

中华蜜蜂的单核苷酸多态性和插入缺失突变位点鉴定*

吴 鹰^{1**} 蔡宗兵^{1**} 许雅静¹ 郭意龙¹ 鲍佳益¹ 康育欣¹
叶亚萍¹ 钱加琚¹ 张凯遥¹ 陈大福^{1,2***} 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)

摘 要 【目的】本研究拟利用已获得的中华蜜蜂 *Apis cerana cerana* 幼虫肠道的转录组数据对单核苷酸多态性 (Single nucleotide polymorphism, SNP) 和插入缺失 (Insertion-Deletion, InDel) 突变位点进行挖掘和分析, 旨在丰富中华蜜蜂的 SNP 和 InDel 信息, 并为新型分子标记的开发提供基础。【方法】根据有效读段与东方蜜蜂 *Apis cerana* 参考基因组的比对情况, 采用 GATK 软件识别单碱基错配和碱基的插入缺失情况, 再利用 ANNOVAR 软件对 SNP 位点和 InDel 位点进行分析。通过相关生物信息学软件将 SNP 和 InDel 位点所在基因分别比对 GO 和 KEGG 数据库, 以获得相应的功能和通路注释。【结果】共鉴定到中华蜜蜂的 58 919 个 SNP 位点, 包括 24 548 个纯合位点和 34 371 个杂合位点; 发生转换和颠换的 SNP 位点分别有 49 102 和 9 817 个; 数量最多和最少的突变类型分别是 C/T 和 T/G; 分布在外显子区的 SNP 位点数量最多, 达到 22 649 个; 此外, 发生同义突变的 SNP 位点数量最多, 其次是非同义突变; SNP 位点所在基因可注释到 46 个 GO 条目和 121 条 KEGG 通路。共鉴定到 6 551 个 InDel 位点, 包括 3 270 个插入突变和 3 281 个缺失突变; 分布在内含子区 InDel 位点最多, 共计 2 793 个; 发生移码插入的 InDel 位点最多; 进一步分析结果显示 InDel 位点所在基因可注释到 27 个 GO 条目和 28 条 KEGG 通路。【结论】本研究鉴定到中华蜜蜂的大量 SNP 位点和 InDel 位点, 解析了 SNP 和 InDel 位点的突变类型、基因组功能元件分布和密码子突变类型, 并揭示 SNP 和 InDel 位点对中华蜜蜂的重要生物学过程具有潜在影响。

关键词 东方蜜蜂, 中华蜜蜂, 单核苷酸多态性, 插入缺失突变, 转录组

Identification of single nucleotide polymorphism (SNP) and insertion- deletion (InDel) mutation loci in *Apis cerana cerana*

WU Ying^{1**} CAI Zong-Bing^{1**} XU Ya-Jing¹ GUO Yi-Long¹ BAO Jia-Yi¹
KANG Yu-Xin¹ YE Ya-Ping¹ QIAN Jia-Jun¹ ZHANG Kai-Yao¹
CHEN Da-Fu^{1,2***} GUO Rui^{1,2***}

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

2. Institute of Apitherapy, Fujian Agriculture and Forestry University, Fuzhou 350002, China)

Abstract 【Objectives】To increase the available information on single nucleotide polymorphism (SNP) and insertion-deletion (InDel) mutation loci in *Apis cerana cerana* by searching for these loci in transcriptome data obtained from the *A. c. cerana* larval gut. 【Methods】Based on mapping information from the *A. cerana* reference genome, GATK software was used to identify single base mismatches and base insertion deletions between clean reads. Annovar software was then used to analyze SNP and InDel loci. In addition, genes containing SNP or InDel loci were aligned to the GO and KEGG databases to

*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建设专项 (CARS-44-KXJ7); 福建农林大学杰出青年科研人才计划 (xjq201814); 福建农林大学硕士生导师团队项目 (郭睿); 福建省病原真菌与真菌毒素重点实验室开放课题 (郭睿); 福建农林大学大学生创新创业训练计划项目 (202210389114, 202210389114)

**共同第一作者 Co-first authors, E-mail: wy569703@163.com; caizongbing@126.com

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

收稿日期 Received: 2021-07-07; 接受日期 Accepted: 2021-11-10

deduce their likely function and pathway annotation. **[Results]** A total of 58 919 SNPs were identified, including 24 548 homozygous and 34 371 heterozygous, sites. Among these, 4 9102 had undergone conversion and 9 817 had undergone transversion. C/T and T/G were the most and least common mutations, and most (22 649) were found in the exon region. Synonymous mutations were more abundant than non-synonymous mutations. Genes containing SNP loci were annotated to 46 GO terms and 121 KEGG pathways. A total of 6 551 InDel loci were identified, including 3 270 insertion mutations and 3 281 deletion mutations. Most (2 793) InDel loci were found in the intronic region. Frameshift insertions were the most common mutation in InDel loci. InDel loci could be annotated to 27 GO terms and 28 KEGG pathways. **[Conclusion]** Over 58 000 SNP and 6 500 InDel loci were identified in *A. c. cerana* and the most common kinds of mutations in these genes, their distribution in various functional elements of the genome, and the codon mutation types, were analyzed. Our findings indicate the likely functions of SNP and InDel loci in *A. c. cerana*.

Key words *Apis cerana*, *Apis cerana cerana*; single nucleotide polymorphism; insertion-deletion mutation; transcriptome

中华蜜蜂 *Apis cerana cerana* 是东方蜜蜂 *Apis cerana* 的指名亚种, 也是我国特有的优良本土蜂种, 经长期适应性进化已高度适应我国的地理和生态环境 (Li *et al.*, 2007)。中华蜜蜂不仅是自然界重要的授粉昆虫, 也是我国养蜂生产使用的主要蜂种之一, 具有特殊的生态和经济价值 (Lin *et al.*, 2016)。

单核苷酸多态性 (Single nucleotide polymorphism, SNP) 是由等位基因位点的单碱基突变而引起的 DNA 序列多态性。SNP 作为一类最主要的可遗传变异是驱动生物进化的关键动力 (赵辉等, 2006)。插入缺失突变 (Insertion-Deletion, InDel) 是另一种新型分子标记, 指基因组中小片段序列的插入或缺失 (Wajnberg and Passetti, 2016), 属于长度多态性标记。SNP 和 InDel 分子标记在基因组中具有数量多、密度大、分布广、重复性好、结果准确可靠、开发成本低及共显性遗传等优点, 因而在遗传育种、遗传图谱绘制、物种亲缘关系鉴定、系统进化分析以及基因定位等方面具有广泛的应用 (Zhang *et al.*, 2019; 蔡宗兵等, 2022)。

