Infection, Genetics and Evolution 67 (2019) 38–43

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Development and evaluation of a core genome multilocus sequence typing (cgMLST) scheme for *Brucella* spp.

Jagadesan Sankarasubramanian[a], Udayakumar S. Vishnu[a], Paramasamy Gunasekaran[b], Jeyaprakash Rajendhran[a,*]

[a] Department of Genetics, School of Biological Sciences, Madurai Kamaraj University, Madurai 625021, Tamil Nadu, India
[b] VIT Bhopal University, Bhopal, Madhya Pradesh, India

ARTICLE INFO

Keywords:
Brucella
cgMLST
wgMLST
Biovars
Sequence type

ABSTRACT

Brucellosis is a zoonotic disease caused by *Brucella* spp. *Brucella* spp. can be sub-typed by multilocus sequence typing (MLST) method, which targets a set of housekeeping genes. We have developed a core genome MLST (cgMLST) typing scheme to distinguish and differentiate species of *Brucella* up to biovar level. A total of 407 whole (complete and draft) genome sequences of different *Brucella* strains were used in this study. Genome sequences were filtered using the BLAST score ratio (BSR)-based allele calling algorithm, and we found that 164 cgMLST target loci are shared in all the 407 genome sequences. These 164 loci were used to develop the cgMLST scheme and further evaluated to sub-type different species of *Brucella*. Based on our cgMLST scheme, *Brucella* spp. were classified into 287 sequence types (STs). A phylogenetic tree was constructed based on the STs derived from the cgMLST analysis. The phylogenetic tree differentiated all the 11 *Brucella* spp. and five biovars of *B. suis*. *B. vulpis* formed the outmost clade followed by *B. inopinata* and *B. microti*. Among the four subgroups of *B. abortus*, group A and B were differentiated based on their geographic origins. Similarly, three subgroups of *B. melitensis* were separated based on their geographical origins with few exceptions. *B. neotomae* that infect rodents were distinguished from other *Brucella* spp. *B. canis* showed the closest relationship with *B. suis* bv. 4, followed by *B. suis* bv. 3 and bv. 1. *Brucella* spp. associated with the marine mammals, such as *B. ceti* and *B. pinnipedialis* were closely related. Of these, *B. ceti* strains isolated from dolphins and porpoise were differentiated into two groups. We incorporated our cgMLST tool in BrucellaBase (http://www.dbtbrucellosis.in/brucella_cgmlst.html), which will be helpful to predict the cgMLST allelic profile and the ST of a newly sequenced genome.

1. Introduction

Brucellosis is one of the world's most important zoonotic diseases and continues to have a significant impact on animals and humans (Pappas et al., 2006). *Brucella* genomes are composed of two circular chromosomes of approximately 2.1 and 1.2 Mb in size (Michaux-Charachon et al., 1997). Both chromosomes share a similar GC content, a similar proportion of coding regions, and an equal housekeeping gene distribution (Ficht, 2011). At present, eleven species of *Brucella* have been recognized with apparent host-preference, which include six classical species, *B. abortus* (cattle), *B. melitensis* (sheep and goats), *B. suis* (pigs, hares, wild boar, reindeer and rodent), *B. canis* (dogs), *B. ovis* (sheep), and *B. neotomae* (rodents), and five newly recognized species, *B. microti* (voles), *B. pinnipedialis* (seals), *B. ceti* (cetaceans), *B. inopinata* (isolated from a human, natural host unknown) (Scholz et al., 2010) and *B. vulpis* (foxes) (Scholz et al., 2016). Among these, *B. melitensis*, *B. abortus*, and *B. suis* are largely responsible for human brucellosis. *B. canis* and *B. ovis* seldom infect humans (Marzetti et al., 2013). *B. neotomae* is a rodent pathogen, and infection by *B. neotomae* was thought to be limited to wood rats. However, isolation of *B. neotomae* from the cerebrospinal fluid of a neurobrucellosis patient, recently, questioned the nonzoonotic status of *B. neotomae* (Suarez-Esquivel et al., 2017). *B. ceti* and *B. pinnipedialis* are associated with marine mammals, and *B. microti* strains were isolated from the common vole and red fox. Human infections by these three species are not reported (Nymo et al., 2011). *B. inopinata* is confined to a single human case. However, the reservoir of infection for this species has not been identified (Kaltungo et al., 2014; Whatmore et al., 2016).

