

Article

Whole Genome Sequence Analysis of *Brucella melitensis* Phylogeny and Virulence Factors

Peter Rabinowitz ^{1,*}, Bar Zilberman ², Yair Motro ² , Marilyn C. Roberts ¹, Alex Greninger ³, Lior Nesher ^{2,4} , Shalom Ben-Shimol ^{2,4}, Yael Yagel ^{2,4}, Michael Gdalevich ^{2,5}, Orly Sagi ^{2,4}, Nadav Davidovitch ², David Kornspan ⁶ , Svetlana Bardenstein ⁶ and Jacob Moran-Gilad ^{2,4}

¹ Department of Environmental and Occupational Health Sciences, School of Public Health, University of Washington, Seattle, WA 98195, USA; marilyn@uw.edu

² Faculty of Health Sciences, Ben Gurion University of the Negev, Beer Sheva 84105, Israel; Barsch@post.bgu.ac.il (B.Z.); motroy@post.bgu.ac.il (Y.M.); Nesherke@bgu.ac.il (L.N.); shalomb2@clalit.org.il (S.B.-S.); Ygrushka@gmail.com (Y.Y.); michael.gdalevich@bsh.health.gov.il (M.G.); Orlisha@clalit.org.il (O.S.); nadavd@bgu.ac.il (N.D.); giladko@post.bgu.ac.il (J.M.-G.)

³ Department of Laboratory Medicine and Pathology, School of Medicine, University of Washington, Seattle, WA 98195, USA; agrening@uw.edu

⁴ Soroka University Medical Center, Beer Sheva 84101, Israel

⁵ Southern District Health Office, Ministry of Health, Beer Sheva 84104, Israel

⁶ Brucellosis Lab, OIE, FAO Reference Laboratory, Kimron Veterinary Institute, Bet Dagan 50250, Israel; Davidko@moag.gov.il (D.K.); Svetab@moag.gov.il (S.B.)

* Correspondence: Peterr7@uw.edu



Citation: Rabinowitz, P.; Zilberman, B.; Motro, Y.; Roberts, M.C.; Greninger, A.; Nesher, L.; Ben-Shimol, S.; Yagel, Y.; Gdalevich, M.; Sagi, O.; et al. Whole Genome Sequence Analysis of *Brucella melitensis* Phylogeny and Virulence Factors. *Microbiol. Res.* **2021**, *12*, 698–710. <https://doi.org/10.3390/microbiolres12030050>

Academic Editor: Beniamino T. Cenci-Goga

Received: 22 July 2021

Accepted: 16 August 2021

Published: 24 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Brucellosis has a wide range of clinical severity in humans that remains poorly understood. Whole genome sequencing (WGS) analysis may be able to detect variation in virulence genes. We used *Brucella melitensis* sequences in the NCBI Sequence Read Archive (SRA) database to assemble 248 whole genomes, and additionally, assembled 27 *B. melitensis* genomes from samples of human patients in Southern Israel. We searched the 275 assembled genomes for the 43 *B. melitensis* virulence genes in the Virulence Factors of Pathogenic Bacteria Database (VFDB) and 10 other published putative virulence genes. We explored pan-genome variation across the genomes and in a pilot analysis, explored single nucleotide polymorphism (SNP) variation among the ten putative virulence genes. More than 99% of the genomes had sequences for all *Brucella melitensis* virulence genes included in the VFDB. The 10 other virulence genes of interest were present across all the genomes, but three of these genes had SNP variation associated with particular *Brucella melitensis* genotypes. SNP variation was also seen within the Israeli genomes obtained from a small geographic region. While the *Brucella* genome is highly conserved, this novel and large whole genome study of *Brucella* demonstrates the ability of whole genome and pan-genome analysis to screen multiple genomes and identify SNP variation in both known and novel virulence genes that could be associated with differential disease virulence. Further development of whole genome techniques and linkage with clinical metadata on disease outcomes could shed light on whether such variation in the *Brucella* genome plays a role in pathogenesis.

Keywords: *Brucella melitensis*; virulence factors; brucellosis/epidemiology; single nucleotide polymorphism; whole genome sequencing

1. Introduction

Brucella spp., a Gram-negative proteobacterium, is a facultative intracellular pathogen that causes brucellosis, one of the most common and severe zoonotic diseases worldwide [1]. *Brucella* is also considered to be a potential bioterrorism agent [2]. North Africa and the Middle East experience the highest known burden of human brucellosis [2], with incidence rates over 250 per 100,000 in some populations [3]. While it causes acute febrile illness, brucellosis can also manifest as a chronic disease resembling common conditions, such

as arthritis [4]. While 50 years ago there were only 3 known species of *Brucella* (*Brucella abortus*, *Brucella melitensis*, and *Brucella suis*), growth in scientific knowledge about the genus *Brucella* in recent decades has resulted in the recognition of multiple additional species including *Brucella canis*, *Brucella ceti*, *Brucella inopinata*, *Brucella microti*, *Brucella neotomae*, *Brucella ovis*, and *Brucella pinnipedialis* [5]. Among different *Brucella* species, *B. melitensis* is considered to have the greatest virulence in humans, evident by its ability to cause life-threatening infections such as endocarditis [5,6]. The current study focuses on this *Brucella* species.

In humans, *Brucella melitensis* infection can cause a wide range of clinical manifestations in both children and adults. Some infected patients may be asymptomatic or experience a mild and self-limited febrile illness, while other patients may develop musculoskeletal involvement including osteomyelitis and osteoarthritis. Uncommonly, there are severe and sometimes fatal cardiac, neurological, or visceral complications, such as endocarditis, neurobrucellosis, or hepatosplenomegaly. While often brucellosis is generally an acute disease, in some patients it relapses and becomes a recurrent and/or chronic condition (1). Brucellosis in pregnancy causes an increased risk of spontaneous abortion, preterm delivery, and miscarriage.

The biological basis of this significant clinical variability of brucellosis remains poorly understood. Theoretically, this could relate to individual differences in host susceptibility, but genomic variation in the pathogen could also play a role [7]. At the same time, the *B. melitensis* genome appears to be well conserved. For example, an analysis of a *Brucella melitensis* outbreak found clonality between the different isolates and no evidence of adaptation to different hosts, including different animal species [8,9]. While *Brucella melitensis* is not considered to have “classical” virulence factors, such as exotoxins or cytotoxins, effective pathogenesis involves its ability to survive intracellularly and evade the host immune system [10].

