



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Holt, DC;Holden, MTG;Tong, SYC;Castillo-Ramirez, S;Clarke, L;Quail, MA;Currie, BJ;Parkhill, J;Bentley, SD;Feil, EJ;Giffard, PM

Title:

A very early-branching staphylococcus aureus lineage lacking the carotenoid pigment staphyloxanthin

Date:

2011-12-01

Citation:

Holt, D. C., Holden, M. T. G., Tong, S. Y. C., Castillo-Ramirez, S., Clarke, L., Quail, M. A., Currie, B. J., Parkhill, J., Bentley, S. D., Feil, E. J. & Giffard, P. M. (2011). A very early-branching staphylococcus aureus lineage lacking the carotenoid pigment staphyloxanthin. *Genome Biology and Evolution*, 3 (1), pp.881-895. <https://doi.org/10.1093/gbe/evr078>.

Persistent Link:

<https://hdl.handle.net/11343/256567>

License:

CC BY-NC

# A Very Early-Branching *Staphylococcus aureus* Lineage Lacking the Carotenoid Pigment Staphyloxanthin

Deborah C. Holt<sup>1</sup>, Matthew T.G. Holden<sup>2</sup>, Steven Y.C. Tong<sup>1</sup>, Santiago Castillo-Ramirez<sup>3</sup>, Louise Clarke<sup>2</sup>, Michael A. Quail<sup>2</sup>, Bart J. Currie<sup>1</sup>, Julian Parkhill<sup>2</sup>, Stephen D. Bentley<sup>2</sup>, Edward J. Feil<sup>3</sup>, and Philip M. Giffard<sup>\*1</sup>

<sup>1</sup>Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory, Australia

<sup>2</sup>The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

<sup>3</sup>Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, United Kingdom

\*Corresponding author: E-mail: phil.giffard@menzies.edu.au.

**Accepted:** 23 July 2011

## Abstract

Here we discuss the evolution of the northern Australian *Staphylococcus aureus* isolate MSHR1132 genome. MSHR1132 belongs to the divergent clonal complex 75 lineage. The average nucleotide divergence between orthologous genes in MSHR1132 and typical *S. aureus* is approximately sevenfold greater than the maximum divergence observed in this species to date. MSHR1132 has a small accessory genome, which includes the well-characterized genomic islands, vSA $\alpha$  and vSA $\beta$ , suggesting that these elements were acquired well before the expansion of the typical *S. aureus* population. Other mobile elements show mosaic structure (the prophage  $\phi$ Sa3) or evidence of recent acquisition from a typical *S. aureus* lineage (SCCmec, ICE6013 and plasmid pMSHR1132). There are two differences in gene repertoire compared with typical *S. aureus* that may be significant clues as to the genetic basis underlying the successful emergence of *S. aureus* as a pathogen. First, MSHR1132 lacks the genes for production of staphyloxanthin, the carotenoid pigment that confers upon *S. aureus* its characteristic golden color and protects against oxidative stress. The lack of pigment was demonstrated in 126 of 126 CC75 isolates. Second, a mobile clustered regularly interspaced short palindromic repeat (CRISPR) element is inserted into *orfX* of MSHR1132. Although common in other staphylococcal species, these elements are very rare within *S. aureus* and may impact accessory genome acquisition. The CRISPR spacer sequences reveal a history of attempted invasion by known *S. aureus* mobile elements. There is a case for the creation of a new taxon to accommodate this and related isolates.

**Key words:** bacterial species, staphyloxanthin, CRISPR, positive selection.

## Introduction

The bacterial species *Staphylococcus aureus* is a human commensal that commonly colonizes the skin and mucosal surfaces. It is also a major human pathogen that can cause a variety of disease states, including minor skin and soft tissue infections and life threatening systemic and pulmonary infections. *Staphylococcus aureus* is a phylogenetically well-defined species. Orthologous pairs of housekeeping coding sequences (i.e., core genes) exhibit in general <2% nucleotide diversity within *S. aureus*, which is at least 10-fold lower than the diversity between *S. aureus* and the most closely related *Staphylococcus* species (Enright et al. 2000; Poyart et al. 2001; Drancourt and Raoult 2002; Ghebremedhin et al. 2008). Multilocus sequencing typing (MLST)-based studies have revealed that intraspecific

homologous recombination occurs occasionally within the core genome but that this is less frequent than in other bacterial species (Feil et al. 2003; Pearson et al. 2009). Furthermore, a high level of synteny (gene order) is retained between different *S. aureus* strains. Superimposed upon the stable *S. aureus* core genome is a more variable accessory genome composed of phage, genomic islands, transposons, and mobile genetic elements. These accessory genes are rapidly lost or acquired by horizontal gene transfer and play a key role in adaptation and pathogenicity (Lindsay and Holden 2006).

A community-associated lineage of *S. aureus* termed “clonal complex 75” (CC75), was recently reported as the dominant community-associated methicillin-resistant *S. aureus* (MRSA) genotype among indigenous communities in

The Author(s) 2011. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

the Northern Territory of Australia (McDonald et al. 2006). The presence of this lineage has subsequently been documented in south east Asia (Ruimy et al. 2009) and South America (Ruimy et al. 2010), and data housed on [sau-reus.mlst.net](http://sau-reus.mlst.net) also point to its presence in Europe. This suggests that CC75 is widespread but has until recently escaped notice, possibly because of difficulties in typing these strains using the standard MLST primers. MLST data revealed an average nucleotide divergence between CC75 and typical *S. aureus* of 9.6% over all seven loci (ranging from 5.8% at *gmk* to 15.8% at *aroE*) (Ng et al. 2009; Ruimy et al. 2009). Despite this very high level of divergence, CC75 remains more closely related to typical *S. aureus* than the next most closely related species (*Staphylococcus simiae*), and it is identical to *S. aureus* by 16S ribosomal RNA (rRNA; Ng et al. 2009; Ruimy et al. 2009). Array-based experiments are also suggestive of high divergence (Monecke et al. 2010).

Here we report the complete genome sequence of strain MSHR1132, a CC75 clinical isolate from the tropical north of the Northern Territory of Australia. These data confirm the high level of core nucleotide divergence between CC75 and typical *S. aureus*. The genome is toward the low end of the size range of other sequenced *S. aureus* strains, and this is due to a small accessory genome that contains no pathogenicity islands (SaPIs) or novel elements. The presence of the genomic islands, vSA $\alpha$  and vSA $\beta$ , suggests that these elements were acquired very early in the evolution of *S. aureus*. Two other notable attributes of the MSHR1132 genome are the absence of the operon encoding staphyloxanthin and the presence of a CRISPR region. We discuss the possible adaptive significance of these findings.

## Methods

### Bacterial Isolates

MSHR1132 was isolated from the blood culture of an indigenous woman with necrotizing fasciitis at the Royal Darwin Hospital (RDH), Darwin, Northern Territory, Australia, in September 2006. The isolate was resistant to oxacillin but susceptible to all other tested non  $\beta$ -lactam antibiotics. Additional CC75 isolates (M34, M70, M180, HS2, HS22, HS42, HS158, SCC1098, SCC1119, SCC1165, SCC1229, SCC1302) from impetigo lesions (McDonald et al. 2006) and hospital clinical specimens (Tong et al. 2009) were identified using a real-time polymerase chain reaction (PCR) single nucleotide polymorphism (SNP) typing approach (McDonald et al. 2006).

### Whole Genome Sequencing

The whole genome of the *S. aureus* isolate MSHR1132 was sequenced using both capillary sequencing (on ABI 3730xl analysers) and pyrosequencing (on 454 instruments; subsidiary of Roche Diagnostics Corporation, Branford, CT). A total of 29,300 high-quality capillary reads were produced mostly

from two subclone libraries (libraries with inserts in the range 2–6 kb using the vector pOTW12 and libraries with inserts in the range 5–12 kb using the vector pMAQ1-Sac\_BstXI). The average read length was 650 bp, and these reads represented 6.8 $\times$  coverage. The 454 sequencing produced 73.9 Mb data in reads with an average length of 245 bp. The assembly of these reads using Newbler 1.1.03.24 gave 70 contigs >500 bp with a combined length of 2,753,553 bp in nine scaffolds.

A combined assembly of the capillary reads, using phrap, and the consensus sequences from the 454 assembly (which were converted into overlapping 500-bp sequences) produced 26 contigs >2 kb with an N50 of 532 kb. A further 2,310 high-quality reads were produced to close gaps and to improve the quality of the sequence to finished standard.

The sequences and annotations of the *S. aureus* strain MSHR1132 chromosome and plasmid have been deposited in the EMBL database under accession numbers FR821777 and FR821778, respectively. The sequence was annotated using Artemis software (Rutherford et al. 2000). Initial coding sequence (CDS) predictions were performed using Orpheus (Frishman et al. 1998), Glimmer2 (Delcher et al. 1999), and EasyGene (Larsen and Krogh 2003) software. These predictions were amalgamated, and codon usage, positional base preference methods, and comparisons with the nonredundant protein databases using BLAST (Altschul et al. 1990) and FASTA (Pearson and Lipman 1988) software were used to refine the predictions. The entire DNA sequence was also compared in all six potential reading frames against UniProt, using BLASTx (Altschul et al. 1990) to identify any possible coding sequences previously missed. Protein motifs were identified using Pfam (Bateman et al. 2002) and Prosite (Falquet et al. 2002), transmembrane domains were identified with TMHMM (Krogh et al. 2001), and signal sequences were identified with SignalP version 2.0 (Nielsen et al. 1997). rRNAs were identified using BLASTn (Altschul et al. 1990) alignment to defined rRNAs from the EMBL nucleotide database; transfer RNAs (tRNAs) were identified using tRNAscan (Lowe and Eddy 1997); stable RNAs were identified using Rfam (Griffiths-Jones et al. 2003).

