ABSTRACT

LAWRIE, ROGER. Pesticide Effects on Gene Expression: From Humans to Insects.
(Under the direction of Dr. R Michael Roe).

Human exposure to environmental chemicals both individually and in combination occurs frequently world-wide most often with unknown consequences. In this study, we examined the impact of two environmental chemicals used in and around homes, the insect repellent DEET (N,N-diethyl-m-toluamide) and the phenylpyrazole insecticide fipronil (fluocyanobenpyrazole) on transcript levels of enzymes potentially involved in xenobiotic metabolism and on long non-coding RNAs (lncRNAs). Primary human hepatocytes were treated with these two chemicals both individually and in combination. Using RNA-Seq, we found that 10 major enzyme categories involved in phase 1 and phase 2 xenobiotic metabolism were significantly ($\alpha = 0.05$) up- and down-regulated (i.e., 100 μM DEET-19 transcripts, 89% up and 11% down; 10 μM fipronil-52 transcripts, 53% up and 47% down; and 100 μM DEET +10 μM fipronil-69 transcripts, 43% up and 57% down). The altered genes were then mapped to the human genome and their proximity to lncRNAs examined. Unique proximities were discovered between altered lncRNA and altered P450s (CYP) and other enzymes (DEET, 2 CYP; Fipronil, 6 CYP and 15 other; and DEET + fipronil, 7 CYP and 21 other). Many of the altered P450 transcripts were in multiple clusters in the genome with proximal altered lncRNAs, suggesting a regulator function for the lncRNA. At the gene level there was high percent identity for lncRNAs near P450 clusters, but this relationship was not found at the transcript level. The role of these altered lncRNAs associated with xenobiotic induction, human diseases and chemical mixtures is discussed.

Several different agricultural insect pests have developed field resistance to Bt (*Bacillus thuringiensis*) proteins (ex. Cry1Ac, Cry1F, etc.) expressed in crops, including corn and cotton. In the bollworm, *Helicoverpa zea*, resistance levels are increasing annually. A common method to analyze global differences in gene expression is RNA-seq. This technique was used to measure differences in global gene expression between a Bt-susceptible and Bt-resistant strain of the bollworm, where the differences in susceptibility to Cry1Ac insecticidal proteins were 100-fold in resistant bollworms. We found expected gene expression differences based on our current understanding of the Bt mode of action, including increased expression of proteases and reduced expression of Bt-interacting receptors in resistant bollworms. We also found expression differences for transcripts that were not previously investigated, i.e., transcripts from three immune pathways-Jak/STAT, Toll, and IMD. Immune pathway receptors (ex. PGRPs) and the IMD pathway demonstrated the highest differences in expression. Our analysis suggested that multiple mechanisms are involved in the development of Bt-resistance, including potentially unrecognized pathways.

Non-protein coding (lncRNAs) mechanisms of resistance are also understudied, the same RNA-seq dataset was used to investigate the long non-coding RNAs (lncRNAs) in Bt-resistance. Significant differential expression of lncRNAs was found in the Cry1Ac-resistant strain (58 up- and 24 down-regulated gene transcripts with an additional 10 found only in resistant and 4 only in susceptible caterpillars). These lncRNAs were examined as potential pseudogenes and for their genomic proximity to coding genes, both of which can be indicative of regulatory relationships between a lncRNA and gene expression. A possible cadherin interacting lncRNA was found. Also, lncRNAs were found proximal to a serine protease, ABC transporter, and cytochrome P450 (CYP) coding genes, potentially involved in the mechanism of Bt and/or chemical insecticide resistance. Characterization of non-coding genetic mechanisms in *Helicoverpa zea* will improve the understanding of the genomic evolution of insect resistance, improve the identification of specific

regulators of coding genes in general (some could be important in resistance), and is the first step for potentially targeting these regulators for pest control and resistance management (using approaches like RNAi). More research is needed to study insect lncRNAs and their role in insect pest biology and resistance.

Pesticide Effects on Gene Expression: From Humans to Insects

by
Roger Duval Lawrie

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctorate of Philosophy

Entomology

Raleigh, North Carolina

2021

APPROVED BY:

_____	_____
R. Michael Roe Committee Chair	Seth Kullman
_____	_____
Greg Cope	Clyde Sorenson

_____
Marce Lorenzen

DEDICATION

I would like to dedicate this thesis to all of my mentors, family, and friends both past and present who have encouraged me to continue pursuing science and to achieve all that I have in science and in life. It wouldn't have been possible without you.

BIOGRAPHY

Roger D. Lawrie grew up in Asheville, NC where he initially found his love for nature and things outdoors. He completed his Bachelors degree in Cell and Molecular Biology in 2015 at Appalachian State University. During his time there he pursued undergraduate research for 3 years and particularly enjoyed invertebrate zoology courses. After pursuing field research and wilderness guiding for two years, he started graduate school in 2017 in the Toxicology program. At the encouragement of his mentor Dr Michael Roe, he changed from a Master's student to a PhD student and also added Entomology as a second major. Since this time he has been involved in many research and teaching opportunities beyond those presented in this thesis including introductory biology labs, repellent bioassays, insect-microbial studies, and others. Following his PhD Roger seeks to continue to be involved in interesting and impactful research in the areas of conservation, sustainability in agriculture, toxicology, entomology, and innovative applied sciences.

ACKNOWLEDGMENTS

I would like to acknowledge the other members past and present of the Roe lab that assisted or guided these research projects and others throughout the PhD process. Michael Roe, Loganathan Ponnusamy, Jean Marcel Deguenon, Bobby Mitchell III, Grayson Cave, Kaiying Chen, Nick Travanty, Jiwei Zhu, and Anirudh Dhammi. I would also like to acknowledge Elizabeth Scholl, formerly of NCSU Bioinformatics core who facilitated bioinformatics and gave assistance with these chapters. I would also like to acknowledge the lab of Dr. Dominic Reisig and Dan Mott for their collaborations with the field-related portions of these projects.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Role of long non-coding RNA in DEET- and fipronil-mediated alteration of transcripts associated with Phase I and Phase II xenobiotic metabolism in human primary hepatocytes

Abstract

Human exposure to environmental chemicals both individually and in combination occurs frequently world-wide most often with unknown consequences. Use of molecular approaches to aide in the assessment of risk involved in chemical exposure is a growing field in toxicology. In this study, we examined the impact of two environmental chemicals used in and around homes, the insect repellent DEET (N,N-diethyl-m-toluamide) and the phenylpyrazole insecticide fipronil (fluocyanobenpyrazole) on transcript levels of enzymes potentially involved in xenobiotic metabolism and on long non-coding RNAs (lncRNAs). Primary human hepatocytes were treated with these two chemicals both individually and in combination. Using RNA-Seq, we found that 10 major enzyme categories involved in phase 1 and phase 2 xenobiotic metabolism were significantly ($\alpha = 0.05$) up- and down-regulated (i.e., 100 μM DEET–19 transcripts, 89% up and 11% down; 10 μM fipronil–52 transcripts, 53% up and 47% down; and 100 μM DEET +10 μM fipronil–69 transcripts, 43% up and 57% down). The altered genes were then mapped to the human genome and their proximity (within 1,000,000 bp) to lncRNAs examined. Unique proximities were discovered between altered lncRNA and altered P450s (CYP) and other enzymes (DEET, 2 CYP; Fipronil, 6 CYP and 15 other; and DEET + fipronil, 7 CYP and 21 other). Many of the altered P450 transcripts were in multiple clusters in the genome with proximal altered lncRNAs, suggesting a regulator function for the lncRNA. At the gene level there was high percent identity for lncRNAs near P450 clusters, but this relationship was not found at the transcript level. The role of these altered lncRNAs associated with xenobiotic induction, human diseases and chemical mixtures is discussed.

Introduction

Human exposure to pesticides and other environmental chemicals is widespread, potentially occurring in the home and as part of regular outdoor activities. There are some occupations where acute or chronic exposure to chemicals are frequent and at high levels, e.g., agricultural workers, foresters, veterinarians, and military personnel. The impact of this exposure at the global molecular level especially for pesticides is poorly understood, and in most cases environmental chemical exposure is assessed one chemical at a time. Also, most often risk assessment is conducted using animal models. To change this paradigm, there has been a growing call for the use of human cell lines and global molecular approaches like RNAseq [1-7]. The only approach to directly assess human variations in risk is the use of primary human cells, also understudied. In this paper, we used primary human hepatocyte exposure to the insect repellent DEET (N,N-diethyl-m-toluamide) and the phenylpyrazole insecticide fipronil (fluocyanobenpyrazole) (separately and as a mixture) as a model system to examine global treatment effects on enzymes potentially involved in xenobiotic metabolism and possible epigenetic effects (using RNAseq). DEET is used on skin and clothing as a repellent for protection against arthropod biting from mosquitoes, fleas, chiggers, ticks and other insects. DEET was first developed for the U.S. military during World War II (1946) and then made available to consumers in 1957 [8]. In commercial products, DEET concentrations range from 5 to 100% [9-10]. Veltri et al. (1994) found that one-third of the U.S. population is exposed to DEET annually [11]. DEET can be detected in drinking water, freshwater, seawater, and

groundwater [12]. DEET is considered safe. However, there are some isolated cases of toxic effects including seizures, cardiovascular toxicity, and Gulf War Syndrome (fatigue, headaches, and dermatitis) [10, 13-14]. Simultaneous application of DEET and oxybenzone-containing sunscreens was shown to affect DEET dermal absorption [15]. The combined exposure of DEET with the insecticide permethrin resulted in biochemical, behavioral, and metabolic changes [16]. Following dermal application, DEET is absorbed into the blood stream (5.61–8.33% for 15% DEET) and transported to the liver, the primary site of DEET metabolism [17-18].

The insecticide fipronil is used in urban settings in over 70 countries and includes topical application to companion animals. Fipronil is also registered for over 100 different crops [19-20]. Fipronil is a $GABA_A$ agonist and elicits uncontrolled neuron firing in the insect central nervous system causing death. Fipronil is known to be highly toxic to aquatic organisms including fish and invertebrates as well as lizards and gallinaceous birds with moderate toxicity to mammals [21]. Water and soil contamination by fipronil have been recorded in several states in the U.S [12]. Chronic exposure to fipronil in rats can induce thyroid cancer, affect fertility, and cause offspring to experience developmental delays and increased mortality [22]. The Environmental Protection Agency (EPA) has approved fipronil for household use on cats and dogs for flea and tick control and in the soil to control termites around homes. The EPA reports that acute fipronil exposure can elicit headaches, dizziness, vomiting, and seizures [12]. Because of the potential (but not documented) simultaneous human exposure to both fipronil and DEET (since both are used in the urban environment), risk assessment of these separately and as a mixture are worthwhile. More importantly, the study of chemical mixtures and especially with pesticide mixtures on primary human cells is under-studied. The primary site of xenobiotic metabolism and the focus of the research here is the human liver [23] and the enzymes, P450s, glutathione-S-transferases (GSTs), sulfotransferases, flavin-containing mono-oxygenases (FMO's), alcohol dehydrogenases, aldehyde dehydrogenases, epoxide hydrolases, glucuronidases, and N-acetyltransferases [24].

Long non-coding RNAs (lncRNA) are strands of RNA over 200 nucleotides in length that lack the capacity to code for proteins [25]. lncRNAs also share similarities with mRNAs, i.e., a 5′ cap with some lncRNAs having a 3′ poly-A tail and similar folding characteristics [26]. Formerly thought of as “junk” DNA, lncRNAs have a variety of regulatory functions including cis or trans gene expression, transcription factor activation, chromatin remodeling, imprinting, and enhancer regulation [25]. lncRNAs are also known to be involved in broader genetic processes such as post transcriptional regulation, protein trafficking, and messenger RNA (mRNA) processing [25]. Genetic proximity of lncRNAs to coding genes (like those for xenobiotic enzymes considered in this paper) is often indicative of cis regulatory mechanisms [25,27]. In a number of different studies, the alteration or disruption of specific lncRNAs was linked to human diseases including Alzheimer's, autoimmune and cardiovascular disease, and cancer [28-31]. Some lncRNAs are altered in mammals after exposure to benzene, cigarette smoke, cadmium, phenobarbital, and arsenic [32-35]. The impact of insecticide exposure on lncRNAs and a study of the association of lncRNAs with xenobiotic enzymes in primary human cells outside our laboratory group has not been considered.

LncRNAs have been established to be important regulators of gene expression and transcription. Regulation via lncRNAs can come from a variety of mechanisms: recruitment of

transcription factors, chromatin remodeling to increase or decrease access to DNA wrapped around histone proteins, as enhancers of genes and other regulatory mechanisms [36-37]. It has also been discovered that dysfunction of lncRNAs is involved in several different medical conditions including cancer. Increased expression of lncRNAs has been linked to drug resistance genes in tumor cells in addition to cardiovascular disease, autoimmune disease, and Alzheimer's disease [25, 28, 29, 30, 38]. Recent literature has also found that lncRNA regulation is impacted by exposure to xenobiotics including polycyclic aromatic hydrocarbons, benzene, cadmium, and phthalates among others [25, 32, 33, 34, 35].

The objective of this study was to examine the impact of DEET and fipronil, both individually and in combination, on gene expression of transcripts potentially involved in xenobiotic metabolism and to determine if there is a possible regulatory role for cis-proximal lncRNAs. Li and Cui (2018) recently suggested that lncRNAs had a regulatory role on protein-coding genes involved in xenobiotic metabolism [39]. The study here is using primary human hepatocytes. Additionally, the mixture of DEET and fipronil will aide in assessing the consequences of chemical combinations on the human liver. This study is a continuation of the work of Mitchell et al., 2016, Mitchell et al., 2017 who characterized the effect of DEET, fipronil, and both in combination on overall changes in transcript levels for coding genes and lncRNAs, respectively [40,41].This paper focuses specifically on the impacts of these chemical exposures on metabolizing genes and these genes' interactions with specific non-coding elements.

Materials and Methods

Cell culture and treatments

Primary human hepatocytes from the liver of a Caucasian female of the age of 62, body mass index (BMI) of 26.3, and no history of smoking or alcohol consumption were harvested within 24 h of the donor's death. It should be noted that human hepatocytes acquired from donors with a BMI about 27.5 show moderate steatosis, a BMI below this amount show little or no steatosis, the condition of which may affect molecular studies [42]. These hepatocytes were delivered to our laboratory in William's E Medium in 12-well culture plates that were coated with a Type 1 Collagen substratum and a Geltrex® overlay at the density of 0.67×10^6 hepatocytes per well. Once the samples were received, we replaced the old William's E Medium with fresh, sterile medium containing 0.292 g/L L-Glutamine (cat. no. W1878, Sigma-Aldrich, St. Louis, MO) which was supplemented with: (1) 10% (v/v) fetal bovine serum (cat. no. S11050, Atlanta Biologicals, Norcross, GA); (2) 10^{-7} M dexamethasone (cat. no. P0500, Steraloids, Inc., Newport, RI); (3) insulin-transferrin-selenium A formulated with 0.17 mM insulin, 0.0069 mM transferrin, 0.0039 mM sodium selenite, and 100.0 mM sodium pyruvate (cat. no. 51300–044, Invitrogen, Carlsbad, CA); (4) a penicillin/streptomycin/amphotericin B solution containing 10,000 units/mL of penicillin, 10,000 µg/mL of streptomycin, and 25 µg/mL of FungizoneR antimycotic (cat. no. 15240–062, Invitrogen, Carlsbad, CA); and (5) a sterile gentamicin/glutamate solution containing 200 mM L-glutamine and 5 mg/mL gentamicinin 0.9% sodium chloride (cat. no. G9654, Sigma-Aldrich, St. Louis, MO). Following this, our plates were incubated at a relative humidity of 95% and 5% CO_2:95% air at 37 °C for 24 h. After this incubation, the William's E Medium was changed for fresh medium and an additional 24 h of incubation was carried out in order to observe viability and quality of hepatocytes.

After 48 h in the lab, treatment with DEET and fipronil began. Contained within one plate, three wells containing primary human hepatocytes were subsequently inoculated with DEET (purity>98%; cat. no. F2284, Chem Service, West Chester, PA) at a final concentration of 100 μM per well; an additional 3 wells were inoculated with fipronil (purity >98%; cat. no. PS2136; Chem Service, West Chester, PA) at a final concentration of 10 μM per well; and lastly, three more wells were treated with a pre-mixed combination of 100 μM DEET and 10 μM fipronil. DEET and fipronil were acquired two weeks prior to the hepatocytes and kept at room temperature until the treatment was conducted. Each chemical treatment was delivered dissolved (wt/vol) in DMSO (dimethyl sulfoxide; > 99.7% pure; cat. no. BP231–100, Fisher Scientific International, Hampton, NH). A DMSO concentration of 0.1% was maintained for each treatment. DMSO has been shown to exhibit minimal cytotoxicity or gene expression alteration in cultured hepatocytes at that concentration [43]. We determined the ideal dose of DEET and fipronil from dose-response data of Das et al. (2006) and Das et al. (2008) [44, 45]. For both DEET (dermal LD_{50} of 5000 mg/kg in rats) and fipronil (dermal LD_{50} of >2000 mg/kg in rats), these doses were chosen to maximize expression and levels of activity for multiple P450s in primary human hepatocytes as well as immortalized HepG2 liver cells with minimal cytotoxic effects [21, 46]. Additionally, this dose in past experiments by our group produced maximum effects on P450 expression/activity without cytotoxicity [44, 45]. Experimental assay conditions here for primary human hepatocytes were identical to the study by Das et al. (2006) and Das et al. (2008) [44, 45]. The DEET concentration of 100 μM is within the concentration range expected in human blood from a dermal treatment 6.7-fold greater than an expected concentration in human blood when DEET is applied at a maximum suggested dose, and approximately one-fifth that for a person introduced to an acute intentional oral overdose of DEET [47]. At the moment, there are no data available on fipronil levels in human blood; one tenth of the treatment level for DEET was used for our fipronil treatments. Three more wells containing only DMSO in the same amount as the other wells was used as a control. Following treatments, all wells were incubated and undisturbed for 72 h in a humidified incubator under the same environmental conditions mentioned above.

RNA isolation, quality assessment and sequencing

Following the 72 h treatment, the media was removed from each well, next each well was washed using 500 μL of 1 x phosphate-buffered saline of the formulation 1 mM KH_2PO_4, 155 mM NaCl, and 3 mM Na_2HPO_4. $7H_2O$ at pH 7.4 (cat. no. 10010–023, Life Technologies, Carlsbad, CA). Then 350 μL of RLT lysis buffer (Qiagen, Valencia, CA) were transferred into each well, and the cells were removed using a sterile, disposable cell scraper resulting in a cell suspension in RLT buffer. The suspensions from each of the 12 wells were stored separately in 1.5 mL microcentrifuge tubes at -80 °C. RNA extraction was conducted using the RNeasy Mini Kit (cat. no. 74104, Qiagen, Valencia, CA) according to the manufacturer's instructions. Each RNA sample was checked for purity using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) by the North Carolina State Genome Sciences Laboratory, Raleigh, NC. Those samples with an RNA Integrity Number (RIN) below 9.0 were not used for sequencing; the lowest RIN found was 9.4. Sequencing was carried out for all treatments, including controls, on the Illumina HiSeq 2000 platform (Illumina, San Diego, CA) at the Beijing Genomics Institute collaborative genome center in the Children's Hospital of Philadelphia (BGI@CHOP, Philadelphia, PA). In order to prepare RNA-Seq libraries, we used a TruSeq RNA Sample Preparation kit following instructions provided by the manufacturer. The libraries were

multiplexed with three samples per lane randomly distributed over four flow cell lanes (12 samples total as defined above) and run on the paired-end read (2 × 100 bp) setting. According to our calculated Phred quality scores (Q scores), approximately 97% of the base reads in each lane had a score above Q30. These quality scores were used to predict probability of errors for base calling. Ideal runs are those where the majority of base scoring is above Q30 for the best possible sequencing applications [48].

Data analysis

The RNA-Seq data that we generated was analyzed using portions of the Tuxedo suite pipeline [49] conducted by the NC State University Bioinformatics Consulting Core. We aligned each fastq file to the hg19 construct of the human genome. This hg19 annotation file was from the University of California, Santa Cruz in conjunction with Cufflinks [50, 51] to guide the assembly for each data set. The CuffMerge program was then used to merge each of the 12 assemblies, and the transcripts produced by this merging were then used to calculate differential expression through the CuffDiff program. During CuffDiff calculations, we used the "rescue method" for multireads and conducted a normalization via the use of geometric means. The Cummerbund package [52] facilitated quality control and resulting plots. In order to ensure that normalization was accurate, we generated scatter plots that indicated no errors. Distribution of fragments per kilobase of exon per million fragments mapped (FPKM) demonstrated similarity across all replicates. The web application Idiographica version 2.4 (http://www.ncrna.org/idiographica) was used to develop chromosomal maps of xenobiotic metabolic transcripts altered by treatments and Venny version 2.1 (http://bioinfogp.cnb.csic.es/tools/venny) to generate Venn diagrams of xenobiotic metabolic transcripts affected by treatments [53, 54]. The GREAT algorithm version 3.0, courtesy of Stanford University (http://great.stanford.edu/public/html/) was used to determine lncRNA transcripts that were within significant proximity (within 1,000,000 base pairs) to xenobiotic metabolic transcripts that were altered on the same chromosome which is indicative of interaction [55]. For those lncRNAs that were significantly proximal to aggregations of cytochrome P450s, a BLAST Alignment was conducted to determine percent identity of these to each other [56]. While it is possible that lncRNA proximity to protein-coding genes is coincidental, oftentimes, genomic proximity of lncRNAs to protein-coding genes can be indicative of interaction between the two via a variety of potential regulatory mechanisms [27]. The differences in expression levels reported here were significant at $\alpha = 0.05$ and were at least 1.4-fold up regulated and 2-fold down-regulated ($\geq$log2 fold change of ± 0.50 and ± 1.00, respectively). A priori exclusion of statistically significant differences based on an arbitrary expression threshold was not conducted since there is no basis to exclude significant results as having no biological importance.

Reverse Transcriptase Quantitative PCR (RT-qPCR) analysis

We reverse transcribed total RNA (1 μg) isolated from a DEET-exposed primary human hepatocyte sample and a control (DMSO) sample in order to produce complementary DNA (cDNA) using the RNeasy mini kit (cat. no. 74104, Qiagen, Valencia, CA) with the SuperScript II Reverse Transcriptase (cat. no. 18064–022, Life Technologies, Carlsbad, CA) following manufacturer's instructions. Hepatocyte cells were plated and treated under the same conditions as described above in the cell culture procedure for this experiment. RNA extraction and reverse transcription protocols and requirements such as RNA integrity thresholds were also under the

same conditions as described in the above methods. Reverse Transcriptase Quantitative PCR (RT-qPCR) was conducted on the Bio-Rad C1000 with a CFX384 Real-time PCR System (Bio-Rad Laboratories, Hercules, CA) using Bio-Rad SsoFast™ EvaGreen® Supermix (cat. no. 1725200, Bio-Rad Laboratories, Hercules, CA) to confirm HiSeq transcriptomics results (all qPCR reagent volumes were used based off of the Supermix guidelines, 1 μL each for forward, reverse primers, and template DNA, 20 μL total volume). We targeted selected transcripts for qPCR based upon previous research where genes were found that were up-regulated by DEET, fipronil, and prototypical inducers like rifampin [47, 49-50]. Primers for the qPCR assays were obtained from Integrated DNA Technologies (IDT, San Jose, CA). Primer and probe sets used were as follows: GAPDH, glyceraldehyde-3-phosphate dehydrogenase variant 1 (NM_002046.7) (For. 5′-GCTGAGAACGGGAAGCTTGTCAT-3′/Rev. 5′-TCTCCATGGTGGTGAAGACGC-3′); CYP 1A1 variant 1 (NM_001319217.2) (For. 5′-GCATGGGCAAGCGGAAGTGTA-3′/Rev. 5′ CATAGATGGGGGTCATGTCCACCT-3′); CYP 1A2 variant 1 (NM_000761) (For. 5′GACGTCCTGCAGATCCGCATT-3′/Rev. 5′-AGGGTGGAGGTGTAGAGGTCAGGC-3′); CYP 2B6 variant 1 (NM_000767) (For. 5′-GGAAACCGCTGGAAGGTGCT-3′/Rev. 5′-AGGAAGGTGGCGTCCATGAG-3′); and CYP 3A4 variant 1 (NM_017460.6) (For. 5′-GCTGGCTATGAAACCACGAGCA-3′/Rev. 5′-GTGGGTGGTGCCTTATTGGGT-3′). NCBI BLAST was used to design all of these primers. Cycling condition began with a hot start at 95 °C for 3 min (initial denaturation of cDNA), then 40 cycles at 95 °C for 10 s (denaturation), 55 °C for 30 s (primer annealing and extension), and finishing with a melt curve analysis with temperature ramps from 65 to 95 °C in 5 °C increments for 5 s each (after amplification cycles). Measurement of amplicon fluorescence was carried out after each amplification cycle with EvaGreen® fluorescent dye. We used a melt curve to determine the presence of single specific products formed by the intercalating dye. We conducted at least 4 replicates and calculations were conducted using delta delta Ct. Lastly, relative expression values were normalized against human GAPDH expression values. All analysis steps were conducted using the Bio-Rad CFX Manager Software v 3.1 (Bio-Rad Laboratories, Inc. Hercules, CA). A statistical significance threshold of $\alpha = 0.05$ was used for all results.

Results

DEET and fipronil treatment effects on transcript alteration for metabolic enzymes

Figure 1 shows the total number of significantly ($\alpha = 0.05$) altered (i.e., up- and down-regulated) transcripts for enzymes potentially involved in phase 1 and phase 2 xenobiotic metabolism after treatment of primary human hepatocytes with 100 μM DEET, 100 μM fipronil, or a mixture of 100 μM DEET plus 10 μM fipronil. Phase 3 mechanisms were not examined. The DEET treatment showed the lowest number of altered genes, fipronil second highest, and the mixture causing the greatest number of altered transcripts. In the DEET treatment, a total of 19 transcripts were altered, with 17 in phase 1 and with 2 in phase 2. In the fipronil treatment, a total of 52 transcripts were altered, 39 in phase 1 and 13 in phase 2. In the DEET + fipronil treatment, a total of 69 transcripts were altered, 46 involved in phase 1 metabolism and 23 in phase 2. Interestingly, there were 2 transcripts that were down-regulated in all treatments, CYP4A11 (−1.11, −2.46, −3.61 log2 fold change, respectively) and CYP4A22 (−1.07, −2.08, and −2.69 log 2 fold change, respectively) and were increasingly down-regulated as discussed before for the total altered transcripts. In terms of the number of altered transcripts found in our analysis, the DEET plus fipronil treatment did not produce a greater than additive effect than each compound

applied separately. In each treatment, the majority of the transcripts altered were phase 1 enzymes compared to phase 2.

Chromosomal distribution of altered transcripts and enzymatic families affected

Figure 2A-C is a chromosomal map of the altered genes found in our studies. For DEET (Figure 2A), 17 transcripts were up-regulated and 2 were down-regulated. For fipronil (Figure 2B), 27 transcripts were up-regulated and 23 down-regulated. For DEET + fipronil (Figure 2C), 29 transcripts were up-regulated and 39 down-regulated. Most of the alteration for DEET was an increase in the number of transcripts (89%), fipronil less up-regulation (53%), and DEET + fipronil only 43% up-regulation. Alteration in our treatments was not specific to a single or even few chromosomes with the mixture affecting more chromosomes than each of the other two treatments. DEET (Figure 2A) altered transcripts on chromosomes 1, 6, 7, 9, 10, 15, 17, and 19 (i.e., 8 separate chromosomes affected). Fipronil (Figure 2B) altered transcripts on all chromosomes with the exception of 11, 13, 14, 16, 18, 21, X, and Y (i.e., 15 separate chromosomes affected). DEET + fipronil (Figure 2C) altered transcripts on all chromosomes with the exception of 11, 13, 16, 18, 21, and Y (i.e., 17 separate chromosomes affected). The number of altered transcripts does not appear to correlate with chromosomal size in any of the three treatments. In the DEET + fipronil treatment (Figure 2C), the largest chromosome (chromosome 1) had 11 altered transcripts while one of the smallest chromosomes (chromosome 19) had 7 altered transcripts. Chromosome 2, the second largest chromosome in the human genome, only had 4 altered transcripts (Figure 2C).

In the DEET treatment, the transcripts altered were aldehyde dehydrogenase (2 up-regulated), epoxide hydrolase (1 up-regulated), cytochrome P450s (12 up- and 2 down-regulated), and glutathione-S-transferase (2 up-regulated) (Table 1). The transcripts altered in the fipronil treatment were cytochrome P450 (17 up, 7 down), aldehyde dehydrogenase (2 up, 3 down), alcohol dehydrogenase (5 down), epoxide hydrolase (1 up, 1 down), flavin-containing monooxygenase (3 down), glutathione-S-transferase (5 up, 1 down), N-acetyltransferase (1 up, 2 down), glucuronidase (2 down), and sulfatase (2 up)(Table 2). The transcripts altered for the DEET + fipronil treatment were cytochrome P450 (16 up, 11 down), aldehyde dehydrogenase (2 up, 5 down), alcohol dehydrogenase (6 down), epoxide hydrolase (1 up, 1 down), flavin-containing monooxygenase (4 down), PON-1 esterase (1 down), glutathione-S-transferase (6 up, 2 down), N-acetyltransferase (1 up, 2 down), glucuronidase (5 down), and sulfatase (3 up) (Table 3). Altered genes in several cases occurred in clusters on different chromosomes. A cluster is defined in this study as 3 or more genes within 1 million base pairs on a single chromosome (discussed in more detail later). The trend for all treatments is that the direction of regulation is similar for a particular cluster. The magnitude of change, however, is variable in these groupings. In Figure 2A, cytochrome P450s CYP4A11 and CYP4A22 are both down-regulated; however, their log 2 fold change differs (−1.11, −1.07, respectively). Even greater differences in log2 fold change appear in the other treatments. In Figure 2B, the flavin-containing monooxygenase family group FMO1, FMO3, and FMO4 are all down-regulated in the same cluster; however their log 2 fold change difference is much greater (−1.2, −0.73, and −0.37, respectively) than the P450s just discussed. Figure 2C also follows this trend of down-regulation within groupings with differential log 2 fold change.

Some of the altered transcripts are known to be important in human health. For example, in all three treatments, members of the CYP2B family were differentially expressed (for DEET, CYP2B6 + 3.82 and CYP2B7P1 + 2.97; for fipronil, CYP2B6 + 3.54 and CYP2B7P1 + 2.42; and DEET + fipronil: +3.84, CYP2B7P1 + 2.79 (Figure 2A-C, Table 1, Table 2, Table 3)). The CYP2B group of enzymes was linked to tumorigenesis and phenobarbital exposure in hepatocytes; CYP2B induction plays a significant role in phenobarbital induced hepatocyte proliferation [57]. In addition, defects in many CYP families have been associated with increased risks of cancer, birth defects, or adverse reactions to xenobiotics [58]. We found in this study that a large number of CYP transcripts were differentially expressed (Figure 2A-C) suggesting potential functions beyond that of xenobiotic metabolism (discussed in more detail later).

Long non-coding RNA proximities with altered transcripts for metabolic enzymes

Significant proximity was determined using the GREAT algorithm [55]; genomic proximity could indicate involvement in the regulation of genes that code for proteins [27]. We found in each treatment that the number of significantly proximal lncRNAs (those associated with altered genes for the metabolic enzymes we investigated) increased from DEET to fipronil to the mixture of DEET + fipronil. In previous studies from our laboratory [41], we found that primary human hepatocyte exposure to a mixture of DEET + fipronil caused a more than additive effect on the number of differentially expressed lncRNAs than that for the sum of the altered lncRNAs for each chemical alone, suggesting a role for lncRNAs in the human response to pesticide exposure. Here, we examined those lncRNAs that were significantly proximal to the protein coding genes for the metabolic enzymes examined. There were 2 proximal lncRNAs in the DEET treatment, 8 proximities in the fipronil treatment, and 11 proximities in the combined treatment. The lncRNAs in proximity to many of the altered P450s were spread among different chromosomes (Figure 3A-C) and both with isolated P450s and with P450s found in clusters. The seven clusters of phase 1 and phase 2 genes are as follows in Figure 2C: chromosome 1 has two clusters (GSTM1, GSTM2, GSTM3 (gluthatione-S-transferases)) and (FMO3, FMO4, FMO1 (flavin-containing monooxygenases)), chromosome 4 has one cluster (ADH6, ADH4, ADH1A, ADH1C, ADH1B (alcohol dehydrogenases)), chromosome 7 has one cluster (CYP3A5, CYP3A4, CYP3A7, CYP3A43 (cytochrome P450s)), chromosome 10 has one cluster (CYP26A1, CYP2C9, CYP2C19, CYP2C8), chromosome 19 has two clusters (CYP4F22, CYP4F12, CYP4F2) and (CYP2A6, CYP2A7, CYP2B6).

Figure 3 shows the relationships we found between altered lncRNAs and both individual and clusters of P450s. Figure 4 (discussed later) shows the same for the other altered metabolic enzymes we examined. Each P450 and lncRNA that had significant proximity and were altered by chemical treatment in this study are described. In Figure 3A (DEET), the lncRNA LINC01004 and CYP3A7 were both altered by exposure and determined to be significantly proximal on chromosome 7 (Q arm). The lncRNA CYP2B7P (a pseudogene) and CYP2A7 and CYP2B6 are within significant genomic proximity on chromosome 19 (Q arm); this particular lncRNA (CYP2B7P) is also included within this cluster of P450s on chromosome 19.

In Figure 3B (fipronil), the lncRNA and P450 relationships were as follows. The lncRNA GCSHP3 was proximal to CYP20A1 on chromosome 2 (Q arm), and the lncRNA ENTPD3-AS1 proximal to CYP8B1 on chromosome 3 (P arm). On chromosome 6 (P arm), there are three lncRNAs in proximity to CYP21A2; these are PSORS1C3, HCP5, and HLA-DRB6. The cluster

of P450s on chromosome 7 (Q arm) including CYP3A5, CYP3A7, CYP3A4, and CYP3A43 are in proximity to the lncRNA AZGP1P1. The lncRNA UCA1 is in proximity to CYP4F22 on chromosome 19 (P arm). Also on chromosome 19 (Q arm), CYP2A6 and CYP2B6 are in proximity to CYP2B7P.

P450 and lncRNA proximities in Figure 3C (DEET + fipronil) are described. The lncRNA F11-AS1 and CYP4V2 are on chromosome 4 (Q arm). On chromosome 6 (P arm), there are three lncRNAs in proximity to CYP21A2; these are HCP5, HCG26, and HLA-DRB6. The cluster of P450s on chromosome 7 (Q arm) including CYP3A5, CYP3A7, CYP3A4, and CYP3A43 are in proximity to the lncRNA AZGP1P1. The lncRNA HECTD2-AS1 and CYP26A1 are on chromosome 10 (Q arm); CYP26A1 is also included in a cluster of P450s (discussed later). The lncRNAs MEG3 and DIO3OS are in proximity to CYP46A1 on chromosome 14 (Q arm). The chromosome 19 (P arm) has a cluster of altered P450s including CYP4F22, CYP4F12, and CYP4F2 that are in proximity to the lncRNA UCA1. Also on chromosome 19 (Q arm) is the cluster of P450s including CYP2A6, CYP2A7, and CYP2B6, which are in significant proximity to two lncRNAs CYP2B7P and PCAT19.

Figure 4A-B show base-pair proximity relationships between altered lncRNAs and non-P450 phase 1 and phase 2 genes. In Figure 4A (fipronil), the phase 1 and phase 2 proximities to lncRNA genes are as follows. On chromosome 1 (P arm), there are two lncRNAs AKR7L and LOC100506730 that are in proximity to ALDH4A1 (aldehyde dehydrogenase). The lncRNA LINC01160 is in proximity to a cluster of three coding genes GSTM2, GSTM1, and GSTM3 (glutathione-S-transferases) on chromosome 1 (P arm). A second cluster on chromosome 1 (Q arm) is also present; FMO1, FMO3, and FMO4 (flavin-containing monooxygenases) are in proximity to the lncRNAs ANKRD36BP1, LOC100505918, and GAS5. The lncRNAs RPL32P3 and TMCC1-AS1 are in proximity to ALDH1L1 (aldehyde dehydrogenase) on chromosome 3 (Q arm). The lncRNA GTF2H2B and GUSBP3 (glucuronidase) are in proximity on chromosome 5 (Q arm). The lncRNA CMAHP is in proximity to ALDH5A1 (aldehyde dehydrogenase) on chromosome 6 (P arm). Two genes in separate families are also in proximity to the same lncRNA SNHG6 on chromosome 8 (Q arm); these are ADHFE1 (alcohol dehydrogenase) and SULF1 (sulfatase). The lncRNA LOC100130075 and GNS (sulfatase) are in proximity on chromosome 12 (Q arm). The lncRNA LINC00673 and NAT9 (N-acetyltransferase) are in proximity to each other on chromosome 17 (Q arm). The lncRNA PI4KAP2 and GSTT2 (glutathione-S-transferase) are in proximity to each other on chromosome 22 (Q arm).

In Figure 4B (DEET + fipronil), the phase 1 and phase 2 gene and lncRNA relationships are as follows. The lncRNA LINC01160 is in proximity to a cluster of three coding genes GSTM2, GSTM1, and GSTM3 (glutathione-S-transferases) on chromosome 1 (P arm). The lncRNA LOC728989 and FMO5 (flavin-containing monooxygenase) are in proximity to each other on chromosome 1 (Q arm). A second cluster on chromosome 1 (Q arm) is also present; FMO1, FMO3, and FMO4 (flavin-containing monooxygenases) are in proximity to the lncRNAs ANKRD36BP1, LOC100505918, and GAS5. The coding gene MGST3 (glutathione-S-transferase) is also included in this base-pair proximity relationship. The lncRNA ESRG and NAT6 (N-acetyltransferase) are in proximity on chromosome 3 (P arm). Also on chromosome 3 (Q arm), the lncRNAs RPL32P3 and TMCC1-AS1 are in proximity to ALH1L1 (aldehyde dehydrogenase). The lncRNAs SNHG8 and LINC01061 are in proximity to ARSJ (sulfatase) on

chromosome 4 (Q arm). The pseudogenes GUSPB3 (glucuronidase), SMA4, and SMA5 (sulfatases) are in significant proximity to each other on chromosome 5 (Q arm). There are three lncRNAs in proximity to ALDH5A1 (aldehyde dehydrogenase) on chromosome 6 (P arm), these are CMAHP, BTN2A3P, and GUSBP2. Two lncRNAs that have been noted earlier, AZGP1P1 and MGC72080, are in proximity to the single coding-gene PON1 (paraoxonase) on chromosome 7 (Q arm). Two genes in separate families are also in proximity to the same lncRNA SNHG6 on chromosome 8 (Q arm); these are ADHFE1 (alcohol dehydrogenase) and SULF1 (sulfatase). The lncRNA LOC100130075 and GNS (sulfatase) are in proximity on chromosome 12 (Q arm). The lncRNA ADAM1A is in proximity to ALDH2 (alcohol dehydrogenase) on chromosome 12 (Q arm). The lncRNA LOC100289511 and ALDH6A1 (aldehyde dehydrogenase) are in proximity on chromosome 14 (Q arm). The lncRNA CCDC144B and ALDH3A1 (aldehyde dehydrogenase) are in proximity on chromosome 17 (P arm). Lastly, the pseudogene GUSBP11 and coding-gene GSTT2 (glutathione-S-transferase) are in significant proximity to each other on chromosome 22 (Q arm).

LncRNAs have been determined to potentially be categorized as intergenic, sense, antisense, intronic, pseudogenic and divergent based upon their proximity to nearby protein-coding genes that have these lncRNAs act in various regulatory roles [59]. However, specific targets and functions of specific lncRNAs are often unclear [59]. A majority of the lncRNAs examined in this study can be categorized as intergenic to the metabolic coding genes we also examined. Of the lncRNAs that we found in proximity in this study (categorized by GeneBank) 15 were pseudogenes, 8 intergenic, 5 antisense, and 8 were uncategorized.

Similarity of altered lncRNAs proximal to P450 clusters for the DEET + fipronil treatment

After determining which lncRNAs were significantly proximal to an assemblage of P450s (an assemblage being defined as 3+ P450s in one location on a single chromosome), nucleotide BLAST alignments were conducted to determine percent identity between lncRNA genes proximal in four P450 clusters (C1-C4) (Figs. 2C and 5) [56]. The BLAST alignment was conducted for the gene sequence not the mRNA transcript sequence. C1, cluster 1 composed of CYP3A5, CYP3A4, CYP3A7, and CYP3A43 (all up-regulated) had the lncRNAs AZGP1P1 and MGC72080 in significant proximity. C2, cluster 2 composed of CYP26A1, CYP2C19, CYP2C9, and CYP2C8 (all up-regulated except CYP2C19) had the lncRNA HECTD2-AS1 in significant proximity. C3, cluster 3 composed of CYP4F22, CYP4F12, and CYP4F2 (down-regulated except for CYP4F22) had the lncRNA UCA1 in significant proximity. C4, cluster 4 composed of CYP2A6, CYP2A7, and CYP2B6 (all up-regulated) had the lncRNAs PCAT19 and CYP2B7P in significant proximity. Percent identities from lncRNA alignments are as follows: MGC72080 x AZGP1P1 (91%), CYP2B7P x PCAT19 (84%), AZGP1P1 x PCAT19 (82%), AZGP1P1 x CYP2B7P (89%), MGC72080 x PCAT19 (84%), MGC72080 x CYP2B7P (90%), HECTD2-AS1 x AZGP1P1 (91%), HECTD2-AS1 x MGC72080 (91%), HECTD2-AS1 x PCAT19 (94%), HECTD2-AS1 x CYP2B7P (90%), UCA1 x CYP2B7P (no significant alignment), UCA1 x PCAT19 (no significant alignment), UCA1 x AZGP1P1 (no significant alignment), UCA1 x MGC72080 (no significant alignment), and UCA1 x HECTD2-AS1 (no significant alignment). Interestingly, UCA1 the lncRNA in proximity to cluster 3 which was in majority down-regulated did not have significant alignment to the other lncRNAs examined. These alignments were conducted using gene sequences from the NCBI database; when aligned using mRNA transcript sequence no significant alignments were found. Using the same technique, the lncRNAs in

proximity to the coding gene clusters in Figure 4 were also compared. On chromosome 1 (P arm), the cluster of glutathione-S-transferases GSTM2, GSTM1, and GSTM3 had the lncRNA LINC01160 in proximity. Also on chromosome 1 (Q arm), the cluster of flavin-containing monooxygenases FMO3, FMO1, and FMO4 had the lncRNAs GAS5, LOC100505918, and ANKRD36BP1 in proximity (Figure 4). A BLAST alignment on these 4 lncRNAs was conducted at the gene level resulting in the following identities: GAS5 x ANKRD36BP1 (no significant alignment), GAS5 x LINC01160 (no significant alignment), GAS5 x LOC100505918 (79%), ANKRD36BP1 x LOC100505918 (no significant alignment), ANKRD36BP1 x LINC01160 (no significant alignment), and LOC100505918 x LINC01160 (86%)

Discussion

DEET and fipronil alteration of phase 1 and 2 transcripts

In this study, we found that mRNA transcript levels associated with possible xenobiotic metabolism for phase 1 and 2 pathways were affected by primary human hepatocyte exposure to DEET, fipronil, and DEET + fipronil (Figure 1). Changes in transcript levels for a particular gene may or may not translate to changes in protein levels and/or changes in metabolic activity; our analysis only measured transcript levels. In addition, some of the altered transcripts coding for metabolic proteins may be involved in xenobiotic metabolism and/or the metabolism of endogenous substrates; for example, P450s that were altered might be involved in steroid hormone synthesis.

There is supporting data suggesting that at least some of the alteration of transcripts observed in this study are relevant to xenobiotic metabolism. Das et al. (2008), for example, reported changes in transcript levels for P450s correlated with similar directional changes in both P450 protein levels and enzymatic activity after human hepatocyte exposure to DEET [45]. DEET affected significant changes in the mRNA transcript levels (using branched DNA assays) for CYP3A4, CYP2B6, CYP2A6 and CYP1A2 (all up-regulated). For CYP3A4, the increase in mRNA levels also resulted in increased protein levels as determined by Western Blotting and an increase in testosterone metabolism [45]. In this study, we confirmed the results of Das et al. (2008) and found that transcript levels for CYP3A4 had a + 3.74 log 2 fold change, CYP2B6 + 3.84, CYP1A2 + 1.05, and CYP2A6 + 1.75 (Table 3) when exposed to DEET under very similar study conditions to that conducted by Das et al [45]. It is also noteworthy that Das et al. (2008) found that directional changes in P450 levels and activity translated into changes in xenobiotic metabolism; for example, P450 down-regulation resulted in the reduced metabolism of DEET [45]. We assume the same in our studies.

