# Shotgun assembly of the assassin bug Brontostoma colossus mitochondrial genome (Heteroptera, Reduviidae)

Arthur Kocher, Maria Kamilari, Emeline Lhuillier, Eric Coissac, Julie Péneau, Jérôme Chave, Jerôme Murienne

# Shotgun assembly of the assassin bug *Brontostoma colossus* mitochondrial genome (Heteroptera, Reduviidae)

Arthur Kocher[a], Maria Kamilari[ab], Emeline Lhuillier[c], Eric Coissac[d], Julie Péneau[e], Jérôme Chave[a], Jerôme Murienne[a].

a. Laboratoire EDB (UMR 5174, CNRS/UPS/ENFA) ; Université Paul Sabatier, Bât. 4R1, 118 route de Narbonne, 31062 Toulouse cedex 9, France ; arthur.kocher@gmail.com; jerome.murienne@univ-tlse3.fr; jerome.chave@univ-tlse3.fr; mariakamilari@gmail.com;

b. Division of Animal Biology, Department of Biology, School of Natural Sciences, University of Patras, GR-26500 Patras, Greece.

c. INRA, GeT–PlaGe, UAR 1209 Département de génétique animale, INRA Auzeville, 31326 Castanet-Tolosan, France ; emeline.lhuillier@udear.cnrs.fr

d. Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, 38041 Grenoble, France ; eric.coissac@inrialpes.fr

e. Équipe « Écosystèmes Amazoniens et Pathologie Tropicale » (EA 3593), Faculté de Médecine Hyacinthe Bastaraud de l'Université des Antilles et de la Guyane, Campus Saint Denis, av. d'Estrées, F-97306 Cayenne, French Guiana ; peneau.julie@wanadoo.fr

Corresponding author : Arthur Kocher, Laboratoire EDB (UMR 5174, CNRS/UPS/ENFA) ; address: Université Paul Sabatier, Bât. 4R1, 118 route de Narbonne, 31062 Toulouse cedex 9, France ; email : arthur.kocher@gmail.com; phone : +33 6.14.30.79.46

**Abstract**

The complete mitochondrial genome of the assassin bug *Brontostoma colossus* (Distant, 1902) (Heteroptera: Reduviidae) has been sequenced using a genome-skimming approach on an Illumina Hiseq 2000 platform. Fifty-four additional heteropteran mitogenomes, including five assassin bug species, were retrieved to allow for comparisons and phylogenetic analyses. The mitochondrial genome of *B. colossus* was determined to be 16,625 bp long, and consists of 13 protein-coding genes (PCGs), 23 transfer-RNA genes (tRNAs), two ribosomal-RNA genes (rRNAs), and one control region. The nucleotide composition is biased toward adenine and thymine (A+T=73.4%). Overall, architecture, nucleotide composition and genome asymmetry are similar among all available assassin bugs mitogenomes. All PCGs have usual start-codons (Met and Ile). Three T and two TA incomplete termination codons were identified adjacent to tRNAs, which was consistent with the punctuation model for primary transcripts processing followed by 3' polyadenylation of mature mRNA. All tRNAs exhibit the classic clover-leaf secondary structure except for $tRNA_{Ser(AGN)}$ in which the DHU arm forms a simple loop. Two notable features are present in the *B. colossus* mitogenome: *(i)* a 131 bp duplicated unit including the complete $tRNA_{Arg}$ gene, resulting in 23 potentially functional tRNAs in total, and (ii) a 857 bp duplicated region comprising 277 bp of the srRNA gene and 580 bp of the control region. A phylogenetic analysis based on 55 true bugs mitogenomes confirmed that *B. colossus* belongs to Reduviidae, but contradicted widely accepted hypothesis. This highlights the limits of phylogenetic analyses based on mitochondrial data only.

**Keywords:** genome skimming; next-generation sequencing; phylogenomic.

## *1.* Introduction

Mitochondrial DNA has various interesting properties such as abundance in animal tissues, small size, relatively simple genomic structure, fast rate of evolution, and a straightforward mode of transmission with a low level of recombination (due to its maternal inheritance). This makes it a valuable tool for comparative genomics, population genetics and phylogenetics at various taxonomic resolutions (Avise et al., 1987; Moritz et al., 1987). An increasing number of complete mitochondrial genomes has been made available in the past decade, relying on long range Polymerase Chain Reaction (PCR), but this approach is difficult to perform and time-consuming. The immense yield now provided by Next Generation Sequencing (NGS) helps resolve these issues. The sequencing of the complete nuclear genome remains expensive because it requires a deep sequencing, but a relatively shallow sequencing can be used to recover the high copy fraction of mitochondrial DNA. This «genome skimming» approach, originally developed for plant organelles assemblage (Besnard et al., 2013; Malé et al., 2014; Straub et al., 2012), has been successfully used to assemble a wide variety of animal mitochondrial genomes (Besnard et al., 2014; Doyle et al., 2014; Thompson et al., 2014; Veale et al., 2014). However, mitogenome-based studies have been unbalanced among taxa, and the amount of available data for insects remains limited in comparison with that of vertebrates (Gissi et al., 2008; Salvato et al., 2008).

As for most metazoans, the mitogenome of insects is a circular double-stranded molecule of 14-20kb in size and exhibits a typical set of 13 protein-coding genes (PCGs), 22 transfer RNA genes (tRNAs) and 2 ribosomal RNA genes (rRNAs) (Boore, 1999; Wolstenholme, 1992), even though variations in gene content exist (Gissi et al., 2008; Junqueira et al., 2004; Shao and Barker, 2003; Thao et al., 2004). In addition, it contains a large non-coding region, the control region (CR), which is implicated in the initiation of transcription and replication processes (Bernt et al., 2013a; Clayton, 1992; Saito, 2005; Zhang and Hewitt, 1997).

Heteroptera (true bugs) contains over 40,000 described species to date, constituting one of the most diverse group of non-holometabolous insects (Weirauch and Schuh, 2011). Assassin bugs (Reduviidae) are a large family of mostly predatory land bugs belonging to the infraorder Cimicomorpha. It currently comprises close to 7,000 species worldwide, that exhibit a remarkable diversity in morphological traits and life habits (Weirauch and Munro, 2009; Wheeler, 1997). Some of them are of agricultural or medical importance, the most notorious being part of the hematophagous Triatominae subfamily, known as vectors of Chagas disease in Central and South America (Lent and Wygodzinsky, 1979). Five assassin bugs mitogenomes have been sequenced so far: *Agriosphodrus dohrni* (Li et al., 2011), *Oncocephalus breviscutum* (Li et al., 2013), *Sirthenea flavipes* (Gao et. al, 2013), *Valentia hoffmanni* (Hua et. al, 2009) and *Triatoma dimidiata* (Dotson and Beard, 2001). Among Reduviidae, the neotropical genus *Brontostoma* currently includes around 20 species (Dougherty, 1995; Gil-Santana and Baena, 2009; Maldonado Capriles, 1990). It is characterized by a bright coloration with red, yellow, black and brown. Like all members of the Ectrichodiinae subfamily, they are predators specialized on millipedes.

In this paper, we describe a genome-skimming approach using Illumina technology to assemble the complete mitochondrial genome of *Brontostoma colossus* (Distant, 1902). Its organization and features are compared to five other mitogenomes of Reduviidae. Fifty-four additional heteropteran mitochondrial genomes are used to perform a phylogenetic analysis.

## 2. Material and method

### 2.1 Specimen, DNA extraction and sequencing

One specimen of *B. colossus* was collected in French Guiana (RN2 Roura-Saint Georges) on April 20th 2010. Total genomic DNA was extracted from leg muscle tissue using the DNeasy Blood and Tissue kit (Qiagen, Valencia, CA, USA), following a protocol adapted from the manufacturer's instructions (Supplementary Data 1). The quality and quantity of extracted genomic DNA was evaluated using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and a PicoGreen double-stranded DNA quantitation assay kit (Life Technologies, Carlsbad, CA, USA).

The genomic DNA was sent for library construction and sequencing to the GeT-PlaGe core facilities of Genotoul (Toulouse, France). 288 ng of DNA were used for library construction using the Illumina TruSeq Nano DNA Sample Prep Kit following the instructions of the supplier (Illumina Inc., San Diego, CA, USA). After shearing by ultrasonication with a Covaris M220 (Covaris Inc., Woburn, MA, USA), purified fragments were A-tailed and ligated to sequencing indexed adapters. Fragments with an insert size around 450 bp were selected with Agencourt Ampure XP beads (Beckman Coulter, Inc.), and enriched with 8 cycles of PCR before library quantification and validation. The library was multiplexed with 23 other libraries (generated in other projects). The pool of libraries was then hybridized on one lane of Hiseq 2000 flow cell using the Illumina TruSeq PE Cluster Kit v.3, and paired-end reads of 100 nucleotides were collected on the HiSeq 2000 sequencer using the Illumina TruSeq SBS Kit v.3 (200 cycles). Quality filtering was performed by the Consensus Assessment of Sequence and Variation (CASAVA) pipeline. Sequence data were stored on the NG6 platform (Mariette et al., 2012) and all the computations were performed on the computer cluster of the Genotoul bioinformatic platform (Toulouse, France).

### 2.2 Sequence assembly

Mitochondrial genome and nuclear ribosomal clusters were assembled using a previously described strategy (Besnard et al 2013; Besnard et al 2014; Malé et al. 2014). It is in essence similar to that proposed by Hahn et al. (2013). Reads aligning with the mitochondrial protein sequences of the closely related species *Triatoma dimidiata* (NC 002609, Dotson and Beard, 2001) were identified using the PLAST program (Nguyen and Lavenier, 2009). For the nuclear ribosomal cluster, we started from reads aligning with the 28S and 18S rRNA genes of *Eurydema maracandica* (Yu et al., 2013).

Reads with a match of at least 90% were assembled into contigs using the Velvet assembler (Zerbino and Birney, 2008) with a k-mer length of 81 and all the remaining parameters left at their default values. The resulting contigs were used as seeds to initiate the genome walking strategy (iterative mapping) using the extractreads2 program (included in the Obitools package; http://metabarcoding.org/obitools). Reads sharing at least 80 consecutive bp with the seeds were selected and subsequently used as seed to repeat the operation until no new read was identified. The newly selected reads were assembled with the Velvet assembler. The few resulting contigs were assembled using Geneious 6.0.6 Pro (Biomatters, Auckland, New Zealand). Two regions of the genome were not assembled using this procedure, due to ambiguities in the assembling process. At these locations, repeated regions were revealed by the coexistence of two assembling paths with significant coverage support (i.e. of the same magnitude than the average coverage), one path being

at the junction between two copies of the repeated element, the other being at the end of the last copy (see Figure 1). We inferred the repeated copy number by comparing the number of reads mapping on a DNA fragment present only once in the mitogenome to that of DNA fragment belonging to the repeated element, assuming that the read coverage of a particular genomic region is proportional to its copy fraction in the sample. This coverage analysis is detailed in the Supplementary Material.