### Comparative Genomics

Comparison of the genome sequences was facilitated by using the Artemis Comparison Tool (Carver et al. 2005), which enabled the visualization of BLASTn and tBLASTx comparisons (Altschul et al. 1990) between the genomes. Orthologous proteins were identified as reciprocal best matches using FASTA (Pearson and Lipman 1988) with subsequent manual curation. Pseudogenes had one or more mutations that would prevent correct translation; each of the inactivating mutations was subsequently checked against the original sequencing data.

*Staphylococcus aureus* sequences used for comparative genomic analysis were MRSA252 (accession number BX571856) (Holden et al. 2004), MSSA476 (BX571857) (Holden et al. 2004), MW2 (BA000033) (Baba et al. 2002), N315 (BA000018) (Kuroda et al. 2001), Mu50 (BA000017) (Kuroda et al. 2001), Mu3 (AP009324) (Neoh et al. 2008), COL (CP000046) (Gill et al. 2005), NCTC8325 (CP000253) (Gillaspy et al. 2006), USA3000 FPR3757 (CP000255) (Diep et al. 2006), JH9 (CP000703) (Mwangi et al. 2007), Newman (AP009351) (Baba et al. 2008), and RF122 (AJ938182) (Herron-Olson et al. 2007). The sequences were also compared with *Staphylococcus epidermidis* RP62a (CP000029) (Gill et al. 2005), *S. epidermidis* ATC12228 (AE015929) (Zhang et al. 2003), *Staphylococcus haemolyticus* JCSC1435 (AP006716) (Takeuchi et al. 2005), *Staphylococcus saprophyticus* ATCC 15305 (AP008934) (Kuroda et al. 2005), *Staphylococcus carnosus* TM300 (AM295250) (Rosenstein et al. 2009), and *Staphylococcus lugdunensis* HKU09-01 (CP001837) (Tse et al. 2010).

### Phylogeny Based on Genetic Content Distance Matrix

As part of the genome alignment process, progressiveMauve (Darling et al. 2010) generates a pairwise genome content distance matrix that reflects the proportions of the genomes that are included in the initial set of local alignments. Because the progressiveMauve algorithm only requires that a sequence block be included in at least two of the genomes for it to be included in the multiple alignment, this analysis yields the alignable blocks of sequence between each pair of genomes. This extent of the alignable blocks is an excellent measure of shared content. The pairwise distances may be calculated by dividing the number total extent of the aligned blocks by the average of the sizes of genomes. In this way, progressiveMauve was used to generate a distance matrix, which was converted into a tree using the Neighbor-Joining algorithm, implemented in PHYLIP (Felsenstein 1989).

### Phylogeny Based on SNPs

ProgressiveMauve produces several out files, which have data related to the genome alignment. One of these contains every SNP in the alignment. Using an ad hoc perl script, we parsed this file to get only those SNPs that were located in homologous regions across the genomes of MSHR1132, MRSA252, USA300\_FPR3757, ED98, and *S. epidermidis* RP62a. With those SNPs, the percentage of identity between any two genomes was established, and an identity matrix was assembled and converted to a tree using the Neighbor-Joining algorithm implemented in PHYLIP (Felsenstein 1989).

### Identification of Genes Under Positive Selection

We utilized the branch-site models implemented in PAML (Yang 2007), which aim to detect positive selection that

has affected only a few sites on some lineages. The lineages under positive selection are known as foreground branches, whereas all other branches are the background branches; this distinction between foreground and background branches should be done a priori (in this case, the branch leading to MSHR1132 was the foreground branch, whereas the rest of the branches were background branches). According to this model, only foreground branches may have experienced positive selection. The model assumes four classes of sites (here sites mean codons rather than single nucleotides and “w” equals  $dN/dS$ ): Class 0 includes codons that are conserved throughout the tree (with  $0 < w_0 < 1$ ,  $w_0$  estimated from the data); Class 1 has codons that are evolving neutrally throughout the tree ( $w_1 = 1$ ); Class 2a has codons that are conserved on the background branches ( $0 < w_0 < 1$ ), but on the foreground, branches are under positive selection ( $w_2 > 1$ ); and Class 2b includes codons that are evolving neutrally ( $w_1 = 1$ ) but become under positive selection on the foreground branches ( $w_2 > 1$ ). Whereas the alternative hypothesis uses the four classes of sites as described above thus allowing positive selection on the foreground branch, in the null hypothesis,  $w_2 = 1$  is fixed for Classes 2a and 2b. A simple likelihood ratio test (LRT) is constructed by comparing the null and alternative hypotheses (Zhang et al. 2005). The LRT was applied to a set of 1776 MSHR1132 genes for which orthologues were identified in each of the MRSA252, USA300\_FPR3757 and *S. epidermidis* RP62a genome sequences in the comparative genomic analyses.

### Phenotypic and Genotypic Analysis of Isolates

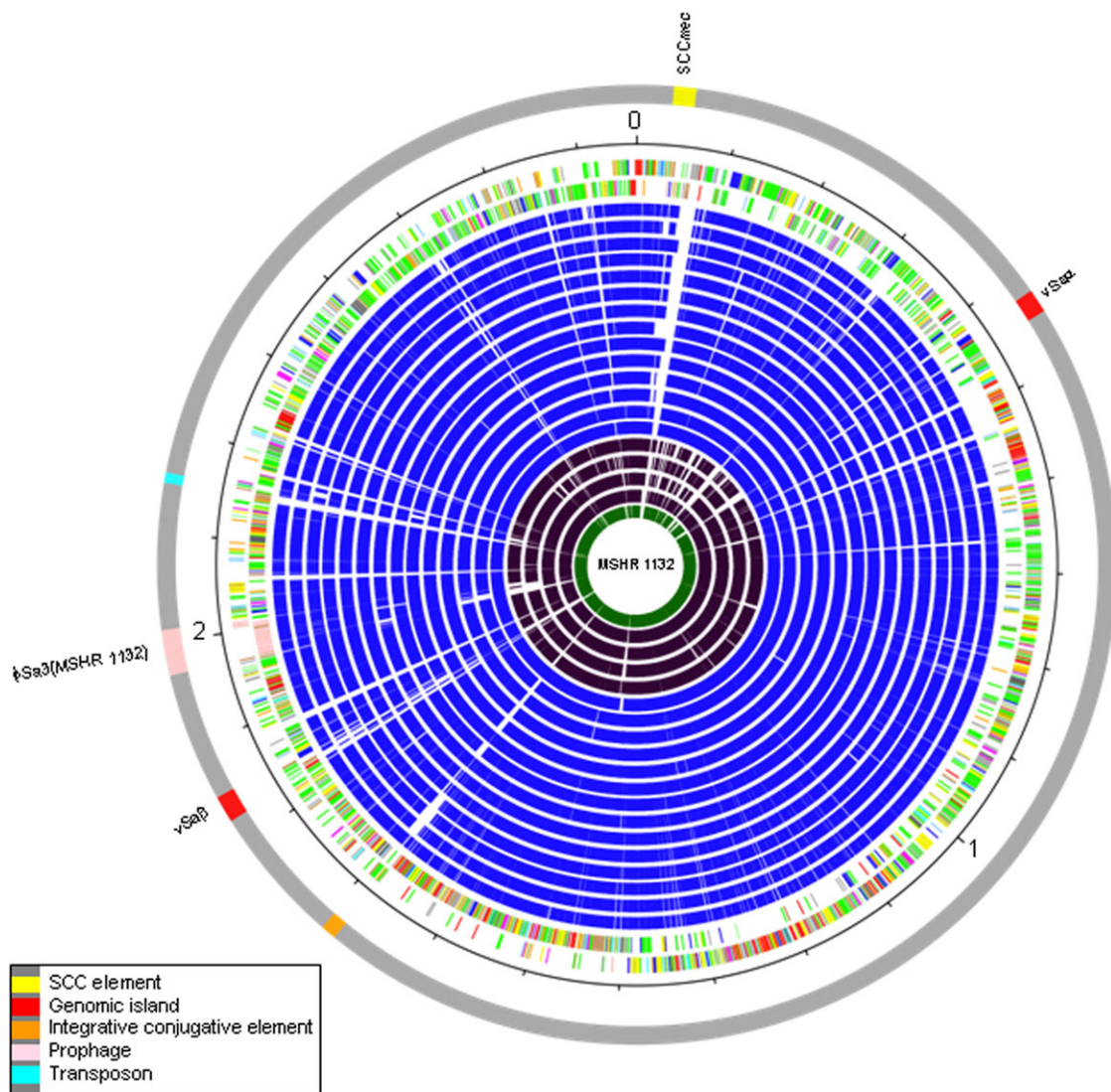
Isolates were tested for the production of staphyloxanthin by incubation on chocolate agar plates (Oxoid) for 48 h at 37 °C. Screening for *lukPV* was conducted by real-time PCR as previously described (McDonald et al. 2006). Multi-locus sequence typing of CC75 isolates was conducted using modified primers for *aroE* (Ruimy et al. 2009), *glpF*, *gmk*, *tpi*, and *yqjL* (Ng et al. 2009) with respect to the standard *S. aureus* MLST scheme (Enright et al. 2000).

Biochemical tests were carried out using a VITEK 2 device (bioMérieux—Australia Pty. Ltd., Baulkham Hills NSW, Australia) according to the manufacturer’s instructions.

