

RESEARCH ARTICLE

Phylogenetic Variants of *Rickettsia africae*, and Incidental Identification of "*Candidatus Rickettsia Moyalensis*" in Kenya

Gathii Kimita^{1,2}, Beth Mutai¹, Steven Ger Nyanjom², Fred Wamunyokoli², John Waitumbi^{1*}

1 Walter Reed Project/ Kenya Medical Research Institute, Kisumu, Kenya, **2** Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya

* john.waitumbi@usamru-k.org



Abstract

Background

Rickettsia africae, the etiological agent of African tick bite fever, is widely distributed in sub-Saharan Africa. Contrary to reports of its homogeneity, a localized study in Asembo, Kenya recently reported high genetic diversity. The present study aims to elucidate the extent of this heterogeneity by examining archived *Rickettsia africae* DNA samples collected from different eco-regions of Kenya.

Methods

To evaluate their phylogenetic relationships, archived genomic DNA obtained from 57 ticks *a priori* identified to contain *R. africae* by comparison to *ompA*, *ompB* and *gltA* genes was used to amplify five rickettsial genes i.e. *gltA*, *ompA*, *ompB*, 17kDa and *sca4*. The resulting amplicons were sequenced. Translated amino acid alignments were used to guide the nucleotide alignments. Single gene and concatenated alignments were used to infer phylogenetic relationships.

Results

Out of the 57 DNA samples, three were determined to be *R. aeschlimanii* and not *R. africae*. One sample turned out to be a novel rickettsiae and an interim name of "*Candidatus Rickettsia moyalensis*" is proposed. The bonafide *R. africae* formed two distinct clades. Clade I contained 9% of the samples and branched with the validated *R. africae str ESF-5*, while clade II (two samples) formed a distinct sub-lineage.

Conclusions

This data supports the use of multiple genes for phylogenetic inferences. It is determined that, despite its recent emergence, the *R. africae* lineage is diverse. This data also provides evidence of a novel *Rickettsia* species, *Candidatus Rickettsia moyalensis*.

OPEN ACCESS

Citation: Kimita G, Mutai B, Nyanjom SG, Wamunyokoli F, Waitumbi J (2016) Phylogenetic Variants of *Rickettsia africae*, and Incidental Identification of "*Candidatus Rickettsia Moyalensis*" in Kenya. PLoS Negl Trop Dis 10(7): e0004788. doi:10.1371/journal.pntd.0004788

Editor: Ruifu Yang, Beijing Institute of Microbiology and Epidemiology, CHINA

Received: January 27, 2016

Accepted: May 29, 2016

Published: July 7, 2016

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: The sequences for the study OTUs are publicly available in the Genbank and can be accessed using accession numbers provided in the manuscript ([S3 Table](#)).

Funding: Financial support for this study was from a grant from the US Armed Forces Health Surveillance Center, Division of Global Emerging Infection Surveillance Operations. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors declared that no competing interests exist.

Author Summary

Rickettsia africae is a bacterium mainly vectored by *Amblyomma* and *Rhipicephalus* species of ticks. It is the etiological agent of African tick bite fever (ATBF), a spotted fever rickettsiosis that presents as an acute febrile illness characterized by petechial skin hemorrhages, from which the name is derived. This bacterium is probably the most important in sub-Saharan Africa, including Kenya, in terms of incidence and prevalence. This notwithstanding, the disease is poorly understood and is often mistreated as malaria, and therefore qualifies as a highly neglected disease. This study examined the genetic relationships of *R. africae* collected from diverse eco-regions of Kenya. We present data that indicate high genetic diversity in Kenya's *R. africae* and corroborate a recent study that reported similar genetic diversity in *R. africae* samples collected from a localized area in western Kenya. Importantly, we describe a divergent lineage and propose the name *Candidatus Rickettsia moyalensis*.

Introduction

Rickettsiae are obligate intracellular gram negative bacteria, belonging to the class *alpha-proteobacteria*. They are found in a diverse array of hosts ranging from vertebrates, arthropods, annelids, amoeba and plants. Based on a host perspective, the non-vertebrate-associated *Rickettsia* remain understudied and poorly characterized [1]. In contrast, the vertebrate-associated *Rickettsia* that are vectored by hematophagous arthropods such as ticks, fleas, lice and mites are better studied, are responsible for rickettsial diseases that are important cause of illness and death worldwide [2]. To date, this genus consists of 29 validated species and numerous partially characterized species [3,4], thus illustrating the difficulties of unravelling the composition of this seemingly homogeneous group of bacteria.

Phenotypic characters such as pathogenicity, growth temperature requirements, ability to polymerize host cell actin, and cross reactivity to somatic antigens of *Proteus vulgaris* strains (OX19 and OX2) and *P. mirabilis* OXK have been used to infer evolutionary relationships amongst rickettsiae [5]. From these criteria, the genus *rickettsia* was organized into three biotypes, namely, spotted fever group (SFG), typhus group (TG) and scrub typhus group (STG). The phenotypic characters have been found to be unreliable estimators of their phylogeny [6]. The advent of molecular tools brought major reorganizations in *rickettsia* taxonomy. For instance, by analysis of *Rickettsia* 16S rRNA (*rrs*), STG was removed from the genus *Rickettsia* and placed into its own genus, *Orientia*. This genus currently has only two species, *O. tsutsugamushi* [7] and a recently described *O. chuto* [8].

Several genes have been used for *Rickettsia* phylogenetic systematics: the *rrs* [5], *gltA* [9], 17kDa [10], *ompA* [11], *ompB* [12], *sca4* [13], *sca2* [14], and more recently, complete genomic sequences [15,16]. Currently, the genus delineates into four clades [17–19]: (i) The SFG which is the most derived, and consists of 23 validated species including *R. africae*, and numerous partially characterized species. Using whole genome approach, it has been realized that some of the members of rickettsiae such as *R. helvetica* do not fit in the SFG [16,20]. (ii) The transitional group (TRG) whose members are *R. akari* and *R. felis*. (iii) The TG which has only two members, namely, *R. typhi* and *R. prowazekii*; and (iv), the ancestral group (AG), whose members consists of *R. bellii* and *R. canadensis*. The use of the name TRG has been challenged [15,21]). In addition to the systemic of vertebrate-associated *Rickettsia*, clades associated with non-vertebrate *Rickettsia* have been described [16,19].

Of the rickettsiae, *R. africae* is probably the most important in Africa. It is the aetiological agent of African tick bite fever [22] and the most reported [23]. It has been reported in 22 sub-Saharan countries [3], the West Indies [24] and Oceania [25]. In Kenya, a recent surveillance study of rickettsiae in ticks identified 104 rickettsiae, of which 93% were *R. africae*, clearly, demonstrating its dominance [26]. *R. africae* is vectored by *Amblyomma* ticks, primarily *A. variegatum* and *A. hebraeum*. Infections have also been detected in many other ticks species [3,26–28] by PCR and not by competency studies.

