



Genetic Characterization of *Brucella* spp.: Whole Genome Sequencing-Based Approach for the Determination of Multiple Locus Variable Number Tandem Repeat Profiles

OPEN ACCESS

Edited by:

Martin Maiers,
National Marrow Donor Program,
United States

Reviewed by:

Maryam Dadar,
Razi Vaccine and Serum Research
Institute, Iran
Marcela Suárez-Esquivel
Universidad Nacional de Costa Rica,
Costa Rica

*Correspondence:

João Paulo Gomes
j.paulo.gomes@insa.min-saude.pt

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 12 July 2021

Accepted: 18 October 2021

Published: 12 November 2021

Citation:

Pelerito A, Nunes A, Grilo T, Isidro J,
Silva C, Ferreira AC, Valdezate S,
Núncio MS, Georgi E and
Gomes JP (2021) Genetic
Characterization of *Brucella* spp.:
Whole Genome Sequencing-Based
Approach for the Determination of
Multiple Locus Variable Number
Tandem Repeat Profiles.
Front. Microbiol. 12:740068.
doi: 10.3389/fmicb.2021.740068

Ana Pelerito^{1†}, Alexandra Nunes^{2,3,4†}, Teresa Grilo¹, Joana Isidro², Catarina Silva^{5,6},
Ana Cristina Ferreira^{4,7}, Sylvia Valdezate⁸, Maria Sofia Núncio¹, Enrico Georgi⁹ and
João Paulo Gomes^{2,4*}

¹Emergency Response and Biopreparedness Unit, Department of Infectious Diseases, National Institute of Health, Lisbon, Portugal, ²Bioinformatics Unit, Department of Infectious Diseases, National Institute of Health, Lisbon, Portugal, ³CBIOS – Universidade Lusófona's Research Center for Biosciences & Health Technologies, Lisbon, Portugal, ⁴Faculty of Veterinary Medicine, Lusófona University, Lisbon, Portugal, ⁵Technology and Innovation Unit, Department of Human Genetics, National Institute of Health, Lisbon, Portugal, ⁶Centre for Toxicogenomics and Human Health (ToxOmics), Faculdade de Ciências Médicas, Nova Medical School, Universidade Nova de Lisboa, Lisbon, Portugal, ⁷National Institute for Agrarian and Veterinary Research, I.P. (INIAV, IP), Oeiras, Portugal, ⁸ISCIII Reference and Research Laboratory for Taxonomy, National Centre of Microbiology, Instituto de Salud Carlos III, Madrid, Spain, ⁹Bundeswehr Institute of Microbiology, Munich, Germany

Brucellosis is an important zoonosis that is emerging in some regions of the world, gaining increased relevance with the inclusion of the causing agent *Brucella* spp. in the class B bioterrorism group. Until now, multi-locus VNTR Analysis (MLVA) based on 16 loci has been considered as the gold standard for *Brucella* typing. However, this methodology is laborious, and, with the rampant release of *Brucella* genomes, the transition from the traditional MLVA to whole genome sequencing (WGS)-based typing is on course. Nevertheless, in order to avoid a disruptive transition with the loss of massive genetic data obtained throughout the last decade and considering that the transition timings will vary considerably among different countries, it is important to determine WGS-based MLVA alleles of the nowadays sequenced genomes. On this regard, we aimed to evaluate the performance of a Python script that had been previously developed for the rapid *in silico* extraction of the MLVA alleles, by comparing it to the PCR-based MLVA procedure over 83 strains from different *Brucella* species. The WGS-based MLVA approach detected 95.3% of all possible 1,328 hits (83 strains × 16 loci) and showed an agreement rate with the PCR-based MLVA procedure of 96.4% for MLVA-16. According to our dataset, we suggest the use of a minimal depth of coverage of ~50x and a maximum number of ~200 contigs as guiding “boundaries” for the future application of the script. In conclusion, the evaluated script seems to be a very useful and robust tool for the *in silico* determination of MLVA profiles of *Brucella* strains, allowing retrospective and prospective molecular

epidemiological studies, which are important for maintaining an active epidemiological surveillance of brucellosis.

Keywords: *Brucella* spp., MLVA, whole-genome sequencing, zoonosis, genotyping, Python script

INTRODUCTION

Brucellosis is one of the most common bacterial zoonosis causing great damage to the farming industry and public health (Franc et al., 2018). The brucellosis burden specifically on low-income countries has led the World Health Organization (WHO) to classify it as one of the world's leading neglected zoonotic diseases (WHO, 2020). However, given the absence of specific signs and symptoms, the disease is commonly underdiagnosed (Valdezate et al., 2010).

Brucellosis is transmitted to humans by ingestion of unpasteurized dairy products or by direct contact with infected animals, placentas, or aborted fetuses (Young, 2005). It can constitute a severely debilitating illness, with diverse symptoms ranging from fever, sweating, fatigue, weight loss, headache, and joint pain. Neurological complications, such as personality changes, meningitides, encephalitis, and peripheral neuropathy, can also occur (Dean et al., 2012).

The interest in human brucellosis has been boosted due to its recent re-emergence and enhanced surveillance worldwide and from the inclusion of the causing agent *Brucella* spp. in the group of class B bioterrorism agent (Franco et al., 2007). A low infectious dose of 100–1,000 organisms is sufficient to cause an infection. The mechanisms of transmission, through aerosols or food chains, make them easily transmissible to both humans and animals (Tan et al., 2015). Thus, the distinction between natural outbreaks and/or intentional release of microorganisms may be of fundamental importance in the context of the bioterrorism.