## Results

### General Characteristics of the Genome

The genome of CC75 isolate MSHR1132 consists of a single circular chromosome of 2,762,762 bp containing 2,578 coding sequences (CDSs) and a single circular plasmid (pMSHR1132) of 24,853 bp, containing 23 CDSs. The chromosome is colinear in comparison with other sequenced *S. aureus* chromosomes and has a very similar gene content (fig. 1). The 16S rRNA sequence is confirmed as identical to that of other *S. aureus*. The MSHR1132 genome



**Fig. 1.**—Circular diagram of the *S. aureus* MSHR1132 chromosome. Key for the circular diagram (outer to inner): Outer colored segments on the gray outer ring represent genomic islands and horizontally acquired DNA (see figure for key); scale (in Mb); annotated CDSs colored according to predicted function are shown on a pair of concentric circles, representing both coding strands; *S. aureus* reciprocal fasta matches shared with the *S. aureus* strains (blue): MRSA252, MSSA476, MW2, N315, Mu50, Mu3, COL, NCTC8325, USA3000 FPR3757, JH9, Newman, RF122, LGA251, TW20; *Staphylococcus* reciprocal fasta matches shared with the staphylococcal species (purple): *S. epidermidis*, *S. saprophyticus*, *S. haemolyticus*, *S. carnosus*, reciprocal fasta matches shared with *Macrococcus caseolyticus* (green); CDS functions: dark blue, pathogenicity/adaptation; black, energy metabolism; red, information transfer; dark green, surface associated; cyan, degradation of large molecules; magenta, degradation of small molecules; yellow, central/intermediary metabolism; pale green, unknown; pale blue, regulators; orange, conserved hypothetical; brown, pseudogenes; pink, phage and IS elements; and gray, miscellaneous.

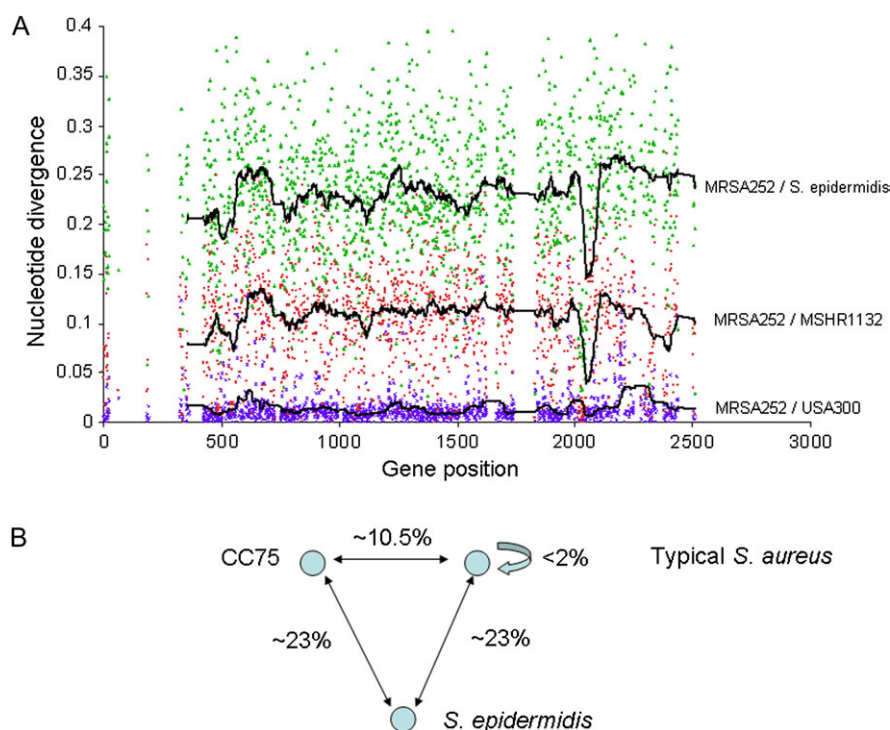
possesses a novel *spa* type (repeat numbers 259, 31, 17, 17, 17, 22, 17, 17, 23, 17, and 22), a distinct *coa* type (type XII), *agr* type I, and MLST sequence type (ST) 1850. The capsule gene cluster of 15 genes (SAMSHR1132\_01230 to SAMSHR1132\_01380) closely resembles the capsule gene cluster that encodes serotype 8 in other *S. aureus* strains.

The accessory genome is small. Readily identified elements consist of the genomic islands  $\alpha$  and  $\beta$ , a type IVa Staphylococcal cassette chromosome *mec* (SCC*mec*) in-

serted into *orfX*, a single integrative conjugative element (ICE), a single putative transposon, a single prophage, and a plasmid pMSHR1132. *Staphylococcus aureus* pathogenicity islands (SaPIs) were not detected.

### MSHR1132 is Divergent from Other *S. aureus* Throughout the Core Genome

Sequencing of the MLST loci and a small sample of other housekeeping genes had previously demonstrated that



**FIG. 2.**—Nucleotide divergence plot of MRSA252 compared with USA300\_FPR3757, MSHR1132 and *S. epidermidis* RP62a. (A) Plot of the nucleotide divergence against gene position between the genomes of MRSA252 and USA300\_FPR3757 (purple), MSHR1132 (red) and *S. epidermidis* RP62a (green). (B) The divergence between MRSA252 and MSHR1132 is approximately seven times greater than that between MRSA252 and USA300\_FPR3757 and slightly less than half the divergence between *S. aureus* and *S. epidermidis*.

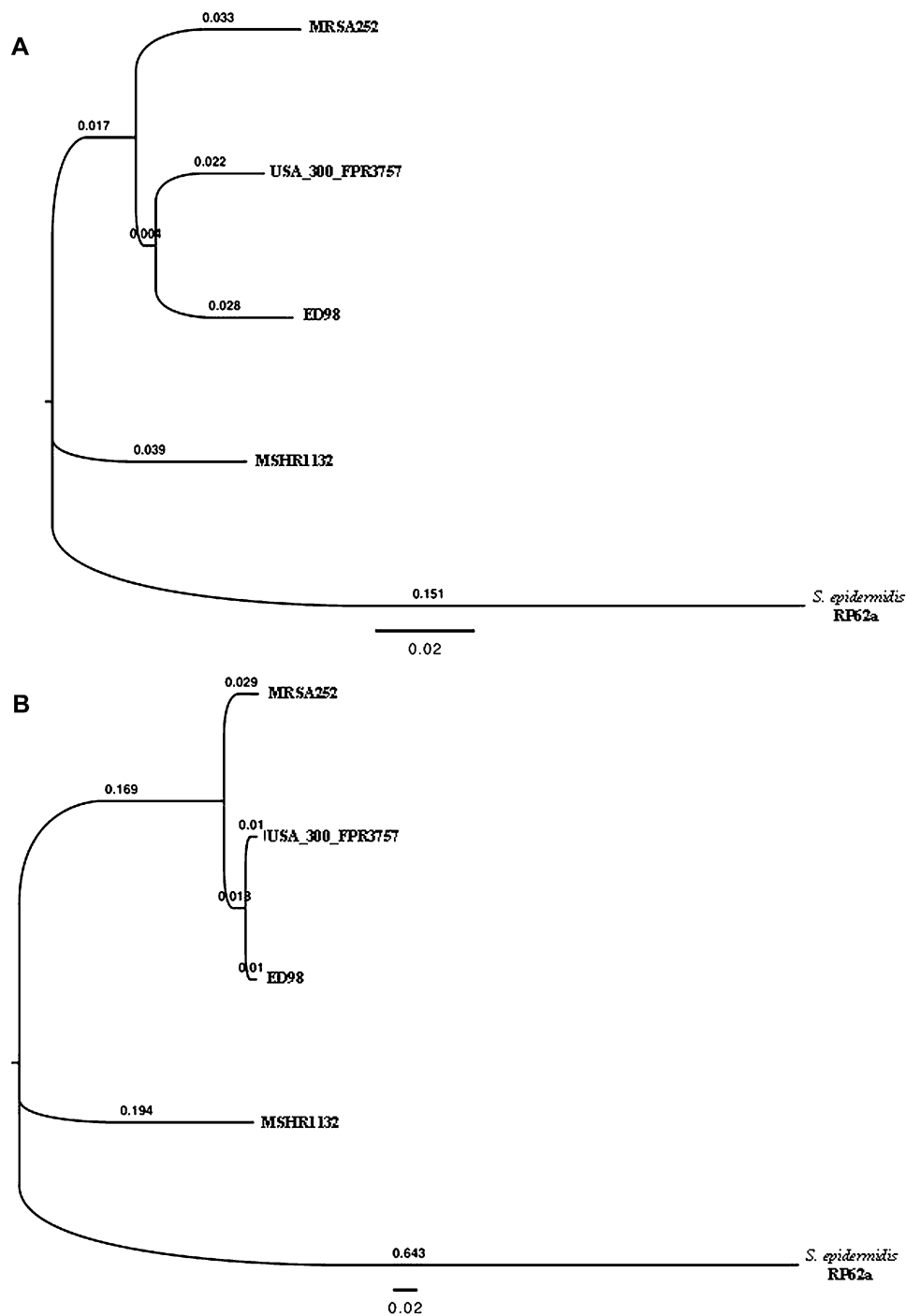
these genes were significantly diverged in CC75 isolates compared with other *S. aureus* (Ruimy et al. 2009; Ng et al. 2009). To refine the phylogenetic position of MSHR1132, the identities between 1498 orthologous CDSs in the core genomes of MRSA252 and MSHR1132, USA300\_FPR3757, and *S. epidermidis* RP62A were determined. MRSA252 and USA300\_FPR3757 were chosen because they are from different phylogenetic groups within *S. aureus* (CC30/group 1a and CC8/group 2, respectively), so their divergence represents the upper range of the divergence between non-CC75 *S. aureus* strains (Cooper and Feil 2006). The divergence of MSHR1132 from MRSA252 is approximately seven times greater than that between MRSA252 and USA300\_FPR3757 and slightly less than half the divergence between *S. aureus* and *S. epidermidis* (fig. 2). There is no clear evidence for recent core genome horizontal gene transfer between MSHR1132 and any of the other strains.