In general, *Rickettsia* species are very closely related. Within the SFG where *R. africae* belongs, the mean nucleotide homogeneity with *rrs*, *gltA*, *ompA*, *ompB* and *sca4* genes ranges from 82.2% to 99.8% [29]. Considering this high interspecies homogeneity, intraspecies differences are even smaller. For example, using the more variable multi spacer typing that combined *dksA-xerC*, *mppA-purC*, and *rpmE-tRNA^{fMet}* spacer sequences, it was impossible to discriminate *R. africae* strains [30]. Using *ompA* and *ompB* genes, many groups have however reported heterogeneity [25,31,32]. A recent study reported an even higher heterogeneity of *R. africae* samples collected from a localized area in Western Kenya [33]. The study reported here sought to determine how widespread the heterogeneity of Kenya's *R. africae* is by examining DNA samples collected from different eco-regions of Kenya.

Methods

Ethics statement

This study was carried out using ticks collected from domestic animals presented for slaughter. The tick samples were collected under protocol SSC#1248 that was reviewed and approved by the Animal Use Committee of the Kenya Medical Research Institute.

Sample acquisition and study site

Details of the areas the tick were collected from, method of collection, DNA extraction and preliminary genotyping have been published before [26] and are summarized in [S1 Table](#).

Amplification of target genes

Sequence data for 57 tick-extracted DNA samples that had been identified as *R. africae* by comparison to 385 bp citrate synthase (*gltA*) gene, 530 bp outer membrane protein A gene (*ompA*) and 444 bp outer membrane protein B gene (*ompB*) were obtained from our laboratory's database from a previous study [26]. The sequences of both strands were re-checked for correctness and errors cleaned. Samples with short or missing sequences were re-amplified and re-sequenced. Two additional genes: the 450 bp 17kDa and 2700 bp *sca4* genes were also amplified and sequenced.

Primers used to amplify target genes are listed in [Table 1](#) and are previously described [34]. PCR reagents were obtained from Applied Biosystems (CA, USA) and reactions performed in a 25 µL reaction volume containing 10 µM of each primer, 200 µM of dNTP mix, 1.5U Taq polymerase and 2 mM MgCl₂. Amplification was carried out in a DNA thermal cycler (HID Veriti) from Applied Biosystems (CA, USA). The following conditions were used for amplification: For *ompA* and *gltA* genes: 3 min of initial denaturation at 94°C, then 40 cycles at 94°C for 30 sec, 53°C for 30 sec, 68°C for 1 min. For the *ompB* gene: 3 min of initial denaturation at 95°C, then 40 cycles at 95°C for 30 sec, 50°C for 30 sec, 68°C for 1 min 30 sec. For the 17kDa gene: 5 min of initial denaturation at 94°C, then 35 cycles at 94°C for 30 sec, 50°C for 1 min, 68°C for 1 min. For *sca4*: 5 min of initial denaturation at 95°C, then 40 cycles at 95°C for

Table 1. Primers used for PCR and sequencing.

Gene	Oligo Sequence (5'-3')	Amplicon size
<i>gltA</i>	^{1,2} CS49Fw: ACCTATACTTAAAGCAAGTATYGGT	385bp
	¹ CS1234Rv: TCTAGGTCTGCTGATTTTTTGTTC	
	² CS1258Rv: ATTGCAAAAAGTACAGTGAACA	
<i>ompB</i>	^{1,2} RAK1009Fw: ACATKGTATACARAGTGYTAATGC	444bp
	¹ OmpB1902Rv: CCGTCATTTCCAATAACTAACTC	
	² RAK1452Rv: SGTTAACTTKACCGYTTATAACTGT	
<i>ompA</i>	¹ OmpAM50Fw: TTGCGTTATAACACTTTTTAAGTGA	530bp
	¹ OmpA642Rv: ATTACCTATTGTTCCGTTAATGGCA	
	² 190-70Fw: ATGGCGAATATTTCTCCAAA	
	² 190-701Rv: GTTCCGTTAATGGCAGCATCT	
<i>sca4</i>	¹ D1Fw: ATGAGTAAAGACGGTAACCT	2700bp
	¹ D3069Rv: TCAGCGTTGTGGAGGGGAAG	
	² RrD749Fw: TGGTAGCATTAAAAGCTG	
	² RrD2685Rv: TTCAGTAGAAGATTTAGT	
17kDa	¹ Rp17kFw: AATGAGTTTTATACTTTACAAAATTCTAAAAACCA	450bp
	^{1,2} Rr2608Rv: CATTGTTTCGTCAGGTTGGCG	
	² Rr1175Fw: GCTCTTGCAACTTCTATGTT	

Fw = Forward primer; Rv = Reverse primer;

¹ = primary PCR;

² = secondary PCR and sequencing

doi:10.1371/journal.pntd.0004788.t001

45 sec, 60°C for 30 sec and 68°C for 3 min. All the amplifications were then completed by holding for 7 min at 72°C. To ascertain correct product sizes, a portion of the amplicons (5µL) was run on a 1% (w/v) agarose gel containing ethidium bromide. Product sizes were estimated by comparing with a molecular mass standard (1kb plus ladder, Invitrogen, (CA, USA).

Gene sequencing

The PCR products were purified using Isolate II PCR and Gel Kit (Bioline, UK) as recommended by the manufacturer. The purified PCR products were sequenced in both directions using the Big Dye Terminator Cycle Sequencing Kit v 3.1 (Applied Biosystems, CA, USA) and the sequences analyzed by capillary electrophoresis in a 3130 Genetic Analyzer (Applied Biosystems). The sequences were proofread, edited and assembled into consensus sequences using CLC Main Workbench v 7 (CLC Inc, Aarhus, Denmark), and used to query GenBank using the nucleotide Basic Local Alignment Search Tool (BLAST) [35].

Phylogenetic data analysis

Six different alignments were generated: Five of them corresponded to sequences of each target gene (*gltA*, *ompA*, *ompB*, 17kDa and *sca4*), and one corresponding to the concatenated sequence of all the five genes as well as the validated rickettsia strains derived from GenBank (see S2 Table for names of reference strains and their accession numbers). To ensure the accuracy of these alignments, nucleotide sequences were translated to their respective amino acids using the translate tools in the CLC Main Workbench v7. Amino acid alignments were made using Muscle v 3.8 software [36]. The protein alignments were then used to guide the corresponding nucleotide alignments using TOPALi V2 software [37].

