Next-generation sequencing as means to retrieve tick systematic markers, with the focus on *Nuttalliella namaqua* (Ixodoidea: Nuttalliellidae)

Ben J. Mans[1,3,4*], Daniel de Klerk[1], Ronel Pienaar[1], Minique H. de Castro,[1,2] and Abdalla A. Latif[1,3]

[1]Parasites, Vectors and Vector-borne Diseases, Agricultural Research Council-Onderstepoort Veterinary Institute, Onderstepoort 0110, South Africa

[2]The Biotechnology Platform, Agricultural Research Council-Onderstepoort Veterinary Institute, Onderstepoort 0110, South Africa

[3]Department of Veterinary Tropical Diseases, University of Pretoria, Pretoria, South Africa

[4]Department of Life and Consumer Sciences, University of South Africa, South Africa

*Corresponding author: mansb@arc.agric.za

Abstract

Nuclear ribosomal RNA (18S and 28S rRNA) and mitochondrial genomes are commonly used in tick systematics. The ability to retrieve these markers using next-generation sequencing was investigated using the tick *Nuttalliella namaqua*. Issues related to nuclear markers may be resolved using this approach, notably, the monotypic status of *N. namaqua* and its basal relationship to other tick families. Four different Illumina datasets (~55 million, 100 bp reads each) were generated from a single tick specimen and assembled to give 350k-390k contigs. A genome size of ~1 Gbp was estimated with low levels of repetitive elements. Contigs (>1000 bp, >50-fold coverage) present in most assemblies (n=69), included host-derived 18S and 28S rRNA, tick and host-derived transposable elements, full-length tick 18S and 28 rRNA, the mitochondrial genome in single contig assemblies and the histone cassette. Coverage for the nuclear rRNA genes was above 1000-fold confirming previous sequencing errors in the 18S rRNA gene, thereby maintaining the monotypic status of this tick. Nuclear markers for the soft tick *Argas africolumbae* were also retrieved from next–generation data. Phylogenetic analysis of a concatenated 18S-28S rRNA dataset supported the grouping of *N. namaqua* at the base of the tick tree and the two main tick families in separate clades. This study confirmed the monotypic status of *N. namaqua* and its basal relationship to other tick families. Next-generation sequencing of genomic material to retrieve high quality nuclear and mitochondrial systematic markers for ticks is viable and may resolve issues around conventional sequencing errors when comparing closely related tick species.

Keywords: Genomics, Phylogenetics, Molecular systematics, Next-generation sequencing

Introduction

Molecular systematics of arthropods makes use of nuclear and mitochondrial genes as markers. Most common markers include the full mitochondrial genome, mitochondrial 16S ribosomal RNA, nuclear 18S and 28S ribosomal RNA and their ITS regions (Giribet and Edgecombe, 2012). These genes are generally abundant and amenable to samples with limited available biological material; they have no introns and are highly conserved allowing PCR amplification using universal primers. In most cases several independent PCR amplification steps need to be performed and several clones need to be sequenced, making targeting of independent genes cumbersome and increasing the probability for sequencing errors. For tick systematics these basic approaches have been extensively used (Black and Piesman, 1994; Black et al. 1997; Black and Roehrdanz, 1998; Klompen et al. 2000; Shao et al. 2004; Shao et al. 2005; Klompen et al. 2007). Next-generation sequencing has been used to sequence amplified nuclear and mitochondrial PCR products (Burger et al. 2012; Burger et al. 2013; Xiong et al. 2013; Burger et al. 2014). However, recently next generation sequencing of total genomic DNA without amplification of PCR products was used to assemble the mitochondrial genomes from the ticks *Nuttalliella namaqua* and *Argas africolumbae* (Mans et al. 2013). The possibility of using this approach to find other markers for tick systematics were investigated in the current study, focusing specifically on *N. namaqua*, since some biological questions regarding its monotypic status and its relationship to the other tick families exist that may be amenable to next generation sequencing.

Ticks constitute three families, Argasidae (soft ticks ~200 species), Ixodidae (hard ticks ~ 700 species) and the Nuttalliellidae (Guglielmone et al. 2010). The latter family comprise one species, *N. namaqua*, and has been considered to be the "missing link" between the hard and soft tick families as well as being a living fossil (Mans et al. 2011). Differences in the 18S rRNA gene were detected for geographically isolated *N. namaqua* populations (Horak et al. 2012). This is of interest, since it would negate the monotypic status of *N. namaqua* and could suggest that many undiscovered species of this family still exist, given its general host preference (skinks, geckos, girdled lizards, hyrax, meerkat, murid rodents and possibly birds), and its wide distribution from southern Africa (Namibia and South Africa) to Tanzania in East Africa (Keirans et al. 1976; Mans et al. 2014).

Analysis of the nuclear 18S rRNA and mitochondrial genes indicated that *N. namaqua* group basal to the Argasidae and Ixodidae (Mans et al. 2011; Mans et al. 2012; Gu et al. 2014; Chen et al. 2014). Other studies using concatenated 18S-28S rRNA gene sets suggested that *N. namaqua* does not group basal to the other tick families, but shows a closer affinity to the Argasidae (Burger et al. 2013), or that its relationship to the other families was unresolved (Burger et al. 2014), or that it shows a closer relationship to the Ixodidae using mitochondrial data (Burger et al. 2014). No 28S data were included for *N. namaqua* or Argasidae in these latter studies and raised the question whether its inclusion would affect the systematic analysis (Wiens and Morrill, 2011; Roure et al. 2013), since its placement at the base of the tick tree or as part of the main tick families affects conclusions regarding the ancestral tick lineage (Mans et al. 2012).

The current study shows that a next-generation sequencing approach is useful to retrieve the major nuclear and mitochondrial systematic markers from a single tick specimen. This includes full-length 28S rRNA sequences for *N. namaqua* and *A. africolumbae*, histone and transposable element sequences not previously reported. Differences in 18S rRNA sequences could be attributed to conventional sequencing errors in previous reported sequences, thereby maintaining the monotypic status of *N. namaqua*, while inclusion of the 28S rRNA sequences in a nuclear phylogenetic analysis support the basal position of *N. namaqua* in the tick tree. Host markers were also retrieved by next generation sequencing.

Materials and Methods

Next generation sequencing of tick genomic DNA

Genomic DNA from a single *N. namaqua* (40 ng) was submitted to the Biotechnology Platform Next Generation Sequencing Service of the Agricultural Research Council (South Africa). Samples were processed using the Nextera DNA sample preparation kit (Epicentre) and sequenced using the Illumina HiScanSQ (Illumina). Four different datasets of ~ 5 Gbp (~55 million reads) paired-end reads with lengths of 100 bp were analysed and these were unique datasets from that previously reported (Mans et al. 2012). Data were processed using the CLC Genomics Workbench v5.1 software package, imported using a range of 100-500 bp, quality trimmed (0.05 quality limit), Nextera adapters removed and the last 19 bp trimmed to give reads with an average length of 78 bp. Reads were *de novo* assembled using assembly parameters: mismatch cost-2, insertion cost-3, deletion cost-3, length fraction-0.9, similarity-0.9, minimum contig length-200, word size-23, bubble size-50. Data were filtered to give contigs with molecular sizes >1kb and average coverage >50. Contigs with multiple regions with no coverage (multiple N's due to paired read assembly) were discarded. Reciprocal BLAST analysis were used to determine best orthologous hits between the datasets and contigs were retained if they found hits in at least three of the databases. Consensus sequences were derived by multiple alignment of the reciprocal best BLAST hits using ClustalX alignment (Jeanmougin et al. 1998) and Genedoc (Nicholas et al. 1997). Consensus sequences were trimmed to only include consensus regions from all contigs and were analysed using BLASTN or BLASTX analysis (Altschul et al. 1990), and hits with E-values below E-5 were considered significant. To obtain final mapping statistics, data from the different assemblies were mapped to the consensus sequences using the same parameters previously used for assembly. Consensus sequences were deposited in Genbank (KF925832-KF925880). A dataset for *A. africolumbae* (Mans et al. 2012) were mined in a similar manner to obtain full-length 18S rRNA (JQ731646) and 28S rRNA genes (KF984488).

Estimating repetitive elements

Repetitive element content was estimated using RepeatMasker (Smit et al. 1996-2004), by analysing 700 000 reads (1.2%) from each dataset. The option "–species all" was used to screen all possible repeat structures present in the RepeatMasker databases (RepeatMasker and RepBase version 20120418).

