

## Hybrid assembly using ultra-long reads resolves repeats and completes the genome sequence of a laboratory strain of *Staphylococcus aureus* subsp. *aureus* RN4220

Suresh Panthee<sup>1,2</sup>#, Hiroshi Hamamoto<sup>2,3</sup>#, Atmika Paudel<sup>3</sup>, Chikara Kaito<sup>4</sup>, Yutaka Suzuki<sup>5</sup>, Kazuhisa Sekimizu<sup>1,\*</sup>

<sup>1</sup> Drug Discoveries by Silkworm Models, Faculty of Pharma-Science, Teikyo University, Tokyo, Japan

<sup>2</sup> Teikyo University Institute of Medical Mycology, 359 Otsuka, Hachioji, Tokyo, 192-0395, Japan

<sup>3</sup> Division of Infection and Immunity, International Institute for Zoonosis Control, Hokkaido University, North 20, West 10, Kita-ku, Sapporo, Hokkaido 001-0020, Japan

<sup>4</sup> Graduate School of Medicine Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, Japan

<sup>5</sup> Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan

#Contributed equally to this manuscript.

Email address of all authors:

SP: [supanthee@main.teikyo-u.ac.jp](mailto:supanthee@main.teikyo-u.ac.jp)

HH: [hamamotoh@main.teikyo-u.ac.jp](mailto:hamamotoh@main.teikyo-u.ac.jp)

AP: [atmikapd@czc.hokudai.ac.jp](mailto:atmikapd@czc.hokudai.ac.jp)

YS: [ysuzuki@k.u-tokyo.ac.jp](mailto:ysuzuki@k.u-tokyo.ac.jp)

CK: [ckaito@okayama-u.ac.jp](mailto:ckaito@okayama-u.ac.jp)

\*Correspondence to: [sekimizu@main.teikyo-u.ac.jp](mailto:sekimizu@main.teikyo-u.ac.jp)

## Abstract

*Staphylococcus aureus* RN4220 has been extensively used by staphylococcal researchers as an intermediate strain for genetic manipulation due to its ability to accept foreign DNA. Despite its wide use in laboratories, its complete genome is not available. In this study, we used the hybrid genome assembly approach using the minION long reads and Illumina short reads to sequence the complete genome of *S. aureus* RN4220. The comparative analysis of the annotated complete genome showed the presence of 39 genes fragmented in the previous assembly, many of which were located near the repeat regions. Using RNA-Seq reads, we showed that a higher number of reads could be mapped to the complete genome than the draft genome and the gene expression profile obtained using the complete genome also differs from that obtained from the draft genome. Furthermore, by comparative transcriptomic analysis, we showed the correlation between expression levels of staphyloxanthin biosynthetic genes and the production of yellow pigment. This study highlighted the importance of long reads in completing the microbial genomes, especially those possessing repetitive elements.

**Keywords:** *Staphylococcus aureus*, RN4220, complete genome, hybrid assembly, RNA-Seq, repeat regions, staphyloxanthin

## Introduction

*Staphylococcus aureus* is a Gram-positive bacterium capable of opportunistic infections, which can sometimes be fatal. Genetic manipulation of *S. aureus* was limited until *S. aureus* strain RN4220 was obtained by chemical mutagenesis of *S. aureus* NCTC8325-4 strain[1]. NCTC8325-4 is a derivative of a clinical isolate NCTC8325 obtained by curing the three prophages  $\Phi$ 11,  $\Phi$ 12, and  $\Phi$ 13[2]. Therefore, both NCTC8325-4 and RN4220 lack the three prophages. In addition, RN4220 can accept foreign DNA and is characterized by a mutation in the *saul hsdR* gene belonging to the restriction-modification system[3]. Due to this property, RN4220 is routinely used in the laboratories as an intermediate for genetic manipulation; plasmids from *Escherichia coli* are electro-transformed into RN4220, and the plasmids from RN4220 are then transformed to another *S. aureus* strain by suitable methods such as phage transduction.

Despite its wide use, the complete genome of this strain is not available. With the recent development in next-generation sequencing technologies, there have been attempts to sequence the genome. Apart from our assembly, there are two deposited assemblies of RN4220 in NCBI. The first assembly was done in 2011 using Illumina GA II[4] (accession: GCA\_000212435.2), and the second was performed in 2020 using BGISEq (accession: GCA\_011751615.1), which generated 118 and 27 contigs, respectively. Whereas the assemblies primarily provided valuable information regarding the genetic make-up of this strain, we still need a complete genome sequence to make the most out of this laboratory strain. Short reads sequencing of the genome can be attributed to the large number of contigs generated from these

assemblies. Short read assemblies are challenged by the presence of identical sequences at more than one locus of the chromosome called the repetitive DNA sequences, or repetitive elements or repeated regions. Larger organisms such as eukaryotes have many repetitive elements throughout the genome; for instance, nearly half of the human genome consists of repetitive elements[5]. Bacteria such as *Orientia tsutsugamushi* possess 37% repetitive elements throughout its genome[6]. In this study, by hybrid assembly using both long and short reads, we completed the genome of RN4220. Upon further analysis, we found many repetitive elements and several fragmented genes in the previous assembly. Our approach using long reads was suitable for covering those repetitive regions, and we were successful in obtaining a complete genome.