Coverage statistics were computed on the assembled genome with Geneious 6.0.6, by mapping the reads using the following mapping parameter: a minimum overlap of 100 bp, a minimum overlap identity of 95%, a word length of 50 and a maximum mismatch per read of 5%.

### *2.3 Genome annotation*

The mitochondrial genome was first annotated using the MITOS web server (Bernt et al., 2013b) applying the invertebrate mitochondrial genetic code (NCBI code Table 5). The annotations of tRNA genes were kept unchanged. The annotations of PCGs were refined by checking manually for consistent start/stop codons and reading frames. The annotations of rRNA genes were extended until adjacent tRNAs ($tRNA_{Ser(TCN)}$, $tRNA_{Leu(CUN)}$ and $tRNA_{Val}$), following the punctuation model of mtDNA transcription (Ojala et al., 1981; Stewart and Beckenbach, 2009). The 5' end of srRNA is not flanked by a tRNA. We adjusted it by mapping the srRNA of the *Agriosphodrus dohrni* mitogenome for which the secondary structured was predicted, and presented all expected domains (Li et al. 2011). We used Geneious 6.0.6 Pro with the following parameters: a word length of seven; maximum gap size of 15 and maximum mismatches of 40%. This approach conducted to extend lrRNA's annotation by 33 bp until $tRNA_{Leu}$ at its 3' end and by 578 bp until tRNA-Val at its 5' end. srRNA's annotation was extended by 2 bp until $tRNA_{Val}$ and by 47 bp at its 5' end. We further verify the consistency of these new annotations by mapping the rRNAs of *Agriosphodrus dohrni* as described above. The remaining large non-annotated sequence was annotated as the control region in homology with other insect mitogenomes.

The 18S and 28S rRNAs were annotated in comparison with that of *Eurydema maracandica* (Yu et al., 2013). The 5.8S rRNA was annotated in comparison with that of *Triatoma dimidiata* (accession number: KF142517).

### *2.4 Sequence analyses and phylogenetics*

Base composition and codon usage where computed with MEGA6 (Tamura et al., 2013). AT-skew [(A-T)/(A+T)] and GC-skew [(G-C)/(G+C)] where used to measure nucleotide compositional differences between genes (Perna and Kocher, 1995) Relative synonymous codon usage (RSCU) where used to described bias in synonymous codon composition.

Tandem repeats were identified using Tandem Repeat Finder webserver (Benson, 1999). The secondary structure of tRNA's was inferred via the MITOS web server pipeline. Putative stem-loop structures were inferred using the RNAstructure web server (Bellaousov et al., 2013). We looked for structures conserved among the six assassin bugs within the 100 bp of the control region flanking $tRNA_{Ile}$ where stem loops structures have already been reported in Reduviidae (Dostson and Beard, 2001; Gao et al. 2013). We used the Turbofold algorithm (Harmanci et al.

2011), which infers secondary structure from high base pairing probabilities using the information derived from the sequence itself via the nearest neighbour thermodynamic model and also the information computed by using pairwise-sequence-alignment-based probabilities.

Fifty-four addditional heteropteran mitogenomes, including five assassin bugs, were downloaded from Genbank (Table 1). Two mitogenomes of Auchenorrhyncha were used as outgroups. The 13 PCGs were used for the analysis to allow for comparison with previous studies (Li et al., 2011; Yang et al., 2013). They were first aligned separately based on amino-acids translation with translatorX (Abascal et al., 2010). Divergent regions were removed with Gblocks.0.91b before back-translation in order to conserve reading frames. All resulting alignments were then concatenated using FASconCAT (Kück and Meusemann, 2010).The best partitioning scheme and substitution model were inferred with PartitionFinder.1.1.1 (Lanfear et al., 2012), using the greedy algorithm for scheme search and the Bayesian information criterion for scheme selection. A maximum-likelihood (ML) analysis was performed with RAxML 8.0 (Stamatakis, 2014), using the rapid bootstrap analysis option with the majority-rule tree based bootstopping criteria. Bootstrap support values were printed on the best ML tree. A Bayesian analysis was conducted using Mr.Bayes 3.2 (Ronquist and Huelsenbeck, 2003), starting from four random trees with 10 Markov chains (nine heated chain and 1 cold chain), 2,000,000 generations and all other parameters set to default. Each set was sampled every 200 generations with a burn-in of 25% of the sampled trees. At the end of the analysis, the average standard deviation of split frequencies was below the recommended 0.01.

## *3* Results and discussion

### *3.1 Genome sequencing, assembly and annotation*

After filtering 4.74% of the initial reads, raw sequence data represented a total of 7,831,929 paired-end reads (15,663,858 reads in total). Among the remaining reads, 34,224 were assembled into a 16,625 bp circular sequence, representing the complete mitochondrial genome with an average sequencing depth of 209.6. A circular map of the mitogenome and the assembly coverage are presented on Figure 2. The sequence was deposited in Genbank under the accession number KM044501. A total of 38 genes (13 PCGs, 23 tRNAs, two rRNAs) and one control region were identified. Twenty-four genes are encoded on the majority strand and the others mapped to the minority strand. Seven gene overlaps where observed, the longest being a 8-bp region between $tRNA_{Cys}$ and $tRNA_{Trp}$ (Table 2), which is a peculiar feature in Arthropoda (Bernt et al., 2013a). Apart from the control region, 13 non-coding regions ranging from 1 bp to 46 bp where identified.

Three repeated regions were identified: a 131-bp element containing $tRNA_{Arg}$; an 857-bp element consisting of 277 bp of srRNA and 580 bp of the control region and a 74-bp repeated element followed by a 39 partial copy located in the control region were tandem repeats are also found in other assassin bugs (Li et al., 2011). The existence of these duplications was supported by assembling paths with high coverage (over 100 reads). Details on the results of the coverage analysis are given in the Supplementary material. However, the sequencing technology used in this study does not allow inferring repeats copy numbers with high certainty, especially if a polymorphism exists in the sample. Indeed, heteroplasmy is often associated with tandem repetition (Zhang, 1997).

The complete nuclear ribosomal gene cluster was recovered. A total of 21,514 reads were assembled into a 8,287 bp sequence comprising of 18S rRNA (1,893 bp), ITS1 (1,141 bp), 5.8S rRNA (155 bp), ITS2 (941 bp) and 28S rRNA (4,077 bp). The sequence was deposited on Genbank under the accession number KM278219.

The mitogenome of *B. colossus* shares the same architecture and orientation as the other six mitogenomes of assassin bugs, except for the presence of an additional $tRNA_{Arg}$ gene, as will be described below. This gene arrangement (without the additional $tRNA_{Arg}$) is also found in *Drosophila melanogaster* and was the first to be determined, differing by a single tRNA translocation, from that of the chelicerate *Limulus polyphemus,* which is considered ancestral for Arthropoda (Boore et al., 1995; Lavrov et al., 2000). This mitogenome organisation is also found in crustaceans, and is thought to be ancestral for the insect-crustacean clade (Boore, 1999; Boore et al., 1998). Among the available data for heteropterans, seven out of 54 species have been found to present a different gene arrangement: *Nabicalis apicalis* mitogenome miss the cluster containing $tRNA_{Ile}$, $tRNA_{Gln}$ and $tRNA_{Met}$ (Li et al., 2012a). $tRNA_{Ile}$ and $tRNA_{Gln}$ are also missing in *Urochela quadrinotata* (Yuting et al., 2012). The positions of $tRNA_{Thr}$ and $tRNA_{Pro}$ are inverted in *Physopelta gutta* (Hua et al., 2008). The gene order of the *Stenopirates* sp. mitogenome differs largely with the inversion of two tRNA genes ($tRNA_{Thr}$ and $tRNA_{Pro}$) and translocations of five gene clusters ($tRNA_{Thr}$-$tRNA_{Pro}$ -ND6,CYTB-$tRNA_{Ser(TCN)}$, ND1- $tRNA_{Leu(CUN)}$, l-rRNA- $tRNA_{Val}$ -s-rRNA and the control region) between ND4L and $tRNA_{Ile}$ (Li et al., 2012b). The positions of $tRNA_{Cys}$ and $tRNA_{Trp}$ are exchanged in *Aradacanthia heissi* (Shi et al., 2012). In the latter, as well as in the two other Aradoidea represented (*Brachyrhynchus hsiaoi* and *Neuroctenus parus*); the positions of $tRNA_{Ile}$ and $tRNA_{Gln}$ are exchanged (Li et al., 2014; Hua et al., 2008).

The size of the six assassin bug mitogenomes ranges from 15,625 bp in *V. hoffmannii* to 17,019 bp in *T. dimidiata*. These differences are mostly due to variations in the size of the control region, which is generally observed for all insects. Previous studies have reported control region size ranging from 70 bp in *Ruspolia dubia* (Orthoptera) to 4,599bp in *Drosophila melanogaster* (Diptera) (Garesse, 1988; Zhou et al., 2007).

### *3.2 Protein-coding genes*

The total length of the 13 PCGs was 11,041 bp. Their nucleotide composition is strongly biased toward AT with an overall AT content of 72.8% (Supplementary Table 1). All PCGs have an ATN start codon (Table 2). Six PCGs initiated with ATT (ND2, COX2, ATP8, ND5, ND4L and ND1), five initiated with ATG (COX1, ATP6, COX3, ND4 and CYTB) and two initiated with ATA (ND3 and ND6). Four genes share the same ATG start codon in the six assassin bugs mitogenomes (COXI, ATP6, COXIII and ND4). No GTG start codon was found in *B. colossus* in contrast with other assassin bugs for ND5, ND4L and ND1 genes (Supplementary Table 2). Other unconventional start codons were described in insects such as TTG in heteropterans (Yang et al., 2013), CGA and TTAG in lepidopterans (Lee et al., 2006; Yukuhiro et al., 2002), or ATAA, GTAA and TTAA in dipterans (Ballard, 2000; Clary and Wolstenholme, 1985) but none of them were found in assassin bugs.

