Title
Transcriptome analysis unravels RNAi pathways genes and putative expansion of CYP450 gene family in cotton leafhopper *Amrasca biguttula biguttula*(Ishida)

Authors
Mridula Gupta§#, Satnam Singh§*, Gurmeet Kaur§, Suneet Pandher§, Noorpreet Kaur*, Neha Goel‡, Ramandeep Kaur§, and Pankaj Rathore§

Affiliations
§ Punjab Agricultural University, Regional Station, Faridkot-151203, Punjab (India)
# Department of Veterinary Sciences, Texas A & M University, Tx, USA
* University of Connecticut, Storrs, Connecticut 06269
‡Forest Research Institute, Dehradun-248195, Uttarakhand (India)

* Corresponding author: Satnam Singh (satnam@pau.edu)

Short Running title: RNA-seq analysis of cotton leafhopper

* Corresponding author

Satnam Singh
Regional Research Station
Punjab Agricultural University
Circular Road
Faridkot-151203,
Punjab, India
e-mail:satnam@pau.edu
Telephone +91-9876962774
Abstract

*Amrasca biguttula biguttula* is an important pest of cotton and okra in the Indian subcontinent. Presently limited genomic/ transcriptomic information is available for this insect in any open source databases. To initiate molecular studies in this insect, we report first assembled and annotated *de novo* transcriptome of cotton leafhopper. Out of 75,551 transcripts, 39613 CDS (Coding Sequence) were predicted with 35282 showing positive blast hits with NCBI nr database. From the Gene ontology (GO) analysis, 7431 CDS were annotated. KEGG pathway analysis categorized CDS into 22 different functional categories. The majority of CDS were annotated in signal transduction and transport catabolism pathways. The sequence data was screened for RNAi pathway genes and presence of 37 transcripts associated with this process confirmed the existence of robust RNAi machinery in this insect. The role of core RNAi machinery genes (*Dicer-2*, *Ago-2*, *Piwi* and *Staufen*) has been validated through dsRNA feeding studies. The data resource has also been used to identify potential RNAi targets and genes associated with insecticide detoxification specifically CYP 450 family.

Data Records

Raw sequencing data in FASTQ was submitted to NCBI Sequence Read Archive (SRA) (SRR6308174) under BioProject (PRJNA417995) BioSample (SAMN08013300), and computationally assembled sequences were submitted under TSA (Transcriptome Sequence Assembly) accession number GGAG00000000.

Key words: *de novo* Transcriptome, cotton leafhopper, RNAi, CYP450, insecticide resistance, phylogenetic analysis
Introduction

Numerous hemipteran insects such as cotton leafhoppers, bugs (mealybug, red cotton bug and dusky cotton bug) and aphids are economically important pests and major threat to agricultural crops. Cotton leafhopper or jassid, *Amrasca biguttula biguttula* (Ishida) (Hemiptera: Cicadellidae), is among the most damaging sucking pests of cotton and okra. Other major hosts for this insect include groundnut, jute, soybean, niger, sunflower, aubergine, potato, mung bean and cowpea. Its severe infestation in the crops causes bronzing or brick red coloring of leaves, which is known as “hopper burn” (Iqbal et al., 2008). Apart from drying of leaves, the leafhopper damage induces shortening of internodes and downward curling of leaves which significantly reduces plant vigour, growth and yield (Somnuck, 2001). This insect also infests many malvaceous and solanaceous crops, however in cotton alone its damage accounts up to 25-45% loss in yield (Dhawan & Sidhu, 1986). Since the mainstay of leafhopper management mainly relies on synthetic insecticides (Gallun et al., 1975), their excessive use has led to the resistance development against major insecticidal groups in this insect (Saeed et al., 2017). However, the molecular basis of resistance needs complete information on the various genes associated with the xenobiotic detoxification. Reportedly, there are three foremost xenobiotic detoxifying enzyme families which play role in metabolic resistance to insecticides are the carboxylesterases (COEs), glutathione transferases (GSTs), and cytochrome P450s (CYP450s) (Ranson et al., 2002). The proteins of these families are associated with the synthesis and breakdown of a large amount of endogenous metabolic moieties, protection against oxidative stress, detoxification of insecticides and the movement of molecules through cells (Le Goff et al., 2003). In-depth knowledge of molecular and biochemical basis of insecticide resistance development would help to devise strategies for its future management which is quite challenging as the various subfamilies of these enzymes are associated with numerous metabolic pathways. To proceed with this research, we have used
the present transcriptome to identify potential GSTs, CYP450s, COEs, of *A. biguttula biguttula* using BLASTx for the first time, which can be used in future to interpret molecular basis of insecticide resistance. Bioassays based few studies have been conducted for continuing insecticide resistance but lack of sequences data resources have restricted the validation of results through molecular tools. Thus, to decipher evolved insecticide resistance at molecular level as well as to develop novel eco-friendly strategies for pest management, genome or transcriptome information is a pre-requisite (Sagar & Balikai, 2014).

Apart from a few studies on mitochondrial-COI based genetic diversity, a recent study on reference genes validation and the demonstration of feeding RNAi in this insect, no genomic or transcriptomic resource is available for *A. biguttula biguttula* to best of our knowledge (Kranthi et al., 2018; Singh et al., 2018). Since RNAi is a potential present-day tool for exploring functional genomics, holds wide scope for developing new generation insect-pest management strategies. In this endogenous cellular process, the double stranded RNA (dsRNA) when binds to complementary mRNA, directs its cleavage, and eventually blocks the expression of targeted gene which may lead to phenotypic and physiological changes, and insect mortality in some cases (Singh et al., 2019b; Rodrigues et al., 2018; Burand & Hunter, 2013; Perrimon, Ni, & Perkins, 2010). There are a series of pathway genes that govern this Ai mechanism in an organism starting from uptake of dsRNA by the cell till knockdown of targeted mRNA and disruption of protein synthesis. Therefore, sequence information availability is the prerequisite for initiating RNAi studies in any organism. We have tried to mine the *A. biguttula biguttula* transcriptome for identification of RNAi pathway genes and validated the same through dsRNA feeding assays. Additionally, we have also identified vital and unique targets amenable for RNAi based on earlier identified targets from other insects, which can be exploited in future to develop management tools in the form of dsRNA based sprays or insect resistant transgenic crops (Artymovich, 2009). The *de novo* transcriptome of *A. biguttula biguttula* is the first ever assembled and annotated sequence data resource
submitted in NCBI repository. To compare our results, we assembled and annotated the available *A. biguttula biguttula* transcriptomes (Adult male: SRR9705737, SRR9705738; Adult female: SRR9705739, SRR9705740; Late Nymph: SRR9705741, SRR9705742 and Early Nymph: SRR9705743, SRR9705744). Therefore, this study intends to unveil detailed information about the Cytochrome P450 (CYP450) gene family and RNAi pathway genes, which may be useful for the groups interested to take up molecular studies in this insect.

**Experimental Procedures**

**Transcriptome assembly and annotation**

**Sample collection and RNA isolation:**

The cotton leafhopper was collected from the insect culture maintained on cotton plants in walk-in environmental chambers of Punjab Agricultural University, Regional Research Station, Faridkot, Punjab, India. Samples for *de novo* transcriptome included 15-20 individuals of *A. biguttula biguttula* collected from these cultures. The male and female individuals (males and females equal in number) were pooled in 0.2 ml TRI-Reagent (Sigma, St. Louis, USA), stored at −80 °C. Total RNA extractions were performed using TRI-reagent (Sigma, St. Louis, USA) as per manufacturer’s protocol followed by quality evaluation on a Qubit Fluorometer for the generation of RNA-seq libraries (Supplementary File S1: Figure S1 and Figure S2).

