

2016

## Transmission of *Staphylococcus aureus* from humans to green monkeys in the Gambia as revealed by whole-genome sequencing

George Weinstock  
*Washington University School of Medicine in St. Louis*

et al

Follow this and additional works at: [https://digitalcommons.wustl.edu/open\\_access\\_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

---

### Recommended Citation

Weinstock, George and et al, "Transmission of *Staphylococcus aureus* from humans to green monkeys in the Gambia as revealed by whole-genome sequencing." *Applied and Environmental Microbiology*. 82,19. 5910-5917. (2016).  
[https://digitalcommons.wustl.edu/open\\_access\\_pubs/5293](https://digitalcommons.wustl.edu/open_access_pubs/5293)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact [vanam@wustl.edu](mailto:vanam@wustl.edu).

# Transmission of *Staphylococcus aureus* from Humans to Green Monkeys in The Gambia as Revealed by Whole-Genome Sequencing

Madikay Senghore,<sup>a,b</sup> Sion C. Bayliss,<sup>c</sup> Brenda A. Kwambana-Adams,<sup>a</sup> Ebenezer Foster-Nyarko,<sup>a</sup> Jainaba Manneh,<sup>a</sup> Michel Dione,<sup>a,\*</sup> Henry Badji,<sup>a</sup> Chinelo Ebruke,<sup>a</sup> Emma L. Doughty,<sup>b</sup> Harry A. Thorpe,<sup>c</sup> Anna J. Jasinska,<sup>d</sup> Christopher A. Schmitt,<sup>d,\*</sup> Jennifer D. Cramer,<sup>e</sup> Trudy R. Turner,<sup>f,g</sup> George Weinstock,<sup>h,\*</sup> Nelson B. Freimer,<sup>d</sup> Mark J. Pallen,<sup>b</sup> Edward J. Feil,<sup>c</sup> Martin Antonio<sup>a,b,i</sup>

Medical Research Council Unit, Fajara, The Gambia<sup>a</sup>; University of Warwick, Coventry, United Kingdom<sup>b</sup>; University of Bath, Bath, United Kingdom<sup>c</sup>; University of California Los Angeles, Los Angeles, California, USA<sup>d</sup>; American Military University and American Public University, Charles Town, West Virginia, USA<sup>e</sup>; University of Wisconsin—Milwaukee, Milwaukee, Wisconsin, USA<sup>f</sup>; University of the Free State, Bloemfontein, South Africa<sup>g</sup>; Washington University, St. Louis, Missouri, USA<sup>h</sup>; London School of Hygiene and Tropical Medicine, London, United Kingdom<sup>i</sup>

## ABSTRACT

*Staphylococcus aureus* is an important pathogen of humans and animals. We genome sequenced 90 *S. aureus* isolates from The Gambia: 46 isolates from invasive disease in humans, 13 human carriage isolates, and 31 monkey carriage isolates. We inferred multiple anthroponotic transmissions of *S. aureus* from humans to green monkeys (*Chlorocebus sabaeus*) in The Gambia over different time scales. We report a novel monkey-associated clade of *S. aureus* that emerged from a human-to-monkey switch estimated to have occurred 2,700 years ago. Adaptation of this lineage to the monkey host is accompanied by the loss of phage-carrying genes that are known to play an important role in human colonization. We also report recent anthroponotic transmission of the well-characterized human lineages sequence type 6 (ST6) and ST15 to monkeys, probably because of steadily increasing encroachment of humans into the monkeys' habitat. Although we have found no evidence of transmission of *S. aureus* from monkeys to humans, as the two species come into ever-closer contact, there might be an increased risk of additional interspecies exchanges of potential pathogens.

## IMPORTANCE

The population structures of *Staphylococcus aureus* in humans and monkeys in sub-Saharan Africa have been previously described using multilocus sequence typing (MLST). However, these data lack the power to accurately infer details regarding the origin and maintenance of new adaptive lineages. Here, we describe the use of whole-genome sequencing to detect transmission of *S. aureus* between humans and nonhuman primates and to document the genetic changes accompanying host adaptation. We note that human-to-monkey switches tend to be more common than the reverse and that a novel monkey-associated clade is likely to have emerged from such a switch approximately 2,700 years ago. Moreover, analysis of the accessory genome provides important clues as to the genetic changes underpinning host adaptation and, in particular, shows that human-to-monkey switches tend to be associated with the loss of genes known to confer adaptation to the human host.

*Staphylococcus aureus* is an important pathogen of humans, causing a range of conditions from serious invasive diseases such as meningitis, pneumonia, and bacteremia to less severe skin and soft tissue infections (1). *S. aureus* is among the top five most common causes of bacteremia in sub-Saharan Africa and the second leading cause of bacteremia in The Gambia (2–4). *S. aureus* thus poses a serious public health burden in The Gambia, yet little is known about the population structure and dynamics of this pathogen in sub-Saharan Africa. In other parts of the world, interest has focused on the role of nonhuman hosts (mostly livestock) as reservoirs of infection and drug resistance relevant to humans (5–7). In addition, it is clear that *S. aureus* can switch host species, sometimes resulting in adaptation to the new host and onward transmission in the new host species (8).

Interspecies transmission and adaptive host switching are known to occur between humans and nonhuman primates. Human-associated *S. aureus* lineages readily colonize and infect nonhuman primates in captivity and in the wild (9–12). In remote regions of Africa, wild monkeys are mainly colonized by *S. aureus* isolates belonging to uncharacterized clonal complexes that rarely colonize or infect humans, with one highly divergent clade isolated from monkeys in sub-Saharan Africa now classified as a new species, *Staphylococcus schweitzeri* (13, 14).

Received 19 May 2016 Accepted 13 July 2016

Accepted manuscript posted online 29 July 2016

**Citation** Senghore M, Bayliss SC, Kwambana-Adams BA, Foster-Nyarko E, Manneh J, Dione M, Badji H, Ebruke C, Doughty EL, Thorpe HA, Jasinska AJ, Schmitt CA, Cramer JD, Turner TR, Weinstock G, Freimer NB, Pallen MJ, Feil EJ, Antonio M. 2016. Transmission of *Staphylococcus aureus* from humans to green monkeys in The Gambia as revealed by whole-genome sequencing. *Appl Environ Microbiol* 82:5910–5917. doi:10.1128/AEM.01496-16.

**Editor:** J. Björkroth, University of Helsinki

Address correspondence to Mark J. Pallen, m.pallen@warwick.ac.uk.

\* Present address: Michel Dione, International Livestock Research Institute, Kampala, Uganda; Christopher A. Schmitt, Boston University, Boston, Massachusetts, USA; George Weinstock, The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, USA.

M.S. and S.C.B. are co-first authors who contributed equally to this article.

M.J.P., E.J.F., and M.A. are co-senior authors who contributed equally to this article.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.01496-16>.

