

# 意大利蜜蜂工蜂中肠的长链非编码 RNA 的预测、分析及鉴定\*

熊翠玲\*\* 耿四海\*\* 王心蕊 刘思亚 陈大福 郑燕珍  
付中民 杜宇 王海朋 陈华枝 周丁丁 郭睿\*\*\*

(福建农林大学蜂学院, 福州 350002)

**摘要** 【目的】预测、分析和鉴定意大利蜜蜂 *Apis mellifera ligustica* 工蜂中肠的长链非编码 RNA (lncRNA), 探究 lncRNA 的作用。【方法】结合 lncRNA-seq 技术和链特异性 cDNA 建库方法对意大利蜜蜂工蜂中肠进行深度测序; 利用 Perl 脚本对原始数据进行过滤; 联用 CPC 和 CNCI 软件对测序数据进行 lncRNA 预测, 并比较 lncRNA 基因与其邻近的蛋白编码基因的结构特征; 通过 RT-PCR 对随机选取的 lncRNA 进行鉴定; 利用相关生物信息学软件对 lncRNA、的上下游基因进行 GO 分类和 KEGG 代谢通路富集分析。【结果】意大利蜜蜂工蜂中肠样品的 lncRNA-seq 共得到 1 956 118 858 原始读段, 经过滤得到 1 946 489 304 有效读段, 共预测出 6 353 个 lncRNA, 它们较蛋白编码基因含有较少的外显子数、较短的转录本长度。通过 RT-PCR 验证了 9 个 lncRNA 的表达。lncRNA 的上下游基因富集在 42 个 GO 条目和 256 条代谢通路, 进一步分析结果表明这些 lncRNA 参与调控意大利蜜蜂工蜂中肠的新陈代谢和细胞生命活动等生物学过程。【结论】研究结果丰富了蜜蜂的 lncRNA 信息, 也为阐明 lncRNA 在意大利蜜蜂工蜂中肠发育及胁迫响应过程中的作用打下了基础。

**关键词** 意大利蜜蜂, 中肠, lncRNA-seq 技术, 长链非编码 RNA, 上下游基因

## Prediction, analysis and identification of long, non-coding RNA in the midguts of *Apis mellifera ligustica* workers

XIONG Cui-Ling\*\* GENG Si-Hai\*\* WANG Xin-Rui LIU Si-Ya CHEN Da-Fu ZHENG Yan-Zhen  
FU Zhong-Min DU Yu WANG Hai-Peng CHEN Hua-Zhi ZHOU Ding-Ding GUO Rui\*\*\*

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

**Abstract** [Objectives] To predict, analyze and identify lncRNAs in the midguts of *Apis mellifera ligustica* workers, and investigate their functions. [Methods] The midguts of *A. m. ligustica* workers were sequenced with a combination of lncRNA-seq technology and strand-specific cDNA library construction method; the raw data was filtered with perl script; lncRNAs were predicted using CPC and CNCI softwares, and their characteristics were compared with those of adjacent protein-coding genes; RT-PCR was used to identify randomly selected lncRNAs; GO classification and KEGG pathway enrichment analyses of lncRNAs' upstream and downstream genes were carried out using the related bioinformatic software. [Results] A total of 1 956 118 858 raw reads were obtained from the deep sequencing of midgut samples, and 1 946 489 304 clean reads were obtained after filtering. In total, 6 353 lncRNAs were predicted. Compared with protein-coding genes, these lncRNAs genes had fewer exons and shorter transcripts. The expression of nine lncRNAs was identified. Further investigation

\*资助项目 Supported projects: 国家自然科学基金 (31702190); 现代农业产业技术体系建设专项资金 (CARS-44-KXJ7); 福建省科技计划项目 (2018J05042); 福建省教育厅中青年教师教育科研项目 (JAT170158); 福建农林大学科技创新专项基金 (CXZX2017342); 福建农林大学科技发展基金 (KF2015123)

\*\*同等贡献作者 Authors contributed equally, E-mail: pandaxiongcl@sina.com; 15737313592@163.com

\*\*\*通讯作者 Corresponding author, E-mail: ruiguo@fafu.edu.cn

收稿日期 Received: 2018-07-23, 接受日期 Accepted: 2018-10-09

showed that up- and down-stream genes of *A. m. ligustica* lncRNAs were enriched in 42 GO terms and 256 pathways. Moreover, these lncRNAs are found to be engaged in the metabolism and cell-life activity in the midgut of *A. m. ligustica* worker. **[Conclusion]** These results not only enrich the information of honeybee lncRNAs, but also provide a foundation for revealing the role of lncRNAs during the development and stress response processes in the *A. m. ligustica* worker midgut.

**Key words** *Apis mellifera ligustica*, midgut, lncRNA-seq technology, long non-coding RNA, upstream and downstream genes

蜜蜂作为社会行为学研究的模式昆虫,在神经生物学和行为学等方面的研究中应用广泛 (Begna *et al.*, 2012; Foret and Maleszka, 2012; Zayed and Robinson, 2012; Kurze *et al.*, 2016)。此外,蜜蜂也是自然界中最重要的授粉昆虫,在农业生产和生态维持中发挥无法替代的作用 (Committee on the status of pollinators in north America, 2007)。意大利蜜蜂 *Apis mellifera ligustica* 是一种优良的西方蜜蜂 *Apis mellifera* 亚种,具有优越的采集和产浆能力,自 19 世纪引入我国以来已在养蜂生产中占主导地位。

Okazaki 等 (The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team, 2002) 于 2002 年首次在小鼠 *Mus musculus* 中发现并鉴定一类长度大于 200 个核苷酸且不具备编码蛋白能力的转录本,并将其命名为长链非编码 RNA (long non-coding RNA, lncRNA)。此后的较长时间里, lncRNA 被认为是基因组的“暗物质”和转录的“噪音”。近年来,随着高通量测序技术、生物信息学软件和分析方法的迅速发展、高度融合及广泛应用,人类、动物和植物中的大量 lncRNA 被发掘出来 (Pauli *et al.*, 2012; Iyer *et al.*, 2015; Mudge and Harrow, 2015; Qin *et al.*, 2017)。lncRNA 已被证明在细胞分化、生长发育、细胞周期和癌症发生等生物学过程发挥重要的调控功能 (Hainer and Martens, 2011; Cooper *et al.*, 2014; Ouyang *et al.*, 2014)。相较于哺乳动物,昆虫的 lncRNA 研究相对滞后且进展缓慢。虽然在黑腹果蝇 *Drosophila melanogaster* (Young *et al.*, 2012; Chen *et al.*, 2016)、家蚕 *Bombyx mori* (Wu *et al.*, 2016)、埃及伊蚊 *Aedes aegypti* (Kayvan *et al.*, 2016)、小菜蛾 *Plutella xylostella* (Zhu *et al.*,