For phylogenetic inference, MEGA v7 software [38] was used to estimate the best substitution model as well as estimate for the Maximum Likelihood (ML) trees for the individual genes. The concatenated alignment tree was also estimated with the ML method using a General Time Reversal (GTR) nucleotide substitution model with a gamma distribution (GTR+G). For bayesian probability analysis, jModeltest v 2.1 [39] was used to determine the GTR+G as the best fit model for *gltA*, *ompA*, *ompB*, 17kDa and *sca4* gene alignments. A partitioned analysis was then performed on the concatenated dataset using a bayesian Markov Chain Monte Carlo (MCMC) method implemented with MrBayes software v 3.2 [40]. The cluster confidence was given as posterior probabilities.

Results

Sample description

In total, 57 tick DNA samples were available for analysis. From these and as shown in supplementary information (S3 Table), 45 sequences were obtained for *gltA* (Genbank accession KX368721-KX368765), 57 for *ompA* (Accession: KX368868-KX368924), 44 for *ompB* (Accession: KX368823- KX368867), 57 for 17kDa (Accession: KX368766-KX368822) and 40 for *sca4* (Accession: KX368925-KX368964). The gene sequences were subjected to BLAST analysis for a preliminary verification of their identity. BLAST results are shown in S1 Table.

Single gene topology trees and their ability to resolve study samples

The topology of the *gltA* gene tree inferred with the ML method is shown in Fig 1, panel A. The tree shows unresolved evolutionary relationships (nodes with <50% bootstrap support values). Overlooking these clade credibility values, a major cluster (clade I) consisting of 89% of the sequences was observed in the more derived parts of the tree. The other two distinct clades were samples from the North Eastern part of Kenya. Clade II consisted of samples (044 and 045 from Wajir) that diverged as sister operational taxonomic units (OTUs). Clade III samples clustered at the basal part of the tree and came from Moyale (139, 135, 136 and 176) and Wajir (577).

In comparison to *gltA*, the *ompA* tree resolved majority of the nodes into three well supported groups (Fig 1, panel B). As in *gltA*, clade I had the majority of OTUs (52/57, 91%), and was the most derived. Majority of the nodes were unresolved, and the members clustered in a polytomy. As in *gltA*, clade II formed a dichotomy made of samples 044 and 045 from Wajir. Clade III constituted a cluster of 176_Moyale, 164_Wajir and 195_Machakos in the basal parts of the tree. In this clade, only 176_Moyale was shared with the *gltA* gene tree.

As shown in Fig 1, panel C, *ompB* also resolved the study OTUs into three groups. Clade I contained 40/44 (91%) members in a polytomy. Clade II contained the same sister OTUs (044 and 045) from Wajir. Clade III had the same samples as in the *ompA* gene tree. As with *gltA*, *ompA* and *ompB* genes, the topology of the 17kDa gene tree was consistent in delineating three clades (Fig 1, panel D). Clade I contained 52/57 (91%) of unresolved OTUs. Clade II consisted of sister OTUs (044 and 045) from Wajir, while clade III consisted of 164_Wajir, 195_Machakos and 176_Moyale

Compared to *gltA*, *ompA*, *ompB* and 17kDa, the *sca4* gene tree was better resolved especially in delineation of clade I (Fig 1, panel E) which consisted of 35/40 (86%) OTUs. Clade II consisted of sister OTUs (044 and 045) from Wajir. Clade III consisted of 176_Moyale, 164_Wajir, and an additional member 293_Migori.

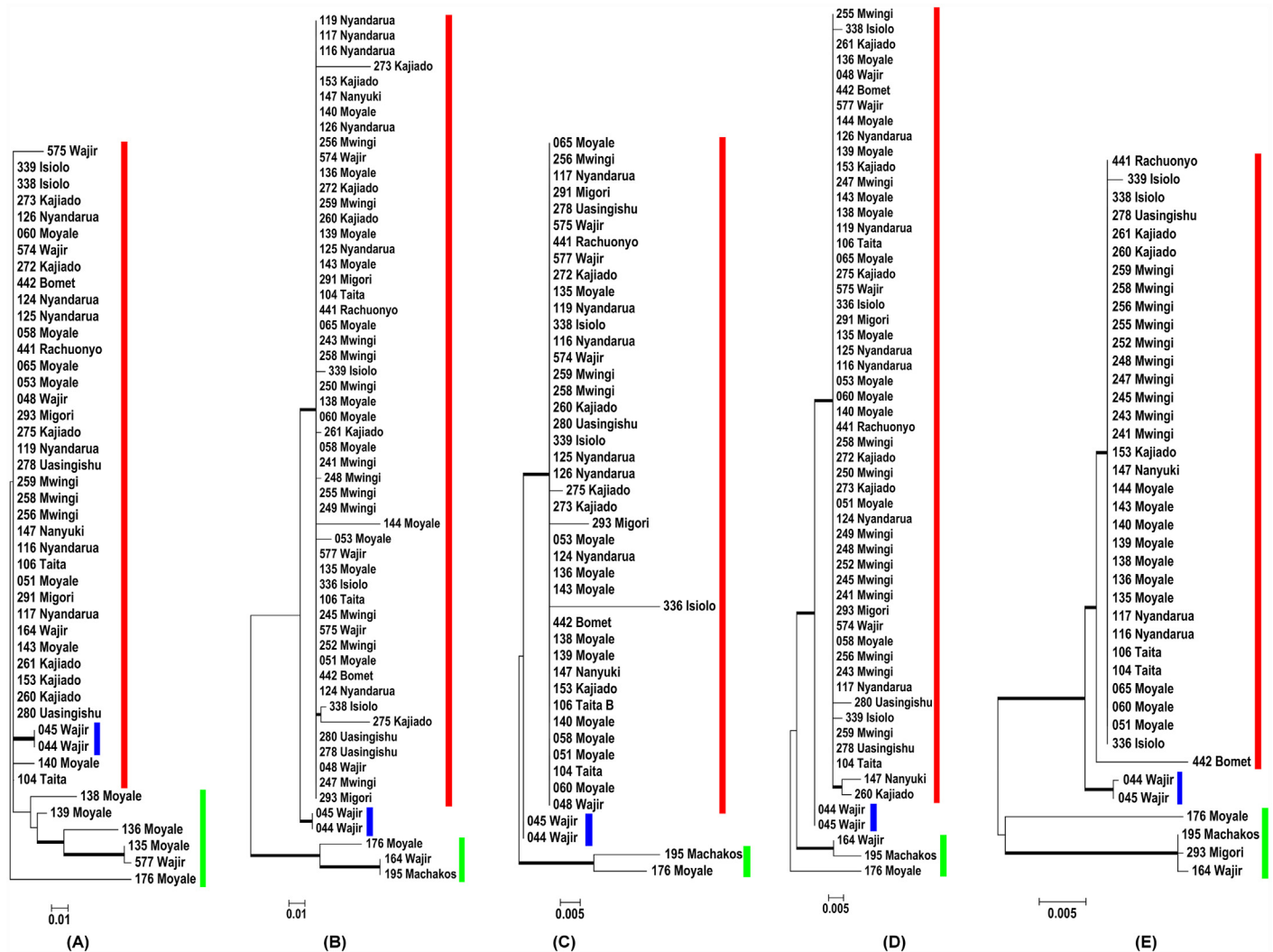


Fig 1. Phylogeny of *Rickettsia* study samples isolated from diverse eco-regions of Kenya. Maximum Likelihood trees were obtained from (A) *gltA*, (B) *ompA*, (C) *ompB*, (D) 17kDa and (E) *sca4* partial nucleotide sequences. Members of clade I, II and III are shown beside the bolded red, blue and green lines respectively. Numbers at the nodes are bootstrap proportions with 1000 replicates. Only bootstrap values >50% are shown. The scale bar indicates the number of substitutions per nucleotide position.

