Pakistan Veterinary Journal

ISSN: 0253-8318 (PRINT), 2074-7764 (ONLINE)
Accessible at: www.pvj.com.pk

RESEARCH ARTICLE

Whole-Genome Sequencing of *Brucella melitensis* from Xinjiang Provides Insights into its Genetic Features

Xiaowen Yang[1], Hao Yan[2], Ning Wang[1], Yexia Li[1], Pengfei Bie[1], Juan Zang[1], Yajie Li[1], Yanli Lu[1]* and Qingmin Wu[1]*

[1]Key Laboratory of Animal Epidemiology and Zoonosis of the Ministry of Agriculture, College of Veterinary Medicine, China Agricultural University, Beijing 100193, China; [2]Animal Health Supervision Station in XinJiang Uyghur Autonomous Region, Urumqi 830065, China
*Corresponding author: luyanli@cau.edu.cn; wuqm@cau.edu.cn

ARTICLE HISTORY (16-085)

Received: April 06, 2016
Revised: September 08, 2016;
Accepted: May 31, 2017
Published online: July 03, 2017

Key words:
COG analysis
Missense variant
SNP
Whole-genome sequence

ABSTRACT

Brucellosis is a widespread zoonotic infection, and in recent years human brucellosis emerged as one of the most significant public health concerns in Xinjiang, China. In these cases, *B. melitensis* is the major pathogen. In order to explore the genetic features of *B. melitensis* in Xinjiang, five suspected *B. melitensis* strains isolated in Northern Xinjiang were selected for whole-genome sequencing and analysis in this study. The results showed that all five strains belonged to *B. melitensis*. Phylogenetic analysis based on whole genomes showed that the five samples had significant differences from other *B. melitensis* strains from Europe and other Asian countries, but these strains were highly homologous with strains previously isolated in China. Among them, BY38 and BY72 had the closest phylogenetic relationship with *B. melitensis* M28, while AKS, ALT2 and ALT3 had the closest phylogenetic relationship with *B. melitensis* NI. Meanwhile, the SNPs and InDels that exist in the five strains make traceability possible, and the missense variant genes in COG analysis can help researchers understand the genetic and/or phenotypic differences among the strains, such as amino acid transport and metabolism (COG category E) indentified in this study. This study showed the genetic characteristics of *B. melitensis* strains from Xinjiang, and also provided basic data for researches on the prevention and control brucellosis in China.

To Cite This Article: Yang X, Yan H, Wang N, Li Y, Bie P, Zang J, Li Y, Lu Y and Wu Q, 2017. Whole-genome sequencing of *Brucella melitensis* from Xinjiang provides insights into its genetic features. Pak Vet J, 37(3): 251-256.

INTRODUCTION

Brucella spp., facultative intracellular pathogens that can persistently colonize animal host cells, cause a zoonosis, that is one of the most widespread diseases in both livestock and humans (Gul *et al.*, 2015; Keriel *et al.*, 2015). The National Health and Family Planning Commission of the People's Republic of China (http://www.nhfpc.gov.cn/) showed that the incidence of human brucellosis in China had increased each year since 2003. The number of newly diagnosed cases of human brucellosis in China was 57,222 cases in 2014, compared to 27,767 cases in 2008 and 6,448 in 2003. Although ten *Brucella* spp. are identified-*B. melitensis*, *B. abortus*, *B. suis*, *B. ovis*, *B. canis*, *B. ceti*, *B. pinnipedialis*, *B. neotomae*, *B. microti*, and *B. inopinata* most of the newly diagnosed human brucellosis cases in China were caused by *B. melitensis*, which had been divided into *B. melitensis* biovars 1-3 based on biochemical characteristics and host preferences. Traditionally, the characterization of *Brucella* into species and biovars depended on differential tests, such as phage typing, dye sensitivity, CO_2 requirement, H_2S production and so on. However, these methods limited the epidemiological investigation of *Brucella* traceability (De Santis *et al.*, 2013). In the last decade, analysis based on whole-genome sequencing had been used extensively to identify *Brucella* at species levels. This method had many advantages over traditional methods, including direct comparison with any reference genomes (Shallom *et al.*, 2012) and identification of the origin of epidemic strains through analysis of single nucleotide polymorphisms (SNPs) in their genomes. Since SNPs are the most common variation in the genomes, they represent an important bioinformatics marker in the relatively stable species. Therefore, SNPs detection has been widely used to identify genetic relationships among different strains (Pearson *et al.*, 2004). For instance, A study compared the

published *Bacillus anthracis* genomes from other countries with 122 sequenced genomes in France and were able to identify the origination of *B. anthracis* in France via phylogenetic trees based on genome SNPs (Girault *et al.*, 2014). Also, the potential origination of *B. melitensis* in the Mediterranean region and its global spread through infected goat and sheep or their milk products were determined by analyzing the genome SNPs of different strains (Tan *et al.*, 2015). Some researchers revealed that the MR Xichang lineage of *Schistosoma japonicum* was phylogenetically closer to LR Guich lineage than to other two LR lineages, and the MR lineage might be evolved from LR lineages by SNPs tree (Yin *et al.*, 2016).