The *Brucella melitensis* genome consists of two circular chromosomes, one 1.1 and one 2.2 Mb, for a genome of ≈ 3.3 Mb [11]. Numerous *B. melitensis* genes have been identified as potential virulence factors [6,12], but variation between strains of *B. melitensis* regarding these genes have not been well described. Multiple genes manipulate the macrophage response to prevent apoptosis, and others assist with intracellular invasion and survival [6,12]. For example, the type IV secretion system (T4SS), encoded by the *virB* operon, is essential for intracellular survival and replication [13]. Urease affects the persistence of *Brucella* spp. in the low pH intracellular environment [14]. While recombination is rare, genomic islands may play a role in virulence and fitness variation through horizontal gene transfer [15]. While Ding et al. found SNP differences between a virulent strain of *B. melitensis* and some attenuated vaccine strains [16], it is unclear how much difference in virulence occurs naturally. A study of 60 *B. melitensis* isolates from Iran reported that *btpA* (also known as *TcpB*), *btpB*, *virB5*, *vceC*, *bpe275*, *bspB*, and *virB2* genes were present in 100% of the isolates, while *prpA* and *betB* were detected in 86% and 97%, respectively [6]. However, using PCR methods, an analysis of 57 *B. melitensis* and 21 *B. abortus* isolates from Iran found variable frequencies of six virulence genes, Urease (*ure*), Mannosyl-transferase (*wbkA*), Outer membrane protein 19 (*omp19*), Membrane-bound protein (*mviN*), Mannose-6-phosphate isomerase (*manA*) and Perosamine-synthetase (*perA*), (74.4%, 89.7%, 93.6%, 94.9%, 100% and 92.3% respectively) [17]. Although these studies examined only a limited number of genes, their findings suggest that it is possible that many virulence genes may be present across strains, and that consequently examination of single nucleotide polymorphism (SNP) may be important (16). Techniques that allow examination of multiple genes and polymorphisms in key genes will be important to better elucidate questions about determinations of variation in *Brucella* virulence.

Advances in whole genome sequencing (WGS) technology allow for pan-genome assessment of variability in *Brucella* genomes, including virulence genes, which also conducts a rapid comparison to virulence factor databases and exploration of how these genes vary between strains. WGS has revealed that while the genome is highly conserved, there are

strains that cluster in different geographic regions, allowing strain variation to be used for traceback analysis [18]. WGS allows for a more precise definition of such strain variation than previous classification methods such as MLVA [19] and has been used for detection of outbreak patterns in particular countries [20], as well as the movement of the pathogen between regions of a country [21]. Whole genome sequencing has been used to define five distinct genotypes (I–V) of *B. melitensis* as well as several subgenotypes [22]. Genotype I was the most basal lineage and clustered in the Mediterranean region. Genotype II was associated with several different regions, including the Middle East, while genotype III appears to cluster strongly in Africa, genotype IV is associated with Europe, and genotype V is associated with the Americas. These genotype classifications were confirmed in several recent WGS studies, including one that compared whole genomes of 11 *B. melitensis* isolates from Russia to 87 *B. melitensis* isolates in the NCBI Genbank, a study of 57 imported cases in Germany, and a comparison of 13 NCBI *B. melitensis* genomes with genomes of 25 *B. melitensis* isolates from patients in Norway [18,23,24].

While WGS holds the potential to rapidly examine variation in multiple *Brucella* virulence genes in a more comprehensive manner than with previous techniques, the use of WGS to examine genetic variation in virulence genes of *Brucella* has been limited. Whole genome sequencing was used to investigate phenotypic resistance to rifampin of a *B. melitensis* strain from a single human patient [25]. This allowed the authors to identify a lack of mutations in a single resistance gene, but further analysis was not reported. In another study, a whole genome comparison of the Rev1 attenuated vaccine strain of *B. melitensis* to the virulent 16M *B. melitensis* reference strain revealed regions of insertion and deletion in the Rev1 genome, as well as several missense mutations in several virulence genes [26].

To explore the potential of whole genome sequencing to investigate the genomic basis of clinical variation in virulence of *B. melitensis* infection, we performed a study of *B. melitensis* sequences in NCBI, as well as additional sequences from a series of human *Brucella* isolates in a limited geographical region (Southern Israel).

2. Materials and Methods

2.1. Search and Accession of Sequences

We searched the NCBI Sequence Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra/> (accessed on 12 September 2019)) all sequences of *B. melitensis* isolates that had been generated by Illumina methodology. We used the following search query: ("*Brucella melitensis*" [Organism] OR *brucella melitensis* [All Fields]) AND wgs [All Fields] AND illumina [All Fields] AND paired [All Fields]).

This search allowed the selection of sequences that were generated using comparable technology. All publicly available raw reads for *B. melitensis* meeting the search definition were downloaded from the SRA database along with the available metadata, using the reads_download component of the Flowcraft pipeline (release 1.3.1, <https://github.com/assemblerflow/flowcraft> (accessed on 12 September 2019), using default parameters, unless stated otherwise).

2.2. *Brucella* Isolates Recovered in Israel

In addition to the NCBI data, we included in our analysis 27 *B. melitensis* isolates from human patients with *Brucella* bacteremia seen at the Soroka University Medical Center in Southern Israel in 2017. Blood cultures from that hospital presumptively growing *Brucella* are sent to the National *Brucella* Reference Laboratory for confirmatory culture and serotyping. DNA was extracted in the reference laboratory from clinical *Brucella* isolates using a two-step approach (heat killing at 80 degrees C for 10 min, followed by extraction using the Qiagen Blood and Tissue kit; Qiagen Sciences, Germantown, MD, USA). DNA extracts were confirmed as non-infectious and then sent on dry ice without personal identifiers to the lab of one of the authors (AG) at the University of Washington, where whole genome sequencing was carried out using an Illumina MiSeq. Briefly, 1 ng of

DNA was tagged using Nextera XT transposase with 15 cycles of PCR amplification. Libraries were paired-end 2×125 bp and sequenced using the Illumina HiSeq system, aiming for a minimum coverage $>100\times$. Additional MiSeq 2×300 bp sequencing was performed on any isolates for which additional sequencing reads were required to increase coverage and contiguity. Quality control for sequencing runs was $>99\%$ of Q30 bases after trimming with $>100\times$ minimum coverage and assembly N50 > 100 kb. Reads are available in BioProject PRJNA608659. Supplementary Table S1 lists the sequences used for this study.

2.3. Assembly of Whole Genomes

Raw Reads QC, Filtering, and Assembly

Raw sequencing reads underwent Quality Control (QC), assembly, and species identification using the flowcraft pipeline, including the integrity_coverage and fastqc_trimmomatic components for raw reads QC, filtering, and trimming. Read depth estimation and down-sampling to $100\times$ depth (if required) were performed with the components check_coverage and downsample_fastq (using the estimated *B. melitensis* genome size of 3.3 Mb, and minimum coverage of $15\times$). Species identification for contamination and sequence quality from the reads was performed using the Kraken (version 1) component with the Minikraken database (minikrakenDB2017). The filtered and trimmed reads were assembled using the spades component (SPAdes Version 3.12) and the assemblies were corrected using the process_spades, assembly_mapping, and pilon components. QC details for the genome assemblies (using QUAST v5.0.2, <https://doi.org/10.1093/bioinformatics/bty266> (accessed on 12 September 2019)) are available as a supplementary table (Supplementary Table S5). The reference used was Bm 16M (GCA_000007125.1). BUSCO (version 3.0.2, <https://doi.org/10.1093/molbev/msx319> (accessed on 5 July 2021)) was used with the bacteria_odb9 dataset.

SRA sequences failing QC or assembly were not analyzed further. In addition, several published whole genomes for *B. melitensis* did not have sequences in SRA and were therefore not included in the analysis.