**cDNA synthesis and Sequencing**

The total RNA extracted in the previous step was subjected to the construction of PE (Paired End) libraries using TruSeq stranded mRNA Library Prep Kit (Illumina) as per manufacturer’s instructions. To proceed, the mRNA was converted into First Strand cDNA using poly-T attached magnetic beads and enzymatic fragmentation followed by First double stranded cDNA synthesis using second strand mix and Act-D mix. This double stranded cDNA was purified using Ampure XP beads followed by A-tailing, adapter ligation and enrichment by 10 PCR cycles (10 cycles). These PCR libraries were further evaluated in a 4200 TapeStation system (Agilent Technologies) with High sensitivity D1000 Screen tape as per manufacturer
instructions. Afterwards, the PE Illumina library was loaded onto NextSeq 500 for clustering and sequencing which allows it to be read in both forward and reverse directions.

**Sequence assembly**

To retrieve high quality reads, the sequenced raw data was processed using Trimmomatic v0.35 (Bolger, Lohse, & Usadel, 2014) which involved the removal of adapter sequences, ambiguous reads (reads with unknown nucleotides “N” larger than 5%), and low-quality sequences (reads with more than 10% quality threshold (QV) > 20 phred score). The threshold length of 100 nt (nucleotide) after trimming was applied to eliminate the sequences below this size. The reads were assembled using Velvet v1.2.10 (Zerbino & Birney, 2008) and Oases v0.2.09 (Schulz et al., 2012) (*De novo* transcriptome assembler for the velvet package) on optimized *Kmer* 57. The reads were mapped and aligned back to their reference assembled transcripts using BWA v0.7.12 (Burrows-Wheeler Aligner) for validation of transcriptome (Li & Durbin, 2009). The ORFs were predicted from these transcripts using TransDecoder (Haas et al., 2013) and further annotated using various annotation tools.

**Functional Annotation**

The sequences (ORFs) obtained were annotated using BLASTx against NCBI nr databases. The BLASTx resulted accession IDs were used to retrieve gene names or symbols, which were then searched in the species-specific entries of the gene-product tables of GO database. BLASTx analysis of the resulting proteins was performed with the UniProt, Swiss-Prot, PSD, TrEMBL, RefSeq, GenPept and PDB to have access to UniProt IDs. Accession IDs were searched directly in the dbxref table of GO database. Gene Ontology (GO) of the predicted CDS were determined by the Blast2GO program (Conesa et al., 2005). Afterwards, GO assignments were used to classify the functions of the predicted CDS. GO mapping was carried out in order to retrieve GO terms for all the BLASTx functionally annotated CDS. The completeness and contiguity of the assembled transcriptome was further validated using BUSCO v1.1b1 (Benchmarking Universal Single-Copy Orthologs) (Waterhouse et al., 2017).
This analysis was performed using insect (1658) and arthropoda (1066) BUSCO dataset, which includes BUSCO lineages of 42 insects and 60 arthropods (‘http://busco.ezlab.org/v2/datasets/insecta_odb9.tar.gz’; ‘BUSCO: https://busco.ezlab.org/datasets/arthropoda_odb9.tar.gz’).

RNAi in leafhopper

Identification of RNAi machinery genes and RNAi targets: The assembled transcriptome was mined for the presence of RNAi machinery in *A. biguttula biguttula*. BLAST program was used to authenticate the sequences retrieved from the annotated transcriptome. Furthermore, the phylogenetic analysis of core RNAi genes was done using a neighbour joining method by MEGA X 10.1 with 1000 bootstrap value (1000). Similarly, novel and vital gene targets amenable for RNAi based on earlier reports from other insect species (Khan et al., 2017; Rodrigues et al., 2017) were identified in this insect from the annotated sequences.

Expression of core RNAi genes across different stages of leafhopper: To validate the identified RNAi machinery genes from *A. biguttula biguttula* transcriptome, 1µg total RNA isolated from 15 adults, 20 younger nymphs (1st, 2nd and 3rd instars) and 15 late nymphs (4th and 5th instars) in three biological replicates as per earlier described protocols was used to synthesize cDNA with First Strand cDNA Synthesis Kit (ThermoFisher Scientific) as per manufacturer’s instructions. The expression of core RNAi genes was evaluated across different stages of leafhopper using RT-qPCR and qPCR forward and reverse primers designed using Primer3 software (Untergasser et al., 2007) (Supplementary File S1: Table S1). Quantitative expression studies were carried out using 5µl of 5x SYBR Premix Ex Taq (Tli RNase H Plus) (Clontech takara), 0.2 µl of 10pmol primers and 1µl of 1/10 cDNA template in a 10µl reaction in Lightcyler (Roche).

Functional validation of core RNAi genes: *SNF7* functions as a constituent of *ESCRT* (Endosomal Sorting Complex Required for Transport) pathway, which has role in cellular functions such as internalization, transport, sorting and lysosomal degradation of
transmembrane proteins (Huseth et al., 2016). To validate the functioning of core RNAi genes, Dicer-2, Ago-2, Piwi, and Staufen were knocked-down and their impact on RNAi efficiency of SNF7 was compared with the dsGFP fed control insects. Gene-specific primers with T7 promoter sequence at 5’ ends of both forward and reverse primers were custom synthesized (Eurofin Genomics, Bangalore India) to amplify the template fragments for dsRNA synthesis of Dicer-2, Ago-2, Piwi, Staufen and SNF7 genes (Supplementary File S1: Table S1). The respective gene specific PCR product was amplified using leafhopper cDNA followed by purification using Nucleospin Gel and PCR Cleanup kit (Macherey–Nagel) as per instruction manual. The eluted PCR product was used to synthesize respective gene specific dsRNA using MEGAscript™ RNAi Kit (Thermo Fisher Scientific). Biological samples in triplicate each comprising 15 individuals (5/ tube) of 5th instar nymphs were released in 1.5 ml tube (lower 1/4th part of the tube was cut and covered with muslin cloth) (Brar et al., 2018). The dsSNF7 (500 ng/μl) was fed along with a diet stretched between two layers of parafilm as per previously described methodology (Brar et al., 2018). The dsRNAs against Dicer-2, Ago-2, Piwi and Staufen genes were evaluated to examine their respective knockdown impact on RNAi of SNF7 gene. For this, 500 ng/μl each of dsDicer-2, dsAgo-2, dsPiwi and dsStaufen was administered in 100 μl of diet (Singh et al., 2018) separately for each experimental setup. After 24 hours, the above dsRNAs against respective RNAi gene was replaced by dsSNF7 in the diet of cotton leafhoppers as per methodology described (Supplementary File S1: Figure S3). Live cotton leafhoppers were collected post 48 h from respective feeding chambers, and total RNA was extracted and subjected to expression studies as described in the earlier section. The relative expression of the SNF7 under the influence of each core RNAi pathway gene was calculated using ΔΔCT method in comparison with dsGFP fed control insects after normalization with RPL9 as housekeeping gene (Singh et al., 2018).

