De novo transcriptome assembly of *Premnotrypes vorax* (Coleoptera: Curculionidae)

Luisa-Fernanda Velásquez C.
Universidad Nacional de Colombia Instituto de Biotecnologia
Pablo Emiliano Canton
University of Florida
Alejandro Sánchez-Flores
Universidad Nacional Autonoma de Mexico Instituto de Biotecnologia
Alejandra Bravo
Universidad Nacional Autonoma de Mexico Instituto de Biotecnologia
Jairo Cerón (jacerons@unal.edu.co)
Universidad Nacional de Colombia Instituto de Biotecnologia https://orcid.org/0000-0001-9301-3772

Research note

Keywords: arthropods, pest, potato, RNAseq, de novo transcriptome assembly

Posted Date: January 17th, 2020

DOI: https://doi.org/10.21203/rs.2.21229/v1

Abstract

Objective: *Premnotrypes vorax (P. vorax)* is an insect pest that causes significant losses to potato crops in Colombia. Currently, the insect control is mainly done by using highly toxic chemical insecticides and there are no reports of any commercial biological control strategy against this pest. Hence, the objective of this study was to characterize the insect genetic expression to search for genes that could codify for *Bacillus thuringiensis* Cry toxin receptors. Using an RNA-seq approach, we sequenced the mRNA from the insect tissue, performed a *de novo* assembly and analyzed the reconstructed transcriptome of *P. vorax*. To our knowledge, this is the first genetic report of this endemic insect which will set the basis of a possible biological control strategy.

Results: The transcriptome data was obtained from dissected midgut tissue samples of *P. vorax* larvae. The isolated RNA was isolated and sequenced using the Illumina HiSeq platform with a configuration of 2x150pb reads. A total of 383,552,246 reads were obtained and subsequently a quality and cleaning process was performed through FastQC and Trimmomatic software, respectively. A *novo* assembly was done using the Trinity software, obtaining a transcriptome assembly with 25,631 genes that showed at least one annotation record, resulting in 74,984 transcript isoforms.

Introduction

*Premnotrypes vorax (*Hustache) (Coleoptera: Curculionidae) is one of the 15 species of insects that form part of the complex known as “Andean potato weevil” (Figure 1). *P. vorax* is distributed in South America, principally in Colombia, Ecuador, Venezuela and Peru with registries as one of the most important pests of potato (*Solanum tuberosum*) crops (Pérez et al, 2009). Adult insects feed on the plants and cause damage along the edges of the leaves, but the larvae also make tunnel-shaped lesions in the tubers causing externally visible damage (Pérez-Álvarez et al., 2010). This insect can cause commercial loses up to 80% of damaged tubers or the complete destruction of the potato crop, especially with high larvae populations (ICA, 2011).

Historically, insect control by indigenous farmers involves extended crop rotation, special separation between fields and chemical control with highly toxic insecticides applied at planting. These insecticides usually fail to penetrate the soil, which is where the tuber pest of interest is able to survive and grow (Alcázar and Cisneros, 1998). In Colombia, more than 22 million $US dollars are spent each year for spraying insecticides against *P. vorax* (CIP, 1984; Perez et al., 2018). An additional alternative is the use of specific bioinsecticides such as the insecticidal proteins produced by *Bacillus thuringiensis*, called Cry toxins, which have been used as biological pesticides in different insect orders (Pardo et al., 2013. These Cry insecticidal proteins interact with specific receptors located on the surface of the insect midgut epithelial cells. The toxin inserts into the membrane and forms pores, causing osmotic shock and death of the insect (Bravo et al., 2007). Previously, Cry3 proteins have been reported as active against

Coleoptera insect pests such as *Tenebrio molitor (*Fabrick *et al., 2009)* and *Leptinotarsa decemlineata (*Park *et al., 2009;* van Frankenhuyzen*, 2009)* suggesting that it could be used as bioinsecticide against *P. vorax.* The objective of this work was to make *de novo* transcriptome assembly of *P. vorax* (Coleoptera: Curculionidae) that would set the basis for future development of biological control strategies.