近十多年来, 以 Illumina HiSeq 为代表的二代测序技术在动植物和微生物的转录组研究领域广泛应用, 取得了一系列重要的研究进展 (Dillies *et al.*, 2013; Schirmer *et al.*, 2015; Pichler *et al.*, 2018)。RNA-seq 具有通量高、周期短、灵敏度强和成本低等优势, 已成为大规模挖掘 SNP 和 InDel 位点的有效工具 (Briese *et al.*, 2016; Ungaro *et al.*, 2017)。目前, 基于二代测序数据鉴定 SNP 和 InDel 位点已在玉米 *Zea mays*

L. (刘小红, 2021)、太平洋长牡蛎 *Crassostrea gigas* (Juárez *et al.*, 2021) 和东方蜜蜂微孢子虫 *Nosema ceranae* (张文德等, 2022) 等物种中已见诸报道。虽然前人 (李志勇等, 2009; 吴贵辉, 2019; 李延璨, 2020) 对中华蜜蜂的 SNP 进行了一些研究, 但总体仍较为有限。Park 等 (2014) 在中华蜜蜂的蜂毒素基因中鉴定到自然发生的 9 个新 SNP 位点, 为昆虫源抗菌肽 (AMPs) 的治疗及应用提供了重要线索。Shi 等 (2013) 在构建中华蜜蜂的连锁图谱时鉴定出 126 990 个 SNP 位点, 通过基因分型和质量筛选最终利用 1 535 个具有丰富信息的 SNP 构建了中华蜜蜂的连锁图谱。相比于果蝇 *Drosophila* (Housden and Perrimon, 2016; Kane *et al.*, 2017) 和家蚕 *Bombyx mori* (Xia *et al.*, 2009; Zhang *et al.*, 2015) 等模式生物, 蜜蜂的 InDel 研究较为匮乏。中华蜜蜂的 InDel 研究迄今未见报道。

本研究利用已获得的转录组数据对中华蜜蜂的 SNP 位点和 InDel 位点进行全基因组鉴定, 并分析其突变类型、基因组功能元件分布及密码子突变类型, 进而对 SNP 和 InDel 位点所在基因进行功能和通路注释, 以期丰富中华蜜蜂的 SNP 和 InDel 位点信息, 并为中华蜜蜂的基因图位克隆、基因型分型、遗传多样性研究及遗传图谱绘制等提供候选分子标记。

1 材料与方法

1.1 供试生物材料

中华蜜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。

1.2 转录组数据来源

前期研究中,笔者所在课题组利用基于链特异性 cDNA 建库的 RNA-seq 技术对中华蜜蜂 6 日龄幼虫肠道样品进行测序,测得 213 479 408 条原始读段,经质控得到 213 458 846 有效读段(clean reads),Q30 达到 96.65%;此外,有 134 564 975 条 clean reads 可比对到东方蜜蜂参考基因组(Assembly ACSNU-2.0, <https://www.ncbi.nlm.nih.gov/genome/?term=apis+cerana>)。测序原始数据已上传 NCBI SRA 数据库, BioProject 号: PRJNA560730。高质量的转录组数据可为本研究中华蜜蜂 SNP 和 InDel 位点的鉴定和分析提供可靠的数据基础。

1.3 SNP 位点和 InDel 位点的鉴定及分析

根据 clean reads 与东方蜜蜂参考基因组的比对情况,使用 GATK 软件(Van der Auwera *et al.*, 2013)识别错配碱基,并采用 ANNOVAR 软件(Wang *et al.*, 2010)分析 SNP 位点和 InDel 位点,步骤为:(1)使用 GATK 软件识别 SNP 位点的同时将低质量的 SNP 过滤;(2)对于 InDel 位点附近的 SNP 位点,进一步通过 GATK 软件进行校正;

(3)筛选基因不重叠的 UTR 和 Exon 区域的位点;(4)筛选参考 reads ≥ 2 且突变 reads ≥ 3 的位点;(5)筛选突变频率在 0.1 到 0.9 之间的位点。

1.4 SNP 位点和 InDel 位点所在基因的功能及通路注释

采用 OmicShare 平台(<https://www.omicshare.com/tools/home/index/index>)的相关生物信息学工具将中华蜜蜂的 SNP 位点所在基因和 InDel 位点所在基因分别比对 GO (Gene Ontology) 数据库(<https://www.omicshare.com/tools/Home/Soft/gogsea>)和 KEGG (Kyoto Encyclopedia of Genes and Genomes) 数据库(<https://www.omicshare.com/tools/Home/Soft/pathwaygsea>),从而获得相应的 GO 条目(term)和 KEGG 通路(Pathway)注释。

2 结果与分析

2.1 中华蜜蜂 SNP 位点的鉴定及分析

共鉴定到中华蜜蜂的 58 919 个 SNP 位点,占基因突变总数的 89.99%。上述 SNP 位点包括 24 548 个纯合位点和 34 371 个杂合位点。部分 SNP 位点的详细信息如表 1 所示。

表 1 中华蜜蜂部分 SNP 位点的详细信息
Table 1 Detailed information about partial SNP loci in *Apis cerana cerana*

染色体 ID Chromosome ID	起始位点 Start site	终止位点 End site	参考碱基 Reference base	突变碱基 Mutation base	位置类型 Structure type	位置注释 Structure gene
NW_016019397.1	829083	829083	T	C	Exonic	ncbi_107999943
NW_016018678.1	48144	48144	A	G	Intronic	ncbi_107995327
NW_016019863.1	184291	184291	T	C	Intergenic	ncbi_108003287; ncbi_108003289
NW_016019220.1	927934	927934	A	G	UTR3	ncbi_107998305
NW_016019108.1	2499150	2499150	C	T	UTR5	ncbi_107996881
NW_016018767.1	135083	135083	G	A	Downstream	ncbi_107995467
NW_016018122.1	2237522	2237522	A	G	Upstream	ncbi_107993530
NW_016017690.1	274612	274612	G	A	Upstream; Downstream	ncbi_108004126; ncbi_108004119
NW_016019186.1	154578	154578	T	C	Splicing	ncbi_107997717
NW_016019441.1	2033762	2033762	T	G	UTR5; UTR3	ncbi_108000475; ncbi_108000474
NW_016018011.1	2973041	2973041	C	G	Exonic; Splicing	ncbi_107993146