In order to confirm the phylogenetic position of CC75 relative to typical *S. aureus* and *S. epidermidis*, we constructed a tree based on nucleotide divergence of the core genes and a second tree based on differences in genome content. Although the topologies of the two trees are identical and each places CC75 as intermediate between typical *S. aureus* and *S. epidermidis* with respect to time of divergence, there are large differences in the relative branch lengths (fig. 3;

note that the two trees are drawn to different scales, and branch lengths are given on the tree). In the nucleotide divergence tree, the branch lengths within the typical *S. aureus* clade are much shorter than the branch leading to CC75. This reflects the fact that the divergence between unrelated lineages of typical *S. aureus* genomes is ~2%, whereas CC75 is ~10% diverged from typical *S. aureus*. However in the genome content tree, the branch lengths leading to the typical *S. aureus* and CC75 are much more similar. For example, the distance between MRSA252 and ED98 is 85% of the distance between MSHR1132 and ED98. The simplest explanation of these data is rapid evolution of the accessory genome, with constraints both on total genome size and on core genome size that result in accessory genome differences between strains rapidly becoming saturated.

### Variation within the CC75 Lineage

In order to gauge the diversity within the CC75 lineage, we characterized 12 further CC75 isolates recovered from the northern part of the Northern Territory of Australia using a modified MLST scheme. The Australian isolates were resolved by MLST into five STs. A neighbor joining tree that included the CC75 ST ST1223 present on the *S. aureus* MLST database resolved the STs into three distinct lineages represented by STs 1824, 1848, and 1849, STs 1223 and 1823, and ST 1850 (that includes MSHR1132)

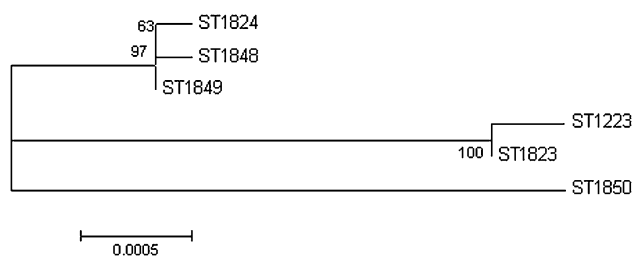


**Fig. 3.**—Nucleotide divergence and genome content neighbor joining trees. The neighbor joining tree was generated from distances matrices based on (A) genome content and (B) nucleotide divergence between MRSA252, USA300\_FPR3757, ED98, MSHR1132, and *S. epidermidis* RP62a.

(fig. 4). The average divergence between these STs was approximately 0.5%, and the maximum divergence, between STs 1850 and 1223, was 0.94%. This is an approximately 10-fold higher level of divergence than is seen within typical *S. aureus* clonal complexes by MLST (e.g., CC8 and CC30). Thus, this lineage exhibits a level of divergence more typical of an entire staphylococcal species. There is no evidence for

high rates of recombination between these isolates, either using the phi test (implemented in SplitsTree4.0) (Huson and Bryant 2006) (which was not significant) or the Recombination Detection Program suite of programs (Martin and Rybicki 2000) (which did not identify any recombination events).

Originally ST1223 was identified in isolates recovered from Cambodia (Ruimy et al. 2009); however, more recently



**FIG. 4.**—Variation within the CC75 lineage. Analysis of six CC75 multilocus sequence types by neighbor joining, which resolves three groups.

isolates from this ST have been reported at a high frequency in a sample of asymptotically carried isolates from Wayampi Amerindians living in a remote Amazonian village in French Guiana in South America (Ruimy et al. 2010). Although ST1223 is distinct from most of the isolates from Australia, it is very closely related to ST1823 that was represented by the Australian isolate HS2.

Six isolates representing the three lineages (M34 [ST1824], MSHR1132 [ST1850], SCC1165 [ST1848], SCC1302 [ST1850], HS2 [ST1823], and HS22 [ST1849]) were subjected to phenotypic analysis and identification using a VITEK2 device. All were identified as *S. aureus* with a confidence of either “very good” or “excellent.” This is essentially identical to the findings of Monecke et al. (2010). *Staphylococcus aureus* ATCC29213 is the quality control strain recommended by the VITEK2 manufacturer. The CC75 isolates differed from this only in their being positive for trehalose utilization and variable for urease production and O/129 resistance. *Staphylococcus aureus* ATCC29213 is negative, negative and positive for these tests, respectively. GenBank searches indicated that *S. aureus* commonly carries genes for urease production and trehalose utilization, so for these results do not correspond with any obvious genetic differences between CC75 and other *S. aureus*. It was concluded that commonly used biochemical tests will not reliably discriminate CC75 isolates from other *S. aureus*.

### Evidence for Positive Selection

In order to identify genes, which may experience positive selection in CC75 and which may therefore provide clues as to the adaptation of this lineage, we used the branch-site models implemented in PAML. A total of 1776 orthologous genes common to the genome sequences of MSHR1132, USA300\_FPR3757, MRSA252 (*S. aureus*), and *S. epidermidis* RP62a were identified and used in this analysis. The MSHR1132 sequence was defined as the foreground branch. Twenty-three of the 1,776 genes were identified as having experienced positive selection in the branch leading to the CC75 lineage (table 1). These include five genes encoding proteins associated with the cell envelope (including two putative membrane proteins, Riley’s category 4) and the agr (accessory gene regulator) protein B and a number of

metabolic genes (Riley’s category 3). The analysis was repeated twice using MRSA252 or USA300 as the foreground branch, and this identified 33 and 29 genes under positive selection respectively (supplementary tables S1 and S2, Supplementary Material online). The relatively low number of genes identified to be under positive selection in CC75 is consistent with the suggestion of lower propensity to cause disease than typical *S. aureus*. Interestingly, we note that although 11 cell membrane genes (Riley’s category 4) in MRSA252 were detected as positively selected, only 2 cell membrane genes were identified in USA300. However, nine metabolic genes appear to be under positive selection in USA300. This analysis therefore suggests quite different adaptive paths between CC75, the hospital-associated MRSA252 and the community-acquired USA300, and also challenges the widely held assumption that the evolution of metabolic genes is predominantly driven by purifying selection.

### CC75 Is Nonpigmented

A defining feature of *S. aureus* is the production of the membrane-bound triterpenoid carotenoid pigment staphyloxanthin that confers upon *S. aureus* its characteristic golden color (Marshall and Wilmoth 1981). Staphyloxanthin protects against oxidative stress and scavenging reactive oxygen substances, making *S. aureus* more resistant to hydrogen peroxide, superoxide radical, hydroxyl radical, hypochloride, and neutrophil killing (Liu et al. 2005; Clauditz et al. 2006). It may also contribute to intracellular survival after phagocytosis (Olivier et al. 2009). A *S. aureus* ΔCrtM mutant is more susceptible to oxidant killing, has impaired neutrophil survival, and is less pathogenic in a mouse subcutaneous abscess model (Liu et al. 2005). Inhibition of staphyloxanthin synthesis in vivo resulted in increased susceptibility of *S. aureus* to killing by human blood and to innate immune clearance in a mouse infection model (Liu 2008).

Staphyloxanthin is encoded by CDSs SAR2642–SAR2647 in the MRSA252 chromosome, and orthologs have been found in all other *S. aureus* genome sequences to date. Orthologous CDSs were not found in the MSHR1132 genome. The ability of 126 of our collection of CC75 isolates to produce staphyloxanthin was tested by extended incubation on chocolate agar. None produced pigment, with the colonies being a brilliant white color. In contrast, randomly selected non-CC75 *S. aureus* clinical isolates from our collection were all clearly pigmented (fig. 5). It was therefore concluded that CC75 is nonpigmented.

### The orfX Region of MSHR1132 Encompasses a Type IVa SCCmec Element, Putative hsd Genes, and CRISPR/cas Genes

MSHR1132 carries a 24,181-bp type IVa SCCmec element, typical of community-acquired MRSA (fig. 6). As is seen with other staphylococci, it is inserted into *orfX*. The element is

**Table 1**

Identification of Genes under Positive Selection in MSHR1132

MSHR1132 CDS	Riley's Category <sup>a</sup>	Predicted Function
MSHR1132_01070	4	putative membrane protein
MSHR1132_02940	4	lipase precursor
MSHR1132_03360	7	putative GTP-binding protein
MSHR1132_04100	4	putative membrane protein
MSHR1132_04330	7	tetrapyrrole (corrin/porphyrin) methylase family protein
MSHR1132_06960	0	conserved hypothetical protein
MSHR1132_10340	7	cell division initiation protein, putative
MSHR1132_11510	3	glutamine synthetase
MSHR1132_11740	3	threonine synthase
MSHR1132_12220	1	putative oligopeptide transporter ATPase
MSHR1132_12530	3	dihydrolipoamide succinyltransferase E2 component of 2-oxoglutarate dehydrogenase complex
MSHR1132_13190	4	cell surface elastin binding protein
MSHR1132_13700	7	putative peptidase
MSHR1132_13730	7	lipoate-protein ligase A protein
MSHR1132_14220	6	heat-inducible transcription repressor
MSHR1132_15910	7	FtsK/SpoIIIE family protein
MSHR1132_18500	1	putative sodium transport protein
MSHR1132_18580	4	accessory gene regulator protein B
MSHR1132_18760	3	acetolactate synthase large subunit
MSHR1132_21220	0	hypothetical protein
MSHR1132_22600	3	putative glycerate kinase
MSHR1132_22710	7	2-dehydropantoate 2-reductase
MSHR1132_23700	1	putative ferrous iron transport protein B
Riley's category	Function	Number of genes
7	Not classified (includes putative assignments)	7
4	Cell envelope	5
3	Metabolism of small molecules	5
1	Cell processes	3
0	Unknown function, no known homologs	2
6	Regulatory functions	1

NOTE.—A total of 1,776 orthologous genes common to MSHR1132, MRSA252, USA300\_ FPR3757, and *S. epidermidis* RP62a were analyzed using the branch-site models implemented in PAML with the MSHR1132 sequence was defined as the foreground branch. Twenty-three of the 1,776 genes were identified as having experienced positive selection in the branch leading to the CC75 lineage.