Studies enrolling DNA–DNA hybridization procedures and comparative genomics revealed that *Brucella* species are characterized by >80% interspecies homology and >98% sequence similarity (Whatmore et al., 2006; Kattar et al., 2008). Indeed, the sequencing of 16S rRNA gene showed 100% identity between all *Brucella* spp. (Gee et al., 2004). The genus presently encloses 12 genetically highly related species. Human brucellosis can be caused by various *Brucella* species; however, *Brucella melitensis* is the most virulent and by far the most frequently observed causative agent of human infection (Young, 2005; Georgi et al., 2017). On this regard, the identification of the circulating *Brucella* species, biovar, and genotype is very important, mainly for tracking back infectious sources and monitoring transmission routes (Pisarenko et al., 2018). The species identification by PCR assays is sufficient for the purposes of diagnosis of human/animal disease or the detection of food contamination but not for the tracing of outbreaks or bioterrorism events (De Santis et al., 2011).

To achieve the goal of sub-species discrimination, Variable Number Tandem Repeats (VNTR) have been investigated in multi-locus VNTR analysis (MLVA) by various research groups since 2003 (Bricker et al., 2003; Le Fleche et al., 2006; Whatmore et al., 2006). This MLVA *Brucella* typing scheme has proved to have the ability to differentiate *Brucella* species, biovar, and

even some isolates. This is facilitated do to the creation of an online database of MLVA-16 profiles (MLVA Bank, n.d.)¹ that is available to all laboratories, allowing the comparison of *Brucella* strains worldwide (Le Fleche et al., 2006; Mambres et al., 2017; Sun et al., 2017). The recent implementation of whole genome single nucleotide polymorphism (SNP)-based typing, associated with its decreasing costs, has led to substantial improvements of both molecular subtyping and phylogenetic analyses in microbiology. The development of core- and whole-genome multilocus sequence typing (MLST) schemes has been focused on the restrict number of bacterial pathogens, including *Brucella* spp. but their application may be tricky (Tan et al., 2015; Janowicz et al., 2018; Sankarasubramanian et al., 2019). In fact, the creation of universal intra- or inter-species schemes needs to overcome some genetic hurdles such as the existence of paralogous genes, annotation issues, the accessory genome, and nomenclature-associated difficulties. Nevertheless, public databases for molecular typing and microbial genome diversity (PubMLST) are already available,² allowing the use of whole genome sequences for typing purposes of multiple bacterial species. The same scenario is seen in the viral field as bioinformatics platforms were already developed, allowing the genotype determination from viral complete genomes, as for influenza virus and SARS-CoV-2 (Borges et al., 2018).³ Meanwhile, until whole genome data is fully established and accepted by the scientific community for classification/typing purposes in *Brucella*, the *in silico* determination of MLVA schemes can be of extreme utility. In fact, not only it overcomes the laborious PCR-based MLVA assessment but it also allows the dynamic cross-comparison with the typing-associated genetic data determined during the last decade. On this regard, a Python script has been developed focusing on the *in silico* determination of *Brucella* MLVA schemes taking advantage of the increasing number of sequenced genomes (Georgi et al., 2017). As no experimental validation of such script was performed, we now aimed to evaluate the agreement among the MLVA profiles determined through PCR- and WGS-based approaches for strains from several *Brucella* species, in order to check the validity of such technological transition underlying the genetic characterization of *Brucella*.

MATERIALS AND METHODS

Samples

Eighty-three *Brucella* strains isolated in Portugal, Spain, Germany, Hungary, and Belgium were used in this study. This set comprises

¹<https://microbesgenotyping.i2bc.paris-saclay.fr>

²<https://pubmlst.org/>

³<https://insaflu.insa.pt/>

essentially not only *B. melitensis* but also some representatives of *B. suis*, *B. abortus*, and *B. ovis* (**Supplementary Table S1**). Unfortunately, we had no access to the isolates of low prevalent species such as *B. ceti*, *B. canis*, and *B. pinnipedialis*. *Brucella melitensis* 16M strain (NC_003317 and NC_003318) was used as a reference strain.

All samples were handled in a BSL-3 biocontainment laboratory at the Portuguese National Institute of Health. *Brucella* isolates were cultured on blood agar for 3–5 days at 37°C under 5% CO₂, and total DNA was extracted from fresh cultures on the NucliSens easyMAG platform (Biomérieux), according to the manufacturer's instructions.

All strains were identified as *Brucella* species by real-time PCR, using a previously published assay (Pelerito et al., 2017). The molecular methods used up to species differentiation were performed in a tandem fashion. Firstly, an “in house” real-time PCR using hydrolysis probes was used to detect and identify *Brucella* genus. Secondly, for species differentiation, primers and Taqman probes were designed within the BMEI10466 gene for *B. melitensis* and BruAb2_0168 gene for *B. abortus* (Gopaul et al., 2008; Pelerito et al., 2017).

