Genomic Characteristics of Emerging Intraerythrocytic *Anaplasma capra* and High Prevalence in Goats, China

Zhe-Tao Lin, Li-Feng Du, Ming-Zhu Zhang, Xiao-Yu Han, Bai-Hui Wang, Jiao Meng, Fu-Xun Yu, Xiao-Quan Zhou, Ning Wang, Cheng Li, Xiao-Yang Wang, Jing Liu, Wan-Ying Gao, Run-Ze Ye, Luo-Yuan Xia, Yi Sun, Na Jia, Jia-Fu Jiang, Lin Zhao,[1] Xiao-Ming Cui,[1] Lin Zhan,[1] Wu-Chun Cao[1]

Anaplasma capra is an emerging tickborne human pathogen initially recognized in China in 2015; it has been reported in ticks and in a wide range of domestic and wild animals worldwide. We describe whole-genome sequences of 2 *A. capra* strains from metagenomic sequencing of purified erythrocytes from infected goats in China. The genome of *A. capra* was the smallest among members of the genus *Anaplasma*. The genomes of the 2 *A. capra* strains contained comparable G+C content and numbers of pseudogenes with intraerythrocytic *Anaplasma* species. The 2 *A. capra* strains had 54 unique genes. The prevalence of *A. capra* was high among goats in the 2 endemic areas. Phylogenetic analyses revealed that the *A. capra* strains detected in this study were basically classified into 2 subclusters with those previously detected in Asia. Our findings clarify details of the genomic characteristics of *A. capra* and shed light on its genetic diversity.

Anaplasma capra is an emerging tickborne zoonotic pathogen in the genus *Anaplasma*, family Anaplasmataceae, and was initially identified in blood samples from asymptomatic goats (*Capra aegagrus hircus*) and a febrile human patient with tick-bite history in China in 2015 (*1*). The patient infected with *A. capra* had fever, headache, malaise, dizziness, myalgia, gastrointestinal symptoms, rash, lymphadenopathy, and abnormalities in cerebrospinal fluid pleocytosis and hepatic aminotransferase. Since then, *A. capra* has been detected in various domestic animals (e.g., goats, sheeps, cattle, yaks, and dogs) (*2*–*5*) and wild animals (e.g., takins, muntjacs, water deer, musk deer, onagers, serows, and brown hares) (*6*–*10*), and in a wide range of ticks (e.g., *Ixodes persulcatus*, *Haemaphysalis longicornis*, *H. qinghaiensis*, *Dermacentor abaensis*, *D. nuttalli*, *and Rhipicephalus microplus* [*1*,*11*–*14*]) across China and around the world (*2*,*7*–*10*,*15*,*16*), posing a potential threat to the health of humans and animals.

Members of the family Anaplasmataceae have complex life cycles involving vertebrate hosts and hematophagous ticks, many of which have emerged as human pathogens. The genus *Anaplasma* was proposed according to the phylogenetic analyses based on 16S rRNA and *groEL* sequences (*17*) and initially encompassed 6 species: *A. phagocytophilum*, *A. marginale*, *A. centrale*, *A. ovis*, *A. platys*, and *A. bovis*. Subsequently, 2 candidate novel species (*A. capra* and *A. odocoilei*) and other unclassified genovariants (*1*,*18*–*20*) were included in the List of Prokaryotic Names with Standing in Nomenclature (https://www.bacterio.net) pending validation. To date, 5 *Anaplasma* species have been known to infect humans: *A. phagocytophilum*, *A. capra*, *A. ovis*, *A. platys*, and *A. bovis* (*21*). Since the *A. marginale* genome sequence was reported in 2005 (*22*), a total of 24 *A. marginale* genomes (*23*), 32 *A. phagocytophilum* genomes (*24*,*25*), 1 *A. centrale* genome

Author affiliations: State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China (Z.-T. Lin, L.-F. Du, M.-Z. Zhang, X.-Y. Han, Y. Sun, N. Jia, J.-F. Jiang, X.-M. Cui, W.-C. Cao); Institute of EcoHealth, School of Public Health, Shandong University, Jinan, China (L.-F. Du, M.-Z. Zhang, B.-H. Wang, N. Wang, C. Li, X.-Y. Wang, J. Liu, W.-Y. Gao, R.-Z. Ye, L.-Y. Xia, L. Zhao); National Health Commission Key Laboratory of Pulmonary Immunological Diseases, Guizhou Provincial People's Hospital, Guiyang, China (J. Meng, F.-X. Yu, L. Zhan); Guizhou Provincial Blood Center, Guiyang (X.-Q. Zhou)

DOI: https://doi.org/10.3201/eid2909.230131

[1]These senior authors contributed equally to this article.

(*26*), 2 *A. ovis* genomes (*27*), and 1 *A. platys* genome (*28*) have been sequenced and deposited in GenBank. Although *A. capra* has been extensively detected in ticks and animal hosts worldwide, no genome of the emerging pathogen has been determined so far, which has hindered us from better understanding its genetic features and pathogenesis. Considering *A. capra* is an intraerythrocytic pathogen and abundant in blood samples of host goats (*1*,*29*), we separated erythrocytes from the blood of infected goats to enrich the bacteria and generated the entire genome of *A. capra* using metagenome assembly to promote better understanding of this emerging pathogen, to compare the characteristics of *A. capra* genomes with previously published genomes of other *Anaplasma* and related species, and to evaluate intraspecies genetic diversity of *A. capra* in different geographic locations and tick species across China.

Materials and Methods

Sample Collection and Preparation

We collected EDTA blood samples from 3 flocks of goats in Shandong Province and a flock of goats in Guizhou Province, China (Appendix Figure 1, https://wwwnc.cdc.gov/EID/article/29/9/23-0131-App1.pdf), during September 2021–July 2022. Meanwhile, we prepared blood smears for some goats. We collected host-seeking ticks in the same areas where the infected goats lived by dragging white flags over vegetation. An entomologist (Y.S.) identified all ticks to the species level and developmental stage. We extracted DNA from each goat blood sample or tick by using a High Pure PCR Template Preparation Kit (Roche, https://www.roche.com) according to the manufacturer's instructions.

PCRs and Sequencing

We conducted a nested PCR specific for the citrate synthase (*gltA*) gene of *A. capra* (Appendix Table 1) to screen all goat blood and tick samples, as previously described (*1*). We amplified all the positive samples for *gltA* by specific PCRs targeting the 16S rRNA, *msp4*, and *groEL* genes of *A. capra* (Appendix Table 1). We sequenced all amplicons to confirm the correctness of PCR results and conducted a SYBR Green–based quantitative PCR (qPCR) targeting different regions of the *gltA* gene by using a specific primer (Appendix Table 1).

Fluorescence In Situ Hybridization

We used fluorescence in situ hybridization (FISH) to observe the *A. capra* on blood smears. We designed the probe on the basis of the 16S rRNA full-length sequence of *A. capra* (Appendix Table 2) and labeled it with Quasar 570. We resuspended the pooled FISH probes in a final concentration of 25 μmol/L in RNase-free storage buffer, which we protected from light and stored at −20°C. We performed FISH on the prepared blood smear with a commercial kit (Biosearch Technologies, https://www.biosearchtech.com), according to the manufacturer's instructions.

Enrichment of *A. capra* for Genomic Sequencing

We separated erythrocytes from infected goats by conducting gradient centrifugation using cell separation solution (Eppendorf, https://www.eppendorf.com) for 20 min at 200 × *g* at 4°C. Then, we added 4 times volume of precooled (4°C) erythrocyte lysis buffer (Solarbio, http://www.solarbio.net) to the isolated erythrocytes by gentle pipetting to ensure adequate mixing. After placing the lysis solution at 4°C for 10 min, we centrifuged the solution at 350 × *g* for 10 min to remove residual blood cells. After that treatment, we maximally removed the host DNA in samples. Finally, we centrifuged the supernatant at 20,000 × *g* at 4°C for 30 min. We resuspended the pooled *A. capra* for DNA extraction by using the High Pure PCR Template Preparation Kit (Roche). We then constructed a sequencing library by using the AxyPrep MAG PCR Clean Up Kit (Fisher Scientific, https://www.fishersci.com) for an MGI sequencing set (https://en.mgi-tech.com). We prepared the sequencing library according to the Whole Genome Sequencing Library Preparation Protocol (MGI). We sequenced the paired-end libraries with a read length of 2 × 150 bp on a DNBseq-T7 platform at Grandomics Gene Technology Beijing Co. Ltd (Beijing, China).

Genome Assembly and Comparative Analyses

We mapped the clean reads to the goat (*Capra hircus*) reference genome (GenBank accession no. GCF_001704415) by using SAMtools 1.14 (*30*) to discard host-derived reads. We de novo assembled contigs from the unmapped reads by using metaSPAdes 3.15.3 (*31*). We performed contig binning by using MetaBAT 2.15 (*32*) and evaluated assembly quality by using CheckM version 1.1.3 in linage_wf mode, which searches for universal single-copy marker genes and deduces completeness and contamination on the basis of presence and absence of these genes (*33*). We generated G+C content, genome completeness, and annotation information and depicted them by using an approach described previously (*34*,*35*). We estimated average nucleotide identity (ANI) and DNA–DNA

hybridization (DDH) by using fastANI 1.32 (*36*) and GGDC (https://ggdc.dsmz.de/ggdc.php).

Phylogenetic Analyses

We deposited in GenBank the results of the phylogenetic analysis of the whole genomes of the 2 *A. capra* strains and all the genomes of *Anaplasma* species by using Orthofinder 2.5.4 (*37*), after eliminating the poorly aligned positions and divergent regions by using Gblocks 0.91b. We aligned trimmed sequence by using Muscle 5.1 (R.C. Edgar, unpub. data, https://doi.org/10.1101/2021.06.20.449169) and constructed the phylogenetic tree by using iqtree 2.2.0.3 (*38*). Furthermore, we conducted phylogenetic analyses on *A. capra gltA*, *groEL*, 16S rRNA, and *msp4* genes obtained from infected goats and ticks by using the maximum-likelihood method in MEGA11 (*39*).

