

RESEARCH ARTICLE

# Whole genome sequencing of *Brucella melitensis* isolated from 57 patients in Germany reveals high diversity in strains from Middle East

Enrico Georgi<sup>1</sup>, Mathias C. Walter<sup>1</sup>, Marie-Theres Pfalzgraf<sup>1</sup>, Bernd H. Northoff<sup>2</sup>, Lesca M. Holdt<sup>2</sup>, Holger C. Scholz<sup>1,3</sup>, Lothar Zoeller<sup>1,3</sup>, Sabine Zange<sup>1,3</sup>, Markus H. Antwerpen<sup>1</sup>✉\*

**1** Bundeswehr Institute of Microbiology, Munich, Germany, **2** Institute of Laboratory Medicine, Ludwig-Maximilians University, Munich, Germany, **3** National Consultant Laboratory for *Brucella*, Munich, Germany

✉ These authors contributed equally to this work.

\* [markusantwerpen@bundeswehr.org](mailto:markusantwerpen@bundeswehr.org)



**OPEN ACCESS**

**Citation:** Georgi E, Walter MC, Pfalzgraf M-T, Northoff BH, Holdt LM, Scholz HC, et al. (2017) Whole genome sequencing of *Brucella melitensis* isolated from 57 patients in Germany reveals high diversity in strains from Middle East. PLoS ONE 12 (4): e0175425. <https://doi.org/10.1371/journal.pone.0175425>

**Editor:** Roy Martin Roop, II, East Carolina University Brody School of Medicine, UNITED STATES

**Received:** December 29, 2016

**Accepted:** March 24, 2017

**Published:** April 7, 2017

**Copyright:** © 2017 Georgi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data are uploaded to the NCBI Bioproject database (Bioproject PRJNA347914) together with the sequenced reads.

**Funding:** This study was supported by internal funding of the German Ministry of Defence Medical Service as part of project no. 57-2014-03. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Brucellosis, a worldwide common bacterial zoonotic disease, has become quite rare in Northern and Western Europe. However, since 2014 a significant increase of imported infections caused by *Brucella (B.) melitensis* has been noticed in Germany. Patients predominantly originated from Middle East including Turkey and Syria. These circumstances afforded an opportunity to gain insights into the population structure of *Brucella* strains. *Brucella*-isolates from 57 patients were recovered between January 2014 and June 2016 with culture confirmed brucellosis by the National Consultant Laboratory for *Brucella*. Their whole genome sequences were generated using the Illumina MiSeq platform. A whole genome-based SNP typing assay was developed in order to resolve geographically attributed genetic clusters. Results were compared to MLVA typing results, the current gold-standard of *Brucella* typing. In addition, sequences were examined for possible genetic variation within target regions of molecular diagnostic assays. Phylogenetic analyses revealed spatial clustering and distinguished strains from different patients in either case, whereas multiple isolates from a single patient or technical replicates showed identical SNP and MLVA profiles. By including WGS data from the NCBI database, five major genotypes were identified. Notably, strains originating from Turkey showed a high diversity and grouped into seven subclusters of genotype II. MLVA analysis congruently clustered all isolates and predominantly matched the East Mediterranean genetic clade. This study confirms whole-genome based SNP-analysis as a powerful tool for accurate typing of *B. melitensis*. Furthermore it allows special allocation and therefore provides useful information on the geographic origin for trace-back analysis. However, the lack of reliable meta-data in public databases often prevents a resolution below geographic regions or country levels and corresponding precise trace-back analysis. Once this obstacle is resolved, WGS-derived bacterial typing adds an important method to complement epidemiological surveys during outbreak investigations. This is the first report of a detailed genetic

**Competing interests:** The authors have declared that no competing interests exist.

investigation of an extensive collection of *B. melitensis* strains isolated from human cases in Germany.

## Introduction

On a global scale human brucellosis is one of the most common bacterial zoonotic diseases [1]. However, its occurrence greatly differs between geographic areas throughout the world. The disease has its highest incidence and prevalence in countries of the Mediterranean basin, Middle East, some parts of Central and South America, Africa and Asia [2]. The highest annual incidences of human brucellosis per million of the population are observed in Syria and in Mongolia [3]. In contrast, countries of Northern and Western Europe are considered free of autochthonous human brucellosis. However, each year a relatively small number of cases are reported in Germany, with most of them having a history of travelling to or immigrating from endemic regions like the Mediterranean basin. In total, 267 cases of human brucellosis with a median number of 24 cases per annum were reported during the last decade (2005–2015) to the national surveillance system at the Robert Koch Institute in Berlin (SurvStat@RKI 2.0, <https://survstat.rki.de>, data as of 09/05/2016). Notably, within the last two years an increase of cases was observed. In 2014 and 2015, the number of annual cases nearly doubled to 47 and 44, respectively.

The most important route of transmission is the ingestion of contaminated raw milk and other unpasteurized dairy products, but the disease may also be acquired by handling infected animals, animal discharges, or cultures of the pathogen [4]. Human-to-human transmissions are rare events. Clinical manifestations of brucellosis are versatile and comprise acute systemic infections with undulating fever as cardinal symptom as well as localized inflammatory processes and site-specific manifestations in chronically infected patients. A 6-week regimen comprising rifampin and doxycycline is commonly applied in Germany to eradicate the pathogen [5]. Patients with complications often need longer treatment and additional antimicrobial substances like aminoglycosides [6].

Human brucellosis can be caused by various *Brucella* species. The genus currently comprises 12 validly published species which are genetically highly related to each other [7,8]. *Brucella melitensis* is by far the most frequently observed causative agent of human infection. Based on DNA-DNA hybridization studies that revealed DNA-DNA homologies of >80%, all classical species might be attributed to a single species with several biovars [9]. However, due to historic reasons, species-specific predilections for particular animal hosts, biochemical features etc. the Subcommittee on the taxonomy of *Brucella* agreed in 2003 on a return to the pre-1986 taxonomic opinion with a six species concept (*B. abortus*, *B. canis*, *B. melitensis*, *B. neotomae*, *B. ovis*, *B. suis*) [10].

Multiple Locus Variable Number Tandem-Repeat (VNTR) Analysis (MLVA) has become a major molecular typing method to characterize several pathogenic bacterial species. *Brucella* MLVA-16 scheme has been proven to be a valuable tool in epidemiological trace-back investigations with high discriminatory power in several studies [11–17]. Recently, a full genome SNP-based phylogenetic analysis was published and could provide another powerful tool for intraspecies discrimination of *Brucella* [18].

Increasing numbers of brucellosis cases in Germany in 2014 and 2015 triggered us to investigate the geographic origin of strains isolated in Germany. Using the National Consultant Lab's collection of clinical isolates of *B. melitensis* originating from 57 patients diagnosed as

brucellosis cases in Germany since 2014, our study was designed (i) to trace back the strains to their geographic origin; (ii) to compare the resolution of WGS-based SNP-typing with the standard MLVA-16 genotyping; (iii) to determine genetic variation within the genomes of the strains that could affect molecular diagnostic or typing assays using next-generation sequencing technology.