doi:10.1371/journal.pntd.0004788.g001

Concatenating *gltA*, *ompA*, *ompB*, 17kDa and *sca4* genes allows better phylogenetic resolution

In order to improve the phylogenetic resolution of individual genes, the five genes were concatenated. Out of the 57 samples initially available for analysis, only 39 were included in the concatenation. The choice of the 39 was influenced by availability of the limiting gene (39 *sca4* samples), and not missing more than one of the other four gene sequences. The concatenated sequences also included validated *Rickettsia* species available in GenBank. The concatenated tree constructed with Bayesian method is shown in Fig 2. The validated *Rickettsia* sequences delineated into the three known *Rickettsia* clades: TG, TRG and SFG. All the 39 study OTUs lay within the SFG clade, of which 33/39 (84%) clustered with the validated *R. africae* str ESF-5 shown as clade I. The two sister OTUs (044 and 045) from Wajir formed a distinct clade (clade II) that shared the most recent common ancestor with clade I. Two other samples (164_Wajir

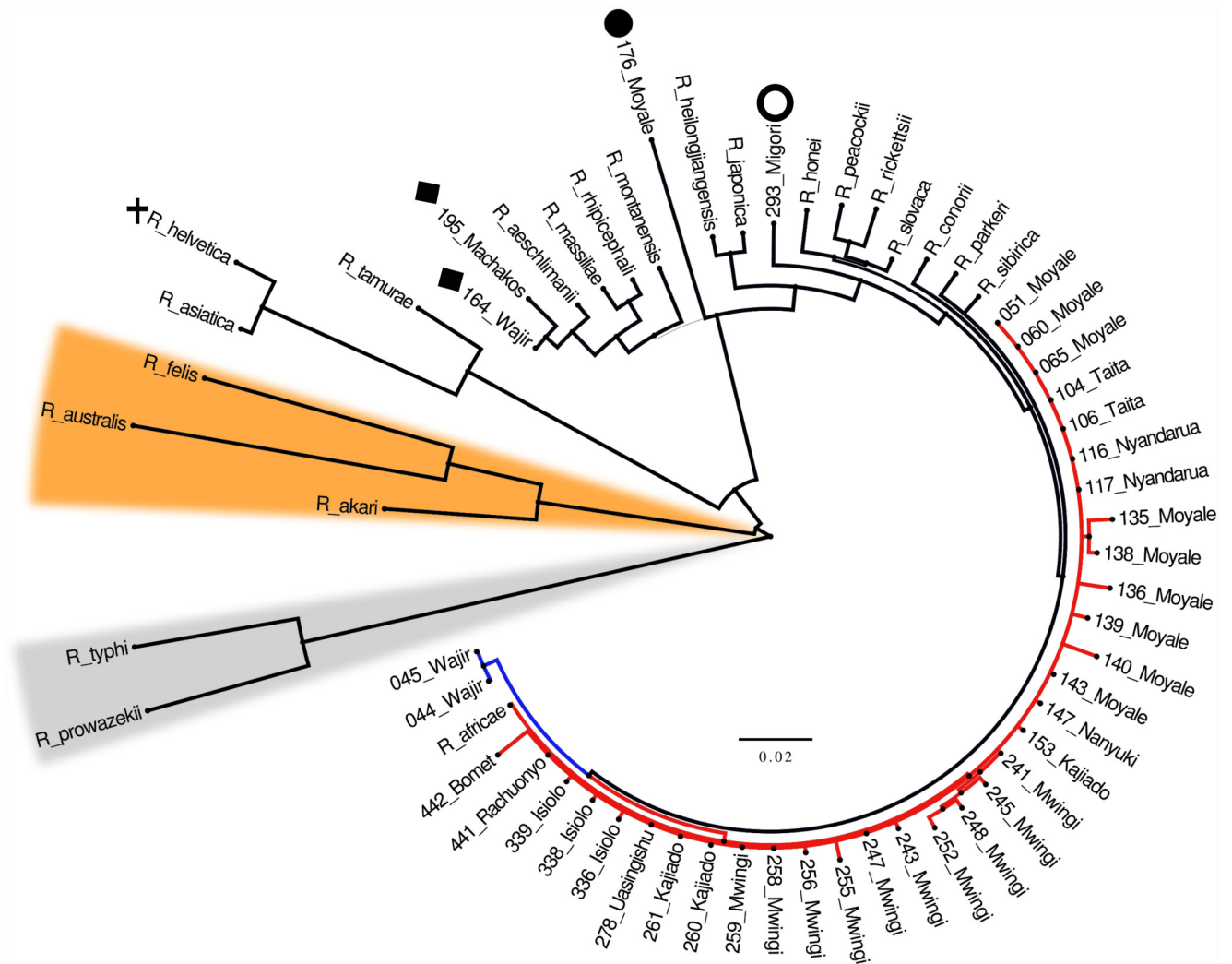


Fig 2. Bayesian probability tree of study samples with validated *Rickettsia* species. The tree is based on partitioned concatenated datasets of *gltA*, *ompA*, *ompB*, 17kDa and *sca4* partial nucleotide sequences. Amino acid alignments were used to guide the nucleotide alignments. The tree is estimated using a GTR+G substitution model as implemented in MrBayes v3.2. The tree is a consensus of 15,002 trees (post burn-in) pooled from two independent Markov Chain run in parallel. Thin lines indicate posterior probability values of < 1. Lineage diversity within the *R. africae* study samples is highlighted in red and blue to indicate clades i and ii respectively. Samples previously misclassified as *R. africae* are now classified as *R. aeschlimanii* (black diamond). Study sample 176_Moyale branches distinctly from other rickettsiae and is considered a novel rickettsia species and a provisional name "*Candidatus rickettsia moyalensis*" (black circle) is proposed. NB: Although 293_Migori (open circle) branched as a lone taxon, it clustered with *R. aeschlimanii* by Maximum Likelihood method. Non-spotted fever group lineages are highlighted orange for transition group and grey for typhus group. The status of *R. helvetica* (shown in black cross), originally in spotted fever group is now uncertain [20].

doi:10.1371/journal.pntd.0004788.g002

and 195_Machakos) previously identified as *R. africae* [26] delineated with the validated *R. aeschlimanii*. The positions of these OTUs held when tested by ML method (S1 Fig).