<sup>a</sup> Genes were classified according to cellular function (78).

most similar to the *SCCmec* in isolate JKD6159, a ST93 MRSA (Chua et al. 2010), with only three SNPs over the entire length, and with a 55-bp duplication present in JKD6159. ST93 is a dominant strain of Panton Valentin Leukocidin (PVL)+ community-acquired MRSA in Australia (Nimmo et al. 2006) having recently emerged in northern or eastern Australia. It is undergoing radiation and currently coexists with CC75 in the north of the Northern Territory (Tong et al. 2010). Recent horizontal transfer of this *SCCmec* element between CC75 and ST93 would simply explain their close similarity. The MSHR1132 *SCCmec* is also very similar to the *SCCmec* in the Japanese MRSA isolate JCSC4744 (ST391 and CC91) (Kishii et al. 2008); there is only one SNP in the 18 kb that is available for this region in JCSC4744. There are also only seven SNPs compared with the *SCCmec* element in the USA300 isolate TCH1516 (Highlander et al. 2007).

Three *hsd* genes or pseudogenes are located immediately adjacent to *SCCmec* and distal to the chromosomal origin of

replication. Genes encoding restriction modification systems have previously been observed in this location in both *S. aureus* and *S. epidermidis* (Mongkolrattanohai et al. 2004; Gill et al. 2005; Noto et al. 2008). They have also been mobilized as part of *SCCmec* type V element (Ito et al. 2004). However, the *hsd* clusters at this location in different strains and species are only very distantly related to each other, and much closer relatives of these genes can be found at other locations. This implies multiple parallel events of *hsd* locus insertion into *orfX*.

Immediately adjacent to the *hsd* gene cluster in MSHR1132 is the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas) (CRISPR/cas) locus (fig. 6). CRISPR/cas has been found in a variety of bacteria and Archaea, where it mediates defense against mobile genetic elements (Horvath and Barrangou 2010). The only *S. aureus* lineage that has previously been noted to contain a CRISPR/cas locus is the livestock-associated



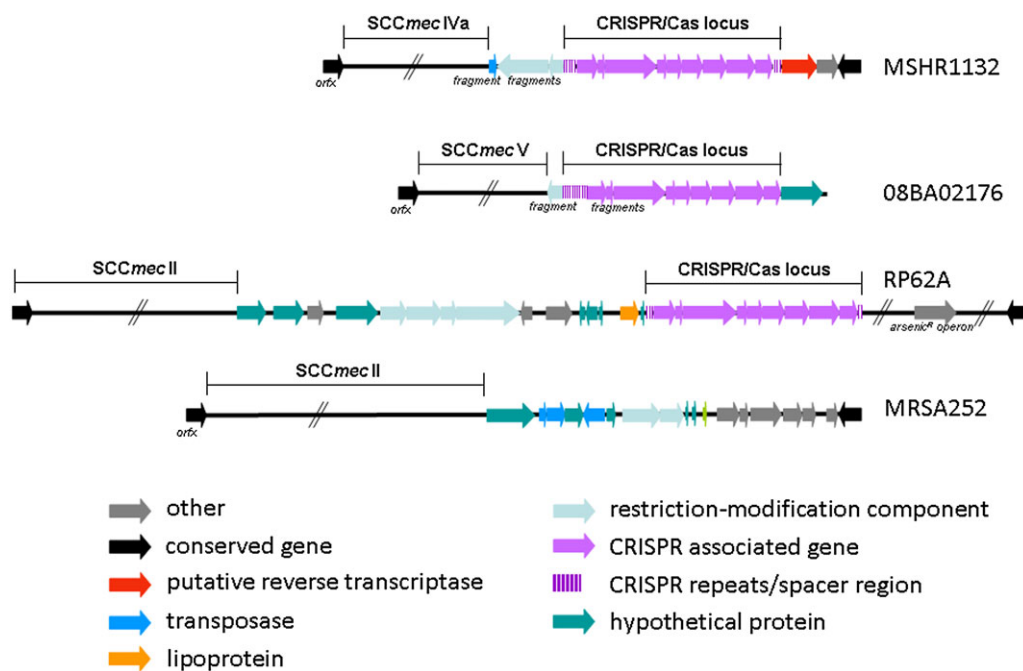
**FIG. 5.**—CC75 isolates lack staphyloxanthin. The *S. aureus* clinical isolate SCC1007 (ST93) (Tong et al. 2009) (left) produces staphyloxanthin, whereas the CC75 isolate MSHR1132 (right) does not.

ST398 clone (Golding et al. 2010). CRISPR/cas is present in at least some strains of *S. epidermidis* (Gill et al. 2005) where it has been demonstrated to limit plasmid conjugation (Marraffini and Sontheimer 2008) and *S. lugdunensis* (Tse et al. 2010) but has not been observed to date in *S. saprophyticus*, *S. haemolyticus*, *S. carnosus*, or *Macrocooccus*

*caseolyticus*. In general, CRISPR/cas elements have been identified in ~40% of bacterial genomes sequenced to date (van der Oost et al. 2009).

CRISPR loci typically consist of repeat sequences interspersed with variable spacer sequences, which are generally segments of DNA captured from viral or plasmid sequences. These acquired and heritable DNA spacer sequences are utilized by the Cas-encoded proteins in a defense system against mobile genetic elements, although intriguingly the CRISPR/cas elements are themselves associated with plasmids and phage (Touchon and Rocha 2010). The CRISPR loci are most often located adjacent to the cas CDSs, which encode a heterogeneous family of proteins including nucleases, helicases, and polymerases (Horvath and Barrangou 2010). Forty-five different protein families have been identified within the cas CDSs within prokaryotic genomes (Haft et al. 2005). The number and type of cas CDSs can vary considerably and appear to be linked to the sequence and length of the repeat unit (Haft et al. 2005).

The cas CDSs in MSHR1132 are homologous and syntenic with those found in the sequenced genomes of *S. epidermidis* RP62A and *S. lugdunensis* HKU09-01 and have similar levels of sequence divergence as other orthologous CDSs in these comparisons. Thus, the typical nucleotide divergence in the CRISPR/cas of CC75 suggests that there is no need to evoke recent horizontal acquisition to explain the presence of this element, the simplest explanation being that CRISPR/cas was present in a common ancestor of *S. epidermidis*, *S. lugdunensis*, and *S. aureus* and has been lost in



**FIG. 6.**—Structure of SCCmec and CRISPR/Cas region. Diagram of the structure of the region including SCCmec and CRISPR/Cas in MSHR1132, *S. aureus* 08BA02176 (ST398), *S. epidermidis* RP62a, and *S. aureus* MRSA252.

conventional *S. aureus* since it diverged from CC75. However, the patchy distribution of CRISPR/cas throughout the *Staphylococcus* genus as a whole and within single species (notably the presence of the element in *S. aureus* strain ST398 and its absence in *S. epidermidis* ATC12228) points to the mobility of this element. We also note that there is strong evidence for horizontal gene transfer of CRISPR/cas in other genera (Haft et al. 2005; Touchon and Rocha 2010).

In summary, there is conflicting evidence as to whether the presence of CRISPR/Cas is an ancestral or acquired characteristic in MSHR1132. Although sequence similarities suggest that it is ancestral, its patchy distribution and close association with *SCCmec* in the staphylococci suggest that it is either mobile or easily able to be mobilized by other elements.

The CRISPR spacer sequences are of interest because they provide a record of past challenges from mobile genetic elements and thus potentially indicate whether the presence of CRISPR in MSHR1132 has the potential to impede the acquisition of genes with clinical significance. The MSHR1132 CRISPR/cas locus possesses two CRISPR repeat/spacer arrays; the left contains six spacers (L1–L6) and the right four spacers (R1–R4) (supplementary table S3, Supplementary Material online). Spacers L5, L6, and R2 show similarity to small hypothetical genes observed in multiple members of the siphoviridae that are known to infect *S. aureus* and similar prophages in *S. aureus* genomes. Spacer L6 is of particular interest because it is identical to phage that carry the *lukPV* genes that encode the PVL toxin. We therefore screened 126 isolates of our collection of CC75 isolates for the presence of the *lukPV* genes, and as expected, all were negative. Spacers L4 and R1 are similar to *S. aureus* phages in the myoviridae family, whereas spacer R4 is similar to *S. aureus* plasmid sequences. The “WBG” plasmids listed (supplementary table S3, Supplementary Material online) are resident in *S. aureus* isolated from the remote north of Western Australia, which is consistent with the northern Australian origin of MSHR1132.

