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:

_____ _____

Abraham Moller Date

Understanding phage-host interactions in *Staphylococcus aureus* through population genomics and bioinformatics

By

Abraham Moller
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Microbiology and Molecular Genetics

Timothy Read, Ph.D.
Advisor

Nicole Gerardo, Ph.D.
Committee Member

Bruce Levin, Ph.D.
Committee Member

Phil Rather, Ph.D.
Committee Member

Yih-Ling Tzeng, Ph.D.
Committee Member

Accepted:

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

Date

Understanding phage-host interactions in *Staphylococcus aureus* through population genomics and bioinformatics

By

Abraham Moller
B.S., B.A., M.S., Miami University (Ohio), 2016

Advisor: Timothy 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

2021

Abstract

Understanding phage-host interactions in *Staphylococcus aureus* through population genomics and bioinformatics

By

Abraham Moller

This dissertation seeks to understand the host basis of phage host range in *Staphylococcus aureus* with the ultimate goal of improving phage therapy against *S. aureus* infections. Increasing antibiotic resistance, high prevalence, and failure to develop vaccines makes alternative *S. aureus* therapies such as phage therapy crucial. However, while *S. aureus* phages have long been known to have broad host ranges, the mechanism behind the resistant exceptions remains unknown. This paradox justified comprehensive, population-wide studies to understand determinants of phage host range in *S. aureus* on a species rather than strain-wide level. In my dissertation, I not only reviewed the literature to identify what is known about host range and resistance factors in the species but also used population genomics to discover new host range factors and examine the evolution of known factors in the species. I found that host resistance factors have been identified and characterized at three stages of the infection cycle (adsorption, biosynthesis, and assembly) but neither uptake nor lysis stages. I also hypothesized that these factors determine host range in a phylogenetically hierarchical manner given their respective conservation in the species. GWAS identified novel core genes involved in host range and known host range determinants but significant determinants only partially explained the phenotype in predictive modeling. Molecular genetic techniques (backcrossing and complementation) confirmed a subset of novel determinants did have causal associations with the phenotype. Regarding phage resistance evolution, I found adsorption genes to be the most conserved in the database, core phage resistance genes functional, diverse, and under purifying selection like core genes overall. Only superinfection immunity correlated with temperate phage resistance or accessory genome content, but overall phage resistance gene presence never correlated with virulent phage resistance. All genes exhibited some level of phylogenetic signal, but it was weakest amongst the assembly genes. I also developed a sequence-based assay with the future potential to rapidly determine phage susceptibility of clinical isolates. I found that the limit of detection was close to minor allele frequency average (~5-7%), I could detect VISA-causing mutations in clinical strains, and I could discriminate clinical VSSA from VISA strains.

Understanding phage-host interactions in *Staphylococcus aureus* through population genomics and bioinformatics

By

Abraham Moller

Advisor: Timothy 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

2021

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

p003p, p0040, and pyo).

Supplemental Figure S3: GWAS approach and significant SNP annotations.

Supplemental Figure S4: Growth curves of USA300, USA300 transposon mutants, transposon mutants electroporated with the empty pOS1 vector, and transposon mutants complemented with vectors containing respective genes.

Supplemental Figure S5: Bacterial survival after completing the high-throughput host range assay (p003p against trpA strains).

Supplemental Figure S6: High-throughput host range assay phenotypes demonstrating genetic validation of novel GWAS phage host range determinants.

Supplemental Figure S7: Efficiency of plating (EOP) phenotypes demonstrating genetic validation of phage host range determinants.

Supplemental Figure S8: Construction of neural network predictive models for each ternary phage resistance phenotype.

Supplemental Figure S9: Evaluation of ternary phage resistance phenotype predictive models through receiver operating characteristic-area under the curve.

Supplemental Figure S10: Evaluation of ternary phage resistance phenotype neural network predictive models through receiver operating characteristic-area under the curve.

Supplemental Table S1: Diversity of the phages used in this study.

Supplemental Table S2: Excel spreadsheet including tested *S. aureus* strain names, quantitative phenotypes, qualitative phenotypes (quantitative phenotypes from Supplemental Table S1 converted to qualitative by the OD600 scale 0.1-0.4 for S - sensitive, 0.4-0.7 for SS - semi-sensitive, and 0.7 or more for R - resistant), BioProject, BioSample, and SRA accessions, sequence types (STs), clonal complexes (CCs), isolation dates, and isolation locations.

Supplemental Table S3: Primers used to amplify wild-type genes corresponding to transposon mutants and clone them into pOS1-P*lgt* with splicing overlap extension (SOE)-PCR or HiFi/Gibson assembly.

Supplemental Table S4: Significant differences ($p<0.05$; p-values listed in parentheses) in phage host range phenotypes between tested strains' CCs based on Tukey HSD/one-way ANOVA tests.

Supplemental Table S5: Summary statistics for protein-protein interaction networks identified with STRING amongst genes corresponding to significant SNPs or k-mers (inside or adjacent to genes).

Supplemental Table S6: Functions enriched amongst gene sets analyzed with STRING databases.

Supplemental Table S7: Functions enriched amongst gene sets analyzed with PANTHER databases.

Chapter 4

Figure 1: Conservation of examined phage resistance genes in the species based on a search of 40,000+ *S. aureus* genomes.

Figure 2: Core and extended core phage resistance genes do not differ from core and extended core genes overall in terms of diversity, functionality, or selection.

Figure 3: All non-extended core phage resistance genes have phylogenetic signal, but assembly genes have the least phylogenetic signal amongst all.

Figure 4: Relationship between clonal complex (CC) and accessory genome and non-extended core phage resistance gene content.

Figure 5: Non-extended core phage resistance genes are encoded together based on calculating genomic overlaps from a search of 40,000+ *S. aureus* genomes.

Figure 6: Superinfection immunity but neither adsorption nor assembly genes correlates with empirical temperate phage resistance.

Figure 7: Superinfection immunity but neither adsorption nor assembly genes correlates with accessory genome content.

Figure 8: Superinfection immunity does not correlate with non-extended core adsorption, non-superinfection immunity biosynthesis, or assembly gene counts.

Figure 9: Correlations between non-extended core antibiotic resistance or virulence genes and non-extended core phage resistance genes.

Figure 10: Phylogenetic overlap between classes of non-extended core genes determined from a search of 40,000+ *S. aureus* genomes.

Table 1: List of curated phage resistance genes with strains, genome coordinates, accessions, classes (e.g., adsorption), and subclasses (e.g., receptor).

Chapter 5

Figure 1: Overview of the VISA amplicon sequencing process, from sample isolation to SNP calling and comparison to known VISA mutations.

Figure 2: Evidence that our 10 VISA regions can be amplified from N384-3 mutant DNA extracted from pure culture.

Figure 3: Gene coverage analyses for simulation, DNA, and cell parent/mutant mixtures.

Figure 4: Evaluating limit of detection over a range of mutant/parent mixes (cultures, amplicon DNA, and simulated sequence reads).

Table 1: Strains used in this study.

Table 2: Primers used for amplifying each of 10 regions containing genes likely to contain VISA-conferring mutations.

Table 3: Mutation calls (whether walK 645C>G was called) by bwa/bcftools and medaka for simulated, amplicon DNA, and culture mixtures.

Table 4: Mutation calls by bwa/bcftools and medaka for test strains EUH15 and 107.

Chapter 1: Introduction

Staphylococcus aureus

The opportunistic Gram-positive bacterial pathogen *Staphylococcus aureus* is a substantial public health burden. It colonizes the skin, nasal passages, mucous membranes, and oral cavity of between 20 and 30% of the world's population persistently and up to 70% intermittently (1). It causes illnesses ranging from abscesses and other skin and soft tissue infections (SSTIs) to food poisoning, toxic shock syndrome, osteomyelitis, endocarditis, and septicemia. The troubling rise of antibiotic resistance has worsened mortality, with the CDC reporting 11,285 deaths due to MRSA in 2013 (2). A combined rise in antibiotic resistance (3), failure to develop a vaccine (4), and lack of alternative therapies makes developing new *S. aureus* therapeutics urgent. In this thesis, I present my work aiming to improve development of at least one class of therapeutics, phages, by improving our understanding of the bacterial determinants of *S. aureus* phage killing specificity, or host range.

History - initial discovery and naming

The bacterial species *S. aureus* was first designated by the German scientist Anton Friedrich Julius Rosenbach in 1884. Rosenbach separated white and yellow colonies isolated from aseptically collected pus (5), naming the respective bacterial species *Staphylococcus albus* (white; now *Staphylococcus epidermidis*) and *Staphylococcus aureus* (golden) (6). Previously, in 1881, the Scottish surgeon Sir Alexander Ogston had identified *Staphylococcus* bacteria by their cellular morphology ("Staphylo" bunches of grapes; "coccus" berries) (7). Ogston observed the bacteria in pus from a surgical abscess in a patient's knee joint (7) which he stained and examined under a microscope (8). It was thus from these initial discoveries of colony and cellular morphology, respectively, that the organism was named.

Colony morphology, diagnostic characteristics, and cellular structure

S. aureus is a non-spore forming, non-motile bacterium that grows in groups of slightly-oblong spherical cocci and forms characteristic golden colonies on agar growth medium (9). While not actively motile, it does undergo a form of passive motility called spreading, in which surfactants break surface tension that holds cells to each other and the colony surface, causing cells to fall outwards as the colony grows (10, 11). It is a facultative anaerobe that grows by aerobic respiration or fermentation, but it also forms small colony variants (SCVs) that only metabolize by fermentation (12). Unlike *S. epidermidis*, it is often but not always beta-hemolytic on blood agar (13). It is both salt tolerant and ferments mannitol, justifying the use of mannitol salt agar (MSA) for its detection. Mannitol salt agar is a selective and differential medium for *S. aureus* that contains mannitol, phenol red, and 7.5% NaCl (14). The high salt concentration selects for *Staphylococci*, while the phenol red indicator changes from red to yellow upon the pH decrease that coincides with mannitol fermentation, thus differentiating the mannitol-fermenting *S. aureus*. *S. aureus* can further be distinguished by catalase and coagulase tests (9). *S. aureus* produces a catalase that converts hydrogen peroxide to water and oxygen, leading to oxygen bubbles forming upon addition of 3% hydrogen peroxide solution to solid or broth culture in a lab test. It also produces a coagulase that converts fibrinogen into fibrin, causing *S. aureus* culture to coagulate plasma. This allows differentiation from Coagulase-Negative staphylococci (CoNS) such as *S. epidermidis*.

S. aureus has a single membrane surrounded by a thick cell wall. Like other Gram-positive cell walls, the *S. aureus* cell wall consists of multiple cross-linked peptidoglycan layers. It is estimated by cryo-EM to be 35 nm thick, consisting of a 16 nm low density inner zone followed by a 19 nm high density outer zone (15). The peptidoglycan consists of N-acetylmuramic acid (NAM)-N-acetylglucosamine (NAG) polymers cross-linked by short peptides connected through pentaglycine bridges (16). The cell wall is also embedded with surface proteins, wall and lipoteichoic acid polysaccharides that weave through the peptidoglycan, and (sometimes)

capsular polysaccharide outside the cell wall (Figure 1A). Proteins with C-terminal LPXTG amino-acid motifs are attached to peptidoglycan through sortase proteins (17). Wall teichoic acid (WTA) is a cell-wall (NAM) anchored chain polysaccharide that consists of N-acetylglucosamine, N-acetylmannosamine, and two glycerol phosphate monomers followed by 40-60 repeats of ribitol phosphate (Figure 1A and B) (18). Lipoteichoic acid (LTA), on the other hand, is a lipid anchored chain polysaccharide that consists of two glucose monomers followed by many glycerol phosphate repeats (18). LTA and WTA can be modified with D-alanine and N-acetylglucosamine (19, 20). Capsular polysaccharide is also attached to N-acetylmuramic acid and has been categorized into four types (1, 2, 5, and 8) based on serological properties (21). Type 1 and 2 strains are mucoid and heavily encapsulated (22), while type 5 and 8 strains are encapsulated but not mucoid. Type 1 and 2 capsule consists of (→4)-α-D-GalNAcA-(1→4) -α-D-GalNAcA-(1→3)-α-D-FucNAc-(1→) and (→4)-β-D-GlcNAcA-(1→4)-β-D-GlcNAcA-(L-alanyl) -(1→) monomers, respectively (21). Type 5 and 8 capsule consists of (→4)-3-O-Ac- β-D-ManNAcA-(1→4)-α-L-FucNAc-(1→3)-β-D-FucNAc-(1→) monomers and (→3)-4-O-Ac-β-D -ManNAcA-(1→3)-α-L-FucNAc-(1→3)-β-D-FucNAc-(1→) monomers, respectively (21). Type 5 and 8 monomers only differ in glycosidic linkages between sugars and O-acetylation positions on D-ManNAcA.

Typing and known diversity

Early *S. aureus* genotyping schemes depended on differential sensitivity to *S. aureus*-specific viruses (bacteriophages). Currently, multilocus sequence typing (MLST) and clonal complexes (CC) distinguish *S. aureus* strains below the species level. *S. aureus* multilocus sequence typing (MLST) depends on sequencing seven housekeeping gene markers (*arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, and *yqiL*) (23). Some 6,583 *S. aureus* sequence types have been classified (24). A clonal complex (CC) is a group of strains in which each strain's sequence type differs from another only by an allele at a single gene. Clonal complexes are thought to explain the diversity

of *S. aureus*, with clones displacing each other in different geographic areas (25). Amongst MRSA strains, ST22, 30, 36, 80, 239, and 398 are most common in Europe; ST5 and 8 in North America; ST5 in South America; and ST5, ST8, ST59, ST72, and ST239 in Asia; while ST93 and 121 are most common in Australia and ST5 in Africa. Healthcare-associated (HA)-MRSA, community-associated (CA)-MRSA, and livestock- associated (LA)-MRSA most commonly belong to CCs 22, 30, or 8 (ST239), 5 or 8, and 398, respectively (26).

Horizontal gene transfer in the species

While *S. aureus* reproduces asexually like other bacterial species, it can also transfer DNA from strain to strain via three well-known mechanisms of horizontal gene transfer (HGT): transformation, conjugation, and transduction (27). HGT is responsible for the spread of virulence factors and antibiotic resistance genes through the population.

Natural DNA transformation, while shown to be possible in the laboratory, is probably extremely rare in nature. Molecular genetic studies indicated *S. aureus* transformation requires duplication of the sigma factor σ^H and even then occurs at rates of 1e-9 transformants/cfu recipient (28). Conjugation, in which DNA is transferred from donor to recipient during contact to form a mating pair, is thought to be the likely source of vancomycin resistance genes through transfer from *Enterococcus faecalis* (29–31). While most plasmids are carried by phages (transduction), a subset moves from strain to strain via conjugation (32), including the vancomycin resistance plasmid pBRZ01 (33). Chromosomal DNA elements called integrative conjugative elements (ICEs), such as Tn*916*, Tn*5801*, and ICE*6013* (34), also move via conjugation and carry genes to evade restriction systems in the case of Tn*5801* (35).

Transduction in the species occurs primarily through three mechanisms - generalized transduction (27), island-mediated generalized transduction (36), and the recently discovered, *S. aureus*-specific lateral transduction (37). Transducing phages are a subset of the temperate *Siphoviridae* known to infect the species (38, 39). Generalized transduction occurs when phages

take up genomic or plasmid DNA and undergo a second abortive infection to deliver their cargo into a new host (38), where the transferred DNA either exists freely as a plasmid or recombines with the homologous section of the host chromosome. A subset of generalized transduction called autotransduction was also recently discovered in *S. aureus* in which lysogens release prophages that go on to infect neighboring phage-sensitive cells and transfer new antibiotic resistance genes back to the original donor (40). Island-mediated generalized transduction is known to occur during the transfer of *S. aureus* pathogenicity islands (SaPIs). SaPIs only excise from the chromosome upon infection with a respective, derepressing helper phage. The excised SaPIs then abortively infect a new host and deliver their cargo. Lateral transduction is the transfer of DNA from genome to genome downstream of an integrated prophage at higher rates (1e6 transducing units/mL) than average across the genome (1e3 transducing units/mL) (37). This phenomenon happens because prophages excise after DNA replication begins, giving time to package amplified DNA downstream of the phage origin of replication. Up to seven successive headfuls of DNA downstream of the prophage are cleaved at *pac*-like sites and packaged into phage particles along with phage and host genomic DNA (37).

Antibiotic resistance

Some *S. aureus* strains have become resistant to antibiotics of the most common classes (cell envelope disruptors, cell wall biosynthesis, nucleic acid biosynthesis, DNA replication, transcription, and translation inhibitors) with resistance gained through both gene acquisition and mutation. Resistance mechanisms to each class have been previously reviewed (3, 41). Cell wall biosynthesis inhibitors are perhaps the most famous antibiotics used against *S. aureus* due to their early and widespread development of clinical resistance. Methicillin resistance (MRSA) was famously first discovered clinically in 1961 at the Staphylococcal Reference Laboratory in Colindale, England (42, 43). MRSA occurs when strains acquire the staphylococcal cassette chromosome-mec (SCC*mec*) which encodes the *mecA* (*pbp2a*) gene (26). This gene encodes an

altered version of the PBP2 transpeptidase that does not bind methicillin. General beta-lactam resistance also occurs through the secreted, Tn552-encoded lipoprotein beta-lactamase *blaZ* (3, 44).

Full vancomycin resistance (VRSA; MIC of 16 µg/mL or higher) is quite rare and occurs due to transmission of a resistance gene from *Enterococcus faecalis*, most likely through conjugation (29–31). Substitution of D-Ala-D-Ala, the vancomycin target, with D-Ala-D-lactate synthesized by VanA causes full vancomycin resistance (45). Vancomycin-intermediate resistance (VISA), marked by an MIC between 4 and 8 µg/mL, on the other hand, is a far more common phenotype, first described in Japan in 1997 (17). VISA is a polygenic phenotype depending on changes in cell wall turnover (combination of synthesis and degradation). Increased net cell wall accumulation due to activation of the cell wall stimulon (GraSR, WalKR, and VraSR) results in increased free unlinked D-Ala-D-Ala that absorbs vancomycin, reducing its diffusion through the cell wall (46). The antibiotic daptomycin, on the other hand, is a cationic antimicrobial peptide that acts on the *S. aureus* cell membrane itself in complex with calcium ions. Daptomycin resistance occurs due to altered cell membrane charge. *mprF* lysylates phosphatidylglycerol, increasing positive charge on the cell surface and repelling daptomycin from its target. The *dlt* operon in addition is responsible for transferring positively charged alanines to wall and lipoteichoic acids, also repelling daptomycin (19, 47–49).

Nucleic acid biosynthesis inhibitors used against *S. aureus* include sulfonamides and trimethoprim, which inhibit precursor folic acid biosynthesis (50). Chromosomal dihydropteroate synthase (DHPS) mutations preventing sulfamethoxazole binding and chromosomal or plasmid-borne dihydrofolate reductase mutations preventing trimethoprim binding cause resistance to each respective antibiotic class (3). DNA replication and transcription inhibitors used against *S. aureus* include fluoroquinolones and rifampin, respectively. Fluoroquinolone resistance occurs due to DNA topoisomerase (*parC*) target site mutations and efflux pump expression (*norA*, *norB*, and *norC*). Rifampin resistance mutations occur due to mutations in a RNA polymerase gene.

These mutations in *S. aureus* map to a 87 amino acid segment (463-550) of the RNA polymerase beta-subunit (*rpoB*) (51, 52) and reduce rifampin affinity to the target site.

Translation inhibitors used against *S. aureus* act on the 30S subunit, 50S subunit, or other protein/RNA translation components (3). Those in clinical use include mupirocin, lincosamides, streptogramins, fusidic acid, linezolid, gentamicin, and tigecycline. Inhibitors acting on the 30S subunit include tetracyclines and aminoglycosides. Tetracyclines cause amino acid-tRNAs to fall off the A site, while aminoglycosides cause mRNA misreading. Those acting on the 50S subunit include linezolid, chloramphenicol, lincosamides, macrolides, and streptogramins. These compounds block the polypeptide exit tunnel or interfere with aminoacyl-tRNA binding at the peptidyl transferase center. Other translation inhibitors used on *S. aureus* include isoleucine-tRNA synthetase inhibitor mupirocin and elongation factor G inhibitor fusidic acid. Chromosomal mutations conferring resistance against these inhibitors occur in 23S rRNA or L3/L4 (linezolid; reducing target binding), rRNA or upstream of efflux pump *mepA* (tigecycline; reducing target binding and increasing efflux, respectively), *fusA* (fusidic acid; reducing target binding), or *ileRS* (mupirocin; reducing target binding). Horizontally transferred genetic determinants conferring such resistance include target modification (*erm* - erythromycin, *cfr* - linezolid), target protection (*vga*, *lsa*, *sal* - streptogramin A, *msr* - streptogramin B, *optrA* - linezolid ; *fusB* - fusidic acid; *tetO/M* - tetracycline), efflux (*tet* - tetracycline and *fex* - florfenicol), drug modification (*aac*, *aph*, *aad* - aminoglycosides, *vat* - streptogramin A, *vgb* - streptogramin B, *cat* - chloramphenicol), or encoding an alternative target (*mupA* - mupirocin)

Bacteriophages

(Note - some text in this section has been adapted from my recently published review article (53), which is chapter 2 of this dissertation)

Classes, morphologies, genome structures, and possible life cycles

Bacteriophages, (greek for “bacteria eaters”) are viruses that infect bacteria. They are the most abundant living organisms on Earth (1e31 total) and are highly diverse (54). Their genomes can be either single- or double-stranded, circular or linear, DNA or RNA, and even some phages

have multi-segmented dsRNA genomes (*Pseudomonas* phage phi-6). They can be enveloped (*Cystoviridae*), nonenveloped (*Microviridae*), filamentous (*Inoviridae*), or tailed (*Caudovirales*), or non-tailed (*Leviviridae*) (55). ssDNA phages include the circular genome, icosahedral *Microviridae* and the circular genome, filamentous *Inoviridae*. dsDNA phages include the linear genome, tailed *Caudovirales* (*Myoviridae*, *Siphoviridae*, and *Podoviridae*), the circular genome, enveloped, double-encapsidated, icosahedral *Corticoviridae*, linear genome, icosahedral *Tectiviridae*, and the circular genome, enveloped, pleomorphic *Plasmaviridae* and *Fuselloviridae*. ssRNA phages include the linear genome, icosahedral *Leviviridae*. dsRNA phages include the segmented genome, enveloped *Cystoviridae*, the only class of phages known to have segmented genomes (55). Phages can also be temperate, in which they either infect and kill a cell (lysis) or integrate into the chromosome and remain dormant (lysogeny), or virulent, in which they solely kill cells they infect. This section will focus on life cycle mechanisms of the *Caudovirales,* as they are the taxon associated with *S. aureus* and they are the best characterized from extensive Gram-negative studies, along with host resistance that emerges at each stage.

Life cycles - lytic and lysogenic

Adsorption

In the first stage of phage infection, viruses must bind specifically to the surface of bacteria using receptor-binding proteins (RBPs) exposed on their tail fibers or baseplates. Adsorption to carbohydrate or protein receptors involves reversible binding followed by irreversible attachment to the final receptor (56). Random reversible binding to cell wall carbohydrate components keeps phages in close proximity to the cell surface to increase the chances of binding the protein receptor irreversibly. *E. coli* phage T5, for example, binds reversibly to LPS O-antigen with its tail fibers and irreversibly to outer membrane protein FhuA with tail protein pb5 (57, 58). Receptors in the Gram-negatives include LPS, capsule, efflux pumps, and other outer membrane proteins.

Receptors in the Gram-positives include wall teichoic acid (WTA), phage infection proteins (Pips), and cell wall peptidoglycan.

Resistance at the adsorption stage occurs through receptor alteration, occlusion, removal, or competitive inhibition (59). Protein receptors can be deleted or mutated at regions that bind to the phage receptor binding protein, while carbohydrate receptors can be eliminated or undergo further modification. Either thick capsular polysaccharide, extracellular biofilm polysaccharides, or increased surface protein content can prevent receptor binding. Outer membrane vesicles (OMVs) containing phage receptor (60) or other proteins binding the phage receptor (e.g., microcins binding FhuA in *E. coli*) (59) can competitively inhibit phages binding receptors on the bacterial cell surface.

Uptake

After adsorption, the phage DNA must be injected into the cell. First either the cell wall or capsular polysaccharide is degraded by phage baseplate enzymes, such as peptidoglycan hydrolases (61, 62) and capsule depolymerases (63) like poly-glutamate depolymerase in *B. subtilis* phages (64–66). The sheath then contracts and the tube protrudes into the cell (in the case of *Myoviridae*) and DNA is injected into the cell. DNA enters the cell due to 1) the pressure of DNA packed into the capsid forcing it out and 2) DNA-binding proteins cell drawing the DNA inside (67).

Resistance at the uptake stage occurs through superinfection exclusion. This phenomenon occurs when a temperate phage has integrated into the host chromosome. The resulting prophage prevents DNA uptake from an identical infecting phage by expressing superinfection exclusion (Sie) proteins (68) that interact with the phage tape measure protein to prevent channel formation and DNA injection.

Biosynthesis

After phage DNA injection into the cell, the phage genome must be replicated (or integrated, in the case of lysogeny), transcribed, and translated into phage proteins to assemble

new particles. Phage DNA genome replication mechanisms vary but must overcome the loss of genetic material near the 5’ end of the genome in the case of linear dsDNA phages like the *Caudovirales* (69). Such phages have several ways of solving this problem after uptake and before replication - 1) protein primer covalently attached to each end of the genome during and after replication, 2) direct terminal repeats reformed after concatemer processing during phage packaging, 3) protelomerase to reform covalently closed ends of the linear dsDNA genome after replication of a circular intermediate, 4) conversion of a linear phage genome to a circular form after uptake (69). Phage DNA then replicates by either rolling circle or theta mechanisms (69). During phage genome transcription and protein synthesis, the phage must redirect host cell transcription and translation machinery away from the host and towards the propagating phage.

Resistance at the biosynthesis stage includes mechanisms that restrict phage DNA replication, transcription, and translation. Such mechanisms include restriction-modification, abortive infection, superinfection immunity, and CRISPR-Cas. Restriction-modification systems destroy phage DNA that has not been methylated appropriately by cognate modification enzymes, thus restricting phage infection and distinguishing self from nonself DNA (70). Abortive infection (Abi) systems halt phage propagation by inhibiting DNA replication, transcription, and translation upon detection of phage proteins (71). Superinfection immunity is the process that occurs when temperate phages lysogenize a host cell and repress the lytic cycle genes of subsequently infecting identical phage, in effect halting biosynthesis of these phages (72, 73). Clustered regularly interspaced short palindromic repeats (CRISPR)-Cas systems form arrays of short spacer sequences acquired from infecting phages (74–76). These spacers form a memory of past infections and consequently initiate DNA and RNA cleavage of infecting phages and their transcripts.

Assembly

Phages in general vary greatly in overall morphology, ranging from filamentous to tailed, and this has consequences for phage assembly. In filamentous phages, single stranded DNA or

RNA is coated helically with capsid proteins. Indeed, in tailed phages, both the head and tail must be formed and the head (or filament) packaged with the genetic material. The head capsid is icosahedral with cubic symmetry and is composed of scaffolding, capsid, neck, and decoration proteins. After a spherical head forms, it is filled with DNA, matured into the final icosahedral structure, and attached to the neck and assembled tail (77). Tail shafts can either be flexible and noncontractile in the case of *Siphoviridae* or inflexible and contractile (correspondingly sheathed) in the case of *Myoviridae*. dsDNA fragments are packaged via terminases that cut successive headfuls of DNA from a replicating phage genome concatemer as they are driven into assembled heads. Either *pac* or *cos* sites determine terminase cleavage specificity. In the case of *pac* phages, concatemers are cleaved before individual *pac* sites once heads are filled, whereas in *cos* phages, concatemers are cleaved at dual, subsequent cohesive ends to exact unit lengths (78).

Resistance at the assembly stage occurs through the process of assembly interference by parasitic satellite phages. Phage-inducible chromosomal islands (PICIs), which have mainly been discovered in Gram-positives, become lytic upon induction by helper phages (79). The PICIs in turn assemble phage particles at the cost of the helper phage, altering the capsid to fit the smaller PICI genome, blocking the terminase protein of the helper phage with phage packaging interference (Ppi) proteins, and repressing helper phage late genes.

Lysis

After phage particles have assembled, the cell must be lysed to release them and let them initiate new infections. Lysis is a two step process that involves holins which permeabilize the membrane, lysins that cleave the cell wall (80), and spanins that cross both membranes in Gram-negatives (81). By permeabilizing one (Gram-positive) or both (Gram-negative) membranes, holins both release lysins to cleave cell wall peptidoglycan and adjust the proton gradient to favor lysin activity.

Resistance at the lysis step can occur through interference with phage dissemination (73). Extracellular polysaccharides can reduce phage spread after lysis. Other resistance mechanisms at this step can manifest as delayed lysis or reduced burst size.

Lysogeny

While virulent phages only kill host bacteria upon infection, temperate phages can also integrate into the host chromosome, or lysogenize the host, after infection, thus becoming a prophage. A set of inputs determines whether bacteria establish lysogeny - appropriate integration site, physiological state (e.g., nutrient depletion favors lysogeny), and infecting phage density (higher MOI favors lysogeny). Upon integration, a complex regulatory system is responsible for maintaining lysogeny. The cI (clear I) repressor discovered in *E. coli* phage lambda and its homologs are the central regulators of temperate phage lysogeny (82). Lysogens prevent superinfection with the same phage through the process of superinfection immunity, in which existing cI repressor prevents infection with the same phage by transcriptionally repressing lysis genes. Prophages exit lysogeny upon stressors such as DNA damage, antibiotics, hydrogen peroxide, and changes to either nutrients, pH, or temperature (83). DNA damage triggers the SOS response, in which RecA binds to unwound ssDNA after DNA damage and cleaves cI repressor proteins to thus derepress genes responsible for the lytic cycle (84).

Resistance and host range determination

In the laboratory, several methods measure bacterial strain resistance to phages and corresponding host range for one phage against many strains (73). Host range is the number of strains a phage can infect out of those tested. These include the efficiency of plating (EOP), efficiency of centers of infection (ECOI), spot, and mixed-indicator assays. Spot assays, by far the simplest, involve dropping a small volume of phage lysate upon a soft agar bacterial lawn. It measures phage killing but not necessarily that due to phage propagation. Efficiency of plating (EOP) assays measure the ability of phages to plaque on a particular bacterial strain, which

involves bacterial killing by phage propagation. EOP is calculated as the ratio of plaques formed on a strain compared to those formed on a strain lacking phage resistance. The mixed indicator assay, on the other hand, assesses plaque formation on two indicator bacteria simultaneously in the same lawn. Clear plaques indicate host range expansion while turbid plaques indicate host range shifts. The failure to form plaques, however, does not exclude phage propagation. The "efficiency of center of infection" (ECOI) assay is a method to detect phage propagation that does not form plaques (85). After preadsorbing phage to the test bacterium and washing away free phage, the formed infective centers are plated on a lawn of known permissive bacterial strain. Plaque formation on the second strain only requires successful infection of the first. This assay thus most accurately measures host range in terms of productive phage infection.

Population dynamics with bacterial hosts

Unlike bacteria, phages grow in a stepwise fashion with their bacterial hosts. Such phage growth dynamics are captured in an experiment called a one-step growth curve. Phages infect bacteria at a multiplicity of infection (MOI) of 0.01-0.1 to prevent phage reinfection and plaque-forming units (PFUs) are measured at regular times after infection (86). The one-step growth curve splits into three periods - the latent period, during which no phage has been released from the cell, the eclipse phase, during which no fully matured phage exists in the cell, and the burst, in which phages are released after lysis (87). Latent period and burst size define the kinetics of a particular phage's infection. Unlike antibiotics, phages depend on the bacteria they kill for their own propagation. They thus exhibit predator-prey (Lotka-Volterra) dynamics with infectable host cells (88).

Therapeutic potential

Almost since their discovery, phages have been considered for therapeutic use against bacterial infections. Phage therapy was first developed in the 1910s as a treatment for dysentery (89–92). Phage therapy declined after the discovery and application of antibiotics, but continued

to be studied in eastern Europe and the former Soviet Union (Eliava Institute in Tbilisi, Georgia, for example) (89–91). There, phage cocktails were developed for sepsis, osteomyelitis, and burn wounds, among other infections, with complete recovery reported in some cases (93). More recently, the emergence of multi-drug resistance in bacterial pathogens has renewed interest in phage therapy and understanding underlying phage biology to make therapy viable (93–96). Phages have also been recently approved by the FDA as a treatment to clear *Listeria monocytogenes* from meats (Listex and ListShield), the first phage products to have been (97, 98). Bacteriophages are promising alternatives to antibiotics because of the large number of diverse natural phages, their low toxicity to non-target species, and their self-limiting spread. However, multiple challenges - standardizing formulation, selection of appropriate phages, lack of host range knowledge, potential for phage resistance, and phage neutralization by the immune system, among others - must be overcome for phage therapy to be successful (99, 100).

Phages that infect *S. aureus*...don't follow the book of lambda*

*acknowledgement to B. Levin for the title of this section

Classes, morphologies, and genome structures

All phages known to infect *S. aureus* belong to the order *Caudovirales*, or tailed dsDNA phages, and have been previously reviewed (39, 101). This order is divided into three classes based on virus morphology - the *Siphoviridae*, *Myoviridae*, and *Podoviridae*. Phages belonging to each class differ by average genome length, genome structure, and morphology - of the tail, most specifically. The temperate *Siphoviridae* have long, flexible, noncontractile tails (100-200 nm) and genomes roughly 39-43 kb in length. These phages are known to carry virulence genes such as the immune evasion cluster. The virulent *Myoviridae* have long, contractile tails (100 nm) and genomes roughly 120-140 kb in length. The virulent *Podoviridae* have short, noncontractile tails and genomes roughly 16-18 kb in length. Genomes of the *Siphoviridae* infecting this species are linearly and chronologically arranged by function (lysogeny, replication, packaging, head, tail, and lysis), while those of the *Myoviridae* and *Podoviridae* infecting the species are not (39).

Host ranges

S. aureus phages have broad host ranges relative to those infecting other species, but they still vary significantly amongst and between classes. The *Myoviridae* have the broadest host ranges by far, as determined by simple spot assays. Two lytic phages – the myovirus K and Stau2 – have extensive host ranges within *S. aureus*. One study of phage K found it infected all but 2 of 95 tested strains representing diverse *S. aureus* lineages (102). The lytic phage Stau2, isolated from a patient endotracheal tube, showed an even broader host range than phage K. Stau2 killed 80% of 205 *S. aureus* strains isolated from Taiwanese hospitals, compared to 47% for phage K (103). *Podoviridae*, on the other hand, have more limited host ranges (104, 105) due to their less common receptors, as we explain in the next section. The *Siphoviridae* being temperate have far more limited host ranges, though our own study and another (106, 107) found they could have wider host ranges, perhaps by spontaneous deletion of the lysogeny module. This module is responsible for resistance to respective phages as we also explain later.

Host factors affecting phage host range in *S. aureus*

Host genes have been found to affect *S. aureus* phage host range at three of the five steps of the lytic cycle - adsorption, biosynthesis, and assembly (Figure 3). I have previously reviewed this topic (53), with the findings paraphrased in the following paragraphs.

Adsorption

As in other bacterial species, *S. aureus* adsorption resistance involves receptor removal, alteration, or occlusion. *S. aureus* phages bind either to wall teichoic acid (WTA) ribitol phosphate (Rbo-P) backbone (*Myoviridae*) or backbone N-acetylglucosamine (GlcNAc) modifications (either α-O-GlcNAc or β-O-GlcNAc for *Siphoviridae*; for *Podoviridae*) (39). Another subset of phages bind a small subset of strains (ST395) that synthesize a different WTA (108) - glycerol phosphate (Gro-P) modified with N-acetylgalactosamine (GalNAc). The α-O-GlcNAc glycosyltransferase *tarM* is

present in 30% of *S. aureus* strains while the β-O-GlcNAc glycosyltransferase *tarS* is present throughout the species (109). *tarM*+ *tarS*+ strains are *Podoviridae* resistant as *tarM* has a dominant effect over *tarS*, making strain WTA α-O-GlcNAc-modified. However, both *tarM* and *tarS* must be knocked out for *Siphoviridae* resistance, while *tagO* (first step of WTA biosynthesis) knockout causes resistance to all classes of *S. aureus* phages, including the *Myoviridae*. On the other hand, a prophage-encoded enzyme (*tarP*) also confers *Podoviridae* resistance by moving O-GlcNAc from the 2 to the 3 position on ribitol phosphate (Rbo-P) (110). Two (1 and 2) of the four (1, 2, 5, and 8) known types of capsule in *S. aureus* are known to occlude phage receptors and reduce adsorption (22). High surface protein concentrations (surface protein A) are also known to prevent phage adsorption (111).

Biosynthesis

Resistance at the biosynthesis level in the species occurs through several known mechanisms - restriction-modification, CRISPRs, superinfection immunity, and abortive infection systems. The most common restriction-modification systems in *S. aureus* are type I, type II, and type IV systems. Type I systems (Sau1 primarily in *S. aureus*) cleave as far as 1000 bp from the binding and methylation site (112). Type I specificity (*hsdS*), defined by two variable target recognition domains (TRDs) in the HsdS protein sequence (113), correlates strongly with clonal complex (CC). Type II systems, on the other hand, cut at the restriction site. Type I and IV systems are conserved throughout the species, while type II systems are rare and strain-specific because they are carried on mobile genetic elements (114). Type IV systems (SauUSI) are responsible for cross-species restriction (115), while type I and II restrict within the species. Some *S. aureus* strains are known to encode type I anti-restriction proteins (ArdA) within Tn*5801* ICEs (34, 35), however. CRISPR systems are nearly nonexistent in the species. Abortive infection systems are quite rare in the species. The eukaryote-like serine/threonine kinase Stk2 kills host cells upon detection of a siphovirus protein of unknown function by phosphorylating host proteins with

diverse core cellular functions (116). However, Stk2 has been identified in only a few *S. aureus* strains.

Assembly

Resistance at the assembly level in the species involves assembly interference by *Staphylococcus aureus* pathogenicity islands (SaPIs) - phage-inducible chromosomal islands that carry virulence genes (such as toxic shock syndrome toxin) and are only excised upon cognate, helper *Siphoviridae* phage infection (36, 79). SaPIs require these helper phages as they lack lysis and structural genes. SaPIs interfere with helper phage assembly in three ways - 1) capsid remodeling, 2) helper phage DNA packaging interference, and 3) helper phage late gene repression. SaPIs remodel the helper phage capsid proteins to fit the small SaPI genome (117). To selectively package capsids with their own DNA, phage package interference (Ppi) proteins that inhibit helper phage but not SaPI terminases (118). Finally, phage transcription inhibition (Pti) proteins inhibit the helper phage late gene operon (packaging and lysis) to interfere with later steps of the helper phage life cycle (119).

Phage host range in *S. aureus* is determined by interaction of phage defenses expressed at different phylogenetic levels

A combination of both host- and phage-encoded genes as well as the epigenetic DNA methylation patterns conferred on phage DNA from the last strain it infected determine *Staphylococcus aureus* phage host range. Bacterium-encoded factors can be conceived of as affecting host range at different levels within the species (Figure 3). At the highest level, most phages' target for receptor binding (WTA) is highly conserved across *Staphylococcus aureus*. Strains with unusual WTAs, such as *S. aureus* ST395 and CoNS strains with poly-GroP WTA (108, 120), would be expected to be genetically isolated within the species. Type I R-M *hsdS* allotypes and capsule type are conserved between most strains of the same clonal complex (CC)

but differ between isolates of different CC groups and thus contribute to defining the host range in a large subset of *S. aureus* strains. At the level of individual strains, inserted prophages and SaPIs, Stk2, type II systems acquired by HGT, and other as yet unknown functions may all serve to limit the host range. We know even less about phage-encoded systems that counteract host resistance. The finding that lytic phages (*Myoviridae* and *Podoviridae*) tend to have broader host ranges than *Siphoviridae* when challenged against the same set of *Staphylococcus* strains suggests that the former encode an array of uncharacterized genes that work against host defenses.

S. aureus phages don't follow the book of *E. coli* lambda phage, nor *B. subtilis* phages

S. aureus phage-host interactions fundamentally differ from both Gram-negative (*E. coli*) and Gram-positive (*B. subtilis*) paradigms. Unlike phages infecting either of these other organisms, *S. aureus* phages are neither filamentous nor encode single strand genomes, bind reversibly and irreversibly solely to carbohydrate receptors, and have broad host ranges. S. aureus phages bind reversibly to lipoteichoic acid (LTA) (121) and irreversibly to WTA, while *B. subtilis* phages like SPP1 bind reversibly to WTA (122) and irreversibly to protein receptors like the phage interaction protein (Pip) homolog YueB (123). *E. coli* phages bind to a range of receptors from carbohydrate lipopolysaccharide (LPS) to outer membrane proteins like maltose-binding protein (MBP; lambda) and efflux pumps (TolC) (56). Additionally, CRISPRs and abortive infection systems appear to be much more prominent in *E. coli* and *B. subtilis* (71, 124–127), and more such systems have been discovered in these species. Furthermore, unlike most *E. coli* or *B. subtilis* temperate phages, *S. aureus* temperate *Siphoviridae* are known to derepress and correspondingly face resistance from chromosomal phage-like elements (SaPIs) that compete for the same assembly pathways (36, 79).

Therapeutic potential

Phage therapy has long been known to be particularly effective against staphylococcal infections (128). The famous Eaton and Bayne-Jones phage therapy review and meta-analysis suggested phage therapy had the most efficacy against staphylococcal infections (129). Nonetheless, lytic phages have varied in potential effectiveness for treatment and prevention of infection. Stau2, previously discussed with regards to host range, protected 100% of tested mice against lethal *S. aureus* infection at a multiplicity of infection (MOI) of 10 (103). Phages ΦMR11 and LS2a protected against lethal *S. aureus* infection in mice and abscess formation in a rabbit model, respectively (130, 131). On the other hand, despite wide host range and high levels of bacterial killing, some phages have shown weak effectiveness in controlling colonization. Phage K was shown to decrease *S. aureus* titer on skin in hand wash tests, but not eradicate it completely. In addition, a cocktail of phages P68 and K*710 effectively killed tested methicillin-resistant (MRSA) strains in vitro but did not protect against nasal colonization in a pig model (132). It has not yet been approved for clinical use in the United States, however. Phage lysates have been used for veterinary treatment as part of a STAPHAGE cocktail (Dumont Labs) (128).

Personalized phage therapy through rapid genome-based diagnostics

The goal of the research presented in this thesis is to improve detection of phage resistance through genome-based diagnostics. Phage therapy, not only against staphylococcal infections, has been hampered by lack of host-side host range knowledge. With the ability to curate and discover phage host range determinants, it will one day be possible to predict the efficacy of phage therapy based on infecting strain genome sequence. This will give rise to personalized, efficacious phage treatments for staphylococcal infections. Prediction will especially be possible because of technologies that give clinically important sequence information very fast - the most promising of which is nanopore sequencing.

Nanopore sequencing and its applications

Principles behind method

Nanopore sequencing determines the order of nucleic acid bases in a sample fragment by measuring electrical signal (current) as nucleic acids are drawn through membrane-embedded protein nanopores after unwinding and application of a voltage (Figure 4A and B). This process enables much longer reads and rapid, local, real-time (minute or hour) collection of sequence data, unlike sequencing by synthesis (SBS; exemplified by the ubiquitous Illumina sequencing technology), in which reads are sequenced by bases added, which requires weeks for preparation and analysis and is limited in read length in the case of Illumina sequencing (comparisons described in Table 1). Nanopore sequencing was first patented in 1995; purine and pyrimidine sequences were distinguished in single RNA molecules in 1999, solid-state nanopores produced in 2001, single nucleotides distinguished in 2005, DNA translocation processively controlled by 2012, and nanopore sequencing first released for commercial use in 2014 (133).

DNA is prepared as nanopore libraries suitable for pore and DNA motor protein binding rather than synthesis via PCR and chain termination, with megabase long reads possible with careful sample prep. DNA is sheared, end repaired, and then ligated to double stranded adapters with attached motor proteins (134, 135). After prepared DNA is unwound by a helicase that itself binds the pore, a single DNA strand passes processively through the nanopore under a voltage (Figure 4A), and blocked pore current is collected over time as the DNA passes through the pore (Figure 4B). Current traces are converted to DNA sequence through a process called basecalling. There are several approaches to basecalling but typically neural network models are trained to match impedance profiles to strings of nucleotides of a certain sequence (134, 136). Basecalling accuracy has increased in recent years from 70 to 98% (137) with advances in sequencing chemistry and algorithms. While all input sequence must be basecalled, input sequence can be selected at the pore via computational control. Such functions (read.Until) select DNA driven into

or out of the pore based on recently recorded sequence in real time, making selective sequencing possible (138, 139).

As the basecalling procedure demonstrates, nanopore sequencing is intrinsically dependent on complex algorithmic interpretation. Nonetheless, downstream sequence processing follows similar pipelines to other types of sequencing data (Figure 4C), with nanopore-specific nuances. Such software tools include sequence data simulators (SiLiCON, NanoSim, DeepSimulator), base modification (e.g., methylation) detection tools (DeepSignal, DeepMod, Nanopolish, SignalAlign), nanopore-specific assemblers (LQS, PBcR, Falcon, canu, miniasm, Unicycler, flye), long (Nanopolish, racon) and short (pilon) read polishers, and mutation callers (medaka, Nanopolish). Sequence data simulators use either existing models for different nanopore chemistries or models trained based on test data to generate synthetic nanopore data. Assemblers combine long reads into complete contiguous sequences, but are adjusted to account for long, error-prone reads (140). Long (racon) and short (pilon) read polishers align respective reads against a nanopore assembly to bridge contigs and correct nanopore systematic errors, respectively. The mutation caller and read polisher medaka identifies true mutations from sequencing error based on nanopore-specific error models.

Strengths and weaknesses relative to sequencing by synthesis (SBS) technologies

Nanopore sequencing has several strengths and weaknesses relative to sequencing by synthesis (SBS) methods that directly detect each base as it is added to the new strand. Speed is one of several advantages of nanopore sequencing relative to existing Illumina and PacBio sequencing technologies, which both rely on SBS (Table 1). Illumina sequencing relies on tracing the order of nucleotide base dye terminators to determine base sequence, while PacBio sequencing traces light pulses during single molecule synthesis with zero-mode waveguides, which hold the smallest possible volume in which light can be detected (141). Unlike these existing technologies, which report sequencing results in several hours to days, and often are only present in sequencing centers, nanopore sequencing returns data immediately after sample preparation

into a sequencing library, which itself can take 10-90 minutes. Unlike Illumina and PacBio instruments, nanopore sequencers are as small as a candybar in their standard form (MinION) and only need a desktop computer for collecting data. Read lengths are similar between nanopore and PacBio sequencing technology (5-10 kb vs. 10-15 kb average, respectively) but much longer for nanopore than Illumina (5-10 kb average vs. 300 bp maximum). Nanopore and PacBio sequencing also both natively detect nucleotide modifications (e.g., methyl-6-adenine - m6A) either via current trace differences (nanopore) or kinetics of base translocation (PacBio). However, Illumina offers far higher accuracy (99.9% vs. 90-98%) and PacBio typically offers more data and higher net assembly accuracy. Price per gigabase (estimated as of 2016) is highest for nanopore (~$1000) relative to Illumina (~$30) and PacBio (~$500) (142–144). However, the release of Flongle flow cells - smaller (128 vs. 512 pore), less expensive (~$90) flow cells that still maintain 2-3 Gb max output - in 2019 would bring the nanopore cost down to an estimated ~$100 per Gb, making it the second most expensive per output (145).

Recent applications

Nanopore sequencing applications include rapid pathogen detection, resolving repeats to help complete large genomes (of eukaryotes, especially), and phasing haplotypes. It was used to generate an *E. coli* K12 reference genome (146) in an original proof of concept study, cover a bacterial pathogenicity island (147), complete 170 kb bacterial plasmids containing repetitive insertion sequences (148), and perform epidemiological analysis of Zika virus outbreaks in Brazil (149, 150). It has also been used as a diagnostic tool, detecting bacterial pathogens directly from the urine of patients with urinary tract infections (151) and respiratory samples from patients with TB (152). Because reads are long (5-10 kb average, but including 20 kb+ reads), nanopore data can resolve repeats that short read sequencing would not. This has made it possible to complete large, eukaryotic organism assemblies (e.g., *Saccharomyces cerevisiae*) (153), or even repetitive bacterial plasmids (148), as previously mentioned. The long reads also solve phasing problems in both small and large genomes, or the correct association of multiple alleles into contiguous

haplotypes. Short-read sequencing fails to associate SNPs beyond certain lengths (i.e., of assembled contigs). Long reads can fully complete genomes and thus exceed these thresholds, making it possible to stitch together all alleles into a full haplotype. This makes it possible to potentially determine all alleles in a eukaryotic chromosome or a minority subpopulation of a eukaryotic or prokaryotic sample (e.g., tissue sample, bacterial colony, bacterial culture). Such nanopore haplotyping was recently demonstrated on the human MHC/HLA alleles (154).

Nanopore sequencing for antibiotic resistance detection

Rationale for use and therapeutic potential

Rapid, long-read sequencing has the potential to be a game changer for diagnosing and treating antibiotic resistant bacterial infections. Existing culture-based methods slow down diagnosis while organisms grow (24-48 hours typically but longer for TB and other fastidious microbes). Culture also could miss resistant subpopulations, for example, heteroresistant strains that contain sensitive majority and super-resistant minority subpopulations. Nanopore sequencing on the other hand could detect resistance faster (even 15 minutes or less after sequencing begins, not counting DNA and library preparation time) and would tell exactly what gene or mutation is responsible for resistance. Also, due to its read length, it could detect expansions in tandem repeat copy number that can lead to antibiotic heteroresistance, as recently shown for arylomycin (155). Nanopore sequencing detected copy number amplifications of arylomycin target gene *lepB* ranging from 4.8 to 50.0 kb, which in turn leads to variation in arylomycin resistance (155). This would make it possible to treat infections far earlier during pathogenesis and more precisely, possibly determining treatment outcome. Studies have indicated mortality increases 10% for every hour septic shock infections are left untreated, for example (156).

Current and past efforts

Efforts to use nanopore sequencing for antibiotic resistance detection have identified resistance gene presence alone, used bacterial genotype to predict resistance, or attempted to

detect causative resistance mutations (SNPs). These efforts have sequenced either genomic (culture) or metagenomic (direct environmental sample) inputs.

