RESEARCH PAPER

Virulence 6:8, 745–754; November/December 2015; ©2015 Taylor & Francis Group, LLC

# Comparative genomic analysis of *Brucella abortus* vaccine strain 104M reveals a set of candidate genes associated with its virulence attenuation

Dong Yu[1,†], Yiming Hui[2,†], Xiaodong Zai[1], Junjie Xu[1], Long Liang[1], Bingxiang Wang[2,*], Junjie Yue[1,*], and Shanhu Li[1,*]

[1]Beijing Institute of Biotechnology; Beijing, PR China; [2]LanZhou Institute of Biological Products; Lanzhou, PR China

[†]These authors equally contributed to this work.

**Keywords:** *Brucella abortus*, comparative genomics, vaccine, virulence, virulence attenuation

The *Brucella abortus* strain 104M, a spontaneously attenuated strain, has been used as a vaccine strain in humans against brucellosis for 6 decades in China. Despite many studies, the molecular mechanisms that cause the attenuation are still unclear. Here, we determined the whole-genome sequence of 104M and conducted a comprehensive comparative analysis against the whole genome sequences of the virulent strain, A13334, and other reference strains. This analysis revealed a highly similar genome structure between 104M and A13334. The further comparative genomic analysis between 104M and A13334 revealed a set of genes missing in 104M. Some of these genes were identified to be directly or indirectly associated with virulence. Similarly, a set of mutations in the virulence-related genes was also identified, which may be related to virulence alteration. This study provides a set of candidate genes associated with virulence attenuation in B.abortus vaccine strain 104M.

## Introduction

*Brucella*e are Gram-negative, facultative intracellular bacteria that can cause brucellosis in many animals and humans.[1] Brucellosis is among the most common zoonotic infectious disease epidemics worldwide, leading to great economic and public health problems,[2] particularly in developing countries. Human brucellosis is transmitted through direct animal contact and the consumption of contaminated food products of animal origin. Over the past 100 years, human brucellosis has been controlled by vaccination and culling animals,[2-4] but more than 500,000 new cases are reported annually worldwide.[5] China is a major stock-raising country. Over the last decade, human brucellosis has increased quickly, and human cases have been reported in all provinces.[6] This disease causes chronic infections with common animal outcomes of abortion and sterility. In humans, it can cause a severe acute fever and a febrile illness if untreated, producing focal lesions in joints, the genitourinary tract, and other organs.[7]

The most rational approach for preventing and controlling brucellosis is vaccination. Some live, attenuated vaccine strains are useful for the control and elimination of animal infections to decrease the rate of human brucellosis, such as *Brucella abortus* S19[8,9] and RB51,[10,11] which are used in cattle, and *Brucella* melitensis Rev1[12] and M5,[13] which are used in sheep and goats.

In addition to these vaccine strains, *Brucella abortus* (*B. abortus*) 104M [14] is a vaccine strain that is used in cattle and is widely applied for brucellosis prevention in China. This strain was first isolated from a sick cow's placenta in a former Soviet republic. It exhibits typical properties of biotype I, low and stable virulence on experimental animals, and a strong immunogenicity.[15-18] In certain conditions, the strain 104M was used as a live attenuated vaccine for humans. After 1965, the Chinese FDA approved the live, attenuated strain of 104M as a vaccine for use in human. Though these vaccine strains play an irreplaceable role in the control and prevention of brucellosis, the underlying molecular and physiological mechanisms causing attenuated virulence are unclear. Recently, several comparative genomic analyses have been conducted to detect the factors affecting virulence in *Brucella* vaccine strains, such as S19[19] and M5,[13] but none were related to 104M.

We sequenced the *Brucella abortus* vaccine strain 104M genome with the goal of elucidating the genetic mechanisms underlying virulence attenuation. As a result, a set of genes, which were associated with virulence or virulence attenuation, was identified through the comparison of this newly sequenced genome with its virulent counterparts.

## Results and Discussion

### General genome features

The complete genome of the vaccine strain, 104M, was sequenced as described in the Material and Methods, and 2

*Correspondence to: Bingxiang Wang; Email: yd2011start@gmail.com; Junjie Yue; Email: yue_junjie@126.com; Shanhu Li; Email: leeshanhu@hotmail.com
Submitted: 10/22/2014; Revised: 04/01/2015; Accepted: 04/01/2015
http://dx.doi.org/10.1080/21505594.2015.1038015

**Table 1.** Genome properties of the newly sequenced *B. abortus* vaccine strain 104M genome in comparison with the known genome sequences of 3 strains

| | B. abortus 104M | | B. abortus A13334 | | B. abortus S19 | | B. abortus 9–941 | | B. abortus 2308 | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Feature** | **Chr1** | **Chr2** | **Chr1** | **Chr2** | **Chr1** | **Chr2** | **Chr1** | **Chr2** | **Chr1** | **Chr2** |
| Size (Mbp) | 2.12 | 1.16 | 2.12 | 1.16 | 2.12 | 1.16 | 2.12 | 1.16 | 2.12 | 1.16 |
| ORFs | 2098 | 1094 | 2185 | 1153 | 2005 | 1057 | 2030 | 1055 | 2000 | 1182 |
| tRNA | 40 | 15 | 40 | 14 | 41 | 14 | 41 | 14 | 41 | 14 |
| rRNA | 7 | 3 | 6 | 3 | 6 | 3 | 6 | 3 | 6 | 3 |
| GC(%) | 57.2 | 57.3 | 57.2 | 57.3 | 57.2 | 57.3 | 57.2 | 57.3 | 57.2 | 57.3 |
| Average length | 859.5 | 917.8 | 830.9 | 883.5 | 890.4 | 942.6 | 845.7 | 900 | 852 | 904.2 |

complete chromosome sequences were obtained: one is 2122848 bp long, and the other is 1162581 bp long. The average GC content of the 2 chromosomes is 57.2%. This newly sequenced genome was annotated using the RAST server,[20] and 3192 protein-encoding genes, 65 RNA genes, including 10 rRNAs and 55 tRNAs, were identified. The detailed genome properties and 3 other *Brucella abortus* genomes are listed in **Table 1**. The genome of strain 104M showed remarkable similarity in various properties to those of its relatives, *B. abortus* A13334, S19, 9–941 and 2308.

