Received: 7 June 2022 | Revised: 7 October 2022 | Accepted: 11 November 2022

DOI: 10.1111/tbed.14768

Transboundary and Emerging Diseases

WILEY

ORIGINAL ARTICLE

Tracking the distribution, genetic diversity and lineage of *Brucella melitensis* recovered from humans and animals in Egypt based on core-genome SNP analysis and in silico MLVA-16

Katharina Holzer[1] | Gamal Wareth[2,4] | Mohamed El-Diasty[3] | Nour H. Abdel-Hamid[3] | Mahmoud E. R. Hamdy[3] | Shawky A. Moustafa[4] | Jörg Linde[2] | Felix Bartusch[5] | Mostafa Y. Abdel-Glil[2] | Ashraf E. Sayour[3] | Essam M. Elbauomy[3] | Mohamed Elhadidy[6,7] | Falk Melzer[2] | Wolfgang Beyer[1]

[1]Department of Livestock Infectiology and Environmental Hygiene, Institute of Animal Science, University of Hohenheim, Stuttgart, Germany

[2]Friedrich-Loeffler Institut, Institute for Bacterial Infections and Zoonoses, Jena, Germany

[3]Animal Health Research Institute, Agricultural Research Center, Cairo, Egypt

[4]Department of Bacteriology, Immunology, and Mycology, Faculty of Veterinary Medicine, Benha University, Toukh, Moshtohor, Egypt

[5]IT Center (ZDV), High-Performance and Cloud Computing Group, University of Tuebingen, Tuebingen, Germany

[6]Biomedical Sciences Program, University of Science and Technology, Zewail City of Science and Technology, Giza, Egypt

[7]Department of Bacteriology, Mycology, and Immunology, Faculty of Veterinary Medicine, Mansoura University, Mansoura, Egypt

Correspondence
Gamal Wareth, Friedrich-Loeffler Institut, Institute for Bacterial Infections and Zoonoses, Naumburger Str. 96a, 07743 Jena, Germany.
Email: gamal.wareth@fli.de
Katharina Holzer, Department of Livestock Infectiology and Environmental Hygiene, Institute of Animal Science, University of Hohenheim, Stuttgart, Germany.
Email: katharina.holzer@uni-hohenheim.de

Funding information
LEAP-AGRI project MUSCBCEA, Grant/Award Number: 2817LEAP01; German Research Foundation, Grant/Award Number: INST 37/935-1 FUGG; Emerging Infectious Disease Research, Grant/Award Number: GTR4-17-63138-0)

Abstract

Brucellosis is one of the most common neglected zoonotic diseases globally, with a public health significance and a high economic loss in the livestock industry caused by the bacteria of the genus *Brucella*. In this study, 136 Egyptian *Brucella melitensis* strains isolated from animals and humans between 2001 and 2020 were analysed by examining the whole-core-genome single-nucleotide polymorphism (cgSNP) in comparison to the in silico multilocus variable number of tandem repeat analysis (MLVA-16). Almost all Egyptian isolates were belonging to the West Mediterranean clade, except two isolates from buffalo and camel were belonging to the American and East Mediterranean clades, respectively. A significant correlation between the human case of brucellosis and the possible source of infection from animals was found. It seems that several outbreak strains already existing for many years have been spread over long distances and between many governorates. The cgSNP analysis, in combination with epidemiological metadata, allows a better differentiation than the MLVA-16 genotyping method and, hence, the source definition and tracking of outbreak strains. The MLVA based on the currently used 16 markers is not suitable for this task. Our results revealed 99 different cgSNP genotypes with many different outbreak strains, both older and widely

distributed ones and rather newly introduced ones as well. This indicates several different incidents and sources of infections, probably by imported animals from other countries to Egypt. Comparing our panel of isolates to public databases by cgSNP analysis, the results revealed near relatives from Italy. Moreover, near relatives from the United States, France, Austria and India were found by in silico MLVA.

KEYWORDS
Brucella melitensis, genotyping, MLVA, outbreak analysis, SNP analysis

1 | INTRODUCTION

Brucellosis is a globally distributed zoonotic disease caused by the bacteria of the genus *Brucella*, infecting mammals, including humans (Cutler et al., 2005). Out of the accepted 12 nomo-species, *Brucella melitensis*, *Brucella abortus* and *Brucella suis* are the most important human pathogens with varying pathogenicity (Chiliveru et al., 2015; Dentinger et al., 2015). *B. melitensis* and *B. suis*, except for biovar 2, are the most virulent species, whereas *B. abortus* provokes milder illness (Galińska & Zagórski, 2013). In general, all brucellae are genetically highly related to each other (Ficht, 2010). Small ruminants and bovines are the predominant hosts for *B. melitensis* and *B. abortus*, respectively, while cross-species transmission has been proven (Hashemifar et al., 2017; De Massis et al., 2019). *B. melitensis* is by far the most observed causative agent of human infection. Human infection is dependent on the animal reservoir and mainly occurs by direct contact with infected animals or by consuming unpasteurized dairy products and raw milk (Pappas et al., 2005). Because *Brucella* can be transmitted via aerosols, the fear exists that it might be used in bioterrorism events (Godfroid et al., 2011). Control of livestock and human infection requires an appropriately structured brucellosis surveillance programme which is based on highly discriminatory methods to trace back and determine the sources of infection. In Egypt, the incidence had increased steadily since 1960 when Friesian cows were imported (Wareth et al., 2020). Since then, the disease has been recognized as one of the endemic and most important livestock diseases in the country, despite the establishment of a brucellosis eradication programme in 1981 (Refai, 2002). Reasons are low-level biosafety in farms, high livestock and human density, traditional food habits and limited success in the official control programme (Hegazy et al., 2011). The uncontrolled movement of animals for grazing or trade further enhances the spread of the disease throughout the country (Wareth et al., 2014). Clarifying the status quo of brucellosis has a high impact on the health sector.

To get an overview of the epidemiological situation of this notorious bacterial agent, a total of 136 *B. melitensis* isolates were used in this study to compare outbreak strains in Egypt. The diversity, distribution and possible spread of outbreak strains of *B. melitensis* in Egypt from 2006, 2007 and 2010 to 2020, plus one isolate each from 2001 and 2002, were investigated using both whole-genome sequencing (WGS)-based single-nucleotide polymorphism (SNP) analysis and in silico MLVA-16. This also enabled comparing both methods for the suitability for outbreak investigation. Additionally, isolates were compared to entries from other countries in public databases. As the first *B. abortus* data from Egypt based on WGS are available (Holzer et al., 2021; Khan et al., 2021), only a few data have been published about *B. melitensis* (Khan et al., 2021).

2 | MATERIALS AND METHODS

2.1 | Origin of *Brucella* isolates

A total of 136 *B. melitensis* isolates were recovered from both animals and humans from various governorates from 2006 to 2020, except in 2008 and 2009, plus one isolate each from 2002 and 2003. A total of 13 human isolates were obtained from the blood of *Brucella*-infected patients. The remaining 123 isolates were isolated from animals, including buffalos ($n = 22$), camels ($n = 2$), cows ($n = 66$), goats ($n = 13$) and sheep ($n = 20$). Recovered isolates are listed in Tables 1 and S1, including the metadata.

The Animal Health Research Institute ethical committee in Giza approved the protocol of this study (ethical code Ref. No. 165870) in alignment with WHO guidelines and the Helsinki Declaration. The consent of the patient was sought. Written informed consent was obtained from all participants. All patient data were blinded.