## Material and methods

### Strains

All available strains from culture-positive cases of human brucellosis that had been processed by the National Consultant Laboratory for *Brucella* and that had been identified as *B. melitensis* by real-time and conventional PCR methods between January 2014 and June 2016 (n = 57) were subjected to whole genome sequencing (WGS). Patient histories were evaluated for travel to or residence in countries known to be endemic for brucellosis. To enlarge the covered region of geographic origin as well as to consider genomic variation over time, six strains of cases from previous years were also added to the strain collection. To assess the reproducibility of applied methods, biological (i.e. one patient, but different isolates) and technical (i.e. one isolate, but different sequence libraries) replicates were included. Overall, isolates from 63 individual patients from 17 countries were analysed (see [Table 1](#) and [S1 Table](#)).

### Whole genome sequencing

Strains were re-cultivated from MicroBank™ vials by standard procedures. Pure colonies were picked and inactivated using an ethanol-based in-house protocol. Genomic DNA was isolated

**Table 1. Summary of samples grouped by country of origin.**

Country	Samples*
Afghanistan	4
Albania	1
Algeria	1
Bulgaria	1
India	1
Iran	2
Iraq	6
Italy	2
Jordan	1
Kuwait	1
Morocco	1
Romania	1
Saudi Arabia	1
Somalia	2
Syria	16
Turkey	27
Turkmenistan	1
unknown	5

\*This study comprises 74 samples (including biological and technical replicates) from 63 individual patients with reference to 17 countries. Genomes were uploaded to NCBI Bioproject database (Bioproject PRJNA347914).

<https://doi.org/10.1371/journal.pone.0175425.t001>

according to manufacturer's tissue protocol of QIAamp DNA Mini Kit (Qiagen). DNA concentration was determined by use of a Qubit<sup>®</sup> 2.0 Fluorometer (Thermo Fisher Scientific) and the Qubit<sup>®</sup> dsDNA high sensitivity assay kit (Thermo Fisher Scientific). Nextera<sup>®</sup> XT DNA Library Preparation kit (Illumina) with an input DNA amount of 1 to 3 ng was used for library preparation. Finally, whole genome sequencing was performed on a MiSeq instrument (Illumina) with corresponding MiSeq Reagent Kit v3 (600 cycle) chemistry.

## Post-sequencing data processing

The sequenced paired-end reads were assembled with SPAdes v3.9 [19]. Even with DNA fragment sizes of about 550 bp, short reads will not assemble into a single contig for each of the two chromosomes, as the *B. melitensis* genome contains rRNA operons as well as insertion sequence elements. For identification of a reference sequence with the most similar nucleotide identity we randomly picked five assemblies of our sequenced strains and compared them with all nine already available complete genomes (NCBI) using blastn. *B. melitensis* M5-90 (GenBank accessions CP001851 and CP001852) showed the highest similarity rather than the commonly used reference strain *B. melitensis* 16M.

Hence, resulting contigs were aligned against *B. melitensis* M5-90 using blastn [20] to determine their order. Contigs with multiple coverage of the average read coverage were duplicated and inserted between the appropriate contigs to get a valid and continuous mapping to the reference sequence. Afterwards, an in-house developed Python script was used to concatenate and circularize the ordered contigs into the two chromosomes by resolving overlaps or creating gaps filled with Ns of a length estimated from the mapping. BWA [21] was subsequently used to map the raw reads to the chromosomal contigs. Finally, Pilon [22] was applied to undo mis-assemblies, polish the sequence by correcting SNPs and small InDels, as well as to close gaps. In some cases, several iterations of BWA and Pilon were necessary to close longer gaps or gaps newly introduced by undoing the mis-assemblies. The polished contigs were manually checked after each cycle, corrected if necessary and uploaded to the NCBI Bioproject database (Bioproject [PRJNA347914](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA347914)) together with the sequenced reads. Annotation was automatically performed by the NCBI Prokaryotic Genome Annotation Pipeline [23].

## Whole genome SNP analysis

For a rapid whole genome alignment of all sequenced strains, all publicly available completed *B. melitensis* genomes and some incomplete genome assemblies from selected strains against the reference genome *B. melitensis* M5-90, we used Parsnp (parameters -c -e -u -C 1000) from the Harvest suite [24]. The HarvestTools from the same suite were used to extract the SNP positions as a VCF file (parameter -V, [24]). These positions were filtered to exclude clusters of SNPs as well as SNPs located in annotated repeat regions or rRNAs as well as predicted tandem repeats (using [25] with default parameters) by creating a BED file and using it as input to the HarvestTools to extract concatenated artificial SNP sequences as a multiple FASTA file (parameters -b <BED file> -S). This FASTA file was then analyzed with the R Analysis package ape [26]. A maximum parsimony tree was created based on the Kimura nucleotide distance [27], ACCTRAN, UPGMA clustering and 1,000 bootstrap replicates using PAUP\* [28]. The R Analysis package phangorn [29] was used to determine homoplastic SNPs. The VCF file was further analysed using snpEff [30] to predict and annotate the coding effects of the SNPs found (see [S1 File](#)).

## In silico PCR

Following sequence assembly, several in silico PCRs were performed using an in-house Python script to confirm species identification as well as to look out for SNPs in genomic regions

usually targeted by molecular assays. In short, *bcp31*, a single-copy locus for identification, *IS711*, a multi-copy locus for identification, *recA* [31], a target to discriminate closely related *Ochrobactrum* from *Brucella*, BruceLadder [32], a multiplex PCR with specific band patterns after gel electrophoresis to distinguish *Brucella* species from each other, as well as *rpoB* [33], a hotspot region for rifampin resistance, were included.

### In silico MLVA

All sequences also underwent in silico MLVA with 16 loci (MLVA-16, [11,12]), a recognized tool to perform epidemiological trace-back investigations of *B. melitensis*. As determining repeat numbers from WGS data may be error-prone, we carefully checked each locus in respect of expected total length, internal repeat homogeneity or probability to get collapsed VNTRs during the assembly. The analysis was performed by an in-house Python script and all resulting MLVA-16 genotypes were compared to a public database with 2,215 entries of *B. melitensis* strains that can be assessed online (<http://microbesgenotyping.i2bc.paris-saclay.fr/>, [34]).

