Comparative Genomic Analysis of *Brucella melitensis* Vaccine Strain M5 Provides Insights into Virulence Attenuation

Hai Jiang1, Pengcheng Du1, Wen Zhang1, Heng Wang2, Hongyan Zhao1, Dongri Piao1, Guozhong Tian1, Chen Chen1*, Buyun Cui1*

1 State Key Laboratory for Infectious Disease Prevention and Control, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Beijing, China. 2 Department of Endemic and Parasitic Diseases Control and Prevention, Hangzhou Center for Disease Control and Prevention, Hangzhou, China

Introduction

Brucellosis is one of the most common bacterial zoonoses endemic in many countries, particularly those with developing livestock industries [1]. Among these, China is a large stock-raising country with approximately 300 million sheep [2]. Along with the increased number of animals, the incidence of human brucellosis has also increased rapidly over the last 10 years, which has negatively affected the public health and the livestock industry [3]. *B. melitensis* was the predominant strain associated with human *B. melitensis* negatively affected the public health and the livestock industry [3].

Addition, a genome fragment with a 56 kb rearrangement was determined to be consistent with previous studies. Comparative genomic analysis indicated that genomic island inversion in *Brucella* was widely present. With the genetic pattern common among all strains analyzed, these 2 RDs, 1 RI, and one inversion region are potential sites for detection of genomic differences. Several SNPs of important virulence-related genes (*motB, dhbC, sfuB, dsbAB, aidA, aroC*, and *lysR*) were also detected, and may be used to determine the mechanism of virulence attenuation. Collectively, this study reveals that comparative analysis between wild-type and vaccine strains can provide resources for the study of virulence and microevolution of *Brucella*.

Abstract

The *Brucella melitensis* vaccine strain M5 is widely used to prevent and control brucellosis in animals. In this study, we determined the whole-genome sequence of M5, and conducted a comprehensive comparative analysis against the whole-genome sequence of the virulent strain 16 M and other reference strains. This analysis revealed 11 regions of deletion (RDs) and 2 regions of insertion (RIs) within the M5 genome. Among these regions, the sequences encompassed in 5 RDs and 1 RI showed consistent variation, with a large deletion between the M5 and the 16 M genomes. RD4 and RD5 showed the large diversity among all *Brucella* genomes, both in RD length and RD copy number. Thus, RD4 and RD5 are potential sites for typing different *Brucella* strains. Other RD and RI regions exhibited multiple single nucleotide polymorphisms (SNPs). In addition, a genome fragment with a 56 kb rearrangement was determined to be consistent with previous studies. Comparative genomic analysis indicated that genomic island inversion in *Brucella* was widely present. With the genetic pattern common among all strains analyzed, these 2 RDs, 1 RI, and one inversion region are potential sites for detection of genomic differences. Several SNPs of important virulence-related genes (*motB, dhbC, sfuB, dsbAB, aidA, aroC*, and *lysR*) were also detected, and may be used to determine the mechanism of virulence attenuation. Collectively, this study reveals that comparative analysis between wild-type and vaccine strains can provide resources for the study of virulence and microevolution of *Brucella*.
associated with virulence in 16 M or lack of virulence in M5. Using bioinformatics, this study is the first to perform whole genome analysis directed toward identifying the sequence differences between the \textit{B. melitensis} vaccine strain M5 and the virulent strain 16 M and reveals a mechanism of attenuation genes.

**Results and Discussion**

**General features of the \textit{B. melitensis} vaccine strain M5 genome**

High-throughput bioinformatics revealed 26 genes present in the M5 genome. These genes were compared to those of the 16 M strain, and all were found to be homologues of the 16 M genome. These data suggest that the genomes of M5 and 16 M have a very similar genomic structure.

**Genome assembly and annotation.** The draft-quality genome of the vaccine strain M5 was sequenced and assembled into 95 contigs with N50 = 775,745 bp (Table S1). Scaffold 1, the largest scaffold, comprised most of chromosome II. Further analysis by PCR sequencing showed that this scaffold contained a complete 16S rRNA gene which was different from the copy in chromosome I. However, we did not continue to close the whole gap caused by duplication of the 16S-23S-5S rRNA operon. Composed to the sequenced genome of \textit{B. melitensis} 16 M, the size, number of open reading frames (ORFs), guanine–cytosine (GC) content, gene length, and gene density were largely conserved in the genome of M5 (Table 1).

In the M5 genome, we identified 3305 ORFs that were larger than 300 bp (100aa), covering 86.5% of the genome, of which 90.9% matched the NCBI COG database with an e-value less than 1e-5 (Table S2). Many genes of these ORFs were located in common pathogenic pathways, such as \textit{virB} and \textit{bcsr/bcrR}. In addition, the previously reported genomic island containing a type IV secretion system [14] was also identified in the vaccine strain M5. Moreover, two additional GC islands with unknown functions were identified, and appeared to be possible transfer elements by horizontal transfer with extremely low GC content. We also detected 43 rRNA operons in the genome, which was consistent with the \textit{B. melitensis} 16 M. However, 2 SNPs in the rRNA sequences within chromosome II of M5 were identified when compared to those of chromosome I.

**Genome region differences**

**Genomic arrangement.** Although there was homology between the genes of M5 and 16 M, the genome layout of M5 suggests a ~56 kb inversion or rearrangement in the head of scaffold 1, which mapped to chromosome II of 16 M. This region is flanked by two duplicate sequences of the insertion sequence (IS) IS1953. Two genes located upstream (BMEI0183/BMII0184) and downstream (BMEI0227/BMII0228) of these inversion sequences were associated with transposes. In addition, inactivated derivatives were found in the IS elements. These findings suggest that the inversion was due to the duplication of insertion elements.

