JIDC | THE JOURNAL OF INFECTION IN DEVELOPING COUNTRIES

Brief Original Article

Geographical distribution of *Brucella melitensis* inferred from *rpoB* gene variation

Kim-Kee Tan[1,2], Yung-Chie Tan[3], Li-Yen Chang[1,2], Kok Wei Lee[3], Siti Sarah Nore[1], Wai-Yan Yee[3], Chee-Choong Hoh[3], Sazaly AbuBakar[1,2]

[1] Tropical Infectious Diseases Research and Education Centre (TIDREC), University of Malaya, Kuala Lumpur, Malaysia
[2] Department of Medical Microbiology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
[3] Department of Science and Technology, Codon Genomics S/B, Seri Kembangan, Malaysia

Abstract

Introduction: Currently available tests have limitations for the identification of *Brucella* species and strains, and their genetic lineage. The genome sequence of the *rpoB* gene encoding the β-subunit of DNA-dependent RNA polymerase was investigated for its use in genotyping *Brucella melitensis*.

Methodology: Complete *rpoB* gene sequences of globally distributed *Brucella melitensis* strains were analyzed. Single nucleotides polymorphisms (SNPs) of the *rpoB* gene sequences were identified and used to type *Brucella melitensis* strains.

Results: Six DNA polymorphisms were identified, of which two (nucleotides 3201 and 558) were novel. Analysis of the geographical distribution of the strains revealed a spatial clustering pattern with *rpoB* type 1 representing European and American strains, *rpoB* type 2 representing European, African, and Asian strains, *rpoB* type 3 representing Mediterranean strains, and *rpoB* type 4 representing African (C3201T) and European (C3201T/T558A) strains.

Conclusions: We report the discovery of two novel SNPs of *rpoB* gene that can serve as useful markers for epidemiology and geographical tracking of *B. melitensis*.

Key words: *Brucella*; *rpoB*; SNPs.

J Infect Dev Ctries 2017; 11(5):420-425. doi:10.3855/jidc.7598

(Received 26 August 2015 – Accepted 25 March 2016)

Introduction

Brucella spp. is small Gram-negative coccobacillus that lives as a facultative intracellular pathogen in the host. It is the causative agent of brucellosis, which primarily affects livestock and wildlife [1]. Infection in animals causes abortion, low milk production, and stillbirth, affecting the output of animal husbandry and leading to economic losses [2]. Humans usually acquire the infection through contact with infected animals or by ingestion of unpasteurized dairy products. Human infection can result in a chronic debilitating disease with non-specific symptoms affecting multiple organs [3,4].

Brucella spp. is a category B potential biological warfare agent that can be transmitted through aerosols [5]. Rapid detection and precise identification of *Brucella* spp. is, therefore, essential to determine the possible native geographical and host origins. Diagnostics and identification of brucellosis based on biochemical tests are inadequate for the differentiation of *Brucella* spp. In a clinical setting, general identification of *Brucella* spp. can be performed using 16s rRNA gene sequencing. The method however, is still inadequate for differential identification of *Brucella* species and strains important for the determination of the origin of the infection.

Owing to the high genomic similarity among the *Brucella* spp., identification methods that target multiple genes, including multi-locus variable-number tandem-repeat analysis and multilocus sequencing, have been developed [6-8]. In the meantime, use of the *rpoB* gene sequence for bacteria identification, especially between closely related species, has been reported [9]. Thus far, this approach demonstrated high sensitivity in differentiating *Brucella* spp. down to their genotype through single-gene sequencing prodecure [10,11]. In the present study, we explore the use of the *rpoB* gene to study intraspecies variation on a large set of diverse *B. melitensis* strains collected globally.

Methodology

Two *Brucella melitensis* strains isolated in Malaysia, namely MY/2009/1483 and BM/Phil/2012/1136, were previously recovered from brucellosis patients who had consumed unpasteurized cow milk in Malaysia and goat milk in the Philippines, respectively [12]. Complete sequences of the *rpoB* gene were retrieved from draft genomes of previously reported isolates [12].

