UNIVERSAL WISER
PUBLISHER

Complete Genome Sequence and Evaluation of Expression Level for Genes Associated with Virulence for a Clinical *Brucella Melitensis* Strain C-573 Isolated in Russia

Dmitry A. Kovalev, Sergey V. Pisarenko*, Anna Yu. Evchenko, Dmitry G. Ponomarenko, Olga V. Bobrysheva, Andrei M. Zhirov, Irina V. Kuznetsova, Diana V. Rusanova, Alexander N. Kulichenko
Laboratory of Biochemistry, Stavropol Research Anti-Plague Institute, 355035, Stavropol, Russia
Email: pisarenko_sv@mail.ru

Abstract: Brucellosis is one of the most pressing global zoonotic diseases, which is endemic in many regions of the world. It is believed that *Brucella melitensis* is the most pathogenic species of *Brucella* genus for humans. However, the processes underlying the pathogenicity of this pathogen remain not fully understood. In our study, we report on the first complete genome of the clinical *B. melitensis* strain isolated in Russia, perform structural and functional analysis of the genomic sequence, and evaluate the expression level of genes associated with virulence based on Next Generation Sequencing (NGS) data. The obtained information on the genetic similarities and differences between *B. melitensis* strains can be used to study the mechanisms responsible for the pathogenicity of *Brucella* spp., as well as in the process of developing new therapeutic and preventive strategies for controlling brucellosis.

***Keywords*:** *Brucella melitensis*, complete genome, NGS, gene expression

1. Introduction

Brucellosis is a particularly dangerous zoonotic disease, which in the acute form typically presents with high, undulating fever, while chronic brucellosis may affect many host organs[1]. *Brucella* classification is based on the host preference of the pathogen and its phenotypic and biochemical properties[2-3].

B. melitensis and *B. abortus* cause most serious forms of disease, and therefore their epidemiological significance is quite high[4]. Human infection occurs via alimentary, contact or air-dust mode of transmission when human contacts infected animals or consumes products of animal origin, especially unpasteurized milk products contaminated with *Brucella* spp.[5].

Nowadays brucellosis remains an urgent problem for the whole world with 500,000 cases reported annually. The highest incidence of brucellosis is registered in the countries of South America, Africa, the Middle East and most of Asia[6-7]. In Russia, the epidemic situation remains unstable: while the number of cases of human brucellosis is decreased compared to that of previous years, the epizootic situation in livestock exacerbates. The most endemic regions, mostly due to the high development of agriculture, are North Caucasus, Southern, Siberian and Volga Federal districts. According to the data of the Reference center that monitoring the causative agent of brucellosis (Stavropol Research Anti-Plague Institute), 90% of all cases registered in Russia occur on these territories[8]. At that, most cases of human infection are associated with *B. melitensis*.

The *Brucella* spp. lack classic virulence factors[9]; however, these bacteria are able to adapt to various environmental conditions, and their virulence is determined by their ability to survive and reproduce in the macrophages of the host organism. The key role in the pathogenesis process belongs to the genes involved in survival processes in host macrophages. Although these factors are only indirectly associated with clinical manifestations of brucellosis, they are crucial for surviving and reproducing inside the host cells. Thus, the accumulation of data on the structure and functional features of the genomes of *Brucella* strains is an important step in the process of studying the mechanisms that determine their pathogenicity.

There is very little published information in international databases about the *Brucella* spp. strains isolated on the territory of Russia. The available data on the genomic sequences of Russian strains are presented in the form of fragmented

DOI: https://doi.org/10.37256/amtt.122020626

draft assemblies, while only complete genome assemblies may provide full information on the structure of the genome and its functional features.

In the previous research[10], we were able to establish the phylogenetic relationship of *B. melitensis* strain C-573 with other species of *Brucella* genus. Further research revealed that *B. melitensis* strain C-573 has physiological, morphological and biochemical properties typical for strains isolated in the southern part of European Russia. In the current paper, we provide a complete genomic sequence of a clinical isolate *B. melitensis* C-573, isolated from human in 2014, describe structural and functional features of the genome and evaluate the level of expression of genes associated with virulence. New information on the genetic similarities and differences between *B. melitensis* strains, in addition to studying the processes causing pathogenicity of *Brucella* spp., can be used in the development of antimicrobial agents and vaccines by identifying common conservative targets for all strains of the species.

2. Materials and methods

2.1 Bacterial strain

B. melitensis strain C-573 used in this study was isolated from a person diagnosed with brucellosis in the Stavropol Territory in 2014. The culture was obtained through a direct inoculation of mammary gland punctate to a biphasic medium (Castaneda method). The strain was identified via standard biochemical methods in accordance with "Procedure for organizing and conducting laboratory diagnostics of brucellosis for laboratories at the territorial, regional and federal levels" (MUK 4.2.3010-12). Morphological and biochemical properties of the strain state that this strain belongs to *B. melitensis* biovar II. According to antimicrobial susceptibility testing results, the strain is sensitive to gentamicin, amikacin, kanamycin, streptomycin, doxycycline, tetracycline, nalidixic acid, ciprofloxacin, ofloxacin, pefloxacin, levofloxacin, and rifampicin. Antimicrobial susceptibility testing was performed using disc diffusion method with "Set of discs for antimicrobial susceptibility testing-1" (Institute Pasteur in Saint-Petersburg for Research in Epidemiology and Microbiology of Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing) in accordance with manufacturer's instructions.

Bacterial strain *B. melitensis* C-573 used in this study, was obtained from the State Collection of Pathogenic Microorganisms of Stavropol Research Anti-Plague Institute, where it was stored in a lyophilized state.

2.2 Preparation of nucleic acid samples

Bacteria were cultivated on Brucella Agar at 37°C for 48 hours. The microbial suspension of 2×10^9 m.sub/ml was decontaminated by adding thiomersal to a final concentration of 0.01% and subsequent incubation at 56°C for 30 minutes; after that sterile DNA was isolated. Genomic DNA was isolated from 0.5 ml of decontaminated microbial suspension with the use of PureLink Genomic DNA Kits (Life Technologies, USA). RNA was isolated using 1 ml of microbial suspension with concentration of 2×10^9 mc/ml and TRIzol Max Bacterial RNA Isolation Kit (Life Technologies, USA); manipulations were performed as described in the manufacturer's protocol.

