# The mitochondrial genome of a Texas outbreak strain of the cattle tick, *Rhipicephalus* (*Boophilus*) *microplus*, derived from whole genome sequencing Pacific Biosciences and Illumina reads.

John K. McCooke [a], Felix D. Guerrero [b^], Roberto A Barrero [a], Michael Black [a], Adam Hunter [a], Callum Bell [c], Faye Schilkey [c], Robert J Miller [d], Matthew I Bellgard [a, c]

[a] Centre for Comparative Genomics, Murdoch University, Murdoch, Australia

[b] United States Department of Agriculture, Agricultural Research Service, Knipling-Bushland U. S. Livestock Insects Research Laboratory, Kerrville, USA

[c] National Center for Genome Resources, Santa Fe, NM, USA

[d] United States Department of Agriculture, Agricultural Research Service, Cattle Fever Tick Research Laboratory, Edinburg, USA

**^ Corresponding Author:**

Felix D. Guerrero

Felix.Guerrero@ars.usda.gov

1-830-792-0327

1-830-792-0314 FAX

**Working title:** Assembly of *Rhipicephalus microplus* mitogenome

**Keywords:** Cattle Tick, mitochondrial genome, Pac Bio

**Abstract**

The cattle fever tick, *Rhipicephalus* (*Boophilus*) *microplus* is one of the most significant medical veterinary pests in the world, vectoring several serious livestock diseases negatively impacting agricultural economies of tropical and subtropical countries around the world. In our study, we assembled the complete *R. microplus* mitochondrial genome from Illumina and Pac Bio sequencing reads obtained from the ongoing *R. microplus* (Deutsch strain from Texas, USA) genome sequencing project. We compared the Deutsch strain mitogenome to the mitogenome from a Brazilian *R. microplus* and from an Australian cattle tick that has recently been taxonomically designated as *Rhipicephalus australis* after previously being considered *R. microplus*. The sequence divergence of the Texas and Australia ticks is much higher than the divergence between the Texas and Brazil ticks. This is consistent with the idea that the Australian ticks are distinct from the *R. microplus* of the Americas.

## 1. Introduction

The cattle tick, *Rhipicephalus* (*Boophilus*) *microplus* is a significant pest of cattle in tropical and subtropical regions of the world. The economic impact of this ectoparasite has been estimated to be 3.2$US billion annually in Brazil alone (Grisi et al., 2014). The huge economic toll this tick places upon the agricultural economy of many countries has driven efforts to discover effective control measures that cattle producers can implement to maximize the productivity of their livestock operations. Current control methods rely almost exclusively upon chemical acaricides and tick populations have developed multiple mechanisms of acaricide resistance to overcome this control method. Emerging DNA technologies have made sequencing the complex cattle tick genome a possibility and the ongoing project to sequence the genome of the Texas outbreak strain of *R. microplus*, designated the Deutsch strain, is hosted at CattleTickBase (Bellgard et al., 2012). As an early outcome of this sequencing project, mitochondrial gene sequences from the Deutsch strain were obtained from data generated by both Pacific Biosciences- and Illumina-based sequencing protocols. We report the complete mitochondrial sequence of the Deutsch Texas strain of *R. microplus*. This is the first complete mitochondrial sequence from a North American population of cattle tick. The long Pac Bio reads enabled us to resolve the sequence of the variable tandem repeat region of the mitochondrial genome that had proven difficult to determine in earlier studies based on Sanger (Campbell et al., 1999) or short read (Burger et al., 2014) technologies. Comparisons between the Deutsch tick mitogenome and those from

Brazil and Australia reveal a much closer relationship between the Texas and Brazilian cattle ticks than the Texas and Australian cattle ticks. Our findings are consistent with the molecular phylogenetic analysis by Burger et al. (2014) and our Texas tick strain is apparently from their study's *R. microplus* Clade A. Our results also add support for the recent reclassification of *R. microplus* populations from Australia to *Rhipicephalus australis* (Estrada-Pena et al., 2012).

## 2. Materials and Methods

### *2.1. Source of tick materials*

Genomic DNA was extracted from eggs of the Deutsch strain of *R. microplus*. This strain has been maintained in colony at the Cattle Fever Tick Research Laboratory, Edinburg, TX since its collection during an outbreak in South Texas which occurred in 2001 (Davey et al., 1980). Tick eggs are collected at each generation and frozen and stored at -80°C as part of our laboratory tick-rearing protocols. The frozen eggs from the f7, f10, f11, and f12 generation collections served as the source of genomic DNA. The strain was started from only a few engorged females and has been inbred since its collection, but is not genetically homogeneous. Eggs from this group of 50 females were pooled in a petri plate, mixed with a flat spatula, and weighed into vials to contain a final amount of approximately 3 g each.