突变类型为转换和颠换的 SNP 位点分别有 49 102 和 9 817 个, 占总数的 83.34% 和 16.66% (图 1: A)。SNP 位点的突变类型可分为 C/T、G/A、A/G、T/C、G/T、A/T、C/A、T/A、T/G、A/C、C/G 和 G/C, 其中数量最多和最少的突变类型分别是 C/T 和 T/G (图 1: A)。此外, 分布在外显子区的 SNP 位点数量最多, 共计 22 649 个 (图 1:

B); 其次是内含子区、基因间隔区、3' UTR、5' UTR、下游区、上游区、上下游重叠区、可变剪切位点、3' UTR 与 5' UTR 重叠区和外显子与剪切位点重叠区 (图 1: B)。进一步对 SNP 位点涉及的密码子突变类型进行分析, 结果显示发生同义突变的 SNP 位点数量最多, 其次是非同义突变; 发生终止子增加和终止子减少的 SNP 数量最少 (图 1: C)。

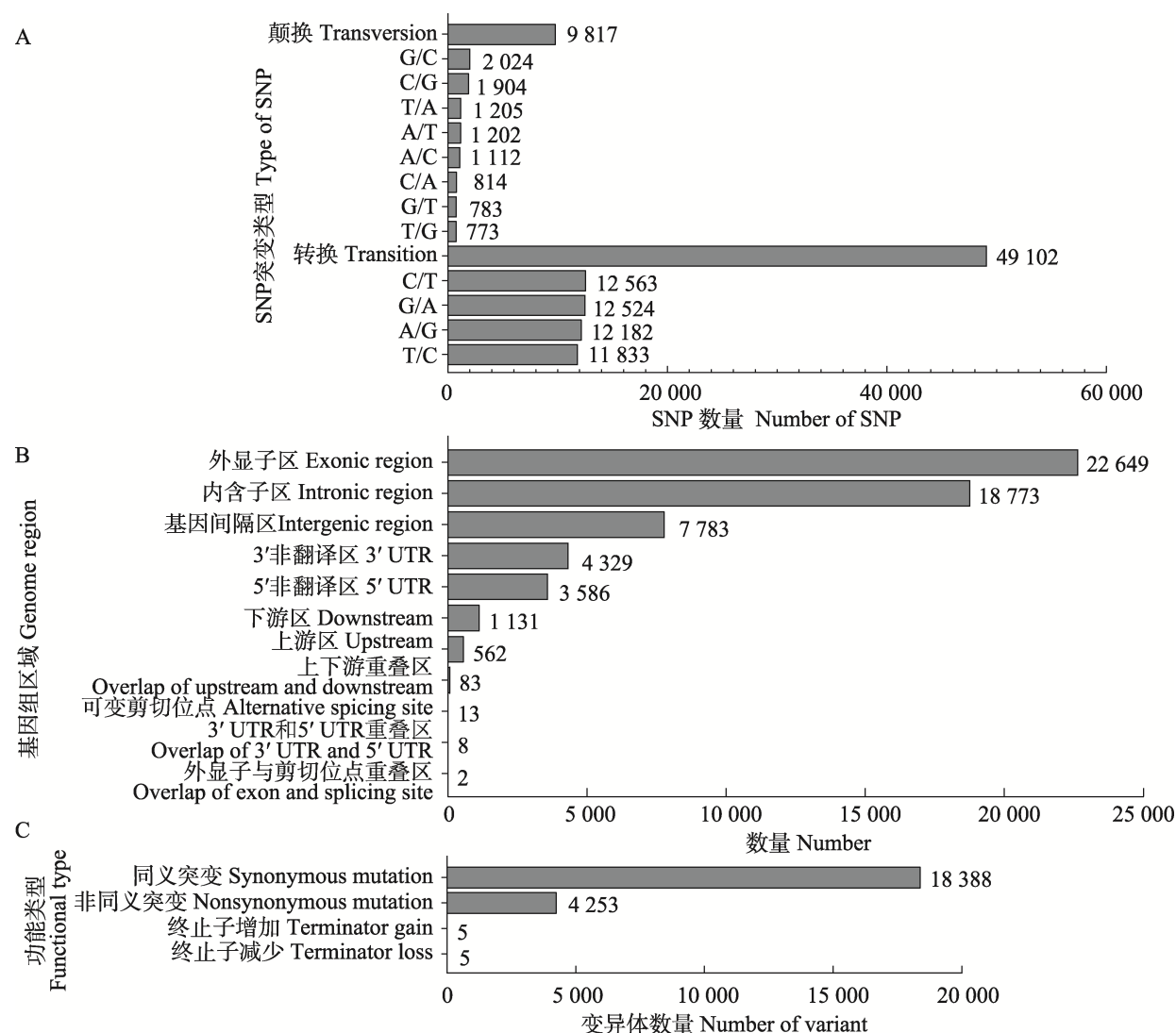


图 1 中华蜜蜂 SNP 位点的突变类型 (A)、基因组区域分布 (B) 及密码子突变类型 (C)
Fig. 1 Mutation type (A), distribution in genomic regions (B) and codon mutation type (C) of SNP loci in *Apis cerana cerana*

GO 数据库注释结果显示, 中华蜜蜂 SNP 位点所在基因可注释到代谢进程和细胞进程等 20 个生物学进程相关条目, 细胞组件和细胞等 17 个细胞组分相关条目, 结合和催化活性等 9 个分子功能相关条目 (图 2: A)。KEGG 数据库注释

结果显示, 上述 SNP 位点所在基因还可以注释到细胞进程、环境信息处理、遗传信息处理、代谢及有机系统相关的 121 条通路, 包括代谢途径、次生代谢物的生物合成、抗生素的生物合成、胞吞作用及内质网中的蛋白质加工等 (图 2: B)。

