RE# MJVR - 0007-2016

WHOLE GENOME SEQUENCE OF *Brucella melitensis* LOCAL ISOLATE FROM AN INFECTED GOAT IN MALAYSIA

OMER KHAZAAL SALLOU[1,2,4*], RAMLAN BIN MOHAMED[2], MOHD MOKHTAR B. ARSHAD[1], HIRZAHIDA MOHD. PADIL[3], SHUHAILA MAT SHARANI[3], HARDY ABU DAUD[1] AND MOHD AZAM KHAN B. GORIMAN KHAN[1]

1 Faculty of Veterinary Medicine, University Malaysia Kelantan, Karung Berkunci 36, Pengkalan Chepa, 16100 Koto Bharu, Kelantan, Malaysia.

2 Veterinary Research Institute, 59 Jalan Sultan Azlan Shah, 31400 Ipoh, Perak Darul Ridzuan, Malaysia.

3 Malaysia Genome Institute, Ministry of Science Technology And Innovation, Jalan Bangi, 43000 Kajang, Selangor Darul Ehsan, Malaysia.

4 Department of Internal and Preventive Medicine, College of Veterinary Medicine, University of Mosul, Iraq

* Corresponding author: khazaalvet79@yahoo.com

ABSTRACT. Brucellosis in goats is mainly caused by the bacterium *Brucella melitensis,* which is one of the most important pathogenic species in the world. In Malaysia, the annual prevalence data of brucellosis was recorded in goats and the control strategy of the disease based on test and cull of infected animals. This strategy has caused huge economic losses to farmers and government alike. Therefore, whole genome sequencing of *B. melitensis* local strain is essential for improving the current vaccine.

B. melitensis strain VRI 6530/11 was obtained from veterinary research institute biobank, Ipoh. The strain was submitted for classical identification procedures and the total genomic DNA was extracted by using DNeasy blood and tissue kit (QIAGEN). The concentration and purity of DNA were determined by using agarose gel electrophoresis and spectrophotometer (DNA/RNA) assay respectively. The genome was sequenced by using Illumina HiSeq platform with insert size ~200 bp.

A total of 1.0 Gb data was generated from the sample. More than 95% of sequencing data was retained in the sample after quality filtering, this indicate the sequencing reads are of high quality. Final assembly had 33 scaffolds with total size ~3.28 Mb, 44 contigs, GC content is 57.25%, N50 is 293,291. A total of 3,238 protein coding genes, 48 tRNAs and 3 rRNAs were predicted and over 87% of the genes were functionally annotated.

Genome sequencing of a local *B. melitensis* strain is the first of its kind in Malaysia and work from this study can contribute towards the development of a new effective vaccine for the control of the disease in the country.

Keywords: Brucella melitensis, genome sequencing, zoonotic

INTRODUCTION

Brucella melitensis is a Gram-negative and facultative intracellular pathogen, which can infect human beings and various species of animals. It is responsible for ovine and caprine brucellosis (Alton *et al*., 1988; Manish *et al*., 2013). In some countries, brucellosis is considered a major endemic zoonotic disease with high prevalence rates among sheep and goats (Shang *et al*., 2002; Mantur and Amarnath, 2008). The organism was first isolated in Malaysia from an ovine brucellosis outbreak in Johor in 1994 (Moktar *et al*., 1995). Not much work was carried out to map the prevalence of the disease among ruminants between that period and the current study period. In a more recent brucellosis surveillance of goats in Malaysia in 2014, Pulau Pinang recorded the highest reactor rate (2.18%) compared to other states. In that surveillance, four *B. melitensis* strain were isolated from 300 vaginal samples taken from goats (Al-Garadi *et al*., 2011; Omer *et al*., 2014). In Malaysia, infected animals are culled as a control strategy of the disease. This strategy has caused huge economic losses to farmers and government alike (Bahaman *et al*., 2007; Bamaiyi *et al*., 2012; Bamaiyi *et al*., 2015). Vaccination policies using the Rev. 1 vaccine, is currently the only approved and appropriate method of controlling the disease in sheep and goats (Elberg, 1981; Al Khalaf *et al*., 1992). It is a live attenuated vaccine derived from a virulent *B. melitensis* strain. This vaccine however had shown a considerable degree of virulence and it has been reported to induce abortion when administered during pregnancy (Blasco and Díaz, 1993; Banai, 2002).