Similar studies to that of DEET were also conducted with fipronil by Das et al. (2006) [44]. CYP3A4 and CYP1A1 in response to human hepatocyte exposure were up-regulated at the transcript level (using branched DNA assays), the protein level (determined by Western blots) and activity level (measured by luciferin and testosterone metabolism). These same P450s were altered in the same direction (up-regulated) in our current study as determined by RNAseq (Figure 2B and C). In order to further confirm gene expression changes for our RNAseq results, quantitative PCR was conducted for CYP1A1, CYP1A2, CYP2B6, and CYP3A4 and similar directional changes were found [40, 41].

It appears in summary that transcript alteration by exposure to DEET and fipronil in primary human hepatocytes reported here using RNAseq correlates to similar P450 results in previous work determined by DNA branched assays and also translates to changes in P450 protein and activity levels including metabolism of DEET [44, 45]. Celius et al. (2010) found that TCDD increased the mRNA levels of FMOs in human hepatoma cells and mouse livers but the level of increase was not reflected to the same degree for the protein concentration or catalytic activity [60]. FMO3 was increased in this study in the fipronil and DEET + fipronil treatments (Figure 2C). DEET and fipronil elicited increased P450 transcript levels reported here (14 in the DEET treatment, 23 in the fipronil treatment, and 27 in the DEET + fipronil treatment; Table 1, Table 2, Table 3) with evidence of increased P450 protein levels and activity as well [44, 45]. This potentially could result in changes in metabolism for a variety of substrates [61, 62]. It was reported before that pesticide exposure (for example to organophosphates and fipronil) can inhibit the metabolism of drugs and hormones [63-65]. Usmani et al. (2006) discovered that chemical mediated alteration of CYP1A2 and CYP3A4 was linked to the inhibition of estradiol in human liver cells [63]. These same P450s have been altered in our work (Figure 2C). Therefore, it is likely that mRNA transcript alteration of P450s and the other metabolic gene families examined in this study is having an effect on xenobiotic metabolism. More research is needed to understand the short term and long-term effects of exposure to DEET and fipronil but also other pesticides alone and in mixtures on their possible impact on human physiology and drug-drug interactions.

Global gene alteration versus alteration of transcripts for metabolic genes

Our group (Mitchell et al., 2016) found that primary human hepatocytes treated with 100 μM DEET, 10 μM fipronil, and 100 μM DEET +10 μM fipronil caused a more than additive response in global transcript alteration for the mixture [40]. A more than additive response is defined as the combined effect of the mixture being greater than the simple additive effect of each chemical individually. Mitchell et al. (2016) using the same treatment levels and conditions as that used in this current work reported that DEET altered 172 global transcripts, fipronil 3674 global transcripts, and DEET + fipronil 5104 global transcripts; the simple additive effect would have been 3846 altered transcripts for the DEET + fipronil treatment [40]. This more than additive effect for global gene alteration did not occur for the phase 1 and 2 genes (Figure 1), i.e., DEET had 19 altered transcripts, fipronil had 52, and DEET + fipronil had 69; a simple additive effect would be 71 altered metabolic transcripts. This suggests that different regulatory mechanisms were involved in global alteration versus that for xenobiotic metabolism, and there were no interactions from the mixture of DEET + fipronil on human hepatocytes to enhance the treatment effect unlike that for global alteration.

Potential impact of chemical mixtures on function

Combined exposure to pesticides occurs frequently. Bergmann et al. (2017) found that volunteers over the course of one-month had the highest pesticide exposure, 106 different chemicals from 8 different classes [66]. Another vulnerable demographic to combinatorial exposure are children. One epidemiological study conducted in the United States found children to be particularly prone to multiple pesticide exposures including organochlorines, organophosphates, carbamates, and neonicotenoids from food, water, and the agricultural and urban environment [67]. It has been noted that increased exposure to xenobiotics causes cellular

stress leading to down-regulation of general transcription and in some cases up-regulation of stress related transcripts such as heat shock proteins and cytokines [68, 69].

There was no evidence of cytotoxicity in the studies reported here. The combined treatment of DEET + fipronil resulted in the highest number of altered metabolic-associated transcripts than each pesticide alone but interestingly, also the highest number of down-regulated transcripts (Figure 2C, Table 3). Most often risk assessments for pesticide exposure are conducted one chemical at a time. The results shown here highlight the importance of investigating chemical mixtures. One explanation for the increased alteration in the DEET + fipronil treatment in comparison to the other two treatments alone is that DEET can have a potentiation effect on fipronil. Potentiation is an observed phenomena when one chemical, in particular insecticides, serves to increase the overall toxicity of another chemical when combined. DEET and fipronil have been demonstrated to elicit this response in a variety of models [37, 70-71]. DEET (10 nM compared to 100 μM in this study) was observed to be a potentiator of carbamate insecticides through interactions with muscarinic receptors which effectively increased the potency of carbamates in mosquito control [70]. Fipronil (50 g/ha) has been observed to have potentiative characteristics when sequentially exposed to neonicotinoid insecticides on non-target invertebrates, e.g., ecologically important annelids [71].

Potential role for long non-coding RNAs in exposure to environmental chemicals

We reported before that lncRNAs across the genome were altered when primary human hepatocytes were exposed to DEET and fipronil, alone and in combination [41]. Figure 3 and Figure 4 show lncRNA genes that were altered by our pesticide treatments that were significantly proximal to metabolic genes that we examined (P450s and the other 9 xenobiotic metabolic groups we examined). When examining the proximity relationships of altered P450s to altered lncRNAs, the DEET treatment had 2 proximal lncRNAs, fipronil had 6, and DEET + fipronil had 7 (Figure 3 A-C). When examining the other 9 enzymatic groups in our study, there were 10 altered lncRNAs proximal to altered coding-genes in the fipronil treatment and 15 in the DEET + fipronil treatment (Figure 4 A-B). The DEET treatment did not have any altered lncRNAs in proximity to altered non-P450 genes examined in this study. In both Figure 3 and Figure 4, altered lncRNAs were found to have genomic proximity to altered solitary coding-genes and groups of altered coding-genes (a group being defined as 3 or more genes). For example, the lncRNA AZGP1P1 was in significant proximity to CYP3A5, CYP3A7, CYP3A4, and CYP3A43 (Figure 3C). The lncRNA F11-AS1 was in proximity to the solitary P450 CYP4V2 (Figure 3C). This trend is also seen in Figure 4 where the lncRNA LINC01160 is in genomic proximity to GSTM2, GSTM1, and GSTM3 (Figure 4B). Additionally, the lncRNA ADAM1A is in proximity to the solitary gene ALDH2 (Figure 4B). For each enzyme group that we examined in this study, there were lncRNAs in significant genomic proximity, the only exception being the epoxide hydrolases. Each group had the following number of lncRNA-coding gene proximities (for all treatments): P450s-15, flavin-containing monooxygenases-3, aldehyde dehydrogenases-8, alcohol dehydrogenases-2, epoxide hydrolases-0, glutathione-S-transferases-5, glucuronidases-1, paraoxonases-1, sulfatases-6, and N-acetyltransferases-2 (Figure 3, Figure 4). These results suggest a possible role for these lncRNAs in the regulation of genes involved in xenobiotic metabolism and/or other cellular metabolic activity.

Genomic proximity of lncRNAs to protein-coding genes often indicates a cis-regulatory role; however, it is also possible for lncRNAs to act in a trans-regulatory function proximal to protein-coding genes [27]. LncRNAs have been determined to potentially be categorized as intergenic, sense, antisense, intronic, pseudogenic and divergent based upon their proximity to nearby protein-coding genes [59]. However, specific targets and functions of specific lncRNAs are often unclear [59]. Of the lncRNAs that we found in proximity to coding genes in this study 15 were pseudogenes, 8 intergenic, 5 antisense, and 8 were uncategorized (Figure 3, Figure 4; Table 1, Table 2, Table 3, categorized by GeneBank).

Interestingly, the majority of lncRNAs we examined in proximity to coding genes in this study were pseudogenes (15 lncRNAs). Pseudogenes are a type of lncRNA that predominantly facilitates a regulatory role on their target gene, particularly when these pseudogenes are transcribed [40, 41]. Most often pseudogenes act as genetic regulators through interactions with promoter regions of the target coding gene [41, 72]. One potential mechanism that could apply to down-regulated transcripts could be a single lncRNA being fragmented into small interfering RNAs (siRNA) in order to reduce the transcription of the target gene; this process is known as RNA interference [73]. This lncRNA to coding gene interaction could potentially relate to the high percent identities in Figure 5 where small portions of the lncRNAs compared had high percent identities to each other. In Figure 5, after running a nucleotide BLAST alignment on the gene sequence of those lncRNAs in proximity to 4 clusters of P450s (C1-C4), a high percent identity (>80%) but low query coverage was found when all lncRNAs in proximity to the 4 clusters were aligned. Two high identities were found also in lncRNAs in proximity to FMO and glutathione-S-transferase clusters (GAS5 x LOC100505918 79% and LOC100505918 x LINC01160 86%). When the mRNA transcript sequence was used for alignment, no significant alignments were found. While the high percent identities found at the genomic level is interesting, these data alone are not conclusive in establishing an overall link among the different altered lncRNAs associated with altered metabolic enzymes shown in Figure 5.

Several of the lncRNAs we found to be altered by treatment with DEET and fipronil have also been implicated in disease and increased risk of disease (the lncRNAs UCA1 and GAS5, Figure 3, Figure 4). The lncRNA UCA1 is involved in suppression of p27, increased expression of cell-cycle progression, increased expression of antiapoptotic protein Bcl-2, and increased expression of c-myc, all of these being linked to resistance to anti-cancer treatments and promotion of tumor growth [38, 74]. One problem that may stem from exposures to environmental chemicals both individually and in combination besides direct toxicity could be increased vulnerability to other diseases such as cancer or complications including resistance to anti-cancer drugs due to lncRNA alteration [38]. When a particular lncRNA becomes altered by chemical exposure, for example UCA1 or GAS5, this may increase the overall risk a person experiences for such medical complications as drug resistance. Therefore, for these reasons further investigation is needed to elucidate the role of lncRNAs in toxicological exposures and xenobiotic metabolism.

In summary, the evidence we found in this study suggests that P450s in addition to other important xenobiotic metabolism transcripts are significantly up- and down-regulated by exposure to DEET and fipronil, alone and in a mixture, in primary human hepatocytes. DEET and fipronil caused alteration in the RNA transcripts of 10 major enzyme families that we

examined, 10 major families potentially responsible for phase 1 reactions and 6 in phase 2. Included in these altered transcripts were the P450s. An increased level of alteration occurred between treatments with DEET, fipronil, and DEET + fipronil; however, unlike global expression levels where a more than additive effect on transcription occurred, this was not the case for the transcripts possibly coding for proteins involved in xenobiotic metabolism. This is suggesting a difference in the regulation of these enzymes. We found a number of altered lncRNAs in close genomic proximity to xenobiotic metabolism transcripts that were altered and in the cases examined, the lncRNAs had a high identity at the genomic level but not the transcript level. This may be a suggestion of a “master” mechanism for the regulation of xenobiotic metabolism but needs much more work to understand. Many of the proximal lncRNAs found were pseudogenes or intergenic, suggesting an interaction between the coding and non-coding genes characterized. Among these lncRNAs, we found UCA1 and GAS5 were altered, which have been implicated in other studies to be associated with Alzheimer's and resistance to anti-cancer drugs when altered. Much more research is needed to understand the role of lncRNAs in metabolic responses to environmental chemical exposure.

REFERENCES

1. Kullman, S.W.; Mattingly, C.J.; Meyer, J.N.; Whitehead, A. Perspectives on informatics in toxicology. In: Hodgson E., editor. *A textbook of modern toxicology.* Hoboken, NJ: Wiley; **2010**, *pp 593–605.*

2. Judson, R.S.; Houck, K.A.; Kavlock, R.J.; Knudsen, T.B.; Martin, M.T.; Mortensen, H.M.; Reif, D.M.; Rotroff, D.M.; Shah, I.; Richard, A.M. In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. *Environ. Health Perspect*. **2010**. *118, 485-485.*

3. National Research Council. Committee on Toxicity Testing and Assessment of Environmental Agents, National Research Council. Toxicity testing in the 21st century: a vision and a strategy. Washington, DC: National Academies Press. **2007**.

4. Schmidt, C.W. Tox21: new dimensions of toxicity testing. *Environ. Health. Perspec.* **2009**. *117,* 348–353.

5. National Toxicology Program. Tox21: transforming environmental health. **2011**. Web. Accessed May 3, 2015, http://www.niehs.nih.gov/research/resources/assets/docs/tox21_transforming_environmental_health_508.pdf.

6. Gibbs, R.A.; Belmont, J.W.; Hardenbol, P.; Willis, T.D.; Yu, F.; Yang, H.; Ch'ang, L.-Y.; Huang, W.; Liu, B.; Shen, Y. The international HapMap project. *Nature*. **2003**. 426, 789–796.

7. Thorisson, G.A.; Smith, A.V.; Krishnan, L.; Stein, L.D. The international HapMap project web site. *Gen. Res*. **2005**. 15, 1592–1593.

8. Schoenig, G.P.; Osimitz, T.G.; Gabriel, K.L.; Hartnagel, R.; Gill, M.W.; Goldenthal, E.I. Evaluation of the chronic toxicity and oncogenicity of N, N-diethyl-m-toluamide (DEET). *Toxicol. Sci.* **1999**. 47, 99–109.

9. Arthur, O.; Maciarello, J. Essential oil analysis and field evaluation of the citrosa plant "*Pelargonium citrosum*" as a repellent against populations of *Aedes* mosquitoes. *J. Am. Mosq. Control Assoc.* **1996**. 12, 69–74.

10. Osimitz, T.G.; Murphy, J.V.; Fell, L.; Page, B.C. Adverse events associated with the use of insect repellents containing *N,N*-diethyl-m-toluamide (DEET). *Regul. Toxicol. Pharmacol.* **2010**. *56*, 93–99.

11. Veltri, J.C.; Osimitz, T.G.; Bradford, D.C.; Page, B.C. Retrospective analysis of calls to poison control centers resulting from exposure to the insect repellent *N,N* diethyl-m-toluamide (DEET) from 1985–1989. *Clin. Toxicol.* **1994**. *32*, 1–16.

12. Slaninova, A.; Modra, H.; Hostovsky, M. Effects of subchronic exposure to N,N-diethyl-m-toluamide on selected biomarkers in common carp (*Cyprinus carpio L*.). *Biomed. Res. Int.* **2014**. *828515*. Online Journal.

13. Steele, L.; Sastre, A.; Gerkovich, M.M.; Cook, M.R. Complex factors in the etiology of Gulf War illness: wartime exposures and risk factors in veteran subgroups. *Environ. Health Perspect*. **2012**. *120*, 112-118.

14. Osimitz, T.G.; Murphy, J.V. Neurological effects associated with use of the insect repellent N,N-diethyl-m-toluamide (DEET). *J. Toxicol. Clin. Toxicol*. **1997**. *35*, 435–441.

15. Yiin, L.-M.; Tian, J.-N.; Hung, C.-C. Assessment of dermal absorption of DEET-containing insect repellent and oxybenzone-containing sunscreen using human urinary metabolites. *Environ. Sci. Poll. Res.* **2015**. *22*, 7062–7070.

16. Abu-Qare, A.W.; Abou-Donia, M.B. Combined exposure to DEET (*N*,*N*-diethyl-m-toluamide) and permethrin: pharmacokinetics and toxicological effects. *J. Toxicol. Environ. Health, Part B*. **2003**. *6*, 41–53.

17. Lerapetritou, M.G.; Georgopoulos, P.G.; Roth, C.M.; Androulakis, L.P. Tissue-level modeling of xenobiotic metabolism in liver: an emerging tool for enabling clinical translational research. *Clin. Transl. Sci.* **2009**. *2*, 228–237.

18. Selim, S.; Hartnagel, R.E. Jr.; Osimitz, T.G.; Gabriel, K.L.; Schoenig, G.P. Absorption, metabolism, and excretion of N,N-Diethyl-m-toluamide following dermal application to human volunteers. *Fundam. Appl. Toxicol.* **1995**. *25*, 95-100.

19. Carrington, D. EU to ban fipronil to protect honeybees. **2013**. Web. Accessed May 8, 2015, http://www. theguardian.com/environment/2013/jul/16/eu-fipronil-ban-bees.

20. Hamon, N.; Gamboa, H.; Melgarejo Garcia, J.E. Fipronil: a major advance for the control of boll weevil in Colombia. In: *Proceedings Beltwide Cotton Conferences*, Nashville, TN, USA. **1996**. *2*, 990–994.

21. EPA 737-F-96-005, New Pesticide Fact Sheet–Fipronil. U.S. Environmental Protection Agency, Office of Prevention, Pesticides and Toxic Substances, Office of Pesticide Programs, U.S. Government Printing Office, Washington, DC, **1996**, pp 1–10.

22. Hurley, P.M. Mode of carcinogenic action of pesticides inducing thyroid follicular cell tumors in rodents. *Environ. Health Perspect.* **1998**. *106*, 437-445.

23. Patterson, A.D.; Gonzalez, F.J.; Idle, J.R. Xenobiotic metabolism: a view through the metabolometer. *Chem. Res. Toxicol*. **2010**. *23*, 851–860.

24. Wallace, B.D.; Redinbo, M.R. Xenobiotic-sensing nuclear receptors involved in drug metabolism: a structural perspective. *Drug. Metab. Rev.* **2012**. *45*, 79-100.

25. Dempsey, J.L.; Cui, J.Y. Long non-coding RNAs: a novel paradigm for toxicology. *Toxicol. Sci.* **2016**. *155*, 3–21.

26. Lai, W.C.; Kayedkhordeh, M.; Cornell, E.V.; Farah, E.; Bellaousov, S.; Rietmeijer, R.; Salsi, E.; Mathews, D.H.; Ermolenko, D.N. mRNAs and lncRNAs intrinsically form secondary structures with short end-to-end distances. *Nat. Commun.* **2018**. *9*.

27. Ponting, P.P.; Oliver, O.L.; Reik W. Evolution and functions of long noncoding RNAs. *Cell*. **2009**. *136*, 629–641.

28. Chen, S.; Shao, C.; Xu, M.; Ji, J.; Xie, Y.; Lei, Y.; Wang, X. Macrophage infiltration promotes invasiveness of breast cancer cells via activating long non-coding RNA UCA1. *Int. J. Clin. Exp. Pathol.* **2015**. *8*, 9052–9061.

29. Faghihi, M.A.; Modarresi, F.; Khalil, A.M.; Wood, D.E.; Sahagan, B.G.; Morgan, T.E.; Finch, C.E.; St Laurent, G., 3rd; Kenny, P.J.; Wahlestedt, C. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat. Med.* **2008**. *14*, 723–730.

30. Marchese, F.P. ; Raimondi, I. ; Huarte, M. The multidimensional mechanisms of long noncoding RNA function. *Genome. Biol.* **2017**. *18*, 206-206.

31. Mayama, T.; Marr, A.K.; Kino, T. Differential expression of glucocorticoid receptor noncoding RNA repressor Gas5 in autoimmune and inflammatory diseases. *Horm. Metab. Res.* **2016**. *48*, 550–557.

32. Bai, W.; Yang, J.; Yang, G.; Niu, P.; Tian, L.; Gao, A. Long non-coding RNA NR_045623 and NR_028291 involved in benzene hematotoxicity in occupationally benzene-exposed workers. *Exp. Mol. Pathol.* **2014**. *96*, 354–360.

33. Lu, Z.; Xiao, Z.; Liu, F.; Cui, M.; Li, W.; Yang, Z.; Li, J.; Ye, L.; Zhang, X. Long non-coding RNA HULC promotes tumor angiogenesis in liver cancer by up-regulating sphingosine kinase 1 (SPHK1). *Oncotarget.* **2016**. *7*, 241–254.

34. Lempiainen, H.; Couttet, P.; Bolognani, F.; Muller, A.; Dubost, V.; Luisier, R.; Del Rio Espinola, A.; Vitry, V.; Unterberger, E.B.; Thomson, J.P. Identification of Dlk1-Dio3 imprinted gene cluster noncoding RNAs as novel candidate biomarkers for liver tumor promotion. *Toxicol. Sci.* **2013**. *131*, 375–386.

35. Tani, H.; Takeshita, J.I.; Aoki, H.; Abe, R.; Toyoda, A.; Endo, Y.; Miyamoto, S.; Gamo, M.; Torimura, M. Genome-wide gene expression analysis of mouse embryonic stem cells exposed to p-dichlorobenzene. *J. Biosci. Bioeng.* **2016**. *122*, 329–333.

36. Karapetyan, A.R.; Buiting, C.; Kuiper, R.A.; Coolen, M.W. Regulatory roles for long ncRNA and mRNA. *Cancers.* **2013**. *5*, 462–490.

37. Smart, R.C.; Hodgson, E. Molecular and Biochemical Toxicology fifth edition. **2018**. Wiley Publishing. Indianapolis, Indiana, USA. Pages *196-210*.

38. Hahne, J.C.; Valeri, N. Non-coding RNAs and resistance to anticancer drugs in gastrointestinal tumors. *Frontiers in Oncology*. **2018**. *8*, 226-226.

39. Li, C.Y.; Cui, J.Y. Regulation of protein-coding gene and long noncoding RNA pairs in liver of conventional and germ-free mice following oral PBDE exposure. *PLoS One*. **2018**. *13*, e0201387.

40. Mitchell, R.D.; Dhammi, A.; Wallace, A.; Hodgson, E.; Roe, R.M. Impact of environmental chemicals on the transcriptome of primary human hepatocytes: potential for health effects. *J. Biochem. Molec. Tox.* **2016**. *30*, 8, 375-395.

41. Mitchell, R.D.; Wallace, A.D.; Hodgson, E.; Roe, R.M. Differential expression profile of lncRNAs from primary human hepatocytes following DEET and fipronil exposure. *Int. J. Mol. Sci.* **2017**. *18*, 2104-2104.

42. Peng, C.J.; Yuan, D.; Li, B.; Wei, Y.G.; Yan, L.N.; Wen, T.F.; Zhao, J.; Yang, J.; Xu, M.Q. Body mass index evaluating donor hepatic steatosis in living donor liver transplantation. *Elsevier Online Journal*. **2009**. *41*, 9, 3556–3559.

43. LeCluyse, E.L.; Madan, A.; Hamilton, G.; Carroll, K.; DeHaan, R.; Parkinson, A. Expression and regulation of cytochrome P450 enzymes in primary cultures of human hepatocytes. *J. Biochem. Mol. Toxicol.* **2000**. *14*, 177–188.

44. Das, P.C.; Cao, Y.; Cherrington, N.; Hodgson, E.; Rose, R.L. Fipronil induces CYP isoforms and cytotoxicity in human hepatocytes. *Chem. Biol. Interact*. **2006**. *164*, 200–214.

45. Das, P.C.; Cao, Y.; Rose, R.L.; Cherrington, N.; Hodgson, E. Enzyme induction and cytotoxicity in human hepatocytes by chlorpyrifos and N,N-diethyl-m-toluamide (DEET). *Drug Metabol. Drug Interact.* **2008**. *23*, 237–260.

46. Tice, R.; Brevard, B. N,N-Diethyl-m-toluamide (DEET) [134-62-3]. A Review of Toxicological Literature. National Institute of Environmental Health Sciences. *NIEHS Literature Review*. **1999**. Online. https://ntp.niehs.nih.gov/ntp/htdocs/chem_background/exsumpdf/deet_508

47. Baselt, R.C.; Cravey, R.H. Disposition of toxic drugs and chemicals in man. Biomedical Publications, Seal Beach, California (USA), **2014**, hardbound, ISBN: 978-0-9626523-9-4.

48. Illumina Inc. Quality scores for next-generation sequencing. **2011**. Web. Accessed June 6, 2015, http://www. illumina.com/documents/products/technotes/technote_Q-Scores.pdf.

49. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D.R.; Pimentel, H.; Salzberg, S.L.; Rinn, J.L.; Pachter, L. Differential gene and transcript expression analysis of RNAseq

experiments with TopHat and Cufflinks. *Nature Protocols*. **2012**. *7*, 562–578.

50. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; Van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**. *28*, 511–515.

51. Trapnell, C.; Hendrickson, D.G.; Sauvageau, M.; Goff, L.; Rinn, J.L.; Pachter, L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **2013**. *31*, 46–53.

52. Goff, L.; Trapnell, C.; Kelley, D. cummeRbund: analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. **2013**. R package version 2.13.0.

53. Kin, T.; Ono, Y. Idiographica: a general-purpose web application to build idiograms on-demand for human, mouse and rat. *Bioinformatics.* **2007**. *23*, 2945–2946.

54. Oliveros, J. VENNY. An interactive tool for comparing lists with Venn diagrams. **2007**. Web. Accessed July 14 2018. http://bioinfogp. cnb.csic.es/tools/venny/.

55. McLean, C.Y.; Bristor, D.; Hiller, M.; Clarke, S.L.; Schaar, B.T.; Lowe, C.B.; Wenger, A.M.; Bejerano, G. GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotech*. **2010**. *28*, 495–501.

56. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**. *215*, 403–410.

57. Li, L.; Bao, X.; Zhang, Q.Y.; Negishi, M.; Ding, X. Role of CYP2B in phenobarbital-induced hepatocyte proliferation in mice. *Drug Metab. Dispos.* **2017**. *45*, 977–981.

58. Nebert, D.W.; Wikvall, K.; Miller, W.L. Human cytochromes P450 in health and disease. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. **2013**. *368*, 1612-1612.

59. Zhao, X.Y.; Lin, J.D. Long noncoding RNAs: a new regulatory code in metabolic control. *Trends Biochem Sci.* **2015**. *40*, 10, 586-596.

60. Celius, T.; Pansoy, A.; Matthews, J.; Okey, A. B.; Henderson, M. C.; Krueger, S. K.; Williams, D.E. Flavin-containing monooxygenase-3: induction by 3-methylcholanthrene and complex regulation by xenobiotic chemicals in hepatoma cells and mouse liver. *Toxicology and applied pharmacology*, **2010**. *247*(1), 60–69.

61. Caballero, M.V.; Ares, I.; Martinez, M.; Martinez-Larranaga, M.R.; Anadon, A.; Martinez, M.A. Fipronil induces CYP isoforms in rats. *Food Chem. Toxicology*. **2015**. *83*, 215–221.

62. Jin, Y.; Cong, B.; Wang, L.; Gao, Y.; Zhang, H.; Dong, H.; Lin, Z. Differential gene expression analysis of the *Epacromius coerulipes* (Orthoptera: Acrididae) transcriptome.

Journal of Insect Science. **2016**. *16*, 1, 42-42.

63. Usmani, K.; Rose, R.L.; Hodgson, E. Inhibition of the human liver microsomal and human cytochrome P450 1A2 and 3A4 metabolism of estradiol by deployment-related chemicals. *Drug Metabolism Dispos.*, **2006**. *34*, 1606-1614.

64. Di Consiglio, E.; Meneguz, A.; Testai, E. Organophosphorothionate pesticides inhibit the bioactivation of imipramine by human hepatic cytochrome P450s. *Toxicol. Appl. Pharmacol.*, **2005**. *205*, 237-246.

65. Tang, J.; Usmani, K.A.; Hodgson, E.; Rose, R.L. In vitro metabolism of fipronil by human and rat cytochrome P450 and its interactions with testosterone and diazepam. *Chemico-Biol. Interactions.*, **2004**. *147*, 319-329.

66. Bergmann, A.J.; North, P.E.; Vasquez, L.; Bello, H.; Del Carmen Gastañaga Ruiz, M.; Anderson, K.A. Multi-class chemical exposure in rural Peru using silicone wristbands. *J. Expo. Sci. Environ. Epidemiol.* **2017**. *27*, 560–568.

67. Roberts, J.R.; Karr, C.J. Council on environmental health. Pesticide exposure in children [published correction appears in Pediatrics 131, 1013-1014]. *Pediatrics*, **2012**. *130*, 1765–1788.

68. Crapeau, M.; Merhi, A.; André, B. Stress conditions promote yeast Gap1 permease ubiquitylation and down-regulation via the arrestin-like Bul and Aly proteins. *J. Biol. Chem.* **2014**. *289*, 22103–222116.

69. Pawar, H.N.; Kumar, R.; Narang, R.; Agrawal, R.K. Heat and cold stress enhances the expression of the heat shock protein 70, heat shock transcription factor 1 and cytokines (IL-12, TNF-α, and GMCSF) in buffaloes. *Intl. J. Current Microbiol. Appl. Sci*. **2014**. *3*, 307–317.

70. Abd-Ella, A.; Stankiewicz, M.; Mikulska, K. The repellent DEET potentiates carbamate effects via insect muscarinic receptor interactions: an alternative strategy to control insect vector-borne diseases. *PLoS One*. **2015**. *10*, e0126406.

71. Pisa, L.W.; Amaral-Rogers, V.; Belzunces, L.P.; Bonmatin, J.M.; Downs, C.A.; Goulson, D.; Kreutzweiser, D.P.; Krupke, C.; Liess, M.; McField, M.; Morrissey, C.A.; Noome, D.A.; Settele, J.; Simon-Delso, N.; Stark, J.D.; Van der Sluijs, J.P.; Van Dyck, H.; Wiemers, M. Effects of neonicotinoids and fipronil on non-target invertebrates. *Environ. Sci. Pollut. Res. Intl.* **2014**. *22*, 68–102.

72. Milligan, M.J.; Harvey, E.; Yu, A.; Morgan, A.L.; Smith, D.L.; Zhang, E.; Berengut, J.; Sivananthan, J.; Subramaniam, R.; Skoric, A. Global intersection of long non-coding RNAs with processed and unprocessed pseudogenes in the human genome. *Front. Genet*. **2016**. *7*, 26-26.

73. Reynolds, A.; Leake, D.; Boese, Q.; Scaringe, S.; Marshall, W.S.; Khvorova, A. Rational siRNA design for RNA interference. *Nature Biotech.* **2004**. *22*, 326–330.

74. Huang, J.; Zhou, N.; Watabe, K. Long non-coding RNA UCA1 promotes breast tumor growth by suppression of p27 (Kip1). *Cell Death Dis.* **2014**. *5*, e1008.

FIGURES AND TABLES

TABLES

Table 1. Effect of 100 µM DEET treatment on transcript levels for Phase 1 and 2 metabolism in primary human hepatocytes[a].

Phase	Enzyme	Gene Symbol[b]	Chromosome[c]	Log 2-Fold Change[d]
1	P450s	CYP4A11	chr1:47394845-47407156	-1.11
		CYP4A22	chr1:47603106-47614526	-1.07
		CYP3A5	chr7:99245812-99277621	+1.15
		CYP3A7	chr7:99282301-99332819	+3.39
		CYP3A43	chr7:99425635-99464173	+3.25
		CYP2C9	chr10:96698414-96749148	+0.93
		CYP2C8	chr10:96796528-96829254	+2.29
		CYP1A1	chr15:75011882-75017877	+1.82
		CYP1A2	chr15:75041183-75048941	+1.23
		CYP2A6	chr19:41349442-41356352	+2.21
		CYP2A7	chr19:41381343-41388657	+2.44
		CYP2B6	chr19:41497203-41524301	+3.82
		CYP2B7P1	chr19:41430169-41456565	+2.97
		CYP2A13	chr19:41594355-41602100	+2.34
	aldehyde dehydrogenases	ALDH1B1	chr9:38392660-38398662	+0.66
		ALDH3A1	chr17:19641297-19651746	+2.76
	epoxide hydrolases	EPHX1	chr1:225997796-226070420	+2.30
2	glutathione-S-transferases	GSTA2	chr6:52614884-52628361	+2.51
		GSTA1	chr6:52656177-52668664	+1.90

[a] Transcripts that were not significantly up- or down-regulated were not included. Phase 3 enzymes are also not included.
[b] Gene symbol annotations are as follows: CYP=Cytochrome P450, chr=Chromosome, ALDH=Aldehyde Dehydrogenase, EPHX= Epoxide Hydrolase, GST=Glutathione-S-Transferase.
[c] Chromosomal coordinates (chromosome + transcriptional start and stop sites) and gene symbols pair with maps illustrated in Figure 2A-C.
[d] Log 2-fold change was statistically significant (α=0.05), (+) indicates up-regulation and (–) indicates down-regulation. Log 2-fold change details magnitude of gene expression.

Table 2. Effect of 10 µM fipronil treatment on transcript levels for Phase 1 metabolism in primary human hepatocytes[a].

Phase	Enzyme	Gene Symbol[b]	Chromosome[c]	Log 2-Fold Change[d]
1	P450s	CYP4A11	chr1:47394845-47407156	-2.46
		CYP4A22	chr1:47603106-47614526	-2.08
		CYP27C1	chr2:127941411-127963343	-0.85
		CYP20A1	chr2:204103163-204170563	-0.56
		CYP26B1	chr2:72356366-72374963	+2.15
		CYP8B1	chr3:42913683-42917633	-0.84
		CYP4V2	chr4:187112673-187134617	-0.46
		CYP21A2	chr6:32006092-32077151	+0.88
		CYP51A1	chr7:91741462-91764059	+0.62
		CYP3A5	chr7:99245812-99277621	+1.17
		CYP3A7	chr7:99282301-99332819	+3.71
		CYP3A4	chr7:99425635-99464173	+3.86
		CYP3A43	chr7:99425635-99464173	+2.27
		CYP26A1	chr10:94833231-94837641	+3.33
		CYP2C9	chr10:96698414-96749148	+0.75
		CYP2C8	chr10:96796528-96829254	+2.39
		CYP2E1	chr10:135340866-135352620	-1.59
		CYP1A1	chr15:75011882-75017877	+1.09
		CYP4F22	chr19:15619335-15663128	+1.25
		CYP2A6	chr19:41349442-41356352	+1.31
		CYP2B7P	chr19:41430169-41456565	+2.42
		CYP2B6	chr19:41497203-41524301	+3.54
		CYP2D7P	chr22:42536213-42540575	+0.39
	FMOs	FMO3	chr1:171060017-171086959	-0.73
		FMO1	chr1:171217662-171255113	-1.2
		FMO4	chr1:171283485-171311223	-0.37
	aldehyde dehydrogenases	ALDH4A1	chr1:19197923-19229293	-0.37
		ALDH1L1	chr3:125822403-125929011	-3.2
		ALDH5A1	chr6:24495196-24537435	-0.41
		ALDH8A1	chr6:135238527-135271260	-1.07
		ALDH1A1	chr9:75515577-75568233	+0.55
		ALDH3A1	chr17:19641297-19651746	+1.28
	alcohol dehydrogenases	ADH1B	chr4:100227526-100242572	-3.35
		ADH1C	chr4:100257648-100273917	-1.37
		ADH6	chr4:100010007-100222513	-1.67
		ADH4	chr4:100010007-100222513	-2.09
		ADHFE1	chr8:67344717-67381044	-0.59
	epoxide hydrolases	EPHX1	chr1:225997796-226070420	+1.72
		EPHX2	chr8:27348518-27402439	-0.76

Table 2 (continued).

Phase	Enzyme	Gene Symbol[b]	Chromosome[c]	Log 2-Fold Change[d]
2	glutathione-S-transferases	GSTM2	chr1:110210643-110226619	-0.76
		GSTM1	chr1:110230417-110236367	+0.34
		GSTM3	chr1:110276553-110283660	+0.62
		GSTA2	chr6:52614884-52628361	+2.94
		GSTA1	chr6:52656177-52668664	+2.50
		GSTT2	chr22:24322313-24326106	+1.69
	N-acetyltransferases	NAT8B	chr2:73927635-73928467	-1.76
		NAT8	chr2:73927635-73928467	+0.36
		NAT9	chr17:72766685-72772470	-0.34
	glucuronidases	GUSBP3	chr5:68935289-69006272	-2.59
		GUSBP2	chr6:26839265-26924333	-0.61
	sulfatases	SULF1	chr8:70378858-70573147	+2.20
		GNS	chr12:65107221-65153226	+0.45

[a]Transcripts that were not significantly up- or down-regulated were not included. Phase 3 enzymes are also not included.
[b]Gene symbol annotations are as follows: chr=Chromosome, GST=Glutathione-S-Transferase, NAT=N-Acetyltransferase, GUS=Glucuronidase, SULF=Sulfatase, GNS= Sulfatase
[c] Chromosomal coordinates (chromosome + transcriptional start and stop sites) and gene symbols pair with maps illustrated in Figure 2A-C.
[d]Log 2-fold change was statistically significant (α=0.05), (+) indicates up-regulation and (–) indicates down-regulation. Log 2-fold change details magnitude of gene expression.

Table 3. Effect of 100 µM DEET plus 10 µM fipronil treatment on transcript levels for Phase 1 metabolism in primary human hepatocytes[a].

Phase	Enzyme	Gene Symbol[b]	Chromosome[c]	Log 2-Fold Change[d]
1	P450s	CYP4A11	chr1:47394845-47407156	-3.61
		CYP4A22	chr1:47603106-47614526	-2.69
		CYP27C1	chr2:127941411-127963343	-1.38
		CYP20A1	chr2:204103163-204170563	-0.66
		CYP26B1	chr2:72356366-72374963	+2.41
		CYP8B1	chr3:42913683-42917633	-1.01
		CYP4V2	chr4:187112673-187134617	-0.56
		CYP21A2	chr6:32006092-32077151	+0.76
		CYP3A5	chr7:99245812-99277621	+1.21
		CYP3A7	chr7:99282301-99332819	+3.57
		CYP3A4	chr7:99354582-99381811	+3.72
		CYP3A43	chr7:99425635-99464173	+1.47
		CYP26A1	chr10:94833231-94837641	+4.23
		CYP2C19	chr10:96522462-96612671	-0.35
		CYP2C9	chr10:96698414-96749148	+0.62
		CYP2C8	chr10:96796528-96829254	+2.04
		CYP2E1	chr10:135340866-135352620	-1.87
		CYP46A1	chr14:100150754-100193638	-0.98
		CYP1A1	chr15:75011882-75017877	+2.02
		CYP1A2	chr15:75041183-75048941	+1.05
		CYP2A6	chr19:41349442-41356352	+1.75
		CYP2A7	chr19:41381343-41388657	+1.09
		CYP2B7P	chr19:41430169-41456565	+2.79
		CYP2B6	chr19:41497203-41524301	+3.84
		CYP4F22	chr19:15619335-15663128	+0.95
		CYP4F12	chr19:15783827-15807984	-0.38
		CYP4F2	chr19:15988833-16008884	-0.36
	FMOs	FMO5	chr1:146655883-146697230	-0.62
		FMO3	chr1:171060017-171086959	-0.77
		FMO1	chr1:171217662-171255113	-1.76
		FMO4	chr1:171283485-171311223	-0.65
	epoxide hydrolases	EPHX1	chr1:225997796-226070420	+1.76
		EPHX2	chr8:27348518-27402439	-0.89
	alcohol dehydrogenases	ADH6	chr4:100010007-100222513	-1.75
		ADH4	chr4:100010007-100222513	-1.82
		ADH1A	chr4:100010007-100222513	-3.34
		ADH1B	chr4:100227526-100242572	-3.56
		ADH1C	chr4:100257648-100273917	-1.39
		ADHFE1	chr8:67344717-67381044	-0.71
	aldehyde dehydrogenases	ALDH1L1	chr3:125822403-125929011	-0.58
		ALDH5A1	chr6:24495196-24537435	-0.52
		ALDH8A1	chr6:135238527-135271260	-1.49
		ALDH1A1	chr9:75515577-75568233	+0.73
		ALDH2	chr12:112204690-112247789	-0.42
		ALDH6A1	chr14:74486058-74551196	-0.36
		ALDH3A1	chr17:19641297-19651746	+2.08

Table 3 (continued).

Phase	Enzyme	Gene Symbol[b]	Chromosome[c]	Log 2-Fold Change[d]
2	glutathione-S-transferases	GSTM2	chr1:110210643-110226619	-1.02
		GSTM1	chr1:110230417-110236367	+0.51
		GSTM3	chr1:110276553-110283660	+0.69
		MGST3	chr1:165600109-165625372	+0.41
		GSTA2	chr6:52614884-52628361	+3.80
		GSTA1	chr6:52656177-52668664	+2.99
		GSTO2	chr10:106028630-106059176	-0.59
		GSTT2	chr22:24322313-24326106	+1.63
	glucuronidases	GUSBP3	chr5:68935289-69006272	-2.09
		SMA4	chr5:69423288-69586004	-0.98
		GUSBP9	chr5:69776869-69881549	-0.55
		GUSBP2	chr6:26839265-26924333	-1.17
		GUSBP11	chr22:23980674-24059610	-0.57
	sulfatases	SUMF1	chr3:4402828-4508966	+0.33
		SULF1	chr8:70378858-70573147	+1.79
		GNS	chr12:65107221-65153226	+0.42
	N-acetyltransferases	NAT8B	chr2:73927635-73928467	-2.22
		NAT6	chr3:50330258-50336899	-0.45
		GNPNAT1	chr14:53241910-53258386	+0.36
	paraoxonases	PON1	chr7:94927668-94953884	-0.69

[a]Transcripts that were not significantly up- or down-regulated were not included. Phase 3 enzymes are also not included.
[b]Gene symbol annotations are as follows: CYP=Cytochrome P450, chr=Chromosome, ALDH=Aldehyde Dehydrogenase, EPHX= Epoxide Hydrolase, ADH=Alcohol Dehydrogenase, FMO=Flavin-containing Monooxygenase.
[c] Chromosomal coordinates (chromosome + transcriptional start and stop sites) and gene symbols pair with maps illustrated in Figure 2A-C.
[d]Log 2-fold change was statistically significant (α=0.05), (+) indicates up-regulation and (–) indicates down-regulation. Log 2-fold change details magnitude of gene expression.

FIGURES

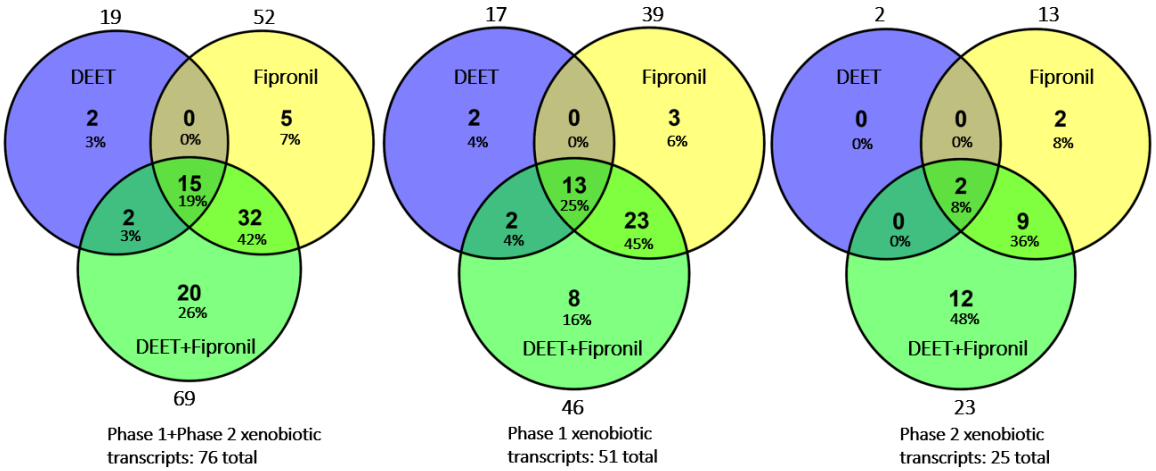


Figure 1. Total number of altered transcripts associated with possible phase 1 and 2 metabolism in primary human hepatocytes treated for 72 h with (A) 100 µM DEET, (B) 10 µM fipronil, or (C) a mixture of 100 µM DEET and 10 µM fipronil. From left to right, the Venn diagrams illustrate phase 1 + phase 2 transcripts, only phase 1 transcripts, and only phase 2 transcripts. Numbers on the outside of each circle represent the total number of transcripts altered for that particular treatment. Numbers and percentages on the inside of each circle represent the number and percentage of transcripts for each treatment that were only found in one treatment, two treatments, or all three of the treatments.

