RNA-Seq analysis and de novo transcriptome assembly of Cry toxin susceptible and tolerant Achaea janata larvae

Narender K. Dhania¹, Vinod K. Chauhan¹, R. K. Chaitanya² & Aparna Dutta-Gupta¹

Larvae of most lepidopteran insect species are known to be voracious feeders and important agricultural pests throughout the world. Achaea janata larvae cause serious damage to Ricinus communis (Castor) in India resulting in significant economic losses. Microbial insecticides based on crystalline (Cry) toxins of Bacillus thuringiensis (Bt) have been effective against the pest. Excessive and indiscriminate use of Bt-based biopesticides could be counter-productive and allow susceptible larvae to eventually develop resistance. Further, lack of adequate genome and transcriptome information for the pest limit our ability to determine the molecular mechanisms of altered physiological responses in Bt-exposed susceptible and tolerant insect strains. In order to facilitate biological, biochemical and molecular research of the pest species that would enable more efficient biocontrol, we report the midgut de novo transcriptome assembly and clustering of susceptible Cry toxin-exposed and Cry toxin tolerant Achaea janata larvae with appropriate age-matched and starvation controls.

Background & Summary

Bacillus thuringiensis (Bt) insecticidal proteins used in sprayable formulations and transgenic crops are the most promising alternatives to synthetic insecticides. However, the evolution of resistance in the field, as well as laboratory insect populations is a serious roadblock to this technology. Achaea janata, a major castor crop pest in India, is controlled using Bt-based formulation¹ comprising of Cry1 (Cry1Aa, Cry1Ab, and Cry1Ac) and Cry2 (Cry2Aa and Cry2Ab) genes. Recent studies from our group reported extensive changes at the cellular and molecular level in the midgut of A. janata exposed to a sublethal dose of the Bt formulation³–⁵. Since a decade, reports of resistance against Bt toxins and their mechanisms have been emerging⁶–⁷. Long term exposure to Cry toxin formulations promotes tolerance in larvae which eventually leads to resistance⁸–¹¹. Development of Bt resistance could be due to alterations in proteolytic cleavage of the Cry toxin, altered receptor binding or enhanced midgut regeneration responses⁹–¹¹. With the advent of next generation sequencing technology it is now possible to characterize the entire repertoire of transcripts under different conditions and predict pathways involved in various molecular mechanisms. The RNA sequencing study presented here generated the first de novo transcriptome assembly of castor semilooper, Achaea janata (Noctuidae: Lepidoptera), and compared gene expression signatures between toxin-exposed susceptible and tolerant larvae. This article, is a first step in determining the primary basis for Cry tolerance in the pest, which could facilitate new long term management strategies.

Methods

Toxin administration and sample preparation. Wild population of A. janata larvae, unexposed to pesticides, was field collected from the Indian Institute of Oil Seed Research, Hyderabad, India. Further, the larvae were reared and maintained on castor leaves at 27 ± 2 °C, 14:10h (light: dark) photoperiod and 60–70% relative humidity for three generations at the insectary of School of Life Sciences, University of Hyderabad, India. In the present de novo transcriptome analysis for the sublethal toxin exposure, 1/10 of LD₅₀ was used (Group ii) (Fig. 1),