## Materials and Methods

### Genome sequencing, assembly, and annotation

*S. aureus* RN4220 was routinely cultured on tryptic soy broth at 37°C without antibiotic selection. Genome sequencing and assembly was performed as previously explained[7-10]. Briefly, genomic DNA was isolated from overnight culture using Qiagen DNA-blood Mini Kit (Qiagen, Hilden, Germany) and lysostaphin for bacterial lysis. Construction of short-read single-end libraries and sequencing was performed using Illumina HiSeq2000 (Illumina, San Diego, CA, USA)[11]. MinION long reads sequencing was performed using 1 µg genomic DNA. Hybrid error correction of the long reads was performed by LoRDEC[12] using the short reads, and final assembly of the circular chromosome was performed using Flye 2.3.3[13]. Short reads were then mapped to the chromosome, and the consensus was generated to obtain the final assembly. The final assembly was then annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP).

### Comparative genomic analysis

The complete genome sequence of the parent strain NCTC8325 and two draft assemblies of RN4220 strain were downloaded from NCBI. The draft assemblies were first aligned using Mauve Contig Mover[14]. The ordered contigs were then submitted to the CLC Genomics workbench for whole-genome alignment. To identify the genes fragmented in the previous assembly, the annotations were checked manually. Repeat finding of the genome was performed using Unipro UGENE v.39.0[15].

### Staphyloxanthin biosynthesis pathway analysis

Raw RNA-Seq reads for *S. aureus* RN4220, NCTC8325, and SH1000 were downloaded from NCBI SRA. The SRA accession numbers and BioProject details are summarized in **Table 1**. The reads were then mapped to their respective complete genome using the CLC Genomic Workbench ver 20.0.4 (CLC bio, Aarhus, Denmark). Since the complete genome sequence of the SH1000 strain is not publicly available, the SH1000 reads were mapped to the NCTC8325 genome. Transcripts Per Million (TPM), a sequence

depth normalized indicator for expression analysis, was used to compare the expression level among the study samples.

**Table 1: SRA accession numbers used in this study**

<i>S. aureus</i>	BioProject	Sample	SRA accession	Sequencing method	Institute
RN4220	PRJDB5479	1_RN4220-1	DRR084259	Illumina HiSeq 2500	University of Tokyo
		2_RN4220-2	DRR084260		Graduate School of
		3_RN4220-3	DRR084261		Pharmaceutical Sciences
		4_DcvfE-1	DRR084262		
		5_DcvfE-2	DRR084263		
		6_DcvfE-3	DRR084264		
NCTC8325	PRJNA433003[16]	wt1	SRR6674886	Illumina HiSeq 2500	University of
		wt2	SRR6674887		British
		wt3	SRR6674888		Columbia
SH1000	PRJNA682641[17]	PL_TSB_1	SRR13200515	Illumina NextSeq 500	Université de Sherbrooke
		PL_TSB_2	SRR13200516		
		PL_TSB_3	SRR13200517		
		Bio_TSB_1	SRR13200521		
		Bio_TSB_2	SRR13200522		
		Bio_TSB_3	SRR13200523		
		PL_BHI_1	SRR13200518		
		PL_BHI_2	SRR13200519		
		PL_BHI_3	SRR13200520		
		Bio_BHI_1	SRR13200524		
		Bio_BHI_2	SRR13200525		
		Bio_BHI_3	SRR13200526		
PRJNA472336[18]		WT1_S7	SRR7189482	Illumina NextSeq 500	University of
		WT3_S9	SRR7189483		Leeds, UK

## Results and Discussions

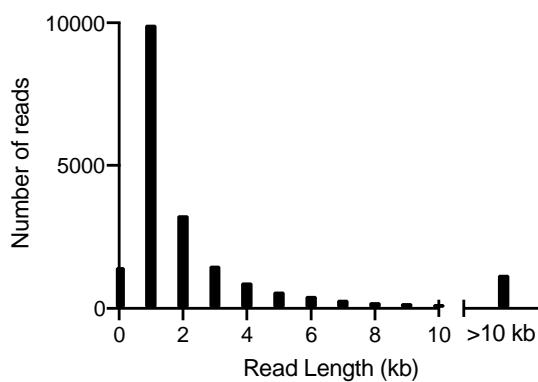
### Completion of the genome sequence of *S. aureus* RN4220

Previous attempts to sequence the RN4220 genome used the sequencers that produced short reads. *S. aureus* NCTC8325, the parent strain of RN4220, also possesses repetitive elements, which create a challenge in assembling the genome using short reads. To cover the repetitive elements while sequencing a genome, either the reads longer than the repetitive elements or alternative approaches to overcoming this problem are essential. In this study, we took the advantage of our hybrid genome assembly approach[7-10] using long reads from ONT MinION and short reads from Illumina to complete the RN4220 genome. As high-quality long reads are much crucial than coverage, we could complete the genome of this bacterium with low coverage long reads, and high coverage short reads. The summary of reads obtained from both the MinION and Illumina platforms is shown in **Figure 1A, B**.

