

Attenuation mechanism of *Brucella melitensis* M5-10, implications for vaccine development and differential diagnosis

Yuanqiang ZHENG^{1,4*}, Yanchun SHI^{2*}, Chang AN^{4*}, Ruisheng LI^{3*}, Dongjun LIU¹, Yuehua KE⁴, Kairong MAO⁵, Mingjuan YANG⁴, Zeliang CHEN (✉)^{2,4,6}, Shorgan BOU (✉)¹

1 National Research Center for Animal Transgenic Biotechnology, Inner Mongolia University, Hohhot 010021, China

2 Inner Mongolia Key Laboratory of Molecular Biology, Inner Mongolia Medical University, Hohhot 010058, China

3 Animal Laboratory Center, 302 Hospital of PLA, Beijing 100039, China

4 Institute of Disease Control and Prevention, Academy of Military Medical Sciences, Beijing 100071, China

5 China Institute of Veterinary Drug Control, Center for Veterinary Drug Evaluation, Ministry of Agriculture, Beijing 100081, China

6 College of Medicine, Shihezi University, Shihezi 832003, China

Abstract Brucellosis is a worldwide zoonosis. Vaccination is the most efficient means to prevent and control brucellosis. The current licensed attenuated vaccines for animal use were developed by sequential passage in non-natural hosts that decreased virulence in its original hosts. The attenuation mechanism of these strains remains largely unknown. In the present study, we sequenced the genome of *Brucella melitensis* vaccine strain M5-10. Sequence analysis showed that a large number of genetic changes occurred in the vaccine strains. A total of 2854 genetic polymorphic sites, including 2548 SNP, 241 INDEL and 65 MNV were identified. Of the 2074 SNPs in coding regions, 1310 (63.2%) were non-synonymous SNPs. Gene number, percent and N/S ratios were disproportionally distributed among the cog categories. Genetic polymorphic sites were identified in genes of the *virB* operon, flagella synthesis, and virulence regulating systems. These data indicate that changes in some cog categories and virulence genes might result in the attenuation. These attenuation mechanisms also have implications for screening and development of new vaccine strains. The genetic changes in the genome represent candidate sites for differential diagnosis between these vaccine strains and other virulence ones. Transcription analysis of virulence genes showed that expression of *dnaK*, *vjbR* were reduced in M5-10 strain when compared with that in 16M. A duplex PCR targeting *virB6* and *dnaK* was successfully used to differentiate between M5-10 and the virulent 16M strain. The genome re-sequencing technique represents a strong

strategy not only for evaluation of vaccines, but also for development of new vaccines.

Keywords *Brucella*, live attenuated vaccine, attenuation, differential diagnosis

1 Introduction

Brucellosis is one of the most common zoonoses worldwide. About half a million new cases of human brucellosis are reported each year. This remains only part of the total cases, because human brucellosis is highly under-diagnosed or under-reported [1]. As a zoonosis, brucellosis is an important cause of veterinary morbidity and mortality that cause great economic losses. Human brucellosis is mainly derived from animal brucellosis, and human-to-human transmission of brucellosis is rarely reported [2]. Therefore, control of brucellosis is considered to be dependent on control and eradication of animal brucellosis.

Brucellosis is preventable by vaccination. Vaccination with live attenuated vaccine has proven to be the most efficient means for prevention and control of animal brucellosis and has significantly decreased the incidence rate of animal brucellosis in many countries. With the employment of eradication programs for tens of years, incidence rates were reduced to such a very low level that some countries declared eradication of brucellosis. For these countries, human brucellosis cases are mainly imported from other countries through international travel or animal transactions [3]. As a consequence of brucellosis occurring in wildlife and being transferred from wildlife to economic animals, a surveillance, inspection and slaughter

Received November 9, 2014; accepted December 20, 2014

Correspondences: zeliangchen.amms@gmail.com, xurg@cae.cn

*These authors contributed equally to the work

program is important for keeping the *Brucella*-free status [4]. Actually, the complete elimination of animal Brucellosis is nearly impossible.

Although live attenuated vaccine is essential in decreasing incidence and eradication programs, the characteristics of the vaccine strains have restricted their extensive application [5]. There are several limitations for the present live attenuated vaccine. Vaccine strains still have some virulence that might cause disease after immunization. Another limitation is that the immune response induced by the vaccine strains is difficult to differentiate from that of natural infection, which makes it nearly impossible to differentiate immunization and infection through standard serological methods. Overcoming these limitations will improve and extend the applications of these vaccine strains.

Several vaccine strains have been licensed worldwide for prevention and control of animal brucellosis. The licensed vaccine strains include Rev. 1, RB51 and S19 have been applied in many countries [6]. In addition to these strains, some other vaccine strains, including *Brucella melitensis* M5 and *B. suis* S2, were licensed and used for prevention of animal brucellosis in China [7]. Vaccine strain *B. melitensis* M5 was derived a virulent strain M28 isolated in China. These vaccine strains were screened through sequential passage in non-natural hosts. Passage in non-natural hosts usually results in attenuation of the strain in its original hosts. Attenuation raises two problems. On the one hand, the virulence is reduced, but the attenuation mechanism remains unclear. On the other hand, these strains have similar antigenicity to their parent virulent strains, and it is difficult to develop a serological method to differentiate immunization and infection.

Live attenuated vaccine strain M5 has been extensively applied in China. Analysis of the attenuation mechanism will contribute to further evaluation, but also development of new vaccine. In the present study, we sequenced the genome of M5-10, a derivative strain of M5 stored under laboratory condition. Genome characteristics and genetic changes were analyzed to uncover the attenuation mechanism. This will provide information about *Brucella* attenuation, and for future development of live attenuated vaccines. Based on the genetic differences, expression of virulence genes and differential diagnosis methods were analyzed.

