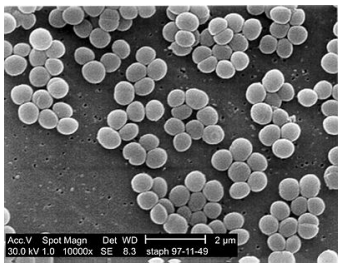
Introduction



Staphylococcus (from the Greek: σταφυλή, *staphylē*, "grape" and κόκκος, *kókkos*, "granule") is a genus of Gram-positive bacteria. Under the microscope, they appear round (cocci), and form in grape-like clusters.

The *Staphylococcus* genus includes at least 40 species. Of these, nine have two subspecies, one has three subspecies, and one has four subspecies. Most are harmless and reside normally on the skin and mucous membranes of humans and other organisms. Staphylococcus has been found to be a nectar-inhabiting microbe. Found worldwide, they are a small component of soil microbial flora.

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 Staphylococcus aureus subsp. aureus strain VB4283

Contigs	78	**GC Content**	32.7
Plasmids		**Contig L50**	9
Genome Length	2796422	**Contig N50**	105363
Chromosomes			

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 >> Terrabacteria group >> Firmicutes >> Bacilli >> Bacillales >> Staphylococcaceae >> Staphylococcus >> Staphylococcus aureus >> Staphylococcus aureus subsp. aureus*

The annotations includes 3349 CDS, 50 tRNAs and 3 rRNAs. All annotated genome features for this genome are summarized in Table 2.

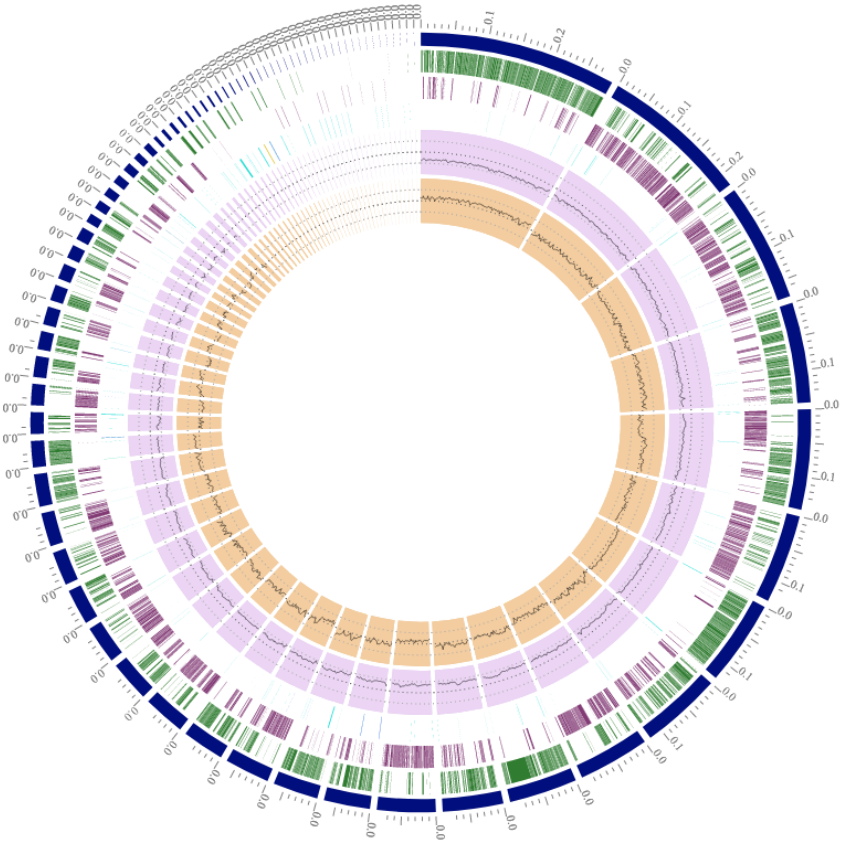
The functional annotation included 655 hypothetical proteins and 2694 proteins with functional assignments. Furthermore, 963 proteins were assigned EC numbers and 749 proteins were mapped to KEGG pathways. A breakdown of the proteins that have been annotated for this isolate is provided in Table 3.

Table 2. Annotated Genome Features

	PATRIC
CDS	2903
repeat_region	127
classifier_predicted_region	66
tRNA	56
misc_RNA	12
regulatory	10
rRNA	6

Table 3. Protein Features	
	PATRIC
Hypothetical proteins	740
Proteins with functional assignments	2163
Proteins with EC number assignments	947
Proteins with GO assignments	640
Proteins with Pathway assignments	561
Proteins with PATRIC genus-specific family (PLfam) assignments	2845
Proteins with PATRIC cross-genus family (PGfam) assignments	2845

A graphical display of the distribution of the CDS on each strand, the repeat_regions, GC content and GC skew are provided (Figure 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 4).

