Spring 2021

Genome evolution and population dynamics of Staphylococcus bacteria

Joshua Smith
University of New Hampshire, Durham

Genome evolution and population dynamics of *Staphylococcus* bacteria

By

Joshua Thomas Smith

B.S. Biology, Franklin Pierce University, 2017

DISSERTATION

Submitted to the University of New Hampshire

In Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

In

Molecular and Evolutionary Systems Biology

September 2021

This dissertation was examined and approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Molecular and Evolutionary Systems Biology by:

Dissertation Director, Cheryl P. Andam, Ph.D.,
Primary advisor
Affiliate Faculty, Molecular, Cellular, and Biomedical Sciences
University of New Hampshire
(Assistant Professor, Biological Sciences University at Albany, State University of New York)

W. Kelley Thomas, Ph.D.,
Secondary advisor
Professor, Molecular, Cellular, and Biomedical Sciences
University of New Hampshire

David Plachetzki, Ph.D.,
Associate Professor, Molecular, Cellular, and Biomedical Sciences
University of New Hampshire

Anissa Poleatewich, Ph.D.,
Assistant Professor, Agriculture, Nutrition, and Food Systems
University of New Hampshire

Louis Tisa, Ph.D.,
Professor, Molecular, Cellular, and Biomedical Sciences
University of New Hampshire

On June 8, 2021

Approval signatures are on file with the University of New Hampshire Graduate School.

DEDICATION

For my wonderful wife Tahmineh, who has been right by my side throughout this work, keeping me happy and focused on our future.

عاشقتم عزیزم

ACKNOWLEDGEMENTS

First and foremost, I want to thank my advisor Dr. Cheryl Andam, who could not have been more supportive, encouraging, or positive if she tried. I am truly humbled by your unwavering commitment to helping me reach success. Our conversations on a wide range of topics have not only kept me on track professionally but have given me a social outlet during an incredibly isolating time. You are an amazing mentor and friend. I want to thank my committee members Dr. W. Kelley Thomas, Dr. David Plachetzki, Dr. Anissa Poleatewich, and Dr. Louis Tisa for providing their time and input to improve my thesis and challenge my thinking.

I want to thank my best friend and wife, Tahmineh Rahmani Eliato, who made the consequential decision of sitting next to me at orientation. Without you being by my side through all the ups and downs, I don't know how I would have made it here. Your willingness to go for what makes you happy despite enormous hurdles is an inspiration. This has been such a happy time, and I look forward to the next stage for us. I must also thank my mom, dad, and brother, Matthew, for shaping me into who I am today. Mom, you were the most patient person in the world to allow me to take my time to figure out what to do with my life. A mere eight years ago, who knew this is where it would lead? Matthew, watching you grow up has been one of my greatest joys, particularly during the most challenging times. I cannot wait to see where you go next. Dad, I owe my positive outlook on life to your ability to remain good natured and positive in the face of the sometimes-difficult world we live in. I'm sorry I haven't been able to see you all very frequently, but I hope you're proud of my progress. I would also like to thank my new brother-in-law, Kiarash, who is super smart and helped a lot with the math.

I will be eternally grateful to my biology professors at Franklin Pierce who truly ignited my love for learning about the wonderous life on this planet. Dr. Amy Piispanen, Dr. Susan Arruda, Dr. Jacques Veilleux, and Dr. Fred Rogers (among several others) helped me to see the world around me through a lens that just makes so much sense. All of my professors at FPU and UNH inspire me to keep learning. I hope to model their teaching approach myself for younger generations of college students.

I owe a great deal of gratitude to the friends and collaborators I have gotten to know at UNH, who have helped me a multitude of ways. First, thank you to the other members of the Andam lab, including Dr. Cooper Park, Isaiah Lee, and Kayti Belknap for being amazing friends and for providing perfect examples of how to be a great graduate student. Thanks to the undergraduates who were kind enough to let me teach and mentor them. I hope at least a fraction of what I said was helpful. Specifically, I would like to thank Sharlene Amador, Colin McGonagle, Jennifer Mydosh, and John Ball for their hard work and contributions to my projects. You certainly made my job much easier. Game nights with Eric English, Clay Ludwig, Will Hardesty-Dyck, and others were always great breaks in the research. Some of the collaborators I would like to thank include Dr. David Needle, Rob Gibson, and staff at the New Hampshire Veterinary Diagnostic Laboratory, who provided me and others in the lab with so much awesome data, valuable points of view, and guidance. I would like to thank Dr. Isabella Martin and her lab from Dartmouth Hitchcock Medical Center for letting me raid her freezer for the bacteremia samples. It was a really fun experience and I wish I had the time to collect every last isolate.

There are several people from COLSA/MCBS at UNH that I would like to thank as well. Dr. Toni Westbrook was crucial to helping me learn how to code when I was floundering early on. Everyone at the Hubbard Center for Genome Studies and the Research Computing Center, including Joe Sevigny, Steve Simpson, and Toni did a great job with genome sequencing and managing the Premise cluster. Regarding the administrative tasks, I would like to thank Paul Boisselle, Pam Wildes, and Matthew James, who always had the answer I needed and were incredibly helpful.

Of course, there are several funding sources that I would like to acknowledge for facilitating all of the great research I was able to do. These include UNH COLSA and UAlbany College of Arts and Sciences start-up funds, the UNH Graduate School Dissertation Year and Summer Teaching Assistant Fellowships, and the New Hampshire Agricultural Experiment Station Hatch Award. I was able to go to lots of interesting and productive meetings due to generous contributions from the Robert and Ruth Zsigray Academic Enrichment Travel Fund, the UNH Graduate School Travel Grant, and the MCBS Departmental Travel Fund.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

ABSTRACT

Genome evolution and population dynamics of *Staphylococcus* bacteria

by

Joshua T. Smith

University of New Hampshire

The gram-positive bacterial genus *Staphylococcus* (phylum Firmicutes) is a heterogeneous group that consists of commensals and opportunistic pathogens of humans and animals. To date, there are 57 validly published species according to the List of Prokaryotic names with Standing in Nomenclature (LPSN). The most well-studied and widely sampled species is *Staphylococcus aureus*, which causes a wide range of diseases in humans from relatively minor skin infections to much more invasive and serious sepsis, pneumonia, or endocarditis. For many years, *Staphylococcus* other than *S. aureus* have been regarded as non-pathogenic, but these bacteria are now being recognized as potential pathogens in humans. Other species, mainly *Staphylococcus epidermidis* and *Staphylococcus haemolyticus*, are members of the skin flora of human; however, they can invade deeper tissues through infection of inserted medical devices. Many other species are frequent colonizers of various species of mammals and birds, yet little is known about the diversity and evolutionary history of these non-*aureus* species. In this dissertation, I use concepts, approaches and methods in bacterial population genomics to investigate the evolutionary mechanisms that shape the diversity and population structure of *Staphylococcus* species across different ecological niches: *Staphylococcus pseudintermedius* in companion animals (Chapter 1), *S. aureus* in bloodstream infections in humans (Chapter 2), and coagulase-negative Staphylococcus (Chapter 3). Overall, I found that many genomic elements,

such as antimicrobial resistance and virulence genes, substantially vary within and between species. Frequent horizontal gene transfer (HGT) shapes this diversity and the adaptation of certain species and strains to certain niches. HGT in Staphylococcus is facilitated by a variety of mobile genetic elements, such as insertion sequences (Chapter 4). Many non-*aureus* species likely act as major reservoir of genetic variants that can be acquired by *S. aureus*. Findings from this dissertation will provide important insights to managing and controlling staphylococcal diseases in humans and animals.

INTRODUCTION

The *Staphylococcus* genus is heterogeneous group of gram-positive coccoid bacteria, a fact often overshadowed by the well-known commensal and pathogen *Staphylococcus aureus* (Becker, Heilmann, & Peters, 2014). The majority of these species naturally occur as skin and mucus membrane colonizers of humans and animals, often with the capacity to become opportunistic pathogens (Parlet, Brown, & Horswill, 2019). Over the past several decades, numerous novel species have been discovered and species boundaries have been clarified (Parte et al. 2020). Many species are not constrained to specific ecological niches, as certain species have the ability to adapt and thrive in several different animal species or host tissues (Becker, Heilmann, & Peters, 2014). Horizontal gene transfer (HGT) plays a major role in staphylococcal ecology and evolution (Méric et al. 2015, Langhanki et al. 2018). HGT allows *Staphylococcus* to rapidly acquire antimicrobial resistance and virulence determinants that have made infections caused by these pathogens incredibly difficult to treat. The overall goal of this dissertation is to elucidate the diversity of *Staphylococcus* bacteria and the genetic basis for adaptation and pathogenesis in different ecological niches.

In this dissertation, I present three genetic studies and a literature review that describe the high level of diversity and pathogenic and epidemiologic traits of different *Staphylococcus* populations, as well as bring attention to many non-*aureus* species. As these species also exist in a variety of niches, the bacteria studied herein have been obtained from many different sources including humans, animals, and the environment to give important ecological context. In Chapter 1, I describe *Staphylococcus pseudintermedius*, a colonizer and opportunistic pathogen of dogs

(Bannoehr & Guardabassi, 2012). Results revealed the co-circulation of phylogenetically diverse lineages in New Hampshire that have access to a large pool of accessory genes. Many antimicrobial resistant clones have emerged through multiple independent, horizontal acquisition of resistance determinants and frequent genetic exchange that disseminate DNA to the broader population and over large geographical distances. In Chapter 2, I genomic analysis and epidemiological concepts to study the population structure and dynamics of *S. aureus* from patients diagnosed bloodstream infections in New Hampshire, USA. I found that the *S. aureus* population was shaped mainly by the clonal expansion, recombination and co-dominance of two major methicillin-resistant clones in the last five decades In Chapter 3, I analyzed a dataset of 1,876 publicly available named coagulase-negative *Staphylococcus* (CoNS) genomes. I found variable rates of recombination and biases in recombination partners imply that certain CoNS species function as hubs of gene flow and major reservoir of genetic diversity for the entire genus. Finally, in chapter 4, I present a literature review of the known insertion sequences of *Staphylococcus* species, which are integral elements in shaping the genomic structure and facilitating DNA mobility (Siguier, Gourbeyre, and Chandler, 2014).

References

Bannoehr, J. and Guardabassi, L. (2012). Staphylococcus *pseudintermedius* in the dog: taxonomy, diagnostics, ecology, epidemiology and pathogenicity. *Veterinary Dermatology, 23*(4), 253–266, e51-52. https://doi.org/10.1111/j.1365-3164.2012.01046.x

Becker, K., Heilmann, C., and Peters, G. (2014). Coagulase-negative staphylococci. *Clinical Microbiology Reviews, 27*(4), 807-926; doi: 10.1128/CMR.00109-13

Langhanki, L., Berger, P., Treffon, J., Catania, F., Kahl, B.C., and Mellmann, A. (2018). In vivo competition and horizontal gene transfer among distinct *Staphylococcus aureus* lineages as major drivers for adaptational changes during long-term persistence in humans. *BMC Microbiology, 18*(152), 1-12; doi: 10.1186/s12866-018-1308-3.

Méric, G., Miragaia, M., de Been, M., Yahara, K., Pascoe, B., Mageiros, L., Mikhail, J., Harris, L. G., Wilkinson, T. S., Rolo, J., Lamble, S., Bray, J. E., Jolley, K. A., Hanage, W. P., Bowden, R., Maiden, M. C., Mack, D., de Lencastre, H., Feil, E. J., Corander, J., … Sheppard, S. K. (2015). Ecological Overlap and Horizontal Gene Transfer in *Staphylococcus aureus* and *Staphylococcus epidermidis*. *Genome biology and evolution, 7*(5), 1313–1328. https://doi.org/10.1093/gbe/evv066

Parlet, C.P., Brown, M.M., and Horswill, A.R. (2019). Commensal staphylococci influence *Staphylococcus aureus* skin colonization and disease. *Trends in Microbiology, 27*(6), 497-507; DOI: https://doi.org/10.1016/j.tim.2019.01.008

Parte, A.C., Sardà Carbasse, J., Meier-Kolthoff, J.P., Reimer, L.C. and Göker, M. (2020). List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology, 70*, 5607-5612; Database accessed May 19, 2021; DOI: 10.1099/ijsem.0.004332

Siguier, Patricia, Gourbeyre, E., and Chandler, M. (2014). Bacterial insertion sequences: Their genomic impact and diversity. *FEMS Microbiology Reviews, 38*(5), 865–891. https://doi.org/10.1111/1574-6976.12067

CHAPTER 1

Population genomics of *Staphylococcus pseudintermedius* in companion animals in the United States

Joshua T. Smith[a], Sharlene Amador[a], Colin J. McGonagle[a], David Needle[b], Robert Gibson[b], Cheryl P. Andam[a].

Article published in *Communications Biology*

Presented here with permission from publisher (See Appendix 1)

[a] University of New Hampshire, Department of Molecular, Cellular and Biomedical Sciences, Durham, *New Hampshire*, USA

[b] New Hampshire Veterinary Diagnostic Laboratory, Durham, NH 03824, USA

Abstract

Staphylococcus pseudintermedius is a commensal bacterium and a major opportunistic pathogen of dogs. The emergence of methicillin resistant *S. pseudintermedius* (MRSP) is also becoming a serious concern. We carried out a population genomics study of 130 clinical *S. pseudintermedius* isolates from dogs and cats in the New England region of the United States. Results revealed the co-circulation of phylogenetically diverse lineages that have access to a large pool of accessory genes. Many MRSP and multidrug resistant clones have emerged through multiple independent, horizontal acquisition of resistance determinants and frequent genetic exchange that disseminate DNA to the broader population. When compared to a Texas population, we found evidence of clonal expansion of MRSP lineages that have disseminated over large distances. These findings provide unprecedented insight into the diversification of a common cutaneous colonizer of

man's oldest companion animal and the widespread circulation of multiple high-risk resistant clones.

Introduction

Staphylococcus pseudintermedius is a commensal bacterium of the skin and mucous membrane and is frequently found in companion animals. Carriage of *S. pseudintermedius* often reaches more than 80% in some populations of healthy dogs (Bannoehr & Guardabassi, 2012). It is also an opportunistic pathogen responsible for severe and necrotizing infections, and is frequently isolated in the skin, ears, bones and post-surgical abscesses (Bannoehr & Guardabassi, 2012). *S. pseudintermedius* has been found in other animals such as cats and horses, although they are not considered its natural hosts (Ruscher et al., 2009). The bacterium shares several features with *Staphylococcus aureus*, the most important staphylococcal species in humans, including the capacity to express a range of virulence factors such as coagulase and other proteolytic enzymes, as well as a variety of toxins such as haemolysins, exfoliative toxins, enterotoxins and leucotoxins (Bannoehr & Guardabassi, 2012; Maali et al., 2018). The emergence of methicillin resistant *S. pseudintermedius* (MRSP) is becoming a serious concern in veterinary medicine (McCarthy et al., 2015; Nisa et al., 2019; Wegener et al., 2018) and highlights the need for accurate long-term surveillance. Furthermore, there has been recent reports of *S. pseudintermedius* being isolated in human carriage and infections, mainly associated with contacts with dogs (e.g., pet owners, veterinary staff) (Paul et al., 2011; Van Hoovels et al., 2006). While traditionally *S. pseudintermedius* has not been considered a risk for humans, recent reports have suggested that it is emerging as a zoonotic diagnosis and it may have been

previously misidentified as *S. aureus* in human infections (Börjesson et al., 2015; Somayaji et al., 2016).

Microbial population genomics involves sequencing the genomes of hundreds or even thousands of closely related strains within and between environments (Sheppard et al., 2018). It has been instrumental in revolutionizing the epidemiology, surveillance and control strategies of infectious diseases that threaten human health, and has been widely used to investigate common bacterial species such as *S. aureus*, *Klebsiella pneumoniae* and *Streptococcus pneumoniae* (Tagini & Greub, 2017). Remarkable levels of genomic variation have been widely reported in bacterial pathogen populations, which have critical implications in understanding the origins of highly virulent and resistant lineages, successful colonization, transmission and instances of host switching (Andam et al., 2017; Chewapreecha et al., 2014; Croucher et al., 2014). However, the application of large-scale whole genome sequencing of bacterial species that naturally inhabit animal hosts remains limited. Even more problematic is when strains with multiple resistant phenotypes become prevalent in the population, but the lack of precise diagnostic and surveillance techniques makes them invisible to clinical and epidemiological studies. This gap in our knowledge ultimately limits our understanding of the underlying genetic characteristics of bacterial lineages that cause infections in animals and impacts the likelihood of precisely identifying new and emerging clones, including potential zoonoses.

The few genomic studies of *S. pseudintermedius* have offered a glimpse of its diversity, epidemiological characteristics and antimicrobial resistance (AMR) profile. A study of 12 genomes revealed that the multidrug resistant and MRSP phenotypes in sequence types (ST) 71

and 68 in Europe evolved through a stepwise accumulation of *mec*-carrying chromosomal cassette (SCC*mec*), transposon Tn*5405* and single nucleotide polymorphisms (SNPs) conferring fluoroquinolone resistance (McCarthy et al., 2015). In the Netherlands, *S. pseudintermedius* is dominated by STs 45, 71 and 258, and genome sequencing of 50 isolates showed different resistance profiles among members of each clone (Wegener et al., 2018). These results suggest that the AMR phenotype is acquired and maintained through both vertical inheritance and horizontal gene transfer (HGT) (Wegener et al., 2018). In New Zealand, ST 71-dominated MRSP population has undergone infrequent ancestral acquisition events of the mobile chromosomal cassette SCC*mec*, which was followed by its subsequent widespread dispersal (Nisa et al., 2019). However, genomic studies aimed at elucidating the structure and dynamics of *S. pseudintermedius* populations remain scarce. Such information will be critical to addressing the question of whether the multidrug resistant and MRSP phenotypes of *S. pseudintermedius* are restricted to one or few clones, the factors that drive their evolution, and how the species will respond to different selective pressures. It will also help us clarify the breadth of potentially useful adaptive variants that the species possesses and the underlying mechanisms that allow it to successfully switch between hosts. In this study, we carried out a population genomics study of 130 clinical *S. pseudintermedius* isolates from dogs and cats across five states in the New England region of the United States sampled from 2017-2018. Results reveal the diversification and widespread circulation of multiple high-risk resistant clones.

Methods

Bacterial sample collection

The *S. pseudintermedius* isolates were acquired as culture swabs from routine clinical specimen submissions to NHVDL, Durham, New Hampshire, USA, from October 2017 to October 2018. The clinical specimens were received from veterinary practices in the New England region that includes the states of Connecticut, New Hampshire, Maine, Massachusetts, and Vermont. No live vertebrates were used in this study; therefore, the NHVDL is exempt from the IACUC approval process. All isolates were from disease cases. Pure isolates of *S. pseudintermedius* were cultured on commercially prepared tryptic soy agar with 10% sheep red blood cells. Initial species identification was carried out using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), which has been previously demonstrated to be a sensitive and specific species-level identification tool for *S. pseudintermedius* (Little et al., 2019; Nisa et al., 2019). We used the Bruker Biotyper MALDI-TOF MS following manufacturer's protocols. The addition of formic acid in the extended direct technique was used when necessary to ensure a reliable log(score), which refers to the level of similarity between an unknown tested specimen and a reference sample. Species assignments were made when log(score) values were ≥ 2.0 and identical species were included in the top two database matches when compared to the library of reference spectra available in the Bruker MBT Compass, RUO library 6903(V6) and 7311(V7) (Bruker Daltonics, Bremen, Germany). The most common sites of isolation included skin, ears, urine, and wounds. Metadata information of isolates are displayed in Table S1.1. All isolates were stored in DMSO solution in -80^{o}C.

Methicillin susceptibility screening

Isolates were initially screened *in vitro* at the NHVDL with the Kirby Bauer disc diffusion technique using both the cefoxitin and oxacillin discs. Whereas cefoxitin predicts methicillin resistance in *S. aureus*, oxacillin is used as the official predictor of methicillin resistance for S. *pseudintermedius* following breakpoint guidelines of the most current Clinical and Laboratory Standards Institute (*CLSI Performance Standards for Antimicrobial Disk and Dilution Susceptibility Tests for Bacteria Isolated from Animals. VET01*, 2018). For methicillin resistant isolates, we further tested for the presence of the penicillin-binding protein PBP2 using a commercial latex agglutination test (MASTALEX MRSA Latex Kit, MAST, UK) following manufacturer's protocols. Verification for the presence of the *mecA* gene was done using whole genome sequencing (described below).

DNA extraction and whole genome sequencing

Cultures were grown in brain heart infusion broth at 37°C for 24-48 hrs. DNA was extracted and purified from cultures using the Zymo Research Quick-DNA Fungal/Bacterial Miniprep Kit (Irvine, California) following manufacturer's protocols. DNA concentration was measured using a Qubit fluorometer (Invitrogen, Grand Island, NY). DNA libraries were prepared using the Nextera XT protocol (as per the manufacturer's instructions) with 1 ng of genomic DNA/isolate. Samples were sequenced as multiplexed libraries on the Illumina HiSeq platform operated per the manufacturer's instructions to produce paired end reads of 250 nucleotides in length. Sequencing was done at the UNH Hubbard Center for Genome Studies, Durham, NH, USA.

Genome assembly, annotation, and pan-genome analysis

Reads were assembled into contigs using the *de novo* assembler SPAdes v.3.13.1 (Bankevich et al., 2012). The resulting contigs were annotated using Prokka, a stand-alone tool specifically developed for annotating bacterial genomes (Seemann, 2014). Genome assembly quality was assessed using Quast (Gurevich et al., 2013). Sequencing failures and poor overall sequence qualities are known to occur with any sequencing platform. We used CheckM to assess the quality of our sequences and excluded genomes with <90% completeness and >5% contamination (Parks et al., 2015). After filtering out the genomes with low coverage and of poor quality and exclusion of any assemblies with >200 contigs and an N50 <40,000 bp, we used a total of 130 genomes for downstream analysis, with the numbers of contigs ranging from 25 to 196 and N50 ranging from 50,676 to 351,605 bp (Figure S1.1 and Table S1.1). We also confirmed the species identity using a BLASTN (Altschul et al., 1990) search of the NCBI non-redundant (nr) database using the annotated 16S rDNA sequences of each genome in our study. To determine the degree of genomic relatedness and hence clarify whether these genomes belong to the same species, we calculated the genome-wide ANI for all possible pairs of genomes using the program FastANI v.1.0 (Jain et al., 2018). ANI is a robust similarity metric that has been widely used to resolve strain relatedness and determine whether any two strains belong to the same or different species (Jain et al., 2018). We used the program Roary (Page et al., 2015) to characterize the pan-genome of the 130 New England *S. pseudintermedius*. Roary iteratively pre-clusters protein sequences using CD-HIT (Fu et al., 2012), all-against-all BLASTP (Altschul et al., 1990) and Markov clustering (Enright et al., 2002). Roary then sorts all orthologous gene families identified in the pan-genome into four categories: core (genes found in ≥ 99% strains), soft core (found in 95% ≤ strains < 99%), shell (found in 15% ≤ strains < 95%) and cloud (found

in less than 15% of strains) (Page et al., 2015). We used the default parameters for all programs used.

Comparative population genomics of New England and Texas

We compared 126 New England genomes with 107 previously published genomes from Texas for a total of 233 genomes (Little et al., 2019). For both datasets, we included only those clinical isolates from dogs (disease cases). The four genomes from cats in the New England dataset were excluded. A list of accession numbers and metadata for the Texas *S. pseudintermedius* genomes were included in Table S1.1. To maintain consistency in gene annotations, we used Roary to characterize the pan-genome of this merged dataset and re-annotated them using Prokka with default parameters (Seemann, 2014).

Phylogeny and population structure analyses

For each of the two datasets (i.e., the 130 New England genomes and the 233 combined New England + Texas genomes), each single-copy orthologous gene family obtained from the Roary analysis was aligned using MAFFT (Katoh et al., 2002). The alignments were concatenated to give a single core alignment, and a maximum-likelihood phylogeny was then generated using the program Randomized Axelerated Maximum Likelihood (RAxML) v.8.2.11 (Stamatakis, 2006) with a general time-reversible nucleotide substitution model (Tavaré, 1986) and four gamma categories for rate heterogeneity. Genetic population structure analysis was performed using R-implemented hierarchical Bayesian analysis of population structure (RhierBAPS) with default parameters using nested clustering and with the core genome alignment as input (Tonkin-Hill et al., 2018). All phylogenies were visualized using the

Interactive Tree of Life (Letunic & Bork, 2016). To further elucidate the population structure of *S. pseudintermedius* based on the divergence of both shared sequence and gene content in a population, we used PopPUNK (Population Partitioning Using Nucleotide *K*-mers) with default parameters (Lees et al., 2019). PopPUNK compares all possible pairs of genomes by calculating the proportion of shared k-mers of different lengths to determine core and accessory genome distances. It then generates a scatterplot of the two distances to reveal the predicted clustering of isolates (Lees et al., 2019).

In silico molecular typing and detection of AMR genes

ST identification of isolates was confirmed using the program MLST (https://github.com/tseemann/mlst), which extracts the sequences of seven housekeeping genes (*tuf, cpn60, pta, purA, fdh, ack, sar* (Solyman et al., 2013)) from the contig files and compares them to the *S. pseudintermedius* MLST database (https://pubmlst.org/spseudintermedius/). We screened all of the genomes for known accessory element resistance genes using two programs. For the New England dataset, we used a direct read mapping approach implemented in ARIBA (Inouye et al., 2014) that can identify resistance determinants due to allelic variants and horizontally acquired AMR genes. For the combined New England and Texas genomes, we used the contig-based search method ABRicate v.0.8.13 (https://github.com/tseemann/abricate). The resistance allele sequences used for comparison were retrieved from the CARD database (Zankari et al., 2012). A query gene is accepted as a true AMR gene if it reaches a threshold of 95% sequence identity and 95% coverage when compared to the database. We used the default parameters for each program.

Detection of virulence genes and SCCmec

We screened all New England genomes for known virulence genes using ARIBA to query the Virulence Factor Database (VFDB) (Liu et al., 2019). Genomes carrying the *mecA*-carrying chromosomal cassette SCC*mec* were identified using SCCmecFinder v.1.2 (Kaya et al., 2018) with minimum thresholds of >60% for sequence coverage and >90% sequence identity. We used the default parameters for each program.

Estimating recombination rates

We used three approaches to detect recombination in *S. pseudintermedius*. First, we calculated the PHI (pairwise homoplasy index) to determine the statistical likelihood of recombination being present in our dataset (Bruen et al., 2006). If recombination is absent, the genealogical correlation or similarity of neighboring sites remains unchanged to permutations of the sites as all sites have identical evolutionary history (Bruen et al., 2006). We then visualized potential recombination events using Splitstree v.4.14.6 which shows the presence of conflicting phylogenetic signals for different genes (Huson, 1998). Second, we ran fastGEAR (Mostowy et al., 2017) with default parameters to detect evidence for recombination in core genes and shared accessory genes. FastGEAR first infers the population structure of individual sequence alignments using a Hidden Markov Model to identify lineages, which are groups of strains that are genetically distinct in at least 50% of the alignment (Mostowy et al., 2017). Within each lineage, recombinations are identified by determining sequence similarity between the target sequence and all other remaining lineages (Mostowy et al., 2017). Third, we used mcorr with defaulty parameters to calculate the correlation profile and different evolutionary parameters using the core gene alignment as input and with 1000 bootstrapped replicates (Lin & Kussell,

2019). Mcorr estimates six evolutionary parameters: θ – mutational divergence; ϕ – recombinational divergence; c – recombination coverage or proportion of sites in the genome whose diversity was derived from outside the sample through recombination; d – diversity; mean fragment size ($\bar{f}$) of a recombination event; and θ/ϕ (or γ/μ) - relative rate of recombination to mutation (Lin & Kussell, 2019).

Statistics and reproducibility

We used Welch's t test and z score test to compare the different genomic components between the New England and Texas populations. Significance of the inferred PHI was obtained using a permutation test (Bruen et al., 2006). We used a diversity test implemented in fastGEAR to determine the significance of the inferred recombinations and uncover false positives (Mostowy et al., 2017). We used Welch's t test to compare the two *S. pseudintermedius* populations for each of the six parameters calculated by mcorr (Lin & Kussell, 2019). A p value of < 0.05 was considered statistically significant. We used the default parameters for all programs used unless otherwise stated and are then described. Exact sample sizes for each group were described. Source data used to plot **Figures 1.2d-g, 1.3a-b, 1.4b-h** are archived in Table S1.9.

Data availability

All supporting data are included in this published article and its supplementary material and are available from the corresponding author upon request. Genome sequence data of the New England samples have been deposited in the NCBI Sequence Read Archive under BioProject accession number PRJNA563147 with BioSample accession numbers listed in Table

S1.1. Allelic profiles of the 79 STs newly identified to our knowledge from the New England genomes were submitted to the *S. pseudintermedius* database in the MLST website (https://pubmlst.org/spseudintermedius/). Accession numbers of the Texas genomes (Little et al., 2019) are listed in Table S1.1 and have been downloaded from the PATRIC database (https://patricbrc.org).

Code availability

All tools and R packages used for the analysis are publicly available and fully described in the method sections.

Results

Diverse lineages are co-circulating in New England

We obtained a total of 162 *S. pseudintermedius* isolates from routine diagnostic tests of clinical specimens submitted to the New Hampshire Veterinary Diagnostic Laboratory (NHVDL) from October 2017 to October 2018. Of the 162 *S. pseudintermedius* isolates, we retrieved high-quality draft genome sequences for 130 isolates, composed of 126 genomes from dogs and four from cats (**Figure 1.1** and Table S1.1). More than half of the isolates came from New Hampshire (n = 84), while the remainder were from neighboring states in New England (Massachusetts, Connecticut, Maine, and Vermont). Initial *in vitro* screening of methicillin and other beta-lactam antibiotic resistance was performed using disc diffusion testing. Results revealed the presence of 36 MRSP isolates and 94 methicillin-susceptible *S. pseudintermedius* (MSSP) isolates throughout the entire study period. We found MRSP isolates in almost every

month of sampling, with as many as six out of 12 isolates being MRSP in a single month (**Figure 1.1a**).

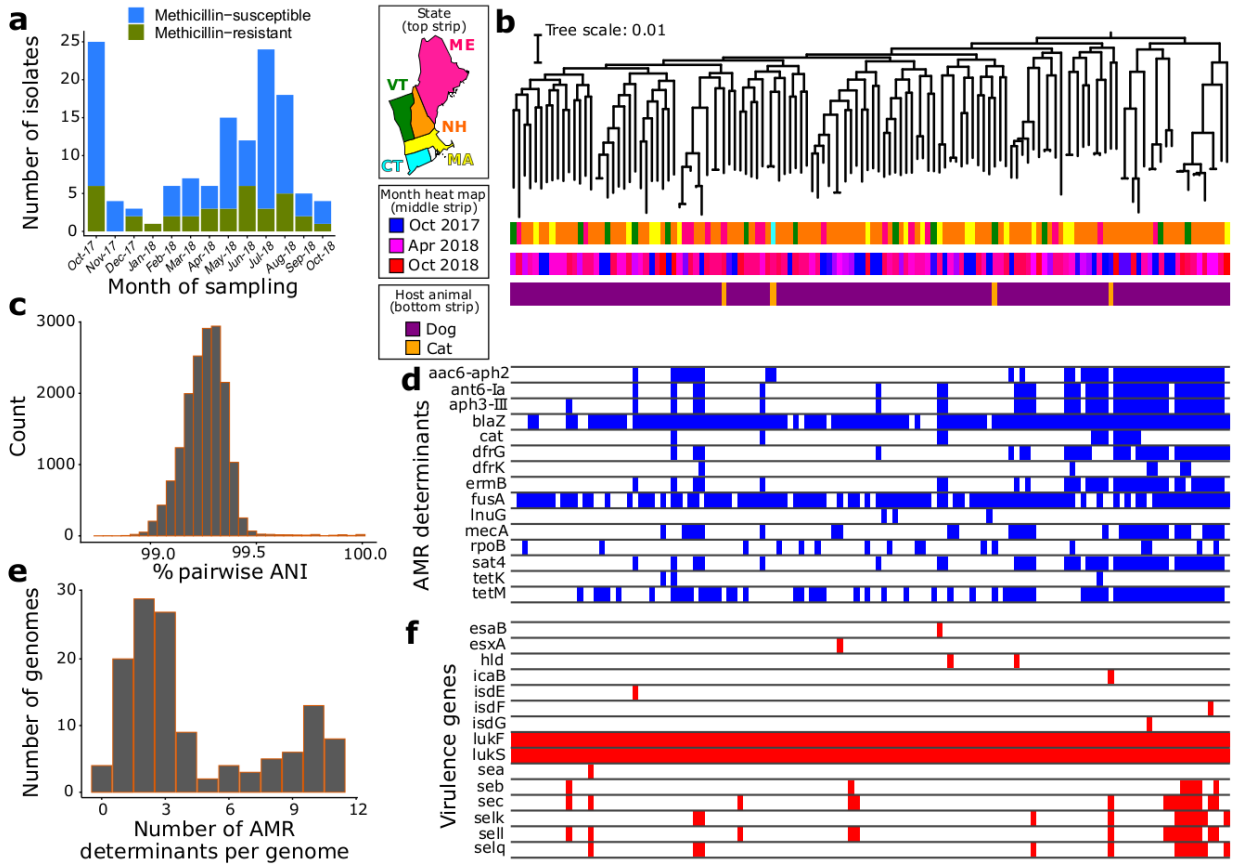

Figure 1.1 Characteristics of the New England *S. pseudintermedius* population (n = 130 genomes). **a** Monthly sampling of MRSP and MSSP isolates based on results of disk diffusion test. **b** The midpoint-rooted maximum likelihood phylogenetic tree was calculated using sequence variation in 1906 core genes. Scale bar represents nucleotide substitutions per site. For visual clarity, not all months of sampling are shown on the color legend of month heat map; however, all months are shown on the middle strip. **c** Frequency distribution of pairwise genome-wide ANI values. **d** Gene presence–absence matrix showing the distribution of AMR determinants present in each genome (blue blocks–present; white–absent). AMR determinants refer to both horizontally acquired genes and allelic variants of a gene. **e** Distribution of the number of AMR determinants per genome. **f** Gene presence–absence matrix showing the distribution of virulence-associated genes present in each genome (red blocks–present; white–absent). Each column in the two matrices correspond to a genome on the phylogeny. NH New Hampshire, MA Massachusetts, CT Connecticut, ME Maine, VT Vermont.

De novo genome assembly generated sequences of sizes ranging from 2.44 to 2.93 Mb (mean = 2.61 Mb) and the number of predicted genes ranged from 2,260 to 2,850 (mean = 2,449) (Table S1.1). We used Roary (Page et al., 2015) to estimate the pan-genome of the entire New England *S. pseudintermedius* population. Roary identified a total of 8,207 orthologous gene families in the pan-genome, which can be classified into core genes (n = 1,906 genes), soft core genes (n = 108 genes), shell genes (n = 741 genes) and cloud genes (n = 5,452) (Figure S1.2 and Table S1.2). The combined core and soft core genes comprised 24.5% of the pan-genome, while the combined shell and cloud genes (which together make up the accessory genome) comprised 75.5% of the pan-genome. It is notable that many accessory genes were unique to a single strain (2,797 genes, representing 34.1% of the pan-genome). The large accessory genome in our results is not uncommon and has been reported in different bacterial pathogens, including *S. aureus* (Bosi et al., 2016; Park et al., 2019) and *Staphylococcus haemolyticus* (Pain et al., 2019). However, we recognize that pan-genome sizes and precise characterization of core and accessory genes are greatly influenced by the phylogenetic distance, sequence quality of genomes, the number of genomes being compared, assembly and annotation methods, and choice of threshold values to define orthologous genes. The maximum likelihood phylogenetic tree based on the sequence alignment of the 1,906 core genes reveal many deep branching lineages that have relatively little structure relative to sampling location (state) or month of sampling (**Figure 1.1b**). The four isolates from cats are intermingled among the dog isolates throughout the tree. To further ensure that all isolates are from the same species and determine the degree of overall genomic relatedness among the *S. pseudintermedius* isolates, we calculated the pairwise average nucleotide identity (ANI) of all orthologous genes shared between any two genomes (Jain et

al., 2018). Genome-wide ANI values for every possible pair of *S. pseudintermedius* genomes range from 98.8-100% (mean = 99.3%) (**Figure 1.1c**, Figure S1.3 and Table S1.3).

We considered bioinformatic evidence for the presence of different AMR determinants in our samples. We used ARIBA to detect the presence of horizontally acquired AMR genes and resistance alleles due to chromosomal mutations (Hunt et al., 2017). The distribution of AMR determinants varied substantially among the genomes and we did not find evidence for phylogenetic or geographical clustering of any one of the AMR genes (**Figure 1.1d**, Table S1.4). Two mechanisms confer penicillin resistance in *Staphylococcus*: the production of beta-lactamase encoded by *blaZ*, which inactivates penicillin by hydrolysis of its beta-lactam ring, and the penicillin-binding protein PBP2a encoded by *mecA* (Llarrull et al., 2009). We detected *blaZ* in 111 isolates representing 85.4% of the population, and the *mecA* gene in 32 genomes representing 24.6% of the population. The genes *tetK* and *tetM*, which confer resistance to broad-spectrum tetracyclines, were detected in 3 and 59 genomes representing 2.3% and 45.4% of the population, respectively. We also detected resistance determinants for other antimicrobial classes. Many genomes carry genes that encode resistance against aminoglycosides [aac(6)-aph(2), aph(3')-III, ant(6')-Ia, sat4a; n= 38, 38, and 39 genomes respectively], chloramphenicols [*cat*; n= 12 genomes], daptomycin [*rpoB* allelelic variant; 17 genomes], diaminopyrimidines [*dfrG*, *dfrK*; n= 31 and 6 genomes respectively], fusidic acid [*fusA* allelelic variant; 99 genomes], lincosamides/macrolides/streptogramin B [*lnuG*, *ermB*; n= 3 and 34 genomes respectively], and streptothricin [sat4; n=38 genomes]. Only four isolates out of the 130 genomes do not contain any known AMR determinant, whether it is a resistance allele or a horizontally acquired gene. Although none of the isolates from the four cats are MRSP, all four genomes harbor DNA that

encode resistance against other antibiotic classes. Overall, we found that a total of 126 *S. pseudintermedius* genomes carry at least one AMR determinant, and a remarkable 41 genomes contain ≥ 5 AMR determinants per genome (**Figure 1.1e**).

Mobile genetic elements (MGEs) play an important role in the evolution and diversity of *Staphylococcus* and facilitate the dissemination of AMR genes within and between species (Firth et al., 2018). Here, we explored the presence and types of the chromosomal cassette SCC*mec*, which can facilitate the mobilization and distribution of *mecA* and other AMR genes in *Staphylococcus* (International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC), 2009; Rolo et al., 2017). SCCmec elements are highly variable in terms of their structural organization and gene content, but are classified mainly based on the *ccr* and *mec* gene complexes, the key elements of the cassette responsible for integration and excision of SCC*mec* and the beta-lactam resistance phenotype, respectively (Kaya et al., 2018). A total of 13 SCC*mec* types and numerous subtypes have been identified in *S. aureus* to date (Kaya et al., 2018). Of the 32 *S. pseudintermedius* genomes that carry the SCC*mec*, we identified three known types [Type III (also known as Type II-III in MRSP (Worthing et al., 2018)), n= 2 genomes; Type IV, n= 10 genomes; Type V, n= 15 genomes] and five genomes with unknown type (Figure S1.4 and Table S1.1). These results were similar to those identified in veterinary MRSP in Australia (Worthing et al., 2018). Their locations in the phylogenetic tree show that the MRSP phenotype has been derived from at least 13 independent acquisitions of different SCC*mec* elements. In *S. aureus*, Type III is found predominantly in hospital-associated methicillin resistant *S. aureus* (MRSA) and carry resistance genes for cadmium, tetracycline, mercury, erythromycin and spectinomycin, in addition to *mecA*

(Lakhundi & Zhang, 2018). Hence, a single transfer event of Type III can turn a susceptible strain into a multidrug-resistant strain. Types IV and V are often found in community-associated MRSA and carry only the *mecA* gene (Lakhundi & Zhang, 2018). In *S. pseudintermedius*, SCC*mec* III tended to be healthcare-associated MRSP lineages whereas isolates with SCC*mec* V tended to be community-associated MRSP lineages (Kasai et al., 2016). The five untypeable SCC*mec* highlight the need to continue efforts to discover and characterize undiscovered variants of SCC*mec* elements, especially in non-*aureus* species.

We identified several virulence genes commonly found in *S. aureus* in the New England *S. pseudintermedius* population (**Figure 1.1f** and Table S1.5). The two-component pore-forming leukocidin genes *lukF* and *lukS* (Abouelkhair et al., 2018) were present in all genomes. The synergistic action of these genes produces a toxin that damages membranes of host defense cells and erythrocytes (Maali et al., 2018; Pinchuk et al., 2010). Genes associated with other virulence factors were also detected from the genomic sequences, although present in lower frequencies. These include enterotoxins encoded by *sec* (n = 15 genomes), *selK* (n = 11 genomes), *selI* (n = 15 genomes), and *selq* (n = 12 genomes). Staphylococcal enterotoxins encoded by *sec* and *sel* cause superantigenic and emetic activities in the bacterium (Pinchuk et al., 2010).

Overall, we found high levels of phylogenetic and genomic diversity in *S. pseudintermedius* isolates co-circulating in New England only within a single year of sampling. The population consists of several methicillin resistant, multidrug resistant and virulent clones present in dogs and cats.

Comparison of two S. pseudintermedius populations

To place our dataset in a country-wide context, we compared the genome sequences of New England *S. pseudintermedius* to a previously published genomic dataset from Texas (Little et al., 2019) (**Figure 1.2a**, Figure S1.3, and Table S1.1). New England, located in the northeast corner of the United States, is approximately 2,770 km from Texas, which is in the southern part of the country (**Figure 1.2b**). We chose this dataset because its sampling strategy was most similar to our study, i.e., only isolates from clinical specimens from dogs collected by the Texas Veterinary Medical Teaching Hospital and a mix of MRSP and MSSP (Little et al., 2019). Of the 160 Texas genomes sequenced in that study, we did not include genomes of isolates from healthy dogs and those with low quality sequences. To minimize the confounding effect of the host animal, we also excluded the four isolates from cats in the New England dataset. Overall, we compared 126 New England and 107 Texas genomes, all from disease cases in dogs.

We next re-ran Roary on each dataset. The composition of the pan-genome varied between the two populations. The New England population has a total of 8,122 genes (1,910 core, 107 soft core, 765 shell, 5,340 cloud genes), while the Texas population has a total of 6,845 genes (1,836 core, 133 soft core, 1,020 shell, 3,856 cloud genes) (Table S1.2). The combined New England and Texas dataset has a total of 9,817 genes (1,776 core, 166 soft core, 905 shell, 6,970 cloud genes) (Figure S1.5 and Table S1.2). We used the concatenated sequences of 1,776 core genes that are common to both New England and Texas populations, generating a 1.60 Mb sequence alignment, to build a maximum likelihood phylogeny (**Figure 1.2a**). We observed frequent intermingling of isolates from each population across the tree, suggesting widespread dissemination across a large geographical distance.

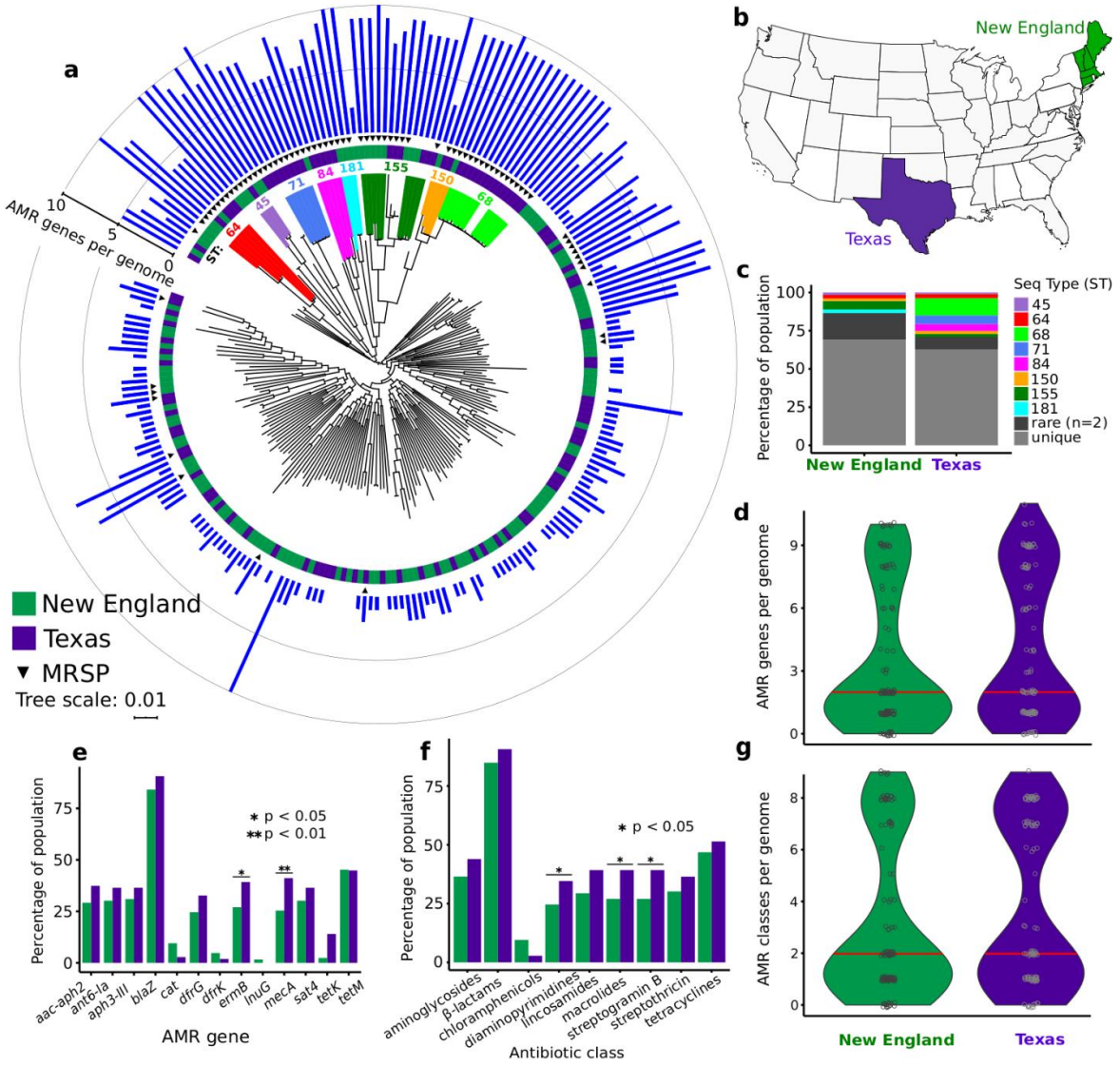

Figure 1.2 Phylogenetic relationship and AMR profiles of 126 New England genomes and 107 Texas genomes of *S. pseudintermedius*. All isolates in this tree were sampled from clinical specimens from dogs submitted to the veterinary laboratory in each region. **a** Midpoint-rooted maximum likelihood tree built using concatenated sequence alignments of 1776 core genes present in both New England and Texas populations. Scale bar represents nucleotide substitutions per site. Upside down triangles indicate the presence of *mecA* in genome sequences. Blue bar plots represent number of horizontally acquired AMR genes per genome. **b** Map of the United States showing the New England and Texas regions. **c** Distribution of STs per region. Comparison between the New England and Texas populations in terms of the total number of AMR genes per genome **d**, percentage of each population that carry specific AMR genes **e**, percentage of the population that carry at least one AMR gene for each major class of antibiotic **f**, and the total number of genomes that are resistant to zero, one or multiple classes of antibiotics **g**. Red horizontal lines in **d** and **g** represent median values. **e**, **f** Only statistically significant differences are shown for visual clarity. **g** Counts of genomes are based on the presence of at least one AMR gene that represent each major class of antibiotic.

We carried out a second AMR analysis to compare horizontally acquired AMR genes in the two populations using ABRicate (Table S1.6). In this analysis, we did not use ARIBA because ARIBA requires the raw reads as input, which were not available for the Texas genomes. Based on *in silico* detection of the *mecA* gene, ABRicate identified 44 MRSP and 63 MSSP in the Texas population (**Figure 1.2a**). In both New England and Texas populations, the MRSP lineages are mostly clonal, indicated by the extremely short branches at the tip of each clonal branch. The most common MRSP STs that had representatives from both regions were STs 45, 64, 150 and 155. There is also evidence for clonal expansion of some MRSP lineages within each region. For example, STs 68, 71 and 84 were found only in Texas, while ST 181 was found only in New England. Moreover, in both populations, the MSSP lineages are genetically diverse and are found in particularly long branches. There were also numerous STs that were either rare (here defined as present in only two genomes) or unique (present only in a single genome) (**Figure 1.2c**). STs that are intermingled with another may reflect recent divergence or recombination of the MLST genes.