2 Materials and methods

2.1 Bacterial culture and genomic DNA isolation

B. melitensis M5-10 was obtained from China Institute of Veterinary Drug Control and stored in our laboratory. The strain was cultured in tryptic soy broth (TSB) at 37°C. Genomic DNA was extracted using genomic DNA isolation kit (Promega).

2.2 Genome sequencing and contig assembly

Extracted DNA was subjected to genome sequencing by High-Seq 2000 at Shanghai Genomics Center (Shanghai). The generated raw sequences were trimmed by removing reads with low quality and then the clear data were used for assembly and polymorphism analysis. Contigs were assembled from the clear data with *de novo* assembly module CLCbio of Genomics Workbench (version 5.5). Open reading frames in contigs were predicted and annotated by using online RAST (<http://rast.nmpdr.org>).

2.3 Sequence polymorphism analysis

The clear reads data were mapped to genome sequences of *B. melitensis* 16M. Polymorphism sites, including SNP (single nucleotide polymorphism), INDEL (insertion and deletion), and MNV (multiple nucleotide variation), were identified based on the mapping results. The polymorphism data were further filtered to remove sites with low quality. Synonymous and non-synonymous SNPs were analyzed for each of the SNPs located in coding regions.

2.4 Genetic changes among cog categories

Distribution of genetic polymorphism sites in different cog categories were analyzed as follows. SNPs (both synonymous and non-synonymous) were calculated for each gene. All the genes were grouped according to their cog annotation. For each of the cog categories, the total number of SNP sites and genes, and SNP sites per gene were calculated. N/S ratio was calculated for each of the COG categories.

2.5 Genetic changes in virulence genes

Genes with genetic changes were investigated to evaluate their roles in *Brucella* virulence. Virulence genes are those directly involved in *Brucella* virulence. Virulence related genes are those regulating *Brucella* adaptation to environments.

2.6 Expression analysis of virulence genes

Total RNA was extracted from liquid cultures of *B. melitensis* using Trizol reagent (Invitrogen) as recommended by the manufacturer. RNA samples were then treated with DNase I (Promega) to remove any contaminating genomic DNA. RNA quantity and quality were assessed using ND-1000 Spectrophotometer Nanodrop (Technologies) and agarose gel electrophoresis. Samples were run in triplicate and amplified in a 25 µL volume containing 12.5 µL of 2 × SYBR Green I Master Mix (Takara Biochemicals), 100 nmol·L⁻¹ each primer, and 1 µL of cDNA sample. Thermocycling conditions were as follows: 10 min at 95°C for pre-incubation, and then 45

cycles of amplification (95°C for 30 s, 60°C for 30 s, and 72°C for 30 s). The primers used for qRT-PCR are listed in Appendix A (Table S1). All primer sets showed standard curves with $R^2 > 0.98$, 90% to 110% reaction efficiencies, and only one peak in dissociation curves. Relative transcriptional level was determined by the methods of $2^{-\Delta\Delta C_T}$ as described previously. The level of 16S rRNA was used as a reference gene to normalize the expression data for target genes. The average expression levels and standard deviations were calculated using data from three technical replicates of three independent experiments.

Different cDNA samples were amplified with primers for 16S rRNA, and the cDNA samples were normalized by differential dilutions according to the quantity of 16S rRNA products. Then, selected genes were amplified from normalized cDNA samples with specific primers (Appendix A, Table S1). The PCR products were analyzed on 1.2% agarose gels and visualized by ethidium bromide staining.

2.7 Multiplex PCR for differential detection of M5-10

A pair of primers targeting the genetic changes of *virB6* were designed and synthesized: forward primer VirB6-dif-F (5'-GCATGAATGCCAGACCGTCA-3') and reverse primer VirB6-dif-R (5'-TTGATGACGCGCTTGACGCG-3'), which generates a 360 bp product. The base at the 3' end of the primer is located at the genetic polymorphism sites. Other pairs of primers, DnaK-dif-F (5'-CATGATCCTTCAGAAGATG-3') and DnaK-dif-R (5'-CAGGCTTTCAAACCTTGG-3'), were synthesized that

generate a 550 bp product. A duplex PCR assay was developed by using the two pairs of primers. When the template is M5-10, only a 550 bp product will be generated while other virulent strain will generate both the 550 and 360 bp products.

3 Results

3.1 Genome sequencing of *B. melitensis* M5-10

DNA samples were prepared from M5-10 and subjected to sequencing with HiSeq 2000. A total of 1.5 Gb raw data were generated, consisting of 15.6 Mb reads, of which 99.5% were successfully mapped to the *B. melitensis* 16M genome. The average length of the reads was 64.8 (Appendix A, Table S1). The reads were assembled into contigs with *de novo* assembly method and a total of 47 contigs were generated, ranging in length from 0.282 to 296 kb. Average coverage for these contigs ranged from 261 to 3058 (Fig. 1). A total of 3214 ORFs were predicted from the contigs.