CYP450 gene family analysis in A. biguttula biguttula

Annotation, identification and classification of CYP450 gene family
The predicted protein data resource from sequenced transcriptome was used for identification of transcripts associated with CYP450 (cytochrome P450) gene family in *A. biguttula biguttula* using Hidden Markov Model (HMM) based HMMER v3 (Eddy & Wheeler, 2015). HMM profile was built using raw sequences corresponding to PFAM CYP (PF00067) and 50 protein seed sequences having unique CYP domains. HMM profile of the CYP gene family was used in local search against *A. biguttula biguttula* protein sequences with e-value $\geq 1e^{-5}$. The identified CYP genes in leafhopper were reconfirmed through PFAM database search and BLASTx search against NCBI nr database. Alignment of predicted CYP450 protein sequences was done using MUSCLE algorithm along with 76 genes representing different CYP clans from *Bombyx mori*, *Apis mellifera*, *Tribolium castaneum* and *Acyrthosiphon pisum* (“http://drnelson.uthsc.edu/CytochromeP450.html,” n.d.). The aligned sequences were used for phylogenetic analysis with the help of Maximum parsimony method (bootstrap 500) in MEGAX. The Newick output of the tree was viewed and edited in Figtree (“http://tree.bio.ed.ac.uk/software/figtree/,” n.d.) for the better classification of different clades. We also analysed the number of CYP450 genes present in the available transcriptomes for different developmental stages of the cotton leafhopper (Adult male: SRR9705737, SRR9705738; Adult female: SRR9705739, SRR9705740; Late Nymph: SRR9705741, SRR9705742; Early Nymph: SRR9705743, SRR9705744). After pulling out these sequences from all available transcriptomes as well as our lab’s transcriptome, venn diagram for those sequences was generated by orthoMCL based OrthoVenn tool with the e-value $1e^{-10}$. The identified CYP genes were also validated using Multiple Em for Motif Elicitation (MEME; version 4.9.1) at http://meme.nbcr.net/meme and SALAD database (http://salad.dna.affrc.go.jp/CGViewer/en/howto/multi_motif_nj.html with default parameters to classify different clans on the basis of conserved sequence motifs (CSMs) (Mihara, Itoh, & Izawa, 2009). The CSMs obtained from SALAD were compared with those obtained with MEME to validate and find the most common CSM in a particular clan of CYP450.
Additionally, the other genes belonging to carboxylesterases (COEs), monooxygenases, and glutathione transferases (GSTs) associated with xenobiotic detoxification were also identified and annotated from this transcriptome using BLASTx search against NCBI nr database.

**Expression level of CYP450 genes of A. biguttula biguttula using RT-qPCR.**

F5\(^{th}\) instar nymphs (15 individuals per replicate) of two Faridkot (Punjab-India) populations (AbbFdk1 and AbbFdk2) were collected from the cotton fields using aspirators in triplicates. The two locations were selected based on differential insecticide usage patterns against cotton leafhopper. The Fdk 1 (30.4032° N, 74.4456° E) was a low insecticide usage location while Fdk2 (30.4906° N, 74.6386° E) was a comparatively higher insecticide using area. The collected live individuals were brought to the laboratory and processed for RNA isolation and cDNA synthesis using earlier described methodologies. Primers specific to each selected CYP gene sequence were designed and custom synthesized (Eurofin Genomics, Bangalore India) (Supplementary File S1: Table S1). Randomly selected 9-10 genes representing each clan were evaluated for their expression across these two populations of *A. biguttula biguttula*. Fifteen genes (3-4 per clan), which showed comparable expression were selected for differential expression analysis between two field populations of leafhopper. The comparative mRNA expression of these genes among two populations was evaluated using Real-time PCR (Lightcycler, Roche) as described in the earlier section.

**Results**

**Transcriptome assembly and annotation**

The libraries sequenced on NexSeq 500 (Illumina) using 2 x150 bp PE (paired end) library generated ~9.8 GB raw data (SRR6308174) after eradicating adapter sequences and ambiguous reads. After removing uncertain sequences from the raw data, 28,400,988 (2 x 150 bp) high quality PE reads (QV>20) were obtained, which were used for *de novo* assembly (Table 1). The reads were assembled using high throughput Illumina NexSeq500 platform and *de novo* transcriptome assembly generated with the help of Velvet (Zerbino & Birney, 2008) and Oases
(Schulz et al., 2012) softwares yielded 75,551 transcripts from our lab transcriptome (SRR6308174), 59,663 from Adult female (SRR9705739, SRR9705740), 121865 from Early nymph (SRR9705743, SRR9705744), 115950 from late nymph (SRR9705741, SRR9705742) and 100790 from Adult male (SRR9705737, SRR9705738). The open reading frames (ORFs) of these transcripts were predicted using TransDecoder to understand the protein coding sequences (Tang, Lomsadze, & Borodovsky, 2015). The predicted 39,613 CDS sequences with average length 3080 bp, yielded 35,282 annotated sequences after BLASTx against NCBI nr database. Similarly, the total number of CDS in Adult male (SRR9705737, SRR9705738) were 29212, in Adult female (SRR9705739, SRR9705740) were 47958, Late Nymph (SRR9705741, SRR9705742) were 43747 and Early Nymph (SRR9705743, SRR9705744) were 43747. The majority of hits were found against Z. nevadensis, Athalia rosae, T. castaneum and Cimex lectularius. The total of 35,282 positive BLAST hits showed comparable matrices to Z. nevadensis (23%), C. lectularius (15%), Halymorpha halys (14.3%), T. castaneum (3.4%), Diaphorina citri (3.22%), A. pisum (2.59%), A. rosae (2.29%), Pediculus humanus (1.95%), Diuraphis lugens (1.8%), Nilaparvata lugens (1.5%) and others (18%) with cut off e-value of $10^{-5}$. A total of 10.9 % of predicted CDS went uncharacterized, which may be a unique gene resource for future studies on the cotton leafhopper (Supplementary File S1, Fig. S4).

Gene ontology (GO) analysis using Blast2GO annotated 7,431 CDS into Biological processes, Molecular function and Cellular component (Supplementary File S1, Fig. S5). The most enriched terms for Biological processes were organic substance metabolic process (16.36% of total number of sequences), primary metabolic process (15.55 % of total CDS) and cellular metabolic process (15.24% of total sequences). In Molecular function, main pathways were organic cyclic compound binding (19.88% of total CDS) and heterocyclic compound binding (19.87% of total CDS). Under Cellular component, primary pathways were covered by intracellular pathways (23.53 % of all the sequences) followed by intracellular part (19.45%) (Supplementary File S1, Fig. S5). Gene ontology of predicted CDS sequences has been
represented with the help of WEGO plot, which shows the percentage participation of these genes in different pathways (Supplementary File S1, Fig. S5). The transcripts were functionally annotated using KEGG analysis and KO IDs were assigned to four biological processes, i.e., Metabolism, Cellular, Genetic and Environmental information processing (Supplementary File S1, Fig. S6). The majority of transcripts were grouped under metabolism (33.92 %) followed by cellular processes (20.61%). Individually among these major processes, the majority of transcripts were associated with carbohydrate metabolism (17.71%), folding, sorting and degradation (34.76%), signal transduction pathway (89.18%) and transport and catabolism (39.52%) under metabolism, genetic information processing, environmental information processing and cellular, respectively. Out of 39613 CDS, only 21.42 % were assigned the KO IDs using KEGG pathway analysis (Supplementary File S3, Fig. S6). The completeness of all transcriptomes was assessed using BUSCO (Waterhouse et al., 2017) by validating it against insect and arthropod lineages which revealed that 82.6-91.6% and 86.3-94.4% BUSCO genes were “complete” compared to insect and arthropod lineages, respectively. The remaining 4.9-9.0% and 4.6-8.0% were “fragmented”, while 2.1-8.8% and 1.2-6.8% were missing compared to insect and arthropod lineages, respectively (Table 2). The quality of our assembled transcripts was comparable to the earlier assemblies listed in Simao et al (Simão et al., 2015)

**Identification of RNAi targets and RNAi machinery genes in leafhopper**

The assembled transcriptome can be used as a future resource to identify genes associated with various physiological, biochemicals, hormone biosynthesis and many other processes, which are important from the pest-management perspective in this insect. In the current study, we explored RNAi pathway genes, potential RNAi targets as well genes associated with insecticide resistance in cotton leafhopper. The transcriptome screening suggests the existence of robust RNAi machinery in this insect with 175 identified transcripts associated with 27 RNAi processing genes. After, multiple sequence alignment and removal of duplicates, only 89 transcripts were found to be related with 27 RNAi machinery genes. The details of different
transcripts associated with each gene have been summarized in Supplementary File S1: Table S2. The prime siRNA machinery genes identified in this insect includes \textit{SID1}, \textit{Argonaute-1}, \textit{Argonaute-2}, \textit{Argonaute-3}, \textit{Dicer-1}, \textit{Dicer-2}, \textit{Piwi}, \textit{Aubergine}, \textit{RNaseIII}, helicases and dsRNA binding proteins. The expression of the core RNAi genes showed variation across different insect stages (Fig. 1), however it reconfirmed the presence of these genes in cotton leafhopper. The expression of \textit{SID1}, \textit{Dicer-2}, \textit{Ago-1}, \textit{Piwi}, \textit{PCDG6}, \textit{Ago-2} and \textit{RNaseIII} was significantly higher in adult compared to early and late nymphal instars. However, the expression of few genes such as \textit{Dicer-1}, \textit{Asp} and \textit{Dwil} in late nymph was significantly higher in late nymphs as compared to early nymph and adult. The expression of all the RNAi genes identified was either significantly or numerically lower in early nymph compared to other two stages. The BLASTx analysis for these genes revealed their notable similarity with homologs from other insects groups. Furthermore, the phylogenetic analysis using neighbour joining method of MEGAX (500 bootstrap value) confirmed that the depicted genes were involved in core RNAi pathways through matching metrics with their orthologs from other insect species (Fig. 2).