Aligning this 56 kb inversion region with other \textit{Brucella} strains revealed that genomic inversions vary among available genomes including; \textit{B. microti} CCM 4915, \textit{B. canis} ATCC 23365, \textit{B. suis} ATCC 23445, and \textit{B. suis} 1330; similar to \textit{B. ovis} ATCC 2540 and \textit{B. melitensis} ATCC 23457; and missing in \textit{B. abortus} bv.1 str.9–941, \textit{B. abortus} S19, and \textit{B. Abortus} 2308 (Table 2). Functional analysis of this 56 kb inversion indicated that the region contains a large number of peptide ABC transporter genes (19 of 42 genes) such as spermidine–putrescine (\textit{psq}), oligopeptidase (\textit{opp}), and dipeptide (\textit{dpp}) systems. These ABC transporter genes play an important role the virulence of \textit{Brucella} strains, and can influence trafficking of \textit{Brucella} to compartments associated with endoplasmic reticulum where peptides and amino acids can be found in large numbers. However, absence of these transporters can be negative effects. For example, loss of the \textit{opp} system in \textit{B. ovis} may deregulate peptide uptake resulting in increased peptide levels and reduced intracellular survival [18].

**Genome deletions.** This inversion region revealed a potential marker for distinguishing infected animals from animals vaccinated with M5. To explore this possibility, we designed primers and verified the results by testing various strains. Comparative genomic analysis identified 11 low read coverage regions and 2 high coverage regions of more than 300 bp each in the 16 M genome. Except for deletion, the other possible reason previous SOLEXA reads matching failed to identify these regions is that they were side effects of genomic rearrangement, where reads only match references with half-lengths. However, these characteristics may not produce a gap over 300 bp (100aa, 4 folds of reads length). Among these regions, we hypothesized that the low coverage regions in the 16 M genome comprised the majority of the genomic differences observed in the M5 genome. To test this, we examined with the assembled genome of M5 to verify whether these regions were missing. However, the large regions were not missing in the M5 draft genome, and only 11 large RDs

![Table 2](https://example.com/Table2.png)

**Table 1.** Genomic features of the newly sequenced genome of the \textit{B. melitensis} vaccine strain M5 compared with the known genomic sequences of the virulent strain 16 M.

| Genome Feature       | M5   | 16M  |
|----------------------|------|------|
| G+C content(%)       | 57.25| 57.22|
| # of ORFs            | 3305 | 3198 |
| % of ORFs in Genome  | 86.51| 87.60|
| Average Length of ORFs| 862  | 894  |
| # of rRNA            | 49   | 54   |
| # of rRNA operons    | 3    | 3    |

**Table 2.** Inversion on the \textit{B. melitensis} vaccine strain M5 chromosome II aligned to reference sequences.

| Accession Number | Strain     | Biovar | Inversion* |
|------------------|------------|--------|------------|
| NC_003318        | \textit{Brucella melitensis} 16M | biovar 1 | +          |
| NC_013118        | \textit{Brucella microti} CCM 4915 | –      | –          |
| NC_010104        | \textit{Brucella canis} ATCC 23365 | –      | +          |
| NC_010167        | \textit{Brucella suis} ATCC 23445 | biovar 2 | +          |
| NC_004311        | \textit{Brucella suis} 1330 | biovar 1 | +          |
| NC_009504        | \textit{Brucella ovis} ATCC 25840 | –      | –          |
| NC_012442        | \textit{Brucella melitensis} ATCC 23457 | biovar 2 | –          |
| NC_006933        | \textit{Brucella abortus} 9–941 | biovar 1 | No hit     |
| NC_010740        | \textit{Brucella abortus} 519 | biovar1 | No hit     |
| NC_006924        | \textit{Brucella abortus} 2308 | biovar1 | No hit     |

*++", inversion; "--", the same direction; "No hit", no alignment (but no deletion).
were found: RD1 to RD11, 5 in chromosome I, and 6 in chromosome II (Table 3).

The differences between the whole genomes of M5, 16 M, and other Brucella strains, are presented in Table 3 and Figure 1. Sixteen genes of 16 M genome were located in the RDs. The accurate genomic positions of these RDs were determined using the assembled M5 genome data. After the assembled M5 genome was compared with the 16 M genome, 5 of 11 RDs (RD1, RD2, RD4, RD5, and RD7) were confirmed to vary in the 16 M genome. The variation of these 5 RDs are described as follows:

RD1 included two ORFs, BMEI0194 and BMEI0195. BMEI0194 encodes a hypothetical cytosolic protein, whereas BMEI0195 encodes the ATP-dependent chaperone ClpB, which is essential for heat-shock response and belongs to the Class 1 family of Clp/Hsp100 AAA+ ATPases. It is involved in the recovery of the cell from heat-induced damage in cooperation with DnaK, DnaJ and GrpE.

RD2 contained 1 ORF, BMEI0934, encoding the ATP-dependent RNA helicase, which catalyzes the unwinding of double-stranded nucleic acids. Helicases are tightly integrated (or coupled) components of various macromolecular complexes involved in processes such as DNA replication, recombination, and nucleotide excision repair, as well as RNA transcription and splicing.

RD4 contained 1 ORF, BMEI0407, which encodes a hypothetical protein. A 59 bp repeat region was found in the flank of this region in the 16 M genome, and each of the two flanking repeats was duplicated on each side (Fig S1). This region is a candidate for multi-locus variable-number tandem-repeat analysis.

