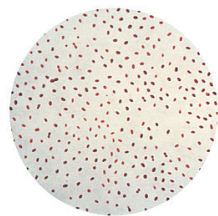


## Introduction



***Brucella*** is a genus of Gram-negative bacteria, named after David Bruce (1855–1931). They are small (0.5 to 0.7 by 0.6 to 1.5  $\mu\text{m}$ ), nonencapsulated, nonmotile, facultatively intracellular coccobacilli.

*Brucella* is the cause of brucellosis, which is a zoonosis transmitted by ingesting contaminated food (such as unpasteurized milk products), direct contact with an infected animal, or inhalation of aerosols. Transmission from human to human, for example through sexual intercourse or from mother to child, is exceedingly rare, but possible. Minimum infectious exposure is between 10 and 100 organisms.

The different species of *Brucella* are genetically very similar, although each has a slightly different host specificity. Hence, the NCBI taxonomy includes most *Brucella* species under *B. melitensis*.

The many names of brucellosis include (human disease/animal disease):

- Malta fever/Bang's disease
- Undulant fever/enzootic abortion
- Mediterranean fever/epizootic abortion
- Rock fever of Gibraltar/slinking of calves
- Gastric fever/ram epididymitis
- Contagious abortion/spontaneous abortion

*Brucella neotomae* BNWLG2-S2 was isolated from the urine of a 49 year-old female who presented to a clinic in Pennsylvania on April 26th 2016 with symptoms indicative of a UTI. It harbors mcr-1 and blaCTX-M on a Novel IncF Plasmid, First report of mcr-1 in the USA.

## Genome Assembly

*Brucella neotomae* BNWLG2-S2 ran under job ID 2fee1c1b-f8bb-4781-bb20-f08fa66398e1 at PATRIC (1). The assembly job started at 10/6/17, 2:53 PM and completed at 10/7/17, 4:41 PM, after 25h 47m 90011s. The auto assembly strategy was selected, and it runs BayesHammer [2] on short reads, followed by three assembly strategies that include Velvet [3], IDBA [4] and Spades [5], each of which is given an assembly score by ARAST, an in-house script. The minimum contig length was 120bp, and smaller contigs were not included in the assembly. The minimum contig coverage was 5, and contigs with less coverage were not included in the assembly. Also add reference to QUAST. ARAST ranked the Spades assembly best (Table 1).

The assembled genome has 33 contigs, with the total length of 3.31 Mbp and %GC of 57.26%.

Table 1. Assembly details for *Brucella neotomae* 5K33

Contigs	11	GC Content	57.3
Plasmids	0	Contig L50	1
Genome Length	3329623	Contig N50	1923503
Chromosomes	0		

## Genome Annotation

The Genome Annotation Service in PATRIC [1] uses the RAST tool kit (RASTtk) [6] to provide annotation of genomic features. The job for *B. neotomae* BNWLG2-S2 ran under job number 685fb8c1-348a-4cea-a298-496b037af1f8. The annotation job started at 10/9/17, 7:10 AM and completed at 10/9/17, 8:16 AM, after 1h 5m 3623s. The selected domain was Bacteria, the Taxonomy ID was 234.128. The Genetic code was 11. The taxonomy of *B. neotomae* BNWLG2-S2 is:

Bacteria >> Proteobacteria >> Alphaproteobacteria >> Rhizobiales >> Brucellaceae >> *Brucella* >> *Brucella neotomae* >> *Brucella neotomae* 5K33

The annotations includes 3349 CDS, 50 tRNAs and 3 rRNAs. All annotated genome features for this genome are summarized in Table 3.

The functional annotation included 655 hypothetical proteins and 2694 proteins with functional assignments. Furthermore, 963 proteins were assigned EC numbers and 749 proteins were mapped to KEGG pathways. A breakdown of the proteins that have been annotated for this isolate is provided in Table 4.