Many genomes of the eight most common MRSP STs and their close relatives also carry the highest number of horizontally acquired AMR genes per genome, ranging from 2-11 AMR genes (median = 9 AMR genes per genome) (Table S1.6). We did not find significant differences between the two populations in terms of the total number of AMR genes per genome (p-value = 0.14, Welch's t test) (**Figure 1.2d**). We further queried the genomes to compare the distribution of specific AMR genes between New England and Texas (**Figure 1.2e**). Both had comparable proportions of genomes that carry many of these AMR genes, such as *aac-aph2*, *blaZ*, and *tetM*. However, we found significant differences between the two populations in

those genomes that carry the gene *ermB* (p = 0.0233, z score test) and *mecA* (p = 0.0054, z score test). For both genes, the Texas population exhibited higher number of genomes as a proportion of their respective population size. In terms of the major classes of antibiotics, both populations have high numbers of genomes that carry genes conferring resistance against beta-lactams, but they differ in the proportions of genomes that carry resistance genes for macrolides (p = 0.0233, z score test), streptogramin B (p = 0.0233, z score test) and diaminopyrimidines (p = 0.0475, z score test) (**Figure 1.2f**). We also found that a total of 45 genomes representing 35.7% of the New England population and 48 genomes representing 44.9% of the Texas population carry resistance genes to ≥ 3 antibiotic classes (**Figure 1.2g**). We did not detect any horizontally acquired AMR genes in 15/126 (11.9%) and 10/107 (9.35%) of the New England and Texas genomes, respectively.

Within-species diversity stems from allelic differences in core genes as well as gene content variation in the accessory genomes among strains (Croucher et al., 2014), which may partly stem from differential response to local selective pressures through frequent DNA gain and loss (Chewapreecha et al., 2014; Croucher et al., 2014). We therefore sought to determine to what extent the genome structures of *S. pseudintermedius* in New England and Texas have diverged. First, using the 1.60 Mb core genome alignment, we calculated the total number of SNPs between each pair of genomes from each population. We did not find significant differences in the number of SNPs in the core genome (p-value = 0.46, Welch's t test) (**Figure 1.3a**), but found significant differences in the number of accessory genes per genome between the two populations (p-value <0.00016, Welch's t test) (**Figure 1.3b**). To gain insight on how the accessory genome has diverged in relation to the core genome (Croucher et al., 2014), we

used PopPUNK which employs pairwise nucleotide k-mer comparisons to distinguish shared sequence and gene content (Lees et al., 2019). We found similar patterns in the distribution of pairwise genomic distances between New England and Texas, with larger genetic distances concentrated away from the origin (**Figure 1.3c**). This pattern is indicative of the presence of multiple genetically distinct clusters that are diverging in both core sequences and accessory gene content. These results expand the core-genome-based phylogenetic analysis described above (**Figure 1.2a**). Overall, these data show that the each of the two geographically distinct populations of *S. pseudintermedius* in the United States is composed of numerous genetically distinct lineages that can be distinguished in both their core and accessory genome divergence patterns.

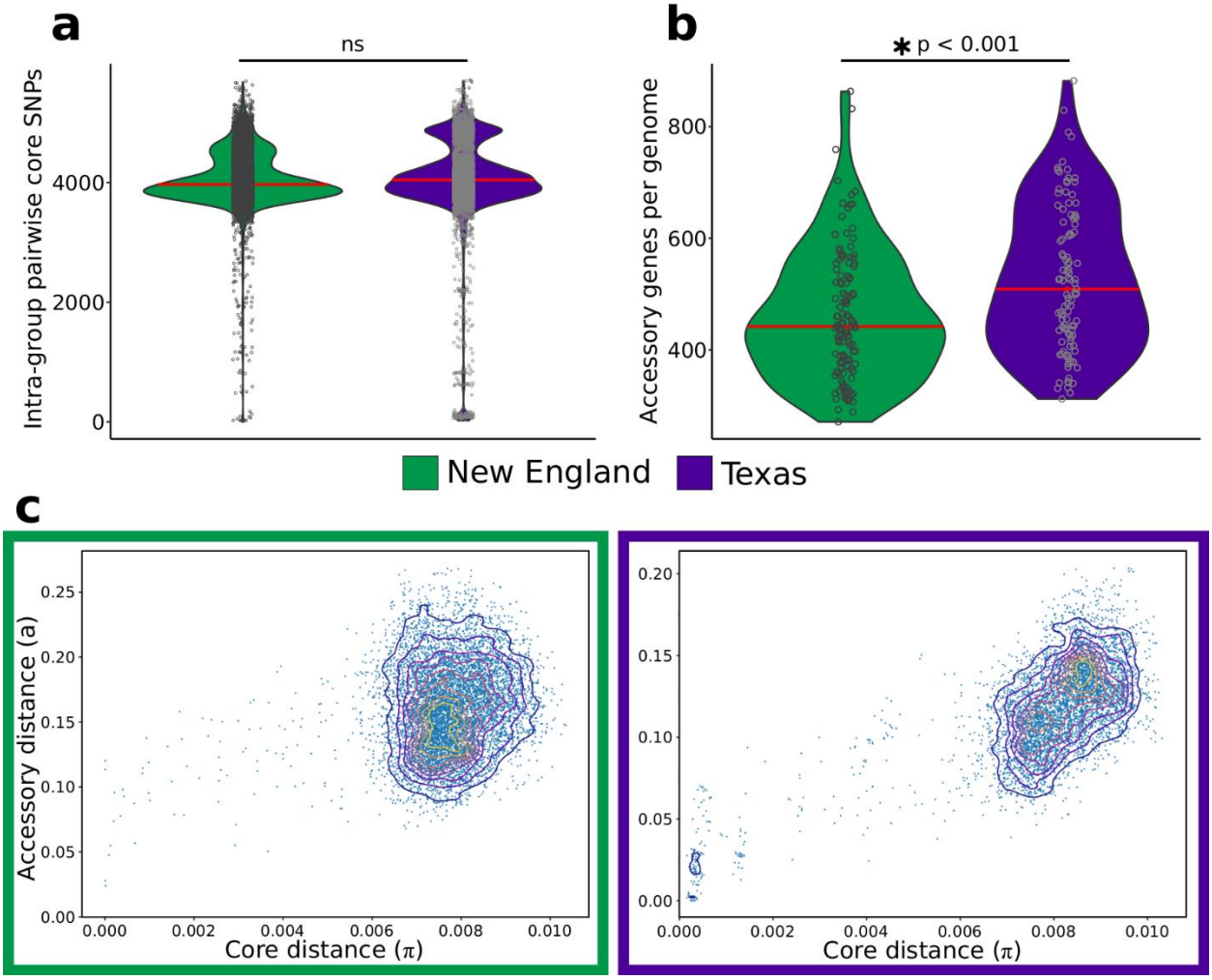

Figure 1.3 Pan-genome differences between New England ($n = 126$ genomes) and Texas ($n = 107$ genomes) populations. **a** Pairwise core SNPs. **b** Number of accessory genes per genome. Red horizontal lines represent median values. **c** Contour plots of pairwise distances between genomes in terms of their core genome divergence (measured by SNP density across the core genome) and the difference in their accessory genomes (measured by the Jaccard distance based on the variation in the gene content of their sequences).

Heterogeneity in the frequency of recombination

Many bacterial species are known to receive DNA fragments from other taxa and integrate them into their genomes through recombination (Vos, 2009). The process of recombination plays a fundamental role in the evolution of many bacterial pathogens and has been implicated in the emergence of new phenotypic traits such as AMR, ecological adaptation

and host colonization (Vos, 2009). Recombination is known to contribute to the evolution of *Staphylococcus*, as in the case of *S. aureus* and *Staphylococcus epidermidis*, which have been reported to frequently recombine within and between species (Méric et al., 2015). Here, we sought to determine the frequency and characteristics of recombination in *S. pseudintermedius* because this process may likely contribute to its genomic variation and diversification.

We used different methods to elucidate the impact of recombination on the population structure and evolution of *S. pseudintermedius*. First, under the null hypothesis of no recombination, we calculated the pairwise homoplasy index (PHI) (Bruen et al., 2006) and detected evidence for significant recombination in the core genome of the combined New England and Texas dataset ($p = 0.0$, permutation test). Using the sequence alignment of the 1,776 core genes shared by the New England and Texas genomes, we used the SplitsTree network (Huson, 1998) to visualize the reticulated, star-like evolutionary history of the combined *S. pseudintermedius* populations (**Figure 1.4a**). We next sought to identify frequently recombining genes in each population by running fastGEAR (Mostowy et al., 2017) on individual sequence alignments of core and shared accessory genes. The program fastGEAR identifies both recent (i.e., recombination events affecting a few strains) and ancestral (i.e., recombination events affecting entire lineages) recombinations (**Figure 1.4b**, Table S1.7). In the New England population, we found that a total of 961 genes have had a history of recombination. A total of 922 and 199 genes were involved in recent recombination and ancestral recombination, respectively. In contrast, we identified a total of 869 genes that have experienced recombination in the Texas population, with 799 and 184 genes that have experienced recent and ancestral recombination, respectively. Although many of the frequently recombined genes have

hypothetical or unknown functions, we found that *copB*, which plays a role as a copper-translocating P-type ATPase (Zapotoczna et al., 2018), is frequently recombined in both populations. In *S. aureus*, this gene is part of the operon that promotes copper hypertolerance and enhanced resistance to phagocytic killing (Zapotoczna et al., 2018). Metal resistance genes, including *copB*, are carried by some SCC*mec* variants in both human and animal MRSA and other species of *Staphylococcus* (Argudín & Butaye, 2016). Hence, *copB* is likely co-translocated and/or co-selected with AMR genes in the SCC*mec*. Another gene that has experienced frequent recombination is *fnbA*, which encodes a fibronectin-binding protein and thus promotes colonization to different anatomical sites of the eukaryotic host (Loughman et al., 2008). The gene *bca*, which encodes the surface-associated C protein alpha antigen (Michel et al., 1992), is also frequently recombined in both populations. In *S. aureus*, sequence variation in *fnbA* has been previously reported to contribute to differences in antigenicity between strains (Loughman et al., 2008; Murai et al., 2016).

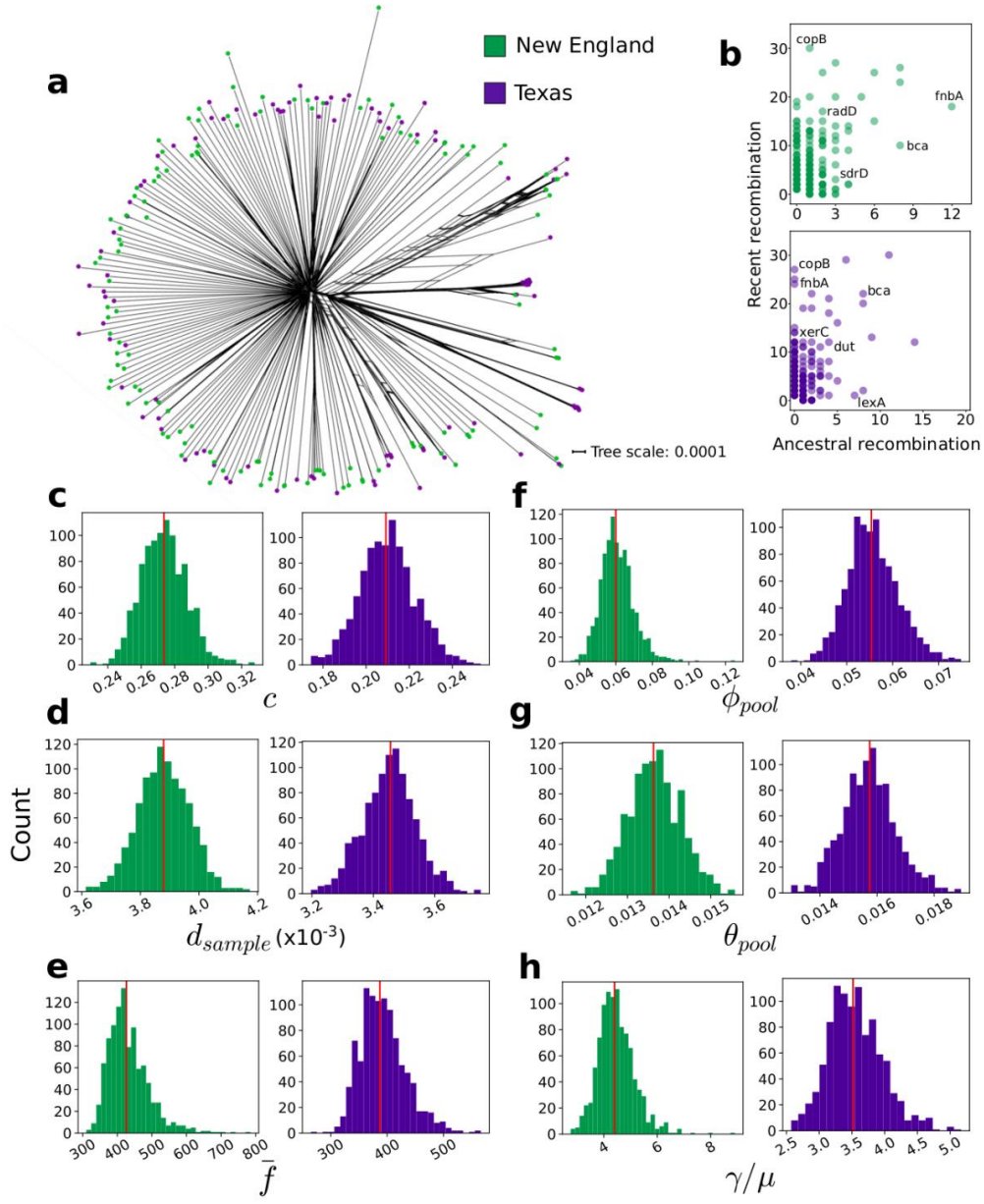

Figure 1.4 Recombination in *S. pseudintermedius* genomes from New England and Texas. **a** A phylogenetic SplitsTree network of the core genome. Scale bar represents nucleotide substitutions per site. **b** Genes that have undergone recent and ancestral recombination. Horizontal axis shows the estimated number of ancestral recombinations, and vertical axis shows the estimated number of recent recombinations. Names of some of the most frequently recombined genes with known functions are shown. List of all recombined genes are shown in Table S1.7. **c**–**g** Evolutionary and recombination parameters calculated by mcorr.: **c** recombination coverage; **d** diversity; **f** mean fragment size of a recombination event; ϕ recombinational divergence; θ mutational divergence; γ/μ relative rate of recombination to mutation (equivalent to ratio of ϕ/θ). Red vertical line in each plot represents the inferred value of each parameter calculated with 1000 bootstrapped replicates.

We also wanted to compare different genome-wide evolutionary and recombination parameters between New England and Texas populations. We used the core genome alignment of each dataset as an input to mcorr (Lin & Kussell, 2019) (**Figure 1.4c-g**; Table S1.8). The recombination coverage (c), which indicates the fraction of the genome whose diversity was derived from recombination events since the last common ancestor of the population, was estimated to be 0.274 (standard deviation [sd] = 0.014) and 0.210 (sd = 0.013) in New England and Texas, respectively. Recombination coverage ranges from 0 (clonal evolution) to 1 (complete recombination) (Lin & Kussell, 2019). These values indicate that 27.4% and 21.0% of sites in any one genome from New England and Texas, respectively, originated from recombination events. The diversity (d) is the probability that a pair of genomes will differ at any locus and is estimated from the diversity generated from both recombination and accumulation of mutations of the clonal lineage (Lin & Kussell, 2019). This parameter was estimated to be 3.88×10^{-3} (sd = 8.80×10^{-5}) and 3.46×10^{-3} (sd = 8.73×10^{-5}) in New England and Texas, respectively. The mean fragment size ($\bar{f}$) of a recombination event was estimated to be 426.14 (sd = 60.22) bp and 387.09 (sd = 40.55) bp in New England and Texas, respectively. Recombinational divergence (ϕ) was estimated to be 0.060 (sd = 0.009) and 0.055 (sd = 0.006) in New England and Texas, respectively. The mutational divergence (θ), which refers to the mean number of mutations per locus since divergence of a pair of homologous sites in the pool, was estimated to be 0.014 (sd = 6.39×10^{-4}) and 0.016 (sd = 9.48×10^{-4}) in New England and Texas, respectively. Lastly, the ratio ϕ/θ (or γ/μ), which gives the relative rate of recombination to mutation, was estimated to be 4.40 (sd = 0.683) and 3.52 (sd = 0.413) in New England and Texas, respectively. For all parameters, we found significant differences between the two populations (p-value < 0.001 for all six parameters; Welch's t test). For comparison, *S. aureus*

has lower values for recombination divergence (0.042) and ϕ/θ (1.0), and higher values for recombination coverage (0.36), diversity (0.015), mean fragment size (550 bp) and mutational divergence (0.042) (Lin & Kussell, 2019). Overall, we found that heterogeneity in rates and characteristics of recombination has contributed to shaping the genomic structure and diversity of the New England and Texas populations of *S. pseudintermedius*.

Discussion

We present a comparative population genomics study of *S. pseudintermedius* in the United States. Results revealed the co-circulation of phylogenetically diverse lineages that have access to a large pool of accessory genes, including AMR genes. For *S. pseudintermedius*, the long term clinical and public health significance of MRSP and multidrug resistant lineages remains unclear. Suffice to say that companion animals can act as a reservoir of resistant strains that can be transmitted to humans or other animals, as has been reported for other *Staphylococcus* species that colonize multiple eukaryotic hosts (Guardabassi et al., 2004; Richardson et al., 2018). *S. pseudintermedius* also appears to be undergoing a considerable amount of genetic exchange which can lead to the emergence of high-risk clones and those with hitherto uncharacterized virulence capacity. The origins of recombined and horizontally acquired DNA remains unclear, but they were likely derived from other *S. pseudintermedius* strains, other *Staphylococcus* species that they co-inhabit their niche with, and even distantly related organisms. This has been demonstrated in a recent study that reported that many resistance genes have been frequently exchanged between different *Staphylococcus* species and between *Staphylococcus* and other non-*Staphylococcus* taxa (John et al., 2019). The ability of *S. pseudintermedius* to take up and integrate large DNA segments highlights the pathogen's

capacity for rapid evolution and adaptation to primary and secondary hosts, similar to what has been observed in *S. aureus* (Méric et al., 2015; Richardson et al., 2018). Frequent recombination may also contribute in part to the pathogenicity of *S. pseudintermedius* and its ability to colonize other host species. Recombination is not the only process that contributes to pathogen diversity. Extensive phylogenetic intermixing of strains spanning a large geographical distance suggests substantial translocation and/or exchange of strains. This may likely be a result of either widespread pet transport and distribution across the country or that there is a country-wide pool of circulating MSSP and MRSP lineages, with indication of local expansion and diversification of certain MRSP clones in each region.

A few limitations need to be acknowledged. The retrospective nature of the isolate collection we used and sampling bias due to the inclusion of only clinical submissions mean that sampling may have missed some less prevalent genotypes and counties in New England. We also did not have access to data regarding the antibiotic history, treatment received or clinical outcomes of the dogs and cats from which isolates were obtained. Another limitation is our use of databases for querying AMR genes, virulence genes and SCC*mec* types. These databases were developed primarily for *S. aureus*, of which thousands of genomes have already been sequenced (for example, see Ref. (Richardson et al., 2018)). Less is known of other *Staphylococcus* species however and this gap in current knowledge may have important consequences. For example, newly discovered genetic variants (as in the case of the five unknown SCC*mec* types) may be present in other species but remain invisible from current *in silico* detection methods and databases, which make them difficult to track over the long term. This also means that clinical decisions and treatment options made when treating diseases caused by non-*aureus* species are

based on assumptions from studies of *S. aureus*. Estimates of evolutionary and recombination parameters show considerable differences not only between the two *S. pseudintermedius* populations but also between *S. pseudintermedius* and *S. aureus*, which can impact population dynamics and the emergence of favorable phenotypes. Hence, more extensive and large-scale genome sequencing of non-*aureus* species will greatly augment this gap in our knowledge of the diversity, pathogenicity and resistance mechanisms of other *Staphylococcus* species that can potentially develop into formidable high-risk pathogens. Notwithstanding these limitations, we obtained sufficient representation of the diversity of *S. pseudintermedius* and the widespread distribution of MRSP and multidrug resistant clones even within a few years of sampling, which was the main focus of this study.

The results presented here open multiple avenues for future research. First, the integration of genomic epidemiology in veterinary medicine will advance our ability to identify targets for strain differentiation, surveillance and diagnostics. It is imperative that long-term monitoring of the eight major STs with the highest number of horizontally acquired AMR genes (STs 45, 64, 68, 71, 84, 150, 151 and 181) should be carried out, as they appear to be expanding in New England and Texas and likely in other parts of the country. Future work using functional assays is also needed to precisely characterize the virulence factors of *S. pseudintermedius*, instead of simply relying on *S. aureus* studies. Our results on the most frequently recombining genes (e.g., *copB*) provides an important starting point to understand the genetic basis of *S. pseudintermedius* pathogenicity. Second, local and global surveillance of resistant phenotypes is particularly critical to detect and limit the emergence of new resistant clones with the capacity to spread and switch between animals and humans. Future work should include determining the prevalence of

MRSP and multidrug resistant strains across a range of domestic and wildlife animals, and specific groups of people who are in constant proximity to animals (e.g., veterinary personnel, pet owners, breeders, pet shop owners, pet adoption workers, farmers). Close contact between animals and humans has facilitated multiple host-switching events in *S. aureus* (Shepheard et al., 2013; Spoor et al., 2013; Weinert et al., 2012) and it is not unreasonable to assume that it can also occur in *S. pseudintermedius*. Investigation into the extent of host switching (i.e., animal-to-human and between dogs and other animal species) and the adaptive changes associated with radical changes in host ecology will reveal key host–pathogen interactions that could be targeted for novel clinical interventions and therapies. Third, the precise origins of MRSP (whether from human-associated methicillin-resistant *S. aureus*, methicillin-susceptible animal-associated *S. pseudintermedius* or other staphylococcal species) remains unclear and genomic sequencing approaches will certainly advance our understanding of how and when it emerged in companion animals. We also encourage the application of population genomics as an integral component of the One Health initiative which emphasizes that the health of humans, animals and the environmental are inextricably linked (Destoumieux-Garzón et al., 2018). Taken together, our findings provide a genomic framework that will provide a critical foundation and practical support for future studies investigating the population structure and dynamics, drug resistance and diversification of *S. pseudintermedius* as well as the risks to pets, pet owners and veterinary personnel. It will facilitate more deeply informed genomic tracking and surveillance for the emergence and convergence of virulence and AMR in certain genotypes of *S. pseudintermedius* and in different geographical regions. In summary, we underscore the value of elucidating the population genomic structure and evolution of this increasingly important pathogen to advance the health of man's favorite companion and oldest domestic animal.

References

Abouelkhair, M. A., Bemis, D. A., Giannone, R. J., Frank, L. A., & Kania, S. A. (2018). Characterization of a leukocidin identified in *Staphylococcus pseudintermedius*. *PloS One*, *13*(9), e0204450. https://doi.org/10.1371/journal.pone.0204450

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andam, C. P., Mitchell, P. K., Callendrello, A., Chang, Q., Corander, J., Chaguza, C., McGee, L., Beall, B. W., & Hanage, W. P. (2017). Genomic epidemiology of penicillin-nonsusceptible pneumococci with nonvaccine serotypes causing invasive disease in the United States. *Journal of Clinical Microbiology*, *55*(4), 1104–1115. https://doi.org/10.1128/JCM.02453-16

Argudín, M. A., & Butaye, P. (2016). Dissemination of metal resistance genes among animal methicillin-resistant coagulase-negative Staphylococci. *Research in Veterinary Science*, *105*, 192–194. https://doi.org/10.1016/j.rvsc.2016.02.009

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bannoehr, J., & Guardabassi, L. (2012). *Staphylococcus pseudintermedius* in the dog: Taxonomy, diagnostics, ecology, epidemiology and pathogenicity. *Veterinary Dermatology*, *23*(4), 253–266, e51-52. https://doi.org/10.1111/j.1365-3164.2012.01046.x

Börjesson, S., Gómez-Sanz, E., Ekström, K., Torres, C., & Grönlund, U. (2015). *Staphylococcus pseudintermedius* can be misdiagnosed as *Staphylococcus aureus* in humans with dog bite wounds. *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology*, *34*(4), 839–844. https://doi.org/10.1007/s10096-014-2300-y

Bosi, E., Monk, J. M., Aziz, R. K., Fondi, M., Nizet, V., & Palsson, B. Ø. (2016). Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(26), E3801-3809. https://doi.org/10.1073/pnas.1523199113

Bruen, T. C., Philippe, H., & Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics*, *172*(4), 2665–2681. https://doi.org/10.1534/genetics.105.048975

Chewapreecha, C., Harris, S. R., Croucher, N. J., Turner, C., Marttinen, P., Cheng, L., Pessia, A., Aanensen, D. M., Mather, A. E., Page, A. J., Salter, S. J., Harris, D., Nosten, F., Goldblatt, D., Corander, J., Parkhill, J., Turner, P., & Bentley, S. D. (2014). Dense genomic sampling identifies highways of pneumococcal recombination. *Nature Genetics*, *46*(3), 305–309. https://doi.org/10.1038/ng.2895

CLSI Performance standards for antimicrobial disk and dilution susceptibility tests for bacteria isolated from animals. VET01 (5th ed.). (2018). Clinical and Laboratory Standards Institute.

Croucher, N. J., Coupland, P. G., Stevenson, A. E., Callendrello, A., Bentley, S. D., & Hanage, W. P. (2014). Diversification of bacterial genome content through distinct mechanisms over different timescales. *Nature Communications*, *5*, 5471. https://doi.org/10.1038/ncomms6471

Destoumieux-Garzón, D., Mavingui, P., Boetsch, G., Boissier, J., Darriet, F., Duboz, P., Fritsch, C., Giraudoux, P., Le Roux, F., Morand, S., Paillard, C., Pontier, D., Sueur, C., & Voituron, Y. (2018). The One Health concept: 10 Years old and a long road ahead. *Frontiers in Veterinary Science*, *5*, 14. https://doi.org/10.3389/fvets.2018.00014

Enright, A. J., Van Dongen, S., & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, *30*(7), 1575–1584.

Firth, N., Jensen, S. O., Kwong, S. M., Skurray, R. A., & Ramsay, J. P. (2018). Staphylococcal plasmids, transposable and integrative elements. *Microbiology Spectrum*, *6*(6). https://doi.org/10.1128/microbiolspec.GPP3-0030-2018

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics (Oxford, England)*, *28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Guardabassi, L., Schwarz, S., & Lloyd, D. H. (2004(21–23)). Pet animals as reservoirs of antimicrobial-resistant bacteria. *The Journal of Antimicrobial Chemotherapy*, *54*(2), 321–332. https://doi.org/10.1093/jac/dkh332

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hunt, M., Mather, A. E., Sánchez-Busó, L., Page, A. J., Parkhill, J., Keane, J. A., & Harris, S. R. (2017). ARIBA: Rapid antimicrobial resistance genotyping directly from sequencing reads. *Microbial Genomics*, *3*(10), e000131. https://doi.org/10.1099/mgen.0.000131

Huson, D. H. (1998). SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics (Oxford, England)*, *14*(1), 68–73.

Inouye, M., Dashnow, H., Raven, L.-A., Schultz, M. B., Pope, B. J., Tomita, T., Zobel, J., & Holt, K. E. (2014). SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Medicine*, *6*(11), 90. https://doi.org/10.1186/s13073-014-0090-6

International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC). (2009). Classification of staphylococcal cassette chromosome mec (SCCmec): Guidelines for reporting novel SCCmec elements. *Antimicrobial Agents and Chemotherapy*, *53*(12), 4961–4967. https://doi.org/10.1128/AAC.00579-09

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 5114. https://doi.org/10.1038/s41467-018-07641-9

John, J., George, S., Nori, S. R. C., & Nelson-Sathi, S. (2019). Phylogenomic analysis reveals the evolutionary route of resistant genes in *Staphylococcus aureus*. *Genome Biology and Evolution*, *11*(10), 2917–2926. https://doi.org/10.1093/gbe/evz213

Kasai, T., Saegusa, S., Shirai, M., Murakami, M., & Kato, Y. (2016). New categories designated as healthcare-associated and community-associated methicillin-resistant *Staphylococcus pseudintermedius* in dogs. *Microbiology and Immunology*, *60*(8), 540–551. https://doi.org/10.1111/1348-0421.12401

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066.

Kaya, H., Hasman, H., Larsen, J., Stegger, M., Johannesen, T. B., Allesøe, R. L., Lemvigh, C. K., Aarestrup, F. M., Lund, O., & Larsen, A. R. (2018). SCCmecFinder, a web-based tool for typing of Staphylococcal Cassette Chromosome mec in *Staphylococcus aureus* using whole-genome sequence data. *MSphere*, *3*(1). https://doi.org/10.1128/mSphere.00612-17

Lakhundi, S., & Zhang, K. (2018). Methicillin-resistant *Staphylococcus aureus*: Molecular characterization, evolution, and epidemiology. *Clinical Microbiology Reviews*, *31*(4). https://doi.org/10.1128/CMR.00020-18

Lees, J. A., Harris, S. R., Tonkin-Hill, G., Gladstone, R. A., Lo, S. W., Weiser, J. N., Corander, J., Bentley, S. D., & Croucher, N. J. (2019). Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Research*, *29*(2), 304–316. https://doi.org/10.1101/gr.241455.118

Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, *44*(W1), W242–W245. https://doi.org/10.1093/nar/gkw290

Lin, M., & Kussell, E. (2019). Inferring bacterial recombination rates from large-scale sequencing datasets. *Nature Methods*, *16*(2), 199–204. https://doi.org/10.1038/s41592-018-0293-7

Little, S. V., Bryan, L. K., Hillhouse, A. E., Cohen, N. D., & Lawhon, S. D. (2019). Characterization of agr groups of *Staphylococcus pseudintermedius* isolates from dogs in Texas. *MSphere*, *4*(2). https://doi.org/10.1128/mSphere.00033-19

Liu, B., Zheng, D., Jin, Q., Chen, L., & Yang, J. (2019). VFDB 2019: A comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Research*, *47*(D1), D687–D692. https://doi.org/10.1093/nar/gky1080

Llarrull, L. I., Fisher, J. F., & Mobashery, S. (2009). Molecular basis and phenotype of methicillin resistance in *Staphylococcus aureus* and insights into new beta-lactams that meet the challenge. *Antimicrobial Agents and Chemotherapy*, *53*(10), 4051–4063. https://doi.org/10.1128/AAC.00084-09

Loughman, A., Sweeney, T., Keane, F. M., Pietrocola, G., Speziale, P., & Foster, T. J. (2008). Sequence diversity in the A domain of *Staphylococcus aureus* fibronectin-binding protein A. *BMC Microbiology*, *8*, 74. https://doi.org/10.1186/1471-2180-8-74

Maali, Y., Badiou, C., Martins-Simões, P., Hodille, E., Bes, M., Vandenesch, F., Lina, G., Diot, A., Laurent, F., & Trouillet-Assant, S. (2018). Understanding the virulence of *Staphylococcus pseudintermedius*: A major role of pore-forming toxins. *Frontiers in Cellular and Infection Microbiology*, *8*, 221. https://doi.org/10.3389/fcimb.2018.00221

McCarthy, A. J., Harrison, E. M., Stanczak-Mrozek, K., Leggett, B., Waller, A., Holmes, M. A., Lloyd, D. H., Lindsay, J. A., & Loeffler, A. (2015). Genomic insights into the rapid emergence and evolution of MDR in *Staphylococcus pseudintermedius*. *The Journal of Antimicrobial Chemotherapy*, *70*(4), 997–1007. https://doi.org/10.1093/jac/dku496

Méric, G., Miragaia, M., de Been, M., Yahara, K., Pascoe, B., Mageiros, L., Mikhail, J., Harris, L. G., Wilkinson, T. S., Rolo, J., Lamble, S., Bray, J. E., Jolley, K. A., Hanage, W. P., Bowden, R., Maiden, M. C. J., Mack, D., de Lencastre, H., Feil, E. J., … Sheppard, S. K. (2015). Ecological overlap and horizontal gene transfer in *Staphylococcus aureus* and *Staphylococcus epidermidis*. *Genome Biology and Evolution*, *7*(5), 1313–1328. https://doi.org/10.1093/gbe/evv066

Michel, J. L., Madoff, L. C., Olson, K., Kling, D. E., Kasper, D. L., & Ausubel, F. M. (1992). Large, identical, tandem repeating units in the C protein alpha antigen gene, bca, of group B streptococci. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(21), 10060–10064. https://doi.org/10.1073/pnas.89.21.10060

Mostowy, R., Croucher, N. J., Andam, C. P., Corander, J., Hanage, W. P., & Marttinen, P. (2017). Efficient inference of recent and ancestral recombination within bacterial

populations. *Molecular Biology and Evolution*, *34*(5), 1167–1182. https://doi.org/10.1093/molbev/msx066

Murai, M., Moriyama, H., Hata, E., Takeuchi, F., & Amemura-Maekawa, J. (2016). Variation and association of fibronectin-binding protein genes fnbA and fnbB in *Staphylococcus aureus* Japanese isolates. *Microbiology and Immunology*, *60*(5), 312–325. https://doi.org/10.1111/1348-0421.12377

Nisa, S., Bercker, C., Midwinter, A. C., Bruce, I., Graham, C. F., Venter, P., Bell, A., French, N. P., Benschop, J., Bailey, K. M., & Wilkinson, D. A. (2019). Combining MALDI-TOF and genomics in the study of methicillin resistant and multidrug resistant *Staphylococcus pseudintermedius* in New Zealand. *Scientific Reports*, *9*(1), 1271. https://doi.org/10.1038/s41598-018-37503-9

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

Pain, M., Hjerde, E., Klingenberg, C., & Cavanagh, J. P. (2019). Comparative genomic analysis of *Staphylococcus haemolyticus* reveals key to hospital adaptation and pathogenicity. *Frontiers in Microbiology*, *10*, 2096. https://doi.org/10.3389/fmicb.2019.02096

Park, S.-C., Lee, K., Kim, Y. O., Won, S., & Chun, J. (2019). Large-scale genomics reveals the genetic characteristics of seven species and importance of phylogenetic distance for estimating pan-genome size. *Frontiers in Microbiology*, *10*, 834. https://doi.org/10.3389/fmicb.2019.00834

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Paul, N. C., Moodley, A., Ghibaudo, G., & Guardabassi, L. (2011). Carriage of methicillin-resistant *Staphylococcus pseudintermedius* in small animal veterinarians: Indirect evidence of zoonotic transmission. *Zoonoses and Public Health*, *58*(8), 533–539. https://doi.org/10.1111/j.1863-2378.2011.01398.x

Pinchuk, I. V., Beswick, E. J., & Reyes, V. E. (2010). Staphylococcal enterotoxins. *Toxins*, *2*(8), 2177–2197. https://doi.org/10.3390/toxins2082177

Richardson, E. J., Bacigalupe, R., Harrison, E. M., Weinert, L. A., Lycett, S., Vrieling, M., Robb, K., Hoskisson, P. A., Holden, M. T. G., Feil, E. J., Paterson, G. K., Tong, S. Y. C., Shittu, A., van Wamel, W., Aanensen, D. M., Parkhill, J., Peacock, S. J., Corander, J., Holmes, M., & Fitzgerald, J. R. (2018). Gene exchange drives the ecological success of a

multi-host bacterial pathogen. *Nature Ecology & Evolution*, *2*(9), 1468–1478. https://doi.org/10.1038/s41559-018-0617-0

Rolo, J., Worning, P., Nielsen, J. B., Bowden, R., Bouchami, O., Damborg, P., Guardabassi, L., Perreten, V., Tomasz, A., Westh, H., de Lencastre, H., & Miragaia, M. (2017). Evolutionary origin of the Staphylococcal Cassette Chromosome mec (SCCmec). *Antimicrobial Agents and Chemotherapy*, *61*(6). https://doi.org/10.1128/AAC.02302-16

Ruscher, C., Lübke-Becker, A., Wleklinski, C.-G., Soba, A., Wieler, L. H., & Walther, B. (2009). Prevalence of methicillin-resistant *Staphylococcus pseudintermedius* isolated from clinical samples of companion animals and equidaes. *Veterinary Microbiology*, *136*(1–2), 197–201. https://doi.org/10.1016/j.vetmic.2008.10.023

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Shepheard, M. A., Fleming, V. M., Connor, T. R., Corander, J., Feil, E. J., Fraser, C., & Hanage, W. P. (2013). Historical zoonoses and other changes in host tropism of *Staphylococcus aureus*, identified by phylogenetic analysis of a population dataset. *PloS One*, *8*(5), e62369. https://doi.org/10.1371/journal.pone.0062369

Sheppard, S. K., Guttman, D. S., & Fitzgerald, J. R. (2018). Population genomics of bacterial host adaptation. *Nature Reviews. Genetics*, *19*(9), 549–565. https://doi.org/10.1038/s41576-018-0032-z

Solyman, S. M., Black, C. C., Duim, B., Perreten, V., van Duijkeren, E., Wagenaar, J. A., Eberlein, L. C., Sadeghi, L. N., Videla, R., Bemis, D. A., & Kania, S. A. (2013). Multilocus sequence typing for characterization of *Staphylococcus pseudintermedius*. *Journal of Clinical Microbiology*, *51*(1), 306–310. https://doi.org/10.1128/JCM.02421-12

Somayaji, R., Priyantha, M. a. R., Rubin, J. E., & Church, D. (2016). Human infections due to *Staphylococcus pseudintermedius*, an emerging zoonosis of canine origin: Report of 24 cases. *Diagnostic Microbiology and Infectious Disease*, *85*(4), 471–476. https://doi.org/10.1016/j.diagmicrobio.2016.05.008

Spoor, L. E., McAdam, P. R., Weinert, L. A., Rambaut, A., Hasman, H., Aarestrup, F. M., Kearns, A. M., Larsen, A. R., Skov, R. L., & Fitzgerald, J. R. (2013). Livestock origin for a human pandemic clone of community-associated methicillin-resistant *Staphylococcus aureus*. *MBio*, *4*(4). https://doi.org/10.1128/mBio.00356-13

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford, England)*, *22*(21), 2688–2690. https://doi.org/10.1093/bioinformatics/btl446

Tagini, F., & Greub, G. (2017). Bacterial genome sequencing in clinical microbiology: A pathogen-oriented review. *European Journal of Clinical Microbiology & Infectious*

Diseases: Official Publication of the European Society of Clinical Microbiology, *36*(11), 2007–2020. https://doi.org/10.1007/s10096-017-3024-6

Tavaré, S. (1986). Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. In *American Mathematical Society: Lectures on Mathematics in the Life Sciences* (Vol. 17, pp. 57–86). Amer Mathematical Society. http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0821811673

Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W., & Corander, J. (2018). RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. *Wellcome Open Research*, *3*, 93. https://doi.org/10.12688/wellcomeopenres.14694.1

Van Hoovels, L., Vankeerberghen, A., Boel, A., Van Vaerenbergh, K., & De Beenhouwer, H. (2006). First case of *Staphylococcus pseudintermedius* infection in a human. *Journal of Clinical Microbiology*, *44*(12), 4609–4612. https://doi.org/10.1128/JCM.01308-06

Vos, M. (2009). Why do bacteria engage in homologous recombination? *Trends in Microbiology*, *17*(6), 226–232. https://doi.org/10.1016/j.tim.2009.03.001

Wegener, A., Broens, E. M., Zomer, A., Spaninks, M., Wagenaar, J. A., & Duim, B. (2018). Comparative genomics of phenotypic antimicrobial resistances in methicillin-resistant *Staphylococcus pseudintermedius* of canine origin. *Veterinary Microbiology*, *225*, 125–131. https://doi.org/10.1016/j.vetmic.2018.09.013

Weinert, L. A., Welch, J. J., Suchard, M. A., Lemey, P., Rambaut, A., & Fitzgerald, J. R. (2012). Molecular dating of human-to-bovid host jumps by *Staphylococcus aureus* reveals an association with the spread of domestication. *Biology Letters*, *8*(5), 829–832. https://doi.org/10.1098/rsbl.2012.0290

Worthing, K. A., Schwendener, S., Perreten, V., Saputra, S., Coombs, G. W., Pang, S., Davies, M. R., Abraham, S., Trott, D. J., & Norris, J. M. (2018). Characterization of Staphylococcal cassette chromosome mec elements from methicillin-resistant *Staphylococcus pseudintermedius* infections in Australian animals. *MSphere*, *3*(6). https://doi.org/10.1128/mSphere.00491-18

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F. M., & Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. *The Journal of Antimicrobial Chemotherapy*, *67*(11), 2640–2644. https://doi.org/10.1093/jac/dks261

Zapotoczna, M., Riboldi, G. P., Moustafa, A. M., Dickson, E., Narechania, A., Morrissey, J. A., Planet, P. J., Holden, M. T. G., Waldron, K. J., & Geoghegan, J. A. (2018). Mobile-genetic-element-encoded hypertolerance to copper protects *Staphylococcus aureus* from killing by host phagocytes. *MBio*, *9*(5). https://doi.org/10.1128/mBio.00550-18

Appendix 1

Supplementary Tables S1.1-S1.9 listed below can be found at:
https://www.nature.com/articles/s42003-020-1009-y

Table S1.1 Accession numbers, metadata and genome characteristics of the 130 New England *S. pseudintermedius* genomes sequenced in this study and the 107 previously published genomes from Texas.

Table S1.2 List of genes (core and accessory genes) identified by Roary in New England genomes (n = 130 genomes), Texas genomes (n = 107 genomes), and in the combined New England and Texas genomes (n = 233 genomes).

Table S1.3 Pairwise ANI values (%) in New England genomes and in the combined New England and Texas genomes. These numbers correspond to Figure S1.3.

Table S1.4 Distribution of ABR genes in New England genomes identified by ARIBA and CARD database. This list includes horizontally acquired ABR genes and resistance alleles due to chromosomal mutations.

Table S1.5 Distribution of virulence genes in New England *S. pseudintermedius* genomes identified by ARIBA and VFDB database.

Table S1.6 Distribution of ABR genes in Texas genomes identified by ABRicate and CARD database. This list includes only horizontally acquired ABR genes.

Table S1.7 List of recombined genes in New England and Texas genomes inferred by fastGEAR.

Table S1.8 Evolutionary and recombination parameters that characterize New England and Texas genomes inferred by mcorr.

Table S1.9 Source data for Figures 2d-g, 3a-b, 4b-h

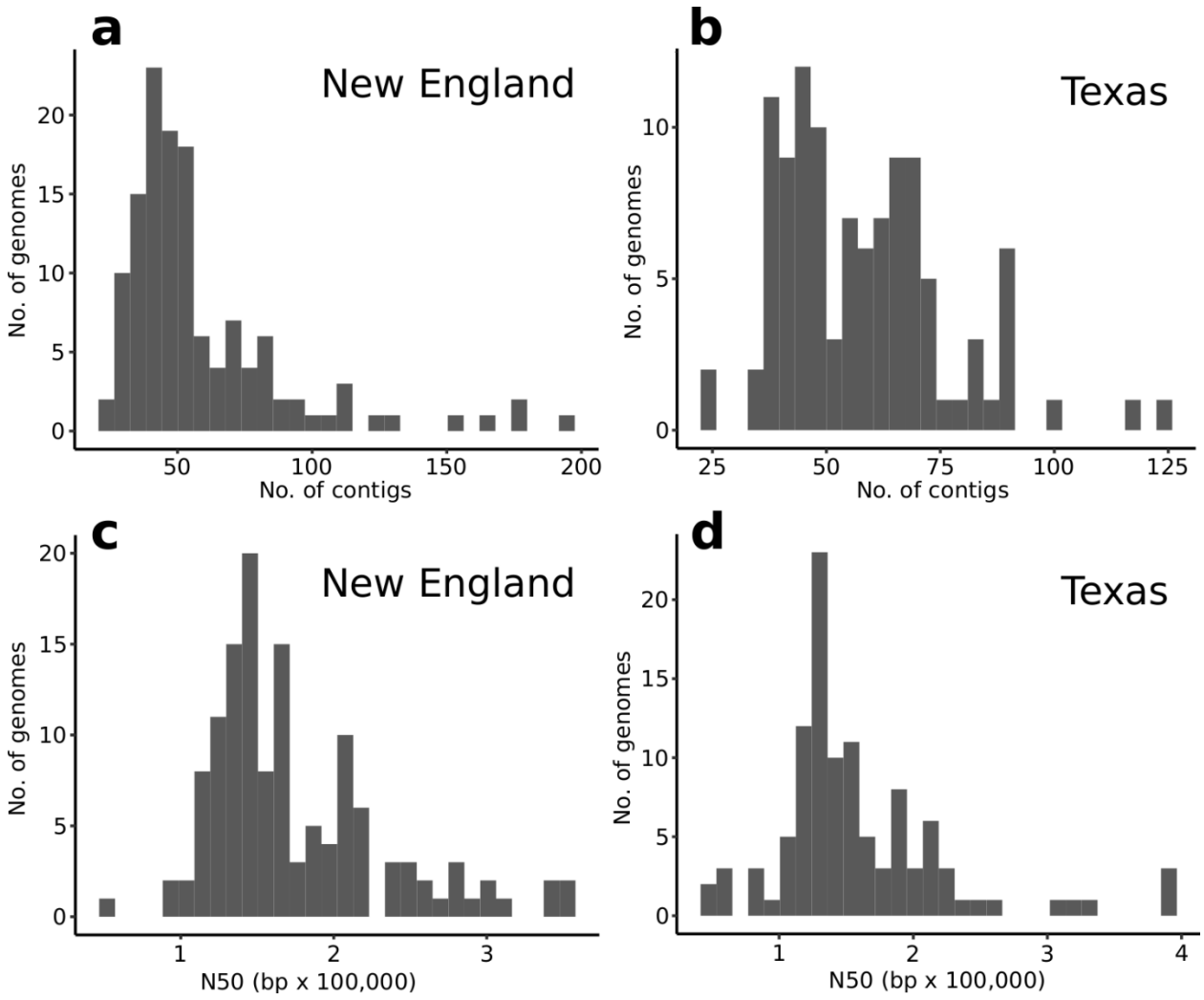

Figure S1.1 Distribution of number of contigs and N50. Histograms indicating contigs per genome for (a) New England and (b) Texas genomes; and N50 values for (c) New England and (d) Texas genomes.

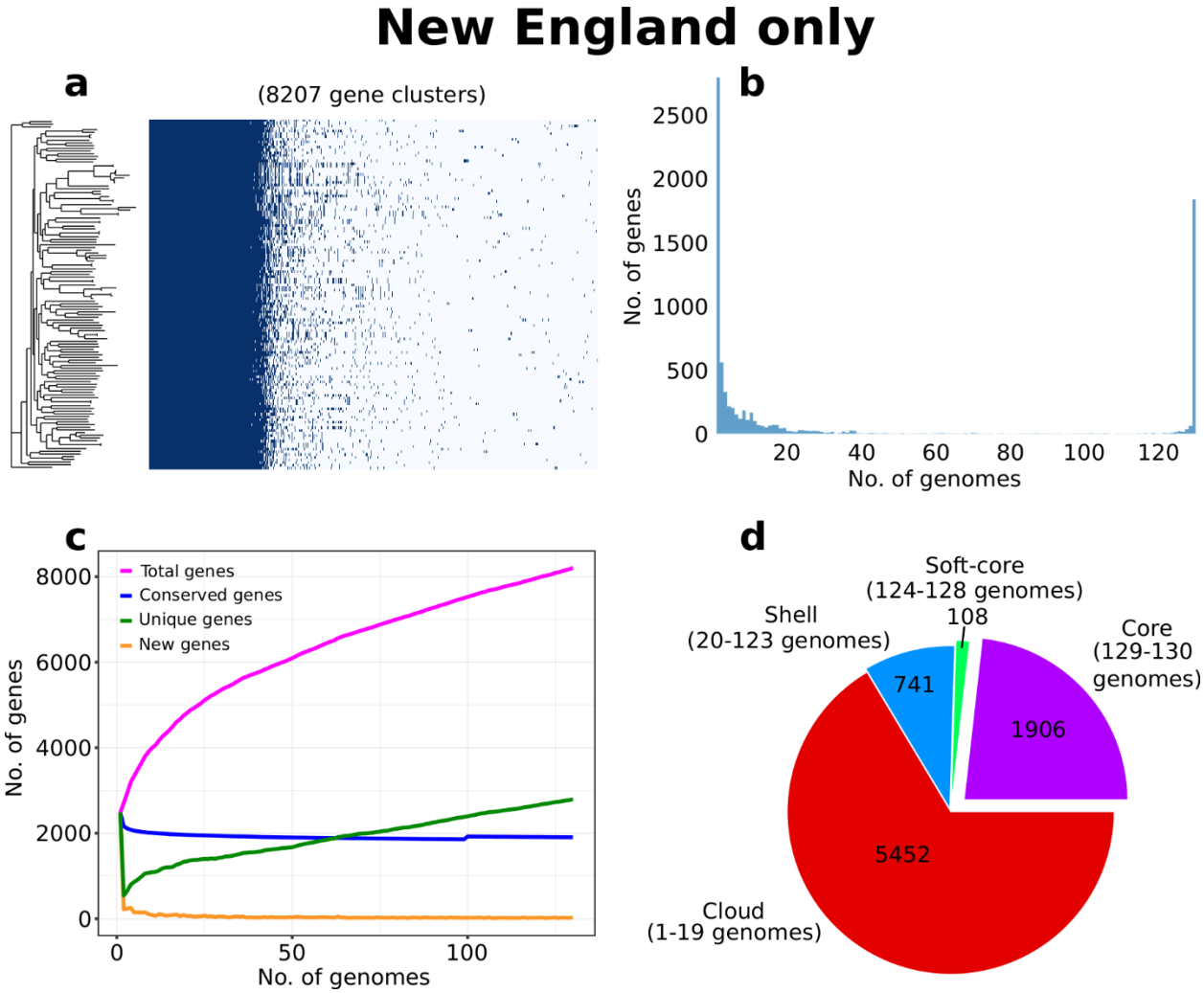

Figure S1.2 Pan-genome analyses of New England *S. pseudintermedius* (n = 130 genomes). (a) Matrix showing gene presence or absence per genome. Each row corresponds to a genome on the phylogeny. Each column represents a unique gene family. (b) The number of unique genes that are shared by any given number of genomes or unique to a single genome. (c) The size of the pan-genome, i.e., the totality of unique genes present in the population (pink line), the size of the core genome, i.e., genes that are present in at least 99% of the strains (blue line), the number of unique genes, i.e., genes unique to an individual strain (green line), and new genes, i.e., genes not found in the previously compared genomes (orange line) in relation to numbers of genomes compared. (d) Pie chart showing the distribution of core, soft core, shell and cloud genes.