The sequence specific knockdown of core RNAi pathway genes such as \textit{Dicer-2}, \textit{Piwi}, \textit{Argonaute-2} and \textit{Staufen} resulted in 59.33 to 99.01\% decrease in mRNA expression level of these genes. The functional validation of RNAi pathway genes demonstrated through consecutive feeding of dsRNA against the RNAi pathway gene and \textit{SNF7} revealed 44.64 to 97.71\% hampering of RNAi efficiency of this gene. The knockdown of \textit{Dicer-2} deterred 44.6 \% of RNAi efficiency of \textit{SNF7} in \textit{A. biguttula biguttula}. The knockdown of \textit{Piwi} resulted in almost no RNAi effect with expression level of \textit{SNF7} being 97.7\% similar to that of GFP control. Knockdown of \textit{Piwi} resulted in maximum hampering of RNAi efficiency of \textit{SNF7} compared to other core RNAi genes and this may be attributed to 99.01\% decrease in mRNA level of \textit{Piwi} gene in dsPiwi fed insects (Fig. 3). Consecutive knockdown of \textit{Staufen} followed by \textit{SNF7} resulted in 46.7\% hampering of RNAi efficiency compared to only \textit{dsSNF7} fed insects .
The transcriptome analysis also revealed the existence of novel and potential RNAi targets in *A. biguttula biguttula* such as *chickadee* (responsible for female fertility, nourishment of oocytes (Shields et al., 2014) and regulation of germline stem cells (Kiger et al., 2001)), *snap* (major component in neurotransmitter machinery (Lin & Scheller, 2000)), GATA transcription factors (important for cardiogenesis and haematopoiesis (Sorrentino, Gajewski, & Schulz, 2005) and many others which have been earlier identified in other insect species. Additional novel targets such as *tubulin*, *cyclophilin*, *heat shock proteins*, *crumbs*, *tetraspanin*, *apoptosis inhibitor* and *serine protease* were also screened and mined from the *A. biguttula biguttula* transcriptome (Rodrigues et al., 2018; Li et al., 2017) (Supplementary File S1: Table S3). Thus the transcriptome data may also be useful for searching potential RNAi targets which can be utilized in future to develop novel and eco-friendly pest-management strategies for *A. biguttula biguttula*.

**CYP450 gene family identification**

(Supplementary File S1: Table S5). Similarly, the total 171 unique transcripts CYP450 genes were identified and pulled out from the available assembled transcriptomes when the transcripts of all developmental stages were merged using orthoMCL i.e. Adult male-SRR9705737, SRR9705738; Adult female- SRR9705739, SRR9705740; Late nymph- SRR9705741, SRR9705742; Early nymph- SRR9705743, SRR9705744. In addition, there were only 38 common CYP450 transcripts in all transcriptomes (Fig 4). The phylogenetic tree analysis based on Maximum parsimony method of 158 putative CYP450 transcript sequences of our lab’s *A. biguttula biguttula* with 76 CYP450 genes of *B. mori*, *A. mellifera*, *T. castaneum* and *A. pisum* (‘http://drnelson.uthsc.edu/CytochromeP450.html’) segregated these sequences into different CYP clans (CYP2, CYP3, CYP4 and mitochondrial CYP) through clade classification. The insect CYPs have been classified into four clans with majority of genes falling in Clan 3 in the present studies as well as in *Aedes aegypti* (Issa, 2014), *T. castaneum* (Zhu et al., 2013) and *Erthesina fullo* (Y. Liu et al., 2015). In the case of cotton leafhopper, the
CYP genes belonging to respective clans are clustered as one clade with maximum number of transcripts (79) in Clan 3 (orange). The remaining transcripts grouped into Clan 2 (red) (30), Clan 4 (yellow) (13) and mitochondrial Clan (violet) (11) [(Fig. 5) (Supplementary File 1: Table S6). The remaining uncolored sequences were not grouped into any clan as were too divergent. In our studies we also identified some CSMs such as C-helix, I-helix, K-helix, Heme binding region and PERF motifs by MEME and SALAD (Supplementary File 1: Fig.S5- S8). The predicted CSMs from different CYP clans will be helpful in understanding structure and function of CYP family proteins in *A. biguttula biguttula*. In addition, the transcriptional analysis of 15 representative CYP450 genes of different clans revealed the differential expression between two different insecticide exposed leafhopper populations (Fig. 6). However, this needs to be studied in detail considering more CYP genes and resistant and baseline susceptible populations of leafhopper.

The gene ontology with Blast2GO was conducted to assign GO terms to CYP450s. Out of 182 CYP450s transcripts 116 were mapped and annotated to GO terms, of which 90 belonged to biological process, 140 to molecular function and 84 to cellular component. Among these categories, catalytic activity, binding, membrane and metabolic process were the largest subcategories (Fig. 7). A subsequent analysis of these transcripts with KEGG revealed that four clans contained multiple CYP450 families such as CYP4, CYP6, CYP15, CYP18, CYP49, CYP302, CYP306, CYP307, CYP314, and CYP15. However, CYP4 and CYP6 were found to be largest CYP450 families in this CYPome (Supplementary File S1: Table S5). About 124 out of 235 CYP putative transcripts were annotated in KEGG database and were assigned to six pathways. These pathways include lipid metabolism (Fatty acid biosynthesis), metabolism of terpenoids and polyketides (Insect hormone biosynthesis), environmental information processing (AMPK signalling pathway), Endocrine system (Insulin signalling pathway), metabolism (Cytochrome P450), signalling and cellular processes (Exosome) (Fig. 7). Further, “metabolism of terpenoids and polyketides” pathways included seven CYP450s from
Mitochondrial as well as Clan 2. Similarly, four CYP genes from ‘fatty acid biosynthesis’, two from ‘Insulin signaling pathway’ and two from ‘AMPK signaling pathway’ belonged to Clan 2 as well as Clan 3. About 107 CYP genes were found to be involved in ‘metabolism by Cytochrome P450’ pathway.

Discussion

The de novo transcriptome of A. biguttula biguttula generates the first assembled and annotated sequence [1]source for the future pest management strategies at molecular level. Without availability of reference genome sequences, profound transcriptome analysis can be a foundation for future gene expression and functional analysis on A. biguttula biguttula and can provide better understanding of its biological processes and molecular mechanisms. About 39 thousands of contigs were generated with average length (N50) of predicted CDS was 3080 bp, which was approximately two-three times to the average CDS length predicted in earlier transcriptomes of Homalodisca liturata (1650bp), Clastoptera arizonana (2510bp), Cuerna arida (1560bp), Graphocephala atropunctata (1692bp) and Leptinotarsa decemlineata (570bp) (Tassone, Cowden, & Castle, 2017; Kumar et al., 2014). The A. biguttula biguttula transcriptome showed the highest similarity with Z. Nevadensis (Harrison et al., 2018), C. Lectularius (Bai et al., 2011) and Halyomorpha halys (Sparks et al., 2014).