Copyright © 2016 Senghore et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

The gain or loss of genes associated with mobile genetic elements is thought to be the primary driver of host adaptation following interhost transmission (15). Nonhuman hosts provide an environment for the acquisition of novel virulence and resistance determinants (16). For example, clones of methicillin-resistant *Staphylococcus aureus* (MRSA) from human-associated lineages such as clonal complex 5 (CC5), CC9 and sequence type 88 (ST88) have been reported in livestock (17) and livestock-associated MRSA lineages, most notably CC97 and CC398, have overcome the species barrier to infect humans (15, 18).

In The Gambia, increasing urbanization and tourism have meant that wild green monkeys have become habituated to humans, resulting in increased opportunities for interhost transmission of potential pathogens. In particular, free-ranging wild monkeys inhabit the Bijilo Forest Park which is close to the Senegambia tourist area and serves as a tourist attraction where locals and tourists go to visit the monkeys. Although feeding the animals is prohibited, people take bags of peanuts into the park and feed the monkeys by hand. Here, we describe the use of whole-genome sequencing to detect transmission of *S. aureus* between humans and nonhuman primates and to document the genetic changes accompanying host adaptation.

## MATERIALS AND METHODS

**Study isolates.** We pooled isolates from three previous studies that characterized *S. aureus* from monkeys, human carriage, and invasive disease by multilocus sequence typing (MLST) and antimicrobial susceptibility testing (19, 20). The first study was conducted in 2011 by the International Vervet Research Consortium on simian immunodeficiency virus (SIV) infection in green monkeys (*Chlorocebus sabaeus*) in The Gambia (20). Thus, we were able to collect nasopharyngeal swabs (NPS) and oropharyngeal swabs (OPS) from the monkeys. Eighty-two *S. aureus* isolates were cultured from 64 NPS and 63 OPS collected from 64 human-habituated wild monkeys in The Gambia. In the second study, 100 *S. aureus* strains were isolated from human NPS as part of a carriage study conducted between December 2005 and April 2005 in Sibanor (19). The third study analyzed a selection of 116 *S. aureus* strains isolated from archived clinical specimens of patients who reported to the Medical Research Council (MRC) clinic in Fajara with invasive bacterial disease between 2002 and 2010 (unpublished data). Table S1 and Fig. S4 in the supplemental material show the temporal spread of sampling in the different epidemiological classes and the spatial distribution of the sites where the samples were collected.

To isolate *S. aureus*, specimens were plated on mannitol salt agar (MSA) (Oxoid, Basingstoke, United Kingdom) and 5% sheep blood agar (BA) (Oxoid, Basingstoke, United Kingdom) plates. The specimens were incubated at 37°C for 24 h on BA plates and 48 h on MSA plates under aerobic conditions. Suspected *S. aureus* colonies were subcultured on BA plates for 24 h and confirmed by a coagulase test using the Slidex Staph kit (bioMérieux, Basingstoke, Hampshire, United Kingdom). Antimicrobial susceptibility testing was performed by the disk diffusion method on BA plates for the following antibiotics: penicillin, co-trimoxazole, tetracycline, chloramphenicol, gentamicin, cloxacillin, erythromycin, and ceftioxin (Oxoid, Basingstoke, United Kingdom). Results were interpreted according to the guidelines of the Clinical and Laboratory Standards Institute (CLSI) (21).

**DNA extraction and MLST analysis.** Genomic DNA was extracted from fresh overnight pure cultures of *S. aureus* strains using the Qiagen genomic DNA extraction kit (Qiagen, United Kingdom) according to the manufacturer's protocol. MLST was performed on *S. aureus* isolates targeting seven housekeeping genes (*aroE*, *ptA*, *glp*, *arcC*, *gmK*, *tpi*, and *yqiL*) as described previously (22). An eBURST (23) analysis was performed on all STs of *Staphylococcus aureus* in the MLST database (<http://saureus.mlst.net>).

STs were assigned to clonal complexes where they had 6 identical alleles with at least one other ST within the clonal complex. eBURST groups sequence types based on shared alleles, but it does not take into consideration the existing knowledge of the clonal population structure of *S. aureus*. As a result, in some instances, eBURST merged two or more clonal complexes that are well described in the literature into one eBURST group. In these cases, the clonal complex designations from the literature were used instead of the eBURST grouping in order to maintain consistency with the literature.

**Whole-genome sequencing, assembly, and annotation.** Ninety isolates were analyzed for whole-genome sequencing: 46 isolates from invasive disease in humans, 13 human carriage isolates, and 31 monkey carriage isolates. These isolates included at least one representative of all the major clonal complexes inferred from MLST. Whole-genome sequencing was carried out on the Illumina MiSeq system with the Nextera X preparation kit. *De novo* contigs were generated for each genome using SPAdes (kmers 21, 33, 55, 77, 99, and 127) (24). Contigs shorter than 300 bp and with kmer coverage of <2 were removed from the assemblies. Coding sequences (CDSs) were predicted and annotated by Prokka (version 1.11) (25).

**MLST from the whole genome.** To determine whether any isolates might have been misassigned in the MLST or whole-genome sequencing workflows, we used the draft genomes to predict multilocus sequence types. Two methods were employed: an assembly-based approach and a mapping-based approach. Assemblies were subjected to BLAST searches against all of the alleles in the MLST database (<http://saureus.mlst.net>). Alleles were called whenever there was 100% sequence coverage and 100% nucleotide identity. The mapping approach used SRST2 to map the reads to the seven housekeeping loci and identify the ST (26). The two sets of results were combined and manually curated. STs presented in Table S1 in the supplemental material are STs inferred from the whole genome that share at least 5 identical loci with the MLST results from the laboratory and belong to the same clonal complex. Eleven isolates were excluded from further analysis because the ST derived from the conventional MLST analysis differed from the ST inferred from the whole genome by more than two loci, leading us to believe that isolates had been misassigned in one analysis or the other.

**Phylogenetic analysis.** Sequencing reads were mapped to the EMRSA15 HO 5096 0412 reference genome (accession number HE681097) using SMALT (<http://www.sanger.ac.uk/resources/software/smalt/>). Single nucleotide polymorphisms (SNPs) were called using SAMtools 0.1.18, the Genome Analysis Toolkit (GATK), and in-house scripts (27, 28). SNPs were called from the core genome after exclusion of known repeat regions, insertion sequences, and known horizontally acquired elements. An approximate maximum likelihood phylogenetic tree was reconstructed using FastTree (29). Where appropriate, we included a reference genome belonging to each CC in the phylogenetic analysis. Isolates of *Staphylococcus argenteus* and *Staphylococcus schweitzeri* were included to ensure correct species identification (14).

