Comparative genomic analysis between newly sequenced *Brucella suis* Vaccine Strain S2 and the Virulent *Brucella suis* Strain 1330

Dong-dong Di†, Hai Jiang†, Li-li Tian, Jing-li Kang, Wen Zhang, Xin-ping Yi, Feng Ye, Qi Zhong, Bo Ni, You-yu He, Lin Xia, Yao Yu, Bu-yun Cui, Xiang Mao, and Wei-xing Fan*

**Abstract**

**Background:** Brucellosis is a bacterial disease caused by *Brucella* infection. In the late fifties, *Brucella suis* vaccine strain S2 with reduced virulence was obtained by serial transfer of a virulent *B. suis* biovar 1 strain in China. It has been widely used for vaccination in China since 1971. Until now, the mechanisms underlie virulence attenuation of S2 are still unknown.

**Results:** In this paper, the whole genome sequencing of S2 was carried out by Illumina Hiseq2000 sequencing method. We further performed the comparative genomic analysis to find out the differences between S2 and the virulent *Brucella suis* strain 1330. We found premature stops in outer membrane autotransporter omaA and eryD genes. Single mutations were found in phosphatidylcholine synthase, phosphorglucosamine mutase, pyruvate kinase and FliF, which have been reported to be related to the virulence of *Brucella* or other bacteria. Of the other different proteins between S2 and 1330, such as Omp2b, periplasmic sugar-binding protein, and oligopeptide ABC transporter, no definitive implications related to bacterial virulence were found, which await further investigation.

**Conclusions:** The data presented here provided the rational basis for designing *Brucella* vaccines that could be used in other strains.

**Keywords:** *Brucella suis* vaccine strain S2, *Brucella suis* virulent strain 1330, Genome sequence, Comparative genomic analysis

**Background**

Brucellosis is a worldwide zoonotic disease caused by the infection of *Brucella* species, a Gram-negative and facultative intracellular bacterium. Brucellosis can be acquired by direct contact with infected animals including cattle, sheep, goats and swine. *Brucella* infects approximately 500,000 humans worldwide annually according to the World Health Organization (WHO) [1]. The infection can cause reproductive failure in food animals and even the loss of human productivity. The clinical symptoms of *Brucella* infection include undulant fever, abortion, asthenia, endocarditis, and encephalitis. Brucellosis in food animals could be controlled by vaccination, while human brucellosis can be treated with antibiotics.

Currently, there are ten recognized *Brucella* species based on their host preference and phenotypic differences: *B. melitensis* (goats), *B. abortus* (cattle), *B. suis* (swine), *B. canis* (dogs), *B. ovis* (sheep), *B. neotomae* (desert mice), *B. ceti* (cetaceans), *B. pinnipedialis* (pinnipeds), *B. microti* (common voles) and *B. inopinata* (unknown) [2]. The genome of *B. melitensis* strain 16M is the first one to
be sequenced [3], followed by *B. suis* strain 1330 [4, 5] and *B. abortus* strain 2308 and 9–941 [6]. New *Brucella* genome sequences are increasing, and at least 39 *Brucella* strains have been sequenced. These genomic resources can be served for identifying *Brucella* gene variability that associated with their host preference, pathogenesis, and virulence.

*B. suis* strain 1330 (hereafter referred to as 1330) is a swine isolate and used as the standard reference strain for *B. suis* biovar 1 [7]. According to phylogenetic analysis of 10 *Brucella* strains, *B. suis* has a broader host specificity without any identified species-specific markers [8].

The *B. suis* strain S2 is isolated from swine fetus by the Institute of Inspection for Veterinary Medicine in 1952 in China. It possesses all of the characteristics of *B. suis* biovar 1 and a smooth colonial morphology [9]. After serial transfer on media for years, *B. suis* strain S2 is naturally attenuated to a vaccine strain designated as *B. suis* vaccine “strain 2” or S2 (hereafter referred to as S2). The avirulent property of S2 is stable when inoculated into susceptible animals, and it does not cause the abortion of pregnant female [9]. Therefore, it has been widely used in China since 1971 to vaccinate sheep and goats [9]. The vaccine is administered through drinking water, a dose of 10 billion bacteria provides the protection for 2–3 years. In the middle of 1980s, S2 is introduced to the other countries [10]. Although studies suggest that the protection rate of S2 is less than that of *B. abortus* “strain 19” or S19 (hereafter referred to as S19) and *B. melitensis* Rev. 1 (hereafter referred to as Rev. 1) [11, 12], a large-scale animal test shows that S2 can provide a satisfactory protection rate in the field [9, 13]. Based on these encouraging results, the low virulence compared with S19 and Rev.1, the low cost of production, the safety for pregnant sheep and the feasibility of administration by the oral route in cattle, pigs, sheep and goats, S2 is used successfully in China to vaccinate all target species against brucellosis by administration through drinking water [9–12]. In mice, S2 has low residual virulence and less persistence time compared with S19 and Rev.1. It provides short-time protection against virulence *B. melitensis* challenge compared with the other two vaccines [14]. However, it does not show significant protection against *B. melitensis* 53H38 challenge when tested in lambs [11].

The underlying molecular and physiological mechanisms of the low virulence of S2 are not well understood so far. In this study, we performed the paired-end sequencing combined with Sanger sequencing method to determine the complete genome sequences of S2. We further identified the different genes between S2 and 1330 by genome comparison. Several putative factors were identified as the candidate genes that might be related to the virulence attenuation of S2. Further investigation of these target genes will be important for designing better vaccines to provide efficient control for brucellosis.