Designing a new vaccine against *B. melitensis* requires an understanding of the mechanism that is used by the bacterium in causing the disease. The genome sequencing and bioinformatics analysis of genome data provide novel information about the bacterium such as, identification of novel virulence genes, novel metabolic pathways and novel cell wall proteins that represent new targets for vaccine development (Ribeiro *et al*., 2012).

In the present study, we analyzed the complete genome sequence of local *B. melitensis* strain to help in the identification of suitable antigens for the development of a new vaccine, which can be used to control the disease in small ruminants in Malaysia.

MATERIALS AND METHODS

A local *B. melitensis* strain (VRI 6530/11) was obtained from the Veterinary Research Institute in Ipoh, Malaysia. The strain was submitted for classical identification procedures. Total genomic DNA was extracted and purified using DNeasy blood and tissue kit (QIAGEN). A spectrophotometer (DNA/RNA) assay was used to determine DNA concentration and purity. Agarose gel electrophoresis was used to separate DNA for analysis. The genome was sequenced using the illumina

HiSeq platform supporting a read length of 200 bp. Quality of the raw sequencing reads was assessed by FastQC v0.10.1. and low quality reads, bases (lesser than Qv20), ambiguous bases (Ns), and artifacts were removed before assembly by using Fastx toolkit v0.0.13.2. Velvet assembler v1.2.10 was used for *de novo* assembly of the reads. Empirical test with different values of k-mer (hash length) was used to choose the k-size. Contigs from reads assembly were scaffold by SSPACE v2.0. The gaps were filled by Gapfiller v1.10. Genes content of the genome were predicted by using GeneMarks v4.10d. The predicted genes were functionally annotated by Kyoto Encyclopedia of genes and genomes (KEGG), non-redundant (NR) and Swiss-Prot (SP) protein databases at maximum expected value (E-value) threshold of $1e^{-5}$. Structural RNA was identified by using tRNAscan-SE v1.3.1 and rRNAmmer v1.2. Potential virulence genes were identified by blast search against PHI-base and VFDB databases. The genomic circular map was drawn by using Circos v0.64.

RESULTS

Quality assessment of raw and clean reads

Quality assessment is always performing before analyzing the sequence for drawing biological conclusions and produce better mapping results. This assessment includes: per base sequence quality, per sequence quality scores, per base sequence content and sequence length distribution. Raw reads produced from sequencing machine contain dirty reads (reads are contain adapters or unknown and low quality bases). These reads will affect negatively, the bioinformatic analysis of genome sequencing data. Fastx toolkit v0.0.13.2 was used for filtering and trimming these reads from the sequencing data. The results of quality assessment of raw data and filtering reads of *B. melitensis* (VRI 6530/11) genome revealed that over 95% of sequencing data was retained after the quality filtering process in both phases of Illumina sequencing platform; paired-end sequencing (PE1) and paired PCR enrichment (PE2). This indicates the sequencing reads are of very good quality. The results of raw read (pre-filter) and clean read (post-filter) are depicted in Table 1.

De novo whole genome assembly

The randomized subset of clean reads with 200× coverage of estimated genome size 3.28 Mb was used for de novo assembly. An empirical test with different values of hash length was used to select best assembly based on the assembly statistics. The assembly from hash-57 was selected as depicted in Figure 1 and the results of final assembly are listed in Table 2.