3.2 Identification of polymorphism sites

Trimmed sequence reads were mapped to the genome sequence of *B. melitensis* 16M, and polymorphism sites were then identified by the probability method in CLCbio Genomic Workbench software. After removal of low quality mappings, SNP, INDEL and MNV, were defined. A total of 2854 polymorphism sites, including 2548 SNPs

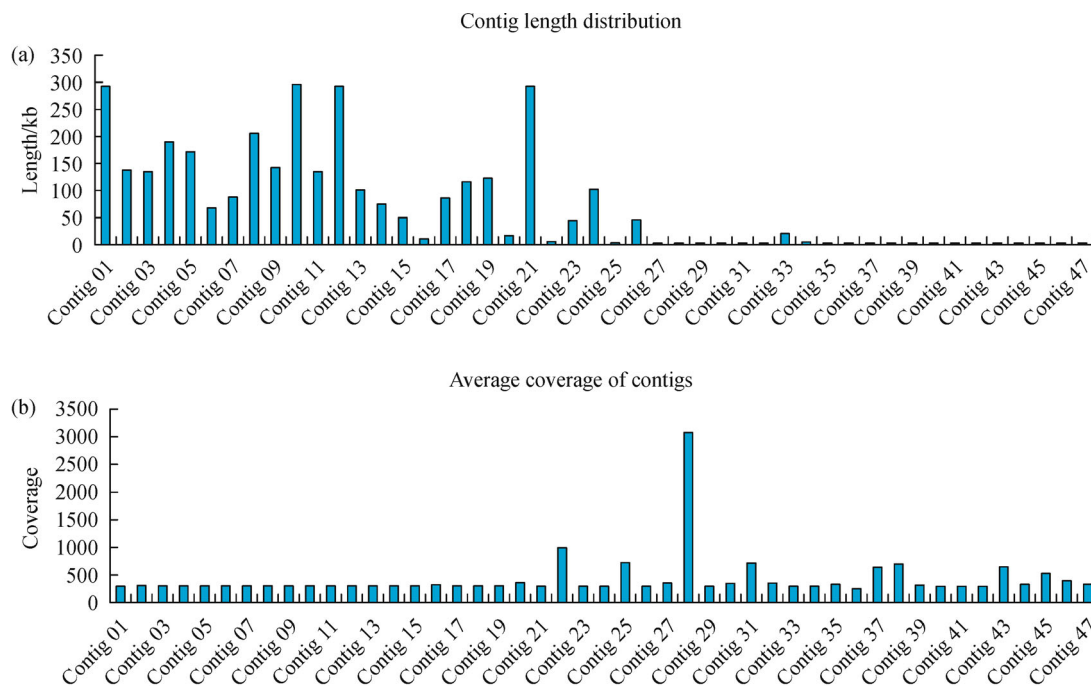


Fig. 1 Length and coverage distribution of the assembled contigs. (a) Length distribution of the assembled contigs; (b) average reads coverage of the contigs.

sites, 241 INDELs and 65 MNV, were identified. 2074 of the SNPs were located in CDS and 474 located in intergenic regions. SNP frequency intensity per gene was calculated as SNP number per kilobase. SNP intensity varied from 0.02 to 14 per kilobase in length, representing a variation of 200 fold. Of the MNVs, 50 were located in the coding regions, and 15 in the intergenic regions. For the INDELs, 151 were located in the coding regions, and 90 in the intergenic regions (Table 1). For all the variant types, the site numbers in the coding regions were higher than those in the intergenic regions (Table 1).

Insertion and deletion analysis identified 241 INDEL sites, of which 145 were small deletions, and 96 were insertions. For the small deletions, 126 were 1 bp deletion, 12 were 2 bp deletion. The 3 bp deletion occurred in 5 sites, and 4 bp deletion in only 2 sites. Base insertions ranged from 1 to 3 bp, with 91, 4 and 1 sites for 1, 2 and 3 bp deletion, respectively (Table 2). Both deletion and insertions occurred at higher frequency in chromosome II than in chromosome I. These small insertions and deletions were mainly located in coding regions. Multi nucleic acid variants (MNV) are sites where sequential bases are mutated. A total of 65 MNV sites were identified, 50 of them were located in coding regions. Two sites were changed from 1 base to 2 bases. MNV of 2 bases occurred in 55 sites, and 3 bases occurred in 8 sites (data not shown).

3.3 Nucleic acid replacement frequency of SNPs

The single nucleotide polymorphism (SNP) is one of the

commonest and most important polymorphism types in bacterial genomes. We analyzed the base replacement frequency among the SNPs. As shown in Table 2, of the 2074 SNPs, the original A (513, 24.73%), T (497, 23.96%), G (506, 24.39%) and C (558, 26.90%) were similarly distributed ($P > 0.05$). The base changes were divided into transition (A ↔ G or C ↔ T change) and transversion (A, G ↔ C, T). The base A was changed to T, C and G with frequency of 7.80%, 17.9% and 74.3% and base C was changed to G, A and T with frequency of 17.7%, 17.7% and 64.5%, respectively (Table 3). These results showed that transition occurred at a significantly higher frequency than transversion.

3.4 Differential distribution of the SNPs among COG categories

To analyze further whether the genetic changes caused functionally significant alterations, functional categories of the genes with polymorphism sites were defined and we analyzed the distribution of CDS with SNPs to attempt to identify any functional effects of the mutations. A total of 2074 SNPs were located in CDS, 1310 (63.2%) were non-synonymous, and 764 (36.8%) were synonymous. These SNPs were distributed in 1480 genes, and each of the genes had 1.40 SNP sites, on average. Then, each of the CDS was categorized according to their cog division. As shown in Fig. 2, SNPs occurred in genes of all the 21 functional categories, but genes with mutation were distributed differentially among these categories. The most repre-

Table 1 SNP, MNV and INDEL sites identified in the genome sequence

Variant type	Number	Coding region	Non-coding region	Non-synonymous	Synonymous
SNP	2548	2074	474	1310	764
MNV	65	50	15	–	–
INDEL	241	151	90	–	–

Table 2 Base change frequency of SNPs

Reference	Base no.	Variant Base	Variant no.	Syn no.	Non-syn no.	Frequency/%	Transition/%	Transversion/%
A	513	C	92	24	68	17.9	–	17.9
		G	381	152	229	74.3	–	74.3
		T	40	10	30	7.80	7.80	–
C	558	A	99	28	71	17.7	–	17.7
		G	99	20	79	17.7	17.7	–
		T	360	153	207	64.5	–	64.5
G	506	A	341	141	200	67.4	–	67.4
		C	81	26	55	16.0	16.0	–
		T	84	18	66	16.6	–	16.6
T	497	A	36	7	29	7.24	7.24	–
		C	373	161	212	75.1	–	75.1
		G	88	24	64	17.7	–	17.7

Note: no., number; Syn, synonymous; Non-syn, non-synonymous.

