基于 PacBio Iso-Seq 红棕象甲全长转录组测序分析*

杨红军[1, 2**] 胡佳萌[2] 王治博[2] 徐丹萍[1] 卓志航[1, 2, 3***]

（1. 西华师范大学，南充 637002；2. 海南大学林学院，海口 570228；3. 农业部华南作物有害生物综合治理重点实验室，广州 510642）

摘 要 **【目的】** 建立红棕象甲 *Rhynchophorus ferrugineus* 全长转录组数据库，深入挖掘红棕象甲基因数据信息。**【方法】** 采用高通量测序平台，利用二代测序（Illumina RNA-seq）校正三代测序（PacBio Iso-Seq）的方法对红棕象甲进行全长转录组测序，并对转录组数据进行生物信息学分析。**【结果】** 红棕象甲全长转录组平均长度为 2 302 bp，N90 长度为 1 321 bp，N50 长度为 2 785 bp；经 CD-Hit 程序去冗余，获得转录本 63 801 条，主要长度范围为 0.5-6 k。基因功能注释表明，在 NR、Swiss-Prot、KEGG、KOG、GO、NT 和 Pfam 数据库中，分别有 50 280、40 109、47 197、33 511、27 707、27 253 和 27 707 条转录本被注释；其中，12 508 条转录本均在 7 个数据库中有注释，54 999 条转录本至少在一个数据库有注释。此外，经鉴定或预测，获得 2 184 个可变剪接（AS）、66 230 个 SSR、2 084 个转录因子（TFs）和 9 618 条长链非编码 RNA（LncRNA）。CDS 长度的主要分布范围为 0-2 500 nt。**【结论】** 本研究获得了红棕象甲全长转录组数据库，为红棕象甲后续分子生物学基础研究奠定基础。

关键词 红棕象甲；转录组；基因组注释；高通量测序

Using PacBio Iso-Seq to determine the transcriptome of *Rhynchophorus ferrugineus*

YANG Hong-Jun[1, 2**] HU Jia-Meng[2] WANG Zhi-Bo[2] XU Dan-Ping[1] ZHUO Zhi-Hang[1, 2, 3***]

(1. College of Life Science, China West Normal University, Nanchong 637002, China; 2. College of Forestry, Hainan University, Haikou 570228, China; 3. The Key Laboratory of Integrated Pest Management on Crops in South China, Ministry of Agriculture, P. R. China, Guangzhou 510642, China)

Abstract **[Objectives]** To obtain genetic data on, and establish a transcriptome database for, *Rhynchophorus ferrugineus*. **[Methods]** The transcriptome of *R. ferrugineus* was sequenced using a high-throughput sequencing platform with second-generation sequencing (Illumina RNA-seq) to correct third-generation sequencing (PacBio Iso-Seq) data, and analyzed with bioinformatic software. **[Results]** The mean number of *R. ferrugineus* transcripts was 2 302 bp and the N90 and N50 lengths were 1 321 bp and 2 785 bp, respectively. An additional 63 801 transcripts were obtained using the CD-Hit program with the majority ranging in length between 0.5 and 6 k. Of the transcripts found, 50 280 were annotated in the NR database, 27 253 in the NT database, 47 197 in the KEGG database, 40 109 in the Swiss-Prot database, 27 707 in the Pfam database, 27 707 in the GO database and 33 511 in the KOG database. 54 999 were annotated in at least one database and 12 058 were annotated in all databases. Furthermore, 2 184 AS, 66 230 SSR, 20 84 TFs and 9 618 LncRNA were predicted, or identified, respectively, and the main CDS length range was 0-2 500 nt. **[Conclusion]** The transcriptome database of *R. ferrugineus* was successfully obtained. The results provide a foundation for further study of the molecular biology of this species.

Key words *Rhynchophorus ferrugineus*; transcriptome; genome annotation; high-throughput sequencing

---

*资助项目 Supported projects：海南省自然科学基金（319QN163）；农业部华南作物有害生物综合治理重点实验室 2017/2018 年度开放基金（SCIPM2018-04）；海南大学科研启动项目[KYQD(ZR)1823]；海南大学大学生创新创业训练计划项目（201810589250）；四川省一流本科专业；四川省第二批地方普通本科高校应用型示范专业（川教函[2019]31 号）

**第一作者 First author，E-mail：yanghongjun92@foxmail.com

***通讯作者 Corresponding author，E-mail：zhuozhihang@foxmail.com

收稿日期 Received：2020-04-24；接受日期 Accepted：2020-07-20

红棕象甲 *Rhynchophorus ferrugineus* 属鞘翅目 Coleoptera 象甲科 Curculionidae，是棕榈科 Palmae 植物主要害虫（Vatanparast *et al.*，2014；Wakil *et al.*，2015）。该虫起源于东南亚，之后迅速蔓延到亚洲和大洋洲等地，严重危害棕榈科植物（Soroker *et al.*，2005）。红棕象甲在我国属外来入侵害虫，危害面积超 1 万 km^2，严重威胁沿海地区生态安全（Shi *et al.*，2014；Ge *et al.*，2015）。该虫主要以幼虫钻蛀危害，通过破坏寄主顶端分生组织，威胁植物机械稳定，因其生活隐蔽且前期症状不明显，往往造成后期防治效果不佳，树木风折坍塌严重（Ferry and Gomez，2002；Sacchetti，2006；Hussain *et al.*，2013）。目前，高通量转录组测序研究在农林学科应用已较为成熟，在昆虫学方面，则主要集中于抗药性、生长发育和遗传变异等方面，这将为红棕象甲挖掘新的药效靶标、利用分子设计实现害虫防治等提供理论依据。