The position of 293_Migori was however tenuous, as it branched with *R. aeschlimanii* on ML method, but as a lone taxon between *R. heilongjiangensis* and *R. slovaca* on Bayesian analysis. Another interesting sample 176_Moyale branched as a lone taxon with total statistical support (a posterior probability value of 1), thus raising questions concerning its taxonomic status. We consider this sample as a novel rickettsiae and an interim name of "*Candidatus Rickettsia Moyalensis*" is proposed.

Phylogeny of *R. africae* study OTUs and those collected previously in Kenya

The *ompA* gene sequences obtained from our study samples and those published previously collected in Kenya [23,31] were used to generate a phylogenetic tree (Fig 3). With this tree, three clades were discernible. Clade I contained majority of the OTUs 61/66 (92.4%) and a cluster of other 7 that formed a subclade within clade I, that consisted of OTUs published in previous studies. Clade II consisted of sister OTUs (044 and 045) from Wajir, and a basal clade III that consisted of members 176_Moyale, 164_Wajir, and 195_Machakos. Clearly, five of our sequences (044, 045 and 164 from Wajir, 176 Moyale and 195 Machakos) are distinct from those described previously.

Discussion

In this study, a combination of relatively conserved (*gltA*, 17kD) and variable (*ompA*, *ompB*, and *sca4*) genes were used to infer the evolutionary topology of *R. africae* using DNA samples obtained from ticks [26]. This work expands on a recent study that reported a significant heterogeneity of *R. africae* samples collected from a localized area in Western Kenya [33]. We also report on possible existence of a sub-lineage within the *R. africae* samples, as well as identify a putative novel *Rickettsia* species that was associated with *R. appendiculatus* tick collected from a cow in Moyale County, in Northern Kenya.

Due to its intracellular lifestyle, rickettsiae are highly dependent on their primary tick vectors and tend to be selective for the tick species they infect. For example, *R. africae* are primarily vectored by the *Amblyomma* species [41] but infections have also been found in *Rhipicephalus* and *Hyalomma* ticks [3]. As shown in S1 Table, 84% of the *R. africae* sequences came from *Amblyomma* and *Rhipicephalus* ticks. The remaining 16% came from *H. truncatum* and other unspiciated *Hyalomma* ticks. Nevertheless, without competency studies, it is difficult to say which of these findings were true infections.

The *gltA* gene codes for citrate synthase, an enzyme ubiquitously found in nearly all living cells, and is central in energy metabolism [42]. Evolutionary history inferred from this gene demonstrated low sequence divergence and yielded poorly supported clades with unresolved nodes (Fig 1, panel A). This is expected considering that, even within the *Rickettsia* genus, *gltA* gives poor interspecies resolution especially for the more derived branches of the SFG [9]. Nevertheless, three unsupported groupings (Fig 1, panel A, shown by red, blue and green lines) were discernible. It could be argued that the lack of resolution emanated from using a small fragment (385 bp) compared to 1234 bp that was used by Roux et al [9]. The Roux study aimed to develop *gltA* gene as a phylogenetic marker for *Rickettsia* species. From their generated phylogenies, the resolution decreased within the more recently emerged species, of which *R. africae* is a member. Our study that focused on intra-species variation within the *R. africae* lineage had similar problems with *gltA* and it is doubtful that a longer *gltA* that failed to resolve the recently evolved rickettsiae would have resolved variation within the study OTUs. We think longer fragments may have increased the number of variable characters within the gene but not the resolution.

As shown in Fig 1 (panels B, C and D), gene trees derived from *ompA*, *ompB* and 17kDa had very similar topologies and the three groupings seen in *gltA* tree were better supported. This topological concordance gives credence to the derived gene trees. By BLAST analysis, majority of the OTUs clustering in clade I were identical to *R. africae* reference strain ESF (S1 Table).

Tree topologies derived from individual genes identified clades II and III as outliers (Fig 1, blue and green lines). Interestingly, for all the genes, clade II had only two members

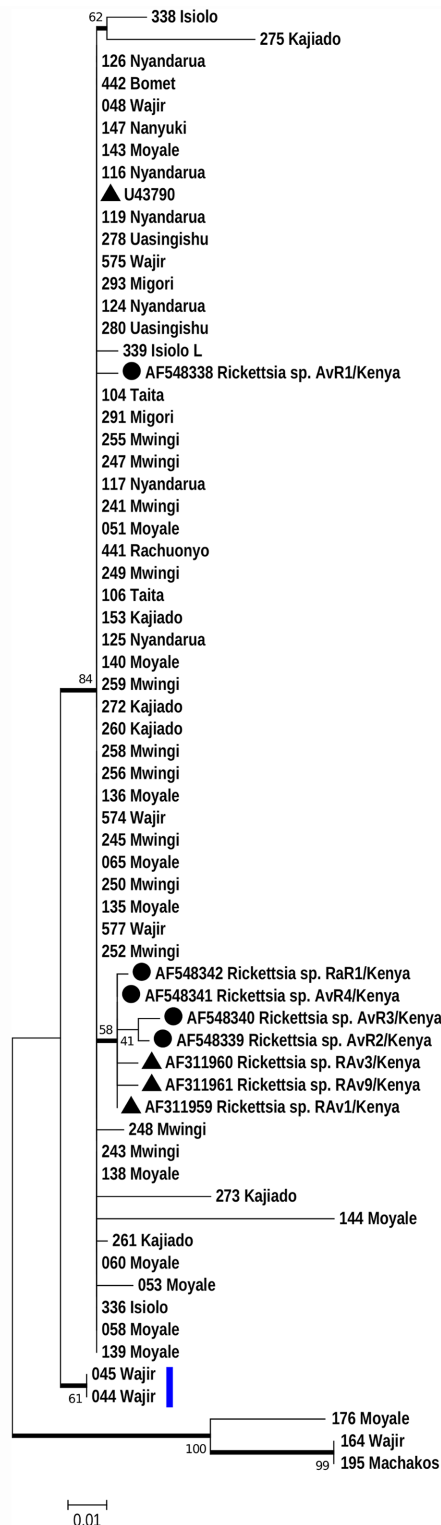


Fig 3. Phylogeny of *Rickettsia* sequences from this study and those collected previously in Kenya [23, 31]. *ompA* nucleotide sequences of study isolates and other *R. africae* reported from previous studies [23,31] were analysed by Maximum Likelihood method using MEGA v7 based on the Hasegawa-Kishino-Yano (HKY) model of substitution. The tree has a log likelihood ratio of -1049 and involved all codon positions. Members of clade I, II and III are shown beside the bolded red, blue and green lines respectively. Sequences from Parola et al 2001 [23] are shown as black triangles and those from Macaluso et al 2003 [31]

by black circles. Numbers at the nodes are bootstrap proportions with 1000 replicates. Only bootstrap values >50% are shown. The scale bar indicates the number of substitutions per nucleotide position. Clearly, five of our sequences (044, 045 and 164 from Wajir, 176 Moyale and 195 Machakos) are distinct from those described previously.