Briefly, after centrifugation, the microbial suspension precipitate was resuspended in 200 µl of pre-warmed TRIzol Reagent MAX to 95°C and incubated at 95°C for 4 min. Then, 1 ml of TRIzol Reagent was added to the suspension and incubated for 5 min at room temperature, after which 200 µl of chloroform was added and incubated for another 2-3 min. After centrifugation for 15 min at 12000 g and 4°C, the aqueous (transparent) phase was collected. RNA was precipitated with 100% isopropanol and 1 ml of 75% ethanol. The finished RNA sample was dissolved in 50 µl of RNase-free water and stored at -80°C.

The concentration of nucleic acids was determined using Qubit dsDNA HS Assay Kit and Qubit PHK HS Assay Kit (Invitrogen, Life Technologies, USA). The purity of the genomic DNA was evaluated with a NanoDrop 2000 spectrophotometer (Thermo Scientific, USA). In addition, the size distribution of RNA fragments was evaluated using a reagent kit for visualizing fragments of the RNA library: Experion RNA Analysis Kit and Experion™ Automated Electrophoresis System. Sample preparation was carried out in the Laboratory of Brucellosis Research of the Stavropol AntiPlague Institute of Rospotrebnadzor.

2.3 Genome and transcriptome sequencing

Preparation of DNA libraries with a read length of 400 bp. was performed with the use of Ion XpressTM Plus Fragment Library Kit (Life Technologies, USA) in accordance with the manufacturer's protocol. DNA library fragments were separated with the use of 2% E-Gel SizeSelect agarose gel (Invitrogen, USA). The finished libraries were purified with the use of Agencourt AMPure XP magnetic particles (Beckman Coulter, USA). Libraries quality and concentration were determined using the Experion™ Automated Electrophoresis System, Experion DNA 1K Reagents and Supplies and

Experion DNA Chips kits (Bio-Rad, USA).

For the RNA library preparation, we have used a sample with RNA Quality Indicator (RQI) > 9. rRNA was removed from the RNA sample with the use of RiboMinus™ Transcriptome Isolation Kit bacteria (Life Technologies, USA) as described in the manufacturer's protocol. Quality and quantity of the RNA after the removal of rRNA was evaluated with the use of Experion™ Automated Electrophoresis System and Experion RNA Analysis Kit. cDNA library was prepared using Ion Total RNA-Seq kit v2 in accordance with the manufacturer's protocol. After evaluating fragment size distribution and quantitative analysis of the concentration of cDNA library, we have performed monoclonal amplification on microspheres.

Monoclonal amplification on microspheres was performed using Ion OneTouch 400 Template Kit (Life Technologies, USA) according to the manufacturer's protocol. Microspheres were enriched with Dynabeads MyOne Streptavidin C1 magnetic particles (Invitrogen, Life Technologies, USA). The efficiency of the enrichment process was evaluated using Ion Sphere Quality Control Kit (Life Technologies, USA). Nucleic acid sequencing was performed using Ion Torrent PGM sequencer and Ion 318 Chips Kit V2 (Life Technologies, USA).

2.4 Post-sequencing data processing

Evaluation of the quality of the obtained data was carried out using the FastQC program version 0.11.3[11]. Reads containing nucleotides with quality of reading, $Q < 15$ were removed in Trimmomatic version 0.33[12]. Reads with an average quality of $Q < 20$, as well as reads less than 75 nucleotides long were removed.

2.5 Genome de novo assembly

Genome *de novo* assembly was performed via Newbler v 3.0 (Roche, Switzerland); the extent of the genome coverage was more than 80×. To finish the complete assembly, we have performed mapping of the contigs obtained during *de novo* assembly on reference genome of *B. melitensis* 16 M (GeneBank: NC_003317.1, NC_003318.1) in Mauve v. 2.4.0 software[13]. After that, primers complementary to the ends of each contig were selected so that during PCR the sections corresponding to gaps in the draft assembly were amplified. The same DNA sample was used as a template for PCR as was used in the preparation of fragment libraries for genome sequencing. The obtained DNA fragments were sequenced by Sanger method using an Applied Biosystems 3500 Genetic Analyzer sequencer (Applied Biosystems, USA). The process of assembly, which consists of aligning and merging of the contigs and DNA fragments obtained during Sanger sequencing, was performed in Contig Express app from Vector NTI Suite software package[14].

2.6 Annotation and analysis of orthologous gene clusters

Genome annotation was performed using NCBI Prokaryotic Genome Annotation Pipeline[15]. An analysis of the distribution of overlapping orthologous gene clusters was performed using OrthoVenn web interface[16].

2.7 Gene expression level analysis

We used HISAT2 software to map the reads to the reference genome[17]. HTSeq-count v. 0.6.1 was used to count readings mapped to coding regions of the genome[18]. Mapping results were normalized to compare gene coverage values by calculating the RPKM (Reads per kilobase per million mapped reads) values for each gene[19].

3. Results and discussion

3.1 Structural features of the genome of the *B. melitensis* C-573 strain

The genome of *B. melitensis* C-573 is a plasmid-free genome with two circular chomosomes of 2,125,194 and 1,185,595 bp. The total length is 3,310,789 bp, and the G + C content of the assembly corresponds to a species-specific value of 57.22%. Of the total 3245 genes, 3168 are protein coding genes, 77 encode RNA and 217 appeared to be pseudogenes. For 2676 (82.47%) genes, functions were predicted and 2427 genes could be assigned to a COG functional category. The genome statistics are provided in Table 1 and Figure 1.