2.2 | *Brucella* isolation from sample materials and DNA extraction

Swab samples were smeared on *Brucella*-selective agar (Oxoid GmbH, Wesel, Germany) containing 2500 IU polymyxin B, 12,500 IU bacitracin, 2.5-mg nalidixic acid, 50,000 IU nystatin, 10-mg vancomycin and 25-mg natamycin in a total volume of 500-ml *Brucella*-selective agar and incubated at 37°C in the presence of 5%–10% CO_2 and examined daily for bacterial growth for 5 days. *Brucella* isolates were subcultured three times to obtain pure *Brucella* colonies. Suspected *Brucella* colonies from a selective agar medium were picked from a selective agar medium and grown in *Brucella* broth for around 3 days for DNA extraction. DNA was extracted using Qiagen DNeasy Blood & Tissue Kits (Qiagen, Hilden, Germany) according to the manufacturer's guidelines. The DNA of 48 isolates was extracted directly after

TABLE 1 List of number of isolates, type of samples and host species of each strain

Number of isolates	Type of sample	Host
30	Milk	2 buffaloes, 3 sheep, 4 goats, 5 cows, 16 cattle and 1 camel
5	Foetal abomasal content	2 sheep, 2 cows and 1 buffalo
1	Foetal liver	Sheep
1	Liver	Sheep
1	Lung	Goat
26	Unspecified lymph nodes	10 buffaloes, 10 cows, 4 goats and 2 sheep
9	Retropharyngeal lymph nodes	3 sheep and 6 cows
27	Supramammary lymph nodes	5 buffaloes, 18 cows, 1 goat and 3 sheep
10	Unspecified organs	2 buffaloes, 3 sheep, 4 cows and 1 goat
4	Spleen	2 goats and 2 cows
4	Stomach content	2 sheep and 2 cows
4	Uterine discharge	1 cow, 1 camel and 2 buffaloes

cultivation without preheating, and the remaining 89 isolates were heat-inactivated before.

2.3 | Whole-genome sequencing and bioinformatics procedure of the raw reads

Default genomic library preparation and total genomic DNA sequencing were performed by Eurofins Genomics GmbH (Konstanz, Germany). The libraries were sequenced using Illumina NovaSeq 6000, theoretically producing at least 5 million 151-bp paired-end reads. High-throughput short-read sequencing yielded an average of 11,626,432 reads per isolate (min 1,136,924, max 25,638,672), leading to a mean coverage of 525 (min 51, max 1160). For analysing the WGS data in a standardized and automated manner, the Linux-based bioinformatics WGSBAC (v.2.1) pipeline (https://gitlab.com/FLI_Bioinfo/WGSBAC/-/tree/version2, accessed on 22 February 2022) was used. The pipeline input consisted of a metadata file and the Illumina paired-end FASTQ files. Raw coverage for each data set was calculated by the number of reads multiplied by their average read length and divided by the genome size. Contigs were assembled using Shovill v.1.0.4 (https://github.com/tseemann/shovill, accessed on 24 February 2022), an optimizer for SPAdes assembler (Bankevich et al., 2012). Quality control of the assembled contigs was performed by QUAST v. 5.0.2 (Gurevich et al., 2013). The genomic characteristics are shown in Table S2. All FASTQ files were submitted to the National Center for Biotechnology Information (NCBI) under the BioProject number PRJNA773199 (https://ncbinlm.nih.gov.bioproject/PRNJA773199, accessed on 22 February 2022).

2.4 | Species determination

The species of *Brucella* isolates were determined by Bruce-ladder PCR (García-Yoldi et al., 2006; Kang et al., 2011; López-Goñi et al., 2008) in silico using the programme Geneious v.11.1.5 (https://www.geneious.com/, accessed on 16 May 2022) based on assembled contigs as calculated by the programme Shovill in the bioinformatics pipeline. The primers used are listed in Table S3.

2.5 | Core-genome SNP genotyping

In silico SNP calling was performed using Snippy (v. 4.6.0) with default parameters (https://github.com/tseemann/snippy, accessed on 22 February 2022) for the Egyptian isolates. *Detection* of SNPs was done on the core-genome, including the removal of rRNA genes. Core-genome SNP (cgSNPs) (Table S4) were called based on the alignment to the *B. melitensis* 16-M reference strain (GenBank accession numbers NC_003317.1 and NC_003318.1). cgSNP-based genotypes (cgSNPGTs) were defined using a maximum of one cgSNP difference.

2.6 | Calculation of trees and SDI

Cluster analysis, phylogenetic analysis, creation of minimum spanning trees (MSTs) and the determination of Simpson's diversity index (SDI) were performed with BioNumerics version 8.0 (Applied Maths, Belgium). One SNP and one Bruce marker were used to define a genotype for SDI. For the cgSNP dendrogram, a maximum parsimony (MP) tree was created based on the character data. For MLVA, a neighbour-joining (NJ) tree based on categorical data was created. Both trees were permutated 1000 times and rooted by maximum branch length. The MSTs are presented with logarithmic scaling.

2.7 | Canonical SNP assay

To determine the African, American, East Mediterranean or a not yet defined clade (probably the West Mediterranean [WM] clade according to personal communication with Jeffrey T. Foster, developer of the canonical SNP [canSNP] assay, University of Arizona) of *B. melitensis*, the canSNP assay was performed according to Foster et al. (2017), in silico using Geneious v.11.1.5. The yet undefined branch has been adopted as the WM branch in this work. Primer were matched to the complementary genome sequence using the contigs provided by the bioinformatics pipeline described in Section 2.3. As described by Foster et al. (2017), the GC-overhang for the primers is unnecessary for the in silico analysis and has therefore been omitted. After primer

annealing, the respective nucleotide position of the isolate can be determined.

2.8 | In silico MLVA-16

Multilocus variable number of tandem repeats (MLVA-16) genotyping was carried out in silico using MISTReSS (https://github.com/Papos92/MISTReSS, accessed on 22 February 2022) with primers adapted for *Brucella* (Sacchini et al., 2019). To avoid multiple primer binding sites, the forward primer sequence of Bruce21 was extended to (5′GGCAGTGGGGCAGTGAAGAATATGGTCGCTGCGCTCATGCGCAACCAAAACA-3′). The number of repeats at each locus was determined by the fragment size according to the published *Brucella* allele assignment table (Al Dahouk et al., 2007). All in silico MLVA results of the isolates mentioned in Section 2.1 are listed in Table S5. All MLVA profiles have been submitted to the MLVA database (https://microbesgenotyping.i2bc.paris-saclay.fr/databases/, accessed on 22 February 2022). MLVA-based genotypes (MLVAGTs) were defined using one marker difference.

2.9 | Comparison with entries in public databases

To compare the 136 Egyptian *B. melitensis* isolates, as described in Section 2.1, entries of *B. melitensis* strains were taken from the public databases and processed bioinformatically as described in Section 2.3. To ensure that the reads downloaded from NCBI Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra, accessed on 22 February 2022) and all assemblies downloaded from GenBank (https://www.ncbi.nlm.nih.gov/genbank/, accessed on 22 February 2022) originate from *B. melitensis* isolates, the average nucleotide identity (ANI) to the reference (*B. melitensis* 16 M) was computed. The corresponding assemblies calculated by Shovill were used for the reads downloaded from SRA. The tool pyANI v. 0.2.10 (https://github.com/widdowquinn/pyani#conda, accessed on 22 February 2022) (Pritchard et al., 2016) was used to compute a pairwise ANI and other metrics between all assemblies. MUMmer (NUCmer) v. 3.23 (https://github.com/mummer4/mummer, accessed on 22 February 2022) (Marçais et al., 2018) was used by pyani to align the input sequences. At least 99% nucleotide identity was accepted for ongoing analysis when aligned to the *B. melitensis* 16-M reference strain. Before compiling cgSNP analysis with snippy, the quality of the genomes downloaded from the SRA database was controlled by calculating average sequencing depth and percentage of bases with a Phred quality score of Q30. In addition, species assignment to *B. melitensis* was confirmed with the Mash software (Ondov et al., 2016) using the NCBI reference database. Snippy was then applied, and SNP positions that were conserved in at least 98% of all analysed genomes were included in the core-genome (Perrin et al., 2017). For MST calculation, gaps in the cgSNP matrix were ignored. Because our cgSNP analysis is based on the programme Snippy, which uses the raw Illumina reads for analysis, 704 available *B. melitensis* data sets with known origin could be used from SRA. The appropriate cgSNP distance table comprising 841 data sets in total (with our Egyptian isolates and the reference strain *B. melitensis* 16 M) is provided in Table S6.