¹Department of Animal Biology, School of Life Sciences, University of Hyderabad, Hyderabad, 500046, India. ²Department of Animal Sciences, School of Basic and Applied Sciences, Central University of Punjab, Bathinda, 151001, India. Correspondence and requests for materials should be addressed to A.D.-G. (email: aparnaduttagupta@gmail.com) or R.K.C. (email: chaitanyark@cup.ac.in)
while for the generation of a tolerant population (Group iv) (Fig. 1) an LD50 dose of DOR Bt-1 formulation was administered. Larval batches (n = 100) designated as Cry-susceptible larvae and control larvae were exposed to toxin-coated and distilled water coated leaves respectively. The midgut was isolated from 15 randomly selected surviving larvae from each batch after every 12 h till 48 h. In earlier study we noticed that larvae probably sense the toxin and avoid feeding on toxin coated leaves after a short exposure. Hence, to eliminate any effect induced by starvation, an additional batch (Group iv) of 3rd instar larvae was maintained on moist filter paper and collected for the midgut isolation every 12 h till 48 h. All the midgut dissections were carried out in ice-cold insect Ringer solution (130 mM NaCl, 0.5 mM KCl, and 0.1 mM CaCl2). The experiment was performed in duplicates. For the Cry tolerant larval population, larvae (n = 100) in each generation were exposed to LD50 dose and the surviving insects were maintained for larval development, pupal molting, adult emergence and egg laying. The larvae hatched from the eggs were collected and reared till 3rd larval instar larvae and exposed to LD50 Bt dose once again. This schedule was carried out for fifteen generations. The batch (n = 100) of Cry tolerant larvae thus generated were exposed to toxin-coated leaves and the midguts were isolated from randomly selected fifteen larvae after 24 h. Total RNA was isolated from the midgut samples using Trizol-based method. The RNA was quantified using NanoDrop™ 8000 spectrophotometer and the quality was assessed using 1% formaldehyde denaturing agarose gel.

Library preparation. Illumina 2 × 150 pair end libraries were prepared as follows. Briefly, mRNA was enriched from isolated total RNA and fragmented. The fragmented mRNA was used for first-strand cDNA synthesis, followed by second-strand generation, A-tailing and adapter ligation. Adapter ligated products were purified and PCR amplification was carried out. PCR amplified cDNA libraries were assessed for quality and quantity using DNA High Sensitivity Assay Kit (Agilent Technologies).

Quality assessment prior to cluster generation and sequencing. The amplified libraries were analyzed using Bioanalyzer 2100 and High Sensitivity DNA chip (Agilent Technologies). After obtaining the Qubit concentration for each of the libraries, it was loaded on Illumina platform (2 × 150 bp chemistry) for cluster generation and sequencing. Data was generated on Illumina HiSeq. 2500 system and paired-end sequencing allowed the template fragments to be sequenced in both the forward and reverse directions. The library molecules bind to complementary adapter oligos on paired-end flow cell. The adapters were designed to allow selective cleavage of the forward strands after re-synthesis of the reverse strand during sequencing. The copied reverse strand was then used to sequence from the opposite end of the fragment. Total RNA was subjected to pair-end library preparation with Illumina TruSeq Stranded Total RNA Library Preparation Kit. The mean size of the libraries was between 357 bp to 567 bp for the 28 samples. The libraries were sequenced and high quality data was generated for ~ 3.05 GB data per sample (Online-only Table 1).

Sequence analysis. Illumina 2 × 150 pair end libraries were prepared using the Illumina TruSeq stranded mRNA Library Preparation Kit and as per the firm’s protocol (Illumina Inc.). The amplified libraries were analyzed on the Bioanalyzer 2100 with a High Sensitivity DNA chip (Agilent Technologies). The de novo master assembly was generated using “SOAP-deno-Trans (v1.03)” assembler (Short Oligonucleotide Analysis Package)12. For each data set, raw quality was assessed and filtered with Trimmomatic (v.0.36)13. Transcripts were clustered using the CD-HIT (Cluster Database at High Identity with Tolerance) package14. The predicted proteins...
of CDS (Coding sequence) were subjected to similarity search against NCBI’s non-redundant (nr) database using
the BLASTP (Basic Local Alignment Search Tool) algorithm.