Table 3 Distribution of small deletions and insertions

Indel type	Number	CDS	Intergenic
Deletion 1	126	79	47
Deletion 2	12	4	8
Deletion 3	5	1	4
Deletion 4	2	1	1
Insertion 1	91	3	88
Insertion 2	4	2	2
Insertion 3	1	0	1

sented categories were no cluster of homologous groups (with unknown function), followed by that of predicted function. The percent of changed genes in each of the COG divisions was calculated. Change was highest for the functional category of secondary metabolite biosynthesis, followed by cell envelope biogenesis and amino acid transport and metabolism. In contrast, change was low in categories of transcription, cell motility and secretion, signal transduction mechanism, and unknown function.

We then asked whether these changes resulted in functional alterations. Synonymous and non-synonymous SNPs were analyzed and compared between different functional groups. SNP numbers were divided by gene number for each group to calculate SNP site per gene. The SNPs per gene ranged from 2.17 (function unknown) to 3.46 (secondary metabolite biosynthesis) for different COGs. The ratio of non-synonymous to synonymous SNPs (N/S) was also calculated for each category. N/S ratio for cell motility and secretion categories was 5.00, being the highest. The ratios for intracellular traffic, inorganic ion transport and metabolism, secondary metabolite biosynthesis, lipid metabolism and transcription were more than 2-fold. Only the ratio for the signal transduction mechanism category was lower than 1 (0.81). That is, signal transduction mechanism showed more synonymous mutation than non-synonymous ones. The N/S ratio was also calculated for each of the genes. Genes related to defense, metabolism and energy productions showed higher N/S ratio changes.

3.5 Genetic changes in virulence related genes

Previously, a number of genes involved in virulence of *Brucella* had been identified and confirmed experimentally. We analyzed the distribution of SNPs among these virulence or virulence-related genes and identified SNPs in known virulence and virulence-related genes. The *virB* operon encodes the type IV secretion system that plays an essential role in *Brucella* intracellular survival. Several *virB* genes, including *virB2*, *virB4*, *virB6* and *virB9*, had SNPs. The effector protein, VceC also had 2 non-synonymous SNPs. SNPs were identified in 13 flagella-

related genes. Ten of these genes had non-synonymous SNPs. Outer membrane protein Omp25 is involved in *Brucella* virulence. A total of 4 *omp25* genes exists in the *Brucella* genome, 2 of them, BMEI1830 and BMEI1249, had SNPs. Some virulence regulating genes, such as *gntR* and *lysR*, had SNP sites. The SNP intensity for the virulence genes ranged from 0.01 to 6.00 per kilobase with *virB6* and *virB9* having 1.92 and 2.69 non-synonymous SNP per kilobase, respectively.

3.6 Reduced expression of virulence genes in M5-10

The genetic changes and reduced virulence of M5-10 implied that expression of virulence genes might be reduced. To test this, M5-10 and virulent strain 16M were culture in TSB and then RNA was isolated and transcription of several important virulence genes was analyzed by real time-PCR. As shown in Fig. 3, compared with that in 16M, expression of *dnaK*, *vjbR* and *omp25* was greatly reduced in M5-10. The reduction in expression of these virulence genes indicated that this might be the reason for the reduced virulence of M5-10.

3.7 Signature sequences for M5-10

Genetic changes occur at random, and the reversion of a mutation is very rare. It is possible that the genetic changes, including SNP, INDEL and MNV, might be specific for M5-10, and if so, could be candidate signature sequences for the changes in this strain. To test whether a sequence could be used as signature sequence, the candidate sequences could be compared with sequences from other strains. Any unique distribution of the genetic changes found could be used as a signature for the identification of a strain. Small deletions and insertions of M5-10 are good candidates for differentiating this vaccine strain from other *Brucella* strains. To develop a differential diagnosis method for M5-10, a duplex PCR, which targeted the *virB6* and *dnaK*, was developed. The 3' end of the primers for *virB6* was located at the SNP, and only wild type sequences will generate product. At the same time, another conserved gene, *dnaK*, was used to identify *Brucella*. As shown in Fig. 3, the product of *virB6* could be amplified from 16M but not from M5-10. *DnaK* could be amplified from each of the strains. Therefore, duplex PCR could be used to differentiate M5-10 from other virulent strains.

4 Discussion

Brucellosis is a common zoonosis that is difficult to diagnose and treat. Human brucellosis is derived from animal brucellosis. Therefore, control and/or eradication is dependent on control of animal brucellosis. The use of live