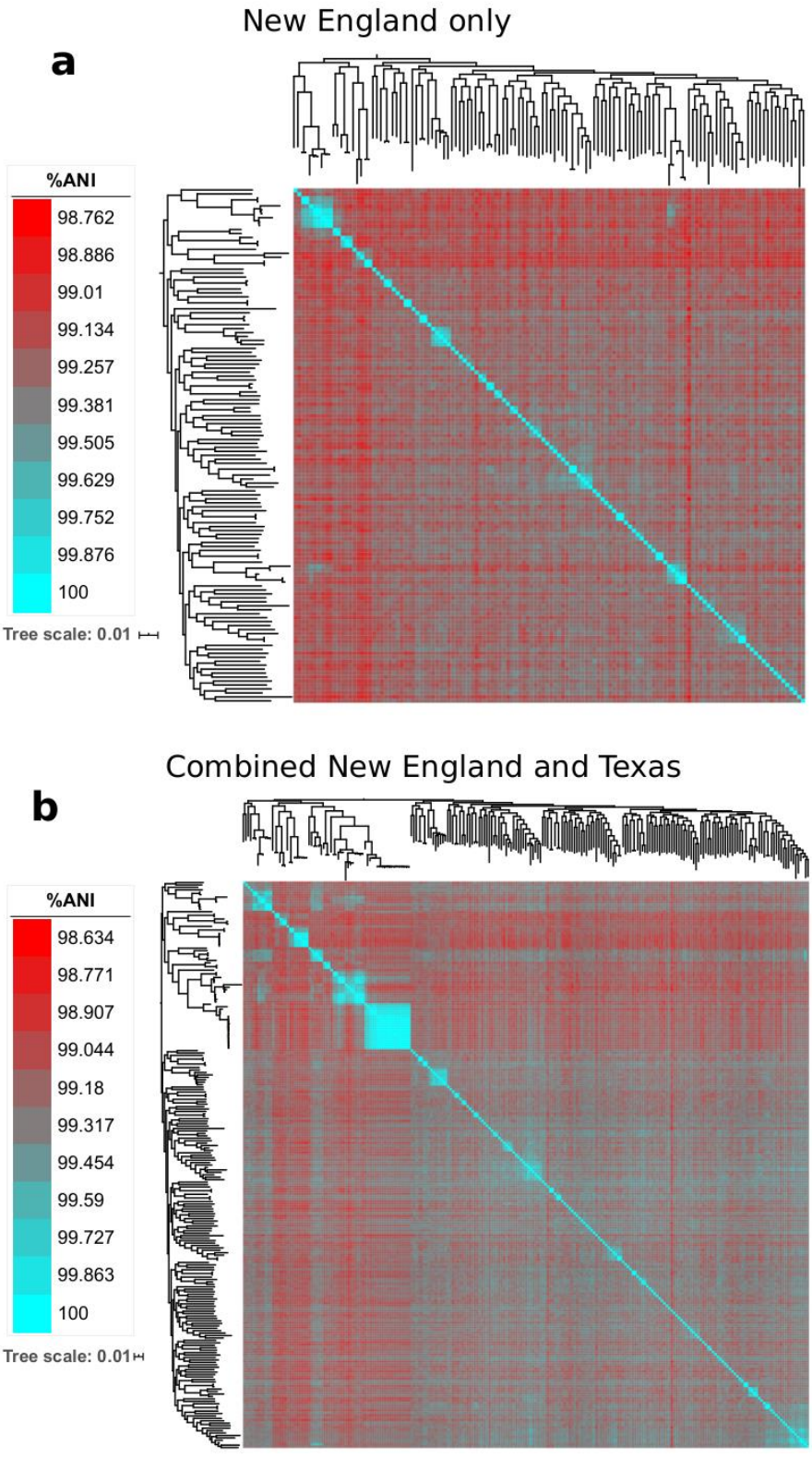

Figure S1.3 Heatmap matrix showing pairwise ANI comparison among all New England genomes (a) and among New England and Texas genomes (b).

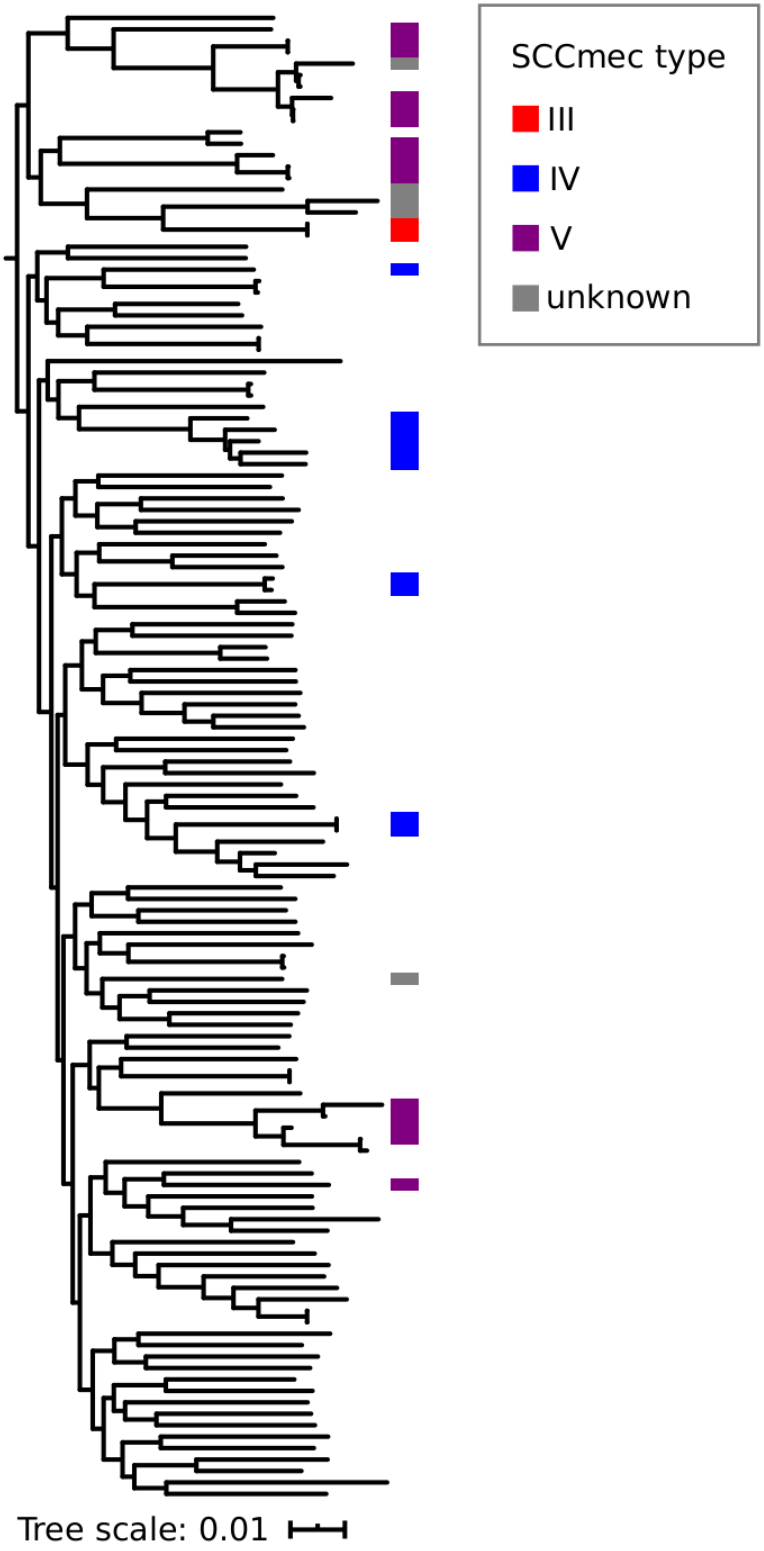

Figure S1.4 Distribution of SCCmec types in the New England population.

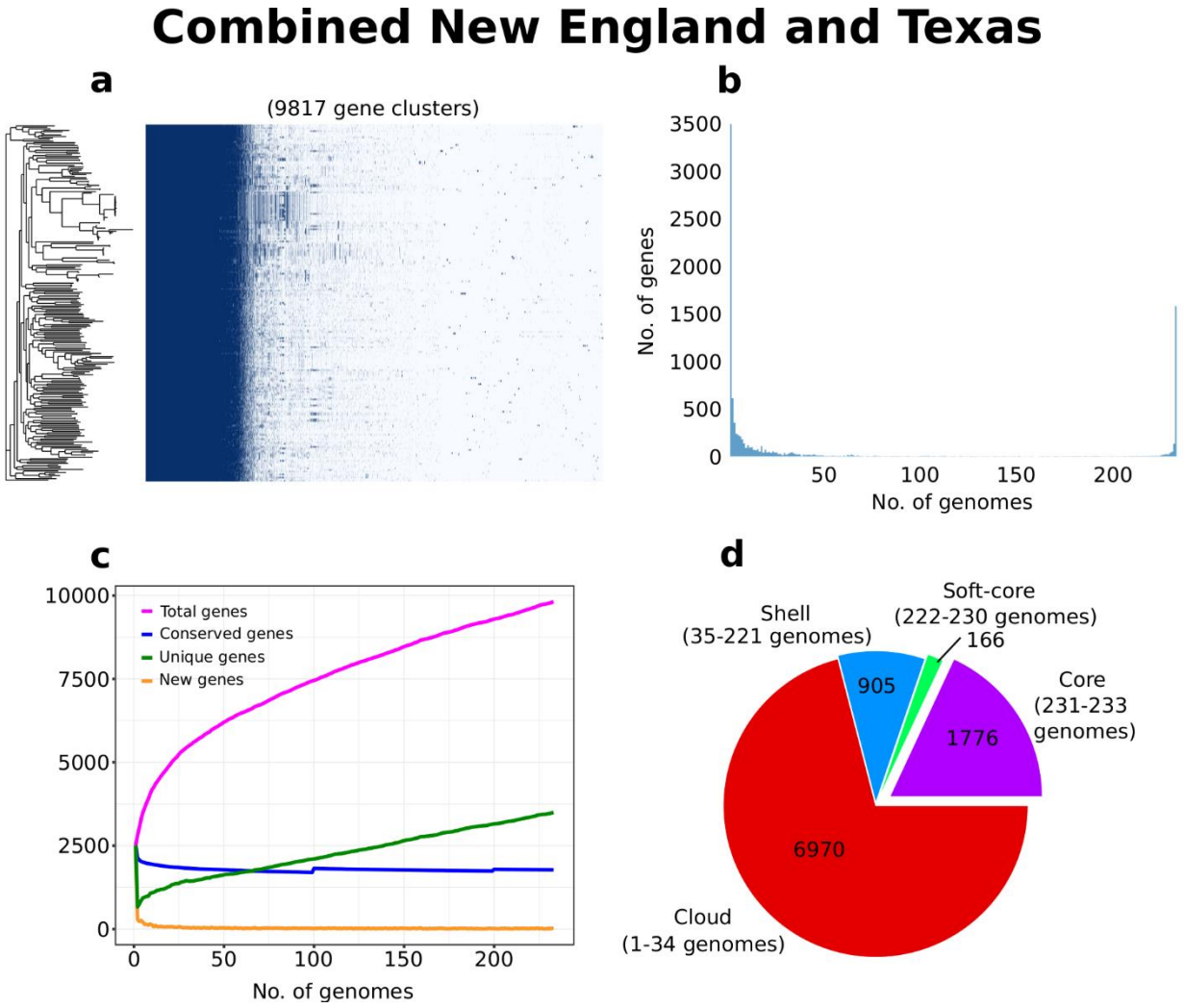

Figure S1.5 Pan-genome analyses of the combined New England and Texas genomes (126 New England and 107 Texas genomes). These are all from dogs. Color legend and definitions are identical to those described in Figure S1.2.

CHAPTER 2

Genomic epidemiology of methicillin-resistant and -susceptible *Staphylococcus aureus* from bloodstream infections

Joshua T. Smith[a], Elissa M. Eckhardt[b], Nicole B. Hansel[b], Tahmineh Rahmani Eliato[c], Isabell[a] W. Martin[b], Cheryl P. Andam[d]

Article accepted for publication in *BMC Infectious Diseases*

[a] Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, New Hampshire, USA 03824

[b] Dartmouth-Hitchcock Medical Center and Dartmouth College Geisel School of Medicine, Lebanon, New Hampshire, USA 03756

[c] Department of Chemical Engineering, University of New Hampshire, Durham, New Hampshire, USA 03824

[d] Department of Biological Sciences, University at Albany, State University of New York, Albany, New York, USA 12222

Abstract

Bloodstream infections due to *Staphylococcus aureus* cause significant patient morbidity and mortality worldwide. Of major concern is the emergence and spread of methicillin resistant *S. aureus* (MRSA) in bloodstream infections, which are associated with therapeutic failure and increased mortality. We generated high quality draft genomes from 323 *S. aureus* blood culture isolates from patients diagnosed with bloodstream infection at the Dartmouth-Hitchcock Medical

Center, New Hampshire, USA in 2010-2018. *In silico* detection of antimicrobial resistance genes revealed that 133/323 isolates (41.18%) carry horizontally acquired genes conferring resistance to at least three antimicrobial classes, with resistance determinants for aminoglycosides, beta-lactams and macrolides being the most prevalent. The most common resistance genes were *blaZ* and *mecA*, which were found in 262/323 (81.11%) and 104/323 (32.20%) isolates, respectively. Majority of the MRSA (102/105 isolates or 97.14%) identified using *in vitro* screening were related to two clonal complexes (CC) 5 and 8. The two CCs emerged in the New Hampshire population at separate times. We estimated that the time to the most recent common ancestor of CC5 was 1973 (95% highest posterior density (HPD) intervals: 1966 - 1979) and 1946 for CC8 (95% HPD intervals: 1924 - 1959). The effective population size of CC8 increased until the late 1960s when it started to level off until late 2000s. The levelling off of CC8 in 1968 coincided with the acquisition of SCC*mec* Type IV in majority of the strains. The plateau in CC8 also coincided with the acceleration in the population growth of CC5 carrying SCC*mec* Type II in the early 1970s and which eventually leveled off in the early 1990s. Lastly, we found evidence for frequent recombination in the two clones during their recent clonal expansion, which has likely contributed to their success in the population. We conclude that the *S. aureus* population was shaped mainly by the clonal expansion, recombination and co-dominance of two major MRSA clones in the last five decades. These results have important implications on the development of effective and robust strategies for intervention, control and treatment of life-threatening bloodstream infections.

Introduction

Bacteremia refers to the presence of viable bacteria in the blood (Christaki & Giamarellos-Bourboulis, 2014). Asymptomatic bacteremia can occur from normal daily activities, such as vigorous toothbrushing or from minor medical and dental procedures (Lockhart et al., 2008). Bacteremia is often transient and causes no further sequelae in healthy individuals (Christaki & Giamarellos-Bourboulis, 2014). However, when bacteria are able to breach innate host defenses and gain access to deeper tissues and subcutaneous sites, they can cause debilitating and life-threatening infections such as systemic inflammatory response syndrome, septic shock and organ failure (Thomer et al., 2016). Multiple bacterial species are implicated in bloodstream infections, of which *Staphylococcus aureus* cause 20.7% of total cases (Diekema et al., 2019). The incidence of bloodstream infections due to *S. aureus* is 20-50 per 100,000 population per year, with case-fatality rates of approximately 10–30% (van Hal et al., 2012). Bloodstream infection is notoriously hard to treat because it requires prompt source control and prolonged antimicrobial therapy (Thomer et al., 2016). Of major concern is the emergence and spread of multidrug resistant and methicillin resistant *S. aureus* (MRSA) in bloodstream infections, which are associated with therapeutic failure and increased mortality (Akova, 2016).

Understanding host and pathogen factors that impact bloodstream infections are crucial to disease control and treatment options. Host-related factors that determine outcome from bloodstream infections due to *S. aureus* have been widely studied and include age, comorbidities, immune status and primary focus of infection (i.e., the site of infection considered most likely to be responsible for seeding bacteria into the bloodstream) (Thomer et al., 2016; van Hal et al.,

2012). The most prominent risk factor for bloodstream infections due to *S. aureus* is the presence of prosthetic devices and other kinds of medical implantation of foreign bodies that allow the bacterium to invade the intravascular space (Thomer et al., 2016; van Hal et al., 2012). However, the contribution of the bacterium in bloodstream infections is less understood. Previous studies report that *S. aureus* in the blood often derive from colonizing isolates from other parts of the body acquired prior to bloodstream infection such as the nasopharynx and gastrointestinal tract (Benoit et al., 2018; Méric et al., 2018; Young et al., 2012). Different *S. aureus* clones in bloodstream infections may have adopted different strategies to overcome host responses such as cytolytic toxicity and biofilm formation (Recker et al., 2017). An important gap in current knowledge is that relatively little is known of the genetic basis that underlies the successful adaptation and persistence of certain *S. aureus* lineages, in particular MRSA, in the larger population over the long term.

We generated high quality draft genomes from 323 *S. aureus* blood culture isolates from patients diagnosed with bloodstream infection at the Dartmouth-Hitchcock Medical Center, New Hampshire, USA in 2010-2018. We conclude that the *S. aureus* population was shaped mainly by the clonal expansion, recombination and co-dominance of two major MRSA clones in the last five decades.

Methods

Ethics approval

Ethical approval was granted by the Committee for the Protection of Human Subjects of Dartmouth-Hitchcock Medical Center and Dartmouth College. This study protocol was deemed

not to be human subjects research. Samples used in the study were subcultured bacterial isolates that had been archived in the routine course of clinical laboratory operations. No patient specimens were used, and patient protected health information was not collected. Therefore, informed consent was not required.

Bacterial isolates

Archived *S. aureus* isolates from 325 unique patient blood stream infections were included in this study. These archived isolates grew from clinical blood culture specimens submitted to the Department of Pathology and Laboratory Medicine at the Dartmouth-Hitchcock Medical Center, New Hampshire, USA from December 2010 – August 2018. The first significant blood culture isolate from each patient is routinely archived (freezer space permitting) in case of future need for patient care, epidemiologic, public health or laboratory quality studies. Upon subculture, isolates were assigned a study number and all patient identifiers were removed with only the date of collection and results of clinical antimicrobial susceptibility testing linked to the study number. All isolates had previously been tested *in vitro* on an automated broth microdilution testing platform (MicroScan Walkaway, Beckman Coulter, Inc., La Brea, CA) against a panel of antimicrobial agents including cefoxitin (screening well), daptomycin, oxacillin, penicillin and vancomycin. Results were interpreted per Clinical Laboratory Standards Institute (CLSI) guidelines (CLSI, 2020). Methicillin resistance was determined per the manufacturer's package insert and CLSI guidelines by growth in the cefoxitin screening well (> 4 μg/mL) and/or by an oxacillin minimum inhibitory concentration (MIC) of > 2 μg/mL. Methicillin susceptible isolates were tested for beta-lactamase production using a nitrocefin disk (Remel, Lenexa, KS) using bacterial growth nearest to a 30 μg cefoxitin disk. Confirmatory

penicillin zone-edge testing was not performed on beta-lactamase-negative isolates. Clinical breakpoints of other antimicrobials tested were interpreted per CLSI guidelines (CLSI, 2020). All isolates were stored in DMSO solution in −80 °C.

DNA extraction and whole genome sequencing

Isolates were subcultured from glycerol stocks onto commercially prepared tryptic soy agar with 10% sheep red blood cells (Remel, Lenexa, KS) and in brain heart infusion broth (BD Difco, Franklin Lakes, NJ) at 37ºC for 24 hrs. DNA was extracted and purified from the liquid culture using the Zymo Research QuickDNA Fungal/Bacterial Miniprep Kit (Irvine, CA) following manufacturer's protocol. We used Qubit fluorometer (Invitrogen, Grand Island, NY) to measure DNA concentration. DNA libraries were prepared using the RipTide High Throughput Rapid DNA Library Prep kit (iGenomX, Carlsbad, CA). DNA samples were sequenced as multiplexed libraries on the Illumina HiSeq platform operated per the manufacturer's instructions. Sequencing resulted to 250-nt long paired end reads. Sequencing was carried out at the UNH Hubbard Center for Genome Studies, Durham, NH, USA.

De novo genome assembly, annotation and pan-genome construction

Reads were assembled into contigs using the Shovill v.1.1.0 pipeline (https://github.com/tseemann/shovill) with the option --trim to produce high quality draft genomes. Shovill is a series of methods that includes subsampling read depth down to 150X, trimming adapters, correcting sequencing errors and assembling using SPAdes (Bankevich et al., 2012). We used the scaffolding and gap-filling programs SSPACE (Boetzer et al., 2011) and GapFiller (Nadalin et al., 2012) to improve the assemblies. Genome quality was assessed using

the programs QUAST (Gurevich et al., 2013) and CheckM (Parks et al., 2015). Overall, genomes were at least 97.7% complete with no more than 2.86% contamination. Only genomes with < 200 contigs and N50 of > 40 kb were included. The number of contigs range from 16-115 and N50 of 77-1,011 kb (Table S2.1 and Figure S2.1). After filtering out the genomes with low coverage and of poor quality, a total of 323 genomes were used for downstream analyses. We calculated the genome-wide average nucleotide identity (ANI) for all possible pairs of genomes using the program FastANI v.1.0 (Jain et al., 2018) and compared them to the reference strain (strain NCTC 8325; RefSeq assembly accession no. GCF_000013425.1). The resulting contigs were annotated using Prokka v.1.14.6 (Seemann, 2014). We used Roary v.3.13.0 (Page et al., 2015) to characterize the totality of genes of all strains in a dataset or pan-genome (Medini et al., 2005).

In silico sequence typing and detection of antimicrobial resistance

The sequence type (ST) of each isolate was confirmed using the program MLST v.2.19.0 (https://github.com/tseemann/mlst), which extracts seven housekeeping genes (*arcC, aroE, glpF, gmk, pta, tpi* and *yqiL*) from the sequence contigs and compares sequence variation against previously characterized STs in the *S. aureus* PubMLST database (https://pubmlst.org/organisms/staphylococcus-aureus/). We used ARIBA v.2.14.5 (Hunt et al., 2017) and the ResFinder database (Zankari et al., 2012) to identify resistance determinants due to horizontally acquired antimicrobial resistance genes. We also used VirulenceFinder (Joensen et al., 2014) and the Virulence Factor Database (VFDB) (B. Liu et al., 2019) to screen all genomes for known virulence genes. We used SCCmecFinder v.1.2 (Kaya et al., 2018) to identify the presence and type of *mecA*-carrying staphylococcal chromosomal cassette (SCC*mec*) using minimum thresholds of > 60% for sequence coverage and > 90% sequence identity.

Phylogenetic and population structure analyses

Using the core genome alignment produced by Roary (Page et al., 2015), we generated a maximum likelihood phylogeny using IQ-Tree v1.6.12 (L.-T. Nguyen et al., 2015). We used ModelFinder to select the most appropriate nucleotide substitution model (Kalyaanamoorthy et al., 2017). We applied the general time reversible model of evolution (Tavaré, 1986) using nucleotide frequencies calculated directly from the sequence alignment and a FreeRate model of rate heterogeneity with three categories based on the ModelFinder results. We performed 1000 bootstrap replicates using ultrafast bootstrap approximation approach UFBoot2 (Hoang et al., 2018). We partitioned the strains on the core genome phylogeny into sequence clusters consisting of genetically similar individuals using the Bayesian hierarchical clustering algorithm RhierBAPS v1.1.2 (Tonkin-Hill et al., 2018).

Construction of time-calibrated phylogenies

We identified single nucleotide polymorphisms (SNPs) in the full core genome length alignment using Snippy v.4.6.0 (https://github.com/tseemann/snippy) and mapped them to a reference genome (strain NCTC 8325). Nucleotide sites on the core SNP alignment of each sequence cluster that were identified by Gubbins v.2.4.1 (Croucher et al., 2015) as having experienced recombination were removed to generate a phylogeny free of the distorting signal of recombination. To test for the presence of a temporal signal, we used the recombination-free phylogenies as input in the roottotip function in BactDating v1.0.12 (Didelot et al., 2018) to perform a root-to-tip linear regression analysis and random permutation of sampling dates to assess significance. We used BactDating to estimate the dates of the most recent common

ancestor. We carried out 10^8 iterations, removed the first half as burnin and the remainder was sampled every 100 iterations. Markov Chain Monte Carlo (MCMC) chain lengths were carried out until effective sample sizes of the inferred parameters α (coalescent time unit), μ (mean substitution rate) and σ (standard deviation of the per-branch substitution rates) parameters reached >200 as suggested by the authors. To estimate the effective population size (Ne) for the time-scaled phylogenies, we used the R package (R Core Team, 2013) Skygrowth v. 0.3.1 that implements a Bayesian MCMC and maximum a posteriori algorithms (Volz & Didelot, 2018). Skygrowth models the growth rate of Ne as a simple Brownian motion process.

Estimating recombination rates

Using the core genome alignment as input and 1000 bootstrapped replicates, we calculated the recombination rate using mcorr with default parameters (Lin & Kussell, 2019). Mcorr uses a coalescent-based model to measure the degree to which any two loci separated by N bp have correlated substitutions and estimates six evolutionary parameters: θ – mutational divergence; ϕ – recombinational divergence; c – recombination coverage or proportion of sites in the genome whose diversity was derived from outside the sample through recombination; d – diversity; mean fragment size ($\bar{f}$) of a recombination event; and θ/ϕ (or γ/μ) - relative rate of recombination to mutation (Lin & Kussell, 2019). We used Welch's t-test to compare the mcorr parameters between samples.

Availability of data and materials

The dataset supporting the conclusions of this article is included within the article and its additional files. Genome sequence data of the *S. aureus* blood culture isolates have been

deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under BioProject accession number PRJNA673382 with BioSample accession numbers for each genome listed in Table S2.1. The reference strain NCTC 8325 (RefSeq assembly accession no. GCF_000013425.1) is available in NCBI.

Unless otherwise noted, default parameters were used for all programs.

Results

Phylogenetic relationships and population structure

We obtained a total of 325 convenience *S. aureus* isolates recovered from unique patient blood cultures at the Dartmouth-Hitchcock Medical Center, USA which were sampled from December 2010 to August 2018. *In vitro* susceptibility testing revealed a total of 105 isolates (32.31%) that were MRSA. We observed the occurrence of MRSA every year, with prevalence ranging from 14.3% in 2014 to 48.8% in 2018 (**Figure 2.1a**). We retrieved high-quality draft genome sequences from 323 isolates (Table S2.1 and Figure S2.1). *De novo* genome assembly produced sequences ranging from 2.67 to 2.97 Mb in length. Annotation of assembled genomes revealed a total of 8,430 families of orthologous genes, which can be classified into 1,698 core genes (genes present in 320-323 genomes), 171 soft core genes (present in 307-319 genomes), 1,369 shell genes (present in 49-306 genomes) and 5,192 cloud genes (present in ≤ 48 genomes) (Table S2.2 and Figure S2.2). The number of genes per genome spanned from 2,440 to 2,814. The sizes of the pan- and core genomes were consistent with findings from previous genomic studies of *S. aureus* in clinical and non-clinical settings (Bosi et al., 2016; Manara et al., 2018).

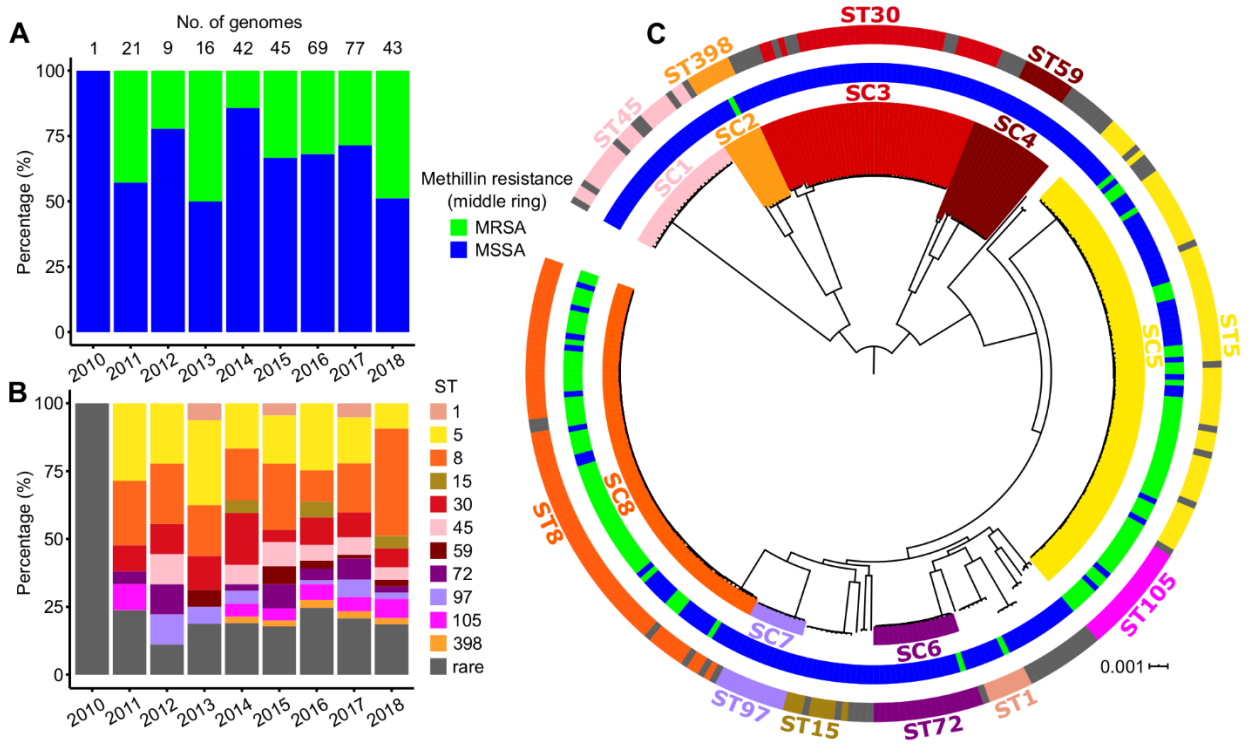

Figure 2.1 Methicillin resistance and phylogenetic relationships of *S. aureus* isolates. (a) Yearly sampling of methicillin resistance based on either cefoxitin or oxacillin phenotypic screenings. (b) Yearly distribution of sequence types (ST) throughout the study period. Gray bars indicate STs either previously unknown or rare (< 3 individuals). (c) Midpoint-rooted maximum likelihood tree showing phylogenetic structure of 323 *S. aureus* isolates. Clade colors on the branches represent sequence clusters defined by the BAPS program. Inner ring shows the phenotypic results of methicillin resistance screening as described above. Outer ring shows the distribution of different STs.

Multilocus sequence typing (MLST) revealed high *S. aureus* diversity throughout the duration of our study period, with at least five sequence types (STs) present every year and several rare and previously unidentified STs (**Figure 2.1b** and Table S2.1). Those STs that were consistently observed from 2011 to 2018 were STs 5, 8 and 30, while STs 1 and 15 were observed in alternating years. ST398 appeared first in 2014 and continued until 2018. Other STs such as STs 45, 59, 72, 97 and 105 were less common and appeared intermittently over eight years of our sampling period.

Population structure analysis using Bayesian hierarchical clustering of the core genome alignment showed eight distinct sequence clusters, ranging in size from 8 to 89 isolates (**Figure 2.1c**). Sequence clusters consists of genetically related strains of the same or closely related STs. The population was dominated by two large clusters. Sequence cluster 5 consisted mainly of STs 5 and 105 (which are members of clonal complex or CC5 (Challagundla, Reyes, et al., 2018)), while sequence cluster 8 consisted mostly of ST 8 (which is a member of CC8 (Bowers et al., 2018)). Majority of the MRSA identified using *in vitro* screening were present in these two clusters [50/89 (56.18%) in CC5 and 52/76 (68.42%) in CC8]. We also identified one MRSA isolate in each of sequence clusters 2 (ST398), 6 (ST72) and an undelineated cluster consisting of ST1. Isolates that did not group with the eight sequence clusters made up the low frequency genotypes that were divergent from the rest of the population and with each other.

Resistance to multiple antimicrobial classes

We considered *in silico* evidence to identify resistance determinants in each isolate (**Figure 2.2a** and Table S2.3). We detected a total of 20 horizontally transferrable resistance genes. Two mechanisms mediate penicillin resistance in *S. aureus*. The *blaZ* gene encodes a narrow-spectrum penicillinase, which inactivates the penicillinase-labile penicillins (penicillin, amoxicillin, ampicillin, carbenicillin, ticarcillin, azlocillin and piperacillin) by hydrolysis of its β-lactam ring (Fuda et al., 2005). The *mecA* gene encodes an alternative penicillin-binding protein PBP2a and confers resistance to all beta-lactam antimicrobial agents with the exception of ceftaroline (Hartman & Tomasz, 1984). We detected *blaZ* and *mecA* in 262/323 isolates (81.11%) and 104/323 isolates (32.20%), respectively. We found a slight discrepancy between

the *in silico* detection of the *mecA* gene and *in vitro* phenotypic testing for methicillin resistance. There were two isolates whose genomes contained the *mecA* gene but phenotypically tested as methicillin susceptible. There were three isolates whose genomes did not contain the *mecA* gene but that phenotypically tested as MRSA. These phenotype/genotype discrepancies may be due to sequencing errors, the existence of novel *mecA* homologs that are not recognized as yet in current databases or the presence of other yet unknown determinants of methicillin resistance. The *mecA* gene is carried by the mobile element staphylococcal cassette chromosome (SCC*mec*), of which there are currently 14 recognized types (International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC), 2009). In our study, SCC*mec* type II is associated mainly with CC5 (43/89 or 48.31% of total isolates in the cluster), while type IV with CC 8 (52/76 or 68.42% of total isolates in the cluster). Each of the two CCs also contains isolates with SCC*mec* type from the other CC (i.e., five isolates in CC5 have SCC*mec* type IV and one isolate in CC8 have SCC*mec* type II). We also detected SCC*mec* type V in sequence cluster 6 (ST72). The three SCC*mec* types vary in terms of their gene content (J. Liu et al., 2016). SCC*mec* type II is a relatively large element (~52 kb) carrying a type 2 recombinase-encoding *ccr* gene complex and is commonly found among healthcare-acquired MRSA strains, including CC5 (J. Liu et al., 2016; Monecke et al., 2011). It carries transposon Tn554, which contains the antimicrobial resistance genes *ermA* (erythromycin resistance) and *spc* (spectinomycin resistance). SCC*mec* Type IV is the smallest type (~24 kb), often circulating in community settings (J. Liu et al., 2016). Often, this element contains transposon Tn4001, carrying genes conferring resistance to aminoglycosides (J. Liu et al., 2016). Like type IV, SCC*mec* type V is smaller in length (~27 kb) and is often found in community-associated MRSA

(J. Liu et al., 2016). It carries other resistance genes aside from *mecA*, but has a restriction modification system that may play a role in stabilizing the SCC*mec* (J. Liu et al., 2016).

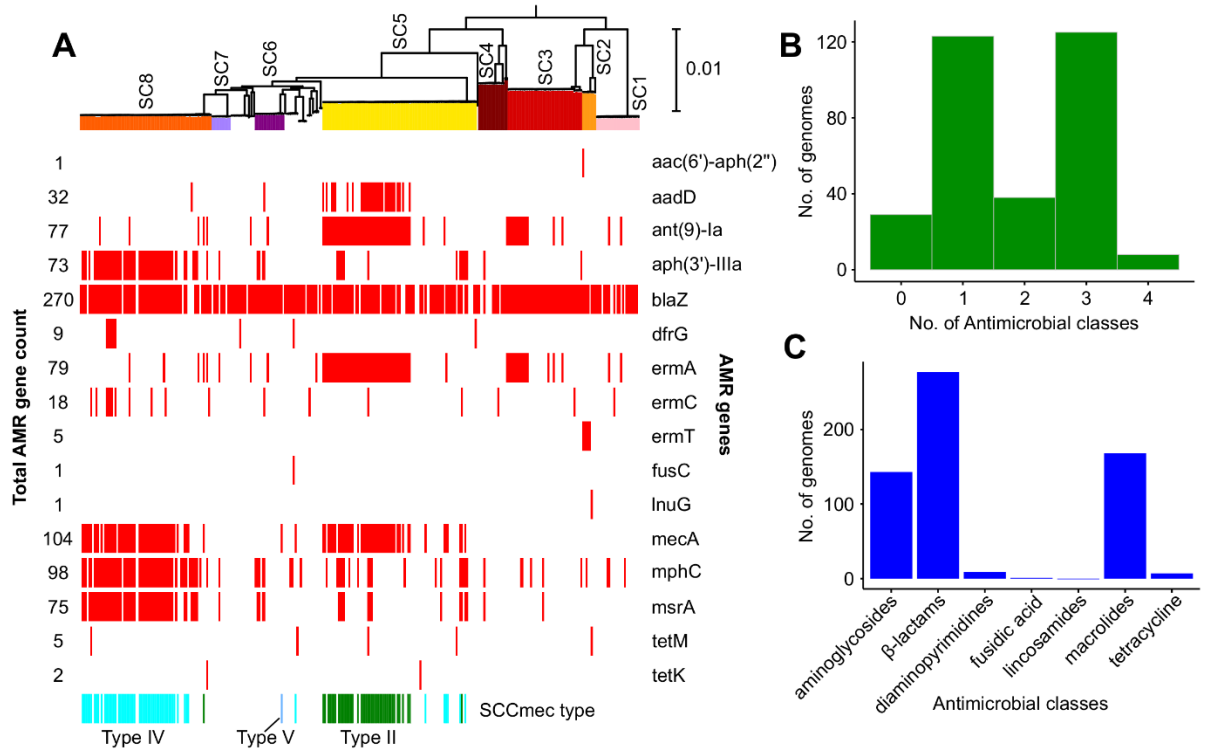

Figure 2.2 Antimicrobial resistance gene and class distribution among 323 isolates. (a) Gene presence-absence matrix showing the distribution of antimicrobial resistance genes and SCC*mec* types across the phylogeny (tree is identical to that in Figure 2.1c). Red blocks indicate presence of gene listed to the right of the panel. SCC*mec* types are color-coded as green = Type II, aqua = Type IV, blue = Type V. (b) Histogram showing distribution of the number of antimicrobial classes each genome is resistant to. (c) Bar chart showing the number of genomes resistant to each individual antimicrobial class.

We identified other horizontally acquired resistance genes (Haaber et al., 2017) from the genome sequences. These were differentially distributed across the population and in some cases were associated mainly with specific sequence clusters. For example, the genes *aadD* (kanamycin resistance), *ant(9)-Ia* (aminoglycoside resistance) and *ermA* (erythromycin resistance) were mainly found in CC 5 [30/89, 53/89 and 52/89 genomes, respectively]. The

genes *aph(3')-IIIa* (aminoglycoside resistance), *mphC* (macrolide resistance) and *msrA* (macrolide and streptogramin B resistance) were most frequently observed in isolates from CC 8 (54/76, 56/76 and 55/76 genomes, respectively) (Haaber et al., 2017). Although the prevalence of vancomycin resistance had been reported to be increasing in recent years (Shariati et al., 2020), we did not detect vancomycin-resistant isolates using either phenotypic assays or *in silico* methods. Overall, we found that 133/323 isolates or 41.18% carry horizontally acquired genes conferring resistance to at least three antimicrobial classes (**Figure 2.2b**), with resistance determinants for aminoglycosides (143/323), beta-lactams (277/323) and macrolides (168/323) being the most prevalent (**Figure 2.2c**).

Virulence characteristics of S. aureus genomes

We used *in silico* screening of virulence genes that may play critical role in bloodstream infections (Figure S2.3 and Table S2.4). Several of these were detected among different sequence clusters. The *hlgABC* pore-forming leucotoxin genes were found in every genome, with the exception of two genomes missing the *hlgC* gene. We found the *sak* gene, which encodes a plasminogen activator protein staphylokinase (L. T. Nguyen & Vogel, 2016), in 297/323 genomes (or 91.95%). The gene product of *sak* has two known roles in virulence. First, it forms a complex with plasmin to prevent biofilm formation in the blood, allowing for access to deeper tissues (L. T. Nguyen & Vogel, 2016). It also prevents antimicrobial peptides excreted by the host to effectively target bacterial cells (L. T. Nguyen & Vogel, 2016). Found in 309/323 genomes (or 95.67%), the virulence factor staphylococcal complement inhibitor encoded by the *scn* gene allows *S. aureus* to evade phagocytosis and killing by human neutrophils (Rooijakkers et al., 2005). The metalloproteinase-encoding aureolysin gene (*aur*) was found in 322/323

genomes (99.69%). Aureolysin has been implicated in the inactivation of host protease inhibitors and degradation of host antimicrobial peptides (Pietrocola et al., 2017). Two allelic variants of *aur* have been previously reported (Sabat et al., 2000) and we found both variants in our dataset: *aur* in 246/323 genomes (76.16%) and *aur1* in 76/323 genomes (23.53%). We detected the *aur* variant in sequence clusters 4, 5, 6, 7, 8 and the unclustered strains, while *aur1* was found in sequence clusters 1, 2 and 3. There appears to be little functional difference between the two *aur* variants and it remains unclear why they continue to persist in distinct *S. aureus* lineages (Sabat et al., 2000).

We found differences in the distribution of virulence genes between CC5 and CC8. For example, CC5 has a much larger number of enterotoxin genes compared to CC8. Enterotoxins are proteins secreted by primarily by *Staphylococcus* spp. that target intestinal tissues and are commonly associated with food poisoning and toxic shock syndrome (Pinchuk et al., 2010). Among these enterotoxin-encoding genes, we detected *seg, sei, sem, sen, seo* and *seu* in every CC5 genome with very few found in CC8 genomes. Other genes such as *sed, sej* and *ser* were also found more frequently in CC5 (46/89, 49/89 and 48/89 genomes, respectively). In contrast, the enterotoxin genes *sek* and *seq* were more common in CC8 (48/76 genomes for both). The gene *splE*, which encodes a serine protease (Paharik et al., 2016), was also common in CC8 (71/76 genomes). The Panton-Valentine leukocidin (PVL) genes *lukF* and *lukS* were frequently associated with CC8 (48/76 and 47/76 genomes, respectively). The prophage-encoded PVL produces a pore-forming toxin capable of killing neutrophils, monocytes and macrophages and is closely linked to severe *S. aureus* infections (Spaan et al., 2013). We also detected the arginine

catabolic mobile element genomic island that can enhance the growth and survival of *S. aureus* (Diep et al., 2008) in 47/76 genomes in CC8.

Among the other sequence clusters, one notable finding is the presence of the gene that encodes the superantigen toxic-shock syndrome toxin-1 (TSST-1) among 39/43 or 90.70% of the genomes in ST30. TSST-1 activates T-cells, which results to the induction of a cytokine storm and can become fatal (Sharma et al., 2019).

Estimating the date of clonal origins and effective population size

We sought to determine the dates of clonal origin of CCs 5 and 8, which make up majority of the MRSA genomes. Using the recombination-free alignment of core SNPs as input to BactDating (Didelot et al., 2018), we first determined the presence of a temporal signal in each CC (Figure S2.4). We observed a significant positive correlation between the dates of isolation and root-to-tip distances (CC 5: $R^2 = 0.06$ and p value = 1.4 e-2 and CC 8: $R^2 = 0.15$ and p value = 3.0 e-4) indicating the presence of a clock-like signal. Results showed that the two CCs emerged in the New Hampshire population at separate times. We estimated that the time to the most recent common ancestor of CC5 was around 1973 (95% highest posterior density (HPD) intervals: 1966 - 1979) and around 1946 for CC8 (95% HPD intervals: 1924 - 1959) (**Figure 2.3ab**). For each CC, we also estimated the change in the effective population size (Ne), which is a measure of the rate of change in population composition due to genetic drift (Nei & Tajima, 1981). The effective population size of CC8 increased until the late 1960s when it started to level off until late 2000s. The onset of the levelling off of CC8 in 1968 coincided with the acquisition of SCC*mec* Type IV in majority of the strains. The plateau in CC8 also coincided

with the acceleration in the population growth of CC5 in the early 1970s and which eventually leveled off in the early 1990s. The beginning of the levelling off of CC5 in 1993 overlaps with the acquisition of SCC*mec* Type II (**Figure 2.3cd**). These results indicate that the emergence of lineages each carrying distinct SCC*mec* types contributed to the success of the two MRSA CCs in the local population.

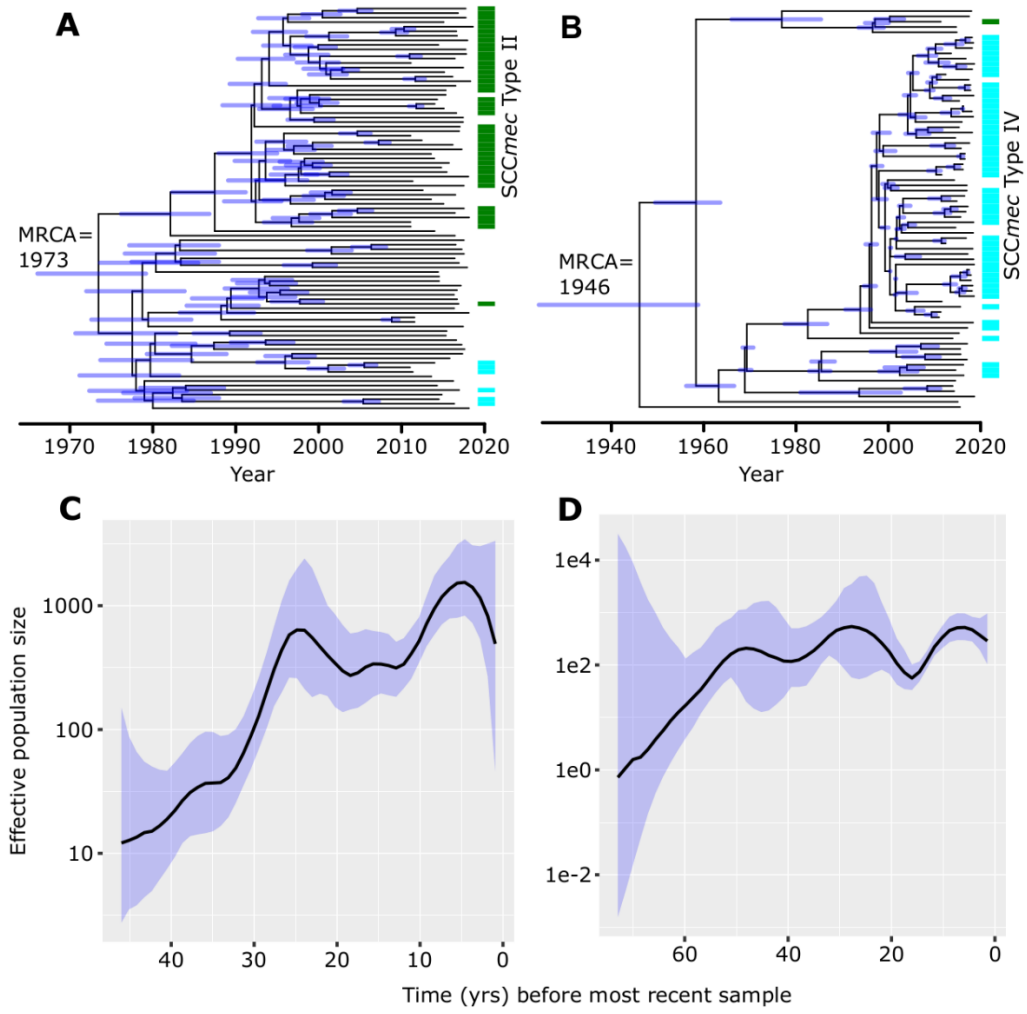

Figure 2.3 Bayesian phylogeny and population dynamics of CC5 and CC8. Bayesian maximum clade credibility time-calibrated phylogeny of CC5 (a) and CC8 (b) based on non-recombining regions of the core genome. Divergence date (median estimate with 95% highest posterior density dates in brackets) is indicated in blue on the tree. Bayesian skygrowth plot showing changes in effective population size (Ne) over time of CC5 (c) and CC8 (d). Median is represented by a black line and 95% confidence intervals are in blue.