RD5 contained 2 ORFs, BMEI1661 and BMEI1662. BMEI1661 is part of an operon, whereas BMEI1661 belongs to the resolvase family (Fig S1). Notably, BMEI0902 in chromosome I of 16 M also belongs to the resolvase family. Located in a region with high read coverage, this gene is considered a duplicate in the M5 genome. Of the whole 231-bp sequence of BMEI0902, 180 bp was identical to BMEI1661. This observation suggests that the two genes are possible homologs, although BMEI1661 contained other sequences more than 500 bp that were not present in any other species. This kind of use bias in different site-specific recombinases may lead to various acquisition abilities among different DNA fragments.

RD7 contained 1 ORF, BMEI0259, which encodes the Asp-tRNAAsn/Glu-tRNAGln aminotransferase A subunit and related amidases. In this gene, a 21 bp fragment in M5 replaced a 6 bp fragment in 16 M. This mutation may minimally affect the protein function because a 36 nucleotide insertion in a gene does not cause a downstream frame shift. Thus, we considered this region a normal mutation causing the low read coverage.

The other 6 RDs may be caused by large numbers or continual SNPs in the core regions (Figure S1). These regions were compared to all available Brucella genomes. In addition to RD4 and RD5, all other RDs were specific in the vaccine strain M5 compared with the virulent strain 16 M. RD4 and RD5 showed large divergence among Brucella, suggesting that they are active mobile elements and can be targets for typing and detection of Brucella. RD4 is a tandem repeat with different lengths in each genome ranging from 38 bp to 342 bp. A similar method was investigated in all Brucella genomes. Figure 1 shows all the RDs compared with the M5 genome.

### Table 3. Genes deleted from RDs of the B. melitensis vaccine strain M5.

| Chromosome | Log2(CDM*) | RD # | RD Length | Gene name | % in RD | Annotation |
|------------|------------|------|-----------|-----------|---------|------------|
| I          | >−2, <−1   | RD1  | 337       | BMEI0194  | 33.2    | Putative cytoplasmic protein |
|            |            | RD1  | 337       | BMEI0195  | 2.07    | ATPases with chaperone activity, ATP-binding subunit |
|            |            | RD2  | 309       | BMEI0934  | 14.55   | Superfamily II DNA and RNA helicases |
|            |            | RD3  | 906       | BMEI1965  | 6.54    | Translation initiation factor 2 (IF-2; GTPase) |
|            | ≤−2        | RD1a | 461       | BMEI0194  | 53.89   | Putative cytoplasmic protein |
|            |            | RD1a | 461       | BMEI0195  | 0.89    | ATPases with chaperone activity, ATP-binding subunit |
|            |            | RD4  | 342       | BMEI0407  | 39.24   | Unknown function |
|            |            | RD5  | 883       | BMEI1661  | 74.97   | Site-specific recombinases, DNA invertase Pin homologs |
| III         | >−2, <−1   | RD5  | 883       | BMEI1662  | 73.89   | Hypothetical protein BMEI1662 |
|            |            | RD6  | 306       | BMEI0253  | 29.91   | Murein endopeptidase |
|            |            | RD7  | 330       | BMEI0289  | 23.76   | Asp-tRNAAsn/Glu-tRNAGln aminotransferase A subunit and related amidases |
|            |            | RD8  | 719       | BMEI0527  | 46.63   | Exonuclease VII large subunit |
|            |            | RD9  | 335       | BMEI0566  | 8.29    | ABC-type Fe3+ transport system, permease component |
|            |            | RD10 | 327       | BMEI1039  | 19.26   | 16S RNA uridine-516 pseudouridylate synthase and related pseudouridylate synthases |
|            | ≤−2        | RD11 | 321       | BMEI0124  | 23.46   | NAD-dependent aldehyde dehydrogenases |
|            |            | RD11 | 321       | BMEI0125  | 21.31   | NAD-dependent aldehyde dehydrogenases |

*CDM: Coverage Divided by Mean coverage.*

*RD1a: included in RD1.*

[doi:10.1371/journal.pone.0070852.t003]
Unique genome regions. Chromosome I of M5 contained two regions with high read coverage, suggesting duplication. These regions were identified as RI1 and RI2. RI1 contained the ORF BMEI0200, which encodes a transposase insertion sequence in the transposon Tn4003. This transposon was also found in 16 M, where another copy was also found. RI2 contained two ORFs, BMEI0901 and BMEI0902. BMEI0901 encodes a resolvase similar to phage Mu DNA invertase Pin homologs, whereas BMEI0902 encodes a site-specific recombinase similar to an anti-repressor (ant) gene. This gene encodes a phage-related DNA binding protein and is homologous to BMEI1661 in RD5. Comparative genome analysis revealed that most available Brucella genomes contained only one copy of BMEI0901/BMEI0902. However, B. canis ATCC 23365, B. suis 1330, B. melitensis ATCC 23457, B. abortus bv.1 str.9−941, B. abortus S19, and B. Abortus 2308 as well as the M5 strain, have two copies of these ORFs in each genome. Unlike other genomes, we did not find BMEI0901/BMEI0902 homologues in the genome of B. ovis ATCC 25840. Thus, this site-specific recombinase widely exists in almost every Brucella genome, suggesting that it originates from a Brucella ancestor. Interestingly, the recombinase appears to the same family to RD5, suggesting that it complements the absent genes in that region. Similar to RI2, RD5 exhibits different patterns of these genes in different genomes (Table 4).

Genome SNPs. Beyond the 11 RDs, M5 and 16 M exhibit considerable general genetic homology. Pairwise analysis demonstrated that there were 2819 SNPs between the genomes of 16 M and M5. Among these SNPs, 585 were in the intergenic regions and 2234 were in the genetic regions. Although the genome divergence time can be calculated by non-synonymous intra-species mutations, the splice time of these two genomes is difficult to estimate because the genome is under manual selection pressure. The 2234 SNPs in the genetic regions included 798 synonymous SNPs and 1436 non-synonymous SNPs (Table S3). The non-synonymous SNPs only resulted in amino acid changes, with no frame shift or nonsense mutations.