The transcriptome was studied to curate the potential RNAi genes that exist in cotton leafhopper transcriptome. The expression studies revealed variability in the core RNAi genes across developmental stages, however the variation in the expression of RNAi genes among different developmental stages has also been reported in earlier studies with T. castaneum (Perkin & Oppert, 2019), T. tabaci (Singh et al., 2019) and P. solenopsis (Singh et al., 2019b).

In depth studies are required to understand the reasons for differential expression of RNAi pathway genes across developmental stages of a particular insect species. The genes from A. biguttula biguttula such as SID1 and Clathrin heavy chain responsible for dsRNA uptake and commencement of RNAi pathway showed similarity with Z. nevadensis (Rocha et al., 2011;
Shih & Hunter, 2011). Two different paralogs of dicers i.e. \textit{Dicer-1} and \textit{Dicer-2} were identified in the transcriptome which showed high resemblance to hemipteran insects \textit{N. lugens} and \textit{Graminella nigrifrons}, respectively. Three paralogs of \textit{Argonautes} identified from this insect were closely related to \textit{Cerapachys biroi}, \textit{N. lugens} and \textit{G. nigrifrons}. Current study also focuses on functional validation of core RNAi genes through dsRNA mediated knockdown of \textit{Dicer-2}, \textit{Piwi}, \textit{Argonaute-2} and \textit{Staufen}. Their knockdown decreased the RNAi efficiency of the target gene \textit{SNF7}. Many studies have reported that \textit{Dicers} produce small RNAs that includes siRNA as well as miRNA, these 20-23 nt RNA fragments are important component of target mRNA degrading RISC (RNA-induced silencing complex) (Lee et al., 2004). Similar studies in \textit{L. decemlineata} demonstrated the role of \textit{Dicer-2} in RNAi pathway (Yoon et al., 2016). Since \textit{Piwi} knockdown led to more efficient RNAi hinderance showed its significant role in RNAi mechanism. \textit{PAZ} and \textit{Piwi} are the two domains present in the middle and C-terminal region of argonaute proteins, respectively. \textit{Piwi} sub family has been characterized from different insect species and consists of 300 aa similar in structure to catalytic domain of RNaseH, which performs a crucial role in slicer (Zhou et al., 2007). The knockdown of \textit{Argonaute-2} led to 70.6% blockage of RNAi, which primarily mediate target recognition and form silencing complex (Gu & Knipple, 2013). Previously, it has been reported that in the cell line of \textit{L. decemlineata} (coleopterans), the knockdown of \textit{Dicer-2} and \textit{Argonaute} partially blocked RNAi in the cells (Yoon et al., 2016). \textit{Staufen} is a dsRNA-binding protein initially discovered in \textit{Drosophila} (St Johnston, Beuchle, & Nüsslein-Volhard, 1991) and its role in RNAi machinery has been demonstrated in \textit{C. elegans} (LeGendre et al., 2013). \textit{Staufen} has not been much studied in insects in context to RNAi. However, a recent report in \textit{L. decemlineata} reveals specific \textit{Staufen (StauC)} in coleopterans, which predominantly contributes to the RNAi machinery and enhances knockdown efficiency (Yoon et al., 2018). The \textit{Staufen} bearing five dsRBD (dsRNA Binding Domain) has also been predicted from \textit{A. biguttula biguttula} transcriptome using sequence analysis software SMART (Letunic & Bork, 2018).
Staufen may be contributing to RNAi in insect’s species other than coleopterans, which lack specific StauC or its homologue as its knockdown has hampered the RNAi mechanism in our studies. Earlier studies with *Thrips tabaci* (Singh et al., 2019a) and *P. solenopsis* (Singh et al., 2019b) also reveal the role of Staufen protein in RNAi mechanism.

The variation in dsRNA mediated knockdown efficiency of different genes within the same organism may be attributed to many known and unknown factors such as length of dsRNA, delivery method, target gene etc. (Burand & Hunter, 2013). Our earlier studies also demonstrated the functional feeding RNAi in *A. biguttula biguttula* through the differential level of knockdown of *SNF7, VATPase, AQP* and *IAP* genes (Singh et al., 2018).

Insecticide resistance is the major issue in insects of agricultural importance due to continuous and sometimes indiscriminate use of insecticides for their management. The information is lacking on contribution of CYP 450, voltage gated Na channels, GABA receptors and many other genes towards insecticide resistance in cotton leafhopper. Our first sequence data resource provides an opportunity for better understanding of insecticide resistance mechanism at system biology level for this economically important non-model organism.

Resistance at molecular level has been differentially associated with various gene families such as point mutations in ion channel of GABA receptor (cyclodiene insecticides) and vicinity of the acetylcholinesterase active site (organophosphorus and carbamate insecticide resistance), mutations linked to a sodium channel gene (DDT and pyrethroid insecticides), amplification of esterase genes (organophosphorus and carbamate insecticides) as well as uncharacterized mutations leading to up-regulation of detoxification enzymes, such as cytochrome P450 and glutathione S-transferases (many classes of insecticides) (Feyereisen 1995). These gene families are known to be related to insecticide resistance mechanism in insects (Li, Schuler, & Berenbaum, 2006). Thus taking into account the current information generated for various gene families, the detailed studies with resistant and susceptible populations of *A. biguttula*
*biguttulla* will help to elucidate and validate the indepth role of various genes in insecticide resistance. Our findings suggest that there is vast diversity in the CYP genes of the cotton leafhopper with different stages and change of fields. Our major focus was on identification of Cytochrome P450 transcripts, which is the vast gene superfamily consisting of numerous families and subfamilies. These genes are majorly associated with detoxification of synthetic chemicals consumed by the pests thus playing a key role in imparting resistance against insecticides due to evolving nature of the CYPome (Li, Schuler, & Berenbaum, 2006). The insect CYPs have been classified into four clans with majority of genes falling in Clan 3 in the present studies as well as in *Aedes aegypti* (Issa, 2014), *T. castaneum* (Zhu et al., 2013) and *Erthesina fullo* (Y. Liu et al., 2015). The CSMs (Conserved Sequence Motifs) present in CYPs are involved in opening and closing of the substrate access/egress channel by modulating the flexible/plastic region of the protein and overall protein stability that affects enzyme function (Oezguen & Kumar, 2011). To date only five motifs i.e. C-helix, I-helix, K-helix, Heme binding region and PERF motifs are known in insect CYP genes (Feyereisen, 2005). The distinct motif located in the C-helix is WxxxR, and the Arg (R residue) is known to form a charge pair with the propionate of the Heme. The CSM present in the middle of the long helix I is GxE/DTT/S, which surrounds a conserved threonine (T-residues) that is responsible for oxygen activation. The PERF motif PxxFxPxRF contributes towards stabilization of the overall structure of CYP protein through a set of salt bridge interactions (E—R—R) with the CSM ExLR of K-helix. The fifth conserved motif is Heme binding region i.e PFxxGxRxCxG/A which carries the cysteine (C residues) to bind with Heme iron on the opposite side of helix I(Sezutsu, le Goff, & Feyereisen, 2013). CYP450s CSMs are important for biocatalyst designing and specifically in context to insects, they have an important role in insecticide discovery (Chandor-Proust et al., 2013). On the whole the information about CYP450 genes, their domains and CSMs will also be useful in unravelling the gene structural changes occurring due to development of insecticide resistance in this insect. In an hemipteran insect *E.*
fullo (yellow spotted stink bug) the CYP genes from Clan 2 and mitochondrial Clan were also involved in Insect hormone biosynthesis and those from Clan3 as well as mitochondrial clan participated in fatty acid biosynthesis (Y. Liu et al., 2015). This indicated that genes from different CYP clans are associated with one or multiple biological pathways. The P450 enzymes in insects are associated with diverse functions ranging from the synthesis and degradation of ecdysteroids and juvenile hormones to the metabolism of xenobiotics (Y. Liu et al., 2015; Feyereisen, 2005). The enzyme family plays a key role in adaptation of insects to toxic compounds in their host plants as well as metabolism of commonly used insecticides (Hodgson 1985). In some insects the detoxification is so active that the insecticide is metabolized before reaching the target site thus rendering organisms resistant to these molecules (N. Liu et al., 2015). The role of various CYP450 genes in insecticide resistance has been reported in many insect species such as Culex pipiens pallens (Chang et al., 2017), Culex quinquefasciatus (Liu et al., 2011), Myzus persicae (Puinean et al., 2010), Tribolium castaneum (Zhu et al., 2010), Drosophila (Le Goff et al., 2003), Anopheles gambiae (Nikou, Ranson, & Hemingway, 2003). It has been evident in many studies that the cotton leafhopper has developed many fold resistance to various classes of insecticides used for its management (Sagar & Balikai, 2014). Thus, identification of P450 genes in this insect will help to understand the molecular mechanisms underlying the insecticide resistance.

