Evidence confirming the phylogenetic position of *Anaplasma centrale* (*ex* Theiler 1911) Ristic and Kreier 1984

Running title: Confirming the phylogenetic position of *Anaplasma centrale*

Zamantungwa T. H. Khumalo[1], Kelly A. Brayton[1,2], Nicola E. Collins[1], Mamohale E. Chaisi[1], Melvyn Quan[1], and Marinda C. Oosthuizen[1*]

[1] Vectors and Vector-borne Diseases Research Programme, Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Private Bag X04, Onderstepoort, 0110, South Africa

[2] Program in Genomics, Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA, 99164-7040, USA

*Author for correspondence: Marinda C. Oosthuizen.

Tel: +27 12 529 8390. Fax: +27 12 529 8312. Email: marinda.oosthuizen@up.ac.za

Keywords: *Anaplasma centrale*, *A. marginale* subspecies *centrale*, *Anaplasma marginale*, Anaplasmosis, South Africa, Phylogenetic analysis

Subject category: Evolution, Phylogeny and Biodiversity

Paper type: Research

Footnote: The GenBank accession numbers for the 16S rRNA gene, *groEL* and *msp4* gene sequences are given in Table S2. Two supplementary tables are available as online supplementary material.

LIST OF NON-STANDARD ABBREVIATIONS

AEP	Addo Elephant National Park
Berg	Bergville
BLAST	Basic Local Alignment Search Tool
CNP	Camdeboo National Park
HiP	Hluhluwe iMfolozi Park
KNP	Kruger National Park
MEGA	Molecular Evolutionary Genetics Analysis
MNP	Mokala National Park
NZG	National Zoological Gardens
qPCR	Quantitative real-time PCR
WBRC	Wildlife Biological Resource Center

ABSTRACT

In 1911, Sir Arnold Theiler isolated and described a parasite that was very similar to *A. marginale* but which was more centrally located within the erythrocytes of the host cells, and was much less pathogenic than *A. marginale*. He named the parasite *A. marginale* variety *centrale*. The name *A. centrale*, referring to the same organism, was published in Validation List no. 15 in 1984, but the publication was based on an erroneous assumption that Theiler had indicated that it was a separate species. Many authors have subsequently accepted this organism as a separate species, but evidence to indicate that it is a distinct species has never been presented. The near full length 16S rRNA gene sequence, and the deduced amino acid sequences for *groEL* and *msp4* from several isolates of *A. marginale* and *A. centrale* from around South Africa were compared with those of the *A. marginale* type strain, St Maries, and the *A. centrale* Israel strain and other reference sequences. Phylogenetic analyses of these sequences demonstrated that *A. centrale* consistently forms a separate clade from *A. marginale*, supported by high bootstrap values (≥90%), revealing that there is divergence between these two organisms. In addition, we discuss distinctive characteristics which have been published recently, such as differences in Msp1a/Msp1aS gene structure, as well as genome architecture that provide further evidence to suggest that *A. centrale* is, in fact, a separate species. Our results, therefore, provide evidence to support the existing nomenclature, and confirm that *A. centrale* (*ex* Theiler 1911) (Ristic and Kreier, 1984) is, indeed, a distinct species.

INTRODUCTION

In 1896, a point-like pathogen in blood smears of cattle was reported and described as a "very minute roundish body which is stained blue to bring it into view. The body as a rule is situated near the edge of the corpuscle" [1]. Fourteen years later, after extensive microscopic examination of infected red blood cells, Sir Arnold Theiler described this minute roundish body as *Anaplasma marginale*; referring to the pathogen as having "marginal points" in bovine

erythrocytes, and being the causative agent of gallsickness or bovine anaplasmosis [2]. A year later, Theiler isolated and described a very similar parasite which was more centrally located within the erythrocytes of the host cells. He named the parasite *A. marginale* variety *centrale* [3]. The latter was found to be less pathogenic in domestic animals and conferred immunity against infection by *A. marginale* [3].

Anaplasma marginale variety *centrale* is often referred to as a separate *Anaplasma* species [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. Ristic [15] erroneously stated that "In 1911, Theiler, who first described *A. centrale*, indicated that it was a separate species and thus distinct from *A. marginale*". This resulted in the inclusion of the name *A. centrale* in List No.15 of Validation of the Publication of New Names and New Combinations Previously Effectively Published Outside the International Journal of Systematic Bacteriology [16]. In conjunction, the organism was listed as a separate species in Bergey's Manual of Systematic Bacteriology [17]. Thus, many authors refer to *A. marginale* variety *centrale* as a separate species. However, others have recognized that the description by Ristic and Kreier [17] was flawed, and that, since a formal species description is lacking, the official taxonomic classification should revert to its original designation as a variety of *A. marginale* [18, 19].

In 2001, Dumler *et al*. [18] reorganized the order Rickettsiales, based on phylogenetic analyses of the 16S rRNA and *groEL* genes. These authors indicated that the 16S rRNA gene sequences of strains of *A. marginale*, *A. ovis* and *A. centrale* are nearly identical with 99.1% similarity, supporting Theiler's original description of *A. centrale* being a variant of *A. marginale* [3]. However, they noted the existence of a strain of *A. centrale* (the Aomori strain) with a 16S rRNA gene sequence (1292 bp) that had 1.8% nucleotide difference from other phenotypically characterised strains of *A. centrale*. Sequence data from other genes was not available at the time to enable these authors to resolve this taxonomic fine point.

To date, the taxonomic status of *A. centrale* and its relationship to *A. marginale* sensu stricto remains unclear. In this study, a comparative phylogenetic analysis was performed for three conserved genes of *A. marginale* and *A. centrale* obtained from cattle, buffalo, and black wildebeest samples originating from different geographical areas of South Africa. The full length 16S rRNA (coding for the small subunit ribosomal RNA gene), *groEL* (encoding a chaperone) and *msp4* (encoding major surface protein 4) genes were amplified and sequenced. Phylogenetic analysis was employed to resolve the appropriate phylogenetic position of *A. centrale* and coupled with other retrospective genomic analyses to provide evidence that *A. centrale* is indeed a separate species.

MATERIALS AND METHODS

A total of 380 blood samples from African buffalo, waterbuck, eland, black wildebeest, blue wildebeest and cattle were obtained from the Wildlife Biological Resource Center and Biobank South Africa under the auspices of the National Zoological Gardens of South Africa, from the South African National Parks Biobank and from two commercial farms in Bergville, KwaZulu-Natal. The samples were previously screened for the presence of *A. marginale* (targeting the *msp1ß* gene) and *A. centrale* (targeting the *groEL* gene) using a duplex quantitative real-time PCR (qPCR) assay; qPCR results indicated that these infections were common in both wildlife and cattle, and often occurred as mixed infections (ranging from 28 to 100%) in cattle and buffalo (see Khumalo *et al* [20] for details).

Selection of samples for amplification, cloning and sequencing

At least three samples were selected from each National Park and the two Bergville farms for targeted gene sequencing (Table 1). National Parks included: Mokala National Park (MNP), Camdeboo National Park (CNP), Kruger National Park (KNP), Hluhluwe iMfolozi Park (HiP)

and Addo Elephant National Park (AEP). Only samples that were co-infected with *A. marginale* and *A. centrale* were chosen for further analysis so that that the full gene set could be amplified from both species from individual animals.