Estimating genome size

A) Using the average coverage peak obtained for the different contigs (3.7 coverage), which resemble coverage of unique genes and dividing the total size of the reads used in each assembly by this number (Fig. 1). B) The combined datasets (~16.5 Gbp) were analysed in Kmergenie (Chikhi and Medvedev, 2014) using a variety of kmers (15, 20, 25, 30, 35, 40, 45) and assuming a diploid model, to estimate the unique haploid average coverage, which was used to estimate genome size by dividing the total length of the reads (16552689327 bp) by the unique average coverage (peak height ranging from 15-20). C) Mapping the combined datasets to 18737 open reading frames derived from a salivary gland transcriptome (manuscript in preparation) using CLC Genomics Workbench. The frequency distribution for the average coverage was plotted to determine the coverage depth with highest frequency (peak height=15). Genome size was then determined using the formula: Genome size = Genome reads x Read length/Highest frequency coverage depth (Hu et al. 2011).

Phylogenetic analysis

Acarine 28S and 18S rRNA sequences were retrieved from the non-redundant database by BLASTN analysis (Altschul et al. 1990). Only sequences from species with representatives of both sequences were used for downstream analysis. Sequences for the different datasets were aligned separately using ClustalX (Jeanmougin et al. 1998). Alignments were manually inspected, adjusted and trimmed and then concatenated to yield the final super-matrix used for phylogenetic analysis.

Bayesian analysis was performed using MrBayes 3.1.2 (Ronquist et al. 2003). A general time reversible (GTR) model of nucleotide substitution with a proportion of invariant sites and a gamma distribution of among site heterogeneity using the nst = 6 rates = ingamma command was used. Four categories were used to approximate the gamma distribution and two runs were performed simultaneously, each with four Markov chains (one cold, three heated) which ran for 5,000,000 generations. The first 2,000,000 generations were discarded from the analysis (burnin) and every 100th tree was sampled to calculate a 50% majority-rule consensus tree. Nodal values represent the posterior probability that the recovered clades exist, given the sequence dataset and were considered significant above 95% (Alfaro et al. 2003).

Maximum-likelihood analysis was performed in Mega 5 (Tamura et al. 2011). The Kimura 2-parameter substitution model was used with a Gamma distribution of 5 discrete Gamma categories. The initial tree was generated automatically using neighbor-joining and trees were searched using Nearest-Neighbor-Interchange. All sites were used (1985) and nodal support was estimated using 1000 bootstrap replications. Bootstrap support above 50% was considered significant for the current analysis (Hillis and Bull, 1993).

Results

Analysis of data generated

Four different datasets were generated (~5 Gbp each) and assembled, each with ~55 million independently generated reads with average length of 78 bp after quality and adapter trimming (Table 1). Greater than 90% of all the reads had PHRED scores above 30 (99.9% accuracy) with a median quality distribution of 36 over the total read length. The %GC content for the different datasets was ~45%. Less than 0.6% coverage was observed for enrichment of 5mers across the total read length, indicating that reads were not enriched in low complexity sequences.

Assembly of next generation sequencing data

Approximately 20% of reads could be assembled from each dataset resulting in ~350-390 thousand contigs/assembly. Contigs ranged in length from ~20-28000 bp with N50 values of ~350 bp. All contigs above 1000 bp (9183 contigs) were analysed using BLASTX analysis of the non-redundant protein database. Of these, 3431 gave significant hits (<E-5) and 83% of the hits were for arthropod proteins, with 63% for ticks as best hit (results not shown). The majority of hits are for proteins with general house-keeping functions. This suggests that the majority of assembled contigs code for the tick genome. Coverage ranged from 0.6-7500, while ~89% of all contigs (coverage 2-6) had an average coverage of 3.72 ± 0.95 with a distinct peak for the coverage frequency distribution that represent the coverage for non-repetitive unique regions in the genome (Fig. 1).

Estimating genome size

Since no biological material was available to perform biophysical estimates of the genome size, it was estimated using three different bioinformatics approaches that all gave estimates ranging from 0.92-1.1 Gbp. The first used the unique coverage peak observed in the assembled data (Fig. 1) and estimated the genome at 1.1±0.23 Gbp. The second used the program Kmergenie which yielded an estimated genome size of 0.955±0.103 Gbp. The third determined average coverage by mapping data to an independent salivary gland transcriptome and yielded a genome size of 1.103 Gbp.

Repetitive regions in genome

Contigs with high coverage can be detected based on the slope of the curves with high coverage regions indicated by the exponential rise in the slope (Fig. 1). Reads above 10 fold coverage (repetitive or duplicated regions) comprised ~2.6% of all contigs (Table 1), but 33-39% of the assembled reads, suggesting that a large portion of the sequenced genome might be composed of repetitive regions. However, if the genome is ~1 Gbp in size and the coverage sequenced was ~4, it might not be surprising that the majority of assembled reads were for repetitive sequences, since only these would have sufficient coverage to enable assembly. In this regard, the repetitive sequences only comprised ~6.9% of all reads (Table 1). No contig comprised more than 0.14% of all reads and the dominant contigs coded for protein coding regions or systematic markers which comprised ~1.4% of all reads (Table 1; Table 2). To investigate the makeup of repetitive sequences, ~1% of the data from each dataset was analysed using RepeatMasker (Fig. 2). This indicated that ~4% of the data comprised interspersed elements that included SINEs, LINEs, LTR elements and transposons, while ~1.5% included small RNA, simple repeats, regions of low complexity and satellites.

Screening for markers

Large (>1000 bp) and abundant contigs (>50 fold coverage) from the initial assemblies were selected to identify systematic markers, since these contigs should be the most consistent in regard to reproducibility and assembly quality. The filtered datasets (contigs >1000bp, average coverage >50-fold) gave ~130 contigs (<0.05% of total contigs) for each assembly with N50 values of ~3800 bp. Of these, 69 were represented in at least three of the databases and consensus sequences were derived for these. Fifty of the consensus sequences gave significant hits using BLASTN and BLASTX analysis (Table 2). Hits were found for vertebrate, arthropod or tick homologs and will be discussed in the sections below.

Vertebrate homologs: Potential host-derived genomic material

Twenty of the contigs found hits to vertebrate sequences that included mainly transposable elements (retroviral, transposons, transposases and retrotransposons) common in vertebrate genomes. A number of these found as best BLAST hit sequences from reptile origin (*Phyton*, *Anolis*, *Vipera*). Notable, is that the highest coverage obtained was for two potential host-derived transposable elements, a transposase and a BovB long interspersed element (LINE). BLASTN hits to ribosomal sequences for the girdled lizard genus *Cordylus* (18S rRNA) and the turtle genus *Chrysemys* (28S rRNA) were found (Table 2). A previous gut meal analysis of the 16S rRNA gene for reptiles found DNA for at least 4 different members of the *Cordylus* family in this tick specimen (Mans et al. 2011). It can therefore be assumed that the 18S rRNA sequence detected, codes for one of the *Cordylus* species, while the 28S rRNA represent the first sequence for this genus in the database. The contig for the *Cordylus* 18S rRNA spans the full-length 18S rRNA gene (1744 bp). Mapping of the original reads to the consensus sequence obtained mappings for all four datasets with average coverage of 314-329 that spanned the whole sequence (Fig. 3). The consensus sequence for the 28S rRNA gene spanned 2361 bp and mapping indicated an average coverage of 255-269 that spanned the whole sequence (Fig. 3).

Arthopod homologs: Potential tick-derived genomic material

Nineteen contigs found hits to non-tick arthropod homologs that included mostly mobile elements (transposons, transposases and retrotransposons) (Table 2; Supplementary material S1). In addition, seven hits retrieved tick proteins annotated as transposons, transposases or retrotransposons. Since all of these hits are for arthropod-derived genes, they probably derive from the tick genome.

Tick-derived systematic markers

Except for mobile elements, four notable hits associated with molecular systematics were found and included the nuclear histone genes, 18S rRNA, 28S rRNA and the mitochondrial genome (Table 2; Supplementary material). These contigs were present in all assemblies and mapping of reads back to the consensus sequences indicated high and consistent coverage across the full-length of the genes as well as the mitochondrial genome (Fig. 4). In the case of the histone genes, the H3-H2A-H2B cassette was assembled into a single contig that yielded the full-length translated sequences for all three genes, since these do not possess introns. BLASTP analysis of the protein sequences retrieved tick proteins with >95% identity and 98% similarity.

Comparison of 18S rRNA sequences to those in the databank

In each assembly a full-length contig that codes for the 18S rRNA gene was obtained that spanned a length of 1974 bp (Fig. 4). Each contig had an average excess of 1000-fold coverage for each nucleotide position and when reads were mapped back onto the 18S gene alone, coverage increased above an average of ~1100-fold. Coverage ranged from a minimum to maximum of 16-3233 and the low coverage regions corresponded with the ends of the contigs that does not fall within the full-length 18S gene (Fig. 4). Of interest is the similarity in coverage obtained for the different assemblies and is probably due to the fact that the same fragmented DNA sample were used for the generation of each dataset and that each dataset gives a representative coverage of all generated fragments. Comparison of the 18S rRNA gene generated by next-generation sequencing with the sequences in Genbank, indicate that the differences previously observed between the two submitted *N. namaqua* sequences (Horak et al. 2012) were due to sequencing errors, specifically that the sequence from the current study is identical to those reported for larvae collected from rodents (Fig. 5).