MLVA Assays

Single locus amplification of the eight minisatellite loci (panel 1) and eight microsatellite loci (panels 2A and 2B), that constitute the MLVA-16 assay, was performed as describe by Le Fleche (Le Fleche et al., 2006), with modifications by Garofolo (2015). MLVA PCRs were performed in four multiplex reactions in a final volume of 10 µl. The reactions contained: 1× Type-it Multiplex PCR Master Mix (Qiagen), 0.5× solution buffer, primers at appropriate concentrations, and 5 µl of DNA. The thermocycling conditions were as follows: 96°C for 5 min followed by either 30 (for multiplex 1, 3, and 4) or 24 cycles (for multiplex 2) of: 95°C for 30s, 60°C for 90s, and 72°C for 30s; followed by 60°C for 30 min. Multiplex 2 was run for 24 cycles in order to contain VNTR amplification artifacts (Garofolo et al., 2013). A 1.1 µl of each MLVA PCR product was mixed with 15 µl of formamide-diluted GeneScan 500 LIZ dye or GeneScan 1,200 LIZ dye size standards (Applied Biosystems), depending on the expected size of the fragments, and denatured at 96°C for 3 or 5 min, respectively. The mixtures were electrophoresed on an 8-capillary 3500 Genetic Analyzer equipped with 50 cm-long capillaries and POP7 polymer (Applied Biosystems). Estimation of molecular sizes of PCR products was obtained using GeneMapper software 6 (Applied Biosystems) with default analysis parameters. The reference *B. melitensis* 16M strain, for which the expected size is known for each VNTR locus, was used as control for allele's assignment.

Whole Genome Sequencing

For each strain, WGS was performed as previously described (Pinto et al., 2018). Briefly, quantification and quality assessment of the purified DNA was performed using the DNA HS Assay Kit (Thermo Fisher Scientific) in the Qubit Fluorometer and agarose gel electrophoresis (0.8%), respectively. High-quality DNA samples were then used to prepare dual-indexed Nextera

XT Illumina libraries that were subsequently subjected to cluster generation and paired-end sequencing (2 × 250 bp and 2 × 300 bp) on a MiSeq Illumina platform (Illumina Inc.), according to the manufacturer's instructions.

Reads quality control and bacterial *de novo* assembly were performed using the INNUca v4.0.1 pipeline,⁴ which consists of several integrated modules for reads QA/QC, *de novo* assembly, and post-assembly optimization steps. Briefly, after reads' quality analysis (FastQC v0.11.5)⁵ and cleaning (Trimmomatic v0.36; Bolger et al., 2014), genomes were assembled with SPAdes 3.11 (Bankevich et al., 2012) and subsequently improved using Pilon v1.18 (Walker et al., 2014), with genome coverage being monitored and reported after each processes. In order to evaluate the impact of the “post-assembly polishing” on the assembled genomes and subsequently on the *in silico* MLVA analyses, the SPAdes assemblies were also performed skipping the Pilon step. A final check was also performed. Considering that the *in silico* extraction of loci may be influenced by the quality of the assembled genomes, another largely used *de novo* assembler – Velvet (Zerbino and Birney, 2008) was applied through VelvetOptimiser v2.2.5,⁶ for comparative purposes, with and without Pilon. The VelvetOptimiser script was run using trimmed reads for odd k-mer values ranging from 31 to 127 (highest k-mer used in SPAdes), with all program default settings unchanged apart from the minimum output contig size, which was the same as used by SPAdes.

WGS-Based MLVA

Bacterial draft genomes were subjected to a Python script for *in silico* extraction of *Brucella* MLVA scheme (with 16 loci) as previously described (Georgi et al., 2017).⁷ This script is based on the count of repetitive DNA stretches contained within conserved DNA boxes that are upstream and downstream to the repetitions. As determining numbers of repeated stretches from WGS data may be error-prone, we carefully checked each locus in respect to the expected total length, internal repeat homogeneity or probability to get collapsed VNTRs during the assembly. All resulting MLVA 16 genotypes were compared to a public database with 2,215 entries of *B. melitensis* strains that can be assessed online (Grissa et al., 2008).⁸

Accuracy Evaluation of the WGS-Based MLVA Approach

To assess the performance of the WGS-based MLVA approach, we determined the percentage of agreement between PCR- and WGS-based MLVA methods by calculating the number of identical results (i.e., identical called alleles), divided by the total number of hits that were detected simultaneously by both approaches. Although the PCR-based MLVA approach is considered as the gold standard (with obvious accuracy increment after the optimization of the multiplex PCR with fluorescent

⁴<https://github.com/B-UMMI/INNUca>

⁵<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

⁶<https://github.com/tseemann/VelvetOptimiser>

⁷<https://github.com/dinergate42/in-silico-typing>

⁸<http://microbesgenotyping.i2bc.paris-saclay.fr/>

dyes for capillary electrophoresis), we believe the use of the “total number of hits that were detected simultaneously by both approaches” as the denominator, as the most reasonable and cautious procedure. In fact, although some hits were exclusively detected by the PCR-based MLVA approach, the opposite scenario was also observed. The maximum number of possible hits is 1,328 (i.e., 83 strains \times 16 loci).

Finally, for all strains, the performance of the bioinformatics script in extracting all MLVA loci was also evaluated by taking into account the quality of the draft genome generated by two assemblers (SPAdes and VelvetOptimiser) with and without “post-assembly polishing.” Briefly, for each condition, both the mean coverage depth and the number of contigs of each draft genome were correlated with the number of extracted alleles. Pearson’s coefficients (r) were measured to see potential linear associations. Nevertheless, as these final evaluations were done as complements of the major strategy, for the sake of clarity, whenever the text refers “WGS-based MLVA approach” it refers to the approach that used SPAdes with Pilon.