doi:10.1371/journal.pntd.0004788.g003

(044_Wajir and 045_Wajir) that appeared as a sub-lineage within *R. africae*. With four of the five genes analysed, BLAST analysis of the two samples showed highest identity to *R. africae* (accession no. HQ335126) with 99% for *gltA*, >99% for *ompA* (accession no. HQ335132), 99% for 17kDa (accession no. KF646137) and 100% for *sca4* (accession no. CP001612). BLAST analysis of the *ompB* gene showed highest homology (98%) to *R. mongolotimonae* (accession no. DQ097083). Within our study OTUs, members of clade III were the most genetically distant. With the *Sca4*, one member (293_Migori) that had consistently been in clade I was placed in clade III.

From the foregoing, the limitations of gene trees constructed from single genes are evident and give credence to recommendations to use a variety of genes sampled from different regions of the genome as the best practice for phylogeny assignment [43]. Fig 2 shows a Bayesian tree of study OTUs and the validated *Rickettsia* species derived from a concatenated sequence of *gltA*, *ompA*, *ompB*, 17kDa and *sca4*. The concatenated tree confirms the branching orders of clade I and authenticates members of clade II (044_ and 045_Wajir) as being a sub-lineage of *R. africae*. This derivation was confirmed by ML method (S1 Fig). Only after concatenation is the picture clearer that, the majority of members populating clade III in single gene analysis are not *R. africae*. Samples 164_Wajir and 195_Machakos (both associated with *H. truncatum* ticks from a cow) delineate with *R. aeschlimanii*, while 293_Migori appeared as a lone taxon (Fig 2).

Since sample 176_Moyale did not branch with any of the validated species in the concatenated tree (Fig 2), its sequence was compared to isolates available in GenBank. With the 276 bp *gltA* gene, it was most identical to *R. heilongjiangensis* (97.0%). With this gene, a homology of 99.9% is required in order for the sample to qualify as *R. heilongjiangensis* [29]. With the 489 bp fragment of *ompA* gene, the DNA was most identical to *Candidatus R. amblyommii* (97.0%) against the required homology of 98.8%. With the *ompB* gene, the 267 bp fragment was most identical to *Rickettsia rhipicephali* (99.0%) against a required homology of 99.2% to qualify as *R. rhipicephali*. With the *sca4* gene, the 1846 bp was most similar to *R. africae str ESF5* (97.0%) against the required homology of 99.3%. For the 17kDa, the closest homology was with *Rickettsia raoultii str Khabarovsk* (97.0%) identity. There are no published homology requirements for 17kDa gene. Given this level of nucleotide sequence divergence from validated *Rickettsia* species, our results support the consideration of 176_Moyale as a sample from a new *Rickettsia* species. Until grown in culture, and its biology elucidated, we propose that this sample be identified as “*Candidatus R. moyalensis*”.

Two other studies in Kenya have reported genetic heterogeneity within the *R. africae* lineage, one in Maasai Mara game reserve [31] and another in rural farming community in Asembo, Nyanza province [33]. The current study extends these findings and identifies Northern Kenya (Moyale and Wajir) as harbouring more heterogeneous *R. africae* or completely new species (Fig 2 and S1 Fig) compared to other regions. To determine homology of our study OTUs and those described as *R. africae* variants in previous studies in Kenya [23, 31], *ompA* gene sequences were compared. As shown in Fig 3, five of our sequences (044, 045 and 164 from Wajir, 176 Moyale and 195 Machakos) are distinct from those described previously. Unfortunately, we did not sequence a second region of *ompA* and *ompB* genes that were associated with significant sequence variation [33]. The *Rickettsia* species dynamics in Kenya are

probably being moulded by: 1) the highly mobile nomadic populations that move across wide geographical borders that create opportunities for mixing of rickettsiae from different livestock species (goats, sheep, cows, donkeys and camels). 2) The encroachment of wildlife habitats by nomadic pastoralists that introduce previously isolated *Rickettsia* species in wildlife to livestock. We speculate that these factors could be responsible for the observed genetic variations within the Kenyan *R. africae*, as well providing opportunity for genetic mixing that may result in creation of new lineages.

Conclusion

This study provides new information regarding the phylogenetic relationships of the *R. africae* lineage. This lineage, though only recently emerged [30] is clearly undergoing diversification, as observed in the branching order of the samples studied. A definite sub-lineage of *R. africae* (samples 044 and 045 from Wajir) was identified. It was impossible to confirm the placement of sample 293_Migori. Additional sequence data will be required to resolve the ambiguity. Lastly, a putative novel rickettsiae (sample 176_Moyale) with a proposed name of “*Candidatus Rickettsia moyalensis*” was identified. Further work will be needed to determine its prevalence in Kenya and its implications to human and/or animal disease.

Supporting Information

S1 Table. *Rickettsia* samples used in the study.

(DOCX)

S2 Table. Names and accession numbers for validated strains used in this study.

(DOCX)

S3 Table. List of Genbank accession numbers of study OTUs.

(DOCX)

S1 Fig. Maximum Likelihood tree of study samples and validated *Rickettsia* species. A General Time Reversal with Gamma distribution (GTR+G) model was used to infer phylogeny of concatenated partial sequences of *gltA*, *ompA*, *ompB*, 17kDa and *sca4* nucleotide sequences. Amino acid alignments were used to guide the nucleotide alignments. The tree with the highest log likelihood (-8252.6475) is shown. Study samples that are bonafide *R. africae* aggregate in clades I and II. Samples previously misclassified as *R. africae* are now classified as *R. aeschlimanii* (black diamond). Study sample 176_Moyale branches distinctly from other rickettsiae and is considered a novel rickettsia species provisionally named “*Candidatus rickettsia moyalensis*” (black circle). With this method, 293_Migori (open circle) clusters with *R. aeschlimanii*. Numbers at the nodes are bootstrap proportions with 1000 replicates. Only bootstrap values >50% are shown. SFG = spotted fever group, TRG = transition group, TG = typhus group. The status of *R. helvetica* (shown in black cross), originally in spotted fever group is now uncertain [20].

(TIF)

Acknowledgments

We thank the staff of the Basic Science Laboratory, Walter Reed Project for their considerable help with this study. We are grateful to Ms Elizabeth Vitalis for reading the manuscript and making useful remarks. This work is published with the permission of the Director, Kenya Medical Research Institute.