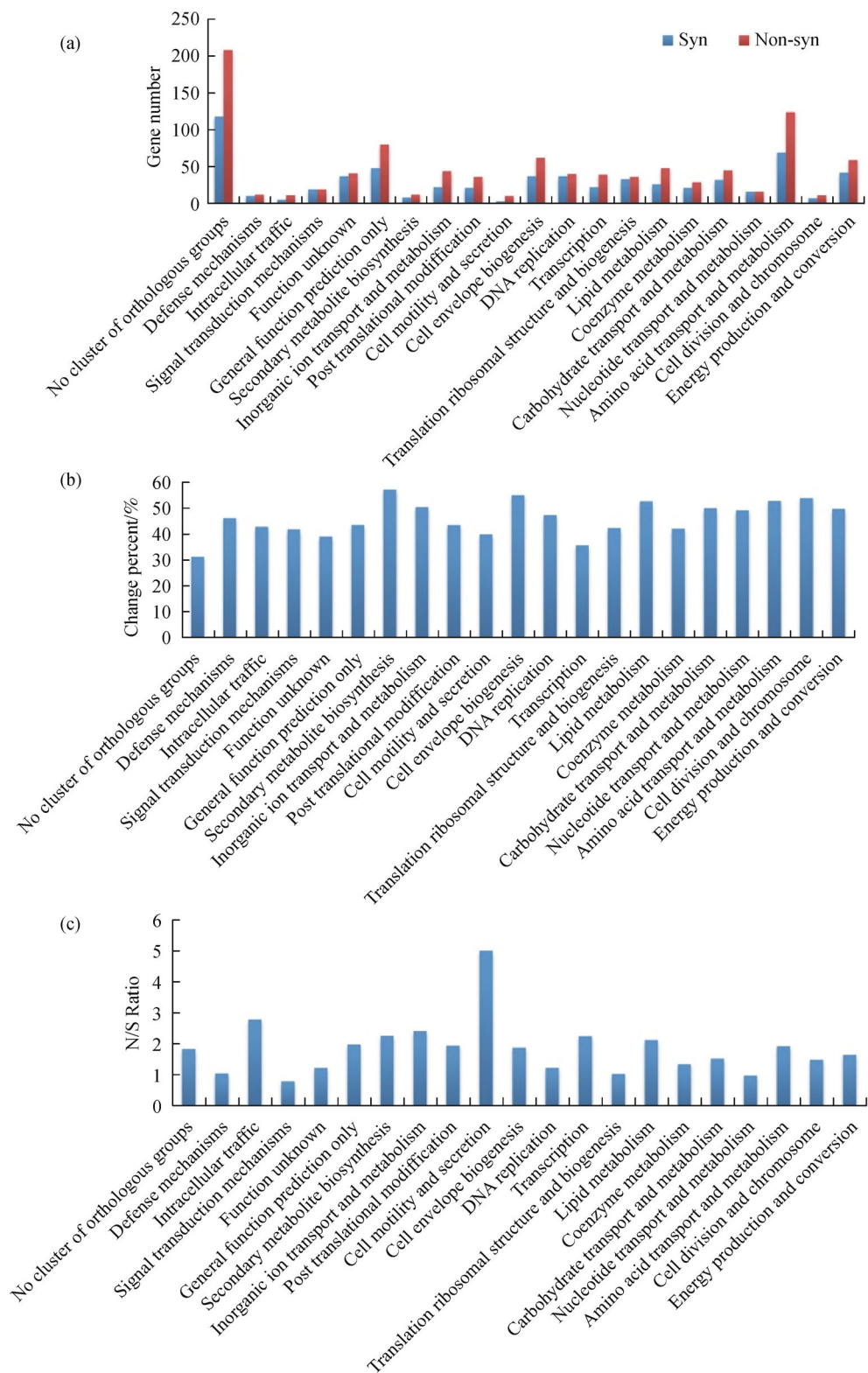


Fig. 2 Distribution of SNPs in genes of different functional categories. (a) Gene number of different COG categories; (b) percent of genes with genetic changes in each of the cog categories; (c) N/S ratio for each of the cog categories. N/S: the ratio of non-synonymous to synonymous SNPs; Syn: synonymous; Non-syn: non-synonymous.

attenuated vaccine is one of the most important strategies for prevention of animal brucellosis. Since development

and license of the first vaccine, vaccination has been applied in many countries for decades and great successes

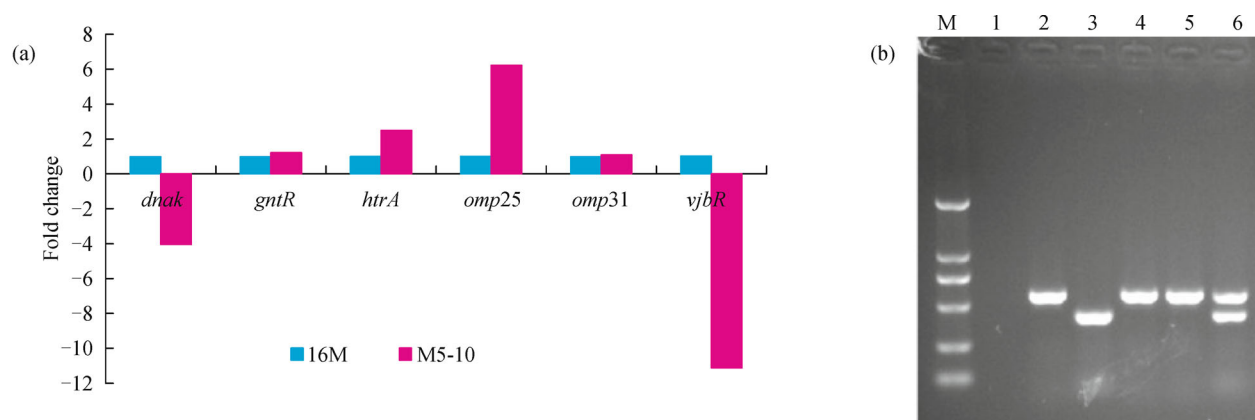


Fig. 3 Reduced expressions of virulence genes and differential diagnosis by PCR. (a) Relative expression of selected virulence genes in M5-10 compared with 16M; (b) duplex amplification of M5-10 and 16M. Lane 1, 3: amplification of *virB6*; lanes 2, 4: amplification of *dnak*; lanes 5, 6: duplex amplification of *virB6* and *dnak*. Lanes 1, 2 and 5: M5-10; lane 3, 4 and 6: 16M.

have been achieved in these programs. For the present, all the licensed vaccines are made with live attenuated strains, which have been developed by conventional strategies of sequential passages in non-natural hosts. Genome sequences and their genetic characteristics will provide information about their attenuation mechanism(s), which will also contribute to understanding the virulence mechanism of *Brucella*. Till now, genome sequences for some of these vaccine strains had been decoded and possible attenuation mechanisms have been predicted [8,9].