2017) 和褐飞虱 *Nilaparvata lugens* (Xiao *et al.*, 2015) 等少数昆虫中预测出一些 lncRNA, 但对于包括蜜蜂在内的多数昆虫, lncRNA 的相关信息极为匮乏。Chen 等 (2017) 对西方蜜蜂蜂王卵巢进行测序,筛选出与卵巢发育和产卵相关的 lncRNA。Jayakodi 等 (2015) 对东方蜜蜂和西方蜜蜂的 lincRNA 分析,表明 lincRNA 可通过调控激素信号通路和代谢从而抑制病毒入侵。蜜蜂中肠是食物消化、营养吸收和能量吸收的主要场所。然而到目前为止,有关蜜蜂中肠 lncRNA 的研究未见报道,中肠相关 lncRNA 的数量、结构特征及功能仍未可知。

本研究率先结合 lncRNA-seq 技术和链特异性建库方法对意大利蜜蜂 7 d 和 10 d 工蜂中肠进行深度测序,通过生物信息学方法对中肠的 lncRNA 进行预测和全面分析,研究结果提供了意大利蜜蜂工蜂中肠 lncRNA 的数量、结构特征及功能预测信息,为探究 lncRNA 在中肠发育及胁迫应答过程中的功能打下了基础。

## 1 材料与方法

### 1.1 生物材料

本研究使用的意大利蜜蜂工蜂取自福建农林大学蜂学学院教学蜂场。

### 1.2 测序样品的制备

选择群势较强的健康蜂群作为实验蜂群,将封盖子脾迅速提至实验室,放入 ( $34 \pm 0.5$ ) 培养箱培养,将刚出房的工蜂 (当天记为 0 d) 放入干净的塑料盒 (塑料盒四周打孔以通风, 35 只/盒), 每个盒子上方插入一支装有 50% (w/v) 无菌糖水的饲喂器。 ( $34 \pm 0.5$ ) 培养 24 h, 饥饿处理 2 h, 第一组 (AmNc) 工蜂单只饲喂 5  $\mu$ L

含有  $10^6$  个东方蜜蜂微孢子虫 *Nosema ceranae* 孢子的糖水,第二组(Am)工蜂饲喂不含 *N. ceranae* 孢子的糖水 5  $\mu$ L。每 24 h 检查并清理死去的蜜蜂。冰上操作快速拉取各组 7 d 和 10 d 工蜂的中肠,3 只中肠放入 RNA-Free 的 1.5 mL 离心管,迅速投到液氮速冻后保存于  $-80^{\circ}\text{C}$  超低温冰箱。每组各日龄工蜂制备 18 只中肠样品,其中 9 只用于高通量测序,另外 9 只用于后续分子生物学实验。本次实验进行 3 个生物学重复,第一组包括 AmNc7 : AmNc7-1、AmNc7-2、AmNc7-3,AmNc10 : AmNc10-1、AmNc10-2、AmNc10-3;第二组包括 Am7 : Am7-1、Am7-2、Am7-3,Am10 : Am10-1、Am10-2、Am10-3。本研究旨在对意大利蜜蜂工蜂中肠的 lncRNA 进行预测、分析及鉴定,因而将上述 12 个样品的测序数据混合后用于后续分析,之所以设置正常组和感染组,是因为在本研究的基础上,下一步将对意大利蜜蜂工蜂中肠响应东方蜜蜂微孢子虫 *N. ceranae* 胁迫过程的差异表达 lncRNA (DElncRNA) 进行深入分析。

### 1.3 RNA 提取、链特异性 cDNA 文库构建及 lncRNA-seq

首先用 RNA 抽提试剂盒 (AxyPrep<sup>TM</sup> Multisource Total RNA Miniprep Kit)(TaKaRa 公司,日本)抽提蜜蜂中肠样品的总 RNA,为最大限度地保留所有非编码 RNA(ncRNA),去除核糖体 RNA 后的 mRNA 和 ncRNA 用裂解缓冲液随机打断为小片段作为模板,用六碱基随机引物合成 cDNA 第一链,并加入缓冲液、dNTP、RNase H 和 DNA polymerase I 合成 cDNA 第二链。经过 QiaQuick PCR 试剂盒(Qiagen 公司,德国)纯化并加 EB 缓冲液洗脱经末端修复、加碱基 A,加测序接头,然后通过尿嘧啶-N-糖基化酶(UNG)降解 cDNA 第二链。消化产物经琼脂糖凝胶电泳和 PCR 扩增,PCR 产物经高碘酸钠(Sigma 公司,美国)处理。委托广州基迪奥生物科技有限公司对上述 cDNA 文库进行双端(paired-end)测序,测序平台为 Illumina HiSeq 4000。测序数据已上传 NCBI SRA 数据库,

BioProject 号:PRJNA406998。

### 1.4 lncRNA 的预测与特征分析

利用 perl 脚本对于下机的原始读段(raw data)进行过滤处理,去除低质量数据和接头序列得到有效读段(clean reads),再去除比对到 rRNA 数据库上的 reads,将过滤完 rRNA 的 reads 比对到西方蜜蜂参考基因组(Amel\_4.5)。利用 FPKM(Fragments Per Kilobase of transcript per Million mapped reads)法计算基因表达量。利用 R 语言包(<http://www.r-project.org/>)计算各样品之间的相关性系数。通过与已知的 mRNA 和 lncRNA 参考序列比较,得到所有新的转录本,最后联用 CPC(Kong *et al.*, 2007)和 CNCI(Sun, 2013)软件对转录本的编码能力进行预测,取二者交集作为可靠的预测结果,并对 lncRNA 编码基因的内含子数、外显子数、外显子长度及转录本长度进行统计。

### 1.5 lncRNA 的 RT-PCR 验证

随机选取 10 个预测出的 lncRNA,参照相应的核酸序列,利用 DNAMAN 软件(Lynnon Biosof 公司,美国)设计引物特异性上游引物和下游引物。利用 RNA 抽提试剂盒(Axygen 公司,美国)提取意大利蜜蜂 7 d 和 10 d 工蜂中肠的总 RNA,等摩尔比混合后作为模板进行 RT-PCR 扩增。PCR 反应体系为 20  $\mu$ L,包括 Mixture 10  $\mu$ L,上下游引物各 1  $\mu$ L,cDNA 模板 1  $\mu$ L,无菌水补至 20  $\mu$ L;PCR 程序如下:94  $^{\circ}\text{C}$  预变性 5 min;94  $^{\circ}\text{C}$  变性 50 s,55  $^{\circ}\text{C}$  退火 30 s,72  $^{\circ}\text{C}$  延伸 1 min,共 34 个循环;最后 72  $^{\circ}\text{C}$  延伸 10 min。PCR 扩增产物经 1.5%的琼脂糖凝胶电泳检测和分析。