Gene prediction and annotation

A total of 3,238 protein coding genes were predicted by GeneMarks. The annotation of

Table 1. Sequencing statistics of raw reads (pre-filter) and clean reads (post-filter)

	Reads Number of reads (%)	Bases Total bases (%)
Raw Reads	PE1 5,490,328	549,032,800
	PE2 5,490,328	549,032,800
	Total 10,980,656 100%	1,098,065,600 100%
Clean Reads	PE1 5,453,428	523,273,827
	PE2 5,453,428	526,407,591
	Total 10,906,856 99.3%	1,049,681,418 95.6%

Table 2. Final assembly statistics of *B. melitensis* (VRI 6530/11) genome

Basic statistics	***B. melitensis* VRI 6530/11 genome**
Number of scaffolds (>=1000bp)	33
Number of contigs	44
Largest scaffolds	609,228
N50	293,291
GC (%)	57.25
Number of Ns	653
Total length	3,283,458

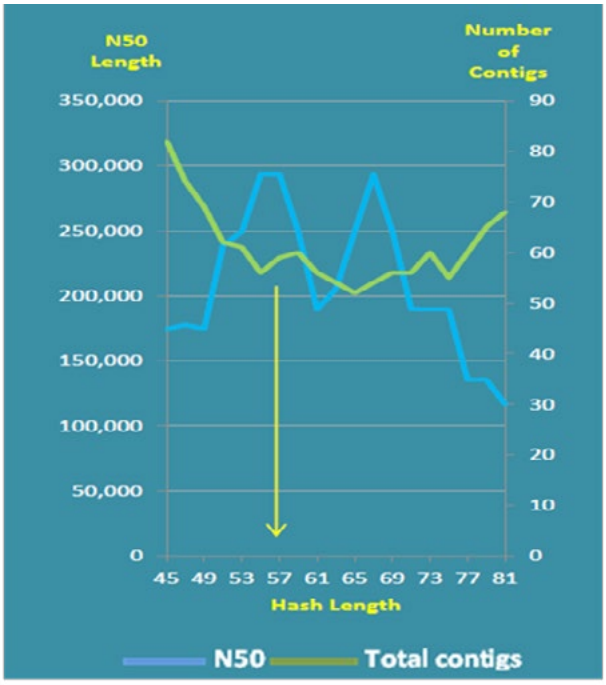

Figure 1. Empirical test with different values of hash lengths.

putative protein encoding genes was shown 2,842 genes have functional assignments, 396 hypothetical proteins, 1,210 Proteins with EC number assignments and 2,499 Proteins with GO assignments. As well as, 48 tRNAs and 3 rRNAs were identified in the genome of *B. melitensis* (VRI 6530/11). The results of gene prediction and annotation are listed in Table 3.

Identification of plasmid sequence

Scaffolds of genome assembly were used as query in blastn search against the plasmid database. No significant plasmid hit was observed from the results at cut-off values of (1) E-value ≤ $1e^{-5}$, (2) identity ≥ 50%, (3) query coverage ≥ 50% and (4) subject coverage ≥ 50%.

Identification of phage or prophage sequence

One intact and one incomplete prophage regions have been identified by phage search tool (PHAST) for assembly. The details of the prophage region are listed in Table 4 and depicted in Figure 2, 3 and 4.

Table 3. Gene prediction and annotation statistics of whole genome sequence of *B. melitensis* VRI 6530/11

Gene prediction and annotation	*B. melitensis* VRI 6530/11 genome
Number of predicted protein-coding gene (≥ 33 aa)	3,238
Hypothetical proteins*	396
Proteins with functional assignments	2,842
Protein with EC number assignments	1,210
Protein with GO assignments	2,499
tRNA	48
rRNA	3

*Include proteins with keywords unknown, hypothetical, predicted, unnamed and uncharacterized.