The Xinjiang Uygur Autonomous Region is located in the northwestern border region of China and boasts a segment of the ancient Silk Road. The region shares a border with eight nations: Mongolia, Russian Federation, Kazakhstan, Kyrgyzstan, Tajikistan, Afghanistan, Pakistan and India. In the past ten years, Xinjiang had been one of the most seriously affected provinces based on the number of newly diagnosed cases of human brucellosis each year, with most cases caused by *B. melitensis.* In order to understand the genetic characteristics of the *B. melitensis* strain in Xinjiang and the relationship between *B. melitensis* from Xinjiang and from other countries, this study selected and isolated five strains isolated from different regions for analysis using whole-genome sequencing and comparative analysis. Exploring the potential origin of the five *B. melitensis* strains and their genetic characteristics will provide basic data for understanding the epidemic status of *B. melitensis* in China.

MATERIALS AND METHODS

Genome sequences: In this study, we chose and downloaded the whole genomes of *B. melitensis* 16M, *B. melitensis* M28, *B. melitensis* 63/9, *B. melitensis* Ether, *B. melitensis* NI, *B. melitensis* M5-90, *B. abortus* 2308, *B. abortus* S19, *B. suis* 1330 and *B. suis* S2 and the draft genomes of *B. melitensis* 66/59, *B. melitensis* F5/07-239A, *B. melitensis* UK24/06, *B. melitensis* UK19/04, *B. melitensis* UK14/06, *B. melitensis* REV.1, *B. melitensis* CNGB 1076, *B. melitensis* CNGB 1120, *B. melitensis* F8/01-155, *B. melitensis* F9/05, *B. melitensis* UK37/05, *B. melitensis* F10/06-16 and *B. melitensis* BG2 that had clear geographic information. These genomes were downloaded from NCBI GenBank FTP (ftp://ftp.ncbi.nlm.nih.gov/genomes/all).

Sequencing of *B. melitensis:* Strains of BY72 and BY38 were randomly isolated from sheep in Bayingolin Mongol Autonomous Prefecture and identified in 2011 and 2013; a strain of AKS was randomly isolated from sheep in Aksu Prefecture and identified in 2013 and strains of ALT2 and ALT3 were randomly isolated from Altay Prefecture and identified in 2010 and 2012. Cultured and inactivated samples were completed in the National Institute for Communicable Disease Control and Prevention of the Chinese Center for Disease Control and Prevention. Samples were sequenced using the Illumina/Solexa sequencing analyzer through at least 100-fold (100×) coverage to get the raw data. We used fastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to test the sequencing quality, and removed redundant and low-quality sequences until we obtained clean data.

Identification of the species: High-quality paired-end clean data were mapped for the reference genomes using BWA software (Li and Durbin, 2009). We used Samtools (Li *et al.*, 2009) to output the depth of sequencing and we used Perl (Stajich *et al.*, 2002) to output the 4× and 20× coverage of sequencing. By comparing the sequencing depth and coverage, we were able to identify species and potential biovars of *Brucella* (Shallom *et al.*, 2012).

SNPs and InDel calling: We got the best alignments in the SAM format, and then we filtered the results. Only high-quality SNPs were retained for subsequent analysis. We used Samtools and GATK (McKenna *et al.*, 2010) to call both SNPs and InDel at the same time. By analyzing the different algorithms to identify SNPs and InDel, we obtained highly credible final results.

Phylogenetic tree by SNPs: In this study, we selected one genome that had the highest coverage for all sequencing samples as the reference for SNP calling. We used Samtools and GATK to call SNPs and InDel for the strains, while we used MAUVE (Darling *et al.*, 2004) to call SNPs for the complete genomes. We used phyloSNP (Faison *et al.*, 2014) to output the final results, and the phylogenetic trees were constructed using Seaview (Gouy *et al.*, 2010). We used the neighbor-joining (NJ) method and the bootstrap was 1000.

SNP and InDel annotation: We used snpEff (Cingolani *et al.*, 2012) software to identify the annotation for the variants. We also used Excel to collect the variant information, and we used R program to draw variant sites between the strains.