Acknowledgement

We thank Punjab Agricultural University, Ludhiana for the support to conduct this study.

Conflict of interest

Authors declare no conflict of interest

References

Artymovich, K. A. (2009). Using RNA interference to increase crop yield and decrease pest damage RNAi: A conserved mechanism across different organisms. Basic Biotechnology.
Bolger, A. M., Lohse, M. & Usadel, B. (2014). Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics (Oxford, England)*, 30(15), 2114–2120.

Brar, A. S., Gupta, M., Pandher, S. & Singh, S. (2018). Modified diet assay for studying feeding RNAi in cotton leafhopper, *Amrasca biguttula biguttula*. protocols.io. 10-11 Retrieved from https://www.protocols.io/view/modified-diet-assay-for-studying-feeding-rnai-in-c-ksgcwbw.pdf

Burand, J. P. & Hunter, W. B. (2013). RNAi: Future in insect management. *Journal of Invertebrate Pathology*, 112 (SUPPL.1), 9–14. Retrieved from http://dx.doi.org/10.1016/j.jip.2012.07.012

Chandor-Proust, A., Bibby, J., Régent-Kloeckner, M., Roux, J., Guittard-Crilat, E., Poupardin, R., Riaz, M. A., Paine, M., Dauphin-Villemant, C., Reynaud, S. & David, J.-P. (2013). The central role of mosquito Cytochrome P450 cyp6zs in insecticide detoxification revealed by functional expression and structural modelling. *Biochemical Journal*, 455(1), 75–85. Retrieved from http://biochemj.org/lookup/doi/10.1042/BJ20130577

Chang, K.-S., Kim, H.-C., Klein, T. A. & Ju, Y. R. (2017). Insecticide resistance and Cytochrome-P450 activation in unfed and blood-fed laboratory and field populations of *Culex pipiens pallens*. *Journal of Pest Science*, 90(2), 759–771.

Conesa, A., Götz, S., Garcia-Gómez, J. M., Terol, J., Talón, M. & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18), 3674–3676.

Dhawan, A. & Sidhu, A. (1986). Assessment of losses due to attack of Cotton Jassid on Hirsutum cotton. *Indian Journal of Plant Protection*, 14, 45–50.

Eddy, S. R. & Wheeler, T. J. (2015). *HMMER User Guide*. Retrieved from http://eddylab.org/software/hmmer3/3.1b2/Userguide.pdf

Feyereisen, R. (2005). Insect Cytochrome P450., in: Gilbert, L.I., Iatrou, K., Gill, S. S. (Ed.), *Comprehensive Molecular Insect Science*, (pp. 1–77). Elsevier, Oxford.
Le Goff, G., Boundy, S., Daborn, P. J., Yen, J. L., Sofer, L., Lind, R., Sabourault, C., Madi-Ravazzi, L. & Ffrench-Constant, R. H. (2003). Microarray analysis of cytochrome p450 mediated insecticide resistance in \textit{Drosophila}. \textit{Insect Biochemistry and Molecular Biology}, 33(7), 701–708.

Gu, L. & Knipple, D. C. (2013). Recent advances in RNA interference research in insects: implications for future insect pest management strategies. \textit{Crop Protection}, 45(2013), 36–40. Retrieved from http://dx.doi.org/10.1016/j.cropro.2012.10.004

Haas, B. J. et al. (2013). \textit{De novo} transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. \textit{Nature Protocols}, 8(8), 1494–1512.

Huseth, A. S., Chappell, T. M., Langdon, K., Morsello, S. C., Martin, S., Greene, J. K., Herbert, A., Jacobson, A. L., Reay-Jones, F. P., Reed, T., Reisig, D. D., Roberts, P. M., Smith, R. & Kennedy, G. G. (2016). \textit{Frankliniella fusca} resistance to neonicotinoid insecticides: an emerging challenge for cotton pest management in the eastern united states. \textit{Pest Management Science}, 72(10), 1934–1945.

Iqbal, J., ul Hasan, M., Ashfaq, M., Sahi, S. T. & Ali, A. (2008). Screening of okra genotypes against jassid, \textit{Amrasca biguttula biguttula} (Ishida) (Homoptera: Cicadellidae). \textit{Pakistan Journal Agricultural Sciences}, 45(5), 448–451.

Issa, M. S. (2014). Molecular characterization and functional analysis of cytochrome p450 genes in the yellow fever mosquito \textit{Aedes aegypti} (Diptera: Culicidae). \textit{Manhattan}. 1-151. Retrieved from http://krex.k-state.edu/dspace/bitstream/handle/2097/18253/MoustaphaSoumailaIssa2014.pdf;sequence=1

Khan, A. M., Ashfaq, M., Khan, A. A., Naseem, M. T. & Mansoor, S. (2017). Evaluation of potential RNA-interference-target genes to control cotton mealybug, \textit{Phenacoccus solenopsis} (Hemiptera: Pseudococcidae). \textit{Insect Science}, 25(5), 778–786.
Kiger, A. A., Jones, D. L., Schulz, C., Rogers, M. B. & Fuller, M. T. (2001). Stem cell self-renewal specified by JAK-STAT activation in response to a support cell cue. *Science (New York, N.Y.)*, 294(5551), 2542–2545.

Kranthi, S., Ghodke, A. B., Puttuswamy, R. K., Mandle, M., Nandanwar, R., Satija, U., Pareek, R. K., Desai, H., Udikeri, S. S., Balakrishna, D. J., Hugar, B. M., Monga, D. & Kranthi, K. R. (2018). Mitochondria COI-based genetic diversity of the cotton leafhopper *Amrasca biguttula biguttula* (Ishida) populations from India. *Mitochondrial DNA Part A*, 29(2), 228–235. Retrieved from https://doi.org/10.1080/24701394.2016.1275595

Kumar, A., Congiu, L., Lindström, L., Piirainen, S., Vidotto, M. & Grapputo, A. (2014). Sequencing, *de novo* assembly and annotation of the colorado potato beetle, *Leptinotarsa decemlineata*, transcriptome. *PLoS ONE*, 9(1), e86012.

Lee, Y. S., Nakahara, K., Pham, J. W., Kim, K., He, Z., Sontheimer, E. J. & Carthew, R. W. (2004). Distinct roles for *Drosophila dicer-1* and *dicer-2* in the siRNA/miRNA silencing pathways. *Cell*, 117(1), 69–81.

LeGendre, J. B., Campbell, Z. T., Kroll-Conner, P., Anderson, P., Kimble, J. & Wickens, M. (2013). RNA targets and specificity of Staufen, a double-stranded RNA-binding protein in *Caenorhabditis elegans*. *Journal of Biological Chemistry*, 288(4), 2532–2545.

Letunic, I. & Bork, P. (2018). 20 Years of the SMART protein domain annotation resource. *Nucleic Acids Research*, 46(D1), D493–D496.