Table 1. Genome statistics

Attribute	Value	% of Total
Genome size (bp)	3310789	100
DNA coding (bp)	2874326	86.82
DNA G + C (bp)	1894445	57.22
DNA scaffolds	2	100
Total genes	3245	97.63
Protein coding genes	3168	91.60
RNA genes	77	2.37
Pseudo genes	217	6.4
Genes in internal clusters	NA	NA
Genes with function prediction	2676	82.47
Genes assigned to COGs	2427	74.79
Genes with Pfam domains	2756	84.93
Genes with signal peptides	273	8.41
Genes with transmembrane helices	799	24.62
CRISPR repeats	NA	NA

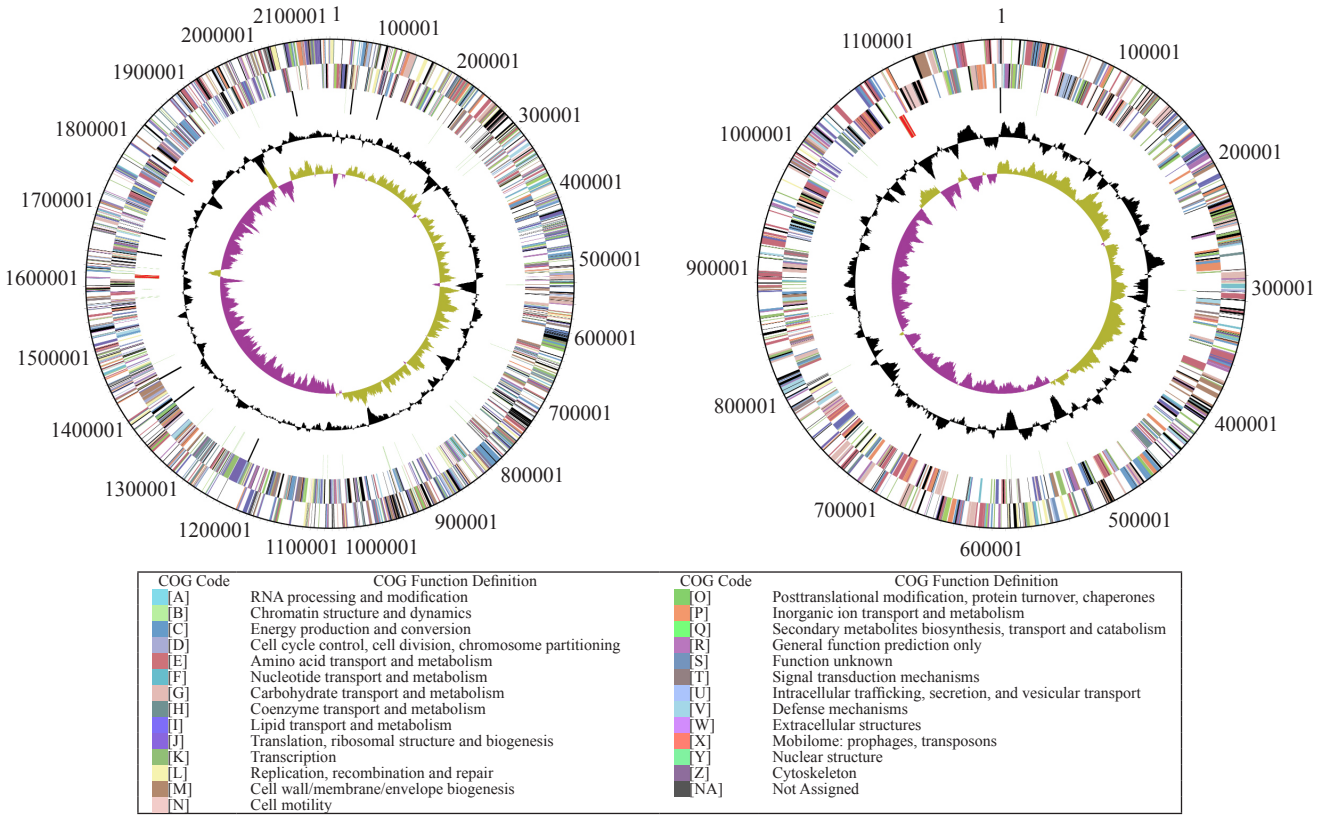

Figure 1. Circular map of two chromosomes
From outside to the center: Genes on forward strand (colored by COG categories), Genes on reverse strand (colored by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew

Functional annotations were predicted for 2676 coding sequences, which is 82.47% of the total number of genes. 2427 coding sequences (74% of the total number of genes) were assigned to the functional categories of COG; their distribution by functional groups is presented in Table 2.

Table 2. Distribution of genes by functional categories of COG

COG Code	COG Function Definition	Count	Percent
[A]	RNA processing and modification	NA	NA
[B]	Chromatin structure and dynamics	1	0.04
[C]	Energy production and conversion	171	6.34
[D]	Cell cycle control, cell division, chromosome partitioning	30	1.11
[E]	Amino acid transport and metabolism	308	11.42
[F]	Nucleotide transport and metabolism	75	2.78
[G]	Carbohydrate transport and metabolism	180	6.68
[H]	Coenzyme transport and metabolism	150	5.56
[I]	Lipid transport and metabolism	129	4.78
[J]	Translation, ribosomal structure and biogenesis	185	6.86
[K]	Transcription	174	6.45
[L]	Replication, recombination and repair	93	3.45
[M]	Cell wall/membrane/envelope biogenesis	175	6.49
[N]	Cell motility	31	1.15
[O]	Posttranslational modification, protein turnover, chaperones	126	4.67
[P]	Inorganic ion transport and metabolism	160	5.93
[Q]	Secondary metabolites biosynthesis, transport and catabolism	70	2.60
[R]	General function prediction only	224	8.31
[S]	Function unknown	184	6.82
[T]	Signal transduction mechanisms	94	3.49
[U]	Intracellular trafficking, secretion, and vesicular transport	39	1.45
[V]	Defense mechanisms	59	2.19
[W]	Extracellular structures	3	0.11
[X]	Mobilome: prophages, transposons	35	1.30
[Y]	Nuclear structure	NA	NA
[Z]	Cytoskeleton	NA	NA
[NA]	Not Assigned	NA	NA

The genomic sequence of the *B. melitensis* C-573 isolate is 99.9% similar to the genome of the reference *B. melitensis* strain 16 M (GeneBank: NC_003317.1, NC_003318.1). When comparing the genomes of *B. melitensis* C-573 and *B. melitensis* 16M strains, we have detected genomic rearrangements in both chromosomes. In the first chromosome, we have found two insertions, while one insertion and one inversion were found on the second.

