

Comparison of the mitochondrial genomes of the Old and New World strains of the legume pod borer, *Maruca vitrata* (Lepidoptera: Crambidae)

Tolulope A. Agunbiade^{1*}, Brad S. Coates², Weilin Sun³,
Mu-Rou Tsai⁴, Maria Carmen Valero⁵, Manuele Tamò⁶
and Barry R. Pittendrigh³

¹Section of Infectious Diseases, Department of Internal Medicine, Yale University, New Haven, CT 06519, USA; ²United States Department of Agriculture-Agricultural Research Service, Corn Insect and Crop Genetics Research Unit, Genetics Laboratory, Iowa State University, Ames, IA, 50011, USA; ³Department of Entomology, Michigan State University, East Lansing, MI, 48824, USA; ⁴Graduate Institute of Biomedical Science, Chung Hwa University of Medical Technology, Tainan, 71703, Taiwan; ⁵Computing Genomes for Reproductive Health, Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, 61801, USA; ⁶International Institute of Tropical Agriculture, 08 BP 0932 Tri Postal, Cotonou, Benin

(Accepted 29 June 2016)

Abstract. *Maruca vitrata* (Fabricius, 1787) is a cryptic pantropical species of Lepidoptera that are comprised of two unique strains that inhabit the American continents (New World strain) and regions spanning from Africa through to Southeast Asia and Northern Australia (Old World strain). In this study, we *de novo* assembled the complete mitochondrial genome sequence of the New World legume pod borer, *M. vitrata*, from shotgun sequence data generated on an Illumina HiSeq 2000. Phylogenomic comparisons were made with other previously published mitochondrial genome sequences from crambid moths, including the Old World strain of *M. vitrata*. The 15,385 bp *M. vitrata* (New World) sequence has an 80.7% A+T content and encodes the 13 protein-coding, 2 ribosomal RNA and 22 transfer RNA genes in the typical orientation and arrangement of lepidopteran mitochondrial DNAs. Mitochondrial genome-wide comparison between the New and Old World strains of *M. vitrata* detected 476 polymorphic sites (4.23% nucleotide divergence) with an excess of synonymous substitution as a result of purifying selection. Furthermore, this level of sequence variation suggests that these strains diverged from ~1.83 to 2.12 million years ago, assuming a linear rate of short-term substitution. The *de novo* assemblies of mitochondrial genomes from next-generation sequencing (NGS) reads provide readily available data for similar comparative studies.

Key words: Lepidoptera, *Maruca vitrata*, mitogenome, New World

*E-mail: tolulope.agunbiade@yale.edu

Introduction

The mitochondrial genome (mitogenome) encodes proteins that are involved in electron transport chain and oxidative phosphorylation that are essential for the energy production of the cell (Wolstenholme, 1992; Cameron, 2014). Insect mitogenomes are circular double-stranded DNA (dsDNA) molecules that typically range in size from 14 to 20 kb (Boyce *et al.*, 1989; Wolstenholme, 1992) and contain a conserved set of 13 protein coding genes (PCGs), 22 transfer RNAs (tRNAs), and 2 ribosomal RNA genes (rRNAs) that are required for the translation of mitogenome-encoded proteins (Boore, 1999). Insect mitogenomes usually contain a large A+T-rich, non-coding control region that functions in the initiation of transcription and the replication of the mtDNA (Wolstenholme, 1992; Cameron, 2014), where the length is highly variable among different lineages, due to high rates of nucleotide substitution, insertions/deletions, and variable numbers of tandem repeats (Fauron and Wolstenholme, 1980; Inohira *et al.*, 1997). Animal mitogenomes have a highly conserved gene content, maternal inheritance, and lack recombination (Avise, 2000), which have facilitated their use in the studies of phylogenetic relationships, population genetic structuring, as well as comparative evolution (Zhang *et al.*, 1995; Nardi *et al.*, 2003; Arunkumar *et al.*, 2006). Comparisons between various insect mitogenomes have enabled the estimation of short-term evolutionary patterns within and between closely related species (Ballard, 2000; Coates *et al.*, 2005). Although the gene content of the metazoan mitogenomes is highly conserved, variable lengths of A+T-rich coding regions, gene order, and orientation of tRNAs and PCGs are observed among insects (Hong *et al.*, 2009).

To date, complete or near complete mitogenomes have been sequenced from more than 600 arthropod species (Cameron, 2014). Although the insect order Lepidoptera contains approximately 200,000 species, 72 complete or near complete mitogenomes representing 17 families in the lepidopteran lineage Ditrysia are available (Cameron, 2014). Among the family Crambidae, mitogenomes have been obtained for eight species: *Ostrinia nubilalis* and *O. furnacalis* (Coates *et al.*, 2005), *Diatraea saccharalis* (Li *et al.*, 2011), *M. vitrata* (Old World) (Margam *et al.*, 2011b), *Cnaphalocrocis medinalis* (Chai *et al.*, 2012), *Chilo suppressalis* (Chai *et al.*, 2012), *Paracymoriza prodigalis* (Ye *et al.*, 2013), and *Spoladea recurvalis* (He *et al.*, 2015). Complete mitogenome sequences for Lepidoptera have been used to study the divergences between sibling and congeneric species (Yukuhiro *et al.*, 2002; Coates *et al.*, 2005) and to facilitate population-level studies (Kim *et al.*, 2006). Furthermore, full mitogenomes for Lepidoptera are

important for the determination of phylogenetic relationships and the identification of cryptic species (Lee *et al.*, 2006; Chai *et al.*, 2012) at various taxonomic levels.

Obtaining full mitogenome sequences has been a challenging and resource demanding task when sequenced by long-range PCR and subsequent primer walking (Huyse *et al.*, 2007) or overlapping short polymerase chain reaction (PCR) products (Coates *et al.*, 2005). Recently, next-generation sequencing (NGS) technologies have led to more straightforward pipelines for the assembly of complete mitogenome sequences (Jex *et al.*, 2010a; Knaus *et al.*, 2011). Mitogenomes have been sequenced from NGS libraries prepared from long-PCR amplified sequences (Maricic *et al.*, 2010; Morin *et al.*, 2010; Horn *et al.*, 2011) or from sequence captured DNA molecules (Vasta *et al.*, 2009). Long-range PCR strategies may remain useful for NGS-based sequencing, especially when limited genetic material can be recovered from small invertebrates (Jex *et al.*, 2010b). Ultra-deep sequencing on NGS platforms can acquire data that are sufficient for the subsequent assembly of full mitogenomes at high read depth (Knaus *et al.*, 2011). Also, due to the high copy numbers of mitochondria in most eukaryotic cells, NGS can reduce the impact of rare nuclear-integrated copies of mitochondrial DNAs or other non-target products (Gilbert *et al.*, 2007; Ho and Gilbert, 2010).

