IMPROVING THE BOLL WEEVIL MOLECULAR TOOLBOX: A BETTER TRANSCRIPTOME FOR THE ADVANCEMENT OF GENE-LEVEL CONTROL

L. Perkin
C.P.-C. Suh
USDA-ARS ICCDRU
College Station, TX

Abstract

The boll weevil is a destructive pest of commercial cotton throughout the Americas. An eradication program in the United States has removed the majority of boll weevil from its range. However, weevil populations in south Texas remain a threat to eradicated areas. Monitoring and prevention techniques include boll weevil pheromone traps and cultural control practices such as delayed planting and early stalk destruction. However, eradication programs rely primarily on the pesticide malathion for control. The goal of this project is to improve the current molecular tools available for the boll weevil so that new control methods can be explored that are specific to the boll weevil and are environmentally friendly. We sequenced the major life stages of the boll weevil and assembled an improved transcriptome. We then used that transcriptome in an RNA-seq experiment to identify genes involved in male pheromone production.

Introduction

The boll weevil, *Anthonomus grandis grandis*, is a pest of commercial cotton throughout the Americas. The United States has had a very successful eradication program that eliminated the boll weevil from most of its range. Weevil populations remaining in south Texas continue to pose a threat to other cotton growing areas. The eradication program relies on pheromone traps for monitoring boll weevil populations and malathion to treat areas with boll weevil activity (Texas Boll Weevil Eradication Foundation, 2021). In order to develop gene-level control methods, foundational molecular resources for the boll weevil are needed. One such foundational resource is an updated transcriptome. The currently available transcriptome was assembled almost a decade ago and while a good resource of its time, an update is needed because this transcriptome only accounts for approximately 50% of boll weevil genes (Firmino et al., 2013).

Here we present an updated boll weevil transcriptome that was created using next-generation RNA sequencing and the Trinity assembler (Grabherr et al., 2011). The new transcriptome is representative of 98% of the genes in the genome. We then show how this improved transcriptome can aid in the advancement of gene-level control tools by using RNA-seq analysis to identify genes involved in pheromone production in male boll weevils.

Materials and Methods

Boll Weevil Collections and Experimental Design

Boll weevils were collected as larvae in infested squares from a field northwest of Edinburg, TX. The major life stages (early instar larvae, late instar larvae, pupae, adult male and adult female) were collected, placed in RNAlater (Invitrogen, Carlsbad, CA), and stored at -20°C until total RNA was extracted. A collection of pupae was kept on vermiculite in a 29.4 ± 1°C degree incubator to obtain adults for the pheromone experiment. Newly enclosed adults (<24 h-old) were sexed and only males were kept. Forty males were divided into two groups. One group was held without food for 24 hours. The other group was fed squares (5-8 mm in diameter) daily for 6-8 days to promote pheromone production. Weevils in both groups were tested for pheromone production using gas chromatography mass spectrometry using procedures described by Suh and Spurgeon (2016). Weevils in each group that expressed the correct phenotype were placed in a 2 ml snap cap tubes with RNAlater and stored at -20°C until nucleic acid extraction.

RNA extraction and sequencing

RNA was extracted from weevils for each life stage and pheromone experiment using the RNeasy extraction kit (Qiagen, Hilden, Germany). Samples were checked for quantity and quality using a Tapestation 4200 (Agilent, Santa Clara, CA). Total RNA from 40 life stage samples (2, 2nd instar larvae; 8, 3rd instar larvae; 10 pupae; 10 adult females; and 10 adult males) and 20 pheromone samples (10 pheromone producing males and 10 males with no pheromone production) were taken to Texas A&M AgriLife Genomics and Bioinformatics Service (TxGen) for purification, library preparation, and sequencing on an Illumina NovaSeq (Illumina, San Diego, CA).

Transcriptome assembly and annotation

A total of 20 million reads per sample of data were downloaded from TxGen and used to assemble a *de novo* transcriptome using the TRINITY assembler (Grabherr et al. 2011). Default parameters were used including the option for pair-wise reads, non-strand specific assembly, and minimum codon length of 200 bp. Completeness of the transcriptome was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO; Seppey et al., 2019) selected from OrthoDB (Kriventseva et al., 2010). The transcriptome was annotated using BLAST via OmicsBox (Valencia, Spain). Parameters were limited to the arthropod database and E-value ≤ $e10^{-5}$.

RNA-seq analysis

RNA-seq analysis was accomplished using edgeR as part of the OmicsBox software (Robinson et al., 2010). Workflow included mapping the reads to the *de novo* transcriptome described above and pairwise differential expression analysis, which included quantifying the mapped reads at each locus and statistically assessing fold change (FDR < 0.05) differences between pheromone producing and non-pheromone producing weevils.