*2.2. DNA Sequencing*

A total of 10 g of tick eggs was used to extract very high molecular weight genomic DNA following the protocol from Sambrook et al. (1989). The extraction protocol consists of grinding material in aqueous buffer, RNAse treatment, digestion by proteinase K, phenol extraction, followed by 4 d of dialysis in 50 mM Tris, 10 mM EDTA, pH 8.0 dialysis buffer, changing buffer twice daily. A total of 15.6 mg of genomic DNA was recovered and stored at 4 $^{o}$C. RNA-free status, DNA size and integrity were verified by agarose gel electrophoresis and the size was determined to be >200kb (Guerrero et al., 2010). A portion of the genomic DNA was processed by Cot filtration as described in Guerrero et al. (2010) to a Cot cloning value of 69.6 M. s. (Lamoureux et al., 2005) to enrich for single, low copy, and moderately repetitive DNAs. Input DNA for both the Pacific Biosciences and Illumina sequencing was quality checked for quantity and size using a Qubit fluorometer (Life Technologies, Grand Island, NY, USA) and Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

The Pacific Biosciences sequencing was performed at National Center for Genome Resources (Santa Fe, NM, USA). Two μg of genomic DNA (not Cot-selected) was sheared to ~8kb with the Covaris G-tube according to manufacturer's instructions (Covaris Inc., Woburn, MA, USA). Five DNA libraries were prepared according to the Pacific Biosciences low-input 10kb library preparation and sequencing protocol which includes DNA damage repair, end repair, SMRTbell adapter ligation, and an exonuclease step to remove failed ligation products. The

libraries were sequenced on 178 SMRT cells using C2 chemistry and XL polymerase, yielding 10,499,989 reads with a minimum length of 1 kb, representing 31,054,642,957 bases. The average and median read length was 2,957 bases and 2,234 bases, respectively. The longest read was 26, 364 bases. Overall, the Pac Bio reads represent ~4X coverage of the cattle tick genome.

The Illumina-based sequencing was performed at the National Center for Genome Resources (Santa Fe, NM, USA) using the standard Illumina DNA library preparation protocol. The TruSeq DNA Sample Preparation V2 kit (Illumina, CA, USA) was used to generate sequencing libraries from the Cot-selected genomic DNA described above. Two μg of DNA was fragmented using the Covaris S2 system according to the manufacturer's protocol (Covaris Inc.). The resulting overhang of the dsDNA fragments was end-repaired with the End Repair Mix of the V2 kit by incubating at 30°C for 30 min. The polished fragments were then phosphorylated by T4 polynucleotide kinase, followed by the addition of a single A nucleotide to the 3′ end by incubating the end repaired fragments with A-Tailing mix at 37°C for 30min. The fragment-adapter ligation occurred at 30°C for 10 min, after which the ligated product was size-selected by gel electrophoresis, the library fragment range was visualized under brief ultraviolet light and the desired size range of 300-400 bp excised and subjected to a final PCR amplification step of 10 cycles. All amplified libraries were quantitatively and qualitatively assessed by Nanodrop ND-1000 (Thermo Scientific, DE, USA) and DNA bioanalyzer 2100 (Agilent, CA, USA), respectively. The Cot-selected genomic DNA library was sequenced as 100 nt-paired ends on three lanes in a flowcell using HiSeq2000. Following sequencing, the raw

reads were processed by the Illumina pipeline and further by the NCGR contaminant filtering pipeline to remove adapter dimers, PCR primers, unused indexes, and Illumina PhiX control sequences, among others. This process yielded approximately 185 million high-quality reads per lane for a total of approximately 555 million reads and 111 billion bases.

*2.3 Bioinformatics*

To *de novo* assemble the *R. microplus* mitochondrial genome, we first identified conserved mitochondrial sequence motifs by the comparison of *Rhipicephalus sanguineus* (NC_002074), *Haemaphysalis flava* (NC_005292) and *Amblyomma triguttatum* (NC_005963) mitochondrial sequences sourced from NCBI (www.ncbi.nlm.nih.gov), as these tick mitochondria share a common genome organization. At the time of our analysis, these were the most complete tick mitochondrial genomes available for public download at NCBI. Conserved mitochondrial sequence motifs were then used to retrieve Illumina reads from the Cot-selected DNA sequence dataset (described above) with sequence similarity to the mitochondrial genome (Langmead and Salzberg, 2012). Mitochondrial genome-enriched Illumina reads were *de novo* assembled using Velvet (Zerbino and Birney, 2008) with an optimal k-mer size determined using VelvetOptimizer set to iterate over a k-mer range from 31 to 63 with a step of 4. Assembled Illumina contigs were initially validated against the *R. sanguineus* genome. To further extend the *R. microplus* mitochondrial genome coverage, we used the Illumina assembled contigs

to identify raw Pacbio reads with sequence similarity to validated mitochondrial contigs. These raw Pacbio reads were error-corrected using LSC 1.alpha (Au et al., 2012) and then utilized to resolve repeats within the *R. microplus* Texas mitochondrial genome.