The first nanopore antibiotic resistance strategy reviewed is the most direct - identification of causative resistance genes. Resistance genes have been rapidly detected in mobile genetic elements (MGEs) such as plasmids (157), bacterial strains isolated from blood culture, shellfish, and sewage, and metagenomic samples from lung tissue, urine, and wastewater. *E. coli* and *Klebsiella pneumoniae* plasmid sequencing studies indicated reads collected in only 20 minutes of sequencing were sufficient to provide an antibiotic resistance gene-annotated assembly and that plasmid resistance genes could be fully annotated in as little as 6 hours after processing a subcultured isolate. Nanopore plasmid sequencing also identified the spread of resistance through plasmids in hospital settings (158, 159). Time frames for resistance gene detection have ranged from minutes to hours. Long reads from nanopore sequencing provide the added benefit of placing antibiotic resistance genes into their genomic contexts. Genomic studies indicated that nanopore data helped properly assemble diverse *E. coli* tigecycline resistance plasmids and beta-lactamase-carrying genomes and antibiotic resistant-carrying MGEs in hospital microbiomes.

An alternative strategy has extrapolated antibiotic resistance from lineage instead. Closely related strains often share antibiotic resistance phenotypes, so strain lineage can predict antibiotic resistance. Břinda et al. cleverly used rapid *Streptococcus pneumoniae* sequence type (ST) detection to determine antibiotic resistance phenotype, taking 10 minutes or less to detect resistance after sequencing begins and 4 hours or less after clinical sputum collection (160). They used this “genomic neighbors” technique to predict antibiotic resistance in both *Streptococcus pneumoniae* and *Neisseria gonorrhoeae*.

A third strategy, direct resistance mutation detection, has rarely been implemented to date. It has rarely been implemented because of the high levels of error and more specifically, systematic error, that nanopore sequencing exhibits per read. Homopolymer regions (e.g., AAAAA) are most affected by this systematic error. Only in 2020 was nanopore sequencing

reported to detect antibiotic resistance SNPs in metagenomes containing *Neisseria gonorrhoeae* (161).

Another recent development has taken a completely different approach to detecting antibiotic resistance by discriminating live from dead bacteria under antibiotic treatment (162). Bacterial cultures (*E. coli* and *Pseudomonas aerguinosa*) were grown without (control) or with (treatment) antibiotic, cultured for 1 hour, treated with propidium monoazide (PMA) to stain dead bacterial DNA and prevent PCR, DNA extracted, and then sequenced for 16S rRNA after PCR. Only live bacteria would be thus subject to PCR, and levels of death under antibiotic treatment could be measured by PMA fluorescence under blue light. The assay would be expected to find drug-resistant bacteria in 4 hours. This technique thus both detects antibiotic resistance and the members of a bacterial community possessing it, without identifying the causative determinant itself. A similar technique was used to distinguish carbapenem resistant from sensitive *Klebsiella pneumoniae*, but instead by comparing 16S rRNA - not RT-PCR amplified DNA - nanopore/probe detection between control and antibiotic-treated cultures (163).

Strengths and weaknesses relative to culture-based techniques

Nanopore sequencing has multiple strengths and weaknesses relative to culture-based techniques for detecting antibiotic resistance. It is quite rapid (sequencing is complete in under 24 hours), but it is also expensive, difficult to set up, and per-read error is high (2-10%). In addition, finding resistance genes or alleles doesn't necessarily indicate actual antibiotic resistance, as previous studies have found (164). Culture-based tests like broth microdilution minimum inhibitory concentration (MIC) determination remain the gold standard for antibiotic resistance determination. Coupling short culturing followed by detection of live cell DNA as performed in the PMA dye/nanopore drug resistance detection study could strengthen the accuracy of genome-based tests while still reducing culture and thus test time considerably.

Future applications

Future nanopore sequencing applications include better detecting resistance mutations, exploring resistant subpopulations and epigenetic changes conferring antibiotic resistance, determining resistance based on real-time mutation detection, and pushing limits of sample preparation and diversity (e.g., working with DNA input less than the 0.25-1 µg typically used for library prep, large numbers of samples, and many genes of interest). Nanopore detection also must overcome non-random sequencing errors such as in homopolymer regions. In addition, so far, resistance detection has been limited to classes of antibiotic resistance caused by single genes or mutations in individual genes. The haplotyping ability, long read length, and base modification identification that nanopore sequencing offer may increase our knowledge of epigenetics-dependent and subpopulation-level resistance in the laboratory and clinic. Finally, the most direct goal nanopore sequencing must meet in future AMR detection is simple improvement of timing, accuracy, and sampling constraints such as numbers of genes analyzed. Most studies have examined individual genes sufficient to cause resistance, while other phenotypes like vancomycin-intermediate *Staphylococcus aureus* (VISA) involve multiple independent genes containing causative mutations (165). Regarding sample number, barcoding kits (allowing six samples per flow cell) could boost throughput, but would need to be validated to see if they still provided the coverage necessary to identify causative mutations especially. Such validation would involve aligning the single flow cell's output against a reference and calling mutations and demultiplexing the barcoded samples and aligning each against respective references. The multiplexed flow cell would need to provide high enough coverage per sample to call causative mutations over sequencing error, especially at systematically error-prone sites.

Nanopore sequencing for personalizing phage therapy

Current limitations of phage therapy

Phage therapy currently suffers from many limitations hindering its clinical use. Host range knowledge makes it possible to predict therapeutic efficacy, but such knowledge remains minimal. It also remains difficult to standardize phage therapeutic propagation and preparation. The underpinnings of phage resistance development during infection remain largely unknown in many cases. Possible immune system phage neutralization and unknown consequences on the local microbiome remain further challenges. Nanopore sequencing could help with standardization, efficacy prediction, and phage resistance detection, especially in contexts that benefit from real-time sequencing.

Genomic prediction of phage efficacy and potential for nanopore impact

As stated earlier, the goal of the presented dissertation research is to improve genomic prediction of phage therapy efficacy against staphylococcal infections. Nanopore sequencing has the advantage of rapid throughput relative to other sequencing techniques. Unfortunately, as our recent paper indicates, phage host range in *S. aureus* is a polygenic phenotype and the determinants we discovered only account for 60-90% of host range variability (166). There is still much more that needs to be discovered on the host side to make genomic host range prediction effective.

It may one day be possible to sequence a clinical sample or bacterial isolate through nanopore and phage host range from genome sequence quickly in real time. Based on our *S. aureus* phage resistance studies, this is probably extremely difficult for similar hosts where causative determinants are numerous and complexly interrelated (possibly through epistatic interactions). This may be reasonable for hosts where a small subset of determinants explains the majority of resistance, such as protein receptor mutations or integrated prophages causing superinfection immunity. However, even in such cases there may be phase variable leaky

resistance that is more difficult to detect. Phase variable gene expression due to slipped strand mispairing at homopolymer regions, for example, would be difficult to pick up by nanopore sequencing.

Figures

Figure 1: Cell wall structure and wall teichoic acid (WTA) biosynthesis in *S. aureus*. (A) Structure of the staphylococcal cell envelope. Lipoteichoic acid is shown in orange (glycerol phosphate), a surface protein is in black, wall teichoic acid is in orange (glycerol phosphate) and yellow (ribitol phosphate), capsule is in blue, and cell wall carbohydrates are in green (*N*-acetylglucosamine [GlcNAc]) and purple (*N*-acetylmuramic acid [MurNAc]). Staphylococcal phages bind WTA and/or its ribitol phosphate modifications (i.e., GlcNAc). (B) Outline of the wall teichoic acid (WTA) biosynthesis pathway, with the proteins corresponding to each step listed in the blue arrows. Abbreviations are defined as follows: C_{55}-P, undecaprenyl phosphate; GlcNAc, *N*-acetylglucosamine; UDP-GlcNAc, uridine-5-diphosphate-*N*-acetylglucosamine; ManNAc, *N*-acetylmannosamine; UDP-ManNAc, uridine-5-diphosphate-*N*-acetylmannosamine; Gro-P, glycerol phosphate; CDP-Gro, cytidyl diphosphate-glycerol; Rbo-P, ribitol phosphate; CDP-Rbo, cytidyl diphosphate-ribitol; ABC, ATP-binding cassette; and LCP, LytR-CpsA-Psr. Figure originally published in Moller et al., AEM, 2019.

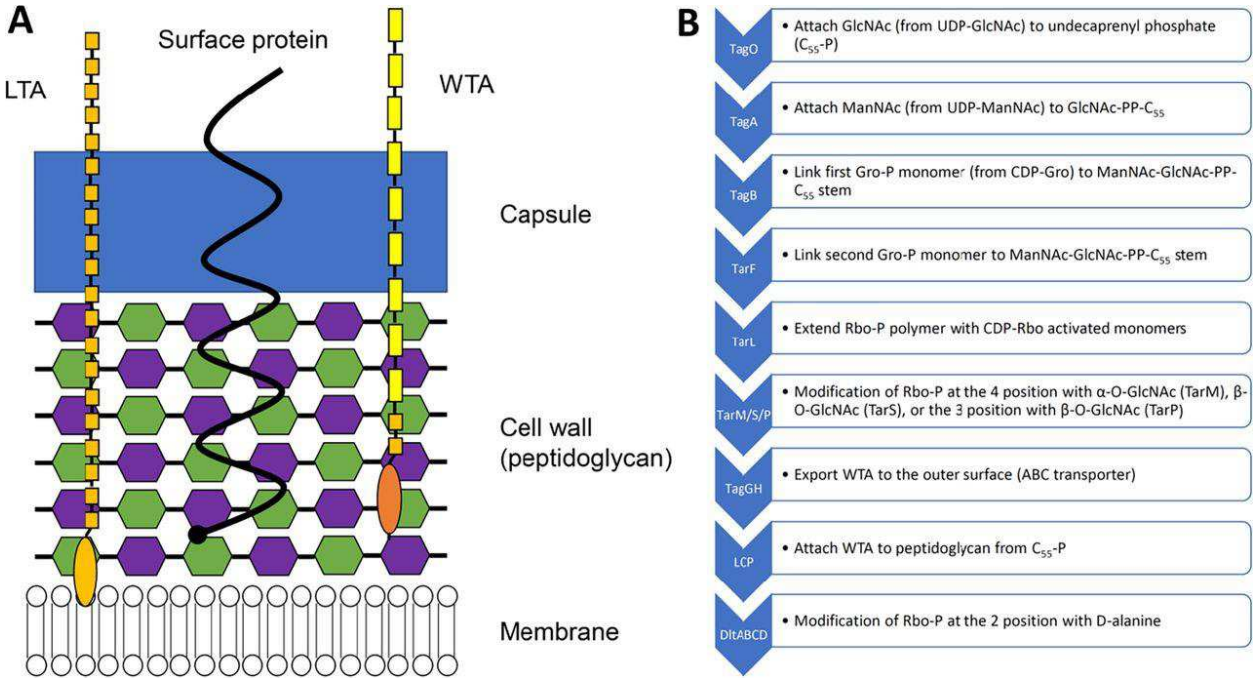


Figure 2: Stages of phage infection and corresponding examples of resistance mechanisms at each stage. Examples not yet identified in the staphylococci are listed in red. Figure originally published in Moller et al., AEM, 2019.

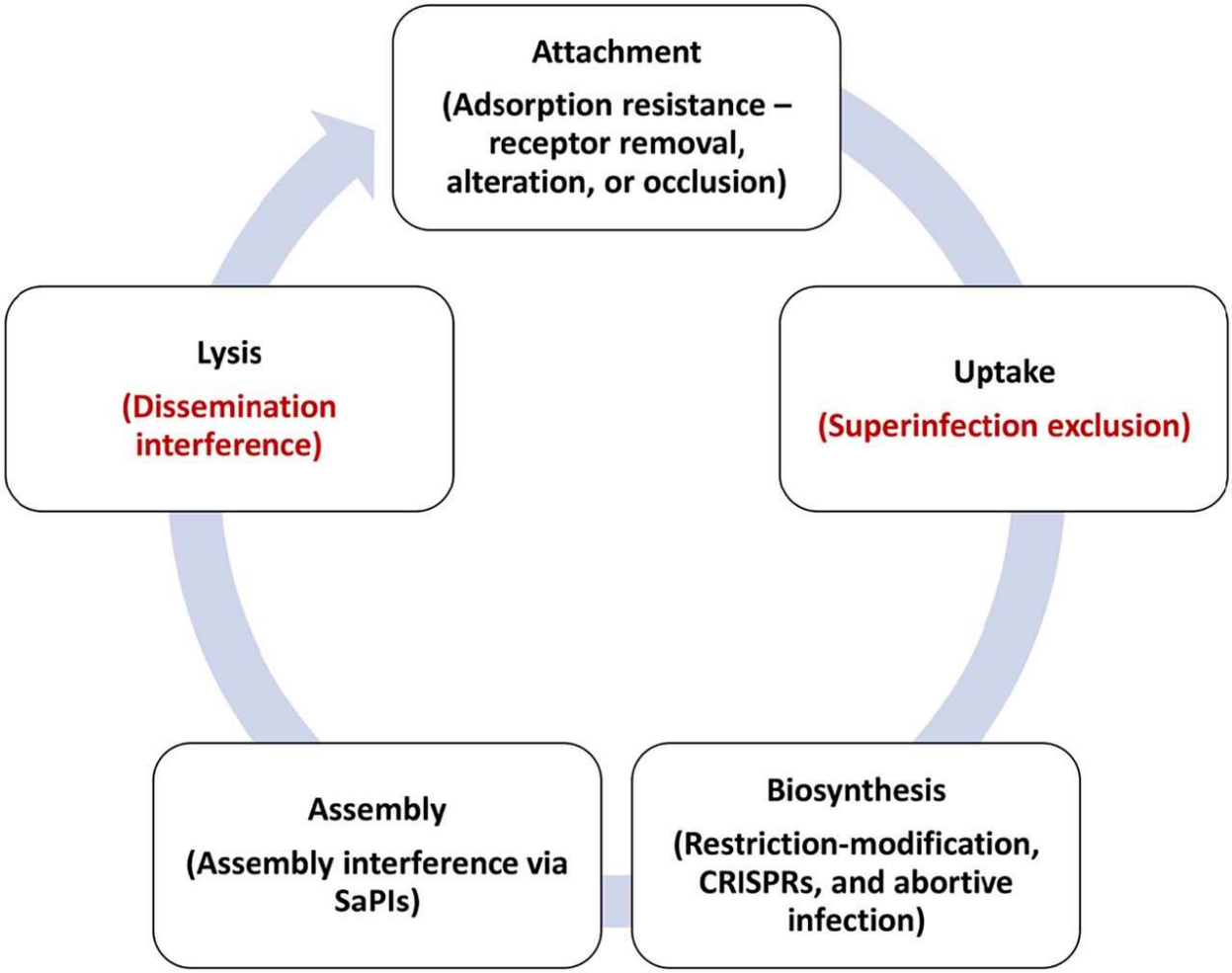


Figure 3: *S. aureus* host range determination is phylogenetically hierarchical. Phage host range for an individual strain is the combination of multiple factors that have different levels of conservation within the species. This is illustrated by a hypothetical phylogenetic tree. Mechanisms can be present throughout strains (1, most conserved; red), present in many strains but with considerable allelic variation (2, conserved but polymorphic; shades of green), or present in a few strains, possibly with allelic variation (3a to 3c, less conserved with potential polymorphism; blue, purple, and yellow, respectively). Branches where mechanisms evolved by mutation or homologous recombination, in the case of mechanisms 1 and 2, or were acquired by HGT, in the case of mechanisms 3a to 3c, are annotated with colored stars. The table on the right summarizes the mechanisms (1 to 3c) present in each strain (strains A to J) using shaded boxes with corresponding colors. Strain J has a mutation that results in the null phenotype for the red mechanism. Host range is the result of the combination of mechanisms present, so strains A to C as well as F, H, and I would be predicted to have identical host ranges, but phage-specific factors could also introduce variability. Figure originally published in Moller et al., AEM, 2019.

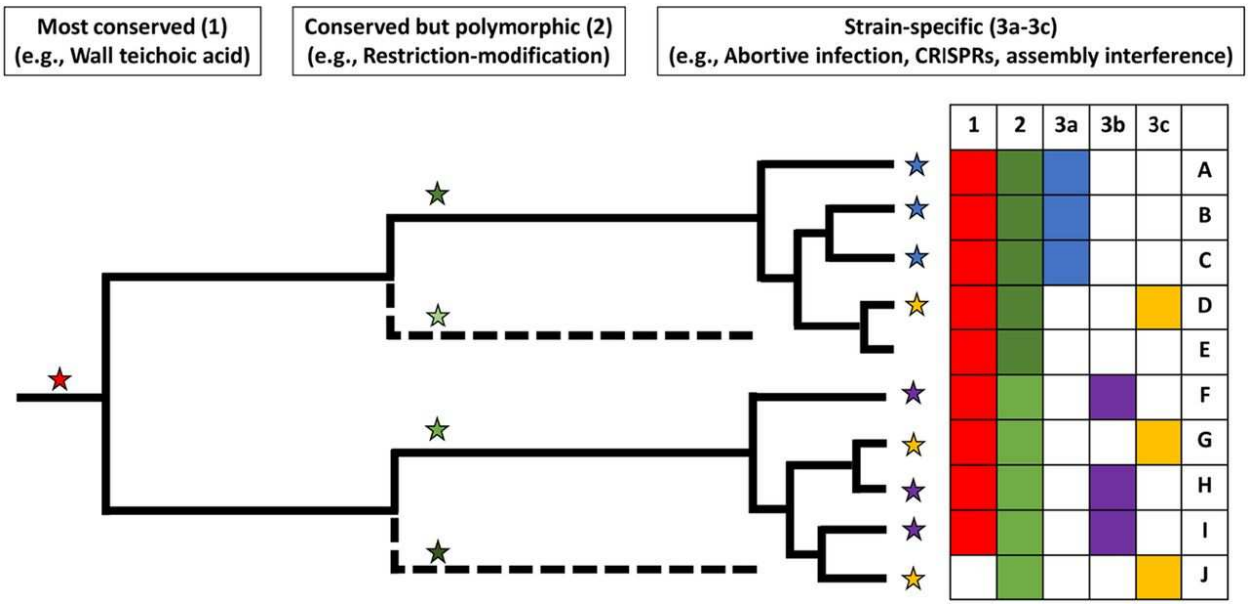
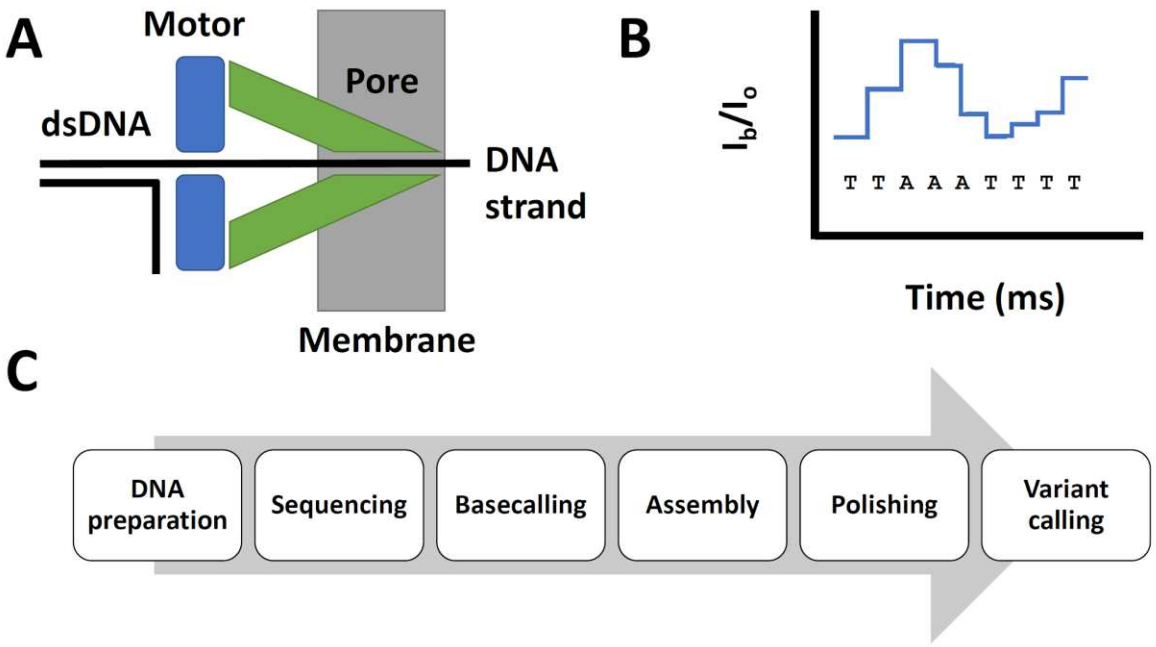


Figure 4: Nanopore sequencing mechanism (A), reading current traces (B), and pipeline for resistance detection (C). A) DNA is first unwound by a motor protein, allowing a single strand to pass through the pore embedded in the membrane. B) Blocked pore current (I_b) is compared to open pore current (I_o) over time to generate current trace that is then converted to a sequence by a neural network algorithm. C) The most common sequence of steps for processing and analyzing nanopore data starts with DNA preparation and ends with variant calling.



Tables

Table 1: Comparison of nanopore, Illumina, and PacBio sequencing technologies (142–144). Best technologies in each metric (e.g., average read length) are highlighted in bold. Oxford Nanopore is listed as fastest because the speed listed does not include sample delivery time, and only nanopore sequencing is available in compact devices usable outside a sequencing facility.

Characteristic/ sequencing method	**Illumina**	**PacBio**	**Oxford Nanopore**
Sequencing method	Read synthesis with dye terminators	Single-molecule real time synthesis via zero mode waveguides	Measuring electrical currents through protein nanopores
Average read length	300 bp maximum	**10-15 kbp**	5-10 kbp
Speed (per Gb) (not counting delivery time)	6 days	2-3 hours	**6 hr-2 days**
Raw error rate	**0.1%**	10-15%	2-30%
Cost (per Gb)	**~$30**	~$500	~$100-1000 (high estimate)
Applications	Whole genome sequencing, RNA-seq, mutation detection	Whole genome sequencing, methylation detection	Infectious disease diagnostics, mutation detection, metagenomics

Chapter 2: Determinants of Phage Host Range in *Staphylococcus* Species

Published in AEM

Abraham G. Moller[a,c], Jodi A. Lindsay[b], and Timothy D. Read[c*]

a. Program in Microbiology and Molecular Genetics (MMG), Graduate Division of Biological and Biomedical Sciences (GDBBS), Emory University School of Medicine, Atlanta, Georgia, 30322

b. Institute of Infection and Immunity, St George's, University of London, Cranmer Terrace, London, SW17 0RE, UK

c. Division of Infectious Diseases, Department of Medicine, Emory University School of Medicine, Atlanta, Georgia, 30322

*Corresponding author – tread@emory.edu

Abstract

Bacteria in the genus *Staphylococcus* are important targets for phage therapy due to their prevalence as pathogens and increasing antibiotic resistance. Here we review *Staphylococcus* outer surface features and specific phage resistance mechanisms that define host range - the set of strains an individual phage can potentially infect. Phage infection goes through five distinct phases - attachment, uptake, biosynthesis, assembly and lysis. Adsorption inhibition, encompassing outer surface teichoic acid receptor alteration, elimination, or occlusion, limits successful phage attachment and entry. Restriction-modification systems (in particular, type I and IV systems), which target phage DNA inside the cell, serve as the major barriers to biosynthesis as well as transduction and horizontal gene transfer between clonal complexes and species. Resistance to late stages of infection occurs through mechanisms such as assembly interference, in which staphylococcal pathogenicity islands siphon away superinfecting phage proteins to

package their own DNA. While genes responsible for teichoic acid biosynthesis, capsule, and restriction-modification are found in most *Staphylococcus* strains, a variety of other host-range determinants (e.g., CRISPRs, abortive infection, and superinfection immunity) are sporadic. Fitness costs of phage resistance through teichoic acid structure alteration could make staphylococcal phage therapies promising, but host range prediction is complex because of the large number of genes involved, many with unknown roles. In addition, little is known about genetic determinants that contribute to host range expansion in the phages themselves. Future research must identify host range determinants, characterize resistance development during infection and treatment, and examine population-wide genetic background effects on resistance selection.

Keywords: staphylococci, phage resistance, host range, phage therapy, CRISPR

Introduction

The *Staphylococcus* genus includes commensals and pathogens of humans and animals. *S. aureus* and *S. epidermidis*, in particular, cause diverse infections in humans and have become increasingly antibiotic resistant over the past seventy years. Diseases range from food poisoning to skin and soft tissue infections, pneumonia, osteomyelitis, endocarditis, and septic shock. *S. aureus* is carried by between 20% (persistently) and 60% (intermittently) of the human population (1), primarily on the skin and upper respiratory tract. Methicillin-resistant *S. aureus* (MRSA) emerged in the mid-1960s (167) and has reduced the options for treatment with beta-lactam antibiotics. The combination of high carriage rates, diverse pathologies, prevalent antimicrobial resistance, and lack of a licensed vaccine (4) makes staphylococcal species important targets for new therapies.

Bacteriophage (phages) are natural killers of *Staphylococcus* bacteria lysing bacterial cells through expression of holins, which permeabilize the membrane and release endolysins (168,

169) that degrade the peptidoglycan of the cell wall (170). Phage therapy is a promising alternative to antibiotics for treating infections because of the large number of diverse phages with low toxicity to humans and non-target species (96, 171).

Phage therapy has a long history, reaching back before the antibiotic era to shortly after the discovery of phages themselves by Frederick Twort and Felix d'Herelle in the 1910s (90–92). While overshadowed by the subsequent discovery of antibiotics and generally abandoned in the West for many years, phage therapy persisted as a bacterial treatment in eastern Europe and the nations that composed the former Soviet Union (90, 91). There, phage cocktails were developed for sepsis, osteomyelitis, and burn wounds, among other staphylococcal diseases, with complete recovery reported in some cases (93). Polish and Soviet studies showed that phage lysates effectively treated staphylococcal skin and lung infections (89). More recently, the emergence of multi-drug resistance in bacterial pathogens has renewed interest in phage therapy and phage biology (95, 96). Safety studies on the staphylococcal phage lysate (SPL) as well as phage cocktails containing *S. aureus*-specific phages indicated that they had no adverse effects when administered intranasally, intravenously, orally, topically, or subcutaneously (95). Phages have also been recently approved by the FDA as a treatment to clear another Gram-positive species (*Listeria monocytogenes*) present in food (97) and approved as personalized treatment for burn wound infections (172).

All known staphylococcal phages are members of the order *Caudovirales* with linear dsDNA virion genomes. Staphylococcal phages are divided into three families with distinctive morphologies – the long, noncontractile-tailed *Siphoviridae*, the contractile-tailed *Myoviridae*, and the short, noncontractile-tailed *Podoviridae* (39, 101). *Siphoviridae* genomes are 39-43 kb in size, while those of the *Myoviridae* are 120-140 kb and *Podoviridae* are 16-18 kb (39). Currently reported *Siphoviridae* are typically temperate phages that encode lysogeny functions within a genomic module, while reported *Myoviridae* and *Podoviridae* are virulent. The virulent phages are

the strongest potential candidates for phage therapy, given that they are not known to lysogenize and thus obligately kill their targets. Lytic staphylococcal phages have surprisingly broad host ranges (102, 103, 173, 174), anti-biofilm activity (102, 175), and varying effectiveness against infection (130–132). The *Siphoviridae* are agents of horizontal gene transfer (HGT) through transduction (37) into recipient strains (39) and activation of staphylococcal pathogenicity islands (SaPIs) (79). The *Siphoviridae* have been subdivided into “integrase types” based on the sequence of the integrase gene necessary for lysogenic insertion into the chromosome (39, 176). Certain integrase type phages introduce specific virulence factors (39). Integrase type 3 (Sa3int) phages encode the immune evasion cluster (IEC), which includes the staphylokinase (*sak*), staphylococcal complement inhibitor (*scn*), chemotaxis inhibitory protein (*chp*), and enterotoxin S (*sea*). In addition, Sa2int phage often encode Panton-Valentine leukocidin (*lukFS-PV*), while Sa1int phages often encode exfoliative toxin A (*eta*). Temperate staphylococcal phages can also disrupt chromosomal virulence factors (39). Sa3int and Sa6int phages, for example, integrate into sites in the beta-hemolysin (*hlb*) or lipase (*geh*) genes, respectively (177, 178).

No single phage can kill every *Staphylococcus* strain. Instead, each phage has a particular host range, defined as the set of strains permissive for its infection. Host range can be limited by active host resistance mechanisms such as CRISPR or restriction-modification that actively suppress phage infection or by passive mechanisms such as loss of receptors for phage adsorption. It is unclear whether these host range limiting factors have arisen through specific adaptation against phage infection or are byproducts of selection against other stresses. There are, however, specific phage counteracting mechanisms to host resistance that serve to broaden phage host range. Phage host range has great importance to phage therapy because it defines the potential scope of treatable strains, thus informing selection of phages for rational, personalized cocktail development.

Mechanisms of resistance to phage have been reviewed previously across bacteria generally (59, 179) and in lactic acid bacteria (180), but this is the first article to focus on the particular features of *Staphylococcus* (**Figure 1**). By far, the majority of the literature has focused on two species: *S. epidermidis*, and especially, *S. aureus*. However, we include studies on other species (e.g. *S. simulans*) where appropriate. We then reflect on possible consequences of resistance on phage host range and potential phage therapy for staphylococcal infections, given that phage resistance elements determine host range and thus provide one criterion for phage efficacy in therapy. We also consider the evolutionary trade-offs of phage resistance in a therapeutic context due to the potential effects of phage resistance on either virulence or antibiotic resistance.

Host resistance can occur at different points in the phage life cycle (**Figure 1**) (59, 179). There are no reports in *Staphylococcus* of mechanisms that limit host range at the uptake and host lysis phases. We therefore concentrate on the attachment, biosynthesis, and assembly phases.

Attachment

Wall teichoic acid is the primary staphylococcal phage receptor

Attachment of phages to the outside of the *Staphylococcus* cell (**Figure 2A**) is the first stage of infection (**Figure 1**). *Staphylococcus* may be resistant to phage adsorption if the receptor molecule is not present, not recognized by the phage, or blocked. Mutations that alter components of the outer surface can have the effect of inhibiting adsorption and thus conferring resistance. Through genetic and biochemical studies on a small range of staphylococcal phages, the polyribitol phosphate (poly-RboP) polymer of wall teichoic acid (WTA) or N-acetylglucosamine (GlcNAc) modifications at the 4 positions of ribitol phosphate monomers in WTA appear to be the primary targets (18, 181–186).

In an early *S. aureus* phage resistance study published in 1969, N-methyl-N'-nitro-N-nitrosoguanidine-mutagenized strain H (Multi Locus Sequence Type 30; ST30) (187) phage-resistant mutants were selected by plating on agar plates containing lawns of 52A (siphovirus) (185). Mutants also found resistant to phage K (myovirus) were deficient in N-acetylglucosamine, cell wall phosphorus, and ester-linked D-alanine in their envelopes, presumably due to a loss of wall teichoic acid production. Further biochemical characterization showed that the mutants lacked UDP-GlcNAc:polyribitol phosphate transferase activity and WTA. Counterintuitively, they did show the relevant biochemical activity for the last known step in WTA biosynthesis (phosphoribitol transferase – TarL, **Figure 2B**) (183). This surprising result suggested the double resistant mutants produced ribitol phosphate but either failed to properly polymerize WTA or attach it to the cell wall. These mutants had pleiotropic phenotypic differences from their parent strain (186), including a longer generation time than its parent; cell growth in clumps; irregular, rough, gray colonies; and increased levels of wall-bound autolysin. A later study characterizing spontaneous *S. aureus* strain A170 (ST45) mutants resistant to siphovirus M^{Sa} found similar phenotypic defects (188) and biochemical assays also showed that resistance was likely due to the lack of GlcNAc-modified WTA.

Peschel and colleagues identified genes responsible for phage adsorption in a series of elegant molecular genetic studies in the RN4220 (ST8) (189) background (18, 109, 181). Deletion of undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase (*tagO*), the first gene involved in WTA biosynthesis, conferred resistance and reduced adsorption to tested *Myoviridae* (Φ812 and ΦK), while a transposon insertion mutant in the *tarM* gene had resistance and reduced adsorption to *Siphoviridae* (ΦSa2mw, Φ47, Φ13, and Φ77). Complementation of wild-type alleles rescued these phenotypes (181). TarM is a glycosyltransferase responsible for attaching α-O-GlcNAc to the 4 position of the ribitol phosphate WTA monomer (20, 190). The *tarM* mutant was previously shown to lack GlcNAc-modified WTA in its envelope (20). TarS, the glycosyltransferase

responsible for attaching β-O-GlcNAc to the 4 position of the ribitol phosphate WTA monomer (191), was specifically required for podovirus adsorption (109). Deletion of *tarS* conferred resistance and reduced adsorption to tested *Podoviridae* (Φ44AHJD, Φ66, and ΦP68) (109), but only deletion of both *tarS* and *tarM* conferred reduced adsorption to tested *Siphoviridae* (Φ11) in the same RN4220 background used in prior studies (192, 193). On the other hand, even *tarS*+, *tarM*+ strains were resistant to *Podoviridae*, suggesting WTA decorated with α-O-GlcNAc by TarM impeded podovirus adsorption (109). Taken together, these findings suggested, for the small number of representatives that were tested, elimination of WTA confers resistance to all classes of phage, elimination of GlcNAc modifications confers resistance to the *Siphoviridae* and *Podoviridae*, and elimination of β-O-GlcNAc modification confers resistance specifically to the *Podoviridae*. Given the conservation of wall teichoic acid biosynthesis genes amongst *S. aureus* genomes (194) and the cross-species activity of staphylococcal phages such as phage K (195), these conclusions could be expected to hold in staphylococci beyond *S. aureus*.

Recent studies have suggested that as the number of strains and phages expands we may find a larger number of genes influencing host range through attachment. Azam et al. conducted a long-term evolution experiment in which they selected *S. aureus* SA003 (ST352) mutants resistant to myovirus ΦSA012 (196). Resistant mutants gained missense mutations in five genes (*tagO*, RNase adapter protein *rapZ*, putative membrane protein *yozB*, guanylate kinase *gmk*, and alpha subunit of DNA-dependent RNA polymerase *rpoA*), a nonsense mutation in one gene (UDP-N-acetylglucosamine 1-carboxyvinyltransferase *murA2*), and a 1,779 bp deletion that included the C-terminal region of the teichoic acid glycosyltransferase *tarS*, a non-coding region, and the N-terminal region of the iron-sulfur repair protein *scd*. Complementation of mutations in genes *scd*, *tagO*, *rapZ*, and *murA2* restored ΦSA012 sensitivity and adsorption, while only complementation of mutations in *tarS* restored sensitivity and adsorption of another myovirus, ΦSA039. The results suggested that while ΦSA012 recognized the WTA backbone, ΦSA039 was

unusual in recognizing β-O-GlcNAc-modified WTA, hinting that there may be more variability in receptor targets within phage groups than the limited number of earlier studies suggested.

The carriage of a prophage in certain *S. aureus* CC5 and CC398 strains that encodes alternative WTA glycosyltransferase *tarP* (110) adds further complications. TarP attaches GlcNAc to the 3 position of ribitol phosphate rather than the 4 position, thus conferring *Siphoviridae* (Φ11, Φ52a, Φ80) sensitivity but *Podoviridae* (Φ44, Φ66, and ΦP68) resistance. It is interesting in the light of host range evolution that a gene carried on a prophage can change the properties of the *S. aureus* surface and thus affect the host ranges of other phages.

Although the majority of staphylococcal phage tested bind WTA and GlcNAc receptors, there is one known exception. Siphovirus Φ187 binds WTA glycosylated with N-acetyl-D-galactosamine (GalNAc), the unusual WTA synthesized by *S. aureus* ST395 (108). The α-O-GalNAc transferase *tagN*, the nucleotide sugar epimerase *tagV*, and the short GroP WTA polymerase *tagF* genes are required specifically for synthesis of ST395 WTA. Homologs of these genes were found in genomes of multiple Coagulase-Negative *Staphylococci* (CoNS) strains, such as *S. pseudointermedius* ED99, *S. epidermidis* M23864:W1, and *S. lugdunensis* N920143. Complementation of a *S. aureus* PS187 *tagN* C-terminal glycosyltransferase deletion with the wild-type *tagN* gene or that from *S. carnosus* (*tagN*-Sc) successfully restored the wild-type phenotype, suggesting *tagN* homologs in other CoNS genomes had similar functions to that in *S. aureus* PS187 (ST395). Complementation of the *tagN* C-terminal deletion with either PS187 or *S. carnosus tagN* also restored wild-type Φ187 sensitivity. This difference in WTA structure was shown to prohibit transduction between ST395 and other *S. aureus* lineages (120). Staphylococcal pathogenicity island (SaPI) particles prepared in a ST1, 5, 8, 22, 25, or 30 strain with phages Φ11 or Φ80α failed to transduce any ST395 strains. SaPI particles prepared in a ST395 strain, on the other hand, transduced other ST395 strains as well as CoNS species and *Listeria monocytogenes*. These findings suggest the unique ST395 WTA restricts phage host

range to strains of the same sequence type or Gram-positives with a related WTA structure, such as *Listeria monocytogenes*.

There has been one study showing that staphylococcal phages (siphovirus ΦSLT) can bind lipoteichoic acid (LTA), the lipid-anchored, polyglycerol phosphate (GroP) TA polymer (121) (**Figure 2A**). However, subsequent elimination of LTA biosynthesis through *ltaS* deletion had no effect on phage adsorption or sensitivity (181) and therefore the potential significance of LTA as an alternative receptor is currently unknown.

The effects of surface proteins and extracellular polysaccharides on attachment

Although proteins serve as receptors for many Gram-positive phages (for example, the YueB receptor for *Bacillus subtilis* phage SPP1 (197)), there is no evidence to suggest *S. aureus* proteins serve as its phage receptors. Phage interaction protein (Pip) homologs exist throughout the Gram-positives, serving as protein receptors to which phage irreversibly bind (198). There are Pip surface protein homologs anchored to the staphylococcal cell wall through the action of the sortase enzyme in *Staphylococcus* (17, 199). However, neither deletion of the Pip homologs in RN4220 (ST8) (192) nor sortase A in Newman (ST254) (200, 201) affected sensitivity to phage Φ11 and phages ΦNM1, ΦNM2, and ΦNM4, respectively.

Some classes of proteins or extracellular polysaccharides have been shown to block phage adsorption in the staphylococci through occlusion of the WTA receptors. Overproduction of surface protein A in *S. aureus* was shown to reduce phage adsorption through this mechanism (111), but work on surface protein occlusion remains limited. Capsule types 1 and 2 - strains M (ST1254) (187) and Smith diffuse (ST707) (187), respectively - were shown to occlude adsorption (22), but the most common capsule types, 5 and 8, showed inconclusive results (202, 203). Differences in capsule thickness between strains may account for these variable results. Type 1 and 2 strains are mucoid and heavily encapsulated, while type 5 and 8 are non-mucoid despite

encapsulation (21). The CoNS species *Staphylococcus simulans* also showed capsule-dependent inhibition of phage adsorption (204).

The exopolysaccharides (EPS) of staphylococcal biofilms have not been shown to occlude adsorption. Surface proteins, such as biofilm-associated protein (Bap), exopolysaccharides (polysaccharide intercellular adhesin - PIA - composed of poly-N-acetylglucosamine – PNAG – and synthesized by the products of the *icaADBC* operon), and extracellular DNA (eDNA) compose staphylococcal biofilms, which can form by PIA-dependent or protein (Bap)-dependent mechanisms (205, 206). Other surface proteins more common than Bap can also mediate biofilm formation, such as FnbA/FnbB (207, 208) and SasG (209) in *S. aureus* and Aap in *S. epidermidis* (205). Both *S. aureus* (102, 210) and *S. epidermidis* (195, 211, 212) biofilms are susceptible to phage predation. Phage resistance in staphylococcal biofilms may instead be associated with altered biofilm diffusion or metabolism, the latter of which resembles stationary phase growth. Studies on *S. epidermidis* suggested phage susceptibility was similar in biofilms and stationary phase cultures (195). Phages may in fact promote bacterial persistence in *S. aureus* biofilms by releasing nutrients from lysed cells for remaining live ones to utilize (213).

Biosynthesis

Superinfection immunity

Staphylococcal temperate phages encode homologs of the cI repressor (39, 101). In *E. coli*, this protein represses expression of the lytic cycle in newly infecting phages with the same cI protein-binding sites, thus stopping new infections through a mechanism called superinfection immunity. Molecular and evolutionary studies on the *E. coli* phage lambda model suggest many superinfection immunity groups (in which member temperate phages confer immunity to each other upon integration) coexist in nature (214), with cI repressor – operator coevolution driving the emergence of new immunity groups (72). Superinfection immunity as a determining factor in

phage host range in staphylococcal species appears not to have been studied yet, but since prophages are common (most sequenced *S. aureus* genomes contain 1-4 prophages) (101, 215), it may be a significant barrier to phage infection.

Restriction-modification (R-M) systems

Bacteria can resist phage infection by degrading injected phage DNA before it has the chance to replicate and enter the lytic or lysogenic cycle (**Figure 1**). Restriction-modification (R-M) is a prominent phage infection barrier in the *Staphylococcus* genus. R-M systems are modular operons containing combinations of host specificity determinant (*hsd*) genes encoding three types of functions: restriction endonuclease activity (*hsdR*) responsible for destroying unmodified DNA, DNA adenosine or cytosine methyltransferase activity (*hsdM*) responsible for modifying host DNA so that it is not cleaved by restriction endonucleases, and specificity DNA-binding proteins (*hsdS*) responsible for recognizing sequence motifs targeted for cleavage or modification (70).

There are four known types of R-M systems in bacteria, all of which have been found in the staphylococci (216). In type I systems, the restriction enzyme cleaves unmodified DNA adjacent to its binding site, sometimes separated by as much as 1000 bp from the binding site, while the modification enzyme methylates host DNA at the target site specified by the specificity protein. A complex containing all three types of subunits restricts unmodified exogenous DNA, while HsdSHsdM complexes only modify DNA. In type II systems, the restriction enzyme ($HsdR_2$) cleaves unmodified DNA at its binding site, while the modification enzyme (HsdM) modifies DNA at this site. In type III systems, the restriction enzyme cleaves unmodified DNA roughly 24-28 bp downstream from its asymmetric target site, while the modification enzyme methylates a single strand of host DNA at the target site. The modification subunit (Mod) modifies one strand of DNA either by itself (Mod_2) or in complex with the restriction subunit (e.g., Mod_2Res_1 or Mod_2Res_2), while the restriction subunit (Res) cleaves unmodified DNA only in complex with modification subunits (Mod_2Res_1 or Mod_2Res_2). In type IV systems, the restriction enzyme only cleaves

modified, methylated DNA. Type IV systems do not include a modification enzyme. These systems have been well studied in *S. aureus* (and in *S. epidermidis*, to a more limited extent) due to their role in restricting natural horizontal gene transfer and genetic manipulation of the organism (216–219).

Type I R-M systems are the most abundant class of R-M systems reported in *S. aureus*, followed by type IV and then type II systems (216). Type III systems appear to be rare, with only two described in the genus (216). Analyses of the restriction enzyme genomic database REBASE in 2014 showed that all completed *S. aureus* genomes encode a type I R-M system and that most *S. aureus* genomes annotated with R-M genes encode a type I system (216, 220). The most common type I R-M locus found in *S. aureus* is Sau1 (112). Expressing a functional Sau1 *hsdR* gene in restriction-deficient *S. aureus* strain RN4220 greatly reduced electroporation, conjugation, and transduction frequencies (112). *S. aureus* genomes generally encode two Sau1 *hsdS* genes that specify two distinct DNA motif targets for restriction or modification (113). The Sau1 HsdS subunit determines target specificity through its two target recognition domains (TRDs), which each bind to one part of the target sequence (221). TRDs are the least conserved portions of the HsdS amino acid sequences (112), and vary in carriage between strains with lineage and/or clonal complex-specific variant associations, as microarray hybridization studies indicate (112, 113). The Sau1 system prevented transfer of plasmid DNA from one clonal complex (CC5) to another (CC8) with a different target recognition site (113), showing that restriction defines barriers between clonal complexes. Sau1 also affected susceptibility of two CC8 strains (NCTC8325-4 and RN4220 p*hsdR*) but not the *hsdR*-deficient RN4220 to phage Φ75 (siphovirus) propagated in a CC51 strain (879R4RF), suggesting Sau1 can control phage host range (112). Sau1 variation is a powerful marker of lineage/clonal complex (112, 222) and likely drives the independent evolution of clonal complexes. Sau1 would therefore be predicted to be a major host range limitation to phages grown in a strain of a different clonal complex. Since the target sites of nearly

all *S. aureus* Sau1 R-Ms from each of the different clonal complexes have now been identified (221), it should be possible to bioinformatically predict the Sau1-defined clonal complex host range of any sequenced bacteriophage.

Type IV R-M system SauUSI is estimated to be found in 90% of *S. aureus* strains (115, 216) and, in combination with Sau1, presents an effective restriction barrier for resisting phage infection (223). SauUSI specifically restricts DNA methylated or hydroxymethylated at the C5 position of cytosine (115). The preferred binding site for SauUSI is Sm5CNGS, where S represents either cytosine or guanine (115). Type II R-M systems have been estimated to be in ~33% of strains and display a range of target sites (114, 216, 224, 225). The most common type II R-M system found in *S. aureus* is called Sau3A (224). The Sau3A restriction enzyme cleaves 5' to the guanine in unmodified 5'-GATC-3' sequences. The Sau3A modification enzyme, on the other hand, methylates the restriction site at the C5 position of cytosine (226). Some type II systems, such as Sau42I, are encoded by phages. Sau42I is an example of a type IIS R-M system, which binds asymmetric DNA sequences and cleaves outside the recognition site, unlike most type II systems (70). Unlike type I and type IV, type II systems are often carried on mobile genetic elements which are capable of frequent transfer between strains and are not conserved amongst all members of the same clonal complex, so they present a more strain-specific and variable limit to host range (220). Certain *S. aureus* type II R-M systems (e.g., Sau96I) serve to negate the Type IV SauUSI system because they methylate cytosines and guanines in sequences SauUSI targets for cleavage. This is an interesting example of how R-M systems acquired by HGT can have unpredictable interactions with existing systems.

If unmodified phages can survive restriction enzyme degradation upon cell entry, the phage DNA molecules acquire protective DNA methylation as they replicate. While survival of restriction can happen stochastically at high multiplicities of infection, phages have also been shown to have evolved or acquired adaptations for restriction evasion. Anti-restriction

mechanisms include restriction site alteration, restriction site occlusion, indirect subversion of restriction-modification activity, and direct inhibition of restriction-modification systems (227). Restriction site alteration can include both incorporation of alternative bases, such as 5-hydroxymethyluracil (5hmU) and 5-hydroxymethylcytosine (5hmC), and loss of restriction sites through selection. A clear example of the latter in the staphylococci is the elimination of GATC sites in the 140 kb phage K genome, enabling its avoidance of Sau3A restriction (228). Another example is the evolution of particular antimicrobial resistance-carrying conjugative plasmids which have lost specific Sau1 R-M sites allowing their transfer between common MRSA lineages (112). Restriction site occlusion refers to DNA-binding proteins preventing restriction enzymes from binding and digesting DNA (227, 229, 230). R-M subversion either occurs through stimulation of host modification enzymes or destruction of restriction cofactors (e.g., SAM) (227, 231, 232). R-M inhibition occurs most often in type I systems (but also in some type II systems) through the binding of specific anti-restriction proteins, such as ArdA, ArdB, and Ocr (227, 233, 234). There is no literature specifically characterizing anti-restriction in *Staphylococcus*, but an *E. coli ardA* homolog has been identified in the staphylococcal Tn916 and Tn5801 transposons (34).

Clustered regularly interspaced short palindromic repeat (CRISPR) systems

CRISPRs confer immunity to phage infection through the cleavage of extrinsic DNA in a sequence-specific manner. Unlike R-M systems, which target specific DNA sequence motifs, CRISPRs adaptively incorporate target sequences from phages they have destroyed to increase the efficiency of protection. After integrating short segments of foreign DNA as spacers of CRISPR arrays, CRISPR-associated (Cas) nucleases process the transcribed CRISPR array RNA into CRISPR RNAs (crRNAs) used to target new incursions of identical foreign DNA elements for destruction (74, 235). Surveys of *S. aureus* and *S. epidermidis* genomes indicate CRISPRs are not common in these species (236, 237). These surveys looked for the presence of *cas6* and *cas9* genes, which are nucleases required for Type I/III and Type II CRISPR-mediated resistance,

respectively. Cas6 is an endoribonuclease found in Type I and III CRISPR systems that cleaves pre-crRNA transcripts within the 3' end of the repeat region to produce mature guide crRNAs (238, 239), while Cas9 is an endonuclease found in Type II CRISPR systems that cleaves DNA in a crRNA-guided manner (239, 240). Only 12 of 300 published *S. epidermidis* genomes searched encoded the Cas6 nuclease, 18 of 130 *S. epidermidis* isolates from Denmark (Copenhagen University Hospital) tested positive for *cas6* via PCR, and 14 of nearly 5000 published *S. aureus* genomes encoded CRISPR/Cas systems (236). Another study specifically examining *S. aureus* found that 2 of 32 *S. aureus* strains encoded CRISPR/Cas systems (237). These CRISPRs were similar to those found in two *S. lugdunensis* strains, suggesting they were recombined with *S. lugdunensis* or derived from a common ancestor (237). CRISPR/Cas systems have also occasionally been reported in strains of other species (*S. capitis*, *S. schleifer*, *S. intermedius*, *S. argenteus*, and *S. microti*) (236). Only a single *S. aureus* strain has been reported to encode Cas9, which is found in an SCCmec-like region (241). Nonetheless, CRISPR systems have been shown to be important in resisting introduction of foreign DNA in *S. epidermidis* RP62a (242, 243). Anti-CRISPR mechanisms, such as proteins that prevent CRISPR-Cas systems from binding DNA target sites, are being discovered in many phages (244–246), although not yet in those specific for staphylococci. Currently discovered anti-CRISPR mechanisms have been shown to target both type I and type II CRISPR systems (244–247).

Assembly

Assembly interference is the parasitization of superinfecting phage by chromosomal phage-like elements and has been demonstrated experimentally in *S. aureus* pathogenicity island (SaPI)-helper phage interactions. SaPIs encode important virulence factors, such as toxic shock syndrome toxin (TSST), but are only mobilized by superinfecting helper siphoviruses (79, 248). The Dut dUTPase protein expressed by helper phages derepresses the Stl SaPI repressor, activating the SaPI lytic cycle (79). The derepressed SaPIs then take advantage of the

superinfection to proliferate at the expense of the helper phage. SaPIs interfere with helper phage assembly through several mechanisms (60) - remodeling phage capsid proteins to fit the small SaPI genome (117, 118, 249–251), encoding phage packaging interference (Ppi) proteins that prevent helper phage DNA packaging into new SaPI particles (118), and disrupting phage late gene activation (119). All known SaPIs encode phage packaging interference (Ppi) proteins, which divert phage DNA packaging toward SaPIs by inhibiting helper phage terminase small subunits ($TerS_P$) but not corresponding SaPI subunits ($TerS_S$) (118). Ppi proteins are divided into two classes based on sequence that differ in helper phage specificity – Class I interferes with Φ80α and Φ11, while Class II interferes with Φ12 (118). The PtiM-modulated PtiA and the PtiB SaPI2 proteins inhibit expression of the LtrC-activated phage 80 late gene operon (packaging and lysis genes), thus interfering with later steps of the helper phage life cycle (119). The SaPI particles then go on to infect new *S. aureus* hosts, integrating their DNA into the chromosome instead of killing the cell. Helper phages and SaPIs are thought to gain and lose resistance to each other in a 'Red Queen' scenario, given the observed rapid co-evolution of their respective *dut* and *stl* genes (252). SaPIs are found throughout *Staphylococcus* species and beyond; therefore, they may be a common strain-specific modifier of siphovirus infection potential.