### Results and discussion

#### Genome sequence properties

The S2 genome was comprised of two circular chromosomes (Table 1). One chromosome was 2,107,842 bp and the other was 1,207,433 bp long in length. The average GC content of the two chromosomes was 57 %. Unsurprisingly, The S2 genomes exhibited significant similarities in size and structure compared to the genomic sequences of 1330. The S2 genome sequences showed over 99.9 % similarities with the genomes of 1330. The diagram of S2 genomes is shown in Fig. 1. Circos was employed to construct the diagram [15]. A total of 2119 and 1139 open reading frames (ORFs) were identified on chromosome I and II, respectively. Sequences of these two chromosomes can be accessed from GenBank (CP006961 and CP006962). More than 99.5 % of ORFs in S2 were identical to those found in 1330 (CP002997.1 and CP002998.1). S2 exhibited highly genome-wide collinearity with 1330 (Fig. 2 and Additional file 1: Table S1). More than 38 % of the predicted ORFs in S2 had rpsblast hits to the cluster of orthologous groups (COG) database with an e-value less than 1e-2. A total of 1147 ORFs (35.2 %) were hypothetical proteins (Table 1).

Genome-wide comparison between S2 and 1330 genomes identified all the single nucleotide polymorphisms (SNPs), a total of 72 SNPs were identified (Table 2). The SNPs were not equally distributed between the two chromosomes, 42 SNPs (58 %) were located on chromosome I, while the remaining 30 SNPs (42 %) were located on chromosome II. The exact positions of all of the SNPs at the nucleotide level and protein level (if the SNP is within an ORF) are shown in Additional file 2: Table S2. Twenty-nine SNPs (42 %) were located on chromosome II. The exact positions of all of the SNPs at the nucleotide level and protein level (if the SNP is within an ORF) are shown in Additional file 2: Table S2. Twenty-nine SNPs (42 %) were located on chromosome II. The exact positions of all of the SNPs at the nucleotide level and protein level (if the SNP is within an ORF) are shown in Additional file 2: Table S2. Twenty-nine SNPs (42 %) were located on chromosome II. The exact positions of all of the SNPs at the nucleotide level and protein level (if the SNP is within an ORF) are shown in Additional file 2: Table S2. Twenty-nine SNPs (42 %) were located on chromosome II. The exact positions of all of the SNPs at the nucleotide level and protein level (if the SNP is within an ORF) are shown in Additional file 2: Table S2. Twenty-nine SNPs (42 %) were located on chromosome II. The exact positions of all of the SNPs at the nucleotide level and protein level (if the SNP is within an ORF) are shown in Additional file 2: Table S2.
non-synonymous substitutions; one SNP caused either frameshifting or premature stop in protein coding region.

Putative protein-encoding genes, tRNA and rRNA in S2 genome were predicted with RAST (http://rast.nmpdr.org). The predicted protein sequences were compared with non-redundant protein database. The protein sequences were also compared with COG and CDD database, the threshold is E < 1e-2.

Identification of virulence associated differences between the attenuated strain S2 and virulent strain 1330

The pairwise comparison was carried out to identify the non-identical ORFs between S2 and 1330 (Table 3), which provided useful information on the underlying mechanisms for the attenuation of S2.

Most of the different ORFs between 1330 and S2 in chromosome I and II had only 1 bp difference in nucleotide...
sequence. Six ORFs had 2, 6, 8, 16, 69 and 78 bp insertion or deletion in the genes respectively (Table 3). The insertion or deletion in the gene caused either frameshifting (including premature stop) or residue insertion in the coding protein. The different ORFs in S2 that encoded the same protein sequences in other Brucella strains were labeled ‘found’ and omitted in discussions. The different ORFs labeled with ‘surface’ suggested that the different residue was located on the surface of the modeled structure, which were also omitted from the analysis (Table 3).

Known virulence associated differences

In chromosome I of S2, a 2 bp deletion caused the premature stop of an outer membrane autotransporter OmaA. The autotransporters are secreted proteins that can translocate themselves through the membranes to the cell surface or to the extracellular environment. Bandara et al. studied the function of this OmaA [16]. In mice, the OmaA-deficient strain provides greater protection against the challenge of 1330 than that of B. suis mutant strain VTRS1. The deficient strain can induce a significantly higher level of serum IgG1 and IgG2a antibodies and can be cleared quickly from the mice. Although the OmaA activity is not necessary in the course of the acute phase of infection of B. suis, it might still be a critical virulent factor during the chronic phase of the infection. Therefore, we suspected the lack of intact OmaA protein might be the major cause for S2 virulence attenuation.

The second difference between 1330 and S2 was found in the erythritol (ery) operon. The erythritol operon contains 4 ORFs (eryA, eryB, eryC and eryD) [17]. Erythritol metabolism by Brucella has been known to be associated with the abortion in livestock. The high concentration of erythritol in the foetal tissue of the animal (cattle, sheep, goat and pig) might be advantageous for the growth of Brucella. On the other hand, erythritol catabolism pathway might play the important physiological roles in bacteria growth. It has been suggested that erythrose 4-phosphate, the precursor for the biosynthesis of aromatic compounds, may be easily obtained from the intermediates of erythritol catabolism. Compared to 1330, a 16 bp insertion caused the premature stop of the eryD gene in S2. The studies on B. abortus S19 and S2308 showed that eryD might act as a repressor of erythritol operon [17]. The consequence of eryC and eryD ORFs deletion in S19 has been studied, which reveals that they are not sufficient or required for virulence in a murine model. No direct correlation between erythritol metabolism and in vivo colonization was found when S19 and 2308 strains were compared with three genetically engineered strains related to ery operon [18]. Considering its repressive control on ery operon, the lack of eryD might cause the uncontrolled erythritol catabolism. Nevertheless, its consequence on virulence attenuation needs to be further explored.