Table 3. Annotated Genome Features

	PATRIC
CDS	3434
tRNA	48
pseudogene	23
rRNA	3
misc_RNA	1

## Genome Quality

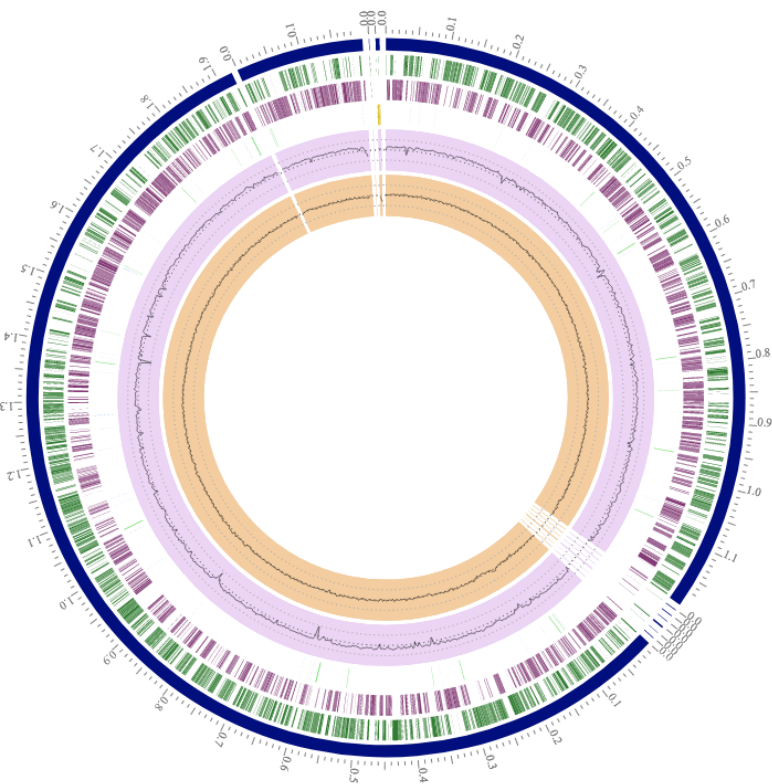
Based on the analysis of genome assembly and annotation statistics and comparing them to other closely related genomes available in PATRIC for the same species, the overall genome seems to be of [high/medium/low] quality. The reasons affecting the quality of genome are: [genome qc flags].

## Specialty Genes

PATRIC [1] annotation also includes identification of proteins that have homology to known virulence factors, drug targets, and antibiotic resistance genes (Reference the sources). A breakdown of the distribution is provided (Table 5).

Table 3. Annotated Genome Features		
	Source	Genes
Essential Gene	PATRIC	224
Virulence Factor	Victors	224
Human Homolog	Human	79
Virulence Factor	VFDB	44
Transporter	TCDB	25
Antibiotic Resistance	PATRIC	11
Drug Target	DrugBank	4
Antibiotic Resistance	CARD	3
Drug Target	TTD	1
Virulence Factor	PATRIC_VF	1

A graphical display of the distribution of the CDS on each strand, the repeat\_regions, GC content and GC skew are provided (Figure 1).

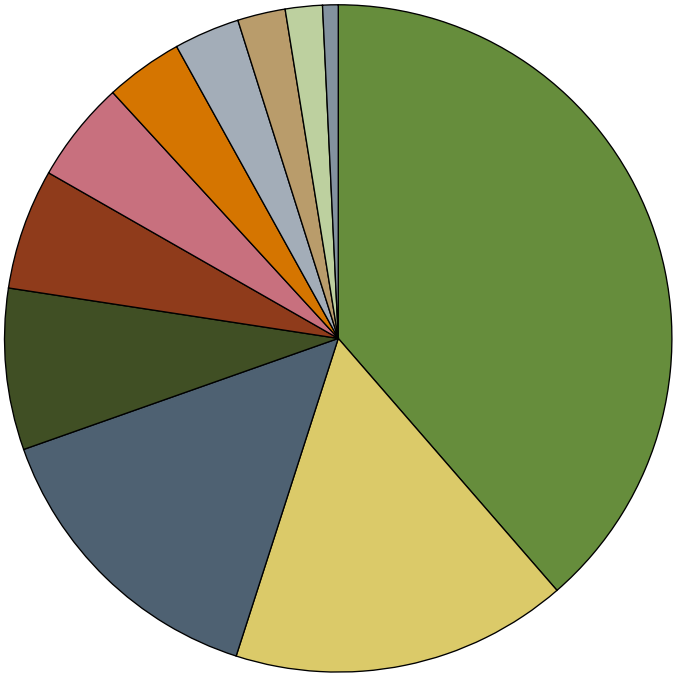


## Subsystem Analysis

A subsystem is a set of protein functions that are related. Frequently, subsystems represent the collection of functional roles that make up a metabolic pathway, a complex (e.g., the ribosome), or a class of proteins (e.g., two-component signal-transduction proteins or AMR genes) [Overbeek]. An overview of the subsystems for this genome is provided in Figure 2.

Subsystem Category Distribution - Brucella neotomae 5K33

Subsystem Coverage



Subsystem Counts (Subsystems, Genes)

- Metabolism (84, 667)
- Energy (29, 282)
- Protein Processing (45, 253)
- Membrane Transport (18, 135)
- Cellular Processes (13, 101)
- Stress Response, Defense, Virulence (18, 135)
- DNA Processing (14, 65)
- RNA Processing (14, 55)
- Cell Envelope (7, 40)
- Miscellaneous (6, 31)
- Regulation And Cell Signaling (3, 13)