Other phage host range limiting factors

Several uncommon or less well-understood mechanisms may contribute to phage host range limitation in *Staphylococcus*. One abortive infection (Abi) system, the eukaryotic-like serine/threonine kinase Stk2, has been characterized in *S. aureus* and *S. epidermidis* (116). In this case, siphovirus infection results in self-induced killing of the host cell, preventing the amplification and spread of phages in the population. Stk2 was found to be activated by a phage protein of unknown function and caused cell death by phosphorylating host proteins involved in diverse core cellular functions. Only *S. epidermidis* RP62A and a few *S. aureus* strains encode Stk2, however, suggesting limited genus-wide importance. The recent long-term evolution study

on *S. aureus* strain SA003 uncovered two genes involved in post-adsorption resistance to myovirus ΦSA012 (196). Missense mutations in guanylate kinase and the alpha subunit of DNA-dependent RNA polymerase conferred resistance but not corresponding decreases in adsorption rate, suggesting some post-adsorption role in resisting infection. More phage resistance systems likely remain undiscovered. A genome-wide association study of 207 clinical MRSA strains and 12 phage preparations identified 167 gene families putatively associated with phage-bacterial interactions (107). While these families included restriction-modification genes, transcriptional regulators, and genes of prophage and SaPI origin, most were accessory gene families of unknown function.

Phage host range in *Staphylococcus* is determined by a hierarchical combination of host factors

In summary, we have described how host range of a *Staphylococcus* phage is determined by a combination of both host and phage-encoded genes, as well as the epigenetic DNA methylation patterns conferred on its DNA from the last strain it infected. Bacterial encoded factors can be conceived as affecting host range at different levels within the species (**Figure 3**). At the highest level, most phages' target for receptor binding (WTA) is highly conserved across *Staphylococcus* species. Strains with unusual WTAs, such as *S. aureus* ST395 and CoNS strains with poly-GroP WTA (108, 120), would be expected to be genetically isolated within the genus. Type I and IV R-M HsdS allotypes and capsule type are conserved between most strains of the same CC but differ between isolates of different CC groups and thus contribute to defining host range in a large subset of *S. aureus* strains. At the level of individual strains, inserted prophages and SaPIs, Stk2, type II systems acquired by HGT, and other as yet unknown functions may all serve to limit host range. We know even less about phage-encoded systems that counteract host resistance. The finding that lytic phages (*Myoviridae* and *Podoviridae*) tend to have broader host

ranges than *Siphoviridae* when challenged against the same set of *Staphylococcus* strains suggests the former encode an array of uncharacterized genes that work against host defenses.

Future directions

Although much progress has been made in the past five decades toward understanding the mechanisms that define staphylococcal phage host range, numerous important questions remain. We need to know more about species other than *S. aureus* and *S. epidermidis*, and even within these species, we need to make sure that rarer and non-methicillin resistant strains are included in studies (253). We also need to ensure that our collections reflect the true diversity of phages that infect *Staphylococcus* species. Even within the two main species only a relatively small number of phages have been tested. This will lead us to consider the questions of phage ecology when understanding what types of phages are found in different environments and with what abundance.

Discovering novel phage resistance mechanisms would aid the effort to understand determinants of host range. Many phage resistance mechanisms have been identified and characterized in other Gram-positives and other bacteria generally but not in the staphylococci. Superinfection exclusion (Sie) and abortive infection (Abi) systems, for example, are well-characterized in the lactococci (68, 254, 255). In addition, a recent publication describes some 26 new anti-phage defense systems identified in bacteria (256), not including the recently discovered bacteriophage exclusion (BREX) and defense island system associated with restriction-modification (DISARM) phage defenses (257–259). Six of the ten verified, newly discovered anti-phage defense systems (Thoeris, Hachiman, Gabija, Septu, Lamassu, and Kiwa) have orthologs in staphylococcal genomes (256).

Understanding phage host range to the point that we can make accurate predictions based on the host genome will be important for developing phage therapies against *Staphylococcus*

strains. Ideally, cocktail formulations for therapy consist of phages with broad, non-overlapping host ranges against the target species (or clonal complex) to be treated. As there are many more genome sequences available than strains that can be tested for sensitivity in the laboratory (e.g > 40,000 for *S. aureus*) (260), with a predictive model we could run *in silico* tests on genome sequences to model the efficacy of the cocktail. With the potential for genome sequencing to be used in the future as a primary clinical diagnostic, we could modify the cocktail to contain phages that specifically target the bacterium causing the infection.

Knowledge of phage host range will also lead us to understand the fitness costs of resistance and its potential trade-offs with virulence and antibiotic resistance of *Staphylococcus*. Strains with null mutations in biosynthetic genes are rare, given WTA's roles in cell division, autolysis, virulence, and antibiotic resistance (18, 182). Although knocking out the genes involved in the first two steps of WTA biosynthesis has no fitness cost in *S. aureus* (at least in laboratory conditions) (261, 262), WTA has many critical physiological roles, especially in environments subject to phage therapy. Staphylococcal WTA is required for nasal colonization (261, 263), cell division (186, 188), regulating autolysis (264, 265), lysozyme resistance through cell wall crosslinking (253, 266), resistance to cationic antimicrobial peptides and fatty acids (49, 267), and biofilm formation (268). WTA-altered or negative phage-resistant mutants would in turn become less virulent (188) and even antibiotic sensitive – highly unfit in the natural habitat colonizing mammalian hosts or in an infection site subject to treatment. Given that methicillin resistance requires WTA (193), phage/beta-lactam combination therapies could be particularly promising. Mutants resistant to either phage or beta-lactams would be sensitive to the other treatment, assuming the infecting strain is sensitive to the phage treatment. Nonetheless, as we note for host range, strains containing minor but fitness-neutral resistance mechanisms, such as R-M systems – rather than costly mutations – may be the most recalcitrant to phage therapy. Staphylococcal phage therapies must then overcome both immediate, emerging mutational resistance and

intrinsic resistance mechanisms (e.g., R-M systems) specific to strains or clonal complexes. These resistance limitations, however, could be overcome by selecting phage host range mutants that escaped host resistance mechanisms, thus isolating more useful phages that would form more effective phage cocktails (269, 270).

Phage-resistant mutants isolated so far, such as those described in the adsorption studies, were typically selected in rich, aerated laboratory medium. The consequences for fitness of the same mutations occurring during *in vivo* infection might be more severe. In addition, both the relevance of various resistance mechanisms *in vivo* and the effect of strain genetic background on resistance selection - especially on a species-wide scale – have been left unexamined in most previous work. One study in mammalian hosts showed that environment altered phage transfer frequency and selection (271), leading to spread of prophage and selection of phage resistance by superimmunity. In laboratory media, phage transfer frequency was lower and spread of prophage was less pronounced (271). It will be important to know both how quickly and in which loci mutations emerge as well as the more general distribution of resistance gene families.

Finally, it is interesting to consider what phage host range studies reveal about the hosts themselves. Staphylococci seem to be unusual among Gram-positives in requiring conserved WTA receptors for attachment and having no reported role for protein receptors. Differences in the outer surface of *Staphylococcus* and/or a feature of the phage ecology within the genus requiring highly conserved receptors may account for this fact. Another interesting question is why CRISPRs play a much-reduced role for intercepting extrinsic phage DNA than R-M systems in this genus compared to other bacteria. It could be that CRISPR systems have a finite capacity for carrying fragments of mobile genetic elements, while R-M systems can attack a wider range of incoming DNA, relevant to rapidly evolving populations. Future studies that probe these questions may reveal some of the differential evolutionary forces that shape the genomes of pathogenic bacteria.

Conclusions

Staphylococcal phage resistance mechanisms have been identified at three stages of infection (attachment, biosynthesis, and assembly) and regulate host range in a hierarchical manner depending on mechanism conservation. Nonetheless, staphylococcal phage-bacterial interactions certainly present many open questions that must be addressed to accurately develop and evaluate possible phage therapies. We need further studies to objectively identify the contribution of individual phage resistance mechanisms to host range. Such work would provide the information needed not only to formulate phage cocktails effective against a wide variety of strains but also to overcome remaining obstacles to cocktail development (e.g., highly effective R-M or Abi systems). Future studies relevant to phage therapy should also characterize phage resistance development during infection and therapy as well as the effects of resistance on mutant fitness. Taken together, this future work will inform the rational design of phage cocktails to treat staphylococcal infections alone or in combination with antibiotics.

Acknowledgments

We thank Michelle Su and Robert Petit for critically reading the manuscript and providing helpful comments. AGM was supported by the National Science Foundation (NSF) Graduate Research Fellowship Program (GRFP). JAL was supported by the Medical Research Council (grant MR/P028322/1). TDR was supported by the National Institutes of Health (NIH) grant R21 AI121860.

Figures

Figure 1: Stages of phage infection and corresponding examples of resistance mechanisms at each stage. Examples not yet identified in the staphylococci are listed in red.

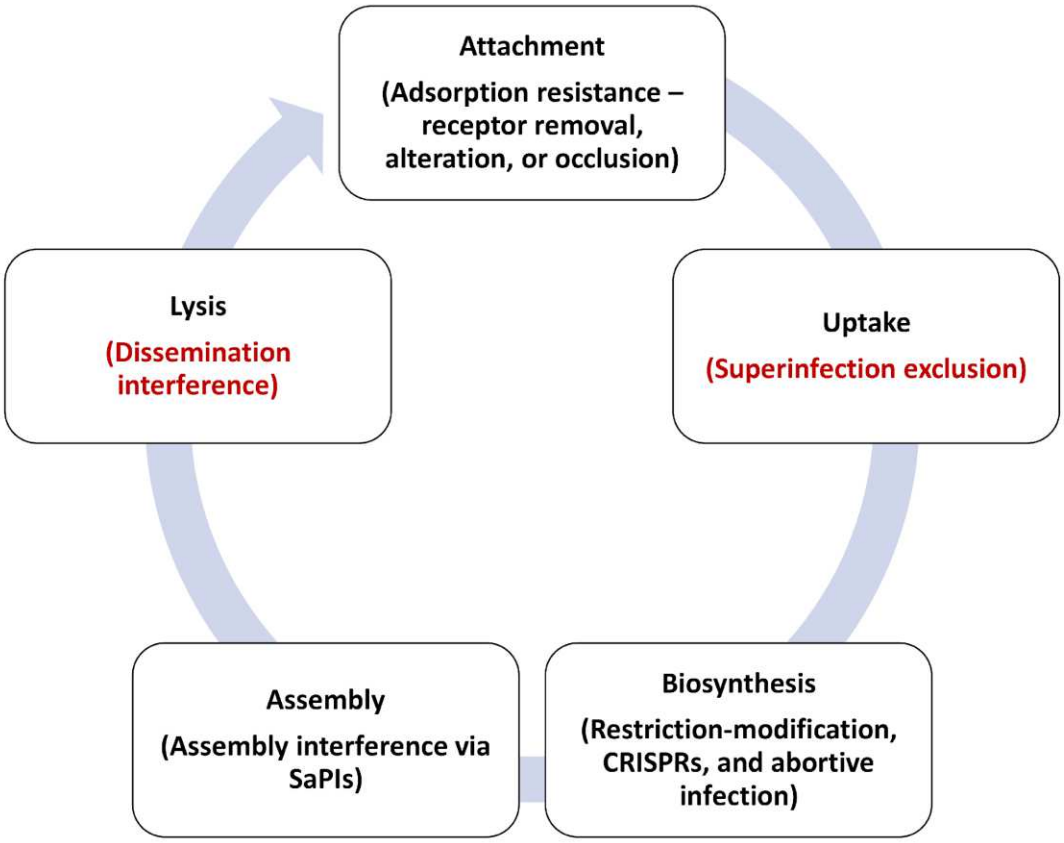


Figure 2: A - The structure of the staphylococcal cell envelope. Lipoteichoic acid is shown in orange (glycerol phosphate), a surface protein in black, wall teichoic acid in orange (glycerol phosphate) and yellow (ribitol phosphate), capsule in blue, and cell wall carbohydrates in green (N-acetylglucosamine – GlcNAc) and purple (N-acetylmuramic acid – MurNAc). Staphylococcal phages bind WTA and/or its ribitol phosphate modifications (i.e., GlcNAc). B – Outline of the wall teichoic acid (WTA) biosynthesis pathway with proteins corresponding to each step listed in the blue arrows. Abbreviations are defined as follows - C_{55}-P, undecaprenyl phosphate; GlcNAc, N-acetylglucosamine; UDP-GlcNAc, uridine-5-diphosphate-N-acetylglucosamine; ManNAc, N-

acetylmannosamine; UDP-ManNAc, uridine-5-diphosphate-N-acetylmannosamine; Gro-P, glycerol phosphate; CDP-Gro, cytidyl diphosphate-glycerol; Rbo-P, ribitol phosphate; CDP-Rbo, cytidyl diphosphate-ribitol; ABC, ATP-binding cassette; and LCP, LytR-CpsA-Psr.

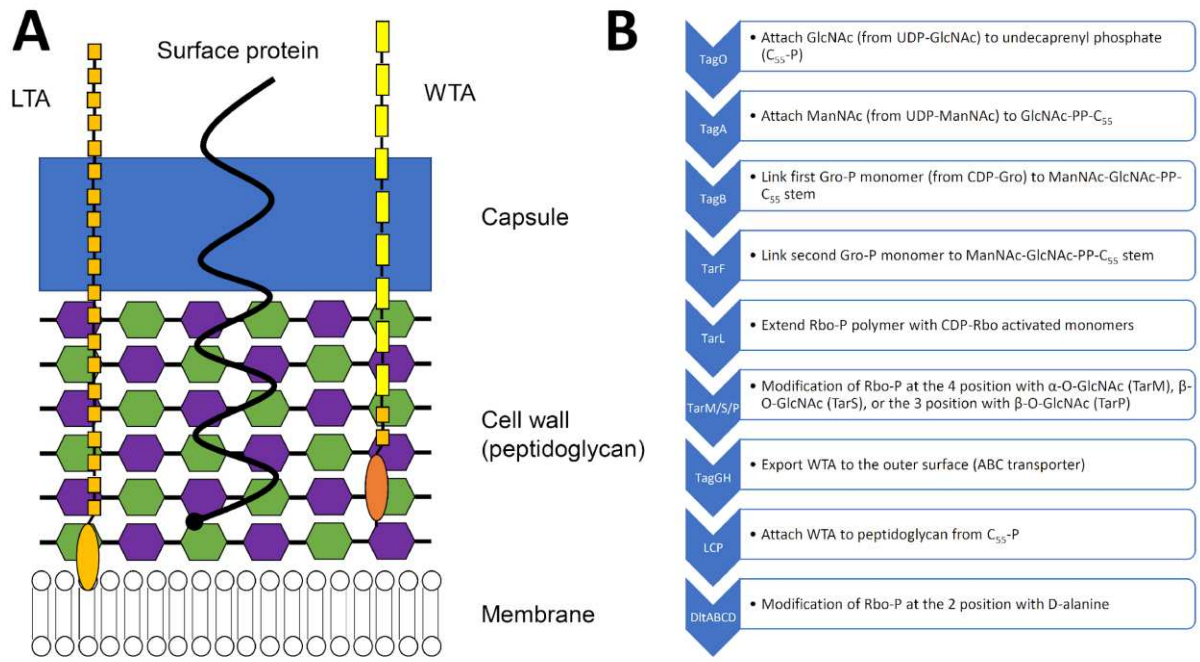
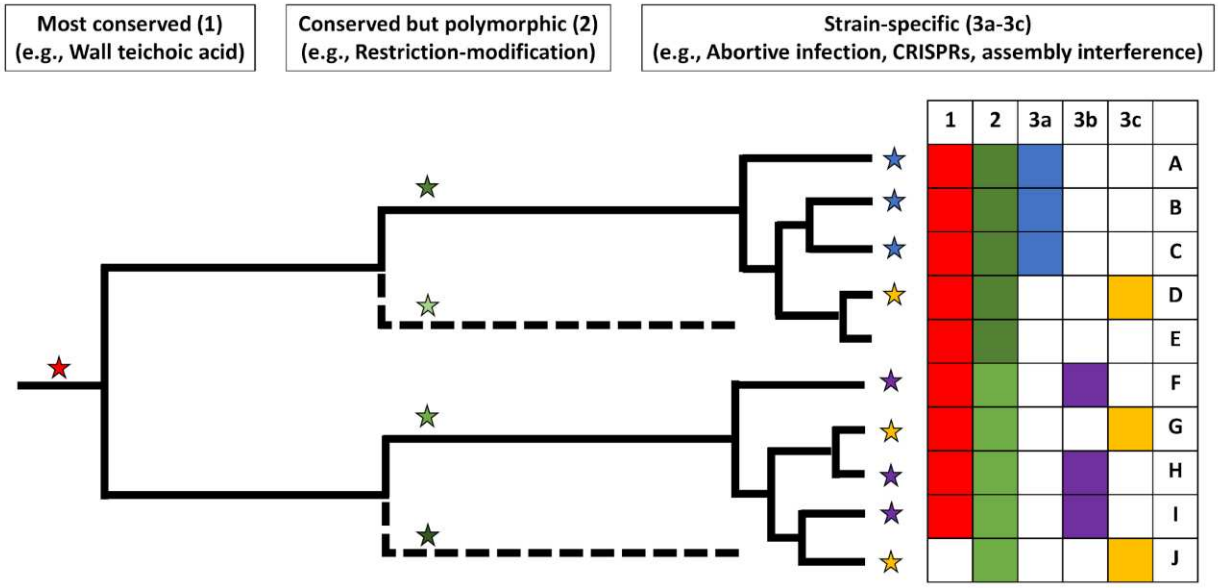


Figure 3: Phage host range for an individual strain is the combination of multiple factors that have different levels of conservation within the species. This is illustrated by a hypothetical phylogenetic tree. Mechanisms can be present throughout strains (1, most conserved – red), present in many strains but with considerable allelic variation (2, conserved but polymorphic – shades of green), or present in a few strains, possibly with allelic variation (3a-3c, less conserved with potential polymorphism – blue, purple, and yellow). Branches where mechanisms evolved by mutation or homologous recombination in the case of 1 and 2 or were acquired by HGT, in the case of 3a-3c, are annotated with colored stars. The table on the right summarizes the mechanisms (1-3c) present in each strain (A-J) using shaded boxes with corresponding colors. Strain J has a mutation that results in the null phenotype for the red mechanism. Host range is the result of the

combination of mechanisms present, so strains A-C as well as F, H, and I would be predicted to have identical host ranges, but phage-specific factors could also introduce variability.



Chapter 3: Genes influencing phage host range in *Staphylococcus aureus* on a species-wide scale

Published in mSphere

Abraham G. Moller[1,2], Kyle Winston[3], Shiyu Ji[4], Junting Wang[5], Michelle N. Hargita Davis[2], Claudia R. Solis-Lemus[6], and Timothy D. Read[2]*

1. Microbiology and Molecular Genetics (MMG) Program, Graduate Division of Biological and Biomedical Sciences (GDBBS), Emory University, Atlanta, GA.

2. Division of Infectious Diseases, Department of Medicine, Emory University, Atlanta, GA.

3. Department of Epidemiology, Rollins School of Public Health (RSPH), Emory University, Atlanta, GA.

4. Eugene Gangarosa Laboratory Research Fellowship, Emory College Online & Summer Programs, Emory College of Arts and Sciences, Atlanta, GA.

5. Department of Statistics, University of Wisconsin-Madison, Madison, WI.

6. Wisconsin Institute for Discovery, Department of Plant Pathology, University of Wisconsin-Madison, Madison, WI.

*corresponding author (tread@emory.edu)

Abstract

Staphylococcus aureus is a human pathogen that causes serious diseases ranging from skin infections to septic shock. Bacteriophages ("phages") are both natural killers of *S. aureus*, offering therapeutic possibilities, as well as important vectors of horizontal gene transfer in the species. Here, we used high-throughput approaches to understand the genetic basis of strain-to-strain

variation in sensitivity to phages, which defines the host range. We screened 259 diverse *S. aureus* strains covering more than 40 sequence types for sensitivity to eight phages, which were representatives of the three phage classes that infect the species. The phages were variable in host range, each infecting between 73 and 257 strains. Using genome-wide association approaches, we identified putative loci that affect host range and validated their function using USA300 transposon knockouts. In addition to rediscovering known host range determinants, we found several previously unreported genes affecting bacterial growth during phage infection, including *trpA*, *phoR*, *isdB*, *sodM*, *fmtC*, and *relA*. We used the data from our host range matrix to develop predictive models that achieved between 40 and 95% accuracy. This work illustrates the complexity of the genetic basis for phage susceptibility in *S. aureus* but also shows that with more data, we may be able to understand much of the variation. With a knowledge of host range determination, we can rationally design phage therapy cocktails that target the broadest host range of *S. aureus* strains and address basic questions regarding phage-host interactions, such as the impact of phage on *S. aureus* evolution.

Importance

Staphylococcus aureus is a widespread, hospital- and community-acquired pathogen, many strains of which are antibiotic resistant. It causes diverse diseases ranging from local to systemic infection and affects both the skin and many internal organs, including the heart, lungs, bones, and brain. Its ubiquity, antibiotic resistance, and disease burden make new therapies urgent. One alternative therapy to antibiotics is phage therapy, in which viruses specific to infecting bacteria clear infection. In this work, we identified and validated *S. aureus* genes that influence phage host range - the number of strains a phage can infect and kill - by testing strains representative of the diversity of the *S. aureus* species for phage host range and associating strain's genome sequences with host range. These findings together improved our understanding of how phage

therapy works in the bacterium and improve prediction of phage therapy efficacy based on the infecting strain's predicted host range.

Introduction

There is no licensed vaccine for *Staphylococcus aureus* and many clinical strains are resistant to multiple antibiotics. For these reasons, alternative treatments such as bacteriophage therapy are being actively investigated (53, 272). Phage therapy has some advantages over using antibiotics. Phages show little or no human toxicity and the high diversity of natural phages available to be isolated for treatment suggests that complete resistance would be hard to evolve (96, 171). However, there is no natural phage known to kill all *S. aureus* strains and for that reason phage cocktails (mixtures of phages with non-overlapping host ranges) are necessary. Rational cocktail formulation requires comprehensive knowledge of the genetic factors that influence phage host range.

S. aureus phages and corresponding known host mechanisms regulating phage resistance and host range have previously been reviewed (39, 53, 101). Known *S. aureus* phages belong to the order *Caudovirales* (tailed phages) and are further divided into three morphological classes - the long, noncontractile-tailed *Siphoviridae*, the long, contractile-tailed *Myoviridae*, and the short, noncontractile-tailed *Podoviridae* (39). The *Siphoviridae* are temperate, while the *Myo-* and *Podoviridae* are virulent (39). The *Siphoviridae* bind either α-O-GlcNAc or β-O-GlcNAc attached at the 4 positions of wall teichoic acid (WTA) ribitol phosphate monomers, while the *Podoviridae* bind only β-O-GlcNAc-decorated WTA, and the *Myoviridae* bind either the WTA ribitol-phosphate backbone or β-O-GlcNAc-decorated WTA (53, 196, 273). *S. aureus* is known to produce poly-ribitol-phosphate rather than poly-glycerol-phosphate WTA (18). WTA biosynthesis genes are conserved throughout the species, with the exception of the unusual sequence type

ST395 (120), as are WTA glycosyltransferases *tarM* and *tarS*, but occasional *tarM* inactivation or absence provides *Podoviridae* susceptibility (109).

Currently identified resistance mechanisms in *Staphylococcus* species act at the adsorption, biosynthesis, and assembly stages of infection (53). Adsorption resistance mechanisms include receptor alteration, removal, or occlusion by large surface proteins or polysaccharides (capsule) (22, 109–111, 181, 185, 196). Biosynthesis resistance mechanisms include halting the infection process through metabolic arrest (abortive infection) and adaptive (CRISPR) or innate (restriction-modification) immunity to phage infection through phage DNA degradation (116, 237, 241, 274, 275). Temperate phage and staphylococcal pathogenicity islands (SaPIs) inserted in the genome may also offer barriers to *Siphoviridae*, through superinfection immunity and assembly interference, which occurs through SaPI parasitization of the infecting viruses packaging machinery (79, 117–119, 251, 276).

While previous studies have identified numerous individual host resistance mechanisms in *S. aureus*, few have examined the importance of these mechanisms on a species-wide scale. In addition, although many *S. aureus* phages are reported to have wide host ranges (102, 103, 173, 174, 277–279), and even early studies suggested staphylococcal phage therapies to be highly effective (129), experiments conducted thus far have failed to explain the genetic bases of host range or resistance development in a species-wide manner. Only one previous study has associated genetic factors (gene families) with phage resistance using a hypothesis-free method (107). This work used a two-step linear regression model to associate some 167 gene families, mostly of unknown function, with resistance assessed in 207 clinical MRSA strains and 12 phage preparations. However, the study did not associate any other types of genetic changes with host range and examined only MRSA strains.

In this study, we associated multiple genetic factors - gene presence/absence, point mutations, and more complex polymorphisms - with *S. aureus* phage host range and resistance in a hypothesis-free, species-wide, and genome-wide manner. We used a novel high-throughput

assay to determine resistance phenotypes of 259 strains challenged with eight *S. aureus* phages belonging to all three morphological categories (*Siphoviridae*, *Myoviridae*, and *Podoviridae*). We then used two bacterial genome-wide association study techniques to identify core genome single nucleotide polymorphisms (SNPs), and subsequences of length k (k-mers) significantly associated with each phenotype and used these significant features to develop predictive models for each phenotype. We also tested for associations between phenotypes and phylogeny, clonal complex (CC), and methicillin resistance (MRSA) and validated novel genes found to be associated with sensitivity or resistance in the GWAS through molecular genetics, thus complementing the hypothesis-free GWAS approach with hypothesis-driven experiments and demonstrating that GWAS-discovered determinants have causative effects on phage resistance.

Materials and Methods

Strains, media, and phage propagation

Phages used in this study were phage p0045 (80α-like), p0017S, p002y-DI p003p-Mourad 87, and p0040-Mourad 2 (*Siphoviridae*); p0006-K and pyo (*Myoviridae*); and p0017-HER49/p66 (*Podoviridae*). All phage genomic DNA was isolated with the bioWORLD Phage DNA Isolation Kit following manufacturer's directions after phage precipitation by a previously described protocol. The corresponding genomes were prepared for sequencing with a 1D ligation sequencing kit (SQK-LSK109) or 1D rapid sequencing kit (SQK-RAD004) and sequenced with an Oxford Nanopore MinION using a Flongle flow cell (FLO-FLG001). Phage p0045, p0017S, p002y-DI, p003p-Mourad 87, p0040-Mourad 2, and p0006-K genomes were also sequenced with Illumina technology by the Microbial Genome Sequencing Center (MiGS) at the University of Pittsburgh.

All *Siphoviridae* and *Myoviridae* were propagated in *S. aureus* RN4220, while the sole podovirus was propagated in *S. aureus* RN4220 *tarM*::Tn, which was constructed by transducing strain RN4220 with Nebraska Transposon Mutant Library (NTML) (280) strain USA300 JE2

tarM::Tn (NE611) phage 0045 lysate. Strains, phages, and plasmids used for phage propagation and molecular genetic validation of GWAS results are listed in **Table 1**. Transduction was performed according to a previously published protocol (281). All overnight cultures were grown in LB/TSB 2:1 supplemented with 5 mM $CaCl_2$ to promote phage adsorption.

Phages were propagated by inoculating a chunk of soft agar containing a plaque and surrounding bacteria into liquid medium. Phage lysates in TMG (Tris-magnesium-gelatin) buffer were spotted (4 μL) on a top agar (0.8% agar, 0.8% NaCl) lawn (5 mL) containing 0.2 mL of a 1:10 dilution of a RN4220 or RN4220 *tarM*::Tn overnight culture (18 hr growth, 37°C, 250 rpm). After overnight growth at 37°C, a chunk of soft agar containing a plaque and surrounding bacteria was inoculated in 35 mL of LB/TSB 2:1 with 5 mM $CaCl_2$. This phage-bacterium co-culture was grown overnight at 37°C and 250 rpm, centrifuged for 20 minutes at 4,000 rpm, and filtered with a 0.45 μm syringe filter before being stored at 4°C. The resulting lysate was titered on RN4220 (*Siphoviridae* or *Myoviridae*) or RN4220 *tarM*::Tn (*Podoviridae*).

Phage resistance/host range assays

259 previously genome-sequenced *S. aureus* strains consisting of 126 from the Network on Antimicrobial Resistance in Staphylococcus aureus (NARSA) repository (NCBI BioProject accession PRJNA289526) (282), 69 strains previously sequenced in a vancomycin-intermediate *S. aureus* (VISA) study (165) (PRJNA239001), and 64 strains previously sequenced in a cystic fibrosis (CF) lung colonization study (283) (PRJNA480016) were rapidly profiled for resistance to the eight phages using a high-throughput assay. Arrayed glycerol (50%) stocks of the strains were used to inoculate 96-well plates containing 200 μL of LB/TSB 2:1 with 5 mM $CaCl_2$ in each well using a 96-pin replicator. Cultures were grown overnight at 37°C and 225 rpm. The following day, overnight cultures were diluted 1:10 in ddH_2O. In order to permit phage adsorption, 10 μL of each phage lysate (~1e9 pfu/mL) was co-incubated with 10 μL of each overnight culture dilution for 30 minutes at room temperature in 96-well plates. 200 μL of molten LB/TSB/$CaCl_2$ agar (LB/TSB 2:1

with 5 mM $CaCl_2$ and 0.4% agar) was then added to each well containing the culture-phage mixtures and allowed to solidify. After incubation overnight (37°C), plates were photographed and final optical densities at 600 nm (OD_{600}) per well measured using a plate reader (BioTek Eon). Strains were categorized as sensitive (0.1-0.4), semi-sensitive (0.4-0.7), or resistant (0.7 or greater) based on classifying average final OD_{600} from at least six replicates into three equal bins (with the third bin counting outlier resistant strains with OD_{600}s above 1). Strains and host range phenotypes (quantitative and quantitative converted to ternary) are listed in Supplemental Tables S1 and S2.

High-throughput assays were also calibrated against a standard spot assay. 108 NARSA strains were tested for resistance to five of the eight phages listed previously (Phage p0045, p0006, p0017S, p002y, and p003p). Briefly, an overnight culture of each strain was diluted 1:10 in ddH_2O and a top agar lawn (0.2 mL dilution per 5 mL molten top agar) was poured on a TSA plate. After solidification, each of the five lysates were spotted (4 μL) twice on the top agar lawn and let to dry. The plates were then incubated face up overnight at 37°C and the spots evaluated for clearing (sensitive), turbid clearing (semi-sensitive), or no clearing (resistant) the following day. High-throughput assay and spot assay phenotypes were compared in boxplots made with ggplot2 (284). The statistical significance of high-throughput assay phage resistance differences between all possible pairs of sensitive (S), semi-sensitive (SS), and resistant (R) strains were assessed with Wilcoxon signed-rank tests.

Bioinformatic processing

Phage p0017 and pyo genomes were assembled from Oxford Nanopore reads with canu 2.0 (140). Hybrid Illumina/nanopore phage genome assemblies were constructed using Unicycler 0.4.8, filtering for contigs with coverage higher than 5x (285). Average nucleotide identity (ANI) was then determined amongst all phage contigs using fastANI 1.31 (286), which is shown as a lower-triangle identity matrix in **Supplemental Table S1**

(https://figshare.com/account/projects/94019/articles/13355909). All *S. aureus* genomes were processed using the Staphopia analysis pipeline (260), which included *de novo* assembly using SPAdes (287) and annotation using Prokka (288). The core-genome phylogenetic tree was constructed by first determining the core genome alignment for all tested strains with Roary (289), correcting for recombination with Gubbins (290), and then generating a maximum-likelihood phylogenetic tree with IQ-TREE (291). Strains (253 total) for which there are corresponding phage resistance phenotypes (quantitative and qualitative), BioProject, BioSample, and SRA accessions, sequence types, clonal complexes, isolation years, and isolation locations are listed in **Supplemental Table S2** (https://figshare.com/account/projects/94019/articles/13355933). MLST (Multi-Locus Sequence Typing) Sequence types were identified for each genome with the mlst command line tool (292), which uses the PubMLST website (https://pubmlst.org/) (24). Quantitative phage resistance phenotypes were annotated on the tree using the Interactive Tree of Life (iTOL) (293).

Preliminary phenotype analysis

Phage resistance phenotypes were initially placed on a core-genome phylogenetic tree and were associated with two factors - clonal complex (CC) and MRSA/MSSA genetic background. Phage resistance associations with CC and MRSA/MSSA were visualized in boxplots made with ggplot2 (284). Statistical significance of phage resistance differences between MRSA/MSSA was determined with Wilcoxon signed-rank tests. Statistical significance of overall phage resistance differences between represented CCs was determined using one-way analysis of variance (ANOVA) tests with or without phylogenetic correction.

Measuring phylogenetic signal

Four different measures of phylogenetic signal were calculated for each phenotype - Abouheif's C_{mean}, Moran's I, Pagel's λ, and Blomberg's K (294). Abouheif's C_{mean} and Moran's I

were calculated with the abouheif.moran function from the adephylo R package (295) while Pagel's λ and Blomberg's K were calculated using the phylosig function from the phytools R package (296). Phylogenetic signal was determined using the core-genome phylogenetic tree annotated with quantitative phage resistance data previously described. Randomization tests for phylogenetic signal calculation were performed with 999 permutations of the data.

Genome-wide association studies (GWAS)

Genotypes were associated with phage host range phenotype data using two different GWAS pipelines - pyseer 1.2.0 (297) and treeWAS 1.0 (298). Pyseer associated clusters of orthologous genes (COGs), core genome single nucleotide polymorphisms (SNPs), and k-mers between lengths 6 and 610 with each phenotype, while treeWAS only associated biallelic core genome SNPs with the phenotype. TreeWAS used the recombination-corrected core-genome phylogeny for population structure correction while pyseer used a conversion of the phylogeny into a kinship matrix. The core genome alignment was rearranged to set N315 as the reference (first sequence). We chose N315 as reference because it was used as a global *S. aureus* reference for the Staphopia project (260). SNPs were called from the core genome alignment with snp-sites (299). For identifying significantly associated genetic determinants, a Bonferroni correction of 0.05/6,058 or 8.25e-6 was set for COG GWAS, 0.05/15,557 or 3.21398e-6 for SNP GWAS, and 0.05/2,304,257 or 2.17e-8 for k-mer GWAS, counting the numbers of intermediate-frequency COGs, biallelic core genome SNPs, and unique k-mers as hypotheses to be tested, respectively.

Pyseer SNP and COG association analyses performed multidimensional scaling (MDS) on a mash distance matrix between tested strains to correct for population structure. Pyseer SNP association was performed with a fixed effect (for variant and covariate lineage) model, the default 10 multidimensional scaling (MDS) dimensions retained, and lineage effect testing on each quantitative phage resistance/host range phenotype for all biallelic core genome SNPs. Pyseer

COG association was performed with a fixed effects model on each phenotype and 9 MDS dimensions retained for intermediate frequency COGs (**Supplemental Figure S1**). Pyseer k-mer association was performed with a FaST-LMM linear mixed (combined fixed variant/covariate lineage and random kinship effects) model on each quantitative phenotype for unique k-mers between 6 and 610 bp in length extracted from genomes of all tested strains. Pyseer k-mer association analyses used a kinship matrix between tested strains constructed from the core-genome phylogeny to correct for population structure and set a minor allele frequency cutoff for analysis of 1%, like SNP and COG analyses. SNP and k-mer association p-values were visualized relative to genetic coordinates using Manhattan plots (with phandango) (300). Associations for all k-mers were assessed for p-value inflation (exceeding the observed/expected p-value diagonal below 1e-2) using Q-Q plots (**Supplemental Figure S2**). Significant SNPs and k-mers were annotated using SnpEff (301) (relative to the Roary N315 core genome sequence) and downstream analysis scripts included with pyseer, respectively, identifying the genes containing the genetic elements (or near the genetic elements, in the case of k-mers) and mutation effects, in the case of SNPs.

TreeWAS was performed for each phage resistance phenotype using the R package with core genome alignment, IQ-TREE core-genome phylogeny, and quantitative phage resistance phenotype as inputs and with default parameters. Significant treeWAS SNPs were annotated using SnpEff (301) relative to the core genome sequence of strain N315 (302).

Functional annotation and network analysis of significantly associated genes

Genes with significant association from the GWAS study (containing SNPs, and either near or overlapping with k-mers) were then used to identify enriched protein functions or possible protein-protein interactions. Gene name lists for each phage were converted to NCTC 8325 RefSeq protein accession lists for use with STRING (303) and PANTHER (304), which depend on NCTC 8325 *S. aureus* accessions. To convert genes containing significant SNPs to NCTC

8325 accessions, Roary N315 core genes were aligned against NCTC 8325 RefSeq proteins with NCBI blastx (1 maximum target sequence, 1 maximum high scoring pair, default e-value). Gene names matching NCTC 8325 RefSeq accessions were converted for each significant SNP using these alignment results. To convert genes containing significant k-mers to NCTC 8325 accessions, all significant genes were aligned against NCTC 8325 RefSeq proteins with blastx (1 maximum target sequence, 1 maximum high scoring pair, default e-value). Gene names matching NCTC 8325 RefSeq accessions were converted for each significant k-mer using these alignment results. Any gene names not mapped to any NCTC 8325 RefSeq protein accessions after this procedure were left unchanged. Lists of significant genes for each phage, for all phage morphological classes (*Siphoviridae*, *Myoviridae*, and *Podoviridae*), and for each life cycle type (virulent or temperate) were used as inputs for STRING and PANTHER. STRING network properties (nodes, edges, average node degree, average local clustering coefficient, expected number of edges, and PPI enrichment p-value) were saved for each input, while PANTHER functional classification and statistical overrepresentation test analyses were performed for each input with respect to molecular function, biological process, cellular component, protein class, and pathway.

Genetic validation of novel phage resistance mechanisms

Six genes (*trpA*, *phoR*, *isdB*, *sodM*, *fmtC*, and *relA*) found to contain significantly associated SNPs or k-mers for any phage resistance phenotype were validated to cause phage resistance changes when knocked out in a single *S. aureus* genetic background (USA300 JE2). Transposon insertion mutants in each gene were selected from the Nebraska Transposon Mutant Library (NTML) (280) and backcrossed into USA300 JE2 through the transduction method previously described (281) to eliminate any possible secondary acquired mutations. Backcrossed mutants were then complemented with each gene cloned into the vector pOS1-P*lgt* (305). Relevant strains (selected mutants and complemented strains) are listed in **Table 1**. Growth

curves were performed on all listed strains (**Supplemental Figure S4**). USA300 JE2, respective transposon mutants, empty vector controls, or complemented mutants were inoculated with a 96-pin replicator from arrayed frozen glycerol stocks into 96-well plates containing 200 µL LB/TSB 2:1 with 5 mM $CaCl_2$ or the same medium supplemented with 10 µg/mL chloramphenicol in each well. We then diluted each culture 1:100 in fresh LB/TSB 2:1 with 5 mM $CaCl_2$ or the same medium supplemented with 10 µg/mL chloramphenicol and collected growth curves on a BioTek Eon plate reader (37°C, 225 rpm agitation, OD_{600} measured every 10 minutes).

Genes were cloned into pOS1-P*lgt* either through splicing overlap extension (SOE) PCR (*trpA*, *phoR*, and *sodM*) or through NEB HiFi assembly (*isdB*, *fmtC*, and *relA*). Each gene and pOS1-P*lgt* were amplified with the primers listed in **Supplemental Table S3** (https://figshare.com/account/projects/94019/articles/13355939) to create overlap into the corresponding fragment using NEB Q5 High-Fidelity DNA Polymerase according to manufacturer's directions. All genes were amplified from USA300 JE2 genomic DNA except for fmtC, which was amplified both from USA300 JE2 and NRS209. Genes were cloned into the same site downstream of the P*lgt* promoter. For SOE PCR, AMpure XP bead-purified gene and vector fragments were mixed together at a ratio of 1:59 and amplified for 20 cycles with NEB Q5 High-Fidelity polymerase at an annealing temperature of 60°C. For HiFi assembly, purified gene and vector fragments were mixed together at a ratio of 1:2 (less than 0.2 pmol DNA total) and incubated with NEBuilder HiFi DNA Assembly Master Mix for 3 hours at 50°C. SOE PCR and HiFi assembly products were transformed into NEB DH5α competent cells (High Efficiency), plated on LB agar with ampicillin (100 µg/mL), and grown overnight at 37°C. Transformants were verified by colony PCR with respective LF and RR primers listed in **Supplemental Table S3** (https://figshare.com/account/projects/94019/articles/13355939). Plasmids were extracted from verified transformant overnight cultures with the Promega PureYield Plasmid Miniprep System.

These plasmids were then transformed into *E. coli* IM08B (306) to improve electroporation efficiency into the USA300 JE2 transposon mutants.

Electrocompetent *S. aureus* cells (USA300 JE2 transposon mutants) were prepared as previously described (307). *S. aureus* electrocompetent cells were electroporated with 2 µg of ethanol-precipitated plasmid DNA (empty vector and vector with insert corresponding to transposon insertion). Electrocompetent cells were first thawed, centrifuged, and resuspended in 50 µL 10% glycerol/0.5 M sucrose. After adding plasmid DNA, cells were transferred to 0.1 cm electroporation cuvettes and pulsed at 2.1 kV, 100 Ω, and 25 µF. Immediately after electroporation, 1 mL of TSB/0.5 M sucrose was added to the cuvette and the culture was transferred to an Eppendorf tube to recover for 90 minutes at 37°C and 250 rpm. Dilutions of the outgrowth were plated on TSA with chloramphenicol (10 µg/mL) and grown overnight at 37°C. Electroporants were verified by colony PCR with respective LF and RR primers listed in **Supplemental Table S3** (https://figshare.com/account/projects/94019/articles/13355939).

pOS1 *fmtC* and *relA* were introduced into USA300 JE2 transposon mutants, however, by transduction from RN4220. *S. aureus* RN4220 was electroporated with pOS1 *fmtC* (USA300), pOS1 *fmtC* (NRS209), and pOS1 *relA* plasmids according to the procedure described previously. Plasmids were then transduced from RN4220 to USA300 JE2 transposon mutants according to a procedure previously published (281). Briefly, a recipient strain was infected with donor phage at a MOI of 0.1 after supplementing with $CaCl_2$. The infected culture was then outgrown in TSB supplemented with sodium citrate to prevent phage lysogeny. The outgrowth culture was plated on TSA supplemented with both chloramphenicol (10 µg/mL) and sodium citrate (40 mM) to select for plasmids and inhibit lysogeny, respectively.

Mutants and their complemented derivatives were assessed for phage resistance and host range both through the high-throughput assay described previously and the efficiency of plating (EOP) assay (179) to assess bacterial growth in the presence of phage and phage plaquing efficiency, respectively. The high-throughput host range assay was performed as described

earlier, but strain overnight cultures were grown in LB/TSB 2:1 with 5 mM $CaCl_2$ supplemented with chloramphenicol (10 µg/mL) to maintain plasmid selection in the case of complemented strains for this and the EOP assay. The EOP assay was performed by spotting 4 µL of neat through 1e-8 dilutions of phages p0045, p0006, p0017, p0017S, p002y, p003p, p0040, and pyo on lawns (0.2 mL of a 1:10 overnight culture dilution mixed with 5 mL of top agar) of a test and reference (USA300) strain. Lawns were poured on TSA plates. EOP was calculated by dividing phage titer on the test strain by that on the reference strain.

Additional experiments on the *trpA* mutant set and phage p003p examined bacterial survival after performing the phage/culture soft agar coincubation of the high-throughput assay. The high-throughput assay was performed as described earlier for six replicates of USA300, USA300 *trpA*::Tn, USA300 *trpA*::Tn pOS1, and USA300 *trpA*::Tn pOS1 *trpA* strains. Corresponding ODs were recorded as described for the high-throughput phage host range assay (**Supplemental Figure S5A**). Agar plugs were then removed with toothpicks, placed in 0.8 mL volumes of sterile TMG, and broken apart by vortexing. The resuspensions were then serially diluted in TMG and 4 uL of 1e-1 through 1e-6 dilutions were spotted four times on TSA plates. Dilution plates were grown overnight at 37°C and colonies counted the following day to determine surviving CFU in each condition (**Supplemental Figure S5B**).

Construction of phage resistance phenotype predictive models

Phage resistance predictive models were constructed using three methods - random (decision) forests, gradient-boosted decision trees, and neural networks. Random forests were generated using the randomForest R package, gradient-boosted decision trees were generated with the XGBoost R package (308). Ternary (S, SS, or R) phenotypes converted from the original high-throughput assay quantitative phenotypes (described in the phage host range assay methods) were set as the response variable, while either presence/absence of each significant genetic element, each k-mer, or one of the previous two sets (all elements or just k-mers) and

both strain sequence type (ST) and clonal complex (CC) were set as predictor variables. Random forest and XGBoost predictive accuracy and receiver operating characteristic (ROC) area under the curve (AUC) was determined on the validation set through multiple replicates of 10-fold cross-validation, in which alternating tenths of the data are used for validation while the model is trained on the remaining data. The optimal number of rounds (iterations) for XGBoost was determined for each phage and set of input predictor variables with 5-fold cross-validation. XGBoost model training also used the softmax objective for multiclass (three classes - S, SS, and R) classification.

Neural network model construction was more complicated as it involved a preprocessing step to balance datasets where necessary. Oversampling or a combination of over- and undersampling methods was performed to balance specific datasets. For the oversampling method, new samples of the minority classes were randomly generated with replacement so that the number of samples for each class would be equal to that of the majority class in the original dataset. For the combination method, the Synthetic Minority Over-sampling Technique (SMOTE) for over-sampling and Tomek links for under-sampling were performed together. However, for phages with limited cases for one class type, such as p002y, we cannot conduct undersampling. Therefore, for such datasets, only the oversampling method was performed. Then the new balanced datasets were split into training and validation sets with 30% validation. Random splits were performed four times to generate four replicates for evaluation, each with different train and test datasets. Each replicate was evaluated as before with validation set prediction accuracy and ROC AUC.

Neural network models were constructed three ways - 1) with or without oversampling or an over-/undersampling combination alone, 2) also with a regularizer and dropout layer, or 3) also with lasso regression for feature selection. All methods use ADAM (309) for optimizing and sparse categorical cross-entropy for loss. For imbalanced datasets, the oversampling and combination over/undersampling methods were used as well, if possible. The fully connected neural network was constructed based on the selected, balanced dataset. We then found both training and

prediction accuracy to evaluate performance for each network model. We note that network models were optimized for each replicate training set, which means there may be different network models for the four replicates. In the first method, fully connected neural network models were constructed on datasets either originally balanced or balanced after over-/combination methods, with no further correction. Since some network models have high prediction accuracies, it is possible that these models are overfitting, so the second method adds a regularizer and a dropout layer to fully connected neural networks as new models. Finally, for some network models, the prediction accuracies were not as high as others. Thus, in the third method, lasso regression was performed to select important features and improve performance. A neural network model was constructed on the new dataset based on these selected features.

Information entropy was compared to average randomForest and XGBoost 10-fold cross-validation and neural network predictive accuracies and ROC AUCs after calculation through using the following equation (310), where $P_x(x_i)$ is the probability of event x_i, and the three possible events are S, SS, and R phenotypes:

$$H(x) = -\sum_{i=1}^{n} P_X(x_i)\, ln\, (P_X(x_i))$$

Results

Development of a novel high-throughput host range assay

In order to evaluate host range on a large number of *S. aureus* strains in a quantitative manner, we developed a high-throughput host range assay (**Figure 1**) described in the Materials and Methods section. This assay measures the extent that phages cause retardation of growth compared to a control. Before using data from the high-throughput assay for further analysis, we calibrated it against the traditional spot assay (**Figure 1A**), which measures whether phages

cause lysis in a lawn of bacterial cells. We compared spot assay results (sensitive - S, semi-sensitive - SS, or resistant - R) for 108 strains and five phages to the strains' average final soft agar turbidity (OD_{600}) in the high-throughput assay (**Figure 1B**). For all phages tested, turbidity was significantly higher ($p<0.05$, Wilcoxon signed-rank test) for spot-resistant strains relative to spot-sensitive strains. For all phages tested but p003p, the turbidity was significantly higher for spot-semi-sensitive strains relative to spot-sensitive strains. However, for only phages p0006 and p003p were turbidities significantly higher for spot-resistant strains relative to spot-semi-sensitive strains. Thus, for all phages but p003p, it was possible to tell spot-sensitive from spot-semi-sensitive strains by the high-throughput assay, but only for phages p0006 and p003p was it possible to tell spot-semi-sensitive from spot-resistant by the new assay. Overall, these results showed strong agreement between the lysis-based spot assay and the high-throughput growth-based assay for differentiating between sensitive and resistant/semi-sensitive phenotypes.

Host range is associated with clonal complex but not methicillin resistance

We evaluated the host range of eight phages belonging to the *Siphoviridae*, *Myoviridae*, and *Podoviridae*. *Siphoviridae* (p0045, p0017S, p002y, p003p, and p0040), *Myoviridae* (p0006 and pyo), and *Podoviridae* (p0017) were most closely related to others of the same class, but not related at all to those of other classes (**Supplemental Table S1** - https://figshare.com/account/projects/94019/articles/13355909). Amongst the *Siphoviridae*, p003p was the most divergent from the others (between 97.75 and 97.83% similar to the others). On the host side, host range was determined on a set of 259 *S. aureus* strains representing 47 already-defined sequence types (STs) and 17 already-defined clonal complexes (CCs) against eight phages (253 strains with sequence data are included in **Supplemental Table S2** - https://figshare.com/account/projects/94019/articles/13355933). The most common STs were 5 (25.69%), 8 (13.04%), 30 (6.72%), 105 (4.35%), and 121 (3.16%), while the most common CCs were 5 (37.15%), 8 (23.32%), 30 (12.25%), 121 (5.14%), and 1 (4.74%), respectively. The most

common strain isolation years were 2005 (31.92%), 2012 (14.08%), 2002 (12.68%), 2017 (7.51%), and 2018 (7.04%), while the most common isolation locations were the United States (61.26%), France (19.76%), the United Kingdom (11.46%), and Japan (1.19%). Strain isolation years ranged from 1935 to 2018.