A multiple genomes alignment was constructed among 104M and the other sequenced *Brucella* genomes to detect this newly sequenced strain's phylogenetic relationships. A maximum likelihood phylogenetic analysis of the 18 strains using SNP data was conducted, as shown in **Figure 1**. All *Brucella* strains were divided into 5 clades, as previously reported:[21] *B. melitensis-B. abortus* clade; *B. ovis* clade; *B. pinnipedialis* clade; *B. microti* clade; and *B. suis-B. canis* clade. Each internal node received 100% bootstrap support. In the tree, strain 104M was located in the *B. abortus* sub-clade and diverged earlier than other *B. abortus* strains. Despite an evolutionary branching order, strain 104M appeared to be very close to other *B. abortus* strains for the short internal branch length.

A genome-wide comparison was conducted among 104M and its relatives, *B. abortus* A13334, 9–941, S19 and 2308, to detect the whole-genome similarities. The result is showed in **Figure 2**: 104M exhibits a perfect genome-wide colinearity with other B. abortus strains. The only difference is the location of blue block in each genome, which could be due to the different starting loci of genome sequencing. Thus, further comparisons in gene content between 104M and other strains were needed. The results, including the number of strain-specific genes, were shown in **Figure 3**. It is clear to see that the smallest differences lie in the comparison of 104M and A13334.

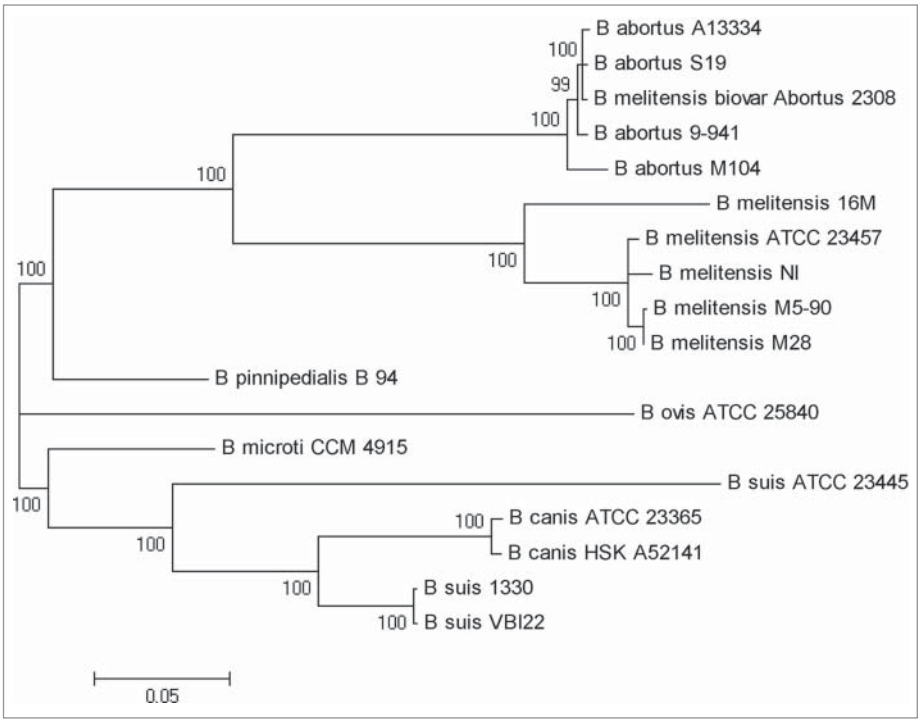


**Figure 1.** Phylogenetic tree of 18 *Brucella* strains. The maximum likelihood tree was constructed using SNP data. Almost all nodes received 100% bootstrap support and were labeled on each node.

The virulent strain A13334 is derived from the foetal gastric fluid of an infected dairy cow in Gyeonggi, Yangpyeong, Republic of Korea,[22] which is near China. Considering the high similarity in genome and close distance in geography, we conducted the comparative genomic analysis between 104M and A13334 to detect the virulence or virulence-attenuated associated genes.

### Identification of virulence-associated gene missing between 104M and A13334

Although the genome alignment results showed high similarity between the genomes of 104M and A13334, there are genetic differences leading to the attenuated virulence in 104M. The main target of this work was to identify all of the genetic differences between 104M and its virulent relative A13334, which would provide a complete basis for the virulence attenuation. Pairwise and reciprocal comparison was conducted between the 2 proteomes to identify genes that are 100% identical,


**Figure 2.** Mauve alignment of both chromosomes from the 5 complete *Brucella abortus* genomes. Each genome is laid out horizontally and homologous segments are shown as colored blocks: red and olivine blocks constitute the sequence of chromosome I, and blue and green blocks constitute the sequence of chromosome II.

non-identical with any differences and unique in either of the strains. A total of 99 and 52 genes, as showed in **Figure 3**, were identified to be unique in A13334 and 104M, respectively. The remaining genes in either of the genomes had their corresponding homologs in the other genome.

### Gene loss associated with virulence

The unique genes in either of the genomes might imply the loss of the same genes in the other genome, suggesting a set of potential targets leading to the 2 strain's virulence differences. In addition to deletion and horizontal gene transfer (HGT), the other possible reasons for the existence of the unique genes could be de novo gene formation or frameshift mutation. However, the unique genes in either genome did not show a continuous distribution along the genome. Considering the close evolutionary relationship between the 2 strains, they may not be the result of repeated gene gain, gene loss or de novo gene formation events. Additionally, the gene sequences were used to perform a genes-to-genome search, and nearly all of the unique genes were identical to those in the corresponding genome. Thus, the unique genes might be the results of frameshift mutations.

However, 62 strain-specific genes in A13334 and 45 strain-specific genes in 104M were annotated as hypothetical. For the remaining genes with definite functional assignments, literature searches were conducted to determine whether the genes were related to virulence.[23,24]

For the A13334-specific genes, the detailed functional annotation for each gene was listed in **Table 2**. Among the 37 A13334-specific genes with definite function assignment, more than half encode various enzymes that are involved in the fundamental life cycle, such as ATP-dependent ligase, diguanylate cyclase and fumarase. Among them, several genes might be associated with virulence.

Gene BAA13334_I03390 encodes a superoxide dismutase, which can directly detoxify the reactive oxygen intermediates (ROIs) in bacteria. The ROIs, including superoxide, hydrogen peroxide and hydroxyl radicals, damage macromolecular structure and macrophages and utilize this process to limit the intracellular replication of the *Brucellae*.[25,26] Thus, the absence of this gene in 104M could have an effect on the intracellular replication, which might indirectly decrease the strain's virulence.