Table 1. Origin of samples used for sequencing and phylogenetic analysis.

No.	Sample	Origin	Province	Host
1	MNP958	Mokala National Park	Northern Cape	Black wildebeest
2	MNP999			Buffalo
3	MNP1000			Buffalo
4	MNP1021			Buffalo
5	CNP976	Camdeboo National Park	Eastern Cape	Buffalo
6	CNP978			Buffalo
7	CNP979			Buffalo
8	CNP985			Buffalo
9	KNP581	Kruger National Park	Mpumalanga	Buffalo
10	KNP582			Buffalo
11	KNP584			Buffalo
12	HiP2	Hluhluwe iMfolozi Park	KwaZulu-Natal	Buffalo
13	HiP3			Buffalo
14	HiP4			Buffalo
15	HiP5			Buffalo
16	HiP6			Buffalo
17	HiP7			Buffalo
18	AEP1002	Addo Elephant National Park	Eastern Cape	Buffalo
19	AEP1003			Buffalo
20	AEP1007			Buffalo
21	AEP1013			Buffalo
22	Berg19	Bergville Farm 1	KwaZulu-Natal	Cattle
23	Berg25	Bergville Farm 2	KwaZulu-Natal	Cattle
24	Berg27	Bergville Farm 2		Cattle

Amplification of the 16S rRNA, groEL *and* msp4 *genes*

The near full-length genes were amplified using the primers described in Table 2. The PCR was performed in a final reaction volume of 25 µl, containing 1X Phusion Flash High-Fidelity PCR Master Mix (includes Phusion Flash II DNA Polymerase, reaction buffer, dNTPs, and $MgCl_2$) (Thermo Fisher Scientific, South Africa), 0.1 µM of each primer and 10 to 25 ng total genomic DNA. The thermal cycling programme was as follows: an initial denaturation at 98ºC for 10 s, 30 cycles of denaturation at 98ºC for 1 s, annealing temperature as shown in Table 2

for 5 s, and extension at 72ºC for 15 s, followed by a final extension at 72ºC for 1 min and a hold at 4ºC.

Table 2. Oligonucleotide primers used in this study.

Gene	Oligo	Sequence (5′–3′)	Annealing temperature	Size (bp)	Reference
16S rRNA	fD1	AGAGTTTGATCCTGGCTCAG	55 ˚C	1470	[58]
	rP2	ACGGCTACCTTGTTACGACTT			
groEL	GroELF	GCGCATTCTGGAGGCTG	64 ˚C	1482	This study
	GroELR	GCGTTTGACTTGGCTGTGTC			
A. marginale msp4	MSP45	GGGAGCTCCTAATTACAGAGAATTGTTTAC	60 ˚C	868	[37]
	MSP43	GCAAGATTCCTGTTCAGCTAAGGATCCGG			
A. centrale msp4	MSP4ACF	GCTCCCTACTTGTCAGTGGGCCTG	67 ˚C	800	This study

Sequencing and phylogenetic analysis of 16S rRNA, groEL *and* msp4 *genes*

Amplicons of the correct sizes were purified, ligated into the pJET vector (Thermo Fisher Scientific) and transformed into *E. coli* JM109 High Efficiency Competent cells (Promega, Madison, WI). Recombinant plasmids were purified using a High Pure Plasmid Isolation Kit (Roche Diagnostics, Mannheim, Germany) and sequenced (5–12 recombinants per sample) using 0.2 µM M13 primers and ABI Big Dye V3.1 on an ABI 3500XL Genetic Analyzer at Inqaba Biotechnical Industries (Pty) Ltd (Pretoria, South Africa). Sequences were assembled and edited using CLC Main Workbench 7 (Qiagen, http://www.clcbio.com). Searches of databases for homologous sequences were performed using BLASTn [21]. A multiple sequence alignment was performed for each gene or deduced amino acid sequence, along with sequences of related genera available in GenBank (Table S1), using MAFFT (multiple sequence alignment programme) v6 employing the FFT-NS-1 algorithm [22]. The alignments were truncated to the size of the shortest sequence using BioEdit v7 [23].

Phylogenetic trees were constructed by the neighbour-joining, maximum likelihood and maximum parsimony methods as implemented by the Molecular Evolutionary Genetics

Analysis version 7.0 (MEGA7) software package [24]. The maximum likelihood tree was inferred based on the Poisson correction model for Msp4 and GroEL, while the 16S rDNA tree was based on the Jukes-Cantor model and GTR model; this was in combination with the bootstrap method [25] using 1000 replicates/tree for each method. The genetic distances between the sequences were estimated by determining the number of nucleotide/amino acid differences between sequences using MEGA7 [24]. Bayesian phylogenetic trees were constructed using MrBayes v3.1.2 [26]. The WAG+G+F and JTT+G+F model were determined to be the best-fit for the Msp4 and GroEL amino acid sequence data, respectively. This was determined by the ProtTest [27], while the 16S rRNA nucleotide data best-fit model was determined to be invgamma using the Modeltest v3.7 software package [28]. All consensus trees were edited using MEGA7. The GenBank accession numbers of reference sequences used in this study are reported in Table S1, while the 16S rRNA, *groEL* and *msp4* gene sequences used to construct the phylogenetic trees were submitted to GenBank, and these accession numbers are reported in Table S2.

RESULTS

16S rRNA, groEL *and* msp4 *gene sequence and phylogenetic analysis*

We analysed the 16S rRNA, *groEL* and *msp4* gene sequences from 24 animals co-infected with *A. marginale* and *A. centrale*; these included three cattle, 20 African buffalo and one black wildebeest. We found four (Ac1, Ac2, Ac3 and Ac4) 16S rRNA gene sequence types for *A. centrale* and three (Am1, Am2 and Am3) for *A. marginale*. Except for Ac4, which was identical to the *A. centrale* Israel strain (CP001759, ACIS_01352), the *A. centrale* 16S rRNA sequences obtained in this study differed from the *A. centrale* Israel strain by 1 to 2 nucleotides. The *A. marginale* 16S rRNA sequences differed from the *A. marginale* St Maries strain (CP000030, AMr3) by 1 to 2 nucleotides.

Table 3. *A. centrale* and *A. marginale* 16S rRNA, GroEL and Msp4 genotypes.

Sample	**16S rRNA**							**GroEL**			**Msp4**						
	Ac*				**Am†**			**Ac**	**Am**		**Ac**				**Am**		
	1	**2**	**3**	**4**	**1**	**2**	**3**	**1**	**1**	**2**	**1**	**2**	**3**	**4**	**1**	**2**	**3**
MNP958										X							
MNP999			X	X	X		X	X					X				X
MNP1000			X				X	X		X				X			X
MNP1021							X			X							X
CNP976				X			X	X	X					X	X		X
CNP978				X				X						X			
CNP985				X				X									
CNP979														X			
KNP581		X		X			X	X			X						X
KNP582										X						X	X
KNP584		X		X			X	X		X			X				X
HiP2										X				X			X
HiP3		X						X		X				X			
HiP4	X									X				X			
HiP5												X	X				
HiP6			X	X			X	X		X				X			X
HiP7										X				X			X
AEP1002										X							
AEP1003								X									
AEP1007				X		X	X	X	X					X			X
AEP1013			X	X			X	X									
Berg19												X		X			
Berg25													X	X			
Berg27			X	X			X	X		X			X	X			X

*Ac, *A . centrale.*

tblfn2†Am, *A . marginale.*

The *A. centrale* (Ac1) GroEL deduced amino acid sequence was completely conserved and identical to the *A. centrale* Israel vaccine strain (CP001759, ACIS_00394) and the recently published *groEL* sequences from KNP [29] (data not shown), but differed from other previously published South African and Italian isolates by 1 and 2 amino acids, respectively. Two GroEL sequence types were identified for *A. marginale* (Am1 was identical to the St Maries sequence (CP000030, AM944), while Am2 differed from it by 3 amino acids). We found four (Ac1, Ac2, Ac3 and Ac4) *A. centrale* Msp4 deduced amino acid sequence types that differed by 1 to 4 amino acid residues from the *A. centrale* Israel strain (CP001759, ACIS_01187), and three (Am1, Am2 and Am3) sequence types for *A. marginale* that differed by 1 to 2 amino acid from the *A. marginale* St Maries strain (CP000030, AM090).