## Results

### Whole genome assemblies

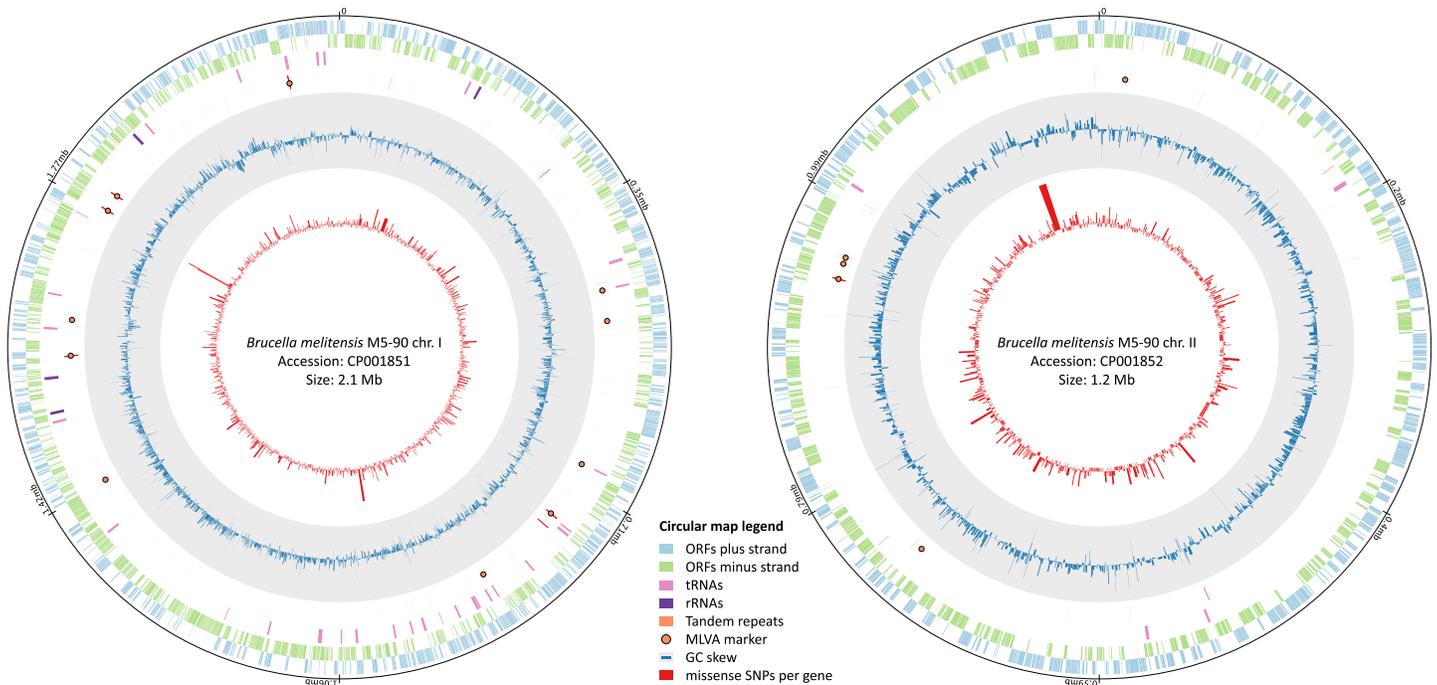
The median MiSeq library output was 2x 906,000 reads (ranging from 290,252 to 2,126,996) resulting in an average coverage of 75-fold (27–159). De novo assembly using SPAdes ended in about 21 initial contigs (14–31) for each strain. Ordered scaffolding and polishing using two iterations of Pilon resulted in continuous chromosomes with an average of three gaps mostly located in tandem repeat or homopolymer regions summing up to 200 missing bases. None of the gaps were located in MLVA marker regions. The genome of five strains resulted in completely resolved final assemblies.

### Alignment and SNP analysis

The multiple whole genome alignment of all samples together with 6 complete genomes and 14 genome assemblies from the NCBI database covered 95% of the reference genome *B. melitensis* M5-90 and can be assumed to represent the core genome. Choosing strain *B. melitensis* 16M (bv1) as reference did not increase the core genome and did not affect the phylogenetic tree. We found significant differences between the two available genome assemblies (GCA\_000007125 [Sanger + Primer walking] and GCA\_000740415 [Illumina + 454]) of 16M and therefore excluded the assembly based on Sanger sequencing.

The SNP analysis revealed 9,525 filtered core SNPs whereof 5,964 are located on chromosome 1 and 3,561 are on chromosome 2 (excluding those only present in the outgroup, see Fig 1). Out of the 7,752 intragenic SNPs, 2,942 were synonymous mutations, 88 SNPs introduced a premature stop and 4,722 led to amino acid changes (see S1 File). Technical as well as inpatient replicates were identical according to the core genome SNPs.

Phylogenetic analysis based on all whole-genome SNPs revealed spatial clustering (Fig 2). Only 42 SNPs are homoplasious; hence, the consistency index is 0.4%. Five main lineages are clearly distinguished according to Tan et al [18]: Genotype I comprising strains from the Western Mediterranean Region and Egypt, Genotype II with strains from regions of the Eastern Mediterranean Region and the Middle and Far East, as well as Genotype III with strains from the African continent. Genotypes IV and V are assigned to strains from Malta, Portugal and the American continent. Furthermore, several clades and subclades can be discriminated within the lineages and are associated with geographic attributions (see Figs 2 and 3).



**Fig 1. Circular view of the two chromosomes with SNP frequencies and MLVA loci.** 5,964 and 3,561 SNPs on chromosome 1 and 2 used in the subsequent analysis are scattered throughout the genome and located, respectively. In contrast, MLVA loci cover only a minor portion of the genome.