Functional Analysis of Predicted Genes

To find difference in the Kyoto Encyclopedia of Genes and Genomes (KEGG) between the 2 strains of *A. capra* and other species in the genus *Anaplasma*, were annotated orthogroup sequences by using KOfam 1.4.0 (*40*) and illustrated them using a Venn diagram. We used the software eggNOG-Mapper 2.1.7 to determine the Clusters of Orthologous Group (COG) categories for protein encoding regions (*41*).

Results

Forty-three (59.7%) of 72 goat blood samples were positive for *gltA* gene of *A. capra*. We chose 2 blood samples (1 from a 2-year-old female goat in Shandong Province and another from a 10-month-old female goat in Guizhou Province) (Appendix Figure 1) for next-generation sequencing because they had high bacterial loads (8.4×10^6 *gltA* gene copies/mL blood for the goat in Shandong Province and 2.0×10^6 *gltA* gene copies/mL blood for the goat in Guizhou Province) as estimated by qPCR (Appendix Table 1). In addition, we visualized *A. capra* by specific FISH in erythrocytes on the blood smear prepared from the goat in Shandong Province for next-generation sequencing (Figure 1).

The metagenome sequencing resulted in >38 million 150-bp clean reads from each sample. Despite primary removing of host DNA, 95.9% and 93.3% of reads in the 2 samples were mapped to the goat genome and discarded. The remaining reads were subsequently de novo assembled into contigs by using the SPAdes 3.15.3 with meta parameters (*31*). The 2 assembled *A. capra* genomes were named *A. capra* str. BIME1 (GenBank accession no. GCA_025628785.1) and *A. capra* str. BIME2 (GenBank accession no. GCA_025628805.1), and had a higher level of completeness (99.79% for BIME1 and 99.36% for BIME2). The genome of *A. capra* was the smallest (≈1.07 Mb) among those in the genus *Anaplasma* and the second smallest genome of the family Anaplasmataceae, just after *Neorickettsia sennetsu* (0.859 Mb) (*24*). The genome sequences of the 2 strains shared 99.89% nucleotide similarity with each other.

We compared the 2 *A. capra* genomes with other representative species strains in the genus *Anaplasma* (Appendix Table 3). The G+C content (48.3% for both) of the 2 *A. capra* genomes was similar to those of *A. ovis*, *A. marginale*, *and A. centrale*, which are all intraerythrocytic pathogens. The *A. capra* genomes yielded a total of 929 and 932 genes, of which 862 and 863, respectively, represented coding sequences. They possessed 37 tRNAs and a complete ribosomal RNA operon, in which the 16S rRNA gene was separated from the 23S-5S rRNA gene pair (Figure 2) as displayed by other members of the order Rickettsiales (*42*). The 2 strains of *A. capra* and other intraerythrocytic *Anaplasma* species, including *A. ovis*, *A. centrale*, and *A. marginale*, contained comparable numbers of pseudogenes that have lost functions owing to mutation accumulation and are observed more frequently in obligate intracellular bacteria where the lost gene functions are compensated by the host cells (*43*). Of note, *A. phagocytophilum* has ≈4-fold more pseudogenes than the other *Anaplasma* species (Appendix Table 3).

The estimated values of ANI and DDH between *A. capra* and other *Anaplasma* species suggested that *A. capra* were distinct from the other species. On the basis of ANI values, *A. capra* str. BIME1 was most similar to *A. marginale*, whereas *A. capra* str. BIME2 was most similar to *A. ovis*. The DDH results revealed that both *A. capra* strains were most close to *A. marginale* (Appendix Table 4). The phylogenetic analysis based on the single copy genes revealed that the 2 *A. capra* strains together occupied a distinct branch and were more closely related to *A. ovis*, *A. marginale*, and *A. centrale* than to *A. phagocytophilum* and *A. platys* in the genus *Anaplasma* (Figure 3, panel A). To explore the gene differences in species in the genus *Anaplasma*, we used Orthofinder (*37*) to identify the homologous genes. All species in the genus *Anaplasma* shared 643 genes in common, and the 2 *A. capra* strains together with other intraerythrocytic *Anaplasma* species (*A. ovis*, *A. centrale*, and *A. marginale*) shared 75 genes that are not present in the other 2 species, *A. phagocytophilum* and *A. platys*. Compared with other members of the genus *Anaplasma*, 14 genes were not possessed by *A. capra*. Of note, a total of 54 genes were only shared by the 2 *A. capra* strains, which had other 14 distinct genes in BIME1 and 10 in

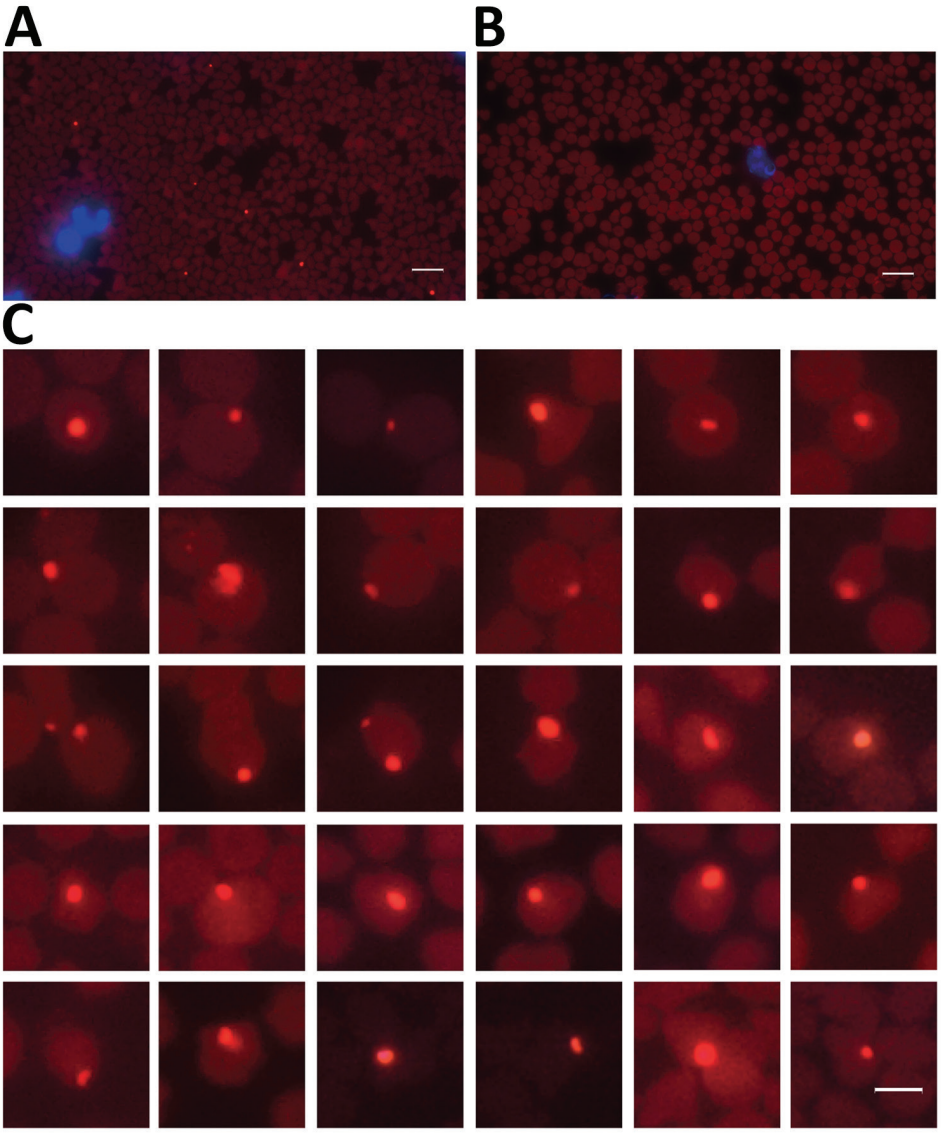


Figure 1. *Anaplasma capra* in the erythrocytes of an infected goat detected by fluorescence in situ hybridization (FISH) in study of emerging intraerythrocytic *A. capra* and high prevalence in goats, China. Glowing red indicates *A. capra*; blue indicates leukocyte nucleus stained with fluorescent antibody blocker containing DAPI. A) FISH results under fluorescence microscope of *A. capra*. B) FISH results of *A. capra*–negative blood smear. C) FISH results showing different shapes and sizes of *A. capra* in erythrocytes.

BIME2 (Figure 3, panel B). In addition, we identified 25 virulent genes in the 2 *A. capra* strains that were shared by all the species in the genus of *Anaplasma*, including *virB2* gene family, *virB6* gene family, *virB4* gene family, *virB8* gene family, *virB9* gene family, and *virB3*, *virB7*, *virB10*, *virB11*, *virD4*, *and Ats-1* genes that encode the type 4 secretion system and membrane protein-encoding genes (Appendix Table 5).

Among the 54 unique genes of *A. capra*, a total of 37 were unclassified, none of which was assigned to any KEGG category. Six of the remaining 17 genes were associated with metabolic processing, 5 genes were related to genetic information processing, and 6 were involved in signaling and cellular processing (Appendix Table 6). Among them, the most noteworthy of genes were *RSF1*, a gene related to the repair of DNA double-strand breaks (*44*), and *desk*, which encodes a protein acting as a kinase at cold temperatures in *Bacillus subtilis* (*45*).

We classified the coding proteins of the 2 *A. capra* strains (BIME1 and BIME2) into functional clusters of orthologous group (COG) categories and compared them with those of representative species strains in the genus *Anaplasma* (Appendix Table 7). Most proteins were involved in translation, ribosomal structure and biogenesis, energy production and conversion, and nutrient (including amino acid, nucleotide, carbohydrate, coenzyme, and lipid) transport and metabolism, all of which were essential for bacterial survival. Of note, the number of genes encoding cell wall and membrane in *A. platys* was substantially lower than those of other *Anaplasma* species. In addition, ≈10% of the proteins did not assign to any COG category and were classified as function unknown in each species.

