Complete Mitochondrial Genome of *Helicoverpa zea* (Lepidoptera: Noctuidae) and Expression Profiles of Mitochondrial-Encoded Genes in Early and Late Embryos

Omaththage P. Perera,1,2 Thomas K. Walsh,3 and Randall G. Luttrell1

1Southern Insect Management Research Unit, USDA-ARS, Stoneville, MS 38776 (op.perera@ars.usda.gov; tom.walsh@csiro.au; randy.luttrell@ars.usda.gov), 2Corresponding author, e-mail: op.perera@ars.usda.gov, and 3Land and Water Flagship, Commonwealth Scientific and Industrial Research Organization, Clunies Ross Street, GPO Box 1700, Canberra, ACT 2601, Australia

Subject Editor: Qili Feng

Received 19 November 2015; Accepted 6 March 2016

Abstract

The mitochondrial genome (mitogenome) of the bollworm, *Helicoverpa zea* (Boddie), was assembled using paired-end nucleotide sequence reads generated with a next-generation sequencing platform. Assembly resulted in a mitogenome of 15,348 bp with greater than 17,000-fold average coverage. Organization of the *H. zea* mitogenome (gene order and orientation) was identical to other known lepidopteran mitogenome sequences. Compared with *Helicoverpa armigera* (Hübner) mitogenome, there were a few differences in the lengths of gaps between genes, but the lengths of nucleotide overlaps were essentially conserved between the two species. Nucleotide composition of the *H. zea* mitochondrial genome was very similar to those of the related species *H. armigera* and *Helicoverpa punctigera* Wallengren. Mapping of RNA-Seq reads obtained from 2-h eggs and 48-h embryos to protein coding genes (PCG) revealed that all *H. zea* PCGs were processed as single mature gene transcripts except for the bicistronic atp8 + atp6 transcript. A tRNA-like sequence predicted to form a hammer-head-like secondary structure that may play a role in transcription start and mitogenome replication was identified within the control region of the *H. zea* mitogenome. Similar structures were also found within the control regions of several other lepidopteran species. Expression analysis revealed significant differences in levels of expression of PCGs within each developmental stage, but the pattern of variation was similar in both developmental stages analyzed in this study. Mapping of RNA-Seq reads to PCG transcripts also identified transcription termination and polyadenylation sites that differed from the sites described in other lepidopteran species.

Key words: bollworm, mtDNA, transcript processing

*Helicoverpa zea* (Boddie), commonly known as the bollworm or corn earworm, is a polyphagous insect in the family Noctuidae and the only species of *Helicoverpa* in the subfamily Heliothentinae (Hardwick 1996) in North America. Its geographic range includes most temperate and tropical regions of the Americas with the exception of Alaska and northern Canada (Metcalf and Flint 1962, Hardwick 1965) and overlaps with the range of *Helicoverpa gelotopoec* Dyar in South America. Both have very similar host ranges, and morphology of the larval stages is similar. Corn earworm damages a large number of agricultural crops and estimated annual losses in the United States alone exceed one billion dollars (Light et al. 1993). Control costs for this pest are a significant component of crop production costs in the United States and insecticides are routinely used for suppression of *H. zea* in cotton (Williams 2014) and soybean (Musser et al. 2013). *Helicoverpa zea* is one of the primary targets of transgenic corn hybrids and cotton varieties that express insecticidal proteins derived from *Bacillus thuringiensis* Berliner (Luttrell and Jackson 2012). Of the 10 other species of Heliothentinae (Hardwick 1996) in North America, *Heliothis virescens* (Fabricius) is the only other significant pest of agricultural crops.

Both *H. zea* and *Helio. virescens* are polyphagous and often attack the same crops at the same time. Eggs and larvae are almost indistinguishable in the field. Morphological traits can be used to identify both species, but the process is labor intensive and requires inspection under magnification. Diagnostic tools for larvae based on monoclonal antibodies and genetic differences are available and have been used on a limited scale to characterize mixed populations in field situations (Zeng et al. 1999, 1998). An expanding base of genetic knowledge and genetic markers (Grasela and McIntosh...
The recent introduction of *Helicoverpa armigera* (Hubner) into Brazil (Czechak et al. 2013, Tay et al. 2013), further spread into Argentina, Paraguay, and Uruguay (Muria et al. 2014), and potential future movement into North America further complicates identification of the Heliothinae pest species. The establishment of *H. armigera* in the United States represents a significant risk of increased economic costs to agriculture, both in control costs and in direct crop losses. Inter-specific hybridizations between heliothine species that yield fertile progeny have been documented. Partial sterility was observed between *Heliothis virescens* and *Heliothis subflexa* Guenée (Laster and Hardee 1995), *H. armigera* and *Helicoverpa assulta* (Guenée) (Wang 2007), and *H. zea* and *H. assulta* (Guenée) (Degragillier and Newman 1993). However, no evidence of sterility or incompatibility was identified between *H. armigera* and *H. zea* (Laster and Harde 1995, Laster and Sheng 1995). Any hybridization between invasive *H. armigera* and the two *Helicoverpa* species native to the Americas resulting in nuclear gene introgression would further complicate taxonomic identification based on morphology as well as nuclear genes. Mitogenome analysis facilitates the positive identification of the maternal lineage of any suspect hybrid and can be compared with nuclear markers.

Insect mitochondrial (mt) genomes tend to range from 14 to 18 kb in size and usually possess 37 genes; 13 protein coding genes (PCGs), 22 tRNAs, 2 rRNA genes, and several noncoding regions (Boore 1999). Mitochondrial DNA (mtDNA) sequences have been the tool of choice in a large number of studies on population genetics and phylogenetics of insects (Castro et al. 1998, Catterino et al. 2000). This is largely due to the highly conserved nature of certain regions that have allowed the development of universal primers. However, these highly conserved regions of the mitochondrial genomes are not ideal for intraspecific genetic studies as they may underestimate actual genetic diversity. The use of entire mitogenome sequences is increasing in intraspecies genetic studies as they can provide greater resolution (Yu et al. 2007, Teacher et al. 2012, Pabijan et al. 2013).

