The Genome Sequence of Brucella abortus Vaccine Strain A19 Provides Insights on its Virulence Attenuation Compared to Brucella Abortus Strain 9-941

shuyi wang (287333611@qq.com)
Inner Mongolia Agricultural University https://orcid.org/0000-0003-4883-3370
Xueliang Zhao
Inner Mongolia Agricultural University
HUHE Bateer
Inner Mongolia Agricultural University
Ke Sun
Inner Mongolia Agricultural University
Wenlong Wang
Inner Mongolia Agricultural University

Research article

Keywords: Brucella abortus vaccine strain A19, Brucella abortus virulent strain 9-941, Comparative genomic analysis, Virulence attenuation

DOI: https://doi.org/10.21203/rs.3.rs-55444/v1

Abstract

Background: Brucellosis is a widespread disease that affects animals and humans. The live attenuated *Brucella abortus* A19 strain is used for vaccination against brucellosis in China. In addition, the main mechanisms supporting the residual toxicity of A19 have not been elucidated. Here, we performed a comprehensive comparative analysis of the genome-wide sequence of A19 against the whole genome sequences of the published virulent reference strain 9-941. The primary objective of this study was to identify candidate virulence genes by systematically comparing the genomic sequences between the two genomes.

Results: This analysis revealed two deletion regions in the A19 genome, in which all included large fragments of 63 bp, and one of their gene function is related to ABC transporter permease protein. In addition, we have identified minor mutations in important virulence-related genes that can be used to determine the underlying mechanisms of virulence attenuation. The function of its virulence gene covers LysR family transcriptional regulator, outer membrane, MFS transporter and oxidoreductase etc. At the same time, a PCR differential diagnosis method was constructed, which can distinguish A19, S19 and most other commonly used *Brucella* viruent strains and vaccine strains.

Conclusion: The data may help to provide resources for further detailed analysis of mechanisms for other *Brucella* vaccines. It laid the foundation for further distinguishing between vaccine immunity and virulent strains infection.

Background

Brucellosis remains the most crucial zoonosis infection worldwide caused by Gram-negative and small bacteria which is exist in the subphylum α- Proteobacteria [1, 2]. It reports nearly 500,000 new cases per year in disease-endemic areas (Herrick JA et al., 2014). New species of *Brucella* are increasing in recent years, there are currently 12 species in the genus *Brucella* [3]. Three of the them can cause serious Brucellosis in humans, including *Brucella abortus(B. abortus), Brucella melitensis* (*B. melitensis*) and *Brucella suis(B. suis)* [4]. And *B. abortus* which infects cattle is the most frequent in brucellosis of the world. *Brucella* leads to infertility and abortion in infected animals and causes significant economic losses [5]. Human infection with brucellosis is usually accompanied by chronic complications resulting from consumption of unpasteurized dairy products and uncooked meat or direct contact with tissues of infected animals. The spread between people has also been identified, although the probability is very low. [6, 7].

Although attenuated vaccines have various disadvantages, such as pathogenicity to humans and interference diagnostic tests, it is still an effective method for preventing and controlling brucellosis [8]. The *B.abortus* vaccine plays a core role in control and elimination brucellosis program and has been globally used for decades [9]. But common diagnostic methods cannot distinguish between natural infections of *Brucella* and artificial immune, affecting quarantine and herd purification. At home and

abroad, there is no report about the attenuation mechanism of A19 vaccine and the differential diagnosis method to distinguish between A19 vaccine immunity and natural infection. There are different strains used for immunization of different animal species, *B. melitensis* strain Rev. 1 and *Brucella abortus* strain S19 and RB51 are all animal brucellosis vaccines used for commercial [10]. With the advancement of society, the whole genome sequencing technology has also developed rapidly. Genomic sequence analysis of *B. abortus* 9-941, *B. suis* 1330 and *B. melitensis* 16M demonstrates the close correlation of these organisms [11]. So far, the *Brucella* genomes consist of two circular chromosomes, which have probably 3.3 Mb size of many genomes [12]. It is an exception that *B. suis* biovariant 3 only has one chromosome. In general, the nucleotide similarities of all *Brucella* species greater than 90% are genetically highly related to each other.

Brucella abortus (*B. abortus*) S19 strain was firstly attenuated initially by natural passage and isolated from the milk sample by Buck in 1923 [13]. Subsequently, S19 was screened, which was sensitive to erythritol, and the US Department of Agriculture began replacing the originally used strain with the screened erythritol-sensitive strain from 1956[14].S19 is still used as the main measure to control brucellosis in several countries were the disease is endemic (India, Brazil, Mexico, among others). At present, the S19 vaccine used in most countries and regions are erythritol metabolic gene deletion strains, that is, a 702 bp sequence deletion between the erythritol metabolism genes *eryC* and *eryD* [15]. The *B. abortus* A19 has been widely used for vaccination since 1956 in China which low virulence in cattle, and does not have this *ery* gene deletion [16]. Therefore, some scholars speculate that China did not introduce the gene-deficient S19 vaccine at that time, and the *B. abortus* strain A19 used in China was widely used by the *B. abortus* vaccine strain S19 before 1956.

In this article, the main goal was to compare the completely sequenced A19 genome with the genome of *B. abortus* virulent strain 9-941. Identified some candidate genes associated with virulence and established a PCR diagnosis method. Comparative genomics analysis between the two genomes gained insights into mechanism of attenuation and provided efficient data to design better vaccines for brucellosis.

Results

General Genomic Features

The size of previous completely sequenced *B. abortus* A19 genome was 3.3 Mb, and obtained 2 complete circular chromosome sequences: one was 2123997 bp long and the other was 116170 bp long in length. The average G + C content of 2 chromosomes was 57.24% [17]. The detailed genome characteristics of A19 genome and its comparison with the *B. abortus* 9-941 are listed in Table 1. Not surprisingly, the A19 genomes appeared significant similarity in structure and size compared to another virulent genomic sequence of 9-941. The genomes sequence of A19 showed over 99.6% similarity with the genomes of 9-941. The genome size of the A19 (3.286 Mbp) and 9-941 (3.286 Mbp) was within 5 kb.

Table 1
Genomic features of the newly sequenced *Brucella abortus* strain A19 in comparison with *Brucella abortus* strain 9-941.

	***Brucella abortus* A19**		***Brucella abortus* 9-941**	
Feature/Property	Chrl 1	Chrl 2	Chrl 1	Chrl 2
ORFs	2189	1182	2219	1157
tRNAs	41	14	41	14
rRNAs	8	12	6	3
Size(bp)	2123997	1162170	2124241	1162204
GC (%)	57.2	57.3	57.2	57.3
Average length(bp)	836	885	845.7	900
Hypothetical proteins	477	208	708	296

In the *B. abortus* A19 genome, we obtained 3,371 open reading frames (ORFs) annotated as predictive genes. A total of 2,189 and 1,182 ORFs were identified on the chromosomes I and II, respectively. More than 99.5% of the ORFs in A19 were found to those identical to 9-941 (NC_0006932 - NC_0006933). A19 exhibited very high genome wide collinearity with the whole 9-941genome. The two chromosomes of A19 were compared with the corresponding chromosomes of 9-941 separately (Figure. 1). A genome-wide alignment of *B. abortus* A19 and virulent strain 9-941 was performed, and it was found that the similarity between the two genomes was high and there was no large inversion region (Figure. 2). The predicted protein sequence hits the cluster of the COG database with an E-value less than 1e-5[18]. A total of 685 ORFs are hypothetical proteins, which account for 20.3% of the genomic ORFs.

Brucella abortus 9-941 is a highly virulent strain isolated from an infected cow in northwestern Wyoming [19]. It is a typical representative strain of *B. abortus*, and has been widely reported and cited in the research in recent decades. Drawing evolutionary trees found that 9-941 and A19 have a close relationship. Therefore, in this study, 9-941 was selected as the representative strain for genome-wide comparison with A19.

Although the global characteristics of the 9-941 genome alignment showed a high degree of similarity to the recently published A19 genomes, many strain-specific genetic differences also have been found. A comprehensive analysis revealed that A19 contained 47 indels (1–50 bp) with structural variations between the two genomes. A total of 215 SNPs were identified. To better approximate the functional annotation of the SNPs described below, the SNPs were classified according to their GO and then their frequency was evaluated. From these analyses, we observed that GO database is divided into three categories (Fig. 3). Among them, the proportion of catalytic activity, metabolic process and cellular process is high. The extracellular region, structural molecule activity, antioxidant activity, detoxification,

reproductive process, reproduction, negative regulation of biological process and carbon utilization all have lower SNP contents.