Delrue et al. presented a list of 196 unique virulence-related genes divided into 12 functional classes [19]. Our comparative analysis of these two genomes suggests that SNPs widely exist in these genes. A total of 183 SNPs were present in 197 genes, among which 101 SNPs were non-synonymous mutations. Compared with SNPs occurring at other sites of the genome, these genes presented higher positive selection pressure, with Ka/Ks of 1.23. Detailed function analysis of these virulence-related genes revealed that most functional categories were involved in fast evolution (Figure 2 and Table 5). Within all known functional groups, both total SNP rates and non-synonymous SNP rates were higher in the genes of "metal acquisition" and "nitrogen metabolism," and both Ka/Ks values were greater than 1. The Ka/Ks values were also large in classical virulence factors, including genes of envelope molecules and secretion or transport system. These data suggest that mutations in these genes can play important roles in virulence attenuation of the vaccine strain. However, the SNP rates and Ka/Ks were also high in the groups of unknown function and other genes, suggesting the need for further investigation into the functions of these genes.

Virulence-associated metabolism and virulence factors. Comprehensive analysis of the 16 M and M5 genomes revealed variations in genome deletion, insertion, and SNPs. Given these genomic differences, we next sought to identify...
candidate ORFs associated with virulence in 16 M or lack of virulence in M5. To improve the accuracy of our analysis, we first performed an independent literature search to identify published differences between virulence factors and important metabolic pathways. Following this, large-scale screens and testing in the model systems were performed in Brucella to confirm and/or identify virulence factors associated with metabolism and virulence. The mutations of each virulence-associated gene between M5 and 16 M are listed in Table S4.

Envelope molecule. The LPS O-polysaccharide (O-PS) is the first molecule clearly involved in Brucella intracellular entry [20,21]. Genetic studies demonstrated that the O-PS is synthesized by the so-called ABC transporter-dependent pathway, which characterizes homopolymeric polysaccharides. In Brucella, the genes involved in the synthesis of the O-PS and its translocation to the periplasm are clustered in the Wbk region of chromosome I (ORF BMEI1393 to BMEI1427) [22]. In our analysis of M5, no non-synonymous mutations were found in the core genes of this region (wbpZ, wbkA, rfbD, and others); however, several mutations were present in the ORFs BMEI1399 [G238C(A80P)] and BMEI400 [G617T(P21S)], which are transposases in the cluster. In addition, other mutations involved in the synthesis of LPS were identified and included pmm [T397C(F133L)], pgm [A752G(H251R)], wbdA [T347C(L116S), G1103C(R368P)], lpsA [C278T(S93L)], mgtA [C142T(P48S), G380A(R127K)], and amiC [A545T(N182I)]. These data suggest that the reduced virulence of M5 is due, in part, to defects in O-PS synthesis.

Secretion or transport system. In addition to envelope molecules, the type IV secretion system (T4SS) of Brucella encoded by the VirB operon is a major virulence factor [23]. This operon is composed of 12 ORFs clustered in chromosome II (BMEII0025 to BMEII0036). We identified three ORFs associated with the M5 T4SS that contained single or two sites of non-synonymous mutations. These included: A463C (K155Q) in inner membrane ATPases (virB4, BMEII0028), A622G (T208A) and T1007A (F336Y) in the channel protein (virB6, BMEII0030), and C700G (R234G) in the channel protein (virB9, BMEII0033). The virB mutants cannot reach the replicative niche and reside in a

### Table 4. Site-specific recombinase genes on Brucella strains.

| Accession Number | Strain                         | Biovar  | Site specific recombinase genes |
|------------------|-------------------------------|---------|---------------------------------|
| AONT00000000.1   | Brucella melitensis M5        | biovar1 | +                               |
| NC_003317        | Brucella melitensis 16M       | biovar1 | +                               |
| NC_013119        | Brucella microti CCM 4915     |         | +                               |
| NC_010103        | Brucella canis ATCC 23365     |         | +                               |
| NC_010169        | Brucella suis ATCC 23445      | biovar2 | +                               |
| NC_004310        | Brucella suis 1330            | biovar1 | +,2                             |
| NC_009505        | Brucella ovis ATCC 25840      |         |                                 |
| NC_012441        | Brucella melitensis ATCC 23457| biovar2 | +,2                             |
| NC_006932        | Brucella abortus 9–941        | biovar1 | +,2                             |
| NC_010742        | Brucella abortus S19          | biovar1 | +,2                             |
| NC_007618        | Brucella abortus 2308         | biovar1 | +,2                             |

Figure 2. Synonymous and non-synonymous SNPs rates among the virulence-associated genes. The x-axis represents the functional groups of virulence-associated genes, and the y-axis represents the SNPs in 1 kbs. The blue and red bars represent synonymous and nonsynonymous SNPs, respectively.

doi:10.1371/journal.pone.0070852.g002
membrane-bound vacuole [24]. In addition to mutations in the T1SS mutations C613G(P205A) in the ABC transporter (BMEI1092), C180T(A63V) and C659T(T220I) in the flagellar gene BMEI0154, and A247G(H143V) in the flagellar gene BMEI0159 were also identified. Recent studies showed that some proteins involved in the export of flagellar components are similar to the components of the type III secretion machinery, and flagellar apparatus itself can secrete virulence factors [23]. These data suggest that defects in virulence factor secretion may affect the replicative ability of M5.

