Infection, Genetics and Evolution 43 (2016) 38–42

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Short communication

BrucellaBase: Genome information resource

Jagadesan Sankarasubramanian, Udayakumar S. Vishnu, L.K.M. Abdul Khader, Jayavel Sridhar, Paramasamy Gunasekaran, Jeyaprakash Rajendhran *

Department of Genetics, School of Biological Sciences, Madurai Kamaraj University, Madurai – 625021, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 18 June 2015
Received in revised form 15 April 2016
Accepted 5 May 2016
Available online 7 May 2016

Keywords:
BrucellaBase
MLST
Phylogeny
BLAST
VFDB
CARD

ABSTRACT

Brucella sp. causes a major zoonotic disease, brucellosis. *Brucella* belongs to the family Brucellaceae under the order *Rhizobiales* of Alphaproteobacteria. We present BrucellaBase, a web-based platform, providing features of a genome database together with unique analysis tools. We have developed a web version of the multilocus sequence typing (MLST) (Whatmore et al., 2007) and phylogenetic analysis of *Brucella* spp. BrucellaBase currently contains genome data of 510 *Brucella* strains along with the user interfaces for BLAST, VFDB, CARD, pairwise genome alignment and MLST typing. Availability of these tools will enable the researchers interested in *Brucella* to get meaningful information from *Brucella* genome sequences. BrucellaBase will regularly be updated with new genome sequences, new features along with improvements in genome annotations. BrucellaBase is available online at http://www.dbtbrucellosis.in/brucellabase.html or http://59.99.226.203/brucellabase/homepage.html.

1. Introduction

Brucellosis is a zoonotic disease endemic in many parts of the world. Brucellosis is caused by *Brucella*, a Gram-negative, non-motile, facultative intracellular coccobacillus (Corbel, 1997). Chronic *Brucella* infections lead to abortion and sterility in infected animals (Wattam et al., 2009). Currently, ten species of *Brucella*, designated based on the differences in pathogenicity and host specificity, are recognized (Audic et al., 2009). The *Brucella* genome has two circular chromosomes of ~2.1 MB and ~1.2 MB in size and has an overall GC content of 57.2%. Approximately, 3200–3400 ORFs are encoded in a *Brucella* genome.

Presently, 510 whole genome sequences (complete and draft) of *Brucella* spp. are publicly available. We have established a dedicated *Brucella* genomic information resource to store and analyze these genome data. Multilocus sequence typing (MLST) is a robust method of *Brucella* typing using multiple housekeeping genes (Whatmore et al., 2007). No specialized MLST web server is available hitherto for the typing *Brucella* spp. Therefore, we implemented the web-version of the MLST in BrucellaBase, as described by Whatmore et al. (2007). Thus, the major objectives of this resource are to provide access to the genome sequences of *Brucella*, genome typing and phylogeny analysis under one portal.

2. Material and methods

2.1. Collection of genome sequences and annotation

A total of 510 genome sequences of *Brucella* strains that have been reviewed and deposited in public databases were collected. Of these, 360 *Brucella* genome sequences were retrieved from the National Center for Biotechnology Information (NCBI) GenBank database (Benson et al., 2015) and 150 genomes were retrieved from the Pathosystems Resource Integration Center (PATRIC) (Wattam et al., 2014) database (Supplementary file 1, Table S1). Annotations of protein coding genes (PEGs) and genes coding for ribosomal ribonucleic acid (rRNAs) and transfer ribonucleic acid (tRNAs) are stored in BrucellaBase. Also, subcellular localization of each protein was predicted using PSORTb version 3.0.2 (Yu et al., 2010) and listed.

2.2. Prediction of virulence genes in Brucella genomes

Amino acid sequences of all the *Brucella* proteins were downloaded from the BrucellaBase, and homology search was performed against Virulence Factor Database (VFDB) using Basic Local Alignment Search Tool (BLASTp) program (Chen et al., 2012). We used default parameters of E-value threshold 10, percentage identity filter (70%) and coverage identity (70%) for the prediction of virulence genes.

* Corresponding author.
E-mail address: jrajendhran@gmail.com (J. Rajendhran).

http://dx.doi.org/10.1016/j.meegid.2016.05.006
1567-1348/

2.3. Prediction of antibiotic resistance genes in Brucella genomes

Amino acid sequences of all the *Brucella* proteins were searched against Comprehensive Antibiotic Resistance Database (CARD) (McArthur et al., 2013). The proteins showing homology to any antibiotic resistance proteins are stored in the BrucellaBase, which can be searched through CARD BLAST. We used the default parameters of E-value threshold 10, percentage identity filter (70%) and coverage identity (70%) for the prediction of antibiotic resistance genes.

2.4. MLST and phylogeny analysis

A unique web version of MLST was implemented in BrucellaBase for the prediction of allelic profiles of nine housekeeping genes and the sequence types (STs) of *Brucella* spp., using the method proposed by Whatmore et al. (2007). The STs proposed by Whatmore et al. (2007) and the newly defined STs in this study were created as a dataset. We also incorporated an MLST sequence alignment and phylogenetic analysis tool based on the STs using SeaView (version 4) (Gouy et al., 2010). The phylogenetic tree was constructed using the concatenated sequences of 4396 bp from nine loci using neighbor-joining approach.