ClustalW from the MacVector 12.7.5 software suite (MacVector, Inc., Cary, NC, USA) was used to align the the mitochondrial genomes of *R. microplus* Brazil (GenBank Accession No. KC503261), *R. microplus* Cambodia (GenBank Accession No. KC503260), *R. microplus* China (GenBank Accession No. KC503259), *R. australis* (GenBank Accession No. KC503255), *R. microplus* Texas (this study, GenBank Accession No. KP143546), and a partial mitochondrial genome of *Rhipicephalus annulatus* Romania (GenBank Accession No. KC503256), using a gap penalty of 10.0, an extend gap penalty of 5.0, a delay divergent of 40%, and weighted transitions. The phylogram was constructed from the ClustalW using MacVector and Neighbor Joining with 1000 reps in Bootstrap mode, the Tamura-Nei distance setting, and ignoring all gaps. Custom Blast datasets were created using BLAST+ version 2.2.29 (http://blast.ncbi.nlm.nih.gov; Altschul et al., 1990). The tRNA predictions were done with Aragorn at parameters of -t -mt -gcinvert -seq -fasta -jr4 -o (Laslett and Canback, 2004). MITOS was also used for tRNA prediction confirmation (Bernt et al., 2013). For conceptual translations of open reading frames, the MacVector 12.7.5 software suite was used (MacVector, Inc.) with the invertebrate mitochondrial codon table.

The ratio of nonsynonymous (Ka) and synonymous (Ks) substitution rates (Ka/Ks) between *R. microplus* and *R. australis* mitochondrial genes was calculated

with the KaKs Calculation v2.0 program (http://sourceforge.net/projects/kakscalculator2). A maximum likelihood approach was taken with optimal model selection for each gene via Akaike information criterion corrected for finite sample sizes. Preliminary Needleman-Wunsch global pairwise alignment was performed with the needle algorithm from EMBOSS tools with FASTA format output. This output was adjusted to the required axt format for Kaks Calculation 2.0 and Ka/Ks values graphed according to gene order in the tick mtDNA genome.

## 3. Results and Discussion

### *3.1. Sequencing and de novo assembling the mitogenome*

The complete mitochondrial genome of the Deutsch strain of *R. microplus* was assembled via a combination of sequence sources and assembly approaches (GenBank Accession No. KP143546; Supplemental File 1). Unselected and Cot-selected genomic DNA were used as the template for Pac Bio and Illumina sequencing, respectively. As our first step, the mitochondrial sequences of *R. sanguineous, H. flava* and *A. triguttatum* were obtained from GenBank and aligned to determine regions of conservation between these three species from different metastriate tick genera. Sequence regions greater than 7 continuous bases that were strictly conserved between the 3 species were used to extract reads with sequence identity from the 100 nt paired-end Cot-selected genomic DNA sequence data set (Guerrero et al. 2010). This step created a mitochondrial sequence-enriched paired

end sequence data set. We used this data set and bowtie2 to perform our initial assembly of the cattle tick mitochondrial genome, aligning to the *R. sanguineous* mitochondrial genome as a reference. An allowed alignment error of 22% was used for the initial alignment and an initial consensus sequence from the alignment was generated using a 75% base call identity. All gaps and N's were removed, however, the length of the initial consensus sequence was not trimmed and was used at a later stage to circularize the sequence. The initial consensus sequence was used to reanalyze with a new iteration under more restrictive parameters, producing a new alignment. A total of 6 iterations of consensus calling and correcting was used to generate the final consensus sequence. Upon each iteration, the allowed mismatch ratio was reduced by 4 mismatches until a final alignment allowing 2 mismatches was achieved. After the final iteration, a total of 60-80 million Illumina reads (count was different for each iteration number and the percentage of mismatch allowed) aligned to the reference R. *sanguineous* mitogenome and the final assembly consisted of 20 mitochondrial genome sequence contigs which covered approximately 72% of the mitochondrial genome. These 20 contigs were used to search CattleTickBase (Bellgard et al., 2012) for sequences with nucleotide identity. CattleTickBase is a database containing all the sequences from the *R. microplus* genome sequencing project (presently at ~4X coverage with Pac Bio) and would include mitochondrion-derived sequences from the Deutsch Texas cattle tick studied here. Following this process, the entire set of the 20 contigs and the new sequences identified from CattleTickBase were reassembled. The resulting assembly

covered approximately 80% of the mitochondrial genome when compared to the *R. sanguineus* reference mitogenome.

The final stage involved using the 20 contigs from the Illumina-derived data set to screen Pac Bio sequence reads. Ninety-six Pac Bio reads were identified with significant sequence identity to the 20 Illumina mitochondrial contigs representing approximately 14X coverage of the mitochondrial genome. Because of the relatively high error rate of Pac Bio sequences, we aligned the Pac Bio and Illumina reads and used the Illumina reads to error-correct the Pac Bio sequences. Forty-eight of the Pac Bio reads could be corrected with a confidence/coverage score > 0.90 and these were used along with manual curation to complete the final mitochondrial genome sequence alignment and assembly. The resulting sequence was circularized to create the final mitochondrial genome sequence of *R. microplus* Deutsch Texas (GenBank Accession No. KP143546).