高通量转录组测序根据碱基滚动读取长度，分为二代测序和三代测序。二代测序广泛用于基因表达水平研究（Nagalakshmi *et al.*，2008；Ekblom and Galindo，2011；Djebali *et al.*，2012），如家蚕 *Bombyx mori*（Li *et al.*，2020）、赤拟谷盗 *Tribolium castaneum*（Guo *et al.*，2019a）、大蜡螟 *Galleria mellonella*（杨爽等，2019）、棉铃虫 *Helicoverpa armigera*（Wang *et al.*，2020）和椰子二疣犀 *Oryctes rhinoceros* L.（Arvind *et al.*，2020）等，但二代测序读取长度短，不能跨越整个转录本（Koren *et al.*，2012）。与第二代测序相比，三代测转录组测序无需组装即可捕获全长转录本，近年来三代测序逐渐应用于多种昆虫学研究，如灰飞虱 *Laodelphax striatellus*（Zhu *et al.*，2017）、白星花金龟 *Protaetia brevitarsis Lewis*（Wang *et al.*，2019）、窗萤 *Pyrocoelia pectoralis*（Fu *et al.*，2017）等。目前，关于红棕象甲二代测序虽有部分研究（Antony *et al.*，2019；Habineza *et al.*，2019；Muhammad *et al.*，2019），但三代全长测序的报道相对较少。综上，红棕象甲作为棕榈科主要蛀干害虫，本文采用二代测序（Illumina RNA-seq）校正三代测序（PacBio Iso-Seq）方法，对红棕象甲进行全长转录组测序，并通过基因功能注释、CDS 预测、SSR 分析和 TFs 分析等，为其相关基因互作、转录表达、功能注释及分子标记和开发提供理论依据，为进一步从分子水平开展红棕象甲防治奠定基础，同时也丰富鞘翅目昆虫全长转录组数据库。

1 材料与方法

1.1 供试昆虫

供试红棕象甲采自中国热带农业科学院椰子研究所（海南文昌），且虫体健康活跃；采集后，迅速用液氮冷冻并保存于 - 80 ℃低温冰箱中备用。

1.2 Total RNA 提取和高通量测序

RNeasy Plus Mini Kit（Qiagen，Valencia，CA，USA）提取样品 Total RNA；1%琼脂糖凝胶电泳检测 RNA 降解程度及污染；Nanodrop（NanoDrop products，USA）检测 RNA 的纯度（$OD_{260/280}$）；Qubie 对 RNA 浓度进行精确定量；Agilent 2100 精确检测 RNA 的完整性。PacBio Iso-Seq 高通量测序采用红棕象甲 Total RNA 混合样。Illumina RNA-Seq 高通量测序采用红棕象甲单个 Total RNA 样品。红棕象甲全长转录组测序工作主要由北京诺禾致源科技股份有限公司协助完成。

1.3 PacBio Iso-Seq 数据处理和校正

SMRTlink 6.0 软件对高通量序列数据进行处理，从 subread BAM files 获得循环一致性序列（Cyclic consensus sequence，CCS）并生成 CCS.BAM files；通过 pbclassify 将 CCS.BAM files 分为全长读取和非全长读取，并将全长和非全长的 fasta 文件执行 isoform 聚类，最后使用 Quiver 进行快速校正。PacBio Iso-Seq 数据通过 Illumina RNA-seq 数据和 LoRDEC 软件校正（Salmela and Rivals，2014）。LoRDEC 采用混合纠错的模式，首先读取 RNA-seq 测序的 reads，利用 reads 数据构建 DBG（de Bruijn Graph）图；然后依次读取 PacBio Iso-Seq 的 reads，在构建的 DBG 图中判断三代数据中是否有二代数据支

持；对 DBG 图中没有二代数据支持的数据进行校正，输出校正后的序列。并经 CD-HIT-EST 程序去冗余后，获得红棕象甲全长转录组（Fu *et al.*，2012）。

1.4　生物信息学分析

功能注释：通过 NR、NT、Pfam、KOG、Swiss-Prot、KEGG 和 GO 数据库分别对红棕象甲全长转录组进行功能注释（E-value≤1×10^{-5}）。

转录因子（TFs）鉴定：通过动物转录因子数据库（Animal TFDB 2.0 database）进行转录因子鉴定（Zhang *et al.*，2015）。不在数据库中的物种，由 HMMSEARCH 软件根据转录因子家族的蛋白质家族数据库文件（Pfam files）鉴定。

长链非编码 RNA（lncRNAs）预测：通过 Coding-Non-Coding-Index（CNCI）（Altschul *et al.*，1997），Coding Potential Calculator（CPC）（Kong *et al.*，2007），Pfam-scan（Finn *et al.*，2016）和 PLEK（Li *et al.*，2014a）鉴定转录组 lncRNAs。

可变剪接（AS）分析：使用编码基因组重建工具（Cogent v3.1，https://github.com/Magdoll/Cogent）和 SUPPA（https://github.com/comprna/SUPPA，默认设置）对红棕象甲全长转录组 AS 进行分析。