### 1.6 lncRNA 的上下游基因的预测及分析

lncRNA 上下游基因分析即 *cis* 作用靶基因分析,基本原理为 lncRNA 的功能与其坐标临近的蛋白编码基因相关,位于上游的 lncRNA 可能与启动子或共表达基因的其他顺式作用元件有交集,从而在转录或者转录后水平对基因的表达

进行调控;位于 3' UTR 或者基因下游的 lncRNA 可能参与其他调控作用 (Flirian and Joshua, 2018)。我们对 lncRNA 进行注释, 如果其编码基因位于一个基因的上游或者下游, 则有可能与顺式作用元件所在区域有交集, 从而参与转录调控。对于预测出的 lncRNA 的上下游基因, 利用 OmicsShare (<http://www.omicsshare.com/>) 上的 GO 和 KEGG pathway 富集分析工具进行相关分析。

## 2 结果与分析

### 2.1 数据质控与评估

本研究中, 意大利蜜蜂工蜂中肠样品的转录组测序共得到 1 956 118 858 条 raw reads, 经过滤得到 1 946 489 304 条 clean reads, 两端 Q20 和 Q30 均在 96.98% 和 93.34% 以上 (表 1)。各中肠样品的组内 Pearson 相关系数均在 0.91 以上,

说明样本重复性较好 (图 1)。上述结果说明本研究的转录组测序数据质量良好, 可用于进一步分析。

表 1 LncRNA-seq 数据统计  
Table 1 Overview of lncRNA-seq data

| 样品<br>Samples | 原始读段<br>Raw reads | 有效读段<br>Clean reads | 99% 碱基<br>正确率<br>Q20 | 99.9% 碱基<br>正确率<br>Q30 |
|---------------|-------------------|---------------------|----------------------|------------------------|
| Am7-1         | 160 844 082       | 160 049 106         | 97.41%               | 94.00%                 |
| Am7-2         | 129 878 194       | 129 283 918         | 97.56%               | 94.19%                 |
| Am7-3         | 113 683 898       | 113 165 446         | 97.52%               | 94.03%                 |
| Am10-1        | 160 537 248       | 159 765 346         | 97.27%               | 93.84%                 |
| Am10-2        | 149 230 808       | 148 494 716         | 97.28%               | 93.77%                 |
| Am10-3        | 131 386 354       | 130 619 802         | 96.98%               | 93.34%                 |
| AmNc7-1       | 152 313 278       | 151 668 484         | 97.55%               | 94.15%                 |
| AmNc7-2       | 200 417 896       | 199 313 090         | 97.11%               | 93.35%                 |
| AmNc7-3       | 126 667 596       | 126 053 962         | 97.38%               | 93.85%                 |
| AmNc10-1      | 249 472 666       | 248 333 982         | 97.32%               | 93.81%                 |
| AmNc10-2      | 208 589 832       | 207 574 770         | 97.34%               | 93.77%                 |
| AmNc10-3      | 173 097 006       | 172 166 682         | 97.38%               | 93.74%                 |

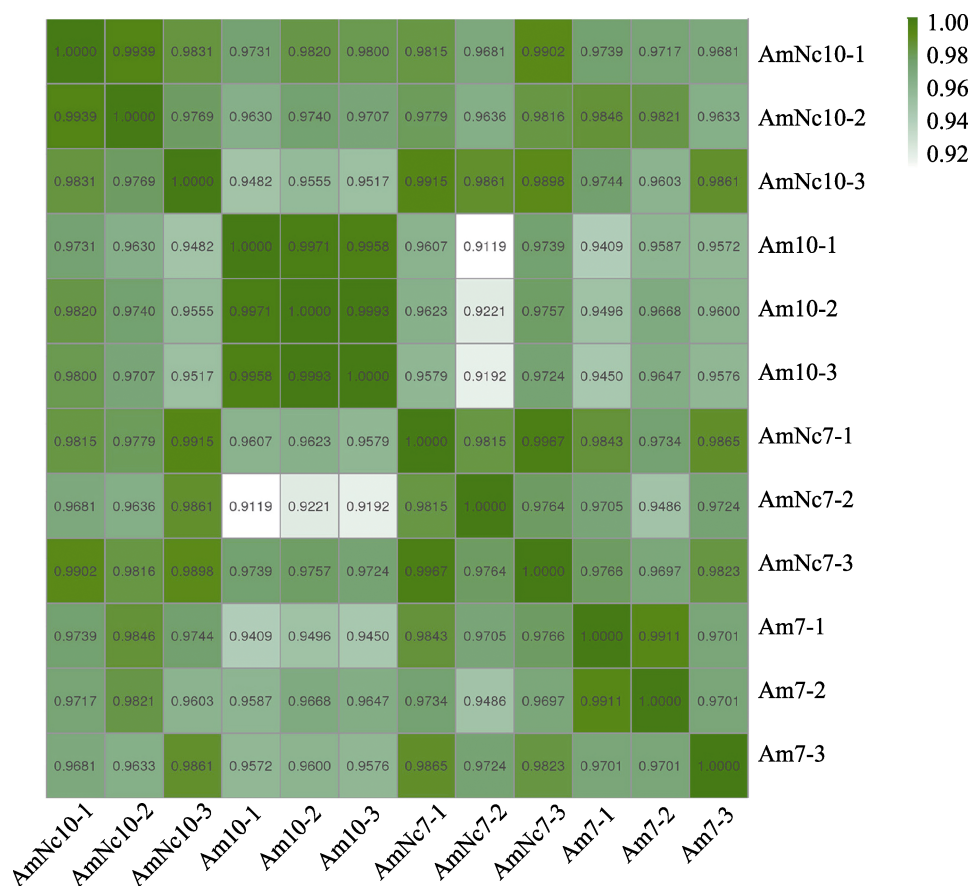


图 1 各意大利蜜蜂工蜂中肠样品不同生物学重复间的相关性

Fig. 1 Pearson correlations between every two biological repeats within each *Apis mellifera ligustica* midgut group

## 2.2 意大利蜜蜂工蜂中肠 lncRNA 的预测及特征分析

利用软件 CPC 和 CNCI 对所有转录本进行蛋白编码能力的预测,共预测出 6 353 个 lncRNA,其中包括 1 604 个未知 lncRNA (novel lncRNA)。上述 lncRNA 基因和蛋白编码基因的内含子长度差别不大,多集中在 2 500 nt 内

(图 2:A)。相较于蛋白编码基因,lncRNA 基因的外显子数多为 2-4 个,其中含 2 个外显子的 lncRNA 最多 (30%);蛋白编码基因的外显子数多为 2 个以上,其中外显子数多于 11 个的蛋白编码基因最多 (28%) (图 2:B)。此外,lncRNA 基因与蛋白编码基因的的外显子长度分布相似,多数分布在 0-2 500 nt 之间 (图 2:C);lncRNA 的转录本长度较蛋白编码基因更短 (图 2:D)。

