AMERICAN SOCIETY FOR MICROBIOLOGY

Microbiology® Resource Announcements

Draft Genome Sequences of 62 *Staphylococcus aureus* Isolates Associated with Four Foodborne Outbreaks in the United States

Jennifer Hait,[a] Guojie Cao,[a] Maria Balkey,[a] George Kastanis,[a] Phillip Curry,[a] Lanlan Yin,[b] James Pettengill,[b] Sandra Tallent[a]

aCenter for Food Safety and Applied Nutrition, Office of Regulatory Science, Division of Microbiology, U.S. Food and Drug Administration, College Park, Maryland, USA
bCenter for Food Safety and Applied Nutrition, Office of Analytics and Outreach, U.S. Food and Drug Administration, College Park, Maryland, USA

Jennifer Hait and Guojie Cao contributed equally to the manuscript. Jennifer Hait is the lead of this project.

ABSTRACT *Staphylococcus aureus* bacteria are ranked among the top five foodborne pathogens in the United States. Here, we report the draft genome sequences of 62 *S. aureus* isolates that originated from the manufacturing environment of an Illinois bakery and were associated with outbreaks between 2010 and 2011 in the United States.

Staphylococci are ubiquitous, and *Staphylococcus aureus* bacteria are commonly found in foods due to environmental, human, and animal contamination (1). According to the Centers for Disease Control and Prevention, staphylococcal foodborne illness has been ranked among the top five foodborne pathogens in the United States, resulting in more than 240,000 individual cases annually (2). Although foodborne illness associated with staphylococcal enterotoxins commonly has self-limiting symptoms, including nausea, abdominal cramping, diarrhea, and vomiting (1), it is also the cause of superficial infections and life-threatening diseases (3). *S. aureus* was associated with outbreaks when high pathogen doses were found in the food worker population (4). Importantly, *S. aureus* can asymptomatically colonize in the throat and nasal cavity of food workers, allowing frequent contamination to hands and arms (4). Foods commonly associated with staphylococcal food poisoning include cream-filled bakery products, dairy products, salads, puddings, pastries, sandwiches, and other ready-to-eat (RTE) foods, which are contaminated during preparation in homes or food-service establishments. Although *S. aureus* can be killed by the heating process in cooking, their enterotoxins are not destroyed and will be able to cause diseases. The 62 selected outbreak-associated *S. aureus* isolates originated from the manufacturing environment of an Illinois bakery, which were implicated in four U.S. outbreaks in 2010 to 2011 resulting in more than 100 individuals reporting illness (Table 1) (5). The detailed analysis of enterotoxin gene contents of the selected isolates was discussed (1, 5). These individuals were sickened after ingesting an assortment of desserts linked to the products manufactured in this bakery. The isolates sequenced in the current study will provide genomic information for implicated outbreaks and will contribute to a better understanding of the genomic diversity of *S. aureus* in the United States.

The *S. aureus* isolates are maintained in the strain collection of the FDA Center for Food Safety and Applied Nutrition. These isolates were confirmed phenotypically using Baird-Parker with rabbit plasma fibrinogen (RPF) agar as the selective plating substrate (item number 43531; bioMérieux, Marcy-l'Etoile/France) and biochemical testing using the Vitek 2 Gram-positive identification card (bioMérieux). Genomic DNA was extracted after following a 16-hour culture incubation at 37°C in Trypticase soy broth (Becton, Dickinson, Franklin Lakes, NJ) using the DNeasy blood and tissue kit (Qiagen, Inc., Valencia, CA). DNA concentration was measured using a Qubit 3.0 fluorometer (Life Technologies, MD). Libraries were prepared according to Nextera XT or Nextera Flex

TABLE 1 Summary characteristics of whole-genome sequencing of *Staphylococcus aureus* isolates[a]