Brucella vaccine strains are usually developed by sequential passage in non-natural host. During the selection, bacterial adaptation leads to genetic changes that make the strain more adaptive to the non-natural host, but also cause the loss of the capability to survive in its natural hosts. Many bacterial vaccines were developed by this mechanism, however, because this is a results driven process, many of the mechanisms and genetic characteristics of attenuation remains unknown. With the development of genome sequencing technology, many bacterial strains, including vaccine strains, were sequenced and their possible pathogenesis mechanisms putatively defined [10].

Two live attenuated vaccines, *B. melitensis* M5 and *B. suis* S2, were licensed in China [7]. The two vaccines have been extensively applied in China and have been important in reducing the incidence of brucellosis. In the present study, we sequenced and analyzed vaccine strain M5-10, a derivative of M5 that stored in laboratory. First, we obtained genome sequence of M5-10 with Hiseq2000. A high sequencing coverage was obtained and this will generate higher fidelity, which is important for mutation detection. With the generated reads, contigs were assembled, and the coverage of these contigs ranged from 260 to 3000 times, with a coverage of about 300 times for most contigs. This high coverage makes polymorphism detection more specific and accurate. The differences in coverage might have resulted from the copy

number of the target regions, or differences in amplification efficiency of the genomic DNA during the library construction.

Genetic polymorphism includes mainly SNP, INDEL and MNV. In the genome sequence of M5-10, we observed a total of 2854 sites. Interestingly, a significantly higher percent of polymorphism sites were observed in coding regions than in intergenic ones. In addition, the change ratios, which exclude influence of length, were also higher in coding regions than intergenic ones. The greater occurrence of mutations in coding regions implied that they lead to functional changes. *Brucella* has 2 chromosomes, polymorphic sites occurred at higher frequency in chromosome II than in chromosome I. Chromosome II has been considered as a large plasmid or part of the chromosome I [11]. Significantly, a high proportion of the genes located on chromosome II are involved in environmental adaptation. All these data implied that the attenuation of the vaccine strains could be considered as a process of adaptation to the new environments encountered in the non-natural hosts.

Nucleic acid replacements include two types, transition and transversion [12]. Transversion refers to the substitution of a purine for a pyrimidine or vice versa. The consequence of this type of mutation is a more dramatic change in the chemical structure, than occurs with transitions. Transition is a point mutation that changes a purine nucleotide to another purine ($A \leftrightarrow G$) or a pyrimidine nucleotide to another pyrimidine ($C \leftrightarrow T$). Approximately two out of three single nucleotide polymorphisms (SNPs) are transitions. The higher frequency of transition over transversion indicates it is an efficient strategy used by bacteria for causing functional changes.

Changes in different functional categories also indicated the adaptation of the vaccine strain. Of the 2074 SNPs, 1310 were non-synonymous and 764 were synonymous. That is, a significantly higher percentage of the SNPs resulted in amino acid changes, and accordingly potential

functional changes. When the functional categories were analyzed, genes of all the categories had genetic changes. The most highly represented category is the group with unknown function. A possible reason for this is that a significant proportion of the *Brucella* genes were predicted with unknown function. To exclude the influence of gene number for each of the categories, change percent of the genes for each category was calculated. Interestingly, changes in the functional category of secondary metabolites biosynthesis, cell envelope biogenesis and amino acid transport and metabolism, were significantly higher than in other categories. In contrast, transcription, cell motility and signal transduction showed a lower level of changes.

Brucella is a facultative intracellular bacterium. Virulence genes of *Brucella* are usually those involved in environmental adaptation and intracellular survival. We found that many of the virulence and virulence related genes were also mutated (Table 4). Of these virulence genes, the *virB* operon represents one of the most important virulent determinants. Importantly, 4(36.4%) genes of the *virB* operon have mutated sites. Non-

synonymous sites were also identified in the *virB* genes. Although the definite effects of these mutations in *virB* genes need to be further defined, it could be predicted that these modifications result in functional changes. It is possible that mutation in the *virB* will affect the function of the type IV secretion system, and accordingly the intracellular survival. Flagella have not been observed in *Brucella* [13]. However, disruption of flagellar genes resulted in attenuation of the mutant, implying that flagellar genes were involved in *Brucella* virulence [14]. Omp25 is an important outer membrane protein that plays great roles in integrity of the membrane [15]. Mutations identified in this gene implied possible membrane changes in the vaccine strain. Another type of gene, the virulence-related genes, which affect virulence by regulating expression of other genes involved in adaptation or virulence, also showed genetic changes. The *gntR* genes represent one of the important regulatory genes contributing to bacterial adaptation to environments [16]. Genetic changes in these genes might also result in attenuation. Expression analysis indicated that, in addition to the