2.4. Phylogenomic Analyses

Phylogenetic analyses were based on multi-loci sequence typing (MLST), core genome MLST (cgMLST), and single nucleotide polymorphism (SNP) analysis methods. The assembled genomes were grouped by assigning sequence types (STs) using a publicly available pubMLST 9-loci schema for *Brucella* [27] in silico with the MLST component of flowcraft. The creation of an ad hoc cgMLST schema for *B. melitensis* was performed with chewB-BACA (as per published methods) [28] using the assembled *B. melitensis* genomes (with the Minimum BSR at a default 0.6, Prodigal training file trained on the *B. melitensis* 16M reference strain complete genome [GenBank assembly accession ID: GCA_000007125.1]), producing a final cgMLST schema consisting of 2652 loci (at 95% genome presence).

Core SNPs were identified using the software tool snippy (version 4.1.0, <https://github.com/tseemann/snippy> (accessed on 12 September 2019)), where all genome assemblies, with nine outliers [see Table S2], were compared to the BM2_63_9 (SRR4038986) assembly as a reference genome with the core SNPs isolated with snippy-core, and potential genome recombination sites masked using the software tool Gubbins (version 2.3.4, <https://github.com/sanger-pathogens/gubbins> (accessed on 12 September 2019)).

We included previously published assignment of genotypes to individual strains, as well as host species information as part of our metadata (see Supplementary Table S1)

2.5. Analysis of Virulence Genes

Virulence-associated genes appearing in the Virulence Factors of Pathogenic Bacteria Database (VFDB; updated October 2018, <http://www.mgc.ac.cn/VFs/main.htm> (accessed on 12 September 2019)) as well as 10 additional targets were analyzed in all of the assembled *B. melitensis* genomes (including the 27 Israeli and 248 SRA assembled genomes). We used the ABRicate component of the flowcraft pipeline with the VFDB, and the parameters of a minimum identity at 85% and hit coverage at 80% [29]. VFDB (October 2018

version) consists of 3202 putative bacterial virulence genes, of which 43 are identified with *B. melitensis*. All genes in the database are labeled with the source bacterial species.

We searched for 10 other previously published putative virulence genes (*ure*, *omp19*, *mviN*, *vceC*, *prpA*, *betB*, *bpe275*, *bspB*, *perA*, *manA*) using the primers described in the relevant publication (see Supplementary Table S4) by performing in silico PCR (using the program ipress from the exonerate package: <https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate> (accessed on 3 October 2019) [30]).

2.6. Visualizations

The MLST, cgMLST, core SNPs, and virulome analyses results were visualized with appropriate metadata as a minimum spanning tree (MST), using GrapeTree (version 1.5, doi:10.1101/gr.232397.117) with the MSTreeV2 method. Also using GrapeTree, a Neighbour-Joining (NJ) phylogenetic tree was generated from the ad hoc cgMLST profile and visualized with metadata using the R package ggtree (doi:10.18129/B9.bioc.ggtree).

2.6.1. Pangenome Analysis

The pangenome for all 275 genome assemblies was analyzed using the tool panaroo [31]. A maximum-likelihood tree was constructed from the pangenome using the tool Iqtree2 (v2.1.2) [32] and visualized using GrapeTree. Hierarchical clustering of pangenome results was performed in R (v4.0.3) using the pheatmap function (with the “complete” clustering method) from the package COMPASS [33].

2.6.2. SNP Variation Analysis

For this pilot analysis of SNP variation among virulence genes, we examined the 10 virulence genes mentioned in recent papers as above (*ure*, *omp19*, *mviN*, *vceC*, *prpA*, *betB*, *bpe275*, *bspB*, *perA*, *manA*) for SNP variation. Genes were extracted via in silico PCR (as described above) and aligned with the Mafft (v7.397) tool [34]. Single nucleotide polymorphisms (SNPs) were identified from the multiple sequence alignment with the tools Snp-sites [35] and Bcftools [36].

3. Results

3.1. Genome Assembly

We were able to assemble a total of 275 genomes, including 248 from the NCBI SRA and 27 from the sequences of isolates from Israel. The host species for the 275 genomes included humans (66%), sheep (25%), goats (5%), cattle (3%), and unknown/other (1%). Supplementary Table S1 shows the metadata associated with these 275 genomes, including in silico genotype, host species, previously assigned genotype (when available), country, and region. Supplementary Table S2 indicates reference *Brucella* genomes used for the analysis.

3.2. Segregation by Genotype and Region

3.2.1. Global Clustering of Genotypes I–V

As Figures 1 and 2 show, the assembled genomes distribute into five major phylogenetic clusters that correspond to previously described genotypes [22], with genotypes III, IV, and V clustering more closely together, and several genotypes having defined subgenotypes. The majority of genomes assigned to genotypes were of genotype II. Many of the 275 isolates had not previously been assigned to genotypes in published literature, but Figure 1 shows that genotype information, when available, indicated clustering by geographical areas, with genotype I associated with Europe, genotype II predominating in the middle east, genotype III in Africa, and genotype V in South America. This pattern resembles previously published spatial analyses of *B. melitensis* phylogeny [18,19,24,37]. All of the Israeli strains mapped to the genotype II that has been associated with the Middle East.

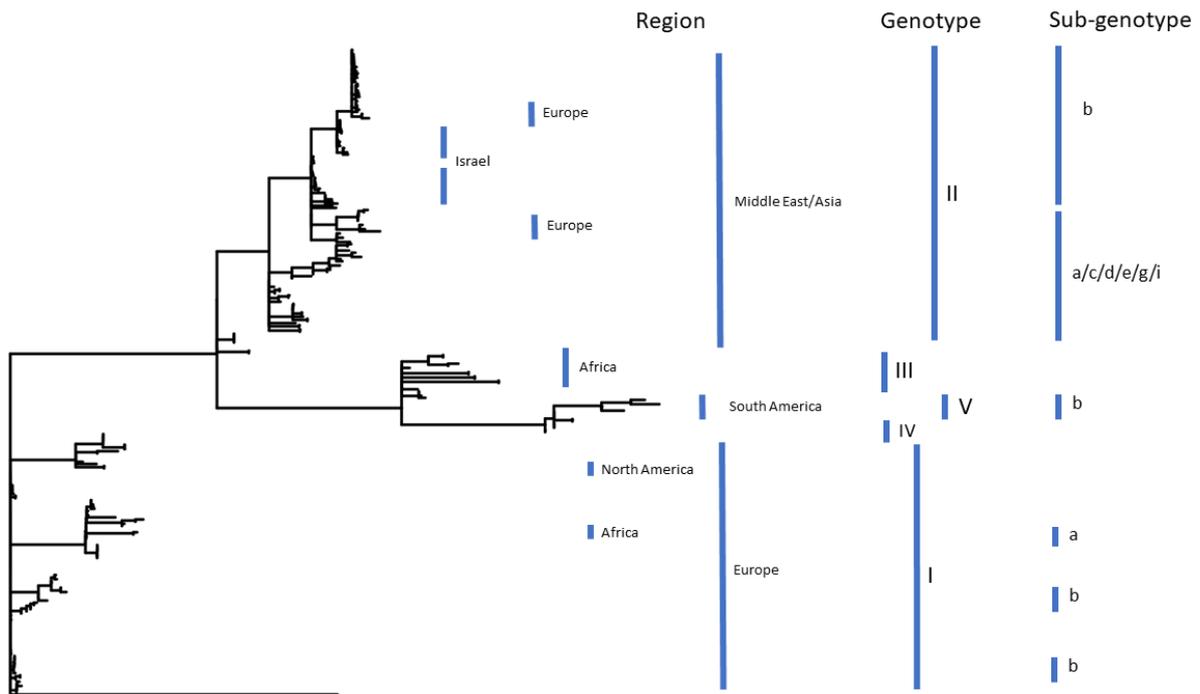


Figure 1. The phylogenetic groupings of *B. melitensis* isolates by their genotype and region.