**A**

Properties	Platform	
	MinION	Illumina
Total number of reads	20138	38648219
Average read length (bp)	3189	35.8
Longest read (bp)	147164	-
Shortest read (bp)	57	-
Total bases (bp)	64220140	1382108289
Coverage (fold)	23.8	512.4

**B**

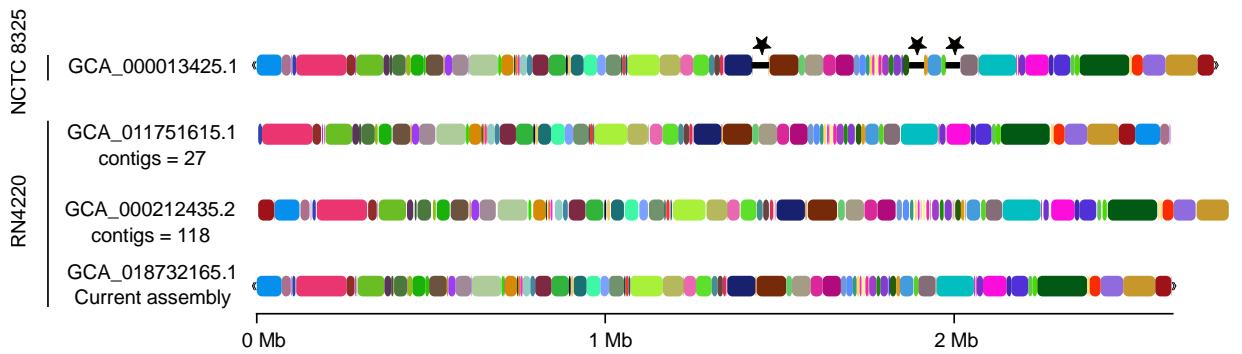


**Figure 1. Genome sequencing of *S. aureus* RN4220 using hybrid-genome assembly approach. (A)** Summary of sequence reads from MinION and Illumina sequencers. **(B)** Histogram of read length obtained from the MinION sequencer.

The complete genome sequence of *S. aureus* RN4220 was 2.7 Mb in length, harbored 2654 genes, including 19 rRNAs and 59 tRNAs (Table 2). To identify the genomic difference among the parent strain and RN4220 assemblies, we performed the whole genome alignment of the parent strain NCTC8325[19] (assembly accession: GCA\_000013425.1). We found that large regions from NCTC8325 were deleted in the RN4220 genome, which has also been noted by the previous analysis using the draft genome [4].

**Table 2:** Analysis and comparison of general features of the current, complete *S. aureus* RN4220 genome with previous draft assemblies.

Features	Current assembly GCA_018732165.1	GCA_011751615.1	GCA_000212435.2[4]
Assembly release date	2021/06/07	2020/03/25	2011/05/05
Sequencing technology	ONT minION Illumina HiSeq	BGIseq	Illumina GA II
Genome coverage	ONT minION: 23x Illumina HiSeq: 512x	372x	77x
Total Sequence length (bp)	2,697,195	2,657,542	2,663,395
No of contigs	1	27	118
Contig N50	-	174,720	80,460
Contig L50	-	4	13
Protein coding genes	2508	2517	2540
rRNA genes	78	-	-
CDS	2572	2571	2604
Gene	2654	2626	2661
Misc. binding	3	3	3
Misc. feature	3	3	3
ncRNA	3	3	3
Regulatory	10	10	10
rRNA	19	-	8
tmRNA	1	1	1
tRNA	59	51	45



**Figure 2. Comparison of *S. aureus* RN4220 genome.** Whole-genome alignment of the parent strain NCTC8325 with RN4220 assemblies. The three regions indicated by the stars were present in the NCTC8325 but not in the RN4220 genomes. The homologous regions were randomly colored for ease of distinction. The assembly GCA\_000212435.2 appears longer due to the large number of contigs.