3952 bp insertion in the first chromosome (555133-559085, the coordinates of the genomic rearrangements are given in accordance with the complete assembly of the genome of the *B. melitensis* C-573 strain) includes the genes of the prokaryotic DNA-binding protein, which belongs to the family of transcriptional regulators of xenobiotic response elements, the amino acid ABC transporter, AzlD protein and transposase IS2020 from the family of transposases IS6. The genes in the structure of this insertion are involved in the processes of branched-chain amino acid transport.

The second insertion (3538 bp, 1481462-1485000) on this chromosome contains the coding sequences of the phage main capsid protein, hypothetical protein, adapter protein, DUF3168 domain-containing protein. The main phage capsid protein is species-specific protein for *B. melitensis*, the remaining listed proteins are found in different species of the *Brucella* genus. The listed genes encode proteins involved in the processes that provide the cell's resistance to bacterial pathogens.

6815 bp insertion on the second chromosome (590980-597795) includes the genes of FAD-binding oxidoreductase, a hypothetical protein, sugar-dehydrogenase, a transcription regulator of the Lrp/AsnC family, amidohydrolase, and 3-oxoadipyl-CoA thiolase. The encoded proteins are involved in the formation of cellular response to exogenous amino acid effectors by regulating lysine metabolism, amino acid transport, transcription activation of oxygen for oxidation and hydroxylation processes.

Inverted region of the second chromosome with a length of 45656 bp (1035284-1080940) includes 42 genes, among them are genes encoding transport proteins, proteins of the dehydrogenase family, transferases, transcriptional regulators, penicillin acylases, hypothetical proteins, the PemK/MazF toxin-antitoxin system gene, and others. Functional annotation of all genes located in the regions of genomic rearrangements are presented in Table 3.

Table 3. Functional annotation of all genes located in the regions of genomic rearrangements

Chromosome	Rearrangement type	Size, b.p.	Coordinates	Locus tag	Protein name
1	insertion	3952	555133-559085	ADS42_RS02860	helix-turn-helix domain-containing protein
				ADS42_RS02865	branched-chain amino acid ABC transporter permease
				ADS42_RS02870	AzlD domain-containing protein
				ADS42_RS02875	IS6 family transposase IS2020
	insertion	3538	1481462-1485000	ADS42_RS07650	phage major capsid protein
				ADS42_RS07655	hypothetical protein
				ADS42_RS07660	head-tail adaptor protein
				ADS42_RS07665	DUF3168 domain-containing protein
	insertion	6815	590980-597795	ADS42_RS13980	FAD-binding oxidoreductase
				ADS42_RS13985	amino acid ABC transporter
				ADS42_RS13990	saccharopine dehydrogenase family protein
				ADS42_RS13995	Lrp/AsnC family transcriptional regulator
				ADS42_RS14000	amidohydrolase
				ADS42_RS14005	3-oxoadipyl-CoA thiolase
				ADS42_RS16225	lactate dehydrogenase
				ADS42_RS16230	homoserine O-succinyltransferase
				ADS42_RS16235	type II toxin-antitoxin system PemK/MazF family toxin
				ADS42_RS16240	hypothetical protein
				ADS42_RS16245	four-carbon acid sugar kinase family protein
				ADS42_RS16250	class II aldolase/adducin family protein
				ADS42_RS16255	hypothetical protein
				ADS42_RS16260	hypothetical protein
				ADS42_RS16265	aspartate aminotransferase family protein
				ADS42_RS16270	ABC transporter ATP-binding protein
				ADS42_RS16275	ABC transporter permease
				ADS42_RS16280	ABC transporter permease
				ADS42_RS16285	extracellular solute-binding protein
				ADS42_RS16290	amidase
				ADS42_RS16295	SMP-30/gluconolactonase/LRE family protein
				ADS42_RS16300	ABC transporter ATP-binding protein
				ADS42_RS16305	ABC transporter ATP-binding protein
				ADS42_RS16315	ABC transporter permease
2				ADS42_RS16320	ABC transporter substrate-binding protein
				ADS42_RS16325	GntR family transcriptional regulator
				ADS42_RS16330	hypothetical protein
	inversion	45656	1035284-1080940	ADS42_RS16335	ABC transporter ATP-binding protein
				ADS42_RS16340	ABC transporter ATP-binding protein
				ADS42_RS16345	ABC transporter permease
				ADS42_RS16350	peptide ABC transporter
				ADS42_RS16355	ABC transporter substrate-binding protein
				ADS42_RS16360	penicillin acylase
				ADS42_RS16365	penicillin acylase
				ADS42_RS16370	acyl-CoA dehydrogenase
				ADS42_RS16375	enoyl-CoA hydratase
				ADS42_RS16380	3-hydroxyacyl-CoA dehydrogenase
				ADS42_RS16385	MFS transporter
				ADS42_RS16390	ABC transporter substrate-binding protein
				ADS42_RS16395	acetoin dehydrogenase dihydrolipoyllysine-residue acetyltransferase subunit
				ADS42_RS16400	IclR family transcriptional regulator
				ADS42_RS16405	ABC transporter permease
				ADS42_RS16410	ABC transporter permease
				ADS42_RS16415	ABC transporter ATP-binding protein
				ADS42_RS16420	dipeptide ABC transporter ATP-binding protein
				ADS42_RS16425	CoA transferase
				ADS42_RS16430	NAD-dependent succinate-semialdehyde dehydrogenase
				ADS42_RS16435	GntR family transcriptional regulator