编码序列（CDS）预测及微卫星序列 SSR 分析：ANGEL pipeline 可实现 ANGEL 长读取，能确定全长互补脱氧核糖核酸（cDNAs）的蛋白质编码序列（CDS）。本文采用红棕象甲及其近缘物种的蛋白序列进行 ANGLE 测试，然后对给定序列进行 ANGLE 预测（Shimizu *et al.*，2006）。同时，通过 MIS（http://pgrc.ipk-gatersleben.de/misa/misa.html）分析红棕象甲全长转录组的 SSR。

2　结果与分析

2.1　全长转录组数据统计

PacBio Iso-Seq 测序结果表明，红棕象甲全长转录组数据库包含 454 369 条 CCSs，362 466 条全长序列（Full length reads），81 424 条非全长序列（Nonfull-length reads），330 973 条全长非嵌合体序列（Ull-length non chimera），平均长度为 2 332 bp。Illumina RNA-seq 测序结果表明，红棕象甲转录组数据库包含 642 179 304 条原始序列（Raw reads）和 625 983 256 条干净序列（Clean reads）。此外，PacBio Iso-Seq 测序结果含有 10 172 136 条子序列（Subreads），181 405 条共识序列（Consensus reads）；其中，共识序列平均长度 2 305 bp，N90 长度 1 327 bp，N50 长度 2 790 bp。共有 181 405 条 PacBio Iso-Seq 序列经过 Illumina RNA-seq 序列校正后，共识序列平均长度 2 302 bp，N90 长度 1 321 bp，N50 长度 2 785 bp。通过 CD-Hit 程序去冗余和校正共识序列再聚类，获得转录本 63 801 条，主要长度分布范围 0.5-6 k（图 1）。

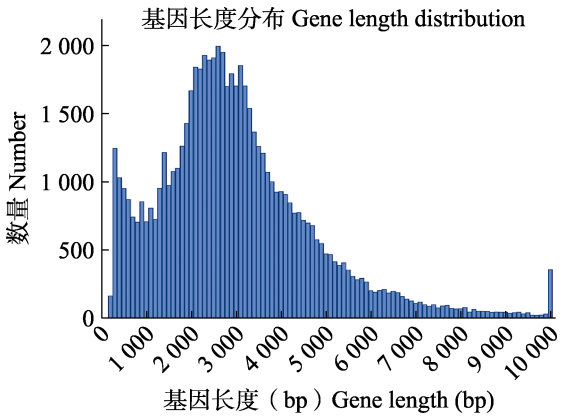

图 1　PacBio Iso-Seq 获得红棕象甲基因的长度分布

Fig. 1　Length distribution of *Rhynchophorus ferrugineus* unigenes obtained by PacBio Iso-Seq

2.2　功能注释

为获得红棕象甲全长转录组功能注释，本文通过 7 个数据库对 63 801 条转录本进行注释，结果表明：在 NR、Swiss-Prot、KEGG、KOG、GO、NT 和 Pfam 数据库中，分别有 50 280、40 109、47 197、33 511、27 707、27 253 和 27 707 条转录本被注释；其中，12 508 条转录本均在 7 个数据库中有注释，54 999 条转录本至少在一个数据库有注释（图 2）。

2.3　TFs 鉴定

动物转录因子数据库及 HMMSEARCH 软件

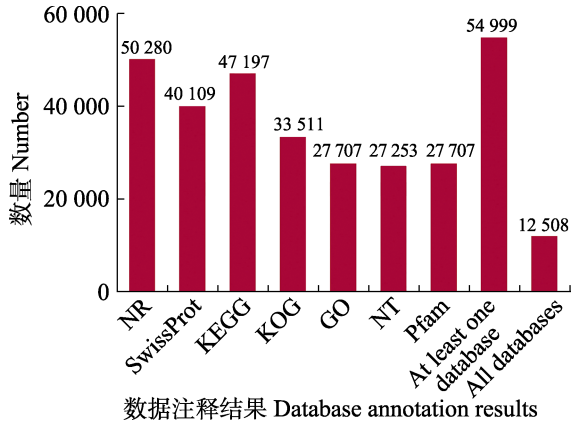

图 2　红棕象甲转录本功能注释

Fig. 2　Function annotation of *Rhynchophorus ferrugineus* transcripts

NR：NCBI 官方的蛋白序列数据库；SwissProt：蛋白序列数据库；KEGG：京都基因和基因组百科全书；KOG：真核直系同源组；GO：基因功能描述的分类系统；NT：NCBI 核酸序列数据库；Pfam：蛋白结构域注释的分类系统；At least one database：至少在一个数据库注释成功的转录本数目；All databases：在所有数据库中注释成功的转录本数目。

NR: NCBI non-redundant protein sequences; SwissProt: Protein sequence database; KEGG: Kyoto Encyclopedia of Genes and Genomes; KOG: euKaryotic Ortholog Groups; GO: Gene function description classification system; NT: NCBI nucleotide sequences; Pfam: Classification system for annotation of protein family; At least one database: Number of transcripts successfully annotated in at least one database; All databases: Number of transcripts successfully annotated in all databases.