The *A. centrale* and *A. marginale* genotypes obtained did not show any specific geographic distribution pattern. The most common *A. centrale* 16S rRNA sequence genotypes were Ac3 and Ac4; Ac4 was common to all of the study sites. The sequence of the Ac4 genotype was identical to that of the *A. centrale* Israel vaccine strain (CP001759, ACIS_01352) over a 1116 bp region of the 16S rRNA gene. The *A. marginale* 16S Am3 genotype was identical to the St Maries strain (CP000030, AMr3) and was found in all the study sites. Notably, a sequence named *A. centrale* Aomori strain (AF283007) that was described by Inokuma *et al*. [4] grouped with "*A. capra*" sequences (KP314237, KX417207, KX417195, KY007144) (Fig. 2). The *groEL* gene was completely conserved, yielding only one *A. centrale* GroEL genotype (Ac1). *A. marginale* had two GroEL genotypes of which Am2 was more common than Am1 and was distributed in all of the study sites except CNP. The *A. centrale* Msp4 Ac4 genotype was found to be present in all study sites except in KNP where Ac1 and Ac3 were present. The *A. centrale* Israel vaccine strain Msp4 sequence (CP001759, ACIS_01187) differed by 1 amino acid from that of Ac4 over the 212 amino acid region that was sequenced. As for *A. marginale*, Msp4

genotype Am3 was found to be the most common and was present in all the study sites (Fig. 1).

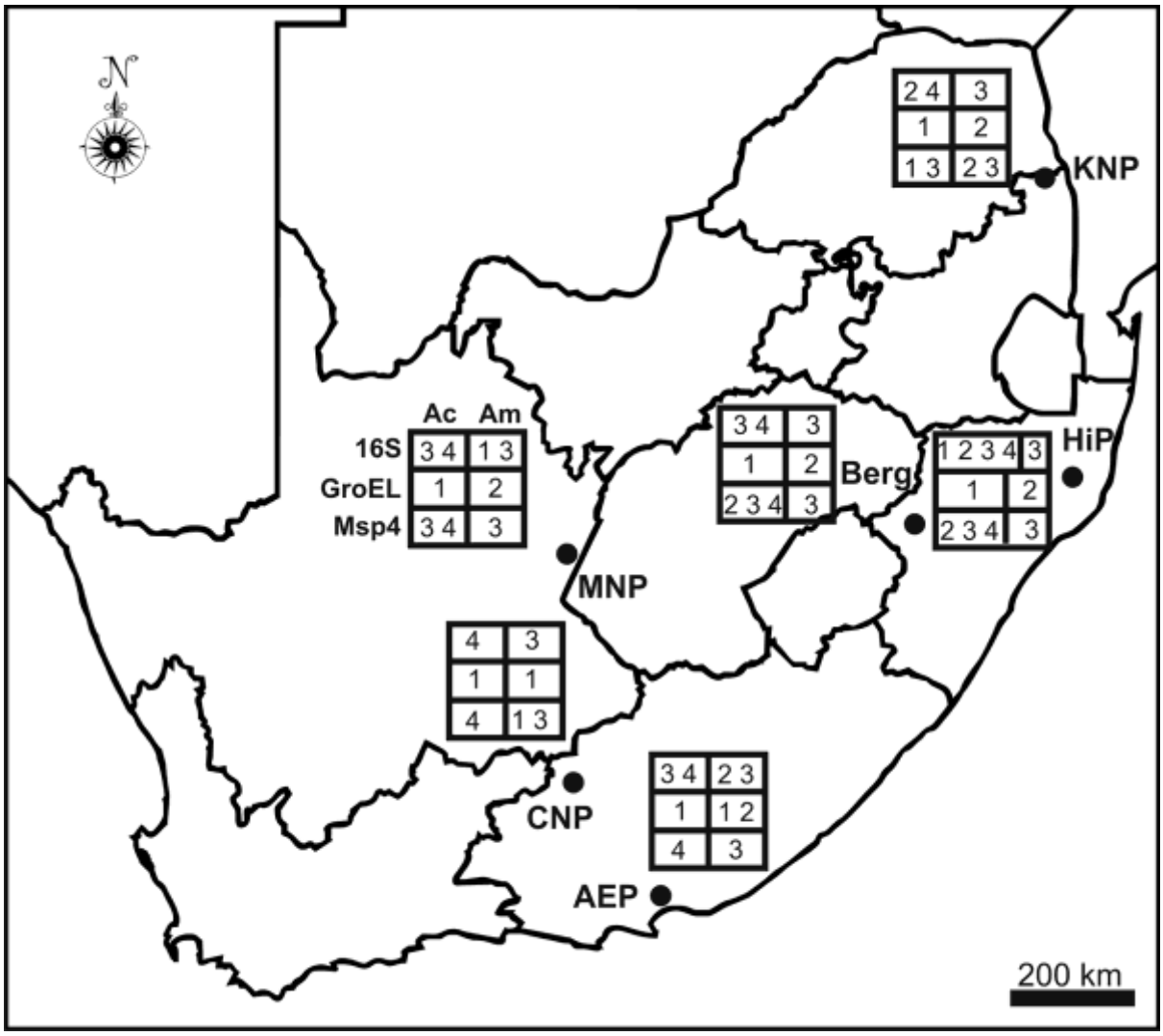

Figure1. Map of South Africa showing the *A. centrale* and *A. marginale* 16S rRNA, GroEL and Msp4 genotypes in selected study areas.

All of the phylogenetic tree topologies obtained using all four tree algorithms were similar, and the maximum likelihood tree was chosen as a representative tree. The trees inferred using 16S rDNA (Fig. 2), GroEL (Fig. 3) and Msp4 (Fig. 4) sequences always grouped *A. marginale* and *A. centrale* into two distinct clades, indicative of a divergence between the two organisms. This was supported by high bootstrap values of 91% (16S rRNA), 92% (GroEL) and 97% (Msp4).