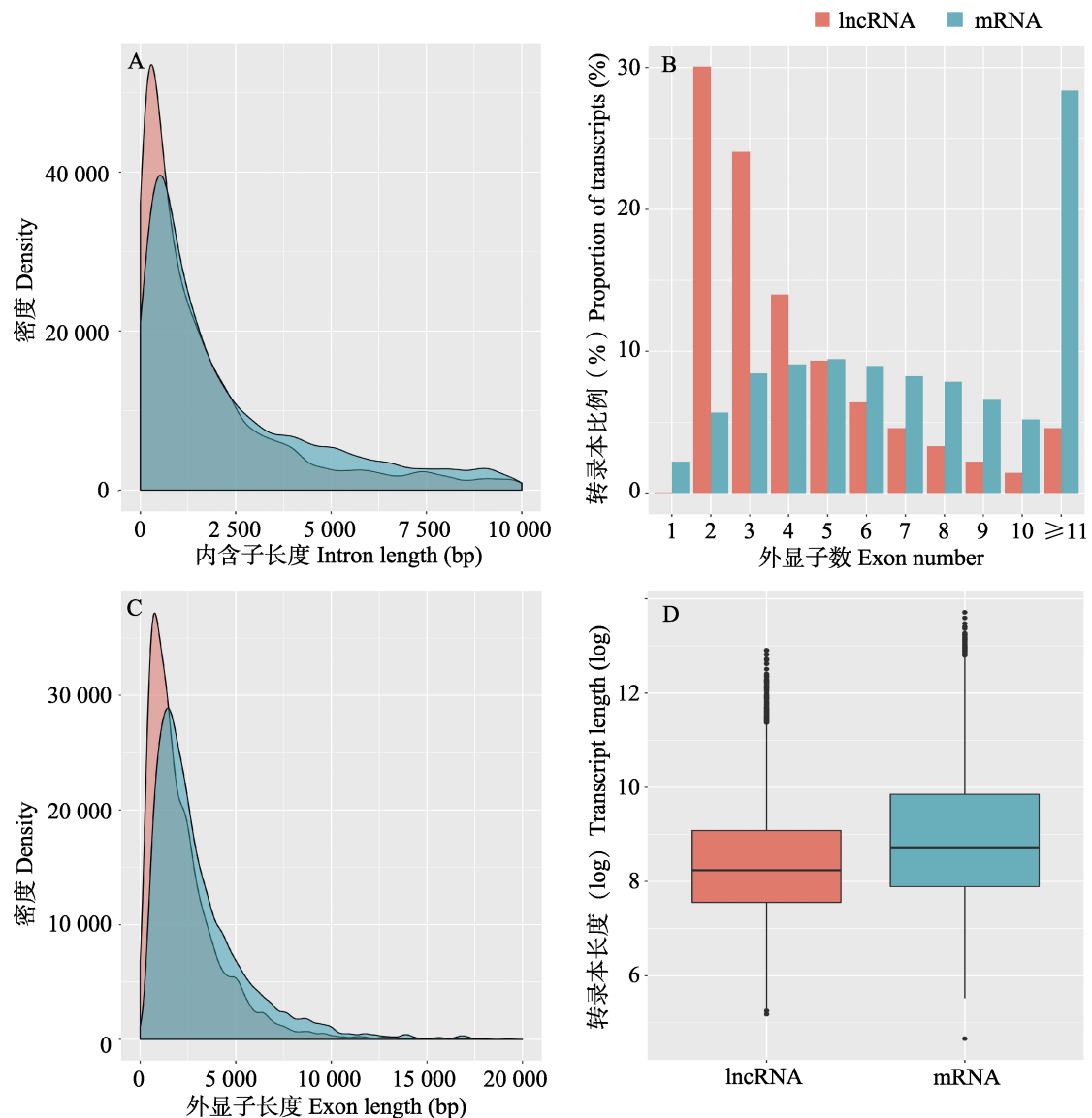


图 2 意大利蜜蜂工蜂中肠 lncRNA 的特征分析

Fig. 2 Characteristic analysis of lncRNAs in the midgut of *Apis mellifera ligustica* worker

A. lncRNA 与 mRNA 的内含子长度比较; B. lncRNA 与 mRNA 的外显子数比较;

C. lncRNA 与 mRNA 外显子长度比较; D. lncRNA 与 mRNA 的转录本长度比较。

A. Intron length comparison of lncRNA and protein-coding gene; B. Exon number comparison of lncRNA and protein-coding gene; C. Exon length comparison of lncRNA and protein-coding gene;

D. Length comparison of lncRNA and protein-coding gene.

## 2.3 意大利蜜蜂工蜂中肠 lncRNA 的 RT-PCR 鉴定

随机选取 10 个 lncRNA 进行 RT-PCR 鉴定, 电泳结果显示成功扩增出 9 个符合预期大小的目的片段(图 3), 说明本研究的 lncRNA-seq 数据真实可靠, 预测的多数 lncRNA 真实存在。相关引物信息详见表 2。

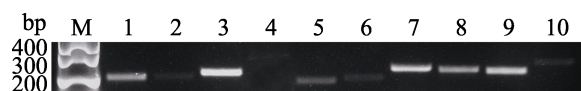


图 3 意大利蜜蜂工蜂中肠 lncRNA 的 RT-PCR 验证

Fig. 3 RT-PCR verification of lncRNAs in the midgut of *Apis mellifera ligustica* worker

Lane M: DNA Marker; Lane1: TCONS\_00013922; Lane 2: TCONS\_00013941; Lane 3: TCONS\_00013942; Lane 4: TCONS\_00015752; Lane 5: TCONS\_00016999; Lane 6: TCONS\_00017000; Lane 7: TCONS\_00015837; Lane 8: TCONS\_00017073; Lane 9: TCONS\_00019496; Lane10: TCONS\_00019499.