Variation in recombination rates

Homologous recombination, whereby similar tracts of DNA are acquired from closely related strains (Didelot & Maiden, 2010; Hanage, 2016), is known to impact the evolution and adaptation of *S. aureus* to its hosts (Castillo-Ramírez et al., 2012; Driebe et al., 2015; Spoor et al., 2015). The genetic diversity generated through recombination may be further disseminated in the population via subsequent recombination events with other strains or may be passed on to descendants through clonal descent. Here, we sought to determine the extent in which recombination has affected the genome evolution of the two dominant MRSA clones CCs 5 and 8. Using the core genome alignment as input and 1000 bootstrapped replicates, we used mcorr to calculate six recombination parameters (**Figure 2.4** and Table S2.5) (Lin & Kussell, 2019). The diversity (d) is the probability that a pair of genomes will differ at any locus and is estimated from the diversity generated from both recombination and accumulation of mutations of the clonal lineage. This parameter was estimated to be 7.63×10^{-4} and 5.51×10^{-4} in CC5 and CC8, respectively (p value = 0.0; Welch's t test). The mutational divergence (θ), which refers to the mean number of mutations per locus since divergence of a pair of homologous sites, was estimated to be 0.020 for both CC5 and CC8 (p value = 0.120; Welch's t test). Recombinational divergence (ϕ) was estimated to be 0.073 and 0.041 in CC5 and CC8, respectively (p value = 4.19×10^{-58}; Welch's t test). The ratio ϕ/θ (or γ/μ), which gives the relative rate of recombination to mutation, was estimated to be 3.63 and 2.06 in CC5 and CC8, respectively (p value = 1.28×10^{-53}; Welch's t test). The mean fragment size ($\bar{f}$) of a recombination event was estimated to be 407.45 bp and 530.55 bp in CC5 and CC8, respectively (p value = 4.30×10^{-107}; Welch's t test). Lastly, we calculated the recombination coverage c, which indicates the fraction of the genome whose diversity was derived from recombination events since the last common ancestor of the

population, and which ranges from 0 (indicating clonal evolution) to 1 (indicating complete recombination) (Lin & Kussell, 2019). Recombination coverage was estimated to be 0.035 and 0.026 in CC5 and CC8, respectively (p value = 1.10×10^{-231}; Welch's t test), indicating that 3.5% and 2.6% of sites in any one genome from CC5 and CC8, respectively, originated from recombination events. Except for the mutational divergence (θ), we found significant differences in five recombination parameters between the two clonal complexes. We compared these values to those calculated by the authors of mcorr for a *S. aureus* genomic dataset of 308 invasive isolates from 26 countries in Europe and representing 10 CCs and six minor STs (Aanensen et al., 2016). Both CC5 and CC8 in the New Hampshire population have lower diversity (0.015 in Europe), mutational divergence (0.042 in Europe) and coverage (0.36 in Europe) than the European *S. aureus*, but higher γ/μ (1.0 in Europe). These were expected because the European dataset consisted of more diverse genotypes. Remarkably, the recombinational divergence and mean recombination size of the European *S. aureus* were 0.042 and 550, respectively, which were similar to the values of the New Hampshire CC8. CC5 had higher recombinational divergence and lower mean recombination size than the European *S. aureus*. Overall, these results indicate that the two MRSA clones in New Hampshire have experienced frequent recombination during their recent clonal expansion.

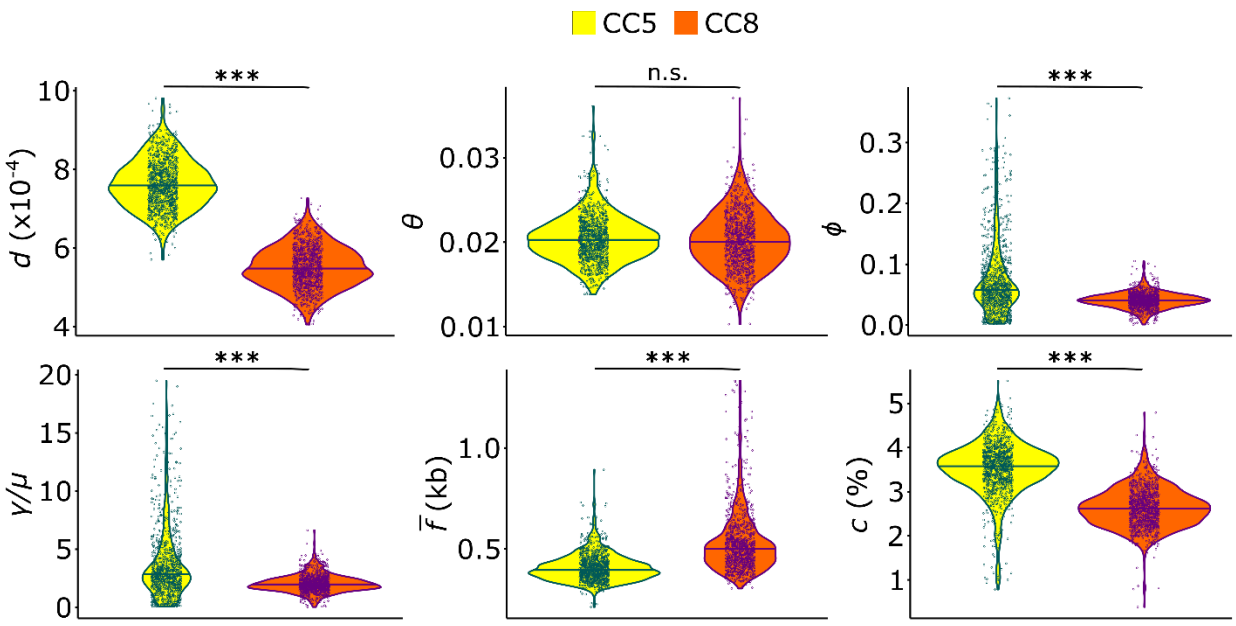

Figure 2.4 Recombination parameters comparing CC5 and CC8. Core genome alignment of each CC was used as input in mcorr with 1000 bootstrapped replicates. d – diversity brought into the population by recombination and clonal diversity; θ – mutational divergence; ϕ – recombinational divergence; γ/μ relative rate of recombination to mutation (equivalent to ratio of ϕ/θ); $\bar{f}$ – mean fragment size of a recombination event; c – recombination coverage. For all panels, brackets indicate results of Welch's t-test of group comparisons; n.s. - not significant; *** - $p < 0.0001$.

Discussion

Elucidating the evolutionary context that has selected certain methicillin resistant *S. aureus* genotypes is critical to developing effective surveillance, control strategies and treatment options for bloodstream infections. Our rationale is that, once we understand the drivers that underlie the population and genome dynamics of bacteria causing bloodstream infections, it would be possible to minimize them from two different facets of the biology of bloodstream infections: the bacterial clones that are identified as having high risk for invasion of the blood or the chromosomal regions of the bacterium that can be manipulated or targeted through therapeutic options. We highlight both aspects in this study using genome sequences of 323 *S. aureus* blood culture isolates spanning at least eight years of sampling.

Two mechanisms explain the spread of drug resistance in bacterial populations: clonal expansion of resistant lineages (Challagundla, Luo, et al., 2018; Gomez-Simmonds et al., 2015; Lam et al., 2018; Ward et al., 2014) and horizontal gene transfer of resistance genes (Dolejska & Papagiannitsis, 2018; Evans et al., 2020; Lerminiaux & Cameron, 2019). Our study showed that both processes have shaped the long-term population dynamics of MRSA in New Hampshire. Our first major finding is that both CC5 and CC8 (each carrying different suites of resistance genes) have undergone clonal expansion in the last 50 years, which would partly explain the broad occurrence of multidrug and methicillin resistance in the *S. aureus* population. The clones that comprise these two lineages are the most prominent MRSA lineages worldwide and have been implicated in numerous healthcare-associated and community-acquired infections and outbreaks (Monecke et al., 2011). CC8, which includes ST8, consists of several pandemic MRSA clones associated with healthcare- and community-associated infections (Bowers et al., 2018; Monecke et al., 2011). ST8 includes the notorious pandemic clone USA300 that emerged in Central Europe in the mid-19th century and was exported to North America and elsewhere in the early 20th century (Strauß et al., 2017). Community and household transmission facilitated the widespread distribution of this clone in the United States (Alam et al., 2015; Jamrozy et al., 2016; Uhlemann et al., 2014). ST8 also includes USA500 that emerged in the past 20 years and causes invasive infections in North America (Frisch et al., 2018). On the other hand, CC5 includes multiple MRSA clones that cause hospital-associated infections in the Western Hemisphere (Challagundla, Reyes, et al., 2018). It is also the principal genetic background in *S. aureus* upon which full resistance to vancomycin, a key last-line bactericidal drug for treating MRSA infections, has arisen by acquisition of the Tn1546 carrying the *van* operon from

Enterococcus donors (Kos et al., 2012). Although none of the New Hampshire isolates were vancomycin-resistant, reduced vancomycin susceptibility has been increasing in *S. aureus* causing bloodstream infections in other parts of the United States (Canty et al., 2020; Shariati et al., 2020) and it may therefore just be a matter of time before it emerges in New Hampshire.

The predominance of CCs 5 and 8 in bloodstream infections has been previously reported in other places. In Minnesota USA, the most common MRSA clones were CC5 (ST5, all SCC*mec* type II) and clonal complex 8 (ST8 and ST3342, mostly SCC*mec* type IV) (Park et al., 2017). In a Colorado USA study that tracked *S. aureus* colonizing the nares and progressing to bacteremia within individual patients, STs 5 and 8 were also the most frequently observed (Benoit et al., 2018). In a study of MRSA isolates causing bloodstream infections in England sampled from 2012-2013, ST22 (clonal complex 22) make up the majority of *S. aureus* isolates, while STs 5 and 8 were less common (Toleman et al., 2019). In nine Latin American countries sampled in 2011 – 2014, the population of *S. aureus* from bloodstream infections was dominated by STs 5 and 8 as well as minor STs such as STs 30 and 72 (Arias et al., 2017). In contrast to our results, the Latin American study reported evidence for frequent clonal replacement of previously prevalent hospital-associated clones: ST8 replaced ST5 in Colombia and Ecuador, while ST5 replaced ST239 in Brazil (Arias et al., 2017). In our New Hampshire study, co-dominance rather than clonal replacement enabled both methicillin resistant CC5 (ST5) and CC8 (ST8) to persist in the last five decades. We can only postulate the reasons for such remarkable contrast in *S. aureus* population dynamics – clonal replacement in South America versus clonal co-dominance in North America – in a relatively short time span. Selective pressures due to ecological differences may explain this variation. Differences in the population dynamics of

resistant bacteria could be due to regional differences in host immunity that selected for specific genes that happened to be resistant, differences in antimicrobial use or different patterns of transmission among specific host groups that receive more antimicrobial prescriptions such as children (Andam et al., 2017). The co-dominance of two high-risk clones with distinct resistance and virulence characteristics is an important consideration in the epidemiology of *S. aureus* causing bloodstream infections in New Hampshire.

Our second major finding is that the second mechanism, i.e., horizontal acquisition of resistance genes, was also a major driver in the spread of resistance in *S. aureus* causing bloodstream infections. Although the mobile *mecA* was primarily found in CC5 and CC8, numerous horizontally-acquired resistance genes were also differentially distributed in other lineages. Two possible scenarios can happen. These rare lineages (i.e., isolates that were not part of CCs 5 and 8) can act as reservoir of resistance genes and their allelic variants from which the dominant lineages can draw from (Monecke et al., 2011; Nimmo et al., 2015; Planet et al., 2013). It may also be possible that the presence of these resistance genes can facilitate the clonal expansion and dissemination of the rare genotypes in the future (Copin et al., 2019; Packer et al., 2019). In either case, these rare genotypes carrying horizontally mobile resistance genes must be observed over the long term to ensure that risks due to either scenario are reduced. Moreover, both CC5 and CC8 appear to have had a history of frequent recombination during their recent clonal expansion. Frequent or large-scale genetic recombination events can significantly alter the genome structure of a strain, which can drive the emergence of lineages with unique traits such as ability to switch between animal and human hosts, epidemic features or enhanced growth in

new ecological niches (Nimmo et al., 2015; Richardson et al., 2018; Robinson & Enright, 2004; Spoor et al., 2015).

We acknowledge several limitations of our study. First, only one isolate was obtained per patient. Multiple infections (either repeated or mixed infections) and the diversity of the *S. aureus* inoculum at the site of infection can greatly influence the *S. aureus* population within a single patient as the pathogen evolves during the course of infection (Didelot et al., 2016). Second, characterization of mobile elements, resistance genes and virulence genes depend greatly on the composition and quality of existing databases used in *in silico* methods. This means that novel genetic variants (including unknown types of SCC*mec*) may not be recognized by current methods. Lastly, as in any phylogenetic and population demographic method, estimation of the dates of the last common ancestor is made difficult with increasing time between the ancestor and the observed descendants as well as the range of diversity of missing descendants. Hence, the dates of the last common ancestors of CC5 and CC8 represent only an approximation and may vary with the inclusion of additional isolates.

The results presented here open multiple avenues for future research. Continuous surveillance over many years is essential for documenting the long-term dynamics and drug resistance of *S. aureus*, including the potential for clonal replacement by other, less common lineages. Low frequency clones, which may harbor unique genomic elements or phenotypic features, may remain hidden from epidemiological and phylogenetic studies. They also have the capacity to replace one or both co-dominant lineages (CCs 5 and 8), which may be facilitated by changes in ecological conditions (e.g., change in antimicrobial use and demographic changes in

the region). Future surveillance strategies will also be enhanced by the inclusion of other hospitals in the region. It is likely that a more heterogenous assembly of *S. aureus* lineages are present in New Hampshire, but the lack of a systematic state- or region-wide *S. aureus* surveillance in bloodstream infections may lead to a subpopulation of genetic variants that we may have overlooked in the present analyses.

Second, future work should also focus on the relationship of clone-specific genomic features and the clinical outcomes of patients with bloodstream infection, which can be used as predictive factors in *S. aureus* mortality, clinical manifestation and disease severity. Once *S. aureus* enters the bloodstream, it replicates and disseminates to many different sites, causing severe disease manifestations such as sepsis, infective endocarditis, and deep-seated abscesses in virtually every organ tissue (Thomer et al., 2016). It remains unclear if there is site-specific selection occurring for specific strains or genetic variants. This information will be particularly useful to precisely characterize protective vaccine antigens and the development of immune therapeutics to prevent disease or improve clinical outcomes. It will also inform appropriate patient management and therapeutics for persistent, relapse, metastatic or complicated *S. aureus* bloodstream infections. Unfortunately, our data set does not include an extensive amount of clinical, phenotypic or other epidemiological information for each isolate. Hence, we strongly advocate for the inclusion of such pertinent associated data in sampling and surveillance schemes of bloodstream infections in the state of New Hampshire and elsewhere.

We conclude that the *S. aureus* population causing bloodstream infections was shaped mainly by the clonal expansion, recombination and long-term co-dominance of two high-risk

lineages with distinct genomic features, resistance characteristics and evolutionary histories. These results have important implications on the development of effective and robust approaches to identifying new bacterial targets for intervention, control and treatment strategies of life-threatening bloodstream infections.

References

Aanensen, D. M., Feil, E. J., Holden, M. T. G., Dordel, J., Yeats, C. A., Fedosejev, A., Goater, R., Castillo-Ramírez, S., Corander, J., Colijn, C., Chlebowicz, M. A., Schouls, L., Heck, M., Pluister, G., Ruimy, R., Kahlmeter, G., Åhman, J., Matuschek, E., Friedrich, A. W., … European SRL Working Group. (2016). Whole-genome sequencing for routine pathogen surveillance in public health: A population snapshot of invasive *Staphylococcus aureus* in Europe. *MBio*, *7*(3). https://doi.org/10.1128/mBio.00444-16

Akova, M. (2016). Epidemiology of antimicrobial resistance in bloodstream infections. *Virulence*, *7*(3), 252–266. https://doi.org/10.1080/21505594.2016.1159366

Alam, M. T., Read, T. D., Petit, R. A., Boyle-Vavra, S., Miller, L. G., Eells, S. J., Daum, R. S., & David, M. Z. (2015). Transmission and microevolution of USA300 MRSA in U.S. households: Evidence from whole-genome sequencing. *MBio*, *6*(2), e00054. https://doi.org/10.1128/mBio.00054-15

Andam, C. P., Worby, C. J., Gierke, R., McGee, L., Pilishvili, T., & Hanage, W. P. (2017). Penicillin resistance of nonvaccine type pneumococcus before and after PCV13 introduction, United States. *Emerging Infectious Diseases*, *23*(6), 1012–1015. https://doi.org/10.3201/eid2306.161331

Arias, C. A., Reyes, J., Carvajal, L. P., Rincon, S., Diaz, L., Panesso, D., Ibarra, G., Rios, R., Munita, J. M., Salles, M. J., Alvarez-Moreno, C., Labarca, J., Garcia, C., Luna, C. M., Mejia-Villatoro, C., Zurita, J., Guzman-Blanco, M., Rodriguez-Noriega, E., Narechania, A., … Seas, C. (2017). A prospective cohort multicenter study of molecular epidemiology and phylogenomics of *Staphylococcus aureus* bacteremia in nine Latin american countries. *Antimicrobial Agents and Chemotherapy*, *61*(10). https://doi.org/10.1128/AAC.00816-17

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Benoit, J. B., Frank, D. N., & Bessesen, M. T. (2018). Genomic evolution of *Staphylococcus aureus* isolates colonizing the nares and progressing to bacteremia. *PloS One*, *13*(5), e0195860. https://doi.org/10.1371/journal.pone.0195860

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., & Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics (Oxford, England)*, *27*(4), 578–579. https://doi.org/10.1093/bioinformatics/btq683

Bosi, E., Monk, J. M., Aziz, R. K., Fondi, M., Nizet, V., & Palsson, B. Ø. (2016). Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(26), E3801-3809. https://doi.org/10.1073/pnas.1523199113

Bowers, J. R., Driebe, E. M., Albrecht, V., McDougal, L. K., Granade, M., Roe, C. C., Lemmer, D., Rasheed, J. K., Engelthaler, D. M., Keim, P., & Limbago, B. M. (2018). Improved subtyping of *Staphylococcus aureus* clonal complex 8 strains based on whole-genome phylogenetic analysis. *MSphere*, *3*(3). https://doi.org/10.1128/mSphere.00464-17

Canty, E., Carnahan, B., Curley, T., Anususinha, E., Hamdy, R. F., & Ericson, J. E. (2020). Reduced vancomycin susceptibility, MRSA and treatment failure in pediatric *Staphylococcus aureus* bloodstream infections. *The Pediatric Infectious Disease Journal*. https://doi.org/10.1097/INF.0000000000002992

Castillo-Ramírez, S., Corander, J., Marttinen, P., Aldeljawi, M., Hanage, W. P., Westh, H., Boye, K., Gulay, Z., Bentley, S. D., Parkhill, J., Holden, M. T., & Feil, E. J. (2012). Phylogeographic variation in recombination rates within a global clone of methicillin-resistant *Staphylococcus aureus*. *Genome Biology*, *13*(12), R126. https://doi.org/10.1186/gb-2012-13-12-r126

Challagundla, L., Luo, X., Tickler, I. A., Didelot, X., Coleman, D. C., Shore, A. C., Coombs, G. W., Sordelli, D. O., Brown, E. L., Skov, R., Larsen, A. R., Reyes, J., Robledo, I. E., Vazquez, G. J., Rivera, R., Fey, P. D., Stevenson, K., Wang, S.-H., Kreiswirth, B. N., … Robinson, D. A. (2018). Range expansion and the origin of USA300 North American epidemic methicillin-resistant *Staphylococcus aureus*. *MBio*, *9*(1). https://doi.org/10.1128/mBio.02016-17

Challagundla, L., Reyes, J., Rafiqullah, I., Sordelli, D. O., Echaniz-Aviles, G., Velazquez-Meza, M. E., Castillo-Ramírez, S., Fittipaldi, N., Feldgarden, M., Chapman, S. B., Calderwood, M. S., Carvajal, L. P., Rincon, S., Hanson, B., Planet, P. J., Arias, C. A., Diaz, L., & Robinson, D. A. (2018). Phylogenomic classification and the evolution of clonal complex 5 methicillin-resistant *Staphylococcus aureus* in the western hemisphere. *Frontiers in Microbiology*, *9*, 1901. https://doi.org/10.3389/fmicb.2018.01901

Christaki, E., & Giamarellos-Bourboulis, E. J. (2014). The complex pathogenesis of bacteremia: From antimicrobial clearance mechanisms to the genetic background of the host. *Virulence*, *5*(1), 57–65. https://doi.org/10.4161/viru.26514

CLSI. (2020). *Performance Standards for Antimicrobial Susceptibility Testing. CLSI supplement M100* (30th ed.). Clinical and Laboratory Standards Institute.

Copin, R., Sause, W. E., Fulmer, Y., Balasubramanian, D., Dyzenhaus, S., Ahmed, J. M., Kumar, K., Lees, J., Stachel, A., Fisher, J. C., Drlica, K., Phillips, M., Weiser, J. N., Planet, P. J., Uhlemann, A.-C., Altman, D. R., Sebra, R., van Bakel, H., Lighter, J., …

Shopsin, B. (2019). Sequential evolution of virulence and resistance during clonal spread of community-acquired methicillin-resistant *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(5), 1745–1754. https://doi.org/10.1073/pnas.1814265116

Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., Parkhill, J., & Harris, S. R. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Research*, *43*(3), e15. https://doi.org/10.1093/nar/gku1196

Didelot, X., Croucher, N. J., Bentley, S. D., Harris, S. R., & Wilson, D. J. (2018). Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Research*, *46*(22), e134. https://doi.org/10.1093/nar/gky783

Didelot, X., & Maiden, M. C. J. (2010). Impact of recombination on bacterial evolution. *Trends in Microbiology*, *18*(7), 315–322. https://doi.org/10.1016/j.tim.2010.04.002

Didelot, X., Walker, A. S., Peto, T. E., Crook, D. W., & Wilson, D. J. (2016). Within-host evolution of bacterial pathogens. *Nature Reviews. Microbiology*, *14*(3), 150–162. https://doi.org/10.1038/nrmicro.2015.13

Diekema, D. J., Hsueh, P.-R., Mendes, R. E., Pfaller, M. A., Rolston, K. V., Sader, H. S., & Jones, R. N. (2019). The microbiology of bloodstream infection: 20-year trends from the SENTRY antimicrobial surveillance program. *Antimicrobial Agents and Chemotherapy*, *63*(7). https://doi.org/10.1128/AAC.00355-19

Diep, B. A., Stone, G. G., Basuino, L., Graber, C. J., Miller, A., des Etages, S.-A., Jones, A., Palazzolo-Ballance, A. M., Perdreau-Remington, F., Sensabaugh, G. F., DeLeo, F. R., & Chambers, H. F. (2008). The arginine catabolic mobile element and staphylococcal chromosomal cassette mec linkage: Convergence of virulence and resistance in the USA300 clone of methicillin-resistant *Staphylococcus aureus*. *The Journal of Infectious Diseases*, *197*(11), 1523–1530. https://doi.org/10.1086/587907

Dolejska, M., & Papagiannitsis, C. C. (2018). Plasmid-mediated resistance is going wild. *Plasmid*, *99*, 99–111. https://doi.org/10.1016/j.plasmid.2018.09.010

Driebe, E. M., Sahl, J. W., Roe, C., Bowers, J. R., Schupp, J. M., Gillece, J. D., Kelley, E., Price, L. B., Pearson, T. R., Hepp, C. M., Brzoska, P. M., Cummings, C. A., Furtado, M. R., Andersen, P. S., Stegger, M., Engelthaler, D. M., & Keim, P. S. (2015). Using whole genome analysis to examine recombination across diverse sequence types of *Staphylococcus aureus*. *PloS One*, *10*(7), e0130955. https://doi.org/10.1371/journal.pone.0130955

Evans, D. R., Griffith, M. P., Sundermann, A. J., Shutt, K. A., Saul, M. I., Mustapha, M. M., Marsh, J. W., Cooper, V. S., Harrison, L. H., & Van Tyne, D. (2020). Systematic

detection of horizontal gene transfer across genera among multidrug-resistant bacteria in a single hospital. *ELife*, *9*. https://doi.org/10.7554/eLife.53886

Frisch, M. B., Castillo-Ramírez, S., Petit, R. A., Farley, M. M., Ray, S. M., Albrecht, V. S., Limbago, B. M., Hernandez, J., See, I., Satola, S. W., & Read, T. D. (2018). Invasive methicillin-resistant *Staphylococcus aureus* USA500 strains from the U.S. emerging infections program constitute three geographically distinct lineages. *MSphere*, *3*(3). https://doi.org/10.1128/mSphere.00571-17

Fuda, C. C. S., Fisher, J. F., & Mobashery, S. (2005). Beta-lactam resistance in *Staphylococcus aureus*: The adaptive resistance of a plastic genome. *Cellular and Molecular Life Sciences: CMLS*, *62*(22), 2617–2633. https://doi.org/10.1007/s00018-005-5148-6

Gomez-Simmonds, A., Greenman, M., Sullivan, S. B., Tanner, J. P., Sowash, M. G., Whittier, S., & Uhlemann, A.-C. (2015). Population structure of *Klebsiella pneumoniae* causing bloodstream infections at a New York City tertiary care hospital: Diversification of multidrug-resistant isolates. *Journal of Clinical Microbiology*, *53*(7), 2060–2067. https://doi.org/10.1128/JCM.03455-14

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Haaber, J., Penadés, J. R., & Ingmer, H. (2017). Transfer of antibiotic resistance in *Staphylococcus aureus*. *Trends in Microbiology*, *25*(11), 893–905. https://doi.org/10.1016/j.tim.2017.05.011

Hanage, W. P. (2016). Not so simple after all: Bacteria, their population genetics, and recombination. *Cold Spring Harbor Perspectives in Biology*, *8*(7). https://doi.org/10.1101/cshperspect.a018069

Hartman, B. J., & Tomasz, A. (1984). Low-affinity penicillin-binding protein associated with beta-lactam resistance in *Staphylococcus aureus*. *Journal of Bacteriology*, *158*(2), 513–516. https://doi.org/10.1128/JB.158.2.513-516.1984

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hunt, M., Mather, A. E., Sánchez-Busó, L., Page, A. J., Parkhill, J., Keane, J. A., & Harris, S. R. (2017). ARIBA: Rapid antimicrobial resistance genotyping directly from sequencing reads. *Microbial Genomics*, *3*(10), e000131. https://doi.org/10.1099/mgen.0.000131

International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC). (2009). Classification of staphylococcal cassette chromosome mec

(SCCmec): Guidelines for reporting novel SCCmec elements. *Antimicrobial Agents and Chemotherapy*, *53*(12), 4961–4967. https://doi.org/10.1128/AAC.00579-09

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 5114. https://doi.org/10.1038/s41467-018-07641-9

Jamrozy, D. M., Harris, S. R., Mohamed, N., Peacock, S. J., Tan, C. Y., Parkhill, J., Anderson, A. S., & Holden, M. T. G. (2016). Pan-genomic perspective on the evolution of the *Staphylococcus aureus* USA300 epidemic. *Microbial Genomics*, *2*(5), e000058. https://doi.org/10.1099/mgen.0.000058

Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., & Aarestrup, F. M. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *Journal of Clinical Microbiology*, *52*(5), 1501–1510. https://doi.org/10.1128/JCM.03617-13

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kaya, H., Hasman, H., Larsen, J., Stegger, M., Johannesen, T. B., Allesøe, R. L., Lemvigh, C. K., Aarestrup, F. M., Lund, O., & Larsen, A. R. (2018). SCCmecFinder, a web-based tool for typing of Staphylococcal Cassette Chromosome mec in *Staphylococcus aureus* using whole-genome sequence data. *MSphere*, *3*(1). https://doi.org/10.1128/mSphere.00612-17

Kos, V. N., Desjardins, C. A., Griggs, A., Cerqueira, G., Van Tonder, A., Holden, M. T. G., Godfrey, P., Palmer, K. L., Bodi, K., Mongodin, E. F., Wortman, J., Feldgarden, M., Lawley, T., Gill, S. R., Haas, B. J., Birren, B., & Gilmore, M. S. (2012). Comparative genomics of vancomycin-resistant *Staphylococcus aureus* strains and their positions within the clade most commonly associated with Methicillin-resistant *S. aureus* hospital-acquired infection in the United States. *MBio*, *3*(3). https://doi.org/10.1128/mBio.00112-12

Lam, M. M. C., Wyres, K. L., Duchêne, S., Wick, R. R., Judd, L. M., Gan, Y.-H., Hoh, C.-H., Archuleta, S., Molton, J. S., Kalimuddin, S., Koh, T. H., Passet, V., Brisse, S., & Holt, K. E. (2018). Population genomics of hypervirulent *Klebsiella pneumoniae* clonal-group 23 reveals early emergence and rapid global dissemination. *Nature Communications*, *9*(1), 2703. https://doi.org/10.1038/s41467-018-05114-7

Lerminiaux, N. A., & Cameron, A. D. S. (2019). Horizontal transfer of antibiotic resistance genes in clinical environments. *Canadian Journal of Microbiology*, *65*(1), 34–44. https://doi.org/10.1139/cjm-2018-0275

Lin, M., & Kussell, E. (2019). Inferring bacterial recombination rates from large-scale sequencing datasets. *Nature Methods*, *16*(2), 199–204. https://doi.org/10.1038/s41592-018-0293-7

Liu, B., Zheng, D., Jin, Q., Chen, L., & Yang, J. (2019). VFDB 2019: A comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Research*, *47*(D1), D687–D692. https://doi.org/10.1093/nar/gky1080

Liu, J., Chen, D., Peters, B. M., Li, L., Li, B., Xu, Z., & Shirliff, M. E. (2016). Staphylococcal chromosomal cassettes mec (SCCmec): A mobile genetic element in methicillin-resistant *Staphylococcus aureus*. *Microbial Pathogenesis*, *101*, 56–67. https://doi.org/10.1016/j.micpath.2016.10.028

Lockhart, P. B., Brennan, M. T., Sasser, H. C., Fox, P. C., Paster, B. J., & Bahrani-Mougeot, F. K. (2008). Bacteremia associated with toothbrushing and dental extraction. *Circulation*, *117*(24), 3118–3125. https://doi.org/10.1161/CIRCULATIONAHA.107.758524

Manara, S., Pasolli, E., Dolce, D., Ravenni, N., Campana, S., Armanini, F., Asnicar, F., Mengoni, A., Galli, L., Montagnani, C., Venturini, E., Rota-Stabelli, O., Grandi, G., Taccetti, G., & Segata, N. (2018). Whole-genome epidemiology, characterisation, and phylogenetic reconstruction of *Staphylococcus aureus* strains in a paediatric hospital. *Genome Medicine*, *10*(1), 82. https://doi.org/10.1186/s13073-018-0593-7

Medini, D., Donati, C., Tettelin, H., Masignani, V., & Rappuoli, R. (2005). The microbial pan-genome. *Current Opinion in Genetics & Development*, *15*(6), 589–594. https://doi.org/10.1016/j.gde.2005.09.006

Méric, G., Mageiros, L., Pensar, J., Laabei, M., Yahara, K., Pascoe, B., Kittiwan, N., Tadee, P., Post, V., Lamble, S., Bowden, R., Bray, J. E., Morgenstern, M., Jolley, K. A., Maiden, M. C. J., Feil, E. J., Didelot, X., Miragaia, M., de Lencastre, H., … Sheppard, S. K. (2018). Disease-associated genotypes of the commensal skin bacterium *Staphylococcus epidermidis*. *Nature Communications*, *9*(1), 5034. https://doi.org/10.1038/s41467-018-07368-7

Monecke, S., Coombs, G., Shore, A. C., Coleman, D. C., Akpaka, P., Borg, M., Chow, H., Ip, M., Jatzwauk, L., Jonas, D., Kadlec, K., Kearns, A., Laurent, F., O'Brien, F. G., Pearson, J., Ruppelt, A., Schwarz, S., Scicluna, E., Slickers, P., … Ehricht, R. (2011). A field guide to pandemic, epidemic and sporadic clones of methicillin-resistant *Staphylococcus aureus*. *PloS One*, *6*(4), e17936. https://doi.org/10.1371/journal.pone.0017936

Nadalin, F., Vezzi, F., & Policriti, A. (2012). GapFiller: A de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics*, *13 Suppl 14*, S8. https://doi.org/10.1186/1471-2105-13-S14-S8

Nei, M., & Tajima, F. (1981). Genetic drift and estimation of effective population size. *Genetics*, *98*(3), 625–640.

Nguyen, L. T., & Vogel, H. J. (2016). Staphylokinase has distinct modes of interaction with antimicrobial peptides, modulating its plasminogen-activation properties. *Scientific Reports*, *6*, 31817. https://doi.org/10.1038/srep31817

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nimmo, G. R., Steen, J. A., Monecke, S., Ehricht, R., Slickers, P., Thomas, J. C., Appleton, S., Goering, R. V., Robinson, D. A., & Coombs, G. W. (2015). ST2249-MRSA-III: A second major recombinant methicillin-resistant *Staphylococcus aureus* clone causing healthcare infection in the 1970s. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *21*(5), 444–450. https://doi.org/10.1016/j.cmi.2014.12.018

Packer, S., Pichon, B., Thompson, S., Neale, J., Njoroge, J., Kwiatkowska, R. M., Oliver, I., Telfer, M., Doumith, M., Buunaaisie, C., Heinsbroek, E., Hopewell-Kelly, N., Desai, M., Hope, V., Williams, O. M., Kearns, A., Hickman, M., & Gobin, M. (2019). Clonal expansion of community-associated meticillin-resistant *Staphylococcus aureus* (MRSA) in people who inject drugs (PWID): Prevalence, risk factors and molecular epidemiology, Bristol, United Kingdom, 2012 to 2017. *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, *24*(13). https://doi.org/10.2807/1560-7917.ES.2019.24.13.1800124

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics (Oxford, England)*, *31*(22), 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

Paharik, A. E., Salgado-Pabon, W., Meyerholz, D. K., White, M. J., Schlievert, P. M., & Horswill, A. R. (2016). The Spl serine proteases modulate *Staphylococcus aureus* protein production and virulence in a rabbit model of pneumonia. *MSphere*, *1*(5). https://doi.org/10.1128/mSphere.00208-16

Park, K.-H., Greenwood-Quaintance, K. E., Uhl, J. R., Cunningham, S. A., Chia, N., Jeraldo, P. R., Sampathkumar, P., Nelson, H., & Patel, R. (2017). Molecular epidemiology of *Staphylococcus aureus* bacteremia in a single large Minnesota medical center in 2015 as assessed using MLST, core genome MLST and spa typing. *PloS One*, *12*(6), e0179003. https://doi.org/10.1371/journal.pone.0179003

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Pietrocola, G., Nobile, G., Rindi, S., & Speziale, P. (2017). Staphylococcus aureus manipulates innate immunity through own and host-expressed proteases. *Frontiers in Cellular and Infection Microbiology*, *7*, 166. https://doi.org/10.3389/fcimb.2017.00166

Pinchuk, I. V., Beswick, E. J., & Reyes, V. E. (2010). Staphylococcal enterotoxins. *Toxins*, *2*(8), 2177–2197. https://doi.org/10.3390/toxins2082177

Planet, P. J., LaRussa, S. J., Dana, A., Smith, H., Xu, A., Ryan, C., Uhlemann, A.-C., Boundy, S., Goldberg, J., Narechania, A., Kulkarni, R., Ratner, A. J., Geoghegan, J. A., Kolokotronis, S.-O., & Prince, A. (2013). Emergence of the epidemic methicillin-resistant *Staphylococcus aureus* strain USA300 coincides with horizontal transfer of the arginine catabolic mobile element and speG-mediated adaptations for survival on skin. *MBio*, *4*(6), e00889-00813. https://doi.org/10.1128/mBio.00889-13

R Core Team. (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. http://www.R-project.org/

Recker, M., Laabei, M., Toleman, M. S., Reuter, S., Saunderson, R. B., Blane, B., Török, M. E., Ouadi, K., Stevens, E., Yokoyama, M., Steventon, J., Thompson, L., Milne, G., Bayliss, S., Bacon, L., Peacock, S. J., & Massey, R. C. (2017). Clonal differences in *Staphylococcus aureus* bacteraemia-associated mortality. *Nature Microbiology*, *2*(10), 1381–1388. https://doi.org/10.1038/s41564-017-0001-x

Richardson, E. J., Bacigalupe, R., Harrison, E. M., Weinert, L. A., Lycett, S., Vrieling, M., Robb, K., Hoskisson, P. A., Holden, M. T. G., Feil, E. J., Paterson, G. K., Tong, S. Y. C., Shittu, A., van Wamel, W., Aanensen, D. M., Parkhill, J., Peacock, S. J., Corander, J., Holmes, M., & Fitzgerald, J. R. (2018). Gene exchange drives the ecological success of a multi-host bacterial pathogen. *Nature Ecology & Evolution*, *2*(9), 1468–1478. https://doi.org/10.1038/s41559-018-0617-0

Robinson, D. A., & Enright, M. C. (2004). Evolution of *Staphylococcus aureus* by large chromosomal replacements. *Journal of Bacteriology*, *186*(4), 1060–1064. https://doi.org/10.1128/jb.186.4.1060-1064.2004

Rooijakkers, S. H. M., Ruyken, M., Roos, A., Daha, M. R., Presanis, J. S., Sim, R. B., van Wamel, W. J. B., van Kessel, K. P. M., & van Strijp, J. A. G. (2005). Immune evasion by a staphylococcal complement inhibitor that acts on C3 convertases. *Nature Immunology*, *6*(9), 920–927. https://doi.org/10.1038/ni1235

Sabat, A., Kosowska, K., Poulsen, K., Kasprowicz, A., Sekowska, A., van Den Burg, B., Travis, J., & Potempa, J. (2000). Two allelic forms of the aureolysin gene (aur) within *Staphylococcus aureus*. *Infection and Immunity*, *68*(2), 973–976. https://doi.org/10.1128/iai.68.2.973-976.2000

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Shariati, A., Dadashi, M., Moghadam, M. T., van Belkum, A., Yaslianifard, S., & Darban-Sarokhalil, D. (2020). Global prevalence and distribution of vancomycin resistant, vancomycin intermediate and heterogeneously vancomycin intermediate *Staphylococcus aureus* clinical isolates: A systematic review and meta-analysis. *Scientific Reports*, *10*(1), 12689. https://doi.org/10.1038/s41598-020-69058-z

Sharma, H., Turner, C. E., Siggins, M. K., El-Bahrawy, M., Pichon, B., Kearns, A., & Sriskandan, S. (2019). Toxic shock syndrome toxin 1 evaluation and antibiotic impact in a transgenic model of staphylococcal soft tissue infection. *MSphere*, *4*(5). https://doi.org/10.1128/mSphere.00665-19

Spaan, A. N., Henry, T., van Rooijen, W. J. M., Perret, M., Badiou, C., Aerts, P. C., Kemmink, J., de Haas, C. J. C., van Kessel, K. P. M., Vandenesch, F., Lina, G., & van Strijp, J. A. G. (2013). The staphylococcal toxin Panton-Valentine Leukocidin targets human C5a receptors. *Cell Host & Microbe*, *13*(5), 584–594. https://doi.org/10.1016/j.chom.2013.04.006

Spoor, L. E., Richardson, E., Richards, A. C., Wilson, G. J., Mendonca, C., Gupta, R. K., McAdam, P. R., Nutbeam-Tuffs, S., Black, N. S., O'Gara, J. P., Lee, C. Y., Corander, J., & Ross Fitzgerald, J. (2015). Recombination-mediated remodelling of host-pathogen interactions during *Staphylococcus aureus* niche adaptation. *Microbial Genomics*, *1*(4), e000036. https://doi.org/10.1099/mgen.0.000036

Strauß, L., Stegger, M., Akpaka, P. E., Alabi, A., Breurec, S., Coombs, G., Egyir, B., Larsen, A. R., Laurent, F., Monecke, S., Peters, G., Skov, R., Strommenger, B., Vandenesch, F., Schaumburg, F., & Mellmann, A. (2017). Origin, evolution, and global transmission of community-acquired *Staphylococcus aureus* ST8. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(49), E10596–E10604. https://doi.org/10.1073/pnas.1702472114

Tavaré, S. (1986). Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. In *American Mathematical Society: Lectures on Mathematics in the Life Sciences* (Vol. 17, pp. 57–86). Amer Mathematical Society. http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0821811673

Thomer, L., Schneewind, O., & Missiakas, D. (2016). Pathogenesis of *Staphylococcus aureus* bloodstream infections. *Annual Review of Pathology*, *11*, 343–364. https://doi.org/10.1146/annurev-pathol-012615-044351

Toleman, M. S., Reuter, S., Jamrozy, D., Wilson, H. J., Blane, B., Harrison, E. M., Coll, F., Hope, R. J., Kearns, A., Parkhill, J., Peacock, S. J., & Török, M. E. (2019). Prospective genomic surveillance of methicillin-resistant *Staphylococcus aureus* (MRSA) associated

with bloodstream infection, England, 1 October 2012 to 30 September 2013. *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, *24*(4). https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800215

Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W., & Corander, J. (2018). RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. *Wellcome Open Research*, *3*, 93. https://doi.org/10.12688/wellcomeopenres.14694.1

Uhlemann, A.-C., Dordel, J., Knox, J. R., Raven, K. E., Parkhill, J., Holden, M. T. G., Peacock, S. J., & Lowy, F. D. (2014). Molecular tracing of the emergence, diversification, and transmission of *S. aureus* sequence type 8 in a New York community. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(18), 6738–6743. https://doi.org/10.1073/pnas.1401006111

van Hal, S. J., Jensen, S. O., Vaska, V. L., Espedido, B. A., Paterson, D. L., & Gosbell, I. B. (2012). Predictors of mortality in *Staphylococcus aureus* Bacteremia. *Clinical Microbiology Reviews*, *25*(2), 362–386. https://doi.org/10.1128/CMR.05022-11

Volz, E. M., & Didelot, X. (2018). Modeling the growth and decline of pathogen effective population size provides insight into epidemic dynamics and drivers of antimicrobial resistance. *Systematic Biology*, *67*(4), 719–728. https://doi.org/10.1093/sysbio/syy007

Ward, M. J., Gibbons, C. L., McAdam, P. R., van Bunnik, B. a. D., Girvan, E. K., Edwards, G. F., Fitzgerald, J. R., & Woolhouse, M. E. J. (2014). Time-scaled evolutionary analysis of the transmission and antibiotic resistance dynamics of *Staphylococcus aureus* clonal complex 398. *Applied and Environmental Microbiology*, *80*(23), 7275–7282. https://doi.org/10.1128/AEM.01777-14

Young, B. C., Golubchik, T., Batty, E. M., Fung, R., Larner-Svensson, H., Votintseva, A. A., Miller, R. R., Godwin, H., Knox, K., Everitt, R. G., Iqbal, Z., Rimmer, A. J., Cule, M., Ip, C. L. C., Didelot, X., Harding, R. M., Donnelly, P., Peto, T. E., Crook, D. W., … Wilson, D. J. (2012). Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(12), 4550–4555. https://doi.org/10.1073/pnas.1113219109

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F. M., & Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. *The Journal of Antimicrobial Chemotherapy*, *67*(11), 2640–2644. https://doi.org/10.1093/jac/dks261

Appendix 2

Due to size, Tables S2.1-S2.5 were excluded. Requests for these tables can be made to joshuasmithmicro@gmail.com.

Table S2.1 Genome assembly, MLST, accession numbers and associated metadata of 323 *S. aureus* isolates

Table S2.2 Distribution of genes in the *S. aureus* pan-genome identified using Roary

Table S2.3 Distribution of horizontally acquired antimicrobial resistance genes identified using ResFinder

Table S2.4 Distribution of virulence genes identified using VirulenceFinder and the VFDB

Table S2.5 Recombination parameters inferred by mcorr for each of CC5 and CC8

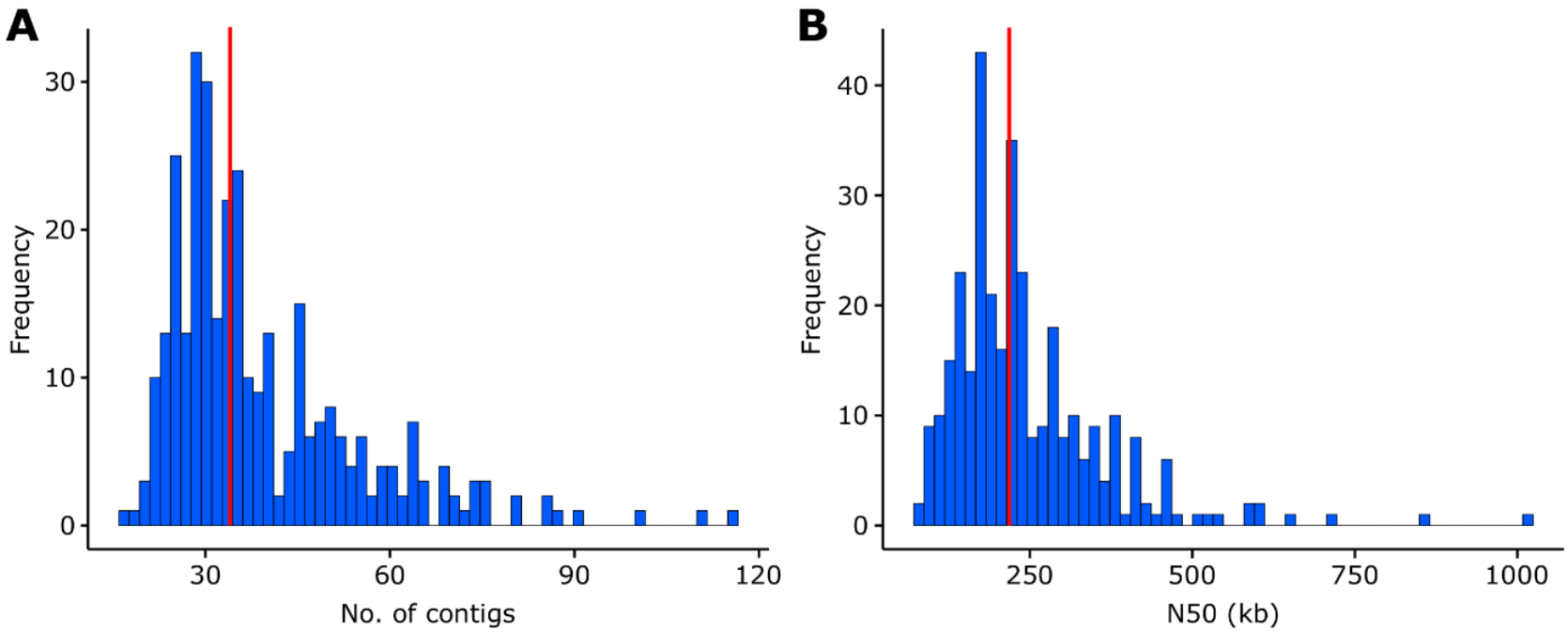

Figure S2.1 Assembly statistics of the *S. aureus* genomes

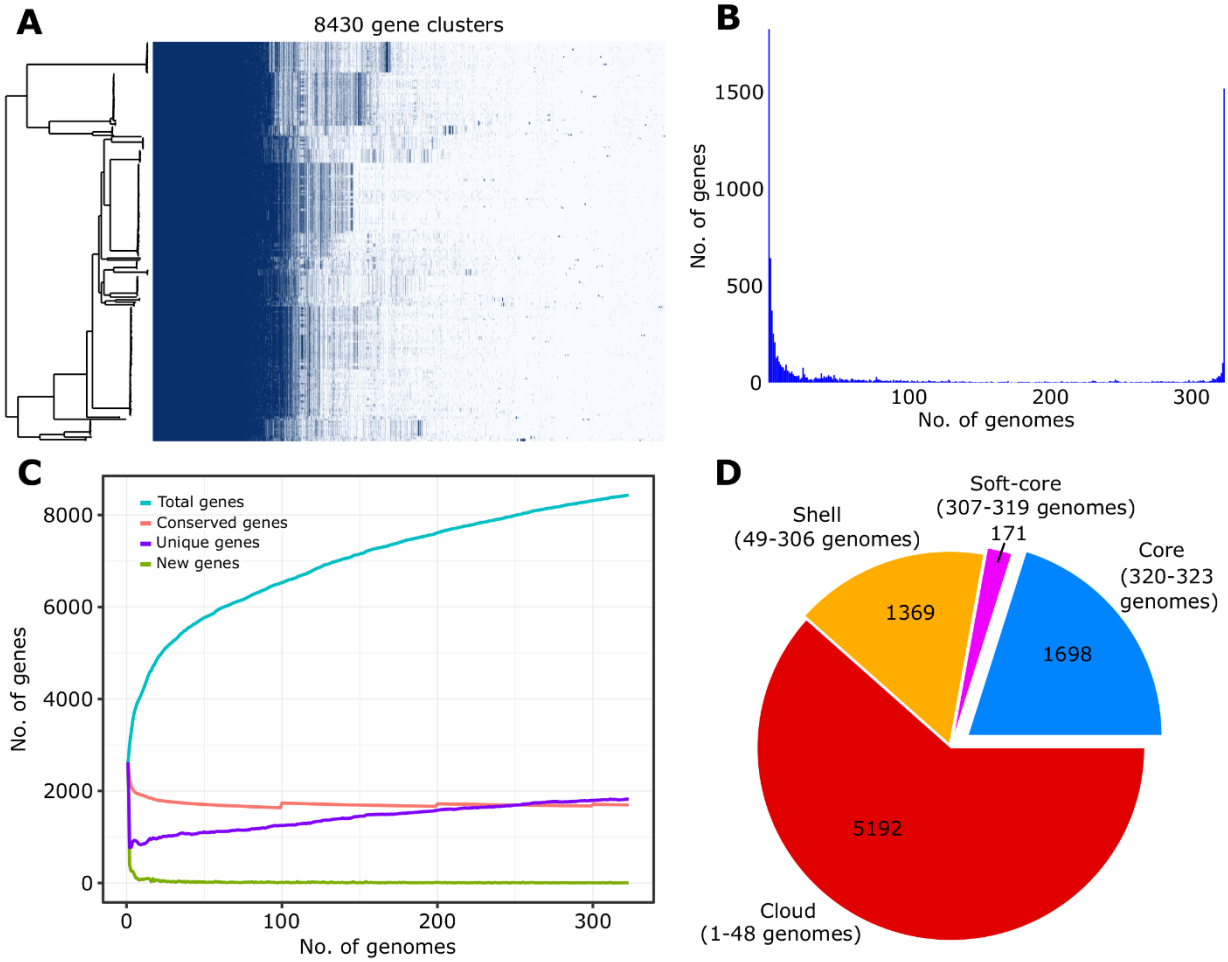

Figure S2.2 Pan-genome analysis of 323 *S. aureus* genomes. (a) Presence-absence matrix of gene clusters as determined by Roary, aligned to the phylogeny. Blue indicates presence. (b) Gene frequency histogram indicating the number of genomes each gene is present in. (c) The size of the pan-genome (blue), core genome (red), unique gene additions (purple) and new genes (green) as related to the number of individuals in the population. Proportional illustration of core, soft-core, shell, and cloud genes across the pangenome.