Figure 2A-C. Chromosomal map of altered transcripts possibly involved in xenobiotic metabolism for the treatment of primary human hepatocytes with 100 µM DEET (A), 10 µM Fipronil (B), and a mixture of 100 µM DEET and 10 µM fipronil (C). Chromosomal maps were made using the program, Idiographica (Kin and Ono, 2007). The maps display the presence and location of significantly (α=0.05) up-regulated (green) or down-regulated (red) transcripts. Green lines indicate and display the chromosomal locations of up-regulated transcripts. Red lines indicate and display the chromosomal locations of down-regulated transcripts. Included in Figure 2A-C are also those transcripts (gray) that were detected by RNASeq but not significantly altered (α<0.05). Each chromosomal map also shows the overall number and percentage of up-regulated and down-regulated transcripts.

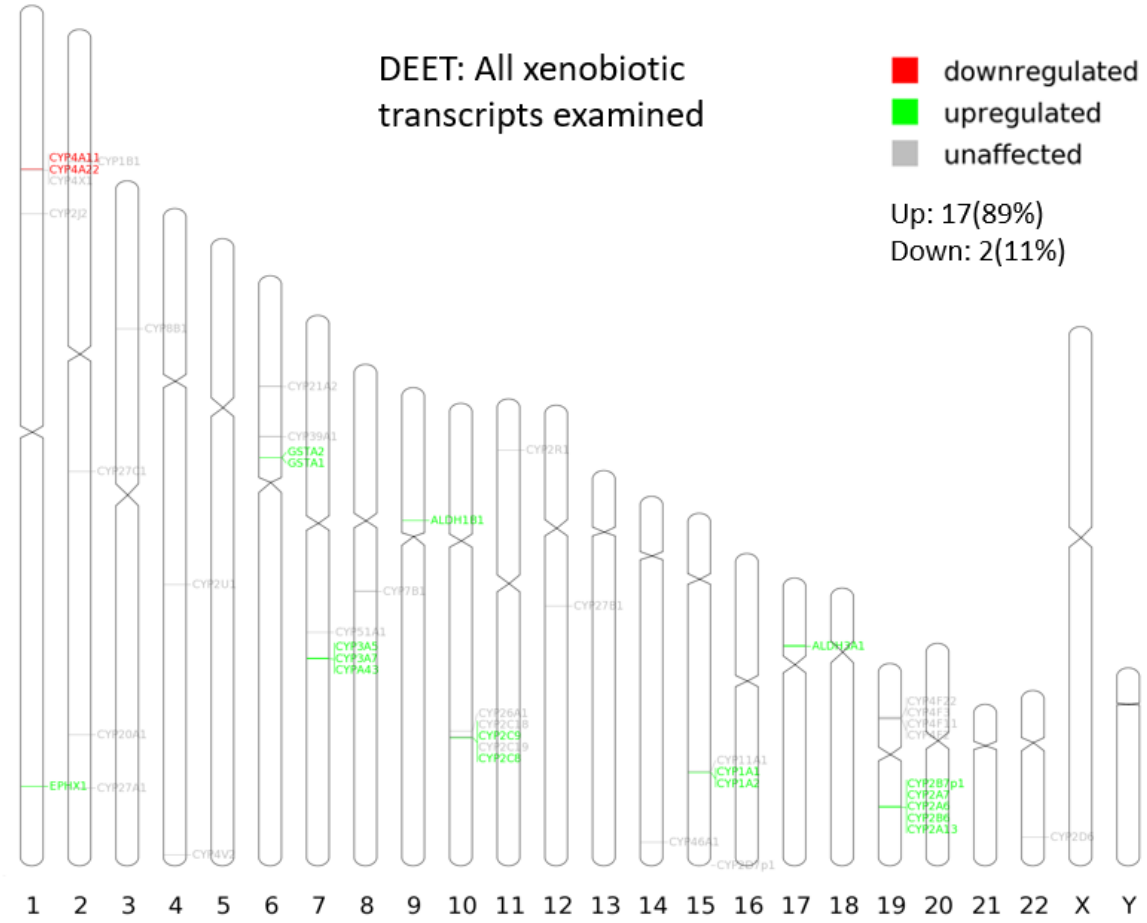

DEET: All xenobiotic transcripts examined
downregulated
upregulated
unaffected
Up: 17(89%)
Down: 2(11%)
CYP4A11
CYP4A22
CYP1B1
CYP4X1
CYP2J2
CYP8B1
CYP27C1
CYP21A2
CYP39A1
GSTA2
GSTA1
CYP2U1
CYP7B1
CYP51A1
CYP3A5
CYP3A7
CYP3A43
ALDH1B1
CYP2R1
CYP27B1
CYP20A1
EPHX1
CYP27A1
CYP4V2
CYP26A1
CYP2C18
CYP2C9
CYP2C8
CYP46A1
CYP11A1
CYP1A1
CYP1A2
CYP2B7P1
ALDH3A1
CYP4F22
CYP4F3
CYP4F11
CYP2B7p1
CYP2A7
CYP2A6
CYP2B6
CYP2A13
CYP2D6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y

Figure 2A.

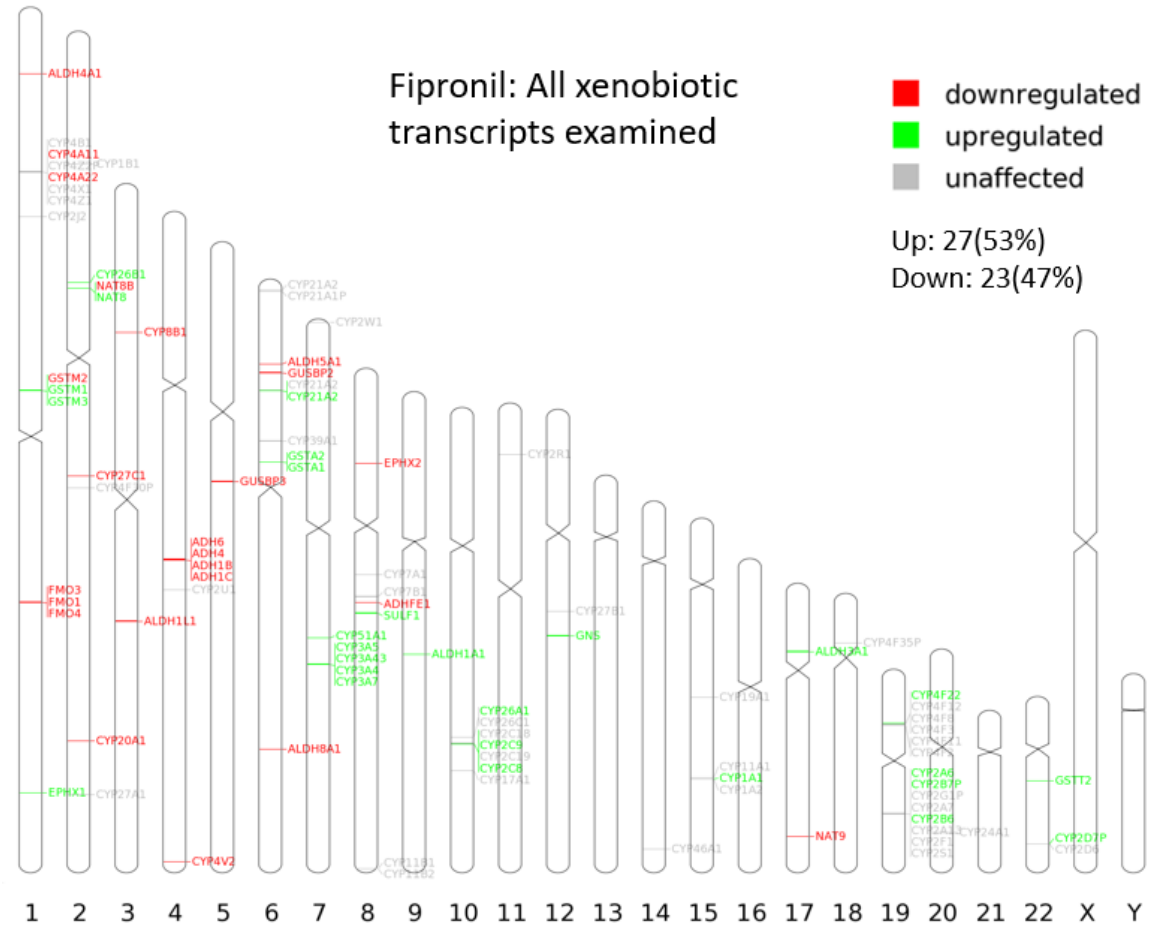

Fipronil: All xenobiotic transcripts examined
downregulated
upregulated
unaffected
Up: 27(53%)
Down: 23(47%)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y

Figure 2B.

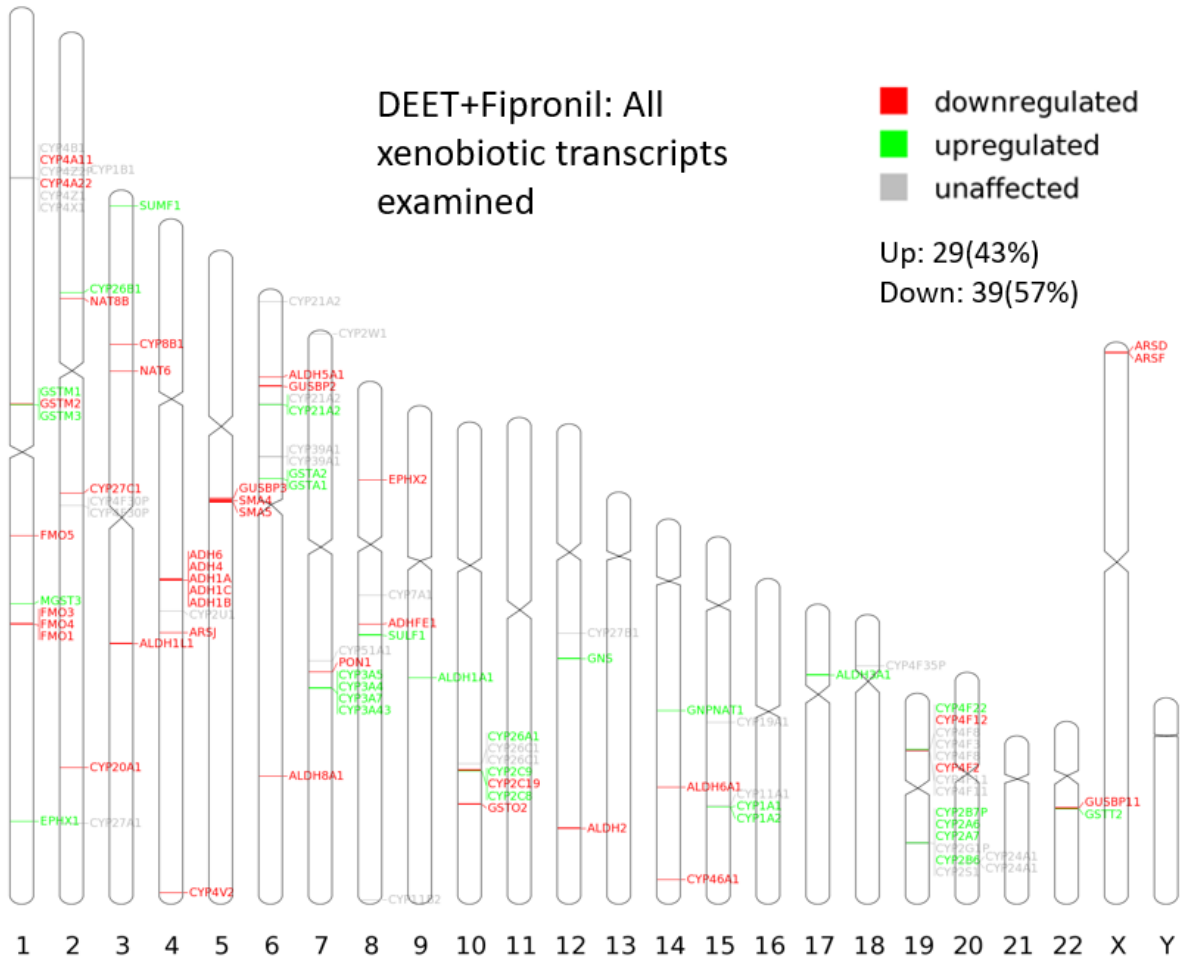

DEET+Fipronil: All xenobiotic transcripts examined
downregulated
upregulated
unaffected
Up: 29(43%)
Down: 39(57%)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y

Figure 2C.

Figure 3A-C. lncRNAs altered by treatment of primary human hepatocytes with 100 µM DEET (A), 10 µM Fipronil (B), and a mixture of 100 µM DEET and 10 µM fipronil (C) that are in significant proximity to P450s also altered by the same treatment. Significant proximity was determined by the GREAT algorithm (McLean et al. 2010). Base-pair (bp) coordinate numbers for P450s and lncRNAs are included to calculate base-pair distance. Green indicates a coding gene for a P450 while blue indicates a lncRNA. The stars (blue=lncRNA, green=P450) indicate which P450 and lncRNA on the chromosome are illustrated with gene location and coordinates to the right. Chromosome number and respective arm of the chromosome (q or y) is also included on the left. Arrows indicate direction of transcription. Gene symbols for coding-genes are as follows: CYP=cytochrome P450.

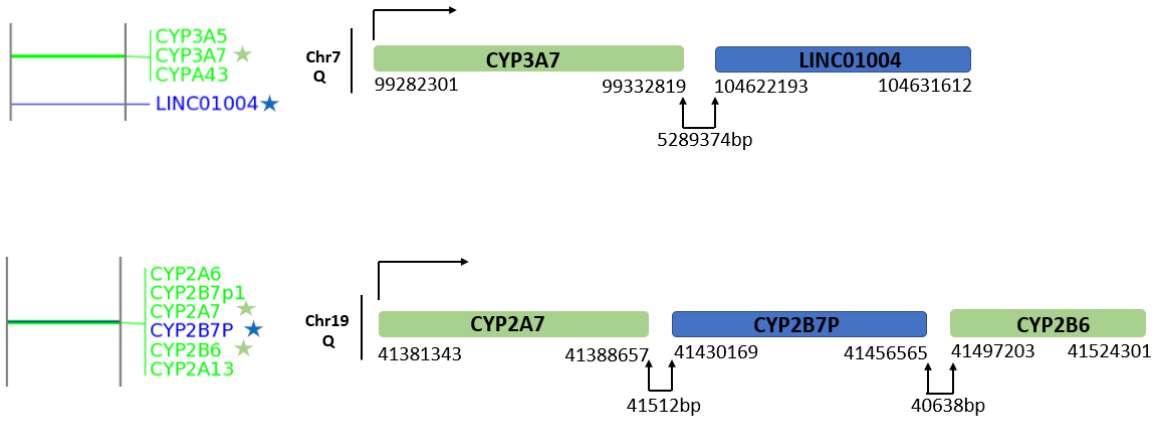


Figure 3A.

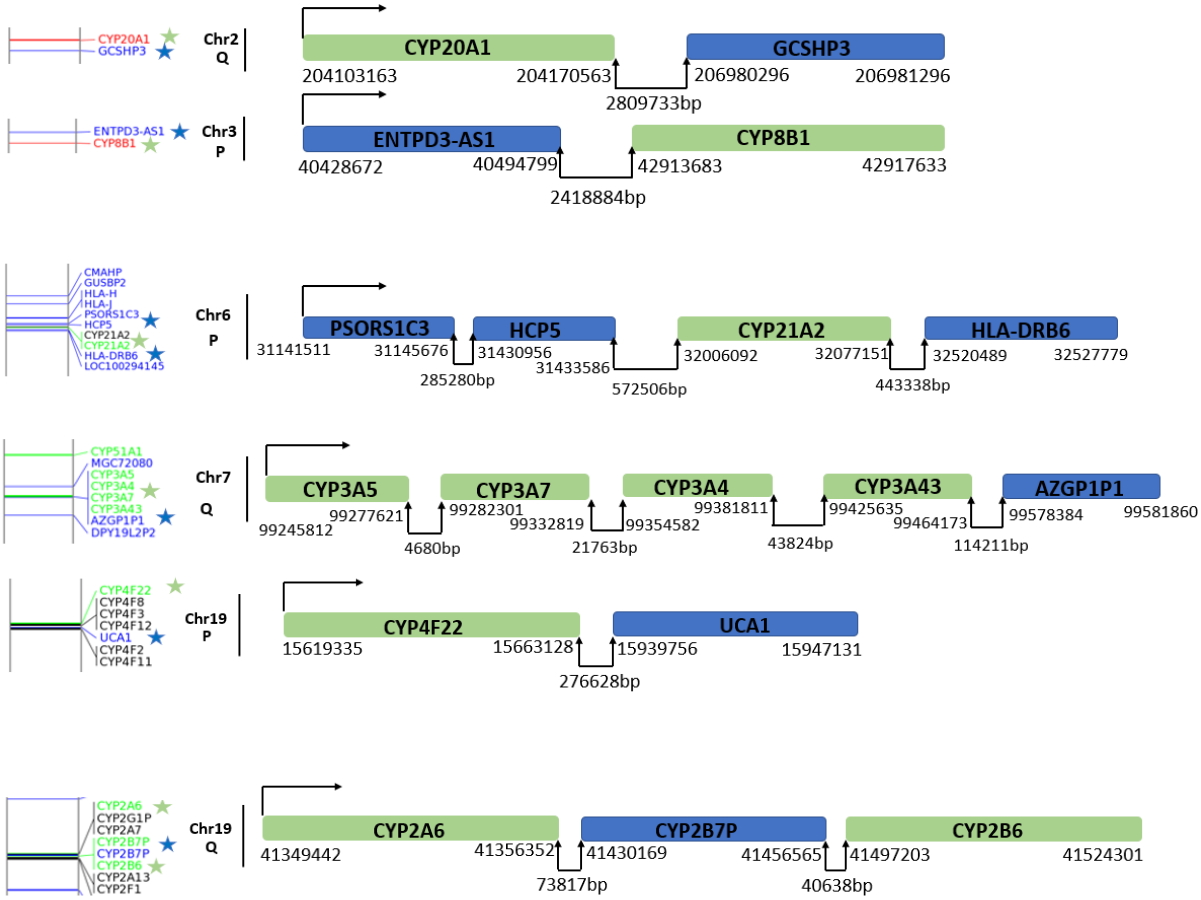


Figure 3B.

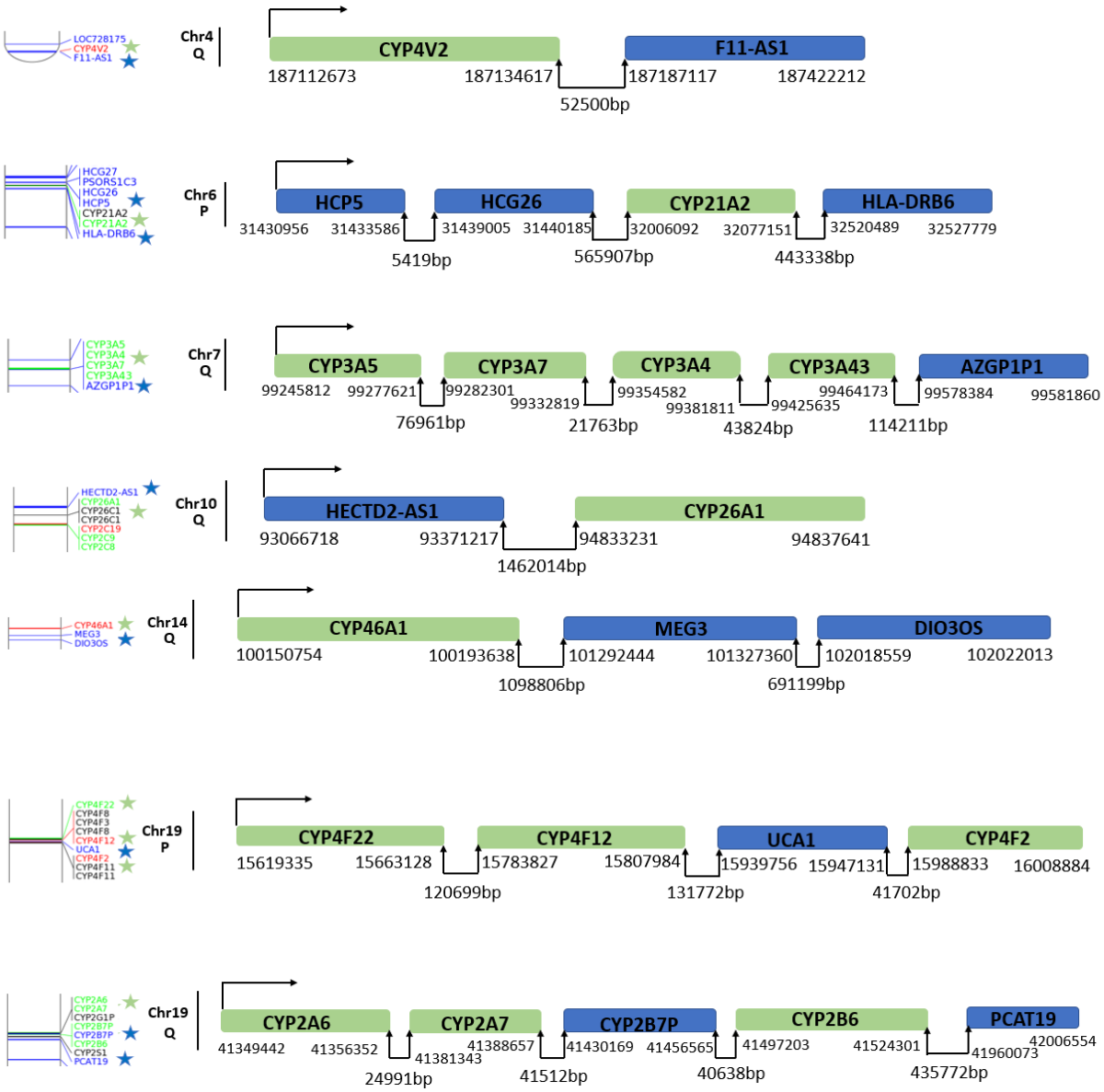

LOC728175
CYP4V2
F11-AS1
Chr4
Q
CYP4V2
F11-AS1
187112673
187134617
187187117
187422212
52500bp
HCG27
PSORS1C3
HCG26
HCP5
CYP21A2
HLA-DRB6
Chr6
P
HCP5
HCG26
CYP21A2
HLA-DRB6
31430956
31433586
31439005
31440185
32006092
32077151
32520489
32527779
5419bp
565907bp
443338bp
CYP3A5
CYP3A4
CYP3A7
CYP3A43
AZGP1P1
Chr7
Q
CYP3A5
CYP3A7
CYP3A4
CYP3A43
AZGP1P1
99245812
99277621
99282301
99332819
99354582
99381811
99425635
99464173
99578384
99581860
76961bp
21763bp
43824bp
114211bp
HECTD2-AS1
CYP26A1
CYP26C1
CYP26C1
CYP2C19
CYP2C9
CYP2C8
Chr10
Q
HECTD2-AS1
CYP26A1
93066718
93371217
94833231
94837641
1462014bp
CYP46A1
MEG3
DIO3OS
Chr14
Q
CYP46A1
MEG3
DIO3OS
100150754
100193638
101292444
101327360
102018559
102022013
1098806bp
691199bp
CYP4F22
CYP4F8
CYP4F3
CYP4F8
CYP4F12
UCA1
CYP4F2
CYP4F11
CYP4F11
Chr19
P
CYP4F22
CYP4F12
UCA1
CYP4F2
15619335
15663128
15783827
15807984
15939756
15947131
15988833
16008884
120699bp
131772bp
41702bp
CYP2A6
CYP2A7
CYP2G1P
CYP2B7P
CYP2B6
CYP2S1
PCAT19
Chr19
Q
CYP2A6
CYP2A7
CYP2B7P
CYP2B6
PCAT19
41349442
41356352
41381343
41388657
41430169
41456565
41497203
41524301
41960073
42006554
24991bp
41512bp
40638bp
435772bp

Figure 3C.

Figure 4A-C. lncRNAs altered by treatment of primary human hepatocytes with 10 µM fipronil (A) and a mixture of 100 µM DEET and 10 µM fipronil (B) that are in significant proximity to non-P450 metabolic transcripts also altered by the same treatment. No significant relationships were found for the 100 µM DEET treatment. Significant proximity was determined by the GREAT algorithm (McLean et al. 2010). Base-pair (bp) coordinate numbers for transcripts potentially involved in xenobiotic metabolism (other than P450s) and lncRNAs are included to calculate base-pair distance. Green indicates a coding gene for a transcript for a metabolic gene while blue indicates a lncRNA. The stars (blue=lncRNA, green=coding-gene) indicate which genes on the chromosome (on the left) are illustrated on the right with a gene symbol and chromosomal coordinates. Chromosome number and respective arm of the chromosome (q or y) is also included on the left. Arrows indicate direction of transcription. Gene symbols for coding-genes are as follows: ADH=alcohol dehydrogenase, ALDH=aldehyde dehydrogenase, GST=glutathione-S-transferase, GUS/SMA=glucuronidase, FMO=flavin-containing monooxygenase, NAT=N-acetyltransferase, SULF/GNS/ARS=sulfatase, PON=Paraoxonase.

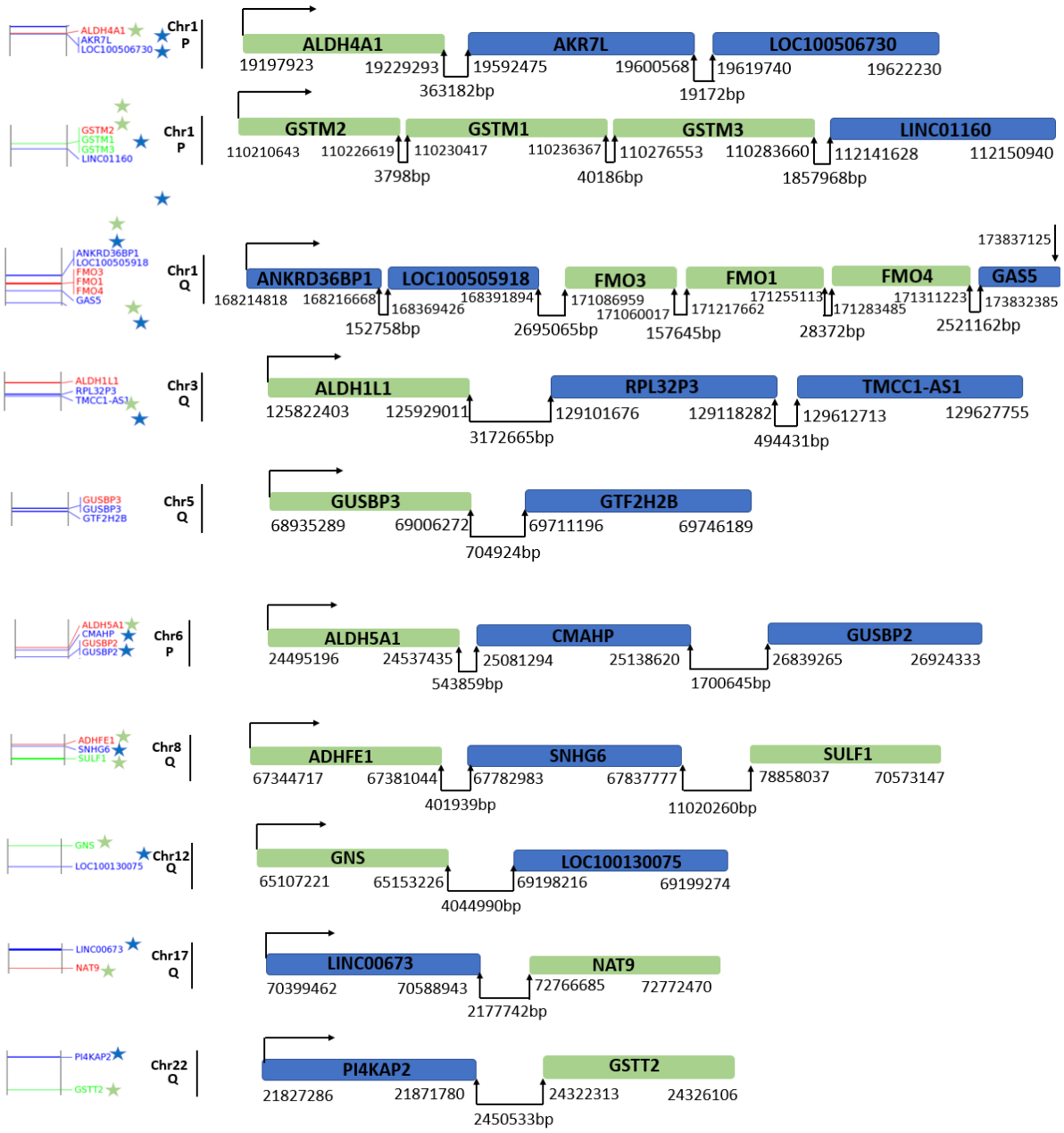

Chr1 P
ALDH4A1
AKR7L
LOC100506730
19197923
19229293
19592475
19600568
19619740
19622230
363182bp
19172bp
Chr1 P
GSTM2
GSTM1
GSTM3
LINC01160
110210643
110226619
110230417
110236367
110276553
110283660
112141628
112150940
3798bp
40186bp
1857968bp
Chr1 Q
ANKRD36BP1
LOC100505918
FMO3
FMO1
FMO4
GAS5
173837125
168214818
168216668
168369426
168391894
171086959
171060017
171217662
171255113
171283485
171311223
173832385
152758bp
2695065bp
157645bp
28372bp
2521162bp
Chr3 Q
ALDH1L1
RPL32P3
TMCC1-AS1
125822403
125929011
129101676
129118282
129612713
129627755
3172665bp
494431bp
Chr5 Q
GUSBP3
GTF2H2B
68935289
69006272
69711196
69746189
704924bp
Chr6 P
ALDH5A1
CMAHP
GUSBP2
24495196
24537435
25081294
25138620
26839265
26924333
543859bp
1700645bp
Chr8 Q
ADHFE1
SNHG6
SULF1
67344717
67381044
67782983
67837777
78858037
70573147
401939bp
11020260bp
Chr12 Q
GNS
LOC100130075
65107221
65153226
69198216
69199274
4044990bp
Chr17 Q
LINC00673
NAT9
70399462
70588943
72766685
72772470
2177742bp
Chr22 Q
PI4KAP2
GSTT2
21827286
21871780
24322313
24326106
2450533bp

Figure 4A.

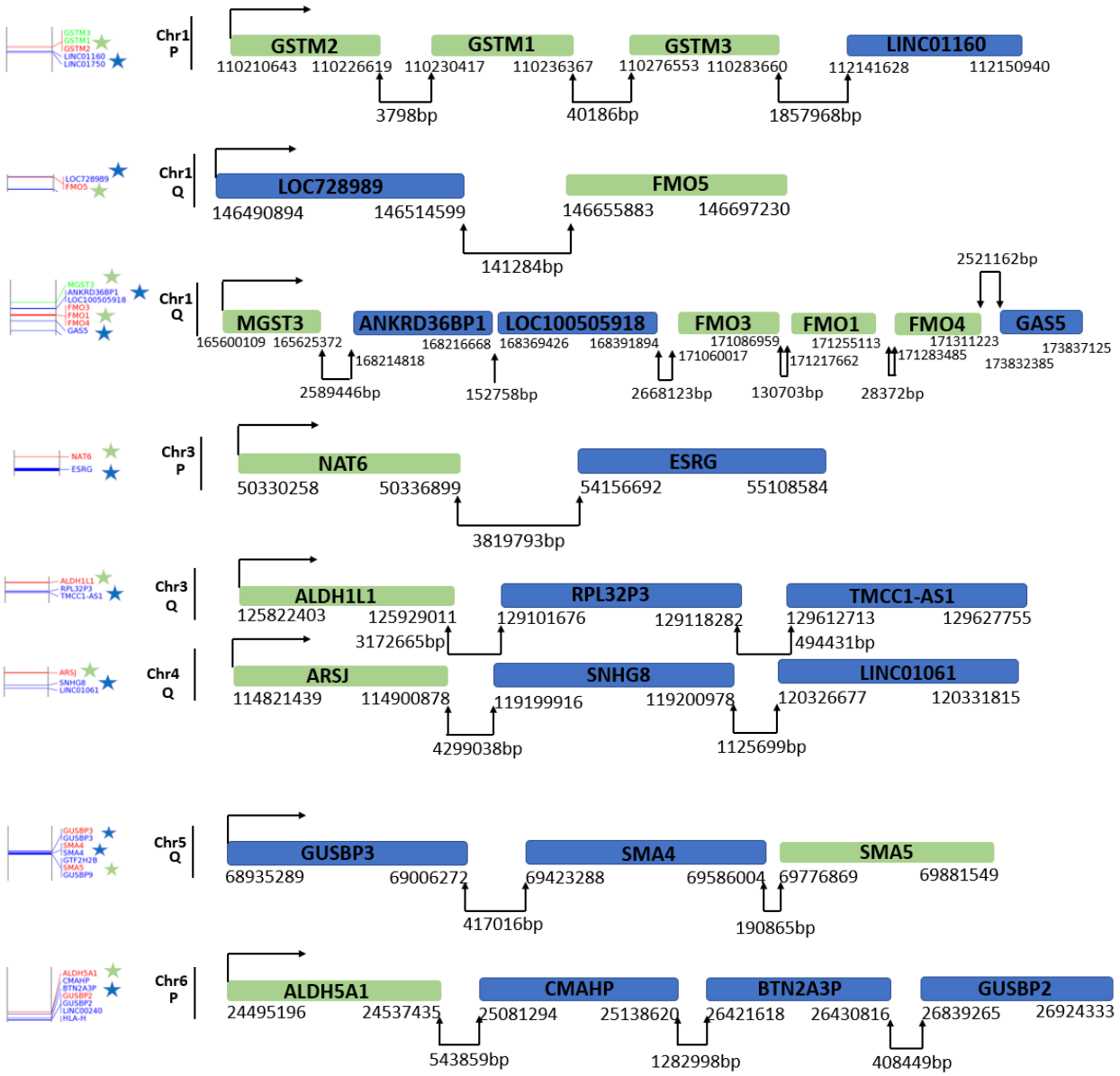

Chr1 P
GSTM2
GSTM1
GSTM3
LINC01160
110210643 110226619 110230417 110236367 110276553 110283660 112141628 112150940
3798bp
40186bp
1857968bp
Chr1 Q
LOC728989
FMO5
146490894 146514599 146655883 146697230
141284bp
Chr1 Q
MGST3
ANKRD36BP1
LOC100505918
FMO3
FMO1
FMO4
GAS5
165600109 165625372 168214818 168216668 168369426 168391894 171060017 171086959 171217662 171255113 171283485 171311223 173832385 173837125
2521162bp
2589446bp
152758bp
2668123bp
130703bp
28372bp
Chr3 P
NAT6
ESRG
50330258 50336899 54156692 55108584
3819793bp
Chr3 Q
ALDH1L1
RPL32P3
TMCC1-AS1
125822403 125929011 129101676 129118282 129612713 129627755
3172665bp
494431bp
Chr4 Q
ARSJ
SNHG8
LINC01061
114821439 114900878 119199916 119200978 120326677 120331815
4299038bp
1125699bp
Chr5 Q
GUSBP3
SMA4
SMA5
68935289 69006272 69423288 69586004 69776869 69881549
417016bp
190865bp
Chr6 P
ALDH5A1
CMAHP
BTN2A3P
GUSBP2
24495196 24537435 25081294 25138620 26421618 26430816 26839265 26924333
543859bp
1282998bp
408449bp

Figure 4B.

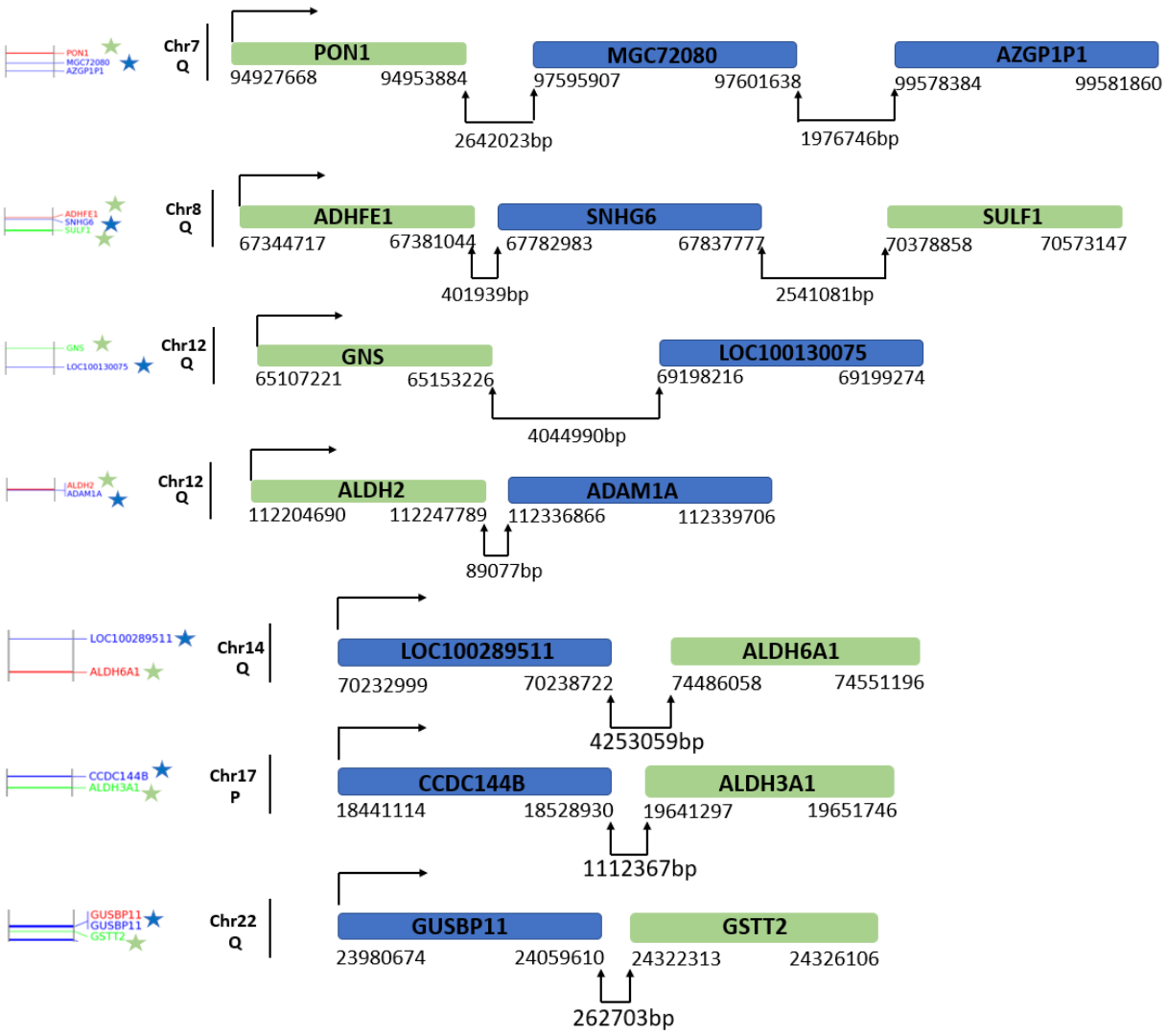

Chr7
Q
PON1
MGC72080
AZGP1P1
94927668
94953884
97595907
97601638
99578384
99581860
2642023bp
1976746bp
Chr8
Q
ADHFE1
SNHG6
SULF1
67344717
67381044
67782983
67837777
70378858
70573147
401939bp
2541081bp
Chr12
Q
GNS
LOC100130075
65107221
65153226
69198216
69199274
4044990bp
Chr12
Q
ALDH2
ADAM1A
112204690
112247789
112336866
112339706
89077bp
Chr14
Q
LOC100289511
ALDH6A1
70232999
70238722
74486058
74551196
4253059bp
Chr17
P
CCDC144B
ALDH3A1
18441114
18528930
19641297
19651746
1112367bp
Chr22
Q
GUSBP11
GSTT2
23980674
24059610
24322313
24326106
262703bp

Figure 4C.

lncRNA Alignment	Percent Identity
MGC72080 x AZGP1P1	91%
CYP2B7P x PCAT19	84%
AZGP1P1 x PCAT19	82%
AZGP1P1 x CYP2B7P	89%
MGC72080 x PCAT19	84%
MGC72080 x CYP2B7P	90%
HECTD2-AS1 x AZGP1P1	91%
HECTD2-AS1 x MGC72080	91%
HECTD2-AS1 x PCAT19	94%
HECTD2-AS1 x CYP2B7P	90%
UCA1 x CYP2B7P	No significant alignment
UCA1 x PCAT19	No significant alignment
UCA1 x AZGP1P1	No significant alignment
UCA1 x MGC72080	No significant alignment
UCA1 x HECTD2-AS1	No significant alignment

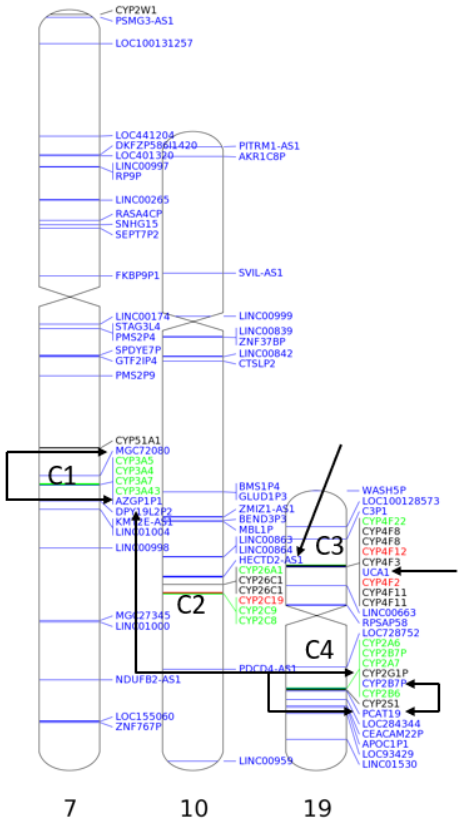


Figure 5. Percent identity (BLAST alignments; Altschul et al., 1990) between different altered lncRNA genes found within 10^6 bp to different altered P450 clusters (cluster = 3 or more P450s per 10^6 bp) for the 100 µM DEET + 10 µM fipronil treatment of primary human hepatocytes. The P450 clusters (C1-4) are shown on chromosomes 7, 10, and 19. Gene symbols in blue indicate a lncRNA. Green and red gene symbols correspond to P450 transcripts that were either up- or down-regulated, respectively. The arrows indicate the chromosomal locations of those lncRNAs that were compared to one another (see table on left). The table displays the percent identity for each comparison. No significant alignment indicates that alignment values are below the required threshold to be considered significant.

Chapter 2: Multiple Known Mechanisms and a Possible Role of an Enhanced Immune System in Bt-Resistance in a Field Population of the Bollworm, *Helicoverpa zea*: Differences in Gene Expression with RNAseq

Abstract

Several different agricultural insect pests have developed field resistance to Bt (*Bacillus thuringiensis*) proteins (ex. Cry1Ac, Cry1F, etc.) expressed in crops including corn and cotton. In the bollworm, *Helicoverpa zea*, resistance levels are increasing; recent reports in 2019 show up to 1000-fold levels of resistance to Cry1Ac, a major insecticidal protein in Bt-crops. A common method to analyze global differences in gene expression is RNA-seq. This technique was used to measure differences in global gene expression between a Bt-susceptible and Bt-resistant strain of the bollworm, where the differences in susceptibility to Cry1Ac insecticidal proteins were 100-fold. We found expected gene expression differences based on our current understanding of the Bt mode of action including increased expression of proteases (trypsins and serine proteases) and reduced expression of Bt-interacting receptors (aminopeptidases and cadherins) in resistant bollworms. We also found additional expression differences for transcripts that were not previously investigated, i.e, transcripts from 3 immune pathways--Jak/STAT, Toll, and IMD. Immune pathway receptors (ex. PGRPs) and the IMD pathway demonstrated the highest differences in expression. Our analysis suggested that multiple mechanisms are involved in the development of Bt-resistance, including potentially unrecognized pathways.

Introduction

The bollworm, *Helicoverpa zea* (Lepidoptera: Noctuidae), is a pest of cotton in the United States (US) [1]. Besides cotton, *H. zea* is an economically important pest of corn (also commonly named the corn earworm and tobacco fruitworm), sorghum, wheat, soybeans, and other crops [2,3]. Integrated Pest Management (IPM) uses several different management techniques to control the bollworm. However, an essential component of cotton IPM has been the use of insecticides [2,4]. Because of the intensive use of foliar, chemical insecticides in cotton for many decades to control insect pests, there has been significant public and industry pressure to develop alternative control methods to minimize the environmental impact of cotton production. The outcome was the development of Bt-cotton expressing proteins from bacteria that are toxic to insects; this technology is widely used today for pest control in row crops.