2.5. Web implementation

We developed BrucellaBase to accommodate the annotated *Brucella* genome sequences and to provide an user-friendly interface capable of accessing *Brucella* genome information. The system design is based on secure web application architecture of client workstation, web server, application server and database server. BrucellaBase was implemented by PHP: Hypertext Preprocessor (PHP), Hyper Text Markup Language (HTML) 5 and Cascading Style Sheets (CSS). An Apache web server is dedicated to handling the requests from web clients and to interact with the back-end servers to serve the requests. Server-side operations are performed on a Linux server (CentOS 6.4) for creating complex pipelines of inputs and outputs for the necessary programs. My Structured Query Language (MySQL) server (5.1.17) was used to construct a relational database to store annotated sequence data.

3. Results and discussion

3.1. Overview of BrucellaBase

BrucellaBase is freely accessible at http://www.dbtbrucellosis.in/brucellabase.html with any modern web browser. Overview of BrucellaBase is shown in Fig. 1. The 'Genomes' feature begins with a summary of genome sequences available for each of 10 *Brucella* species and *Brucella* sp. not classified to species level. All existing complete and draft genomes of each strain along with the information on genome size in bp, GC content, number of contigs, open reading frames (ORFs), tRNAs, rRNAs, and MLST sequence types (STs) are provided. STs provide the allelic profiles of 9 genetic loci and the corresponding sequences of the individual strains. 'ORF Details' icon provided for each strain on the right will lead to a complete list of ORFs of the selected strain. Each nucleotide and protein accession numbers are hyperlinked with their respective sequences. Also, subcellular localization of each annotated protein predicted using PSORTb is provided. Details of annotations and sequences are available for downloading from the 'Download' page as tab-delimited FASTA files.

3.2. Brucella ORF search engine

A search engine is provided under the 'ORF Search' tab in BrucellaBase to find the ORF of interest directly. Users can search the ORF of interest by entering the ORF ID such as BAB1_0001, BMEA_A0001, and BS1330_I0001, or using keywords such as ABC transporter, and transglycosylase, in the search box. Search result for ORF ID provides the details about the strain name, accession no, ORF-type, start, and end position, strand, functional classification, sub-cellular localization, nucleotide and protein sequences. Search by keyword will list all

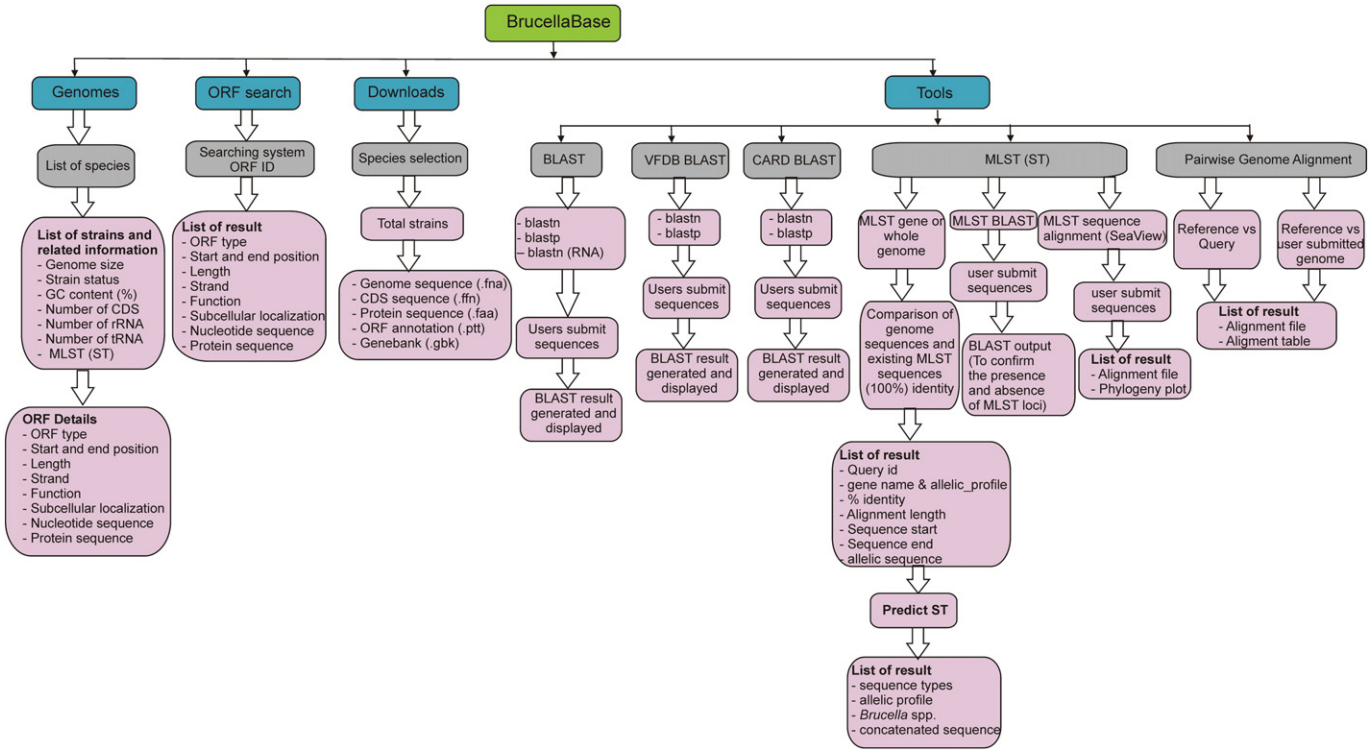

Fig. 1. Overview of BrucellaBase. BrucellaBase has four main sections, i.e., Genomes, ORF Search, Tools, and Download. Under 'Tools', standard BLAST, VFDB BLAST, CARD BLAST, MLST typing, phylogeny construction and pairwise genome alignment utilities are incorporated.