**Accessory genome analysis.** Representative core and accessory genomes for the data set were identified using Roary on default settings (30). A pairwise matrix was generated, showing the proportion of shared accessory CDSs (see Fig. 2). Finding the set differences between clade 2 and the remaining isolates identified the accessory genome content specific to clade 2. CDSs found in more than 20 of the 22 clade 2 isolates were considered to potentially contribute to monkey-specific host adaptation.

The presence or absence of a number of genes associated with virulence in *S. aureus* was inferred by a BLAST search of the assembled genomes. Nucleotide sequences for virulence-associated genes alpha-hemolysin (*hla*), beta-hemolysin (*hlyB*), delta-hemolysin (*hlyD*), staphylococcal enterotoxins A (*seA*), B (*seB*), C (*seC*), G (*seG*), H (*seH*), and I (*seI*), toxic shock syndrome toxin gene 1 (*tstI*), and Panton-Valentine leukocidin (PVL) genes (*lukF-PV* and *lukS-PV*) were identified from the Virulence Factors of Pathogenic Bacteria database (<http://www.mgc.ac.cn/VFs/>). The presence of the virulence genes in isolates used in this study was

identified by using BLAST against the whole-genome assemblies with a cutoff of >90% base identity and length similarity to the reference gene.

**Genomic divergence.** To visualize genetic divergence across the genomes in clade 2, monkey isolates H7, F2, G2, G11, F7, and H10 were compared the finished closed genome USA300\_FPR3757 using BLAST Ring Image Generator (BRIG) (31). Query genomes represented subclusters within the monkey-associated clade 2. USA300\_FPR3757 was selected for comparison because of the presence of  $\nu$ Sa $\alpha$  and  $\nu$ Sa $\beta$  and the high quality of its annotation. The lower bounds of nucleotide similarity were set to 85%. Well-characterized mobile genetic elements were annotated for comparison.

**Ethical approval.** The Gambia Government/MRC Joint Ethics Committee approved the carriage study and the sampling of biological samples from the green monkeys. The Gambia Government/MRC Joint Ethics Committee gave subsequent approval to send genomic DNA of 96 *S. aureus* isolates to the University of Warwick, United Kingdom, for whole-genome sequencing.

**Accession number(s).** The genome sequence data from this study have been uploaded to the European Nucleotide Archive under the study accession number PRJEB12419.

## RESULTS

**Genomic analysis.** We analyzed 90 genomes of *Staphylococcus aureus*, representing 46 isolates from human invasive disease, 13 human carriage isolates, and 31 monkey pharyngeal isolates (see Table S1 in the supplemental material). Draft assemblies ranged in size from 2.7 to 3.4 Mb with a range of 2,462 to 2,921 coding sequences. When mapped to the EMRSA-15 reference genome, the mean average coverage was 61.5-fold (range, 12.5- to 126.2-fold).

**Phylogenetic analysis.** Single nucleotide polymorphisms (SNPs) were called from 232,276 sites in the core genome after exclusion of all known repeat regions and transposable mobile elements. We created a maximum-likelihood phylogenetic tree linking our isolates with reference strains. This tree resolved five major *S. aureus* clades that are consistent with previous species-wide phylogenetic analyses (Fig. 1; see also Fig. S2 in the supplemental material). Clade 1 encompasses the well-described CCs 8, 1, 5, 15, and 25. Clade 2 is a novel monkey-associated clade encompassing six subclusters. Clade 3 encompasses CC30 and CC45. Clade 4 corresponds to CC121 (22) and various animal-associated reference genomes. Finally, clade 5 corresponds to ST152 and related genotypes.

Current estimates of intraclonal mutation rates are consistent in *S. aureus* (see the supplemental material in reference 32). Based on the intraclonal mutation rate for *S. aureus* of  $\sim 2 \times 10^{-6}$  per site per year (32) and a core genome size of  $\sim 2.5$  Mb, we were able to estimate an upper bound date for the transmission events. The short-term mutation rate equates to about 5 SNPs per year. Thus, simply dividing the total number of SNPs between any pair of contemporary *S. aureus* isolates by 10 gives the approximate number of years since they shared a common ancestor.

The single monkey-derived ST15 isolate differs by only 71 core genome SNPs from the most closely related human-derived ST15. Assuming equal rates of mutation in both lineages, this represents approximately 7 years divergence. However, in this case, the estimated time of interhost transmission is likely to represent a misleadingly high upper bound, as we are very unlikely to have sampled the human-associated relative closest to the transmitted strain. Furthermore, the fact that ST15 was only found once in monkeys and that there is no evidence of onward transmission is consistent with a very recent transmission event.

We identified a cluster of seven ST6 isolates derived from monkeys. An interrogation of the MLST database ([www.mlst.net](http://www.mlst.net)) revealed that this genotype is occasionally recovered from cases of bovine mastitis but is predominantly human associated. Therefore, the cluster of seven ST6 isolates recovered from monkeys is also likely to represent recent anthroponotic transmission. The closest pairwise SNP distance between a monkey ST6 isolate and the ST6 reference strain was 311 SNPs, suggesting that the host switch happened no more than 3 decades ago. The most divergent pair of monkey-derived ST6 isolates differed by 270 SNPs, which also equates to around 3 decades of divergence. However, if this cluster represents multiple transmission events from humans to monkeys, then the resulting host switches might have occurred more recently.

The third host switching event evident from our analysis is the much more ancient event that gave rise to the diverse monkey-associated clade 2, which was characterized by isolates with novel STs not found in humans. As this clade is nested within predominantly human-associated lineages (Fig. 1), it is likely to have resulted from an ancient human-to-monkey transmission event. The smallest SNP difference between a clade 2 isolate and a non-clade 2 isolate (in this case the reference MSSA476) is 26,968 SNPs, which equates to  $\sim 2,700$  years of divergence. This may be an underestimate, since our rate estimate is based on divergence over much shorter time scales (which is likely to be more rapid due to the lag time of purifying selection).

**Accessory genome variation and host adaptation.** Genomes corresponding to the same clonal complex also share similar accessory gene content (Fig. 2). Moving out to a broader phylogenetic scale, related CCs share more similar accessory gene content than unrelated CCs (e.g., the large square corresponding to the CC15/1/8/97/6/25/5 clade). This confirms previous observations using microarray data, which led to the concept of the “core variable” genome, meaning those genes that are stably present or absent within a given clonal complex but vary in their distribution between clonal complexes (33).

Variations in accessory gene contents were observed within the different clade 2 subgroups (Fig. 3). Regions of difference mostly represented horizontally acquired elements such as the genomic islands  $\nu$ Sa $\alpha$  and  $\nu$ Sa $\beta$ , *SCCmec*, and phage and pathogenicity islands. Genes present in  $\nu$ Sa $\alpha$  but absent or in low frequency within clade 2 include a variant of *tst1* (toxic shock syndrome precursor) and genes encoding two superantigen-like proteins, two putative leukocidins, and a 65-kDa membrane protein. The *ltrA* gene was present in all strains except those in clade 2 and the CC152 strains of clade 5 (see the supplemental material). This gene encodes a “low temperature requirement” protein (Pfam: PF06772.5, COG4292) found to be essential for growth at low temperatures (4°C) in *Listeria monocytogenes* (34), but its function in *S. aureus* is unknown.