Methods

Total RNA extraction, sequencing library preparation and sequencing

The RNA was extracted from larvae midgut tissue with no biological replicates since pests were collected from the wild. A total of 20 larvae were used where 20 mg of midgut tissue were extracted and the tissue was separated in two Eppendorf tubes. The total RNA was extracted using the kit Agentcourt RNAdvance Cell v2 (Beckman Coulter) and quantified using Nanodrop 2000 and Qubit 2.0 system. The RNA integrity number (RIN) for each sample was calculated using the Bioanalyzer2100 (Agilent) system. All samples presented a RIN >= 7 (8.5 and 8.8), indicating enough quality and integrity for library preparation. Illumina sequencing libraries were prepared using the TruSeq stranded mRNA following the vendor's protocol and obtaining an average library fragment size of 500 bp. RNA-seq libraries were sequenced using the Hiseq platform with a configuration of 300 cycles to generate pair-end reads of 150 bp. General statistics for the sequencing can be found in Table 1.

Preprocess, *de novo* Assembly, Annotation and Mapping

For all samples, the quality of sequencing results was analyzed using the FASTQC software using default parameters for fastq format. Due to the excellent quality, no reads were removed from the dataset and they were assembled using Trinity v2.6.5 (Grabherr et al., 2011) under default parameters. The statistics for the generated transcriptome are presented in Table 1. The resulting transcripts were used to predict open reading frames (ORFs) for the probable protein products using TransDecoder v5.3. 0 0 (Haas et al., 2013) with default parameters. The ORFs were annotated using Trinotate v3.1.1 (Bryant et al., 2017) to integrate the results from blastx and blastp against the Uniprot databases provided with the software; the hmmsearch against the PfamA database; the signal peptide predictions using SignalP v4.1 and the transmembrane domain prediction using tmhmm v2.0c. Reads from each sample were mapped back to the transcriptome to determine the abundance of each transcript and detect those that were tissue-specific.

Completeness Evaluation and Transcriptome filtering

The software BUSCO v3.0.2 (Waterhouse et al., 2017) was used to evaluate the completeness of the assembled transcriptome. In a nutshell, BUSCO performs a blastx search using all transcripts against a database of conserved orthologous proteins from a certain taxonomic clade. For this analysis we used

the arthropoda_odb9 database included with the software using the -m transcriptome parameter. The transcriptome filtering was performed based on the annotation results where transcripts having a transcript or ORF with a blastx, blastp or a Pfam result, were retained. The evaluation was performed on the Trinity filtered version and the results are depicted in Table 1.

Results

We used an RNA-seq approach to characterize the transcriptome of P. vorax (Hustache) (Coleoptera: Curculionidae). Total RNA was processed and sequenced using the Illumina technology and the statistics for the sequencing can be found at Table 1. The quality control evaluation for the sequencing data was performed using FASTQC and the reconstruction and annotation of the transcriptome was achieved with the Trinity and Trinotate pipelines. Raw and processed data were deposited at NCBI public repositories under the accession number PRJNA506951.

Discussion

To our knowledge, this is the first characterization of the gut transcriptome profile of P. vorax (Hustache) (Coleoptera: Curculionidae). We are presenting a de novo assembly and annotation of the transcriptome achieved with the Trinity and Trinotate pipelines, respectively. All raw and processed data have been deposited at NCBI public repositories (PRJNA506951) for public accession to be used for different purposes. There is an urgent need for development of effective control strategies since this pest is a severe problem in Colombia where currently it occupies the rank 36th out of 183 countries that produce potatoes worldwide with 60 varieties. Since potato is the third most important crop and the average consumption in the country is 60 Kg per person per year (Fedepapa, 2018) an effective, ecofriendly and human harmless pest control strategy is needed. This is the first transcriptome data of this insect pest that could contribute to the understanding of the genetic expression of this organism and guide the research towards finding targets for pest control.

Limitations

The data presented here is a first approach to this species. Annotation quality is dependent on closely related species, which currently are few. Additional sequencing for replicates of midgut tissue and other organs will help refine bona fide transcript isoforms. Experimental work being performed at the moment will also help validate the expression of the transcripts of interest for biological control.

Abbreviations

P. vorax Premnotrypes vorax

Declarations

Ethics approval and consent to participate

No ethical concerns in this work. All authors reviewed the final document and accepted to participate. This work complies with Colombian legal requirements: Genetic resources access contract (No. 121 January 22, 2016) and Colombian biological diversity export permit (No. 06470 November 8, 2019)

Consent to publish

All authors accepted to publication of these data

Availability of data and material

The datasets generated and analyzed in this study can be found in the Bioproject number **PRJNA**506951. https://www.ncbi.nlm.nih.gov/

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This work was supported by the Institute of Biotechnology from National University of Colombia (UNAL). Was funded by the Project 35953 (UNAL) and the Colciencias scholarship 2014 (Resolution No. 038).