Members from the genus *Maruca* (Lepidoptera: Crambidae) are serious insect pests that feed on cowpea crops throughout the tropical and subtropical regions, from northern Australia and East Asia through sub-Saharan Africa to the Caribbean, Central America and Hawaii (Sharma, 1998). Feeding by *M. vitrata* causes considerable crop loss and reduced yields to cultivated cowpea in many African countries, and is a target for improved and sustainable control methods (Agunbiade *et al.*, 2012). The comparison of *M. vitrata* mitochondrial cytochrome c oxidase subunit I (*cox1*) sequence data obtained from samples collected in Africa and Puerto Rico revealed unexpectedly high levels of sequence variation, which suggested the existence of cryptic species or undefined strains (Margam *et al.*, 2011a). A recent study also confirmed the existence of different subspecies of *M. vitrata* in Asia and Africa and *Maruca* spp. in Oceania and Latin America based on *cox1* data (Malini *et al.*, 2015). An analogous relationship among New and Old World species has also been predicted between *Helicoverpa zea* and *H. armigera*, the corn earworm, and cotton bollworm (Behere *et al.*, 2007), but accurate estimates of time since divergence between these species are lacking. As a result of the lack of morphological characters for subspecies sorting between the Old and New World strains of *M. vitrata*, Illumina HiSeq

2000 sequencing reads were obtained from *M. vitrata* (New World) and used to successfully assemble the full mitogenome sequence. These data were applied to refine previous haplotype divergence estimates and provide an estimated time of divergence between the New and Old World strains, in ongoing efforts to describe the evolutionary divergence in *M. vitrata*.

Materials and methods

Sample collection and DNA extraction

Insect samples of *M. vitrata* adults were collected from white bean (*Phaseolus vulgaris*) near Lares, Puerto Rico. All specimens were preserved in 100% ethanol and stored at 4 °C. Total genomic DNA was extracted from the adult specimens using the DNeasy animal tissue kit following the manufacturer's instructions (QIAGEN, Valencia, California). The quality of the DNA was assessed through electrophoresis on a 1% agarose gel before submission for library construction and sequencing.

Illumina HiSeq 2000™ library construction and sequencing

A shotgun genomic DNA library was constructed using the TruSeq DNA sample prep kit (Illumina, San Diego, California). Briefly, 1 µg of genomic DNA was sonicated on a Covaris M220 (Covaris, Woburn, Massachusetts) using the default methods for sonication to a size of 500 bp library, as described by the manufacturer. Sonicated DNA was blunt-ended, 3'-end A-tailed and ligated to an indexed adaptor. The adaptor-ligated gDNA was amplified by PCR to selectively enrich for those fragments that have adapters on both ends. Amplification was carried out for eight cycles with the Kapa HiFi polymerase (Kapa Biosystems, Woburn, Massachusetts), to reduce the likeliness of multiple identical reads, due to preferential clonal amplification. The library was size selected on a 2% agarose gel for fragments 600 bp to 800 bp in length. The size-selected library was run on Agilent bioanalyzer DNA 7500 LabChips (Agilent, Santa Clara, California), to determine the average fragment size and to confirm the presence of DNA of the expected size range, and quantitated by qPCR on an ABI 7900. The indexed library was loaded onto a single lane of an eight-lane flowcell for cluster formation and paired-end sequenced at 100-bp lengths on an Illumina HiSeq 2000™ using the TruSeq SBS sequencing kits version 3. The raw .bcl files were converted into FASTQ files using the software Casava 1.8.2 (Illumina). All shotgun genomic DNA library construction and Illumina HiSeq 2000™ sequencing were carried out at the W. M. Keck Center for Comparative and Functional

Genomics, Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign.

Mitogenome assembly

FASTQ output was trimmed of read data with PHRED quality scores < 20 using the script TrimmingReads.pl (Patel and Jain, 2012; NGSToolKit). A total of 30 million reads from the trimmed FASTQ sequence were loaded into the Velvet Assembler program v.1.2.10 (Zerbino and Birney, 2008) and assembled using the de Bruijn graph method with a hash size (k -mer) = 21, no coverage cutoff, and a minimum contig length ($-\text{min_contig_lgth}$) = 10,000. Querying the National Center for Biotechnology Information (NCBI) non-redundant (nr) nucleotide database with the blastn algorithm (Altschul *et al.*, 1997; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) identified resulting contigs containing mitogenome sequence. The contigs that produced database 'hits' were filtered for those with E -values $\geq 10^{-40}$ and percent similarities ≥ 75 to identify contigs with encoded mitochondrial gene sequences.

Sequence annotation

Coding sequences for each *M. vitrata* (New World) mitochondrial gene were aligned to mitochondrial gene sequences from crambid mitogenomes (GenBank Accessions: JF339041, JN246082, FJ240227, HM751150, AF467260, NC_020094, and KJ739310) (Supplementary Table 1) using the CLC Genomics Workbench v6.5.2. The 5' end of the mitochondrial genes was inferred to be the first legitimate start codon in the open reading frame, and manual adjustments were made when needed. The 3' end was inferred to be at the first in-frame stop codon encountered. When the stop codon was located within the sequence of a downstream gene encoded on the same strand, a truncated stop codon (T) adjacent to the beginning of the downstream gene was designated as the termination codon. The positions and secondary structures for the tRNA genes were confirmed using ARWEN v1.2 (Laslett and Canbäck, 2008; <http://mbio-serv2.mbioekol.lu.se/ARWEN/index.html>). The graphical circular structure of the mitogenome was drawn using GenomeVx, a suite of tools for generating physical maps of plastid and mitogenomes (Conant and Wolfe, 2008; <http://wolfe.ucd.ie/GenomeVx/>). All annotations were transferred to Sequin 13.05 (<http://www.ncbi.nlm.nih.gov/Sequin/>) and submitted to the GenBank database under the accession number KJ466365.

Genome sequence analysis

The nucleotide sequences, codon usage, and relative synonymous codon usage (RSCU) in the *M.*

vitrata (New World) mitogenome were calculated with MEGA 5.2.2 (Tamura *et al.*, 2004). Nucleotide composition skew analysis was carried out by the formula: AT skew = $[A-T]/[A+T]$ and GC skew = $[G-C]/[G+C]$, respectively (Perna and Kocher, 1995).

The DNA and protein sequence identities were calculated across the Crambidae using the FASTA program, a tool for biological sequence comparison (Pearson and Lipman, 1988; http://fasta.bioch.virginia.edu/fasta_www2/). The online program, microsatellite repeats finder (http://insilico.ehu.es/mini_tools/microsatellites/) was used to analyse the A+T-rich control region for tandem repeats (repeat units ≥ 2 nucleotides and arrays of $4 \geq$ units).

Estimation of divergence times

The nucleotide sequence of *M. vitrata* (Old World; GenBank accession number HM751150) was downloaded in the FASTA format and aligned with that of the *M. vitrata* New World sample (GenBank accession number KJ466365) using MEGA 5.2.2 (Tamura *et al.*, 2004). Gaps in the resulting alignment due to ambiguous bases (Ns) in HM751150 overhangs were deleted. The number of nucleotide differences between the two aligned sequences were calculated using DiffSeq from the EMBOSS Package (Rice *et al.*, 2000). A clock-like rate of 2% mitochondrial DNA sequence change per million years has been estimated among *Drosophila* species (Powell *et al.*, 1986) and a 1.115% per million year estimate was proposed among Lepidoptera (Brower, 1994), which to date is the divergence time based on the substitution rate across a 11,250 bp of the consensus alignment between *M. vitrata* strains (inter-species substitution rate/clock rate).