The majority of PCGs have a usual TAA stop codon, but three T and two TA stop codons were identified (ND2, COX1, COX2 and ND5, CYTB respectively). These incomplete stop codons are immediately adjacent to tRNA genes encoded on the same strand, consistent with the punctuation model for primary transcripts processing followed by 3' polyadenylation of mature mRNA that will allow the completion of termination codons (Nagaike, 2005; Ojala et al., 1981; Stewart and Beckenbach, 2009). Incomplete stop codons can be found in all six assassin bugs mitogenomes and are shared with many arthropods (Boore, 2000).

### *3.3 Ribosomal and transfer RNA genes*

rRNA genes locations and lengths are similar to those of other insects mitogenomes. lrRNA is located between $tRNA_{Leu(CUN)}$ and $tRNA_{Val}$ and is 1,251 bp-long. srRNA is located between $tRNA_{Val}$ and the control region and is 793 bp-long. Their AT content is respectively 79.0% and 74.0%.

The classical set of 22 tRNAs found in arthropods is present in *B. colossus,* but an additional copy of the $tRNA_{Arg}$ gene was identified (see below). Their lengths vary between 61 bp ($tRNA_{Ala}$) and 72 bp ($tRNA_{Lys}$). Secondary structures of tRNAs are schematized in Supplementary Figure 1. The classical clover leaf structure was observed for each of them, except for $tRNA_{Ser(AGN)}$, in which the D arm is reduced to a simple loop, as in many insects, and more generally, in most bilaterians (Bernt et al., 2013a; Wolstenholme, 1992).

Flanking $tRNA_{Ala}$, we identified a duplicated element consisting of two identical 131-bp copies separated by eight non-coding base pairs. It includes the entire $tRNA_{Arg}$ and short-flanking regions corresponding to 28 bp of $tRNA_{Ala}$ and 38 bp of $tRNA_{Asn}$ (Figure 3). This unusual feature results in two copies of $tRNA_{Arg}$ for a total of 23 tRNA genes which is relatively rare in insect mitogenomes, even though more than 22 tRNAs were already observed in other species such as *Coreana raphaelis* (Lepidoptera; Kim et al., 2006)), *Thrips imaginis* (Thysanoptera; Shao and Barker, 2003), *Chrysomya chloropyga* (Diptera; Junqueira et al., 2004) and *Trialeurodes vaporariorum* (Hemiptera; Thao et al., 2004). To our knowledge, the presence of an additional $tRNA_{Arg}$ has only been described in Porifera and Placozoa (Lavrov and Lang, 2005; Signorovitch et al., 2007). The two copies of this duplicated element are strictly identical, which suggests a recent origin of the duplication event. Interestingly, it could lead to gene rearrangement through a duplication/deletion mechanism involving the random deletion of the original copy of the gene (Boore, 2000; Moritz and Brown, 1986; Shao et al., 2006). Most repeated elements are located close to the replication origin, supporting the idea that mitogenomic duplication events are mainly due to replication slippage mechanisms (Macey et al., 1998; Zhang and Hewitt, 1997). However, to our knowledge, no putative replication origin located close to this duplicated region has been mentioned so far.

### *3.4 Non-coding regions*

Thirteen short intergenic spacers (IGS) and a long control region were identified, matching the usual organisation of insect mitogenomes. Most IGS are very short, with less than four base pairs, and seven overlapping sequences are found (Table 2). The longest IGS are the 41 and 46 bp-flanking regions of the first copy of $tRNA_{Arg}$ and the 38-bp IGS found between ND1 and $tRNA_{Ser2}$. The latter is

remarkably long in every assassin bug mitogenome, ranging from 23 bp in *V. hoffmanni* to 309 bp in *T. dimidiata*. It exhibits tandem repeats in *A. dohrni* and *T. dimidiata* and has been suspected to be one of the replication origins (Dotson and Beard, 2001; Li et al., 2011) comparably to the 193 bp region located between the $tRNA_{leu(UUR)}$ and COXII genes in the honeybee *A. mellifera* (Crozier and Crozier, 1993).

The control region is 1,896-bp long and is located between the srRNA and $tRNA_{Ile}$ genes. In *B. colossus*, as well as in the five other Reduviidae, it exhibits a higher G+C content than that of the whole mitogenome, in contrast with other insect species in which the control region was found to be remarkably A+T rich (Zhang and Hewitt, 1997).

The alignment of the six Reduviidae control regions reveals a conserved sequence block (CSB) of 40 bp, including a string of 13 Gs (Figure 4a; Li et al., 2011). CSBs have been identified in the control region of various metazoans and are generally thought to play a role in the replication mechanism (Lee et al., 1995; Walberg and Clayton, 1981; Zhang and Hewitt, 1997). However, we did not find similarity between the CSB described here and those reported for other taxa, even though G islands have already been described in other insects (Oliveira et al., 2008). More studies are needed to identify precisely replication origins and to speculate on the role of adjacent sequences, as was proposed by (Saito, 2005) who strongly suspected a "T-strech" sequence conserved in *Drosophila* and other insect species to be involved in the replication process.

The control regions of the six assassin bugs present a similar organization (Figure 4b). In all of them, except for *S. flavipes,* tandem-repeats were identified between the CSB and $tRNA_{Ile}$. In *B. colossus,* they consist of four 74 bp units and one 39 bp unit, the latter corresponding to a partial copy of the 74 bp unit. The 100 bp preceding the CSB are remarkably G+C rich (42%) in *B. colossus*, but the nucleotide composition of the whole control region does not show any clear pattern, in contrast with other assassin bugs in which long G+C and A+T rich regions were found preceding and following the CSB respectively (Dotson and Beard, 2001; Gao et al., 2013; Li et al., 2011). In *B. colossus* control region, a notable feature is a large duplicated region consisting of two identical copies of 857 bp separated by seven non-coding nucleotides. It includes 277 bp of srRNA and 580 bp of the control region, comprising the CSB. It is remarkably long and results in two copies of the CSB, which could therefore have implications on the mitogenome replication process.

Within the 60 bp of the control region flanking the $tRNA_{Ile}$, DNA segments have the potential to form stem-loop structures involving at least 11 base-pairings in the six assassin bugs mitogenomes except for *V. hoffmanni* (Figure 5). Such features may be involved in the replication mechanism (Song and Liang, 2009; Zhang et al., 1995). However, this is only speculative. More comprehensive studies would be required to assess the significance of these inferences and the potential role of these structures.

### *3.5 Nucleotide content and codon usage*

The nucleotide composition is strongly biased toward adenine and thymine in the mitogenome of *B. colossus*, with A+T representing 73.5% of the whole sequence and ranging from 70.2% in the control region, 70.8% in protein-coding genes, 76.8% in tRNA genes to 77.1% in rRNA genes. AT-rich codons are predominant, with the most prevalent being in order ATT (Ile), TTA (Leu), TTT (Phe) and ATA

(Met). The relative synonymous codon usage (RSCU) clearly indicates that AT rich codons are favoured among synonymous codons (Supplementary Table 3). At the third codon position, AT content is particularly high (82.8%), and G nucleotides are under-represented (GC skew= -0.20). AT content, as well as A-T and G-C skews patterns, are similar among the six assassin bugs mitochondrial genomes (Supplementary Table 1).

### *3.6 Phylogenetic analysis*

Bayesian inference and Maximum Likelihood analysis (ML) generated phylogenetic trees with very similar topologies. The tree inferred by the Bayesian method is presented in Figure 6 with nodes posterior probabilities and ML bootstrap support values. The topology of the best ML tree for Reduviidae is also presented. The relationships among Reduviidae are conserved in both analyses except for the position of *B. colossus*, which is placed as a sister group of *V. hoffmanni* under Bayesian inference whereas it is the early lineage of Reduviidae in the ML tree. Our results are not well supported and are hardly comparable with those of recent studies that have addressed the relationships among assassin bugs based on nuclear and mitochondrial DNA as well as morphological data for a large number of taxa (Weirauch and Munro, 2009; Weirauch, 2008). However, the higher-level relationships of Reduviidea remain poorly resolved and the addition of mitogenomic data for more taxa will surely provide useful phylogenetic information in the future.

The 20 superfamilies represented in our dataset are monophyletic except for Miroidea. At the infra-order level, Pentatomorpha is monophyletic with the following relationships between the five superfamilies: Aradoidea + (Pentatomoidea + (Lygaeoidea + (Pyrrhocoroidea + Coreoidea))). These relationships are strongly supported in our analyses and do not confirm the results of previous studies based on a subset of the present mitogenomic data that placed Coreoidea and Lygaeoidea as sister groups (Hua et al., 2008; Yang et al., 2013). Based on the tree generated by Bayesian inference, Nepomorpha is monophyletic with the following relationships between the six superfamilies: (Pleoidea + Corixoidea) + ((Notonectoidea + Naucoroidea) + (Nepoidea + Ochteroidea)). ML analysis differs in positioning Pleoidea as a sister group of the remaining Nepomorpha. These results are inconsistent with those of a recent study based on molecular and morphological data (Hebsgaard et al., 2004). Interestingly, they also contradict a previous analysis of a subset of the present mitogenomic dataset by confirming the monophyly of Nepomorpha including Pleoidea, for which an infraordinal status was proposed (Hua et al., 2009). Gerromorpha and Leptopodomorpha are also monophyletic, but only two species of each were included in the study. Enicocephalomoprha was only represented by *Stenopirates* sp.

Our analysis supports the paraphyly of Cimicomorpha that consisted of four different clades: (Cimicoidea + Naboidea), Reduvioidea, Miridae and Tingidae. Tingidae is placed as a sister group to all remaining Heteroptera. However, the monophyly of Cimicomorpha has been widely accepted and is supported by various analyses that have been using mitochondrial and nuclear data for a larger number of taxa (Li et al., 2012; Schuh et al., (2009); Tian et al., 2008). Infraordinal relationships are conserved in both ML and Bayesian analyses: Tingidae + (Miridae + (Pentatomomorpha + ((Cimicoidea + Naboidea) + (Leptopodomorpha + ((Enicocephalimorpha + Gerromorpha) + Nepomorpha))))). These results are questionning the general consensus that considers Enicocephalomorpha as the early infraorder of Heteroptera (Weirauch et al., 2011). However, the relationships between the infraorders of Heteroptera remain controversial. Only few phylogenetic studies have addressed the question and most of these have only included a small number of taxa (Mahner, 1993; Wheeler et al., 1993; Xie et al., 2008). Mitogenomic data provide a new insight in this regard, but more taxa should be added to the current database, especially in the poorly represented infra-orders Enicocephalomorpha, Gerromorpha, Leptopodomorpha and Dipsocoromorpha.