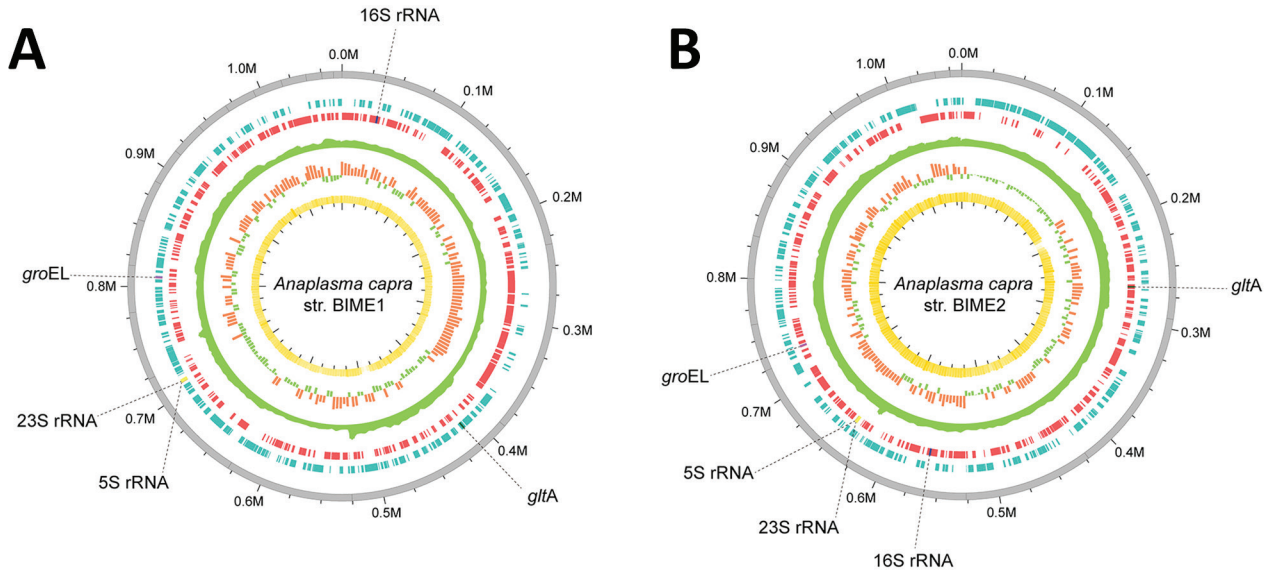


Figure 2. Circular map of *Anaplasma capra* strains BIME1 and BIME2 genomes in study of emerging intraerythrocytic *A. capra* and high prevalence in goats, China. The outermost ring shows the genome size in 100-kb increments. Moving inward, the blue-green and red marks indicate the coding sequences on the reverse and forward strands. The fourth ring represents the sequencing depth. The fifth ring shows the G+C skew, and the sixth rings show and G+C content. The location of *groEL and gltA* genes and the complete ribosomal RNA genes (5S rRNA, 16S rRNA, and 23S rRNA) within the genome are indicated.

We screened blood samples from 3 flocks of 54 goats in Shandong Province and a flock of 18 goats in Guizhou Province (Appendix Figure 1) by using nested PCR and qPCR targeting different regions of the *gltA* gene (Appendix Table 1). The overall positive rate was 59.7% (95% CI 48.4%–71.0%), and the positive rate was significantly higher among goats in Guizhou Province than in Shandong Province (77.8% vs. 53.7%; p<0.001). Accordingly, among the *H. longicornis* ticks collected from the same sites of the positive goats, the overall positive rate was 8.0% (95% CI 4.2%–11.8%), and the *A. capra* infection rate was significantly higher among ticks in Guizhou Province than that in Shandong Province (15.8% vs. 4.9%; p<0.001) (Appendix Table 8). To understand the genetic diversity, we amplified *A. capra* 16S rRNA (1,500 bp), *groEL* (1264 bp), and *msp4* (799 bp) genes from those positive samples. We compared the nucleotide identities for each gene sequence and (Appendix Figures 2–5; GenBank accession numbers are provided).

The *gltA* genes amplified from either goats or ticks in this study had 99.7%–100% identity with each other and with the strain that infected humans (Appendix Figure 2). The phylogenetic analysis based on *gltA* gene revealed that the *A. capra* sequences in this study were in an independent cluster from those previously reported in various animals from China and South Korea but distinct from those detected in wild and domestic animals from Europe and Kyrgystan. The South Korea water deer seemed to be capable of carrying both variants of *A. capra* (Figure 4, panel A). No *A. capra groEL* gene was acquired from tick samples, and the sequences from goats shared 99.4%–100% identity with each other and 99.8%–100% with sequences from humans (Appendix Figure 3). Similarly, the phylogenetic analyses based on the *groEL* gene revealed that *A. capra* strains of this study clustered with those from humans, dogs, and domestic ruminants in Asia but were distinguished from those in Europe (Figure 4, panel B). The entire 16S rRNA gene sequences (1,500 bp) of *A. capra* detected in goats and *H. longicornis* ticks from either Shandong or Guizhou Province shared average similarity of >99.7% from each other and from the sequence detected in humans (Appendix Figure 4). The phylogenetic tree based on 16S rRNA gene sequences indicated that all the *A. capra* strains detected in this study were in the same clade with previously reported strains in Asia (Figure 4, panel C). The *A. capra msp4* gene sequences were also relatively conserved (Appendix Figure 5) among the goats and ticks, and the topology of phylogenetic tree based on *msp4* gene were similar to that based on the 16S rRNA gene, in which all *A. capra* sequences clustered in the clade different from other members of *Anaplasma* species (Figure 4, panel D).

Discussion

Whole-genome assembly of obligate intracellular bacteria has usually been hindered by the DNA presence

of host cells. In this study, we first assembled 2 complete genomes of *A. capra* from the red blood cells of infected goats by using the metagenomic sequencing strategy. Because *A. capra* is an intraerythrocytic pathogen (*1*,*29*), we separated erythrocytes from the periphery blood of the infected goats and then lysed them for maximum removal of goat DNA. After metagenomic next-generation sequencing, we discarded the remaining goat genomic sequences and successfully assembled the *A. capra* genomes from 2 infected goats. The high percentage of reads from goat could be attributable to the low abundance of *A. capra* in erythrocytes or the fact that all other host cells rather than erythrocytes were not totally removed during the isolation of erythrocytes. In any case, the completeness of the 2 *A. capra* genomes are up to 99.79% for BIME1 and 99.36% for BIME2. The genome sizes obtained in this study reach 1,066,874 bp for BIME1 and 1,059,758 bp for BIME2. Therefore, their predicted sizes are ≈1.07 Mbp, which remain the smallest genome in the genus of *Anaplasma*. The phylogenetic analysis based on genome sequences and the comparative analyses of genomic characteristics provide the evidence that *A. capra* is closely related to other intraerythrocytic *Anaplasma* species, including *A. ovis*, *A. centrale*, and *A. marginale*.

The genome of *A. capra* consists of a single circular chromosome with a total size of 1.07 Mbp and has 862 protein-coding genes, which is smaller than other *Anaplasma* species. In fact, all the *Anaplasma* genomes sequenced so far are relatively small compared with free-living bacteria. The small genome size might be because a part of the intracellular bacterial functions has been compensated by the host cells, a process of reductive evolution that has occurred in the order Rickettsiales because of long-term intracellular association with eukaryotic hosts (*46*). This reductive evolution is associated with the frequent formation of pseudogenes, affecting distinct loci in different species (*47*). Moreover, we found that the G+C content of *A. capra* is close to that of *A. ovis*, *A. marginale*, and *A. centrale*. Of note, their relatedness also seems to be closest according to the phylogenetic analysis. The common invasiveness of erythrocytes also accounts for their high similarity.

A limitation of this study is that both the *A. capra* genomes were directly derived from the blood samples of infected goats through metagenomic next-generation sequencing. Unfortunately, we did not obtain the genomes at chromosome level, which usually relies on 3rd-generation sequencing of an isolate. In any case, this study reveals the genomic characteristics of *A. capra* and sheds light on its genetic diversity.

The high prevalence of *A. capra* in goats from Shandong and Guizhou Provinces in this study

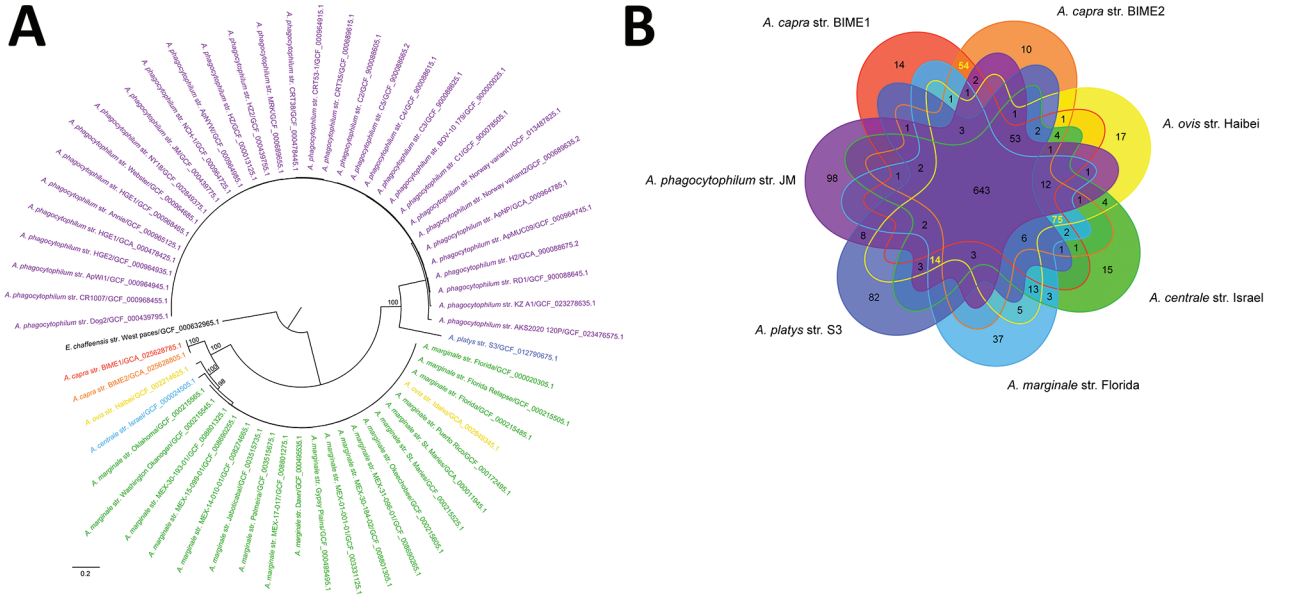


Figure 3. Phylogenetic tree and genomic comparison among *Anaplasma* species in study of emerging intraerythrocytic *A. capra* and high prevalence in goats, China. A) Phylogenetic tree of *Anaplasma* species based on all the genomic sequences deposited in GenBank, constructed by using maximum-likelihood method with *Ehrlichia chaffeensis* as an outgroup. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches. B) Differences in gene contents among *Anaplasma* species strains. Venn diagrams show the distribution of shared and unique gene clusters among representative *Anaplasma* species.