Methods based on the genome sequence analysis are promising for accurate, and reproducible typing of bacteria. Multilocus sequence typing (MLST) methods based on 9 or 21 loci (Whatmore et al., 2007, 2016), and multilocus VNTR analysis (MLVA) using 16 loci (Le Fleche

* Corresponding author at: Department of Genetics, School of Biological Sciences, Madurai Kamaraj University, Madurai 625021, Tamil Nadu, India.
E-mail address: jrajendhran@gmail.com (J. Rajendhran).

https://doi.org/10.1016/j.meegid.2018.10.021
Received 27 December 2017; Received in revised form 29 September 2018; Accepted 27 October 2018
Available online 30 October 2018
1567-1348/

et al., 2006) are currently used for typing *Brucella* spp. Due to the developments in DNA sequencing technologies, genomes of several bacteria, including various strains of the same species are sequenced and made available in the databases. Therefore, the genome-wide comparison is frequently used to analyze the whole genome sequence (WGS) data for molecular typing (Collins et al., 1999; Schork et al., 2001). The whole genome multilocus sequence typing (wgMLST) is an extended concept of the traditional MLST (Jolley and Maiden, 2010), developed to use WGS data and to produce a reproducible phylogeny. Recently, the core-genome multilocus sequence typing (cgMLST) schemas were developed for few bacterial pathogens such as *Campylobacter jejuni* and *Acinetobacter baumannii* (Cody et al., 2017; Higgins et al., 2017). In this study, we performed a WGS-based typing using 407 WGS data. From the wgMLST, we extracted the core gene set and developed the cgMLST scheme for the phylogeny analysis of *Brucella* spp.

2. Materials and methods

2.1. Retrieval of whole genome sequences

A total of 685 *Brucella* genome sequences were available in NCBI and BrucellaBase (Sankarasubramanian et al., 2016a,b) on 15th March 2018. We have selected all the available 161 complete genome sequences and 245 draft genome sequences with < 25 contigs and > 60× sequence coverage. Since only two draft genomes were available for *B. inopinata*, we have included one of the draft genomes with 55 contigs. Thus, a total of 407 WGS of *Brucella* strains belonging to 11 different species (192 *B. abortus*, 18 *B. canis*, 7 *B. ceti*, 1 *B. inopinata*, 121 *B. melitensis*, 1 *B. microti*, 1 *B. neotomae*, 11 *B. ovis*, 3 *B. pinnipedialis*, 50 *B. suis* and 2 *B. vulpis*) were used in this study (Supplementary Table S1). The genome sequence of *B. abortus* 2308 (chromosome I -NC_007618 and Chromosome II - NC_007624) was used as the reference genome (Chain et al., 2005) to predict the wgMLST loci.

2.2. Development of cgMLST scheme

We performed WGS-based typing using a BLAST Score Ratio-Based Allele Calling Algorithm (chewBBACA) available at (https://github.com/INNUENDOCON/chewBBACA). The chewBBACA is a complete pipeline for creating and validating wgMLST/cgMLST schemas. All the 407 genome sequences were re-annotated using Prodigal 2.6.0 (Hyatt et al., 2010). The CDS files were further used to identify the wgMLST loci using chewBBACA with the default parameter (minimum BSR locus similarity of 0.6). The created alignment files were used to determine the cgMLST loci, which are conserved in all the 407 genomes used in this study. Based on the SNPs in the core loci, allelic profiles were generated, and the sequence types (STs) were assigned to each unique allelic profile. The detailed workflow for the development of cgMLST is given in Fig. 1.

2.3. Phylogenetic tree construction based on cgMLST

For each genome, sequences of the 164 loci were concatenated using EMBOSS union (Rice et al., 2000), and aligned using Multiple Alignment using Fast Fourier Transform (MAFFT) (Katoh and Standley, 2013). Based on the aligned sequences, FastTree2 was used to construct a cgMLST based phylogenetic tree using the maximum-likelihood method with 1000 bootstrap replicates. FastTree2 uses the Jukes-Cantor of generalized time reversible models of nucleotide evolution (Price et al., 2010). The phylogenetic tree was visualized using Interactive Tree of Life (iTOL) (Letunic and Bork, 2016).

2.4. Phylogenetic tree construction based on MLST

We have also predicted the allelic profiles of nine house-keeping genes for all the 407 genomes using BrucellaBase (Sankarasubramanian et al., 2016a,b). The predicted MLST sequences were aligned using MAFFT, and the MLST-based phylogenetic tree was constructed using FastTree2 employing the maximum-likelihood method with default bootstrap values. The phylogenetic tree was visualized using iTOL.

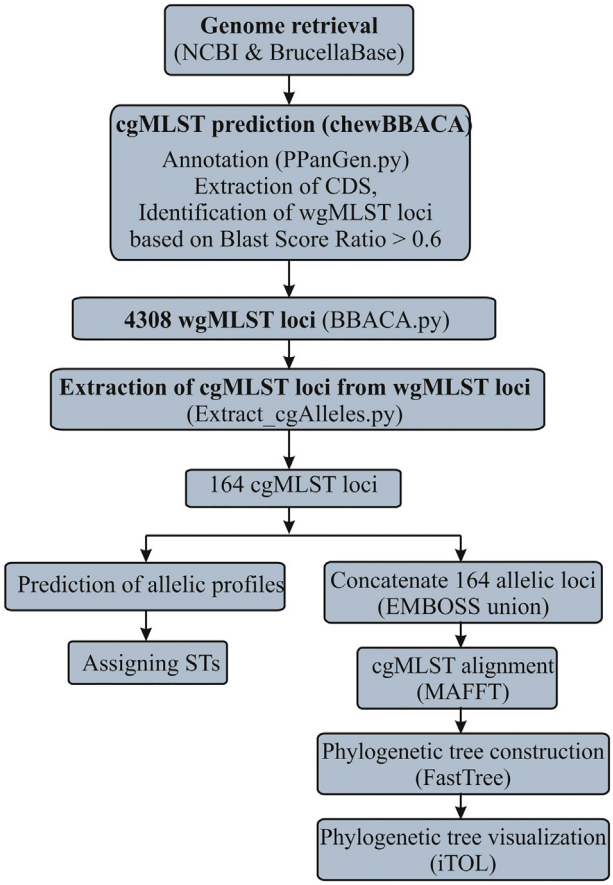

Fig. 1. Detailed workflow for the development of cgMLST scheme for *Brucella* spp.

