Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Michelle Su Date

Evolution of *Staphylococcus aureus* phenotypes

By

Michelle Su
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Microbiology and Molecular Genetics

Timothy D. Read
Advisor

Joanna B. Goldberg
Committee Member

Levi T. Morran
Committee Member

Cassandra L. Quave
Committee Member

Jorge E. Vidal
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Evolution of *Staphylococcus aureus* phenotypes

By

Michelle Su
B.S., Georgia Institute of Technology, 2013

Advisor: Timothy D. Read, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Graduate Division of Biological and Biomedical Sciences
Microbiology and Molecular Genetics
2020

Abstract

Evolution of *Staphylococcus aureus* phenotypes
By Michelle Su

Staphylococcus aureus poses a significant threat partly due to its incredible ability to adapt and evolve. Its evolution can be defined by parallel evolution, epistasis, and/or trade-offs between phenotypes. Understanding the selective pressures and constraints that influence the fitness landscape and mutational space *S. aureus* has access to will be a step towards better infection management and treatment. Antimicrobial resistance phenotype prediction is largely accurate due to well-characterized resistance determinants. Errors have been mainly attributed to the variability of culture-based antimicrobial susceptibility tests, but undefined epistasis can also dramatically affect the phenotypic manifestation of a resistance gene. Delta-toxin production in *S. aureus*, which is linked to atopic dermatitis, was found to be highly associated with clonal complex, indicating epistasis is the primary constraint is its evolution. Additionally, a genome-wide association study identified *carA*, carbamoyl-phosphate synthase small chain, as necessary for the expression of delta-toxin, confirming the link between virulence and metabolic status. Vancomycin resistance in *S. aureus* was confirmed to be driven largely by parallel evolution using experimental evolution in three backgrounds (CC1, CC5, CC8), but epistasis nonetheless affected the mutational patterns observed in these strains. In addition, differential fitness costs were seen in different genetic backgrounds during the progression from lower levels of resistance to higher levels of resistance. Virulence evolution of *S. aureus* in *Caenorhabditis elegans* was determined to exert selective pressure on virulence regulators and metabolic loci. In addition, there were trade-offs found between virulence and antibiotic exposure, which could be potentiated by host exposure. These studies together highlight the complex interplay between genetic background and outside selective forces and furthers our understanding the evolutionary dynamics of *S. aureus* phenotypes.

Evolution of *Staphylococcus aureus* phenotypes

By

Michelle Su
B.S., Georgia Institute of Technology, 2013

Advisor: Timothy D. Read, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Graduate Division of Biological and Biomedical Sciences
Microbiology and Molecular Genetics
2020

Acknowledgements

As this five year undertaking comes to a close, I'd like to take time to thank the people that have helped shape my scientific and personal discovery of what kind of adult I want to be when I grow up. It is the kind that gets to make people call me Doctor!

To my advisor Tim Read, thank you for being a beacon of light during what I would often describe as a nebulous trek through science and towards my Ph.D. Not everyone has the good fortune to find a scientific mentor that believes in their potential and gives them the space to grow and explore. Thank you for allowing me to be human during these times, for championing a healthy lab culture filled with drinking and free food, and proving that science does not have to come at the cost of self. I could not ask for a more supportive advisor. Thank you for your investment in me.

To the members of the Read laboratory, current and past, as well as the Satola laboratory, it has been a blast. First and foremost, thank you to Sarah Satola for the additional scientific and career guidance and for inspiring my love for New York Times crosswords. Thank you to Jessica Peterson for training me when I first started. Thank you to Robert Petit for being an accomplice to our lab happy hours and for lending an ear when I needed it. Thank you to Jon Moller for providing another brain to troubleshoot my experiments and Michelle Davis for providing another set of hands to do all those tedious experiments I kept planning. To all the others, Ashley Alexander, Emily Wissel, Vishnu Raghuram, and Alex Page, thank you for helping make graduate school as fun as it can be.

To the MMG program, I hope you never change. A special thanks to my committee members, Joanna, Levi, Cassie, and Jorge for being a sounding board for my scientific ideas and career goals and also to Bill Shafer, who gave me the opportunity to foster my interest in public health and antimicrobial

resistance and has provided invaluable mentorship. To all the friends I have made along the way, thank you and I wish you all the best.

To my friends, especially Faridah, Jairo, Big Joe, and Poprocks, thank you for making Atlanta special and providing a reprieve from the sometimes-all-consuming nature of graduate school. I know I can always rely on you all to hype up my crazy schemes.

To my father Zhan and mother Xue, who has always dreamed of this Ph.D., maybe even to a fault, I offer you congratulations by association and thank you for instilling in me a love for education and perseverance. To my brothers David and Steven, thank you and I follow in your footsteps. To the remainder of my extended family, who are too numerous to name here, thank you for all your support. A special thanks to Shirley for not thinking her Pharm.D. is a doctorate and letting me be the first doctor of the So family. I dedicate this dissertation to the next generation, especially my nephews Euler and Maxwell.

Table of Contents

List of Figures and Tables

Chapter I. Introduction

As pathogens have continued to evolve, so too has our understanding of how they cause disease. With the advent of genome sequencing, we have been able to ask broader questions about how *Staphylococcus aureus* has evolved over time and how the evolution of phenotypes such as antibiotic resistance and virulence is achieved. What has emerged is a complex picture of how different evolutionary forces interact with genetic variation to define the relationship from genotype to phenotype. In addition, trade-offs between different phenotypes govern the adaptation of *S. aureus* in different niches. This dissertation aims to investigate the evolution of *S. aureus* using traditional laboratory approaches and genomic approaches with an eye towards the prediction of phenotypes and the future evolution of *S. aureus*.

Staphylococcus aureus

Staphylococcus aureus is a Gram-positive coccus that can live as a human commensal or a pathogen. Throughout its history, it has shown remarkable ability to adapt to the changing environments caused by human healthcare. Although originally susceptible to most antibiotics, *S. aureus* has been the cause of many epidemic waves of antibiotic resistant infections (Chambers & Deleo, 2009). In the 1950s and 1960s, penicillin resistant *S. aureus* dominated in hospitals and communities (Rountree & Freeman, 1955). Shortly after the introduction of methicillin in 1961, there was an emergence of methicillin-resistant *S. aureus* (MRSA) in hospitals (Crossley, Landesman & Zaske, 1979; Crisostomo et al., 2001). Increasing use of vancomycin against MRSA strains saw the appearance of vancomycin-intermediate *S. aureus* (VISA) (Hiramatsu et al., 1997a) and vancomycin-resistant *S. aureus* (VRSA) (Weigel et al., 2003). In the 1990s, MRSA infections, which had been largely confined to hospitals, began to appear in the community (Herold et al., 1998; "From the Centers for Disease Control and Prevention. Four pediatric deaths from community-acquired methicillin-resistant Staphylococcus aureus--Minnesota

and North Dakota, 1997-1999," 1999). As *S. aureus* spread globally, there has been a continued evolution and stratification of the species: Major MRSA clones have been shown to descend from successful epidemic Methicillin-sensitive *S. aureus* (MSSA) strains with the majority of infections being attributed to a few clonal complexes (CC): CC1, CC5, CC8, CC9, CC12, CC15, CC22, CC25, CC30, CC45, and CC51/121 (Enright et al., 2002; Deurenberg & Stobberingh, 2008), though the incidence of each has varied dramatically over time and may be linked to the acquisition of new traits such antibiotic resistance. Thus, understanding how *S. aureus* evolves and adapts to changing environments is crucial.

How do phenotypes evolve?

Gene variation and population dynamics

The five major mechanisms of evolution are mutation, genetic drift, gene flow, non-random mating, and natural selection. As bacteria reproduce asexually, I will focus on the mechanisms other than non-random mating in discussing the evolutionary dynamics of *S. aureus*.

The *S. aureus* genome consists of a 2.8 Mbp chromosome and a variable number of episomal elements (plasmids). Gene flow or horizontal gene transfer (HGT) represents a powerful force in driving quick adaptation. Mobile genetic elements (MGEs) comprise on average 15-20% of the *S. aureus* genome (Lindsay, 2014), but lineages vary dramatically (Lindsay et al., 2006; Goerke et al., 2009). MGEs, which include bacteriophages, pathogenicity islands (SaPI), plasmids, transposons, and staphylococcal cassette chromosomes (SCC), are known contributors to virulence and resistance. Bacteriophages and SaPIs carry superantigens, toxins, and immune-modulatory proteins such as enterotoxins A-C (Betley & Mekalanos, 1985), Panton-Valentine leukocidin (Kaneko et al., 1998), toxic shock syndrome toxin-1, complement inhibitory protein, chemotaxis inhibitory protein, and staphylokinase (van Wamel et

al., 2006). SCC*mec* (SCC elements containing the mecA gene encoding PBP2a) introduction into MSSA strains and their stable maintenance generates MRSA (Katayama, Ito & Hiramatsu, 2000). Transposon insertions in these SCC elements can also contribute other antibiotic resistance genes. VRSA (Weigel et al., 2003) is also caused by acquisition of a transposon, which carries *vanA*, from enterococci.

Genetic drift in *S. aureus* is dominated by two phenomenon: transmission and infection bottlenecks and geographic expansion. Host species jumps are rare but evolutionary reconstruction suggests several events in the past 10,000 years (Sakwinska et al., 2011; Richardson et al., 2018). Transmission bottlenecks seem to be of variable size (Hall et al., 2019) and likely depend on several factors such as host susceptibility, duration of contact, and bacterial burden. Infections require large inoculums for initiation and are a stringent bottleneck as bacteria need to survive host immune responses (Prajsnar et al., 2012; McVicker et al., 2014; Pollitt et al., 2018). A notable example of *S. aureus* expansion is that of USA300 in North America, first detected in 1999 in a Mississippi prison (Centers for Disease Control and Prevention (CDC), 2001) and now the predominant MRSA clone in the United States (Diekema et al., 2014). As USA300 spread from the east coast to west coast, there was a drop in genetic diversity due to serial founder effects (Challagundla et al., 2018).

Mutation and natural selection are intimately linked. Mutation gives natural selection the population variation to act on. Mutations occur during errors in DNA replication, and mutation rate estimates in *S. aureus* vary from 2.0×10^{-6} to 4.6×10^{-6} per site per year (Smyth et al., 2010; Harris et al., 2010; Nübel et al., 2010; Young et al., 2012). A mechanism for increased mutation is knockout of DNA repair pathways (O'Neill & Chopra, 2002; Prunier & Leclercq, 2005), but environmental factors such as nutrient limitation (Shapiro, 1984; Mittler & Lenski, 1990) and SOS induction by quinolone antibiotics (Phillips et al., 1987) or other chemicals have also been shown to increase adaptive mutation rates

(Martinez & Baquero, 2000). Natural selection of *S. aureus* acts on a fitness differential between strains. However, fitness is a non-static determinant is a combination of environmental conditions, growth rates, antibiotic resistance, phage resistance, among others.

Epistasis and how it affects evolution: convergent vs lineage-centric

Discussion of fitness differentials between mutations requires acknowledgement of the relationship between genotype and phenotype. Epistasis, defined as the interaction of two or more genes in determining phenotype (Bateson, 1909), plays a key role in defining the fitness landscape (Wright, 1932) and available mutational space. The sensitivity of each phenotype to epistasis has to be determined empirically. In cases of parallel evolution where phenotypes have emerged in multiple lineages of *S. aureus*, the effect of epistasis to natural selection is low. An example of such a phenotype is antibiotic resistance. Methicillin resistance via acquisition of SCC*mec* has occurred multiple times in *S. aureus* history (Enright et al., 2002). VISA has been found in all major lineages of *S. aureus* (Hiramatsu et al., 1997b; Howden et al., 2006, 2008; Mwangi et al., 2007; Alam et al., 2014; Gardete & Tomasz, 2014). Other phenotypes are defined strongly by lineage. Delta-toxin production evolution in *S. aureus* shows a strong phylogenetic signal (Su et al., 2020). Other phenotypes likely to be strongly affected by epistasis include those that are unusually elevated in certain clonal complexes such as persister formation in CC30 (Liu et al., 2020).

Long-term evolution and persistence of *S. aureus* strains

Fitness

Fitness measurements can help elucidate the probabilities for long term persistence of evolved strains outside of direct selection by antibiotic therapy. Antibiotic resistance is generally a costly trait to evolve (Andersson & Levin, 1999; MacLean & Vogwill, 2014; Melnyk, Wong & Kassen, 2015), though there

have been reports of cost-free or low-cost resistance (Wichelhaus et al., 2002). In the case of vancomycin resistance in *S. aureus*, the prevalence of VISA and VRSA is low due to the high burden of resistance. VISA is often the result of within-host evolution from antibiotic therapy of a MRSA infection (Hu, Peng & Rao, 2016), and there have only been 14 cases of VRSA in the United States (McGuinness, Malachowa & DeLeo, 2017) since the first documented case in 2002 (Chang et al., 2003; Weigel et al., 2003). In addition, it is believed that MRSA strains were limited to hospital settings in the past due to the fitness costs of SCC*mec*. The appearance of SCC*mec* type IV, which is smaller and thus imposes a lower fitness cost, has contributed to the success of USA300 as a community-associated pathogen (Lee et al., 2007; Diep et al., 2008).

However, the determination of resistance maintenance is complicated by compensatory mutation to reduce fitness costs (Hernando-Amado et al., 2017) as well as co-selection by metals and biocides (Baker-Austin et al., 2006; Wales & Davies, 2015), genetic linkage to other resistance markers or beneficial genes (Enne et al., 2004; Dionisio et al., 2005; Yates et al., 2006), and selection from subminimal concentrations of antibiotics (Thiele-Bruhn, 2003; Cabello, 2006; Drlica & Zhao, 2007; Li et al., 2008; Kümmerer, 2008). These factors help explain the persistence of resistance after discontinuation of antibiotics (trimethoprim resistance in Sweden (Sundqvist et al., 2010), sulfonamide resistance (Enne et al., 2001) and streptomycin resistance (Bean et al., 2005) in the UK). Thus, predictions for the persistence of antibiotic resistances genes must rely not only on fitness costs but must be combined with a more holistic view of the selection pressures bacteria face.

Constraints to evolution: Opposed selection pressures (virulence vs antibiotic resistance, immune system vs virulence)

Often, evolution of one phenotype is at the expense of another. Trade-offs have been characterized in many organisms, *S. aureus* notwithstanding (Geisinger & Isberg, 2017). Null mutations in virulence regulators are common in *S. aureus* clinical infections, suggesting it is an adaptive response to the host. These mutations appear to be an evolutionary trade-off to reduce the damage to the host caused by toxins, which elicits strong immune response, in favor of slower, persistent growth. Mutations in the *agr* operon, which controls extracellular toxins, are well documented (Traber et al., 2008) in chronic infections such as osteomyelitis (Suligoy et al., 2018), cystic fibrosis (Kahl et al., 2003), atopic dermatitis (Soong et al., 2015), and bacteremia (Fowler et al., 2004). These mutants trigger less inflammatory immune responses than *agr*+ strains and form more robust biofilms (Vuong et al., 2000). Mutation in another virulence regulator *rsp* (Li et al., 2015; Liu & Sun, 2020) has also been shown to allow prolonged survival in the host at the cost of cytotoxicity (Young et al., 2012; Das et al., 2016). Similar to *agr*, mutation in *rsp* has a profound impact on biofilm formation (Lei et al., 2011). Lastly, a mechanism agnostic approach found that lower toxicity in *S. aureus* is associated with a higher within-host fitness and persistence in the host (Laabei et al., 2015).

Some antibiotic resistant *S. aureus* phenotypes are attenuated in virulence. This has been demonstrated for vancomycin and other glycopeptides (McCallum et al., 2006; Majcherczyk et al., 2008; Peleg et al., 2009). However, the lower virulence likely is not wholly attributable to the development of resistance and its associated fitness costs as other phenotypic changes are seen in these isolates (Howden et al., 2010). Nonetheless, trade-offs between virulence and antibiotic resistance due to direct epistatic interactions have been demonstrated. Methicillin resistance induces cell wall changes that interfere with quorum sensing through *agr* and thus lowers toxicity relative to methicillin resistance level (Collins et al., 2010; Rudkin et al., 2012). Thus, understanding the trade-offs that exist between these phenotypes will be key to the development of better therapies.

Traditional laboratory approaches

Traditional molecular genetic lab techniques presuppose a simple causal relationship between gene and phenotype. Typically, the relationships between genes and phenotypes are probed by experimentally introducing a single mutation at a time. However, this ignores the varying selective pressures that exist in the environment, in the host, and in transmission that determine whether mutations are fit enough to survive. In addition, experiments are typically limited in scope: using rich media or only one genetic background. The *S. aureus* genome is remarkably diverse, and epistasis, as discussed above, is an important determinant in the evolution of phenotypes.

Experimental evolution

Experimental evolution is a powerful tool that can be used to investigate the mechanisms of clinically relevant phenotypes, and in the case of antimicrobial treatment development, the frequency. It is a top-down approach that selects for the phenotype of interest and then finds the genetic determinants. Thus far, it has been used to elucidate important genes in daptomycin resistance (Friedman, Alder & Silverman, 2006; Lasek-Nesselquist et al., 2019) and compensatory rescue of fusidic acid resistance (Nagaev et al., 2001) as well assess the potential for cross-resistance to the innate immune system if antimicrobial peptides are used as a treatment (Dobson, Purves & Rolff, 2014; Johnston, Dobson & Rolff, 2016; Makarova et al., 2018). Though rarer, experimental evolution has also been used to explore population dynamics. Alternating antibiotic treatments have been demonstrated to slow the evolution of multidrug resistance (Kim, Lieberman & Kishony, 2014). New host adaptation is possible due to mutation accumulation during infection that can rapidly sweep the population (Bacigalupe et al., 2019). Finally, it was shown that spontaneous intra-species competition can lead to rapid evolution of VISA-like strains (Koch et al., 2014).

Genomic approaches

Unlike traditional lab approaches, genomic approaches use standing natural variation to investigate associations retrospectively and are only limited by the granularity of data. However, the breadth of data underscores one of the primary challenges, which is having sufficient power to detect meaningful associations. Appropriate statistical models must also be developed that reflect the real world so as to avoid spurious conclusions. In the end, all associations found by bioinformatic approaches must be validated in the laboratory.

Ancestral reconstruction, GWAS, GWES

Ancestral reconstruction extrapolates backwards in time from current measured traits to shared ancestors. It can be used to test hypotheses about adaptation (Coddington, 1988), characterize the number of times a trait has evolved independently (Donoghue, 1989), and elucidate whether there has been differential selection among clades of the phylogenetic tree. To determine ancestral states, however, an appropriate model of evolution must be chosen *a priori*. In addition, accuracy decreases with time elapsed from the ancestor states being reconstructed or with sufficiently fast trait evolution, which obscure underlying phylogenetic relationships. Several other issues limit its applicability of use: accuracy of phylogeny (except in Bayesian methods), need for independent probability estimates for the gain or loss of traits, and complete taxon sampling (Omland, 1999).

Genome-wide association studies (GWAS) are a potentially unbiased method to identify genes associated with disease or a particular phenotype. Genetic variants, commonly single nucleotide polymorphisms or k-mers, in a genome are tested for significant association with a trait of interest. The inclusion of the whole genome makes GWAS more powerful than traditional linkage studies in

studying multi-factorial illnesses while still being able to detect strong signals within a population. GWAS has been used successfully in human populations to study susceptibility to myocardial infarction (Ozaki et al., 2002) and age-related macular degeneration (Eagle, 2006). In bacterial populations, it has been used successfully to identify a phage associated with hypervirulence in *Neisseria meningitidis (Bille et al., 2005)* as well as a host specificity factor in *Campylobacter (Sheppard et al., 2013)*. In those cases and in the use of GWAS to identify antibiotic resistance determinants, the studied phenotypes have been highly penetrant with large effect sizes, allowing for the easy discrimination of the variant from other variants tested. In studying more complex traits, maintaining power while accounting population structure, multiple testing, and linkage disequilibrium remains a significant obstacle in elucidating genetic determinants (Power, Parkhill & de Oliveira, 2017). Nonetheless, at the time of writing, several softwares exist that aim to tackle these unique problems in bacterial GWAS: bugwas (Earle et al., 2016), DBGWAS (Jaillard et al., 2018), SEER/pyseer (Lees et al., 2016, 2018), hogwash (Saund & Snitkin, 2020), treeWAS (Collins & Didelot, 2018).

As genes and proteins are not isolated and independent entities within cells but exist as broad and vast networks (Boucher & Jenna, 2013), genome-wide epistasis and co-selection study (GWES) aim to reveal sites that are co-evolving under shared selection pressures or due to epistatic interactions (effect of one gene is dependent on other genes). It has been used to identify epistasis between virulence and biofilm genes in *Vibrio parahaemolyticus* (Cui et al., 2015) as well as interacting networks of antibiotic resistance and virulence in *Streptococcus (Skwark et al., 2017)*. The main challenges with GWES are related to the fact as the number of variant pairs increases quadratically with the number of variants included, and similar to GWAS, population structure, linkage disequilibrium and multiple testing can confound analyses and limit power. Unlike GWAS, it is phenotype-free and may uncover insights about the evolutionary path of microbes outside of traditionally studied traits.

Prediction

Classification algorithms and machine learning approaches have been used for antibiotic resistance prediction in a variety of organisms (Su, Satola & Read, 2019) as well as toxicity (Laabei et al., 2014; Su et al., 2020) with great success. Regression has long been used in clinical settings to try to determine predictors for patient outcomes. These approaches have also recently been used to determine the importance of genetic determinants for a variety of phenotypes. A plasmid encoding enterotoxin genes *sed*, *sej*, and *ser* was found to have a protective effect against embolism in endocarditis, though whether it is due to unknown functions of the enterotoxin genes or the additional uncharacterized genes on the plasmid remains to be determined (Rasigade et al., 2018). Cytotoxicity and biofilm formation were found to be associated with a higher infection mortality in CC22 but not CC30, uncovering the role of epistasis in the evolution of pathogen virulence (Recker et al., 2017).

Accurate prediction of a phenotype ab initio from genome sequences is a testament to our understanding of the genetic basis of the phenotype and can be used to improve patient outcomes. Inaccurate predictions are an impetus to investigate further, though some phenotypes may prove to be too complicated to accurately model. The knowledge of which phenotypes are amenable to genome-based prediction will allow for appropriate use cases of empiric therapy in contrast with its current use stemming from a lack of information.

Specific aims

This dissertation aims to elucidate the dynamics of *S. aureus* evolution and the capability to predict phenotypes by incorporating both traditional and genomic approaches. I analyze the current status of the field of genome-based prediction of bacterial antibiotic resistance and its potential utility in the

clinic. Further, I investigate the genetic determinants of two phenotypes, delta-toxin and vancomycin-intermediate resistance, and the importance of genetic background and epistasis to the evolution of the phenotypes. Finally, trade-offs between virulence and antibiotic resistance were explored using a model host, *Caenorhabditis elegans*.

Chapter II. Genome-Based Prediction of Bacterial Antibiotic Resistance

Michelle Su[1,2,3], Sarah W. Satola[1,3,4], Timothy D. Read[1,3,4]

[1] Department of Infectious Diseases, Emory University, Atlanta, Georgia, USA

[2] Antimicrobial Resistance and Therapeutic Discovery Training Program, Atlanta, Georgia, USA

[3] Antibiotic Resistance Center, Emory University, Atlanta, Georgia, USA

[4] Emory Investigational Clinical Microbiology Laboratory, Atlanta, Georgia, USA

This work was published in Journal of Clinical Microbiology in 2019.

Citation:

Su M, Satola SW, Read TD. 2019. Genome-based prediction of bacterial antibiotic resistance. J Clin Microbiol 57:e01405-18. https://doi.org/10.1128/JCM.01405-18.

Abstract

Clinical microbiology has long relied on growing bacteria in culture to determine antimicrobial susceptibility profiles, but the use of whole genome sequencing for antibiotic susceptibility testing (WGS-AST) is now a powerful alternative. This review discusses the technologies that made this possible and presents results from recent studies to predict resistance based on genome sequences. We examine differences between calling antibiotic resistance profiles by simple presence or absence of previously known genes and SNPs against using approaches that deploy machine-learning and statistical models. Often the limitations to genome-based prediction arise from limitations of accuracy of culture-based AST in addition to an incomplete knowledge of the genetic basis of resistance. However, we need to maintain widespread phenotypic testing even as genome-based prediction becomes more widespread to ensure that the results do not diverge over time. We argue that standardization of WGS-AST by challenge with consistently phenotyped strain sets of defined genetic diversity is necessary to compare the efficacy of methods of prediction of antibiotic resistance based on genome sequences.

Introduction: the importance of AST

Antibiotic resistant bacterial infections are a global threat. Many previously manageable bacterial infections are becoming increasingly hard to treat, and the CDC has recently estimated that in the United States alone, at least two million people will be infected by a drug resistant bacterium each year, and at least 23,000 will die as a result("Antibiotic Resistance Threats in the United States, 2013 | Antibiotic/Antimicrobial Resistance | CDC"). Rising rates of resistance amplify the morbidity and economic burden associated with infections. Even successful treatment can come with increased complications as more toxic drugs of last resort like colistin are being used more frequently because bacteria are not susceptible to less toxic antibiotics. For management of infections in both the clinic

and community, accurate detection of antimicrobial resistance is necessary to guide treatment decisions.

Culture-based antimicrobial susceptibility testing (AST) is still the primary method employed by clinical laboratories. While there are other promising approaches for phenotypic detection and rapid non-sequencing genetic methods currently in use (e.g. PCR for resistance determinants)(van Belkum & Dunne, 2013), dramatic progress over the past five years in the applications of genomics has caught the attention of the clinical microbiology community. Whole genome sequencing for antimicrobial susceptibility testing (WGS-AST) offers the potential to provide rapid, consistent and accurate predictions of every known resistance phenotype for a strain, as well as simultaneously providing rich surveillance data. Recent reviews of the subject have focused on clinical standardization (van Belkum & Dunne, 2013; Goldberg et al., 2015; Ellington et al., 2017). Here we concentrate on the problem of prediction of resistance based only on genome sequence and consider the future symbiotic relationship between genomics and phenotypic-based AST.

Potential advantages of WGS-AST

WGS-AST follows selective culturing of the bacterium of interest from a clinical sample (**Figure 1**). AST following direct shotgun sequencing of clinical samples (metagenomic-AST) is also possible (e.g. Bradley et al., 2015), and is the subject of intense current research, but is more complex, expensive and prone to false negative results due to the potentially low abundance of the pathogen of interest relative to host DNA. Metagenomic-AST also includes approaches that enrich a library of antibiotic-resistance DNA fragments from complex clinical samples before sequencing (Doyle et al., 2018). Slow-growing or hard-to-culture bacteria (such as *M. tuberculosis* (Votintseva et al., 2017; Doyle et al., 2018)), are important early targets for metagenomic-AST because DNA sequencing may be easier and

faster than obtaining enough culture growth for phenotypic testing. This review will focus on the WGS-AST application, but many principles may also apply to other sequence-based AST approaches. Compared to culture-based AST or nucleic acid amplification tests (NAATs), which are often limited by the number of resistant phenotypes that can be determined from one test, WGS-AST can ascertain the antibiotic resistance phenotypes of the entire genome simultaneously, and phenotypes where multiple loci contribute can be easily screened (instead of performing multiplex PCR reactions). Once collected, the genome sequence data is stored digitally and can be queried for other purposes (e.g complete genome multi-locus sequence typing (cgMLST) genotype (Feijao et al., 2018) and virulence (Chen et al., 2016)). Genomes can be sequenced to very high levels of depth, giving very accurate sequence data. Unlike NAATs, there is no reliance on primers specificity for template amplification, reducing the possibility of false negative results. The accumulation of genomes in clinical laboratories creates a data source that can be used to survey the evolution of pathogens (Gardy & Loman, 2017). If new antibiotic resistance loci are discovered, these databases can be immediately scanned to understand how long these genes have been circulating and how they may have entered the clinical setting. One of the first examples of this type of retrospective surveillance investigation was performed for the emergence of mcr-1 colistin resistance in Germany (Falgenhauer et al., 2016; Schürch & van Schaik, 2017) where a database of 577 *Enterobacteriaceae* genomes from animals and humans was searched to find four previously undiagnosed colistin-resistant isolates.

Next generation sequencing technologies driving WGS-AST

Two waves of sequencing technology innovation, the "second generation" starting in the mid-2000s and the "third generation" from about 2010, transformed all aspects of genomics and set the stage for WGS-AST (Goodwin, McPherson & McCombie, 2016; Quince et al., 2017) (**Box 1; Figure 1**). Second generation instruments, represented by the currently dominant Illumina sequencing-by-synthesis

technology, sharply reduced the cost to generate data ($/Mbase), which led to large scale sequencing of thousands of pathogen genomes and the use of shotgun metagenomics for clinical diagnostics. Illumina sequence reads are short (<= 300 bp) (Reinert et al., 2015), typically paired-end and with low per base error rates (typically < 0.1%). Illumina sequencing allows deep shotgun coverage with high consensus accuracy, but *de novo* assembly typically results in genomes fragmented into multiple contigs and collapsed repeat regions (**Figure 1**).

Third generation, single-molecule sequencing, exemplified by Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) technologies, produces much longer reads (typically 5-100 kbp but there are recent reports of reads > 2Mb(Payne et al., 2018)). Genome assemblies generated with long reads have fewer gaps and often span lengthy repeat regions, allowing resolution of complex structural features such as tandem repeats and nested insertions (Giordano et al., 2017). Third generation technologies have higher cost per base and higher per-base error rates than Illumina (5-15%), although improvements to chemistry and base calling algorithms are reducing the differential (Rhoads & Au, 2015; Lu, Giordano & Ning, 2016). Error rates can in part be compensated by increasingly achievable higher read depths. For example, recent ONT studies generated >100x genome coverage and consensus error rates of < 0.08% (Cao et al., 2016; Lemon et al., 2017). This level of error is adequate for most WGS-AST methods based on gene-detection but may still be too high for many SNP-based methods. Hybrid assemblies (Wick et al., 2017) combine the accuracy of Illumina data and the gap-free assembly of ONT/PacBio and can achieve accuracies of > 99.9% but rely on creating multiple libraries per sample, hence increasing cost and data management complexity.

The ONT technology has particular potential for WGS-AST because DNA sequence data becomes available within minutes of starting the sequencing run. A number of clinical gene-based ASTs for

ONT have been piloted (e.g (Judge et al., 2016; Schmidt et al., 2016; van der Helm et al., 2017; Lemon et al., 2017)). At least two publications have demonstrated "streaming" AST, where the antibiotic resistance profile is updated in real time as the sequencing data is produced by the instrument (Cao et al., 2016; Břinda et al., 2018). Břinda et al (Břinda et al., 2018) used an indirect, lineage-based method for predicting the resistance phenotype, which has the potential to mitigate some of the potential errors in ONT data, especially SNP calling.

The outlook for clinical sequencing constantly shifts as existing technologies mature and become more cost effective, and new approaches emerge. Currently, Illumina is the dominant platform for WGS-AST. However, the approximate minimum cost of $80 per genome is still too high for routine use in clinical laboratories. Minimum costs are based on one-week turnaround times and batching many samples together for efficiency. Obtaining results within <24 hours for small numbers of samples is possible but comes at a price penalty. Further decreases in cost and turnaround times will be needed before sequencing will be economically viable for routine WGS-AST. New instrumentation, such as the ONT Flongle disposable flow cells might allow quicker, cheaper sequencing of smaller numbers of strains. Emerging synthetic long read technologies (e.g. 10X Genomics) that rely on the barcoding of long DNA fragments to associate the resulting short reads during assembly (Kuleshov, Snyder & Batzoglou, 2016) may provide the combination of read length, accuracy and reduced cost that opens the door for routine WGS-AST and eventually routine metagenomic-AST.

WGS-AST based on searching catalogs of resistance loci

At its essence WGS-AST attempts to estimate the phenotype that would have been ascertained if the strain were subjected to a gold standard culture-based antibiotic resistance test. The simplest approaches seek to classify the strain as either "susceptible" or "resistant" to specific antibiotics as

defined by CLSI or EUCAST (Ellington et al., 2017). More complex models have further tried to predict the minimum inhibitory concentration (MIC) of an antibiotic for the strain. The most straightforward approach is to use a "rules-based" classification based on the presence of one or more known antimicrobial resistance (AMR) genes or mutations (**Box 2; Table 1**). This requires cross-referencing the genome sequence against databases of antibiotic resistance determinants. Databases have been developed mostly from curation of the literature of molecular genetic studies that link antibiotic resistance phenotypes to genes (Xavier et al., 2016). Multi-species databases include: CARD (McArthur et al., 2013; Jia et al., 2017), ResFinder (Zankari et al., 2012) and its companion PointFinder (Zankari et al., 2017), ARG-ANNOT (Gupta et al., 2014), ARDB (Liu & Pop, 2009), MEGARes (Lakin et al., 2017), Resfams (Gibson, Forsberg & Dantas, 2015), RAST (Antonopoulos et al., 2017), BARRGD (Bacterial Antimicrobial Resistance Reference Gene Database; https://www.ncbi.nlm.nih.gov/bioproject/313047). In addition, there are databases developed for single organisms, such as Dream TB (Sandgren et al., 2009) and MUBII-TB-DB (Flandrois, Lina & Dumitrescu, 2014) for *M. tuberculosis*.

Software tools for rules-based matching of antibiotic resistance catalogs operate on data produced at two points in the workflow for next generation sequencing: raw sequence data and assembled contigs (**Figure 1**). Each has tradeoffs in terms of speed of result and accuracy.

Detecting resistance in raw reads obviates the need for assembly and can therefore reduce time-to-result if the algorithms are efficient. However, false positives may be introduced because of sequencing errors present in individual reads or DNA contamination from other organisms. Setting minimum thresholds for the number of reads to be considered a positive can help overcome read error problems. Software tools use differing strategies for processing raw reads. KmerResistance (Clausen et al., 2016)

matches k-mer subsequences of the resistance locus catalog against raw reads similarly split into the same length subsequence. SRST2 (Inouye et al., 2014) and Genefinder (Mason et al., 2018) use the efficient read alignment program Bowtie (Langmead & Salzberg, 2012) to map genes to the read set as a first step before enumerating SNPS and gene matches. ARIBA (Hunt et al., 2017) uses a partial de novo assembly after first recruiting individual reads that may map to target genes. The webtool Point-Finder (Zankari et al., 2017) identifies known point mutations after mapping reads against reference genomes. Mykrobe (Bradley et al., 2015) creates deBruijn graphs of contigs from raw data and matches against known genes but for speed omits generating a consensus sequence (Mason et al., 2018).

Genome assembly can either be *de novo* or by mapping to a reference strain. *De novo* assembly usually produces more fragmented genome assemblies but avoids the biases of building the assembly on an existing reference template. For reference-based assembly, single nucleotide polymorphism (SNP) detection becomes less accurate the greater the distance between isolate and reference, but de novo assembled DNA sequences are free from this bias. However, an antibiotic resistance gene may be missed when using a *de novo* assembly if it is split across multiple contigs. Multicopy genes associated with antibiotic resistance, such as rRNA, present a particular challenge. Mutation in just one or two copies is sometimes sufficient to impart a resistance phenotype (e.g in the case of azithromycin resistance caused by mutations in Neisseria 23S (Eyre et al., 2017)), yet assembly algorithms seek a consensus sequence and thus important genetic variants within repeats could be missed if repeats "collapse" into a single copy (Sinclair, Arnold & Woodford, 2003; Ellington et al., 2017). Using longer reads as input for assembly can overcome the problem of collapsed repeats. Increasing sequence coverage should also minimize these types of errors but more data comes with increased cost and slower computation time. This can be partially addressed by downsampling coverage, as after about

100x Illumina genome coverage of *Staphylococcus aureus* genomes, adding more data to assembly produced little extra benefit for assembly accuracy at the cost of significant declines in processing speed (Petit & Read, 2018). Most catalog-based software that takes assembled data as an input (e.g Typewriter (Mason et al., 2018), SSTAR (de Man & Limbago, 2016), CARD RGI (Jia et al., 2017), ARG-ANNOT (Gupta et al., 2014), ResFinder (Zankari et al., 2012) and ABRICATE https://github.com/tseemann/abricate) uses some form of BLAST alignment and results parsing, which usually takes a small fraction of the processor time used to construct the de novo assembly.

For many species and antibiotic-resistance phenotypes, there is good concordance between what is known about the genetic basis of resistance and the resistance phenotype. Rules-based WGS-AST has been shown to have high sensitivity and specificity (>95%) for many phenotypes across several pathogen species (**Table 1**) (Eyre et al., 2012; Köser et al., 2012; Holden et al., 2013; Ginn et al., 2013, 2014; Stoesser et al., 2013; Gordon et al., 2014; Leopold et al., 2014; Kos et al., 2015; Lee et al., 2015; Coll et al., 2015; ElMaraachli et al., 2015; Tyson et al., 2015; Zhao et al., 2015; Bradley et al., 2015; Pankhurst et al., 2016; Aanensen et al., 2016; Clausen et al., 2016; McDermott et al., 2016; Deng et al., 2016; Moran et al., 2017; Zankari et al., 2017; Sadouki et al., 2017; Miotto et al., 2017; Day et al., 2017; Quan et al., 2018; Neuert et al., 2018; Macedo et al., 2018; Harris et al., 2018; Mason et al., 2018), although with the caveat that studies varied widely in the number of strains tested and the within-species genetic diversity of the test set. However, there were some cases, e.g levofloxacin resistance in *P. aeruginosa* (**Table 1**) (Kos et al., 2015), where sensitivity and specificity were below 95%. In an extensive survey of the genetic basis of resistance in *M. tuberculosis*, Miotto et al (Miotto et al., 2017), classified the predictability of *M. tuberculosis* mutations into High-, Moderate- and Minimal- confidence. Therefore, rules-based approaches alone may not always be sufficient for accurate WGS-AST.

Model-based antibiotic resistance prediction

Most rules-based methods make a number of (often unacknowledged) assumptions about the phenotypes they attempt to predict. These include: 1) either a single genetic locus is responsible for the phenotype, or, if multiple loci are present, they do not interact in a complex manner (i.e. absence of epistasis (Trindade et al., 2009; Vogwill, Kojadinovic & MacLean, 2016; Knopp & Andersson, 2018)); 2) loci are highly penetrant and are not affected by the strain background; and 3) there is complete knowledge of the genetic basis of the phenotype. For a large number of cases, these assumptions don't completely hold. Some studies have attempted capture uncertainty in the genetic basis of resistance and reduce overfitting, using a variety of statistical modelling and machine learning (ML) approaches (**Table 2**) (Alam et al., 2014; Davis et al., 2016; Pesesky et al., 2016; Eyre et al., 2017; Li et al., 2017; Yang et al., 2017; Chen et al., 2018; Moradigaravand et al., 2018; Nguyen et al., 2018). For simplicity, we have placed them together here as under the term "model-based" prediction.

The most common strategy used is to train a classifier on a set of genomes with known phenotypes. The classifier can be asked to learn which SNPs, indels or other genetic features are important for the phenotype ab initio or be given a set of features already known to be important based on existing databases or a combination of both. As not all AMR determinants contribute equally to the antibiotic resistance of a strain, noise in phenotype prediction can often be reduced and accuracy increased by weighting each locus using a machine learning model. Models can also be trained to take into account potential interactions between loci. The accuracy of the model is determined by predicting resistance in a second set of phenotyped genomes, ideally different strains from the training set. Models can be used to predict either sensitivity or resistance based on an accepted threshold or the MIC level of the strain to the particular antibiotic. As for rule-based methods, data inputs can be reads, k-mers and assembled contigs (**Table 2**). Time-to-result is highly variable and is dependent on factors such as the

number of features used and the complexity of the ML algorithm. Three examples, discussed briefly underneath, illustrate different model-based approaches to prediction and some of the advantages and pitfalls.

Yang et al (Yang et al., 2017) examined the genomes of 1839 strains resistant to eight drugs isolated in the UK, using reference mapping to identify mutations in 23 putative resistance genes identified in earlier experimental studies. Because of the lack of lateral gene transfer in M. tuberculosis, resistance arises primarily through mutations (in this case, SNPs). Seven ML models were built and compared to simple rules-based models, where presence of a known mutation indicated resistance. Different subsets of mutations were also tested to determine the effect on performance. The best ML method for each drug increased sensitivity over the rules-based models by 2-24% with the trade-off of minor losses in specificity due to strains labelled as "susceptible" containing mutations associated with resistance. It is possible that the phenotype of these strains were determined incorrectly. The best model between different drugs varied in both the ML algorithm used and the mutation subset used, indicating that there is no one-size-fits-all solution, and the model chosen for prediction can be optimized for the complexity of resistance phenotype and amount of *a priori* knowledge.

In an alternative approach to the multiple locus problem, Eyre et al (Eyre et al., 2017) predicted *Neisseria gonorrhoeae* MICs to cefixime, penicillin, azithromycin, ciprofloxacin and tetracycline using multivariate linear regression. Multiple genetic loci (SNPs, plasmids, alleles of the *penA* gene) were already known to make contributions to resistance. The presence or absence of each candidate locus was ascertained in 681 *N. gonorrhoeae* genomes by mapping reads to a reference genome. Backwards selection (where each variable is removed until the information lost, as estimated by the Akaike Information Criterion score, is minimized) was used to reduce the number of loci in the model and

limit overfitting. Overall, model-fitted MICs were within two doubling dilutions of the MIC for 98% of strains. Using EUCAST cutoffs, the sensitivity of calling resistance was 98.7%, and the specificity was 98.3%.

The final example of predicting MICs in a complex genetic system is beta lactam resistance in *Streptococcus pneumoniae (*Li et al., 2016, 2017). Most of the strain-specific variation in levels of resistance in *S. pneumoniae* was found to be driven by amino acid sequence variation in the transpeptidase domain of three penicillin binding proteins (PBPs). Proteins with similar sequence signatures in their transpeptidase domains were clustered into "PBP Types". There were high levels of horizontal transfer of genes in the *S. pneumoniae* species, but the strain genomic background outside the PBPs only contributed 1-6% of variation in MIC (Li et al., 2016), and instead the PBP type was the most important variable. In 4,309 *Streptococcus* genome sequences where PBP Type was ascertained from *de novo* assembled data (Metcalf et al., 2016; Li et al., 2017), MIC could be accurately predicted using both a rules-based model ("Mode MIC", where the MIC was the most frequent in the closest known PBP Type) and a machine learning (random forest) classifier using the amino acid sequence of the transpeptidase domains. Mode MIC and random forest false positive rates were <3%, but critically, while the false negative rates for the random forest classifier were low, the rates for Mode MIC were above 50% for some beta lactams. The reason for the poor performance of the rule-based prediction was a subset of PBP Types with significant sequence divergence from the training set. The good performance of the random forest classifier across previously unseen sequences was because the machine learning method discovered the previously unappreciated significance for the resistance of a group of key amino acid residues in the transpeptidase domains supplied in the training set.

Consequences of incorrect AST prediction and methods to increase robustness

Errors in the sensitivity and specificity of genome prediction of antimicrobial phenotypes, either false positives (phenotypically susceptible, WGS-prediction resistant) or false negatives (phenotypically resistant, WGS-prediction susceptible) have different consequences for treatment. False negatives are considered most concerning as they can lead to inadequate treatment of a resistant infection, increasing morbidity and mortality. It is often preferable to reduce false negatives at the cost of increasing false positive rate (Gordon et al., 2014), although false positives results lead to inappropriate antibiotic use, potentially harming the patient and increasing the risk of resistance to last line antibiotics.

How can the accuracy of genetic prediction be improved in the future? It is important in the development of WGS-AST tools to have as large and diverse test and training strain sets as possible. Most studies to date have used a convenience sample of strains based on accessible collections with limited geographic and temporal variability. Even if large numbers of strain genomes are obtained, accuracy statistics can be misleading if isolates from locally abundant clonal lineages are heavily overrepresented. In reporting the isolate collection used for WGS-AST development and testing, we need to develop accepted statistics for assessment of genetic diversity. These could be pairwise average nucleotide identity, number of MLST/cgMLST sequence types, a sequence entropy measure or percentage of known species pangenome represented. Epistatic effects could explain reduced accuracy of WGS-AST across diverse strain backgrounds, especially those not included in test sets. The relatively few studies on epistasis and antibiotic resistance suggest that few generalizations can be made across species and phenotypes. Knopp and Andersson (Knopp & Andersson, 2018) found little variation in phenotypic expression for 13 resistance mutations across 10 strains of *Salmonella enterica* and *Escherichia coli* (even when as many as four mutations were combined in one background). However, there are reports of epistasis effects both between resistance mutations and between the

resistance mutation and the strain genetic background in *Pseudomonas aeruginosa* and other species (Vogwill & MacLean, 2015; Vogwill, Kojadinovic & MacLean, 2016; Wadsworth et al., 2018). Extensive empirical testing will be needed to understand how resistance phenotype expression varies across diverse strain backgrounds in each species. New statistical methods, coupled with very large pathogen genomic datasets can also discover novel epistatic interactions (Skwark et al., 2017).

The finding that rules-based methods for predicting beta lactam MIC from PBP type performed poorly on previously unseen strain types (Li et al., 2017), discussed above, is an example of how incomplete knowledge impacts WGS-AST. We don't yet understand the genetic basis of resistance for some antibiotics. For others, even where there are well-defined canonical mutations or genes, it is becoming apparent that there are a large number of rare genetic events that could lead to resistance. Guéurillot et al (Guérillot et al., 2018a) found through deep sequencing of lab selected colony pools that while eight *rpoB* mutations were responsible for 93% of rifampin resistance in *S. aureus* clinical strains (Guérillot et al., 2018b), the remaining 8% was caused by least 72 rarer mutations (Guérillot et al., 2018a). Wistrand-Yuen et al (Wistrand-Yuen et al., 2018) showed that selection for streptomycin in *Salmonella enterica* at sub-MIC levels resulted in strains with an array of novel low-effect mutations different from classic high-effect drug resistance. Potentially even harder to predict based on sequence alone are resistance phenotypes caused by rare genetic events such as insertion sequence movement into or upstream of open reading frames (McEvoy et al., 2013; Di Gregorio et al., 2016).

In the face of this genetic complexity, how can WGS-AST continue to improve? It is unrealistic to expect 100% accuracy for every drug. We may be able to establish a stable level of uncertainty for each phenotype and learn to accept that some are more genetically plastic than others. Improvements may also be possible by incorporating functional prediction algorithms for mutations in resistance

genes where the functional effect is untested. At its simplest, this could be taking into account frameshift mutations, that, for example, in a *blaZ* gene would be predicted to inactivate translation of the beta lactamase and yield a phenotypically susceptible strain. More sophisticated metrics include delta-bitscore, which uses hidden Markov models to identify functional divergence in orthologous genes and is an improvement on the previous use of dN/dS (used to indicate positive selection), which was prone to false positives (Wheeler et al., 2016). *S. aureus* resistance to trimethoprim has been accurately predicted based on the free energy state of the dihydrofolate reductase amino acid sequence (Fowler et al., 2018), suggesting an avenue for sophisticated structural prediction studies. One straightforward way to increase robustness of WGS-AST prediction is to combine predictions of multiple tools that examine different facets of the WGS dataset (i.e k-mers, assembled contigs) (Mason et al., 2018).