Table 4. The details of prophage regions have been identified in whole genome sequence of *B. melitensis* VRI 6530/11

	Region	Region length	Completeness	No. of CDS	Region position	GC (%)
***B. melitensis* VRI 6530/11**	1	22.6 kb	incomplete	14	850667-873319	58.15%
	2	13.7 Kb	intact	18	2693223-2706963	61.19%

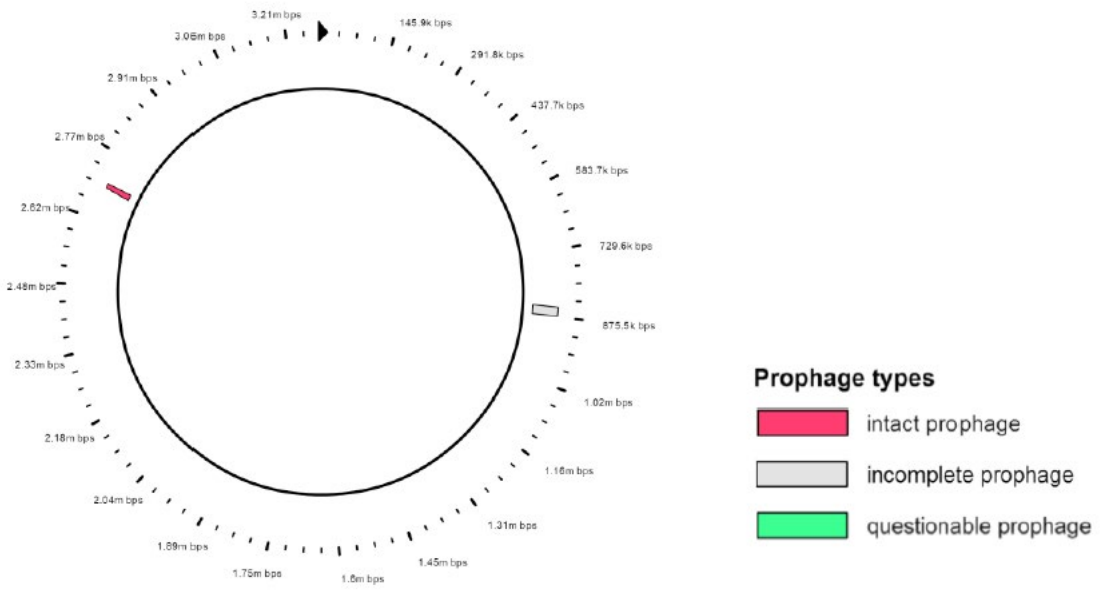

Figure 2. Prophage types have been identified in the genome of *B. melitensis* VRI 6530/11

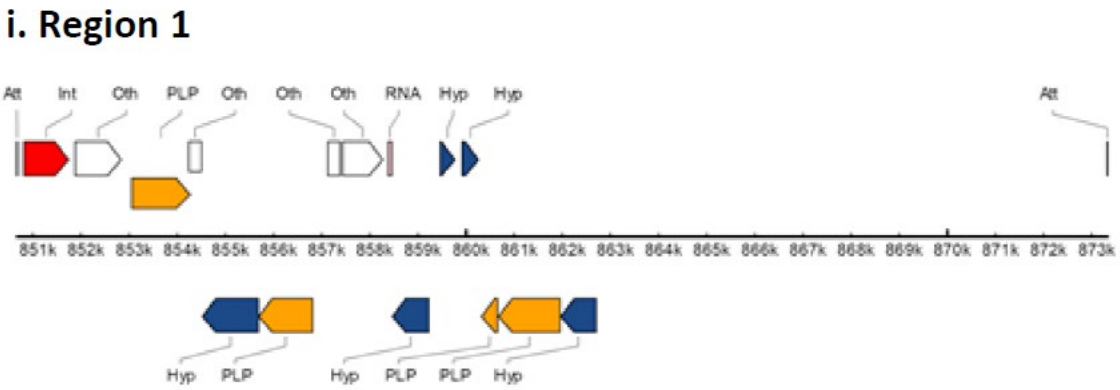

Figure 3. Prophage region 1 have been identified in the genome of *B.melitensis* VRI 6530/11