**Data Records**
The total raw sequencing data from 28 samples (14 biological replicates, where the sequencing experiment was
performed twice and the replicates are derived from different pool of larvae and they are biologically independ-
ent samples) was used for assembly in the present study. They have been deposited in the NCBI SRA database,
with identifier SRP1867015 and accession numbers SRR8617834, SRR8617835, SRR8617836, SRR8617837,
SRR8617838, SRR8617839, SRR8617840, SRR8617841, SRR8617842, SRR8617843, SRR8617844, SRR8617845,
SRR8617846, SRR8617847, SRR8617848, SRR8617849, SRR8617850, SRR8617851, SRR8617852, SRR8617853,
SRR8617854, SRR8617855, SRR8617856, SRR8617857, SRR8617858, SRR8617859, SRR8617860 and
SRR8617861, under BioProject PRJNA523326 and BioSample SAMN09241884. This Transcriptome Shotgun
Assembly project has been deposited at DDBJ/ENA/GenBank under the accession GHGZ00000000. The ver-
sion described in this paper is the first version, GHGZ01000000.

| **Description** | **Master Assembly** |
|-----------------|---------------------|
| Total number of transcripts | 1,74,066 |
| Total transcriptome length in bps | 100,247,510 |
| Average transcript length in bps | 575 |
| N50 | 421 |
| Maximum transcript length in bps | 25,338 |
| Minimum transcript length in bps | 200 |

| **Metrics** | **Master Assembly** |
|--------------|---------------------|
| Length &gt;=200 & &lt;=300 | 85429 |
| Length &gt;300 & &lt;=400 | 32056 |
| Length &gt;400 & &lt;=500 | 13439 |
| Length &gt;500 & &lt;=600 | 8124 |
| Length &gt;600 & &lt;=700 | 5796 |
| Length &gt;700 & &lt;=800 | 4098 |
| Length &gt;800 & &lt;=900 | 3028 |
| Length &gt;900 & &lt;=1000 | 2444 |
| Length &gt;1000 & &lt;=5000 | 18383 |
| Length &gt;5000 | 1269 |

**Table 1.** Statistics of assembled transcripts and transcript length distribution.

| **Description** | **Unigenes** |
|----------------|-------------|
| Total number of unigenes | 1,36,618 |
| Total size of all unigenes in bps | 86,577,226 |
| Average unigene length in bps | 633 |
| N50 | 458 |
| Maximum unigene length in bps | 25,338 |
| Minimum unigene length in bps | 200 |

| **Metrics** | **Unigenes** |
|--------------|-------------|
| Length &gt;=200 & &lt;=300 | 56403 |
| Length &gt;300 & &lt;=400 | 26874 |
| Length &gt;400 & &lt;=500 | 12354 |
| Length &gt;500 & &lt;=600 | 7740 |
| Length &gt;600 & &lt;=700 | 5647 |
| Length &gt;700 & &lt;=800 | 3980 |
| Length &gt;800 & &lt;=900 | 2941 |
| Length &gt;900 & &lt;=1000 | 2366 |
| Length &gt;1000 & &lt;=5000 | 17193 |
| Length &gt;5000 | 1120 |

**Table 2.** Statistics of unigenes and length distribution.
Technical Validation

SOAPdenovo-Trans assembler was used to generate de novo transcriptome assembly from four experimental sets of midgut samples viz. Group (i) susceptible larvae exposed to medium (water), Group (ii) susceptible larvae exposed to 1/10 of LD₅₀ dosage of DOR Bt-1 formulation, Group (iii) susceptible larvae subjected to starvation and Group (iv) tolerant larvae exposed to LD₅₀ dosage of DOR Bt-1 formulation (reared for 15 generations) (Fig. 1). A total of 1,74,066 transcripts were generated for master assembly with a transcriptome length of 10,02,47,510 bps (base pairs). A total of 1,36,818 unigenes were reported using CD-HIT and 35,559 coding sequences were predicted by Transdecoder. The top-hit species distribution revealed that majority (23%) of the CDS aligned with Spodoptera litura followed by Helicoverpa armigera and Heliothis virescens all of which belong to family Noctuidae in the Lepidoptera order.