Two hundred sixteen sequence assemblies of *B. melitensis* entries with known origin were downloaded from the GenBank database. In total, 710 isolates were used for MLVA downloaded from both databases, GenBank and SRA. The list of MLVA-strings is provided in Table S7. All entries of the isolates from both databases and their metadata are provided in Table S8.

2.10 | Geographical Map

The geographical map was created using the Free and Open Source QGIS version 3.12.2 Bucuresti (https://www.qgis.org/de/site/forusers/download.html, accessed on 22 February 2022), generated from GPS data in Google Maps and the layer EPSG: 4326 and WGS: 84. The appropriate QGIS layer download file is available in the Supporting Information section.

3 | RESULTS

3.1 | *Brucella melitensis* isolates

Isolates included in this study were obtained from 17 governorates over 20 years. Figure 1 shows the sampling region, mainly in the northern part of Egypt. An interactive GIS-Map was established, allowing for geographical mapping of the isolates with all available metadata.

A total of 136 outbreak-related isolates of *Brucella*, collected from 5 animal species and humans, were specified as *B. melitensis* by in silico Bruce-ladder PCR. Cross-species infection of *B. melitensis* in non-preferential hosts was according to the metadata detected for 2 camels, 22 buffaloes and 66 cows.

3.2 | Canonical SNP analysis

All collected Egyptian isolates belong to the WM clade except 15707 and HEA143. Isolate 15707 (2017, buffalo, Ismailia) belongs to the American clade, whereas isolate HEA143 (2016, camel, Red Sea governorate) belongs to the East Mediterranean clade. Isolate 15,707, which differs in 2914 cgSNPs from the next isolate of the WM branch, was isolated in the northwest of the country, and HEA143 is the only isolate from the south-eastern part, differing in 3087 cgSNPs to isolate 21,874 and in 2620 cgSNPs to isolate 15,707.

3.3 | Core-genome SNP genotyping for *B. melitensis* isolates assigned to the West Mediterranean Clade

The 135 *B. melitensis* isolates belonging to the WM clade are divided into 97 cgSNPGTs, and possible phylogenetic relationships are

FIGURE 1 The geographic origin of *Brucella melitensis* isolates from Egypt. In this map, the isolates in the associated governorates are shown. If the city within the governorate was available in the metadata (Table S1), the more precise geographic location of the city was taken. Otherwise, the geographic location in the middle of the governorate was taken. This map was established with QGIS v. 3.12.2

demonstrated in Figure 2. The isolates are ordered in 3 main clusters and 15 subclusters (SCs). The SDI based on one cgSNP distance is 0.9921. The cgSNP distances are provided in Table S4. Cluster 1 includes three SCs, SC1a, SC1b and SC1c. SC1a shows a high diversity of 17 isolates from animals collected in the period from 2011 to 2019, originating from different governorates. HEA151 and HEA152 from Faiyum (2018) show the most significant distance (72 cgSNPs) to their closest relative in this cluster (HEA51). HEA70 and 5319, which were isolated in 2013 and 2014, respectively, from Sharqia and Asyut (500 km apart), represent an identical cgSNPGT showing a minimum distance of 58 cgSNPs to their closest relative (HEA51). In SC1b, however, the 11 isolates appear very homogeneous, with a distance of up to 10 cgSNPs to each other, originating from 4 different governorates (Kafr El Sheikh, Faiyum, Gharbia and Ismailia). The oldest isolates in this SC are HEA6 and HEA58 from Faiyum (2013). SC1c can be divided into 2 groups (both from 2019), differing in at least 26 cgSNPs. Cluster 2 includes four SCs: SC2a–SC2d. SC2a, except HEA25, contains a homogeneous group of three human and two animal isolates with a maximum distance of seven cgSNPs to each other. HEA25 shows a distance of 10 cgSNPs to the other both animal isolates and represents the oldest isolate, dating from 2011, in this SC. In SC2b, the minimal difference is 13 cgSNPs between the isolates, among them 1 from a human. SC2c represents a high diversity. Nevertheless, HEA43 and 5317, which were isolated 8 years apart (2006 and 2014) in Beni Suef and Gharbia, respectively, show a completely identical cgSNPGT. SC2d contains five isolates differing from each other in up to seven cgSNPs, although isolated in 2006, 2010, 2012 and 2014 at four different places: Asyut, Qalyubia, Beni Suef and Giza. The single isolate 21,907 (Aswan, 2020) differs in 78 cgSNPs from its closest relative (5314/HEA31) in SC2d. Cluster 3 includes eight SCs, SC3a–SC3h. SC3a contains a group of four isolates from 2015 and 2016 from Damietta and Dakahlia that differ in up to five cgSNPs and another group of three isolates from Beheira and Ismailia in 2020 that differ in up to six cgSNPs. Between both groups, a distance of 10 cgSNPs is observable. HEA27 (cow, 2001, Monufia) differs from all other isolates in this SC in at least 12 cgSNPs. In SC3b, the isolates are very homogeneous. All originate from Monufia, 2018 and 2019, and show a maximum distance of seven cgSNPs. Furthermore, homogeneous groups are present in SC3c, SC3d, SC3e, SC3f and SC3g. SC3c consists of five isolates differing in up to five cgSNPs. All those isolates are from 2019, from Sharqia and Damietta