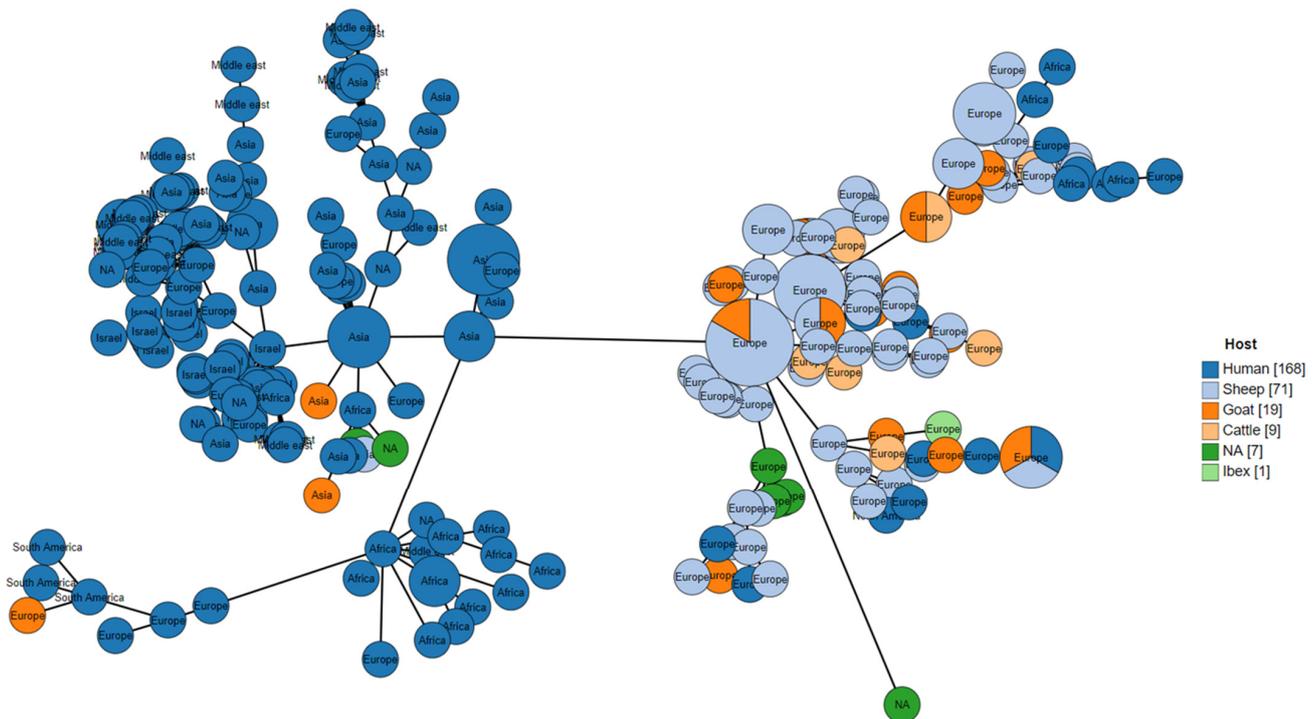


Figure 2. The GrapeTree minimum-spanning tree visualization of cgMLST of SRA public and Israeli WGS using chewBBACA (2652 loci at 95% presence), colored by the host species and labeled by the region. The red arrow indicates the position in the phylogeny of the Israel samples. The distances are the number of loci not shared by the connected nodes.

3.2.2. Host Species

Figure 2 shows the 275 whole genomes in a cluster visualization, with the clusters corresponding to the genotypes previously described. The figure demonstrates that while available *Brucella* genomes in NCBI include genomes from both human and animal hosts, some genotypes (such as genotype II) are predominantly represented in NCBI collections

currently by human isolates, while other genotype collections (such as genotype 1) include a greater number of animal origin genomes.

3.2.3. Local Variation

Figure 3 shows only the Israeli isolate genomes, with core SNPs used to define clustering as a minimum spanning tree visualization. These isolates, even though found in a small geographic region, fall into two major clusters separated by slightly more than one hundred SNPs. The visualization shows that in certain locations (towns), *Brucella* strains from both clusters are circulating, demonstrating some genomic heterogeneity even at a small geographic scale.

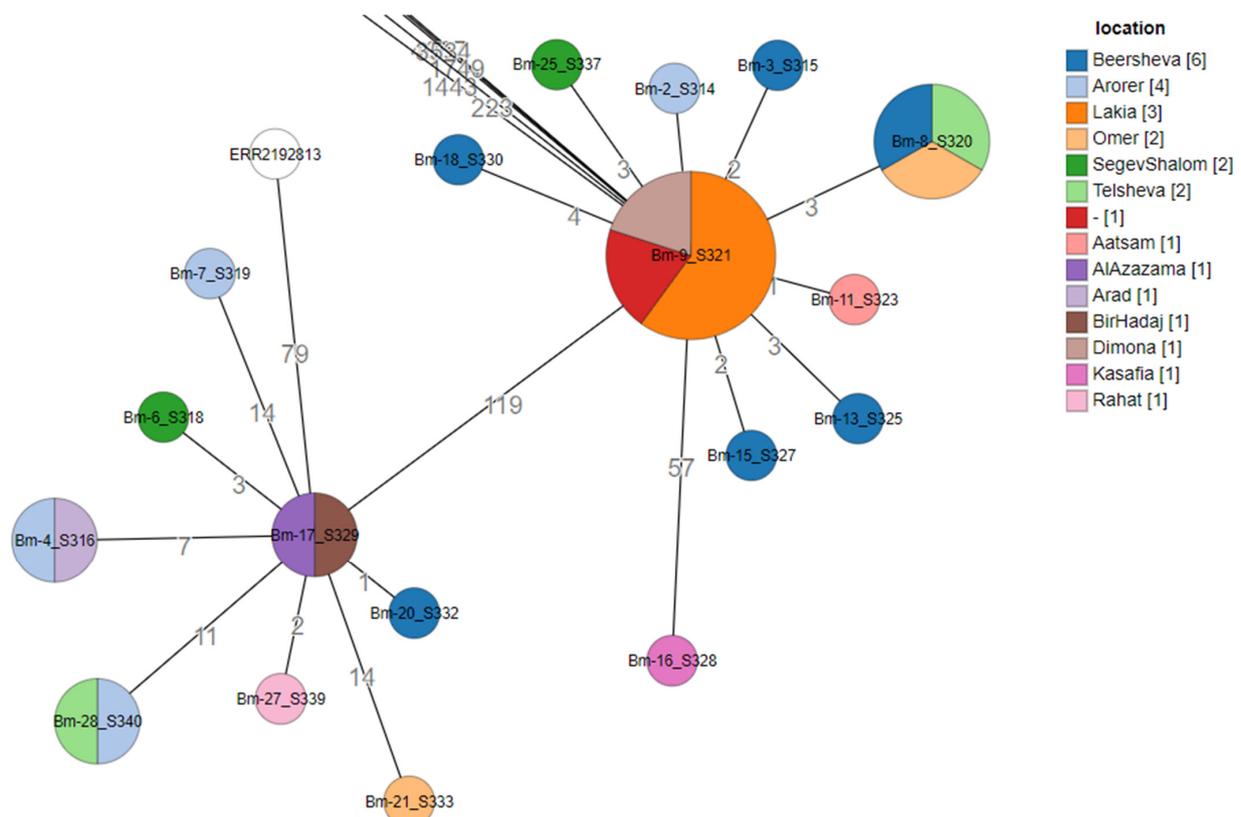


Figure 3. The core SNP analysis of Israeli samples and some outliers (not visible), using BM2_63_9 (SRR4038986) as a reference. The nodes were colored according to the patient's town. The distances are the number of SNPs between two samples (nodes).

