ILLUMINA® SEQUENCING OF GREEN STINK BUG NYMPH AND ADULT cDNA TO IDENTIFY POTENTIAL RNAi GENE TARGETS

J. B. van Kretschmar
K. V. Donohue
A. R. Cabrera
L. C. Magalhaes
C. E. Sorenson
J. S. Bacheler
S. M. S. Khalil
R. M. Roe
North Carolina State University, Department of Entomology
Raleigh, NC

Abstract

Whole-body transcriptomes for nymphs and adults of the green stink bug, *Acrosternum hilare* (Say), were sequenced on an Illumina® Genome Analyzer IIx sequencer. The insects were collected from sites in North Carolina and Virginia, USA. The cDNA library for each sample was sequenced on one lane of an eight-lane flow cell. Total read count was 25,211,411 and 23,601,951 for the nymph and adult libraries, respectively. Mean read length was 124 bps (base pairs) for nymph and 125 bps for adult. Total sequencing data were 3,127,036,121 bps for nymph cDNA and 2,956,794,438 bps for adult. Trinity® software was used to assemble the nymph and adult reads into 79,062 and 74,271 contigs (contiguous sequences), respectively. Blast2GO® software was used to align, map, and annotate the contigs. For the alignment step, the contigs were translated to peptides and compared to the GenBank nr (non-redundant) protein database using the BLASTx (Basic Local Alignment Search Tool) algorithm with E-value cut-off set at E-3 (10^{-3}). The average length of BLASTx translated sequences was 1248 (range = 201 – 7968) amino acids for the nymph and 1320 (range = 201 – 7957) for the adult. Several nymph and adult sequences showed high homology (E-value ≤ 1.5 E-13) with database hormones, neuropeptides, neurotransmitters and/or their receptors. Seventeen nymph and adult sequences were identical (E-value = 0) to *A. pisum, T. castaneum, B. germaninca, and P. humanus corporis* hormones. After alignment of green stink bug contigs with GenBank nr database sequences, all BLAST hits (E-value ≤ E-3) were mapped and annotated with GO (Gene Ontology) terms that assigned the translated sequences to categories of putative protein function. Mapping and annotation steps yielded 13,226 nymph and 14,725 adult sequences assigned to 13 GO functional categories: Catalysis, binding, transport, signal transduction, structural molecule activity, enzyme regulation, electron carrier activity, antioxidant activity, metallochaperone, channel regulation, protein tag, translation regulation and nutrient reservoir. The goal of sequencing and of on-going bioinformatic analysis of these transcriptomes is to identify candidate RNAi gene targets with potential to be developed as insecticidal transgenes for control of stink bugs and other hemipteran pests of cotton. The resulting data can also be applied to other protein function and genetics studies.

Introduction

A consequence of the low-chemical-spray environment that has resulted with the success of the boll weevil eradication program and the widespread adoption of Bt cotton has been emergence of plant bugs (Hemiptera: Miridae) and stink bugs (Hemiptera: Pentatomidae) to become major pests of cotton (Leonard, 2008). Insecticidal transgenes in current strains of genetically modified cotton (Bt cotton) are effective in controlling Lepidoptera but not hemipteran pests. The prospect of developing transgenes insecticidal to plant bugs and stink bugs may lie in manipulating the RNAi (RNA interference) mechanism to suppress or silence critical-functional genes of the pest species (Gordon & Waterhouse, 2007). RNAi disrupts gene expression/protein synthesis in the cell cytoplasm of target organisms at the gene transcript (messenger RNA (mRNA)) level (Huvenne and Smagghe, 2010). The agent of RNAi is dsRNA (double stranded RNA). The mechanism by which this disruption occurs is the cleavage of introduced dsRNA into siRNAs (short-interfering RNAs) which are complementary in their nucleotide sequence to the targeted insect mRNA. These complementary siRNAs interact with the target mRNA and prevent it from being translated into proteins that are essential to the physiology of the pest organism. An alternative approach is the use of siRNA directly.