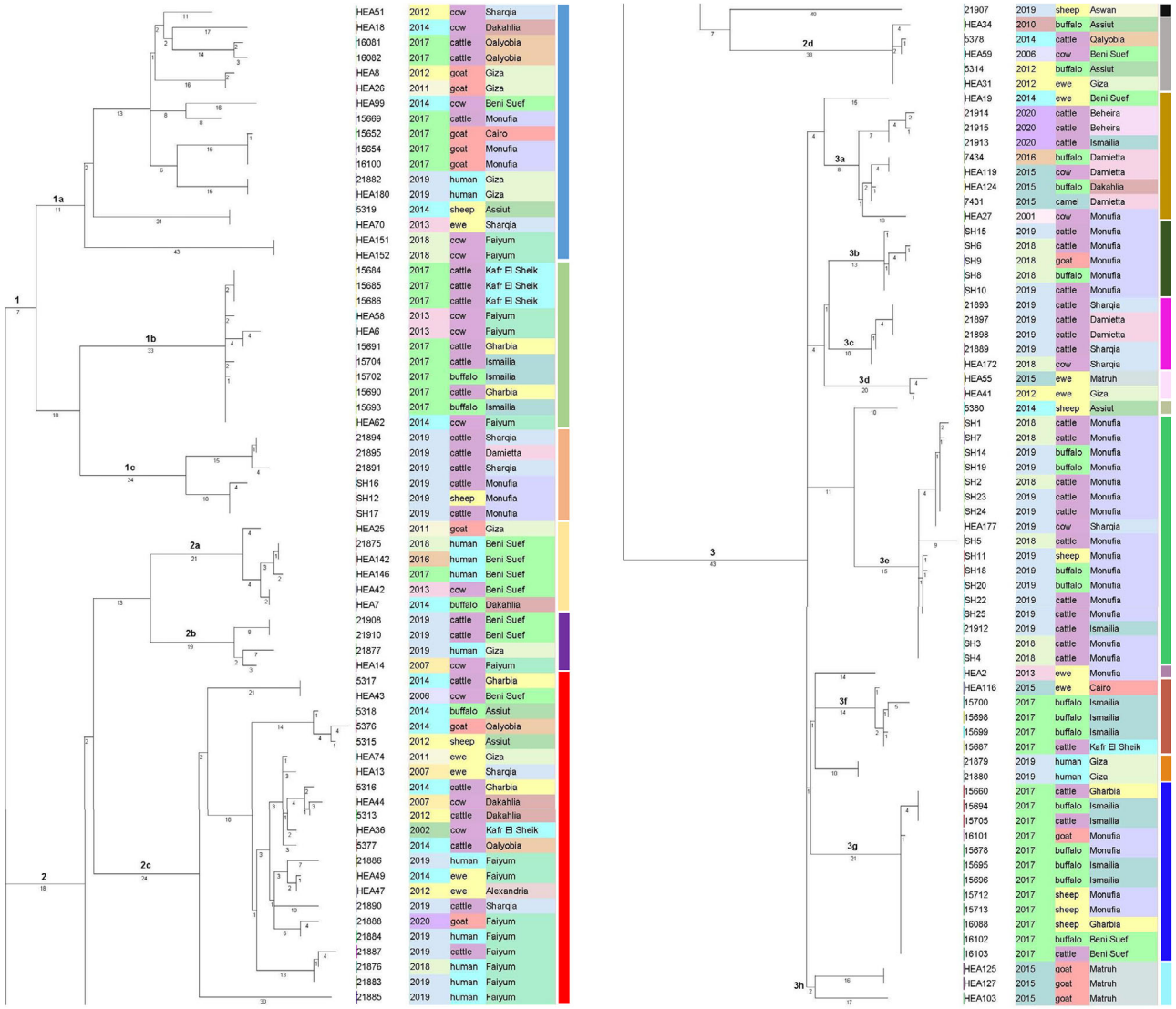

FIGURE 2 Maximum parsimony tree for 122 *Brucella melitensis* field isolates plus 13 human isolates of the West Mediterranean clade (according to the canonical single-nucleotide polymorphism [canSNP] analysis). Isolates 15,707 (American clade) and HEA143 (East Mediterranean clade) are not included. The tree is based on whole-genome sequencing (WGS) core-genome single-nucleotide polymorphism (cgSNP) analysis and was calculated in BioNumerics with 1000 permutations to get the highest resampling support and rooted by maximum branch length. *B. melitensis* 16 M was set as reference. The three columns represent the year of isolation, the host and the governorate, correspondingly. The last column represents the colour of the corresponding cluster. The numbers on the lines represent the cgSNP distance. This automatically generated figure from BioNumerics has been modified by Adobe Acrobat Pro 2017

(100 km apart), except HEA172 (2018). In SC3d, two isolates differing in only five cgSNPs were isolated in 2012 and 2015 from Giza and Matruh (450 km apart). SC3e contains 17 isolates from Monufia and one each from Sharqia and Ismailia, all isolated in 2018 and 2019 from cattle, buffalo and from one sheep. These isolates represent a distance of up to 8 cgSNPs, except SH5 from 2018, which differs in at least 10 cgSNPs from the other isolates. SC3f contains a group of five isolates with a maximum difference of nine cgSNPs from each other, isolated in 2015 and 2017 in Cairo, Ismailia and Kafr El Sheikh. SC3g consists of a homogenous group of 12 isolates, isolated in 2017 from Gharbia, Ismailia, Monufia and Beni Suef, differing in up to 5 cgSNPs, with 8 isolates sharing an identical cgSNPGT. The last SC, SC3h, contains 3 isolates with 2 different cgSNPGTs, differing in 33 cgSNPs. However, all three isolates originated from Matruh in 2015.

3.4 | Comparing the cgSNP analysis to MLVA genotyping of the isolates assigned to the West Mediterranean clade

Based on in silico MLVA with 16 markers, performed on the WGS data using MISTReSS and using 1 marker distance to define a MLVAGT, the panel of isolates represents 64 MLVAGTs when taking isolates 15,707 (American clade) and HEA143 (East Med. clade) into account.

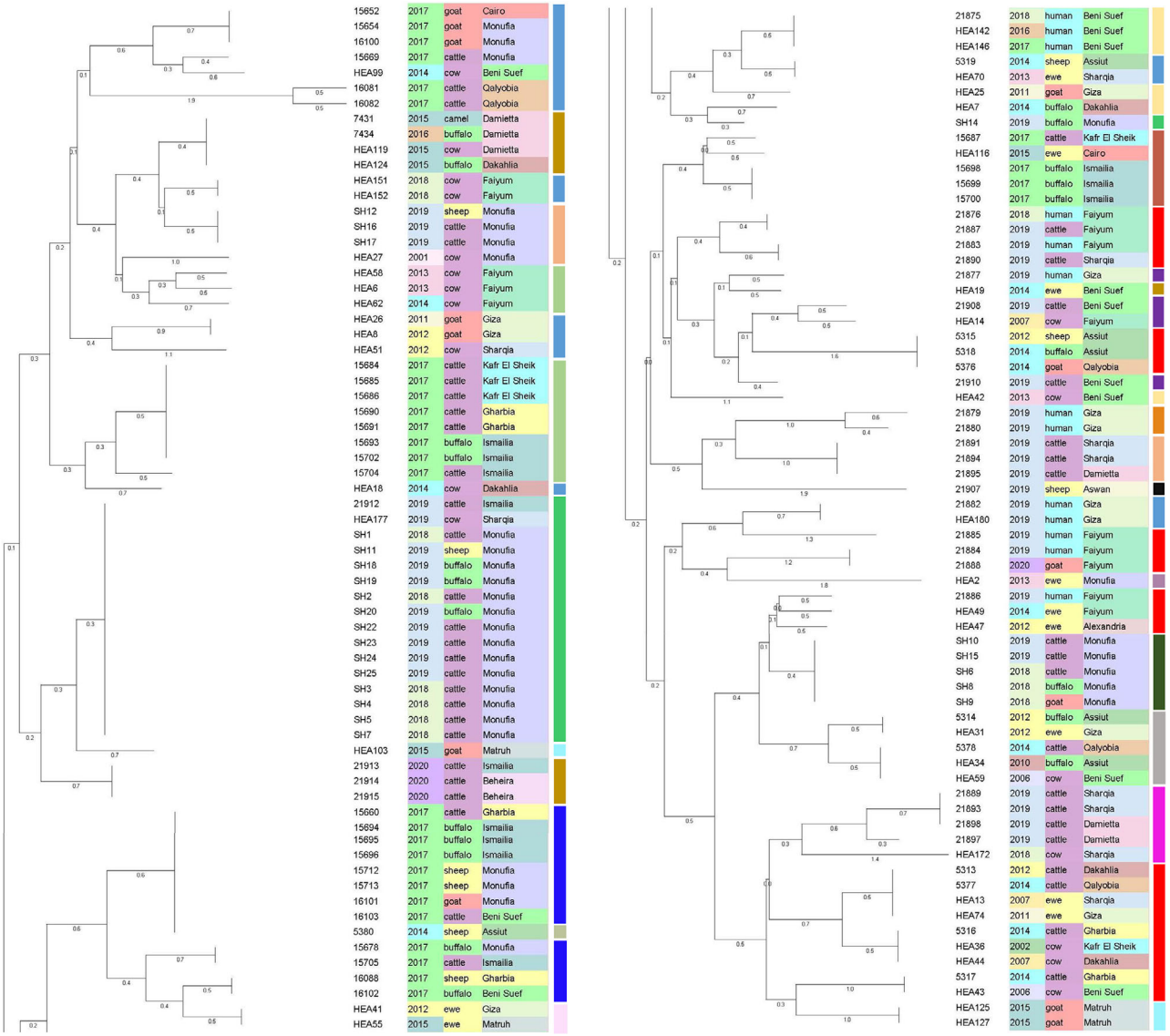

FIGURE 3 Neighbour-joining (NJ) tree for 122 *Brucella melitensis* field isolates plus 13 human isolates of the West Mediterranean clade according to the canonical single-nucleotide polymorphism (canSNP) analysis. The tree is based on multilocus variable number of tandem repeats (MLVA-16) data from MISTReSS, calculated in BioNumerics using categorical values with 1000 permutations to receive the highest resampling support. The root position was set as the maximum branch length. The three columns represent the year of isolation, the host and the governorate, correspondingly. The last column represents the colours from the core-genome single-nucleotide polymorphism (cgSNP) analysis according to the cgSNP clustering in Figure 2

HEA143 and 15,707 show completely different MLVAGTs compared to the remaining isolates. Four of them are identical to former entries from France and Sicily, Italy, whereas 60 of them were not yet found in the MLVA database. The SDI based on one marker distance for the isolates belonging to the WM clade is 0.9719. Two categories could be identified when comparing the cgSNP analysis (Figure 2) to MLVA (Figure 3). The first category represents an identical cgSNPGT but different MLVAGTs, and the second one represents the same MLVAGT but different cgSNPGTs. Because of that and the fact that two different calculation trees were used (MP tree versus NJ tree), there is a quite different clustering in Figures 2 and 3. To facilitate the comparison of the trees, the numbering of clusters from Figure 2 was taken. The examples in Tables 2 and 3 describe the same cgSNPGT but different MLVAGTs (Table 2) and vice versa the same MLVAGT but different cgSNPGTs (Table 3).