TFs 鉴定结果表明：红棕象甲全长转录组含 2 084 个转录因子，其中 Zf-C2H2（27.35%）、ZTBB（22.84%）、TF_bzip（4.85%）和 bHLH（4.08%）是主要转录因子家族（图 3）。

2.4　AS 分析

Cogent 和 SUPPA 分析结果表明，红棕象甲全长转录组包含 2 184 个 AS，其中选择性 3′剪接位点（Alternative 3′ splice sites）、互斥外显子（Mutually exclusive exons）、跳跃外显子（Skipping exon）、选择性 5′剪接位点（Alternative 5′ splice sites）、保留内含子（Retained introns）和选择性第一外显子（Alternative first exons）为主要的 AS。保留内含子最多（134 个，占 6.14%），其他 5 种类型 AS 占比均不到 2%。

2.5　LncRNA 预测

红棕象甲全长转录组 LncRNA 分析结果表明，cnci、pfam、plek 和 cpc 分别预测 lncRNAs：18 897、32 552、17 481 和 34 066 个，同时预测 lncRNAs：9 618 个（图 4）。此外，通过 mRNA 与预测的 lncRNA 长度分布密度比较结果表明：相比 mRNA，lncRNA 更倾向于短转录本分布（图 5）。

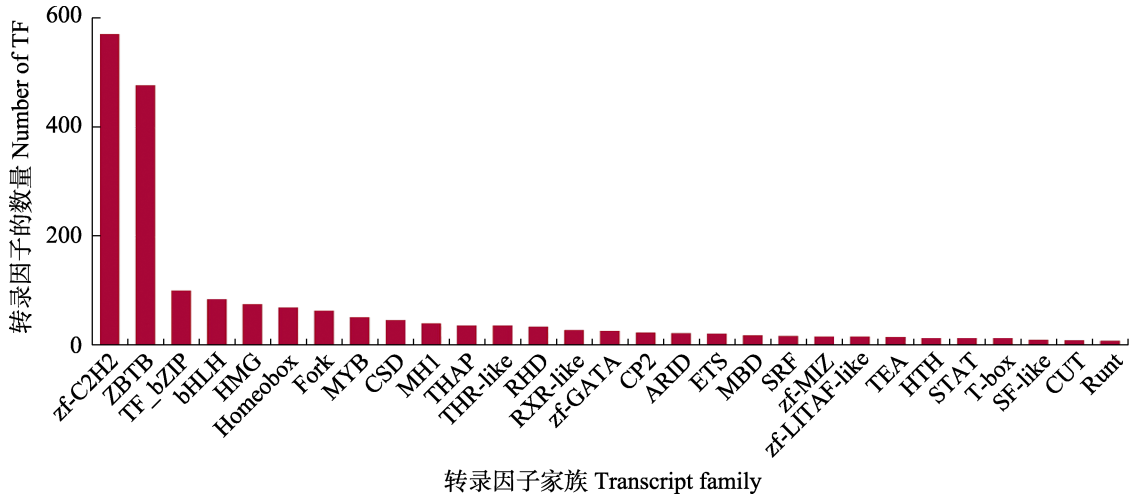

图 3　SMRT 预测的前 29 个转录因子的数量和家族

Fig. 3　Number and family of top 29 transcription factors predicted by SMRT

横坐标代表不同的转录因子家族。

The abscissa represents different transcription factor families.

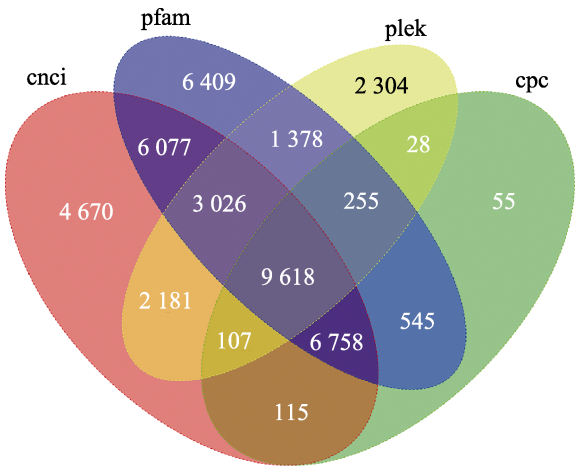

图 4　利用 cnci、pfam、plek 和 cpc 预测的 lncRNA 维恩图

Fig. 4　Venn diagram of lncRNA transcripts identified from cnci, pfam, plek and cpc

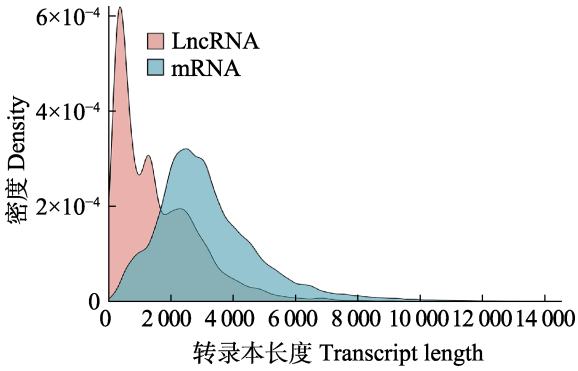

图 5　LncRNA 和 mRNA 在红棕象甲中的长度分布

Fig. 5　Length distribution of LncRNA and mRNA in *Rhynchophorus ferrugineus*

2.6　CDS 预测

CDS 是一种编码蛋白质产物的序列，它与蛋白质的密码子完全一致。在全长转录组的测序结果中，预测蛋白质编码区有助于基因的初步分析，也是后续蛋白质结构分析的基础。基于 PacBio Iso-Seq 红棕象甲全长转录组测序分析，利用 ANGEL 软件对获得的全长转录组进行 CDS 预测，结果表明 CDS 长度的主要分布范围为 0-2 500 nt（图 6）。