RESULTS

The performance of the WGS-based MLVA approach was assessed through the determination of the percentage of agreement with the results obtained by using the gold standard PCR-based MLVA. For the sake of clarity, we defined as “shared hits” the ones that were simultaneously detected by both approaches regardless their correct allele assignment. In this regard, the number of “matching alleles” was estimated using the number of “shared hits” as denominator.

Overall, the WGS-based MLVA approach detected 1,265 of the 1,328 possible hits (95.3%). One, two, and ≥ 3 loci yielded no results for 18 (21.6%), 7 (8.4%), and 7 (8.4%) of the strains, respectively (**Figure 1**). Regarding the PCR-based MLVA approach, it detected 1,269 (95.6%) hits, whereas one, two, and ≥ 3 loci yielded no results for 18 (21.6%), 8 (9.6%), and 6 (7.2%) strains, respectively.

The allelic profiles obtained by both approaches are presented in **Supplementary Table S2**. Of the total 1,328 possible hits, 1,222 (92.0%) were simultaneously detected by both approaches (“shared hits”), whereas 6.8% were differently detected solely by one approach. These discrepancies were mostly observed for loci Bru07 and Bru09 (both from panel 2B) and Bru11 (from panel 1; **Figure 2**). Both MLVA approaches simultaneously failed the detection of 16 out of the 1,328 hits (1.2%), where 12 fall in panel 2B, specially focusing loci Bru04 and Bru07.

The analysis *per locus* showed an agreement rate of MLVA profiles (ratio of “matching alleles” *per* number of “shared hits”) determined through PCR- and WGS-based approaches ranging from 89.9 to 100.0% (**Figure 2**). The discrepancies involved loci from all three panels, being more pronounced in six loci (Bru06, Bru12, Bru55, Bru19, Bru21, and Bru04). In general, the mean of agreement of MLVA profiles for all 16 loci was 96.4%, revealing a high allele concordance between the two approaches.