3.5 | Strain comparison with public database entries

To compare the 136 isolates presented in this study with available entries from GenBank and the NCBI SRA, a cgSNP analysis (Figure S1) and MLVA (Figure S2) were performed. For cgSNP analysis, 626 datasets from the SRA database were available, and for MLVA, 216

TABLE 2 Examples representing the same core-genome canonical single-nucleotide polymorphism-based genotype (cgSNPGT) but different MLVA-based genotypes (MLVAGTs)

Number	Subcluster	Identical cgSNPGTs	Differences in MLVA
1	1b	HEA62, 15690, 15693	HEA62: Br04 and Br16
2	2b	21908, 21910	Br04
3	3c	21893, 21897, 21889	21897: Br07
4	3e	SH14, SH19, SH2, SH23, SH24	SH14: Br07
5	3f	21879, 21880	Br16
6	3g	15678–16103	15678, 16088, 16102; Br16

Note: The last column represents the distinguishing Bruce markers.

TABLE 3 Examples representing the same MLVA-based genotype (MLVAGT) but different core-genome canonical single-nucleotide polymorphism-based genotypes (cgSNPGTs)

Number	Subcluster	Identical MLVAGT	Differences in cgSNPGTs
1	3g	5380, 15660, 15694, 15695, 15696, 15712, 15713, 16101, 16103	5380 differs in 43–46 cgSNPs from the other isolates
2	3e	SH5 and all isolates from SC3e except SH14	SH5 differs in at least 10 cgSNPs

datasets from the GenBank database (see Section 2.9 for explanation). The entire cgSNP analysis based on cgSNP distances is provided in Table S6. The cgSNP analysis (Figure S1) reveals that isolates from Italy, which were assigned to the WM clade, are the closest relatives to the Egyptian isolates, with a distance of at least 11 cgSNPs. The next relatives are from Austria (441 cgSNPs), Sweden (901 cgSNPS), Algeria (928 cgSNPs), Morocco (953 cgSNPs) and USA (1007 cgSNPs, unknown host), whereby isolates from Austria, Sweden, Algeria and Morocco were isolated from humans. One isolate each from Austria and Italy locate within the Egyptian cluster. Both isolates originate from humans who might have aquired the infection while travelling in Egypt. Isolate 15,707 shows just four cgSNPS to one entry from the United States, whereas isolate HEA143 has no appropriate relative. According to the MLVA (Figure S2), our Egyptian isolates show short distances of a maximum of two markers to entries, mainly from Italy and the United States, India, France, Austria and Morocco (Table S7).

4 | DISCUSSION

This study focuses on the analyses of the distribution and possible spread of outbreak strains of Egyptian *B. melitensis* using a panel of 136 isolates collected in 17 governorates between 2001 and 2020, including only 1 isolate each from 2001, 2002 and 2010. Unfortunately, none of the investigated isolates were taken from the years 2003–2005 and 2008–2009. Data from in silico genotyping methods (cgSNP analysis and in silico MLVA) based on WGS were used to analyse a possible phylogenetic relationship among our panel members isolates. Genotyping based on either WGS cgSNP analysis or in silico MLVA-16 resulted in a high-diversity resolution. The SDI, calculated just for the isolates belonging to the WM clade, determined for cgSNP analysis was 0.9921 and 0.9719 for MLVA. Both indices were calculated by accepting just a single-character difference (one cgSNP or one Bruce marker) to define a different genotype. In general, *B. melitensis* appears to be widespread, especially in the northern part of Egypt. All isolates are from the northern part of Egypt except isolates from Asyut, Aswan and Shalateen, which are located in the middle or the South of Egypt.

4.1 | Core-genome SNP analysis as a tool for outbreak investigation

Results of cgSNP analysis revealed many different outbreak strains, both older and widely distributed ones, and rather newly introduced ones, as shown in Figure 2. For example, an older outbreak strain that has spread continuously over a long period can be found in SC2d. This outbreak strain differed in a maximum of seven cgSNPs and was found in 2006, 2010, 2012 and 2014 in four different governorates. On the other hand, both isolates HEA59 (from 2006) and 5378 (from 2014) differentiated in just one single cgSNP within 7 years, given one expect that both isolates descended from each other. Because the isolates of this SC originate from four different governorates, it can be assumed this outbreak strain was spread over time. In SC2c, there are 5 years between the isolation of HEA44 (2007) and 5313 (2012) that differ in just three cgSNPs.

Furthermore, SC2a represents an at least 6-year-old outbreak strain differing in a maximum of seven cgSNPs (except HEA25) which might have originated in the Giza region in 2011 and then spread to Beni Suef and Dakahlia. Taking HEA25 into account as a possible ancestor of this group of isolates, just according to the year of isolation, the number of cgSNPs still defining the same outbreak strain would be ten, presupposing the isolates from the animals are considered only. The year of isolation is, however, not a strong argument to determine phylogenetic relations. Another example of a possible long-term present outbreak strain is represented by SC2b, where HEA14 and 21877 differ in 10 cgSNPs. If one accepts the possibility that HEA14 from Faiyum (2007) could be the ancestor of isolates 21908/21910 (2019) and even

the human isolate in this group, the 10 cgSNPs would still define the same outbreak strain as having been active during a period of 13 years. In addition, in Figure 2 are several diverse outbreak strains; some of them were newly introduced (2020). Such diverse outbreak strains are represented in SC s SC1a, SC2c and SC3a.

While comparing the number of cgSNPs with epidemiological metadata on the source, location and time of isolation, the question arises if and to what extent the number of cgSNPs can be correlated with the definition of an outbreak strain. For example, with regard to our panel of isolates, a number of seven cgSNPs would represent an outbreak strain where isolates are collected within the same year and from the same location, like in SC3f. The same applies to isolates in SC3b when accepting the timeline 2018/2019 as 1 year.

A similar situation is found in SC1b. Here, the isolates are related by six cgSNPs but differ in up to 4 years and four different isolation regions. The only exception is HEA6 isolated in Faiyum in 2013 with 10 cgSNPs different from another isolate from the same year and region. Assuming HEA6 with its up to 10 cgSNPs to its relatives has been the ancestor of this outbreak strain, later on, it might have spread to Kafr El Sheikh, Gharbia and Ismailia. The last example worth discussing is SC3e which contains 17 isolates from Monufia and one each from Sharqia and Ismailia, all isolated in 2018 and 2019 from cattle, buffaloes and from one sheep. These isolates represent two groups with cgSNPs of only three within and nine between them. Because of the same region and the narrow time of isolation, one would expect them to belong to the same outbreak strain. The only exception here is SH5 from Monufia in 2018, which differs in at least 10 and up to 16 cgSNPs from the other isolates. Hence, this single isolate might represent a different outbreak strain isolated by chance in 2018 in Monufia.