The phenotype problem

As mentioned earlier, WGS-AST uses culture-based antibiotic phenotypes as the gold standard but culture-based methods are not themselves free of degrees of uncertainty. Culture-based assays for drugs in some species have been shown to be unreliable: for example, ethambutol, pyrazinamide and rifampicin in *M. tuberculosis* (Zhang & Mitchison, 2003; Van Deun et al., 2009, 2013; Horne et al., 2013). Even for drug resistances that can be tested reliably, there is evidence to suggest that the error in culture-based AST between different reference laboratories is greater than the inaccuracy of WGS-AST (Aanensen et al., 2016; Eyre et al., 2017; Harris et al., 2018). Errors may arise from subtle operator-specific biases or variations between individual automated instruments that perform rapid organismal identification and susceptibility testing. Additionally, unlike standard research laboratory experimentation, it is generally not practical to perform replicate MIC tests in clinical laboratories.

Culture-based AST measures resistance in a highly constrained setting that does not reflect the variance of expression of the phenotype, especially during human infections. Many culture based MICs are not robust to changes in growth media, bacterial cell density or temperature. Antimicrobial susceptibility may be altered due to formation of biofilms and/or the synergistic or antagonistic interactions with other bacterial species (Donlan & Costerton, 2002; Beaudoin et al., 2017; Radlinski et al., 2017). Resistance gene expression can be induced to higher levels *in vivo* than *in vitro*, (Cornforth et al., 2018) small resistant subpopulations (heteroresistant strains) or mixed infections can quickly exhibit phenotypic resistance when exposed to selection, and persisters (Hassan et al., 2018), which are not detected by AST, can survive treatment and cause recurrent infections. This underlies the fundamental issue of using MIC breakpoints to define resistance. Breakpoints are not based on genetic correlates of resistance but factors such as distribution of MICs, pharmacokinetics of the drug, and clinical treatment experience among others. Consequently, there is considerable debate on how to set breakpoints, which can vary across countries and organizations (Kassim et al., 2016). Breakpoints are often revised due to new epidemiological data, but adoption can sometimes be slow, and changes can serve to confound retrospective meta-analyses of antibiotic resistance trends and mechanisms defined in the literature.

Despite culture-based AST being an imperfect gold standard for WGS-AST, it is unlikely that it will be soon replaced. However, due to a revolution in NAAT based typing, clinical laboratories are culturing many fewer isolates, especially hard-to-grow species such as *Legionella spp* and *M. tuberculosis*; or sexually transmitted diseases such as *Neisseria*, which need a fast turnaround time for diagnosis. As sequencing becomes cheaper and more accurate, metagenomic-AST (**Figure 1**) may render culturing of isolates unnecessary for routine treatment. However, as bacterial pathogens continue to evolve, it is possible that new resistance loci may emerge or the relationship between known AMR determinants

and phenotypic resistance change due to epistatic interactions with novel emerging strain backgrounds. Therefore, even when phenotypic testing is no longer used as a primary screen, it should continue to be used as a quality control step to ensure that contemporary clinical strains have not drifted away from predicted phenotypes. Large multi-site surveys that integrate rational strain sampling and centralised gold standards AST such as the recent European *N. gonorrhoea* study (Harris et al., 2018) will be needed to provide the continually update and challenge existing WGS-AST methods.

Concluding Remarks

The adoption of routine use of whole genome sequencing in clinical microbiology would be a revolution in clinical medicine. While this review only discusses bacterial pathogens, there is also progress being made to use WGS to predict drug resistance in viruses, fungi, and eukaryotic parasites (Codoñer et al., 2011; Pou et al., 2014; Runtuwene et al., 2018). The major barriers to adoption for routine use stem from technological limitations that make the process currently too slow and expensive and the need to demonstrate the advantages of a new approach to diagnosis for skeptical end-users in clinical microbiology laboratories (**Box 3**). Along with the development of instrumentation, there is need for efficient data management of WGS data, user-friendly pipelines that do not require specialized bioinformaticians for use, and rapidly interpretable results (Crisan et al., 2018). As has been covered in more depth in other articles (Goldberg et al., 2015; Gargis, Kalman & Lubin, 2016; Ellington et al., 2017), standardization of software, sequence data, genomic test sets and phenotypes and data storage are all needed to transform what is now a research enterprise into a robust clinical tool acceptable to both clinical microbiologists and regulatory agencies.

Acknowledgements

We would like to thank Claudia Solis Lemus for reading and commenting on the draft manuscript. T.D.R was supported by National Institute of Allergy and Infectious Diseases (NIAID) award AI121860. M.S. was supported in part by the Antimicrobial Resistance and Therapeutic Discovery Training Program funded by NIAID T32 award AI106699-05.

Disclaimer

The contents of this article are solely the responsibility of the authors and do not necessarily reflect the official views of the National Institutes of Health or the United States government.

Figures and tables

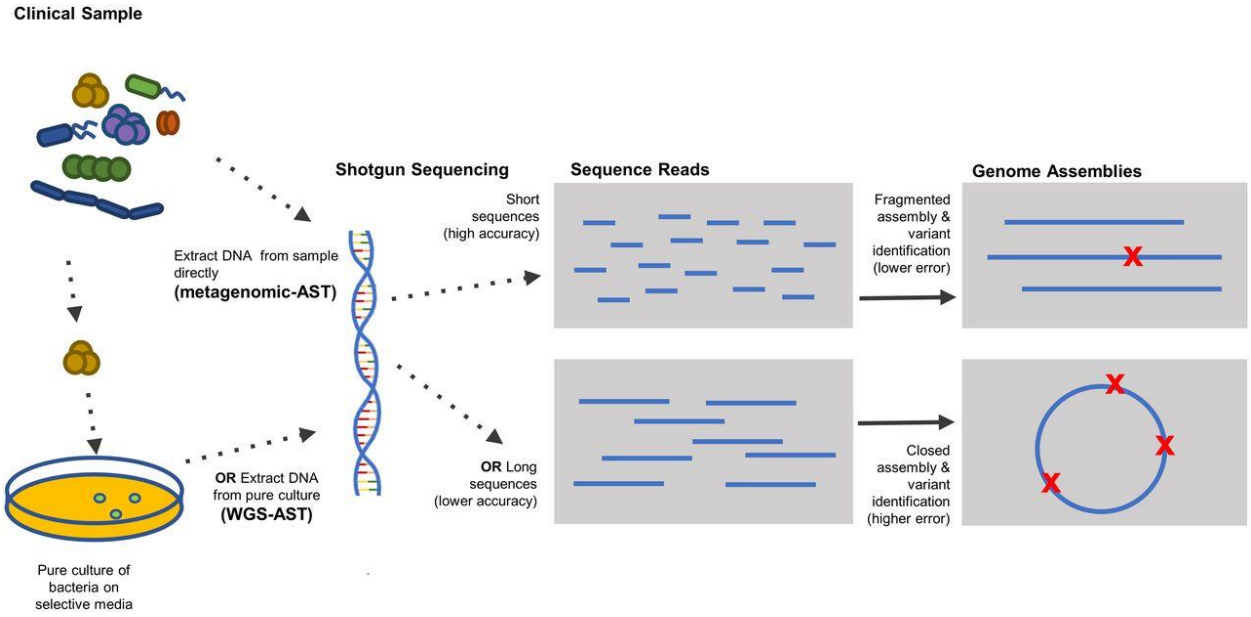

Figure 1. Overview of genome sequencing and how it is used in WGS-AST.

DNA is extracted from directly from bacteria in clinical samples (metagenomics) or more commonly, from cultured bacterial colonies. Sequencing technologies fragment the DNA and then randomly sequence to produce a library of reads (stored in FASTQ files). The reads are assembled into genomic scaffolds *in silico*. Sequencing is performed either using short-read second-generation technology, which tends to produce fragmented whole-genome assemblies of high accuracy, or long-read third-generation technologies that have higher error rates but more complete assemblies. WGS-AST algorithms operate on the raw reads and/or assembled contigs.

Term	Definition
Quality score	A measure of the probability of an inaccurate base call, typically represented by the Phred score [$Q = -10 \log_{10}(P)$] (113). A Q of 10 represents a 1 in 10 chance of error, whereas a Q of 40 is a 1 in 10,000 chance of error.
Coverage	A measure of how many instances a base was sequenced, as quantified by number of unique reads mapped to that position in the genome. A "genome of coverage of 30×" means that on average, each base of the genome has 30 reads mapped.
Sequence read	Inferred nucleotide sequence of a genome fragment. Reads range from short (≤300 bp) to long (5 to 100+ kbp).
Contig	A contiguous sequence created by assembling multiple overlapping sequence reads.
Genome assembly	A singular (complete) or set of contigs after aligning and merging all sequence reads. Assemblies can be created *de novo* (relies only on sequence reads) or by mapping to a reference strain.
First-generation sequencing	Nucleotide sequencing that relied on either chain termination (Sanger) or cleavage (Maxam-Gilbert) methodology in single-tube reactions.
Second-generation sequencing	Nucleotide sequencing methods that sheared the genome and PCR amplified individual DNA fragments to massively parallelize sequencing and detect base identity by monitoring release of pyrophosphate (454), release of hydrogen (Ion Torrent), release of fluorescent reversible-terminator nucleotides (Illumina), or fluorescent ligated probes (SOLiD).
Third-generation sequencing	Nucleotide sequencing that relies on real-time single-molecule sequencing via monitoring of fluorescently labeled nucleotide incorporation (Pacific Biosciences) or ion current after DNA is fed through a channel (Oxford Nanopore).

Table 1. Genomic terms.

Locus type	Description[a]
Gene	Presence of an intact protein-coding gene that confers resistance. For example, a strain that contains *blaZ* is inferred to be resistant to beta-lactams.
Plasmid/mobile element	Presence of a known drug resistance plasmid or mobile genetic cassette (e.g., SCC*mec* [114]) is used to infer that *S. aureus* is resistant to a beta-lactams.
Mutation	A particular SNP or SNV (which encompasses both SNPs and 1-bp indels) that is associated with resistance.
Allele	Nucleotide variant of gene caused by mutation. One sequence variant of a gene may be sensitive to a drug, while another allele may be associated with resistance.
Gene amplification	Increase in gene copy number due to homologous recombination. For example, a single gene in a genome may be sensitive to a drug, but a strain with two or more tandem repeats may be resistant.

Table 2. Types of antibiotic resistance loci.

a SCC*mec*, staphylococcal cassette chromosome *mec* element; SNV, single-nucleotide variant.

Species	Antibiotic(s)	No. of genomes tested	Diversity/no. of STs	Primary database(s)	Software	Input data	Sensitivity (%)	Specificity (%)	Reference
E. coli	Amoxicillin-clavulanate	76	NR	Custom	Blastx, ClustalW	Assembly, FASTQ	100	100	Tyson et al. (71)
	Trimethoprim	48	19+ STs	ResFinder	ResFinder	Assembly or FASTQ	100	100	Zankari et al. (115)
	Gentamicin	74	NR	>100 loci	Blastn	Assembly	100	100	Stoesser et al. (70)
M. tuberculosis	Pyrazinamide	167	NR	PhyResSE	Stampy	FASTQ	88.9	100, 94.9 (with uncharacterized)	Pankhurst et al. (64)
	Isoniazid	693	Lineages 1–4	TBDReaMDB, MUBII-TB-DB minus phylogenetic SNPs	TB Profiler	FASTQ	92.8	100	Coll et al. (63)
	Moxifloxacin	13,424	NR	NR	NR	NR	88.2	90	Miotto et al. (74)
	Amikacin	667	7 clades	Custom	SAMtools, mpileup, Cortex	FASTQ	91.2, 88.1 (with uncharacterized)	99.4, 99.5 (with uncharacterized)	Walker et al. (116)
	Rifampin	1,565	NR	Hain, Cepheid, AID[b] assays, literature	Mykrobe	FASTQ	90.8	99	Bradley et al. (5)
	Ethambutol	752	NR	Mykrobe	Mykrobe	FASTQ	100	98.5, 77.3 (with uncharacterized)	Quan et al. (72)
S. aureus	Fusidic acid	491	61 STs	Custom	Blastn, tblastn	Assembly	91	100	Gordon et al. (66)
	Vancomycin	NR	16 CCs	Custom	Blastn, mapping software	Assembly, FASTQ	100	100	Aanensen et al. (67)
	Mupirocin	340	25 CCs	Modified from Gordon et al. (66)	Mykrobe	FASTQ	100	100	Bradley et al. (5)
	Ciprofloxacin, clindamycin, erythromycin, fusidic acid, gentamicin, methicillin, mupirocin, penicillin, rifampin, tetracycline, trimethoprim, vancomycin	1,379	111 STs	Custom	Mykrobe, GeneFinder, Typewriter	FASTQ, Assembly	97	99	Mason et al. (42)
Salmonella enterica serovar Typhi	Chloramphenicol	332	ST1 and ST2	CARD, ResFinder, literature	GeneFinder	FASTQ	100	100	Day et al. (61)
Non-serovar Typhi *S. enterica*	Ceftriaxone	640	NR	Tyson et al. (71)	Blastx, ClustalW	Assembly, FASTQ	100	99.8	McDermott et al. (59)
	Ciprofloxacin	3,491	227 serovars	CARD, ResFinder	GeneFinder	FASTQ	99.28	99.97	Neuert et al. (60)
S. pneumoniae	Erythromycin	210	90 STs	SRST2	SRST2	FASTQ	100	100	Deng et al. (69)
Campylobacter jejuni,	Erythromycin	32/82	NR	Tyson et al. (71), ARDB, ResFinder	Blastx, ClustalW	Assembly, FASTQ	100	100	Zhao et al. (62)

Species	Antibiotic(s)	No. of genomes tested	Diversity/no. of STs	Primary database(s)	Software	Input data	Sensitivity (%)	Specificity (%)	Reference
Campylobacter coli									
Enterococcus faecalis, Enterococcus faecium	Kanamycin	50	12 STs 17 STs	ResFinder	ResFinder	Assembly or FASTQ	100	100	Zankari et al. (115)
Pseudomonas aeruginosa	Levofloxacin	390	175 STs	Custom	NR	Assembly	91.9	93.7	Kos et al. (68)
Klebsiella pneumoniae	Gentamicin	69	NR	Custom	Blastn	Assembly	96	100	Stoesser et al. (70)
Shigella sonnei	Ampicillin	341	NR	CARD, ResFinder	GeneFinder	FASTQ	100	100	Sadouki et al. (65)

Table 3. Selected rule-based WGS-AST results.[a]

a NR, not reported; ST, sequence type; CC, clonal complex.

b AID, Autoimmun Diagnostika GmbH.

Species	Antibiotic(s)	No. of genomes tested	Diversity[a]	Database[b]	ML algorithm	Input data	Sensitivity (%)	Specificity (%)	Overall accuracy (%)[c]	Reference
E. coli	Amoxicillin	329	7 STs	NA	Gradient-boosted trees	Pangenome, population structure matrix	90	95		Moradigaravand et al. (78)
	Ciprofloxacin	581	7 STs	NA	Gradient-boosted trees	Pangenome, population structure matrix, SNPs	81	99		Moradigaravand et al. (78)
	Gentamicin	564	7 STs	NA	Gradient-boosted trees	Pangenome, population structure matrix	87	99		Moradigaravand et al. (78)
	Trimethoprim	283	7 STs	NA	Gradient-boosted trees	Pangenome, population structure matrix	92	97		Moradigaravand et al. (78)
M. tuberculosis	Isoniazid	1,811 (80% train, 20% test)	7 clades	NA	Random Forest	Variants in 23 genes	97	94		Yang et al. (79)
	Rifampin	1,725 (80% train, 20% test)	7 clades	NA	Class-conditional Bernoulli mixture model	Variants in 23 genes	97	97		Yang et al. (79)
	Ethambutol	3,526 (80% train, 20% test)	5 genetic clusters	NA	Multitask wide and deep neural networks	Variants in 32 regions	91.9	90.3		Chen et al. (81)
	Pyrazinamide	3,147 (train), 567 (test)	5 genetic clusters	NA	Multitask wide and deep neural networks	Variants in 32 regions	75.2	90.1		Chen et al. (81)
	Kanamycin	162 (train), 18 (test)	NR	PATRIC, RAST	AdaBoost	Assembly			88.3 (F1)	Davis et al. (83)
S. pneumoniae	Beta-lactams (PEN, AMO, MER, TAX, CFT, CFX)[d]	2,528 (train), 1,781 (test)	403 STs (train), 299 STs (test)	NA	Random Forest	PBP sequences			>97 (±1 MIC dilution), >93 (category)	Li et al. (80)
	Beta-lactams	1,350 (train), 58 (test)	NR	PATRIC, RAST	AdaBoost	Assembly			87.6 (F1)	Davis et al. (83)
N. gonorrhoeae	Azithromycin	681	NR	NA	Linear regression	Variants in 20 regions	80, 99 (±1 MIC dilution)	83, 94 (±1 MIC dilution)	93 (±1 MIC dilution), 44 (category)	Eyre et al. (45)
	Ciprofloxacin	676	NR	NA	Linear regression	Variants in 20 regions	100	99	94 (±1 MIC dilution), 68 (category)	Eyre et al. (45)
K. pneumoniae	Ampicillin-sulbactam	1,668	>99 STs	PATRIC, RAST	XGBoost	Assembly			99 (F1, ±1 MIC dilution)	Nguyen et al. (117)

Species	Antibiotic(s)	No. of genomes tested	Diversity[a]	Database[b]	ML algorithm	Input data	Sensitivity (%)	Specificity (%)	Overall accuracy (%)[c]	Reference
	Levofloxacin	1,668	>99 STs	PATRIC, RAST	XGBoost	Assembly			93 (F1)	Nguyen et al. (117)
	Meropenem	1,777	>99 STs	PATRIC, RAST	AdaBoost	Assembly			92 (F1)	Long et al. (118)
	Piperacillin-tazobactam	1,777	>99 STs	PATRIC, RAST	AdaBoost	Assembly			76 (F1)	Long et al. (118)
E. coli, *Enterobacter aerogenes*, K. pneumoniae	Ampicillin	78	NR	Resfams	Logistic regression	FASTA, alignments			97.4	Pesesky et al. (85)
	Chloramphenicol	78	NR	Resfams	Logistic regression	FASTA, alignments			89.7	Pesesky et al. (85)
S. aureus	Methicillin	99 (train), 11 (test)	NR	PATRIC, RAST	AdaBoost	Assembly			99.5 (F1)	Davis et al. (83)
	Vancomycin	75	12 STs	Custom	Random Forest	Assembly	73		81	Alam et al. (82)
Acinetobacter baumannii	Carbapenem	99 (train), 11 (test)	NR	PATRIC, RAST	AdaBoost	Assembly			95 (F1)	Davis et al. (83)
Non-serovar Typhi *S. enterica*	Ceftriaxone	5,278		PATRIC, RAST	XGBoost	Assembly			80, 95 (±1 MIC dilution)	Nguyen et al. (84)

Table 4. Selected model-based WGS-AST results.

a ST, sequence type; NR, not reported.

b NA, not applicable.

c F1, harmonic average of the precision (positive predictive value [PPV]) and recall (sensitivity) (84).

d PEN, penicillin; AMO, amoxicillin; MER, meropenem; TAX, ceftriaxone; CFT, cefotaxime; CFX, cefixime.

Question
At what price and turnaround time will WGS-AST replace culture-based sequencing for routine use in clinical microbiology labs?
How do we interpret the presence of an antimicrobial resistance determinant gene if the susceptibility of the strain is below the MIC?
Can genome prediction be used to detect heteroresistance? Or to detect polygenic phenotypes?
How important is epistasis in determining the resistance to different classes of antibiotics?
Can gene amplification as a mechanism of resistance be accurately determined from WGS data?
How efficiently can WGS-AST prediction software be ported to metagenomic-AST data?

Table 5. Some outstanding questions.

Chapter III. Genomic analysis of variability in Delta-toxin levels between *Staphylococcus aureus* strains

Michelle Su[1], James T. Lyles[2], Robert A. Petit III[1], Jessica Peterson[1], Michelle Hargita[1], Huaqiao Tang[2], Claudia Solis-Lemus[3], Cassandra L. Quave[2,4], Timothy D. Read[1,3]

[1] Division of Infectious Diseases, Department of Medicine, School of Medicine, Emory University, Atlanta, GA, United States of America

[2] Center for the Study of Human Health, College of Arts and Sciences, Emory University, Atlanta, GA, United States of America

[3] Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA, United States of America

[4] Department of Dermatology, School of Medicine, Emory University, Atlanta, GA, United States of America

This work was published in PeerJ in 2020.

Citation:

Su M, Lyles JT, Petit III RA, Peterson J, Hargita M, Tang H, Solis-Lemus C, Quave CL, Read TD. 2020. Genomic analysis of variability in Delta-toxin levels between *Staphylococcus aureus* strains. PeerJ 8:e8717 https://doi.org/10.7717/peerj.8717

Abstract

Background

The delta-toxin (δ-toxin) of *Staphylococcus aureus* is the only hemolysin shown to cause mast cell degranulation and is linked to atopic dermatitis, a chronic inflammatory skin disease. We sought to characterize variation in δ-toxin production across *S. aureus* strains and identify genetic loci potentially associated with differences between strains.

Methods

A set of 124 *S. aureus* strains was genome-sequenced and δ-toxin levels in stationary phase supernatants determined by high performance liquid chromatography (HPLC). SNPs and kmers were associated with differences in toxin production using four genome-wide association study (GWAS) methods. Transposon mutations in candidate genes were tested for their δ-toxin levels. We constructed XGBoost models to predict toxin production based on genetic loci discovered to be potentially associated with the phenotype.

Results

The *S. aureus* strain set encompassed 40 sequence types (STs) in 23 clonal complexes (CCs). δ-toxin production ranged from barely detectable levels to >90,000 units, with a median of >8,000 units. CC30 had significantly lower levels of toxin production than average while CC45 and CC121 were higher. MSSA (methicillin sensitive) strains had higher δ-toxin production than MRSA (methicillin resistant) strains. Through multiple GWAS approaches, 45 genes were found to be potentially associated with toxicity. Machine learning models using loci discovered through GWAS as features were able to predict δ-toxin production (as a high/low binary phenotype) with a precision of .875 and specificity of .990 but recall of .333. We discovered that mutants in the *carA* gene, encoding the small chain of carbamoyl phosphate synthase, completely abolished toxin production and toxicity in *Caenorhabditis elegans*.

Conclusions

The amount of stationary phase production of the toxin is a strain-specific phenotype likely affected by a complex interaction of number of genes with different levels of effect. We discovered new candidate genes that potentially play a role in modulating production. We report for the first time that the product of the *carA* gene is necessary for δ-toxin production in USA300. This work lays a foundation for future work on understanding toxin regulation in *S. aureus* and prediction of phenotypes from genomic sequences.

Introduction

Staphylococcus aureus is a common causative agent of nosocomial and community-acquired infections, encoding a wide variety of factors that damage the host and evade immunity. Central to its ability to cause disease is its large repertoire of toxins. *S. aureus* can produce at least 13 extracellular toxins (Grumann, Nübel & Bröker, 2014; Otto, 2014; Laabei et al., 2015), including phenol-soluble modulins (PSMs) (Peschel & Otto, 2013), alpha-toxin (Bhakdi & Tranum-Jensen, 1991), Panton-Valentine Leukocidin (PVL) (Genestier et al., 2005), and δ-toxin (Wang et al., 2007).

Toxin expression levels are subject to evolutionary trade-offs between survival and transmission in different environments (Laabei et al., 2015; Young et al., 2017). Toxins contribute to important biological functions: In *S. aureus*, alpha-toxin is important for initial cell-to-cell contacts in biofilm formation, beta-toxin contributes to biofilm structure and growth via crosslinking, and PSMs are involved in detachment of cells for dispersal (Rudkin et al., 2017). In addition, expression of toxins is essential to skin and soft tissue infections and other common diseases caused by the bacterium (Xu & McCormick, 2012; Otto, 2013; Peschel & Otto, 2013; Kitur et al., 2015). However, during chronic *S. aureus* infections, toxin production is a contra-indication of disease as reduced toxicity mutants may

have situationally increased fitness (Cheung et al., 2014; Soong et al., 2015; Rose et al., 2015; Laabei et al., 2015). Dysfunction in the Agr quorum sensing system (Novick, 2003), central to upregulation of many toxins, has been linked to longer durations of bacteremia (Fowler Jr et al., 2004; Sakoulas et al., 2005). Similarly, mutational inactivation of another regulator, Rsp, which promotes *S. aureus* infection and virulence (Li et al., 2015), allows for prolonged survival in chronic infections (Das et al., 2016).

In this study, we focus on the genetics of strain-specific differences of δ-toxin expression. δ-toxin is an amphipathic peptide in the PSM family. It can form pores on the surface of host cells, eliciting a pro-inflammatory response or cytolysis at high concentrations (Bernheimer & Rudy, 1986; Kasimir et al., 1990; Otto, 2014). δ-toxin is the product of the *hld* gene, which is part of the Agr quorum sensing system. The Agr operon consists of two divergently transcribed operons P2 and P3. The P2 operon encodes the four genes necessary for quorum sensing and activates the P3 operon, which transcribes the main effector of the Agr system, a 514-nucleotide regulatory molecule RNAIII. RNAIII also contains the *hld* gene encoding the 26 amino acid δ-toxin peptide, which has been found only in *S. aureus* and *S. epidermidis* (McKevitt et al., 1990). In a community-associated MRSA (CA-MRSA) bacteremia mouse model, PSM *a* and δ-toxin were shown to be important for disease severity, indicating their importance as virulence factors (Wang et al., 2007; Peschel & Otto, 2013). However, δ-toxin is the only PSM shown to induce mast cell degranulation (Nakamura et al., 2013) and increase the severity of *S. aureus* mediated Atopic Dermatitis (AD), a chronic inflammatory skin disease, affecting 15–30% of children and 5% of adults in the US and industrialized countries (Williams & Flohr, 2006; Pustišek, VurnekŽivković & Šitum, 2016). Despite its importance, we know little of the natural variation in production of the δ-toxin molecule between *S. aureus* strains and the genetic factors that influence this trait. Therefore, we queried the range of δ-toxin production in a diverse set of *S. aureus* strains and attempted to determine if there are genetic loci strongly associated with δ-toxin

production by using bacterial genome-wide association study (GWAS) methods. We then analyzed the performance of identified genome variants and metadata for predicting δ-toxin production.

Materials & Methods

Strains and growth conditions

Network on Antimicrobial Resistance in *Staphylococcus aureus* (NARSA) and Nebraska Transposon Mutant Library (NTML) strains were acquired from BEI resources (https://www.beiresources.org/) (Table S1). For δ-toxin assays, bacteria were grown on tryptic soy agar (TSA) plates overnight (18–24 hours) at 37 °C. TSA plates used for NTML strains had the addition of erythromycin (5 μg/ml). Cultures from a single colony were inoculated and grown overnight in tryptic soy broth (TSB) at 37 °C, a 45° angle, and 200 rpm. Final cultures were standardized to a starting cell density of 5×10^5 CFU/ml of TSB and grown for 15 hours at 37 °C, a 45° angle, and 275 rpm (Quave & Horswill, 2018).

Whole-genome shotgun sequencing

DNA extraction and paired-end library prep were performed as manufacturer's instructions (Wizard Genomic DNA Purification Kit, Promega; Nextera XT DNA Library Prep Kit, Illumina). Genome sequencing was performed using both Illumina HiSeq and MiSeq. Raw read data were deposited in the NCBI Short Read Archive under project accession number PRJNA289526. 102/124 strains had more than 40x average genome coverage, and the minimum coverage of any strain was 33x.

Genome assembly and annotation

Genomes were processed using the Staphopia pipeline (Petit & Read, 2018). BBduk (v37.66) (Bushnell, 2016) was used to eliminate Illumina adapters, trim low quality ends (base quality <20), and filter out low quality reads (mean read PHRED quality <20). Read error correction and *de novo* genome

assembly was performed using SPAdes (v3.11.11) (Bankevich et al., 2012). Genome assemblies were annotated with Prokka (Seemann, 2014) (v1.12) using its default database. SNP-sites was used to call single nucleotide polymorphisms (SNPs) in the core genome alignment with *S. aureus* N315 as the reference strain (Page et al., 2016). Agr type was determined using BLAST to query genome assemblies for the *agrD* nucleotide sequences of defined agr types: I (AB492152.1), II (AF001782), III (AF001783), and IV (AF288215). For all but four samples, 100% coverage and >95% identity were used to identify Agr type. NRS168, NRS182 NRS235, and NRS260 Agr types were determined based on available metadata. Untyped strains (NA) returned no BLAST results likely due to contig boundaries falling in this region. MLST (multilocus sequencing type) was determined using the SRST2 tool (Inouye et al., 2014) with the PubMLST database (Jolley, Bray & Maiden, 2018).

Phylogenetic tree estimation

A core genome alignment of 999,473 base-pairs (bp) from the 124 NARSA strains was obtained from Roary (Page et al., 2015; Tange, 2011) (v3.11.2). Gubbins was used to remove potential recombination regions and to obtain a downsized core genome alignment of 42,406 bp containing only polymorphic sites (Croucher et al., 2015). A final maximum likelihood (ML) tree was obtained with RAxML (v8.2.10) with 100 bootstraps and a GTRGAMMA model (Stamatakis, 2014).

Toxin identification using HPLC

High performance liquid chromatography (HPLC) methods were employed to detect and quantify the levels of δ-toxin present in supernatants of 124 NARSA strains following established procedures (Quave & Horswill, 2018). Briefly, a 1.5 ml *S. aureus* culture grown for 15 hours as described above was centrifuged and the supernatant transferred to an HPLC vial and frozen at −20 °C until ready for HPLC testing. HPLC was performed with the following parameters: 500 µl injection, flow rate of 2

mL/min, and UV/Vis monitored at 214 nm using solvents (a) 0.1% (vol/vol) trifluoracetic acid in water and (b) 0.1% trifluoracetic acid in acetonitrile. Peaks at retention time ~7.2 min and ~7.5 min corresponding to deformylated and formylated δ-toxin respectively, were quantified by taking the sum of the total peak area. Peak areas were normalized using OD_{600} readings of the cultures. Prior studies using this HPLC method confirmed peak identity at these retention by LC-MS (Somerville et al., 2003; Quave, Plano & Bennett, 2011). Analyses were performed on three replicate supernatants per strain. For subsequent analyses treating toxin production as a continuous variable, we used a Box-Cox power transformation to achieve a more symmetric distribution and thus satisfy the normality assumption of phylogenetic regression and other comparative models. For analyses treating toxin production as a binary Low/High, we used a cutoff of 20,000, which clusters strains on the left-side of the distribution and split the data into 109 low and 15 high toxin producers.

Hemolysis assay

TSA with 5% rabbit's blood was used to test transposon strains for reduced hemolysis as rabbit blood is more susceptible to δ-toxin. TSA II with 5% sheep's blood agar was used to test the hemolysis profile of complemented strains. Strains were spotted and incubated at 37 °C for 24 hours before incubation at 4 °C for an additional 24 hours. Photos of the plates were taken with the use of a lightbox to illuminate hemolysis zones. Images were imported into software ImageJ (Schneider, Rasband & Eliceiri, 2012) to increase the contrast of the image and for measurement of hemolysis zones by taking the hemolysis measurement and subtracting colony size.

GWAS

All GWAS analyses were done with 106 *S. aureus* strains. NRS168, NRS252-NRS256, NRS259, NRS260, NRS262, NRS264-NRS266, NRS271, NRS272, NRS275, NRS383, NRS386, NRS387, and NRS408 were later phenotyped and included in all other analyses.

SEER (Lees et al., 2016) (v 1.1.4) was run according to https://github.com/johnlees/seer/wiki/Tutorial. In brief, kmers used for SEER were counted using fsm-lite (https://github.com/nvalimak/fsm-lite) using genome assemblies in fasta format as input. Population structure was estimated using Mash (Ondov et al., 2016) to sketch assemblies and output pairwise distances between all samples. SEER scripts were used to create a distance matrix. Six dimensions was chosen based on scree plot output (Fig. S1). SEER was run using a binary phenotype with *p* value filtering off. QQ plots were made in R to ensure that population structure was properly accounted for. A minor allele frequency (MAF) of .20 was chosen as regression analysis with kmers of lower MAF tend to fail or have high standard errors. Significant kmers were kmers with likelihood ratio test *p*-values lower and equal to the Bonferroni correction of .05/n, where n is the number of kmers tested. Significant kmers were mapped to reference genome N315 (NC_002745.2) using BLAST (Camacho et al., 2009) optimized for short queries. Bedtools (v2.27.1) (Quinlan & Hall, 2010) was then used to annotate the matches.

For treeWAS (Collins & Didelot, 2017), a binary core SNPs matrix was generated from snp-sites output. To account for the right skewed distribution of δ-toxin production, the values were transformed into ranks. treeWAS was run with 3 unrooted trees (NARSA strains alone, NARSA strains plus ST93, NARSA strains plus *S. argenteus*) generated from RAxML (v8.2.10) to limit false positives generated from an incorrect phylogeny, and the intersection of all loci identified was considered significant.

For bugwas (Earle et al., 2016), we used a modified version of GEMMA 0.93 (Zhou & Stephens, 2012) with a centered relatedness matrix (GEMMA option -gk 1) created using BIMBAM files and a binary toxin phenotype file and set Minor Allele Frequency of 0. Biallelic core SNPs were used to create a mean genotype file, and SNP positions were noted in a SNP annotation file with the chromosome number set to 24 to indicate one allele. A nucleotide matrix of core SNPs, a binary phenotype file, and an unrooted phylogenetic tree created by RAxML (v8.2.10) were used to run the bugwas R package. A Bonferroni correction of .05/n was used, where n was the sum of phylogenetic patterns represented by the bi and tri allelic SNPs.

DBGWAS (Jaillard et al., 2018) (v0.5.0) was run using genome assemblies in fasta format, a binary phenotype, unrooted phylogenetic tree created by RAxML (v8.2.10), and DBGWAS resistance and UniProt databases for annotation. A false discovery rate (FDR) of 5% was used to determine significant kmers.

Phylogenetic regression

We fit three phylogenetic comparative models: (1) a phylogenetic regression to study the association between δ-toxin production and several covariates, (2) a Pagel's lambda model to estimate the phylogenetic signal of δ-toxin production, and (3) an ancestral state reconstruction of δ-toxin production along the branches of the *Staphylococcus aureus* phylogeny.

For the phylogenetic regression model, we included δ-toxin level as a continuous response for 106 strains and 11 predictors: clonal complex (CC), methicillin-resistant *S. aureus* (MRSA), Agr type (agr), and variants associated with δ-toxin production in *S. aureus* identified using SEER (Lees et al., 2016)

and bugwas (Earle et al., 2016) and DBGWAS (Jaillard et al., 2018) (Table S3). *isdC* and WP_000894032.1 were excluded from the analysis due to large standard errors while sequence type (ST) was excluded due to similarity to CC. Bugwas variants were represented in the analysis as 6 phylogenetic patterns. We used the julia package PhyloNetworks (Solís-Lemus, Bastide & Ané, 2017; Bastide et al., 2018) to fit the phylogenetic regression, to estimate the phylogenetic signal in δ-toxin through Pagel's λ transformation, and to reconstruct the ancestral states. To tease apart which if any of these factors truly impact δ-toxin production, we performed a phylogenetic regression. First, with Pagel's lambda model, we estimated a strong phylogenetic signal ($\lambda = .504035$) for the δ-toxin production using a rooted (by *S. argenteus*) phylogenetic tree calibrated to be consistent with time. This estimate did not entirely fit under the Brownian Model (BM) assumption, which requires $\lambda \approx 1$. Regardless of this, we assumed a BM for the evolution of δ-toxin production in the phylogenetic regression model. The rationale for the use of BM was its simplicity as well as the shown robustness to model misspecification (Bastide et al., 2018).

Extreme Gradient Boosted Tree Classifier

The R package xgboost (Chen & Guestrin, 2016) was used to create predictive classifiers with strain metadata and genetic features from the GWAS. The predictor was trained using stratified 10-fold cross-validation wherein 90% is used for training and 10% is used for validation. Model performance metrics such Area Under Receiver Operating Characteric (AUROC) and Cohen's Kappa were calculated using R packages pROC and irr respectively.

Other statistical analysis

Association of MRSA/MSSA status, Agr type, and CC to toxin production was performed with Kruskall-Wallis and pairwise Mann-Whitney U tests using continuous δ-toxin as the dependent

variable. A Bonferroni correction was applied to test *p*-values to account for multiple tests. Pagel's lambda and Blomberg's K (Pagel, 1999; Blomberg, Garland Jr & Ives, 2003) were estimated using R package phytools using an unrooted phylogenetic tree obtained from RAxML and no calibration to the branch lengths. All analyses were performed using the R (R Core Team, 2016) and Julia programming language (Bezanson et al., 2017) for statistical computing.

Phenotypic analysis of toxin phenotypes of transposon mutant strains

Nine transposon mutants in genes potentially associated with δ-toxin production were selected from the USA300 Nebraska transposon library (Fey et al., 2013). In addition, we selected an *agrA* mutant as a positive control for δ-toxin disruption and one randomly chosen mutant with no known association from the GWAS experiments as a negative control. The gene disrupted was a phi77 ORF109-like protein, SAUSA300_1928, WP_000582165.1. All transposon mutants were transduced into an isogenic USA300 JE2 background and confirmed by PCR (Table S2). HPLC assays and hemolysis assays for δ-toxin were as previously described. Complementation was performed by cloning PCR fragments containing the USA300 genes into the pOS1-Plgt vector using splicing overlap extension PCR (Bubeck Wardenburg, Patel & Schneewind, 2007). In brief, the plasmid and genes were PCR amplified to contain complementary overhangs. The purified products were then mixed and subject to another round of PCR with no primers. This reaction was used to transform IM08B *E. coli* (Monk et al., 2015). The plasmid was purified and electroporated into the mutant strains as previously described (Monk et al., 2012). A pOS1-Plgt only plasmid was used as a control for complementation experiments. The *C. elegans* virulence assays were performed using *C. elegans* strain N2. Nematode population synchronization was performed as in Penley & Morran (2018). Populations were bleached in 20% household bleach and M9 buffer and plated on OP50 until L4 larval stage (48 hours at 20 °C). Worms were subsequently washed off, counted, and ~200 were plated on control OP50 plates

and *S. aureus* lawns on BHI agar. *S. aureus* plates were prepared 24 hours prior by adding 200 ul of an overnight culture and growing at 24 °C. At 24 hours and 48 hours, plates were scored by counting live worms. Worm counts on OP50 plates were used to normalize mortality calculations and to account for plating efficiency.

Results

δ-toxin production level is highly variable between *S. aureus* strains and is associated with MSSA/MRSA and Clonal Complex

We used high performance liquid chromatography (HPLC) to quantify stationary phase δ-toxin production in 124 publically available *S. aureus* strains from the Network on Antimicrobial Resistance in *Staphylococcus aureus* (NARSA) collection, which represents diverse taxonomic groups within the species (Table S1). The strains, which were shotgun sequenced using Illumina technology, were a diverse representation of the *S. aureus* species, consisting of 40 sequence types (STs) in 23 clonal complexes (CCs). There was considerable variation in the total δ-toxin production (sum of the formylated and deformylated δ-toxin peptides) between strains (Fig. 1, Fig. S2). The distribution most closely fits a gamma model with a strong left skew. Production ranged from zero to 97,235 units, with a median value of 8,295. The majority of strains produced less than 10,000 units; 118 of the 124 strains (95%) produced less than 30,000.

When toxin production was mapped onto the strain phylogeny, it was apparent that there was variation in the average level between clonal complexes even though there was also a large variation in the phenotype within CCs (Fig. 2). Two tests for phylogenetic signal of the trait, Blomberg's K and Pagel's lambda (Pagel, 1999; Blomberg, Garland Jr & Ives, 2003), returned statistically robust scores (K = .019, p = .016/ λ = .99, p = 1.55e−48; for both measures, a value of 1 indicates trait similarity measured by

variance (K) or correlation (λ) as expected under Brownian evolution). A Pagel's lambda value of ~1 indicates a strong phylogenetic signal, while the low Blomberg's K indicates that the variance that exists within δ-toxin production is primarily on the tips of the trees within clades and does not wholly fit a Brownian model of evolution for δ-toxin. When variation between clades was analyzed pair-wise using Kruskal-Wallis and Mann-Whitney tests, CC30 (Average δ-toxin 4299) was found to have significantly lower δ-toxin production than CC45 (Average δ-toxin 30955, $p = .027$) and CC121 (Average δ-toxin 17693, $p = .00042$, Fig. 3C). Ancestral reconstruction of the δ-toxin phenotype (Fig. S3) suggested that high δ-toxin producing clades such as CC45, CC890 and CC72 had arisen independently in the *S. aureus* species from a low producing ancestor.

Agr groups have been suggested to be associated with differences in *S. aureus* cytotoxicity (Jarraud et al., 2002; Collins, Buckling & Massey, 2008). All four Agr groups were present in our samples (I: 50, II: 22, III: 38, IV: 11, NA: 3). We found significant differences in toxin production between Agr I and III ($p = .022$) using Kruskal-Wallis test and between Agr III and IV ($p = .00049$) using pairwise Mann-Whitney U tests. Agr I and IV have higher mean levels than Agr II and III (Fig. 3B). Methicillin resistance has also been previously indicated to interfere with the Agr quorum sensing system and thus toxin production (Rudkin et al., 2012). Within our set of strains, MRSA strains were found to have lower δ-toxin than MSSA strains by Mann-Whitney U ($p = 0.024$, Fig. 3A). Some caution must be used when assigning causality as clonal complex, Agr group and MRSA status are strongly confounded, but we also found that MRSA status and Agr type III were significant negative predictors of δ-toxin level in a phylogenetic regression (see Methods).

Diverse genetic loci are associated with variation in individual strain δ-toxin production levels

To ascertain individual genomic variants that may potentially be associated with the differences in toxin production, we used four recently published programs for bacterial GWAS (genome-wide association studies). The programs differ in the population structure correction, the types variants tested and whether continuous or binary phenotypes could be tested. SEER (Lees et al., 2016) is an alignment-free method that uses kmers as features to create a distance matrix and a fixed effects model to correct for population structure. SEER therefore allows discovery of both core and accessory gene variants associated with the phenotype. For the purposes of GWAS, we defined the binary toxin phenotype at a cutoff of 20,000 units (Fig. 1), which gave a set of 87 strains in the "low" toxin and 19 in the high toxin category. This threshold was chosen to separate the very high-producing strains from the main mass (Fig. 1). Using the binary phenotype, SEER identified three genes having more than ten kmers with statistically significant association (Table S3). Polymorphisms in *isdC*, *glpD*, and a gene encoding YbbR-like domain-containing protein (WP000894032.1) were found to be negatively associated with δ-toxin production. Bugwas (Earle et al., 2016) is an alternative distance-based GWAS program that uses principal components as random effects for population structure control. Bugwas analysis, performed with a binary phenotype, produced six phylogenetic patterns of SNPs (Table S3). DBGWAS (Jaillard et al., 2018) is a kmer based alignment-free method which relies on De Brujin graphs to interpret genomic variation but uses bugwas for downstream analyses. Running DBGWAS with a binary phenotype yielded one hit in an intergenic region near staphopain A (Table S3). treeWAS (Collins & Didelot, 2017) differed from the other three programs in using the strain phylogeny to model changes associated with phenotype changes at the tip and within the structure of the tree. Using a phylogenetic tree can be more accurate than distance estimation if the tree is correct, therefore for robustness we used three separate trees (NARSA strains only, NARSA strains with a ST93 outgroup, NARSA strains with an *S. argenteus* outgroup) and pooled common loci. Using a ranked phenotype,

genes common to all three analyses were *fadD, vraD, degA, gdpP, ggt, sufB, opc*R, *rebM, thiD,* and three uncharacterized proteins (Table S3).

When the results from the four approaches were aggregated (Table S3), we noted the majority of the genetic loci were in genes encoding enzymes that were part of conserved metabolic pathways. Forty-two variants were synonymous mutations, and only nine were non-synonymous (6 loci from the bugwas analysis and 3 from treeWAS). None of the four GWAS approaches found any significant genetic loci in common.

We attempted to determine if machine learning approaches could predict δ-toxin from genome sequences by integrating information from the diverse GWAS analyses with MRSA/MSSA, Agr group and genotype (CC or ST). We chose Extreme Gradient Boosting (XGBoost) (Friedman, 2001; Chen & Guestrin, 2016) which uses decision tree ensembles to predict from the given set of features. XGBoost has been used to successfully predict biologically relevant phenotypes such as antibiotic resistance in *Enterobactericiae* and *Salmonella* (Nguyen et al., 2018; Nguyen et al., 2019) as well as RNA-protein interactions (Jain, Gupte & Aduri, 2018), protein-protein interactions (Wang, Liu & Deng, 2018; Sanchez-Garcia et al., 2019), and RNA methylation (Qiang et al., 2018). An XGBoost model was trained with stratified 10-fold cross-validation. When using a binary δ-toxin phenotype (>20000), the model had a precision of .875 and a recall of .333. Specificity was .990, and the Area Under Receiver Operating Characteristic (AUROC) was .697 (Table 1). Interestingly, excluding MRSA/MSSA status, Agr group and ST/CC had no effect on model performance, suggesting that rare genomic variants are the main driving force of very high δ-toxin production. Recall was poor, suggesting there are yet unfound genomic determinants that contribute strongly to the phenotype. Splitting the toxin levels into 4 categories (0–1000, 1001–7000, 7001–30000, >30000) decreased

performance (With metadata: .423 weighted precision, .443 weighted recall, .664 AUROC; Without metadata: .451 weighted precision, .326 weighted recall, .667 AUROC). However, most of the errors in the 4-category model (~75%) occurred in adjacent categories, suggesting that the classifier was better than random choice with near-misses. ST was the top parameter in prediction, resulting in a loss of ~21% accuracy when omitted. This suggested that the driving force behind differential δ-toxin level in *S. aureus* is interactions between a potentially large number of genes with the potential to affect toxin expression levels.

The small chain of carbamoyl phosphate synthase, encoded by *carA*, is necessary for δ-toxin production in USA300

We screened the δ-toxin production phenotype of transposon mutants of 9 of the 42 genes putatively identified by GWAS (Table S3) as well as *agrA* as a positive control and a randomly chosen gene as negative control (SAUSA300_1928). We used mutants from the USA300 Nebraska Transposon Mutant Library (Fey et al., 2013) that were transduced back into the parental USA300 JE2 strains and validated by PCR (Table S2). Of the eleven mutants tested, transposons in *hemL*, *carA*, *glpD*, *isdC*, *thiD*, and *agrA* significantly reduced δ-toxin production (Fig. 4), but only *carA, agrA*, and *isdC* mutants showed significantly different hemolysis on rabbit blood agar by Mann Whitney U. δ-toxin production in strains containing transposons in *fadD*, *sbnC*, *brnQ*, *hlgB,* and phi77 ORF109-like protein was not different to the parental strain USA300 JE2. *agrA* is necessary to activate RNAIII and *hld* transcription (Janzon, Löfdahl & Arvidson, 1989; Gagnaire et al., 2012), so the transposon knock-out was expected to completely abrogate expression. However, the complete shutdown of toxin expression in the *carA* mutant had not been previously reported. We showed that δ-toxin accumulation by the *carA* mutant could be rescued by a cloned version of the gene on an expression plasmid (71% δ-toxin production restoration compared to USA300 JE2) but not an empty vector (0% δ-toxin production restoration).

carA encodes the carbamoyl-phosphate synthase small chain protein, which is involved in L-arginine biosynthesis and UMP biosynthesis (part of pyrimidine metabolism) and has been shown to potentially regulate nitric oxide resistance (Grosser et al., 2018) and be important for the regulation of PSMα1 expression (Hardy et al., 2019). We confirmed the results of Bae et al. (2004) that a *carA* mutant was defective in killing *Caenorhabditis elegans* and showed this phenotype could be restored by complementation. While hemolysis of the complementation strain on sheep blood agar was less than the original strain (Table 2, Figs. 5A–5E), virulence in *C. elegans* was similar. Additionally, complementation with a *carA* allele from a CC45 high δ-toxin strain that restored hemolysis to comparable levels but not δ-toxin production showed reduced killing of *C. elegans* (Fig. 5F). This result suggests that δ-toxin may play an important role in virulence in this model organism.