*3.2. Annotation of the Texas R. microplus mitogenome*

The complete mitochondrial genome of the Deutsch strain of *R. microplus* was annotated using the *R. sanguineous, H. flava* and *A. triguttatum* annotation files as a template. The annotated sequence is depicted in Figure 1 and given in detail in Supplemental File 2. The gene number and the order of features (genes , RNAs, control regions, and other miscellaneous regions) is conserved as compared with the recently sequenced *R. microplus* Brazil sample and *R. australis* from Australia (Burger et al., 2014). The complete sequence length of *R. microplus* Texas is 15,167 bp and contains a variable tandem repeat region consisting of a “tRNA-Glu” and a

"similar to NAD1" motif (Fig. 2 and 3) that is repeated 4 and 3 times, respectively. This region in the Deutsch Texas *R. microplus* mitogenome consists of nt 5892-6156 and is flanked by tRNA-Ser and the *ND1* gene. By contrast, Burger et al. (2014) reported an inability to discern the number of repeats in this variable tandem repeat region of *R. australis* due to their sequence data originating from short read sequencing technology that has a PCR step in the sample preparation protocol. Campbell et al. (2001) reported that PCR can introduce extra copies of template repeat regions during the amplification process, probably during the annealing step. However, our mitogenome sequence can definitively resolve the variable tandem repeat region because we have individual Pac Bio reads that span the entire 265 bp region. Additionally, PCR amplification is not part of the Pac Bio protocol. In the CattleTickBase Pac Bio dataset, we found 1 Pac Bio read that completely spanned the variable tandem repeat region and 2 Pac Bio reads that partially spanned this region. Following error correction, the mitogenome reads alignment allowed the definitive identification of the number of "tRNA-Glu" and "similar to NAD1" motifs in the Texas cattle tick.

*3.3. Comparison of the Texas R. microplus mitogenome to R. australis*

The recent reclassification of *R. microplus* populations from Australia to *R. australis* (Estrada-Pena et al., 2012), coupled with the phylogenetic analyses of Burger et al. (2014) and our definitive sequence through the variable tandem repeat region, led us to compare the gene coding regions of the mitogenomes of Texas *R. microplus* to the *R. australis*. We also included the mitogenomes from *R. microplus*

populations in Brazil, Cambodia, and China to examine possible variation between *R. microplus* from several continents (Supplemental File 3). Over the entire mitogenomes in the ClustalW alignment, 99.9 % of the Brazilian and Cambodian cattle tick nucleotides have identical aligned nucleotides in the Texas cattle tick, while 96.0 % of the Australian cattle tick (*R. australis*) nucleotides have identical aligned nucleotides in the Texas tick (Table 1). We also aligned the partial genome sequence from the related cattle tick, *Rhipicephalus annulatus* (GenBank Accession No. KC503256), and found 95.0 % nucleotide identity over the region available to align. Interestingly, the *R. microplus* from China seems to possess a similar level of mitogenome sequence divergence (6%) as *R. annulatus* when compared to the Texas *R. microplus* mitogenome. The phylogenetic relationships (Fig. 4) between the various *Rhipicephalus* cattle ticks reflect those reported by Burger et al. (2014). In that study, phylogenetic analyses based on both the *cox1* and 16S rRNA suggested that the China Clade B *R. microplus* (included in our Fig. 4) was likely a cryptic species within the existing *R. microplus* complex. Also, using this general measure, the Australian cattle tick appears to be a different species from the cattle ticks from Texas, Cambodia, and Brazil.

From the ClustalW alignment of the Texas, Brazilian, and Australian cattle ticks, we identified 401 and 9 synonymous nucleotide changes in the gene coding regions comparing the Texas mitogenome to that of the cattle tick from Australia and Brazil, respectively. We also identified 112 and 3 nonsynonymous nucleotide changes in the gene coding regions comparing the Texas mitogenome and that from Australia and Brazil, respectively (Supplemental File 4). The total of 513 and 12

overall nucleotide differences equate to differences of 3.38% and 0.08% between the Texas and Australian and the Texas and Brazilian cattle tick mitogenome gene coding regions, respectively. The nonsynonymous changes between the three cattle ticks are diagrammatically represented in Figure 5. Visual examination of Figure 5 and a check of a non-synonymous/synonymous ratio adjusted by gene length (Supplemental File 4) indicated a non-uniform mutational rate in the genes. Nucleotide differences between the Australian mitogenome and the two from the Americas seem to be concentrated within the *ATP6*, *ATP8*, *ND1*, and C-terminal half of the *ND5* genes. The *ND2* and *Cox1* genes appear to be less affected by nucleotide substitutions, as they had only 6 and 1 nonsynonymous nucleotide difference, respectively, between the Texas and Australian sequence. The nonuniform mutational rate in the mitochondrial gene comparisons was quantified by a Ka/Ks ratio analysis (Supplemental File 5). Figure 6 shows the most variable genes in this analysis were *ND5*, *ND4*, *ND4L*, and *ND1*. The least variable gene was *Cox1*, possessing 72 nucleotide differences between the Texas and Australian cattle ticks, and only 1 difference between the Texas and Brazilian ticks.

Codon usage was very similar between the Texas and Australian cattle tick mitogenomes (Supplemental File 6). The biggest difference was in the greater usage of ATA and a lesser usage of ATT as a start codon in the Texas ticks compared to the Australian. Mono- and di-nucleotide frequencies were almost identical between the two mitogenomes. The % frequency for A, G, C, and T in the Texas mitogenome was 38.8, 11.2, 9.1, and 40.9, while 38.9, 11.3, 8.8, and 41.0 in the Australian

mitogenome, respectively. All the di-nucleotide frequencies were within 0.1 % of each other.