Results and Discussion

Mitogenome assembly and annotation

Paired-end Illumina sequencing generated a total of 336,472,660 reads, of which 30 million were subsampled and used to generate a 15,385 bp *de novo* assembly of the previously uncharacterized *M. vitrata* (New World) mitogenome (Table 1 and Fig. 1). The mitogenome of *M. vitrata* (New World) was similar to the size previously reported from other Lepidoptera, that range from 15,314 bp in *Coreana raphaelis* (Lepidoptera: Lycaenidae) (Kim *et al.*, 2006) to 15,928 bp in *Bombyx mandarina* (Lepidoptera: Bombycidae) (Yukuhiro *et al.*, 2002). The *M. vitrata* (New World) mitogenome contained the 13 PCGs, 22 tRNAs, and 2 rRNA genes that are typical for bilaterian metazoans (Wolstenholme, 1992; Table 1 and Fig. 1). As in all other lepidopteran

mitogenomes, of the total 37 genes, 23 were encoded on the major strand, whereas the 14 others were encoded on the minor strand (Table 1 and Fig. 1). The 13 PCGs spanned 11,369 bp and accounted for 73.90% of the entire sequence. The *M. vitrata* (New World) mitogenome had the typical Ditrysian gene order *trnM-trnI-trnQ-nad2* (Yukuhiro *et al.*, 2002; Coates *et al.*, 2005; Kim *et al.*, 2006; Lee *et al.*, 2006; Cameron and Whiting, 2008; Hong *et al.*, 2008), which differed from the ancestral gene order *trnI-trnE-trnM-nad2* found in other insects (Cameron, 2014). This was first reported by Taylor *et al.* (1993) in lycaenids and noctuids, but has now been found in five lepidopteran superfamilies (Papilionoidea, Noctuoidea, Pyraloidea, Tortricoidea, and Bombycoidea) and a synapomorphy of Ditrysia (Salvato *et al.*, 2008).

The *M. vitrata* (New World) mitogenome showed a high A+T content 80.70% (A = 40.30%, T = 40.4%, G = 7.99%, and C = 11.31%) (Table 2), and this is consistent with the high A+T content of insect (Clary and Wolstenholme, 1985; Crozier and Crozier, 1993) and lepidopteran mitogenomes (Liu *et al.*, 2008). The nucleotide composition of the PCGs and the AT- and GC-skew calculated for all available crambid mtDNA genomes are presented in Table 2. The average AT-skew of the crambid mitogenomes was 0.006, ranging from -0.024 to 0.03 , whereas the average GC-skew was -0.195 , ranging from -0.26 to -0.17 . The AT-skew of *M. vitrata* (New World) was -0.002 . A similar negative AT-skew was also previously reported for *C. medinalis* (Chai *et al.*, 2012) and *Spoladea recurvalis* (He *et al.*, 2015) (Table 2). This indicates a bias for T over A nucleotides. Consistent with other lepidopteran mitogenomes that exhibit negative GC skewness, the GC-skew of *M. vitrata* (New World) was -0.17 , indicating a bias for C over G nucleotides.

Protein coding genes

The putative start codons of PCGs used by *M. vitrata* (New World) (ATN; Table 1) were identical to those previously known in animal mitogenomes (Wolstenholme, 1992). CGA was used as the start codon for *cox1* as previously reported for lepidopterans such as *Bombyx mori* (Yukuhiro *et al.*, 2002), *O. nubilalis* and *O. furnacalis* (Coates *et al.*, 2005), *S. recurvalis* (He *et al.*, 2015), *Adoxophyes honmai* (Lee *et al.*, 2006), *C. raphaelis* (Kim *et al.*, 2006), *Antheraea pernyi* (Liu *et al.*, 2008), *B. mandarina* (Pan *et al.*, 2008), *Ochrogaster lunifer* (Salvato *et al.*, 2008), *Artogeia melete* (Hong *et al.*, 2009), *Eriogyna pyretorum* (Jiang *et al.*, 2009), *Hyphantria cunea* (Liao *et al.*, 2010) and *M. vitrata* (Old World) (Margam *et al.*, 2011b). The termination codon TAA was used in all PCGs with the exception of *nad3*, which had TAG. Also, consistent with some other insect mitogenomes,

Table 1. Annotation of the full mitogenome of *Maruca vitrata* (New World)

Gene	Strand	CDS positions	Size (bp)	Start codon	Stop codon	Anticodon
<i>trnM</i>	J	1..68	68			CAT
<i>trnI</i>	J	69..133	65			GAT
<i>trnQ</i>	N	130..200	71			TTG
<i>nad2</i>	J	269..1255	987	ATA	TAA	
<i>trnW</i>	J	1257..1324	68			TCA
<i>trnC</i>	N	1316..1382	67			GCA
<i>trnT</i>	N	1383..1451	69			GTA
<i>cox1</i>	J	1458..2988	1531	CGA	T-	
<i>trnL1</i>	J	2988..3056	69			TAA
<i>cox2</i>	J	3056..3737	682	ATG	T-	
<i>trnK</i>	J	3737..3809	73			CTT
<i>trnD</i>	J	3811..3878	68			GTC
<i>atp8</i>	J	3879..4040	162	ATC	TAA	
<i>atp6</i>	J	4034..4708	675	ATG	TAA	
<i>cox3</i>	J	4715..5503	789	ATG	TAA	
<i>trnG</i>	J	5505..5573	69			TCC
<i>nad3</i>	J	5570..5926	357	ATA	TAG	
<i>trnA</i>	J	5924..5992	69			TGC
<i>trnR</i>	J	5990..6054	65			TCG
<i>trnN</i>	J	6054..6119	66			GTT
<i>trnS1</i>	J	6122..6189	68			GCT
<i>trnQ</i>	J	6189..6256	68			TTC
<i>trnF</i>	N	6255..6324	70			GAA
<i>nad5</i>	N	6307..8058	1752	ATT	TAA	
<i>trnH</i>	N	8058..8126	69			GTG
<i>nad4</i>	N	8125..9465	1341	ATG	TAA	
<i>nad4L</i>	N	9465..9746	282	ATA	TAA	
<i>trnT</i>	J	9758..9823	66			TGT
<i>trnP</i>	N	9823..9890	68			TGG
<i>nad6</i>	J	9892..10422	531	ATT	TAA	
<i>cytb</i>	J	10426..11574	1149	ATG	TAA	
<i>trnS2</i>	J	11573..11638	66			TGA
<i>nad1</i>	N	11660..12790	1131	ATG	TAA	
<i>trnL2</i>	N	12791..12859	69			TAG
<i>rrnL</i>	N	12894..14197	1304			
<i>trnV</i>	N	14210..14276	67			TAC
<i>rrnS</i>	N	14280..15044	765			
A+T-rich region		15045..15385	341			

J, major strand; N, minor strand.

Table 2. Nucleotide composition in the whole mitogenomes of *Maruca vitrata* (New World) and other available mitogenomes from six additional species from the lepidopteran Family Crambidae

Species	A	T	A+T	AT-Skew	G	C	G+C	GC-Skew
<i>Cnaphalocrocis medinalis</i>	40.36	41.58	81.94	-0.01	7.45	10.61	18.06	-0.17
<i>Chilo suppressalis</i>	40.64	40.03	80.67	0.01	7.39	11.94	19.33	-0.24
<i>Diatraea saccharalis</i>	40.87	39.15	80.02	0.02	7.42	12.56	19.98	-0.26
<i>Ostrinia furnacalis</i> ¹	41.46	38.92	80.38	0.03	7.91	11.71	19.62	-0.19
<i>Ostrinia nubilalis</i> ¹	41.36	38.81	80.17	0.03	8.02	11.82	19.83	-0.19
<i>Paracymoriza prodigalis</i>	40.84	40.71	81.55	0.002	7.54	10.91	18.45	-0.18
<i>Spoladea recurvalis</i> ²	39.48	41.43	80.91	-0.024	7.79	11.3	19.09	-0.18
<i>Maruca vitrata</i> (Old World) ¹	40.14	39.89	80.03	0.003	8.21	11.76	19.97	-0.18
<i>Maruca vitrata</i> (New World)	40.27	40.44	80.7	-0.002	7.99	11.31	19.3	-0.17

¹Partial mitogenome sequences that lacked portions of the AT-rich control region sequence.