A large number of lepidopteran mitogenomes have been sequenced to date (>100 NCBI) from diverse taxonomic groups. Missing from many of the sequenced mitogenomes is information on expression of the genes and confirmation of predicted annotations using expressed gene transcripts. In most vertebrate and invertebrate mitochondrial genomes there is little or no intergenic space and PCGs are interspersed by tRNA genes, the usual transcript cleavage sites (Ojala et al. 1980). However, the mitochondrial genomes can also produce polycistronic protein coding RNA transcripts where there is overlap within the genome sequence. This has been identified in a number of insect species including *Maruca vitrata* (Fabricius) and *Drosophila melanogaster* Meigen but with some differences in the genes identified as polycistronic (Stewart and Beckenbach 2009, Margam et al. 2011). Analysis of transcriptome data can help accurately annotate these regions and identity genes that are polycistronic and those that are not. Mapping of mitochondrial transcripts, expression mechanisms, and gene expression profiles have been studied in various animal species (Costanzo and Fox 1990, Stewart and Beckenbach 2009, Torres et al. 2009, Wang et al. 2013, Marková et al. 2015), but so far only one study has examined the transcript processing in a moth species (Margam et al. 2011). A reliable means of identifying samples to species and performing intragenetic studies would be useful to properly understand the distribution, ecology, agricultural impacts, and the insecticide resistance status in these heliothine pest species. The full-length mitogenome of *H. zea* is a useful tool in developing markers and tests suitable for species-specific and intraspecific studies. In this work, we determined the nucleotide sequences of the complete mitogenome of the moth, *H. zea* and used RNA-Seq data to study processing of polycistronic mitochondrial transcripts.

### Materials and Methods

#### DNA Extraction and Nucleotide Sequencing

A thorax of a single adult female *H. zea* from a laboratory colony maintained at the USDA-ARS Southern Insect Management Research Unit, Stoneville, MS was used in mtDNA extraction. Tissue was homogenized in a buffer containing 90 mM KCl, 55 mM CaCl₂, 30 mM NaCl, 15 mM MgSO₄, and 250 mM sucrose and filtered using a nylon mesh cloth. The filtrate was spun at 200 × g for 25 min. to pellet nuclei and large tissue debris. Supernatant enriched with mitochondria was carefully aspirated and transferred to a new 1.5-ml microcentrifuge tube. The supernatant was centrifuged at 5,000 × g to pellet mitochondria. DNA was extracted from mitochondria using MasterPure DNA extraction kit (Epicentre, Charlotte, NC) following manufacturer’s instructions. Nextera XT (Illumina, San Diego, CA) was used to prepare a small insert sequencing library from the mtDNA enriched nucleic acid and 100 bp paired end reads were generated on a MiSeq (Illumina) at the USDA-ARS Genomics and Bioinformatics Research Unit.

A de novo assembly of the paired-end miSeq reads was performed using SeqMan NGen 12.0 (DNAStar, Madison, WI). Assembly parameters minimum match percentage, match size, match spacing, mismatch penalty, gap penalty, and maximum gap length were set to 93, 50, 10, 20, 30, and 6%, respectively. Consensus sequence was exported and ends were manually edited to remove duplicated nucleotides.

#### Annotation of the Mitogenome

tRNA genes were identified using the tRNAscan-SE Search Server (Lowe and Eddy 1997, Schattner et al. 2005). PCGs, rRNA genes, and the tRNAs that were not identified by the tRNAscan-SE program were annotated by comparing with published mt genomes of *H. armigera* and *Helicoverpa punctigera* Wallengren (Yin et al. 2010, Walsh 2014). Nucleotide compositional skew for the whole genome was calculated using the formula GC-skew = (G – C)/((G + C) and AT-skew = (A – T)/(A + T), where A, C, G, and T are the total number of nucleotides represented by the respective IUPAC letter code in the nucleotide sequence (Perna and Kocher 1995).

#### RNA-Seq Library Preparation

Freshly emerged females and males (*n* = 50) from a laboratory colony were placed in a 4-liter plastic bucket and allowed to mate for 3 d. Eggs were collected on the fourth day by placing a sheet of wax paper over the open end of the container for 1 h. Twenty-five eggs were dislodged from the wax paper using a small brush at the end of the second hour (2-h eggs) and total RNA was extracted using TriZol reagent (Invitrogen, Carlsbad, CA) following manufacturer’s instructions. The remainder of the eggs was incubated at 27°C for a total of 48 h. At the end of the incubation period, 25 eggs containing developed embryos with dark head capsules (48-h embryos) were selected using a microscope and dislodged from the wax paper. TriZol reagent was used to extract total RNA from these eggs. Freshly extracted total RNA was treated with 1 unit of RNase-free DNase per 2 μg of total RNA. Double-stranded cDNA was synthesized from 2 μg of
total RNA using SuperScript III cDNA synthesis reagents (Invitrogen) following manufacturer’s instructions. First strand cDNA synthesis was primed with a custom-designed oligo d(T) primer (5'-GCTTACAAAGCAGATTTTTTTTTTTTTTTTTTTTTTTTTTTTG-3'). After the second strand synthesis, cDNA was purified and concentrated and Illumina sequencing libraries were prepared with 50 ng of double-stranded cDNA using Nextera library construction reagents. Two replicate libraries each were prepared for 2-h eggs and 48-h embryos. Library quality was evaluated by running 1 µl of each library on an Agilent 2200 Tape Station (Agilent Technologies, Santa Clara, CA). The libraries were size fractionated to select 400–500 bp fragments using the Pipin Prep instrument (Sage Science, Inc., Beverly, MA). One-hundred base pair single-end reads were generated using the Illumina HiSeq 2500 instrument at the USDA-ARS Genomics and Bioinformatics Research Unit.

Transcript and Expression Analysis

Complete nucleotide sequences of 13 PCGs were extracted from *H. zea* mitogenome and used in calculating synonymous and nonsynonymous substitution rates compared with orthologous mitochondrial PCGs of *H. armigera*. The number of nonsynonymous substitutions per nonsynonymous site (Ks) and the number of synonymous substitutions per synonymous site (Ks) between the two *Helicoverpa* species were estimated using the software KaKs Calculator (Zhang et al. 2006) with the maximum-likelihood model averaging method (Posada 2003).

