



GENOME ANNOUNCEMENT

Open Access

Genome sequence of type strain of *Staphylococcus aureus* subsp. *aureus*

Bong-Soo Kim¹, Hana Yi^{2,3,4}, Jongsik Chun^{1,5} and Chang-Jun Cha^{6*}

Abstract

Background: *Staphylococcus aureus* is a pathogen that causes food poisoning and community-associated infection with antibiotic resistance. This species is an indigenous intestinal microbe found in infants and not found in adult intestine. The relatively small genome size and rapid evolution of antibiotic resistance genes in the species have been drawing an increasing attention in public health. To extend our understanding of the species and use the genome data for comparative genomic studies, we sequenced the type strain of *S. aureus* subsp. *aureus* DSM 20231^T.

Results: Seventeen contigs were generated using hybrid assembly of sequences derived from the Roche 454 and Illumina systems. The length of the genome sequence was 2,902,619 bases with a G + C content of 32.8%. Among the 2,550 annotated CDSs, 44 CDSs were annotated to antibiotic resistance genes and 13 CDSs were related to methicillin resistance. It is interesting to note that this strain was first isolated in 1884 before methicillin was generally used on patients.

Conclusions: The present study analyzed the genome sequence of *S. aureus* subsp. *aureus* type strain as the reference sequence for comparative genomic analyses of clinical isolates. Methicillin resistance genes found in the genome indicate the presence of antibiotic resistance mechanism prior to the usage of antibiotics. Further comparative genomic studies of methicillin-resistant strains based on this reference genome would help to understand the evolution of resistance mechanism and dissemination of resistance genes.

Keywords: *Staphylococcus aureus* subsp. *aureus*, Genome sequencing, Type strain, Hybrid assembly

Background

Staphylococcus aureus is a member of normal microbiota in human body and also known as an opportunistic pathogen. This species can cause a broad range of nosocomial and community-associated infections, and the antibiotic resistance of the species has been studied for many years [1]. *S. aureus* was also reported as the predominant species in infant feces, and decreased toward adulthood due to the colonization of complex gut microbiota [2,3]. The species can spread through skin-to-skin contact with colonized individuals, and cause a global epidemic as antibiotic resistant strains [4]. Foodborne illness can be caused by enterotoxin-producing *S. aureus* with symptoms such as diarrhea, nausea and abdominal cramps [5]. Recently, *S. aureus* was detected in irritable bowel syndrome (IBS) subjects [6].

Many strains of *S. aureus* subsp. *aureus* were genome-sequenced and submitted to public databases due to the importance in antibiotic resistance and the possibility of nosocomial infections even in health care and community settings [7-9]. However, type strain of this species has not been genome-sequenced yet. Type strain is usually the firstly isolated strain of the species, and exhibits all of the relevant phenotypic and genotypic properties cited in the species circumscriptions. Therefore, the genome sequence of type strain is important to analyze the phenotypic and genotypic characteristics of species. In the present study, we analyzed the whole genome sequence of *S. aureus* subsp. *aureus* type strain as the standard reference genome required for *S. aureus* studies.

Methods

Strain information

Type strain of *S. aureus* subsp. *aureus* (DSM 20231^T) was obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ; Barunshweig, Germany).

* Correspondence: cjcha@cau.ac.kr

⁶Department of Systems Biotechnology, Chung-Ang University, Anseong, Republic of Korea

Full list of author information is available at the end of the article

The strain was known to be non-motile, non-spore-forming, Gram-positive cocci (0.5-1.0 μm in diameter), facultatively anaerobic and producing enterotoxin. Optimal growth is observed at 30-37°C on trypticase soy yeast extract media containing 10% NaCl [10].

Genomic DNA extraction

Genomic DNA was extracted using a Wizard Genomic DNA Isolation kit (Promega, Madison, WI, USA). The concentration of extracted DNA was quantified using a PicoGreen dsDNA Assay kit (Invitrogen, Carlsbad, CA, USA), and the contamination of DNA or cultured strain was checked by sequencing the 16S rRNA gene using the ABI 3730 DNA sequencing machine (Applied Biosystems, Foster City, CA, USA).