Results and Discussion

Transcriptome

The final transcriptome was compiled using over 1 million transcripts with an average length of 766 bp. The BUSCO analysis revealed the transcriptome represented 96% of all genes present in the insect genome database. The BLAST annotation revealed the majority of sequences matched two weevil species, *Dendroctonus ponderosae* and *Sitophilus oryzae,* followed by the tenebrionid, *Tribolium castaneum*. Nearly all sequences in the dataset matched other Coleopteran (Figure 1).

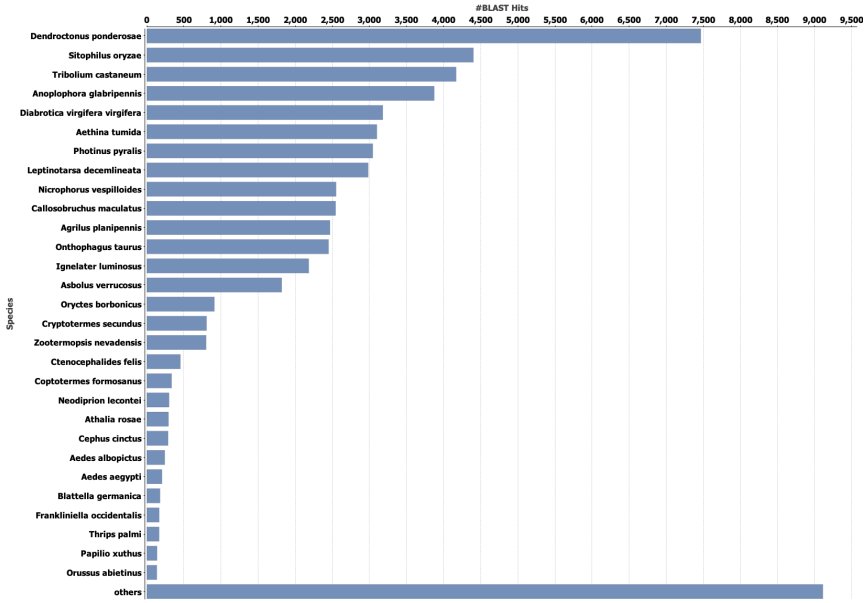

Figure 1. Top species BLAST hit distribution for the boll weevil.

RNA-seq analysis

RNA-seq analysis showed a total of 192 differentially expressed genes between the males that produced pheromone compared to those that did not produce pheromone. Of these 192 genes, 63 genes were up-regulated and 129 genes were down-regulated (Figure 2). The top 50 differentially expressed genes showed clear separation between the two treatment groups (Figure 3). A gene enrichment analysis using Fisher's Exact Test indicated that genes involved in energy metabolism, fatty acid metabolism and terpene biosynthesis were enriched in our dataset. These genes are key components in the production of boll weevil (Tumlinson et al., 1969) and other insect pheromones (Blomquist et al., 2020).

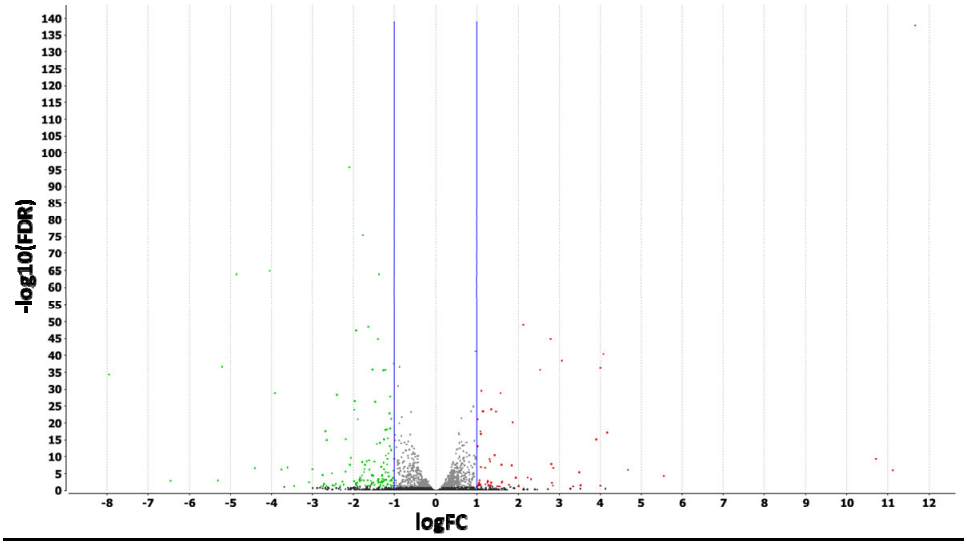

Figure. 2. Volcano plot of all differentially expressed genes in the boll weevil when comparing pheromone producing males to the non-pheromone producing males. Red dots indicate significantly up-regulated genes and green dots represent significantly down-regulated genes (FDR < 0.05).