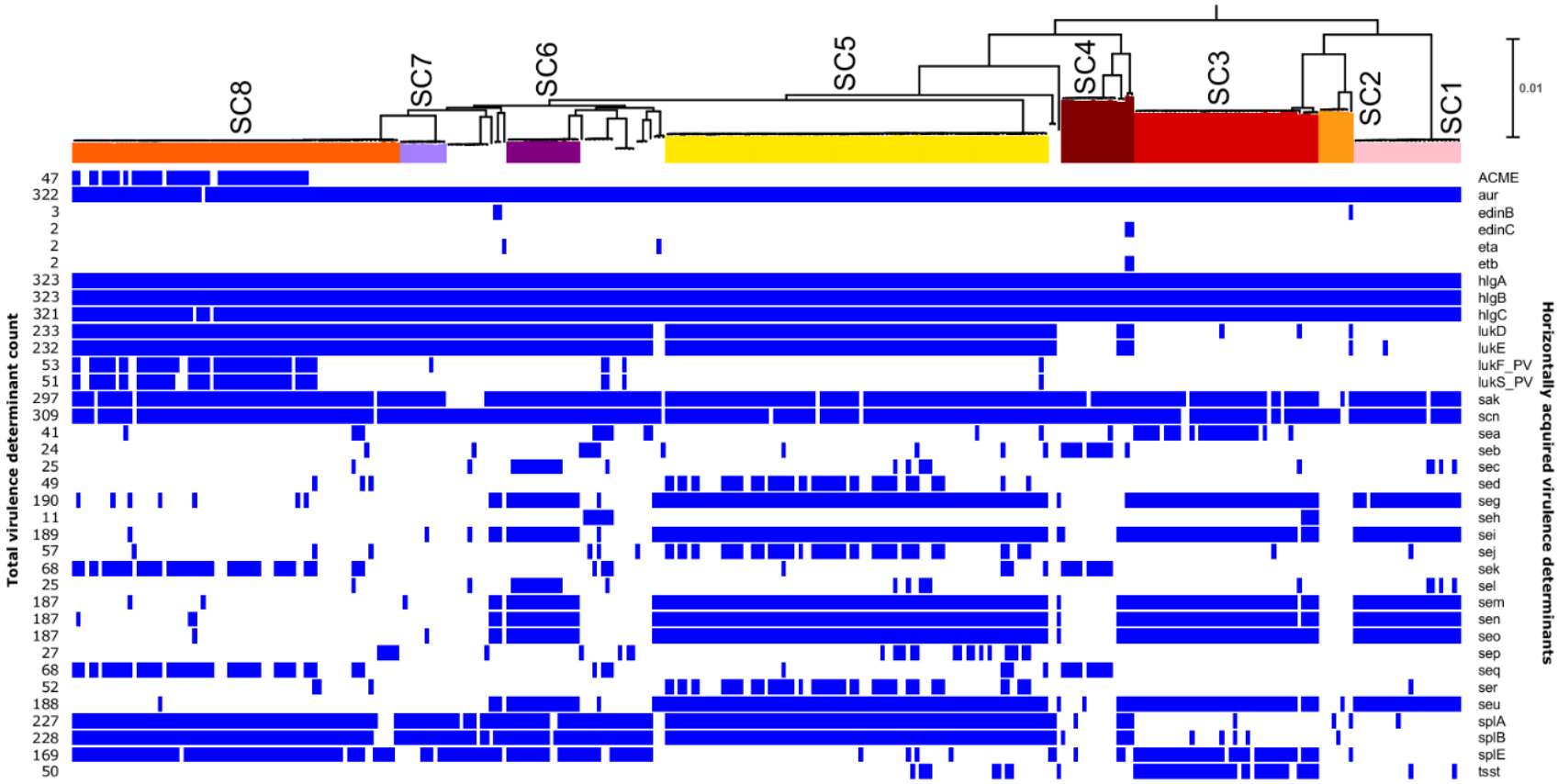

Figure S2.3 Distribution of virulence genes using VirulenceFinder and the Virulence Factor Database (VFDB). Details are shown in Table S2.4.

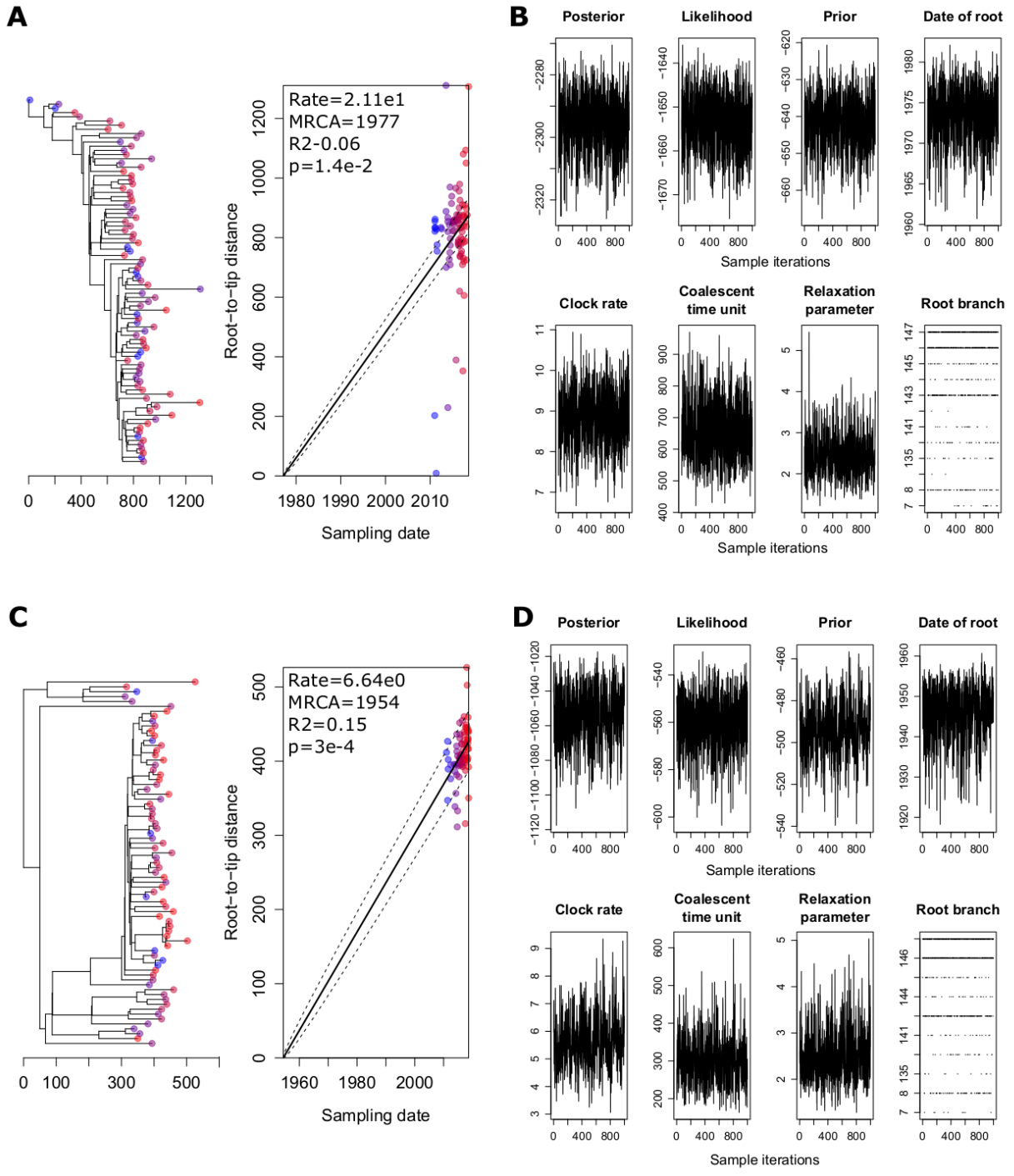

Figure S2.4 Bactdating statistical tests and MCMC trace plots. (a) Initial rooted phylogeny and correlation test between date and root-to-tip distance withing the phylogeny for CC5. (b) Bactdating trace plots constructed by periodic sampling over the MCMC runs for CC5. (c) Initial phylogeny and correlation test for CC8. (d) Bactdating trace plots for CC8.

CHAPTER 3

Highways of inter-species recombination in coagulase-negative *Staphylococcus*

Joshua T. Smith[a], Cheryl P. Andam[b]

Article under review in *Genome Biology and Evolution*

[a] Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, New Hampshire, USA 03824

[b] Department of Biological Sciences, University at Albany, State University of New York, Albany, New York, USA 12222

Abstract

Coagulase-negative *Staphylococcus* (CoNS) is a heterogenous group of bacteria that has been historically defined based on the diagnostic presentation of the coagulase protein. This classification scheme initially proved useful to differentiate the highly pathogenic *Staphylococcus aureus* from other species that were considered to be less clinically relevant. However, little is known of the phylogenetic relationships of CoNS from a genomic standpoint. We compiled a dataset of 1,876 publicly available named CoNS genomes, which can be delineated into 55 species based on allele differences in 462 core genes and variation in accessory gene content. Among the ten species with the highest number of genomes, there is large variation within and between species in terms of gene content. We identified nine structural types of the mobile genetic element SCC*mec*, which carries the methicillin resistance determinant *mecA* and were distributed across multiple species. Frequently recombined genes included those associated with resistance to antimicrobials and heavy metals, surface-associated

proteins related to virulence and biofilm formation, type VII secretion system, iron capture, recombination and metabolic enzymes. Homologous recombination within and between species plays a major role in CoNS evolution, but recombination rates varied between species. The highest frequencies of inter-species receipt and donation of recombined DNA were detected in *Staphylococcus capitis*, *Staphylococcus caprae*, *Staphylococcus hominis*, *Staphylococcus haemolyticus* and members of the Saprophyticus species group. We conclude that certain *Staphylococcus* species and ecological sources (animals and animal-derived food) function as hubs of gene flow and major reservoir of genetic diversity for the entire genus. Little is known of the evolutionary history and genome structure of coagulase-negative *Staphylococcus* (CoNS), which are now being recognized as major opportunistic pathogens of humans. Here, we report that CoNS are more phylogenetically diverse than previously recognized, with frequent but biased recombination contributing to its evolution. Defining the evolutionary processes that contribute to CoNS diversification is the first step towards effective surveillance, diagnostics and control of CoNS infections.

Introduction

Members of the bacterial genus *Staphylococcus* have historically been classified into coagulase-positive *Staphylococcus* (CoPS) and coagulase-negative *Staphylococcus* (CoNS). This binomial classification system is based on whether a species exhibits clotting of plasma by the coagulase enzyme, which converts fibrinogen to fibrin (Crosby et al., 2016). Coagulase production as a diagnostic method was developed in the 1940s to discriminate the highly pathogenic *Staphylococcus aureus* from other species that were considered to be less clinically relevant (Fairbrother, 1940). At that time, CoPS was almost exclusively represented by *S. aureus*. Since then, numerous CoPS species have been identified. There are also coagulase-variable species, i.e., a species composed of strains that are either coagulase-positive or -negative). For example, *Staphylococcus schleiferi* consists of a coagulase-negative subspecies (*S. schleiferi* subsp. *schleiferi*) and a coagulase-positive subspecies (*S. schleiferi* subsp. *coagulans*), but the phenotypic distinction between the two is often blurry (Vandenesch et al., 1994). Two other species are coagulase-variable (*Staphylococcus hyicus* and *Staphylococcus agnetis*) (Hassler et al., 2008; Taponen et al., 2012). As of 2014, the genus consisted of 47 species and 23 subspecies, with 38 belonging to CoNS (Becker et al., 2014). As more *Staphylococcus* species are discovered, the coagulase-based classification scheme proves inadequate to resolve the differences among these less well-known species.

There is mounting evidence from recent diagnostic studies that CoNS species pose a significant health burden as opportunistic pathogens. Because CoNS represent a major component of the microbiota of skin and mucous membranes in humans and animals, colonization appears to be a key source of endogenous infections (Akiyama et al., 1998; Fux et

al., 2005; Zingg et al., 2009). Breaching the skin barrier through use of transiently or permanently inserted medical devices (e.g., intravenous catheters, vascular grafts, prosthetic joints) thus offers CoNS opportunities to gain access to host tissues and form biofilms on foreign body surfaces (von Eiff et al., 2006; Y. Zheng et al., 2018). A few CoNS species are particularly notable. *Staphylococcus epidermidis* and *Staphylococcus haemolyticus* are most commonly associated with infections of indwelling or implanted foreign bodies (Ahmed et al., 2019; Both et al., 2021). *Staphylococcus saprophyticus* is the common cause of urinary tract infections (Lawal et al., 2021; Widerström et al., 2007), while *Staphylococcus lugdunensis* can cause infectious endocarditis (Non & Santos, 2017). *Staphylococcus capitis* has been implicated in sepsis in preterm infants in neonatal intensive care units (Stenmark et al., 2019; Wirth et al., 2020). CoNS are also reported to be responsible for a variety of laryngological diseases (i.e., diseases of the upper respiratory tract and other organs present in the neck and head) (Michalik et al., 2020). Worldwide, increasing antimicrobial resistance in CoNS species in healthcare and community settings further exacerbates the health burden they cause and limits available therapeutic options (Asante et al., 2021; May et al., 2014; Pedroso et al., 2018).

CoNS clearly comprise a remarkably heterogeneous group, ranging from true non-pathogenic to facultative pathogenic species with variable virulence potential (Becker et al., 2014). The binomial coagulase-based classification scheme can obfuscate accurate species identification, which may have important clinical consequences. The lack of a robust evolution-based classification and identification scheme is therefore an important barrier to determining a clear association between CoNS species and specific diseases. Hence, there is a pressing need to monitor these bacteria and identify strains and species that may be pre-disposed to pathogenicity.

In this study, we aimed to elucidate the phylogenetic relationships and genomic characteristics of CoNS species. Using 1,876 publicly available high-quality genomes representing 55 species, we showed that CoNS species are remarkably diverse in terms of their core allelic and gene content variation, including antimicrobial resistance genes. Frequent recombination within and between species plays a major role in CoNS evolution, but frequencies of DNA donation and receipt varied significantly between recombination partners. These results provide important insights into the evolutionary history of this obscure group of *Staphylococcus* species. Our findings are relevant for developing an improved evolution-based classification system, surveillance and management options of CoNS-related infections.

Methods

Dataset

We retrieved 2,258 genome assemblies of all named CoNS species from the National Center for Biotechnology Information (NCBI) Reference Sequence database in November 2020. To reduce ambiguity, we excluded those species with coagulase-variable phenotypes (*S. agentis*, *S. hyicus* and *S. schleiferi*). To ensure that only high-quality genomes were included, we used QUAST v.5.0.2 (Gurevich et al., 2013) to identify and exclude genomes with >200 contigs and <40,000 bp N50 scores. We also used CheckM v.1.1.2 to exclude genomes with <90% completeness and >5% contamination (Parks et al., 2015). A total of 1,876 genomes passed these criteria and were subsequently used in all downstream analyses (Table S3.1). The number of contigs ranged from 1 - 194 and N50 values ranged from 40,254 - 2,941,886 bp. To confirm species identity, we used fastANI v.1.1 to calculate the average nucleotide identity (ANI) for all

pairs of genomes (Jain et al., 2018). We used the ≥95% ANI threshold to define a species (Jain et al., 2018). Genomes identified as a single species in NCBI but exhibited <95% ANI were designated with capital letters (e.g., *Staphylococcus xylosus* A, B or C) to distinguish them as separate species.

Pan-genome analysis and phylogenetic reconstruction

We used Prokka v.1.14.6 to annotate the genome sequences (Seemann, 2014). To estimate the pan-genome of the entire dataset, we used Panaroo v.1.2.3 that employs a graph-based algorithm (Tonkin-Hill et al., 2020) to cluster gene families using CD-HIT (Fu et al., 2012). Panaroo enables stringent quality control checks to correct annotation errors, fragmented assemblies, mistranslation and contamination of the genomes (Tonkin-Hill et al., 2020). We used Panaroo's strict mode for contamination and erroneous annotation removal to ensure that only high-quality genomes were included. The presence or absence of a gene in each genome was used to define the core and accessory genes. The genes in the pan-genome were classified into core genes (present in ≥99% of genomes), soft-core genes (present in 95 to <99% of genomes), shell genes (present in 15 to <95% of genomes) and cloud genes (present in <15% of genomes). Sequences of each gene family were aligned using MAFFT (Katoh et al., 2009). The aligned sequences of the core genes were then concatenated and were used as input to build a maximum likelihood phylogenetic tree using IQTree v.2.0.3 (Minh et al., 2020). We used a general time reversible (GTR) nucleotide substitution model (Tavaré, 1986) and a FreeRate rate heterogeneity model with four categories. We used the IQTree built-in UFBoot2 (Hoang et al., 2018) to carry out 1000 ultrafast bootstrap replicates. We also re-ran the Panaroo pipeline (Tonkin-Hill et al., 2020) for each species that has at least 30 genomes each to determine the gene content of each

species. Phylogenetic trees were visualized using the online platform Interactive Tree of Life (IToL) (Letunic & Bork, 2019).

In silico detection of resistance genes, virulence genes, plasmids and SCCmec

We used ABRicate v.0.8.13 (https://github.com/tseemann/abricate) to detect horizontally acquired genes conferring resistance to different antimicrobial classes and genes associated with virulence. ABRicate employs the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) to screen the Resfinder (Bortolaia et al., 2020) and Virulence Factor databases (Liu et al., 2019) (last updated in January 2021). To increase the likelihood of accurate gene detection, we used 95% sequence identity and coverage thresholds for resistance and virulence gene detection. Plasmids were identified using the PlasmidFinder database (Carattoli & Hasman, 2020). We used SCCmecFinder v.1.2 (Kaya et al., 2018) to identify the presence and type of *mecA*-carrying SCC*mec* using default minimum thresholds of > 60% for sequence coverage and > 90% sequence identity.

Recombination detection

We used two approaches to detect genome-wide recombination. First, we used the coalescent-based mcorr program, which measures the degree to which any two loci separated by defined number of nucleotides have correlated synonymous substitutions (Lin & Kussell, 2019). The correlation profiles were then used to calculate six parameters related to recombination: diversity, mutational divergence, recombinational divergence, proportion of sites in the genome whose diversity was derived from recombination (or recombination coverage), mean

recombination fragment size, and relative rate of recombination to mutation (Lin & Kussell, 2019).

We also used fastGEAR to identify specific core and shared accessory genes that have experienced recent and ancestral recombination (Mostowy et al., 2017). A recent recombination involves a few strains, while an ancestral recombination affects entire lineages (Mostowy et al., 2017). FastGEAR uses a hidden Markov model (HMM) to compare polymorphic sites in the strain's sequence and compares them to those found in members of its own lineage and in other lineages. The comparison is made over multiple HMM iterations, with parameters updated for each iteration based on results from the prior run. To test the significance of the inferred recombinations and identify false-positive results, we used a diversity test implemented in fastGEAR to compare sequence variation of the recombined fragment to its background (Mostowy et al., 2017). By default, fastGEAR determines the recipient genome of a recent recombination sequence and the lineage from which the sequence was donated. To predict the origin of the recently recombined regions, we identified the specific donor genome by developing a custom script to extract the recombined sequence from the recipient genome and then compare it to members of the donor lineage using BLAST (Altschul et al., 1990). The closest matching donor with a BLAST e-value of 1e-5 or less was then paired with the recipient. Recent recombinations were visualized in IToL (Letunic & Bork, 2019) and the postprocessing scripts provided by fastGEAR.

We used the default parameters for each program unless indicated otherwise.

Statistical tests

We used Kruskal-Wallis to compare genome content variation among species. We used ANOVA to determine significant differences in the mcorr parameters among species. P values were adjusted using the Benjamini-Hochberg method. To account for variation in the number of genomes per species, the frequencies of DNA donations were adjusted by dividing the number of observed donations per species by the number of genomes of a species. A similar calculation was done for DNA receipts.

To account for the differences in the number of genomes in each ecological source, we used the combinations formula as a normalizing factor to compare frequencies of recombination between strains from the same source:

$$\frac{2r}{p(p-1)}$$

and different sources:

$$\frac{r}{pq}$$

where r is the number of recombination events, p is the number of genomes from the first source, and q is the number of genomes from the second source.

Data availability

All genome sequences and associated metadata used in this study are publicly available from NCBI. Accession numbers for each genome are listed in Table S3.1.

Results

CoNS is phylogenetically, geographically and ecologically diverse

The early history of *Staphylococcus* discovery was characterized by several instances of taxonomic reclassifications and renaming of species. In 2012, Lamers et al. developed a multilocus sequence classification scheme based on four loci (16S rRNA, *dnaJ, rpoB tuf*) (Lamers et al., 2012). Using this molecular classification scheme and phenotypic properties (oxidase and novobiocin susceptibility), Lamers et al. proposed to classify *Staphylococcus* species into six species groups (Auricularis, Hyicus-Intermedius, Epidermidis-Aureus, Saprophyticus, Simulans and Sciuri) (Becker et al., 2014; Lamers et al., 2012). Here, we sought to determine the phylogenetic relationships of the 1,876 CoNS genomes retrieved from NCBI and ascertain the phylogenetic basis of the six species groups according to their genomic content.

Genome sizes ranged from 2.10 to 3.12 Mb (mean = 2.53 Mb), while the number of orthologous genes per genome ranged from 1,992 to 3,080 (mean = 2,410) (Table S3.1). The entire CoNS pan-genome consisted of 59,238 orthologous gene clusters (Table S3.2), of which 462 genes were present in ≥95% of the genomes (i.e., 239 core + 223 soft core genes). The accessory genome consisted of 2,267 shell genes and 56,509 cloud genes. Of the latter, a total of 10,667 genes were singletons, i.e., detected in only a single genome, and represented 18% of the pan-genome. The average ANI value calculated for all pairs of genomes was 83.47% (range = 77.02 - 100%). Using the 95% ANI threshold to define a species (Jain et al., 2018), we identified 55 species in the dataset.

CoNS genomes came from all seven continents, albeit in highly variable numbers (Figure S3.1). Majority came from North America and Europe, with 825 and 552 genomes, respectively. The dataset also included 83 genomes that were derived from the International Space Station. The strains were derived from a variety of sources (human, animal, environment and food items) (Table S3.1). A total of 994 genomes came from human samples.

We built a maximum likelihood phylogenetic tree using sequence variation in 462 core genes (**Figure 3.1**). The core genome phylogeny showed clearly defined species boundaries among the 55 species and among the six species clusters. The largest species group Epidermidis consisted of 1,357 genomes, which can be delineated into 17 species. *S. epidermidis* represented the species with the highest number of genomes (n = 732). This was not surprising as *S. epidermidis* is an opportunistic pathogen associated with major human infections and thus has been widely sampled (Méric et al., 2018). The midpoint-rooted tree also showed the Sciuri species group, consisting of five species, located at the base of the phylogeny. This result is consistent with previous studies that establish it is the ancestral group of *Staphylococcus* (Lamers et al., 2012; Rolo et al., 2017). The four other species groups Auricularis, Hyicus-Intermedius, Saprophyticus and Simulans were also clearly defined on the phylogeny and consisted of 91 (seven species), 70 (six species), 31 (three species), and 269 (17 species) genomes, respectively. Estimation of the genome-wide ANI values also revealed the boundaries between species that comprise each species group (Figure S3.2).

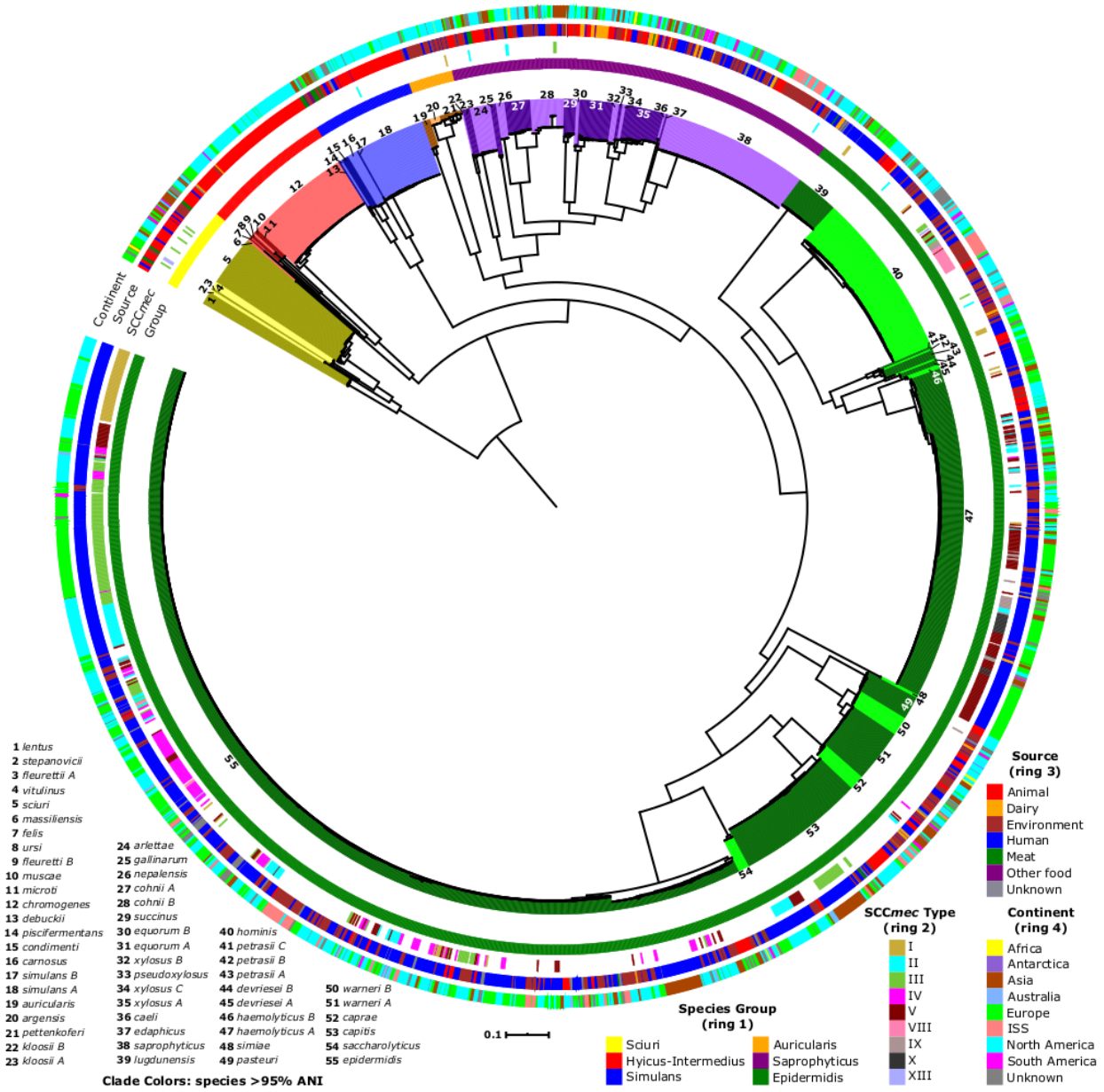

Figure 3.1 Maximum likelihood phylogenetic tree of 1,876 CoNS genomes representing 55 species. The midpoint-rooted tree was built using the concatenation of 462 core genes. Species delineation is based on the ≥95% average nucleotide identity (ANI) threshold. Species are indicated by colored branches and numbered 1-55. Tree scale represents the number of nucleotide substitutions per site. Rings outside the tree (inner to outer) show the species group, SCC*mec* type, source and geographical location. Species group is based on the classification by (Lamers et al., 2012) and member species of each species group are indicated by alternating light and dark shades of the same color. Some species were designated A, B or C if they were named the same species in NCBI but likely represent different species based on phylogenetic position and pairwise ANI values of <95%. ISS – International Space Station.

We detected a few cases of discrepancies in how the genomes were named. Some genomes that were initially identified as belonging to a single species can be delineated into more than one species based on their phylogenetic placement and ANI values. These included *Staphylococcus cohnii, Staphylococcus devriesei, Staphylococcus equorum, Staphylococcus fleurettii, Staphylococcus haemolyticus, Staphylococcus kloosi, Staphylococcus petrasii, Staphylococcus simulans, Staphylococcus warneri* and *Staphylococcus xylosus. S. petrasii* and *S. xylosus* can be differentiated into three distinct species, while the rest into two distinct species. *S. fleurettii* A and B species were found in species groups Sciuri (consistent with the results of reference (Lamers et al., 2012)) and Hyicus-Intermedius, respectively. We also found few instances whereby some genomes that were identified in NCBI as belonging to one species or species group were more closely related to others. For example, two genomes named as *Staphylococcus pasteuri* in NCBI (GCF_000494875.1 and GCF_002276895.1) were more closely related to genomes in *S. warneri* A. Other examples are listed in Table S3.3. Lastly, we found conflicts in the classification of some genomes to the species groups. The two *Staphylococcus massiliensis* genomes were more closely related to the Hyicus-Intermedius species group, rather than the Saprophyticus group as previously reported (Lamers et al., 2012). *Staphylococcus pettenkoferi* clusters within the Auricularis species group, rather than the Saprophyticus species group as previously reported (Lamers et al., 2012).

Overall, we found that CoNS is more diverse than previously recognized, highlighting the need for a more extensive sampling of this obscure bacterial group. Numerous species were poorly represented in genomic sequencing efforts and some species have conflicting taxonomic features.

CoNS species are a reservoir of diverse and transferrable resistance genes

We next sought to compare the genomic content between species. Here, we focused on the ten most common species that consisted of at least 30 genomes each. The ten species also represented the six species groups. All ten species have large and open pan-genomes (Figure S3.3). An open pan-genome is one in which the size of the pan-genome increases with the addition of new genomes (Medini et al., 2005), which may reflect ecological adaptation and/or neutral evolution (Andreani et al., 2017; McInerney et al., 2017). The total number of orthologous genes per species ranged from 3,222 in *S. lugdunensis* to 8,298 in *S. epidermidis* (Figure S3.3 and Table S3.4). The ten species also differed in terms of the number of intra-species pairwise single nucleotide polymorphisms (SNPs) in the core genome and accessory genes per genome (p = 0.0 and 1.96e-111, Kruskall-Wallis test; **Figure 3.2AB**). The median number of SNPs in the core genome among strains of the same species ranged from six in *S. saprophyticus* to 50 in *S. hominis*. The median number of accessory genes ranged from 139 in *S. lugdunensis* to 503 in *S. sciuri*. The ten species also varied in the number of plasmids that each genome carries (p = 1.20e-77, Kruskall-Wallis test; **Figure 3.2C**).

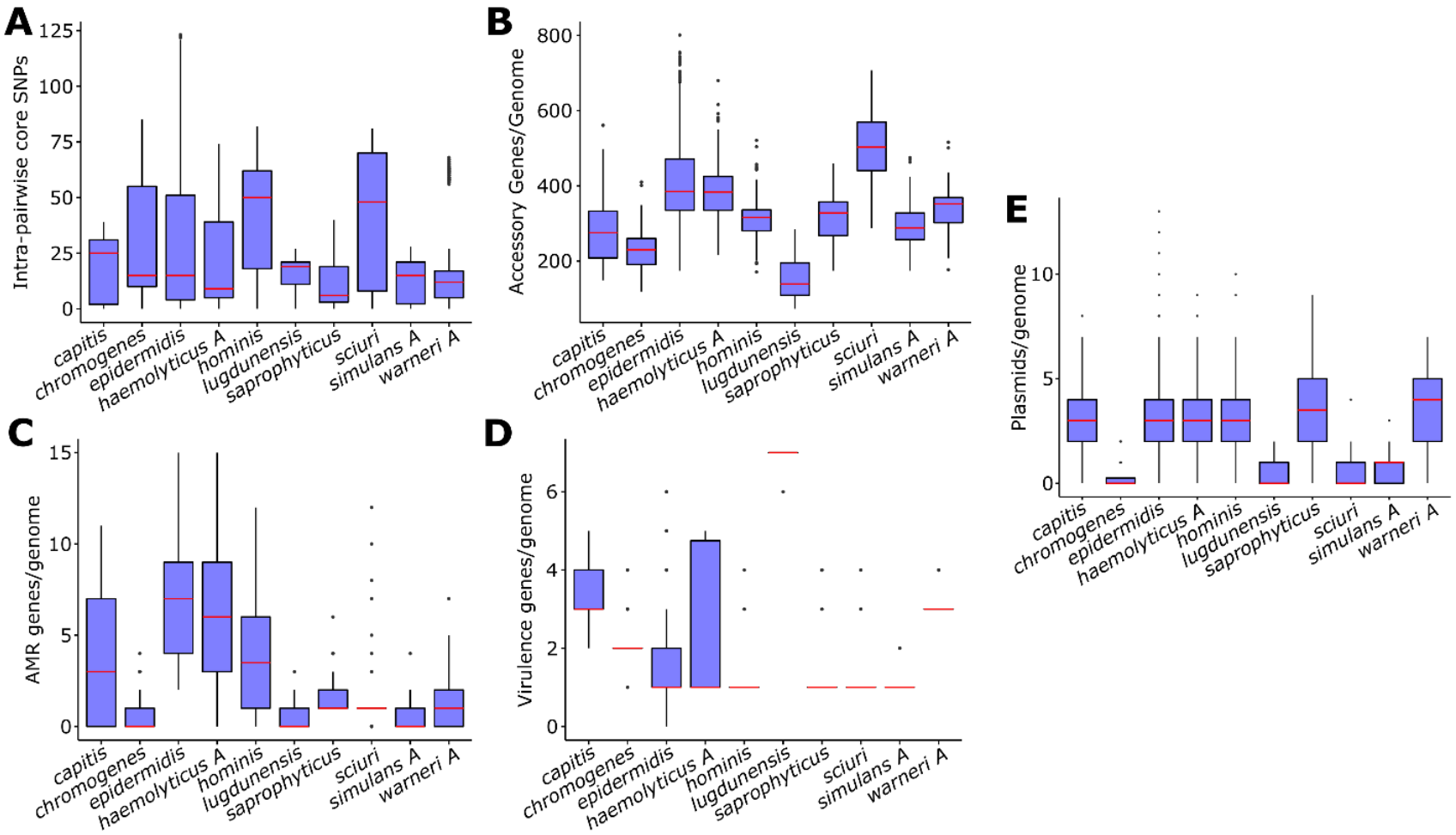

Figure 3.2 Genomic diversity plots of ten species in Figure 3.1 with the highest number of representative genomes. Comparison of (A) intra-species pairwise core SNPs, (B) number of accessory genes per genome, (C) number of plasmids per genome, (D) number of horizontally acquired antimicrobial resistance genes per genome, and (E) number of virulence genes per genome. Box plots depict the first quartile, median, third quartile values, while whiskers represent minimum and maximum values. Outliers are depicted as single points.

The horizontally acquired antimicrobial resistance genes we detected *in silico* represented a total of 16 antimicrobial classes. The ten species differed in the number of resistance genes per genome (p = 4.81e-142, Kruskall-Wallis test; **Figure 3.2D**), ranging from zero in *S. chromogenes*, *S. lugdunensis* and *S. simulans* A to seven in *S. epidermidis*. In *S. epidermidis* (n = 732 genomes), the most frequently detected resistance genes were the *norA* gene (fluoroquinolone resistance; n = present in 731 genomes), *dfrC* (trimethoprim resistance; n = 675 genomes) and *blaZ* (beta-lactam resistance; n = 608 genomes). In *S. haemolyticus* A (n = 254 genomes), the most commonly found resistance genes were *blaZ* (n = 182 genomes), *mecA* (beta-lactam resistance; n = 151 genomes) and *qacA* (fluoroquinolone resistance; n = 148 genomes). The most common resistance genes detected in *S. hominis* (n = 128 genomes) were *blaZ* (n = 90 genomes), *qacA* (n = 74 genomes) and *mecA* (n = 39 genomes). In *S. capitis* genomes (n = 98 genomes) include *blaZ* (n = present in 63 genomes), *mecA* (n = 44 genomes), and aac(6')-aph(2") (aminoglycoside resistance; n = 41 genomes). Finally, in *S. saprophyticus* (n = 112 genomes), the most frequently detected resistance genes were *fusD* (fusidic acid resistance; n = 112 genomes), *tetK* (tetracycline resistance; n = 25 genomes) and *msrA* (erythromycin and streptogramin B resistance; n = 14 genomes).

Several antimicrobial resistance genes were found to be unique to certain species. In addition to *norA* which was identified in almost every *S. epidermidis* strain, we also detected streptogramin A resistance genes *vatB* and *vgaB*, respectively (n = 9 genomes) which are often transferred together on a single plasmid (Haroche et al., 2003). Unique to *S. haemolyticus* A was the clindamycin resistance gene *lsaB* (n = 3 genomes). The beta-lactamase genes arl-1, arl-2, and arl-3 were detected only in *Staphylococcus arlettae* (n = 8, 5, and 3 genomes, respectively). Two

genes which confer resistance to macrolide, lincosamide, and streptogramin B were detected only in *S. lentus* (*erm(43)*; n = 5 genomes) and *S. chromogenes* (*ermT*; n = 5 genomes). Unique to every *S. saprophyticus* genome (n = 112 genomes) was *fusD*. The *fusF* gene (fusidic acid resistance; n = 23) was detected only in the two *S. cohnii* species A and B. Finally, the *salA* gene was found in majority of *S. sciuri* genomes (resistance to lincosamides and streptogramins; n = 28 genomes). The distribution of resistance genes in other CoNS species is shown in Figure S3.4 and Table S3.5).

The protein product of *mecA* is the penicillin-binding protein PBP2a, which has a low affinity to methicillin and other broad-spectrum beta-lactams (Fishovitz et al., 2014). We detected the *mecA* gene in 18 species. The mobile SCC*mec*, which carries the *mecA* gene, can be differentiated into 14 structurally diverse types (International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC), 2009). Across the entire dataset, we detected nine SCC*mec* types (I, II, III, IV, V, VIII, IX, X, XIII). The most common SCC*mec* types were type V (n = 168 genomes), type III (n = 158 genomes), and type IV (n = 107 genomes). The species with the most genomes, *S. epidermidis*, had eight SCC*mec* types (all except type V). The intermingled distribution of different SCC*mec* types across the phylogeny suggested multiple independent acquisition events of SCC*mec* via horizontal gene transfer in different lineages.

CoNS species harbor virulence genes often found in S. aureus

Among the ten most common species with the highest number of representative genomes, the number of virulence-associated genes per genome varied from a median of one in most

species, including *S. epidermidis*, *S. haemolyticus A*, *S. hominis*, *S. saprophyticus*, *S. sciuri* and *S. simulans*, to a median of seven in *S. lugdunensis* (p = 4.72e-116, Kruskall-Wallis test; **Figure 3.2E**). The most common virulence gene *clpP* was detected in 1,194 genomes, representing 100% of genomes from the 54 non-*S. epidermidis* species (n = 1,144 genomes) and 6.83% of *S. epidermidis* genomes (n = 50 genomes). The proteolytic activity of ClpP plays a major role in virulence, stress response and physiology of *S. aureus* (Michel et al., 2006) and other pathogenic species (Knudsen et al., 2013; Zhao et al., 2016; J. Zheng et al., 2020).

Some of the virulence genes we detected were notable. The genes *hld* and *icaA* were found almost exclusively in *S. epidermidis* and closely related species. A total of 520 *S. epidermidis* genomes, three *Staphylococcus saccharolyticus* genomes and two *Staphylococcus simiae* genomes contained the *hld* gene. The *hld* gene codes for a delta-hemolysin that is cytolytic to neutrophils and erythrocytes and can induce chronic inflammatory skin disease by triggering mast cell degranulation in *S. aureus* (Nakamura et al., 2013). The *icA* gene was detected in 330 *S. epidermidis* genomes, 94 *S. capitis* genomes, 11 *Staphylococcus caprae* genomes, three *S. saccharolyticus* genomes and two *S. simiae* genomes. The *ica* operon, of which *icaA* is part of, functions in the production of polysaccharide intercellular adhesin which mediates cell to cell adhesion and biofilm formation (Cue et al., 2012). This adhesin has been implicated in catheter-associated infections (Arciola et al., 2001). Lastly, many non-*epidermidis* species carried the *cap8* genes, which are involved in the synthesis of the serotype 8 capsular polysaccharide and is the most prevalent capsule type in clinical isolates of *S. aureus* (Luong & Lee, 2002; O'Riordan & Lee, 2004). In *S. aureus*, the capsule enhances virulence by impeding phagocytosis and stimulating colonization and persistence on mucosal surfaces (Luong & Lee,

2002; O'Riordan & Lee, 2004). Capsule genes had been previously detected in *S. lugdunensis* (Lebeurre et al., 2019), and our results greatly expand the taxonomic distribution of *cap8* genes. We detected four types of *cap8* genes that were frequently found in *S. sciuri, S. chromogenes, S. lugdunensis* and *S. haemolyticus* A genomes. The distribution of other virulence-associated genes in other CoNS species is shown in Figure S3.5 and Table S3.6.

CoNS species exhibit frequent but biased recombination

Recombination is known to contribute to the evolution and ecology of *S. aureus* (Driebe et al., 2015; Spoor et al., 2015), yet little is known of the frequency of recombination in CoNS species. Inter-species recombination has been reported between *S. aureus* and *S. epidermidis* due to overlapping niche space in the human skin and nasal pharynx (Méric et al., 2015). We therefore hypothesized that CoNS species have had a history of frequent recombination, which may partly explain the remarkable genomic variation we observed.

We first estimated six different evolutionary and recombination parameters for each of the ten most common species using mcorr (**Figure 3.3** and Table S3.7). The species with the lowest diversity were *S. epidermidis* (0.002356) and *S. warneri* A (0.002437), while *S. sciuri* (0.006427) and *S. simulans* (0.00641) have the highest values (p value = 0.0, ANOVA). The recombinational divergence (ϕ; p value = 0.0, ANOVA) and mutational divergence (θ; p value = 0.0, ANOVA) also varied among the species. The recombinational and mutational divergence values refer to the average number of recombination events and mutation events, respectively, per locus since coalescence of a pair of individuals (Lin & Kussell, 2019). The relative rate of recombination to mutation (ϕ/θ or γ/μ) ranged from 2.45 in *S. haemolyticus* A to 8.30 in *S.*

hominis (p value = 0.0, ANOVA). The mean fragment size ($\bar{f}$) of recombination events ranged from 433.44 bp in *S. lugdunensis* to 955.20 bp in *S. capitis* (p value = 3.36E-18, ANOVA). Finally, the recombination coverage (c), which represents the proportion of sites in the genome whose diversity was derived from recombination, were lowest in *S. warneri* A (10.91%) and highest in *S. lugdunensis* (30.72%), *S. simulans* A (30.16%) and *S. sciuri* (27.89%) (p value = 0.0, ANOVA). For comparison, *S. aureus* (n = 308 genomes from reference (Lin & Kussell, 2019)) showed lower values for diversity (0.015), recombination divergence (0.042), and γ/μ (1.0) and higher values for mutational divergence (0.042) and recombination coverage (36%). The mean fragment length of recombined DNA in CoNS is comparable to that found in *S. aureus* (550 bp).

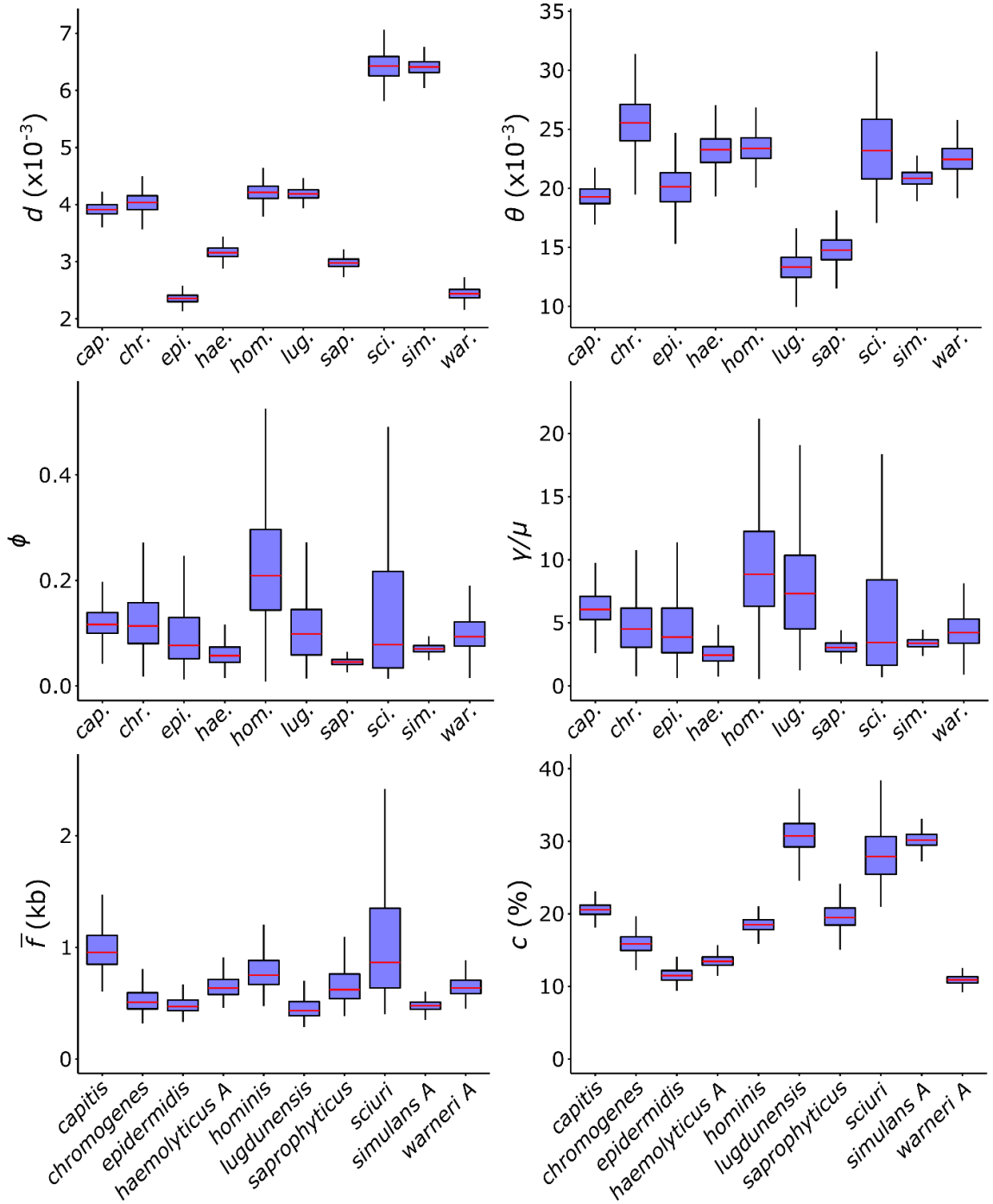

Figure 3.3 Characteristics of recombination in ten species in Figure 3.1 with the highest number of representative genomes. Recombination parameters were calculated by mcorr. Core genome alignment of each species was used as input in mcorr with 1000 bootstrapped replicates. d – diversity brought into the population by recombination and clonal diversity; θ – mutational divergence; ϕ – recombinational divergence; γ/μ relative rate of recombination to mutation (equivalent to ϕ/θ); $\bar{f}$ – mean fragment size of a recombination event; c – recombination coverage.