Our results are incongruent with current phylogenetic hypothesis of Heteroptera (e.g. Schuh et al. 2009). On the other hand, they are in accordance with previous studies based on a subset of the present mitogenomic data (Li et al., 2011; Yang et al., 2013). Analyses performed on individual genes by Tian et al (2008) and Schuh et al. (2009) indicate that the monophyly of Cimicomorpha is only supported by nuclear DNA. The incongruence of phylogenetic analyses among different genomic regions is a well-known issue that can have various biological causes such as incomplete lineage sorting or rate variation among partitions (Som, 2014). Lin and Danforth (2004) studied the differences in the pattern of nucleotide substitution among nuclear and mitochondrial genes and concluded that insect phylogenetic studies should increasingly focus on nuclear data. This raises the limitations of phylogenetic inferences from complete mitochondrial genomes only. While reducing stochastic errors by providing large molecular datasets, this approach is susceptible to potential site-specific bias and would benefit from a conjoint analysis with other genes. One advantage of the genome skimming approach used in this paper is that it allows the recovery of nuclear genes of phylogenetic interest in addition to the full mitogenome sequence. However, 18S and 28S data available to date on Genbank are too scarce (14 species out of 55 represented in our mitochondrial dataset) to perform a combined analysis.

## ***4*** *Concluding remarks*

This study provides further evidence that NGS can be used efficiently to generate mitogenomic data with a low amount of DNA. We successfully recovered the full mitogenome sequence of *Brontostoma colossus,* which included two unusual duplicated regions. The Illumina technology used in this study identified repeated elements without ambiguities, but their copy number can only be estimated using the sequence coverage information. In a near future, the rapid evolution of sequencing technologies (especially read length) and bioinformatics tools will probably bring improvements in this regard. The increasing number of full mitochondrial genome sequences brings precious phylogenetic information. However, our study highlights the limits of analyses based on mitochondrial DNA only. Currently, there is a lack of correspondence between publicly available mitogenomic and nuclear data. The genome skimming-approach could provide an interesting improvement in this regard. In a single experimentation, it allows to recover the full mitogenome sequence as well as nuclear genes that are classically used for phylogenetic inference. We argue that future studies reporting full organelles sequencing with genome skimming approaches should systematically report the assembly of the nuclear genes of phylogenetic interest.

## Tables

**Table 1: Complete or near-complete mitochondrial genomes used in this study**

| Suborder | Infra-order/ Superfamily | Family | Species | Accession num. | Reference |
|---|---|---|---|---|---|
| **Auchenorryncha** | Fulgoroidea | Fulgoridae | *Lycorma delicatula* | EU909203 | (Song et al., 2012) |
| | | Flatidae | *Geisha distinctissima* | NC_012617 | (Song and Liang, 2009) |
| **Heteroptera** | **Cimicomorpha** | | | | |
| | Cimicoidea | Anthocoridae | *Orius niger* | EU427341 | (Hua et al., 2008) |
| | Miroidea | Miridae | *Adelphocoris fasciaticollis* | NC_023796 | (Wang et al., 2014) |
| | | | *Apolygus lucorum* | NC_023083 | (Wang et al., 2013) |
| | | | *Lygus lineolaris* | NC_021975 | Unpublished |
| | | | *Nesidiocoris tenuis* | NC_022677 | (Dai et al., 2012) |
| | | Tingidae | *Corythucha ciliata* | NC_022922 | (Yang et al., 2013) |
| | Naboidea | Nabidae | *Alloeorhynchus bakeri* | HM235722 | (Li et al., 2012a) |
| | | | *Gorpis annulatus* | JF907591 | (Li et al., 2012a) |
| | | | *Gorpis humeralis* | JF927830 | (Li et al., 2012a) |
| | | | *Himacerus apterus* | JF927831 | (Li et al., 2012a) |
| | | | *Himacerus nodipes* | JF927832 | (Li et al., 2012a) |
| | | | *Nabis apicalis* | JF907590 | (Li et al., 2012a) |
| | Reduvioidea | Reduviidae | *Agriosphodrus dohrni* | NC 015842 | (Li et al., 2011) |
| | | | *Brontostoma colossus* | KM044501 | This study |
| | | | *Oncocephalus breviscutum* | NC 022816 | (Li et al., 2013) |
| | | | *Sirthenea flavipes* | HQ645959 | (Gao et al., 2013) |
| | | | *Triatoma dimidiata* | NC 002609 | (Dotson and Beard, 2001) |
| | | | *Valentia hoffmanni* | NC 012823 | (Hua et al., 2009) |
| | **Enicocephalomorpha** | | | | |
| | Enicocephaloidea | Enicocephalidae | *Stenopirates* sp. | NC_016017 | (Li et al., 2012b) |
| | **Gerromorpha** | | | | |
| | Gerroidea | Gerridae | *Aquarius paludum* | NC_012841 | (Hua et al., 2009) |
| | Hydrometroidea | Hydrometridae | *Hydrometra* sp. | NC_012842 | (Hua et al., 2009) |
| | **Leptopodomorpha** | | | | |
| | Leptopodoidea | Leptopodidae | *Leptopus* sp. | FJ456946 | (Hua et al., 2009) |
| | Saldoidae | Saldidae | *Saldula arsenjevi* | EU427345 | (Hua et al., 2008) |
| | **Nepomorpha** | | | | |
| | Corixoidea | Corixidae | *Sigara septemlineata* | FJ456941 | (Hua et al., 2009) |
| | Naucoroidea | Aphelocheiridae | *Aphelocheirus ellipsoideus* | FJ456939 | (Hua et al., 2009) |
| | | Naucoridae | *Ilyocoris cimicoides* | NC_012845 | (Hua et al., 2009) |
| | Nepoidea | Belostomatidae | *Diplonychus rusticus* | FJ456940 | (Hua et al., 2009) |
| | | Nepidae | *Laccotrephes robustus* | NC_012817 | (Hua et al., 2009) |
| | Notonectoidea | Notonectidae | *Enithares tibialis* | NC_012819 | (Hua et al., 2009) |
| | Ochteroidea | Gelastocoridae | *Nerthra* sp. | NC_012838 | (Hua et al., 2009) |
| | | Ochteridae | *Ochterus marginatus* | NC_012820 | (Hua et al., 2009) |
| | Pleoidea | Helotrephidae | *Helotrephes* sp. | FJ456951 | (Hua et al., 2009) |
| | **Pentatomomorpha** | | | | |
| | Aradoidea | Aradidae | *Aradacanthia heissi* | HQ441233 | (Shi et al., 2012) |
| | | | *Brachyrhynchus hsiaoi* | NC_022670 | (Li et al., 2014) |
| | | | *Neuroctenus parus* | EU427340 | (Hua et al., 2008) |
| | Coreoidea | Alydidae | *Riptortus pedestris* | EU427344 | (Hua et al., 2008) |
| | | Coreidae | *Hydaropsis longirostris* | EU427337 | (Hua et al., 2008) |
| | | Rhopalidae | *Aeschyntelus notatus* | EU427333 | (Hua et al., 2008) |
| | | | *Stictopleurus subviridis* | NC_012888 | (Hua et al., 2009) |
| | Lygaeoidea | Berytidae | *Yemmalysus parallelus* | EU427346 | (Hua et al., 2008) |
| | | Colobathristidae | *Phaenacantha marcida* | EU427342 | (Hua et al., 2008) |
| | | Geocoridae | *Geocoris pallidipennis* | EU427336 | (Hua et al., 2008) |
| | | Malcidae | *Chauliops fallax* | NC_020772 | (Hua et al., 2008) |
| | | | *Malcus inconspicuus* | EU427339 | (Hua et al., 2008) |
| | Pentatomoidea | Cydnidae | *Macroscytus gibbulus* | EU427338 | (Hua et al., 2008) |
| | | Dinidoridae | *Coridius chinensis* | JQ739179 | (Liu et al., 2012) |
| | | Pentatomidae | *Dolycoris baccarum* | NC_020373 | (Zhang et al., 2013) |
| | | | *Halyomorpha halys* | NC_013272 | (Lee et al., 2009) |
| | | | *Nezara viridula* | NC_011755 | (Hua et al., 2008) |
| | | Plataspidae | *Coptosoma bifaria* | EU427334 | (Hua et al., 2008) |
| | | | *Megacopta cribraria* | NC_015842 | (Hua et al., 2008) |
| | | Tessaratomidae | *Eusthenes cupreus* | NC_022449 | (Song et al., 2013) |
| | | Urostylididae | *Urochela quadrinotata* | NC_020144 | (Yuting et al., 2012) |
| | Pyrrhocoroidea | Largidae | *Physopelta gutta* | EU427343 | (Hua et al., 2008) |
| | | Pyrrhocoridae | *Dysdercus cingulatus* | EU427335 | (Hua et al., 2008) |

413 **Table 2: Summary of the mitochondrial genome of *Brontostoma colossus***

| Locus | Direction | Location (bp) | Size (bp) | Anticodon | Start codon | Stop codon | Interlocus nucleotides |
|---|---|---|---|---|---|---|---|
| $tRNA_{Ile}$ | F | 1-63 | 63 | GAT | | | 0 |
| $tRNA_{Gln}$ | R | 66-134 | 69 | TTG | | | 2 |
| $tRNA_{Met}$ | F | 133-202 | 70 | CAT | | | -2 |
| ND2 | F | 203-1193 | 991 | | ATT | T-- | 0 |
| $tRNA_{Trp}$ | F | 1194-1260 | 67 | TCA | | | 0 |
| $tRNA_{Cys}$ | R | 1253-1317 | 65 | GCA | | | -8 |
| $tRNA_{Tyr}$ | R | 1319-1382 | 64 | GTA | | | 1 |
| COX1 | F | 1384-2917 | 1534 | | ATG | T-- | 1 |
| $tRNA_{Leu(UUR)}$ | F | 2918-2982 | 65 | TAA | | | 0 |
| COX2 | F | 2983-3661 | 679 | | ATT | T-- | 0 |
| $tRNA_{Lys}$ | F | 3662-3733 | 72 | CTT | | | 0 |
| $tRNA_{Asp}$ | F | 3734-3796 | 63 | GTC | | | 0 |
| ATP8 | F | 3797-3955 | 159 | | ATT | TAA | 0 |
| ATP6 | F | 3949-4620 | 672 | | ATG | TAA | -7 |
| COX3 | F | 4620-5408 | 789 | | ATG | TAA | -1 |
| $tRNA_{Gly}$ | F | 5413-5477 | 65 | TCC | | | 4 |
| ND3 | F | 5478-5831 | 354 | | ATA | TAA | 0 |
| $tRNA_{Arg}$ | F | 5873-5937 | 65 | TCG | | | 41 |
| $tRNA_{Ala}$ | F | 5984-6044 | 61 | TGC | | | 46 |
| $tRNA_{Arg}$ | F | 6049-6113 | 65 | TCG | | | 4 |
| $tRNA_{Asn}$ | F | 6118-6187 | 70 | GTT | | | 4 |
| $tRNA_{Ser(AGN)}$ | F | 6187-6255 | 69 | GCT | | | -1 |
| $tRNA_{Glu}$ | F | 6258-6319 | 62 | TTC | | | 2 |
| $tRNA_{Phe}$ | R | 6318-6388 | 71 | GAA | | | 2 |
| ND5 | R | 6389-8094 | 1706 | | ATT | TA- | 0 |
| $tRNA_{His}$ | R | 8095-8158 | 64 | GTG | | | 0 |
| ND4 | R | 8159-9493 | 1335 | | ATG | TAA | 0 |
| ND4L | R | 9487-9771 | 285 | | ATT | TAA | -7 |
| $tRNA_{Thr}$ | F | 9774-9837 | 64 | TGT | | | 2 |
| $tRNA_{Pro}$ | R | 9838-9903 | 66 | TGG | | | 0 |
| ND6 | F | 9906-10397 | 492 | | ATA | TAA | 2 |
| CYTB | F | 10397-11529 | 1133 | | ATG | TA- | -1 |
| $tRNA_{Ser(TCN)}$ | F | 11530-11598 | 69 | TGA | | | 0 |
| ND1 | R | 11637-12548 | 912 | | ATT | TAA | 38 |
| $tRNA_{Leu(CUN)}$ | R | 12549-12613 | 65 | TAG | | | 0 |
| lrRNA | R | 12614-13864 | 1251 | | | | 0 |
| $tRNA_{Val}$ | R | 13865-13935 | 71 | TAC | | | 0 |
| srRNA | R | 13936-14728 | 793 | | | | 0 |
| Control region | | 14729-16625 | 1896 | | | | 0 |