Gene BAA13334_I01801 (*ureF*) encodes a urease accessory protein. In bacteria, ureases are multi-subunit metalloenzymes that hydrolyse urea to form carbonic acid and 2 molcules of ammonia, the latter of which could cause the pH to increase.[27] Recent studies have found that urease protects *B. abortus* during their passage through the stomach when acquired by the oral route in BALB/c mice, which is the major route of infection in human brucellosis.[28] Urease accessory protein is essential for

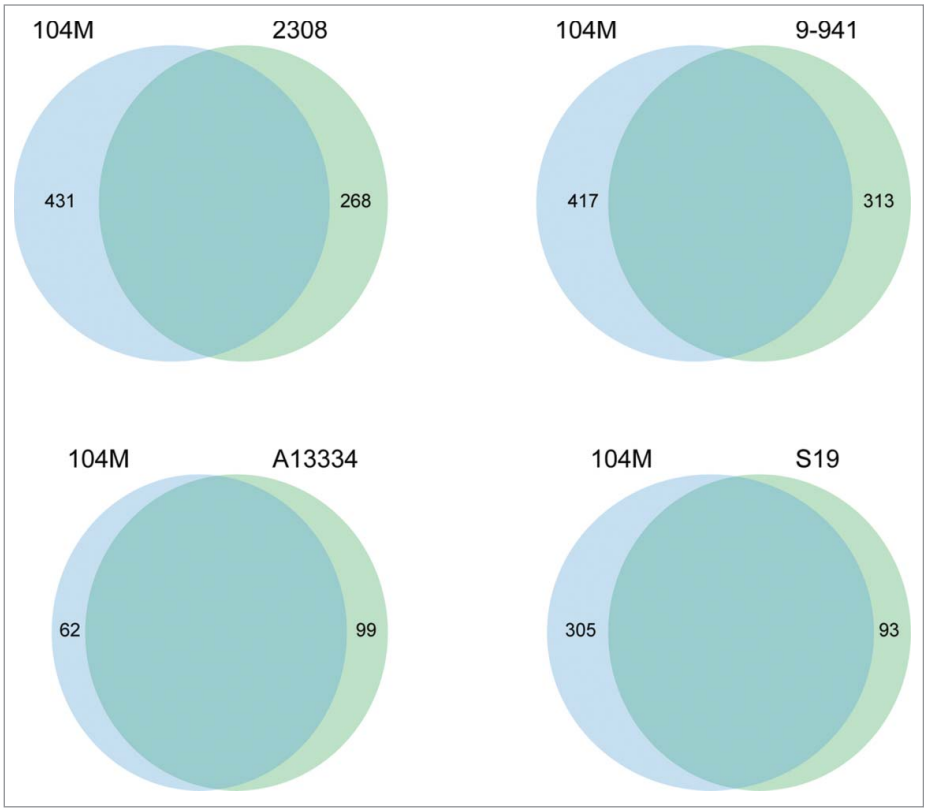


**Figure 3.** The Venn diagrams showing the distribution of strain-specific genes between the genomes of 104M and other B.abortus strains.

incorporation of nickel into the active center of urease.[29] At least 3 urease accessory proteins, including UreF, are required for the activation of urease.[30-32] Thus, the lack of *ureF* can lead to the defective in urease activity, which might contribute to the attenuated virulence in 104M.

Gene BAA13334_I03416 encodes a 2-component transcription regulator, belonging to the LuxR family. The two component regulatory system in *Brucella* is predicted to be essential for sensing the phagosomal environment and changing from an extracellular to intracellular life style.[33] Recent experiments showed that the *bvrS/bvrR* mutants are attenuated in mice and show reduced invasiveness in cells.[33,34] Though this gene is not affiliated with the system, it is related with the system. The absence of this gene in 104M might have an effect on 2-component regulatory system functions.

In addition to the enzyme-coding genes, there are some genes encoding membrane proteins or transporters. For example, 2 genes (BAA13334_I00134 and BAA13334_II01500) encode the ABC transporter, which can play important roles in the *Brucella* strain virulence and influence the *Brucella* trafficking to compartments associated with the endoplasmic reticulum.[35] There are also other genes encoding the $Mg^{2+}$ and $Co^{2+}$ transporter, nitrite transporter and sulfate permease, all of which are involved in iron transport. These genes might not be directly related to the *Brucella* virulence, but all are related to the fundamental metabolism of the cell life cycle, contributing much to the survival in niches. These genes might play a role in the virulence attenuation of 104M.

There are 3 genes directly associated with the classical bacterial virulence: the gene BAA13334_I03052 encodes a MFS transporter, which functions as part of the bacterial efflux system and has an apparent effect on the virulence and resistance to antimicrobial agents. The gene BAA13334_II01445 encodes a Tellurite resistance protein-related permease, and the last gene BAA13334_II00407 encodes the flagellar motor switch protein G, which plays a central role in switching.

Thus, some of the A13334-specific genes are directly or indirectly associated with virulence. The absence of these genes might be one of the reasons for the virulence attenuation in 104M.

For the 104M-specific genes, only 7 genes were annotated, which encode an acid-shock protein, an AzlC family protein, a diguanylate cyclase, a non-classified regulator and 3 mobile element proteins. None of the functions were associated with bacterial virulence.

As a result, the absence of A13334-specific genes in 104M provides a candidate set of genes that resulted in the attenuated virulence in 104M.

**Different virulence attenuation mechanisms between the vaccine strains 104M and S19**

S19 was the other *B. abortus* vaccine strain. To determine whether the genes identified above are also the potential reason for virulence attenuation in S19, we conducted a search of the 99 genes against the genome of S19. As a result, 23 genes were found to be homologous in S19, which were marked with asterisk in **Table 2**. These genes contained most of the virulence-associated genes described above, suggesting that S19 possessed the candidate genes that might result in the attenuated virulence in 104M. It also suggested that there existed different sets of genes leading to the attenuation of 104M and S19.