Regulation. Brucella must survive under varying conditions ranging from the open environment to the intracellular medium. Thus, the bacterium must coordinate an intricate network of factors to generate a suitable adaptive response to various signals. The adaptive response is mediated by two-component regulatory systems (TCSs) and transcriptional regulators (TRs) [26]. Moreover, BvrS/R is the best characterized BvrS/R is the best characterized Brucella TCS [27]. In addition, BvrS/R regulates several critical cellular functions, including cell division, chemotaxis, metabolism, and expression of toxins and other proteins important for pathogenesis. TRs important for Brucella survival are the GntR and LysR families. In addition, LysR-type transcriptional regulators (LITTRs) are widely distributed in the diverse genera of prokaryotes [28]. Many virulence factors of Brucella are regulated by LITTRs [29,30]. Although no mutations were observed in the common TCSs or TRs, we did identify a non-synonymous mutation of T176C(L59P) and a 3 bp deletion from 391 to 393 at the M5 locus encoding LTTR (BMEII0159). Recent studies showed that some proteins involved in the export of flagellar components are similar to the components of the type III secretion machinery, and flagellar apparatus itself can secrete virulence factors [23]. These data suggest that defects in virulence factor secretion may affect the replicative ability of M5.

Table 5. Distribution of SNPs in virulence-related genes.

| Functional group             | Gene                  | Non-synonymous SNP | Synonymous SNP | All SNP | Ka/Ks |
|------------------------------|-----------------------|--------------------|---------------|--------|------|
| Envelope molecules           |                       |                    |               |        |      |
| Secretion or transport system|                       |                    |               |        |      |
| Regulation                   |                       |                    |               |        |      |
| Metal acquisition            |                       |                    |               |        |      |
| Amino acid metabolism        |                       |                    |               |        |      |
| Sugar metabolism             |                       |                    |               |        |      |
| DNA/RNA metabolism           |                       |                    |               |        |      |
| Vitamins/cofactors           |                       |                    |               |        |      |
| Stress proteins/Chaperones   |                       |                    |               |        |      |
| Oxidoreduction               |                       |                    |               |        |      |
| Nitrogen metabolism          |                       |                    |               |        |      |
| Other genes                  |                       |                    |               |        |      |
| Unknown function             |                       |                    |               |        |      |
| Total                        | 197                   | 197.00             | 101.00        | 0.51   | 0.42 |

doi:10.1371/journal.pone.0070852.t005

Amino acid metabolism. Among the 19 synthesis subclass genes of these two strains, 10 genes of M5 contained point mutations. These data suggest that M5 may have some amino acid metabolism defects; yet failure of M5 to resist a sudden lack of nutrients requires further investigation.

Sugar metabolism. Maltose, ribose, arabinose, galactose, glucose, glyceral, erythritol, and inositol degradation pathways are essential for the intracellular survival of Brucella [37]. For example, mutations of genes encoding sugar transport systems, such as pyruvate carboxylase (BMEI0266), ribokinase (BMEI0089), 6-phosphogluconate dehydrogenase (BMEI1124), inosamine methylase (BMEI1301), glycosyl transferase (BMEI1101), ribose transport (BMEI0300), and glucose/galactose transport (BMEI1053) limit carbon or energy supply and affect gene regulation. In addition, the production of cyclic β-1,2 glucan (BMEI1837) is necessary for the intracellular replication of Brucella [38]. Upon examination of these pathways, SNP mutations in the ORF BMEI1837 of M5 caused five residue mutations at the positions 3587 to 3594, 3631 to 3638, and 3774 to 3782, and another non-synonymous mutation at position 4996. These data suggest that M5 may have defects in cyclic β-1,2 production.
DNA/RNA. Purine/pyrimidine synthesis, repair, and regulatory genes are essential for the intracellular replication of Brucella. A study by Crawford RM indicated that deletion of the purE gene from B. melitensis dramatically reduced bacterial virulence and suggested that this gene should be further evaluated as a potential vaccine [39]. Other genes under the same classification, such as xseA, purL, purD, mgsS, bltD, and cobB mutants, appear to be targets for attenuation vaccine strains.

Stress proteins/Chaperones. Genes encoding stress proteins have been an obvious choice for directed mutagenesis for virulence studies and vaccine generation. Within the M5 genome, genes encoding chaperones (BMEI2002 and BMEI1513) showed one non-synonymous mutation, namely, A37C(M13L) and G398A(G133D), respectively.

Oxidoreduction and nitrogen metabolism. DsbA has been shown to be involved in toxin production, adhesion, motility, extracellular enzyme production, and type III secretion system [40]. Few studies have investigated the role of DsbB in pathogenicity [41]. A G120A(M40I) mutation was detected in the DsbB (BMEI384) gene of M5. In addition, published genome analysis indicated that B. melitensis can adapt to either aerobic or anaerobic environments. Among the genes involved in the process, we detected non-synonymous mutations in FdhA, glnA, mfhS, and glnD of M5.

Unknown function. Seven ORFs encoding hypothetical protein were slightly different between the attenuated M5 strain and the virulent strain 16 M. Among these ORFs, BMEI0128 was identified as a virulence-associated ORF in another attenuated vaccine strain S19. Mutations of A31C(M11L), C589T(H197S), and D605A(H202S) were also identified in the vaccine strain M5.

Conclusions
The vaccine strain M5 has been widely used in Asia for decades to vaccinate sheep against brucellosis [2]. While extensive studies have been performed on the phenotype and physiology of this strain, the molecular mechanisms responsible for loss of virulence are unclear. In this study, we determined the whole-genome sequence of M5 and conducted a comprehensive comparative analysis against the whole-genome sequence of the virulent strain 16 M. The whole-genome analysis revealed that the genome sequences of M5 strain and the virulent strain 16 M. Among these ORFs, BMEI0128 was identified as a virulence-associated ORF in another attenuated vaccine strain S19. Mutations of A31C(M11L), C589T(H197S), and D605A(H202S) were also identified in the vaccine strain M5.