Table 4 Selected virulence genes and virulence related genes with SNPs

Synonym	Gene	Product	Non-syn	Syn	Total	Length	SNPs/kb	nSNPs/kb	sSNPs/kb	cog single
BMEI0026	<i>virB2</i>	VirB2	–	2	2	318	6.29	0.00	6.29	–
BMEI0033	<i>virB9</i>	Channel protein VirB9 homolog	1	1	2	744	2.69	1.34	1.34	U
BMEI0030	<i>virB6</i>	Channel protein VirB6 homolog	2		2	1044	1.92	1.92	0.00	U
BMEI0948	<i>vceC</i>	Hypothetical protein	2		2	1086	1.84	1.84	0.00	–
BMEI0028	<i>virB4</i>	ATPase VirB4 homolog	1		1	2496	0.40	0.40	0.00	U
BMEI1830	<i>omp25</i>	25 kDa outer-membrane immunogenic protein precursor	1	1	2	693	2.89	1.44	1.44	M
BMEI1249	<i>omp25</i>	25 kDa outer-membrane immunogenic protein precursor	–	1	1	642	1.56	0.00	1.56	M
BMEI0748	<i>L7/L12</i>	50S ribosomal protein L7/L12	–	1	1	375	2.67	0.00	2.67	J
BMEI1048	<i>GroEL</i>	Chaperonin GroEL	1	1	2	1641	1.22	0.61	0.61	O
BMEI2002	<i>dnaK</i>	Molecular chaperone dnaK	1	–	1	1926	0.52	0.52	0.00	O
BMEI0754	<i>EF-2</i>	Elongation factor EF-2	2	–	2	2085	0.96	0.96	0.00	J
BMEI1086	<i>flgG</i>	Flagellar basal-body rod protein flgG	3	–	3	789	3.80	3.80	0.00	N
BMEI0154	–	Flagellar motor protein	3	–	3	1101	2.72	2.72	0.00	N
BMEI1085	–	Flagellar basal body P-ring biosynthesis protein	1	–	1	414	2.42	2.42	0.00	N
BMEI1110	<i>flim</i>	Flagellar motor switch protein flim	1	1	2	951	2.10	1.05	1.05	N
BMEI1107	<i>flgF</i>	Flagellar basal-body rod protein flgF	1	–	1	522	1.92	1.92	0.00	N
BMEI1692	<i>flgJ</i>	Flagellar protein flgJ	3	–	3	1917	1.56	1.56	0.00	N
BMEI0151	<i>flfF</i>	Flagellar m-ring protein flfF	–	1	1	729	1.37	0.00	1.37	N
BMEI1080	–	Flagellar biosynthesis protein	1	–	1	741	1.35	1.35	0.00	N
BMEI1109	–	Flagellar motor protein	1	–	1	972	1.03	1.03	0.00	N
BMEI0152	–	Flagellar m-ring protein flfF	1	–	1	1002	1.00	1.00	0.00	N
BMEI0324	–	Flagellar motor protein	–	1	1	1032	0.97	0.00	0.97	N
BMEI0159	–	Flagellar hook protein	1	–	1	1191	0.84	0.84	0.00	N
BMEI0167	–	Flagellar biosynthesis protein	–	1	1	1770	0.56	0.00	0.56	N

(Continued)

Synonym	Gene	Product	Non-syn	Syn	Total	Length	SNPs/kb	nSNPs/kb	sSNPs/kb	cog single
BMEI0881	<i>gntR5</i>	Transcriptional regulator, gntR family	–	2	2	762	2.62	0.00	2.62	K
BMEI1913	<i>lysR13</i>	Transcriptional regulator, lysR family	2	–	2	900	2.22	2.22	0.00	K
BMEI0305	<i>gntR2</i>	Transcriptional regulator, gntR family	1	–	1	633	1.58	1.58	0.00	K
BMEI0320	<i>gntR17</i>	Transcriptional regulator, gntR family	–	1	1	693	1.44	0.00	1.44	K
BMEI2036	–	Transcriptional regulatory protein chvi	–	1	1	720	1.39	0.00	1.39	T
BMEI0390	<i>lysR12</i>	Transcriptional regulator, lysR family	1	–	1	933	1.07	1.07	0.00	K
BMEI1709	–	Oxidoreductaseucpa	3	1	4	735	5.44	4.08	1.36	I
BMEI1305	–	Porin	2	3	5	1128	4.43	1.77	2.66	-
BMEI0322	–	Cytochrome c oxidase polypeptide iv	1	–	1	240	4.17	4.17	0.00	-
BMEI1980	–	DNA protection during starvation conditions	1	–	1	498	2.01	2.01	0.00	P
BMEI1483	–	50S ribosomal protein L9	–	1	1	570	1.75	0.00	1.75	J
BMEI0319	–	Bioy protein	–	1	1	576	1.74	0.00	1.74	R
BMEI1053	<i>gluP</i>	Glucose/galactose transporter	2	–	2	1239	1.61	1.61	0.00	G
BMEI1985	–	Phosphate transport system protein phou	–	1	1	723	1.38	0.00	1.38	P
BMEI0330	–	Opgc	1	–	1	1203	0.83	0.83	0.00	S
BMEI2035	–	Sensor protein chvg	–	1	1	1806	0.55	0.00	0.55	T

genetic changes, expression of the virulence genes were also reduced in the vaccine strains, further indicating the attenuation of M5-10.

Distinguishing between immunization and natural infection represents a great challenge for live attenuated vaccine. There are no ideal methods available for differential diagnosis of immunization and natural infection. The polymorphic sites in the vaccine strains might represent signature candidates for such a differential diagnosis. As shown in Appendix B, polymorphic sites identified in the vaccine strains could be candidate sites for molecular methods to differentiate vaccine strains from other virulent strains. It may be possible to develop a duplex PCR assay based on sequence polymorphism for such differential diagnosis.

5 Conclusions

The licensed attenuated vaccine, *B. melitensis* M5, has been widespread used in China. However, the attenuation mechanism is still unknown. In our study, the genome of M5-10 strain, a derivative of M5, was sequenced and mapped to that of the virulent 16M strain. A total of 2854 genetic polymorphic sites, including 2548 SNP, 241 INDEL and 65 MNV were identified. Of the 2074 SNPs in coding regions, 1310 (63.2%) were non-synonymous SNPs. The genetic polymorphic sites in *virB* operon, flagella synthesis, and virulence regulating systems coding genes indicate changes in some cog categories and virulence genes, which might result in the attenuation.