To confirm it, a pairwise and reciprocal comparison between A13334 and S19 was conducted. A total of 428 and 95 genes were identified to be unique in A13334 and S19, respectively. The number of strain-specific genes was greater in the comparison between A13334 and S19 than in the comparison between A13334 and 104M. Moreover, the genes specific in A13334 and missing in S19, listed in **Table S1**, included a set of genes associated with virulence, such as multidrug resistance transporters, *flgI* and *flhA*, most of which were present in the genome of 104M. These differences

**Table 2.** A list of genes missing in the attenuated strain 104M compared with the virulent strain, A13334

| Gene | Function | KO | Pathway |
|---|---|---|---|
| BAA13334_I00134 | ABC transporter | K06147 | ATP-binding cassette |
| BAA13334_II01500* | ABC transporter | K02032 | Peptides/nickel transport system |
| BAA13334_II00879 | ATP-dependent DNA ligase | | |
| BAA13334_I02291 | Diguanylate cyclase | | |
| BAA13334_I01435 | Fe-S protein | | |
| BAA13334_II00407* | Flagellar motor switch protein G<br>Flagellar assembly | K02410 | Bacterial chemotaxis |
| BAA13334_I01019 | Fumarase | | |
| BAA13334_I01734* | Gluconolactonase | | |
| BAA13334_I01361* | Glucose-methanol-choline oxidoreductase | K00108 | Glycine, serine and threonine metabolism |
| BAA13334_II00868* | Glyceraldehyde-3-phosphate dehydrogenase/<br>erythrose-4-phosphate dehydrogenase | | |
| BAA13334_II01395 | Glycosyl hydrolase family protein | | |
| BAA13334_I03093 | Integrase, catalytic domain-containing protein | | |
| BAA13334_II01601 | Leucyl-tRNA synthetase | | |
| BAA13334_I03416* | LuxR family 2 component transcriptional regulator | | |
| BAA13334_I03052 | Major facilitator superfamily transporter | K03449 | MFS transporter |
| BAA13334_I02644 | Membrane protein related to metalloendopeptidase | | |
| BAA13334_I00309 | Membrane-associated Zn-dependent proteases 1 | | |
| BAA13334_I01622 | Mg2+ and Co2+ transporter | | |
| BAA13334_I01425* | Myosin-14 | | |
| BAA13334_I01865 | N-acetylglucosamine-6-phosphate deacetylase | | |
| BAA13334_I01533* | N-acetyltransferase GCN5 | | |
| BAA13334_II01619* | Nitrite transporter | K02575 | Nitrogen metabolism |
| BAA13334_I01778* | NTP pyrophosphohydrolase | | |
| BAA13334_II00969* | Outer membrane receptor protein | | |
| BAA13334_I00149 | Phosphatidylglycerophosphate synthase | | |
| BAA13334_I00176* | Phosphoadenosine phosphosulfate reductase | K00390 | Sulfur metabolism |
| BAA13334_I02247 | Prophage antirepressor | | |
| BAA13334_I00833* | Radical SAM domain-containing protein | K10026 | Folate biosynthesis |
| BAA13334_II00036 | Retron-type reverse transcriptase | | |
| BAA13334_II01216* | Ribonuclease R | K12573 | RNA degradation |
| BAA13334_I02254* | Site-specific recombinase, phage integrase family protein | | |
| BAA13334_I01159* | Sulfate permease | | |
| BAA13334_I03390* | Superoxide dismutase | | |
| BAA13334_II01445 | Tellurite resistance protein-related permease | | |
| BAA13334_I02277 | UDP-glucose 4-epimerase | | |
| BAA13334_I01801* | Urease accessory protein UreF | K03188 | urease accessory protein |
| BAA13334_II00218* | WD40 repeat protein | | |
| BAA13334_I00025 | Hypothetical protein | | |
| BAA13334_I00054 | Hypothetical protein | | |
| BAA13334_I00130 | Hypothetical protein | | |
| BAA13334_I00328 | Hypothetical protein | | |
| BAA13334_I00441 | Hypothetical protein | | |
| BAA13334_I00661 | Hypothetical protein | | |
| BAA13334_I00777 | Hypothetical protein | | |
| BAA13334_I00781 | Hypothetical protein | | |
| BAA13334_I00883* | Hypothetical protein | | |
| BAA13334_I00994* | Hypothetical protein | | |
| BAA13334_I01009 | Hypothetical protein | | |
| BAA13334_I01155 | Hypothetical protein | | |
| BAA13334_I01185 | Hypothetical protein | | |
| BAA13334_I01224 | Hypothetical protein | | |
| BAA13334_I01234 | Hypothetical protein | | |
| BAA13334_I01281 | Hypothetical protein | | |
| BAA13334_I01332 | Hypothetical protein | | |
| BAA13334_I01544 | Hypothetical protein | | |
| BAA13334_I01558 | Hypothetical protein | | |
| BAA13334_I01585 | Hypothetical protein | | |
| BAA13334_I01834 | Hypothetical protein | | |

(*Continued on next page*)

**Table 2.** A list of genes missing in the attenuated strain 104M compared with the virulent strain, A13334 *(Continued)*

| Gene | Function | KO | Pathway |
|---|---|---|---|
| BAA13334_I01840 | Hypothetical protein | | |
| BAA13334_I01849 | Hypothetical protein | | |
| BAA13334_I01922 | Hypothetical protein | | |
| BAA13334_I01965 | Hypothetical protein | | |
| BAA13334_I01967* | Hypothetical protein | | |
| BAA13334_I01998 | Hypothetical protein | | |
| BAA13334_I02213* | Hypothetical protein | | |
| BAA13334_I02339 | Hypothetical protein | | |
| BAA13334_I02438 | Hypothetical protein | | |
| BAA13334_I02447 | Hypothetical protein | | |
| BAA13334_I02529 | Hypothetical protein | | |
| BAA13334_I02539 | Hypothetical protein | | |
| BAA13334_I02563 | Hypothetical protein | | |
| BAA13334_I02641 | Hypothetical protein | | |
| BAA13334_I02697 | Hypothetical protein | | |
| BAA13334_I02716 | Hypothetical protein | | |
| BAA13334_I02805 | Hypothetical protein | | |
| BAA13334_I02838 | Hypothetical protein | | |
| BAA13334_I03094 | Hypothetical protein | | |
| BAA13334_I03113 | Hypothetical protein | | |
| BAA13334_I03253 | Hypothetical protein | | |
| BAA13334_I03297 | Hypothetical protein | | |
| BAA13334_I03335 | Hypothetical protein | | |
| BAA13334_II00091 | Hypothetical protein | | |
| BAA13334_II00282 | Hypothetical protein | | |
| BAA13334_II00299 | Hypothetical protein | | |
| BAA13334_II00408 | Hypothetical protein | | |
| BAA13334_II00469 | Hypothetical protein | | |
| BAA13334_II00654 | Hypothetical protein | | |
| BAA13334_II00682 | Hypothetical protein | | |
| BAA13334_II00816 | Hypothetical protein | | |
| BAA13334_II01072 | Hypothetical protein | | |
| BAA13334_II01240 | Hypothetical protein | | |
| BAA13334_II01292 | Hypothetical protein | | |
| BAA13334_II01303 | Hypothetical protein | | |
| BAA13334_II01316 | Hypothetical protein | | |
| BAA13334_II01365 | Hypothetical protein | | |
| BAA13334_II01482 | Hypothetical protein | | |
| BAA13334_II01580 | Hypothetical protein | | |
| BAA13334_II01804 | Hypothetical protein | | |
| BAA13334_II01806 | Hypothetical protein | | |

*The genes that are found to be homologous in the vaccine strain S19.

further indicated that the 2 vaccine strains had different attenuation mechanisms. Thus, this paper would focus on the differences between 104M and A13334 to detect the potential genes associated with virulence attenuation in 104M.