²Partial mitogenome sequences for which the *cox1* gene sequence is publicly unavailable.

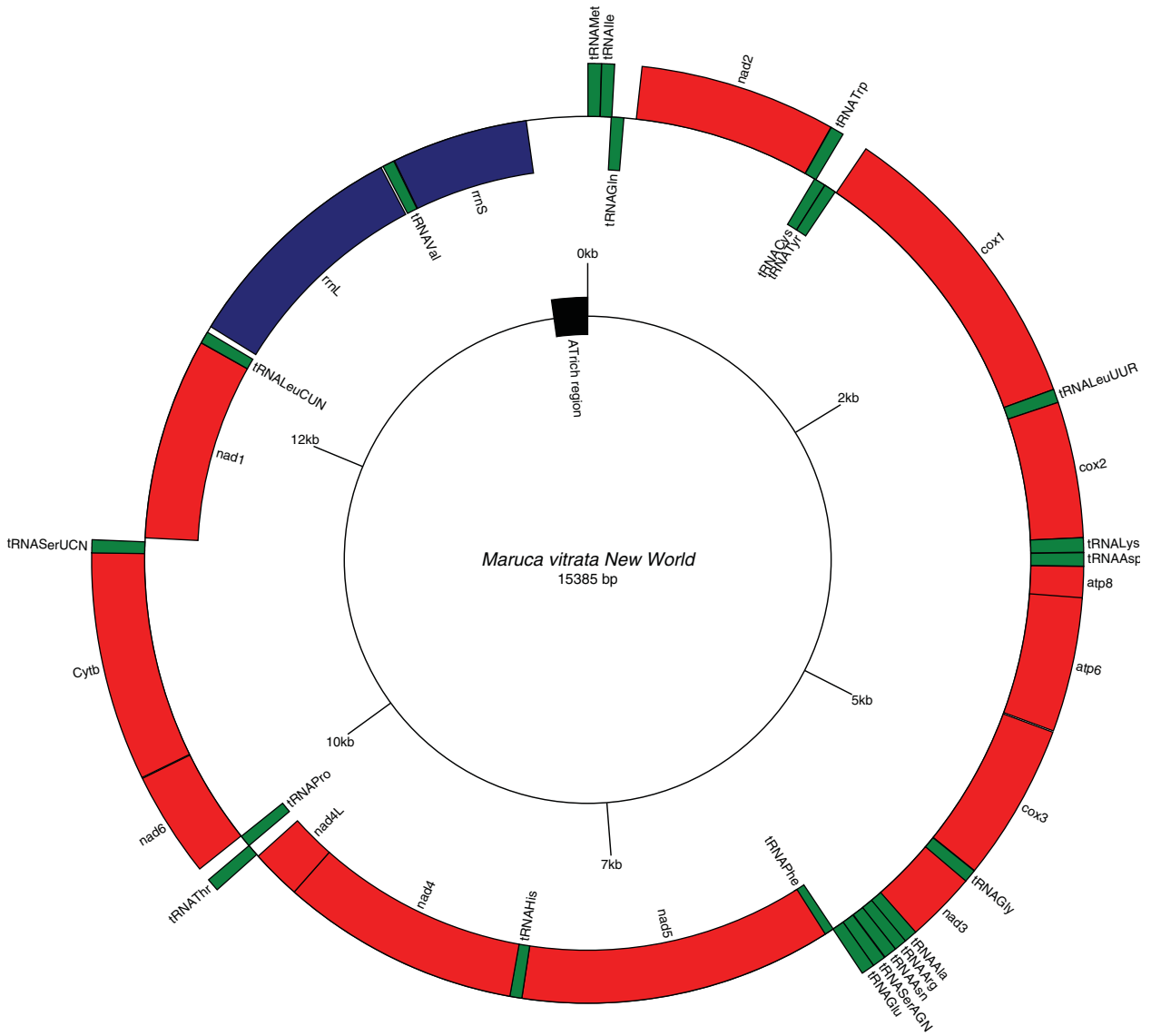


Fig. 1. Gene arrangements in the circular mitogenome of *Maruca vitrata* (New World). Red colour indicates the protein coding genes; green, tRNAs; and blue, rRNAs. Inner and outer circles represent genes oriented on major and minor strands, respectively.

cox1 and *cox2* both had incomplete (T-) stop codon (Coates *et al.*, 2005; Salvato *et al.*, 2008) (Table 1). The presence of incomplete stop codons is a molecular feature shared with all lepidopteran mitogenomes sequenced to date (Brower, 1994; Yukuhiro *et al.*, 2002; Coates *et al.*, 2005; Kim *et al.*, 2006; Lee *et al.*, 2006; Cameron and Whiting, 2008; Hong *et al.*, 2008) and with many arthropod mitogenomes (Boore, 1999). The motif TTAG has been observed immediately upstream of the putative *cox1* CGA start codon in some lepidopteran mitogenomes and has been proposed to serve in non-standard initiation (Yukuhiro *et al.*, 2002). We observed the motif ATAG immediately upstream of the CGA

start codon of the *M. vitrata* (New World) *cox1* gene. This motif is conserved in some lepidopteran mitogenomes and lacking in others (Margam *et al.*, 2011b). Among the available crambid mitogenomes, *O. furnacalis* and *O. nubilalis* both have the motif TTAG upstream of the start codon of the *cox1* gene, while *M. vitrata* (Old World) has ATAG. Reduction in mitogenome size has led to minimal intergenic space as well as overlap between adjacent genes in *M. vitrata* (New World). *Atp8* and *atp6* PCGs had a seven-nucleotide overlap (Table 1). This feature is common to all lepidopteran mitogenomes known (Brower, 1994; Yukuhiro *et al.*, 2002; Coates *et al.*, 2005; Kim *et al.*, 2006; Lee *et al.*, 2006; Cameron

Table 3. Codon usage and relative synonymous codon usage (RSCU) in *Maruca vitrata* (New World) mitogenome