Discussion

Unique Genome Differences

The interspecies comparison is discussed between the two sequenced genomes of *B.abortus* vaccine A19 and virulent strain 9-941. Comparative analysis revealed differences in strain-specific and may help explain its virulence attenuation mechanisms. Although the highly conserved genomic skeleton shared by two *Brucella* strains, several important gene differences have been identified in the context. and are described the details below.

Genome Structural Variation

Previously reported sequences larger than 100 bp unique between the two genomes were defined as Structural variation. The A19 were aligned to the 9-941 genomes to determine their presence or absence. Ten unique fragments have been found. Three of the special sequences were located at the intergenic region. In the remaining seven special regions located in the gene coding region, two deletions sequence may resulted in changes in their gene function with virulence.

In *B.abortus* A19, a deletion region has led to the loss of 63bp sequence, which interrupts the coding regions of an ABC transporter permease. It is play an important role in maintaining the osmotic pressure balance inside and outside the cell, cell differentiation, antigen presentation, bacterial immunity and the like [20]. The deletion occurred on chromosome 2 of A19 at the position of 371693 to 371756, which its corresponding gene was *WP_002965788.1*. Compared with 9-941, A19 has sequence variation results a 31 amino acid removed of the ORF length, which may results its gene function changed from ABC transporter protein to hypothetical protein. According to reports, ABC transporter protein was predicted to related with virulence in the *Brucella* genome.

Using the above gene data with MEGA7, a phylogenetic tree was constructed, showing the evolutionary relationship of this gene between the sequenced *Brucella* genomes. Clustering includes vaccine and virulent strain of *B. abortus, B. suis* and *B. melitensis* data that can be clustered into different groups (Figure 4). This is consistent with other analyses showing that *B. abortus* A19 was most closely related to *B. abortus* S19, and A19 was more closely associated with *B. abortus* 9-941, 2308, A13334 than *B. suis* and *B. melitensis*. From the results of cluster analysis, it can be concluded that the genetic relationship between the different species and the classification of *Brucella* species are similar. However, there are different close relationships between vaccine strains and virulent strains of the same type of *Brucella*.

To further investigate the relationship of this deletion sequence in different species of *Brucella*, we performed sequence analysis of 9 complete genomes of the deleted sequence to provide more detailed information on changes in the genome region. As shown in Figure 5, the region marked with red box

represents the detailed sequence alignment of the deleted gene. As can be seen from the comparison results, the sequence was deleted in both the A19 and S19 genomes, and was present in the other *B. abortus, B. suis* and *B. melitensis*. However, these existing genes are identical in *B. suis* and *B. melitensis*, and there are some differences in *B. abortus*. This corresponds to the results of the close relationship we analyzed in Figure 4. We also compared this sequence with the *Brucella* genome-wide library on NCBI and found that there are many different species of *Brucella* that contain this sequence, so this deletion sequence on A19 is significant and it is related to virulence. It also provides a research hotspot for the study of subsequent attenuation mechanisms and the establishment of diagnostic methods.

In order to analysis the structural features of the above gene, a conservative motif was determined based on evolutionary relationship of the conserved motifs. Using the MEME database to search for its conserved sequences, a total of 20 conserved motifs were identified. These motifs were then further aligned in different types of *Brucella*. As shown in Figure 6, most of the identified *B. suis, B. melitensis* and *B. abortus* contain 20 motifs, except that these motifs were ordered differently in different types. The ordering in *B. suis* and *B. melitensis* was the same. Surprisingly, we found that motifs 17 and 18 were missing in A19 and S19. After the InterProScan search, we screened aligned sequences with an E-value of less than 0.05, and motifs 17 and 18 aligned the 23 and 14 nucleotide sequences, respectively, and labeled with Nuclear receptors with C4 zinc fingers and DNA-Binding of Transcription Factors. Taken together, these results indicate that most species have the same protein motif, further supporting their phylogenetic classification.

The zinc finger (ZnF) domain is a relatively small set of protein motifs that contain fingertips stabilized by zinc ions and function to bind DNA, RNA, proteins and lipid substrates. It is also widely involved in gene transcription and translation, mRNA transport and processing, and chromatin remodeling [21]. The Cys4 (C4) zinc finger motif (class II ZnF motif) consists of eight conserved cysteine residues bound to two zinc atoms, usually in a single form, and is involved in DNA binding, and in many Nuclear receptors are ubiquitous in transcription factors. The SZnF protein is mainly located in the cytoplasm and is also present in a small amount in the nucleus.

Proteins that read gene regulatory codes, transcription factors (TFs), have been extensively identified. Affinity for all possible DNA sequences is described based on the affinity and biochemical principles of the TF-DNA binding model. The overall analysis of the data indicates that the orientation and spacing preferences of homodimers and base stacking interactions play a large role in TF-DNA binding. In this model, transcription factors control gene expression by dynamically binding and generating partial possession of the same locus [22].

The high affinity binding protein-dependent ABC transporter composed of a high affinity periplasmic substrate binding subunit, two hydrophobic membrane subunits and two other cytoplasmic subunits. ATP hydrolysis by the related ATPase provides energy for the substrate to accumulate across the intima over a range of concentrations. The periplasmic substrate binding subunit consists of two separate but

similarly folded globular domains joined by a hinge region consisting of two or three short polypeptide fragments.

Next, in order to evaluate the possible effect of the gene on protein function, we performed a structural analysis of the amino acid sequence of ABC transporter protein, and constructed a three-dimensional structural model of this gene using Blast and MOD (Fig.7). It can be seen from the figure that the red region represents the deletion fragment of A19, and the tertiary structure analysis of the protein indicates that the structure of the deletion fragment is irregular curl. The three-dimensional structure of the protein in the blue region is an alpha helix and a random coil. Our analysis indicated that the deleted region is located within the alpha helix and irregular curl of the gene. Irregular curl is a loose peptide chain structure with no regularity, often the active site of the enzyme. Moreover, these sites are often important regions for the functional and conformational functions of protein molecules. Therefore, the *B. abortus* A19 vaccine lacking this fragment may result in the loss of the active region of its active center, resulting in loss of function. We hypothesized that this sequence affects protein activity, although further studies are needed to determine its possible link to A19 decay.

The second variable deletion region in A19 occured of Hypothetical protein compared to 9-941. Chrl of A19 also contained 63bp changes, the position of the missing sequence starts at 435999 and ends at 436062, which corresponds to gene *WP_002963568.1*. This gene encodes a hypothetical protein in 9-941, resulting in the deletion of 63 nucleotide sequences in the noncoding region of A19. We continued to observe the upstream and downstream genes of this gene and found that the upstream and downstream genes of the two genomes are identical, encoding ribbon-helix-helix protein%2C CopG family and hypothetical proteins, respectively. After a structural similarity search in the protein database, a major hit was found in the hypothetical protein BR0409 of *Brucella suis* 1330.

By comparison with other species of *Brucella*, the sequence was found in *B. suis* (1330, VBI22), *B. abortus* (9-941) and *B. melitensis* (16M, M28), and were missing in *B. abortus* vaccine A19, S19 and virulent strain 2308. Drawing the phylogenetic tree shows that the close relationship between A19 and 2308 is closer than that of 9-941, so we speculate that the virulence strains 2308 and 9-941 also have differences in virulence.

Genome Indels

Indels refers to the insertion or deletion of one or more nucleotides in one genome relative to another, which can be used as a sequence feature to characterize the evolution of different organisms [23]. Indels can have a dramatic effect on the gene, resulting in frameshift, extension or truncation of the encoded protein. Pairwise analysis identified that there were 47 indels in A19 genome compared to 9-941, 8 of which are located in the intergenic region. A total of 29 indels are located in 22 different ORFs of the coding region. We assume it results in changes in gene function. Of the 22 different genes, 3 are related to virulence in 9-941 and may missing their functions in A19. The specific Indels are shown in Table 2. These data suggest that mutations in these genes can play a central role in virulence attenuation mechanisms and future vaccine research.