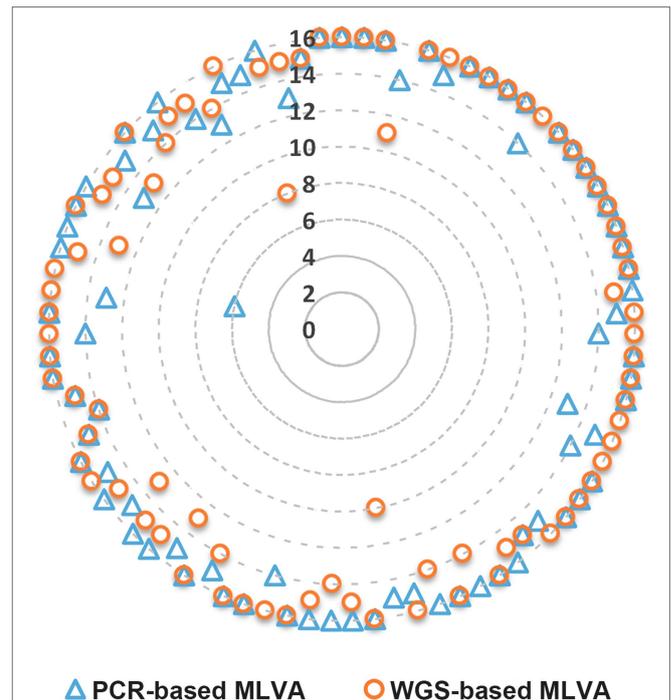
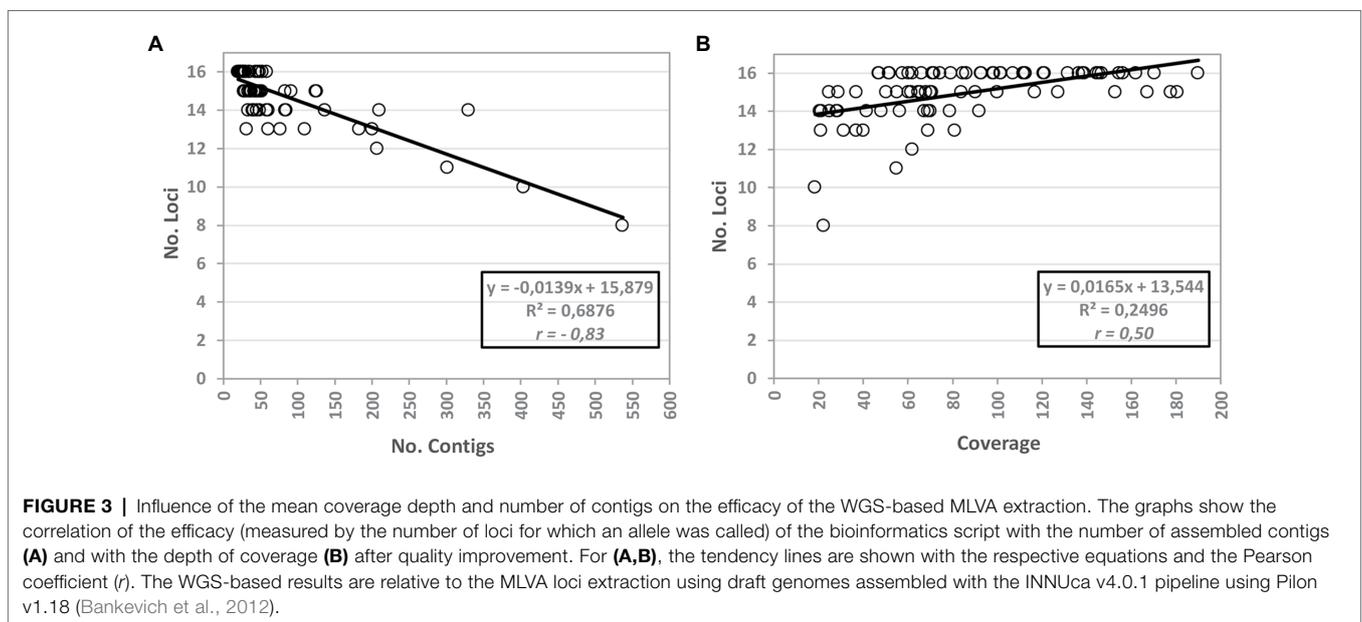
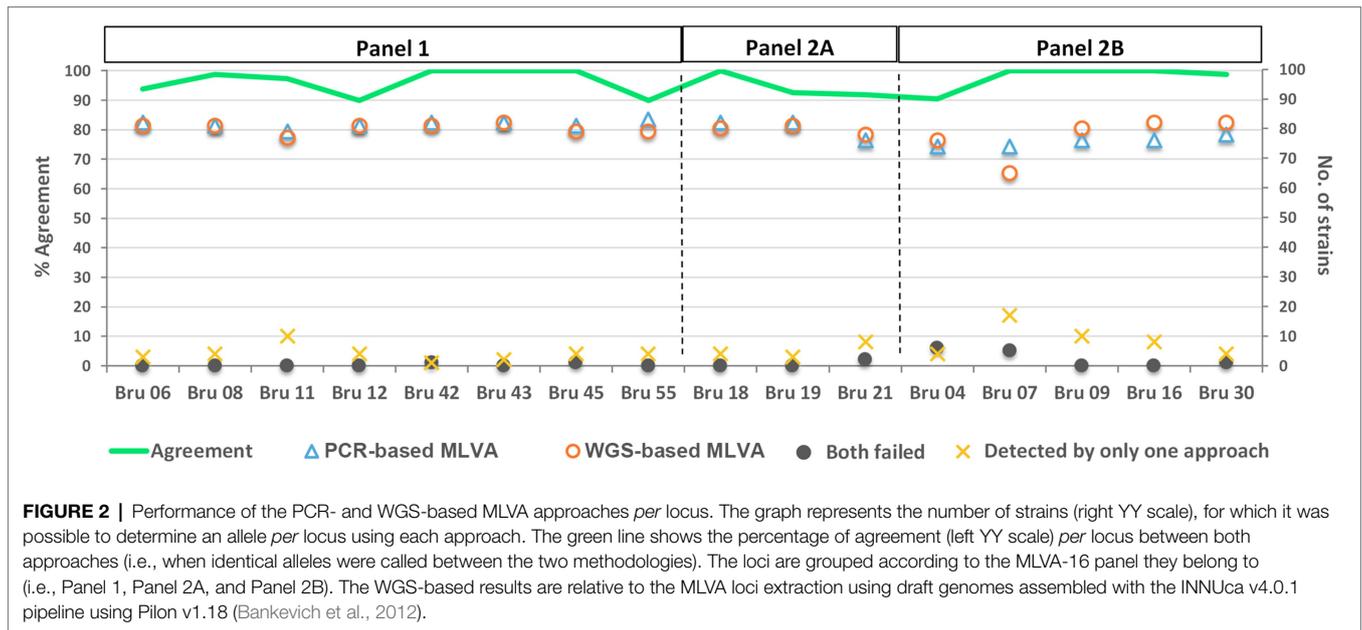


FIGURE 1 | Number of MLVA loci detected per strain through PCR- and WGS-based MLVA approaches. The data for each approach are disposed along the imaginary radius of the circle graph (one radius *per* strain). The WGS-based results are relative to the MLVA loci extraction using draft genomes assembled with SPAdes 3.11 (Walker et al., 2014) and subsequently improved with Pilon v1.18 (Bankevich et al., 2012) from the INNUca v4.0.1 pipeline (see methods for details).

We also inspected the quality of the draft sequences used as input because the Python script for *in silico* extraction of *Brucella* MLVA schemes is applied after the genome assembly and thus may be dependent on the quality of the “reads.” The influence of the mean depth of coverage and number of contigs on the efficacy of the bioinformatics script is illustrated in **Figure 3**. As expected, a negative linear correlation was observed among the efficacy of the *in silico* extraction and the number of assembled contigs, with less partitioned genomes allowing the detection of a higher number of alleles. On the other hand, higher genome mean coverage depth seem to favour the *in silico* extraction of MLVA loci.

As a final assessment, the performance of the bioinformatics script was also evaluated by using as input draft genome sequences assembled with different assemblers (SPAdes versus VelvetOptimiser; **Supplementary Figure S1**). Curiously, although no significant differences were observed regarding the number of loci extracted both with and without “post-assembly polishing” (data not shown), we observed that VelvetOptimiser was particularly affected by a low depth of coverage and genome fragmentation. In fact, for a mean depth of coverage < 50 as well as for high fragmented genomes, the number of detected MLVA loci decreased more sharply when we used VelvetOptimiser than when SPAdes assemblies were used.