It has been reported that S19 does not grow on TSA (Ery). To test the growth capacity of S2 on erythritol, we chose S19 as the control. We found that S2 was able to grow on erythritol agar, suggesting it has the ability to metabolize erythritol (Table 4). We compared the genomes of 1330, S2 and S19 around the ery operon. S2 and 1330 genomes have the same eryA, eryB and eryC genes, while eryD in S2 genome encodes a premature protein. Meanwhile, a 68 bp deletion in the region of eryC and eryD results in an ORF encoding a putative sugar binding protein in S19. The other genes upstream and downstream of ery operon in these three genomes are highly conserved, except for some mutations in the encoded proteins (Fig. 3). Based on the analysis, the inviability of S19 on erythritol agar might be due to the lack of eryC and eryD genes in the genome. A recent paper describes a new pathway on erythritol metabolism. Except for the known eryA, eryB, eryC and eryD genes, two critical genes downstream of the eryD gene, TpiA2 and RpiB, in erythritol metabolism are demonstrated. The new model strongly suggests that erythritol metabolism needs three isomerization reactions sequentially catalyzed by EryC, TpiA2, and RpiB to produce D-erythrose-4-P, which will be converted into glyceraldehyde-3-P and fructose-6-P in the pentose phosphate pathway [19]. This new model clarifies why S19 could not grow on erythritol agar since functional eryC is lost in the genome.

Phosphatidylcholine synthase gene (PCS) in chromosome 2 of S2 strain had one different residue (S32L) compared with PCS protein in 1330. In bacteria, phosphatidylcholine can be synthesized through either the phospholipid N-methylation pathway or the phosphatidylcholine synthase pathway [20, 21]. The correlation

| Chromosome I | S2 nt | A | C | G | T | Total |
|--------------|-------|---|---|---|---|-------|
| 1330 nt      | -     | 0 | 0 | 1 | 1 | 2     |
| A            | 1     | 3 | 12| 1 | 17|       |
| C            | 1     | 2 | 0 | 9 | 12|       |
| G            | 0     | 5 | 0 | 0 | 5 |       |
| T            | 0     | 1 | 5 | 0 | 6 |       |
| Chromosome II|       |   |   |   |   |       |
| S2 nt        | -     | 0 | 1 | 1 | 0 | 2     |
| A            | 0     | 1 | 3 | 0 | 4 |       |
| C            | 2     | 1 | 1 | 3 | 7 |       |
| G            | 2     | 3 | 1 | 3 | 9 |       |
| T            | 0     | 1 | 7 | 0 | 8 |       |
| Total        | 7     | 12| 17| 18| 17| 72    |

**Table 2** Single nucleotide polymorphism comparison between B. suis strain S2 and 1330.
| slno | locus  | S2 vs. 1330 | Residue change | annotation |
|------|--------|-------------|----------------|------------|
| 1    | BS1330_I1369  | 69<sup>a</sup> | H → R | hypothetical protein (found) |
| 2    | BS1330_I0234  | 8<sup>a</sup> | F → S | icIR family transcriptional regulator (paper) |
| 3    | BS1330_I0875  | 6<sup>a</sup> | V → A | glycosyl hydrolase family protein (found) |
| 4    | BS1330_I2007  | 2<sup>a</sup> | V → A | outer membrane autotransporter (paper) |
| 5    | BS1330_I0045  | 1<sup>b</sup> | H → R | qoxD (found) |
| 6    | BS1330_I0093  | 1<sup>b</sup> | V → A | acnA (found) |
| 7    | BS1330_I0235  | 1<sup>b</sup> | S → Y | sugar ABC transporter, periplasmic sugar-binding protein (paper) |
| 8    | BS1330_I0501  | 1<sup>b</sup> | Q → R | pppK (surface) |
| 9    | BS1330_I0635  | 1<sup>b</sup> | D → A | omp2b (paper) |
| 10   | BS1330_I0045  | 1<sup>b</sup> | R → H | omp2b (paper) |
| 11   | BS1330_I0712  | 1<sup>b</sup> | Y → H | hypothetical protein (found) |
| 12   | BS1330_I0775  | 1<sup>b</sup> | K → T | hypothetical protein (found) |
| 13   | BS1330_I0803  | 1<sup>b</sup> | V → A | nufO (paper) |
| 14   | BS1330_I0912  | 1<sup>b</sup> | R → Q | penicillin-binding protein 1A (unknown) |
| 15   | BS1330_I1054  | 1<sup>b</sup> | H → R | acetyl-CoA hydrolase/transferase family protein (surface) |
| 16   | BS1330_I1203  | 1<sup>b</sup> | S → G | rplR (found) |
| 17   | BS1330_I1213  | 1<sup>b</sup> | P → S | serine protease (unknown) |
| 18   | BS1330_I1329  | 1<sup>b</sup> | H → R | dedA family protein (paper) |
| 19   | BS1330_I1458  | 1<sup>b</sup> | H → R | pncA (found) |
| 20   | BS1330_I1642  | 1<sup>b</sup> | A → V | glcB (surface) |
| 21   | BS1330_I1655  | 1<sup>b</sup> | V → A | hypothetical protein (found) |
| 22   | BS1330_I1684  | 1<sup>b</sup> | L → R | glmM (no direct interaction with substrate) |
| 23   | BS1330_I1742  | 1<sup>b</sup> | T → A | pyk (no direct interaction with substrate) |
| 24   | BS1330_I1896  | 1<sup>b</sup> | Y → H | sdhA (found) |
| 25   | BS1330_I2038  | 1<sup>b</sup> | E → K | GMC family oxidoreductase (surface) |
| 26   | BS1330_I2085  | 1<sup>b</sup> | A → V | sensor histidine kinase BvrS (unknown) |
| 27   | BS1330_I2149  | 1<sup>b</sup> | A → D | PhoH family protein (found) |
| 28   | BS1330_I0973  | 1<sup>b</sup> | A | IS711, transposase orfA |
| 29   | BS1330_I1108  | 1<sup>b</sup> | P | luciferase family protein |
| 30   | BS1330_I1108  | 1<sup>b</sup> | E | tgt |
| 31   | BS1330_I1130  | 1<sup>b</sup> | I | cobN |
| 32   | BS1330_I1147  | 1<sup>b</sup> | L | short chain dehydrogenase |
| 33   | BS1330_I2107  | 1<sup>b</sup> | I | hypothetical protein |
| 34   | BS1330_I0476  | 78<sup>a</sup> | F → S | pyridine nucleotide-disulfide oxidoreductase family protein (found) |
| 35   | BS1330_I0860  | 16<sup>a</sup> | F → S | eryD (paper) |
| 36   | BS1330_I0148  | 1<sup>a</sup> | F | 3-hydroxy/isobutyrate dehydrogenase (paper) |
| 37   | BS1330_I1053  | 1<sup>a</sup> | S → G | sugar ABC transporter ATPase (found) |
| 38   | BS1330_I0999  | 1<sup>a</sup> | S → G | amidase (unknown) |
| 39   | BS1330_I1074  | 1<sup>b</sup> | S → G | IclR family transcriptional regulator (paper) |
| 40   | BS1330_I0278  | 1<sup>b</sup> | A → T | nnnR (surface) |
| 41   | BS1330_I0440  | 1<sup>b</sup> | A → E | glnK (surface) |
| 42   | BS1330_I0441  | 1<sup>b</sup> | T → A | major facilitator family transporter (found) |
| 43   | BS1330_I0533  | 1<sup>b</sup> | P → Q | oligopeptide ABC transporter, periplasmic oligopeptide-binding protein (paper) |
| 44   | BS1330_I0973  | 1<sup>b</sup> | S → G | oligopeptide ABC transporter, periplasmic oligopeptide-binding protein (paper) |
between bacterial virulence and membrane phosphatidylcholine has been reported in *L. pneumophila* and *A. tumefaciens* [22–24]. The cell envelope of the Brucella contains phosphatidylethanolamine and phosphatidylcholine. Although the phosphatidylcholine level in bacterial membrane does not affect some major virulence properties of *Brucella* such as invasion, intracellular traffic, and intracellular replication, it is necessary to maintain a chronic infection in mice [25, 26]. A systematic study on phosphatidylcholine synthase from *S. meliloti* was performed by alanine scanning mutagenesis. Protein sequence alignment of PCSs from several bacteria species showed that only five of eight residues at position 32 are serine. Changing the conserved residues close to position 32 causes about 20 % decrease in PCS activity [27]. Further studies are needed to explore if the mutation at position 32 affects the PCS activity, which might decrease S2 virulence.