A notable Indel was located in the *Ceg34* gene, and the function of its annotation is gamma-glutamyl-gamma-aminobutyrate hydrolase. Deletion of 16-nucleotide was found at the position 247339 of chromosome 2 in *Ceg34* gene. By aligning the results, the region between the two genes *WP_002965668.1* and *WP_002965669.1* in 9-941 was the intergenic region, which was located from 975 to 1077. In A19, the location of the two intergenic regions was from 975 to 1061. Due to the 16 bp deletion in A19, a hypothetical protein appeared in the interdigit region of 981-1076, may resulting in a change in gene function. To further explore the reason, we analyzed the upstream and downstream of *WP_002965669.1* gene. As shown in Figure 9, the blue arrow to the right represents the forward annotation result, while the yellow arrow to the left represents the reverse annotation result. It can be seen that the upstream and downstream of this gene did not change in A19 and 9-941, but only one hypothetical protein was added. Therefore, we speculate that the promoter only encodes and translates the hypothetical protein, which in turn inhibits the function of the WP_002965669.1 gene, rendering it incapable of expression.

Through the alignment of nucleotide and protein sequences, we conducted an in-depth study of the deleted sequence of A19. Surprisingly, it was found that the deleted 16-nucleotide sequence (16bp) is a tandem repeat in 9-941. And this sequence is located at the end of the gene *BRUAB_RS11615.* Two copies of this repeat appeared in the full length ORF of strain 9-941. A variant ORF or protein is produced by deleting a repeat of strain A19. The absence of *B.abortus* vaccine and virulent strains is different. We compared the S19 vaccine strain and found that S19 also contains only one copy. The possibility of this deletion may be related to the host preference of a particular species.

γ-glutamyl-γ-aminobutyrate hydrolase (PuuD) identified its active center as Cys-114 by site-directed mutagenesis. Expression of PuuD is induced by putrescine and O2 [24]. Strains lacking puuD accumulateγ-glutamyl -aminobutyric acid (g-Glu-GABA) and could not grow on the putrescine as the sole nitrogen source (Kurihara S and Oda S, 2005) [25]. This finding indicates that PuuD has important physiological significance of g-Glu-GABA hydrolase.

The protein sequence of the *BRUAB_RS11615* gene was compared to the VFDB database and aligned with the ceg34 gene belonging to the Dot / Icm Type IV secretion system. Intracellular replication of Chlamydia pneumoniae requires the Dot / Icm Type IV secretion system, which consists of approximately 27 proteins that may cross the bacterial membrane and the phagocytic membrane [26]. The ability of bacteria to survive in phagocytic hosts depends on the Dot / Icm Type IV secretion system (T4SS), which transports multiple effector proteins into host cells. The perturbation of host killing capacity is largely mediated by the collective function of the protein substrate injected into the host cell by the Dot/Icm transporter [27]. This gene may be a virulence factor for *Brucella* A19, and we speculate that this may be the reason for the virulence of the A19 vaccine.

Another 1bp insertion was occured in A19 resulting in loss of encoded protein of the amino acid ABC transporter permease compared to 9-941. The *BRUAB_RS09305* gene was located on the ChrI, and insertion of the C base at position 1901322. We further compared the upstream and downstream genes

of this gene. As can be seen from Fig. 10, the red position in the figure represents the position of base insertion, and the upstream and downstream of the *BRUAB_RS09305* genes are identical in 9-941 and A19. The *WP_002965020.1* gene in 9-941 became a hypothetical protein in A19 due to the insertion of the C base. The membrane-associated complex of Salmonella typhimurium periplasmic histidine permease is a member of the ABC transporter or transport ATPase superfamily, a copy of two integral membrane proteins HisQ and HisM and two ATP-binding subunits HisP composition [28]. HisP is absolutely essential for ATP hydrolysis and HisQ could not. HisP is dependent on HisQ to deliver an induction signal from the soluble receptor HisJ, and HisQ regulates the ATPase activity of HisP, while HisP changes conformation after exposure to phospholipids. This ensures the normal operation of the virulence factor ABC transport system.

In chromosome 2 of A19 genome, a 1 bp difference in nucleotide sequence. Insertion of an adenosine residue at position 215674 of the *BRUAB_RS00955* gene results in a frameshift. It may caused the changes in gene function of the alcohol dehydrogenase. Alcohol dehydrogenase D (*AdhD*) is a monomeric thermostable alcohol dehydrogenase derived from the protein aldose-keto reductase (AKR) superfamily [29]. *AdhD* has a fairly broad base specificity. And it is a metal ion-independent enzyme that has been isolated from thermophilic and archaea and successfully expressed in *E. coli*. It catalyzes the oxidation–reduction reaction of NAD / NADH linked short chain alcohols and aldehydes (or ketones). This enzyme is involved in the synthesis of ethanol. It has been reported that constitutive expression of the AADH2 gene can be used by Agrobacterium to produce 1-butanol, but overexpression of this gene prevents cell growth. The resulting *Adh* acts synergistically with other enzymes involved in 1-butanol metabolism, and *Adh* is an important and irreversible enzyme for the synthesis of 1-butanol [30]. This deficiency of ADH in the A19 strain does not prevent the growth of body cells. Therefore, we suspect that this mutation may be the cause of A19 virulence attenuation.

Previously, this study also compared the genomic differences between A19 and S19. It was found that in chromosome I of A19, a 1 bp insertion results in aldehyde dehydrogenase to become dehydrogenase. They are all a kind of dehydrogenase. But the function of this gene also exists in 9-941. Therefore, we speculate that there are dehydrogenases in other positions of the chromosomes of A19 and S19, which may be a factor that they are still virulent.

Genome SNPs

SNPs are powerful tools for describing the phylogenetic framework of species [31]. SNPs data will help develop novel high resolution molecular typing techniques for intra- and interspecies identification of pathogenic microorganisms. We compared the analysis of *B. abortus* vaccine A19 and virulent strain 9-941 to show that SNPs are widespread in these genes.

The distribution of SNPs between the two chromosomes was not equally, one hundred and forty seven SNPs (68%) and the remaining sixty-eight SNPs (32%) were located on chromosome I and chromosome II respectively. Among these total 215 SNPs, Fifty-six SNPs were synonymously substituted and encode the

same amino acid (aa). The 117 SNPs were missense substitutions, conserved, encoding different aa with similar properties; Two SNPs cause frame changes or protein coding regions to stop prematurely.

A total of 117 SNPs were present in 112 different genes, and these genes have changed their functions due to SNP. Among the 112 well defined genes, more than half of the genes encode various enzymes involved in the basic life cycle, such as RNA methyltransferase, nitrite/sulfite reductase, 6S rRNA-dimethyltransferase RsmA, 23S rRNA-methyltransferase RlmJ, arginyltransferase and CTP synthase. In addition, there were always 9 genes whose function may be related to virulence in 9-941, but in A19 the function of this virulence gene is missing. The correlation with virulence is described in the following section (Table 2).

Gene *BRUAB_RS00275* and *BRUAB_RS15760* encode LysR and IclR family transcriptional regulator respectively. The two regulatory genes in 9-941 were different from the A19 genes. The LysR regulatory gene located on chromosome 1, and another different *IclR* gene located on chromosome 2. Mutation of the *LysR* gene results in the replacement of arginine by leucine, and the *IclR* gene is a substitution of proline for serine. Some study describes *LysR* family transcriptional regulator in *B. abortus*, which activates abcR2 but does not activate abcR1, and more importantly, is critical for the ability of bacteria to successfully infect macrophages and mice [32]. The *IclR* gene affects stress survival and host infection in the zoonotic pathogen *Brucella abortus*. In one study, *LysR* and *GntR* mutants show reduced virulence [33]. Members of the *IclR, Crp, LysR* and *TetR* family of transcriptional regulators are involved in

virulence and symbiosis. Therefore, the lack of these two *LysR* and *IclR* regulatory genes may be the cause of A19 attenuation.

Gene *BRUAB_RS14580* encodes MFS transporter. The mutation site of this gene is located at position 875262 of chromosome 2, and its mutation causes arginine to replace histidine. It is worth noting that the major facilitator superfamily (MFS) transporter is located in the plasma membrane of the cell and is critical for the pathogenicity of plants and many other organisms (Michael Hohl and Sille Remm, 2019) [34]. Their active efflux is resistant to a variety of secondary metabolites and antibiotics. It can be a virulence factor by protecting pathogens from virulent compounds produced by host phagocytic cells.