For each of the ten species, we also identified the specific genes that were frequently recombined (Figure S3.6 and Table S3.8). Five species were characterized as having a quarter of their pan-genome consisting of genes that have had a history of recent and ancestral recombination. These included *S. epidermidis* (2,077 genes, representing 25.03% of its pan-genome), *S. chromogenes* (1,025 genes; 27.63%), *S. hominis* (1,449; 24.55%), *S. haemolyticus* A (1,570; 23.73%) and *S. sciuri* (1,152; 21.42%). The other five species have fewer genes that had experienced recombination: *S. lugdunensis* (185 genes, representing 5.74% of its pan-genome), *S. warneri* A (345 genes; 8.41%), *S. saprophyticus* (642; 10.99%), *S. capitis* (711; 16.57%), and *S. simulans* A (773; 18.44%). Some of the most common functions of highly recombined genes across the ten species included those associated with response and/or resistance to antimicrobials (*bceB*, *bmrA*, *swrC*, *gyrA*, *lnrL*) and heavy metals (*arsB*), surface-associated and adherence-related proteins related to virulence and biofilm formation (*ebh*, *fbe*, *sdrF*, *sraP*, *sdrG*, *sasG*, *uafA*, *bca*, *sdrC*), type VII secretion system (*essC*), iron capture (*isdB*), recombination process (*bin3*, *hin*, *sbcC*, *xerC*) and metabolic processes that involve ion binding (*bioB*, *gltA*, *ipdC*, *lip*, *lipA*, *pgcA*, *phoB*, *ppdK*).

Highways of inter-species recombination in CoNS

Mapping the recent recombination events across the CoNS phylogeny allowed us to ask whether there was heterogeneity in the frequency of recent recombination between species. Using the fastGEAR results, we found that although inter-species recombination was widespread, species tended to recombine with certain partners that were members of their own species group (**Figure 3.4** and Table S3.8). This was particularly notable for species that comprise the species groups Epidermidis, Saprophyticus and Simulans. For example, the most

common species pairs with the highest frequencies of inter-species recombination were *S. capitis* and *S. caprae*, *S. haemolyticus* A and *S. haemolyticus* B, and *S. saprophyticus* and *S. edaphicus* (**Figure 3.4**). For each pair of species, we detected a total of 889, 616, and 381 recombination events, respectively. The species pairs with frequent recombination events traversing species group boundaries were *S. hominis* and *S. cohnii* B (128 recombination events), *S. capitis* and *S. arlettae* (31 recombination events), and *S. haemolyticus* A and *S. equorum* A (28 recombination events).

We also calculated the frequencies of DNA donations and receipts across the entire CoNS phylogeny using the fastGEAR results for recent recombination events (**Figure 3.4** and Figure S3.7). The most frequent donors to other species were *S. caprae* (971 donations) and two of the ten species with the highest number of representative genomes [*S. haemolyticus* A (693 donations) and *S. saprophyticus* (644 donations)]. When adjusted for the number of genomes per species, those with the highest number of donated DNA per genome were *S. devriesei* B (275 donations per genome), *S. haemolyticus* B (126 donations per genome) and *S. caprae* (88 donations per genome). In contrast, the ten major species with the highest number of representative genomes had fewer donations per genome (Figure S3.7). On the other hand, the most frequent recipients of recombined DNA were *S. capitis* (1,627 receipts), *S. haemolyticus* A (667 receipts) and *S. edaphicus* (506 receipts). When adjusted for the number of genomes per species, the most common recipients were *S. edaphicus* (506 receipts per genome), *S. caeli* (217 receipts per genome) and *S. simulans* B (187 receipts per genome). In contrast, those ten species with the highest number of representative genomes had lower number of receipts per genome.

Overall, these results show that certain CoNS species were frequent participants (either as donors or recipients) in inter-species recombination events.

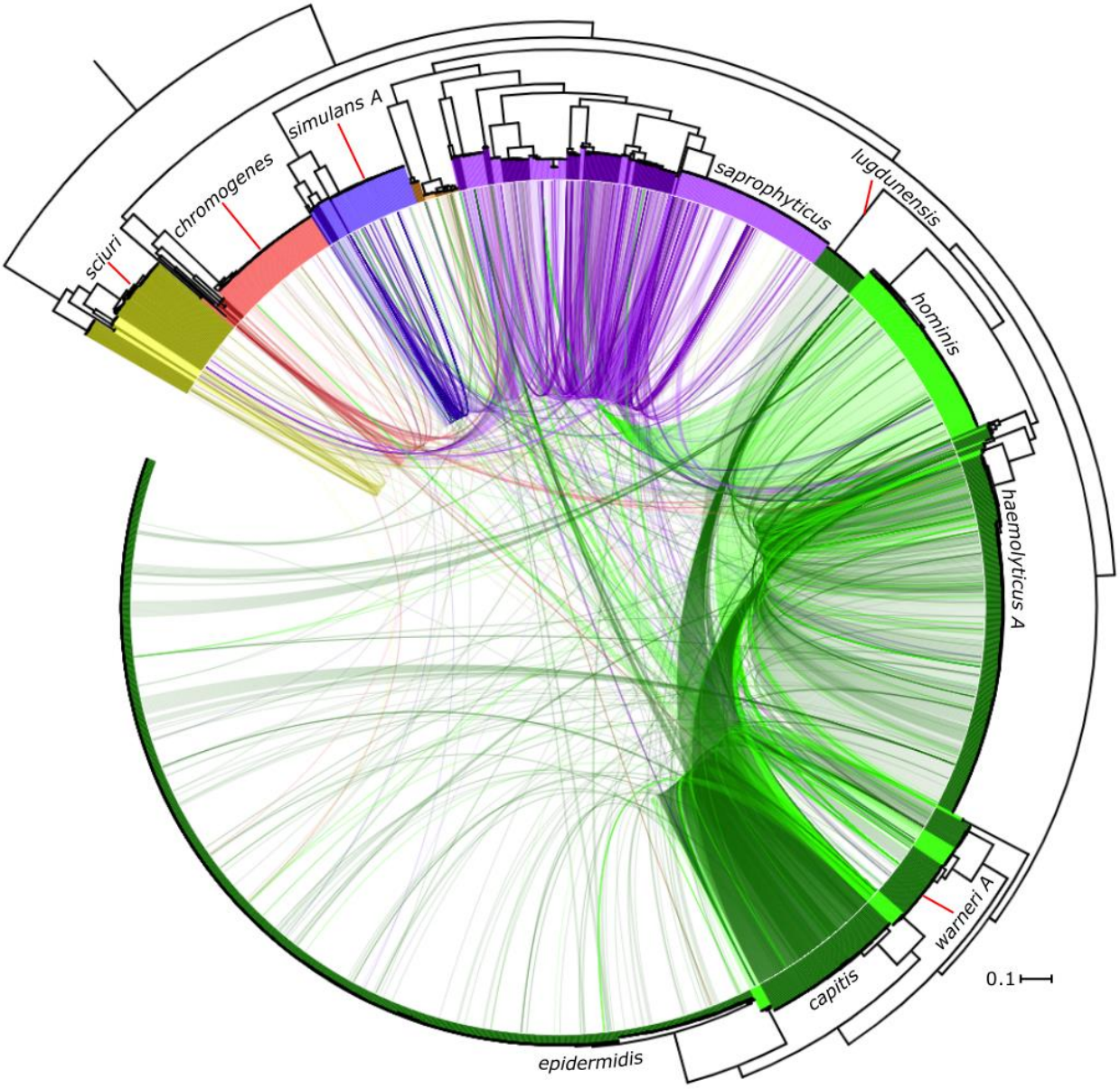

Figure 3.4 Highways of recombination between species. Results of the fastGEAR recombination detection program were mapped to the core genome phylogenetic tree. For each recombination event, the donor and recipient genomes were linked with a colored line. The color of the connecting line is based on the color of the recipient genome. The tree is identical to that in Figure 3.1. For visual clarity, the ten major species in the tree are labeled. Recombination between genomes within a species were not mapped onto the tree. Only recent recombinations (and not ancestral recombinations) were included.

To explore the contributions of ecology to the observed biases in inter-species recombination, we compared the frequencies of recombination between genomes that came from humans, live animals (domestic and wild), environmental (including plants) and animal-derived foods (consisting of milk, cheese, poultry, beef, pork, and fermented seafood and meats) (Figure S3.8). When the total number of recombination events was considered, frequent inter-species recombination involved human-associated strains, whether between strains from humans and animals (n = 1,793 recombinations), between strains from humans only (n = 1,154 recombinations) and between strains from humans and environment (n = 973 recombinations). However, this result is likely due to human-associated strains making up more than half the dataset, thus providing more potential recombination partners. When normalizing the ecological group sizes for more even comparisons (i.e., the number of observed recombination events relative to the total number of possible recombining pairs), we found that strains from animal-derived foods recombined with each other at a frequency of 0.0182, indicating that 1.82% of all possible recombinations were detected. Frequent recombination also occurred between strains from animals (n = 0.0108) and between strains from animal and animal-derived foods (n = 0.0080).

Discussion

CoNS species were historically defined by delimitation from the more pathogenic coagulase-producing *S. aureus* (Fairbrother, 1940). While coagulase production was established initially for diagnostic procedure, it gradually became a clinical classification scheme to exclude the purportedly non-pathogenic species of *Staphylococcus*. This consequently led to the lack of recognition for the clinical importance and poor taxonomic sampling of CoNS. As such, the

evolutionary relationships and genomic characteristics among CoNS species have not been confidently established.

Our study offers a robust and systematic analysis of CoNS evolutionary relationships from a genomic standpoint. There are two major findings in this study. First, our findings expand the known CoNS species to 55 based on core allele differences and accessory gene content variation. Our results showing the monophyly of the six species groups were consistent with the 4-locus sequence analysis of Lamers et al. (Lamers et al., 2012). Within each species group, genomic analysis allowed us to precisely infer the phylogenetic relationships of closely related species, further refining the current knowledge of the systematics and nomenclature for this important genus. Second, our study revealed the importance of recombination in CoNS evolution. The variable rates of recombination and biases in recombination partners imply that certain CoNS species function as hubs of gene flow and major reservoir of genetic diversity for the entire genus. We detected biases in recombination between certain species within and between species groups. The highest frequencies of inter-species receipt and donation of recombined DNA were detected in *S. capitis, S. caprae, S. hominis, S. haemolyticus* and members of the Saprophyticus species group. CoNS strains from animal-derived food items and live animals may act to facilitate recombination with strains from other sources, including humans. This has important implications to designing agricultural practices to reduce the emergence of high-risk bacterial clones and to managing foodborne disease transmission. These data support the concept that surveillance in both human and non-human sources could play a critical role in the early identification of bacterial clones that have acquired novel genetic material and/or phenotypes.

A pair of lineages or species that exchange DNA more frequently with each other than they do with others is said to be linked by a highway of gene sharing (Beiko et al., 2005). These highways have been previously reported in higher taxonomic groups, as in the case of phyla (e.g., horizontal gene transfer between thermophilic Thermotogales, Firmicutes and Archaea (Zhaxybayeva, Swithers, et al., 2009)) and genera (e.g., horizontal gene transfer between *Synechococcus* and *Prochlorococcus* (Zhaxybayeva, Doolittle, et al., 2009)). Frequent recombination between distantly related taxa may originate from co-occupancy of the same ecological niche over a long period of time or to syntrophic or other close symbiotic relationships (Arroyo et al., 2019; Russell & Cavanaugh, 2017; Zeng et al., 2017). At lower taxonomic levels, highways of recombination have also been previously reported. For example, in *Streptococcus pneumoniae,* frequent DNA donors and recipients were those strains that lack genes for capsule biosynthesis (Chewapreecha et al., 2014). Non-capsulated *S. pneumoniae* are considered to be less associated with disease than the capsulated strains and thus are excluded from the targets of currently available polysaccharide vaccines (Bradshaw & McDaniel, 2019; Daniels et al., 2016). Heterogeneity in recombination frequency between capsulated and non-capsulated *S. pneumoniae* may be a reflection of differential rates of response and adaptation to selection pressures from the environment or host (Chewapreecha et al., 2014). Highways of recombination have also been observed between different subspecies of *Salmonella enterica*, including those subspecies that are not often implicated in disease (Park & Andam, 2020). In these two pathogenic species, it appears then that the limited association with disease or clinical interventions may have facilitated greater opportunities for genetic recombination. For CoNS, inter-species highways of recombination are likely shaped by ecology, with animal hosts and

animal-derived foods acting as major hubs of genetic exchange. Our findings suggest that certain CoNS species can act as a major reservoir for many transferrable antimicrobial resistance determinants and variants of virulence genes. These results also suggest that CoNS species are an important reservoir of genetic diversity for the wider *Staphylococcus* genus, including *S. aureus* and other CoPS.

We recognize the limitations of this study. First, our reliance on publicly available genomes means that some CoNS species and geographical regions were disproportionately represented in our analysis. Rare species such as those consisting of less than ten genomes remain poorly characterized in terms of within-species genomic variation and population structure. Limited ecological data on many non-*aureus* species can obfuscate inferences of the contributions of shared or overlapping ecological niches to their evolution, which may also have clinical implications. Second, existing databases for querying antimicrobial resistance genes, virulence-associated genes and SCC*mec* types are limited to those built for the more intensively studied *S. aureus*. Hence, CoNS-specific genetic variants are likely to exist but remain invisible from current *in silico* detection methods and databases. This makes it problematic to track their prevalence and spread over the long term. Notwithstanding these limitations, we obtained sufficient representation of CoNS that can be used as basis for developing genus-wide databases for taxonomy and strain typing on non-*aureus* species in the future. Third, recombination between closely related strains with very similar DNA sequences is difficult to detect using current methods. Hence, our results are likely an underestimation of the true extent of recombination between CoNS found in nature. Last, inferring highways of recombination between ancestral lineages remains challenging. This is because of the uncertainty due to missing

lineages that can obscure true donors and those DNA segments that have been horizontally acquired and subsequently lost.

The results presented here spawn outstanding questions and open multiple avenues for future investigations. First, the ecology of specific CoNS species remains poorly studied. This includes the genetic basis of adaptation to different ecological niches within the human body and specific clinical presentations. Many CoNS species are also known to colonize a diverse array of animals (Bhargava & Zhang, 2012; Kern & Perreten, 2013; Mama et al., 2019), yet the range of host preferences and instances of switching between hosts (including animal to human) remain poorly studied. A systematic surveillance and sampling from both human and animals will greatly advance our understanding CoNS ecology. Morever, overlapping niches between CoNS and CoPS may reveal previously unrecognized ecological interactions, including recombination and horizontal gene transfer (Méric et al., 2015), that contribute to each other's success as commensals and pathogens. Second, our results on species-specific genomic variation will form the foundation of *in vitro* and *in vivo* investigations of clinically relevant phenotypes, such as adhesion, biofilm formation and internalization by and persistence in host cells. This knowledge will be particularly useful for developing effective therapeutic options for treatment of CoNS infections. Third, species-specific genetic variants can aid in precise laboratory detection in carriage and disease. Difficulties in species identification may lead to a lack of noted infections caused by *Staphylococcus* other than *S. aureus*. Hence, the real impact of CoNS species as etiological factors of human infections may remain underreported or overlooked. The implementation of reliable genetic methods in clinical practice will therefore improve the identification process and result in more precise diagnosis of staphylococcal infections.

In conclusion, our phylogenomic analysis of globally distributed CoNS provides important and robust basis for further study of their taxonomy, resistance mechanisms and pathogenic potential. Defining the evolutionary processes that contribute to CoNS diversity and speciation is the first step towards effective surveillance, diagnostics and control of CoNS infections.

References

Ahmed, A., Satti, L., Zaman, G., Gardezi, A., Sabir, N., & Khadim, M. T. (2019). Catheter related recurrent blood stream infection caused by linezolid-resistant, methicillin resistant *Staphylococcus haemolyticus*; an emerging super bug. *JPMA. The Journal of the Pakistan Medical Association*, *69*(2), 261–263.

Akiyama, H., Kanzaki, H., Tada, J., & Arata, J. (1998). Coagulase-negative staphylococci isolated from various skin lesions. *The Journal of Dermatology*, *25*(9), 563–568. https://doi.org/10.1111/j.1346-8138.1998.tb02459.x

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andreani, N. A., Hesse, E., & Vos, M. (2017). Prokaryote genome fluidity is dependent on effective population size. *The ISME Journal*, *11*(7), 1719–1721. https://doi.org/10.1038/ismej.2017.36

Arciola, C. R., Baldassarri, L., & Montanaro, L. (2001). Presence of icaA and icaD genes and slime production in a collection of staphylococcal strains from catheter-associated infections. *Journal of Clinical Microbiology*, *39*(6), 2151–2156. https://doi.org/10.1128/JCM.39.6.2151-2156.2001

Arroyo, F. A., Pawlowska, T. E., Choat, J. H., Clements, K. D., & Angert, E. R. (2019). Recombination contributes to population diversification in the polyploid intestinal symbiont *Epulopiscium* sp. Type B. *The ISME Journal*, *13*(4), 1084–1097. https://doi.org/10.1038/s41396-018-0339-y

Asante, J., Hetsa, B. A., Amoako, D. G., Abia, A. L. K., Bester, L. A., & Essack, S. Y. (2021). Multidrug-resistant coagulase-negative staphylococci isolated from bloodstream in the uMgungundlovu district of KwaZulu-Natal Province in South Africa: Emerging pathogens. *Antibiotics (Basel, Switzerland)*, *10*(2). https://doi.org/10.3390/antibiotics10020198

Becker, K., Heilmann, C., & Peters, G. (2014). Coagulase-negative staphylococci. *Clinical Microbiology Reviews*, *27*(4), 870–926. https://doi.org/10.1128/CMR.00109-13

Beiko, R. G., Harlow, T. J., & Ragan, M. A. (2005). Highways of gene sharing in prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(40), 14332–14337. https://doi.org/10.1073/pnas.0504068102

Bhargava, K., & Zhang, Y. (2012). Multidrug-resistant coagulase-negative Staphylococci in food animals. *Journal of Applied Microbiology*, *113*(5), 1027–1036. https://doi.org/10.1111/j.1365-2672.2012.05410.x

Bortolaia, V., Kaas, R. S., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R. L., Rebelo, A. R., Florensa, A. F., Fagelhauer, L., Chakraborty, T., Neumann, B., Werner, G., Bender, J. K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B. B., … Aarestrup, F. M. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. *The Journal of Antimicrobial Chemotherapy*, *75*(12), 3491–3500. https://doi.org/10.1093/jac/dkaa345

Both, A., Huang, J., Qi, M., Lausmann, C., Weißelberg, S., Büttner, H., Lezius, S., Failla, A. V., Christner, M., Stegger, M., Gehrke, T., Baig, S., Citak, M., Alawi, M., Aepfelbacher, M., & Rohde, H. (2021). Distinct clonal lineages and within-host diversification shape invasive *Staphylococcus epidermidis* populations. *PLoS Pathogens*, *17*(2), e1009304. https://doi.org/10.1371/journal.ppat.1009304

Bradshaw, J. L., & McDaniel, L. S. (2019). Selective pressure: Rise of the nonencapsulated pneumococcus. *PLoS Pathogens*, *15*(8), e1007911. https://doi.org/10.1371/journal.ppat.1007911

Carattoli, A., & Hasman, H. (2020). PlasmidFinder and in silico pMLST: Identification and typing of plasmid replicons in whole-genome sequencing (WGS). *Methods in Molecular Biology (Clifton, N.J.)*, *2075*, 285–294. https://doi.org/10.1007/978-1-4939-9877-7_20

Chewapreecha, C., Harris, S. R., Croucher, N. J., Turner, C., Marttinen, P., Cheng, L., Pessia, A., Aanensen, D. M., Mather, A. E., Page, A. J., Salter, S. J., Harris, D., Nosten, F., Goldblatt, D., Corander, J., Parkhill, J., Turner, P., & Bentley, S. D. (2014). Dense genomic sampling identifies highways of pneumococcal recombination. *Nature Genetics*, *46*(3), 305–309. https://doi.org/10.1038/ng.2895

Crosby, H. A., Kwiecinski, J., & Horswill, A. R. (2016). *Staphylococcus aureus* aggregation and coagulation mechanisms, and their function in host-pathogen interactions. *Advances in Applied Microbiology*, *96*, 1–41. https://doi.org/10.1016/bs.aambs.2016.07.018

Cue, D., Lei, M. G., & Lee, C. Y. (2012). Genetic regulation of the intercellular adhesion locus in staphylococci. *Frontiers in Cellular and Infection Microbiology*, *2*, 38. https://doi.org/10.3389/fcimb.2012.00038

Daniels, C. C., Rogers, P. D., & Shelton, C. M. (2016). A review of pneumococcal vaccines: Current polysaccharide vaccine recommendations and future protein antigens. *The Journal of Pediatric Pharmacology and Therapeutics*, *21*(1), 27–35. https://doi.org/10.5863/1551-6776-21.1.27

Driebe, E. M., Sahl, J. W., Roe, C., Bowers, J. R., Schupp, J. M., Gillece, J. D., Kelley, E., Price, L. B., Pearson, T. R., Hepp, C. M., Brzoska, P. M., Cummings, C. A., Furtado, M. R., Andersen, P. S., Stegger, M., Engelthaler, D. M., & Keim, P. S. (2015). Using whole genome analysis to examine recombination across diverse sequence types of *Staphylococcus aureus*. *PloS One*, *10*(7), e0130955. https://doi.org/10.1371/journal.pone.0130955

Fairbrother, R. (1940). Coagulase production as a criterion for the classification of the staphylococci. *Journal of Pathology and Bacteriology*, *50*, 83–88.

Fishovitz, J., Hermoso, J. A., Chang, M., & Mobashery, S. (2014). Penicillin-binding protein 2a of methicillin-resistant *Staphylococcus aureus*. *IUBMB Life*, *66*(8), 572–577. https://doi.org/10.1002/iub.1289

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics (Oxford, England)*, *28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Fux, C. A., Uehlinger, D., Bodmer, T., Droz, S., Zellweger, C., & Mühlemann, K. (2005). Dynamics of hemodialysis catheter colonization by coagulase-negative staphylococci. *Infection Control and Hospital Epidemiology*, *26*(6), 567–574. https://doi.org/10.1086/502586

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Haroche, J., Morvan, A., Davi, M., Allignet, J., Bimet, F., & El Solh, N. (2003). Clonal diversity among streptogramin A-resistant *Staphylococcus aureus* isolates collected in French hospitals. *Journal of Clinical Microbiology*, *41*(2), 586–591. https://doi.org/10.1128/jcm.41.2.586-591.2003

Hassler, C., Nitzsche, S., Iversen, C., Zweifel, C., & Stephan, R. (2008). Characteristics of *Staphylococcus hyicus* strains isolated from pig carcasses in two different slaughterhouses. *Meat Science*, *80*(2), 505–510. https://doi.org/10.1016/j.meatsci.2008.02.001

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC). (2009). Classification of staphylococcal cassette chromosome mec (SCCmec): Guidelines for reporting novel SCCmec elements. *Antimicrobial Agents and Chemotherapy*, *53*(12), 4961–4967. https://doi.org/10.1128/AAC.00579-09

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 5114. https://doi.org/10.1038/s41467-018-07641-9

Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. *Methods in Molecular Biology (Clifton, N.J.)*, *537*, 39–64. https://doi.org/10.1007/978-1-59745-251-9_3

Kaya, H., Hasman, H., Larsen, J., Stegger, M., Johannesen, T. B., Allesøe, R. L., Lemvigh, C. K., Aarestrup, F. M., Lund, O., & Larsen, A. R. (2018). SCCmecFinder, a web-based tool for typing of Staphylococcal Cassette Chromosome mec in *Staphylococcus aureus* using whole-genome sequence data. *MSphere*, *3*(1). https://doi.org/10.1128/mSphere.00612-17

Kern, A., & Perreten, V. (2013). Clinical and molecular features of methicillin-resistant, coagulase-negative staphylococci of pets and horses. *The Journal of Antimicrobial Chemotherapy*, *68*(6), 1256–1266. https://doi.org/10.1093/jac/dkt020

Knudsen, G. M., Olsen, J. E., Aabo, S., Barrow, P., Rychlik, I., & Thomsen, L. E. (2013). ClpP deletion causes attenuation of *Salmonella* Typhimurium virulence through mis-regulation of RpoS and indirect control of CsrA and the SPI genes. *Microbiology (Reading, England)*, *159*(Pt 7), 1497–1509. https://doi.org/10.1099/mic.0.065797-0

Lamers, R. P., Muthukrishnan, G., Castoe, T. A., Tafur, S., Cole, A. M., & Parkinson, C. L. (2012). Phylogenetic relationships among *Staphylococcus* species and refinement of cluster groups based on multilocus data. *BMC Evolutionary Biology*, *12*, 171. https://doi.org/10.1186/1471-2148-12-171

Lawal, O. U., Fraqueza, M. J., Bouchami, O., Worning, P., Bartels, M. D., Gonçalves, M. L., Paixão, P., Gonçalves, E., Toscano, C., Empel, J., Urbaś, M., Domínguez, M. A., Westh, H., de Lencastre, H., & Miragaia, M. (2021). Foodborne origin and local and global spread of *Staphylococcus saprophyticus* causing human urinary tract infections. *Emerging Infectious Diseases*, *27*(3), 880–893. https://doi.org/10.3201/eid2703.200852

Lebeurre, J., Dahyot, S., Diene, S., Paulay, A., Aubourg, M., Argemi, X., Giard, J.-C., Tournier, I., François, P., & Pestel-Caron, M. (2019). Comparative genome analysis of *Staphylococcus lugdunensis* shows clonal complex-dependent diversity of the putative virulence factor, ess/type VII locus. *Frontiers in Microbiology*, *10*, 2479. https://doi.org/10.3389/fmicb.2019.02479

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, *47*(W1), W256–W259. https://doi.org/10.1093/nar/gkz239

Lin, M., & Kussell, E. (2019). Inferring bacterial recombination rates from large-scale sequencing datasets. *Nature Methods*, *16*(2), 199–204. https://doi.org/10.1038/s41592-018-0293-7

Liu, B., Zheng, D., Jin, Q., Chen, L., & Yang, J. (2019). VFDB 2019: A comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Research*, *47*(D1), D687–D692. https://doi.org/10.1093/nar/gky1080

Luong, T. T., & Lee, C. Y. (2002). Overproduction of type 8 capsular polysaccharide augments *Staphylococcus aureus* virulence. *Infection and Immunity*, *70*(7), 3389–3395. https://doi.org/10.1128/iai.70.7.3389-3395.2002

Mama, O. M., Ruiz-Ripa, L., Lozano, C., González-Barrio, D., Ruiz-Fons, J. F., & Torres, C. (2019). High diversity of coagulase negative staphylococci species in wild boars, with low antimicrobial resistance rates but detection of relevant resistance genes. *Comparative Immunology, Microbiology and Infectious Diseases*, *64*, 125–129. https://doi.org/10.1016/j.cimid.2019.03.006

May, L., Klein, E. Y., Rothman, R. E., & Laxminarayan, R. (2014). Trends in antibiotic resistance in coagulase-negative staphylococci in the United States, 1999 to 2012. *Antimicrobial Agents and Chemotherapy*, *58*(3), 1404–1409. https://doi.org/10.1128/AAC.01908-13

McInerney, J. O., McNally, A., & O'Connell, M. J. (2017). Why prokaryotes have pangenomes. *Nature Microbiology*, *2*, 17040. https://doi.org/10.1038/nmicrobiol.2017.40

Medini, D., Donati, C., Tettelin, H., Masignani, V., & Rappuoli, R. (2005). The microbial pan-genome. *Current Opinion in Genetics & Development*, *15*(6), 589–594. https://doi.org/10.1016/j.gde.2005.09.006

Méric, G., Mageiros, L., Pensar, J., Laabei, M., Yahara, K., Pascoe, B., Kittiwan, N., Tadee, P., Post, V., Lamble, S., Bowden, R., Bray, J. E., Morgenstern, M., Jolley, K. A., Maiden, M. C. J., Feil, E. J., Didelot, X., Miragaia, M., de Lencastre, H., … Sheppard, S. K. (2018). Disease-associated genotypes of the commensal skin bacterium *Staphylococcus epidermidis*. *Nature Communications*, *9*(1), 5034. https://doi.org/10.1038/s41467-018-07368-7

Méric, G., Miragaia, M., de Been, M., Yahara, K., Pascoe, B., Mageiros, L., Mikhail, J., Harris, L. G., Wilkinson, T. S., Rolo, J., Lamble, S., Bray, J. E., Jolley, K. A., Hanage, W. P., Bowden, R., Maiden, M. C. J., Mack, D., de Lencastre, H., Feil, E. J., … Sheppard, S. K. (2015). Ecological overlap and horizontal gene transfer in *Staphylococcus aureus* and *Staphylococcus epidermidis*. *Genome Biology and Evolution*, *7*(5), 1313–1328. https://doi.org/10.1093/gbe/evv066

Michalik, M., Samet, A., Podbielska-Kubera, A., Savini, V., Międzobrodzki, J., & Kosecka-Strojek, M. (2020). Coagulase-negative staphylococci (CoNS) as a significant etiological factor of laryngological infections: A review. *Annals of Clinical Microbiology and Antimicrobials*, *19*(1), 26. https://doi.org/10.1186/s12941-020-00367-x

Michel, A., Agerer, F., Hauck, C. R., Herrmann, M., Ullrich, J., Hacker, J., & Ohlsen, K. (2006). Global regulatory impact of ClpP protease of *Staphylococcus aureus* on regulons involved in virulence, oxidative stress response, autolysis, and DNA repair. *Journal of Bacteriology*, *188*(16), 5783–5796. https://doi.org/10.1128/JB.00074-06

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mostowy, R., Croucher, N. J., Andam, C. P., Corander, J., Hanage, W. P., & Marttinen, P. (2017). Efficient Inference of recent and ancestral recombination within bacterial populations. *Molecular Biology and Evolution*, *34*(5), 1167–1182. https://doi.org/10.1093/molbev/msx066

Nakamura, Y., Oscherwitz, J., Cease, K. B., Chan, S. M., Muñoz-Planillo, R., Hasegawa, M., Villaruz, A. E., Cheung, G. Y. C., McGavin, M. J., Travers, J. B., Otto, M., Inohara, N., & Núñez, G. (2013). *Staphylococcus* δ-toxin induces allergic skin disease by activating mast cells. *Nature*, *503*(7476), 397–401. https://doi.org/10.1038/nature12655

Non, L. R., & Santos, C. A. Q. (2017). The occurrence of infective endocarditis with *Staphylococcus lugdunensis* bacteremia: A retrospective cohort study and systematic review. *The Journal of Infection*, *74*(2), 179–186. https://doi.org/10.1016/j.jinf.2016.10.003

O'Riordan, K., & Lee, J. C. (2004). *Staphylococcus aureus* capsular polysaccharides. *Clinical Microbiology Reviews*, *17*(1), 218–234. https://doi.org/10.1128/cmr.17.1.218-234.2004

Park, C. J., & Andam, C. P. (2020). Distinct but intertwined evolutionary histories of multiple *Salmonella enterica* subspecies. *MSystems*, *5*(1). https://doi.org/10.1128/mSystems.00515-19

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Pedroso, S. H. S. P., Sandes, S. H. C., Filho, R. A. T., Nunes, A. C., Serufo, J. C., Farias, L. M., Carvalho, M. A. R., Bomfim, M. R. Q., & Santos, S. G. (2018). Coagulase-negative staphylococci isolated from human bloodstream infections showed multidrug resistance profile. *Microbial Drug Resistance (Larchmont, N.Y.)*, *24*(5), 635–647. https://doi.org/10.1089/mdr.2017.0309

Rolo, J., Worning, P., Nielsen, J. B., Bowden, R., Bouchami, O., Damborg, P., Guardabassi, L., Perreten, V., Tomasz, A., Westh, H., de Lencastre, H., & Miragaia, M. (2017). Evolutionary origin of the staphylococcal cassette chromosome mec (SCCmec). *Antimicrobial Agents and Chemotherapy*, *61*(6). https://doi.org/10.1128/AAC.02302-16

Russell, S. L., & Cavanaugh, C. M. (2017). Intrahost genetic diversity of bacterial symbionts exhibits evidence of mixed infections and recombinant haplotypes. *Molecular Biology and Evolution*, *34*(11), 2747–2761. https://doi.org/10.1093/molbev/msx188

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Spoor, L. E., Richardson, E., Richards, A. C., Wilson, G. J., Mendonca, C., Gupta, R. K., McAdam, P. R., Nutbeam-Tuffs, S., Black, N. S., O'Gara, J. P., Lee, C. Y., Corander, J., & Ross Fitzgerald, J. (2015). Recombination-mediated remodelling of host-pathogen interactions during *Staphylococcus aureus* niche adaptation. *Microbial Genomics*, *1*(4), e000036. https://doi.org/10.1099/mgen.0.000036

Stenmark, B., Hellmark, B., & Söderquist, B. (2019). Genomic analysis of *Staphylococcus capitis* isolated from blood cultures in neonates at a neonatal intensive care unit in Sweden. *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology*, *38*(11), 2069–2075. https://doi.org/10.1007/s10096-019-03647-3

Taponen, S., Supré, K., Piessens, V., Van Coillie, E., De Vliegher, S., & Koort, J. M. K. (2012). *Staphylococcus agnetis* sp. Nov., a coagulase-variable species from bovine subclinical and mild clinical mastitis. *International Journal of Systematic and Evolutionary Microbiology*, *62*(Pt 1), 61–65. https://doi.org/10.1099/ijs.0.028365-0

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *American Mathematical Society: Lectures on Mathematics in the Life Sciences*, *17*, 57–86.

Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J. A., Gladstone, R. A., Lo, S., Beaudoin, C., Floto, R. A., Frost, S. D. W., Corander, J., Bentley, S. D., & Parkhill, J. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biology*, *21*(1), 180. https://doi.org/10.1186/s13059-020-02090-4

Vandenesch, F., Lebeau, C., Bes, M., Lina, G., Lina, B., Greenland, T., Benito, Y., Brun, Y., Fleurette, J., & Etienne, J. (1994). Clotting activity in *Staphylococcus schleiferi* subspecies from human patients. *Journal of Clinical Microbiology*, *32*(2), 388–392. https://doi.org/10.1128/JCM.32.2.388-392.1994

von Eiff, C., Arciola, C. R., Montanaro, L., Becker, K., & Campoccia, D. (2006). Emerging *Staphylococcus* species as new pathogens in implant infections. *The International Journal of Artificial Organs*, *29*(4), 360–367. https://doi.org/10.1177/039139880602900405

Widerström, M., Wiström, J., Ferry, S., Karlsson, C., & Monsen, T. (2007). Molecular epidemiology of *Staphylococcus saprophyticus* isolated from women with uncomplicated

community-acquired urinary tract infection. *Journal of Clinical Microbiology*, *45*(5), 1561–1564. https://doi.org/10.1128/JCM.02071-06

Wirth, T., Bergot, M., Rasigade, J.-P., Pichon, B., Barbier, M., Martins-Simoes, P., Jacob, L., Pike, R., Tissieres, P., Picaud, J.-C., Kearns, A., Supply, P., Butin, M., Laurent, F., International Consortium for *Staphylococcus capitis* neonatal sepsis, & ESGS Study Group of ESCMID. (2020). Niche specialization and spread of *Staphylococcus capitis* involved in neonatal sepsis. *Nature Microbiology*, *5*(5), 735–745. https://doi.org/10.1038/s41564-020-0676-2

Zeng, Q., Wang, J., Bertels, F., Giordano, P. R., Chilvers, M. I., Huntley, R. B., Vargas, J. M., Sundin, G. W., Jacobs, J. L., & Yang, C.-H. (2017). Recombination of virulence genes in divergent *Acidovorax avenae* strains that infect a common host. *Molecular Plant-Microbe Interactions: MPMI*, *30*(10), 813–828. https://doi.org/10.1094/MPMI-06-17-0151-R

Zhao, B.-B., Li, X.-H., Zeng, Y.-L., & Lu, Y.-J. (2016). ClpP-deletion impairs the virulence of *Legionella pneumophila* and the optimal translocation of effector proteins. *BMC Microbiology*, *16*(1), 174. https://doi.org/10.1186/s12866-016-0790-8

Zhaxybayeva, O., Doolittle, W. F., Papke, R. T., & Gogarten, J. P. (2009). Intertwined evolutionary histories of marine *Synechococcus* and *Prochlorococcus marinus*. *Genome Biology and Evolution*, *1*, 325–339. https://doi.org/10.1093/gbe/evp032

Zhaxybayeva, O., Swithers, K. S., Lapierre, P., Fournier, G. P., Bickhart, D. M., DeBoy, R. T., Nelson, K. E., Nesbø, C. L., Doolittle, W. F., Gogarten, J. P., & Noll, K. M. (2009). On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(14), 5865–5870. https://doi.org/10.1073/pnas.0901260106

Zheng, J., Wu, Y., Lin, Z., Wang, G., Jiang, S., Sun, X., Tu, H., Yu, Z., & Qu, D. (2020). ClpP participates in stress tolerance, biofilm formation, antimicrobial tolerance, and virulence of *Enterococcus faecalis*. *BMC Microbiology*, *20*(1), 30. https://doi.org/10.1186/s12866-020-1719-9

Zheng, Y., He, L., Asiamah, T. K., & Otto, M. (2018). Colonization of medical devices by staphylococci. *Environmental Microbiology*, *20*(9), 3141–3153. https://doi.org/10.1111/1462-2920.14129

Zingg, W., Demartines, N., Imhof, A., Senn, G., & Ruef, C. (2009). Rapid colonization with methicillin-resistant coagulase-negative staphylococci after surgery. *World Journal of Surgery*, *33*(10), 2058–2062. https://doi.org/10.1007/s00268-009-0167-3

Appendix 3

Due to size, Tables S3.1, S3.2, S3.4-S3.6, and S3.8 were excluded. Requests for these tables can be made to joshuasmithmicro@gmail.com.

Table S3.1 Accession numbers, relevant metadata and genomic characteristics of 1,876 CoNS strains used in this study

Table S3.2 Pan-genome of the entire dataset (n = 1,876 strains) estimated using Panaroo

Table S3.3 List of strains showing discrepancies in the classification of 16 CoNS strains. Second column shows the strain names in NCBI. The third column shows their closest relative based on phylogenetic placement and fastANI results.

Accession_ID	original_sp	reclassified_sp
GCF_000494875.1	pasteuri	warneri_A
GCF_001074355.1	saprophyticus	kloosii_B
GCF_001940625.1	epidermidis	warneri_B
GCF_002018435.1	fleurettii	fleurettii_A
GCF_002276895.1	pasteuri	warneri_A
GCF_003850005.1	xylosus	pseudoxylosus
GCF_003857215.1	warneri	pasteuri
GCF_003968855.1	schleiferi	sciuri
GCF_003968865.1	fleurettii	fleurettii_B
GCF_003970495.1	saccharolyticus	pasteuri
GCF_004569575.1	saprophyticus	xylosus_C
GCF_007666705.1	warneri	pasteuri
GCF_007667785.1	warneri	pasteuri
GCF_007677205.1	haemolyticus	petrasii_C
GCF_009730555.1	fleurettii	fleurettii_B
GCF_900458505.1	fleurettii	fleurettii_A

Table S3.4 Pan-genome of each of the ten major CoNS species estimated using Panaroo. Individual species are shown in different tabs.

Table S3.5 Presence-absence matrix showing the distribution of horizontally acquired antimicrobial resistance genes in 1,876 CoNS genomes

Table S3.6 Presence-absence matrix showing the distribution of virulence genes in 1,876 CoNS genomes

Table S3.7 Results of the mcorr calculation of the six evolutionary and recombination parameters for each of the ten major CoNS species. Bootstrapping results are shown in the second tab.

	sp.	**d_sample**	**theta_pool**	**phi_pool**	**ratio**	**fbar**	**c**
value	capitis	0.00392	0.0193232	0.117893	6.101	960.7	0.205
	chromogenes	0.004038	0.0256646	0.105245	4.101	502.8	0.158
	epidermidis	0.002361	0.0199369	0.075217	3.773	479.7	0.117
	haemolyticus_A	0.003159	0.0229818	0.057222	2.49	654.1	0.136
	hominis	0.004229	0.023556	0.189692	8.053	739.5	0.184
	lugdunensis	0.00419	0.0133794	0.094046	7.029	438.8	0.307
	saprophyticus	0.002983	0.0148039	0.045296	3.06	624.4	0.195
	sciuri	0.006413	0.0201965	0.077223	3.823	1587	0.316
	simulans_A	0.006403	0.0208715	0.070247	3.366	475.5	0.301
	warneri_A	0.002436	0.0223521	0.091437	4.091	648.5	0.109
	sp.	**d_sample**	**theta_pool**	**phi_pool**	**ratio**	**fbar**	**c**
mean	capitis	0.003916	0.01928047	0.121728	6.273	1026	0.206
	chromogenes	0.004036	0.0256013	0.128079	5.037	539.9	0.159
	epidermidis	0.002356	0.02010305	0.102687	4.974	488.3	0.116
	haemolyticus_A	0.003161	0.02322477	0.068669	2.901	657.4	0.135
	hominis	0.004218	0.02335649	0.21006	8.878	2596	0.186
	lugdunensis	0.004189	0.01328806	0.109447	8.012	481.8	0.31
	saprophyticus	0.002983	0.01478214	0.045654	3.098	684.8	0.197
	sciuri	0.006429	0.02329592	0.130106	5.137	4102	0.282
	simulans_A	0.006407	0.02085457	0.070849	3.398	481	0.302
	warneri_A	0.00244	0.02247353	0.104401	4.59	655.1	0.109
	sp.	**d_sample**	**theta_pool**	**phi_pool**	**ratio**	**fbar**	**c**
stdev	capitis	0.000117	0.00092395	0.032229	1.467	406	0.01
	chromogenes	0.000181	0.00224692	0.07144	2.803	141.6	0.015
	epidermidis	8.34E-05	0.00183711	0.078692	3.575	82.32	0.01

	haemolyticus_A	0.00011	0.00156056	0.045621	1.742	120.2	0.009
	hominis	0.000157	0.00147039	0.093908	3.758	23018	0.013
	lugdunensis	0.000104	0.00131322	0.066987	4.547	209.9	0.027
	saprophyticus	9.07E-05	0.00125918	0.008076	0.545	235.3	0.017
	sciuri	0.000248	0.00303898	0.116427	4.198	27273	0.034
	simulans_A	0.00014	0.00073797	0.009124	0.424	50.48	0.011
	warneri_A	0.000106	0.00131901	0.047556	1.911	102.2	0.007
	sp.	**d_sample**	**theta_pool**	**phi_pool**	**ratio**	**fbar**	**c**
median	capitis	0.003913	0.01926717	0.116599	6.062	955.2	0.205
	chromogenes	0.004037	0.02555736	0.113813	4.504	506.7	0.159
	epidermidis	0.002356	0.02013238	0.076703	3.87	468.9	0.115
	haemolyticus_A	0.003157	0.02327585	0.057458	2.449	634.1	0.135
	hominis	0.004215	0.0233844	0.199507	8.296	750.9	0.185
	lugdunensis	0.004189	0.0133297	0.098698	7.313	433.4	0.307
	saprophyticus	0.002976	0.01475701	0.045354	3.045	620.5	0.195
	sciuri	0.006427	0.02318399	0.07844	3.456	863.7	0.279
	simulans_A	0.00641	0.02085181	0.07018	3.377	478	0.302
	warneri_A	0.002437	0.02244449	0.093542	4.229	636	0.109
	sp.	**d_sample**	**theta_pool**	**phi_pool**	**ratio**	**fbar**	**c**
lower_bound_5%	capitis	0.003909	0.0192232	0.11973	6.182	1001	0.205
	chromogenes	0.004025	0.02546204	0.123651	4.863	531.1	0.158
	epidermidis	0.002351	0.01998919	0.09781	4.753	483.2	0.115
	haemolyticus_A	0.003155	0.02312805	0.065841	2.793	649.9	0.134
	hominis	0.004208	0.02326536	0.20424	8.645	1169	0.185
	lugdunensis	0.004183	0.01320667	0.105295	7.73	468.8	0.309
	saprophyticus	0.002977	0.01470409	0.045154	3.064	670.2	0.195
	sciuri	0.006414	0.02310757	0.12289	4.877	2412	0.28
	simulans_A	0.006399	0.02080883	0.070283	3.372	477.9	0.301
	warneri_A	0.002434	0.02239178	0.101453	4.472	648.8	0.109
	sp.	**d_sample**	**theta_pool**	**phi_pool**	**ratio**	**fbar**	**c**
upper_bound_95%	capitis	0.003923	0.01933774	0.123725	6.364	1052	0.207
	chromogenes	0.004047	0.02574056	0.132507	5.211	548.6	0.16
	epidermidis	0.002361	0.02021692	0.107564	5.196	493.4	0.116
	haemolyticus_A	0.003168	0.0233215	0.071496	3.009	664.8	0.135
	hominis	0.004228	0.02344763	0.215881	9.111	4022	0.187
	lugdunensis	0.004196	0.01336945	0.113599	8.294	494.8	0.312
	saprophyticus	0.002988	0.01486018	0.046155	3.132	699.4	0.198
	sciuri	0.006444	0.02348427	0.137322	5.398	5793	0.284
	simulans_A	0.006416	0.02090031	0.071414	3.424	484.1	0.303
	warneri_A	0.002447	0.02255528	0.107348	4.709	661.4	0.11

Table S3.8 Number of recent and ancestral recombination events calculated by fastGEAR for each of the ten major species. Individual species are shown in different tabs.

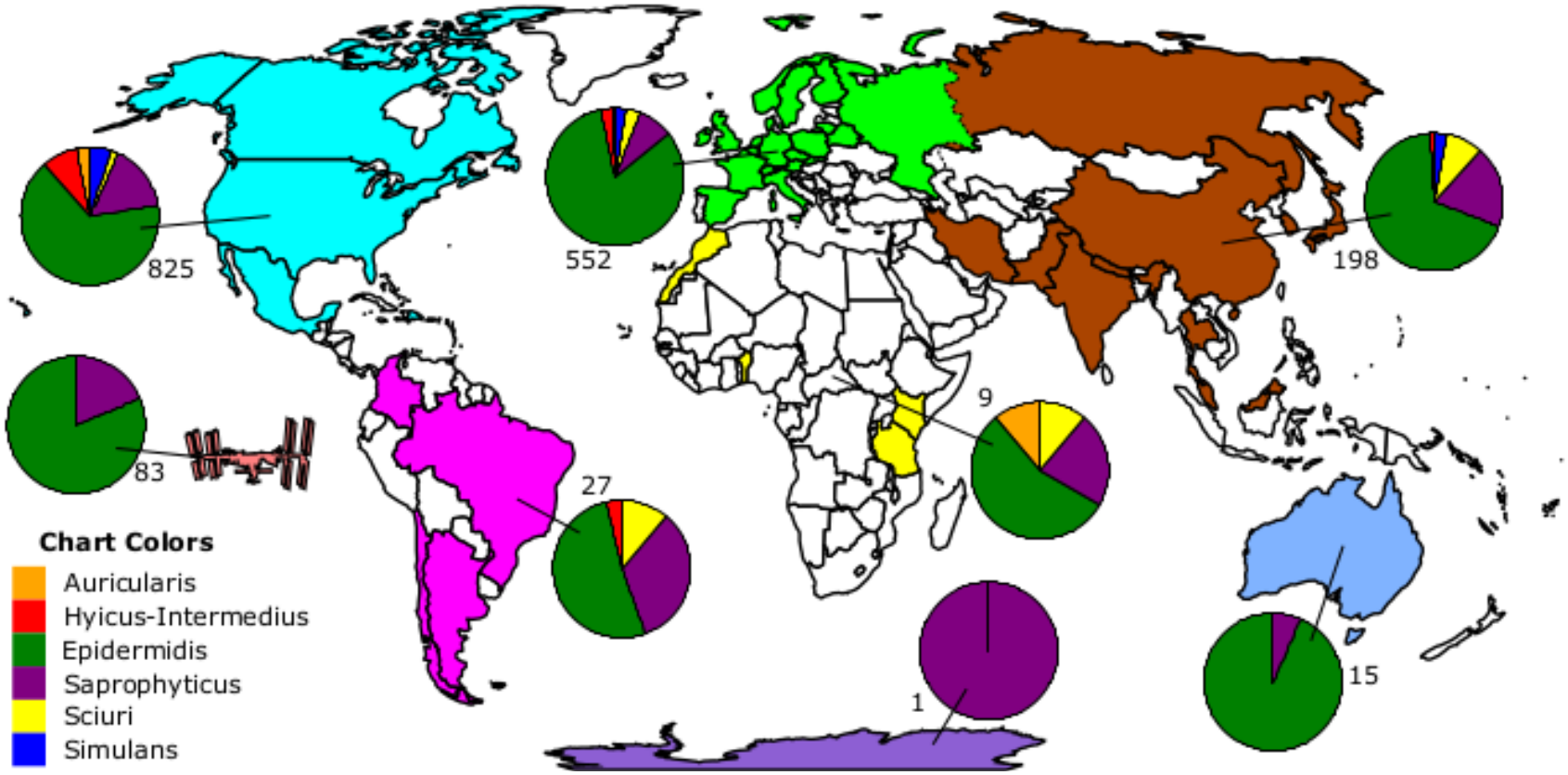

Figure S3.1 Global map showing the distribution of the 1,876 CoNS genomes. Number next to the pie charts show the number of genomes in each region. Colors in pie chart represent the six species groups. Note that Russia was split into east and west, with two genomes sampled from each region.