Table 4. Annotated Genome Features		
	Source	**Genes**
Virulence Factor	VFDB	70
Virulence Factor	Victors	35
Antibiotic Resistance	CARD	32
Drug Target	DrugBank	30
Human Homolog	Human	20
Antibiotic Resistance	PATRIC	19
Antibiotic Resistance	ARDB	10
Drug Target	TTD	10

Antimicrobial Ressistance Analysis

[some text will go here]

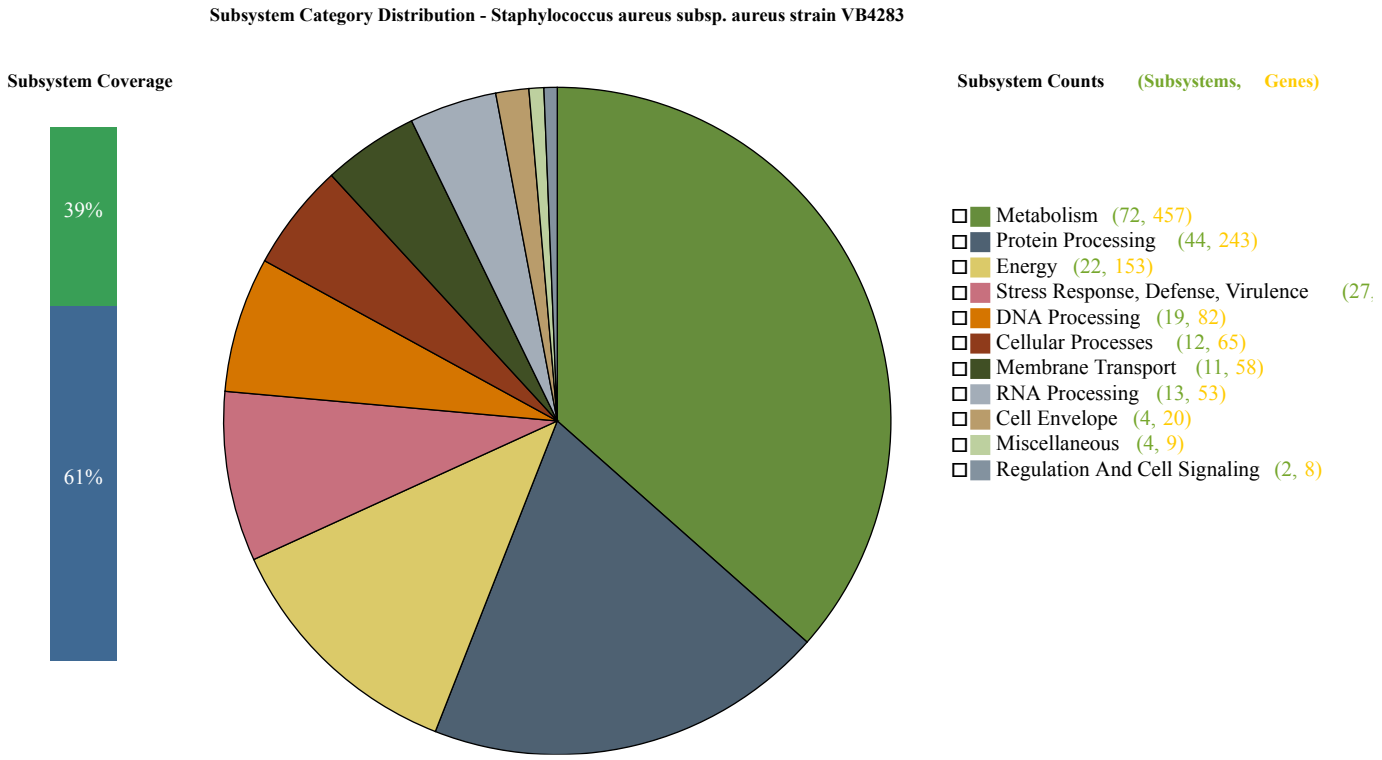
Table 3. Protein Features	
	Antibiotics
Resistant	ciprofloxacin, erythromycin, gentamicin, methicillin, penicillin, trimethoprim/sulfamethoxazole
susceptible	clindamycin, tetracycline

Antimicrobial Ressistance Genes

[Incomplete]

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.



Phylogenetic Analysis

PATRIC [1] provides a phylogentic tree building service. The makes use of several third-party tools. These include, BLAST[1], MCL[2], Muscle[3], hmmbuild[4], hmmsearch[4], Gblocks[5], FastTree[6], and RAxML[7]. The pipeline begins with thegi set of ingroup genome protein files. These are filtered to remove duplicate species, resulting in a distinct-species subset of the ingroup genomes. This is done to reduce

biasing the homolog sets with overrepresented species. BLAST searches are used to find bi-directional best hit protein pairs between genomes, and these bidirectional best hit pairs are clustered using MCL. Clusters containing members from at least half of the distinct genomes are chosen as initial, or seed, homolog sets. These seed sets are expanded to include members from all ingroup and outgroup taxa using tools from the HMMer suite. A hidden Markov Model (HMM) is built from each seed set using hmmbuild. These HMMs are used to search each genome, with hmmsearch, to find the best match from each genome for each homolog set model. The final, expanded, homolog sets are created from the hmmsearch results. Homolog sets representing fewer than 80% of ingroup genomes are removed. The remaining sets are aligned using Muscle, and the alignments are trimmed using GBlocks. The trimmed alignments are concatenated and this concatenated alignment is used to build the main tree with either RAxML or FastTree.

Comparison to Reference Genomes in the Same Genus