In conclusion, according to the correlation of cgSNPs with epidemiological metadata in our panel of *B. melitensis* isolates, a number of around 10 cgSNPs looks like a reasonable approximation of a bioinformatics cut-off for the discrimination of outbreak strains. However, one would need to accept that such an approximation applies only to the panel of isolates used here and, therefore, is not a general rule.

4.2 | Human isolates can be assigned to possible animal infectious sources

Of the 136 isolates included in our analysis, 13 are of human origin. In two cases, a significant correlation between a human case of disease and the possible source of infection was found (SC2c). Human isolates from Faiyum in 2018 and 2019 (21876, 21883) were found to be close relatives to an animal isolate from cattle in Faiyum (2019), differing in four cgSNPs only. One can speculate that the outbreak strain isolated from cattle has already been active in the same region for some years, leading to human infections in 2018 and 2019. The same might apply to the human isolate 21884, Faiyum 2019, and an animal isolate from goat in Faiyum 2020, with four cgSNPs only. Again, the outbreak strain isolated from the goat might have already been active in the same region some years before, leading to the human infection in 2019.

Less convincing but still relevant are the following correlations: Isolates HEA142, HEA146 and 21,875 from 2016, 2017 and 2018 (SC2a), respectively, are very homogeneous; the closest animal relatives are HEA42 isolated in 2013 from a cow of the same region, Beni Suef, and HEA7, isolated in 2014 from Dakahlia. Although the time and region of the human and animal isolates are not identical, the human isolate might originate from a hospital not related to the point of infection; the disease causing outbreak strain might have been around between 2013 and 2016 in Beni Suef. This assumption is supported by the fact that fever hospitals across the country pooling hospitals and are not restricted to patients from specific governorates (Abd El-Wahab et al., 2019). The same might apply for the human isolate 21,877 (SC2b) from a hospital in Giza (2019), showing a distance of 10 cgSNPs to an isolate (HEA14) of a cow in Faiyum (2007) and, again, to the human isolate 21,886 from Faiyum (2019) which is related to HEA49 from Faiyum (2014) by 10 cgSNPs (SC2c). According to the cgSNP analysis, isolates 21,882 and HEA180 (SC1a) from 2020 and 2019 are identical (Figure 2) but cannot be assigned to any source of infection. Both human samples were isolated at the central hospital in Giza, and the source of infection is unknown. No sources of infection were found for isolates 21,885, 21,879 and 21,880.

4.3 | MLVA-16 genotyping has great value for differentiation but not for epidemiological investigation

The comparison of the cgSNP analysis and MLVA revealed some fundamental differences. Outbreak strains with identical cgSNPGTs show different MLVAGTs and vice versa, as indicated in Tables 2 and 3. The cgSNP analysis provides a much more precise clustering of the outbreak strains when including epidemiological metadata. One of the striking examples is SC3g, where isolates indicate a distance of at least 43 cgSNPs to isolate 5380 (SC3e), which was shown to have the same MLVAGT as isolates in SC3g. In the aforementioned examples, the differences are based on the MLVA markers Bruce04, Bruce07, Bruce09, Bruce16, Bruce18 and/or Bruce19, representing microsatellite markers belonging to panel 2 of VNTR markers. Their discriminative value is of great importance for a pure differentiation of isolates but not informative for epidemiological analyses of outbreaks. The reason is that microsatellites are susceptible to homoplasia (Le Flèche et al., 2006), that is the possibility of multiple mutations and the associated convergence of the markers. This means that a given genotype based on copy numbers of MLVA markers can occur more than once at different locations or even arise during subculture in the laboratory. After only 10 passages, it has already been observed that the addition or deletion of a repeat unit occurred in Bruce07 of *B. abortus* S19 and RB51 vaccine strains (Dorneles et al., 2013; Wareth et al., 2020). This makes outbreak analyses and the tracking of the spread of outbreak strains unreliable. However, SNPs are very stable and are thought to have a phylogenetic character as long as highly repetitive SNPs like single-nucleotide repeats or rRNA sequences are omitted.

4.4 | Outbreak strains of *B. melitensis* might have been imported to Egypt from other countries

Because the isolates are obviously from at least three different governorates of origin according to the canSNP assay, it is more likely that the *Brucella* infections were originally introduced to Egypt than vice versa, at least as far as the East Mediterranean and American origins are concerned. To search for possible correlations between our panel of 136 *B. melitensis* isolates with isolates from other countries, a new cgSNP analysis (Figure S1, Table S6) and MLVA (Figure S2, Table S7) with entries from public databases were performed. Performing a new cgSNP analysis with a large set of sequences from international strains resulted in a slightly smaller core genome as some SNP positions were not available in all analysed genomes. For such large sequencing data sets defining core positions in less than 100% of the samples prevents strong shrinking of the core-genome due to missing bases in the samples (Perrin et al., 2017). The cgSNP analysis with entries from public databases confirmed that the closest relatives of the WM group (2001–2020) are those from Italy (2006–2018). This is in concordance with recent previous publications (Abdel-Hamid et al., 2020; Hegazy et al., 2022; Sayour et al., 2020; Wareth et al., 2020) that showed by MLVA-16 that the *B. melitensis* strains circulating in Egypt are close to the Italian ones. Our single isolate 15,707, belonging to canSNP clade America, is very close to one entry from the USA (2005). No relatives were found for the single isolate HEA143 belonging to the canSNP clade East Mediterranean. Concerning the data of MLVA (Figure S2), a very similar result as in the cgSNP analysis (Figure S1) was found except for additional entries from France showing a maximum distance of two markers to the isolates from Egypt. However, the both analyses cannot be directly compared with one another here, because the numbers of entries from the databases differ. To substantiate the possible interpretation of the data as indicating a long-lasting and continuing import of *B. melitensis* to Egypt from other countries, the report of the Food and Agriculture Organization of the United Nations (FAOSTAT) on the import of animals for breeding or slaughter purposes from several countries to Egypt was used as the context (http://www.fao.org/faostat/en/#data/TM, accessed on 24/04/2021), provided in Table S9. Among others, this table shows the import of animals which took place in 1986–2018 from Italy (cattle and buffaloes). Although animals from Italy have been imported for breeding, animals from France were imported for breeding and slaughter purposes. Although some animals were imported for slaughtering, a possible source of infection from these countries should not be ruled out.

4.5 | Limitations of the study

For the interpretation of our bioinformatics data, some obvious limitations of the study should be taken into consideration. First, there is a sampling bias in data collection, both temporally and geographically. The majority of samples are from the north of Egypt, whereas other country regions are not represented at all. One should also keep in mind that the available metadata on the region of isolation does not represent the exact location of sample collection. The same issue applies to human isolates as it remains unknown where the patients eventually came from.

Furthermore, there were no isolates available from 2003–2005, 2008 and 2009, and only one isolate originates from 2001 and 2002, respectively. Moreover, the exact isolation dates are unknown; hence, only full years had to be used as sampling dates. As our analysis starts in 2001, only and no isolates from earlier years are available, evolutionary analyses of outbreak strains were limited and did not reveal meaningful results.

5 | CONCLUSIONS

B. melitensis causing brucellosis outbreaks in both animals and humans is widespread in at least the northern part of Egypt. Several quite different outbreak strains indicated a long-lasting and probably continuing import of the infectious agent to Egypt from abroad. By comparing our panel of isolates with entries from public databases, the results show possible relationships with entries from Italy. To track the history of distribution more in detail, many more isolates are needed, preferably from worldwide. At the same time, outbreak strains existing for already many years have been spread over long distances and between many governorates. All but two of the isolates analysed belong to the WM clade. One isolate each was assigned to the American and the East Mediterranean clades. This study shows that cgSNP analysis can reliably be used to define and track outbreak strains of *B. melitensis* when data are correlated to epidemiological metadata on the source, time and site of isolation. MLVA-16 showed a high discriminatory power in the resolution of different isolates but was misleading for the definition and tracking of an outbreak strain.