#### Identification of fragmented genes and presence of repetitive elements

We compared our current assembly with the first genome assembly and found that many genes were fragmented and possibly not detected earlier. We found 39 new genes among the fragmented regions. Some notable genes identified included KMZ21\_00310: *spa*; KMZ21\_02625: *sdrC*; KMZ21\_02630: *sdrD*; KMZ21\_06705: *ebh*; and KMZ21\_08630: *spfF*. As these proteins are known to be involved in *S. aureus* pathogenesis through processes such as immune evasion, adhesion, complement resistance, and substrate acquisition, it is speculated that the virulence potential, based on the draft genome sequence,

might have been overlooked. Furthermore, we found that 14 of these genes were located around the repetitive elements, suggesting the importance of long reads in the genome assembly (**Table 3**).

**Table 3:** Fragmented genes in the previous assembly [4], identified through the complete genome analysis. (-) in the position, column indicates gene in complementary strand, and + or - sign in the repeat column indicates a presence or absence of repetitive elements in the proximity of the gene, respectively.

SN	locus_tag	product	position	repeat
1	KMZ21_00310	spa: Staphylococcal protein A	72912..74462(-)	-
2	KMZ21_00450	deoC: deoxyribose phosphate aldolase	103991..104653	-
3	KMZ21_00505	tnpA: IS200/IS605 family transposase	115089..115573	-
4	KMZ21_01125	transposase	263883..264233	-
5	KMZ21_01235	TIGR01741 family protein	288910..289401	+
6	KMZ21_01240	DUF5079 family protein	289609..290292	+
7	KMZ21_01265	TIGR01741 family protein	292558..293053	+
8	KMZ21_01275	DUF5079 family protein	293765..294448	+
9	KMZ21_01285	TIGR01741 family protein	295034..295534	+
10	KMZ21_01295	antitoxin YezG family	296056..296556	+
11	KMZ21_01300	TIGR01741 family protein	296567..297067	+
12	KMZ21_01855	restriction-modification system subunit M	403369..404925	-
13	KMZ21_02625	sdrC	559794..560643	-
14	KMZ21_02630	sdrD	561010..565159	-
15	KMZ21_04460	mgtE	925408..926793	-
16	KMZ21_04700	hypothetical protein	972806..972916	-
17	KMZ21_05015	pyruvate carboxylase	1037702..1041154	-
18	KMZ21_05785	proline tRNA ligase	1196448..1198151	-
19	KMZ21_06475	transposase	1341780..1342073	-
20	KMZ21_06560	transposase	1357791..1358271(-)	-
21	KMZ21_06705	ebh: hyperosmolarity resistance protein	1384986..1413593(-)	-
22	KMZ21_06985	DUF	1469250..1470200(-)	+
23	KMZ21_07000	DUF	1473421..1474350(-)	+
24	KMZ21_07575	Acetyl-CoA carboxylase biotin carboxylase subunit	1578089..1579450(-)	-
25	KMZ21_08010	IS1182 transposase	1667723..1669374(-)	-
26	KMZ21_08245	hypothetical protein	1721716..1723233(-)	+
27	KMZ21_08370	transposase	1757495..1758156(-)	-
28	KMZ21_08480	IS1182 transposase	1775882..1777506	-
29	KMZ21_08525	menC	1785343..1786344(-)	-
30	KMZ21_08620	type I restriction-modification system subunit M	1802604..1804160(-)	+
31	KMZ21_08630	serine protease splF	1804523..1805242(-)	+
32	KMZ21_08640	hypothetical protein	1806211..1806318	+
33	KMZ21_08930	ISL3-like element IS1181 family transposase	1859410..1860729	-
34	KMZ21_09435	IS5/IS1182 family transposase	1936612..1936820(-)	-
35	KMZ21_10360	LmrS: multidrug efflux MFS transporter	2121681..2123123(-)	-
36	KMZ21_10440	IS1182 transposase	2140090..2141714	-
37	KMZ21_11355	transposase	2301164..2301720	-
38	KMZ21_12080	E domain-containing protein	2445649..2450148(-)	+
39	KMZ21_12100	fibronectin-binding protein FnbB	2453135..2455957(-)	-

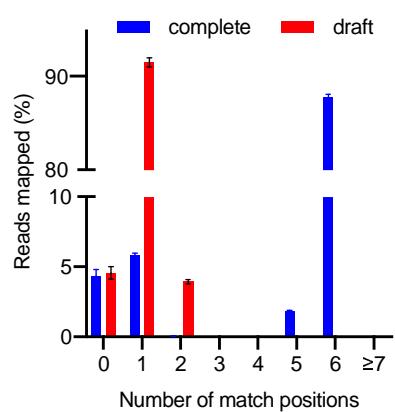
### Complete genome facilitates the RNA-Seq analysis