Mitochondrial PCG sequences were also used for mapping RNA-Seq reads generated from 2-h eggs and 48-h embryos. SeqMan NGen 4.0 assembly parameters used in this reference assembly were identical to those used in the reference assembly of the mitogenome. Reads mapping to each gene were manually examined to verify ends of the genes and the presence or absence of polyadenylation. Coding sequences of the 13 PCG were also used in expression profiling using ArrayStar 5.0 software (DNAStar). Reads per kilobase per million was used to normalize the expression levels of the PCG. Normalized expression values and the standard deviations of the PCG were exported from ArrayStar 5.0 software and each expression value was divided by the value of the gene with the lowest expression level (nad4L). Covariance (CV) of the quotient was calculated by the formula CV = √(SD1² + SD2²), where SD1 and SD2 are the standard deviations of the two expression values used to calculate the relative expression level. Standard deviations of the relative expression values were calculated by multiplying the relative expression value by the CV of each relative expression value.

In addition, all Illumina sequence reads from 2-h eggs and 48-h embryos were mapped separately to full-length sequence using CLC Genome Workbench 8.5 (Qiagen, Redwood City, CA) to obtain sequence read density profiles across the complete mitogenome. Mapped reads were analyzed for sequence polymorphisms using a basic variant detection tool implemented in CLC Genome Workbench.

Results

Nucleotide Composition and Genome Organization

The mtDNA enriched *H. zea* sequencing library yielded 2,226,753 sequence read pairs with Phred quality values ≥35. Nucleotide sequence reads were deposited in the NCBI SRA database (Submission no. SRX690295). De novo assembly of paired-end miSeq reads yielded a contig of 15,984 bp with 17,173-fold average coverage from a total of 1,447,494 read pairs consistent within the contig.

Manual editing of duplicated sequences at the ends resulted in a final *H. zea* mitogenome of 15,348 bp (Accession no. KJ930516). The mitogenome of *H. zea* is 81.00% A+T, similar to 80.97% in *H. armigera* and 81.35% in *H. punctigera* (Yin et al. 2010, Walsh 2014). The GC-skew and the AT-skew in the complete mitogenome of *H. zea* were −0.2032 and 0.0024, respectively, indicating that this mitogenome contains more C than G and more A than T. This nucleotide skew pattern is similar to that of closely related species *H. armigera* and *H. punctigera*. Overall nucleotide identity between *H. zea* mitogenome and those from *H. armigera* and *H. punctigera* were 97.4 and 94.6%, respectively.

Annotation of the mt genome of *H. zea* identified 13 PCG, 2 rRNA genes, 22 tRNA genes, and a control region with high A+T content and repetitive sequences. The gene order and the orientation of the genes in *H. zea* mitogenome were identical to those of *H. armigera* and other lepidopteran species. Organization of the genes in *H. zea* mitogenome is shown in Fig. 1 and the coordinates of the genes and control region are given in Table 1. Although there are a few differences in the lengths of space (i.e., number of nucleotides) between some of the mitochondrial genes, the lengths of overlaps between trnI and trnQ, trnW and trnC, atp8 and atp6, atp6 and cox3, trnE and trnF are conserved between *H. armigera* and *H. zea* (Table 1). One exception noted in this study is the processing of nad3 and trnA genes. Although there is a native in-frame TAG stop codon (nucleotides 5923–5925) in the nad3 gene, the stop codon in all processed transcripts of (i.e., all polyadenylated transcripts) found in this study contained a TAA stop codon created by polyadenylation. The trnA sequence starts at the nucleotide position 5924 and the excision of the trnA at this nucleotide would essentially remove the TAG stop codon, and generation of the TAA stop codon when the polyA+ tail was added. There were one to two nucleotide
Table 1. Organization of *Helicoverpa* zea mitogenome

| Gene name | Start | Stop | Length | Start | Stop | tRNA | Overlap/anticond gap (bp) |
|-----------|-------|------|--------|-------|------|------|--------------------------|
| trnM      | 1     | 69   | 68     | CAT   | 0    |
| trnl      | 70    | 133  | 64     | GAT   | -3   |
| trnQ      | 199   | 131  | 69     | TTG   | 45   |
| nad2      | 245   | 1,255| 1,011  | ATT   | TAA  | 2    |
| trnW      | 1,258 | 1,325| 68     | TCA   | -8   |
| trnC      | 1,386 | 1,318| 69     | GCA   | 10   |
| trnY      | 1,462 | 1,397| 66     | GGA   | 9    |
| cox1      | 1,465 | 2,995| 1,531  | CGA   | T    |
| trnL2     | 2,996 | 3,062| 67     | TGG   | 7    |
| cox2      | 3,063 | 3,744| 682    | ATG   | PolyA |
| trnK      | 3,745 | 3,815| 71     | CTG   | 4    |
| trnD      | 3,820 | 3,885| 66     | GTC   | 0    |
| atp8      | 3,886 | 4,047| 162    | ATC   |
| atp6      | 4,041 | 4,718| 678    | ATG   |
| cox3      | 4,718 | 5,503| 786    | ATG   |
| nadG      | 5,506 | 5,571| 66     | TCC   | 0    |
| nad3      | 5,572 | 5,925| 354    | ATT   |
| trnA      | 5,924 | 5,992| 69     | TGC   | 0    |
| trnR      | 5,993 | 6,059| 67     | TCG   | 27   |
| trnN      | 6,087 | 6,152| 66     | GGT   | 1    |
| trnS1     | 6,154 | 6,221| 68     | GCT   |
| trnE      | 6,222 | 6,288| 67     | TTC   | -2   |
| trnF      | 6,353 | 6,287| 67     | GAA   | 0    |
| nad5      | 8,099 | 8,534| 1,436  | ATG   |
| trnH      | 8,166 | 8,107| 60     | GTC   | 0    |
| nad4      | 9,508 | 8,167| 1,339  | ATG   |
| nad4L     | 9,840 | 9,550| 291    | TAG   |
| trnT      | 9,844 | 9,909| 66     | TGT   | 0    |
| trnP      | 9,974 | 9,910| 65     | TGG   |
| nad6      | 9,982 | 10,515| 534    | ATT   |
| cob       | 10,516| 11,667| 1,152  | ATG   |
| trnS2     | 11,673| 11,738| 66     | TGA   |
| nad1      | 12,700| 11,762| 939    | ATG   |
| trnl      | 12,769| 12,702| 68     | TAG   | 0    |
| rnlL      | 14,159| 12,770| 3,390  | TAC   |
| trnV      | 14,225| 14,160| 66     | TAC   |
| rnsS      | 15,019| 14,226| 794    | TAC   |
| Control   | 15,020| 15,348| 329    | TAC   |
| trnl-like*| 15,173| 15,268| 95     | AAT   |