2.7　SSR 分析

MISA 软件 SSR 分析结果表明：每个单位尺寸的最小重复次数为 1-10、2-6、3-5、4-5、5-5、6-5。此外，红棕象甲全长转录组共鉴定出 66 230 个 SSR 位点；其中，单核苷酸重复（Mono nucleotide motifs）最多（75.34%），其后依次为双核苷酸重复（Di nucleotide motifs，19.12%）、三核苷酸重复（Tri nucleotides，5.09%）、四核苷酸重复（Tetra nucleotides，0.29%）、五核苷酸重复（Penta nucleotides，0.05%）和六核苷酸重复（Hexa-nucleotide motifs，0.10%）（图 7）。

图 6　红棕象甲转录本编码序列的数量、百分比和长度分布图

Fig. 6　Number, percentage and length distributions of coding sequences of *Rhynchophorus ferrugineus* transcripts

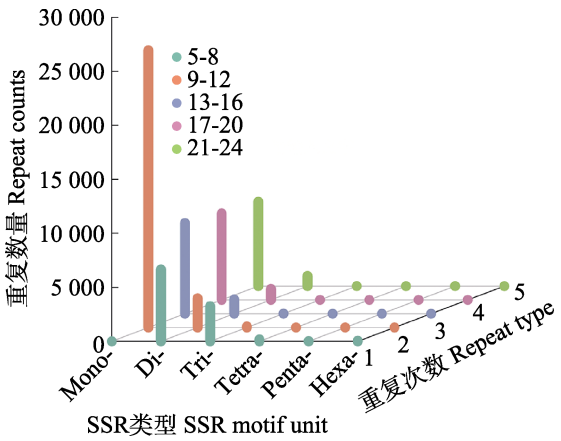

图 7　红棕象甲转录本的简单序列重复散点图

Fig. 7　Scattergram of simple sequence repeats of *Rhynchophorus ferrugineus* transcripts

Mono-：单核苷酸重复；Di-：双核苷酸重复；Tri-：三核苷酸重复；Tetra-：四核苷酸重复；Penta-：五核苷酸重复；Hexa-：六核苷酸重复。

Mono-: Mono nucleotide motifs; Di-: Di nucleotide motifs; Tri-: Tri nucleotides; Tetra-: Tetra nucleotides; Penta-: Penta nucleotides; Hexa-: Hexa-nucleotide motifs.

3 讨论

目前，RNA-seq 二代转录组测序广泛运用于昆虫学相关研究，特别是基因表达水平和差异基因表达模式等（Fatih and Patrice，2010），如家蚕（Li *et al.*，2020）、赤拟谷盗（Guo *et al.*，2019a）、白蜡虫 *Ericerus pela*（赵遵岭等, 2019）、棉铃虫（Wang *et al.*, 2020）和椰子二疣犀（Arvind *et al.*，2020）等。但二代测序具有读取长度短、剪接转录不完整、难以准确预测基因结构及功能注释错误等缺点（Coghlan *et al.*，2008；Koren *et al.*，2012；Lin *et al.*，2017；Li *et al.*，2018）。与第二代测序相比，三代转录组测序无需组装即可捕获全长转录本，具有可识别多种可变剪切形式和剪接位点、检测新功能基因、补充基因组注释等优点，逐渐应用于多种昆虫学研究。如小菜蛾 *Plutella xylostella*（Zhao *et al.*，2019）、日本栎蚕蛾 *Antheraea yamamai*（Kim *et al.*，2018）、家蚕（Kawamoto *et al.*，2019）等。然而，三代测序虽然具有超长读取优势，其单碱基读取错误率较高，需要第二代测序校正（Au *et al.*，2013；Li *et al.*，2014b）。