Discussion

This study revealed the complex relationship between strain phylogeny and δ-toxin accumulation at stationary phase in *S. aureus*. The phenotype had a strikingly left-skewed distribution, with a minority of strains having >5-fold the median value in toxin units (Fig. 1). Pathoadaptation probably plays a major role in generating this diversity: δ-toxin has been shown to be an important virulence factor in skin and soft tissue infection (Berlon et al., 2015), but in bacteremia *agr*-regulated toxins may be under negative selection (Fowler Jr et al., 2004; Sakoulas et al., 2005). Toxin expression levels may also change through neutral genetic variation or selection on other metabolic pathways, especially if the levels of the toxin are ultimately determined by the interaction of multiple complex regulatory pathways (Priest et al., 2012). We showed there was a strong relationship between phylogeny and δ-toxin expression (Fig. 3C). Ancestral reconstruction of δ-toxin levels (Fig. S3) suggested higher expression has evolved several times independently but in a minority of clades, indicative of the fitness trade-offs that can exist with increased virulence. Strains also vary considerably within CCs, suggesting

within-clade mutations affect the level of expression. An example of this is the NRS22 strain in CC45, which had a more 4-fold less production than the average for the CC (NRS22 = 6,686; CC45 average = 30,955) (Fig. 2). The association of higher or lower levels with particular CCs is likely due to epistatic interaction between rare mutations and variants shared between clade members. As the strains in this study originate from a wide range of infections (Table S1), it was not possible to associate δ-toxin production in *S. aureus* with a particular disease (such as atopic dermatitis).

We used multiple bacterial GWAS approaches to produce a list of candidate loci that may be affecting δ-toxin production at different phylogenetic levels. GWAS looks for homoplasic genetic variants produced by recombination or parallel evolution that can be associated with phenotypic variation (Read & Massey, 2014; Power, Parkhill & De Oliveira, 2016). Methods vary by the types of variant tested (SNPs, kmers, indels), whether continuous or discrete phenotypes are used and methods for controlling non-independence of samples due to shared ancestry and typically widespread linkage disequilibrium. Two main approaches have been implemented to determine the underlying population structure of tested bacteria. The most common is a form of principal components correction whereby the genomes of strains are used to create a matrix of linearly uncorrelated variables which can then be included in as either fixed or random effects in a regression model (e.g DBWAS, bugwas, SEER). The second approach uses a phylogenetic tree as input to determine genetic relatedness between strains and can be fed into a regression model as with principal components or used to simulate null data to determine a cut-off for true associations (e.g treeWAS). Although all the variants listed in Table S3 passed the family-wise correction cutoff by their respective programs, many are likely false positives due to the presence of linkage disequilibrium or from underestimation of the underlying population structure, especially since the majority were synonymous substitutions. SEER and treeWAS appeared to be finding SNPs that were common in the *S. aureus* population (in ~1,500–20,000 of the ~44,000

strains in the Staphopia database (Petit & Read, 2018)). In contrast, bugwas found SNPs strongly associated with CC45 (the most toxic CC), while DBGWAS found a variant associated with low toxin production that excludes all high toxin strains in CC45. Of note, *mecA* kmers were not discovered by GWAS methods, although phylogenetic regression pointed to MSSA strains having higher toxicity. Similarly, we didn't find any variants within *agr* genes associated with differences in the phenotype, possibly because they were too rare in our population. Although, it is possible to use GWAS to find novel SNPs with large effect sizes using a relatively small number of genomes within one CC (Laabei et al., 2014), the conclusion to be drawn from this pilot study is that larger numbers of *S. aureus* genomes will be needed to understand the δ-toxin phenotype across multiple CCs.

Two other GWAS studies have focused on toxicity in *S. aureus*. Laabei et al were able to build a random forest predictor using 31 SNPs and 21 indels to predict low, medium, and high toxicity in MRSA with an accuracy of >85% (Laabei et al., 2014). Recker et al. (2017) clarified the role of toxicity by determining factors associated with bacteremia-associated mortality. Five genes, including two in the *agr* operon, were selected by random forest to be predictive of mortality in CC22 and CC30 *S. aureus* bacteremia. None of the genes putatively associated with variation in toxin production from these two studies overlap ours, although our work differed in having a focus specifically on δ-toxin. Genes from our list that have been implicated in other work are *hemL1,* which is part of the *agrA* transcriptional pathway (Das et al., 2016; Young et al., 2017) and *brnQ*. Other genes found to have reduced rabbit blood hemolysis that did not overlap our set were *hemB, qoxA-C* and *hlgA*. Our GWAS results suggested variants in *qoxD, hemC* and *hlgB* affected in delta toxin production, but these have not been shown to have any toxin-related phenotype in any study we have seen. Mutations in *clpX* and *walK*, which were found in our GWAS results, have also been shown to affect hemolysis (Frees et al., 2003; Delauné et al., 2012; Jacquet et al., 2019).

We validated nine candidate genes for the effect that transposon mutations had on δ-toxin production. Some candidate genes essential for cellular survival (e.g., *clpX* and *walK*) cannot be tested using knockout mutants. Transposons in 5 genes (two predicted by bugwas, two by SEER and one using treeWAS) had no effect on production, indicating they were likely false positive calls. The finding that *hemL, glpD, isdC*, and *thiD* knockouts resulted in a small but significant reduction δ-toxin levels suggests that they may have a significant functional role. The USA300 *carA* knockout had the most dramatic phenotype as the gene was found to be indispensable for δ-toxin production, a result not previously reported. A non-δ-toxin producing *carA* mutant was found to have reduced virulence in *C. elegans,* suggesting a role for δ-toxin in infection. Further mechanistic studies are now needed to understand why *carA* is necessary. Strikingly, the variant discovered through the GWAS screen, and all mutations in the gene in high production CCs such as CC45 were synonymous. There is a growing body of literature documenting differences in protein function caused by synonymous mutations that impact RNA toxicity (Mittal et al., 2018) or protein folding (Walsh, Bowman & Clark, 2019). Given the involvement of the Agr system, there is possibly a role for mutations to change RNAIII binding specificity and influence gene regulation.

This work (and other GWAS studies) suggest that strain-to-strain variation in δ-toxin production is governed by complex genetic interactions. The high number of significant but probably low effect variants discovered in this analysis highlights the complex regulation of the δ-toxin phenotype and may parallel models proposed for the genetic basis of some traits in eukaryotes (Boyle, Li & Pritchard, 2017). Nevertheless, we showed that we can train a classifier (XGBoost) using only genome features with prediction accuracies of 87.9% (binary categories) and 43.5% (4 categories). We found that the most important predictive feature was ST in the non-binary model, which reflects how much of the

variation in δ-toxin production between strains is dependent on phylogeny. The ability to predict phenotypes of toxicity based on sequence data is likely to become an important diagnostic tool as medicine increasingly adopts genome-based technologies (Laabei et al., 2014). Going forward, we can improve genome-based predictors and gain mechanistic insights that may lead to development of anti-toxin drugs through a combination of efforts to expand collection of phenotypic variation in natural strains and molecular genetic studies targeted at high-effect loci.

Conclusions

δ-toxin production in *S. aureus* is a strain-specific phenotype likely affected by a complex network of genes. GWAS and machine learning approaches have proved successful in determining genetic determinants underlying this phenotype and using them for genome-based prediction. While most genes discovered by GWAS modify δ-toxin production, *carA* was found to be essential. Differences in *carA* function may contribute to virulence by modulating δ-toxin production. Further studies are needed to understand toxin regulation in *S. aureus* and to predict phenotypes from genomic sequences.

Additional Information and Declarations

Competing Interests

Timothy D. Read is an Academic Editor for PeerJ.

Author Contributions

Michelle Su conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft. James T. Lyles conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

Robert A. Petit III analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

Jessica Peterson, Michelle Hargita and Huaqiao Tang performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.

Claudia Solis-Lemus analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Cassandra L. Quave and Timothy D. Read conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The sequenced S aureus strains described here are available at NCBI Short Read Archive: PRJNA289526.

Data Availability

The following information was supplied regarding data availability:

The raw data is available at Figshare: Su, Michelle (2019): GWAS of Delta toxin in *S. aureus* PeerJ #41844. Figshare. Dataset. https://figshare.com/projects/GWAS_of_Delta_toxin_in_S_aureus/69566.

Network on Antimicrobial Resistance in *Staphylococcus aureus* (NARSA) and Nebraska Transposon Mutant Library (NTML) strains were acquired from BEI resources.

Funding

Timothy D. Read and Michelle Su were supported by the National Institute of Allergy and Infectious Diseases (NIAID) award AI121860. Michelle Su was also supported by the Antimicrobial Resistance and Therapeutic Discovery Training Program funded by NIAID T32 award AI106699-05. There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We would like to thank Levi Morran and McKenna Penley for help with *C. elegans* experiments and John Lees and Ruth Massey for discussion about running GWAS analysis. The discussed strains were provided by the Network on Antimicrobial Resistance in *Staphylococcus aureus* (NARSA) for distribution by BEI Resources, NIAID, NIH: *Staphylococcus aureus*.

Figures and tables

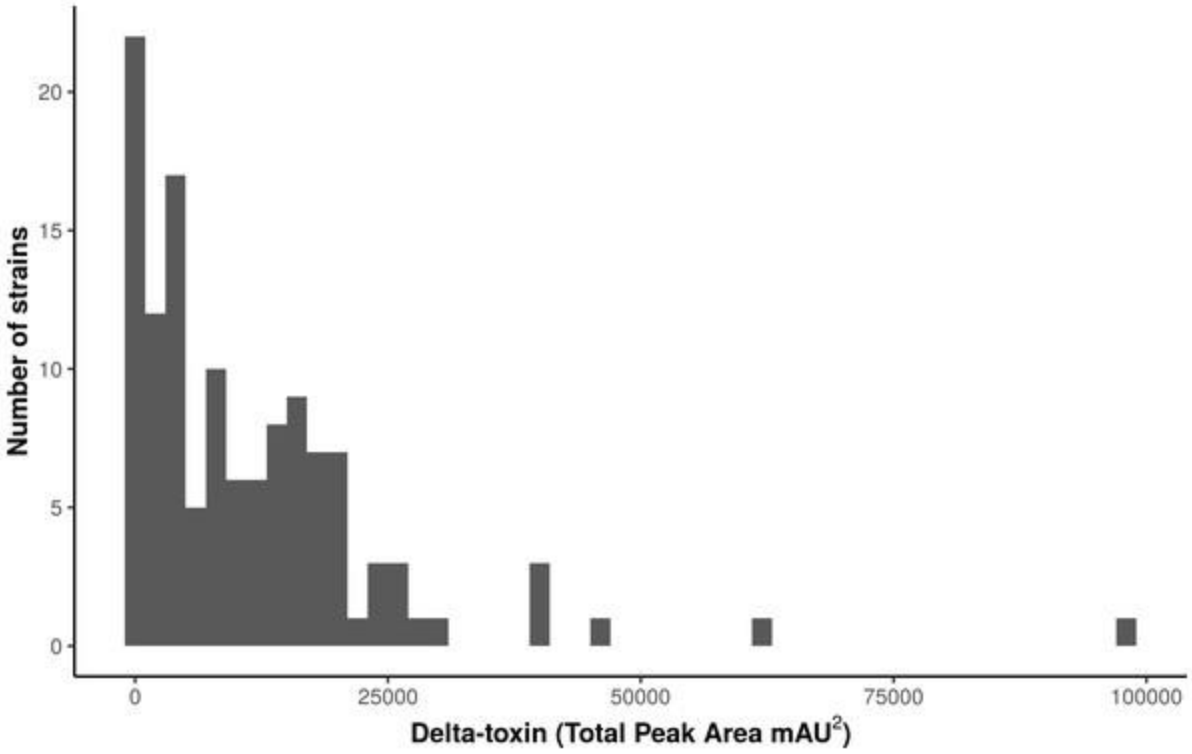

Figure 1: Characterization of *δ*-toxin production in 124 *S. aureus* strains.

Supernatants from 124 *S. aureus strains* from the Network on Antimicrobial Resistance in Staphylococcus aureus (NARSA) repository were subjected to high performance liquid chromatography (HPLC). The amount of *δ*-toxin present in samples is proportional to the area under the peak generated by UV absorbance when passing through the HPLC detector. Production ranged from 0 to 97,234 mAU^2.

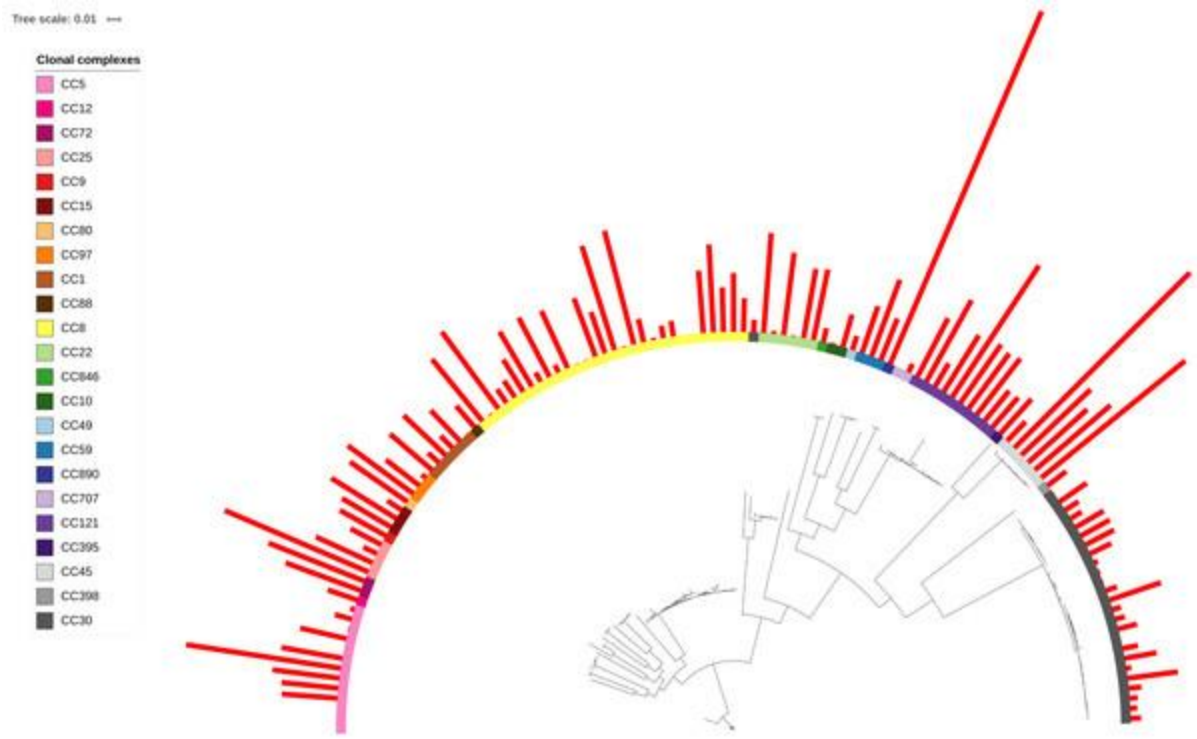

Figure 2: *δ*-toxin production across the *S. aureus* phylogeny.

A total of 124 *S. aureus* strains in 23 clonal complexes (CC) were used to create a core genome phylogeny using RAxML. Clonal complexes are represented in the inner ring and color coded. Red bars represent *δ*-toxin values from HPLC. Missing bars indicate *δ*-toxin was undetectable by HPLC.

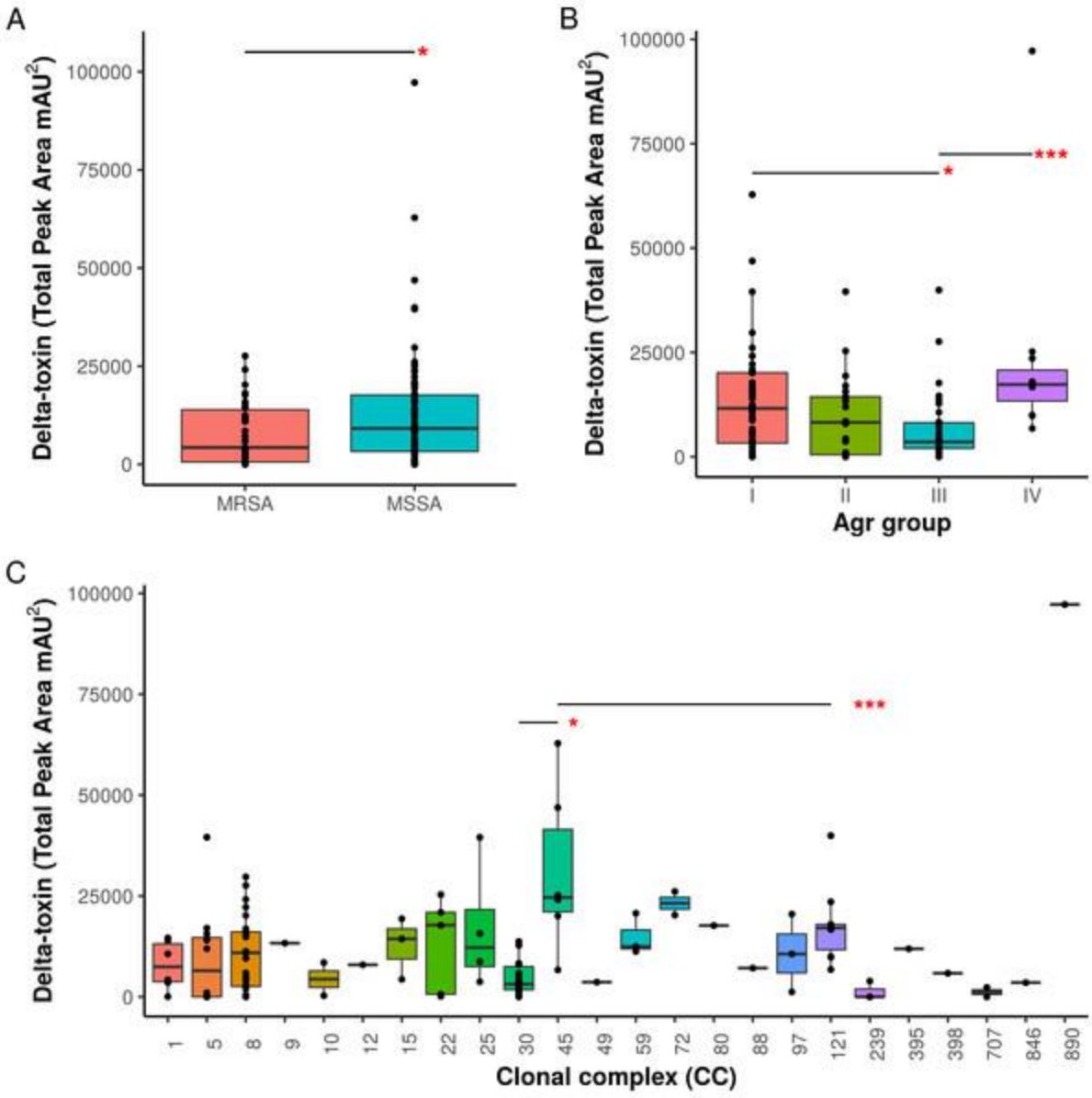

Figure 3: Associations of *δ*-toxin production to methicillin resistance, Agr type and clonal complex.

(A) Differences in *δ*-toxin production between MSSA and MRSA strains. (B) Differences in *δ*-toxin production between agr types. (C) Differences in *δ*-toxin production between CCs. *$p < .05$, **$p < .01$, ***<.001.

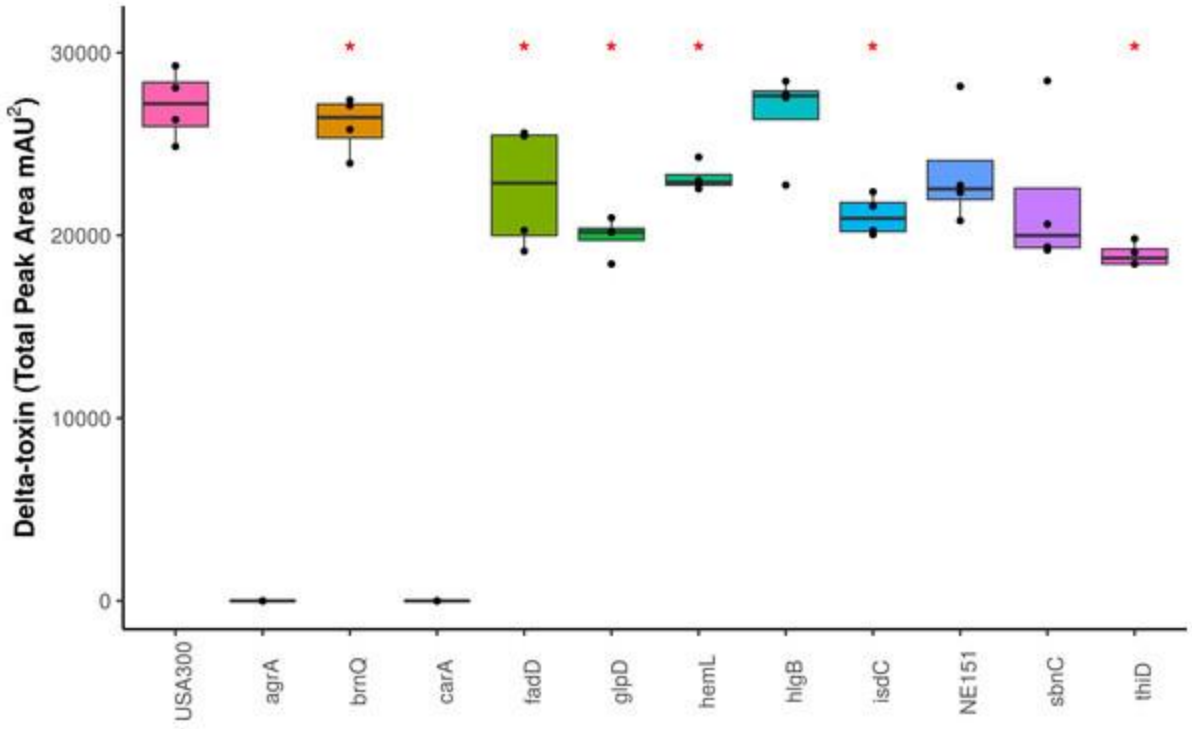

Figure 4: Impact of gene knockouts on *δ*-toxin production.

A subset of genes that were found to be significantly associated with *δ*-toxin by GWAS were tested for their effect of *δ*-toxin production. *δ*-toxin from transposon mutants from the Nebraska Transposon Mutant Library (NTML) was measured via HPLC. *$p < .05$.

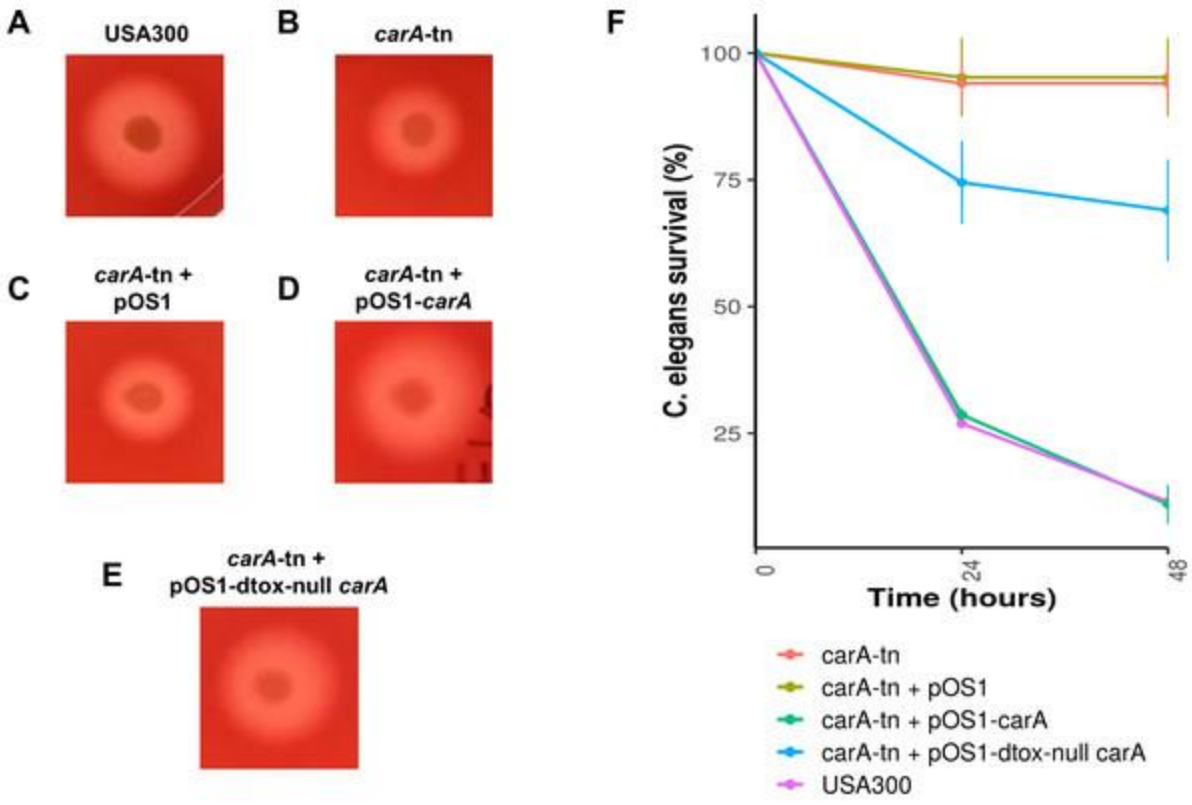

Figure 5: Impact of *carA* on hemolysis and virulence of *S. aureus* in *C. elegans*.

(A–E) *S. aureus* strains were spotted on 5% sheep's blood TSA II and incubated at 37 °C for 24 hours followed by 24 hours at 4 °C. Representative colonies are being shown. (F) L4 *C. elegans* were fed either USA300 *S. aureus* (purple), a *carA* transposon mutant USA300 (red), a *carA* transposon mutant with vector (yellow), a *carA* transposon mutant complemented with the native USA300 *carA* allele (green), or a *carA* transposon mutant complemented with a *carA* allele that produces no *δ*-toxin (blue). Survival was scored at 24 and 48 hours. Survival counts are normalized against C. elegans fed on OP50 *E. coli*.

Model	Precision	Recall	Specificity	AUROC(95% CI)	Cohen's Kappa(*p*-value)
Binary predictor +/− metadata	.875	.333	.990	.697 (.553, .840)	.429 (.000000037)
Four category predictor + metadata	.423 (weighted)	.443 (weighted)	N/A	.664 (.597, .731)	.255 (.00441) (weighted)
Four category predictor − metadata	.451 (weighted)	.326 (weighted)	N/A	.667 (.576, .758)	.133 (.131) (weighted)

Table 1: XGBoost models were trained with and without metadata (ST, CC, Agr group, and MSSA/MRSA status).

Ten-fold cross-validation was used to assess model performance. For all measures, the average performance across cross-validation is reported. For non-binary classification, individual precision and recall measures are weighted according to its proportion in the overall dataset during averaging, and Cohen's Kappa is calculated using squared weights. Specificity is not measured for non-binary classification models as it is included in the weighted precision and recall measures.

Strain	Average Hemolysis (mm)	Standard deviation (mm)
USA300	9.0945	.508
carA-tn	5.6450	.360
carA-tn + pOS1	5.0459	.446
carA-tn + pOS1-carA	7.3305	.536
carA-tn + pOS1-dtox null carA	7.0114	.422

Table 2: Characterization of hemolysis activity of USA300, *carA* mutant, and complemented strains.

S. aureus strains were spotted on 5% sheeps blood TSA II and incubated at 37 °C for 24 hours followed by 24 hours at 4 °C. Plates were photographed using a lightbox and processed in ImageJ. Hemolysis rings and colonies (n = 10) for each strain were measured using ImageJ.

Supplemental Information

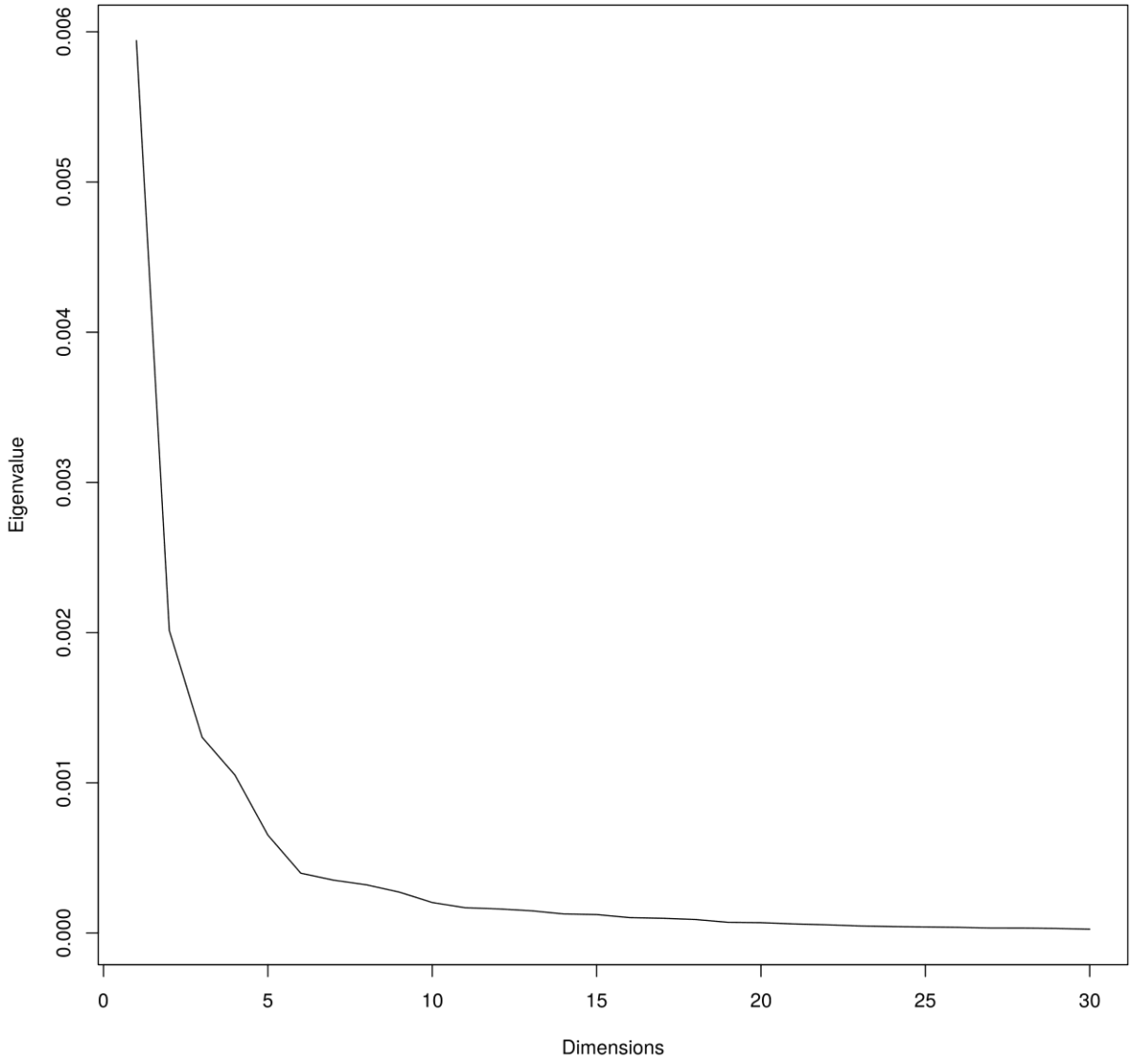

Figure S1. SEER Scree plot

For SEER, a distance matrix was estimated from kmers from *S. aureus* genome assemblies. To determine the number of dimensions to project the distance matrix into to estimate population

structure, eigenvalues were plotted against dimensions. Six dimensions was chosen based on the second "elbow" of the graph.

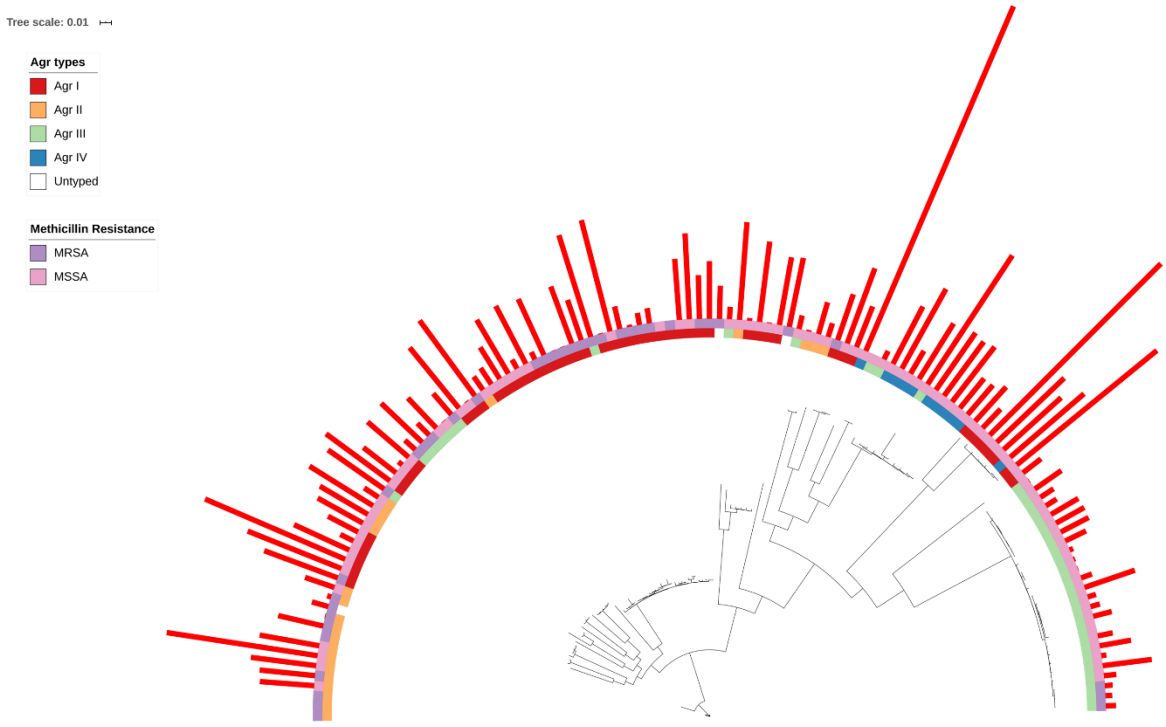

Figure S2. *δ*-toxin production in different Agr types and between MRSA/MSSA *S. aureus* strains

124 *S. aureus* strains in 23 clonal complexes (CC) were used to create a core genome phylogeny using RAxML. Agr type is represented in the inner ring and color coded. MRSA/MSSA status is represented in the outer ring and color coded. Red bars represent *δ*-toxin values from HPLC. Missing bars indicate *δ*-toxin was undetectable by HPLC.

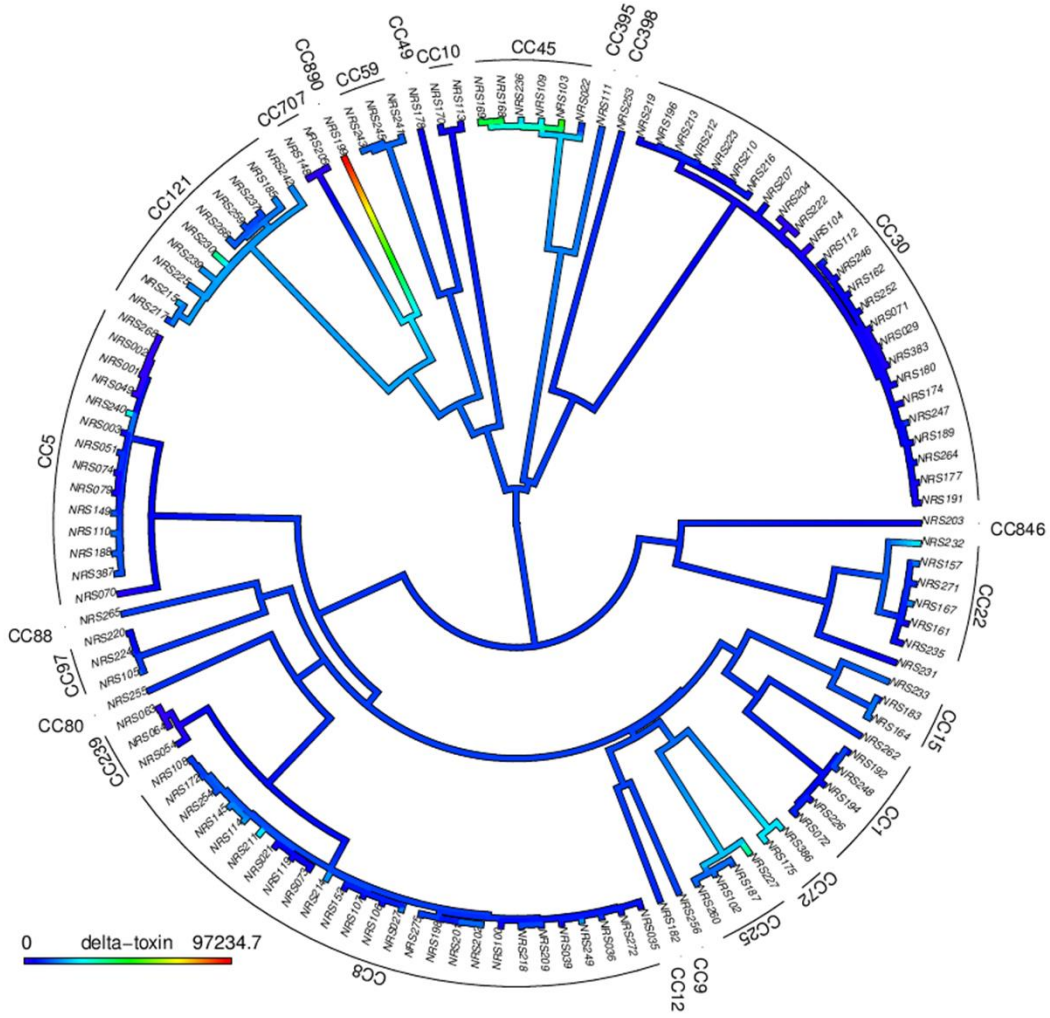

Figure S3. Ancestral state reconstruction of *δ*-toxin

The julia package PhyloNetworks was used to reconstruct ancestral states of *δ*-toxin based on 124 *S. aureus* strains.

NID	SRA Experiment Accession	Sequence Model	Total Reads	Total Basepair	Estimated Coverage (x)	Total Contigs	N50	ST	CC	MRSA	Agr	Strain	Delta toxin (Total Peak Area mAU²)	Year	Sex	Age	Locale	Country
NRS001	SRX11144452;SRX7142943	Illumina MiSeq; HiSeq X Ten	37,179,004	5,166,695,149	1835.54	66	103013	5	5	TRUE	2	Mu50	0	1996	M	4 mo		Japan
NRS002	SRX11160071	Illumina MiSeq	627,596	136,755,297	48.58	786	6493	5	5	TRUE	2	Mu3	0	1996				Japan
NRS003	SRX11160072	Illumina MiSeq	497,886	105,365,042	37.43	872	5245	5	5	TRUE	2	HIP5827	0	1997	M	59	Michigan	USA
NRS021	SRX11160073	Illumina MiSeq	924,558	211,054,750	74.98	264	54067	8	8	TRUE	1	HIP07920	473.1516775	1999	M	43	Rhode Island	USA
NRS022	SRX11160074	Illumina MiSeq	822,092	176,832,644	62.82	391	23471	45	45	FALSE	1	HIP07930	6686.805947	1999	F		New York	USA
NRS027	SRX11160075	Illumina MiSeq	987,608	207,190,825	73.61	287	24415	8	8	TRUE	1	HIP09433	24153.65577	2000	F	77	Michigan	USA
NRS029	SRX11160076	Illumina MiSeq	774,288	153,217,266	54.43	608	10052	36	30	TRUE	3	HIP09735	1703.411573	2000	M	64	North Carolina	USA
NRS035	SRX11160077	Illumina MiSeq	692,992	144,288,908	51.26	689	7757	572	8	TRUE	1	LIM1	41.99747447	1995	F	2		France
NRS036	SRX11160078;SRX7142944	Illumina MiSeq; HiSeq X Ten	23,250,748	3,614,552,538	1284.11	270	69327	247	8	TRUE	1	LIM2	10925.00867	1995	F	2		France
NRS039	SRX11160079	Illumina MiSeq	7,513,948	747,767,438	265.65	173	59824	247	8	TRUE	1	99.3795.V	142.5134834	1999			Scotland	United Kingdom

NRS049	SRX11160080	Illumina MiSeq	989,194	189,877,003	67.46	285	20231	5	5	TRUE	2	NRS49	0	1997	M	45		South Korea
NRS051	SRX11160081	Illumina MiSeq	889,862	187,694,313	66.68	72	179023	5	5	TRUE	2	HIP09740	11951.94796	2000	F	27	California	USA
NRS054	SRX11160083	Illumina MiSeq	842,692	169,238,745	60.12	535	10068	239	239	TRUE	1	BR15	3907.393244	1998	M	8		Brazil
NRS063	SRX11160084	Illumina MiSeq	694,706	145,635,109	51.74	604	10789	372	239	FALSE	1	NRS63	0					
NRS064	SRX11160085	Illumina MiSeq	762,262	145,134,315	51.56	805	6452	372	239	TRUE	1	LY-199906 20-02	0	1998	F	50		Oman
NRS070	SRX11160086	Illumina MiSeq	919,642	166,198,304	59.04	1184	5221	5	5	TRUE	2	N315	1006.177869	1982				Japan
NRS071	SRX11160087	Illumina MiSeq	667,582	139,550,716	49.58	358	15994	36	30	TRUE	3	MRSA252	1980.884961	1997	F	64		
NRS072	SRX11160088	Illumina MiSeq	1,129,846	235,383,117	83.62	147	53225	1	1	FALSE	3	MSSA476	10650.96425	1998	M	9		
NRS073	SRX11160089;SRX7142946	Illumina MiSeq; HiSeq X Ten	73,638,570	11,247,46,685	3995.80	312	62456	8	8	TRUE	1	HIP10540	5931.409935	2000	M		Ohio	USA
NRS074	SRX11160090	Illumina MiSeq	9,735,952	993,717,678	353.03	122	103526	105	5	TRUE	2	HIP10267	105.8070596	2000	M	30	Maryland	USA
NRS079	SRX11160091	Illumina MiSeq	497,768	96,579,659	34.31	941	4813	231	5	TRUE	NA	IL (Isolate F)	4336.958368	1999	F	63	Illinois	USA
NRS100	SRX11160092	Illumina MiSeq	1,120,942	253,338,402	90.00	263	86825	250	8	TRUE	1	COL	2036.396645	1961			London	United Kingdom
NRS102	SRX11160093	Illumina MiSeq	1,087,952	204,896,800	72.79	570	11777	25	25	FALSE	1	Reynolds	15695.69516	1979			California	USA

NRS103	SRX11160094	Illumina MiSeq	892,592	165,293,366	58.72	663	7654	508	45	FALSE	1	Becker	62816.6 1608	1979				USA
NRS104	SRX11160095	Illumina MiSeq	1,054,232	213,203,500	75.74	343	16111	30	30	FALSE	3	Cowan I, NC TC8530	67.8931 3473	1935			New York	USA
NRS105	SRX11160096	Illumina MiSeq	712,016	155,634,499	55.29	336	47796	97	97	FALSE	1	Wood 46	20522.3 1777					
NRS106	SRX11160097	Illumina MiSeq	1,203,294	249,731,415	88.72	183	62450	8	8	FALSE	1	RN4220/pG01	49700.07 7122					
NRS107	SRX11160098	Illumina MiSeq	1,109,600	198,161,964	70.40	676	14547	8	8	FALSE	2	RN4220/pG0400	4138.51 2711					
NRS108	SRX11160099	Illumina MiSeq	634,378	134,984,987	47.96	521	10411	8	8	TRUE	NA	A96064 9	8502.70 2183					France
NRS109	SRX11161 00	Illumina MiSeq	6,947,400	680,507,504	241.76	23	206809	47	45	FALSE	1	FRI361	24092.0 451	1962			England	United Kingdom
NRS110	SRX11161 01	Illumina MiSeq	1,268,998	169,230,493	60.12	160	38960	5	5	FALSE	2	FRI472	17026.4 8334					USA
NRS111	SRX11161 02	Illumina MiSeq	2,842,724	297,558,250	105.71	535	21977	395	395	FALSE	1	FRI913	11908.5 5423					USA
NRS112	SRX11161 03	Illumina MiSeq	837,092	190,320,425	67.61	179	53641	30	30	FALSE	3	MN8	12758.9 7106	1980			Minnesota	USA
NRS113	SRX11161 04	Illumina MiSeq	855,038	193,062,408	68.59	72	118858	10	10	FALSE	2	MNDON	305.524 1879					
NRS114	SRX11161 05	Illumina MiSeq	509,804	111,111,985	39.47	304	21324	8	8	FALSE	1	MNHOCH	22232.4 3453					USA
NRS119	SRX11161 06;SRX7142947	Illumina MiSeq; HiSeq X Ten	75,165,988	11,187,102,000	397 4.37	144	100273	507	8	TRUE	1	SALinR#12	3091.42 0143	2001	M	85	Massachusetts	USA

NRS1455	SRX11161607	Illumina MiSeq	492,894	114,311,924	40.61	247	29251	8	8	FALSE	1	RN4282	15778.2152				Minnesota	USA
NRS1488	SRX11161608	Illumina MiSeq	429,006	101,357,082	36.01	162	44271	121	707	FALSE	3	RN6432, Smith Diffuse	2326.340585					
NRS1499	SRX11161609	Illumina MiSeq	790,792	169,649,431	60.27	241	28385	5	5	FALSE	2	RN6607, 502A	15645.40123					
NRS1522	SRX11161610	Illumina MiSeq	830,508	181,427,498	64.45	162	46817	8	8	FALSE	1	RN7044, WGB4316	142.7192163					
NRS1577	SRX11161611	Illumina MiSeq	899,562	191,842,936	68.15	557	26118	22	22	FALSE	1	A980592	17770.65493		F	14		France
NRS1611	SRX11161612	Illumina MiSeq	797,006	179,482,237	63.76	148	42414	22	22	FALSE	1	HT20000509	151.7066926		M	26		France
NRS1622	SRX11161613	Illumina MiSeq	1,233,314	272,861,168	96.94	173	83775	30	30	FALSE	3	HT20000328	3876.021547	2000	F	11		France
NRS1644	SRX11161614	Illumina MiSeq	1,002,546	214,279,544	76.13	191	51911	15	15	FALSE	2	A890259	19366.1052					France
NRS1677	SRX11161615	Illumina MiSeq	736,400	163,139,102	57.96	212	44047	1446	22	FALSE	1	A970675	20944.78505		F			France
NRS1688	SRX11161616;SRX7142948	Illumina MiSeq; HiSeq X Ten	50,142,776	7,540,879,464	2679.00	545	11123	45	45	FALSE	4	A850375	25184.2255	1985	M			France
NRS1699	SRX11161617	Illumina MiSeq	696,626	151,548,946	53.84	227	25404	45	45	FALSE	1	A920222	46904.41482	1992				France
NRS1700	SRX11161618	Illumina MiSeq	665,990	143,590,154	51.01	170	39198	10	10	FALSE	2	A960562	8498.369052	2002	M	37		France
NRS1722	SRX11161619	Illumina	514,402	115,815,999	41.15	196	33667	8	8	TRUE	1	A970230	14848.45078	1997	M			France