The *spl* operon resides in the  $\nu$ Sa $\beta$  pathogenicity island and encodes extracellular serine proteases. The distribution of genes within this operon is consistent with a role in host adaptation. The variants of the *splA* to *splE* genes present in the USA300\_FPR3757 reference are missing in all clade 2 isolates. However, clade 2 isolates contain novel variants of *spl* genes that are missing in all other isolates.

The immune evasion cluster (IEC1) proteins *sak*, *scn*, and *chp*, which are harbored on the phage  $\phi$ SA3 (Fig. 3), are absent from clade 2. All three genes are absent in the ST6 cluster, with the

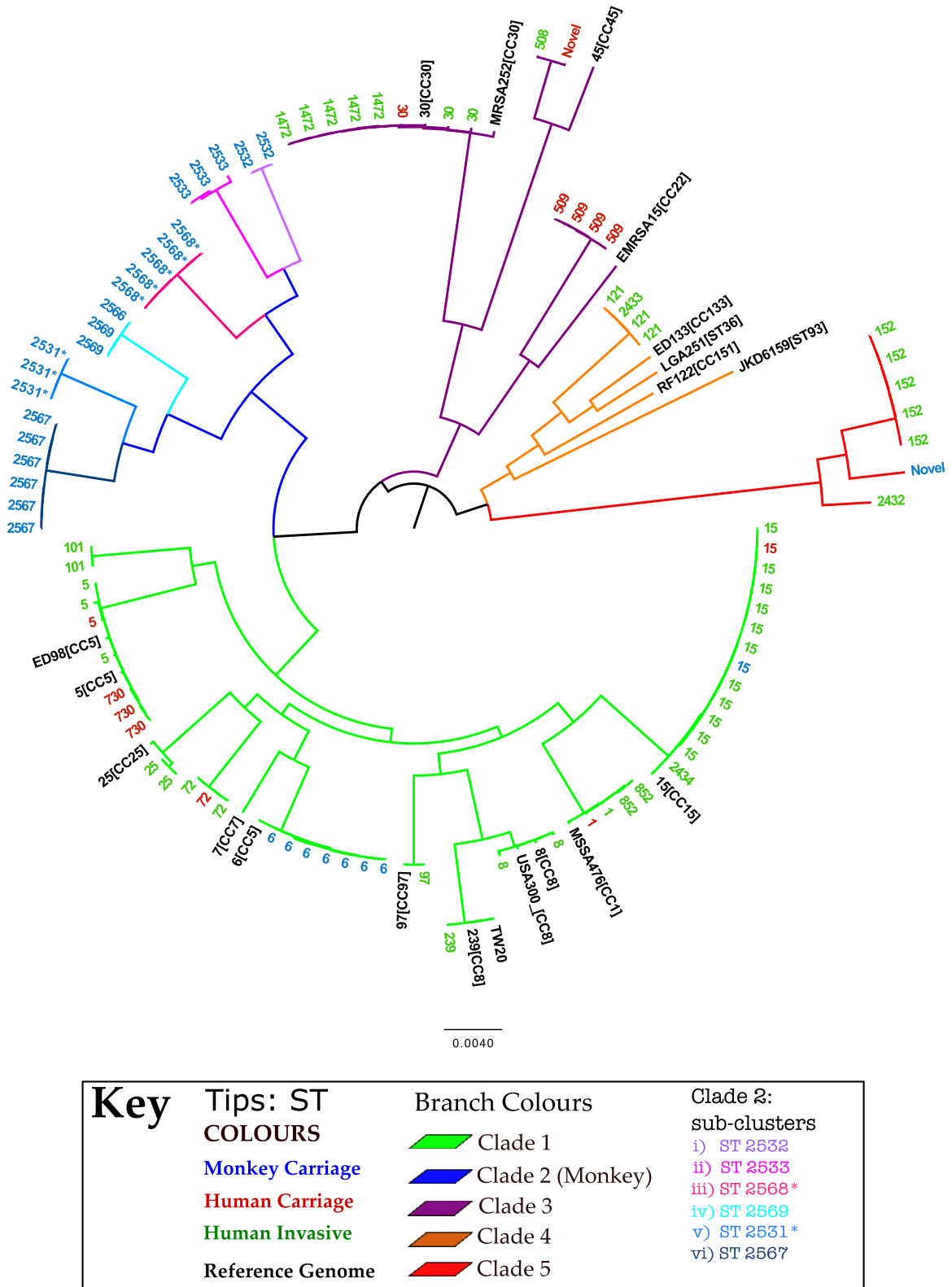


FIG 1 A maximum likelihood phylogenetic tree showing 5 major clades; branches are colored based on clade assignment. Tips are annotated by ST and colored by host. Reference genomes are annotated by name and CC and colored black. \*, single locus variant of given ST.

Downloaded from http://aem.asm.org/ on October 4, 2016 by Washington University in St. Louis