Gene *BRUAB_RS04595* encodes outer membrane subunit *BepC*. Lipopolysaccharide (LPS), outer membrane proteins (OMP) and adhesins are all virulence factors of *Brucella ovis* and may be of greater importance in the host and in interaction with pathogens. Several OMPs associated with the virulence of *Brucella* include *Omp10, Omp16, Omp19, SP41* and *BepC*, which have previously been associated with the virulence of *Brucella*. And *BepC* is a TolC homologous protein that was found to be essential for the complete virulence of *S. suis* 1330 in a mouse model. Heparin Bartonella (*Bhe*) can invade human endothelial cells (EC) by invading body-mediated internalization as a large bacterial aggregate. The process depends on the functional type IV secretion system (T4SS) and the thereby translocated *Bep* effector proteins. *Bep* effector protein is one of the virulence factors of bacteria invading the body [35].

There are other three genes (*BRUAB_RS07715, BRUAB_RS14740* and *BRUAB_RS14975*) that encode ABC transporter protein, which their corresponding virulence genes are *hitB, ddrA* and *oppA* respectively. Bacteria can use peptides as a source of amino acids, nitrogen, carbon and energy. A variety of peptide uptake systems mediate the transport of peptides across the plasma membrane of bacteria. Recent studies have shown that *OppA* can preferentially select substrate peptides. *E. coliOppA* prefers a positively charged peptide with three or four amino acids in length. However, other studies have also shown that the substrate preferences of peptide uptake systems are more complex [36].

In addition, gene *BRUAB_RS06870* and *BRUAB_RS07680* were also related to virulence, the functions of the two genes are cell division protein *FtsA* and *Gfo/Idh/MocA* family oxidoreductase. In bacterial models, including *Streptococcus pneumoniae*, *Escherichia coli* and *Bacillus subtilis,FtsZ* and *FtsA* proteins assemble into loops at midcell and are specifically used for spacer peptidoglycan (PG) synthesis. Therefore, inactivation of *FtsZ* or *FtsA* results in the inability of filamentous cells to divide. Moreover, complete depletion of *FtsA* results in dislocation of the *FtsZ* loop and ultimately leads to cell expansion and lysis [37]. The Gfo / Idh / MocA protein family contains many different proteins, almost all of which are composed of NAD(P)-dependent oxidoreductases (or dehydrogenases), which have multiple substrates. The earliest mentioned Gfo / Idh / The history of the MocA family dates back to 2001, and the publication describes whole-genome sequencing of a virulent isolate of *Streptococcus pneumonia* [38].

In the comparison of SNPs, many virulence genes have their functions in 9-941, but are missing in A19. When comparing A19 and S19, we focused on the reason why A19 is more virulent than S19. Although there is no clear report on which of the two is more virulent, it is well known that the immune effect of S19 is better than that of A19. Some virulent genes have functions such as LysR family transcriptional regulator and ABC transporter permease. When these genes were exist in A19, the comparison with 9-941 is lost, but the comparison with S19 does exist. Therefore, it is first explained that these genes are related to virulence, and then it is also the reason why the A19 vaccine still has virulence. Finally, the number of functional genes in virulent strains and vaccines is also an important indicator for judging its virulence.

In summary, the genomes of A19 and 9-941 were very similar in sequence, organization and structure. Few fragments in the genome are unique. Even some genes are closely related to virulence. Our future research will focus on in vitro and in vivo infection studies using the above virulence genes to analyze their virulence characteristics with the aim of developing a new live attenuated vaccine for brucellosis in livestock.

PCR Verification

In this study, the *WP_002965788.1* gene screened was A19-specific deletion by genomic alignment, but exists in 9-941. Further BLAST alignment of this gene sequence in NCBI found that it was also deleted in S19, but it was present in most other commonly used *Brucella* vaccine strains and virulent strains. This experiment designed primers for this gene to verify (Table 3). Using PCR differential diagnosis, the sequence was amplified in the genomes of A19, S19, M28, S2 and Rev.1, respectively. The results showed that the *WP_002965788.1* gene was not amplified in the A19 and S19 genomes, and could be amplified in S2, M28 and Rev.1 (Figure 11). Indirectly indicates that the gene sequence can distinguish A19, S19 and most other commonly used *Brucella* viruent strains and vaccine strains, and the results are in line with expectations.

Vaccine immunization is the main measure for the prevention and treatment of brucellosis, which has been widely used in many herds in China. Different types of vaccines have differences in virulence, protection, immune protection period, applicable animal species, etc. The successful construction of PCR differential diagnosis method can effectively distinguish artificial immune(A19, S19) and natural infections of *Brucella*, so as to better purify the herd and prevent and control disease.

Conclusion

Brucellosis poses a serious hazard to global economic and public health problems. The *B. abortus* A19 strain was a spontaneously attenuated strain that has been used for vaccination against brucellosis [39]. In this study, we used a comparative genomics approach to perform genome-wide differential analysis and differential annotation for A19 and 9-941. Although the genomic similarity revealed a high identical and perfect genome-wide collinearity of the two strains, there are also a set of genomic structure differential was found.

Compared to 9-941, we found two deletion regions in the A19 genome that may be associated with virulence. The two regions are located in ORF *WP_002965788.1* and *WP_002963568.1*, each with a reduction of 63 amino acids. The function of gene *WP_002965788.1* is ABC transporter. We also analyzed the conserved sequence motif of this gene, and found two motifs deleted in A19, the corresponding structures are Nuclear receptors with C4 zinc fingers and DNA-Binding of Transcription Factors. At the same time, a three-dimensional structural model was constructed, and its missing fragments were predicted to be related to the active center.

Indel and SNP analysis revealed that non-synonymous and frameshift mutations of important virulence-related genes are associated with *LysR* family transcriptional regulator, outer membrane, MFS transporter, *Gfo/Idh/MocA* family oxidoreductase, *IclR* family transcriptional regulator, subunit *BepC*, γ-glutamyl-γ-aminobutyrate hydrolase and alcohol dehydrogenase *AdhP.* These mutations are compared with the VFDB database, and the obtained virulence genes are *YPA, adhD, ddrA, oppA* and *hitB* etc. We suggest that these genes be involved in the molecular mechanism of A19 attenuation.

After that, the deletion sequence of *WP_002965788.1* gene was amplified, and PCR method was used to distinguish A19, S19 and most other commonly used *Brucella* virulent strains and vaccine strains. As a result of the establishment of PCR differential diagnosis method, it provides data support for the subsequent distinction between virulent strains infection and artificial immunity.

Although some virulence-differentiating genes were found, the target genes were screened for the virulence-attenuating mechanism of the *B. abortus* A19 vaccine, but most of the identified genes and proteins have not yet been studied in *Brucella.* Therefore, subsequent experimental verification is imperative. Overall, the study proposes a set of candidate genes that may be associated with virulence and will help elucidate the mechanisms by which *Brucella* virulence attenuation. The comparison of the two genomic sequences of *Brucella* laid the foundation for further understanding of *Brucella* and gives us a thought for studying the contribution of various pathways to the relative pathogenicity and virulence of these bacteria. It also provides new data support and a reference for the establishment of diagnostic methods and the study of new vaccines.

Methods

Strain Information and Genomic DNA Preparation

The *B. abortus* A19 in sequencing was obtained from the Qilu Animal Health Products Co., Ltd. (Shandong, China). The DNA samples of A19 used in this study is previously deposited genome sequence in our laboratory [17]. Dissolve the A19 vaccine with 5ml of normal saline, centrifuge at 5000rpm for 5min, and then according to the manufacturers' instructions,genomic DNA of A19 was extracted and purified by the TIANamp Bacteria DNA Kit (Tiangen Biotech (Beijing) Co., Ltd., Beijing, China). After verifying that the DNA sample had not degraded (OD260/280 1.8-2.0, total amount is not less than 10 μg), 30 μg aliquots of DNA were analyzed using a Qubit 3.0 fluorometer (Q33216; Thermo Fisher Scientific).

The sequencing of A19 genome sequence has been done in the previous manuscript [17]. Screening differential genes, comparing gene structures and analyzing the virulence attenuation mechanism is done by Inner Mongolia Benniu Technology Co., Ltd.

Genome Sequencing and Assembly

The complete genome was sequenced using Illumina HiSeq 4000 platform. The technicians first fragmented the genomic DNA (300-500 bp) and an Illumina PE library was constructed with DNA adapters sequences (“A” and “B”) added to both ends of the DNA fragment. Amplification was subsequently performed by bridge PCR to generate DNA clusters. Then analyzed the small fragment DNA ranging from 8 to 10 kb and constructed a PacBio library in conjunction with a third generation single molecule PacBio technique. The quality of the obtained data is controlled.