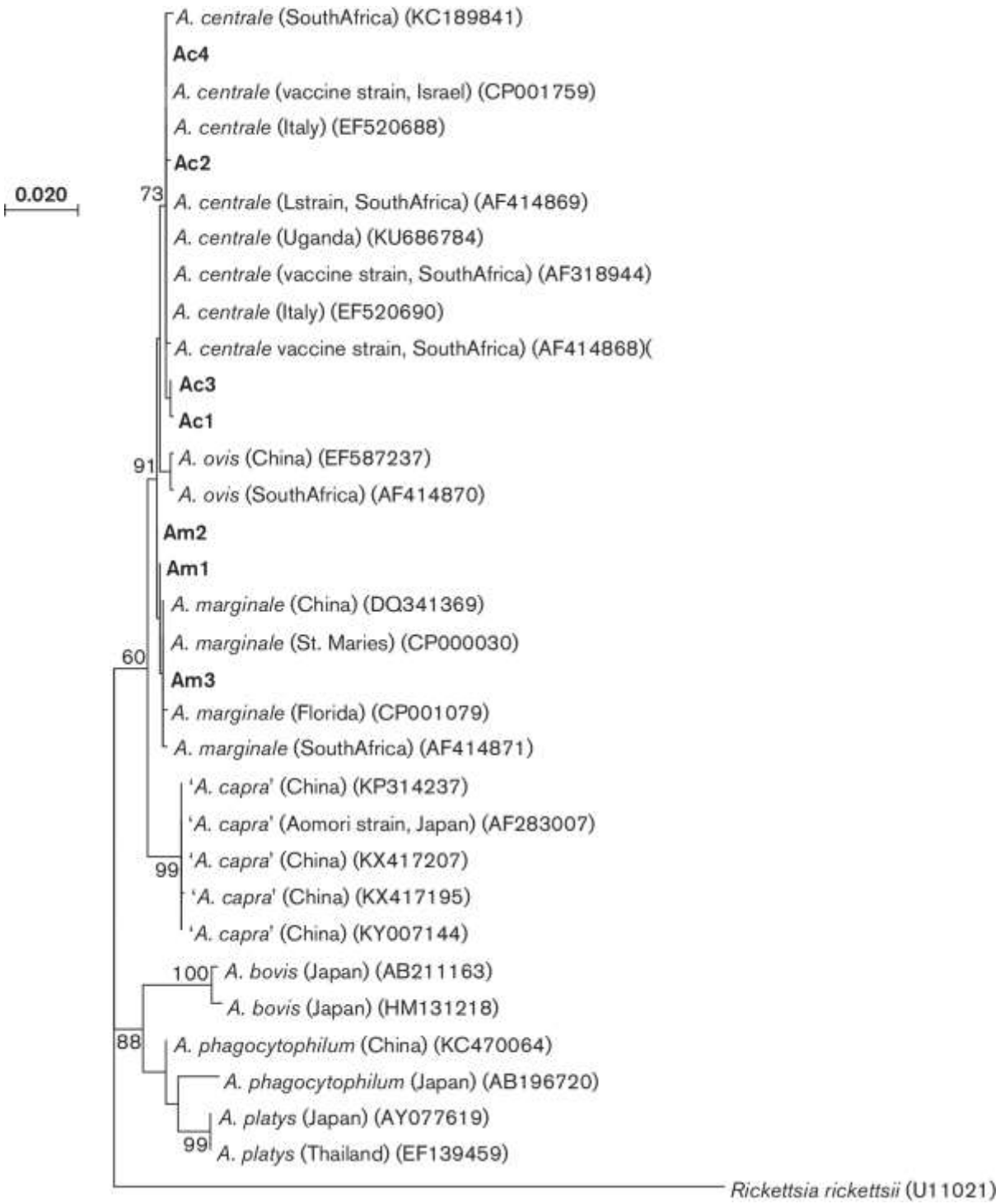

Figure 2. Maximum likelihood tree based on 16S rRNA nucleotide sequences. The tree shows the phylogenetic relationship between *A. centrale, A. marginale* and other *Anaplasma* species. The numbers at the internal nodes represent the percentage of 1000 replicates (bootstrap) for which the same branching patterns were obtained. There were a total of 1108 positions in the final dataset. As some sequences obtained in this study were identical to each other we have coded them with a genotype name (i.e. "Am1") and used a single representative of the genotype to construct the tree. Genotype representation is provided in Table 3.

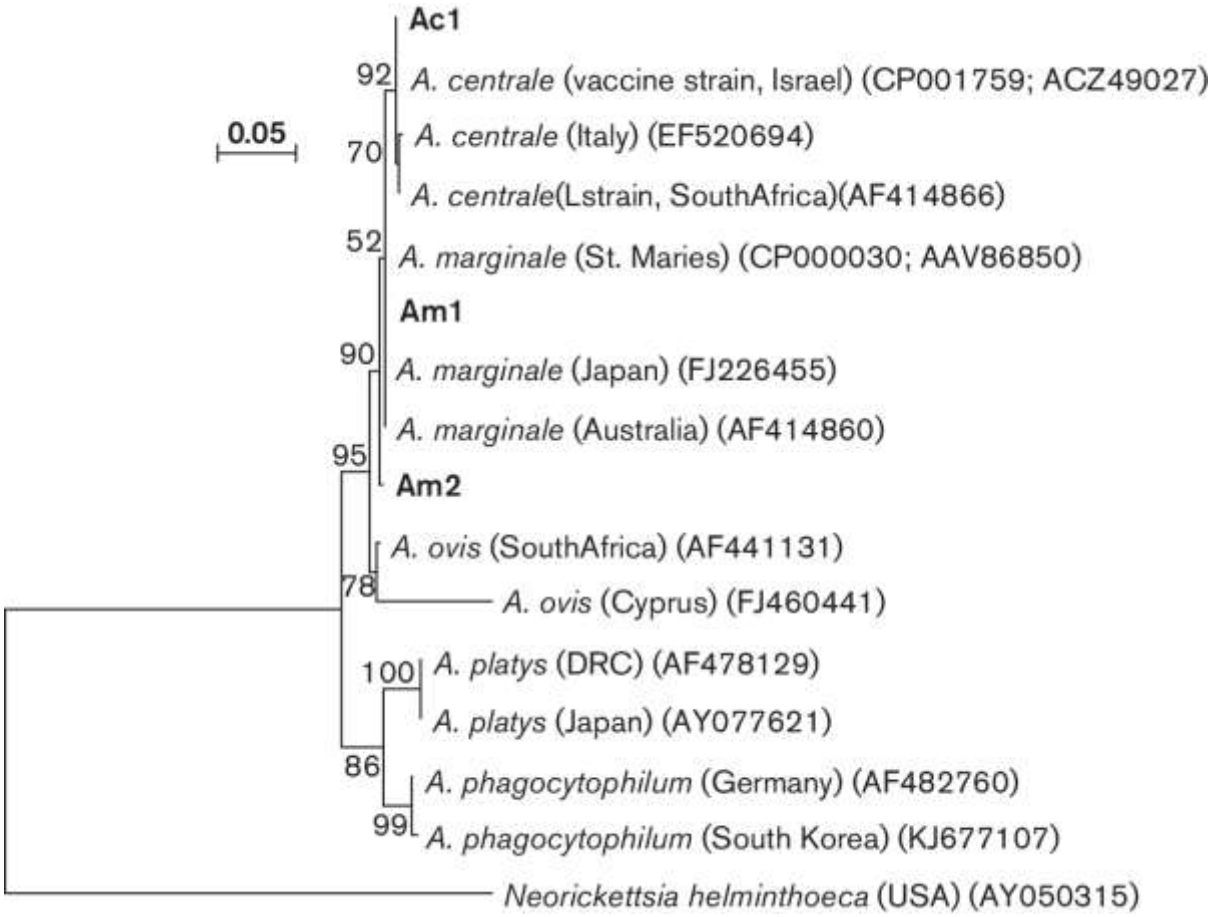

Figure 3. Maximum likelihood tree based on GroEL deduced amino acid sequences. The tree shows the phylogenetic relationship between *A. centrale*, *A. marginale* and other *Anaplasma* species. The numbers at the internal nodes represent the percentage of 1000 replicates (bootstrap) for which the same branching patterns were obtained. There were a total of 488 positions in the final dataset. As some sequences obtained in this study were identical to each other we have coded them with a genotype name (i.e. "Am1") and used a single representative of the genotype to construct the tree. Genotype representation is provided in Table 3.

