAPK signaling pathway; full-length transcript; alternative splicing; alternative polyadenylation

蜜蜂是自然界中不可或缺的授粉者，其生态、经济及科研价值均无法替代。意大利蜜蜂 *Apis mellifera ligustica* 作为西方蜜蜂 *Apis mellifera* 亚种之一，因具有卓越的繁殖能力，丰富的蜂产品产量以及出色的环境适应性，而成为养蜂生产的首选蜂种，不仅为养蜂从业者带来了可观的经济回报，更为维护生态平衡、促进生物多样性作出了重要贡献（曾志将，2017）。

丝裂原活化蛋白激酶（Mitogen Activated Protein Kinase，MAPK）级联反应是真核生物的一条关键信号转导途径，其作为信息传递的重要桥梁，有效地将外界刺激转化为细胞内信号，通过调节细胞应激、免疫防御及物质代谢等一系列生命活动，从而引发相应的生理和生化反应，确保细胞在复杂多变的环境中保持动态平衡和行使正常功能（Asl *et al.*，2021）。MAPK 家族主要包括 ERKs（Extracellular signal regulated kinases）、JUKs（Jun amino terminal kinases）和 p38 MAPK。MAPK 信号通路相关研究多集中在人类等哺乳动物，昆虫方面的研究比较有限。Ryan 等（2021）研究发现果蝇 *Drosophila* 中的 p38Kb 可通过与 starvin 和 BAG-3 互作进而调节年龄依赖的蛋白质稳态。目前，蜜蜂的 MAPK 信号通路研究十分滞后，相关信息匮乏。Wang 等（2018）曾在中华蜜蜂 *Apis cerana cerana* 工蜂中预测并克隆到丝裂原活化蛋白激酶激酶 6 基因 *AccMKK6*，并通过 RNAi 揭示沉默 *AccMKK6* 能够导致工蜂体内丙二醛含量及超氧化物歧化酶（SOD）和过氧化物酶（POD）活性显著降低。

以 Illumina 测序为代表的二代测序技术具有较高的碱基正确率，在基因表达定量方面的准确度较高，因而广泛应用于基因组和转录组研究，已成功生命科学领域不可或缺的技术手段（Meyer and Kircher，2010）。作为但二代测序产生的读段较短，长度一般介于 100-400 bp，在全长转录本的鉴定和分析方面受到极大限制（Yoshinaga *et al.*，2018）。目前，三代测序的主流代表是 PacBio 单分子实时（Single molecule

real-time，SMRT）测序技术和纳米孔（Nanopore）长读段测序技术，前者通过记录荧光标记的核苷酸在 DNA 聚合酶作用下依次掺入、合成和释放的过程，实现对 DNA 序列的实时读取，后者基于纳米孔电信号检测原理检测和分析 DNA 分子通过纳米孔时的电信号变化，从而实现 DNA 的实时测序。相比于 PacBio SMRT 测序，Nanopore 除了能较为准确地鉴定全长转录本和结构变异，还能进行基因组装和剪接体的表达量分析（van Dijk *et al.*，2018）。此外，Nanopore 测序设备小巧便携，连接笔记本电脑即能工作，因而实用场景更为广泛（van Dijk *et al.*，2018）。Nanopore 测序已成功应用于人类 *Homo sapiens*、小鼠 *Mus musculus*、拟南芥 *Arabidopsis thaliana*、果蝇、酵母 *Saccharomyces* 等动植物和微生物的相关研究，在优化基因结构（Treitli *et al.*，2021），发掘可变多聚腺苷酸化（Alternative polyadenylation，APA）位点（Wang *et al.*，2022）和鉴定可变剪接（Alternative splicing，AS）事件（Shu *et al.*，2022）等方面展现出强大能力。截至目前，在西方蜜蜂全长转录组的研究领域中，利用 Nanopore 测序技术进行的探索相对稀少。其中，He 等（2022）利用 Nanopore 测序技术对西方蜜蜂中的蜂王和工蜂在幼虫发育的三个关键时间点进行了转录本的分析和比较。通过深入的研究，他们成功鉴定并发现了数千个显著差异表达的剪接异构体。这些发现不仅丰富了我们对蜜蜂基因表达多样性的理解，还揭示了这些剪接异构体的表达模式与蜜蜂种姓分化之间的动态关联，为进一步揭示蜜蜂的生物学特性和行为机制提供了宝贵的线索；基于高质量的 Nanopore 长读段测序数据系统鉴定了意大利蜜蜂丝氨酸/苏氨酸蛋白激酶、细胞内吞和 Hippo 信号通路相关基因和全长转录本，并对 AS 事件的真实性进行了分子验证（范小雪等，2022，2023；张佳欣等，2023）。

相较于传统二代测序及其他三代测序技术，Nanopore 测序具有成本更低、读段更长和无需 PCR 扩增等优势（van Dijk *et al.*，2018）。本研究对意大利蜜蜂 MAPK 信号通路相关基因和全长转录本进行系统鉴定，优化参考基因组已注释基因的结构，分析和验证相关基因的 AS 事件，发掘和分析相关基因的 APA 位点，以期丰富意大利蜜蜂 MAPK 信号通路相关基因和全长转录本信息，完善西方蜜蜂参考基因组相关注释，并为持续深入开展功能研究奠定基础。

