Data Article

Dataset of *de novo* assembly and functional annotation of the transcriptome of certain developmental stages of coconut rhinoceros beetle, *Oryctes rhinoceros* L.

Kumar Arvind [a], M.K. Rajesh [b], A. Josephrajkumar [c], Tony Grace [a],*

[a] Central University of Kerala, Kasaragod, 671320, Kerala, India
[b] ICAR-Central Plantation Crops Research Institute, Kasaragod, 671124, Kerala, India
[c] ICAR-Central Plantation Crops Research Institute, Regional Station, Kayamkulam, 690533, India

ARTICLE INFO

ABSTRACT

The coconut rhinoceros beetle, *Oryctes rhinoceros* L. (Insecta: Coleoptera: Scarabaeidae: Dynastinae) is one of the world's most important endemic and incessant pests of coconut (particularly in India and Southeast Asia), causing an estimated 10% yield loss in the crop. Various management strategies formulated and implemented to control this pest include bioagents, insecticide sprays, liquid formulations, pheromone traps, and botanical formulations. Also, potential microbial bioagents *viz., Oryctes rhinoceros nudivirus (OrNV)* and *Metarhizium anisopliae* have been implemented as biological control agents and this has led to a beneficial reduction of the pest population unless significant immigration occurs. To date, research and development activities are still on-going for the successful management of the pest; yet advances in understanding at the molecular level have been limited because basic genomic information is lacking for this cosmopolitan pest. Transcriptome approach has been proved extremely useful in finding potential genes for pest control. Transcriptome analysis aids in gaining insights into the transcriptional changes which occur during different developmental stages of an organism. We have performed RNA sequencing of certain different developmental stages of *O. rhinoceros viz.*, early instar larva, late instar larva, pupa, and

---

* Corresponding author.
  E-mail address: tonygrace@cukerala.ac.in (T. Grace).

adult, in an Illumina HiSeq™ 2500 platform. Due to the unavailability of *O. rhinocero*s genome, the RNA-seq data generated were assembled *de novo* using Trinity and annotated following redundancy removal. A dataset of 87,451 transcripts, which resulted after redundancy removal, were annotated using the NCBI non-redundant (nr) protein and Uniprot databases. The data furnished could be used by others working in the development of pest management strategies, especially the identification of molecular targets for effective pest control. This information allows a better understanding of *O. rhinoceros* biology which would contribute to outlining a new generation of stage-specific, environmentally friendly pest management techniques.

Specifications Table

Subject	Biology
Specific subject area	Transcriptomics
Type of data	Tables, figures
How data were acquired	Illumina Hiseq™ 2500 sequencing platform
Data format	Raw sequencing data and analyzed data
Parameters for data collection	Certain life stages (early instar larva, late instar larva, pupa and adult) of *Oryctes rhinoceros*.
Description of data collection	Different developmental stages of the coconut rhinoceros beetle, *O. rhinoceros* that included early instar larva, late instar larva, pupa pupa and adult were sampled.
Data source location	Central University of Kerala, Kasaragod 671,320, Kerala, India
Data accessibility	Repository name: NCBI SRA Data identification number: PRJNA486419 Direct URL to data: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA486419/

Value of Data

- The coconut rhinoceros beetle, *O. rhinoceros* L. is an indigenous pest of coconut reported from most of the coconut growing regions of the world.
- The damage by the pest causes significant reduction in coconut production.
- The transcriptome data of the certain developmental stages presented represents the first comprehensive molecular resource for this species.
- These datasets may be used to identify differentially expressed genes among the different life stages of the beetle.
- The dataset may be used for deciphering relevant candidate proteins critical for evading and damaging the host as well as resistance against the various control strategy.
- The RNA-seq and assembled transcriptome datasets would provide evidence of gene expression for use of the researchers for gene prediction and functional annotation of the *Oryctes rhinoceros* genome as and when it becomes available.

1. Data

This data article reports the first comprehensive deep-sequencing transcriptome datasets of certain different developmental stages of *Oryctes rhinoceros* L. comprising of four life stages *viz.*, early instar larva (EIL), late instar larva (LIL), pupa, and adult. Table 1 provides the RNA-seq statistics (both raw reads and clean reads) of the four stages of *O. rhinoceros* sampled. Table 2 provides the statistics of the raw transcriptome assembly and quality assessment. Fig. 1 displays the functional classification of *O.*

Table 1

RNA-seq statistics of four developmental stages of *O. rhinoceros*: (a) Raw reads (b) Clean reads.

(a) Raw reads				
Sl. No.	Sample name	Number of paired-end reads	Number of bases (Gb)	GC%
1.	Elder instar larva	22,746,266	4.55	44.29
2.	Late instar larva	21,694,513	4.34	41.17
3.	Pupa	22,704,107	4.54	44.44
4.	Adult	33,613,928	6.72	44.43
(b) Clean reads				
Sl. No.	Sample name	Number of paired-end reads	Number of bases (Gb)	Average read length (bases)
1.	Elder instar larva	18,563,480	3.03	81
2.	Late instar larva	16,789,020	2.75	82
3.	Pupa	18,969,930	3.09	81
4.	Adult	28,432,393	4.65	81

Table 2

Statistics of the raw transcriptome assembly and quality assessment.