*Start and stop nucleotide positions are given for all genes. Translation start and stop codons for open reading frames of PCGs and the anti-codons for tRNA genes are given in respective columns. The number of nucleotide in a start or stop nucleotide positions are given for all genes. Translation start and stop codons for open reading frames of PCGs and the anti-codons for tRNA genes are given in respective columns. The number of nucleotide in a start or stop codon positions for PCGs are given in Table 2. Alignment of the transcripts and translated peptide sequences of the PCGs of two species are given in Fig. S1. The ML models tested whether the Kd/Ks ratio for a given sequence pair significantly deviate from 1 (i.e., neutral selection). Significantly low Kd/Ks ratios (P < 0.00001) were observed for all PCG comparisons except for atp8 and nad4L, indicating that stabilizing or purifying selection pressure on most mitochondrial PCGs. The highest number of nonsynonymous substitutions was found in nad5 (6), nad6 (6), nad1 (4), nad4L (4), and cob (3). Nonsignificant test results indicated that the Kd/Ks ratios of atp8 and nad4L genes did not deviate from Kd/Ks ratio of 1 and may be selection neutral. The power of Kd/Ks ratio test was dependent on the sequence length and the degree of divergence between the sequences being compared (Nekrutenko et al. 2002). Therefore, short sequence length of atp8 and nad4L genes may have contributed to the nonsignificant test results.

Mitochondrial Transcripts

In the present study, a cDNA library produced by reverse transcription using an oligo d(T) primer was used to generate sequence reads from the polyadenylated 3’-ends, facilitating identification of polyadenylation sites of the PCG transcripts. Illumina sequence reads obtained from both developmental stages were submitted to NCBI SRA database (Project no. PRJNA259919). Reference assembly of RNA-Seq reads from 2-h eggs and 48-h embryos yielded 411,145 and 168,606 reads mapped to the reference mitochondrial gene set, respectively. Examination of the RNA-Seq reads mapped to the PCG revealed that in addition to polycistronic primary transcripts, both 2-h eggs and 48-h embryos produced mature polyadenylated transcripts for all genes except for atp8. All the reads mapped to atp8 extended beyond the putative translation end site into the atp6 gene, indicating that polypeptides of the gene pair atp8 and atp6 are translated from bicistronic mature transcripts. In *H. zea*, we did not find any evidence for tricistronic atp8-atp6-cox3 transcripts reported in *M. vitrata* (Margam et al. 2011).

Bicistronic nad4L-nad4 transcripts may be an evolutionarily conserved feature of mitogenome transcription in other species studied so far (Berthier et al. 1986, Scheibye-Alsing et al. 2007, Stewart and Beckenbach 2009, Margam et al. 2011, Wang et al. 2013). However, we found polyadenylated transcripts of nad4L and nad4 in both 2-h eggs and 48-h embryos, indicating that the transcripts are processed into single transcripts in *H. zea*. Moreover, the polyadenylation most likely started 42–44 bp downstream of the nad4L translation stop codon (at the nucleotide position 9,506–9,508), but the precise location could not be predicted because we could not determine whether the two adenosines present at the nucleotide positions 9,506–9,507 were part of the poly(A)+ tail or not. Unlike overlapping nad4L-nad4 genes in *M. vitrata* and *D. melanogaster*, these two genes in *H. zea* were separated by a 44-bp gap and processing them into individual mature transcripts would not affect the translation into polypeptides.

The mitochondrial genes nad6 and cob were considered as a bicistronic transcript in other insect species. We detected polyadenylated nad6 transcripts in both 2-h eggs and 48-h embryos, clearly indicating that nad6 and cob are processed as independent mature transcripts in all developmental stages of *H. zea*. Differences in processing of the cob transcript were also observed between 2-h eggs and 48-h embryos. In 48-h embryos, 40.5% of 121 reads mapped to the 3’-end of the cob transcript were polyadenylated, indicating that the transcript was processed as independent mature transcripts. In *H. zea*, however, we did not detect any evidence for polyadenylation in the 3’-ends of the cob transcript.

Although the transcription termination and polyadenylation sites of mitochondrial PCG are commonly inferred by comparison with a conserved feature of mitogenome transcription in other species studied so far (Berthier et al. 1986, Scheibye-Alsing et al. 2007, Stewart and Beckenbach 2009, Margam et al. 2011, Wang et al. 2013), we detected polyadenylated nad4L-nad4 transcripts in both 2-h eggs and 48-h embryos, indicating that the transcripts are processed as independent mature transcripts. In *H. zea*, differences in processing of the cob transcript were also observed between 2-h eggs and 48-h embryos. In 48-h embryos, 40.5% of 121 reads mapped to the 3’-end of the cob transcript were polyadenylated. However, we did not detect any evidence for polyadenylation in the 3’-ends of the cob transcript.

Although the transcription termination and polyadenylation sites of mitochondrial PCG are commonly inferred by comparison with a closely related species, our study found that some of the *H. zea* PCG have entirely different or multiple alternative polyadenylation sites compared with those established for *H. armigera* and other related species. In this study, we found polyadenylation of cob transcripts started at a site different from predicted by comparison with other closely related species. Transcripts of cox3, nad1, nad2, and nad5 had alternative polyadenylation sites...
Table 2. Total nucleotide substitutions, maximum-likelihood (ML) estimates of synonymous (S) and nonsynonymous (N) substitutions, proportion of synonymous substitutions per synonymous site (Ks), proportion of nonsynonymous substitutions per nonsynonymous site (Ka), the Ka/Ks ratio, and the P-value of the Fisher exact test for the 13 mitochondrial PCGs of Helicoverpa zea compared with the orthologous genes of H. armigera