COG analysis of missense variant genes: We used BLASTALL (Altschul *et al.*, 1990) to build COG database in our computer, and then we chose which missense genes to find based on their COG numbers and classification. The minimum E-value in BLASTALL was 1e-10, and we used "/" instead when the genes didn't match any COGs.

RESULTS

Sequencing and mapping of samples: Sequencing of the five strains generated 6.38Gb of paired-end DNA sequences, of which 6.36Gb of high-quality paired-end reads were mapped to the *B. melitensis* 16M complete genome. Meanwhile, for each strain, 98.71% (average) of reads mapped to 99.92% of the reference complete genome with 171- fold average depth.

Identification of the species and phylogenetic tree: *B. melitensis* 16M, *B. abortus* 2308 and *B. suis* 1330 were selected as reference genomes to identify the species by coverage and depth (Table 1). The results showed that the coverage and depth of *B. melitensis* 16M were higher than those of *B. abortus* 2308 and *B. suis* 1330, indicating that the five samples belonged to *B. melitensis*. Next, we

Table 1: The table of depth and coverage of five strains

Name	Reference genomes	Depth	Coverage (4×, %)	Coverage (20×, %)
AKS	*B. melitensis* 16M	157.30	99.91	99.87
	B. abortus 2308	156.71	99.72	99.60
	B. suis 1330	155.80	99.17	99.07
	B. melitensis M28	157.28	99.97	99.94
	B. melitensis 63/9	157.19	99.94	99.81
	B. melitensis ether	157.15	99.87	99.82
ALT2	*B. melitensis* 16M	176.71	99.92	99.88
	B. abortus 2308	176.08	99.71	99.62
	B. suis 1330	175.00	99.14	99.09
	B. melitensis M28	176.68	99.97	99.95
	B. melitensis 63/9	176.59	99.95	99.85
	B. melitensis ether	176.54	99.88	99.84
ALT3	*B. melitensis* 16M	149.94	99.91	99.87
	B. abortus 2308	149.40	99.72	99.61
	B. suis 1330	148.56	99.19	99.08
	B. melitensis M28	149.92	99.97	99.95
	B. melitensis 63/9	149.83	99.93	99.81
	B. melitensis ether	149.72	99.88	99.83
BY72	*B. melitensis* 16M	177.55	99.94	99.92
	B. abortus 2308	176.89	99.72	99.63
	B. suis 1330	175.78	99.18	99.12
	B. melitensis M28	177.52	100.00	99.99
	B. melitensis 63/9	177.42	99.97	99.89
	B. melitensis ether	177.38	99.89	99.87
BY38	*B. melitensis* 16M	193.67	99.94	99.92
	B. abortus 2308	192.95	99.71	99.63
	B. suis 1330	191.79	99.19	99.12
	B. melitensis M28	193.65	100.00	99.99
	B. melitensis 63/9	193.55	99.97	99.89
	B. melitensis ether	193.49	99.90	99.87

selected *B. melitensis* M28, *B. melitensis* 63/9 and *B. melitensis* Ether to represent *B. melitensis* biovars 1-3 to do the alignment. The results showed that the reference genome of *B. melitensis* M28 was the best alignment of these genomes, so we selected it as the reference to call SNPs and InDel to construct phylogenetic trees.

To explore relatedness among the sequenced strains and other strains of *B. melitensis* distributed worldwide, we conducted phylogenetic analysis using genomic SNPs. The phylogenetic tree showed significant differences in the strains in comparison with other strains of *B. melitensis* from Europe and other Asian countries, while they were highly homologous with strains isolated from China. Among these, strains of BY38 and BY72 shared the closest phylogenetic relationship with *B. melitensis* M28; while strains of AKS, ALT2 and ALT3 shared the closest phylogenetic relationship with *B. melitensis* NI (Fig. 1). Because *B. melitensis* M28 and *B. melitensis* M5-90 were highly similar, we selected the genome of *B. melitensis* M5-90 to do the alignment again for the samples of BY38 and BY72. The results showed that the depths were 193.62 and 177.50, respectively, and the coverage for both was 99.98%, which were less than *B. melitensis* M28.

SNPs and In Del calling: Based on the previous results of the phylogenetic tree and coverage, the genomes of *B. melitensis* M28 and *B. melitensis* NI were selected as the reference genomes. We performed SNPs and InDels calling and annotation, and counted SNPs per kb. The SNPs per kb were 0.068, 0.066, 0.065, 0.066 and 0.069 for the strains of BY38, BY72, ALT3, ALT2 and AKS. The results showed that the sequencing strains had low variance ratio and high similarity to *B. melitensis* M28 or *B. melitensis* NI.