DISCUSSION

The control of brucellosis requires an accurate surveillance and the use of high discriminatory methods to characterize outbreak strains and determine the infection source and transmission routes. For many years, multiple typing methods were used for *Brucella* characterization at both species and biovar levels. These relied on host specificity, growth features, biochemical reactions, serotyping and bacteriophage typing, but they lacked discriminatory power (Sun et al., 2017). There are some studies focusing the application of WGS-based approaches to *Brucella* spp., where strains from multiple species and locations are commonly included

(Pisarenko et al., 2018; Ashford et al., 2020; Esquivel et al., 2020; Pelerito et al., 2020). Nevertheless, currently, PCR-based MLVA is the most widely used approach for outbreak investigations and is still considered the gold standard for *Brucella* typing. Although the traditional MLVA assay relies on singleplex PCR followed by gel electrophoresis, some laboratories have already adopted a less error-prone approach based on multiplex PCRs and multicolor capillary electrophoresis (Garofolo, 2015; Vergnaud et al., 2018). The sixteen markers enrolled in the MLVA-16 scheme are a combination of moderately variable (minisatellites, panel 1) and highly discriminatory (microsatellites, panels 2A and 2B) loci (Al Dahouk et al., 2007). A MLVA typing

assay depends on the selection of markers which individually would not provide a relevant clustering. Taken separately, the Tandem Repeat markers are either not informative enough, are too variable or show a high level of homoplasmy. As such, the combination of well selected independent loci may be highly discriminatory as previously shown for other species (Le Fleche et al., 2006).

On behalf of the unavoidable transition from the classical typing to the WGS-based approaches, Python scripts were recently developed for the rapid *in silico* extraction of the *Brucella* MLVA alleles (Georgi et al., 2017; Vergnaud et al., 2018). This will allow the assignment of the MLVA types in the genomic era, avoiding the undesirable loss of genetic information that has been provided during more than 10 years by using the gold standard PCR-based MLVA typing. This is also important because the timings for the technology transition will vary considerably among different countries. As no experimental evaluation had been performed so far for one of those scripts (Georgi et al., 2017), our main goal was to access its performance against the gold standard method.

Overall, the WGS-based MLVA approach detected 95.3% of all possible 1,328 hits and the agreement with the experimental method in detecting the correct alleles was found to be 96.4%. Despite this high level of agreement, some loci could not be detected either by both or by solely a single approach. Although most of the loci did not simultaneously fail for several strains, it is worth highlighting Bru04 and Bru07 (both from panel 2B), for which the PCR-based MLVA approach failed for nine out of the 83 strains under study. We may hypothesize the occurrence of experimental difficulties associated with a less optimized PCR for these specific loci. However, we observed that Bru04 and Bru07 were also the loci for which the WGS-based MLVA approach most failed (for seven and 18 strains, respectively). We speculate that this relies on the fact that these loci are among the ones with the shortest repetitive sequence stretches. This could hypothetically be an obstacle for a more precise experimental distinction between the number of repeats. Bioinformatically, short repetitive stretches may eventually impact MLVA loci determination mainly when high genome fragmentation occurs.

According to the results obtained for the present dataset, we observed that the performance of the WGS-based MLVA approach does not seem to be dependent on the post-assembly polishing but is clearly dependent on the depth of coverage and the degree of assembly fragmentation (where SPAdes performed better). Still, a minimal depth of coverage of ~50x and a maximum number of ~200 contigs (a range where both assemblers behaved similarly) seem to constitute guiding “boundaries” for the future application of the script. Of note, some failures underlying the WGS-based MLVA approach could hypothetically be solved by the long-reads technology as it usually overcomes the assembly problems associated with the repeated sequences, although this technology is more prone to yield sequencing errors. On the other hand, the use of very short reads (e.g., 100 nt) may yield highly fragmented genomes and thus a more erroneous determination

of the repeat number. In this study, we used the Illumina technology with reads of 250–300 nt, which, together with a minimum depth of coverage of ~50x, yielded highly solid results.

In conclusion, although this study focused the most prevalent species (with emphasis in *B. melitensis*) the evaluated script seems to be a very useful and robust tool for *in silico* extraction of MLVA types of *Brucella* strains, dealing with a large number of samples in a short time period, and allowing retrospective and prospective molecular epidemiological studies. This allows a continuous and non-disruptive transition to a new typing era by putting the newly sequenced strains in the frame of the genetic characterization obtained for thousands of isolates collected worldwide throughout the last decade. This will certainly be important for public health reference laboratories to maintain an active epidemiological surveillance of brucellosis.

DATA AVAILABILITY STATEMENT

All raw sequence reads used in the present study were deposited in the European Nucleotide Archive under the study accession number PRJEB30030.

AUTHOR CONTRIBUTIONS

AP, AN, and JG conceived and designed the experiments, performed the analysis, interpreted the results, and wrote the manuscript. AP, JL, CS, AF, SV, and MN performed the experiments. EG performed the script optimization and implementation. All authors contributed to the article and approved the submitted version.