This mitogenome analysis has revealed interesting differences in cattle tick samples from different countries. Some of these differences raise questions about existing species classifications of the *R. microplus* ticks. It is important to correctly identify and classify cattle tick populations because some tick control methodologies, particularly the use of anti-tick vaccines in cattle, show inconsistent results (Almazán et al. 2010). Genomic variation among tick population that have been misclassified as a single species might be the reason for this inconsistency.

**Conflict of interest statement**

The authors have no conflicts of interest to declare.

**Acknowledgements**

We wish to thank the sequencing team at NCGR, specifically Patricia Mena, Jennifer Jacobi, Nico Devitt, Peter Ngam, for their work in library preparation and sequencing. We also wish to acknowledge the assistance provided by Drs. Steven Kappes, Dan Strickman, and Adalberto Perez de Leon in securing funding for the research. M. Bellgard conducted part of this research via a fellowship under the OECD Co-operative Research Programme: Biological Resource Management for Sustainable Agriculture Systems. This work was supported by funding from Bioplatforms Australia Pty Ltd, through the National Collaborative Research

Infrastructure Strategy of the Australian Government and by the USDA-ARS Knipling Bushland US Livestock Insects Research Laboratory CRIS Project No. 6205-32000-031-00. For the computational analysis, supercomputing resources were made available through iVEC, the Western Australian hub for supercomputing. This article reports the results of research only. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. USDA is an equal opportunity provider and employer.

**References**

Almazán, C., Lagunes, R., Villar, M., Canales, M., Rosario-Cruz, R., Jongejan, F., de la Fuente, J., 2010. Identification and characterization of *Rhipicephalus* (*Boophilus*) *microplus* candidate protective antigens for the control of cattle tick infestations. Parasitology Research 106, 471-479.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403-410.

Au, K.F., Underwood, J.G., Lee, L., Wong, W.H., 2012. Improving PacBio long read accuracy by short read alignment. PLoS One 7, e46679.

Bellgard, M., Moolhuijzen, P.M., Guerrero, F.D., Schibeci, D., Rodriguez-Valle, M., Peterson, D.G., Dowd, S.E., Barrero, R., Hunter, A., Miller, R.J., Lew-Tabor, A.E., 2012. CattleTickBase: An integrated Internet-based bioinformatics resource for *Rhipicephalus* (*Boophilus*) *microplus*. International Journal for Parasitology 42, 161-169.

Bernt, M., Donath, A., Juhling, F., Externbrink, F., Florentz, C., Fritzsch, G., Putz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: Improved *de novo* metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution 69, 313-319.

Burger, T.D., Shao, R., Barker, S.C., 2014. Phylogenetic analysis of mitochondrial genome sequences indicates that the cattle tick, *Rhipicephalus* (*Boophilus*) microplus, contains a cryptic species. Molecular Phylogenetics and Evolution 76, 241-253.

Campbell, N.J., Barker, S.C., 1999. The novel mitochondrial gene arrangement of the cattle tick, *Boophilus microplus*: fivefold tandem repetition of a coding region. Molecular Biology and Evolution 16, 732-740.

Campbell, N.J.H., Sturm, R.A., Barker, S.C., 2001. Large mitochondrial repeats multiplied during the polymerase chain reaction. Molecular Ecology Notes 1, 336-340.

Davey, R.B., Garza, J. Jr., Thompson, G.D., 1980. Ovipositional biology of the cattle tick, *Boophilus annulatus* (Acari: Ixodidae), in the laboratory. Journal of Medical Entomology 17, 287-289.

Estrada-Peña, A., Venzal, J.M., Nava, S., Mangold, A., Guglielmone, A.A., Labruna, M.B., de la Fuente, J., 2012. Reinstatement of *Rhipicephalus* (*Boophilus*) *australis* (Acari: Ixodidae) with redescription of the adult and larval stages. Journal of Medical Entomology 49, 794-802.

Grisi, L., Leite, R.C., Martins, J.R., Barros, A.T.M., Andreotti, R., Cancado, P.H.D., Perez de Leon, A.A., 2014. Reassessment of the potential economic impact of cattle parasites in Brazil. Brazilian Journal of Veterinary Parasitology 23, 150-156.

Guerrero, F.D., Moolhuijzen, P., Peterson, D.G., Bidwell, S., Caler, E., Bellgard, M., Nene, V.M., Djikeng, A., 2010. Reassociation kinetics-based approach for partial genome sequencing of the cattle tick, *Rhipicephalus Boophilus microplus*. BMC Genomics, 11, 374.

Lamoureux, D., Peterson, D.G., Li, W., Fellers, J.P., Gill, B.S., 2005. The efficacy of Cot-based gene enrichment in wheat (*Triticum aestivum* L.). Genome, 48, 1120-1126.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357-359.

Laslett, D., Canback, B., 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Research 32, 11-16.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. In *Molecular Cloning. A Laboratory Manual* 2nd edition. pp. 9.17-9.19. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Research 18, 821-829.