Author's contributions Authors' information

LFV collected the samples and perform experiments and analyzed RNA-Seq data, PEC analyzed RNA-Seq data, ASF designed the experiments, AB and JC coordinated the research and wrote the manuscript.

Acknowledgements

We thank Jerome Verleyen, and Veronica Jimenez-Jacinto for technical bioinformatic support and access to the high-performance computer cluster from the “Unidad de Secuenciación Masiva y Bioinformática” of the “Laboratorio Nacional de Apoyo Tecnológico a las Ciencias Genómicas” CONACyT #260481.

References

1. Alcazar, J., Cisneros, F. 1998. Taxonomy and bionomics of the Andean potato weevil complex: *Premnotrypes* and related genera.
2. Bravo, A. Gill, S. Sorberón, M. 2007. Mode of action of *Bacillus thuringiensis* Cry and Cyt toxins an their potential for insect control. Toxicon 49, 423-435
3. Bryant, DM., Johnson, K., DiTommaso, T., Tickle, T., Couger, MB., Payzin-Dogru, D,. Lee, TJ., Leigh, ND., Kuo, TH., Davis, FG., Bateman, J., Bryant, S., Guzikowski, AR., Tsai, SL., Coyne, S., Ye, WW., Freeman, RM Jr., Peshkin, L., Tabinm, CJ., Regev, A., Haas, BJ., Whited, JL. 2017. A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. Cell Rep. 18, 762-776. doi: 10.1016/j.celrep.2016.12.063.
4. Fabrick, J., Oppert, C., Lorenzen, M.D., Morris, K., Oppert, B., Jurat-Fuentes, J.L. 2009. A novel *Tenebrio molitor* cadherin is a functional receptor for *Bacillus thuringiensis* Cry3Aa toxin. J Biol Chem. 284, 18401-18410.
5. 2018. ¿Cuánta papa se consume en Colombia? Revista Papa No. 43. Federación Colombiana de Productores de Papa. Bogotá, Colombia. p. 20. http://www.fedepapa.com/wp-content/uploads/2017/01/REVISTA-43-OK.pdf
6. Gomez, S., Constanza- Mateus, A., Hernandez, J., Zimmermann, B. 2000. Recombinant Cry3Aa has insecticidal activity against the andean potato weevil, *Premnotrypes vorax*. Biochem Biophysl Res Com 279, 653–656.
7. Grabherr, MG., Haas, BJ., Yassour, M., Levin, JZ., Thompson, DA., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29, 644-652.

8. Haas, BJ., Papanicolaou, A., Yassou, M., Grabherr, M., Blood, PD., Bowden, J., Couger, MB., Eccles, D., Li, B., Lieber, M., Macmanes, MD., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, CN., Henschel, R., Leduc, RD., Friedman, N., Regev, A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 8, 1494-512.
9. 2011. Manejo fitosanitario del cultivo de la papa (*Solanum tuberosum subsp. andigena y S. phureja*). Medidas para la temporada invernal. Bogotá D.C. Colombia. https://www.ica.gov.co/getattachment/b2645c33-d4b4-4d9d-84ac-197c55e7d3d0/Manejo-fitosanitario-del-cultiva-de-la-papa-nbsp;-.aspx
10. CIP International Potato Center, 1984 Andean potato weevil (*Premnotrypexs vorax*). Plantwise Technical Factseet. Rep. 1997-98, 141-151 https://www.plantwise.org/KnowledgeBank/Datasheet.aspx?dsid=44044
11. Pardo-López, L., Soberón, M., and Bravo, A. 2013. *Bacillus thuringiensis* insecticidal toxins: Mode of action, insect resistance and consequences for crop protection. FEMS Microbiol Rev 37, 3-22
12. Park, Y., Amir, M., Abdullah, F., Taylor, M., Rahman, K., and Adang, M. 2009. Enhancement of *Bacillus thuringiensis* Cry3Aa and Cry3Bb toxicities to Coleopteran larvae by a toxin-binding fragment of an insect Cadherin. Appl Environ Microbiol. 75,3086–3092.
13. Pérez-Álvarez, R., Garza, J., Argüelles-Cárdenas, J., Garramuño, E. 2010. Distribución espacial de *Premnotrypes vorax* (Hustache) (Coleoptera: Curculionidae) en cultivos de papa. Revista Corpoica CienTec Agrop. 11, 11-20.
14. Pérez, R., Garza, J., Argüelles-Cárdenas, J. 2009. Método de cría en laboratorio del gusano blanco de la papa *Premnotrypes vorax* (Coleoptera: Curculionidae). Revista Corpoica – Cien Tec Agrop 1, 16-23.
15. Perez, W., Valverde, M., Barreto, M., Andrade-Piedra, J., Forbes, G. Replication data for: Pest and diseases affecting potato landraces and breb varieties grow in Peru under indigenous farming system. doi/10.21223/P3/B6ACYT International Potato Center
16. Van Frankenhuyzen, K. 2009. Insecticidal activity of *Bacillus thuringiensis* crystal proteins. J Invertebr Pathol. 101, 1-16.
17. Waterhouse, R. Seppey, M. Simão, FA. Manni, M. Ioannidis, P. Klioutchnikov, G. Kriventseva, EV. Zdobnov, EM. 2017. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol Biol Evol. Mar; 35(3): 543–548.