Disclaimer

The opinions or assertions contained herein are the private views of the authors, and are not to be construed as official, or as reflecting true views of the Department of the Army or the Department of Defense. The U.S. Government has the right to retain a nonexclusive, royalty free license and to any copyright covering this paper.

Author Contributions

Conceived and designed the experiments: JW GK. Performed the experiments: GK BM. Analyzed the data: JW GK. Contributed reagents/materials/analysis tools: JW. Wrote the paper: JW GK SGN FW.

References

1. Perlman SJ, Hunter MS, Zchori-Fein E. The emerging diversity of *Rickettsia*. *Proc Biol Sci*. 2006; 273: 2097–2106. PMID: [16901827](#)
2. Raoult D, Roux V. Rickettsioses as paradigms of new or emerging infectious diseases. *Clin Microbiol Rev*. 1997; 10: 694–719. PMID: [9336669](#)
3. Parola P, Paddock CD, Socolovschi C, Labruna MB, Mediannikov O, Kernif T, et al. Update on tick-borne rickettsioses around the world: A geographic approach. *Clin Microbiol Rev*. 2013; 26: 657–702. doi: [10.1128/CMR.00032-13](#) PMID: [24092850](#)
4. LPSN. List of prokaryotic names with standing in nomenclature [Internet]. [cited 18 Mar 2016]. Available: <http://www.bacterio.net/rickettsia.html>
5. Roux V, Raoult D. Phylogenetic analysis of the genus *Rickettsia* by 16S rDNA sequencing. *Res Microbiol*. 1995; 146: 385–396. PMID: [8525055](#)
6. Fournier PE, Raoult D. Current knowledge on phylogeny and taxonomy of rickettsia spp. *Ann N Y Acad Sci*. 2009; 1166: 1–11. doi: [10.1111/j.1749-6632.2009.04528.x](#) PMID: [19538259](#)
7. Tamura a, Ohashi N, Urakami H, Miyamura S. Classification of *Rickettsia tsutsugamushi* in a new genus, *Orientia* gen. nov., as *Orientia tsutsugamushi* comb. nov. *Int J Syst Bacteriol*. 1995; 45: 589–591. PMID: [8590688](#)
8. Izzard L, Fuller A, Blacksell SD, Paris DH, Richards AL, Aukkanit N, et al. Isolation of a novel *Orientia* species (*O. chuto* sp. nov.) from a patient infected in Dubai. *J Clin Microbiol*. 2010; 48: 4404–4409. doi: [10.1128/JCM.01526-10](#) PMID: [20926708](#)
9. Roux V, Rydkina E, Eremeeva M, Raoult D. Citrate synthase gene comparison, a new tool for phylogenetic analysis, and its application for the rickettsiae. *Int J Syst Bacteriol*. 1997; 47: 252–261. PMID: [9103608](#)
10. Anderson BE, Tzianabos T. Comparative sequence analysis of a genus-common Rickettsial antigen gene. *J Bacteriol*. 1989; 171: 5199–5201. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=210341&tool=pmcentrez&rendertype=abstract> PMID: [2768201](#)
11. Fournier PE, Roux V, Raoult D. Phylogenetic analysis of spotted fever group rickettsiae by study of the outer surface protein rOmpA. *Int J Syst Bacteriol*. 1998; 48 Pt 3: 839–849. PMID: [9734038](#)
12. Roux V, Raoult D. Phylogenetic analysis of members of the genus *Rickettsia* using the gene encoding the outer-membrane protein rOmpB (ompB). *Int J Syst Evol Microbiol*. 2000; 50: 1449–1455. PMID: [10939649](#)
13. Sekeyova Z, Roux V, Raoult D. Phylogeny of *Rickettsia* spp. inferred by comparing sequences of “gene D”, which encodes an intracytoplasmic protein. *Int J Syst Evol Microbiol*. 2001; 51: 1353–1360. PMID: [11491333](#)
14. Ngwamidiba M, Blanc G, Ogata H, Raoult D, Fournier PE. Phylogenetic study of *Rickettsia* species using sequences of the autotransporter protein-encoding gene sca2. *Ann N Y Acad Sci*. 2005; 1063: 94–99. PMID: [16481498](#)
15. Merhej V, Raoult D. Rickettsial evolution in the light of comparative genomics. *Biol Rev*. 2011; 86: 379–405. doi: [10.1111/j.1469-185X.2010.00151.x](#) PMID: [20716256](#)
16. Murray GGR, Weinert LA, Rhule EL, Welch JJ. The Phylogeny of *Rickettsia* Using Different Evolutionary Signatures: How Tree-Like is Bacterial Evolution? *Syst Biol*. 2016; 65: 265–279. doi: [10.1093/sysbio/syv084](#) PMID: [26559010](#)
17. Gillespie JJ, Beier MS, Rahman MS, Ammerman NC, Shallom JM, Purkayastha A, et al. Plasmids and Rickettsial evolution: Insight from *Rickettsia felis*. *PLoS One*. 2007; 2: e266. PMID: [17342200](#)