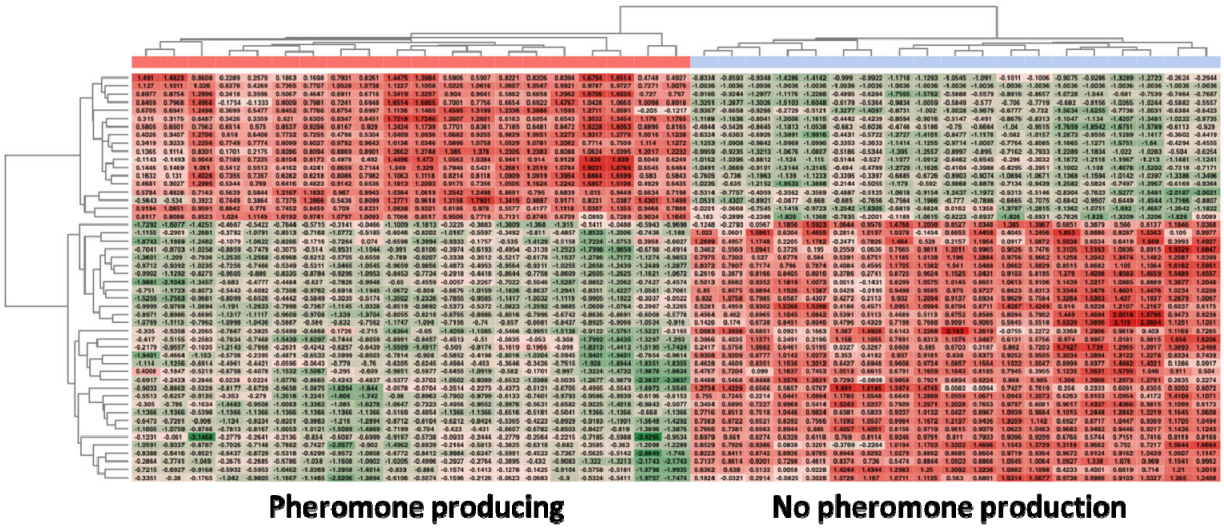

Figure. 3. Heatmap depicting the up (red) and down (green) differentially expressed genes between the pheromone producing and non-pheromone producing boll weevil groups. Numbers within each tile indicate fold change values between groups.

Summary

We provide an updated boll weevil *de novo* transcriptome that shows a completeness BUSCO score of 96%. We then used this assembly to perform an RNA-seq analysis comparing gene expression between pheromone producing and non-pheromone producing weevils. We ultimately identified 50 differentially expressed genes that likely have a key function in pheromone metabolic pathways. This foundational transcriptome will be a building block for production of gene-level control strategies for the boll weevil.

Acknowledgements

We thank Jason Carlson for collecting infested squares and Jose and Mercedes Perez for boll weevil rearing and GC-MC analysis. This research was funded through Cotton Inc. (Project # 20-216). Mention of trade names or commercial products does not imply recommendation or endorsement by the USDA. USDA is an equal opportunity provider and employer.

References

Blomquist, G.J., C.T. Tittiger, and R. Jurenka. 2020. Circular hydrocarbons and pheromones of arthropods. p. 213-244. *In* H. Wilkes (ed.), Hydrocarbons, oils, and lipids: Diversity, origin, chemistry, and fate, handbook of hydrocarbon and lipid microbiology, Springer Nature Switzerland.

Firmino, A.A.P., F.C. de Assis Fonseca, L.L.P de Macedo, R.R. Coelho, J.D.A. de Souza Jr, R.C. Togawa, O.B. Silva-Junior, G.J. Pappas Jr, M.C.M. da Silva, G. Engler, M.F. Grossi-de-Sa. 2013. Transcriptome analysis in cotton boll weevil (*Anthonomus grandis*) and RNA interference in insect pests. PLOS ONE. 8:e85079.

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotech, 29:644-52.

Kriventseva, E.V., D. Kuznetsov, F. Tegenfeldt, M. Manni, R. Dias, F.A. Simao, and E.M. Zdobnov. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic acids res, 47: D807-D811.

Robinson, M.D., D.J. McCarthy, and G.K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:1.

Seppey, M., M. Manni, and E.M. Zdobnov. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods in molecular biology. 1962: 227-245.

Suh, C.P.-C. and D.W. Spurgeon. 2016. Continued release by boll weevils (Coleoptera: Curculionidae) following host removal. J. Entomol. Sci. 51: 332-335.

Texas Boll Weevil Eradication Foundation. 2021. www.txbollweevil.org.

Tumlinson, J.H., D.D. Hardee, R.C. Geuldner, A.C. Thompson, P.A. Hedin, and J.P. Minyard. 1969. Sex pheromones produced by male boll weevil: Isolation, identification, and synthesis. Science. 166: 1010-1012.