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

Dhanalakshmi Nair  
*Dartmouth College*

Guido Memmi  
*Dartmouth College*

David Hernandez  
*University of Geneva*

Jonathan Bard  
*New York State Center of Excellence for Bioinformatics and Life Sciences*

Marie Beaume  
*University of Geneva*

Follow this and additional works at: [https://digitalcommons.dartmouth.edu/facoa](https://digitalcommons.dartmouth.edu/facoa)

Part of the Bacteriology Commons, Infectious Disease Commons, and the Medical Microbiology Commons

Dartmouth Digital Commons Citation  
Nair, Dhanalakshmi; Memmi, Guido; Hernandez, David; Bard, Jonathan; Beaume, Marie; Gill, Steven; Francois, Patrice; and Cheung, Ambrose L., "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" (2011). *Dartmouth Scholarship*. 1057.  
[https://digitalcommons.dartmouth.edu/facoa/1057](https://digitalcommons.dartmouth.edu/facoa/1057)

This Article is brought to you for free and open access by the Faculty Work at Dartmouth Digital Commons. It has been accepted for inclusion in Dartmouth Scholarship by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.
Authors
Dhanalakshmi Nair, Guido Memmi, David Hernandez, Jonathan Bard, Marie Beaume, Steven Gill, Patrice Francois, and Ambrose L. Cheung

This article is available at Dartmouth Digital Commons: https://digitalcommons.dartmouth.edu/facoa/1057
NOTES

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†‡

Dhanalakshmi Nair,1 Guido Memmi,1 David Hernandez,2 Jonathan Bard,3 Marie Beaume,2 Steven Gill,4 Patrice Francois,2 and Ambrose L. Cheung1*

Department of Microbiology, Dartmouth Medical School, Hanover, New Hampshire 03755; Genomic Research Laboratory, Service of Infectious Diseases, University of Geneva Hospitals, Rue Gabrielle-Perret-Gentil, CH-1211, Geneva 14, Switzerland; New York State Center of Excellence for Bioinformatics and Life Sciences, 701 Ellicott Street, Buffalo, New York 14203; and Department of Microbiology, University of Rochester, Rochester, New York 14620

Received 6 January 2011/Accepted 7 February 2011

* Corresponding author. Mailing address: Department of Microbiology, Dartmouth Medical School, Hanover, NH 03755. Phone: (603) 650-1340. Fax: (603) 650-1362. E-mail: ambrose.cheung@dartmouth.edu.
† Supplemental material for this article may be found at http://jb.asm.org/.
‡ Published ahead of print on 4 March 2011.

*Staphylococcus aureus* is a versatile pathogen that is responsible for the majority of nosocomial and community-acquired infections. One of the biggest challenges in treating staphylococcal infections is that many *S. aureus* strains have developed resistance against various antibiotics. In contrast to clinical isolates, RN4220 is a commonly used laboratory strain that is characterized by a mutation in the *sauI hsdR* gene, making it restriction deficient and hence an ideal intermediate cloning host. RN4220 was originally derived from NCTC8325-4 using UV and chemical mutagenesis (10). NCTC8325-4, in turn, was derived from an early clinical isolate, NCTC8325 (also known as PS47 or RN1), cured of three prophages (16). Indeed, RN4220 is not a suitable candidate for the study of antibiotic resistance, because newer methicillin-resistant *S. aureus* (MRSA) lineages have evolved [e.g., ST239 (hospital), ST80, and ST59 (community)], presumably due to recombination events between lineages. RN4220 is also known to harbor a small deletion in *rsbU*, a gene within the stress-induced *sigB* operon, which renders it deficient in α-expression. Additionally, RN4220 shows a Δ*agr* mutant phenotype and does not produce α-hemolysin despite producing small amounts of RNAIII in late log phase (24). Recent studies revealed that a deletion of *cvfB* (encoding conserved virulence factor B) in RN4220 resulted in diminished *agr* expression, with a reduction in *hla* expression, protease production, and virulence, in a silkworm model of systemic infection (14), but the loss of hemolytic activity in the *cvfB* mutant of RN4220 was found to be due to a defective *agr* locus and not attributable to the *cvfB* mutation (14). In another study, SrrAB, a two-component regulatory system, was found to repress the transcription of RNAIII of the *agr* locus in RN4220 (20, 27). However, the interpretation of virulence and of the associated regulatory data in these studies with RN4220 is suspect due to an inherent *agr* defect in this strain (24). Finally, O’Neill showed by comparative sequencing that NCTC8325-4, which was thought to be identical to its parent NCTC8325 except for the deletion of three prophages, possesses previously undescribed polymorphisms that may influence the virulence and pathogenicity of NCTC8325-4 (18). As a result of these issues, it is extremely important to delineate an accurate picture of the mutations in RN4220, given the polymorphisms in this strain which can have an impact upon virulence and resistance phenotype.