Total genes from Trinity assembly	65,716
Total Trinity transcripts	87,451
Median contig length (bp)	491
Average contig length (bp)	1021.94
Contig N50 length (bp)	1979
Total assembled bases	89,369,562
Longest transcript length (bp)	27,421
Mean GC% of transcripts	39.65
Number of transcripts with significant BLASTX	39,606
Number of transcripts with UniProt annotation	39,165

rhinoceros in three Gene Ontology (GO) categories viz., biological processes, molecular functions and cellular components.

We found that 40,107 (68.79%) of assembled transcripts possessed at least one significant hit in the NCBI database. Also, 39,165 transcripts (among the significant BLASTX hits) were annotated using the UniProt database. Around 67% of the transcripts, found using BLASTX, had a confidence level of at least 1e-5, which indicates high protein conservation. Further, gene ontology (GO) was performed to assign GO identity to the annotated transcripts, resulting in 10,534 transcripts with assigned GO identities (Biological processes = 4758; Molecular functions = 3748; Cellular components = 3028) (Fig. 1). We provide the first molecular resource that integrates the assembly and annotation of different developmental stages of *O. rhinoceros.* The analyses undertaken using the RNA-seq data would be helpful in gene identification and annotation of the *O. rhinoceros* genome.

2. Experimental design, materials, and methods

2.1. Insects rearing and sampling

Rearing of insects was carried out using standardized procedures. Fifty freshly laid eggs of *O. rhinoceros* were collected from the farmyard manure heaps. Eggs were reared in sterile farmyard manure added with 15% moisture in standard plastic containers in the laboratory maintained at 27±2 °C temperature with 75 ± 5% RH in a BOD incubator (Analab, India). The incubation period was found to be about 9−11 days. Twenty-five days old grubs, as well as 85 days old grub after eclosion, constituted the early larval instar (EIL) and late larval instar (LIL), respectively. The sterile farmyard manure was changed once in seven days period for the accessibility of good quality food by the feeding grubs. Prior to pupation, the grubs wandered and constructed earthen cocoon and underwent pupation. Pupation
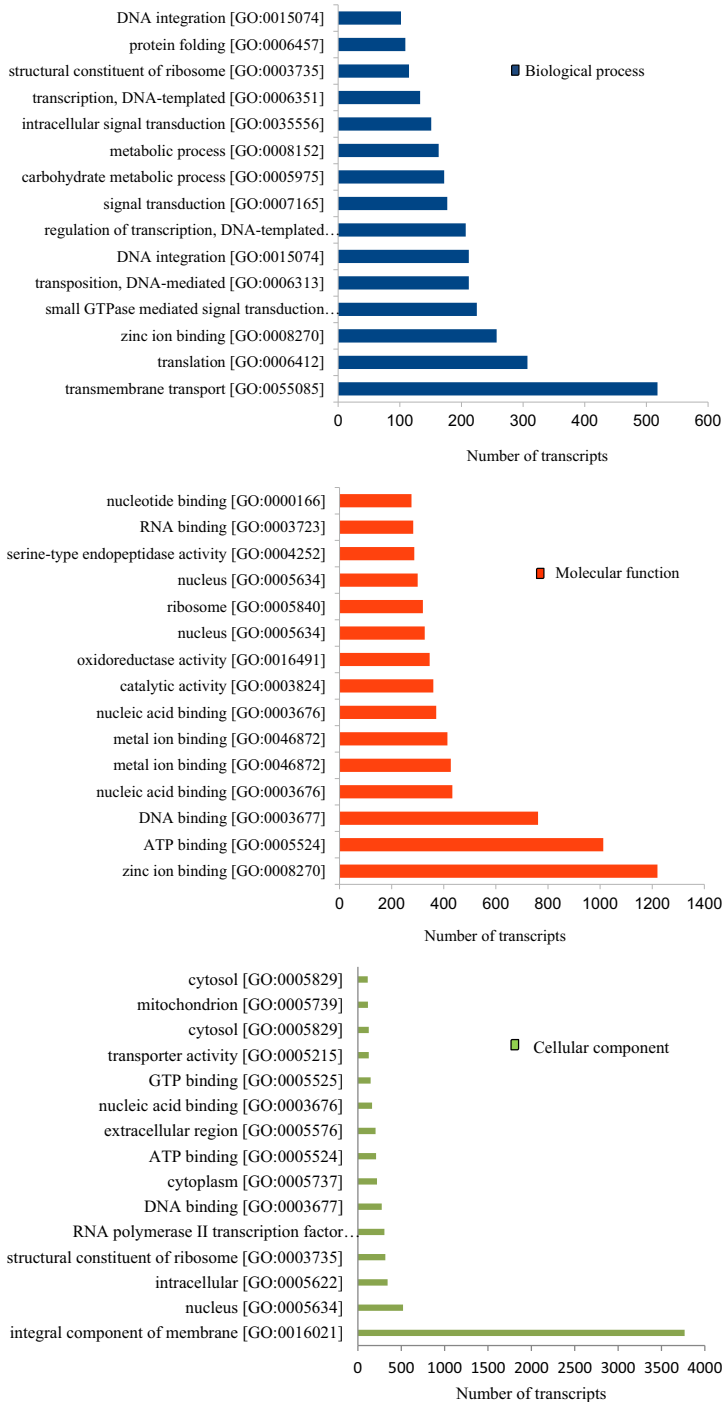
Fig. 1. Functional classification of *Oryctes rhinoceros* in three Gene Ontology (GO) categories - biological process (blue), molecular function (orange), and cellular component (green).