Transgenic crops have been an effective alternative to chemical insecticides in IPM. Transgenic Bt-cotton, as one example, has been used commercially for over two decades (since 1996) to control caterpillars (including *H. zea*, the focus of this paper) [3]. Transcripts coding for insecticidal proteins from the bacteria, *Bacillus thuringiensis* (Bt), Cry1Ac, Cry1F, Cry1Ab, and other Cry proteins, were inserted into cotton plants and into other crops [3,5]. More recently, Bt-crops have added the Vip family of proteins [3,5]. The latest generation (2015) of US transgenic cotton produces Cry1Ac, Cry1F, and Vip3A [3]. Major benefits of using these insecticidal proteins is their lack of toxicity to humans and other animals including beneficial insects [6,7]. Effective pest control in cotton is crucial and of special concern now because resistance has been detected for the Cry toxins [8].

The mechanism of Bt protein toxicity involves ingestion by the insect followed by protein solubilization, activation via protease cleavage, binding to low-affinity sites, binding to high-

affinity sites, toxin insertion into the midgut leading to pore formation, and death by sepsis from gut bacteria invading the insect hemocoel (the insect's circulatory system) [5,9]. These processes involve interactions with specific proteins found in the insect digestive system including serine proteases and Bt-binding receptors (cadherins, alkaline phosphatases, and aminopeptidases) required for pore formation [5,9]. Bt-transgenic crops and Bt toxic proteins in general have been effective in the control of a variety of pest species such as Coleoptera (beetles), Diptera (flies and mosquitoes), and Lepidoptera (moths and butterflies) [5,9], either genetically engineered into plants or by topical application.

A major concern in pest management is the development of insecticide resistance. Resistance now to the Cry proteins in multiple pest species has become widespread. In Puerto Rico, populations of the fall armyworm, *Spodoptera frugiperda*, have become resistant to Cry1F in Bt-modified corn [10], and now resistant insects have been found in North Carolina, USA [11,12]. In South Africa, the maize stalk borer, *Busseoloa fusca*, developed resistance to Cry1Ab [13]. Resistance was found for the western corn rootworm, *Diabrotica v. virgifera*, and resistance was found in the pink bollworm, *Pectinophora gossypiella*, in both India and the US [14-16]. The first suggestion that caterpillars had the potential to develop resistance to Bt came from the development of a Bt-resistant, tobacco budworm, *Heliothis virescens*, laboratory strain by Dr. Fred Gould at North Carolina State University (NCSU) which was 32,000-fold resistance to Cry1Ac [17,18]. Multi-protein resistance to Cry1A and Cry2A2 was also developed in the laboratory for the fall armyworm at NCSU [17]. The model organism in the studies reported here, *H. zea*, now has developed resistance in the field to transgenic cotton (Cry1Ac [19,20]) and transgenic sweet corn (Cry1A.105, Cry1Ab, and Cry2Ab2 [21,22]) in the US. Resistance to Cry1Ac and Cry1F proteins expressed in recent strains of Bt-cotton have also been reported [3].

Several different mechanisms for insect Bt resistance have been reported for *Heliothis virescens* and in other caterpillars involving changes in toxin activation in the midgut and toxin binding [23,24]. Changes in midgut cadherin receptors, proteases, GPI-anchored alkaline phosphatases, and glycolipid synthesis were shown [23-25]. Functional changes in the ATP-binding cassette (ABC) transport proteins was reported [26,27], and a dominant point mutation in tetraspanin proteins (TSPAN) in field resistant *Helicoverpa armigera* was found in China [28]. Decreases is Cry1Ac midgut protease cleavage was described in resistant *H. zea* [29]. A shotgun, global transcriptomics approach was used in this study to assess differences in gene expression between field-obtained, Bt-resistant versus laboratory Bt-susceptible (unfed) neonates of *H. zea* both reared under the same laboratory conditions.

Materials and Methods

Sample collection and preparation

Helicoverpa zea Bt-resistant (Cry1Ac, 100-fold resistant; see results section) eggs were obtained from a colony established at NCSU. The susceptible insects were from a laboratory strain reared with no exposure to Bt for 18 years while the resistant strain was originally collected from the field in NC and reared in the laboratory for 2 generations. The resistant colony was collected from Wake Forest, North Carolina, USA in non-Bt corn (2017). The susceptible colony was obtained from Benzon Research, Inc. (Carlisle, PA). Both colonies were reared in the lab for 2 generation on artificial diet [30]. To eliminate strain differences as much as possible, both the resistant and susceptible bollworms were reared in the same laboratory at NCSU and on

the same artificial diet, under the same environmental conditions, and using the same rearing methods. Rearing conditions were as follows in a growth chamber: 14:10 L:D, 27 ^{o}C:24 ^{o}C L:D, and 60% RH and were mated to conspecifics for each colony. The artificial diet was *H. zea* diet (Southland Products, Lake Village, AR, USA). The same rearing methods were used as described in Reisig et al. for both resistant and susceptible colonies [30]. Neonates less than 6 h after hatching from both colonies were then used for diet-based susceptibility bioassays and also RNA extraction.

Cry1Ac susceptibility bioassays

For both bioassays, 128-well plastic trays (Bio-assay tray bio-ba-128, Frontier Agricultural Sciences, Newark, DE, USA) were used. The overlay method with Cry1Ac protein (94%-96% pure, trypsin activated, ion exchange HPLC purified, desalted, freeze dried, provided by Marriane Pusztai-Carey, Case Western Reserve University) dissolved in Triton X-100 buffer (0.1%) was used for toxin application. For each well, 200 µL of Cry1Ac were added at the following concentrations: 0 µg/cm^2, 0.1 µg/cm^2, 1 µg/cm^2, 5 µg/cm^2, 10 µg/cm^2, 25 µg/cm^2, and 100 µg/cm^2. Each concentration was placed in 64 individual wells per bioassay tray where 1 neonate insect was placed using a fine tip paintbrush immediately after drying. In total, 128 neonates were used for the Bt-susceptible strain assays and 448 neonates for the Bt-resistant assays. Trays were then returned to the growth chamber for 7d. Mortality was then recorded, which was determined by whether or not neonates moved upon prodding with a brush. Data from each well and concentration of Cry1Ac was then pooled and used to calculate the LC_{50} and 95% CIs using a SAS probit analysis [PROC PROBIT, SAS Institute 2008]. The OPTC and LOG10 options were used to model the responses. The same bioassay and LC_{50} calculation methods were used as described in Reisig et al. [30].

RNA extraction

From these colonies, 5 Bt-resistant samples and 5 Bt-susceptible samples were prepared, each of these samples made from 10 neonate *H. zea*. All neonates were lab-reared and unfed prior to RNA extraction as described earlier. Neonates were mechanically homogenized into one DNAse and RNAse free tube for each sample. From each sample, total RNA was extracted using the RNeasy Mini Kit following the manufacturer's protocol (Qiagen, Valencia, CA, USA). Purity of total RNA in each sample was then assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) by the NC State University Genomics Core Facility (Raleigh, NC, USA). Sequencing was then only conducted on samples that had sufficient purities (RNA Integrity Number >9.0).

RNA sequencing

The NCSU Genomics Core Facility also conducted RNA-seq for this experiment. cDNA libraries for each sample (using 500 µg of total RNA each) were constructed to prepare for RNA-seq using the TruSeq RNA Library Prep Kit v2 (Illumina, San Diego, CA, USA) following the manufacturer's protocol. Transcriptome sequencing was performed on the NextSeq 500 System (Illumina, San Diego, CA, USA) using a paired end setting and read length of 2 x 150 base pairs. A sequencing depth of >25 million reads per library was obtained using a High Output Flow Cell. A total of 10 mRNA libraries were then prepared, 5 each for resistant and susceptible. The SRA Toolkit v2.9.2 was used to convert raw reads to fastq files [50]. Fastq file read quality was then assessed using the FastQC tool v0.11.7 [51]. A Phred score of >30 was required for a

majority of the sequencing reads to establish a baseline for quality. Fastq files with appropriate quality then proceeded to assembly and quality control steps.

Transcript assembly and quality control

The NC State Bioinformatics Core (Raleigh, NC, USA) conducted transcript assembly and quality control. Reads were assembled using the StringTie program (v1.3.5, John Hopkins University, Baltimore, Maryland, USA) with 45224 primary transcripts assembled into transcript set 1 with the *Helicoverpa zea* reference genome [52]. The program Trinity (v2.8.4, Broad Institute and Hebrew University of Jerusalem, Jerusalem, Israel) was used to assemble an alternate set of transcripts (set 2) that did not align with StringTie in order maximize transcript assemblies [53]. For transcript assembly, there were 149108 transcripts assembled and then processed through the Blobology program (v2.15.2, University of Edinburgh, UK) in order to determine if contaminants were present [54]. Transcripts that matched to Lepidoptera were then saved (108867 transcripts). From these, all ribosomal RNA transcripts were deleted from this transcript set. The Evigene program (v1.0, University of Indiana, Indiana, USA) was then used to cluster the remaining 108841 transcripts which resulted in 34059 transcripts in set 1 [55]. Transcript sets 1 and 2 were then combined and clustered using Evigene resulting in 26800 primary and 12095 alternate transcripts. Primary and alternate sets were then run through Blobology to check for contaminates once again. Ribosomal RNA transcripts were removed from these sets as well. Primary and alternate transcripts were then clustered and combined with Evigene.

Fastq files for each replicate were trimmed for adapter sequence and quality using the TrimmoMatic sequence trimmer (v0.39, Max Planck Institute, Germany) [56]. The *Helicoverpa zea* reference genome (NCBI), was used to map each trimmed file to the reference genome using HiSat2 [57]. StringTie was then used to assemble resulting mapped files in order to assemble RNA-seq alignments into potential transcripts. All transcript annotations from each replicate were then merged into one “expressed transcriptome” file. This was then used to guide gene boundaries when calculating differential expression values (log2 fold change) between the susceptible and resistant strains via CuffDiff (v7.0, Cambridge, Massachusetts, USA) [58]. Statistical significance was determined using the Tuxedo Pipeline (in CuffDiff, which assigned transcript q-values, α=0.05). Only statistically significant transcripts were included in later data analysis [59]. These results were then imported into the R statistical software platform for quality control checks and visualization of results [60]. The sequence of transcripts that were determined to be differentially expressed were extracted from the reference genome and used in a BLAST search against insects to provide initial annotations. Quality control steps were conducted with FPKM, boxplots, MDS plot, PCA plot, normalization, heatmap, and volcano plots. All quality control steps were passed by all replicates. After all transcripts were assembled and quality control steps passed, 6098 transcripts were identified as differentially expressed in this experiment. Of these 6098, 3042 transcripts had higher expression in the susceptible strain with 267 being found only in this strain. The remaining 3056 had higher expression in the resistant strain with 323 being only expressed in the resistant strain. Blast2GO (v5.2.4) was implemented to functionally annotate open reading frame assignments [34]. Gene ID and function was determined using BLASTx (E- value cut off 10^{-5}), using lepidopteran taxonomy to filter results, running against the nr and swissprot databases [34]**.**

Data Analysis and Figure Construction

Figures and tables for this paper were constructed using Microsoft Excel, PowerPoint, Word (2018), and SigmaPlot (v14.0, SigmaPlot, Systat Software, San Jose, CA). Venn diagrams were constructed using Venny 2.1 (http://bioifogp.cnb.csic.es/tools/venny) to depict transcript differences between strains. OmicsBox (v1.3.3)(BLAST2GO) was also used to construct gene ontologies (https://www.biobam.com/omicsbox-apps/) [34].

Results

Cry1Ac Susceptibility Bioassays

Cry1Ac feeding bioassays were conducted to establish the susceptibility of our field collected Wake Forest (resistant) versus Benzon (susceptible) strains prior to RNA seq. The resistant strain had an LC_{50} of 43.79 µg/cm^2 while the susceptible strain had an LC_{50} of 0.43 µg/cm^2 for Cry1Ac [30]. This was a resistance ratio of 100-fold. Statistical analysis (SAS probit analysis) and pertinent data were as follows for both bioassays: susceptible (slope=2.1, 95% CL= 0.26-0.46, Chi-square= 59.44) and resistant (slope=2.04, 95% CL= 19.36-85.47, Chi-square=11.01). Confidence limits do not overlap, and therefore the difference in LC_{50} between strains was considered statistically significant.

Genome-wide differential expression in Helicoverpa zea

After RNA-Seq, statistically differential expression levels were determined between the resistant versus susceptible strains at α=0.05. All statistically significant differences were included in our analysis so as not to exclude any possible mechanisms of resistance that might occur when excluding data based on an arbitrary established, minimum fold change. Overall, there were 3056 transcripts with increased expression (up-regulated) and 3042 with decreased expression (down-regulated) in the resistant strain (Figure 1). Additionally, there were 323 expressed transcripts that were found in only the resistant strain and 267 only in susceptible bollworms. The resistant strain had both a higher number of differentially expressed transcripts as well as novel expressed transcripts.

We also examined differential expression between the two strains at different log2 fold change (Figure 2A-C). Those transcripts with a (+) log2 fold change were up-regulated in the resistant strain and those with a (-) log2 fold change were down-regulated. Those transcripts unique to each strain were not included in this analysis. Of the differentially expressed transcripts at the α=0.05 threshold, 2325 (48.1%) were up-regulated, 2420 (50%) down-regulated, and 93 (1.9%) shared (defined in Figure 2 caption) in the resistant strain (Figure 2A). Examining differential expression with a log2 fold change of $\geq$ 2.0, 225 (42%) were up-regulated in the resistant strain, 290 (54.1%) down-regulated, and 21 (3.9%) shared (Figure 2B). Examining differentially expression with a log2 fold change of $\geq$ 5.0, 18 (62.1%) were up-regulated, 10 (34.5%) down-regulated, and 1 (3.4%) shared in the resistant strain (Figure 2C). The up versus down regulated transcripts were about the same except when the log2 fold change was $\geq$ 5.0, when there was more up-regulation in the resistant strain (Figure 2A-C).

Table 1 shows the top 50 transcripts with the highest degree of up-regulation (log2 fold change) in the resistant strain. Uncharacterized transcripts or those transcripts with no significant matches after a NCBI BLAST search were not included in the top 50 (see supplemental data for this sequence information). The highest top 50 log2 fold changes ranged from +9.08 (highest

level of up-regulation) to +3.51 log2 fold change (Table 1). General functions of these highest upregulated transcripts were variable. These included but were not limited to bromodomain, pupation, serine endopeptidase, metabolic (a broad variety of processes), replication, and binding proteins (Table 1). These top 50 transcripts also contained messages previously suggested to be involved in Bt resistance including tetraspanin 1 (+4.9 log2 fold change), serine protease (+5.13), trypsin 3A1 (+5.3), gamma-secretase (+4.7), chymotrypsin 1 (+4.4), and other trypsins (+4.12, +3.94) (Table 1). In general, a number of transcripts implicated before in Bt-resistance were in the top 50 upregulated transcripts; however some resistance-associated transcripts were not in the top 50 (Table 1).

Table 2 shows the top 50 transcripts with the highest degree of down-regulation (log2 fold change) in the resistant strain (Table 2). Again, uncharacterized transcripts with no matches to *H. zea* after a BLAST search were not included in the top 50 (see supplemental data for this sequence information). The highest top 50 log2 fold changes ranged from -7.18 (highest degree of down-regulation) to -3.48 log2 fold change (Table 2). General functions of the top 50 down regulated transcripts include but were not limited to xenobiotic metabolism, glucose metabolism, transcriptional modification, binding proteins, transporter proteins, and pupal proteins (Table 2). Present in these top 50 are transcript copies of genes known to be involved in resistance including beta secretase 1 (-3.6 log2 fold change), cytochrome P450s (-7.18, -5.56, -4.54, -4.32, -4.19, -3.86), and alkaline phosphatase (-3.77) (Table 2). Overall, the highest degree of log2 fold change was in the up-regulated category (+9.08) compared to the down-regulated (-7.18) (Tables 1-2).

Global gene functional annotations for differentially expressed transcripts

In order to examine functional annotations for all transcripts differentially expressed, OmicsBox was used to construct gene ontologies (Figure 3A-C). For all differentially expressed transcripts annotated under Biological Processes, transcripts with the greatest number of annotations were defined as follows: organic substance metabolic process (17%), primary metabolic process (16%), nitrogen compound metabolic process (14%), cellular metabolic process (14%), and biosynthetic process (6%) (Figure 3A). Metabolic processes (organic substance, primary metabolism, nitrogen compound, and cellular) were the most prevalent types of processes that were annotated and differentially expressed (Figure 3A).

For all differentially expressed transcripts annotated under Molecular Functions, transcripts with the greatest number of annotations were defined as follows: organic cyclic compound binding (15%), heterocyclic compound binding (15%), ion binding (13%), hydrolase activity (10%), and transferase activity (7%)(Figure 3B). The most prevalent types were binding activity (organic cyclic compound, heterocyclic compound, and ions) and enzymatic activity (hydrolase and transferase) (Figure 3B).

For all differentially expressed transcripts annotated under Cellular Components, transcripts with the greatest numbers of annotations were defined as follows: membrane (30%), intrinsic component of membrane (26%), organelles (10%), intracellular organelle (9%), and cytoplasm (6%)(Figure 3C). Under this category, the most prevalent categorizations were transcripts involved in membrane, an intrinsic component of the membrane, or portions of an organelle (Figure 3C).

Bt-resistance associated differential expression

Transcripts previously associated with Bt-resistance in lepidopterans were examined for their presence in our bollworm RNAseq results. Figure 4 shows the numbers of transcripts for each category of resistance associated transcripts and their direction of log2 fold change. Table 3 shows specific log2 fold changes for these resistance-associated transcripts with a log2 fold change ≥ 2.0. Numbers of transcripts and the highest log2 fold change transcripts in these categories are as follows: serine proteases [4 up-regulated (+5.13, +6.86, +3.1, and +2.4 log2 fold change), 2 down-regulated, 1 only in resistant], tetraspanins [3 up-regulated (+4.92, +3.37, and +2.88 log2 fold change), 1 down-regulated, 1 only in resistant, 1 only in susceptible], secretase proteins [3 up-regulated (+4.7, +3.04, and +2.07 log2 fold change), 1 down-regulated, 6 only in resistant, 3 only in susceptible], trypsin proteins [6 up-regulated (+4.44, +5.31, +4.12, +3.94, +3.62, and +2.17 log2 fold change), 1 only in susceptible], Bt-receptors [1 up-regulated (+2.93 log2 fold change), 6 down-regulated (-3.77, -6.26, -2.29, -2.89, -2.02, and -2.21 log2 fold change)], ABC transporters [4 up-regulated (+3.36, +3.06, +2.6, and +2.5 log2 fold change), 2 down-regulated, 11 only in resistant, 8 only in susceptible], and cytochrome P450s (CYPs) [10 up-regulated (+4.23, +4.1, +4, +3.37, +2.43, +2.31, +2.22, +2.16, +2.04, and +2.01 log2 fold change), 10 down-regulated, 15 only in resistant, 14 only in susceptible] (Figure 4, Table 3). CYPs were included in this analysis due to evidence showing involvement of these proteins in cross-resistance in insects [31-33]. For all the gene families, there were increased expression levels in our resistant bollworm strain with the exception of intestinal receptors (aminopeptidases, alkaline phosphatases, and cadherins) which had decreased expression in resistant bollworms. Beyond these broader protein families, there were also other important potential resistance associated transcripts with high levels of log2 fold change, i.e., carboxypeptidases (+3.04 and -2.03 log2 fold change), chitin synthase (+3.11), heat shock protein (+5.02), carboxyl/choline esterase (+6.17), and E3 ubiquitin-protein ligase associated protein (+5.47) (Table 3).

Figure 5 depicts three generalized immune pathways found in the bollworm: (i) Jak/STAT, (ii) Toll, and (iii) IMD pathways. Table 4 details additional transcripts involved in immunity in *H. zea* and their respective log2 fold changes. In the Jak/STAT pathway, 3 transcripts were differentially expressed: Hopscotch kinase (+1.1 log2 fold increase), STAT92E (+0.59 log2 fold increase), and PIAS (a negative regulator) (-0.79 log2 fold decrease). In the Toll pathway, 3 transcripts were differentially expressed: beta-1,3, glucanase (+1.47 log2 fold increase) and 2 Toll receptor proteins (+1.29 and +0.32 log2 fold increase). In the IMD immunity pathway, 5 transcripts were differentially expressed: PGN (+1.43 log2 fold increase), PGRP-LC (+1.86 log2 fold increase), PGRP-LB (+1.11 log2 fold increase), and 2 NF-kappa-beta proteins (+0.88 and +0.74 log2 fold increase). In each of these three pathways, the proteins involved in cell membrane receptors or protein activation (kinases) were the main types of proteins observed to be differentially expressed. The effector proteins for each pathway (ex. Defensins, attacins) were not found to be differentially expressed (Figure 5). Transcripts involved in the IMD pathway were also observed to have the most differential expression (10 individual transcripts) and also to have the highest degree of increased expression (+3.75 log2 fold change) (Table 4). Beyond the IMD pathway there were 4 different transcripts differentially expressed involved in the JAK/STAT pathway, 4 involved in the Toll pathway, 4 antimicrobial/bacterial proteins, 2 pathogen defense or recognition proteins, 2 immune signaling proteins, 6 autophagy, and 10 involved in the general immune response (Table 4). Some transcripts involved in immunity also

had decreased expression in the resistant strain (ex. Cecropin (-3.39 log2 fold change) and lysozyme (-3.34)); however overall the majority of immune-associated transcripts analyzed were found to have increased expression in the resistant strain (Table 4).

Discussion

Resistant versus susceptible bollworms

There was a 100-fold difference in susceptibility to Cry1Ac between susceptible and resistant bollworms. The optimum comparison would have been the study of both a susceptible and resistant field strain collected from the same location. However, at this juncture, the deployment of GMO crops that express Cry toxins is widespread, Bt resistance is widespread [13-16], and obtaining a Bt-susceptible field population of bollworms never exposed to Bt selection is not possible. Using a susceptible laboratory strain that is commercially available is advantageous as a standardized reference and was used before to study new mechanisms of Bt resistance [35,36]. However, some of the differences we found in this study could be unrelated to Bt resistance but a function of genome selection in a laboratory versus the field. To eliminate strain differences as much as possible, both the resistant and susceptible bollworms were reared in the same laboratory at NCSU and on the same artificial diet, under the same environmental conditions, and using the same rearing methods. In addition, we used newly hatched, unfed neonates so any differences in developmental polymorphisms between the two strains would be minimized. Using unfed neonates also allowed us to examine the constitutive differences in gene expression between the two strains before stadium develop is initiated by feeding.

Differential expression between Bt resistant versus susceptible bollworms

Initial analysis of the resulting RNA-seq data showed global differences in gene expression between strains. After removing all non-differentially expressed transcripts and transcript copies, in the resistant strain there were 2325 up-regulated transcripts, 2420 down-regulated, and 93 shared transcripts. Shared transcripts were transcript variants (different mRNA sequences that code for the same protein) or gene isoforms (mRNA sequences coding for the same protein with differing transcriptional start sites, untranslated regions, or protein coding regions). Additional copies or variants of resistance associated transcripts (ex. proteases and transporters) may help explain how a particular Bt-resistance associated transcript is impacting resistance. In some cases, transcript variants with different mRNA sequences for a gene that code for the same protein were differentially expressed but in opposite directions, for example with cytochrome P450s. Potentially, some transcript variants are impacting Bt-resistance and some are not (discussed in more detail later). When examining the numbers of transcripts with high levels of log2 fold change in resistant neonates, it was found that 225 transcripts were up-regulated, 290 were down-regulated, and 21 were shared using a threshold of ≥ 2.0 log2 fold change in either direction. Using a threshold of ≥ 5.0 log2 fold change, 18 transcripts were up-regulated, 10 down-regulated, and 1 was shared. This indicates that when comparing each of these different thresholds, only when examining transcripts with the highest degrees of log2 fold change does up-regulation become dominant. An explanation for this could be that the genetic changes that are linked to Bt-resistance are predominantly occurring on specific transcripts instead of global shifts in the genome, which does adhere to the current understanding of Bt-resistance [24,25,28].

Results also indicated a higher overall degree of log2 fold change in the resistant strain (ex. highest log2 fold change +9.08 compared to -7.18 in susceptible neonates). Uncharacterized transcripts with even higher log2 fold change are shown in the supplement section. While the functions of transcripts with the highest magnitudes of log2 fold change do not appear to have a known or obvious connection with Bt-resistance (ex. bromodomain protein and cytochrome P450), it may be that the high degree of differential expression is a side-effect of other genetic shifts occurring when Bt-resistance develops in *H. zea* (or a result of strain differences unrelated to Bt resistance). The gene with the highest degree of down-regulation, a cytochrome P450 (CYPs), normally is involved in xenobiotic metabolism and other processes. It would be interesting to examine cross-resistance to insecticide chemistry for Bt resistant bollworms (discussed in more detail later).

Differential expression of Cytochrome P450s in Bt-resistant Helicoverpa zea

In examining the CYPs that were differentially expressed, there were 10 up-regulated, 10 down-regulated, 15 only in resistant, and 14 only in susceptible neonates. CYPs are a broad family of proteins involved in the metabolism of many different substrates, including endogenous chemicals and xenobiotics [37]. The variation in differential expression for this group of transcripts in this study suggest in part that some of these CYPs may have a role in Bt resistance. For example, CYP337Bv1, the gene exhibiting the highest down-regulation may have no role while CYP6B5 with a high degree of up-regulation in the resistant strain (log2 fold change +4.44) may be important in Bt resistance. P450s are not known to metabolize proteins. However, these enzymes are involved in the metabolism of plant secondary compounds and metabolic products from bacteria, among many other substrates [38-40]. Maybe an increase in some P450s is important in Bt resistance because of their role in the metabolism of secondary metabolites produced from the proliferation of bacteria and fungi in the insect hemocoel and/or by the metabolism of secondary plant compounds that reduce insect fitness along with Bt poisoning. However, our study was a comparison between a field vs lab strain; P450 transcript differences could simply reflect differences in xenobiotic exposure and natural selection between living in the field versus being reared on artificial diet. Transcriptional differences in metabolizing enzymes between a laboratory vs a field strain was shown before [41] but cross-resistance between Cry1Ac and the chemical insecticide deltamethrin also was found in the diamondback moth, *Plutella xylostella* L. [42]. Bollworms collected from Alachua County, FL (USA) (before the deployment of Bt crops) and reared in the laboratory for 3 generations had a much higher tolerance to Cry1Ac than other caterpillars collected throughout the SE US (and treated the same) and also demonstrated cross tolerance to Orthene (Roe, personal communication). There were clear differences in CYP transcript levels between Bt resistant and susceptible bollworms in this study, the consequences of which we do not yet understand but which raise interesting possibilities.

Genome characterization of global differentially expressed transcripts

For the differential expression between the resistant and susceptible strains, gene ontologies were constructed to categorize annotated transcripts by function (Figure 3A-C). Of the differentially expressed transcripts annotated under Biological Processes, the highest proportions were categorized as being involved in organic substance metabolic processes, primary metabolic processes, nitrogen compound metabolism, and cellular metabolic processes. Of the differentially expressed transcripts annotated under Molecular Function, the highest

proportions were categorized as being involved in organic cyclic compound binding, heterocyclic compound binding, ion binding, and hydrolase activity. Of the differentially expressed transcripts annotated under Cellular Components, the highest proportions were categorized as being involved in the cell membrane, a component of the cell membrane, organelles, and intracellular organelle. These categories include different types of metabolism, binding, and cellular membrane components. Analysis of differential expression levels between Bt-resistant and susceptible *Plutella xylostella* showed similar Gene Ontology results including metabolic processes, binding, cellular components, and binding processes [42]. A large number of metabolizing transcripts (cytochrome P450s, 49 transcripts) were differentially expressed in our data. Cytochrome P450s are involved in a broad category of metabolic processes and contribute to the high degree of functional categories annotating to metabolic processes. Possible explanations for differential expression in CYPs were discussed in the above section. Further characterization of functional categories for differences in gene expression between resistant and susceptible insects will help illuminate potential new areas of investigation for resistance mechanisms in the future, especially as more global gene expression studies are conducted related to Bt resistance.

Role of proteases, receptors, and transporters in resistance

Former studies investigating Bt-resistance in *H. zea* discovered several different mechanisms of resistance. Most importantly, alterations in proteases (secretases, chymotrypsins, and trypsins), midgut Bt-interacting receptors (cadherins, aminopeptidases, and alkaline phosphatases), transporters (ABC), and tetraspanins (TSPAN) have all previously been associated with Bt-resistance [23-28]. In this particular study, we found supporting evidence that all of these changes were present in the same resistant strain (increases in proteases, transporters, and tetraspanins; decreases in midgut receptors). Additionally, a number (9) of these important resistance-associated transcripts were found among the top 50 up-regulated transcripts with the highest degree of increased expression in the resistant strain. A mid-gut Bt-interacting receptor (1) was also found among the top 50 down-regulated transcripts with decreased expression in the resistant strain. While each of these mechanisms of Bt-resistance have been recognized individually and across several different model organisms, this study represents one of the first to find all of the discussed mechanisms of Bt-resistance present in a single population of *H. zea*. We do not know yet if they all occurred in the same insect. Our results suggest that the resistant population studied evolved Bt-resistance in the field via a wide array of different mechanisms, potentially indicating that the genomic control mechanisms for insecticide resistance may fall under the same control pathway. Ideally, as investigators gain a greater understanding of these different mechanisms of Bt resistance and maybe discover more, this information could lead to improved resistance management and better decision making in the development of the next generation of biopesticides.

Role of insect immunity in resistance

When a susceptible insect consumes plant tissue and therefore Cry and Vip families of insecticidal proteins, the hypothesized ultimate cause of death is sepsis caused by gut bacteria invading the body cavity. We hypothesized from this study that an additional potential mechanism of Bt-resistance may involve an enhanced immune system. We found three differentially expressed immune related pathways in the bollworm, JAK/STAT, Toll, and IMD. The primary function of the IMD pathway is to respond to infection by gram-negative bacteria.

Major components of this pathway include but are not limited to Peptidoglycan recognition proteins (PGRPs), Peptidoglycan binding proteins, Fas-associated death domains (FADD), DREDD, Relish, Transforming Growth factors (TAK or TAB), Nuclear factors kappa beta (nf-kb), immunoglobulin binding proteins, Fas binding factors, caspases, and defensive proteins [43,44]. In this study, a number of transcripts (9) involved in the IMD pathway were found to have increased expression in the resistant strain of the bollworm. These were a Fas-binding factor (+3.75 log2 fold change), immunoglobulin binding protein (+3.41 log2 fold change), Peptidoglycan recognition protein C (+1.86 log2 fold change), Peptidoglycan binding protein (+1.43 log2 fold change), Nuclear factor kappa betas (+0.88, +0.74 log2 fold change), and transforming growth factor betas (+0.54, +0.48 log2 fold change). A recent study by Liu et. al. (2019), also correlated PGRP expression to Cry1Ac proteins [45]. This may indicate that the IMD pathway, in particular PGRP proteins, are an important mechanism for Bt-resistance.

Another immune-response pathway in *H. zea* is the Toll pathway. This pathway responds to gram-positive bacterial and fungal infections in addition to being involved in developmental processes. Some components of the Toll pathway are shared with the IMD pathway. Major protein components of the Toll pathway are Spatzle, Toll receptors, Beta-1,3, glucanases, Cactus proteins, Dorsal proteins, death associated protein kinases (DAP), and cecropin proteins [43,44]. We found transcripts involved in the Toll pathway were found to have increased expression in the resistant strain. These included beta-1,3, glucanase (+1.47 log2 fold change), DAP kinases (+0.68 log2 fold change), Toll protein (+0.33 log2 fold change), and Toll receptor (+1.30 log2 fold change).

A third immune pathway in insects is JAK/STAT, and components of this pathway demonstrate differential expression between the resistant and susceptible bollworms studied. This pathway deals with more generalized immune responses and developmental processes rather than specific types of bacterial infection. Major components of this pathway include Domeless receptors, Hopscotch kinases, STAT proteins, PIAS regulators, SOCS proteins, and defensive proteins [44]. We found differential expression levels in the following transcripts: Hopscotch kinase (+1.1 log2 fold change), STAT5B (+0.59 log2 fold change), SH3 binding protein (+0.68 log2 fold change), and PIAS3 (a negative regulator) (-0.79 log2 fold change).

Of the three immune pathways examined, the IMD pathway had the greatest number of transcripts with increased expression in the resistant strain (9 transcripts) as well as the highest degree of log2 fold change (Fas-binding factor (+3.75 log2 fold change), immunoglobulin binding protein (+3.41 log2 fold change), and Peptidoglycan recognition protein C (+1.86 log2 fold change)). This could potentially be explained by the fact that gram-negative bacteria commonly colonize gut-cavities of lepidopterans [46]. Since the recognized cause of death by Cry proteins is from sepsis when bacteria move from the gut to the hemocoel, this could explain why the IMD pathway saw the most differences in expression. It is important to note that both gram-negative and gram-positive bacteria are present in lepidopterans with gram-positive the most dominant [46].

Increased expression in these three immune pathways could be a mechanism of insecticide resistance through an enhanced ability to fight bacterial infection. This study provides some of the first insight into the possible role of the insect immune system in Bt-resistance. Previous

research has found that the gut-microbiome and immune activity in insects is linked to chlorpyrifos resistance; potentially similar interactions are happening in Bt-resistance [46]. However, since this study only considered the correlation between gene expression and Bt-resistance, further investigation is needed to support the hypothesis of enhanced immunity have a role in insect resistance to transgenic crops.

In summary, this study has confirmed most of the previous work on identifying possible mechanisms for caterpillar Bt-resistance, but in this case all of these mechanisms appear to be in play in a single resistant strain, to an extent not shown before. In the resistant strain, we found increased expression of multiple proteases, transporters, tetraspanins, and secretases and decreased expression of Cry midgut receptors. Additional differential expression was found for enzymes typically involved in resistance to chemical insecticides. This could be a result of differences in a laboratory versus field strain or maybe a mechanism by which resistant caterpillars detoxify increasing levels of microbial metabolites in the hemocoel during the advancement of sepsis. Furthermore, there were clear changes in three of the insect's major immune pathways which could be providing an enhanced mechanism to fight infection from Bt induced sepsis. These pathways were characterized for the first time for bollworms. Further research is needed to confirm a role of the immune system in Bt resistance. While other studies have examined Bt-resistance in lepidopterans (Asian corn borer, Ostrinia furnacalis, diamondback moth, Plutella xylostella, and the old world bollworm, *Helicoverpa armigera* [47-49]) using RNAseq and lab-selected resistance, this study with *Helicoverpa zea* represents the first work that used RNAseq to examine expression levels in a field-caught resistant strain with no selection for resistance prior to sequencing. This study also supports the findings in these other Lepidoptera including decreased expression levels for Bt-receptors as well as increased trypsin and other mid-gut proteases and increases in detoxifying/metabolic enzymes (P450s).

REFERENCES

1. USDA. *Pests and Diseases of Cotton*; United States Department of Agriculture: Washington D.C. USA, 1925.

2. University of California Agriculture, *UC IPM Pest Management Guidelines: Cotton.* UC ANR Publication 3444. Available online: http://ipm.ucanr.edu/PMG/r114300511.html (accessed June 19 2019)

3. Fleming, D.; Musser, F.; Reisig, D.; Greene, J.; Taylor, S.; Parajulee, M.; Lorenz, G., Catchot, A.; Gore, J.; Kerns, D.; Stewart, S.; Boykin, D.; Caprio, M.; Little, N. Effects of transgenic *Bacillus thuringiensis* cotton on insecticide use, heliothine counts, plant damage, and cotton yield: A meta-analysis, 1996-2015. *PLoS ONE* **2018**, *13*, e0200131.

4. Zhu, B.; Xu, M.; Shi, H.; Gao, X.; Liang, P. Genome-wide identification of lncRNAs associated with chlorantraniliprole resistance in diamondback moth *Plutella xylostella* (L.). *BMC Genomics* **2017**, *18*, 380.

5. Palma, L.; Muñoz, D.; Berry, C.; Murillo, J.; Caballero, P. *Bacillus thuringiensis* toxins: an overview of their biocidal activity. *Toxins* **2014**, *6*, 3296–3325.

6. Qaim, M.; De Janvry, A. Bt cotton and pesticide use in Argentina: Economic and environmental effects. *Environment and Development Economics* **2005,** *10*, 179-200.

7. United States Environmental Protection Agency (EPA). EPA's Regulation of Bacillus thuringiensis (Bt) Crops. Available online: https://archive.epa.gov/pesticides/biopesticides/web/html/regofbtcrops.html (accessed June 19 2019

8. Statista. Cotton production by country worldwide in 2018/2019. Available online: https://www.statista.com/statistics/263055/cotton-production-worldwide-by-top-countries/ (accessed June 19 2019)

9. Bravo, A.; Gill, S.S.; Soberón, M. Mode of action of *Bacillus thuringiensis* Cry and Cyt toxins and their potential for insect control. *Toxicon* **2007**, *49*, 423–435.

10. Moar, W.; Roush, R.; Shelton, A.; Ferre, J.; MacIntosh, S.; Leonard, B.R.; Abel, C. Field-evolved resistance to Bt toxins. *Nat. Biotechnol*. **2008**, *26*, 1072.

11. Tabashnik, B.E.; Van Rensburg, J.; Carrière, Y. Field-evolved insect resistance to Bt crops: definition, theory, and data. *J. Econ. Entomol*. **2009**, *102*, 2011-2025.

12. Huang, F.; Qureshi, J.A.; Meagher, R.L.Jr.; Reisig, D.D.; Head, G.P.; Andow, D.A.; Ni, X.; Kerns, D.; Buntin, G.D.; Niu, Y. Cry1F resistance in fall armyworm *Spodoptera frugiperda*: single gene versus pyramided Bt maize. *PLoS ONE* **2014** *9*, e112958.

13. van Rensburg, J.B.J. First report of field resistance by the stem borer, *Busseola fusca* (Fuller) to Bt-transgenic maize. *South Afr. J. Plant Soil* **2007**, *24*, 147-151.

14. Gassmann, A.J.; Petzold-Maxwell, J.L.; Keweshan, R.S.; Dunbar, M.W. Field-evolved resistance to Bt maize by western corn rootworm. *PLoS ONE* **2011**, *6*, e22629.

15. Gassmann, A.J.; Petzold-Maxwell, J.L.; Clifton, E.H.; Dunbar, M.W.; Hoffmann, A.M.; Ingber, D.A.; Keweshan, R.S. Field-evolved resistance by western corn rootworm to multiple *Bacillus thuringiensis* toxins in transgenic maize. *Proc. Natl. Acad. Sci. USA* **2014**; *111*, 5141– 5146.

16. Dhurua, S.; Gujar, G.T. Field-evolved resistance to Bt toxin Cry1Ac in the pink bollworm, *Pectinophora gossypiella* (Saunders) (Lepidoptera: Gelechiidae), from India. *Pest Manage. Sci.* **2011**, *67*, 898–903.

17. Jackson, R.; Marcus, M.; Gould, F.; Bradley, J.Jr.; Van Duyn, J. Cross-resistance responses of Cry1Ac-selected *Heliothis virescens* (Lepidoptera: Noctuidae) to the *Bacillus thuringiensis* protein Vip3A. *J. Econ. Entomol.* **2007**, *100*, 180-186.

18. Gould, F.; Anderson, A., Jones, A.; Sumerford, D.; Heckel, D.G.; Lopez, J.; Micinski, S.; Leonard, R.; Laster, M. Initial frequency of alleles for resistance to *Bacillus thuringiensis* toxins in field populations of *Heliothis virescens*. *Proc. Natl. Acad. Sci. USA* **1997,** *94*, 3519-3523.

19. Tabashnik, B.E.; Fabrick, J.A.; Unnithan, G.C.; Yelich, A.J.; Masson, L.; Zhang, J.; Bravo, A.; Soberón, M. Efficacy of genetically modified Bt toxins alone and in combinations against pink bollworm resistant to Cry1Ac and Cry2Ab. *PLoS ONE* **2013**, *8*, e80496.

20. Tabashnik, B.E.; Zhang, M.; Fabrick, J.A.; Wu, Y.; Gao, M.; Huang, F.; Wei, J.; Zhang, J.; Yelich, A.; Unnithan, G.C. Dual mode of action of Bt proteins: protoxin efficacy against resistant insects. *Sci. Rep*. **2015**, *5*, 15107.

21. Dively, G.P.; Venugopal, P.D.; Finkenbinder, C. Field-evolved resistance in corn earworm to Cry proteins expressed by transgenic sweet corn. *PLoS ONE* **2016**, *11*, e0169115.

22. Yang, F.; Santiago González, J.C.; Williams, J.; Cook, D.C.; Gilreath, R.T.; Kerns, D.L. Occurrence and ear damage of *Helicoverpa zea* on transgenic *Bacillus thuringiensis* maize in the field in Texas, U.S. and its susceptibility to Vip3A protein. *Toxins*. **2019**, *11*, 1-13.

23. Soberón, M.; Pardo, L.; Muñóz-Garay, C.; Sánchez, J.; Gómez, I.; Porta, H.; Bravo, A. Pore formation by Cry toxins. In *Proteins membrane binding and pore formation*, Anderluh, G., Lakey, J., Eds.; Landes Bioscience: Texas, USA, 2010.; pp. 127-142.

24. Pardo-Lopez, L.; Soberon, M.; Bravo, A. *Bacillus thuringiensis* insecticidal three domain Cry toxins: mode of action, insect resistance and consequences for crop protection. *FEMS*

Microbiol. Rev. **2013**, *37*, 3-22.

25. Jurat-Fuentes, J.L.; Karumbaiah, L.; Jakka, S.R.K.; Ning, C.; Liu, C.; Wu, K.; Jackson, J.; Gould, F.; Blanco, C.; Portilla, M. Reduced levels of membrane-bound alkaline phosphatase are common to lepidopteran strains resistant to Cry toxins from Bacillus thuringiensis. *PLoS ONE* **2011,** *6*, e17606.

26. Gahan, L.J.; Pauchet, Y.; Vogel, H.; Heckel, D.G. An ABC transporter mutation is correlated with insect resistance to *Bacillus thuringiensis* Cry1Ac toxin. *PLoS Genet.* **2010**, *6*, e1001248.

27. Atsumi, S.; Miyamoto, K.; Yamamoto, K.; Narukawa, J.; Kawai, S.; Sezutsu, H.; Kobayashi, I.; Uchino, K.; Tamura, T.; Mita, K. Single amino acid mutation in an ATP-binding cassette transporter gene causes resistance to Bt toxin Cry1Ab in the silkworm, *Bombyx mori*. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E1591-E1598.

28. Jin, L.; Wang, J.; Guan, F.; Zhang, J.; Yu, S.; Liu, S.; Xue, Y.; Li, L.; Wu, S.; Wang, X. Dominant point mutation in a tetraspanin gene associated with field-evolved resistance of cotton bollworm to transgenic Bt cotton. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 11760-11765.

29. Zhang, M.; Wei, J.; Ni, X.; Zhang, J.; Jurat-Fuentes, J.L.; Fabrick, J.A.; Carrière, Y.; Tabashnik B.E.; Li, X. Decreased Cry1Ac activation by midgut proteases associated with Cry1Ac resistance in *Helicoverpa zea*. *Pest Manage. Sci.* **2019**, *75*, 10199-1106.

30. Reisig, D.D.; Huseth, A.S.; Bacheler, J.S.; Aghaee, M.A.; Braswell, L.; Burrack, H.J.; Flanders, K.; Greene, J.K.; Herbert, D.A.; Jacobson, A. Long-term empirical and observational evidence of practical *Helicoverpa zea* resistance to cotton with pyramided Bt toxins. *J. Econ. Entomol.* **2018**, *111*, 1824-1833.

31. Sayyed, A.H.; Moores, G.; Crickmore, N.; Wright, D.J. Cross-resistance between a *Bacillus thuringiensis* Cry toxin and non-Bt insecticides in the Diamondback Moth. *Pest Manage. Sci.* **2008**, *64*, 813-819.

32. Hafeez, M.; Liu, S.; Jan, S.; Shi, L.; Fernández-Grandon, G. M.; Gulzar, A.; Ali, B.; Rehman, M.; Wang, M. Knock-Down of gossypol-Inducing cytochrome P450 genes reduced deltamethrin sensitivity in *Spodoptera exigua* (Hübner). *Int. J. Mol. Sci.* **2019**, *20*, 2248.

