



# Whole-Genome Sequences of *Staphylococcus aureus* Isolates from Cystic Fibrosis Lung Infections

Eryn E. Bernardy,<sup>a,b</sup> Robert A. Petit III,<sup>c</sup> Abraham G. Moller,<sup>c</sup> Jennifer A. Blumenthal,<sup>d,e,f</sup> Alexander J. McAdam,<sup>g,h</sup> Gregory P. Priebe,<sup>d,e,f</sup> Aron T. Chande,<sup>i,j,k</sup> Lavanya Rishishwar,<sup>i,j,k</sup> I. King Jordan,<sup>i,j,k</sup> Timothy D. Read,<sup>c</sup> Joanna B. Goldberg<sup>a,b</sup>

<sup>a</sup>Department of Pediatrics, Division of Pulmonology, Allergy/Immunology, Cystic Fibrosis and Sleep, Emory University, Atlanta, Georgia, USA

<sup>b</sup>Emory-Children's Center for Cystic Fibrosis Research, Children's Healthcare of Atlanta, Atlanta, Georgia, USA

<sup>c</sup>Department of Medicine, Division of Infectious Diseases, Emory University, Atlanta, Georgia, USA

<sup>d</sup>Division of Critical Care Medicine, Department of Anesthesiology, Critical Care and Pain Medicine, Boston Children's Hospital, Boston, Massachusetts, USA

<sup>e</sup>Division of Infectious Diseases, Department of Pediatrics, Boston Children's Hospital, Boston, Massachusetts, USA

<sup>f</sup>Departments of Anaesthesia and Pediatrics, Harvard Medical School, Boston, Massachusetts, USA

<sup>g</sup>Department of Laboratory Medicine, Boston Children's Hospital, Boston, Massachusetts, USA

<sup>h</sup>Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA

<sup>i</sup>School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA

<sup>j</sup>IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, Georgia, USA

<sup>k</sup>PanAmerican Bioinformatics Institute, Cali, Valle del Cauca, Colombia

**ABSTRACT** *Staphylococcus aureus* is an early colonizer in the lungs of individuals with cystic fibrosis (CF), but surprisingly, only a limited number of genomes from CF-associated *S. aureus* isolates have been sequenced. Here, we present the whole-genome sequences of 65 *S. aureus* isolates obtained from 50 individuals with CF.

Cystic fibrosis (CF) is a genetic disease that affects over 70,000 people worldwide. The major cause of death is chronic bacterial lung infections. Early colonization is often with *Staphylococcus aureus*, and *Pseudomonas aeruginosa* emerges later as the major cause of mortality. While many researchers focus on *P. aeruginosa*, the importance of *S. aureus* in the history of CF lung infections remains understudied (1). Only a small number of whole-genome sequences of CF-associated *S. aureus* clinical isolates have been reported (1–4). We have started to fill that gap by performing whole-genome sequence analysis of 65 CF-associated *S. aureus* isolates, obtained mostly from sputum samples, after cultivation on blood agar. This collection is composed of 50 isolates (from 36 individuals) from the Emory Cystic Fibrosis Biospecimen Registry (CFBR) and 15 isolates (from 14 individuals) from Boston Children's Hospital. Metadata associated with the isolates (e.g., methicillin resistance) and the deidentified patient information (e.g., age and sex) were recorded and made available.

DNA was extracted using the Promega Wizard genomic DNA purification kit, and paired-end libraries were constructed for each isolate using the Nextera XT DNA library kit, with a fragment size of 1,000 bp, and sequenced on an Illumina MiSeq platform using version 3 chemistry. One isolate (CFBR\_EB\_Sa105, BioSample number SAMN09847825), the first sample of six longitudinal samples from a single patient, also underwent long-read genomic DNA sequencing using the Oxford Nanopore MinION sequencer. DNA was extracted in the same manner as before, and sequencing libraries were prepared using the SQK-RAD003 1D rapid sequencing kit and sequenced on a FLO-MIN106 R9.4 flow cell. The raw Illumina reads for each isolate were processed with the Staphopia analysis pipeline (Docker tag 112017, default parameters) (5). Staphopia performed sequence quality control by removing Illumina adapters and low-quality bases and reads ( $Q < 20$ ). The remaining high-quality reads for each isolate

**Citation** Bernardy EE, Petit RA, III, Moller AG, Blumenthal JA, McAdam AJ, Priebe GP, Chande AT, Rishishwar L, Jordan IK, Read TD, Goldberg JB. 2019. Whole-genome sequences of *Staphylococcus aureus* isolates from cystic fibrosis lung infections. *Microbiol Resour Announc* 8:e01564-18. <https://doi.org/10.1128/MRA.01564-18>.