RD4 is a typical random repeat region consisting of the sequence: $^*_{GGGGATCAAGA-GATCGCCCCCGTCTCGTCCCGCCCTTCCGTGTCGCC-GAAGGCACAGAAAGAGAAACATGGCCGCGAGCGCAGATGGTAAAGGCGCCATCATTTTTCGTGGTGACG-GATGAC}$, which repeats 1 to 2 times in each genome and can be considered the candidate VNTR loci for MLVA genotyping. RD5 is 883 bp and overlaps with two resolvable family genes in 16 M. The ORFs BMEI1661 and BMEI0902 belong to this family, which are 231 and 747 bp respectively. The proteins encoded by these ORFs have the same core functions but different protein and structure combinations. Comparative genome analysis in Brucella revealed two typical sequence types present in this region. The first sequence was approximately 139 bp and contained a gene similar to BMEI0902. The second was 683 bp and contained a gene similar to BMEI1661. In our study, RI2 belongs to the first sequence. Compared with the 16 M genome, the M5 strain had an additional RI2 region but the RD5 was missing. Further analysis in all Brucella strains presented different patterns of these two regions in the various genomes. This bias of the selected site-specific recombinase has been shown to be involved in the acquisition of drug resistance genes and alteration of gene expression [42], suggesting that this unique gene affects virulence.

Considering the genetic patterns widely present in all strains, we consider the 2 RDs, 1 RI, and the inversion region as candidates for detection of genome differences. Our analysis was also aimed at identifying 1) the genes or genome fragments that only existed in our genome, 2) the mutations specific toward the virulent strain and the attenuated strain, and 3) the molecular mechanisms of M5 strain attenuation. Among the 11 RDs identified, RD1, RD2, and RD7 were in M5, whereas they are consistent with other strains in 16 M. Furthermore, RD3, RD6, RD8, RD9, RD10, and RD11 were highly differentiated regions with a large number of single nucleotide differences. All of these regions are possible candidates for distinguishing M5 from 16 M. Notably, RD9 includes an ORF encoding an ABC-type Fe$^{3+}$ transport system (sfuB, BMEI0566).

In Brucella, SfuA binds free Fe$^{3+}$ in the periplasm, and SfuB and C transport this cation across the cytoplasmic membrane. Several studies suggested that the ferric dicitrate complex itself represents important substrates for the SfuABC type transporters [32,33]. On the basis of the potential relevance of ferric dicitrate as an iron source for Brucella in vitro, the sfuB mutation seemingly is responsible for the loss of virulence in vaccine strain M5. Surveys of the genome sequences of B. melitensis 16 M and B. suis 1330 suggest that transport systems that allow the Brucella to utilize heme and ferric dicitrate as iron sources are important for sustaining the intracellular lifestyle of the Brucella in host macrophages [43].

The whole-genome sequences analyzed in this study clearly revealed large genome differences between 16 M and M5. Many of these differences were in genes associated with Brucella metabolism and survival, and thus serve as potential detection sites for distinguishing between the virulent 16 M and attenuated M5 strains. In addition, our analysis revealed two novel sites for typing Brucella. Furthermore, 3 RDs and SNPs of important virulence-related genes such as mgbB, dhhB, sfuB, dshAB, aidA, avc, and yfiR, were considered to be sites causing virulence loss in the M5 strain. Collectively, this study reveals that comparative analysis between the wild-type and vaccine strains can provide resources for the study of Brucella virulence and Brucella microevolution. Further detailed analysis of the gene function of deletion genes can elucidate their roles in virulence attenuation of B. melitensis.

Materials and Methods
Isolation extraction and identification
B. melitensis M5 was obtained from the National Institute for Communicable Disease Control and Prevention Center. This strain was submitted for classical identification procedures as follows: CO$_2$ requirement, H$_2$S production, and inhibition of growth by basic fuchsin and thionin, agglutination with monospecific antisera, and phage typing. Total genomic DNA was
extracted using the DNaseasy Blood and Tissue Kit (QIAGEN, China Ltd., China). An aliquot of the DNA was subjected for analysis using the Bioanalyzer (Agilent Technologies) and was confirmed for absence of DNA degradation.

**Genome sequencing, assembly, and annotation**

A genome shotgun method was used to collect the genome sequence. We constructed two DNA libraries: 500 bp and 3 kb–5 kb DNA fragments for high-throughput genome sequencing with SOLEXA. Pair-end 75-bp reads were collected. We obtained 518 Mb and 7,000,000 reads covering ~157 folds for the 16 M genome. The draft genome sequence of *B. melitensis* vaccine strain M5 is available in GenBank under accession number AONT00000000.

The SOLEXA data were assembled using the SOAP de novo software. Genome sequences were analyzed using an automatic genome analysis pipeline. Gene prediction was conducted following the method used in the study by Glimmer [44] with default parameters. Gene annotation was performed with COG and InterproScan. Gene pathways were annotated based on Kyoto Encyclopedia of Genes and Genomes.

**Comparative genomic analyses**

High-throughput SOLEXA reads mapping to 16 M was used in the whole-genome comparison because that the draft genome had limits in gene prediction and alignment. Genome differences were valued by each site of SOLEXA coverage, which is standardized by calculating the CDM (read coverage of each site/mean coverage of the whole genome). The log2 value of the CDM (Log2CDM) was used to improve the description.

We compared the sequences of two strains by MUMmer [45] and removed unclear fragments by BLAST. The IS elements were identified and compared based on the IS database (http://www-is.biotoul.fr/). Genes lost or gained were selected by PERL scripts, and their functions were determined. A synteny map between the two strains was constructed by the OrthoMCL pipeline [46]. Whole-genome raw SNPs were detected following the method published by Mummer et al. with default parameters. The detailed variations in virulence-associated genes between M5 and 16 M, including SNPs, deletion, and insertion, were further identified by alignment using CLUSTALW and then calculated by PERL scripts. A Ka/Ks calculator was used [47].