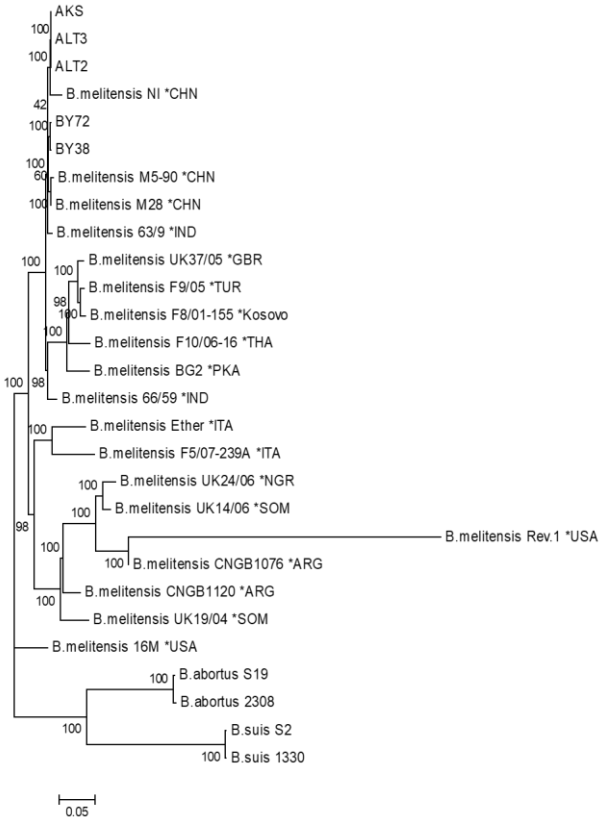

Fig. 1: Phylogenetic tree of five strains with the reference genome of *B. melitensis* M28 based on NJ method. * represents the abbreviation of the country where the strain was isolated, and CHN indicates the strain was isolated in China.

The genome of *B. melitensis* M28 was the reference genome of BY38 and BY72, whereas the genome of *B. melitensis* NI was the reference genome of ALT2, ALT3 and AKS. We counted the variant sites of BY38 and BY72, and the results showed that the two strains shared more than 88% of the variant sites. The specific sites of BY38 were 25 and BY72 were 20 (Fig. 2A). Meanwhile, we counted the variant sites of ALT2, ALT3 and AKS, and the results showed that the three strains shared more than 89% of the variant sites. The specific sites of ALT2 were 8, ALT3 were 8 and AKS were 18 (Fig. 2B). Based on the number of variant-specific sites, the strains from the same region were highly homologous, which provided some clues for the epidemiological investigation of *Brucella* traceability.

The variant sites were marked in chromosomes I and II. While the results showed that the variant sites of *B. melitensis* M28 were not relatively concentrated (Fig. 3A and B), the variant sites of *B. melitensis* NI were relatively concentrated in chromosome I at the sites of 542615-547638 and 647126-649472.

The annotation of variant sites showed that the sample of BY38 had 48 synonymous mutations, 95 missense mutations and 1.92 Ts/Tv (transitions/ transversions) ratio, while BY72 had 48 synonymous mutations, 96 missense mutations and 1.86 Ts/Tv ratio. ALT2 had 26 synonymous mutations, 80 missense mutations and 0.31 Ts/Tv ratio; ALT3 had 25 synonymous mutations, 78 missense mutations and 0.29 Ts/Tv ratio; and AKS had 26 synonymous mutations, 83 missense mutations and 0.30 Ts/Tv ratio.