1 材料与方法

1.1 供试生物材料

本研究所用的意大利蜜蜂工蜂取自福建农林大学蜂学与生物医药学院教学蜂场的实验蜂群。

1.2 MAPK 信号通路相关基因和全长转录本的筛选及分析

笔者团队前期已制备意大利蜜蜂 7 日龄和 10 日龄工蜂中肠样品（n=3），并完成 RNA 提取、cDNA 建库和 Nanopore 测序，上述实验设置 1 次生物学重复（范小雪等，2022）。获得的高质量长读段数据可用于本研究中 MAPK 信号通路相关基因和全长转录本的鉴定、分析及验证。通过 Blast 工具将鉴定到的全长转录本序列比对到 Nr 数据库（https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/）和 KEGG 数据库（https://www.kegg.jp/），再根据数据库注释结果筛选 MAPK 信号通路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 MAPK 信号通路相关基因的结构优化

参照本团队前期已建立的技术流程（杜宇等，2020；范小雪等，2022；郭思佳等，2023）的方法对 MAPK 信号通路相关全长转录本进行发掘。采用 gffcompare 工具（Pertea and Pertea，2020）将识别到的转录本与西方蜜蜂参考基因组（版本号：Amel_HAv3.1）上已注释转录本进行比对分析，如果有读段比对上注释基因边界外区域，则通过向上游或下游延伸非翻译区（Untranslated regions，UTR）对基因边界进行修正。

1.4 MAPK 信号通路相关基因的 APA 位点鉴定与分析

采用 TAPIS pipeline（Abdel-Ghany *et al.*，2016）鉴定 MAPK 信号通路相关基因的 APA 位

点，参照杜宇等（2021）搭配的方法设置参数。通过 MEME 软件（Bailey *et al.*，2006）对 MAPK 信号通路相关基因 APA 位点上游 50 bp 的序列特征进行分析以鉴定 motif。

1.5　MAPK 相关基因的 AS 分析

AS 事件类型主要包括：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）及可变 5′ 端剪接（Alternative 5′ splice site，A5SS），参照陈华枝等（2020）的方法，使用 Astalavista 软件（Foissac and Sammeth，2007）鉴定意大利蜜蜂 MAPK 信号通路相关基因的 AS 事件类型，然后根据预测结果统计各类型的 AS 事件数目。

1.6　AS 事件的 PCR 验证

随机选择涉及 1 种 AS 类型（ES）的 2 个基因进行 PCR 验证，包括骨形态发生蛋白受体 1b（Bone morphogenetic protein receptor type-1B）基因 *BMPR-1B*（GeneBank ID: LOC408442）和 Ets DNA 结合蛋白 pokkuri（Ets DNA-binding protein pokkuri）基因 *AOP*（GeneBank ID：LOC100576369）。利用 Primer premier 5 软件（Singh *et al.*，1998）设计跨剪接位点的特异性引物（表 1），并委托生工生物（上海）公司合成。以肌动蛋白基因 *actin*（GeneBank ID：LOC107999330）作为内参。使用 Promega（美国）提供的 RNA 抽提试剂盒提取上述工蜂中肠样品的总 RNA。通过 Hifair® Ⅱ 1st Strand cDNA Synthesis Kit（Yeasen，上海）将上述的总 RNA 逆转录成 cDNA。利用跨剪接位点引物进行 PCR，反应体系和条件参照范小雪等（2022）的报道。反应完成后，利用 3%琼脂糖凝胶电泳检测扩增产物，再使用核酸凝胶成像仪（Bio-red，美国）进行观察和拍照。

表 1　引物序列信息

Table 1　Information about primer sequences

名称 Name	序列（5′-3′） Sequence (5′-3′)
ONT.6802.5-ONT.6802.1-F	CAGGTCGGTGCGGATGTAAATA
ONT.6802.5-ONT.6802.1-R	TCTGGTGGTAAACAGCCAAATG
ONT.6802.5-ONT.6802.7-F	GCTGCGCTCTTTGCTACCCCAG
ONT.6802.5-ONT.6802.7-R	TTTGCATTGCATGAAACCTTGT
ONT.6802.3-rna25751-F	TGGCTGCGCTCTTTGCTACCCC
ONT.6802.3-rna25751-R	GCATGAAACCTTGTTCATCTGG
ONT.6802.3-ONT.6802.9-F	GGTCGGTGCGGATGTAAATACG
ONT.6802.3-ONT.6802.9-R	ATTTTTTCCTTGTAAATGTGGC
ONT.4398.1-ONT.4398.2-F	TGGGTAAAAACATGTTATACTT
ONT.4398.1-ONT.4398.2-R	GTCTTTCTCCATTACATTCTGA
actin-F	TTATATGCCAACACTGTCCTTT
actin-R	AGAATTGATCCACCAATCCA

2　结果与分析

2.1　意大利蜜蜂 MAPK 信号通路相关基因和全长转录本鉴定

共鉴定到 MAPK 信号通路相关的 83 个基因与 443 条转录本，包括 7 个新基因和 207 条新转录本。其中注释到促分裂原活化蛋白激酶（Mitogen-activated protein kinase，MAPK）的基因和全长转录本数量最多，分别为 1 个和 35 条；其次为注释到 14-3-3ζ 蛋白（14-3-3 protein zeta，

14-3-3ζ）的基因（2 个）和全长转录本（32 条）及注释到 Ras 相关 GTP 结合蛋白（Ras-like GTP-binding protein，Ras-like GTPases）的基因（1 个）和全长转录本（25 条）。MAPK 信号通路基因和全长转录本的详细信息见表 2。