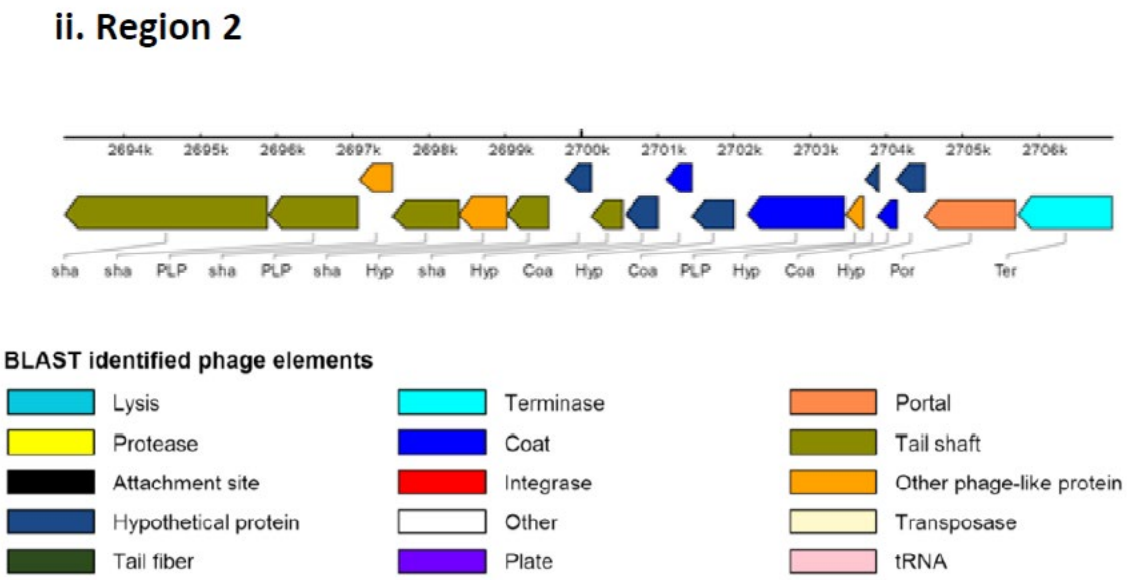

Figure 4. Prophage region 2 have been identified in the genome of *B. melitensis* VRI 6530/11

Gene ontology

The gene ontology (GO) software level 2 was used to provide a system of hierarchically classifying genes or gene products to terms organized in a graph structure. The terms of GO included three categories: (1) molecular function (describing the molecular activity of a gene), (2) biological process (describing the larger cellular or physiological role carried out by the gene, coordinated with other genes) and (3) cellular process (describing the location of gene in the cell where the gene product executes its function) as depicted in Figure 5.

Genomic circular map

The whole genome sequencing data of *B. melitensis* VRI 6530/11 was visualized as a circular map by using Clico software. The bars of the map are marked from outmost to the inner ones. The outmost bar is considering 33 scaffolds in green color. Subsequently, two regions of prophage sequence are in red color. The forward and reverse directions of putative genes are in blue and purple color respectively. 48 tRNA and 3 rRNA are in dark blue and dark red color respectively. The GC plot is in green color and the GC skew is in orange and black color as depicted in Figure 6.