| PCG  | Actual substitutions | ML estimated S-substitutions | ML estimated N-substitutions | Ks  | Ka  | P-value (Fisher) |
|------|----------------------|------------------------------|-----------------------------|-----|-----|----------------|
| nad2 | 24                   | 21.124                       | 2.876                       | 0.00536 | 0.64407 | 0.00833 1.8014E−27 |
| cob  | 39                   | 38.529                       | 0.472                       | 0.00064 | 0.47556 | 0.00134 1.6880E−42 |
| cox2 | 21                   | 20.728                       | 0.272                       | 0.00069 | 0.69224 | 0.00100 2.0111E−27 |
| atp8 | 3                    | 0.659                        | 2.341                       | 0.02107 | 0.04452 | 0.47329 5.6586E−01 |
| atp6 | 16                   | 15.908                       | 0.092                       | 0.00022 | 0.21872 | 0.00100 6.774E−15 |
| cox3 | 21                   | 20.861                       | 0.139                       | 0.00031 | 0.31421 | 0.00100 1.3740E−20 |
| nad3 | 13                   | 12.928                       | 0.072                       | 0.00035 | 0.34691 | 0.00100 2.0000E+00 |
| nad5 | 35                   | 31.137                       | 3.863                       | 0.00412 | 0.39477 | 0.00104 4.0051E−34 |
| nad4 | 27                   | 26.787                       | 0.213                       | 0.00197 | 1.77085 | 0.00111 1.3641E−26 |
| nad4l| 5                    | 1.554                        | 3.446                       | 0.01416 | 0.08010 | 0.17675 4.5499E−02 |
| nad6 | 18                   | 11.706                       | 6.294                       | 0.01829 | 0.31909 | 0.05732 2.1480E−09 |
| cob  | 39                   | 36.792                       | 2.208                       | 0.00385 | 0.46802 | 0.00823 5.7313E−34 |
| nad1 | 31                   | 29.368                       | 1.633                       | 0.00490 | 1.05782 | 0.00464 1.539E−35 |

Table 3. Protein coding sequences in Helicoverpa zea mitogenome with different polyadenylation start sites

| Gene | Positions of alternative Ply A+/site(s) |
|------|---------------------------------------|
| nad2 | 1255; 1257                            |
| cox3 | 5005; 5007; 5011                       |
| nad5 | 6351; 6354                             |
| cob  | 11677                                 |
| nad1 | 11757; 11759; 11760                    |

Table 4. Analysis of control regions of selected lepidopteran mitogenomes with tRNAscan-SE for putative tRNA-like sequences

| Species                  | Control region length | tRNA       | Intron Accession number |
|--------------------------|-----------------------|------------|-------------------------|
| Agrotis ipsilon          | 332                   | −          | −                       |
| Athyma sulphita          | 349                   | −          | −                       |
| H. punctigera            | 327                   | trnN +     | −                       |
| Bombyx mandarina         | 747                   | trnN +     | −                       |
| Bombyx mori              | 499                   | trnY −     | −                       |
| Glyphodes quadramaculalis| 327                   | −          | −                       |
| Helicoverpa armigera     | 328                   | −          | −                       |
| Helicoverpa zea          | 329                   | trn1 +     | −                       |
| Manduca sexta            | 324                   | −          | −                       |
| Mythimma separata        | 374                   | −          | −                       |
| Paracyrumoniza distincti| 352                   | −          | −                       |
| Spodoptera exigua        | 331                   | trn1 +     | −                       |
| Spodoptera litura        | 326                   | sup(UUA) + | −                       |

A “+” or “−” in the intron column indicates presence or absence of a possible intron within the predicted tRNA-like structure.

in addition to the site predicted by comparison with other mitogenomes (Table 3). These changes in polyadenylation sites also affect the start sites of the gene immediately downstream of the affected gene.

Other Features

Analysis of the control region of H. zea mitogenome with tRNAscan-SE Search program indicated the presence of a putative trnI-like sequence with a 23-bp intron. Control regions of H. punctigera and Bombyx mandarina mitogenomes also contained putative trnN-like structures with possible intron sequences (Table 4). Mitogenome control regions of Spodoptera exigua and Spodoptera litura contained trnI-like (AAU) and an Ochre suppressor tRNA-like (UAA) sequences with possible intron sequences, respectively. A putative trnY-like structure without an intron was predicted in the Bombyx mori control region. Control regions of H. armigera and a few other insects did not have tRNA-like structures detectable by the tRNAscan-SE program (Table 4). Vertebrate mitogenome control regions contain nucleotide sequences forming tRNA-like secondary structures that may play a role in DNA replication (Shadel and Clayton 1997). Evaluation of the 96 bp putative H. zea trnL-like sequence with mFold (http://mfold.rna.albany.edu/?q=DINAMelt/Quickfold) indicated the presence of a hammerhead-like structure (Fig. 2). Examination of the putative tRNA-like sequences from H. punctigera, B. mandarina, B. mori, S. exigua, and S. litura with mFold revealed secondary structures similar to that of predicted trnL-like sequence in H. zea. Mapping of RNA-Seq reads from 2-h eggs and 48-h embryos to the 329 bp control region resulted in a total of 269 mapped reads yielding an average coverage of 47.6. However, no transcripts were mapped to the nucleotides between 15,186 and 15,257 of the H. zea mitogenome. This stretch of control region overlaps with the predicted tRNA-like sequence. The presence of transcripts flanking this region may indicate that the polycistronic transcripts of mitochondrial genes start in this region and the secondary structure possibly plays a role in transcription initiation as well as replication initiation.