Name	Yr	Genome size (bp)	GC content (%)	No. of genes	No. of RNA genes	No. of reads	Coverage (x)	N_{50} (bp)	No. of contigs	SRA accession no.	GenBank accession no.
CFSAN007820	2011	2,661,394	32.74	2,509	61	2,383,434	108	423,141	16	SRR12980320	DACXVD010000000
CFSAN007821	2011	2,661,540	32.74	2,506	62	1,660,544	86	423,085	15	SRR12981341	DACXWL010000000
CFSAN007822	2011	2,661,109	32.74	2,507	61	1,703,756	87	423,141	16	SRR12980429	DACXVG010000000
CFSAN007823	2011	2,740,527	32.66	2,565	61	991,248	48	226,122	23	SRR12979474	DACXVC010000000
CFSAN007824	2010	2,639,316	32.77	2,478	57	3,028,672	109	411,661	19	SRR12980427	DACYBA010000000
CFSAN007825	2010	2,639,946	32.77	2,480	61	2,151,232	84	423,140	15	SRR12979362	DACXVS000000000
CFSAN007826	2010	2,639,476	32.77	2,475	57	1,919,904	70	423,184	20	SRR12980213	DACXWH010000000
CFSAN007827	2010	2,640,521	32.77	2,481	62	974,448	50	423,142	16	SRR12981014	DACXXL010000000
CFSAN007828	2010	2,640,452	32.77	2,480	61	1,525,036	79	423,140	14	SRR12980215	DACXVH010000000
CFSAN007829	2010	2,639,847	32.77	2,483	61	685,280	36	423,142	17	SRR12980837	DACXUZ010000000
CFSAN007830	2010	2,640,061	32.77	2,481	60	740,832	38	423,140	15	SRR12981340	DACXUY010000000
CFSAN007832	2010	2,639,963	32.77	2,483	61	778,256	40	423,139	16	SRR12979398	DACXUW010000000
CFSAN007833	2010	2,640,323	32.77	2,480	61	1,129,168	59	423,142	15	SRR12980634	DACXVE010000000
CFSAN007834	2010	2,641,320	32.78	2,419	63	1,965,680	102	423,142	14	SRR12979708	DACXWR010000000
CFSAN007836	2010	2,641,420	32.78	2,419	60	1,696,800	88	423,142	16	SRR12981432	DACXUX010000000
CFSAN007837	2010	2,645,111	32.81	2,419	65	1,715,808	88	423,140	17	SRR12981213	DACXVY010000000
CFSAN007838	2011	2,801,714	32.75	2,627	65	1,246,640	65	174,711	40	SRR12980896	DACXVB010000000
CFSAN007839	2010	2,754,596	32.69	2,562	63	1,448,480	75	174,710	36	SRR12981261	DACXWF010000000
CFSAN007841	2011	2,797,508	32.72	2,629	62	1,700,064	88	174,710	38	SRR12981214	DACXVV010000000
CFSAN007848	2010	2,769,241	32.69	2,610	60	1,145,248	60	225,596	32	SRR12980212	DACXVJ010000000
CFSAN007849	2010	2,773,128	32.70	2,559	67	1,365,536	71	325,333	24	SRR12980118	DACXWJ010000000
CFSAN007852	2011	2,738,393	32.66	2,561	61	2,989,040	132	116,016	44	SRR12979463	DACXXM010000000
CFSAN007853	2011	2,723,124	32.72	2,534	59	2,081,102	94	41,512	138	SRR12979401	DACXWT010000000
CFSAN007854	2011	2,732,449	32.69	2,553	61	3,517,628	153	52,793	102	SRR12979400	DACXWK010000000
CFSAN007855	2011	2,728,363	32.69	2,543	62	2,182,078	101	56,037	89	SRR12981431	DACXVU010000000
CFSAN007856	2011	2,659,802	32.75	2,503	61	2,037,890	96	388,909	29	SRR12980839	DACXWM010000000
CFSAN007857	2011	2,734,626	32.66	2,558	62	2,296,136	107	114,046	36	SRR12979368	DACXWV010000000
CFSAN007858	2011	2,742,336	32.66	2,566	61	1,180,636	61	314,472	17	SRR12980105	DACXWN010000000
CFSAN007859	2011	2,742,189	32.66	2,567	61	1,338,120	69	314,472	18	SRR12980104	DACXVK010000000
CFSAN007860	2011	2,741,845	32.66	2,567	61	1,442,512	76	314,472	18	SRR12980101	DACXVQ010000000
CFSAN007861	2011	2,742,195	32.66	2,565	62	1,280,800	67	314,472	19	SRR12980218	DACXUV010000000
CFSAN007862	2011	2,743,173	32.65	2,558	61	1,559,386	82	314472	19	SRR12980226	DACXWG010000000
CFSAN007863	2011	2,737,381	32.66	2,559	61	1,582,678	83	314,472	19	SRR12980428	DACXWO010000000
CFSAN007867	2010	2,737,639	32.66	2,559	61	1,500,732	79	314,472	18	SRR12979475	DACYFK010000000
CFSAN007868	2010	2,738,091	32.66	2,560	61	1,399,056	74	314,472	18	SRR12979885	DACXWQ010000000
CFSAN007869	2010	2,737,894	32.66	2,560	61	1,434,126	74	314,472	16	SRR12981263	DACXWW010000000
CFSAN007870	2010	2,737,426	32.66	2,559	60	1,393,334	72	314,472	18	SRR12981013	DACXZS010000000
CFSAN007872	2010	2,771,090	32.60	2,601	61	1,923,118	100	314,472	18	SRR12980227	DACXWP010000000
CFSAN007873	2010	2,738,000	32.66	2,559	62	1,186,244	62	359,244	17	SRR12979396	DACXWE010000000
CFSAN007874	2010	2,766,466	32.60	2,597	61	1,539,540	81	314,472	17	SRR12980638	DACXWA010000000
CFSAN007875	2010	2,716,949	32.69	2,536	61	2,002,936	105	314,472	17	SRR12979886	DACXXW010000000
CFSAN007876	2010	2,738,108	32.66	2,560	61	2,400,532	126	314,472	16	SRR12979466	DACXVP010000000
CFSAN007877	2010	2,737,702	32.66	2,560	61	2,201,998	115	314,472	17	SRR12979404	DACXVR010000000

