Agri Gene 5 (2017) 27–36

Contents lists available at ScienceDirect

Agri Gene

journal homepage: www.elsevier.com/locate/aggene

The mitogenome of the brown pod-sucking bug *Clavigralla tomentosicollis* Stäl (Hemiptera: Coreidae)

Laura D. Steele[a,*], Weilin Sun[b], M. Carmen Valero[a,c], James Adebayo Ojo[d], Keon Mook Seong[a], Brad S. Coates[e], Venu M. Margam[f], Manuele Tamò[g], Barry R. Pittendrigh[b]

[a] Department of Entomology, University of Illinois, Urbana-Champaign, IL 61801, USA
[b] Department of Entomology, Michigan State University, East Lansing, MI 48824, USA
[c] Carl R. Woese Institute of Genomic Biology, University of Illinois at Urbana-Champaign, IL 61801, USA.
[d] Department of Crop Production, Kwara State University, Malete, Ilorin, Nigeria
[e] United States Department of Agriculture – Agricultural Research Service, Corn Insect and Crop Genetics Research Unit, Genetics Laboratory, Iowa State University, USA
[f] Aspire Food Group, Kumasi, Ghana
[g] International Institute of Tropical Agriculture, Cotonou, Benin

ARTICLE INFO

Keywords:
Clavigralla tomentosicollis
Pod-sucking bug
Cowpea
Vigna unguiculata
mitogenome
Coreoidea

ABSTRACT

The brown pod-sucking bug, *Clavigralla tomentosicollis* Stäl (Hemiptera: Coreidae), causes significant damage to cultivated cowpea, *Vigna unguiculata* Walp, a staple crop in sub-Saharan Africa. *C. tomentosicollis* pierce and suck sap from cowpea pods, resulting in reduced grain yield and quality. The complete, 16,089 bp mitogenome of *C. tomentosicollis* encodes 13 protein-coding genes (PCGs), 22 transfer RNAs (tRNAs), two ribosomal RNAs (rRNAs) and an A + T rich control region, with gene order and orientation identical to that of the insect ancestral gene order. The initiation and termination codons for the PCGs used standard ATN codons and TAA or TAG codons respectively. All predicted tRNAs fold into a clover-leaf secondary structures with the exception of tRNA-Ser (AGN) with a semi-loop dihydrouridine arm. The 1509 bps A + T rich region contains a single 89 bp tandem repeat unit duplicated 3.7 times. When compared with other published Coreoidea mitogenomes, *C. tomentosicollis* was also highly A − T skewed, and similar in both size and A − T%; however, its longer tandem repeat within the A + T rich region was unique. The *C. tomentosicollis* mitogenome can serve as a foundation to combine molecular marker data with pest monitoring strategies to better understand the population dynamics of this species.

1. Introduction

Cowpea (*Vigna unguiculata* Walp.) is a legume crop that serves as a major food staple in sub-Saharan Africa, grown for both human and animal consumption (Singh and Singh, 2015). Cowpea, owing to its high protein content (23–25%), drought tolerance and nitrogen-fixing ability, is an important crop in many developing nations (Singh and Singh, 2015; Devi et al., 2015). A pest insect complex in Africa causes severe damage to cowpea that can result in yield losses of up to 70% (Adati et al., 2008; Aliyu et al., 2007). Notable pest species include the legume pod borer (*Marcua vitrata* Fabricius), the cowpea aphid (*Aphis craccivora* Koch), the flower thrips (*Megalurothrips sjostedti* Trybom), the cowpea weevil (*Callosobruchus maculatus* Fabricius) (a storage pest), and a complex of pod-sucking insects including the brown pod-sucking bug *Clavigralla tomentosicollis* Stäl (Coreidae), *Clavigralla shadabi* Dolling (Coreidae), and *Riptortus dentipes* Fabricius (Alydidae) (Koona et al., 2004; Soyelu et al., 2007; Dreyer and Baumgartner, 1994). The most damaging of the pod-sucking bugs, *C. tomentosicollis* (Hemiptera: Coreidae) can dramatically decrease yields in cowpea crops (Dreyer and Baumgartner, 1994; Jackai, 1990; Koona et al., 2002), with levels of damage ranging from 20–100% (Singh and Allen, 1980; Aliyu et al., 2007). Both nymphs and adults suck the sap from the pods causing premature pod drying and shriveling (Jackai et al., 2001). Pest management efforts have included the planting of resistant cowpea varieties (Olatunde et al., 2007; Dabire-Binso et al., 2010), the use of botanical pesticides (Oparaeke, 2006a, 2006b) and biocontrol methods among others.