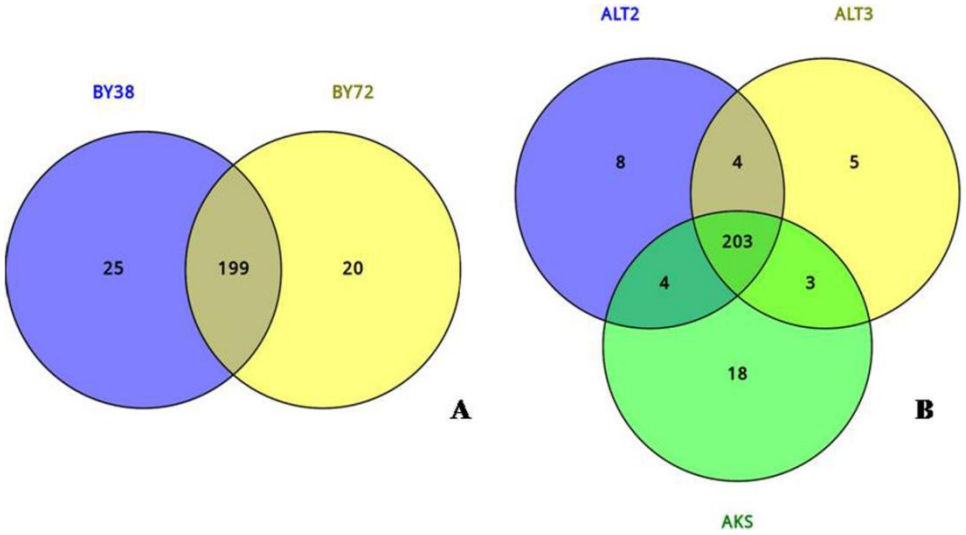

Fig. 2: Venn diagram for the variant sites on the genomes. A was the common variant sites about strains of BY38 and BY72, the reference genome was *B. melitensis* M28; B was the common variant sites among strains of ALT2, ALT3 and AKS, and the reference genome was *B. melitensis* NI.

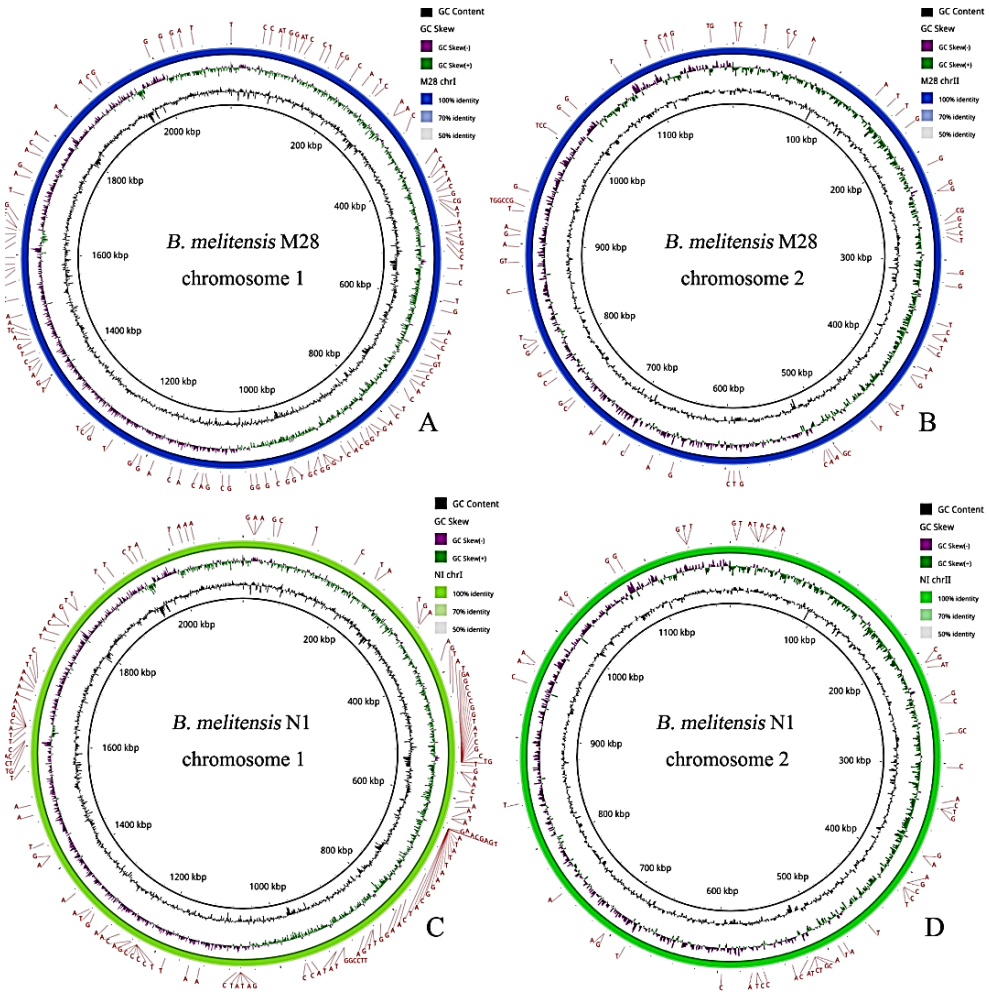

Fig. 3: Distribution of common variant sites. A indicates the common variant sites of BY38 and BY72 on the *B. melitensis*M28 chromosome I; B indicates the common variant sites on the *B. melitensis* M28 chromosome 2; C indicates the common variant sites of AKS, ALT2 and ALT3 on the *B. melitensis* NI chromosome I and D indicates the common variant sites on the *B. melitensis* NI chromosome 2.