Table 1

Table 1. General statistics for the sequencing, Trinity assembly, annotation, filtering and completeness evaluation.

Basic statistics	
Sequencing yield (paired reads/Gbases)	383,552,246 / 115.83
Trinity predicted genes	232,969
Trinty reconstructed transcript isoforms	517,235
Total bases of reconstructed transcript isoforms	390.49 Mbases
N50	566 bases
Median transcript length	298 bases
Average transcript length	477.80 bases
Predicted ORFs	870,310
BUSCO completeness (%)	94.1
Filtering and completeness evaluation	
Trinity genes with at least one annotation record	25,631
Trinity transcript isoforms with at least one annotation record	74,984
Total bases of reconstructed transcript isoforms	76.82 Mbases
N50	1,508 bases
Median transcript length	728 bases
Average transcript length	978 bases

Figures

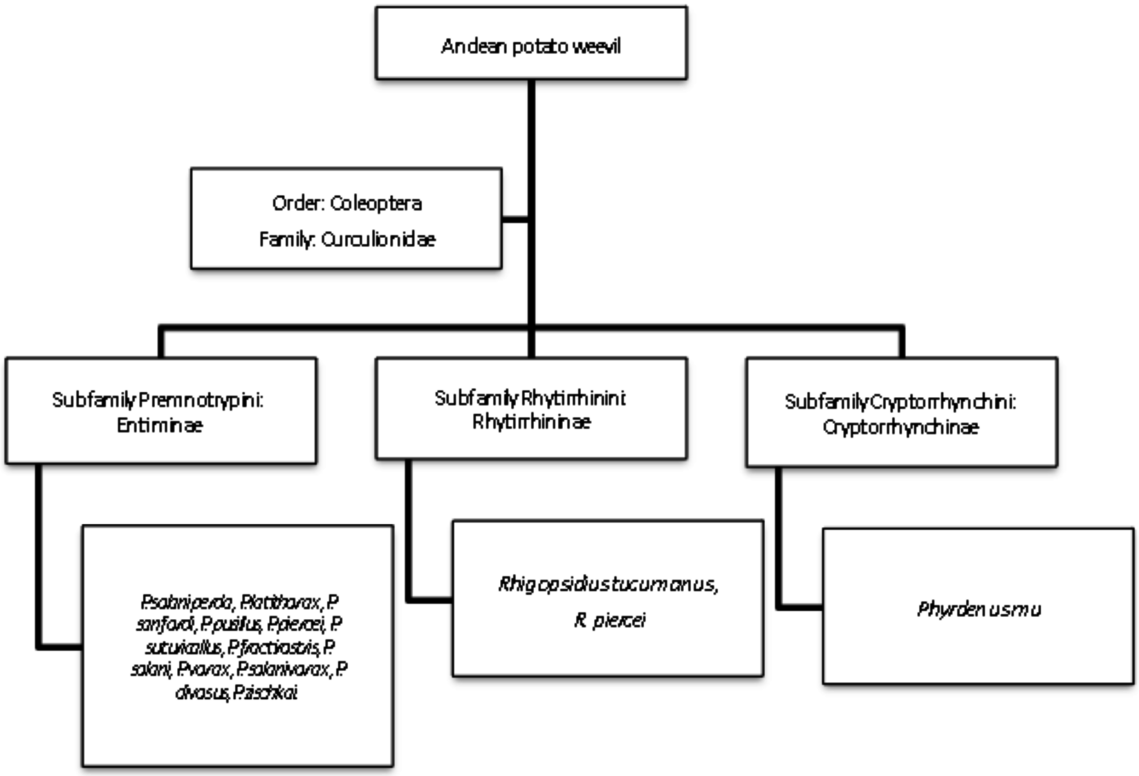

Figure 1

Taxonomy of “Andean potato weevil” complex (Kühne, 2007)