**Figure Legends**

**Fig. 1.** Representation of the mitochondrial genome from Texas *R. microplus*. Gene coding regions are noted in green with direction of arrow indicating transcription direction. The tRNA-encoding regions and their direction of transcription are noted by purple text and arrows. Regions containing the rRNAs are noted in bright red arrows, while the control region and the control region duplicate are noted in gray. The region from approximately 5,700 to 6,300 encompasses the tandem repeat region analyzed in Burger et al. (2014).

**Fig. 2.** Representation of the variable tandem repeat region in Texas *R. microplus* compared to Brazilian *R. microplus* and *R. australis*. The *R. australis* sequence and the *R. microplus* sequence from the Brazilian population was from Burger et al. (2014) with GenBank IDs of KC503255 and KC503261, respectively. The Texas *R. microplus* sequence was assembled from Pac Bio- and Illumina-derived sequence described in this report (GenBank Accession No. KP143546). Linear regions of sequence are noted in gray, while sequence gaps are represented by solid black lines. Coding regions for tRNAs are represented by magenta arrows, the 3' end of the NAD1 gene coding region by a green arrow, and a NAD1-like repeat by gray arrows.

**Fig. 3.** Aligned sequences from the variable tandem repeat region. Sequences from the Texas Deutsch (GenBank Accession No. KP143546) and Brazilian (GenBank

Accession No. KC503261) cattle tick variable tandem repeat regions were manually aligned to the corresponding region from *R. australis* (GenBank Accession No. KC503255) to visualize the similarities and differences within this area of the mitochondrial sequence. The sequence region encoding the tRNA-serine, tRNA glutamic acid, a NAD1-like repeat, and the 3' end of the NAD1 gene are indicated by solid, dotted, bold solid, and double underlining, respectively. Gray shading was used to indicate nucleotide differences among the aligned sequences. Colons are used to indicate identity with the Australian nucleotide. Gaps inserted to optimize the alignment are noted by a dash (-). There is an insertion of 265 and 5 nucleotides in this region from the Texas Deutsch and Brazilian cattle ticks, respectively. The location of this insertion is between the AA noted in bold at nts 5891-5892 of the Australian sequence.

**Fig. 4.** Phylogram of the mitogenomes from the various *Rhipicephalus* cattle ticks. ClustalW was used to align the the mitochondrial genomes of *R. microplus* Brazil (GenBank Accession No. KC503261), *R. microplus* Cambodia (GenBank Accession No. KC503260), *R. microplus* China (GenBank Accession No. KC503259), *R. australis* (GenBank Accession No. KC503255), *R. microplus* Texas (this study, GenBank Accession No. KP143546), and a partial mitochondrial genome of *Rhipicephalus annulatus* Romania (GenBank Accession No. KC503256). A phylogram from the resulting alignment was derived using Neighbor Joining and 1000 bootstraps.

**Fig. 5.** Nucleotide differences between the mitochondrial genomes of *R. microplus* Texas, *R. microplus* Brazil, and *R. australis*. The mitochondrial genomes of *R. microplus* Brazil (GenBank Accession No. KC503261), and *R. australis* (GenBank Accession No. KC503255) were compared to *R. microplus* Texas (This study, GenBank Accession No. KP143546), so as to identify nonsynonymous nucleotide differences between the Texas cattle tick mitochondrial genome and those from Brazil and Australia. Genes are represented by dark arrows indicating the direction of transcription. Gaps in the alignment are indicated by a light gray line and nucleotide changes in each sequence, in reference to *R. microplus* (Texas), that result in nonsynonymous changes are represented by red positional markers.

**Fig. 6.** Ratios of nonsynonymous (Ka) to synonymous (Ks) substitution rates for each mitochondrial gene of *R. microplus* Texas and *R. australis*. The $\log_{10}$ of the Ka/Ks values have been plotted according to gene order in the tick mitochondrial genome.

**Table 1**

Nucleotide differences in sequenced mitogenomes from *Rhipicephalus* ticks compared to the Deutsch Texas outbreak *R. microplus*.

| ID | Variant nt No.[a] | Total nt No.[b] | Difference %[c] |
|---|---|---|---|
| *R. microplus* Brazil KC503261 | 18 | 14,905 | 0.12 |
| *R. microplus*Cambodia KC503260 | 59 | 14,903 | 0.40 |
| *R. australis* KC503255 | 605 | 14,891 | 4.06 |
| *R. microplus* China KC503259 | 868 | 14,864 | 5.84 |
| *R. annulatus* KC503256 | 469 | 8,549 | 5.49 |

[a] Number of nucleotides that differ from the sequence in question and the Deutsch mitogenome sequence in the ClustalW alignment of Supplemental File 3.

[b] Total number of nucleotides that are in the GenBank Accession file.

[c] Percentage difference = [(Variant No.)/Total nt No.] X 100.