ACKNOWLEDGEMENTS

The authors acknowledge the support of the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tuebingen, the state of Baden-Wuerttemberg through bwHPC and the German Research Foundation (DFG). Furthermore, the authors thank Prof. Dr. Hoelzle for providing the premises and for the support. Further, the authors thank Prof. Dr. Heinrich Neubauer (FLI, Jena) for making possible the collaboration with the Egyptian partners, for reading the manuscript, and for his valuable discussion and for his enormous support. Furthermore, we thank Dr. Hosny El Adawy (FLI, Jena) for organizing 42 isolates from Egypt, Prof. Dr. Sascha Al-Dahouk from the Bundesinstitut für Risikobewertung (BfR) Berlin for training in essential bacteriological methods, Dr. Anne Mayer-Scholl (BfR) for kindly providing the Bruce-ladder PCR protocol, Silvia Speck (University of Hohenheim) for the excellent technical support in the laboratory and Dr. Katleen Vranckx for the support in BioNumerics. Finally, special thanks go to Prof. Dr. Jeff Foster (University of Arizona) for giving a deeper understanding of the canSNP assay developed by himself. Funding by LEAP-AGRI project MUSCBCEA number 2817LEAP01 supported this work. High Computing and Cloud Computing was funded by the German Research Foundation (DFG) through

grant no INST 37/935-1 FUGG. The resources for isolates were funded by CRDF Global's 2016 Emerging Infectious Disease Research Grant (agreement number GTR4-17-63138-0).

Open access funding enabled and organized by Projekt DEAL.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Conceptualization: Wolfgang Beyer and Katharina Holzer; resources: Mohamed El-Diasty, Gamal Wareth, Shawky A. Moustafa, Nour H. Abdel-Hamid, Mahmoud E. R. Hamdy, Ashraf E. Sayour, Mohamed Elhadidy and Essam M. Elbauomy; methodology: Wolfgang Beyer, Katharina Holzer and Jörg Linde; software: Katharina Holzer, Jörg Linde, Felix Bartusch and Mostafa Y. Abdel-Glil; validation: Wolfgang Beyer, Katharina Holzer and Jörg Linde; investigation: Katharina Holzer; data curation: Wolfgang Beyer, Katharina Holzer, Jörg Linde, Felix Bartusch, Mostafa Y. Abdel-Glil, Gamal Wareth, Mohamed El-Diasty, Nour H. Abdel-Hamid, Mahmoud E. R. Hamdy, Shawky A. Moustafa, Ashraf E. Sayour and Essam M. Elbauomy; writing-original draft preparation: Katharina Holzer and Wolfgang Beyer; writing-review and editing: Gamal Wareth, Jörg Linde, Falk Melzer, Felix Bartusch, Mostafa Y. Abdel-Glil, Ashraf E. Sayour, Mohamed Elhadidy and Nour H. Abdel-Hamid; project administration: Wolfgang Beyer; funding acquisition: Wolfgang Beyer and Mohamed Elhadidy. All authors have read and agreed to the published version of the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Katharina Holzer https://orcid.org/0000-0002-8258-3910
Nour H. Abdel-Hamid https://orcid.org/0000-0001-5554-7984
Mostafa Y. Abdel-Glil https://orcid.org/0000-0003-2068-6081
Mohamed Elhadidy https://orcid.org/0000-0002-6153-1302

REFERENCES

Abdel-Hamid, N. H., El-Bauomy, E. M., Ghobashy, H. M., & Shehata, A. A. (2020). Genetic variation of *Brucella* isolates at strain level in Egypt. *Veterinary Medicine and Science*, *6*(3), 421–432. https://doi.org/10.1002/vms3.260

Abd El-Wahab, E. W., Hegazy, Y. M., El-Tras, W. F., Mikheal, A., Kabapy, A. F., Abdelfatah, M., Bruce, M., & Eltholth, M. M. (2019). A multifaceted risk model of brucellosis at the human-animal interface in Egypt. *Transboundary and Emerging Diseases*, *66*(6), 2383–2401. https://doi.org/10.1111/tbed.13295

Al Dahouk, S., Le Flèche, P., Nöckler, K., Jacques, I., Grayon, M., Scholz, H. C., Tomaso, H., Vergnaud, G., & Neubauer, H. (2007). Evaluation of *Brucella* MLVA typing for human brucellosis. *Journal of Microbiological Methods*, *69*(1), 137–145. https://doi.org/10.1016/j.mimet.2006.12.015

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, S. N., Phan, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyes, M., & Pevzner, P. A. (2012). Spades: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Chiliveru, S., Appari, M., & Suravajhala, P. (2015). On *Brucella* pathogenesis: Looking for the unified challenge in systems and synthetic biology. *Systems and Synthetic Biology*, *9*(1–2), 73–75. https://doi.org/10.1007/s11693-014-9158-2

Cutler, S. J., Whatmore, A. M., & Commander, N. J. (2005). Brucellosis—New aspects of an old disease. *Journal of Applied Microbiology*, *98*(6), 1270–1281. https://doi.org/10.1111/j.1365-2672.2005.02622.x

Dentinger, C. M., Jacob, K., Lee, L. V., Mendez, H. A., Chotikanatis, K., McDonough, P. L., Chico, D. M., De, B. K., Tiller, R. V., Traxler, R. M., Campagnolo, E. R., Schmitt, D., Guerra, M. A., & Slavinski, S. A. (2015). Human *Brucella canis* infection and subsequent laboratory exposures associated with a puppy, New York City, 2012. *Zoonoses and Public Health*, *62*(5), 407–414. https://doi.org/10.1111/zph.12163

Dorneles, E. M. S., de Faria, A. P. P., Pauletti, R. B., Santana, J. A., Caldeira, G. A. V., Heinemann, M. B., Titze-de-Almeida, R., & Lage, A. P. (2013). Genetic stability of *Brucella abortus* S19 and RB51 vaccine strains by multiple locus variable number tandem repeat analysis (MLVA16). *Vaccine*, *31*(42), 4856–4859. https://doi.org/10.1016/j.vaccine.2013.07.063

Ficht, T. (2010). *Brucella* taxonomy and evolution. *Future Microbiology*, *5*(6), 859–866. https://doi.org/10.2217/fmb.10.52

Foster, J. T., Walker, F. M., Rannals, B. D., Hussain, M. H., Drees, K. P., Tiller, R. V., Hoffmaster, A. R., Al-Rawahi, A., Keim, P., & Saqib, M. (2017). African lineage *Brucella melitensis* isolates from Omani livestock. *Frontiers in Microbiology*, *8*, 2702. https://doi.org/10.3389/fmicb.2017.02702

Galińska, E. M., & Zagórski, J. (2013). Brucellosis in humans—Etiology, diagnostics, clinical forms. *Annals of Agricultural and Environmental Medicine: AAEM*, *20*(2), 233–238.