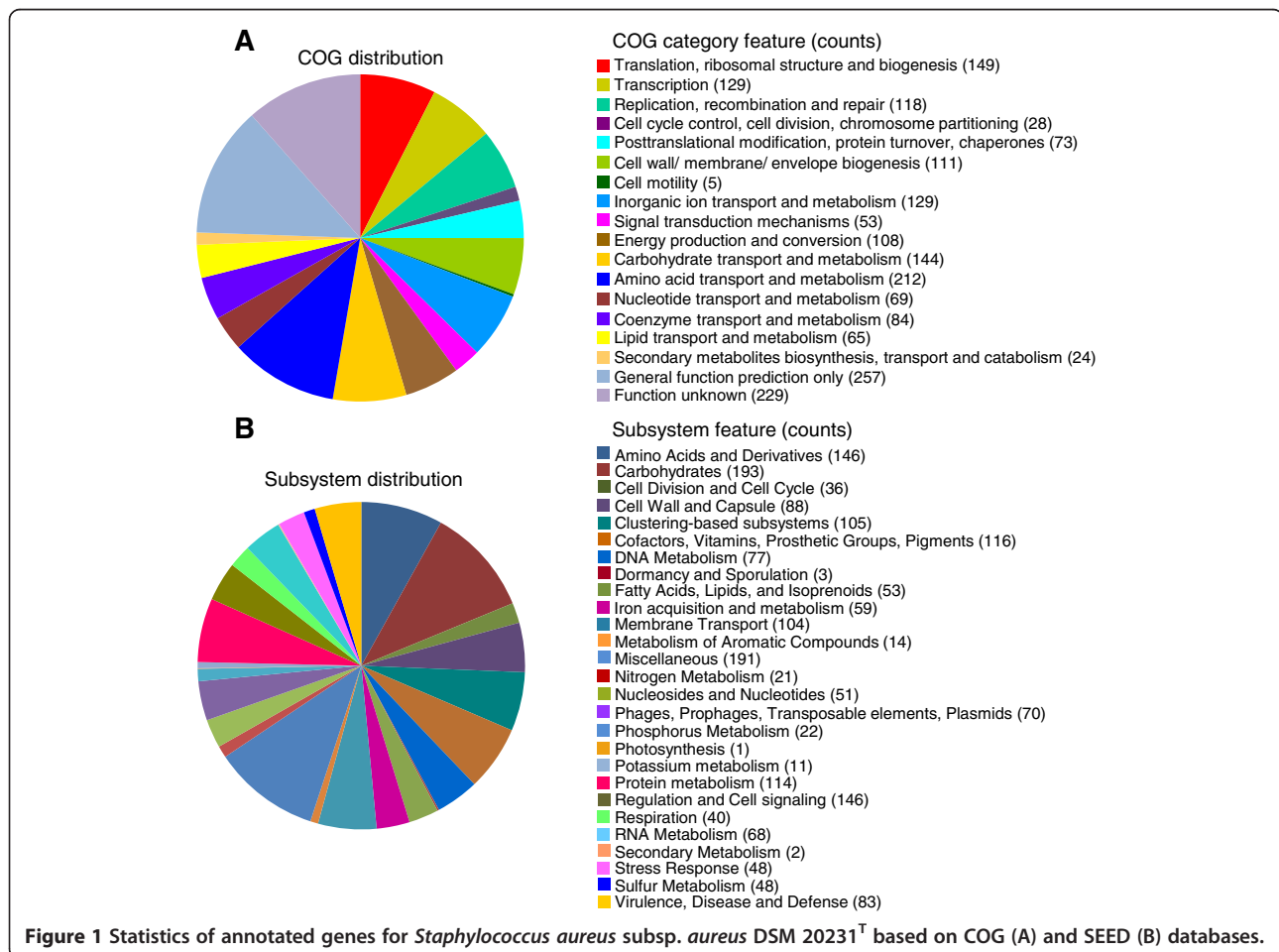
Whole genome sequencing

The draft genome sequence of strain DSM 20231^T was determined by a combination of Illumina Genome Analyzer IIx (150 bp paired end) and Roche 454 (8-kb insert paired end) sequencing systems. The sequencing library was prepared with the TruSeq DNA LT Sample Prep kit (Illumina, San Diego, CA, USA) for the Illumina

system, and the library for the Roche 454 system was prepared using the GS FLX Titanium Rapid Library Preparation kit (Roche Diagnostics, Branford, CT, USA).