The *rpoB* gene sequences of other global *B. melitensis* strains were retrieved from the GenBank and the PATRIC databases [13]. Complete *rpoB* gene sequences from 56 globally distributed *B. melitensis* strains, including two human *B. melitensis* MY1483/09 (accession number: PRJEB7499) [14] and Phil1136/12 (accession number: PRJEB7504), were aligned. SNPs from the sequence alignment were identified and compared. For phylogenetic analysis, multiple sequence alignments of the 56 *rpoB* gene sequences from *B. melitensis* and two from *B. abortus* (as outgroups) were carried out using ClustalW 2.0.12 [15]. The phylogenetic tree of the *rpoB* genes was constructed using MrBayes v3.2.1 [16]. Bayesian MCMC analysis was conducted by sampling across the entire general time reversible (GTR) model space. One million generations were run with a sampling frequency of 500, and diagnostics were calculated for every 5,000 generations. A burn-in setting of 25% was used to discard the first 500 trees. Convergence was assessed manually, with the standard deviation of split frequencies falling below 0.01. There was no obvious trend for the plot of the generation versus the log probability of the data (the log likelihood values), and the potential scale reduction factor (PSRF) was reasonably close to 1.0 for all parameters. The discriminatory power of the *rpoB* typing method was calculated as described previously [17] and implemented in insilico.ehu.eus (http://insilico.ehu.es/mini_tools/discriminatory_power/index.php) [18].

Results

Nucleotide differences in the *B. melitensis rpoB* gene variants are shown in Table 1. Using *B. melitensis* 16M as reference, six DNA polymorphisms were observed, two of which were novel (Table 1). The first novel DNA polymorphism, located at residue 3201, involved C to T substitution. It was present in most of the African isolates in the study and in one *B. melitensis* strain (UK22_04.201744) collected in the UK. This polymorphism enabled the segregation of 3201T-*B. melitensis* strains from other *B. melitensis* strains (*rpoB* types 1, 2 and 3). Isolates with the 3201T SNP are designated as *rpoB* type 4 in the present study. Within the *rpoB* type 4 isolates, an additional SNP at residue 558, T to A was observed exclusively in the UK22_04.201744 strain.

Figure 1. Geographical distribution of *B. melitensis* strains according to *rpoB* typing.

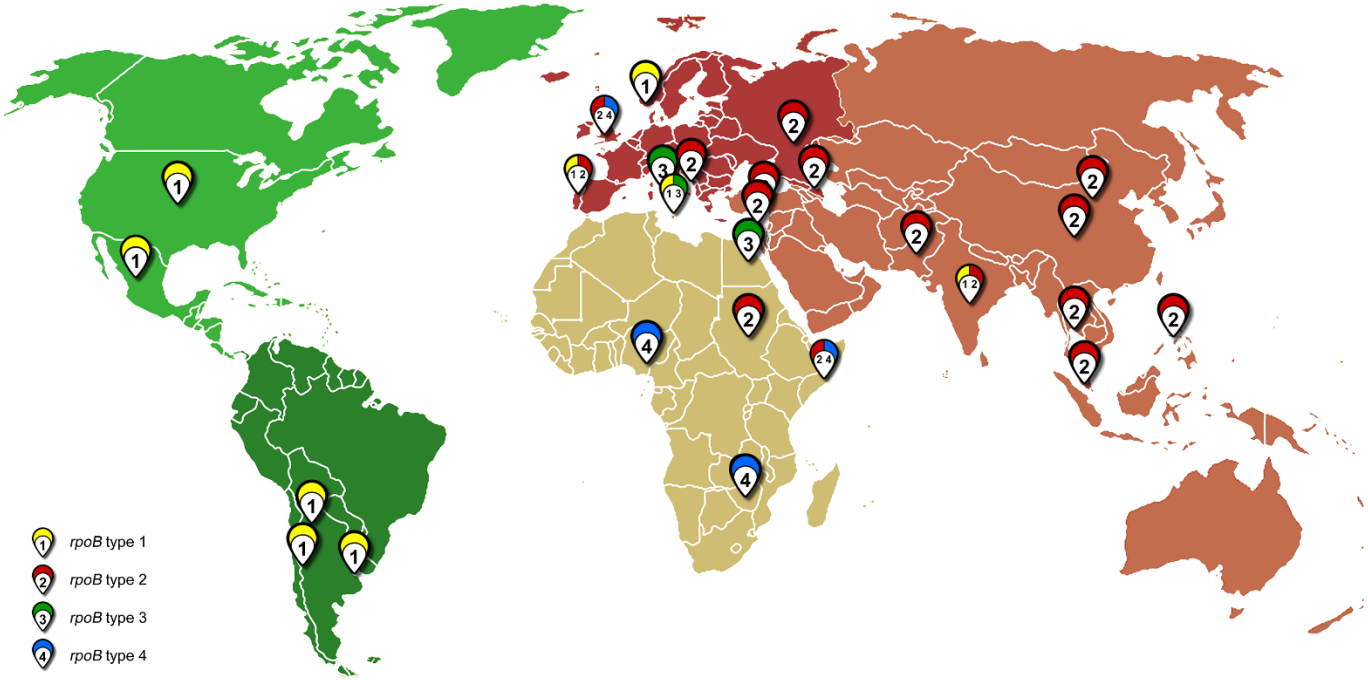


Table 1. Nucleotide in the *rpoB* of 56 *B. melitensis* strains used in this study.