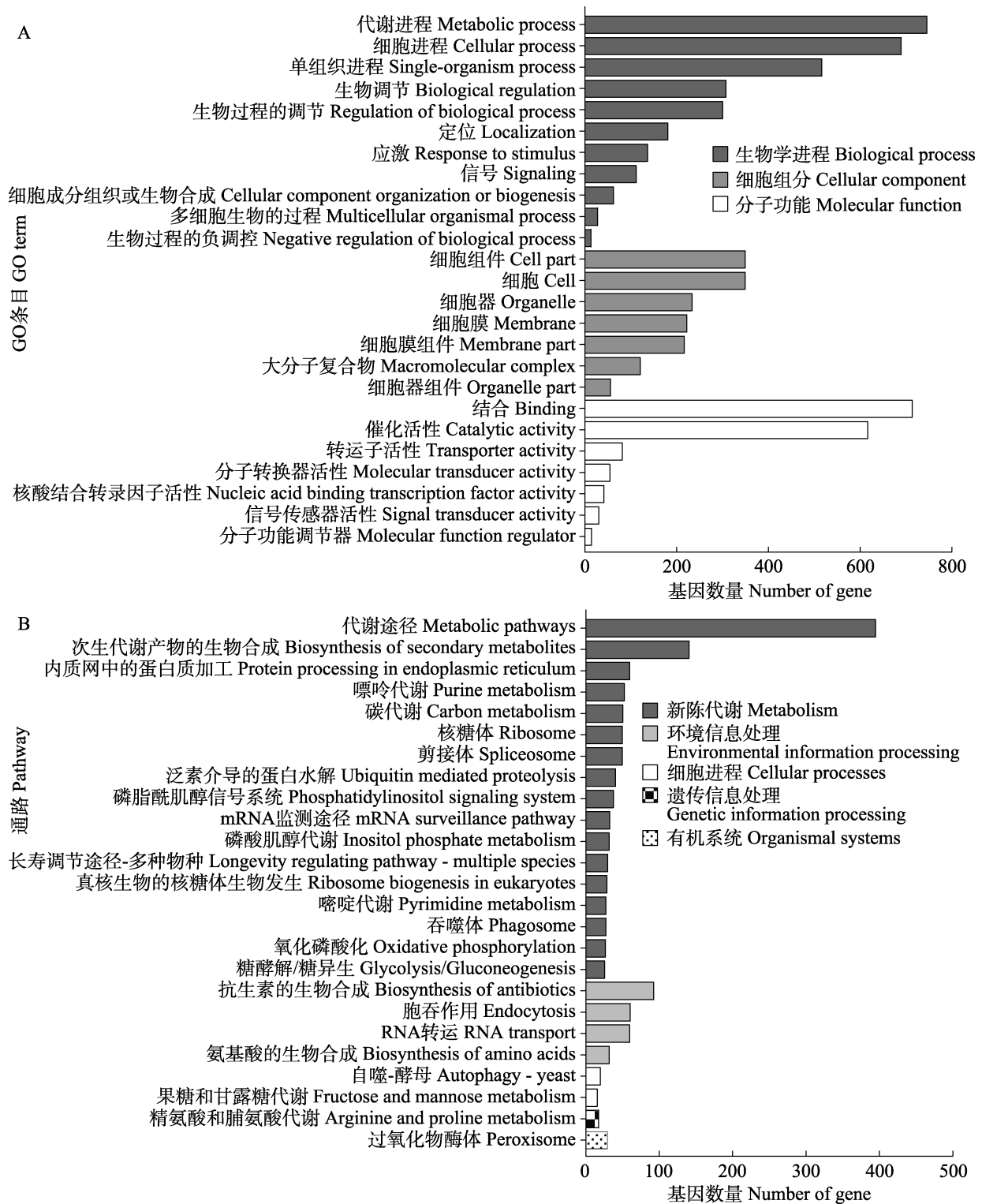


图2 中华蜜蜂 SNP 位点所在基因的 GO (A) 和 KEGG (B) 数据库注释

Fig. 2 GO (A) and KEGG (B) database annotation of genes containing *Apis cerana cerana* SNP loci

2.2 中华蜜蜂 InDel 位点的鉴定及分析

共鉴定到 6 551 个 InDel 位点, 占基因突变

总数的 10.01%, 包括 3 270 个插入突变和 3 281 个缺失突变。部分 InDel 位点的详细信息如表 2 所示。

表 2 中华蜜蜂部分 InDel 位点的详细信息
Table 2 Detailed information of partial InDel loci in *Apis cerana cerana*

染色体 ID Chromosome ID	起始位点 Start site	终止位点 End site	参考碱基 Reference base	突变碱基 Mutation base	位置类型 Structure type	位置注释 Structure gene
NW_016019397.1	831308	831308	-	T	Intronic	ncbi_107999943
NW_016019220.1	39433	39433	-	T	Intergenic	ncbi_107998294; ncbi_107998319
NW_016019430.1	791436	791436	A	-	UTR3	ncbi_108000173
NW_016017678.1	486046	486046	-	G	UTR5	ncbi_108003995
NW_016017456.1	2652728	2652728	T	-	Downstream	ncbi_107995143; ncbi_107996097
NW_016017690.1	264048	264049	TC	-	Exonic	ncbi_108004123
NW_016019552.1	378268	378268	A	-	Upstream	ncbi_108001271
NW_016017456.1	3468590	3468590	-	T	Upstream; Downstream	ncbi_107994792; ncbi_107994809; ncbi_107994838
NW_016019441.1	418456	418456	-	A	Splicing	ncbi_108000335

InDel 位点分布最多的基因组功能元件为内含子区, 其次是基因间区、3' UTR、5' UTR、下游区、外显子区、上游区、上下游重叠区和剪接区 (图 3: A)。进一步对 InDel 位点涉及的密码子突变类型进行分析, 结果显示数量最多的密码子突变类型为移码插入, 其次是移码缺失、非移码缺失和非移码插入; 数量最少的是终止子增加和同义核苷酸突变 (图 3: B)。

数据库注释结果显示, 上述 InDel 位点所在

基因可注释到代谢进程和细胞进程等 14 个生物学进程相关条目, 细胞组件和细胞等 7 个细胞组分相关条目, 结合和催化活性等 6 个分子功能相关条目 (图 4: A); 以及细胞进程、环境信息处理、遗传信息处理、代谢和有机系统相关的 28 条通路, 包括代谢途径、剪接体、胞吞作用、内质网中的蛋白质加工、各种类型的 N-聚糖生物合成、N-聚糖生物合成、氨基糖和核苷酸糖代谢及氨酰基-tRNA 的生物合成等 (图 4: B)。