		MiSeq																
NRS174	SRX11161120	Illumina MiSeq	2,025,338	412,279,960	146.47	598	9188	30	30	FALSE	3	A900507	28724.883878	1990	F			France
NRS175	SRX11161121	Illumina MiSeq	725,356	163,036,586	57.92	126	52062	72	72	FALSE	1	A910565	26117.92337	1991	F	15		France
NRS177	SRX11161122	Illumina MiSeq	1,017,772	218,545,371	77.64	88	116200	34	30	FALSE	3	A960197	2415.348873	1996	F			France
NRS178	SRX11161123	Illumina MiSeq	1,101,598	212,068,790	75.34	675	8725	49	49	FALSE	2	A910469	3654.438394	1991	M	2		France
NRS180	SRX11161124	Illumina MiSeq	463,284	97,134,440	34.51	423	14677	30	30	FALSE	3	A960254	3163.001663	1996	F	4		France
NRS182	SRX11161125	Illumina MiSeq	459,796	94,005,322	33.40	882	4897	12	12	FALSE	2	A950085	7945.397683	1995	F	34		France
NRS183	SRX11161126	Illumina MiSeq	774,384	161,418,155	57.35	229	26240	15	15	FALSE	2	A980101	4343.311022	1998	F	16		France
NRS185	SRX11161127	Illumina MiSeq	634,846	138,880,268	49.34	348	19366	121	121	FALSE	4	A890511	17317.85693	1989		8		France
NRS187	SRX11161128	Illumina MiSeq	797,444	182,540,828	64.85	198	33591	25	25	FALSE	1	A860325	8728.541158	1986	F			France
NRS188	SRX11161129	Illumina MiSeq	684,686	148,158,401	52.64	329	18639	1447	5	FALSE	2	A950206	14072.70622	1995	M	42		France
NRS189	SRX11161130	Illumina MiSeq	506,220	114,385,960	40.64	228	41699	34	30	FALSE	3	A910371	4874.710165	1991	M	16		France
NRS191	SRX11161131	Illumina MiSeq	772,346	143,503,137	50.98	854	5532	34	30	FALSE	3	A970698	2155.740408	1997	F	45		France
NRS192	SRX11161132	Illumina MiSeq	550,970	104,825,831	37.24	1207	3229	1	1	TRUE	3	C19980003 70	3565.49066	1997	F	7	Minnesota	USA
NRS194	SRX11161133	Illumina	474,384	96,259,905	34.20	444	13085	1	1	TRUE	3	C19990005 29	4238.089625	1999	M	1	North Dakota	USA

		MiSeq																
NRS1996	SRX11161134	Illumina MiSeq	689,196	130,484,864	46.36	404	13189	30	30	FALSE	3	No.49	8106.487992					United Kingdom
NRS1998	SRX11161135	Illumina MiSeq	649,896	151,151,519	53.70	96	70029	254	8	FALSE	1	No.56	2597.217219	1949				United Kingdom
NRS1999	SRX11161136	Illumina MiSeq	1,030,912	232,648,983	82.65	117	86292	890	890	FALSE	4	No.66	97234.77109	1947				United Kingdom
NRS2001	SRX11161137	Illumina MiSeq	658,434	146,815,322	52.16	255	26145	254	8	FALSE	1	No.150	16117.82279					United Kingdom
NRS2002	SRX11161138	Illumina MiSeq	544,588	122,210,894	43.42	150	84153	254	8	FALSE	1	No.152	16863.86132					
NRS2003	SRX11161139	Illumina MiSeq	962,272	193,844,903	68.87	185	36375	846	846	FALSE	3	No.153	3531.545938					United Kingdom
NRS2004	SRX11161140	Illumina MiSeq	496,202	100,496,950	35.70	529	10405	30	30	FALSE	3	No.167, NCTC6571	889.9088611	1940				United Kingdom
NRS2005	SRX11161141	Illumina MiSeq	940,904	183,622,996	65.23	99	142251	707	707	FALSE	3	No.208	0					United Kingdom
NRS2007	SRX11161142	Illumina MiSeq	799,164	165,276,117	58.72	206	38927	30	30	FALSE	3	No.229	5870.899579					United Kingdom
NRS2009	SRX11161143	Illumina MiSeq	473,916	103,264,073	36.69	625	8216	247	8	TRUE	1	No.315	17.3077404	1971	M			United Kingdom
NRS2010	SRX11161144	Illumina MiSeq	1,086,412	252,219,683	89.60	139	48124	30	30	FALSE	3	No.326, KCM187	3202.453643					United Kingdom
NRS2111	SRX11161145	Illumina MiSeq	829,392	183,360,952	65.14	222	43065	8	8	FALSE	1	No.333	29741.05787					United Kingdom
NRS2112	SRX11161146	Illumina MiSeq	556,018	125,270,639	44.50	123	65031	30	30	FALSE	3	No.344	8055.582957	1957				United Kingdom
NRS2113	SRX11161147	Illumina	1,368,268	274,537,776	97.53	137	43665	30	30	FALSE	3	No.348, 605	8289.914317	1958				United Kin

		MiSeq										E, G2						gdom
NRS214	SRX11161148	Illumina MiSeq	835,416	178,629,506	63.46	174	43862	8	8	FALSE	1	No. 359	20168.42654					United Kingdom
NRS215	SRX11161149	Illumina MiSeq	600,294	125,613,094	44.63	242	22325	121	121	FALSE	4	No. 425	23560.6258					United Kingdom
NRS216	SRX11161150	Illumina MiSeq	898,028	181,940,515	64.64	176	35868	30	30	FALSE	3	No. 426	8300.131003					United Kingdom
NRS217	SRX11161151	Illumina MiSeq	591,146	134,430,217	47.76	204	36358	121	121	FALSE	4	No. 430	6777.79962					United Kingdom
NRS218	SRX11161152	Illumina MiSeq	1,485,076	298,579,617	106.07	350	20484	247	8	TRUE	1	No. 437	15900.09437					United Kingdom
NRS219	SRX11161153	Illumina MiSeq	522,698	102,082,564	36.27	221	30058	30	30	FALSE	3	No. 536, NCTC9789, PS80	59.01670047	1974				United Kingdom
NRS220	SRX11161154	Illumina MiSeq	783,902	160,175,690	56.90	76	128476	97	97	FALSE	1	No. 611	1210.00498					United Kingdom
NRS222	SRX11161155	Illumina MiSeq	574,180	119,212,701	42.35	294	24554	30	30	FALSE	3		291.8344614	1978				United Kingdom
NRS223	SRX11161156	Illumina MiSeq	724,500	121,229,208	43.07	790	5887	30	30	FALSE	3	No. 784	3967.660063					United Kingdom
NRS224	SRX11161157	Illumina MiSeq	794,338	163,745,138	58.17	294	60643	97	97	FALSE	1	No. 690, NAG9	10629.94725					United Kingdom
NRS225	SRX11161158	Illumina MiSeq	478,316	109,207,362	38.80	449	11338	121	121	FALSE	4	No. 691	16851.8608					United Kingdom
NRS226	SRX11161159	Illumina MiSeq	436,598	101,476,507	36.05	289	27374	1	1	FALSE	3	HT200200028	55.7836392	2002	M	3 mo		France
NRS227	SRX11161160	Illumina MiSeq	597,794	140,168,394	49.80	196	42168	25	25	FALSE	1	HT200200030	39538.64819		M	4 mo		France

NRS230	SRX11161661	Illumina MiSeq	670,790	153,518,067	54.54	455	25943	121	121	FALSE	3	HT200200057	39950.5776 7	2002	M	2 yr 4 mo		France
NRS231	SRX11161662	Illumina MiSeq	604,622	138,185,825	49.09	370	38976	30	30	FALSE	3	HT200200058	3148.02758 8	2002	F	7		France
NRS232	SRX11161663	Illumina MiSeq	809,886	180,448,779	64.11	3140	2578	22	22	FALSE	2	HT200200065	25353.8975 3	2002	M	4		France
NRS233	SRX11161664	Illumina MiSeq	540,764	115,114,564	40.90	370	31627	582	15	FALSE	2	HT200200067	14365.3260 5	2002	F	1 mo		France
NRS235	SRX11161665	Illumina MiSeq	531,166	115,308,205	40.96	560	15294	22	22	FALSE	1	HT200200075	634.538349 1	2002	M	5		France
NRS236	SRX11161666	Illumina MiSeq	539,158	117,064,667	41.59	423	20842	45	45	FALSE	1	HT200201041	20045.4704 6	2002	M	1 mo		France
NRS237	SRX11161667	Illumina MiSeq	421,256	94,113,904	33.44	371	17510	121	121	FALSE	4	HT200201067	9992.64675 6		F	5 yr 5 mo		France
NRS239	SRX11161668;SRX7142949	Illumina MiSeq; HiSeq X Ten	47,404,582	7,259,650,861	257 9.08	103	76574	121	121	FALSE	4	HT200202004	17973.7092		M	3 mo		France
NRS240	SRX11161669	Illumina MiSeq	700,170	154,011,393	54.71	364	38123	5	5	FALSE	2	HT200202029	39564.6843 7	2002	M	1 yr 6 mo		Fra bce
NRS241	SRX11161670	Illumina MiSeq	550,060	124,287,539	44.15	178	41759	59	59	TRUE	1	HT200202033	12480.1503 8	2002	M	2 mo		France
NRS242	SRX11161671	Illumina MiSeq	579,286	123,848,461	44.00	787	11490	121	121	FALSE	4	HT200202038	16706.013	2002	F	1 yr 2 mo		France
NRS243	SRX11161672	Illumina MiSeq	594,574	128,056,586	45.49	256	31573	59	59	FALSE	1	HT200202052	20713.4778	2002	F	3 yr 8 mo		France
NRS245	SRX11161673	Illumina MiSeq	429,048	95,403,139	33.89	193	35189	59	59	TRUE	1	HT200203020	11264.0488 2	2002	F	8		France

NRS246	SRX11161174	Illumina MiSeq	477,014	102,671,011	36.48	496	10151	30	30	FALSE	3	HT200203307	12403.0937237	2002	M	1 mo		France
NRS247	SRX11161175	Illumina MiSeq	466,736	100,487,348	35.70	615	8305	34	30	FALSE	3	HT200203317	13707.8969	2002	F	5 mo		France
NRS248	SRX11161176	Illumina MiSeq	621,182	137,130,321	48.72	274	31054	1	1	TRUE	3	HT200203415	14588.58705	2002	M	5		France
NRS249	SRX11161177	Illumina MiSeq	556,684	127,048,439	45.14	658	20174	247	8	TRUE	3	HT200203656	27604.16016	2002	F	8		France
NRS252	SRX11161178	Illumina MiSeq	445,572	100,505,962	35.71	388	15767	30	30	FALSE	3	HT200203512	8172.815262					
NRS253	SRX14229004	Illumina MiSeq	2,300,194	317,101,488	112.65	74	118060	398	398	FALSE	1	HT200203542	5845.298322	2002	M	4 yr 2 mo		France
NRS254	SRX14229005	Illumina MiSeq	2,120,104	249,405,656	88.60	114	74643	8	8	TRUE	1	HT200203763	11266.03564	2002	F	5 mo		France
NRS255	SRX14229006	Illumina MiSeq	1,623,336	221,426,448	78.66	139	56092	80	80	TRUE	3	HT200203717	17668.37047	2002	M	3 mo		France
NRS256	SRX11161179	Illumina MiSeq	429,286	92,928,505	33.01	720	6517	109	9	FALSE	2	HT200203726	13315.71146	2002	F	2 mo		France
NRS259	SRX11161180	Illumina MiSeq	514,540	113,989,671	40.50	776	6307	121	121	FALSE	4	HT200203814	9829.176694	2002	F	6 mo		France
NRS260	SRX11161181	Illumina MiSeq	411,306	93,325,893	33.16	462	12682	25	25	FALSE	1	HT200204553	3745.491083	2002	F	5 mo		France
NRS262	SRX11161182	Illumina MiSeq	427,566	99,475,889	35.34	227	29664	188	1	FALSE	1	HT200204202	13984.0786	2002	M			France
NRS264	SRX11161183	Illumina MiSeq	2,196,388	261,897,409	93.04	393	67244	34	30	FALSE	3	HT200204380	0	2002	M	8 yr 2 mo		France
NRS265	SRX11161184	Illumina MiSeq	582,610	124,680,617	44.29	817	5979	88	88	TRUE	3	HT200204449	7130.797529	2002				Switzerland

NRS2666	SRX11161185	Illumina MiSeq	674,036	148,632,836	52.80	476	11476	121	121	FALSE	4	HT20020470	17971.95212	2002	F	1 mo		France
NRS2711	SRX11161187	Illumina MiSeq	530,738	110,386,652	39.22	1162	3352	22	22	TRUE	NA	P1V44	18169.61253	1999	F	18		Belgium
NRS2722	SRX11161188	Illumina MiSeq	591,446	122,770,514	43.62	1048	4276	247	8	TRUE	1	160013	15672.55116	2002	M	46		United Kingdom
NRS2755	SRX11161189	Illumina MiSeq	500,420	114,279,893	40.60	513	13729	254	8	FALSE	1	No. 55-2	9560.675684					United Kingdom
NRS3833	SRX11161190	Illumina MiSeq	545,130	123,997,268	44.05	164	40509	36	30	TRUE	3	USA300-0114	26000.718086				Mississippi	USA
NRS3866	SRX11161191	Illumina MiSeq	3,034,776	293,832,635	104.39	88	89196	72	72	TRUE	1	1045	20275.3957				Washington	USA
NRS3877	SRX11161192	Illumina MiSeq	7,374,422	697,613,810	247.84	62	106614	5	5	TRUE	2	HIP12864	14382.51805	2003	M	78	Oklahoma	USA

Table S1. *S. aureus* dataset metadata*

Metadata for *S. aureus* strains including sequencing and genome assembly statistics as well as ST/CC, Agr type, MRSA status, delta-toxin HPLC values and any available data on isolation.

***** Table has been modified to fit. Isolation site column has been removed

Primer	Sequence (5' -> 3')	Description
Tn_Up	CTCGATTCTATTAACAAGGG	Designed by authors of NTML. For use if transposon is in plus orientation with gene specific primer. 464 bp to transposon end.
Tn_Buster	GCTTTTTCTAAATGTTTTTAAGTAAATCAAGTAC	Designed by authors of NTML. For use if transposon is in minus orientation with gene specific primer to transposon end.
fadD_tn	ATCAGAGAAGAAACGTGC	Gene specific primer for *fadD* (SAUSA300_0227). Used to check transposon insertion in NE260.
agrA_tn	TTTTTAACGTTTCTCACCGAT	Gene specific primer for *agrA* (SAUSA300_1992). Used to check transposon insertion in NE1532.
brnQ_tn	TTAAGTTGTCGCTTGTTTCG	Gene specific primer for *brnQ* (SAUSA300_0306). Used to check transposon insertion in NE605.
carA_tn	TTAGGCATTGATATGACGC	Gene specific primer for *carA* (SAUSA300_1095). Used to check transposon insertion in NE1526.
glpD_tn	TTTTGCTCTACAAACGCAT	Gene specific primer for *glpD* (SAUSA300_1193). Used to check transposon insertion in NE233.
hemL_tn	TACGACTTAAAGCCGTATC	Gene specific primer for *hemL* (SAUSA300_1614). Used to check transposon insertion in NE303.
hlgB_tn	TCACTTTGTGATTTTCCCAA	Gene specific primer for *hlgB* (SAUSA300_2367). Used to check transposon insertion in NE1682.
isdC_tn	TTATTCCACATTGCCTTTAGAT	Gene specific primer for *isdC* (SAUSA300_1030). Used to check transposon insertion in NE557.

NE151_tn	TTATCCTGCATTCTTTGACTC	Gene specific primer for hi77 ORF109-like protein (SAUSA300_1928). Used to check transposon insertion in NE151.
sbnC_tn	GCATTTTTGTACACGTCC	Gene specific primer for *sbnC* (SAUSA300_0120). Used to check transposon insertion in NE1031.
thiD_tn	AATTCATCGTCTAATCCCTCT	Gene specific primer for *thiD* (SAUSA300_2049). Used to check transposon insertion in NE1428.
carA_pOS1_rev	CACTAGATAACGTTTGCTTTGCAT-GCATATGTTCACCTCAATTGTATTTATCCC	SOE primer to amplify linearized pOS1-Plgt with *carA* overhangs
carA_pOS1_fwd	GGAGCGTCATATCAATGCCTAA-TCGAGGATCCAAACAAGGGGG	SOE primer to amplify linearized pOS1-Plgt with *carA* overhangs
pOS1_carA_fwd	GGGATAAATACAATTGAGGTGAACATATGC-ATGCAAAGCAAACGTTATCTAGTG	SOE primer to amplify *carA* with pOS1-Plgt overhangs
pOS1_carA_rev	CCCCCTTGTTTGGATCCTCGA-TTAGGCATTGATATGACGCTCC	SOE primer to amplify *carA* with pOS1-Plgt overhangs

Table S2. Primers used to verify Nebraska Transposon Mutant Library (NTML) strains and used to create complementation plasmids are listed

Sequence	Amino acid	N315 mapping	Protein/Gene	GWAS method
CAAATACAATACCA ATGACACGTCGATT GCTAATGACTATTT TAATAAACCGGCAA AGTACATTAAGAAA AATGGTAAATTGTA TGTTCAAATAACTG TAAACCACAGTCAT TGGATTACAGGAAT GAGTAT A -> G C -> A T -> A	S47S V72V T79T	1108175 - 1108306 (33 kmers) 1108199 1108274 1108295	WP_000789821.1 isdC	SEER Binary phenotype
CCAGCTCAAATTGA GGCTTATACTGAAG AAGTTGAGCAAGC AATTAAAGAAGCGC AACATGGAAATAAT CAACCAGCAGTTAA AGAATAATTAATTT GTACAATCATAAAC TG A -> G A -> T	E533E A534A	1300876 - 1300988 (14 kmers) 1300890 1300893	WP_001218596.1 glpD	SEER Binary phenotype
ACAGTCTTTTCAGT TGATTCTGAAATAC CATCTAAATCTACT TCTGCATCAACTTC GCTTATATTTTGCA AGTCATCTCGACTA CCGAAGATTTCAAT TTCTTTATCTTCTAA GTCAATCGAACTTA ATTCTTTATCATCT GCTAAACTACCTTT CTGTTTAACATTTA CTTTAACCTTTTTGC TA T -> C A -> G	L264L D249D E244E	2227323 - 2227508 (87 kmers) 2227391 2227436 2227451	WP_000894032.1 YbbR-like domain-containing protein	SEER Binary phenotype

C -> T				
C -> T	N220N	1182697	WP_001190913.1 carA	Bugwas Binary phenotype
T -> C (complement)	D194D	349263	WP_000745087.1 brnQ	Bugwas Binary phenotype
T -> C C -> T G -> A	D426D P643L A663A	473701 474351 474412	WP_000211545.1 DUF2309 family protein	Bugwas Binary phenotype
T -> C	L242L	2481094	WP_000783428.1 hlgB	Bugwas Binary phenotype
G -> A (complement) A -> G (complement)	M435I K436E	2380169 2380168	WP_000451828.1 Na+/H+ antiporter NhaC family protein	Bugwas Binary phenotype
T -> A (complement)	D83E	2532805	WP_001229087.1 opp1A	Bugwas Binary phenotype
T -> A	R282R	26493	WP_000871607.1 walK	Bugwas Binary phenotype
A -> T (complement)	P4P	1037680	WP_000104254.1 qoxD	Bugwas Binary phenotype
A -> G	V84V	299201	WP_000610204.1 tarJ	Bugwas Binary phenotype
C -> T	V467V	132180	WP_001179385.1 sbnC	Bugwas Binary phenotype
T -> C (complement) G -> T (complement)	R259R T321T	1706627 1706441	WP_000472302.1 clpX	Bugwas Binary phenotype
A -> G	D146G	546992	WP_000613722.1 folK	Bugwas Binary phenotype
C -> T (complement) T -> C (complement) G -> A (complement)	G98G V116V T164T	1710281 1710227 1710083	WP_000032657.1 NUDIX domain-containing protein	Bugwas Binary phenotype
G -> A (complement)	V227V	1702275	WP_001230232.1	Bugwas

C -> T (complement) C -> T (complement) C -> A,T (complement)	I251I T265T G285G	1702203 1702161 1702101	hemC	Binary phenotype
G -> A (complement)	T299T	1699418	WP_001270865.1 hemL1	Bugwas Binary phenotype
T -> C (complement)	N263N	272940	WP_001008400.1 fadE	Bugwas Binary phenotype
G -> A (complement)	A102A	1680683	WP_000457386.1 50S ribosomal protein L21	Bugwas Binary phenotype
A -> G (complement)	G277G	1676466	WP_001005768.1 ruvB	Bugwas Binary phenotype
T -> A T -> C	S146S C179C	1697413 1697512	WP_001001237.1 tag	Bugwas Binary phenotype
C -> T, A (complement)	R353R	271053	WP_000142198.1 fadD	treeWAS
A -> G	K116K	2803540	WP_000154162.1 vraD	treeWAS
A -> G (complement)	L69L	316218	WP_000757880.1 degA	treeWAS
T -> C	V549V	19974	WP_001081640.1 cyclic-di-AMP phosphodiesterase	treeWAS
G -> A	G520G	240930	WP_000236427.1 ggt	treeWAS
A -> G, T (complement) C -> T (complement)	V287V Y319Y	2527215 2527119	WP_000675401.1 MFS transporter	treeWAS
T -> G	P347P	1306337	WP_000652047.1 aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme	treeWAS
T -> A	V360V	885317	WP_001074405.1	treeWAS

			sufB	
C -> T	D113D	2698328	WP_000198598.1	treeWAS
C -> A,T (complement)	H181N, H181Y	312161	WP_000031262.1 class I SAM-dependent methyltransferase	treeWAS
G -> T, A (complement)	G196V, G196D	2148703	WP_000594954.1 thiD	treeWAS
GATAGGGGCTATTTTAATAAAATTCGTCCTC	N/A	1383811 - 1383841	Intergenic, 18 bp into end of WP_000523668.1 SWIM zinc finger family protein	DBGWAS

Table S3. Summary of significant GWAS hits

Softwares used were SEER, bugwas, treeWAS, and DBGWAS. *S. aureus* N315 was used as reference genome to map GWAS hits.

Chapter IV. Effect of genetic background on the direction of evolution of Vancomycin-Intermediate *Staphylococcus aureus* (VISA)

Michelle Su, Michelle Hargita, Jessica Peterson, Sarah Satola, Timothy D. Read

Abstract

Vancomycin has been a relatively safe and economical treatment against MRSA for the past three decades but increasing use of the drug has contributed to a growing number of reports of less susceptible strains. High-level vancomycin-resistant *Staphylococcus aureus* (VRSA; minimum inhibitory concentration (MIC) ≥ 16 μg/mL) has occurred through horizontal acquisition of the *vanA* gene but has remained rare. VISA (MIC 4-8 μg/mL), first reported in 1997 arises from vancomycin-susceptible *S. aureus* (VSSA; MIC ≤ 2 μg/mL) through stepwise acquisition of mutations in a limited but not yet completely defined set of genes. Understanding the fitness limitations and genetic background effects on acquisition of VISA mutations is a step toward better treatments and diagnostics. We created a library of laboratory selected VISA strains from three important *S. aureus* backgrounds - (CC1, CC5 and CC8). We detected mutations in more than 150 genes, but only six genes occurred in all three background strains (*walK, prs, rpoB, rpoC, vraS, yvqF*). These six were already known to be commonly VISA targets, but the prevalence, co-occurrence, and SNP patterns within these genes suggested epistatic interactions between VISA loci dependent on genetic background. Linear regression models also indicate that different genes are predictive for different backgrounds, perhaps due to differential selective pressures arising from epistasis. Further, fitness costs were observed with increasing vancomycin resistance regardless of background; however, NRS70 VISA strains were able to achieve resistance at lower levels with lower fitness costs compared to NRS384 VISA strains. These results

together indicate that VISA is a phenotype that has many evolutionary paths, with genetic background skewing strains towards a subset due to epistasis from a vast genetic network.

Introduction

Vancomycin has been a relatively safe and economical treatment against MRSA for the past three decades, but increasing use of the drug has contributed to a growing number of less susceptible strains (Gardete & Tomasz, 2014; Zhang et al., 2015; McGuinness, Malachowa & DeLeo, 2017). High-level vancomycin-resistant S. aureus (VRSA; minimum inhibitory concentrations (MIC) ≥ 16 μg/mL) has occurred through horizontal acquisition of the *vanA* gene from Enterococcus spp (Kobayashi, Musser & DeLeo, 2012). The incidence of VRSA has remained rare, likely due to poor regulation of expression of the *vanA* gene in *S. aureus* imposing a heavy fitness burden on the bacterium (Foucault, Courvalin & Grillot-Courvalin, 2009). VISA (MIC 4-8 μg/mL), first reported in 1997 in the case of a 4 month old infant with a MRSA wound infection following cardiac surgery (Hiramatsu et al., 1997), arises from vancomycin-susceptible *S. aureus* (VSSA; MIC ≤ 2 μg/mL) through mutagenesis. hVISA are susceptible in vitro to vancomycin (MIC ≤ 2 μg/mL) but contain subpopulations, in a ratio of 1 to 10^6 cells, that can grow in the presence of ≥ 4 μg/mL of vancomycin and thus are capable of spontaneous transitions to VISA (Liu & Chambers, 2003; Sakoulas & Moellering, 2008; Deresinski, 2009; El-Halfawy & Valvano, 2015).

Standard microbiology laboratory-based phenotypic tests for VISA are labor intensive, and the incidence may be much higher than currently reported (Marlowe et al., 2001; Charles et al., 2004; Prakash, V, Lewis & Jorgensen, 2008; Satola et al., 2009; Swenson et al., 2009; Vaudaux et al., 2010). hVISA strains are particularly challenging to detect using current standard clinical microbiology methods (automated broth microdilution assays) as MICs can overlap with VSSA (Swenson et al.,

2009). Development of a nucleic acid test using standard PCR based approaches has been confounded to date by the complexity of the genetics of VISA. VISA has emerged independently in strains from each of the major MRSA lineages (Hiramatsu et al., 1997; Howden et al., 2006, 2008; Mwangi et al., 2007; Alam et al., 2014; Gardete & Tomasz, 2014). The classic study by Mwangi et al (Mwangi et al., 2007) showed VISA evolving within a patient on long-term vancomycin therapy through a series of adaptive mutations. Most intermediate resistance is acquired by within-patient evolution (Klevens et al., 2007; Gardete & Tomasz, 2014), although there have been rare reports of person-to-person spread (Cassone et al., 2004; de Lassence et al., 2006). Some patients with MRSA infections failing therapy showed enhanced vancomycin MIC even when the drug was not used (Horne et al., 2009; Lalueza et al., 2010; Holland & Fowler, 2011), suggesting that VISA may overlap with 'persister' phenotypes (Johnson & Levin, 2013) associated with long-term invasive infection.

The genetic basis of VISA has been investigated for several years, primarily through identification of mutations by comparative sequencing of small numbers of isogenic clinical samples (Ohta et al., 2004; Mwangi et al., 2007; Howden et al., 2008) and molecular genetic characterization. VISA cells typically show cell-wall thickening and a reduction of fitness in growth compared to isogenic parents. Mutations in a variety of conserved 'core' *S. aureus* genes have been reported to associate with VISA (Howden et al., 2010; Gardete & Tomasz, 2014; Wang et al., 2016), most commonly those involved in regulation of cell wall architecture (e.g., *graRS* (Neoh et al., 2008; Howden et al., 2008), *vraRS* (Kato et al., 2010; Baek et al., 2017; Asadpour & Ghazanfari, 2019), *yvqF* (Kato et al., 2010; Yoo et al., 2013), *walK* (Shoji et al., 2011; Hafer et al., 2012; McEvoy et al., 2013; Hu et al., 2015), *walR* (Howden et al., 2011)), or pleiotropic effects of certain global transcriptional regulators (e.g. *agr* (Sakoulas et al., 2003), *stp1* (Cameron et al., 2012; Passalacqua et al., 2012), *rpoB* (Cui et al., 2010; Watanabe et al., 2011; Hafer et al., 2012; Katayama et al., 2017). Laboratory selection experiments have shown that an *S. aureus*

genotype can acquire the VISA phenotype through mutations in different genes (Passalacqua et al., 2012; Vidaillac et al., 2013).

In this study, we aim to further evaluate the relationship between genetic determinants of vancomycin resistance and genetic background and determine if epistasis can direct the evolutionary trajectory of this phenotype. We evolved VISA strains from three common genetic backgrounds encountered clinically: Clonal complex (CC) 5 (N315), CC 8 (USA300-0114), CC 1 (MW2) and analyzed the differences in mutation, fitness, and resistance levels.

Methods

Strains

Parental strains used for the evolution experiments were NRS70 (N315), NRS123 (MW2), and NRS384 (USA300-0114) and were obtained from BEI resources (https://www.beiresources.org/).

Evolution experiments

Frozen stocks of the three parent strains were streaked on Brain Heart Infusion (BHI) agar plates and single colonies used to establish shaking overnight cultures of BHI broth. We then propagated independent lines in BHI broth shaking at 37C until cultures were turbid (OD600 > 1.5) before diluting culture into fresh media. Strains were grown initially in 1 μg/mL vancomycin then transferred to broth with 2 μg/mL antibiotic. Once turbid, the cultures were similarly diluted to 4 μg/mL then grown for 3 days before transfer to 8 μg/mL and grown for a final 3 days. After the growth period in broth at the final antibiotic concentration, cultures were plated on BHI agar plates containing 2, 4 and 8 μg/mL vancomycin. A single colony was picked from the highest concentration at 48 hours growth. The colony was passaged twice on selective media then examined by Gram stain to confirm species.

Genomic Analysis

DNA extraction and library prep were performed as manufacturer's instructions (Wizard Genomic DNA Purification Kit, Promega; Nextera XT DNA Library Prep Kit, Illumina). Genome sequencing was performed on Illumina HiSeq and MiSeq with paired-end reads. Raw read data were deposited in the NCBI Short Read Archive under project accession number PRJNA525705.

Mutations were called based on comparison to reference sequences using breseq (Deatherage & Barrick, 2014). Reference genomes used for NRS70, NRS123, and NRS384 VISA strains were NC_002745.2 (BioProject: PRJNA224116, BioSample: SAMD00061099, Assembly: GCF_000009645.1), NC_003923.1 (BioProject: PRJNA224116, BioSample: SAMD00061104, Assembly: GCF_000011265.1), and NZ_CP027476.1 (BioProject: PRJNA224116, BioSample: SAMN07411405, Assembly: GCF_002993865.1) respectively. To compare mutations across the three strains, the protein sequences of NRS123 and NRS384 were blasted against that of NRS70 (Camacho et al., 2009). Genes were considered orthologous to NRS70 if the amino acid sequences had at least 95% identity, 95% coverage, or an evalue=0.

Linear regression analysis

Linear regression models were fitted using the *lm* function in R. *walK, prs, rpoB, rpoC, vraS, and yvqF* as binary (mutated or not) variables were used as predictors for log transformed vancomycin MICs. The final model was chosen by backwards selection with the goal of minimizing Akaike information criterion (AIC) as in Eyre et al (Eyre et al., 2017).

Growth curve analysis

Strains were grown in duplicate in a 96-well plate beginning at an OD < .1 and grown for 24 hours at 37C with constant shaking in a Biotek Eon Microplate Spectrophotometer, with OD measurements every 10 minutes. To assess growth-curves, OD readings were imported into R (Team, 2016) and maximal growth rate (r) calculated using the growthcurver package (Sprouffske & Wagner, 2016). Fitness was calculated as the ratio between the average r of each evolved strain to the average r of the parent strain. Area under the curve (AUC) and final OD600 measurements were also considered as alternative fitness metrics, but the high variation between replicates made them unsuitable for further analyses. Fitness distributions between NRS70 evolved VISA strains and NRS384 evolved VISA strains were compared using a two-sided two-sample Kolmogorov-Smirnov test.

Other statistical analyses

To determine if the difference in the prevalence of mutations across different VISA genes (*walK, prs, rpoB, rpoC, vraS, yvqF*) was significant, a Pearson's Chi-squared test was performed with post hoc pairwise comparisons using Fisher's exact test and a Bonferroni multiple test correction. To analyze if the prevalence of mutations was also affected by background, binomial generalized linear models (GLMs) were fitted for each gene, and pairwise Wald tests were performed to test whether the difference in the coefficients for each background was non-zero. The statistical significance of gene correlation was assessed using a Spearman correlation with a Holm multiple testing correction. Total SNP distributions were compared using a Kruskall-Wallis test followed by pairwise Mann-Whitney U tests. Synonymous and nonsynonymous SNP distributions were compared with a Pearson's Chi-squared test followed by Bonferroni corrected Fisher's exact tests. The vancomycin and daptomycin MIC distributions between NRS70 VISA strains and NRS384 VISA strains were compared using two-sided two-sample Kolmogorov-Smirnov tests. To test if the linear relationship between vancomycin

MIC and daptomycin MIC was dependent on background (NRS70 and NRS384), an Analysis of Covariance (ANCOVA) was performed. All analyses were performed using R.

Antimicrobial susceptibility testing

Antimicrobial susceptibility testing for daptomycin, vancomycin, and methicillin were done according to Clinical Laboratory Standards Institute (CLSI) standards (Clsi, 2013). In brief, cultures were streaked from a frozen stock onto BHI agar and restreaked the next day. From the second day culture, colonies were resuspended in normal saline to a .5 McFarland standard. For plate-based testing, e-test for daptomycin and disc diffusion with cefoxitin, cultures were struck to create a lawn on cation-adjusted Mueller-Hinton broth (CAMHB) agar plates before application. Plates were incubated at 35C and read at 18 hours. For broth microdilution (BMD) to determine strain vancomycin minimum inhibitory concentration (MIC), .5 McFarland standards were diluted 1:20 in normal saline, and 10 ul was added to 90 ul of CAMHB with the appropriate concentration of vancomycin. Plates were incubated at 35C without shaking and read at 24 hours.

Results

Patterns of acquisition of SNPs at VISA endpoints is affected by genetic background

To study the evolution of vancomycin intermediate resistance, we selected three strains, representing major lineages in the US: N315/NRS70 (Kuroda et al., 2001) (CC5); USA300-0114/NRS384 (Diep et al., 2006) (CC8) and MW2/NRS123 (Baba et al., 2002) (CC1). We experimentally evolved *S. aureus* strains with a vancomycin MIC of 4-10 μg/mL from an initial parental MIC of 1 μg/mL. Selection for 120 of these mutants, 40 per genetic background, was achieved by sequentially raising the vancomycin concentration through serial passages of strains in BHI broth. Mutants were sequenced and analyzed with breseq (Deatherage & Barrick, 2014) to ascertain mutations acquired during

vancomycin selection. Thirteen evolved strains are excluded from further analyses due to either low sequence quality or culture contamination. The final number of evolved VISA strains per background was 35 NRS70 strains, 34 NRS123 strains, and 38 NRS384 strains.

Within each background, most deletions, insertions, and substitutions only occurred in one strain, and little parallelism was exhibited. In cases where the deletions are found more commonly within a background, mutations are in intergenic regions and of unclear significance. Of note, large-scale deletions of 42-44kb in the same region were observed in all genetic backgrounds (four NRS70, four NRS123, four NRS384). A list of mutations greater than 2 bp can be found in Tables 1-5. The bulk of the genetic changes that occurred during the selection for vancomycin-intermediate resistance were single nucleotide polymorphisms (SNPs) and will be the focus of the rest of the analyses.

We found between 5 and 25 SNPs in evolved VISA strains compared to their respective parent strains, with non-synonymous mutations in several well-known VISA-associated genes. SNPs fell within 151 coding genes, and no gene was universally mutated in all strains, highlighting the existence of multiple pathways to vancomycin resistance in *S. aureus*. However, of the 151 genes, 121 genes were mutated in only one of the three backgrounds; 24 genes were mutated in two backgrounds; and only 6 genes (*walK, prs, rpoB, rpoC, vraS, yvqF*) were mutated in all three backgrounds. These 6 genes have been previously implicated in VISA from sequencing of clinical isolates, but interestingly, some genes that had been previously reported did not acquire mutations frequently or in all three backgrounds. Notably, *agr*R was not mutated in any strains, *graRS* mutants were found in only NRS384 derived strains and *walR* (Howden, Peleg & Stinear, 2013) only in NRS70 and NRS123 derived strains.
In examining the 6 "universally mutated" genes (i.e. those occurring in all three backgrounds - *walK, prs, rpoB, rpoC, vraS, yvqF*), there was an indication that some genes were more broadly important to

the evolution of VISA, with indications of lineage specific effects. On opposite sides of the spectrum, mutations in *prs* occur in fewer than 10% of the evolved strains compared to mutations in *walk* which occurred in 80-95% of the evolved VISA strains (Table 6). A Pearson's chi-squared test ($p < 2.2 \times 10^{-16}$) and post hoc Fisher's exact tests (all $p < 5.91 \times 10^{-03}$) determined that mutation rates differed significantly between genes. The only genes with comparable mutation rates were *rpoB*, *rpoC*, and *vraS*. By fitting binomial generalized linear models (GLMs) to the gene mutation distributions, mutation rates in *rpoB* (NRS70 vs NRS123, p = .03; NRS70 vs NRS384, p = .015), *rpoC* (NRS70 vs NRS384, p = .019), *vraS* (NRS70 vs NRS384, p = .04), *and yvqF* (NRS70 vs NRS123, p = .035; NRS70 vs NRS384, p = .028) were found to be significantly influenced by genetic background. In the case of *rpoC* and *vraS*, the mutation rates of NRS70 and NRS384 represent the ends of the spectrum with NRS123 as an intermediate. In the case of *rpoB* and *yvqF*, NRS123 and NRS384 share similar mutation rates, which differ significantly from the mutation rates of NRS70.

While selective pressures on individual genes may differ between backgrounds, similar selective pressures may exist on the amino acid level. To examine the mutational profile of SNPs, we focused on *yvqF* and *walK* due to their high mutation rate and 100% conservation across backgrounds. In *yvqF*, there are a total of 32 SNPs spread across 68 strains: 3 SNPs are found in all three backgrounds, 10 SNPs are found in two, and 19 SNPs in only one. Intriguingly, despite *walK* mutations (41 SNPs across 93 strains) being more common, there are no SNPs in common between the three backgrounds and 31 SNPs are found in only one. Furthermore, the most common SNPs within each background for these two genes were found in different amino acid residues (*prs*: NRS70/A152V, NRS123/P126S, NRS384/P174A; *walK*: NRS70/E236D, NRS123/I29M, NRS384/Q216E). The lack of commonality in the SNPs found between these genetic backgrounds indicated a strong effect of epistasis on the protein architectures of *walK* and *yvqF*.

Because VISA is a stepwise process and requires the action of more than one mutation, we next examined the co-occurrence of mutations in the six universally mutated genes (Figure 1a-d). Broadly, mutations in *vraS* and *yvqF* were almost mutually exclusive (coefficient -.8, $p < .0001$), with minor differences between genetic backgrounds (Figure 1b-d, NRS70 and NRS384: -.9; NRS123: -.6). Interestingly, NRS70 was the only background with significant associations between other genes (Figure 1b): *prs*/*vraS* (coefficient .5, $p = .011$) and *prs*/*yvqF* (coefficient -.5, $p = .035$).

Considering the complex genetic networks that exist within cells and the multiple routes available to become VISA, it may not be surprising that only six genes are mutated in all three backgrounds. However, similarities between backgrounds may exist on a smaller scale. Between NRS70 and NRS123, 5 other genes were mutated in both backgrounds. Similarly, between NRS123 and NRS384, 6 other genes were mutated. Finally, between NRS70 and NRS384, 12 other genes were mutated (Table 7). The higher proportion of other shared genes between NRS70 and NRS384 may indicate that these two strains share more similar evolutionary paths toward vancomycin resistance.

In analyzing the multi-mutational nature of VISA, it was of interest to examine how many mutations occurred per strain to gain high intermediate vancomycin resistance. NRS123 VISA strains had fewer mutations than NRS70 ($p = .004$) and NRS384 ($p = 2.8 \times 10^{-5}$) VISA strains (Figure 2). Furthermore, the distribution of synonymous SNPs to nonsynonymous SNPs was not the same within each genetic background (Table 8, all $p < .05$). While synonymous SNPs were often regarded as having little functional impact, mutations in *sdr* genes were enriched for in our evolved VISA strains. *sdrD* was mutated in 8 out of 35 NRS70 VISA strains, accounting for 42% of the total synonymous SNPs found; and *sdrC* was mutated in 13 out of 38 NRS384 VISA strains, accounting for 29% of the total

synonymous SNPs found. The *sdr* locus encodes *sdrC, sdrD, sdrE* which are members of the microbial surface components recognizing adhesive matrix molecules (MSCRAMM) family (Josefsson et al., 1998). These genes are not always conserved together (Liu et al., 2015), and their presence has been associated with bone infection (Sabat et al., 2006), resistance to host immunity (Sitkiewicz, Babiak & Hryniewicz, 2011; Askarian et al., 2017), biofilm formation (Barbu et al., 2014), and host-cell adhesion (Corrigan, Miajlovic & Foster, 2009; Cheng et al., 2009; Barbu et al., 2010; Askarian et al., 2016).

Resistance phenotypes of evolved VISA strains

All evolved VISA strains were able to maintain growth in 8 μg/mL vancomycin BHI broth at the end of selection. However, to ensure strains were truly resistant and not merely tolerant, we performed BMD on NRS70 and NRS384 strains according to CLSI standards (unfortunately, all NRS123 stocks were lost while in -80C storage). All strains had MICs within the range to be considered VISA (Figure 3), but some (20 NRS70 and 28 NRS384 strains) were measured to be below 8 μg/mL. The NRS70 VISA strain MICs ranged from 4 to 10 μg/mL, with a median and mode of 6 μg/mL, while the NRS384 VISA strain MICs ranged from 4 to 8 μg/mL with a median and mode of 6 μg/mL. The difference in the MIC distributions is not statistically significant between NRS70 and NRS384 strains ($p = .323$), however, NRS70 VISA strains were able to achieve 10 μg/mL, while NRS384 strains were not.

The cross-resistance of vancomycin and daptomycin in *S. aureus* is well documented. NRS70 and NRS384 evolved strains were thus tested for development of daptomycin resistance by E-test. Parental strains had daptomycin MICs of .38 μg/mL. Thirty of the 35 NRS70 strains became resistant, and 33 of the 38 NRS384 strains became resistant (Figure 4). The daptomycin distributions between NRS70 and NRS384 were not statistically different ($p = 1$), however, only NRS384 strains were able to achieve

a daptomycin MIC of 6 μg/mL versus 4 μg/mL for NRS70 strains. To test the hypothesis that greater levels of vancomycin resistance correlated to greater levels of daptomycin resistance, we fitted a linear regression on both distributions and found this to be true (Figure 5). The genetic background of the strains did not play a role ($p = .822$).

A less reported phenomenon in VISA strains is collateral sensitivity with beta-lactams. Within our evolved NRS70 and NRS384 strains, only 2 strains lost methicillin resistance. These two strains were in a NRS70 background. Evidence for obvious mechanisms for beta-lactam sensitivity in a MRSA background such as loss of SCCmec or inactivation of *mecA* was not found, and the exact mechanism remains unknown.

Fitness generally decreases during evolution of higher resistance but is modulated by genetic background

Higher resistance in our evolved VISA strains generally equated to lower fitness, measured as the maximum growth rate relative to the parent (Figures 6a-b). However, there was evidence of compensatory mutations by the presence of "V" shaped lines. In the case of right-side up "V" shaped fitness lines, fitness decreased as vancomycin resistance or tolerance continued to rise but was at least partially rescued by further mutations. For upside-down "V" shaped fitness lines, fitness was rescued at an earlier stage of vancomycin resistance or tolerance development before further mutation. Both types of "V" shaped lines are comparatively rare, and the majority of fitness lines proceed in a relatively linear negative trend as vancomycin resistance and tolerance develops. Looking at the fitness distributions at each selection stage (4 μg/mL, 6 μg/mL, 8 μg/mL) between NRS70 and NRS384, there was a striking difference at 4 μg/mL ($p = 7.08 \times 10^{-6}$) and 6 μg/mL ($p = 1.86 \times 10^{-6}$), which

disappears at 8 μg/mL ($p = .273$). NRS70 strains may thus be able to tolerate lower levels of vancomycin resistance with lower fitness cost compared to NRS384 strains.

Genetic predictors of vancomycin MIC

Linear regression models were fit onto the genomic data for NRS70 and NRS384 strains to determine which genes were most predictive of vancomycin resistance. To reduce the number of parameters and for ease of comparison, only the 6 universally mutated genes were used in the initial models (Table 9). To make the models more meaningful, parameters were evaluated for inclusion in the final models using AIC. In the final model for NRS70, *walK, rpoC, vraS,* and *yvqF* were statistically significantly associated with vancomycin resistance, defined as log(MIC) ($R^2 = .583$, $p = 2.15 \times 10^{-6}$). In comparison, only *walK* by itself was statistically significant in the final model of NRS384 ($R^2 = .313$, $p = 9.95 \times 10^{-4}$), although *vraS* and *yvqF* were also included for improved model fit. The difference in models between the two different genetic backgrounds indicated that while these core genes were broadly responsible for vancomycin resistance, they were of differential importance within each background and thus likely under different selective pressures due to epistatic interactions with the background.

Discussion

The emergence of vancomycin-intermediate resistance within many *S. aureus* lineages and the implication of the same genes and operons (e.g *walKR*, *rpoB/C*, *vraTSR*) in many cases indicated that this phenotype is strongly driven by parallel evolution. However, *S. aureus* lineages have been shown to be heterogeneous in several significantly clinically relevant phenotypes such as toxin production, biofilm formation, and host immune resistance (King et al., 2016; Su et al., 2020) as well vary in gene content (Lindsay et al., 2006; McCarthy & Lindsay, 2010). As VISA requires several mutations, which are often in regulatory genes, the probability of epistatic interaction of these mutations with lineage

defining variation is high. In this study, we evolved VISA in three clinically important *S. aureus* lineages (CC1, CC5, CC8) to determine how epistasis shapes the evolution of vancomycin-intermediate resistance and demonstrated that genetic background can play in role in determining what genes are mutated and how genes are mutated.