Assembly and annotation of genome sequence

Sequencing reads obtained from the Illumina system were assembled using the CLC genomic workbench 5.5 (CLC Bio, Denmark), and the reads obtained from the Roche 454 sequencing system were assembled using the GS Assembler 2.6 (Roche Diagnostics). The assembled contigs from each sequencing system were corrected in their order using the published reference genomes. Hybrid assembly of contigs generated by both systems was conducted using the CodonCode Aligner (CodonCode Co. MA, USA). In brief, the contigs generated by each sequencing system were reassembled together using the CodonCode Aligner. Reassembly of hybrid contigs and unassembled contigs were repeated until the number of hybrid contigs did not change. Contigs of short length (<500 bp) were removed from the hybrid result file. Prediction of genes was performed using Glimmer 3 [11], and annotation was conducted by homology search against the Clusters of Orthologous Groups (COG)



and SEED database [12,13]. Prediction of multilocus sequencing typing (MLST) was performed using assembled contigs [14].

Comparative genomics

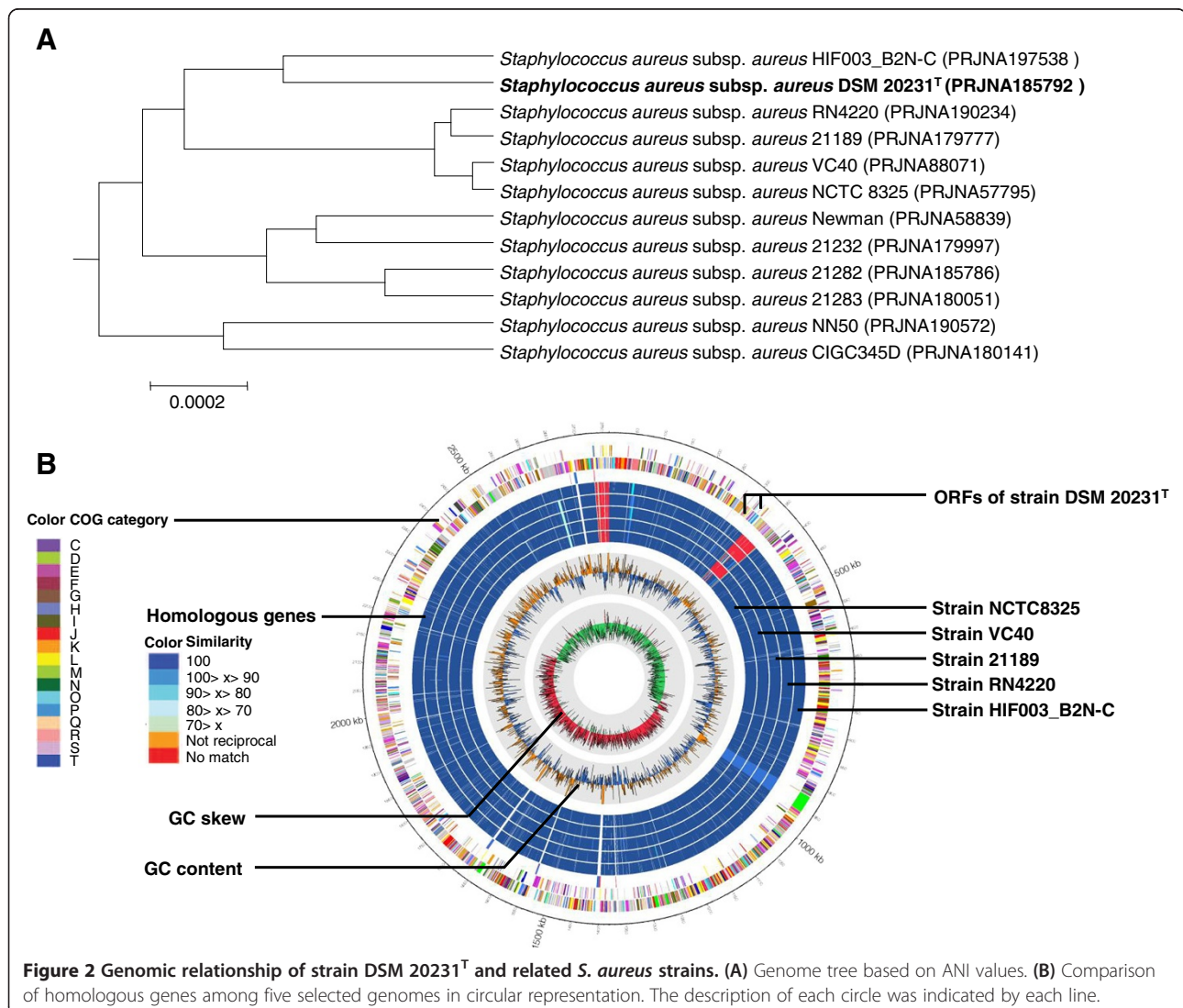
A total of 441 genome sequences of strains that belong to *S. aureus* subsp. *aureus* were obtained from EzGenome database (<http://ezgenome.ezbiocloud.net>), and used to calculate average nucleotide identity (ANI) values [15] with strain DSM 20231^T. For ANI calculation, the query genome was cut into small fragments (1020 bp), and high-scoring pairs between two genome sequences were selected by BLAST algorithm [16]. Then, a dendrogram was constructed using calculated ANI values by the unweighted pair group method.