Phages p0045, p0040, the two temperate phages, and p0017, the sole tested podovirus, had the highest proportions of resistant strains (71.8, 38.2, and 35.9%, respectively) amongst those tested (**Figure 2A** and **Table 2**). The average and median final turbidities amongst tested strains were likewise highest for these phages (0.80/0.88, 0.61/0.60, and 0.56/0.54, for p0045, p0040, and p0017, respectively). On the other hand, phages p0017S, p002y, p003p, and pyo, all virulent *Sipho*- or *Myoviridae*, had the lowest proportions of resistant strains (0.8, 1.2, 1.2, and 1.5%, respectively) and average/median final turbidities (0.31/0.27, 0.27/0.22, 0.32/0.31, and 0.26/0.21, respectively). Phage p0006 had an intermediate proportion of resistant strains (15.4%) and average/median final turbidity (0.49/0.44). Strains were resistant to between zero and six phages (**Figure 2B**), with a median of two. The strains NRS148, NRS209, and NRS255 were resistant to six phages, the most among any strains. Phage host ranges were most similar (concordant - defined by number of strains with identical phenotypes between two phages) between phages p0017S, p002y, p003p, and pyo, but least similar between phage p0045 and the previous set of four phages (**Figure 2C**).

We also examined whether there were significant associations between clonal complex (CC) or methicillin resistant *S. aureus* (MRSA) genetic background and each phage host range phenotype (**Figure 3**). We hypothesized CC would correlate with host range given that type I restriction-modification specificity is strongly associated with CC (112, 113), restricting the infection of a strain by phage propagated in a strain of a different CC. We hypothesized MRSA genetic background may also affect host range, given that the phage receptor WTA is required for methicillin resistance (193) but MRSA strains can tolerate more defects in WTA biosynthesis than MSSA strains (311). However, MRSA/MSSA phenotypic differences were only significant for

phage 0040 ($p<0.001$, Wilcoxon signed-rank test; **Figure 3B**). There were significant differences in phage resistance between individual CCs for all phages ($p<0.05$, Tukey Honest Significant Differences based on one-way ANOVA; **Supplemental Table S4** - https://figshare.com/account/projects/94019/articles/13355942) and significant overall differences amongst all CCs (one-way ANOVA) for all phages ($p<0.05$). Overall, these results indicate MRSA genetic background for the most part is not associated with the host range of these phages, while CC overall affects all tested phages' host ranges.

Resistance to each phage is highly homoplasious, emerging independently in multiple CCs (**Figure 4**). We estimated phylogenetic signal by calculating Moran's I, Abouheif's C_{mean}, Pagel's λ, and Blomberg's K (294) for each phage host range phenotype, which resulted in statistically significant values in every case (**Table 3**). Both Moran's I and Abouheif's C_{mean} values fell between 0.17 and 0.37. Pagel's λ values all were nearly 1, while Blomberg's K values approached 0. Pagel's λ values around 1 and Moran's I/Abouheif's C_{mean} values around 0 support a Brownian motion model (the phylogeny structure alone best explains the trait distribution), but Blomberg's K values around 0 suggest trait variance at the tips is greater than that predicted by the phylogeny under a Brownian motion model. All calculated phylogenetic signal values were statistically significant ($p<0.05$ for randomization tests based on 999 simulations). Taken together, these results suggest the structure of the phylogeny could explain the host ranges of the tested phage as expected under a Brownian motion model (random distribution of phenotypes amongst strains directed by the phylogeny overall). This neutral phylogenetic signal agrees with the previous finding that CC is associated with host range (**Figure 3A** and **Supplemental Table S4** - https://figshare.com/account/projects/94019/articles/13355942). While there is a CC association with host range, strain-specific effects may be even stronger than CC-specific effects, resulting in weak net phylogenetic signals.

GWAS reveals novel genetic determinants of host range

We used the GWAS tools pyseer (297) and treeWAS (298) to identify genetic loci strongly associated with the phage host range phenotype (**Supplemental Figure 3A, Table 4**). We chose these tools because they represent two alternatives for population structure correction - identifying principal components of a distance matrix (pyseer) and testing against phenotypes simulated based on the phylogeny (treeWAS). pyseer identified COGs, SNPs, and k-mers beyond the respective multiple-corrected significance thresholds in all phages. Most phages lacked k-mer p-value inflation with the exceptions of p0017S, p002y, and p003p, based on associated Q-Q plots (scatter above the diagonal at p-values of 1e-2 or more indicated p-value inflation; **Supplemental Figure S2**). The number of significant COGs detected ranged from 48 (p0017S) to 347 (pyo). Significant SNPs were detected for all phages but p0045 and p0017S and ranged from 1 (p0017) to 249 (pyo). Significant SNPs were identified in *tarJ* (pyo - 672A>G synonymous) and *tagH* (p002y - 848T>C missense and 873A>T missense; pyo - 848T>C missense, 873A>T missense, and 876C>T synonymous). TarJ is responsible for activating ribitol phosphate with CTP to form CDP-ribitol (194), while TagH is a component of the ABC transporter that exports WTA to the cell surface (18). A substantial number of the significant p0017 k-mers (1,382, -log(p-value) = 12.259) mapped to the recently discovered host range factor *tarP*. TarP was shown to confer podovirus resistance by transferring N-acetylglucosamine to the C3 position of ribitol phosphate (110). Significant k-mers also mapped to *hsdS* (32 for p002y, -log(p-value) = 9.33; 6 for p003p, -log(p-value) = 8.54), *oatA* (2 for p002y, -log(p-value) = 7.75; 3 for p003p, -log(p-value) = 8.45), and *tagH* (11 for p002y, -log(p-value) = 9.47; 10 for p003p, -log(p-value) = 8.81). HsdS determines the sequence specificity of Sau1 restriction-modification (112), while OatA, or peptidoglycan O-acetyltransferase, is required for phage adsorption at least in *S. aureus* strain H (184). Prophage-associated genes (186 k-mers for phage tail fiber gene SRX477019_02350 for phage p0045, -log(p-value) = 12.21; 37 k-mers for same gene for p0040, -log(p-value) = 8.69) were the most

significantly associated with two of the tested *Siphoviridae* – phage p0045 and p0040. This result agrees with the known temperate phage resistance mechanism of superinfection immunity, in which prophages express a repressor gene that prevents transcription of superinfecting phages' lytic genes (72).

TreeWAS detected 4 or fewer significant SNPs for three phages and none for phages p0045, p0017, p0017S, and p002y. Amongst significant SNPs, the majority were synonymous for each phage, with the exception of phage p0040 (**Supplemental Figure 3B**). A single nonsense mutation was detected for phage p002y. The number of significant k-mers in or near a gene detected ranged from 14 (pyo) to 7078 (p0017).

Searches using the entire set of GWAS loci for potential enriched protein-protein interactions and pathways in the STRING (303) and Gene Ontology (304) databases (using the PANTHER tool) (**Supplemental Figure 3A, Supplemental Tables S5** - https://figshare.com/account/projects/94019/articles/13355945**, S6** - https://figshare.com/account/projects/94019/articles/13355948**, and S7** - https://figshare.com/account/projects/94019/articles/13355951), resulted in a biologically diverse group of functions. These included periplasmic substrate-binding (p0017S, STRING), type I restriction-modification specificity (p0017S, STRING), metal ion binding (p002y, STRING; pyo, STRING and PANTHER), ATP binding (p002y, STRING and PANTHER; pyo, STRING), amino acid metabolism (pyo, STRING and PANTHER), pyrimidine metabolism (pyo, STRING), and RNA metabolism (p0045, PANTHER). We note that the search results are limited to genes present in NCTC 8325 and must be interpreted accordingly.

Confirmation of causal roles for novel determinants of host range

We next used molecular genetic experiments to confirm a causal role for genes discovered in the GWAS where there were no previous references in the literature for a role in *S. aureus* phage host range. The genes (*trpA* - p002y/pyseer, *phoR* - p002y, p003p, p0040/pyseer, *isdB* -

p002y, p0040/pyseer, *sodM* - p002y, p003p/pyseer, *mprF/fmtC* - p002y/pyseer, and *relA* - p003p/pyseer) were selected for validation because there were available transposon mutants in the Nebraska Transposon Mutant Library (NTML) (280) and these mutants could be backcrossed into the wild-type USA300 to eliminate second site mutations. We thus could not use transposon mutants that would confer full resistance (e.g., insertions in wall teichoic biosynthesis genes *tarJ* or *tagH*) as this resistance to phage infection would prevent lysate preparation for backcrossing. Nonetheless, we backcrossed selected mutants into their isogenic background USA300 JE2 and complemented these strains with the multicopy vector pOS1-P*lgt* (305).

We assessed the USA300 JE2 background, transposon mutants, transposon mutants with empty vectors, and complemented transposon mutants for growth defects and phage resistance with the previously described high-throughput (**Figure 5** and **Supplemental Figure S6**) and efficiency of plating (EOP) assays (**Figure 5** and **Supplemental Figure S7**), respectively. No strains had growth defects respect to each other or the wild-type background (**Supplemental Figure S4**). We found significant decreases in phage resistance for all mutants in the presence of phages p0006, p0017S, p003p, and p0040 ($p<0.05$, Wilcoxon signed-rank test). However, when we attempted to rescue the phenotype by complementation, we only found corresponding rescue of phage resistance back towards the wild-type phenotype in *trpA*, *phoR*, *sodM*, and *fmtC* ($p<0.05$, Wilcoxon signed-rank test). Interestingly, the *fmtC* allele from NRS209 did not complement the *fmtC*::Tn insertion, while the *fmtC* allele from the same strain (USA300 JE2) did, suggesting allele specificity for *fmtC* in phage resistance effects. As found in growth curves (**Supplemental Figure S4**), in the high-throughput assay, for the most part, mutations and plasmids did not affect bacterial growth in the absence of phage (no phage panel in **Figure 5** and **Supplemental Figure S6**). We further evaluated bacterial survival after the high-throughput assay by measuring CFUs in assay soft agar after overnight culture for the trpA set of strains and phage p003p. As expected, surviving CFU correlated with final OD, with significantly ($p<0.05$, Wilcoxon signed-rank test) higher CFU and OD for USA300 JE2 than USA300 *trpA*::Tn and

USA300 *trpA*::Tn pOS1 *trpA* than USA300 *trpA*::Tn pOS1, respectively (**Supplemental Figure S5**).

We did not observe any significant changes in phage propagation efficiency when performing the efficiency of plating (EOP) assay on these strains, except for USA300/USA300 *trpA*::Tn pOS1 *trpA*, USA300/USA300 *phoR*::Tn pOS1, and USA300/USA300 *relA*::Tn pOS1 *relA* ($p<0.05$, Wilcoxon signed-rank test). EOP measures differences in plaquing, or actual infection and phage propagation. The growth-based assay measures survival despite infection. We interpreted the different results between the EOP and growth assays to indicate that these genes mostly influence survival post-infection but do not necessarily prevent infection. Taken together, these results confirmed at least six GWAS-significant genes are implicated in phage resistance for some of the eight phages but not necessarily at the level of direct interference with phage propagation.

Host range predictive models based on significant genetic determinants explain most phenotypic variation

In order to determine the extent to which host range is predictable by the loci identified by GWAS, we constructed predictive models for qualitative host range phenotypes using random forests, gradient-boosted decision trees, and neural networks. We determined predictive accuracy for each phage host range phenotype and four different sets of predictors (presence/absence of significant genetic determinants or k-mers from GWAS result, with or without sequence type and clonal complex for corresponding strains) with 10-fold cross-validation (**Figure 6A and Supplemental Figure S8A**). In no cases were there significant differences in 10-fold cross-validation predictive accuracies between model construction methods or predictor sets used, suggesting no combination of method and predictors improved model predictive accuracy relative to another and that there is a limit to the amount of host range variation explained by the predictive models. The phages p0017S (0.83-0.87), p002y (0.81-0.88), p003p (0.83-0.92), and

pyo (0.83-0.91) had the highest average predictive accuracies, followed by p0045 (0.67-0.73), p0006 (0.47-0.61), p0040 (0.42-0.61), and p0017 (0.45-0.54), respectively. We hypothesized that predictive accuracy correlated with host range distribution, expecting simpler distributions to be easier to predict and thus to have higher predictive accuracies. We thus examined the relationship between information entropy (average level of uncertainty or information in a variable's possible outcomes) and predictive accuracy (**Figures 6B and 6C**; **Supplemental Figures S8B and C**). We found that predictive accuracy increased at the extremes of phenotype proportion (S, SS, R) and that information entropy was negatively correlated with predictive accuracy for all models.

We also performed the same analyses on another predictive model statistic - the receiver operating characteristic (ROC) curve area under the curve (AUC), which measures the ability of the model to distinguish between classes (true positive and true negative). We found that gradient-boosted decision trees AUCs held uniform amongst phages, while random forest and neural network AUCs negatively correlated with information entropy (**Supplemental Figures S9** and **S10**), suggesting phenotype complexity (entropy) did not affect the robustness of gradient-boosted decision tree prediction. Taken together, these results show that significant GWAS determinants from this study do not completely predict phage host range and that prediction is most effective for low complexity host range distributions, at least for random forest and neural network models.

Discussion

Through GWAS on a diverse natural set of *S. aureus* strains we discovered numerous genetic determinants of phage host range, many of which had not been reported previously in the scientific literature. This study uses a far more diverse set of strains than the previous hypothesis-free study of *S. aureus* phage host range (107). However, our set of genetic loci still only partially

explains the variation in the overall broad host ranges of our tested phages, as the predictive modeling results indicate.

We found that knockouts of six GWAS-significant genes - *trpA*, *phoR*, *isdB*, *sodM*, *fmtC*, and *relA* - increased phage sensitivity, suggesting these could be targets for phage-therapy adjunctive drugs. TrpA together with TrpB (encoding tryptophan synthase alpha and beta chains, respectively) carries out the last step in L-tryptophan biosynthesis (312). The enzymes convert indole-glycerol phosphate and serine to tryptophan and glyceraldehyde 3-phosphate (312). TrpA inactivation might then sensitize *S. aureus* to phage infection by increasing indole-glycerol phosphate levels at the expense of tryptophan. In the absence of *trpA*, built-up tryptophan biosynthesis intermediates including IGP may sensitize cells to phage infection, making *trpA* necessary for resistance. Alternatively, by reducing the total tryptophan pool, removing tryptophan biosynthesis may increase the proportion of tryptophan used to translate phage relative to host proteins, thus enhancing phage infection at the cost of host growth. Indeed, it is already known that throttling down protein synthesis with sublethal doses of ribosomal active antibiotics enhances plaque formation on MRSA lawns (313).

The PhoPR two-component system is responsible for regulating expression of phosphate uptake systems (ABC transporters) based on phosphate levels. In *S. aureus*, PhoPR is necessary for growth under phosphate-limiting conditions by regulating either phosphate transporters or other factors, depending on the environment (314). In *Bacillus subtilis*, the sensor kinase PhoR senses phosphate limitation through wall teichoic acid intermediates (315) and correspondingly represses WTA biosynthesis gene expression (316). PhoPR also upregulates glycerol-phosphate WTA degradation in *S. aureus* and *B. subtilis* to scavenge phosphate (317, 318). If all these mechanisms are present in *S. aureus*, and if there is also a pathway for degrading *S. aureus* Rbo-P WTA, PhoR activity may lead to reduced WTA under phosphate starvation, thus forming phage resistant cells. On the other hand, as for *trpA*, *phoR* might simply be required for properly inducing phosphate uptake necessary for survival during phage infection.

Superoxide dismutase (SodM) and phosphatidylglycerol lysyltransferase/multiple peptide resistance factor (FmtC/MprF) more likely have direct mechanistic roles in the phage infection process. SodM may be required for tolerance to cell wall stress imposed by phage infection. SodM is a Mn/Fe-dependent superoxide dismutase that converts superoxide into hydrogen peroxide and oxygen. Previous studies have shown that superoxide dismutase has affected tolerance to cell wall active antibiotics in *S. aureus* and *E. faecalis* (319, 320) and phage plaquing in *C. jejuni* (321). Superoxide dismutase transcripts were found to be upregulated upon phage infection in *E. faecalis* (322). FmtC, on the other hand, may affect the lysis step by altering cell surface charge. FmtC (MprF) alters cell surface charge first by attaching the positively charged lysine to phosphatidylglycerol through esterification with glycerol (323, 324). It then flips these modified phospholipids from the inner to the outer leaflet of the cell membrane (325). This resulting positive charge on the outer membrane confers resistance to cationic antimicrobial peptides (CAMPs) but also may alter lysis. Phage lysis depends on holin proteins which form pores in the membrane that dissipate proton-motive force and release endolysins to degrade the cell wall peptidoglycan (80, 326, 327). Because FmtC alters cell surface charge, it also could affect holin-dependent membrane depolarization, endolysin activity, or phage attachment, especially if the phage receptor-binding protein is positively charged. Interestingly enough, the *fmtC* allele from NRS209 did not complement the transposon insertion in USA300 JE2. This could indicate either a loss of function in the allele or incompatibility with some aspect of the USA300 JE2 strain.

Two of the six validated genes did not restore wild-type phenotypes upon complementation (*isdB* and *relA*). RelA, or the *relA*/*spoT* homolog in *S. aureus*, synthesizes (p)ppGpp in response to sensing uncharged tRNAs on the ribosome (328). Transcriptomic studies indicated *S. aureus* upregulates its *relA/spoT* homolog in response to lytic phage predation (329). RelA may contribute to phage-resistant, slow-growing cell (persister) formation (330), although studies indicate ATP depletion rather than (p)ppGpp synthesis accounts for persistence in *S. aureus* (331). IsdB, on the other hand, is part of the iron-regulated surface determinant (isd)

system responsible for scavenging iron from hemoglobin (332). As experiments were conducted in rich media, the hemoglobin-iron scavenging activity of IsdB does not seem relevant, but IsdB may be an abundant surface protein, implicating it in surface occlusion. Neither of these genes are in operons, at least in USA300 JE2. It could be that the native promoters are inherently stronger than the P*lgt* promoter or are strongly upregulated during phage infection thus affecting the efficiency of complementation. We also note for all genes that there was no apparent complementation for phages p002y and pyo (**Supplemental Figure S6**). In the case of the latter two, the parental USA300 JE2 strain was already sensitive to those two phages.

These validated genes along with most other GWAS-detected host range factors were not previously reported as important in *S. aureus* phage infection, but the GWAS did identify some known factors. Such factors included WTA biosynthesis and modification genes, such as *tarP*, *tarJ*, and *tagH*. While TarJ and TagH are involved in WTA biosynthesis, the WTA glycosyltransferase TarP was recently shown to directly confer *Podoviridae* resistance. Capsule biosynthesis (*cap8A* and *cap8I*) (21) and peptidoglycan modification (*oatA*) genes (184) encode surface-associated functions previously implicated in *S. aureus* phage resistance. Capsule or capsule overproduction are known to confer phage resistance in *S. aureus* (22, 196), while peptidoglycan O-acetyl groups are part of the phage receptor (184). Type I restriction-modification (*hsdS*) was implicated as well, and this is a well-known mechanism for suppression of infection across clonal complexes (112). Staphylococcal pathogenicity islands (SaPIs) were not implicated most likely because these are highly specific to siphovirus helper phages, and even for possibly affected helper phages (80α), SaPI interference reduces but does not eliminate helper phage production (248). This means our high-throughput assay may not capture SaPI-level effects, as it does not directly measure phage propagation through plaquing efficiency. CRISPRs were not significant in our study either, because these are rare in *S. aureus* strains (53, 237, 275).

Our study agreed with prior work demonstrating *S. aureus* phages have broad host ranges (102, 103, 173, 174, 277–279). A major goal of our work was to prototype predictive models for

host range based on genome sequence. Genome-based predictions for several antibiotic resistance phenotypes have proven to be of similar accuracy to classic laboratory-based assays (164). We found that *S. aureus* host range prediction accuracy was 40-95% depending on phage. More strains and phages will need to be added to the host range matrix to make genomic host range prediction clinically useful. The difficulty in predicting resistance may come from the large number of genes found to influence the phenotype. Resistant strains may instead have individual, unique mechanisms or other traits that simply confer phage resistance, with the exception of superinfection immunity, in which host-encoded prophages prevent infection of a cognate temperate phage by repressing its lytic genes with their *cI* repressors (72). The two phages with the highest overall resistance (p0045 and p0040 - **Figure 2**) are temperate *Siphoviridae*. Most isolated *S. aureus* strains encode prophages (215), making superinfection immunity and corresponding overall p0045 and p0040 resistance common in the tested strains.

There are limitations to performing phage host range measurement. The high-throughput assay did not measure lysis directly but also did not have the disadvantages of observer bias, low throughput, and qualitative output of the spot assay. In our host range assay, we are measuring the ability of the population overall to survive phage challenge, but this could also indicate the phage suppression of bacterial growth through some level of infection. Likewise, multiple possible sets of population dynamics confound the spot assay. Efficiency of plating (EOP), on the other hand, measures phage propagation efficiency directly by comparing phage titer on a permissive control strain to that on a test strain (179). Nonetheless, factors altering EOP still could affect any stage of the infection cycle, so EOP measurement does not suggest possible phage resistance mechanism. The ambiguity of both assays suggests examining the population dynamics of phages and identified mutants (e.g., *trpA*::Tn) during infection (i.e., adsorption rate; latent period, and burst size from one-step growth curve) would be worthy for future studies to pinpoint the specific mechanism by which that gene affects phage resistance. We also recognize that a

multitude of environmental variables (temperature, multiplicity of infection, growth media) could influence the assay.

There are also some limitations inherent in GWAS approaches. Bacterial GWAS associates homoplasic variants that arise from parallel evolution or recombination with a phenotype of interest (333, 334). While bacterial GWAS can find more types of genetic events (either loss of function or gain of function; mutation, insertion, deletion, recombination, and so on, but not genes with no changes) and more broadly relevant genes and polymorphisms related to a phenotype than screening transposon mutants in a single genetic background, clonal population structure, abundant small effect variants, and genetic interactions hamper it (333). When recombination is relatively rare in a species, like *S. aureus*, large numbers of variants remain in linkage disequilibrium, making it difficult to distinguish lineage from strain-level effects. Variants linked to a causative variant may then be detected as false positives. While we have validated at least a few genes as true positives, and we expect phylogenetically hierarchical effects on host range based on reviewing past work (53), our GWAS methods also include various corrections for clonal population structure when associating variants.

Two recent studies used single gene knockout, overexpression, transcriptional suppression methods as well as global transcriptional profiling to identify phage resistance determinants in *E. coli* (335) and *Enterococcus faecalis* (322). Unlike these previous studies, our findings are not limited to one or a few genetic backgrounds, making them more widely applicable to the species and its underlying evolution. Nonetheless, extensive functional molecular genetics studies will be needed to distinguish genes that truly contribute to host range from false positives. These studies, like those in *E. coli* and *E. faecalis*, would complement the GWAS with global searches for phage resistance genes in a single genetic background, such as Tn-Seq, DUB-Seq, and CRISPR interference to identify genes required for surviving phage infection and RNA-Seq to identify genes differentially regulated in response to phage infection. Such work would both

corroborate GWAS results and fill in the gaps - possible determinants not present or conserved in enough of the resistant or sensitive population.

Our results have important consequences for phage therapy, phage-small molecule combination therapy, and horizontal gene transfer in the species. Genes identified expand the set of potential combination therapies by providing additional targets to which to design small molecules to interfere with phage resistance. Already, combination phage/antibiotic therapies have shown promise for clearing biofilms and reducing emergence of antibiotic resistance in *S. aureus* (336), and ribosomal active antibiotics are known to enhance MRSA phage sensitivity at sublethal doses (313). Additionally, because the phage receptor WTA is necessary for methicillin resistance (193) and WTA inhibition resensitizes MRSA to methicillin (311), phages have the exciting possibility of inducing collateral beta-lactam sensitivity. We also cannot discount the possibility that phage resistance polymorphisms are the result of selection by other stresses besides phage infection, such as immune escape, interbacterial interactions, or antibiotic selection. Wall teichoic acid, the *S. aureus* phage receptor, for example, is also important for colonization, antibiotic resistance, and immune evasion (49, 193, 261, 266, 267, 337, 338). Because we identified phage host range determinants, we also gain insights into the evolution of the *S. aureus* through horizontal gene transfer. Transduction, the transfer of host genetic material between strains by abortive phage infection, is a major mechanism of HGT (27) and recombination (339) in the species. There is a tradeoff between the need to resist phage killing and the need to adapt by gaining new virulence genes (such as Panton-Valentine leukocidin) (340) through HGT. It is possible that the most transducible strains are both more sensitive to killing by phage infection, but also more able to outcompete other strains for advantageous genetic material. The finding that even the most resistant strains (NRS148, NRS209, and NRS255) were still sensitive to two out of the eight phages may be the result of a selection for sensitivity that could be the Achilles heel of *S. aureus* when confronted by phage therapy.

Acknowledgements

We thank Veronique Perrot and Bruce Levin for providing the pyo myophage used for host range evaluation. We also thank Bruce Levin for providing constructive comments on the manuscript. Sarah Satola and Eryn Bernardy provided VISA and CF *S. aureus* strains used for host range testing, respectively. Abraham Moller was supported by the National Science Foundation (NSF) Graduate Research Fellowship Program (GRFP). Timothy Read was supported by the National Institutes of Health (NIH) grant AI121860. Kyle Winston was supported by the Emory REAL fellowship. We thank Michelle Su and Robert Petit for assistance with GWAS methods and constructive criticism of the project.

Figures

Figure 1: Development of the high-throughput phage host range assay. A) Example of fully sensitive (NRS149) and fully resistant (NRS148) spot assay phenotypes for five test phages (p0045, p0006, p0017S, p002y, and p003p). *Siphoviridae* are listed in red, *Myoviridae* in blue, and *Podoviridae* in purple. B) Calibration of the high-throughput assay against qualitative spot assay phenotypes (S - sensitive, complete clearing; SS - semi-sensitive, cloudy clearing; R - resistant, no clearing) collected with the spot assay for 108 NARSA strains and the five phages listed in A. Data represent the distribution of average high-throughput assay measurements for strains evaluated as S, SS, or R in corresponding spot assays. Wilcoxon signed-rank test significance values for each possible comparison are listed at the top of the corresponding boxplots (ns - not significant, * - 0.01 to 0.05, ** - 0.001 to 0.01, *** - 0.0001 to 0.001, **** - less than 0.0001). C) Example high-throughput assay results from one 96-well plate containing 96 NARSA strain overnight cultures co-incubated with phage p0006 (K). D) Example high-throughput assay phenotypes for a sensitive *S. aureus* strain, resistant strain, bacteria without phage, and phage without bacteria.

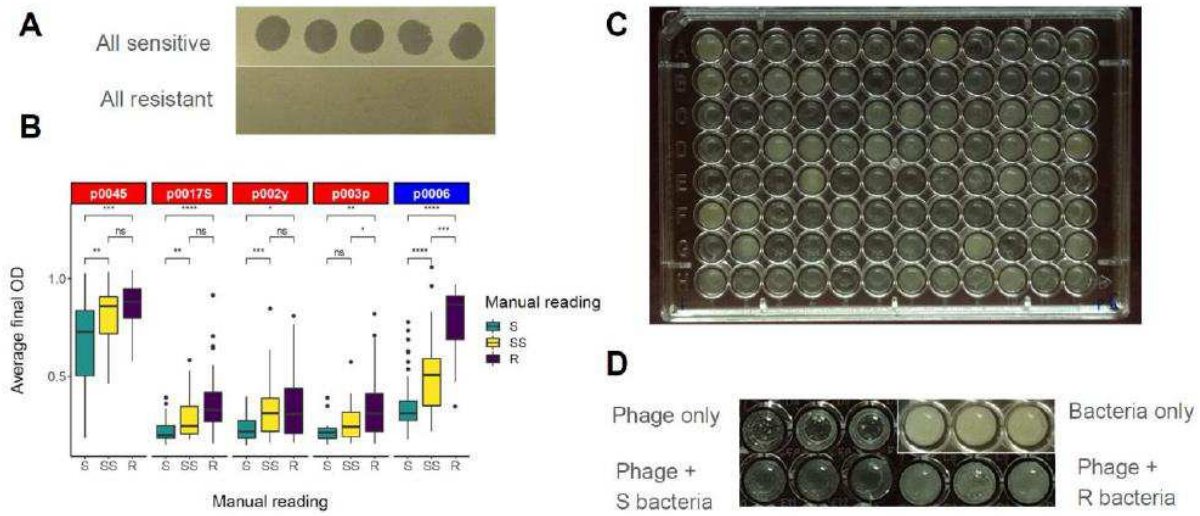


Figure 2: Host range distribution, concordance, and multiple phage resistance. A) Number of strains that fall into host range categories for each phage. Sensitive (S) corresponds to an OD_{600} of 0.1-0.4, semi-sensitive corresponds to an OD_{600} of 0.4-0.7, and resistant corresponds to an OD_{600} of 0.7 or higher. B) Histogram of number of phages to which strains are resistant, by the previous definition. C) Concordance matrix of the tested phages' host ranges. Concordance is defined as the number of strains with identical phenotypes between two phages. *Siphoviridae* are listed in red, *Myoviridae* in blue, and *Podoviridae* in purple.

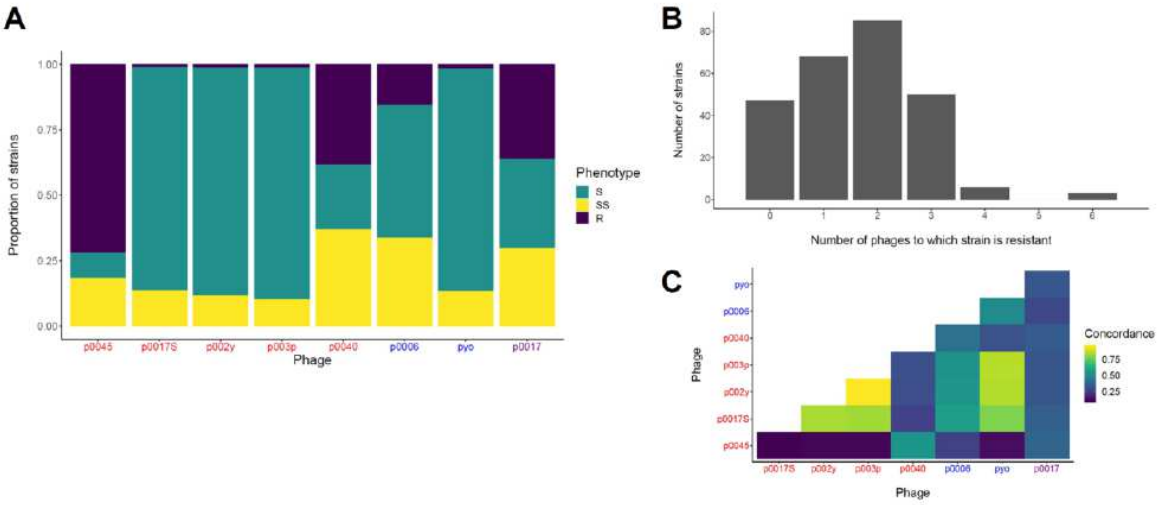


Figure 3: Phage resistance is related to clonal complex (CC) but not MRSA genetic background. Data represent the distribution of average high-throughput assay measurements for strains belonging to each presented CC (all 259 strains) (A) or MRSA/MSSA (126 NARSA strains) (B) genetic background. One-way ANOVA significance values for overall differences amongst CCs presented and Wilcoxon signed-rank test significance values for MRSA/MSSA differences are listed at the top of the corresponding boxplots (ns - not significant, * - 0.01 to 0.05, ** - 0.001 to 0.01, *** - 0.0001 to 0.001, **** - less than 0.0001). *Siphoviridae* are listed in red, *Myoviridae* in blue, and *Podoviridae* in purple.

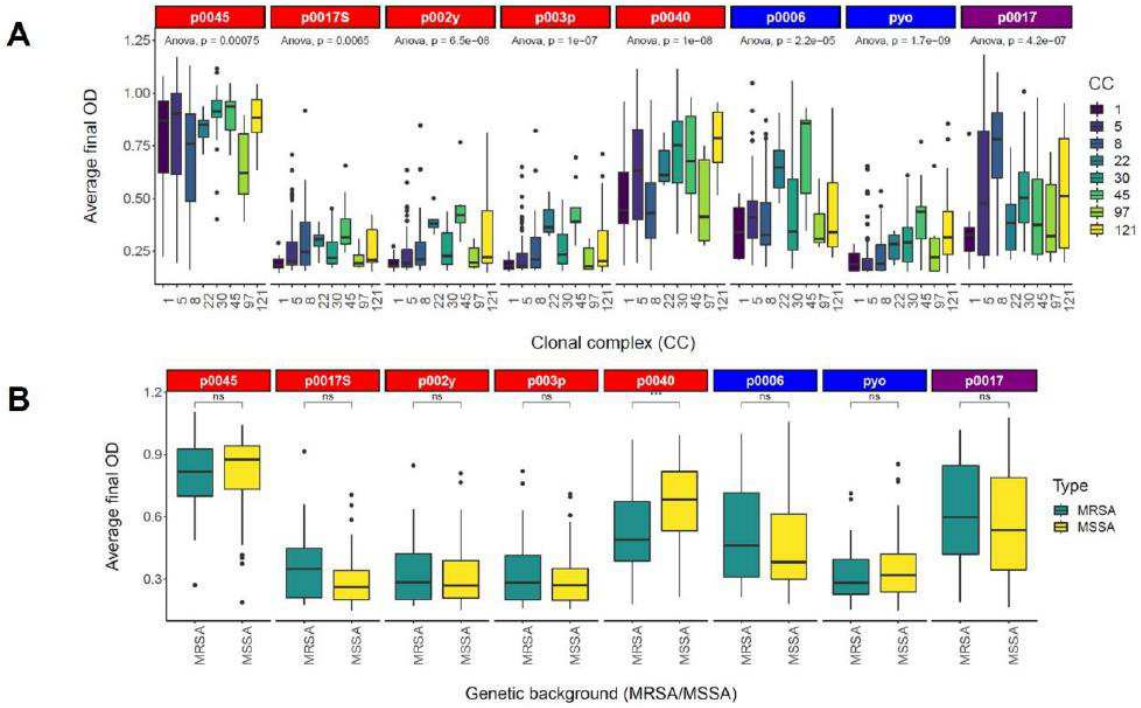


Figure 4: Phage resistance across the *S. aureus* species. Average high-throughput phage host range assay phenotypes (of at least six replicates) and corresponding strain clonal complexes were placed on a maximum-likelihood, midpoint-rooted core-genome phylogeny and visualized with the Interactive Tree of Life (iTOL) (293). Phenotypes are presented on a scale from blue (lowest OD_{600}, most sensitive) to orange (highest OD_{600}, most resistant). Phenotypes from inside to outside correspond to phages p0045, p0006, p0017, p0017S, p002y, p003p, p0040, and pyo. CCs are shaded inside and outside the circumference of the tree.

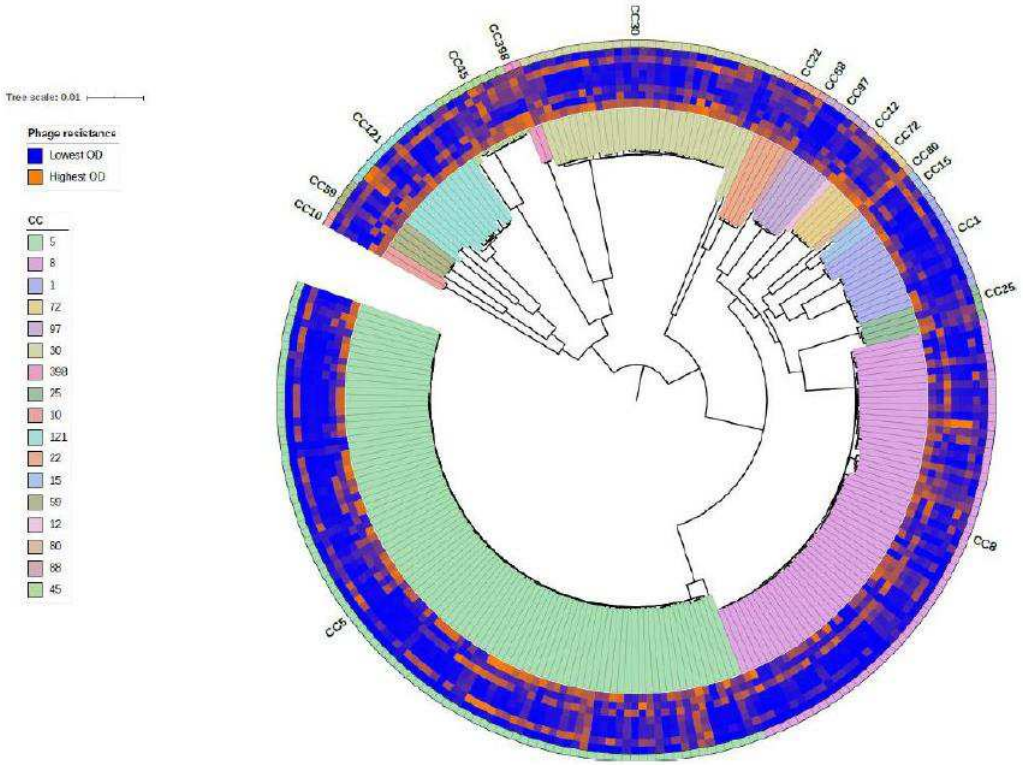


Figure 5: Molecular genetics validates putative phage resistance determinants. High-throughput host range assay (top) and efficiency of plating (EOP; bottom) phenotypes demonstrating genetic validation of novel GWAS phage host range determinants. Results are grouped by gene (*trpA*, *phoR*, *isdB*, *sodM*, *fmtC*, and *relA*). All assays were performed with siphovirus p003p or no phage. Each gene group includes four strains demonstrating complementation with proper controls (USA300, USA300 transposon mutant, USA300 transposon mutant with empty pOS1 vector, and USA300 transposon mutant complemented with gene in pOS1 vector). All significant ($p<0.05$) pairwise differences (Wilcoxon signed-rank test) are shown at the top of the corresponding boxplots.

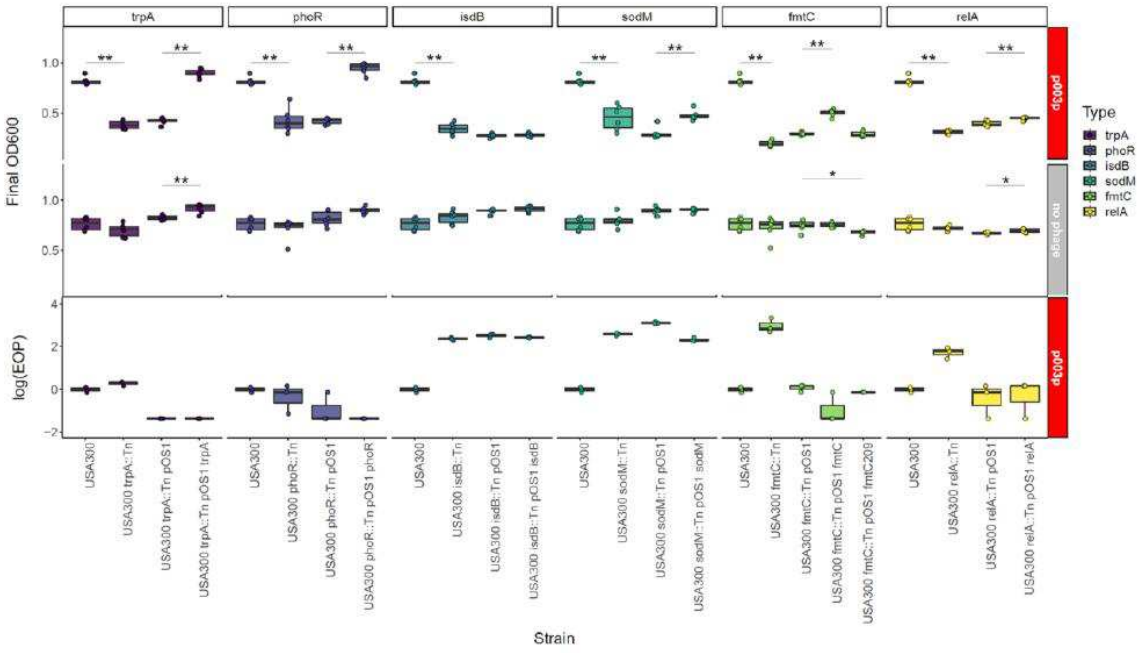
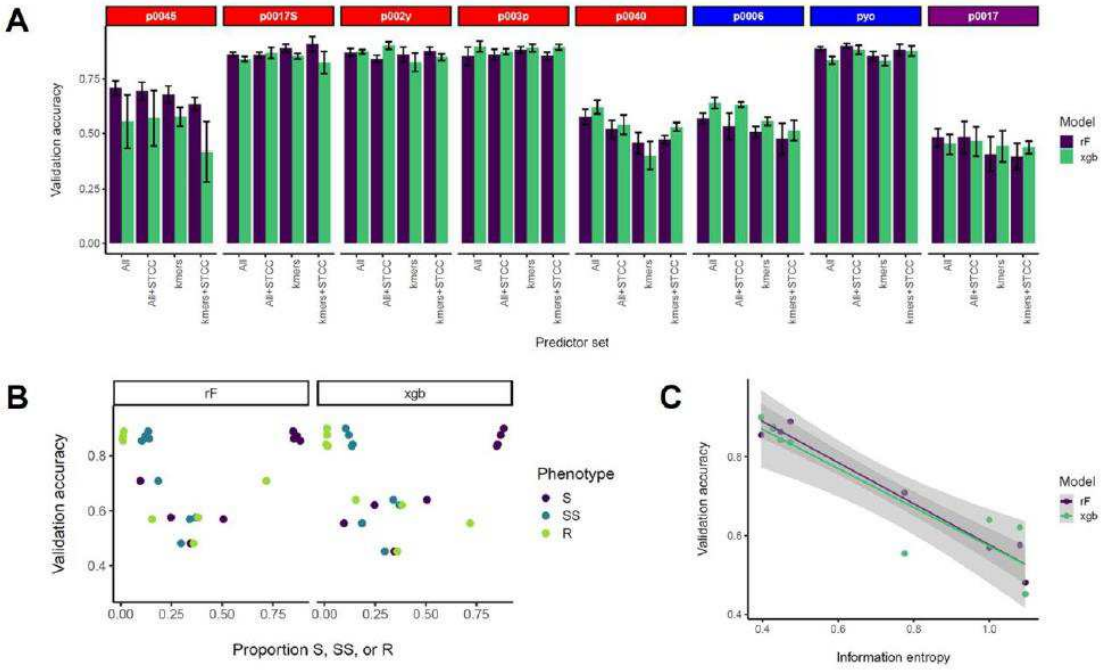


Figure 6: Construction of predictive models for each ternary phage resistance phenotype. Quantitative host range phenotypes were classified as S - sensitive, SS - semi-sensitive, or R - resistant based on the bins (0.1-0.4, 0.4-0.7, and 0.7 or more, respectively). *Siphoviridae* are listed in red, *Myoviridae* in blue, and *Podoviridae* in purple. A) 10-fold cross-validation predictive accuracies for each phage based on two model building methods (randomForest and XGBoost)

and four sets of predictors - all significant GWAS genetic determinants (COGs, SNPs, and k-mers) for a particular phage, all determinants plus corresponding strain sequence type and clonal complex (ST and CC), significant k-mers for a particular phage, and significant k-mers plus strain ST and CC. Average accuracies of four 10-fold CV replicates are presented with one standard error above and below the mean. Validation accuracy represents the proportion of correctly identified ternary phenotypes in the validation set (one-tenth of the strain set). B) Average accuracies from four 10-fold CV replicates for each model building method and all significant GWAS determinants as predictors relative to the proportion of each ternary phenotype (S, SS, or R) amongst tested strains for the corresponding phage. Three points are shown for each validation accuracy (corresponding to each of the three possible phenotypes). C) Average accuracies from four 10-fold CV replicates for each model building method and all significant GWAS determinants as predictors relative to the information entropy for each host range phenotype, which was calculated as described in the Materials and Methods section. Information entropy was calculated with a natural logarithm in natural units (nats).



Tables

Table 1: Strains, phages, and plasmids used for phage propagation and molecular genetic validation of GWAS results.

Strain, phage, or plasmid	Characteristics/description	Reference
***E. coli* Strains**		
DH5α	*E. coli* cloning strain; F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK–mK+), λ-	(341)
IM08B	*E. coli* cloning strain with *S. aureus* CC8 DNA modification; DNA cytosine methyltransferase (dcm) negative mutant of E. coli K12 DH10B; mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL endA1 nupG Δdcm ΩPhelp-hsdMS (CC8-2) ΩPN25-hsdS (CC8-1)	(306)
***S. aureus* strains**		
RN4220	Phage propagation strain; background for transducing *tarM*::Tn; cloning	(189)

	intermediate for pOS1-P*lgt-fmtC*	
RN4220 *tarM*::Tn	Podovirus (p0017) propagation strain; generated by transducing RN4220 with USA300 JE2 *tarM*::Tn (NE611) phage 0045 lysate	This study
USA300 JE2	Wild-type for genetic validation experiments and background for transposon mutant backcrossing	(280)
USA300 JE2 *tarM*::Tn (NE611)	Transposon mutant transduced into RN4220 to make RN4220 *tarM*::Tn by a NE611 phage 0045 lysate	(280)
USA300 JE2 *trpA*::Tn	Backcrossed mutant NE304 back into USA300 JE2	This study
USA300 JE2 *trpA*::Tn pOS1	Complemented backcrossed mutant with empty vector	This study
USA300 JE2 *trpA*::Tn pOS1 *trpA*	Complemented backcrossed mutant with *trpA* from USA300 JE2	This study
USA300 JE2 *phoR*::Tn	Backcrossed mutant NE618 back into USA300 JE2	This study
USA300 JE2	Complemented backcrossed mutant with	This study

phoR::Tn pOS1	empty vector	
USA300 JE2 *phoR*::Tn pOS1 *phoR*	Complemented backcrossed mutant with *phoR* from USA300 JE2	This study
USA300 JE2 *isdB*::Tn	Backcrossed mutant NE1102 back into USA300 JE2	This study
USA300 JE2 *isdB*::Tn pOS1	Complemented backcrossed mutant with empty vector	This study
USA300 JE2 *isdB*::Tn pOS1 *isdB*	Complemented backcrossed mutant with *isdB* from USA300 JE2	This study
USA300 JE2 *sodM*::Tn	Backcrossed mutant NE1224 back into USA300 JE2	This study
USA300 JE2 *sodM*::Tn pOS1	Complemented backcrossed mutant with empty vector	This study
USA300 JE2 *sodM*::Tn pOS1 *sodM*	Complemented backcrossed mutant with *sodM* from USA300 JE2	This study
USA300 JE2 *fmtC*::Tn	Backcrossed mutant NE1360 back into USA300 JE2	This study
USA300 JE2	Complemented backcrossed mutant with	This study

fmtC::Tn pOS1	empty vector	
USA300 JE2 *fmtC*::Tn pOS1 *fmtC*	Complemented backcrossed mutant with *fmtC* from USA300 JE2	This study
USA300 JE2 *fmtC*::Tn pOS1 *fmtC209*	Complemented backcrossed mutant with *fmtC* from NRS209	This study
USA300 JE2 *relA*::Tn	Backcrossed mutant NE1714 back into USA300 JE2	This study
USA300 JE2 *relA*::Tn pOS1	Complemented backcrossed mutant with empty vector	This study
USA300 JE2 *relA*::Tn pOS1 *relA*	Complemented backcrossed mutant with *relA* from USA300 JE2	This study
Phages		
Phage p0045 (80α-like)	*Siphoviridae* phage; also used for backcrossing and pOS1-P*lgt-fmtC* transduction from RN4220 into USA300 *fmtC*::Tn	(39, 101, 342)
Phage p0006 (K)	*Myoviridae* phage; GenBank accession NC_005880.2	(102, 174, 343)

Phage p0017 (HER49/p66)	*Podoviridae* phage; GenBank accession NC_007046.1	This study
Phage p0017S	*Siphoviridae* phage	This study
Phage p002y (DI)	*Siphoviridae* phage	This study
Phage p003p (Mourad 87)	*Siphoviridae* phage	This study
Phage p0040 (Mourad 2)	*Siphoviridae* phage	This study
Phage pyo	*Myoviridae* phage; BioProject accession PRJNA477834	(336, 344)
Plasmids		
pOS1-P*lgt*	Empty complementation vector	(305)
pOS1-P*lgt-trpA*	Complementation vector with *trpA* cloned downstream of P*lgt*	This study
pOS1-P*lgt-phoR*	Complementation vector with *phoR* cloned downstream of P*lgt*	This study
pOS1-P*lgt-isdB*	Complementation vector with *isdB* cloned downstream of P*lgt*	This study
pOS1-P*lgt-sodM*	Complementation vector with *sodM* cloned downstream of P*lgt*	This study

pOS1-P*lgt-fmtC*	Complementation vector with *fmtC* cloned downstream of P*lgt*	This study
pOS1-P*lgt-fmtC209*	Complementation vector with *fmtC* from NRS209 cloned downstream of P*lgt*	This study
pOS1-P*lgt-relA*	Complementation vector with *relA* cloned downstream of P*lgt*	This study

Table 2: Summary statistics of phage host range phenotypes. For each phage, the number of strains falling into each category were counted. These phenotypes were determined for each phage using the high-throughput assay. Sensitive (0.1-0.4), semi-sensitive (0.4-0.7), and resistant (0.7 and higher) strain numbers and percentages are listed first, followed by mean, standard deviation, and median quantitative phenotypes for all tested strains. Statistics summarize at least six biological replicates for each page.

Phage	p0045	p0006	p0017	p0017S	p002y	p003p	p0040	pyo
Sensitive (%)	25 (9.7%)	131 (50.6%)	89 (34.4%)	221 (85.3%)	225 (86.9%)	229 (88.4%)	64 (24.7%)	220 (84.9%)
Semi-sensitive (%)	48 (18.5%)	88 (34.0%)	77 (29.7%)	36 (13.9%)	31 (12.0%)	27 (10.4%)	96 (37.1%)	35 (13.5%)
Resistant (%)	186 (71.8%)	40 (15.4%)	93 (35.9%)	2 (0.7%)	3 (1.2%)	3 (1.2%)	99 (38.2%)	4 (1.5%)
Mean	0.80	0.49	0.56	0.31	0.27	0.32	0.61	0.26

Stdev	0.24	0.20	0.27	0.12	0.12	0.12	0.23	0.14
Median	0.88	0.44	0.54	0.27	0.22	0.31	0.60	0.21

Table 3: Measures of phylogenetic signal for each phage resistance phenotype. Values that are significant are shown in bold. Significance was determined for 999 random permutations of the data.