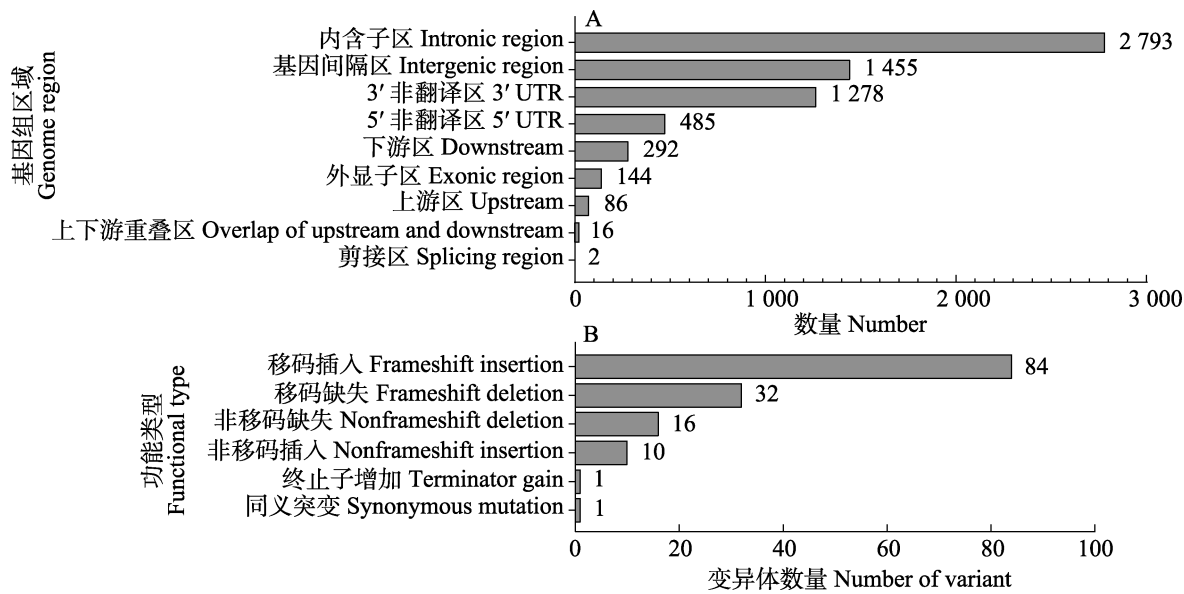


图 3 在中华蜜蜂 InDel 位点的基因组区域分布 (A) 和密码子突变类型 (B)
Fig. 3 Distribution in genomic regions (A) and condonmutation type (B) of InDel loci in *Apis cerana cerana*

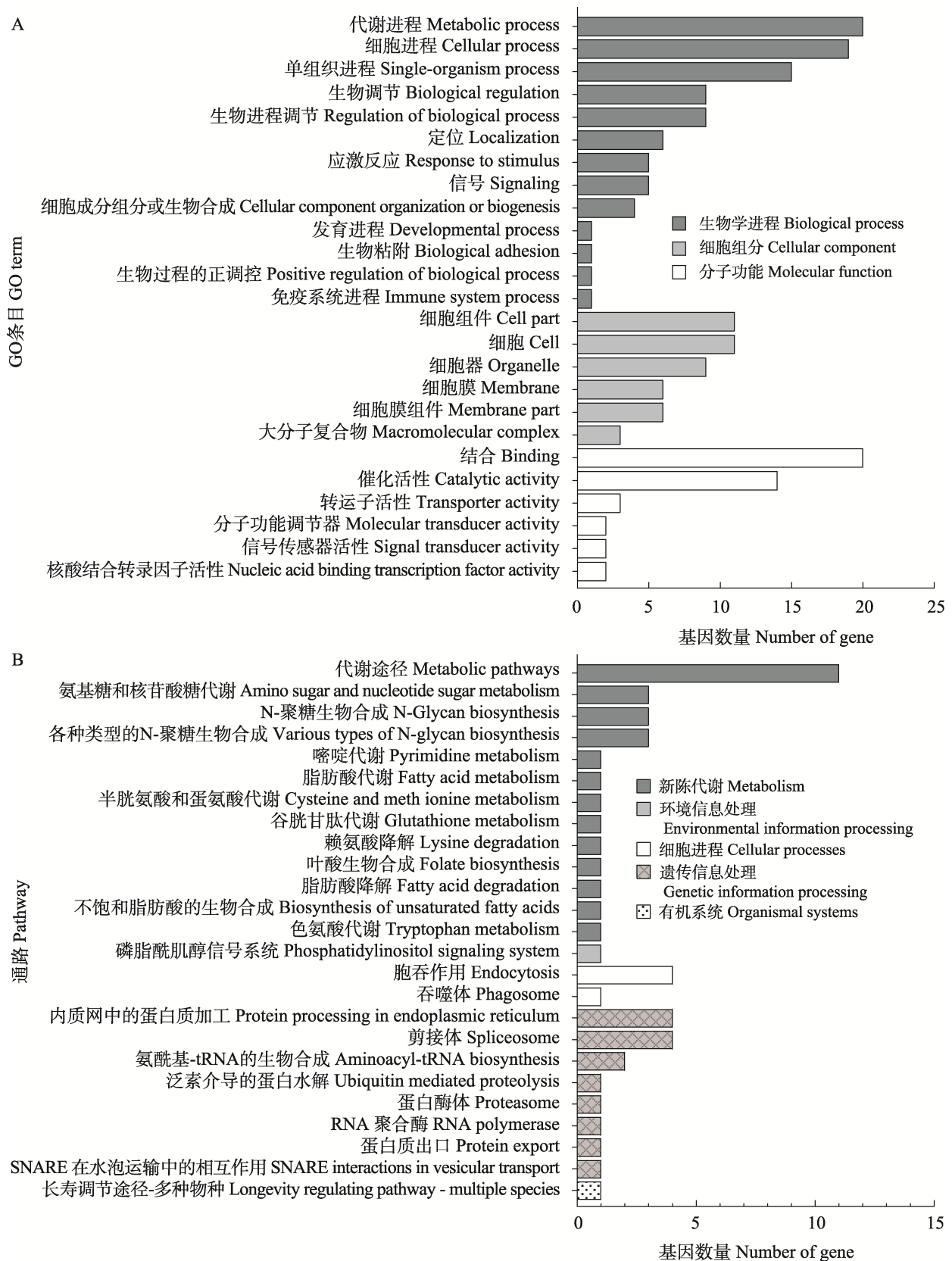


图 4 中华蜜蜂 InDel 位点所在基因的 GO (A) 和 KEGG (B) 数据库注释

Fig. 4 Annotation of *Apis cerana cerana* InDel loci in GO (A) and KEGG (B) databases