33. Chen, A.; Zhang, H.; Shan, T.; Shi, X.; Gao, X. The overexpression of three cytochrome P450 genes CYP6CY14, CYP6CY22 and CYP6UN1 contributed to metabolic resistance to Dinotefuran in melon/cotton Aphid, *Aphis Gossypii* Glover. *Pestic. Biochem. Physiol*. **2020**, *167*, 104601.

34. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.

Bioinformatics **2005**, *21*, 3674–3676.

35. Carriere, Y.; Degain, B.; Unnithan, G.C.; Harpold, V.S.; Li, X.; Tabashnik, B.E. Seasonal Declines in Cry1Ac and Cry2Ab concentration in maturing cotton favor faster evolution of resistance to pyramided Bt cotton in *Helicoverpa zea* (Lepidoptera: Noctuidae). *J. Econ. Entomol.* **2019**, *112*, 2907-2914.

36. Welch, K.L.; Unnithan, G.C.; Degain, B.A.; Wei, J.; Zhang, J.; Li, X.; Tabashnik, B.E.; Carriere, Y.; Cross-resistance to toxins used in pyramided Bt crops and resistance to Bt sprays in helicoverpa zea. *J. Invetrebr. Pathol*. **2015**, *132*, 149-156.

37. Kullman, S.W.; Mattingly, C.J.; Meyer, J.N.; Whitehead, A. Perspectives on informatics in toxicology. In *A textbook of modern toxicology*, 4th ed.; Hodgson E., Ed.; Wiley: New Jersey,USA, **2010**, pp. 593–605.

38. Santos-Garcia, D.; Mestre-Rincon, N.; Zchori-Fein, E.; Morin, S. Inside Out: Microbiota dynamics during host-Plant adaption of whiteflies. *ISME J***. 2020**, *14*, 847-856.

39. Crava, C.M.; Brutting, C.; Baldwin, I.T. Transcriptome profiling reveals differential gene expression of detoxification enzymes in a hemimetabolous tobacco pest after feeding on jasmonate-silenced *Nicotiana attenuata* plants. *BMC Genomics* **2016**, *17*, 1005.

40. Mao, W.; Schuler, M.A.; Berenbaum, M.R. Task-related differential expression of four cytochrome P450 genes in honeybee appendages. *Insect Mol. Biol*. **2015**, *24*, 582-588.

41. Thorpe, P.; Escudero-Martinez, C.M.; Eves-van den Akker, S.; Bos, J.I.B. Transcriptional changes in the aphid species *Myzus cerasi* under different host and environmental conditions. *Insect Mol. Biol.* **2020**, *29*, 271-282.

42. Lei, Y.; Zhu, X.; Xie, W.; Wu, Q.; Wang, S.; Guo, Z.; Xu, B.; Li, X.; Zhou, X.; Zhang, Y. Midgut transcriptome response to a cry toxin in the Diamondback Moth, *Plutella Xylostella* (Lepidoptera: Plutellidae). *Gene* **2014**, *533*, 180-187.

43. Liu, Y; Shen, D; Zhou, F; Wang, G; An, C. Identification of immunity-related genes in *Ostrinia furnacalis* against entomopathogenic fungi by RNA-seq analysis. *PLoS ONE* **2014**, *9*, e86436.

44. Cao, X.; He, Y.; Hu, Y.; Wang, Y.; Chen, Y. R.; Bryant, B.; Clem, R.J.; Schwartz, L.M.; Blissard, G.; Jiang, H. The immune signaling pathways of *Manduca sexta*. *Insect Biochem. Mol. Biol.* **2015**, *62*, 64–74.

45. Liu, A.; Huang, X.; Gong, L.; Guo, Z.; Zhang, Y.; Yang, Z. Characterization of immune-related PGRP gene expression and phenoloxidase activity in Cry1Ac-susceptible and -resistant *Plutella xylostella* (L.). *Pestic. Biochem. Physiol.* **2019**, *160*, 79-86.

46. Xia, X.; Sun, B.; Gurr, G.M.; Vasseur, L.; Xue, M.; You, M.; Gut microbiota mediate insecticide resistance in the diamondback moth, *Plutella xylostella* (L.). *Front Microbiol*.

2018, *9*, 25.

47. Wei, J; Yang, S.; Chen, L.; Liu, X.; Du, M. ; An, S.; Liang, G. ; Transcriptomic Responses to Different Cry1Ac Selection Stresses in *Helicoverpa armigera. Front Physiol*. **2018,** 9, 1653. doi: 10.3389/fphys.2018.01653

48. Xu, L.N.; Wang, Y.Q.; Wang, Z.Y.; Hu, B.J.; Ling, Y.H.; He, K.L.; Transcriptome differences between Cry1Ab resistant and susceptible strains of Asian corn borer. *BMC Genomics*. **2015,** 16(1), 173. Doi: 10.1186/s12864-015-1362-2

49. Gong, L.; Kang, S.; Zhou, J.; Sun, D.; Guo, L. ; Qin, J.; Zhu, L.; Bai, Y.; Ye, F.; Akami, M.; Wu, Q.; Wang, S.; Xu, B.; Yang, Z.; Bravo, A.; Soberon, M.; Guo, Z.; Wen, L.; Zhang, Y. Reduced Expression of a Novel Midgut Trypsin Gene Involved in Protoxin Activation Correlates with Cry1Ac Resistance in a Laboratory-Selected Strain of *Plutella xylostella* (L.). *Toxins (Basel).* **2020**, 12(2), 76. Doi: 10.3390/toxins12020076

50. Leinonen, R.; Sugawara, H.; Shumway, M. International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res*. **2011**, *39*, D19–D21.

51. Babraham Bioinformatics: FastQC: a quality control tool for high throughput sequence data. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed January 8 2019)

52. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L.; StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol*. **2015,** *33*, 290–295.

53. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; MacManes, M.D.; Ott, M.; Orvis, J.; Pochet, N.; Strozzi, F.; Weeks, N.; Westerman, R.; William, T.; Dewey, C.N.; Henschel, R.; LeDuc, R.D.; Friedman, N.; Regev, A. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc*. **2013,** *8*, 1494–1512.

54. Kumar, S.; Jones, M.; Koutsovoulos, G.; Clarke, M.; Blaxter, M. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front. Genet*. **2013**, *4*, 237.

55. Rago, A.; Gilbert, D.G.; Choi, J.H.; Sackton, T.B.; Wang, X.; Kelkar, Y.D.; Werren, J.H.; Colbourne, J.K. OGS2: genome re-annotation of the jewel wasp *Nasonia vitripennis*. *BMC Genomics*. **2016**, *17*, 678.

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

57. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357-360

58. Trapnell, C.; Hendrickson, D.; Sauvageau, S.; Goff, L.; Rinn, J.L.; Pachter, L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnology **2013**, *31*, 46-53.

59. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D.R.; Pimentel, H.; Salzberg S.L.; Rinn, J.L.; Pachter L. Differential gene and transcripts expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols.* **2012**, 7, 562-578. https://doi.org/10.1038/nprot.2012.016

60. R Development Core Team. R: A language and environment for statistical computing; R Foundation for Statistical Computing: Vienna, Austria, 2012.

FIGURES AND TABLES

TABLES

Table 1. Top 50 (highest log2 fold change) up-regulated transcripts in the Bt-resistant strain of the bollworm, *Helicoverpa zea*, including gene identity, general function, and magnitude of log2 fold change.

	Gene[a]	Gene ID[b]	General function[c]	Log2 fold increase[d]
1	Hzea.19884	WD repeat-containing protein on Y chromosome-like	Bromodomain protein	+9.08472
2	Hzea.21942	BAC, pupae DNA	Pupal protein	+7.16661
3	Hzea.16694	BAC, pupae DNA	Pupal protein	+6.87974
4	Hzea.24399	allergen Api m 6-like	Serine Type Endopeptidase, protease inhibitor	+6.86176
5	Hzea.23153	BAC, pupae DNA	Pupal protein	+6.51622
6	Hzea.2403	KGHa033C10 carboxyl/choline esterase	Dietary detoxification, hormone/pheromone degradation, neurodevelopment	+6.17351
7	Hzea.15968	fatty acid binding protein	Binding and transfer of fatty acids	+5.93447
8	Hzea.27515	zonadhesin-like	Facilitates binding of sperm to egg	+5.87046
9	Hzea.2907	sphingolipid delta(4)-desaturase/C4-monooxygenase DES2	Degenerative spermatocyte protein	+5.83895
10	Hzea.2119	serine/threonine-protein kinase dyrk1	Phosphorylation of serines and threonines	+5.51915
11	Hzea.8474	sorbitol dehydrogenase-like	Sorbitol metabolism	+5.51262
12	Hzea.3274	E3 ubiquitin-protein ligase Siah1-like	Proteosome mediated protein degradation	+5.46839
13	Hzea.4257	trypsin 3A1-like	Intestinal protein degradation	+5.31623
14	Hzea.23538	transmembrane protease serine 9-like	Serine protein cleavage	+5.13044
15	Hzea.32418	zinc finger protein 266-like	DNA binding domain protein	+5.0701
16	Hzea.18297	heat shock protein Hsp-12.2-like	Cellular stress response protein	+5.02824
17	Hzea.18477	AY2 tetraspanin 1 (TSPAN1)	Protein stabilization, cell signaling pathways, associated with BT resistance	+4.91784
18	Hzea.13946	facilitated trehalose transporter Tret1	Trehalose transport from fat body	+4.73271
19	Hzea.10453	gamma-secretase subunit pen-2	Cleavage of transmembrane proteins	+4.70584
20	Hzea.31473	JP123 retrotransposon HaRT3	Replication	+4.6664
21	Hzea.16029	nose resistant to fluoxetine protein 6-like	Uptake of lipids and xenobiotics from intestines	+4.64037
22	Hzea.1396	BAC, pupae DNA	Pupal protein	+4.5819
23	Hzea.15497	chymotrypsin-1-like	Intestinal protein degradation	+4.44894
24	Hzea.14146	cytochrome P450 6B5-like	Xenobiotic metabolism	+4.4463
25	Hzea.8807	BAC, pupae DNA	Pupal protein	+4.43852
26	Hzea.20434	stabilizer of axonemal microtubules 2	Microtubule binding	+4.42253
27	Hzea.16622	deoxyribose-phosphate aldolase	Deoxyribose phosphate catalysis	+4.37083
28	Hzea.9448	cytochrome P450 337B2v2	Xenobiotic metabolism, associated with insecticide resistance	+4.23841
29	Hzea.31690	H/ACA ribonucleoprotein complex subunit 1-like	Ribosome biogenesis and telomere maintenance	+4.14773

Table 1 cont.

	Gene[a]	Gene ID[b]	General function[c]	Log2 fold increase[d]
30	Hzea.8806	protein lethal(2)essential for life-like	Embryonic development	+4.14618
31	Hzea.27511	JP151 retrotransposon HaRT2	Replication	+4.14218
32	Hzea.14487	BAC 79L08	Unknown	+4.13627
33	Hzea.15502	trypsin-like protease	Intestinal protein degradation	+4.12705
34	Hzea.29377	cytochrome P450 337B3v1	Xenobiotic metabolism, associated with insecticide resistance	+4.10688
35	Hzea.17522	JP151 retrotransposon HaRT2	Replication	+4.10241
36	Hzea.1891	odorant receptor Or1-like	Odorant receptor	+4.07168
37	Hzea.25785	BAC, pupae DNA	Pupal protein	+4.04936
38	Hzea.391	cytochrome P450 337B2v2	Xenobiotic metabolism, associated with insecticide resistance	+4.0049
39	Hzea.2400	trypsin-1-like	Intestinal protein degradation	+3.94767
40	Hzea.11080	neuroblastoma-amplified sequence-like	Vesicle binding/transport	+3.89792
41	Hzea.18643	H/ACA ribonucleoprotein complex subunit 1-like	Ribosome biogenesis and telomere maintenance	+3.89234
42	Hzea.30733	NAD-dependent protein deacylase sirtuin-5, mitochondrial-like	Protein Deacylation	+3.78646
43	Hzea.10806	BAC, pupae DNA	Pupal protein	+3.77303
44	Hzea.2665	BAC, pupae DNA	Pupal protein	+3.76731
45	Hzea.15446	fas-binding factor 1 homolog	Binding protein, cell stabilization	+3.75496
46	Hzea.19761	BAC, pupae DNA	Pupal protein	+3.67297
47	Hzea.12185	trypsin, alkaline C-like	Intestinal protein degradation	+3.62558
48	Hzea.23659	BAC, pupae DNA	Pupal protein	+3.57875
49	Hzea.29138	zinc finger protein OZF-like	DNA binding domain protein	+3.56187
50	Hzea.15307	pancreatic triacylglycerol lipase-like	Lipid metabolism	+3.50668

a. Gene number corresponds to sequence number in Fastq files. b. Gene ID annotations were found from an NCBI BLAST search top result (query coverage, E-value, percent identity). c. General function determined by NCBI and UniProt database searches. d. Plus indicates increased expression in the resistant strain.

Table 2. Top 50 (highest level of log2 fold change) down-regulated transcripts in the Bt-resistant strain of the bollworm, *Helicoverpa zea*, including gene identity, general function, and magnitude of log2 fold change.

	Gene[a]	Gene ID[b]	General function[c]	Log2 fold decrease[d]
1	Hzea.11969	cytochrome P450 337B3v1	Xenobiotic metabolism, associated with insecticide resistance	-7.18074
2	Hzea.27562	family 31 glucosidase KIAA1161-like	Glucose metabolism	-6.66764
3	Hzea.17125	protein ALP1-like	Cytoskeletal development	-6.26055
4	Hzea.29678	tudor domain-containing protein 7A	Post-transcriptional modification	-5.60295
5	Hzea.812	cytochrome P450 337B3v1	Xenobiotic metabolism, associated with insecticide resistance	-5.56349
6	Hzea.6938	splicing factor 3B subunit 4-like	Gene Splicing	-5.09605
7	Hzea.12845	putative nuclease HARBI1	Nucleic Acid cleavage	-5.01358
8	Hzea.31102	cuticle protein-like	Cuticle structural protein	-4.96903
9	Hzea.15432	transcription factor Adf-1-like	Adh gene expression regulation	-4.93912
10	Hzea.28236	ATP-binding cassette sub-family G member 1	ABC transporter protein	-4.92069
11	Hzea.14683	BAC, pupae DNA	Pupal Protein	-4.82383
12	Hzea.12475	isolate AD2 clone 1 microsatellite D47	Unknown	-4.78341
13	Hzea.7219	BAC, egg DNA	Oval DNA	-4.77338
14	Hzea.4118	multiple inositol polyphosphate phosphatase 1-like	Regulates cellular inositol levels	-4.63011
15	Hzea.28234	serine/threonine-protein kinase RIO1	Ribosomal subunit maturation	-4.61965
16	Hzea.15388	zinc finger protein 628-like	DNA binding domain protein	-4.61078
17	Hzea.13573	glucose dehydrogenase [FAD, quinone]-like	Glucose metabolism	-4.56858
18	Hzea.811	BAC 33J17 cytochrome P450 337B3v1	Xenobiotic metabolism, associated with insecticide resistance	-4.54212
19	Hzea.17013	small nuclear ribonucleoprotein F	Pre-mRNA splicing	-4.54025
20	Hzea.32168	UDP-glucuronosyltransferase 2B4-like	Glucuronidation catalysis	-4.51753
21	Hzea.4237	KGHa033C10 carboxyl/choline esterase CCE001f	Dietary detoxification, hormone/pheromone degradation, neurodevelopment	-4.4949
22	Hzea.21278	facilitated trehalose transporter Tret1-like	Trehalose transport from fat body	-4.46546
23	Hzea.25419	putative nuclease HARBI1	Nucleic Acid cleavage	-4.41916
24	Hzea.9745	BAC 18J13 cytochrome P450 337B2v2 and cytochrome P450 337B1v1	Xenobiotic metabolism, associated with insecticide resistance	-4.32618
25	Hzea.13850	BAC, pupae DNA	Pupal Protein	-4.30511
26	Hzea.12890	fatty acid synthase-like	Fatty acid synthesis	-4.27779
27	Hzea.4150	phospholipase A2-like	Fatty acid cleavage	-4.24604
28	Hzea.3034	BAC, pupae DNA	Pupal Protein	-4.20592
29	Hzea.13370	cytochrome P450 6B6	Xenobiotic metabolism	-4.19982

Table 2 cont.

	Gene[a]	Gene ID[b]	General function[c]	Log2 fold increase[d]
30	Hzea.11776	cuticle protein 65	Cuticle structural protein	-4.10268
31	Hzea.4423	lipase member K-like	Lipid metabolism	-3.98569
32	Hzea.23288	cuticle protein 1	Cuticle structural protein	-3.97585
33	Hzea.25418	putative nuclease HARBI1	Nucleic Acid cleavage	-3.95527
34	Hzea.6856	enoyl-CoA delta isomerase 1, mitochondrial-like	Fatty acid oxidation	-3.95465
35	Hzea.4455	dihydrofolate reductase	Dihydrofolic reduction	-3.931
36	Hzea.21115	zinc finger BED domain-containing protein 1-like	DNA binding domain protein	-3.9258
37	Hzea.14952	guanine nucleotide-binding protein G(q) subunit alpha	Guanine binding	-3.8824
38	Hzea.20452	cytochrome P450 337B2v2	Xenobiotic metabolism, associated with insecticide resistance	-3.8696
39	Hzea.13504	protein deltex	Cell communication, Notch pathway	-3.86954
40	Hzea.11171	organic cation transporter protein-like	Transport protein, cations	-3.81687
41	Hzea.14992	alkaline phosphatase 2	BT receptor in intestines	-3.77113
42	Hzea.30431	BAC, pupae DNA	Pupal Protein	-3.75267
43	Hzea.7746	cytochrome P450 337B2v2	Xenobiotic metabolism, associated with insecticide resistance	-3.71829
44	Hzea.5603	UDP-glucuronosyltransferase 2B7-like	Glucuronidation catalysis	-3.71518
45	Hzea.21691	BAC, pupae DNA	Pupal Protein	-3.69295
46	Hzea.17647	beta-secretase 1-like	Protein Cleavage	-3.64879
47	Hzea.32280	tyrosine--tRNA ligase	Ligation of tRNA and tyrosine, translation	-3.6039
48	Hzea.23371	CD209 antigen-like protein 2	Pathogen recognition receptor	-3.59733
49	Hzea.14510	BAC, pupae DNA	Pupal Protein	-3.5325
50	Hzea.21503	polyribonucleotide nucleotidyltransferase 1	Transferase protein	-3.48953

a. Gene number corresponds to sequence number in Fastq files. b. Gene ID annotations were found from an NCBI BLAST search top result (query coverage, E-value, percent identity). c. General function determined by NCBI and UniProt database searches. d. minus indicates decreased expression.

Table 3. Highly up-regulated (threshold log2 fold change ≥ 2.0) *Helicoverpa zea* transcripts associated with insecticide or Bt-resistance (categorized by gene family and organized by function) including gene identity, general function, and magnitude of log2 fold change.

	Gene[a]	Gene ID[b]	General function[c]	Log2 fold Change[d]
Tetraspanins	Hzea.18477	Tetraspanin 1	Protein stabilization, cell signaling pathways, associated with BT resistance	+4.92
	Hzea.11255	Tetraspanin 1	Protein stabilization, cell signaling pathways, associated with BT resistance	+3.37
	Hzea.498	Tetraspanin 1	Protein stabilization, cell signaling pathways, associated with BT resistance	+2.88
Serine Proteases	Hzea.23538	Serine Protease 9-like	Protein cleavage	+5.13
	Hzea.24399	allergen Api m 6-like	Serine Type Endopeptidase	+6.86176
	Hzea.7824	Serine Protease Snake-like	Protein cleavage	+3.1
	Hzea.5128	Serine Protease 9-like	Protein cleavage	+2.4
	Hzea.16515	Serine Protease 3-like	Protein cleavage	+2.23
Secretase Proteins	Hzea.10453	Gamma-Secretase	Cleavage of transmembrane proteins	+4.7
	Hzea.30068	Beta-Secretase 1-like	Protein Cleavage	+3.04
	Hzea.27773	Beta-Secretase 1-like	Protein cleavage	+2.07
Trypsin Proteins	Hzea.15497	Chymotrypsin 1-like	Intestinal protein cleavage	+4.44
	Hzea.4257	Trypsin 3A1-like	Intestinal protein cleavage	+5.31
	Hzea.15502	Trypsin-like Protease	Intestinal protein cleavage	+4.12
	Hzea.2400	trypsin 1-like	Intestinal protein cleavage	+3.94
	Hzea.12185	Trypsin, Alkaline C like	Intestinal protein cleavage	+3.62
	Hzea.12186	Trypsin, Alkaline C like	Intestinal protein cleavage	+2.17
Bt-Receptors	Hzea.14992	Alkaline Phosphatase 2	Intestinal receptor for Cry1Ac	-3.77
	Hzea.17125	Alkaline Phosphatase 1 like	Intestinal receptor for Cry1Ac	-6.26
	Hzea.11178	Mutant Cadherin (BtR)	Intestinal receptor for Cry1Ac	+2.93
	Hzea.18127	Cadherin-like	Intestinal receptor for Cry1Ac	-2.29
	Hzea.20	Caherin-r15	Intestinal receptor for Cry1Ac	-2.89
	Hzea.8058	Aminopeptidase 1D	Intestinal receptor for Cry1Ac	-2.02
	Hzea.16397	Aminopeptidase 1	Intestinal receptor for Cry1Ac	-2.21
Peptidases	Hzea.12825	Carboxypeptidase B like	Peptide cleavage	+3.04
	Hzea.7667	Carboxypeptidase B like	Peptide cleavage	-2.03

Table 3 cont.

	Gene[a]	Gene ID[b]	General function[c]	Log2 fold Change[d]
Transporters	Hzea.29517	Multidrug Resistance protein 1 like	Efflux transporter	+2.22
	Hzea.3344	Multidrug Resistance protein 4 like	Efflux transporter	-2.06
	Hzea.9148	ABC Transporter ABCC3	Transporter protein	+3.36
	Hzea.10318	ABC Transporter ABCC3	Transporter protein	+3.06
	Hzea.13698	ABC Transporter ABCC2	Transporter protein	+2.6
	Hzea.13541	ABC Transporter ABCC3	Transporter protein	+2.5
Chitin	Hzea.21994	Chitin Synthase A & B	Chitin synthesis	+3.11
	Hzea.18297	Heat Shock Protein 12.2	Cellular stress response	+5.02
Metabolic transcripts	Hzea.9448	CYP337B2v2 & CYP337B1v1	Xenobiotic metabolism, associated with resistance to insecticides	+4.23
	Hzea.29377	CYP337B3v1	Xenobiotic metabolism, associated with resistance to insecticides	+4.1
	Hzea.391	CYP337B2v2 & CYP337B1v1	Xenobiotic metabolism, associated with resistance to insecticides	+4
	Hzea.20633	CYP337B3v1	Xenobiotic metabolism, associated with resistance to insecticides	+3.37
	Hzea.6595	CYP367A2	Xenobiotic metabolism	+2.43
	Hzea.1468	CYP4AU1	Xenobiotic metabolism	+2.31
	Hzea.13660	CYP337B3v1	Xenobiotic metabolism, associated with resistance to insecticides	+2.22
	Hzea.18411	CYP421A5	Xenobiotic metabolism	+2.16
	Hzea.16306	CYP337B3v1	Xenobiotic metabolism, associated with resistance to insecticides	+2.04
	Hzea.29950	CYP337B3v1	Xenobiotic metabolism, associated with resistance to insecticides	+2.01
	Hzea.2403	carboxyl/choline esterase CCE001f	Xenobiotic metabolism, associated with resistance to insecticides	+6.17351
Protease	Hzea.3274	E3 ubiquitin-protein ligase Siah1-like	Proteosome mediated protein degradation	+5.46839

a. Gene number corresponds to sequence number in Fastq files. b. Gene ID annotations were found from an NCBI BLAST search top result (query coverage, E-value, percent identity). c. General function determined by NCBI and UniProt database searches. d. Plus indicates increased expression in the resistant strain, and minus indicates decreased expression.

Table 4. *Helicoverpa zea* differentially expressed transcripts associated with immune functions (organized by immune pathway or general function) including gene identity, strain of bollworm, general function, and magnitude of log2 fold change.

Gene[a]	Gene ID[b]	Strain	General function[c]	Log2 fold Change[d]
Hzea.27139	Peptidoglycan (PGN) Recognition Protein C	Resistant	IMD immune pathway	+1.86
Hzea.24180	lysM & PGN binding protein 2	Resistant	IMD immune pathway	+1.43
Hzea.15134	PGN recognition protein LB	Resistant	IMD immune pathway	+1.11
Hzea.8213	NF-kappa- beta p110 subunit	Resistant	IMD immune pathway	+0.88
Hzea.3860	NF-kappa- beta binding protein	Resistant	IMD immune pathway	+0.74
Hzea.11065	Immunoglobulin binding protein	Resistant	IMD immune pathway	+3.41
Hzea.15446	Fas-binding factor 1	Resistant	IMD immune pathway	+3.75496
Hzea.18171	Transforming Growth Factor Beta 1	Resistant	IMD immune pathway	+0.545215
Hzea.22952	Transforming Growth Factor Beta 1 Receptor	Resistant	IMD immune pathway	+0.486983
Hzea.12555	TGF-Beta Activated Kinase 1	Susceptible	IMD immune pathway	-0.584251
Hzea.2642	Cecropin 1	Susceptible	Antibacterial protein	-3.39285
Hzea.5285	Lysozyme- like	Resistant	Antimicrobial protein	+0.984174
Hzea.8918	Lysozyme-like	Susceptible	Antimicrobial protein	-0.469066
Hzea.5268	Lysozyme	Susceptible	Antimicrobial protein	-3.34181
Hzea.19610	Putative Defense Protein Hdd11	Resistant	Pathogen Defense	+1.11573
Hzea.28790	Beta- 1,3-glucanase protein	Resistant	Pathogen recognition protein	+1.47469
Hzea.30418	Nodulin 75 like	Resistant	General Immune response	+1.55155
Hzea.5226	Adaptor molecule Crk	Resistant	Immune signaling	+0.598283
Hzea.5227	Adaptor molecule Crk	Susceptible	Immune signaling	-0.79207
Hzea.8567	Autophagy protein 5	Resistant	Programmed cell death, removes unnecessary cells. Involved in immunity	+0.984792
Hzea.23318	Autophagy Related protein 16-1	Resistant	Programmed cell death, removes unnecessary cells. Involved in immunity	+0.918917
Hzea.21156	Autophagy Related protein 13	Resistant	Programmed cell death, removes unnecessary cells. Involved in immunity	+0.80002
Hzea.25537	Autophagy Related protein 2 homolog A	Susceptible	Programmed cell death, removes unnecessary cells. Involved in immunity	-0.690981
Hzea.25536	Autophagy related protein 2 homolog A	Susceptible	Programmed cell death, removes unnecessary cells. Involved in immunity	-1.00944
Hzea.25507	Autophagy related protein 2 homolog A	Only in Resistant	Programmed cell death, removes unnecessary cells. Involved in immunity	N/A
Hzea.2633	C-type lectin	Susceptible	Immune response to pathogens	-1.07
Hzea.19146	Death associated protein kinase 1	Resistant	Toll immune pathway	+0.678001
Hzea.15107	Toll like receptor 3	Resistant	Toll immune pathway	+1.29648
Hzea.4906	Toll like protein	Resistant	Toll immune pathway	+0.327495

Table 4 cont.

Gene[a]	Gene ID[b]	Strain	General function[c]	Log2 fold Change[d]
Hzea.24906	CD63 Antigen	Resistant	General Immune response	+1.08759
Hzea.23769	Antigen 8	Resistant	General Immune response	+0.582441
Hzea.25130	CD109 antigen like	Resistant	General Immune response	+0.529484
Hzea.24900	CD63 antigen like	Resistant	General Immune response	+0.513073
Hzea.25569	H13 Antigen	Resistant	General Immune response	+0.334133
Hzea.24905	Antigen CD53 like	Susceptible	General Immune response	-0.8215
Hzea.8177	cell nuclear antigen	Susceptible	General Immune response	-1.15965
Hzea.23371	CD209 antigen like protein 2	Susceptible	General Immune response	-3.59733
Hzea.9589	CD63 antigen	Only in Susceptible	General Immune response	N/A
Hzea.11558	CD109 antigen	Susceptible	General Immune response	-1.03885
Hzea.8646	tyrosine kinase hopscotch	Resistant	JAK/STAT pathway	+1.10303
Hzea.19419	signal transducer and activator of transcription 5B (STAT5B)	Resistant	JAK/STAT pathway	+0.589054
Hzea.17134	SH3 domain-containing kinase-binding protein 1	Resistant	JAK/STAT pathway	+0.676851
Hzea.3426	E3 SUMO-protein ligase PIAS3	Susceptible	JAK/STAT pathway	-0.79

a. Gene number corresponds to sequence number in Fastq files. b. Gene ID annotations were found from an NCBI BLAST search top result (query coverage, E-value, percent identity). c. General function determined by NCBI and UniProt database searches. d. Plus indicates increased expression in the resistant strain, and minus indicates decreased expression.

FIGURES

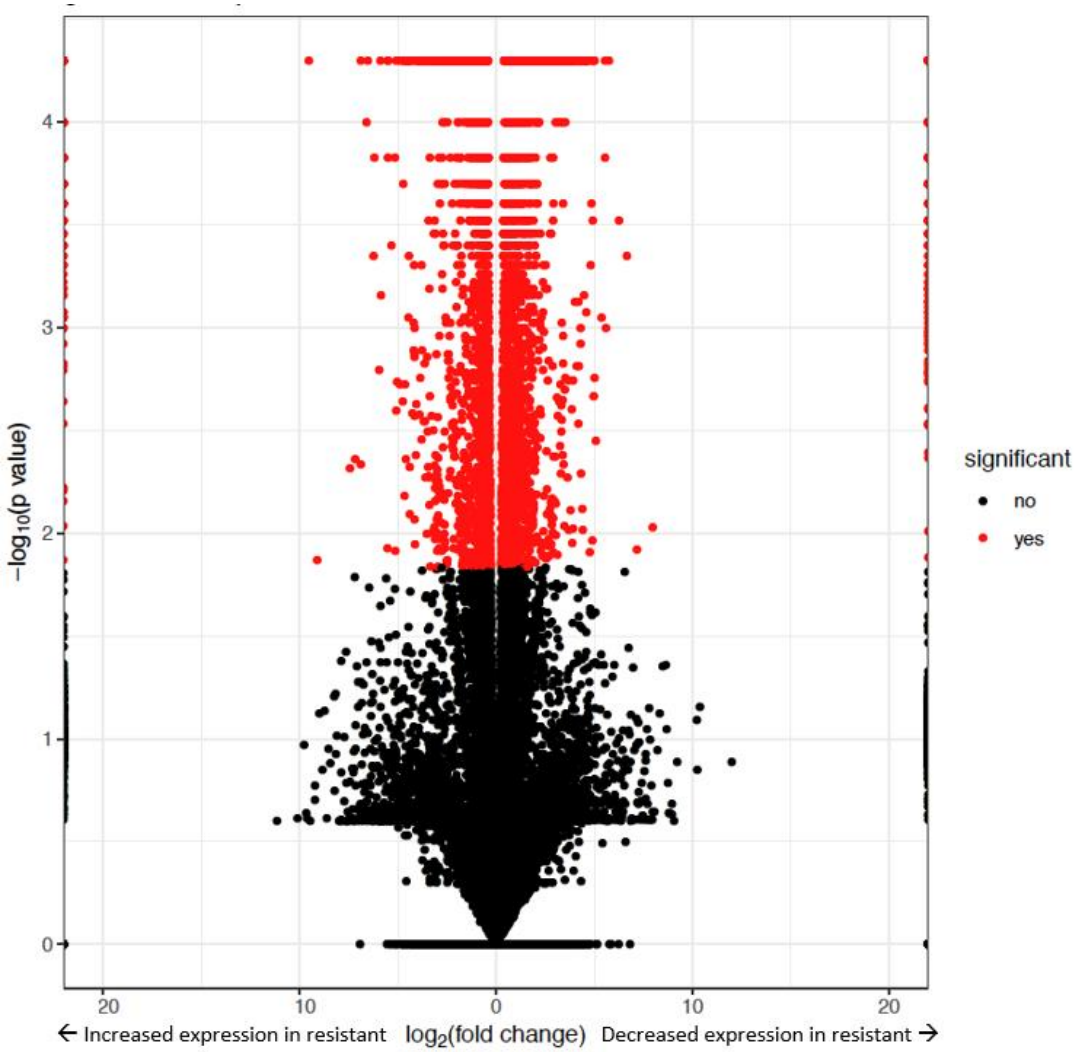


Figure 1. Volcano plot depicting significantly different and insignificantly different transcripts levels determined by RNAseq isolated from *Helicoverpa zea* Bt-resistant and susceptible strains. The Y-axis is a $-\log_{10}$ (p value) scale. The X-axis is a log2 (fold change) scale. Each data point indicates a transcript that was differentially expressed between the two strains of caterpillar. Data points to the left of 0 indicate transcripts with increased expression in the resistant strain. Data points to the right of 0 indicate transcripts with decreased expression in the resistant strain. Those in red had statistically significant differential expression levels (α=0.05). Those in black did not have statistically significant levels of differential expression in these data.

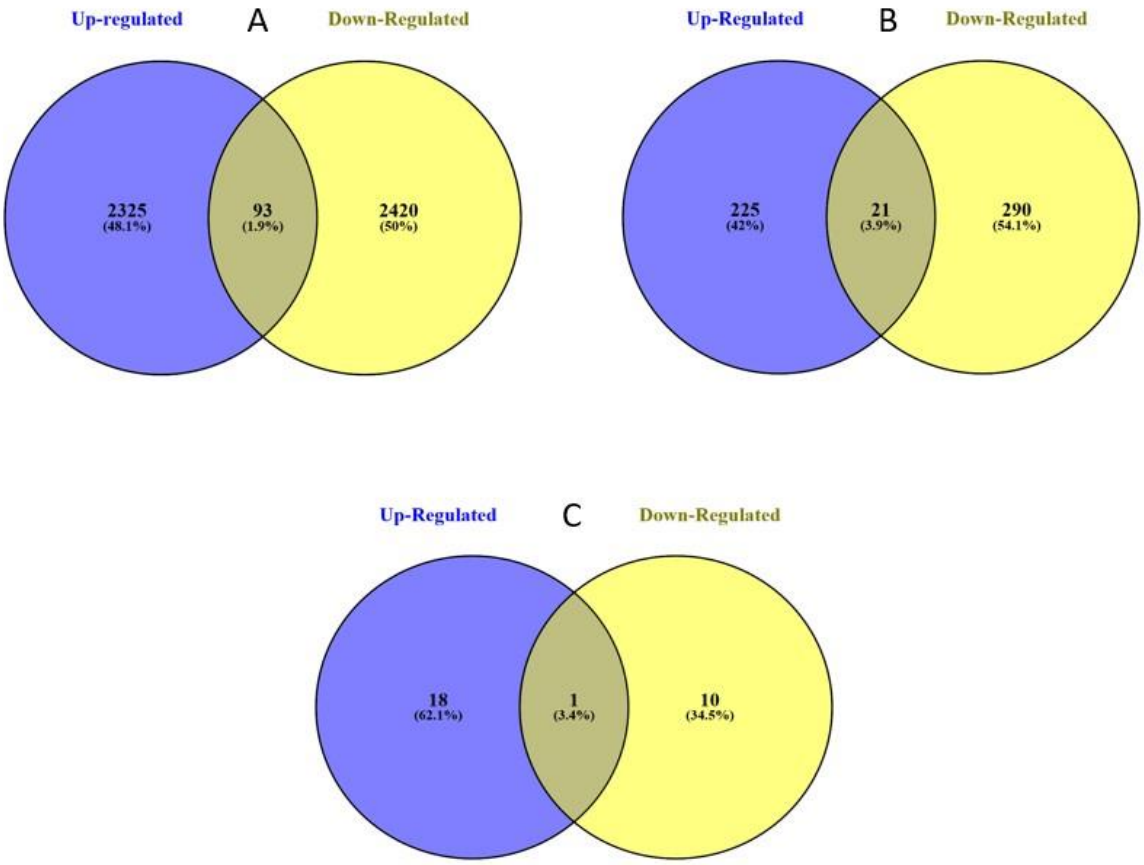
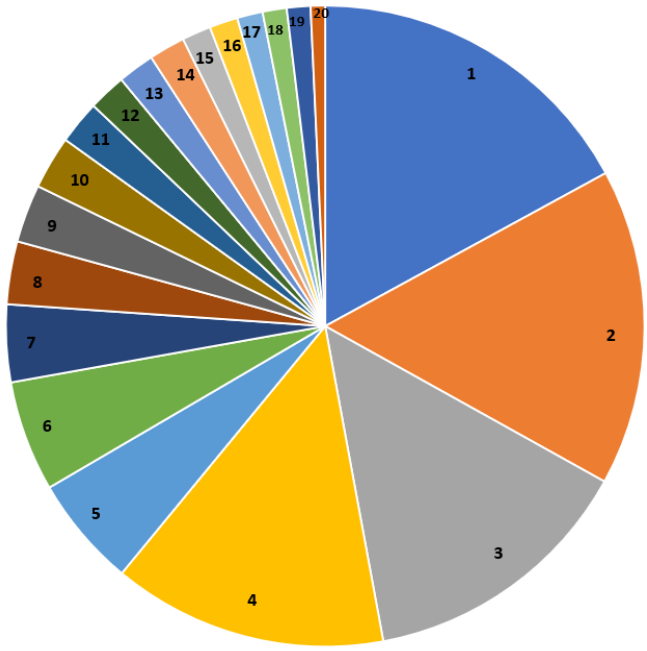


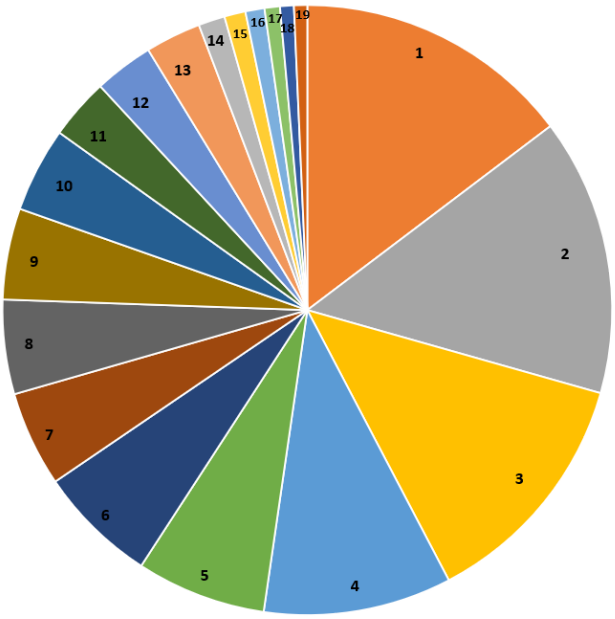
Figure 2. Numbers of statistically significant (α=0.05) differentially expressed transcripts using 3 different thresholds of log2 fold change for transcripts isolated from *Helicoverpa zea* Bt-resistant and susceptible strains. (A) All differentially expressed transcripts, (B) those greater than 2.0 log2 fold change in either direction, and (C) those greater than 5.0 log2 fold change in either direction. Blue indicates those transcripts with increased expression in the resistant strain. Yellow indicates those transcripts with decreased expression in the resistant strain. Numbers and percentages on the inside of each circle represent the total number of transcripts found to be only up- or down-regulated. Those numbers in the shared section of the Venn diagrams represent the numbers and percentages of transcripts that had copies that were both up- and down-regulated in the resistant strain. These were transcript variants (different mRNA sequences that code for the same protein). Additionally, shared transcripts could have been gene isoforms (mRNA sequences coding for the same protein with differing transcriptional start sites, untranslated regions, or protein coding regions). In some cases, a variant or isoform of a transcript was up-regulated, and in other cases was down-regulated.

Figure 3A-C. Biological processes (A), Molecular functions (B), and Cell component (C) gene ontologies for all differentially expressed transcripts isolated from *Helicoverpa zea* Bt-resistant versus susceptible strains. Chart on the left depicts the proportion of each biological process. The table to the right is the functional assignments and percentages for each biological process that were differentially expressed. Transcripts were annotated using the OmicsBox program (BLAST2GO function) and were categorized under "Biological process" [34]. Numbers for each biological process on the right (e.g., 1. Organic substance metabolic process) is shown by the respective number on the pie chart to the left. Numbers for each molecular function (e.g., 1. Organic cyclic compound binding) on the right is shown by the respective number on the pie chart to the left. Numbers for each cellular component (e.g., 1. Membrane) on the right is shown by the respective number on the pie chart to the left.



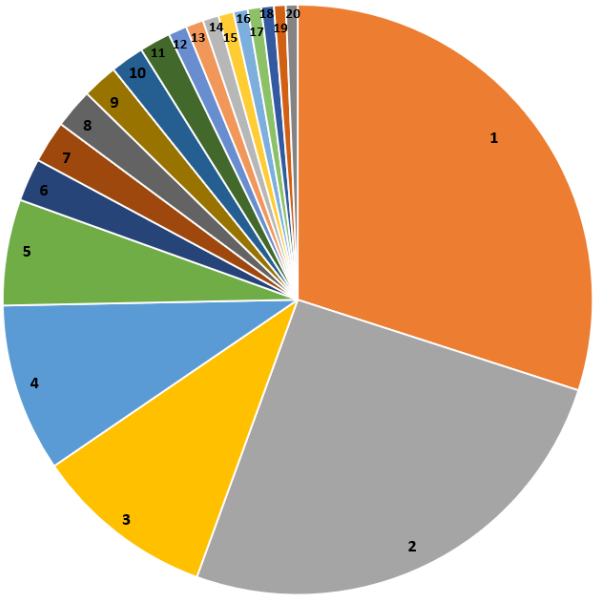
Biological Process	Percentage of Transcripts (%)
1. Organic substance metabolic process	17
2. Primary metabolic process	16
3. Nitrogen compound metabolic process	14
4. Cellular metabolic process	14
5. Biosynthetic process	6
6. Establishment of localization	6
7. Oxidation-Reduction process	4
8. Regulation of cellular process	3
9. Cellular response to stimulus	3
10. Small molecule metabolic process	3
11. Cellular component organization	2
12. Catabolic process	2
13. Cell communication	2
14. Signal transduction	2
15. Macromolecule localization	1
16. Cellular localization	1
17. Cellular component biogenesis	1
18. Regulation of metabolic process	1
19. Response to stress	1
20. Regulation of biological quality	1

Figure 3A.



Molecular Function	Percentage of Transcripts (%)
1. Organic cyclic compound binding	15
2. Heterocyclic compound binding	15
3. Ion binding	13
4. Hydrolase activity	10
5. Transferase activity	7
6. Small molecule binding	6
7. Carbohydrate derivative binding	5
8. Catalytic activity, acting on a protein	5
9. Drug binding	5
10. Oxidoreductase activity	4
11. Transmembrane transporter activity	3
12. Protein binding	3
13. Cofactor binding	3
14. Catalytic activity, acting on DNA	1
15. Catalytic activity, Acting on RNA	1
16. Structural constituent of ribosome	1
17. Signaling receptor activity	1
18. Lyase activity	1
19. Isomerase activity	1

Figure 3B.



Cellular Component	Percentage of Transcripts (%)
1. Membrane	30
2. Intrinsic component of membrane	26
3. Organelle	10
4. Intracellular organelle	9
5. Cytoplasm	6
6. Extracellular region	2
7. Membrane-enclosed lumen	2
8. Catalytic complex	2
9. Ribonucleoprotein complex	2
10. Endomembrane system	2
11. Envelope	2
12. Nucleoplasm	1
13. Cell periphery	1
14. Membrane protein complex	1
15. Sm-like protein family complex	1
16. Mitochondrial protein complex	1
17. Nuclear pore	1
18. Chromatin	1
19. Extracellular space	1
20. RNA polymerase complex	1

Figure 3C.