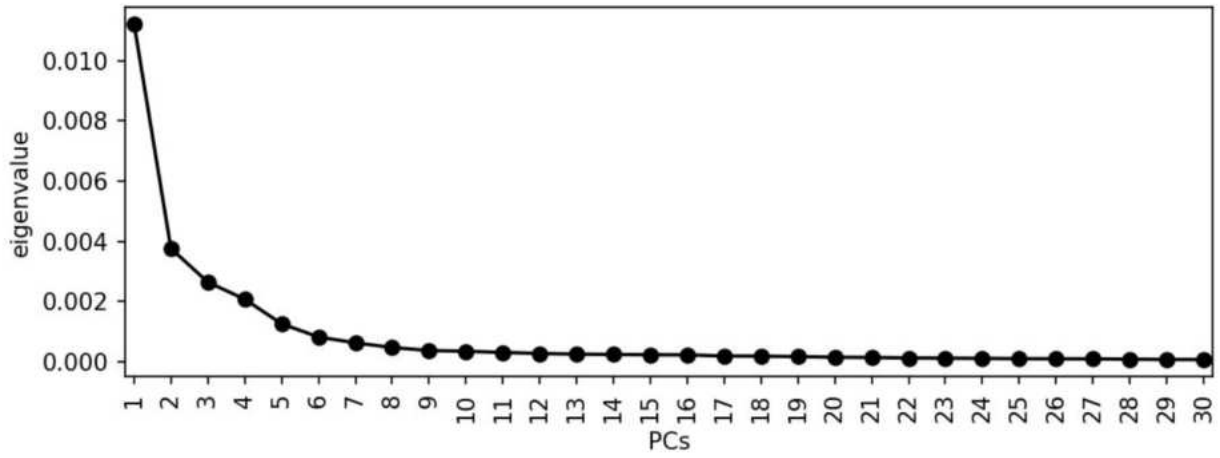
Phage	Moran's I	Abouheif's C_{mean}	Pagel's λ	Blomberg's K
p0045	**0.23**	**0.23**	**1.00**	**0.005**
p0006	**0.17**	**0.17**	**1.00**	**0.008**
p0017S	**0.32**	**0.32**	**1.00**	**0.007**
p002y	**0.23**	**0.23**	**1.00**	**0.008**
p003p	**0.30**	**0.30**	**1.00**	**0.012**
p0040	**0.28**	**0.28**	**1.00**	**0.014**
pyo	**0.36**	**0.37**	**1.00**	**0.006**
p0017	**0.31**	**0.31**	**1.00**	**0.014**

Table 4: GWAS summary statistics for each associated genetic element. Each value represents the number of unique genetic elements of a particular type found to be significantly associated with the phage host range phenotype.

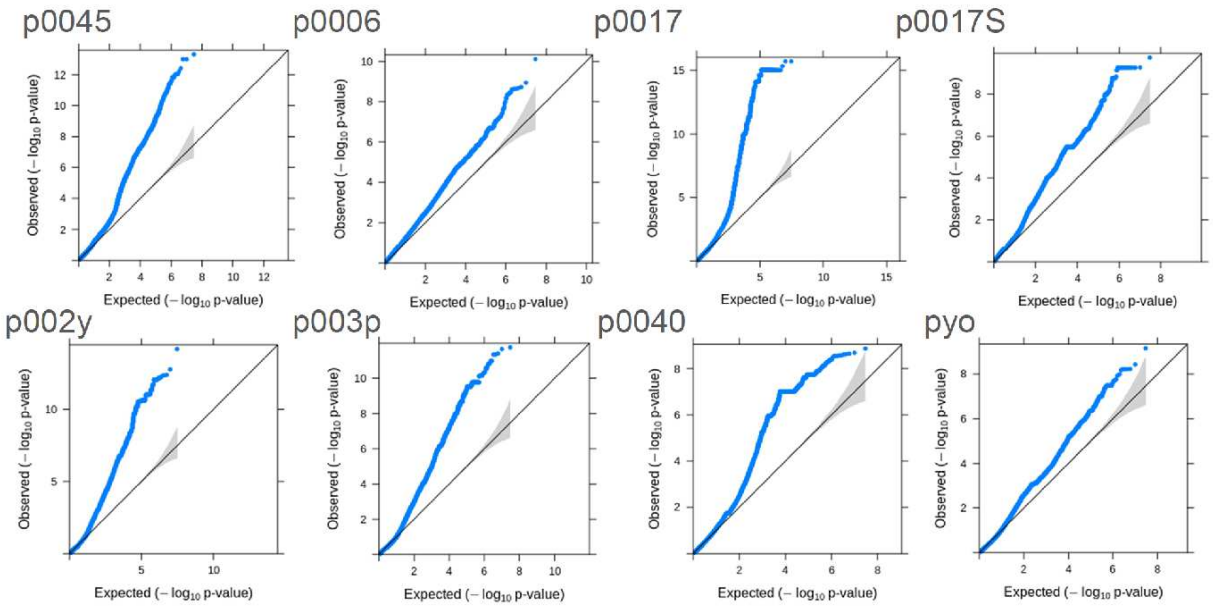
Phage	p0045	p0006	p0017	p0017 S	p002y	p003p	p0040	pyo
COGs (pyseer)	131	49	76	48	163	175	163	347
SNPs (pyseer)	0	27	1	0	134	48	6	249
k-mers (pyseer)	820	18	7078	101	1734	866	180	14
SNPs (treeWAS)	0	1	0	0	0	1	1	4

Supplemental Material for Chapter 3

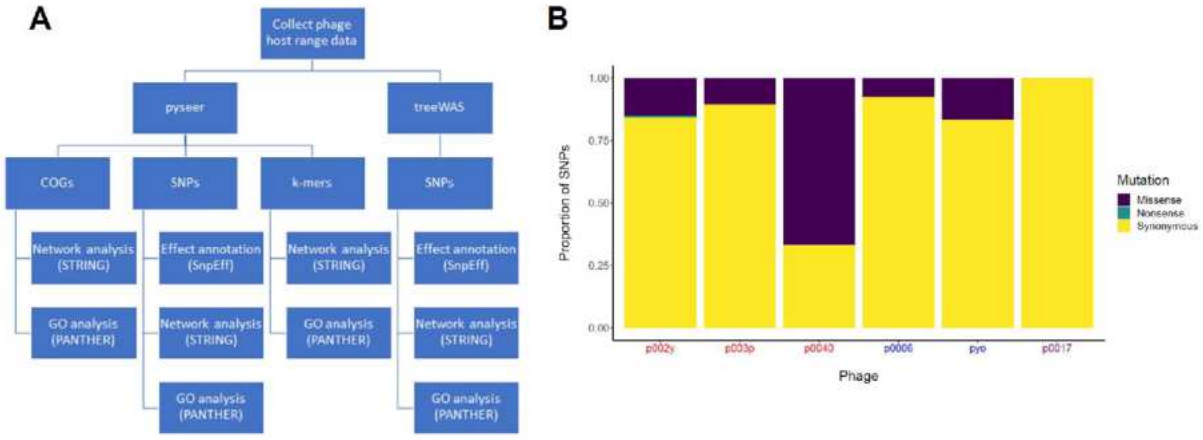
Supplemental Figure S1: Scree plot used to pick the number of dimensions for multidimensional scaling (MDS) in pyseer COG significance analysis. The number of dimensions (PCs) picked was the least possible (297) after which the eigenvalue stabilized with respect to dimension number.



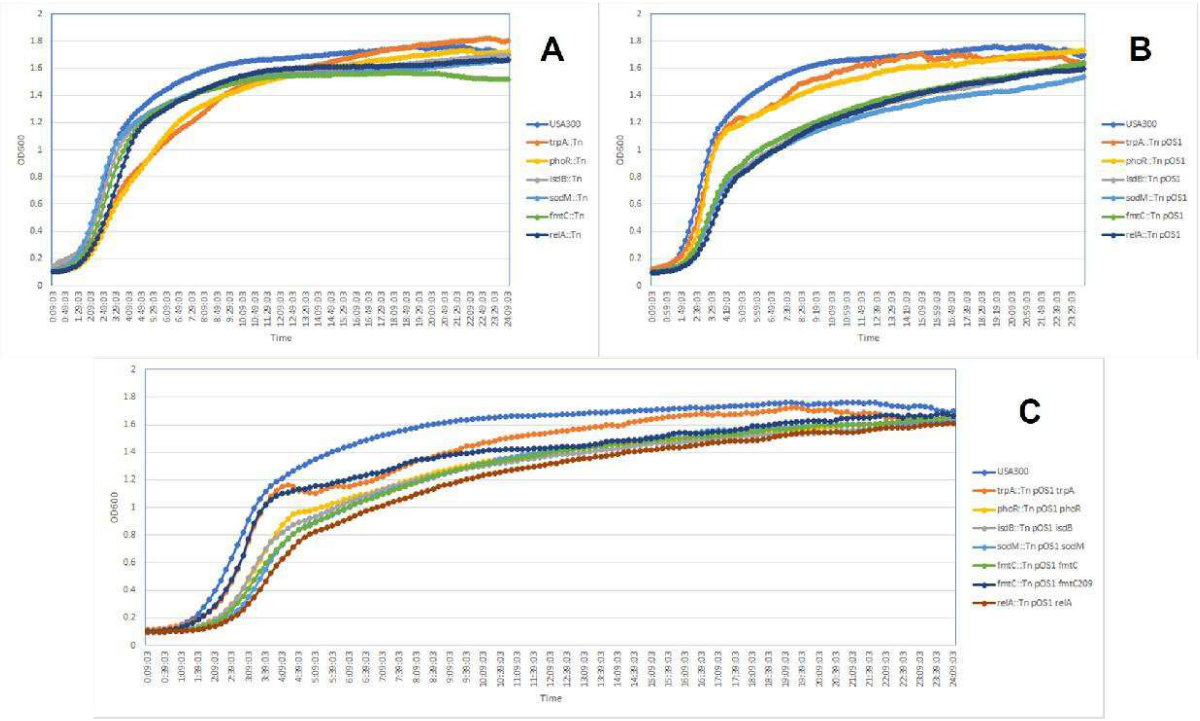
Supplemental Figure S2: Pyseer k-mer Q-Q plots for each phage (p0045, p0006, p0017, p0017S, p002y, p003p, p0040, and pyo). The observed p-values were plotted relative to the expected p-values based on the null distribution. Expected p-values were plotted with a 95% confidence interval on the diagonal. Deviation of the observed/expected curve from the diagonal indicated p-value inflation.



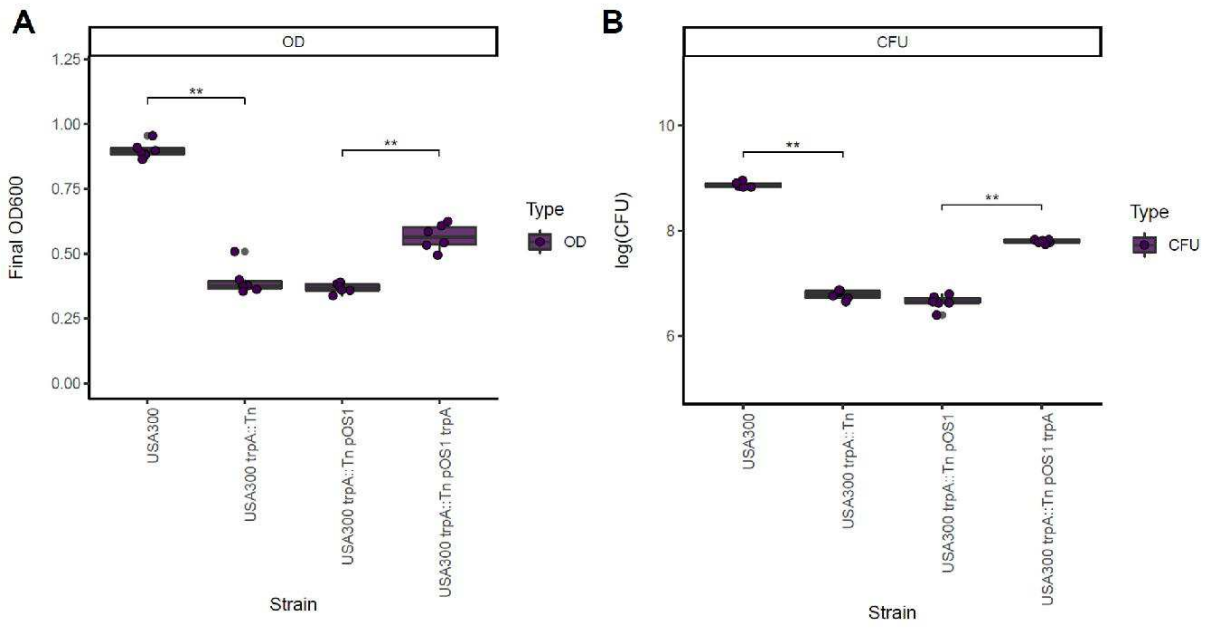
Supplemental Figure S3: GWAS approach and significant SNP annotations. A) Overview of the genome-wide association study (GWAS) workflow. Pyseer (297) associated intermediate-frequency COGs, core-genome SNPs, and k-mers with each host range phenotype, while treeWAS (298) only associated core-genome SNPs with each host range phenotype. SnpEff (301) classified mutation effects (synonymous, missense, or nonsense) from the corresponding Roary (289) gene sequence, while STRING (303) identified putative protein-protein interactions and PANTHER (304) identified enriched functions from lists of genes corresponding to each significant SNP or k-mer. B) Classification of significantly associated pyseer or treeWAS SNPs based on mutational effect (synonymous, missense, or nonsense). SnpEff annotated SNP effects based on corresponding genes identified in the tested strains' core genome with Roary. Phage 0045 was not included as no significant SNPs were detected for its host range phenotype. *Siphoviridae* are listed in red, *Myoviridae* in blue, and *Podoviridae* in purple.



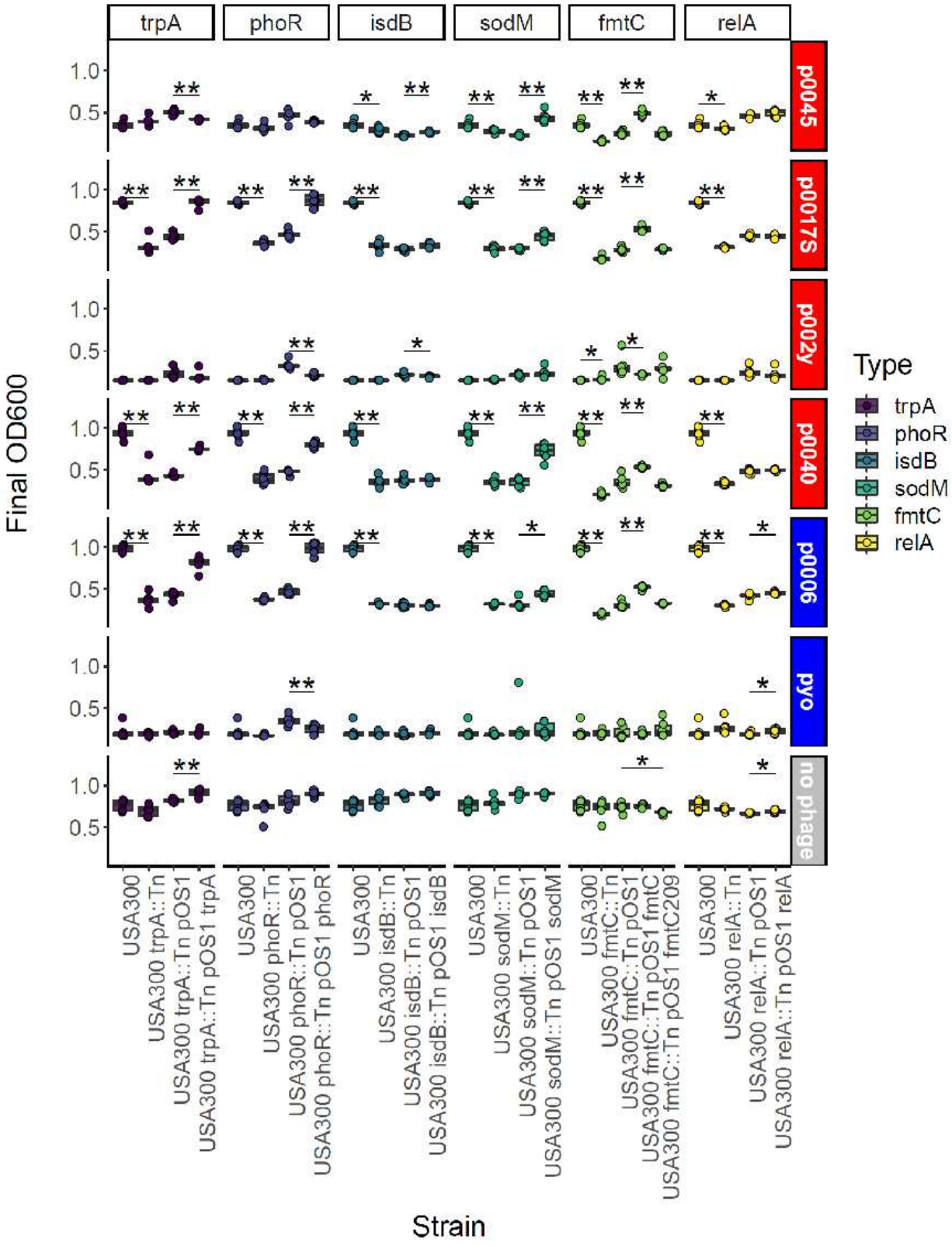
Supplemental Figure S4: Growth curves of USA300, USA300 transposon mutants (A), transposon mutants electroporated with the empty pOS1 vector (B), and transposon mutants complemented with vectors containing respective genes (C; *trpA*, *phoR*, *isdB*, *sodM*, *fmtC*, and *relA*). Strains were inoculated with a 96-pin replicator from arrayed frozen glycerol stocks into 96-well plates containing 200 µL LB/TSB 2:1 with 5 mM $CaCl_2$ or the same medium supplemented with 10 µg/mL chloramphenicol in each well. We then diluted each culture 1:100 in fresh LB/TSB 2:1 with 5 mM $CaCl_2$ or the same medium supplemented with 10 µg/mL chloramphenicol and collected growth curves on a BioTek Eon plate reader (37°C, 225 rpm agitation, OD_{600} measured every 10 minutes).



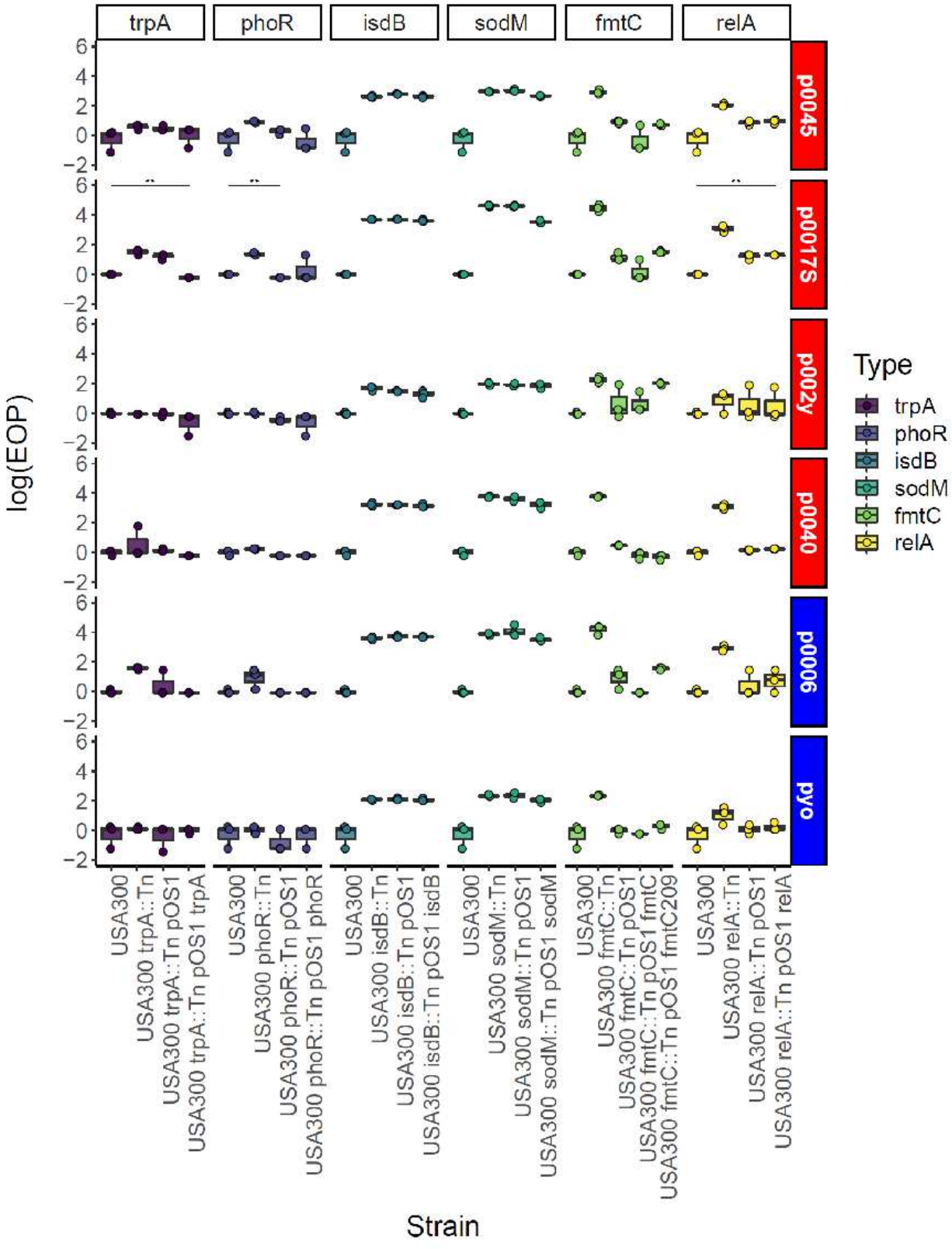
Supplemental Figure S5: Bacterial survival after completing the high-throughput host range assay (p003p against *trpA* strains). The high-throughput assay was performed for six biological replicates of USA300, USA300 *trpA*::Tn, USA300 *trpA*::Tn pOS1, and USA300 *trpA*::Tn pOS1 *trpA* strains. A) ODs were measured for the high-throughput phage host range assay replicates as described previously. B) Agar plugs were removed with toothpicks, transferred to 0.8 mL volumes of sterile TMG, and bacteria resuspended by vortexing. The resuspensions were serially diluted in TMG and 4 uL of 1e-1 through 1e-6 dilutions were spotted four times on TSA plates. Dilution plates were grown overnight at 37°C and colonies counted the following day to determine surviving CFU in each condition.



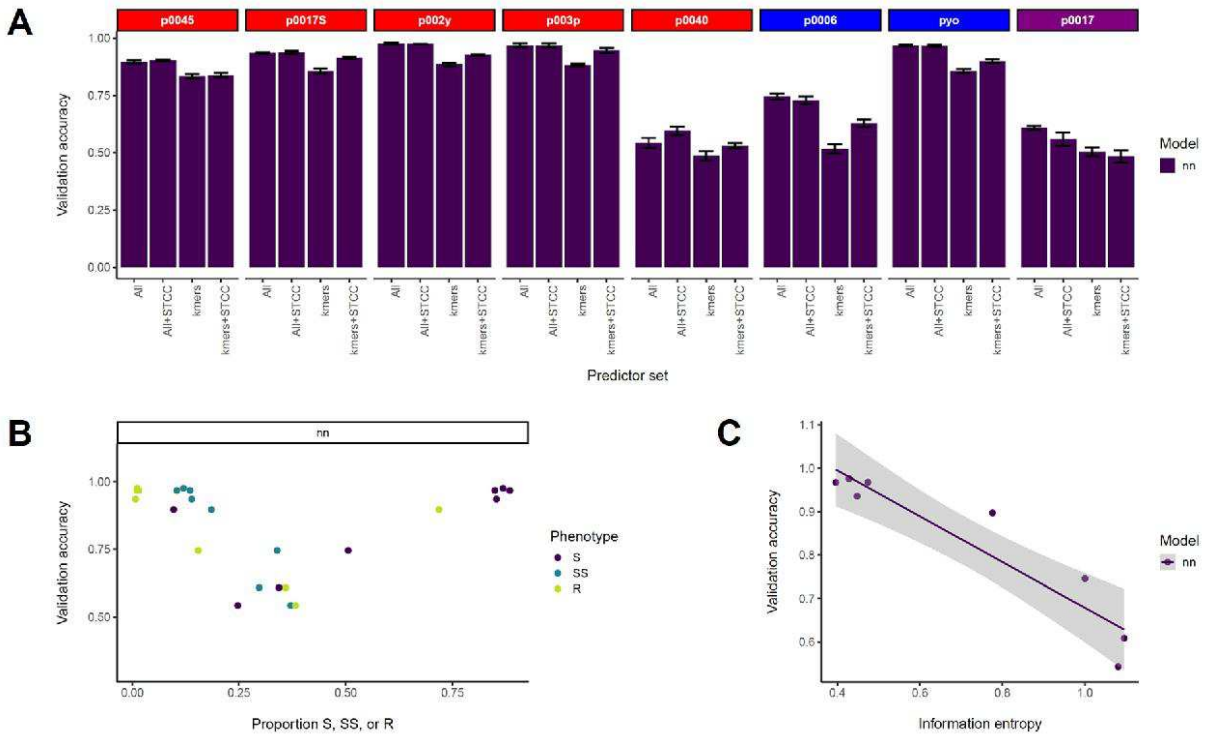
Supplemental Figure S6: High-throughput host range assay phenotypes demonstrating genetic validation of novel GWAS phage host range determinants. Results are grouped by gene (*trpA*, *phoR*, *isdB*, *sodM*, *fmtC*, and *relA*) and phage (p0045, p0017S, p003p, p0040, p0006, p002y, pyo, and no phage). Each group includes four strains demonstrating complementation with proper controls (USA300, USA300 transposon mutant, USA300 transposon mutant with empty pOS1 vector, and USA300 transposon mutant complemented with gene in pOS1 vector). All significant ($p<0.05$) pairwise differences (Wilcoxon signed-rank test) are shown at the top of the corresponding boxplots. *Siphoviridae* are listed in red, *Myoviridae* in blue, and the no phage control in gray.


trpA
phoR
isdB
sodM
fmtC
relA
p0045
p0017S
p002y
p0040
p0006
pyo
no phage
Final OD600
1.0
0.5
Type
trpA
phoR
isdB
sodM
fmtC
relA
USA300
USA300 trpA::Tn
USA300 trpA::Tn pOS1
USA300 trpA::Tn pOS1 trpA
USA300
USA300 phoR::Tn
USA300 phoR::Tn pOS1
USA300 phoR::Tn pOS1 phoR
USA300
USA300 isdB::Tn
USA300 isdB::Tn pOS1
USA300 isdB::Tn pOS1 isdB
USA300
USA300 sodM::Tn
USA300 sodM::Tn pOS1
USA300 sodM::Tn pOS1 sodM
USA300
USA300 fmtC::Tn
USA300 fmtC::Tn pOS1
USA300 fmtC::Tn pOS1 fmtC
USA300 fmtC::Tn pOS1 fmtC209
USA300
USA300 relA::Tn
USA300 relA::Tn pOS1
USA300 relA::Tn pOS1 relA
Strain

Supplemental Figure S7: Efficiency of plating (EOP) phenotypes demonstrating genetic validation of phage host range determinants. Undiluted through 1e-8 dilutions of phage were spotted (4 μL) three times on each top agar lawn, let to dry, incubated face up overnight at 37°C, and plaques counted at the lowest countable dilution. EOP was calculated relative to the average PFU/mL for the control strain, USA300 JE2. Results are grouped by gene (*trpA*, *phoR*, *isdB*, *sodM*, *fmtC*, and *relA*) and phage (p0045, p0017S, p003p, p0040, p0006, p002y, and pyo). *Siphoviridae* are listed in red and *Myoviridae* in blue. Each group includes four strains demonstrating complementation with controls (USA300, USA300 transposon mutant, USA300 transposon mutant with empty pOS1 vector, and USA300 transposon mutant complemented with gene in pOS1 vector). All significant ($p<0.05$) pairwise differences (Wilcoxon signed-rank test) are shown at the top of the corresponding boxplots.

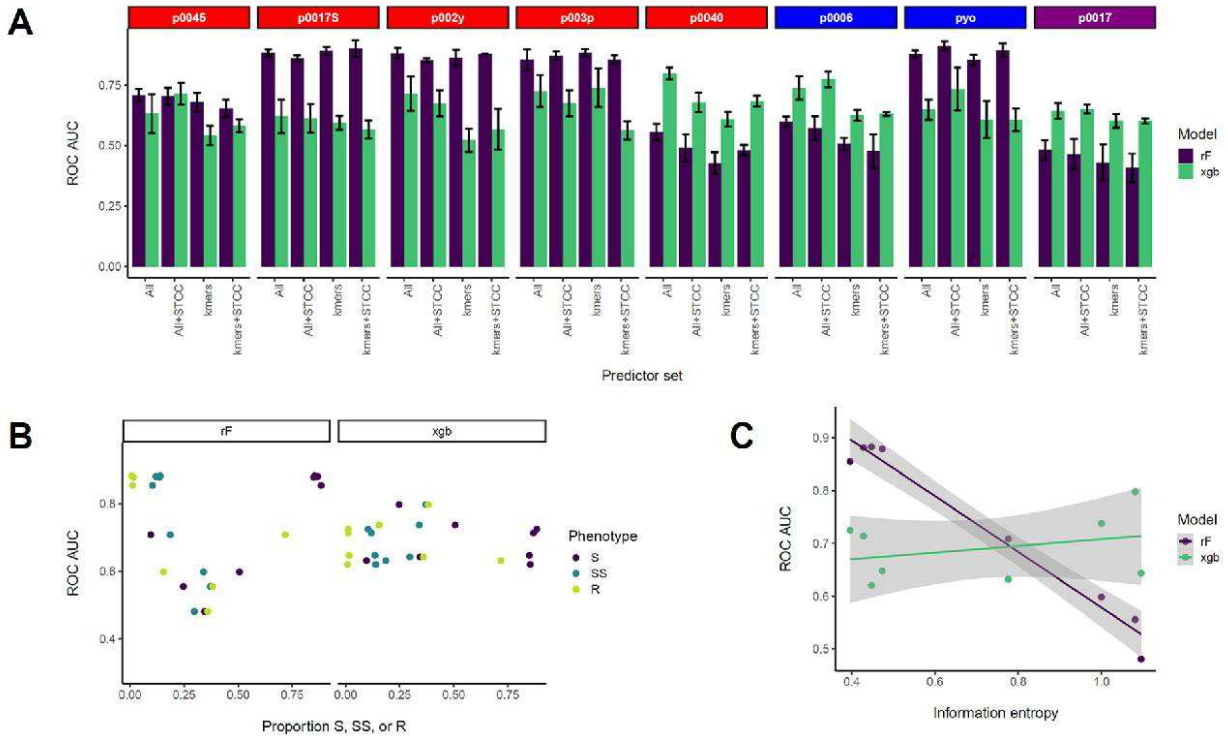

trpA
phoR
isdB
sodM
fmtC
relA
P0045
P0017S
P002y
P0040
P0006
pyo
log(EOP)
Type
trpA
phoR
isdB
sodM
fmtC
relA
USA300
USA300 trpA::Tn
USA300 trpA::Tn pOS1
USA300 trpA::Tn pOS1 trpA
USA300
USA300 phoR::Tn
USA300 phoR::Tn pOS1
USA300 phoR::Tn pOS1 phoR
USA300
USA300 isdB::Tn
USA300 isdB::Tn pOS1
USA300 isdB::Tn pOS1 isdB
USA300
USA300 sodM::Tn
USA300 sodM::Tn pOS1
USA300 sodM::Tn pOS1 sodM
USA300
USA300 fmtC::Tn
USA300 fmtC::Tn pOS1
USA300 fmtC::Tn pOS1 fmtC
USA300 fmtC::Tn pOS1 fmtC209
USA300
USA300 relA::Tn
USA300 relA::Tn pOS1
USA300 relA::Tn pOS1 relA
Strain

Supplemental Figure S8: Construction of neural network predictive models for each ternary phage resistance phenotype. Quantitative host range phenotypes were classified as S - sensitive, SS - semi-sensitive, or R - resistant based on the bins (0.1-0.4, 0.4-0.7, and 0.7 or more, respectively). Data preprocessing included oversampling (p0045, p0017S, p002y, p003p, or pyo), lasso regression (p0017), both (p0006), or neither (p0040). A) Predictive accuracies for each phage based on neural networks and four sets of predictors - all significant GWAS genetic determinants (COGs, SNPs, and k-mers) for a particular phage, all determinants plus corresponding strain sequence type and clonal complex (ST and CC), significant k-mers for a particular phage, and significant k-mers plus strain ST and CC. Average accuracies of four replicates are presented with one standard error above and below the mean. Validation accuracy represents the proportion of correctly identified ternary phenotypes in the validation set (30% of the strain set). B) Average accuracies from four replicates and all significant GWAS determinants as predictors relative to the proportion of each ternary phenotype (S, SS, or R) amongst tested strains for the corresponding phage. Three points on the same horizontal are shown for each validation accuracy (corresponding to each of the three possible phenotypes). C) Average accuracies from four replicates and all significant GWAS determinants as predictors relative to the information entropy for each host range phenotype, which was calculated as described in the Materials and Methods section. Information entropy was calculated with a natural logarithm in natural units (nats). *Siphoviridae* are listed in red, *Myoviridae* in blue, and *Podoviridae* in purple.



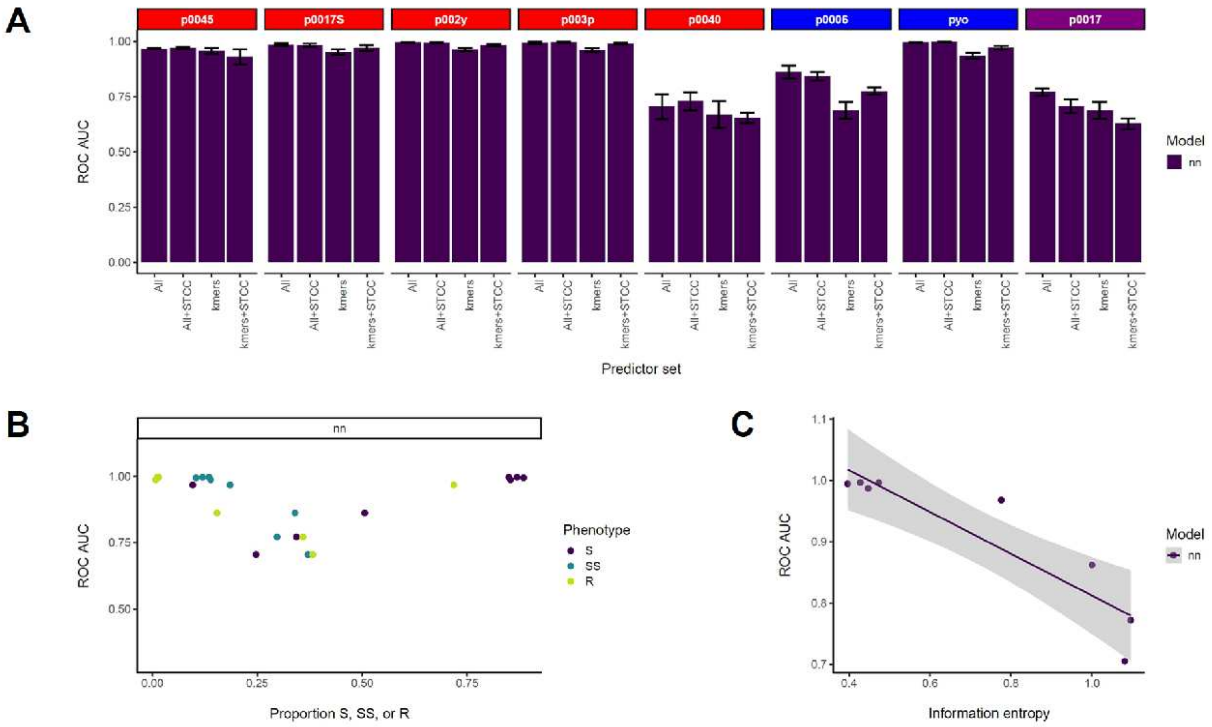
Supplemental Figure S9: Evaluation of ternary phage resistance phenotype predictive models through receiver operating characteristic-area under the curve. Quantitative host range phenotypes were classified as S - sensitive, SS - semi-sensitive, or R - resistant based on the bins 0.1-0.4, 0.4-0.7, and 0.7 or more, respectively. Data preprocessing included oversampling (p0045, p0017S, p002y, p003p, or pyo), lasso regression (p0017), both (p0006), or neither (p0040). A) 10-fold cross-validation ROC AUCs for each phage based on two model building methods (randomForest and XGBoost) and four sets of predictors - all significant GWAS genetic determinants (COGs, SNPs, and k-mers) for a particular phage, all determinants plus corresponding strain sequence type and clonal complex (ST and CC), significant k-mers for a particular phage, and significant k-mers plus strain ST and CC. Average ROC AUCs of four 10-fold CV replicates are presented with one standard error above and below the mean. B) Average ROC AUCs from four 10-fold CV replicates for each model building method and all significant GWAS determinants as predictors relative to the proportion of each ternary phenotype (S, SS, or

R) amongst tested strains for the corresponding phage. Three points are shown for each ROC AUC (corresponding to each of the three possible phenotypes). C) Average ROC AUCs from four 10-fold CV replicates for each model building method and all significant GWAS determinants as predictors relative to the information entropy for each host range phenotype, which was calculated as described in the Materials and Methods section. Information entropy was calculated with a natural logarithm in natural units (nats). *Siphoviridae* are listed in red, *Myoviridae* in blue, and *Podoviridae* in purple.



Supplemental Figure S10: Evaluation of ternary phage resistance phenotype neural network predictive models through receiver operating characteristic-area under the curve. Quantitative host range phenotypes were classified as S - sensitive, SS - semi-sensitive, or R - resistant based on the bins 0.1-0.4, 0.4-0.7, and 0.7 or more, respectively. A) ROC AUCs for each phage based on neural network models and four sets of predictors - all significant GWAS genetic determinants (COGs, SNPs, and k-mers) for a particular phage, all determinants plus corresponding strain

sequence type and clonal complex (ST and CC), significant k-mers for a particular phage, and significant k-mers plus strain ST and CC. Average ROC AUCs of four replicates are presented with one standard error above and below the mean. B) Average ROC AUCs from four replicates and all significant GWAS determinants as predictors relative to the proportion of each ternary phenotype (S, SS, or R) amongst tested strains for the corresponding phage. Three points are shown for each ROC AUC (corresponding to each of the three possible phenotypes). C) Average ROC AUCs from four replicates and all significant GWAS determinants as predictors relative to the information entropy for each host range phenotype, which was calculated as described in the Materials and Methods section. Information entropy was calculated with a natural logarithm in natural units (nats). *Siphoviridae* are listed in red, *Myoviridae* in blue, and *Podoviridae* in purple.



Supplemental Table S1: To determine diversity of the phages used in this study, we calculated average nucleotide identities (ANIs) with fastANI 1.31 (286). The phages were sequenced with Oxford Nanopore or Oxford Nanopore and Illumina technologies. p0017 and pyo genomes were assembled from nanopore reads with canu 2.0 (140) while p0045, p0017S, p002y, p003p, p0040, and p0006 genomes were assembled from Illumina and nanopore reads with Unicycler 0.4.8 (285).

p0045								
p0017S	99.59							
p002y	99.59	99.98						
p003p	97.83	97.80	97.81					
p0040	99.47	99.85	99.72	97.75				
p0006	NA	NA	NA	NA	NA			
pyo (ONT)	NA	NA	NA	NA	NA	98.88		
p0017 (ONT)	NA	NA	NA	NA	NA	NA	NA	
	p0045	p0017S	p002y	p003p	p0040	p0006	pyo (ONT)	p0017 (ONT)

Supplemental Table S2 (https://figshare.com/account/projects/94019/articles/13355933): Excel spreadsheet including tested *S. aureus* strain names, quantitative phenotypes, qualitative phenotypes (quantitative phenotypes from Supplemental Table S1 converted to qualitative by the OD_{600} scale 0.1-0.4 for S - sensitive, 0.4-0.7 for SS - semi-sensitive, and 0.7 or more for R - resistant), BioProject, BioSample, and SRA accessions, sequence types (STs), clonal complexes (CCs), isolation dates, and isolation locations.

Supplemental Table S3: Primers used to amplify wild-type genes corresponding to transposon mutants and clone them into pOS1-P*lgt* with splicing overlap extension (SOE)-PCR or HiFi/Gibson assembly. The annealing portion is **bolded**. Tms were calculated for NEB Q5 HF

polymerase (https://tmcalculator.neb.com/#!/batch). *fmtC* primers were used to amplify *fmtC* from USA300 and NRS209.

Primer	Sequence (5' to 3')	Tm (gene)	Tm (pOS1)
trpA LF	GGGATAAATACAATTGAGGTGAACATATGC**GCAAATGACTAAATTATTTATACC**	55	n/a
trpA LR	GGTATAAATAATTTAGTCATTTGC**GCATATGTTCACCTCAATTGTATTTATCCC**	n/a	65
trpA RF	CCAACAAACATTGAATAATTAAG**TCGAGGATCCAAACAAGGGGG**	n/a	69
trpA RR	CCCCCTTGTTTGGATCCTCGA**CTTAATTATTCAATGTTTGTTGG**	55	n/a
phoR LF	GGGATAAATACAATTGAGGTGAACATATGC**AAGAACAATGATGAAGTTTC**	55	n/a
phoR LR	GAAACTTCATCATTGTTCTT**GCATATGTTCACCTCAATTGTATTTATCCC**	n/a	65
phoR RF	CAAAGTTATTCTAAAAGATTATAAAGAATAA**TCGAGGATCCAAACAAGGGGG**	n/a	69
phoR RR	CCCCCTTGTTTGGATCCTCGA**TTATTCTTTATAATCTTTTAGAATAACTTTG**	55	n/a

isdB LF	GGGATAAATACAATTGAGGTGAACATATGC**TTTCTA CAACATGAACAAAC**	55	n/a
isdB LR	GTTTGTTCATGTTGTAGAAA**GCATATGTTCACCTCA ATTGTATTTATCCC**	n/a	65
isdB RF	CGTAAAAACTAATAAATCGTCT**TCGAGGATCCAAA CAAGGGGG**	n/a	69
isdB RR	CCCCCTTGTTTGGATCCTCGA**AGACGATTTATTAG TTTTTACG**	55	n/a
sodM LF	GGGATAAATACAATTGAGGTGAACATATGC**GAATA TACTTATGGCATTTAAATTAC**	55	n/a
sodM LR	GTAATTTAAATGCCATAAGTATATTC**GCATATGTTC ACCTCAATTGTATTTATCCC**	n/a	65
sodM RF	CAAGCAGCAAAATAATATAACTTAA**TCGAGGATCCA AACAAGGGGG**	n/a	69
sodM RR	CCCCCTTGTTTGGATCCTCGA**TTAAGTTATATTATT TTGCTGCTTG**	57	n/a
fmtC LF	GGGATAAATACAATTGAGGTGAACATATGC**GTGAA AAAATGAATCAGGAAG**	56	n/a
fmtC LR	CTTCCTGATTCATTTTTTCAC**GCATATGTTCACCTC AATTGTATTTATCCC**	n/a	65

fmtC RF	CGTCACAAATAATTAAAATCC**TCGAGGATCCAAAC AAGGGGG**	n/a	69
fmtC RR	CCCCCTTGTTTGGATCCTCGA**GGATTTTAATTATTT GTGACG**	54	n/a
relA LF	GGGATAAATACAATTGAGGTGAACATATGC**GTATC ATATAATGAACAACGAATATCC**	59	n/a
relA LR	GGATATTCGTTGTTCATTATATGATAC**GCATATGTT CACCTCAATTGTATTTATCCC**	n/a	65
relA RF	GTTTGGAACTAGAGGTGCAAAA**TCGAGGATCCAAA CAAGGGGG**	n/a	69
relA RR	CCCCCTTGTTTGGATCCTCGA**TTTTGCACCTCTAGT TCCAAAC**	63	n/a

Supplemental Table S4: Significant differences ($p<0.05$; p-values listed in parentheses) in phage host range phenotypes between tested strains' CCs based on Tukey HSD/one-way ANOVA tests.

Phage	Significant comparisons
p0045	CC30 vs. CC8 (3.25e-3)
p0006	CC45 (1.91e-3) and 72 (6.07e-3) vs. CC1; CC45 (1.95e-3) and 72 (0.0141) vs. CC5; CC45 (4.90e-4) and CC72 (5.05e-3) vs. CC8; CC45 vs. CC30 (0.0193)
p0017	CC1 (9.54e-5), 5 (2.03e-4), 30 (0.0170), and 72 (0.0140) vs. CC8

p0017S	CC1 (0.0155) and 97 (0.0416) vs. CC80
p002y	CC22 (0.0280), 45 (3.21e-5), and 80 (0.0177) vs. CC1; CC22 (0.0457), 45 (4.72e-6), and 80 (0.0365) vs. CC5; CC8 (2.18e-4), 30 (1.28e-3), and 97 (8.88e-3) vs. CC45
p003p	CC22 (0.0142), 45 (4.20e-5), 80 (9.05e-5), 88 (0.0369), and 398 (4.91e-3) vs. CC1; CC22 (0.0227), 45 (7.53e-6), 80 (2.08e-4), and 398 (0.0112) vs. CC5; CC45 (5.12e-4), 80 (7.25e-4), and 398 (0.0431) vs. CC8; CC10 (8.55e-3), 12 (0.0148), 15 (3.39e-3), 25 (0.0483), 30 (1.01e-3), 59 (0.0106), 97 (5.43e-4), and 121 (7.92e-3) vs. CC80; CC30 (2.48e-3) and 97 (0.0102) vs. CC45; CC398 (0.0343) vs. CC97
p0040	CC5 (5.83e-5), 30 (1.95e-6), and 121 (9.04e-5) vs. CC8
pyo	CC45 (2.96e-3), 80 (3.86e-3), 88 (8.90e-3), 121 (8.94e-3), and 398 (0.0167) vs. CC1; CC30 (0.0138), 45 (3.10e-5), 80 (2.47e-3), 88 (6.01e-3), 121 (3.54e-5), and 398 (7.79e-3) vs. CC5; CC45 (1.19e-3), 80 (6.58e-3), 88 (0.0150), 121 (2.58e-3), and 398 (0.0275) vs. CC8; CC22 (0.0482) and 30 (0.0492) vs. CC80

Supplemental Table S5: Summary statistics for protein-protein interaction networks identified with STRING amongst genes corresponding to significant SNPs or k-mers (inside or adjacent to genes). PPI enrichment p-value corresponds to the likelihood nodes and edges would be selected from the *S. aureus* database by chance.

Phage	p0045	p0006	p0017	p0017S	p002y	p003p	p0040	pyo

number of nodes	39	33	49	25	405	214	25	164
number of edges	18	6	26	6	1075	347	10	304
average node degree	0.92	0.36	1.06	0.48	5.31	3.24	0.8	3.71
avg. local clustering coefficient	0.36	0.30	0.33	0.21	0.38	0.38	0.52	0.38
expected number of edges	6	6	18	4	930	287	5	243
PPI enrichment p-value	0.000 15	0.62	0.038	0.23	1.77E-06	3.24E-04	0.036	9.45E-05

Supplemental Table S6: Functions enriched amongst gene sets analyzed with STRING databases. Sets of gene names corresponding to significant COGs, SNPs, and k-mers were created as described in the Materials and Methods section.

Term ID	term description	observed gene count	background gene count	false discovery rate
p0045				
PFAM Protein Domains				

PF10651	Domain of unknown function (DUF2479)	2	2	0.041
p0006				
none				
p0017				
SMART Protein Domains				
SM00062	Bacterial periplasmic substrate-binding proteins	2	2	0.015
SM00287	Bacterial SH3 domain homologues	2	4	0.0183
Reference publications				
PMID:29270158	(2017) Commercial Biocides Induce Transfer of Prophage Phi13 from Human Strains of Staphylococcus aureus to Livestock CC398.	4	6	0.007
PMID:28515479	(2017) Acquisition of virulence factors in livestock-associated MRSA: Lysogenic conversion of CC398 strains by virulence gene-containing phages.	3	4	0.0397
p0017S				
PFAM Protein Domains				
PF01420	Type I restriction modification DNA specificity domain	2	2	0.0149

INTERPRO Protein Domains and Features				
IPR000055	Restriction endonuclease, type I, HsdS	2	2	0.02
p002y				
Biological Process (GO)				
GO:0008152	metabolic process	105	530	0.0464
Molecular Function (GO)				
GO:0046872	metal ion binding	36	135	0.0403
GO:0005488	binding	81	393	0.0403
GO:0003824	catalytic activity	91	451	0.0403
GO:0043167	ion binding	58	270	0.0425
UniProt Keywords				
KW-0067	ATP-binding	56	244	0.0442
p003p				
none				
p0040				
KEGG Pathways				
sauw00051	Fructose and mannose metabolism	3	15	0.0072

PFAM Protein Domains				
PF10651	Domain of unknown function (DUF2479)	2	2	0.0094
PF05031	Iron Transport-associated domain	2	4	0.0116
INTERPRO Protein Domains and Features				
IPR018913	BppU, N-terminal	2	2	0.0208
IPR037250	NEAT domain superfamily	2	4	0.0257
IPR006635	NEAT domain	2	4	0.0257
SMART Protein Domains				
SM00725	NEAr Transporter domain	2	4	0.0107
pyo				
Keyword				
KW-0479	Metal-binding	33	218	3.66E-05
KW-0963	Cytoplasm	38	310	0.00025
KW-0560	Oxidoreductase	20	133	0.004
KW-0460	Magnesium	14	83	0.0131
KW-0808	Transferase	29	279	0.0243
KW-0143	Chaperone	6	19	0.0311
KW-0456	Lyase	12	74	0.0311

KEGG				
sauw01100	Metabolic pathways	57	424	1.00E-08
sauw01110	Biosynthesis of secondary metabolites	29	214	0.00041
sauw00240	Pyrimidine metabolism	9	40	0.0174
sauw01130	Biosynthesis of antibiotics	20	164	0.0185
sauw00130	Ubiquinone and other terpenoid-quinone biosynthesis	4	8	0.0263
sauw00260	Glycine, serine and threonine metabolism	7	29	0.0263
sauw01120	Microbial metabolism in diverse environments	16	123	0.0263
sauw03070	Bacterial secretion system	4	9	0.03
Component				
GO:0005737	cytoplasm	41	334	3.23E-05
GO:0044424	intracellular part	43	351	3.23E-05
GO:0044464	cell part	49	495	0.00034
GO:0044444	cytoplasmic part	16	128	0.0216
Function				
GO:0043167	ion binding	39	270	1.05E-05
GO:0046872	metal ion binding	26	135	1.05E-05

GO:0003824	catalytic activity	51	451	2.17E-05
GO:0005488	binding	45	393	7.96E-05
GO:0036094	small molecule binding	25	191	0.0029
GO:0097159	organic cyclic compound binding	33	305	0.0052
GO:1901363	heterocyclic compound binding	33	305	0.0052
GO:0046914	transition metal ion binding	8	26	0.0056
GO:0016740	transferase activity	20	147	0.006
GO:0048037	cofactor binding	11	52	0.006
GO:0043168	anion binding	22	179	0.0089
GO:0000287	magnesium ion binding	5	10	0.0107
GO:0005515	protein binding	6	17	0.0123
GO:0000166	nucleotide binding	21	177	0.0146
GO:0008144	drug binding	17	147	0.0471
GO:0097367	carbohydrate derivative binding	18	159	0.0471
GO:0005524	ATP binding	16	136	0.0495
Process				
GO:0009987	cellular process	58	519	3.41E-05
GO:0008152	metabolic process	58	530	3.45E-05

GO:0044237	cellular metabolic process	53	469	3.45E-05
GO:0044281	small molecule metabolic process	32	208	3.45E-05
GO:0071704	organic substance metabolic process	55	492	3.45E-05
GO:0006807	nitrogen compound metabolic process	48	423	8.62E-05
GO:0044238	primary metabolic process	49	439	8.62E-05
GO:1901564	organonitrogen compound metabolic process	36	306	0.0011
GO:0034641	cellular nitrogen compound metabolic process	35	316	0.0041
GO:0006082	organic acid metabolic process	20	135	0.0054
GO:0046483	heterocycle metabolic process	29	248	0.006
GO:1901576	organic substance biosynthetic process	35	326	0.006
GO:0019752	carboxylic acid metabolic process	18	118	0.0063
GO:1901565	organonitrogen compound catabolic process	8	25	0.0065
GO:0006725	cellular aromatic compound metabolic process	28	243	0.0067
GO:0044249	cellular biosynthetic process	34	322	0.0067
GO:1901360	organic cyclic compound metabolic process	29	256	0.0067
GO:0044282	small molecule catabolic process	9	34	0.0069

GO:0009056	catabolic process	12	61	0.0072
GO:0044248	cellular catabolic process	11	52	0.0072
GO:0016054	organic acid catabolic process	7	22	0.0118
GO:0008610	lipid biosynthetic process	8	30	0.0122
GO:1901135	carbohydrate derivative metabolic process	16	109	0.0122
GO:0006139	nucleobase-containing compound metabolic process	24	208	0.0131
GO:0009063	cellular amino acid catabolic process	6	16	0.0131
GO:1901606	alpha-amino acid catabolic process	6	16	0.0131
GO:0055086	nucleobase-containing small molecule metabolic process	12	69	0.014
GO:1901575	organic substance catabolic process	11	60	0.0153
GO:0044255	cellular lipid metabolic process	8	33	0.0162
GO:0006520	cellular amino acid metabolic process	13	83	0.0184
GO:0006796	phosphate-containing compound metabolic process	15	110	0.0268
GO:1901605	alpha-amino acid metabolic process	10	58	0.0319
GO:0019637	organophosphate metabolic process	12	83	0.0438
GO:0006457	protein folding	4	9	0.0454

Supplemental Table S7: Functions enriched amongst gene sets analyzed with PANTHER databases. Sets of gene names corresponding to significant COGs, SNPs, and k-mers were created as described in the Materials and Methods section. FDR means False Discovery Rate.