**Editor** Catherine Putonti, Loyola University Chicago

**Copyright** © 2019 Bernardy et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Timothy D. Read, [tread@emory.edu](mailto:tread@emory.edu), or Joanna B. Goldberg, [joanna.goldberg@emory.edu](mailto:joanna.goldberg@emory.edu).

E.E.B. and R.A.P. contributed equally to this work.

**Received** 28 November 2018

**Accepted** 12 December 2018

**Published** 17 January 2019

**TABLE 1** Sequencing and assembly metrics for *Staphylococcus aureus* CF lung infection isolates in this study

| Sample name   | SRA accession no.      | NCBI RefSeq or GenBank assembly accession no. | Genome coverage (x) | Total length (bp) | No. of contigs | N <sub>50</sub> length (bp) | GC content (%) |
|---------------|------------------------|---|---------------------|-------------------|----------------|-----------------------------|----------------|
| CFBR_EB_Sa105 | SRX4563672, SRX4596105 | NZ_CP031779                                   | 147; 250            | 2,782,740         | 1              | 2,782,740                   | 32.78          |
| BCH-SA-01     | SRX4563649             | RIWH000000000                                 | 168                 | 2,788,868         | 43             | 266,842                     | 32.77          |
| BCH-SA-02     | SRX4563648             | RIWI000000000                                 | 173                 | 2,791,081         | 31             | 443,670                     | 32.77          |
| BCH-SA-03     | SRX4563647             | RIWJ000000000                                 | 177                 | 2,695,371         | 47             | 443,662                     | 32.71          |
| BCH-SA-04     | SRX4563646             | RIWK000000000                                 | 220                 | 2,741,344         | 31             | 711,701                     | 32.72          |
| BCH-SA-05     | SRX4563645             | RIWL000000000                                 | 172                 | 2,702,509         | 54             | 137,651                     | 32.76          |
| BCH-SA-06     | SRX4563644             | RIWM000000000                                 | 189                 | 2,743,469         | 51             | 244,024                     | 32.78          |
| BCH-SA-07     | SRX4563643             | RIWN000000000                                 | 128                 | 2,734,154         | 32             | 672,541                     | 32.71          |
| BCH-SA-08     | SRX4563642             | RIWO000000000                                 | 136                 | 2,871,626         | 72             | 150,513                     | 32.74          |
| BCH-SA-09     | SRX4563652             | RIWP000000000                                 | 137                 | 2,738,552         | 61             | 242,262                     | 32.85          |
| BCH-SA-10     | SRX4563651             | RIWQ000000000                                 | 287                 | 2,737,369         | 35             | 585,329                     | 32.68          |
| BCH-SA-11     | SRX4563667             | RIWR000000000                                 | 153                 | 2,843,748         | 80             | 150,770                     | 32.78          |
| BCH-SA-12     | SRX4563666             | RIWS000000000                                 | 179                 | 2,656,854         | 37             | 716,196                     | 32.76          |
| BCH-SA-13     | SRX4563665             | RIWT000000000                                 | 113                 | 2,816,350         | 47             | 806,706                     | 32.65          |
| BCH-SA-14     | SRX4563664             | RIWU000000000                                 | 111                 | 2,733,025         | 36             | 578,911                     | 32.71          |
| BCH-SA-15     | SRX4563671             | RIWV000000000                                 | 268                 | 2,776,591         | 32             | 605,345                     | 32.65          |
| CFBR_EB_Sa101 | SRX4563670             | RIWW000000000                                 | 263                 | 2,822,328         | 67             | 153,724                     | 32.74          |
| CFBR_EB_Sa102 | SRX4563669             | RIWX000000000                                 | 327                 | 2,819,773         | 66             | 136,645                     | 32.72          |
| CFBR_EB_Sa103 | SRX4563668             | RIWY000000000                                 | 295                 | 2,967,979         | 176            | 305,135                     | 32.59          |