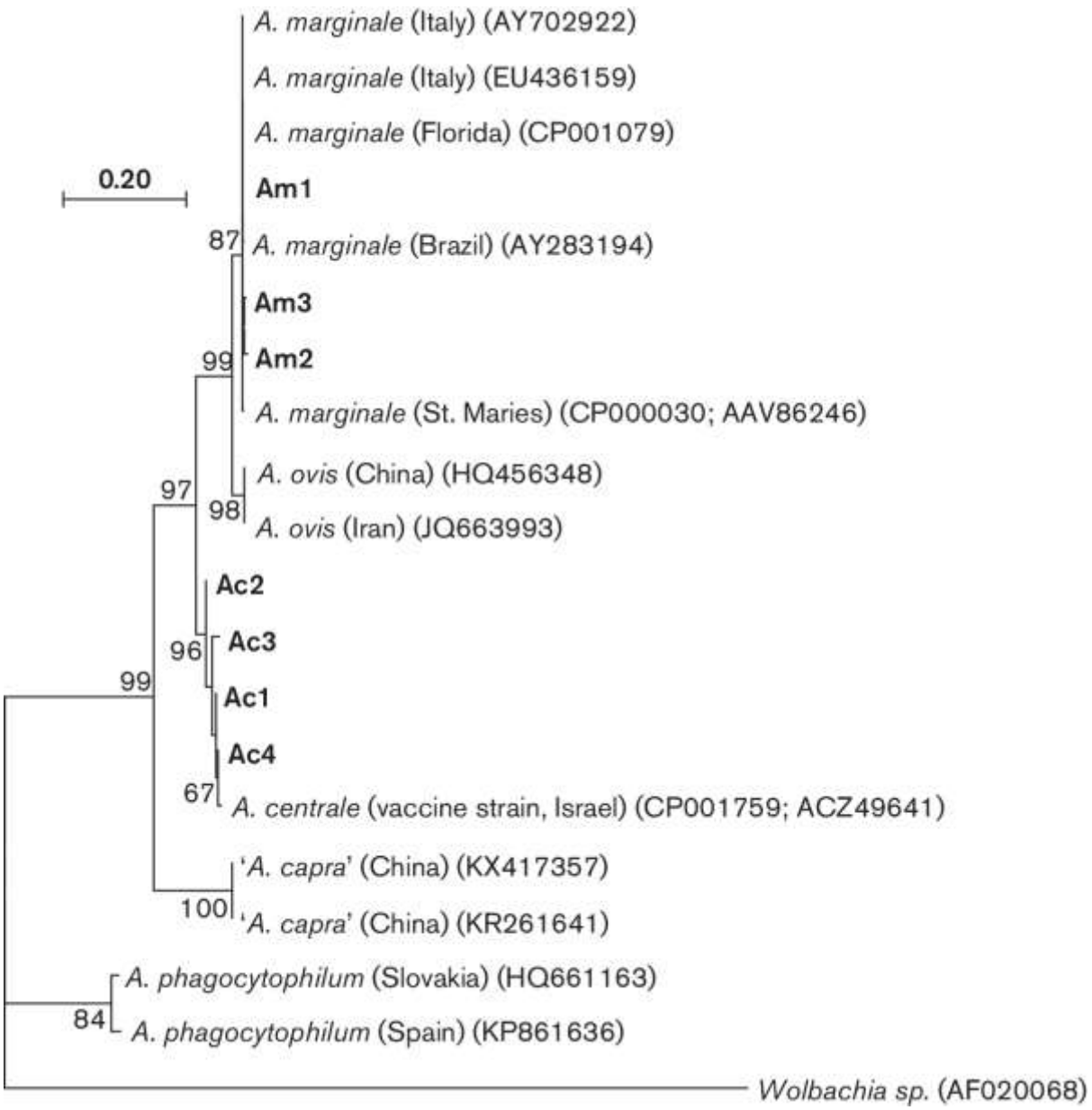

Figure 4. Maximum likelihood tree based on Msp4 deduced amino acid sequences. The tree shows the phylogenetic relationship between *A. centrale, A. marginale* and other *Anaplasma* species. The numbers at the internal nodes represent the percentage of 1000 replicates (bootstrap) for which the same branching patterns were obtained. There were a total of 198 positions in the final dataset. All positions containing gaps and missing data were eliminated. As some sequences obtained in this study were identical to each other we have coded them with a genotype name (i.e. "Am1") and used a single representative of the genotype to construct the tree. Genotype representation is provided in Table 3.

DISCUSSION

The *A. centrale* Israel vaccine strain 16S rRNA, GroEL and Msp4 genotype sequences (Ac4, Ac1 and Ac4, respectively) were found in all of the study sites, with the exception of Msp4 Ac4 that was absent in the KNP. This is in concordance with our previous findings based on *msp1aS* that the *A. centrale* vaccine strain is found widely distributed across South Africa and in animals that do not have a history of vaccination [20].

The pattern of two distinct clades for *A. marginale* and *A. centrale* that was observed from the phylogenetic trees was in concordance with previous findings: Inokuma *et al*. [4], Lew *et al*. [30], Liu *et al*. [31], Carelli *et al*. [10], Yang *et al*., [32] and Sisson *et al*. [29], to name a few. In a study done by Inokuma *et al*. [4], the authors showed that the 16S rRNA sequence of "*A. centrale* Aomori strain" was related to *A. marginale* by both level-of-similarity (98.08% identical) and distance analysis. They concurred that this "*A. centrale*" is an independent species although closely related to *A. marginale*. These findings were based on the Aomori strain, a Japanese isolate which actually appears to be a novel *Anaplasma* species, "*A. capra*" [33]. Phylogenetic analysis of 16S rRNA revealed that "*A. capra*" sequences clustered together in a clade but grouped separately from other *Anaplasma* species [33]. Interestingly, "*A. capra*" is not a formally recognized species, and is not on the List of Prokaryotic Names with Standing in Nomenclature (www.bacterio.net/anaplasma.html). However, it should be noted that when the Inokuma study was done, there were relatively few *A. centrale* 16S rRNA gene sequences available, which led to the misclassification of the Aomori strain as *A. centrale*; and thus some confusion when others have compared their data with this sequence, as relatively speaking, it is somewhat distant from other *A. centrale* 16S rRNA gene sequences.