PANTHER GO-Slim Biological Process	Staphylococcus aureus - REFLIST (2889)	p0045 (24)	p0045 (expected)	p0045 (over/under)	p0045 (fold Enrichment)	p0045 (raw P-value)	p0045 (FDR)
regulation of nucleic acid-templated transcription (GO:1903506)	45	4	0.37	+	10.7	5.87E-04	1.17E-01
regulation of RNA biosynthetic process (GO:2001141)	45	4	0.37	+	10.7	5.87E-04	8.74E-02
nucleic acid-templated transcription (GO:0097659)	46	4	0.38	+	10.47	6.34E-04	7.56E-02
regulation of RNA metabolic process (GO:0051252)	46	4	0.38	+	10.47	6.34E-04	6.30E-02
regulation of nucleobase-containing compound metabolic process (GO:0019219)	47	4	0.39	+	10.24	6.84E-04	5.83E-02

RNA biosynthetic process (GO:0032774)	47	4	0.39	+	10.24	6.84E-04	5.10E-02
regulation of macromolecule biosynthetic process (GO:0010556)	51	4	0.42	+	9.44	9.13E-04	6.05E-02
regulation of biosynthetic process (GO:0009889)	51	4	0.42	+	9.44	9.13E-04	5.44E-02
regulation of cellular biosynthetic process (GO:0031326)	51	4	0.42	+	9.44	9.13E-04	4.95E-02
regulation of cellular macromolecule biosynthetic process (GO:2000112)	51	4	0.42	+	9.44	9.13E-04	4.54E-02
regulation of nitrogen compound metabolic process (GO:0051171)	52	4	0.43	+	9.26	9.78E-04	4.48E-02
regulation of primary metabolic process (GO:0080090)	52	4	0.43	+	9.26	9.78E-04	4.16E-02
regulation of gene expression (GO:0010468)	52	4	0.43	+	9.26	9.78E-04	3.89E-02

regulation of cellular metabolic process (GO:0031323)	53	4	0.44	+	9.08	1.05E-03	3.90E-02
regulation of macromolecule metabolic process (GO:0060255)	53	4	0.44	+	9.08	1.05E-03	3.67E-02
regulation of metabolic process (GO:0019222)	54	4	0.45	+	8.92	1.12E-03	3.70E-02
macromolecule biosynthetic process (GO:0009059)	123	6	1.02	+	5.87	4.68E-04	2.79E-01
cellular macromolecule biosynthetic process (GO:0034645)	123	6	1.02	+	5.87	4.68E-04	1.39E-01
RNA metabolic process (GO:0016070)	104	5	0.86	+	5.79	1.61E-03	4.57E-02
nucleic acid metabolic process (GO:0090304)	148	6	1.23	+	4.88	1.20E-03	3.78E-02
Nucleobase-containing compound metabolic process	211	7	1.75	+	3.99	1.37E-03	4.09E-02

(GO:0006139)							
PANTHER Protein Class	Staphylococcus aureus - REFLIST (2889)	p0017 (38)	p0017 (expected)	p0017 (over/under)	p0017 (fold Enrichment)	p0017 (raw P-value)	p0017 (FDR)
protein class (PC00000)	986	25	12.97	+	1.93	9.62E-05	4.95E-03
Unclassified (UNCLASSIFIED)	1903	13	25.03	-	0.52	9.62E-05	9.91E-03
GO biological process complete	Staphylococcus aureus - REFLIST (2889)	p002y (326)	p002y (expected)	p002y (over/under)	p002y (fold Enrichment)	p002y (raw P-value)	p002y (FDR)
organic substance metabolic process (GO:0071704)	878	143	99.08	+	1.44	1.62E-06	6.35E-04
cellular metabolic process (GO:0044237)	837	136	94.45	+	1.44	4.19E-06	1.31E-03
metabolic process (GO:0008152)	1030	167	116.23	+	1.44	6.30E-08	3.29E-05
primary metabolic process (GO:0044238)	748	121	84.41	+	1.43	3.14E-05	7.01E-03

nitrogen compound metabolic process (GO:0006807)	717	115	80.91	+	1.42	7.77E-05	1.52E-02
biological_process (GO:0008150)	1496	237	168.81	+	1.4	3.16E-13	2.48E-10
cellular process (GO:0009987)	947	146	106.86	+	1.37	2.00E-05	5.21E-03
Unclassified (UNCLASSIFIED)	1393	89	157.19	-	0.57	3.16E-13	4.95E-10
GO molecular function complete	Staphylococcus aureus - REFLIST (2889)	p002y (326)	p002y (expected)	p002y (over/under)	p002y (fold Enrichment)	p002y (raw P-value)	p002y (FDR)
ATP binding (GO:0005524)	275	58	31.03	+	1.87	1.37E-05	1.23E-03
adenyl ribonucleotide binding (GO:0032559)	276	58	31.14	+	1.86	2.01E-05	1.38E-03
adenyl nucleotide binding (GO:0030554)	277	58	31.26	+	1.86	2.07E-05	1.34E-03
DNA binding (GO:0003677)	215	45	24.26	+	1.85	2.28E-04	1.21E-02
carbohydrate derivative binding	320	66	36.11	+	1.83	5.86E-06	8.54E-04

(GO:0097367)							
drug binding (GO:0008144)	311	64	35.09	+	1.82	1.01E-05	1.18E-03
purine ribonucleoside triphosphate binding (GO:0035639)	302	62	34.08	+	1.82	1.80E-05	1.50E-03
ribonucleotide binding (GO:0032553)	312	64	35.21	+	1.82	1.06E-05	1.12E-03
purine ribonucleotide binding (GO:0032555)	303	62	34.19	+	1.81	1.85E-05	1.44E-03
purine nucleotide binding (GO:0017076)	304	62	34.3	+	1.81	1.92E-05	1.40E-03
anion binding (GO:0043168)	378	74	42.65	+	1.73	7.40E-06	9.58E-04
nucleotide binding (GO:0000166)	382	72	43.11	+	1.67	3.55E-05	2.18E-03
nucleoside phosphate binding (GO:1901265)	382	72	43.11	+	1.67	3.55E-05	2.07E-03

ion binding (GO:0043167)	569	107	64.21	+	1.67	1.37E-07	2.29E-05
small molecule binding (GO:0036094)	427	80	48.18	+	1.66	1.33E-05	1.29E-03
nucleic acid binding (GO:0003676)	353	66	39.83	+	1.66	1.20E-04	6.65E-03
hydrolase activity (GO:0016787)	364	65	41.07	+	1.58	4.08E-04	2.07E-02
heterocyclic compound binding (GO:1901363)	710	126	80.12	+	1.57	1.19E-07	2.77E-05
organic cyclic compound binding (GO:0097159)	710	126	80.12	+	1.57	1.19E-07	2.31E-05
binding (GO:0005488)	898	156	101.33	+	1.54	2.60E-09	7.57E-07
catalytic activity (GO:0003824)	1105	189	124.69	+	1.52	1.10E-11	4.27E-09
molecular_function (GO:0003674)	1567	256	176.82	+	1.45	6.25E-18	3.64E-15
Unclassified (UNCLASSIFIED)	1322	70	149.18	-	0.47	6.25E-18	7.28E-15

PANTHER GO-Slim Molecular Function	Staphylococcus aureus - REFLIST (2889)	p002y (326)	p002y (expected)	p002y (over/under)	p002y (fold Enrichment)	p002y (raw P-value)	p002y (FDR)
catalytic activity (GO:0003824)	498	90	56.2	+	1.6	1.44E-05	1.18E-03
molecular_function (GO:0003674)	729	122	82.26	+	1.48	4.48E-06	5.51E-04
Unclassified (UNCLASSIFIED)	2160	204	243.74	-	0.84	4.48E-06	1.10E-03
PANTHER Protein Class	Staphylococcus aureus - REFLIST (2889)	p002y (326)	p002y (expected)	p002y (over/under)	p002y (fold Enrichment)	p002y (raw P-value)	p002y (FDR)
protein class (PC00000)	986	172	111.26	+	1.55	7.79E-11	4.01E-09
metabolite interconversion enzyme (PC00262)	473	78	53.37	+	1.46	1.08E-03	3.69E-02
Unclassified (UNCLASSIFIED)	1903	154	214.74	-	0.72	7.79E-11	8.02E-09
GO biological process complete	Staphylococcus aureus - REFLIST (2889)	p003p (181)	p003p (expected)	p003p (over/under)	p003p (fold Enrichment)	p003p (raw P-value)	p003p (FDR)
Nucleobase-containing compound metabolic	331	42	20.74	+	2.03	1.74E-05	5.45E-03

process (GO:0006139)							
cellular aromatic compound metabolic process (GO:0006725)	394	46	24.68	+	1.86	4.34E-05	1.13E-02
cellular nitrogen compound metabolic process (GO:0034641)	491	56	30.76	+	1.82	8.37E-06	4.36E-03
heterocycle metabolic process (GO:0046483)	408	46	25.56	+	1.8	1.30E-04	2.26E-02
organic cyclic compound metabolic process (GO:1901360)	418	46	26.19	+	1.76	2.34E-04	3.33E-02
nitrogen compound metabolic process (GO:0006807)	717	69	44.92	+	1.54	1.43E-04	2.24E-02
cellular metabolic process (GO:0044237)	837	78	52.44	+	1.49	1.05E-04	2.05E-02
metabolic process (GO:0008152)	1030	95	64.53	+	1.47	1.03E-05	4.05E-03

cellular process (GO:0009987)	947	86	59.33	+	1.45	8.91E-05	1.99E-02
organic substance metabolic process (GO:0071704)	878	79	55.01	+	1.44	3.47E-04	4.52E-02
biological_process (GO:0008150)	1496	131	93.73	+	1.4	5.45E-08	8.54E-05
Unclassified (UNCLASSIFIED)	1393	50	87.27	-	0.57	5.45E-08	4.27E-05
GO molecular function complete	Staphylococcus aureus - REFLIST (2889)	p003p (181)	p003p (expected)	p003p (over/under)	p003p (fold Enrichment)	p003p (raw P-value)	p003p (FDR)
nucleotide binding (GO:0000166)	382	43	23.93	+	1.8	2.13E-04	3.10E-02
nucleoside phosphate binding (GO:1901265)	382	43	23.93	+	1.8	2.13E-04	2.75E-02
heterocyclic compound binding (GO:1901363)	710	74	44.48	+	1.66	3.81E-06	8.89E-04
organic cyclic compound binding (GO:0097159)	710	74	44.48	+	1.66	3.81E-06	7.41E-04
ion binding (GO:0043167)	569	58	35.65	+	1.63	1.33E-04	2.22E-02

binding (GO:0005488)	898	88	56.26	+	1.56	2.31E-06	6.75E-04
catalytic activity (GO:0003824)	1105	103	69.23	+	1.49	9.48E-07	3.68E-04
molecular_function (GO:0003674)	1567	137	98.17	+	1.4	7.99E-09	9.31E-06
Unclassified (UNCLASSIFIED)	1322	44	82.83	-	0.53	7.99E-09	4.66E-06
PANTHER GO-Slim Molecular Function	Staphylococcus aureus - REFLIST (2889)	p003p (181)	p003p (expected)	p003p (over/under)	p003p (fold Enrichment)	p003p (raw P-value)	p003p (FDR)
catalytic activity (GO:0003824)	498	52	31.2	+	1.67	2.01E-04	1.65E-02
molecular_function (GO:0003674)	729	72	45.67	+	1.58	3.56E-05	4.38E-03
Unclassified (UNCLASSIFIED)	2160	109	135.33	-	0.81	3.56E-05	8.76E-03
PANTHER GO-Slim Biological Process	Staphylococcus aureus - REFLIST (2889)	p003p (181)	p003p (expected)	p003p (over/under)	p003p (fold Enrichment)	p003p (raw P-value)	p003p (FDR)
nitrogen compound metabolic process (GO:0006807)	371	42	23.24	+	1.81	2.74E-04	4.08E-02

cellular metabolic process (GO:0044237)	425	48	26.63	+	1.8	7.05E-05	4.20E-02
organic substance metabolic process (GO:0071704)	415	45	26	+	1.73	3.43E-04	4.09E-02
metabolic process (GO:0008152)	443	48	27.75	+	1.73	2.16E-04	4.30E-02
cellular process (GO:0009987)	485	52	30.39	+	1.71	1.10E-04	3.29E-02
PANTHER Protein Class	Staphylococcus aureus - REFLIST (2889)	p003p (181)	p003p (expected)	p003p (over/under)	p003p (fold Enrichment)	p003p (raw P-value)	p003p (FDR)
protein class (PC00000)	986	95	61.77	+	1.54	1.16E-06	5.99E-05
Unclassified (UNCLASSIFIED)	1903	86	119.23	-	0.72	1.16E-06	1.20E-04
GO biological process complete	Staphylococcus aureus - REFLIST (2889)	pyo (146)	pyo (expected)	pyo (over/under)	pyo (fold Enrichment)	pyo (raw P-value)	pyo (FDR)
cellular amino acid metabolic process (GO:0006520)	136	19	6.87	+	2.76	1.11E-04	1.33E-02
small molecule metabolic process	375	45	18.95	+	2.37	4.79E-08	9.37E-06

(GO:0044281)							
carboxylic acid metabolic process (GO:0019752)	198	23	10.01	+	2.3	4.17E-04	3.63E-02
oxoacid metabolic process (GO:0043436)	202	23	10.21	+	2.25	4.69E-04	3.67E-02
organic acid metabolic process (GO:0006082)	222	25	11.22	+	2.23	2.56E-04	2.50E-02
cellular biosynthetic process (GO:0044249)	469	48	23.7	+	2.03	1.45E-06	2.52E-04
organic substance biosynthetic process (GO:1901576)	476	48	24.06	+	2	2.65E-06	4.14E-04
biosynthetic process (GO:0009058)	495	49	25.02	+	1.96	3.75E-06	5.34E-04
Nucleobase-containing compound metabolic process (GO:0006139)	331	32	16.73	+	1.91	5.50E-04	4.10E-02

organic substance metabolic process (GO:0071704)	878	84	44.37	+	1.89	5.65E-11	4.42E-08
cellular metabolic process (GO:0044237)	837	80	42.3	+	1.89	2.49E-10	6.50E-08
cellular aromatic compound metabolic process (GO:0006725)	394	37	19.91	+	1.86	2.30E-04	2.40E-02
primary metabolic process (GO:0044238)	748	70	37.8	+	1.85	2.87E-08	6.41E-06
metabolic process (GO:0008152)	1030	95	52.05	+	1.83	3.57E-12	5.59E-09
cellular process (GO:0009987)	947	87	47.86	+	1.82	1.39E-10	4.35E-08
heterocycle metabolic process (GO:0046483)	408	37	20.62	+	1.79	4.54E-04	3.74E-02
cellular nitrogen compound metabolic process (GO:0034641)	491	43	24.81	+	1.73	3.12E-04	2.87E-02

organonitrogen compound metabolic process (GO:1901564)	514	45	25.98	+	1.73	1.79E-04	2.00E-02
nitrogen compound metabolic process (GO:0006807)	717	62	36.23	+	1.71	6.12E-06	7.98E-04
biological_process (GO:0008150)	1496	115	75.6	+	1.52	7.10E-11	2.78E-08
Unclassified (UNCLASSIFIED)	1393	31	70.4	-	0.44	7.10E-11	3.71E-08
GO molecular function complete	Staphylococcus aureus - REFLIST (2889)	pyo (146)	pyo (expected)	pyo (over/under)	pyo (fold Enrichment)	pyo (raw P-value)	pyo (FDR)
protein dimerization activity (GO:0046983)	11	5	0.56	+	8.99	6.83E-04	3.46E-02
protein binding (GO:0005515)	28	9	1.42	+	6.36	4.22E-05	3.08E-03
oxidoreductase activity, acting on CH-OH group of donors (GO:0016614)	41	11	2.07	+	5.31	2.41E-05	1.87E-03
magnesium ion binding	53	11	2.68	+	4.11	1.80E-04	1.10E-02

(GO:0000287)							
transition metal ion binding (GO:0046914)	83	16	4.19	+	3.81	1.30E-05	1.38E-03
coenzyme binding (GO:0050662)	111	16	5.61	+	2.85	2.92E-04	1.62E-02
cofactor binding (GO:0048037)	154	22	7.78	+	2.83	2.19E-05	2.13E-03
metal ion binding (GO:0046872)	271	37	13.7	+	2.7	5.05E-08	9.82E-06
cation binding (GO:0043169)	276	37	13.95	+	2.65	7.83E-08	1.30E-05
oxidoreductase activity (GO:0016491)	182	22	9.2	+	2.39	2.43E-04	1.42E-02
transferase activity, transferring phosphorus-containing groups (GO:0016772)	161	19	8.14	+	2.34	8.91E-04	4.33E-02
ion binding (GO:0043167)	569	62	28.76	+	2.16	1.59E-09	3.71E-07
small molecule binding	427	46	21.58	+	2.13	8.69E-07	1.27E-04

(GO:0036094)							
nucleotide binding (GO:0000166)	382	39	19.3	+	2.02	2.30E-05	2.06E-03
nucleoside phosphate binding (GO:1901265)	382	39	19.3	+	2.02	2.30E-05	1.91E-03
transferase activity (GO:0016740)	366	37	18.5	+	2	5.00E-05	3.43E-03
drug binding (GO:0008144)	311	31	15.72	+	1.97	3.83E-04	2.03E-02
anion binding (GO:0043168)	378	37	19.1	+	1.94	1.03E-04	6.69E-03
binding (GO:0005488)	898	85	45.38	+	1.87	7.44E-11	2.17E-08
catalytic activity (GO:0003824)	1105	101	55.84	+	1.81	2.43E-13	9.43E-11
heterocyclic compound binding (GO:1901363)	710	62	35.88	+	1.73	5.52E-06	7.15E-04
organic cyclic compound binding (GO:0097159)	710	62	35.88	+	1.73	5.52E-06	6.43E-04
molecular_function	1567	128	79.19	+	1.62	3.24E-17	3.78E-

(GO:0003674)							14
Unclassified (UNCLASSIFIED)	1322	18	66.81	-	0.27	3.24E-17	1.89E-14
GO cellular component complete	Staphylococcus aureus - REFLIST (2889)	pyo (146)	pyo (expected)	pyo (over/under)	pyo (fold Enrichment)	pyo (raw P-value)	pyo (FDR)
cytoplasm (GO:0005737)	332	39	16.78	+	2.32	8.66E-07	1.01E-04
intracellular (GO:0005622)	411	42	20.77	+	2.02	1.21E-05	7.08E-04
PANTHER Pathways	Staphylococcus aureus - REFLIST (2889)	pyo (146)	pyo (expected)	pyo (over/under)	pyo (fold Enrichment)	pyo (raw P-value)	pyo (FDR)
Unclassified (UNCLASSIFIED)	2663	117	134.58	-	0.87	8.09E-06	5.67E-04
PANTHER GO-Slim Molecular Function	Staphylococcus aureus - REFLIST (2889)	pyo (146)	pyo (expected)	pyo (over/under)	pyo (fold Enrichment)	pyo (raw P-value)	pyo (FDR)
catalytic activity (GO:0003824)	498	53	25.17	+	2.11	1.10E-07	2.72E-05
molecular_function (GO:0003674)	729	66	36.84	+	1.79	3.66E-07	4.50E-05
Unclassified (UNCLASSIFIED)	2160	80	109.16	-	0.73	3.66E-07	3.00E-05

PANTHER GO-Slim Biological Process	Staphylococcus aureus - REFLIST (2889)	pyo (146)	pyo (expected)	pyo (over/under)	pyo (fold Enrichment)	pyo (raw P-value)	pyo (FDR)
metabolic process (GO:0008152)	443	44	22.39	+	1.97	1.31E-05	7.82E-03
cellular metabolic process (GO:0044237)	425	40	21.48	+	1.86	1.30E-04	1.55E-02
organic substance metabolic process (GO:0071704)	415	39	20.97	+	1.86	1.82E-04	1.80E-02
cellular process (GO:0009987)	485	45	24.51	+	1.84	4.68E-05	1.40E-02
biological_process (GO:0008150)	606	52	30.63	+	1.7	7.70E-05	1.53E-02
Unclassified (UNCLASSIFIED)	2283	94	115.37	-	0.81	7.70E-05	1.15E-02
PANTHER GO-Slim Cellular Component	Staphylococcus aureus - REFLIST (2889)	pyo (146)	pyo (expected)	pyo (over/under)	pyo (fold Enrichment)	pyo (raw P-value)	pyo (FDR)
cytoplasm (GO:0005737)	265	30	13.39	+	2.24	4.36E-05	4.05E-03
cytoplasmic part (GO:0044444)	184	20	9.3	+	2.15	1.82E-03	4.22E-02

intracellular part (GO:0044424)	288	31	14.55	+	2.13	8.13E-05	3.78E-03
intracellular (GO:0005622)	292	31	14.76	+	2.1	1.46E-04	4.54E-03
PANTHER Protein Class	Staphylococcus aureus - REFLIST (2889)	pyo (146)	pyo (expected)	pyo (over/under)	pyo (fold Enrichment)	pyo (raw P-value)	pyo (FDR)
metabolite interconversion enzyme (PC00262)	473	50	23.9	+	2.09	4.06E-07	1.39E-05
protein class (PC00000)	986	82	49.83	+	1.65	1.68E-07	1.73E-05
Unclassified (UNCLASSIFIED)	1903	64	96.17	-	0.67	1.68E-07	8.64E-06

Chapter 4: Species-scale genomic analysis of known phage resistance mechanisms and their relationships to horizontal gene transfer in *S. aureus*

Abraham G. Moller, Robert A. Petit III, and Timothy D. Read

Abstract

Increasing antibiotic resistance and vaccine development failure for the ubiquitous opportunistic pathogen *Staphylococcus aureus* make alternative therapies necessary. One possible alternative is phage therapy, in which bacteriophages kill infecting *S. aureus*. However, successful phage therapy requires knowing both host and phage genetic factors influencing host range for rational cocktail formulation. In addition, phages play an important role in *S. aureus* evolution through horizontal gene transfer (HGT). To address these questions, we searched 40,000+ *S. aureus* genomes for literature-annotated phage resistance genes to investigate relationships between predicted phage resistance, empirically determined phage resistance, and genes acquired by HGT. We found that phage adsorption targets and genes that block phage assembly were significantly more conserved than genes targeting phage biosynthesis. Core phage resistance genes had significantly similar nucleotide diversity, selection (dN/dS), and functionality (delta-BS) to all core genes in a set of 380 non-redundant *S. aureus* genomes (each from a different MLST sequence type). Non-core phage resistance genes were significantly less consistent with the core genome phylogeny than all non-core genes in this set. Only superinfection immunity genes correlated with empirically determined temperate phage resistance, accessory genome content, and numbers of accessory antibiotic resistance or virulence genes encoded per strain. Taken together, these results suggest that while phage adsorption genes are heavily conserved in the *S. aureus* species, they are not undergoing positive selection, arms race dynamics. They also suggest assembly genes are least phylogenetically constrained and superinfection immunity genes best predict both empirical phage resistance and

levels of phage-mediated HGT. These analyses thus show both that temperate phage host range limits their use as therapeutic agents and also that evolution to escape phage infection has been a major selective pressure driving *S. aureus* evolution.

Introduction

New treatments are needed for *Staphylococcus aureus* infections because of its high prevalence, increasing antibiotic resistance, diverse pathologies, and a lack of available vaccines. One alternative for *S. aureus* treatment is phage therapy - clearing infecting bacteria with bacteriophages. Phage therapy advantages over antibiotics include reduced toxicity and the high diversity of natural phages available to be isolated for treatment (96, 171). However, no natural phage is known to kill all *S. aureus* strains, making phage cocktails (combinations of phages that have non-overlapping host ranges) necessary for successful treatment. Comprehensive knowledge of host and phage genetic factors influencing host range is needed for rational cocktail formulation.

There have been recent reviews of *S. aureus* phages (39, 101) and their corresponding resistance/host range regulating mechanisms (53). Currently known *S. aureus* phages, which all belong to the order *Caudovirales* (tailed phages), are divided into three morphological classes - the long, noncontractile-tailed *Siphoviridae*, the long, contractile-tailed *Myoviridae*, and the short, noncontractile-tailed *Podoviridae* (39). The *Siphoviridae* are both temperate and virulent, while the *Myo*- and *Podoviridae* are virulent (39). The *Siphoviridae* bind either α-O-GlcNAc or β-O-GlcNAc attached at the 4 positions of wall teichoic acid (WTA) ribitol phosphate monomers, the *Podoviridae* bind only β-O-GlcNAc-decorated WTA, and the *Myoviridae* bind either the WTA ribitol-phosphate backbone or β-O-GlcNAc-decorated WTA (53, 196, 273).

Currently identified *Staphylococcus* phage resistance mechanisms act at the adsorption, biosynthesis, and assembly infection stages of the phage life cycle (53). Adsorption resistance mechanisms include phage receptor alteration, removal, or occlusion by large surface proteins or

polysaccharides (capsule) (22, 109–111, 181, 185, 196). Biosynthesis resistance mechanisms involve halting infection through metabolic arrest (abortive infection) or phage DNA degradation by adaptive (CRISPR) or innate (restriction-modification) immunity (116, 237, 241, 274, 275). Assembly resistance occurs by assembly interference, in which *Staphylococcus aureus* pathogenicity islands (SaPIs), chromosomal phage-like elements, divert away the assembly of helper infecting viruses (*Siphoviridae*) toward their own, enabling them to replicate at the cost of helper viruses (79, 117–119, 251, 276).

In this work, we searched 40,000+ annotated *S. aureus* genomes for known phage resistance genes to investigate the relationships between predicted phage resistance, empirically determined phage resistance, and horizontal gene transfer (HGT). We developed scoring metrics for resistance determinant diversity, abundance (frequency amongst strains), functionality, and overlap, as well as overall predicted phage resistance per strain, and accessory gene content as a measure of horizontal gene transfer. We examined phylogenetic signal of phage resistance determinants and levels of overlap between determinants for each strain. We then evaluated the correlation between genome-predicted phage resistance and either empirically determined resistance levels, levels of horizontal gene transfer, or networks of gene transfer amongst strains. We anticipate the conclusions of this work will both improve phage resistance prediction, thus improving phage therapy potential, and understanding the evolution of the *S. aureus* species by elucidation of genetic determinants affecting HGT.

Results

Distribution of phage resistance genes in *S. aureus*

I curated a list of 331 genes implicated from other studies to play a role in phage resistance to search our Staphopia database based on criteria listed in Materials and Methods - namely, 1) show to cause phage resistance in a *S. aureus* strain through molecular genetic studies, 2) site of mutation selected for phage resistance, 3) having causative effects on phage resistance in

other species with *S. aureus* homologs, or 4) having causative effects on phage resistance in other species but lacking known *S. aureus* homologs. The first question we asked about these genes was how well conserved they were in the species, based on a search of the Staphopia database containing 43,000 *S. aureus* genomes. We found that the most conserved (core) genes in the database were mainly adsorption genes (Figure 1). Both adsorption resistance and assembly resistance genes were significantly ($p<0.05$) more conserved than biosynthesis resistance genes (hereafter referred to by resistance category - e.g., adsorption) but adsorption genes did not significantly differ in conservation from assembly genes (Figure 1B) based on non-parametric Wilcoxon tests. Conversely, assembly genes were over-represented in the accessory gene pool. We found the Septu and retron phage defense systems in the species for the first time, amongst other newly discovered phage defense systems.

Diversity, functionality, and selection of core/extended core phage resistance genes was similar to corresponding core/extended core genes

We next assessed the diversity and functionality of core phage resistance genes in the Staphopia non-redundant (NRD) set (260). The NRD set includes one representative genome for every known *S. aureus* sequence type. We asked whether core (present in all genomes of the NRD set) or extended core (present in 80% or more genomes of the NRD set) genes differ in levels of diversity (measured as allele count or translated nucleotide diversity), functionality (measured as delta-bit score - difference between reference and query gene matches to a profile Hidden Markov model, where mutations at conserved amino acids have stronger effects than those at nonconserved amino acids), or selection (measured as nonsynonymous to synonymous change ratio - dN/dS) from respective phage resistance genes. If phages and hosts existed in an arms race scenario, we would expect core phage resistance genes to be undergoing diversifying selection (higher average dN/dS than core genes). We also would expect decreased functionality (increased delta-bit score) in cases where inactivating genes would lead to resistance. We instead

found similar diversity, functionality, and dN/dS between core/extended core genes and corresponding phage resistance gene subsets ($p>0.05$ with non-parametric Wilcoxon tests) (Figure 2). This suggested *S. aureus* was likely not undergoing an arms race dynamic with its phages, as core/extended core phage resistance genes are undergoing purifying selection similar to core/extended core genes in general.

Non-extended core phage resistance genes have phylogenetic signal but assembly genes have the least phylogenetic signal

We also asked whether non-extended core phage resistance genes were unusual in their phylogenetic signal compared to non-extended core genes by calculating consistency indices, which measure how much gene presence/absence diverges from that expected by the phylogeny (in this case, a core genome phylogeny of the NRD set). We define phylogenetic signal as the tendency for a phylogeny to explain a trait at the tips (in this case, gene presence) more than expected at random, such that certain lineages show a trait moreso than other lineages. Homoplasy, on the other hand - the independent loss or gain of a trait in separate lineages during evolution - is inversely proportional to consistency index and phylogenetic signal. Genes on mobile genetic elements that were frequently gained through horizontal gene transfer and lost frequently through deletion/replacement would be expected to have low consistency indices. Extremely rare and extremely common genes would be expected to have high consistency indices, on the other hand, as either circumstance would make apparent gene gains or losses rare. To examine these relationships, we plotted the number of NRD strains encoding the gene of interest against the number of changes necessary, both for actual cases and the average of 999 gene presence-absence permutations per gene (Figure 3A). As expected, the permuted gene presence-absence data generated a parabola with a peak at intermediate-frequency genes. All observed gene change counts were below those expected by the parabola, indicating all genes had a level of phylogenetic signal. However, we found assembly genes had the highest changes

amongst any group considered (Figure 3), though non-extended core adsorption and biosynthesis also were significantly less consistent with the phylogeny than non-extended core genes in general (Figure 3B).

Taken together, these results indicate all non-extended core phage resistance genes had some level of phylogenetic signal, but that assembly genes had the least. Low assembly resistance gene phylogenetic signal is biologically plausible because mobile SaPIs carry assembly resistance genes, leading to extensive horizontal instead of purely vertical transfer. Unlike assembly genes, a substantial number of non-core adsorption and biosynthesis genes approach complete consistency with the tree, but not nearly to the same proportion of non-extended core genes. Nonetheless, we can't discount the possibility that horizontal transfer from other *S. aureus* clades unrepresented in our database or other species could lead to rare genes completely consistent with the phylogeny. For example, a single ancestral strain may have acquired a gene and then descended into strains at the tips that all possess that gene.

Clonal complex (CC) correlates with accessory genome and non-extended core phage resistance genes

We further examined the relationship between phylogeny and phage resistance genes by assessing the differences between clonal complex (CC) in their numbers of non-extended core phage genes. For all gene categories examined, there was statistically significant variation ($p<0.05$) as determined by an analysis of variance (ANOVA) test (Figure 4). This suggests clonal complex is associated with both accessory genome and numbers of phage resistance genes. This result agrees with previous studies correlating type I restriction-modification specificity (*hsdS* gene) alleles with clonal complex (113), which in turn would restrict transduction across clonal complex and make clonal complexes differ from each other in net levels of horizontal gene transfer. Differences in restriction specificity between clonal complexes have been shown to restrict transduction between them in the lab (112), preventing all clonal complexes from

exchanging genetic material evenly when in contact, which would otherwise lead to no variation in accessory genome content amongst clonal complexes.

Non-extended core phage resistance genes are most often co-encoded with those of the same class

We next examined the modularity of phage resistance genes by examining genomic overlap (co-encoding) amongst genes in each category. We calculated genomic overlap as the average number of genes in a particular category (e.g., adsorption genes) encoded by strains containing a query gene (e.g., also adsorption genes, or all non-extended core genes). We then visualized genomic overlap distributions in violin plots for each query (e.g., all phage resistance genes) and each subject (e.g., number of overlapping non-extended core genes). We hypothesized that phage resistance genes with shared functions are encoded together on the genome. We expect that such genes with shared functions are co-encoded more often than any gene in the genome at random. All phage resistance genes (non-extended core, all phage resistance, biosynthesis) or biosynthesis genes (non-extended core, all phage resistance, adsorption, biosynthesis, assembly) were most often significantly different (non-parametric Wilcoxon test, $p<0.05$) from non-extended core genes (Figure 5) in subject genomic overlap distribution (subject listed here in parentheses). In each phage resistance gene category, when query and subject were the same (e.g., adsorption genes against themselves), query genomic overlap was most significantly different ($p < 0.0001$) with non-extended core query overlap, and increased on average in all cases (Figure 5). This supported our hypothesis that genes with shared functions were encoded together, as adsorption, biosynthesis, and assembly genes were encoded together significantly more often than similar non-extended core genes at random.

Superinfection immunity correlated with empirically determined phage resistance and accessory genome content

We next asked whether phage resistance gene presence could explain two relevant phenotypes - empirical phage resistance, as we measured in a previous phage host range GWAS paper (166), and accessory genome content in the set of completely assembled *S. aureus* genomes. While no phage resistance gene categories correlated with measured virulent (p002y and pyo) phage resistance, all phage resistance genes, biosynthesis genes, and superinfection immunity genes, a subset of biosynthesis genes responsible for lysogens repressing superinfecting phages, correlated positively with temperate (p11 and p0040) phage resistance (Figure 6). This result confirmed the hypothesis that prophages confer temperate phage resistance through superinfection immunity. We observed similar patterns for accessory genome content (Figure 7), but here superinfection immunity had the largest correlation with accessory genome content ($R^2 = 0.26$) amongst features examined, while all phage resistance genes had a weak positive correlation ($R^2 = 0.0068$). This finding suggested prophages correlated with accessory genome, as strains subject to extensive transduction would potentially also be expected to carry more prophages given the necessary exposure to temperate phage. However, superinfection immunity did not correlate with non-extended core adsorption, non-superinfection immunity biosynthesis, or assembly gene counts (Figure 8), however, suggesting none of these factors prevented acquisition of superinfection immunity genes through lysogeny.

Relationship between non-extended core phage resistance genes and horizontal gene transfer, antibiotic resistance, and virulence

We then asked whether phage resistance acted as a barrier to antibiotic resistance and virulence gene acquisition by 1) correlating counts of phage resistance genes with non-extended core antibiotic resistance and virulence gene counts amongst complete *S. aureus* genomes, 2) calculating phylogenetic overlap between phage resistance genes of each category and non-extended core genes or subsets of antibiotic resistance or virulence genes, and 3) calculating

genomic overlap between the same sets as 2). As far as the first objective, all phage resistance, biosynthesis, and superinfection immunity genes positively correlated with non-extended core antibiotic resistance genes, but all categories positively correlated with non-extended core virulence genes (Figure 9). This suggested superinfection immunity was not the only phage resistance factor correlated with these subsets of the accessory genome, especially virulence genes. Assembly gene-encoding SaPIs are known to encode virulence factors such as toxic shock syndrome toxin, while VFDB virulence genes include capsule genes that we considered adsorption factors.

Regarding phylogenetic overlap, assembly genes were co-encoded with antibiotic resistance and virulence genes in more strains than non-extended core genes on average, while virulence genes were co-encoded with adsorption genes in far fewer strains than non-extended core genes on average ($p<0.05$, Wilcoxon test, Figure 10). In all other cases (e.g., non-extended core genes, all phage resistance genes as subjects), query adsorption genes had significantly less and query assembly genes had significantly more phylogenetic overlap with subject genes than query non-extended core genes in general ($p<0.05$, Wilcoxon test, Figure 10). Regarding genomic overlap, more antibiotic resistance genes were encoded for adsorption gene-encoding strains than non-extended core gene-encoding strains, while more virulence genes were encoded for all phage resistance and biosynthesis gene-encoding strains than non-extended core gene-encoding strains ($p<0.05$, Wilcoxon test, Figure 5). These phylogenetic and genomic overlap results thus indicate non-extended core antibiotic resistance genes tended to be encoded together with adsorption genes, while non-extended core virulence genes tended to be encoded together with biosynthesis genes. They also indicate assembly genes shared the most common strains with antibiotic resistance and virulence genes, suggesting common networks of horizontal gene transfer. These results thus corroborate the conclusions from correlating non-extended core phage resistance with corresponding antibiotic resistance or virulence gene matches on a per-strain level in the much smaller complete *S. aureus* genome set.

Consequences of type I restriction-modification specificity gene (*hsdS*) allelism on horizontal gene transfer, antibiotic resistance, and virulence

In addition to non-extended core genes, we also analyzed type I restriction-modification specificity gene (*hsdS*) alleles for their relationships with HGT, given their correlation with clonal complex and role in restricting transduction between clonal complexes based on restriction specificity (112). We calculated phylogenetic and genomic overlaps between detected NRD *hsdS* alleles (81 total) and all non-extended core genes or respective antibiotic resistance and virulence gene subsets. In all cases, *hsdS* allele phylogenetic overlap was significantly less ($p<0.05$, Wilcoxon test) than that of non-extended core genes generally (Figure 10). We noticed a bimodal distribution of overlap indicative of alleles associated with more or less non-extended core genes, which possibly suggests alleles associated with strains being open to or restricted from transduction, respectively. However, only for subject non-extended core, adsorption, biosynthesis, and antibiotic resistance genes were *hsdS* allele genomic overlap significantly different ($p<0.05$, Wilcoxon test) than that of non-extended core genes generally (Figure 5). Taken together, these results indicated, as expected, a role for restriction specificity in affecting horizontal gene transfer, especially given that the differences were most pronounced when viewed on the phylogenetic (strains shared between *hsdS* alleles and accessory genes of interest) rather than genomic (number of genes encoded in strains with an *hsdS* allele) level, consistent with the hypothesis that phylogeny (CC)-associated *hsdS* alleles restrict horizontal gene transfer.

Discussion

In this work, I curated a list of known phage resistance-related genes in *S. aureus*, and used them to ask questions about the nature of phage resistance in the species and its relationship to measured phage resistance or horizontal gene transfer. The genes were assigned to three stages of phage infection (adsorption, biosynthesis, and assembly) at which *S. aureus* is known to have developed resistance to phages. Adsorption genes examined include wall teichoic

acid (WTA) and capsule biosynthesis genes. Biosynthesis genes examined include characterized restriction-modification, abortive infection, and CRISPR systems, as well as newly discovered systems such as cyclic oligonucleotide-based anti-phage signaling systems (CBASS), defense island systems associated with restriction modification (DISARM), and Lamassu systems, amongst others. This study is the first to find recently discovered Septu and retron phage defense systems (345) in *S. aureus*. Assembly genes examined include all three main mechanisms known in SaPIs - capsid remodeling, packaging interference, and helper phage late gene repression. Care was taken to include both genes identified and characterized in *S. aureus* (e.g., WTA biosynthesis genes) as well as genes characterized in other species but having *S. aureus* homologs (e.g., Lamassu phage defense system).

From the analyses performed on these genes, we draw four major conclusions. We found that core and extended core phage resistance genes do not significantly differ in diversity, functionality, and selection from corresponding core or extended core genes in general, refuting a possible arms race hypothesis for the evolution of host and phage. Phages specific to strains in species such as *E. coli* often exist in an arms race dynamic with the outer-surface of the bacterium, in which phage and host coevolve to outcompete each other (346, 347). Host receptors and phage receptor-binding proteins undergo diversifying selection during this coevolution. We do not see this pattern in *S. aureus*, however, given that core and extended core phage resistance genes have similar high functionality (low delta-bit score), low diversity, and negative selection to core and extended core genes. We attribute this result to fitness costs of losing capsule and wall teichoic acid (WTA) that these core genes are responsible for synthesizing. Wall teichoic acid is critical for cell division (264, 348, 349), methicillin resistance (193), nasal colonization (261), and antimicrobial peptide resistance (49), amongst other roles. Alternatively, our strain set may not capture transient strains resistant through core gene mutation (especially unstable mutations as in phase variation), or non-core genes may instead be undergoing arms race dynamics through frequent gain or loss.

Looking at non-extended core genes instead, we also found that all curated non-extended core phage resistance genes have measurable phylogenetic signal, but by far assembly genes had the least phylogenetic signal. We found these genes all had phylogenetic signal because their presence/absence pattern was more consistent with the phylogeny than expected by random simulations. Our finding that assembly genes were the least consistent amongst the different categories matches what we expect based on their horizontal transfer by SaPIs. We expect more vertically transmitted genes (e.g., ST395 WTA biosynthesis) to be more consistent with the phylogeny and more horizontally transmitted genes (e.g., prophage and SaPI genes) to be less consistent with the phylogeny. Indeed, as SaPIs carry assembly resistance genes during their transduction, which in turn is dependent on helper phage mobilization (79), we expect their presence and absence to be quite inconsistent with the core genome phylogeny.

Regarding gene clustering and modularity, we found that non-extended core phage resistance genes with similar functions are encoded together (e.g., adsorption) by measuring genomic overlap. If a strain encoded a non-extended core adsorption gene, for example, it was more likely to encode a gene of the same category than a control strain, encoding any non-extended core gene. In other words, encoding a single non-extended core gene increased the chances of encoding another. We note that we did not ask explicitly about genes being encoded together as operons, though recent studies have discovered new phage defense systems based on proximity (but not operonic linkage) to existing systems (256). These results suggest that phage resistance genes tend to be clustered together in strains within the species. This is likely due to functional interactions between different genes (i.e., one gene depends on another for carrying out a common function) preventing loss of either and transfer of such genes together on common genetic elements (e.g., SaPIs carrying assembly interference genes). In the future we will examine phage defense islands that contain multiple defense genes and systems in the same genomic region to further address this clustering.

We also showed that superinfection immunity amongst non-extended core phage resistance genes was the sole class to correlate with empirical (temperate) phage resistance, accessory genome content, and accessory antibiotic resistance/virulence gene content. We note that superinfection immunity by indicating prophages could correlate with lateral transduction in the pac-type phages, which is worth further study population-wide. Having certain prophages (i.e., pac-type) in the genome could enhance transduction of adjacent genomic DNA several orders of magnitude up to 300 kb downstream, according to recent studies (37). This suggests that prophages may not only positively correlate with accessory genome content due to increased transduction but also due to prophage-enhanced lateral transduction.

Our most paradoxical finding was that phage resistance gene presence for the most part did not predict phage resistance nor HGT, especially resistance to the virulent phages. We note that we expect the presence of a phage resistance gene to correlate with resistance to all phages, when some genes may not affect all phage types and phages are known to have defenses against specific barriers (e.g., anti-restriction and anti-CRISPR systems) (227, 350). It is possible that existing studies characterizing phage resistance genes in *S. aureus* do not reflect growth conditions (e.g., rich medium vs. blood or skin) or selection pressures in natural environments where *S. aureus* is present, making laboratory-defined genes poor predictors of phage resistance in the environment. Correspondingly, it may be that cryptic loci not characterized yet with respect to phage resistance and host range are the most important in predicting each phenotype, such as those found in our recent host range GWAS. It is possible that complete resistance to virulent phage infection is extremely rare and instead subtle metabolic or surface protein changes are instead responsible for most variation in host range. On the other hand, we simply may need to adjust our approach to ignore genes with neutral (type 5 or 8 capsule) or negative (WTA) resistance effects or focus on rare, mobile phage defense systems. Recent studies indicate that rapid turnover of phage defense systems carried on mobile genetic elements explains arms race dynamics between phages and hosts, with the gradient of systems carried creating a gradient of

phage sensitivities, at least in the marine *Vibrio* strains examined (351). We may also want to further examine genes for which genetic diversity has effects on phage resistance, such as type I restriction specificity. We also examined phylogenetic overlap between *hsdS* alleles and various non-extended core genes of interest, for example, finding significantly lower overlap than for non-extended core genes in general, and a bimodal distribution with a subset of alleles having lower and higher overlap. These findings support the long-held belief that *hsdS* diversity is a major factor shaping horizontal gene transfer patterns in the species through restriction specificity.

Nonetheless, future efforts from the bioinformatic survey should most likely focus on mobile defense systems and prophage diversity. Combinations of these defense systems may impact host range and horizontal gene transfer phenotypes more strongly than most gene classes considered here. The spectrum from cryptic to complete prophages as well as the types should be classified considering their roles in transduction, resistance, and large-scale levels of horizontal gene transfer that this work brings further to light. We also must continue to conduct laboratory phage resistance studies in the species that focus on what is likely to occur in the natural population. Such work would include further boosting the phage host range GWAS with a large number of diverse strains, to classify as many examples of phage resistance in the species as possible. Resistance evolution studies should also be done in physiologically relevant conditions (e.g., consequences of within-host evolution, biofilm development, and phage challenge at MOIs common in human *S. aureus* niches). All of this future work, together with what has been described here, will enhance our understanding of how phages shape *S. aureus* evolution.

Materials and Methods

Curating list of *S. aureus* phage resistance genes

Genes selected for bioinformatic analysis were manually curated from the existing literature on laboratory-confirmed phage resistance determinants in *Staphylococcus aureus*.

Genes selected were either 1) reported to have causative effects on phage resistance in a *S. aureus* strain through molecular genetic studies, 2) identified through laboratory selection for phage resistance, 3) reported to have causative effects on phage resistance in other species but had homologs in *S. aureus*, or 4) were reported to have causative effects on phage resistance in other species but did not have known *S. aureus* homologs, which were only included when the previously criteria were not met for that phage resistance gene class. An example of the final criterion was the inclusion of the *Lactococcus lactis* superinfection exclusion (*sie*) uptake resistance gene (68), which is not reported in *S. aureus* nor contains a known *S. aureus* homolog. Supplementary Table S1 lists these genes, coordinates and accessions of the associated sequences, the resistance class and subclass, and literature supporting its inclusion in the list, nucleotide sequence selection, and resistance designation.

Determining phage resistance gene conservation in the *S. aureus* species

Nucleotide sequences of the curated phage resistance gene list were matched against our library of 42,949 *S. aureus* genomes (Staphopia) (260) using a BLAST (352) search with default parameters except maximum target sequences of 10,000,000 and maximum high scoring pairs of 1. BLAST output was filtered for unique matches between each gene and each strain. We then counted the number of unique matches per gene to determine the number of strains containing each gene. Gene conservation in the species was then compared between phage resistance gene categories (adsorption, biosynthesis, and assembly) using violin plots and assessed for statistically significant differences between groups with non-parametric Wilcoxon tests.

Determining core/extended core phage resistance gene diversity, functionality, and selection

The pangenome of the Staphopia non-redundant (NRD) set (380 strains representing each sequence type) (260) was constructed using PIRATE (353) run with default parameters.

Core genes were those unique PIRATE gene clusters only present in all NRD genomes, while we defined extended core genes to be those unique PIRATE gene clusters present in 80% of NRD genomes (304) or more. We focused gene diversity, functionality, and selection studies on these core and extended core phage resistance genes, comparing these sets against corresponding total core and extended core genes (excluding core/extended core phage resistance). We evaluated gene diversity both through the number of alleles per corresponding gene in the pangenome and the translated nucleotide sequence amino acid diversity (π) calculated from the corresponding gene's pangenome nucleotide alignment using modified scripts originally written by John Lees. We evaluated functionality using delta-bit score (354) and measured selection by calculating dN/dS for each gene with the package Hypothesis Testing using Phylogenies (HyPhy) using single-likelihood ancestor counting (SLAC). PIRATE output provided the number of alleles at the maximum cutoff (98%) for each gene, so it was directly parsed to get allele counts for each core or extended core gene. Amino acid diversity (π) was calculated from translated PIRATE gene cluster nucleotide alignments. dN/dS, on the other hand, was calculated using individual gene phylogenetic trees inferred from nucleotide alignments with IQ-TREE (291). For delta-bit score analysis, for each tested gene, we calculated average delta-bit scores for all strains encoding protein sequences that matching each HMMs.