Next-generation sequencing for systematic analysis

The power of next-generation sequencing to resolve sequencing mistakes by conventional approaches is apparent in the current study and should have a significant impact on systematic analysis of closely related species. It was therefore of interest to determine whether the mitochondrial genome data previously generated for *N. namaqua* were still valid (Mans et al. 2012). The previous approach assembled ~22 million single end reads and obtained two contigs that coded for the whole mitochondrial genome with an average coverage of ~37-fold. Using the current datasets that used paired-end data, the mitochondrial genome of *N. namaqua* was assembled in a single contig for each assembly with an average coverage of ~87-fold (Fig. 4). Assemblies 1-3 gave full-length mitochondrial genomes, while Assembly 4 lacked 145 bp, but could be obtained by a mapping approach (Fig. 4). The sequences were 100% identical to the sequence previously published (Mans et al. 2012). This indicated that Illumina next-generation sequencing is a viable method to obtain high quality markers.

The position of N. namaqua *at the base of the tick tree*

A concatenated dataset of the 18S-28S nuclear rRNA genes were constructed using Ixodidae and outgroups for which all sequences are available and alignment ends were trimmed before concatenation to include only homologous sites. Since 18S and 28S rRNA data is available for a limited number of soft tick species, the next generation dataset previously generated for *A. africolumbae* (Mans et al. 2012), were mined for 18S and 28S rRNA data and both full-length genes could be obtained from the original assembly (Fig. 6). A 50% consensus tree using Maximum Likelihood analysis indicated that *N. namaqua* grouped at the base of the tick families with 65% support (Fig. 7). The split between the two major families (Argasidae and Ixodidae) were retrieved in the topology, albeit with low support. Using Bayesian analysis, higher than 95% posterior probability was obtained for both nodes, while the majority of other nodes were also supported.

Discussion

Previously the mitochondrial genomes of *N. namaqua* (Nuttalliellidae) and *A. africolumbae* (Argasidae) was determined using next-generation sequencing and comprised the first mitochondrial genomes reported for the Nuttalliella and the Argasinae either tick family (Mans et al. 2012). This latter study concerned the comparative and phylogenetic analysis of mitochondrial data and did not investigate the use of next-generation sequencing to retrieve nuclear systematic markers. The current study attempted a methodical analyses of next-generation sequencing with the aim of determining which contigs are most abundantly represented in such a dataset, with the hypothesis that highly repetitive regions of the genome such as the nuclear ribosomal cistron would be well represented, and that such an approach would be amenable to retrieve nuclear markers such as the 18S and 28S rRNA genes with high coverage. Full-length sequences for systematic markers such as the nuclear 18S and 28S rRNA as well as the full mitochondrial genome were retrieved in the current study. In addition, the histone cassette were also retrieved. To date, few tick histones have been described or used as systematic markers, but may be considered in the future (Edgecombe et al. 2000; Giribet et al. 2001).

Three independent methods estimated the genome size at ~1 Gbp. This is comparable to argasid (1-1.5 Gbp) and Prostriate (2 Gbp) genome sizes (Geraci et al. 2007). In general, there is a trend towards larger genome sizes for more recently derived ixodid genera (Geraci et al. 2007). As such, argasids have smaller sizes than Prostriates, which is smaller than basal metastriates, which is smaller than *Rhipicephalus* (Geraci et al. 2007). The estimated genome size for *N. namaqua* places it closer to argasids than the Prostriata or Metastriata and fits with its proposed position near the base of the phylogenetic tree (Mans et al. 2011; Mans et al. 2012). With regard to next-generation sequencing as universal method to retrieve systematic markers from ticks, genome size may affect the data size required for assembly. As such, smaller datasets such as the 5 Gb generated in this study may be amenable to retrieve systematic markers for argasids and prostriates. For metastriates that have genomes ~3-8X in size (Geraci et al. 2007), larger datasets may need to be generated, although the repetitive nature of these latter genomes may compensate for the nuclear ribosomal markers.

The genome of *N. namaqua* is enriched in transposable elements and this has been observed for other tick genomes (Ullman et al. 2005; Francischetti et al. 2008; Chmelar et al. 2008; Guerrero et al. 2010). Given the ancient proposed origin of *Nuttalliella* in the Late Carboniferous or Early Permian (Mans et al. 2011; Mans et al. 2012), its generalist feeding behaviour on reptiles and mammals (Mans et al. 2014), the ability to maintain nucleated red blood cells intact over prolonged period between multiple feeding events (Mans et al. 2012), with the possibility of blood meal regurgitation, it is a likely vector of vertebrate transposable elements (Walsh et al. 2013). Given that transposable elements are abundant in vertebrate genomes, *N. namaqua* and argasids (which share a similar feeding biology) could have contributed to vertebrate genome evolution since the Permian. In this regard, the BovB LINE gene obtained in this study was proposed to have been widely disseminated among vertebrates via horizontal transfer by tick vectors (Walsh et al. 2013). The high coverage of these elements in the present study may be due to their high abundance in vertebrate genomes and tick hosts (Walsh et al. 2013). The possibility exists that transposable elements might have been the first "infectious" agent transmitted by ticks. Reciprocally, invasion of the tick genome by similar elements could have contributed towards their increased genome size compared to other Acari (Geraci et al. 2007). In this regard, the repetitive content of the genome of *N. namaqua* is lower than genomes from other ticks, where significant proportions of the genome were repetitive, with 66% for *I. scapularis* and 69% for *R. microplus*, (Ullmann et al. 2005), with tandem repeat families comprising a significant part of the genomes of *I. scapularis* (~8%) and *R. microplus* (Hill et al. 2009; Meyer et al. 2010). In contrast, contigs obtained in the present study did not exhibit such high coverage, with the overall repetitive content estimated at ~7%. As such, a lower proportion of the genome of *N. namaqua* may be composed of repetitive regions, making it a potential candidate for genome sequencing as model tick species.

Based on a four times coverage of the genome sequenced, the coverage obtained for the 18S rRNA, 28S rRNA genes and histone cassette suggest that there is ~230 copies for the nuclear rRNA genes and ~180 copies of the histone cassette in the genome of *N. namaqua*. These values compare well with estimates for other arthropods, where ~100-240 copies were estimated for *Drosophila* and 39-1024 copies in mosquitoes for the rRNA cassette (Kumar and Rai, 1990; Lyckegaard and Clark, 1991). The histone cassette also shows high copy numbers depending on the metabolic activity of the species in question. As such, sea urchins possess 300-800 copies and *Drosophila* 100 tandem repeated copies (Fitch et al. 1990). Most transposable elements have coverage ranging from 12-200 (average 45 copies), indicating that while the genome is enriched in these elements, they are not more abundant than systematic markers.

The wide geographic distribution of *N. namaqua* and the proposed ancient origin of this family (Mans et al. 2011; Mans et al. 2012), raised the possibility that it is not monotypic (Horak et al. 2012). The differences previously observed (476/478) in the 18S rRNA gene between ticks collected from the field and from rodents corresponded to the 1093-1571 region of the published sequence and would therefore give support to this notion (Horak et al. 2012). However, BLAST analysis of the 1093-1571 bp region indicated that the 18S sequence for *N. namaqua* (GI accession number: 333109165) had 3 gaps in this region, which seemed peculiar since no other tick sequence including the sequence from *N. aff. namaqua* (GI accession number: 398256856) had any gaps in this region of the 18S rRNA gene. This indicated possible sequence errors in 333109165. This was confirmed in the sequence derived by next-generation sequencing from the same DNA sample as 333109165, that is 100% identical to that of 398256856. There is therefore no difference at the 18S rRNA level that will support the recognition of more than one *Nuttalliella* species and the monotypic status of this family remains unchanged. Many of the members of the Rhipicephalinae have highly similar 18S rRNA sequences with too few differences to be phylogenetically informative (Black et al. 1997), which indicate their relatively recent divergence. Identical 18S rRNA sequences as such does not imply that the monotypic status of *N. namaqua* could not change. However, it has been indicated that the larval morphology of *N. namaqua* from Namaqualand and the Soutpansberg area are identical (I. Horak and D. Apanaskevich, personal communication). No morphological data therefore exist that would support different species. In this regard, specimens collected from Heuningvlei pan (~650 km from Springbok, ~730 km from Soutpansberg) are the same species as those collected at Springbok using 16S rRNA gene analysis (results not shown). A much larger geographical analysis of *N. namaqua* may reveal the presence of cryptic species. However, this future study may be hampered by the secretive nature of this genus (Mans et al. 2011).