Peptidoglycan is a major component of the bacterial cell wall. Its main function is to preserve the cell integrity and resist osmotic pressure [28]. It has been shown that peptidoglycan synthesis genes affect bacterial shape, elongation, and cell division [29]. Therefore, disruption of peptidoglycan synthesis might affect the survival rate of the bacteria in the host. UDP-N-acetylglucosamine (UDP-GlcNAc) is one of the key building blocks in peptidoglycan biosynthesis [30]. The second step in the formation of UDP-GlcNAc is the interconversion of glucosamine-6-phosphate to glucosamine-1-phosphate, which is catalyzed by the enzyme phosphoglucomutase (PNGM). A growing body of research suggests the importance of PNGM in bacterial virulence and infectivity [31–33]. The different residue in PNGM between S2 and 1330 was located at position 250. Although the residue was close to the substrate binding site, it did not interact with the substrate. The biochemical studies are needed to fully define the function of this residue on PNGM activity and bacterial virulence.

Pyruvate kinase (PK), the key enzyme in the glycolysis, is important for controlling the concentrations of glycolytic intermediates, biosynthetic precursors and nucleoside triphosphates in the cells. Pyruvate kinase transfers the phosphoryl group of phosphoenolpyruvate to ADP to yield pyruvate and ATP. Most eukaryotic PKs can be allosterically activated by fructose 1,6-bisphosphate (FBP), while bacterial PK could be activated by either FBP or AMP/sugar monophosphates [34]. The bacteria carrying the impaired PK activity might exhibit severe attenuation as

### Table 3

The identified non-identical proteins between attenuated strain S2 and virulent strain 1330 (Continued)

| No. | Protein Accession | Difference | Function Description |
|-----|-------------------|------------|----------------------|
| 45  | BS1330_II0567     | □          | S → L pcs (paper)    |
| 46  | BS1330_II0571     | □          | A → P peptide ABC transporter, periplasmic peptide-binding protein (found) |
| 47  | BS1330_II0782     | □          | V → A hypothetical protein (found) |
| 48  | BS1330_II0786     | □          | M → T fadA (found)    |
| 49  | BS1330_II0805     | □          | E → K glucuronate isomerase (interface) |
| 50  | BS1330_II0844     | □          | N → D aminotransferase, class III (surface) |
| 51  | BS1330_II0911     | □          | W → R oxidoreductase, molybdoenzyme-binding (unknown) |
| 52  | BS1330_II0980     | □          | S → P WD repeat-containing protein (unknown) |
| 53  | BS1330_II0990     | □          | C → W hypothetical protein (found) |
| 54  | BS1330_II1031     | □          | P → S lepA (unknown) |
| 55  | BS1330_II1137     | □          | M → L fliF (paper)   |
| 56  | BS1330_II0646     | □          | L branched-chain amino acid ABC transporter, ATP-binding protein |
| 57  | BS1330_II0913     | □          | G outer surface protein |
| 58  | BS1330_II0945     | □          | A branched-chain amino acid ABC transporter, periplasmic amino acid-binding protein |
| 59  | BS1330_II1103     | □          | A hypothetical protein |