Amino acid	Codon	Count	RSCU	Amino acid	Codon	Count	RSCU
Ala (A)	GCU	78	2.22	Pro (P)	CCU	75	2.21
	GCC	7	0.21		CCC	9	0.29
	GCA	40	1.21		CCA	43	1.14
	GCG	1	0.05		CCG	2	0.06
Cys (C)	UGU	26	1.31	Gln (Q)	CAA	63	2
	UGC	4	0.23		CAG	0	0
Asp (D)	GAU	65	1.32	Arg (R)	CGU	13	0.9
	GAC	3	0.22		CGC	0	0
Glu (E)	GAA	72	1.72		CGA	41	2.71
	GAG	6	0.13	CGG	1	0.08	
Phe (F)	UUU	344	1.81	Ser1 (S)	AGU	32	0.75
	UUC	32	0.19		AGC	3	0.05
Gly (G)	GGU	67	1.23	Ser2 (S)	AGA	81	1.89
	GGC	4	0.09		AGG	0	0
	GGA	122	2.15		UCU	109	2.77
	GGG	15	0.22		UCC	4	0.06
His (H)	CAU	57	1.34	Thr (T)	UCA	87	2.39
	CAC	12	0.35		UCG	4	0.09
Ile (I)	AUU	443	1.88		ACU	89	2.23
	AUC	26	0.12	ACC	9	0.31	
Lys (K)	AAA	93	1.77	Val (V)	ACA	58	1.12
	AAG	14	0.23		ACG	2	0.04
Leu1 (L)	CUU	27	0.31		GUU	86	2.22
	CUC	3	0.04	GUC	2	0.03	
	CUA	23	0.68	GUA	59	1.71	
	CUG	0	0	GUG	3	0.04	
Leu2 (L)	UUA	481	4.86	Trp (W)	UGA	94	2
	UUG	11	0.12		UGG	0	0
Met (M)	AUA	267	1.83	Try (Y)	UAU	183	1.84
	AUG	23	0.17		UAC	13	0.16
Asn (N)	AAU	238	1.94				
	AAC	8	0.06				

and Whiting, 2008; Hong *et al.*, 2008) and is found in many animal mitogenomes (Boore, 1999). The A+T content of the PCGs was 79.4%, which is lower than the A+T content of the mitogenome as a whole (80.7%). The total number of codons on the majority strand, excluding stop codons, was 3,777 and showed a strong bias toward AT-rich codons, with the four most prevalent codons being *Leu*, UAA (481); *Ile*, AUU (443); *Phe*, UUU (344); and *Met*, AUA (267) (Table 3). The codon usage prevalence was consistent with that found in *M. vitrata* (Old World). Comparison of the 13 PCGs across the eight available crambid mitogenomes with those of *M. vitrata* (New World) resulted in 80.4% and 99.1% identity at the nucleotide level, and 61.8% and 99.4% identity at the amino acid level (Table 4). At both nucleotide and amino acid levels, all the 13 PCGs of *M. vitrata* (New World) showed the highest identity with those of *M. vitrata* (Old World).

Transfer and ribosomal RNA genes

The published mitogenome of *M. vitrata* (Old World) is incomplete and thereby has 19 tRNAs and a part of the rRNA reported. However, the *M. vitrata* (New World) mitogenome encoded the 22 tRNAs typically found in animal mitogenomes (Boore, 1999) and all formed the canonical cloverleaf-like tRNA secondary structure, with the exception of *trnS1*, in which its dihydrouridine (DHU) arm formed a simple loop. The lack of DHU arm in *trnS1* is a common condition in metazoan mitogenomes (Taanman, 1999) but is not found in all lepidopterans. For example, *A. honmai* has all the tRNAs with a complete clover leaf structure (Lee *et al.*, 2006). Of the 22 tRNA genes, 14 genes were encoded on the majority strand, while eight were encoded on the minority strand. The length of tRNAs ranged from 63 bp to 71 bp. Consistent with what is observed in all other insect mitogenomes, the mitogenome of *M.*

Table 4. Comparison of the percent nucleotide and amino acid identity in the 13 protein coding gene sequences of *Maruca vitrata* (New World) with orthologs in 6 other crambid lepidopteran mitogenomes [Mv – *M. vitrata* (Old World), Cs (*C. suppressalis*), Cm (*C. medinalis*), Ds (*D. saccharalis*), Of (*O. furnacalis*), On (*O. nubilalis*), Pp (*Paracymoriza prodigalis*), Sr (*Spoladea recurvalis*)]; *cox1* gene sequence not available)

Genes	Mv		Cs		Cm		Ds	
	DNA (%)	Protein (%)	DNA (%)	Protein (%)	DNA (%)	Protein (%)	DNA (%)	Protein (%)
<i>nad2</i>	97.9	97.9	84.6	80.8	86.5	83.8	78.9	75.0
<i>nad1</i>	95.7	97.7	85.8	87.4	87.5	90.7	82.2	80.8
<i>cox1</i>	94.3	99.4	88.6	95.7	89.5	94.1	86.5	93.9
<i>cox2</i>	94.9	97.8	88.0	94.7	90.0	93.8	85.9	92.5
<i>cox3</i>	94.6	98.5	86.8	92.4	89.0	94.7	86.2	91.6
<i>atp8</i>	93.8	90.6	81.2	64.8	78.6	61.8	75.3	62.3
<i>atp6</i>	93.9	98.7	83.7	85.4	90.4	93.3	79.4	80.5
<i>nad3</i>	94.3	99.1	83.8	83.8	88.4	90.6	80.4	81.2
<i>nad4</i>	95.7	97.5	87.6	84.3	91.6	90.1	84.3	78.3
<i>nad4l</i>	96.1	90.5	89.4	87.1	91.8	88.2	87.9	86.0
<i>nad5</i>	96.0	98.3	85.7	83.6	88.5	87.0	85.3	80.7
<i>nad6</i>	96.8	94.3	82.1	73.4	85.6	80.2	77.7	69.3
<i>cytb</i>	97.5	98.4	86.4	91.1	89.3	91.6	85.7	89.5
Genes	Of		On		Pp		Sr	
	DNA (%)	Protein (%)	DNA (%)	Protein (%)	DNA (%)	Protein (%)	DNA (%)	Protein (%)
<i>nad2</i>	83.9	80.5	83.9	80.8	83.1	78.4	87.3	87.2
<i>nad1</i>	87.4	90.1	87.2	89.7	85.8	84.4	86.9	90.4
<i>cox1</i>	88.6	95.7	87.9	94.3	88.3	93.7	90.2	97.5
<i>cox2</i>	88.1	93.8	87.8	93.0	87.1	91.6	90.9	94.3
<i>cox3</i>	86.4	90.9	86.2	90.9	86.9	89.7	89.0	93.1
<i>atp8</i>	85.8	73.6	85.2	71.7	82.1	66.0	90.1	84.9
<i>atp6</i>	85.6	89.7	85.2	89.7	83.7	86.2	87.4	91.5
<i>nad3</i>	84.4	86.1	84.2	86.0	83.5	80.9	84.4	88.0
<i>nad4</i>	87.9	82.4	87.7	82.6	86.7	81.4	90.2	89.2
<i>nad4l</i>	89.7	82.8	89.0	83.9	88.3	81.7	92.2	89.2
<i>nad5</i>	87.6	84.3	87.2	83.9	86.4	84.1	88.2	88.5
<i>nad6</i>	83.9	75.1	83.3	74.6	83.7	72.9	83.3	77.4
<i>cytb</i>	87.2	90.3	87.5	90.8	86.1	85.6	–	–

*vitrat*a (New World) contained two rRNAs with a total length of 2,069 bp. The large rRNA (*rrnL*) had a length of 1,304 bp, while the small rRNA (*rrnS*) had a length of 765 bp (Table 1).