| CFBR_EB_Sa104 | SRX4563673             | RIWZ000000000                                 | 124                 | 2,818,295         | 60             | 153,724                     | 32.73          |
| CFBR_EB_Sa108 | SRX4563632             | RIXA000000000                                 | 156                 | 2,788,410         | 54             | 242,088                     | 32.80          |
| CFBR_EB_Sa110 | SRX4563633             | RIXB000000000                                 | 124                 | 2,787,015         | 49             | 222,157                     | 32.79          |
| CFBR_EB_Sa112 | SRX4563630             | RIXC000000000                                 | 261                 | 2,787,692         | 49             | 222,157                     | 32.80          |
| CFBR_EB_Sa114 | SRX4563631             | RIXD000000000                                 | 172                 | 2,787,023         | 49             | 242,088                     | 32.78          |
| CFBR_EB_Sa116 | SRX4563628             | RIXE000000000                                 | 303                 | 2,851,502         | 61             | 441,461                     | 32.75          |
| CFBR_EB_Sa117 | SRX4563629             | RIXF000000000                                 | 134                 | 2,723,586         | 49             | 430,410                     | 32.75          |
| CFBR_EB_Sa118 | SRX4563626             | RIXG000000000                                 | 155                 | 2,811,738         | 49             | 441,461                     | 32.80          |
| CFBR_EB_Sa119 | SRX4563627             | RIXH000000000                                 | 146                 | 2,840,258         | 80             | 158,059                     | 32.66          |
| CFBR_EB_Sa121 | SRX4563637             | RIXI000000000                                 | 152                 | 2,861,087         | 47             | 378,754                     | 32.62          |
| CFBR_EB_Sa122 | SRX4563638             | RIXJ000000000                                 | 130                 | 2,866,137         | 148            | 142,136                     | 32.72          |
| CFBR_EB_Sa123 | SRX4563641             | RIXK000000000                                 | 133                 | 2,748,775         | 50             | 222,157                     | 32.76          |
| CFBR_EB_Sa125 | SRX4563640             | RIXL000000000                                 | 264                 | 2,902,592         | 68             | 209,697                     | 32.70          |
| CFBR_EB_Sa126 | SRX4563650             | RIXM000000000                                 | 111                 | 2,793,970         | 44             | 603,205                     | 32.63          |
| CFBR_EB_Sa127 | SRX4563661             | RIXN000000000                                 | 126                 | 2,936,734         | 68             | 345,554                     | 32.65          |
| CFBR_EB_Sa129 | SRX4563677             | RIXO000000000                                 | 166                 | 2,881,105         | 38             | 403,760                     | 32.64          |
| CFBR_EB_Sa130 | SRX4563634             | RIXP000000000                                 | 148                 | 2,879,733         | 41             | 605,839                     | 32.63          |
| CFBR_EB_Sa131 | SRX4563639             | RIXQ000000000                                 | 136                 | 2,769,595         | 42             | 393,217                     | 32.84          |
| CFBR_EB_Sa133 | SRX4563636             | RIXR000000000                                 | 111                 | 2,880,683         | 40             | 605,719                     | 32.63          |
| CFBR_EB_Sa135 | SRX4563675             | RIXS000000000                                 | 149                 | 2,809,235         | 45             | 380,091                     | 32.66          |
| CFBR_EB_Sa138 | SRX4563674             | RIXT000000000                                 | 279                 | 2,813,106         | 35             | 592,726                     | 32.68          |
| CFBRsa03      | SRX4563678             | RIXU000000000                                 | 326                 | 2,714,903         | 43             | 206,745                     | 32.75          |
| CFBRsa04      | SRX4563679             | RIXV000000000                                 | 272                 | 2,790,920         | 49             | 161,318                     | 32.77          |