Carelli and coworkers [10] compared *A. ovis* isolates from China with *A. marginale* and *A. centrale* using 16S rRNA and GroEL peptide sequences; the phylogenetic trees revealed two

distinct clades representative of *A. marginale* and *A. centrale*. Since their study was focused on the taxonomic placement of the Chinese *A. ovis* isolates, no further taxonomic classification or mention of *A. centrale* was made. In the study done by Lew *et al*. [30], 16S rDNA analysis grouped *A. marginale* and *A. centrale* separately; however, it could not delineate *A. ovis* isolates from either *A. marginale* or *A. centrale*. In contrast, *A. ovis* grouped separately in our study, although with very low bootstrap support (28%). Our results were in concordance with the studies of Liu *et al*. [31] and Carelli *et al*. [10] who also found that phylogenetic analyses based on 16S RNA sequences resolved *A. ovis*, *A. marginale* and *A. centrale* into separate groups. Furthermore, Lew *et al*. [30] concluded that GroEL sequences were more reliable for phylogenetic inferences of the species of the erythrocytic *Anaplasma* species (*A. centrale, A. ovis* and *A. marginale*). Our GroEL results were in concordance with those of Lew *et al*. [30] who showed that the differences between *A. marginale* and *A. ovis* were more prominent than those between *A. marginale* and *A. centrale.* This was also consistent with differences previously demonstrated by RFLP and monoclonal antibody studies [29, 35, 36, 37]. Recently Sisson *et al*. [29], used *groEL* to characterise *A. centrale* strains, amplifying *A. centrale groEL* nucleotide sequences from DNA of buffalo collected in KNP. This study confirmed that the *groEL* gene is conserved and can discriminate *A. centrale* strains from *A. marginale* strains; however, phylogenetic analysis of *A. centrale groEL* sequences grouped South African sequences separately from other *A. centrale* strain sequences from Italy and Australia – this is likely due to the very limited number of samples represented in their analysis. The authors used an 881 bp fragment of the *groEL* gene sequence for their phylogenetic analysis. When these nucleotide sequences were translated and added to our GroEL deduced amino acid sequences, they formed one clade with our sequences and those from other countries (i.e. the GroEL Ac1 genotype).

As for phylogenetic trees inferred from Msp4 sequence data, most publications focus on the description of *A. phagocytophilum* isolates [38, 39] with no statements made about the taxonomic position of *A. centrale*. However, it remains clear from these published phylogenetic trees that *A. marginale* and *A. centrale* always group into two distinct clades.

A. centrale is known to be closely related to *A. marginale* based on morphological, protein structural and immunological studies. The size and location of *A. marginale* and *A. centrale* organisms within red-blood cells have traditionally been used to differentiate morphologically between these two organisms. *Anaplasma marginale* is situated on the margins of the red blood cells, appearing as deeply stained "points", round in shape ranging from 0.3 to 1.0 µm in diameter [40], while *A. centrale* is situated towards the centre of the red blood cells, with size variation from 0.4 - 0.95 µm in diameter [41, 42].

A. marginale and *A. centrale* are antigenically related, sharing immunodominant epitopes that play a role in the protection induced by *A. centrale* [43]. However, the protection is partial and varies with *A. marginale* genotype. This is because there is a lower degree of conservation between the deduced amino acid sequences of surface proteins of *A. centrale* and *A. marginale* strains (72.4%) than between any two *A. marginale* strains (95.1%) [44].

Although *A. marginale* and *A. centrale* are detected in similar hosts (i.e. buffalo, black wildebeest, blue wildebeest, eland, waterbuck and cattle) [20], the tick vector and/or the biology of transmission of *A. centrale* appears to differ from that of *A. marginale* sensu stricto. Biological transmission of *A. marginale* is effected by 20 tick species around the world [45], and five tick species have been implicated in the transmission *A. marginale* in South Africa: *Rhipicephalus microplus, R. decoloratus, R. evertsi evertsi, R. simus* and *Hyalomma rufipes* [46]. *Anaplasma centrale* was thought not to be tick transmissible [47], however, experimental

transmission of *A. centrale* by *R. simus* and *D. andersoni* has been demonstrated [48, 49]. Work done in the *D. andersoni* model demonstrated that *A. centrale* infects the midgut and salivary gland at similar rates to *A. marginale*, but *A. centrale* was not transmitted when only a few ticks were used in transmission experiments. Further analysis demonstrated that *A. centrale* resided in a different subcellular location in the salivary gland and was secreted into the saliva at a much lower rate than *A. marginale*; when tick numbers were dramatically increased to compensate for the low pathogen load, transmission was achieved [49]. These two transmission studies are the only successful transmissions of *A. centrale* on record amongst a myriad of failed transmission attempts.

Genomic comparisons of the two organisms have also revealed that the two organisms are divergent [50, 51]: *A. marginale* sensu stricto strains have closed core, highly syntenic genomes [52], while the *A. centrale* genome exhibits a marked lack of synteny with sensu stricto strains. The genome of *A. marginale* is comprised of 949 protein encoding genes and 16 pseudogenes, while that of *A. centrale* is comprised of 925 protein encoding genes and 19 pseudogenes. The genome of *A. marginale* contains 18 putative genes that are absent in *A. centrale,* while the *A. centrale* genome contains 10 putative genes that are absent in *A. marginale* [44]. The *A. centrale* genome also revealed the presence of a homolog of *msp1α,* a gene that was thought to be absent from *A. centrale*. The *A. centrale* homolog, *msp1aS*, was used in a recent study to genotype strains of *A. centrale* in the same manner that *msp1α* is used to genotype strains of *A. marginale* [20]. The *A. centrale* Msp1aS repeats are longer (~51 amino acids in length) than the *A. marginale* Msp1a repeats (28-29 amino acids in length) and there is no sequence similarity in the repeat regions of these proteins, although the carboxy-terminus of the protein has approximately 30% amino acid identity. This genotyping analysis provides clear distinction between *A. marginale* and *A. centrale*. The results further demonstrate the diversity of *A.*

centrale strains from cattle and wildlife hosts from South Africa, also highlighting the significance of wildlife as reservoir hosts for *A. centrale*.

In conclusion, the phylogenetic analysis presented here, together with differences in genome architecture, *msp1α*/*msp1aS* gene sequence, and the biology of tick transmissibility, provide sufficient divergence between *A. centrale* and *A. marginale* to classify them as separate species. Therefore, there are seven officially recognized species of *Anaplasma*: *A. marginale, A. centrale, A. ovis, A. phagocytophilum, A. platys, A. bovis,* and *A. caudatum.* "*A. capra*" represents a potential eighth species and is discussed in this article as a novel *Anaplasma* species whose name has not yet been officially recognized.

***Emended description of* Anaplasma centrale**

***Anaplasma centrale* (*ex* Theiler 1911) Ristic and Kreier 1984**

Etymology: *cen.tra'le.* L. neut. adj. *centrale*, in the middle, central, referring to the location of the organism within erythrocytes.

Synonym: "*Anaplasma marginale* variety *centrale*" Theiler, 1911. Other names that have been used in the literature include: *Anaplasma marginale centrale*; *Anaplasma marginale* subspecies *centrale*; *Anaplasma marginale* vaccine strain.