Using Illumina Solexa-based whole-genome sequencing (paired end) (P. Mayer, L. Farinelli, and E. Kawashima, U.S. patent application WO98/44151), we obtained the whole-genome sequence of strain RN4220 and subsequently identified the polymorphisms in RN4220 compared with the released NCTC8325 genome. Briefly, RN4220 was grown with aeration at 37°C in tryptic soy broth to log phase (optical density at 620 nm [OD₆₂₀], 0.7). Genomic DNA, isolated with a phenol-chloroform extraction method (6), was sent to Ambry Genetics (California) for library preparation and sequencing. The library preparation was carried out by shearing genomic DNA and blunting it, followed by the addition of adenine at the 3’ end. A specific adapter with bar coding was then ligated to these DNA fragments, followed by PCR amplification.
Fragments of ~76 bases were generated and assembled.

Genome assembly, single-nucleotide polymorphism (SNP) calling, and annotation were done as follows. The obtained 3.5 million paired reads, each of which is 68 nt, were de novo assembled using an Edena assembler (9), development version 3.0. The assembly has been slightly refined using the Minimus assembler (23). Final assembly resulted in 179 contigs (sum = 2.67 Mb, \( N_{50} = 80.5 \) kb \([N_{50} \text{ is the contig size such that } 50\% \text{ of the entire assembly is contained in contigs equal to or larger than that size}], \text{max} = 148 \) kb). Contigs were annotated using the RAST server (2). Comparison, SNP calling, SNP annotation, and graphical mapping were performed using the MUMmer software package (12), the CIRCOS visualization engine (11), and applications developed in-house. A map of the genome showing the polymorphisms in RN4220 is shown in Fig. 1. The entire genomic sequence of RN4220 can be found in the supplemental material.

Compared with the published NCTC8325 genome, we identified 121 SNPs and 4 large-scale deletions (see Tables S1 and S2 in the supplemental material). Among the SNPs, 14 were synonymous. The remaining SNPs involve 80 nonhomologous...
substitutions in coding regions and 27 substitutions in the intergenic regions. As anticipated, three of the four large deletions were associated with the absence of Φ11, Φ12, and Φ13 (Table S1). We also confirmed a subset of 29 nonsynonymous mutations (those that might affect virulence or metabolism) by PCR amplification followed by DNA sequencing (Table 1). This subset included nine nonsynonymous mutations that were also identified by O’Neill (18).

Besides the deletions of the three phages, there were also deletions of two hypothetical proteins and an 1,195-bp region that codes for the B subunit of excision endonuclease (also called excinuclease) ABC, which catalyzes the processing of DNA lesions by the UvrABC excinuclease complex for DNA repair (25).