<https://doi.org/10.1371/journal.pone.0175425.g001>

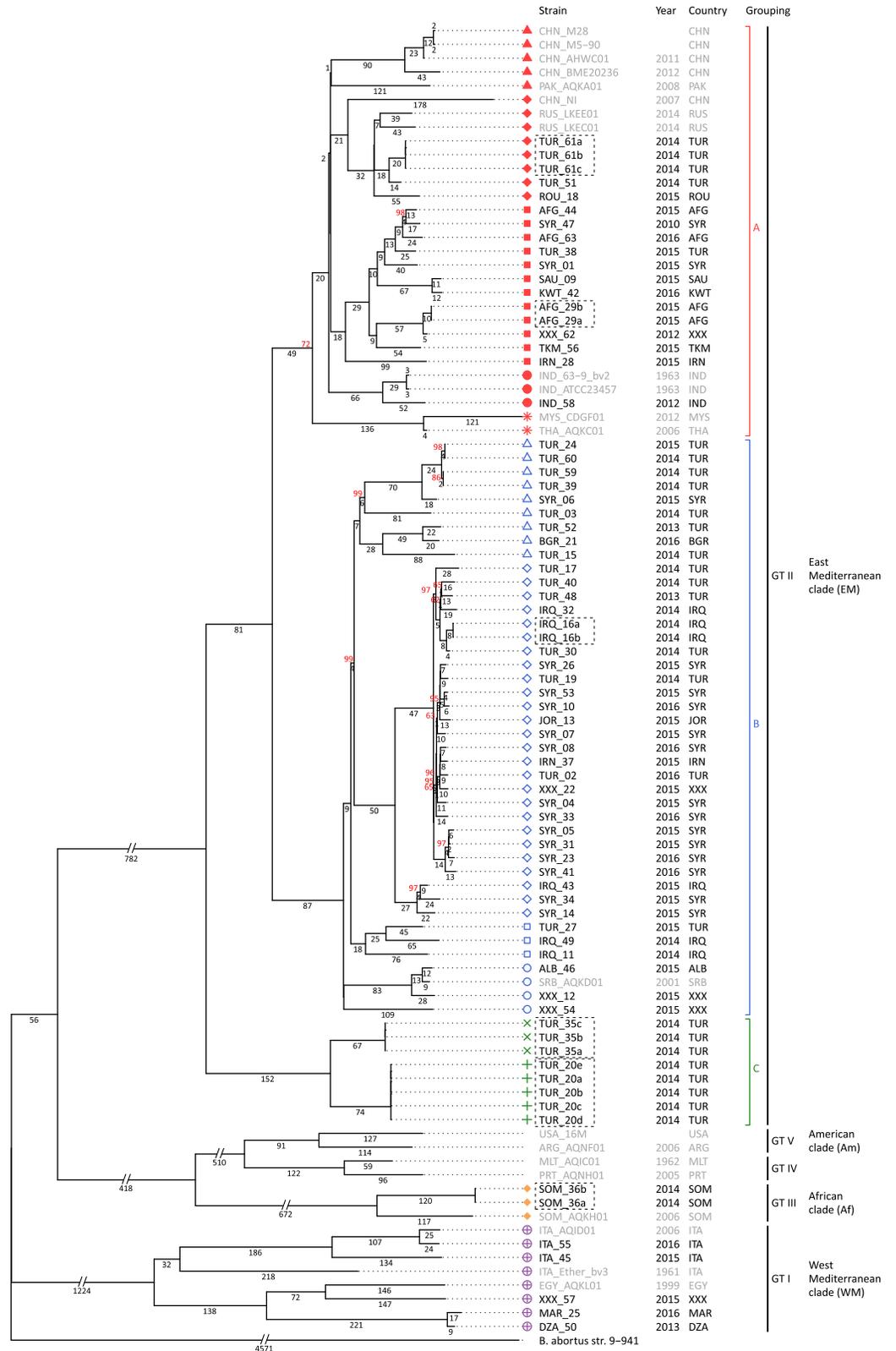
Most of the strains isolated from patients in Germany clumped together in the Eastern Mediterranean clade with three distinct subclusters. Subcluster A covered a broad geographic area and could be subdivided in five clades: (i) Far East (China, but also Pakistan), (ii) South-east Asia (Malaysia, Thailand), (iii) South Asia (India, but also Sudan), (iv) Black Sea area (Turkey, Romania, Russia), (v) Middle East (Iran, Iraq, Turkey, Syria, Turkmenistan, Afghanistan, Kuwait, Saudi Arabia). Subcluster B forms four clades and comprised strains from (i) Turkey-Bulgaria, (ii) Middle East (Syria, Iran, Iraq, Turkey, Jordan) (iii) Turkey-Iraq, and (iv) Southeast Europe (Albania, Serbia). Two of the patients in Germany probably got infected in Turkey and established the third subcluster C.

### In silico PCRs and MLVA

**bcsp31.** As expected, all samples yielded an amplicon of 224 bp size. 3 sequence variants were observed: A dominant genotype with 100 percent conformity of 68 samples, one sample with one SNP variant (52C>T), and 5 samples with another SNP variant (136T>G). The latter demarcated all strains of the Western Mediterranean clade.

**IS711.** The 158 bp multi-copy locus was present in all samples with a median of 9 copies (range 7 to 13). A correlation between copy number and major clade was observed even if the clade specific sample sizes differ from each other: samples from the African and Western Mediterranean clade yielded seven and eight copies, respectively. In contrast, 9 or up to 13 copies of the locus were present in specimen of the Eastern Mediterranean clade.

**recA.** When the analysis was run with *Ochrobactrum*-specific primer sequences (Anth-f/r, Inter-f/r), no amplicon was predicted, whereas genus-specific *Brucella* primers (Bruc-f/r) yielded a 167 bp amplicon with 100% sequence identity for all *B. melitensis* samples.



**Fig 2. Whole genome SNP-based phylogenetic tree discriminates four major lineages.** Clades were inferred using the Maximum Parsimony method. East Mediterranean clade was subdivided in cluster A (red), B (blue), and C (green) with five, four, and two nested clades with unique symbols, respectively. Selected publicly

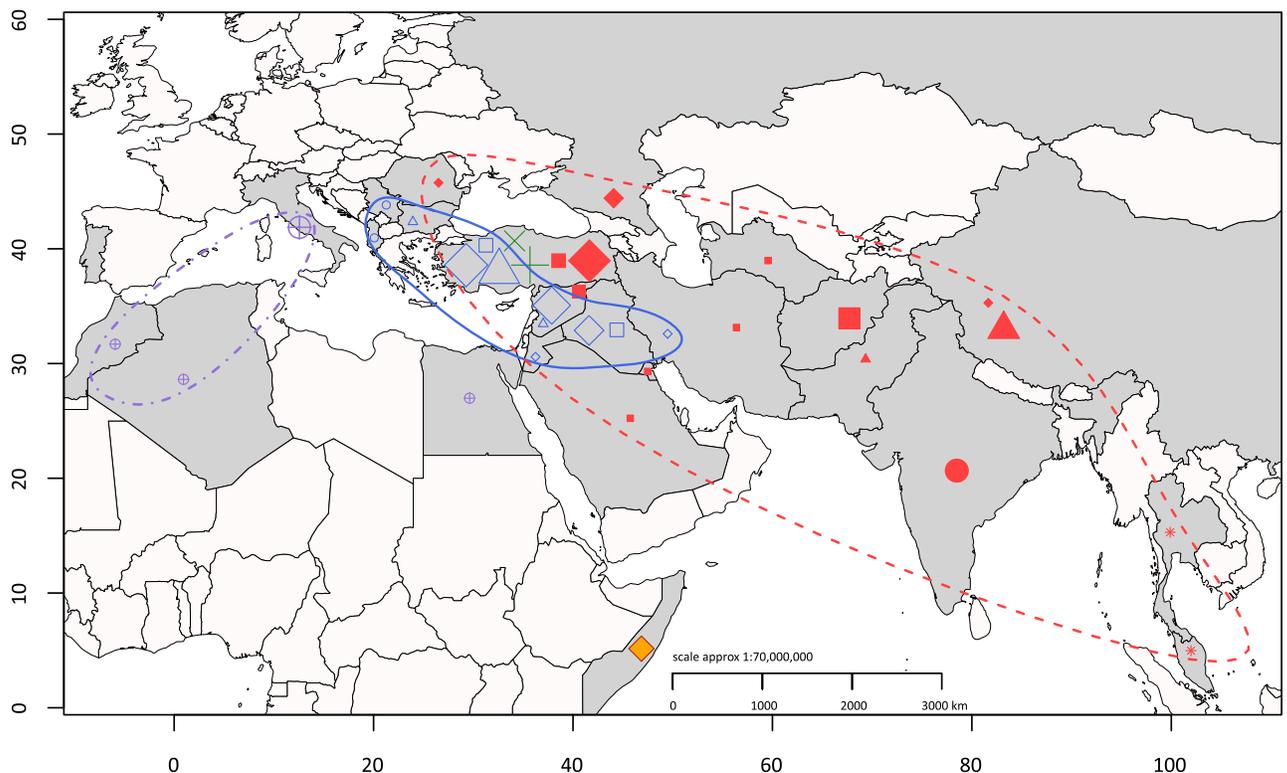
available sequences with geographic attribution were integrated of the dataset, stated with GenBank accession, and marked by grey text colour. Biological or technical replicates were marked with black dotted rectangle. Bootstrap values (based on 1,000 replicates) are shown in red, if less than 100%.