3.3. Pangenome Analysis

Pangenome analysis of the 275 genome assemblies revealed a total of 3383 genes. The analysis revealed that the *Brucella* pangenome was highly conserved in our sample, with 2971 (88%) of the 3383 genes present in at least 95% of the genome assemblies. We classified 2933 genes with 99% or greater presence as Core genes, and 38 genes with presence between 95% and 99% as Soft-core genes. We classified 252 genes (7.5% of the total) that were present in between 15% and 95% of the genomes as Shell genes, and 160 genes that were present in less than 15% of the genomes as Cloud genes. The overall clustering of the 275 genome assemblies by pangenome analysis (Figure 4) showed good agreement with the genotypes already described, resembling the analysis performed using cgSNPs (Figure 1).

Figure 5A shows the heatmap of gene presence [red] and absence [blue] for the 252 genes in the Shell gene accessory pangenome. The shell genes show clustering of the accessory pangenome in a manner that is similar to the genotypes defined by the cgSNP analysis, as depicted at the top of the heatmap. Clustering was also seen when comparing the structural presence/absence of gene triplets in the entire pangenome (Figure 5B). Once

again, this clustering corresponds closely to the previously described genotypes shown at the top of the heat map.

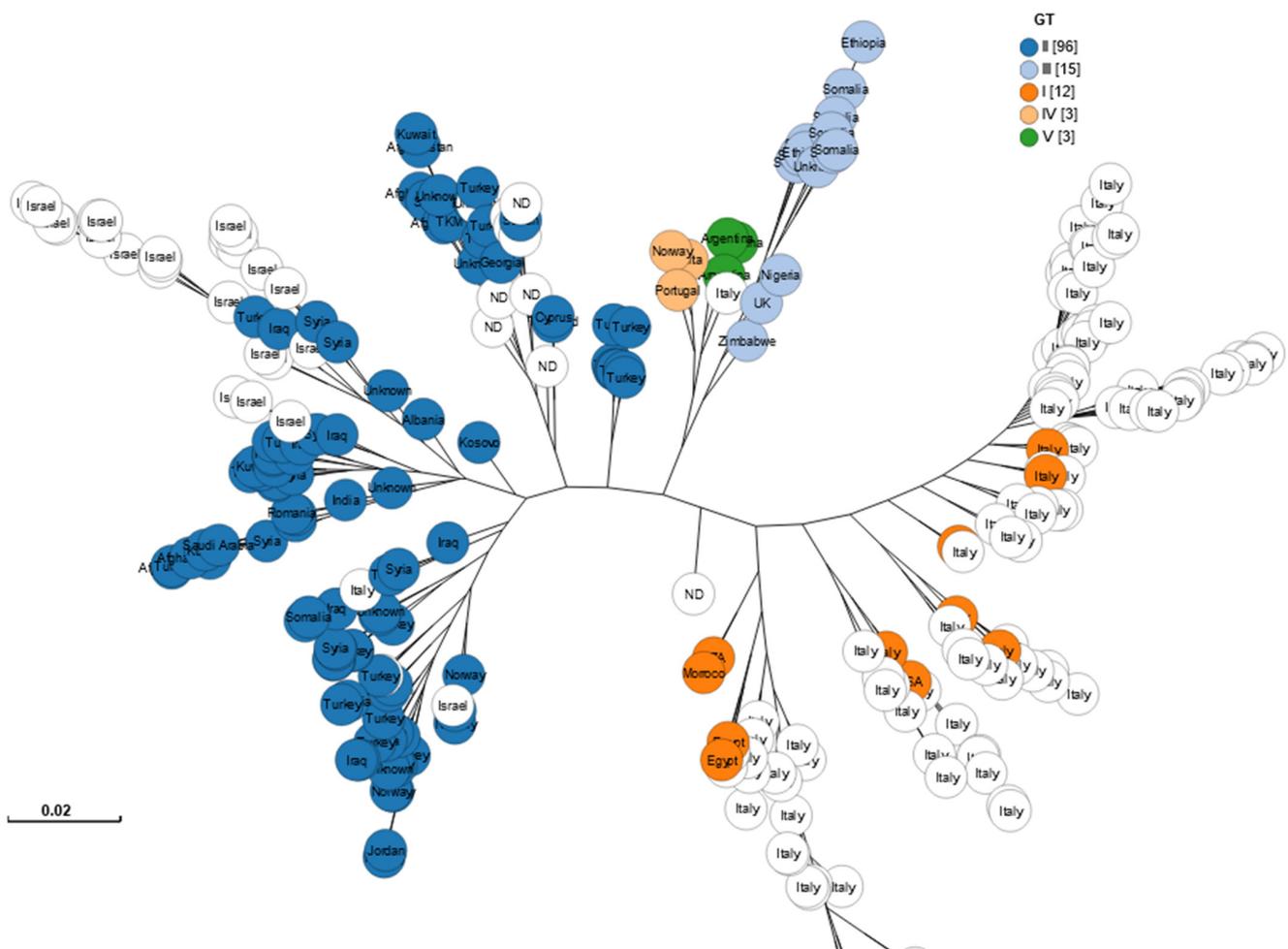


Figure 4. Maximum-likelihood tree based on pangenome. Coloured by genotype (GT) and nodes labelled by country.

3.4. Virulome Analysis

The assembled genomes matched to the 42 *B. melitensis* virulence genes in the VFDB set at levels of 99% or greater, with most genes present in 100% of the isolates. (see Supplemental Table S3). All 10 other virulence genes not included in the VFDB database were present across all the assembled genomes. SNP variation analysis of these 10 virulence genes revealed that 7 of the 10 genes assessed were completely conserved, without any mutations (Table 1). The *vceC* gene had a single mutation (position 216) which was observed specifically in the genomes of the genotype IIa sub-genotype cluster. The *mviN* gene had two mutations (positions 104 and 211) which were both observed only in the genomes of the genotype I cluster. The *ure* gene showed the greatest variation, with 6 mutations. Mutation in position 169 was observed in genomes of the genotype I cluster, while a mutation in position 1188 was observed in genomes of both the genotype I and genotype II clusters. Mutations in positions 328, 515, and 863 were observed in genomes of the genotype II cluster. Finally, a mutation in position 785 was specifically observed in the genomes of the genotype IIb sub-genotype cluster (mainly among the Israeli/Middle east isolates).