2.2 参考基因组已注释 MAPK 信号通路相关基因的结构优化

对 83 个西方蜜蜂参考基因组中已注释的 MAPK 通路基因进行结构优化分析，其中完成 21 个基因结构优化，包括 13 个正链和 9 个负链结构优化的基因；1 个基因的 5′ UTR 和 3′ UTR 分别延长了 48 和 1 054 bp；12 个基因的 5′ UTR 被延长，延长长度介于 3-14 644 bp 之间；10 个基因的 3′ UTR 被延长，延长长度介于 75-2 163 bp。MAPK 通路相关基因的结构优化信息如表 3 所示。

2.3 MAPK 信号通路相关基因的 APA 位点数量及特征

MAPK 信号通路中，52 个相关基因含有至少 1 个 APA 位点，其中有 7 个基因仅有 1 个 APA 位点，6 个基因则含有 6 个 APA 位点，25 个基因包含的 APA 位点超过 6 个（图 1：A）。APA 位点信息详见表 4。此外，在 MAPK 信号通路相

表 2 意大利蜜蜂 MAPK 通路和全长转录本的详细信息（仅展示 10 个）

Table 2 Detailed information of genes and full-length transcripts relative to MAPK in *Apis mellifera* (only 10 presented)

基因 ID Gene ID	转录本 ID Transcript ID	蛋白家族 Protein family	Nr 注释 Nr annotation	KEGG 通路注释 KEGG pathway annotation
LOC413504	ONT.5105.3	—	serine/threonine-protein kinase mig-15 isoform X16	MAPK signaling pathway
LOC552419	ONT.3435.3	Ras family	ras-like GTP-binding protein RhoL-like isoform X5	MAPK signaling pathway
LOC408442	ONT.6802.1	—	bone morphogenetic protein receptor type-1B isoform X5	MAPK signaling pathway; TGF-beta signaling pathway
LOC408442	ONT.6802.3	Protein tyrosine kinase	bone morphogenetic protein receptor type-1B isoform X5	MAPK signaling pathway; TGF-beta signaling pathway
LOC408442	ONT.6802.5	Protein kinase domain	bone morphogenetic protein receptor type-1B isoform X5	MAPK signaling pathway; TGF-beta signaling pathway
LOC408442	ONT.6802.7	Protein tyrosine kinase	bone morphogenetic protein receptor type-1B-like	MAPK signaling pathway; TGF-beta signaling pathway
LOC408442	ONT.6802.9	Activin types I and II receptor domain	bone morphogenetic protein receptor type-1B isoform X5	MAPK signaling pathway; TGF-beta signaling pathway
LOC408442	rna25751	Protein kinase domain	bone morphogenetic protein receptor type-1B isoform X1	MAPK signaling pathway; TGF-beta signaling pathway
LOC100576369	ONT.4398.1	Sterile alpha motif (SAM)/Pointed domain	ets DNA-binding protein pokkuri-like isoform X1	MAPK signaling pathway; Dorso-ventral axis formation
LOC100576369	ONT.4398.2	Ets-domain	ets DNA-binding protein pokkuri-like isoform X3	MAPK signaling pathway; Dorso-ventral axis formation

表 3 意大利蜜蜂 MAPK 信号通路相关基因结构优化详细信息

Table 3 Detailed information on genes' structure optimization in MAPK signaling pathway of *Apis mellifera ligustica*

基因 ID GeneBank ID	优化后的起止位置 Start and end sites after optimization	正负链 Strand	末端 End	原始位置（nt） Original site (nt)	优化后的位置（nt） Optimized site (nt)
LOC105201154	NC_037641.1:5916546-5919072	-	5′	5 917 130	5 916 546
LOC552419	NC_037643.1:16462333-16474082	-	5′	16 467 405	16 462 333
LOC408442	NC_037651.1:10615924-10622270	-	5′	10 616 724	10 615 924

续表 3 (Table 3 continued)

基因 ID GeneBank ID	优化后的起止位置 Start and end sites after optimization	正负链 Strand	末端 End	原始位置（nt） Original site (nt)	优化后的位置（nt） Optimized site (nt)
LOC411173	NC_037643.1:8913923-8936273	+	3′	8 936 198	8 936 273
LOC105149689	NC_037640.1:12014498-12069261	+	5′	12 034 869	12 014 498
LOC727101	NC_037645.1:12367501-12375441	+	3′	12 373 748	12 375 441
LOC409286	NC_037640.1:1491011-1633745	+	5′	1 505 655	1 491 011
LOC550710	NC_037643.1:16587668-16592974	+	3′	16 592 025	16 592 974
LOC406154	NC_037645.1:2752056-2759474	-	5′	2 754 403	2 752 056
LOC410184	NC_037647.1:11430743-11434281	+	3′	11 433 826	11 434 281
LOC726958	NC_037646.1:1299408-1324277	+	3′	1 323 291	1 324 277
LOC410622	NC_037652.1:5093633-5097486	+	5′	5 093 681	5 093 633
LOC410622	NC_037652.1:5093633-5097486	+	3′	5 096 432	5 097 486
LOC410808	NC_037639.1:15609862-15615734	-	5′	15 612 068	15 609 862
LOC409498	NC_037640.1:4307932-4313442	+	3′	4 312 959	4 313 442
LOC413018	NC_037638.1:18436494-18443147	+	3′	18 442 344	18 443 147
LOC725374	NC_037641.1:5889184-5893292	+	3′	5 891 784	5 893 292
LOC724793	NC_037641.1:8528604-8533910	-	5′	8 530 002	8 528 604
LOC552718	NC_037649.1:8084895-8093728	+	5′	8 084 898	8 084 895
LOC410671	NC_037638.1:21701665-21729619	-	3′	21 727 456	21 729 619
LOC409523	NC_037640.1:4939424-4943048	-	5′	4 940 307	4 939 424
LOC411915	NC_037645.1:6317022-6321552	-	5′	6 317 172	6 317 022