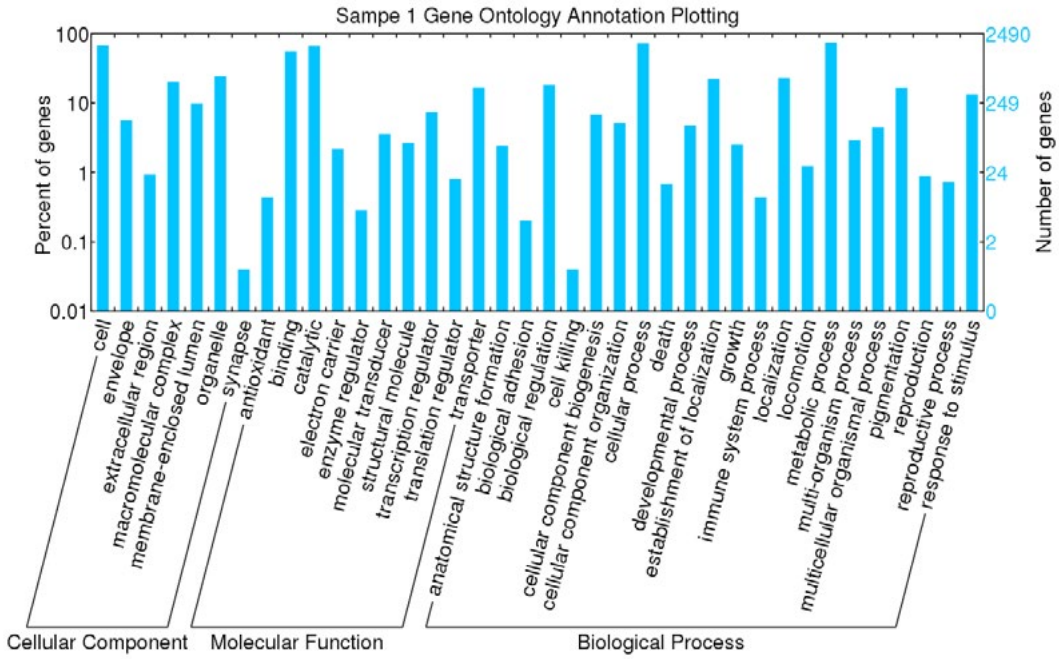

Figure 5. Gene ontology of *B. melitensis* VRI 6530/11 genome

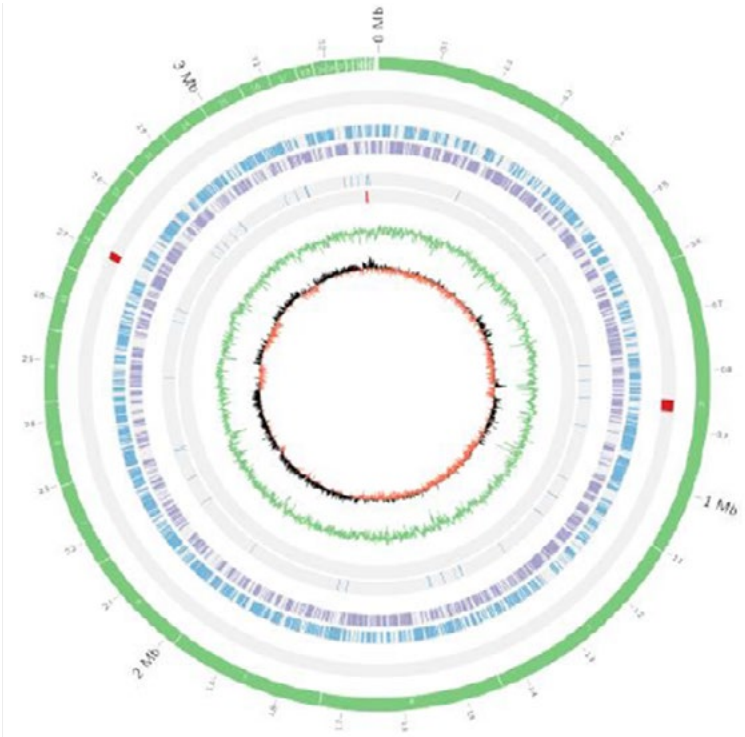

Figure 6. The whole genome sequencing data of *B. melitensis* VRI 6530/11 was visualized as circular map. The bars marked from outmost circle to the inner ones corresponding to scaffolds (green), Prophage sequence (red), putative genes in forward (blue) and reverse (purple) direction, tRNA (dark blue), rRNA (dark red), GC plot (green) and GC skew (orange and black).