Sequencing data was initially assembled using SOAPdenovo (v2.04) software, and then Pacbio sequencing data was compared using blasR. After completing all of the scaffold connections, the single molecule sequencing data was corrected based on the comparison results. Finally, the corrected single-molecule sequencing data was used to assembled, the scaffold connection using the overlap relationship between the sequences, and the subsequent assembly was performed using Celera Assembler 8.0 software. After completing all the scaffold connections, the Illumina data is used again for verification, and the gap closing work is performed to partially assemble and optimize the assembly results.

Gene Prediction and Annotation

The complete genomic sequence of the *B.abortion* A19 vaccine has been uploaded to GenBank under accession number CP030751-CP030752, which corresponds to two circular chromosomal sequences. Gene prediction was established using Glimmer 3.02 software. We classified the protein sequence of the predicted gene with the GO, string, gene and Nr databases, respectively (BLAST 2.2.30+, alignment standard: e-value < 1e-5), and annotation information was obtained.

Comparative Genomics Analysis

The Pairwise comparison and alignment between *B. abortus* A19 vaccine and virulent strain 9-941 was conducted using the progressive Mauve algorithm. The Illumina sequence for the 9-941 whole genome are download from NCBI database, with the accession number of GCA_000008145.1. Colinearity analysis of A19 and 9-941 was performed using MCScanX software (Wang Y et al., 2012), and collinear results were drawn by circos [40]. Sequence alignment of the phylogenetic tree was performed using MEGA7.

Structural variations between the A19 and 9-941 genomes were screened by pairwise comparisons and alignments. Using the MUMmer and Samtoolsr software, the *B. abortus* A19 vaccine was used as a reference to detect detailed changes in SNP and Indel between genomes, to screen for genome-wide SNPs and InDels, and to combine SNP and InDel position information and annotation information of A19, which is annotated and analyzed. The selected genes were compared with the experimentally validated *Brucella* virulence gene database VFDB (pathogenic factor of pathogenic bacteria) to find the virulence

difference genes of A19 and 9-941. The MEME program was used to identify conserved protein motifs and was further annotated with InterProScan [41].

Primer Design and Amplification

Based on the genomic alignment, the specific deleted sequences from A19 and S19 were synthesized by primers, and the restriction sites *EcoR* I and *Xho* I were designed into the primer sequences.

The PCR reaction conditions: pre-denaturation at 95 °C for 5 min; denaturation at 94 °C for 1 min; annealing at 60 °C for 1 min; extension at 72 °C for 1.5 min; 33 cycles; continued extension at 72 °C for 10 min. The PCR product was obtained by PCR amplification and detected by gel electrophoresis.

Abbreviations

*B. abortu*s
Brucella abortus; bp:Base pair; ORF:open reading frame; LPS:lipopolysaccharide; ABC:ATP-binding cassette; ery:erythritol; NCBI:National Center of Biotechnology Information;

Declarations

Ethics approval and consent to participate

All participants provided written informed consent to participate in the study.

Consent for publication

Not applicable

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Agricultural Project of Inner Mongolia Science and Technology Agency_(No.20120244), Project of Inner Mongolia Science and Technology Agency_(No.20140174) and Project of Inner Mongolia Higher Education Researc Agency_(No.NJZZ19041). Funders had no role in research design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

WSY, WWL, and ZXL conceived and designed the experiments. WSY performed the experiments. WSY, WWL, HHBT, SK and ZXL analyzed the data. WSY wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgment

We thank the Inner Mongolia Benniu Technology Co., Ltd for providing performed whole genome sequencing services.

Author details

1Key Laboratory of Clinical Diagnosis and Treatment Technology in Animal Disease, Ministry of Agriculture/ College of Veterinary Medicine, Inner Mongolia Agricultural University,Hohhot, Inner Mongolia, 010018, China

2Inner Mongolia Autonomous Region Comprehensive Center for Disease Control and Prevention, Hohhot, Inner Mongolia, 010031, China

*Corresponding author wwl.imau@163.com (W.W.)

References

1. Ebrahimpour S, Bayani M, Moulana Z, Hasanjani Roushan MR. Skeletal complicati ons of brucellosis: A study of 464 cases in Babol, Iran. Caspian J Intern Med. 2017;8(1):44–8.
2. Tuon FF, Gondolfo RB, Cerchiari N. Human-to-human transmission of *Brucella* - a sy stematic review. Trop Med Int Health. 2017;22(5):539–46.
3. Xu-ming Li Yao-xia, Kang L, Lin, et al. Genomic Characterization Provides Ne w Insights for Detailed Phage-Resistant Mechanism for *Brucella abortus.* Front Mic ro biol. 2019; 10(3): 917.
4. Dabral N, Martha-Moreno-Lafont, Sriranganathan N. Oral immunization of mice wit h gamma-irradiated *Brucella* neotomae induces protection against intraperitoneal and intranasal challenge with virulent *B. abortus* 2308. PLoS ONE. 2014;9(9):e10718 0.
5. Zhan BD, Wang SQ, Lai SM, Lu Y, Shi XG, Cao GP, et al. Outbreak of occupati on al Brucellosis at a pharmaceutical factory in Southeast China. Zoonoses Public Hea lth. 2017;64(4):431–7.
6. Hasanjani MR, Bayani M, Soleimani SA, Mohammadnia-Afrouzi M, Nouri H, Ebra him pour S. Evaluation of CD4 + CD25 + FoxP3 + regulatory T cells during treatme nt of patients with brucellosis. J Biol Regul Homeost Agents. 2016; 30(3): 675 – 68 2.
7. Roushan MRezaH, Ebrahimpour S, Afshar ZM, et al. Cervical Spine Spondylitis with an Epidural Abscess in a Patient with Bruce llos is: A Case Report. The Journal of Critical Medicine. 2019;5(3):103–6.
8. Florencia M, González G, Sycz IM. Alonso Paiva1, et al. The BtaF adhesin is necessary for full virulence during respiratory infection by *Brucella suis* and is a novel immunogen for Nasal

vaccination against *Brucella* Infection. Front I mmunol. 2019; 6(26): 1–10.

9. Elaine MS, Dorneles N, Sriranganathan AP. Lage. Recent advance s in *Brucella aboutus* vaccines. Vet Res. 2015;46(1):76.
10. Dorneles EM, de Faria AP, Pauletti RB, Santana JA, Caldeira GA, Heinemann MB, Titze-de-Almeida R, Lage AP. Genetic stability of *Brucella abortus* S19 and RB51 vaccine strains by multiple locus variable number tandem repeat analysis (MLVA16). Vaccine. 2013;31(42):4856–9.
11. Wattam AR, Williams KP, Snyder EE, et al. Analysis of ten *Brucella* geno mes reveals evidence for horizontal gene transfer despite a preferred intracellular l ifestyle. J Bacteriol. 2009;191(11):3569–79.
12. Scholz HC, Vergnaud G. Molecular characterisation of *Brucella* species. Rev Sci Tech. 2013;32(1):149–62.
13. DING jia-Bo FENG, Zhong-Wu. Current application of brucellosis vaccine and its reaearch advances. Chinese Bulletin of life Sciences. 2013;25:91–9.
14. Thomas EL, Bracewell CD. Characterisation of *Brucella abortus* strain 19 cultures isolated from vaccinated cattle. Vet Rec. 1981;108(5):90–3.
15. Perkins SD, Smither SJ, Atkins HS. Towards a *Brucella* vaccine for humans. FE MS Microbiol Rev. 2010;34(3):379–94.
16. Schurig GG, Sriranganathan N. Corbel MJ. Brucellosis vaccines: past, present and future. Vet Microbiol. 2002;90(1–4):479–96.
17. Shuyi Wang W, Wang, et al, Comparative genomic analysis between newly sequenced *Brucella abortus* vaccine strain A19 and another Brucella abortus vacci ne S19.Genomics, 2020; 112: 1444–1453.
18. Janga SC, Moreno-Hagelsieb G. Conservation of adjacency as evidence of paralog ous operons. Nucleic Acids Res. 2004;32(18):5392–7.
19. Zhong Z, Wang Y, Xu J, Chen Y, Ke Y, Zhou X, Yuan X, Zhou D, Yang Y, Yang R, Peng G, Jiang H, Yuan J, Song H, Cui B, Huang L, Chen Z, et al. Parallel gene loss and acquisition among strains of different *Brucella* species and biovars. J Mic robiol. 2012;50(4):567–74.
20. Teane MA, Silva TA, Paixao EA, Costa, et al. Putative ATP-Bindi ng Cassette Transporter Is Essential for *Brucella ovis* Pathogenesis in Mice. Infect Immun. 2011;79(4):1706–17.
21. Pohlmann T, Baumann S, Haag C, Albrecht M, Feldbrugge M. A FYVE zinc f inger domain protein specifically links mRNA transport to endosome trafficking. Elife. 2015;4:e06041.
22. Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, Bjo¨rklund M, Wei G, Yan J, Niittyma¨ki I, et al. The common colorectal c ancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enh anced Wnt signaling. Nat Genet. 2009;41(8):885–90.
23. Kamneva OK, Liberles DA, Ward NL. Genome-wide influence of indel substitutions on evolution of bacteria of the PVC superphylum, revealed using a novel computational method. Genome Biology Evolution. 2010;2:870–86.

