



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

# Whole-genome sequence data and analysis of a *Staphylococcus aureus* strain SJTUF\_J27 isolated from seaweed



Yanping Xie<sup>a</sup>, Yiping He<sup>a,\*</sup>, Sandeep Ghatak<sup>b</sup>, Peter Irwin<sup>a</sup>,  
Xianghe Yan<sup>c</sup>, Terence P. Strobaugh, Jr.<sup>a</sup>, Andrew Gehring<sup>a</sup>

<sup>a</sup> Molecular Characterization of Foodborne Pathogens Research Unit, Eastern Regional Research Center, Agricultural Research Service (ARS), United States Department of Agriculture (USDA), 600 East Mermaid Lane, Wyndmoor, PA 19038, USA

<sup>b</sup> Division of Animal Health, ICAR Research Complex for NEH Region, Umiam, Meghalaya 793103, India

<sup>c</sup> Environmental Microbial and Food Safety Laboratory, Beltsville Agricultural Research Center, ARS, USDA, 10300 Baltimore Avenue, Beltsville, MD 20705, USA

## ARTICLE INFO

## Article history:

Received 26 June 2018

Received in revised form

14 August 2018

Accepted 23 August 2018

Available online 30 August 2018

## Keywords:

*Staphylococcus aureus*

Genome assembly

Whole genome sequencing (WGS)

Virulence factor

## ABSTRACT

The complete genome sequence data of *S. aureus* SJTUF\_J27 isolated from seaweed in China is reported here. The size of the genome is 2.8 Mbp with 32.9% G + C content, consisting of 2614 coding sequences and 77 RNAs. A number of virulence factors, including antimicrobial resistance genes (fluoroquinolone, beta-lactams, fosfomycin, mupirocin, trimethoprim, and aminocoumarin) and the *egc* enterotoxin cluster, were found in the genome. In addition, the genes encoding metal-binding proteins and associated heavy metal resistance were identified. Phylogenetic data analysis, based upon genome-wide single nucleotide polymorphisms (SNPs), and comparative genomic evaluation with BLAST Ring Image Generator (BRIG) were performed for SJTUF\_J27 and four *S. aureus* strains isolated from food. The completed genome data was deposited in NCBI's GenBank under the accession number CP019117, <https://www.ncbi.nlm.nih.gov/nucleotide/CP019117>.

Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

\* Correspondence to: US Department of Agriculture, 600 E. Mermaid Lane, Wyndmoor, PA 19038, USA. Fax: +215 836 3742.  
E-mail address: [yiping.he@ars.usda.gov](mailto:yiping.he@ars.usda.gov) (Y. He).

<https://doi.org/10.1016/j.dib.2018.08.084>

2352-3409/Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Specifications Table

Subject area	Biology
More specific subject area	Microbial genomics
Type of data	Completed genome sequence in FASTA format, figures
How data was acquired	Illumina Miseq sequencing platform
Data format	Analyzed
Experimental factors	<i>Staphylococcus aureus</i> SJTUF_J27 isolated from seaweed
Experimental features	Whole genome sequencing, <i>de novo</i> assembly, and annotation
Data source location	Shanghai, China (Latitude 31.23 N and Longitude 121.47 E)
Data accessibility	Data is with this article and available online at <a href="https://www.ncbi.nlm.nih.gov/nucleotide/CP019117">https://www.ncbi.nlm.nih.gov/nucleotide/CP019117</a>

## Value of the data

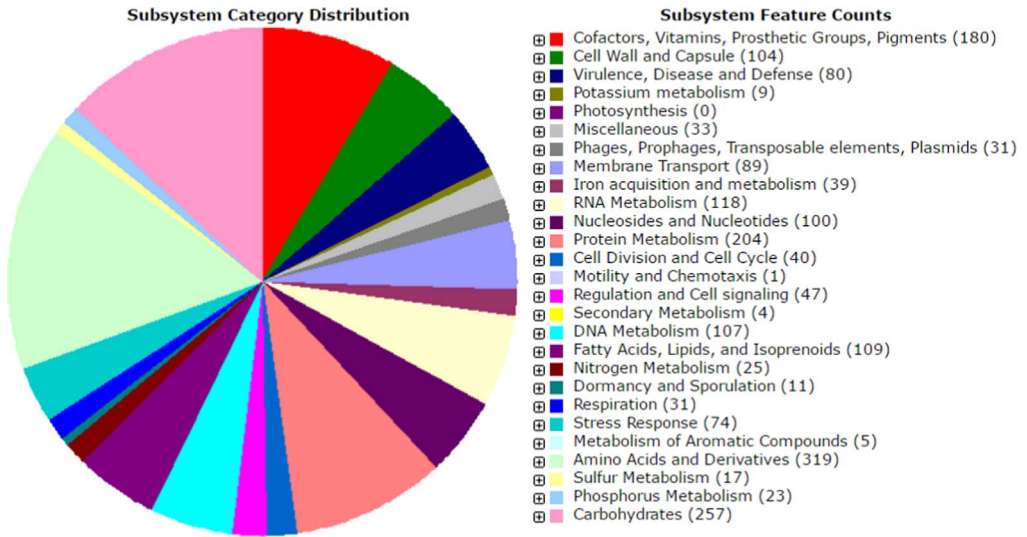
- The complete genome sequence of *S. aureus* SJTUF\_J27, which was isolated from Chinese seaweed, provides a genetic basis for understanding the epidemiology of food-associated staphylococci.
- The sequence data will be useful for comparative genomic study of *S. aureus*.
- Analyses of virulence and antibiotic resistance genes can be used to predict the probability of the organism being a multidrug resistance pathogen.
- The genome-wide SNP analysis generated a high-resolution phylogenetic tree of *S. aureus* food isolates, which is a useful tool for accurately discriminating closely related species.