DISCUSSION

The whole genome sequence of *B. melitensis* has provided significant information on the basic nature of the organism and identification of genes, which are involved in immune evasion, intracellularsurvival, regulation and secretion system.

Regarding the genome structure of the *B. melitensis* local isolate, it was found to have 3,283,458 bases, two chromosomes (chromosome I larger than chromosome II), GC content is 57.25%, 3,238 protein coding genes, and no significant plasmid hit was observed from the results at cut-off values of (1) E-value $\leq 1e^{-5}$, (2) identity $\geq 50\%$, (3) query coverage $\geq 50\%$ and (4) subject coverage $\geq 50\%$. These findings agree with the DNA structure of *B.melitensis* reviewed by DelVecchio *et al*., (2002) and Halling *et al*., (2005). They found that the members of Brucella genus share a similar GC content of 55-58% and it has no further chromosomal duplicating entities such as the plasmids. Furthermore, the genome sequencing of *B. melitensis* strain 16M has confirmed the existence of two chromosomes with estimated size 3.27Mb and the chromosome I being larger than chromosome II with a median length of 2.1 and 1.2 Mb respectively (Michaux *et al*., 1993; Jumas-Bilak *et al*., 1998).

Recently, the genome sequence of *B. melitensis* strain Bm IND1 from India was found with 3,284,360 bases, 3,360 protein-coding genes, and GC content is 57.2% (Sashi *et al*., 2014). These data agree approximately with what this study has obtained about the genome size and the GC content, but disagree about the number of protein coding genes. The difference within the Brucella genome is common due to finding deletion events, more frequently than insertions (Gao *et al*., 2012). The difference could be used as applicant for molecular epidemiology studies among Brucella strains. Furthermore, the differences between Brucella strains may be explained by the bacterial adaptation to host and environment (Atluri *et al*., 2011). In genome sequencing of numerous *B. melitensis* strains around the world and from different hosts, it is very important to find the conserved or different genetic materials among the strains.

Characterizing the genome sequence of a local *B. melitensis* strain is the first of its kind in Malaysia, and work from this study can contribute towards the development of diagnostic tests or vaccine to avoid the epidemiological, public health and economic complications caused by the disease.

REFERENCES

1. Alton G.G., Jones L.M., Angus R.D. and Verger M. (1988). *Techniques for the brucellosis laboratory*. Paris: INRA.
2. Manish K., Chand P., Rajesh C., Teena R. and Sunil K. (2013). Brucellosis: an updated review of the disease. *Indian Journal of Animal Sciences.* **83(1):** 3–16.
3. Shang D.Q., Xiao D.L. and Yin J.M. (2002). Epidemiology and control of brucellosis in China. *Vet Microbiol*. **90:**165–182.
4. Mantur B.G. and Amarnath S.K. (2008). Brucellosis in India a review. *Journal of Biosciences.* **33(4):** 539–547.