24. Shin Kurihara S, Oda H, Kumagai. Gama-Glutamyl-gama-aminobutyrat e hydrolase in the putrescine utilization pathwayofEscherichia coli K-12. FEMS M icrobiol. 2006; 256(2):318–323.
25. Kurihara S, Oda S, Kato K, Kim HG, Koyanagi T, Kumagai H, Suzuki H. A novel putrescine utilization pathway involves g-glutamylated intermediates of Esch erichia coli K-12. J Biol Chem. 2005;280(6):4602–8.
26. Bo Hu. Pratick Khara, Liqiang Song. In Situ Molecular Architecture of the Helico bacter pylori Cag Type IV Secretion System. mBioasmorg. 2019;10(3):e00849-1 9.
27. Hayley J, Newton, Justin A, McDonough. Effector Protein Translocation by the C oxiella burnetii Dot/Icm Type IV Secretion System Requires Endocytic Maturation of the Pathogen-Occupied Vacuole. PLOS ONE. 2013;8(1):e54566.
28. PEI-QI LIU, GIOVANNA FERRO-LUZZI AMES. In vitro disassembly and reasse mbly of an ABC transporter, the histidine permease. Proc Natl Acad Sci U S A. 1998;95(7):3495–500.
29. Kusum, Solanki. Walaa Abdallah and Scott Banta. Extreme makeover: Engineering the activity of a thermostable alcohol dehydrogenase (AdhD) from *Pyrococcus furi osus*. Biotechnology journal. 2016;11(12):1483–97.
30. Marion Rauter J, Kasprzak K, Becker. Aadh2p: an Arxula adeninivorans a lcohol dehydrogenase involved in the first step of the 1–butanol degradation path way. Microb Cell Fact. 2016;15(1):175.
31. Brito PH, Edwards SV. Multilocus phylogeography and phylogenetics usin g sequence-based markers. Genetica. 2009;135(3):439–55.
32. Valerie Haine A, Sinon F, Van Steen. Systematic Targeted Mutagenesi s of *Brucella melitensis* 16M Reveals a Major Role for GntR Regulators in the Control of Virulence. Infect Immun. 2005;73(9):5578–86.
33. Lauren M, Sheehan JA, Budnick. Catlyn Blanchard. A LysR-family transcrip tional regulator required for virulence in *Brucella abortus* is highly conserved am ong the α-proteobacteria. Mol Microbiol. 2015;98(2):318–28.
34. Hohl M, Remm S, Eskandarian HA. Increased drug permeability of a stiffened mycobacterial outer membrane in cells lacking MFS transporter Rv1410 and lipoprotein LprG. Mol Microbiol. 2019;111(5):1263–82.
35. Matthias C. Truttmann1, P, Guye C, Dehio. BID-F1 and BID-F2 Dom ains of Bartonella henselae Effector Protein BepF Trigger Together with BepC the Formation of Invasome Structures. PLoS ONE. 2011; 6(10): e25106.
36. Klepsch MM, Kovermann M, Low C, Balbach J, Permentier HP, Fusetti F, de Gi er JW, Slotboom DJ, Berntsson RP. Escherichia coli peptide binding protein Opp A has a preference for positively charged peptides. J Mol Biol. 2011;414(1):75–85.
37. Mura A, Fadda D, Perez AJ. Roles of the Essential Protein Fts A in Cell Growth and Division in *Streptococcus pneumoniae*. Journal of Bacteriol ogy. 2017;199(3):e00608-16.
38. Pichoff S, Lutkenhaus J. Tethering the Z ring to the membrane through a conserv ed membrane targeting sequence in FtsA. Mol Microbiol. 2005;55(6):1722–34.

39. Tian G, Zhan Z, Zhang A. A case report on mother-to-child trans mission of *Brucella* in human, China. BMC Infectious Disease. 2019;19(1):666.
40. Krzywinski M, Schein J, Birol I, et al. Circos: An information aesthetic for comp arative genomics. Genome Res. 2009;19(9):1639–45.
41. Mulder N, Apweiler R, InterPro. and InterProScan: tools for protein sequence cla ssification and comparison. Methods Mol Biol. 2007;396:59–70.

Figures

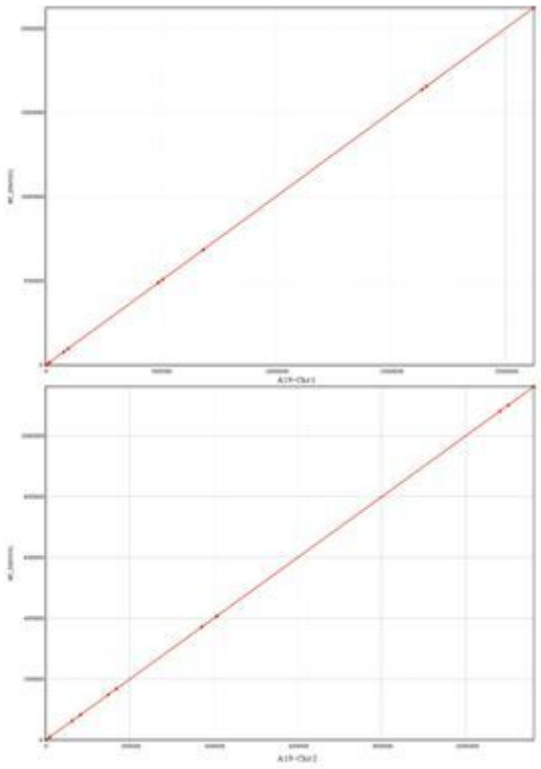

Figure 1

Comparative genomic analysis of the B.abortus A19 with 9-941 using MUMmer.(Red dots indicate missing sequences; straight lines indicate complete genome alignment)

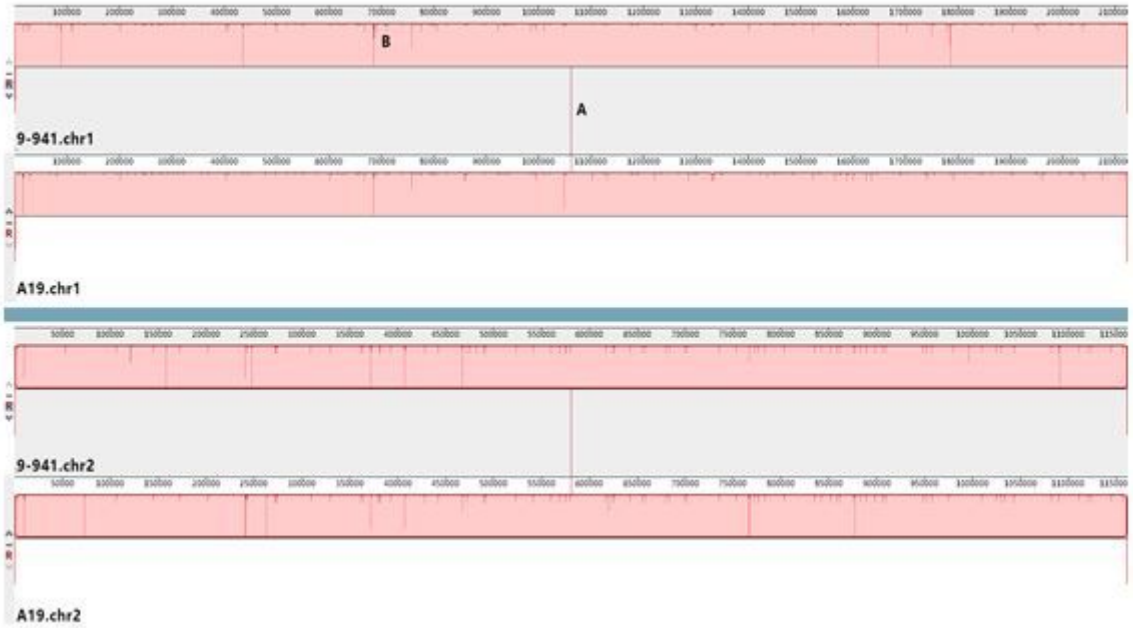

Figure 2

Whole-genome alignment of the B. abortus A19 and virulent strain 9-941.(A：These lines indicate which regions in each genome are homologous; B：These lines indicate which site in each regions are variation.)