表 4　MAPK 信号通路相关基因的 APA 位点信息概要（仅展示 10 个）

Table 4　Summary of APA sites within related genes in MAPK signaling pathway (only 10 presented)

基因 ID GeneBank ID	正负链 Strand	可变多聚腺苷酸化位点数 Number of APA sites	发生可变多聚腺苷酸化的位置 (bp) Position of APA (bp)
LOC411566	-	8	10 560 641,10 557 497, 10 560 663, 10 560 769, 10 560 641, 10 557 491, 10 560 663, 10 557 591
LOC100577174	-	7	4 712 541, 4 710 187, 4 709 893, 4 710 114, 4 710 238, 4 709 895, 4 712 541,
LOC413946	+	1	1 879 811
LOC100576527	+	1	3 593 077
LOC550857	+	4	5 172 180, 5 172 181, 5 171 088, 5 171 807
LOC105201154	-	4	5 915 964, 5 916 544, 5 916 732, 5 916 545
LOC409925	-	1	9 264 608
LOC408442	-	23	10 616 132, 10 616 227, 10 616 252, 10 616 197, 10 616 632, 10 615 924, 10 616 727, 10 615 951, 10 616 571, 10 616 796, 10 618 361, 10 618 380, 10 616 282, 10 616 396, 10 616 543, 10 616 919, 10 618 187, 10 616 229, 10 616 202, 10 616 555, 10 615 946, 10 618 185, 10 61 6791
LOC105247762	-	10	5 007 532, 5 012 253, 5 009 427, 5012628, 5 012 628, 5 007 533, 5 012 387, 5 017 103, 5 012 450, 5 012 602
LOC411173	+	14	8 936 179, 8 936 273, 8 935 888, 8 935 271, 8 935 640, 8 936 604, 8 936 603, 8 935 270, 8 926 299, 8 926 854, 8 935 886, 8 935 627, 8 936 056, 8 935 249

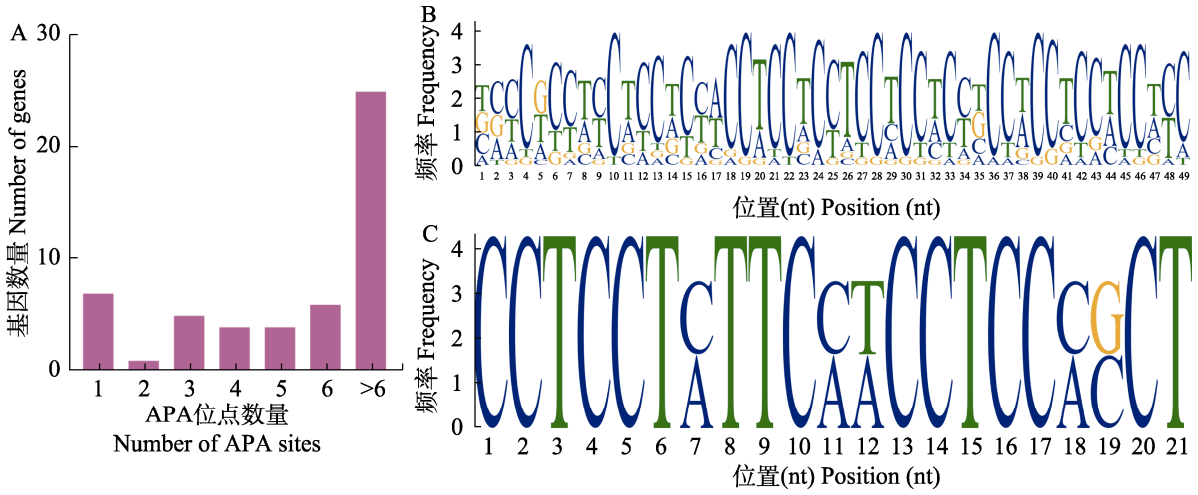

图 1　意大利蜜蜂 MAPK 信号通路相关基因 APA 位点分布特性的统计与分析

Fig. 1　Statistics and analysis of APA sites within genes associated with MAPK signaling pathway in *Apis mellifera ligustica*

A. 含 APA 位点的基因数量统计；B, C. APA 位点上游 50 nt 的 motif 序列。

A. Number statistics of genes with APA sites; B, C. Logo of motifs identified at 50 nt upstream of APA sites.

关的基因 APA 位点上游鉴定到 2 个 motif，一致性序列分别为 NNNBNBNNNYNHNNBNNV-SDHYNMBNBSHSBNHNNMHNSSNHVHHBNHH（图 1：B）和 CCTCCTMTTCMWCCTCC-MSCT（图 1：C）。

2.4　MAPK 信号通路相关基因的 AS 事件鉴定及分析

共鉴定到 MAPK 信号通路相关的 13 个基因的 21 次 AS 事件，其中发生次数最多的 AS 类型为 ES（9 次），其次是 A5SS（7 次）、IR（2 次）、A3SS（2 次）和 MEE（1 次）。上述 13 个基因的 21 次 AS 事件详细信息见表 3。

2.5　MAPK 信号通路相关基因的 AS 事件验证

基于 Nanopore 测序数据的生物信息学预测结果显示，MAPK 信号通路相关基因 *BMPR-1B*（LOC408442）发生 4 次 ES；*AOP*（LOC100576369）