表 2 RT-PCR 引物信息

Table 2 Information of primers used in RT-PCR

| 引物名字<br>Primer name | 引物序列 (5'-3')<br>Primer sequences (5'-3') |
|---------------------|------------------------------------------|
| 1F                  | GAGATAGCAGAAGACAGCGAC                    |
| 1R                  | GCAGAGAATACGAGAGTGAAAG                   |
| 2F                  | ATCAGTGAGTGGAAACACGG                     |
| 2R                  | TATCGGAGAAAGGGACGA                       |
| 3F                  | ATCAGTGAGTGGAAACACGG                     |
| 3R                  | GGAGAAAGGGACGAGATGA                      |
| 4F                  | CCACTGAACTCTCTAATCACCA                   |
| 4R                  | GCGTAACCGACGATAAACA                      |
| 5F                  | TGCGTCAAGGGAACGATA                       |
| 5R                  | GCCAATCAACATCCACTCTC                     |
| 6F                  | TGCGTCAAGGGAACGATA                       |
| 6R                  | GCCAATCAACATCCACTCTCT                    |
| 7F                  | GAAACAGGTAGCACAGTCTTTG                   |
| 7R                  | CACGATACAAGGGAACGAA                      |
| 8F                  | GATAAGGAGACGAGAGGTGG                     |
| 8R                  | ATTGAGAGACCAGAGGACAGA                    |
| 9F                  | CGTTCGTTTGTCTTCTGCC                      |
| 9R                  | GCTTCTCCCTTTCTACTCCC                     |
| 10F                 | GCCGATGTTCTCAAACCT                       |
| 10R                 | GACACACGAGGATGGAAATC                     |

## 2.4 意大利蜜蜂工蜂中肠 lncRNA 的上下游基因分析

GO 分类结果显示, 意大利蜜蜂工蜂中肠 lncRNA 的上下游基因富集在生物学进程、细胞组分和分子功能三大类, 富集在生物学进程的前 5 位 GO term 是细胞进程 (349 gene)、代谢进程 (332 gene)、单一有机体进程 (289 gene)、生物调控 (112 gene)、应激反应 (108 gene); 富集在细胞组分的前 5 位 GO term 是细胞 (158 gene)、细胞组件 (158 gene)、细胞膜 (153 gene)、细胞膜组分 (146 gene)、细胞器 (123 gene); 富集在分子功能的前 5 位 GO term 结合 (400 gene)、催化活性 (283 gene)、分子转导活性 (56 gene)、信号传感活性 (46 gene)、运输活性 (40 gene) (图 4)。上述结果说明 lncRNA 可能在意大利蜜蜂工蜂中肠的新陈代谢、细胞生命活动和环境应激中发挥重要作用。

进一步对 lncRNA 的上下游基因进行 KEGG pathway 富集分析, 结果显示上下游基因富集在 256 个 pathway。其中, 富集基因数最多的前 10 位分别是 Wnt 信号通路 (49 gene)、癌症蛋白聚糖 (42 gene)、神经营养因子信号通路 (41 gene)、cAMP 信号通路 (38 gene)、多巴胺突触 (38 gene)、钙离子信号通路 (37 gene)、胃酸分泌物 (37 gene)、催生素信号通路 (36 gene)、心肌细胞肾上腺素信号 (36 gene) 和光传导 (35 gene) (图 5)。上述结果说明 lncRNA 通过广泛参与各类信号通路的调控对意大利蜜蜂工蜂中肠生长和发育产生影响。

## 3 讨论

果蝇 (Soshnev *et al.*, 2011; Young *et al.*, 2012; Chen *et al.*, 2016) 和家蚕 (Wu *et al.*, 2016; Zhou *et al.*, 2017) 等少数模式昆虫的 lncRNA 研究起步较早且相关信息较多。Chen 等 (2016) 在果蝇中发现 1 077 个 lncRNA, 并发现它们的表达具有高度特异性。Soshnev 等 (2011) 发现一个保守性高的 lncRNA (*yar*), 并证明 *yar* 可以调控果蝇的睡眠行为。Wu 等 (2016) 通过对

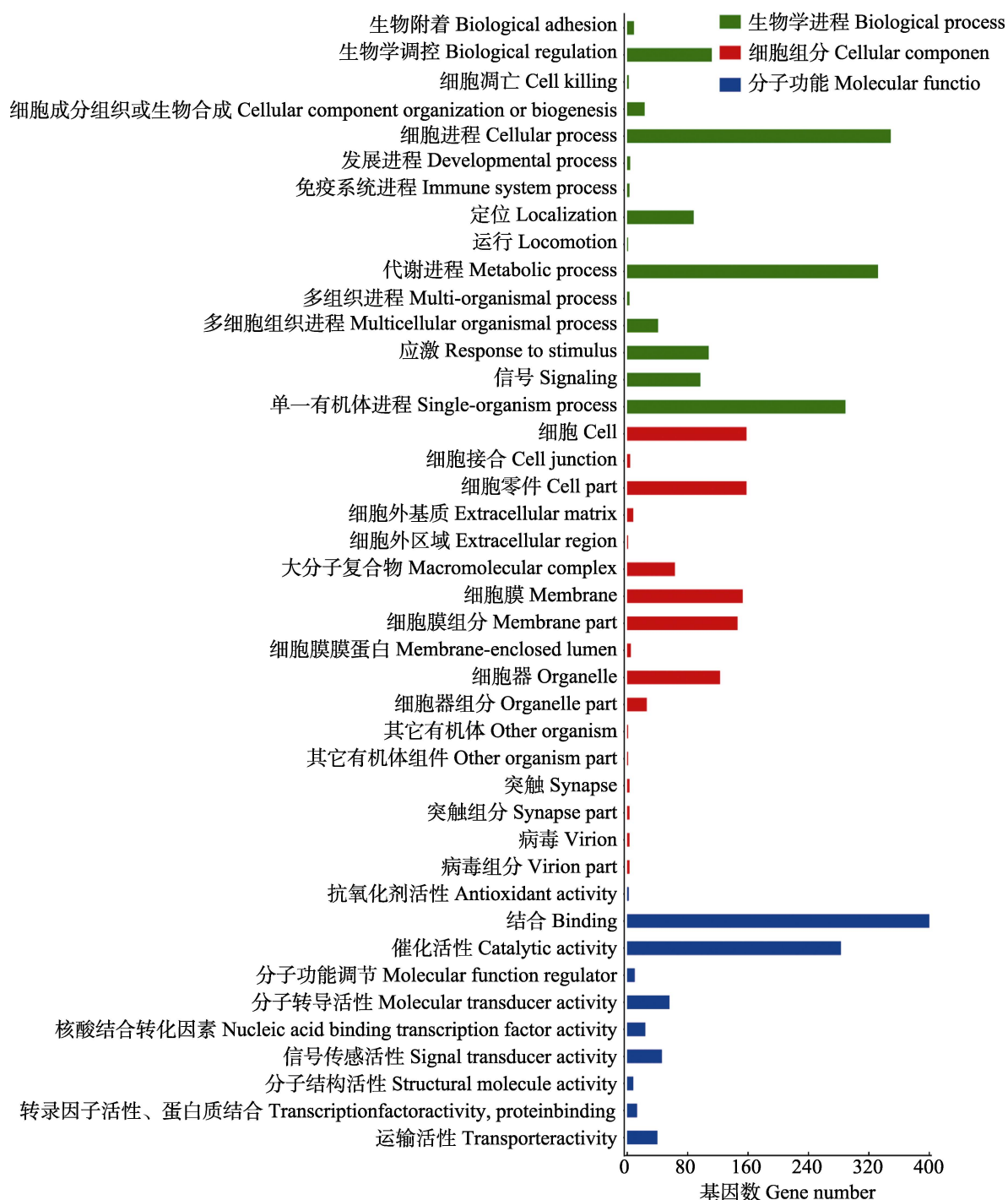


图 4 意大利蜜蜂工蜂中肠 lncRNA 的上下游基因的 GO 富集分析  
Fig. 4 GO enrichment analysis for upstream and downstream genes of lncRNAs in the midgut of *Apis mellifera ligustica* worker