### Other Mobile Elements and Genomic Island Regions

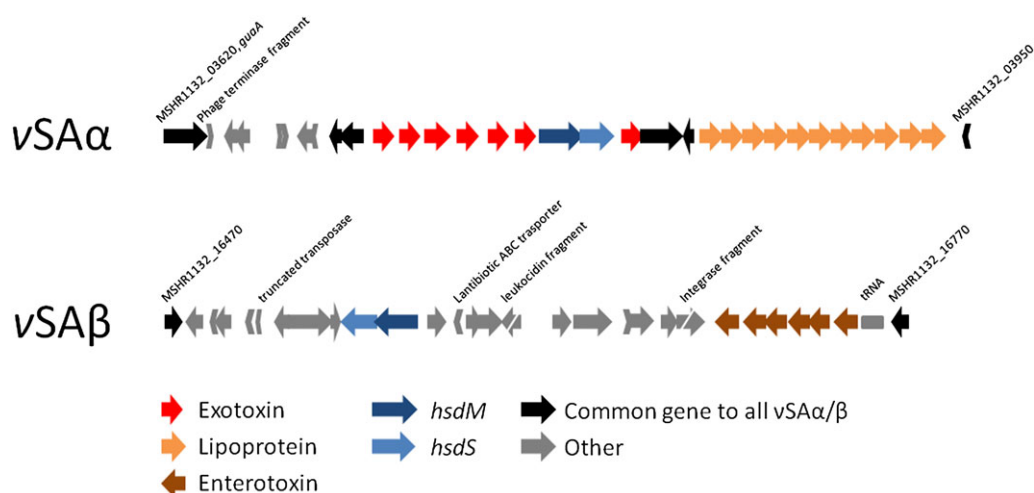
The genomic islands, *vSA $\alpha$*  and *vSa $\beta$* , are present at the same locations as in other *S. aureus* strains (fig. 1). These regions were examined to determine whether they had been acquired independently in MSHR1132 or inherited from a common ancestor. The *vSA $\alpha$*  genomic island in MSHR1132 is approximately 22 kb in size and includes CDSs for 7 exotoxins, type I restriction-modification system M and S subunits, and an array of 11 lipoproteins (fig. 7). As there are a number of differences relative to *vSA $\alpha$*  in other *S. aureus* strains, including the presence of a novel *hsdS* gene sequence, we argue that *vSA $\alpha$*  should be classified into a new type V. The exotoxin and lipoprotein sequences have between 84–97% identity with MRSA252, and no significantly

higher levels of similarity were found with other *S. aureus* strains. As this level of divergence is typical for orthologous genes elsewhere in the genome, these data indicate a long-term stable association with this island rather than a recent independent acquisition.

The *vSa $\beta$*  genomic island in MSHR1132 is approximately 25 kb in size and includes CDSs for six enterotoxins, type I restriction-modification system S and M subunits, and numerous hypothetical proteins (fig. 7). There is an absence of a serine protease gene cluster present in other *S. aureus* isolates. Similar to *vSA $\alpha$* , based on these differences and the presence of a novel *hsdS* gene sequence, we propose that *vSa $\beta$*  is classified into a new type IV. The enterotoxin array is syntenic to MRSA252, and the genes have sequence identities of 92–98%. Again this level of divergence is similar to the genome as a whole; hence, there is no compelling reason to evoke recent acquisition by horizontal transfer.

The 42,445-bp  $\phi$ Sa3(MSHR1132) prophage is inserted at the same site as  $\phi$ Sa3 insertions in other *S. aureus* strains within the phospholipase C precursor gene (SAMSHR1132\_17840).  $\phi$ Sa3(MSHR1132) is largely syntenic with other  $\phi$ Sa3 elements but lacks the gene encoding enterotoxin type A (SAR2043). The nucleotide sequence identity between  $\phi$ Sa3(MSHR1132) CDSs and homologous CDSs in other *S. aureus*  $\phi$ Sa3 prophage is highly variable from gene to gene, consistent with a highly mosaic structure. For example, the most similar public nucleotide database entry to the  $\phi$ Sa3(MSHR1132) portal protein gene (SAMSHR1132\_18060) is from *S. aureus* isolates JH1 and JH9, and there is only 75% identity. In contrast, the 756-bp autolysin-encoding gene (SAMSHR1132\_17880) has just one SNP with homologues in a variety of *S. aureus* isolates. In contrast to the genomic islands discussed above, these data suggest that there have been independent insertions of prophage or homologous recombination of prophage genes since CC75 diverged from other *S. aureus*.

There is a 14,419-bp ICE integrated into, and disrupting, a putative membrane CDS (SAMSHR1132\_15220). This entire element is ~99% identical to ICE6013 (Smyth and Robinson 2009) at the nucleotide level, with the exception that ICE6013<sub>MSHR1132</sub> contains two additional CDS (SAMSHR1132\_15320 and SAMSHR1132\_15330), which encode putative membrane proteins of unknown function. ICE6013 insertions are found in several locations in *S. aureus*, and ICE6013<sub>MSHR1132</sub> is inserted at the same location as ICE6013 in several sequenced strains (fig. 1). An ICE6013 element essentially identical to ICE6013<sub>MSHR1132</sub>, including the additional CDS, has been identified in plasmid pSK53 from *S. aureus* SK1700 that was isolated in Melbourne, Australia, in 1975 (GenBank accession number GQ915270). To date, this structure has not been found elsewhere. The high level of identity between ICE6013<sub>MSHR1132</sub> and ICE6013 in other *S. aureus* strains and the specific similarity with a plasmid-borne form mean that it is likely that



**Fig. 7.**—Structure of the MSHR1132 *vSA $\alpha$*  and *vSA $\beta$*  genomic islands. The MSHR1132 genome contains novel *vSA $\alpha$*  and *vSA $\beta$*  genomic islands, which were assigned types V and IV, respectively.

insertion of ICE6013<sub>MSHR1132</sub> took place after the divergence of CC75 and other *S. aureus*.

There is a putative transposon between bases 2,147,694 and 2,157,171. It is 98.3% identical and entirely syntenic with a putative transposon that has only been previously observed in the genome of *S. aureus* S0385. S0385 belongs to the pig-associated ST398 but is derived from a case of human endocarditis (Schijffelen et al. 2010). Similar to ICE6013, the high level of similarity with the structure in *S. aureus* S0385 suggests that its insertion postdates the divergence of CC75 and other *S. aureus*.

MSHR1132 contains a single plasmid (pMSHR1132) of 24,853 bp. The backbone of the plasmid is similar to other plasmids found in *S. aureus*, including pTW20\_2, which was identified in the TW20 (ST239) genome (Holden et al. 2010). The plasmid encodes a number of proteins involved in antibiotic and antiseptic resistance. An enoyl-acyl-carrier-protein reductase (NADH) 2 (FabI2) is found in an indel region of the plasmid. The protein is associated with hexachlorophene and triclosan resistance. This plasmid copy is highly similar to the chromosomal copy, but notably, the plasmid copy has substitutions associated with resistance (e.g., F204C; Fan et al. 2002). The plasmid also encodes the antiseptic resistance protein QacA in a divergent operon with the transcriptional regulator QacR and a beta-lactamase (BlaZ) on a remnant of a Tn552 transposon.

## Discussion

The analysis of the MSHR1132 genome sequence confirmed the phylogenetic divergence between CC75 and other *S. aureus* indicated by multilocus sequence analysis and array-based experiments (Ng et al. 2009; Ruimy et al. 2009; Monecke et al. 2010). Averaging the data from 1,498 orthologous CDSs revealed that MSHR1132 is approximately seven times more divergent from other *S. aureus* than major phylo-

genetic groups within non-CC75 *S. aureus* are diverged from each other and that the divergence between CC75 and other *S. aureus* is slightly under half the divergence between *S. aureus* and *S. epidermidis*. Non-CC75 *S. aureus* have been very successful and has undergone a global radiation, and it is clear that CC75 diverged from the progenitor of this radiation long before this radiation began.

The virulence of CC75 is far from understood. Essentially, the only systematic study was reported by Tong et al. (2010) and was based upon clinical isolates from RDH in the Australian Northern Territory. The spectrum of disease caused by CC75 was similar to other community-associated *S. aureus* lineages. However, CC75 was less likely to cause abscesses and sepsis. Also, this report outlined evidence that CC75 is underrepresented in clinical isolates from RDH, given its abundance in superficial skin lesions in some communities in the RDH catchment. Our working model is that CC75 has a greater association with minor skin infections and less propensity to cause serious and systemic infections than other *S. aureus*. We speculate that this reflects an ancestral, less invasive, and less versatile virulence phenotype as compared with typical *S. aureus*. The retention of this phenotype may be a cause of the small number of genes in MSHR1132 under positive selection for variation. The golden pigment staphyloxanthin is known to facilitate *S. aureus* invasive infections by providing oxidative defense (Liu et al. 2005; Clauditz et al. 2006; Liu et al. 2008; Olivier et al. 2009). The connection between the CC75 virulence phenotype and its lack of staphyloxanthin is an interesting topic for future research.

The MSHR1132 genome contains a locus specifying a CRISPR/cas system. It is intriguing that CRISPR/cas and the *hsd* genes are so close to each other and may have both inserted into *orfX* (fig. 6). This implies that *orfX* has attracted not only the SCC<sub>mec</sub> mobile elements but also DNA

encoding two separate systems both devoted to defense against mobile elements but that are themselves mobile. Clusters of this nature have been recently noted and termed “defense islands,” although the molecular mechanisms and evolutionary pressures underlying this are not currently understood (Makarova et al. 2009).

The CRISPR spacers provide a record of past encounters with bacteriophage and plasmids closely related to elements known to infect or reside in *S. aureus* and in some cases are identical to highly conserved and widely distributed bacteriophage genes, including those that encode the PVL toxin. The accessory genome of MSHR1132 is at the small end of the range found in *S. aureus* and contains no SaPIs. It is possible that this is a function of the presence of CRISPR/cas and that CC75 has a less plastic genome than other *S. aureus*. This is consistent with the notion that CC75 has an ancestral phenotype that is less impacted by the accessory genome. However, although it is tempting to regard CC75 as an ancestral relic and the absence of pigment and presence of CRISPR as primitive characteristics, a recent loss of pigmentation genes and acquisition of CRISPR cannot be ruled out at this time. Also, the presence of SCCmec, an ICE element, a prophage, a transposon, and a plasmid in the MSHR1132 genome demonstrates that this lineage is not completely resistant to genetic invaders.