García-Yoldi, D., Marín, C. M., de Miguel, M. J., Muñoz, P. M., Vizmanos, J. L., & López-Goñi, I. (2006). Multiplex PCR assay for the identification and differentiation of all *Brucella* species and the vaccine strains *Brucella abortus* S19 and RB51 and *Brucella melitensis* rev1. *Clinical Chemistry*, *52*(4), 779–781. https://doi.org/10.1373/clinchem.2005.062596

Godfroid, J., Scholz, H. C., Barbier, T., Nicolas, C., Wattiau, P., Fretin, D., Whatmore, A. M., Cloeckaert, A., Blasco, J. M., Moriyon, I., Saegerman, C., Muma, J. B., Al Dahouk, S., Neubauer, H., & Letesson, J. J. (2011). Brucellosis at the animal/ecosystem/human interface at the beginning of the 21st century. *Preventive Veterinary Medicine*, *102*(2), 118–131. https://doi.org/10.1016/j.prevetmed.2011.04.007

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). Quast: Quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hashemifar, I., Yadegar, A., Jazi, F. M., & Amirmozafari, N. (2017). Molecular prevalence of putative virulence-associated genes in *Brucella melitensis* and *Brucella abortus* isolates from human and livestock specimens in Iran. *Microbial Pathogenesis*, *105*, 334–339. https://doi.org/10.1016/j.micpath.2017.03.007

Hegazy, Y. M., Abdel-Hamid, N. H., Eldehiey, M., Oreiby, A. F., Algabbary, M. H., Hamdy, M. E. R., Beleta, E. I., Martínez, I., Shahein, M. A., García, N., & Eltholth, M. (2022). Trans-species transmission of Brucellae among ruminants hampering brucellosis control efforts in Egypt. *Journal of Applied Microbiology*, *132*(1), 90–100. https://doi.org/10.1111/jam.15173

Hegazy, Y. M., Molina-Flores, B., Shafik, H., Ridler, A. L., & Guitian, F. J. (2011). Ruminant brucellosis in Upper Egypt (2005-2008). *Preventive Veterinary Medicine*, *101*(3–4), 173–181. https://doi.org/10.1016/j.prevetmed.2011.05.007

Holzer, K., El-Diasty, M., Wareth, G., Abdel-Hamid, N. H., Hamdy, M. E. R., Moustafa, S. A., Linde, J., Bartusch, F., Sayour, A. E., Elbauomy, E. M., Elhadidy, M., Melzer, F., & Beyer, W. (2021). Tracking the distribution of *Brucella abortus* in Egypt based on core genome SNP analysis and in silico MLVA-16. *Microorganisms*, *9*(9), 1942. https://doi.org/10.3390/microorganisms9091942

Kang, S. I., Her, M., Kim, J. W., Kim, J. Y., Ko, K. Y., Ha, Y. M., & Jung, S. C. (2011). Advanced multiplex PCR assay for differentiation of *Brucella* species.

Applied and Environmental Microbiology, *77*(18), 6726–6728. https://doi.org/10.1128/AEM.00581-11

Khan, A. U., Melzer, F., Sayour, A. E., Shell, W. S., Linde, J., Abdel-Glil, M., El-Soally, S. A. G. E., Elschner, M. C., Sayour, H. E. M., Ramadan, E. S., Mohamed, S. A., Hendam, A., Ismail, R. I., Farahat, L. F., Roesler, U., Neubauer, H., & El-Adawy, H. (2021). Whole-genome sequencing for tracing the genetic diversity of *Brucella abortus* and *Brucella melitensis* isolated from livestock in Egypt. *Pathogens (Basel, Switzerland)*, *10*(6), 759. https://doi.org/10.3390/pathogens10060759

Le Flèche, P., Jacques, I., Grayon, M., Al Dahouk, S., Bouchon, P., Denoeud, F., Nöckler, K., Neubauer, H., Guilloteau, L. A., & Vergnaud, G. (2006). Evaluation and selection of tandem repeat loci for a *Brucella* MLVA typing assay. *BMC Microbiology*, *6*, 9. https://doi.org/10.1186/1471-2180-6-9

López-Goñi, I., García-Yoldi, D., Marín, C. M., de Miguel, M. J., Muñoz, P. M., Blasco, J. M., Jacques, I., Grayon, M., Cloeckaert, A., Ferreira, A. C., Cardoso, R., Correa de Sa, M. I., Walravens, K., Albert, D., & Garin-Bastuji, B. (2008). Evaluation of a multiplex PCR assay (Bruce-ladder) for molecular typing of all *Brucella* species, including the vaccine strains. *Journal of Clinical Microbiology*, *46*(10), 3484–3487. https://doi.org/10.1128/JCM.00837-08

Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). Mummer4: A fast and versatile genome alignment system. *PLoS Computational Biology*, *14*(1), e1005944. https://doi.org/10.1371/journal.pcbi.1005944

De Massis, F., Zilli, K., Di Donato, G., Nuvoloni, R., Pelini, S., Sacchini, L., D'Alterio, N., & Di Giannatale, E. (2019). Distribution of *Brucella* field strains isolated from livestock, wildlife populations, and humans in Italy from 2007 to 2015. *PLoS One*, *14*(3), e0213689. https://doi.org/10.1371/journal.pone.0213689

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., & Phillippy, A. M. (2016). Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biology*, *17*(1), 132. https://doi.org/10.1186/s13059-016-0997-x

Pappas, G., Akritidis, N., Bosilkovski, M., & Tsianos, E. (2005). Brucellosis. *The New England Journal of Medicine*, *352*(22), 2325–2336. https://doi.org/10.1056/NEJMra050570

Perrin, A., Larsonneur, E., Nicholson, A. C., Edwards, D. J., Gundlach, K. M., Whitney, A. M., Gulvik, C. A., Bell, M. E., Rendueles, O., Cury, J., Hugon, P., Clermont, D., Enouf, V., Loparev, V., Juieng, P., Monson, T., Warshauer, D., Elbadawi, L., Walters, M. S., … Brisse, S. (2017). Evolutionary dynamics and genomic features of the *Elizabethkingia anophelis* 2015 to 2016 Wisconsin outbreak strain. *Nature Communications*, *8*, 15483. https://doi.org/10.1038/ncomms15483

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., & Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Analytical Methods*, *8*(1), 12–24. https://doi.org/10.1039/C5AY02550H

Refai, M. (2002). Incidence and control of brucellosis in the near east region. *Veterinary Microbiology*, *90*(1–4), 81–110. https://doi.org/10.1016/s0378-1135(02)00248-1

Sacchini, L., Wahab, T., Di Giannatale, E., Zilli, K., Abass, A., Garofolo, G., & Janowicz, A. (2019). Whole genome sequencing for tracing geographical origin of imported cases of human brucellosis in Sweden. *Microorganisms*, *7*(10), 398. https://doi.org/10.3390/microorganisms7100398

Sayour, A. E., Elbauomy, E., Abdel-Hamid, N. H., Mahrous, A., Carychao, D., Cooley, M. B., & Elhadidy, M. (2020). MLVA fingerprinting of *Brucella melitensis* circulating among livestock and cases of sporadic human illness in Egypt. *Transboundary and Emerging Diseases*, *67*(6), 2435–2445. https://doi.org/10.1111/tbed.13581

Wareth, G., El-Diasty, M., Melzer, F., Schmoock, G., Moustafa, S. A., El-Beskawy, M., Khater, D. F., Hamdy, M. E. R., Zaki, H. M., Ferreira, A. C., Ekateriniadou, L. V., Boukouvala, E., Abdel-Glil, M. Y., Menshawy, A. M. S., Sancho, M. P., Sakhria, S., Pletz, M. W., & Neubauer, H. (2020). MLVA-16 genotyping of *Brucella abortus* and *Brucella melitensis* isolates from different animal species in Egypt: Geographical relatedness and the Mediterranean lineage. *Pathogens (Basel, Switzerland)*, *9*(6), 498. https://doi.org/10.3390/pathogens9060498

Wareth, G., Hikal, A., Refai, M., Melzer, F., Roesler, U., & Neubauer, H. (2014). Animal brucellosis in Egypt. *Journal of Infection in Developing Countries*, *8*(11), 1365–1373. https://doi.org/10.3855/jidc.4872

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Holzer, K., Wareth, G., El-Diasty, M., Abdel-Hamid, N. H., Hamdy, M. E. R., Moustafa, S. A., Linde, J., Bartusch, F., Abdel-Glil, M. Y., Sayour, A. E., Elbauomy, E. M., Elhadidy, M., Melzer, F., & Beyer, W. (2022). Tracking the distribution, genetic diversity and lineage of *Brucella melitensis* recovered from humans and animals in Egypt based on core-genome SNP analysis and in silico MLVA-16. *Transboundary and Emerging Diseases*, *69*, 3952–3963. https://doi.org/10.1111/tbed.14768