The primary mode of mutation to achieve VISA in our experiment was through SNPs. The majority of changes were SNPs, and other genetic variation (indels, substitutions) showed no discernable pattern. Six previously implicated VISA genes (*walK, prs, rpoB, rpoC, vraS, yvqF*) (Howden et al., 2010) were also mutated in our three genetic backgrounds, but we found that mutation rates (probability that the gene was mutated in a strain) differed for three out of the six genes. *walK* was the most universally seen mutation, *yvqF* was also commonly mutated and *prs* was the least common. *rpoB, rpoC,* and *vraS* mutation rates were similar and intermediate between the others. Understanding the likelihood of mutation is a step toward PCR tests as a monitoring mechanism for within-host development of vancomycin resistance. This knowledge needs to be informed by genetic background as mutation rates for all but *prs* and *walk* were affected by genetic background. For *rpoB, rpoC, vraS,* and *yvqF*, NRS70 and NRS384 mutation rates were most different, and NRS123 mutation rates were intermediate between the others. These results may be partially explained by the closer genetic relationship between CC1 and CC8 than CC5 and CC8 (Petit & Read, 2018). At the amino acid level of *walK* and *yvqF*, genetic background seemed to dominate, with few SNPs shared across the genetic backgrounds and different amino acid positions mutated in each genetic background. This indicates that differential protein-protein interactions or protein function imposes selection at unique amino acid sites specific to each genetic background (Chi & Liberles, 2016).

Other genes were found to have been mutated in two of the three genetic backgrounds and may have a role in vancomycin resistance or act as compensatory mutations. Several are already validated as being associated with vancomycin resistance (*walR* (Howden et al., 2011), vraR (Kato et al., 2010; Baek et al., 2017; Asadpour & Ghazanfari, 2019), *mprF* (Ruzin et al., 2003; Chen et al., 2018)). NRS70 and NRS384 backgrounds had more of these genes in common than with NRS123, indicating that NRS70 and NRS384 strains may experience more similar selective pressures in the larger mutational landscape outside of the previously discussed six genes.

The overall vancomycin resistance levels achieved after experimental evolution in NRS70 and NRS384 VISA strains (NRS123 VISA strains were not phenotyped due to loss of strains during storage) were similar, suggesting both backgrounds are equally capable of achieving lower and higher levels of vancomycin intermediate resistance. Linear regression demonstrated that genetic predictors of vancomycin resistance differed between backgrounds. *walK, rpoC, vraS*, and *yvqF* were all significantly associated with vancomycin MICs in NRS70, but only *walK* was significant in NRS384. As NRS384 strains carried more mutations than the other two backgrounds, this may indicate that while a subset of the 6 universally mutated genes (*walK, prs, rpoB, rpoC, vraS, yvqF*) are necessary for vancomycin resistance, other genes may play a significant role in determining the level of vancomycin resistance achieved in NRS384 VISA strains but not NRS70 strains.

Daptomycin cross-resistance (Cui et al., 2006; Chen, Huang & Chiu, 2015) occurred in the majority of our VISA strains, and the level of daptomycin resistance correlated with the level of vancomycin resistance regardless of genetic background and is likely related to thickened cell walls and limited cell penetration by both antibiotics (Cui et al., 2003). Two NRS70 VISA strains were phenotyped as methicillin sensitive by cefoxitin disk diffusion while the remaining NRS70 and NRS384 VISA strains

remained resistant. While not directly interrogated, the "seesaw effect" between beta-lactams and vancomycin (Barber et al., 2014) is likely to be seen in our VISA strains and responsible for the loss of resistance in these strains.

Finally, we demonstrated the vancomycin-intermediate resistance imposes a fitness cost (measured by maximal growth rate compared to parent strain) to *S. aureus* that linearly scales with the level of vancomycin resistance in most cases. However, "V" shaped fitness lines as *S. aureus* was grown in 4, 6, and 8 μg/mL of vancomycin suggest that compensatory mutations can rescue fitness at different stages of vancomycin-intermediate resistance evolution. These nonlinear fitness lines were observed in both NRS70 and NRS384 backgrounds. Fitness distributions between these two genetic backgrounds were markedly different at stages 4 μg/mL and 6 μg/mL, with vancomycin resistance imposing less of a fitness cost on NRS70 VISA strains than NRS384 VISA strains. This may have clinical implications as it suggests that NRS70 *S.aureus* strains may evolve vancomycin resistance quickly, not be rapidly selected against due to growth defects, and may reach fixation if vancomycin concentrations are sustained and lead to the extinction of the ancestral vancomycin susceptible strain.

This study has utilized experimental evolution to elucidate relationships between genetic background and mutations related to vancomycin-intermediate resistance in *S. aureus*. Its in vitro nature has given us the power to evolve VISA multiple times independently and achieve statistical significance. Even so, the clinical relevance is not entirely clear as most VISA strains are the result of within-host evolution, and immune pressure is not present within our experimental design. Other factors that are present in human infection that were omitted in our experiment include but are not limited to nutrient limiting conditions and microenvironments. These undoubtedly reduce the mutational space *S. aureus* strains have access to. In addition, the mutations in our studies were grouped by gene for simplicity,

but likely not all mutations in VISA genes contributed meaningfully to vancomycin resistance. The sheer number of mutations makes it impossible to recapitulate these mutations individually. Lastly, fitness determinations by measuring maximal growth rate do not always correlate with competitive indices of evolved strains to ancestor strains and have been shown to be sensitive to environmental conditions. Thus, our measured fitness values may not accurately reflect within-host fitness, especially in a competitive environment.

In summary, the interactions between VISA determinants and genetic background are complex, and the strength of epistasis varies based on what level. At the gene level, there is clear parallel evolution in several genes that is not affected by genetic background, but more accessory genes (those mutated in only 2 backgrounds) show clear signs of differential selective pressure related to genetic background, with NRS70 and NRS384 more similar to one another. At the level of gene mutation frequency and importance, NRS70 and NRS384 are more dissimilar to one another than to NRS123, highlighting epistasis can drive evolution of a phenotype both in similar and dissimilar ways. Finally, at the amino acid level, all genetic backgrounds examined were unique, and epistasis strongly influenced the course of VISA evolution.

Figures and tables

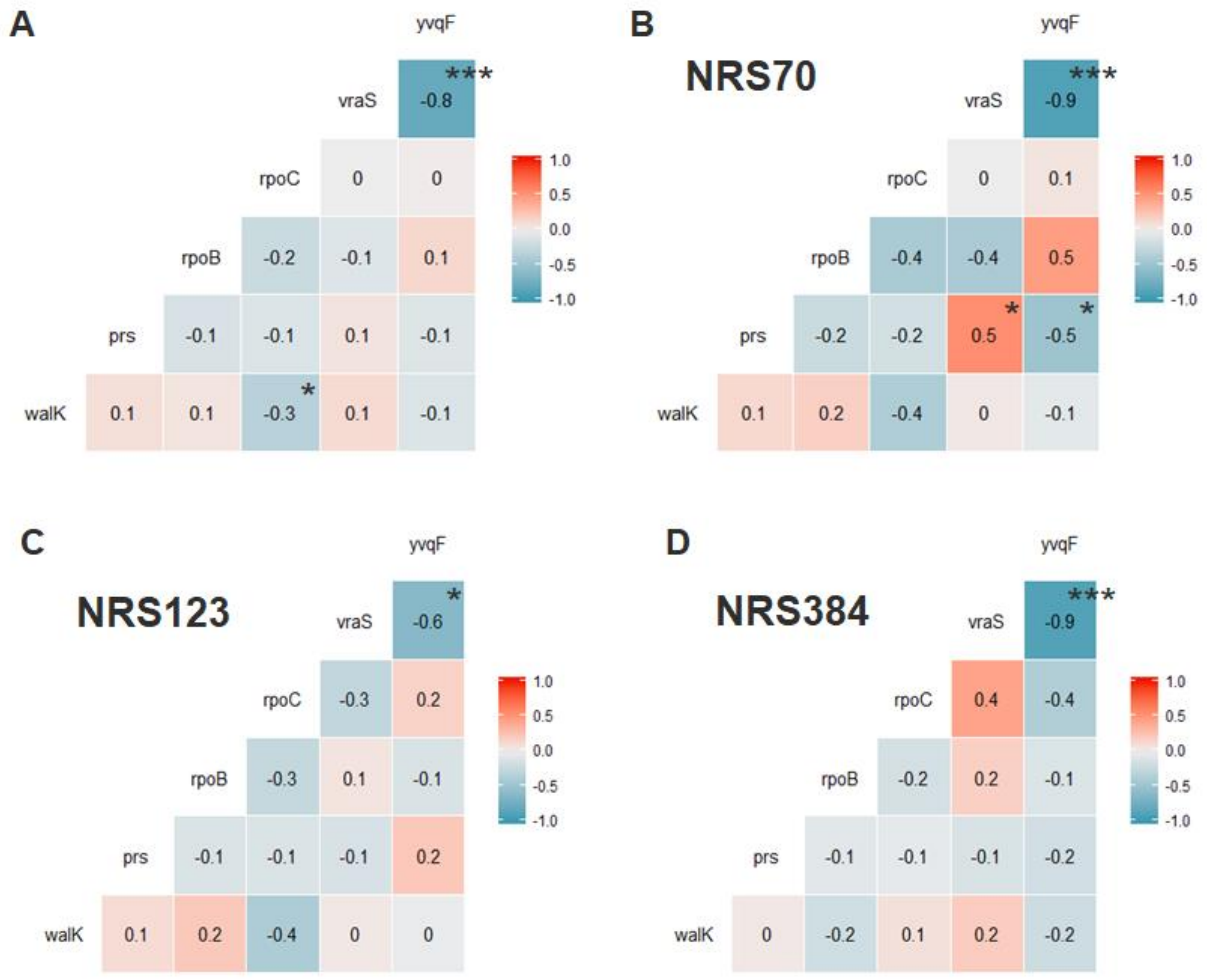

Figure 1. Correlation of mutations between VISA genes.

The NRS70 *walK* mutations considered are SNPs as well as inframe deletions. All other genes as well as *walK* in NRS123 and NRS384 include only SNPs. (a) Correlation matrix of all evolved VISA strains. (b) Correlation matrix of all NRS70 evolved VISA strains. (c) Correlation matrix of all NRS123 evolved VISA strains. (d) Correlation matrix of all NRS384 evolved VISA strains. Significance codes: '***' <0.0001 '**' <0.001 '*' <0.05

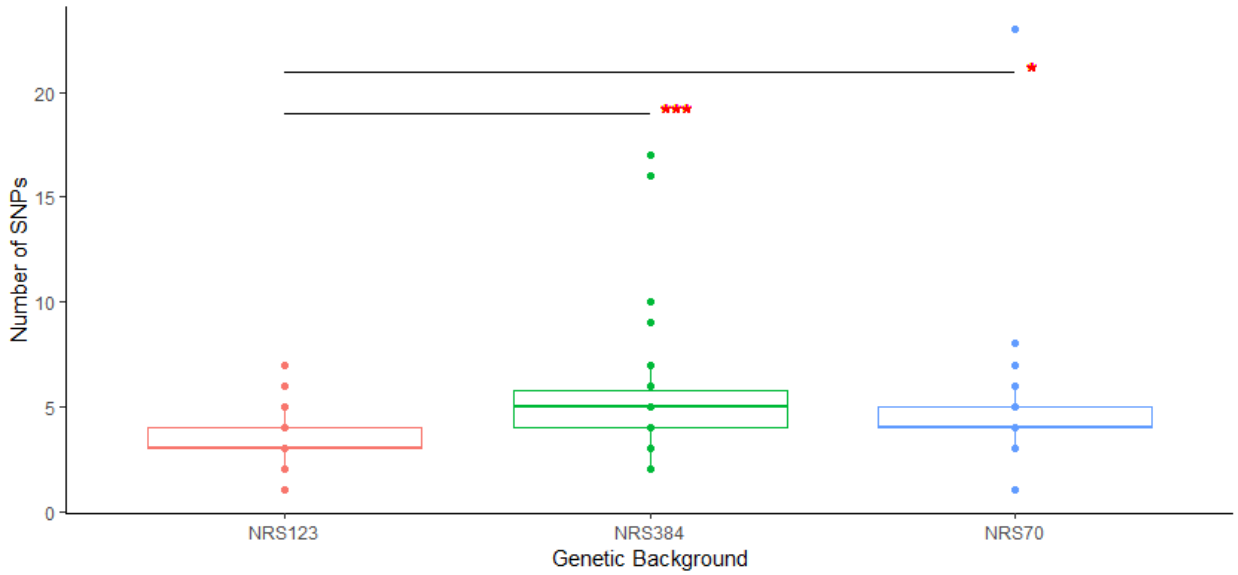

Figure 2. SNP distributions of VISA evolved strains differ by genetic background.

Average SNPs per background: NRS70 (4.26), NRS123 (3.35), NRS384 (4.37). Mode and median of SNPs per background: NRS70 (4), NRS123 (3), NRS384 (5). Significance codes: '***' <0.0001 '**' <0.001 '*' <0.05

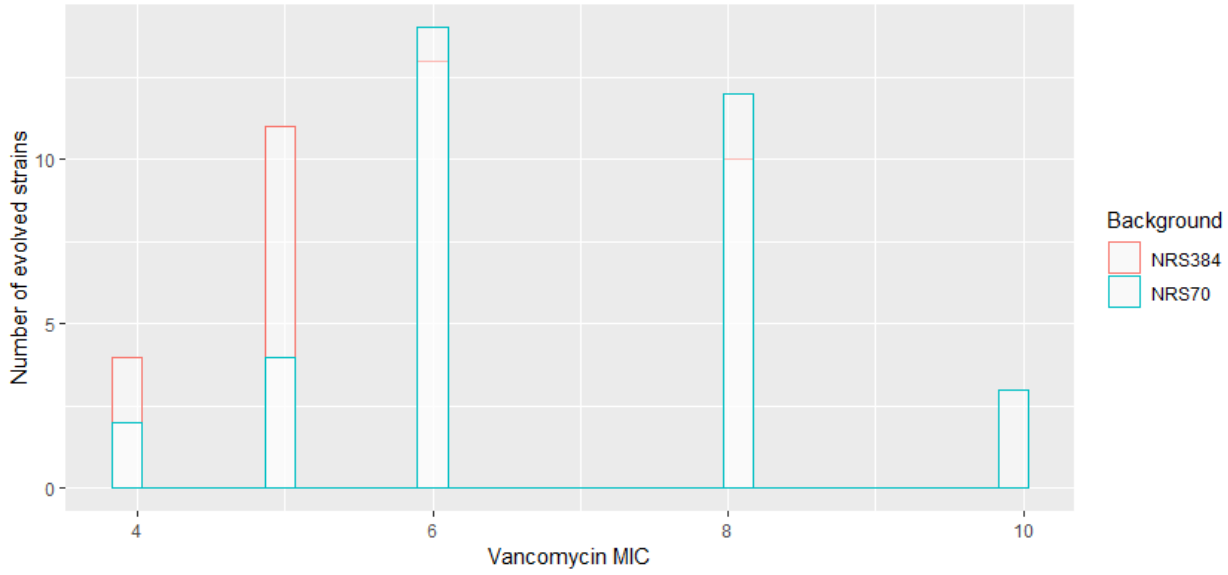

Figure 3. Vancomycin MIC distribution of NRS70 and NRS384 evolved strains.

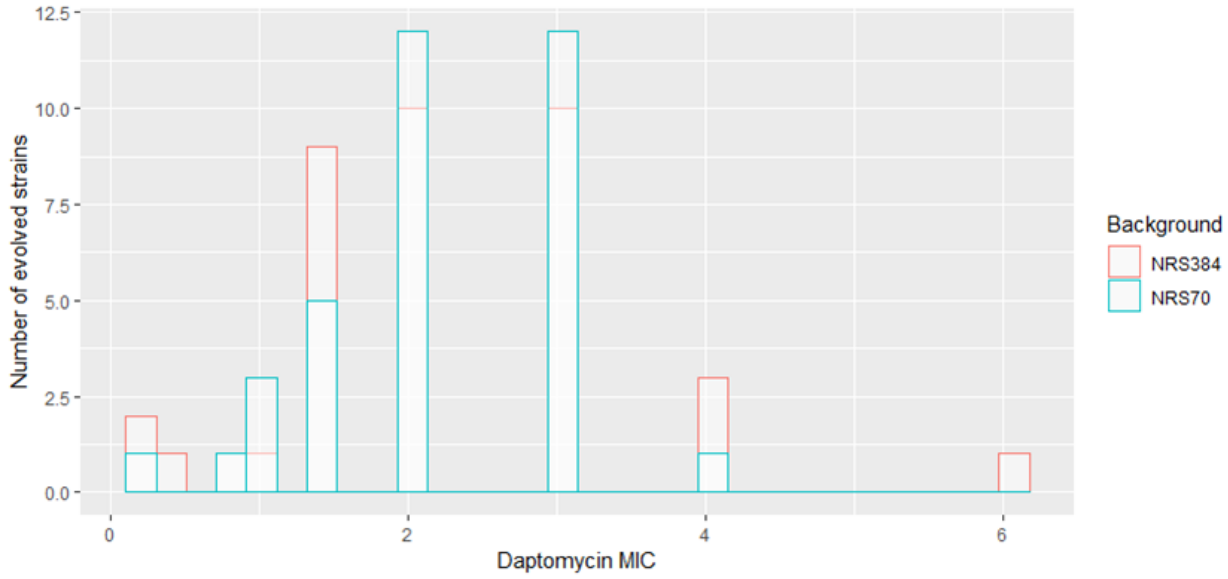

Figure 4. Daptomycin MIC distribution of NRS70 and NRS384 evolved strains.

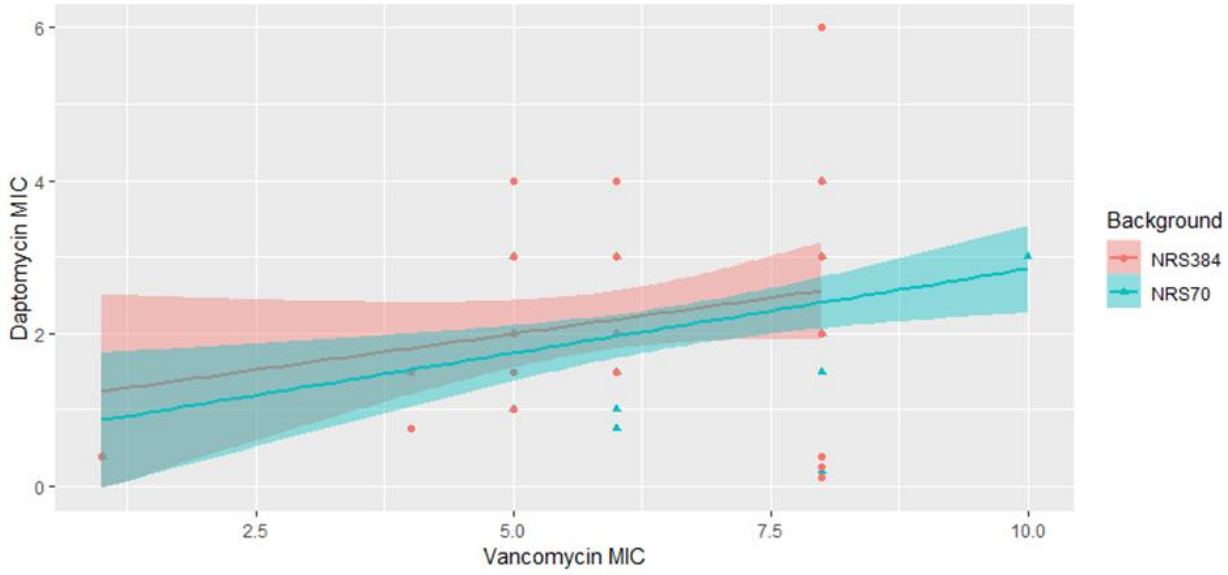

Figure 5. Linear relationship between vancomycin MIC and daptomycin MIC.

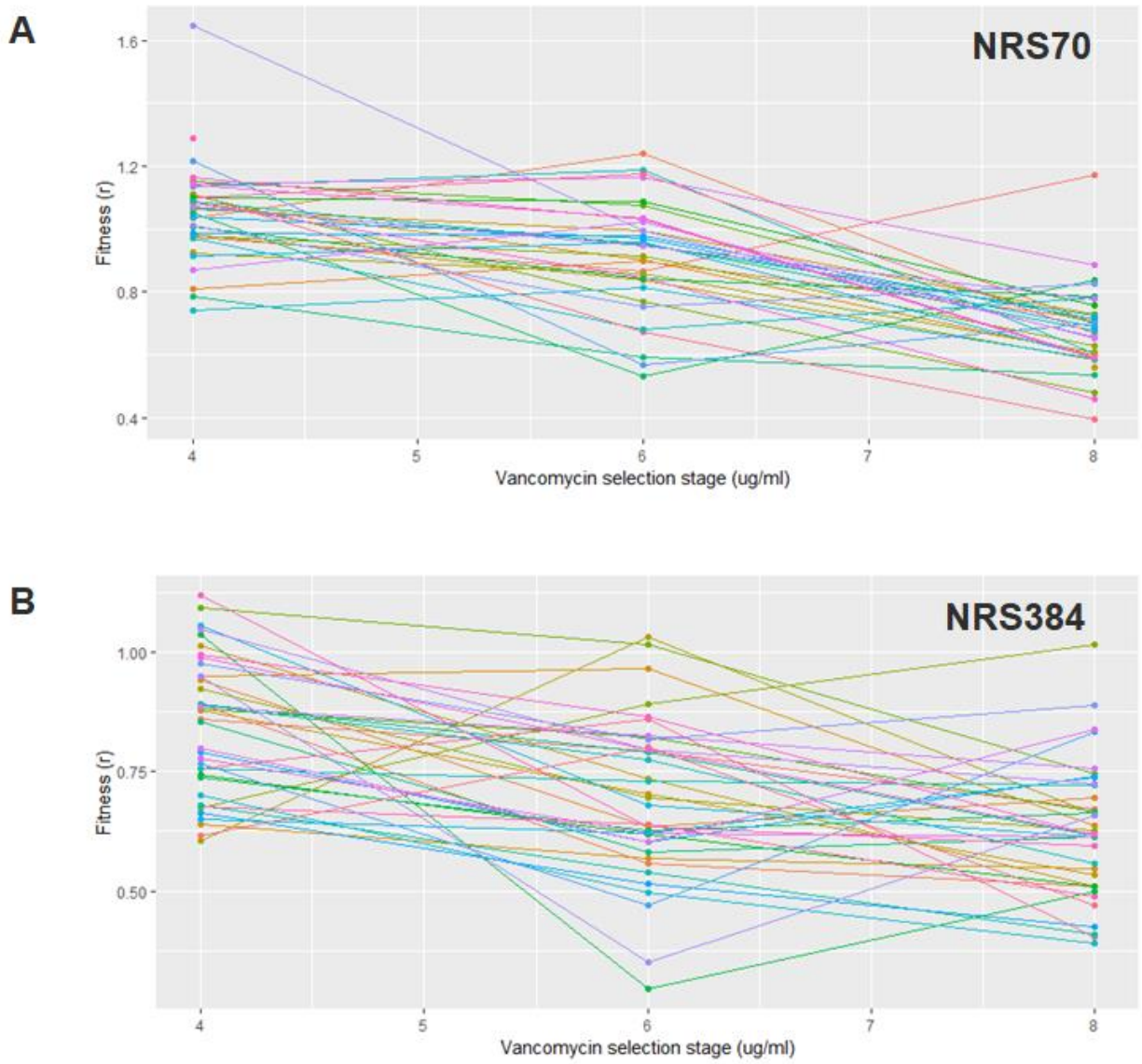

Figure 6. Fitness of NRS70 and NRS384 evolved strains.

(a) Fitness (r) progression of NRS70 evolved strains. NRS70-7 at vancomycin selection stage 6 and 8 and NRS70-14 at vancomycin selection stage 6 were excluded due to lack of growth or too large a variance between replicates. (b) Fitness (r) progression of NRS384 evolved strains.

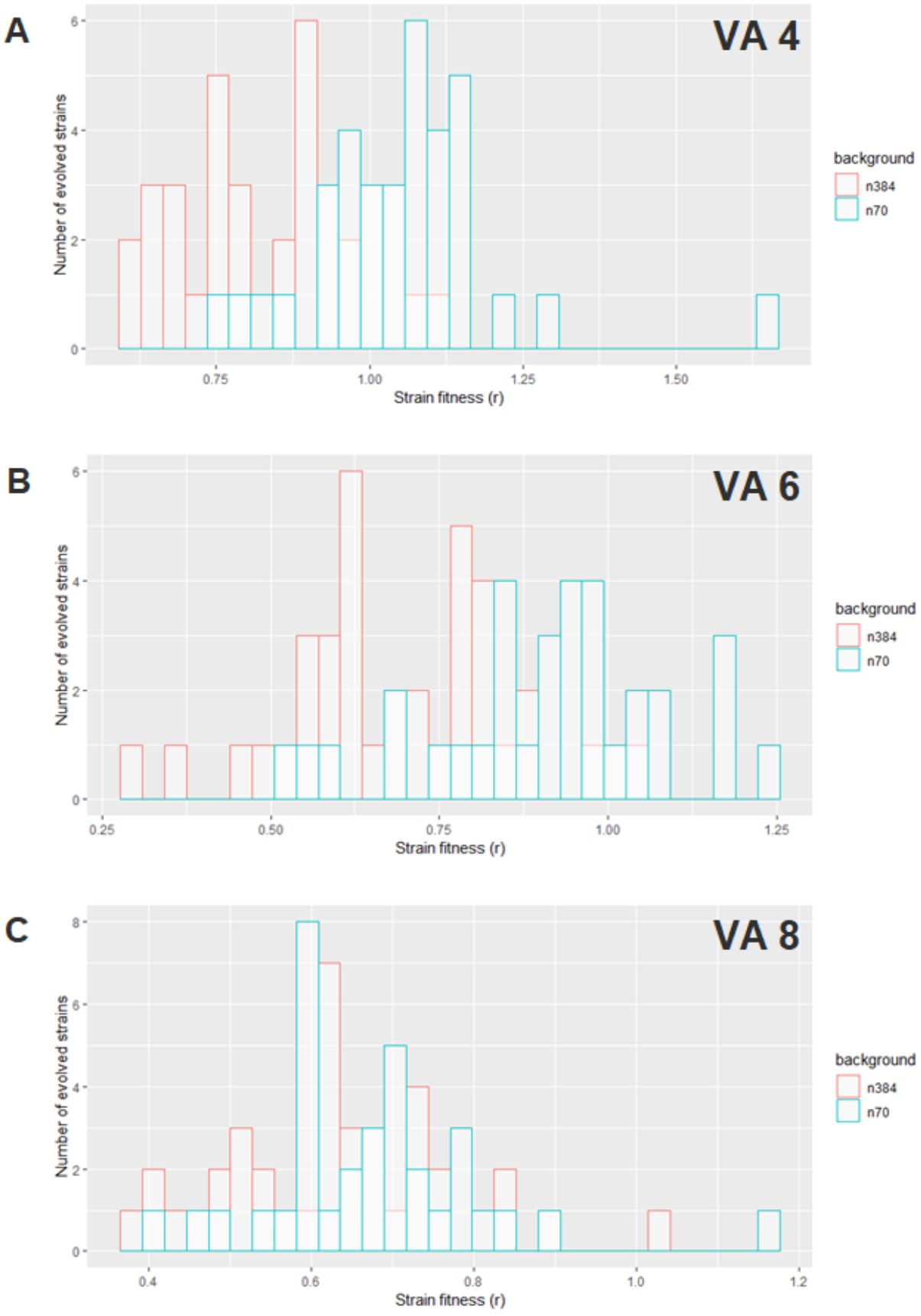
A
VA 4
Number of evolved strains
Strain fitness (r)
background
n384
n70
B
VA 6
Number of evolved strains
Strain fitness (r)
background
n384
n70
C
VA 8
Number of evolved strains
Strain fitness (r)
background
n384
n70

Figure 7. Fitness distributions at each stage of vancomycin selection for NRS70 and NRS384 strains.

NRS70-7 at vancomycin selection stage 6 and 8 and NRS70-14 at vancomycin selection stage 6 were excluded due to lack of growth or too large a variance between replicates. (a) Vancomycin selection stage 4 μg/mL. (b) Vancomycin selection stage 6 μg/mL. (c) Vancomycin selection stage 8 μg/mL.

Gene	Description	Size (bp)
SA_RS04005	*degV*	84
SA_RS14675, SA_RS06275	Hypothetical protein, *rimP*	75
SA_RS07900	*dgkA*	48
SA_RS10700 – SA_RS10405	72 genes	44257
SA_RS00245	*walK*	3
SA_RS05215	*purQ*	3
SA_RS09410	Helix turn helix transcription regulator	281
SA_RS04005	*degV*	156
SA_RS05345, SA_RS05350	Cell wall binding lipoprotein, *pdhA*	1421

Table 1. Deletions greater than 2 bp in NRS70 VISA strains.

Gene	Description	Insertion
SA_RS07890	GTPase era	TTTTCC

Table 2. Insertions greater than 2 bp in NRS70 VISA strains.

Gene	**Description**	**Size (bp)**
MW_RS05925	*stk1*	45
MW_RS10245 - MW_RS10565	66 genes	42602
MW_RS03060	DU443 domain-containing protein	4
MW_RS14510, MW_RS02905, MW_RS02910, MW_RS02915, MW_RS02920	Hypothetical protein, *vraA, vraB, vraC,* hypothetical protein	3842
MW_RS08165, MW_RS08170	*mtaB*, rRNA small subunit methyltransferase	2420
Intergenic MW_RS12225 & MW_RS12230	Intergenic *hutG, sdpC*	135
MW_RS03860	*tagO*	51

Table 3. Deletions greater than 2 bp in NRS123 VISA strains.

Gene	Description	Size (bp)
Intergenic NRS_1842 & NRS_1843	Intergenic IS1181 transposase, tRNA-leu	51
NRS_1469	*srrA*	5
NRS_1951 – NRS_2005	52 genes	43048
NRS_1650	*clpX*	41
NRS_1619	*relA2*	30
NRS_2214	*rpoA*	15
NRS_2117	*rpoE*	126
NRS_2620	*sasF*	12

Table 4. Deletions greater than 2 bp in NRS384 VISA strains.

Gene	Description	Insertion
NRS_261	*tarL*	CCTGAAGATAAGTACTTA
NRS_1202	*rny*	AAAGAA
NRS_1897	*vra*R	TACGGTTACGCATTTTCA
NRS_1939	DUF1700 domain-containing protein	ACTAAT
Intergenic NRS_2521 & NRS_2522	Intergenic *ssaA2_3, mvaA*	TCAG *Two 1 bp insertions also in this region

Table 5. Insertions greater than 2 bp in NRS384 VISA strains.

Gene / Background	NRS70	NRS123	NRS384
walK	80%	85.3%	94.7%
prs	5.7%	5.9%	2.6%
rpoB	45.7%	20.6%	18.4%
rpoC	34.3%	20.6%	10.5%
vraS	17.1%	23.6%	39.5%
yvqF	80%	55.9%	55.3%

Table 6. Prevalence of mutations in six universally mutated genes by genetic background.

The NRS70 *walK* prevalence calculation considers the presence of SNPs as well as inframe deletions. All other prevalence calculations are based solely on SNPs.

	NRS70 and NRS123	NRS123 and NRS384	NRS70 and NRS384
Genes shared	WP_000101976.1 (*walR*) WP_000035320.1 (*pdhA*) WP_001032833.1 (*gad*) WP_000375864.1 (HTH cro/C1-type domain-containing protein) WP_000251253.1 (UPF0374 protein)	WP_001081640.1 (Cyclic-di-AMP phosphodiesterase) WP_000260117.1 (*pdhD*) WP_001071136.1 (*mprF*) WP_000809131.1 (*cspA*) WP_000048060.1 (*rpsU*) WP_000830380.1 (*mgt*)	WP_000631969.1 (*cysS*) WP_001060462.1 (*sdrC*) WP_001273060.1 (*119ag*) WP_000491755.1 (*ykaA*) WP_000120368.1 (*pitA*) WP_000431312.1 (*greA*) WP_000782121.1 (*prsA*) WP_000153535.1 (*vraR*) WP_000159960.1 (*pyrG*) WP_000347896.1 (*dacA*) WP_000004085.1 (*rpsQ*) WP_000008673.1 (*ureC*)

Table 7. Genes with SNPs shared between two genetic backgrounds.

	Synonymous	Nonsynonymous
NRS70	19	149
NRS123	4	114
NRS384	45	166

Table 8. Synonymous SNPs and nonsynonymous SNPs by genetic background.

Background	Gene	Estimate (inclusive)	p (inclusive)	Estimate (final)	p (final)
NRS70	*121 al*	.288	.015	.298	.012
	k	.198	.389	-	-
	prs	.132	.254	-	-
	rpoB	.309	.012	.228	.026
	rpoC	.901	.0005	.989	$< 10^{-4}$
	vraS	.956	$< 10^{-4}$	1.058	$< 10^{-5}$
	yvqF				
NRS384	*121 al*	.629	.004	.633	.002
	k	.255	.515	-	-
	prs	.103	.460	-	-
	rpoB	-.050	.790	-	-
	rpoC	.412	.129	.340	.096
	vraS	.450	.073	.378	.053
	yvqF				

Table 9. Linear regression models of vancomycin MIC.

Chapter V. Trade-offs between virulence and antibiotic resistance in *Staphylococcus aureus* during selection in the model host, *Caenorhabditis elegans*

Michelle Su, McKenna Penley, Michelle Hargita, Levi T. Morran, Timothy D. Read

Abstract

In vitro experimental evolution has historically been used to determine mechanisms of pathogen resistance and the likelihood of emergence during patient therapy. The failure of these experiments to utilize host models means their clinical relevance is limited as host infection can undoubtedly alter the evolutionary trajectories and fitness landscapes of pathogens in addition to other selective pressures such antibiotic exposure. Understanding how pathogens evolve virulence under these conditions will better inform treatment of patients undergoing long-term antibiotic treatment. We experimentally evolved methicillin resistant (USA300 JE2) and methicillin sensitive (USA300 JE2 *mecA*-tn) *Staphylococcus aureus* in a model host, *Caenorhabditis elegans* under differing conditions of antibiotic exposure. Sub-MIC oxacillin exposure selected for extracellular toxicity in *S. aureus*, maintenance of SCC*mec* in USA300 JE2 strains, and higher resistance in USA300 JE2 *mecA*-tn strains exposed to *C. elegans* but not control strains, highlighting the importance of host pressure on pathogen evolution. Genome sequencing revealed that metabolic (*pur*R*, codY, gpmA*), virulence (*agr*, *sae*), and beta-lactam associated (*gra*R*, pbpA, pbpB, gdpP*) loci were under selection within hosts, highly dependent on the antibiotic resistance status of the strain. In USA300 JE2 *mecA*-tn strains (host-free and *C. elegans*), virulence was more phenotypically responsive to sub-MIC oxacillin exposure whereas this was only seen in the absence of host exposure in USA300 JE2 strains, underscoring the strength and importance of host pressure as a selective force that drives differential pathogen evolution.

Introduction

Staphylococcus aureus is a successful nosocomial pathogen, and in recent history, certain lineages have become community associated. It is associated with a variety of infections (osteomyelitis, endocarditis, skin and soft tissue, bacteremia), and in 2017, more than 119,000 bloodstream infections occurred in the United States with a mortality rate of 18% (Kourtis et al., 2019). Paradoxically, *S. aureus* has maintained virulence as a pathogen while becoming increasingly multidrug resistant. However, there are well documented cases of within-host evolution that suggests there are clear trade-offs between virulence, antibiotic resistance, and immune evasion (Ender et al., 2004; Foucault, Courvalin & Grillot-Courvalin, 2009; Laabei et al., 2015; Das et al., 2016; Singh et al., 2017; Li et al., 2017; Choe et al., 2018).

Experimental evolution has the ability to elucidate how different selective pressures impact the evolution of *S. aureus* phenotypes. It has been primarily used to study the mechanisms of antibiotic resistance in vitro. Host infections, however, differ drastically in nutrient availability and add another layer of selection via the immune system. These factors can skew the direction of evolution and limit the evolutionary trajectories possible. In one of few in vivo experiments, Pollitt et al (Pollitt et al., 2014) studied how *S. aureus* strain diversity in a *Galleria mellonella* infection system affected virulence evolution and found that high relatedness favored the maintenance of quorum sensing, loss of *agr* cheats, and higher virulence. Experimental evolution in hosts remains severely underutilized but has the ability to answer key questions regarding the impact of the host environment on the evolution and/or maintenance of virulence and antibiotic resistance.

Caenorhabditis elegans has been evaluated as a model host for *S. aureus* infection (Sifri et al., 2003; Wu et al., 2010). Sifri et al. demonstrated that known virulence regulators (*agr* and *sarA*) were important for

nematode killing as well as extracellular products (V8 protease as alpha-hemolysin). Further, immunocompromised *C. elegans* (*esp-2* and *esp-8* mutants) were more susceptible to *S. aureus* infection, demonstrating the conservation of a p38 MAP kinase pathway in immunity, similar to human neutrophils (McLeish et al., 1998; Lundqvist-Gustafsson et al., 2001). Wu et al were able to demonstrate a correlation between nematocidal activity of different clinical isolates methicillin-resistant *S. aureus* (MRSA) and their site of isolation (invasive disease vs colonization). Lastly, virulence screens in *C. elegans* using *S. aureus* transposon libraries have shown overlapping results to other infection models (Bae et al., 2004; Begun et al., 2005). Together, these studies demonstrate that many aspects of *S. aureus* pathogenesis are maintained in *C. elegans* infection and that the capability exists to differentiate *S. aureus* strains of differential virulence and invasiveness. However, it must be acknowledged that limitations exist within this model system. *C. elegans* growth (15 to 25C) occurs at a much lower temperature than mammalian infection, and while *C. elegans* possesses an immune system, it lacks phagocytotic cells (Cutuli et al., 2019). These factors in turn may affect the evolutionary dynamics of *S. aureus* in *C. elegans* as compared to human infection.

In this study, we aim to characterize how *S. aureus* adapts to host populations by experimentally evolving *S. aureus* in a model system *C. elegans*. A major mechanism for existing antibiotic resistance in *S. aureus* is the carriage of SCC*mec*, which encodes *mecA* and provides resistance against beta-lactams and is often maintained even in the absence of antibiotics. This fixed cost has the potential to change both the fitness landscape and available mutational space during the evolution of traits such as virulence. This evolution may be further constrained by subminimal inhibitory concentrations (sub-MICs) of antibiotics, which may change the fitness costs of antibiotic resistance relative to susceptible strains. Therefore, we explored how two *S. aureus* isogenic strains (*mecA*+ and *mecA*-) differentially evolved virulence to *C. elegans* with or without sub-MIC oxacillin exposure.

Methods

Staphylococcus aureus strains

USA300 JE2 and USA300 JE2 *mecA*-tn from the Nebraska Transposon Mutant Library (NTML) (Fey et al., 2013) were acquired from BEI resources (https://www.beiresources.org/). To ensure true isogenic parents, the *mecA*-tn was transduced into the USA300 JE2 background, selecting for the erythromycin resistance marker on the Mariner transposon, and the insertion site was confirmed by polymerase chain reaction (PCR). Transposon maintenance was confirmed by testing for resistance to erythromycin (5 μg/mL) or by genome sequencing when available. In cases where transposon loss was verified by genome sequencing, strains were tested and shown to maintain sensitivity to oxacillin. As the primary function of the transposon was to knock out *mecA* function, loss of the transposon should not confound subsequent analyses as antibiotic sensitivity was maintained.

Caenorhabditis elegans strains and mortality assays

C. elegans strain N2 was obtained from the Morran laboratory and maintained on nematode growth medium (NGM) according to WormBook protocols (Girard et al., 2007). For mortality assays, nematode population synchronization was performed as in Penley et al (Penley & Morran, 2018). Populations were bleached in 20% household bleach and M9 buffer and plated on *Escherichia* coli strain OP50 until L4 larval stage (48 hours at 20C). Worms were subsequently washed off, counted, and approximately 200 worms were plated on control OP50 NGM plates and *S. aureus* lawns on BHI agar. *S. aureus* plates were prepared 24 hours prior by adding 200 μL of an overnight culture to BHI agar and growing at 24C. At 24 hours and 48 hours, plates were scored by counting live worms. Worm counts on OP50 plates were used to normalize mortality calculations and to account for plating efficiency.

Experimental evolution

Selection for increased virulence of *S. aureus* (USA300 JE2 or USA300 JE2 *mecA-tn*) in *C. elegans* was performed by passaging *S. aureus* from dead hosts 24 hours post-infection (hpi). Half of the *S. aureus* populations were also additionally subjected to antibiotic pressure during the passages before and after *C. elegans* selection with sub-MIC oxacillin exposure (.03125 μg/mL). This concentration was tested against both parent strains and did not substantially inhibit growth of the methicillin-sensitive USA300 JE2 *mecA*-tn strain. Twelve rounds of passage through *C. elegans* N2 were completed. Control *S. aureus* populations, used to account for adaptation to culturing media or genetic drift, were maintained by passaging *S. aureus* in the same manner without exposure to N2.

For each passage of experimental evolution, *S. aureus* was grown in Brain Heart Infusion (BHI) broth with 4 μg/mL colistin (to prevent OP50 contamination from the worm intestine) ± .03125 μg/mL oxacillin and incubated at 37C overnight, shaking at 250 rpm. The next day, 200 μL of the overnight S. aureus culture was plated onto BHI agar and incubated at 24C overnight before worm plating. Approximately 1500 *C. elegans* were plated and allowed to consume *S. aureus* strains for 24 hours before 30 dead worms were picked from each plate. Worms were identified as dead by a lack of response to tapping by a platinum wire (Amrit et al 2014). Bacterial extraction methods were adapted from Vega et al (Vega & Gore, 2017). In brief, picked dead worms were washed once with M9 buffer supplemented with 0.1% Triton X-100, once with M9 buffer, subjected to a 1:1000 bleach M9 buffer solution for 15 minutes, washed twice with M9 buffer, and finally ground with a motorized pestle to extract the *S. aureus* from the worm intestine. Control *S aureus* populations were obtained from worm-free bacterial lawns with a loop. Isolated *S. aureus* were grown as previously before freezing and storage at -80C. Twenty-five percent of each frozen culture was used as inoculum for the next

round of experimental evolution. Preliminary experiments determined colony-forming units per milliliter (CFU/mL) dropped by approximately 50% after freezing. To maintain population diversity, the inoculum amount (25%), adjusted for freezing, was chosen to be within the dilution ratio presented in Wahl et al that minimizes the chance that rare beneficial mutations are lost (Wahl, Gerrish & Saika-Voivod, 2002). After 12 rounds of evolution, single colonies were isolated from each population for further analyses.

Antimicrobial susceptibility assays

Antimicrobial susceptibility testing for oxacillin was done according to Clinical & Laboratory Standards Institute (CLSI) standard protocols for broth microdilution (BMD) (Clsi, 2013). In brief, cultures were streaked from a frozen stock and restreaked the next day. From the second day culture, colonies were resuspended in normal saline to a 0.5 McFarland standard. Then, these cultures were diluted 1:20 in normal saline before 10 μL was added to 90 μL of CAMHB + 2% NaCl with the appropriate concentration of oxacillin. Plates were incubated at 35C without shaking and read at 24 hours.

Hemolysis phenotype

S. aureus strains were streaked onto Trypticase Soy Agar (TSA) II with 5% sheep's blood plates and incubated overnight at 37C. Strains were assessed as either hemolysis positive or negative dependent on whether there was a complete clearing around the colonies. We did not cross-streak with RN4220, which produces beta-hemolysin, and allows for detection of delta-hemolysin. Therefore, our assay mainly detects alpha-toxin and phenol-soluble modulins (PSMs). Statistical significance between populations was assessed using Fisher's exact test.

Whole-genome shotgun sequencing and variant annotation

Bacterial genome DNA was extracted from pure cultures (Wizard Genomic DNA Purification Kit, Promega). Genome sequencing of these DNA samples was performed at the M-Gen center using a paired-end library preparation method based on Illumina Nextera and sequenced on the NextSeq 550.

The raw data was quality trimmed and adapters removed using illumina-cleanup (https://github.com/rpetit3/illumina-cleanup) (Andrews, 2010; Bushnell, 2014; Song, Florea & Langmead, 2014; Adler, 2015; Di Tommaso et al., 2017) or with bactopia (Petit & Read, 2020). Mutations were called based on comparison to the reference USA300 JE2 genome (BioProject: PRJNA224116, BioSample: SAMN06677988, Assembly: GCF_002085525.1) using breseq (Deatherage & Barrick, 2014) or snippy (Seemann, 2015) from bactopia (Petit & Read, 2020). Mutations called by both snippy and breseq were considered true positives. The presence of large scale deletions and other genetic changes were confirmed by analysis of de novo assembled genomes performed by the SKESA assembler software (Souvorov, Agarwala & Lipman, 2018).

Results

Hemolysis and oxacillin resistance phenotypes of evolved *S. aureus*

We experimentally evolved *S. aureus* strains in *C. elegans* while varying the prior antibiotic resistance (methicillin resistance) and antibiotic exposure (sub-MIC oxacillin) of the strains during the course of the experiment. Eight populations (4 in a host (*C. elegans*), 4 host-free controls) with 6 replicates each were evolved for 12 weeks, for a total of 48 evolved *S. aureus* strains (Figure 1).

S. aureus strains were phenotypically characterized to determine changes in hemolytic activity and oxacillin resistance (Table 1). Populations of both genetic backgrounds (USA300 JE2 and USA300

JE2 *mecA*-tn) were found to lose the ability to hemolyse sheep's blood under antibiotic free conditions, except USA300 JE2 *mecA*-tn strains also exposed to *C. elegans*. Hemolysis correlates with the ability of the strain to secrete extracellular toxins. This result is surprising as alpha-toxin and PSMs are regarded as important virulence factors in *S. aureus*, and selection for virulence in *C. elegans* did not overwhelmingly favor maintenance. On the contrary, sub-MIC oxacillin exposure favored maintenance of hemolytic ability in all populations regardless of *C. elegans* virulence selection, suggesting that it may select for retention of extracellular toxicity ($p < .05$ for U vs UO population and MN vs MON population)

To assess the strength of the selection pressure imposed by sub-MIC oxacillin exposure in the presence or absence of a host, we examined populations for changes in antibiotic resistance (measured by MIC). In control populations (those that did not evolve in C. *elegans*) of both genetic backgrounds, there were essentially no changes in the oxacillin MICs of evolved *S. aureus* strains. In USA300 JE2 strains evolved in *C. elegans*, MICs were maintained at 16 μg/mL oxacillin except in the case of two USA300 JE2 strains not exposed to sub-MIC oxacillin. In USA300 JE2 *mecA*-tn strains, MICs did not change significantly in strains not exposed to sub-MIC oxacillin but did so dramatically in strains exposed to sub-MIC oxacillin (4-8 fold the parental MIC). These results suggested sub-MIC exposure during *C. elegans* exposure potentiated the increase in antibiotic resistance in methicillin sensitive strains and maintained high-level resistance in MRSA strains.

Genomic changes associated with evolution of *S. aureus*

Half of the evolved *S. aureus* populations were genome sequenced and analyzed for mutations (Table 2). Of note, parent (non-evolved) strains were sequenced and were almost identical to the reference genome used with the exception of 3 mutations that were not found in any evolved strains. Three

mutations in *bioD1, cls* and *ccdC* were present in all USA300 JE2 *mecA*-tn strains, and a likely explanation is that these mutations occurred in the first passage from the frozen culture before experimental evolution was performed. Therefore, we disregarded these mutations when considering mutations that may have contributed to the phenotypic differences between strains.