Several labs have reported using RNAi to cause varying levels of gene suppression in targeted insects fed or injected with dsRNA or siRNA. Walker and Allen (2010) suppressed expression of a salivary enzyme in tarnished plant bug, *Lygus lineolaris* (Palisot de Beauvois) (Hemiptera: Miridae); Chen et al. (2010) targeted trehalose phosphate synthase in brown planthopper, *Nilaparvata lugens* (Hemiptera: Delphacidae); and Rosa et al. (2010) decreased levels of actin mRNA in the glassy-winged sharpshooter, *Homalodisca vitripennis* (Hemiptera: Cicadellidae).

Working with Lepidoptera, Griebler et al. (2008) targeted JH (juvenile hormone) regulating neuropeptides in fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae); Kumar et al. (2009) suppressed acetylcholinesterase in the cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae); and, Tian et al. (2009) reduced levels of chitin synthase in the beet armyworm, *S. exigua.*

The efficacy of transforming plants to express dsRNAs for pest control has been demonstrated with two non-hemipterans. Baum et al. (2007) reported that corn, *Zea mays*, had been transformed with transgenes expressing dsRNA that reduced expression of a target gene in the western corn rootworm, *Diabrotica virgifera virgifera* (Coleoptera: Chrysomelidae). Mao et al. (2011) reported reduced weights of *H. armigera* larvae fed leaves of cotton transformed to express dsRNA targeting a P450 monooxygenase that allows larvae to feed on gossypol, a phenolic cotton secondary metabolite with insect antifeedant activity.

A first step in the development of an RNAi-based insecticidal transgene is to identify potential genes of the pest species that can be targeted with dsRNA. The objective of the work reported here was to use Illumina® sequencing technology to sequence the transcriptomes of green stink bug nymphs and adults, and then use bioinformatics to analyze the resulting sequences to find genes specific and critical to stink bug physiology. There are a number of uses for these data, e.g., for monitoring the evolution of insecticide resistance, discovery of RNA viruses and basic studies of insect function and population structure.

Materials and Methods

Insects

Green stink bug adults and nymphs, *Acrosternum hilare* (Say), were collected with a sweep net from soybeans at sites in Virginia and North Carolina. Adults and nymphs were kept in 1-gallon plastic tubs covered with one layer of cheesecloth, fed on soybeans and artificial diet (Cohen, 2000) and held in a growth chamber (Percival Scientific Model I-66NL; Percival Scientific, Inc., Perry, IA) at 27 ± 1° C, 65% relative humidity, 14 hours light: 10 hours dark until needed for RNA extraction.

Total RNA isolation

Whole live green stink bug adults (5 insects) or nymphs (10 insects) were placed separately in 17 mm x 100 mm culture tubes (Fisher Scientific) and homogenized in 3 mls of TRI Reagent® (Sigma-Aldrich, St. Louis, MO) according to the manufacturer's recommendations to extract total RNA. Whole-insect homogenization was done using a Brinkmann® Polytron. RNA pellets were rehydrated in 100 mM aurintricarboxylic acid to prevent degradation (Hallick et al., 1977).

cDNA library preparation and Illumina® sequencing

From each whole-body sample (nymph, adult), a cDNA library was prepared from ≥ 5 µg total RNA following vendor recommendations (Illumina, Inc., San Diego, CA) for sequencing on the Illumina® Genome Analyzer IIx sequencer. Each sample was sequenced on one lane of an Illumina eight-lane flow cell. Trinity® software (20 August 2011 release; www.trinityrnaseq.sourceforge.net) was used to assemble the nymph and adult reads into contigs (contiguous nucleotide sequences); k-mer length was 25.