Evaluating phylogenetic associations with non-extended core phage resistance genes

We evaluated phylogenetic associations with phage resistance genes by 1) determining homoplasy for each non-extended core phage resistance gene and 2) correlating clonal complex (CC) with non-extended core phage resistance gene count. Non-extended core genes were defined as those present in less than 80% of the genomes and filtered for redundancy (unique PIRATE gene cluster matches to query phage resistance genes were selected for further analysis). Homoplasy measurement through consistency index (CI) calculation was conducted with HomoplasyFinder (355) given gene presence/absence input and the NRD set phylogenetic

tree. We constructed the NRD set maximum-likelihood phylogenetic tree with IQ-TREE using the gubbins-recombination corrected PIRATE core genome alignment. Consistency index, or the consistency between a character amongst strains and that expected on the tree, was calculated from (number of possible character - 1)/(number of changes necessary to explain the character pattern on the tree) or 1/(number of necessary changes) because only two outcomes were possible for each gene (presence or absence). To assess whether consistency indices were statistically significant, we calculated average CI values and their standard deviations for the original data plus 999 permutations of the gene presence/absence data on the tree. We transformed CI to the number of necessary changes (1/CI) for better data visualization and further comparisons. We compared the number of changes for non-extended-core genes to non-extended-core phage resistance genes of each category with violin plots and assessed significance with non-parametric Wilcoxon tests. We also plotted the number of necessary changes to explain the character pattern against the number of strains encoding the gene to determine a relationship between these factors and to compare the relationships for the actual and permuted data. In addition, we also compared non-extended core phage resistance gene count by clonal complex in the complete genome set (535 genomes). We compared counts for each phage resistance category along with accessory genome content using a boxplot and assessed statistical significance of overall differences with an analysis of variance (ANOVA) statistical test.

Non-extended-core resistance gene correlation analyses with empirical phage resistance and accessory genome content

We used two more sets of *S. aureus* genomes to examine the relationship between non-extended-core phage resistance genes and 1) empirically measured phage resistance phenotypes, 2) accessory genome content, and 3) more specifically than 2), non-extended-core antibiotic resistance or virulence gene content. We used genomes of previously resistance-

phenotyped strains from our *S. aureus* genome-wide host range study (166) and genomes of all completely assembled *S. aureus* genomes (260) to address the first and second objectives, respectively (the third we addressed with both sets). Antibiotic resistance genes searched were previously identified in *S. aureus* genomes (356), whereas virulence genes searched were the Virulence Factor Database (VFDB) set (357). Just as for the NRD set, non-extended-core phage resistance, antibiotic resistance, and virulence genes were defined as those unique PIRATE gene cluster matches present in 80% of the respective genome set (complete or GWAS). We then used BLAST to match these three sets of non-extended-core genes in the complete and GWAS genome sets. BLAST matches were filtered by query coverage relative to subject, only keeping those with 60% or higher. Matches were further filtered for uniqueness. Filtered numbers of matches were then plotted against empirically measured phage resistance (GWAS set), accessory genome content, or non-extended core antibiotic resistance or virulence gene matches. Linear regressions were performed on each distribution to assess correlations between these factors.

Calculating non-extended-core phage resistance gene phylogenetic and genomic overlap

In addition to assessing correlations between non-extended core phage resistance gene counts and accessory genome content on a per-strain level, we also evaluated strain and gene level concordance between non-extended-core phage resistance genes. We did this, as for the correlation analysis, to determine associations among classes of phage resistance genes or between such classes and accessory antibiotic resistance or virulence genes, but with phylogenetic or genomic corrections on a much larger dataset. Unlike the previous analysis, we instead searched phage resistance, antibiotic resistance, and virulence (VFDB) genes against our full Staphopia database with BLAST. BLAST matches were filtered by query coverage relative to subject, only keeping those with 60% or higher. Matches were further filtered for uniqueness and converted to a list of strain-gene pairs. Strain-gene pairs were compared to lists of perfect

strain-gene pairs (all strains matching to each gene) to convert the list into a presence absence matrix. Only Staphopia non-extended-core genes (phage resistance genes from the complete list present in less than 80% of Staphopia genomes) were considered for further analysis. Genomic overlap per gene was calculated as the average number of genes in a category (e.g., adsorption) encoded by strains encoding the gene of interest, while phylogenetic overlap per gene was calculated as the total number of genes in a category encoded by all strains encoding the gene of interest divided by the number of genes in that category. Genomic overlap measures how many genes of a certain type are co-encoded with a gene of interest on average, while phylogenetic overlap measures how many strains co-encode a gene and all those of a certain type on average. Genomic and phylogenetic overlap were compared between groups with violin plots and significant differences assessed with non-parametric Wilcoxon tests. For the *hsdS* gene, this analysis was repeated with the 81 alleles detected in the NRD set PIRATE pangenome. Genomic and phylogenetic overlap distributions were compared with violin plots and significant differences assessed using non-parametric Wilcoxon tests.

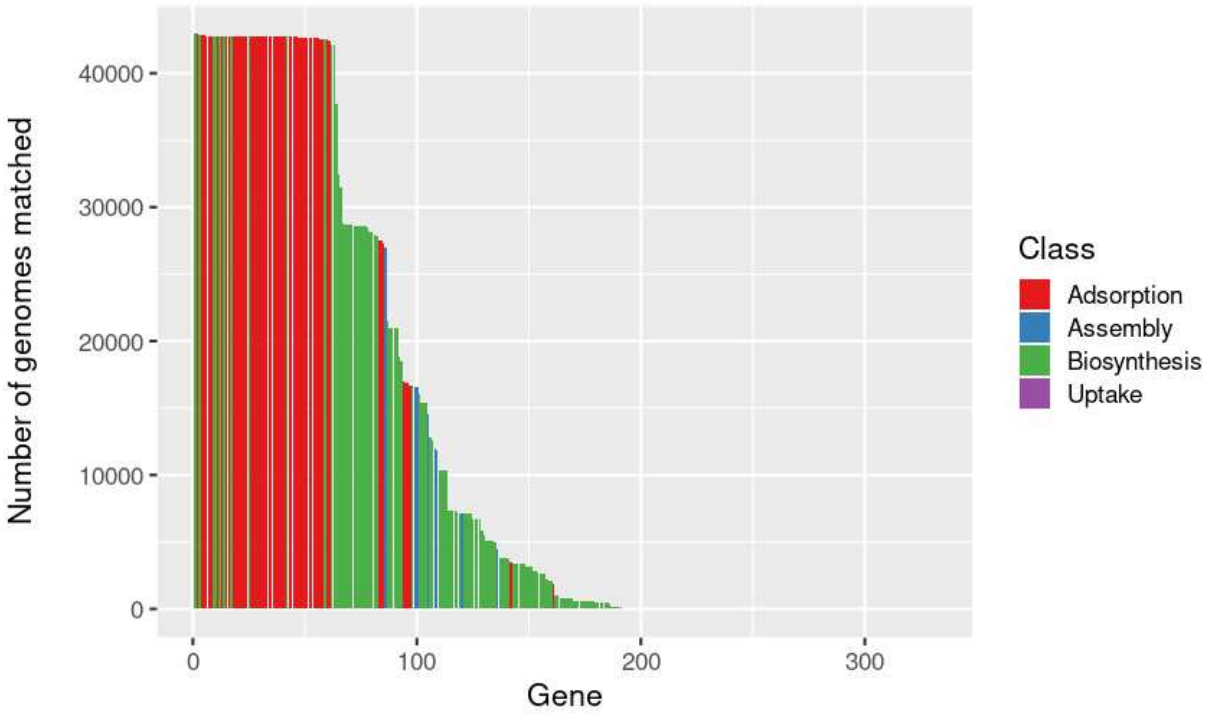
Acknowledgements

We thank members of the lab for constructive criticism. Timothy D. Read was supported by NIH R21 AI 138079-02, while Abraham G. Moller was supported by the NSF GRFP.

Figures

Figure 1: Conservation of examined phage resistance genes in the species based on a search of 40,000+ *S. aureus* genomes. We used BLAST to search for our set of 331 curated phage resistance genes in the Staphopia database. A) Conservation of each gene (y-axis) ranked from most to least conserved on the x-axis. Genes are colored by class - adsorption in red, assembly in blue, and biosynthesis in green. B) Distributions of conservation for each considered category visualized as violin plots. Groups were tested for statistically significant differences with the non-parametric Wilcoxon test (ns, not significant; *, *P* = 0.01 to 0.05; **, *P* = 0.001 to 0.01; ***, *P* = 0.0001 to 0.001; ****, *P* = 0 to 0.0001).

A



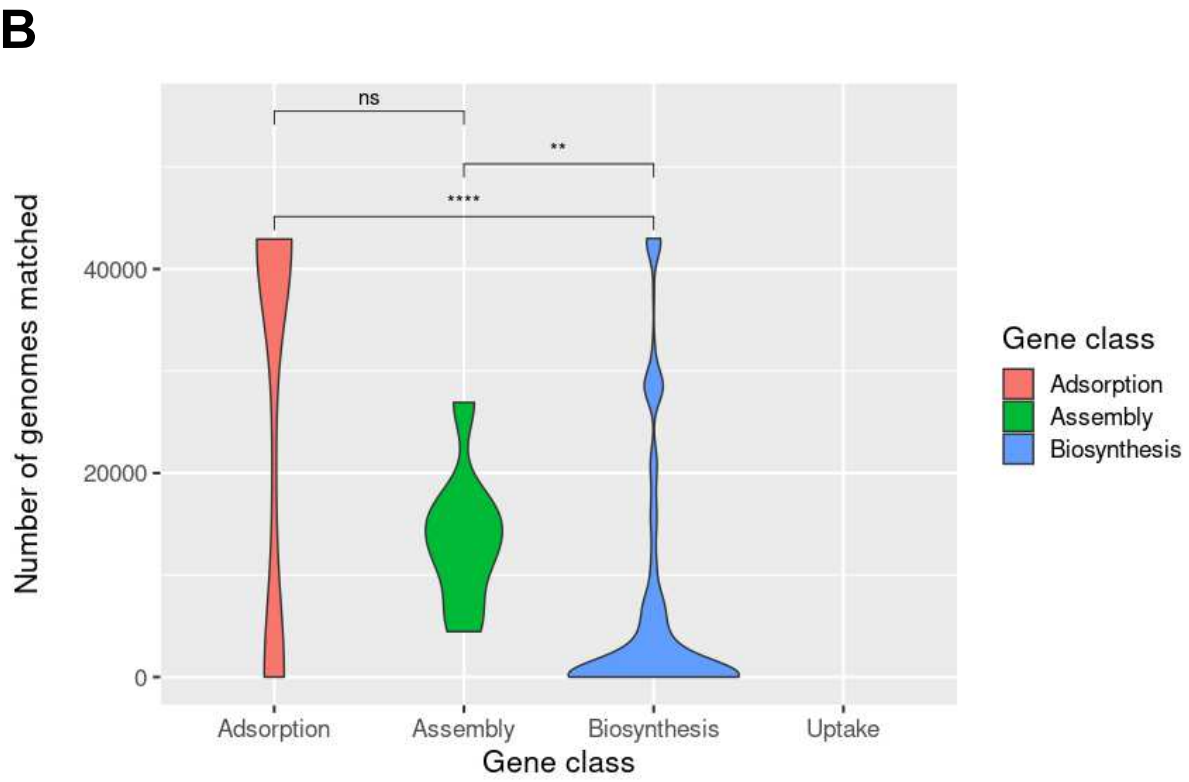
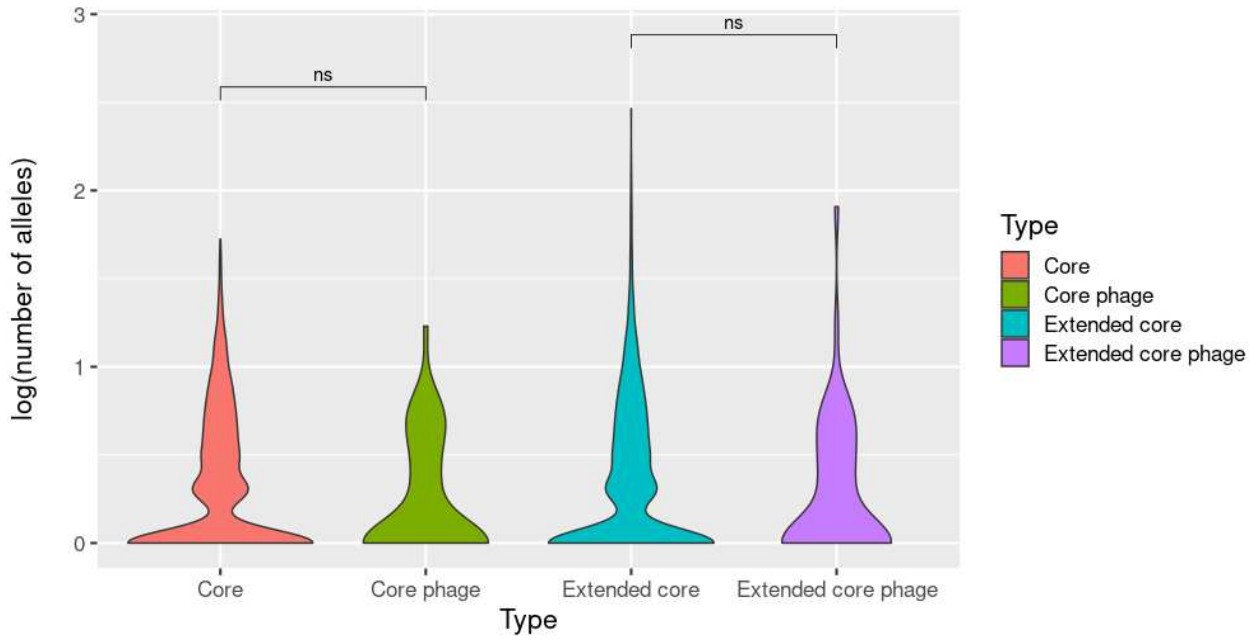


Figure 2: Core and extended core phage resistance genes do not differ from core and extended core genes overall in terms of diversity, functionality, or selection. Core genes were those found in all 380 genomes of the Staphopia non-redundant (NRD) set while extended core genes were those found in 80% or more of the NRD set. Diversity (A, B) was measured as the number of alleles or translated nucleotide diversity (π). Functionality (C) was measured by delta-bit score, or calculated difference from reference profile Hidden Markov models (HMMs), which score changes from conserved amino acids higher than those from non-conserved amino acids. Selection (D) was measured through the dN/dS metric, which was calculated from gene phylogenetic trees using single-likelihood ancestor counting (SLAC). Group distributions were visualized as violin plots and differences were tested for significance using the non-parametric Wilcoxon test (ns, not significant; *, *P* = 0.01 to 0.05; **, *P* = 0.001 to 0.01; ***, *P* = 0.0001 to 0.001; ****, *P* = 0 to 0.0001).

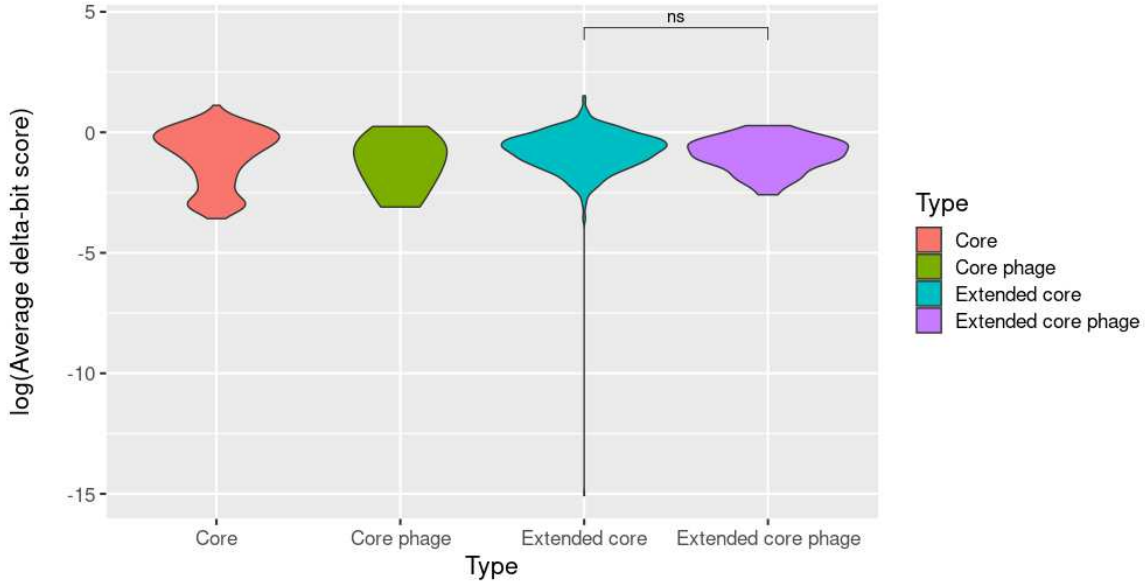
A


log(number of alleles)
3
2
1
0
ns
ns
Core
Core phage
Extended core
Extended core phage
Type
Type
Core
Core phage
Extended core
Extended core phage

B


log(Nucleotide diversity (pi))
0
-1
-2
-3
-4
ns
ns
Core
Core phage
Extended core
Extended core phage
Type
Type
Core
Core phage
Extended core
Extended core phage

C


ns
5
0
-5
-10
-15
log(Average delta-bit score)
Core
Core phage
Extended core
Extended core phage
Type
Type
Core
Core phage
Extended core
Extended core phage

D

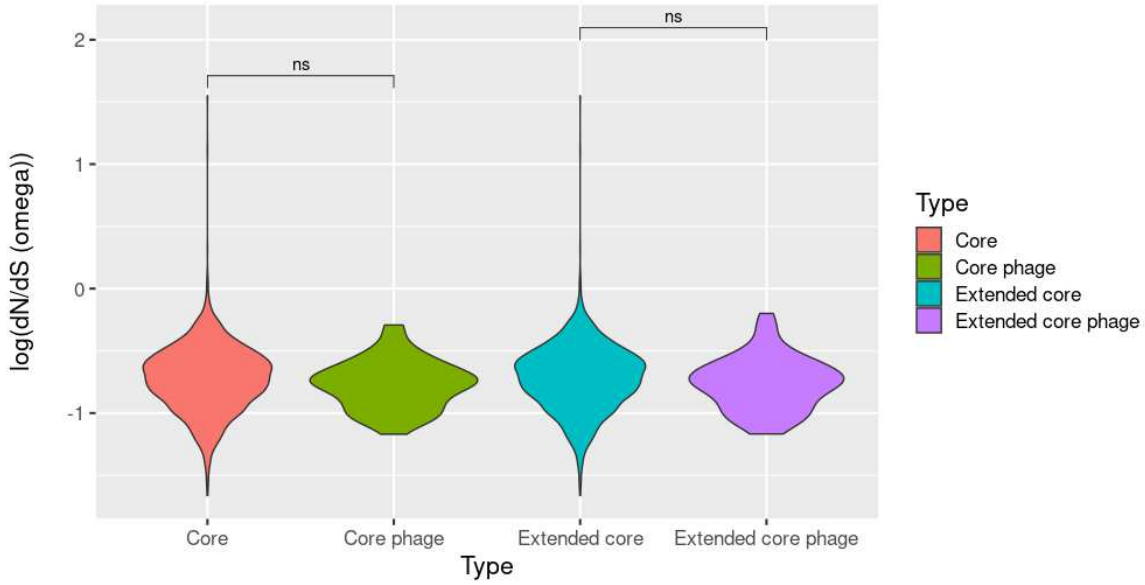
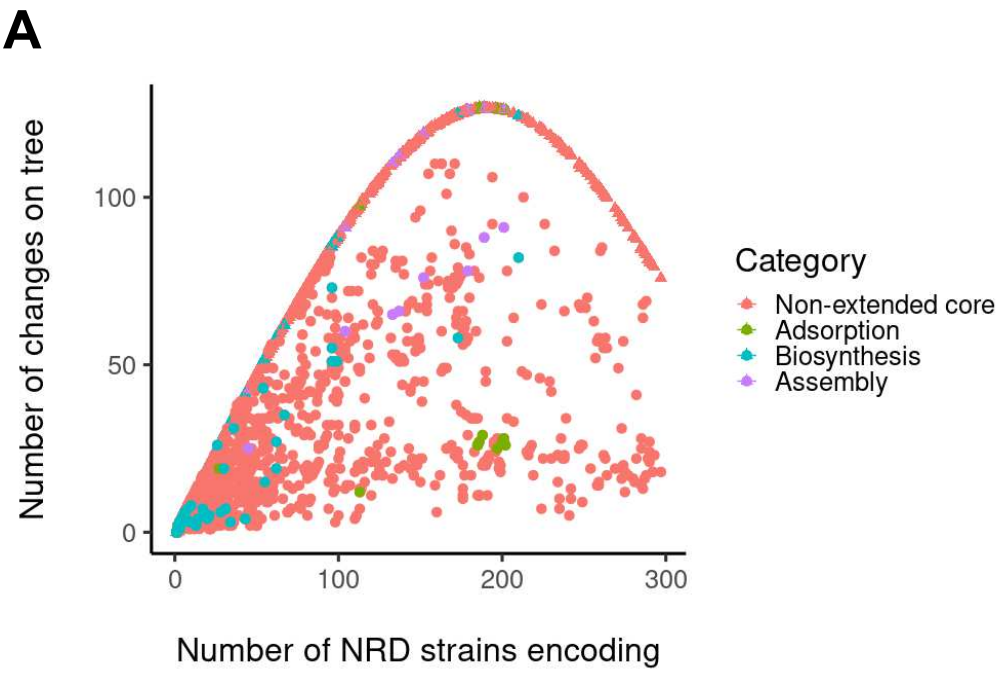

ns
ns
2
1
0
-1
log(dN/dS (omega))
Core
Core phage
Extended core
Extended core phage
Type
Type
Core
Core phage
Extended core
Extended core phage

Figure 3: All non-extended core phage resistance genes have phylogenetic signal, but assembly genes have the least phylogenetic signal amongst all. We calculated phylogenetic signal as consistency index (CI) between non-extended core (genes present in 80% or less than the NRD set) A) The relationship between number of changes necessary to make gene presence/absence patterns consistent with the phylogeny (y-axis) and the number of strains encoding each gene (x-axis). Average change numbers after 999 permutations of gene presence/absence on the tree are shown as triangles with error bars (1 standard error above and below the mean), while change numbers for actual gene presence/absence are shown as circles. Non-extended core genes are colored in salmon red, while adsorption, biosynthesis, and assembly genes are colored in olive green, turquoise, and purple, respectively. B) Distributions of gene presence/absence changes necessary to make them consistent with the phylogeny for each gene category (non-extended core, adsorption, biosynthesis, and assembly) visualized as violin plots. Group differences were tested for significance using the non-parametric Wilcoxon test (ns, not significant; *, *P* = 0.01 to 0.05; **, *P* = 0.001 to 0.01; ***, *P* = 0.0001 to 0.001; ****, *P* = 0 to 0.0001).



B

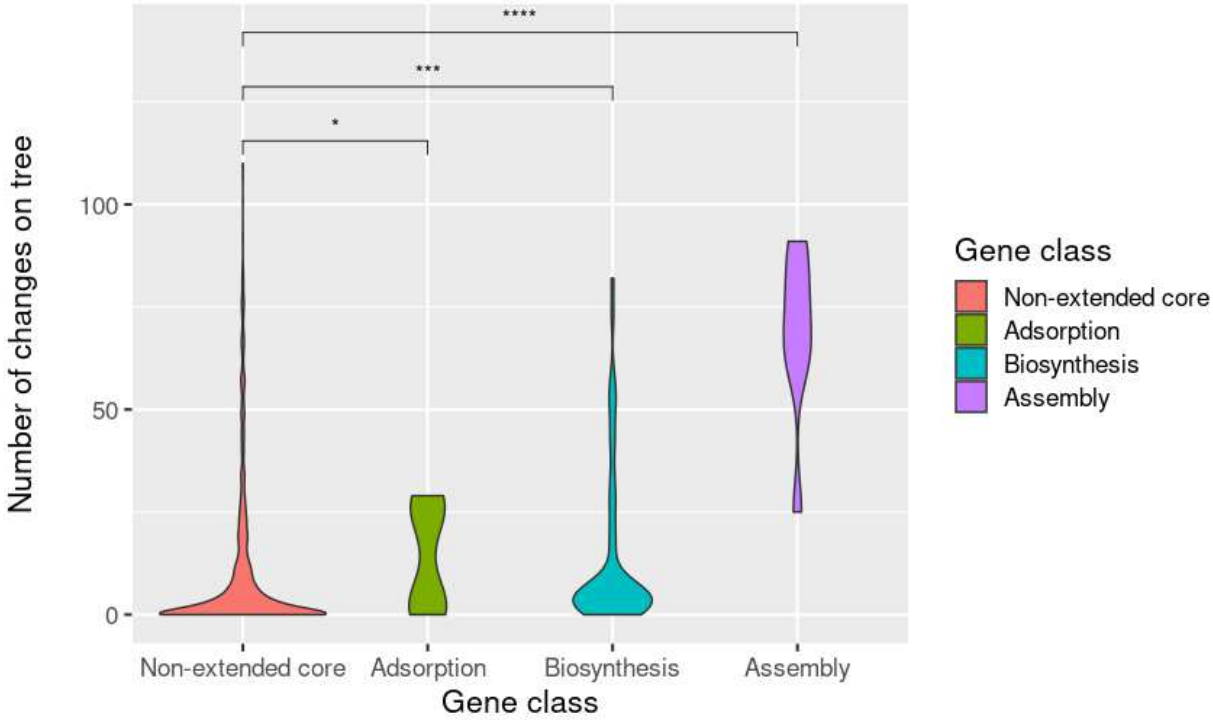

****
***
*
100
50
0
Number of changes on tree
Non-extended core
Adsorption
Biosynthesis
Assembly
Gene class
Gene class
Non-extended core
Adsorption
Biosynthesis
Assembly

Figure 4: Relationship between clonal complex (CC) and accessory genome and non-extended core phage resistance gene content. We used BLAST to search for our set of non-extended core phage resistance genes and accessory genes in the set of 535 complete *S. aureus* genomes in Staphopia. We then visualized the distributions of genes for each CC as boxplots. Strains without a defined CC are listed as NA. Analysis of variance (ANOVA) assessed significant overall differences, with significance values posted on each facet.

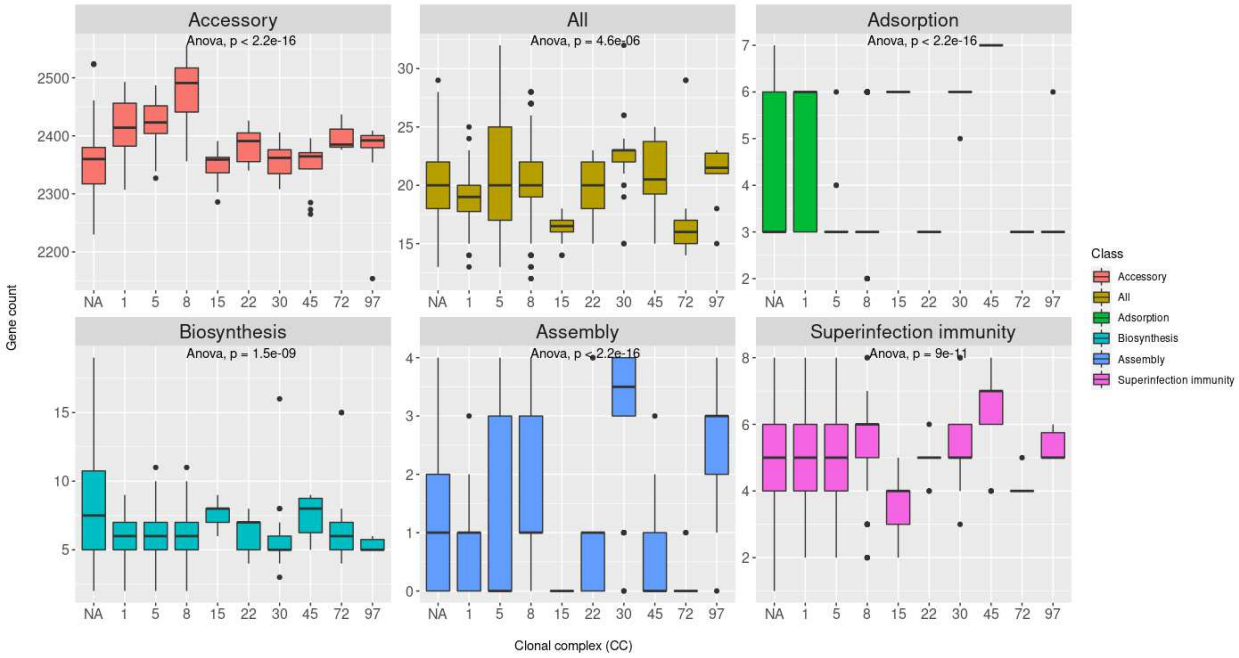


Figure 5: Non-extended core phage resistance genes are encoded together based on calculating genomic overlaps from a search of 40,000+ *S. aureus* genomes. We used BLAST to search for our non-extended core (found in less than 80% of genomes) phage resistance genes in the Staphopia database. We then calculated the genomic overlap as the average number of genes in a subject category for strains encoding a query gene. Genomic overlap (y-axis) distributions for different query sets (x-axis) and subject sets (facets) are visualized as violin plots. Group differences relative to all non-extended core genes were tested for significance with non-parametric Wilcoxon tests (ns, not significant; *, P = 0.01 to 0.05; **, P = 0.001 to 0.01; ***, P = 0.0001 to 0.001; ****, P = 0 to 0.0001).

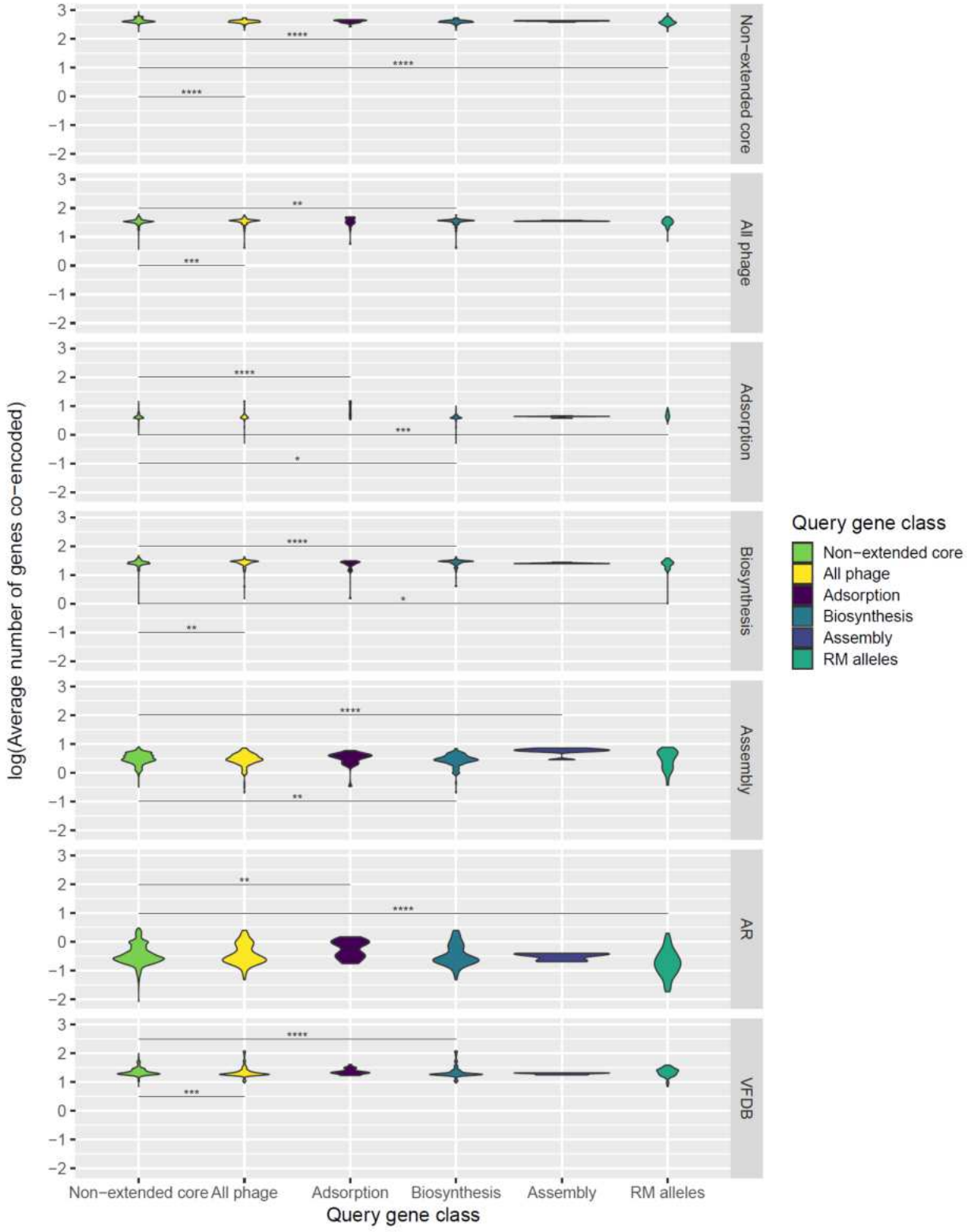

log(Average number of genes co−encoded)
Non−extended core
All phage
Adsorption
Biosynthesis
Assembly
AR
VFDB
Query gene class
Non−extended core
All phage
Adsorption
Biosynthesis
Assembly
RM alleles
Query gene class
Non−extended core
All phage
Adsorption
Biosynthesis
Assembly
RM alleles

Figure 6: Superinfection immunity but neither adsorption nor assembly genes correlates with empirical temperate phage resistance. We used BLAST to search for our set of non-extended core phage resistance genes in the set of 263 *S. aureus* genomes from our recent phage host range study. We then plotted the number of matches to all non-extended core phage resistance, adsorption, biosynthesis, assembly, or superinfection immunity genes (x-axis) against previously measured phage resistance phenotypes (OD_{600} or turbidity after co-culture; y-axis) and calculated correlations (R^2) between each. We present results for two temperate phages (p11 and p0040) and two virulent phages (p002y and pyo).

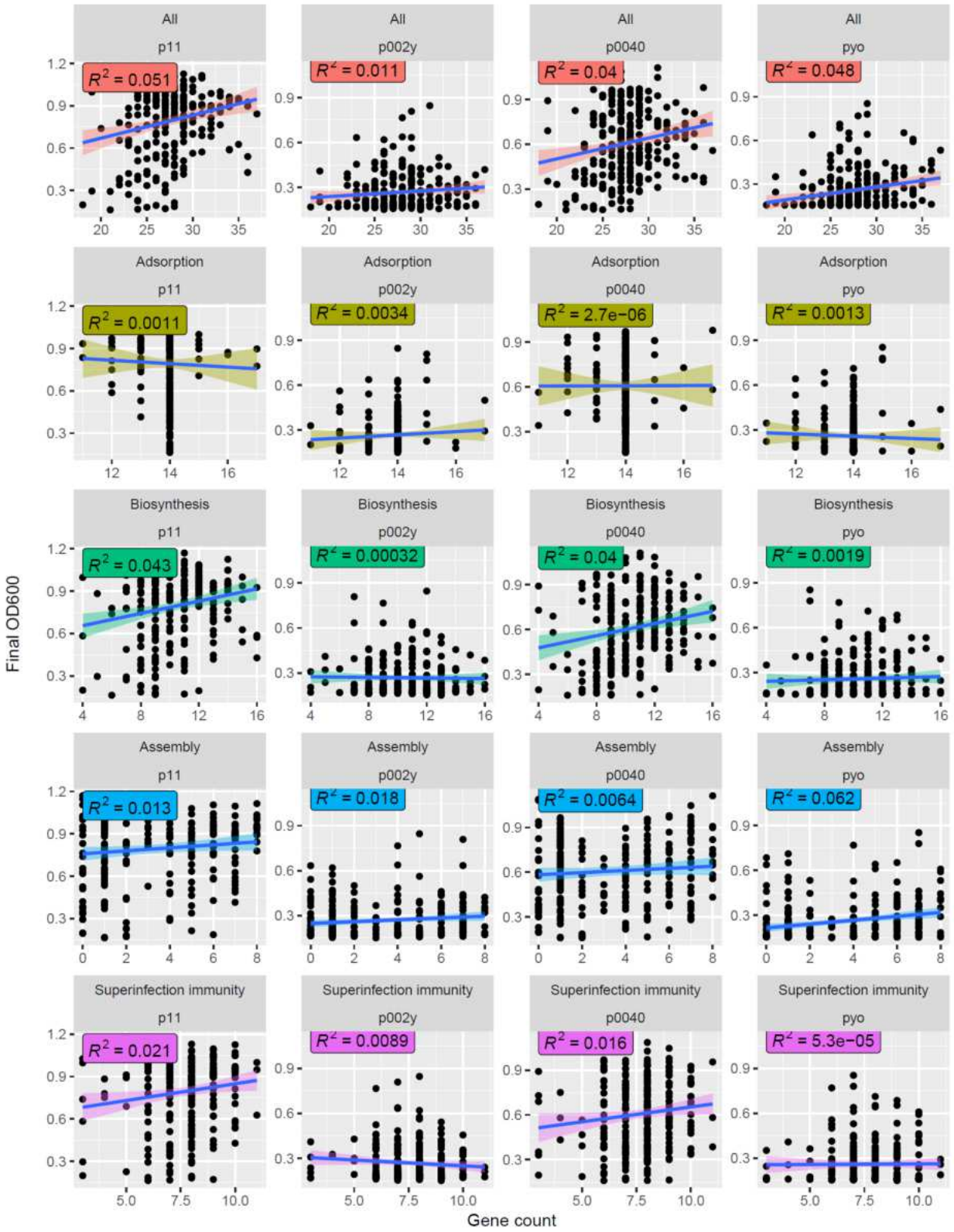

All
p11
R^2 = 0.051
All
p002y
R^2 = 0.011
All
p0040
R^2 = 0.04
All
pyo
R^2 = 0.048
Adsorption
p11
R^2 = 0.0011
Adsorption
p002y
R^2 = 0.0034
Adsorption
p0040
R^2 = 2.7e−06
Adsorption
pyo
R^2 = 0.0013
Biosynthesis
p11
R^2 = 0.043
Biosynthesis
p002y
R^2 = 0.00032
Biosynthesis
p0040
R^2 = 0.04
Biosynthesis
pyo
R^2 = 0.0019
Assembly
p11
R^2 = 0.013
Assembly
p002y
R^2 = 0.018
Assembly
p0040
R^2 = 0.0064
Assembly
pyo
R^2 = 0.062
Superinfection immunity
p11
R^2 = 0.021
Superinfection immunity
p002y
R^2 = 0.0089
Superinfection immunity
p0040
R^2 = 0.016
Superinfection immunity
pyo
R^2 = 5.3e−05
Final OD600
Gene count

Figure 7: Superinfection immunity but neither adsorption nor assembly genes correlates with accessory genome content. We used BLAST to search for our set of non-extended core phage resistance genes in the set of 535 complete *S. aureus* genomes in Staphopia. We then plotted the number of matches to all non-extended core phage resistance, adsorption, biosynthesis, assembly, or superinfection immunity genes (x-axis) against accessory genome content (y-axis) and calculated correlations (R^2) between each.

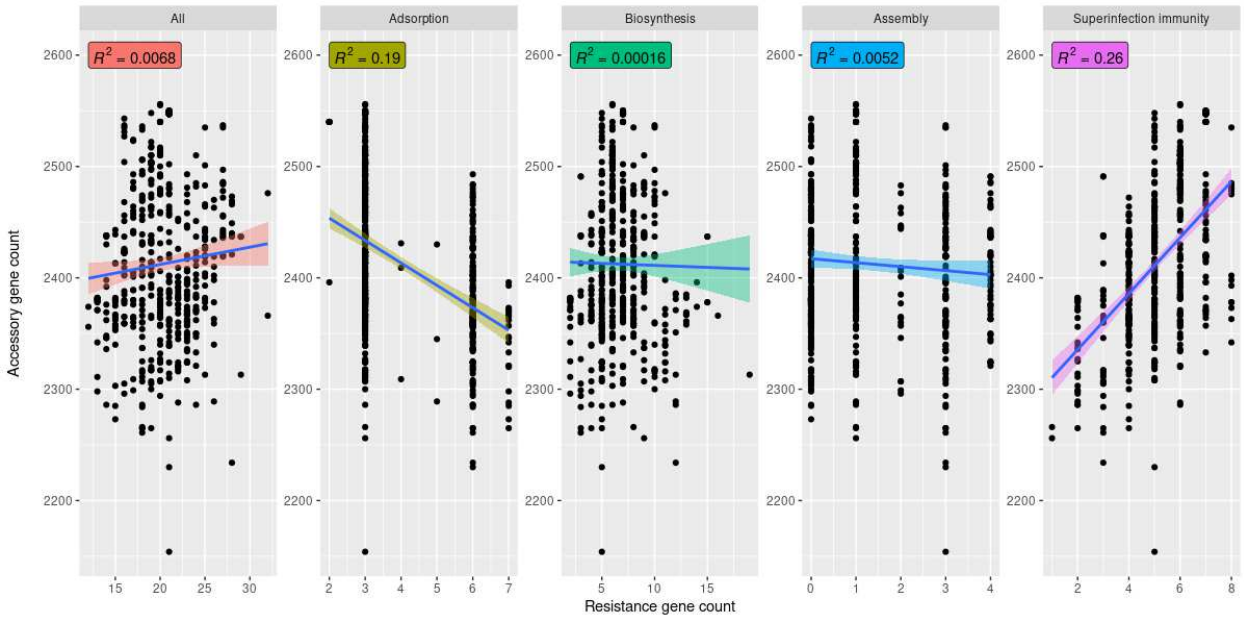


Figure 8: Superinfection immunity does not correlate with non-extended core adsorption, non-superinfection immunity biosynthesis, or assembly gene counts. We used BLAST to search for our set of non-extended core phage resistance genes in the set of 535 complete *S. aureus* genomes in Staphopia. We then plotted the number of matches to all non-extended core adsorption, non-superinfection immunity biosynthesis, or assembly genes (y-axis) against matches to non-extended core superinfection immunity genes (x-axis) and calculated correlations (R^2) between each.

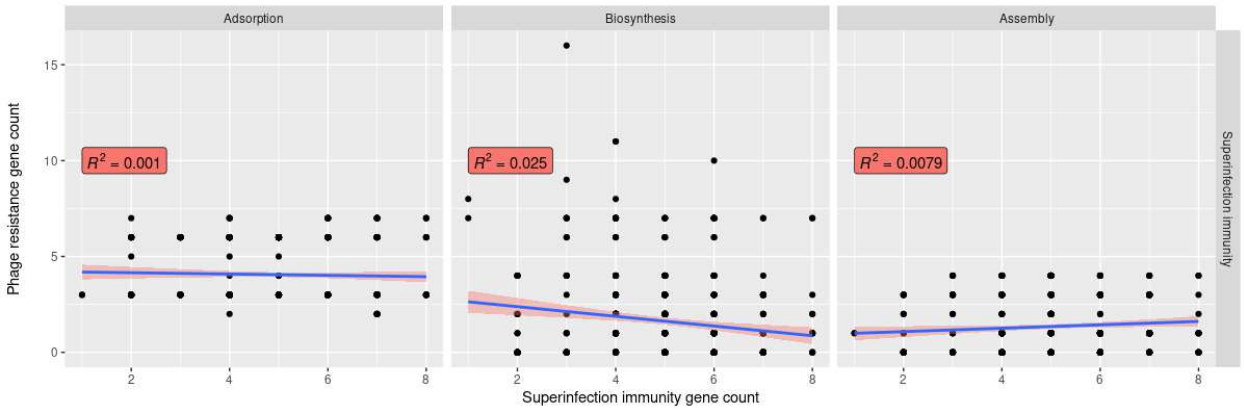


Figure 9: Correlations between non-extended core antibiotic resistance or virulence genes and non-extended core phage resistance genes. We used BLAST to search for our set of non-extended core phage resistance, antibiotic resistance, and virulence genes in the set of 535 complete *S. aureus* genomes in Staphopia. We then plotted the number of matches to all non-extended core phage resistance, adsorption, biosynthesis, assembly, or superinfection immunity genes (x-axis) against matches (y-axis) to non-extended core antibiotic resistance and virulence genes (facets) and calculated correlations (R^2) between each.

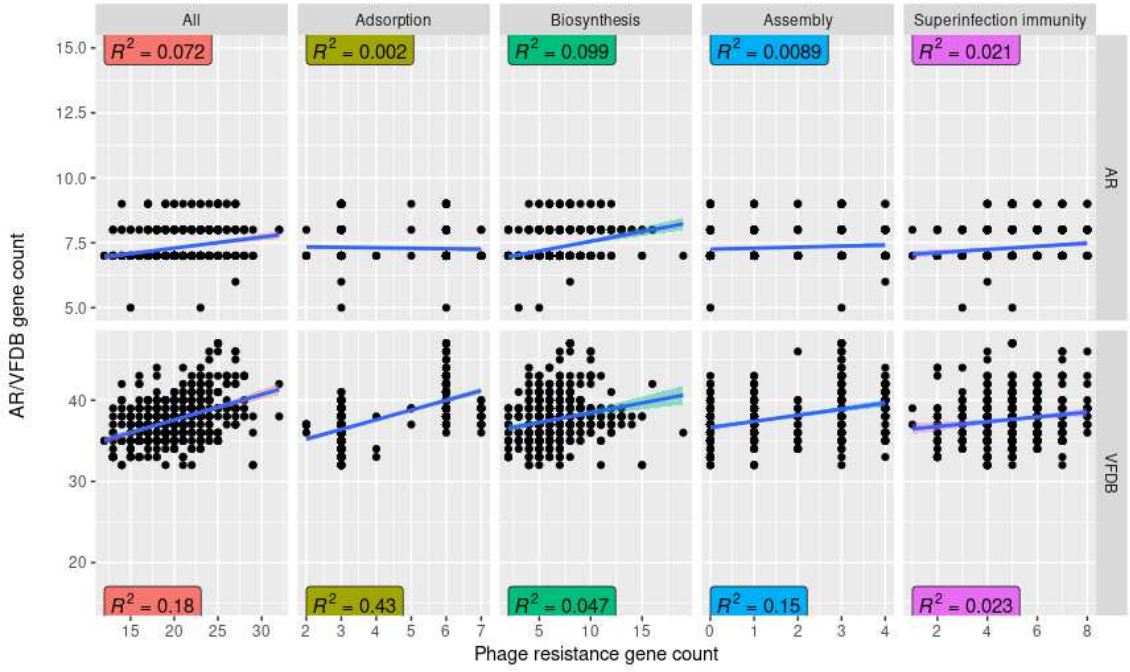
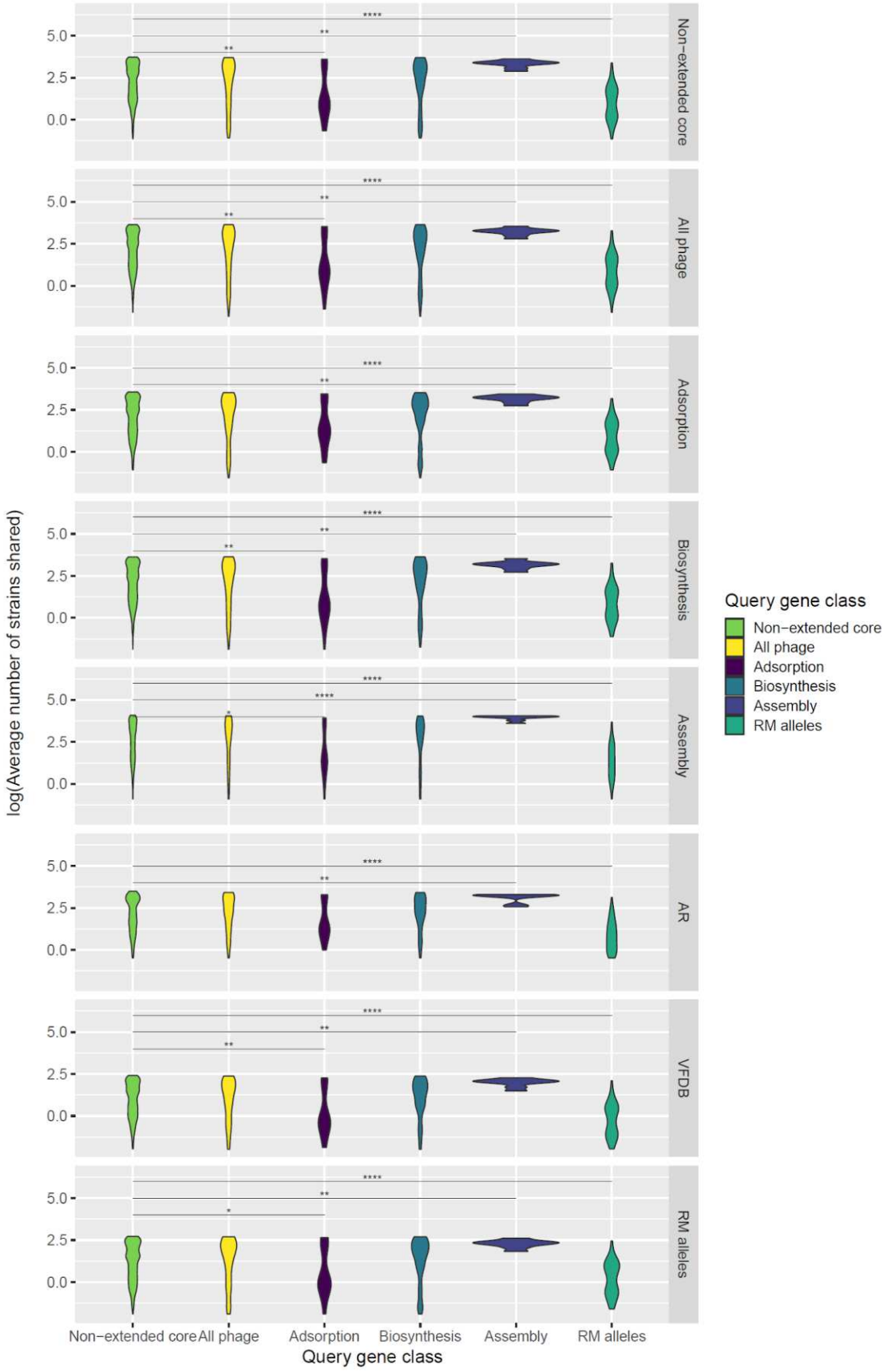


Figure 10 (following page): Phylogenetic overlap between classes of non-extended core genes determined from a search of 40,000+ *S. aureus* genomes. We used BLAST to search for our non-extended core (found in less than 80% of genomes) phage resistance genes in the Staphopia database. We then calculated the phylogenetic overlap as the average number of encoding strains shared between a query gene and those of a subject category. Phylogenetic overlap (y-axis) distributions for different query sets (x-axis) and subject sets (facets) are visualized as violin plots. Group differences relative to all non-extended core genes were tested for significance with non-parametric Wilcoxon tests (ns, not significant; *, P = 0.01 to 0.05; **, P = 0.001 to 0.01; ***, P = 0.0001 to 0.001; ****, P = 0 to 0.0001).


log(Average number of strains shared)
Non−extended core
All phage
Adsorption
Biosynthesis
Assembly
AR
VFDB
RM alleles
5.0
2.5
0.0
****
**
*
Non−extended core
All phage
Adsorption
Biosynthesis
Assembly
RM alleles
Query gene class
Query gene class
Non−extended core
All phage
Adsorption
Biosynthesis
Assembly
RM alleles

Tables

Table 1: List of curated phage resistance genes with strains, genome coordinates, accessions, classes (e.g., adsorption), and subclasses (e.g., receptor).