18. Gillespie JJ, Driscoll TP, Verhoeve VI, Utsuki T, Husseneder C, Chouljenko VN, et al. Genomic Diversification in Strains of *Rickettsia felis* Isolated from Different Arthropods. *Genome Biol Evol.* 2014; 7: 35–56. doi: [10.1093/gbe/evu262](https://doi.org/10.1093/gbe/evu262) PMID: [25477419](https://pubmed.ncbi.nlm.nih.gov/25477419/)
19. Weinert L a, Werren JH, Aebi A, Stone GN, Jiggins FM. Evolution and diversity of *Rickettsia* bacteria. *BMC Biol.* 2009; 7: 6. doi: [10.1186/1741-7007-7-6](https://doi.org/10.1186/1741-7007-7-6) PMID: [19187530](https://pubmed.ncbi.nlm.nih.gov/19187530/)
20. Driscoll T, Gillespie JJ, Nordberg EK, Azad AF, Sobral BW. Bacterial DNA sifted from the *Trichoplax adhaerens* (Animalia: Placozoa) genome project reveals a putative rickettsial endosymbiont. *Genome Biol Evol.* 2013; 5: 621–645. doi: [10.1093/gbe/evt036](https://doi.org/10.1093/gbe/evt036) PMID: [23475938](https://pubmed.ncbi.nlm.nih.gov/23475938/)
21. Fournier P, Belghazi L, Robert C, Elkarkouri K, Richards AL, Ogawa M, et al. Variations of Plasmid Content in *Rickettsia felis*. *PLoS One.* 2008; 3: 3–8.
22. Kelly PJ, Beati L, Mason PR, Matthewman L a, Roux V, Raoult D. *Rickettsia africae* sp. nov., the etiological agent of African tick bite fever. *Int J Syst Bacteriol.* 1996; 46: 611–614. PMID: [8934912](https://pubmed.ncbi.nlm.nih.gov/8934912/)
23. Parola P. Rickettsioses in sub-Saharan Africa. *Ann N Y Acad Sci.* 2006; 1078: 42–47. PMID: [17114679](https://pubmed.ncbi.nlm.nih.gov/17114679/)
24. Kelly PJ. *Rickettsia africae* in the West Indies. *Emerg Infect Dis.* 2006; 12: 224–226. PMID: [16494746](https://pubmed.ncbi.nlm.nih.gov/16494746/)
25. Eldin C, Mediannikov O, Davoust B, Cabre O, Barré N, Raoult D, et al. Emergence of *Rickettsia Africae*, Oceania. *Emerg Infect Dis.* 2011; 17: 100–102. doi: [10.3201/eid1701.101081](https://doi.org/10.3201/eid1701.101081) PMID: [21192865](https://pubmed.ncbi.nlm.nih.gov/21192865/)
26. Mutai BK, Wainaina JM, Magiri CG, Nganga JK, Ithondeka PM, Njagi ON, et al. Zoonotic Surveillance for *Rickettsiae* in Domestic Animals in Kenya. *Vector-Borne Zoonotic Dis.* 2013; 13: 360–366. doi: [10.1089/vbz.2012.0977](https://doi.org/10.1089/vbz.2012.0977) PMID: [23477290](https://pubmed.ncbi.nlm.nih.gov/23477290/)
27. Maina AN. Sero-epidemiology and molecular characterization of *Rickettsiae* infecting humans, selected animals and arthropod vectors in Asembo, western Kenya, 2007–2010 Alice Ngonyo Maina A thesis submitted in partial fulfilment for the degree of Doctor of Philosop [Internet]. Jomo Kenyatta University of Agriculture and Technology. 2012. Available: <http://ir.jkuat.ac.ke/handle/123456789/1016>
28. Socolovschi C, Matsumoto K, Marie J Lou, Davoust B, Raoult D, Parola P. Identification of *Rickettsiae*, Uganda and Djibouti [2]. *Emerg Infect Dis.* 2007; 13: 1508–1509. doi: [10.3201/eid1310.070078](https://doi.org/10.3201/eid1310.070078) PMID: [18257998](https://pubmed.ncbi.nlm.nih.gov/18257998/)
29. Fournier PE, Dumler JS, Greub G, Zhang J, Wu Y, Raoult D. Gene Sequence-Based Criteria for Identification of New *Rickettsia* Isolates and Description of *Rickettsia heilongjiangensis* sp. nov. *J Clin Microbiol.* 2003; 41: 5456–5465. PMID: [14662925](https://pubmed.ncbi.nlm.nih.gov/14662925/)
30. Fournier P-E, Raoult D. Identification of rickettsial isolates at the species level using multi-spacer typing. *BMC Microbiol.* 2007; 7: 72. PMID: [17662158](https://pubmed.ncbi.nlm.nih.gov/17662158/)
31. Macaluso KR, Davis J, Alam U, Korman A, Rutherford JS, Rosenberg R, et al. Spotted fever group *Rickettsiae* in ticks from the masai mara region of Kenya. *Am J Trop Med Hyg.* 2003; 68: 551–553. PMID: [12812343](https://pubmed.ncbi.nlm.nih.gov/12812343/)
32. Parola P, Inokuma H, Camicas JL, Brouqui P, Raoult D. Detection and identification of spotted fever group *Rickettsiae* and *Ehrlichiae* in African ticks. *Emerg Infect Dis.* 2001; 7: 1014–1017. PMID: [11747731](https://pubmed.ncbi.nlm.nih.gov/11747731/)
33. Maina AN, Jiang J, Omulo S a., Cutler SJ, Ade F, Ogola E, et al. High Prevalence of *Rickettsia africae* Variants in *Amblyomma variegatum* Ticks from Domestic Mammals in Rural Western Kenya: Implications for Human Health. *Vector-Borne Zoonotic Dis.* 2014; 14: 693–702. doi: [10.1089/vbz.2014.1578](https://doi.org/10.1089/vbz.2014.1578) PMID: [25325312](https://pubmed.ncbi.nlm.nih.gov/25325312/)
34. Jiang J, Blair PJ, Felices V, Moron C, Cespedes M, Anaya E, et al. Phylogenetic analysis of a novel molecular isolate of spotted fever group *Rickettsiae* from northern Peru: *Candidatus Rickettsia andeanae*. *Ann N Y Acad Sci.* 2005; 1063: 337–342. PMID: [16481537](https://pubmed.ncbi.nlm.nih.gov/16481537/)
35. Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ. Basic Local Alignment Search Tool. *J Mol Biol.* 1990; 215: 403–410. PMID: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/)
36. Edgar RC, Drive RM, Valley M. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32: 1792–1797. PMID: [15034147](https://pubmed.ncbi.nlm.nih.gov/15034147/)
37. Milne I, Lindner D, Bayer M, Husmeier D, Mcguire G, Marshall DF, et al. TOPALi v2: A rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics.* 2009; 25: 126–127. doi: [10.1093/bioinformatics/btn575](https://doi.org/10.1093/bioinformatics/btn575) PMID: [18984599](https://pubmed.ncbi.nlm.nih.gov/18984599/)
38. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016; 1–11.
39. Posada D. jModelTest: Phylogenetic model averaging. *Mol Biol Evol.* 2008; 25: 1253–1256. doi: [10.1093/molbev/msn083](https://doi.org/10.1093/molbev/msn083) PMID: [18397919](https://pubmed.ncbi.nlm.nih.gov/18397919/)
40. Ronquist F, Teslenko M, Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3. 2 : Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol.* 2012; 61: 539–542. doi: [10.1093/sysbio/sys029](https://doi.org/10.1093/sysbio/sys029) PMID: [22357727](https://pubmed.ncbi.nlm.nih.gov/22357727/)

41. Fournier P-E, El Karkouri K, Leroy Q, Robert C, Giumelli B, Renesto P, et al. Analysis of the *Rickettsia africae* genome reveals that virulence acquisition in *Rickettsia* species may be explained by genome reduction. *BMC Genomics*. 2009; 10: 166. doi: [10.1186/1471-2164-10-166](https://doi.org/10.1186/1471-2164-10-166) PMID: [19379498](https://pubmed.ncbi.nlm.nih.gov/19379498/)
42. Wiegand G, Remington IJ. Citrate Synthase: Structure, Control, and Mechanism. *Annu Rev Biophys Chem*. 1986; 15: 97–117. PMID: [3013232](https://pubmed.ncbi.nlm.nih.gov/3013232/)
43. Cummings MP, Otto SP, Wakeley J. Sampling properties of DNA sequence data in phylogenetic analysis. *Mol Biol Evol*. 1995; 12: 814–822. PMID: [7476127](https://pubmed.ncbi.nlm.nih.gov/7476127/)