Using the publicly available data, we calculated the mapping of short RNA-Seq reads and fold expression of genes and compared the data with the complete and draft RN4220 genomes. We found that the number of reads mapped to the complete genome was slightly higher in the complete genome than that of the draft genome (**Figure 3A**). It is well known that the scaffolds in most draft genomes contain gaps[20] which may sometimes lead to the difference in the mapping of the reads and might be misleading while interpreting results. Therefore, we can expect that the increased mapping could be due to the mapping of the additional reads in the “gap” region that lied in between the contigs in the draft genome. Interestingly, the number of reads mapped specifically to the genome drastically reduced in the complete genome (**Figure 3A**). This could be because of the resolved repeat and duplicate regions in the complete genome, which appeared as a single contig in the draft genome. We found more than 80% of the reads were matched to six positions in the complete genome, which in the case of the draft genome was one (**Figure 3B**). To compare the difference between the gene expression analysis obtained using a complete genome and draft genome, we used RNA-seq reads from RN4220 wild-type and  $\Delta$ KMZ21\_09125 (SA1684), involved in virulence of *S. aureus* strains[21]. Using a volcano plot (**Figure 3C**), we could see a vast difference in the differential expression of the genes when using the complete and the draft genome. Besides, the lack of the RN4220 complete genome required researchers to analyze RNA-Seq reads by mapping to the NCTC8325 genome[21, 22], where the results should be carefully interpreted, considering the differences among NCTC8325 and RN4220 genome sequences. In summary, these suggested the importance of a complete genome for omics-based analysis.

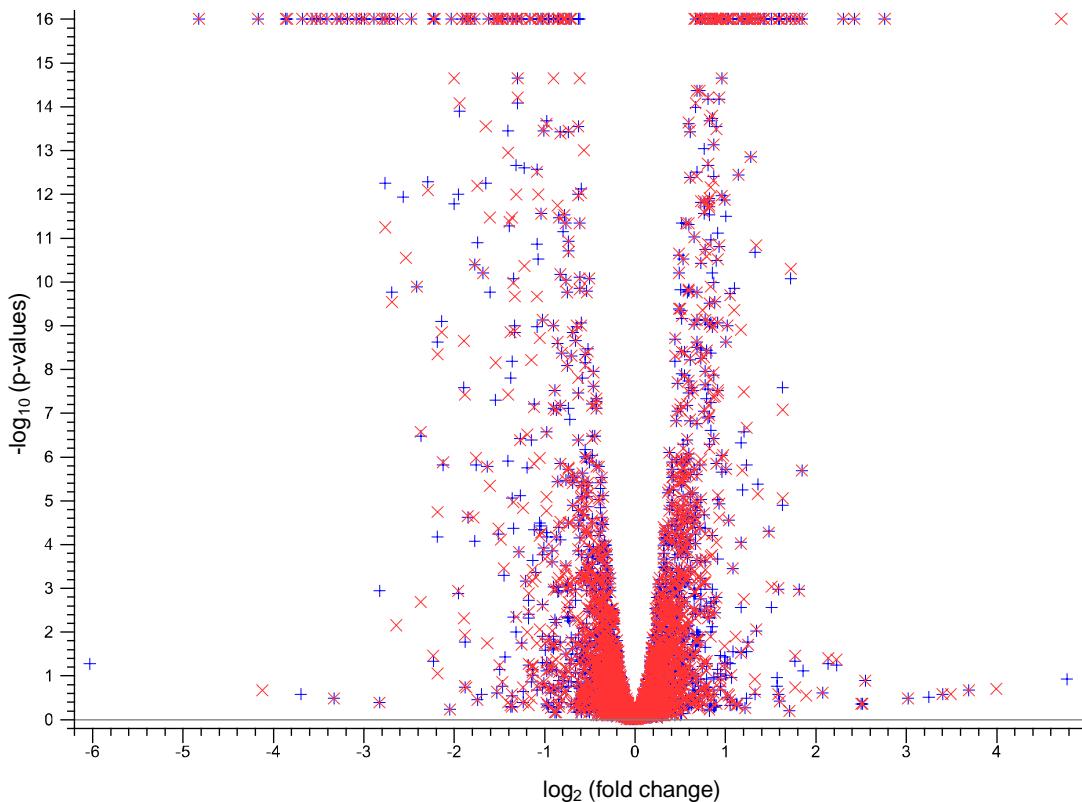
**A**

	SRA		
	DRR084259	DRR084260	DRR084261
No of reads	24,073,750	20,841,626	35,553,574
Reads mapped to genome			
complete	23,081,817	20,050,407	33,704,521
draft	23,035,455	20,010,309	33,637,924
Uniquely mapped to genome			
complete	1,386,202	1,269,534	2,004,865
draft	22,025,437	19,250,167	32,227,969