<https://doi.org/10.1371/journal.pone.0175425.g002>

**BruceLadder PCR and virtual gel electrophoresis.** This assay resulted in six identical fragments for each sample. The fragment sizes of 152; 450; 590; 794; 1,071; and 1,682 bps were in accordance with the unique pattern of *B. melitensis*.

**rpoB.** The 4,134 bp locus was present in all assembled sequences. 7 SNP variants with 6 different combinations were observed: 1,332A>G; 1,821C>T; 1,886T>C; 2,954C>T; 3,201C>T; 3,747G>A; and 3,927A>G (see Table 2). Interestingly these single nucleotide alterations resolved three major SNP-based lineages, but also four subclades within the Eastern Mediterranean clade. Genetically encoded resistance determinants with phenotypic effect on the susceptibility of rifampin as a commonly used antibiotic for brucellosis treatment were not observed. This was in accordance with phenotypic testing that could not detect rifampin-resistant strains.

**MLVA.** All technical and biological replicates resulted in identical MLVA-16 genotypes in each case. 3 out of 16 loci incurred single-nucleotide substitutions within primer binding sites that could lead to misinterpretation by conventional MLVA (bruce11R 21T>C, bruce43R 10C>T, bruce43R 22C>T, bruce55R 7G>A, bruce55R 13G>A). In total, 52 different genotypes from 63 patients were observed (for details see S1 Table). The public external databases used as reference contained 1,206 different MLVA-16 genotypes of *B. melitensis* from 37



**Fig 3. Geographic map of genotypes and probable origin of infection.** Genotypes of all patients were coded corresponding to nested clades in Fig 2 and projected to their probable origin of infection on country level.

<https://doi.org/10.1371/journal.pone.0175425.g003>

**Table 2. Mutations of *rpoB* gene.** SNPs were observed at 7 out of 4,134 positions. *rpoB*-genotypes consistently matched the WGS clustering.

1332	1821	1886	2954	3201	3747	3927	Clade	Cluster
A	C	T	C	C	G	A	EM	B
A	C	T	T	C	G	A	EM	A
G	C	C	C	C	G	A	EM	C
G	T	C	C	C	G	A	EM	C
A	C	C	C	T	G	G	Af	not assigned
A	C	C	C	C	A	G	WM	not assigned

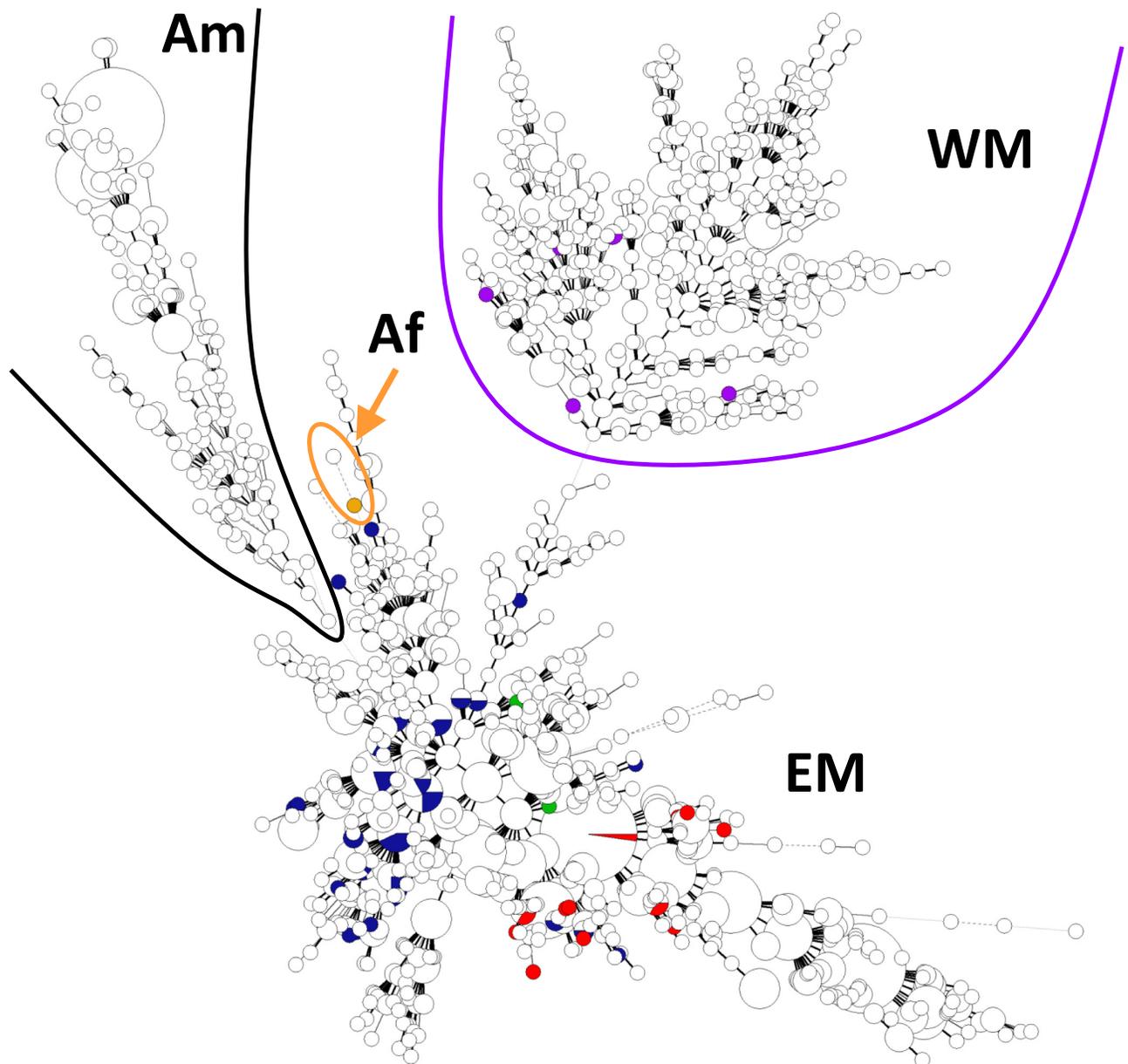
<https://doi.org/10.1371/journal.pone.0175425.t002>

countries, if only complete records (full MLVA-16 panel and geographic origin of specimen, at least at country level) are considered (2,177 entries). This study contributes 28 new genotypes, but no new alleles. A MLVA minimum spanning tree analysis was performed with all *B. melitensis* entries of the database together with all genotypes found in this study (see Fig 4). The illustration reveals that the genotypes are dispersed throughout the phylogeny corresponding to multiple geographic origins of infection. MLVA-16 led to a similar spatial clustering of isolates compared to the SNP analysis with assigning most of the isolates into the Eastern Mediterranean genetic clade.