TABLE 1 (Continued)

Name	Yr	Genome size (bp)	GC content (%)	No. of genes	No. of RNA genes	No. of reads	Coverage (×)	N_{50} (bp)	No. of contigs	SRA accession no.	GenBank accession no.
CFSAN007878	2010	2,737,595	32.66	2,560	61	2,193,328	114	314,472	18	SRR12979390	DACXWD010000000
CFSAN007879	2010	2,737,921	32.66	2,560	61	1,356,376	70	314,472	17	SRR12979361	DACXVQ010000000
CFSAN007880	2010	2,738,112	32.66	2,560	61	2,434,346	126	314,472	16	SRR12980631	DACXVF010000000
CFSAN007881	2010	2,766,326	32.60	2,595	61	1,529,818	79	314,472	17	SRR12980217	DACXWI010000000
CFSAN007882	2010	2,736,721	32.66	2,559	61	2,538,852	130	238,040	25	SRR12980661	DACXWU010000000
CFSAN007884	2011	2,666,118	32.68	2,489	62	1,658,094	87	151,508	29	SRR12980639	DACXVZ010000000
CFSAN007886	2011	2,666,272	32.68	2,489	62	2,849,608	139	151,508	29	SRR12981264	DACXWN010000000
CFSAN007887	2011	2,699,550	32.65	2,509	62	987,038	52	314,472	16	SRR12980632	DACYAZ010000000
CFSAN007888	2011	2,698,922	32.66	2,508	61	2,317,998	119	146,458	22	SRR12980633	DACXVO010000000
CFSAN007889	2011	2,699,909	32.65	2,508	61	1,739,824	89	226,737	17	SRR12980225	DACXVI010000000
CFSAN007892	2011	2,699,475	32.65	2,508	61	1,053,570	55	226,737	17	SRR12980102	DACXVN010000000
CFSAN007893	2011	2,693,689	32.65	2,501	61	4,272,278	197	126,918	20	SRR12980630	DACXVX010000000
CFSAN007895	2011	2,723,594	32.69	2,501	63	1,913,008	100	689,169	21	SRR12980166	DACXWB010000000
CFSAN007897	2011	2,698,606	32.75	2,567	60	1,017,654	52	388,899	18	SRR12979402	DACXWC010000000
CFSAN007898	2011	2,697,014	32.75	2,561	55	465,984	23	162,289	27	SRR12979399	DACXVT010000000
CFSAN007899	2011	2,699,743	32.75	2,507	62	1,657,216	86	388,899	16	SRR12980167	DACXVL010000000
CFSAN007901	2011	2,754,859	32.81	2,655	61	2,958,206	123	235,381	34	SRR12980214	DACXWS010000000
CFSAN007902	2011	2,693,749	32.76	2,554	62	2,186,776	93	192,582	26	SRR12981215	DACXVA010000000
CFSAN007903	2011	2,693,955	32.76	2,556	62	2,736,044	119	441,727	22	SRR12980921	DACXVW010000000