### Identification of Gene change associated with virulence between 104M and A13334

The comprehensive comparison also revealed 465 genes that are non-identical between the 2 genomes. These differences are mainly due to site mutations, premature stops and indels. Similarly, a literature search was also conducted to associate these genes with the published virulence-related genes,[23,24] and the identified virulence-associated genes with the corresponding function information were listed in **Table 3**.

### LPS

Lipopolysaccharide is vital to the structural and functional integrity of the Gram-negative bacteria outer membrane. The LPS of *Brucella* constitutes a key virulence mechanism for intracellular survival and replication. In *Brucella*, the genes in the *wbk* region of chromosome I are involved in the synthesis of the O-PS and its translocation to the periplasm.[36] In the genomes of A13334 and M104, the core genes of this region (*wzm, wzt* and *wbkB*) are missing. In this O-antigen gene cluster, *gmd, per* and *wbkC* use GDP-mannose as substrate to synthesize the GDP-Nf-Per, and here, wbkC acts as an Nf-Per-transferase.[37] However, compared with A13334, several mutations were present in the core genes, *wbkA* [T1012G(S338A), A1047C(K349N)] and *per* [C733A(Q245K)]. These data could suggest that the mutation of these 3 sites might contribute to the reduced virulence of M104.

**Table 3.** A list of virulence-associated genes with mutations in 104M when compared with A13334

| Gene | Mutation | Function category |
|---|---|---|
| wbkA | S228A, K349N | LPS |
| per | Q245K | |
| *virB10* | A100V | Secretion system |
| BAA13334_I02537 | Insertion: 460KKKAAP466 | transporter system |
| BAA13334_I00045 | P619L | Flagellar |
| cgs | V2412A | Intracellular survival |
| BAA13334_I1773 | P76T | Amino acid metabolism |
| BAA13334_I02590 | C-terminus mutation | |
| eyrB | R194S | Sugar metabolism |
| galcD | Gene fusion | |
| rpsA | S552R | DNA/RNA metabolism |
| aidB | A470V | |
| rpoA | A219T | |
| purF | S425G | |
| cobB | K100Q | Vitamins/Cofactors |
| dxpS | S571G | |
| BAA13334_ I02969 | N117T | Stress |
| BAA13334_ I00578 | V203G | |
| dsbA | Q70K | Oxidoreduction |
| BAA13334_ I00370 | N19T | Unknown |
| BAA13334_ II01604 | R181Q | |

**Secretion system VirB or transport system**

The T4SS in *Brucella* is typified by the *virB* operon encoding 12 proteins. *Brucella* uses T4SS for the translocation of virulence factors into mammalian cells. In the comparison, we identified that Both M104 and A13334 lost the OMP gene *virB7*. In addition, there is a mutation occurred in the inner membrane protein *virB10*, C299T(A100V).

In addition to the mutations in T4SS, there were 2 gene mutants in the transporter system: BAA13334_I02537 encodes the trigger factor, and the homologous protein in M104 has a fragment insertion 460KKKAAP466.

One non-synonymous mutation was also identified in the flagellar gene BAA13334_I00045[C1856T(P619L)]. Due to the stop codon mutation, the orthologous genes of BAA13334_II00088, BAA13334_II00089 and BAA13334_ II00090 were merged to form a new gene in 104M. It has also been shown that *Brucella* mutants in *fliF, flhA* (BAA13334_II00089), *flgI, flgE, fliC* and *motB* are attenuated in BALB/mice.[38] These data may suggest that virulence attenuation may be related to the mutations in secretion and flagellar assembly systems.

**Intracellular survival**

Osmoregulated periplasmic glucans (OPGs) can be divided into 4 families (I-IV) on the basis of backbone organization and are constituents of the envelopes of gram-negative bacteria. Mutants deficient in OPG synthesis show altered chemotaxis and motility and reduced outer membrane stability and synthesis of exopolisaccharides.[39] Moreover, the mutants are unable to establish successful pathogenic or symbiotic association with their animal or plant hosts.[40] *Brucella* CβGs encoded by the *cgs* gene are considered family II OPGs and are neither O-substituted nor osmoregulated.[41] In the comparison, a mutation was identified between the gene *cgs* of A13334 and M104, BAA13334_I00290 [T7235C(V2412A)], which suggests that the mutation in *cgs* is among the reasons leading to the attenuated virulence.

**Amino acid metabolism**

Some studies[23,42,43] have reported that *Brucella* encounters an environment poor in nutrients before reaching the replicative compartment, and failure to resist this sudden lack of nutrients would result in attenuation. In the amino acid metabolism pathway, there were 2 genes identified with mutations. One is the gene BAA13334_I1773, which encodes an aminotransferase with a C226A(P76T) mutation. The other is the gene BAA13334_I02590, which encodes an extracellular solute-binding protein. Its homolog in 104M has a different 17 amino acid region in C-terminus, which might be caused by a site mutation and premature stop.

**Sugar metabolism**

Eryrthritol metabolism by *Brucella* has been identified as a trait associated with the capability of the pathogen to cause livestock abortions. The preferential growth of *Brucella* in the foetal tissues of cattle, sheep, goats and pigs was shown to be due to the high erythritol concentration. Although *Brucella* infects and causes brucellosis in other organisms such as humans, rats, rabbits and guinea pigs, overwhelming infection of the placental and foetal tissues is not observed, which is also associated with low concentrations of erythritol. In the M104 genome, the *eryB* gene has a mutation (R194S).

In addition to the erythritol metabolism, the *galcD*, which encodes D-galactarate dehydratase, merged with its upstream gene, which encodes a galactarate dehydratase, to form a new gene in 104M at the same location due to the stop codon mutation.