A recent study analysed a concatenated 18S-28S rRNA gene set in which the *Nuttalliella* did not group at the base of the tick tree, but with the Argasidae, although with low bootstrap support (Burger et al. 2013). Another study placed the *Nuttalliella*, Ixodidae and Argasidae on an unresolved branch (Burger et al. 2014). Both studies included a variety of sequences for which no 28S rRNA data exist and this could have potentially biased the analysis (Wiens and Morrill, 2011; Roure et al. 2013). The current study included the 28S rRNA sequence for *N. namaqua* and it grouped at the base of the tick tree using both Maximum Likelihood and Bayesian analysis. These results correlate with those from previous Bayesian analysis using nuclear markers (Mans et al. 2011; Mans et al. 2012). Even so, most studies of tick phylogeny to date may be biased due to incomplete taxon sampling (Wiens and Morrill, 2011), which will hopefully be resolved in future as more species are added to the database. However, loss of taxa and lineages due to mass extinctions (Mans et al. 2011; Mans et al. 2014), may forever limit our ability to resolve deeper phylogenies for ticks, since at deep phylogeny ticks can only be divided into three families and 5 basal lineages (Nuttalliella; Argasinae; Ornithodorinae; Prostriates and Metastriates). Previously, analysis of the mitochondrial genome using Bayesian analysis of five protein genes indicated that *N. namaqua* grouped basal to the other tick families (Mans et al. 2012). Maximum likelihood analysis using total mitochondrial nucleic acid data observed the same grouping (Gu et al. 2014; Chen et al. 2014). Maximum likelihood analysis using a different mitochondrial protein gene set indicated that *N. namaqua* grouped with the Ixodidae, while the Holothryida as sistergroup to the ticks were replaced by the Mesostigmata (Burger et al. 2014). Maximum likelihood analysis using individual mitochondrial protein genes indicated that there is no congruence between the gene trees (BM, personal observation), suggesting that mitochondrial genes might not be appropriate for resolving deep phylogenies (Dávalos et al. 2012). It is also well recognized that arthropod phylogenies between nuclear and mitochondrial markers do not necessarily correlate and the use of nuclear markers over mitochondrial markers should be considered when resolving deep phylogenies (Giribet and Edgecombe, 2012), possibly above genus level in the case of ticks.

For the current study the generation of a single dataset of ~5Gb cost ~$300. To amplify and sequence the mitochondrial genome (~14 kb), the 18S (~1.7 kb) and 28S (~3.5 kb) ribosomal genes using conventional approaches from both directions at 3X coverage would cost ~$640 making the cost of next-generation sequencing comparable. Given the trend in price reduction for data generation using next-generation sequencing and the general acceptability of high-throughput strategies for systematics and phylogenetics (Lemmon and Lemmon, 2013), it is likely that this approach will become feasible at population level in the near future. The next-generation sequencing approach may also be expanded to specimens preserved in alcohol up to two years if adequate concentrations of DNA (>5 ng) can be retrieved from them (BM, personal observation).

Conclusions

Systematic markers as well as mobile elements are abundant in the genome of *N. namaqua*. In addition, host related information could be mined which indicated a similar trend, i.e. enrichment of the mammalian genomes in nuclear systematic markers and mobile elements. Next-generation sequencing using Illumina HiScan technology retrieved both nuclear and mitochondrial systematic markers in a reproducible manner, which shows that this approach would be amenable for the recovery of markers used in tick systematics.

Acknowledgements

This project was funded by the Joy Liebenberg Trust (21/19/JT02) allocated to BM, incentive funding for rated researchers grant from the National Research Foundation of South Africa (NRF-Mans 76499), and the ARC Tick Vaccine project (30/01/V010).

References

Alfaro, M.E., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20, 255-266.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Black, W.C.4th., Piesman, J., 1994. Phylogeny of hard- and soft-tick taxa (Acari: Ixodida) based on mitochondrial 16S rDNA sequences. Proc. Natl. Acad. Sci. USA 91, 10034-10038.

Black, W.C.4th., Klompen, J.S.H., Keirans, J.E., 1997. Phylogenetic relationships among tick subfamilies (Ixodida: Ixodidae: Argasidae) based on the 18S nuclear rDNA gene. Mol. Phylogenet. Evol. 7, 129–144.

Black, W.C.4th., Roehrdanz, R.L., 1998. Mitochondrial gene order is not conserved in arthropods: prostriate and metastriate tick mitochondrial genomes. Mol. Biol. Evol. 15, 1772-1785.

Burger, T.D., Shao, R., Beati, L., Miller, H., Barker, S.C., 2012. Phylogenetic analysis of ticks (Acari: Ixodida) using mitochondrial genomes and nuclear rRNA genes indicates that the genus *Amblyomma* is polyphyletic. Mol. Phyl. Evol. 64, 45-55.

Burger, T.D., Shao, R., Barker, S.C., 2013. Phylogenetic analysis of the mitochondrial genomes and nuclear rRNA genes of ticks reveals a deep phylogenetic structure within the genus Haemaphysalis and further elucidates the polyphyly of the genus *Amblyomma* with respect to *Amblyomma sphenodonti* and *Amblyomma elaphense*. Ticks Tick Borne Dis. 4, 265-274.

Burger, T.D., Shao, R., Labruna, M.B., Barker, S.C., 2014. Molecular phylogeny of soft ticks (Ixodida: Argasidae) inferred from mitochondrial genome and nuclear rRNA sequences. Ticks Tick Borne Dis. 5, 195-207.

Chen, D.S., Jin, P.Y., Zhang, K.J., Ding, X.L., Yang, S.X., Ju, J.F., Zhao, J.Y., Hong, X.Y., 2014. The complete mitochondrial genomes of six species of *Tetranychus* provide insights into the phylogeny and evolution of spider mites. PLoS One 9, e110625.

Chikhi, R., Medvedev, P., 2014. Informed and automated k-mer size selection for genome assembly. Bioinformatics 30, 31-37.

Chmelar, J., Anderson, J.M., Mu, J., Jochim, R.C., Valenzuela, J.G., Kopecký, J., 2008. Insight into the sialome of the castor bean tick, *Ixodes ricinus*. BMC Genomics 9, 233.

Dávalos, L.M., Cirranello, A.L., Geisler, J.H., Simmons, N.B., 2012. Understanding phylogenetic incongruence: Lessons from phyllostomid bats. Biological Reviews 2012 87: 991-1024.

Edgecombe, G.D., Wilson, G.D.F., Colgan, D.J., Gray, M.R., Cassis, G., 2000. Arthropod cladistics: Combined analysis of histone H3 and U2 snRNA sequences and morphology. Cladistics 16, 155-203.

Fitch, D.H., Strausbaugh, L.D., Barrett, V., 1990. On the origins of tandemly repeated genes: does histone gene copy number in *Drosophila* reflect chromosomal location? Chromosoma 1990 99, 118-124.

Francischetti, I.M., Meng, Z., Mans, B.J., Gudderra, N., Hall, M., Veenstra, T.D., Pham, V.M., Kotsyfakis, M., Ribeiro, J.M., 2008. An insight into the salivary transcriptome and proteome of the soft tick and vector of epizootic bovine abortion, *Ornithodoros coriaceus*. J. Proteomics 71, 493-512.

Geraci, N.S., Johnston, J.S., Robinson, J.P., Wikel, S.K., Hill, C.A. (2007) Variation in genome size of argasid and ixodid ticks. Insect Biochem. Mol. Biol. 37, 399-408.

Giribet, G., Edgecombe, G.D., Wheeler, W.C., 2001. Arthropod phylogeny based on eight molecular loci and morphology. Nature 413, 157-161.

Giribet, G., Edgecombe, G.D., 2012. Reevaluating the arthropod tree of life. Ann. Rev. Entomol. 57, 167-186.

Gu, X.B., Liu, G.H., Song, H.Q., Liu, T.Y., Yang, G.Y., Zhu, X.Q., 2014. The complete mitochondrial genome of the scab mite *Psoroptes cuniculi*

(Arthropoda: Arachnida) provides insights into Acari phylogeny. Parasit. Vectors 7, 340.

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.

Guglielmone, A.A., Robbins, R.G., Apanaskevich, D.A., Petney, T.N., Estrada-Pena, A., Horak, I.G., Shao, R., Barker, S.C., 2010. The Argasidae, Ixodidae and Nuttalliellidae (Acari: Ixodida) of the world: a list of valid species names. Zootaxa 2528, 1–28.

Hill, C.A., Guerrero, F.D., Van Zee, J.P., Geraci, N.S., Walling, J.G., Stuart, J.J., 2009. The position of repetitive DNA sequence in the southern cattle tick genome permits chromosome identification. Chromosome Res. 17, 77-89.