Among the variant sites related to genes that were shared between BY38 and BY72, 8 genes were frameshift mutations, and 4 genes were frameshift mutations among the specific variant sites, with the frameshift mutation of 2 genes affecting the conserved domains among these 12 genes. Among the variant sites related to genes that were shared by ALT2, ALT3 and AKS, 3 genes were frameshift mutations, and 7 genes were frameshift mutations among the specific variant sites, with the frameshift mutations of 4 genes affecting the conserved domains among these 10 genes (Table 2).

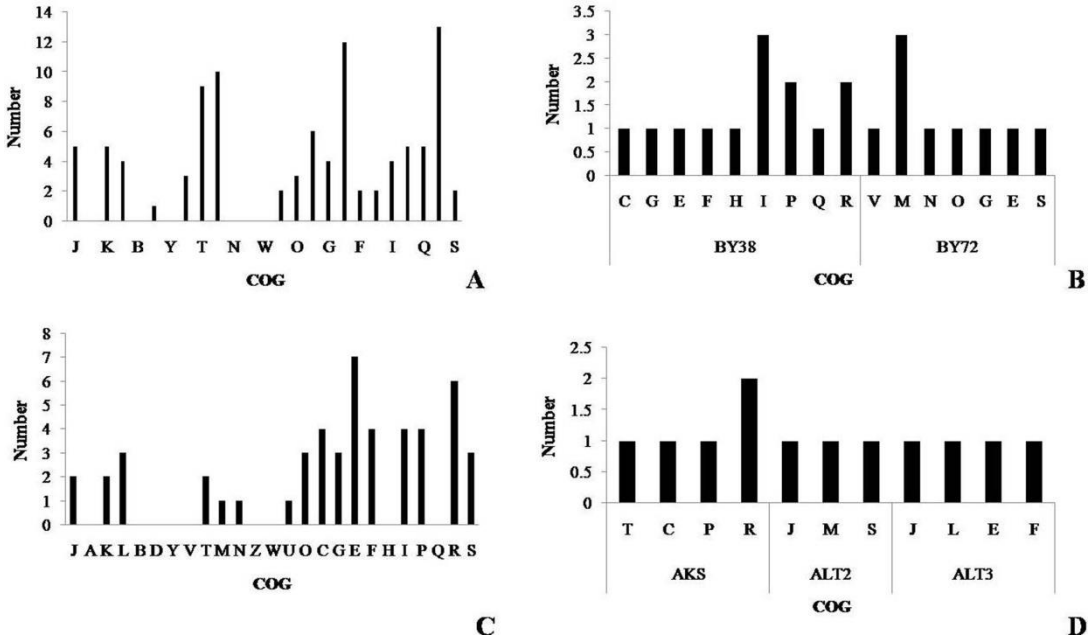

Fig. 4: The COG analysis of common and special missense genes for the sequencing strains. A was the COG of the common variant genes between the strains of BY38 and BY72; B was the COG of the special variant genes between the strains of BY38 and BY72; C was the COG of the common variant genes among the strains of ALT2, ALT3 and AKS; D was the COG of the special variant genes among the strains of ALT2, ALT3 and AKS.

Table 2: The names of genes that were affected by frameshift mutation

Name	Functions
BM28_RS05165	hypothetical protein
BM28_RS09990	oxidoreductase
BMNI_RS02545	transposase
BMNI_RS04140	uracil-DNA glycosylase
BMNI_RS14840	amidase
BMNI_RS08660	hypothetical protein

COG analysis of missense variant genes: A functional classification of the missense variant genes was conducted using COG analysis. The functions of the missense variant genes that were shared between BY38 and BY72 involved amino acid transport and metabolism (COG category E) (Fig. 4A). The specific missense variant genes of BY38 involved lipid transport and metabolism (COG category I), while the specific missense variant genes of BY72 involved cell wall/membrane/envelope biogenesis (COG category M) (Fig. 4B). The functions of missense variant genes which were shared among ALT2, ALT3 and AKS involved amino acid transport and metabolism (COG category E) (Fig. 4C), while the specific missense variant genes had no concentrated function (Fig. 4D).

DISCUSSION

In this study, whole-genome sequencing analysis was used to explore the origin, genetic characteristics, and epidemiologic situation of *B. melitensis* species in Xinjiang. Based on the phylogenetic tree, *B. melitensis*, *B. abortus* and *B. suis* could be well divided. Among the strains of *B. melitensis*, the strains from Asia and Europe were in the same clades, and the samples in this study were highly homologous with the strains from China (Fig. 1). From the phylogenetic tree, we also found that *B. melitensis* Ether (which isolates from Italy) and *B. melitensis* NI are not in the same clades. *B. melitensis* NI and *B. melitensis* M28 (which belongs to biovar 1) are in the same clades. These results were similar to other research (Tan *et al.*, 2015). *B. melitensis* NI, *B. melitensis* M28 and *B. melitensis* M5-90 were not in the same clades with the strains from other countries, suggesting that *B. melitensis* may be quite different in China versus in other countries.