The currently known global distribution of the CC75 lineage poses something of a mystery. Although the presence of this lineage in south east Asia can be broadly explained by geographical proximity to northern Australia, the discovery of ST1223 at a high frequency in carriage in a remote village in French Guiana is far more difficult to explain. Although we note paleontological evidence that the first South American settlers arrived from Australia (Neves and Hubbe 2005), we remain cautious in linking these hypotheses with the current distribution of CC75 in the absence of more extensive sampling. Furthermore, it is striking that the two MLST STs ST1223 and ST1823 are almost identical, differing at only two sites (0.06%), despite the former being recovered from French Guiana and the latter being an Australian isolate. This would argue against the view that *S. aureus* populations at these sites have remained isolated from each other and from the rest of the world since the earliest colonization of South America between 10,000 and 50,000 years ago. It is also noteworthy in this context that many of the globally widespread clonal complexes of typical *S. aureus* are also found both in French Guiana and in northern Australia, which again argues against the isolation of these populations. Thus, the current data remain insufficient to draw strong conclusions relating the distribution of CC75 with ancient patterns of human migration but rather suggest that the global scale dissemination patterns of CC75 may resemble typical *S. aureus*. It is noteworthy that although very closely related CC75 STs may be found on opposite sides of the world, the MSHR1132 SCCmec element is virtually

identical to that found in another community-associated *S. aureus* lineage that has a similar geographical distribution to Australian CC75. Also, the MSHR1132 ICE6013 allotype as also only been found in Australia.

Non-CC75 *S. aureus* is abundant and globally distributed. On the basis of universally high core genome ortholog similarities, it arose from a point source that existed a long time after the divergence of the ancestor of this point source from any other known *S. aureus* species. The identification of CC75 as a group allied to but outside the global radiation of non-CC75 *S. aureus* has the potential to greatly assist the determination of the genetic basis for the global dominance of typical *S. aureus* and the associated burden of human disease. A complete answer will require more CC75 and non-CC75 *S. aureus* genome sequences. However, this study not only revealed that CC75 isolates lack staphyloxanthin but also that MSHR1132 possesses genomic islands, vSa $\alpha$  and vSa $\beta$ . Sequence similarities indicate that the presence of these elements is the ancestral state for CC75 and non-CC75 *S. aureus*. Therefore, the simplest model is that the acquisition of these elements long predated and did not initiate the radiation of non-CC75 *S. aureus*. However, we also note that these regions appear to undergo homologous recombination at much higher rates than core genes within the typical *S. aureus* population as phylogenies based on the allelic variation present in these islands is highly discordant with MLST data (Tsuru et al. 2006; Baba et al. 2008; Tsuru and Kobayashi 2008). These two lines of evidence can be reconciled by arguing that the recent radiation of typical *S. aureus* population has been accompanied by an increase in recombination within these regions. Indeed, it is possible that genes encoded on the islands themselves determine the degree to which they recombine. The islands contain genes for a type I restriction-modification system that have contributed to the subsequent evolution of distinct *S. aureus* lineages and controlled the degree of horizontal transfer between these lineages (Waldron and Lindsay 2006). Consistent with this, MSHR1132 has a unique *hsdS* gene sequence.

MSHR1132, and the CC75 lineage in general, presents a very interesting and challenging case study in bacterial systematics, particularly given the recent interest in identifying operational and conceptual species definitions (Achtman and Wagner 2008). The “lumpers” will point to the fact that CC75 is phenotypically essentially identical to typical *S. aureus*, with the exception of a lack of toxin production and possibly lower virulence, and this phenotypic similarity is consistent with the relatively conserved noncore genome and the absence of novel mobile elements. Furthermore, 16S rRNA data would, similar to classic biochemical criteria, unquestionably place CC75 within the main *S. aureus* population. On the other hand, the “splitters” will argue a strong case for the promotion of CC75 to a separate species or subspecies. In contrast to 16S rRNA, the divergence of the core

genome is striking and places CC75 almost midway between typical *S. aureus* and *S. epidermidis*. CC75 is therefore a clearly distinct lineage from that which has undergone recent global radiation to become the typical *S. aureus*. Following the proposal by Gevers et al. (2005) that core genome divergence can be a powerful way to delineate species, this observation alone should be sufficient to promote this lineage to at least new subspecies. The recent emphasis on the use of average nucleotide identity is also consistent with this view (Richter and Rossello-Mora 2009). Furthermore, the MLST data for 13 CC75 isolates revealed a maximum level of divergence of almost 1% even within this very limited sample. This is a respectable degree of diversity for a separate species and confirms that this lineage is far broader than a single “clonal complex.” Gene flow has also frequently been considered an important criterion for species delineation (Dykhuizen and Green 1991; Fraser et al. 2007). The presence of a CRISPR element, the conserved noncore, and the failure to find evidence for recombination both within the CC75 core genome and between CC75 and typical *S. aureus* all point to the genetic isolation of this lineage. Such taxonomic dilemmas will likely become increasingly common as genetic sampling of the microbiosphere intensifies. For example, a recent study of *Escherichia coli* diversity has revealed the CV lineage that is phenotypically *E. coli* but as phylogenetically divergent from other *E. coli* as is *Escherichia fergusonii* (Walk et al. 2009).

Finally, the apparent restricted distribution of this lineage and possible lower virulence (due to lack of pigment production and other virulence factors) points to an ecological difference. Although further work is needed to confirm this final point, in the end, we side with the splitters and argue that the high degree of core genome divergence outweighs the apparent phenotypic similarity and 16S rRNA uniformity. Serious consideration should therefore be given to formally reclassifying the CC75 lineage as a new species, named *S. argenteus* (silver staph), reflective of the absence of staphyloxanthin.

## Supplementary Material

Supplementary tables S1–S3 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

We are grateful to Jann Hennessey and Rob Baird from Royal Darwin Hospital Pathology for assisting us to acquire clinical isolates and for performing the VITEK 2 assays. This work was supported by an Australian National Health and Medical Research Council Postdoctoral Training Fellowship (508829) awarded to S.Y.C.T. and Translation Research on Combating Antimicrobial Resistance funding awarded to S.C.R. (FP7-HEALTH #223031). D.C.H. and M.T.G.H. had equal inputs into this study. Data deposition: The

chromosome and plasmid sequences reported have been deposited in EMBL with accession numbers FR821777 and FR821778, respectively.

## Literature Cited

- Achtman M, Wagner M. 2008. Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol*. 6:431–440.
- Altschul SF, et al. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403–410.
- Baba T, et al. 2002. Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet* 359:1819–1827.
- Baba T, et al. 2008. Genome sequence of *Staphylococcus aureus* strain Newman and comparative analysis of staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands. *J Bacteriol*. 190:300–310.
- Bateman A, et al. 2002. The Pfam protein families database. *Nucl Acids Res*. 30:276–280.
- Carver TJ, et al. 2005. ACT: the artemis comparison tool. *Bioinformatics* 21:3422–3423.
- Chua K, et al. 2010. Complete genome sequence of *Staphylococcus aureus* strain JKD6159, a unique Australian clone of ST93-IV community methicillin-resistant *Staphylococcus aureus*. *J Bacteriol*. 192:5556–5557.
- Clauditz A, et al. 2006. Staphyloxanthin plays a role in the fitness of *Staphylococcus aureus* and its ability to cope with oxidative stress. *Infect Immun*. 74:4950–4953.
- Cooper JE, Feil EJ. 2006. The phylogeny of *Staphylococcus aureus*—which genes make the best intra-species markers? *Microbiology*. 152:1297–1305.
- Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*. 5:e11147.
- Delcher AL, et al. 1999. Improved microbial gene identification with GLIMMER. *Nucl Acids Res*. 27:4636–4641.
- Diep BA, et al. 2006. Complete genome sequence of USA300, an epidemic clone of community-acquired methicillin-resistant *Staphylococcus aureus*. *Lancet* 367:731–739.
- Drancourt M, Raoult D. 2002. *rpoB* gene sequence-based identification of *Staphylococcus* species. *J Clin Microbiol*. 40:1333–1338.
- Dykhuizen DE, Green L. 1991. Recombination in *Escherichia coli* and the definition of biological species. *J Bacteriol*. 173:7257–7268.
- Enright MC, et al. 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol*. 38:1008–1015.
- Falquet L, et al. 2002. The PROSITE database, its status in 2002. *Nucl Acids Res*. 30:235–238.
- Fan F, et al. 2002. Defining and combating the mechanisms of triclosan resistance in clinical isolates of *Staphylococcus aureus*. *Antimicrob Agents Chemother*. 46:3343–3347.
- Feil EJ, et al. 2003. How clonal is *Staphylococcus aureus*? *J Bacteriol*. 185:3307–3316.
- Felsenstein J. 1989. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* 5:164–166.
- Fraser C, Hanage WP, Spratt BG. 2007. Recombination and the nature of bacterial speciation. *Science* 315:476–480.
- Frishman D, Mironov A, Mewes HW, Gelfand M. 1998. Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. *Nucl Acids Res*. 26:2941–2947.
- Gevers D, et al. 2005. Opinion: re-evaluating prokaryotic species. *Nat Rev Microbiol*. 3:733–739.