Hillis, D.M., Bull, J.J., 1993. An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. Syst. Biol. 42, 182-192.

Horak, I.G., Lutermann, H., Medger, K., Apanaskevich, D.A., Matthee, C.A., (2012) Natural hosts of the larvae of *Nuttalliella* sp. (*N. namaqua*?) (Acari: Nuttalliellidae). Onderstepoort J. Vet. Res. 79, 405.

Hu, H., Bandyopadhyay, P.K., Olivera, B.M., Yandell, M. (2011) Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. BMC Genom. 12, 60.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., Gibson, T.J., 1998. Multiple sequence alignment with Clustal X. Trends Biochem. Sci. 23, 403–405.

Keirans, J.E., Clifford, C.M., Hoogstraal, H., Easton, E.R. 1976. Discovery of *Nuttalliella namaqua* Bedford (Acarina: Ixodoidea: Nuttalliellidae) in Tanzania and redescription of the female based on scanning electron microscopy. Ann. Entomol. Soc. Am. 69, 926–932.

Klompen, J.S.H., Black, W.C.4th., Keirans, J.E., Norris, D.E., 2000. Systematics and biogeography of hard ticks, a total evidence approach. Cladistics 16, 79–102.

Klompen, H., Lekveishvili, M., Black, W.C.4th., 2007. Phylogeny of parasitiform mites (Acari) based on rRNA. Mol. Phylogenet. Evol. 43, 936-951.

Kumar, A., Rai, K.S., 1990. Chromosomal localization and copy number of 18s + 28s ribosomal RNA genes in evolutionarily diverse mosquitoes (Diptera, Culicidae). Hereditas 113, 277-289.

Lemmon, E.M., Lemmon, A.R., 2013. High-Throughput Genomic Data in Systematics and Phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44, 99–121.

Lyckegaard, E.M., Clark, A.G., 1991. Evolution of ribosomal RNA gene copy number on the sex chromosomes of *Drosophila melanogaster*. Mol. Biol. Evol. 8, 458-474.

Mans, B.J., de Klerk, D., Pienaar, R., Latif, A.A., 2011. *Nuttalliella namaqua*: a living fossil and closest relative to the ancestral tick lineage: implications for the evolution of blood-feeding in ticks. PLoS One 6, e23675.

Mans, B.J., de Klerk, D., Pienaar, R., de Castro, M.H., Latif, A.A., 2012. The mitochondrial genomes of *Nuttalliella namaqua* (Ixodoidea: Nuttalliellidae) and *Argas africolumbae* (Ixodoidae: Argasidae): estimation of divergence dates for the major tick lineages and reconstruction of ancestral blood-feeding characters. PLoS One 7, e49461.

Mans, B.J., de Klerk, D.G., Pienaar, R., Latif, A.A., 2014. The host preferences of *Nuttalliella namaqua* (Ixodoidea: Nuttalliellidae): a generalist approach to surviving multiple host-switches. Exp. Appl. Acarol. 62, 233-240.

Meyer, J.M., Kurtti, T.J., Van Zee, J.P., Hill, C.A., 2010. Genome organization of major tandem repeats in the hard tick, *Ixodes scapularis*. Chromosome Res. 18, 357-370.

Nicholas, K.B., Nicholas, H.B., Jr., Deerfield, D.W., 1997. GeneDoc: Analysis and Visualization of Genetic Variation. EMBNEW.NEWS 4, 14.

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

Roure, B., Baurain, D., Philippe, H., 2013. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. Mol. Biol. Evol. 30, 197-214.

Shao, R., Aoki, Y., Mitani, H., Tabuchi, N., Barker, S.C., Fukunaga, M., 2004. The mitochondrial genomes of soft ticks have an arrangement of genes that has remained unchanged for over 400 million years. Insect Mol. Biol. 13, 219-224.

Shao, R., Barker, S.C., Mitani, H., Aoki, Y., Fukunaga, M., 2005. Evolution of duplicate control regions in the mitochondrial genomes of metazoa: a case study with Australasian *Ixodes* ticks. Mol. Biol. Evol. 22, 620-629.

Smit, A., Hubley, R., Green, P., 1996-2004. RepeatMasker Open-3.0.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731-2739.

Ullmann, A.J., Lima, C.M., Guerrero, F.D., Piesman, J., Black, W.C. 4th., 2005. Genome size and organization in the blacklegged tick, *Ixodes scapularis* and the Southern cattle tick, *Boophilus microplus*. Insect Mol. Biol. 14, 217-222.

Walsh, A.M., Kortschak, R.D., Gardner, M.G., Bertozzi, T., Adelson, D.L., 2013. Widespread horizontal transfer of retrotransposons. Proc. Natl. Acad. Sci. U. S. A. 110, 1012-1016.

Wiens, J.J., Morrill, M.C., 2011. Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. Syst. Biol. 60, 719-731.

Xiong, H., Barker, S.C., Burger, T.D., Raoult, D., Shao, R., 2013. Heteroplasmy in the mitochondrial genomes of human lice and ticks revealed by high throughput sequencing. PLoS One 8, e73329.

Figure legends

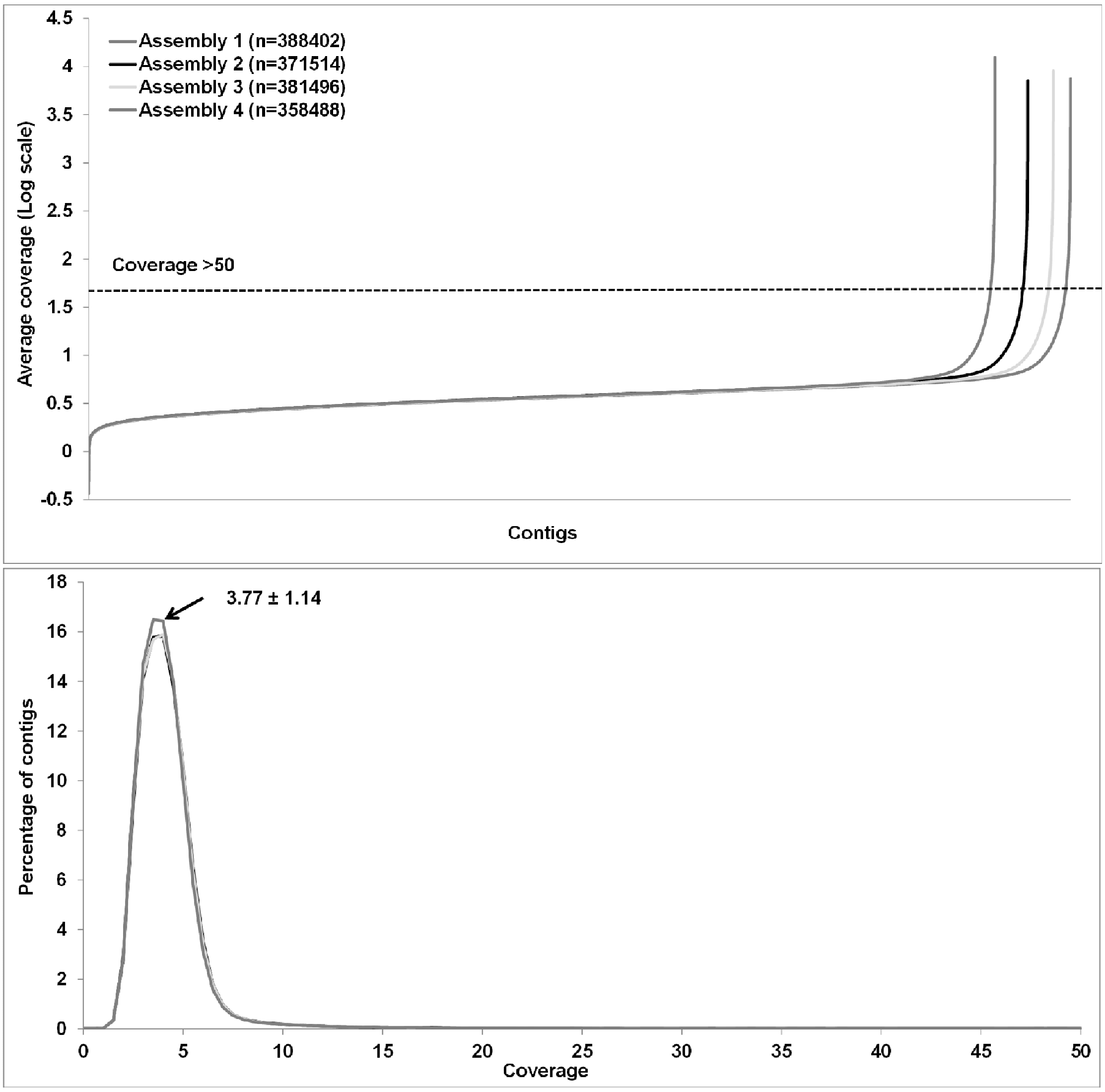


Figure 1: Summary of Nextera read sequence assemblies. The upper graph show average coverage of all contigs obtained from assemblies 1-4. Number of contigs for each assembly is indicated. Data above the upper dotted line represent those contigs with coverage above 50. The lower graph shows a frequency distribution plot of the coverage for the assemblies. The arrow indicates the average coverage (± SD) for contigs with coverage below 10.