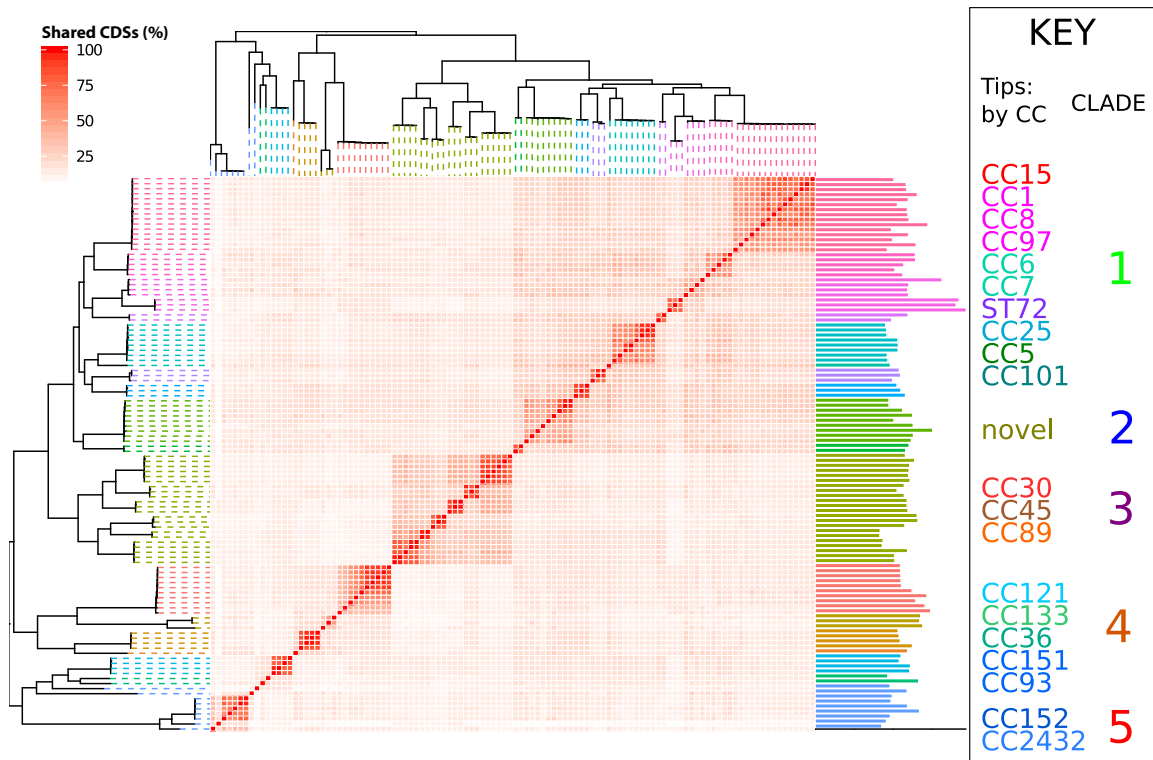


FIG 2 A heat map nested into the phylogeny showing the proportion of accessory genes that are shared between isolates, determined by a pairwise comparison. Branches on the phylogeny are colored based on the clonal complex. Darker squares in the heat map indicate higher shared accessory genome content. To the right of the heat map is a bar chart showing the number of accessory genes for each genome.

exception of *sak* and *scn* present in isolates E3 and E8. Furthermore, *sak* and *scn* are absent in SA29 (the ST15 monkey-derived isolate), and all three genes are absent in the monkey-derived CC152 isolate F5. This last example is particularly notable as the human-associated CC152 isolates in our collection contain these genes (see Data Set S1 and Fig. S1 in the supplemental material). The IEC1 genes are known to be associated only with human isolates and thus are thought to be involved in host-specific functions (35–37). Our analysis also confirms the absence of these genes in the animal-derived reference genomes ED98 (chicken), LG251 (cow), ED133 (sheep), and RF122 (cow) (35).

The *lytN* gene encodes a murein hydrolase thought to contribute to the release of protein A (a major immunoglobulin binding protein) from the cell surface by the removal of sugars (38). We note two variants of this gene in our data. Variant *lytN* is present in 18/22 clade 2 isolates but in only 10/88 non-clade 2 isolates (4 of which are from the monkey-associated ST6 cluster). Variant *lytN\_2* is absent from all 22 clade 2 isolates but present in 68/88 of all other isolates.

The Pantone-Valentine leukocidin (PVL) genes (*lukF-PV* and *lukS-PV*) were absent from all monkey isolates in the data set. This observation supports the adaptation to the new host since it works in humans but not monkeys (36). The staphylococcal enterotoxins A (*seA*), B (*seB*), C (*seC*), G (*seG*), H (*seH*), and I (*seI*), were absent from all monkey isolates. The exception was the presence of enterotoxin A (*seA*) in two ST6 monkey isolates. Beta-hemolysin (*hlyB*) was absent from all human isolates except the ST152 human isolates.

**Antibiotic resistance profiles.** We found no evidence of me-

thicillin-resistant *S. aureus* (MRSA) among our staphylococcal isolates from monkeys in The Gambia. One monkey isolate (G9) was reported as methicillin resistant by the cefoxitin disc diffusion test, but the Etest confirmed that it was susceptible to methicillin. In addition, no genomic evidence of methicillin resistance was found when the genome was analyzed using Mykrobe (39). Two invasive disease isolates were confirmed to be MRSA through phenotypic testing; the *mecA* gene was present in both isolates. To our knowledge this is the first report of MRSA causing human invasive disease in The Gambia.

## DISCUSSION

Using whole-genome sequencing, we infer multiple human-to-monkey transmission events, but no evidence of monkey-to-human transmission. This observation is consistent with the report by Schaumburg et al., who used MLST and *spa* typing to compare human and monkey staphylococcal isolates from three African countries, Côte d'Ivoire, Gabon, and Democratic Republic of Congo. Their findings revealed numerous examples of human-to-monkey transmission but no evidence of the reverse (13).

A consistent picture of a clonal population structure, in which closely related strains cluster into several widespread clonal complexes (CCs) that are clearly delineated from each other, has emerged from our data. The majority of *S. aureus* colonizing in monkeys was due to novel lineages that formed clade 2 in the phylogenetic tree (Fig. 1). The phylogenetic placement of clade 2 suggests that it arose from an ancient human-to-monkey transmission event. This clade is believed to have diverged from human *S. aureus* ~2,700 years ago, long before modern human popula-