The amount of genetic data available for *C. tomentosicollis* is sparse,

Abbreviations: mitogenomes, mitochondrial genomes; *C. tomentosicollis*, *Clavigralla tomentosicollis*; ORF, Open Reading Frame; PCG, protein coding genes; tRNA, transfer ribonucleic acid; rRNA, ribosomal ribonucleic acid; lrRNA, large rRNA; srRNA, small rRNA; cox, cytochrome oxidase c subunit; nad, NADH Dehydrogenase subunit; ATP, *atp* synthase subunit; cytb, cytochrome *b*; PCR, polymerase chain reaction

* Corresponding author.
E-mail address: steele11@illinois.edu (L.D. Steele).

http://dx.doi.org/10.1016/j.aggene.2017.07.002
Received 26 March 2017; Received in revised form 19 June 2017; Accepted 10 July 2017
Available online 16 July 2017

with one study that generated an annotated transcriptome sequence which identified genes of interest for pest control and potential molecular genetic markers (Agunbiade et al., 2013). Insect mitochondrial genomes (mitogenomes) are circular, double-stranded DNA molecules that range in size from 14–20 kb (Kim et al., 2009; Liu et al., 2015; Zhang et al., 2015). Mitogenomes encode 22 transfer RNA (tRNA) genes, two ribosomal RNA (large and small rRNA) genes (*lrRNA* and *srRNA*), which play a role in the translation of seven NADH dehydrogenase subunits (*nad1-6* and *nad4L*), two ATPase genes (*atp6* and *atp8*), three cytochrome *c* oxidase subunits (*cox1-3*), and a cytochrome B subunit (*cytb*) (Moritz et al., 1987; Wolstenholme, 1992; Cameron, 2014a). Insect mitogenomes also contain a large, non-coding, A − T-rich region which serves as the origin for replication and transcription (Wolstenholme, 1992; Zhang and Hewitt, 1997; Cameron, 2014a). Mitogenomes are inherited maternally and undergo very little sequence recombination, all of which are considered advantageous for phylogenetic, comparative, and evolutionary genomics (Harrison, 1989; Boore, 1999; Whinnett et al., 2005; Timmermans et al., 2014; Yuan et al., 2015b). Additionally, mitogenomes serve as molecular markers for molecular evolution and population genetics studies (Harrison, 1989; Boore, 1999; Whinnett et al., 2005; Timmermans et al., 2014; Abascal et al., 2006). The cytochrome *c* oxidase I (*cox1*) gene in particular has aided in differentiating cryptic species of butterflies (Herbert et al., 2004) and subspecies of *M. vitrata* (Margam et al., 2011a, 2011b). Out of the 146 complete hemipteran mitogenomes available on NCBI, there are six from the superfamily Coreoidea: *Riptortus pedestris* (Alydidae), *Aeschyntelus notatus* (Rhopalidae), *Stictopleurus subviridis* (Rhopalidae), *Corizus tetraspilus* (Rhopalidae), *Leptocorisa* sp. (Alydidae), and *Hydaropsis longirostris* (Coreidae), the only one from the family Coreidae (Hua et al., 2008; Tang et al., 2014; Yuan et al., 2015a; Wang et al., 2015).

Herein we describe the sequencing, assembly and annotation of the complete mitogenome of *C. tomentosicollis*, including a comparative analysis with the six other currently available Coreoidea mitogenomes. A complete mitogenome for *C. tomentosicollis* has the potential to provide an important tool for insect comparative and evolutionary genomics, as well as be a tool for understanding population dynamics of this species that might ultimately lead to better pest management decisions.

2. Materials and methods

2.1. Sampling, DNA extraction and sequencing

C. tomentosicollis samples were collected from field grown cowpea plants in Benin, preserved in RNAlater and stored at − 4 °C. The abdomen from a single insect sample was rinsed with 100% ethanol and then ground to a powder in liquid nitrogen. The samples were suspended in Buffer BP1 (Qiaprep Spin Miniprep Kit, QIAGEN, Valencia, CA, USA), incubated while spinning for 20 min, and run through a QiaShredder column (QIAGEN). DNA was extracted using the Qiaprep Spin Miniprep Kit according to the manufacturer's instructions, yielding a mitochondria-enriched sample. DNA quantity was determined using NanoDrop 1000 UV/VIS Spectrophotometer (Thermo Scientific, Serial No. G642) and quality determined by 0.9% agarose gel electrophoresis.