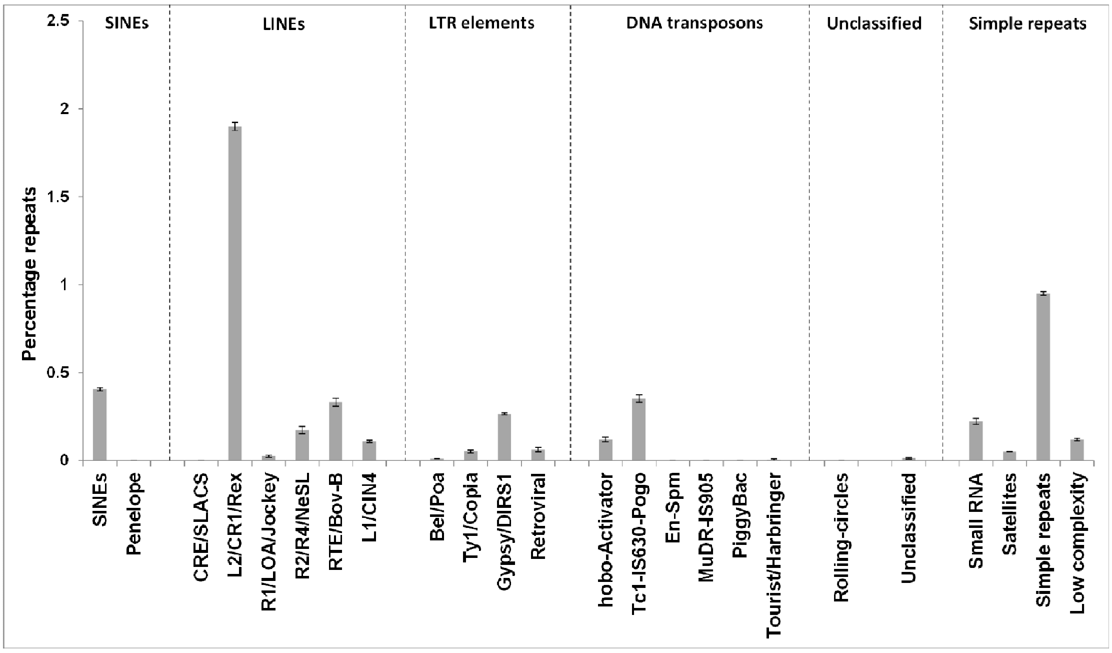


Figure 2: Analysis of repeat elements using RepeatMasker. Randomly selected reads (700 000 reads / dataset) were analysed and the average percentage of reads (± standard deviation) for various repetitive elements are indicated.

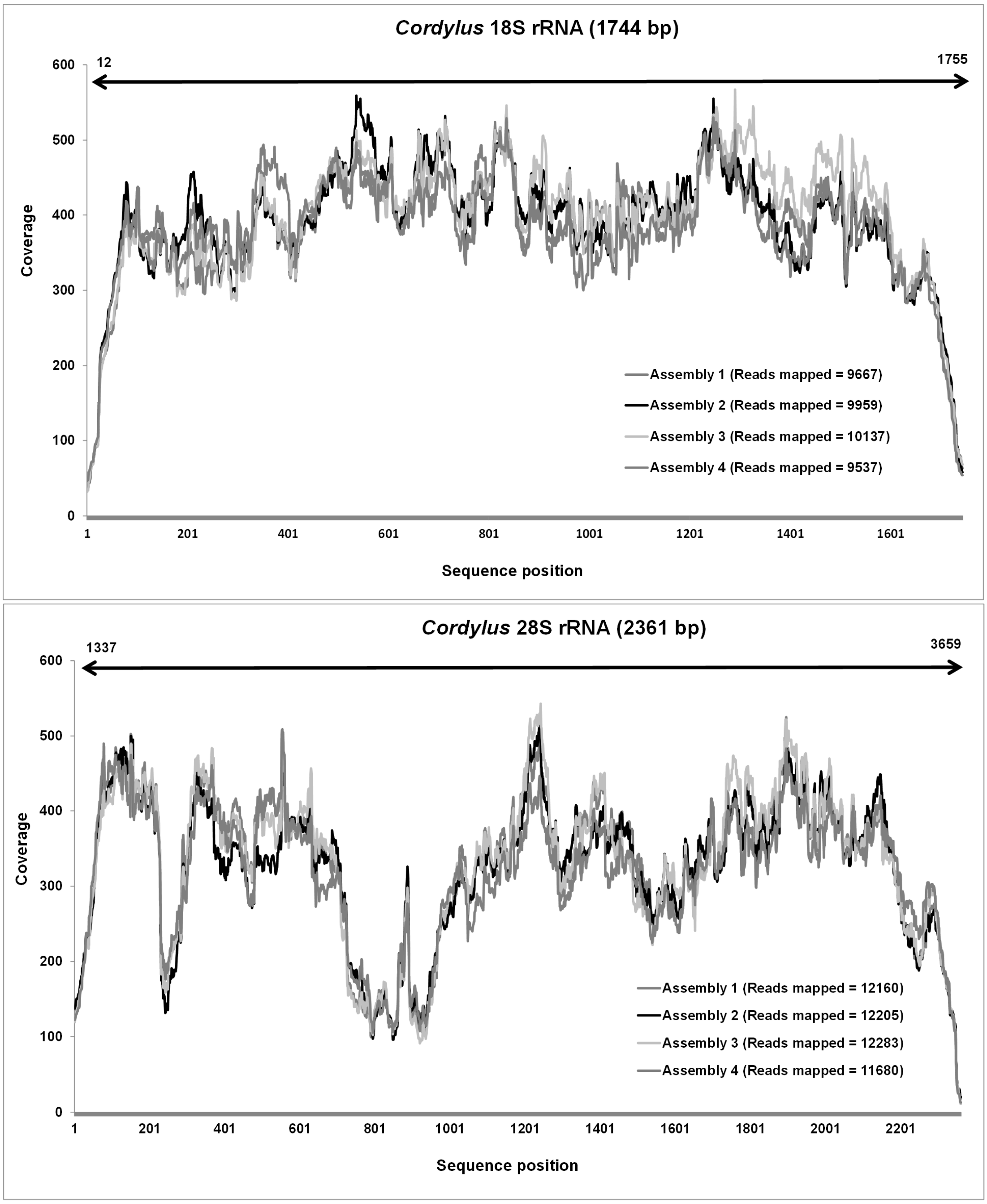


Figure 3: Assembly of the host-derived 18S rRNA and 28S rRNA genes. Indicated are read mapping coverages for the 18S rRNA and 28S rRNA gene assemblies for the various datasets. The number of reads mapped for each dataset is also indicated. The black arrow indicates the conventional full-length gene region.