**Abbreviations and symbols:** □ = bp Difference, △ = bp Deletion, ▽ = bp Insertion, FS frameshift, PS premature stop; paper: see discussion in the paper

*Single nucleotide polymorphism within ORF*

*Representative locus from 1330 genomes*

### Table 4

Effect of erythritol on growth of *Brucella suis* strain S2 and *B. abortus* strain S19

| Strain | TSA | TSA (Ery) |
|--------|-----|-----------|
|        | 10^7 | 10^6 | 10^5 | 10^4 | 10^3 | 10^2 | 10^1 | 10^0 |
| S19    | C   | C   | +++ | +++ | ++ | + | ± | — |
|        | ± | — | — | — | — | — | — | — |
| S2     | C   | C   | +++ | +++ | +++ | ++ | + | C |
|        | C   | C   | +++ | +++ | +++ | ++ | ± | ± |

**Estimate of growth:** C, confluent; ++++, numerous individual colonies; ++, more than 100 colonies; +, 10 to 100 colonies; ±, 1 to 10 colonies
suggested in the study in C. Neoformans [35]. In chromosome I of S2, there was one residue different (T462A) in PK compared to that in 1330. Based on result of the Protein Databank search by blastp program, Brucella PK shared some similarities with PKs from other species, such as human, rabbit, B. Stearothermophilus, S. aureus et al. In the structure of human pyruvate kinase isozyme M2, the corresponding threonine at position 462 was close to the FBP binding site and interacted with FBP through two water molecules [36]. The threonine residue is highly conserved in many PKs, except for the serine in E. coli. Both threonine and serine contain hydroxyl groups, which might be important for its interaction with the activator. A study demonstrates that S2 grows normally in the medium containing glucose as the only carbon source [37]. We, therefore, suspected the mutant on PK might not be a major reason for S2 attenuation. Further biochemical studies are still necessary to clarify this conclusion.

Although Brucella spp. is considered typically as non-motile, a study demonstrates that flagellar genes are necessary for the assembly of a functional flagellum [38, 39]. B. melitensis with fliF mutant is unable to replicate in macrophages and HeLa cells, suggesting the Brucella flagellar genes might be related to virulence [40]. Another study shows that in the early log phase of its growth in 2YT nutrient broth, B. melitensis expresses genes corresponding to the basal (MS ring) and the distal parts of the flagellar apparatus. A polar and sheathed flagellar structure has been observed under transmission electron microscopy. Although the mutations encoding different parts of the flagellum structure do not show a discernible phenotype when compared with the wild-type strain in cellular infection models, all these mutants are attenuated in mice infected via the intraperitoneal route in 4 weeks after infection [41]. We found a mutation in fliF protein (M544L) located in MS ring. Whether the mutation decreases the virulence of S2 awaits further studies.

Other different proteins found between S2 and 1330

Two IclR regulatory genes in S2 were different from those in 1330. The IclR regulatory gene located in chromosome I lacked 8 bp that began at ATG, causing the loss of the coding protein. The other different IclR regulatory gene located in chromosome II had a 1 bp deletion, resulting in the premature stop in the coding protein. Members of the IclR transcription regulator family regulate a variety of metabolic processes in various microorganisms [42]. In one study, B. melitensis 16M mutations in AraC, ArsR, Crp, DeoR, GntR, IclR, LysR, MerR, RpiR and TetR were constructed. GntR and LysR mutants showed the decreased virulence [43]. Two different IclR regulatory genes found in S2 genome were correspondent to IclR5 and IclR6 in the study. Therefore, the lack of these two IclR regulatory genes might not be the major cause for S2 attenuation.

In chromosome I of S2, Omp2b had two residues different from that in 1330. Omp2b is bacterial porins present in the outer membrane of Brucella. The differences in the Omp2b variants are located in the predicted external loops of the porin [44]. The first different residue (D79A) is highly conserved in different Brucella strains, while the second different residue (R225H) is not strictly conserved. Omp2b is present in virulent B. abortus, B. melitensis and B. suis strains that are pathogenic to...
humans, but is absent in non-pathogenic \textit{B. ovis} [45]. Bacterial porsins have been reported to be involved in apoptosis modulation [46–48]. Expression of \textit{B. melitensis} Omp2b in yeast prevents the cell from death induced by the mammalian pro-apoptotic protein Bax [49]. No further functional studies have been reported on \textit{Brucella} Omp2b.

High-affinity binding protein-dependent ABC transporter consists of a high-affinity periplasmic substrate-binding subunit, two hydrophobic membrane subunits and two additional cytoplasmic subunits. ATP hydrolysis by associated ATPase provides the energy for substrate accumulation across the inner membrane against the concentration gradient. Periplasmic substrate-binding subunits are composed of two separate but similarly folded globular domains, which are connected by a hinge region made of two or three short polypeptide segments [50]. In S2 chromosome I, 1 bp difference in periplasmic sugar-binding protein caused one residue difference (S97Y) in protein sequence compared with that in 1330 strain. The sequence alignment with structure-solute receptor GacH of \textit{Streptomyces glaucescens} (Protein Databank accession number: 3JZJ [51]) showed that this residue was located very close to the substrate binding site. Due to the bulky structure of tyrosine, we suspected the mutation might affect the sugar binding affinity or preference of the protein. The comparative proteome study on \textit{B. melitensis} vaccine strain Rev 1 and virulent strain 16 M reveals that four sugar-binding proteins are differentially expressed in Rev 1, although these two strains have similar physical properties in biotyping tests [52]. Therefore, it is possible that the substrate preference change might help bacteria adapt to different growth conditions.