**DNA/RNA metabolism**

Purine/Pyrimidine synthesis, repair and regulatory genes are essential for *Brucella* intracellular replication. Mutants affecting RNA helicase or DNA gyrase activities suggest that *Brucella* might also use other categories as well to regulate virulence genes.[44] In this gene cluster, 4 genes were identified with non-synonymous mutations: gene *rpsA*, encoding the ribosomal protein S1, changed with A1564C(S522R); *aidB*, functioning as protection against to alkylation damage to DNA, changed with C1409T(A470V); the third gene *rpoA*, encoding the RNA polymerase α subunit, changed with G655A(A219T), and the last gene *purF*, encoding the purines synthesis, changes with S425G.

**Vitamins/Cofactors**

There were also mutations occurring in 2 genes involving vitamins or cofactor metabolism. *cobB* is a cobalamin-synthesis gene with the mutation K100Q; and *dxps*, taking part in thiamine synthesis, has a mutation of S571G.

**Stress**

Genes encoding stress proteins have been an obvious choice for directed mutagenesis for virulence studies and vaccine generation. Within the M104 genome, genes encoding protease (BAA13334_ I02969) and chaperone (BAA13334_ I00578) showed one non-synonymous mutation, namely, A350C (N117T) and T608G(V203G), respectively.

**Oxidoreduction**

DsbA has been shown to be involved in toxin production, adhesion, motility, extracellular enzyme production, and the type III secretion system.[45,46] An A208T(Q70K) mutation was detected in the *dsbA* gene (BAA13334_ I02476) of M104.

**Unknown function**

Two ORFs encoding hypothetical proteins were slightly different between the attenuated strain M104 and the virulent strain A13334. One is an orphan gene (BAA13334_ I00370) with the mutation of A56C(N19T), and the other is the unknown function gene (BAA13334_ II01604) with a mutation of G389A (R181Q).

## Conclusion

The *B. abortus* vaccine strain 104M was developed in China, and this vaccine was proven to be stable in antigenic structure, virulence, and immunogenicity. Since 1965, it has been widely used to vaccinate cattle and human against brucellosis in China.

In this study, we determined the whole-genome sequence of this vaccine strain 104M and found high genome similarities between 104M and A13334. A comprehensive comparative analysis against the whole-genome sequence of the virulent strain A13334 reveals that the genome sequences are highly linearly conserved with more than 97% identity. Meanwhile, a set of genome differences, including the difference of gene content and mutation of orthologous genes, was found between both strains, many of which were in genes associated with virulence and cell survival.

The unique genes identified for one genome mean the loss of the same genes in the other genome. Thus, the lack of A13334-specific genes in 104M provides a direct basis for the virulence attenuation. Especially in the 99 genes, some are directly or indirectly related to virulence and survival of the *Brucella* strain. In contrast, the 104M-specific genes have no clear association with virulence, but can act as a set of potential genes for identification of vaccine strain 104M. In addition tothe strain-specific genes, many mutations in the genes shared by both genomes were identified. Similarly, some of them are virulence genes or virulence-associated genes.

Collectively, this study presents a set of candidate genes that might be associated with virulence and will helpful for the elucidation of mechanisms of *Brucella* virulence attenuation.

## Methods and Materials

**Isolation extraction and identification**

*B. abortus* 104M was obtained from the LanZhou Institute of Biological Products in China. The total genomic DNA was extracted and purified by the modification of a previously described method.[21] An aliquot of the DNA was subjected for analysis using the BioAnalyzer (Agilent Technologies) and was confirmed for no degradation.

**Genome sequencing, assembly and annotation**

Paired-end sequencing was conducted through the high throughput Illumina sequencing technology. We first constructed 2 DNA libraries with 500 and 2000 bp fragments. Then, pair-end 90-bp reads were collected. A total of 170 Mb of clean reads were obtained, resulting in more than 47X coverage of this strain's whole genome. These quality filtered reads were then assembled into scaffolds using the SOAP de-novo software and a total of 14 scaffolds were formed.

The 14 scaffolds were aligned to the whole genome sequences of *B. abortus* A13334 to identify the putative gaps not sequenced in the whole genome of 104M. Primers were then designed and the obtained draft genome of 104M was used as a template in PCR to amplify the segments across the gaps. The purified PCR amplicons were used as templates in sequencing. The newly generated sequences, together with the scaffolds, were used to determine the complete genome sequence.

The final complete genome sequences of *B. abortus* vaccine strain 104M is available in GenBank with the accession number CP009625-CP009626, which correspond to the sequences of 2 chromosomes. The following genome sequence annotation was conducted using the RAST server, an automated genome annotation pipeline.

**Comparative genomic analysis**

Whole genome sequences and the corresponding gene sets of *B. abortus* A13334 and other *B. abortus* strains were downloaded from NCBI database. The multiple genome alignment was based on the software of Mauve to detect the SNPs and the genome structure similarity. Pair-wise and reciprocal comparisons were performed to identify potential genes that differ between the attenuated strain, M104 and virulent strain, A13334. The detailed variations in virulence-associated genes, including SNPs, deletions, insertions, and premature stops were further identified by Muscle alignment and then calculated using Python scripts.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Funding**

This work was supported by National Science and Technology Major Project of the Ministry of Science and Technology of China (2012ZX10003008002 to S. Li), National Science and

Technology Major Properties for Major New Drugs Innovation and Development (20132×09304101).

### Authors' Contributions

JY, SL and BW formulated this study. DY performed the research. YH, XZ and JX performed the sequencing of the whole genome of *Brucella abortus* vaccine strain 104M. DY, LL and JY analyzed the data. DY and JY wrote the paper. All authors read and approved the final manuscript.

### Supplemental Material

Supplemental data for this article can be accessed on the publisher's website.