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–wheeler transform. *Bioinformatics*, 25(14), 1754–1760. Retrieved from http://dx.doi.org/10.1093/bioinformatics/btp324

Li, X., Schuler, M. A. & Berenbaum, M. R. (2006). Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual Review of Entomology*, 52(1), 231–253.

Li, Y., Li, N., Wang, Z. K., Li, W. Bin, Ran, R. X., Li, H. Z. & Meng, F. L. (2017). RNA interference mediated serine protease gene (*spbtryl*) knockdown affects growth and mortality
in the soybean pod borer (Lepidoptera: Olethreutidae). Florida Entomologist, 100(3), 607–615.

Lin, R. C. & Scheller, R. H. (2000). Mechanisms of synaptic vesicle exocytosis. Annual Review of Cell and Developmental Biology, 16, 19–49.

Liu, N., Li, M., Gong, Y., Liu, F. & Li, T. (2015). Cytochrome P450s - their expression, regulation, and role in insecticide resistance. Pesticide Biochemistry and Physiology.

Liu, N., Li, T., Reid, W. R., Yang, T. & Zhang, L. (2011). Multiple Cytochrome P450 genes: their constitutive overexpression and permethrin induction in insecticide resistant mosquitoes, Culex quinquefasciatus. PLOS ONE, 6(8), 1–8. Retrieved from https://doi.org/10.1371/journal.pone.0023403

Liu, Y., Wu, H., Xie, Q. & Bu, W. (2015). Novel detection of insecticide resistance related p450 genes and transcriptome analysis of the hemimetabolous pest Erthesina fullo (Thunberg) (Hemiptera: Heteroptera). PLoS ONE, 10(5), 1–13.

Mihara, M., Itoh, T. & Izawa, T. (2009). SALAD Database: a motif-based database of protein annotations for plant comparative genomics. Nucleic Acids Research, 38(D835–D842).

Nikou, D., Ranson, H. & Hemingway, J. (2003). An adult-specific CYP6 P450 gene is overexpressed in a pyrethroid- resistant strain of the malaria vector, Anopheles gambiae. Gene, 318(1–2), 91–102.

Oezguen, N. & Kumar, S. (2011). Analysis of cytochrome p450 conserved sequence motifs between helices E and H: prediction of critical motifs and residues in enzyme functions. Journal of Drug Metabolism & Toxicology, 02(02). Retrieved from https://www.omicsonline.org/2157-7609/2157-7609-2-110.digital/2157-7609-2-110.html

Perkin, L. C. & Oppert, B. (2019). Gene expression in Tribolium castaneum life stages: identifying a species-specific target for pest control applications. PeerJ, 7, e6946.

Perrimon, N., Ni, J.-Q. & Perkins, L. (2010). in vivo RNAi: today and tomorrow. Cold Spring Harbor perspectives in biology, 2(8), a003640.
Puinean, A. M., Foster, S. P., Oliphant, L., Denholm, I., Field, L. M., Millar, N. S., Williamson, M. S. & Bass, C. (2010). Amplification of a cytochrome p450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *PLOS Genetics*, 6(6), 1–11. Retrieved from https://doi.org/10.1371/journal.pgen.1000999

Gallun R L, Starks K J & and Guthrie W D. (1975). Plant resistance to insects attacking cereals. *Annual Review of Entomology*, 20(1), 337–357. Retrieved from https://doi.org/10.1146/annurev.en.20.010175.002005

Ranson, H., Claudianos, C., Ortelli, F., Abgrall, C., Hemingway, J., Sharakhova, M. V, Unger, M. F., Collins, F. H. & Feyereisen, R. (2002). Evolution of supergene families associated with insecticide resistance. *Science (New York, N.Y.)*, 298(5591), 179–181.

Rocha, J. J. E., Korolchuk, V. I., Robinson, I. M. & O’Kane, C. J. (2011). A phagocytic route for uptake of double-stranded RNA in RNAi. *PLoS ONE*, 6(4), e19087.

Rodrigues, T. B., Dhandapani, R. K., Duan, J. J. & Palli, S. R. (2017). RNA interference in the asian longhorned beetle: identification of key RNAi genes and reference genes for RT-qPCR. *Scientific Reports*, 7(1), 1–10. Retrieved from http://dx.doi.org/10.1038/s41598-017-08813-1

Rodrigues, T. B., Duan, J. J., Palli, S. R. & Rieske, L. K. (2018). Identification of highly effective target genes for RNAi-mediated control of emerald ash borer, *Agrilus planipennis*. *Scientific Reports*, 8(1), 5020.

Saeed, R., Razaq, M., Abbas, N., Jan, M. T. & Naveed, M. (2017). Toxicity and resistance of the cotton leaf hopper, *Amrasca devastans* (distant) to neonicotinoid insecticides in Punjab, Pakistan. *Crop Protection*, 93, 143–147. Retrieved from http://dx.doi.org/10.1016/j.cropro.2016.11.032

Sagar, D. & Balikai, R. A. (2014). Insecticide resistance in cotton leafhopper, *Amrasca biguttula biguttula* (Ishida) – a Review. *Biochemical and Cellular Architecture*, 14(2), 283–294.
Schulz, C., Perdiguero, E. G., Chorro, L., Szabo-Rogers, H., Cagnard, N., Kierdorf, K., Prinz, M., Wu, B., Jacobsen, S. E. W., Pollard, J. W., Frampton, J., Liu, K. J. & Geissmann, F. (2012). A lineage of myeloid cells independent of MYB and hematopoietic stem cells. *Science*, 335(6077), 86–90.

Sezutsu, H., le Goff, G. & Feyereisen, R. (2013). Origins of P450 diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368, 1612.

Shields, A. R., Spence, A. C., Yamashita, Y. M., Davies, E. L. & Fuller, M. T. (2014). The actin-binding protein profilin is required for germline stem cell maintenance and germ cell enclosure by somatic cyst cells. *Development (Cambridge, England)*, 141(1), 73–82.

Shih, J. D. & Hunter, C. P. (2011). SID-1 Is a dsRNA-selective dsRNA-gated channel. *RNA*, 1057–1065.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V, Zdobnov, E. M., Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. Retrieved from https://doi.org/10.1093/bioinformatics/btv351

Singh, S., Gupta, M., Pandher, S., Kaur, G., Goel, N. Rathore, P & Palli, S.R. (2019a). RNA Sequencing , selection of reference genes and demonstration of feeding RNAi in *Thrips tabaci* ( Lind .) ( Thysanoptera : Thripidae ). *BMC Molecular Biology*, 1–21. Retrieved from https://doi.org/10.1186/s12867-019-0123-1

Singh, S., Gupta, M., Pandher, S., Kaur, G., Goel, N. & Rathore, P. (2019b). Using de novo transcriptome assembly and analysis to study RNAi in *Phenacoccus Solenopsis* Tinsley (Hemiptera: Pseudococcidae). *Scientific Reports*, 9(1), 13710.
Singh, S., Gupta, M., Pandher, S., Kaur, G., Rathore, P. & Palli, S. R. S. R. (2018). Selection of housekeeping genes and demonstration of RNAi in cotton leafhopper, *Amrasca biguttula biguttula* (Ishida). *PLoS ONE*, 13(1), 1–21.

Sommuck, T. (2001). Yield losses assessment due to pests on cotton lao PDR. *Kasetsart Journal of Natural Sciences*, 35(3), 51-63

Sorrentino, R. P., Gajewski, K. M. & Schulz, R. A. (2005). GATA Factors in *Drosophila* Heart and Blood Cell Development. *Seminars in cell & developmental biology*, 16(1), 107–116.

St Johnston, D., Beuchle, D. & Nüsslein-Volhard, C. (1991). Staufen, a gene required to localize maternal RNAs in the *Drosophila* Egg. *Cell*, 66(1) 51-63.