## Discussion

The overall incidence of human brucellosis in Europe has dramatically decreased over the past 40 years. The decline can be attributed to the implementation of public health measures including testing and slaughtering of cattle, livestock vaccination, border control and surveillance programmes that largely eradicated the causative agent from the animal reservoir and minimized the risk of transmission to humans. Consequently, a brucellosis-free status was granted by the European Union in most countries of Northern and Western Europe. However, a resurgence of case counts has been observed in Germany since 2014, which may be related to increased international travel and trade, rising flow of migrants and refugees entering Europe, or novel foods like raw camel milk. In order to implement appropriate public health countermeasures, knowledge about the origins of infection and transmission routes associated with the cases would be extremely helpful. Molecular typing of strains is an appropriate tool for the investigation of epidemiological relationships. We, therefore, applied whole genome sequencing and subsequent in-silico SNP-typing to 57 *B. melitensis* strains, isolated in Germany from 63 brucellosis patients between 2014 and 2016. Others had already earlier applied the WGS approach to *Brucella*, but primarily just for species discrimination [35,36]. In contrast to such applications, trace-back analysis and the attribution of strains to suspected sources of infection require a high technical discriminatory power. Tan et al. [18] recently published a full genome SNP-based phylogenetic analysis using two own draft genome sequences as well as 49 publicly available genome sequences of *B. melitensis* collected globally. However, due to a lack of available sequences from the Middle East at that time, that region had been omitted, although brucellosis is a major public health concern in that area [37]. Azam et al. [38] sequenced and contributed one single isolate of *B. melitensis* from a naturally infected goat from India to the public databases. Our study almost doubles the number of publicly available *B. melitensis* genomes and also includes isolates from Middle East. This wealth of data will considerably enhance our understanding of the molecular epidemiology of the pathogen.

Our data clearly demonstrate that the increase in brucellosis case numbers in 2014 and 2015 in Germany is associated with strains originating from Middle East. With respect to the limited epidemiological metadata we exclude the possibility that only a few single local sources



**Fig 4. Minimum spanning tree based on MLVA-16 data.** Projection of strains used in this study (coloured circles) to a large collection of genotypes from a public MLVA-database. Genotypes identified in this study are dispersed throughout the phylogeny corresponding to multiple geographic origins of infection.

<https://doi.org/10.1371/journal.pone.0175425.g004>

of infection might have caused this rise. Instead to this, our genetic analysis shows multiple, epidemiological independent strains that obviously had been imported to Germany. This correlates well with the history of most of the patients revealing recent migration from countries where brucellosis is endemic. So, the influx of migrations entering Europe comes along with raised case counts of a hitherto rare disease.

Using a comprehensive collection of strains isolated from patients who acquired brucellosis in various countries, this study also elucidates the distribution of genetic clades, especially of the Eastern Mediterranean lineage, that extends from Turkey to Far East Asia. Whereas strains

of subcluster A occur throughout Asia with distinct nested clades in five dedicated areas, strains of subcluster B and C originated only from Middle East and Turkey, respectively (see Fig 3). However, to consolidate our knowledge of the distribution of *Brucella* genotypes, a representative sampling of strains from human and animal cases in endemic countries would be necessary. In addition this might provide new details, as to which degree the observed genetic diversity of the Turkish isolates is confounded by sampling bias.

As compared to Tan et al. [18], we used different SNP-calling parameters. Instead of removing only paralogous genes, we excluded gaps and ambiguous bases as well as mutations located in annotated regions of rRNAs, tRNAs, repeat regions and tandem repeats. We decided to use the more conservative approach to ensure high reproducibility, which is an important requirement for methods that may as well be used for forensic purposes. Consequently, our technical and biological replicates generated identical results. This finding is contradictory to data from Ke et al., who found 5,019 SNPs in derivative strain 16M13w, isolated from BALB/c mice after infection for 13 weeks, compared to strain 16M used to infect the animal model [39]. The whole-genome-based SNP phylogenetic analysis performed by Tan et al. already suggested, that 16M13w may not have originated from China 16M [18]. Our data set comprising *B. melitensis* strains isolated from follow-up samples of patients in intervals of 7 days (sample SOM\_36) up to 69 days (sample TUR\_61), respectively, showed 100% nucleotide identity, which is in line with the low genetic variability of the pathogen and the high reproducibility of the analysis. This strategy may therefore be suitable to differentiate outbreak from non-outbreak strains or a relapse from re-infection. In addition, it is now possible to predict the origin of unknown *B. melitensis* samples. Thus, we assume Southeast Europe as origin of infection for sample XXX\_12 or the region around Afghanistan and Turkmenistan for sample XXX\_62. Naturally, the resolution of the pathogens' origin is limited to the quality of associated meta-data, in other words unclear correlation of routes of migrations and assumed location of infection obscure phylogeographic attribution.

Even though other methods like multilocus sequence typing (MLST) have been published [40], MLVA is actually the most widely used approach for outbreak investigations [41,42] and is considered to be the gold standard of *Brucella* typing [9]. As it is a comparatively inexpensive and reliable method with high discriminating power and a large publicly accessible database is available, it will certainly remain an important epidemiological tool for trace-back analysis of *B. melitensis* infections in the future. However, our study provides evidence that WGS achieves a higher resolution [43–45]. MLVA discriminated 52 genotypes from 63 patients, but WGS-based SNP analysis distinguished all patients from each other. At least, the spatial clustering of both methods was largely in agreement. As NGS becomes increasingly available in more and more laboratories, in silico PCRs and in silico MLVA will ensure backward compatibility. MLVA data already acquired in the traditional way will, therefore, remain useful for strain comparison.

As the observed SNPs are scattered throughout both chromosomes, clade-specific variation may pertain to genomic regions that are usually considered as targets for typing or identification assays. Actually, such substitutions were detected in our collection and may be useful to develop real-time PCR assays to discriminate major lineages. Marianelli et al. already suggested to use *rpoB* gene as a molecular marker for the genotyping of *Brucella* species [33]. All four SNPs described by Sayan et al. when applying a *rpoB* SNP assay to a collection of strains isolated in Turkey [46] were also present on the genomes studied here. Moreover, three additional SNPs were observed (1332A>G, 1821C>T, 3201C>T), and the latter delineated the African clade. Our results also indicate that WGS is the method of choice to detect mutations within targets of molecular diagnostic assays, as five mutations within three MLVA loci primer

binding sites could be depicted. Furthermore, VNTRs are more susceptible to homoplasy than SNPs in coding regions [43,45].