Within the short timeframe of the experiment, the majority of strains accumulated only 2 to 8 mutations (excluding *bioD1, cls* and *ccdC* present in all USA300 JE2 mecA-tn described above). As expected, all strains sequenced that were hemolysin negative had mutations in the *agr* operon, the likely explanation of the loss of hemolysis. Seven out of nine strains had null or non-synonymous mutations, and two had intergenic mutations. In these two strains, the other mutations are genes unrelated to the hemolysins and PSMs and are likely not responsible for the phenotype. Interestingly, genes associated with beta-lactam resistance (*pbpA, pbpB, gra*R*, gdpP*) were more likely to be mutated in USA300 JE2 strains (5/6 sequenced strains) than USA300 JE2 *mecA*-tn strains (2/6 sequenced strains) exposed to oxacillin, perhaps indicating that the fitness costs of SCC*mec* can be offset with further mutation in other resistance determinants. There is further evidence that SCC*mec* imposed a significant burden in our genomic data: 4 strains (two USA300 JE2 and two USA300 JE2 *mecA*-tn) experienced large scale deletions that included the entire SCC*mec* cassette. Interestingly, two deletions only encompassed SCC*mec* (23.7 kb), and two deletions were much larger (54.7 kb) and included the Arginine Catabolic Mobile Element (ACME) (Diep et al., 2006). In the USA300 JE2 strains, this deletion only occurred in strains not exposed to oxacillin, while in the USA300 JE2 *mecA*-tn strains, it only occurred in strains exposed to oxacillin. *arcA*, part of the arginine-deiminase system encoded by ACME, also has a deletion in one USA300 JE2 strain, suggesting there is selection acting on the ACME locus and its deletion with SCC*mec* may not be accidental. The deletion of ACME is interesting as there have been

studies that indicate that it enhances the fitness of USA300 (Diep et al., 2008; Joshi et al., 2011; Thurlow et al., 2013).

Genes in metabolic operons were often mutated (e.g *brnQ1, pur*R*, codY, gpmA). BrnQ1* mutations are most common in control populations and thus possibly the result of adaptation to culture media or random mutation. The others, however, are found exclusively within the strains exposed to *C. elegans* and have well-known roles in virulence (Pohl et al., 2009; Roux et al., 2014; Goncheva et al., 2019; Radin et al., 2019; Sause et al., 2019; Goncheva, Flannagan & Heinrichs, 2020). *Pur*R mutations were found exclusively in USA300 JE2 oxacillin exposed strains, and *gpmA* were found exclusively in USA300 JE2 *mecA*-tn non-oxacillin exposed strains. *codY* mutations were enriched in USA300 JE2 *mecA*-tn strains. Mutations in another well-known virulence regulator *sae*R*S* (Rogasch et al., 2006; Liu, Yeo & Bae, 2016; Guo et al., 2017; Mashruwala et al., 2017) were also found exclusively in USA300 JE2 strains.

Changes in virulence of evolved *S. aureus* strains

To assess the changes in virulence of the evolved strains, survival assays of *C. elegans* were performed and assessed after 48 hours (Figure 2). The ancestral survival virulence varied from experiment to experiment, contributing significant noise to the calculation of relative virulence of evolved strains. We assumed that the variation was on the side of *C. elegans* and not *S. aureus*, and thus relative strain virulence measurements were assessed using the ancestral survival measured at the same time. Because of this variability and the small sample size, we will focus on qualitative rather than quantitative assessments of the data. Most treatments did not consistently evolve more or less virulence towards *C. elegans*, potentially indicating bottleneck pressures may play a large role in the stochasticity of *S. aureus* evolution. There were indications that sub-MIC oxacillin exerted differential selection

depending on background. In USA300 JE2 *mecA*-tn strains, oxacillin exposed populations tended to become more virulent whereas non-exposed populations were split between strains that became more virulent or less virulent. In contrast, USA300 JE2 oxacillin free control populations evolved further virulence, and oxacillin exposed control populations were split between more virulence and less virulence. This was not true in USA300 JE2 strains exposed to *C. elegans* where the populations were split between more and less virulence, irrespective of oxacillin exposure. Surprisingly, maintenance of hemolysis was not necessary for increased virulence and did not preclude the evolution of increased virulence. This may indicate that other virulence factors may compensate for the loss of hemolysis or may be more important to the pathogenesis of the USA300 background in *C. elegans*.

Discussion

The evolutionary trajectories of pathogen evolution within hosts is a consequence of many factors including but not limited to host immune pressure and antibiotic exposure. Antibiotic resistance is the natural response to antibiotic selection, but maintenance in antibiotic-free environments represents a fitness cost. This cost can subsequently constrain further evolution which requires cellular resources, such as virulence. In this study, we evolved *mecA*+ and *mecA*- *S. aureus* under a variety of conditions (host-free, *C. elegans* exposure and antibiotic-free, sub-MIC oxacillin exposure) and demonstrated the complex interplay of antibiotic exposure and antibiotic resistance determinants in influencing evolution of host killing.

Virulence of *S. aureus* in *C. elegans* is likely determined by a combination of many virulence determinants. In our experiment, *agr* function and hemolysis potential (as a proxy for the activity of membrane-active bacterial toxins) was not necessary for maintenance or evolution of further virulence (Figure 2). While there is evidence that *agr* and hemolysis are important for virulence in *C. elegans* (Sifri et al., 2003;

Bae et al., 2004), the opposite has also been supported by previous research. Wu et al found that *S. aureus* strains of a similar genetic background that possessed alpha-toxin and V8 protease were non-virulent (Wu et al., 2010), and Irazoqui et al found that hemolysin negative isogenic NCTC8325 were as capable of nematode killing as wildtype NCTC8325. Interestingly, Bae et al found alpha-toxin to be necessary for nematode killing, but could not consistently demonstrate that the *agr* locus was also essential despite its primary function of increasing production of extracellular toxins. A caveat is that it is possible that our mortality assay has not truly captured differences in virulence but merely changes to the dynamics. A longer survival assay would be needed to interrogate this possibility. However, experimental evolution may also select for virulence that is not obvious in single generation assays and may be difficult to capture using traditional virulence assays. Cost-benefit ratios of mutations may be different on a short versus long time scale. Assays that focus on *C. elegans* populations over time by analyzing traits such as fecundity will be necessary to determine if this is a potential mode of virulence evolution.

Antibiotic exposure was found to select for *agr* positive strains (Table 1). Because *agr* function and virulence are decoupled in our strains, this potential upregulation of *agr* by sub-MIC oxacillin is not generally associated with virulence in our strains (Figure 2): In USA300 JE2 *mecA*-tn, this may have contributed to the higher virulence of oxacillin exposed strains. However, in USA300 JE2 strains, strains exposed to *C. elegans* demonstrate no virulence differential based on oxacillin exposure, but control strains exposed to oxacillin are less virulent than their counterparts not exposed. The opposing effect of oxacillin on these two genetic backgrounds may be in part explained by the inhibitory effect of *mecA* in *agr* signaling (Rudkin et al., 2012). Thus, the complex virulence regulation of *S. aureus* (governed by *agr, sarA, sae*RS and others) allows for widespread epistasis and a role for genetic background in modulating the importance of virulence factors.

The sub-MIC nature of the oxacillin exposure more accurately reflects the real-world selection that exists outside the context of infection and which may contribute to evolution of antibiotic resistance (Gullberg et al., 2011; Andersson & Hughes, 2014; Wistrand-Yuen et al., 2018). In our control populations, there are essentially no changes in oxacillin MIC, suggesting that without the impact of other forces, there is little selective pressure to evolve resistance in our experimental setup. In two USA300 JE2 populations selected for virulence in *C. elegans*, resistance has been lost. The mechanism of resistance loss has been confirmed to be a SCC*mec* deletion in one strain by genome sequence. Both of these strains were able to evolve further virulence to *C. elegans*, highlighting the trade-offs that exist between virulence and antibiotic resistance. In five out of six USA300 JE2 *mecA*-tn strains selected for virulence and exposed to sub-MIC oxacillin, oxacillin MICs were four to eight-fold the parental MIC. As this change in MIC did not occur in control strains, host selection must have skewed evolution towards further antibiotic resistance. The resistance achieved, however, is still classified as sensitive and thus traditional trade-offs with virulence may not apply at this lower level.

In sequencing half of the evolved strains, we were able to validate the selection pressures on metabolic regulators (*codY, purR, gpmA*) and virulence regulators (*agr*, *sae*RS) in host infection. Mutations in regulatory proteins often have pleiotropic effects, and we expect the phenotypic effects of these mutations will be difficult to predict without methods such as RNA-seq. We also found genomic evidence of adaptation to the sub-MIC exposure of oxacillin (mutations in *graR, pbpA, pbpB, gdpP*), even in methicillin resistant populations. Other loci under selection more broadly (seen in *C. elegans* and control populations) were SCC*mec* and ACME. SCC*mec* was confirmed deleted in four strains by genome sequencing and likely another USA300 JE2 strain that lost methicillin resistance. Two of these strains also deleted ACME with SCC*mec*. While type IV SCC*mec* are smaller than the traditional type

II cassettes found in healthcare-associated MRSA (HA-MRSA) (Collins et al., 2010), fitness costs are still incurred by its maintenance. Its loss was only found in USA300 JE2 strains not also exposed to oxacillin, indicating that its fitness costs can be offset even by low concentrations of antibiotic. ACME deletion and *arcA* mutation suggest that arginine deiminase, which is beneficial to *S. aureus* surviving on acidic skin, may not be crucial or may be costly to survival in *C. elegans*. However, ACME also carries *speG*, which is an important defense against polyamines found in human hosts (Joshi et al., 2011), but this may not be relevant in *C. elegans* infection as it has not been found to be important in a rat pneumonia model or mouse skin infection model (Montgomery, Boyle-Vavra & Daum, 2009). Further, phylogenetic analyses have suggested loss of ACME in several instances, indicating that it may not be beneficial in all circumstances (Uhlemann et al., 2014; Planet et al., 2015; Glaser et al., 2016; Planet, 2017). The common inactivation of *agr* in our *S. aureus* evolved strains reflects clinical experience, where many infections result in the isolation of *agr* mutants (Traber et al., 2008; Altman et al., 2018). This phenomenon may be explained by the presence of *agr* "cheats" that are non-responsive to *agr* signaling but benefit from the public goods production of *agr*+ strains (Pollitt et al., 2014). Because of our isolation of a single colony for analysis, we do not know whether our populations have mixtures of *agr*+ and *agr*- strains.

Of note, while not commonly mutated, mutations in *recA*, *capA* (synonymous mutation), *rsp* (synonymous mutation), *alsT,* and *gltB* were found in *S. aureus* strains evolved in *C. elegans*. In a virulence screen using *C. elegans*, Bae et al found that other DNA helicases, *recG* and *recQ*, were necessary for nematode killing as well as capsular polysaccharide genes, *cap5F, capJ* and *capM (Bae et al., 2004). alsT* and *gltB* are both mutated in a USA300 JE2 *mecA*-tn strain. *alsT* imports glutamine, and *gltB*, together with *gltD*, helps convert excess glutamine to glutamate. This pathway is thought to help regulate bacterial metabolism and potassium intake, but *alsT* may impact c-di-AMP levels by indirectly

inhibiting the c-di-AMP cyclase *DacA* (Zeden et al.). c-di-AMP is required for growth in rich media conditions and osmotic regulation, but it has been shown to contribute to beta-lactam resistance as well (Dengler et al., 2013).

In summary, antibiotic resistance acts as a powerful modulating force on the evolution of bacterial evolution, especially in the context of host infection. There were loci that were under differential selection in the USA300 JE2 and USA300 JE2 *mecA*-tn strains, a sign that antibiotic loci can determine the mutational space for further mutations. In addition, USA300 JE2 *mecA*-tn strains were more phenotypically labile in response to sub-MIC oxacillin exposure, regardless of host exposure. The host exposed USA300 JE2 populations were more phenotypically similar than host-free populations, suggesting host selection is an important and stronger selective force than sub-MIC antibiotic exposure.

Figures and tables

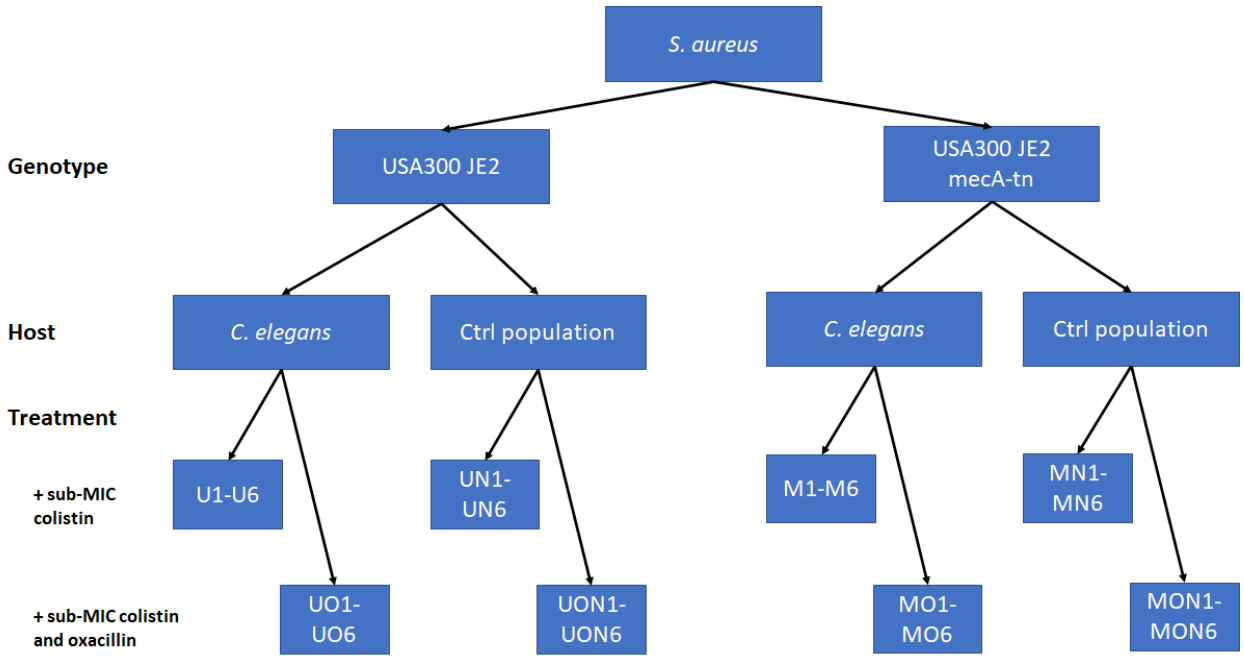

Figure 1. Experimental design and naming scheme of *S. aureus* populations.

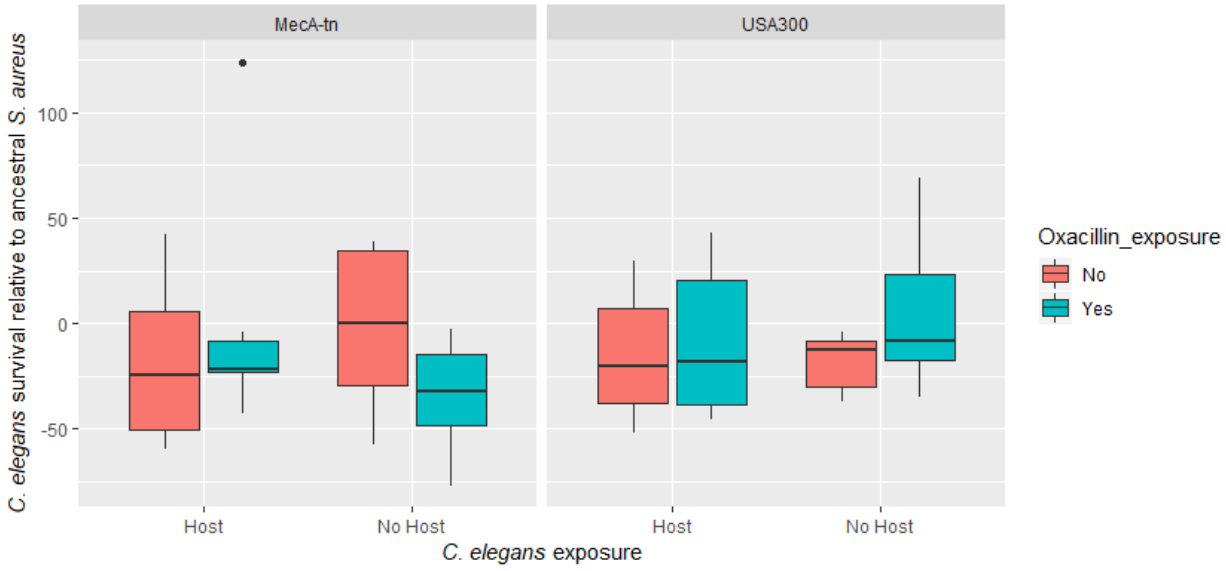

Figure 2. Change in *C. elegans* survival to *S. aureus* infection of evolved strains compared to ancestral *S. aureus.*

	MRSA + worms	MRSA + worms + oxacillin	MSSA + worms	MSSA + worms + oxacillin	MRSA control	MRSA control + oxacillin	MSSA control	MSSA control + oxacillin
Hemolysis positive	1 / 6	6 / 6	3 / 6	5 / 6	0 / 6	4 / 6	0 / 6	6 / 6
Oxacillin resistance (μg/mL)	16(4/6) .25(2/6)	16(6/6)	.25(5/6) 1(1/6)	.5(1/6) 1(3/6) 2(2/6)	16(6/6) *	16(6/6)	.25(6/6)	.25(5/6) .5 (1/6)

Table 1. Hemolysis and oxacillin resistance phenotypes of *S. aureus* strains after passage through *C. elegans*.*

***One of the sequenced strains has lost SCC*mec* and should be re-evaluated.

	Coding gene mutations							Intergenic mutations	
U1	***agrA***							*apt,recJ*	*farE,farR*
U2	***agrA***	***saeS***	pseudo gene					*aur,isaB*	acetyl-CoA synthet., antibiotic monooxyg
U3	**Δ2**							*argR,ispA*	
UO1	***pbpA***	***purR***						*caiA,yut*	
UO2	***codY***	***pbpB***	***saeR***						
UO3	***arcA*(Δ)**	***gdpP***	***graR***	***purR***	hypothet. protein				
M1	***hutG***	***gpmA***						acetyl-CoA synthet., antibiotic monooxyg	
M2	***brnQ1***	***codY***	*capA*	***murP***	***recA***				
M3	***brnQ1***	***alsT***	***gpmA***	***gltB***	**amidase**	***treB***	hypothet. protein	hypothet. protein, *pdxB*	

MO1	***gdpP***	***codY***	hypothet. protein						
MO2	*rsp*	***codY***							
MO3	***brnQ1***	Δ1	***pbpA***	**polysacc. deacetyl.**	**hypothet. protein**				
U1N	***argR***	***alr***							
U2N	***agrC***	***brnQ1***							
U3N	***agrA***	***brnQ1(2)***	Δ1						
UO1N	***pbpB***	***mreD***	nuclease					*agrD,agrC*	
UO2N	***argR***	***alr***							
UO3N	***agrA(2)***	***graR***	**rpoB**	ATPase	pseudo gene	hypothet. protein		*mvaS,ogt*	
M1N	***brnQ1***	***agrC***							
M2N	***brnQ1***	***ebh***						*agrD,agrC*	
M3N	***brnQ1***	***agrA***							
MO1N	***brnQ1***	permease							
MO2N	***brnQ1***	Δ2							

MO3N	***brnQ1***							*relP,ppnk*	

Table 2. List of genomic changes in sequenced evolved *S. aureus* strains.

Genes are color-coded and grouped by operons (e.g *agr* genes). Bold indicates a nonsynonymous mutation, (Δ) indicates a deletion, and (#) indicates the number of SNPs present in the gene. *bioD1, cls,* and *ccdC* mutations in all USA300 JE2 *mecA*-tn strains have been omitted.

Chapter VI. Discussion

In my dissertation, I have focused on the evolution and genetic determinants of *Staphylococcus aureus* virulence and antibiotic resistance phenotypes. Experimental evolution in combination with genome sequencing has allowed me to explore the consequences of epistasis on different *S. aureus* phenotypes.

Chapter II is a survey of where the field currently stands in regards to prediction of antimicrobial resistance phenotypes. As antibiotic resistance is a highly convergent phenotype within species, genetic determinants ascertained by sequencing of clinical isolates or in vitro experiments are highly predictive, and predictions using either presence/absence of genetic determinants, classification algorithms, or machine learning approaches are by and large concordant with phenotypic antimicrobial susceptibility testing results. Inaccuracies have often been traced to errors in variation in phenotypic assays, but cryptic epistasis likely plays a role in determining how a resistance determinant yields resistance at a phenotypic level. However, even in an age where genomic predictions are used as the primary technology to determine antimicrobial susceptibility, active surveillance of new isolates will be necessary to ensure that new genetic mechanisms have not arisen.

In chapter III, I investigated the genetics behind variable delta-toxin production in a diverse set of *S. aureus*. Delta-toxin is a major contributing factor to the disease severity of atopic dermatitis. I determined that the range of delta-toxin production was large, with one strain producing 97,235 mAU^2 (determined by HPLC). The median was 8,295 mAU^2, and most strains produced less than 30,000 mAU^2. Phylogenetic clades were shown to strongly correlate with the level of delta-toxin production, but individual strain variation also played a large role in the variance seen within clades. Interestingly, delta-toxin production of two clades matched the epidemiological data on the causative strains in atopic dermatitis cases. CC30, which is underrepresented in atopic dermatitis relative to its prevalence

in skin colonization, has much lower toxin production than CC45, a clade that is often found in atopic dermatitis (Yeung et al., 2011). As the strains in this study were from a wide range of clinical infections, an investigation into the delta-toxin production of these clonal complexes in atopic dermatitis needs to be completed. Differences of delta-toxin production were also found between Agr groups, but as Agr type is confounded by CC, it is unclear whether variation in the Agr operon plays a role in modulating production or if it represents the closer evolutionary relationship of strains within Agr groups. Determining the delta-toxin production in strains that have undergone Agr switching to another type or performing Agr switching in the laboratory would be a method to disentangle the effects of Agr type with CC. Methicillin resistance status (MRSA) was found to be associated with a decreased capacity to achieve high delta-toxin production. This may be due to methicillin resistance induced cell wall changes interfering with *agr* quorum sensing. To address this, strains can be stratified by SCC*mec* type and re-examined to see if strains with larger SCC*mec* types (which produce more PBP2a and induce larger cell wall changes) have less delta-toxicity than smaller SCC*mec* types.

Little is known about the regulation of delta-toxin aside from its translation is linked with the effector molecule of the Agr quorum sensing system, RNAIII. Delta-toxin is translated with a one hour delay after RNAIII transcription by an unknown mechanism (Balaban & Novick, 1995). The secondary structure of RNAIII acts as a switch to determine whether translation occurs. The cellular input that RNAIII responds to is likely controlled by another genomic region that is not the Agr operon. To determine which genomic regions play a role in delta-toxin production, I performed a GWAS. Unfortunately, due to the statistical challenge of doing these analyses with quantitative phenotypes while appropriately accounting for population structure, a binary phenotype was chosen over a continuous phenotype. The cutoff for the binary phenotype was chosen to capture the highest toxin strains as well as strains at the beginning of the right tail of the distribution. Changing the binary cutoff

to a lower value would increase the number of positive strains, potentially giving the analysis more power to detect genetic determinants at the cost of a higher number of false positives due to linkage disequilibrium. A small subset of genes were chosen for laboratory validation using transposon mutants from the Nebraska Transposon Mutant Library (Fey et al., 2013), and most were found to have an effect on delta-toxin production. This screening excluded essential genes, of which there were several in our GWAS results, but was a quick alternative to allelic replacement, which remains technically difficult in *S. aureus*.

Nevertheless, our GWAS returned a novel finding: the gene *carA* is essential for delta-toxin production. *carA* is the active subunit of carbamoyl-phosphate synthase which is involved in pyrimidine metabolism and L-arginine biosynthesis. The link between *carA* and delta toxin production is unclear but could be investigated by screening for suppressor mutations that restore delta-toxin production. The cost of quantifying delta-toxin using HPLC would be prohibitively expensive, but rescued mutants could potentially be screened using CAMP assays wherein strains are cross-streaked with RN4220, a beta-toxin producing strain, on a blood agar plate. Beta-toxin inhibits alpha-toxin and synergizes with delta-toxin. Phenol soluble modulins have been shown to also synergize with beta-toxin (Cheung, Duong & Otto, 2012), so this should only be done in a background with PSMs knocked out. In addition, the secondary structure of RNAIII should be interrogated using DMS footprinting to assess whether the ribosome binding site is unavailable in a *carA* knockout strain, thus leading to a lack of translation of delta-toxin. Of note, complementation with two alleles with identical protein sequences produced disparate results. Many mechanisms exist by which synonymous mutations are not silent. Initial experiments to narrow down the mechanism would involve checking whether transcription and/or translation occurs efficiently with reporter fusions.

Finally, a much larger sample size would also greatly improve the power of GWAS approaches to detect associations. New genes could then be incorporated into the XGBoost models to increase their predictive power. The current XGBoost models have validated the observations I saw when mapping delta-toxin production on the phylogeny. Sequence types explain a large percentage of the variation in delta-toxin production, indicating this phenotype is strongly influenced by epistasis, but very high-level toxin production is dominated by individual adaptation, potentially due to recent selection.

In chapter IV, I sought to determine the role genetic background played in a vancomycin-intermediate resistance in *S. aureus* by examining the parallel evolution of the phenotype in CC1, CC5, and CC8 backgrounds. I confirmed the role of previously implicated genes (*walK, prs, rpoB, rpoC, vraS, yvqF*), which were generally cell-wall associated or global transcriptional regulators, in all three backgrounds. Some genes were mutated more often than others. It is unclear if mutations in those genes provide the most initial resistance to vancomycin and thus have higher fitness during antibiotic exposure or whether they are more well-tolerated and provide enough tolerance to mutate higher resistance at later stages. Sequencing the earlier stages of vancomycin selection would determine which genes are the initial drivers of resistance. Study of single point mutations in these genes to determine fitness and the effects on tolerance and/or resistance will elucidate whether fitness costs or resistance level is most important in early evolution. Most strains had SNPs in two or three of these genes, indicating while multiple combinations of these genes can be mutated to achieve vancomycin-intermediate resistance, the phenotype is largely driven by parallel evolution. However, epistasis, while not the primary driving force, was shown to play an important role. Mutation patterns (prevalence, co-occurrence, and SNPs) within those six universally mutated genes differed by genetic background. It would be of interest to study how the same mutation in different genetic backgrounds contributed to either fitness or resistance and whether the prevalence of each could be attributed to either. If differences between

genetic backgrounds were observed, it could stem from resulting differences in cell wall thickness, which could be measured and compared using electron microscopy. Further, the pattern of mutation in accessory genes (those mutated in only 2 backgrounds), some of which have also previously been implicated in vancomycin resistance (*walR, vraR, mprF*), indicated there were differential selective pressures on each of the genetic backgrounds.

Accessory genes and genes only found in one background are not consistently mutated within a background, suggesting that these genes may not be under strong selection and may be largely targeted for compensatory mutations. The mutational space for compensatory mutations is much larger and may explain the stochasticity observed. Recapitulation of a defined set of VISA mutations with mutations in these genes would help elucidate their function. Fitness (measured by r - maximum growth rate) was generally found to decrease as vancomycin tolerance and/or resistance increased. However, I measured fitness only in rich media, which is not indicative of within-host environments. Fitness could further be examined using growth curve analysis in whole blood or in vivo by measuring bacterial loads pre- and post- infection.

In chapter V, I studied *S. aureus* virulence evolution within a host model, *Caenorhabditis elegans*. I found there to be significant interaction between oxacillin exposure and hemolysis potential, which genome sequencing indicated was likely *agr* mediated. Sub-MIC oxacillin exposure favored maintenance of hemolysis in our strains. It has been previously shown that sub-MIC oxacillin exposure can upregulate virulence expression of the *agr* operon, alpha-toxin, and Panton-Valentine Leukocidin (Ohlsen et al., 1998; Dumitrescu et al., 2007, 2011; Viedma et al., 2018). However, it can also upregulate *mecA* expression, which interferes with *agr* function (Rudkin et al., 2014). Therefore, oxacillin exposure likely has a pleiotropic effect on virulence expression in our MRSA background but upregulates virulence

in an MSSA background. This will need to be confirmed by transcriptomics and supernatant toxin quantification. I also found interaction between oxacillin exposure and host selection. Control *S. aureus* populations, not exposed to the worms, experienced little change in their oxacillin resistance as measured by BMD. However, five out of six MSSA strains exposed to oxacillin and worms achieved resistance that was 4-8 times the parental resistance, indicating that antibiotic exposure and host selection are synergistic to the development of antibiotic resistance. It is unclear if beta-lactam resistance makes *S. aureus* more resistant to *C. elegans* host defense, thereby selecting for mutations that give higher resistance. This could be examined by electron microscopy to determine the percentage of lysed *S. aureus* in the worm lumen. Two of six MRSA strains not exposed to oxacillin and exposed to worms saw a return to the parental MIC, confirming virulence selection is at odds with antibiotic resistance in the context of a host infection.

Genome sequencing of one colony from half of the evolved *S. aureus* populations confirmed that metabolic genes and virulence regulators are often under selection within hosts. Again, transcriptomics will be necessary to determine what effect the mutations I saw in our *S. aureus* strains have and whether their fitness differs based on existing antibiotic resistance. Within the genome sequence data, I was able to reconfirm the interaction between oxacillin exposure and worm exposure. *S. aureus* strains exposed to both developed mutations in beta-lactam resistance associated genes. Interestingly, this occurred in MRSA strains as well. Resistance by means other than PBP2a may potentially relieve the fitness costs of PBP2a. Competition assays between isogenic MRSA strains with and without further beta-lactam resistance mutations would tell us whether they are fitness rescuing mutations. ACME, which was found mutated or deleted in some strains, should also be interrogated for fitness costs both within and outside of hosts.

Virulence assays qualitatively indicated that oxacillin exposure imposed differential selection on virulence evolution based on pre-existing antibiotic resistance and host exposure. However, the noisiness of the assay makes it necessary to reconfirm these results with more precise assays. Examples include fecundity assays and single worm survival assays. Interestingly, virulence did not depend on *agr* gene maintenance, suggesting virulence is incredibly multifactorial. Transcriptomics will indicate which virulence factors were maintained and contribute to their virulence. PVL, which is likely maintained in *agr*- strains may be important to the virulence I observed. In addition, I observed a large degree of stochasticity in the evolution of these replicate populations. Therefore, it may be of interest to look within treatments at those that were more virulent and less virulent after evolution. As the evolution was over a relatively short time-period (12 weeks), continuing the experiment for further cycles may help glean more insights into the continued evolution of *S. aureus* within *C. elegans*.

In summary, I have shown that the evolution of *S. aureus* phenotypes can be differentially regulated by epistasis and are subject to trade-offs within a host. In the case of delta-toxin production, epistasis is the main evolutionary constraint. In the case of vancomycin-intermediate resistance, a phenotype largely driven by parallel evolution, epistasis plays a role in modulating SNP patterns and other gene mutations as well as fitness. Virulence evolution was found to be dependent on a large repertoire of genes and dominated by trade-offs between virulence and antibiotic exposure, which were stronger within the host.

References

Aanensen DM, Feil EJ, Holden MTG, Dordel J, Yeats CA, Fedosejev A, Goater R, Castillo-Ramírez S, Corander J, Colijn C, Chlebowicz MA, Schouls L, Heck M, Pluister G, Ruimy R, Kahlmeter G, Åhman J, Matuschek E, Friedrich AW, Parkhill J, Bentley SD, Spratt BG, Grundmann H, 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. DOI: 10.1128/mBio.00444-16.

Adler M. 2015. pigz: A parallel implementation of gzip for modern multi-processor, multi-core machines. Jet Propulsion Laboratory.

Alam MT, Petit RA 3rd, Crispelll EK 3rd, Thornton TA, Conneely KN, Jiang Y, Satola SW, Read TD. 2014. Dissecting vancomycin intermediate resistance in Staphylococcus aureus using genome-wide association. *Genome biology and evolution*. DOI: 10.1093/gbe/evu092.

Altman DR, Sullivan MJ, Chacko KI, Balasubramanian D, Pak TR, Sause WE, Kumar K, Sebra R, Deikus G, Attie O, Rose H, Lewis M, Fulmer Y, Bashir A, Kasarskis A, Schadt EE, Richardson AR, Torres VJ, Shopsin B, van Bakel H. 2018. Genome Plasticity of agr-Defective Staphylococcus aureus during Clinical Infection. Infection and immunity 86. DOI: 10.1128/IAI.00331-18.

Andersson DI, Hughes D. 2014. Microbiological effects of sublethal levels of antibiotics. Nature reviews. Microbiology 12:465–478.

Andersson DI, Levin BR. 1999. The biological cost of antibiotic resistance. *Current opinion in microbiology* 2:489–493.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.

Antibiotic Resistance Threats in the United States, 2013 | Antibiotic/Antimicrobial Resistance | CDC. *Available at https://www.cdc.gov/drugresistance/threat-report-2013/index.html* (accessed

December 10, 2017).

Antonopoulos DA, Assaf R, Aziz RK, Brettin T, Bun C, Conrad N, Davis JJ, Dietrich EM, Disz T, Gerdes S, Kenyon RW, Machi D, Mao C, Murphy-Olson DE, Nordberg EK, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Santerre J, Shukla M, Stevens RL, VanOeffelen M, Vonstein V, Warren AS, Wattam AR, Xia F, Yoo H. 2017. PATRIC as a unique resource for studying antimicrobial resistance. *Briefings in bioinformatics*. DOI: 10.1093/bib/bbx083.

Asadpour L, Ghazanfari N. 2019. Detection of vancomycin nonsusceptible strains in clinical isolates of Staphylococcus aureus in northern Iran. *International microbiology: the official journal of the Spanish Society for Microbiology* 22:411–417. DOI: 10.1007/s10123-019-00063-7.

Askarian F, Ajayi C, Hanssen A-M, van Sorge NM, Pettersen I, Diep DB, Sollid JUE, Johannessen M. 2016. The interaction between Staphylococcus aureus SdrD and desmoglein 1 is important for adhesion to host cells. *Scientific reports* 6:22134. DOI: 10.1038/srep22134.

Askarian F, Uchiyama S, Valderrama JA, Ajayi C, Sollid JUE, van Sorge NM, Nizet V, van Strijp JAG, Johannessen M. 2017. Serine-Aspartate Repeat Protein D Increases Staphylococcus aureus Virulence and Survival in Blood. *Infection and immunity* 85. DOI: 10.1128/IAI.00559-16.

Baba T, Takeuchi F, Kuroda M, Yuzawa H, Aoki K-I, Oguchi A, Nagai Y, Iwama N, Asano K, Naimi T, Kuroda H, Cui L, Yamamoto K, Hiramatsu K. 2002. Genome and virulence determinants of high virulence community-acquired MRSA. *The Lancet* 359:1819–1827.

Bacigalupe R, Tormo-Mas MÁ, Penadés JR, Fitzgerald JR. 2019. A multihost bacterial pathogen overcomes continuous population bottlenecks to adapt to new host species. *Science advances* 5:eaax0063.

Bae T, Banger AK, Wallace A, Glass EM, Aslund F, Schneewind O, Missiakas DM. 2004. Staphylococcus aureus virulence genes identified by bursa aurealis mutagenesis and nematode

killing. *Proceedings of the National Academy of Sciences of the United States of America* 101:12312–12317. DOI: 10.1073/pnas.0404728101.

Baek JY, Chung DR, Ko KS, Kim SH, Yang S-J, Kang C-I, Peck KR, Song J-H. 2017. Genetic alterations responsible for reduced susceptibility to vancomycin in community-associated MRSA strains of ST72. *The Journal of antimicrobial chemotherapy* 72:2454–2460. DOI: 10.1093/jac/dkx175.

Baker-Austin C, Wright MS, Stepanauskas R, McArthur JV. 2006. Co-selection of antibiotic and metal resistance. *Trends in microbiology* 14:176–182.

Balaban N, Novick RP. 1995. Translation of RNAIII, the Staphylococcus aureus agr regulatory RNA molecule, can be activated by a 3'-end deletion. *FEMS microbiology letters* 133:155–161.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS. 498 Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, 499 Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A new 500 genome assembly algorithm and its applications to single-cell sequencing. *The Journal* 501:455–477.

Barber KE, Ireland CE, Bukavyn N, Rybak MJ. 2014. Observation of "seesaw effect" with vancomycin, teicoplanin, daptomycin and ceftaroline in 150 unique MRSA strains. *Infectious diseases and therapy* 3:35–43. DOI: 10.1007/s40121-014-0023-0.

Barbu EM, Ganesh VK, Gurusiddappa S, Mackenzie RC, Foster TJ, Sudhof TC, Höök M. 2010. beta-Neurexin is a ligand for the Staphylococcus aureus MSCRAMM SdrC. *PLoS pathogens* 6:e1000726. DOI: 10.1371/journal.ppat.1000726.

Barbu EM, Mackenzie C, Foster TJ, Höök M. 2014. SdrC induces staphylococcal biofilm formation through a homophilic interaction. *Molecular microbiology* 94:172–185. DOI: 10.1111/mmi.12750.

Bastide P, Solís-Lemus C, Kriebel R, William Sparks K, Ané C. 2018. Phylogenetic Comparative Methods on Phylogenetic Networks with Reticulations. *Systematic biology* 67:800–820. DOI: 10.1093/sysbio/syy033.

Bateson W. 1909. Mendel's principles of heredity. Cambridge at the University Press. *März 1909; 2nd Impr* 3:1913.

Bean DC, Livermore DM, Papa I, Hall LMC. 2005. Resistance among Escherichia coli to sulphonamides and other antimicrobials now little used in man. *The Journal of antimicrobial chemotherapy* 56:962–964.

Beaudoin T, Yau YCW, Stapleton PJ, Gong Y, Wang PW, Guttman DS, Waters V. 2017. Staphylococcus aureus interaction with Pseudomonas aeruginosa biofilm enhances tobramycin resistance. *NPJ biofilms and microbiomes* 3:25.

Begun J, Sifri CD, Goldman S, Calderwood SB, Ausubel FM. 2005. Staphylococcus aureus virulence factors identified by using a high-throughput Caenorhabditis elegans-killing model. Infection and immunity 73:872–877.

Berlon NR, Qi R, Sharma-Kuinkel BK, Joo H-S, Park LP, George D, Thaden JT, Messina JA, Maskarinec SA, Mueller-Premru M, Athan E, Tattevin P, Pericas JM, Woods CW, Otto M, Fowler VG Jr. 2015. Clinical MRSA isolates from skin and soft tissue infections show increased in vitro production of phenol soluble modulins. *The Journal of infection* 71:447–457. DOI: 10.1016/j.jinf.2015.06.005.

Bernheimer AW, Rudy B. 1986. Interactions between membranes and cytolytic peptides. *Biochimica et biophysica acta* 864:123–141.

Betley MJ, Mekalanos JJ. 1985. Staphylococcal enterotoxin A is encoded by phage. *Science* 229:185–187.

Bezanson J, Edelman A, Karpinski S, Shah V. 2017. Julia: A Fresh Approach to Numerical Computing. *SIAM Review* 59:65–98. DOI: 10.1137/141000671.

Bhakdi S, Tranum-Jensen J. 1991. Alpha-toxin of Staphylococcus aureus. *Microbiological reviews* 55:733–751.

Bille E, Zahar J-R, Perrin A, Morelle S, Kriz P, Jolley KA, Maiden MCJ, Dervin C, Nassif X, Tinsley CR. 2005. A chromosomally integrated bacteriophage in invasive meningococci. *The Journal of experimental medicine* 201:1905–1913.

Blomberg SP, Garland T Jr, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution; international journal of organic evolution* 57:717–745.

Boucher B, Jenna S. 2013. Genetic interaction networks: better understand to better predict. *Frontiers in genetics* 4:290.

Boyle EA, Li YI, Pritchard JK. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169:1177–1186. DOI: 10.1016/j.cell.2017.05.038.

Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, Earle S, Pankhurst LJ, Anson L, de Cesare M, Piazza P, Votintseva AA, Golubchik T, Wilson DJ, Wyllie DH, Diel R, Niemann S, Feuerriegel S, Kohl TA, Ismail N, Omar SV, Smith EG, Buck D, McVean G, Walker AS, Peto TEA, Crook DW, Iqbal Z. 2015. Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. *Nature communications* 6:10063.

Břinda K, Callendrello A, Cowley L, Charalampous T, Lee RS, MacFadden DR, Kucherov G, O'Grady J, Baym M, Hanage WP. 2018. Lineage calling can identify antibiotic resistant clones within minutes. *bioRxiv*:403204. DOI: 10.1101/403204.

Bubeck Wardenburg J, Patel RJ, Schneewind O. 2007. Surface proteins and exotoxins are required for the pathogenesis of Staphylococcus aureus pneumonia. *Infection and immunity* 75:1040–1044. DOI: 10.1128/IAI.01313-06.

Bushnell B. 2014. BBTools software package. URL http://sourceforge. net/projects/bbmap.

Bushnell B. 2016. BBMap short read aligner. *University of California, Berkeley, California. URL http://sourceforge. net/projects/bbmap*.

Cabello FC. 2006. Heavy use of prophylactic antibiotics in aquaculture: a growing problem for human and animal health and for the environment. *Environmental microbiology* 8:1137–1144.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC bioinformatics* 10:421. DOI: 10.1186/1471-2105-10-421.

Cameron DR, Ward D V, Kostoulias X, Howden BP, Moellering RC, Eliopoulos GM, Peleg AY. 2012. Serine/threonine phosphatase Stp1 contributes to reduced susceptibility to vancomycin and virulence in Staphylococcus aureus. *The Journal of infectious diseases* 205:1677–1687. DOI: 10.1093/infdis/jis252.

Cao MD, Ganesamoorthy D, Elliott AG, Zhang H, Cooper MA, Coin LJM. 2016. Streaming algorithms for identification of pathogens and antibiotic resistance potential from real-time MinION(TM) sequencing. *GigaScience* 5:32.

Cassone M, Campanile F, Pantosti A, Venditti M, Stefani S. 2004. Identification of a variant" Rome clone" of methicillin-resistant Staphylococcus aureus with decreased susceptibility to vancomycin, responsible for an outbreak in an intensive care unit. *Microbial drug resistance* 10:43–49.

Centers for Disease Control and Prevention (CDC). 2001. Methicillin-resistant Staphylococcus aureus skin or soft tissue infections in a state prison--Mississippi, 2000. *MMWR. Morbidity and mortality weekly report* 50:919–922.

Challagundla L, Luo X, Tickler IA, Didelot X, Coleman DC, Shore AC, Coombs GW, Sordelli DO, Brown EL, Skov R, Larsen AR, Reyes J, Robledo IE, Vazquez GJ, Rivera R, Fey PD,

Stevenson K, Wang S-H, Kreiswirth BN, Mediavilla JR, Arias CA, Planet PJ, Nolan RL, Tenover FC, Goering RV, Robinson DA. 2018. Range Expansion and the Origin of USA300 North American Epidemic Methicillin-Resistant Staphylococcus aureus. *mBio* 9. DOI: 10.1128/mBio.02016-17.

Chambers HF, Deleo FR. 2009. Waves of resistance: Staphylococcus aureus in the antibiotic era. *Nature reviews. Microbiology* 7:629–641.

Chang S, Sievert DM, Hageman JC, Boulton ML, Tenover FC, Downes FP, Shah S, Rudrik JT, Pupp GR, Brown WJ, Cardo D, Fridkin SK, Vancomycin-Resistant Staphylococcus aureus Investigative Team. 2003. Infection with vancomycin-resistant Staphylococcus aureus containing the vanA resistance gene. *The New England journal of medicine* 348:1342–1347.

Charles PGP, Ward PB, Johnson PDR, Howden BP, Grayson ML. 2004. Clinical features associated with bacteremia due to heterogeneous vancomycin-intermediate Staphylococcus aureus. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* 38:448–451. DOI: 10.1086/381093.

Chen C-J, Huang Y-C, Chiu C-H. 2015. Multiple pathways of cross-resistance to glycopeptides and daptomycin in persistent MRSA bacteraemia. *The Journal of antimicrobial chemotherapy* 70:2965–2972. DOI: 10.1093/jac/dkv225.

Chen F-J, Lauderdale T-L, Lee C-H, Hsu Y-C, Huang I-W, Hsu P-C, Yang C-S. 2018. Effect of a Point Mutation in mprF on Susceptibility to Daptomycin, Vancomycin, and Oxacillin in an MRSA Clinical Strain. *Frontiers in microbiology* 9:1086. DOI: 10.3389/fmicb.2018.01086.

Chen L, Zheng D, Liu B, Yang J, Jin Q. 2016. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. *Nucleic acids research* 44:D694–7.

Chen ML, Doddi A, Royer J, Freschi L, Schito M, Ezewudo M, Kohane IS, Beam A, Farhat M. 2018. Deep Learning Predicts Tuberculosis Drug Resistance Status from Whole-Genome

Sequencing Data. *bioRxiv*:275628. DOI: 10.1101/275628.

Chen T, Guestrin C. 2016. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. New York, NY, USA: ACM, 785–794. DOI: 10.1145/2939672.2939785.

Cheng AG, Kim HK, Burts ML, Krausz T, Schneewind O, Missiakas DM. 2009. Genetic requirements for Staphylococcus aureus abscess formation and persistence in host tissues. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology* 23:3393–3404. DOI: 10.1096/fj.09-135467.

Cheung GYC, Duong AC, Otto M. 2012. Direct and synergistic hemolysis caused by Staphylococcus phenol-soluble modulins: implications for diagnosis and pathogenesis. *Microbes and infection / Institut Pasteur* 14:380–386.

Cheung GYC, Kretschmer D, Duong AC, Yeh AJ, Ho TV, Chen Y, Joo H-S, Kreiswirth BN, Peschel A, Otto M. 2014. Production of an attenuated phenol-soluble modulin variant unique to the MRSA clonal complex 30 increases severity of bloodstream infection. *PLoS pathogens* 10:e1004298. DOI: 10.1371/journal.ppat.1004298.

Chi PB, Liberles DA. 2016. Selection on protein structure, interaction, and sequence. *Protein science: a publication of the Protein Society* 25:1168–1178. DOI: 10.1002/pro.2886.

Choe D, Szubin R, Dahesh S, Cho S, Nizet V, Palsson B, Cho B-K. 2018. Genome-scale analysis of Methicillin-resistant Staphylococcus aureus USA300 reveals a tradeoff between pathogenesis and drug resistance. Scientific reports 8:2215.

Clausen PTLC, Zankari E, Aarestrup FM, Lund O. 2016. Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole genome data. *The Journal of antimicrobial chemotherapy* 71:2484–2488.

Clsi. 2013. Performance standards for antimicrobial susceptibility testing: CLSI approved standard M100-S23.

Coddington JA. 1988. CLADISTIC TESTS OF ADAPTATIONAL HYPOTHESES. *Cladistics* 4:3–22. DOI: 10.1111/j.1096-0031.1988.tb00465.x.

Codoñer FM, Pou C, Thielen A, García F, Delgado R, Dalmau D, Álvarez-Tejado M, Ruiz L, Clotet B, Paredes R. 2011. Added value of deep sequencing relative to population sequencing in heavily pre-treated HIV-1-infected subjects. *PloS one* 6:e19461.

Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, Mallard K, Nair M, Miranda A, Alves A, Perdigão J, Viveiros M, Portugal I, Hasan Z, Hasan R, Glynn JR, Martin N, Pain A, Clark TG. 2015. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome medicine* 7:51.

Collins C, Didelot X. 2017. A Phylogenetic Method To Perform Genome-Wide Association Studies In Microbes That Accounts For Population Structure And Recombination. *bioRxiv*:140798. DOI: 10.1101/140798.

Collins C, Didelot X. 2018. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. *PLoS computational biology* 14:e1005958.

Collins J, Buckling A, Massey RC. 2008. Identification of factors contributing to T-cell toxicity of Staphylococcus aureus clinical isolates. *Journal of clinical microbiology* 46:2112–2114. DOI: 10.1128/JCM.00156-08.

Collins J, Rudkin J, Recker M, Pozzi C, O'Gara JP, Massey RC. 2010. Offsetting virulence and antibiotic resistance costs by MRSA. *The ISME journal* 4:577–584.

Cornforth DM, Dees JL, Ibberson CB, Huse HK, Mathiesen IH, Kirketerp-Møller K, Wolcott RD, Rumbaugh KP, Bjarnsholt T, Whiteley M. 2018. Pseudomonas aeruginosa transcriptome during

human infection. *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.1717525115.

Corrigan RM, Miajlovic H, Foster TJ. 2009. Surface proteins that promote adherence of Staphylococcus aureus to human desquamated nasal epithelial cells. *BMC microbiology* 9:22. DOI: 10.1186/1471-2180-9-22.

Crisan A, McKee G, Munzner T, Gardy JL. 2018. Evidence-based design and evaluation of a whole genome sequencing clinical report for the reference microbiology laboratory. *PeerJ* 6:e4218.

Crisostomo MI, Westh H, Tomasz A, Chung M, Oliveira DC, de Lencastre H. 2001. The evolution of methicillin resistance in Staphylococcus aureus: Similarity of genetic backgrounds in historically early methicillin-susceptible and -resistant isolates and contemporary epidemic clones. *Proceedings of the National Academy of Sciences* 98:9865–9870. DOI: 10.1073/pnas.161272898.

Crossley K, Landesman B, Zaske D. 1979. An Outbreak of Infections Caused by Strains of Staphylococcus aureus Resistant to Methicillin and Aminoglycosides. II. Epidemiologic Studies. *Journal of Infectious Diseases* 139:280–287. DOI: 10.1093/infdis/139.3.280.

Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic acids research* 43:e15. DOI: 10.1093/nar/gku1196.

Cui L, Isii T, Fukuda M, Ochiai T, Neoh H-M, Camargo ILB da C, Watanabe Y, Shoji M, Hishinuma T, Hiramatsu K. 2010. An RpoB mutation confers dual heteroresistance to daptomycin and vancomycin in Staphylococcus aureus. *Antimicrobial agents and chemotherapy* 54:5222–5233. DOI: 10.1128/AAC.00437-10.

Cui L, Ma X, Sato K, Okuma K, Tenover FC, Mamizuka EM, Gemmell CG, Kim M-N, Ploy M-C, El-Solh N, Ferraz V, Hiramatsu K. 2003. Cell wall thickening is a common feature of

vancomycin resistance in Staphylococcus aureus. *Journal of clinical microbiology* 41:5–14. DOI: 10.1128/jcm.41.1.5-14.2003.

Cui L, Tominaga E, Neoh H-M, Hiramatsu K. 2006. Correlation between Reduced Daptomycin Susceptibility and Vancomycin Resistance in Vancomycin-Intermediate Staphylococcus aureus. *Antimicrobial agents and chemotherapy* 50:1079–1082. DOI: 10.1128/AAC.50.3.1079-1082.2006.

Cui Y, Yang X, Didelot X, Guo C, Li D, Yan Y, Zhang Y, Yuan Y, Yang H, Wang J, Wang J, Song Y, Zhou D, Falush D, Yang R. 2015. Epidemic Clones, Oceanic Gene Pools, and Eco-LD in the Free Living Marine Pathogen Vibrio parahaemolyticus. *Molecular biology and evolution* 32:1396–1410.

Cutuli MA, Petronio Petronio G, Vergalito F, Magnifico I, Pietrangelo L, Venditti N, Di Marco R. 2019. Galleria mellonella as a consolidated in vivo model hosts: New developments in antibacterial strategies and novel drug testing. Virulence 10:527–541.

Das S, Lindemann C, Young BC, Muller J, Österreich B, Ternette N, Winkler A-C, Paprotka K, Reinhardt R, Förstner KU, Allen E, Flaxman A, Yamaguchi Y, Rollier CS, van Diemen P, Blättner S, Remmele CW, Selle M, Dittrich M, Müller T, Vogel J, Ohlsen K, Crook DW, Massey R, Wilson DJ, Rudel T, Wyllie DH, Fraunholz MJ. 2016. Natural mutations in a Staphylococcus aureus virulence regulator attenuate cytotoxicity but permit bacteremia and abscess formation. *Proceedings of the National Academy of Sciences of the United States of America* 113:E3101–10.

Davis JJ, Boisvert S, Brettin T, Kenyon RW, Mao C, Olson R, Overbeek R, Santerre J, Shukla M, Wattam AR, Will R, Xia F, Stevens R. 2016. Antimicrobial Resistance Prediction in PATRIC and RAST. *Scientific reports* 6:27930.

Day MR, Doumith M, Do Nascimento V, Nair S, Ashton PM, Jenkins C, Dallman TJ, Stevens FJ, Freedman J, Hopkins KL, Woodford N, De Pinna EM, Godbole G. 2017. Comparison of

phenotypic and WGS-derived antimicrobial resistance profiles of Salmonella enterica serovars Typhi and Paratyphi. *The Journal of antimicrobial chemotherapy*. DOI: 10.1093/jac/dkx379.

de Lassence A, Hidri N, Timsit J-F, Joly-Guillou M-L, Thiery G, Boyer A, Lable P, Blivet A, Kalinowski H, Martin Y, Lajonchere J-P, Dreyfuss D. 2006. Control and outcome of a large outbreak of colonization and infection with glycopeptide-intermediate Staphylococcus aureus in an intensive care unit. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* 42:170–178. DOI: 10.1086/498898.

de Man TJB, Limbago BM. 2016. SSTAR, a Stand-Alone Easy-To-Use Antimicrobial Resistance Gene Predictor. *mSphere* 1. DOI: 10.1128/mSphere.00050-15.

Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods in molecular biology* 1151:165–188. DOI: 10.1007/978-1-4939-0554-6_12.

Delauné A, Dubrac S, Blanchet C, Poupel O, Mäder U, Hiron A, Leduc A, Fitting C, Nicolas P, Cavaillon J-M, Adib-Conquy M, Msadek T. 2012. The WalKR system controls major staphylococcal virulence genes and is involved in triggering the host inflammatory response. *Infection and immunity* 80:3438–3453. DOI: 10.1128/IAI.00195-12.

Deng X, Memari N, Teatero S, Athey T, Isabel M, Mazzulli T, Fittipaldi N, Gubbay JB. 2016. Whole-genome Sequencing for Surveillance of Invasive Pneumococcal Diseases in Ontario, Canada: Rapid Prediction of Genotype, Antibiotic Resistance and Characterization of Emerging Serotype 22F. *Frontiers in microbiology* 7:2099.

Dengler V, McCallum N, Kiefer P, Christen P, Patrignani A, Vorholt JA, Berger-Bächi B, Senn MM. 2013. Mutation in the C-di-AMP cyclase dacA affects fitness and resistance of methicillin resistant Staphylococcus aureus. PloS one 8:e73512.

Deresinski S. 2009. Vancomycin heteroresistance and methicillin-resistant Staphylococcus aureus. *The Journal of infectious diseases* 199:605–609. DOI: 10.1086/596630.

Deurenberg RH, Stobberingh EE. 2008. The evolution of Staphylococcus aureus. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* 8:747–763.

Di Gregorio S, Fernandez S, Perazzi B, Bello N, Famiglietti A, Mollerach M. 2016. Increase in IS256 transposition in invasive vancomycin heteroresistant Staphylococcus aureus isolate belonging to ST100 and its derived VISA mutants. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* 43:197–202.

Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. 2017. Nextflow enables reproducible computational workflows. Nature biotechnology 35:316–319.

Diekema DJ, Richter SS, Heilmann KP, Dohrn CL, Riahi F, Tendolkar S, McDanel JS, Doern GV. 2014. Continued emergence of USA300 methicillin-resistant Staphylococcus aureus in the United States: results from a nationwide surveillance study. *Infection control and hospital epidemiology: the official journal of the Society of Hospital Epidemiologists of America* 35:285–292.

Diep BA, Gill SR, Chang RF, Phan TH, Chen JH, Davidson MG, Lin F, Lin J, Carleton HA, Mongodin EF, Sensabaugh GF, Perdreau-Remington F. 2006. Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant Staphylococcus aureus. *The Lancet* 367:731–739. DOI: 10.1016/S0140-6736(06)68231-7.

Diep BA, Stone GG, Basuino L, Graber CJ, Miller A, des Etages S-A, Jones A, Palazzolo-Ballance AM, Perdreau-Remington F, Sensabaugh GF, DeLeo FR, Chambers HF. 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* 197:1523–1530.

Dionisio F, Conceição IC, Marques ACR, Fernandes L, Gordo I. 2005. The evolution of a

conjugative plasmid and its ability to increase bacterial fitness. *Biology Letters* 1:250–252. DOI: 10.1098/rsbl.2004.0275.

Dobson AJ, Purves J, Rolff J. 2014. Increased survival of experimentally evolved antimicrobial peptide-resistant Staphylococcus aureus in an animal host. *Evolutionary applications* 7:905–912.

Donlan RM, Costerton JW. 2002. Biofilms: survival mechanisms of clinically relevant microorganisms. *Clinical microbiology reviews* 15:167–193.

Donoghue MJ. 1989. PHYLOGENIES AND THE ANALYSIS OF EVOLUTIONARY SEQUENCES, WITH EXAMPLES FROM SEED PLANTS. *Evolution; international journal of organic evolution* 43:1137–1156.

Doyle RM, Burgess C, Williams R, Gorton R, Booth H, Brown J, Bryant JM, Chan J, Creer D, Holdstock J, Kunst H, Lozewicz S, Platt G, Romero EY, Speight G, Tiberi S, Abubakar I, Lipman M, McHugh TD, Breuer J. 2018. Direct Whole-Genome Sequencing of Sputum Accurately Identifies Drug-Resistant Mycobacterium tuberculosis Faster than MGIT Culture Sequencing. *Journal of clinical microbiology* 56. DOI: 10.1128/JCM.00666-18.

Drlica K, Zhao X. 2007. Mutant selection window hypothesis updated. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* 44:681–688.

Dumitrescu O, Boisset S, Badiou C, Bes M, Benito Y, Reverdy M-E, Vandenesch F, Etienne J, Lina G. 2007. Effect of antibiotics on Staphylococcus aureus producing Panton-Valentine leukocidin. *Antimicrobial agents and chemotherapy* 51:1515–1519.

Dumitrescu O, Choudhury P, Boisset S, Badiou C, Bes M, Benito Y, Wolz C, Vandenesch F, Etienne J, Cheung AL, Bowden MG, Lina G. 2011. Beta-lactams interfering with PBP1 induce Panton-Valentine leukocidin expression by triggering sarA and rot global regulators of Staphylococcus aureus. *Antimicrobial agents and chemotherapy* 55:3261–3271.

Eagle RC. 2006. Complement Factor H Polymorphism in Age-Related Macular Degeneration.

Yearbook of Ophthalmology 2006:245–248. DOI: 10.1016/s0084-392x(08)70394-0.

Earle SG, Wu C-H, Charlesworth J, Stoesser N, Claire Gordon N, Walker TM, Spencer CCA, Iqbal Z, Clifton DA, Hopkins KL, Woodford N, Grace Smith E, Ismail N, Llewelyn MJ, Peto TE, Crook DW, McVean G, Sarah Walker A, Wilson DJ. 2016. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nature Microbiology* 1:16041. DOI: 10.1038/nmicrobiol.2016.41.

El-Halfawy OM, Valvano MA. 2015. Antimicrobial heteroresistance: an emerging field in need of clarity. *Clinical microbiology reviews* 28:191–207. DOI: 10.1128/CMR.00058-14.

Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, Grundman H, Hasman H, Holden MTG, Hopkins KL, Iredell J, Kahlmeter G, Köser CU, MacGowan A, Mevius D, Mulvey M, Naas T, Peto T, Rolain J-M, Samuelsen Ø, Woodford N. 2017. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 23:2–22.

ElMaraachli W, Slater M, Berrada ZL, Lin S-YG, Catanzaro A, Desmond E, Rodrigues C, Victor TC, Crudu V, Gler MT, Rodwell TC. 2015. Predicting differential rifamycin resistance in clinical Mycobacterium tuberculosis isolates by specific rpoB mutations. *The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease* 19:1222–1226.

Ender M, McCallum N, Adhikari R, Berger-Bächi B. 2004. Fitness cost of SCCmec and methicillin resistance levels in Staphylococcus aureus. Antimicrobial agents and chemotherapy 48:2295–2297.

Enne VI, Bennett PM, Livermore DM, Hall LMC. 2004. Enhancement of host fitness by the sul2-coding plasmid p9123 in the absence of selective pressure. *The Journal of antimicrobial chemotherapy*

53:958–963.

Enne VI, Livermore DM, Stephens P, MC Hall L. 2001. Persistence of sulphonamide resistance in Escherichia coli in the UK despite national prescribing restriction. *The Lancet* 357:1325–1328. DOI: 10.1016/s0140-6736(00)04519-0.

Enright MC, Robinson DA, Randle G, Feil EJ, Grundmann H, Spratt BG. 2002. The evolutionary history of methicillin-resistant Staphylococcus aureus (MRSA). *Proceedings of the National Academy of Sciences of the United States of America* 99:7687–7692.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome research* 8:186–194.

Eyre DW, De Silva D, Cole K, Peters J, Cole MJ, Grad YH, Demczuk W, Martin I, Mulvey MR, Crook DW, Walker AS, Peto TEA, Paul J. 2017. WGS to predict antibiotic MICs for Neisseria gonorrhoeae. *The Journal of antimicrobial chemotherapy* 72:1937–1947. DOI: 10.1093/jac/dkx067.

Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, Ip CLC, Wilson DJ, Didelot X, O'Connor L, Lay R, Buck D, Kearns AM, Shaw A, Paul J, Wilcox MH, Donnelly PJ, Peto TEA, Walker AS, Crook DW. 2012. A pilot study of rapid benchtop sequencing of Staphylococcus aureus and Clostridium difficile for outbreak detection and surveillance. *BMJ open* 2. DOI: 10.1136/bmjopen-2012-001124.

Falgenhauer L, Waezsada S-E, Yao Y, Imirzalioglu C, Käsbohrer A, Roesler U, Michael GB, Schwarz S, Werner G, Kreienbrock L, Chakraborty T, RESET consortium. 2016. Colistin resistance gene mcr-1 in extended-spectrum β-lactamase-producing and carbapenemase-producing Gram-negative bacteria in Germany. *The Lancet infectious diseases* 16:282–283.

Feijao P, Yao H-T, Fornika D, Gardy J, Hsiao W, Chauve C, Chindelevitch L. 2018. MentaLiST - A fast MLST caller for large MLST schemes. *Microbial genomics*. DOI: 10.1099/mgen.0.000146.

Fey PD, Endres JL, Yajjala VK, Widhelm TJ, Boissy RJ, Bose JL, Bayles KW. 2013. A genetic resource for rapid and comprehensive phenotype screening of nonessential Staphylococcus aureus genes. *mBio* 4:e00537–12. DOI: 10.1128/mBio.00537-12.

Flandrois J-P, Lina G, Dumitrescu O. 2014. MUBII-TB-DB: a database of mutations associated with antibiotic resistance in Mycobacterium tuberculosis. *BMC bioinformatics* 15:107.

Foucault M-L, Courvalin P, Grillot-Courvalin C. 2009. Fitness cost of VanA-type vancomycin resistance in methicillin-resistant Staphylococcus aureus. *Antimicrobial agents and chemotherapy* 53:2354–2359. DOI: 10.1128/AAC.01702-08.

Fowler PW, Cole K, Claire Gordon N, Kearns AM, Llewelyn MJ, Peto TEA, Crook DW, Sarah Walker A. 2018. Robust Prediction of Resistance to Trimethoprim in Staphylococcus aureus. *Cell Chemical Biology* 0. DOI: 10.1016/j.chembiol.2017.12.009.

Fowler VG Jr, Sakoulas G, McIntyre LM, Meka VG, Arbeit RD, Cabell CH, Stryjewski ME, Eliopoulos GM, Reller LB, Corey GR, Jones T, Lucindo N, Yeaman MR, Bayer AS. 2004. Persistent bacteremia due to methicillin-resistant Staphylococcus aureus infection is associated with agr dysfunction and low-level in vitro resistance to thrombin-induced platelet microbicidal protein. *The Journal of infectious diseases* 190:1140–1149. DOI: 10.1086/423145.

Frees D, Qazi SNA, Hill PJ, Ingmer H. 2003. Alternative roles of ClpX and ClpP in Staphylococcus aureus stress tolerance and virulence. *Molecular microbiology* 48:1565–1578.

Friedman JH. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*:1189–1232.

Friedman L, Alder JD, Silverman JA. 2006. Genetic changes that correlate with reduced susceptibility to daptomycin in Staphylococcus aureus. *Antimicrobial agents and chemotherapy* 50:2137–2145.

From the Centers for Disease Control and Prevention. Four pediatric deaths from community-

acquired methicillin-resistant Staphylococcus aureus--Minnesota and North Dakota, 1997-1999. 1999. *JAMA: the journal of the American Medical Association* 282:1123–1125.

Gagnaire J, Dauwalder O, Boisset S, Khau D, Freydière A-M, Ader F, Bes M, Lina G, Tristan A, Reverdy M-E, Marchand A, Geissmann T, Benito Y, Durand G, Charrier J-P, Etienne J, Welker M, Van Belkum A, Vandenesch F. 2012. Detection of Staphylococcus aureus delta-toxin production by whole-cell MALDI-TOF mass spectrometry. *PloS one* 7:e40660. DOI: 10.1371/journal.pone.0040660.

Gardete S, Tomasz A. 2014. Mechanisms of vancomycin resistance in Staphylococcus aureus. *The Journal of clinical investigation* 124:2836–2840. DOI: 10.1172/JCI68834.

Gardy JL, Loman NJ. 2017. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nature reviews. Genetics*:nrg.2017.88.

Gargis AS, Kalman L, Lubin IM. 2016. Assuring the Quality of Next-Generation Sequencing in Clinical Microbiology and Public Health Laboratories. *Journal of clinical microbiology* 54:2857–2865.

Geisinger E, Isberg RR. 2017. Interplay Between Antibiotic Resistance and Virulence During Disease Promoted by Multidrug-Resistant Bacteria. *The Journal of infectious diseases* 215:S9–S17.

Genestier A-L, Michallet M-C, Prévost G, Bellot G, Chalabreysse L, Peyrol S, Thivolet F, Etienne J, Lina G, Vallette FM, Vandenesch F, Genestier L. 2005. Staphylococcus aureus Panton-Valentine leukocidin directly targets mitochondria and induces Bax-independent apoptosis of human neutrophils. *The Journal of clinical investigation* 115:3117–3127. DOI: 10.1172/JCI22684.

Gibson MK, Forsberg KJ, Dantas G. 2015. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *The ISME journal* 9:207–216.

Ginn AN, Wiklendt AM, Zong Z, Lin RTP, Teo JWP, Tambyah PA, Peterson LR, Kaul K, Partridge SR, Iredell JR. 2014. Prediction of major antibiotic resistance in Escherichia coli and Klebsiella pneumoniae in Singapore, USA and China using a limited set of gene targets.

International journal of antimicrobial agents 43:563–565.

Ginn AN, Zong Z, Wiklendt AM, Thomas LC, Merlino J, Gottlieb T, van Hal S, Harkness J, Macleod C, Bell SM, Leroi MJ, Partridge SR, Iredell JR. 2013. Limited diversity in the gene pool allows prediction of third-generation cephalosporin and aminoglycoside resistance in Escherichia coli and Klebsiella pneumoniae. *International journal of antimicrobial agents* 42:19–26.

Giordano F, Aigrain L, Quail MA, Coupland P, Bonfield JK, Davies RM, Tischler G, Jackson DK, Keane TM, Li J, Yue J-X, Liti G, Durbin R, Ning Z. 2017. De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. *Scientific reports* 7:3935.

Girard LR, Fiedler TJ, Harris TW, Carvalho F, Antoshechkin I, Han M, Sternberg PW, Stein LD, Chalfie M. 2007. WormBook: the online review of Caenorhabditis elegans biology. Nucleic acids research 35:D472–5.

Glaser P, Martins-Simões P, Villain A, Barbier M, Tristan A, Bouchier C, Ma L, Bes M, Laurent F, Guillemot D, Wirth T, Vandenesch F. 2016. Demography and Intercontinental Spread of the USA300 Community-Acquired Methicillin-Resistant Staphylococcus aureus Lineage. mBio 7:e02183–15.

Goerke C, Pantucek R, Holtfreter S, Schulte B, Zink M, Grumann D, Bröker BM, Doskar J, Wolz C. 2009. Diversity of prophages in dominant Staphylococcus aureus clonal lineages. *Journal of bacteriology* 191:3462–3468.

Goldberg B, Sichtig H, Geyer C, Ledeboer N, Weinstock GM. 2015. Making the Leap from Research Laboratory to Clinic: Challenges and Opportunities for Next-Generation Sequencing in Infectious Disease Diagnostics. *mBio* 6. DOI: 10.1128/mBio.01888-15.

Goncheva MI, Flannagan RS, Heinrichs DE. 2020. De Novo Purine Biosynthesis Is Required for Intracellular Growth of Staphylococcus aureus and for the Hypervirulence Phenotype of a purR Mutant. Infection and immunity 88. DOI: 10.1128/IAI.00104-20.

Goncheva MI, Flannagan RS, Sterling BE, Laakso HA, Friedrich NC, Kaiser JC, Watson DW, Wilson CH, Sheldon JR, McGavin MJ, Kiser PK, Heinrichs DE. 2019. Stress-induced inactivation of the Staphylococcus aureus purine biosynthesis repressor leads to hypervirulence. Nature communications 10:775.

Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nature reviews. Genetics* 17:333–351.

Gordon NC, Price JR, Cole K, Everitt R, Morgan M, Finney J, Kearns AM, Pichon B, Young B, Wilson DJ, Llewelyn MJ, Paul J, Peto TEA, Crook DW, Walker AS, Golubchik T. 2014. Prediction of Staphylococcus aureus Antimicrobial Resistance by Whole-Genome Sequencing. *Journal of clinical microbiology* 52:1182–1191.

Grosser MR, Paluscio E, Thurlow LR, Dillon MM, Cooper VS, Kawula TH, Richardson AR. 2018. Genetic requirements for Staphylococcus aureus nitric oxide resistance and virulence. *PLoS pathogens* 14:e1006907. DOI: 10.1371/journal.ppat.1006907.

Grumann D, Nübel U, Bröker BM. 2014. Staphylococcus aureus toxins--their functions and genetics. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* 21:583–592. DOI: 10.1016/j.meegid.2013.03.013.

Guérillot R, da Silva AG, Monk I, Giulieri S, Tomita T, Alison E, Porter J, Pidot S, Gao W, Peleg AY, Seemann T, Stinear TP, Howden BP. 2018b. Convergent Evolution Driven by Rifampin Exacerbates the Global Burden of Drug-Resistant Staphylococcus aureus. *mSphere* 3:e00550–17.

Guérillot R, Li L, Baines S, Howden BO, Schultz M, Seemann T, Monk I, Pidot SJ, Gao W, Giulieri S, da Silva AG, D'Agata A, Tomita T, Peleg AY, Stinear TP, Howden BP. 2018a. Comprehensive antibiotic-linked mutation assessment by Resistance Mutation Sequencing (RM-seq). *bioRxiv*:257915. DOI: 10.1101/257915.

Gullberg E, Cao S, Berg OG, Ilbäck C, Sandegren L, Hughes D, Andersson DI. 2011. Selection of resistant bacteria at very low antibiotic concentrations. PLoS pathogens 7:e1002158.

Guo H, Hall JW, Yang J, Ji Y. 2017. The SaeRS Two-Component System Controls Survival of Staphylococcus aureus in Human Blood through Regulation of Coagulase. Frontiers in cellular and infection microbiology 7:204.

Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, Rolain J-M. 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrobial agents and chemotherapy* 58:212–220.

Hafer C, Lin Y, Kornblum J, Lowy FD, Uhlemann A-C. 2012. Vancomycin-intermediate Staphylococcus aureus: the contribution of selected gene mutations in clinical isolates. *Antimicrobial agents and chemotherapy*:27. DOI: 10.1128/AAC.01139-12.

Hall MD, Holden MT, Srisomang P, Mahavanakul W, Wuthiekanun V, Limmathurotsakul D, Fountain K, Parkhill J, Nickerson EK, Peacock SJ, Fraser C. 2019. Improved characterisation of MRSA transmission using within-host bacterial sequence diversity. *eLife* 8. DOI: 10.7554/eLife.46402.

Hardy BL, Dickey SW, Plaut RD, Riggins DP, Stibitz S, Otto M, Merrell DS. 2019. Corynebacterium pseudodiphtheriticum Exploits Staphylococcus aureus Virulence Components in a Novel Polymicrobial Defense Strategy. *mBio* 10. DOI: 10.1128/mBio.02491-18.

Harris SR, Cole MJ, Spiteri G, Sánchez-Busó L, Golparian D, Jacobsson S, Goater R, Abudahab K, Yeats CA, Bercot B, Borrego MJ, Crowley B, Stefanelli P, Tripodo F, Abad R, Aanensen DM, Unemo M, Euro-GASP study group. 2018. Public health surveillance of multidrug-resistant clones of Neisseria gonorrhoeae in Europe: a genomic survey. *The Lancet infectious diseases*. DOI: 10.1016/S1473-3099(18)30225-1.

Harris SR, Feil EJ, Holden MTG, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A,

Day N, Lindsay JA, Edgeworth JD, de Lencastre H, Parkhill J, Peacock SJ, Bentley SD. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327:469–474.

Hassan MH, Butler MS, Ranzoni A, Cooper MA. 2018. Detection and quantification of the heterogeneity of S. aureus bacterial populations to identify antibiotic-induced persistence. *bioRxiv*:320093. DOI: 10.1101/320093.

Hernando-Amado S, Sanz-García F, Blanco P, Martínez JL. 2017. Fitness costs associated with the acquisition of antibiotic resistance. *Essays in biochemistry* 61:37–48.

Herold BC, Immergluck LC, Maranan MC, Lauderdale DS, Gaskin RE, Boyle-Vavra S, Leitch CD, Daum RS. 1998. Community-acquired methicillin-resistant Staphylococcus aureus in children with no identified predisposing risk. *JAMA: the journal of the American Medical Association* 279:593–598.

Hiramatsu K, Aritaka N, Hanaki H, Kawasaki S, Hosoda Y, Hori S, Fukuchi Y, Kobayashi I. 1997a. Dissemination in Japanese hospitals of strains of Staphylococcus aureus heterogeneously resistant to vancomycin. *The Lancet* 350:1670–1673.

Hiramatsu K, Hanaki H, Ino T, Yabuta K, Oguri T, Tenover FC. 1997b. Methicillin-resistant Staphylococcus aureus clinical strain with reduced vancomycin susceptibility. *The Journal of antimicrobial chemotherapy* 40:135–136.

Holden MTG, Hsu L-Y, Kurt K, Weinert LA, Mather AE, Harris SR, Strommenger B, Layer F, Witte W, de Lencastre H, Skov R, Westh H, Zemlicková H, Coombs G, Kearns AM, Hill RLR, Edgeworth J, Gould I, Gant V, Cooke J, Edwards GF, McAdam PR, Templeton KE, McCann A, Zhou Z, Castillo-Ramírez S, Feil EJ, Hudson LO, Enright MC, Balloux F, Aanensen DM, Spratt BG, Fitzgerald JR, Parkhill J, Achtman M, Bentley SD, Nübel U. 2013. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus

aureus pandemic. *Genome research* 23:653–664.

Holland TL, Fowler VG Jr. 2011. Vancomycin minimum inhibitory concentration and outcome in patients with Staphylococcus aureus bacteremia: pearl or pellet? *The Journal of infectious diseases* 204:329–331. DOI: 10.1093/infdis/jir275.

Horne DJ, Pinto LM, Arentz M, Lin S-YG, Desmond E, Flores LL, Steingart KR, Minion J. 2013. Diagnostic accuracy and reproducibility of WHO-endorsed phenotypic drug susceptibility testing methods for first-line and second-line antituberculosis drugs. *Journal of clinical microbiology* 51:393–401.

Horne KC, Howden BP, Grabsch EA, Graham M, Ward PB, Xie S, Mayall BC, Johnson PDR, Grayson ML. 2009. Prospective comparison of the clinical impacts of heterogeneous vancomycin-intermediate methicillin-resistant Staphylococcus aureus (MRSA) and vancomycin-susceptible MRSA. *Antimicrobial agents and chemotherapy* 53:3447–3452. DOI: 10.1128/AAC.01365-08.

Howden BP, Davies JK, Johnson PDR, Stinear TP, Grayson ML. 2010. Reduced Vancomycin Susceptibility in Staphylococcus aureus, Including Vancomycin-Intermediate and Heterogeneous Vancomycin-Intermediate Strains: Resistance Mechanisms, Laboratory Detection, and Clinical Implications. *Clinical microbiology reviews* 23:99–139. DOI: 10.1128/CMR.00042-09.

Howden BP, Johnson PDR, Ward PB, Stinear TP, Davies JK. 2006. Isolates with low-level vancomycin resistance associated with persistent methicillin-resistant Staphylococcus aureus bacteremia. *Antimicrobial agents and chemotherapy* 50:3039–3047. DOI: 10.1128/AAC.00422-06.

Howden BP, McEvoy CRE, Allen DL, Chua K, Gao W, Harrison PF, Bell J, Coombs G, Bennett-Wood V, Porter JL, Robins-Browne R, Davies JK, Seemann T, Stinear TP. 2011. Evolution of Multidrug Resistance during Staphylococcus aureus Infection Involves Mutation of the

Essential Two Component Regulator WalKR. *PLoS pathogens* 7:e1002359. DOI: 10.1371/journal.ppat.1002359.t006.

Howden BP, Peleg AY, Stinear TP. 2013. The evolution of vancomycin intermediate Staphylococcus aureus (VISA) and heterogenous-VISA. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*:1–9. DOI: 10.1016/j.meegid.2013.03.047.

Howden BP, Stinear TP, Allen DL, Johnson PDR, Ward PB, Davies JK. 2008. Genomic analysis reveals a point mutation in the two-component sensor gene graS that leads to intermediate vancomycin resistance in clinical Staphylococcus aureus. *Antimicrobial agents and chemotherapy* 52:3755–3762. DOI: 10.1128/AAC.01613-07.

Hu J, Zhang X, Liu X, Chen C, Sun B. 2015. Mechanism of Reduced Vancomycin Susceptibility Conferred by walK Mutation in Community-Acquired Methicillin-Resistant Staphylococcus aureus Strain MW2. *Antimicrobial agents and chemotherapy* 59:1352–1355. DOI: 10.1128/AAC.04290-14.

Hu Q, Peng H, Rao X. 2016. Molecular Events for Promotion of Vancomycin Resistance in Vancomycin Intermediate Staphylococcus aureus. *Frontiers in microbiology* 7:1601.

Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, Harris SR. 2017. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *bioRxiv*:118000. DOI: 10.1101/118000.

Inouye M, Dashnow H, Raven L-A, Schultz M, Pope B, Tomita T, Zobel J, Holt K. 2014. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* 6:90. DOI: 10.1186/s13073-014-0090-6.

Jacquet R, LaBauve AE, Akoolo L, Patel S, Alqarzaee AA, Wong Fok Lung T, Poorey K, Stinear TP, Thomas VC, Meagher R, Parker D. 2019. Dual gene expression analysis identifies factors

associated with Staphylococcus aureus virulence in diabetic mice. *Infection and immunity*. DOI: 10.1128/IAI.00163-19.

Jaillard M, Lima L, Tournoud M, Mahé P, van Belkum A, Lacroix V, Jacob L. 2018. A fast and agnostic method for bacterial genome-wide association studies: bridging the gap between kmers and genetic events. *bioRxiv*:297754. DOI: 10.1101/297754.

Jain DS, Gupte SR, Aduri R. 2018. A Data Driven Model for Predicting RNA-Protein Interactions based on Gradient Boosting Machine. *Scientific reports* 8:9552. DOI: 10.1038/s41598-018-27814-2.

Janzon L, Löfdahl S, Arvidson S. 1989. Identification and nucleotide sequence of the delta-lysin gene, hld, adjacent to the accessory gene regulator (agr) of Staphylococcus aureus. *Molecular & general genetics: MGG* 219:480–485. DOI: 10.1007/bf00259623.

Jarraud S, Mougel C, Thioulouse J, Lina G, Meugnier H, Forey F, Nesme X, Etienne J, Vandenesch F. 2002. Relationships between Staphylococcus aureus genetic background, virulence factors, agr groups (alleles), and human disease. *Infection and immunity* 70:631–641.

Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA, Brinkman FSL, Wright GD, McArthur AG. 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic acids research* 45:D566–D573.

Johnson PJT, Levin BR. 2013. Pharmacodynamics, population dynamics, and the evolution of persistence in Staphylococcus aureus. *PLoS genetics* 9:e1003123. DOI: 10.1371/journal.pgen.1003123.

Johnston PR, Dobson AJ, Rolff J. 2016. Genomic Signatures of Experimental Adaptation to Antimicrobial Peptides in Staphylococcus aureus. *G3* 6:1535–1539.

Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome open research* 3:124. DOI: 10.12688/wellcomeopenres.14826.1.

Josefsson E, McCrea KW, Ní Eidhin D, O'Connell D, Cox J, Höök M, Foster TJ. 1998. Three new members of the serine-aspartate repeat protein multigene family of Staphylococcus aureus. *Microbiology* 144 (Pt 12):3387–3395. DOI: 10.1099/00221287-144-12-3387.

Joshi GS, Spontak JS, Klapper DG, Richardson AR. 2011. Arginine catabolic mobile element encoded speG abrogates the unique hypersensitivity of Staphylococcus aureus to exogenous polyamines. Molecular microbiology 82:9–20.

Judge K, Hunt M, Reuter S, Tracey A, Quail MA, Parkhill J, Peacock SJ. 2016. Comparison of bacterial genome assembly software for MinION data and their applicability to medical microbiology. *Microbial Genomics*. DOI: 10.1099/mgen.0.000085.

Kahl BC, Becker K, Friedrich AW, Clasen J, Sinha B, Von Eiff C, Peters G. 2003. agr-dependent bacterial interference has no impact on long-term colonization of Staphylococcus aureus during persistent airway infection of cystic fibrosis patients. *Journal of clinical microbiology* 41:5199–5201.

Kaneko J, Kimura T, Narita S, Tomita T, Kamio Y. 1998. Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage φPVL carrying Panton–Valentine leukocidin genes. *Gene* 215:57–67. DOI: 10.1016/s0378-1119(98)00278-9.

Kasimir S, Schönfeld W, Alouf JE, König W. 1990. Effect of Staphylococcus aureus delta-toxin on human granulocyte functions and platelet-activating-factor metabolism. *Infection and immunity* 58:1653–1659.

Kassim A, Omuse G, Premji Z, Revathi G. 2016. Comparison of Clinical Laboratory Standards Institute and European Committee on Antimicrobial Susceptibility Testing guidelines for the interpretation of antibiotic susceptibility at a University teaching hospital in Nairobi, Kenya: a

cross-sectional study. *Annals of clinical microbiology and antimicrobials* 15:21.

Katayama Y, Azechi T, Miyazaki M, Takata T, Sekine M, Matsui H, Hanaki H, Yahara K, Sasano H, Asakura K, Takaku T, Ochiai T, Komatsu N, Chambers HF. 2017. Prevalence of Slow-Growth Vancomycin Nonsusceptibility in Methicillin-Resistant Staphylococcus aureus. *Antimicrobial agents and chemotherapy* 61. DOI: 10.1128/AAC.00452-17.

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* 44:1549–1555. DOI: 10.1128/aac.44.6.1549-1555.2000.

Kato Y, Suzuki T, Ida T, Maebashi K. 2010. Genetic changes associated with glycopeptide resistance in Staphylococcus aureus: predominance of amino acid substitutions in YvqF/VraSR. *The Journal of antimicrobial chemotherapy* 65:37–45. DOI: 10.1093/jac/dkp394.

Kim D-W, Park J-Y, Park K-D, Kim T-H, Lee W-J, Lee S-J, Kim J. 2009. Are there predominant strains and toxins of Staphylococcus aureus in atopic dermatitis patients? Genotypic characterization and toxin determination of S. aureus isolated in adolescent and adult patients with atopic dermatitis. *The Journal of dermatology* 36:75–81. DOI: 10.1111/j.1346-8138.2009.00592.x.

Kim S, Lieberman TD, Kishony R. 2014. Alternating antibiotic treatments constrain evolutionary paths to multidrug resistance. *Proceedings of the National Academy of Sciences of the United States of America* 111:14494–14499.

King JM, Kulhankova K, Stach CS, Vu BG, Salgado-Pabón W. 2016. Phenotypes and Virulence among Staphylococcus aureus USA100, USA200, USA300, USA400, and USA600 Clonal Lineages. *mSphere* 1. DOI: 10.1128/mSphere.00071-16.

Kitur K, Parker D, Nieto P, Ahn DS, Cohen TS, Chung S, Wachtel S, Bueno S, Prince A. 2015. Toxin-Induced Necroptosis Is a Major Mechanism of Staphylococcus aureus Lung Damage. *PLoS pathogens* 11:e1004820. DOI: 10.1371/journal.ppat.1004820.

Klevens RM, Morrison MA, Nadle J, Petit S, Gershman K, Ray S, Harrison LH, Lynfield R, Dumyati G, Townes JM, Craig AS, Zell ER, Fosheim GE, McDougal LK, Carey RB, Fridkin SK, Active Bacterial Core surveillance (ABCs) MRSA Investigators. 2007. Invasive methicillin-resistant Staphylococcus aureus infections in the United States. *JAMA: the journal of the American Medical Association* 298:1763–1771. DOI: 10.1001/jama.298.15.1763.

Knopp M, Andersson DI. 2018. Predictable Phenotypes of Antibiotic Resistance Mutations. *mBio* 9. DOI: 10.1128/mBio.00770-18.

Kobayashi SD, Musser JM, DeLeo FR. 2012. Genomic analysis of the emergence of vancomycin-resistant Staphylococcus aureus. *mBio* 3. DOI: 10.1128/mBio.00170-12.

Koch G, Yepes A, Förstner KU, Wermser C, Stengel ST, Modamio J, Ohlsen K, Foster KR, Lopez D. 2014. Evolution of resistance to a last-resort antibiotic in Staphylococcus aureus via bacterial competition. *Cell* 158:1060–1071.

Kos VN, Déraspe M, McLaughlin RE, Whiteaker JD, Roy PH, Alm RA, Corbeil J, Gardner H. 2015. The resistome of Pseudomonas aeruginosa in relationship to phenotypic susceptibility. *Antimicrobial agents and chemotherapy* 59:427–436.

Köser CU, Holden MTG, Ellington MJ, Cartwright EJP, Brown NM, Ogilvy-Stuart AL, Hsu LY, Chewapreecha C, Croucher NJ, Harris SR, Sanders M, Enright MC, Dougan G, Bentley SD, Parkhill J, Fraser LJ, Betley JR, Schulz-Trieglaff OB, Smith GP, Peacock SJ. 2012. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *The New England journal of medicine* 366:2267–2275.

Kourtis AP, Hatfield K, Baggs J, Mu Y, See I, Epson E, Nadle J, Kainer MA, Dumyati G, Petit S, Ray SM, Emerging Infections Program MRSA author group, Ham D, Capers C, Ewing H, Coffin N, McDonald LC, Jernigan J, Cardo D. 2019. Vital Signs: Epidemiology and Recent Trends in Methicillin-Resistant and in Methicillin-Susceptible Staphylococcus aureus Bloodstream Infections - United States. MMWR. Morbidity and mortality weekly report 68:214–219.

Kuleshov V, Snyder MP, Batzoglou S. 2016. Genome assembly from synthetic long read clouds. *Bioinformatics* 32:i216–i224.

Kümmerer K. 2008. *Pharmaceuticals in the Environment: Sources, Fate, Effects and Risks*. Springer Science & Business Media.

Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki K, Nagai Y, Lian J, Ito T, Kanamori M, Matsumaru H, Maruyama A, Murakami H, Hosoyama A, Mizutani-Ui Y, Takahashi NK, Sawano T, Inoue R, Kaito C, Sekimizu K, Hirakawa H, Kuhara S, Goto S, Yabuzaki J, Kanehisa M, Yamashita A, Oshima K, Furuya K, Yoshino C, Shiba T, Hattori M, Ogasawara N, Hayashi H, Hiramatsu K. 2001. Whole genome sequencing of meticillin-resistant Staphylococcus aureus. *The Lancet* 357:1225–1240.

Laabei M, Recker M, Rudkin JK, Aldeljawi M, Gulay Z, Sloan TJ, Williams P, Endres JL, Bayles KW, Fey PD, Yajjala VK, Widhelm T, Hawkins E, Lewis K, Parfett S, Scowen L, Peacock SJ, Holden M, Wilson D, Read TD, van den Elsen J, Priest NK, Feil EJ, Hurst LD, Josefsson E, Massey RC. 2014. Predicting the virulence of MRSA from its genome sequence. *Genome research* 24:839–849. DOI: 10.1101/gr.165415.113.

Laabei M, Uhlemann A-C, Lowy FD, Austin ED, Yokoyama M, Ouadi K, Feil E, Thorpe HA, Williams B, Perkins M, Peacock SJ, Clarke SR, Dordel J, Holden M, Votintseva AA, Bowden R, Crook DW, Young BC, Wilson DJ, Recker M, Massey RC. 2015. Evolutionary Trade-Offs

Underlie the Multi-faceted Virulence of Staphylococcus aureus. *PLoS biology* 13:e1002229. DOI: 10.1371/journal.pbio.1002229.

Lakin SM, Dean C, Noyes NR, Dettenwanger A, Ross AS, Doster E, Rovira P, Abdo Z, Jones KL, Ruiz J, Belk KE, Morley PS, Boucher C. 2017. MEGARes: an antimicrobial resistance database for high throughput sequencing. *Nucleic acids research* 45:D574–D580.

Lalueza A, Chaves F, San Juan R, Daskalaki M, Otero JR, Aguado JM. 2010. Is high vancomycin minimum inhibitory concentration a good marker to predict the outcome of methicillin-resistant Staphylococcus aureus bacteremia? *The Journal of infectious diseases* 201:311–2; author reply 312–3. DOI: 10.1086/649572.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* 9:357–359.

Lasek-Nesselquist E, Lu J, Schneider R, Ma Z, Russo V, Mishra S, Pai MP, Pata JD, McDonough KA, Malik M. 2019. Insights Into the Evolution of Staphylococcus aureus Daptomycin Resistance From an in vitro Bioreactor Model. *Frontiers in microbiology* 10:345.

Lee GC, Long SW, Musser JM, Beres SB, Olsen RJ, Dallas SD, Nunez YO, Frei CR. 2015. Comparative whole genome sequencing of community-associated methicillin-resistant Staphylococcus aureus sequence type 8 from primary care clinics in a Texas community. *Pharmacotherapy* 35:220–228.

Lee SM, Ender M, Adhikari R, Smith JMB, Berger-Bächi B, Cook GM. 2007. Fitness cost of staphylococcal cassette chromosome mec in methicillin-resistant Staphylococcus aureus by way of continuous culture. *Antimicrobial agents and chemotherapy* 51:1497–1499.

Lees JA, Galardini M, Bentley SD, Weiser JN, Corander J. 2018. pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* 34:4310–4312.

Lees JA, Vehkala M, Välimäki N, Harris SR, Chewapreecha C, Croucher NJ, Marttinen P, Davies MR, Steer AC, Tong SYC, Honkela A, Parkhill J, Bentley SD, Corander J. 2016. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nature communications* 7:12797. DOI: 10.1038/ncomms12797.

Lei MG, Cue D, Roux CM, Dunman PM, Lee CY. 2011. Rsp inhibits attachment and biofilm formation by repressing fnbA in Staphylococcus aureus MW2. *Journal of bacteriology* 193:5231–5241.

Lemon JK, Khil PP, Frank KM, Dekker JP. 2017. Rapid Nanopore Sequencing of Plasmids and Resistance Gene Detection in Clinical Isolates. *Journal of clinical microbiology* 55:3530–3543.

Leopold SR, Goering RV, Witten A, Harmsen D, Mellmann A. 2014. Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. *Journal of clinical microbiology* 52:2365–2370.

Li D, Yang M, Hu J, Ren L, Zhang Y, Li K. 2008. Determination and fate of oxytetracycline and related compounds in oxytetracycline production wastewater and the receiving river. *Environmental toxicology and chemistry / SETAC* 27:80–86.

Li S, Yin Y, Chen H, Wang Q, Wang X, Wang H. 2017. Fitness Cost of Daptomycin-Resistant Staphylococcus aureus Obtained from in Vitro Daptomycin Selection Pressure. Frontiers in microbiology 8:2199.

Li T, He L, Song Y, Villaruz AE, Joo H-S, Liu Q, Zhu Y, Wang Y, Qin J, Otto M, Li M. 2015. AraC-Type Regulator Rsp Adapts Staphylococcus aureus Gene Expression to Acute Infection. *Infection and immunity* 84:723–734. DOI: 10.1128/IAI.01088-15.

Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE Jr, Walker H, Hawkins PA, Tran T, Whitney CG, McGee L, Beall BW. 2016. Penicillin-Binding Protein Transpeptidase Signatures for Tracking and Predicting β-Lactam Resistance Levels in Streptococcus pneumoniae. *mBio* 7. DOI:

10.1128/mBio.00756-16.

Lindsay JA, Moore CE, Day NP, Peacock SJ, Witney AA, Stabler RA, Husain SE, Butcher PD, Hinds J. 2006. Microarrays reveal that each of the ten dominant lineages of Staphylococcus aureus has a unique combination of surface-associated and regulatory genes. *Journal of bacteriology* 188:669–676. DOI: 10.1128/JB.188.2.669-676.2006.

Lindsay JA. 2014. Staphylococcus aureus genomics and the impact of horizontal gene transfer. *International journal of medical microbiology: IJMM* 304:103–109.

Liu B, Pop M. 2009. ARDB—Antibiotic Resistance Genes Database. *Nucleic acids research* 37:D443–D447.

Liu B, Sun B. 2020. Rsp promotes the transcription of virulence factors in an agr-independent manner in Staphylococcus aureus. *Emerging microbes & infections* 9:796–812.

Liu C, Chambers HF. 2003. Staphylococcus aureus with heterogeneous resistance to vancomycin: epidemiology, clinical significance, and critical assessment of diagnostic methods. *Antimicrobial agents and chemotherapy* 47:3040–3045.

Liu H, Lv J, Qi X, Ding Y, Li D, Hu L, Wang L, Yu F. 2015. The carriage of the serine-aspartate repeat protein-encoding sdr genes among Staphylococcus aureus lineages. *The Brazilian journal of infectious diseases: an official publication of the Brazilian Society of Infectious Diseases* 19:498–502. DOI: 10.1016/j.bjid.2015.07.003.