3.2 Functional features of the genome of *B. melitensis* strain C-573

To describe the functional features of the genome of the *B. melitensis* C-573 strain, we have carried out a comparative analysis of the genes of four strains: *B. melitensis* 16 M (GeneBank ID: GCA_000007125.1), *B. melitensis* Ether (GeneBank ID: GCA_000740355.1), *B. melitensis* ATCC 23457 (GeneBank ID: GCA_000022625.1) and *B. melitensis* C-573. The strains selected for comparison belong to different genotypes established during the phylogenetic analysis performed earlier[10]. *B. melitensis* strain Ether belongs to the first genotype, *B. melitensis* strain ATCC 23457 to the second, and *B. melitensis* strain 16 M to the fifth genotype. All genomes selected from the international GenBank database for analysis were presented as complete assemblies. We did not use fragmented genome assemblies for the analysis as using them could have caused erroneous results. At the time of the analysis, complete assemblies for third and fourth genotype did not exist, so we have excluded these genotypes from the analysis. Results are presented in Figure 2.

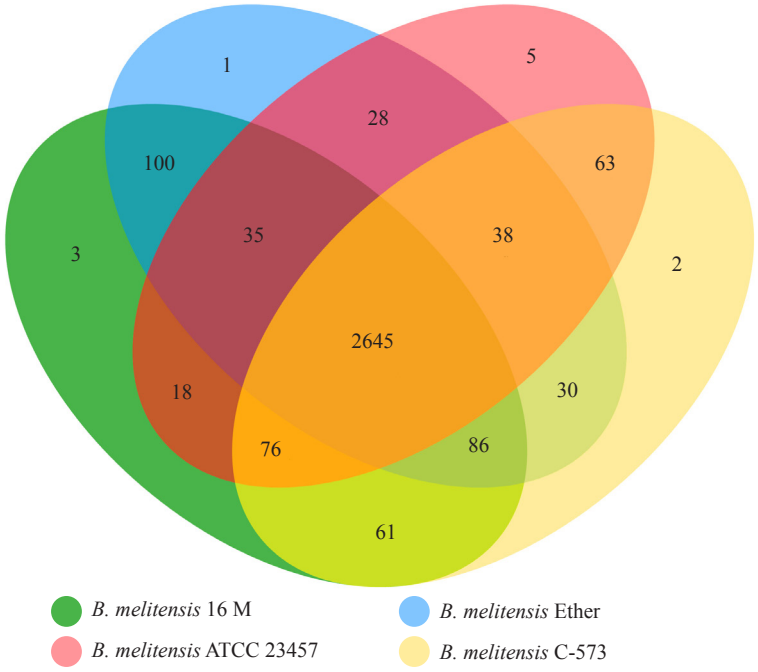

Figure 2. Distribution of overlapping orthologous gene clusters

2645 gene clusters are common to all four strains, 86 gene clusters are common to *B. melitensis* 16 M, *B. melitensis* Ether and *B. melitensis* C-573 strains, 76 clusters are common to *B. melitensis* 16 M, *B. melitensis* ATCC 23457 and *B. melitensis* C-573, another 38 gene clusters are common for *B. melitensis* ATCC 23457, *B. melitensis* Ether and *B. melitensis* C-573 strains. In the pairwise comparison of common gene clusters, we have detected the smallest number of common clusters in *B. melitensis* C-573 and *B. melitensis* Ether strains-30 clusters. Comparison of genes of *B. melitensis* C-573 and *B. melitensis* 16 M strains revealed 61 common gene clusters. The largest number of total gene clusters (63) was found when comparing genes in strains of *B. melitensis* C-573 and *B. melitensis* ATCC 23457. For *B. melitensis* C-573 strain, we have found two unique clusters containing two genes each. The first cluster contains genes encoding hypothetical proteins, second- genes of signal peptidase. All genes included in the unique clusters are located on the first chromosome of the genome of *B. melitensis* strain C-573.

3.3 Assessment of gene expression level

To assess the gene expression level, we have used the method of whole transcriptome sequencing. The reads obtained as a result of the sequencing were filtered and mapped to the genome. Assessment of gene expression level was performed as described in Materials and Methods section. In order to normalize the results of the mapping, we have evaluated RPKM (*Reads per kilobase per million mapped reads*) value for each gene. Average RPKM value was 117.9; maximum and minimum were 266.6 and 0, respectively. Results of the gene expression level assessment are presented in Table S2 «Gene expression level».

100 genes with maximum gene expression level were assigned to functional COG categories, among them for, 37% of genes function is unknown; 9% belong to amino acid transport and metabolism cluster; 9% take part in the process of

replication; 6% are in transcription processes; 4% are genes encoding proteins of carbohydrate transport and metabolism; 3% encode proteins with general function. The least represented were the functional categories of genes involved in the production and conversion of energy, lipid metabolism, translation, ion transport and metabolism-2% of genes for each category. The number of genes belonging to the functional categories of cell cycle control, coenzyme metabolism, cell wall biogenesis, post-translational modification and the secondary structure was 1% for each of these categories. For 15% of the genes, the functional category was not assigned. Detailed information on genes with the highest expression level is given in Table 4.