### References


1. Verger J-M, Grimont F, Grimont PAD, Grayon M. *Brucella*, a monospecific genus as shown by deoxyribonucleic acid hybridization. Int J Syst Bacteriol 1985; 35:292-5; http://dx.doi.org/10.1099/00207713-35-3-292
2. Pappas G, Papadimitriou P, Akritidis N, Christou L, Tsianos EV. The new global map of human brucellosis. Lancet Infect Dis 2006; 6:91-9; PMID:16439329; http://dx.doi.org/10.1016/S1473-3099(06)70382-6
3. Corbel MJ. Brucellosis: an overview. Emerg Infect Dis 1997; 3:213-21; PMID:9204307; http://dx.doi.org/10.3201/eid0302.970219
4. Joint FAO/WHO expert committee on brucellosis. World Health Organ Tech Rep Ser 1986; 740:1-132; PMID:3097967
5. Atluri VL, Xavier MN, de Jong MF, den Hartigh AB, Tsolis RM. Interactions of the human pathogenic *Brucella* species with their hosts. Annu Rev Microbiol 2011; 65:523-41; PMID:21939378; http://dx.doi.org/10.1146/annurev-micro-090110-102905
6. BY C. Endemic surveillance and control of Brucellosis in China. Dis Surveill 2007; 22:649-51
7. EJ Y. *Brucella* species. Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. In: Mandell GL, Bennet JE, Dolin RE, editor. Philadelphia, USA: Elsevier Churchill Livingstone; 2005:2669-74.
8. Corner LA, Alton GG. Persistence of *Brucella abortus* strain 19 infection in adult cattle vaccinated with reduced doses. Res Vet Sci 1981; 31:342-4; PMID:6805054
9. Alton GG, Corner LA, Plackett P. Vaccination of pregnant cows with low doses of *Brucella abortus* strain 19 vaccine. Aust Vet J 1980; 56:369-72; PMID:6776944; http://dx.doi.org/10.1111/j.1751-0813.1980.tb09561.x
10. Schurig GG, Roop RM, Bagchi T, Boyle S, Buhrman D, Sriranganathan N. Biological properties of RB51; a stable rough strain of *Brucella abortus*. Vet Microbiol 1991; 28:171-88; PMID:1908158; http://dx.doi.org/10.1016/0378-1135(91)90091-S
11. Van Metre DC, Kennedy GA, Olsen SC, Hansen GR, Ewalt DR. Brucellosis induced by RB51 vaccine in a pregnant heifer. J Am Vet Med Assoc 1999; 215: 1491–3, 1449; PMID:10579049
12. Blasco JM. A review of the use of *B. melitensis* Rev 1 vaccine in adult sheep and goats. Prev Vet Med 1997; 31:275-83; PMID:9234451; http://dx.doi.org/10.1016/S0167-5877(96)01110-5
13. Jiang H, Du P, Zhang W, Wang H, Zhao H, Piao D, Tian G, Chen C, Cui B. Comparative genomic analysis of *Brucella* melitensis vaccine strain M5 provides insights into virulence attenuation. PLoS One 2013; 8: e70852; PMID:23967122; http://dx.doi.org/10.1371/journal.pone.0070852
14. Gao Y, Ren C, Ying T, Wang X. [Comparative proteomics research on THP-1 cells infected with *Brucella*]. Wei Sheng Wu Xue Bao 2006; 46:629-34; PMID:17037068
15. I.A. K, A.I. P, Sh.A. B. Some results from using the vaccine from strain *B. abortus* 104-M in calves under farm conditions. Res Pap Sib Vet Res Inst 1975; 22:18-22
16. L.N. K. Comparative characteristics of pathomorphological changes caused by *Brucella* vaccine strains No. 19, 19 BA, and M-104 with cutaneous and subcutaneous administration under experimental conditions. J Microbiol Epidemiol Immunol 1963; 3:39-45
17. K.V. S, A.V. A. Study of the vaccine strains *B. abortus* 104-M, *Brucella* melitensis Rev-1, and *B. abortus* 82 in cattle. Res Pap Kovalenko Natl Res Exp Vet Inst 1977; 45:29-36
18. Shumilov KV, Al' bertyan MP, Klochkov AA, Romakhov VA. Biological properties of the vaccine strain 104-M of *Brucella abortus*. Tr Vsesoyuznogo Instituta Eksp noi Vet 1983; 57:42-7
19. Crasta OR, Folkerts O, Fei Z, Mane SP, Evans C, Martino-Catt S, Bricker B, Yu G, Du L, Sobral BW. Genome sequence of *Brucella abortus* vaccine strain S19 compared to virulent strains yields candidate virulence genes. PLoS One 2008; 3:e2193; PMID:18478107; http://dx.doi.org/10.1371/journal.pone.0002193
20. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 2008; 9:75; PMID:18261238; http://dx.doi.org/10.1186/1471-2164-9-75
21. Wattam AR, Williams KP, Snyder EE, Almeida NF, Shukla M, Dickerman AW, Crasta OR, Kenyon R, Lu J, Shallom JM, et al. Analysis of ten *Brucella* genomes reveals evidence for horizontal gene transfer despite a preferred intracellular lifestyle. J Bacteriol 2009; 191:3569-79; PMID:19346311; http://dx.doi.org/10.1128/JB.01767-08
22. Kim H, Jeong W, Jeoung H-Y, Song J, Kim JJW, Beak J, Parisutham V, Lee SK, Jung SC, Her M, et al. Complete genome sequence of *Brucella abortus* A13334, a new strain isolated from the fetal gastric fluid of dairy cattle. J Bacteriol 2012; 194:5444; PMID:22965076; http://dx.doi.org/10.1128/JB.01124-12
23. Delrue R-M, Lestrate P, Tibor A, Letesson J-J, Bolle X. *Brucella* pathogenesis, genes identified from random large-scale screens. FEMS Microbiol Lett 2004; 231:1-12; PMID:14979322; http://dx.doi.org/10.1016/S0378-1097(03)00963-7
24. Seleem MN, Boyle SM, Sriranganathan N. *Brucella*: a pathogen without classic virulence genes. Vet Microbiol 2008; 129:1-14; PMID:18226477; http://dx.doi.org/10.1016/j.vetmic.2007.11.023
25. Jiang X, Leonard B, Benson R, Baldwin CL. Macrophage control of *Brucella abortus*: role of reactive oxygen intermediates and nitric oxide. Cell Immunol 1993; 151:309-19; PMID:8402938; http://dx.doi.org/10.1006/cimm.1993.1241
26. Gee JM, Valderas MW, Kovach ME, Grippe VK, Robertson GT, Ng W-L, Richardson JM, Winkler ME, Roop RM. The *Brucella abortus* Cu, Zn superoxide dismutase is required for optimal resistance to oxidative killing by murine macrophages and wild-type virulence in experimentally infected mice. Infect Immun 2005; 73:2873-80; PMID:15845493; http://dx.doi.org/10.1128/IAI.73.5.2873-2880.2005
27. Cruz-Ramos H, Glaser P, Wray L V, Fisher SH. The Bacillus subtilis ureABC operon. J Bacteriol 1997; 179:3371-3; PMID:9150240
28. Sangari FJ, Seoane A, Rodríguez MC, Agüero J, García Lobo JM. Characterization of the urease operon of *Brucella abortus* and assessment of its role in virulence of the bacterium. Infect Immun 2007; 75:774-80; PMID:17101645; http://dx.doi.org/10.1128/IAI.01244-06
29. Mulrooney SB, Hausinger RP. Nickel uptake and utilization by microorganisms. FEMS Microbiol Rev 2003; 27:239-61; PMID:12829270; http://dx.doi.org/10.1016/S0168-6445(03)00042-1
30. Soriano A, Colpas GJ, Hausinger RP. UreE stimulation of GTP-dependent urease activation in the UreD-UreF-UreG-urease apoprotein complex. Biochemistry 2000; 39:12435-40; PMID:11015224; http://dx.doi.org/10.1021/bi001296o
31. Soriano A, Hausinger RP. GTP-dependent activation of urease apoprotein in complex with the UreD, UreF, and UreG accessory proteins. Proc Natl Acad Sci U S A 1999; 96:11140-4; PMID:10500143; http://dx.doi.org/10.1073/pnas.96.20.11140
32. Kuchar J, Hausinger RP. Biosynthesis of metal sites. Chem Rev 2004; 104:509-25; PMID:14871133; http://dx.doi.org/10.1021/cr020613p
33. Guzman-Verri C, Manterola L, Sola-Landa A, Parra A, Cloeckaert A, Garin J, Gorvel J-P, Moriyon I, Moreno E, Lopez-Goni I. The two-component system BvrR/BvrS essential for *Brucella abortus* virulence regulates the expression of outer membrane proteins with counterparts in members of the Rhizobiaceae. Proc Natl Acad Sci U S A 2002; 99:12375-80; PMID:12218183; http://dx.doi.org/10.1073/pnas.192439399
34. Sola-Landa A, Pizarro-Cerdá J, Grilló MJ, Moreno E, Moriyón I, Blasco JM, Gorvel JP, López-Goñi I. A two-component regulatory system playing a critical role in plant pathogens and endosymbionts is present in *Brucella abortus* and controls cell invasion and virulence. Mol Microbiol 1998; 29:125-38; PMID:9701808; http://dx.doi.org/10.1046/j.1365-2958.1998.00913.x
35. Wang F, Qiao Z, Hu S, Liu W, Zheng H, Liu S, Zhao X, Bu Z. Comparison of genomes of *Brucella* melitensis M28 and the *B. melitensis* M5-90 derivative vaccine strain highlights the translation elongation factor Tu gene tuf2 as an attenuation-related gene. Infect Immun 2013; 81:2812-8; PMID:23716607; http://dx.doi.org/10.1128/IAI.00224-13
36. Zygmunt MS, Blasco JM, Letesson J-J, Cloeckaert A, Moriyón I. DNA polymorphism analysis of *Brucella* lipopolysaccharide genes reveals marked differences in O-polysaccharide biosynthetic genes between smooth and rough *Brucella* species and novel species-specific markers. BMC Microbiol 2009; 9:92; PMID:19439075; http://dx.doi.org/10.1186/1471-2180-9-92
37. Lübeck PS, Hoorfar J, Ahrens P. Chapter 40 cloning and characterization of the yersinia enterocolitica serotype O : 9 lipopolysaccharide O-antigen gene cluster. 1996:207-9.
38. Lestrate P, Dricot A, Delrue R, Lambert C, Martinelli V, De Bolle X, Letesson J, Tibor A. Attenuated signature-tagged mutagenesis mutants of *Brucella* melitensis identified during the acute phase of infection in mice. Infect Immun 2003; 71:7053-60; PMID:14638795; http://dx.doi.org/10.1128/IAI.71.12.7053-7060.2003
39. Bohin JP. Osmoregulated periplasmic glucans in Proteobacteria. FEMS Microbiol Lett 2000; 186:11-9; PMID:10779706; http://dx.doi.org/10.1111/j.1574-6968.2000.tb09075.x
40. Arellano-Reynoso B, Lapaque N, Salcedo S, Briones G, Ciocchini AE, Ugalde R, Moreno E, Moriyón I, Gorvel J-P. Cyclic beta-1,2-glucan is a *Brucella* virulence factor required for intracellular survival. Nat Immunol 2005; 6:618-25; PMID:15880113; http://dx.doi.org/10.1038/ni1202