表 5　MAPK 信号通路相关基因的可变剪接事件详细信息

Table 5　Detail information about AS events of genes relative to MAPK signaling pathway

基因 ID GeneBank ID	转录本 ID Transcript ID	AS 事件类型 AS event type	区域 Region
LOC408442	ONT.6802.5, ONT.6802.1; ONT.6802.5, ONT.6802.7	ES	NC_037651.1:10615947-10622233C (1)
LOC408442	ONT.6802.3, rna25751; ONT.6802.3, ONT.6802.9	ES	NC_037651.1:10615924-10622235C
LOC100576369	ONT.4398.1, ONT.4398.2	ES	NC_037646.1:1373151-1417286C
LOC552419	ONT.3435.32, ONT.3435.6	A5SS	NC_037643.1:16466684-16474043C
LOC552419	ONT.3435.9, ONT.3535.11	A3SS	NC_037643.1:16466684-16474043C
LOC550710	ONT.3443.10, ONT.3543.1	A5SS	NC_037643.1:16587855-16592972W (2)
LOC550710	ONT.3443.9, ONT.3543.1	A5SS	NC_037643.1:16587822-16592974W
LOC406154	ONT.3985.10, ONT.3985.11	A3SS	NC_037645.1:2751939-2759445C
LOC408894	ONT.3204.2, ONT.3204.3	IR	NC_037643.1:7288068-7289375C
LOC105248031	ONT.5069.10, ONT.5069.4	MEE	NC_037647.1:10017432-10036036C

续表 5 (Table 5 continued)

基因 ID GeneBank ID	转录本 ID Transcript ID	AS 事件类型 AS event type	区域 Region
LOC409498	ONT.1616.3, rna6376	A5SS	NC_037640.1:4307956-4313439W
LOC551725	ONT.740.1, ONT.740.4	IR	NC_037638.1:20580996-20587559W
LOC551725	ONT.740.4, rna2962	ES	NC_037638.1:20580999-20587554W
LOC550685	ONT.4912.3, ONT.4912.1	A5SS	NC_037647.1:6400643-6406028W
LOC412916	ONT.4984.1, rna18801	ES	NC_037647.1:8211109-8325208C
LOC412916	ONT.4984.1, ONT.4984.4	ES	NC_037647.1:8211109-8325218C
LOC412916	ONT.4984.1, ONT.4984.3	A5SS	NC_037647.1:8211109-8325208C
LOC412916	ONT.4984.4, ONT.4984.3	A5SS	NC_037647.1:8211109-8325218C
LOC413809	ONT.2488.1, ONT.2488.2	ES	NC_037642.1:564402-571240W
LOC726947	ONT.6727.1, rna25484	ES	NC_037651.1:9363150-9373402W
LOC726947	ONT.6727.1, ONT.6727.5	ES	NC_037651.1:9363150-9373402W

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

发生 1 次 ES（图 2）。PCR 扩增产物的琼脂糖凝胶电泳结果显示，*BMPR-1B* 通过 ES 产生 ONT.6802.1（约 234 bp）和 ONT.6802.5（约 329 bp）（图 2：C），ONT.6802.5（约 329 bp）和 ONT.6802.7（约 278 bp）（图 1：E），ONT.6802.3（约 322 bp）和 rna25751（约 273 bp）（图 2：G），ONT.6802.3（目的条带约 340 bp）和 ONT.6802.9（目的条带约 291 bp）（图 2：I）；*AOP* 通过 ES 产生 ONT.4398.1（约 498 bp）和 ONT.4398.2（约 452 bp）（图 2：K）。该结果说明上述 2 个 MAPK 信号通路相关基因的 AS 事件与生物信息学预测结果相吻合，进一步证实了本研究鉴定到的 AS 事件的真实性和可靠性

3　讨论

Wallberg 等（2019）运用第三代 PacBio 测序技术、10×Chromium、BioNano 和 Hi-C 等最新技术重新组装和注释了西方蜜蜂参考基因组（Amel_HAv3.1），该版本已高度连续并接近染色体水平，为深入开展西方蜜蜂的相关研究奠定了坚实基础。利用 Nanopore 测序产生的长读段数据完善基因组注释已在东方蜜蜂微孢子虫 *Nosema ceranae*（陈华枝等，2020）和蜜蜂球囊菌 *Ascosphaera apis*（杜宇等，2021）等物种中见诸报道，但关于意大利蜜蜂的相关研究仍然缺失。本研究基于已获得的 Nanopore 长读段测序数据鉴定到意大利蜜蜂 MAPK 信号通路相关的 7 个新基因和 207 条新转录本，为西方蜜蜂参考基因组提供了有益补充。

UTR 在基因表达调控方面起到关键作用，例如 miRNA 可通过“种子区”与 3′ UTR 上的 miRNA 反应元件结合进而负调控（或正调控）基因表达（Rouleau *et al.*，2017），5′UTR 能通过募集核糖体和扫描密码子影响蛋白翻译过程（Hinnebusch *et al.*，2016）。在果蝇中，*Dscam1* 基因 3′ UTR 的缺失会严重损害轴突的生长（Zhang *et al.*，2019）。由于测序数据和组装软件的局限性，多数基因组注释基因的 UTR 序列不够完善，本研究对西方蜜蜂参考基因组已注释的 21 个 MAPK 通路相关基因进行了结构优化，其中 1 个基因的 5′ UTR 和 3′ UTR 同时被延长 48 和 1 054 bp；有 12 个基因的 5′UTR 被延长，延长长度介于 3-20 371 bp；有 10 个基因的 3′ UTR 被延长，延长长度介于 75-2 163 bp。这表明即使对于接近染色体水平的高质量基因组，仍能利用 Nanopore 数据进一步完善注释基因的结构。此前，笔者所在团队利用 Nanopore 测序数据优化了西方蜜蜂参