2.5. Creation of cgMLST tool in BrucellaBase

We developed a user-friendly interface for the prediction of allelic profiles of cgMLST loci and the respective STs from the whole genome sequences of *Brucella*. The system design is based on secure web application architecture of client workstation, web server, application server and database server implemented by PHP: Hypertext Preprocessor (PHP), Hyper Text Markup Language (HTML) 5. The cgMLST tool is added to the existing *Brucella* genome resource, BrucellaBase. We also incorporated a cgMLST sequence alignment and phylogenetic analysis tool based on the STs using SeaView (version 4) (Gouy et al., 2010). The phylogenetic tree can be constructed and visualized using the concatenated sequences of 164 loci.

3. Results

3.1. Development of cgMLST scheme

Whole genome sequences of 407 *Brucella* strains (Supplementary Table S1) were used in this study. Based on the whole genome typing using PanGen.py in chewBBACA tool, which uses Prodigal annotation, a total of 4308 genetic loci were predicted as wgMLST targets (Fig. 1). Of these, only 164 loci were selected as the cgMLST targets that are shared in all the 407 *Brucella* genome sequences (Supplementary Table S2). Other genetic loci were missing in at least one of the genome sequences

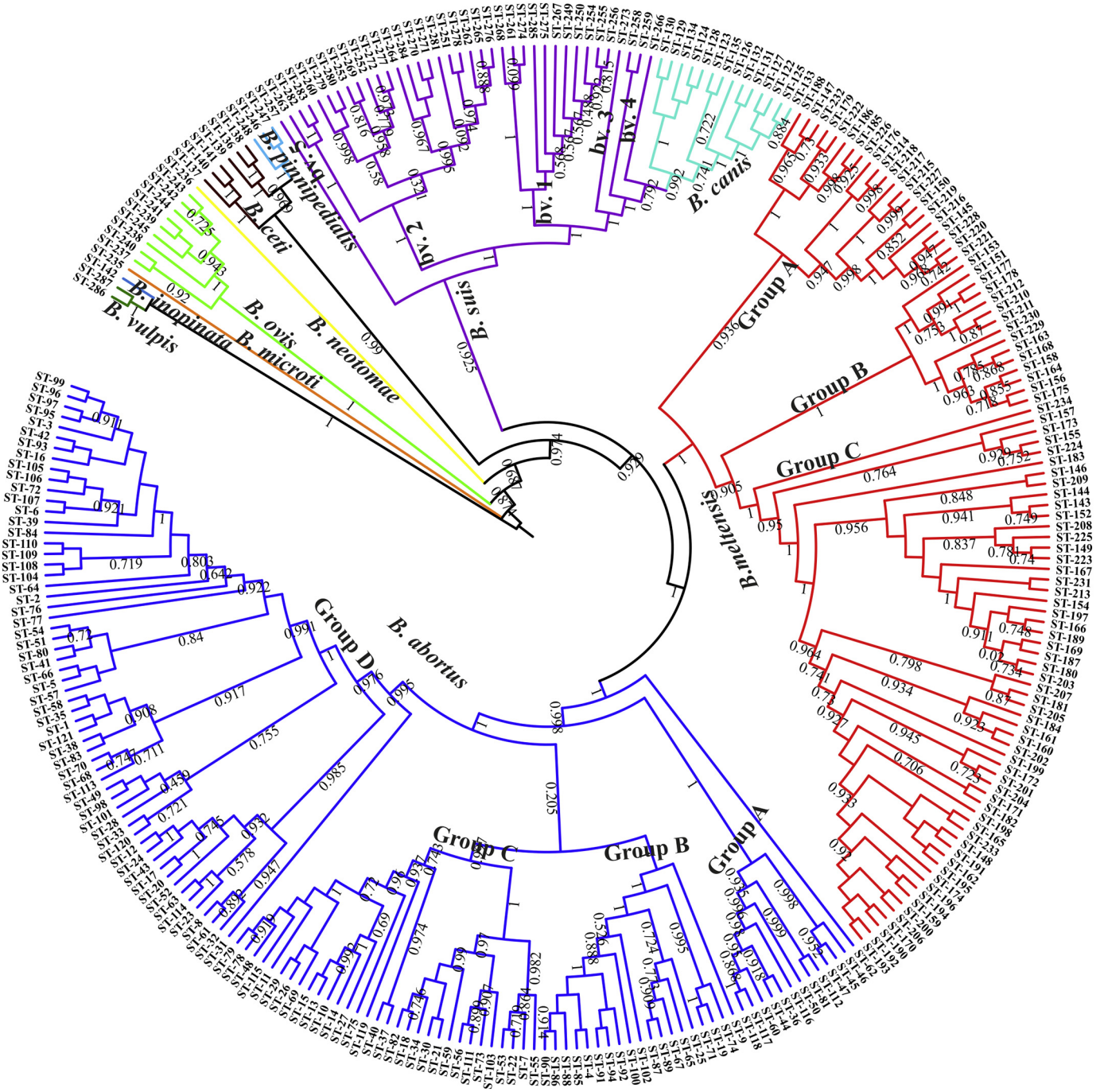

Fig. 2. Predicted STs based on cgMST, which distinguishes *Brucella* sp. and biovar of *B. suis*. The phylogenetic tree was constructed based on the concatenated sequences of 164 cgMLST allelic profiles. The concatenated sequences of STs were aligned using MAFFT and the phylogenetic tree was constructed using FastTree with maximum-likelihood method with 1000 bootstrap replicates and visualized using iTOL.

used in this study. The allelic profiles of 407 genomes are given in the Supplementary Table S3. Based on the allelic profiles of the 164 cgMLST loci, 287 STs were identified in *Brucella* spp. (Table 1). We have constructed a cgMLST-derived phylogenetic tree based on the predicted allelic profiles and their STs. The phylogenetic tree differentiated all the 11 *Brucella* spp. and biovars of *B. suis* (Fig. 2).

3.2. Differentiation of B. abortus

Based on the cgMLST, STs 1 to 121 was assigned to *B. abortus*. All the 192 *B. abortus* strains could be divided into four major groups (A to D) (Fig. 2). The strains belong to each ST, and their geographical regions are shown in Supplementary Fig. S1. Group A consists of 18 *B. abortus* strains with 13 different STs (ST-36, ST-44 to ST-47, ST-50, ST-60, ST-62, ST-81, ST-112, ST-116 to ST-118). The *B. abortus* strains of these STs were isolated from *Bos taurus* in Africa. In this group only four genomes clustered together and showed as ST-81 and the other genomes have their own STs. Group B consist of 26 *B. abortus* strains with 19 different STs. These strains were isolated from various hosts like *B. taurus* and *Bubalus bubalis* and American bison. However, all these strains were isolated from Europe and North America. Group C consisted of 43 strains with 30 different STs isolated from various hosts like *B. taurus*, and *B. bubalis*, which was isolated from Asia and Europe. Group D was the largest group consisting 107 *B. abortus* strains with 60

different STs. These strains were isolated from humans, *Rangifer tarandus, Lama glama, Mus musculus,* caprien and Jerusalem swine. These strains were reported from all over the world. The large numbers of strains were clustered with ST-95 in *B. abortus* group D, which consists of 22 strains and all the genomes were isolated from Europe. The vaccine strain, *B. abortus* S19 was clustered in group D and differentiated from other strains with ST-120. Among 192 *B. abortus* genomes a total of 121 STs were predicted, in this 99 STs were identified as strain- specific STs, only remaining 22 STs were shared among 93 genomes.

3.3. Differentiation of B. melitensis

STs 143 to 235 were assigned to the 121 *B. melitensis* strains (Fig. 2), these strains belonging to each ST, their isolation sources, and geographical regions are shown in Supplementary Fig. S1. *B. melitensis* was divided into three major groups (A to C). Group A consisted of 21 different *B. melitensis* strains with 21 different STs. These strains were isolated from humans, *B. taurus* and the domestic goat, *Capra hircus* from North America, Europe, and Africa. Group B consisted of 18 strains of 15 different STs isolated from humans and *C. hircus* and *B. taurus*. These strains were isolated from all over the world. Group C consists of 82 strains with 52 different STs. These strains isolated from humans and *Ovis aries* from Asia, Europe and Eurasia. In group C, the vaccine strain *B. melitensis* M5–90 (Wang et al., 2013) used in China and its parental strain *B. melitensis* M28 were clustered together. Among the three subgroups of *B. melitensis* group A all the genomes were distinguished with different STs. Likewise, in group B except for five strains, other *B. melitensis* were well differentiated with separate STs. cgMLST analysis showed a better resolution with the existing methods and most of the *B. melitensis* genomes have their STs.