## 1. Data

With an average of 331-fold sequencing coverage, a genome size of 2,804,759 bp constituting 32.9% of G + C content was generated. RAST annotation of the genome revealed a total of 399 subsystems, 2614 coding sequences (80 of them related to virulence, disease and defense), and 77 RNAs (Fig. 1). PathogenFinder showed the probability of this strain being a human pathogen was 98%.

Analysis of the genomic data showed the organism contains several antimicrobial resistance genes, including fluoroquinolone resistance-determining region of *gyrA*, *gyrB*, *parC* and *parE*, teicoplanin-associated operon of *tcaR-tcaA-tcaB*, beta-lactamase genes, and fosfomycin resistance gene *fosB*. Comprehensive Antibiotic Resistance Database (CARD) identified mupirocin resistance mediated by *ileS*, trimethoprim resistance mediated by *dfc*, and aminocoumarin resistance mediated by *alaS*. The strain harbors heavy metal resistance genes and the enterotoxin gene cluster (*egc*) but lacks staphylococcal pathogenicity islands (SaPI).

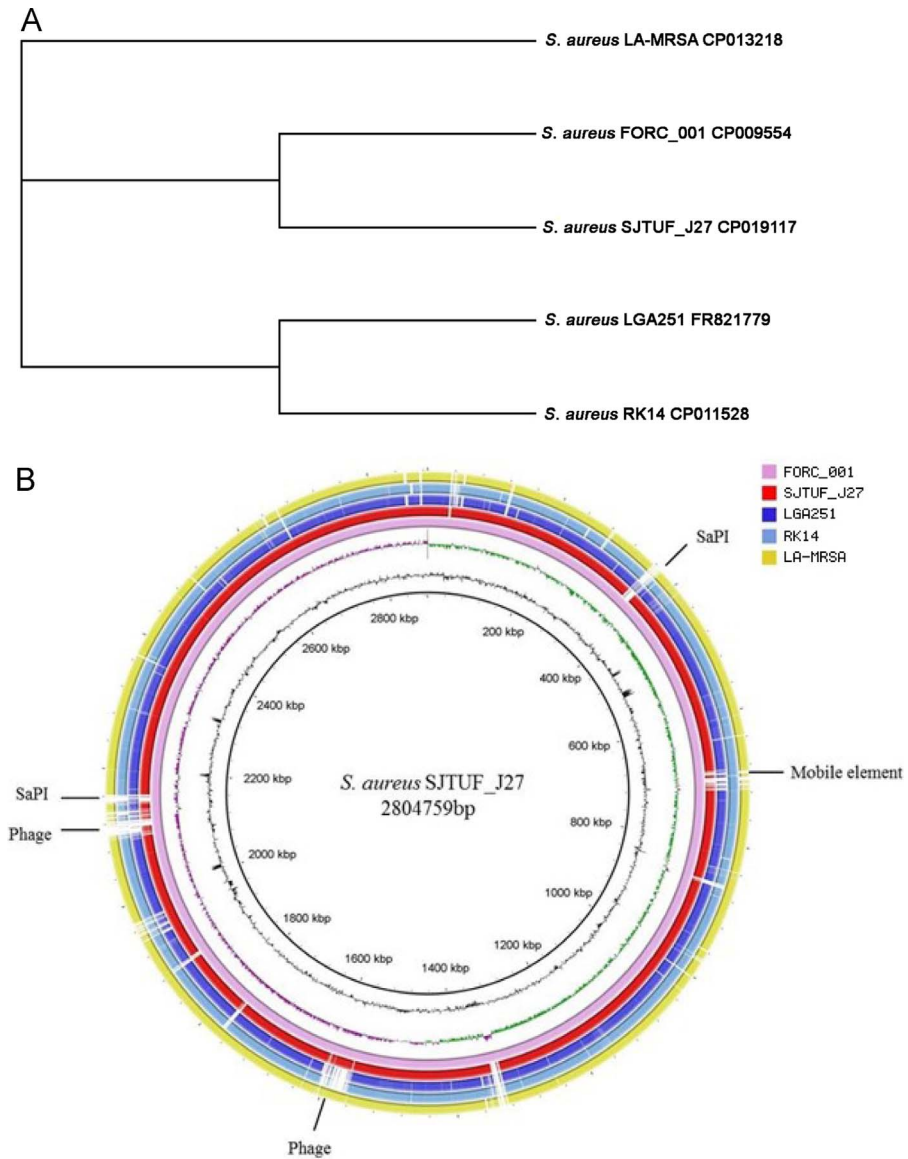
Genome-wide single nucleotide polymorphism analysis revealed the phylogenetic relationships of the strain to four food isolates (*S. aureus* FORC\_001, LGA251, RK14 and LA-MRSA with accession numbers CP009554, FR821779, CP011528 and CP013218, respectively) (Fig. 2A). BRIG analysis showed the differences between these strains mainly in the mobile genetic elements of phage and SaPI (Fig. 2B). Furthermore, MLST (<http://www.mlst.net>) showed that SJTUF\_J27 belongs to sequence type (ST)433, and all these food isolates belong to different STs (FORC\_001 to ST30, LGA251 to ST425, RK14 to ST8, and LA-MRSA to ST398).