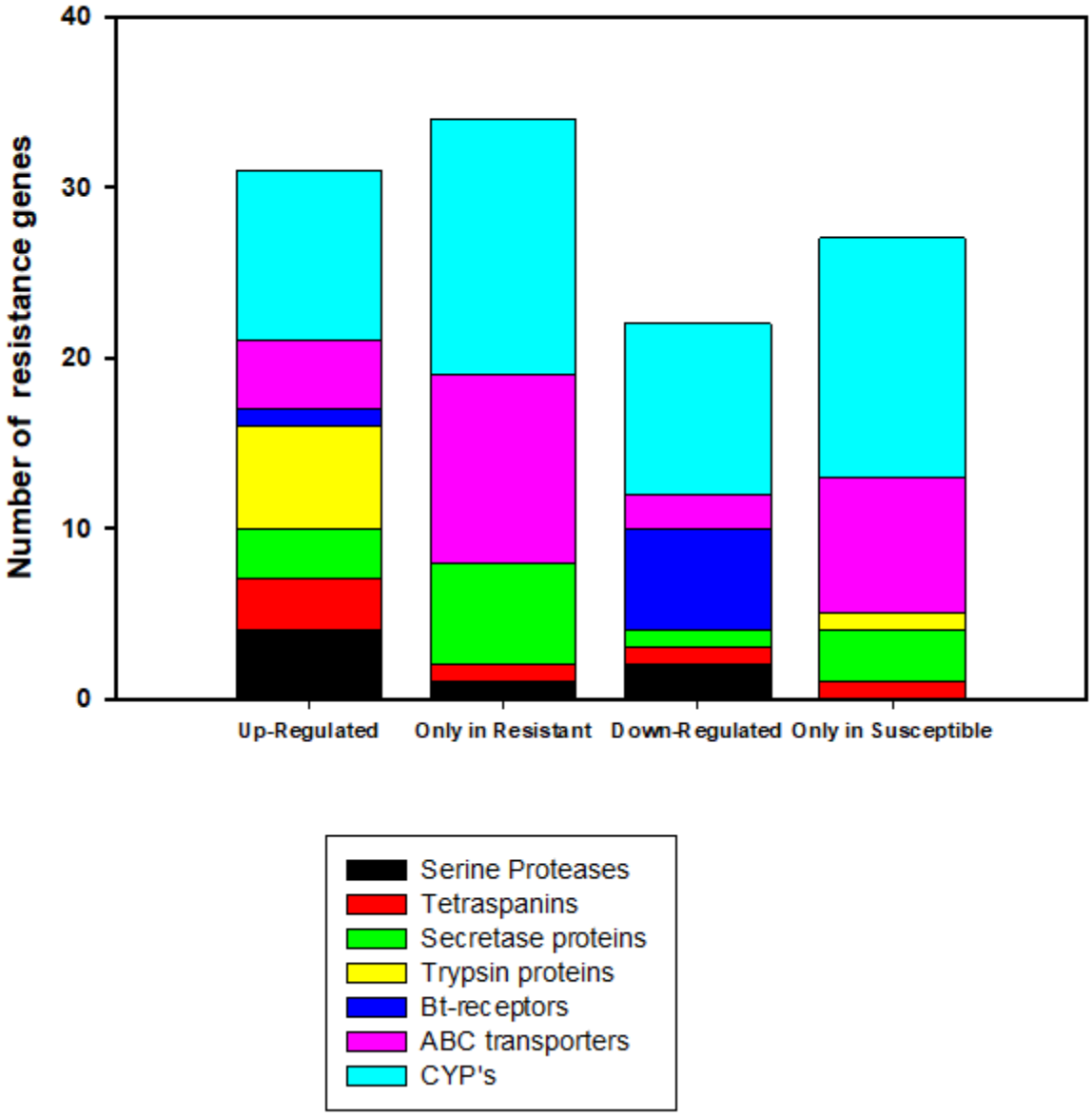


Figure 4. Number of Bt-resistance associated transcripts isolated from *Helicoverpa zea* Bt-resistant and susceptible strains organized by functional category. The Y-axis indicates the total number of transcripts for each category. The X-axis shows the transcripts that were up-regulated or down-regulated in the resistant strain, those found only in the resistant strain, and those found only in the susceptible strain. For those gene families "only in resistant" or "only in susceptible", these categories represent transcripts that were only present in either of these strains, and therefore were not differentially expressed. Gene families depicted in this chart are as follows: serine proteases (Black), tetraspanins (Red), secretase proteins (Green), trypsin proteins (Yellow), Bt-receptors (Blue), ABC transporters (Purple), and cytochrome P450s (CYP)(Turquoise).

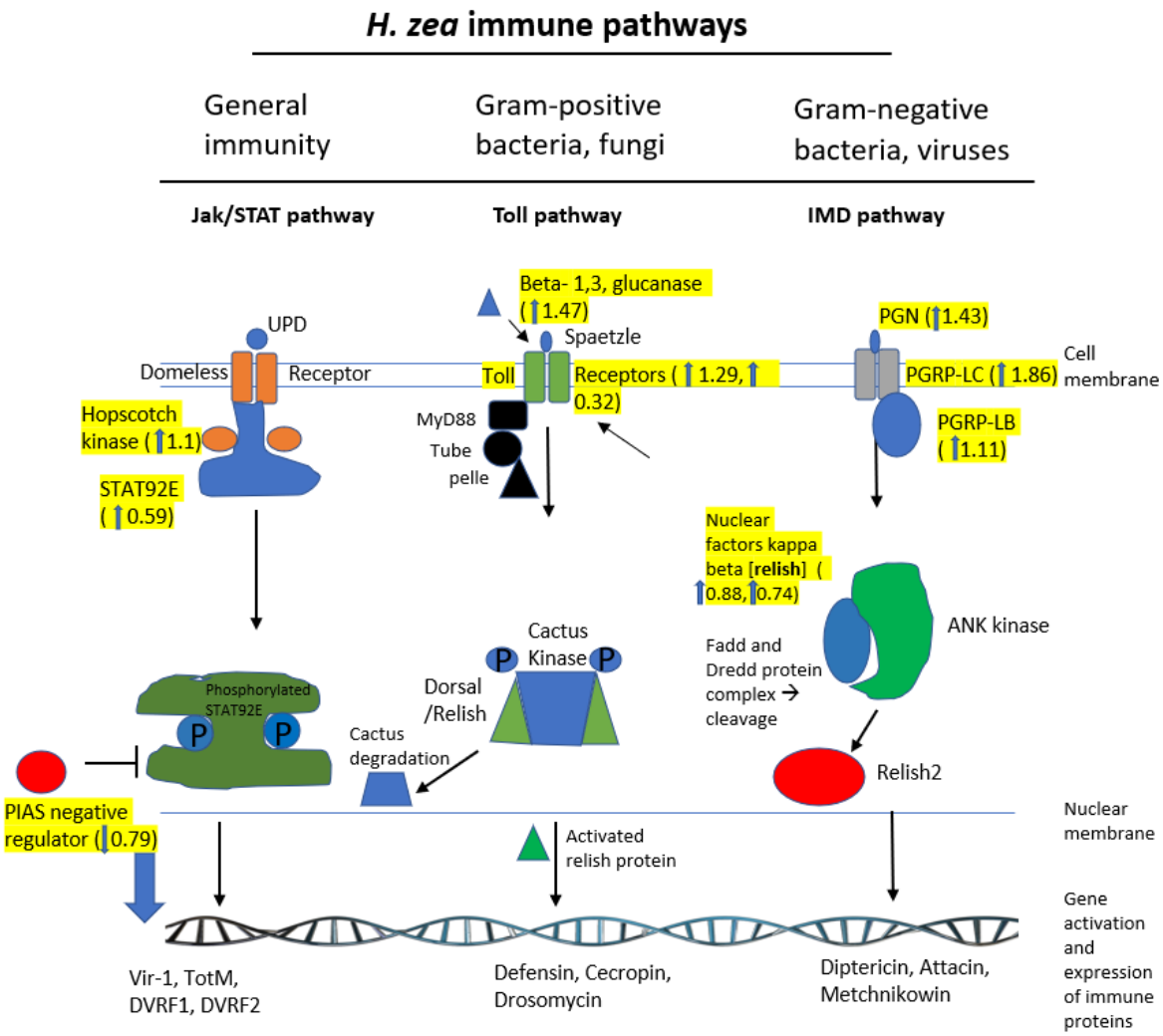


Figure 5. Generalized immune pathways found in *Helicoverpa zea* and differentially expressed transcripts (isolated from Bt-resistant and susceptible strains). Pathway components were found in other research [43-46]. Three immune pathways are shown, from left to right the Jak/STAT, Toll, and IMD pathways. The function of each pathway is shown at the top of the figure. Direction of the pathways move from top to bottom. The top begins with extracellular receptors or signals, the center corresponds to the cytoplasm, and bottom to the nuclear membrane and DNA level. Shapes indicate a protein involved in one of the pathways and are labelled appropriately. Those proteins highlighted in yellow are transcripts involved in any of these immune pathways that were differentially expressed. Magnitude (log2 fold change) and direction (up or down-regulated) of differential expression is indicated in the yellow highlighted section as well.

Chapter 3: Characterization of long non-coding RNA in the bollworm, *Helicoverpa zea*, and their possible role in Cry1Ac-resistance

Abstract

Multiple insect pest species have developed field resistance to Bt-transgenic crops. There has been a significant amount of research on protein-coding genes that contribute to resistance, such as the up-regulation of protease activity or altered receptors. However, our understanding of the role of lncRNA-mediated mechanisms in Bt-resistance is minimal as is also the case for resistance to chemical pesticides. To address this problem, RNA-seq was used to examine statistically significant, differential gene expression between a Cry1Ac-resistant (~100 fold resistant) and Cry1Ac-susceptible strain of *Helicoverpa zea*, a prevalent agricultural pest in the USA. Significant differential expression of long non-coding RNAs (lncRNAs) was found in the Cry1Ac-resistant strain (58 up- and 24 down-regulated gene transcripts with an additional 10 found only in resistant and 4 only in susceptible caterpillars). These lncRNAs were examined as potential pseudogenes and for their genomic proximity to coding genes, both of which can be indicative of a regulatory relationship between a lncRNA and gene expression. A possible cadherin interacting lncRNA was found. Also, lncRNAs were found significantly proximal to a serine protease, ABC transporter, and cytochrome P450 (CYP) coding genes, potentially involved in the mechanism of Bt and/or chemical insecticide resistance. lncRNA mediated mechanisms in *Helicoverpa zea* will improve the understanding of the genomic evolution of insect resistance, improve the identification of specific regulators of coding genes in general (some of which could be important in resistance), and is the first step for potentially targeting these regulators for pest control and resistance management (using molecular approaches like RNAi and others). More research is needed to study agricultural pest lncRNAs and their role in insect pest biology and resistance.

Introduction

In integrated pest management (IPM) practices, an effective method of pest control for many years has been Bt (*Bacillus thuringiensis*)-transgenic crops. Insecticidal proteins (including Cry family proteins) isolated from this bacteria have been cloned into commercial crops (corn, soybeans, cotton, etc.) and have been successful in the control of insect pest species such as lepidopterans [1]. Unfortunately, rapidly increasing levels of resistance have been detected worldwide in multiple pest species in recent years. These include but are not limited to populations of the fall armyworm, *Spodoptera frugiperda*, in Puerto Rico and North Carolina (USA); the maize stalk borer, *Busseola fusca*, in South Africa; the pink bollworm, *Pectinophora gossypiella*, in India and the USA; and the bollworm, *Helicoverpa zea*, in the USA [2-10]. *H. zea*, the model organism in this study, has been reported to have resistance to multiple Cry family proteins, including Cry1Ac, Cry1A.105, Cry1Ab, Cry2Ab2 and Cry1F [8-11]. *H. zea* is a prevalent polyphagous pest that annually causes significant economic damage to crops [8-11].It has also been noted that resistance to Cry1A family proteins has plateaued in *H. zea* but Cry2A family resistance is still being selected for [11]. Recently in *H. zea*, resistance to Vip3Aa has also been detected in the USA, the magnitude of which varies in individual populations [12-13].

There are multiple recognized mechanisms of Bt-resistance in insects. These are alterations in serine proteases, Bt-receptors in the gut cavity (cadherin, alkaline phosphatase, and aminopeptidase), transporters (ABC transporters), tetraspanin, secretase, and trypsin proteins

[14-21]. Our research team recently reported on the potential involvement of the insect immune system in Bt-resistance and changes in cytochrome P450s (CYP, P450s) [22]. Bt-receptors have been linked to resistance via mutations or altered binding of Bt-proteins to the mid-gut receptors [23-24]. It should also be noted that ABCC2 gene has been shown to not act as a Bt-receptor and that alkaline phosphatase 2 is attributed to Cry1Ac but not Cry2Ab resistance [25-26]. Yang et. al, (2020) linked Vip3Aa resistance inheritance to be monogenic, autosomal, and ressesive however the exact gene annotation is unclear [12]. An understudied area of Bt-resistance is the involvement of lncRNAs.

Long non-coding RNAs (lncRNAs) are RNA strands over 200 nucleotides in length that are not translated into proteins but are structurally similar to mRNA [27]. Mechanistically, lncRNAs can act in a variety of roles, most notable in gene regulation. LncRNAs can affect *cis* or *trans* gene expression, transcription factor activation or repression through binding and localization, chromatin remodeling, imprinting, and enhancer regulation [27]. They are also involved in broader regulatory processes such as post-transcriptional regulation, protein trafficking, and mRNA processing [27]. Chromatin remodeling activity occurs via lncRNA mediated histone methylation, which leads to a shift in chromatin structure, either increasing or decreasing access of transcriptional machinery to DNA strands [28]. In this role, the lncRNA acts as a protein scaffold with several binding domains allowing proteins involved in methylation and demethylation to interact with the target histone [28]. Transcriptional suppression can also occur when a lncRNA acts as a "decoy" for RNA polymerases by binding the polymerase to a lncRNA instead of the normal target DNA [29]. LncRNAs can also regulate translation by binding to mRNAs, increasing translation, decreasing translation, or promoting the bound mRNA's degradation [30].

LncRNAs have been studied most in humans and animals used in medical research. Research is minimal by comparison in insects, even less for agricultural pests, and there has been almost no work on their role in pesticide resistance. The lncRNA regulatory role in insect transcriptional processes was studied [31]. In honeybees, *Apis mellifera,* high expression levels of lncRNAs were found in ovaries, likely because they play a role in developmental processes such as social caste determination [32]. In fall armyworms, *Spodoptera frugiperda*, lncRNA expression was correlated to heterochromatin formation [33]. In *Drosophila*, lncRNAs appeared to coordinate sex determination [34].

There is recent evidence that lncRNAs are involved in insecticide resistance, increased fitness, and responsiveness to xenobiotic exposure [31,35]. For example, lncRNAs were associated with chlorpyrifos insecticide resistance in the diamondback moth, *Plutella xylostella.* That study found that lncRNA overexpression likely regulate the increased expression of resistance-associated genes such as those that code for detoxifying enzymes [31]. The brown rice plant hopper, *Nilaparvata lugens*, has become rapidly resistant to many different insecticides, with high fecundity compared to the susceptible strain. It was discovered that significant differences existed in the lncRNA profiles between the two strains, suggesting that lncRNAs had a role in resistance [35]. In the pink bollworm, a specific lncRNA was responsible for transcriptional regulation of the *P. gossypiella* cadherin 1 (*pgCad1)* which encodes for a midgut receptor known to be involved in Bt-resistance. Using RNA interference (RNAi) that targeted the lncRNA, larval Bt-toxin susceptibility was altered [36].

The objective of this study was to examine the role of lncRNA in Bt-resistance in the bollworm, *H. zea*. The insects used in this study were Cry1Ac-resistant or Cry1Ac-susceptible (referred to later as Bt-resistant or Bt-susceptible for brevity). Additionally, this study aims to increase the overall understanding of the functional role of lncRNA in insects. A shotgun RNA-seq approach was used to compare the gene expression profiles of a Bt-resistant and a Bt-susceptible strain of bollworm. From this dataset, lncRNA sequences were isolated for analysis.

Materials and Methods

Sample collection and preparation

Helicoverpa zea eggs were acquired from a Bt-susceptible and Bt-resistant colony. The susceptible insects were from a laboratory strain reared with no Bt exposure for 18 years acquired from Benzon Research, Inc. (Carlisle, PA). The susceptible colony was collected from Wake Forest, North Carolina, USA in non-Bt corn. Both colonies were reared on artificial diet in the lab for 2 generations [50]. To minimize strain differences as much as possible, both the resistant and susceptible bollworms were reared using the same rearing methods in the same laboratory at NCSU on the same artificial diet, under the same environmental conditions. Rearing conditions in the growth chamber were as follows: 14:10 L:D, 27 ºC:24 ºC L:D, and 60% RH, and the moths were mated to conspecifics for each colony. The artificial food was modified *H. zea* diet; the modification is described in Reisig et al. [50]. The same rearing methods were used as described in Reisig et al. for both resistant and susceptible colonies [50]. Cry1Ac susceptibility bioassays were also performed. The observed difference in susceptibility was 100-fold (the Cry1Ac LC_{50} was 43.79 µg/cm^2 for the Bt-resistant strain and 0.43 µg/cm^2 for the Bt-susceptible strain). All rearing and bioassay methods are described in detail in Lawrie et al. and Reisig et al. [22,50].

RNA extraction

F**r**om the colonies, 5 Bt-resistant samples and 5 Bt-susceptible samples were prepared, each replicate sample consisting of 10 neonate *H. zea*. All neonates were unfed and lab-reared preceding RNA extraction as described earlier. Neonates were mechanically homogenized into one DNAse and RNAse free tube for each sample within 6 h of emergence. From each pooled sample, total RNA was extracted using the RNeasy Mini Kit following the manufacturer's protocol (Qiagen, Valencia, CA, USA). Purity of total RNA in each sample was then evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) by the NC State University Genomics Core Facility (Raleigh, NC, USA). Sequencing was conducted on samples that had a RNA Integrity Number >9.0.

RNA sequencing

The NCSU Genomics Core Facility performed RNA-seq for this experiment. DNA libraries for each sample (using 500 µg of total RNA each) were prepared for RNA-seq using the TruSeq RNA Library Prep Kit v2 (Illumina, San Diego, CA, USA) following the manufacturer's protocol. Transcriptome sequencing was conducted on the NextSeq 500 System (Illumina, San Diego, CA, USA) using a paired end setting with a read length of 2 x 150 base pairs. A sequencing depth of >25 million reads per library was obtained using a High Output Flow Cell. A total of 10 TruSeq libraries were then prepared, 5 each for resistant and susceptible. The SRA Toolkit v2.9.2 was used to convert raw reads to fastq files [51]. Read quality was then evaluated

using the FastQC tool v0.11.7 [52]. A Phred score of >30 was required for a majority of the sequencing reads to establish a baseline for quality. Fastq files with suitable quality were then used for assembly and quality control steps.

Transcript assembly and quality control

Transcript assembly and quality control was performed by the NC State Bioinformatics Core (Raleigh, NC, USA). Reads were assembled with the StringTie program (v1.3.5, John Hopkins University, Baltimore, Maryland, USA) with 45,224 primary transcripts assembled into transcript set 1 using the *H. zea* reference genome [53]. The program Trinity (v2.8.4, Broad Institute and Hebrew University of Jerusalem, Jerusalem, Israel) was used to assemble an alternate set of transcripts (set 2) that were not aligned with StringTie in order to maximize transcript assemblies [54]. For transcript assembly, there were 149,108 transcripts assembled and processed using the Blobology program (v2.15.2, University of Edinburgh, UK) to ascertain the presence of contaminants [55]. Transcripts matching to Lepidoptera were then saved (108,867 transcripts). From these, all ribosomal RNA transcripts were deleted. The remaining 108,841 transcripts were clustered with the Evigene program (v1.0, University of Indiana, Indiana, USA) which resulted in 34,059 transcripts in set 1 [56]. Transcript sets 1 and 2 were then combined and clustered using Evigene resulting in 26,800 primary and 12,095 alternate transcripts. Primary and alternate sets were then analyzed with Blobology to check for contaminates again. Ribosomal RNA transcripts were also removed from these sets. Primary and alternate transcripts were subsequently clustered and combined with Evigene.

The TrimmoMatic sequence trimmer (v0.39, Max Planck Institute, Germany) was used to trim fastq files for each replicate for adapter sequence and quality [57]. The *H. zea* reference genome (NCBI), was used to map each trimmed file to the reference genome using HiSat2 [58]. StringTie was then used to assemble resulting mapped files to assemble RNA-seq alignments into potential transcripts. All transcript annotations from each replicate were then combined into one “expressed transcriptome” file. This was used to guide gene boundaries when calculating differential expression values (log2 fold change) between the susceptible and resistant strains with CuffDiff (v7.0, Cambridge, Massachusetts, USA) [59]. Statistical significance was determined using the Tuxedo Pipeline (in CuffDiff, which assigned transcript q-values, α=0.05). Only statistically significant transcripts were included in later data analysis [60]. These results were then imported into the R statistical software platform for quality control checks and visualization of results [61]. The sequence of transcripts that were determined to be differentially expressed were extracted from the reference genome and used in BLASTn searches against insects to provide initial annotations. Quality control steps and data analysis were conducted with volcano plots, FPKM, boxplots PCA plot, MDS plot, normalization, and heatmaps. These steps were performed to ensure replicates were of sufficient quality, mapping rate and variation among replicates. Each replicate passed all quality control steps. After assembly and quality control was conducted for all transcripts, 6,098 transcripts were identified as differentially expressed in this experiment. Of these, 3,042 transcripts had higher expression in the susceptible strain with 267 being found only in this strain. The remaining 3,056 had higher expression in the resistant strain with 323 being only expressed in the resistant strain. Blast2GO (v5.2.4) was utilized to annotate open reading frame assignments [62] and functionally. Gene ID and function was determined using BLASTx (E- value cut off 10^{-5}), using lepidopteran taxonomy to filter results, using the nr and swissprot databases [62].

Data analysis and figure construction

In order to categorize transcripts as "long non-coding RNAs", all transcripts that were annotated as non-coding RNAs were separated from protein coding genes. GenBank was used to examine sequence length from these transcripts defined as non-protein coding genes by NCBI BLAST (BLASTx). All transcripts that were 200 bases or greater, were categorized as long non-coding RNAs (lncRNAs). A lncRNA is a non-protein coding RNA over 200 base-pairs in length [27]. There were 2 non-protein coding RNA transcripts (Hzea.11974 and Hzea.13128) with < 200 base pairs that were excluded. Figure 1 depicts a visualization of the above process. Figures and tables for this paper were prepared using Microsoft Excel, PowerPoint, Word (2018), and SigmaPlot (v14.0, SigmaPlot, Systat Software, San Jose, CA). All sequence alignments were conducted using mega BLAST (BLASTn), which also provided percent alignments, E-values, and query coverages [63]. The R statistical software platform was used to construct a volcano plot [61]. For each table, the fasta sequence for each of the lncRNAs identified were run through BLASTn to analyze the E-values, percent identities, and query coverages to assess candidate lncRNA validity. This was also done to determine if any candidate lncRNAs had been characterized since RNA-sequencing was conducted.

The putative pseudogene analysis was performed by selecting 5 lncRNAs with the highest log2 fold increase, 5 with the highest log2 fold decrease, and 2 at random from the only in resistant and 2 at random from the only in susceptible categories. Ten coding genes with known associations to Bt-resistance were selected (*trypsin, serine protease, tetraspanin, cadherin, and beta-secretase*). The coding gene with the highest log2 fold increase and highest log2 fold decrease for each of the coding gene types listed above was selected for comparison to the lncRNAs. NCBI BLASTn was used to align each lncRNA sequence to each coding gene sequence. E-value, percent identity, query coverage, and score values were assessed.

For proximity analysis, genomic scaffolds were visualized for analysis using the Integrative Genomics Viewer (IGV) program (v2.9.2, Broad Institute, Jerusalem, Israel) (http://software.broadinstitute.org/software/igv/home) to compare genomic proximities of lncRNAs to coding genes. First the *H. zea* reference genome (NCBI) was loaded into IGV followed by all of the differentially expressed transcripts (GTF file) from our RNAseq work. We selected 5 lncRNAs with the highest log2 fold increase, 5 with the highest log2 fold decrease, and 2 at random from the only in resistant and 2 at random from the only in susceptible categories. Each of the lncRNAs were then located in IGV using scaffold numbers and coordinates. Protein coding genes within 1 million bp of each lncRNA were then located. The coding genes were identified for annotation using NCBI BLASTx and the lncRNAs examined as potential pseudogenes proximal to coding genes on each scaffold being studied. All proximal (within 1 million bp) coding genes and lncRNAs were aligned with NCBI BLASTn and E-value, percent identity, query coverage, and score were assessed.

Results

Characterization of lncRNA in H. zea

The workflow for the identification of lncRNAs is shown in Figure 1. A lncRNA was defined as an RNA sequence that annotated as a non-coding sequence using BLASTx and was over 200 base-pairs in length from RNAseq data obtained from replicated Bt-resistant and Bt-

susceptible unfed, neonates of *H. zea*. Only the statistically significant, differentially expressed lncRNAs between the resistant versus susceptible strains are reported, and these are in 2 categories: (i) those found in both strains and (ii) those found only in one strain. In the former, we were able to calculate a log2 fold increase (up-regulation) or decrease (down-regulation) in expression for the resistant strain compared to the susceptible strain (Supplementary Tables S1-2). For lncRNAs only in the resistant or susceptible strains, a log2 fold difference between strains could not be calculated since the expression level was zero in the comparative strain (Supplementary Table S3).

There were 98 non-coding transcripts that were differentially expressed of which 96 were ≥ 200 bases and classified as lncRNA. Figure 2 shows the range of sequence lengths for the differentially expressed lncRNAs. The lengths ranged from 297 to 6719 bp with most sequences between 200-1000 bases. Additionally, lncRNAs that were up-regulated in Bt-resistant bollworms were among the longest lncRNAs in this study, where 8 were above 2500 bases in length; however, most of these lncRNAs did not have a large degree of log2 fold increase (all above 2500 bp had below 2.0 log2 fold increase) (Figure 2). For down-regulated lncRNAs, there was only 1 lncRNA above 2500 bases in length, where again the log2 fold decrease was below 2.0 (Figure 2). For the only in resistant and only in susceptible groups, there was 1 lncRNA for each category over 2500 bases in length; however, the other lncRNAs in these categories were all below 1000 bases (Figure 2).

Differential expression of long non-coding RNAs in Bt-resistant H. zea.

Log2 fold changes for lncRNAs with increased expression in resistant insects ranged from 4.88 to 0.38, and for decreased expression ranged from 4.41 to 0.44 (Supplemental Figure S1). Those lncRNAs with an increased expression not only had a higher maximum log2 fold change in comparison to those with decreased expression (4.88 log2 fold increase vs. 4.44 log2 fold decrease, respectively) but also had a consistently higher number of lncRNAs with higher degrees of expression. There were 33 up-regulated lncRNAs above 1.0 log2 fold change and 13 down-regulated lncRNAs above 1.0 log2 fold change.

In order to further compare the differences in lncRNA expression between the two strains of *H. zea*, different thresholds of expression were compared (Figures 3-4). When examining the 96 differentially expressed lncRNAs identified, there were 58 with increased expression in the Bt-resistant strain, 24 with decreased expression, 10 only in the resistant strain, and 4 only in the susceptible strain (Figure 3A). Using a threshold of ≥1.0 log2 fold change, there were 33 up-regulated lncRNAs and 13 down-regulated lncRNAs (Figure 3B). Using a threshold of ≥2.0 log2 fold change, there were 17 up-regulated lncRNAs and 6 down-regulated lncRNAs (Figure 3C). The 5 lncRNAs with the highest log2 fold change lncRNAs in either direction are shown in Figure 4. The top 5 up-regulated transcripts were LOC113506107 with a 4.88 log2 fold increase, LOC113508874 with a 4.65 log2 fold increase, LOC110372550 at 3.9, LOC110380503 at 3.5, and LOC110371745 at 3.44. The top 5 down-regulated transcripts were LOC110373805 at a 4.41 log2 fold decrease, LOC110373534 at 3.68, LOC110382662 at 3.46, LOC110383440 at 3.05, and LOC110369725 at 2.83. Overall, not only did the Bt-resistant strain have a higher number of up-regulated lncRNAs but also the magnitude of log2 fold changes were consistently higher.

Functional characterization of long non-coding RNAs in Bt-resistant H. zea

A pseudogene is an imperfect copy of a functional coding gene [37]. In order to determine whether there were any pseudogenes present in our data, NCBI BLASTn was used to compare differentially expressed lncRNA sequences to differentially expressed protein coding genes with functions known to be important in Bt-resistance. Often pseudogenes act as regulators for coding genes where they share ancestral traits (ex. cytochrome P450s and cytochrome P450 pseudogenes); we found evidence of this before for primary human liver cells held in short term culture and treated with the pesticides, DEET, and fipronil [37]. Five lncRNAs with the highest log2 fold increase, 5 with the highest log2 fold decrease, 2 found only in the resistant strain and 2 only in the susceptible strain were compared by NCBI BLASTn with 5 coding genes with the highest log2 fold increase and 5 with the highest log2 fold decrease for a serine protease, ABC transporter, trypsin, secretase, and tetraspanin. These proteins have functions known to be important in Bt-resistance (Figure 5). A majority of the sequences did not have any significant alignments. All results are depicted in the supplemental data table (Supplemental Table S4). The best pseudogene candidate was with lncRNA LOC110369725 and cadherin XJ-r15 (Figure 5). The alignment was as follows: E-value=0, percent identity=99.07%, query coverage=81%, max score=950, total score=1002. A BLASTx alignment of cadherin XJ-r15 indicated that multiple exons are present in the sequence, which could indicate that the lncRNA LOC110369725 does not interact with the cadherin coding section. Rather this approach was a first step in characterization.

Genomic proximity also can indicate that two genes have a functional relationship with each other. Significant genomic proximity is defined as a distance *cis* or *trans* within 1000 kb [38]. In our previous work with primary short-term cultures of human liver cells, a number of differentially-regulated lncRNAs affected by exposure to the pesticides, DEET and Fipronil, were found within significant genomic proximity to differentially regulated xenobiotic metabolizing genes [37]. To examine proximity relationships that might be important in Bt-resistance, we identified the genome scaffolds that contained the 5 lncRNAs with the highest log2 fold increase, 5 with the highest log2 fold decrease, 2 found only in the resistant and 2 only in the susceptible bollworm strains (Figure 6). We then located all coding genes within significant proximity up-stream and down-stream of each lncRNA, and these were annotated by NCBI BLASTx. Even though proximity is defined as 1 million base pairs *cis* and *trans* from the lncRNA, proximity measurements were smaller because of the smaller scaffold size. The results of this analysis are shown in Figure 7A-E and Supplemental Figures S2-S5. A wide-variety of coding genes were found in genomic proximity to the lncRNAs we examined. Most interesting, known Bt-resistance associated genes were found in genomic proximity to a number of these lncRNAs. These were a CYP (Hzea.12028, CYP6B7, CYP6B6, CYP6B2) (Figure 7A), an ABC transporter (Hzea.20383, ABCC3, ABCC2, ABCC1) (Figure 7B), and a serine protease (Hzea.7824, LOC110382673, serine protease snake-like) (Figure 7C). Among the lncRNAs we examined, there were also lncRNAs that did not have any genomic proximities (Figure 7D), and those that were uncharacterized or unrelated to Bt-resistance coding genes (Figure 7E). Each proximal Bt-resistance associated genes was less than 100 kb up- or down-stream from the lncRNA (Figure 7A-C). The proximal CYP was 7.868 kb from lncRNA LOC11350610 (this lncRNA was up-regulated) (Figure 7A), the proximal ABC transporter 50.672 kb from lncRNA LOC110369725 (this lncRNA was down-regulated) (Figure 7B), and the serine protease 0.646 kb from lncRNA LOC110382674 (this lncRNA was found only in the resistant strain) (Figure

7C). The lncRNA presented in Figure 8D was down-regulated and the lncRNA presented in Figure 8E was found only in the susceptible strain (Figure 7C-D).

We also looked for putative pseudogenes among the proximal coding genes and lncRNAs (Figure 6) with NCBI BLASTn comparisons. There were no significant alignments found to Bt-resistance associated genes (Figures 7A-E). However, there was a lncRNA (LOC110372708) that aligned to a previously characterized prostaglandin pseudogene (Figure S2D) using the same methodology used to find the putative cadherin pseudogene (Figure 5). The lncRNAs that had significant proximities to Bt-resistance associated genes (Figures 7A-C) were also aligned with each other using BLASTn to see if they had any any significant similar regulatory potential. Each of these 3 lncRNAs did not shown any significant alignment. All of the scaffolds studied were less than 1 million bp in length on either side of the lncRNA. Therefore, it is possible that other significant proximities exist that could not be detected in our research.

Discussion

Characterization of long non-coding RNAs in H. zea

We successfully identified 96 differentially expressed putative non-protein coding sequences in *H. zea* that are candidate lncRNAs. This study focused only on statistically significantly differentially expressed lncRNAs between a Bt-resistant and Bt-susceptible strain of *H. zea* (Figure 2, Supp. Tables S1-3). These 96 transcripts were those that were statistically significantly differentially expressed (Tuxedo suite pipeline determined significance) [22]. In the diamondback moth, *Plutella xylostella*, 3,324 lncRNAs were identified by RNA-seq [39]. In the human genome, 100,000 lncRNAs have been found where differential expressions of several lncRNAs was linked to a wide-variety of pathologies [40]. Therefore, it is likely that there are additional lncRNAs present in *H. zea*; they simply were not significantly differentially expressed between the Bt-resistant and Bt-susceptible strains and were not considered in our study. It is possible that the lncRNAs that were differentially expressed in this study are important in the mechanism of Bt-resistance, and/or differential expression could be caused by insect strain differences. The Bt-resistant strain was obtained from the field and the Bt-susceptible strain was reared in the laboratory for many generations on artificial diet. To minimize strain differences, both the resistant and susceptible bollworms were reared in the same lab under the same conditions with the same artificial diet for 2 generations and the study was conducted on unfed neonates. This approach minimizes any possible differences in developmental rates after hatching and diet preferences. Furthermore, we are measuring constitutive gene expression levels essentially at the time of hatching and before feeding effects can occur. This study was only a first step in characterizing lncRNAs in *H. zea*, more work is needed to confirm these potential lncRNAs presented here.

When we examined lncRNA length (Figure 2), the majority of transcripts ranged between 200-1000 bp, but there were several lncRNAs above 4000 bp. LncRNAs have the capability to exhibit secondary structures such as a stem loop and cloverleaf structure [41]. The type of secondary or tertiary lncRNA structure can provide evidence of it mechanism for regulating a protein coding gene. In the case of the lncRNA HOTAIR (which is 2.2 kb), it was discovered that double stem loops at the 5’ and 3’ ends of the lncRNA are important in chromatin remodeling, a regulatory function of HOTAIR [41]. Another lncRNA, MALAT1 (over 8.7 kb), was discovered to exhibit a cloverleaf secondary structure at the 3’ end; similar to cloverleaf

functions in tRNAs, these are linked to subcellular localization [41]. Sequence length is useful to consider because it is a predictor of possible lncRNA secondary structure. In this study, there were 9 lncRNA candidates above 3 kb (with log2 fold changes below 2.0 in either direction). This could indicate that these particular lncRNAs have secondary structures important in the mechanism of regulating protein-coding genes. Further modeling is needed of these >3kb, differentially expressed lncRNAs in Bt-resistant bollworms, to determine their 3-D structure and possible mechanism of action.

Differential expression of long non-coding RNAs in Bt-resistant H. zea

In the Bt-resistant strain there were a significantly greater overall number of lncRNAs with increased expression levels (58 up- vs. 24 down-regulated) (Supplemental Figure 1 & Figure 3A). Additionally, when examining higher magnitudes of log2 fold change, there were both greater numbers of lncRNAs with high degrees of log2 fold change and an overall average greater magnitude of log2 fold change (Supplemental Figure 1 & Figures 3B-C). It is possible that the magnitude of log2 fold change does not signify importance to Bt-resistance. However, the high degrees of lncRNA up-regulation in Bt-resistant *H. zea* is one argument for their functional role in Bt-resistance. For example, a specific lncRNA up-regulated in this experiment may be acting as an enhancer for a coding gene involved in Bt-resistance such as a serine protease or a ABC transporter. One major functional role of lncRNAs is increasing coding gene expression [27,37]. In the pink bollworm, *Pectinophora gossypiella*, it was discovered that the lncRNA pgCad1 lncRNA is a specific enhancer RNA (eRNA) of the cadherin gene pgCad1 [36]. When pgCad1 lncRNA was silenced using siRNA, the expression levels of pgCad1 (and also Bt-susceptibility) were significantly reduced [36]. Therefore, it is possible that an up-regulated lncRNA identified in this study is linked to up-regulation of a Bt-resistance coding gene. LncRNAs can also act to decrease expression for coding genes causing down-regulation [27].

It is important to note that a lab-reared Bt-susceptible strain was used as a reference strain in these RNA-seq experiments. Ideally, a field-collected Bt-susceptible strain of *H. zea* would be used as a reference for a field-collected Bt-resistant strain. However due to the prevalence of Bt-resistance in wild *H. zea* populations, completely susceptible insects are difficult to collect in sufficient numbers to easily establish a colony. In order to maximize the comparison of the Bt-resistant insects with the suseptible strain, we used the lab-reared susceptible strain (Benzon) which has been used before by multiple research groups [22,42-43]. However, research is needed to develop a better comparative, Bt-suceptible field strain of the bollworm.

Predicting function of long non-coding RNAs in Bt-resistant H. zea

Pseudogenes are imperfect copies of parent coding-genes; often called “genomic fossils”, they can play a key role in regulation of their parent genes (ex., an lncRNA derived from a CYP pseudogene regulates a CYP mRNA) [37,44]. Certain pseudogenes are known to transcribe for non-coding RNAs such as lncRNAs [45-46]. The non-coding RNAs derived from pseudogene sequences can then act as specific regulators of parent coding-genes [45-46]. Previously thought to be untranslated, some pseudogenes are translated in humans; it is unknown if this occurs in insects [45-46]. Pseudogenes can be located anywhere within a genome but often are adjacent to their functional parent genes [37,44]. In this study, we were able to identify one promising candidate for a pseudogene that may be regulating an important Bt-resistance associated gene (lncRNA LOC110369725 and cadherin XJ-r15) (Figure 5). One regulatory function of

pseudogenes is their processing into piRNAs (piwi-interacting RNAs), where the pseudogene, once spliced into smaller piRNAs functions in RNAi-mediated gene silencing [47]. It could be that the proposed pseudogene LOC110369725 is being processed into piRNAs prior to regulatory interactions with cadherin XJ-r15. More research is needed to confirm this hypothesis. The characterization of pseudogene function is a rapidly evolving field where new data is changing our understanding of pseudogenes as new research is produced.

There are likely many more pseudogenes present within *H. zea*. The analysis performed in this study focused on those potential pseudogenes that were differentially expressed in Bt-resistant insects. Pseudogenes unrelated to Bt-resistance are likely to be present in the genome. For example, another pseudogene annotated as a prostaglandin reductase pseudogene was discovered proximal to one of the differentially expressed lncRNAs we examined (Figure S2D). Further identification of pseudogenes in *H. zea* would provide greater insight into the genomic functioning of this important pest species and the potential use of psuedogenes as targets for gene editing and pest management. Characterization of pseudogenes in insects would also be useful in understanding the evolution of genes throughout an insect's natural history.

Significant genomic proximity (within 1,000,000 bp) can be indicative of a relationship between two sequences [37-38,48]. In this study, we examined the genomic scaffolds of some of the greatest differentially expressed lncRNAs. Several significant proximities were discovered for a wide-variety of coding genes (Figure 7A-E, see also supplemental data). Interestingly, 3 lncRNAs with significant proximities to coding genes related to resistance, i.e., CYP, an ABC transporter, and a serine protease (Figure 7A-C). In particular, the serine protease gene was within 1000bp of the proximal lncRNA (Figure 7C). It is possible that due to the significant proximities to these coding genes, the lncRNA LOC113506107 (Figure 8A), lncRNA LOC110369725 (Figure 7B), and lncRNA LOC110382674 (Figure 7C) act as regulators in some capacity to the proximal coding genes with functions at least associated with Bt-resistance. However, it is unlikely that these particular lncRNAs are pseudogenes due to no significant alignments being present after BLASTn. In the pink bollworm, *P. gossypiella*, a specific lncRNA that is intronic to a cadherin gene has been established as an enhancer of that cadherin [36]. Additionally, in human liver cell models, we found a link (high level of similarity) between a wide variety of lncRNAs and proximal coding genes important in drug metabolism [37]. We did not find this to be the case in our bollworm study. By identifying specific regulators of coding genes important to Bt-resistance, it is possible that novel means of resistance management may be developed. For example, RNAi mediated silencing of an lncRNA could be used to enhance expression of a cadherin (or another type of Bt-receptor), increasing Bt-susceptibility. Bt-susceptibility has successfully been altered before using this technique targeting a lncRNA regulating the cadherin gene in *P. gossypiella* [36]. Additionally, gene-editing approaches that target non-coding genes might be more useful in insect resistance management and insect control than targeting coding genes and with greater non-target effects such as RNAi impacting the target species and other closely related species. In plants for example, lncRNAs have a high degree of intraspecies conservation with greater sequence diversity between species [49]. Targeting lncRNAs for resistance management may be more species specific than targeting coding genes; however, much more research and characterization of lncRNAs in insects is needed. This study has only identified a small number of lncRNAs that could be important to Bt-

resistance but is a step towards a greater understanding of how lncRNAs work in insects in general and in Bt-resistance.

In summary, this study examined the differential regulation of lncRNAs in a Bt-resistance strain of unfed neonates of the bollworm, H. *zea*. Overexpression of lncRNAs in other lepidopteran models has been correlated to chemical and Bt-insecticide resistance [31,36]. This study provides a comprehensive list of lncRNAs in *H. zea* associated with Bt-resistance and predicts potential regulatory roles therein for the first time. We characterized a possible pseudogene and multiple examples of genomic proximity between a differentially regulated lncRNAs and differentially regulated protein-coding genes where the protein function is a known mechanism for Bt-resistance. It is likely that additional lncRNAs to those that were differentially expressed between the Bt-resistant and Bt-susceptible strains are present.

REFERENCES

1. Fleming, D.; Musser, F.; Reisig, D.; Greene, J.; Taylor, S.; Parajulee, M.; Lorenz, G., Catchot, A.; Gore, J.; Kerns, D.; Stewart, S.; Boykin, D.; Caprio, M.; Little, N. Effects of transgenic *Bacillus thuringiensis* cotton on insecticide use, heliothine counts, plant damage, and cotton yield: A meta-analysis, 1996-2015. *PLoS ONE* **2018**, *13*, e0200131.

2. Moar, W.; Roush, R.; Shelton, A.; Ferre, J.; MacIntosh, S.; Leonard, B.R.; Abel, C. Field-evolved resistance to Bt toxins. *Nat. Biotechnol*. **2008**, *26*, 1072.

3. Tabashnik, B.E.; Van Rensburg, J.; Carrière, Y. Field-evolved insect resistance to Bt crops: definition, theory, and data. *J. Econ. Entomol*. **2009**, *102*, 2011-2025.

4. Huang, F.; Qureshi, J.A.; Meagher, R.L.Jr.; Reisig, D.D.; Head, G.P.; Andow, D.A.; Ni, X.; Kerns, D.; Buntin, G.D.; Niu, Y. Cry1F resistance in fall armyworm *Spodoptera frugiperda*: single gene versus pyramided Bt maize. *PLoS ONE* **2014,** *9*, e112958.

5. van Rensburg, J.B.J. First report of field resistance by the stem borer, *Busseola fusca* (Fuller) to Bt-transgenic maize. *South Afr. J. Plant Soil* **2007**, *24*, 147-151.

6. Gassmann, A.J.; Petzold-Maxwell, J.L.; Keweshan, R.S.; Dunbar, M.W. Field-evolved resistance to Bt maize by western corn rootworm. *PLoS ONE* **2011**, *6*, e22629.

7. Gassmann, A.J.; Petzold-Maxwell, J.L.; Clifton, E.H.; Dunbar, M.W.; Hoffmann, A.M.; Ingber, D.A.; Keweshan, R.S. Field-evolved resistance by western corn rootworm to multiple *Bacillus thuringiensis* toxins in transgenic maize. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 5141– 5146.

8. Tabashnik, B.E.; Fabrick, J.A.; Unnithan, G.C.; Yelich, A.J.; Masson, L.; Zhang, J.; Bravo, A.; Soberón, M. Efficacy of genetically modified Bt toxins alone and in combinations against pink bollworm resistant to Cry1Ac and Cry2Ab. *PLoS ONE* **2013**, *8*, e80496.

9. Tabashnik, B.E.; Zhang, M.; Fabrick, J.A.; Wu, Y.; Gao, M.; Huang, F.; Wei, J.; Zhang, J.; Yelich, A.; Unnithan, G.C. Dual mode of action of Bt proteins: protoxin efficacy against resistant insects. *Sci. Rep*. **2015**, *5*, 15107.