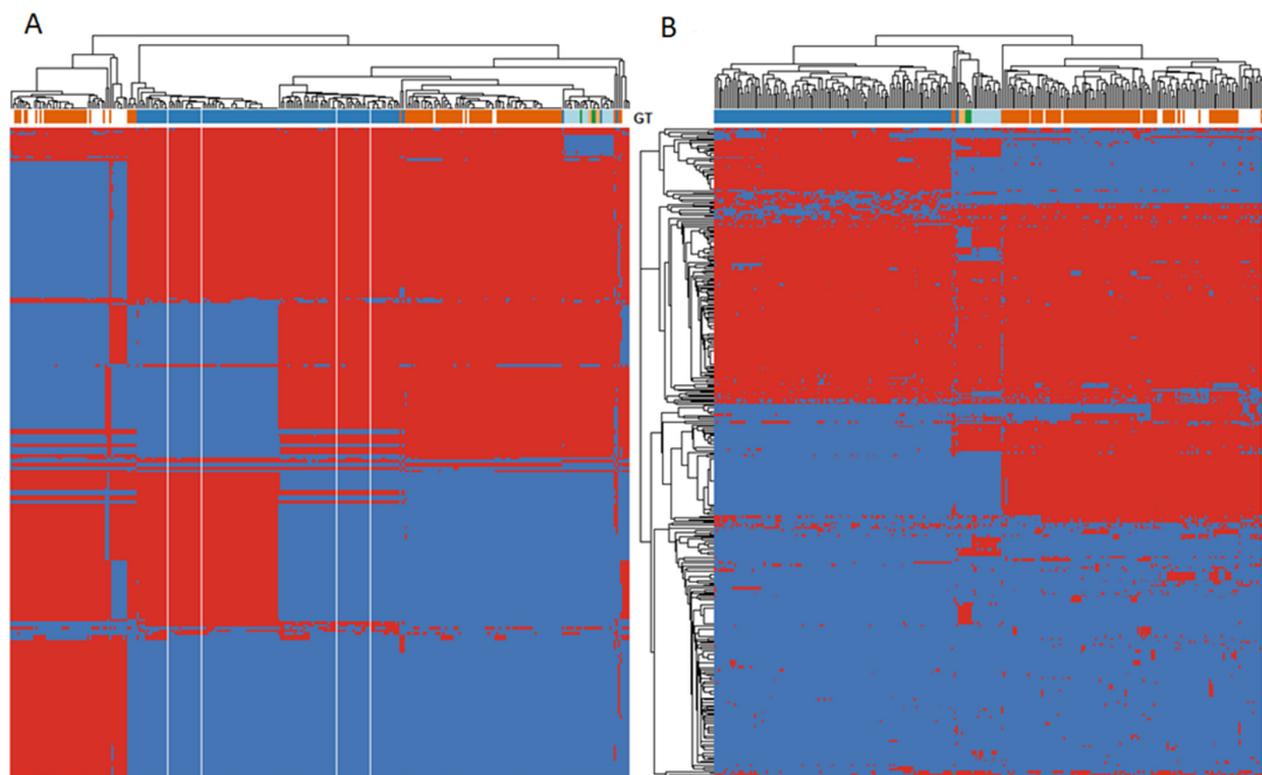


Figure 5. A pangenome analysis of 275 *B. melitensis* genome assemblies. **(A)** clustering of the accessory pangenome by gene presence (red) and absence (blue), with isolate genotypes (GT) highlighted above the heatmap. **(B)** clustering of the whole pangenome structure according to gene triplet presence (red) and absence (blue), with isolate genotypes (GT) highlighted above the heatmap.

Table 1. Mutations in 10 virulence genes and their correlation with the global genotypes.

Gene	Base Position in Gene	Genotype (GT)					Note
		I	II	III	IV	V	
<i>ure</i>	169	T	G	G	G	G	
	328	C	T	C	C	C	
	515	C	A	C	C	C	
	785	C	T	C	C	C	in GTIIb (Israeli/Middle East) sub-genotype
	863	C	T	C	C	C	
	1188	G	G	A	A	A	
<i>mviN</i>	104	A	G	G	G	G	
	211	G	A	A	A	A	
<i>vceC</i>	216	G	C	G	G	G	In GTIIa sub-genotype
<i>omp19</i>	none						
<i>prpA</i>	none						
<i>betB</i>	none						
<i>bpe275</i>	none						
<i>bspB</i>	none						
<i>perA</i>	none						
<i>manA</i>	none						

4. Discussion

This analysis makes use of available short-read sequences for *B. melitensis* available in NCBI Genbank and demonstrates the potential of whole genome sequence analysis of variation in functional *Brucella* genes that could relate to virulence. We expanded on previously published analyses by including more genomes (including new isolates from a geographically restricted region), performing a consistent process of genome assembly, searching for additional metadata on NCBI isolates, and analyzing variation in the virulome across the global pool of genomes. To our knowledge, the comparison of a large number of *B. melitensis* genomes with the VFDB database has not been previously reported. The 275 genomes that we assembled were predominantly from human isolates. Our visualizations of whole genome clustering match closely with previous genotyping schemes for *Brucella melitensis*. We found that the NCBI database has an unequal distribution of samples between human and non-human hosts across regions. In terms of virulence, our comparison to the VFDB database found that the *B. melitensis* genome does not display variability globally with respect to the presence or absence of putative *B. melitensis* virulence genes included in the VFDB database. The same was true when we analyzed the presence versus absence of 10 additional putative virulence genes recently reported in the literature. At the same time, our pilot analysis found SNP variation in three of these ten virulence genes. This variation tended to cluster according to genotype classification.

The finding that known and suspected virulence genes appear to occur at high frequency (>99%) across all genomes we studied agrees with previous studies that also reported a high frequency of virulence gene presence for a limited number of genes studied. At the same time, we were able to demonstrate SNP variation in a minority of potential virulence genes. This reflects the value of whole genome sequence approaches to the question of whether different strains of *Brucella melitensis* could differ in virulence. This approach can screen for variation in a large number of candidate genes, and for *Brucella*, it appears that SNP variation is more frequent than gene absence. While it was beyond the scope of this study to test the functional significance of the identified SNPs, our findings demonstrate the ability of WGS to identify multiple SNPs in *Brucella* virulence genes rapidly. Our analysis also demonstrated that one genotype (genotype II) showed a greater amount of SNP variation in tested virulence genes compared to other genotypes. This suggests that genotype II could potentially show greater variation in phenotypic virulence than other genotypes. At the same time, this study shows that WGS alone has limitations in predicting the function of particular genes, and that there is a need to follow up findings such as those in this study with appropriate *in vitro* and clinical studies.

While *Brucella melitensis* is a pathogen that causes serious disease in both livestock and humans, at present in the NCBI SRA database there are more sequences from human isolates compared to animal origin isolates. In some regions such as Europe, however, the available sequences are predominantly from animal cases. The differential sampling between human and animal origin isolates may reflect differential resources available in human vs. animal surveillance and reference laboratories, as well as the fact that zoonotic transmission from animals to humans may be more common in certain regions. Since understanding transmission and virulence is important in both humans and livestock, further research should aim to ensure adequate sampling and sequencing of *Brucella* isolates from both humans and non-human species.