Expression Profiles in Early Eggs and Late Embryos

In total, 54,570,728 and 39,320,735 reads from 2-h eggs and 48-h embryos, respectively, were used for expression profiling and mapping reads to the complete mitogenome. The number of reads mapped to the mitochondrial PCGs and the two rRNA genes from 2-h eggs and 48-h embryos were 6,361,316 (11.6%) and 1,506,558 (3.8%), respectively. However, relative expression levels of the transcripts, except cox1, did not deviate significantly between two developmental time points. Expression profiles of PCGs indicated only 1.46- to 2.41-fold differences in expression levels of two transcripts (nad2 and cox1) between 2-h eggs and 48-h embryos (Fig. 3), but these differences did not represent physiologically significant differences in mitochondrial respiratory chain complexes between two
developmental stages. Expression levels of different PCG transcripts varied broadly, but in a similar pattern (Fig. 3), in both 2-h eggs and 48-h embryos. Normalized expression levels of cox complex genes varied from 9.49 ± 1.16 for cox2 in 48-h embryos to 90.58 ± 1.28 for cox1 in 2-h eggs. High level of cox gene transcripts in both developmental stages may represent physiologically significant gene expression patterns in developing embryos or an artifact resulting from mapping short nucleotide reads to cox PCGs, especially if nucleotide regions conserved between nuclear cox genes (including pseudo cox genes expressed in the nucleus) were mapped to the reference transcripts. Further studies are needed to fully understand whether these differences represent actual gene expression patterns or experimental artifacts. Sequence read coverage of the large subunit of tRNA (rrnL) was 33- and 65-fold higher in 2-h eggs and 48-h embryos, respectively, compared with the small subunit of rRNA (rrnS). Although the oligo d(T) primer used for reverse transcription was expected to convert mRNA with a poly A-tail, rRNA contamination is common in even in cDNA synthesized with poly(A)⁺ selected mRNA (Sooknanan et al. 2010). Therefore, unequal representation of the two rRNA subunit transcripts was most likely due to uneven conversion of contaminating rRNA transcripts to cDNA.

Mapping the nucleotide sequence reads from two developmental times to the complete mitogenome also resulted in a coverage maps similar to those observed in expression analysis (Fig. 4). Peaks in the coverage maps represent high read coverage, usually in the middle of the transcript, and valleys represent the ends of transcripts with low coverage. This pattern was observed for all individual PCGs except for atp8 and atp6 transcripts where a read mapping pattern consistent with a single bicistronic atp8-atp6 transcript was observed, indicating that atp8-atp6 is the only mature multicistronic transcript present in mitochondrial of H. zea. Evaluation of the mapped reads for nucleotide sequence variants did not identify any polymorphisms in mitochondrial transcripts other than random variations due to sequencing errors. This result is not surprising because the eggs used for generating RNA-Seq libraries were obtained from a highly inbred laboratory colony.

**Discussion**

In this study, we determined the complete nucleotide sequence of the mitogenome of H. zea, identified PCG transcripts that were processed differently compared with other insect species, and evaluated expression levels of PCGs in early eggs and late embryos. Conserved lepidopteran gene organization was preserved in H. zea mitogenome with some species specific characteristics such as differences in gaps or overlaps between mitochondrial genes.

Genetic studies using mitogenomes sequences are abundant in literature (reviewed in: Moritz 1994, Cameron 2014). Segments of mitochondrial genes such as cox1 and cox2 have been used in evolutionary and population genetic studies as well as for identification of Helicoverpa species (Behere et al. 2007, 2008; Tay et al. 2013; Leite et al. 2014; Mastrangelo et al. 2014). Although these highly conserved genes provide sufficient level of divergence between species, they may not provide sufficient resolution for intraspecific genetic studies and multilocus or complete mitogenome-based analyses are more suitable for intraspecific genetic studies (Anstead et al. 2002, Zhang et al. 2012, Cordon-Obras et al. 2014, De Jager et al. 2014, Morales et al. 2015).

Alignment of H. zea and H. armigera mitogenomes indicated 327 nucleotide substitutions between two species (Supplementary data S1 and S2) of which 292 (89.3%) were in the 13 PCGs (Table 2). There were 47 alignment gaps resulting from one to four nucleotide differences (indels) in nonprotein coding regions. Coding regions of cox1 and cox2, two genes commonly used in genetic
studies, had 39 and 21, respectively, nucleotide substitutions between two Helicoverpa species. A recent study of H. armigera and H. zea in Brazil using cox1 gene reported low nucleotide diversity within both species (Leite et al. 2014). Multilocus analysis that included less conserved genes such as nad2, nad6, and cob would have been more suitable for studying these insects due to inherent low genetic diversity (Mallet et al. 1993, Perera and Blanco 2011).

Fig. 3. Expression levels of the protein coding sequences of the mitogenome of Helicoverpa zea in 2-h-old eggs and 48-h-old embryos, relative to the expression level of nad4L. Expression levels of the PCGs were normalized to that of nad4L which had the lowest expression level.

Fig. 4. Coverage of RNA-Seq reads from 2-h eggs (orange) and 48-h embryos (blue) mapped to complete mitogenome of Helicoverpa zea. All PCGs and rRNA subunit genes are shown below the coverage map. Peaks represent high read coverage, usually in the middle of the transcript, and valleys represent the ends of transcripts with low coverage.
Mitogenomes are tRNA punctuated and polycistronic mitochondrial primary gene transcripts cleaved off at the 5’- and 3’-end of each tRNA gene by mitochondrial RNase P and RNase Z, respectively, to produce mature transcripts (Rackham and Filipovska 2013). It has been suggested that when PCGs are not punctuated by tRNA genes, tRNA-like secondary structures at the boundaries of the PCG are recognized by the processing machinery (Ojala et al. 1981, Montoya et al. 1983, Fernandez-Silva et al. 2003, Sanchez et al. 2011, Rackham and Filipovska 2009) and a mitochondrial PCG in H. zea (Hauswirth 1999) no information is available on the mitochondrial gene transcription during insect embryonic development. In reference assembly of RNA-Seq reads, we detected mapped sequence reads without polyA+ tails extending past the 3’-ends all PCGs (i.e., reads originated from partially processed primary transcripts) and polyA+ containing reads at the 3’-ends of all PCGs except atp8 (both 2-h eggs and 48-h embryos) and cob (2-h eggs) indicating that both early and late embryos contain a mixture of mature and primary transcripts at various stages of processing. However, determination of developmental stage at which mitochondrial gene transcription activated was not possible with the data obtained in this study.