10. Dively, G.P.; Venugopal, P.D.; Finkenbinder, C. Field-evolved resistance in corn earworm to Cry proteins expressed by transgenic sweet corn. *PLoS ONE* **2016**, *11*, e0169115.

11. Yu, W.; Lin, S.; Dimase, M. ; Niu, Y.; Brown, S.; Head, G.P.; Price, P.A.; Reay-Jones, F.P.F.; Cook, D.; Reisig, D.; Thrash, B.; Ni, X.; Paula-Moraes, S.V.; Huang, F. Extended investigation of field evolved resistance of the corn earworm *Helicoverpa zea* (Lepidoptera:

Noctuidae) to *Bacillus thuringiensis* Cry1A.105 and Cry2Ab2 proteins in the southeastern United States. *Journal of Invertebrate Pathology*. **2021**. 183:107560.

12. Yang, F.; Santiago Gonzalez, J.C.; Sword, G.A.; Kerns, D.L. Genetic Basis of resistance to the Vip3Aa Bt protein in *Helicoverpa zea*. *Pest Management Science*. **2020**. 77, 3: 1530-1535.

13. Yang, F.; Kerns, D.L.; Little, N.S.; Santiago Gonzalez, J.C.; Tabashnik, B.E. Early warning of resistance to Bt toxin Vip3Aa in *Helicoverpa zea*. *Toxins (Basel).* **2021**. 13(9): 618.

14. Soberón, M.; Pardo, L.; Muñóz-Garay, C.; Sánchez, J.; Gómez, I.; Porta, H.; Bravo, A. Pore formation by Cry toxins. In *Proteins membrane binding and pore formation*, Anderluh, G., Lakey, J., Eds.; Landes Bioscience: Texas, USA, **2010**.; pp. 127-142.

15. Pardo-Lopez, L.; Soberon, M.; Bravo, A. *Bacillus thuringiensis* insecticidal three domain Cry toxins: mode of action, insect resistance and consequences for crop protection. *FEMS Microbiol. Rev.* **2013**, *37*, 3-22.

16. Jurat-Fuentes, J.L.; Karumbaiah, L.; Jakka, S.R.K.; Ning, C.; Liu, C.; Wu, K.; Jackson, J.; Gould, F.; Blanco, C.; Portilla, M. Reduced levels of membrane-bound alkaline phosphatase are common to lepidopteran strains resistant to Cry toxins from Bacillus thuringiensis. *PLoS ONE* **2011,** *6*, e17606.

17. Gahan, L.J.; Pauchet, Y.; Vogel, H.; Heckel, D.G. An ABC transporter mutation is correlated with insect resistance to *Bacillus thuringiensis* Cry1Ac toxin. *PLoS Genet.* **2010**, *6*, e1001248.

18. Atsumi, S.; Miyamoto, K.; Yamamoto, K.; Narukawa, J.; Kawai, S.; Sezutsu, H.; Kobayashi, I.; Uchino, K.; Tamura, T.; Mita, K. Single amino acid mutation in an ATP-binding cassette transporter gene causes resistance to Bt toxin Cry1Ab in the silkworm, *Bombyx mori*. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E1591-E1598.

19. Jin, L.; Wang, J.; Guan, F.; Zhang, J.; Yu, S.; Liu, S.; Xue, Y.; Li, L.; Wu, S.; Wang, X. Dominant point mutation in a tetraspanin gene associated with field-evolved resistance of cotton bollworm to transgenic Bt cotton. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 11760-11765.

20. Zhang, M.; Wei, J.; Ni, X.; Zhang, J.; Jurat-Fuentes, J.L.; Fabrick, J.A.; Carrière, Y.; Tabashnik B.E.; Li, X. Decreased Cry1Ac activation by midgut proteases associated with Cry1Ac resistance in *Helicoverpa zea*. *Pest Manage. Sci.* **2019**, *75*, 10199-1106.

21. Jurat-Fuentes, J.L.; Heckel, D.G.; Ferre, J. Mechanisms of resistance to insecticidal proteins *from Bacillus thuringiensis*. *Annual Review of Entomology*. **2021**. 66: 121-140.

22. Lawrie, R.D.; Mitchell, R.D. iii; Deguenon, J.M.; Ponnusamy, L.; Reisig, D.D.; Del Pozo-Valdivia, A.; Kurtz, R.W.; Roe, R.M. Multiple Known Mechanisms and a Possible Role of Enhanced Immune System in Bt-Resistance in a field Population of the Bollworm, *Helicoverpa zea:* Differences in Gene Expression with RNAseq. *International Journal of Molecular Sciences*. **2020**, 21, 18, 6528. Doi: 10.3390/ijms21186428.

23. Caccia, S.; Moar, W.J. ; Chandrashekhar, J. ; Oppert, C. ; Anilkumar, K.J.; Jurat-Fuentes, J.L.; Ferré, J. Association of Cry1Ac toxin resistance in *Helicoverpa zea* (Boddie) with increased alkaline phosphatase levels in the midgut lumen. *Applied and environmental microbiology*, **2012**. 78(16), 5690-5698.

24. Fritz, M.L.; Nunziata, S.O.; Guo, R.; Tabashnik, B.E.; Carrière, Y. Mutations in a novel cadherin gene associated with Bt resistance in Helicoverpa zea. *G3: Genes, Genomes, Genetics*, **2020**. 10(5), 1563-1574.

25. Perera, O.P.; Little, N.S.; Abdelgaffar, H.; Jurat-Fuentes, J.L.; Reddy, G.V.P. Genetic Knockouts Indicate That the ABCC2 Protein in the Bollworm *Helicoverpa zea* Is Not a Major Receptor for the Cry1Ac Insecticidal Protein. *Genes*, **2021**. 12, 1522. https://doi.org/10.3390/genes12101522

26. Wei, J.; Zhang, M.; Liang, G.; Li, X. Alkaline phosphatase 2 is a functional receptor of Cry1Ac but not Cry2Ab in *Helicoverpa zea*. *Insect molecular biology*, **2018**. 28(3), 372-379.

27. Dempsey, J.L.; Cui, J.Y. Long Non-Coding RNAs: A Novel Paradigm for Toxicology. *Toxicol Sci*. **2016**; *155(1)*:3–21. doi:10.1093/toxsci/kfw203.

28. Wang, K.C.; Chang, H.Y. Molecular mechanisms of long noncoding RNAs. *Mol. Cell.* **2011**, 43, 904–914.

29. Mariner, P.D.; Walters, R.D.; Espinoza, C.A.; Drullinger, L.F.; Wagner, S.D.; Kugel, J.F.; Goodrich, J.A. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol. Cell*. **2008**, *29*, 499–509.

30. Carrieri, C.; Cimatti, L.; Biagioli, M.; Beugnet, A.; Zucchelli, S.; Fedele, S.; Pesce, E.; Ferrer, I.; Collavin, L.; Santoro, C.; Forrest, A.R.R.; Carninci, P.; Biffo, S.; Stupka, E.; Gustincich, S. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* **2012,** *491*, 454–457.

31. Etebari, K.; Furlong, M.J.; Asgari, S. Genome wide discovery of long intergenic non-coding RNAs in Diamondback moth (*Plutella xylostella*) and their expression in insecticide resistant strains. *Sci Rep*. **2015**, *5*, 14642. doi:10.1038/srep14642.

32. Humann, F.C.; Tiberio, G.J.; Hartfelder, K. Sequence and expression characteristics of long noncoding RNAs in honey bee caste development- potential novel regulators for transgressive ovary size. *PLoS One* **2013**, *8*, e78915.

33. Stanojcic, S.; Gimenez, S.; Permal, E.; Cusserans, F.; Quesneville, H.; Fournier, P.; d'Alencon, E. Correlation of LNCR rasiRNAs expression with heterochromatin formation during development of the holocentric insect *Spodoptera frugiperda*. *PLoS One* **2011**, *6*, e24746.

34. Mulvey, B.B.; Olcese, U.; Cabrera, J.R.; Horabin, J.I. An interactive network of long non-coding RNAs facilitates the *Drosophila* sex determination decision. *Biochim. Biophys. Acta*. **2014**, *1839*, 773–784.

35. Xiao, H.; Yuan, Z.; Guo, D.; Hou, B.; Yin, C.; Zhang, W.; Li, F. Genome-wide identification of long noncoding RNA genes and their potential association with fecundity and virulence in rice brown planthopper, *Nilaparvata lugens*. *BMC Genomics*. **2015**, *16*, 749. doi:10.1186/s12864-015-1953-y.

36. Li, S.; Hussain, F.; Unnithan, G.C.; Dong, S.; UlAbdin, Z.; Gu, S.; Mathew, L.G.; Fabrick, J.A.; Ni, X.; Carrière, Y.; Tabashnik, B.E.; Li, X. A long non-coding RNA regulates cadherin transcription and susceptibility to Bt toxin Cry1Ac in pink bollworm, *Pectinophora gossypiella*. *Pestic. Biochem. Physiol*. **2019**, *158*, 54-60. doi: 10.1016/j.pestbp.2019.04.007. PMID: 31378361.

37. Lawrie, R.D.; Mitchell, R.D. iii; Dhammi, A.; Wallace, A.; Hodgson, E.; Roe, R.M. Role of long non-coding RNA in DEET- and fipronil-mediated alteration of transcripts associated with Phase I and Phase II xenobiotic metabolism in human primary hepatocytes. *Pestic. Biochem. Physiol.* **2020**, *167*, 104607. doi: 10.1016/j.pestbp.2020.104607. PMID: 32527422.

38. McLean, C.Y.; Bristor, D.; Hiller, M.; Clarke, S.L.; Schaar, B.T.; Lowe, C.B.; Wenger, A.M.; Bejerano, G. GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotech.* **2010**, *28*, 495–501.

39. Liu, F.; Guo, D.; Yuan, Z.; Chen, C.; Xiao, H. Genome-wide identification of long non-coding RNA genes and their association with insecticide resistance and metamorphosis in diamondback moth, *Plutella xylostella*. *Sci. Rep.* **2017**, 7, 15870. https://doi.org/10.1038/s41598-017-16057-2.

40. Kopp, F.; Mendell, J.T. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell*. **2018**, 172, 3, 393-407. doi:10.1016/j.cell.2018.01.011.

41. Novikova, I.V.; Hennelly, S.P.; Sanbonmatsu, K.Y Sizing up long non-coding RNAs: do lncRNAs have secondary and tertiary structure? *Bioarchitecture*. **2012**, *2*, 6, 189-199. doi:10.4161/bioa.22592.

42. Carriere, Y.; Degain, B.; Unnithan, G.C.; Harpold, V.S.; Li, X.; Tabashnik, B.E. Seasonal Declines in Cry1Ac and Cry2Ab concentration in maturing cotton favor faster evolution of resistance to pyramided Bt cotton in *Helicoverpa zea* (Lepidoptera: Noctuidae). *J Econ Entomol.* **2019**, *112*, 6, 2907-2914. doi: 10.1093/jee/toz236.

43. Welch, K.L.; Unnithan, G.C.; Degain, B.A.; Wei, J.; Zhang, J.; Li, X.; Tabashnik, B.E.; Carriere, Y. Cross-resistance to toxins used in pyramided Bt crops and resistance to bt sprays in *Helicoverpa zea*. *J Inverebr. Pathol*. **2015**, *132*, 149-156. doi: 10.1016/j.jip.2015.10.003. Epub 2015 Oct 10.

44. Tutar, Y. Pseudogenes. *International Journal of Genomics.* **2012**. https://doi.org/10.1155/2012/424526.

45. Ji, Z.; Song, R.; Regev, A.; Struhl, K. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife*, **2015**. DOI: 10.7554/eLife.08890.

46. Cui, H.; Jiang, Z.; Zeng, S. ; Wu, H.; Zhang, Z.; Guo, X.; Dong, K.; Wang, J. Shang, L.; Li, L. A new candidate oncogenic lncRNA derived from pseudogene WFDC21P promotes tumor progression in gastric cancer. *Cell Death and Disease*, **2021**. 12, 903.

47. Wang, C.; Lin, H. Roles of piRNAs in transposon and pseudogene regulation of germline mRNAs and lncRNAs. *Genome Biol*. **2021**, *22*, 1, 27. doi:10.1186/s13059-020-02221-x.

48. Ponting, P.P.; Oliver, O.L.; Reik W. Evolution and functions of long noncoding RNAs. *Cell*. **2009**, *136*, 629–641.

49. Deng, P.; Liu, S.; Nie, X.; Weining, S.; Wu, L. Conservation analysis of long non-coding RNAs in plants. *Science China Life Sciences*. **2018**, *61*, 190-198.

50. Reisig, D.D.; Huseth, A.S.; Bacheler, J.S.; Aghaee, M.A.; Braswell, L.; Burrack, H.J.; Flanders, K.; Greene, J.K.; Herbert, D.A.; Jacobson, A. Long-term empirical and observational evidence of practical *Helicoverpa zea* resistance to cotton with pyramided Bt toxins. *J. Econ. Entomol.* **2018**, *111*, 1824–1833.

51. Leinonen, R.; Sugawara, H.; Shumway, M. International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res*. **2011**, *39*, D19–D21.

52. Babraham Bioinformatics: FastQC: A quality control tool for high throughput sequence data. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed on 8 January 2019).

53. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol*. **2015**, *33*, 290–295.

54. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512.

55. Kumar, S.; Jones, M.; Koutsovoulos, G.; Clarke, M.; Blaxter, M. Blobology: Exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front. Genet*. **2013**, *4*, 237.

56. Rago, A.; Gilbert, D.G.; Choi, J.H.; Sackton, T.B.; Wang, X.; Kelkar, Y.D.; Werren, J.H.; Colbourne, J.K. OGS2: Genome re-annotation of the jewel wasp *Nasonia vitripennis*. *BMC Genom.* **2016**, *17*, 678.

57. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120.

58. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360.

59. Trapnell, C.; Hendrickson, D.; Sauvageau, S.; Goff, L.; Rinn, J.L.; Pachter, L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **2013**, *31*, 46–53.

60. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D.R.; Pimentel, H.; Salzberg, S.L.; Rinn, J.L.; Pachter, L. Differential gene and transcripts expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **2012**, *7*, 562–578, doi:10.1038/nprot.2012.016.

61. R Development Core Team. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2012.

62. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676.

63. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. "Basic local alignment search tool." *J. Mol. Biol.* **1990**, *215*, 403-410.

FIGURES AND TABLES

TABLES

Table S1. Increased expressed lncRNAs in Bt-resistant bollworms where a fold change could be calculated, including their fastq file ID (Figure 2 numerical ID), gene annotation, magnitude of log2 fold increase, sequence length, BLASTn (NCBI) description, and fasta sequence BLASTn result details (E-value, % identity, and query coverage). The table was organized from greatest log2 fold increase to lowest.

Gene ID[a]	Gene Annotation (Accession #)[b]	Log2 Fold Increase[c]	BLASTn Result Description[d]	Sequence Length (bp)[e]	E-value[f]	% Identity[g]	Query Coverage (%)[h]
Hzea.12022 (1)	gi|1496272281|ref|XR_003401692.1|	4.8782	LOC113506107	1128	3.00E-84	85.58	36
Hzea.2506 (2)	gi|1496241937|ref|XR_003402053.1|	4.6556	LOC113508874	893	9.00E-55	80.89	25
Hzea.17450 (3)	gi|1199381004|ref|XR_002429340.1|	3.9043	LOC110372550	573	0.00E+00	95.78	98
Hzea.13715 (4)	gi|1199401540|ref|XR_002429826.1|	3.4978	LOC110380503	650	4.00E-169	94.42	76
Hzea.28004 (5)	gi|1199378765|ref|XR_002429309.1|	3.4416	LOC110371745	584	0.00E+00	95.95	72
Hzea.23814 (6)	gi|1199381581|ref|XR_002429343.1|	3.075	LOC110372708	358	2.00E-36	96.08	10
Hzea.30980 (7)	gi|1199404819|ref|XR_002429930.1|	3.0222	LOC110381881	854	8.00E-32	80	43
Hzea.21858 (8)	gi|1199370955|ref|XR_002430115.1|	2.9312	LOC110383908	1760	0.00E+00	90.3	70
Hzea.13175 (9)	gi|1199369172|ref|XR_002429735.1|	2.9255	LOC110379408	1193	0.00E+00	95.19	93
Hzea.604 (10)	gi|1199385412|ref|XR_002429411.1|	2.8328	LOC110374164	686	0.00E+00	98.83	95
Hzea.29875 (11)	gi|1199397344|ref|XR_002429683.1|	2.7328	LOC110378838	2247	0.00E+00	99.72	97
Hzea.20149 (12)	gi|1199396138|ref|XR_002429657.1|	2.5172	LOC110378368	507	0.00E+00	95.09	44
Hzea.12374 (13)	gi|1199385671|ref|XR_002429418.1|	2.4004	LOC110374252	1254	0.00E+00	90.38	45
Hzea.28389 (14)	gi|1199408314|ref|XR_002430068.1|	2.3533	LOC110383387	426	4.00E-77	94	24
Hzea.14942 (15)	gi|1200717987|ref|XR_002430221.1|	2.2453	LOC110384842	1264	1.00E-77	84.97	37
Hzea.16728 (16)	gi|1199408095|ref|XR_002430048.1|	2.1086	LOC110383295	576	0.00E+00	90.39	97
Hzea.2954 (17)	gi|1199406738|ref|XR_002429977.1|	2.0091	LOC110382662	364	1.00E-176	98.07	42
Hzea.26641 (18)	gi|1199385671|ref|XR_002429418.1|	1.9793	LOC110374252	1254	0.00E+00	90.72	39
Hzea.21239 (19)	gi|1199406154|ref|XR_002429966.1|	1.9431	LOC110382424	1833	0.00E+00	92.89	94
Hzea.20530 (20)	gi|1199384114|ref|XR_002429384.1|	1.8871	LOC110373699	1934	0.00E+00	98.96	96
Hzea.2658 (21)	gi|1199401627|ref|XR_002429835.1|	1.8262	LOC110380550	1351	0.00E+00	91.96	54
Hzea.7449 (22)	gi|1199399535|ref|XR_002429761.1|	1.8078	LOC110379665	1839	0.00E+00	98.36	100
Hzea.24315 (23)	gi|1199406738|ref|XR_002429977.1|	1.6599	LOC110382662	364	2.00E-176	98.08	27
Hzea.5298 (24)	gi|1199398266|ref|XR_002429709.1|	1.6476	LOC110379203	967	0.00E+00	94.94	77
Hzea.10028 (25)	gi|1199386863|ref|XR_002429444.1|	1.644	LOC110374708	522	0.00E+00	94.33	92
Hzea.12850 (26)	gi|1199398466|ref|XR_002429719.1|	1.5387	LOC110379254	3058	0.00E+00	93.57	41
Hzea.31031 (27)	gi|1199369172|ref|XR_002429735.1|	1.4668	LOC110379408	1193	0.00E+00	94.47	100
Hzea.29569 (28)	gi|1199389586|ref|XR_002429504.1|	1.4428	LOC110375772	794	0.00E+00	95.63	30
Hzea.2296 (29)	gi|1199381581|ref|XR_002429343.1|	1.3909	LOC110372708	358	8.00E-121	98.05	23
Hzea.17312 (30)	gi|1496283280|ref|XR_003400832.1|	1.3259	LOC113496559	3955	2.00E-45	92.75	4
Hzea.18336 (31)	gi|1199390745|ref|XR_002429524.1|	1.325	LOC110376247	1010	0.00E+00	96.84	99
Hzea.9709 (32)	gi|1199378738|ref|XR_002429307.1|	1.2877	LOC110371734	345	5.00E-152	95.4	88

Table 1. continued

Hzea.22246 (33)	gi|1199400576|ref|XR_002429791.1|	1.2241	LOC110380105	446	2.00E-173	91.78	78
Hzea.13913 (34)	gi|1199383324|ref|XR_002429366.1|	0.995	LOC110373344	402	0.00E+00	95.77	27
Hzea.15201 (35)	gi|1199372467|ref|XR_002430136.1|	0.9846	LOC110384418	353	2.00E-113	93.1	54
Hzea.30500 (36)	gi|1199366745|ref|XR_002429227.1|	0.9604	LOC110370099	1290	0.00E+00	96.73	76
Hzea.16487 (37)	gi|1200727923|ref|XR_001139805.3|	0.9318	LOC105841990	1689	2.00E-167	83.53	45
Hzea.31235 (38)	gi|1199370720|ref|XR_002430110.1|	0.9309	LOC110383825	745	2.00E-98	82.23	65
Hzea.19358 (39)	gi|1199383525|ref|XR_002429368.1|	0.8859	LOC110373445	6548	0.00E+00	88.73	73
Hzea.18335 (40)	gi|1199390744|ref|XR_002429523.1|	0.8859	LOC110376246	1288	0.00E+00	96.9	76
Hzea.3960 (41)	gi|1199406106|ref|XR_002429964.1|	0.8819	LOC110382392	2574	0.00E+00	95.2	90
Hzea.30284 (42)	gi|1199384299|ref|XR_002429388.1|	0.8088	LOC110373788	528	0.00E+00	99.05	37
Hzea.26682 (43)	gi|1199389560|ref|XR_002429501.1|	0.7866	LOC110375761	449	0.00E+00	98	78
Hzea.5967 (44)	gi|1199398055|ref|XR_002429700.1|	0.7861	LOC110379115	3649	0.00E+00	97.48	55
Hzea.32166 (45)	gi|1199395541|ref|XR_002429643.1|	0.7447	LOC110378136	418	0.00E+00	97.42	88
Hzea.9647 (46)	gi|1199380627|ref|XR_002429335.1|	0.7022	LOC110372433	913	0.00E+00	92.94	75
Hzea.28852 (47)	gi|1199403501|ref|XR_002429876.1|	0.6566	LOC110381286	544	2.00E-143	96.81	59
Hzea.11610 (48)	gi|1199380589|ref|XR_002429333.1|	0.6451	LOC110372421	1125	0.00E+00	84.05	21
Hzea.20367 (49)	gi|1199398907|ref|XR_002429736.1|	0.6257	LOC110379414	297	2.00E-146	98.99	77
Hzea.646 (50)	gi|1199397854|ref|XR_002429694.1|	0.6096	LOC110379030	387	0.00E+00	98.46	83
Hzea.20502 (51)	gi|1199405118|ref|XR_002429943.1|	0.5774	LOC110382011	1835	0.00E+00	90.67	68
Hzea.23711 (52)	gi|1496280423|ref|XR_003400740.1|	0.5217	LOC113495534	6719	6.00E-135	83.86	15
Hzea.22755 (53)	gi|1199407247|ref|XR_002430000.1|	0.4906	LOC110382901	936	0.00E+00	95.05	33
Hzea.24620 (54)	gi|1199382262|ref|XR_002429351.1|	0.4682	LOC110372972	4231	0.00E+00	98.39	100
Hzea.9489 (55)	gi|1199408314|ref|XR_002430068.1|	0.4173	LOC110383387	426	1.00E-87	94.91	19
Hzea.8708 (56)	gi|1199387699|ref|XR_002429464.1|	0.4137	LOC110375033	3217	0.00E+00	96.3	93
Hzea.17033 (57)	gi|1274103780|ref|XR_002696594.1|	0.4135	LOC111347786	1072	4.00E-04	94.74	2
Hzea.8067 (58)	gi|1199389764|ref|XR_002429506.1|	0.3767	LOC110375841	764	0.00E+00	98.17	90

[a]Gene number corresponds to sequence number in fastq files. Number in parentheses corresponds to lncRNA number outlined in Figure 2 x-axis.
[b]Gene annotations are the database accession numbers for each sequence (ncbi).
[c]log2 fold increase indicates the magnitude of increase of expression in the Bt-resistant strain.
[d]Description provides the LOC number for the top ncbi BLASTn match.
[e]Sequence length indicates the number of base-pairs in each lncRNA.
[f]LncRNA fastq sequence BLASTn e-value result.
[g]LncRNA fastq file BLASTn percent identity result.
[h]LncRNA fastq sequence BLASTn query coverage result.

Table S2. Decreased expressed lncRNAs in Bt-resistant bollworms where a fold change could be calculated, including their fastq file ID (Figure 2 numerical ID), gene annotation, magnitude of log2 fold decrease, sequence length, BLASTn (NCBI) description, and fasta sequence BLASTn result details (E-value, % identity, and query coverage). The table was organized from greatest log2 fold decrease to lowest.

Gene ID[a]	Gene Annotation (Accession #)[b]	Log2 Fold Decrease[c]	BLASTn Result Description[d]	Sequence Length (bp)[e]	E-value[f]	% Identity[g]	Query Coverage (%)[h]
Hzea.26537 (59)	gi|1199367711|ref|XR_002429389.1|	4.40755	LOC110373805	659	7.00E-173	92.97	28
Hzea.14205 (60)	gi|1199383740|ref|XR_002429373.1|	3.68166	LOC110373534	4491	3.00E-30	85.62	33
Hzea.3574 (61)	gi|1199406738|ref|XR_002429977.1|	3.46104	LOC110382662	364	2.00E-160	95.33	25
Hzea.17384 (62)	gi|1199408414|ref|XR_002430074.1|	3.05139	LOC110383440	635	3.00E-96	90.55	47
Hzea.20392 (63)	gi|1199373558|ref|XR_002429212.1|	2.83539	LOC110369725	947	0.00E+00	96.23	98
Hzea.13495 (64)	gi|1199369844|ref|XR_002429941.1|	2.48753	LOC110382000	414	2.00E-112	96.12	9
Hzea.10752 (65)	gi|1199370723|ref|XR_002430111.1|	1.68549	LOC110383827	1221	5.00E-175	85.81	42
Hzea.353 (66)	gi|1199378516|ref|XR_002429294.1|	1.54594	LOC110371636	670	8.00E-166	97.69	21
Hzea.31686 (67)	gi|1199391104|ref|XR_002429541.1|	1.48642	LOC110376383	347	1.00E-159	98.18	76
Hzea.28810 (68)	gi|1199371816|ref|XR_002430124.1|	1.43915	LOC110384191	325	4.00E-149	96.62	61
Hzea.29346 (69)	gi|1199374618|ref|XR_002429228.1|	1.34914	LOC110370109	552	0.00E+00	96.9	72
Hzea.4465 (70)	gi|1199374858|ref|XR_002429236.1|	1.21459	LOC110370222	375	7.00E-172	96.55	73
Hzea.9557 (71)	gi|1199393773|ref|XR_002429605.1|	1.20283	LOC110377404	327	2.00E-132	93.13	61
Hzea.22455 (72)	gi|1199398932|ref|XR_002429739.1|	0.982008	LOC110379429	714	0.00E+00	91.97	78
Hzea.23046 (73)	gi|1199371635|ref|XR_002430121.1|	0.937912	LOC110384132	611	0.00E+00	94.03	91
Hzea.12925 (74)	gi|1199404527|ref|XR_002429917.1|	0.902824	LOC110381748	1117	0.00E+00	91.87	94
Hzea.12351 (75)	gi|1199391501|ref|XR_002429554.1|	0.747984	LOC110376565	1450	3.00E-100	78.78	36
Hzea.4323 (76)	gi|1199382408|ref|XR_002429355.1|	0.701294	LOC110373015	1886	4.00E-19	87.38	6
Hzea.9846 (77)	gi|1199408413|ref|XR_002430073.1|	0.669071	LOC110383440	337	5.00E-140	94.38	13
Hzea.7487 (78)	gi|1199391858|ref|XR_002429559.1|	0.635993	LOC110376693	752	0.00E+00	93.96	88
Hzea.4468 (79)	gi|1199374635|ref|XR_002429229.1|	0.575654	LOC110370116	364	3.00E-179	98.35	90
Hzea.26043 (80)	gi|1199369334|ref|XR_002429787.1|	0.561559	LOC110380046	423	0.00E+00	95.52	93
Hzea.26045 (81)	gi|1199369284|ref|XR_002429773.1|	0.449016	LOC110379830	1820	0.00E+00	96.54	83
Hzea.4909 (82)	gi|1199389298|ref|XR_002429490.1|	0.441255	LOC110375656	1081	0.00E+00	97.78	41

[a]Gene number corresponds to sequence number in fastq files. Number in parentheses corresponds to lncRNA number outlined in Figure 2 x-axis.
[b]Gene annotations are the database accession numbers for each sequence (ncbi).
[c]log2 fold decrease indicates the magnitude of decrease of expression in the Bt-resistant strain.
[d]Description provides the LOC number for the top ncbi BLASTn match.
[e]Sequence length indicates the number of base-pairs in each lncRNA.
[f]LncRNA fastq sequence BLASTn e-value result.
[g]LncRNA fastq file BLASTn percent identity result.
[h]LncRNA fastq sequence BLASTn query coverage result.

Table S3. LncRNAs found only in the Bt-resistant or Bt-susceptible bollworms including their fastq file ID (Figure 2 numerical ID), gene annotation, strain of bollworm, sequence length, BLASTn (NCBI) description, and fasta sequence BLASTn result details (E-value, % identity, query coverage). This table was organized by strain (either Bt-susceptible or Bt-resistant).

	Gene ID[a]	Gene Annotation (Accession #)[b]	BLASTn Result Description[c]	Sequence Length (bp)[d]	E-value[e]	% Identity[f]	Query Coverage (%)[g]
Only in Resistant	Hzea.7318 (83)	gi|1199406283|ref|XR_002429971.1	LOC110382476	414	0	97.34	32
	Hzea.7825 (84)	gi|1199406761|ref|XR_002429982.1	LOC110382674	592	0	97.68	99
	Hzea.14864 (85)	gi|1199406977|ref|XR_002429992.1	LOC110382777	864	2.00E-50	99.15	11
	Hzea.3353 (86)	gi|1199383525|ref|XR_002429368.1	LOC110373445	6548	5.00E-51	96.21	16
	Hzea.1819 (87)	gi|1199398970|ref|XR_002429741.1	LOC110379447	377	9.00E-37	90.48	22
	Hzea.14022 (88)	gi|1199395149|ref|XR_002429636.1	LOC110377983	535	0	97.09	32
	Hzea.9756 (89)	gi|1199387541|ref|XR_002429457.1	LOC110374968	346	8.00E-114	94.51	99
	Hzea.27283 (90)	gi|1199367964|ref|XR_002429443.1	LOC110374671	647	5.00E-156	97.85	100
	Hzea.203 (91)	gi|1199378765|ref|XR_002429309.1	LOC110371745	584	3.00E-79	98.84	56
	Hzea.29688 (92)	gi|1199369338|ref|XR_002429789.1	LOC110380067	753	3.00E-57	91.72	81
Only in Susceptible	Hzea.28800 (93)	gi|1496279844|ref|XR_003400705.1	LOC113495312	2497	2.00E-166	89.89	99
	Hzea.4268 (94)	gi|1199385884|ref|XR_002429423.1	LOC110374339	334	6.00E-150	96.91	31
	Hzea.31700 (95)	gi|1199391125|ref|XR_002429544.1	LOC110376393	465	1.00E-73	91.08	61
	Hzea.19053 (96)	gi|1199408314|ref|XR_002430068.1	LOC110383387	426	1.00E-152	95.94	92

[a]Gene number corresponds to sequence number in fastq files. Numbers in parentheses correspond to lncRNA number outlined in Figure 2 x-axis.
[b]Gene annotations are the database accession numbers for each sequence (ncbi).
[c]Description provides the LOC number for the top ncbi BLASTn match.
[d]Sequence length indicates the number of base-pairs in each lncRNA.
[e]LncRNA fastq sequence BLASTn e-value result.
[f]LncRNA fastq file BLASTn percent identity result.
[g]LncRNA fastq sequence BLASTn query coverage result.

Supplementary Table S4. All NCBI BLAST alignments conducted while searching for potential pseudogenes. The top 5 up- and top 5 down-regulated lncRNAs were aligned to the highest log2 fold change (up- and down-regulated) coding-genes for 5 different categories of genes with functions associated with Bt-resistance in *H. zea*. Table depicts from left to right lncRNA direction of expression, fasta ID and accession number, coding-gene direction of expression, coding gene fasta ID and annotation, presence of sequence alignment, E-value, percent identity, and query coverage.

lncRNA direction of expression[a]	**lncRNA ID[b]**	**Coding-gene direction of expression[c]**	**Coding Gene ID[d]**	**Sequence Alignment[e]**	**E-value[f]**	**% Identity[g]**	**Query Coverage (%)[h]**
Up-regulated	Hzea.12022 (*T. ni* uncharacterized LOC113506107)	**Up-regulated**	Hzea.7824 (serine protease snake-like)	No significant alignment			
			Hzea.18477 (AY2 tetraspanin 1)	No significant alignment			
			Hzea.4257 (trypsin 3A1-like)	No significant alignment			
			Hzea.30068 (beta-secretase 1-like)	No significant alignment			
			Hzea.11178 (JP126 mutant cadherin)	No significant alignment			
		Down-regulated	Hzea.15356 (serine protease snake-like)	No significant alignment			
			Hzea.2673 (AY2 tetraspanin 1)	No significant alignment			
			Hzea.17647 (beta-secretase 1-like)	No significant alignment			
			Hzea.15893 (trypsin 5G1-like)	No significant alignment			
			Hzea.20 (XJ-r15 cadherin)	No significant alignment			

Supplementary Table 4. continued							
	Hzea.2506 (*T. ni* uncharacterized LOC113508874)	**Up-regulated**	Hzea.7824 (serine protease snake-like)	No significant alignment			
			Hzea.18477 (AY2 tetraspanin 1)		0.048	93.75	1
			Hzea.4257 (trypsin 3A1-like)	No significant alignment			
			Hzea.30068 (beta-secretase 1-like)	No significant alignment			
			Hzea.11178 (JP126 mutant cadherin)	No significant alignment			
		Down-regulated	Hzea.15356 (serine protease snake-like)	No significant alignment			
			Hzea.2673 (AY2 tetraspanin 1)	No significant alignment			
			Hzea.17647 (beta-secretase 1-like)	No significant alignment			
			Hzea.15893 (trypsin 5G1-like)	No significant alignment			
			Hzea.20 (XJ-r15 cadherin)	No significant alignment			
	Hzea.17450 (*H. armigera* uncharacterized LOC110372550)	**Up-regulated**	Hzea.7824 (serine protease snake-like)	No significant alignment			
			Hzea.18477 (AY2 tetraspanin 1)	No significant alignment			

Supplementary Table 4. continued							
			Hzea.4257 (trypsin 3A1-like)	No significant alignment			
			Hzea.30068 (beta-secretase 1-like)	No significant alignment			
			Hzea.11178 (JP126 mutant cadherin)	No significant alignment			
		Down-regulated	Hzea.15356 (serine protease snake-like)	No significant alignment			
			Hzea.2673 (AY2 tetraspanin 1)	No significant alignment			
			Hzea.17647 (beta-secretase 1-like)	No significant alignment			
			Hzea.15893 (trypsin 5G1-like)	No significant alignment			
			Hzea.20 (XJ-r15 cadherin)	No significant alignment			
	Hzea.13715 (*H. armigera* uncharacterized LOC110380503)	**Up-regulated**	Hzea.7824 (serine protease snake-like)	No significant alignment			
			Hzea.18477 (AY2 tetraspanin 1)	No significant alignment			
			Hzea.4257 (trypsin 3A1-like)	No significant alignment			
			Hzea.30068 (beta-secretase 1-like)	No significant alignment			
			Hzea.11178 (JP126 mutant cadherin)	No significant alignment			

Supplementary Table 4. continued							
		Down-regulated	Hzea.15356 (serine protease snake-like)	No significant alignment			
			Hzea.2673 (AY2 tetraspanin 1)	No significant alignment			
			Hzea.17647 (beta-secretase 1-like)	No significant alignment			
			Hzea.15893 (trypsin 5G1-like)	No significant alignment			
			Hzea.20 (XJ-r15 cadherin)	No significant alignment			
	Hzea.28004 (*H. armigera* uncharacterized LOC110371745)	**Up-regulated**	Hzea.7824 (serine protease snake-like)	No significant alignment			
			Hzea.18477 (AY2 tetraspanin 1)	No significant alignment			
			Hzea.4257 (trypsin 3A1-like)	No significant alignment			
			Hzea.30068 (beta-secretase 1-like)	No significant alignment			
			Hzea.11178 (JP126 mutant cadherin)	No significant alignment			
		Down-regulated	Hzea.15356 (serine protease snake-like)	No significant alignment			
			Hzea.2673 (AY2 tetraspanin 1)	No significant alignment			
			Hzea.17647 (beta-secretase 1-like)	No significant alignment			

Supplementary Table 4. continued							
			Hzea.15893 (trypsin 5G1-like)	No significant alignment			
			Hzea.20 (XJ-r15 cadherin)	No significant alignment			
Down-regulated	Hzea.26537 (*H. armigera* uncharacterized LOC110373805)	**Up-regulated**	Hzea.7824 (serine protease snake-like)	No significant alignment			
			Hzea.18477 (AY2 tetraspanin 1)		0.027	94.12	0
			Hzea.4257 (trypsin 3A1-like)	No significant alignment			
			Hzea.30068 (beta-secretase 1-like)	No significant alignment			
			Hzea.11178 (JP126 mutant cadherin)	No significant alignment			
		Down-regulated	Hzea.15356 (serine protease snake-like)	No significant alignment			
			Hzea.2673 (AY2 tetraspanin 1)	No significant alignment			
			Hzea.17647 (beta-secretase 1-like)	No significant alignment			
			Hzea.15893 (trypsin 5G1-like)	No significant alignment			
			Hzea.20 (XJ-r15 cadherin)	No significant alignment			
	Hzea.14205 (*H. armigera* uncharacterized LOC110373534)	**Up-regulated**	Hzea.7824 (serine protease snake-like)	No significant alignment			

Supplementary Table 4. continued							
			Hzea.18477 (AY2 tetraspanin 1)	No significant alignment			
			Hzea.4257 (trypsin 3A1-like)	No significant alignment			
			Hzea.30068 (beta-secretase 1-like)	No significant alignment			
			Hzea.11178 (JP126 mutant cadherin)	No significant alignment			
		Down-regulated	Hzea.15356 (serine protease snake-like)	No significant alignment			
			Hzea.2673 (AY2 tetraspanin 1)	No significant alignment			
			Hzea.17647 (beta-secretase 1-like)	No significant alignment			
			Hzea.15893 (trypsin 5G1-like)	No significant alignment			
			Hzea.20 (XJ-r15 cadherin)	No significant alignment			
	Hzea.3574 (*H. armigera* uncharacterized LOC110382662)	**Up-regulated**	Hzea.7824 (serine protease snake-like)	No significant alignment			
			Hzea.18477 (AY2 tetraspanin 1)		4.00E-04	85.19	1
			Hzea.4257 (trypsin 3A1-like)		0.018	94.12	1
			Hzea.30068 (beta-secretase 1-like)	No significant alignment			

Supplementary Table 4. continued							
			Hzea.11178 (JP126 mutant cadherin)	No significant alignment			
		Down-regulated	Hzea.15356 (serine protease snake-like)	No significant alignment			
			Hzea.2673 (AY2 tetraspanin 1)		0.002	84.62	3
			Hzea.17647 (beta-secretase 1-like)	No significant alignment			
			Hzea.15893 (trypsin 5G1-like)	No significant alignment			
			Hzea.20 (XJ-r15 cadherin)		0.005	94.44	2
	Hzea.17384 (*H. armigera* uncharacterized LOC110383440)	**Up-regulated**	Hzea.7824 (serine protease snake-like)	No significant alignment			
			Hzea.18477 (AY2 tetraspanin 1)	No significant alignment			
			Hzea.4257 (trypsin 3A1-like)	No significant alignment			
			Hzea.30068 (beta-secretase 1-like)	No significant alignment			
			Hzea.11178 (JP126 mutant cadherin)	No significant alignment			
		Down-regulated	Hzea.15356 (serine protease snake-like)	No significant alignment			
			Hzea.2673 (AY2 tetraspanin 1)	No significant alignment			

Supplementary Table 4. continued							
			Hzea.17647 (beta-secretase 1-like)	No significant alignment			
			Hzea.15893 (trypsin 5G1-like)	No significant alignment			
			Hzea.20 (XJ-r15 cadherin)	No significant alignment			
	Hzea.20392 (*H. armigera* uncharacterized LOC110369725)	**Up-regulated**	Hzea.7824 (serine protease snake-like)	No significant alignment			
			Hzea.18477 (AY2 tetraspanin 1)	No significant alignment			
			Hzea.4257 (trypsin 3A1-like)	No significant alignment			
			Hzea.30068 (beta-secretase 1-like)		0.033	100	1
			Hzea.11178 (JP126 mutant cadherin)	No significant alignment			
		Down-regulated	Hzea.15356 (serine protease snake-like)	No significant alignment			
			Hzea.2673 (AY2 tetraspanin 1)	No significant alignment			
			Hzea.17647 (beta-secretase 1-like)	No significant alignment			
			Hzea.15893 (trypsin 5G1-like)	No significant alignment			
			Hzea.20 (XJ-r15 cadherin)		**0**	**99.07**	**81**

[a]Indicates the direction of expression for the lncRNA. [b]Indicates the fasta ID and the NCBI accession number for the lncRNA. [c]Indicates the coding-gene fasta ID and annotation ID. [d]Indicates whether the lncRNA and coding-gene had ncbi BLAST alignments [e]NCBI assigned E-value for lncRNA to coding-gene alignment. [f]NCBI assigned percent identity for lncRNA to coding-gene alignment. [g]NCBI assigned query coverage for lncRNA to coding-gene alignments.

FIGURES

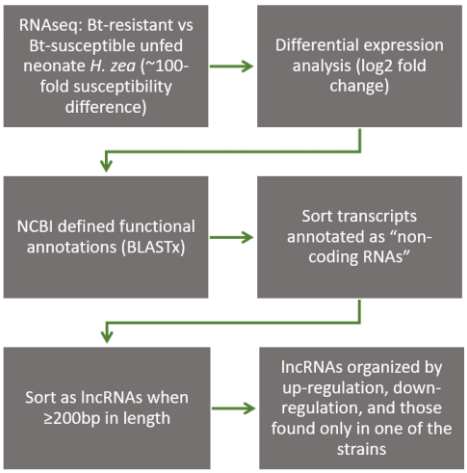


Figure 1. Workflow for identifying statistically significant, differentially expressed lncRNAs between a Bt-resistant and susceptible strain of unfed, neonate *H. zea*.

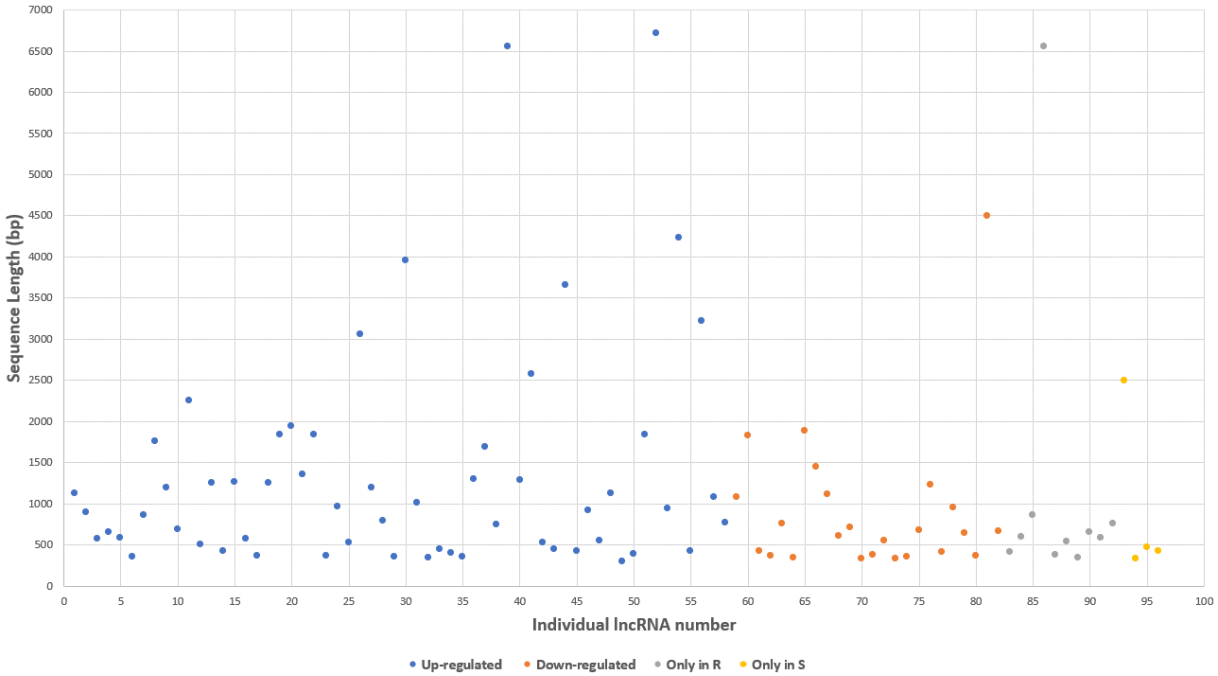


Figure 2. Sequence length in base-pairs (bp) for up-regulated and down regulated lncRNAs where a fold change could be calculated and for lncRNAs found only in the resistant (R) and susceptible (S) strains. lncRNAs were organized left to right from highest to lowest log2 fold change for each direction. LncRNAs in only the R or S categories were randomly organized. The identification numbers on the x-axis for each transcript are also shown in Tables 1-3 along with more details about each transcript.