| CFBRsa05      | SRX4563680             | RIXW000000000                                 | 147                 | 2,791,679         | 37             | 574,969                     | 32.63          |
| CFBRsa06      | SRX4563681             | RIXX000000000                                 | 122                 | 2,898,057         | 83             | 174,877                     | 32.65          |
| CFBRsa07      | SRX4563682             | RIXY000000000                                 | 320                 | 2,817,443         | 62             | 351,360                     | 32.85          |
| CFBRsa21      | SRX4563683             | RIXZ000000000                                 | 170                 | 2,784,304         | 37             | 434,050                     | 32.70          |
| CFBRsa22      | SRX4563684             | RIYA000000000                                 | 111                 | 2,927,132         | 117            | 221,278                     | 32.82          |
| CFBRsa23      | SRX4563685             | RIYB000000000                                 | 138                 | 2,748,954         | 42             | 443,595                     | 32.77          |
| CFBRsa24      | SRX4563686             | RIYC000000000                                 | 119                 | 2,790,775         | 43             | 443,500                     | 32.77          |
| CFBRsa25      | SRX4563687             | RIYD000000000                                 | 167                 | 2,828,883         | 46             | 441,089                     | 32.79          |
| CFBRsa26      | SRX4563660             | RIYE000000000                                 | 243                 | 2,822,821         | 71             | 136,638                     | 32.74          |
| CFBRsa27      | SRX4563659             | RIYF000000000                                 | 113                 | 2,751,330         | 62             | 122,702                     | 32.77          |
| CFBRsa28      | SRX4563658             | RIYG000000000                                 | 303                 | 2,761,455         | 33             | 587,976                     | 32.78          |
| CFBRsa29      | SRX4563657             | RIYH000000000                                 | 258                 | 2,926,156         | 52             | 379,601                     | 32.66          |
| CFBRsa30      | SRX4563656             | RIYI000000000                                 | 140                 | 2,741,496         | 40             | 288,439                     | 32.76          |
| CFBRsa47      | SRX4563655             | RIYJ000000000                                 | 123                 | 2,787,985         | 49             | 242,088                     | 32.79          |
| CFBRsa48      | SRX4563654             | RIYK000000000                                 | 138                 | 2,702,434         | 39             | 304,792                     | 32.76          |
| CFBRsa49      | SRX4563653             | RIYL000000000                                 | 147                 | 2,895,019         | 46             | 302,780                     | 32.64          |
| CFBRsa50      | SRX4563663             | RIYM000000000                                 | 313                 | 2,753,672         | 37             | 500,538                     | 32.81          |
| CFBRsa51      | SRX4563662             | RIYN000000000                                 | 267                 | 2,662,831         | 40             | 262,319                     | 32.76          |
| CFBRsa66A     | SRX4563688             | RIYO000000000                                 | 395                 | 2,783,493         | 37             | 443,669                     | 32.80          |
| CFBRsa66B     | SRX4563689             | RIYP000000000                                 | 114                 | 2,782,586         | 35             | 443,669                     | 32.80          |
| CFBRsa70      | SRX4563676             | RIYQ000000000                                 | 244                 | 2,691,630         | 42             | 500,528                     | 32.72          |
| CFBRsa74      | SRX4563635             | RIYR000000000                                 | 298                 | 2,823,147         | 43             | 321,988                     | 32.81          |
| JE2           | SRX4563690             | RIYS000000000                                 | 319                 | 2,863,521         | 66             | 606,166                     | 32.76          |