further indicate that domestic ruminants might be the main animal hosts, as suggested by previous studies (*2*–*5*). *H. longicornis* ticks collected from the same sites of the positive goats either in Shandong Province or Guizhou Province are naturally infected with *A. capra*, implying the role of the tick species in transmission of the pathogen. Phylogenetic analyses based on the *gltA* and *groEL* genes demonstrate that *A. capra* strains detected from goats and *H. longicornis* ticks in this study are clustered in the same clade with those

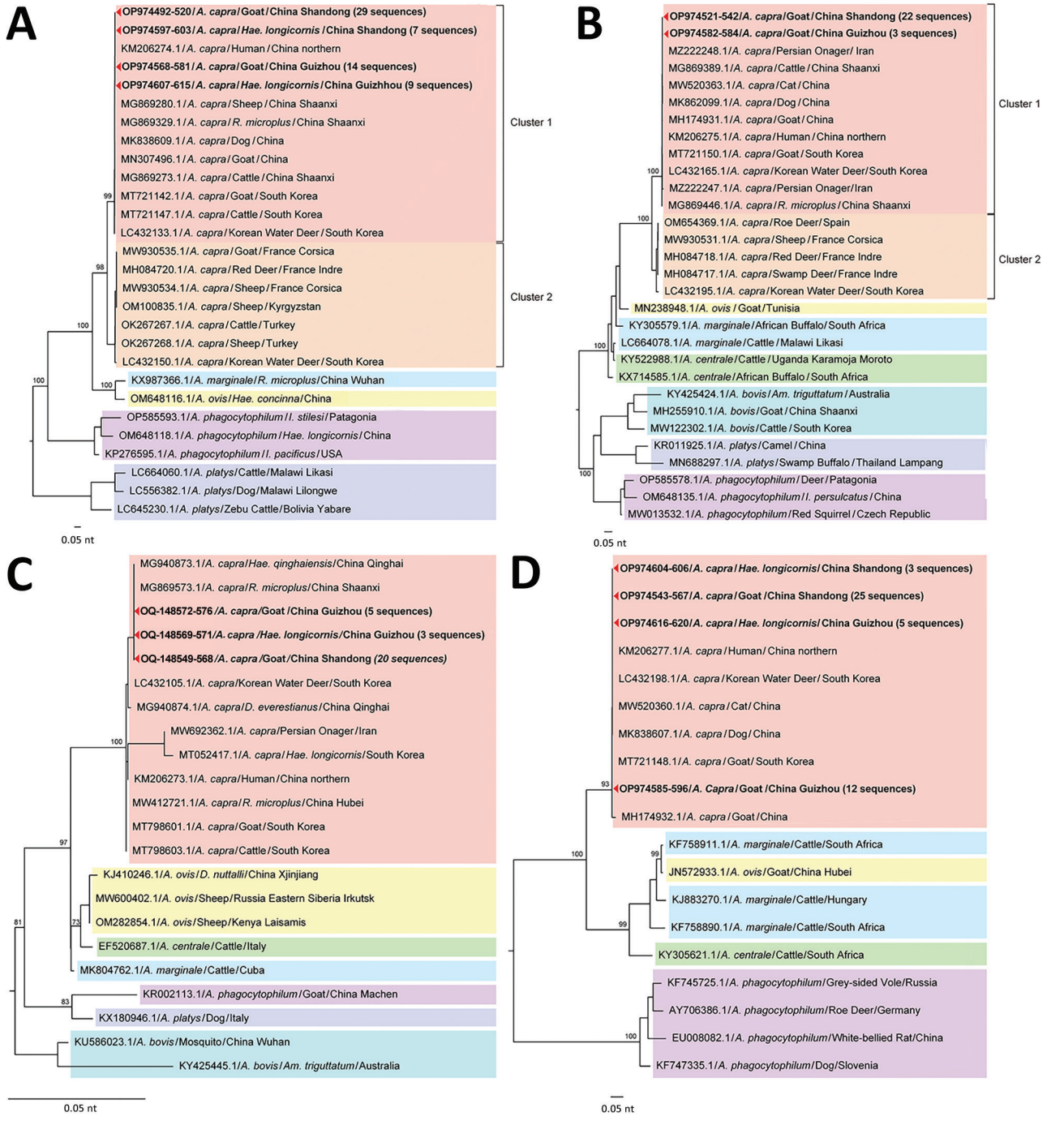


Figure 4. Phylogenetic analysis of *Anaplasma capra* based on nucleotide sequences of 4 genes in study of emerging intraerythrocytic *A. capra* and high prevalence in goats, China. A) Phylogenetic tree based on 536 bp nucleotide sequence of *gltA*. B) Phylogenetic tree based on 620 bp nucleotide sequence of *groEL*. C) Phylogenetic tree based on 860 bp nucleotide sequence of 16S rRNA. D) Phylogenetic tree based on 642 bp nucleotide sequence of *msp4*. We performed bootstrap analysis of 1,000 replicates to assess the reliability of the reconstructed phylogenies. GenBank accession numbers are provided. Scale bars show estimated evolutionary distance.

from humans, domestic ruminants, dogs, and Korean water deer (*2*,*3*,*5*,*10*). Of note, another clade of *A. capra* strains is mainly found in the wild and domestic animals from Europe and Kyrgyzstan (*6*,*10*,*48*). Those findings suggest that the enzootic cycles in various regions of the world might be different. Public health professionals should pay enough attention and formulate prevention and control strategies to reduce the health threat of the emerging tickborne pathogen to humans in other countries besides China.

This study was supported by the Natural Science Foundation of China (grant no. 81621005 to W.-C.C.; grant nos. 81760605, 82160633, and 82103897 to L.Z.), the Natural Science Foundation of Shandong Province, China (grant no. ZR2020QH299 to L.Z.), and the Cheeloo Young Scholar Program of Shandong University.

About the Author

Mr. Lin is an MPH candidate in the State Key Laboratory of Pathogen and Biosecurity at the Beijing Institute of Microbiology and Epidemiology. His primary research interests are the prevention and control of tickborne pathogens.