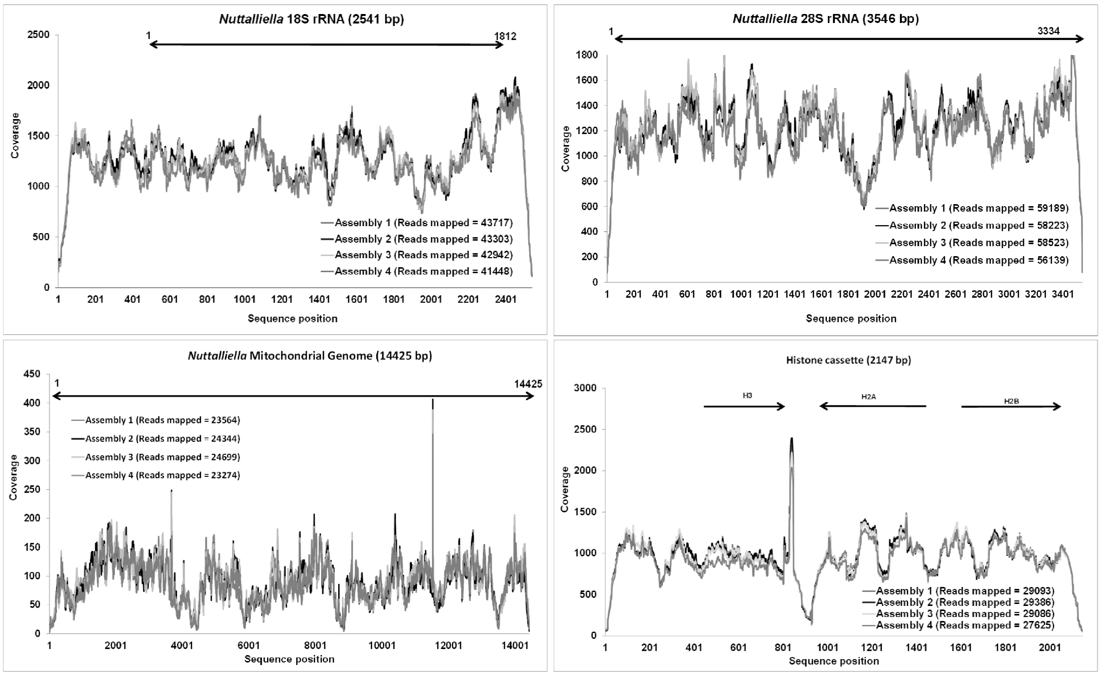
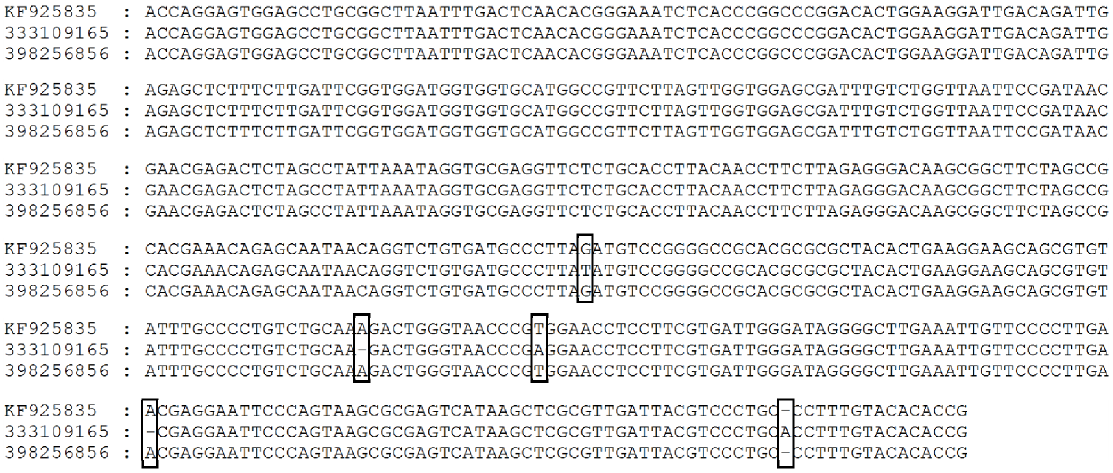


Figure 4: Assembly of tick-derived systematic markers. Indicated are read mapping coverages for the 18S rRNA, 28S rRNA, mitochondrial genome and histone cassette gene assemblies for the various datasets. The number of reads mapped for each dataset is also indicated. The black arrow indicates the conventional full-length gene region, and in the case of the histone cassette, the ORF direction.



```
KF925835   : ACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAATCTCACCCGGCCCGGACACTGGAAGGATTGACAGATTG
333109165 : ACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAATCTCACCCGGCCCGGACACTGGAAGGATTGACAGATTG
398256856 : ACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAATCTCACCCGGCCCGGACACTGGAAGGATTGACAGATTG

KF925835   : AGAGCTCTTTCTTGATTCGGTGGATGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGATAAC
333109165 : AGAGCTCTTTCTTGATTCGGTGGATGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGATAAC
398256856 : AGAGCTCTTTCTTGATTCGGTGGATGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGATAAC

KF925835   : GAACGAGACTCTAGCCTATTAAATAGGTGCGAGGTTCTCTGCACCTTACAACCTTCTTAGAGGGACAAGCGGCTTCTAGCCG
333109165 : GAACGAGACTCTAGCCTATTAAATAGGTGCGAGGTTCTCTGCACCTTACAACCTTCTTAGAGGGACAAGCGGCTTCTAGCCG
398256856 : GAACGAGACTCTAGCCTATTAAATAGGTGCGAGGTTCTCTGCACCTTACAACCTTCTTAGAGGGACAAGCGGCTTCTAGCCG

KF925835   : CACGAAACAGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGGGGCCGCACGCGCGCTACACTGAAGGAAGCAGCGTGT
333109165 : CACGAAACAGAGCAATAACAGGTCTGTGATGCCCTTATATGTCCGGGGCCGCACGCGCGCTACACTGAAGGAAGCAGCGTGT
398256856 : CACGAAACAGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGGGGCCGCACGCGCGCTACACTGAAGGAAGCAGCGTGT

KF925835   : ATTTGCCCCTGTCTGCAAAGACTGGGTAACCCGTGGAACCTCCTTCGTGATTGGGATAGGGGCTTGAAATTGTTCCCCTTGA
333109165 : ATTTGCCCCTGTCTGCAA-GACTGGGTAACCCGAGGAACCTCCTTCGTGATTGGGATAGGGGCTTGAAATTGTTCCCCTTGA
398256856 : ATTTGCCCCTGTCTGCAAAGACTGGGTAACCCGTGGAACCTCCTTCGTGATTGGGATAGGGGCTTGAAATTGTTCCCCTTGA

KF925835   : ACGAGGAATTCCCAGTAAGCGCGAGTCATAAGCTCGCGTTGATTACGTCCCTGC-CCTTTGTACACACCG
333109165 : -CGAGGAATTCCCAGTAAGCGCGAGTCATAAGCTCGCGTTGATTACGTCCCTGCACCTTTGTACACACCG
398256856 : ACGAGGAATTCCCAGTAAGCGCGAGTCATAAGCTCGCGTTGATTACGTCCCTGC-CCTTTGTACACACCG
```

Figure 5: Comparison of the 18S rRNA genes of *Nuttalliella namaqua* and *N. aff. namaqua*. Indicated are the 18S rRNA genes for *N. namaqua* from the present study (KF925835) and from the study by Mans et al. (2011) (333109165). The gene for *N. aff. namaqua* derive from the study by Horak et al. (2012) (398256856). Differences are boxed.

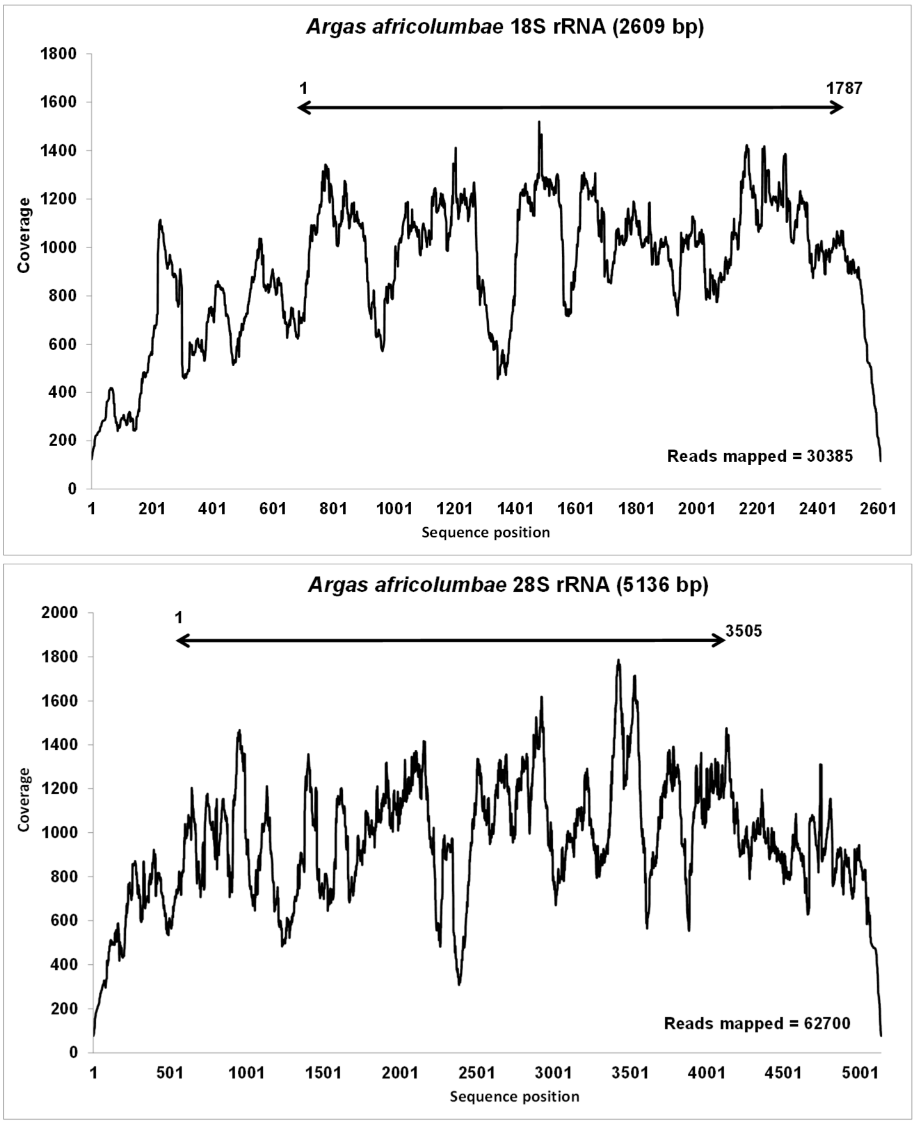


Figure 6: Coverage for the mapping of 18S and 28S rRNA gene assemblies for *A. africolumbae*. The number of reads mapped for each dataset is indicated, while the black arrows indicate the conventional full-length gene region.