Anaplasma centrale is a separate species from *A. marginale*. It is centrally located in the red blood cell. It has been detected in several African countries; i.e. South Africa [9, 53, 54], Uganda [55] and Botswana [56]. It has also been detected in Italy [10], Spain [57] and Brazil [58], usually causing mild disease in cattle. It is prevalent in wildlife, especially in the African buffalo (*Syncerus caffer*). The genome of *A. centrale* contains a homolog of *A. marginale msp1α*, *msp1aS*, which serves as a genetic marker for *A. centrale* strains. *A. centrale* is

antigenically related to *A. marginale*, they both share specific immunological epitopes that enables *A. centrale* to provide limited protection against *A. marginale.*

The mol % G +C of DNA: 50% [50]

Type strain: Israel

Genbank accession number *A. centrale* (genome): CP0001759

AUTHOR STATEMENTS:

Funding

This work is based on the research supported by the National Research Foundation (grant 81840 to Nicola Collins and grants 76529 and 92739 to Marinda Oosthuizen); Technology Innovation Agency, Tshwane Animal Health Cluster (grant TAHC12-00037 to Marinda Oosthuizen) and University of Pretoria Institutional Research Theme (Genomics) (to Marinda Oosthuizen).

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. Any opinion, finding and conclusion or recommendation expressed in this material is that of the author(s) and the NRF does not accept any liability in this regard.

Acknowledgements

We thank B. Christoff from Plaas Hongerspoort and A. Shepherd from Tugela Veterinary Clinic, Bergville, KwaZulu-Natal, who helped us collect blood samples from cattle. Wildlife samples were obtained from the South African National Parks SANParks Biobank and from the Wildlife Biological Resource Center (WBRC) and Biobank SA under the auspices of the

National Zoological Gardens (NZG) of South Africa and Johannesburg Zoo. We thank Dave Cooper for providing buffalo blood samples from Hluhluwe-iMfolozi Park.

Conflicts of interest

All authors declare that there is no conflict of interest.

Ethical statement

The study was approved by the Animal Ethics Committee of the University of Pretoria, South Africa (V085-14), and permission to use wildlife samples was given by SANParks Biobank under reference number LARBJ1118 Conservation Genetics, by the WBRC, and by Biobank SA under the auspices of the NZG of South Africa and the Johannesburg Zoo with project number NZG/P13/05. Collection of cattle samples was approved by the Department of Agriculture, Forestry and Fisheries under section 20 of the Animal Diseases Act of 1984 with reference 12/11/1/1.

REFERENCES

1. **Salmon DE, Smith T.** Infectious diseases of cattle: southern cattle fever (Texas fever), In: *Special report on diseases of cattle and on cattle feeding*. USDA Bureau of Animal Industry. Government Printing Office, Washington. 1896; pp. 428-438.
2. **Theiler A**. *Anaplasma marginale*: the marginal points in the blood of cattle suffering from a specific disease. In: *Report to the Government, Transvaal, South Africa*. Veterinary Bacteriology, Dept. of Agriculture 1908-9. 1910. pp. 7-64.
3. **Theiler A.** Further investigations into anaplasmosis of South African cattle. In: *First Report of the Director of Veterinary Research*, Union of South Africa. 1911. pp.7-46.

4. Inokuma H, Yutaka T, Kamio T, Raoult D, Brouqui P. Analysis of the 16S rRNA gene sequence of *Anaplasma centrale* and its phylogenetic relatedness to other Ehrlichiae. *Clin Diagn Lab Immunol* 2001;8:241–244.

5. Shkap V, Molad T, Brayton KA, Brown WC, Palmer GH. Expression of major surface protein 2 variants with conserved T-cell epitopes in *Anaplasma centrale* vaccinates. *Infect Immun* 2002;70:642–648.

6. Lew AE, Bock RE, Minchin CM, Masaka S. A *msp1alpha* polymerase chain reaction assay for specific detection and differentiation of *Anaplasma marginale* isolates. *Vet Microbiol* 2002;86:325–335.

7. Kocan KM, de la Fuente J, Guglielmone AA, Melendez RD. Antigens and alternatives for control of *Anaplasma marginale* infection in cattle. *Clin Microbiol Rev* 2003;16:698–712.

8. Carelli G, Decaro N, Lorusso A, Elia G, Lorusso E *et al*. Detection and quantification of *Anaplasma marginale* DNA in blood samples of cattle by real-time PCR. *Vet Microbiol* 2007;124:107–114.

9. Mtshali M, de La Fuente J, Ruybal P, Kocan K, Vicente J *et al*. Prevalence and genetic diversity of *Anaplasma marginale* strains in cattle in South Africa. *Zoonoses Public Health* 2007;54: 23–30.

10. Carelli G, Decaro N, Lorusso E, Paradies P, Elia G *et al*. First report of bovine anaplasmosis caused by *Anaplasma centrale* in Europe. *Ann N Y Acad Sci* 2008;1149:107–110.

11. Decaro N, Carelli G, Lorusso E, Lucente MS, Greco G *et al*. Duplex real-time polymerase chain reaction for simultaneous detection and quantification of *Anaplasma marginale* and *Anaplasma centrale*. *J Vet Diagn Invest* 2008;20:606–611.

12. Rymaszewska A, Grenda S. Bacteria of the genus *Anaplasma*–characteristics of *Anaplasma* and their vectors: a review. *Vet Med* 2008;53:573–584.

13. Aubry P, Geale DW. A review of bovine anaplasmosis. *Transbound Emerg Dis* 2011;58:1–30.

14. Bell-Sakyi L, Palomar AM, Bradford EL, Shkap V. Propagation of the Israeli vaccine strain of *Anaplasma centrale* in tick cell lines. *Vet Microbiol* 2015;179:270–6.

15. Ristic M. Anaplasmosis. In: Weinman D & Ristic D (editors). *Infectious Blood Diseases of Man and Animals*. Academic Press, New York, USA.11. 1968; pp. 473-542.

16. International Journal of Systematic and Bacteriology. Validation of the Publication of New Names and New Combinations Previously Effectively Published Outside the IJSB List No.15. *IUMS* 1984;34:355–357.

17. Ristic M, Kreier JP. Family. III. *Anaplasmataceae* Philip 1957, 980^{AL}. In: Krieg NR & Holt JG (editors). *Bergey's Manual of Systematic Bacteriology*. Baltimore: Williams & Wilkins. 1, 1984; pp. 719- 729.

18. Dumler JS, Barbet AF, Bekker CP, Dasch GA, Palmer GH *et al.* Reorganization of genera in the families *Rickettsiaceae* and *Anaplasmataceae* in the order *Rickettsiales*: unification of some species of *Ehrlichia* with *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia*, descriptions of six new species combinations and designation of *Ehrlichia equi* and 'HGE agent' as subjective synonyms of *Ehrlichia phagocytophila*. *Int J Syst Evol Microbiol* 2001;51:2145–2165.

19. Brayton KA, Dark MJ, Palmer GH. *Anaplasma*. In: Nene V & Kole C (editors). *Genome Mapping and Genomics in Animal-Associated Microbes*. Berlin/Heidelberg, Germany: Springer; 2009. pp. 85-116.

20. Khumalo ZTH, Catanese HN, Liesching N, Hove P, Collins NE *et al.* Characterisation of *Anaplasma marginale* subsp. *centrale* strains by use of *msp1aS* genotyping reveals a wildlife reservoir. *J Clin Microbiol* 2016;54:2503–2512.

21. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.

22. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;14:3059–3066.

23. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 1999;41:95–98.

24. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;33:1870–1874.

25. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985;39:783–791.