Non-coding regions

The *M. vitrata* (New World) mitogenome contained characteristics typical of other lepidopteran mitogenomes. The 7 bp intergenic spacer located between *trnS2* and *nad1* contained the ATACTAA motif, which is conserved across the order Lepidoptera (Cameron and Whiting, 2008), including *M. vitrata* (Old World). This motif is possibly fundamental to site recognition by the transcription termination peptide (mtTERM protein) (Taanman, 1999) and is present in most insect mitogenomes (Cameron and Whiting, 2008). Insect mitogenomes often contain a non-coding A+T-rich region that varies considerably in length among insect species, or even within the same species (Zhang and Hewitt,

1997). Consistent with findings from other insects, the largest non-coding region was the A+T-rich control region, although it varies in size in crambids (Hua *et al.*, 2008; Oliveira *et al.*, 2008; Ma *et al.*, 2009). Tandem repeat sequences are common in the control region of insects and can vary in length and copy number (Dotson and Beard, 2001). Repeats in Lepidoptera share greater degrees of conservation than in other insect groups (Cameron and Whiting, 2008). The A+T rich control region in *M. vitrata* (New World) was estimated at 341 bp and contained (AT) tandem repeats. The A+T-rich region contained a conserved ATAG motif followed by an 18 bp poly-T stretch. This feature has been found in *C. medinalis*, *C. suppressalis* and *D. saccharalis*, but the length of the poly-T stretch varies between species. This structure is suggested to function as a signal for mitochondrial DNA replication initiation (Hu *et al.*, 2010; Yin *et al.*, 2010; Kim *et al.*, 2010). A poly-A stretch commonly observed in other lepidopteran mitogenomes was also found immediately upstream of *trnM* in *M.*

*vitrat*a (New World). The (AT)₄, (AATT)₄ and (AT)₉ repeats in the *M. vitrata* control region were predicted (Supplementary Table 2) of which (AT)₉ was similar to an (AT)₈ in *Hyphantria cunea* (Liao *et al.*, 2010) and (AT)₇ in *Ochrogaster lunifer* (Salvato *et al.*, 2008).

Comparative genome analysis and estimation of divergence times

Protein coding sequence was shown to be highly similar between *M. vitrata* (New World) and *M. vitrata* (Old World) mitogenomes (see above; Table 4). Comparison between *M. vitrata* (New World) and *M. vitrata* (Old World) coding sequences found 476 nt substitutions in the 11,250 nt consensus alignment (4.23% divergence), of which 74 (15.5%) were non-synonymous mutations. This suggests that purifying selection may be curbing the rate of amino acid change in *Maruca*. Novel non-synonymous mutations were directly shown to be under negative selection in haplotypes derived from mutator strains of mice (Stewart *et al.*, 2008) such that the accumulation of mildly deleterious mutations might lead to reduced fitness and result in removal from the population. The alignment of entire mitogenome sequences from *M. vitrata* (Old World) and *M. vitrata* (New World) resulted in a 14,039 consensus (gaps excluded) and showed a total of 526 nucleotide differences (3.75%) (Supplementary Table 3). Making a simple assumption of a clock-like rate of 2% change per million years estimated from the *Drosophila* species group (Powell *et al.*, 1986) and with an effective rate of change of 1.15% per million years for Lepidoptera (Brower, 1994) and acknowledging the potential that variance in this rate may have affected the degree of haplotype divergence among *Maruca* mitochondrial genomes, the two strains may have diverged ~1.83 to 2.12 mya. These estimates likely suffer from uneven rates across taxonomic groups and time (Ho and Lo, 2013). Regardless, evolutionary rates have been estimated from calibrated molecular clocks (Papadopoulou *et al.*, 2010) and have resulted in estimates similar to the standard rate (Brower, 1994). With lack of fossil evidence to accurately calibrate the accumulation of substitutions across the lepidopteran mitochondrial genome lineages, the derivation of an accurate divergence time estimate between *M. vitrata* strains remains elusive and beyond the scope of this study.

Conclusion

The 15,385 bp mitogenome of *M. vitrata* (New World) has a size and gene order typical of mitogenomes from the insect order Lepidoptera. The aim of this study was achieved by defining the

previously undetermined levels of mitochondrial genome variation between strains of *M. vitrata*. In doing so, a substitution bias at non-synonymous sites was defined, which is typical for highly conserved mitochondrial PCG, including the previously published mitogenome *coxI* by Margam *et al.* (2011b). This short-term change identified here provides valuable information regarding variation between geographically distinct strains of a species that is analogous to that shown by comparisons between other closely related insect species (Coates *et al.*, 2005; Coates, 2014). Specifically, these whole mitochondrial genome analyses provide estimates of divergence time and short-term patterns of sequence evolution that are more accurate compared to the use of single mitochondrial gene sequence data. Although not explored in the current study, the rapid *de novo* assembly of mitogenomes from NGS sequence data presents future opportunities for phylogenomic studies within and among species of Lepidoptera.

Acknowledgements

This project has been made possible through support provided by the Legumes Innovation Lab, formerly known as the Dry Grains Pulses Collaborative Research Support Program (CRSP), by the Bureau for Economic Growth, Agriculture, and Trade, US Agency for International Development, under the terms of Grant No. EDH-A-00-07-00005. The United States Department of Agriculture, Agricultural Research Service (USDA-ARS: CRIS Project 3625-22000-017-00) and the Iowa Agriculture 567 and Home Economics Experiment Station, Ames, IA (Project 3543) supported a portion of research data collection and analysis. TAA was supported by International Student Research Fellowship from the Howard Hughes Medical Institute, Maryland, USA.

Supplementary material

To view supplementary material for this article, please visit <https://doi.org/10.1017/S1742758417000157>.

References

- Agunbiade T., Steele L., Coates B., Gassmann A., Margam V. M., Ba M., Dabire-Binso C., Baoua I., Bello-Bravo J., Seufferheld F., Sun W., Tamò M. and Pittendrigh B. R. (2012) IPM-omics: from genomics to extension for integrated pest management of cowpea, pp. 231–248. In *Improving Livelihoods in the Cowpea Value Chain through Advancements in Science* (edited by O. Boukar, O. Coulibaly, C. Fatokun, K. Lopez and M. Tamò).