Abbreviations

A: adenine
bp: base pairs
C: cytosine
DNA: deoxyribonucleic acid
ds: double-stranded
G: guanine
kb: kilobase
min: minutes
mt: mitochondria
nt: nucleotide
PCR: polymerase chain reaction
RNA: ribonucleic acid
T: thymine

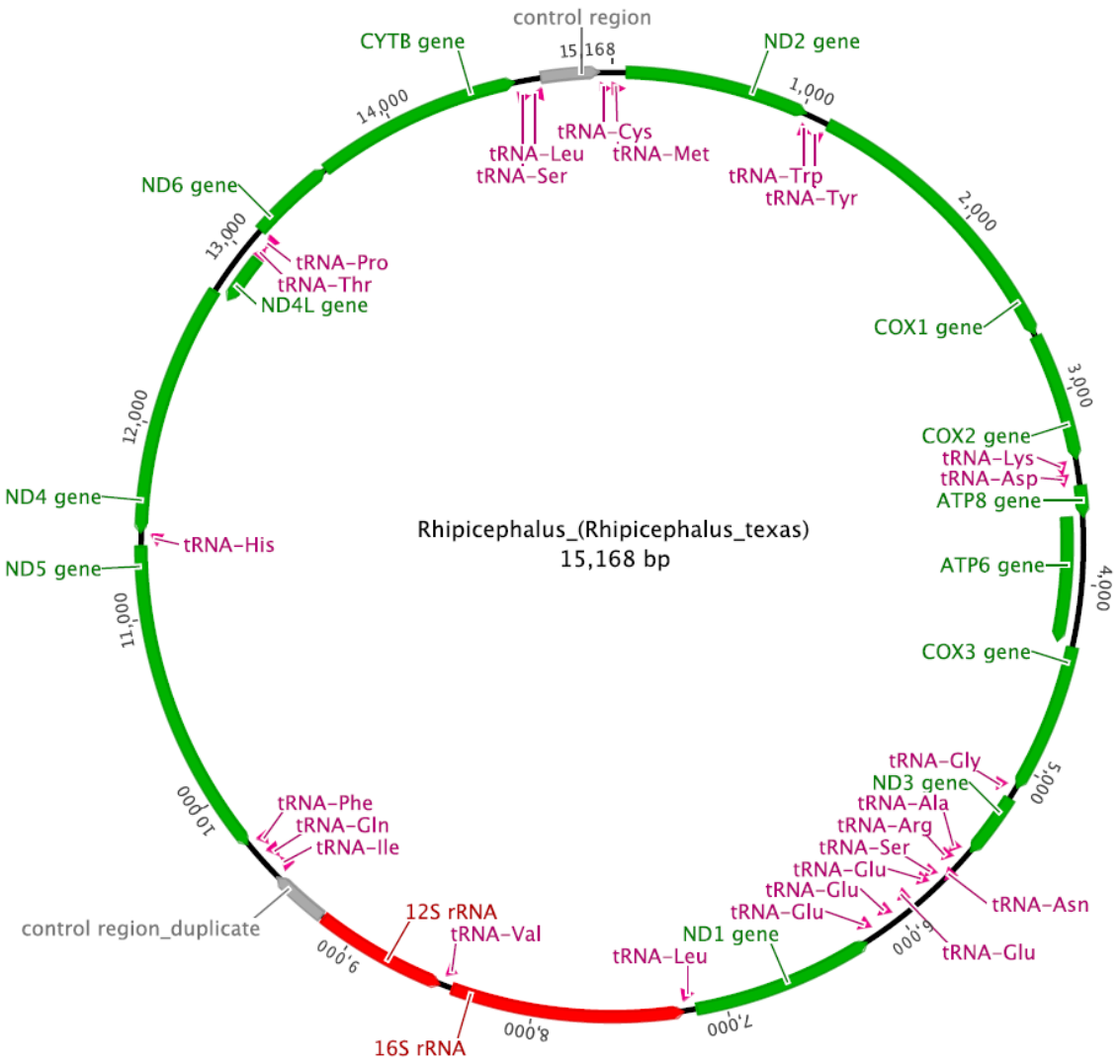


Fig. 1

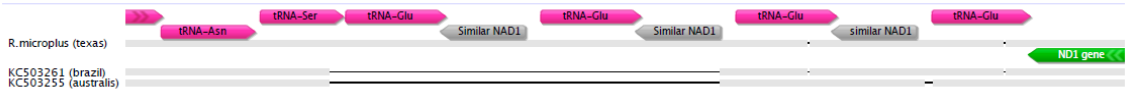


Fig. 2

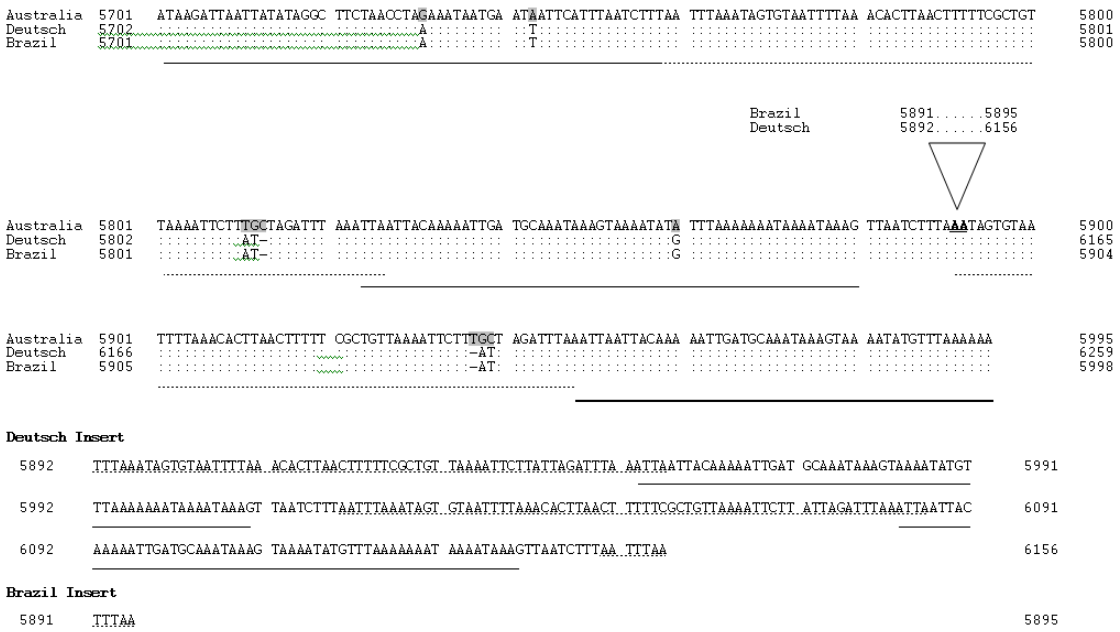


Fig. 3