**Supporting Information**

Figure S1 Alignment of RDs in the *B. melitensis* vaccine strain M5 to the virulent strain 16 M. RD1, RD2, RD4, RD5, and RD7 exhibit obvious genome insertion and deletion in these regions. The insertion and deletion caused low read coverage regions and are denoted as RDs. Sequences of other RDs were identical with the virulent strain 16 M. Low coverage in these regions may be caused by uncertainty in sequencing. (PDF)

Table S1 Genome assembly of the *B. melitensis* vaccine strain M5. (PDF)

Table S2 COG-based function of ORFs identified in the *B. melitensis* vaccine strain M5. (PDF)

Table S3 SNPs detected in the *B. melitensis* vaccine strain M5 compared with the virulent strain 16 M. (PDF)

Table S4 Distribution of SNPs on virulence-related genes. (PDF)

Table S5 RDs aligned to reference sequences. (PDF)

**Acknowledgments**

The authors thank Dr David O’Callaghan for helpful comments on the analysis.

**Author Contributions**

Conceived and designed the experiments: HJ BC. Performed the experiments: HW GT. Analyzed the data: PD WZ. Contributed reagents/materials/analysis tools: HZ DP. Wrote the paper: HJ PD CC.

**References**

1. Food and Agriculture Organization of the United Nations, World Organization for Animal Health, and World Health Organization (2006) Brucellosis in human and animals. Geneva: World Health Organization; WHO/CDS/EP/2006.7.
2. Shang DQ, Xiao DL, Yin JM (2002) Epidemiology and control of brucellosis in China. Vet Microbiol 90:163–182.
3. Cui BY (2007) Endemic surveillance and control of Brucellosis in China. Dis Surf 22:649–651.
4. Zhang WY, Guo WD, Sun SH, Jiang JF, Sun HL, et al. (2010) Human brucellosis, Inner Mongolia, China. Emerg Infect Dis 16:2001–2003.
5. Coutes SJ, Whanmore AM, Commander NJ (2003) Brucellosis—new aspects of an old disease. J Appl Microbiol 98:1270–1281.
6. Graves RR (1943) The Story of John M. Buck’s and Matilda’s Contribution to old disease. J Appl Microbiol 98:1270–1281.
7. Nicosleti P (1990) Vaccination. In: Nelson K, Duncan JR, eds. Animal Brucellosis. Boca Raton: CRC Press. 204–299.
8. Blasco JM(1997) A review of the use of *B. melitensis* Rev. 1 vaccine in adult sheep and goats. Prev Vet Med 31:273–283.
9. Garm-Bastagi B, Blasco JM, Grayson M, Verger JM (1990) *Brucella melitensis* infection in sheep: present and future. Vet Rec 92:255–274.
10. Cloeckaert A, Grayson M, Grépinet OI (2002) Identification of *Brucella melitensis* vaccine strain Rev. 1 by PCR-RFLP based on a mutation in the *spI* gene. Vaccine 20:2546–2550.
11. Sangari FJ, Aggerov J (1994) Identification of *Brucella abortus* B19 vaccine strain by the detection of DNA polymorphism at the ery locus. Vaccine 12:435–438.
12. Sangari FJ, Garcia-Lobo JM, Aggerov J (1994) The *Brucella abortus* vaccine strain B19 carries a deletion in the erythritol catabolic genes. FEBS Microbiol Lett 121:337–342.
13. Vemulapalli R, McQuiston JR, Schurig GG, Siriranganathan N, Halling SM, et al. (1999) Identification of an IS711 element interrupting the *spaI* gene of *Brucella abortus* vaccine strain RB51 and a PCR assay to distinguish strain RB51 from other *Brucella* species and strains. Clin Diagn Lab Immunol 6:760–764.
14. Crasta OR, Folkerts O, Fei Z, Mane SP, Evans C, et al. (2008) Genome Sequence of *Brucella abortus* Vaccine Strain S19 Compared to Virulent Strains Yields Candidate Virulence Genes. PLoS One 3:e2195.
15. Ding JB, Pan YL, Jiang H, Cheng JS, Liu TT, et al. (2011) Whole Genome Sequences of Four *Brucella* Strains. J Bacteriol 194:3674–3675.
16. Wang FK, Hu S, Gao YZ, Qiao ZJ, Liu WQ, et al. (2011) Complete Genome Sequences of *Brucella melitensis* Strains M28 and M3-90, with Different Virulence Backgrounds. J Bacteriol 193:2904–2905.
17. Yang YL, Yin JG, Wang XL, Wang XR, Wang FX, et al. (2012) Immunoproteomic Analysis of Chinese *Brucella* Vaccine Strain M5 and New Immunogenic Proteins. J Data Mining Genomics Proteomics 3:2.
18. Martin-Martín AL, Sanchez P, de Miguel MJ, Fernandez-Lago L, Vizcaíno N(2012) Quorum-Sensing and BvrR/BvrS Regulation, the Type IV Secretion System, Cyclic Glucans, and BacA in the Virulence of *Brucella suis*. Similarities to and Differences from Smooth Brucellae. Infect Immun 80:1783–1793.
19. Delrie RM, Lestrade P, Tiber A, Letesson JJ, De Bolle X (2004) *Brucella* pathogenesis, genes identified from random large-scale screens. FEBS Microbiol Lett 251:1–12.
20. Fernandez-Prada CM, Zebanowiska EB, Nikolich M, Hadfield TL, Roop RM 2nd, et al. (2003) Interactions between Brucella melitensis and human phagocytes: bacterial surface O-polysaccharide inhibits phagocytosis, bacterial killing, and subsequent host cell apoptosis. Infect Immun 71:2110–2119.