**B**



**C**

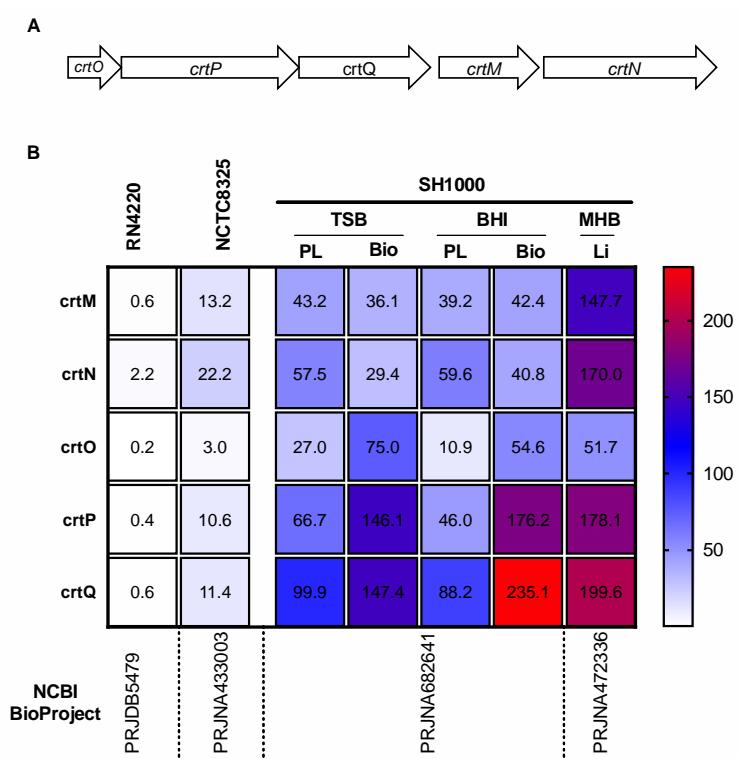


**Figure 3: Comparison of complete (current) and draft[4] genome assemblies to analyze RNA-Seq results.** Mapping (A) and match specificities (B) of RN4220 wild-type reads to the genome. (C) The differential expression of the genes between  $\Delta cvfE$  and wild-type strains using complete and draft genome (+: complete genome; x: draft genome).

#### Downregulation of RN4220 genes involved in staphyloxanthin biosynthesis

*S. aureus* strains are usually distinguishable from other bacteria due to their yellow color, which is because of the production of yellow pigment staphyloxanthin. However, the RN4220 strain does not give

the yellow pigmentation, hence produces very little or no staphyloxanthin. Staphyloxanthin biosynthetic genes are located in an operon *crtMNOPQ* (**Figure 4A**)[23, 24] which is dependent upon the sigma factor B (SigB)[25]. SigB falls in an operon *rsbUVWSigB*, where RsbU and RsbV are the activators, and RsbW is the repressor of SigB[26-28]. It has been known that SigB is also controlled by YjbH[29] and CspA[30]. The parent strain of RN4220, NCTC8325, has a reduced ability to produce staphyloxanthin[31] and this is attributed to a deletion of 11 bp in the *rsbU* gene. Since staphyloxanthin production is more pronouncedly decreased in RN4220 compared with NCTC8325 and NCTC8325-4[31], we expected that the RN4220 strain might have some further alterations within these two operons. We aligned the amino acid sequences of the genes and found that these operons were conserved. Next, we aimed to examine the difference at the gene expression level. We looked for the raw RNA-Seq reads in NCBI SRA for NCTC8325, RN4220, and SH1000 strains. The *S. aureus* SH1000[32] strain is a *rsbU<sup>r</sup>* derivative of NCTC8325 and has the ability to produce staphyloxanthin. The reads were then mapped to complete genomes, and expression was analyzed using transcripts per million (TPM). We found that the *crt* operon was dramatically downregulated in RN4220, consistent with its pigment-less phenotype (**Figure 4B**). As we did not find a genetic level changes within the *crt* operon, it is expected that some other regulatory factors might play a role in the observed difference in staphyloxanthin production.



**Figure 4: Staphyloxanthin biosynthetic gene cluster (A) and expression of the genes among NCTC8325, and SH1000, the *rsbU* repaired NCTC8325-4 strain (B).** The short reads from the NCBI database (indicated Bioproject) were downloaded, mapped to respective genomes, and expression values, expressed as TPMs, were calculated using CLC Genomics Workbench. The mean of TPM values

is shown. For RN4220, wild-type data was used. PL: Planktonic growth; Bio: Biofilm growth; Li: liquid culture.

## Conclusion

In this study, we completed the genome of a popular laboratory strain RN4220 for the first time. We provided an example of the importance of long reads in completing genomes containing repetitive elements, which is not possible by usual short-read sequences. The availability of the complete genome of this widely used strain is expected to serve as a platform for further genetic manipulation in a defined manner and robust omics-based analysis. In addition, we found that although the staphyloxanthin gene cluster was intact in RN4220, transcription of the operon was weak, resulting in a dramatic decrease in staphyloxanthin production and, hence, its pigment-less phenotype. Overall, the findings of this study provide valuable information on *S. aureus* RN4220 strain by completing its genome, which will help interpret results in a defined manner by reducing biases and broadening our understanding of the genetic basis of various phenotypes.

## Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers 19K16653JP to A.P., 19K07140JP to H.H., 21H02733JP to K.S., and 15H05783 to K.S. and H.H., and TBRF, IFO fellowships to S.P.

## Data availability

The complete genome of *S. aureus* RN4220 has been deposited to NCBI GenBank with accession CP076105.

## Author contributions

S.P. and H.H. conceived the idea. S.P. and A.P. performed long-read sequencing, assembled, analyzed the genome, and wrote the manuscript. Y.S. and H.H. performed short-read sequencing. C.K. performed RNA-seq analysis of RN4220 wild-type and  $\Delta$ KMZ21\_09125. H.H., C.K. and K.S. provided critical discussion for writing the manuscript. All authors contributed critically to the interpretation of the results and gave final approval for publication.

## References:

- [1] B.N. Kreiswirth, S. Lofdahl, M.J. Betley, M. O'Reilly, P.M. Schlievert, M.S. Bergdoll, R.P. Novick, The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage, *Nature*, 305 (1983) 709-712.
- [2] R. Novick, Properties of a cryptic high-frequency transducing phage in *Staphylococcus aureus*, *Virology*, 33 (1967) 155-166.
- [3] D.E. Waldron, J.A. Lindsay, Sau1: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into *Staphylococcus aureus* and between *S. aureus* isolates of different lineages, *J Bacteriol*, 188 (2006) 5578-5585.

[4] D. Nair, G. Memmi, D. Hernandez, J. Bard, M. Beaume, S. Gill, P. Francois, A.L. Cheung, Whole-genome sequencing of *Staphylococcus aureus* strain RN4220, a key laboratory strain used in virulence research, identifies mutations that affect not only virulence factors but also the fitness of the strain, *J Bacteriol*, 193 (2011) 2332-2335.

[5] T.J. Treangen, S.L. Salzberg, Repetitive DNA and next-generation sequencing: computational challenges and solutions, *Nature reviews. Genetics*, 13 (2011) 36-46.

[6] N.H. Cho, H.R. Kim, J.H. Lee, S.Y. Kim, J. Kim, S. Cha, S.Y. Kim, A.C. Darby, H.H. Fuxelius, J. Yin, J.H. Kim, J. Kim, S.J. Lee, Y.S. Koh, W.J. Jang, K.H. Park, S.G. Andersson, M.S. Choi, I.S. Kim, The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes, *Proc Natl Acad Sci U S A*, 104 (2007) 7981-7986.

[7] S. Panthee, A. Paudel, H. Hamamoto, A.C. Uhlemann, K. Sekimizu, The Role of Amino Acid Substitution in HepT Toward Menaquinone Isoprenoid Chain Length Definition and Lysocin E Sensitivity in *Staphylococcus aureus*, *Front Microbiol*, 11 (2020) 2076.

[8] S. Panthee, A. Paudel, H. Hamamoto, A.A. Ogasawara, T. Iwasa, J. Blom, K. Sekimizu, Complete genome sequence and comparative genomic analysis of *Enterococcus faecalis* EF-2001, a probiotic bacterium, *Genomics*, 113 (2021) 1534-1542.

[9] S. Panthee, A. Paudel, J. Blom, H. Hamamoto, K. Sekimizu, Complete Genome Sequence of *Weissella hellenica* 0916-4-2 and Its Comparative Genomic Analysis, *Front Microbiol*, 10 (2019) 1619.

[10] S. Panthee, H. Hamamoto, S.A. Ishijima, A. Paudel, K. Sekimizu, Utilization of Hybrid Assembly Approach to Determine the Genome of an Opportunistic Pathogenic Fungus, *Candida albicans* TIMM 1768, *Genome Biol Evol*, 10 (2018) 2017-2022.

[11] S. Panthee, H. Hamamoto, Y. Suzuki, K. Sekimizu, *In silico* identification of lysocin biosynthetic gene cluster from *Lysobacter* sp. RH2180-5, *J. Antibiot.*, 70 (2017) 204-207.

[12] L. Salmela, E. Rivals, LoRDEC: accurate and efficient long read error correction, *Bioinformatics*, 30 (2014) 3506-3514.

[13] M. Kolmogorov, J. Yuan, Y. Lin, P.A. Pevzner, Assembly of long, error-prone reads using repeat graphs, *Nat Biotechnol*, 37 (2019) 540-546.

[14] A.C.E. Darling, B. Mau, F.R. Blattner, N.T. Perna, Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements, *Genome Research*, 14 (2004) 1394-1403.