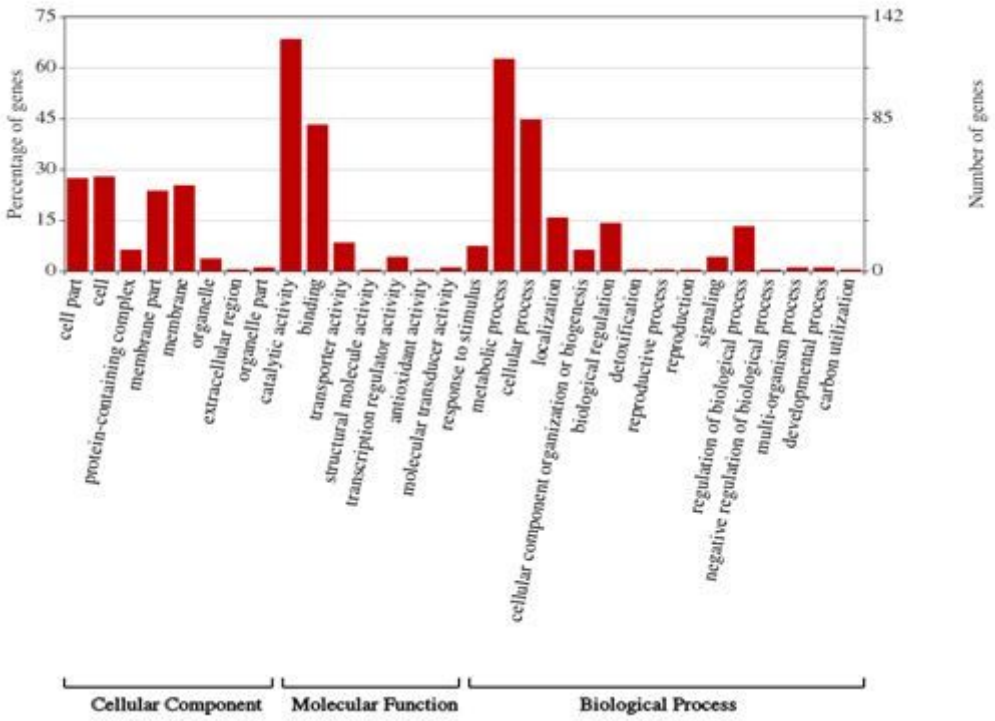

Figure 3

GO-based functional categories of SNPs between the B. abortus A19 and

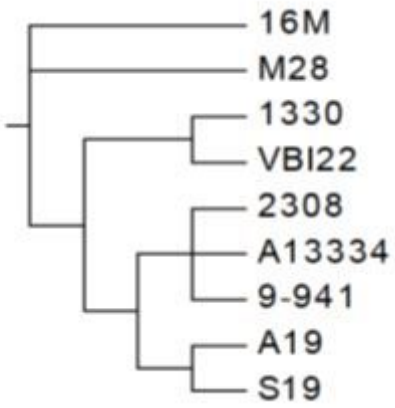

Figure 4

Phylogenetic Tree was built using homologous genes from 9 Brucella genomes in gene WP_002965788.1 using Neighbor-joining method in MEGA7.

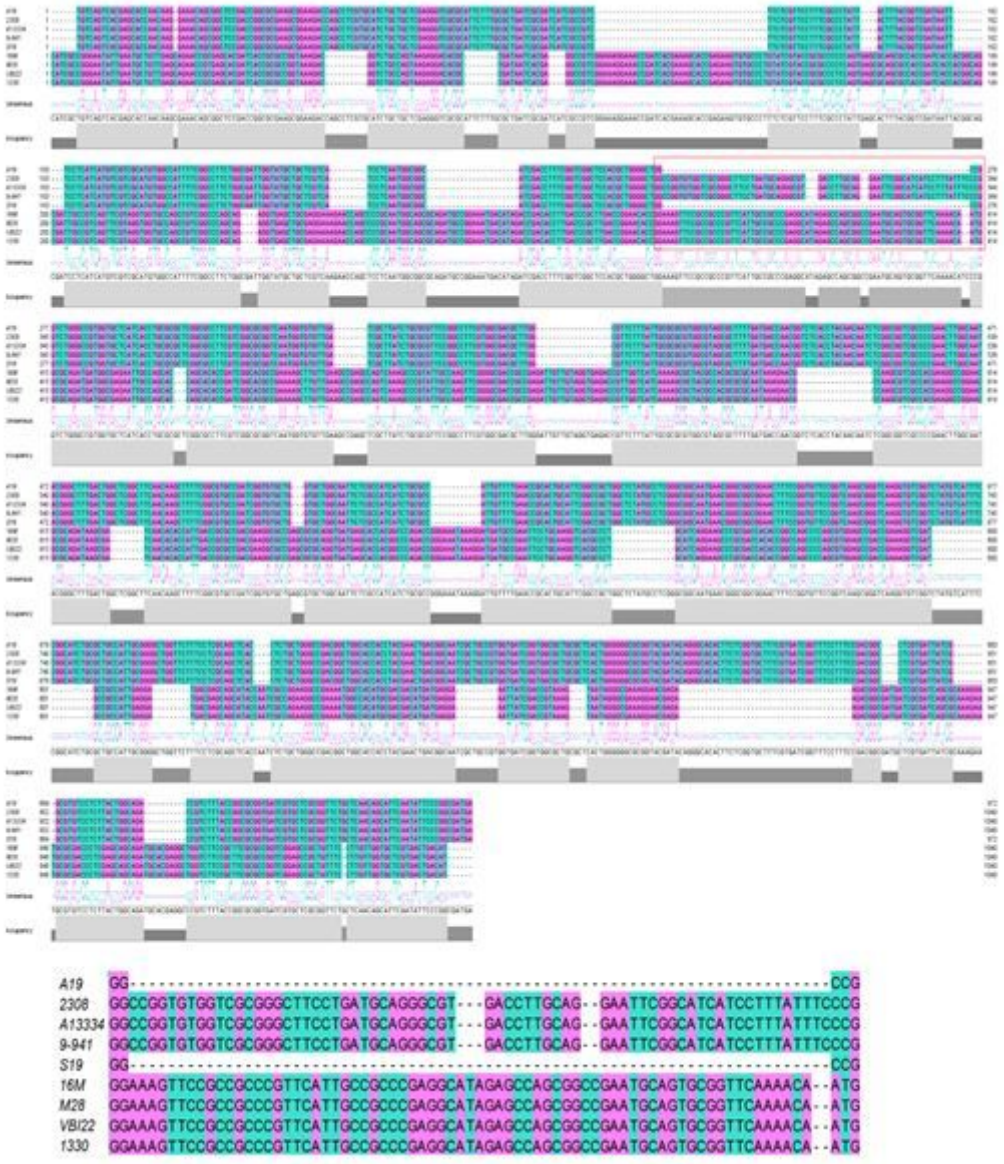

Figure 5

Sequence alignment of gene WP_002965788.1 in different Brucella genomes.