Table 4. Genes with the highest expression level

Locus tag	COG category	RPKM	Protein name
BMEI1814	L	266.6	transposase
BMEI1815	L	238.7	transposase
BMEI1897	S	230.3	hypothetical protein
BMEI1656	S	217.7	hypothetical protein
BMEI0848	S	216.5	probable carnitine operon oxidoreductase caia
BMEI1658	S	215.9	hypothetical protein
BMEI1405	L	215.1	transposase
BMEI0142	S	214.0	hypothetical membrane spanning protein
BMEI1000	S	213.5	hypothetical protein
BMEI0141	C	212.2	dihydrolipoamide succinyltransferase component (e2) of 2-oxoglutarate dehydrogenase complex
BMEI1413	M	211.7	gdp-mannose 4,6-dehydratase
BMEI0570	S	211.5	hypothetical protein
BMEI0059	S	211.1	hypothetical protein
BMEI1896	S	210.7	hypothetical membrane spanning protein
BMEI1386	G	210.6	xylulose kinase
BMEI0585	D	210.5	cell division protein ftsz
BMEI1738	K	210.0	transcriptional activator tena
BMEI0900	S	209.9	hypothetical protein
BMEII1042	S	209.9	hypothetical protein
BMEI1687	not in COG	209.9	hypothetical protein
BMEI1402	L	209.8	transposase
BMEII0912	S	209.6	hypothetical protein
BMEI0058	S	208.2	extracellular serine protease
BMEI0541	S	207.6	hypothetical protein
BMEI1054	J	207.2	bacterial peptide chain release factor 2 (rf-2)
BMEII0182	S	207.0	hypothetical protein
BMEII0470	not in COG	206.9	integral membrane protein
BMEI0994	S	206.4	hypothetical protein
BMEI1657	S	206.0	hypothetical protein
BMEI1686	S	205.9	hypothetical protein
BMEI1680	S	205.7	hypothetical protein
BMEI0389	S	204.9	hypothetical protein
BMEI1664	L	204.6	hypothetical cytosolic protein
BMEI1306	S	204.4	porin
BMEII0044	not in COG	204.3	hypothetical protein
BMEI0847	not in COG	204.0	hypothetical protein
BMEII0559	E	203.7	aminomethyltransferase
BMEI0451	not in COG	203.6	2-isopropylmalate synthase
BMEI0999	S	203.5	hypothetical protein
BMEI0133	E	203.4	diaminopimelate epimerase
BMEI1663	K	203.4	antirepressor protein ant
BMEI0060	S	203.3	hypothetical protein
BMEII1070	S	203.0	adhesin aida-i
BMEI0778	F	202.9	adenylate kinase
BMEI1678	S	202.8	hypothetical protein
BMEI1881	S	202.4	hypothetical protein

BMEI0496	R	201.8	hydroxyacylglutathione hydrolase
BMEI1542	L	201.7	ribonuclease hii
BMEII0878	K	201.7	transcriptional regulator, gntr family
BMEI1397	L	201.1	transposase
BMEI0827	I	201.0	undecaprenyl pyrophosphate synthetase
BMEII0560	E	200.8	glycine cleavage system h protein
BMEII0919	S	200.5	hypothetical protein
BMEI1700	K	200.5	hypothetical protein
BMEI0993	S	200.4	hypothetical protein
BMEI1699	not in COG	200.1	hypothetical protein
BMEI0198	not in COG	199.9	lysostaphin
BMEII0911	E	199.8	glutamate decarboxylase alpha
BMEII0483	G	199.5	ABC transporter integral membrane protein
BMEII0247	R	199.3	exoenzymes regulatory protein aepa precursor
BMEI1696	not in COG	199.3	hypothetical membrane spanning protein
BMEI1416	G, M	199.3	o-antigen export system ATP-binding protein rfbb
BMEII0482	S	199.2	hypothetical protein
BMEI1684	S	199.1	hypothetical protein
BMEII0973	C	199.0	nitrous-oxide reductase
BMEI1674	S	199.0	hypothetical protein
BMEII1133	E	198.9	ornithine decarboxylase
BMEI1486	O	198.8	DNA repair protein rada
BMEII1046	S	198.7	hypothetical protein
BMEII0853	T, K	198.5	two component response regulator
BMEI0434	E	198.4	periplasmic dipeptide transport protein precursor
BMEI0081	H	198.4	methyltransferase
BMEII0902	K	198.1	transcriptional regulatory protein, lysr family
BMEI1916	F	198.1	cytidylate kinase
BMEI1813	S	198.0	hypothetical protein
BMEI0052	S	198.0	hypothetical protein
BMEII0706	R	197.9	hypothetical cytosolic protein
BMEII0571	E	197.9	acetolactate synthase IolD
BMEI0022	I	197.8	3-hydroxybutyryl-CoA dehydratase
BMEII0480	not in COG	197.7	hypothetical protein
BMEI1775	J	197.3	ribonuclease ph
BMEI2024	L	197.2	hypothetical cytosolic protein
BMEI0546	Q	197.2	pyrazinamidase/nicotinamidase
BMEI1690	not in COG	197.1	hypothetical protein
BMEII0100	E	197.1	high-affinity branched-chain amino acid transport ATP-binding protein livg
BMEI1307	L	196.9	integrase
BMEII0439	P	196.7	transporter, dme family
BMEII0478	G	196.6	d-mannonate oxidoreductase
BMEI0375	P	196.5	ferric uptake regulation protein
BMEI0907	S	196.5	hypothetical protein
BMEII0892	not in COG	196.0	hypothetical protein
BMEI0263	E	195.8	hypothetical protein
BMEI0539	not in COG	195.6	hypothetical protein
BMEI0996	not in COG	195.6	hypothetical protein
BMEII1092	G	195.2	hydroxypyruvate isomerase
BMEI0633	not in COG	195.2	integral membrane protein
BMEI1880	S	195.0	hypothetical membrane spanning protein
BMEII0785	K	195.0	transcriptional regulator, marr family
BMEI1220	not in COG	195.0	hypothetical protein
BMEI0502	S	194.7	hypothetical membrane spanning protein

One-letter abbreviations for the functional categories: C-energy production and conversion; D-cell division and chromosome partitioning; E-amino acid metabolism and transport; F-nucleotide metabolism and transport; G-carbohydrate metabolism and transport; H-coenzyme metabolism; I-lipid metabolism; J-translation, including ribosome structure and biogenesis; K-transcription; L-replication, recombination and repair; M-cell wall structure and biogenesis and outer membrane; O-molecular chaperones and related functions; P-inorganic ion transport and metabolism; Q-Secondary metabolites biosynthesis, transport, and catabolism; S-no functional prediction. R-general functional prediction only; T-signal transduction.