3.4. Differentiation of B. suis and B. canis

The STs 249 to 285 were assigned to *B. suis*. The strains belonging to each ST, their isolation sources, and geographical regions are shown in Supplementary Fig. S1. All the biovars of *B. suis* and the *B. canis* strains formed a single clade. This clade could be further differentiated into five groups representing the *B. suis* biovars 5, 2, 1, 3 and 4, and the *B. canis*, respectively. The 26 *B. suis* bv. 2 strains with 21 different STs formed the major group of *B. suis-B. canis* clade. All these strains were isolated from hare and wild boar, and exclusively from European countries. *B. suis* bv. 1 group consisted of 16 strains with STs 249, 250, 254–256, 261, 267, 268, 274, 275 and 285 were isolated from North America, Asia, and Africa. These strains were isolated from various hosts including humans and *B. taurus*. Presently, the genome sequence of only one *B. suis* bv. 3 strain is available, which represent the ST-273. This strain has been isolated from *S. scrofa* in North America. The *B. suis* bv. 4. strains with STs 258, 259 and 266 were isolated from North America, Europe, and Eurasia. The *B. canis* (STs 122 to 135) showed the closest relationship with *B. suis* bv. 4 strains. *B. canis* strains were isolated from the dogs and humans, from various parts of the world. Interestingly, *B. suis* bv. 5 strains with ST-257 were distinct from other biovars of *B. suis*, which causes the infection to humans.

3.5. Differentiation of Brucella spp. associated with marine mammals

Brucella spp. associated with the marine mammals, *B. ceti* (STs 136 to 141) and *B. pinnipedialis* (STs 246 to 248) showed the closest relationship. Based on the STs, *B. ceti* was divided into two groups. *B. ceti* isolated from dolphins (STs 137, 140, and 141), and porpoise (STs 136, 138 and 139) differentiated into two groups. *B. pinnipedialis* (STs 246 to 248) isolated from common seals formed as a separate group (Fig. 2).

Table 1
Predicted ST based on cgMLST.

Organisms	No. of strains	cgMLST
B. abortus	192	ST-1 to ST-121
B. canis	18	ST-122 to ST-135
B. ceti	7	ST-136 to ST-141
B. inopinata	1	ST-142
B. melitensis	121	ST-143 to ST-235
B. microti	1	ST-236
B. neotomea	1	ST-237
B. ovis	11	ST-238 to ST-245
B. pinnipedialis	3	ST-246 & ST-248
B. suis	50	ST-249 to ST-285
B. vulpis	2	ST-286 & ST-287

3.6. Other Brucella spp.

The recently identified species, *B. vulpis* isolated from *Vulpes vulpes* (Red fox), formed a sister clade with the assigned STs 286 and 287. *B. inopinata* strain (ST-142) isolated from humans showed the closest relationship with *B. vulpis*. *B. ovis* strains isolated from sheep, from various countries formed a separated group. *B. ovis* (STs 237 to 245) showed closer relationships with *B. microti* (ST-235) and *B. neotomae* (ST-236), isolated from common vole and desert wood rat, respectively. Only a limited number of genome sequences were available *B. vulpis*, *B. inopinata*, *B. microti* and *B. neotomae*.

3.7. Comparison of cgMLST with MLST

In addition to cgMLST prediction, we have also compared the cgMLST scheme with the publicly available MLST tool. All the 407 genomes were taken for this analysis, and the STs were predicted using BrucellaBase (Supplementary Fig. S2). The MLST typing failed to distinguish the *B. suis* biovar of 3 and 4, but cgMLST clearly distinguished all the biovars of *B. suis*. In MLST-based phylogeny analysis, 111 strains of *B. abortus* was classified as ST-1, whereas these 111 *B. abortus* strains could be differentiated into 60 different STs in the cgMLST scheme. Similarly, the largest clade of *B. melitensis* with 85 strains belonging to the ST-8 could be separated into 52 different STs in the cgMLST. Overall, 71 STs were identified in MLST analysis 286 STs were predicted in cgMLST analysis. Using, cgMLST *B. abortus* and *B. melitensis* could be differentiated according to their geographical origin.

3.8. Incorporation of cgMLST tool in the BrucellaBase

We have incorporated the newly developed cgMLST tool as a web resource in the *Brucella* genome repository, BrucellaBase. The users paste the *Brucella* genome (either complete or draft) sequence in the provided text box in FASTA format. The input sequence is then compared with the cgMLST loci and STs are identified. If the user's sequence is identical to any of the existing sequences in the cgMLST database (STs 1 to 287), it will display the allelic numbers, start and end positions of each gene, the percentage identity, *E*-value, and matched sequences. Thus, the *Brucella* cgMLST is based on BLAST score with 100% identity, alignment without any gaps and/or mismatch. The result page will display the allelic profile numbers of all loci, and the mapped regions in the query genome. After the prediction of allelic profiles, the user can click the “Predict ST” button to identify the ST of the given sequence. In addition to the STs, the species name and the concatenated allelic profile sequences will also be displayed. We have also incorporated a cgMLST BLAST tool to identify the presence or absence of the 164 target loci in the query genome. For instance, if a draft genome is missing any of these 164 loci, STs cannot be predicted. Under such circumstances, cgMLST BLAST tool can be used to check the presence or absence of the cgMLST loci. We also incorporated a phylogenetic tree construction tool in BrucellaBase based on cgMLST sequences, a

minimum of three concatenated sequences of 164 loci can be used for the multiple sequence alignment. The sequence alignment and the constructed phylogenetic tree were displayed on the BrucellaBase.