为建立红棕象甲全长转录组数据库，深入挖掘红棕象甲基因数据，为该虫分子生物学研究提供数据支撑，本文同时采用高通量测序 PacBio Iso-Seq 和 Illumina RNA-seq 对红棕象甲进行全长转录组测序，并对转录组数据进行生物信息学分析。结果表明红棕象甲全长转录组平均长度为 2 302 bp，N90 长度为 1 321 bp，N50 长度为 2 785 bp；经 CD-Hit 程序去冗余，获得转录本 63 801 条，主要长度范围为 0.5-6 k。与二代测序数据量相比，二代测序结合三代测序获得红棕象甲转录组数据高于其它多种二代测序鞘翅目昆虫，如松墨天牛 *Monochamus alternatus*（Li *et al.*，2019a）、暗黑鳃金龟 *Holotrichia parallela*（Yi *et al.*，2018）、咖啡果小蠹 *Hypothenemus hampei*（Noriega *et al.*，2019）等。此外，基因功能注释表明 86.20% 红棕象甲转录本能成功注释，注释率高于其它多种二代测序鞘翅目昆虫，如茄二十八星瓢虫 *Henosepilachna vigintioctopunctata*（Guo *et al.*，2019b）、刀角瓢虫 *Serangium japonicum*（Hu *et al.*，2020）、条背萤 *Sclerotia aquatilis*（Chanchay *et al.*，2019）等。与此同时，相较于其他利用 PacBio Iso-Seq 测序的昆虫，如白背飞虱 *Sogatella furcifera*（Chen *et al.*，2020a）测序获得了 29 700 条转录本，莲草直胸跳甲 *Agasicles hygrophila*（Jia *et al.*，2018）测序获得 28 982 条转录本，家蚕 *Bombyx mori*（Chen *et al.*，2020b）丝腺的测序获得了 11 697 条转录本，本研究获得了更多的昆虫全长转录组数据。通过动物转录因子数据库及 HMMSEARCH 软件分析，本文获得 2 048 个 TFs，而不同 TFs 可能参与不同的代谢过程，并可能具有多种不同的功能（Chen and Rajewsky，2007）。例如 AhR/ARNT 可能调节与农药抗性相关的多个解毒基因的表达（Pan *et al.*，2019）；转录因子 MaPacC 是耐热性的负调节因子，在蝗虫免疫方面发挥着独特的作用（Zhang *et al.*，2020）；转录因子 FTZ-F1 和顺式作用元件介导 CYP6BG1 的表达，能够调控小菜蛾对杀虫剂（氯虫酰胺）的抗药性（Li *et al.*，2019b）。综上，TFs 分析将为进一步研究红棕象甲免疫和抗药性提供数据支撑。此外，本文通过 Cogent 和 SUPPA 分析获得 2 184 个 AS，通过 MISA 软件分析获得 66 230 个 SSR，通过 lncRNAs 预测获得 9 618 个 lncRNAs，这都将为进一步挖掘红棕象甲新的药效靶标、利用分子设计实现害虫防治等提供理论依据。综上，本文基于 PacBio Iso-Seq，成功获得了红棕象甲全长转录组，并成功注释了红棕象甲全长转录组相关基因功能信息，不仅丰富了鞘翅目昆虫基因数据库，也将为下一步红棕象甲的分子标记开发、行为生态学研究、捕获其免疫和抗药性基因等研究提供数据支撑。

参考文献 (References)

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ, 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25(17): 3389–3402.

Antony B, Johny J, Abdelazim M, Jakše J, Al-Saleh M, Pain A, 2019.

Global transcriptome profiling and functional analysis reveal that tissue-specific constitutive overexpression of cytochrome P450s confers tolerance to imidacloprid in palm weevils in date palm fields. *BMC Genomics*, 20(1): 440.

Arvind K, Rajesh MK, Josephrajkumar A, Grace T, 2020. Dataset of *de novo* assembly and functional annotation of the transcriptome of certain developmental stages of coconut rhinoceros beetle, *Oryctes rhinoceros* L. *Data in Brief*, 28: 105036.

Au KF, Sebastiano V, Afshar PT, Durruthy JD, Lee L, Williams BA, van Bakel H, Schadt EE, Reijo-Pera RA, Underwood JG, Wong WH, 2013. Characterization of the human ESC transcriptome by hybrid sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 110(50): E4821–E4830.

Chanchay P, Vongsangnak W, Thancharoen A, Sriboonlert A, 2019. Reconstruction of insect hormone pathways in an aquatic firefly, *Sclerotia aquatilis* (Coleoptera: Lampyridae), using RNA-seq. *Peer J.*, 7: e7428.

Chen K, Rajewsky N, 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nature Reviews Genetics*, 8: 93–103.

Chen J, Yu YY, Kang K, Zhang DW, 2020a. SMRT sequencing of the full-length transcriptome of the white-backed planthopper *Sogatella furcifera*. *Peer J*., 8: e9320.

Chen T, Sun QW, Ma Y, Zeng WH, Liu RP, Qu DW, Huang LH, Xu HF, 2020b. A transcriptome atlas of silkworm silk glands revealed by PacBio single-molecule long-read sequencing. *Molecular Genetics and Genomics*, 295(5): 1227–1237.

Coghlan A, Fiedler TJ, McKay SJ, Flicek P, Harris TW, Blasiar D, Stein LD, Consortium TN, 2008. nGASP-the nematode genome annotation assessment project. *BMC Bioinformatics*, 9: 549.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Der MR, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See L, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigo R, Gingeras TR, 2012. Landscape of transcription in human cells. *Nature*, 489(7414): 101–108.

Ekblom R, Galindo J, 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107(1): 1–15.

Fatih O, Patrice MM, 2010. RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics*, 12(2): 87–98.

Ferry M, Gomez S, 2002. The red palm weevil in the mediterranean area. *Palms*, 46(4): 172–178.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A, 2016. The pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, 44(D1): D279–D285.

Fu L, Niu B, Zhu Z, Wu S, Li W, 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23): 3150–3152.

Fu X, Li J, Tian Y, Quan W, Zhang S, Liu Q, Liang F, Zhu X, Zhang L, Wang D, Hu J, 2017. Long-read sequence assembly of the firefly *Pyrocoelia pectoralis* genome. *GigaScience*, 6(12): 1–6.