Five genome sequences (an ANI value of > 99.78% with strain DSM 20231^T) were selected as the closest genomes and compared with strain DSM 20231^T by

using comparative genomic method as described previously [17]. Briefly, homologous regions in a target genome to query ORFs were determined using the BLASTN program, and aligned using a pairwise global alignment. The matched region in the subject contig was extracted and saved as a homolog.

Quality assurance

A potential contamination was evaluated by identification of 16S rRNA gene amplified from extracted DNA before the whole genome sequencing and by comparison of 16S rRNA gene in draft genome after sequencing. 16S rRNA genes in assembled contigs were found using the rRNA selector [18] and identified using the EzTaxon-e database [19]. Bioinformatic assembly was checked by a comparison of the obtained genome sequence with published genomes of the same species using ANI values [15].



Initial findings

A total of 17 contigs (N50 = 313,118 bp) were generated from a hybrid assembly of reads from Illumina (6,413,077 reads of 150 bp paired end; > 350-fold coverage) and Roche 454 (240,863 reads of 8Kb-insert paired end; > 14-fold coverage) systems. The genome size of strain DSM 20231^T was 2,761,522 bases with 32.8% G + C content. The genome contained 2,550 predicted protein-coding sequences (CDSs), 57 tRNA genes and 12 rRNA genes. Results of the genome annotation are shown in Figure 1. For the COG distribution, R (General function prediction only; 257 ORFs), S (Function unknown; 211 ORFs), and E (Amino acid transport and metabolism; 212 ORFs) were abundant categories (over 10% of total COG matched counts). Genes responsible for carbohydrates (193 ORFs), miscellaneous (191 ORFs), amino acid metabolism (146 ORFs) and cell signaling (146 ORFs) were abundant among the SEED subsystem categories.

The genome tree of *S. aureus* strains was constructed by using ANI calculation (Figure 2A), and strains HIF003-B2N-C, RN4220, 21189, VC40, and NCTC 8325 were chosen based on ANI values for the comparative analysis. Strain HIF003_B2N-C was recovered as the closest genome of the sequenced genome in the genome tree. The number of different gene contents between strains of DSM 20231^T and HIF003-B2N-C was 35 ORFs, and the highest different number between them was observed in K (Transcription) and L (Replication, recombination and repair) categories. Genome sequences among selected strains for comparison were similar to each other, and most of the different ORFs were hypothetical proteins, replication-associated proteins, and transposases. Comparison of homologous genes among the selected genomes is given in Figure 2B.

In subsystem distribution of the sequenced genome, 83 genes (4.6% of total subsystem counts) were annotated

to virulence, disease and defense category, and 91.6% of genes (76 ORFs) in this category were annotated to be responsible for adhesion and antibiotic resistance. Adhesion to human intestinal mucus and antibiotic resistance of *S. aureus* are important characteristics of pathogens. The highest number of predicted protein sequences (13 CDSs) among 44 CDSs in antibiotic resistance subcategory was annotated to methicillin resistance-related genes (Table 1). In the case of the five related genomes, the numbers of these genes were much smaller than that in the type strain (3 CDSs in HIF003_B2N-C and RN4220, 2 CDSs in NCTC 8325, 1 CDS in VC40, and no hit in strain 21189). The methicillin resistance of *S. aureus* was first reported in 1961 [20]. However, strain DSM 20231^T was isolated in 1884 from human pleural fluid [21]. This implies that *S. aureus* had already possessed potential genes for methicillin resistance before methicillin was introduced in 1960. Our finding can provide an insight into history and evolution of methicillin resistance. The predicted MLST of strain DSM 20231^T and the five selected strains for comparison were all ST8 in the clonal complex (CC) 8, where the first MRSA clinical isolate is ST250 [1]. In addition, genes related to arsenic resistance, fluoroquinolone resistance, fosfomycin resistance, vancomycin resistance, multidrug resistance efflux pumps, cobalt-zinc-cadmium resistance, and copper homeostasis were annotated in the genome sequence of strain DSM 20231^T. The presence of antibiotic resistance genes in the genome of this strain implies that antibiotic resistances of this species have evolved for a long time before synthetic antibiotics were used.