Bioinformatics

Blast2GO® software (Conesa et al., 2005) was used to align, map, and annotate the contigs. Blast2GO analysis was concluded 13 December 2011 for adult sequences and 27 December 2011 for nymphs. For the alignment step, the contigs were translated to proteins in all six reading frames and compared to the GenBank nr (non-redundant) protein database using the BLASTx (Basic Local Alignment Search Tool) algorithm with E-value cut-off set at E-3 (10^{-3}). BLAST hits (green stink bug query contigs with database-sequence matches where E-value ≤ E-3) were

mapped and annotated with GO (Gene Ontology) terms. These GO terms assigned the translated query sequences to categories of putative protein function (GO level 2 functional categories) on the basis of sequence and functional conservation among organisms represented in publicly-accessible protein/gene-product sequence databases (Gene Ontology Consortium, Ashburner et al., 2000).

Results and Discussion

Total RNA isolated

From nymphs and adults, 106.6 µg and 121.2 µg total RNA, respectively, were prepared and submitted to the North Carolina State University Genomics Sciences Laboratory 14 June 2011 for sequencing.

Sequencing of reads and assembly of contigs

Sequencing and assembly were completed 28 October 2011. Results are presented in Table 1. Total read count was 25,211,411 and 23,601,951 for the nymph and adult cDNA libraries, respectively. Average read length was 124 bps (base pairs) for nymph and 125 bps for adult. Total number of bps of data obtained was 3,127,036,121 for nymphs and 2,956,794,438 for adults. Nymph and adult reads were assembled into 79,062 and 74,271 contigs, respectively.

Table 1. Results for green stink bug nymph and adult whole-body cDNA libraries sequenced on an Illumina® Gene Analyzer IIx. Sequencing reads were assembled into contigs with Trinity® software.

	Nymph	Adult
Total read count	25,211,411	23,601,951
Mean read length (base pairs)	124	125
Total number of base pairs	3,127,036,121	2,956,794,438
Number of contigs	79,062	74,271

Bioinformatics (alignment of green stink bug contigs with database sequences)

Blast2GO® analysis of adult contigs were concluded on 13 December 2011 and of nymph contigs, 27 December 2011. Results of using the BLASTx algorithm to translate and align our green stink bug contigs with sequences available (up to 27 December 2011) in the GenBank nr database are summarized in Table 2. There were 27,650 nymph contigs and 30,147 adult sequences with BLAST hits (E-value ≤ E-3). The average length of translated sequences was 1248 (range = 201 – 7968) amino acids for nymph contigs and 1320 (range = 201 – 7957) for the adult.

Table 2. Results of using Blast2GO® software to align BLASTx-translated green stink bug nymph and adult contigs with GenBank nr (non-redundant) database sequences.

	nymph	adult
Number of contigs	79,062	74,271
Number of contigs with BLAST hits (database match E-value ≤ E-3)	27,650	30,147
Range in translated sequence length (amino acid residues)	201 – 7968	201 – 7957
Average translated sequence length (amino acid residues)	1248	1320

The number of top BLAST hits (E-value ≤ 1.5 E-13) with homology to GenBank nr database hormones, neuropeptides and neurotransmitters is presented in Table 3. These are of interest as targets because of their critical role in regulation and coordination of insect physiology.

Table 3. Numbers of green stink bug translated contig top BLAST hits (E-value ≤ 1.5 E-13) homologous to reference database (GenBank non-redundant) hormones, neuropeptides, and neurotransmitters.

	nymph	adult
Hormones	131	126
Neuropeptides	10	15
Neurotransmitters	8	10

Among the contigs with homology to database hormones or hormone receptors, 8 nymph and 9 adult contigs were identical (E-value = 0) to hormones of species presented in Table 4. These include 3 green stink bug sequences identical to pea aphid, *Acyrthosiphon pisum* (Hemiptera: Aphididae), hormones or hormone receptors. Among these is a nuclear hormone receptor for FTZ-F1, a DNA transcription factor associated with embryogenesis.

Table 4. Numbers of green stink bug nymph and adult contig BLAST hits identical (E-value = 0) to hormones or hormone receptors of database (GenBank non-redundant) reference species.