<table>
<thead>
<tr>
<th>Genome position</th>
<th>Putative gene product and function</th>
<th>Nucleotide change</th>
<th>Amino acid changea</th>
<th>Locus tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>174867</td>
<td>HsdR family type I site-specific DNase</td>
<td>G</td>
<td>A W197∗</td>
<td>SAOUHSC_00162</td>
</tr>
<tr>
<td>205335</td>
<td>EssC (DNA segregation FtsK/SpoIIIE, S-DNA-T family)</td>
<td>G</td>
<td>A W52∗</td>
<td>SAOUHSC_00262</td>
</tr>
<tr>
<td>229206</td>
<td>Hypothetical protein</td>
<td>G K</td>
<td>Y</td>
<td>SAOUHSC_00274</td>
</tr>
<tr>
<td>292179</td>
<td>Hypothetical protein</td>
<td>C</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>292199</td>
<td>Hypothetical protein</td>
<td>A R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>292238</td>
<td>Hypothetical protein</td>
<td>G R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>388693</td>
<td>Conserved hypothetical protein</td>
<td>C T</td>
<td>P134S</td>
<td>SAOUHSC_00383</td>
</tr>
<tr>
<td>590042</td>
<td>Conserved hypothetical protein</td>
<td>G</td>
<td>Frameshift</td>
<td>SAOUHSC_00591</td>
</tr>
<tr>
<td>751285</td>
<td>SecA preprotein translocase subunitb</td>
<td>A T</td>
<td>E449V</td>
<td>SAOUHSC_00769</td>
</tr>
<tr>
<td>795429</td>
<td>Clumping factor CflA</td>
<td>C T</td>
<td>S815L</td>
<td>SAOUHSC_00812</td>
</tr>
<tr>
<td>827849</td>
<td>Hypothetical protein</td>
<td>A T</td>
<td>T164S</td>
<td>SAOUHSC_00859</td>
</tr>
<tr>
<td>939304</td>
<td>Competence transcription factor, putativeb</td>
<td>A T</td>
<td>E53K</td>
<td>SAOUHSC_00961</td>
</tr>
<tr>
<td>1016979</td>
<td>Spermidine/putrescine ABC transporter, putativeb</td>
<td>A G</td>
<td>E220K</td>
<td>SAOUHSC_01048</td>
</tr>
<tr>
<td>1020577</td>
<td>Manganese transport protein MntHb</td>
<td>C T</td>
<td>S286∗</td>
<td>SAOUHSC_01053</td>
</tr>
<tr>
<td>1063555</td>
<td>UvrC, excinuclease ABC subunit C</td>
<td>G A</td>
<td>G42S</td>
<td>SAOUHSC_01</td>
</tr>
<tr>
<td>1123048</td>
<td>Orotate phosphoribosyltransferaseb</td>
<td>G</td>
<td>A106T</td>
<td>SAOUHSC_01209</td>
</tr>
<tr>
<td>1160531</td>
<td>RimM, 16S rRNA-processing protein</td>
<td>A G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1358230</td>
<td>Kgd, alpha-ketoglutarate decarboxylaseb</td>
<td>A C</td>
<td>D590N</td>
<td>SAOUHSC_01418</td>
</tr>
<tr>
<td>1632629</td>
<td>(5-Methylaminomethyl-2-thiouridylate)methyltransferase</td>
<td>A</td>
<td>Frameshift</td>
<td>SAOUHSC_01726</td>
</tr>
<tr>
<td>1733572</td>
<td>Septation ring formation regulator EzrA</td>
<td>G T</td>
<td>T3N</td>
<td>SAOUHSC_01827</td>
</tr>
<tr>
<td>2087725</td>
<td>GroEL chaperone</td>
<td>C A</td>
<td>F54S</td>
<td></td>
</tr>
<tr>
<td>2096628</td>
<td>AgrA</td>
<td>A</td>
<td>F218I</td>
<td>SAOUHSC_02254</td>
</tr>
<tr>
<td>2106539</td>
<td>ABC transporter, ATP-binding protein, putative</td>
<td>A T</td>
<td>L602F</td>
<td>SAOUHSC_02274</td>
</tr>
<tr>
<td>2446162</td>
<td>Phosphotransferase system sucrose-specific IIBC component, putative</td>
<td>C G</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Mutation identified by O’Neill (18). |

TABLE 1. Nonsynonymous, PCR-verified SNPs identified in RN4220 relative to NCTC8325
cell wall-active antibiotic results in increased transcription of groEL, indicating a role for GroEL in protein folding under heat and antibiotic stress (17, 22). Mutations in these three proteins presumably would thus have an effect on the general fitness (slower growth and translational deficiency) of the strain, especially under stress. The mutation of these genes might result in an inadequate response to antibiotics, which would actually be a false-positive effect of the antibiotic in question. With regard to the virulence factors, the agrA mutation results in the insertion of an extra adenine residue at the 3' end of agrA, leading to a run of eight adenines and a frameshift that adds three amino acids to the C terminus of AgrA. This finding with agrA in RN4220, resulting in delayed activation of agr and a failure to synthesize delta and alpha hemolysins, has been described by Traber and Novick (24). The membrane protein EssC is one product of the eight-gene cluster of the ESAT-6-like secretion system (Ess), which is essential for the secretion of EssA-EssB (5). The essC mutation in RN4220 results in a truncated EssC. S. aureus mutants that fail to secrete EssA and EssB display significantly reduced virulence, dissemination, and colonization in mice (5). The S. aureus clumping factor CiaA, a surface protein belonging to the MSCRAMM family, binds the γ-chain of fibrinogen (15) and induces platelet aggregation (3); it also mediates adherence of S. aureus to fibrinogen-coated surfaces and contributes to protection against phagocytosis by neutrophils (19). The S815L mutation in CiaA in RN4220 results in the substitution of a polar hydrophilic amino acid with a nonpolar hydrophobic residue. This may conceivably disrupt the protein fold and reduce its affinity for fibrinogen. The superantigen-like protein is similar to exotoxin Set6, which is a virulence-associated protein (8).

Based on the above analysis, investigators using RN4220 in virulence studies should proceed with caution, since mutations identified in this paper show that the virulence genes as well as those involved in fitness and numerous stress-associated putative regulators are altered in RN4220.

This work was supported in part by NIH grant AI37142 to A.L.C.

REFERENCES