[a]All strains were isolated by environmental swab in Illinois.

protocols and sequenced on the MiSeq platform (Illumina, San Diego, CA) using MiSeq reagent kit v2 (500 cycles) or v3 (600 cycles). Paired-end read quality was assessed by the following parameters: cluster density of 1,200 to 1,400 K/mm^2 and >80% clusters passing filters. Raw reads were trimmed using Trimmomatic with default parameters (6) and assembled *de novo* using SKESA v2.2 (7) with default settings, and the minimal contig length was reported in output as 500 bp. QUAST was used to assess the quality of each assembly with default settings (Table 1) (8). Annotations of assemblies were processed using Prokka (Galaxy v1.14.5) (9). Assemblies, which were processed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (10) were subsequently deposited at DDBJ/EMBL/GenBank.

Data availability. The genome sequences of the 62 *S. aureus* isolates were deposited in DDBJ/ENA/GenBank, and detailed information is listed in Table 1.

REFERENCES

1. Hait JM, Tallent SM, Bennett RW. 2014. Screening, detection, and serotyping methods for toxin genes and enterotoxins in Staphylococcus strains. J AOAC Int 97:1078–1083. https://doi.org/10.5740/jaoacint.13-211.
2. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM. 2011. Foodborne illness acquired in the United States—major pathogens. Emerg Infect Dis 17:7–15. https://doi.org/10.3201/eid1701.p11101.
3. Mason A, Foster D, Bradley P, Golubchik T, Doumith M, Gordon NC, Pichon B, Iqbal Z, Staves P, Crook D, Walker AS, Kearns A, Peto T. 2018. Accuracy of different bioinformatics methods in detecting antibiotic resistance and virulence factors from Staphylococcus aureus whole-genome sequences. J Clin Microbiol 56:e01815-17. https://doi.org/10.1128/JCM.01815-17.
4. Todd EC, Greig JD, Bartleson CA, Michaels BS. 2008. Outbreaks where food workers have been implicated in the spread of foodborne disease. Part 4. Infective doses and pathogen carriage. J Food Prot 71:2339–2373. https://doi.org/10.4315/0362-028x-71.11.2339.
5. Hait JM, Tallent SM, Melka D, Keys C, Bennett R. 2012. Staphylococcus aureus outbreak investigation of an Illinois bakery. J Food Saf 32:435–444. https://doi.org/10.1111/jfs.12002.
6. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
7. Souvorov A, Agarwala R, Lipman DJ. 2018. SKESA: strategic k-mer extension for scrupulous assemblies. Genome Biol 19:153. https://doi.org/10.1186/s13059-018-1540-z.
8. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
9. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
10. Tatusova T, DiCuccio M, Badretdin A, Chetvernin 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.