Reference species	nymph	adult
Tribolium castaneum	4	6
Acyrthosiphon pisum	2	1
Blatella germanica	1	2
Pediculus humanus corporis	1	0

Numbers of green stink bug contigs with good homology (E-value ≤ 5.55 E-13) to database neuropeptides or neuropeptide receptors are presented in Table 5. These include 10 nymph and 11 adult sequences with good homology to pea aphid and to *Rhodnius prolixus* (Hemiptera: Reduviidae) neuropeptides. Among these are somatostatin, capa, neuropeptide Y, bombesin, neuromedin, and allatostatin-A.

Table 5. Numbers of green stink bug contig top BLAST hits homologous to reference database neuropeptides.

Reference species	nymph	adult
	(E-value ≤ 5.55 E-13)	(E-value ≤ 1.5 E-13)
Acyrthosiphon pisum	7	6
Rhodnius prolixus	3	5
Harpegnathos saltator	0	1
Bombus terrestris	0	1
Nasonia vitripennis	0	1
Apis mellifera	0	1

Numbers of green stink bug contigs with good homology (E-value ≤ 1.25 E-16) to reference database neurotransmitters and/or neurotransmitter receptors are presented in Table 6. These include 1 nymph and 3 adult sequences with good homology to pea aphid and/or *Triatoma rubida* (Hemiptera: Reduviidae) neurotransmitters, including a GABA-receptor subunit.

Table 6. Numbers of green stink bug contig top BLAST hits homologous to reference database neurotransmitters.

	nymph (E-value ≤ 1.25 E-16)	adult (E-value ≤ 2.65 E-66)
Tribolium castaneum	3	3
Acyrthosiphon pisum	1	2
Blatella germanica		2
Harpegnathos saltator	2	1
Bombus terrestris		1
Aedes aegypti	1	2
Camponotus floridanus	1	
Triatoma rubida		1

Bioinformatics (mapping and annotation of green stink bug contigs with putative biological function)

Blast2GO® mapping and annotation steps resulted in 13,226 nymph and 14,725 adult sequences being assigned to GO categories of putative protein function (Table 7). More than 97% of both nymph and adult contigs were associated with catalysis, binding, transport, signal transduction, and structural molecule activity. Contigs associated with enzyme regulation, electron carrier activity, and antioxidant activity accounted for approximately 2.6% of both nymph and adult contigs. The remaining 0.05% of nymph contigs and 0.06% of adult contigs were assigned to the remaining categories, including channel regulation and translation regulation (Table 7).

Table 7. Numbers of green stink bug translated contig BLAST hits assigned to GO level 2 functional categories.

Category	GSB nymph contigs	GSB adult contigs
catalysis	5837	6524
binding	5689	6334
transport	725	798
signal transduction	331	372
structural molecule activity	290	312
enzyme regulation	214	226
electron carrier activity	99	108
antioxidant activity	34	42
metallochaperone	2	4
channel regulation	2	2
protein tag	1	2
translation regulation	1	1
nutrient reservoir	1	0

The hope of developing an insecticidal transgene for cotton that is as effective in controlling hemipteran pests as Bt cotton transgenes are in controlling lepidopteran pests may lie in manipulating the pest RNAi mechanism to suppress or silence critical-functioning pest genes. The first step in developing an RNAi-based insecticidal transgene is to identify genes essential and specific to the physiology of the targeted pest. Our use of Illumina® technology to sequence green stink bug nymph and adult transcriptomes produced more than 153,000 sequences for bioinformatic analysis. Our on-going bioinformatics effort has already identified green stink bug sequences associated with a broad range of biological function. These sequences include a significant number with good homology to reference-database hemipteran hormones, neuropeptides, and neurotransmitters critical to endocrine regulation and physiological coordination. The work also shows the power of the use of short reads as opposed to the longer reads of 454 in the characterization of a whole body insect transcriptome, with the obvious advantage of a lower cost and the ability to identify small and low abundant proteins like neuropeptides.

Summary

Transcriptomes for nymphs and adults of green stink bug, *Acrosternum hilare* (Say)**,** were recently sequenced using Illumina® technology. In addition to identifying sequences associated with a broad range of biological function, on-going bioinformatic analysis has thus far shown a significant number of both nymph and adult contigs with homology to reference-database hormones, neuropeptides, neurotransmitters and/or their receptors.