Guo SH, Yu L, Liu YM, Wang FF, Chen YC, Wang Y, Qiu BL, Sang W, 2019a. Digital gene expression profiling in larvae of *Tribolium castaneum* at different periods post UV-B exposure. *Ecotoxicology and Environmental Safety*, 174: 514–523.

Guo W, Lü J, Guo MJ, Chen SM, Qiu BL, Sang W, Yang CX, Zhang YJ, Pan HP, 2019b. *De Novo* transcriptome analysis reveals abundant gonad-specific genes in the ovary and testis of *Henosepilachna vigintioctopunctata*. *International Journal of Molecular Sciences*, 20(17): 4084.

Ge X, He S, Wang T, Yan W, Zong S, 2015. Potential distribution predicted for *Rhynchophorus ferrugineus* in China under different climate warming scenarios. *PLoS ONE*, 10(10): e0141111.

Habineza P, Muhammad A, Ji T, Xiao R, Yin X, Hou Y, Shi Z, 2019. The promoting effect of gut microbiota on growth and development of red palm weevil, *Rhynchophorus ferrugineus* (Olivier) (Coleoptera: Dryophthoridae) by modulating its nutritional metabolism. *Frontiers in Microbiology*, 10: 1212.

Hu YH, Liu Y, Wei L, Chen HT, 2020. *De Novo* transcriptome sequencing of *Serangium japonicum* (Coleoptera: Coccinellidae) and application of two assembled unigenes. *G3-Genes Genomes Genetics*, 10(1): 247–254.

Hussain A, Rizwan-ul-Haq M, Al-Jabr AM, Al-Ayied HY, 2013. Managing invasive populations of red palm weevil: A worldwide perspective. *Journal of Food Agriculture and Environment*, 11(2): 456–463.

Jia D, Wang YX, Liu YH, Hu J, Guo YQ, Gao LL, Ma RY, 2018. SMRT sequencing of full-length transcriptome of flea beetle *Agasicles hygrophila* (Selman and Vogt). *Scientific Reports*, 8(1): 2197.

Kawamoto M, Jouraku A, Toyoda A, Yokoi K, Minakuchi Y, Katsuma S, Fujiyama A, Kiuchi T, Yamamoto K, Shimada T, 2019. High-quality genome assembly of the silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, 107: 53–62.

Kim S, Kwak W, Kim H, Caetano-Anolles K, Kim K, Kim S, Choi K, Kim S, Hwang J, Kim M, Kim I, Goo T, Park S, 2018. Genome sequence of the Japanese oak silk moth, *Antheraea yamamai*: The first draft genome in the family Saturniidae. *GigaScience*, 7(1): gix113.

Kong L, Zhang Y, Ye Z, Liu X, Zhao S, Wei L, Gao G, 2007. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*, 35: W345–W349.

Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, Phillippy AM, 2012. Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nature Biotechnology*, 30(7): 693–700.

Li A, Zhang J, Zhou Z, 2014a. PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved *k-mer* scheme. *BMC Bioinformatics*, 15(1): 311.

Li H, Zhao X, Qiao H, He X, Tan J, Hao D, 2019a. Comparative transcriptome analysis of the heat stress response in *Monochamus alternatus* Hope (Coleoptera: Cerambycidae). *Frontiers in Physiology*, 10: 1568.

Li J, Li LP, Yu YC, Thakur K, Fei H, Shun MT, Zhao JW, 2020. Transcriptome analysis reveals gene expression changes of the fat body of silkworm (*Bombyx mori* L.*)* in response to selenium treatment. *Chemosphere*, 245: 125660.

Li Q, Li Y, Song J, Xu H, Xu J, Zhu Y, Li X, Gao H, Dong L, Qian J, Sun C, Chen S, 2014b. High-accuracy *de novo* assembly and SNP detection of chloroplast genomes using a SMRT circular consensus sequencing strategy. *New Phytologist*, 204(4): 1041–1049.

Li Y, Fang C, Fu Y, Hu A, Li C, Zou C, Li X, Zhao S, Zhang C, Li C, 2018. A survey of transcriptome complexity in *Sus scrofa* using single-molecule long-read sequencing. *DNA Research*, 25(4): 421–437.

Li X, Shan C, Li F, Liang P, Smagghe G, Gao X, 2019b. Transcription factor FTZ-F1 and cis-acting elements mediate expression of CYP6BG1 conferring resistance to chlorantraniliprole in *Plutella xylostella*. *Pest Management Science*, 75(4): 1172–1180.

Lin H, Lin X, Zhu J, Yu X, Xia X, Yao F, Yang G, You M, 2017. Characterization and expression profiling of serine protease inhibitors in the diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae). *BMC Genomics*, 18(1): 162.

Muhammad A, Habineza P, Ji T, Hou Y, Shi Z, 2019. Intestinal microbiota confer protection by priming the immune system of red palm weevil *Rhynchophorus ferrugineus* Olivier (Coleoptera: Dryophthoridae). *Frontiers in Physiology*, 10: 1303.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M, 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320(5881): 1344–1349.

Noriega DD, Arias PL, Barbosa HR, Arraes FBM, Ossa GA, Villegas B, Coelho RR, Albuquerque EVS, Togawa RC, Grynberg P, Wang H, Vélez AM, Arboleda JW, Grossi-de-Sa MF, Silva MCM, Valencia-Jiménez A, 2019. Transcriptome and gene expression analysis of three developmental stages of the coffee berry borer, *Hypothenemus hampei*. *Scientific Reports*, 9: 12804–12813.