was completed in about 115−120 days and the pupal period extended for 24−28 days. Adult beetles, which emerged afterwards, were sexed out and were confined to iron-meshed cages with farmyard manure for mating and oviposition. The developmental stages of *O. rhinoceros* sampled include EIL, LIL, pupa, and adult. Five insects from each developmental stage were sampled.

2.2. RNA extraction and sequencing

Total RNA isolated from the whole body tissue of each individual stage was pooled (four samples/ stage) and snap-frozen immediately in liquid nitrogen and ground to a fine powder using mortar and pestle. The powder was directly transferred into Tri-Reagent (Sigma-Aldrich) and RNA was extracted with the Direct-zol™ RNA MiniPrep (Zymo, Germany). The quality and the purity of the extracted RNA were assessed by OD 260/280 ratio and RNA integrity number was analyzed using an Agilent Technologies 2100 Bioanalyzer with the Agilent RNA chip with RIN (RNA Integrity Number) > 8.0.

From the various life stages, approximately 5−10 μg of total RNA was used to prepare the RNA-Seq library using TruSeq RNA Sample Prep Kits (Illumina). From the resulting total RNAs, sequencing libraries were prepared using the TruSeq mRNA-Seq kit and protocol from Illumina, Inc. (San Diego, USA). As per Hull et al. [1], total RNA extraction and library generation (TruSeq RNA Sample Preparation Kit v2; Illumina Inc., San Diego, USA) were performed. All the four samples were sequenced using an Illumina HiSeq2500 in rapid run mode (paired-end 100-bp reads).

2.3. Data analysis (de novo assembly and annotation)

Sequencing data from all four samples (paired-end reads) were used individually for processing. Raw data passed through two stages of read trimming: first, we performed quality-based trimming using Trimmomatic v0.33 [2]. FastQC [3] was then used to check data quality before and after trimming. Filtered high quality read pairs were considered for assembly by the virtue of Trinity [4] with min_-kmer_cov set to two and all other parameters set to default. A total of 87, 451 transcripts were obtained and found to be of length more than 200 bp. Annotation of the assembled transcript was performed using CANoPI − (Contig Annotator Pipeline; AgriGenome, India) for *de novo* transcriptome assembly. For annotation, we performed the following steps for assembled transcripts: (1) Comparison with NCBI non-redundant (nr) protein and Uniprot databases using BLASTX program, (2) Organism annotation, (3) Gene and protein annotation to the matched transcript, and (4) Gene Ontology (GO) annotation [5].

Author statements

Tony Grace, M.K. Rajesh: Conceptualization, Methodology. **Kumar Arvind, M.K. Rajesh, Josephrajkumar A.**: Data generation, Curation. **Josephrajkumar A.**: Resources. **Kumar Arvind**: Writing-Original draft preparation. **Kumar Arvind, M.K. Rajesh, Josephrajkumar A.**: Investigation. **Tony Grace, M.K. Rajesh:** Supervision. **Tony Grace, M.K. Rajesh, Josephrajkumar A.**: Writing- Reviewing and Editing.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J.J. Hull, K. Chaney, S.M. Geib, J.A. Fabrick, C.S. Brent, D. Walsh, L.C. Lavine, Transcriptome-based identification of ABC transporters in the western tarnished plant bug *Lygus hesperus*, PLoS One 9 (2014), e113046, https://doi.org/10.1371/journal.pone.0113046.
[2] A.M. Bolger, M. Lohse, M. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114−2120, https://doi.org/10.1093/bioinformatics/btu170.
[3] S. Andrews, FastQC: a quality control tool for high throughput sequence data, Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc, 2010.

[4] M.G. Grabherr, B.J. Haas, M. Yassour, et al., Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data, Nat. Biotechnol. 29 (2011) 644–652, https://doi.org/10.1038/nbt.1883.

[5] H. Yang, Y. Cai, Z. Zhuo, W. Yang, C. Yang, J. Zhang, Y. Yang, B. Wang, F. Guan, Transcriptome analysis in different developmental stages of *Batocera horsfieldi* (Coleoptera: cerambycidae) and comparison of candidate olfactory genes, PLoS One 13 (2018), e0192730, https://doi.org/10.1371/journal.pone.0192730.