Isolates	*rpoB* type	subtype	Country	Continent	Single nucleotide polymorphisms*					
					558	1886	2954	3201	3747	3927
16M	1	-	USA	North America	T	C	C	C	G	G
bv1strRev1	1	-	Mexico	North America	.	.	.	.	.	.
bv1str16M 1	1	-	USA	North America	.	.	.	.	.	.
CNGB1076	1	-	Argentina: San Juan	South America	.	.	.	.	.	.
CNGB1120	1	-	Argentina: Buenos Aires	South America	.	.	.	.	.	.
CNGB290	1	-	Argentina: Jujuy	South America	.	.	.	.	.	.
F10 05 2	1	-	Portugal	Europe	.	.	.	.	.	.
B115	1	-	Malta	Europe	.	.	.	.	.	.
F3 02	1	-	Norway	Europe	.	.	.	.	.	.
ADMAS-G1	1	-	India	Asia	.	.	.	.	.	.
UK3 06	2	a	Cyprus	Europe	.	.	.	.	.	A
F8 01 155	2	b	Kosovo	Europe	.	T	.	.	.	A
F9 05	2	b	Turkey	Asia	.	T	.	.	.	A
UK22 06	2	b	Somalia	Africa	.	T	.	.	.	A
UK37 05	2	b	UK	Europe	.	T	.	.	.	A
UK29 05	2	b	UK	Europe	.	T	.	.	.	A
11-1823-3434	2	b	Unknown	Unknown	.	T	.	.	.	A
F10 06 16	2	c	Thailand	Asia	.	T	T	.	.	A
MY/2009/1483	2	c	Malaysia	Asia	.	T	T	.	.	A
BM/Phil/2012/1136	2	c	Philippines	Asia	.	T	T	.	.	A
66 59	2	c	India	Asia	.	T	T	.	.	A
bv2str63 9	2	c	India	Asia	.	T	T	.	.	A
ATCC23457	2	c	India	Asia	.	T	T	.	.	A
BM IND-1	2	c	India	Asia	.	T	T	.	.	A
F6 05 6	2	c	Sudan	Africa	.	T	T	.	.	A
BG2 S27	2	c	Pakistan	Asia	.	T	T	.	.	A
F2 06 6	2	c	Portugal	Europe	.	T	T	.	.	A
NI.158853	2	c	Inner Mongolia	Asia	.	T	T	.	.	A
M5 10	2	c	China	Asia	.	T	T	.	.	A
bv1str16M 2	2	c	China	Asia	.	T	T	.	.	A
bv1strM28 12	2	c	China	Asia	.	T	T	.	.	A
16M1W	2	c	China	Asia	.	T	T	.	.	A
bv1strM111	2	c	China	Asia	.	T	T	.	.	A
bv1strBCB028	2	c	China	Asia	.	T	T	.	.	A
bv1strBCB033	2	c	China	Asia	.	T	T	.	.	A
bv1str133	2	c	China	Asia	.	T	T	.	.	A
bv3str128	2	c	China	Asia	.	T	T	.	.	A
bv1strM5	2	c	China	Asia	.	T	T	.	.	A
M5	2	c	China	Asia	.	T	T	.	.	A
M5 90	2	c	China	Asia	.	T	T	.	.	A
M28	2	c	China	Asia	.	T	T	.	.	A
548	2	c	Russia: Saratov region	Asia	.	T	T	.	.	A
C-554	2	c	Russia: Republic of Dagestan	Asia	.	T	T	.	.	A
C-555	2	c	Russia: Republic of Dagestan	Asia	.	T	T	.	.	A
C-558	2	c	Russia: Republic of Dagestan	Asia	.	T	T	.	.	A
02-7258	2	c	Unknown	Unknown	.	T	T	.	.	A
UK31 99	3	-	Egypt	Africa	.	.	.	.	A	.
bv3strEther	3	-	Italy	Europe	.	.	.	.	A	.
F15 06 7	3	-	Italy: Sicily	Europe	.	.	.	.	A	.
F5 07 239A	3	-	Sicily	Europe	.	.	.	.	A	.
F1 06 B10	4	a	Zimbabwe	Africa	.	.	.	T	.	.
UK24 06	4	a	Nigeria	Africa	.	.	.	T	.	.
UK19 04	4	a	Somalia	Africa	.	.	.	T	.	.
UK14 06	4	a	Somalia	Africa	.	.	.	T	.	.
UK23 06	4	a	UK (Somalia)	Africa	.	.	.	T	.	.
UK22 04	4	b	UK	Europe	A	.	.	T	.	.