家蚕的转录本深度测序, 预测出 11 810 个 lncRNA。Zhou 等 (2017) 在家蚕和野生蚕的丝腺中鉴定出 600 个 lncRNA, 筛选出两个 lncRNA (*dw4sg\_0040* 和 *dw4sg\_0048*) 可以调控蚕丝蛋白。相比较而言, 蜜蜂的 lncRNA 研究相对滞后

且进展缓慢。Humann 等 (2013) 通过对意大利蜜蜂工蜂 4-6 d 幼虫的卵巢中的 *lncov2* 和 *lncov1* 的功能研究, *lncov2* 对意大利蜜蜂卵巢发育起促进作用, 而 *lncov1* 表达抑制意大利蜜蜂卵巢发育。郭昱等 (2015) 通过对 4-6 d 的工蜂和蜂王

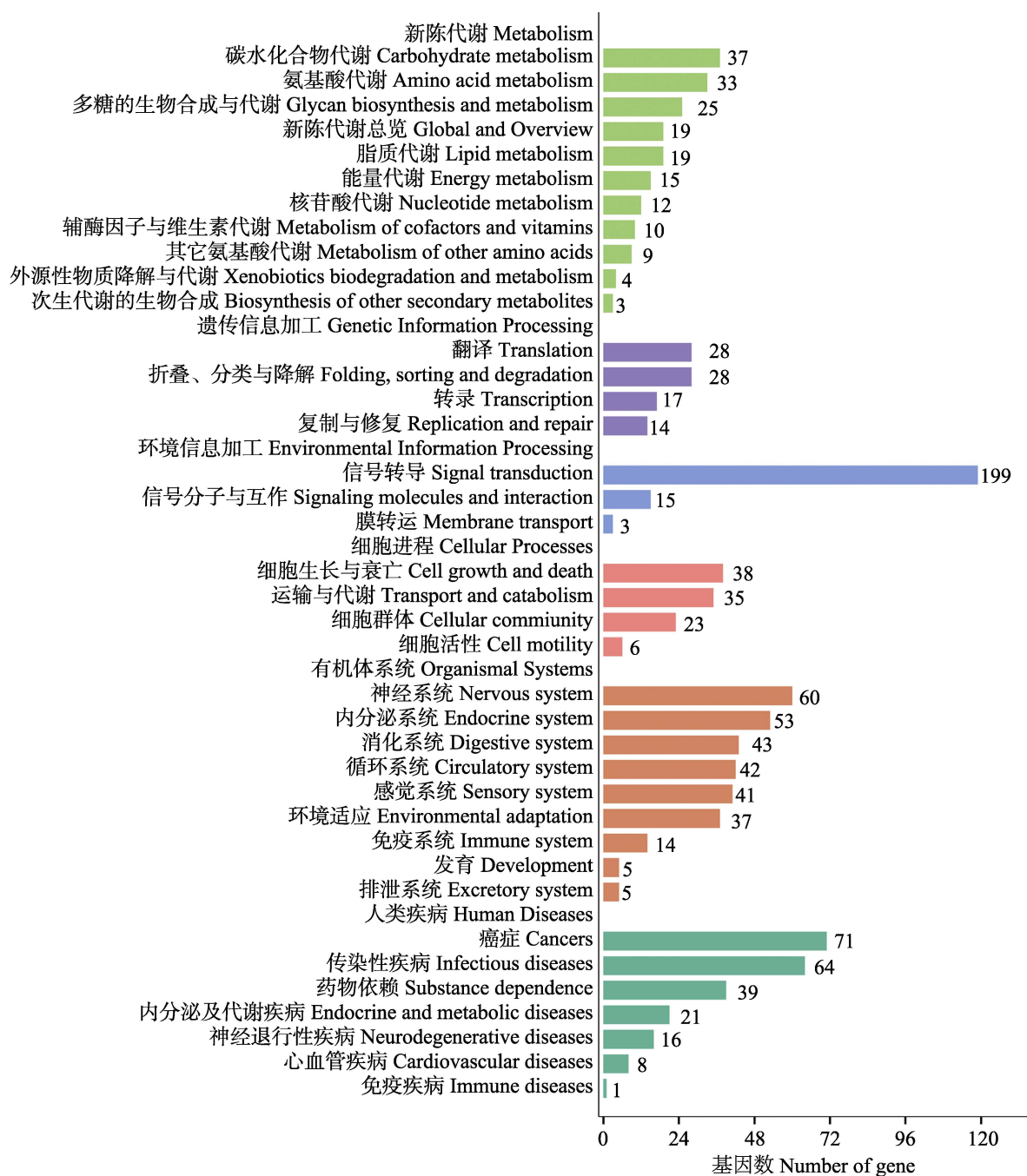


图 5 意大利蜜蜂工蜂中肠 lncRNA 的上下游基因的 KEGG pathway 富集分析

Fig. 5 KEGG pathway enrichment analysis for upstream and downstream genes of lncRNAs in the midgut of *Apis mellifera ligustica* worker

幼虫进行高通量测序预测出 3 993 个 lncRNA, 通过功能分析发现 lncRNA 可能参与调控蜜蜂级型分化。Chen 等 (2017) 对意大利蜜蜂处女王、产卵蜂王、抑制产卵的蜂王和产卵恢复的蜂王进行深度测序, 鉴定到 14 个位于控制卵巢大小数量性状位点 (QTL) 的 lncRNA, 筛选出与意大

利蜜蜂卵巢发育和产卵相关的 2 个 lncRNA (XLOC\_073978 和 XLOC\_081294)。Jayakodi 等 (2015) 在意大利蜜蜂和中蜂分别鉴定出 1 514 和 2 470 个基因间区 lncRNA (lincRNA), 进一步分析发现分别有 863 (57%) 和 464 个 (18%) lincRNA 参与意大利蜜蜂和中蜂的组织发育, 另