4. Discussion

Microbial sub-typing is used in tracking the spread of infectious diseases. Sequence-based typing methods are most reproducible than the gel-based methods. Sequence typing based on the differences in the conserved genes such as 16 rRNA gene usually classify the bacterial strains into fewer and larger phylogenetic groups when compared to the whole-genome based typing. The discriminatory power of a single locus-specific sequence analysis is generally lesser since it involves only a limited region of the genome. In *Brucella* spp., the phylogenetic markers such as 16S rRNA genes are highly conserved with almost 100% identity. Therefore, sequencing of several conserved genes (multi-locus sequence typing, MLST) had been proposed to improve the discriminatory power. MLST is one of the most widely used pathogen sub-typing methods developed in the 1990s (Maiden et al., 1998).

For *Brucella* typing, Whatmore et al. (2007) have proposed an MLST scheme employing nine loci such as *gap*, *aroA*, *glk*, *dnaK*, *gyrB*, *rtpE*, *cobQ*, *omp25*, and int-*hyp* and identified 27 STs based on 160 *Brucella* isolates (Whatmore et al., 2007). Ma et al. (2016) have typed 65 isolates from China using the MLST scheme and identified five known STs: ST2, ST5, ST7, ST8 and ST14. They have also identified a novel STs and reported as a mutant type of ST8. Chawjiraphan et al. (2016) reported the MLST analysis of 36 *Brucella* isolates from Thailand. MLST analysis of these 36 isolates revealed 34 novel STs, while only two isolates matched with the known type, ST8. By analyzing 510 *Brucella* genome sequences, we have reported 43 additional STs (Sankarasubramanian et al., 2016a,b). Subsequently, Whatmore et al. (2016) have proposed an extended MLST scheme with 21 loci. By analyzing 500 *Brucella* isolates, they have reported 101 STs (designated as BruMLSA21 STs). We have identified 287 cgMLST STs in *Brucella*, and the phylogeny analysis showed region-wide discrimination of *B. abortus* and *B. melitensis* strains.

The detailed information on the characteristics of different strains can be obtained only by the complete genome sequence analysis. Therefore, the genome-based analysis may give better insights into the microbial taxonomy, evolution, and phylogenetic analysis. In this direction, Foster et al. (2009) have compared the whole genomes 13 strains from five different species such as *B. abortus*, *B. melitensis*, *B. suis*, *B. canis*, and *B. ovis* and identified 20,154 orthologous SNPs (core SNPs). Genome-wide SNP based phylogeny analysis showed that *B. suis* was paraphyletic with the closest relationship with *B. canis*. Similarly, Wattam et al. (2014) performed a comparative phylogenomic analysis and reported that *Brucella* spp. could be classified into two broad groups, such as atypical strains and a highly conserved classical core clade containing all the major pathogenic species. The outmost group included *Brucella* strains from Australian rodents (83/13 and NF2653) and two recent atypical human isolates, *B. inopinata* BO1 and BO2. Further, they have reported the identification of 2672 SNPs unique to the classic *Brucella* strains and 1172 SNPs unique to the atypical group. Based on the genome-wide SNPs-based analysis of 54 *B. suis* strains, we have reported the biovar-specific SNPs of *B. suis* (Sankarasubramanian et al., 2016a). Recently, whole genome sequences of 57 *B. melitensis* strains were compared by Georgi et al. (2017), and the genome-wide SNP-based analysis differentiated *B. melitensis* into four major lineages such as East Mediterranean clade, American clade, African clade, and West Mediterranean clade.

In this study, we have developed a cgMLST scheme for *Brucella* typing based on the core loci found in 407 whole genome sequences. We identified 164 loci conserved in all 407 genomes (100% consensus). Thus, our cgMLST scheme will be faster and computationally less intensive than cgMLST schemes of other bacteria. We have identified 287 cgMLST STs in *Brucella*, and the phylogeny analysis showed region-wide discrimination of *B. abortus* and *B. melitensis* strains. Initially, the MLST tool for *Brucella* was developed with 9 housekeeping genes, which differentiated the *Brucella* spp. into 27 STs (ST-1 to ST-27) (Whatmore et al., 2007). The MLST tool with nine loci was not sufficient to type all the *Brucella* strains. We have identified newer allelic profiles from the *Brucella* WGS data, and the STs from 28 to 71 were incorporated in the BrucellaBase. The BruMLSA21 is an updated version of previous MLST tool with 21 housekeeping genes, which could differentiate the *Brucella* spp. into 100 STs (Whatmore et al., 2016). With decreasing sequencing costs and increased ability for many groups to sequence genomes, the cost of a whole genome won't be much more than the cost of sequencing 21 genes/loci, and there will be far more information than the standard MLST. Therefore, we made an effort to develop a cgMLST for typing Brucella spp. using the whole genome sequences. Our cgMLST tool distinguished the *Brucella* spp. into 287 STs (ST-1 to ST-287). Here, it should be noted that the STs predicted by different methods are not the same. For example, the ST1 might be assigned to different strains of *Brucella* when we use different typing tools. We have established a web-based nomenclature server that can be used to query, compare and analyze sequence data with the existing cgMLST STs predicted from 407 whole genomes. The database will be continuously updated with newer STs, if any, from newly sequenced genomes in the future.