Acknowledgements

We gratefully acknowledge financial support from Cotton Inc. We thank Dr. Ames Herbert of Virginia Tech for providing insects, Dr. Jenn Schaff of the NCSU Genomics Sciences Laboratory for sequencing services, and Mr. Jonathan Keebler of the NCSU Genomics Sciences Laboratory for assembly of sequenced nucleotide reads.

References

Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A. P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin and G. Sherlock. 2000. Gene ontology: tool for the unification of biology. Nat. Genet., 25: 25–29.

Baum, J.A., T. Bogaert, W. Clinton, G.R. Heck, P. Feldmann, O. Ilagan, S. Johnson, G. Plaetinck, T. Munyikwa, M. Pleau, T. Vaughn and J. Roberts. 2007. Control of coleopteran insect pests through RNAi interference. Nat. Biotech. 25:1322-1326.

Chen, J., D. Zhang, Q. Yao, J. Zhang, X. Dong, H. Tian, J. Chen and W. Zhang. 2010. Feeding-based RNA interference of a trehalose phosphate synthase gene in the brown planthopper, *Nilaparvata lugens*. Insect Mol. Biol. 19:777-786.

Cohen, A.C. 2000. New oligidic production diet for *Lygus hesperus* Knight and *L. lineolaris* (Palisot de beauvois). J. Entomol. Sci. 35:301-310.

Conesa, A., S. Gotz, J.M. Garcia-Gomez, J. Terol, M. Talon and M. Robles. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676.

Gordon, K.H. and P.M. Waterhouse. 2007. RNAi for insect-proof plants. Nat. Biotech. 25:1231-1232

Griebler, M., S.A. Westerlund, K.H. Hoffmann and M. Meyering-Vos. 2008. RNA interference with the allatoregulating neuropeptide genes from the fall armyworm *Spodoptera frugiperda* and its effects on the JH titer in the hemolymph. J. Insect Phys. 54:997-1007.

Hallick, R.B., B.K. Chelm, P.W. Gray and E.M. Orozco, Jr. 1977. Use of aurintricarboxylic acid as an inhibitor of nucleases during nucleic acid isolation. Nucleic Acids Res. 4:3055–3064.

Huvenne, H. and G. Smagghe. 2010. Mechanisms of dsRNA uptake in insects and potential of RNAi for pest control:A review. J. Insect Phys. 56:227-235.

Kumar, M., G.P. Gupta and M.V. Rajam. 2009. Silencing of acetylcholinesterase gene of *Helicoverpa armigera* by siRNA affects larval growth and its life cycle. J. Insect Phys. 55:273-278.

Leonard, B.R. 2008. Management of the sucking bug complex across the Cotton Belt: Getting a handle on the problem!, p. 923. *In* Proc. Beltwide Cotton Conferences. National Cotton Council, Memphis, TN.

Mao, Y., X. Tao, X. Xue, L. Wang and X. Chen. 2011. Cotton plants expressing CYP6AE14 double-stranded RNA show enhanced resistance to bollworms. Transgenic Res. 20:665-673.

Rosa, C., S.G. Kamita, H. Dequine, H. Wuriyanghan, J.A. Lindbo and B. W. Falk. 2010. RNAi effects on actin mRNAs in *Homalodisca vitripennis* cells. J. RNAi Gene Silencing 6:361-366.

Tian, H., H. Peng, Q. Yao, H. Chen, Q. Xie, B. Tang and W. Zhang. 2009. Developmental control of a lepidopteran pest *Spodoptera exigua* by ingestion of bacteria expressing dsRNA of a non-midgut gene. PLoS ONE 4:e6225.

Walker W. and M.L. Allen. 2010. Expression and RNA interference of salivary polygalacturonase genes in the tarnished plant bug, *Lygus lineolaris.* J. Insect Sci. 10:173.