FUNDING

This work is a result of the GenomePT project (POCI-01-0145-FEDER-022184), supported by the COMPETE 2020 – Operational Program for Competitiveness and Internationalisation (POCI), Lisboa Portugal Regional Operational Program (Lisboa2020), Algarve Portugal Regional Operational Program (CRESC Algarve2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and by the Fundação para a Ciência e a Tecnologia (FCT). The studies have arisen from the Project QUANDHIP (Chafea Grant Agreement no. 2010 21 02), which has been funded by the European Commission in the framework of the Health Program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.740068/full#supplementary-material>

REFERENCES

- Al Dahouk, S., Flèche, P. L., Nöckler, K., Jacques, I., Grayon, M., Scholz, H. C., et al. (2007). Evaluation of *Brucella* MLVA typing for human brucellosis. *J. Microbiol. Methods* 69, 137–145. doi: 10.1016/j.mimet.2006.12.015
- Ashford, R. T., Muchowski, J., Koylass, M., Scholz, H. C., and Whatmore, A. M. (2020). Application of whole genome sequencing and pan-family multi-locus sequence analysis to characterize relationships within the family Brucellaceae. *Front. Microbiol.* 11:1329. doi: 10.3389/fmicb.2020.01329
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Borges, V., Pinheiro, M., Pechirra, P., Guiomar, R., and Gomes, J. P. (2018). INSAFLU: an automated open web-based bioinformatics suite “from-reads” for influenza whole-genome-sequencing-based surveillance. *Genome Med.* 10:46. doi: 10.1186/s13073-018-0555-0
- Bricker, B. J., Ewalt, D. R., and Halling, S. M. (2003). *Brucella* 'HOOF-Prints': strain typing by multi-locus analysis of variable number tandem repeats (VNTRs). *BMC Microbiol.* 3:15. doi: 10.1186/1471-2180-3-15
- De Santis, R., Ciammaruconi, A., Faggioni, G., Fillo, S., Gentile, B., di Giannatale, E., et al. (2011). High throughput MLVA-16 typing for *Brucella* based on the microfluidics technology. *BMC Microbiol.* 11:60. doi: 10.1186/1471-2180-11-60
- Dean, A. S., Crump, L., Greter, H., Schelling, E., and Zinsstag, J. (2012). Global burden of human brucellosis: a systematic review of disease frequency. *PLoS Negl. Trop. Dis.* 6:e1865. doi: 10.1371/journal.pntd.0001865
- Franc, K. A., Krecke, R. C., Hasler, B. N., and Arenas-Gamboa, A. M. (2018). Brucellosis remains a neglected disease in the developing world: a call for interdisciplinary action. *BMC Public Health* 18:125. doi: 10.1186/s12889-017-5016-y
- Franco, M. P., Mulder, M., Gilman, R. H., and Smits, H. L. (2007). Human Brucellosis. *Lancet Infect. Dis.* 7, 775–786. doi: 10.1016/S1473-3099(07)70286-4
- Garofolo, G. (2015). Multiple-locus variable-number tandem repeat (VNTR) analysis (MLVA) using multiplex PCR and multicolor capillary electrophoresis: application to the genotyping of *Brucella* species. *Methods Mol. Biol.* 1247, 335–347. doi: 10.1007/978-1-4939-2004-4_24
- Garofolo, G., Ancora, M., and Giannatale, E. (2013). MLVA – 16 loci panel on *Brucella* spp. using multiplex PCR and multicolor capillary electrophoresis. *J. Microbiol. Methods* 92, 103–107. doi: 10.1016/j.mimet.2012.11.007
- Gee, J. E., de, B. K., Levett, P. N., Whitney, A. M., Novak, R. T., and Popovic, T. (2004). Use of 16s rRNA gene sequencing for rapid confirmatory identification of *Brucella* isolates. *J. Clin. Microbiol.* 42, 3649–3654. doi: 10.1128/JCM.42.8.3649-3654.2004
- Georgi, E., Walter, M. C., Pfalzgraf, M. T., Northoff, B. H., Holdt, L. M., Scholz, H. C., et al. (2017). Whole genome sequencing of *Brucella melitensis* isolated from 57 patients in Germany reveals high diversity in strains from Middle East. *PLoS One* 12:e0175425. doi: 10.1371/journal.pone.0175425
- Gopaul, K. K., Koylass, M. S., Smith, C. J., and Whatmore, A. M. (2008). Rapid identification of *Brucella* isolates to the species level by real time PCR based single nucleotide polymorphism (SNP) analysis. *BMC Microbiol.* 8:14. doi: 10.1186/1471-2180-8-86
- Grissa, I., Bouchon, P., Pourcel, C., and Vergnaud, G. (2008). On-line resources for bacterial micro-evolution studies using MLVA or CRISPR typing. *Biochimie* 90, 660–668. doi: 10.1016/j.biochi.2007.07.014
- Hanot Mambres, D., Boarbi, S., Michel, P., Bouker, N., Escobar-Calle, L., Desqueper, D., et al. (2017). Imported human brucellosis in Belgium: BIO and molecular typing of bacterial isolates, 1996–2015. *PLoS One* 12:e0174756. doi: 10.1371/journal.pone.0174756