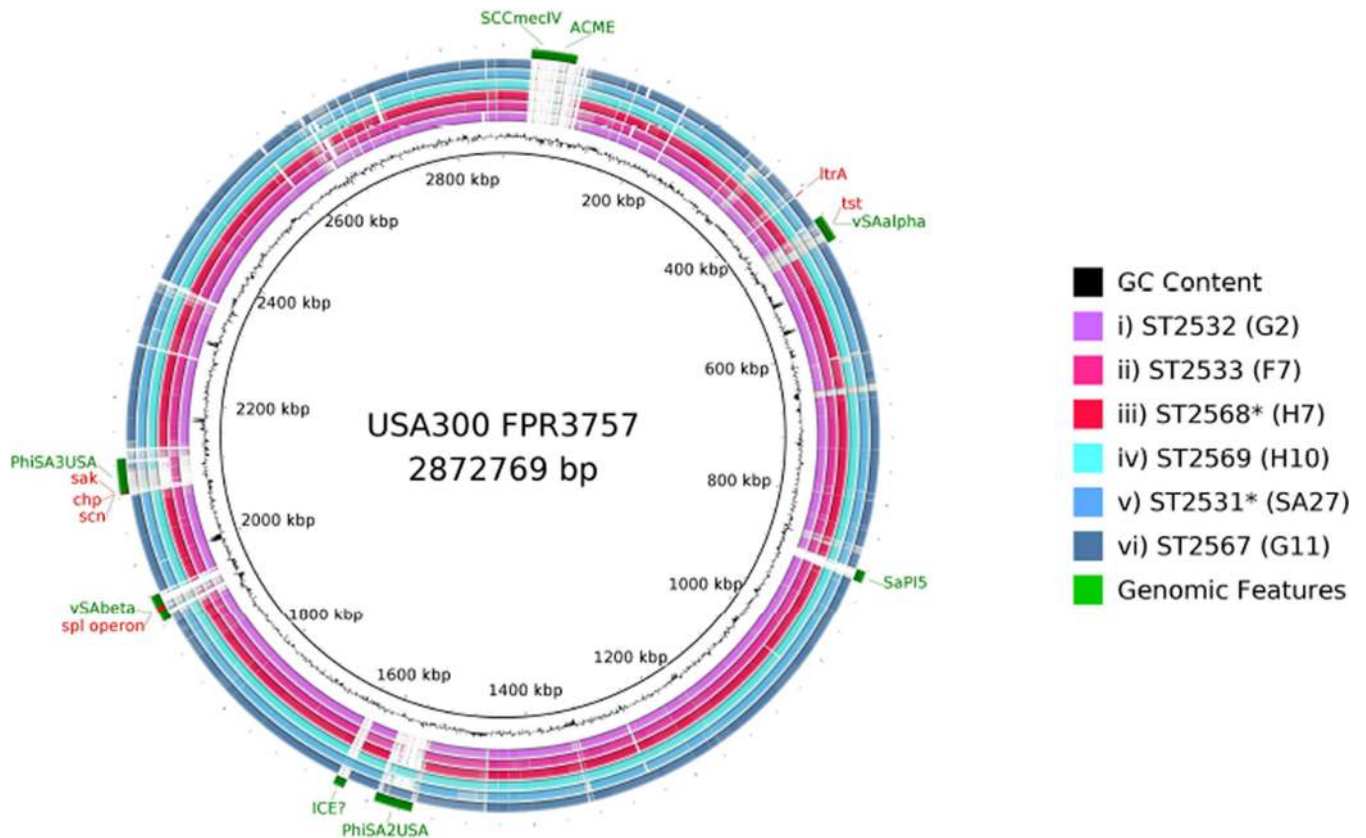


FIG 3 BRIG. Whole-genome sequence analysis and comparison of USA300\_FPR3757 to isolate a representative from each of the six subclusters in the monkey-associated clade 2. Circular diagram of the USA300 chromosome showing (from inner to outer), percent G+C, GC skew, and the homology based on BLASTN analysis of USA300\_FPR3757 to H7, F2, G2, G11, F7, and H10. The sequence similarity ranges from 85% to 100%, as regions of homology of <85% were excluded. The outermost circle shows the location of large horizontally acquired pathogenicity islands and well-characterized phages present in USA300. \*, single locus variant of given ST.

tion expansion and its ecological consequences. Over time, this clade appears to have adapted to the monkey host and has undergone clonal expansion.

The more recent transmission of human-associated lineages ST15 and ST6 to monkeys is believed to be a product of human encroachment into the natural habitat of monkeys and probably a result of transfer of bacteria from human hands to food, which is then fed to the monkeys. These two STs have been previously detected in African monkeys from remote regions of sub-Saharan Africa (13).

Analysis of the distribution of accessory genes between the monkey- and human-associated isolates confirmed the previous suggestion that genes carried on mobile genetic elements play a key role in host adaptation (40). Of particular note is the roles of the two genomic islands  $\nu$ Sa $\alpha$  and  $\nu$ Sa $\beta$  that encode superantigens, lipoproteins, and proteases. Gene contents within these islands differ markedly between strains as these islands recombine at high rates and are transferable by transducing phage particles (40), so that they have been referred to as “enterotoxin nurseries” (41, 42).

Gene loss may be as important as gene acquisition in the context of host adaptation. For example, isolates recovered from non-human hosts often harbor truncated variants of surface proteins present within closely related human isolates (43, 44). The phage-borne genes *chp*, *sac*, and *scn* that constitute the immune evasion

cluster IEC1 have been previously noted to be exclusively associated with human-associated isolates, and our data are consistent with this view. Assuming that gene loss is an evolutionarily more parsimonious event than gene gain, this striking association may help to explain why *S. aureus* anthroponoses are more common than zoonosis, although we note not all human isolates harbor these genes (for example, they are absent within CC15). The association between host and variants of the *spl* operon is intriguing since it encodes serine proteases, as well as genes (*lytN*) involved in the processing of the major surface antigen staphylococcal protein A.

Reassuringly, we find no evidence of transmission of *S. aureus* from monkeys to humans. An analysis of MLST data has shown that for *S. aureus* there are generally higher rates of anthroponoses (human-to-animal transmission,  $n = 13$ ) than zoonosis (animal-to-human transmission,  $n = 2$ ) (8).

**Limitations of this study.** In this study, we did not perform *de novo* sample collection over a standard harmonized time frame but instead made use of existing sets of isolates collected at various times and in various places within The Gambia. Therefore, it seems unlikely that the humans in closest contact with the monkeys will have been sampled as part of this study. This has led us to temper our conclusions on the frequency of and direction of transmission, which could only be established in detail by a longitudinal study.

## ACKNOWLEDGMENTS

The Department of Parks and Wildlife Management, Ministry of Forestry and the Environment, The Gambia, enabled sample collection from free-ranging monkeys. Samples used in this study were collected as a part of the Systems Biology Sample Repository. We thank the staff of MRC The Gambia Unit for help with organizing field sample collection and providing administrative support and transportation, in particular, Sanneh Mamkumba for administrative help, Ousman Secka for help with supplies and sample storage and shipment, and drivers Ousman Bah and Lamin Gibba. We gratefully acknowledge the expertise and assistance of Oliver (Pess) Morton, Ebou Jarjou, and Katherine Camfield during the field work and Ben Kigbu and Toye Adegboye for veterinary care. This work benefitted from use of the MRC Cloud Infrastructure for Microbial Bioinformatics (CLIMB), funded by grant MR/L015080/1. We thank Gemma Kay for help in setting up sequencing on the Illumina MiSeq.

Sample collection from the monkeys was performed through the UCLA Systems Biology Sample Repository funded by NIH grants R01RR016300 and R01OD010980 to N.B.F. S.C.B. and E.J.F. are supported by a grant from the UKCRC Translational Infection Research (TIR) Initiative and the Medical Research Council (grant G1000803) with contributions from the Biotechnology and Biological Sciences Research Council, the National Institute for Health Research on behalf of the Department of Health, and the Chief Scientist Office of the Scottish Government Health Directorate. H.A.T. is funded by a University of Bath Research Studentship.

We declare no conflicts of interest.

## FUNDING INFORMATION

This work, including the efforts of Nelson Freimer, was funded by HHS | National Institutes of Health (NIH) (R01RR016300 and R01OD010980). This work, including the efforts of Sion C. Bayliss, Mark J. Pallan, and Edward J. Feil, was funded by Medical Research Council (MRC) (G1000803 and MR/L015080/1).