- International Institute of Tropical Agriculture, Ibadan, Nigeria.
- Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W. and Lipman D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25, 3389–3402.
- Arunkumar K. P., Metta M. and Nagaraju J. (2006) Molecular phylogeny of silkmoths reveals the origin of domesticated silkmoth, *Bombyx mori* from Chinese *Bombyx mandarina* and paternal inheritance of *Antheraea proylei* mitochondrial DNA. *Molecular Phylogenetics and Evolution* 40, 419–427.
- Avise J. C. (2000) *Phylogeography—The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts. 447 pp.
- Ballard J. W. O. (2000) Comparative genomics of mitochondrial DNA in *Drosophila simulans*. *Journal of Molecular Biology* 51, 64–75.
- Behere G. T., Tay W. T., Russell D. A., Heckel D. G., Appleton B. R., Kranthi K. R. and Batterham P. (2007) Mitochondrial DNA analysis of field populations of *Helicoverpa armigera* (Lepidoptera: Noctuidae) and of its relationship to *H. zea*. *BMC Evolutionary Biology* 7, 117.
- Boore J. L. (1999) Animal mitochondrial genomes. *Nucleic Acids Research* 27, 1767–1780.
- Boyce T. M., Zwick M. E. and Aquadro C. F. (1989) Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. *Genetics* 123, 825–836.
- Brower A. V. (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the USA* 91, 6491–6495.
- Cameron S. L. (2014) Insect mitochondrial genomics: implications for evolution and phylogeny. *Annual Review of Entomology* 59, 95–117.
- Cameron S. L. and Whiting M. F. (2008) The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta* (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. *Gene* 408, 112–123.
- Chai H. N., Du Y. Z. and Zhai B. P. (2012) Characterization of the complete mitochondrial genomes of *Cnaphalocrocis medinalis* and *Chilo suppressalis* (Lepidoptera: Pyralidae). *International Journal of Biological Sciences* 8, 561–579.
- Clary D. O. and Wolstenholme D. R. (1985) The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *Journal of Molecular Evolution* 22, 252–271.
- Coates B. S. (2014) Assembly and annotation of full mitochondrial genomes for the corn rootworm species, *Diabrotica virgifera virgifera* and *Diabrotica barberi* (Insecta: Coleoptera: Chrysomelidae), using next generation sequence data. *Gene* 542, 190–197.
- Coates B. S., Sumerford D. V., Hellmich R. L. and Lewis L. C. (2005) Partial mitochondrial genome sequences of *Ostrinia nubilalis* and *Ostrinia furnicalis*. *International Journal of Biological Sciences* 1, 13–18.
- Conant G. C. and Wolfe K. H. (2008) GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* 24, 861–862.
- Crozier R. H. and Crozier Y. C. (1993) The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* 133, 97–117.
- Dotson E. M. and Beard C. B. (2001) Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. *Insect Molecular Biology* 10, 205–215. doi:10.1046/j.1365-2583.2001.00258.x.
- Fauron C. M. and Wolstenholme D. R. (1980) Extensive diversity among *Drosophila* species with respect to nucleotide sequences within the adenine + thymine-rich region of mitochondrial DNA molecules. *Nucleic Acids Research* 8, 2439–2452.
- Gilbert M. T. P., Tomsho L. P., Rendulic S., Packard M., Drautz D. I., Sher A., Tikhonov A., Dalén L., Kuznetsova T., Kosintsev P., Campos P. F., Higham T., Collins M. J., Wilson A. S., Shidlovskiy F., Buigues B., Ericson P. G. P., Germonpré M., Götherström A., Iacumin P., Nikolaev V., Nowak-Kemp M., Willerslev E., Knight J. R., Irzyk G. P., Perbost C. S., Fredrikson K. M., Harkins T. T., Sheridan S., Miller W. and Schuster S. C. (2007) Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* 317, 1927–1930. doi: 10.1126/science.1146971.
- He S.-L., Zou Y., Zhang L.-F., Ma W.-Q., Zhang X.-Y. and Yue B.-S. (2015) The complete mitochondrial genome of the beet webworm, *Spoladea recurvalis* (Lepidoptera: Crambidae) and its phylogenetic implications. *PLoS ONE* 10(6), e1029355.
- Ho S. Y. W. and Gilbert M. T. P. (2010) Ancient mitogenomics. *Mitochondrion* 10, 1–11.
- Ho S. Y. W. and Lo N. (2013) The insect molecular clock. *Austral Entomology* 52, 101–105. doi:10.1111/aen.12018.
- Hong G., Jiang S., Yu M., Yang Y., Li F., Xue F. and Wei Z. (2009) The complete nucleotide sequence of the mitochondrial genome of the cabbage butterfly, *Artogeia melete* (Lepidoptera: Pieridae). *Acta Biochimica et Biophysica Sinica (Shanghai)* 41, 446–455.
- Hong M. Y., Lee E. M., Jo Y. H., Park H. C., Kim S. R., Hwang J. S., Jin B. R., Kang P. D., Kim K.-G., Han Y. S. and Kim I. (2008) Complete nucleotide sequence and organization of the mitogenome of the silk moth *Caligula boisduvalii* (Lepidoptera: Saturniidae) and comparison with other lepidopteran insects. *Gene* 413, 49–57.
- Horn S., Durka W., Wolf R., Ermala A., Stubbe A., Stubbe M. and Hofreiter M. (2011) Mitochondrial genomes reveal slow rates of molecular evolution and the timing of speciation in beavers (*Castor*), one of the largest rodent species. *PLoS ONE* 6(1), e14622. doi:10.1371/journal.pone.0014622.

- Hu J., Zhang D., Hao J., Huang D., Cameron S. and Zhu C. (2010) The complete mitochondrial genome of the yellow coaster, *Acraea issoria* (Lepidoptera: Nymphalidae: Heliconiinae: Acraeini): sequence, gene organization and a unique tRNA translocation event. *Molecular Biology Reports* 37, 3431–3438.
- Hua J., Li M., Dong P., Cui Y., Xie Q. and Bu W. (2008) Comparative and phylogenomic studies on the mitochondrial genomes of Pentatomomorpha (Insecta: Hemiptera: Heteroptera). *BMC Genomics* 9, 610. doi: [10.1186/1471-2164-9-610](https://doi.org/10.1186/1471-2164-9-610).
- Huyse T., Plaisance L., Webster B. L., Mo T. A., Bakke T. A., Bachmann L. and Littlewood D. T. (2007) The mitochondrial genome of *Gyrodactylus salaris* (Platyhelminthes: Monogenea), a pathogen of Atlantic salmon (*Salmo salar*). *Parasitology* 134, 739–747.
- Inohira K., Hara T. and Matsuura E. T. (1997) Nucleotide sequence divergence in the A+T-rich region of mitochondrial DNA in *Drosophila simulans* and *Drosophila mauritiana*. *Molecular Biology and Evolution* 14, 814–822.
- Jex A. R., Hall R. S., Littlewood D. T. J. and Gasser R. B. (2010a) An integrated pipeline for next-generation sequencing and annotation of mitochondrial genomes. *Nucleic Acids Research* 38, 522–533.
- Jex A. R., Littlewood D. T. J. and Gasser R. B. (2010b) Toward next-generation sequencing of mitochondrial genomes—focus on parasitic worms of animals and biotechnological implications. *Biotechnology Advances* 28, 151–159.
- Jiang S. T., Hong G. Y., Yu M., Li N., Yang Y., Liu Y. Q. and Wei Z. J. (2009) Characterization of the complete mitochondrial genome of the giant silkworm moth, *Eriogyna pyretorum* (Lepidoptera: Saturniidae). *International Journal of Biological Sciences* 5, 351–365.
- Kim I., Lee E. M., Seol K. Y., Yun E. Y., Lee Y. B., Hwang J. S. and Jin B. R. (2006) The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis* (Lepidoptera: Lycaenidae). *Insect Molecular Biology* 15, 217–225.
- Kim M. J., Wan X., Kim K.-G., Hwang J. S. and Kim I. (2010) Complete nucleotide sequence and organization of the mitogenome of endangered *Eumenis autonoe* (Lepidoptera: Nymphalidae). *African Journal of Biotechnology* 9, 735–754.
- Knaus B. J., Cron R., Liston A., Pilgrim K. and Schwartz M. K. (2011) Mitochondrial genome sequences illuminate maternal lineages of conservation concern in a rare carnivore. *BMC Ecology* 11, 10. doi: [10.1186/1472-6785-11-10](https://doi.org/10.1186/1472-6785-11-10).
- Laslett D. and Canbäck B. (2008) ARWEN: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* 24, 172–175.
- Lee E.-S., Shin K. S., Kim M.-S., Park H., Cho S. and Kim C.-B. (2006) The mitochondrial genome of the smaller tea tortrix, *Adoxophyes honmai* (Lepidoptera: Tortricidae). *Gene* 373, 52–57.
- Li H., Gao J., Liu Ha., Liu Hu, Liang A., Zhou X. and Cai W. (2011) The architecture and complete sequence of mitochondrial genome of an assassin bug *Agriosphodrus dohrni* (Hemiptera: Reduviidae). *International Journal of Biological Sciences* 7, 792–804.
- Liao F., Wang L., Wu S., Li Y. P., Zhao L., Huang G.-M., Niu C.-J., Liu Y.-Q. and Li M.-G. (2010) The complete mitochondrial genome of the fall webworm, *Hyphantria cunea* (Lepidoptera: Arctiidae). *International Journal of Biological Sciences* 6, 172–186.
- Liu Y., Li Y., Pan M., Dai F., Zhu X., Lu C. and Xiang Z. (2008) The complete mitochondrial genome of the Chinese oak silkworm, *Antheraea pernyi* (Lepidoptera: Saturniidae). *Acta Biochimica et Biophysica Sinica (Shanghai)* 40, 693–703.
- Ma C., Liu C., Yang P. and Kang L. (2009) The complete mitochondrial genomes of two band-winged grasshoppers, *Gastrimargus marmoratus* and *Oedaleus asiaticus*. *BMC Genomics* 10, 156. doi: [10.1186/1471-2164-10-156](https://doi.org/10.1186/1471-2164-10-156).
- Malini P., Schafleitner R., Muthukalingan K. and Ramasamy S. (2015) Phylogeographical structure in mitochondrial DNA of legume pod borer (*Maruca vitrata*) population in tropical Asia and sub-Saharan Africa. *PLoS ONE* 10(4), e0124057.
- Margam V. M., Coates B. S., Ba M. N., Sun W., Binso-Dabire C. L., Baoua I., Ishiyaku M. F., Shukle J. T., Hellmich R. L., Covas F. G., Ramasamy S., Armstrong J., Pittendrigh B. R. and Murdock L. L. (2011a) Geographic distribution of phylogenetically-distinct legume pod borer, *Maruca vitrata* (Lepidoptera: Pyraloidea: Crambidae). *Molecular Biology Reports* 38, 893–903.
- Margam V. M., Coates B. S., Hellmich R. L., Agunbiade T., Seufferheld M. J., Sun W., Ba M. N., Sanon A., Binso-Dabire C. L., Baoua I., Ishiyaku M. F., Covas F. G., Srinivasan R., Armstrong J., Murdock L. L. and Pittendrigh B. R. (2011b) Mitochondrial genome sequence and expression profiling for the legume pod borer *Maruca vitrata* (Lepidoptera: Crambidae). *PLoS ONE* 6(2), e16444. <https://doi.org/10.1371/journal.pone.0016444>.
- Maricic T., Whitten M. and Pääbo S. (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* 5(11), e14004. <https://doi.org/10.1371/journal.pone.0014004>.
- Morin P. A., Archer F. I., Foote A. D., Vilstrup J., Allen E. E., Wade P., Durban J., Parsons K., Pitman R., Li L., Bouffard P., Abel Nielsen S. C. A., Rasmussen M., Willerslev E., Gilbert M. T. P. and Harkins T. (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research* 20, 908–916.
- Nardi F., Carapelli A., Dallai R. and Frati F. (2003) The mitochondrial genome of the olive fly, *Bactrocera oleae*: two haplotypes from distant geographic locations. *Insect Molecular Biology* 12, 605–611.