3 讨论

此前, 中华蜜蜂的分子标记研究主要集中在 SSR 方面 (Liu *et al.*, 2016, 2017; 徐细建等, 2017), SNP 和 InDel 的相关研究较为缺乏。本研究利用已获得的转录组数据共鉴定到中华蜜蜂的 58 919 个 SNP 位点和 6 551 个 InDel 位点。由于上述转录组数据来源于中华蜜蜂 6 日龄幼虫肠道组织, 推测中华蜜蜂基因组实际包含的 SNP 位点和 InDel 位点应多于本研究鉴定到的数量。随着二代测序成本的大幅下降, 存储于公共数据库 (如 NCBI SRA 数据库) 的各物种的转录组数据快速增加。未来可利用存储于公共数据库的中华蜜蜂不同虫态、同一虫态不同日龄、同一虫态不同器官和组织的转录组数据进一步鉴定中华蜜蜂的 SNP 位点和 InDel 位点。

本研究中, 鉴定到的 SNP 位点中杂合位点和纯合位点的占比分别为 58.34% 和 41.66%, 涉及 12 种突变类型, 其中最丰富的类型为 C/T 型, 这与猪 (Pawlina-Tyszkó *et al.*, 2021)、蚕豆 *Vicia faba* L. (Ocaña *et al.*, 2015)、玉米 (刘小红, 2021) 和椰心叶甲啮小蜂 *Tetrastichus brontispae* (刘华伟等, 2021) 等物种的研究结果一致; 发生转换和颠换的 SNP 分别为 49 102 和 9 817 个, 其中转换类型占比高达 83.34%, 这与前人在大骨鸡 (张洁慧等, 2021)、向日葵锈菌 *Puccinia helianthi* Schw (王妍等, 2018)、伊比利亚蜜蜂 *Apis mellifera iberiensis* (Henriques *et al.*, 2018) 和椰心叶甲啮小蜂 (刘华伟等, 2021) 等物种中的研究报道相似。另外, 中华蜜蜂的 SNP 位点分布最丰富的基因组功能元件为外显子区 (22 649 个, 38.44%)。外显子区的碱基序列以密码子的形式翻译为对应的氨基酸, 进而形成具有复杂空间结构的活性蛋白质, 故外显子区的碱基变化易造成编码氨基酸及蛋白的改变, 进而影响生物体的性状 (蔡宗兵等, 2022)。因此, 推测外显子区的单碱基突变可为中华蜜蜂的进化提供驱动力, 从而更加适应环境。进一步对上述 SNP 位点涉及的密码子突变类型进行分析, 结果显示同义突变 (18 388 个) 的占比高达 81.18%, 这

与蜜蜂球囊菌的研究结果一致 (蔡宗兵等, 2022)。鉴于同义突变不会导致对应氨基酸序列的变化, 因而也不会引起生物体性状的改变 (董昂, 2018), 这或许与基因的容错性有关, 中华蜜蜂在其进化过程中需要保证蛋白的稳定性。

本研究中, 共鉴定到中华蜜蜂的 6 551 个 InDel 位点, 包括 3 270 个插入突变和 3 281 个缺失突变。上述 InDel 位点与 SNP 位点的基因组功能元件分布特点有所不同, 最丰富的分布区域为内含子区 (2 793, 42.63%), 分布数量最少的区域为剪接区, 仅为 2 个 (0.03%), 这与前人在鸡 *Gallus gallus* (Brandström and Ellegren, 2007)、玉米 (Bhattaramakki *et al.*, 2002) 和毛竹 *Phyllostachys edulis* (牟少华等, 2021) 等物种中的研究报道类似。究其原因, 不同于 SNP 的单碱基突变, InDel 意味着大片段的插入或缺失, 如发生在外显子区极易造成移码突变, 进而会导致转录出错或提前终止, 也可导致翻译出的氨基酸和蛋白序列及功能发生改变。因此, 在动植物的进化过程中, InDel 突变较少发生于外显子区, 这是为了保证基因及编码蛋白功能的稳定性, 既是物种适应环境的表现, 也是长期自然选择的结果 (刘小红, 2021)。本研究的结果为上述结论提供了又一例证。

进一步对中华蜜蜂的 SNP 位点所在基因进行数据库注释, 结果显示这些基因涉及生物学进程、细胞组分和分子功能相关的 46 个功能条目, 包括代谢进程和细胞进程等 20 个生物学进程相关条目, 细胞组件和细胞等 17 个细胞组分相关条目, 结合和催化活性等 9 个分子功能相关条目 (图 2: A); 此外还涉及代谢途径、次生代谢物的生物合成及内质网中的蛋白质加工等 121 条通路 (图 2: B)。类似地, 中华蜜蜂的 InDel 位点所在基因可注释到生物学进程、细胞组分和分子功能相关的 27 个功能条目 (图 4: A), 以及代谢途径、剪接体和胞吞作用等 28 条通路 (图 4: B)。这一结果表明, SNP 和 InDel 突变与中华蜜蜂的诸多生物学过程具有潜在关联, 值得进一步深入研究。

综上所述, 本研究鉴定到中华蜜蜂的大量 SNP 位点和 InDel 位点, 并解析了 SNP 和 InDel

位点突变类型、基因组功能元件分布和密码子突变类型,为揭示 SNP 和 InDel 位点对中华蜜蜂的重要生物学过程具有重要意义。

参考文献 (References)