414

415

## Figures

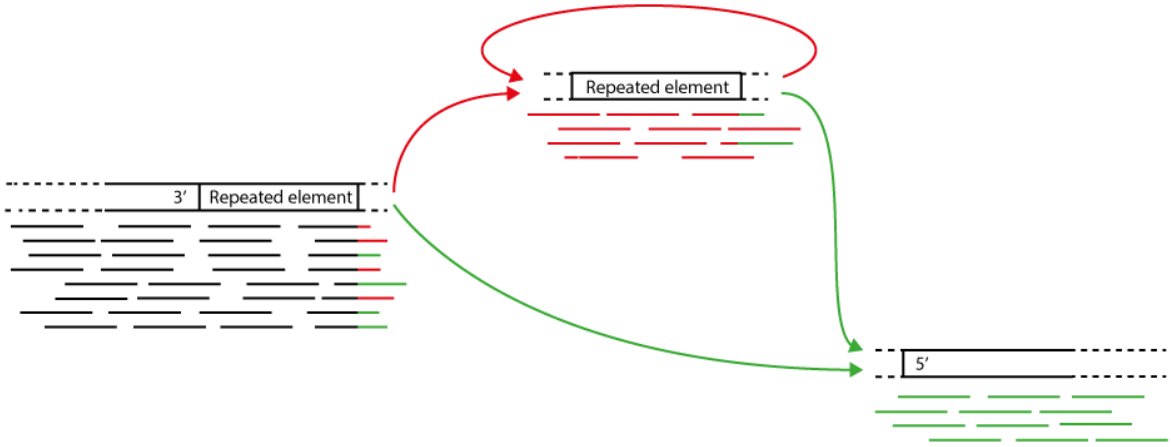



**Figure 1 : Schematization of an assembly ambiguity revealing a repeated element in the sequence. When assembling the 5' end of the repeated element, two different "assembly paths" are supported by a significant number of reads: (i) a path leading to the beginning of a novel repeated element (depicted in red) (ii) a path leading to the region flanking the 5' end of the last repetition (depicted in green).**

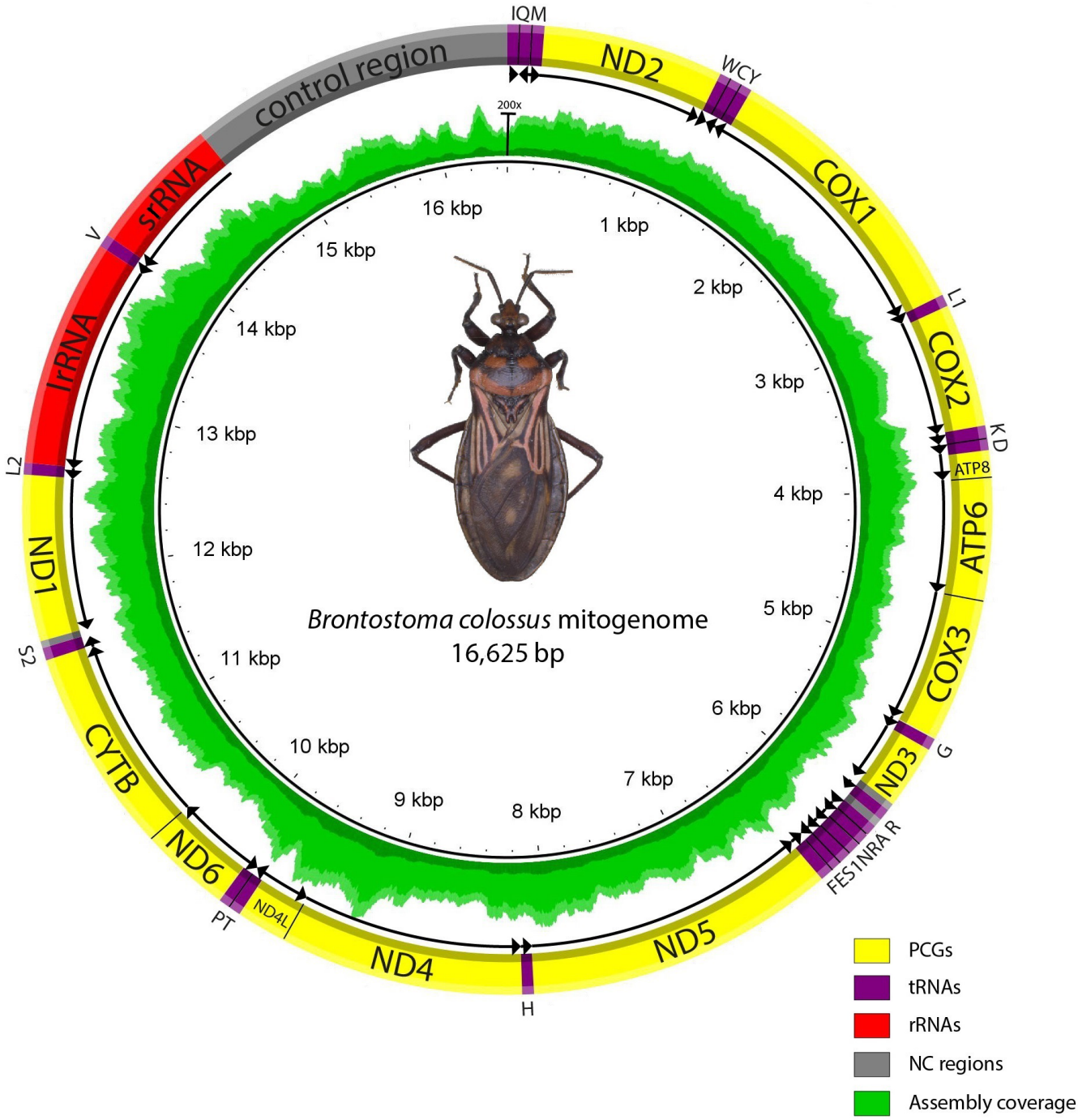



**Figure 2: Schematic representation of *Brontostoma colossus* mitogenome. tRNAs are labelled according to the IUPAC-IUB single-letter amino acid codes. Arrows indicate directions of genes. The scale on the assembly coverage ring indicates 200x coverage.**

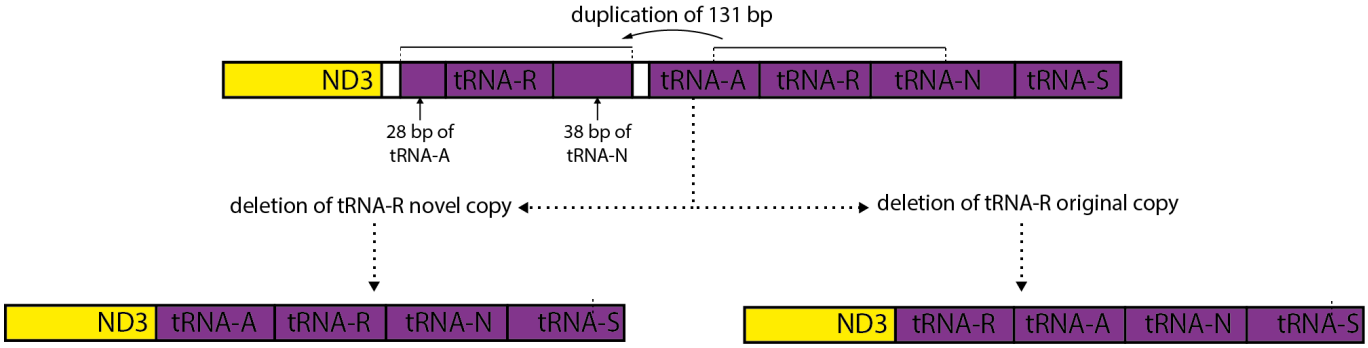



**Figure 3: Organization of the 131 bp duplicated sequence comprising tRNA-Arg and potential gene rearrangements after random deletion of one tRNA-arg copy.**