In S2 chromosome II, 2 bp differences resulted in 2 residues difference in oligopeptide ABC transporter. The first different residue (S136G) existed in other \textit{Brucella} strains, while the other different residue was only seen in S2 (P284Q). Bacteria utilize peptides as a source of amino acids, carbon, nitrogen, and/or energy. A variety of peptide uptake systems mediate the translocation of peptides across the bacterial cytoplasmic membrane. Recent studies show that OppA possesses preference in substrate peptide. \textit{E. coli} OppA prefers positively charged peptides with three or four amino acids in length. However, other studies also suggest that the substrate preference of the peptide uptake systems are more complicated [53, 54]. The sequence alignment with the homologous structure of oligopeptide-binding protein from \textit{Salmonella typhimurium} [55] showed that residue 284 was located close to the substrate binding site, but not involved in direct interaction with the substrate. Compared with other available structures of oligopeptide-binding proteins, the residue 284 of OppA protein can be glutamate (\textit{S. typhimurium}), serine (\textit{E. coli}), alanine (\textit{E. coli}) or tyrosine (\textit{B. Pseudomallei}). Meanwhile, two oligopeptide-binding proteins were encoded in 1330 genome, which shared 84 % similarity in protein sequences. The residue at 284 is aspartate in the other sequence, suggesting this residue was not conserved well in different bacteria. ATP-Binding Cassette Transporter in \textit{Brucella ovis} is shown to be essential in mice [56]. The \textit{ΔabcAB} mutant strain can trigger host serologic responses similar to the WT strain and a significant cellular host response [57].

One residue different in S2 was found in a DedA family protein. The DedA family proteins are integral inner membrane proteins, which are present in nearly all species of bacteria. \textit{E. coli} encodes 8 DedA proteins. Although each of them is nonessential, they are collectively essential [58]. An \textit{E. coli} mutant with two \textit{dedA} genes deletion (\textit{ΔyghB}/\textit{ΔyqjA}) fails to complete cell division or grow at elevated temperatures [59]. \textit{B. burgdorferi} possesses only one \textit{dedA} gene. The gene deletion results in imbalanced membrane phospholipid composition, which is required for proper cell division and envelope integrity [60]. No functional clue was suggested for \textit{Brucella} \textit{DedA} so far.

We identified a difference at residue 576 between S2 and 1330 in a putative molybdopterin-binding oxidoreductase. Members of the molybdopterin oxidoreductase family include formate dehydrogenase, nitrate reductase, DMSO reductase, TMAO reductase, pyrogallol hydroxyltransferase and arsenate reductase. The protein in S2 showed about 40 % similarity with formate dehydrogenase H from \textit{E. coli} at residues 54–420, as suggested by the protein databank search with blastp. The residues involved in Mo binding are conserved [61]. No significant sequence similarity was found between these two proteins around residue 576.

A 1 bp insertion caused the premature stop of 3-hydroxyisobutyrate dehydrogenase in S2. 3-hydroxyisobutyrate dehydrogenase, a key enzyme for the metabolism of valine and some keto-bodies, exists widely in bacteria, yeast, and mammalian tissue. It catalyzes the reversible conversion of 3-hydroxyisobutyrate to methylmalonate semialdehyde [62]. Although the enzymatic properties of 3-hydroxyisobutyrate dehydrogenase have been studied in several organisms, its relevance to bacterial virulence awaits further studies.

NADH-quinone oxidoreductase subunit 1 (NuoF) in S2 had one different residue compared with that of 1330 strain. NuoF is part of the hydrophilic fragment of NADH dehydrogenase I, which represents the electron input part of the enzyme NADH:ubiquinone oxidoreductase. It couples electron transport to proton translocation across the membrane, which is partially responsible for generating the proton gradient necessary for ATP production. NADH:ubiquinone oxidoreductase serves as both a proton pump and an entry point for electrons into the respiratory chain [63]. Currently, no bacterial virulence is associated with this gene.
One residue difference was found in lepA gene between S2 and 1330. LepA protein encoded by this gene exists in most of the bacterial genomes [64]. Although lepA knock-out strain of *E. coli* [65] and *Staphylococcus aureus* [66] are viable, the gene product might be necessary for bacteria survival under certain growth conditions [67].

There are other different proteins between S2 and 1330 (labeled unknown in Table 3). When we analyzed the sequences, we did not find any functional implications of the proteins in bacteria. Therefore, further studies are needed to clarify their links to virulence attenuation of S2 strain.

**Conclusion**

In this report, we sequenced the attenuated strain *Brucella suis* S2 genome and performed comparative genome analysis between S2 and virulent strain 1330. During the analysis, we found out 59 different ORFs between these two strains. Of these different ORFs, several have been reported to be related to *Brucella* virulence, such as outer membrane autotransporter, eryD. The proteins encoded by some of the different genes have been suggested to be essential for bacteria growth, such as dedA, phosphoglucosamine mutase and phosphatidylcholine synthase. These gene products might be responsible for the utilization of the materials from the environment, which are necessary for bacteria growth. We found several mutations in peptide-binding ATP transporter and sugar-binding ATP transporter, but their roles on S2 attenuation could not be ruled out without further investigation.

It has been shown that the high concentration of erythritol in the foetal tissues of cattle, sheep, goats and pigs might be beneficial for the growth of *Brucella*. Erythritol catabolism pathway may play important physiological roles in *Brucella*. The studies in S19 and S2308 showed that eryD might act as the repressor of erythritol operon [17]. In our experiment, we found that S2 but not S19 could grow in the media supplemented with erythritol, suggesting S2 has the ability to metabolize erythritol. Therefore, further investigation on the function of the eryD gene is necessary to clarify its relevance to virulence attenuation.