*Nucleotide position based on 16M genome

By incorporating geographical data (Figure 1), *rpoB* type 1 corresponded to strains collected from European and American countries, while *rpoB* type 2 comprised strains collected from Europe, Africa, and Asia. Both the Malaysian isolates used in this study were included in this group. Within the *rpoB* type 2 group, a few variants possessed only one or two unique type-2 SNPs out of the three (at codon 629, 985, and 1309). The UK3_06.201748 strain isolated in Cyprus was the only isolate to have a single SNP of *rpoB* type 2 at codon 1309. The strain was designated subtype a. The second *rpoB* type 2 variant possessing two SNPs located at codons 629 and 1309 was designated subtype b. These variants comprised strains from diverse geographical locations including Kosovo, Turkey, Somalia, and the United Kingdom. The remaining isolates, designated subtype c, possessed all three SNPs unique to *rpoB* type 2. The *rpoB* type 3 strain corresponded to isolates collected from Mediterranean region, Egypt, and Italy. Finally, *rpoB* type 4, which was newly coined in the present study, encompassed isolates collected from Africa (subtype a) and the United Kingdom (subtype b).

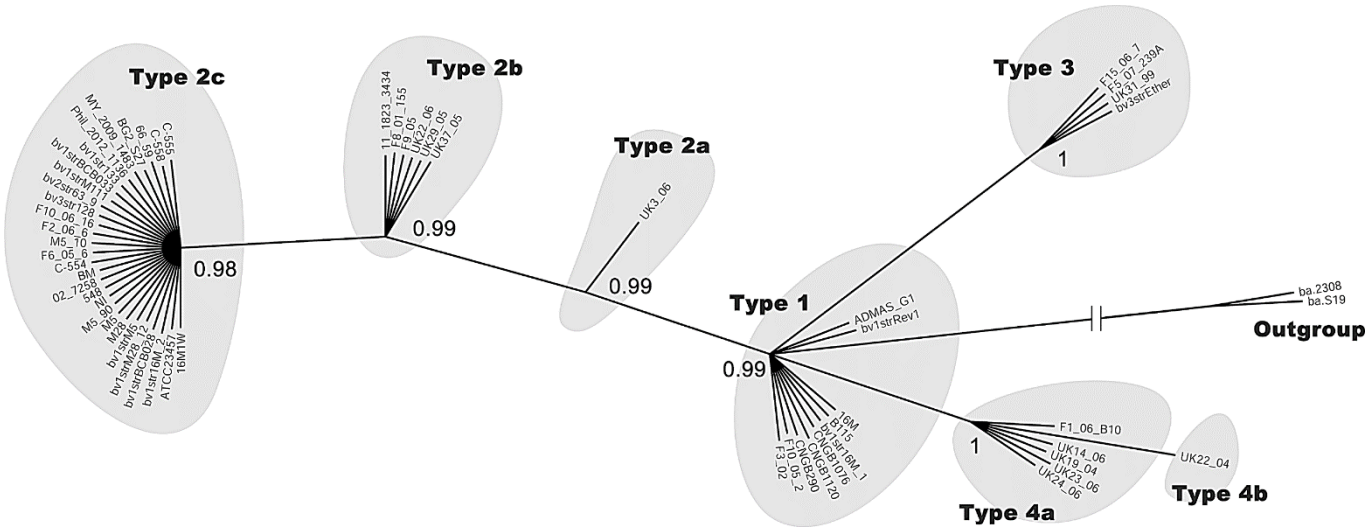
The phylogenetic analysis of the *rpoB* gene revealed a grouping similar to that proposed above. All the *B. melitensis* strains were segregated into four major groups (Figure 2) that corresponded well with the *rpoB* types assigned in the current study. The phylogenetic tree was first delineated into two major groups that corresponded to *rpoB* type 2 and *rpoB* types 1/3/4. The *rpoB* type 2 was then further delineated into three sub-groups that corresponded well with *rpoB* type 2 subtypes a, b, and c. The group that corresponded with *rpoB* types 1/3/4 was further delineated into 3 sub-groups: *rpoB* type 1 (yellow), *rpoB* type 3 (green), and *rpoB* type 4 (blue). The discriminatory power of this *rpoB* typing method was 0.6917.