Transcript mapping analysis also revealed differences in transcript termination in some H. zea PCGs compared with published data for other lepidopteran species. Annotation of most mitogenomes has been carried out not by analyzing actual transcripts, but by using annotations from a mitogenome of a closely related species. For example, stop codon of the nad3 gene was considered the native TAG of all other lepidopteran species. Overlap of two-nucleotides between nad3 and trnA would require that the mitochondrial transcript processing machinery excise the trnA between the T and AG of the native stop codon, leaving the T at the 3’-end of the nad3. Because PCGs and tRNAs do not remain as bicistronic transcripts, it is only logical to assume that all species with overlapping nad3 and trnA would be processed similarly and the annotations in those mitogenomes were simply an overlook of the transcript processing requirements. Similarly, novel transcription end sites we described here for some PCGs may not be novel after all; detailed transcript analysis may reveal similar transcription termination or polyadenylation sites in other closely related lepidopteran species. It is also possible that the pattern of transcript processing observed developing embryos may be different from other life stages. Therefore, further studies will be necessary to determine the patterns of mitochondrial transcripts in other life stages. RNA-Seq libraries used in this analysis were not suitable to determine actual transcription start sites or 5’-ends of mature transcripts. Sanger dideoxy sequencing of a 5’-RACE cDNA library would be the most suitable method to identify the 5’-ends of the mitochondrial genes.

**Supplementary Data**

Supplementary data are available at Journal of Insect Science online.

**Acknowledgments**

We thank Calvin Pierce and Padmariya Chatakondi for technical assistance and Drs Paul Shirk, Steve Valles (Center for Medical and Veterinary Entomology, USDA-ARS), and Kent Shelby (Biological Control of Insects Research Unit, Columbus, MO) for critically evaluating an earlier version of this manuscript. United States Department of Agriculture is an equal opportunity employer. The mention of trade names or commercial products in this report is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.
References Cited

Ammini, C. V., and W. W. Hauswirth. 1999. Mitochondrial gene expression is regulated at the level of transcription during early embryogenesis of Xenopus laevis. J. Biol. Chem. 274: 6265–6271.

Anstead, J., J. Burd, and K. Shufran. 2002. Mitochondrial DNA sequence divergence among Schizaphis graminum (Hemiptera: Aphididae) clones from cultivated and non-cultivated hosts: haplotype and host associations. Bull. Entomol. Res. 92: 17–24.

Behere, G., W. Tay, D. Russell, D. Heckel, B. Appleton, K. Kranthi, and P. Batterham. 2007. Mitochondrial DNA analysis of field populations of Helicoverpa armigera (Lepidoptera: Noctuidae) and of its relationship to H. zeas. BMC Evol. Biol. 7: 117.

Behere, G., W. Tay, D. Russell, and P. Batterham. 2008. Molecular markers to discriminate among four pest species of Helicoverpa (Lepidoptera: Noctuidae). Bull. Entomol. Res. 98: 599–603.

Berthier, F., M. Renaud, S. Alziari, and R. Durand. 1986. RNA mapping on Drosophila mitochondrial DNA: precursors and template strands. Nucleic Acids Res. 14: 4519–4533.

Boore, J. L. 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27: 1767–1770.

Cameron, S. L. 2014. Insect mitochondrial genomics: implications for evolution and phylogeny. Annu. Rev. Entomol. 59: 95–117.

Castro, J. A., A. Picornell, and M. Ramon. 1998. Mitochondrial DNA: a tool for populational genetics studies. Int. Microbiol. 1: 327–332.

Caterino, M. S., S. Cho, and F. A. Spering. 2000. The current state of insect molecular systematics: a thriving Tower of Babel. Annu. Rev. Entomol. 45: 1–54.

Cordon-Obras, C., J. Cano, J. Knapp, P. Nebreda, N. Ndong-Mabale, P. R. Boore, J. L. 1999. Xenopus laevis is regulated at the level of transcription during early embryogenesis of Helicoverpa armigera (Lepidoptera: Noctuidae) in crosses between the North American Heliothis (Hemiptera: Aphididae) clones from Russia. J. Econ. Entomol. 88: 77–80.

American Heliothis? Chromatin-associated virus-like particles in testes of six species and transcription in mammalian mitochondria. Annu. Rev. Biochem. 76; 1–54.

Degrugillier, M. E., and S. M. Newman, Jr. 1993. Hereditary viruses of Heliothis? Chromatin-associated virus-like particles in testes of six species and transcription in mammalian mitochondria. Annu. Rev. Biochem. 76: 1–54.

Czech, G., K. C. Albenz, L. M. Vivian, H. O. Guimaraes, and T. Carvalhais. 2013. First reported occurrence of Helicoverpa armigera (Hübner)(Lepidoptera: Noctuidae) in Brazil. Pesq. Agropec. Trop. 43: 110–113.

De Jager, L., N. Burger, and A. Botha. 2014. Complete mitochondrial genome of Duaraaphis moza (Hemiptera: Aphididae) from nine populations, SNP variation between populations, and comparison with other Aphididae species. Afric. Entomol. 22: 847–862.

Mastrangelo, T., D. Paulo, L. Bergamo, E. Morais, M. Silva, G. Bezerra-Silva, and A. Azeredo-Espin. 2014. Detection and genetic diversity of a Heliothine invader (Lepidoptera: Noctuidae) from north and northeast of Brazil. J. Econ. Entomol. 107: 970–980.

Metcalfe, C. L., and W. P. Hinton. 1962. Destructive and useful insects: their habits and control. McGraw-Hill, New York.

Montoya, J. G., L. G. Gaines, and G. Attardi. 1983. The pattern of transcription of the human mitochondrial tRNA genes reveals two overlapping transcription units. Cell 34: 151–159.

Morales, H. E., A. Pavlova, L. Joseph, and P. Sunnucks. 2015. Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. Mol. Ecol. 24: 2820–2837.

Moritz, C. 1994. Applications of mitochondrial DNA analysis in conserva- tion: a critical review. Mol. Ecol. 3: 401–411.

Mundry, M., E. Bornberg-Bauer, M. Sammeth, and P. G. Feulner. 2012. Evaluating characteristics of de novo assembly software on 454 transcriptome data: a simulation approach. PLoS One 7: e31410 (doi: 10.1371/journal.pone.0016444).

Murría, M. G., F. S. Scalaor, F. R. Navarro, L. E. Cazado, A. Casmuz, M. E. Villagrán, E. Lobos, and G. Gastaminza. 2014. First record of Helicoverpa armigera (Lepidoptera: Noctuidae) in Argentina. Florida Entomol. 97: 854–856.

Muser, F. R., A. L. Catchot, J. A. Davis, D. A. Herbert, G. M. Lorenz, T. Reed, D. D. Reisig, and S. D. Stewart. 2013. 2012 soybean insect losses in the southern US. Midsouth Entomol. 6: 12–24.