were then subsampled to 100× coverage by Staphopia and assembled using the *de novo* assembler SPAdes (version 3.11.1, default parameters) (6). Illumina reads and Nanopore reads for isolate CFBR\_EB\_Sa105 were assembled using the hybrid assembler Unicycler (version 0.4.0, default parameters) (7).

CFBR\_EB\_Sa105 was sequenced to a total 397× genome coverage (Nanopore, 250×; Illumina, 147×) and assembled into a single 2,782,740-bp-long contig. Genome coverage for the remaining 64 isolates with only Illumina sequencing ranged from 111× to 395×, with an average of 192× coverage, which is sufficient to produce a reliable draft assembly. The  $N_{50}$  values ranged from 122,702 bp to 806,706 bp, with an average  $N_{50}$  value of 370,074 bp. The assembled genome sizes ranged from 2,656,854 bp to 2,967,979 bp, with an average size of 2,799,404 bp. The GC contents for these assemblies ranged from 32.59% to 32.85%, with an average GC content of 32.74%. The genome size and GC content values are consistent with what is expected for *S. aureus*. All assemblies were annotated with the Prokaryotic Genome Annotation Pipeline available from the National Center for Biotechnology Information (NCBI) (8, 9).

Genome sequence analysis along with the metadata of these *S. aureus* isolates can be used to understand phenotypic adaptations of this pathogen that are required for survival within the multispecies community of the CF lung.

**Data availability.** The complete assembly for CFBR\_EB\_Sa105 and draft assemblies for the remaining samples, as well as the raw Illumina reads, have been deposited in NCBI and are available under BioProject accession number [PRJNA480016](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA480016). Information for each isolate, including accession numbers, is shown in Table 1.

## ACKNOWLEDGMENTS

This work was supported in part by a Pediatric Research Alliance Pilot Project (grant 00068914 to J.B.G. and T.D.R.) from Cystic Fibrosis and Airways Disease (CF-AIR) and Children's Healthcare of Atlanta. A.G.M. was supported by the National Science Foundation (NSF) Graduate Research Fellowship Program (GRFP). Bacterial isolates were obtained from the Cystic Fibrosis Biospecimen Registry, which is supported in part by the CF Discovery Core of the CF@LANTA RDP Center, and by the Center for CF and Airways Disease Research, components of the Emory+Children's CF Center of Excellence at Emory University and Children's Healthcare of Atlanta. We also thank the NIH IRACDA Fellowships in Research and Science Teaching program at Emory for additional financial support (project number 5K12GM000680-19). This work was also supported in part by the Richard A. and Susan F. Smith President's Innovation Award (to G.P.P.) and by funds from the Translational Research for Infection Prevention in Pediatric Anesthesia and Critical Care (TRIPPACC) Program of the Department of Anesthesiology, Critical Care and Pain Medicine at Boston Children's Hospital (to G.P.P.).

## REFERENCES

- McAdam PR, Holmes A, Templeton KE, Fitzgerald JR. 2011. Adaptive evolution of *Staphylococcus aureus* during chronic endobronchial infection of a cystic fibrosis patient. *PLoS One* 6:e24301. <https://doi.org/10.1371/journal.pone.0024301>.
- Ankrum A, Hall BG. 2017. Population dynamics of *Staphylococcus aureus* in cystic fibrosis patients to determine transmission events by use of whole-genome sequencing. *J Clin Microbiol* 55:2143–2152. <https://doi.org/10.1128/JCM.00164-17>.
- Lima DF, Cohen RW, Rocha GA, Albano RM, Marques EA, Leao RS. 2017. Genomic information on multidrug-resistant livestock-associated methicillin-resistant *Staphylococcus aureus* ST398 isolated from a Brazilian patient with cystic fibrosis. *Mem Inst Oswaldo Cruz* 112:79–80. <https://doi.org/10.1590/0074-02760160342>.
- Tan X, Coureuil M, Ramond E, Euphrasie D, Dupuis M, Tros F, Meyer J, Nemanzny I, Chhuon C, Guerrero IC, Ferroni A, Sermet-Gaudelus I, Nassif X, Charbit A, Jamet A. 2018. Chronic *Staphylococcus aureus* lung infection correlates with proteogenomic and metabolic adaptations leading to an increased intracellular persistence. *bioRxiv* <https://doi.org/10.1101/414409>.
- Petit RA, III, Read TD. 2018. *Staphylococcus aureus* viewed from the perspective of 40,000+ genomes. *PeerJ* 6:e5261. <https://doi.org/10.7717/peerj.5261>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
- Tatsova T, DiCuccio M, Badretdin A, Chetverin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.
- Haft DH, DiCuccio M, Badretdin A, Brover V, Chetverin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* 46:D851–D860. <https://doi.org/10.1093/nar/gkx1068>.