- Bhatramakki D, Dolan M, Hanafey M, Wineland R, Vaske D, Register JC, 3rd Tingey SV, Rafalski A, 2002. Insertion-deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers. *Plant Mol. Biol.*, 48(5/6): 539–547.
- Brandström M, Ellegren H, 2007. The genomic landscape of short insertion and deletion polymorphisms in the chicken (*Gallus gallus*) genome: A high frequency of deletions in tandem duplicates. *Genetics*, 176(3): 1691–1701.
- Briese M, Saal L, Appenzeller S, Moradi M, Baluapuri A, Sendtner M, 2016. Whole transcriptome profiling reveals the RNA content of motor axons. *Nucleic Acids Res.*, 44(4): e33.
- Cai ZB, Zhang WD, Long Q, Yu KJ, Sun MH, Wu Y, Xu YJ, Liu JM, Guo YL, Xu XJ, Chen D, Guo Rui, 2022. Exploration and analysis of SNP and InDel sites in *Ascosphaera apis* based on PacBio sequencing data. *Acta Entomol. Sin.*, 65(2): 187–196. [蔡宗兵, 张文德, 隆琦, 余尚骏, 孙明会, 吴鹰, 许雅静, 刘佳美, 郭意龙, 徐细建, 陈大福, 郭睿, 2022. 基于 PacBio 测序数据的蜜蜂球囊菌 SNP 与 InDel 位点发掘及分析. 昆虫学报, 65(2): 187–196.]
- Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J, Guernec G, Jagla B, Jouneau L, Laloë D, Le Gall C, Schaëffer B, Le Crom S, Guedj M, Jaffrézic F, French SC, 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief. Bioinform.*, 14(6): 671–683.
- Dong A, 2018. Construction of genetic linkage maps and analysis of DNA methylation-related gene expression in hickory. Master dissertation. Hangzhou: Zhejiang A&F University. [董昂, 2018. 山核桃遗传图谱构建及甲基化相关基因表达分析. 硕士学位论文. 杭州: 浙江农林大学.]
- Henriques D, Parejo M, Vignal A, Wragg D, Wallberg A, Webster MT, Pinto MA, 2018. Developing reduced SNP assays from whole-genome sequence data to estimate introgression in an organism with complex genetic patterns, the Iberian honeybee (*Apis mellifera iberiensis*). *Evol. Appl.*, 11(8): 1270–1282.
- Housden BE, Perrimon N, 2016. Detection of Indel mutations in *Drosophila* by high-resolution melt analysis (HRMA). *Cold Spring Harb. Protoc.*, doi: 10.1101/pdb.prot090795.
- Juárez OE, Escobedo FC, Arredondo ER, Ibarra AM, 2021. Development of SNP markers for identification of thermo-resistant families of the Pacific oyster *Crassostrea gigas* based on RNA-seq. *Aquaculture*, doi: 10.1016/j.aquaculture.2021.736618.
- Kane NS, Vora M, Varre KJ, Padgett RW, 2017. Efficient screening of CRISPR/Cas9-induced events in *Drosophila* using a Co-CRISPR strategy. *G3 (Bethesda)*, 7(1): 87–93.
- Li HL, Lou BG, Cheng JA, Gao QK, 2007. The chemosensory protein of Chinese honeybee, *Apis cerana cerana*: Molecular cloning of cDNA, immunocytochemical localization and expression. *Chinese Science Bulletin*, 2007(10): 1355–1364.
- Li YC, 2020. Morphological and population genomics features underlying the germplasm resource characteristics of *Apis cerana cerana* in east China. Master dissertation. Taian: Shandong Agricultural University. [李延臻, 2020. 形态与群体基因组测定研究华东地区中华蜜蜂种质资源特性. 硕士学位论文. 泰安: 山东农业大学.]
- Li ZY, Xue YB, Zhao HY, 2009. Amplification and sequence analysis of mtDNA fragment from non-coding region to COII gene region of Chinese bee (*Apis cerana cerana*) in Changbai Mountain Areas. *Hubei Agric. Sci.*, 48(5): 1051–1054. [李志勇, 薛运波, 赵惠燕, 2009. 长白山区中华蜜蜂 mtDNA 非编码区至 COII 基因 PCR 及序列分析. 湖北农业科学, 48(5): 1051–1054.]
- Lin Z, Page P, Li L, Qin Y, Zhang Y, Hu F, Neumann P, Zheng H, Dietemann V, 2016. Go east for better honey bee health: *Apis cerana* is faster at hygienic behavior than *A. mellifera*. *PLoS ONE*, 11(9): e0162647.
- Liu F, Shi T, Huang S, Yu L, Bi S, 2016. Genetic structure of Mount Huang honey bee (*Apis cerana*) populations: Evidence from microsatellite polymorphism. *Hereditas*, 153: 8.
- Liu HW, Li CX, Li F, Lu CJ, Wu SY, Qin WQ, 2021. SSR, SNP and InDel analysis based on *Tetrastichus brontispae* transcriptome. *Chinese Journal of Tropical Crops*, 42(10): 2828–2833. [刘华伟, 李朝绪, 李芬, 吕朝军, 吴少英, 覃伟权, 2021. 基于转录组测序的椰心叶甲啮小蜂 SSR、SNP 和 InDel 位点分析. 热带作物学报, 42(10): 2828–2833.]
- Liu L, Qin M, Yang L, Song Z, Luo L, Bao H, Ma Z, Zhou Z, Xu J, 2017. A genome-wide analysis of simple sequence repeats in *Apis cerana* and its development as polymorphism markers. *Gene*, 599: 53–59.
- Liu XH, 2021. Analysis on SNP and InDel markers based on RNA-seq technology in Maize. *Molecular Plant Breeding*, 19(4): 1182–1189. [刘小红, 2021. 基于 RNA-seq 技术的玉米 SNP 和 InDel 标记分析. 分子植物育种, 19(4): 1182–1189.]
- Mu SH, Li J, Li XP, Gao J, 2021. Genomic sequence analysis of four culm variants of Moso bamboo (*Phyllostachys edulis*) on culm. *Guihaia*, doi:10.11931/guihaia.gxzw202005047. [牟少华, 李娟, 李雪平, 高健, 2021. 四个竹秆变异毛竹变型的全基因组序列分析. 广西植物, doi:10.11931/guihaia.gxzw202005047.]
- Ocaña S, Seoane P, Bautista R, Palomino C, Claros GM, Torres AM, Madrid E, 2015. Large-scale transcriptome analysis in faba bean (*Vicia faba* L.) under ascochyta fabae infection. *PLoS ONE*, 10(8): e0135143.
- Park D, Jung JW, Lee MO, Lee SY, Kim B, Jin HJ, Kim J, Ahn YJ,