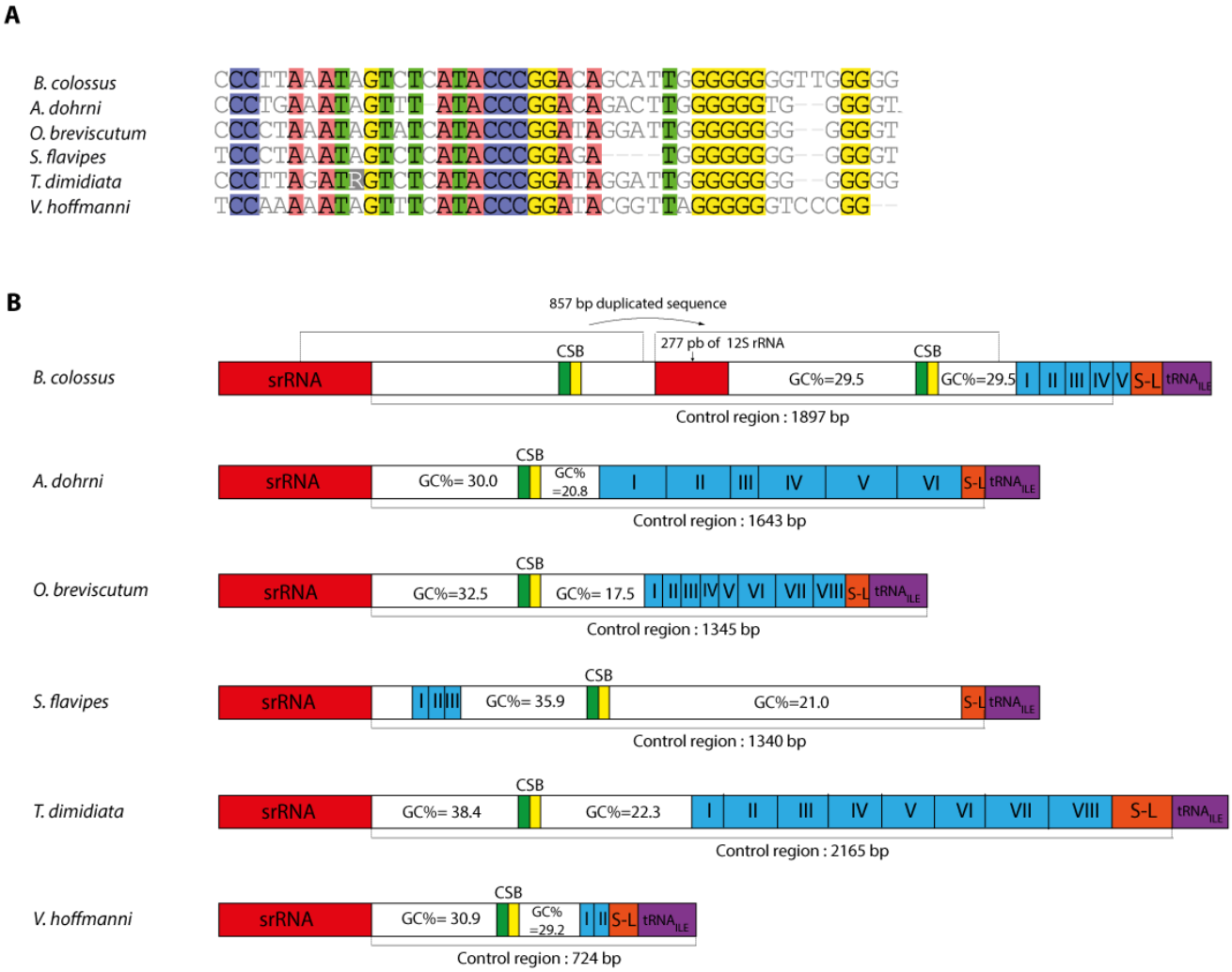



**Figure 4: (A) Alignment of the conserved sequence blocks identified in the mitochondrial control region of the six assassin bugs. Nucleotide positions that are conserved among the six assassin bugs are highlighted. (B) Structural organization of the mitochondrial control region of the six assassin bugs. The blue boxes with roman numerals indicate tandem repeat units. The CSB box indicates the conserved sequence block, and the yellow part indicates the "G element". The orange box "S-L" indicates the region where potential stem loops are found. GC content is shown in the white boxes previously described as GC-rich and AT-rich regions.**

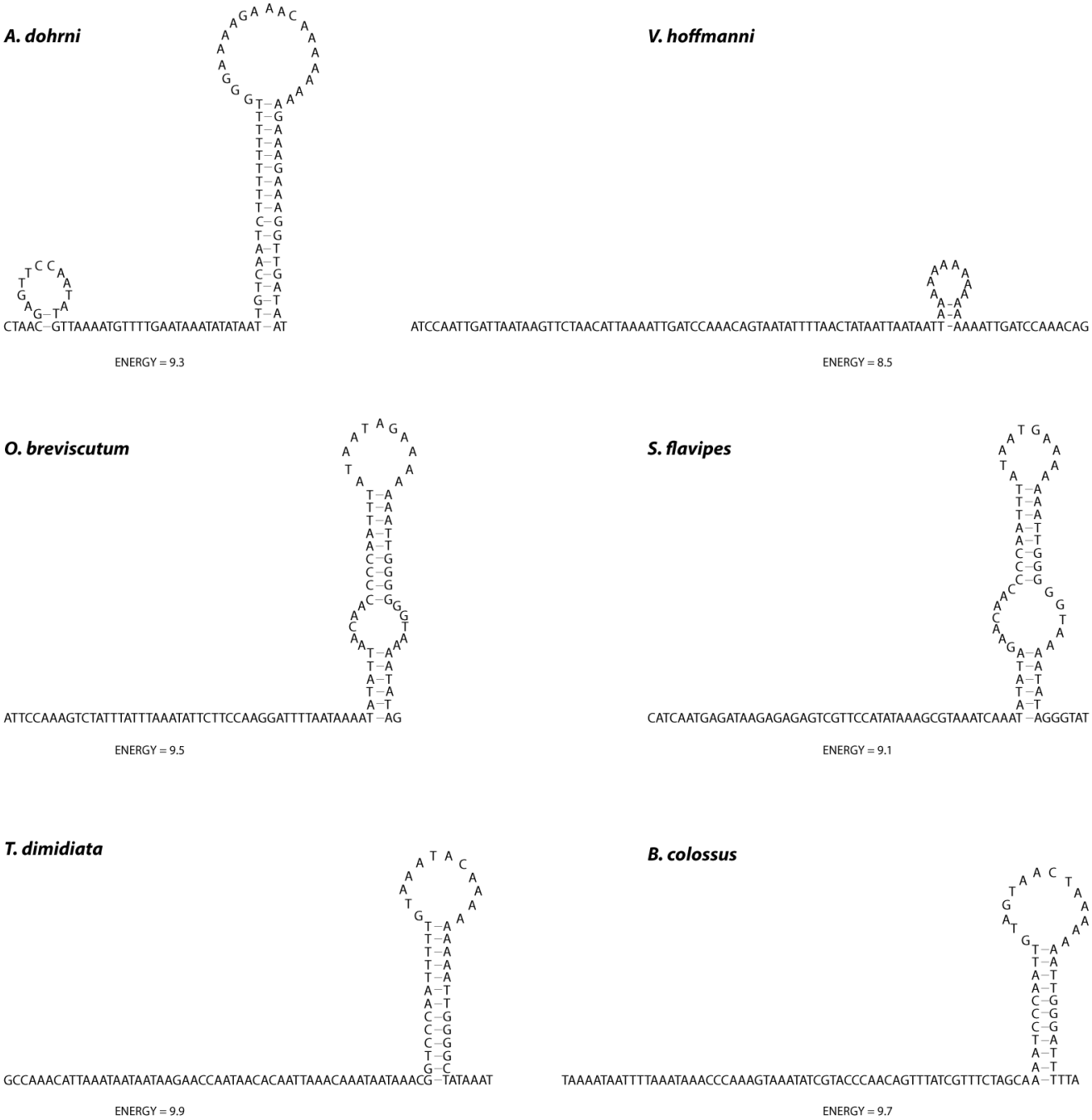
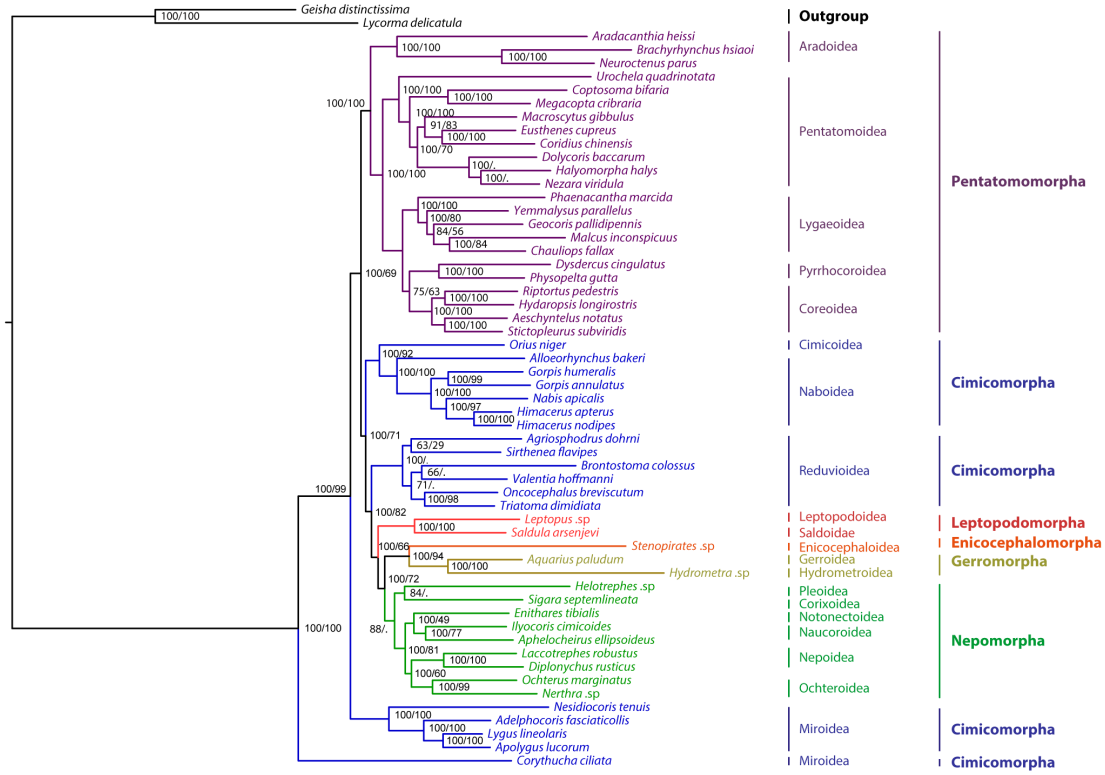



**Figure 5: Conserved stem loop structures inferred by the Turbofold algorithm on the the 100 bp of the control region flanking $tRNA_{Ile}$ of the six assassin bugs. Values of free energy are indicated above each structure.**

A




B


**Figure 6: (A) Phylogenetic tree inferred by Bayesian analysis from 55 heteropteran mitogenomes. Left numbers at the nodes indicate Bayesian posterior probabilities expressed in percentages. When the node was also present on the tree inferred by ML analysis, right numbers indicate bootstrap support values. A dot indicates that the node was absent on the tree inferred by ML analysis (B) Maximum likelihood (ML) sub-tree corresponding to Reduviidae family, with bootstrap values depicted on nodes.**

**Acknowledgements**

This work was supported by PO-FEDER TIMGED N°30195 (Trypanosomes d'Intérêt Médical en Guyane française – Epidémiologie et Diagnostic) and "Investissement d'Avenir" grants managed by Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX-25-01; TULIP, ANR-10-LABX-41, ANR-11-IDEX-0002-02) as well as project METABAR (ANR-11-BSV7-0020). We would like to thank Christine Aznar, Denis Blanchet, Jean-Pierre Dujardin and Jean-Michel Bérenger for the help with obtaining and identifying the specimen. We are grateful to the Genotoul bioinformatics platform Toulouse Midi-Pyrenees for providing computing and storage resources.