- Oliveira D. C. S. G., Raychoudhury R., Lavrov D. V. and Werren J. H. (2008) Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp *Nasonia* (Hymenoptera: Pteromalidae). *Molecular Biology and Evolution* 25, 2167–2180.
- Pan M., Yu Q., Xia Y., Dai F., Liu Y., Lu C., Zhang Z. and Xiang Z. (2008) Characterization of mitochondrial genome of Chinese wild mulberry silkworm, *Bombyx mandarina* (Lepidoptera: Bombycidae). *Science in China Series C: Life Science* 51, 693–701.
- Papadopoulou A., Anastasiou I. and Vogler A. P. (2010) Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution* 27, 1659–1672. doi:10.1093/molbev/msq051.
- Patel R. K. and Jain M. (2012) NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS ONE* 7(2), e30619. <https://doi.org/10.1371/journal.pone.0030619>.
- Pearson W. R. and Lipman D. J. (1988) Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the USA* 85, 2444–2448.
- Perna N. T. and Kocher T. D. (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of Molecular Evolution* 41, 353–358.
- Powell J. R., Caccone A., Amato G. D. and Yoon C. (1986) Rates of nucleotide substitution in *Drosophila* mitochondrial DNA and nuclear DNA are similar. *Proceedings of the National Academy of Sciences of the USA* 83, 9090–9093.
- Rice P., Longden I. and Bleasby A. (2000) EMBOSS (The European molecular biology open software suite). *Trends in Genetics* 16, 276–277.
- Salvato P., Simonato M., Battisti A. and Negrisol E. (2008) The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). *BMC Genomics* 9, 331. doi: 10.1186/1471-2164-9-331.
- Sharma H. C. (1998) Bionomics, host plant resistance, and management of the legume pod borer, *Maruca vitrata*—A review. *Crop Protection* 17, 373–386.
- Stewart J. B., Freyer C., Elson J. L., Wredenberg A., Cansu Z., Trifunovic A. and Larsson N.-G. (2008) Strong purifying selection in transmission of mammalian mitochondrial DNA. *PLoS Biology* 6(1), e10. <https://doi.org/10.1371/journal.pbio.0060010>.
- Taanman J.-W. (1999) The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta (BBA)—Bioenergetics* 1410, 103–123.
- Tamura K., Nei M. and Kumar S. (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences of the USA* 101, 11030–11035.
- Taylor M. F., McKechnie S. W., Pierce N. and Kreitman M. (1993) The lepidopteran mitochondrial control region: structure and evolution. *Molecular Biology and Evolution* 10, 1259–1272.
- Vasta V., Ng S. B., Turner E. H., Shendure J. and Hahn S. H. (2009) Next generation sequence analysis for mitochondrial disorders. *Genome Medicine* 1, 100.
- Wolstenholme D. R. (1992) Animal mitochondrial DNA: structure and evolution. *International Review of Cytology* 141, 173–216.
- Ye F., Shi Y., Xing L., Yu H. and You P. (2013) The complete mitochondrial genome of *Paracymoriza prodigalis* (Leech, 1889) (Lepidoptera), with a preliminary phylogenetic analysis of Pyraloidea. *Aquatic Insects* 35, 71–88.
- Yin J., Hong G. Y., Wang A. M., Cao Y. Z. and Wei Z. J. (2010) Mitochondrial genome of the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) and comparison with other lepidopterans. *Mitochondrial DNA* 21, 160–169. doi: 10.3109/19401736.2010.503242.
- Yukuhiro K. S., Sezutsu H., Itoh H., Shimizu K. and Banno Y. (2002) Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silkworm, *Bombyx mandarina*, and its close relative, the domesticated silkworm, *Bombyx mori*. *Molecular Biology and Evolution* 19, 1385–1389.
- Zerbino D. R. and Birney E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18, 821–829.
- Zhang D.-X. and Hewitt G. M. (1997) Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology* 25, 99–120.
- Zhang R., Lu Z., Zhao H., Zhang X., Diasio R. B., Habus I., Jiang Z., Iyer R. P., Yu D. and Agrawal S. (1995) *In vivo* stability, disposition and metabolism of a “hybrid” oligonucleotide phosphorothioate in rats. *Biochemical Pharmacology* 50, 545–556.