5. Conclusion

We devised a cgMLST scheme from WGS based typing of *Brucella* spp., using BSR-Based Allele Calling Algorithm in chewBBACA. It distinguishes different species of the *Brucella* and the biovars of *B. suis*. Also, *B. abortus* and *B. melitensis* could be differentiated into several groups based on the geographical regions. A web server based cgMLST typing tool is incorporated in the existing genome database, BrucellaBase.

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2018.10.021.

Acknowledgements

The work was financially supported by the Department of Biotechnology, New Delhi, Govt. of India through the DBT-Network Project on Brucellosis. Authors also acknowledge the UGC-CAS, NRCBS, DBT-IPLS, and DST-PURSE Programs of the School of Biological Sciences, Madurai Kamaraj University.

References

Chain, P.S.G., Comerci, D.J., Tolmasky, M.E., Larimer, F.W., Malfatti, S.A., Vergez, L.M., Aguero, F., Land, M.L., Ugalde, R.A., Garcia, E., 2005. Whole-genome analyses of speciation events in pathogenic brucellae. Infect. Immun. 73, 8353–8361. https://doi.org/10.1128/IAI.73.12.8353-8361.2005.

Chawjiraphan, W., Sonthayanon, P., Chanket, P., Benjathummarak, S., Kerdsin, A., Kalambhaheti, T., 2016. Multilocus sequence typing of *Brucella* isolates from Thailand. Southeast Asian J Trop Med Public Health. 47, 1270–1287.

Cody, A.J., Bray, J.E., Jolley, K.A., McCarthy, N.D., Maidena, M.C.J., 2017. Core genome multilocus sequence typing scheme for stable, comparative analyses of *Campylobacter jejuni* and *C. coli* human disease isolates. J. Clin. Microbiol. 55, 2086–2097. https://doi.org/10.1128/JCM.00080-17.

Collins, A., Lonjou, C., Morton, N.E., 1999. Genetic epidemiology of single-nucleotide polymorphisms. Proc. Natl. Acad. Sci. U. S. A. 96, 15173–15177. https://doi.org/10.1073/pnas.96.26.15173.

Ficht, T., 2011. *Brucella* taxonomy and evolution. Future Microbiol 5, 859–866. https://doi.org/10.2217/fmb.10.52.Brucella.

Foster, J.T., Beckstrom-Sternberg, S.M., Pearson, T., Beckstrom-Sternberg, J.S., Chain, P.S.G., Roberto, F.F., Hnath, J., Brettin, T., Keim, P., 2009. Whole-genome-based phylogeny and divergence of the genus *Brucella*. J. Bacteriol. 191, 2864–2870. https://doi.org/10.1128/JB.01581-08.

Georgi, E., Walter, M.C., Pfalzgraf, M.T., Northoff, B.H., Holdt, L.M., Scholz, H.C., Zoeller, L., Zange, S., Antwerpen, M.H., 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. https://doi.org/10.1371/journal.pone.0175425.

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224. https://doi.org/10.1093/molbev/msp259.

Higgins, P.G., Prior, K., Harmsen, D., Seifert, H., 2017. Development and evaluation of a core genome multilocus typing scheme for whole-genome sequence-based typing of *Acinetobacter baumannii*. PLoS One 12, e0179228. https://doi.org/10.1371/journal.pone.0179228.

Hyatt, D., Chen, G.-L., Locascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119. https://doi.org/10.1186/1471-2105-11-119.

Jolley, K.A., Maiden, M.C.J., 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11, 595. https://doi.org/10.1186/1471-2105-11-595.

Kaltungo, B.Y., A Saidu, S.N., Musa, I.W., Baba, A.Y., Lucio Azevedo, J., 2014. Brucellosis: a neglected zoonosis. Br. Microbiol. Res. J. 4, 1551–1574. https://doi.org/10.9734/BMRJ/2014/11061.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.

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

Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242–W245. https://doi.org/10.1093/nar/gkw290.

Ma, J.Y., Wang, H., Zhang, X.F., Xu, L.Q., Hu, G.Y., Jiang, H., Zhao, F., Zhao, H.Y., Piao, D.R., Qin, Y.M., Cui, B.Y., Lin, G.H., 2016. 2016. MLVA and MLST typing of *Brucella* from Qinghai. China Infect. Dis. Poverty 13 (5), 26. https://doi.org/10.1186/s40249-016-0123-z.

Maiden, M.C.J., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., Spratt, B.G., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. 95, 3140–3145. https://doi.org/10.1073/pnas.95.6.3140.

Marzetti, S., Carranza, C., Roncallo, M., Escobar, G.I., Lucero, N.E., 2013. Recent trends in human *Brucella canis* infection. Comp. Immunol. Microbiol. Infect. Dis. 36, 55–61. https://doi.org/10.1016/j.cimid.2012.09.002.

Michaux-Charachon, S., Bourg, G., Jumas-Bilak, E., Guigue-Talet, P., Allardet-Servent, A., O'Callaghan, D., Ramuz, M., 1997. Genome structure and phylogeny in the genus *Brucella*. J. Bacteriol. 179, 3244–3249.