Our analysis confirmed the findings of other investigators showing clustering of genomes by geographic region, and in addition, we found that even in a small geographic area, such as Southern Israel, the origin of the 27 novel isolates, there was further clustering into at least two major groups separated by slightly more than 100 SNPs. This indicates that there could be some variation in phenotypic qualities even in a small area that could have clinical significance. Such variation could also allow for epidemiological tracking of disease spread.

These novel findings demonstrate the feasibility of using WGS and pan-genome analysis to identify potential virulome variation that may have clinical significance for

the management and prevention of disease in both humans and animals. We demonstrate that while at the global scale, many virulence genes are present without significant presence/absence variation, some genes exhibit SNP variation and that even at a small geographic scale, there can be variation between strains of the organisms. These methods can now be applied to look at other candidate virulence genes and identify novel virulence genes as well.

A limitation of genomic analysis is that genotypic characterization alone may not fully predict phenotypic aspects such as virulence. To further explore the clinical significance of genomic variability in the pathogen (as well as host determinants of susceptibility) for this challenging zoonotic infection, there is a need to collect and associate more complete metadata with genomic information, including clinical data regarding disease severity and involved organ systems, as well as demographic factors and outcomes such as treatment failure and relapse. Currently, accessions for genome sequences in NCBI GenBank contain only minimal metadata. Therefore, there is a need for scientific efforts to assemble such clinical data, in large patient cohorts, and combine this information with whole genome and pan-genome analyses to better understand drivers of virulence and opportunities for improved management and prevention of disease.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/microbiolres12030050/s1>, Table S1—Sequences, Table S2—Reference Strains, Table S3—Virulome Analysis, Table S4—Virulence Genes, Table S5—Genome Statistics.

Author Contributions: Conceptualization, J.M.-G. and P.R.; methodology, A.G., Y.M., M.C.R., B.Z. and J.M.-G.; software, Y.M., A.G.; validation, Y.M., M.C.R.; formal analysis, P.R., Y.M., S.B.; investigation, L.N. and S.B.; resources, P.R., A.G. and J.M.-G.; data curation, Y.M. and A.G.; writing—original draft preparation, P.R., J.M.-G. and Y.M.; writing—review and editing, B.Z., L.N., S.B.-S., Y.Y., M.G., O.S., N.D., and D.K.; visualization, Y.M.; supervision, J.M.-G. and P.R.; project administration, J.M.-G. and P.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part from the Chief Scientist Office of the Ministry of Agriculture and Rural Development, Israel.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and the protocols for the Israeli study were reviewed and approved by the Institutional Review Board of the Soroka University Medical Center (protocol #0292-17-SOR, 4 January 2018).

Informed Consent Statement: Not Applicable.

Data Availability Statement: Whole genomes produced by this research are being deposited into NCBI Genbank. Reads are available in BioProject PRJNA608659.

Acknowledgments: The authors express appreciation to Vickie Ramirez for assistance with the preparation of the manuscript, and Lu Wang, Theo Bammler, Gemina Garland Lewis, Nicholas Yang, and Gillian Tarr for assistance in the preliminary stages.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dean, A.S.; Crump, L.; Greter, H.; Hattendorf, J.; Schelling, E.; Zinsstag, J. Clinical Manifestations of Human Brucellosis: A Systematic Review and Meta-Analysis. *PLoS Negl. Trop. Dis.* **2012**, *6*, e1929. [[CrossRef](#)] [[PubMed](#)]
2. Doganay, G.D.; Doganay, M. Brucella as a Potential Agent of Bioterrorism. *Recent Pat. Antiinfect. Drug Discov.* **2013**, *8*, 27–33. [[CrossRef](#)]
3. Dean, A.S.; Crump, L.; Greter, H.; Schelling, E.; Zinsstag, J. Global Burden of Human Brucellosis: A Systematic Review of Disease Frequency. *PLoS Negl. Trop. Dis.* **2012**, *6*, e1865. [[CrossRef](#)] [[PubMed](#)]
4. Bechtol, D.; Carpenter, L.R.; Mosites, E.; Smalley, D.; Dunn, J.R. *Brucella melitensis* Infection Following Military Duty in Iraq. *Zoonoses Public Health* **2011**, *58*, 489–492. [[CrossRef](#)] [[PubMed](#)]
5. Olsen, S.C.; Palmer, M.V. Advancement of Knowledge of *Brucella* Over the Past 50 Years. *Vet. Pathol.* **2014**, *51*, 1076–1089. [[CrossRef](#)] [[PubMed](#)]