Discussion

Analysis of the complete *rpoB* gene sequence of globally diverse *B. melitensis* isolates revealed the existence of four distinct groups within the species. Here, we report, for the first time, a new SNP at residue 3201 within the *B. melitensis* species which functions as a canonical SNP to allow the delineation of *B. melitensis rpoB* type 4 strains corresponding to African strains that have not been previously well studied. One isolate with an additional SNP at residue 558 from the United Kingdom was grouped within *rpoB* type 4. Our study suggests that the SNP at residue 3201 could be an adaptation of *B. melitensis* after its introduction into Africa. It accumulated in the ancestral genome of the African lineage before subsequently spreading within the African continent. The nucleotide change at residue 558 might possibly represent a canonical SNP that defines the European *B. melitensis* of African origin. Nevertheless, it remains inconclusive whether the SNP at residue 558 is geographically informative, as more isolates are required to enable a better segregation analysis.

A previous study assigned the strains of *rpoB* type 2 according to three SNPs [10]. With the current analysis, we note that the acquisition of the three nucleotide polymorphisms in the *rpoB* gene by type 2 strains reflects the evolution of this strain. The UK3_06 isolated from Cyprus was the only strain from this study that displayed one out of three SNPs specific for *rpoB* type 2 *B. melitensis*, whereas the strains characterized

Figure 2. Phylogenetic tree of *rpoB* gene sequences from *B. melitensis* isolates.



by two of the three SNPs had diverse geographical distributions, including Kosovo, Turkey, Somalia, and the United Kingdom. The co-circulation of the *rpoB* type 2 strains with two and three SNPs in Turkey [11] suggests that the *B. melitensis* strains (with two SNPs) could have persisted for a period of time in Turkey, thus allowing the accumulation of the third SNP (2954T) before spreading into other Asian countries.

Conclusions

In summary, our findings show a high degree of *rpoB* gene similarity among *B. melitensis* isolated from geographically closely related regions. Identification of SNPs in the *rpoB* gene allows rapid monitoring of the transcontinental spread of the *B. melitensis* strains. The *rpoB* gene could, therefore, serve as an alternative marker for differentiation and geographical identification of *B. melitensis* strains.

Acknowledgements

We thank the following departments and institutes for the *Brucella* genome sequences that were used in this study: Ministry of Science and Technology of the People's Republic of China (accession number: GCA 000331655), Academy of Military Medical Science, China (accession Number: GCA 000250835, GCA_000250815, GCA 000250795, GCA 000250775), Virginia Bioinformatics Institute (accession number: GCA 000022625, GCA 000018725), Beijing Institute of Disease Control and Prevention (accession number: GCA 000298595, GCA 000292165, GCA 000292085, GCA 000298615, GCA 000292065), Integrated Genomics (accession number GCA 000007125), Chinese Academy of Science (accession number: GCA 000209615, GCA 000209595, GCA 000209575), Chinese Center for Disease Control and Prevention (accession number: GCA 000348645), Chinese Academy of Agricultural Sciences (accession number: GCA 000192725, GCA 000192885), Chinese Agricultural University (accession number: GCA 000227645), Oak Ridge National Laboratory (accession number: GCA 000054005), Madurai Kamaraj University (accession number: GCA_000444515.1), National Institute of Animal Biotechnology (accession number: GCA_000685375.1), The Federal Government of Health Institute (accession number: GCA_000591005.1, GCA_000705095.1, GCA_000705115.1, GCA_000714755.1) and Broad Institute (accession number: GCA 000366865, GCA 000365865, GCA 000370625, GCA 000160295, GCA 000158695, GCA 000182235, GCA 000158735, GCA 000366625, GCA 000366885, GCA 000366905, GCA 000370645, GCA 000366925, GCA 000370665, GCA 000365845, GCA 000366945, GCA 000366965, GCA 000366985, GCA 000367005, GCA 000370685, GCA 000370705, GCA 000370725, GCA 000367045, GCA 000370745, GCA 000366585, GCA 000370765, GCA 000370785, GCA 000370805, GCA 000370825, GCA 000370845, GCA 000370865, GCA_000480215.1, GCA_000479975.1). The research leading to these results received funding from the University of Malaya High Impact Research (HIR) grant scheme through agreement no. E000013-20001. The sequencing was made possible using sequencing infrastructure funded by the Ministry of Science, Technology, and Innovation, Malaysia Genome Institute initiative grant (grant agreement 07-05-MGI-GMB015). Preliminary results were presented as an abstract at the Fifth International Meeting on Emerging Diseases and Surveillance, Vienna, 31 October – 3 November 2014.