Reduced expression of the *dnaK* and *yjbR* genes were proved in M5-10 strain. The two strains of M5-10 and 16M have been successfully differentiated by a duplex PCR targeting *virB6* and *dnaK*. Our findings might provide clues to vaccine development and differential diagnosis.

Acknowledgements This work was supported by grants from National Twelfth Five-Year Plan for Science & Technology Support of China (2014BAI13B03), National Natural Science Foundation of China (81460248, 81260457, 81271899, 31272592, 81071320), Inner Mongolia Natural Science Funds (2013MS1138, 2012MS1121, 2011MS1110), Beijing Natural Science Foundation (6122030, 7132153) and Inner Mongolia Science & Technology Plan (20120101, 20120402, 20110502), by the National Key Program for Infectious Diseases of China (2013ZX10004-203, 2013ZX10004-217-002, 2013ZX10004805-006).

Supplementary materials The online version of this article at <http://dx.doi.org/10.15302/J-FASE-2014032> contains supplementary material (Appendix A and B).

Compliance with ethics guidelines Yuanqiang Zheng, Yanchun Shi, Chang An, Ruisheng Li, Dongjun Liu, YuehuaKe, Kairong Mao, Mingjuan Yang, Zeliang Chen and Shorgan Bou declare that they have no conflict of interest or financial conflicts to disclose.

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

1. Pappas G, Papadimitriou P, Akritidis N, Christou L, Tsianos E V. The new global map of human brucellosis. *The Lancet Infectious Diseases*, 2006, 6(2): 91–99
2. Mesner O, Riesenber K, Biliar N, Borstein E, Bouhnik L, Peled N,

- Yagupsky P. The many faces of human-to-human transmission of brucellosis: congenital infection and outbreak of nosocomial disease related to an unrecognized clinical case. *Clinical Infectious Diseases*, 2007, 45(12): e135–140
3. Al Dahouk S, Neubauer H, Hensel A, Schöneberg I, N°Ckler K, Alpers K, Merzenich H, Stark K, Jansen A. Changing epidemiology of human brucellosis, Germany, 1962-2005. *Emerging Infectious Diseases*, 2007, 13(12): 1895–1900
 4. McGiven J A. New developments in the immunodiagnosis of brucellosis in livestock and wildlife. *Revue Scientifique et Technique (International Office of Epizootics)*, 2013, 32(1): 163–176
 5. Wang Y, Bai Y, Qu Q, Xu J, Chen Y, Zhong Z, Qiu Y, Wang T, Du X, Wang Z, Yu S, Fu S, Yuan J, Zhen Q, Yu Y, Chen Z, Huang L. The 16MΔvjbR as an ideal live attenuated vaccine candidate for differentiation between *Brucella* vaccination and infection. *Veterinary Microbiology*, 2011, 151(3–4): 354–362
 6. Ficht T A, Kahl-McDonagh M M, Arenas-Gamboa A M, Rice-Ficht A C. Brucellosis: the case for live, attenuated vaccines. *Vaccine*, 2009, 27(Suppl 4): D40–D43
 7. Deqiu S, Donglou X, Jiming Y. Epidemiology and control of brucellosis in China. *Veterinary Microbiology*, 2002, 90(1–4): 165–182
 8. Crasta O R, Folkerts O, Fei Z, Mane S P, Evans C, Martino-Catt S, Bricker B, Yu G, Du L, Sobral B W. Genome sequence of *Brucella abortus* vaccine strain S19 compared to virulent strains yields candidate virulence genes. *PLoS ONE*, 2008, 3(5): e2193
 9. Eschenbrenner M, Wagner M A, Horn T A, Kraycer J A, Mujer C V, Hagius S, Elzer P, DelVecchio V G. Comparative proteome analysis of *Brucella melitensis* vaccine strain Rev 1 and a virulent strain, 16M. *Journal of Bacteriology*, 2002, 184(18): 4962–4970
 10. Ding J, Pan Y, Jiang H, Cheng J, Liu T, Qin N, Yang Y, Cui B, Chen C, Liu C, Mao K, Zhu B. Whole genome sequences of four *Brucella* strains. *Journal of Bacteriology*, 2011, 193(14): 3674–3675
 11. Rajashekara G, Glasner J D, Glover D A, Splitter G A. Comparative whole-genome hybridization reveals genomic islands in *Brucella* species. *Journal of Bacteriology*, 2004, 186(15): 5040–5051
 12. Yang Z, Yoder A D. Estimation of the transition/transversion rate bias and species sampling. *Journal of Molecular Evolution*, 1999, 48(3): 274–283
 13. Ferooz J, Letesson J J. Morphological analysis of the sheathed flagellum of *Brucella melitensis*. *BMC Research Notes*, 2010, 3(1): 333
 14. Fretin D, Fauconnier A, Köhler S, Halling S, Léonard S, Nijskens C, Ferooz J, Lestrade P, Delrue R M, Danese I, Vandenhoute J, Tibor A, DeBolle X, Letesson J J. The sheathed flagellum of *Brucella melitensis* is involved in persistence in a murine model of infection. *Cellular Microbiology*, 2005, 7(5): 687–698
 15. Edmonds M D, Cloeckert A, Elzer P H. *Brucella* species lacking the major outer membrane protein Omp25 are attenuated in mice and protect against *Brucella melitensis* and *Brucella ovis*. *Veterinary Microbiology*, 2002, 88(3): 205–221
 16. Haine V, Sinon A, Van Steen F, Rousseau S, Dozot M, Lestrade P, Lambert C, Letesson J J, De Bolle X. Systematic targeted mutagenesis of *Brucella melitensis* 16M reveals a major role for GntR regulators in the control of virulence. *Infection and Immunity*, 2005, 73(9): 5578–5586