有 10 个 lincRNA 在病毒感染时起到关键作用。目前,有关蜜蜂中肠 lincRNA 的研究未见报道。本研究利用 lincRNA-seq 技术对意大利蜜蜂工蜂中肠进行测序,通过生物信息学方法预测出 6 353 个 lincRNA,其中包括 1 604 个 novel lincRNA,为蜜蜂的 lincRNA 信息提供了重要补充。lincRNA 具有发育阶段和组织特异性表达,本研究预测出的 lincRNA 是否在意大利蜜蜂工蜂的其他组织和器官中表达,以及是否在蜂王或雄蜂中表达需要进一步研究,推测意大利蜜蜂 lincRNA 的总数多于 6 353 个。

Ruiz-Orera 等 (2014) 通过对小鼠、人类、斑马鱼 *Danio rerio*、黑腹果蝇、拟南芥 *Arabidopsis thaliana* 和啤酒酵母 *Saccharomyces cerevisiae* 的 lincRNA 的特征分析,发现它们的 lincRNA 相对于蛋白编码基因都具有较短的转录本长度。Tang 等 (2017) 的研究结果表明野猪的 lincRNA 的表达量较低,且较之蛋白编码基因其转录本长度更短、外显子数更少。Wu 等 (2016) 对家蚕的研究结果表明其 lincRNA 较蛋白编码基因具有更少的外显子数、较短的转录本长度和较短的外显子和内含子长度。本研究中,意大利蜜蜂工蜂中肠的 lincRNA 与其蛋白编码基因相比,具有较少的外显子数和较短的转录本长度,这与其他物种的 lincRNA 特征相似。

本研究中,意大利蜜蜂工蜂中肠 lincRNA 的上下游基因涉及 42 个 GO 条目,其中分别有 349、332 和 283 个上下游基因分布在细胞进程、代谢进程和催化活性,表明相应的 lincRNA 参与意大利蜜蜂工蜂中肠的新陈代谢、细胞生命活动的调控。此外,还发现有 108、3 和 2 个上下游基因富集在应激反应、免疫系统进程和细胞凋亡,表明相应的 lincRNA 参与对意大利蜜蜂工蜂中肠的免疫防御的调控。此外,意大利蜜蜂工蜂中肠的上下游基因涉及 256 条 pathway,其中有 239 个上下游基因分布在 74 条代谢通路,包括碳代谢 (13 gene)、嘌呤代谢 (12 gene)、肌醇磷酸盐代谢 (11 gene)、丙氨酸、天冬氨酸和谷氨酸代谢 (9 gene) 等 69 条物质代谢通路;甲烷代谢 (5 gene)、氧化磷酸化 (5 gene)、真菌固碳

作用 (2 gene)、硫代谢 (2 gene) 和光合生物固碳作用 (1 gene) 等 5 个能量代谢通路;再次表明相应的 lincRNA 参与对意大利蜜蜂工蜂中肠的物质和能量代谢的调控。Wnt 信号通路在动物和昆虫中的胚胎发育和组织发育起到重要的作用 (Zachary and Angers, 2018)。Li 等 (2018) 的研究结果显示 Hippo 信号通路调控家蚕的身体发育和细胞凋亡;Elshaer 等 (2015) 发现 Hippo 通路可以调控蟑螂的卵巢发育;Huang 等 (2017) 研究结果表明果蝇的 Hedgehog 和 Hippo 信号通路限制 BMP 受体产生,从而引起卵巢细胞分化。本研究发现,部分上下游基因富集在 Wnt (49 gene)、Hippo (36 gene) 和 Hedgehog (3 gene) 信号通路,表明相应的 lincRNA 参与调控此三条信号通路,从而影响意大利蜜蜂工蜂中肠的生长和发育。还发现部分上下游基因富集在溶酶体 (12 gene)、泛素蛋白水解 (10 gene)、内吞作用 (9 gene) 和吞噬作用 (5 gene) 等细胞免疫通路,以及 MAPK (15 gene)、Jak-STAT (3 gene)、Toll-like 受体 (1 gene) 和 NF-kappa B (1 gene) 信号通路等体液免疫通路,暗示相应的 lincRNA 在意大利蜜蜂工蜂中肠的免疫防御中扮演着重要的调控角色。

本研究结合 lincRNA-seq 技术和生物信息学方法对意大利蜜蜂工蜂中肠的 lincRNA 进行系统的预测、结构特征和功能分析,研究结果不仅丰富了蜜蜂的 lincRNA 信息,也为探究 lincRNA 在意大利蜜蜂工蜂中肠发育及胁迫应答过程中的作用打下基础。下一步将对意大利蜜蜂工蜂中肠响应 *N. ceranae* 胁迫的 lincRNA 差异表达谱进行深入分析,进而揭示差异表达 lincRNA (DElincRNA) 在宿主胁迫应答中的作用,并通过构建和分析 DElincRNA 与其他 ncRNA、mRNA 之间的调控网络筛选出关键 lincRNA。

## 参考文献 (References)

- Begna D, Han B, Feng M, Fang M, Li J, 2012. Differential expressions of nuclear proteomes between honeybee (*Apis mellifera* L.) queen and worker larvae: a deep insight into caste pathway decisions. *Journal of Proteome Research*, 11(2):