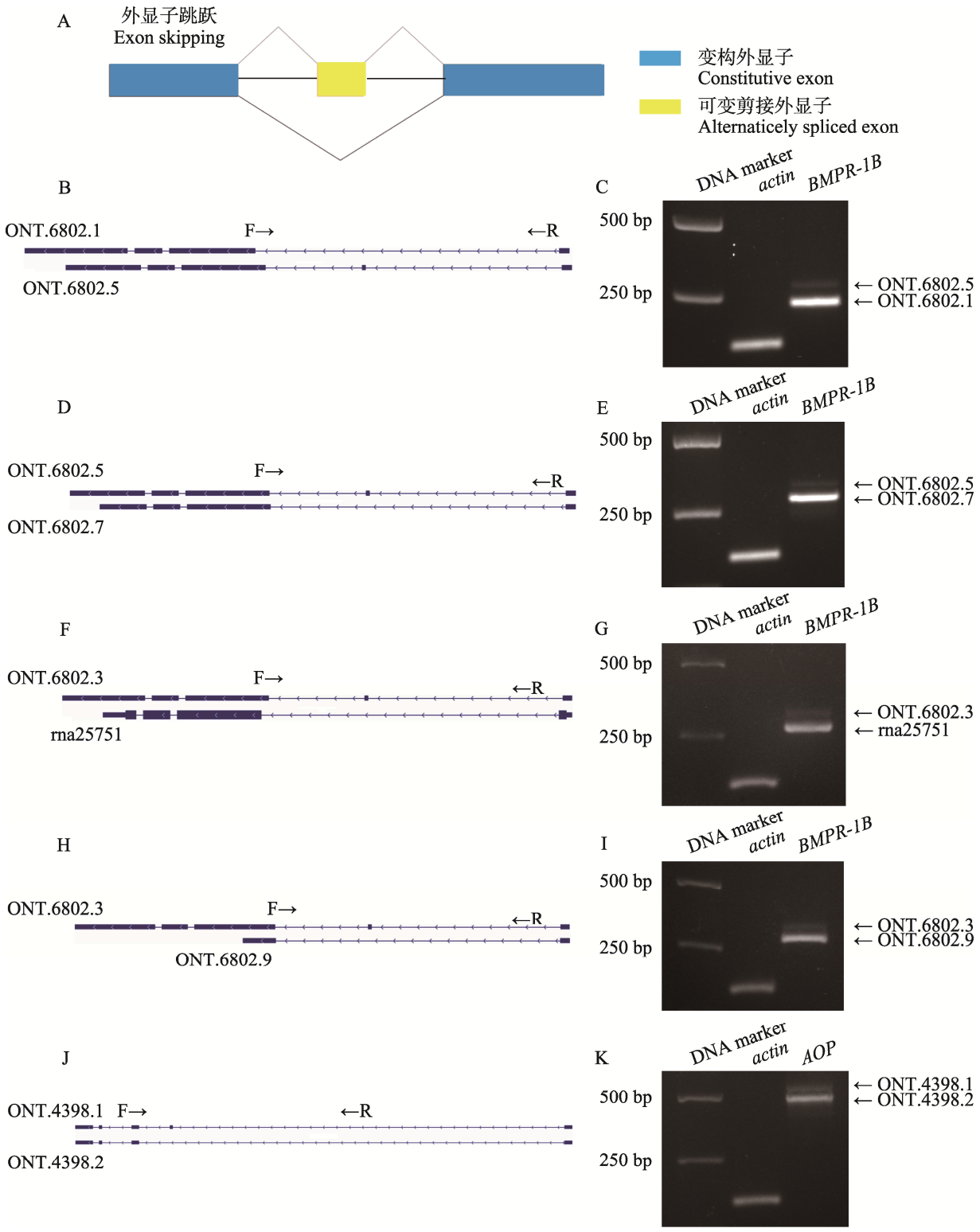

图 2　意大利蜜蜂 MAPK 信号通路相关的 2 个基因的 AS 事件验证

Fig. 2　Validation of AS events of two genes of *Apis mellifera ligustica* MAPK signaling pathway