**Fig. 1.** Subsystem categories and distribution of the *S. aureus* SJTUF\_J27 genome annotated by RAST.

## 2. Experimental design, materials and methods

*S. aureus* SJTUF\_J27 was isolated from seaweed in China. Identification of the strain was carried out using the API Staph-Ident system (bioMerieux, Shanghai, China). The result was confirmed by 16s rRNA sequencing [1]. For genome sequencing, DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA), quantified by a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA), and then subjected to library construction using the Nextera XT sample preparation kit (Illumina, San Diego, CA). Next-generation sequencing was performed in Illumina Miseq platform with  $2 \times 300$  paired-end sequencing chemistry. A total of 1,547,292 raw sequence reads were automatically generated, trimmed for quality, and then *de novo* assembled using the CLC genomics workbench v 9.5 and SPAdes 3.9. The assembled genome was validated by Sanger sequencing and mapping reads back to the assembly. The complete genome sequence of *S. aureus* SJTUF\_J27 was deposited to NCBI under the accession number CP019117. Annotation of the genome was performed using the Rapid Annotation Subsystem Technology (RAST) sever (<http://rast.nmpdr.org/>) [2]. Pathogenicity and antibiotic resistance were predicted using PathogenFinder (<https://cge.cbs.dtu.dk/services/PathogenFinder/>) and Comprehensive Antibiotic Resistance Database (CARD) (<https://card.mcmaster.ca/>) [3], respectively. Single nucleotide polymorphism (<https://cge.cbs.dtu.dk/services/CSIPhylogeny/>) [4], BLAST Ring Image Generator (BRIG) (<https://sourceforge.net/projects/brig/>) [5], and MLST (<http://www.mlst.net>) were used for comparative analyses of the *S. aureus* food isolates.



**Fig. 2.** A. Phylogenetic tree of *S. aureus* strains based on whole genome single-nucleotide polymorphisms. B. BRIG ring comparisons of the *S. aureus* strains SJTUF\_J27, FORC\_001, LGA251, RK14, and LA-MRSA. The main divergent regions are labeled with SaPI, phage and mobile element. The innermost rings represent the GC content (black) and GC skew (purple/green) of FORC\_001.

### Acknowledgement

This research was supported by the US Department of Agriculture, 8072-42000-071-00D Agricultural Research Service (USDA-ARS). We thank Dr. Xianming Shi at Shanghai Jiaotong University, Shanghai, China for providing the strain. SG is thankful to, the Department of Biotechnology (DBT), Ministry of Science & Technology, Government of India for Overseas Associateship Grant for the North Eastern Region (2015–16) and to the Indian Council of Agricultural Research for necessary support.

## Transparency document. Supplementary material

Transparency data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2018.08.084>.

## References

- [1] D.J. Lane, 16S/23S rRNA sequencing, in: E. Stackebrandt, M. Goodfellow (Eds.), *Nucleic Acid Techniques in Bacterial Systematics*, John Wiley & Sons, Chichester, UK, 1991, pp. 115–175.
- [2] R. Overbeek, R. Olson, G.D. Pusch, G.J. Olsen, J.J. Davis, T. Disz, R.A. Edwards, S. Gerdes, B. Parrello, M. Shukla, et al., The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST), *Nucleic Acids Res.* 42 (2014) D206–D214. <https://doi.org/10.1093/nar/gkt1226>.
- [3] A.G. McArthur, N. Wagleichner, F. Nizam, A. Yan, M.A. Azad, A.J. Baylay, K. Bhullar, M.J. Canova, G. de Pascale, L. Ejim, et al., The comprehensive antibiotic resistance database, *Antimicrob. Agents Chemother.* 57 (2013) 3348–3357. <https://doi.org/10.1128/AAC.00419-13>.
- [4] R.S. Kaas, P. Leekitcharoenphon, F.M. Aarestrup, O. Lund, Solving the problem of comparing whole bacterial genomes across different sequencing platforms, *PLoS One* 9 (2014) e104984. <https://doi.org/10.1371/journal.pone.0104984>.
- [5] N. Alikhan, N.K. Petty, N.L.B. Zakour, S.A. Beatson, BLAST ring image generator (BRIG): simple prokaryote genome comparisons, *BMC Genom.* 12 (2011) 402–412. <https://doi.org/10.1186/1471-2164-12-402>.