21. Porte F, Naroeni A, Ouahrani-Bettache S, Lautard JP (2003) Role of the Brucella suis lipopolysaccharide O antigen in phagosomal genesis and in inhibition of phagosome-lysosome fusion in murine macrophages. Infect Immun 71:1481–1490.

22. Zygmunst MS, Blasco JM, Letesson JJ, Cloeckaert A, Moriyo ´n I (2009) DNA

23. Boschiroli ML, Ouahrani-Bettache S, Foulongne V, Michaux-Charachon S, Brucella

24. Comerci DJ, Martı ´nez-Lorenzo MJ, Sieira R, Gorvel JP, Ugalde RA (2001) Essential role of the VirB machinery in the maturation of the Brucella abortus-containing vacuole. Cell Microbiol 3:159–168.

25. Blocker A, Komoriya K, Aizawa S (2003) Type III secretion systems and bacterial flagella: insights into their function from structural similarities. Proc Natl Acad Sci USA 100: 3027–3030.

26. Kohler S, Foulongne V, Ouahrani-Bettache S, Bourg G, Teyssier J, et al. (2002) The Brucella suis virB operon is induced intracellularly in macrophages. Proc Natl Acad Sci U S A99:1544–1549.

27. Viadas C, Rodrı´guez MC, Sangari FJ, Gorvel JP, Garcı´a-Lobo JM, et al. (2010) The analysis of the intramacrophagic virulome of Brucella suis reveals marked differences in O-polysaccharide biosynthetic genes between smooth and rough Brucella species and novel species-specific markers. BMC Microbiol 9:92.

28. Schell MA (1993) Molecular biology of the LysR family of transcriptional regulators. Annu Rev Microbiol 37:597–626.

29. Foulongne V, Bourg G, Cazevieille C, Michaux-Charachon S, O’Callaghan D (2005) Characterization of new DsbB-like thiol-oxidoreductases of Brucella melitensis and classification of the DsbB family based on phylogenomic, structural and functional criteria. Microbiology 151:219–231.

30. Escriva L, Canavessi A, Carey M, Spliter G (2006) Brucella abortus and Brucella suis siderophore 2,3-dihydroxybenzoic acid (DHBA) facilitates intracellular survival of the bacteria. Microb Pathog 32:239–248.

31. Tisd R, Baumber A, Heffron F, Stojilkovic I (1996) Contribution of TonBand Feo-mediated iron uptake to growth of Salmonella typhimurium in the mouse. Infect Immun 64: 4549–4556.

32. Reeves SA, Torres AG, Payne SM (2000) TonB is required for intracellular growth and virulence of Shigella dysenteriae. Infect Immun 68: 6329–6336.

33. Parent MA, Bellaire BH, Murphy EA, Roop RM 2nd, Elzer PH, et al. (2002) Brucella abortus siderophore 2,3-dihydroxybenzoic acid (DHBA) facilitates intracellular survival of the bacteria. Microb Pathog 32:239–248.

34. Tisd R, Baumber A, Heffron F, Stojilkovic I (1996) Contribution of TonBand Feo-mediated iron uptake to growth of Salmonella typhimurium in the mouse. Infect Immun 64: 4549–4556.

35. Reeves SA, Torres AG, Payne SM (2000) TonB is required for intracellular growth and virulence of Shigella dysenteriae. Infect Immun 68: 6329–6336.

36. Danese I, Hane V, Delrue RM, Tilbor A, Lestrade F, et al. (2004) The Ton system, an ABC transporter, and a universally conserved GTPase are involved in iron utilization by Brucella melitensis 16 M. Infect Immun 72: 5783–5790.

37. Eisenberg RC, Seshadri R, Nelson K, Paulsen I (2002) Sugar metabolism by Brucella. Vet Microbiol 90: 249–261.

38. Briones G, Ión de Iñamino N, Roset M, Viguocco A, Paolo PS, et al. (2001) Brucella abortus cyclic beta-1,2-glucan mutants have reduced virulence in mice and are defective in intracellular replication in HeLa cells. Infect Immun 69: 4528–4535.

39. Crawford RM, Van De Verg L, Yuan I, Hadfield TL, Warren RL, et al. (1996) Deletion of pure attenuates Brucella melitensis infection in mice. Infect Immun 64:2188–2192.

40. Jiang BL, Liu J, Chen LF, Ge YY, Hang XH, et al. (2008) DsbB Is Required for the Pathogenesis Process of Xanthomonas campestris pv. Campestris Mel Plant Microbe Interact 21:1036–1045.

41. Raczkowski AM, Bujiicki JM, Pawloski M, Godlewkska R, Lewandowska M, et al. (2005) Characterization of new DsbB-like thiol-oxidoreductases of Campylobacter jejuni and Helicobacter pylori and classification of the DsbB family based on phylogenomic, structural and functional criteria. Microbiology 151:219–231.

42. Escriva L, Canavessi A, Carey M, Spliter G (2001) Brucella abortus genes identified following constitutive growth and macrophage infection. Infect Immun 69:7736–7742.

43. Braun V, Hantske K, Koster W (1996) Bacterial iron transport: mechanisms, genotypes, and regulation. Mol Biol Evol 13:67–145.

44. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) Improved microbial gene identification with GLIMMER. Nucleic Acids Res 27: 4636–4641.

45. Delcher AL, Phillippy A, Carlson J, Salzberg SL (2002) Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res 30: 2478–2483.

46. Li L, Stoeckert CJ Jr, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13: 2178–2189.

47. Zhang Z, Li J, Zhao X, Wang R, et al. (2006) KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics 4: 259–263.