In recent years, studies have shown that *B. melitensis* in Xinjiang belongs to biovar 3 (Wang *et al.*, 2007, Yi *et al.*, 2015). Some researchers isolated and identified, using traditional methods, 11 strains of *B. melitensis* biovar 3 from Bayingolin Mongol Autonomous Prefecture, Hotan Prefecture, Altay Prefecture, Aksu Prefecture and Ili Kazak Autonomous Prefecture (Yi *et al.*, 2015). However, in our study, the strains form Bayingolin Mongol Autonomous Prefecture shared the closest phylogenetic relationship with *B. melitensis* M28, and the strains from Altay Prefecture and Aksu Prefecture shared the closest phylogenetic relationship with *B. melitensis* NI. On a map of China, we found that *B. melitensis* biovar 3 was mainly distributed near the border of China, while the area that was closest to other Chinese provinces had *B. melitensis* biovars 1 and 3. These finding should be tested again using more isolated *B. melitensis* strains. Studies have shown that the Mediterranean and the Middle Eastern strains were mainly *B. melitensis* biovar 3 (Benkirane, 2006), but our study showed that the strains of *B. melitensis* biovar 3 from Xinjiang shared the closest phylogenetic relationship with *B. melitensis* NI which was isolated from Inner Mongolia in China.

The number of SNPs differed between strains, which could provide the basis for dividing different strains; for instance, *B. anthracis* in France could be divided by 1581 SNPs (Girault *et al.*, 2014). Researchers compared *Staphylococcus aureus* NC_017340.1 and *S. aureus* spa-type t003, and found that the strains had 534 variant sites and 478 SNPs (Hamed *et al.*, 2015). It was found that compared with *B. melitensis* 16M, *B. suis* 1330 and *B. abortus* 9-941, strains *B. melitensis* M5-10, *B. suis* S2-30 and *B. abortus* 104M had 2859, 7456 and 3368 SNPs,

respectively (Wang *et al.*, 2012). The number of SNPs in this study was similar to the numbers in that research; the strains in this study had 2556-2584 SNPs compared with *B. melitensis* 16M (data not shown).

Compared with *B. melitensis* NI, we found the samples had concentrated variant sites in chromosome I. The variant region between the sites of 542615 and 547638 belonged to the BMNI_RS02545-encoding transposase. In this gene, the variant sites included 6 synonymous mutations, 9 missense mutations and 1 frameshift mutation that affected the conserved domain; the variant sites at the sites of 674126-649472 were between BMNI_RS03045 and BMNI_RS03065. The genes encoded the transposase and hypothetical protein. We used this sequence to do the alignment by BLAST and found the identity of the sequence was more than 98% in *B. melitensis.*

The frameshift mutation that affected the gene's conserved domain had homologous in *B. melitensis* 16M, which were BME_RS04520, BME_RS09915, BME_RS13690, BME_RS05500, BME_RS11610 and BME_RS01010. These genes were not the known virulence genes (He, 2012). These results were, however, consistent with the actual station, as the strains had not reduced the virulence (data not shown). However, BME_RS13690 (old locus was BMEII0710) belonged to GI06 (genomic island, GI) (Rajashekara *et al.*, 2004), and the genome of *B. ovis* NCTC 63/290 was missing this gene (Zhong *et al.*, 2012). *Brucella* spp. may accept this gene by HGT (Horizontal gene transfer). The function of other genes may need further research.

Conclusions: Phylogenetic analysis showed that five *B. melitensis* strains from Xinjiang were highly homologous with the strains isolated in China. Since Xinjiang is located in a border region of China, these findings suggest that the epidemic strains of *B. melitensis* in China are relatively specific and conserved in their genomic sequences. Additionally, these identified strains had low variant ratio compared with the reference genomes. Meanwhile, COG analysis indicated that the missense variant genes were devoted to metabolism, which were mainly amino acid transport and metabolism (COG category E), suggesting that the energy and amino acid metabolism may play a more important role in the process of *Brucella*. This study revealed the genetic features of *B. melitensis* in Xinjiang, which will provide the basic data for further tracking the epidemic status of *B. melitensis* in China.

Acknowledgements: This work was funded by the National Special Foundation for Science & Technology Basic Research (No. 2012FY111000) and National Natural Science Foundation of China (No. 31372446).