41. Briones G, Iñón De Iannino N, Steinberg M, Ugalde RA. Periplasmic cyclic 1,2-β-glucan in *Brucella* spp. is not osmoregulated. Microbiology 1997; 143:1115-24; PMID:9141674; http://dx.doi.org/10.1099/00221287-143-4-1115
42. Delrue RM, Martinez-Lorenzo M, Lestrate P, Danese I, Bielarz V, Mertens P, De Bolle X, Tibor A, Gorvel JP, Letesson JJ. Identification of *Brucella* spp. genes involved in intracellular trafficking. Cell Microbiol 2001; 3:487-97; PMID:11437834; http://dx.doi.org/10.1046/j.1462-5822.2001.00131.x
43. Kohler S, Foulongne V, Ouahrani-Bettache S, Bourg G, Teyssier J, Ramuz M, Liautard J-P. The analysis of the intramacrophagic virulome of *Brucella* suis deciphers the environment encountered by the pathogen inside the macrophage host cell. Proc Natl Acad Sci U S A 2002; 99:15711-6; PMID:12438693; http://dx.doi.org/10.1073/pnas.232454299
44. Crawford RM, Van De Verg L, Yuan L, Hadfield TL, Warren RL, Drazek ES, Houng HH, Hammack C, Sasala K, Polsinelli T, et al. Deletion of purE attenuates *Brucella* melitensis infection in mice. Infect Immun 1996; 64:2188-92; PMID:8675325
45. Collet J, Bardwell JCA. MicroReview Oxidative protein folding in bacteria. 2002; 44:1-8; PMID:11967064
46. Jiang B-L, Liu J, Chen L-F, Ge Y-Y, Hang X-H, He Y-Q, Tang D-J, Lu G-T, Tang J-L. DsbB is required for the pathogenesis process of Xanthomonas campestris pv. campestris. Mol Plant Microbe Interact 2008; 21:1036-45; PMID:18616400; http://dx.doi.org/10.1094/MPMI-21-8-1036