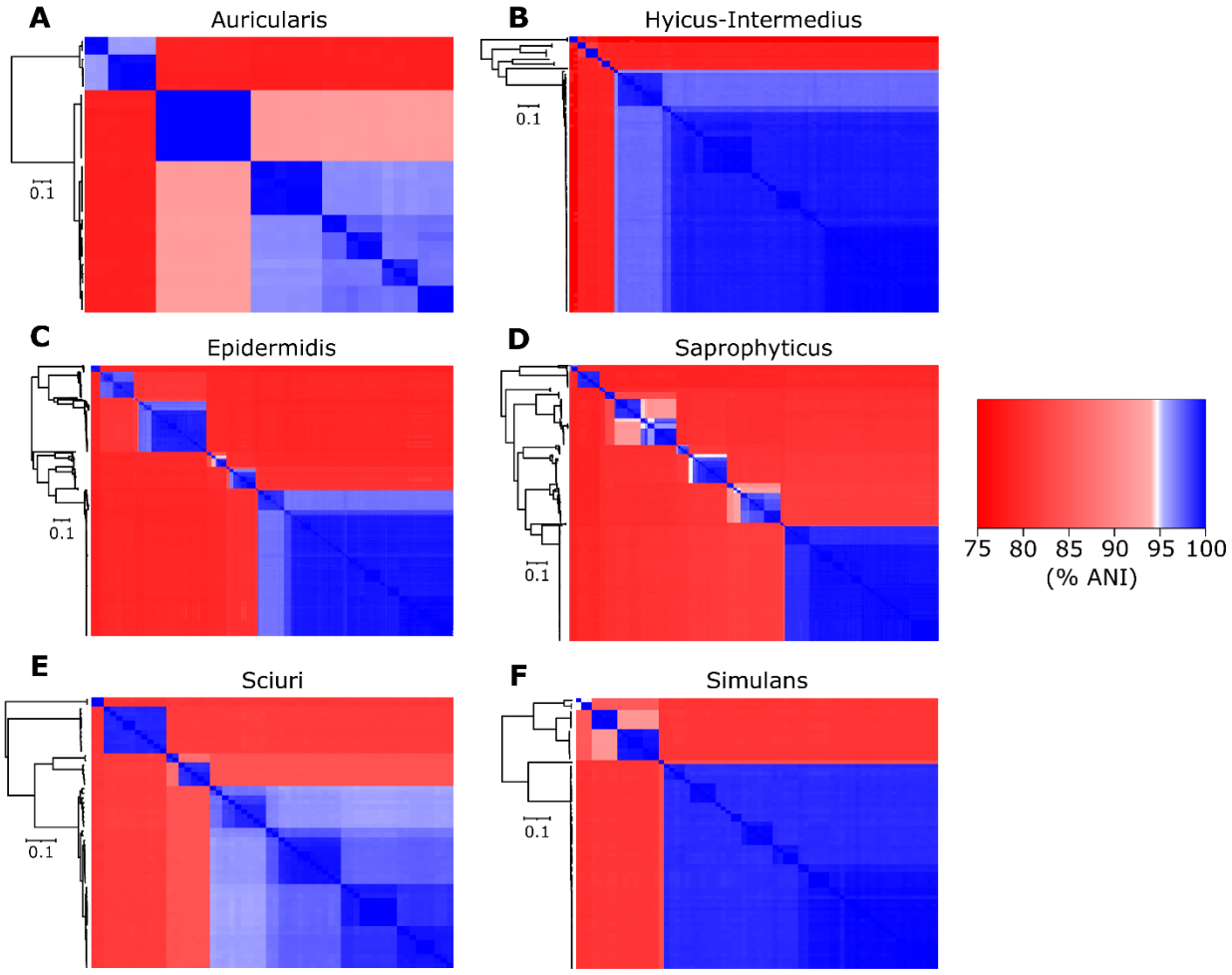

Figure S3.2 Pairwise fastANI values of genomes within each of the six species groups

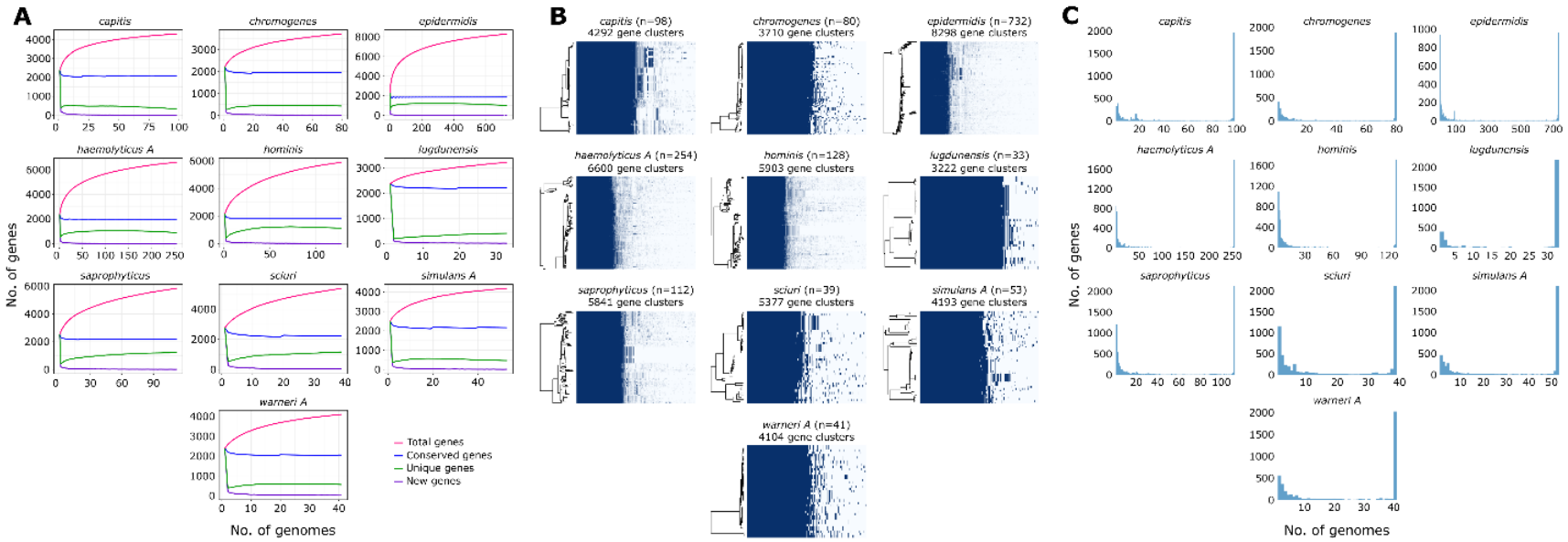

Figure S3.3 Panaroo plots showing the pan-genome characteristics of the ten major species

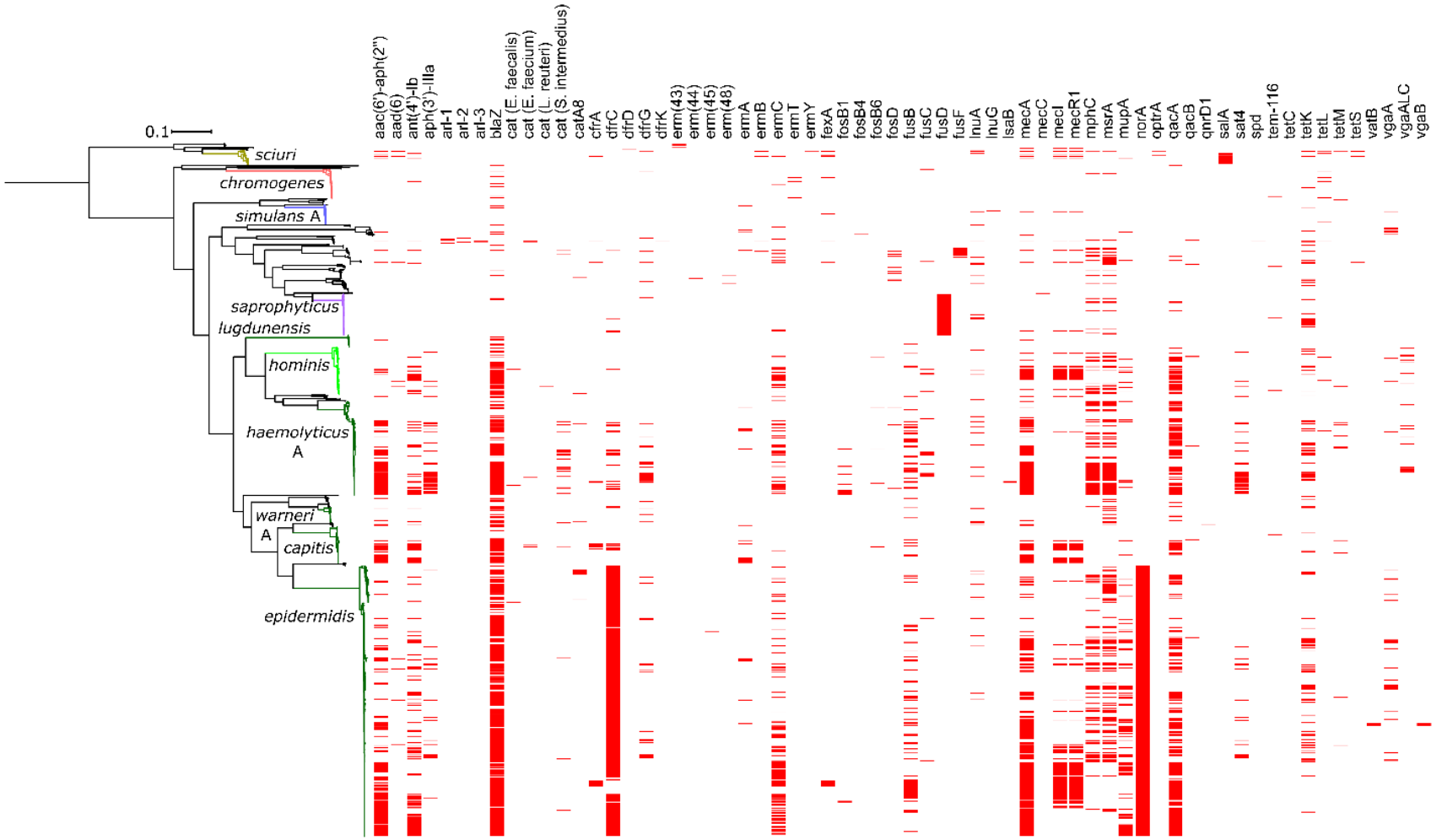

Figure S3.4 Phylogenetic distribution of horizontally acquired antimicrobial resistance genes in CoNS 1,876 genomes

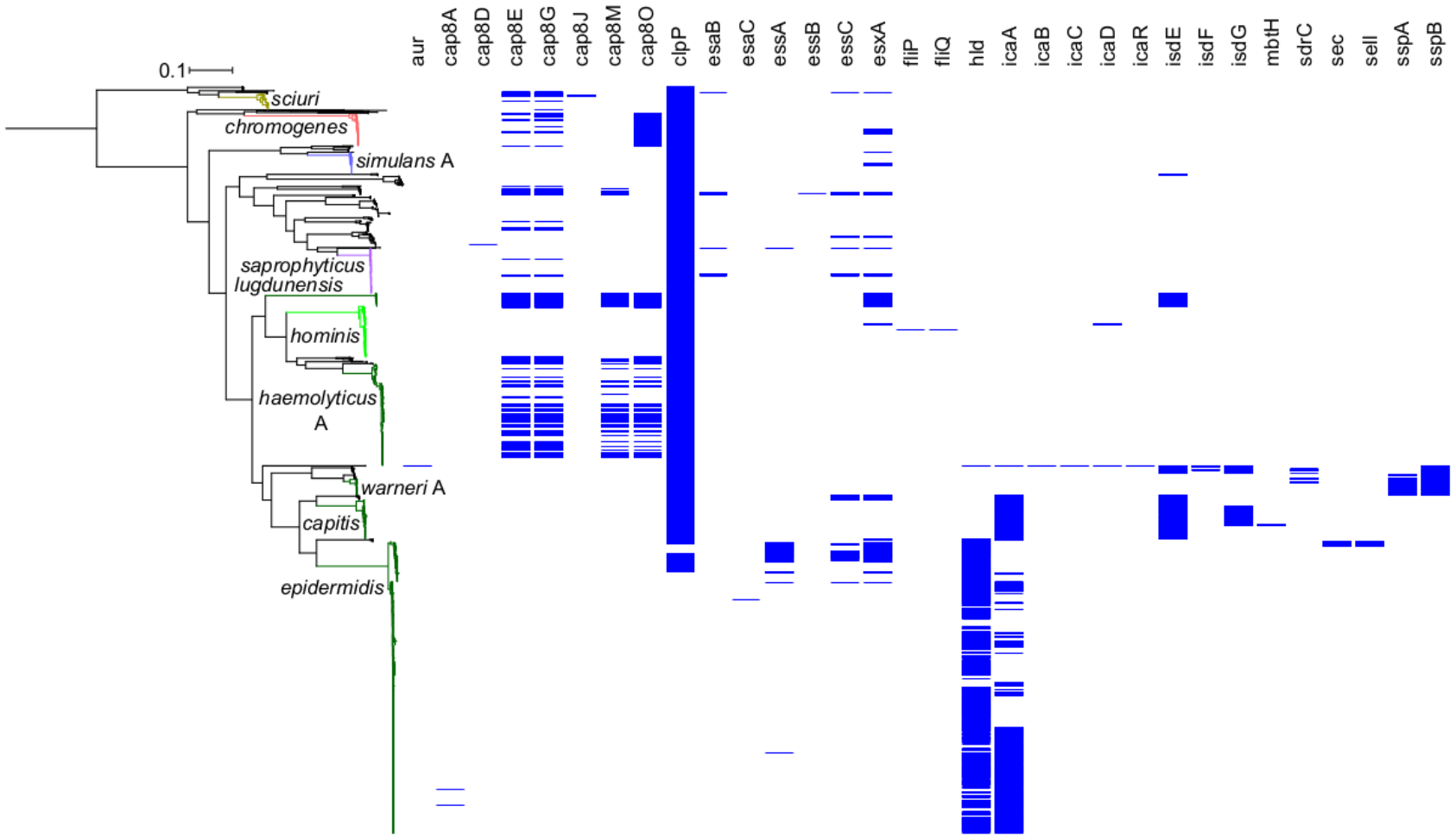

Figure S3.5 Phylogenetic distribution of virulence genes in 1,876 CoNS genomes

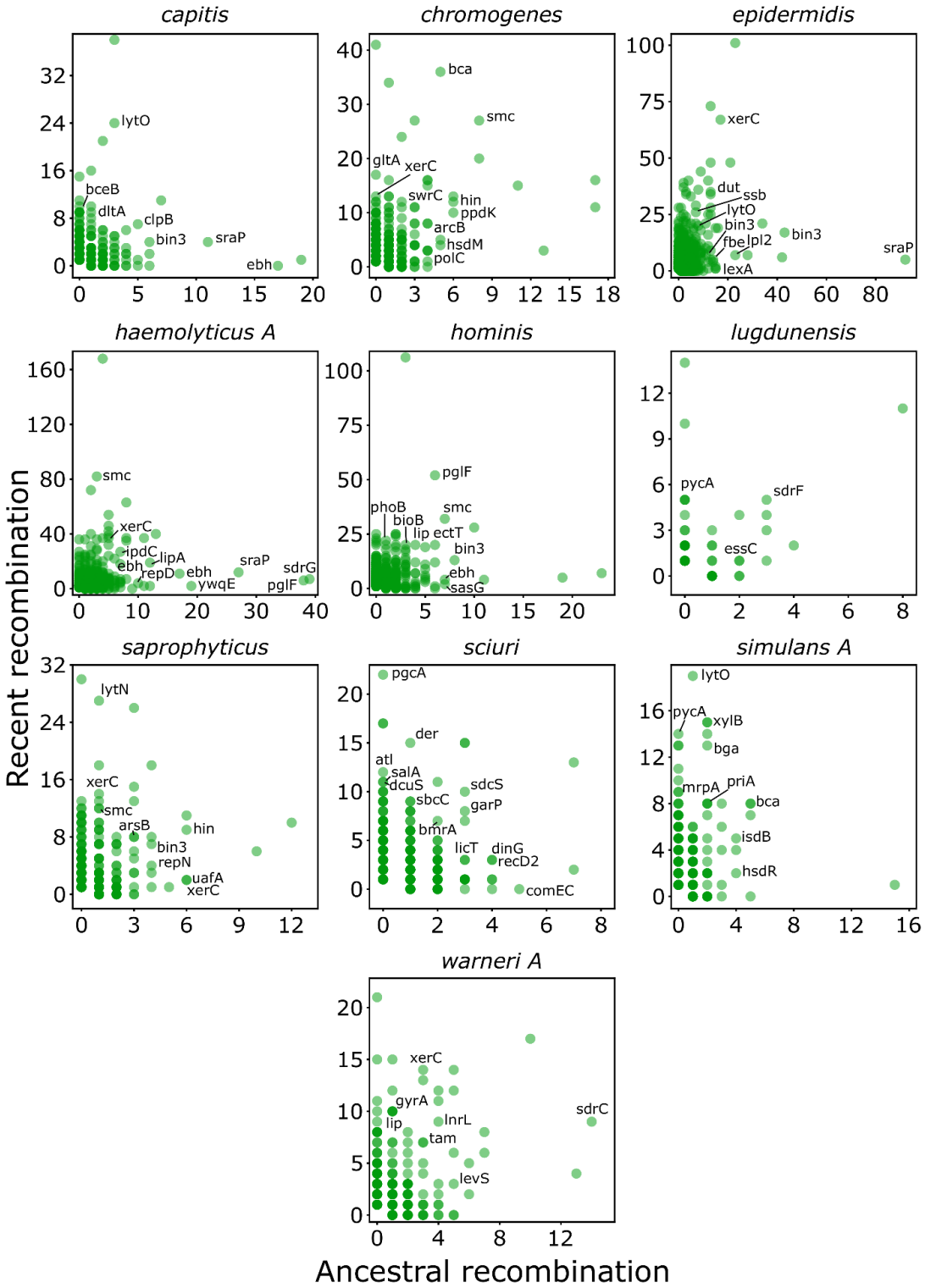

Figure S3.6 Number of recent and ancestral recombination events calculated by fastGEAR in each of the ten species. For visual clarity, only some of the highly recombining genes with known functions are labeled. The range of values in x and y axes in each plot varies between species.

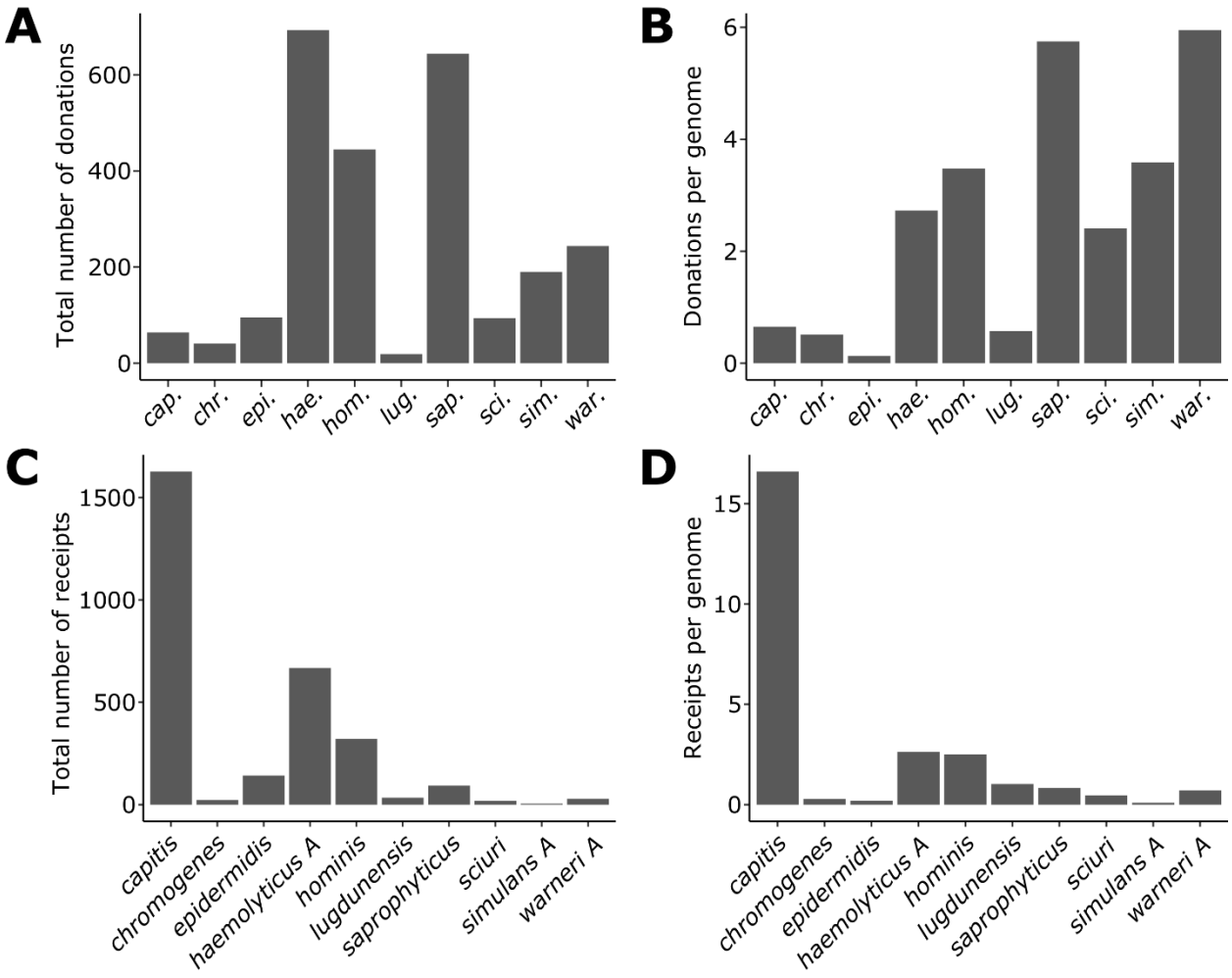

Figure S3.7 Frequencies of DNA donations and receipts in recent recombination events inferred by fastGEAR. Only those species with the highest number of representative genomes are shown. The left panels show the total number of inter-species donations (A) and receipts (C) per species. The right panels show the mean number of donations per genome (B) and receipts per species (D).

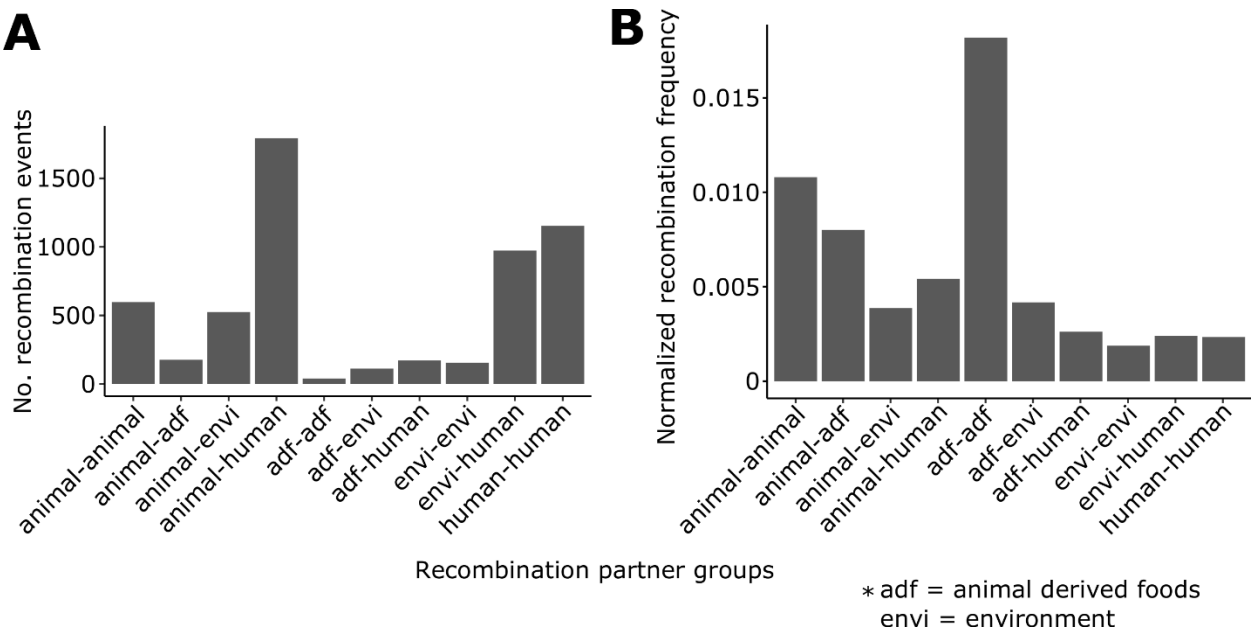

Figure S3.8 Inter-species recombinations between strains from the same or different ecological sources. (A) Total number of recombination events. (B) Number of recombination events normalized to account for the number of genomes per species.

CHAPTER 4

Insertion Sequences as drivers of *Staphylococcus* genome evolution

Joshua T. Smith[a], Cheryl P. Andam[b]

[a] Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, New Hampshire, USA 03824

[b] Department of Biological Sciences, University at Albany, State University of New York, Albany, New York, USA 12222

Introduction

Staphylococcus species are widely distributed in nature, colonizing the skin and mucous membranes of humans and animals. (Becker et al., 2014; Haag et al., 2019). They are commensal organisms making up an important component of a healthy microflora of their hosts (Nakatsuji et al., 2018). However, compromised immune conditions and exposure to internal tissues of the body, e.g. device implantation, allow opportunities for pathogenic strains to arise (Arciola et al., 2018; Lata et al., 2016). Prior to the advent of antimicrobial therapies in the mid-20th century, the most common pathogenic species in the genus, *Staphylococcus aureus*, caused severe infections and high mortality rates (Goto et al., 2017). The number of deaths was reduced dramatically with the discovery and use of drugs such as penicillin and vancomycin; however, the spread of antimicrobial resistance (AMR) combined with a lack of novel treatment options has limited the ability for doctors to effectively treat infections (Foster, 2017). Many other species in the *Staphylococcus* genus have emerged as significant pathogens of humans and animals, with

highly diverse virulence capabilities (Becker et al., 2014; Foster, 2017). For example, *Staphylococcus epidermidis* and *Staphylococcus haemolyticus* are highly prevalent pathogens of humans, causing various soft tissue and blood infections and playing central roles in biofilm production.

Genomic alterations in many *Staphylococcus* species result in diversification, adaptation to new hosts, and acquisition of novel phenotypic characteristics (Argemi et al., 2019; Richardson et al., 2018). One important driver of their genome evolution is a collection of mobile genetic elements that facilitate both intra- and intercellular transport (Firth et al., 2018). These include insertion sequences (IS), transposons, prophages, plasmids, integrative mobilization and conjugative elements, integrons, and genomic islands. ISs are typically very small mobile genetic elements (MGEs) between 0.7 and 2.5 kb in length, most often only encoding one or sometimes two open reading frames (ORF) for transposase (*tnp*) genes, which facilitate the IS's own transposition (P. Siguier et al., 2012; Patricia Siguier et al., 2014, 2015). These *tnp* genes are usually flanked on both ends by terminal inverted repeats. They are classified into different types based primarily on four structural attributes (Patricia Siguier et al., 2014). These include the conserved amino acid sequence within the active site of the Tnp enzyme, the length and size of the inverted repeats, the target site sequence into which they incorporate, and the length and sequence of target site duplications that can occur in the flanking regions upon transposition. However, as more ISs have been discovered, especially with the onset of the genomics era, exceptions to almost every defined characteristic have been found (P. Siguier et al., 2012).

Currently, there are >5,600 classified ISs, which has expanded dramatically with the onset of the genomics era (P. Siguier et al., 2012). As transposases are considered the largest and most widespread protein family found in nature, their huge genetic diversity has resulted in many different types of ISs with considerable differences in activity (Aziz et al., 2010). For example, ISs may carry nearby genes or fragments of genes from the donor DNA to new locations or they may alter gene function in either the donor or the recipient locations upon transposition (Patricia Siguier et al., 2014). If an IS is inserted into a gene within a multigene transcript, a polar effect can occur in which the expression of downstream genes are disrupted (Hutchison et al., 2019). Two copies of similar ISs can also flank both the 5' and 3' ends of one or multiple genes to form a composite transposon that can be mobilized as a single transposable unit (Tansirichaiya et al., 2016). The level of complexity of MGEs formed from ISs can be extensive and ultimately blurs the lines between the defined characteristics of different MGEs (Partridge et al., 2018). For example, the most well studied MGE of *Staphylococcus*, SCC*mec*, is a large chromosomal cassette that carries the methicillin-conferring gene, *mecA*, but can also carry many other antimicrobial and heavy metal resistance genes, plasmids, transposons, and most often, several copies of ISs (Liu et al., 2016). Despite the complexity different combinations of MGEs bring to SCC*mec*, the entire element can be mobilized into new host bacteria, through IS activity, bringing those AMR or virulence determinants with it (Ray et al., 2016).

This review highlights advances in investigations of the diversity and impact of IS in *Staphylococcus* genome evolution. We will describe major ISs found within the genus in terms of how they are classified, their mobilization activity, and how they facilitate adaptation to environmental pressures. This review will focus on the 24 defined *Staphylococcus* ISs in the

ISfinder database (**Table 4.1**) with several isoforms, which for the purposes of this review, will be consolidated where appropriate.

Table 4.1: Characterized Insertion Sequences in *Staphylococcus*

Name	Family	Origin species	Length (bp)	Inverted repeat length (bp)*	Target site duplication length (bp)	ORF**	Accession Number
IS1181	ISL3	*S. aureus*	1513	17/23	8	439 (120-1439)	L14544
IS1182	IS1182	*S. aureus*	1864	14/16		533 (158-1762)	L43082
IS1272	IS1182	*S. haemolyticus*	1935	15/16		273 (272-1093) 236 (1090-1800) 0 (272-1800)	U35635
IS256	IS256	*S. aureus*	1324	17/26	8	390 (102-1274)	M18086
IS257/ IS431	IS6	*S. aureus*	788-791	17-23/20-27		224 (57-731)	X53952 M18437
ISSau1	IS30	*S. aureus*	1070	20/26	4	315 (114-1061)	NC_002952
ISSau2	IS3	*S. aureus*	1660	29/40	3	251 (34-789) 297 (705-1601) 521 (34-1601)	NC_002952
ISSau3	IS1182	*S. aureus*	1946	13/16		548 (280-1926)	NC_002952
ISSau4	IS3	*S. aureus*	1261	27/35		104 (81-395) 298 (380-1222) 380 (81-1222)	NC_002952
ISSau5	IS30	*S. aureus*	1136	7/12		163 (104-595) 175 (597-1124) 340 (104-1124)	NC_002952
ISSau6	IS6	*S. aureus*	793	21/22		224 (59-733)	NC_002952
ISSau8	ISL3	*S. aureus*	1498	19/24	7	438 (141-1457)	NC_007622
ISSau9	IS21	*S. aureus*	2446	21/27		445 (313-1650) 250 (1643-2395)	AM086211
ISSep1	IS1182	*S. epidermidis*	1936	19/28		559 (271-1950)	NC_004461
ISSep2	IS110	*S. epidermidis*	1564			337 (442-1455)	NC_004461
ISSep3	IS200/IS605	*S. epidermidis*	740			160 (133-615)	NC_004461
ISSha1	ISL3	*S. haemolyticus*	1489	22/26	8	438 (132-1448)	NC_007168

* left of slash is number of matching nucleotides, right of slash is total length of inverted repeat
** first number is amino acid length, range in parentheses is transcription start and end positions

IS256

The first discovered and well-studied IS in *Staphylococcus*, IS256 was initially described as a 1,324 bp sequence flanking both ends of a larger aminoglycoside resistance-conferring transposon, TN*4001*, in *S. aureus* (B. R. Lyon et al., 1987). This IS consists of a single ORF coding for a DDE transposase (characterized by a conserved Asp-Asp-Glu amino acid sequence in the active site (Patricia Siguier et al., 2014)), with short imperfect inverted repeats at each end. As IS256 is copied from its original position, it forms a small circularized molecule that is then

re-linearized before insertion into a new site (Loessner et al., 2002). Target sites for insertion appear to be nonrandom with certain hotspots being more common than others, although the molecular mechanisms for this process have yet to be fully explored (Costa et al., 2015; Giulieri et al., 2018).

Since IS256 was originally described, it has also been found within genomes of both closely and distantly related genera such as *Enterococcus* (Kim et al., 2019), *Mycobacterium* (Deng et al., 2018), *Bacillus* (Akashi et al., 2017), and *Edwardsiella* (Tekedar et al., 2020). It is also closely related to the Mutator superfamily of transposases in plants (Guérillot et al., 2014). The ISFinder database lists 244 individual IS sequences within the IS256 family showing widespread distribution of related elements across the bacterial domain (P. Siguier et al., 2012). The mobilization of this IS in *Staphylococcus*, either independently or as a component of a larger element, appears to fill a variety of adaptive roles, most notably by influencing biofilm formation and AMR (Frisch et al., 2018; Hennig & Ziebuhr, 2008). For example, in *S. epidermidis* it has been shown to induce biofilm production by reversibly transposing into biofilm-associated genes and regulatory elements. As such, it is often used as a marker for multidrug-resistant and biofilm forming strains of S. epidermidis acquired in hospital settings (Mertens & Ghebremedhin, 2013). Giulieri et al. have demonstrated *in vivo* that as invasive bacteremia infections progress in response to antimicrobial and immune system stresses, activity of IS256 increases dramatically, allowing for greater genome plasticity in response (Giulieri et al., 2018). In another study, Frisch et al searched a database of 3,755 *S. aureus* assemblies and identified IS256 in several clonal sequence types (ST) known to be expanding in prevalence (Frisch et al., 2018). For example, 41/41 isolates of an emerging strain of community acquired MRSA in India (ST772) contained

copies of IS256. It was also found in every genome of ST239, an emerging global strain. Very often, IS256 is present in many copies in the genome, with up to 45 in total (Kleinert et al., 2017). This contributes to great adaptability, flexibility, and opportunities for homologous recombination. The apparent high transmissibility of IS256 across species and genus boundaries points to the possibility that it may play a role across many bacterial species.

IS256 is a major contributor to the emergence of AMR

With the strong environmental pressures that antimicrobials apply to *Staphylococcus*, the bacteria respond with high IS256 mobility and expansion, which can result in favorable resistance adaptations. They can mobilize transposable elements carrying resistance genes, induce pathways to increase mutation rate, or splice hybrid promoters that influence transcription of genes related to resistance. Having the capability to use IS256 for rapid adaptation plays a significant role in the development of AMR strains in new ecological niches (Frisch et al., 2018; Giulieri et al., 2018).

The first description of IS256 demonstrated that two inverted copies form the flanking ends of Tn*4001* (B. R. Lyon et al., 1987), an AMR composite transposon. This transposable unit can be found either within the chromosome or on a transferable pSK1 plasmid and contains the *aacA-aphD* gene, the product of which confers resistance to aminoglycoside antimicrobials, gentamycin, tobramycin, and kanamycin (Bruce R. Lyon et al., 1984). The expression of *aacA-aphD* is regulated by a promoter within the IS256 element. In one study of hospital-associated outbreaks of MRSA in Japan, researchers found Tn*4001* in 56% of the 161 MRSA isolates in their dataset (Tanabe & Udou, 1995). Since other gram-positive isolates were also found to be

resistant to aminoglycosides, a second retrospective study in the same hospital found that many isolates of *S. epidermidis, S. haemolyticus, S. capitis,* and *Enterococcus faecalis* contained Tn*4001* as well (Udou, 2004). The authors believe that a combination of aminoglycoside use with the high self-mobilization activity of IS256 contributed to interspecific and intergeneric integration and to high prevalence of this resistance-conferring transposon.

The origin of IS256 in staphylococci is mostly unknown. It was initially believed to have originated in *S. aureus,* as that was the organism it was identified in (B. R. Lyon et al., 1987). However, while IS256 presence in *S. aureus* is highly correlated with aminoglycoside resistance, it is present in a remarkably high proportion of enterococci isolates (85%) independent of resistance phenotype (Rice & Thorisdottir, 1994). This led the researchers to believe that IS256 likely originated in *Enterococcus* and eventually crossed into *Staphylococcus* as a component of Tn*4001*. This is supported by a different study of 70 *S. epidermidis* isolates from orthopedic infections in Italy (Montanaro et al., 2007). The authors found a 100% association between gentamycin-resistant and IS256-positive isolates. In addition, 56% of the IS256-positive isolates were resistant to ≥ 5 antimicrobials. More recent research has identified many instances of IS256 acting independently or as a component of other genetic elements in certain strains of *Staphylococcus*, potentially highlighting an expansion within the genus (Frisch et al., 2018; McEvoy et al., 2013).

Induction of the SOS response in bacteria often occurs with exposure to antimicrobials, resulting in halting of cell cycle, repairing DNA damage and inducing mutagenesis (Michel, 2005). However, in addition to inducing the SOS response, one study also found that when

certain strains of *S. aureus* are exposed to ciprofloxacin and vancomycin, IS256 can spontaneously insert into *mutS*, a critical component of the mismatch repair system (Nagel et al., 2011). The authors determined that insertion disrupts normal repair function and induces a mutator phenotype in the strain as the SOS response occurs. Therefore, the strain becomes much more sensitive to antimicrobial stresses and can more easily develop AMR (Schaaff et al., 2002).

One of the most significant effects that IS256 has in influencing gene expression is its ability to form hybrid promoters with endogenous regulatory sequences (Couto et al., 2003; Kuroda et al., 2019; Maki & Murakami, 1997; Yin et al., 2019). IS256 often inserts into regions directly upstream to -10 Pribnow box for specific genes with the adjacent end sequence of the IS containing a suitable -35 site. The location of insertion causes variable nucleotide distance between the two sites, which can directly influence the amount of gene expression that occurs (Maki & Murakami, 1997). One example of this is the regulation of *llm*, an important but little studied gene involved in methicillin resistance (Maki & Murakami, 1997). Insertion of IS256 into different regions of the promoter region results in variable increases in *llm* expression. This then contributes to heteroresistance phenotypes in affected strains.

Increased expression caused by a strong hybrid promoter may also have led to the origin of *mecA*, which is the major methicillin resistance determinant (Couto et al., 2003). A hypothetical progenitor to *mecA*, *mecA1*, is a homolog originally described in *Staphylococcus sciuri* and normally does not cause a methicillin resistance phenotype independently (Wu et al., 1998). However, Couto et al have presented a potential interaction between *S. sciuri* and a clinical strain of *S. aureus* or *S. epidermidis* that led to the acquisition of IS256 by *S. sciuri*

(Couto et al., 2003). This was then ultimately inserted upstream of the *mecA1* gene, forming a strong hybrid promoter. This increases expression of *mecA1* beyond normal levels, resulting in the methicillin resistance phenotype that we see in isolates with *mecA*. While we have not definitively shown that this process occurred, there still appears be an association between *mecA* and IS256 in other coagulase-negative staphylococci like *S. sciuri*. (Rolo et al., 2017). One study from India found that of the 55 methicillin-resistant isolates of *S. haemolyticus, Staphylococcus saprophyticus, Staphylococcus hominis, S. sciuri,* and *Staphylococcus cohnii*, 95% were positive for both *mecA* and IS256 (Soumya et al., 2016).

A third example of hybrid promoter construction due to IS256 insertion is the increase in vancomycin-intermediate *S. aureus* (VISA). As the cases of MRSA began rising in prevalence in the 1960's, vancomycin often became the next drug of choice in many healthcare settings (Kuroda et al., 2019). With the more widespread use of vancomycin, resistance to it began to increase as well (Zhang et al., 2015). One such resistance mechanism to induce reduced susceptibility to vancomycin involves alterations to the *walKR* two-component regulatory system. This system regulates cell wall metabolism-associated genes, such as those associated with peptidoglycan biosynthesis during cross bridge hydrolysis (Dubrac et al., 2007). Several studies have measured how the expression of the *walKR* genes is impacted by IS256 insertion in the promoter region (Jansen et al., 2007; Kuroda et al., 2019; McEvoy et al., 2013; Yin et al., 2019). The two most recent studies conducted *in vivo* show a strong hybrid promoter is formed that increases expression of both genes (Kuroda et al., 2019; Yin et al., 2019). This then leads to a much thicker cell wall, removes targets for vancomycin binding, and depolarizes the surface of the bacterium preventing the cationic killing action of daptomycin and vancomycin. While both

studies show increasing non-susceptibility of vancomycin and daptomycin due to increased expression of *walKR*, Yin at al actually show that IS256 insertion into the promoter of *mprF*, a phosphatidylglycerol lysyltransferase gene, increases its expression and ultimately also contributes to daptomycin non-susceptibility (Yin et al., 2019).

Because IS256 transposes with a copy-paste mechanism, increased insertions during chronic Staphylococcal disease states with prolonged antimicrobial treatments can eventually burden the bacterium with many copies and result in a small-colony variant (SCV) phenotype (Kleinert et al., 2017). Such strains are often slow growing, less susceptible to antimicrobials, and unstable (Kahl et al., 2016; Ou et al., 2016; Tande et al., 2014). As a result, they are more difficult to treat. One study has demonstrated that the high number of IS256 elements present in certain strains can contribute to high transposase activity, which can cause widespread rearrangements and mutations in certain genes such as *hemY* and *guaA* (Kleinert et al., 2017). This activity can then lead to the SCV phenotype. However, this state is reversible as IS256 copies can subsequently be excised as environmental conditions favor a return to rapid growth (Kleinert et al., 2017).

IS256 also moderates virulence phenotypes

IS256 can have a large impact on staphylococcal virulence, especially relating to biofilm formation (Hennig & Ziebuhr, 2008). Polysaccharide intercellular adhesin (PIA) is an important polymer for the formation of biofilms and its expression is regulated by the *ica* operon (Ziebuhr et al., 1999). Reduced activity of *ica* genes ultimately results in a reduction in biofilm forming ability, and vice versa. *S. epidermidis* bacteria are able to modulate their ability to form biofilms

by alternating insertions and excisions of IS256 to and from different regions in the *ica* operon (Ziebuhr et al., 1999), with an apparent hotspot in the *icaC* gene (Loessner et al., 2002). This alternating behavior causes what is known as phase variation, where *S. epidermidis* isolates are capable of transitioning between forming biofilms and not. The *ica* operon expression is also regulated through a separate mechanism by which IS256 inserts into global regulators *rsbU* and *sarA*, reducing their expression (Conlon et al., 2004). The functional benefit of this switching is that when bacteria locate a favorable environment in their host, they can attach and form biofilm (Ziebuhr et al., 1999). However, in later stages of infection when resources are limited, or the immune system targets biofilm-forming isolates, they can switch off PIA production by inserting copies of IS256 into the *ica* operon or its upstream regulators *rsbU* and *sarA* and disseminate to new habitats (Conlon et al., 2004).

In *S. aureus*, IS256 influences virulence by disrupting genes associated with toxins and efflux pump development (Costa et al., 2015; Di Gregorio et al., 2016). This is in contrast to the generally accepted view that the presence or absence of specific virulence genes in bacteria is the main factor in virulence (Benson et al., 2014). The Rot transcription factor is known as an inhibitor of toxin production, although the exact mechanism is not fully understood. Deletion of the *rot* gene induces a hypervirulent state, yet no wildtype strains have been found without it (Alonzo III et al., 2012). One study of the highly virulent USA500 *S. aureus* strain has found a likely explanation for *rot* maintenance, wherein IS256 has been inserted into the promoter region of the gene, reducing but not eliminating its expression (Benson et al., 2014). This is an important finding for *S. aureus* pathogenesis because USA500 did not show any additional virulence factors compared to closely related, yet less virulent strains (M. Li et al., 2010). In

some strains that produce toxins, there is likely some balance in Rot production to prevent counterproductive killing of the host while simultaneously allowing for high transmissibility/colonization (Benson et al., 2014). This may be controlled by IS256 insertion.

Large-scale genomic inversion caused by IS256

In some cases, IS256 has been shown to cause even more dramatic alterations to many different regions of a single *Staphylococcus* genome. In one ST8 CA-MRSA strain isolated from a fatal pediatric case from Russia, the genome responsible contained 19 independent sequences of IS256 across different regions (Wan et al., 2016). Three discrete regions in the genome contained 2-5 copies of IS256 in very close proximity, indicating potential recombination hotspots. Between two of these hotspots, there was a remarkable >1 Mb inversion in the genome, which may provide a selective advantage to bacterial pathogens as previously shown in *Pseudomonas aeruginosa* (Kresse et al., 2002). The authors of this study also found that the circular intermediate form of elements containing IS256 may be capable of cell-to-cell transfer as what would be known as a transmissible insertion sequence (Wan et al., 2016). More research will be necessary to verify this process. Overall, the results indicate that IS256 has played a significant role in the evolution of this MRSA strain in Russia in a variety of ways, from being present in multiple copies of all MRSA isolates, different genomic islands, or upstream of a putative virulence factor *rsp*, while also inducing the largest genomic rearrangement in *Staphylococcus* by more than 2X.

IS257/IS431

Shortly after the discovery of IS256, two isoform IS elements, IS257 and IS431, were characterized at roughly the same time and have nearly identical sequences (Barberis-Maino et al., 1987; Gillespie et al., 1987). IS257 was tentatively described in an Australian population after it was found in several distinct plasmids containing resistance determinants for aminoglycosides, tetracycline, heavy metals, and most notably for *S. aureus*, methicillin (Gillespie et al., 1987). Later that year, a different research group described IS431 and confirmed its presence within a "*mec*-associated" element that would later be known as a major component of SCC*mec* (Barberis-Maino et al., 1987). For the purposes of this review, these two IS elements will be referred to as IS257.

IS257 is considerably shorter than IS256 at approximately 790 bp in length (Gillespie et al., 1987). Like IS256, it was also discovered as flanking regions to a transposon, Tn*4003*, One of the IS257 copies forms a hybrid promoter to express the trimethoprim resistance gene, *dfrA* (Rouch et al., 1989). Since its discovery, IS257 has been associated with different classes of MGEs and other AMR determinants and phenotypes (Firth et al., 2018). IS257 is capable of transposing through two different means (Harmer & Hall, 2020). First, it can use a copy-in mechanism in which a donor molecule with IS257 cointegrates at a random site in the recipient DNA, is copied, and resolves via homologous recombination (Needham et al., 1995). This method, like most simple transpositions, produces target site duplications flanking both ends of the resultant insertion. However, if there is already a second IS257 sequence already present in the recipient, it can instead use a conservative targeted approach to attack one end of the IS in the recipient DNA and transpose into the recipient strand, carrying flanking genes such as AMR

determinants (Harmer & Hall, 2020). This method has been harder to detect due to target site duplications not being produced. The role and contribution each of these mechanisms plays in *Staphylococcus* evolution is not yet well studied; however, they highlight the flexibility of this element to act in either a stochastic or targeted manner.