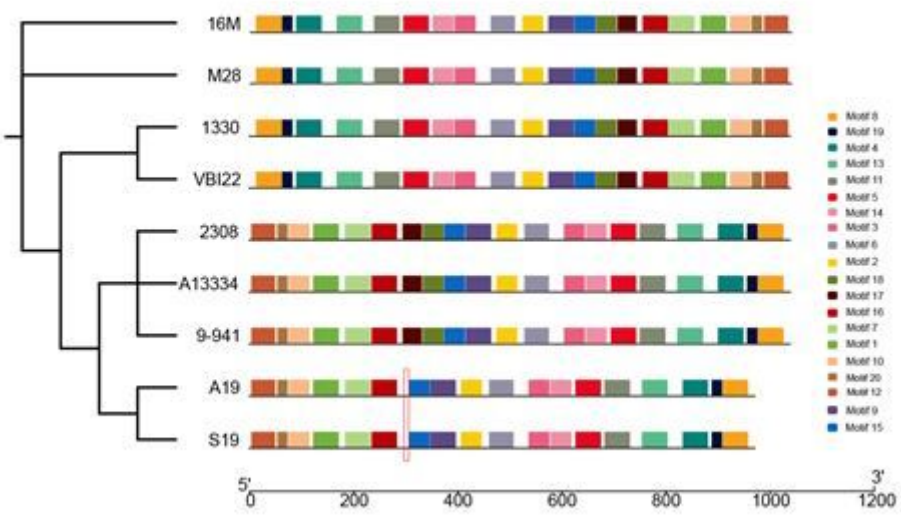

Figure 6

The conserved motifs of gene WP_002965788.1 according to phylogenetic relationship. All motifs were identified by MEME database with the complete amino acid sequences.

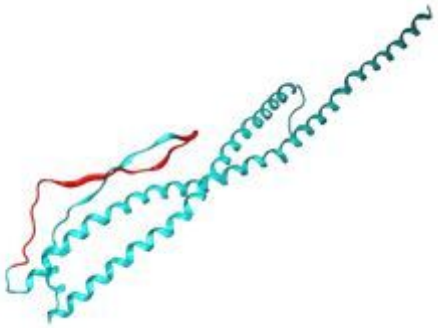

Figure 7

Regions of insertions within the ABC transporter proteins of the A19 strain Constructed a three-dimensional structural model of WP_002965788.1 gene using Blast and MOD

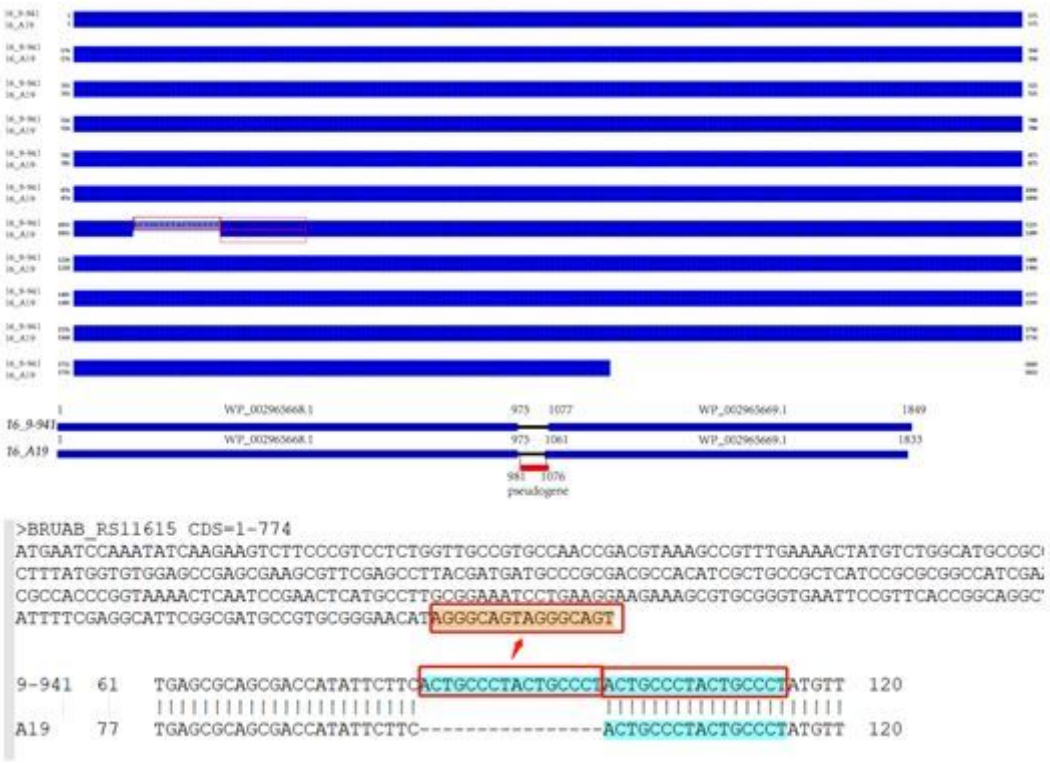

Figure 8

Nucleic acid alignment of the BRUAB_RS11615 genes between B. abortus A19 and 9-941.

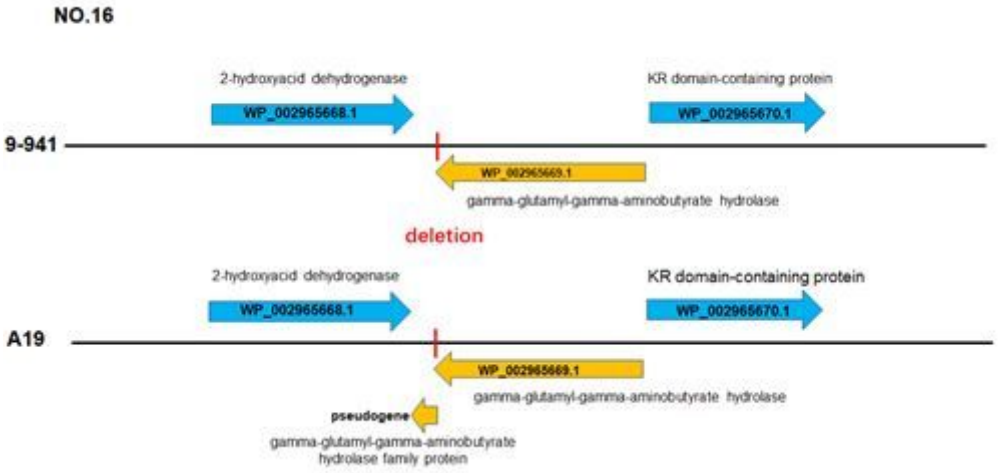

Figure 9

Upstream and downstream alignment of the BRUAB_RS11615 between A19 and 9-941.

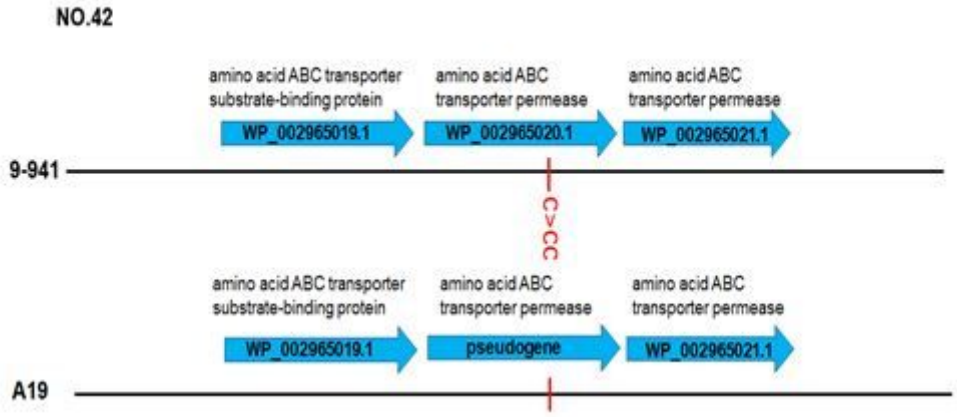

Figure 10

Upstream and downstream alignment of the BRUAB_RS09305 between A19 and 9-941

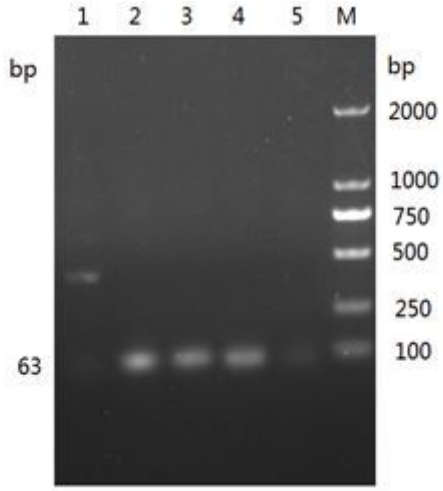

Figure 11

Agarose gel electrophoresis of A19 and S19 deletion gene □M DNA Marker DL2000□1 S19□2 M28□3 Rev.1□4 S2□5 A19□