References

1. Li H, Zheng YC, Ma L, Jia N, Jiang BG, Jiang RR, et al. Human infection with a novel tick-borne *Anaplasma* species in China: a surveillance study. Lancet Infect Dis. 2015;15:663–70. https://doi.org/10.1016/S1473-3099(15)70051-4
2. Yang J, Liu Z, Niu Q, Liu J, Han R, Guan G, et al. A novel zoonotic *Anaplasma* species is prevalent in small ruminants: potential public health implications. Parasit Vectors. 2017;10:264. https://doi.org/10.1186/s13071-017-2182-9
3. Seo MG, Ouh IO, Lee H, Geraldino PJL, Rhee MH, Kwon OD, et al. Differential identification of *Anaplasma* in cattle and potential of cattle to serve as reservoirs of *Anaplasma capra*, an emerging tick-borne zoonotic pathogen. Vet Microbiol. 2018;226:15–22. https://doi.org/10.1016/j.vetmic.2018.10.008
4. Wang Y, Zhang Q, Han S, Li Y, Wang B, Yuan G, et al. *Ehrlichia chaffeensis* and four *Anaplasma* species with veterinary and public health significance identified in Tibetan sheep (*Ovis aries*) and yaks (*Bos grunniens*) in Qinghai, China. Front Vet Sci. 2021;8:727166. https://doi.org/10.3389/fvets.2021.727166
5. Shi K, Li J, Yan Y, Chen Q, Wang K, Zhou Y, et al. Dogs as new hosts for the emerging zoonotic pathogen *Anaplasma capra* in China. Front Cell Infect Microbiol. 2019;9:394. https://doi.org/10.3389/fcimb.2019.00394
6. Yang J, Liu Z, Niu Q, Mukhtar MU, Guan G, Liu G, et al. A novel genotype of "*Anaplasma capra*" in wildlife and its phylogenetic relationship with the human genotypes. Emerg Microbes Infect. 2018;7:210. https://doi.org/10.1038/s41426-018-0212-0
7. Staji H, Yousefi M, Hamedani MA, Tamai IA, Khaligh SG. Genetic characterization and phylogenetic of *Anaplasma capra* in Persian onagers (*Equus hemionus onager*). Vet Microbiol. 2021;261:109199. https://doi.org/10.1016/j.vetmic.2021.109199
8. Rocchigiani G, Ebani VV, Nardoni S, Bertelloni F, Bascherini A, Leoni A, et al. Molecular survey on the occurrence of arthropod-borne pathogens in wild brown hares (*Lepus europaeus*) from central Italy. Infect Genet Evol. 2018;59:142–7. https://doi.org/10.1016/j.meegid.2018.02.005
9. Sato M, Nishizawa I, Fujihara M, Nishimura T, Matsubara K, Harasawa R. Phylogenetic analysis of the 16S rRNA gene of *Anaplasma* species detected from Japanese serows (*Capricornis crispus*). J Vet Med Sci. 2009;71:1677–9. https://doi.org/10.1292/jvms.001677
10. Amer S, Kim S, Yun Y, Na KJ. Novel variants of the newly emerged *Anaplasma capra* from Korean water deer (*Hydropotes inermis argyropus*) in South Korea. Parasit Vectors. 2019;12:365. https://doi.org/10.1186/s13071-019-3622-5
11. Yang J, Liu Z, Niu Q, Liu J, Han R, Liu G, et al. Molecular survey and characterization of a novel *Anaplasma* species closely related to *Anaplasma capra* in ticks, northwestern China. Parasit Vectors. 2016;9:603. https://doi.org/10.1186/s13071-016-1886-6
12. Zhang H, Sun Y, Jiang H, Huo X. Prevalence of severe febrile and thrombocytopenic syndrome virus, *Anaplasma* spp. and *Babesia microti* in hard ticks (Acari: Ixodidae) from Jiaodong Peninsula, Shandong Province. Vector Borne Zoonotic Dis. 2017;17:134–40. https://doi.org/10.1089/vbz.2016.1978
13. Han R, Yang JF, Mukhtar MU, Chen Z, Niu QL, Lin YQ, et al. Molecular detection of *Anaplasma* infections in ixodid ticks from the Qinghai-Tibet Plateau. Infect Dis Poverty. 2019;8:12. https://doi.org/10.1186/s40249-019-0522-z
14. Guo WP, Zhang B, Wang YH, Xu G, Wang X, Ni X, et al. Molecular identification and characterization of *Anaplasma capra* and *Anaplasma platys*–like in *Rhipicephalus microplus* in Ankang, Northwest China. BMC Infect Dis. 2019;19:434. https://doi.org/10.1186/s12879-019-4075-3
15. Remesar S, Prieto A, García-Dios D, López-Lorenzo G, Martínez-Calabuig N, Díaz-Cao JM, et al. Diversity of *Anaplasma* species and importance of mixed infections in roe deer from Spain. Transbound Emerg Dis. 2022;69:e374–85. https://doi.org/10.1111/tbed.14319
16. Elhachimi L, Rogiers C, Casaert S, Fellahi S, Van Leeuwen T, Dermauw W, et al. Ticks and tick-borne pathogens abound in the cattle population of the Rabat-Sale Kenitra Region, Morocco. Pathogens. 2021;10:1594. https://doi.org/10.3390/pathogens10121594
17. Dumler JS, Barbet AF, Bekker CP, Dasch GA, Palmer GH, Ray SC, 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;51:2145–65. https://doi.org/10.1099/00207713-51-6-2145
18. Tate CM, Howerth EW, Mead DG, Dugan VG, Luttrell MP, Sahora AI, et al. *Anaplasma odocoilei* sp. nov. (family Anaplasmataceae) from white-tailed deer (*Odocoileus virginianus*). Ticks Tick Borne Dis. 2013;4:110–9. https://doi.org/10.1016/j.ttbdis.2012.09.005
19. Calchi AC, Vultão JG, Alves MH, Yogui DR, Desbiez ALJ, De Santi M, et al. *Ehrlichia* spp. and *Anaplasma* spp. in Xenarthra mammals from Brazil, with evidence of novel '*Candidatus Anaplasma* spp.'. Sci Rep. 2020;10:12615. https://doi.org/10.1038/s41598-020-69263-w
20. Vanstreels RET, Yabsley MJ, Parsons NJ, Swanepoel L, Pistorius PA. A novel candidate species of *Anaplasma* that infects avian erythrocytes. Parasit Vectors. 2018;11:525. https://doi.org/10.1186/s13071-018-3089-9

21. Rar V, Tkachev S, Tikunova N. Genetic diversity of *Anaplasma* bacteria: twenty years later. Infect Genet Evol. 2021;91:104833. https://doi.org/10.1016/j.meegid. 2021.104833
22. Brayton KA, Kappmeyer LS, Herndon DR, Dark MJ, Tibbals DL, Palmer GH, 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 U S A. 2005;102:844–9. https://doi.org/10.1073/pnas.0406656102
23. Dall'Agnol B, Webster A, Souza UA, Barbieri A, Mayer FQ, Cardoso GA, et al. Genomic analysis on Brazilian strains of *Anaplasma marginale*. Rev Bras Parasitol Vet. 2021;30:e000421. https://doi.org/10.1590/s1984-29612021043
24. Dunning Hotopp JC, Lin M, Madupu R, Crabtree J, Angiuoli SV, Eisen JA, et al. Comparative genomics of emerging human ehrlichiosis agents. PLoS Genet. 2006;2:e21. https://doi.org/10.1371/journal.pgen.0020021
25. Barbet AF, Al-Khedery B, Stuen S, Granquist EG, Felsheim RF, Munderloh UG. An emerging tick-borne disease of humans is caused by a subset of strains with conserved genome structure. Pathogens. 2013;2:544–55. https://doi.org/10.3390/pathogens2030544
26. Herndon DR, Palmer GH, Shkap V, Knowles DP Jr, Brayton KA. Complete genome sequence of *Anaplasma marginale* subsp. *centrale*. J Bacteriol. 2010;192:379–80. https://doi.org/10.1128/JB.01330-09
27. Liu Z, Peasley AM, Yang J, Li Y, Guan G, Luo J, et al. The *Anaplasma ovis* genome reveals a high proportion of pseudogenes. BMC Genomics. 2019;20:69. https://doi.org/10.1186/s12864-018-5374-6
28. Llanes A, Rajeev S. First whole genome sequence of *Anaplasma platys*, an obligate intracellular rickettsial pathogen of dogs. Pathogens. 2020;9:277. https://doi.org/10.3390/pathogens9040277
29. Peng Y, Lu C, Yan Y, Song J, Pei Z, Gong P, et al. The novel zoonotic pathogen, *Anaplasma capra*, infects human erythrocytes, HL-60, and TF-1 cells in vitro. Pathogens. 2021;10:600. https://doi.org/10.3390/pathogens10050600
30. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10:giab008.
31. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27:824–34. https://doi.org/10.1101/gr.213959.116
32. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. Curr Protoc Bioinformatics. 2020;70:e102. https://doi.org/10.1002/cpbi.102
33. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55. https://doi.org/10.1101/gr.186072.114
34. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9. https://doi.org/10.1093/bioinformatics/btu153
35. Syberg-Olsen MJ, Garber AI, Keeling PJ, McCutcheon JP, Husnik F. Pseudofinder: detection of pseudogenes in prokaryotic genomes. Mol Biol Evol. 2022;39:msac153.
36. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9:5114. https://doi.org/10.1038/s41467-018-07641-9
37. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238. https://doi.org/10.1186/s13059-019-1832-y
38. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4. https://doi.org/10.1093/molbev/msaa015
39. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis version 11. Mol Biol Evol. 2021;38:3022–7. https://doi.org/10.1093/molbev/msab120
40. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020;36:2251–2. https://doi.org/10.1093/bioinformatics/btz859
41. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38:5825–9. https://doi.org/10.1093/molbev/msab293
42. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UC, Podowski RM, et al. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature. 1998;396:133–40. https://doi.org/10.1038/24094
43. Beare PA, Sandoz KM, Omsland A, Rockey DD, Heinzen RA. Advances in genetic manipulation of obligate intracellular bacterial pathogens. Front Microbiol. 2011;2:97. https://doi.org/10.3389/fmicb.2011.00097
44. Helfricht A, Wiegant WW, Thijssen PE, Vertegaal AC, Luijsterburg MS, van Attikum H. Remodeling and spacing factor 1 (RSF1) deposits centromere proteins at DNA double-strand breaks to promote non-homologous end-joining. Cell Cycle. 2013;12:3070–82. https://doi.org/10.4161/cc.26033
45. Albanesi D, Martín M, Trajtenberg F, Mansilla MC, Haouz A, Alzari PM, et al. Structural plasticity and catalysis regulation of a thermosensor histidine kinase. Proc Natl Acad Sci U S A. 2009;106:16185–90. https://doi.org/10.1073/pnas.0906699106
46. Darby AC, Cho NH, Fuxelius HH, Westberg J, Andersson SG. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. Trends Genet. 2007;23:511–20. https://doi.org/10.1016/j.tig.2007.08.002
47. Ogata H, Audic S, Renesto-Audiffren P, Fournier PE, Barbe V, Samson D, et al. Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. Science. 2001;293:2093–8. https://doi.org/10.1126/science.1061471
48. Jouglin M, Rispe C, Grech-Angelini S, Gallois M, Malandrin L. *Anaplasma capra* in sheep and goats on Corsica Island, France: a European lineage within *A. capra* clade II? Ticks Tick Borne Dis. 2022;13:101934. https://doi.org/10.1016/j.ttbdis.2022.101934

Address for correspondence: Wu-Chun Cao, State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, 20 Dong-Da St, Fengtai District, Beijing 100071, China; email: caowuchun@126.com; Lin Zhao, Institute of EcoHealth, School of Public Health, Shandong University, 44 Wenhuaxi St, Jinan 250012, Shandong, China; email: zhaolin1989@sdu.edu.cn; Lin Zhan, NHC Key Laboratory of Pulmonary Immunological Diseases, Guizhou Provincial People's Hospital, 83 Zhongshan East Rd, Nanming District 550002, Guiyang, Guizhou, China; email: zhanlin300@hotmail.com

Article DOI: https://doi.org/10.3201/eid2909.230131

Genomic Characteristics of Emerging Intraerythrocytic *Anaplasma capra* and High Prevalence in Goats, China

Appendix

Appendix Table 1. Nucleotide sequence of primers used in the study

Target	Primer name	Primer sequence (5′-3′)	Annealing temperature (°C)	Amplicon size (bp)	References
rrs	Eh-out1	TTGAGAGTTTGATCCTGGCTCAGAACG	50	1500	(*1*)
	3–17U	WAAGGWGGTAATCCAGC			
gltA	Outer-f	GCGATTTTAGAGTGYGGAGATTG	55	1076	(*2*)
	Outer-r	TACAATACCGGAGTAAAAGTCAA			
	Inner-f	GGGTTCMTGTCYACTGCTGCGTG	51	792	(*2*)
	Inner-r	TTGGATCGTARTTCTTGTAGACC			
groEL	Forward	GCGAGGCGTTAGACAAGTCCATT	56	1264	(*2*)
	Reverse	TCCAGAGATGCAAGCGTGTATAG			
msp4	Forward	CAGTCTGCGCCTGCTCCCTAC	55	799	(*2*)
	Reverse	AGGAATCTTGCTCCAAGGTTA			
msp2	Forward	GCGTGTTGATGGCTCTGGT	51	1139	(*2*)
	Reverse	ACCAGTATCCTTATTTTTACC			
*gltA**	Forward	CGAATCTATTTGCCTGCTT	60	200	This study
	Reverse	ATCGTAATTCTTGTAGACCCT			

*This pair of primers is used for the quantitative PCR.