## REFERENCES

- Tong SY, Davis JS, Eichenberger E, Holland TL, Fowler VG, Jr. 2015. *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clin Microbiol Rev* 28:603–661. <http://dx.doi.org/10.1128/CMR.00134-14>.
- Musiime V, Cook A, Bakeera-Kitaka S, Vhembo T, Lutakome J, Keishanyu R, Prendergast AJ, Lubwama S, Robertson V, Hughes P, Nathoo K, Munderi P, Klein N, Musoke P, Gibb DM, ARROW Trial Team. 2013. Bacteremia, causative agents and antimicrobial susceptibility among HIV-1-infected children on antiretroviral therapy in Uganda and Zimbabwe. *Pediatr Infect Dis J* 32:856–862. <http://dx.doi.org/10.1097/INF.0b013e31828c3991>.
- Hill PC, Onyema CO, Ikumapayi UN, Secka O, Ameyaw S, Simmonds N, Donkor SA, Howie SR, Tapgun M, Corrah T, Adegbola RA. 2007. Bacteremia in patients admitted to an urban hospital in West Africa. *BMC Infect Dis* 7:2. <http://dx.doi.org/10.1186/1471-2334-7-2>.
- Mulholland EK, Ogunlesi OO, Adegbola RA, Weber M, Sam BE, Palmer A, Manary MJ, Secka O, Aidoo M, Hazlett D, Whittle H, Greenwood BM. 1999. Etiology of serious infections in young Gambian infants. *Pediatr Infect Dis J* 18:S35–S41. <http://dx.doi.org/10.1097/00006454-199910001-00007>.
- Fluit AC. 2012. Livestock-associated *Staphylococcus aureus*. *Clin Microbiol Infect* 18:735–744. <http://dx.doi.org/10.1111/j.1469-0691.2012.03846.x>.
- Huber H, Koller S, Giezendanner N, Stephan R, Zweifel C. 2010. Prevalence and characteristics of methicillin-resistant *Staphylococcus aureus* in humans in contact with farm animals, in livestock, and in food of animal origin, Switzerland, 2009. *Euro Surveill* 15(21):pii=19542. <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19542>.
- Price LB, Stegger M, Hasman H, Aziz M, Larsen J, Andersen PS, Pearson T, Waters AE, Foster JT, Schupp J, Gillece J, Driebe E, Liu CM, Springer B, Zdvoc I, Battisti A, Franco A, Zmudzki J, Schwarz S, Butaye P, Jouy E, Pomba C, Porrero MC, Ruimy R, Smith TC, Robinson DA, Weese JS, Arriola CS, Yu F, Laurent F, Keim P, Skov R, Aarestrup FM. 2012. *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. *mBio* 3(1):e00305-11. <http://dx.doi.org/10.1128/mBio.00305-11>.
- Shepherd MA, Fleming VM, Connor TR, Corander J, Feil EJ, Fraser C, Hanage WP. 2013. Historical zoonoses and other changes in host tropism of *Staphylococcus aureus*, identified by phylogenetic analysis of a population dataset. *PLoS One* 8:e62369. <http://dx.doi.org/10.1371/journal.pone.0062369>.
- Schaumburg F, Mugisha L, Peck B, Becker K, Gillespie TR, Peters G, Leendertz FH. 2012. Drug-resistant human *Staphylococcus aureus* in sanctuary apes pose a threat to endangered wild ape populations. *Am J Primatol* 74:1071–1075. <http://dx.doi.org/10.1002/ajp.22067>.
- Nagel M, Dischinger J, Turck M, Verrier D, Oedenkoven M, Ngoubangoye B, Le Flohic G, Drexler JF, Bierbaum G, Gonzalez JP. 2013. Human-associated *Staphylococcus aureus* strains within great ape populations in Central Africa (Gabon). *Clin Microbiol Infect* 19:1072–1077. <http://dx.doi.org/10.1111/1469-0691.12119>.
- Lee JI, Kim KS, Oh BC, Kim NA, Kim IH, Park CG, Kim SJ. 2011. Acute necrotic stomatitis (noma) associated with methicillin-resistant *Staphylococcus aureus* infection in a newly acquired rhesus macaque (*Macaca mulatta*). *J Med Primatol* 40:188–193. <http://dx.doi.org/10.1111/j.1600-0684.2011.00470.x>.
- García A, Nambiar PR, Marini RP, Fox JG. 2009. Staphylococcal meningoenzephalitis, nematodiasis, and typhlocolitis in a squirrel monkey (*Saimiri sciureus*). *J Med Primatol* 38:377–381. <http://dx.doi.org/10.1111/j.1600-0684.2009.00363.x>.
- Schaumburg F, Pauly M, Anoh E, Mossoun A, Wiersma L, Schubert G, Flammen A, Alabi AS, Muyembe-Tamfum JJ, Grobusch MP, Karhemere S, Akoua-Koffi C, Couacy-Hymann E, Kremsner PG, Mellmann A, Becker K, Leendertz FH, Peters G. 2015. *Staphylococcus aureus* complex from animals and humans in three remote African regions. *Clin Microbiol Infect* 21:345.e1–8. <http://dx.doi.org/10.1016/j.cmi.2014.12.001>.
- Tong SY, Schaumburg F, Ellington MJ, Corander J, Pichon B, Leendertz F, Bentley SD, Parkhill J, Holt DC, Peters G, Giffard PM. 2015. Novel staphylococcal species that form part of a *Staphylococcus aureus*-related complex: the non-pigmented *Staphylococcus argenteus* sp. nov. and the non-human primate-associated *Staphylococcus schweitzeri* sp. nov. *Int J Syst Evol Microbiol* 65:15–22. <http://dx.doi.org/10.1099/ijs.0.062752-0>.
- Spoor LE, McAdam PR, Weinert LA, Rambaut A, Hasman H, Aarestrup FM, Kearns AM, Larsen AR, Skov RL, Fitzgerald JR. 2013. Livestock origin for a human pandemic clone of community-associated methicillin-resistant *Staphylococcus aureus*. *mBio* 4(4):e00356-13. <http://dx.doi.org/10.1128/mBio.00356-13>.
- Sung JM, Lloyd DH, Lindsay JA. 2008. *Staphylococcus aureus* host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiology* 154:1949–1959. <http://dx.doi.org/10.1099/mic.0.2007/015289-0>.
- Köck R, Schaumburg F, Mellmann A, Koksäl M, Jurke A, Becker K, Friedrich AW. 2013. Livestock-associated methicillin-resistant *Staphylococcus aureus* (MRSA) as causes of human infection and colonization in Germany. *PLoS One* 8:e55040. <http://dx.doi.org/10.1371/journal.pone.0055040>.
- Lewis HC, Molbak K, Reese C, Aarestrup FM, Selchau M, Sorum M, Skov RL. 2008. Pigs as source of methicillin-resistant *Staphylococcus aureus* CC398 infections in humans, Denmark. *Emerg Infect Dis* 14:1383–1389. <http://dx.doi.org/10.3201/eid1409.071576>.
- Ebruke C, Dione MM, Walter B, Worwui A, Adegbola RA, Roca A, Antonio M. 2016. High genetic diversity of *Staphylococcus aureus* strains colonising the nasopharynx of Gambian villagers before widespread use of pneumococcal conjugate vaccines. *BMC Microbiol* 16:38. <http://dx.doi.org/10.1186/s12866-016-0661-3>.
- Clinical and Laboratory Standards Institute. 2008. Performance standards for antimicrobial susceptibility testing; 17th informational supplement. CLSI M100-S17. Clinical and Laboratory Standards Institute, Wayne, PA.
- Ma D, Jasinska AJ, Feyertag F, Wijewardana V, Kristoff J, He T, Raetz K, Schmitt CA, Jung Y, Cramer JD, Dione M, Antonio M, Tracy R, Turner T, Robertson DL, Pandrea I, Freimer N, Apetrei C, International Vervet Research Consortium. 2014. Factors associated with simian immunodeficiency virus transmission in a natural African nonhuman primate host in the wild. *J Virol* 88:5687–5705. <http://dx.doi.org/10.1128/JVI.03606-13>.
- Feil EJ, Cooper JE, Grundmann H, Robinson DA, Enright MC, Berendt