References

1. Glynn MK, Lynn TV (2008) Brucellosis. J Am Vet Med Assoc 233: 900-908.
2. Cutler SJ, Whatmore AM, Commander NJ (2005) Brucellosis--new aspects of an old disease. J Appl Microbiol 98: 1270-1281.
3. Corbel MJ (1997) Brucellosis: an overview. Emerg Infect Dis 3: 213-221.
4. Buzgan T, Karahocagil MK, Irmak H, Baran AI, Karsen H, Evirgen O, Akdeniz H (2010) Clinical manifestations and complications in 1028 cases of brucellosis: a retrospective evaluation and review of the literature. Int J Infect Dis 14: e469-e478.
5. Song L, Ahn S, Walt DR (2005) Detecting biological warfare agents. Emerg Infect Dis 11: 1629-1632.
6. Smits HL, Espinosa B, Castillo R, Hall E, Guillen A, Zevaleta M, Gilman RH, Melendez P, Guerra C, Draeger A, Broglia A, Nockler K (2009) MLVA genotyping of human Brucella isolates from Peru. Trans R Soc Trop Med Hyg 103: 399-402.
7. Bricker BJ, Ewalt DR, Halling SM (2003) *Brucella* 'HOOF-Prints': strain typing by multi-locus analysis of variable number tandem repeats (VNTRs). BMC Microbiol 3: 15.
8. Whatmore AM, Perrett LL, MacMillan AP (2007) Characterisation of the genetic diversity of *Brucella* by multilocus sequencing. BMC Microbiol 7: 34.
9. Adekambi T, Drancourt M, Raoult D (2009) The *rpoB* gene as a tool for clinical microbiologists. Trends Microbiol 17: 37-45.
10. Marianelli C, Ciuchini F, Tarantino M, Pasquali P, Adone R (2006) Molecular characterization of the *rpoB* gene in *Brucella* species: new potential molecular markers for genotyping. Microbes Infect 8: 860-865.
11. Sayan M, Yumuk Z, Bilenoglu O, Erdenlig S, Willke A (2009) Genotyping of *Brucella melitensis* by *rpoB* gene analysis and re-evaluation of conventional serotyping method. Jpn J Infect Dis 62: 160-163.
12. Tan KK, Tan YC, Chang LY, Lee KW, Nore SS, Yee WY, Mat Isa MN, Jafar FL, Hoh CC, AbuBakar S (2015) Full genome SNP-based phylogenetic analysis reveals the origin and global spread of Brucella melitensis. BMC Genomics 16: 93.
13. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJ, Yoo HS, Zhang C, Zhang Y, Sobral BW (2014) PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res 42 (Database issue): D581-D591.

14. Sam IC, Karunakaran R, Kamarulzaman A, Ponnampalavanar S, Syed Omar SF, Ng KP, Mohd Yusof MY, Hooi PS, Jafar FL, AbuBakar S (2012) A large exposure to *Brucella melitensis* in a diagnostic laboratory. J Hosp Infect 80: 321-325.
15. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.
16. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61: 539-542.
17. Hunter PR (1990) Reproducibility and indices of discriminatory power of microbial typing methods. J Clin Microbiol 28: 1903-1905.
18. Bikandi J, San Millan R, Rementeria A, Garaizar J (2004) In silico analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction. Bioinformatics 20: 798-799.

Corresponding author
Sazaly AbuBakar
Director of Tropical Infectious Diseases Research and Education Centre (TIDREC)
Level 4, Block N & O
Faculty of Medicine, Jalan University
University of Malaya
50603 Kuala Lumpur, Malaysia
Phone: +603-79675756
Fax: +603-79675757
Email: sazaly@um.edu.my

Conflict of interests: No conflict of interests is declared.