Nymo, I.H., Tryland, M., Godfroid, J., 2011. A review of *Brucella* infection in marine mammals, with special emphasis on *Brucella pinnipedialis* in the hooded seal (*Cystophora cristata*). Vet. Res. 42, 93. https://doi.org/10.1186/1297-9716-42-93.

Pappas, G., Papadimitriou, P., Akritidis, N., Christou, L., Tsianos, E.V., 2006. The new global map of human brucellosis. Lancet Infect. Dis. 6, 91–99. https://doi.org/10.1016/S1473-3099(06)70382-6.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2 - approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490. https://doi.org/10.1371/journal.pone.0009490.

Rice, P., Longden, L., Bleasby, A., 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277. https://doi.org/10.1016/S0168-9525(00)02024-2.

Sankarasubramanian, J., Vishnu, U.S., Gunasekaran, P., Rajendhran, J., 2016a. A genome-wide SNP-based phylogenetic analysis distinguishes different biovars of *Brucella suis*. Infect. Genet. Evol. 41, 213–217. https://doi.org/10.1016/j.meegid.2016.04.012.

Sankarasubramanian, J., Vishnu, U.S., Khader, L.K.M.A., Sridhar, J., Gunasekaran, P., Rajendhran, J., 2016b. BrucellaBase: Genome information resource. Infect. Genet. Evol. 43, 38–42. https://doi.org/10.1016/j.meegid.2016.05.006.

Scholz, H.C., Nockler, K., Gollner, C., Bahn, P., Vergnaud, G., Tomaso, H., Al Dahouk, S., Kampfer, P., Cloeckaert, A., Maquart, M., Zygmunt, M.S., Whatmore, A.M., Pfeffer, M., Huber, B., Busse, H.J., De, B.K., 2010. *Brucella inopinata* sp. nov., isolated from a breast implant infection. Int. J. Syst. Evol. Microbiol. 60, 801–808. https://doi.org/10.1099/ijs.0.011148-0.

Scholz, H.C., Revilla-Fernández, S., Al Dahouk, S., Hammerl, J.A., Zygmunt, M.S., Cloeckaert, A., Koylass, M., Whatmore, A.M., Blom, J., Vergnaud, G., Witte, A., Aistleitner, K., Hofer, E., 2016. *Brucella vulpis* sp. nov., isolated from mandibular lymph nodes of red foxes (*Vulpes vulpes*). Int. J. Syst. Evol. Microbiol. 66, 2090–2098. https://doi.org/10.1099/ijsem.0.000998.

Schork, N.J., Fallin, D., Lanchbury, J.S., 2001. Single nucleotide polymorphisms and the future of genetic epidemiology. Clin. Genet. 58, 250–264. https://doi.org/10.1034/j.1399-0004.2000.580402.x.

Suarez-Esquivel, M., Ruiz-Villalobos, N., Jimenez-Rojas, C., Barquero-Calvo, E., Chacon-Diaz, C., Viquez-Ruiz, E., Rojas-Campos, N., Baker, K.S., Oviedo-Sánchez, G., Amuy, E., Chaves-Olarte, E., Thomson, N.R., Moreno, E., Guzman-Verri, C., 2017. *Brucella neotomae* infection in humans, Costa Rica. Emerg. Infect. Dis. 23, 997–1000. https://doi.org/10.3201/eid2306.162018.

Wang, F., Qiao, Z., Hu, S., Liu, W., Zheng, H., Liu, S., Zhao, X., Bu, Z., 2013. Comparison of genomes of Brucella melitensis M28 and the B. melitensis M5-90 derivative vaccine strain highlights the translation elongation factor Tu gene tuf2 as an attenuation-related gene. Infect. Immun. 81, 2812–2818. https://doi.org/10.1128/IAI.00224-13.

Wattam, A.R., Foster, J.T., Mane, S.P., Beckstrom-Sternberg, S.M., Beckstrom-Sternberg, J.M., Dickerman, A.W., Keim, P., Pearson, T., Shukla, M., Ward, D.V., Williams, K.P., Sobral, B.W., Tsolis, R.M., Whatmore, A.M., O'Callaghan, D., 2014. Comparative phylogenomics and evolution of the *Brucellae* reveal a path to virulence. J. Bacteriol. 196, 920–930. https://doi.org/10.1128/JB.01091-13.

Whatmore, A.M., Perrett, L.L., MacMillan, A.P., 2007. Characterisation of the genetic diversity of *Brucella* by multilocus sequencing. BMC Microbiol. 7, 34. https://doi.org/10.1186/1471-2180-7-34.

Whatmore, A.M., Koylass, M.S., Muchowski, J., Edwards-Smallbone, J., Gopaul, K.K., Perrett, L.L., 2016. Extended multilocus sequence analysis to describe the global population structure of the genus *Brucella*: Phylogeography and relationship to biovars. Front. Microbiol. 7, 2049. https://doi.org/10.3389/fmicb.2016.02049.