Future directions

The genome sequence of *S. aureus* subsp. *aureus* type strain can be used as a standard reference genome

Table 1 Summary of CDSs annotated to methicillin resistance

Contig number	Length (bp)	Seed subsystem	Seed function
2	1,194	Methicillin_resistance_in_Staphylococci	FmtA protein involved in methicillin resistance
2	1,245	Methicillin_resistance_in_Staphylococci	FmhC protein of FemAB family
2	369	Methicillin_resistance_in_Staphylococci	FemC factor involved in methicillin resistance
2	2,523	Methicillin_resistance_in_Staphylococci	FmtC (MrpF) protein involved in methicillin resistance
2	1,263	Methicillin_resistance_in_Staphylococci	FmhA protein of FemAB family
2	1,260	Methicillin_resistance_in_Staphylococci	FmhA protein of FemAB family
3	876	Methicillin_resistance_in_Staphylococci	LytH protein involved in methicillin resistance
6	465	Methicillin_resistance_in_Staphylococci	HTH domain protein SA1665, binds to mecA promoter region
9	1,185	Methicillin_resistance_in_Staphylococci	HmrA protein involved in methicillin resistance
9	7,437	Methicillin_resistance_in_Staphylococci	FmtB (Mrp) protein involved in methicillin resistance and cell wall biosynthesis
11	1,266	Methicillin_resistance_in_Staphylococci	tRNA-dependent lipid II-glycine ligase (FmhB)
12	1,251	Methicillin_resistance_in_Staphylococci	FmhA protein of FemAB family
16	675	Methicillin_resistance_in_Staphylococci	Transposase for insertion sequence-like element IS431mec

sequence for studying *S. aureus* strains including MRSA. Further comparative genome analyses of *S. aureus* strains will provide differences in the genomic contents found in this species and evolutionary information on resistance developments via horizontal gene transfer and mutation. These studies will also help to understand the pathogenesis of the staphylococcal diseases for infection preventions.

Availability of supporting data

The draft genome sequence of *Staphylococcus aureus* subsp. *aureus* DSM 20231^T was deposited at DDBJ/EMBL/GenBank under the accession AMYL00000000. The version described in this paper is the first version AMYL01000000.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JC and CC designed the study. BK and HY performed experiments. BK, HY and JC analyzed the sequencing data. BK, HY and CC contributed to the writing of manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by Industrial Strategic Technology Development Program (No. 10040176) funded by the Korean Ministry of Knowledge Economy (MKE), by Basic Health Science Research Programs through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, and Future Planning (NRF-2013R1A1A3010041) and by Korea Ministry of Environment as "The Environmental Health Action Program".

Author details

¹ChunLab, Inc., Seoul, Republic of Korea. ²Division of Biosystems Biomedical Science, College of Health Science, Korea University, Seoul, Republic of Korea. ³Department of Public Health Science, Graduate School, Korea University, Seoul, Republic of Korea. ⁴Korea University Guro Hospital, Korea University, Seoul, Republic of Korea. ⁵School of Biological Sciences and Bioinformatics Institute, BIO-MAX, Seoul National University, Seoul, Republic of Korea. ⁶Department of Systems Biotechnology, Chung-Ang University, Anseong, Republic of Korea.