- T, Peacock SJ, Smith JM, Murphy M, Spratt BG, Moore CE, Day NPJ. 2003. How clonal is *Staphylococcus aureus*? *J Bacteriol* 185:3307–3316. <http://dx.doi.org/10.1128/JB.185.11.3307-3316.2003>.
23. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* 186: 1518–1530. <http://dx.doi.org/10.1128/JB.186.5.1518-1530.2004>.
  24. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.
  25. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <http://dx.doi.org/10.1093/bioinformatics/btu153>.
  26. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE. 2014. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 6:90. <http://dx.doi.org/10.1186/s13073-014-0090-6>.
  27. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993. <http://dx.doi.org/10.1093/bioinformatics/btr509>.
  28. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303. <http://dx.doi.org/10.1101/gr.107524.110>.
  29. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <http://dx.doi.org/10.1371/journal.pone.0009490>.
  30. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31:3691–3693. <http://dx.doi.org/10.1093/bioinformatics/btv421>.
  31. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. <http://dx.doi.org/10.1186/1471-2164-12-402>.
  32. Uhlemann AC, Dordel J, Knox JR, Raven KE, Parkhill J, Holden MT, Peacock SJ, Lowy FD. 2014. Molecular tracing of the emergence, diversification, and transmission of *S. aureus* sequence type 8 in a New York community. *Proc Natl Acad Sci U S A* 111:6738–6743. <http://dx.doi.org/10.1073/pnas.1401006111>.
  33. Lindsay JA, Moore CE, Day NP, Peacock SJ, Witney AA, Stabler RA, Husain SE, Butcher PD, Hinds J. 2006. Microarrays reveal that each of the ten dominant lineages of *Staphylococcus aureus* has a unique combination of surface-associated and regulatory genes. *J Bacteriol* 188:669–676. <http://dx.doi.org/10.1128/JB.188.2.669-676.2006>.
  34. Zheng W, Kathariou S. 1995. Differentiation of epidemic-associated strains of *Listeria monocytogenes* by restriction fragment length polymorphism in a gene region essential for growth at low temperatures (4 degrees C). *Appl Environ Microbiol* 61:4310–4314.
  35. McCarthy AJ, Lindsay JA. 2013. *Staphylococcus aureus* innate immune evasion is lineage-specific: a bioinformatics study. *Infect Genet Evol* 19:7–14. <http://dx.doi.org/10.1016/j.meegid.2013.06.012>.
  36. Löffler B, Hussain M, Grundmeier M, Bruck M, Holzinger D, Varga G, Roth J, Kahl BC, Proctor RA, Peters G. 2010. *Staphylococcus aureus* Panton-Valentine leukocidin is a very potent cytotoxic factor for human neutrophils. *PLoS Pathog* 6:e1000715. <http://dx.doi.org/10.1371/journal.ppat.1000715>.
  37. van Wamel WJ, Rooijackers SH, Ruyken M, van Kessel KP, van Strijp JA. 2006. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of *Staphylococcus aureus* are located on beta-hemolysin-converting bacteriophages. *J Bacteriol* 188: 1310–1315. <http://dx.doi.org/10.1128/JB.188.4.1310-1315.2006>.
  38. Becker S, Frankel MB, Schneewind O, Missiakas D. 2014. Release of protein A from the cell wall of *Staphylococcus aureus*. *Proc Natl Acad Sci U S A* 111:1574–1579. <http://dx.doi.org/10.1073/pnas.1317181111>.
  39. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, Earle S, Pankhurst LJ, Anson L, de Cesare M, Piazza P, Votintseva AA, Golubchik T, Wilson DJ, Wyllie DH, Diel R, Niemann S, Feuerriegel S, Kohl TA, Ismail N, Omar SV, Smith EG, Buck D, McVean G, Walker AS, Peto TE, Crook DW, Iqbal Z. 2015. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat Commun* 6:10063. <http://dx.doi.org/10.1038/ncomms10063>.
  40. Moon BY, Park JY, Hwang SY, Robinson DA, Thomas JC, Fitzgerald JR, Park YH, Seo KS. 2015. Phage-mediated horizontal transfer of a *Staphylococcus aureus* virulence-associated genomic island. *Sci Rep* 5:9784. <http://dx.doi.org/10.1038/srep09784>.
  41. Lindsay JA, Holden MT. 2004. *Staphylococcus aureus*: superbug, super genome? *Trends Microbiol* 12:378–385. <http://dx.doi.org/10.1016/j.tim.2004.06.004>.
  42. Jarraud S, Peyrat MA, Lim A, Tristan A, Bes M, Mougel C, Etienne J, Vandenesch F, Bonneville M, Lina G. 2001. egc, a highly prevalent operon of enterotoxin gene, forms a putative nursery of superantigens in *Staphylococcus aureus*. *J Immunol* 166:669–677. <http://dx.doi.org/10.4039/jimmunol.166.1.669>.
  43. Uhlemann AC, Porcella SF, Trivedi S, Sullivan SB, Hafer C, Kennedy AD, Barbian KD, McCarthy AJ, Street C, Hirschberg DL, Lipkin WI, Lindsay JA, DeLeo FR, Lowy FD. 2012. Identification of a highly transmissible animal-independent *Staphylococcus aureus* ST398 clone with distinct genomic and cell adhesion properties. *mBio* 3(2):e00027-12. <http://dx.doi.org/10.1128/mBio.00027-12>.
  44. Lowder BV, Guinane CM, Ben Zakour NL, Weinert LA, Conway-Morris A, Cartwright RA, Simpson AJ, Rambaut A, Nubel U, Fitzgerald JR. 2009. Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. *Proc Natl Acad Sci U S A* 106:19545–19550. <http://dx.doi.org/10.1073/pnas.0909285106>.