In summary, our study corroborates WGS to be a suitable tool for trace-back analysis of *B. melitensis* suggesting the potential geographic origin of a given strain. The new methodology achieves a higher resolution compared to common typing approaches, but still ensures full backward compatibility. However, the lack of reliable metadata in public databases often hinders an actual resolution below geographic regions or country level and corresponding precise trace-back analysis. Despite the novel role of WGS-derived bacterial typing in discovery genetic relationships, epidemiological investigations (geographic origin and source of infection, transmission route) still are and will be crucial and a requirement, that genomic sequences are meaningful to future investigations. The expansion of WGS and typing databases in the context of clinical microbiology requires the commitment of all contributors to add reliable and comprehensive metadata. Such metadata should also comprise methods used to generate and process the data. In addition, clinicians in Europe should be aware of brucellosis and consider the disease as a differential diagnosis when febrile illness is linked to residence in endemic countries (e.g. travel history, visit of relatives or immigration).

## Supporting information

**S1 Table. List of MLVA-profiles.** This table comprises the determined MLVA-profiles of all samples attached to available epidemiological metadata.

(XLSX)

**S1 File. Variant calling file for genome comparison.** This file contains a list of all observed single point mutations used in the analyses including their annotation using software snpEff.

(TXT)

## Acknowledgments

The authors thank Claudia Kahlhofer and Philipp Vette for excellent technical assistance, and Dr Joachim J. Bugert for comments that greatly improved the manuscript. This study was supported by internal funding of the German Ministry of Defence Medical Service as part of project no. 57-2014-03. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author Contributions

**Conceptualization:** MHA SZ LZ HCS.

**Data curation:** SZ LZ HCS MTP.

**Formal analysis:** MCW EG.

**Funding acquisition:** LZ MHA.

**Investigation:** BHN EG LMH.

**Project administration:** MHA.

**Resources:** SZ HCS LZ.

**Supervision:** MHA.

**Validation:** MHA EG MCW.

**Visualization:** MCW EG.

**Writing – original draft:** EG MCW MHA.

**Writing – review & editing:** MHW EG MHA SZ HCS LZ.

## References

1. Seleem MN, Boyle SM, Sriranganathan N. Brucellosis: a re-emerging zoonosis. *Vet Microbiol.* 2010; 140: 392–398. <https://doi.org/10.1016/j.vetmic.2009.06.021> PMID: 19604656
2. Moreno E. Retrospective and prospective perspectives on zoonotic brucellosis. *Front Microbiol.* 2014; 5: 213. <https://doi.org/10.3389/fmicb.2014.00213> PMID: 24860561
3. Pappas G, Papadimitriou P, Akritidis N, Christou L, Tsianos EV. The new global map of human brucellosis. *Lancet Infect Dis.* 2006; 6: 91–99. [https://doi.org/10.1016/S1473-3099\(06\)70382-6](https://doi.org/10.1016/S1473-3099(06)70382-6) PMID: 16439329
4. Godfroid J, Al Dahouk S, Pappas G, Roth F, Matope G, Muma J, et al. A “One Health” surveillance and control of brucellosis in developing countries: moving away from improvisation. *Comp Immunol Microbiol Infect Dis.* 2013; 36: 241–248. PMID: 23044181
5. Vollmar P, Zange S, Zöller L, Erkel J, Robert Thoma B. [Brucellosis, an overview and current aspects]. *Dtsch Med Wochenschr* 1946. 2016; 141: 1014–1018.
6. Ariza J, Bosilkovski M, Cascio A, Colmenero JD, Corbel MJ, Falagas ME, et al. Perspectives for the treatment of brucellosis in the 21st century: the Ioannina recommendations. *PLoS Med.* 2007; 4: e317. <https://doi.org/10.1371/journal.pmed.0040317> PMID: 18162038
7. Whatmore AM, Davison N, Cloeckeaert A, Al Dahouk S, Zygmunt MS, Brew SD, et al. *Brucella papionis* sp. nov., isolated from baboons (*Papio* spp.). *Int J Syst Evol Microbiol.* 2014; 64: 4120–4128. <https://doi.org/10.1099/ijs.0.065482-0> PMID: 25242540
8. Scholz HC, Revilla-Fernández S, Al Dahouk S, Hammerl JA, Zygmunt MS, Cloeckeaert A, et al. *Brucella vulpis* sp. nov., isolated from mandibular lymph nodes of red foxes (*Vulpes vulpes*). *Int J Syst Evol Microbiol.* 2016; 66: 2090–2098. <https://doi.org/10.1099/ijssem.0.000998> PMID: 26928956
9. Scholz HC, Vergnaud G. Molecular characterisation of *Brucella* species. *Rev Sci Tech Int Off Epizoot.* 2013; 32: 149–162.
10. Osterman B. International Committee on Systematics of Prokaryotes; Subcommittee on the taxonomy of *Brucella*: Minutes of the meeting, 17 September 2003, Pamplona, Spain. *Int J Syst Evol Microbiol.* 2006; 56: 1173–1175.
11. Le Flèche P, Jacques I, Grayon M, Al Dahouk S, Bouchon P, Denoeud F, et al. Evaluation and selection of tandem repeat loci for a *Brucella* MLVA typing assay. *BMC Microbiol.* 2006; 6: 9. <https://doi.org/10.1186/1471-2180-6-9> PMID: 16469109
12. Al Dahouk S, Flèche PL, Nöckler K, Jacques I, Grayon M, Scholz HC, et al. Evaluation of *Brucella* MLVA typing for human brucellosis. *J Microbiol Methods.* 2007; 69: 137–145. <https://doi.org/10.1016/j.mimet.2006.12.015> PMID: 17261338
13. Aftab H, Dargis R, Christensen JJ, Le Flèche P, Kemp M. Imported brucellosis in Denmark: molecular identification and multiple-locus variable number tandem repeat analysis (MLVA) genotyping of the bacteria. *Scand J Infect Dis.* 2011; 43: 536–538. <https://doi.org/10.3109/00365548.2011.562531> PMID: 21375425
14. Jiang H, Fan M, Chen J, Mi J, Yu R, Zhao H, et al. MLVA genotyping of Chinese human *Brucella melitensis* biovar 1, 2 and 3 isolates. *BMC Microbiol.* 2011; 11: 256. <https://doi.org/10.1186/1471-2180-11-256> PMID: 22108057
15. Gyuranecz M, Wernery U, Kreizinger Z, Juhász J, Felde O, Nagy P. Genotyping of *Brucella melitensis* strains from dromedary camels (*Camelus dromedarius*) from the United Arab Emirates with multiple-locus variable-number tandem repeat analysis. *Vet Microbiol.* 2016; 186: 8–12. <https://doi.org/10.1016/j.vetmic.2016.02.009> PMID: 27016751
16. Garofolo G, Fasanella A, Di Giannatale E, Platone I, Sacchini L, Persiani T, et al. Cases of human brucellosis in Sweden linked to Middle East and Africa. *BMC Res Notes.* 2016; 9: 277. <https://doi.org/10.1186/s13104-016-2074-7> PMID: 27188619
17. Garofolo G, Di Giannatale E, De Massis F, Zilli K, Ancora M, Cammà C, et al. Investigating genetic diversity of *Brucella abortus* and *Brucella melitensis* in Italy with MLVA-16. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis.* 2013; 19: 59–70.
18. Tan K-K, Tan Y-C, Chang L-Y, Lee KW, Nore SS, Yee W-Y, et al. Full genome SNP-based phylogenetic analysis reveals the origin and global spread of *Brucella melitensis*. *BMC Genomics.* 2015; 16: 93. <https://doi.org/10.1186/s12864-015-1294-x> PMID: 25888205

19. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol J Comput Mol Cell Biol.* 2012; 19: 455–477.
20. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009; 10: 421. <https://doi.org/10.1186/1471-2105-10-421> PMID: 20003500
21. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl.* 2010; 26: 589–595.
22. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* 2014; 9: e112963. <https://doi.org/10.1371/journal.pone.0112963> PMID: 25409509
23. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, et al. Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. *Omics J Integr Biol.* 2008; 12: 137–141.
24. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 2014; 15: 524. <https://doi.org/10.1186/s13059-014-0524-x> PMID: 25410596
25. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999; 27: 573–580. PMID: 9862982
26. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinforma Oxf Engl.* 2004; 20: 289–290.
27. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* 1980; 16: 111–120. PMID: 7463489
28. Wilgenbusch JC, Swofford D. Inferring evolutionary trees with PAUP\*. *Curr Protoc Bioinforma.* 2003; Chapter 6: Unit 6.4.
29. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinforma Oxf Engl.* 2011; 27: 592–593.
30. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin).* 2012; 6: 80–92.
31. Scholz HC, Pfeffer M, Witte A, Neubauer H, Al Dahouk S, Wernery U, et al. Specific detection and differentiation of *Ochrobactrum anthropi*, *Ochrobactrum intermedium* and *Brucella* spp. by a multi-primer PCR that targets the *recA* gene. *J Med Microbiol.* 2008; 57: 64–71. <https://doi.org/10.1099/jmm.0.47507-0> PMID: 18065669
32. Mayer-Scholl A, Draeger A, Göllner C, Scholz HC, Nöckler K. Advancement of a multiplex PCR for the differentiation of all currently described *Brucella* species. *J Microbiol Methods.* 2010; 80: 112–114. <https://doi.org/10.1016/j.mimet.2009.10.015> PMID: 19887090
33. Marianelli C, Ciuchini F, Tarantino M, Pasquali P, Adone R. Molecular characterization of the *rpoB* gene in *Brucella* species: new potential molecular markers for genotyping. *Microbes Infect Inst Pasteur.* 2006; 8: 860–865.
34. Grissa I, Bouchon P, Pourcel C, Vergnaud G. On-line resources for bacterial micro-evolution studies using MLVA or CRISPR typing. *Biochimie.* 2008; 90: 660–668. <https://doi.org/10.1016/j.biochi.2007.07.014> PMID: 17822824
35. Foster JT, Beckstrom-Sternberg SM, Pearson T, Beckstrom-Sternberg JS, Chain PSG, Roberto FF, et al. Whole-genome-based phylogeny and divergence of the genus *Brucella*. *J Bacteriol.* 2009; 191: 2864–2870. <https://doi.org/10.1128/JB.01581-08> PMID: 19201792
36. Foster JT, Price LB, Beckstrom-Sternberg SM, Pearson T, Brown WD, Kiesling DM, et al. Genotyping of *Brucella* species using clade specific SNPs. *BMC Microbiol.* 2012; 12: 110. <https://doi.org/10.1186/1471-2180-12-110> PMID: 22712667
37. Musallam II, Abo-Shehada MN, Hegazy YM, Holt HR, Guitian FJ. Systematic review of brucellosis in the Middle East: disease frequency in ruminants and humans and risk factors for human infection. *Epidemiol Infect.* 2016; 144: 671–685. <https://doi.org/10.1017/S0950268815002575> PMID: 26508323
38. Azam S, Rao SB, Jakka P, NarasimhaRao V, Bhargavi B, Gupta VK, et al. Genetic Characterization and Comparative Genome Analysis of *Brucella melitensis* Isolates from India. *Int J Genomics.* 2016; 2016: 3034756. <https://doi.org/10.1155/2016/3034756> PMID: 27525259
39. Ke Y, Yuan X, Wang Y, Bai Y, Xu J, Song H, et al. Genome sequences of *Brucella melitensis* 16M and its two derivatives 16M1w and 16M13w, which evolved in vivo. *J Bacteriol.* 2012; 194: 5489. <https://doi.org/10.1128/JB.01293-12> PMID: 22965104
40. Whatmore AM, Perrett LL, MacMillan AP. Characterisation of the genetic diversity of *Brucella* by multilocus sequencing. *BMC Microbiol.* 2007; 7: 34. <https://doi.org/10.1186/1471-2180-7-34> PMID: 17448232

41. De Massis F, Ancora M, Atzeni M, Rolesu S, Bandino E, Danzetta ML, et al. MLVA as an Epidemiological Tool To Trace Back *Brucella melitensis* Biovar 1 Re-Emergence in Italy. *Transbound Emerg Dis*. 2015; 62: 463–469. <https://doi.org/10.1111/tbed.12397> PMID: 26194658
42. Dorneles EMS, Santana JA, Alves TM, Pauletti RB, JP da S Mol, Heinemann MB, et al. Genetic stability of *Brucella abortus* isolates from an outbreak by multiple-locus variable-number tandem repeat analysis (MLVA16). *BMC Microbiol*. 2014; 14: 186. <https://doi.org/10.1186/1471-2180-14-186> PMID: 25015840
43. Ahlstrom C, Barkema HW, Stevenson K, Zadoks RN, Biek R, Kao R, et al. Limitations of variable number of tandem repeat typing identified through whole genome sequencing of *Mycobacterium avium* subsp. *paratuberculosis* on a national and herd level. *BMC Genomics*. 2015; 16: 161. <https://doi.org/10.1186/s12864-015-1387-6> PMID: 25765045
44. Sahl JW, Pearson T, Okinaka R, Schupp JM, Gillece JD, Heaton H, et al. A *Bacillus anthracis* Genome Sequence from the Sverdlovsk 1979 Autopsy Specimens. *mBio*. 2016; 7.
45. Homolka S, Projahn M, Feuerriegel S, Ubben T, Diel R, Nübel U, et al. High resolution discrimination of clinical *Mycobacterium tuberculosis* complex strains based on single nucleotide polymorphisms. *PloS One*. 2012; 7: e39855. <https://doi.org/10.1371/journal.pone.0039855> PMID: 22768315
46. Sayan M, Yumuk Z, Bilenoglu O, Erdenlig S, Willke A. Genotyping of *Brucella melitensis* by *rpoB* gene analysis and re-evaluation of conventional serotyping method. *Jpn J Infect Dis*. 2009; 62: 160–163. PMID: 19305062