Received: 28 January 2014 Accepted: 11 March 2014

Published: 17 March 2014

References

1. Chambers HF, Deleo FR: **Waves of resistance: *Staphylococcus aureus* in the antibiotic era.** *Nat Rev Microbiol* 2009, **7**:629–641.
2. Lindberg E, Nowrouzian F, Adlerberth I, Wold AE: **Long-time persistence of superantigen-producing *Staphylococcus aureus* strains in the intestinal microflora of healthy infants.** *Pediatr Res* 2000, **48**:741–747.
3. Bjorksten B, Sepp E, Julge K, Voor T, Mikelsaar M: **Allergy development and the intestinal microflora during the first year of life.** *J Allergy Clin Immunol* 2001, **108**:516–520.
4. Muto CA, Jernigan JA, Ostrowsky BE, Richet HM, Jarvis WR, Boyce JM, Farr BM: **SHEA guideline for preventing nosocomial transmission of multidrug-resistant strains of *Staphylococcus aureus* and *Enterococcus*.** *Infect Cont Hosp Ep* 2003, **24**:362–386.
5. Tortora GJ: ***Staphylococcus* food poisoning (Staphylococcal enterotoxigenesis).** In *Microbiology An introduction*. 1st edition. Edited by Tortora GJ, Funke BR, Case CL. San Francisco, CA, USA: Benjamin Cummings Publishing Company, Inc; 1995:616–618.
6. Rinttila T, Lyra A, Krogius-Kurikka L, Palva A: **Real-time PCR analysis of enteric pathogens from fecal samples of irritable bowel syndrome subjects.** *Gut Pathog* 2011, **3**:6.
7. Kaplan SL, Hulten KG, Gonzalez BE, Hammerman WA, Lamberth L, Versalovic J, Mason EO: **Three-year surveillance of community-acquired *Staphylococcus aureus* infections in children.** *Clin Infect Dis* 2005, **40**:1785–1791.

8. Grundmann H, Aires-De-Sousa M, Boyce J, Tiemersma E: **Emergence and resurgence of methicillin-resistant *Staphylococcus aureus* as a public-health threat.** *Lancet* 2006, **368**:874–885.
9. Holden MTG, Feil EJ, Lindsay JA, Peacock SJ, Day NPJ, Enright MC, Foster TJ, Moore CE, Hurst L, Atkin R, Barron A, Bason N, Bentley SD, Chillingworth C, Chillingworth T, Churcher C, Clark L, Corton C, Cronin A, Doggett J, Dowd L, Feltwell T, Hance Z, Harris B, Hauser H, Holroyd S, Jagels K, James KD, Lennard N, Line A, et al: **Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance.** *Proc Natl Acad Sci* 2004, **101**:9786–9791.
10. Vos PD, Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer K, Whitman WB: **The Firmicutes.** In *Bergey's Manual of Systematic Bacteriology*. 2nd edition. New York, NY, USA: Springer; 2009:400–401.
11. Delcher AL, Bratke KA, Powers EC, Salzberg SL: **Identifying bacterial genes and endosymbiont DNA with glimmer.** *Bioinformatics* 2007, **23**:673–679.
12. Tatusov RL, Koonin EV, Lipman DJ: **A genomic perspective on protein families.** *Science* 1997, **278**:631–637.
13. Disz T, Akhter S, Cuevas D, Olson R, Overbeek R, Vonstein V, Stevens R, Edwards RA: **Accessing the SEED genome databases via web services API: Tools for programmers.** *BMC Bioinforma* 2010, **11**:319.
14. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Ponten T, Ussery DW, Aarestrup FM, Lund O: **Multilocus sequence typing of total-genome-sequenced bacteria.** *J Clin Microbiol* 2012, **50**:1355–1361.
15. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM: **DNA-DNA hybridization values and their relationship to whole-genome sequence similarities.** *Int J Syst Evol Microbiol* 2007, **57**:81–91.
16. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389–3402.
17. Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, Haley BJ, Taviani E, Jeon YS, Kim DW, Lee JH, Brettin TS, Bruce DC, Challacombe JF, Detter JC, Han CS, Munk AC, Chertkov O, Meincke L, Saunders E, Walters RA, Huq A, Nair GB, Colwell RR: **Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*.** *Proc Natl Acad Sci* 2009, **106**:15442–15447.
18. Lee JH, Yi H, Chun J: **rRNASelector: a computer program for selecting ribosomal RNA encoding sequences from metagenomic and metatranscriptomic shotgun libraries.** *J Microbiol* 2011, **49**:689–691.
19. Kim OS, Cho YJ, Lee K, Yoon SH, Kim M, Na H, Park SC, Jeon YS, Lee JH, Yi H, Won S, Chun J: **Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species.** *Int J Syst Evol Microbiol* 2012, **62**:716–721.
20. Barber M: **Methicillin-resistant *Staphylococci*.** *J Clin Pathol* 1961, **14**:385–393.
21. Rosenbach FJ: **Microorganismen bei den Wund-Infektions-Krankheiten des Menschen.** In *Microorganismen bei den Wund-Infektions-Krankheiten des Menschen*. Edited by Bergmann JF. Wiesbaden, Germany; 1884:1–122.

doi:10.1186/1757-4749-6-6

Cite this article as: Kim et al.: Genome sequence of type strain of *Staphylococcus aureus* subsp. *aureus*. *Gut Pathogens* 2014 **6**:6.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