Transcriptome assembly. The de novo master assembly of high quality reads of 28 processed samples was accomplished using “SOAP-denovo-Trans (v1.03)” assembler. For each data set, raw quality (phred40) was assessed and filtered with Trimmomatic (v.0.36) using the parameters ILLUMINACLIP:adapter.fasta:2:30:8 MINLEN:40 to remove adaptor sequence and filter by quality score. An average of 19 million clean reads were obtained. Statistics of high quality reads with total reads, base count and data size are summarized in Online-only Table 1 and statistics of assembled transcripts as well as length distribution is presented in Table 1.
Clustering. To filter the redundancy or the noise, it was required to select one representative transcript for transcripts clusters. Transcripts were clustered using CD-HIT (Cluster Database at High Identity with Tolerance) package\(^1^4\). CD-HIT-EST v4.6.1 was used to remove the shorter redundant transcripts when they were 100% covered by other transcripts with more than 90% identity. The non-redundant clustered transcripts were then designated as unigenes (Table 2). CDS were predicted from the unigene sequences with Transdecoder at default parameters which resulted in the identification of 35,559 CDS (Table 3).

Annotation. The predicted proteins of CDS were subjected to similarity search against NCBI’s non-redundant (nr) database using the BLASTP algorithm. Out of total 35,559 proteins, 32,561 proteins were captured with hits and 2,998 with no hits (Annotation of each transcript of the assembled transcriptome)\(^1^7\). The top-hit species distribution revealed that majority of the hits were found to be against the species *Spodoptera litura* followed by *Helicoverpa armigera* and *Heliothis virescens* (Fig. 2). Simultaneously, all protein sequences were searched for similarity against NR, UniProt (Universal Protein Resources), KOG (EuKaryotic Orthologous Groups) and Pfam database using BLASTP with an e-value threshold of 1e\(^{-5}\). The BLAST result of four databases has resulted in Fig. 3.

Differential expression. In this work we compared the control and Cry toxin tolerant larval transcript map reads for the differential expression analysis. Analysis of count data was done using DESeq. 2 in RStudio platform\(^1^8\). Differential expression analysis shows significant differences in the tolerant larval population as compared to the susceptible population (Differential expression analysis)\(^1^7\). Out of 35,559 CDS analysed, 320 CDS show significant variation (padj < 0.05). Few of these genes like (i) gi|1131919362| Ca\(^{2+}\)-binding protein, RTX toxin-related, (ii) gi|1199381583| superoxide dismutase [Cu-Zn] 2-like, (iii) gi|315139350| serine protease 63, (iv) gi|1274141826| trypsin, alkaline C-like and (v) gi|1274136486| apolipophorins isoform X2 were shown to be upregulated, while (i) gi|123995301| ribosomal protein S4, (ii) gi|744619941| predicted: 60 S ribosomal protein L8, (iii) gi|45219787| ribosomal protein S3A, (iv) gi|1344818460| alanine aminotransferase 1-like and (v) gi|501300966| ubiquitin were downregulated.

Code availability

The following software version/script were used in the current manuscript. The RStudio software packages are available open-source from the repository at https://www.rstudio.com/. SOAPdenovo-Trans (v.1.03)\(^1^2\). Trimmomatic (v.0.36)\(^1^3\). CD-HIT-EST (v.4.6.1)\(^1^4\). BlastP (v.2.2.30). The DESeq. 2 scripts were used for plotting the differential expression data. https://bioconductor.org/packages/release/bioc/vignettes/DESeq. 2/html. As an input we have used- (1) a table having RAW read counts and (2) metadata, that is, each line contains description of one of the samples. See example below: #SampleName Condition. C1_raw_read_count control. D1_raw_read_count tolerant.