6. Hashemifar, I.; Yadegar, A.; Jazi, F.M.; Amirzozafari, N. Molecular prevalence of putative virulence-associated genes in *Brucella melitensis* and *Brucella abortus* isolates from human and livestock specimens in Iran. *Microb. Pathog.* **2017**, *105*, 334–339. [[CrossRef](#)] [[PubMed](#)]
7. Mancilla, M. Smooth to Rough Dissociation in *Brucella*: The Missing Link to Virulence. *Front. Cell Infect. Microbiol.* **2016**, *5*, 98. [[CrossRef](#)]
8. Holzapfel, M.; Girault, G.; Keriell, A.; Ponsart, C.; O’Callaghan DM, V. Comparative Genomics and in vitro Infection of Field Clonal Isolates of *Brucella melitensis* Biovar 3 Did Not Identify Signature of Host Adaptation. *Front. Microbiol.* **2018**, *9*, 2505. [[CrossRef](#)]
9. Yang, X.; Piao, D.; Mao, L.; Pang, B.; Zhao, H.; Tian, G.; Jiang, H.; Kan, B. Whole-genome sequencing of rough *Brucella melitensis* in China provides insights into its genetic features. *Emerg. Microbes Infect.* **2020**, *9*, 2147–2156. [[CrossRef](#)]
10. Głowacka, P.; Żakowska, D.; Naylor, K.; Niemcewicz, M.; Bielawska-Drózd, A. *Brucella*—Virulence Factors, Pathogenesis and Treatment. *Pol. J. Microbiol.* **2018**, *67*, 151–161. [[CrossRef](#)]
11. O’Callaghan, D.; Whatmore, A.M. *Brucella* genomics as we enter the multi-genome era. *Brief. Funct. Genomics* **2011**, *10*, 334–341. [[CrossRef](#)]
12. Brambila-Tapia, A.J.; Armenta-Medina, D.; Rivera-Gomez, N.; Perez-Rueda, E. Main functions and taxonomic distribution of virulence genes in *Brucella melitensis* 16 M. *PLoS ONE* **2014**, *9*, e100349. [[CrossRef](#)] [[PubMed](#)]
13. Cannella, A.P.; Tsolis, R.M.; Liang, L.; Felgner, P.; Saito, M.; Gotuzzo, E.; Sette, A.; Vinetz, J.M. Antigen-specific acquired immunity in human brucellosis: Implications for diagnosis, prognosis, and vaccine development. *Front. Cell Infect. Microbiol.* **2012**, *2*, 1. [[CrossRef](#)]
14. Sangari, F.J.; Cayón, A.M.; Seoane, A.; García-Lobo, J.M. *Brucella abortus* ure2 region contains an acid-activated urea transporter and a nickel transport system. *BMC Microbiol.* **2010**, *10*, 1–12. [[CrossRef](#)]
15. Mancilla, M.; López-Goñi, I.; Moriyón, I.; Zárraga, A.M. Genomic Island 2 is an unstable genetic element contributing to *Brucella* lipopolysaccharide spontaneous smooth-to-rough dissociation. *J. Bacteriol.* **2010**, *192*, 6346–6351. [[CrossRef](#)]
16. Ding, J.; Pan, Y.; Jiang, H.; Cheng, J.; Liu, T.; Qin, N.; Yang, Y.; Cui, B.; Chen, C.; Liu, C.; et al. Whole genome sequences of four *Brucella* strains. *J. Bacteriol.* **2011**, *193*, 3674–3675. [[CrossRef](#)]
17. Mirnejad, R.; Jazi, F.M.; Mostafaei, S.; Sedighi, M. Molecular investigation of virulence factors of *Brucella melitensis* and *Brucella abortus* strains isolated from clinical and non-clinical samples. *Microb. Pathog.* **2017**, *109*, 8–14. [[CrossRef](#)]
18. Georgi, E.; Walter, M.C.; Pfalzgraf, M.T.; Northoff, B.H.; Holdt, L.M.; Scholz, H.C.; Zoeller, L.; Zange, S.; Antwerpen, M.H. Whole genome sequencing of *Brucella melitensis* isolated from 57 patients in Germany reveals high diversity in strains from Middle East. *PLoS ONE* **2017**, *12*, e0175425. [[CrossRef](#)]
19. Janowicz, A.; De Massis, F.; Ancora, M.; Cammà, C.; Patavino, C.; Battisti, A.; Prior, K.; Harmsen, D.; Scholz, H.; Zilli, K.; et al. Core genome multilocus sequence typing and single nucleotide polymorphism analysis in the epidemiology of *Brucella melitensis* infections. *J. Clin. Microbiol.* **2018**, *56*, e00517-18. [[CrossRef](#)] [[PubMed](#)]
20. Pelerito, A.; Nunes, A.; Nuncio, M.S.; Gomes, J.P. Genome-scale approach to study the genetic relatedness among *Brucella melitensis* strains. *PLoS ONE* **2020**, *15*, e0229863. [[CrossRef](#)] [[PubMed](#)]
21. Zhu, X.; Zhao, Z.; Ma, S.; Guo, Z.; Wang, M.; Li, Z.; Liu, Z. *Brucella melitensis*, a latent “travel bacterium,” continual spread and expansion from Northern to Southern China and its relationship to worldwide lineages. *Emerg. Microbes Infect.* **2020**, *9*, 1618–1627. [[CrossRef](#)] [[PubMed](#)]
22. Tan, K.K.; Tan, Y.C.; Chang, L.Y.; Lee, K.W.; Nore, S.S.; Yee, W.Y.; Isa, M.N.; Jafar, F.L.; Hoh, C.C.; AbuBakar, S. Full genome SNP-based phylogenetic analysis reveals the origin and global spread of *Brucella melitensis*. *BMC Genomics* **2015**, *16*, 1–11. [[CrossRef](#)] [[PubMed](#)]
23. Johansen, T.B.; Scheffer, L.; Jensen, V.K.; Bohlin, J.; Feruglio, S.L. Whole-genome sequencing and antimicrobial resistance in *Brucella melitensis* from a Norwegian perspective. *Sci. Rep.* **2018**, *8*, 1–9. [[CrossRef](#)]
24. Pisarenko, S.V.; Kovalev, D.A.; Volynkina, A.S.; Ponomarenko, D.G.; Rusanova, D.V.; Zharinova, N.V.; Khachaturova, A.A.; Tokareva, L.E.; Khvoynova, I.G.; Kulichenko, A.N. Global evolution and phylogeography of *Brucella melitensis* strains. *BMC Genomics* **2018**, *19*, 1–10. [[CrossRef](#)]
25. Liu, Z.G.; Cao, X.A.; Wang, M.; Piao, D.R.; Zhao, H.Y.; Cui, B.Y.; Jiang, H.L.Z. Whole-Genome Sequencing of a *Brucella melitensis* Strain (BMWS93) Isolated from a Bank Clerk and Exhibiting Complete Resistance to Rifampin. *Microbiol. Resour. Announc.* **2019**, *8*, e01645-18. [[CrossRef](#)]
26. Salmon-Divon, M.; Yeheskel, A.; Kornspan, D. Genomic analysis of the original elberg *Brucella melitensis* rev.1 vaccine strain reveals insights into virulence attenuation. *Virulence* **2018**, *9*, 1436–1448. [[CrossRef](#)]
27. Jolley, K.A.; Maiden, M.C.J. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **2010**, *11*, 1–11. [[CrossRef](#)]
28. Silva, M.; Machado, M.P.; Silva, D.N.; Rossi, M.; Moran-Gilad, J.; Santos, S.; Ramirez, M.; Carriço, J.A. chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. *Microb. Genomics* **2018**, *4*, e000166. [[CrossRef](#)]
29. NHC Key Laboratory of Systems Biology of Pathogens. VFDB Virulence Factors Database [Internet]. 2003. Available online: <http://www.mgc.ac.cn> (accessed on 12 September 2019).
30. Slater, G.S.C.; Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* **2005**, *6*, 1–11. [[CrossRef](#)]

31. Tonkin-Hill, G.; MacAlasdair, N.; Ruis, C.; Weimann, A.; Horesh, G.; Lees, J.A.; Gladstone, R.A.; Lo, S.; Beaudoin, C.; Floto, R.A.; et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **2020**, *21*, 1–21. [[CrossRef](#)]
32. Nguyen, L.T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)]
33. Ushey, K.; Lin, L.; Finak, G. COMPASS: Combinatorial Polyfunctionality Analysis of Single Cells. Available online: <https://www.bioconductor.org/packages/release/bioc/html/COMPASS.html> (accessed on 9 March 2021).
34. Nakamura, T.; Yamada, K.D.; Tomii, K.; Katoh, K. Parallelization of MAFFT for Large-Scale Multiple Sequence Alignments. *Bioinformatics* **2018**, *34*, 2490–2492. [[CrossRef](#)]
35. Page, A.J.; Taylor, B.; Delaney, A.J.; Soares, J.; Seemann, T.; Keane, J.A.; Harris, S.R. SNP-Sites: Rapid Efficient Extraction of SNPs from Multi-FASTA Alignments. *Microb. Genom.* **2016**, *2*, e000056. [[CrossRef](#)]
36. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve Years of SAMtools and BCFtools. *GigaScience* **2021**, *10*, giab008. [[CrossRef](#)] [[PubMed](#)]
37. Sun, M.; Jing, Z.; Di, D.; Yan, H.; Zhang, Z.; Xu, Q.; Zhang, X.; Wang, X.; Ni, B.; Sun, X.; et al. Multiple locus variable-number tandem-repeat and single-nucleotide polymorphism-based *Brucella* typing reveals multiple lineages in *Brucella melitensis* currently endemic in China. *Front. Vet. Sci.* **2017**, *4*, 215. [[CrossRef](#)] [[PubMed](#)]