Liu J, Chen D, Peters BM, Li L, Li B, Xu Z, Shirliff ME. 2016. Staphylococcal chromosomal cassettes mec (SCCmec): A mobile genetic element in methicillin-resistant Staphylococcus aureus. *Microbial pathogenesis* 101:56–67.

Liu L, Wang Y, Bojer MS, Andersen PS, Ingmer H. 2020. High persister cell formation by clinical Staphylococcus aureus strains belonging to clonal complex 30. *Microbiology*. DOI: 10.1099/mic.0.000926.

Liu Q, Yeo W-S, Bae T. 2016. The SaeRS Two-Component System of Staphylococcus aureus. Genes 7. DOI: 10.3390/genes7100081.

Lu H, Giordano F, Ning Z. 2016. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, proteomics & bioinformatics* 14:265–279.

Lundqvist-Gustafsson H, Norrman S, Nilsson J, Wilsson Å. 2001. Involvement of p38-mitogen-activated protein kinase in staphylococcus aureus-induced neutrophil apoptosis. Journal of leukocyte biology 70:642–648.

Macedo R, Nunes A, Portugal I, Duarte S, Vieira L, Gomes JP. 2018. Dissecting whole-genome sequencing-based online tools for predicting resistance in Mycobacterium tuberculosis: can we use them for clinical decision guidance? *Tuberculosis* 110:44–51.

MacLean RC, Vogwill T. 2014. Limits to compensatory adaptation and the persistence of antibiotic resistance in pathogenic bacteria. *Evolution, medicine, and public health* 2015:4–12.

Majcherczyk PA, Barblan J-L, Moreillon P, Entenza JM. 2008. Development of glycopeptide-intermediate resistance by Staphylococcus aureus leads to attenuated infectivity in a rat model of endocarditis. *Microbial pathogenesis* 45:408–414.

Makarova O, Johnston P, Rodriguez-Rojas A, El Shazely B, Morales JM, Rolff J. 2018. Genomics of experimental adaptation of Staphylococcus aureus to a natural combination of insect antimicrobial peptides. *Scientific reports* 8:15359.

Marlowe EM, Cohen MD, Hindler JF, Ward KW, Bruckner DA. 2001. Practical strategies for detecting and confirming vancomycin-intermediate Staphylococcus aureus: a tertiary-care hospital laboratory's experience. *Journal of clinical microbiology* 39:2637–2639. DOI: 10.1128/JCM.39.7.2637-2639.2001.

Martinez JL, Baquero F. 2000. Mutation frequencies and antibiotic resistance. *Antimicrobial agents and chemotherapy* 44:1771–1777.

Mashruwala AA, Gries CM, Scherr TD, Kielian T, Boyd JM. 2017. SaeRS Is Responsive to Cellular Respiratory Status and Regulates Fermentative Biofilm Formation in Staphylococcus aureus. Infection and immunity 85. DOI: 10.1128/IAI.00157-17.

Mason A, Foster D, Bradley P, Golubchik T, Doumith M, Gordon NC, Pichon B, Iqbal Z, Staves P, Crook D, Walker AS, Kearns A, Peto T. 2018. Accuracy of different bioinformatics methods in detecting antibiotic resistance and virulence factors from Staphylococcus aureus whole genome sequences. *Journal of clinical microbiology*. DOI: 10.1128/JCM.01815-17.

McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJV, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD. 2013. The Comprehensive Antibiotic Resistance Database. *Antimicrobial agents and chemotherapy* 57:3348–3357.

McCallum N, Karauzum H, Getzmann R, Bischoff M, Majcherczyk P, Berger-Bächi B, Landmann R. 2006. In vivo survival of teicoplanin-resistant Staphylococcus aureus and fitness cost of teicoplanin resistance. *Antimicrobial agents and chemotherapy* 50:2352–2360.

McCarthy AJ, Lindsay JA. 2010. Genetic variation in Staphylococcus aureus surface and immune evasion genes is lineage associated: implications for vaccine design and host-pathogen interactions. *BMC microbiology* 10:173. DOI: 10.1186/1471-2180-10-173.

McDermott PF, Tyson GH, Kabera C, Chen Y, Li C, Folster JP, Ayers SL, Lam C, Tate HP, Zhao S. 2016. Whole-Genome Sequencing for Detecting Antimicrobial Resistance in Nontyphoidal Salmonella. *Antimicrobial agents and chemotherapy* 60:5515–5520.

McEvoy CRE, Tsuji B, Gao W, Seemann T, Porter JL, Doig K, Ngo D, Howden BP, Stinear TP. 2013. Decreased vancomycin susceptibility in Staphylococcus aureus caused by IS256

tempering of WalKR expression. *Antimicrobial agents and chemotherapy* 57:3240–3249. DOI: 10.1128/AAC.00279-13.

McGuinness WA, Malachowa N, DeLeo FR. 2017. Vancomycin Resistance in Staphylococcus aureus. *The Yale journal of biology and medicine* 90:269–281.

McKevitt AI, Bjornson GL, Mauracher CA, Scheifele DW. 1990. Amino acid sequence of a deltalike toxin from Staphylococcus epidermidis. *Infection and immunity* 58:1473–1475.

McLeish KR, Klein JB, Coxon PY, Head KZ, Ward RA. 1998. Bacterial phagocytosis activates extracellular signal-regulated kinase and p38 mitogen-activated protein kinase cascades in human neutrophils. Journal of Leukocyte Biology 64:835–844. DOI: 10.1002/jlb.64.6.835.

McVicker G, Prajsnar TK, Williams A, Wagner NL, Boots M, Renshaw SA, Foster SJ. 2014. Clonal expansion during Staphylococcus aureus infection dynamics reveals the effect of antibiotic intervention. *PLoS pathogens* 10:e1003959.

Melnyk AH, Wong A, Kassen R. 2015. The fitness costs of antibiotic resistance mutations. *Evolutionary Applications* 8:273–283. DOI: 10.1111/eva.12196.

Metcalf BJ, Gertz RE Jr, Gladstone RA, Walker H, Sherwood LK, Jackson D, Li Z, Law C, Hawkins PA, Chochua S, Sheth M, Rayamajhi N, Bentley SD, Kim L, Whitney CG, McGee L, Beall B, Active Bacterial Core surveillance team. 2016. Strain features and distributions in pneumococci from children with invasive disease before and after 13-valent conjugate vaccine implementation in the USA. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 22:60.e9–60.e29.

Miotto P, Tessema B, Tagliani E, Chindelevitch L, Starks AM, Emerson C, Hanna D, Kim PS, Liwski R, Zignol M, Gilpin C, Niemann S, Denkinger CM, Fleming J, Warren RM, Crook D, Posey J, Gagneux S, Hoffner S, Rodrigues C, Comas I, Engelthaler DM, Murray M, Alland D, Rigouts L, Lange C, Dheda K, Hasan R, Ranganathan UDK, McNerney R, Ezewudo M, Cirillo

DM, Schito M, Köser CU, Rodwell TC. 2017. A standardised method for interpreting the association between mutations and phenotypic drug resistance in Mycobacterium tuberculosis. *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology* 50. DOI: 10.1183/13993003.01354-2017.

Mittal P, Brindle J, Stephen J, Plotkin JB, Kudla G. 2018. Codon usage influences fitness through RNA toxicity. *Proceedings of the National Academy of Sciences of the United States of America*:201810022. DOI: 10.1073/pnas.1810022115.

Mittler JE, Lenski RE. 1990. New data on excisions of Mu from E. coli MCS2 cast doubt on directed mutation hypothesis. *Nature* 344:173–175.

Monk IR, Shah IM, Xu M, Tan M-W, Foster TJ. 2012. Transforming the untransformable: application of direct transformation to manipulate genetically Staphylococcus aureus and Staphylococcus epidermidis. mBio 3. DOI: 10.1128/mBio.00277-11.

Monk IR, Tree JJ, Howden BP, Stinear TP, Foster TJ. 2015. Complete Bypass of Restriction Systems for Major Staphylococcus aureus Lineages. mBio 6:e00308–15. DOI: 10.1128/mBio.00308-15.

Montgomery CP, Boyle-Vavra S, Daum RS. 2009. The arginine catabolic mobile element is not associated with enhanced virulence in experimental invasive disease caused by the community-associated methicillin-resistant Staphylococcus aureus USA300 genetic background. Infection and immunity 77:2650–2656.

Moradigaravand D, Palm M, Farewell A, Mustonen V, Warringer J, Parts L. 2018. Precise prediction of antibiotic resistance in Escherichia coli from full genome sequences. *bioRxiv*:338194. DOI: 10.1101/338194.

Moran RA, Anantham S, Holt KE, Hall RM. 2017. Prediction of antibiotic resistance from antibiotic resistance genes detected in antibiotic-resistant commensal Escherichia coli using PCR or WGS.

The Journal of antimicrobial chemotherapy 72:700–704.

Mwangi MM, Wu SW, Zhou Y, Sieradzki K, de Lencastre H, Richardson P, Bruce D, Rubin E, Myers E, Siggia ED, Tomasz A. 2007. Tracking the in vivo evolution of multidrug resistance in Staphylococcus aureus by whole-genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 104:9451–9456. DOI: 10.1073/pnas.0609839104.

Nagaev I, Björkman J, Andersson DI, Hughes D. 2001. Biological cost and compensatory evolution in fusidic acid-resistant Staphylococcus aureus. *Molecular microbiology* 40:433–439.

Nakamura Y, Oscherwitz J, Cease KB, Chan SM, Muñoz-Planillo R, Hasegawa M, Villaruz AE, Cheung GYC, McGavin MJ, Travers JB, Otto M, Inohara N, Núñez G. 2013. Staphylococcus δ-toxin induces allergic skin disease by activating mast cells. *Nature* 503:397–401. DOI: 10.1038/nature12655.

Neoh H-M, Cui L, Yuzawa H, Takeuchi F, Matsuo M, Hiramatsu K. 2008. Mutated response regulator graR is responsible for phenotypic conversion of Staphylococcus aureus from heterogeneous vancomycin-intermediate resistance to vancomycin-intermediate resistance. *Antimicrobial agents and chemotherapy* 52:45–53. DOI: 10.1128/AAC.00534-07.

Neuert S, Nair S, Day MR, Doumith M, Ashton PM, Mellor KC, Jenkins C, Hopkins KL, Woodford N, de Pinna E, Godbole G, Dallman TJ. 2018. Prediction of Phenotypic Antimicrobial Resistance Profiles From Whole Genome Sequences of Non-typhoidal Salmonella enterica. *Frontiers in microbiology* 9:592.

Nguyen M, Brettin T, Long SW, Musser JM, Olsen RJ, Olson R, Shukla M, Stevens RL, Xia F, Yoo H, Davis JJ. 2018. Developing an in silico minimum inhibitory concentration panel test for Klebsiella pneumoniae. *Scientific reports* 8:421. DOI: 10.1038/s41598-017-18972-w.

Nguyen M, Long SW, McDermott PF, Olsen RJ, Olson R, Stevens RL, Tyson GH, Zhao S, Davis JJ. 2019. Using Machine Learning To Predict Antimicrobial MICs and Associated Genomic

Features for Nontyphoidal Salmonella. *Journal of clinical microbiology* 57. DOI: 10.1128/JCM.01260-18.

Novick RP. 2003. Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Molecular microbiology* 48:1429–1449.

Nübel U, Dordel J, Kurt K, Strommenger B, Westh H, Shukla SK, Zemlicková H, Leblois R, Wirth T, Jombart T, Balloux F, Witte W. 2010. A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant Staphylococcus aureus. *PLoS pathogens* 6:e1000855.

O'Neill AJ, Chopra I. 2002. Insertional inactivation of mutS in Staphylococcus aureus reveals potential for elevated mutation frequencies, although the prevalence of mutators in clinical isolates is low. *The Journal of antimicrobial chemotherapy* 50:161–169.

Ohlsen K, Ziebuhr W, Koller KP, Hell W, Wichelhaus TA, Hacker J. 1998. Effects of subinhibitory concentrations of antibiotics on alpha-toxin (hla) gene expression of methicillin-sensitive and methicillin-resistant Staphylococcus aureus isolates. *Antimicrobial agents and chemotherapy* 42:2817–2823.

Ohta T, Hirakawa H, Morikawa K, Maruyama A, Inose Y, Yamashita A, Oshima K, Kuroda M, Hattori M, Hiramatsu K, Kuhara S, Hayashi H. 2004. Nucleotide substitutions in Staphylococcus aureus strains, Mu50, Mu3, and N315. *DNA research: an international journal for rapid publication of reports on genes and genomes* 11:51–56.

Omland KE. 1999. The Assumptions and Challenges of Ancestral State Reconstructions. *Systematic biology* 48:604–611.

Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome biology* 17:132. DOI: 10.1186/s13059-016-0997-x.

Otto M. 2013. Community-associated MRSA: what makes them special? *International journal of medical microbiology: IJMM* 303:324–330. DOI: 10.1016/j.ijmm.2013.02.007.

Otto M. 2014. Staphylococcus aureus toxins. *Current opinion in microbiology* 17:32–37. DOI: 10.1016/j.mib.2013.11.004.

Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T. 2002. Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nature genetics* 32:650–654.

Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31:3691–3693. DOI: 10.1093/bioinformatics/btv421.

Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial genomics* 2:e000056. DOI: 10.1099/mgen.0.000056.

Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884. DOI: 10.1038/44766.

Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, Lemaitre N, Niemann S, Paul J, Rogers TR, Roycroft E, Smith EG, Supply P, Tang P, Wilcox MH, Wordsworth S, Wyllie D, Xu L, Crook DW, COMPASS-TB Study Group. 2016. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. *The Lancet. Respiratory medicine* 4:49–58.

Passalacqua KD, Satola SW, Crispell EK, Read TD. 2012. A Mutation in the PP2C Phosphatase Gene in a Staphylococcus aureus USA300 Clinical Isolate with Reduced Susceptibility to

Vancomycin and Daptomycin. *Antimicrobial agents and chemotherapy* 56:5212–5223. DOI: 10.1128/AAC.05770-11.

Payne A, Holmes N, Rakyan V, Loose M. 2018. Whale watching with BulkVis: A graphical viewer for Oxford Nanopore bulk fast5 files. *bioRxiv*:312256. DOI: 10.1101/312256.

Peleg AY, Monga D, Pillai S, Mylonakis E, Moellering RC Jr, Eliopoulos GM. 2009. Reduced susceptibility to vancomycin influences pathogenicity in Staphylococcus aureus infection. *The Journal of infectious diseases* 199:532–536.

Penley M, Morran L. 2018. Assessment of Caenorhabditis elegans Competitive Fitness in the Presence of a Bacterial Parasite. *BIO-PROTOCOL* 8. DOI: 10.21769/bioprotoc.2971.

Peschel A, Otto M. 2013. Phenol-soluble modulins and staphylococcal infection. *Nature reviews. Microbiology* 11:667–673. DOI: 10.1038/nrmicro3110.

Pesesky MW, Hussain T, Wallace M, Patel S, Andleeb S, Burnham C-AD, Dantas G. 2016. Evaluation of Machine Learning and Rules-Based Approaches for Predicting Antimicrobial Resistance Profiles in Gram-negative Bacilli from Whole Genome Sequence Data. *Frontiers in microbiology* 7:1887.

Petit RA 3rd, Read TD. 2018. Staphylococcus aureus viewed from the perspective of 40,000+ genomes. *PeerJ* 6:e5261. DOI: 10.7717/peerj.5261.

Petit RA, Read TD. 2020. Bactopia: a flexible pipeline for complete analysis of bacterial genomes. bioRxiv.

Phillips I, Culebras E, Moreno F, Baquero F. 1987. Induction of the SOS response by new 4-quinolones. *The Journal of antimicrobial chemotherapy* 20:631–638.

Planet PJ, Diaz L, Kolokotronis S-O, Narechania A, Reyes J, Xing G, Rincon S, Smith H, Panesso D, Ryan C, Smith DP, Guzman M, Zurita J, Sebra R, Deikus G, Nolan RL, Tenover FC, Weinstock GM, Robinson DA, Arias CA. 2015. Parallel Epidemics of Community-Associated

Methicillin-Resistant Staphylococcus aureus USA300 Infection in North and South America. The Journal of infectious diseases 212:1874–1882.

Planet PJ. 2017. Life After USA300: The Rise and Fall of a Superbug. The Journal of infectious diseases 215:S71–S77.

Pohl K, Francois P, Stenz L, Schlink F, Geiger T, Herbert S, Goerke C, Schrenzel J, Wolz C. 2009. CodY in Staphylococcus aureus: a regulatory link between metabolism and virulence gene expression. Journal of bacteriology 191:2953–2963.

Pollitt EJG, Szkuta PT, Burns N, Foster SJ. 2018. Staphylococcus aureus infection dynamics. *PLoS pathogens* 14:e1007112.

Pollitt EJG, West SA, Crusz SA, Burton-Chellew MN, Diggle SP. 2014. Cooperation, quorum sensing, and evolution of virulence in Staphylococcus aureus. Infection and immunity 82:1045–1051.

Pou C, Noguera-Julian M, Pérez-Álvarez S, García F, Delgado R, Dalmau D, Álvarez-Tejado M, Gonzalez D, Sayada C, Chueca N, Pulido F, Ibáñez L, Rodríguez C, Casadellà M, Santos JR, Ruiz L, Clotet B, Paredes R. 2014. Improved prediction of salvage antiretroviral therapy outcomes using ultrasensitive HIV-1 drug resistance testing. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* 59:578–588.

Power RA, Parkhill J, de Oliveira T. 2016. Microbial genome-wide association studies: lessons from human GWAS. *Nature reviews. Genetics*. DOI: 10.1038/nrg.2016.132.

Prajsnar TK, Hamilton R, Garcia-Lara J, McVicker G, Williams A, Boots M, Foster SJ, Renshaw SA. 2012. A privileged intraphagocyte niche is responsible for disseminated infection of Staphylococcus aureus in a zebrafish model. *Cellular microbiology* 14:1600–1619.

Prakash, V, Lewis JS, Jorgensen JH. 2008. Vancomycin MICs for methicillin-resistant Staphylococcus aureus isolates differ based upon the susceptibility test method used. *Antimicrobial agents and chemotherapy* 52:4528. DOI: 10.1128/AAC.00904-08.

Priest NK, Rudkin JK, Feil EJ, van den Elsen JMH, Cheung A, Peacock SJ, Laabei M, Lucks DA, Recker M, Massey RC. 2012. From genotype to phenotype: can systems biology be used to predict Staphylococcus aureus virulence? *Nature reviews. Microbiology* 10:791–797. DOI: 10.1038/nrmicro2880.

Prunier A-L, Leclercq R. 2005. Role of mutS and mutL genes in hypermutability and recombination in Staphylococcus aureus. *Journal of bacteriology* 187:3455–3464.

Pustišek N, Vurnek Živković M, Šitum M. 2016. Quality of Life in Families with Children with Atopic Dermatitis. *Pediatric dermatology* 33:28–32. DOI: 10.1111/pde.12698.

Qiang X, Chen H, Ye X, Su R, Wei L. 2018. M6AMRFS: Robust Prediction of N6-Methyladenosine Sites With Sequence-Based Features in Multiple Species. *Frontiers in genetics* 9:495. DOI: 10.3389/fgene.2018.00495.

Quan TP, Bawa Z, Foster D, Walker T, Del Ojo Elias C, Rathod P, MMM Informatics Group, Iqbal Z, Bradley P, Mowbray J, Walker AS, Crook DW, Wyllie DH, Peto TEA, Smith EG. 2018. Evaluation of Whole-Genome Sequencing for Mycobacterial Species Identification and Drug Susceptibility Testing in a Clinical Setting: a Large-Scale Prospective Assessment of Performance against Line Probe Assays and Phenotyping. *Journal of clinical microbiology* 56. DOI: 10.1128/JCM.01480-17.

Quave CL, Horswill AR. 2018. Identification of Staphylococcal Quorum Sensing Inhibitors by Quantification of ô-Hemolysin with High Performance Liquid Chromatography. In: Leoni L, Rampioni G eds. *Quorum Sensing: Methods and Protocols*. New York, NY: Springer New York, 363–370. DOI: 10.1007/978-1-4939-7309-5_27.

Quave CL, Plano LRW, Bennett BC. 2011. Quorum sensing inhibitors of Staphylococcus aureus from Italian medicinal plants. *Planta medica* 77:188–195. DOI: 10.1055/s-0030-1250145.

Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. 2017. Shotgun metagenomics, from sampling to analysis. *Nature biotechnology* 35:833–844.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. DOI: 10.1093/bioinformatics/btq033.

R Core Team, R Foundation for Statistical Computing, Vienna, Austria. 2016.R: A language and environment for statistical computing

Radin JN, Kelliher JL, Solórzano PKP, Grim KP, Ramezanifard R, Slauch JM, Kehl-Fie TE. 2019. Metal-independent variants of phosphoglycerate mutase promote resistance to nutritional immunity and retention of glycolysis during infection. PLoS pathogens 15:e1007971.

Radlinski L, Rowe SE, Kartchner LB, Maile R, Cairns BA, Vitko NP, Gode CJ, Lachiewicz AM, Wolfgang MC, Conlon BP. 2017. Pseudomonas aeruginosa exoproducts determine antibiotic efficacy against Staphylococcus aureus. *PLoS biology* 15:e2003981.

Rasigade J-P, Leclère A, Alla F, Tessier A, Bes M, Lechiche C, Vernet-Garnier V, Laouénan C, Vandenesch F, Leport C, AEPEI Study Group. 2018. Staphylococcus aureus CC30 Lineage and Absence of sed,j,r-Harboring Plasmid Predict Embolism in Infective Endocarditis. *Frontiers in cellular and infection microbiology* 8:187.

Read TD, Massey RC. 2014. Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies: a new direction for bacteriology. *Genome medicine* 6:109. DOI: 10.1186/s13073-014-0109-z.

Recker M, Laabei M, Toleman MS, Reuter S, Saunderson RB, Blane B, Török ME, Ouadi K, Stevens E, Yokoyama M, Steventon J, Thompson L, Milne G, Bayliss S, Bacon L, Peacock SJ,

Massey RC. 2017. Clonal differences in Staphylococcus aureus bacteraemia-associated mortality. *Nature microbiology* 2:1381–1388. DOI: 10.1038/s41564-017-0001-x.

Reinert K, Langmead B, Weese D, Evers DJ. 2015. Alignment of Next-Generation Sequencing Reads. *Annual review of genomics and human genetics* 16:133–151.

Rhoads A, Au KF. 2015. PacBio Sequencing and Its Applications. *Genomics, proteomics & bioinformatics* 13:278–289.

Richardson EJ, Bacigalupe R, Harrison EM, Weinert LA, Lycett S, Vrieling M, Robb K, Hoskisson PA, Holden MTG, Feil EJ, Paterson GK, Tong SYC, Shittu A, van Wamel W, Aanensen DM, Parkhill J, Peacock SJ, Corander J, Holmes M, Fitzgerald JR. 2018. Gene exchange drives the ecological success of a multi-host bacterial pathogen. *Nature ecology & evolution*. DOI: 10.1038/s41559-018-0617-0.

Rogasch K, Rühmling V, Pané-Farré J, Höper D, Weinberg C, Fuchs S, Schmudde M, Bröker BM, Wolz C, Hecker M, Engelmann S. 2006. Influence of the two-component system SaeRS on global gene expression in two different Staphylococcus aureus strains. Journal of bacteriology 188:7742–7758.

Rose HR, Holzman RS, Altman DR, Smyth DS, Wasserman GA, Kafer JM, Wible M, Mendes RE, Torres VJ, Shopsin B. 2015. Cytotoxic Virulence Predicts Mortality in Nosocomial Pneumonia Due to Methicillin-Resistant Staphylococcus aureus. *The Journal of infectious diseases* 211:1862–1874. DOI: 10.1093/infdis/jiu554.

Rountree PM, Freeman BM. 1955. Infections caused by a particular phage type of Staphylococcus aureus. *The Medical journal of Australia* 42:157–161.

Roux A, Todd DA, Velázquez JV, Cech NB, Sonenshein AL. 2014. CodY-mediated regulation of the Staphylococcus aureus Agr system integrates nutritional and population density signals. Journal of bacteriology 196:1184–1196.

Rudkin JK, Edwards AM, Bowden MG, Brown EL, Pozzi C, Waters EM, Chan WC, Williams P, O'Gara JP, Massey RC. 2012. Methicillin resistance reduces the virulence of healthcare-associated methicillin-resistant Staphylococcus aureus by interfering with the agr quorum sensing system. *The Journal of infectious diseases* 205:798–806.

Rudkin JK, Laabei M, Edwards AM, Joo H-S, Otto M, Lennon KL, O'Gara JP, Waterfield NR, Massey RC. 2014. Oxacillin alters the toxin expression profile of community-associated methicillin-resistant Staphylococcus aureus. *Antimicrobial agents and chemotherapy* 58:1100–1107.

Rudkin JK, McLoughlin RM, Preston A, Massey RC. 2017. Bacterial toxins: Offensive, defensive, or something else altogether? *PLoS pathogens* 13:e1006452. DOI: 10.1371/journal.ppat.1006452.

Runtuwene LR, Tuda JSB, Mongan AE, Makalowski W, Frith MC, Imwong M, Srisutham S, Nguyen Thi LA, Tuan NN, Eshita Y, Maeda R, Yamagishi J, Suzuki Y. 2018. Nanopore sequencing of drug-resistance-associated genes in malaria parasites, Plasmodium falciparum. *Scientific reports* 8:8286.

Ruzin A, Severin A, Moghazeh SL, Etienne J, Bradford PA, Projan SJ, Shlaes DM. 2003. Inactivation of mprF affects vancomycin susceptibility in Staphylococcus aureus. *Biochimica et biophysica acta* 1621:117–121. DOI: 10.1016/s0304-4165(03)00028-x.

Sabat A, Melles DC, Martirosian G, Grundmann H, van Belkum A, Hryniewicz W. 2006. Distribution of the serine-aspartate repeat protein-encoding sdr genes among nasal-carriage and invasive Staphylococcus aureus strains. *Journal of clinical microbiology* 44:1135–1138. DOI: 10.1128/JCM.44.3.1135-1138.2006.

Sadouki Z, Day MR, Doumith M, Chattaway MA, Dallman TJ, Hopkins KL, Elson R, Woodford N, Godbole G, Jenkins C. 2017. Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of Shigella sonnei isolated from cases of diarrhoeal disease in England and Wales, 2015. *The Journal of antimicrobial chemotherapy* 72:2496–2502.

Sakoulas G, Eliopoulos GM, Fowler VG Jr, Moellering RC Jr, Novick RP, Lucindo N, Yeaman MR, Bayer AS. 2005. Reduced susceptibility of Staphylococcus aureus to vancomycin and platelet microbicidal protein correlates with defective autolysis and loss of accessory gene regulator (agr) function. *Antimicrobial agents and chemotherapy* 49:2687–2692. DOI: 10.1128/AAC.49.7.2687-2692.2005.

Sakoulas G, Eliopoulos GM, Moellering RC Jr, Novick RP, Venkataraman L, Wennersten C, DeGirolami PC, Schwaber MJ, Gold HS. 2003. Staphylococcus aureus accessory gene regulator (agr) group II: is there a relationship to the development of intermediate-level glycopeptide resistance? *The Journal of infectious diseases* 187:929–938. DOI: 10.1086/368128.

Sakoulas G, Moellering RC. 2008. Increasing antibiotic resistance among methicillin-resistant Staphylococcus aureus strains. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* 46 Suppl 5:S360–7. DOI: 10.1086/533592.

Sakwinska O, Giddey M, Moreillon M, Morisset D, Waldvogel A, Moreillon P. 2011. Staphylococcus aureus host range and human-bovine host shift. *Applied and environmental microbiology* 77:5908–5915.

Sanchez-Garcia R, Sorzano COS, Carazo JM, Segura J. 2019. BIPSPI: a method for the prediction of partner-specific protein–protein interfaces. *Bioinformatics* 35:470–477. DOI: 10.1093/bioinformatics/bty647.

Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. 2009. Tuberculosis drug resistance mutation database. *PLoS medicine* 6:e2.

Satola SW, Caliendo AM, Farley MM, Patel JB, Burd EM. 2009. Lack of heteroresistance among Staphylococcus aureus isolates with vancomycin MICs of 2 micrograms per milliliter by automated testing. *Journal of clinical microbiology* 47:2680–2681. DOI: 10.1128/JCM.01184-09.

Saund K, Snitkin ES. 2020. hogwash: Three Methods for Genome-Wide Association Studies in

Bacteria. *bioRxiv*:2020.04.19.048421. DOI: 10.1101/2020.04.19.048421.

Sause WE, Balasubramanian D, Irnov I, Copin R, Sullivan MJ, Sommerfield A, Chan R, Dhabaria A, Askenazi M, Ueberheide B, Shopsin B, van Bakel H, Torres VJ. 2019. The purine biosynthesis regulator PurR moonlights as a virulence regulator in Staphylococcus aureus. Proceedings of the National Academy of Sciences of the United States of America 116:13563–13572.

Schmidt K, Mwaigwisya S, Crossman LC, Doumith M, Munroe D, Pires C, Khan AM, Woodford N, Saunders NJ, Wain J, O'Grady J, Livermore DM. 2016. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *The Journal of antimicrobial chemotherapy*. DOI: 10.1093/jac/dkw397.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* 9:671–675.

Schürch AC, van Schaik W. 2017. Challenges and opportunities for whole-genome sequencing–based surveillance of antibiotic resistance. *Annals of the New York Academy of Sciences* 1388:108–120.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. DOI: 10.1093/bioinformatics/btu153.

Seemann T. 2015. Snippy: fast bacterial variant calling from NGS reads.

Shapiro JA. 1984. Observations on the formation of clones containing araB-lacZ cistron fusions. *Molecular & general genetics: MGG* 194:79–90.

Sheppard SK, Didelot X, Meric G, Torralbo A, Jolley KA, Kelly DJ, Bentley SD, Maiden MCJ, Parkhill J, Falush D. 2013. Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in Campylobacter. *Proceedings of the National Academy of Sciences of the United States of America* 110:11923–11927.

Shoji M, Cui L, Iizuka R, Komoto A, Neoh H-M, Watanabe Y, Hishinuma T, Hiramatsu K. 2011. walK and clpP Mutations Confer Reduced Vancomycin Susceptibility in Staphylococcus aureus. *Antimicrobial agents and chemotherapy* 55:3870–3881. DOI: 10.1128/AAC.01563-10.

Sifri CD, Begun J, Ausubel FM, Calderwood SB. 2003. Caenorhabditis elegans as a model host for Staphylococcus aureus pathogenesis. Infection and immunity 71:2208–2217.

Sinclair A, Arnold C, Woodford N. 2003. Rapid detection and estimation by pyrosequencing of 23S rRNA genes with a single nucleotide polymorphism conferring linezolid resistance in Enterococci. *Antimicrobial agents and chemotherapy* 47:3620–3622.

Singh A, Singh S, Singh J, Rahman M, Pathak A, Prasad KN. 2017. Survivability and fitness cost of heterogeneous vancomycin-intermediate Staphylococcus aureus. Indian journal of medical microbiology 35:415–416.

Sitkiewicz I, Babiak I, Hryniewicz W. 2011. Characterization of transcription within sdr region of Staphylococcus aureus. *Antonie van Leeuwenhoek* 99:409–416. DOI: 10.1007/s10482-010-9476-7.

Skwark MJ, Croucher NJ, Puranen S, Chewapreecha C, Pesonen M, Xu YY, Turner P, Harris SR, Beres SB, Musser JM, Parkhill J, Bentley SD, Aurell E, Corander J. 2017. Interacting networks of resistance, virulence and core machinery genes identified by genome-wide epistasis analysis. *PLoS genetics* 13:e1006508.

Smyth DS, McDougal LK, Gran FW, Manoharan A, Enright MC, Song J-H, de Lencastre H, Robinson DA. 2010. Population structure of a hybrid clonal group of methicillin-resistant Staphylococcus aureus, ST239-MRSA-III. *PloS one* 5:e8582.

Solís-Lemus C, Bastide P, Ané C. 2017. PhyloNetworks: A Package for Phylogenetic Networks. *Molecular biology and evolution* 34:3292–3298. DOI: 10.1093/molbev/msx235.

Somerville GA, Cockayne A, Dürr M, Peschel A, Otto M, Musser JM. 2003. Synthesis and deformylation of Staphylococcus aureus delta-toxin are linked to tricarboxylic acid cycle activity. *Journal of bacteriology* 185:6686–6694.

Song L, Florea L, Langmead B. 2014. Lighter: fast and memory-efficient sequencing error correction without counting. Genome biology 15:509.

Soong G, Paulino F, Wachtel S, Parker D, Wickersham M, Zhang D, Brown A, Lauren C, Dowd M, West E, Horst B, Planet P, Prince A. 2015. Methicillin-Resistant Staphylococcus aureus Adaptation to Human Keratinocytes. *mBio* 6. DOI: 10.1128/mBio.00289-15.

Souvorov A, Agarwala R, Lipman DJ. 2018. SKESA: strategic k-mer extension for scrupulous assemblies. Genome biology 19:153.

Sprouffske K, Wagner A. 2016. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC bioinformatics* 17:172. DOI: 10.1186/s12859-016-1016-7.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. DOI: 10.1093/bioinformatics/btu033.

Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo Elias C, Johnson JR, Walker AS, Peto TEA, Crook DW. 2013. Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella pneumoniae isolates using whole genomic sequence data. *The Journal of antimicrobial chemotherapy* 68:2234–2244.

Su M, Lyles JT, Petit RA Iii, Peterson J, Hargita M, Tang H, Solis-Lemus C, Quave CL, Read TD. 2020. Genomic analysis of variability in Delta-toxin levels between Staphylococcus aureus strains. *PeerJ* 8:e8717. DOI: 10.7717/peerj.8717.

Su M, Satola SW, Read TD. 2019. Genome-Based Prediction of Bacterial Antibiotic Resistance. *Journal of clinical microbiology* 57. DOI: 10.1128/JCM.01405-18.

Suligoy CM, Lattar SM, Noto Llana M, González CD, Alvarez LP, Robinson DA, Gómez MI,

Buzzola FR, Sordelli DO. 2018. Mutation of Agr Is Associated with the Adaptation of Staphylococcus aureus to the Host during Chronic Osteomyelitis. *Frontiers in cellular and infection microbiology* 8:18.

Sundqvist M, Geli P, Andersson DI, Sjölund-Karlsson M, Runehagen A, Cars H, Abelson-Storby K, Cars O, Kahlmeter G. 2010. Little evidence for reversibility of trimethoprim resistance after a drastic reduction in trimethoprim use. *The Journal of antimicrobial chemotherapy* 65:350–360.

Swenson JM, Anderson KF, Lonsway DR, Thompson A, McAllister SK, Limbago BM, Carey RB, Tenover FC, Patel JB. 2009. Accuracy of Commercial and Reference Susceptibility Testing Methods for Detecting Vancomycin-Intermediate Staphylococcus aureus. *Journal of clinical microbiology* 47:2013–2017. DOI: 10.1128/JCM.00221-09.

Tange O. 2018. *GNU Parallel 2018*. Lulu.com.

Team RC. 2016. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016. R Foundation for Statistical Computing.

Thiele-Bruhn S. 2003. Pharmaceutical antibiotic compounds in soils – a review. *Journal of Plant Nutrition and Soil Science* 166:145–167. DOI: 10.1002/jpln.200390023.

Thurlow LR, Joshi GS, Clark JR, Spontak JS, Neely CJ, Maile R, Richardson AR. 2013. Functional modularity of the arginine catabolic mobile element contributes to the success of USA300 methicillin-resistant Staphylococcus aureus. Cell host & microbe 13:100–107.

Traber KE, Lee E, Benson S, Corrigan R, Cantera M, Shopsin B, Novick RP. 2008. agr function in clinical Staphylococcus aureus isolates. *Microbiology* 154:2265–2274.

Trindade S, Sousa A, Xavier KB, Dionisio F, Ferreira MG, Gordo I. 2009. Positive epistasis drives the acquisition of multidrug resistance. *PLoS genetics* 5:e1000578.

Tyson GH, McDermott PF, Li C, Chen Y, Tadesse DA, Mukherjee S, Bodeis-Jones S, Kabera C, Gaines SA, Loneragan GH, Edrington TS, Torrence M, Harhay DM, Zhao S. 2015. WGS

accurately predicts antimicrobial resistance in Escherichia coli. *The Journal of antimicrobial chemotherapy* 70:2763–2769.

Uhlemann A-C, Dordel J, Knox JR, Raven KE, Parkhill J, Holden MTG, Peacock SJ, Lowy FD. 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 111:6738–6743.

van Belkum A, Dunne WM Jr. 2013. Next-generation antimicrobial susceptibility testing. *Journal of clinical microbiology* 51:2018–2024.

van der Helm E, Imamovic L, Hashim Ellabaan MM, van Schaik W, Koza A, Sommer MOA. 2017. Rapid resistome mapping using nanopore sequencing. *Nucleic acids research*. DOI: 10.1093/nar/gkw1328.

Van Deun A, Aung KJM, Bola V, Lebeke R, Hossain MA, de Rijk WB, Rigouts L, Gumusboga A, Torrea G, de Jong BC. 2013. Rifampin drug resistance tests for tuberculosis: challenging the gold standard. *Journal of clinical microbiology* 51:2633–2640.

Van Deun A, Barrera L, Bastian I, Fattorini L, Hoffmann H, Kam KM, Rigouts L, Rüsch-Gerdes S, Wright A. 2009. Mycobacterium tuberculosis strains with highly discordant rifampin susceptibility test results. *Journal of clinical microbiology* 47:3501–3506.

van Wamel WJB, Rooijakkers SHM, Ruyken M, van Kessel KPM, van Strijp JAG. 2006. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of Staphylococcus aureus are located on beta-hemolysin-converting bacteriophages. *Journal of bacteriology* 188:1310–1315.

Vaudaux P, Huggler E, Bernard L, Ferry T, Renzoni A, Lew DP. 2010. Underestimation of Vancomycin and Teicoplanin MICs by Broth Microdilution Leads to Under-detection of

Glycopeptide-Intermediate Isolates of Staphylococcus aureus. *Antimicrobial agents and chemotherapy*. DOI: 10.1128/AAC.00269-10.

Vega NM, Gore J. 2017. Stochastic assembly produces heterogeneous communities in the Caenorhabditis elegans intestine. PLOS Biology 15:e2000633. DOI: 10.1371/journal.pbio.2000633.

Vidaillac C, Gardete S, Tewhey R, Sakoulas G, Kaatz GW, Rose WE, Tomasz A, Rybak MJ. 2013. Alternative Mutational Pathways to Intermediate Resistance to Vancomycin in Methicillin-Resistant Staphylococcus aureus. *The Journal of infectious diseases*. DOI: 10.1093/infdis/jit127.

Viedma E, Pérez-Montarelo D, Villa J, Muñoz-Gallego I, Larrosa N, Fernández-Hidalgo N, Gavaldà J, Almirante B, Chaves F. 2018. Sub-inhibitory concentrations of oxacillin modify the expression of agr locus in Staphylococcus aureus clinical strains belonging to different clonal complexes. *BMC infectious diseases* 18:177.

Vogwill T, Kojadinovic M, MacLean RC. 2016. Epistasis between antibiotic resistance mutations and genetic background shape the fitness effect of resistance across species of Pseudomonas. *Proc. R. Soc. B* 283:20160151.

Vogwill T, MacLean RC. 2015. The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. *Evolutionary applications* 8:284–295.

Votintseva AA, Bradley P, Pankhurst L, Del Ojo Elias C, Loose M, Nilgiriwala K, Chatterjee A, Smith EG, Sanderson N, Walker TM, Morgan MR, Wyllie DH, Walker AS, Peto TEA, Crook DW, Iqbal Z. 2017. Same-day diagnostic and surveillance data for tuberculosis via whole genome sequencing of direct respiratory samples. *Journal of clinical microbiology*. DOI: 10.1128/JCM.02483-16.

Vuong C, Saenz HL, Götz F, Otto M. 2000. Impact of the agr quorum-sensing system on adherence to polystyrene in Staphylococcus aureus. *The Journal of infectious diseases* 182:1688–1693.

Wadsworth CB, Arnold BJ, Sater MRA, Grad YH. 2018. Azithromycin resistance through interspecific acquisition of an epistasis dependent efflux pump component and transcriptional regulator in Neisseria gonorrhoeae. *bioRxiv*:309294. DOI: 10.1101/309294.

Wahl LM, Gerrish PJ, Saika-Voivod I. 2002. Evaluating the impact of population bottlenecks in experimental evolution. Genetics 162:961–971.

Wales AD, Davies RH. 2015. Co-Selection of Resistance to Antibiotics, Biocides and Heavy Metals, and Its Relevance to Foodborne Pathogens. *Antibiotics (Basel, Switzerland)* 4:567–604.

Walsh IM, Bowman MA, Clark PL. 2019. Synonymous codon substitutions perturb co-translational protein folding and significantly impair cell fitness. *bioRxiv*:666552. DOI: 10.1101/666552.

Wang H, Liu C, Deng L. 2018. Enhanced Prediction of Hot Spots at Protein-Protein Interfaces Using Extreme Gradient Boosting. *Scientific reports* 8:14285. DOI: 10.1038/s41598-018-32511-1.

Wang R, Braughton KR, Kretschmer D, Bach T-HL, Queck SY, Li M, Kennedy AD, Dorward DW, Klebanoff SJ, Peschel A, DeLeo FR, Otto M. 2007. Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA. *Nature medicine* 13:1510–1514. DOI: 10.1038/nm1656.

Wang Y, Li X, Jiang L, Han W, Xie X, Jin Y, He X, Wu R. 2016. Novel Mutation Sites in the Development of Vancomycin- Intermediate Resistance in Staphylococcus aureus. *Frontiers in microbiology* 7:2163. DOI: 10.3389/fmicb.2016.02163.

Watanabe Y, Cui L, Katayama Y, Kozue K, Hiramatsu K. 2011. Impact of rpoB Mutations on Reduced Vancomycin Susceptibility in Staphylococcus aureus. *Journal of clinical microbiology* 49:2680–2684. DOI: 10.1128/JCM.02144-10.

Weigel LM, Clewell DB, Gill SR, Clark NC, McDougal LK, Flannagan SE, Kolonay JF, Shetty J, Killgore GE, Tenover FC. 2003. Genetic analysis of a high-level vancomycin-resistant isolate of Staphylococcus aureus. *Science* 302:1569–1571.

Wheeler NE, Barquist L, Kingsley RA, Gardner PP. 2016. A profile-based method for identifying functional divergence of orthologous genes in bacterial genomes. *Bioinformatics* 32:3566–3574.

Wichelhaus TA, Böddinghaus B, Besier S, Schäfer V, Brade V, Ludwig A. 2002. Biological cost of rifampin resistance from the perspective of Staphylococcus aureus. *Antimicrobial agents and chemotherapy* 46:3381–3385.

Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS computational biology* 13:e1005595.

Williams H, Flohr C. 2006. How epidemiology has challenged 3 prevailing concepts about atopic dermatitis. *The Journal of allergy and clinical immunology* 118:209–213. DOI: 10.1016/j.jaci.2006.04.043.

Wistrand-Yuen E, Knopp M, Hjort K, Koskiniemi S, Berg OG, Andersson DI. 2018. Evolution of high-level resistance during low-level antibiotic exposure. *Nature communications* 9:1599.

Wright S. 1932. *The roles of mutation, inbreeding, crossbreeding, and selection in evolution*. na.

Wu K, Conly J, McClure J-A, Elsayed S, Louie T, Zhang K. 2010. Caenorhabditis elegans as a host model for community-associated methicillin-resistant Staphylococcus aureus. Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases 16:245–254.

Xavier BB, Das AJ, Cochrane G, De Ganck S, Kumar-Singh S, Aarestrup FM, Goossens H, Malhotra-Kumar S. 2016. Consolidating and Exploring Antibiotic Resistance Gene Data Resources. *Journal of clinical microbiology* 54:851–859.

Xu SX, McCormick JK. 2012. Staphylococcal superantigens in colonization and disease. *Frontiers in cellular and infection microbiology* 2:52. DOI: 10.3389/fcimb.2012.00052.

Yang Y, Niehaus KE, Walker TM, Iqbal Z, Walker AS, Wilson DJ, Peto TE, Crook DW, Smith EG, Zhu T, Clifton DA. 2017. Machine Learning for Classifying Tuberculosis Drug-Resistance from

DNA Sequencing Data. *Bioinformatics* . DOI: 10.1093/bioinformatics/btx801.

Yates CM, Shaw DJ, Roe AJ, Woolhouse MEJ, Amyes SGB. 2006. Enhancement of bacterial competitive fitness by apramycin resistance plasmids from non-pathogenicEscherichia coli. *Biology Letters* 2:463–465. DOI: 10.1098/rsbl.2006.0478.

Yeung M, Balma-Mena A, Shear N, Simor A, Pope E, Walsh S, McGavin MJ. 2011. Identification of major clonal complexes and toxin producing strains among Staphylococcus aureus associated with atopic dermatitis. *Microbes and infection / Institut Pasteur* 13:189–197. DOI: 10.1016/j.micinf.2010.10.023.

Yoo JI, Kim JW, Kang GS, Kim HS, Yoo JS, Lee YS. 2013. Prevalence of amino acid changes in the yvqF, vraSR, graSR, and tcaRAB genes from vancomycin intermediate resistant Staphylococcus aureus. *The Journal of Microbiology* 51:160–165. DOI: 10.1007/s12275-013-3088-7.

Young BC, Golubchik T, Batty EM, Fung R, Larner-Svensson H, Votintseva AA, Miller RR, Godwin H, Knox K, Everitt RG, Iqbal Z, Rimmer AJ, Cule M, Ip CLC, Didelot X, Harding RM, Donnelly P, Peto TE, Crook DW, Bowden R, Wilson DJ. 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* 109:4550–4555.

Young BC, Wu C-H, Claire Gordon N, Cole K, Price JR, Liu E, Sheppard A, Perera S, Charlesworth J, Golubchik T, Iqbal Z, Bowden R, Massey R, Paul J, Crook DW, Peto TEA, Sarah Walker A, Llewelyn M, Wyllie DH, Wilson D. 2017. Severe infections emerge from the microbiome by adaptive evolution. *bioRxiv*:116681. DOI: 10.1101/116681.

Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM. 2017. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *The Journal of antimicrobial chemotherapy*. DOI: 10.1093/jac/dkx217.

Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. *The Journal of antimicrobial chemotherapy* 67:2640–2644.

Zeden MS, Kviatkovski I, Schuster CF, Thomas VC, Fey PD, Gründling A. Identification of the main glutamine and glutamate transporters in Staphylococcus aureus and their impact on c-di-AMP production. DOI: 10.1101/754309.

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:e0136082. DOI: 10.1371/journal.pone.0136082.

Zhang Y, Mitchison D. 2003. The curious characteristics of pyrazinamide: a review. *The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease* 7:6–21.

Zhao S, Tyson GH, Chen Y, Li C, Mukherjee S, Young S, Lam C, Folster JP, Whichard JM, McDermott PF. 2015. Whole-Genome Sequencing Analysis Accurately Predicts Antimicrobial Resistance Phenotypes in Campylobacter spp. *Applied and environmental microbiology* 82:459–466.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature genetics* 44:821–824. DOI: 10.1038/ng.2310.