IS257 is ubiquitous in SCCmec elements

Due to a near universal association with SCC*mec*, IS257 likely plays a significant role in its development and/or mobilization (Ray et al., 2016). SCC*mec* is a diverse mosaic island that is most often characterized by the structure of two components, the *ccr* recombinase gene complex and the methicillin resistance-conferring *mec* complex (Katayama et al., 2000). There are many different insertions, deletions, and rearrangements that distinguish different SCC*mec* types (Liu et al., 2016). Out of 13 known types (Baig et al., 2018), all except for Type XI have between one and four copies of IS257 in various combinations of positions and orientations, but they are often directly flanking other major elements (Ray et al., 2016). A recent study of the origin of SCC*mec* places the introduction of IS257 into SCC*mec* after its initial development in the *S. sciuri* group (Miragaia, 2018). Over time, the number of IS257 copies has varied among different types, resulting in differences in movement and regulation of genetic elements within the overarching MGE.

One major role of IS257 is the deletion of different regulatory elements within SCC*mec* (Jiang et al., 2019). In *S. haemolyticus*, IS257 is implicated in the deletion of *mecI* and partial deletion of *mecR1*, two genes involved in regulating expression of *mecA* (Suzuki et al., 1993). This been demonstrated *in vitro* and consequently leads to derepressed expression of *mecA*,

which can then increase methicillin resistance (Katayama et al., 2000). A study of 60 multidrug-resistant *S. haemolyticus* isolates from an Iranian hospital identified 11 isolates with *mecR1* and *ccr* genes deleted (Hosseinkhani et al., 2018). In these strains, *mecA* is flanked directly by two IS257 copies with intact target site duplications, indicating active transposable elements. These isolates are presumed to have deregulated expression of *mecA*, which is likely responsible for high levels of methicillin resistance. Conversely, there is also evidence that IS257 activity can result in a deletion of *mecA* in response to vancomycin serial passage, which ultimately increased vancomycin resistance and decreased methicillin resistance in these strains (A. Wang et al., 2017). It would be reasonable to conclude that this process may be occurring simultaneously with the IS256-mediated modification to *walKR* expression mentioned previously to result in a "see-saw" effect between methicillin and vancomycin resistance phenotypes (Giulieri et al., 2020).

Due to the clear association of IS257 and SCC*mec*, it has also been hypothesized that IS257 may play a role in mobilizing SCC*mec* (Ray et al., 2016). IS257 is known to integrate and excise AMR genes into/from conjugative plasmids (Baines et al., 2019). A study by Ray et al demonstrated a potential mechanism by which SCC*mec* elements may be horizontally transferred in this manner (Ray et al., 2016). They were successfully able to show that IS257 copies in flanking regions of SCC*mec* can be used as target sites of recombination to excise the entire element and integrate it into a conjugative plasmid. This is then transferred horizontally to a secondary strain to ultimately be incorporated into its chromosome. However, evidence of this process has not yet been elucidated in clinical strains, as the element was always lost from plasmids unless under specific selection (Ray et al., 2016). The authors feel this conjugation

method is more likely over transduction due to a lack of homology in the flanking regions of SCC*mec* between different *Staphylococcus* species. The length of SCC*mec* is also generally longer than known staphylococcal-associated transducing phages can carry (Deghorain & Van Melderen, 2012).

IS1181 and ISSha1

IS1181 is a 1,513 bp IS coding for a single *tnp* gene with 23 bp inverted repeats (Derbise et al., 1994). It is often detected with 8-bp target site duplications, likely indicating replicative transposition. IS1181 is unique compared to IS256 and IS257 in that it has only been found in *S. aureus* and is not directly associated with any AMR determinants (Symms et al., 1998). However, it is often found with many copies within a *S. aureus* genome, irrespective of resistance phenotype, which could make it a useful marker for strains independent of AMR (Symms et al., 1998).

The activity of IS1181 is poorly understood with few experimental studies. It does appear to play roles in regulating virulence not seen in other ISs of *Staphylococcus*. RpiRc is a recently described transcription factor that modulates virulence through a linkage with central metabolism (Gaupp et al., 2016). Higher expression of this protein leads to a repression of certain virulence determinants such as haemolysins. Menendez-Gil et al demonstrated that IS1181 insertions into the 3'UTR of *rpiRc* results in a significant increase in expression, and thereby reducing the production of virulence factors (Menendez-Gil et al., 2020). IS1181 also appears to play a role in metabolism unrelated to pathogenesis. Insertions of IS1181 into the *ptsH* gene disrupts its expression, which initiates a switch from proline auxotrophy to prototrophy (C. Li et al., 2010).

Regulation of this switch in proline production is likely influenced by amino acid sources present in the environment. Since IS1181 often appears many times within a genome (C. Li et al., 2010), it likely plays other important roles beyond what is described here. Some of these may be more subtle, or they may be upstream of more complicated pathways, and, in some cases, likely not directly related with virulence or AMR. Further research into these roles and the factors that prevent cross-species transfer would be beneficial.

ISSha1 is in the same family as IS1181, sharing ~60% amino acid homology (Takeuchi et al., 2005). This IS plays a significant role in *S. haemolyticus* evolution, as it is highly prevalent in this species. In one study, it was present in all 74 clinical isolates in the dataset, with as many as 28 copies per genome (Watanabe et al., 2007). Like IS1181, there is not an apparent connection between ISSha1 and AMR, however, it is known to cause large genomic inversions and deletions in *S. haemolyticus* (Watanabe et al., 2007). The results of such genome fluidity could potentially play roles in pathogenicity.

IS1272 and IS1182

Originally described in *S. haeomlyticus*, IS1272 is a larger element at 1,934 bp in length including terminal inverted repeats of 16 bp, and unlike previously described ISs, contains two ORFs (Archer et al., 1996). Its mechanism of transposition is not fully understood; however, based on related ISs it may use a conservative mechanism in which a stem-loop formed on the end of the 3' inverted repeat attacks palindromic sequences in the recipient strand and is pasted without the formation of target site duplications (Wan et al., 2017). Several subtypes with shared

homology have been identified in other *Staphyococcus* species such as *S. epidermidis* and *S. aureus* (Wan et al., 2017).

A dataset of both pathogenic and commensal *S. haemolyticus* isolates from 16 different countries over a 12-year period showed that 132/133 (99.2%) contained at least one copy of IS1272 (Bouchami et al., 2016). Of these, 59% had between 7-8 copies distributed throughout the genome. This is a lower estimation than a previously found (Archer et al., 1996), but still highlights the ubiquity of IS1272. Bouchami et al also determined that 65.4% of the isolates had non-typable SCC*mec* types, which may have resulted from genomic rearrangement caused by ISs like IS1272 (Bouchami et al., 2016). Type IV SCC*mec* is known to have an IS1272 fragment in the *mec* region, supporting this hypothesis. The singular study describing a specific role of IS1272 determined it is capable of mobilizing a triclosan resistance gene, *sh-fabI*, in *S. aureus* (Furi et al., 2016). It is speculated that like other ISs, IS1272 may play a role in many reversible processes such as mannitol fermentation, haemolysis, biofilm formation, and switching between pathogenic and non-pathogenic lifestyles (Bouchami et al., 2016). This IS, like ISSha1 (above), is likely important to *S. haemolyticus* evolution with potential roles in pathogenicity or AMR, but more research is needed to find these links (Pain et al., 2019).

IS1182 is an IS in the same family as IS1272 but is most often identified in *S. aureus* (Derbise et al., 1996). Like IS256, it is associated with aminoglycoside resistance due to the presence of flanking copies around a composite transposon, Tn*5405*. Within this transposon are two resistance genes, *aphA-3* and *aadE*. A more recent study has demonstrated that IS1272 likely

plays a role in increasing methicillin resistance by inserting into and inactivating the *lytH* gene, which encodes a putative cell wall hydrolyzing enzyme (Fujimura & Murakami, 2008).

ISSau2

Common only *to S. aureus*, ISSau2 is 1,660 bp in length and contains two overlapping ORFs (P. Siguier et al., 2012). Transposase activity does not occur unless there is a -1 translational frameshift, which is presumed to play a role in regulating its activity (L. Wang et al., 2018, p. 2). Compared to closely related ISs in other species, a conservative cut and paste transposition mechanism is assumed to be used by ISSau2. Little is known about specific effects on genome evolution of *S. aureus*, but it appears to be present in high copy numbers in clonal complex 30 as well as several strains causing bovine and ovine infections (McGavin et al., 2012).

Other less known ISs of *Staphylococcus*

The remaining ISs have only been found through distinctive sequence homology, have not been formally described, or little is known about their activity (Gill et al., 2005; Lee et al., 2016; Wilson et al., 2016). These include: ISSau1, ISSau3, ISSau4, ISSau5, ISSau6, ISSau8, ISSau9, ISSep1, ISSep2, and ISSep3. Some notable findings have been presented for some, which provide some starting points for future research. First, ISSep1, a member of the IS1182 family, has been found within an arginine catabolic mobile element (ACME) in *S. epidermidis* (Wilson et al., 2016). This ACME island encodes several genes that help *Staphylococcus* species survive and colonize their hosts, e.g. *speG* is present and confers resistance to polyamines during infection (Joshi et al., 2011). ISSep1 may possibly play a role in mobilizing either the island or

genes within it. Another IS, ISSau4 has been detected in a unique SCC*mec* varient in *Staphylococcus lugdunensis*, indicating a potential role in methicillin resistance (Chang et al., 2017). Finally, ISSau9 sequences with intact target site duplications have been found nearby to a phenicol resistance-conferring *cfr* gene, indicating it likely plays a role in its mobilization (Kehrenberg et al., 2007).

Conclusions

ISs are among the smallest and simplest mobile elements of *Staphylococcus*, and yet have some the most diverse and consequential properties. This review highlights the importance of ISs in the evolution and ecology of *Staphylococcus* species. Among the 17 named ISs above, there are considerable differences in the DNA sequences, transposition mechanism, preferred host species, and the adaptive advantage they confer. ISs contribute to rapid responses to environmental stresses and infection of host tissues. The structural changes these elements make in the genome are often reversible, providing another level of flexibility as their environment fluctuates back and forth. They also play roles in forming larger and more complex MGEs such as composite transposons and genomic islands. In the clinical sense, they are major drivers of pathogenicity by mobilizing and regulating virulence and AMR determinants. Much of this ability is the result of the variable and active nature of ISs.

Bacterial genomes are continually changing, with regular gains, losses and inversions which blur the lines between ISs and other MGEs. Further research into staphylococcal ISs is still needed to explore more about their impact. While genomic methods have helped tremendously in detecting the presence and diversity of ISs, there is still a great deal we do not

know. The exact transposition mechanisms, target sites, secondary molecules are still unclear for many ISs (Patricia Siguier et al., 2014). Additionally, we need to learn more at the population level of bacteria like *Staphylococcus* to understand the dynamics of events, such as IS expansion (e.g. IS256, above) or host adaptation, beyond the level of a single individual. This progress will rely on both *in vivo* functional studies and improved DNA sequencing and assembly strategies. The common sequencing approach using short read draft genomes presents major limitations as they may be too fragmented to accurately detect small and often repetitive elements like ISs (Yui Eto et al., 2019). Advancements in the cost and quality of long read sequencing technology will further improve our understanding of the role and impact of ISs in the diversification, host adaptation and genome evolution of clinical and commensal strains of *Staphylococcus*.

References

Akashi, M., Harada, S., Moki, S., Okouji, Y., Takahashi, K., Kada, S., Yamagami, K., Sekine, Y., Watanabe, S., Chibazakura, T., & Yoshikawa, H. (2017). Transposition of insertion sequence IS*256Bsu1* in *Bacillus subtilis* 168 is strictly dependent on *recA*. *Genes & Genetic Systems*, *92*(2), 59–71. https://doi.org/10.1266/ggs.16-00071

Alonzo III, F., Benson, M. A., Chen, J., Novick, R. P., Shopsin, B., & Torres, V. J. (2012). *Staphylococcus aureus* leucocidin ED contributes to systemic infection by targeting neutrophils and promoting bacterial growth in vivo: LukED contributes to *S. aureus* bacteraemia. *Molecular Microbiology*, *83*(2), 423–435. https://doi.org/10.1111/j.1365-2958.2011.07942.x

Archer, G. L., Thanassi, J. A., Niemeyer, D. M., & Pucci, M. J. (1996). Characterization of IS1272, an insertion sequence-like element from *Staphylococcus haemolyticus*. *Antimicrobial Agents and Chemotherapy*, *40*(4), 924–929. https://doi.org/10.1128/AAC.40.4.924

Arciola, C. R., Campoccia, D., & Montanaro, L. (2018). Implant infections: Adhesion, biofilm formation and immune evasion. *Nature Reviews Microbiology*, *16*(7), 397–409. https://doi.org/10.1038/s41579-018-0019-y

Argemi, X., Hansmann, Y., Prola, K., & Prévost, G. (2019). Coagulase-Negative Staphylococci Pathogenomics. *International Journal of Molecular Sciences*, *20*(5), 1215. https://doi.org/10.3390/ijms20051215

Aziz, R. K., Breitbart, M., & Edwards, R. A. (2010). Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Research*, *38*(13), 4207–4217. https://doi.org/10.1093/nar/gkq140

Baig, S., Johannesen, T. B., Overballe-Petersen, S., Larsen, J., Larsen, A. R., & Stegger, M. (2018). Novel SCC mec type XIII (9A) identified in an ST152 methicillin-resistant *Staphylococcus aureus*. *Infection, Genetics and Evolution*, *61*, 74–76. https://doi.org/10.1016/j.meegid.2018.03.013

Baines, S. L., Jensen, S. O., Firth, N., Gonçalves da Silva, A., Seemann, T., Carter, G. P., Williamson, D. A., Howden, B. P., & Stinear, T. P. (2019). Remodeling of pSK1 Family Plasmids and Enhanced Chlorhexidine Tolerance in a Dominant Hospital Lineage of Methicillin-Resistant *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, *63*(5), e02356-18, /aac/63/5/AAC.02356-18.atom. https://doi.org/10.1128/AAC.02356-18

Barberis-Maino, L., Berger-Bächi, B., Weber, H., Beck, W. D., & Kayser, F. H. (1987). IS431, a staphylococcal insertion sequence-like element related to IS26 from *Proteus vulgaris*. *Gene*, *59*(1), 107–113. https://doi.org/10.1016/0378-1119(87)90271-X

Becker, K., Heilmann, C., & Peters, G. (2014). Coagulase-Negative Staphylococci. *Clinical Microbiology Reviews*, *27*(4), 870–926. https://doi.org/10.1128/CMR.00109-13

Benson, M. A., Ohneck, E. A., Ryan, C., Alonzo, F., Smith, H., Narechania, A., Kolokotronis, S.-O., Satola, S. W., Uhlemann, A.-C., Sebra, R., Deikus, G., Shopsin, B., Planet, P. J., & Torres, V. J. (2014). Evolution of hypervirulence by a MRSA clone through acquisition of a transposable element: Rot dysregulation in MRSA-USA500 leads to hypervirulence. *Molecular Microbiology*, *93*(4), 664–681. https://doi.org/10.1111/mmi.12682

Bouchami, O., de Lencastre, H., & Miragaia, M. (2016). Impact of Insertion Sequences and Recombination on the Population Structure of *Staphylococcus haemolyticus*. *PLOS ONE*, *11*(6), e0156653. https://doi.org/10.1371/journal.pone.0156653

Chang, S.-C., Lee, M.-H., Yeh, C.-F., Liu, T.-P., Lin, J.-F., Ho, C.-M., & Lu, J.-J. (2017). Characterization of two novel variants of staphylococcal cassette chromosome mec elements in oxacillin-resistant *Staphylococcus lugdunensis*. *Journal of Antimicrobial Chemotherapy*, *72*(12), 3258–3262. https://doi.org/10.1093/jac/dkx291

Conlon, K. M., Humphreys, H., & O'Gara, J. P. (2004). Inactivations of rsbU and sarA by IS256 Represent Novel Mechanisms of Biofilm Phenotypic Variation in *Staphylococcus epidermidis*. *Journal of Bacteriology*, *186*(18), 6208–6219. https://doi.org/10.1128/JB.186.18.6208-6219.2004

Costa, S. S., Viveiros, M., Rosato, A. E., Melo-Cristino, J., & Couto, I. (2015). Impact of efflux in the development of multidrug resistance phenotypes in *Staphylococcus aureus*. *BMC Microbiology*, *15*(1), 232. https://doi.org/10.1186/s12866-015-0572-8

Couto, I., Wu, S. W., Tomasz, A., & de Lencastre, H. (2003). Development of Methicillin Resistance in Clinical Isolates of *Staphylococcus sciuri* by Transcriptional Activation of the mecA Homologue Native to the Species. *Journal of Bacteriology*, *185*(2), 645–653. https://doi.org/10.1128/JB.185.2.645-653.2003

Deghorain, M., & Van Melderen, L. (2012). The Staphylococci Phages Family: An Overview. *Viruses*, *4*(12), 3316–3335. https://doi.org/10.3390/v4123316

Deng, D., Li, F., & Li, M. (2018). A Novel Propane Monooxygenase Initiating Degradation of 1,4-Dioxane by *Mycobacterium dioxanotrophicus* PH-06. *Environmental Science & Technology Letters*, *5*(2), 86–91. https://doi.org/10.1021/acs.estlett.7b00504

Derbise, A., Dyke, K. G. H., & El Solh, N. (1994). Isolation and Characterization of IS1181, an Insertion Sequence from *Staphylococcus aureus*. *Plasmid*, *31*(3), 251–264. https://doi.org/10.1006/plas.1994.1027

Derbise, A., Dyke, K. G. H., & El Solh, N. (1996). Characterization of a *Staphylococcus aureus* Transposon, Tn5405,Located within Tn5404and Carrying the Aminoglycoside

Resistance Genes, aphA-3 and aadE. *Plasmid*, *35*(3), 174–188. https://doi.org/10.1006/plas.1996.0020

Di Gregorio, S., Fernandez, S., Perazzi, B., Bello, N., Famiglietti, A., & Mollerach, M. (2016). Increase in IS 256 transposition in invasive vancomycin heteroresistant *Staphylococcus aureus* isolate belonging to ST100 and its derived VISA mutants. *Infection, Genetics and Evolution*, *43*, 197–202. https://doi.org/10.1016/j.meegid.2016.05.001

Dubrac, S., Boneca, I. G., Poupel, O., & Msadek, T. (2007). New Insights into the WalK/WalR (YycG/YycF) Essential Signal Transduction Pathway Reveal a Major Role in Controlling Cell Wall Metabolism and Biofilm Formation in *Staphylococcus aureus*. *Journal of Bacteriology*, *189*(22), 8257–8269. https://doi.org/10.1128/JB.00645-07

Firth, N., Jensen, S. O., Kwong, S. M., Skurray, R. A., & Ramsay, J. P. (2018). Staphylococcal Plasmids, Transposable and Integrative Elements. *Microbiology Spectrum*, *6*(6). https://doi.org/10.1128/microbiolspec.GPP3-0030-2018

Foster, T. J. (2017). Antibiotic resistance in *Staphylococcus aureus*. Current status and future prospects. *FEMS Microbiology Reviews*, *41*(3), 430–449. https://doi.org/10.1093/femsre/fux007

Frisch, M. B., Castillo-Ramírez, S., Petit, R. A., Farley, M. M., Ray, S. M., Albrecht, V. S., Limbago, B. M., Hernandez, J., See, I., Satola, S. W., & Read, T. D. (2018). Invasive Methicillin-Resistant *Staphylococcus aureus* USA500 Strains from the U.S. Emerging Infections Program Constitute Three Geographically Distinct Lineages. *MSphere*, *3*(3), e00571-17, /msphere/3/3/mSphere571-17.atom. https://doi.org/10.1128/mSphere.00571-17

Fujimura, T., & Murakami, K. (2008). *Staphylococcus aureus* Clinical Isolate with High-Level Methicillin Resistance with an lytH Mutation Caused by IS1182 Insertion. *Antimicrobial Agents and Chemotherapy*, *52*(2), 643–647. https://doi.org/10.1128/AAC.00395-07

Furi, L., Haigh, R., Al Jabri, Z. J. H., Morrissey, I., Ou, H.-Y., León-Sampedro, R., Martinez, J. L., Coque, T. M., & Oggioni, M. R. (2016). Dissemination of Novel Antimicrobial Resistance Mechanisms through the Insertion Sequence Mediated Spread of Metabolic Genes. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.01008

Gaupp, R., Wirf, J., Wonnenberg, B., Biegel, T., Eisenbeis, J., Graham, J., Herrmann, M., Lee, C. Y., Beisswenger, C., Wolz, C., Tschernig, T., Bischoff, M., & Somerville, G. A. (2016). RpiRc Is a Pleiotropic Effector of Virulence Determinant Synthesis and Attenuates Pathogenicity in *Staphylococcus aureus*. *Infection and Immunity*, *84*(7), 2031–2041. https://doi.org/10.1128/IAI.00285-16

Gill, S. R., Fouts, D. E., Archer, G. L., Mongodin, E. F., DeBoy, R. T., Ravel, J., Paulsen, I. T., Kolonay, J. F., Brinkac, L., Beanan, M., Dodson, R. J., Daugherty, S. C., Madupu, R., Angiuoli, S. V., Durkin, A. S., Haft, D. H., Vamathevan, J., Khouri, H., Utterback, T., …

Fraser, C. M. (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. *Journal of Bacteriology*, *187*(7), 2426–2438. https://doi.org/10.1128/JB.187.7.2426-2438.2005

Gillespie, M. T., Lyon, B. R., Loo, L. S. L., Matthews, P. R., Stewart, P. R., & Skurray, R. A. (1987). Homologous direct repeat sequences associated with mercury, methicillin, tetracycline and trimethoprim resistance determinants in *Staphylococcus aureus*. *FEMS Microbiology Letters*, *43*(2), 165–171. https://doi.org/10.1111/j.1574-6968.1987.tb02117.x

Giulieri, S. G., Baines, S. L., Guerillot, R., Seemann, T., Gonçalves da Silva, A., Schultz, M., Massey, R. C., Holmes, N. E., Stinear, T. P., & Howden, B. P. (2018). Genomic exploration of sequential clinical isolates reveals a distinctive molecular signature of persistent *Staphylococcus aureus* bacteraemia. *Genome Medicine*, *10*(1), 65. https://doi.org/10.1186/s13073-018-0574-x

Giulieri, S. G., Tong, S. Y. C., & Williamson, D. A. (2020). Using genomics to understand meticillin- and vancomycin-resistant *Staphylococcus aureus* infections. *Microbial Genomics*, *6*(1). https://doi.org/10.1099/mgen.0.000324

Goto, M., Schweizer, M. L., Vaughan-Sarrazin, M. S., Perencevich, E. N., Livorsi, D. J., Diekema, D. J., Richardson, K. K., Beck, B. F., Alexander, B., & Ohl, M. E. (2017). Association of Evidence-Based Care Processes With Mortality in *Staphylococcus aureus* Bacteremia at Veterans Health Administration Hospitals, 2003-2014. *JAMA Internal Medicine*, *177*(10), 1489. https://doi.org/10.1001/jamainternmed.2017.3958

Guérillot, R., Siguier, P., Gourbeyre, E., Chandler, M., & Glaser, P. (2014). The Diversity of Prokaryotic DDE Transposases of the Mutator Superfamily, Insertion Specificity, and Association with Conjugation Machineries. *Genome Biology and Evolution*, *6*(2), 260–272. https://doi.org/10.1093/gbe/evu010

Haag, A. F., Fitzgerald, J. R., & Penadés, J. R. (2019). *Staphylococcus aureus* in Animals. *Microbiology Spectrum*, *7*(3). https://doi.org/10.1128/microbiolspec.GPP3-0060-2019

Harmer, C. J., & Hall, R. M. (2020). IS *26* Family Members IS *257* and IS *1216* Also Form Cointegrates by Copy-In and Targeted Conservative Routes. *MSphere*, *5*(1), e00811-19, /msphere/5/1/mSphere811-19.atom. https://doi.org/10.1128/mSphere.00811-19

Hennig, S., & Ziebuhr, W. (2008). A Transposase-Independent Mechanism Gives Rise to Precise Excision of IS256 from Insertion Sites in *Staphylococcus epidermidis*. *Journal of Bacteriology*, *190*(4), 1488–1490. https://doi.org/10.1128/JB.01290-07

Hosseinkhani, F., Tammes Buirs, M., Jabalameli, F., Emaneini, M., & van Leeuwen, W. B. (2018). High diversity in SCCmec elements among multidrug-resistant *Staphylococcus*

haemolyticus strains originating from paediatric patients; characterization of a new composite island. *Journal of Medical Microbiology*, *67*(7), 915–921. https://doi.org/10.1099/jmm.0.000776

Hutchison, C. A., Merryman, C., Sun, L., Assad-Garcia, N., Richter, R. A., Smith, H. O., & Glass, J. I. (2019). Polar Effects of Transposon Insertion into a Minimal Bacterial Genome. *Journal of Bacteriology*, *201*(19), e00185-19, /jb/201/19/JB.00185-19.atom. https://doi.org/10.1128/JB.00185-19

Jansen, A., Türck, M., Szekat, C., Nagel, M., Clever, I., & Bierbaum, G. (2007). Role of insertion elements and yycFG in the development of decreased susceptibility to vancomycin in *Staphylococcus aureus*. *International Journal of Medical Microbiology*, *297*(4), 205–215. https://doi.org/10.1016/j.ijmm.2007.02.002

Jiang, N., Li, J., Feßler, A. T., Wang, Y., Schwarz, S., & Wu, C. (2019). Novel pseudo-staphylococcal cassette chromosome *mec* element (φSCC *mec* T55) in MRSA ST9. *Journal of Antimicrobial Chemotherapy*, *74*(3), 819–820. https://doi.org/10.1093/jac/dky457

Joshi, G. S., Spontak, J. S., Klapper, D. G., & Richardson, A. R. (2011). Arginine catabolic mobile element encoded speG abrogates the unique hypersensitivity of *Staphylococcus aureus* to exogenous polyamines: ACME speG confers spermine resistance in *S. aureus*. *Molecular Microbiology*, *82*(1), 9–20. https://doi.org/10.1111/j.1365-2958.2011.07809.x

Kahl, B. C., Becker, K., & Löffler, B. (2016). Clinical Significance and Pathogenesis of Staphylococcal Small Colony Variants in Persistent Infections. *Clinical Microbiology Reviews*, *29*(2), 401–427. https://doi.org/10.1128/CMR.00069-15

Katayama, Y., Ito, T., & Hiramatsu, K. (2000). A New Class of Genetic Element, *Staphylococcus* Cassette Chromosome mec, Encodes Methicillin Resistance in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, *44*(6), 1549–1555. https://doi.org/10.1128/AAC.44.6.1549-1555.2000

Kehrenberg, C., Aarestrup, F. M., & Schwarz, S. (2007). IS21-558 Insertion Sequences Are Involved in the Mobility of the Multiresistance Gene cfr. *Antimicrobial Agents and Chemotherapy*, *51*(2), 483–487. https://doi.org/10.1128/AAC.01340-06

Kim, Y. B., Seo, K. W., Son, S. H., Noh, E. B., & Lee, Y. J. (2019). Genetic characterization of high-level aminoglycoside-resistant *Enterococcus faecalis* and *Enterococcus faecium* isolated from retail chicken meat. *Poultry Science*, *98*(11), 5981–5988. https://doi.org/10.3382/ps/pez403

Kleinert, F., Kallies, R., Hort, M., Zweynert, A., Szekat, C., Nagel, M., & Bierbaum, G. (2017). Influence of IS256 on Genome Variability and Formation of Small-Colony Variants in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, *61*(8), e00144-17, e00144-17. https://doi.org/10.1128/AAC.00144-17

Kresse, A. U., Dinesh, S. D., Larbig, K., & Römling, U. (2002). Impact of large chromosomal inversions on the adaptation and evolution of *Pseudomonas aeruginosa* chronically colonizing cystic fibrosis lungs: Diversification of *Pseudomonas* in cystic fibrosis. *Molecular Microbiology*, *47*(1), 145–158. https://doi.org/10.1046/j.1365-2958.2003.03261.x

Kuroda, M., Sekizuka, T., Matsui, H., Ohsuga, J., Ohshima, T., & Hanaki, H. (2019). IS256-Mediated Overexpression of the WalKR Two-Component System Regulon Contributes to Reduced Vancomycin Susceptibility in a *Staphylococcus aureus* Clinical Isolate. *Frontiers in Microbiology*, *10*, 1882. https://doi.org/10.3389/fmicb.2019.01882

Lata, C., Girard, L., Parkins, M., & James, M. T. (2016). Catheter-Related Bloodstream Infection in End-Stage Kidney Disease: A Canadian Narrative Review. *Canadian Journal of Kidney Health and Disease*, *3*, 115. https://doi.org/10.1186/s40697-016-0115-8

Lee, J. Y. H., Monk, I. R., Pidot, S. J., Singh, S., Chua, K. Y. L., Seemann, T., Stinear, T. P., & Howden, B. P. (2016). Functional analysis of the first complete genome sequence of a multidrug resistant sequence type 2 *Staphylococcus epidermidis*. *Microbial Genomics*, *2*(9). https://doi.org/10.1099/mgen.0.000077

Li, C., Sun, F., Cho, H., Yelavarthi, V., Sohn, C., He, C., Schneewind, O., & Bae, T. (2010). CcpA Mediates Proline Auxotrophy and Is Required for *Staphylococcus aureus* Pathogenesis. *Journal of Bacteriology*, *192*(15), 3883–3892. https://doi.org/10.1128/JB.00237-10

Li, M., Cheung, G. Y. C., Hu, J., Wang, D., Joo, H., DeLeo, F. R., & Otto, M. (2010). Comparative Analysis of Virulence and Toxin Expression of Global Community-Associated Methicillin-Resistant *Staphylococcus aureus* Strains. *The Journal of Infectious Diseases*, *202*(12), 1866–1876. https://doi.org/10.1086/657419

Liu, J., Chen, D., Peters, B. M., Li, L., Li, B., Xu, Z., & Shirliff, M. E. (2016). Staphylococcal chromosomal cassettes mec (SCCmec): A mobile genetic element in methicillin-resistant *Staphylococcus aureus*. *Microbial Pathogenesis*, *101*, 56–67. https://doi.org/10.1016/j.micpath.2016.10.028

Loessner, I., Dietrich, K., Dittrich, D., Hacker, J., & Ziebuhr, W. (2002). Transposase-Dependent Formation of Circular IS256 Derivatives in *Staphylococcus epidermidis* and *Staphylococcus aureus*. *Journal of Bacteriology*, *184*(17), 4709–4714. https://doi.org/10.1128/JB.184.17.4709-4714.2002

Lyon, B. R., Gillespie, M. T., & Skurray, R. A. (1987). Detection and Characterization of IS256, an Insertion Sequence in *Staphylococcus aureus*. *Microbiology*, *133*(11), 3031–3038. https://doi.org/10.1099/00221287-133-11-3031

Lyon, Bruce R., May, J. W., & Skurray, R. A. (1984). Tn4001: A gentamicin and kanamycin resistance transposon in *Staphylococcus aureus*. *Molecular and General Genetics MGG*, *193*(3), 554–556. https://doi.org/10.1007/BF00382099

Maki, H., & Murakami, K. (1997). Formation of potent hybrid promoters of the mutant llm gene by IS256 transposition in methicillin-resistant *Staphylococcus aureus*. *Journal of Bacteriology*, *179*(22), 6944–6948. https://doi.org/10.1128/JB.179.22.6944-6948.1997

McEvoy, C. R. E., Tsuji, B., Gao, W., Seemann, T., Porter, J. L., Doig, K., Ngo, D., Howden, B. P., & Stinear, T. P. (2013). Decreased Vancomycin Susceptibility in *Staphylococcus aureus* Caused by IS *256* Tempering of WalKR Expression. *Antimicrobial Agents and Chemotherapy*, *57*(7), 3240–3249. https://doi.org/10.1128/AAC.00279-13

McGavin, M. J., Arsic, B., & Nickerson, N. N. (2012). Evolutionary blueprint for host- and niche-adaptation in *Staphylococcus aureus* clonal complex CC30. *Frontiers in Cellular and Infection Microbiology*, *2*. https://doi.org/10.3389/fcimb.2012.00048

Menendez-Gil, P., Caballero, C. J., Catalan-Moreno, A., Irurzun, N., Barrio-Hernandez, I., Caldelari, I., & Toledo-Arana, A. (2020). Differential evolution in 3′UTRs leads to specific gene expression in *Staphylococcus*. *Nucleic Acids Research*, *48*(5), 2544–2563. https://doi.org/10.1093/nar/gkaa047

Mertens, A., & Ghebremedhin, B. (2013). Genetic determinants and biofilm formation of clinical *Staphylococcus epidermidis* isolates from blood cultures and indwelling devises. *European Journal of Microbiology and Immunology*, *3*(2), 111–119. https://doi.org/10.1556/EuJMI.3.2013.2.4

Michel, B. (2005). After 30 Years of Study, the Bacterial SOS Response Still Surprises Us. *PLoS Biology*, *3*(7), e255. https://doi.org/10.1371/journal.pbio.0030255

Miragaia, M. (2018). Factors Contributing to the Evolution of mecA-Mediated β-lactam Resistance in Staphylococci: Update and New Insights From Whole Genome Sequencing (WGS). *Frontiers in Microbiology*, *9*, 2723. https://doi.org/10.3389/fmicb.2018.02723

Montanaro, L., Campoccia, D., Pirini, V., Ravaioli, S., Otto, M., & Arciola, C. R. (2007). Antibiotic multiresistance strictly associated with IS256 and ica genes in *Staphylococcus epidermidis* strains from implant orthopedic infections. *Journal of Biomedical Materials Research Part A*, *83A*(3), 813–818. https://doi.org/10.1002/jbm.a.31399

Nagel, M., Reuter, T., Jansen, A., Szekat, C., & Bierbaum, G. (2011). Influence of ciprofloxacin and vancomycin on mutation rate and transposition of IS256 in *Staphylococcus aureus*. *International Journal of Medical Microbiology*, *301*(3), 229–236. https://doi.org/10.1016/j.ijmm.2010.08.021

Nakatsuji, T., Chen, T. H., Butcher, A. M., Trzoss, L. L., Nam, S.-J., Shirakawa, K. T., Zhou, W., Oh, J., Otto, M., Fenical, W., & Gallo, R. L. (2018). A commensal strain of

Staphylococcus epidermidis protects against skin neoplasia. *Science Advances*, *4*(2), eaao4502. https://doi.org/10.1126/sciadv.aao4502

Needham, C., Noble, W. C., & Dyke, K. G. H. (1995). The Staphylococcal Insertion Sequence IS257 Is Active. *Plasmid*, *34*(3), 198–205. https://doi.org/10.1006/plas.1995.0005

Ou, J. J. J., Drilling, A. J., Cooksley, C., Bassiouni, A., Kidd, S. P., Psaltis, A. J., Wormald, P. J., & Vreugde, S. (2016). Reduced Innate Immune Response to a *Staphylococcus aureus* Small Colony Variant Compared to Its Wild-Type Parent Strain. *Frontiers in Cellular and Infection Microbiology*, *6*. https://doi.org/10.3389/fcimb.2016.00187

Pain, M., Hjerde, E., Klingenberg, C., & Cavanagh, J. P. (2019). Comparative Genomic Analysis of *Staphylococcus haemolyticus* Reveals Key to Hospital Adaptation and Pathogenicity. *Frontiers in Microbiology*, *10*, 2096. https://doi.org/10.3389/fmicb.2019.02096

Partridge, S. R., Kwong, S. M., Firth, N., & Jensen, S. O. (2018). Mobile Genetic Elements Associated with Antimicrobial Resistance. *Clinical Microbiology Reviews*, *31*(4), e00088-17, /cmr/31/4/e00088-17.atom. https://doi.org/10.1128/CMR.00088-17

Ray, M. D., Boundy, S., & Archer, G. L. (2016). Transfer of the methicillin resistance genomic island among staphylococci by conjugation: Transfer of the methicillin resistance. *Molecular Microbiology*, *100*(4), 675–685. https://doi.org/10.1111/mmi.13340

Rice, L. B., & Thorisdottir, A. S. (1994). The Prevalence of Sequences Homologous to IS256 in Clinical Enterococcal Isolates. *Plasmid*, *32*(3), 344–349. https://doi.org/10.1006/plas.1994.1076

Richardson, E. J., Bacigalupe, R., Harrison, E. M., Weinert, L. A., Lycett, S., Vrieling, M., Robb, K., Hoskisson, P. A., Holden, M. T. G., Feil, E. J., Paterson, G. K., Tong, S. Y. C., Shittu, A., van Wamel, W., Aanensen, D. M., Parkhill, J., Peacock, S. J., Corander, J., Holmes, M., & Fitzgerald, J. R. (2018). Gene exchange drives the ecological success of a multi-host bacterial pathogen. *Nature Ecology & Evolution*, *2*(9), 1468–1478. https://doi.org/10.1038/s41559-018-0617-0

Rolo, J., Worning, P., Nielsen, J. B., Bowden, R., Bouchami, O., Damborg, P., Guardabassi, L., Perreten, V., Tomasz, A., Westh, H., de Lencastre, H., & Miragaia, M. (2017). Evolutionary Origin of the Staphylococcal Cassette Chromosome mec (SCCmec). *Antimicrobial Agents and Chemotherapy*, *61*(6), e02302-16, e02302-16. https://doi.org/10.1128/AAC.02302-16

Rouch, D. A., Messerotti, L. J., Loo, L. S. L., Jackson, C. A., & Skurray, R. A. (1989). Trimethoprim resistance transposon Tn4003 from *Staphylococcus aureus* encodes genes for a dihydrofolate reductase and thymidylate synthetase flanked by three copies of IS257. *Molecular Microbiology*, *3*(2), 161–175. https://doi.org/10.1111/j.1365-2958.1989.tb01805.x

Schaaff, F., Reipert, A., & Bierbaum, G. (2002). An Elevated Mutation Frequency Favors Development of Vancomycin Resistance in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, *46*(11), 3540–3548. https://doi.org/10.1128/AAC.46.11.3540-3548.2002

Siguier, P., Varani, A., Perochon, J., & Chandler, M. (2012). Exploring Bacterial Insertion Sequences with ISfinder: Objectives, Uses, and Future Developments. In Y. Bigot (Ed.), *Mobile Genetic Elements* (Vol. 859, pp. 91–103). Humana Press. https://doi.org/10.1007/978-1-61779-603-6_5

Siguier, Patricia, Gourbeyre, E., & Chandler, M. (2014). Bacterial insertion sequences: Their genomic impact and diversity. *FEMS Microbiology Reviews*, *38*(5), 865–891. https://doi.org/10.1111/1574-6976.12067

Siguier, Patricia, Gourbeyre, E., Varani, A., Ton-Hoang, B., & Chandler, M. (2015). Everyman's Guide to Bacterial Insertion Sequences. *Microbiology Spectrum*, *3*(2). https://doi.org/10.1128/microbiolspec.MDNA3-0030-2014

Soumya, K. R., Sugathan, S., Mathew, J., & Radhakrishnan, E. K. (2016). Studies on coexistence of mec gene, IS256 and novel sasX gene among human clinical coagulase-negative staphylococci. *3 Biotech*, *6*(2), 233. https://doi.org/10.1007/s13205-016-0549-9

Suzuki, E., Kuwahara-Arai, K., Richardson, J. F., & Hiramatsu, K. (1993). Distribution of mec regulator genes in methicillin-resistant Staphylococcus clinical strains. *Antimicrobial Agents and Chemotherapy*, *37*(6), 1219–1226. https://doi.org/10.1128/AAC.37.6.1219

Symms, C., Cookson, B., Stanley, J., & Hookey, J. V. (1998). Analysis of methicillin-resistant *Staphylococcus aureus* by IS *1181* profiling. *Epidemiology and Infection*, *120*(3), 271–279. https://doi.org/10.1017/S0950268898008796

Takeuchi, F., Watanabe, S., Baba, T., Yuzawa, H., Ito, T., Morimoto, Y., Kuroda, M., Cui, L., Takahashi, M., Ankai, A., Baba, S., Fukui, S., Lee, J. C., & Hiramatsu, K. (2005). Whole-Genome Sequencing of *Staphylococcus haemolyticus* Uncovers the Extreme Plasticity of Its Genome and the Evolution of Human-Colonizing Staphylococcal Species. *Journal of Bacteriology*, *187*(21), 7292–7308. https://doi.org/10.1128/JB.187.21.7292-7308.2005

Tanabe, T., & Udou, T. (1995). Molecular epidemiologic analysis of plasmids in clinical isolates of methicillin-resistant *Staphylococcus aureus*. *Japanese Journal of Environmental Infections*, *10*, 18–22.

Tande, A. J., Osmon, D. R., Greenwood-Quaintance, K. E., Mabry, T. M., Hanssen, A. D., & Patel, R. (2014). Clinical Characteristics and Outcomes of Prosthetic Joint Infection Caused by Small Colony Variant Staphylococci. *MBio*, *5*(5), e01910-14. https://doi.org/10.1128/mBio.01910-14

Tansirichaiya, S., Mullany, P., & Roberts, A. P. (2016). PCR-based detection of composite transposons and translocatable units from oral metagenomic DNA. *FEMS Microbiology Letters*, *363*(18), fnw195. https://doi.org/10.1093/femsle/fnw195

Tekedar, H. C., Blom, J., Kalindamar, S., Nho, S., Karsi, A., & Lawrence, M. L. (2020). Comparative genomics of the fish pathogens *Edwardsiella ictaluri* 93-146 and *Edwardsiella piscicida* C07-087. *Microbial Genomics*, *6*(2). https://doi.org/10.1099/mgen.0.000322

Udou, T. (2004). Dissemination of nosocomial multiple-aminoglycoside-resistant *Staphylococcus aureus* caused by horizontal transfer of the resistance determinant (aacA/aphD) and clonal spread of resistant strains. *American Journal of Infection Control*, *32*(4), 215–219. https://doi.org/10.1016/j.ajic.2003.11.002

Wan, T.-W., Higuchi, W., Khokhlova, O. E., Hung, W.-C., Iwao, Y., Wakayama, M., Inomata, N., Takano, T., Lin, Y.-T., Peryanova, O. V., Kojima, K. K., Salmina, A. B., Teng, L.-J., & Yamamoto, T. (2017). Genomic comparison between *Staphylococcus aureus* GN strains clinically isolated from a familial infection case: IS1272 transposition through a novel inverted repeat-replacing mechanism. *PLOS ONE*, *12*(11), e0187288. https://doi.org/10.1371/journal.pone.0187288

Wan, T.-W., Khokhlova, O. E., Iwao, Y., Higuchi, W., Hung, W.-C., Reva, I. V., Singur, O. A., Gostev, V. V., Sidorenko, S. V., Peryanova, O. V., Salmina, A. B., Reva, G. V., Teng, L.-J., & Yamamoto, T. (2016). Complete Circular Genome Sequence of Successful ST8/SCCmecIV Community-Associated Methicillin-Resistant *Staphylococcus aureus* (OC8) in Russia: One-Megabase Genomic Inversion, IS256's Spread, and Evolution of Russia ST8-IV. *PLOS ONE*, *11*(10), e0164168. https://doi.org/10.1371/journal.pone.0164168

Wang, A., Zhou, K., Liu, Y., Yang, L., Zhang, Q., Guan, J., Zhong, N., & Zhuo, C. (2017). A potential role of transposon IS431 in the loss of mecA gene. *Scientific Reports*, *7*(1), 41237. https://doi.org/10.1038/srep41237

Wang, L., Si, W., Xue, H., & Zhao, X. (2018). Characterization of a functional insertion sequence ISSau2 from *Staphylococcus aureus*. *Mobile DNA*, *9*(1), 3. https://doi.org/10.1186/s13100-018-0108-5

Watanabe, S., Ito, T., Morimoto, Y., Takeuchi, F., & Hiramatsu, K. (2007). Precise Excision and Self-Integration of a Composite Transposon as a Model for Spontaneous Large-Scale Chromosome Inversion/Deletion of the *Staphylococcus haemolyticus* Clinical Strain JCSC1435. *Journal of Bacteriology*, *189*(7), 2921–2925. https://doi.org/10.1128/JB.01485-06

Wilson, L. K., Coombs, G. W., Christiansen, K., Grubb, W. B., & O'Brien, F. G. (2016). Characterization of a novel staphylococcal cassette chromosome composite island from

community-associated MRSA isolated in aged care facilities in Western Australia. *Journal of Antimicrobial Chemotherapy*, *71*(12), 3372–3375. https://doi.org/10.1093/jac/dkw317

Wu, S., de Lencastre, H., & Tomasz, A. (1998). Genetic Organization of the mecA Region in Methicillin-Susceptible and Methicillin-Resistant Strains of *Staphylococcus sciuri*. *Journal of Bacteriology*, *180*(2), 236–242. https://doi.org/10.1128/JB.180.2.236-242.1998

Yin, Y., Chen, H., Li, S., Gao, H., Sun, S., Li, H., Wang, R., Jin, L., Liu, Y., & Wang, H. (2019). Daptomycin resistance in methicillin-resistant *Staphylococcus aureus* is conferred by IS256 insertion in the promoter of mprF along with mutations in mprF and walK. *International Journal of Antimicrobial Agents*, *54*(6), 673–680. https://doi.org/10.1016/j.ijantimicag.2019.08.021

Yui Eto, K., Firth, N., Davis, A. M., Kwong, S. M., Krysiak, M., Lee, Y. T., O'Brien, F. G., Grubb, W. B., Coombs, G. W., Bond, C. S., & Ramsay, J. P. (2019). Evolution of a 72-Kilobase Cointegrant, Conjugative Multiresistance Plasmid in Community-Associated Methicillin-Resistant *Staphylococcus aureus* Isolates from the Early 1990s. *Antimicrobial Agents and Chemotherapy*, *63*(11), e01560-19, /aac/63/11/AAC.01560-19.atom. https://doi.org/10.1128/AAC.01560-19

Zhang, S., Sun, X., Chang, W., Dai, Y., & Ma, X. (2015). Systematic Review and Meta-Analysis of the Epidemiology of Vancomycin-Intermediate and Heterogeneous Vancomycin-Intermediate *Staphylococcus aureus* Isolates. *PLOS ONE*, *10*(8), e0136082. https://doi.org/10.1371/journal.pone.0136082

Ziebuhr, W., Krimmer, V., Rachid, S., Lossner, I., Gotz, F., & Hacker, J. (1999). A novel mechanism of phase variation of virulence in *Staphylococcus epidermidis*: Evidence for control of the polysaccharide intercellular adhesin synthesis by alternating insertion and excision of the insertion sequence element IS256. *Molecular Microbiology*, *32*(2), 345–356. https://doi.org/10.1046/j.1365-2958.1999.01353.x

CONCLUSIONS

The outcomes of this research provide profound understanding of the evolutionary context for how staphylococcal pathogens adapt to their environments. This knowledge is a result of the ever-improving next generation sequencing technologies and bioinformatic tools that allow us to sequence hundreds of bacterial isolates. Overall, I have shown that different species of *Staphylococcus*, which thrive in different environments from pets to humans to animal-associated foods, vary greatly in their genomic composition: virulence potential, antimicrobial resistance, mutation rates and recombination rates. Data from whole genome sequences are critical to make accurate predictions of the emergence of high-risk bacterial clones and thus inform health professionals on effective diagnostic and treatment options.

In Chapter 1, I demonstrate that *S. pseudintermedius* populations in the United States are composed of several distinct and deeply branching lineages that colocalize together and primarily cause various skin, ear, and urinary infections in dogs and cats. Due to a high degree of antimicrobial resistance in these populations, particularly in Texas, the results support the hypothesis that strains present in pets may act as reservoirs of resistance that could be transmitted between pets and humans or between animals.

Chapter 2 describes the population structure and dynamics of *S. aureus* from bloodstream infections. My work showed that the emergence and persistence of two lineages (CC5 and CC8) over eight years in New Hampshire were driven by two different mechanisms. First, phylogenetic and demographic analysis indicates that these lineages arose during periods of significant clonal expansion. Second, high levels of recombination allowed for the exchange of

genetic material such as the mobile element SCC*mec* that confer resistance to methicillin and other antimicrobials.

In Chapter 3, I present the evolutionary relationships among CoNS, which make up a large majority of the *Staphylococcus* genus. Many of these species are known to be clinically relevant yet are poorly studied compared to *S. aureus*, which is likely due to a historically perceived lower pathogenic potential. Homologous recombination within and between species plays a major role in CoNS evolution, but recombination rates varied between species. Hence, certain *Staphylococcus* species and ecological sources (animals and animal-derived food) function as hubs of gene flow and major reservoir of genetic diversity for the entire genus.

Finally, I present a review of the most current literature on the diversity and functions of insertion sequences (ISs) in *Staphylococcus*. There are currently 17 known ISs of Staphylococcus, many of which have unique mechanisms for transposing between different regions of the genome. These ISs also function in modulating gene expression. Many ISs are tied to the mobility of antimicrobial resistance genes, often in the form of composite transposons.

Collectively, this dissertation highlights the application of population genomics at both local and global levels. It considers a variety of ecological niches that *Staphylococcus* species are able to successfully colonize. These results will be useful for future studies of staphylococcal taxonomy, antimicrobial resistance, and virulence, which will be critical to accurately diagnosing and treating infections caused by these organisms.