It is a known fact that the *Brucella* spp. do not have common bacterial virulence factors, such as cytolysins, exotoxins, secreted proteases, fimbriae, phage-encoded toxins, and virulence plasmids[9]. *Brucella* virulence is determined by their ability to survive and reproduce in the macrophages of the host organism. Nowadays factors that are crucial for ensuring the intracellular processes of *Brucella* cells and affecting their survival inside macrophages have been identified; these include lipopolysaccharide (LPS), brucebactin which is involved in the iron uptake, cyclic *β*-1,2-glucans (C*β*G), type IV secretion system and a two-component BvrR/BvrS regulation system[20-25]. Although these virulence factors are only indirectly associated with the clinical manifestations of brucellosis, they are crucial for the survival and reproduction of the *Brucella* cells within host cells.

We have performed gene expression level analysis for genes associated with virulence via comparing median values for each gene group. The results of evaluating the level of expression of *B. melitensis* C-573 genes associated with virulence are shown in Table 5.

Table 5. The expression level of *B. melitensis* C-573 genes associated with virulence

Gene	Symbol	RPKM	Virulence factor	Protein
BMEII0076	vibH/entF	178	Brucebactin	enterobactin synthetase component F
BMEII0077	dhbC	180.4	Brucebactin	isochorismate synthase
BMEII0078	dhbE	48.1	Brucebactin	2,3-dihydroxybenzoate-amp ligase
BMEII0079	dhbB	34.9	Brucebactin	isochorismatase
BMEII0080	dhbA	42.5	Brucebactin	2,3-dihydroxybenzoate-2,3-dehydrogenase
BMEII0081	entD	44.2	Brucebactin	enterobactin syntetase component D
BMEI2035	bvrS	18.7	BvrR/BvrS	sensor protein BvrS
BMEI2036	bvrR	4.7	BvrR/BvrS	transcriptional regulatory protein BvrR
BMEI1837	cgs	159.2	CβG	cyclic beta 1-2 glucan synthetase
BMEI0509	lpsB/lpcC	129.6	LPS	lipopolysaccharide core biosynthesis mannosyltransferase LpcC
BMEI0586	lpxC	120.5	LPS	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase
BMEI0831	lpxD	137.5	LPS	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
BMEI0832	fabZ	130	LPS	(3R)-hydroxymyristoyl ACP dehydratase
BMEI0833	lpxA	126	LPS	UDP-N-acetylglucosamine acyltransferase
BMEI0835	lpxB	125.6	LPS	lipid-A-disaccharide synthase
BMEI0850	kdsA	98.4	LPS	2-dehydro-3-deoxyphosphooctonate aldolase
BMEI0997	wbdA	179.2	LPS	mannosyltransferase
BMEI0998	wboA	176.7	LPS	glycosyltransferase
BMEI1111	acpXL	174.1	LPS	acyl carrier protein
BMEI1115	htrB	78.2	LPS	lipid A biosynthesis lauroyl acyltransferase
BMEI1212	lpxE	131.9	LPS	phosphatidylglycerophosphatase B
BMEI1326	lpsA	144.8	LPS	glycosyltransferase
BMEI1393	wbpZ	180.4	LPS	mannosyltransferase C
BMEI1394	manAoAg	157.3	LPS	mannose-6-phosphate isomerase
BMEI1395	manCoAg	147.6	LPS	mannose-1-phosphate guanylyltransferase
BMEI1396	pmm	183.8	LPS	phosphomannomutase
BMEI1404	wbkA	193.5	LPS	mannosyltransferase
BMEI1413	gmd	211.7	LPS	GDP-mannose 4,6-dehydratase
BMEI1414	per	181.3	LPS	perosamine synthetase
BMEI1415	wzm	189.3	LPS	O-antigen export system permease protein
BMEI1416	wzt	199.3	LPS	O-antigen export system ATP-binding protein
BMEI1417	wbkB	185.8	LPS	perosamine synthetase WbkB
BMEI1418	wbkC	170	LPS	GDP-mannose 4,6-dehydratase/GDP-4-amino-4,6-dideoxy-D-mannose formyltransferase
BMEI1426	wbpL	186.1	LPS	undecaprenyl-phosphate alpha-N-acetylglucosaminyltransferase
BMEI1475	acpXL	164.2	LPS	acyl carrier protein
BMEI1886	pgm	171.9	LPS	phosphoglucomutase

BMEI1904	kdsB	169.5	LPS	3-deoxy-manno-octulosonate cytidylyltransferase
BMEII0899	manBcore	146.9	LPS	phosphomannomutase
BMEII0900	manCcore	157.1	LPS	mannose-6-phosphate isomerase/mannose-1-phosphate guanylyltransferase (GDP)
BMEII1028	lpxK	37.9	LPS	tetraacyldisaccharide 4'-kinase
BMEII1029	waaA/kdtA	16.6	LPS	3-deoxy-D-manno-octulosonic-acid transferase
BMEII0025	virB1	164.4	T4SS	lytic transglycosylase
BMEII0026	virB2	183.3	T4SS	Pilus
BMEII0027	virB3	156.2	T4SS	channel protein
BMEII0028	virB4	173.4	T4SS	ATPase
BMEII0029	virB5	177.5	T4SS	Attachment mediating protein
BMEII0030	virB6	182.9	T4SS	IM Channel protein
BMEII0031	virB7	185.4	T4SS	channel protein
BMEII0032	virB8	189.5	T4SS	channel protein
BMEII0033	virB9	190.7	T4SS	channel protein
BMEII0034	virB10	171.6	T4SS	channel protein
BMEII0035	virB11	167.8	T4SS	ATPase
BMEII0036	virB12	183.7	T4SS	outer membrane protein

For genes involved in the iron uptake RPKM value is 46, for CβG gene-159.2; for genes of BvrR/BvrS regulation system-11.35; for T4SS genes-179.5; for LPS group-170. Low level of gene expression for BvrR/BvrS regulation system and iron uptake genes may be explained by the fact that nutrient medium provided comfortable conditions for the growth of the *B. melitensis* C-573 culture, such as the absence of stress factors at low pH value, presence of antibiotics, low iron content. The expression level of genes associated with virulence factors T4SS and LPS significantly exceeds average RPKM values estimated for all other *B. melitensis* C-573 genes, which indicates that the protein products of these genes play a great role not only in successful survival in the host organism, but also in survival on a nutrient medium.