- Janowicz, A., de Massis, F., Ancora, M., Cammà, C., Patavino, C., Battisti, A., et al. (2018). Core genome multilocus sequence typing and single nucleotide polymorphism analysis in the epidemiology of *Brucella melitensis* infections. *J. Clin. Microbiol.* 56, e00517–e00518. doi: 10.1128/JCM.00517-18
- Kattar, M. M., Jaafar, R. F., Araj, G. F., Le Flèche, P., Matar, G. M., Abi Rached, R., et al. (2008). Evaluation of a multilocus variable number tandem repeat analysis scheme for a typing human *Brucella* isolates in a region of brucellosis endemicity. *J. Clin. Microbiol.* 46, 3935–3940. doi: 10.1128/JCM.00464-08
- Le Flèche, P., Jacques, I., Grayon, M., Al Dahouk, S., Bouchon, P., Denoeud, F., et al. (2006). Evaluation and selection of tandem repeat loci for a *Brucella* MLVA typing assay. *BMC Microbiol.* 6:9. doi: 10.1186/1471-2180-6-9
- MLVA Bank (n.d.). *MLVA Bank for Microbes Genotyping*. Available at: <http://microbesgenotyping.i2bc.paris-saclay.fr/>
- Pelerito, A., Matos, R., Santos, M. A., Soeiro, S., Santos, J., Manita, C., et al. (2017). Human brucellosis in Portugal – Retrospective analysis of suspected clinical cases of infection from 2009 to 2016. *PLoS One* 12:e0179667. doi: 10.1371/journal.pone.0179667
- Pelerito, A., Nunes, A., Nuncio, M. S., and Gomes, J. P. (2020). Genome-scale approach to study the genetic relatedness among *Brucella melitensis* strains. *PLoS One* 15:e0229863. doi: 10.1371/journal.pone.0229863
- Pinto, M., González-Díaz, A., Machado, M. P., Duarte, S., Vieira, L., Carriço, J. A., et al. (2018). Insights into the population structure and pan-genome of *Haemophilus influenzae*. *Infect. Genet. Evol.* 67, 126–135. doi: 10.1016/j.meegid.2018.10.025
- Pisarenko, S. V., Kovalev, D. A., Volynkina, A. S., Ponomarenko, D. G., Rusanova, D. V., Zharinova, N. V., et al. (2018). Global evolution and phylogeography of *Brucella melitensis* strains. *BMC Genomics* 19:353. doi: 10.1186/s12864-018-4762-2
- Sankarasubramanian, J., Vishnu, U. S., Gunasekaran, P., and Rajendhran, J. (2019). Development and evaluation of a core genome multilocus sequence typing (cgMLST) scheme for *Brucella* spp. *Infect. Genet. Evol.* 67, 38–43. doi: 10.1016/j.meegid.2018.10.021
- Suárez-Esquivel, M., Hernández-Mora, G., Ruiz-Villalobos, N., Barquero-Calvo, E., Chacón-Díaz, C., Ladner, J. T., et al. (2020). Persistence of *Brucella abortus* lineages revealed by genomic characterization and phylogenetic analysis. *PLoS Negl. Trop. Dis.* 14:e0008235. doi: 10.1371/journal.pntd.0008235
- Sun, M., Jing, Z., di, D., Yan, H., Zhang, Z., Xu, Q., et al. (2017). 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.* 4:215. doi: 10.3389/fvets.2017.00215
- Tan, K. K., Tan, Y. C., Chang, L. Y., Lee, K. W., Nore, S. S., Yee, W. Y., et al. (2015). Full genome SNP-based phylogenetic analysis reveals the origin and global spread of *Brucella melitensis*. *BMC Genomics* 16:93. doi: 10.1186/s12864-015-1294-x
- Valdezate, S., Navarro, A., Villalón, P., Carrasco, G., and Saéz-Nieto, J. A. (2010). Epidemiological and phylogenetic analysis of spanish human *Brucella melitensis* strains by multiple locus variable number tandem repeat typing, hypervariable octameric oligonucleotide fingerprinting, and rpoB typing. *J. Clin. Microbiol.* 48, 2734–2740. doi: 10.1128/JCM.00533-10
- Vergnaud, G., Hauck, Y., Christiany, D., Daoud, B., Pourcel, C., Jacques, I., et al. (2018). Genotypic expansion within the population structure of classical *Brucella* species revealed by MLVA16 typing of 1404 *Brucella* isolates from different animal and geographic origins. *Front. Microbiol.* 9:1545. doi: 10.3389/fmicb.2018.01545
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963
- Whatmore, A. M., Shankster, S. J., Perrett, L. L., Murphy, T. J., Brew, S. D., Thirlwall, R. E., et al. (2006). Identification and characterization of VARIABLE-number tandem – REPEAT MARKERS for typing of *Brucella* spp. *J. Clin. Microbiol.* 44, 1982–1993. doi: 10.1128/JCM.02039-05
- WHO (2020). *Neglected Tropical Diseases*. Available at: http://www.who.int/neglected_diseases/zoonoses/other_NZDs/en/ (Accessed October 06, 2021).
- Young, E. J. (2005). “*Brucella* species” in *Principles and Practice of Infectious Diseases*. 6th Edn. eds. G. L. Mandell, J. E. Bennet and R. Dolin (Philadelphia: Churchill Livingstone), 2669–2772.

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may

be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Pelerito, Nunes, Grilo, Isidro, Silva, Ferreira, Valdezate, Nuncio, Georgi and Gomes. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.