Although we identified some target proteins that might be related to the virulence, most of the identified proteins have not been investigated in *Brucella*. Therefore, all these genes have to be examined individually to ensure their relationships with the virulence. The modification of the genes identified in our study will help to reduce residual virulence, which is critical to developing more efficient *Brucella* vaccine strains.

**Methods**

**Strain and genomic DNA preparation**

The S2 used in sequencing was obtained from the China Institute of Veterinary Drugs Control. Total genomic DNA was extracted using the DNeasy Blood and Tissue Kit (QIAGEN, China Ltd., China). The DNA was fragmented by nebulization, and fragments ranging from 400 to 600 bp were extracted from an agarose gel after size-fractionation, and then adaptors were added according to Illumina Sample Preparation Guide.

**Genome sequencing and assembly**

After sequencing with Illumina Hiseq2000, 50,941,654 paired-end (2*101 bp) reads were obtained. Base calling was performed with the software CAVASA. There were 41,608,314 paired reads passed the filter, which provided coverage of chromosome of ~1000-fold. Reads were assembled into contigs and scaffolds using SOAP de-novo (Release 1.04) [68]. Assembled contigs were compared to the published genome sequence of *B. suis* strain 1330 (NC_017250.1, NC_017251.1) using BLAT v.34 [69]. The order and the orientation of the assembled contigs were determined in accordance with the control genome. Based on the extensive similarities between the published genomes of *Brucella*, PCR primers were designed to link the gaps between two neighboring contigs. The leftover gaps were sequenced separately by conventional Sanger sequencing. The Illumina sequence reads for the genomes are deposited to NCBI, with the accession number of SAMN03068300.

**Gene prediction and annotation**

Gene prediction was established with Glimmer 2.0 using the default settings [70]. All the predicted CDS and putative intergenic sequences were subjected to further manual inspections. Exhaustive BLAST searches with an incremental stringency against the NCBI non-redundant protein database were performed to determine the homology of the predicted coding sequences. Translational start codons were identified based on protein homology, proximity to the ribosome-binding site, relative positions to predicted signal peptide and putative promoter sequences. We also classified the putative proteins according to the COG database search results. Transfer RNAs were predicted with the tRNAscan-SE software. The repeat sequences were predicted by Trf program [71]. The results showed 55 and 23 tandem sequences in chromosome I and II. The average copy number is 2.79 and 2.86.

All the S2 proteins predicted have been compared to KEGG bacterial gene set with KAAS (KEGG Automatic Annotation Server http://www.genome.jp/tools/kaas/) program by BBH (Bi-directional best hit) method to map the genes into bacterial pathways automatically.

**Comparative genomic analysis**

MUMmer 3.0 program was used to investigate the differences between the genome sequences of 1330 and S2 with the default parameters [72]. The genomes
were first automatically compared for SNPs and indels by MUMmer3.0. Then SNPs between 1330 and S2 were called by home-brew perl scripts. We classified SNPs as intergenic, frameshift and premature stop according to their positions and functions, and also counted the numbers of synonymous substitutions and non-synonymous substitutions. To identify potential genes that were different between 1330 and S2, we compared S2 genome sequences using blast program with proteins download from NCBI RefSeq database.

**S2 and S19 erythritol sensitivity test**

The Trypticase-soy agar plates in the presence of 1 mg/ml erythritol [TSA (Ery)] plates were prepared [73]. The autoclaved TSA was cooled to 47 °C, and the appropriate amount of erythritol was added and mixed. The plates were poured. S2 and S19 were grown on TSA slants at 37 °C for 24 h and resuspended in saline, and were tested by placing 20 μl of tenfold dilutions (10^4, 10^6, 10^8, 10^9, 10^10 and 10^11 CFU/ml) of the bacteria resuspension on TSA and TSA (Ery) plates and incubated at 37 °C. Plates were examined for 2 to 5 days. The effect of erythritol on the growth of S2 was estimated from the size of the colonies on the test plates in comparison with the control plates (S2 and S19 inoculated on TSA plates and S19 inoculated on TSA (Ery) plates). The experiments were repeated three times.

**Additional files**

Additional file 1: Table S1. Results of genome-wide collinear analysis between attenuated strain S2 and virulent strain 1330 by Mummer. (DOC 56 kb)

Additional file 2: Table S2. Single nucleotide polymorphism difference between attenuated strain S2 and virulent strain 1330. (DOC 146 kb)

**Acknowledgement**

This study was funded by the Science and Technology Basic Work Program (2012FY111000) and the National Natural Science Foundation of China (81271900). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Authors' contributions**

DDD, WXF and XM conceived and designed the experiments; DDD performed the experiments; DDD, LJT, JKL, YHY, LKX, NB and YY analyzed the data; BYC, HJ, XY, and XM wrote the manuscript. All authors read and approved the final manuscript.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

1Laboratory of Zoonoses, China Animal Health and Epidemiology Center, Qingdao, Shandong, China. 2Chinese Academy of Agricultural Sciences, Shanghai Veterinary Research Institute, Shanghai, China. 3State 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. 4Xinjiang Academy of Animal Science, Institute of Veterinary Research, Urumuqi, Xinjiang, China. 5ZheJinXin Biotechnology Shanghai Co, Ltd. 12F, Building 1, 100 Qinzhou Road, Shanghai, China.