5. Moktar A.M., Sivanandan M., Shafri A.M. and Norlida A.B. (1995). Case report of isolation of *Brucella melitensis* from sheep in Malaysia. In: *Proceedings of the 7th Veterinary Association Malaysia Scientific Congress* in Seremban, Malaysia. 65-67.
6. Al-Garadi M.A., Khairani-Bejo S., Zunita Z. and Omar A.R. (2011). Isolation and Identification of *Brucella melitensis* in Goats. *Journal of Animal and Veterinary Advances.* **10(8):** 972-979.
7. Omer K.S., Mazlan L., Norazura A., Chin S.W., Nurulfatiha A.S., Lily Suhaida M.S., Yeoh N.N., Magendren S., Mohammad M.A., Ramlan M., Mohd Azam K.G., Mokhtar A., Kalthum H., Hardy D. and Mohd Riduan M.H. (2014). Surveillance Of Brucellosis In Small Ruminants In Malaysia During 2003-2013. In: Proceedings of 26th Veterinary Association Malaysia Congress, 1 June 2014. *Malaysian Journal Of Veterinary Research.* **5(suppl 1)**: 92-94. .
8. Bahaman A.R., Joseph P.G. and Siti-Khairani B. (2007). A review of the epidemiology and control of Brucellosis in Malaysia. *J Vet Malaysia.* **19(1):** 1-6.
9. Bamaiyi P.H., Abd-Razak N.S. and Zainal M.A. (2012). Seroprevalence and economic impact of eradicating zoonotic brucellosis in Malaysia: Acase study of Melaka state of Malaysia. *Veterinary World.* **5(7):** 398-404.
10. Bamaiyi P., Khairani-Bejo S. and Zainal Abidin M. (2015). The economic impact attributable to brucellosis among goat farms in Peninsula Malaysia and cost benefit analysis. *Research Opinions in Animal & Veterinary Sciences.* **5(2):** 57-64.
11. Elberg S.S. (1981). Rev 1 *Brucella melitensis* vaccine. Part II: 1968–1980. *Vet Bull.* **51:** 67–73.
12. Al Khalaf S.A., Mohamad B.T. and Nicoletti P. (1992). Control of brucellosis in Kuwait by vaccination of cattle, sheep and goats with *Brucella abortus* strain 19 or *Brucella melitensis* strain Rev 1. *Trop Anim Hlth Prod.* **24:** 45–9.
13. Blasco J.M. and Díaz R. (1993). *Brucella melitensis* Rev-1 vaccine as a cause of human brucellosis. *Lancet.* **342 (8874):** 805.
14. Banai M. (2002). Control of small ruminant brucellosis by use of *Brucella melitensis* Rev.1 vaccine: laboratory aspects and field observations. *Vet Microbiol.* **90(1-4):** 497–519.
15. Ribeiro F.J., Przybylski D., Yin S., Sharpe T., Gnerre S., Abouelleil A., Berlin A.M., Montmayeur A., Shea T.P. and Walker B.J. (2012). Finished bacterial genomes from shotgun sequence data. *Genome Res.* **22:** 2270–2277.
16. DelVecchio V.G., Kapatral V. and Redkar R.J. (2002). The genome sequence of the facultative intracellular pathogen *Brucella melitensis*. *Proc Natl Acad Sci USA.* **99:** 443-8.
17. Halling S.M., Peterson-Burch B.D., Bricker B.J., Zuerner R.L., Qing Z., Li L.L., Kapur V., Alt D.P. and Olsen S.C. (2005). Completion of the genome sequence of *Brucella abortus* and comparison to the highly similar genomes of *Brucella melitensis* and *Brucella suis*. *J. Bacteriol.* **187:** 2715–2726.
18. Michaux S., Paillisson J. and Carles-Nurit M.J. (1993). Presence of two independent chromosomes in the *Brucella melitensis* 16M genome. *J Bacteriol.* **175:** 701-5.
19. Jumas-Bilak E., Charachon M.S. and Bourg G. (1998). Differences in chromosome number and genome rearrangements in the genus *Brucella*. *Mol Microbiol.* **27:** 99-106.
20. Sashi B.R., Vivek K.G., Mukesh K., Nagendra R.H., Gary A.S., Pallu R. and Girish K.R. (2014). Draft Genome Sequence of the Field Isolate *Brucella melitensis* Strain Bm IND1 from India. *Genome Announcements.* **2(3):** 497-14.
21. Gao G., Li J., Li T., Zhang Z., Wang L. and Yuan X. (2012). Complete genome sequence of *Brucella canis* strain 118, a strain isolated from canine. *J. Bacteriol.* **194:** 6680.
22. Atluri V.L., Xavier M.N., De Jong M.F. and Den Hartigh A.B. (2011). Interactions of the human pathogenic *Brucella* species with their hosts. *Annu. Rev. Microbiol.* **65:** 523–541.

ACKNOWLEDGMENT. The authors thank the Department of Veterinary Services and the Veterinary Research Institute, Ipoh for sharing and providing the data and facilities used in this study.