A. ES 原理图；B, D, F, H, J. 剪接体结构图；C, E, G, I, K. 剪接体扩增产物的凝胶电泳图。

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

因组已注释的丝氨酸/苏氨酸蛋白激酶相关的 14 个基因的结构（范小雪等，2022）。以上结果共同为西方蜜蜂参考基因组的基因序列注释提供了有价值的补充。

前体 mRNA 在其加工过程中可通过 APA 形成不同的转录本或不同长度的 3′ UTR，进而大幅增加转录组和蛋白组的复杂性（Elkon *et al.*，2013）。研究表明在核定位的剪接和多聚腺苷酸化（Cleavage and polyadenylation，CPA）复合物作用下，果蝇的 *DmAtg1* 和 *DmAtg8a* 发生 APA 进而促进自噬作用（Li *et al.*，2022）。本研究共鉴定到意大利蜜蜂 MAPK 信号通路相关的 52 个基因含有 1 个及以上的 APA 位点，其中含有 1-6 个 APA 位点的基因占比分别为 13.46%、1.92%、9.62%、7.69%、7.69%和 11.54%，另有 25 个基因（48.08%）含有 6 个以上的 APA 位点（图 1：A）。这与蜜蜂球囊菌（杜宇等，2021）和果蝇（Vallejos Baier *et al.*，2017）等物种的研究报道相似，体现出真核生物 APA 现象的广泛性和 APA 位点的丰富性。Motif 可影响蛋白翻译后修饰、蛋白质互作及蛋白质运输能力（Rajasekaran，2009）。Zhao 和 Zhou（2019）研究发现果蝇 ZIP 家族成员之一，ZIP13，含有 1 个特殊的 motif（DNXXH），该 motif 使 ZIP13 运输铁离子的能力显著强于其他家族成员。本研究中，在意大利蜜蜂 MAPK 信号通路相关基因的 APA 位点上游共鉴定到 2 个 motif，一致性序列分别为 NNNBNBNNNYNHNNBNNVSDHYNMBNBSHSBNHNNMHNSSNHVHHBNHH（图 1：B）和 CCTCCTMTTCMWCCTCCMSCT（图 1：C），与前人在果蝇（Bandyopadhyay *et al.*，2016）和埃及伊蚊 *Aedes aegypti*（Wu and Cho，2014）等昆虫中鉴定到的 MAPK 通路中的 motif 存在差异，说明 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 MAPK 信号通路值得进一步研究。

真核生物基因组包含有限数量的基因，但通过 AS 可编码数量可观的蛋白质，从而极大地增加转录组和蛋白质组的复杂性（Ule and Blencowe，2019）。ES 是由于功能位点丢失或开放阅读框（Open reading frame，ORF）移动而导致的最常见一种 AS 类型（Kim *et al.*，2020）。Tong 等（2022）研究发现灰飞虱 *Laodelphax striatellus* 的 *JNK2* 基因通过 ES 产生的 3 条剪接体在水稻条纹叶枯病毒感染过程中及灰飞虱的不同组织或器官中具有不同的表达规律，并且发现通过注射特异性靶向 *JNK2* 剪接体的 dsRNA 能促进水稻条纹叶枯病毒在灰飞虱体内的积累。本研究共鉴定到意大利蜜蜂 MAPK 信号通路相关的 2 个基因（LOC408442 和 LOC100576369）的 5 次 ES 形成 8 条剪接体。下一步拟对上述 8 条剪接体进行克隆、原核表达及多克隆抗体制备，为深入探究它们的功能提供基础。

综上，本研究首次鉴定到意大利蜜蜂 MAPK 信号通路相关的 83 个基因和 446 条转录本，优化了西方蜜蜂参考基因组已注释的 MAPK 信号通路相关的 21 个基因结构，发掘出 MAPK 信号通路相关基因的 52 次 AS 事件和 406 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 ASN, 2016. A survey of the sorghum transcriptome using single-molecule long reads. *Nature Communication*, 7: 11706.

Asl ER, Amini M, Najafi S, Mansoori B, Mokhtarzadeh A, Mohammadi A, Lotfinejad P, Bagheri M, Shirjang S, Lotfi Z, Rasmi Y, Baradaran B, 2021. Interplay between MAPK/ERK signaling pathway and microRNAs: A crucial mechanism regulating cancer cell metabolism and tumor progression. *Life Sciences*, 278: 119499.

Bailey TL, Williams N, Misleh C, Li WW, 2006. MEME: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research*, 34 (web server issue): W369–W373.

Bandyopadhyay M, Bishop CP, Bidwai AP, 2016. The conserved MAPK site in E(spl)-M8, an effector of *Drosophila* Notch signaling, controls repressor activity during eye development. *PLoS ONE*, 11(7): e0159508.

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, 蒋海宾, 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.]

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, 蒋海宾, 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, Zhu ZW, Wang J, Wang XN, Jiang HB, Fan YC, Fan XX, Chen HZ, Long Q, Cai ZB, Xiong CL, Zhen YZ, Fu ZM, Chen DF, Guo R, 2021. 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, 2021. 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.]

Elkon R, Ugalde AP, Agami R, 2013. Alternative cleavage and polyadenylation: Extent, regulation and function. *Nature Reviews Genetics*, 14(7): 496–506.

Fan XX, Wang SY, Zhu LR, Jing X, Guo SJ, Niu QS, Jiang HB, Xu XJ, Chen DF, Fu ZM, Xu GJ, Guo R, 2023. Identification of endocytosis-associated genes, and full-length transcripts, in *Apis mellifera ligustica*. *Chinese Journal of Applied Entomology*, 60(3): 825–834. [范小雪, 王思懿, 朱乐冉, 荆欣, 郭思佳, 牛庆生, 蒋海宾, 徐细建, 陈大福, 付中民, 徐国钧, 郭睿, 2023. 意大利蜜蜂细胞内吞相关基因与全长转录本鉴定. 应用昆虫学报, 60(3): 825–834.]

Fan XX, Zhang KY, Wang ZX, Zhu LR, Zhang KH, Niu QS, Xu XJ, Luo Q, Chen DF, Guo R, 2022. Identification and verification of genes and full-length transcripts associated with serine/threonine protein kinases in *Apis mellifera ligustica*. *Acta Entomologica Sinica*, 66(4): 478–485. [范小雪, 张凯遥, 王紫馨, 朱乐冉, 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.]

Foissac S, Sammeth M, 2007. ASTALAVISTA: Dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Research*, 35 (web server issue): W297–W299.