References

1. Vimala Devi, P. S. & Sudhakar, R. Effectiveness of a local strain of *Bacillus thuringiensis* in the management of castor semilooper, *Achaea janata* on castor (*Ricinus communis*). Indian J Agr Sci 156, 447–449 (2006).
2. Reddy, V. P., Rao, N. N., Devi, P. V., Narasu, M. L. & Kumar, V. D. PCR-based detection of cry genes in local *Bacillus thuringiensis* DOR Bt-1 isolate. Pest Technol. 6, 79–82 (2012).
3. Ningshen, T. J., Chauhan, V. K., Dhania, N. K. & Dutta-Gupta, A. Insecticidal effects of hemocoelic delivery of *Bacillus thuringiensis* Cry toxins in *Achaea janata* larvae. Front Physiol. 8(629), 1–10 (2017).
4. Chauhan, V. K., Dhania, N. K., Chaitanya, R. K., Senthilkumaran, B. & Dutta-Gupta, A. Larval mid-gut responses to sub-lethal dose of cry toxin in lepidopteran pest *Achaea janata*. Front Physiol. 8(662), 1–11 (2017).
5. Dhania, N. K., Chauhan, V. K., Chaitanya, R. K. & Dutta-Gupta, A. Midgut de novo transcriptome analysis and gene expression profiling of Achaea janata larvae exposed with Bacillus thuringiensis (Bt)-based biopesticide formulation. *Comp Biochem Physiol Part D Genomics Proteomics* 30, 81–90 (2019).

6. Tabashnik, B. E., Mota-Sanchez, D., Whalon, M. E., Hollingworth, R. M. & Carrière, Y. Defining terms for proactive management of resistance to Bt crops and pesticides. *J Econ Entomol* 107, 496–507 (2014).

7. Badran, A. H. et al. Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance. *Nature* 533(7601), 58 (2016).

8. Melo, A. L. D. A., Soccol, V. T. & Soccol, C. R. *Bacillus thuringiensis*: mechanism of action, resistance, and new applications: a review. *Crit Rev Biotechnol* 36, 317–326 (2016).

9. Bretschneider, A., Heckel, D. G. & Pauchet, Y. Three toxins, two receptors, one mechanism: Mode of action of Cry1A toxins from *Bacillus thuringiensis in Heliothis virescens*. *Insect Biochem Mol Bio* 76, 109–117 (2016).

10. Jurat-Fuentes, J. L. & Crickmore, N. Specificity determinants for Cry insecticidal proteins: Insights from their mode of action. *J Invert Path* 142, 5–10 (2017).

11. Carrière, Y., Fabrick, J. A. & Tabashnik, B. E. Can pyramids and seed mixtures delay resistance to Bt crops? *Trends Biotechnol.* 34, 291–302 (2016).

12. Xie, Y. et al. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30, 1660–1666 (2014).

13. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120 (2014).

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

15. NCBI Sequence Read Archive, https://identifiers.org/insdc.SRP186750 (2019).

16. GenBank, https://identifiers.org/ncbi/insdc:GHGZ00000000.1 (2019).

17. Dhania, N. K., Chauhan, V. K., Chaitanya, R. K. & Dutta-Gupta, A. RNA-Seq analysis and de novo transcriptome assembly of Cry toxin susceptible and tolerant Achaea janata larvae. figshare, https://doi.org/10.6084/m9.figshare.c.4436612 (2019).

18. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15(12), 550 (2014).

**Acknowledgements**

This research was supported by the grant from the Council of Scientific and Industrial Research (Grant No. 37(1709)/18/EMR-11), UGC-BSR Faculty Fellowship to ADG and UGC research grant to RKC. Financial support in form of senior research fellowship to NKD by UGC and VC by DBT, India are acknowledged. Special thanks to Dr. Vivek, Department of System & Computational Biology, School of Life Sciences in the help provided in DESeq analysis.

**Author Contributions**

N.K.D., V.K.C. and R.K.C. designed the study and A.D.G. contributed to the project coordination. N.K.D. and V.K.C. performed the experiments and collected sample; N.K.D. and RKC analyzed the data and evaluated; N.K.D. wrote the paper which was critically evaluated and edited by R.K.C. and A.D.G. The research funding was procured by R.K.C. and A.D.G. All authors read and approved the manuscript.

**Additional Information**

**Competing Interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

The Creative Commons Public Domain Dedication waiver http://creativecommons.org/publicdomain/zero/1.0/ applies to the metadata files associated with this article.

© The Author(s) 2019