- Lee KW, Song YS, Hong S, Womack JE, Kwon HW, 2014. Functional characterization of naturally occurring melittin peptide isoforms in two honey bee species, *Apis mellifera* and *Apis cerana*. *Peptides*, 53: 185–193.
- Pawlina-Tyszkó K, Semik-Gurgul E, Gurgul A, Oczkiewicz M, Szmatoła T, Bugno-Poniewierska M, 2021. Application of the targeted sequencing approach reveals the single nucleotide polymorphism (SNP) repertoire in microRNA genes in the *pig* genome. *Sci. Rep.*, 11(1): 9848.
- Pichler M, Coskun ÖK, Ortega-Arbulú AS, Conci N, Wörheide G, Vargas S, Orsi WD, 2018. A 16S rRNA gene sequencing and analysis protocol for the Illumina MiniSeq platform. *Microbiology Open*, 7(6): e00611.
- Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C, 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.*, 43(6): e37.
- Shi YY, Sun LX, Huang ZY, Wu XB, Zhu YQ, Zheng HJ, Zeng ZJ, 2013. A SNP based high-density linkage map of *Apis cerana* reveals a high recombination rate similar to *Apis mellifera*. *PLoS ONE*, 8(10): e76459.
- Ungaro A, Pech N, Martin JF, McCairns R, Mévy JP, Chappaz R, Gilles A, 2017. Challenges and advances for transcriptome assembly in non-model species. *PLoS ONE*, 12(9): e0185020.
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA, 2013. From FastQ data to high confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics*, 43(1110): 11.10.1–11.10.33.
- Wajnberg G, Passetti F, 2016. Using high-throughput sequencing transcriptome data for INDEL detection: Challenges for cancer drug discovery. *Expert Opin. Drug Discovery*, 11(3): 257–268.
- Wang K, Li M, Hakonarson H, 2010. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, 38(16): e164.
- Wang Y, Jing L, LI XC, 2018. SNP mining and gene functional annotation in *Puccinia helianthi* transcriptome group. *Acta Agric. Boreali-Sin.*, 33(3): 92–98. [王妍, 景岚, 李鑫淳, 2018. 向日葵锈菌转录组 SNP 位点挖掘及所在基因功能注释. 华北农学报, 33(3): 92–98.]
- Wu GH, 2019. The effects of reproductive performance and ovarian transcriptome analysis of Queen bees of *Apis cerana* after fed by the worker bees of *Apis mellifera*. Master dissertation. Chengdu: Sichuan Agricultural University. [吴贵辉, 2019. 蜜蜂哺育对中华蜜蜂蜂王繁殖性能的影响及其卵巢转录组分析. 硕士学位论文. 成都: 四川农业大学.]
- Xia QY, Guo YR, Zhang Z, Li D, Xuan ZL, Li Z, Dai FY, Li YR, Cheng DJ, Li RQ, Cheng TC, Jiang T, Becquet C, Xu X, Liu C, Zha XF, Fan W, Lin Y, Shen YH, Jiang L, Jensen J, Hellmann I, Tang S, Zhao P, Xu HF, Yu C, Zhang GJ, Li J, Cao JJ, Liu SP, He NJ, Zhou Y, Liu H, Zhao J, Ye C, Du ZH, Pan GQ, Zhao AC, Shao HJ, Zeng W, Wu P, Li CF, Pan MH, Li JJ, Yin XY, Li DW, Wang J, Zheng HS, Wang W, Zhang XQ, Li SG, Yang HM, Lu C, Nielsen R, Zhou ZY, Wang J, Xiang ZH, Wang J, 2009. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). *Science*, 326(5951): 433–436.
- Xu XJ, Guo R, Luo Q, Xiong CL, Liang Q, Zhang CL, Zheng YZ, Zhang ZN, Huang ZJ, Zhang L, Li WD, Chen DF, 2017. *De novo* transcriptome assembly for *Apis cerana cerana* larval gut and identification of SSR molecular markers. *Sci. Agric. Sin.*, 50(6): 1157–1167. [徐细建, 郭睿, 骆群, 熊翠玲, 梁勤, 张串联, 郑燕珍, 张翌楠, 黄积健, 张璐, 李汶东, 陈大福, 2017. 中华蜜蜂幼虫肠道参考转录组的 *de novo* 组装及 SSR 分子标记鉴定. 中国农业科学, 50(6): 1157–1167.]
- Zhang JH, Dai TS, Zou S, Wang CQ, Ma W, 2021. SNP/InDel analysis of pituitary and hypothalamus of Daggu chicken based on RNA-seq technology. *Biotic. Resources*, 43(2): 160–165. [张洁慧, 戴天姝, 邹爽, 王春强, 马巍, 2021. 基于 RNA-seq 技术对大骨鸡垂体及下丘脑组织的 SNP/InDel 分析. 生物资源, 43(2): 160–165.]
- Zhang K, Kuraparthi V, Fang H, Zhu L, Sood S, Jones DC, 2019. High-density linkage map construction and QTL analyses for fiber quality, yield and morphological traits using CottonSNP63K array in upland cotton (*Gossypium hirsutum* L.). *BMC Genomics*, 20(1): 889.
- Zhang WD, Cai ZB, Long Q, Wu Y, Sun MH, Kang YX, Hu Y, Zhao XM, Chen DF, Guo R, 2022. Identification and analysis of SNP and InDel sites in *Nosema ceranae*. *Chinese Journal of Applied Entomology*, 59(3): 618–626. [张文德, 蔡宗兵, 隆琦, 吴鹰, 孙明会, 康育欣, 胡颖, 赵晓明, 陈大福, 郭睿, 2022. 东方蜜蜂微孢子虫的 SNP 与 InDel 位点鉴定及分析. 应用昆虫学报, 59(3): 618–626.]
- Zhang X, Nie M, Zhao Q, Wu Y, Wang G, Xia Q, 2015. Genome-wide patterns of genetic variation among silkworms. *Mol. Genet. Genomics*, 290(4): 1575–1587.
- Zhao H, Li QZ, Li J, Zeng CQ, Hu SN, Yu J, 2006. Research on the transition or transversion of adjacent base components and SNP produced in plant genome. *Science in China Series C: Life Sciences*, 36(1): 1–8. [赵辉, 李启寨, 李俊, 曾长青, 胡松年, 于军, 2006. 相邻碱基组分与产生 SNP 的转换或颠换在植物基因组中的研究. 中国科学 C 辑: 生命科学, 36(1): 1–8.]
- Zhao Y, Wang K, Wang WL, Yin TT, Dong WQ, Xu CJ, 2019. A high-throughput SNP discovery strategy for RNA-seq data. *BMC Genomics*, 20(1): 160.