Guo SJ, Zhang KY, Jing X, Gao XZ, Feng PL, Zou PY, Zhang HY, Chen DF, Guo R, 2024. Identification and analysis of full-length transcripts of genes relative to phagocytosis and capsulation in *Apis cerana cerana*. *Journal of Northwest A & F University* (*Natural Science Edition*), 52(3): 1–10. [郭思佳, 张凯遥, 荆欣, 高旭泽, 冯佩林, 邹培缘, 张浩宇, 陈大福, 郭睿, 付中民, 2024. 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. 西北农林科技大学学报(自然科学版), 52(3): 1–10.]

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.

Hinnebusch A, Ivanov I, Sonenberg N, 2016. Translational control by 5′-untranslated regions of eukaryotic mRNAs. *Science*, 352(6292): 1413–1416.

Kim P, Yang MY, Ke YY, Zhao WL, Zhou XB, 2020. ExonSkipDB: Functional annotation of exon skipping event in human. *Nucleic Acids Research*, 48(D1): D896–D907.

Li RS, Xiao Y, Li K, Tian L, 2022. Transcription and post-translational regulation of autophagy in insects. *Frontiers in Physiology*, 13: 825202.

Meyer M, Kircher M, 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010(6): pdb.prot5448.

Pertea G, Pertea M, 2020. GFF utilities: GffRead and GffCompare. *F1000Res.*, 9: ISCBCommJ–ISCBCom304.

Rajasekaran S, 2009. Computational techniques for motif search. *Frontiers in Bioscience*, 14(13): 5052–5065.

Rouleau S, Glouzon JPS, Brumwell A, Bisaillon M, Perreault JP, 2017. 3′ UTR G-quadruplexes regulate miRNA binding. *RNA*, 23(8): 1172–1179.

Ryan SM, Almassey M, Burch AM, Ngo G, Martin JM, Myers D, Compton D, Archie S, Cross M, Naeger L, Salzman A, Virola-Iarussi A, Barbee SA, Mortimer NT, Sanyal S, Vrailas-Mortimer AD, 2021. *Drosophila* p38 MAPK interacts with BAG-3/starvin to regulate age-dependent protein homeostasis. *Aging Cell*, 20(11): e13481.

Shu Z, Wang LG, Wang BJ, Zhang LC, Hou XH, Yan H, Wang LX, 2022. Integrative analysis of nanopore and illumina sequencing reveals alternative splicing complexity in pig longissimus dorsi muscle. *Frontiers in Genetics*, 13: 877646.

Singh VK, Mangalam AK, Dwivedi S, Naik S, 1998. Primer premier:

Program for design of degenerate primers from a protein sequence. *Biotechniques*, 24(2): 318–319.

Sun X, Song L, Yang WJ, Zhang LL, Liu M, Li XS, Tian G, Wang WW, 2020. Nanopore sequencing and its clinical applications. *Methods in Molecular Biology*, 2204: 13–32.

Tong L, Chen XF, Wang W, Xiao Y, Yu JT, Lu H, Cui F, 2022. Alternative splicing landscape of small brown planthopper and different response of *JNK2* isoforms to rice stripe virus infection. *Journal of Virology*, 96(2): e0171521.

Treitli SC, Peña-Diaz P, Hałakuc P, Karnkowska A, Hampl V, 2021. High quality genome assembly of the amitochondriate eukaryote *Monocercomonoides exilis*. *Microbial Genomics*, 7(12): 000745.

Ule J, Blencowe BJ, 2019. Alternative splicing regulatory networks: Functions, mechanisms, and evolution. *Molecular Cell*, 76(2): 329–345.

Vallejos Baier R, Picao-Osorio J, Alonso CR, 2017. Molecular regulation of alternative polyadenylation (APA) within the *Drosophila* nervous system. *Journal of Molecular Biology*, 429(21): 3290–3300.

van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C, 2018. The third revolution in sequencing technology. *Trends In Genetics*, 34(9): 666–681.

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 JM, Xi Y, Ma SC, Qi JJ, Li JP, Zhang RP, Han CC, Li L, Wang JW, Liu HH, 2022. Single-molecule long-read sequencing reveals the potential impact of posttranscriptional regulation on gene dosage effects on the avian Z chromosome. *BMC Genomics*, 23(1): 122.

Wang XX, Wang C, Cui XP, Wang LJ, Liu ZG, Xu BH, Li H, 2018. Molecular mechanism by which *Apis cerana cerana* MKK6 (*AccMKK6*)-mediated MAPK cascades regulate the oxidative stress response. *Bioscience Reports*, 38(6): BSR20181301.

Wu RCC, Cho WL, 2014. Cloning and characterization of microbial activated *Aedes aegypti* MEK4 (*AaMEK4*): Influences of noncatalytic domains on enzymatic activity. *Insect Mol. Biol.*, 23(5): 644–655.

Yoshinaga Y, Daum C, He GF, O'Malley R, 2018. Genome Sequencing. *Methods in Molecular Biology*, 1775: 37–52.

Zeng ZJ, 2017. Apiculture. Beijing: China Agricultural Press, 10–15. [曾志将, 2017. 养蜂学. 北京: 中国农业出版社. 10–15.]

Zhang ZP, So K, Peterson R, Bauer M, Ng H, Zhang Y, Kim JH, Kidd T, Miura P, 2019. Elav-mediated exon skipping and alternative polyadenylation of the *Dscam1* gene are required for axon outgrowth. *Cell Reports*, 27(13): 3808–3817.

Zhao MR, Zhou B, 2020. A distinctive sequence motif in the fourth transmembrane domain confers ZIP13 iron function in *Drosophila melanogaster*. *Biochimica et Biophysica Acta. Molecular Cell Research*, 1867(2): 118607.