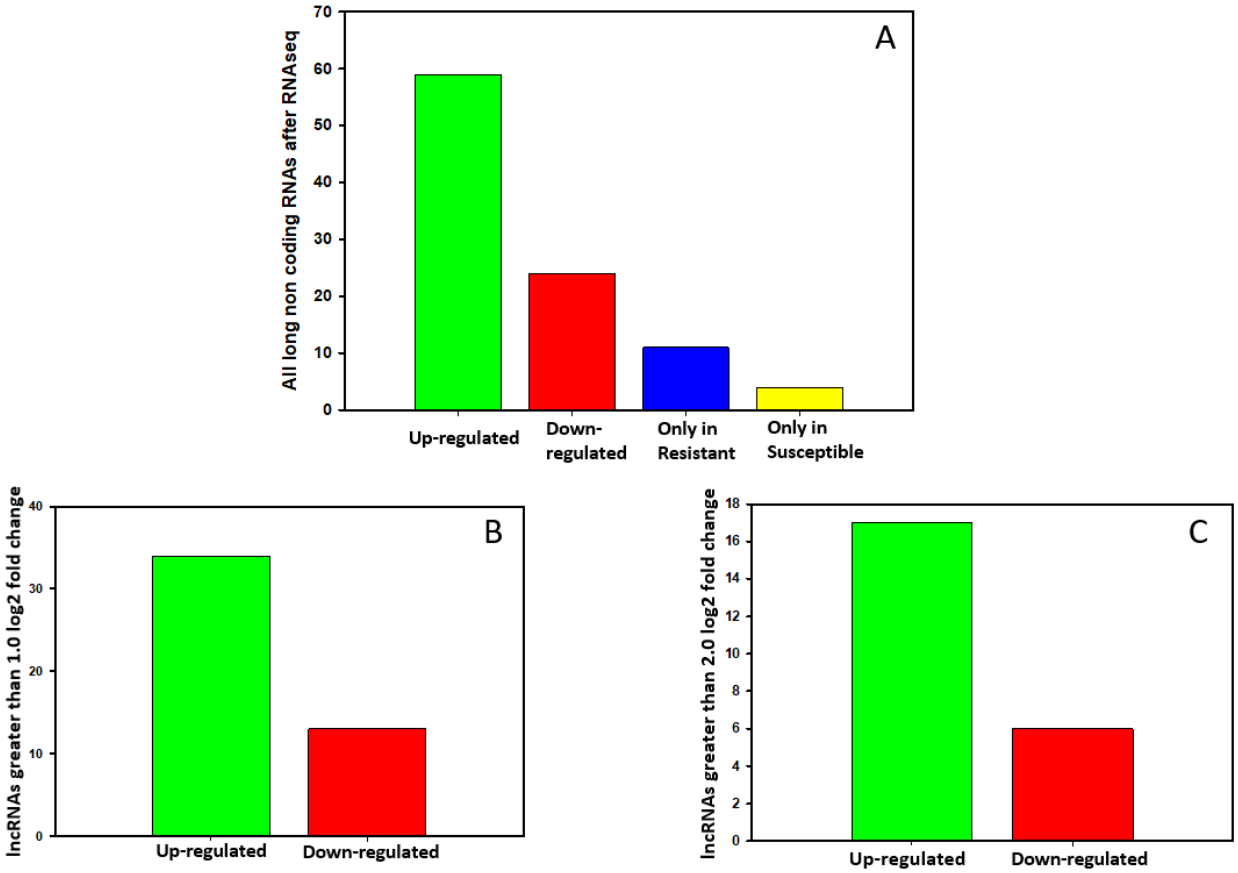


Figure 3. Number of significantly differentiated lncRNA transcripts (up and down regulated) where a fold change could be calculated and those unique to either the Bt-resistant (R) or Bt-susceptible (S) larval *H. zea*. A, All statistically significant differentially expressed lncRNAs between the R and S strains. B, The number of statistically significant differentially expressed lncRNAs with a greater than 1.0 log2 fold change. C, The number of statistically significant differentially expressed lncRNAs with a greater than 2.0 log2 fold change. No log2 fold change could be calculated for lncRNAs found only in one strain.

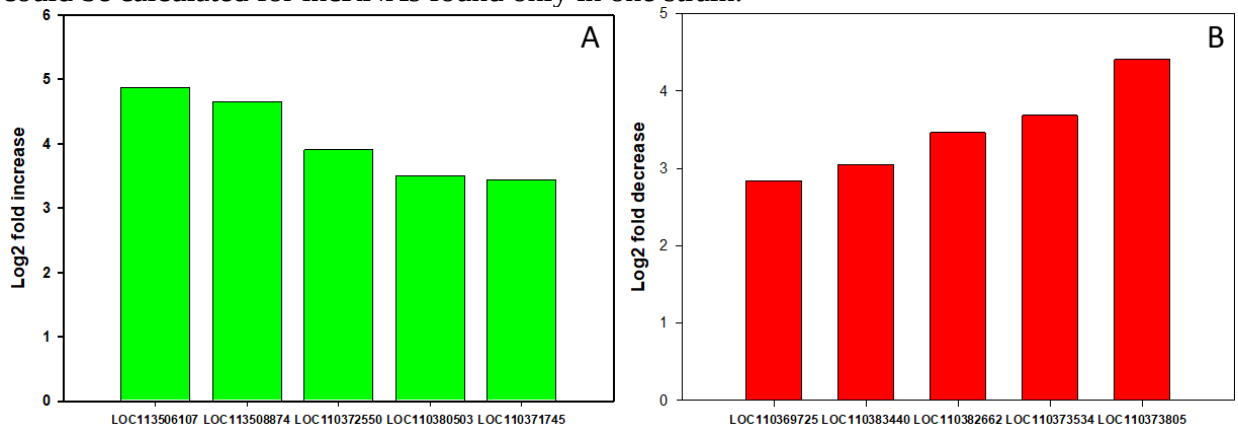


Figure 4. Top 5 differentially expressed lncRNAs with the highest log2 fold changes. A, The top 5 up-regulated lncRNAs (increased expression) in the Bt-resistant *H. zea*. B, Top 5 down-regulated lncRNAs (decreased expression) in the Bt-resistant strain. The X-axis depicts the LOC ID number (NCBI defined) with more information describing these transcripts in Tables 1-2.

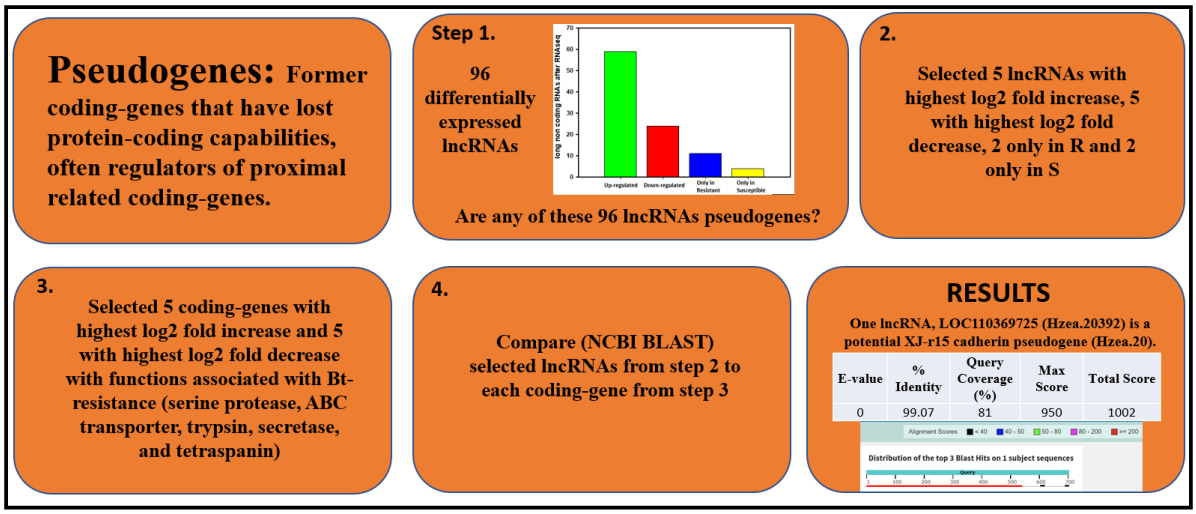


Figure 5. Workflow to identify statistically differentiated lncRNAs as putative pseudogenes.

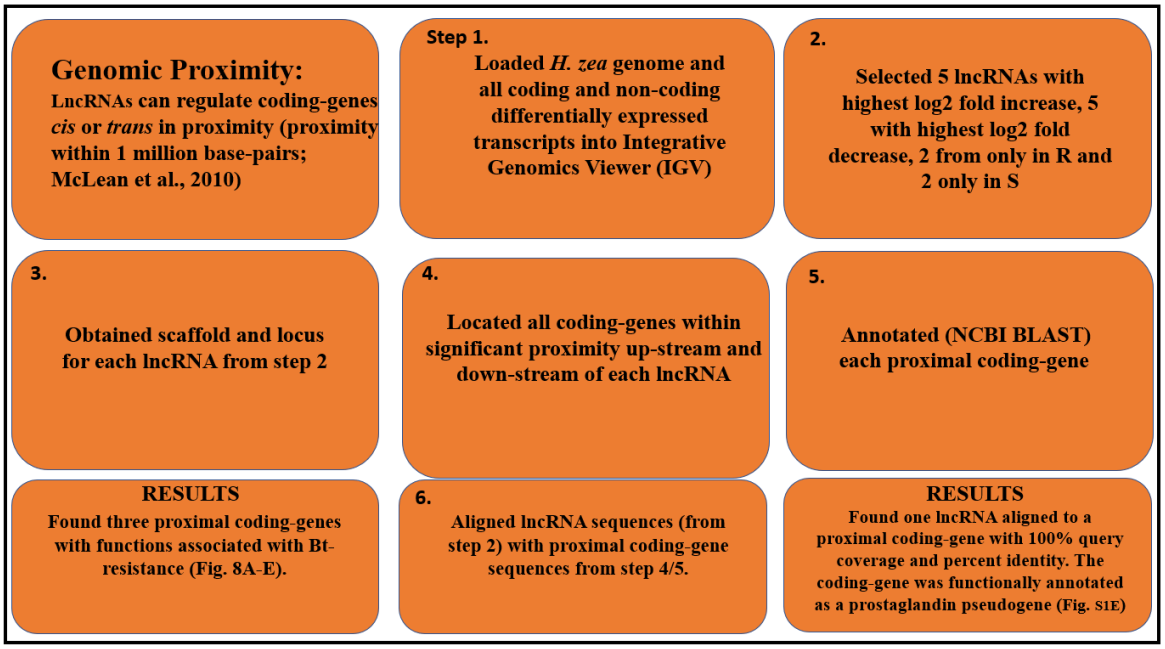


Figure 6. Workflow for identifying statistically differentiated lncRNAs coding genes *in toto* and those with functions known to have a role in Bt-resistance that are proximal to statistically differentiated lncRNAs. Proximity measurments were limited by the size of the scaffolds, even though proximity is defined as 1 million base pairs *cis* and *trans* from the lncRNA. For each proximal coding gene and lncRNA, a BLASTn alignment was also conducted to assess potential pseudogenes.

Figure 7A-E. Genomic scaffold for lncRNAs and identification of proximal protein coding genes. The scaffold at the top of each A-E depicts the range of the scaffold in kilobases (kb). The bars in blue indicate sequences present on each scaffold, with gene ID numbers below each. The red stars indicate a lncRNA, the black stars indicate a protein-coding gene. The scaffold ID number is placed directly below the legend on the left side. The table below the scaffold includes the following information about the lncRNA and coding genes found in the scaffold from left to right: the lncRNA ID number and annotation, lncRNA loci coordinates, gene ID number of proximal coding gene, coding gene loci coordinates, coding gene annotation (NCBI defined), coding gene log2 fold change, and BLASTn alignment results (% identity, E-value, and query coverage) comparing the lncRNA and the protein coding gene. Each subfigure depicts the following: A, lncRNA LOC113506107 proximal to a CYP coding gene; B, lncRNA 110369725 proximal to a ABC transporter coding gene; C, lncRNA LOC110382674 proximal to a serine protease coding gene; D, lncRNA LOC110373534 with no significant proximal coding genes; and E, lncRNA LOC110383387 proximal to non Bt-associated coding genes. All other proximity analyses can be found in Supplemental Figure S1-4.

=lncRNA

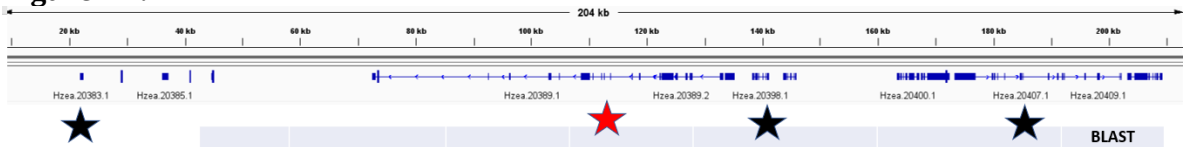
=coding-gene

Scaffold: KZ117479.1

lncRNA ID and annotation[A]	lncRNA locus[B]	Proximal coding-gene ID[C]	Proximal transcript locus[D]	Proximal coding-gene annotation[E]	Proximal coding-gene log2 fold change[F]	BLAST alignment (% identity, E-value, Query coverage)[G]
Hzea.12022 (LOC113506107)	899-12026	Hzea.12028	19894-27603	*Helicoverpa armigera* voltage-gated sodium channel alpha subunit gene, partial cds; and cytochrome p450 6B7 (CYP6B7), cytochrome p450 6B6 (CYP6B6), and cytochrome p450 6B2 (CYP6B2) genes, complete cds	0.423492 log2 fold increase	No significant alignment

[A]lncRNA ID indicates fasta file ID number, annotation indicates NCBI defined annotation.
[B]locus number indicates base-pair coordinates for lncRNA within scaffold.
[C]coding-gene ID indicates fasta file ID number.
[D]locus number indicates base-pair coordinates for coding-gene within scaffold.
[E]coding-gene annotation indicates NCBI defined annotation.
[F]Indicates magnitude of log2 fold change and direction of expression.
[G]NCBI BLAST alignment of lncRNA to proximal coding-gene in following order: % ID, E-value, and % Query coverage.

Figure 7A.



=lncRNA

=coding-gene

Scaffold: KZ118217.1

lncRNA ID and annotation[A]	lncRNA locus[B]	Proximal coding-gene ID[C]	Proximal transcript locus[D]	Proximal coding-gene annotation[E]	Proximal coding-gene log2 fold change[F]	BLAST alignment (% identity, E-value, Query coverage)[G]
Hzea.20392 (LOC110369725)	72493-133287	Hzea.20383	21821-22446	*Heliothis subflexa* clone 89F08 ABC transporter family C protein ABCC3 (ABCC3) gene, partial cds; ABC transporter family C protein ABCC2 (ABCC2) and SMC domain protein (nachbar) genes, complete cds; and ABC transporter family C protein ABCC1 (ABCC1) gene, partial cds	N/A	No significant alignment
		Hzea.20398	138240-141169	*Helicoverpa armigera* uncharacterized LOC110377300 (LOC110377300), mRNA	2.44635 log2 fold decrease	No significant alignment
		Hzea.20407	173316-202347	*Helicoverpa armigera* tyrosine-protein kinase Btk29A (LOC110377297), transcript variant X4, mRNA	0.681871 log2 fold increase	No significant alignment

[A]lncRNA ID indicates fasta file ID number, annotation indicates NCBI defined annotation.
[B]locus number indicates base-pair coordinates for lncRNA within scaffold.
[C]coding-gene ID indicates fasta file ID number.
[D]locus number indicates base-pair coordinates for coding-gene within scaffold.
[E]coding-gene annotation indicates NCBI defined annotation.
[F]Indicates magnitude of log2 fold change and direction of expression.
[G]NCBI BLAST alignment of lncRNA to proximal coding-gene in following order: % ID, E-value, and % Query coverage.

Figure 7B.

=lncRNA

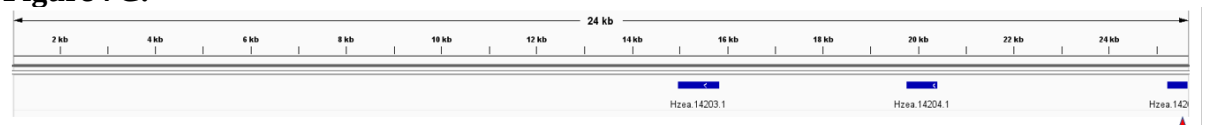
=coding-gene

Scaffold: KZ117066.1

lncRNA ID and annotation[A]	lncRNA locus[B]	Proximal coding-gene ID[C]	Proximal transcript locus[D]	Proximal coding-gene annotation[E]	Proximal coding-gene log2 fold change[F]	BLAST alignment (% identity, E-value, Query coverage)[G]
Hzea.7825 (LOC110382674)	61267-62283	Hzea.7824	50730-60621	*Helicoverpa armigera* serine protease snake-like (LOC110382673), transcript variant X1, mRNA	3.10712 log2 fold increase	No significant alignment

[A]lncRNA ID indicates fasta file ID number, annotation indicates NCBI defined annotation.
[B]locus number indicates base-pair coordinates for lncRNA within scaffold.
[C]coding-gene ID indicates fasta file ID number.
[D]locus number indicates base-pair coordinates for coding-gene within scaffold.
[E]coding-gene annotation indicates NCBI defined annotation.
[F]Indicates magnitude of log2 fold change and direction of expression.
[G]NCBI BLAST alignment of lncRNA to proximal coding-gene in following order: % ID, E-value, and % Query coverage.

Figure 7C.



=lncRNA

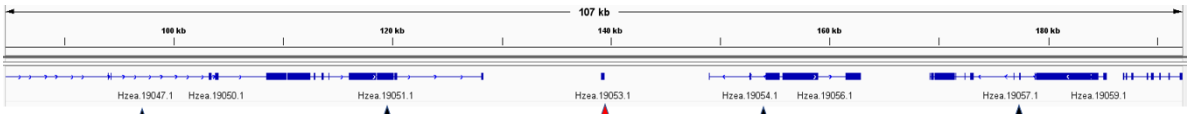
=coding-gene

Scaffold: KZ117729.1

lncRNA ID and annotation[A]	lncRNA locus[B]	Proximal coding-gene ID[C]	Proximal transcript locus[D]	Proximal coding-gene annotation[E]	Proximal coding-gene log2 fold change[F]	BLAST alignment (% identity, E-value, Query coverage)[G]
Hzea.14205 (LOC110373534)	25239-25652	N/A	N/A	N/A	N/A	N/A

[A]lncRNA ID indicates fasta file ID number, annotation indicates NCBI defined annotation.
[B]locus number indicates base-pair coordinates for lncRNA within scaffold.
[C]coding-gene ID indicates fasta file ID number.
[D]locus number indicates base-pair coordinates for coding-gene within scaffold.
[E]coding-gene annotation indicates NCBI defined annotation.
[F]Indicates magnitude of log2 fold change and direction of expression.
[G]NCBI BLAST alignment of lncRNA to proximal coding-gene in following order: % ID, E-value, and % Query coverage.

Figure 7D.



=lncRNA

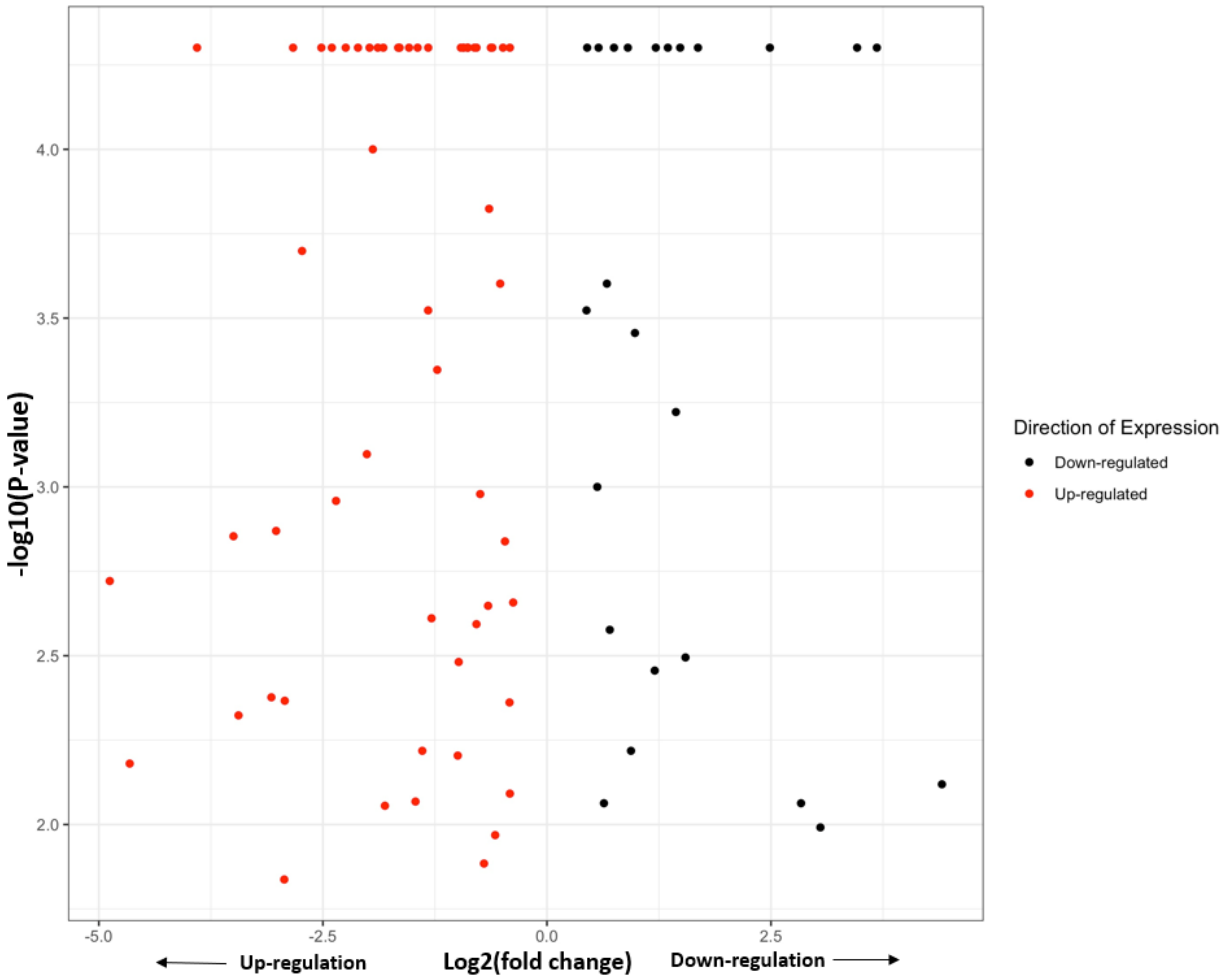
=coding-gene

Scaffold: KZ118171.1

lncRNA ID and annotation[A]	lncRNA locus[B]	Proximal coding-gene ID[C]	Proximal transcript locus[D]	Proximal coding-gene annotation[E]	Proximal coding-gene log2 fold change[F]	BLAST alignment (% identity, E-value, Query coverage)[G]
Hzea.19053 (LOC110383387)	139116-139479	Hzea.19047	29864-110333	*Helicoverpa armigera* uncharacterized LOC110377082 (LOC110377082), transcript variant X2, mRNA	0.872421 log2 fold decrease	No significant alignment
		Hzea.19051	110427-128368	*Helicoverpa armigera* uncharacterized LOC110377082 (LOC110377082), transcript variant X1, mRNA	1.14398 log2 fold decrease	No significant alignment
		Hzea.19054	149019-162910	*Helicoverpa armigera* retinol dehydrogenase 14 (LOC110377063), transcript variant X1, mRNA	0.581699 log2 fold increase	No significant alignment
		Hzea.19057	169217-185466	*Helicoverpa armigera* uncharacterized LOC110377071 (LOC110377071), mRNA	0.457049 log2 fold increase	No significant alignment

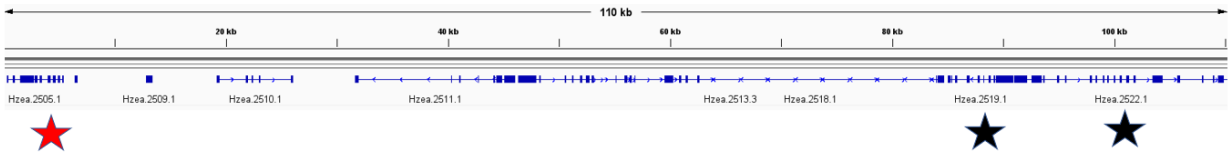
[A]lncRNA ID indicates fasta file ID number, annotation indicates NCBI defined annotation.
[B]locus number indicates base-pair coordinates for lncRNA within scaffold.
[C]coding-gene ID indicates fasta file ID number.
[D]locus number indicates base-pair coordinates for coding-gene within scaffold.
[E]coding-gene annotation indicates NCBI defined annotation.
[F]Indicates magnitude of log2 fold change and direction of expression.
[G]NCBI BLAST alignment of lncRNA to proximal coding-gene in following order: % ID, E-value, and % Query coverage.

Figure 7E.



Supplemental Figure S1. Log2 fold change of differentially expressed up- (red dots) or down- (black dots) regulated lncRNAs. The X-axis depicts the magnitude of log2 fold change in Bt-resistant *H. zea* where values on the left of 0.0 indicate the fold up-regulation and on the right of 0.0 the log2 fold down-regulation relative to the Bt-susceptible *H. zea*. The Y-axis is the p-values for statistical significance.

Supplementary Figure S2 (A-D). Genomic scaffold for up-regulated lncRNAs and identification of proximal protein coding-genes. The scaffold at the top of each A-D, depicts the range of the scaffold in kilobases (kb). The bars in blue indicate sequences present on each scaffold, with gene ID numbers below each. The red stars indicate a lncRNA, the black stars indicate a protein-coding gene. The scaffold ID number is placed directly below the legend on the left side. The table below the scaffold includes the following information about the lncRNA and coding-genes found in the scaffold from left to right: the lncRNA ID number and annotation, lncRNA loci coordinates, gene ID number of proximal coding-gene, coding-gene loci coordinates, coding-gene annotation (NCBI defined), coding-gene log2 fold change, and BLASTn alignment results (% identity, E-value, and query coverage) comparing the lncRNA and the protein coding-gene.



=lncRNA

=coding-gene

Scaffold: KZ116624.1

lncRNA ID and annotation[A]	lncRNA locus[B]	Proximal coding-gene ID[C]	Proximal transcript locus[D]	Proximal coding-gene annotation[E]	Proximal coding-gene log2 fold change[F]	BLAST alignment (% identity, E-value, Query coverage)[G]
Hzea.2506 (LOC113508874)	294-5519	Hzea.2519	85052-90956	*Helicoverpa armigera* mitochondrial import inner membrane translocase subunit TIM44 (LOC110369848), transcript variant X2, mRNA	0.334603 log2 fold increase	No significant alignment
		Hzea.2522	90996-153962	*Helicoverpa armigera* echinoderm microtubule-associated protein-like 2 (LOC110369847), transcript variant X1, mRNA	1.42313 log2 fold increase	No significant alignment

[A]lncRNA ID indicates fasta file ID number, annotation indicates NCBI defined annotation.
[B]locus number indicates base-pair coordinates for lncRNA within scaffold.
[C]coding-gene ID indicates fasta file ID number.
[D]locus number indicates base-pair coordinates for coding-gene within scaffold.
[E]coding-gene annotation indicates NCBI defined annotation.
[F]Indicates magnitude of log2 fold change and direction of expression.
[G]NCBI BLAST alignment of lncRNA to proximal coding-gene in following order: % ID, E-value, and % Query coverage.

Supplementary Figure S2A.

=lncRNA

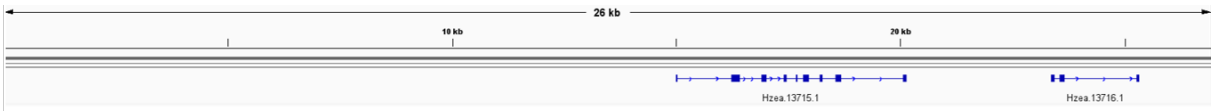
=coding-gene

Scaffold: KZ118099.1

lncRNA ID and annotation[A]	lncRNA locus[B]	Proximal coding-gene ID[C]	Proximal transcript locus[D]	Proximal coding-gene annotation[E]	Proximal coding-gene log2 fold change[F]	BLAST alignment (% identity, E-value, Query coverage)[G]
Hzea.17450 (LOC110372550)	185452-187367	Hzea.17453	200236-202568	*Helicoverpa armigera* uncharacterized LOC110372536 (LOC110372536), transcript variant X4, mRNA	0.618834 log2 fold increase	No significant alignment
		Hzea.17436	91939-134182	*Helicoverpa armigera* CAP-Gly domain-containing linker protein 1 (LOC110372605), transcript variant X15, mRNA	0.844544 log2 fold increase	No significant alignment
		Hzea.17455	207667-223349	*Helicoverpa armigera* uncharacterized LOC110372536 (LOC110372536), transcript variant X3, mRNA	0.400524 log2 fold decrease	No significant alignment
		Hzea.17459	236185-257407	*Helicoverpa armigera* receptor-type tyrosine-protein phosphatase N2 (LOC110372540), transcript variant X3, mRNA	0.58871 log2 fold increase	No significant alignment

[A]lncRNA ID indicates fasta file ID number, annotation indicates NCBI defined annotation.
[B]locus number indicates base-pair coordinates for lncRNA within scaffold.
[C]coding-gene ID indicates fasta file ID number.
[D]locus number indicates base-pair coordinates for coding-gene within scaffold.
[E]coding-gene annotation indicates NCBI defined annotation.
[F]Indicates magnitude of log2 fold change and direction of expression.
[G]NCBI BLAST alignment of lncRNA to proximal coding-gene in following order: % ID, E-value, and % Query coverage.

Supplementary Figure S2B.



=lncRNA

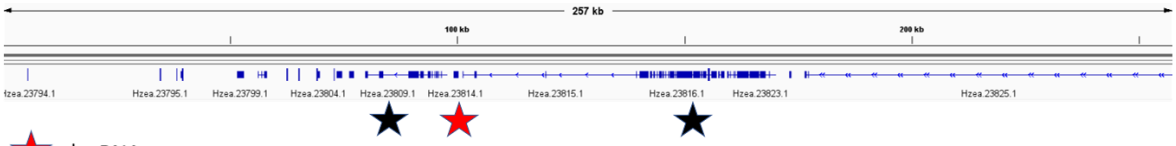
=coding-gene

Scaffold: KZ117672.1

lncRNA ID and annotation[A]	lncRNA locus[B]	Proximal coding-gene ID[C]	Proximal transcript locus[D]	Proximal coding-gene annotation[E]	Proximal coding-gene log2 fold change[F]	BLAST alignment (% identity, E-value, Query coverage)[G]
Hzea.13715 (LOC110380503)	15001-20158	N/A	N/A	N/A	N/A	N/A

[A]lncRNA ID indicates fasta file ID number, annotation indicates NCBI defined annotation.
[B]locus number indicates base-pair coordinates for lncRNA within scaffold.
[C]coding-gene ID indicates fasta file ID number.
[D]locus number indicates base-pair coordinates for coding-gene within scaffold.
[E]coding-gene annotation indicates NCBI defined annotation.
[F]Indicates magnitude of log2 fold change and direction of expression.
[G]NCBI BLAST alignment of lncRNA to proximal coding-gene in following order: % ID, E-value, and % Query coverage.

Supplementary Figure S2C.



=lncRNA

=coding-gene

Scaffold: KZ118374.1

lncRNA ID and annotation[A]	lncRNA locus[B]	Proximal coding-gene ID[C]	Proximal transcript locus[D]	Proximal coding-gene annotation[E]	Proximal coding-gene log2 fold change[F]	BLAST alignment (% identity, E-value, Query coverage)[g]
Hzea.23814 (LOC110372708)	99334-100312	Hzea.23809	79865-90017	*Helicoverpa armigera* prostaglandin reductase 1 pseudogene (LOC110372096), misc_RNA	0.944219 log2 fold increase	**100, 0, 100%**
		Hzea.23816	142574-154397	*Helicoverpa armigera* small subunit processome component 20 homolog (LOC110372067), mRNA	1.5187 log2 fold decrease	No significant alignment

[A]lncRNA ID indicates fasta file ID number, annotation indicates NCBI defined annotation.
[B]locus number indicates base-pair coordinates for lncRNA within scaffold.
[C]coding-gene ID indicates fasta file ID number.
[D]locus number indicates base-pair coordinates for coding-gene within scaffold.
[E]coding-gene annotation indicates NCBI defined annotation.
[F]Indicates magnitude of log2 fold change and direction of expression.
[G]NCBI BLAST alignment of lncRNA to proximal coding-gene in following order: % ID, E-value, and % Query coverage.

Supplementary Figure S2D.

Supplementary Figure S3(A-C). Genomic scaffold for down-regulated lncRNAs and identification of proximal protein coding-genes. The scaffold in A-C depicts the range of the scaffold in kilobases (kb). The bars in blue indicate sequences present on each scaffold, with gene ID numbers below each. The red stars indicate a lncRNA, the black stars indicate a protein-coding gene. The scaffold ID number is placed directly below the legend on the left side. The table below the scaffold includes the following information about the lncRNA and coding-genes found in the scaffold from left to right: the lncRNA ID number and annotation, lncRNA loci coordinates, gene ID number of proximal coding-gene, coding-gene loci coordinates, coding-gene annotation (NCBI defined), coding-gene log2 fold change, and BLASTn alignment results (% identity, E-value, and query coverage) comparing the lncRNA and the protein coding-gene.

=lncRNA

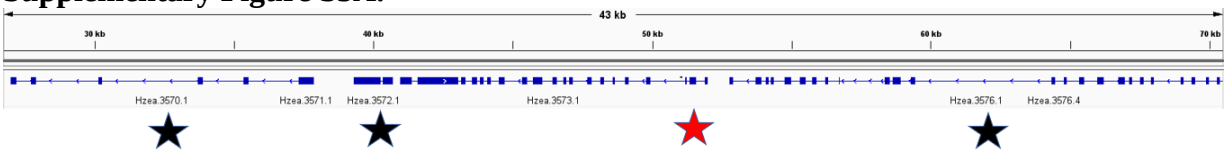
=coding-gene

Scaffold: KZ118515.1

lncRNA ID and annotation[A]	lncRNA locus[B]	Proximal coding-gene ID[C]	Proximal transcript locus[D]	Proximal coding-gene annotation[E]	Proximal coding-gene log2 fold change[F]	BLAST alignment (% identity, E-value, Query coverage)[G]
Hzea.26537 (LOC110373805)	503539-507800	Hzea.26523	421869-460342	*Helicoverpa armigera* unconventional myosin-XV (LOC110372091), mRNA	0.466472 log2 fold increase	No significant alignment
		Hzea.26531	495604-500397	*Helicoverpa armigera* developmentally-regulated GTP-binding protein 2 (LOC110372053), mRNA	0.852664 log2 fold decrease	No significant alignment

[A]lncRNA ID indicates fasta file ID number, annotation indicates NCBI defined annotation.
[B]locus number indicates base-pair coordinates for lncRNA within scaffold.
[C]coding-gene ID indicates fasta file ID number.
[D]locus number indicates base-pair coordinates for coding-gene within scaffold.
[E]coding-gene annotation indicates NCBI defined annotation.
[F]Indicates magnitude of log2 fold change and direction of expression.
[G]NCBI BLAST alignment of lncRNA to proximal coding-gene in following order: % ID, E-value, and % Query coverage.

Supplementary Figure S3A.



=lncRNA

=coding-gene

Scaffold: KZ116697.1

lncRNA ID and annotation[A]	lncRNA locus[B]	Proximal coding-gene ID[C]	Proximal transcript locus[D]	Proximal coding-gene annotation[E]	Proximal coding-gene log2 fold change[F]	BLAST alignment (% identity, E-value, Query coverage)[G]
Hzea.3574 (LOC110382662)	40967-52014	Hzea.3570	26998-37884	*Helicoverpa armigera* uncharacterized LOC110381452 (LOC110381452), mRNA	0.573782 log2 fold decrease	No significant alignment
		Hzea.3572	39328-40721	*Helicoverpa armigera* peptidyl-prolyl cis-trans isomerase NIMA-interacting 4 (LOC110370269), transcript variant X1, mRNA	1.58316 log2 fold decrease	No significant alignment
		Hzea.3576	52787-73597	*Helicoverpa armigera* mediator of RNA polymerase II transcription subunit 12-like (LOC110374056), transcript variant X3, mRNA	0.593304 log2 fold increase	No significant alignment

[A]lncRNA ID indicates fasta file ID number, annotation indicates NCBI defined annotation.
[B]locus number indicates base-pair coordinates for lncRNA within scaffold.
[C]coding-gene ID indicates fasta file ID number.
[D]locus number indicates base-pair coordinates for coding-gene within scaffold.
[E]coding-gene annotation indicates NCBI defined annotation.
[F]Indicates magnitude of log2 fold change and direction of expression.
[G]NCBI BLAST alignment of lncRNA to proximal coding-gene in following order: % ID, E-value, and % Query coverage.

Supplementary Figure S3B.

=lncRNA

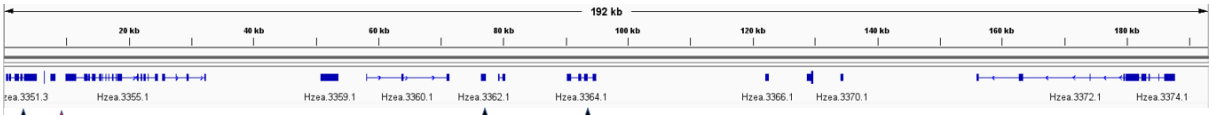
=coding-gene

Scaffold: KZ118094.1

lncRNA ID and annotation[A]	lncRNA locus[B]	Proximal coding-gene ID[C]	Proximal transcript locus[D]	Proximal coding-gene annotation[E]	Proximal coding-gene log2 fold change[F]	BLAST alignment (% identity, E-value, Query coverage)[G]
Hzea.17384 (LOC110383440)	214203-214786	Hzea.17373	444-3087	*Spodoptera litura* 60S ribosomal protein L18a (LOC111357503), mRNA	0.873101 log2 fold increase	No significant alignment

[A]lncRNA ID indicates fasta file ID number, annotation indicates NCBI defined annotation.
[B]locus number indicates base-pair coordinates for lncRNA within scaffold.
[C]coding-gene ID indicates fasta file ID number.
[D]locus number indicates base-pair coordinates for coding-gene within scaffold.
[E]coding-gene annotation indicates NCBI defined annotation.
[F]Indicates magnitude of log2 fold change and direction of expression.
[G]NCBI BLAST alignment of lncRNA to proximal coding-gene in following order: % ID, E-value, and % Query coverage.

Supplementary Figure S3C.



=lncRNA

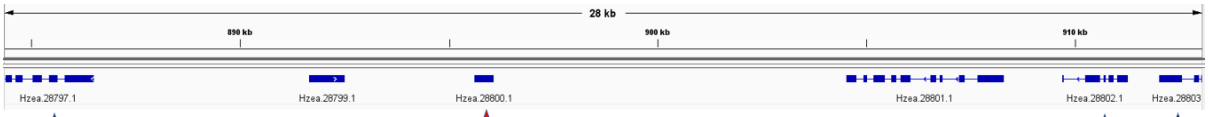
=coding-gene

Scaffold: KZ116682.1

lncRNA ID and annotation[A]	lncRNA locus[B]	Proximal coding-gene ID[C]	Proximal transcript locus[D]	Proximal coding-gene annotation[E]	Proximal coding-gene log2 fold change[F]	BLAST alignment (% identity, E-value, Query coverage)[G]
Hzea.3353 (LOC110373445)	7483-8292	Hzea.3351	490-5342	*Spodoptera litura* protein groucho-like (LOC111355272), partial mRNA	0.676627 log2 fold decrease	No significant alignment
		Hzea.3362	76493-77379	Unknown Gene	1.58121 log2 fold increase	No significant alignment
		Hzea.3364	90221-94993	*Helicoverpa armigera* uncharacterized LOC110370405 (LOC110370405), mRNA	0.496072 log2 fold increase	100, 0.02, 1%

[A]lncRNA ID indicates fasta file ID number, annotation indicates NCBI defined annotation.
[B]locus number indicates base-pair coordinates for lncRNA within scaffold.
[C]coding-gene ID indicates fasta file ID number.
[D]locus number indicates base-pair coordinates for coding-gene within scaffold.
[E]coding-gene annotation indicates NCBI defined annotation.
[F]Indicates magnitude of log2 fold change and direction of expression.
[G]NCBI BLAST alignment of lncRNA to proximal coding-gene in following order: % ID, E-value, and % Query coverage.

Supplementary Figure S4. Genomic scaffold for lncRNAs found only in the R strain and identification of proximal protein coding-genes. The scaffold at the top depicts the range of the scaffold in kilobases (kb). The bars in blue indicate sequences present on each scaffold, with gene ID numbers below each. The red stars indicate a lncRNA, the black stars indicate a protein-coding gene. The scaffold ID number is placed directly below the legend on the left side. The table below the scaffold includes the following information about the lncRNA and coding-genes found in the scaffold from left to right: the lncRNA ID number and annotation, lncRNA loci coordinates, gene ID number of proximal coding-gene, coding-gene loci coordinates, coding-gene annotation (NCBI defined), coding-gene log2 fold change, and BLASTn alignment results (% identity, E-value, and query coverage) comparing the lncRNA and the protein coding-gene.



=lncRNA

=coding-gene

Scaffold: KZ118643.1

lncRNA ID and annotation[A]	lncRNA locus[B]	Proximal coding-gene ID[C]	Proximal transcript locus[D]	Proximal coding-gene annotation[E]	Proximal coding-gene log2 fold change[F]	BLAST alignment (% identity, E-value, Query coverage)[G]
Hzea.28800 (LOC113495312)	895622-896089	Hzea.28797	883429-886507	*Helicoverpa armigera* putative ATPase N2B (LOC110384128), mRNA	0.321567 log2 fold increase	No significant alignment
		Hzea.28802	909695-911283	*Helicoverpa armigera* prothoracicotropic hormone-like (LOC110384160), mRNA	1.44582 log2 fold increase	No significant alignment
		Hzea.28803	912030-915696	*Helicoverpa armigera* zinc finger protein 239-like (LOC110384141), transcript variant X1, mRNA	0.729596 log2 fold increase	No significant alignment

[A]lncRNA ID indicates fasta file ID number, annotation indicates NCBI defined annotation.
[B]locus number indicates base-pair coordinates for lncRNA within scaffold.
[C]coding-gene ID indicates fasta file ID number.
[D]locus number indicates base-pair coordinates for coding-gene within scaffold.
[E]coding-gene annotation indicates NCBI defined annotation.
[F]Indicates magnitude of log2 fold change and direction of expression.
[G]NCBI BLAST alignment of lncRNA to proximal coding-gene in following order: % ID, E-value, and % Query coverage.

Supplementary Figure S5. Genomic scaffold for lncRNAs found only in the S strain and identification of proximal protein coding-genes. The scaffold at the top depicts the range of the scaffold in kilobases (kb). The bars in blue indicate sequences present on each scaffold, with gene ID numbers below each. The red stars indicate a lncRNA, the black stars indicate a protein-coding gene. The scaffold ID number is placed directly below the legend on the left side. The table below the scaffold includes the following information about the lncRNA and coding-genes found in the scaffold from left to right: the lncRNA ID number and annotation, lncRNA loci coordinates, gene ID number of proximal coding-gene, coding-gene loci coordinates, coding-gene annotation (NCBI defined), coding-gene log2 fold change, and BLASTn alignment results (% identity, E-value, and query coverage) comparing the lncRNA and the protein coding-gene.