Nekrutienko, A., K. D. Makova, and W.-H. Li. 2002. The K(A)/K(S) ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study, Genome Res. 12. 198–202.

Ojala, D., C. Merkel, R. Gelfand, and G. Attardi. 1980. The tRNA genes punctuation the reading of genetic information in human mitochondrial DNA. Cell 22: 393–403.

Ojala, D., J. Montoya, and G. Attardi. 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature 290: 470–474.

Pabijan, M., A. Wandycz, S. Hofman, K. Wcek, M. Pwczynski, and J. M. Szymura. 2013. Complete mitochondrial genomes resolve phylogenetic relationships within Bombina (Anura: Bombinatoridae). Mol. Phylogenet. Evol. 69: 63–74.

Perera, O. P., and C. Blanco. 2011. Microsatellite variation in Helicoverpa zea (Boddie) populations in the southern United States. Southwest. Entomol. 36: 271–286.
Perera, O. P., C. A. Blanco, B. E. Scheffler, and C. A. Abel. 2007. Characteristics of 13 polymorphic microsatellite markers in the corn earworm, Helicoverpa zea (Lepidoptera: Noctuidae). Mol. Ecol. Notes 7: 1132–1134.

Perna, N. T., and T. D. Kocher. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41: 353–358.

Posada, D. 2003. Using MODELTEST and PAUP* to select a model of nucleotide substitution. Curr. Protoc. Bioinform. 6.5.1–6.5.14.

Presser, B. D., and C. W. Rutschky. 1957. The embryonic development of the corn earworm, Heliothis zea (Boddie) (Lepidoptera, Phalaenidae). Ann. Entomol. Soc. Am. 50: 133–164.

Rackham, O., and A. Filipovska. 2013. Tiny but mighty: expression of the mitochondrial genome. Aust. Biochem. 44: 5–13.

Sanchez, M. I., T. R. Mercer, S. M. Davies, A. M. Shearwood, K. K. Nygard, T. R. Richman, J. S. Mattick, O. Rackham, and A. Filipovska. 2011. RNA processing in human mitochondria. Cell Cycle 10: 2904–2916.

Schattner, P., A. N. Brooks, and T. M. Lowe. 2005. The tRNAscan-SE, sno-scan and snoGPS web servers for the detection of tRNAs and snorRNAs. Nucleic Acids Res. 33: W686–W689.

Scheibye-Alsing, K., S. Cirera, M. J. Gilchrist, M. Fredholm, and J. Gorodkin. 2007. EST analysis on pig mitochondria reveal novel expression differences between developmental and adult tissues. BMC Genomics 8: 367.

Shadel, G. S., and D. A. Clayton. 1997. Mitochondrial DNA maintenance in vertebrates. Annu. Rev. Biochem. 66: 409–435.

Sooknanan, R., J. Pease, and K. Doyle. 2010. Novel methods for rRNA removal and directional, ligation-free RNA-seq library preparation. Nat. Methods 7: 7.

Stewart, J. B., and A. T. Beckenbach. 2009. Characterization of mature mitochondrial transcripts in Drosophila, and the implications for the tRNA punctuation model in arthropods. Gene 445: 49–57.

Tay, W. T., M. F. Soria, T. Walsh, D. Thomazoni, P. Silvie, G. T. Behere, C. Anderson, and S. Downes. 2013. A brave new world for an old world pest: Helicoverpa armigera (Lepidoptera: Noctuidae) in Brazil. PLoS One 8: e80134 (doi: 10.1371/journal.pone.0080134).

Teacher, A. G., C. André, J. Merilä, and C. W. Wheat. 2012. Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. BMC Evol. Biol. 12: 248.

Torres, T. T., M. Dolezal, C. Schlötner, and B. Ottenwalder. 2009. Expression profiling of Drosophila mitochondrial genes via deep mRNA sequencing. Nucleic Acids Res. gkp856.

Walsh, T. K. 2014. Characterization of the complete mitochondrial genome of the Australian Heliothis moth, Helicoverpa punctigera (Lepidoptera: Noctuidae). Mitochondrial DNA 27: 167–168.

Wang, C. 2007. Interpretation of the biological species concept from interspecific hybridization of two Helicoverpa species. Chinese Sci. Bull. 52: 284–286.

Wang, H.-L., J. Yang, L. M. Boykin, Q.-Y. Zhao, Q. Li, X.-W. Wang, and S.-S. Liu. 2013. The characteristics and expression profiles of the mitochondrial genome for the Mediterranean species of the Bemisia tabaci complex. BMC Genomics 14: 401.

Williams, M. R. 2014. Cotton insect losses—2013. In Williams M. R. (ed.), Mississippi State University, Cooperative Extension Service, Mississippi State University, Mississippi State University (http://www.entomology.msstate.edu/resources/croplosses/2013loss.asp).

Yin, J., G. Y. Hong, A. M. Wang, Y. Z. Cao, and Z. J. Wei. 2010. Mitochondrial genome of the cotton bollworm Helicoverpa armigera (Lepidoptera: Noctuidae) and comparison with other Lepidopterans. Mitochondrial DNA 21: 160–169.

Yu, L., Y. W. Li, O. A. Ryder, and Y. P. Zhang. 2007. Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. BMC Evol. Biol. 7: 198.

Zeng, F., S. B. Ramaswamy, R. G. Luttrell, J. Reed, C. D. Parker, S. Stewart, A. Harris, K. Knighten, J. Robbins, J. Xia, et al. 1999. Comparison of monoclonal antibody and laboratory rearing techniques to identify heliothentinae (Lepidoptera : Noctuidae) eggs from Mississippi cotton fields. Environ. Entomol. 28: 275–281.

Zeng, F., S. B. Ramaswamy, and S. Pruett. 1998. Monoclonal antibodies specific to tobacco budworm and bollworm eggs. Ann. Entomol. Soc. Am. 91: 677–684.

Zhang, B., O. Edwards, L. Kang, and S. Fuller. 2012. Russian wheat aphids (Diuraphis noxia) in China: native range expansion or recent introduction? Mol. Ecol. 21: 2130–2144.

Zhang, Z., J. Li, X.-Q. Zhao, J. Wang, G.K.-S. Wong, and J. Yu. 2006. KaKs Calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics 4: 259–263.