**References**

Abascal, F., Zardoya, R., Telford, M.J., 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38, W7–13.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., Saunders, N.C., 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. 18, 489–522.

Ballard, J.W.O., 2000. Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. J. Mol. Evol. 51, 48–63.

Bellaousov, S., Reuter, J.S., Seetin, M.G., Mathews, D.H., 2013. RNAstructure: web servers for RNA secondary structure prediction and analysis. Nucleic Acids Res. 41, W471–W474.

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573.

Bernt, M., Braband, A., Schierwater, B., Stadler, P.F., 2013a. Genetic aspects of mitochondrial genome evolution. Mol. Phylogenet. Evol. 69, 328–338.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013b. MITOS: Improved *de novo* metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69, 313–319.

Besnard, G., Christin, P.-A., Malé, P.-J.G., Coissac, E., Ralimanana, H., Vorontsova, M.S., 2013. Phylogenomics and taxonomy of Lecomtelleae (Poaceae), an isolated panicoid lineage from Madagascar. Ann. Bot. 112, 1057–1066.

Besnard, G., Jühling, F., Chapuis, É., Zedane, L., Lhuillier, É., Mateille, T., Bellafiore, S., 2014. Fast assembly of the mitochondrial genome of a plant parasitic nematode (*Meloidogyne graminicola*) using next generation sequencing. C. R. Biol. 337, 295–301.

Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27, 1767–1780.

Boore, J.L., 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. In: Sankoff, D. and Nadeau, J.H. eds. Comparative Genomics. Springer Netherlands, pp. 133–147.

Boore, J.L., Collins, T.M., Stanton, D., Daehler, L.L., Brown, W.M., 1995. Deducing the pattern of arthropod phylogeny from mitochondrial-DNA rearrangements. Nature 376, 163-165.

490 Boore, J.L., Lavrov, D.V., Brown, W.M., 1998. Gene translocation links insects and crustaceans.
491 Nature 392, 667–668.

492 Clary, D.O., Wolstenholme, D.R., 1985. The mitochondrial DNA molecule of *Drosophila yakuba*:
493 nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. 22, 252–271.

494 Clayton, D.A., 1992. Transcription and replication of animal mitochondrial DNAs. Int Rev Cytol 141,
495 217–232.

496 Crozier, R.H., Crozier, Y.C., 1993. The mitochondrial genome of the honeybee *Apis mellifera*:
497 complete sequence and genome organization. Genetics 133, 97–117.

498 Dai, X., Xun, H., Chang, J., Zhang, J., Hu, B., Li, H., Yuan, X., Cai, W., 2012. The complete mitochondrial
499 genome of the plant bug *Nesidiocoris tenuis* (Reuter) (Hemiptera: Miridae: Bryocorinae: Dicyphini).
500 Zootaxa (3554), 30–44.

501 De Bruijn, M.H., 1983. *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic
502 code. Nature 304, 234-241.

503 Dotson, E.M., Beard, C.B., 2001. Sequence and organization of the mitochondrial genome of the
504 Chagas disease vector, *Triatoma dimidiata*. Insect Mol. Biol. 10, 205–215.

505 Dougherty, V., 1995. A review of the new world Ectrichodiinae genera (Hemiptera: Reduviidae).
506 Trans. Am. Entomol. Soc. 121,173–225.

507 Doyle, S.R., Griffith, I.S., Murphy, N.P., Strugnell, J.M., 2014. Low-coverage MiSeq next generation
508 sequencing reveals the mitochondrial genome of the Eastern Rock Lobster, *Sagmariasus verreauxi*.
509 Mitochondrial DNA 0,1–2.

510 Gao, J., Li, H., Truong, X.L., Dai, X., Chang, J., Cai, W., 2013. Complete nucleotide sequence and
511 organization of the mitochondrial genome of *Sirthenea flavipes* (Hemiptera: Reduviidae: Peiratinae)
512 and comparison with other assassin bugs. Zootaxa 3669, 1-16.

513 Garesse, R., 1988. *Drosophila melanogaster* mitochondrial DNA: gene organization and evolutionary
514 considerations. Genetics 118, 649–663.

515 Gil-Santana, H., Baena, M., 2009. Two new species of *Brontostoma* Kirkaldy (Hemiptera: Heteroptera:
516 Reduviidae: Ectrichodiinae) from Bolivia, with description of the male genitalia of two other species
517 of the genus, and description of the female of *B. doughertyae* Gil-Santana, Lopes, Marques &
518 Jurberg. Zootaxa 1979, 41–52.

519 Gissi, C., Iannelli, F., Pesole, G., 2008. Evolution of the mitochondrial genome of Metazoa as
520 exemplified by comparison of congeneric species. Heredity 101, 301–320.

521 Hahn C, Bachmann L, Chevreux B, 2013. Reconstructing mitochondrial genomes directly from
522 genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids
523 Research, 41, e129.

524 Harmanci, A.O., Sharma, G., Mathews, D.H., 2011. TurboFold: Iterative probabilistic estimation of
525 secondary structures for multiple RNA sequences. BMC Bioinformatics 12, 108.

526 Hebsgaard, M.B., Andersen, N.M., Damgaard, J., 2004. Phylogeny of the true water bugs
527 (Nepomorpha: Hemiptera–Heteroptera) based on 16S and 28S rDNA and morphology. Syst. Entomol.
528 29, 488–508.

Hua, J., Li, M., Dong, P., Cui, Y., Xie, Q., Bu, W., 2008. Comparative and phylogenomic studies on the mitochondrial genomes of Pentatomomorpha (Insecta: Hemiptera: Heteroptera). BMC Genomics 9, 610.

Hua, J., Li, M., Dong, P., Cui, Y., Xie, Q., Bu, W., 2009. Phylogenetic analysis of the true water bugs (Insecta: Hemiptera: Heteroptera: Nepomorpha): evidence from mitochondrial genomes. BMC Evol. Biol. 9, 134.

Junqueira, A.C.M., Lessinger, A.C., Torres, T.T., da Silva, F.R., Vettore, A.L., Arruda, P., Azeredo Espin, A.M.L., 2004. The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). Gene 339, 7–15.

Kim, I., Lee, E.M., Seol, K.Y., Yun, E.Y., Lee, Y.B., Hwang, J.S., Jin, B.R., 2006. The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis* (Lepidoptera: Lycaenidae). Insect Mol. Biol. 15, 217–225.

Kopp, A., True, J.R., 2002. Phylogeny of the Oriental *Drosophila melanogaster* species group: a multilocus reconstruction. Syst. Biol. 51, 786–805.

Kück, P., Meusemann, K., 2010. FASconCAT: Convenient handling of data matrices. Mol. Phylogenet. Evol. 56, 1115–1118.

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.

Lavrov, D.V., Boore, J.L., Brown, W.M., 2000. The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. Mol. Biol. Evol. 17, 813–824.

Lavrov, D.V., Lang, B.F., 2005. Transfer RNA gene recruitment in mitochondrial DNA. Trends Genet. 21, 129–133.

Lee, W.-J., Conroy, J., Howell, W.H., Kocher, T.D., 1995. Structure and evolution of teleost mitochondrial control regions. J. Mol. Evol. 41, 54–66.

Lee, E.-S., Shin, K.S., Kim, M.-S., Park, H., Cho, S., Kim, C.-B., 2006. The mitochondrial genome of the smaller tea tortrix *Adoxophyes honmai* (Lepidoptera: Tortricidae). Gene 373, 52–57.

Lee, W., Kang, J., Jung, C., Hoelmer, K., Lee, S.H., Lee, S., 2009. Complete mitochondrial genome of brown marmorated stink bug *Halyomorpha halys* (Hemiptera: Pentatomidae), and phylogenetic relationships of hemipteran suborders. Mol. Cells 28, 155–165.

Lent, H., Wygodzinsky, P.W., 1979. Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. Bulletin of the AMNH; v. 163, article 3.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Gao, J., Cai, W., 2013. Complete mitochondrial genome of the assassin bug *Oncocephalus breviscutum* (Hemiptera: Reduviidae). Mitochondrial DNA 0,1–2.

Li, H., Gao, J., Liu, H., Liu, H., Liang, A., Zhou, X., Cai, W., 2011. The architecture and complete sequence of mitochondrial genome of an assassin bug *Agriosphodrus dohrni* (Hemiptera: Reduviidae). Int. J. Biol. Sci. 7, 792.

Li, H., Liu, H., Song, F., Shi, A., Zhou, X., Cai, W., 2012a. Comparative mitogenomic analysis of damsel bugs representing three tribes in the family Nabidae (Insecta: Hemiptera). PloS One 7, e45925.

Li, H., Liu, H., Shi, A., Štys, P., Zhou, X., Cai, W., 2012b. The complete mitochondrial genome and novel gene arrangement of the unique-headed bug *Stenopirates* sp. (Hemiptera: Enicocephalidae). PloS One 7, e29419.

Li, M., Tian, Y., Zhao, Y., Bu, W., 2012c. Higher Level Phylogeny and the First Divergence Time Estimation of Heteroptera (Insecta: Hemiptera) Based on Multiple Genes. PLoS ONE 7, e32152.

Li, H., Shi, A., Song, F., Cai, W., 2014. Complete mitochondrial genome of the flat bug *Brachyrhynchus hsiaoi* (Hemiptera: Aradidae). Mitochondrial DNA 0,1–2.

Lin, C.-P., Danforth, B.N., 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Mol. Phylogenet. Evol. 30,

Liu, L., Li, H., Song, F., Song, W., Dai, X., Chang, J., Cai, W., 2012. The mitochondrial genome of *Coridius chinensis* (Hemiptera: Dinidoridae). Zootaxa 3537, 29–40.

Macey, J.R., Schulte, J.A., Larson, A., Papenfuss, T.J., 1998. Tandem duplication via light-strand synthesis may provide a precursor for mitochondrial genomic rearrangement. Mol. Biol. Evol. 15, 71–75.