[15] K. Okonechnikov, O. Golosova, M. Fursov, U. team, Unipro UGENE: a unified bioinformatics toolkit, *Bioinformatics*, 28 (2012) 1166-1167.

[16] S.C. Mansour, A novel peptide-based treatment for bacterial abscess infections, in, PhD Thesis, University of British Columbia, 2018.

[17] M. Gélinas, L. Museau, A. Milot, P.B. Beauregard, Cellular adaptation and the importance of the purine biosynthesis pathway during biofilm formation in Gram-positive pathogens, *bioRxiv*, (2020).

[18] C.P. Randall, A. Gupta, B. Utley-Drew, S.Y. Lee, G. Morrison-Williams, A.J. O'Neill, Acquired Nisin Resistance in *Staphylococcus aureus* Involves Constitutive Activation of an Intrinsic Peptide Antibiotic Detoxification Module, *mSphere*, 3 (2018).

[19] A.F. Gillaspy, V. Worrell, J. Orvis, B.A. Roe, D.W. Dyer, J.J. Iandolo, The *Staphylococcus aureus* NCTC 8325 genome, in: V. Fischetti, R. Novick, J. Ferretti, D. Portnoy, J. Rood (Eds.) *Gram - Positive Pathogens*, ASM Press, Washington DC, 2006, pp. 381-412.

[20] A.M. Giani, G.R. Gallo, L. Gianfranceschi, G. Formenti, Long walk to genomics: History and current approaches to genome sequencing and assembly, *Comput Struct Biotechnol J*, 18 (2020) 9-19.

[21] K. Imae, Y. Saito, H. Kizaki, H. Ryuno, H. Mao, A. Miyashita, Y. Suzuki, K. Sekimizu, C. Kaito, Novel Nucleoside Diphosphatase Contributes to *Staphylococcus aureus* Virulence, *J Biol Chem*, 291 (2016) 18608-18619.

[22] J. Osmundson, S. Dewell, S.A. Darst, RNA-Seq reveals differential gene expression in *Staphylococcus aureus* with single-nucleotide resolution, *PLoS One*, 8 (2013) e76572.

[23] L. Xue, Y.Y. Chen, Z. Yan, W. Lu, D. Wan, H. Zhu, Staphyloxanthin: a potential target for antivirulence therapy, *Infect Drug Resist*, 12 (2019) 2151-2160.

[24] A. Pelz, K.P. Wieland, K. Putzbach, P. Hentschel, K. Albert, F. Gotz, Structure and biosynthesis of staphyloxanthin from *Staphylococcus aureus*, *J Biol Chem*, 280 (2005) 32493-32498.

[25] M. Bischoff, P. Dunman, J. Kormanec, D. Macapagal, E. Murphy, W. Mounts, B. Berger-Bachi, S. Projan, Microarray-based analysis of the *Staphylococcus aureus* sigmaB regulon, *J Bacteriol*, 186 (2004) 4085-4099.

- [26] M. Palma, A.L. Cheung, sigma(B) activity in *Staphylococcus aureus* is controlled by RsbU and an additional factor(s) during bacterial growth, *Infect Immun*, 69 (2001) 7858-7865.
- [27] P. Giachino, S. Engelmann, M. Bischoff, Sigma(B) activity depends on RsbU in *Staphylococcus aureus*, *J Bacteriol*, 183 (2001) 1843-1852.
- [28] J. Pane-Farre, B. Jonas, S.W. Hardwick, K. Gronau, R.J. Lewis, M. Hecker, S. Engelmann, Role of RsbU in controlling SigB activity in *Staphylococcus aureus* following alkaline stress, *J Bacteriol*, 191 (2009) 2561-2573.
- [29] A. Paudel, S. Panthee, H. Hamamoto, T. Grunert, K. Sekimizu, YjbH regulates virulence genes expression and oxidative stress resistance in *Staphylococcus aureus*, *Virulence*, 12 (2021) 470-480.
- [30] S. Katzif, E.H. Lee, A.B. Law, Y.L. Tzeng, W.M. Shafer, CspA regulates pigment production in *Staphylococcus aureus* through a SigB-dependent mechanism, *J Bacteriol*, 187 (2005) 8181-8184.
- [31] K.T. Baek, D. Frees, A. Renzoni, C. Barras, N. Rodriguez, C. Manzano, W.L. Kelley, Genetic variation in the *Staphylococcus aureus* 8325 strain lineage revealed by whole-genome sequencing, *PLoS One*, 8 (2013) e77122.
- [32] M.J. Horsburgh, J.L. Aish, I.J. White, L. Shaw, J.K. Lithgow, S.J. Foster, sigmaB modulates virulence determinant expression and stress resistance: characterization of a functional rsbU strain derived from *Staphylococcus aureus* 8325-4, *J Bacteriol*, 184 (2002) 5457-5467.