References

1. Pappas G, Papadimitriou P, Akritidis N, Christou L, Tsianos EV. The new global map of human brucellosis. Lancet Infect Dis. 2006;6(2):91–9.
2. O’Callaghan D, Whatmore AM. Brucella genetics as we enter the multi-genome era. Brief Funct Genomics. 2011;10(6):334–41.
3. DeVecchio VG, Kapral V, Redkar RJ, Patra G, Mujer C, Los T, Ivanova N, Anderson I, Bhattacharyya A, Lykidis A, et al. The genome sequence of the facultative intracellular pathogen Brucella melitensis. Proc Natl Acad Sci U S A. 2002;99(1):443–8.
4. Paulsen IT, Seshadri R, Nelson KE, Eisen JA, Heidelberg JF, Read TD, Dodson RJ, Umayam L, Brinkac LM, Beanan MJ, et al. The Brucella suis genome reveals fundamental similarities between animal and plant pathogens and symbionts. Proc Natl Acad Sci U S A. 2002;99(20):13148–53.
5. Tae H, Shalomy S, Settage R, Preston D, Adams LG, Garner HR. Revised genome sequence of Brucella suis 1330. J Bacteriol. 2011;193(22):6410.
6. Cretsa CR, Folkerts O, Fei Z, Mane SP, Evans C, Martino-Catt S, Bricke B, Yu G, Du L, Sobral BW. Genome sequence of Brucella abortus vaccine strain S19 compared to virulent strains yields candidate virulence genes. PLoS One. 2008;3(5):e2193.
7. Morgan WJ, Corbel MI. Recommendations for the description of species and biotypes of the genus Brucella. Dev Biol Stand. 1976;31:27–37.
8. Moreno E, Cloeckaert A, Morion Y. Brucella evolution and taxonomy. Vet Microbiol. 2002;90(1–4):209–27.
9. Xin X. Orally administrable brucellosis vaccine: Brucella suis strain 2 vaccine. Vaccine. 1986;4(4):212–6.
10. Deqiu S, Donglou X, Jiming Y. Epidemiology and control of brucellosis in China. Vet Microbiol. 2002;90(1–4):165–82.
11. Verger JM, Grayon M, Zundel E, Lechopier P, Olivier-Bernardin V. Comparison of the efficacy of Brucella suis strain 2 and Brucella melitensis Rev 1 live vaccines against a Brucella melitensis experimental infection in pregnant ewes. Vaccine. 1995;13(2):191–6.
12. Blasco JM, Marin C, De Bagues Jimenez MP, Barberan M. Efficacy of Brucella suis strain 2 vaccine against Brucella ovis in rams. Vaccine. 1993;11(13):1291–4.
13. Mustafa AA, Abusowa M. Field-oriented trial of the Chinese Brucella suis strain 2 vaccine on sheep and goats in Libya. Vet Res. 1993;24(5):422–9.
14. Bossay N, Plommet M, Brucella suis S2, Brucella melitensis Rev. 1 and Brucella abortus S19 living vaccines: residual virulence and immunity induced against three Brucella species challenge strains in mice. Vaccine. 1990(8):462–8.
15. Krywinski M, Schein J, Birol I, Conners J, Gascoyne R, Hornman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2005;15(10):1476–84.
16. Bandara AB, Srijananganath S, Schurig GG, Boyle SM. Putative outer membrane autotransporter protein influences survival of Brucella suis in BALB/c mice. Vet Microbiol. 2005;109(1–2):95–104.
17. Sangari FJ, Aguero J, Garcia-Lobo JM. The genes for erythritol catabolism are organized as an inducible operon in Brucella abortus. Microbiology. 2000;146(Pt 2):487–95.
18. Sangari FJ, Grillo MJ, De Bagues Jimenez MP, González-Carrero MJ, Garcia-Lobo JM, Blasco JM, Aguero J. The defect in the metabolism of erythritol of the Brucella abortus B19 vaccine strain is unrelated with its attenuated virulence in mice. Vaccine. 1998;16(17):1640–5.
19. Barbier T, Collard F, Zuniga-Ripa A, Morion Y, Godard T, Becker J, Wittmann C, Van Schaftingen E, Letesson JJ. Erythritol feeds the pentose phosphate pathway via three new isomerases leading to D-erythrose-4-phosphate in Brucella. Proc Natl Acad Sci U S A. 2011;108(50):20171–20175.
20. Söhlenkamp C, de Rudder KE, Rohrs V, Lopez-Lara IM, Geiger O. Cloning and characterization of the gene for phosphatidylcholine synthase. J Biol Chem. 2000;275(23):18919–25.
21. de Rudder KE, Söhlenkamp C, Geiger O. Plant-exuded choline is used for rhizobial membrane lipid biosynthesis by phosphatidylcholine synthase. J Biol Chem. 1999;274(28):20011–6.
65. Dibb NJ, Wolfe PB. lep operon proximal gene is not required for growth or secretion by Escherichia coli. J Bacteriol. 1986;166(1):83–7.

66. Colca JR, McDonald WS, Waldon DJ, Thomasco LM, Gadwood RC, Lund ET, Cavey GS, Mathews WR, Adams LD, Cecil ET, et al. Cross-linking in the living cell locates the site of action of oxazolidinone antibiotics. J Biol Chem. 2003;278(24):21972–9.

67. Bijlmer J, Lee ALM, Nootenboom IC, Vandenbroucke-Grauls CM, Kusters JG. Identification of loci essential for the growth of Helicobacter pylori under acidic conditions. J Infect Dis. 2000;182(5):1566–9.

68. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, et al. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 2010;20(2):265–72.

69. Kent WJ. BLAT—the BLAST-like alignment tool. Genome Res. 2002;12(4):656–64.

70. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 1999;27(23):4636–41.

71. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.

72. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res. 2002;30(11):2478–83.

73. Jones LM, Montgomery V, Wilson JB. Characteristics of Carbon Dioxide-Independent Cultures of Brucella Abortus Isolated from Cattle Vaccinated with Strain 19. J Infect Dis. 1965;115:312–20.