Mahner, M., 1993. Systema Cryptoceratorum Phylogeneticum (Insecta, Heteroptera). Stuttgart, Germany: Eds. Schweizerbart.

Maldonado Capriles, J., 1990. Systematic catalogue of the Reduviidae of the world (Insecta: Heteroptera). Caribb. J. Sci. Spec. Ed.

Malé, P.-J.G., Bardon, L., Besnard, G., Coissac, E., Delsuc, F., Engel, J., Lhuillier, E., Scotti-Saintagne, C., Tinaut, A., Chave, J., 2014. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Mol. Ecol. Resour.

Mariette, J., Escudié, F., Allias, N., Salin, G., Noirot, C., Thomas, S., Klopp, C., 2012. NG6: Integrated next generation sequencing storage and processing environment. BMC Genomics 13, 462.

Moritz, C., Brown, W.M., 1986. Tandem duplication of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science 233, 1425–1427.

Moritz, C., Dowling, T.E., Brown, W.M., 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. 269–292.

Nagaike, T., 2005. Human Mitochondrial mRNAs Are Stabilized with Polyadenylation Regulated by Mitochondria-specific Poly(A) Polymerase and Polynucleotide Phosphorylase. J. Biol. Chem. 280, 19721–19727.

Nguyen, V.H., Lavenier, D., 2009. PLAST: parallel local alignment search tool for database comparison. BMC Bioinformatics 10, 329.

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

Oliveira, M.T., Barau, J.G., Junqueira, A.C.M., Feijão, P.C., Rosa, A.C. da, Abreu, C.F., Azeredo-Espin, A.M.L., Lessinger, A.C., 2008. Structure and evolution of the mitochondrial genomes of *Haematobia*

606 *irritans* and *Stomoxys calcitrans*: the Muscidae (Diptera: Calyptratae) perspective. Mol. Phylogenet.
607 Evol. 48, 850–857.

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

610 Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed
611 models. Bioinformatics 19, 1572–1574.

612 Saito, S., 2005. Replication origin of mitochondrial DNA in insects. Genetics 171, 1695–1705.

613 Salvato, P., Simonato, M., Battisti, A., Negrisolo, E., 2008. The complete mitochondrial genome of the
614 bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). BMC Genomics 9, 331.

615 Schuh, R.T., Weirauch, C., Wheeler, W.C., 2009. Phylogenetic relationships within the Cimicomorpha
616 (Hemiptera: Heteroptera): a total-evidence analysis. Syst. Entomol. 34, 15–48.

617 Shao, R., Barker, S.C., 2003. The highly rearranged mitochondrial genome of the plague thrips, *Thrips*
618 *imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary
619 arrangement of rRNA genes. Mol. Biol. Evol. 20, 362–370.

620 Shao, R., Barker, S.C., Mitani, H., Takahashi, M., Fukunaga, M., 2006. Molecular mechanisms for the
621 variation of mitochondrial gene content and gene arrangement among chigger mites of the genus
622 *Leptotrombidium* (Acari: Acariformes). J. Mol. Evol. 63, 251–261.

623 Shi, A.M., Li, H., Bai, X.S., Dai, X., Chang, J., Guilbert, E., Cai, W., 2012. The complete mitochondrial
624 genome of the flat bug *Aradacanthia heissi* (Hemiptera: Aradidae). Zootaxa 3238, 23–38.

625 Signorovitch, A.Y., Buss, L.W., Dellaporta, S.L., 2007. Comparative Genomics of Large Mitochondria in
626 Placozoans. PLoS Genet. 3, e13.

627 Som, A., 2014. Causes, consequences and solutions of phylogenetic incongruence. Brief. Bioinform.
628 doi:10.1093/bib/bbu015

629 Song, N., Liang, A.-P., 2009. Complete Mitochondrial Genome of the Small Brown Planthopper,
630 *Laodelphax striatellus* (Delphacidae: Hemiptera), with a Novel Gene Order. Zoolog. Sci. 26, 851–860.

631 Song, N., Liang, A.-P., Bu, C.-P., 2012. A molecular phylogeny of Hemiptera inferred from
632 mitochondrial genome sequences. PloS One 7, e48778.

633 Song, W., Li, H., Song, F., Liu, L., Wang, P., Xun, H., Dai, X., Chang, J., Cai, W., 2013. The complete
634 mitochondrial genome of a tessaratomid bug, *Eusthenes cupreus* (Hemiptera: Heteroptera:
635 Pentatomomorpha: Tessaratomidae). Zootaxa 3620, 260–272.

636 Stamatakis, A., 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large
637 Phylogenies. Bioinformatics btu033.

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

640 Straub, S.C., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., Liston, A., 2012. Navigating the tip of
641 the genomic iceberg: Next-generation sequencing for plant systematics. Am. J. Bot. 99, 349–364.

642 Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary
643 Genetics Analysis Version 6.0. Mol. Biol. Evol. 30, 2725–2729.

644 Thao, M.L., Baumann, L., Baumann, P., 2004. Organization of the mitochondrial genomes of
645 whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). BMC Evol. Biol. 4, 25.

646 Thompson, K.F., Patel, S., Williams, L., Tsai, P., Constantine, R., Baker, C.S., Millar, C.D., 2014. High
647 coverage of the complete mitochondrial genome of the rare Gray's beaked whale (*Mesoplodon grayi*)
648 using Illumina next generation sequencing. Mitochondrial DNA 0,1–2.

649 Tian, Y., Zhu, W., Li, M., Xie, Q., Bu, W., 2008. Influence of data conflict and molecular phylogeny of
650 major clades in Cimicomorphan true bugs (Insecta: Hemiptera: Heteroptera). Mol. Phylogenet. Evol.
651 47, 581–597.

652 Veale, A.J., Williams, L., Tsai, P., Thakur, V., Lavery, S., 2014. The complete mitochondrial genomes of
653 two chiton species (*Sypharochiton pelliserpentis* and *Sypharochiton sinclairi*) obtained using Illumina
654 next generation sequencing. Mitochondrial DNA 0,1–2.

655 Walberg, M.W., Clayton, D.A., 1981. Sequence and properties of the human KB cell and mouse L cell
656 D-loop regions of mitochondrial DNA. Nucleic Acids Res. 9, 5411–5421.

657 Wang, P., Li, H., Wang, Y., Zhang, J.-H., Dai, X., Chang, J., Hu, B.-W., Cai, W.-Z., 2013. The
658 mitochondrial genome of the plant bug *Apolygus lucorum* (Hemiptera: Miridae): Presently known as
659 the smallest in Heteroptera. Insect Sci. 21, 159-173.

660 Wang, Y., Li, H., Xun, H., Cai, W., 2014. Complete mitochondrial genome sequence of the plant bug
661 *Adelphocoris fasciaticollis* (Hemiptera: Heteroptera: Miridae). Mitochondrial DNA 0,1–2.

662 Weirauch, C., 2008. Cladistic analysis of Reduviidae (Heteroptera: Cimicomorpha) based on
663 morphological characters. Syst. Entomol. 33, 229–274.

664 Weirauch, C., Munro, J.B., 2009. Molecular phylogeny of the assassin bugs (Hemiptera: Reduviidae),
665 based on mitochondrial and nuclear ribosomal genes. Mol. Phylogenet. Evol. 53, 287–299.

666 Weirauch, C., Schuh, R.T., 2011. Systematics and Evolution of Heteroptera: 25 Years of Progress.
667 Annu. Rev. Entomol. 56, 487–510.

668 Wheeler, A.G., 1997. Zoological Catalogue of Australia 27.3 A: Hemiptera: Heteroptera
669 (Coleorrhyncha to Cimicomorpha). Ann. Entomol. Soc. Am. 90, 703–704.

670 Wheeler, W.C., Schuh, R.T., Bang, R., 1993. Cladistic relationships among higher groups of
671 Heteroptera: congruence between morphological and molecular data sets. Insect Syst. Evol. 24, 121–
672 137.

673 Wolstenholme, D.R., 1992. Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. 141,
674 173–216.

675 Xie, Q., Tian, Y., Zheng, L., Bu, W., 2008. 18S rRNA hyper-elongation and the phylogeny of
676 Euhemiptera (Insecta: Hemiptera). Mol. Phylogenet. Evol. 47, 463–471.

677 Yang, W., Yu, W., Du, Y., 2013. The complete mitochondrial genome of the sycamore lace bug
678 *Corythucha ciliata* (Hemiptera: Tingidae). Gene 532, 27–40.

679 Yu, S., Wang, Y., Rédei, D., Xie, Q., Bu, W., 2013. Secondary structure models of 18S and 28S rRNAs of
680 the true bugs based on complete rDNA sequences of *Eurydema maracandica* Oshanin, 1871
681 (Heteroptera, Pentatomidae). ZooKeys 319, 363–377.

682 Yukuhiro, K., Sezutsu, H., Itoh, M., Shimizu, K., Banno, Y., 2002. Significant levels of sequence
683 divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild
684 Mulberry silkmoth, *Bombyx mandarina*, and its close relative, the domesticated silkmoth, *Bombyx*
685 *mori*. Mol. Biol. Evol. 19, 1385–1389.

686 Yuting, D., Li, Jiang, P., Song, F., Ye, Z., Yuang, X., Dai, X., Chang, J., Cai, W., 2012. Sequence and
687 organization of the mitochondrial genome of an urostylidid bug, *Urochela quadrinotata* Reuter
688 (Hemiptera: Urostylididae). Entomotaxonomia 34, 613–623.

689 Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn
690 graphs. Genome Res. 18, 821–829.

691 Zhang, D.-X., Hewitt, G.M., 1997. Insect mitochondrial control region: a review of its structure,
692 evolution and usefulness in evolutionary studies. Biochem. Syst. Ecol. 25, 99–120.

693 Zhang, D.-X., Szymura, J.M., Hewitt, G.M., 1995. Evolution and structural conservation of the control
694 region of insect mitochondrial DNA. J. Mol. Evol. 40, 382–391.

695 Zhang, Q.-L., Yuan, M.-L., Shen, Y.-Y., 2013. The complete mitochondrial genome of *Dolycoris*
696 *baccarum* (Insecta: Hemiptera: Pentatomidae). Mitochondrial DNA 24, 469–471.

697 Zhou, Z., Huang, Y., Shi, F., 2007. The mitochondrial genome of *Ruspolia dubia* (Orthoptera:
698 Conocephalidae) contains a short A+T-rich region of 70 bp in length. Genome Natl. Res. Counc. Can.
699 50, 855–866.

700