- Ghebremedhin B, Layer F, Konig W, Konig B. 2008. Genetic classification and distinguishing of *Staphylococcus* species based on different partial *gap*, 16S rRNA, *hsp60*, *rpoB*, *sodA*, and *tuf* gene sequences. *J Clin Microbiol.* 46:1019–1025.
- Gill SR, et al. 2005. Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. *J Bacteriol.* 187:2426–2438.
- Gillaspay AF, et al. 2006. The *Staphylococcus aureus* NCTC8325 genome. In: Fischetti V, et al editors Gram positive pathogens. Washington (DC): ASM Press. p 381–412.
- Golding GR, et al. 2010. Livestock-associated methicillin-resistant *Staphylococcus aureus* sequence type 398 in humans, Canada. *Emerg Infect Dis.* 16:587–594.
- Griffiths-Jones S, et al. 2003. Rfam: an RNA family database. *Nucl Acids Res.* 31:439–441.
- Haft DH, Selengut J, Mongodin EF, Nelson KE. 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol.* 1:e60.
- Herron-Olson L, Fitzgerald JR, Musser JM, Kapur V. 2007. Molecular correlates of host specialization in *Staphylococcus aureus*. *PLoS One.* 2:e1120.
- Highlander SK, et al. 2007. Subtle genetic changes enhance virulence of methicillin resistant and sensitive *Staphylococcus aureus*. *BMC Microbiol.* 7:99.
- Holden MT, et al. 2004. Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. *Proc Natl Acad Sci USA.* 101:9786–9791.
- Holden MT, et al. 2010. Genome sequence of a recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant of methicillin-resistant *Staphylococcus aureus*, sequence type 239 (TW). *J Bacteriol.* 192:888–892.
- Horvath P, Barrangou R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327:167–170.
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 23:254–267.
- Ito T, et al. 2004. Novel type V staphylococcal cassette chromosome *mec* driven by a novel cassette chromosome recombinase, *ccrC*. *Antimicrob Agents Chemother.* 48:2637–2651.
- Kishii K, et al. 2008. Recurrence of heterogeneous methicillin-resistant *Staphylococcus aureus* (MRSA) among the MRSA clinical isolates in a Japanese university hospital. *J Antimicrob Chemother.* 62:324–328.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 305:567–580.
- Kuroda M, et al. 2001. Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet* 357:1225–1240.
- Kuroda M, et al. 2005. Whole genome sequence of *Staphylococcus saprophyticus* reveals the pathogenesis of uncomplicated urinary tract infection. *Proc Natl Acad Sci USA.* 102:13272–13277.
- Larsen TS, Krogh A. 2003. EasyGene—a prokaryotic gene finder that ranks ORFs by statistical significance. *BMC Bioinformatics.* 4:21.
- Lindsay JA, Holden MT. 2006. Understanding the rise of the superbug: investigation of the evolution and genomic variation of *Staphylococcus aureus*. *Funct Integr Genomics.* 6:186–201.
- Liu CI, et al. 2008. A cholesterol biosynthesis inhibitor blocks *Staphylococcus aureus* virulence. *Science* 319:1391–1394.
- Liu GY, et al. 2005. *Staphylococcus aureus* golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. *J Exp Med.* 202:209–215.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucl Acids Res.* 25:955–964.
- Makarova KS, Wolf YI, van der Oost J, Koonin EV. 2009. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol Direct.* 4:29. doi:1745-6150-4-29 [pii]10.1186/1745-6150-4-29
- Marraffini LA, Sontheimer EJ. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322:1843–1845.
- Marshall JH, Wilmoth GJ. 1981. Pigments of *Staphylococcus aureus*, a series of triterpenoid carotenoids. *J Bacteriol.* 147:900–913.
- Martin D, Rybicki E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16:562–563.
- McDonald M, et al. 2006. Use of a single-nucleotide polymorphism genotyping system to demonstrate the unique epidemiology of methicillin-resistant *Staphylococcus aureus* in remote aboriginal communities. *J Clin Microbiol.* 44:3720–3727.
- Monecke S, et al. 2010. Characterisation of Australian MRSA strains ST75- and ST883-MRSA-IV and analysis of their accessory gene regulator locus. *PLoS One.* 5:e14025. doi:10.1371/journal.pone.0014025
- Mongkolrattanothai K, Boyle S, Murphy TV, Daum RS. 2004. Novel non-*mecA*-containing staphylococcal chromosomal cassette composite island containing *pbp4* and *tagF* genes in a commensal staphylococcal species: a possible reservoir for antibiotic resistance islands in *Staphylococcus aureus*. *Antimicrob Agents Chemother.* 48:1823–1836.
- Mwangi MM, et al. 2007. Tracking the in vivo evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proc Natl Acad Sci USA.* 104:9451–9456.
- Neoh HM, et al. 2008. Mutated response regulator *graR* is responsible for phenotypic conversion of *Staphylococcus aureus* from heterogeneous vancomycin-intermediate resistance to vancomycin-intermediate resistance. *Antimicrob Agents Chemother.* 52:45–53.
- Neves WA, Hubbe M. 2005. Cranial morphology of early Americans from Lagoa Santa, Brazil: implications for the settlement of the New World. *Proc Natl Acad Sci USA.* 102:18309–18314.
- Ng JW, et al. 2009. Phylogenetically distinct *Staphylococcus aureus* lineage prevalent among indigenous communities in northern Australia. *J Clin Microbiol.* 47:2295–2300. doi:JCM.00122-09 [pii]10.1128/JCM.00122-09.
- Nielsen H, Engelbrecht J, Brunak S, von Heijne G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10:1–6.
- Nimmo GR, et al. 2006. Methicillin-resistant *Staphylococcus aureus* in the Australian community: an evolving epidemic. *Med J Aust.* 184:384–388.
- Noto MJ, Kreiswirth BN, Monk AB, Archer GL. 2008. Gene acquisition at the insertion site for SCC<sub>mec</sub>, the genomic island conferring methicillin resistance in *Staphylococcus aureus*. *J Bacteriol.* 190:1276–1283.
- Olivier AC, et al. 2009. Role of *rsbU* and staphyloxanthin in phagocytosis and intracellular growth of *Staphylococcus aureus* in human macrophages and endothelial cells. *J Infect Dis.* 200:1367–1370. doi:10.1086/606012.

- Pearson T, et al. 2009. Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. *BMC Biol.* 7:78.
- Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA.* 85:2444–2448.
- Poyart C, Quesne G, Boumaila C, Trieu-Cuot P. 2001. Rapid and Accurate species-level identification of coagulase-negative *Staphylococci* by using the *sodA* gene as a target. *J Clin Microbiol.* 39:4296–4301.
- Richter M, Rossello-Mora R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA.* 106:19126–19131.
- Rosenstein R, et al. 2009. Genome analysis of the meat starter culture bacterium *Staphylococcus carnosus* TM300. *Appl Environ Microbiol.* 75:811–822.
- Ruimy R, et al. 2009. Comparisons between geographically diverse samples of carried *Staphylococcus aureus*. *J Bacteriol.* 191:5577–5583 doi:JB.00493-09 [pii]10.1128/JB.00493-09.
- Ruimy R, et al. 2010. Are host genetics the predominant determinant of persistent nasal *Staphylococcus aureus* carriage in humans? *J Infect Dis.* 202:924–934.
- Rutherford K, et al. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–945.
- Schijffelen MJ, Boel CH, van Strijp JA, Fluit AC. 2010. Whole genome analysis of a livestock-associated methicillin-resistant *Staphylococcus aureus* ST398 isolate from a case of human endocarditis. *BMC Genomics.* 11:376.
- Smyth DS, Robinson DA. 2009. Integrative and sequence characteristics of a novel genetic element, ICE6013, in *Staphylococcus aureus*. *J Bacteriol.* 191:5964–5975.
- Takeuchi F, et al. 2005. Whole-genome sequencing of *Staphylococcus haemolyticus* uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species. *J Bacteriol.* 187:7292–7308.
- Tong SY, et al. 2009. Community-associated strains of methicillin-resistant *Staphylococcus aureus* and methicillin-susceptible *S. aureus* in indigenous Northern Australia: epidemiology and outcomes. *J Infect Dis.* 199:1461–1470.
- Tong SY, et al. 2010. Clinical correlates of Pantone-Valentine leukocidin (PVL), PVL isoforms, and clonal complex in the *Staphylococcus aureus* population of Northern Australia. *J Infect Dis.* 202:760–769 doi:10.1086/655396.
- Touchon M, Rocha EP. 2010. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. *PLoS One.* 5:e11126.
- Tse H, et al. 2010. Complete genome sequence of *Staphylococcus lugdunensis* strain HKU09-01. *J Bacteriol.* 192:1471–1472.
- Tsuru T, Kobayashi I. 2008. Multiple genome comparison within a bacterial species reveals a unit of evolution spanning two adjacent genes in a tandem paralog cluster. *Mol Biol Evol.* 25:2457–2473.
- Tsuru T, et al. 2006. Evolution of paralogous genes: reconstruction of genome rearrangements through comparison of multiple genomes within *Staphylococcus aureus*. *Mol Biol Evol.* 23:1269–1285.
- van der Oost J, et al. 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci.* 34:401–407.
- Waldron DE, Lindsay JA. 2006. 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.* 188:5578–5585.
- Walk ST, et al. 2009. Cryptic lineages of the genus *Escherichia*. *Appl Environ Microbiol.* 75:6534–6544 doi:AEM.01262-09 [pii] 10.1128/AEM.01262-09.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
- Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 22:2472–2479.
- Zhang YQ, et al. 2003. Genome-based analysis of virulence genes in a non-biofilm-forming *Staphylococcus epidermidis* strain (ATCC 12228). *Mol Microbiol.* 49:1577–1593.

**Associate editor:** Emmanuelle Lerat