- 1317–1329.
- Chen B, Zhang Y, Zhang X, Jia S, Chen S, Kang L, 2016. Genome-wide identification and developmental expression profiling of long noncoding RNAs during *Drosophila metamorphosis*. *Scientific Reports*, 6: 23330.
- Chen X, Ma C, Chen C, Lu Q, Shi W, Liu ZG, Wang HW, Guo HK, 2017. Integration of lncRNA–miRNA–mRNA reveals novel insights into oviposition regulation in honey bees. *Peer Journal*, 5(12. 9): e3881.
- Committee on the status of pollinators in north America, 2007. Status of Pollinators in North America. Washington, D. C.: National Academies Press. 11–14.
- Cooper DR, Carter G, Li PF, Patel R, Watson JE, Patel NA, 2014. Long non-coding RNA NEAT1 associates with SRp40 to temporally regulate PPAR $\gamma$ 2 splicing during adipogenesis in 3T3-L1 cells. *Genes*, 5(4): 1050–1063.
- Elshaer N, Piulachs MD, 2015. Crosstalk of EGFR signaling with Notch and Hippo pathways to regulate cell specification, migration and proliferation in cockroach panoistic ovaries. *Biology of the Cell*, 107(8): 273–285.
- Flirian K, Joshua TM, 2018. Functional classification and experimental dissection of long noncoding RNAs. *Cell*, 172(3): 393–407.
- Foret S, Maleszka R, 2012. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proceedings of the National Academy of Sciences of the United States of America*, 109(13): 4968–4973.
- Guo Y, Su SK, Chen SL, Zhao SW, Chen RS, 2015. The function of lncRNAs in the caste determination of the honeybee. *Progress in Biochemistry and Biophysics*, 42(8): 750–757. [郭昱, 苏松坤, 陈盛禄, 张少吾, 陈润生, 2015. LncRNA 在蜜蜂级型分化中的功能研究. *生物化学与生物物理进展*, 42(8): 750–757.]
- Hainer SJ, Martens JA, 2011. Transcription of ncDNA: many roads lead to local gene regulation. *Transcription*, 2(3): 120–123.
- Huang J, Reilein A, Kalderon D, 2017. Yorkie and Hedgehog independently restrict BMP production in escort cells to permit germline differentiation in the *Drosophila* ovary. *Development*, 144(14): 2584–2594.
- Humann FC, Tiberio GJ, Hartfelder K, 2013. Sequence and expression characteristics of long noncoding RNAs in honey bee caste development-potential novel regulators for transgressive ovary size. *PLoS ONE*, 8(10): e78915.
- Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S, Poliakov A, Cao X, Dhanasekaran SM, Robinson DR, Feng FY, Iyer HK, Chinnaiyan AM, 2015. The landscape of long noncoding RNAs in the human transcriptome. *Nature Genetics*, 47(3): 199–208.
- Jayakodi M, Jung JW, Papk D, Ahn YJ, Lee SC, Shin SY, Shin C, Yang TJ, Kwon HW, 2015. Genome-wide characterization of long intergenic non-coding RNAs (lincRNAs) provides new insight into viral diseases in honey bees *Apis cerana* and *Apis mellifera*. *BMC Genomics*, 16(1): 680.
- Kayvan E, Sultan A, Zhang G, Asgari S, 2016. Identification of *Aedes aegypti* long intergenic non-coding RNAs and their association with *wolbachia* and dengue virus infection. *PLoS Neglected Tropical Diseases*, 10(10): 1–18.
- Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao L, Wei L, Gao G, 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*, 35(Suppl. 2): 345–349.
- Kurze C, Routtu J, Moritz RFA, 2016. Parasite resistance and tolerance in honeybees at the individual and social level. *Zoology*, 119(4): 290–297.
- Li N, Tong X, Zeng J, Meng G, Sun F, Song J, Lu C, Dai F, 2018. Hippo pathway regulates somatic development and cell proliferation of silkworm. *Genomics*, doi: 10. 1016/j. ygeno. 2018. 02. 014.
- Mudge JM, Harrow J, 2015. Creating reference gene annotation for the mouse C57BL/6/J genome assembly. *Mammalian Genome*, 26(9/10): 366–378.
- Ouyang J, Zhu X, Chen Y, Wei H, Chen Q, Chi X, Qi B, Zhang L, Zhao Y, Gao GF, Wang G, Chen JL, 2014. NRAV, a long noncoding RNA, modulates antiviral responses through suppression of interferon-stimulated gene transcription. *Cell Host Microbe*, 16(5): 616–626.
- Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, Fan L, Sandelin A, Rinn JL, Regev A, Schier AF, 2012. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Research*, 22(3): 577–591.
- Qin T, Zhao H, Cui P, Albeshier N, Xiong L, 2017. A nucleus-localized long non-coding RNA enhances drought and salt stress tolerance. *Plant Physiology*, 175(3): 1321.
- Ruiz-Orera J, Messeguer X, Subirana JA, Alba MM, 2014. Long non-coding RNAs as a source of new peptides. *Elife*, 3(e03523): e03523.
- Soshnev AA, Ishimoto H, Mcallister BF, Li X, Wehling MD, Kitamoto T, Geyer PK, 2011. A conserved long noncoding RNA affects sleep behavior in *Drosophila*. *Genetics*, 189(2): 455–468.
- Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y, 2013. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Research*, 41(17): e166.
- Tang ZL, Wu Y, Yang YL, Yang YC, Wang ZS, Yuan JP, Yang Y, Hua CJ, Fan XH, Niu GL, Zhang YB, Lu Zj, Li K, 2017.

- Comprehensive analysis of long non-coding RNAs highlights their spatio-temporal expression patterns and evolutionary conservation in *Sus scrofa*. *Scientific Reports*, 7: 43166.
- The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team, 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, 420(6915): 563.
- Wu Y, Cheng T, Liu C, Liu D, Zhang Q, Long R, Zhao P, Xia Q, 2016. Systematic identification and characterization of long non-coding RNAs in the Silkworm, *Bombyx mori*. *PLoS ONE*, 11(1): e0147147.
- Xiao HM, Yuan ZT, Guo DH, Hou BF, Yin CL, Zhang WQ, Li F, 2015. Genome-wide identification of long noncoding RNA genes and their potential association with fecundity and virulence in rice brown planthopper, *Nilaparvata lugens*. *BMC Genomics*, 16(1): 1–16.
- Young RS, Marques AC, Tibbit C, Haerty W, Bassett AR, Liu JL, Ponting CP, 2012. Identification and properties of 1, 119 candidate lincRNA loci in the *Drosophila melanogaster* genome. *Genome Biology Evolution*, 4(4): 427–442.
- Zachary S, Stephane A, 2018. Wnt signaling in development and tissue homeostasis. *Development*, 145: dev146589. doi: 10.1242/dev.146589.
- Zayed A, Robinson GE, 2012. Understanding the relationship between brain gene expression and social behavior: lessons from the honey bee. *Annual Review of Genetics*, 46(6): 591–615.
- Zhou QZ, Fang SM, Zhang Q, Yu QY, Zhang Z, 2017. Identification and comparison of long non-coding RNAs in the silk gland between domestic and wild silkworms. *Insect Science*, doi.org/10.1111/1744-7917.12443.
- Zhu B, Xu M, Shi H, Gao X, Liang P, 2017. Genome-wide identification of lncRNAs associated with chlorantraniliprole resistance in diamondback moth *Plutella xylostella* (L.). *BMC Genomics*, 18(1): 380.

\*\*\*\*\*

## 封面介绍

### 毛蕊花球象 *Cionus olivieri* Rosenschöld

毛蕊花球象 *Cionus olivieri* Rosenschöld, 隶属于鞘翅目 Coleoptera 象虫科 Curculionidae, 身体球形, 前胸背板比鞘翅基部明显窄, 触角索节 5 节, 体壁黑色, 胫节和附节红色。全身覆盖着致密的黄白色鳞片, 鞘翅缝中部和端部有黑色宽鳞片组成的大黑斑点, 行间散布较小的黑斑点。幼虫体壁非常柔软, 黄白色, 常分泌粘液在体表, 取食毛蕊花属植物的子房组织。该属昆虫的寄主专一性比较强, 在北美已经作为控制入侵的毛蕊花属植物的生物防治天敌。图片于 2018 年 8 月 12 日拍摄于哈萨克斯坦 Altyn Emel 保护区, 寄主植物为毛蕊花 *Verbascum thapsus* L.

(中国科学院动物研究所 姜春燕)