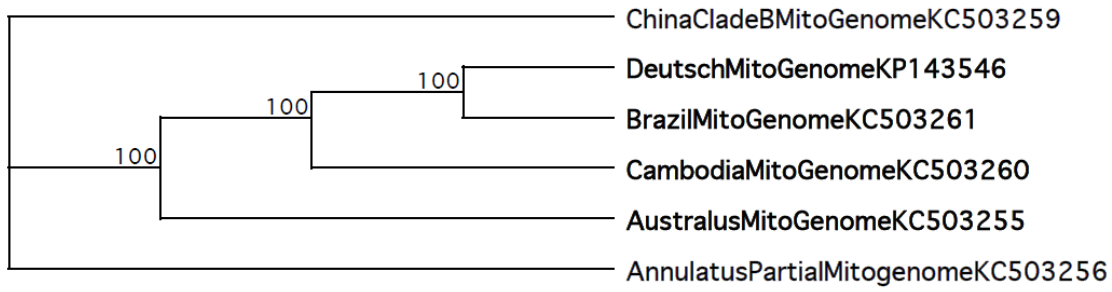


Fig. 4

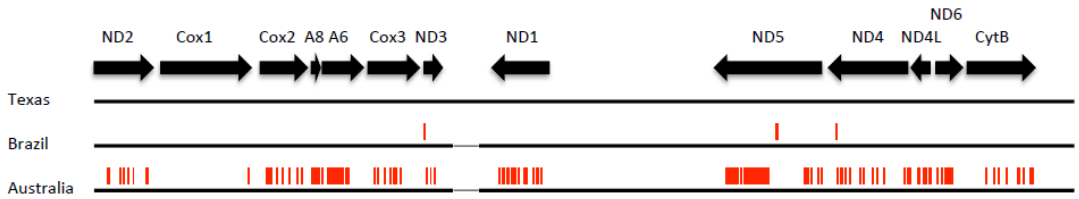


Fig. 5

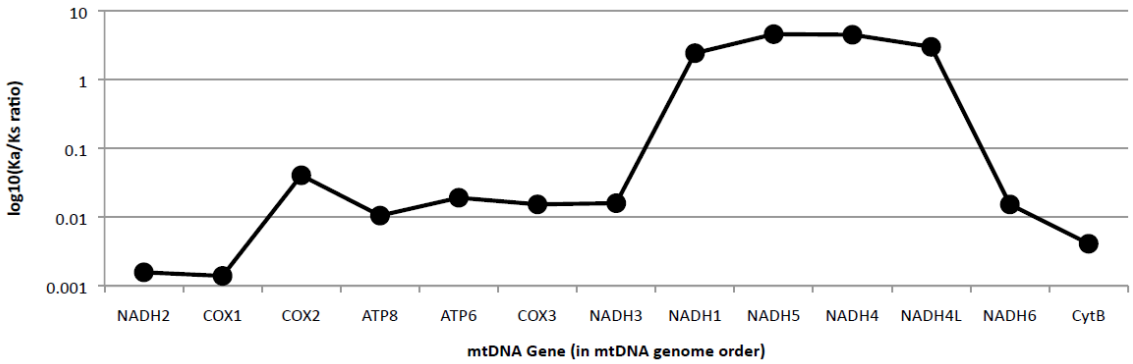


Fig. 6

Highlights

- We sequenced and assembled the mitochondrial genome of the Texas cattle tick.
- We describe phylogenetic associations of cattle ticks from several countries.
- Mitogenome of China tick differs from the Texas tick mitogenome.