26. Ronquist F, Huelsenbeck JP. Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003;19:1572–1574.

27. Abascal F, Zardoya R, Posada D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*. 2005;21:2104–2105.

28. Posada D, Crandall KA. Modeltest: testing the model of DNA substitution. *Bioinformatics* 1998;14:817–818.

29. Sisson D, Hufschmid J, Jolles A, Beechler B, Jabbar A. Molecular characterisation of *Anaplasma* species from African buffalo (*Syncerus caffer*) in Kruger National Park, South Africa. *Ticks Tick Borne Dis* 2017;8:400–406.

30. Lew AE, Gale KR, Minchin CM, Shkap V, de Waal TD. Phylogenetic analysis of the erythrocytic *Anaplasma* species based on 16S rDNA and GroEL (HSP60) sequences of *A. marginale, A. centrale,* and *A. ovis* and the specific detection of *A. centrale* vaccine strain. *Vet Microbiol* 2003;92:145–160.

31. Liu Z, Luo J, Bai Q, Ma M, Guan G *et al*. Amplification of 16S rRNA genes of *Anaplasma* species in China for phylogenetic analysis. *Vet Microbiol* 2005;107:145–148.

32. Yang J, Liu Z, Niu Q, Liu J, Han R *et al*. A novel zoonotic *Anaplasma* species is prevalent in small ruminants: potential public health implications. *Parasit Vectors* 2017;10:1–6.

33. Li H, Zheng YC, Ma L, Jia N, Jiang BG *et al*. Human infection with a novel tick-borne *Anaplasma* species in China: a surveillance study. *Lancet Infect Dis* 2015;15:663–670.

34. List Of Prokaryotic Names With Standing In Nomenclature. Last updated in 2015. http://www.bacterio.net/anaplasma.html. Last viewed on the 27 March 2018.

35. Palmer GH, Barbet AF, Musoke AJ, Katende JM, Rurangirwa F *et al*. Recognition of conserved surface protein epitopes on *Anaplasma centrale* and *Anaplasma marginale* isolates from Israel, Kenya, and the United States. *Int J Parasitol* 1988;18:33–38.

36. Visser ES, Ambrosio RE, de Waal DT. An *Anaplasma centrale* DNA probe that differentiates between *Anaplasma ovis* and *Anaplasma marginale* DNA. *Vet Microbiol* 1991;28:313–325.

37. Ngeranwa JJN, Venter EH, Penzhorn BL, Soi RK, Mwanzia J *et al*. Characterisation of *Anaplasma* isolates from eland (*Taurotragus oryx*) pathogenicity in cattle and sheep and DNA profiles analysis. *Vet Parasitol* 1998;74:109–122.

38. de la Fuente J, Massung RF, Wong SJ, Chu FK, Lutz H *et al*. Sequence analysis of the *msp4* gene of *Anaplasma phagocytophilum* strains. *J Clin Microbiol* 2005;43:1309–1317.

39. Zhan L, Cao WC, Jiang JF, Zhang XA, Liu YX *et al*. 2010. *Anaplasma phagocytophilum* from rodents and sheep, China. *Emerg Infect Dis* 2010;16:764–768.

40. Bruner DW, Gillespie JH. Hagan's Infectious Diseases of Domestic Animals, Ithaca, New York: Comstock; 1966;5:344.

41. Henning MW. *Animal diseases in South Africa. South Africa*, Central News Agency Ltd. 3rd ed. 1949;398.

42. Waddel AH. Observations on the morphology of *Anaplasma centrale*. *Aust Vet J* 1964;40:53–54.

43. Shkap V, Pipano E, Mcguire TC, Palmer GH. Identification of immunodominant polypeptides common between *Anaplasma centrale* and *Anaplasma marginale*. *Vet Immunol Immunopathol* 1991;29:31–40.

44. Agnes JT, Brayton KA, Lafollett M, Norimine J, Brown WC *et al*. Identification of *Anaplasma marginale* Outer Membrane Protein Antigens Conserved between *A. marginale* sensu stricto strains and the Live *A. marginale* subsp. *centrale*. *Vaccine*. 2011;73:1311–1318.

45. Kocan KM, de la Fuente J, Blouin EF, Coetzee JF, Ewing SA. The natural history of *Anaplasma marginale*. *Vet Parasitol* 2010;167:95–107.

46. de Waal DT. Anaplasmosis control and diagnosis in South Africa. *Ann N Y Acad Sci* 2000;916:474–483.

47. Shkap V, Kocan K, Molad T, Mazuz M, Leibovich B *et al*. Experimental transmission of field *Anaplasma marginale* and the *A. centrale* vaccine strain by *Hyalomma excavatum*, *Rhipicephalus sanguineus* and *Rhipicephalus* (*Boophilus*) *annulatus* ticks. *Vet Microbiol* 2009;134:254–260.

48. Potgieter FT, Van Rensburg L. Tick transmission of *Anaplasma centrale*. *Onderstepoort J Vet Res* 1987;54:5–7.

49. Ueti MW, Knowles DP, Davitt CM, Scoles GA, Baszler TV *et al*. Quantitative differences in salivary pathogen load during tick transmission underlie strain-specific variation in transmission efficiency of *Anaplasma marginale*. *Infect Immun* 2009;77:70-75.

50. Herndon DR, Palmer GH, Shkap V, Knowles DP Jr, Brayton KA. Complete genome sequence of *Anaplasma marginale* subsp. *centrale. J Bacteriol* 2010. 192:379–380.

51 Brayton KA, Kappmeyer LS, Herndon DR, Dark MJ, Tibbals DL *et al*. Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins. *Proc Natl Acad Sci* 2005;102:844–849.

52. Dark MJ, Herndon DR, Kappmeyer LS, Gonzales MP, Nordeen E *et al*. Conservation in the face of diversity: multistrain analysis of an intracellular bacterium. *BMC Genomics* 2009;10:16.

53. Debeila EM. Occurrence of *Anaplasma* and *Ehrlichia* species in African buffalo (*Syncerus caffer*) in Kruger National Park and Hluhluwe-iMfolozi Park in South Africa. MSc dissertation. Department of Veterinary Tropical Diseases, University of Pretoria, Pretoria, South Africa. 2012.

54. Mutshembele AM, Cabezas-Cruz A, Mtshali MS, Thekisoe OMM, Galindo RC *et al.* Epidemiology and evolution of the genetic variability of *Anaplasma marginale* in South Africa. *Ticks Tick Borne Dis* 2014;5:624–631.

55. Oura CAL, Tait A, Asiimwea B, Lubega GW, Weir W. Haemoparasite prevalence and *Theileria parva* strain diversity in Cape buffalo (*Syncerus caffer*) in Uganda. *Vet Parasitol* 2011;175:2–9.

56. Eygelaar D, Jori F, Mokopasetso M, Sibeko KP, Collins NE *et al.* Tick-borne haemoparasites in African buffalo (*Syncerus caffer*) from two wildlife areas in Northern Botswana. *Parasit Vectors* 2015;8:1–11.

57. Palomar AM, Portillo A, Santibáñez P, Mazuelas D, Roncero L *et al.* Detection of tick-borne *Anaplasma bovis*, *Anaplasma phagocytophilum* and *Anaplasma centrale* in Spain. *Med Vet Entomol* 2015;29:349–353.

58. Joazeiro AC, Martins J, Masuda A, Seixas A, Junior IDSV. A PCR for Differentiate between *Anaplasma marginale* and *A. centrale*. *Acta Sci Vet* 2015;43:1270.

59. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* 1991;173:697–703.