4. Conclusion

The results of our study expand the current understanding of the structural and functional organization of the *B. melitensis* genome. This is the first analysis of the completed genome of a clinical *B. melitensis* strain isolated in Russia. The created genomic project is a valuable reference material for future studies of the virulence of the pathogen. The genomic data we obtained in combination with the genetic data of other virulent and vaccine strains can make a significant contribution to the study of the pathogenicity of *Brucella* and the development of effective therapeutic and preventative means for controlling brucellosis.

Funding

No funding was received for this research.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgments

Not applicable

This article contains supplementary information online at:
http://ojs.wiserpub.com/index.php/AMTT/article/view/626/407

References

[1] P. de Figueiredo, T. A. Ficht, A. Rice-Ficht, C. A. Rossetti, L. G. Adams. Pathogenesis and immunobiology of brucellosis: Review of brucellae host interactions. *The American Journal of Pathology*. 2015; 185(6): 1505-1517.

[2] A. M. Whatmore, N. Davison, A. Cloeckaert, et al. *Brucella papionis* sp. nov., isolated from baboons (*Papio* spp.). *International Journal of Systematic and Evolutionary Microbiology*. 2014; 64: 4120-4128.
[3] H. C. Scholz, S. Revilla-Fernández, S. Al Dahouk, et al. *Brucella vulpis* sp. nov., isolated from mandibular lymph nodes of red foxes (*Vulpes vulpes*). *International Journal of Systematic and Evolutionary Microbiology*. 2016; 66: 2090-2098.
[4] E. M. Galinska, J. Zagorski. Brucellosis in humans-etiology, diagnostics, clinical forms. *Annals of Agricultural and Environmental Medicine*. 2013; 20: 233-238.
[5] B. Mantur, S. Amarnath, R. Shinde. Review of clinical and laboratory features of human Brucellosis. *Indian Journal of Medical Microbiology*. 2007; 25: 188-202.
[6] G. Pappas, P. Papadimitriou, N. Akritidis, L. Christou, E. V. Tsianos. The new global map of human brucellosis. *The Lancet Infectious Diseases*. 2006; 6: 91-99.
[7] K. A. Franc, R. C. Krecek, B. N. Häsler, A. M. Arenas-Gamboa. Brucellosis remains a neglected disease in the developing world: a call for interdisciplinary action. *BMC Public Health*. 2018; 18: 125.
[8] D. G. Ponomarenko, D. V. Rusanova, T. V. Berdnikova, A. A. Khachaturova, E. A. Manin, A. N. Kulichenko. Overview of epizootological and epidemiological situation on Brucellosis in the Russian federation in 2017 and prognosis for 2018. *Problemy Osobo Opasnykh Infektsii*. 2018; 2: 23-29.
[9] V. G. DelVecchio, V. Kapatral, R. J. Redkar, et al. The genome sequence of the facultative intracellular pathogen *Brucella melitensis*. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99: 443-448.
[10] S. V. Pisarenko, D. A. Kovalev, A. S. Volynkina, et al. Global evolution and phylogeography of *Brucella melitensis* strains. *BMC Genomics*. 2018; 19(1): 353.
[11] S. Andrews. FastQC: a quality control tool for high throughput sequence data. 2010. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
[12] A. M. Bolger, M. Lohse, B. Usadel. Trimmomatic: A flexible trimmer for Illumina sequence data. *BioInformatics*. 2014; 30: 2114-2120.
[13] A. C. Darling, B. Mau, F. R. Blattner, N. T. Perna. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*. 2004; 14(7): 1394-1403.
[14] G. Lu, E. N. Moriyama. Vector NTI, a balanced all-in-one sequence analysis suite. *Briefings in bioinformatics*. 2004; 5(4): 378-88.
[15] National Center for Biotechnology Information Search Database. Available from: https://www.ncbi.nlm.nih.gov/ [Accessed 20th November 2020].
[16] Y. Wang, D. Coleman-Derr, G. Chen, Y. Q. Gu. OrthoVenn: A web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research*. 2015; 43(W1):W78-84.
[17] D. Kim, B. Langmead, S. L. Salzberg, HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*. 2015; 12(4): 357-60.
[18] S. Anders, P. T. Pyl, W. Huber HTSeq-A python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015; 31(2): 166-169.
[19] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*. 2008; 5(7): 621-628.
[20] Z. Xiang, W. Zheng, Y. He. BBP: *Brucella* genome annotation with literature mining and curation. *BMC Bioinformatics*. 2006; 7: 347.
[21] N. Lapaque, I. Moriyon, E. Moreno, J. P. Gorvel. *Brucella* lipopolysaccharide acts as a virulence factor. *Current Opinion in Microbiology*. 2005; 8: 60-66.
[22] L. Eskra, J. Covert, J. Glasner, G. Splitter. Differential expression of iron acquisition fenes by *Brucella melitensis* and *Brucella canis* during macrophage infection. *PLoS One*. 2012; 7(3): e31747.
[23] B. Arellano-Reynoso, N. Lapaque, S. Salcedo, et al. Cyclic *β*-1,2-glucan is a *Brucella* virulence factor required for intracellular survival. *Nature Immunology*. 2005; 6: 618-625.
[24] D. O'Callaghan, C. Cazevieille, A. Allardet-Servent, et al. A homologue of the *Agrobacterium tumefaciens* VirB and *Bordetella pertussis* Ptl type IV secretion systems is essential for intracellular survival of *Brucella suis*. *Molecular Microbiology*. 1999; 33: 1210-1220.
[25] C. Guzman-Verri, L. Manterola, A. Sola-Landa, et al. The two-component system BvrR/BvrS essential for *Brucella abortus* virulence regulates the expression of outer membrane proteins with counterparts in members of the *Rhizobiaceae*. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99: 12375-12380.