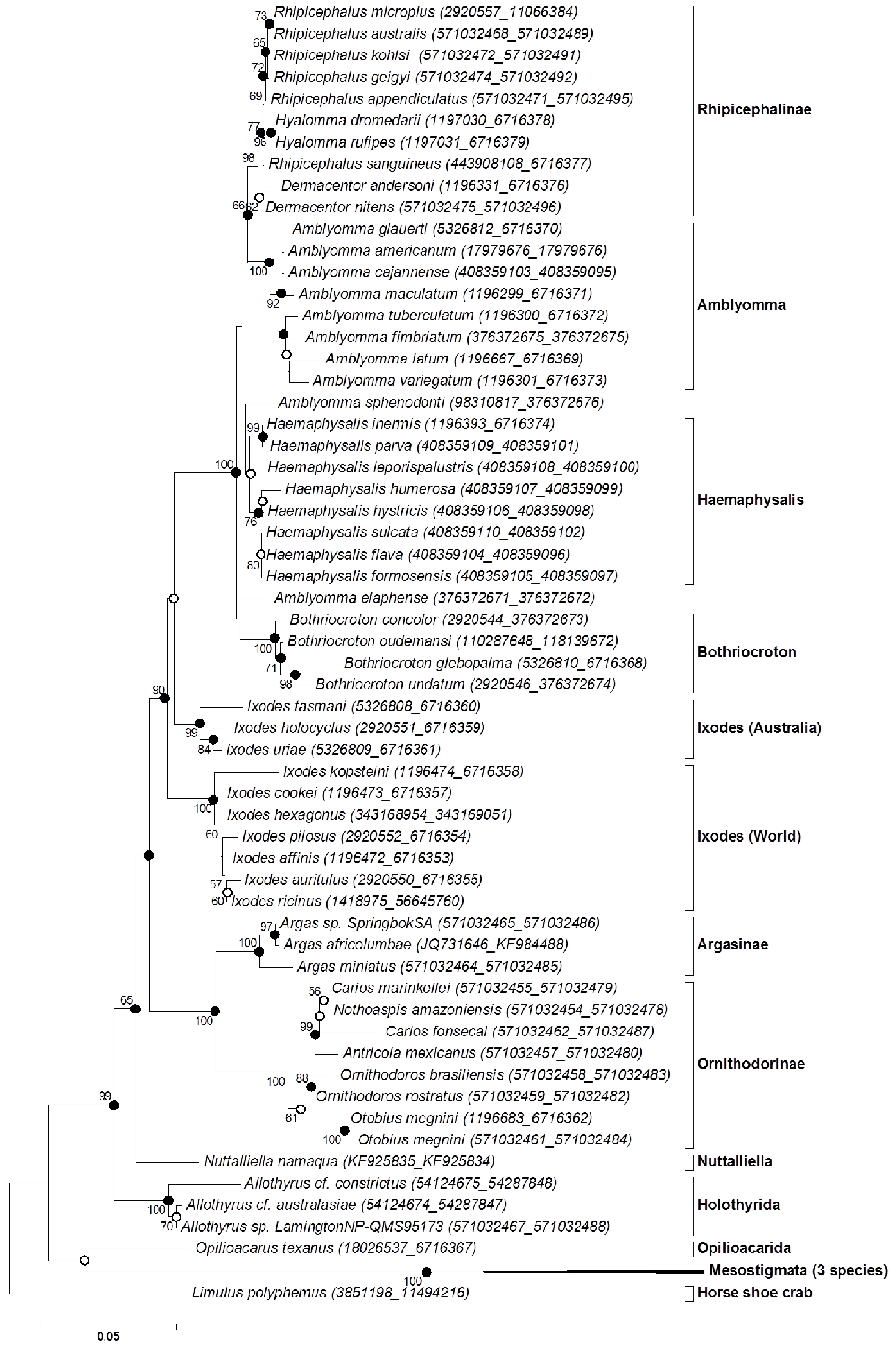

73 Rhipicephalus microplus (2920557_11066384)
Rhipicephalus australis (571032468_571032489)
65 Rhipicephalus kohlsi (571032472_571032491)
72 Rhipicephalus geigyi (571032474_571032492)
69 Rhipicephalus appendiculatus (571032471_571032495)
77 Hyalomma dromedarii (1197030_6716378)
96 Hyalomma rufipes (1197031_6716379)
98 Rhipicephalus sanguineus (443908108_6716377)
Dermacentor andersoni (1196331_6716376)
66 62 Dermacentor nitens (571032475_571032496)
Amblyomma glauerti (5326812_6716370)
Amblyomma americanum (17979676_17979676)
100 Amblyomma cajannense (408359103_408359095)
92 Amblyomma maculatum (1196299_6716371)
Amblyomma tuberculatum (1196300_6716372)
Amblyomma fimbriatum (376372675_376372675)
Amblyomma latum (1196667_6716369)
Amblyomma variegatum (1196301_6716373)
Amblyomma sphenodonti (98310817_376372676)
99 Haemaphysalis inermis (1196393_6716374)
Haemaphysalis parva (408359109_408359101)
100 Haemaphysalis leporispalustris (408359108_408359100)
Haemaphysalis humerosa (408359107_408359099)
76 Haemaphysalis hystricis (408359106_408359098)
Haemaphysalis sulcata (408359110_408359102)
80 Haemaphysalis flava (408359104_408359096)
Haemaphysalis formosensis (408359105_408359097)
Amblyomma elaphense (376372671_376372672)
Bothriocroton concolor (2920544_376372673)
100 Bothriocroton oudemansi (110287648_118139672)
71 Bothriocroton glebopalma (5326810_6716368)
98 Bothriocroton undatum (2920546_376372674)
90 Ixodes tasmani (5326808_6716360)
99 Ixodes holocyclus (2920551_6716359)
84 Ixodes uriae (5326809_6716361)
Ixodes kopsteini (1196474_6716358)
Ixodes cookei (1196473_6716357)
100 Ixodes hexagonus (343168954_343169051)
60 Ixodes pilosus (2920552_6716354)
Ixodes affinis (1196472_6716353)
57 Ixodes auritulus (2920550_6716355)
60 Ixodes ricinus (1418975_56645760)
97 Argas sp. SpringbokSA (571032465_571032486)
100 Argas africolumbae (JQ731646_KF984488)
Argas miniatus (571032464_571032485)
56 Carios marinkellei (571032455_571032479)
65 Nothoaspis amazoniensis (571032454_571032478)
100 99 Carios fonsecai (571032462_571032487)
Antricola mexicanus (571032457_571032480)
100 88 Ornithodoros brasiliensis (571032458_571032483)
99 Ornithodoros rostratus (571032459_571032482)
61 Otobius megnini (1196683_6716362)
100 Otobius megnini (571032461_571032484)
Nuttalliella namaqua (KF925835_KF925834)
Allothyrus cf. constrictus (54124675_54287848)
100 Allothyrus cf. australasiae (54124674_54287847)
70 Allothyrus sp. LamingtonNP-QMS95173 (571032467_571032488)
Opilioacarus texanus (18026537_6716367)
100 Mesostigmata (3 species)
Limulus polyphemus (3851198_11494216)
Rhipicephalinae
Amblyomma
Haemaphysalis
Bothriocroton
Ixodes (Australia)
Ixodes (World)
Argasinae
Ornithodorinae
Nuttalliella
Holothyrida
Opilioacarida
Horse shoe crab
0.05

Figure 7: Phylogenetic analysis of the nuclear 18S-28S rRNA concatenated dataset. Indicated is the topology of the Maximum Likelihood tree, with nodal support values for maximum likelihood (1000 bootstraps) indicated. Posterior probability support for Bayesian analysis (26136 trees) are indicated with black (>95%) and white (<95%) dots for nodes which retrieved the same topology.