2.2. Illumina HiSeq2500™ sequencing and assembly

A shotgun genomic library was constructed using the Hyper Library Construction Kit (Kapa Biosystems, Wilmington, MA, USA). The library was loaded on a single lane and DNA was sequenced at 2 × 160 nt using a HiSeq SBS Kit v4 (Illumina, San Diego, CA, USA). Illumina shotgun sequencing was completed on an Illumina HiSeq2500 at the W.M. Keck Center for Comparative and Functional Genomics at the University of Illinois Urbana-Champaign. Resulting fastq files were input to CLC Genomics Workbench 8.0.2 (Cambridge, MA, USA) as paired reads (minimum distance 160 bp, maximum distance 825 bp), and trimmed for reads with quality score (q) < 20 and the two adapter sequences (5′-AGA TCG GAA GAG CAC ACG TCT GAA CTC CAG TCA C-3′; 5′-AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT-3′) removed. A de novo assembly was performed using the following parameters: mapping mode-create simple contig sequences (fast) contig length ≥ 200; automatic word size = yes; perform scaffolding = yes; auto-detect paired distances = yes.

2.3. Mitogenome annotation and analysis

All resulting contigs from the assembly of between 10,001–30,005 bps were extracted and used as blast queries (blastn; Match/Mismatch and Gap Costs = Match 2 Mismatch 3 Existence 5 Extension 2; Expectation value = 1.0E-6; Word size 11) against the mitogenomes of three other Coreoidea insect species: *A. notatus* (Accession number: NC_012446.1), *H. longirostris* (Accession number: NC_012456.1), and *R. pedestris* (Accession number: NC_012462.1). The BLAST results for contigs with an *E*-value of 0.00 were retained for further PCR validation. The NCBI ORF Finder (Open Reading Frame Finder) with invertebrate mitochondrial genetic codes was used to identify the protein coding genes (PCGs), which were confirmed via protein BLAST tool on NCBI (http://www.ncbi.nlm.nih.gov) and alignment with corresponding PCGs from six other Coreoidea (*A. notatus*, *H. longirostris*, *R. pedestris*, *Stictopleurus subivridis* [Accession number: NC_012888.1], *Corizus tetraspilus* [Accession number: KM983397], *Leptocorisa* sp. [Accession number: KM244663.1]). Three PCGs (*nad6*, *nad4L*, and *cox3*) were verified using PCR and additional Sanger sequencing as portions of these genes contained ambiguity. The *cox1* gene was analyzed for species identification using the Barcode of Life Data System (BOLD) (Ratnasingham and Hebert, 2007).

Transfer RNAs (tRNAs) and tRNA secondary structures were identified using ARWEN v1.2 (Laslett and Canback, 2008). The two ribosomal RNAs (rRNAs), *lrRNA* and *srRNA*, were determined via sequence comparison, using CLUSTAL Omega (Goujon et al., 2010; Sievers et al., 2011) with the six other published Coreoidea insect mitogenomes. Nucleotide composition was determined using MEGA v6.06 (Tamura et al., 2011) and strand asymmetry was calculated using the formulae: AT skew = [A% − T%] / [A% + T%] and GC skew = [G% − C%] / [G% ± C%] (Perna and Kocher, 1995). Codon usage and the relative synonymous codon frequencies (RSCU) were calculated using DnaSP 5.10.01 (Librado and Rozas, 2009). The Tandem Repeats Finder program (https://tandem.bu.edu/trf/trf.html; Benson, 1999) was used to predict tandem repeats in the A + T-rich region. DNA sequence and protein sequence percentage identities were calculated using CLUSTAL Omega (Goujon et al., 2010; Sievers et al., 2011).

2.4. Sanger sequence confirmation of nad6, nad4L, cox3 and the A + T-rich region

Additional verification of the sequence was performed for the genes *nad6*, *nad4l*, *cox3*, as each gene contained a small region of ambiguities, and the A + T-rich region, with its' high A − T content and repeat units making Illumina sequencing alone insufficient. Genes were amplified via PCR using primers designed using CLC Genomics Workbench and SnapGene Viewer Version 3.2.1 (GSL BioTech) (Supplementary Table 1). PCR was performed with GoTaq® Green Master Mix (Promega Corporation, USA) with the following cycle conditions: initial denaturation step: 2 min at 95 °C, then 35 cycles of 30 s denaturation at 95 °C, 30 s annealing at 49 °C, 30 s extension at 68 °C, and final extension step 5 min at 72 °C. Cleaned PCR products were Sanger sequenced at the W.M. Keck Center for Comparative and Functional Genomics at the University of Illinois Urbana-Champaign. Sequencing