References

1. Wen B, Jian R, Zhang Y, Chen R. Simultaneous detection of *Anaplasma marginale* and a new *Ehrlichia* species closely related to *Ehrlichia chaffeensis* by sequence analyses of 16S ribosomal DNA in *Boophilus microplus* ticks from Tibet. J Clin Microbiol. 2002;40:3286–90. PubMed https://doi.org/10.1128/JCM.40.9.3286-3290.2002

2. Li H, Zheng YC, Ma L, Jia N, Jiang BG, Jiang RR, et al. Human infection with a novel tick-borne Anaplasma species in China: a surveillance study. Lancet Infect Dis. 2015;15:663–70. PubMed https://doi.org/10.1016/S1473-3099(15)70051-4

Appendix Table 2. Probe sequences of fluorescence in situ hybridization (FISH)

Probe sequence (5′ to 3′)	Probe sequence name
TTCTGAGCCAGGATCAAACT	AC16S-1
TCGACTTGCATGTGTTAAGC	AC16S-2
AGCAAGCTACAGATTTGGTC	AC16S-3
CGTCTGCCACTAACCAAATC	AC16S-4
AGATTCCTATGCATTACTCA	AC16S-5
TGGCTATCCCATACTACTAG	AC16S-6
GGATTATACGGTATTACCCA	AC16S-7
ATAGCGATAAATCTTTCCCC	AC16S-8
CCAACTAGCTAATCCGACAT	AC16S-9
ACAGATCACTGCCTTGGTAG	AC16S-10
TGATCATCCTCTCAGACCAG	AC16S-11
CATTGTCCAATATTCCCCAC	AC16S-12
CATAGCTGGATCAGGCTTGC	AC16S-13
TTTTACAACCCTAAGGCCTT	AC16S-14
TCATTATCTTCCCTACTGAA	AC16S-15
GGGACTTCTTCTGTAGGTAC	AC16S-16
CGCCCAATAATTCCGAACAA	AC16S-17
TTAACTTACCAAACCGCCTA	AC16S-18
GTTAAGCCCTGGTATTTCAC	AC16S-19
TGCAGTATTAAAAGCAGCCC	AC16S-20
TATCCTCTTCCGGACTCTAG	AC16S-21
ATTTCACCTCTACACTAGGA	AC16S-22
GTGTTCCTCCTAATATCTAC	AC16S-23
CAGGGTATCTAATCCTGTTT	AC16S-24
TCAGCACTCATCGTTTACAG	AC16S-25
CAACACAGAGGCAAAAGCCC	AC16S-26
CGGAGTGCTTAACGCGTTAG	AC16S-27
CCTTTGAGTTTTAGTCTTGC	AC16S-28
CGAATTAAACCACATGCTCC	AC16S-29
TGGTAAGGTTTTTCGCGTTG	AC16S-30
ATCTAACCTCCATGTCAAGA	AC16S-31
AACTGCGCCTTTCTGTTAAG	AC16S-32
ACGAGCTGACGACAGCCATG	AC16S-33
ACTTAACCCAACATCTCACG	AC16S-34
ATGAGGGTTACGCTCGTTGC	AC16S-35
CATTACCCGCTGGTAACTAA	AC16S-36
CACCGGCAGTTTCCTTAAAG	AC16S-37
CGTGCTGACTTGACATCATC	AC16S-38
CATTGTAGCACGTGTGTAGC	AC16S-39
CGACGTTGCAACCTATTGTG	AC16S-40
CTTTTACGGATTAGCTCAGC	AC16S-41
CTCGAGTTGCAGAGGACAAT	AC16S-42
TCCACGATTACTAGCGATTC	AC16S-43
CGAGAACGTATTCACCGTGG	AC16S-44
TGACGGGCAGTGTGTACAAG	AC16S-45
TTTGAGTTAAGCCAATTCCC	AC16S-46
CACCGACCCAACCTTAAATG	AC16S-47
TACAGCTACCTTGTTACGAC	AC16S-48

Appendix Table 3. Genomic characteristics of *Anaplasma capra* strains BIME1 and BIME2 compared with that of representative *Anaplasma* species strains*

	Strain (GenBank accession no.)						
Characteristic	*A. capra* str. BIME1 (GCA_025628785.1)	*A. capra* str. BIME2 (GCA_025628805.1)	*A. ovis* str. Haibei (NZ_CP015994.1)	*A. centrale* str. Israel (NC_013532.1)	*A. marginale* str. Florida (NC_012026.1)	*A. platys* str. S3 (NZ_CP046391.1)	*A. phagocytophilum* str. JM (NC_021880)
Size (bp)	1,066,874	1,059,758	1,214,674	1,206,806	1,202,435	1,196,811	1,481,598
GC rate (%)	48.32	48.32	48.9	50.0	49.8	45.5	41.6
Gene counts (n)	929	932	1021	993	992	940	1155
CDS counts (n)	862	863	945	922	913	882	997
Pseudogenes (n)	27	29	32	27	35	17	114
rRNAs (n)	3	3	3	3	3	3	3
tRNAs (n)	37	37	37	37	37	34	37
Completeness (%)	99.79	99.36	NA	NA	NA	NA	NA

*bp, base pair; CDS, coding sequence; rRNA, ribosomal ribonucleic acid; tRNA, transfer ribonucleic acid; NA, not applicable.

Appendix Table 4. The estimated values of average nucleotide identity (ANI) and DNA-DNA hybridization (DDH) between *Anaplasma capra* and the other *Anaplasma* species

	ANI		DDH	
Species	*A. capra* str. BIME1	*A. capra* str. BIME2	*A. capra* str. BIME1	*A. capra* str. BIME2
A. ovis	78.0783 (GCA_002849345.1)	78.0878 (GCA_002214625.1)	17.4 (GCA_002849345.1)	17.5 (GCA_002849345.1)
A. marginale	78.2897 (GCA_008801305.1)	77.9471 (GCA_000020305.1)	17.9 (GCA_000172475.1)	17.9 (GCA_000172475.1)
A. centrale	77.9688 (GCA_000024505.1)	77.8613 (GCA_000024505.1)	17.4 (GCA_000024505.1)	17.4 (GCA_000024505.1)
A. phagocytophilum	(-)	(-)	13.0 (GCA_023476575.1)	13.0 (GCA_023278635.1)
A. platys	(-)	(-)	13.1 (GCA_012790675.1)	13.1 (GCA_012790675.1)

Appendix Table 5. Virulence genes in *Anaplasma capra* str. BIME1 and BIME2

		GenBank accession number	
Gene	Description	*A. capra* str. BIME1	*A. capra* str. BIME2
virB2	type IV secretion system protein VirB2 family	MCU7611221.1	MCU7612774.1
		MCU7611222.1	MCU7612775.1
		MCU7611775.1	MCU7612776.1
		MCU7611780.1	
		MCU7611781.1	
		MCU7611782.1	
virB3	type IV secretion system protein VirB3	MCU7611541.1	MCU7612020.1
virB4	type IV secretion system protein VirB4 family	MCU7611542.1	MCU7612019.1
		MCU7611779.1	MCU7612773.1
virB6	type IV secretion system protein VirB6 family	MCU7611543.1	MCU7612018.1
		MCU7611544.1	MCU7612017.1
		MCU7611545.1	MCU7612016.1
		MCU7611546.1	MCU7612015.1
virB7	type IV secretion system protein VirB7	MCU7611364.1	MCU7612438.1
virB8	type IV secretion system protein VirB8 family	MCU7611203.1	MCU7612293.1
		MCU7611581.1	MCU7611980.1
virB9	type IV secretion system protein VirB9 family	MCU7611202.1	MCU7612294.1
		MCU7611762.1	MCU7612488.1
virB10	type IV secretion system protein VirB10	MCU7611201.1	MCU7612295.1
virB11	type IV secretion system ATPase VirB11	MCU7611200.1	MCU7612296.1
virD4	type IV secretion system component VirD4	MCU7611199.1	MCU7612297.1
Ats-1	*Anaplasma* T4SS translocated substrate-1	MCU7611426.1	MCU7612135.1
ompA	outer membrane protein OmpA	MCU7611514.1	MCU7612047.1
Asp14	14-kDa *Anaplasma* surface protein Asp14	MCU7611843.1	MCU7612563.1

Appendix Table 6. Genes predicted to be unique in *Anaplasma capra* str. BIME1 and BIME2

Gene	Protein	Function	
menA	1,4-dihydroxy-2-naphthoate polyprenyltransferase	Metabolic processing	Menaquinone (vitamin K2) biosynthesis
unknown	Glycosyltransferase 2 family protein		Glycan metabolism
MKK9	Mitogen-activated protein kinase kinase 9		Ethylene and camalexin biosynthesis
MqnX	Aminodeoxyfutalosine deaminase		Menaquinone Biosynthesis
CPS1	Peregrinol diphosphate synthase		Metabolism
atuF	Geranyl-CoA carboxylase α subunit		Geraniol degradation
Zbtb46	Zinc finger and BTB domain-containing protein 46	Genetic information processing	Transcription factor
UFL1	E3 UFM1-protein ligase 1		Cellular regulation
Hmbox1	Homeobox-containing protein 1		Transcription factor
RSF1	Remodeling and spacing factor 1		DNA repair
Ara54	E3 ubiquitin-protein ligase		Cellular regulation

Gene	Protein	Function	
MACC1	Metastasis-associated in colon cancer protein 1	Signaling and cellular processing	Signaling regulator
dia	Diaphanous protein		Cytokinesis
desK	Membrane-associated kinase DesK		Membrane-associated kinase
hbhA	Heparin binding hemagglutinin		Virulence factor
Mrgprg	Mas-related G protein-coupled receptor member G		G protein-coupled receptor
unknown	Membrane protein		Protein with domain of unknown function
unknown*	unclassified protein		Function unknown

*Including 37 unclassified genes.