Tang, S., Lomsadze, A. & Borodovsky, M. (2015). Identification of protein coding regions in RNA Transcripts. *Nucleic Acids Research*, 43(12), e78. Retrieved from http://dx.doi.org/10.1093/nar/gkv227

Tassone, E. E., Cowden, C. C. & Castle, S. J. (2017). *De novo* transcriptome assemblies of four xylem sap-feeding insects. *GigaScience*, 6, 1–4.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. & Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to primer3. *Nucleic Acids Research*, 35(SUPPL.2).

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V & Zdobnov, E. M. (2017). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, 35(April), 543–548. Retrieved from http://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msx319/4705839

Yoon, J.-S., Mogilicherla, K., Gurusamy, D., Chen, X., Chereddy, S. C. R. R. & Palli, S. R. (2018). Double-stranded RNA Binding Protein, Staufen, is required for the initiation of rnai in coleopteran insects. *Proceedings of the National Academy of Sciences*, 115(33) 8334-8339.

Yoon, J. S., Shukla, J. N., Gong, Z. J., Mogilicherla, K. & Palli, S. R. (2016). RNA Interference in the colorado potato beetle, Leptinotarsa Decemlineata: Identification of key contributors.
Zerbino, D. R. & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de bruijn graphs. Genome Research, 18(5), 821–829.

Zhou, X., Liao, Z., Jia, Q., Cheng, L. & Li, F. (2007). Identification and characterization of piwi subfamily in insects. Biochemical and Biophysical Research Communications 362, 362, 126–131.

Zhu, F., Moural, T. W., Shah, K. & Palli, S. R. (2013). Integrated Analysis of cytochrome p450 gene superfamily in the red flour beetle, Tribolium castaneum. BMC Genomics, 14(1).

Zhu, F., Parthasarathy, R., Bai, H., Woithe, K., Kaussmann, M., Nauen, R., Harrison, D. A. & Palli, S. R. (2010). A brain-specific cytochrome p450 responsible for the majority of deltamethrin resistance in the qtc279 strain of Tribolium castaneum. Proceedings of the National Academy of Sciences of the United States of America, 107(19) 8557-8562

Table 1: Summary statistics of A. biguttula biguttula transcriptome

| Parameters            | Statistics            |
|-----------------------|-----------------------|
| No. of Reads          | 28,400,988            |
| Number of bases       | 8,376,627,673         |
| Total data in Gb      | 8.3 Gb                |
| No. of Transcripts    | 75,551                |
| Total transcript length(bases) | 125,532,225       |
| N50                   | 3,080                 |
| Maximum transcript length | 26,458               |
| Minimum transcript length | 200                  |
| Mean transcript length | 1,661                 |
| No. of CDS            | 39,613                |
| Total CDS length(bases) | 54,969,138            |
| Maximum CDS length    | 25,074                |
| Minimum CDS length | 297 |
|--------------------|-----|
| Mean CDS length    | 1,387 |

**Range of transcripts**

| 200 ≤ transcript< 500 | 25,045 |
|------------------------|--------|
| 500 ≤ transcript< 1000 | 13,046 |
| 1000 ≤ transcript< 2000 | 15,161 |
| 2000 ≤ transcript< 3000 | 8,924 |
| 3000 ≤ transcript< 4000 | 5,474 |
| 4000 ≤ transcript< 5000 | 3,406 |
| transcript ≥ 5000      | 4,495 |

**Range of CDS**

| 200 ≤ CDS < 500     | 7916 |
|---------------------|------|
| 500 ≤ CDS < 1000    | 11480 |
| 1000 ≤ CDS < 2000   | 12608 |
| 2000 ≤ CDS < 3000   | 4265 |
| 3000 ≤ CDS < 4000   | 1726 |
| 4000 ≤ CDS < 5000   | 769 |
| CDS ≥ 5000          | 849 |

Table 2 BUSCO statistics for *A. biguttula biguttula* transcriptome assembly against Insecta and Arthropoda taxa

| BUSCO Notation                                      | Insecta | Arthropoda |
|-----------------------------------------------------|---------|------------|
| Complete BUSCOs (C)                                 | 1519    | 1007       |
| Complete and single-copy BUSCOs (S)                  | 1133    | 756        |
| Complete and duplicated BUSCOs (D)                   | 386     | 251        |
| Fragmented BUSCOs (F)                                | 104     | 47         |
| Missing BUSCOs (M)                                   | 35      | 12         |
| Total BUSCO groups searched                          | 1658    | 1066       |

Table 3 The cluster analysis of the CYP

| Species | Proteins | Clusters | Singletons |
|---------|----------|----------|------------|
| AF      | 162      | 88       | 41         |
| AM      | 227      | 105      | 65         |
| EN      | 306      | 130      | 89         |
| LN      | 265      | 126      | 76         |

Figure legends
Fig. 1 A. Top hits of *Amrasca biguttula biguttula* Illumina sequences across different insects and other species on the basis of highest score in BLASTx using homology searches against nr database B. Annotation distribution of 39,613 predicted CDS of transcriptome of *A. biguttula biguttula*

Fig. 2 A. Classification of *A. biguttula biguttula* transcripts based on predicted Gene Ontology terms via WEGO plot associated with Cellular component, Molecular function and Biological process. B. GO (Gene Ontology) distribution of predicted CDS of Biological processes, Molecular functions and Cellular component in *A. biguttula biguttula*

Fig. 3: KEGG pathway classification of predicted CDS in *de novo* transcriptome of *A. biguttula biguttula*

*Represents overall count for the following pathways: Carbon metabolism, 2-Oxocarboxylic acid metabolism, Fatty acid metabolism, Biosynthesis of amino acids, Degradation of aromatic compounds.

Fig. 4 Expression level of core RNAi genes among different stages of *A. biguttula biguttula* quantified using Real-time quantitative PCR. Error bar represents standard deviation (n=15) with 3 replicates each. Bars with similar alphabets are statistically par with each other (P ≤ 0.05, Student’s t-test).

Fig. 5 Phylogenetic analysis of core RNAi genes of *A. biguttula biguttula* with other insect species using neighbour joining method of in MEGAX at 500 bootstrap value.

Fig. 6 RNAi or RNAi genes: A. Ago-2 knockdown (59.3%) and SNF7 expression levels after Ago-2 knockdown. B. Dicer-2 knockdown (61.6%) and SNF7 expression levels after Dicer-2 downregulation C. Staufen knockdown (95.8%) and SNF7 expression levels D.Piwi knockdown (99.01%) and SNF7 expression levels after its knockdown. The error bars represent the standard deviation (n=10) with 3 replicates each and * represents significant differences in mRNA transcripts compared to the control (P ≤ 0.05, Student’s t-test).

Ago-2: Argonaute-2; SNF7: SWItch/Sucrose Non-Fermentable7

Fig. 7 The Cytochrome P450 (CYP450) genes compared in by orthoMCL based OrthoVenn tool indicating 38 major common transcripts in five transcriptomes

*AM- Adult male, AF- Adult Female, EN- Early nymph, LN- Late nymph, RM- Our lab’s *Amrasca biguttula biguttula* transcriptome

Fig. 8 Phylogenetic analysis of *A. biguttula biguttula* cytochrome P450s using Geneious v11.1.4 using representative sequences (coloured text) from other insect species. Sequences in violet colour belong to mitochondrial CYPs, red colored to CYP2 clan, yellow coloured are from CYP4 clan and remaining orange colored transcripts fall under CYP3 clan.

Fig. 9 Differential expression of mRNA transcripts of selected CYP450s genes in two differentially insecticide exposed late nymphal populations of *Amrasca biguttula biguttula* using Real Time- PCR. The error bars represent the standard deviation (n=15) with 3 replicates each.

Fig. 10 Gene Ontology and Pathway analysis of CYP450 genes of *Amrasca biguttula biguttula* using BLAST2Go (A) and KEGG(B.)