Authors contribution: QMW, LYL and HY conceived and designed the experiments. XWY, NW, YXL, JZ and PFB executed the experiment. XWY analyzed the data and prepare this manuscript. All authors approved the final manuscript and agreed to publication.

REFERENCES

Altschul SF, Gish W, Miller W, *et al.*, 1990. Basic local alignment search tool. J Mol Biol 215:403-10.

Benkirane A, 2006. Ovine and caprine brucellosis: World distribution and control/eradication strategies in West Asia/North Africa region. Small Rum Res 62:19-25.

Cingolani P, Platts A, Wang le L, *et al.*, 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6:80-92.

Darling AC, Mau B, Blattner FR, *et al.*, 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394-403.

De Santis R, Ancora M, De Massis F, *et al.*, 2013. Molecular strain typing of Brucella abortus isolates from Italy by two VNTR allele sizing technologies. Mol Biotechnol 55:101-10.

Faison WJ, Rostovtsev A, Castro-Nallar E, *et al.*, 2014. Whole genome single-nucleotide variation profile-based phylogenetic tree building methods for analysis of viral, bacterial and human genomes. Genomics 104:1-7.

Girault G, Blouin Y, Vergnaud G, *et al.*, 2014. High-throughput sequencing of Bacillus anthracis in France: investigating genome diversity and population structure using whole-genome SNP discovery. BMC Genomics 15:288-97.

Gouy M, Guindon S and Gascuel O, 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:221-4.

Gul ST, Khan A, Maqbool A, *et al*., 2015. Epidemiology of brucellosis at different livestock farms in the Punjab, Pakistan. Pak Vet J 35:309-14.

Hamed M, Nitsche-Schmitz DP, Ruffing U, *et al.*, 2015. Whole genome sequence typing and microarray profiling of nasal and blood stream methicillin-resistant Staphylococcus aureus isolates: Clues to phylogeny and invasiveness. Infect Genet Evol 36:475-82.

He Y, 2012. Analyses of Brucella pathogenesis, host immunity, and vaccine targets using systems biology and bioinformatics. Front Cell Infect Microbiol 2:1-17.

Keriel A, Botella E, Estrach S, *et al.*, 2015. Brucella intracellular life relies on the transmembrane protein CD98 heavy chain. J Infect Dis 211:1769-78.

Li H and Durbin R, 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-60.

Li H, Handsaker B, Wysoker A, *et al.*, 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-9.

McKenna A, Hanna M, Banks E, *et al.*, 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297-303.

Pearson T, Busch JD, Ravel J, *et al.*, 2004. Phylogenetic discovery bias in Bacillus anthracis using single-nucleotide polymorphisms from whole-genome sequencing. Proc Natl Acad Sci USA 101:13536-41.

Rajashekara G, Glasner JD, Glover DA, *et al*., 2004. Comparative whole-genome hybridization reveals genomic islands in Brucella species. J Bacteriol 186:5040-51.

Shallom SJ, Tae H, Sarmento L, *et al.*, 2012. Comparison of genome diversity of Brucella spp. field isolates using universal bio-signature detection array and whole genome sequencing reveals limitations of current diagnostic methods. Gene 509:142-8.

Stajich JE, Block D, Boulez K, *et al.*, 2002. The Bioperl toolkit: Perl modules for the life sciences. Genome Res 12:1611-8.

Tan KK, Tan YC, Chang LY, *et al.*, 2015. Full genome SNP-based phylogenetic analysis reveals the origin and global spread of Brucella melitensis. BMC Genomics 16:93-103.

Wang Y, Ke Y, Wang Z, *et al.*, 2012. Genome sequences of three live attenuated vaccine strains of Brucella species and implications for pathogenesis and differential diagnosis. J Bacteriol 194:6012-3.

Wang YZ, Chen CF, Cui BY, *et al.*, 2007. Isolation and identification of *Brucela melitensis* biovar 3. Chin J Prev Vet Med 29:753-56 (in Chinese).

Yi XP, Ye F, Li JP, *et al.*, 2015. Species and biotype analysis of isolated strains of brucella in Xinjiang. Xingjiang Agri Sci 52:339-43 (in Chinese).

Yin M, Liu X, Xu B, *et al.*, 2016. Genetic variation between *Schistosoma japonicum* lineages from lake and mountainous regions in China revealed by resequencing whole genomes. Acta Trop 161:79-85.

Zhong Z, Wang Y, Xu J, *et al.*, 2012. Parallel gene loss and acquisition among strains of different Brucella species and biovars. J Microbiol 50:567-74.