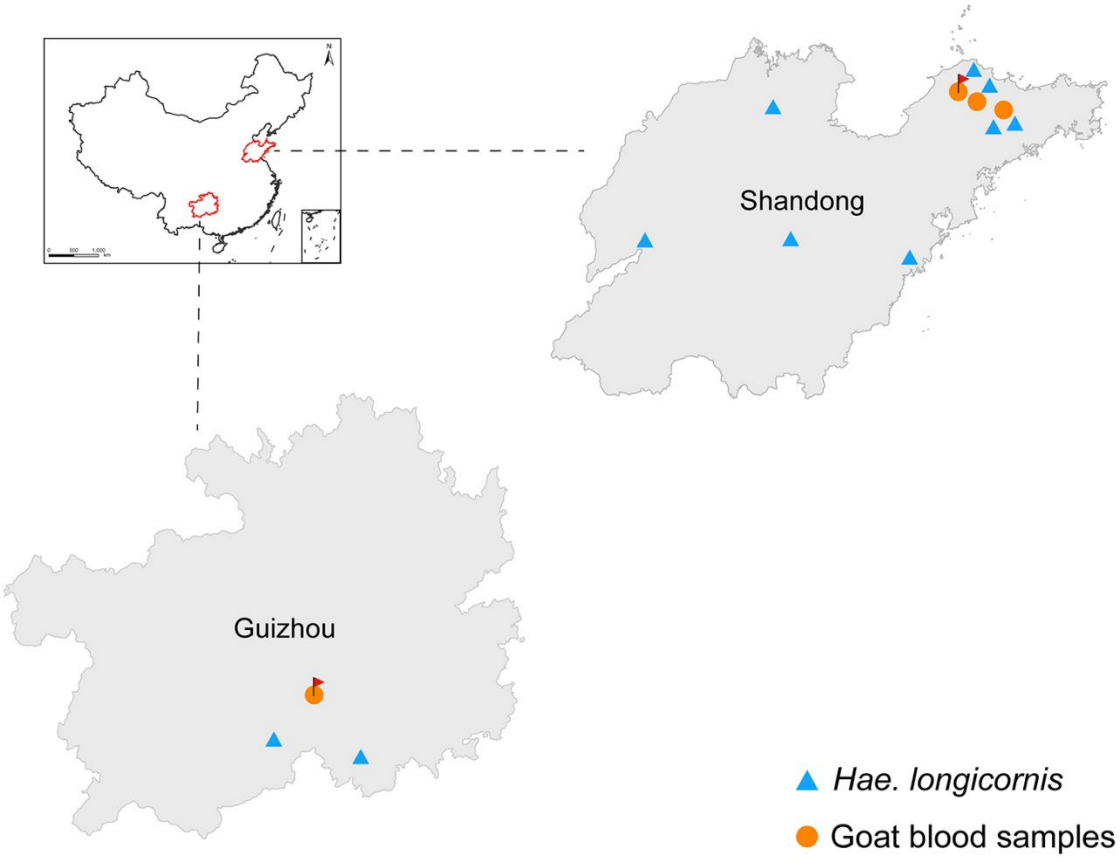
Appendix Table 7. Functional Clusters of Orthologous Groups of protein-coding genes from the representative *Anaplasma* species strains

Category	Functional category	*A. capra* str. BIME1	*A. capra* str. BIME2	*A. ovis* str. Haibei	*A. centrale* str. Israel	*A. marginale* str. Florida	*A. platys* str. S3	*A. phagocytophilum* str. JM
		Number of genes						
A	RNA processing and modification	1	1	1	1	1	1	1
B	Chromatin structure and dynamics	0	0	0	0	0	0	0
C	Energy production and conversion	67	67	74	70	68	71	74
D	Cell cycle control, cell division, chromosome partitioning	16	15	18	17	18	11	14
E	Amino acid transport and metabolism	31	31	35	34	35	30	25
F	Nucleotide transport and metabolism	54	54	55	54	54	52	56
G	Carbohydrate transport and metabolism	21	20	26	26	25	27	22
H	Coenzyme transport and metabolism	65	65	67	64	66	58	68
I	Lipid transport and metabolism	28	28	30	29	28	29	29
J	Translation, ribosomal structure and biogenesis	126	127	131	130	130	128	137
K	Transcription	18	17	21	20	21	19	21
L	Replication, recombination and repair	52	53	53	54	52	51	57
M	Cell wall/membrane/ envelope biogenesis	48	47	62	58	60	35	62
N	Cell motility	2	2	2	2	2	2	2
O	Posttranslational modification, protein turnover, chaperones	44	44	43	43	43	43	45

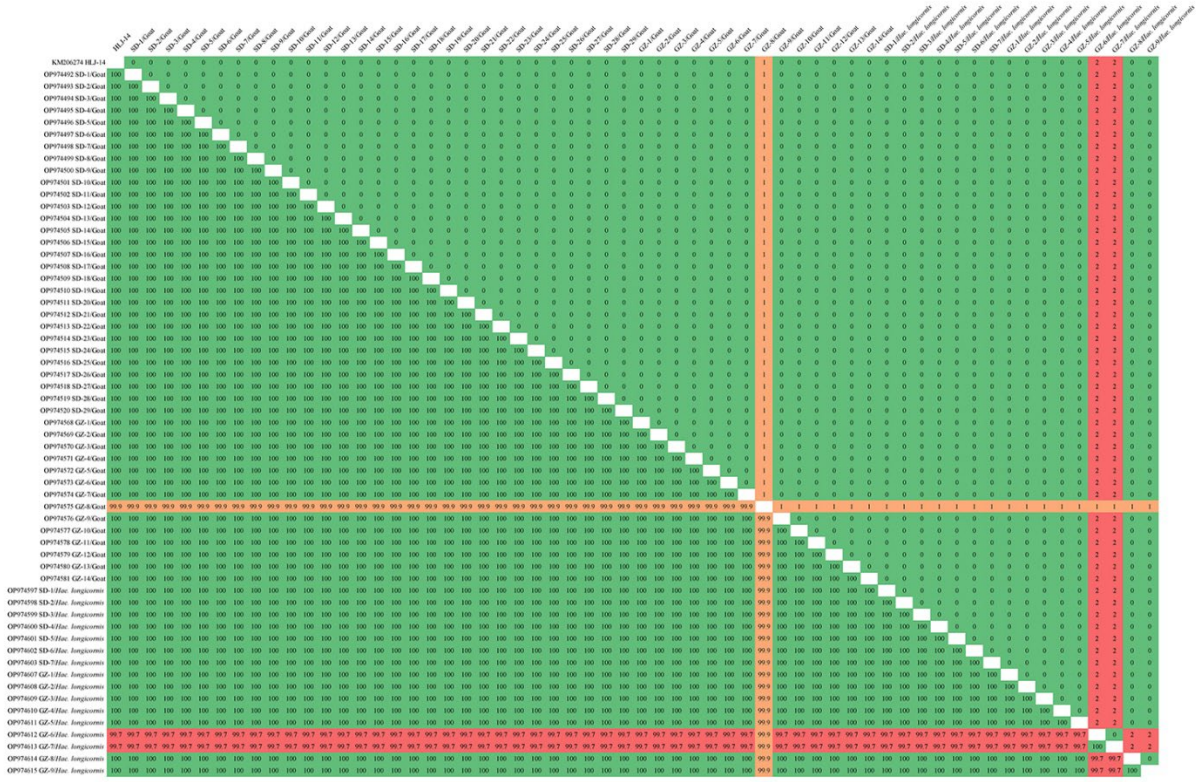
Category	Functional category	*A. capra* str. BIME1	*A. capra* str. BIME2	*A. ovis* str. Haibei	*A. centrale* str. Israel	*A. marginale* str. Florida	*A. platys* str. S3	*A. phagocytophilum* str. JM
		Number of genes						
P	Inorganic ion transport and metabolism	31	31	36	35	34	35	34
Q	Secondary metabolites biosynthesis, transport and catabolism	12	12	12	12	12	11	12
R	General function prediction only	0	0	0	0	0	0	0
S	Function unknown	90	91	89	91	92	83	88
T	Signal transduction mechanisms	9	9	10	10	10	10	9
U	Intracellular trafficking, secretion, and vesicular transport	38	38	38	38	38	40	40
V	Defense mechanisms	4	4	4	4	4	4	4
W	Extracellular structures	0	0	0	0	0	0	0
X	Mobilome: prophages, transposons	0	0	0	0	0	0	0
Y	Nuclear structure	0	0	0	0	0	0	0
Z	Cytoskeleton	0	0	0	0	0	0	0

Appendix Table 8. The *Anaplasma capra*-positive numbers of goats and *Hae. Longicornis* in this study by PCR toward different gene loci