Pan Y, Peng T, Xu P, Zeng X, Tian F, Song J, Shang Q, 2019. Transcription factors AhR/ARNT regulate the expression of CYP6CY3 and CYP6CY4 switch conferring nicotine adaptation. *International Journal of Molecular Sciences*, 20(18): E4521.

Sacchetti P, 2006. Identificazione, biologia e diffusione delcurculionide delle palme, *Rhynchophorus ferrugineus* (Olivier). *Informatore Fitopatologico*, 20: 35–40.

Salmela L, Rivals E, 2014. LoRDEC: Accurate and efficient long read error correction. *Bioinformatics*, 30(24): 3506–3514.

Shi ZH, Lin YT, Hou YM, 2014. Mother-derived trans-generational immune priming in the red palm weevil, *Rhynchophorus ferrugineus* Olivier (Coleoptera: Dryophthoridae). *Bulletin of Entomological Research*, 104(6): 742–750.

Shimizu K, Adachi J, Muraoka Y, 2006. ANGLE: A sequencing errors resistant program for predicting protein coding regions in unfinished cDNA. *Journal of Bioinformatics and Computational Biology*, 4(3): 649–664.

Soroker V, Blumberg D, Haberman A, Hamburger-Rishard M,

Reneh S, Talebaev S, Anshelevich L, Harari AR, 2005. Current status of red palm weevil infestation in date palm plantations in Israel. *Phytoparasitica*, 33(1): 97–106.

Vatanparast M, Hosseininaveh V, Ghadamyari M, Sajjadian SM, 2014. Plant cell wall degrading enzymes, pectinase and cellulase, in the digestive system of the red palm weevil, *Rhynchophorus ferrugineus* (Coleoptera: Curculionidae). *Plant Protection Science*, 26(1): 7–10.

Wakil W, Faleiro JR, Miller TA, 2015. Sustainable Pest Management in Date Palm: Current Status and Emerging Challenges, Sustainability in Plant and Crop Protection. Switzerland: Springer International Press. 1–11.

Wang K, Li P, Gao Y, Liu C, Wang Q, Yin J, Zhang J, Geng L, Shu C, 2019. De *novo* genome assembly of the white-spotted flower chafer (Protaetia brevitarsis). *GigaScience*, 8(4): 1–9.

Wang M, Zhang SS, Shi Y, Yang YH, Wu YD, 2020. Global gene expression changes induced by knockout of a protease gene cluster in *Helicoverpa armigera* with CRISPR/Cas9. *Journal of Insect Physiology*, 122: 104023.

Yang S, Zhao HT, Xu K, Guo LY, Du YL, Li XY, Su WT, Feng YJ, Long DL, Jiang YS, 2019. Analysis of the antennal transcriptome and olfaction-related genes of the greater wax moth *Galleria mellonella* (Lepidoptera: Pyralidae). *Chinese Journal of Applied Entomology*, 56(6): 1279–1291. [杨爽, 赵慧婷, 徐凯, 郭丽娜, 杜亚丽, 李新宇, 苏文婷, 冯宇佳, 龙登隆, 姜玉锁, 2019. 大蜡螟触角转录组及嗅觉相关基因分析. 应用昆虫学报, 56(6): 1279–1291.]

Yi JK, Yang S, Wang S, Wang J, Zhang XX, Liu Y, Xi JH, 2018. Identification of candidate chemosensory receptors in the antennal transcriptome of the large black chafer *Holotrichia parallela* Motschulsky (Coleoptera: Scarabaeidae). *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 28: 63–71.

Zhang M, Wei Q, Xia Y, Jin K, 2020. MaPacC, a pH-responsive transcription factor, negatively regulates thermotolerance and contributes to conidiation and virulence in metarhizium acridum. *Current Genetics*, 66(2): 397–408.

Zhang H, Liu T, Liu C, Song S, Zhang X, Liu W, Jia H, Xue Y, Guo A, 2015. AnimalTFDB 2. 0: A resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Research*, 43: D76–D81.

Zhao Q, Zhong W, He W, Li Y, Li Y, Li T, Vasseur L, You M, 2019. Genome-wide profiling of the alternative splicing provides insights into development in *Plutella xylostella*. *BMC Genomics*, 20(1): 414–463.

Zhao ZL, Yu SH, Zhou WW, Zhu ZR, Guo ZJ, Yang P, 2019. The composition and function of the honeydew secreted by *Ericerus pela* as deduced from metabolome and transcriptome data. *Chinese Journal of Applied Entomology*, 56(5): 1007–1018. [赵遵岭, 于淑惠, 周文武, 祝增荣, 郭忠建, 杨璞, 2019. 基于代谢组和转录组的白蜡虫吊糖成分及其作用分析. 应用昆虫学报, 56(5): 1007–1018.]

Zhu J, Jiang F, Wang X, Yang P, Bao Y, Zhao W, Wang W, Lu H, Wang Q, Cui N, Li J, Chen X, Luo L, Yu J, Kang L, Cui F, 2017. Genome sequence of the small brown *planthopper*, *Laodelphax striatellus*. *GigaScience*, 6(12): 1–6.