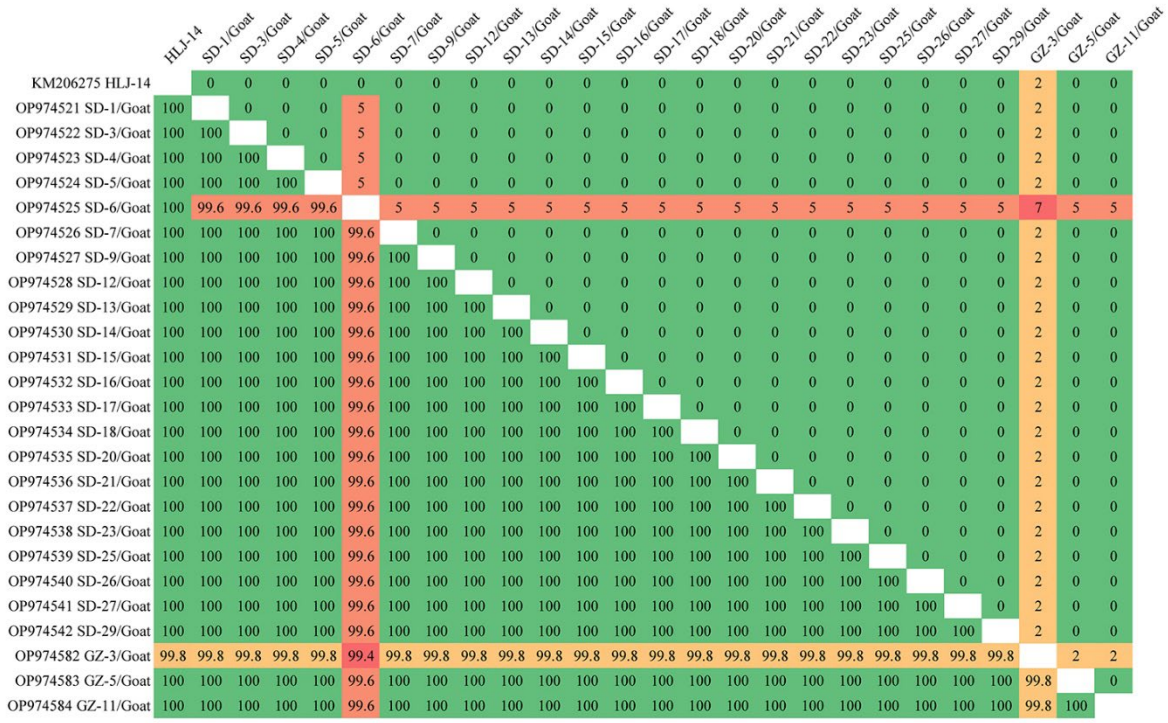
Gene loci	Goats from Shandong	Goats from Guizhou	*Hae. longicornis* from Shandong	*Hae. longicornis* from Guizhou
No. of tested	54	18	144	57
16S rRNA、*gltA*、*groEL*、*msp4*(+)	14	1	0	0
16S rRNA、*gltA*、*msp4*(+)	4	4	0	3
16S rRNA、*gltA*、*groEL*(+)	1	0	0	0
gltA、*groEL*、*msp4*(+)	7	2	0	0
16S rRNA、*gltA*(+)	1	0	0	0
*glt*A、*msp*4(+)	0	5	3	2
gltA(+)	2	2	4	4



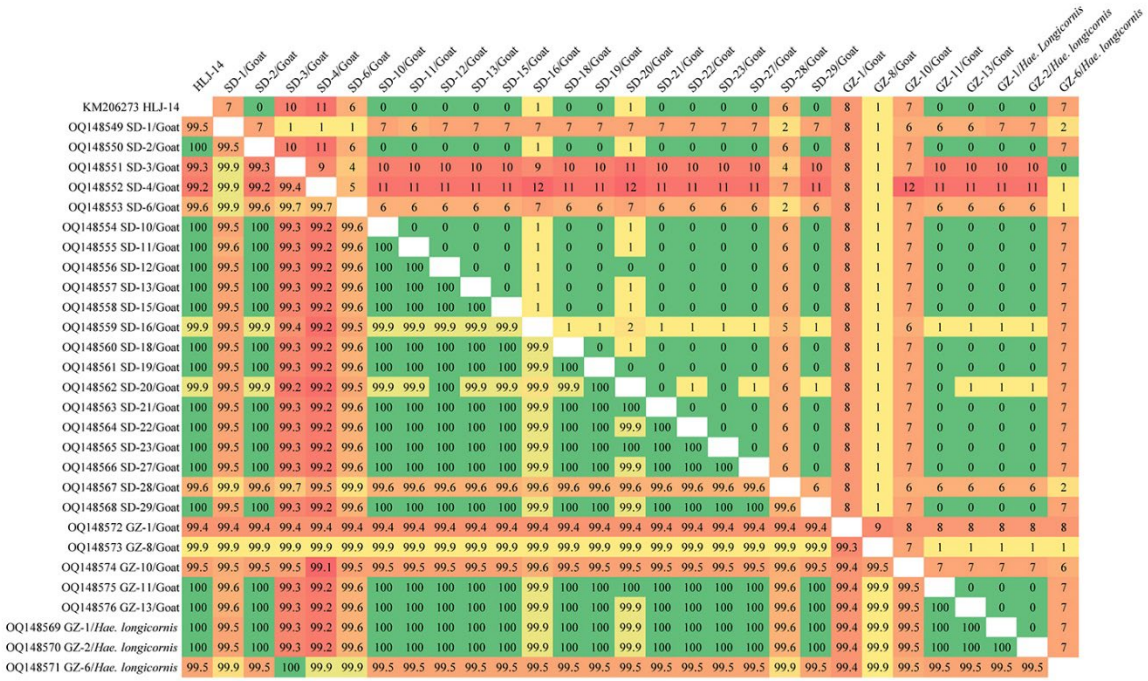
Appendix Figure 1. The sampling sites where ticks and blood samples were collected in this study. Different color and marks represent the types of samples collected in different areas. The flags indicate the locations, where the goat blood samples were collected for next-generation sequencing of *Anaplasma capra* genomes.



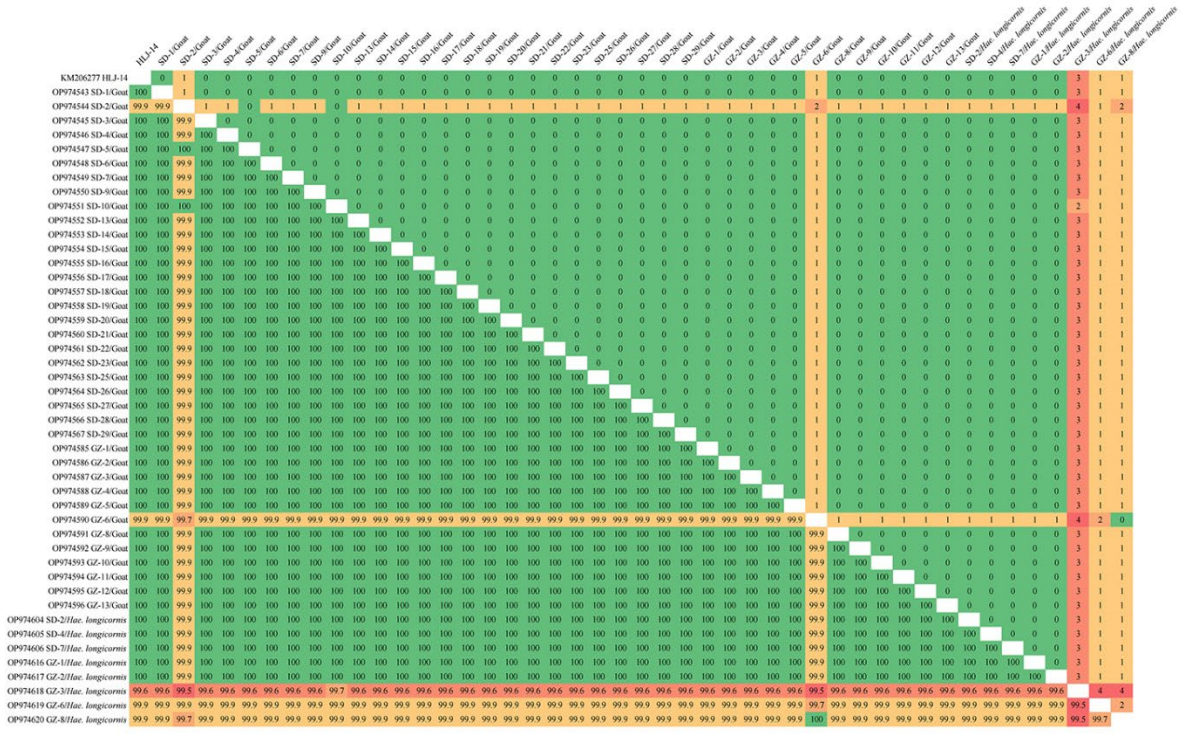
Appendix Figure 2. The comparison between each *gltA* gene sequences of *Anaplasma capra* this study and sequence from human. The upper right part represents the number of bases that differ from each sequence and the lower left part represents the nucleotide identity (%) between each sequence from others.



Appendix Figure 3. The comparison between each *groEL* gene sequences of *Anaplasma capra* this study and sequence from human. The upper right part represents the number of bases that differ from each sequence and the lower left part represents the nucleotide identity (%) between each sequence from others.



Appendix Figure 4. The comparison between each 16S rRNA gene sequences of *Anaplasma capra* this study and sequence from human. The upper right part represents the number of bases that differ from each sequence and the lower left part represents the nucleotide identity (%) between each sequence from others.



Appendix Figure 5. The comparison between each *msp4* gene sequences of *Anaplasma capra* this study and sequence from human. The upper right part represents the number of bases that differ from each sequence and the lower left part represents the nucleotide identity (%) between each sequence from others.