Gene name	Taxon	Coordinates	Accession	Class	Subclass
tagO	S. aureus N315	800380-801435	NC_002745.2	Adsorption	Receptor
tagA/tarA	S. aureus N315	685072-685836	NC_002745.2	Adsorption	Receptor
tagB/tarB	S. aureus N315	687949-689052	NC_002745.2	Adsorption	Receptor
tagF/tarF	S. aureus N315	296796-297965	NC_002745.2	Adsorption	Receptor
tarI1	S. aureus N315	298241-298957	NC_002745.2	Adsorption	Receptor
tarJ1	S. aureus N315	298950-299975	NC_002745.2	Adsorption	Receptor
tarL	S. aureus N315	299997-301685	NC_002745.2	Adsorption	Receptor
tarI2 (tarI')	S. aureus N315	292918-293634	NC_002745.2	Adsorption	Receptor
tarJ2 (tarJ')	S. aureus N315	293627-294652	NC_002745.2	Adsorption	Receptor
tarK	S. aureus N315	294674-296368	NC_002745.2	Adsorption	Receptor
tagG	S. aureus N315	687038-687850	NC_002745.2	Adsorption	Receptor
tagH_1	S. aureus N315	685897-686691	NC_002745.2	Adsorption	Receptor
tagH_2	S. aureus N315	1935669-1937186	NC_002745.2	Adsorption	Receptor
tagD/tarD	S. aureus N315	690173-690571	NC_002745.2	Adsorption	Receptor
msrR (LCP)	S. aureus N315	1366983-1367966	NC_002745.2	Adsorption	Receptor
SA0908 (LCP)	S. aureus N315	1031423-1032640	NC_002745.2	Adsorption	Receptor
tagU (lytR; LCP) (SA2103)	S. aureus N315	2365947-2366870	NC_002745.2	Adsorption	Receptor
tarM (not in N315) (SA21178_0837)	S. aureus 21178	29451-30932	AGRN01000099.1	Adsorption	Receptor
tarS (SA0248) (SA_RS01460)	S. aureus N315	301718-303439	NC_002745.2	Adsorption	Receptor
tarP	S. aureus N315	2047184-2048167	NC_002745.2	Adsorption	Receptor
tagV	S. aureus PS187	24672-27177	ARPA01000002.1	Adsorption	Receptor
tagN	S. aureus PS187	23709-24656	ARPA01000002.1	Adsorption	Receptor
tagD1/tarD1	S. aureus PS187	33581-34753	ARPA01000002.1	Adsorption	Receptor

tagF/tarF	S. aureus PS187	29221-29541	ARPA01000002.1	Adsorption	Receptor
oatA	S. aureus N315	2645038-2646849	NC_002745.2	Adsorption	Receptor
cap1 locus	S. aureus M	1-46392	U10927	Adsorption	Capsule/Occlusion
cap2 locus	S. aureus Smith diffuse	35263-42369	NZ_UHCU01000003.1	Adsorption	Capsule/Occlusion
cap5 locus	S. aureus Reynolds and Newman	1-18131	U81973	Adsorption	Capsule/Occlusion
cap8 locus	S. aureus Becker	1-17539	U73374	Adsorption	Capsule/Occlusion
cap1O	S. aureus M	26015-26878	U10927	Adsorption	Capsule/Occlusion
cap1N	S. aureus M	25458-26015	U10927	Adsorption	Capsule/Occlusion
cap1M	S. aureus M	24316-25458	U10927	Adsorption	Capsule/Occlusion
cap1L	S. aureus M	23014-24288	U10927	Adsorption	Capsule/Occlusion
cap1K	S. aureus M	21652-23001	U10927	Adsorption	Capsule/Occlusion
cap1J	S. aureus M	20484-21659	U10927	Adsorption	Capsule/Occlusion
cap1I	S. aureus M	19460-20464	U10927	Adsorption	Capsule/Occlusion
cap1H	S. aureus M	18374-19441	U10927	Adsorption	Capsule/Occlusion
cap1G	S. aureus M	17856-18374	U10927	Adsorption	Capsule/Occlusion
cap1F	S. aureus M	16655-17845	U10927	Adsorption	Capsule/Occlusion
cap1E	S. aureus M	15331-16653	U10927	Adsorption	Capsule/Occlusion
cap1D	S. aureus M	13504-15303	U10927	Adsorption	Capsule/Occlusion
cap1C	S. aureus M	12712-13479	U10927	Adsorption	Capsule/Occlusion
cap1B	S. aureus M	12023-12709	U10927	Adsorption	Capsule/Occlusion
cap1A	S. aureus M	11343-12008	U10927	Adsorption	Capsule/Occlusion
capA/cap8A_1	S. aureus Smith diffuse	35263-35925	NZ_UHCU01000003.1	Adsorption	Capsule/Occlusion
capB/ywqD	S. aureus Smith diffuse	35937-36644	NZ_UHCU01000003.1	Adsorption	Capsule/Occlusion
capC_1/ywqE	S. aureus Smith diffuse	36637-37404	NZ_UHCU01000003.1	Adsorption	Capsule/Occlusion
capD_1	S. aureus Smith diffuse	37437-39260	NZ_UHCU01000003.1	Adsorption	Capsule/Occlusion
capL_1	S. aureus Smith diffuse	39253-39747	NZ_UHCU01000003.1	Adsorption	Capsule/Occlusion

capL_2	S. aureus Smith diffuse	39780-40541	NZ_UHCU01000003.1	Adsorption	Capsule/Occlusion
arnB	S. aureus Smith diffuse	40560-41789	NZ_UHCU01000003.1	Adsorption	Capsule/Occlusion
wcaJ_1	S. aureus Smith diffuse	41773-42369	NZ_UHCU01000003.1	Adsorption	Capsule/Occlusion
cap5A	S. aureus Reynolds and Newman	376-1044	U81973	Adsorption	Capsule/Occlusion
cap5B	S. aureus Reynolds and Newman	1060-1746	U81973	Adsorption	Capsule/Occlusion
cap5C	S. aureus Reynolds and Newman	1749-2513	U81973	Adsorption	Capsule/Occlusion
cap5D	S. aureus Reynolds and Newman	2533-4356	U81973	Adsorption	Capsule/Occlusion
cap5E	S. aureus Reynolds and Newman	4346-5374	U81973	Adsorption	Capsule/Occlusion
cap5F	S. aureus Reynolds and Newman	5381-6496	U81973	Adsorption	Capsule/Occlusion
cap5G	S. aureus Reynolds and Newman	6500-7624	U81973	Adsorption	Capsule/Occlusion
cap5H	S. aureus Reynolds and Newman	7627-8253	U81973	Adsorption	Capsule/Occlusion
cap5I	S. aureus Reynolds and Newman	8258-9367	U81973	Adsorption	Capsule/Occlusion
cap5J	S. aureus Reynolds and Newman	9381-10547	U81973	Adsorption	Capsule/Occlusion
cap5K	S. aureus Reynolds and Newman	10540-11745	U81973	Adsorption	Capsule/Occlusion
cap5L	S. aureus Reynolds and Newman	11746-12951	U81973	Adsorption	Capsule/Occlusion

cap5M	S. aureus Reynolds and Newman	12962-13519	U81973	Adsorption	Capsule/Occlusion
cap5N	S. aureus Reynolds and Newman	13519-14406	U81973	Adsorption	Capsule/Occlusion
cap5O	S. aureus Reynolds and Newman	14460-15722	U81973	Adsorption	Capsule/Occlusion
cap5P	S. aureus Reynolds and Newman	15769-16944	U81973	Adsorption	Capsule/Occlusion
cap8A	S. aureus Becker	446-1114	U73374	Adsorption	Capsule/Occlusion
cap8B	S. aureus Becker	1130-1816	U73374	Adsorption	Capsule/Occlusion
cap8C	S. aureus Becker	1819-2583	U73374	Adsorption	Capsule/Occlusion
cap8D	S. aureus Becker	2603-4426	U73374	Adsorption	Capsule/Occlusion
cap8E	S. aureus Becker	4416-5444	U73374	Adsorption	Capsule/Occlusion
cap8F	S. aureus Becker	5451-6566	U73374	Adsorption	Capsule/Occlusion
cap8G	S. aureus Becker	6570-7694	U73374	Adsorption	Capsule/Occlusion
cap8H	S. aureus Becker	7697-8776	U73374	Adsorption	Capsule/Occlusion
cap8I	S. aureus Becker	8769-10163	U73374	Adsorption	Capsule/Occlusion
cap8J	S. aureus Becker	10160-10717	U73374	Adsorption	Capsule/Occlusion
cap8K	S. aureus Becker	10726-11964	U73374	Adsorption	Capsule/Occlusion
cap8L	S. aureus Becker	11998-13203	U73374	Adsorption	Capsule/Occlusion
cap8M	S. aureus Becker	13214-13771	U73374	Adsorption	Capsule/Occlusion
cap8N	S. aureus Becker	13771-14658	U73374	Adsorption	Capsule/Occlusion

cap8O	S. aureus Becker	14713-15975	U73374	Adsorption	Capsule/Occlusion
cap8P	S. aureus Becker	16022-17197	U73374	Adsorption	Capsule/Occlusion
rapZ	S. aureus N315	824108-825019	NC_002745.2	Adsorption	Capsule/Occlusion
yozB	S. aureus N315	1097121-1097582	NC_002745.2	Adsorption	Capsule/Occlusion
murA2	S. aureus N315	2174362-2175621	NC_002745.2	Adsorption	Cell wall
sie2009	Lactococcal phage Tuc2009	1339-1860	AF109874.2	Uptake	Superinfection exclusion
hsdR	S. aureus N315	222427-225216	NC_002745.2	Biosynthesis	Restriction-modification
hsdM1	S. aureus N315	451000-452556	NC_002745.2	Biosynthesis	Restriction-modification
hsdM2	S. aureus N315	1859152-1860891	NC_002745.2	Biosynthesis	Restriction-modification
hsdS1	S. aureus N315	452549-453760	NC_002745.2	Biosynthesis	Restriction-modification
hsdS2	S. aureus N315	1857930-1859159	NC_002745.2	Biosynthesis	Restriction-modification
Sau3A (R) (Sau3AI)	S. aureus 3AI	448-1917	M32470.1	Biosynthesis	Restriction-modification
Sau3A (M) (M.Sau3AI)	S. aureus 3AI	2017-3255	M32470.1	Biosynthesis	Restriction-modification
Sau42I	S. aureus 42CR3-L	49-1968	X94423.1	Biosynthesis	Restriction-modification
S.Sau42I	S. aureus 42CR3-L	1949-2962	X94423.1	Biosynthesis	Restriction-modification
Sau96I	S. aureus PS96	1524-2309	X53096.1	Biosynthesis	Restriction-modification
M.Sau96I	S. aureus PS96	168-1460	X53096.1	Biosynthesis	Restriction-modification
SauUSI	S. aureus USA300 FPR3757	2614823-2617684	NC_007793.1	Biosynthesis	Restriction-modification
ardA	S. aureus Mu50	449748-450248	NC_002758.2	Biosynthesis	Restriction-modification
cas1	S. aureus 08BA02176	56694-57599	NC_018608.1	Biosynthesis	CRISPR/Cas
cas2	S. aureus 08BA02176	57599-57904	NC_018608.1	Biosynthesis	CRISPR/Cas
cas10/csm1	S. aureus 08BA02176	57918-60191	NC_018608.1	Biosynthesis	CRISPR/Cas
csm2	S. aureus 08BA02176	60194-60619	NC_018608.1	Biosynthesis	CRISPR/Cas

csm3	S. aureus 08BA02176	60621-61265	NC_018608.1	Biosynthesis	CRISPR/Cas
csm4	S. aureus 08BA02176	61276-62184	NC_018608.1	Biosynthesis	CRISPR/Cas
csm5	S. aureus 08BA02176	62187-63209	NC_018608.1	Biosynthesis	CRISPR/Cas
csm6	S. aureus 08BA02176	63209-64477	NC_018608.1	Biosynthesis	CRISPR/Cas
cas6	S. aureus 08BA02176	64474-65208	NC_018608.1	Biosynthesis	CRISPR/Cas
cas3	S. aureus 08BA02176	1638819-1640165	NC_018608.1	Biosynthesis	CRISPR/Cas
cas9	S. aureus M06/0171	45349-48510	HE980450.1	Biosynthesis	CRISPR/Cas
cI	Staphylococcal phage 23MRA	13413-14126	NC_028775	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage 2638A	24962-25582	NC_007051	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage 3MRA	36304-36627	NC_028917	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage B166	2374-3012	NC_028859	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage B236	2029-2748	NC_028915	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage DW2	3460-3669	NC_024391	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage JS01	13078-13794	NC_021773	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage LH1	29577-30191	JX174275	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage P954	3387-4103	NC_013195	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage phi2958PVL	4317-4931	NC_011344	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage phi5967PVL	3183-3890	NC_019921	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage phi7401PVL	5124-5447	NC_020199	Biosynthesis	Superinfection immunity

rep1	Staphylococcal phage phiJB	3010-3324	NC_028669	Biosynthesis	Superinfection immunity
rep2	Staphylococcal phage phiJB	3476-3712	NC_028669	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage phiNM3	3400-4113	NC_008617	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage phiSa119	3385-4101	NC_025460	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage PVL	26588-27358	NC_002321	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage ROSA	29422-29745	NC_007058	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage SA12	40279-40998	NC_021801	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage SA13	3749-4081	NC_021863	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage SMSAP5	17208-17537	NC_019513	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage SP5	2340-3071	JX274646	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage SP6	2339-3058	JX274647	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage StauST398-1	42375-43088	NC_021326	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage StauST398-2	3696-4409	NC_021323	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage StauST398-3	2700-3026	NC_021332	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage StauST398-4	3447-4160	NC_023499	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage StauST398-5	3092-3415	NC_023500	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage TEM123	6364-7002	NC_017968	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage tp310-1	3111-3881	NC_009761	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage tp310-2	4440-5126	NC_009762	Biosynthesis	Superinfection immunity

cI	Staphylococcal phage tp310-3	4005-4229	NC_009763	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage vB_SauS_phi2	28221-28550	NC_028862	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage YMC/09/04/R1988	27073-27843	NC_022758	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage phiPV83	2923-3633	NC_002486	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage phiPVL108	3434-4150	NC_008689	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage 187	25564-25896	NC_007047	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage 29	29764-30393	NC_007061	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage 37	30480-31106	NC_007055	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage 3a	25656-26342	NC_007053	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage 42e	29939-30658	NC_007052	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage 47	27163-27777	NC_007054	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage 52a	30937-31215	NC_007062	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage 53	30606-31325	NC_007049	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage 55	29298-29627	NC_007060	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage 69	29406-30125	NC_007048	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage 77	24441-25208	NC_005356	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage 80α	3185-3901	NC_009526	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage 85	31388-31714	NC_007050	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage 88	29382-29696	NC_007063	Biosynthesis	Superinfection immunity

cI	Staphylococcal phage 92	28556-29275	NC_007064	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage 96	29505-29831	NC_007057	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage EW	29684-30013	NC_007056	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage IPLA35	3579-3833	NC_011612	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage IPLA88	2196-2834	NC_011614	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage phiETA	2342-3058	NC_003288	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage phiETA2	3733-4065	NC_008798	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage phiETA3	3732-4064	NC_008799	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage phiMR11	2438-2752	NC_010147	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage phiMR25	3275-3601	NC_010808	Biosynthesis	Superinfection immunity
cI*	Staphylococcal phage phiSLT	3641-3955	NC_002661	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage SAP26	27269-27907	NC_014460	Biosynthesis	Superinfection immunity
cI	Staphylococcal phage X2	30223-30825	NC_007065	Biosynthesis	Superinfection immunity
AbiA	Lactococcus lactis subsp. lactis ME2	6967-8853	U17233	Biosynthesis	Abortive infection
AbiB	Lactococcus lactis subsp. lactis strain IL416	903-1655	M77708	Biosynthesis	Abortive infection
AbiC	Lactococcus lactis subsp. lactis ME2	25-1056	M95956	Biosynthesis	Abortive infection
AbiD	Lactococcus lactis subsp. lactis KR5	765-1865	U10992	Biosynthesis	Abortive infection
AbiD1	Lactococcus lactis subsp. lactis IL964	1230-2285	L35176	Biosynthesis	Abortive infection

AbiEi	Lactococcus lactis subsp. lactis DRC3	198-1061	U36837	Biosynthesis	Abortive infection
AbiEii	Lactococcus lactis subsp. lactis DRC3	1058-1954	U36837	Biosynthesis	Abortive infection
AbiF	Lactococcus lactis subsp. lactis DRC3	5296-6324	U36837	Biosynthesis	Abortive infection
AbiGi	Lactococcus lactis subsp. cremoris UC653	4376-5125	U60336	Biosynthesis	Abortive infection
AbiGii	Lactococcus lactis subsp. cremoris UC653	5128-6321	U60336	Biosynthesis	Abortive infection
AbiH	Lactococcus lactis subsp. lactis S94	1095-2135	X97651	Biosynthesis	Abortive infection
AbiI	Lactococcus lactis subsp. lactis M138	1238-2236	U38973	Biosynthesis	Abortive infection
AbiJ	Lactococcus lactis subsp. lactis UK12922	174-1022	U41294	Biosynthesis	Abortive infection
AbiK	Lactococcus lactis subsp. lactis W-1	3297-5096	U35629	Biosynthesis	Abortive infection
AbiLi	Lactococcus lactis subsp. lactis UK19161	550-1926	U94520	Biosynthesis	Abortive infection
AbiLii	Lactococcus lactis subsp. lactis UK19161	1939-2832	U94520	Biosynthesis	Abortive infection
AbiN	Lactococcus lactis subsp. cremoris S114	4036-4572	Y11901	Biosynthesis	Abortive infection
AbiO	unknown	714-2333	I61427	Biosynthesis	Abortive infection
AbiP	Lactococcus lactis subsp. lactis IL1403	9239-9973	U90222	Biosynthesis	Abortive infection
AbiQ	Lactococcus lactis subsp. lactis W-37	3611-4162	AF001314	Biosynthesis	Abortive infection
AbiR	Lactococcus lactis subsp. lactis KR2	1-660	AF216814	Biosynthesis	Abortive infection

AbiR	Lactococcus lactis subsp. lactis KR2	1116-1610	AF216814	Biosynthesis	Abortive infection
AbiR	Lactococcus lactis subsp. lactis KR2	1820-2374	AF216814	Biosynthesis	Abortive infection
AbiR	Lactococcus lactis subsp. lactis KR2	2804-3565	AF216814	Biosynthesis	Abortive infection
AbiR	Lactococcus lactis subsp. lactis KR2	5277-6167	AF216814	Biosynthesis	Abortive infection
AbiR	Lactococcus lactis subsp. lactis KR2	7568-10120	AF216814	Biosynthesis	Abortive infection
AbiR	Lactococcus lactis subsp. lactis KR2	10122-11318	AF216814	Biosynthesis	Abortive infection
AbiR	Lactococcus lactis subsp. lactis KR2	11315-13777	AF216814	Biosynthesis	Abortive infection
AbiR	Lactococcus lactis subsp. lactis KR2	14002-14892	AF216814	Biosynthesis	Abortive infection
AbiR	Lactococcus lactis subsp. lactis KR2	15865-16152	AF216814	Biosynthesis	Abortive infection
AbiTi	Lactococcus lactis subsp. lactis WS1	1175-1558	AF483000	Biosynthesis	Abortive infection
AbiTii	Lactococcus lactis subsp. lactis WS1	1569-2210	AF483000	Biosynthesis	Abortive infection
AbiU	Lactococcus lactis subsp. lactis UK21371	1071-2840	AF188839	Biosynthesis	Abortive infection
gmk	S. aureus N315	1191032-1191655	NC_002745.2	Biosynthesis	Abortive infection
rpoA	S. aureus N315	2295111-2296055	NC_002745.2	Biosynthesis	Abortive infection
stk2	S. aureus N315	85585-87093	NC_002745.2	Biosynthesis	Abortive infection
dut	phage 80α	13894-14406	NC_009526	Assembly	Assembly interference
stl	N315 SaPln1	2070412-2071086	NC_002745.2	Assembly	Assembly interference
ppi	N315 SaPln1	2065678-2066058	NC_002745.2	Assembly	Assembly interference
cpmA	N315 SaPln1	2063659-2064237	NC_002745.2	Assembly	Assembly interference
cpmB	N315 SaPln1	2063423-2063641	NC_002745.2	Assembly	Assembly interference
ptiA	N315 SaPln1	2062501-2062842	NC_002745.2	Assembly	Assembly interference

ptiB	N315 SaPIn1	2064249-2064590	NC_002745.2	Assembly	Assembly interference
ptiM	N315 SaPIn1	2062845-2063372	NC_002745.2	Assembly	Assembly interference
drmD (type I)	Bacillus sp. 278922_107	333607-336789	NZ_KI911354.1	Biosynthesis	DISARM
drmMI (type I)	Bacillus sp. 278922_107	327382-330867	NZ_KI911354.1	Biosynthesis	DISARM
drmA (type I)	Bacillus sp. 278922_107	311695-315003	NZ_KI911354.1	Biosynthesis	DISARM
drmB (type I)	Bacillus sp. 278922_107	309903-311690	NZ_KI911354.1	Biosynthesis	DISARM
drmC (type I)	Bacillus sp. 278922_107	309131-309913	NZ_KI911354.1	Biosynthesis	DISARM
drmA (type II)	Bacillus sp. MSP5.4	69987-73346	NZ_JXAP01000007.1	Biosynthesis	DISARM
drmB (type II)	Bacillus sp. MSP5.4	68246-70015	NZ_JXAP01000007.1	Biosynthesis	DISARM
drmC (type II)	Bacillus sp. MSP5.4	67523-68245	NZ_JXAP01000007.1	Biosynthesis	DISARM
drmMII (type II)	Bacillus sp. MSP5.4	66174-67511	NZ_JXAP01000007.1	Biosynthesis	DISARM
drmE (type II)	Bacillus sp. MSP5.4	73351-75747	NZ_JXAP01000007.1	Biosynthesis	DISARM
brxA (type I)	Bacillus cereus H3081.97	102409-103011	NZ_ABDL02000007.1	Biosynthesis	BREX
brxB (type I)	Bacillus cereus H3081.97	101829-102407	NZ_ABDL02000007.1	Biosynthesis	BREX
brxC (type I)	Bacillus cereus H3081.97	98223-101801	NZ_ABDL02000007.1	Biosynthesis	BREX
pglX (type I)	Bacillus cereus H3081.97	94624-98163	NZ_ABDL02000007.1	Biosynthesis	BREX
pglZ (type I)	Bacillus cereus H3081.97	91980-94544	NZ_ABDL02000007.1	Biosynthesis	BREX
brxL (type I)	Bacillus cereus H3081.97	89926-91956	NZ_ABDL02000007.1	Biosynthesis	BREX
ThsA	Staphylococcus aureus S1	6531-7982	AUPS01000015.1	Biosynthesis	Thoeris
ThsB (i)	Staphylococcus aureus S1	9138-9746	AUPS01000015.1	Biosynthesis	Thoeris
ThsB (ii)	Staphylococcus aureus S1	8076-9110	AUPS01000015.1	Biosynthesis	Thoeris

HamA	Staphylococcus aureus BSAR58	393245-394159	CHEJ01000001.1	Biosynthesis	Hachiman
HamB	Staphylococcus aureus BSAR58	392400-393245	CHEJ01000001.1	Biosynthesis	Hachiman
SduA	Bacillus cereus B4264	955148-956290	CP001176.1	Biosynthesis	Shedu
GajA	Staphylococcus aureus subsp. aureus Mu50	457910-460099	BA000017.4	Biosynthesis	Gabija
GajB	Staphylococcus aureus subsp. aureus Mu50	456120-457913	BA000017.4	Biosynthesis	Gabija
PtuA	Bacillus cereus m1293	91939-93198	ACLS01000104.1	Biosynthesis	Septu
PtuB	Bacillus cereus m1293	91175-91753	ACLS01000104.1	Biosynthesis	Septu
LmuA	Staphylococcus aureus subsp. aureus TCH130	10338-11273	ACHD01000027.1	Biosynthesis	Lamassu
LmuB	Staphylococcus aureus subsp. aureus TCH130	9566-10345	ACHD01000027.1	Biosynthesis	Lamassu
ZorA	Pseudomonas aeruginosa PA7	6238298-6240424	CP000744.1	Biosynthesis	Zorya
ZorB	Pseudomonas aeruginosa PA7	6237534-6238298	CP000744.1	Biosynthesis	Zorya
ZorC	Pseudomonas aeruginosa PA7	6235525-6237471	CP000744.1	Biosynthesis	Zorya
ZorD	Pseudomonas aeruginosa PA7	6232483-6235515	CP000744.1	Biosynthesis	Zorya
ZorE	Enterobacter cloacae MNCRE12	158296-159387	JYME01000006.1	Biosynthesis	Zorya
KwaA	Staphylococcus epidermidis NIHLM018	22218-22778	AKGY01000003.1	Biosynthesis	Kiwa

KwaB	Staphylococcus epidermidis NIHLM018	22790-23740	AKGY01000003.1	Biosynthesis	Kiwa
DruA	Acidibacillus ferrooxidans ITV01	23658-24893	LPVJ01000027.1	Biosynthesis	Druantia
DruB	Bacillus cereus LK9	12597-14105	LDUP01000008.1	Biosynthesis	Druantia
DruC	Bacillus cereus LK9	10588-12576	LDUP01000008.1	Biosynthesis	Druantia
DruD	Bacillus cereus LK9	9624-10574	LDUP01000008.1	Biosynthesis	Druantia
DruE	Bacillus cereus LK9	4236-9563	LDUP01000008.1	Biosynthesis	Druantia
DruM	Clostridia bacterium BRH_c25	13070-14206	LOES01000090.1	Biosynthesis	Druantia
DruF	Clostridia bacterium BRH_c25	2777-4804	LOES01000090.1	Biosynthesis	Druantia
JetA	Bacillus cereus Q1	1011449-1012831	CP000227.1	Biosynthesis	Wadjet
JetB	Bacillus cereus Q1	1012841-1013434	CP000227.1	Biosynthesis	Wadjet
JetC	Bacillus cereus Q1	1013391-1016726	CP000227.1	Biosynthesis	Wadjet
JetD	Bacillus cereus Q1	1016686-1017915	CP000227.1	Biosynthesis	Wadjet
cyclase	Staphylococcus aureus #32S	2638440625	Ga0078613_106571	Biosynthesis	CBASS
type I cyclase	Staphylococcus aureus 144_S7	2667274917	Ga0111770_1149	Biosynthesis	CBASS
type I effector	Staphylococcus aureus 144_S7	2667274918	Ga0111770_1150	Biosynthesis	CBASS
type III cyclase	Staphylococcus aureus DEU28	2736304366	Ga0130311_101824	Biosynthesis	CBASS
type III HORMA1	Staphylococcus aureus DEU28	2736304367	Ga0130311_101825	Biosynthesis	CBASS
type III TRIP13	Staphylococcus aureus DEU28	2736304368	Ga0130311_101826	Biosynthesis	CBASS

type III effector	Staphylococcus aureus DEU28	2736304370	Ga0130311_101828	Biosynthesis	CBASS
type III exonuclease	Staphylococcus aureus DEU28	2736304371	Ga0130311_101829	Biosynthesis	CBASS
type III HORMA1	Staphylococcus aureus DEU35	2736313196	Ga0130313_103581	Biosynthesis	CBASS
type III cyclase	Staphylococcus aureus DEU35	2736313197	Ga0130313_103582	Biosynthesis	CBASS
Retron-TIR	Shigella dysenteriae NCTC2966	n/a	WP_005025120.1*	Biosynthesis	Retron-TIR
Ec67	Escherichia coli NCTC8623	n/a	WP_000169432.1	Biosynthesis	Retron-TOPRIM
Nuc_deoxy + retron	Escherichia coli BL21	n/a	WP_001034589.1	Biosynthesis	Nuc_deoxy + retron
Ec86	Escherichia coli BL21	n/a	WP_001320043.1	Biosynthesis	Nuc_deoxy + retron
Ec78	Escherichia coli ECONIH5	n/a	WP_001549208.1	Biosynthesis	Retron + ATPase + HNH
ptuA	Escherichia coli ECONIH5	n/a	WP_001549209.1	Biosynthesis	Retron + ATPase + HNH
ptuB	Escherichia coli ECONIH5	n/a	WP_001549210.1	Biosynthesis	Retron + ATPase + HNH
drt1a	Klebsiella pneumoniae NCTC9143	n/a	WP_115196278.1	Biosynthesis	RT-nitrilase (UG1)
drt1b	Klebsiella pneumoniae NCTC9143	n/a	WP_040189938.1	Biosynthesis	RT-nitrilase (UG1)
drt2	Salmonella enterica NCTC8273	n/a	WP_012737279.1	Biosynthesis	RT (UG2)
drt3a	Escherichia coli ECOR12	n/a	WP_087902017.1	Biosynthesis	RT (UG3) + RT (UG8)
drt3b	Escherichia coli ECOR12	n/a	WP_062891751.1	Biosynthesis	RT (UG3) + RT (UG8)
drt4	Escherichia coli 21-C8-A	n/a	GCK53192.1	Biosynthesis	RT (UG15)
drt5	Escherichia coli KTE25	n/a	WP_001524904.1	Biosynthesis	RT (UG16)

rdrA	Citrobacter rodentium DBS100	n/a	WP_012906049.1	Biosynthesis	ATPase + deaminase
rdrB	Citrobacter rodentium DBS100	n/a	WP_012906048.1	Biosynthesis	ATPase + deaminase
rdrA	Pluralibacter gergoviae ATCC33028	n/a	WP_155731552.1	Biosynthesis	ATPase + deaminase
rdrB	Pluralibacter gergoviae ATCC33028	n/a	WP_064360593.1	Biosynthesis	ATPase + deaminase
rdrD	Pluralibacter gergoviae ATCC33028	n/a	WP_064360592.1	Biosynthesis	ATPase + deaminase
apeA	Escherichia coli NCTC8008	n/a	WP_000706972.1	Biosynthesis	ApeA (HEPN)
avs1a	Erwinia piriflorinigrans CFBP5888	n/a	WP_023654314.1	Biosynthesis	MBL + protease-STAND
avs1b	Erwinia piriflorinigrans CFBP5888	n/a	WP_084007836.1*	Biosynthesis	MBL + protease-STAND
avs1c	Erwinia piriflorinigrans CFBP5888	n/a	WP_023654316.1	Biosynthesis	MBL + protease-STAND
avs2	Escherichia coli NCTC9087	n/a	WP_063118745.1	Biosynthesis	STAND
avs3a	Salmonella enterica NCTC13175	n/a	WP_126523998.1	Biosynthesis	DUF4297-STAND
avs3b	Salmonella enterica NCTC13175	n/a	WP_126523997.1*	Biosynthesis	DUF4297-STAND
avs4	Escherichia coli NCTC11132	n/a	WP_044068927.1	Biosynthesis	Mrr-STAND
avs5	Escherichia coli NCTC13384	n/a	WP_001515187.1	Biosynthesis	SIR2-STAND
dsr1	Escherichia coli NCTC9112	n/a	WP_029488749.1	Biosynthesis	SIR2-DUF4020

dsr2	Cronobacter sakazakii NCTC8155	n/a	WP_015387030.1*	Biosynthesis	SIR2
SIR2 + HerA	Escherichia coli NCTC11129	n/a	WP_021577683.1	Biosynthesis	SIR2 + HerA
herA	Escherichia coli NCTC11129	n/a	WP_021577682.1	Biosynthesis	SIR2 + HerA
DUF4297 + HerA	Escherichia coli NCTC11131	n/a	WP_016239654.1	Biosynthesis	DUF4297 + HerA
herA	Escherichia coli NCTC11131	n/a	WP_016239655.1	Biosynthesis	DUF4297 + HerA
tmn	Escherichia coli ECOR25	n/a	WP_001683567.1	Biosynthesis	Transmembrane ATPase
qatA	Escherichia coli NCTC9009	n/a	STG85056.1	Biosynthesis	ATPase + QueC + TatD
qatB	Escherichia coli NCTC9010	n/a	STG85057.1	Biosynthesis	ATPase + QueC + TatD
qatC	Escherichia coli NCTC9011	n/a	STG85058.1	Biosynthesis	ATPase + QueC + TatD
qatD	Escherichia coli NCTC9012	n/a	STG85059.1	Biosynthesis	ATPase + QueC + TatD
hhe	Escherichia coli ATCC43886	n/a	WP_032200272.1	Biosynthesis	DUF4011-helicase-Vsr
mzaA	Salmonella enterica NCTC5773	n/a	VEA06816.1*	Biosynthesis	MutL + Z1 + DUF + AIPR
mzaB	Salmonella enterica NCTC5773	n/a	VEA06814.1	Biosynthesis	MutL + Z1 + DUF + AIPR
mzaC	Salmonella enterica NCTC5773	n/a	VEA06812.1	Biosynthesis	MutL + Z1 + DUF + AIPR
mzaD	Salmonella enterica NCTC5773	n/a	VEA06810.1	Biosynthesis	MutL + Z1 + DUF + AIPR
mzaE	Salmonella enterica NCTC5773	n/a	VEA06808.1	Biosynthesis	MutL + Z1 + DUF + AIPR

terY	Citrobacter gillenii NCTC9094	n/a	WP_115257868.1	Biosynthesis	vWA + PP2C + STK-OB
vWA + PP2C + STK-OB	Citrobacter gillenii NCTC9094	n/a	WP_115257869.1	Biosynthesis	vWA + PP2C + STK-OB
vWA + PP2C + STK-OB (2)	Citrobacter gillenii NCTC9094	n/a	WP_115257870.1	Biosynthesis	vWA + PP2C + STK-OB
upx	Salmonella enterica NCTC6026	n/a	WP_060647174.1	Biosynthesis	DUF1887
ppl	Escherichia coli NCTC8620	n/a	STM52149.1	Biosynthesis	Phosphoesterase-ATPase
ietA	Escherichia coli ECOR52	n/a	WP_000385105.1	Biosynthesis	ATPase + protease
ietS	Escherichia coli ECOR52	n/a	WP_001551050.1	Biosynthesis	ATPase + protease
Restriction-like system	Escherichia coli ECOR58	n/a	WP_000860009.1	Biosynthesis	Restriction-like system
Restriction-like system	Escherichia coli ECOR58	n/a	WP_001044652.1	Biosynthesis	Restriction-like system
Restriction-like system	Escherichia coli ECOR58	n/a	WP_001207938.1	Biosynthesis	Restriction-like system
Restriction-like system	Escherichia coli ECOR58	n/a	WP_000985714.1	Biosynthesis	Restriction-like system

Chapter 5: Development of an amplicon nanopore sequencing strategy for detection of mutations conferring intermediate resistance to vancomycin in *Staphylococcus aureus* strains

Abraham G. Moller, Robert A. Petit III, Michelle N. Hargita, and Timothy Read

Abstract

Staphylococcus aureus is a major nosocomial pathogen diverse in pathologies and increasingly antibiotic resistant. Resistance is emerging to important drugs against *S. aureus* infections, such as the cell-wall active antibiotic vancomycin. While full vancomycin resistance is rare, vancomycin intermediate *S. aureus* (VISA) has been detected since the 1990s. Here we developed a combined PCR/long-read sequencing-based method to detect previously known VISA-causing mutations. We amplified 16 genes (*walR*, *walK*, *rpoB*, *graR*, *graS*, *vraF*, *vraG*, *stpI*, *vraR*, *vraS*, *agrA*, *sarA*, *clpP*, *ccpA*, *prsA*, and *yvqF*) known to contain most VISA-conferring mutations as 10 amplicons and sequenced amplicon pools as long-reads with Oxford Nanopore adapter ligation on inexpensive Flongle flow cells. We then detected mutations by mapping reads against a parental consensus or known reference sequence and comparing called variants against a database of known VISA mutations from laboratory selection. There was high (>1000x) coverage of each amplicon in the pool, no relationship between amplicon length and coverage, and the ability to detect the causative mutation (*walK* 645C>G) in a VISA mutant derived from the USA300 methicillin-resistant (MRSA) strain (N384-3 from parental strain N384). Mixing mutant (N384-3) and parental (N384) DNA at various ratios from 0 to 1 mutant suggested a mutation detection threshold of roughly the average minor allele frequency of 6.5% at 95% confidence (two standard error above mean mutation frequency). Future work in this area will further develop the assay for direct phenotype calling from clinical samples.

Introduction

Staphylococcus aureus is a major nosocomial pathogen carried by 30-70% of the world's population (1) that causes diverse pathologies. It is also increasingly antibiotic resistant, with methicillin resistance (MRSA) first reported in the 1960s (358). The glycopeptide vancomycin, first released in 1958, has long been used for the treatment of severe MRSA infections (46). However, even resistance to vancomycin has emerged in *S. aureus*. Vancomycin-intermediate resistance (VISA) was first reported in the 1990s (46). Full vancomycin resistance (VRSA), though detected in *S. aureus* in 2002 (30, 31), fortunately remains rare.

VISA is defined by a vancomycin MIC between 4 and 8 µg/mL, while heterogeneous VISA (hVISA) strains have a vancomycin MIC between 2 and 4 µg/mL Mutations associated with VISA are found in multiple sites in a cluster of key genes (165), while the *vanA* gene, which is carried on an integrative conjugative element (ICE) (29), is responsible for full vancomycin resistance (VRSA) (45). The VISA phenotype includes a thickened cell wall, reduced autolysis, increased capsule, and increased D-alanylation of teichoic acids (46). The thickened VISA cell wall both reduces vancomycin diffusion and contains more free D-alanyl-D-alanine that can bind vancomycin, leading to less ability for vancomycin to reach the site of cell wall synthesis (the septum), and thus reduced sensitivity to the antibiotic (46). VISA strains furthermore have been shown to have reduced vancomycin susceptibility *in vivo* (46, 359) and to be associated with persistent bacteremia in clinical studies (46, 360–362). Despite its prevalence and significance, current VISA detection methods remain laborious and time-consuming (46, 363–365). VISA detection requires either E broth microdilution determination of minimum inhibitory concentration (MIC), while hVISA detection requires the more complex population analysis profile-area under the curve (PAP-AUC) assay (46). MIC and Etest assays require at least 24 hours and the PAP-AUC assay requires 48 hours incubation before resistance determination (46). Improving VISA

treatment will thus require more rapid detection methods reducing or eliminating the time taken culturing.

Recent advances in sequencing technology offer a quicker alternative to culture-based tests for antibiotic resistance. DNA can be sequenced by pulling it through a protein nanopore, measuring the corresponding changes in current, and converting this current trace into a DNA sequence (134). While it has a higher per-base error rate (~10%) than short-read sequencing (e.g., Illumina), nanopore sequencing not only results in longer read length relative to other sequencing technologies (over 1 kb per read) but also speed - sequencing data can be collected in real time, as early as the beginning of the sequencing run. Improvements in base calling have also led to reduced error rates, addressing its major weakness (136).

Recent studies have demonstrated that nanopore sequencing can detect antibiotic resistance genes in metagenomic samples (151, 366–368) and distinguish antibiotic-sensitive from resistant strains far faster than culture-based methods, which typically require overnight growth (163, 369). A method for distinguishing carbapenem-resistant from sensitive *Klebsiella pneumoniae* through 16S rRNA present in culture under imipenem treatment only requires four hours of culturing (163). Another method that extrapolates antibiotic resistance from pneumococcal sequence type takes only five minutes (369). No nanopore method has yet been developed for direct detection of resistance-causing mutations from PCR amplicons, but such mutations have been detected in metagenomic sequence data collected from urine containing *Neisseria gonorrhoeae* (161). Additionally, genomic prediction of antibiotic resistance must still be calibrated thoroughly against culture-based methods (164).

Here we present a method to rapidly determine likely vancomycin-intermediate *Staphylococcus aureus* (VISA)-conferring mutations in *S. aureus* strains or clinical samples by coupling PCR and nanopore sequencing. This method is better than any other possible approach (e.g., qPCR for resistance genes or mutations, culture-based testing) because there are many possible mutations in many genes that may cause VISA, including mutations we have never seen

before, which makes sequencing followed by bioinformatic mutation detection and interpretation necessary. To our knowledge, this is the first study aimed at detecting mutations (SNPs and small indels) linked to antibiotic resistance rather than rather than whole genes through nanopore sequencing. We developed a PCR scheme to amplify 10 regions containing the 16 genes (*walR*, *walK*, *rpoB*, *graR*, *graS*, *vraF*, *vraG*, *stpI*, *vraR*, *vraS*, *agrA*, *sarA*, *clpP*, *ccpA*, *prsA*, and *yvqF*) most likely to contain VISA-conferring mutations based on previous work (165). We then developed a method to detect such mutations without culturing through sequencing, alignment against a reference, and comparison to a database of VISA-conferring mutations, taking far less than the 24+ hours necessary for culture-based detection. As a proof of principle, we have sequenced VISA amplicons from a parent strain (N384) and one mutant (N384-3) to detect VISA-conferring mutations. We also detected the mutation when its abundance was as low as 1% relative to the parent and when multiple samples were barcoded and sequenced on the same flow cell. We also detected VISA mutations directly from clinical samples. Future projects will attempt to amplify all regions together through a multiplex PCR reaction and detect VISA mutations in the shortest possible time through machine learning methods.

Results

Development of a nanopore sequencing VISA detection pipeline

We decided to detect VISA mutations through a strategy combining DNA extraction from a *S. aureus* isolate, PCR amplification of *S. aureus* genes often implicated in VISA, and nanopore sequencing of these amplicons to identify mutations. Our proposed pipeline is outlined in Figure 1. *S. aureus* may be isolated either directly (isolation from culturing) or indirectly (metagenomic DNA extraction) from a clinical sample such as a blood bottle. *S. aureus* populations cannot be assumed to be clonal; instead they may include a mixture of the VSSA parental strain and at least one VISA mutant, if not more. Challenges must be addressed in four general areas - 1) sample isolation, 2) PCR of multiple unlinked genomic regions, 3) sequencing, 4) mutation calling and

calling of VISA/VSSA based on mutation patterns. Sample isolation must both produce enough DNA for PCR and remove any contaminating non-bacterial DNA that would lead to subsequent nonspecific PCR, which we found in amplification from mouse blood spiked with 1e2 or 1e4 CFU/mL N384 (data not shown). PCR must efficiently and specifically - as this previous issue indicates - amplify all 10 markers. Error and time of sequencing must be reduced to the least possible and the least necessary, respectively. Finally, mutation calling must successfully distinguish sequencing errors from VISA-causing mutations. This will require thorough scrutiny of the parental strain sequencing read-parental reference genome alignment to find all detectable mutations. Any mutations called from this alignment would represent systematic sequencing errors that must not be considered as potential VISA-causing mutations. These could lead to false positive errors if they were called as VISA mutations in unknown clinical samples, or false negative errors if they made it impossible to properly detect a true VISA mutation at the same location.

Steps 1 and 2 - sample isolation and PCR

Our efforts found that it was only possible to extract appropriate DNA template for PCR from bacterial cultures themselves with a DNA extraction kit. We attempted spiking CD1 mouse blood with 1e2 or 1e4 CFU/mL N384 or N384-3, but in either case upon DNA extraction with the Qiagen DNeasy Blood and Tissue Mini Kit, we only obtained nonspecific amplification with PCR for all amplicons, presumably from mouse DNA (data not shown). We also found that microwaving culture pellets to extract DNA resulted in template too fragmented to amplify our long regions of interest. However, when we extracted DNA from isolated bacterial cultures, we did manage to amplify our regions of interest (Figure 2). We thus proceeded to extract DNA from bacterial cultures with the modified Qiagen DNeasy Blood and Tissue Mini Kit protocol to serve as a PCR template in subsequent experiments (extracting DNA from mixed N384 and N384-3 cultures).

Steps 3 and 4 - sequencing and mutation detection: limit of detection approaches average minor allele frequency

To evaluate the limit of detection of our assay, we prepared three different mixes of parental VSSA (N384) and VISA (N384-3) mutant DNA - 1) culture mixtures, 2) DNA amplicon mixtures, and 3) simulated amplicon sequencing read mixes. We then called mutations through two different methods (bwa alignment (370) followed by bcftools consensus calling (371); medaka (372) alignment, consensus determination, and consensus variant calling) and analyzed aligned nucleotide counts for each mixture. Culture and DNA amplicon mixtures had similar coverages to each other but not to simulated DNA read mixtures based on non-parametric Wilcoxon tests (Figure 3A and 3B). The same pattern was present at a per-amplicon level (Figure 3A), except for *walRK*, where all three sets overlapped, and *prsA*, where the simulated coverage was significantly lower than that of either. In all cases (cell, DNA, or simulated mixtures), coverage was in excess of 3000-fold, which is well above the inverse of the error rate (~1/0.05 or 20-fold), suggesting coverage is high enough to compensate for at least random errors. Additionally, in all cases, there was no significant relationship between amplicon length and amplicon coverage (Figure 3C).

Given that we achieved high coverage for all amplicons in all cases, we proceeded to call mutations in each mixture (Table 3) and analyze detected mutation frequency relation with mutant proportion in the mixture (Figure 4). Amongst the simulated mixtures, medaka only detected the N384-3 *walK* 645C>G mutation at 60% mutant proportion in the mutant/parent mixture or higher, while bwa/bcftools detected the mutation at a proportion of 50% mutant or higher (Table 3). For the cell and DNA mixtures, on the other hand, medaka detected this mutation at 50% mutant proportion or higher while bwa/bcftools detected it at 10% mutant or higher (Table 3). This suggests the standard variant calling pipeline (bwa/bcftools) was more sensitive than the faster medaka. Regarding the correlation between introduced and detected mutant proportion, we found the greatest deviation from a perfect correlation with the culture mixtures (Figure 4A). At 50%

mutant, the detected mutant proportion was below that detected from all other mixtures at the same introduced mutant proportion. We hypothesize that differences in DNA extraction efficiency between the VISA mutant and VSSA strain explain this deviation, as the thicker VISA cell walls would make DNA harder to extract, leading to a lower detected mutation proportion than expected based on how the cultures were mixed.

We also examined the Z-score for each mixture and each set to determine the limit of detection based on average and standard error of inherent sequencing variation found in the alignments (Figure 4B). In all cases, average minor (non-consensus) allele frequency was roughly 6.5% and tested mutant proportions above this threshold were well more than two standard errors (in fact, at least 100 standard errors) above the mean. Z-scores were highest for cell mixtures amongst the sets of mixtures evaluated (DNA and simulated reads). This suggests that the limit of detection based on our method is close (standard errors of minor allele frequency were roughly 1e-4) to the average minor allele frequency.

Successful identification of VISA-associated mutations in a clinical VISA strain

To show our assay could identify VISA mutations in clinical strains, we sequenced two known clinical strains - one VSSA (EUH15) and one VISA (107) - characterized in a previous study. Strain 107 is known to contain four missense mutations in VISA-associated genes (WalK A243T at 26374, GraR E15K at 708287, VraG T217I at 711484, VraS A314V at 1947464) relative to the N315 reference, while strain EUH15 contains none of these four. Mutation calling results from two different methods (bwa/bcftools and medaka) are shown in Table 4. The first method (bwa/bcftools) identified all four mutations in strain 107 but not strain EUH15, but the second method (medaka) identified only three (in *walK*, *graR*, and *vraS*). This indicates that we can distinguish VISA from VSSA in at least two unknown clinical strains with our assay and bioinformatic pipeline. Standard variant calling (bwa/bcftools) outperformed rapid variant calling (medaka) in sensitivity as we observed in our limit of detection study.

Step 5 - bioinformatic analysis: construction of a VISA-associated mutation database to enhance detection

In the future, we plan to build a curated database of all VISA-associated mutations to make clinical VISA detection as sensitive as possible. We will collect mutations in the ten genes of interest with four levels of support for VISA causation - 1) laboratory selection for VISA, 2) identification in clinical VISA strains but not VSSA strains, 3) variation in our curated Staphopia database that shows the potential for VISA causation (e.g., nonsynonymous mutations), and 4) simulated nonsynonymous mutations in our VISA-associated genes with potential for VISA causation. The critical control to evaluate our ability to discriminate VISA from VSSA based on this database is the overlap between VISA mutations and mutations called from parental N384 nanopore reads against the N384 reference genome. Such mutations would likely represent systematic sequencing errors that would be called as false positives.

Discussion

Nanopore sequencing has the potential to revolutionize clinical diagnostics, but first must overcome multiple challenges, systematic sequencing error primarily. Nanopore sequencing may in particular revolutionize antimicrobial resistance detection as it removes the need for multiple bacterial culturing steps, each taking 24-48 hours. Indeed, the time lost to culturing has been suggested to be a cause of mortality itself, as mortality increases 10% with each hour in septic shock cases (156). Multiple studies have shown that such sequencing can rapidly (15 minutes or less after sequencing begins) identify antibiotic resistance genes in plasmids (157) or antibiotic resistant strain lineages (160) and detect antibiotic resistance mutations in metagenomic sequencing from clinical samples (161). However, nanopore sequencing has only recently become inexpensive, requires new expertise different from what is common in clinical labs, and has higher sequencing error (~3-5% average per read per base for the newest versions) than other sequencing methods. In addition, this sequencing error is often systematic, focused in loci

such as homopolymer regions. The critical problem to overcome for these nanopore applications is distinguishing sequencing errors from causative resistance mutations when analyzing sequence traces.

These results suggest that our assay could identify VISA mutations in unknown clinical strains, but work still must be done to further improve the limit of detection and accurately call mutations. Medaka, a rapid neural network-based alternative to standard alignment and variant calling (372), proved ineffective both in sensitively detecting mutations in the mutant/parent mixtures and in comprehensive detection of the strain 107 VISA mutations. Optimizing 1) bacterial isolation, 2) DNA extraction, 3) PCR scheme, 4) sequencing (time), 4) basecalling, and 5) alignment method (Figure 1) will be necessary to make our assay as useful as possible. We must reduce culturing time to as little as necessary (24 hours conventionally) to get enough isolated bacterial culture (plate or liquid) for DNA extraction, followed by quick enough (~1 hr, including 30 min lysostaphin/lysozyme treatment) DNA extraction. The PCR scheme strongly amplifies all regions, but takes at least two hours. Regarding sequencing time, we need to determine the minimum number of reads necessary to call a mutation in our mutant/parent mixtures by randomly subsampling various numbers of reads and repeating our variant calling pipelines. We also need to see whether fast basecalling is sufficiently accurate to call VISA mutations, as we used the high accuracy, neural network-dependent bonito basecaller (373) with external GPUs after sequencing. We attempted to speed up alignment with medaka, but our standard protocol (bwa/bcftools) sensitively and completely called VISA mutations. Overall, we are likely time-limited by alignment, the PCR step, which must amplify regions longer than 3 kb, and more importantly, culturing, which requires a single day after obtaining the clinical sample.

We must consider that this assay only is clinically useful if a patient is known to have a *S. aureus* infection. In other words, if there is no amplification of the ten VISA regions, there is no *S. aureus* present, and the assay should not proceed. In addition, results suggest amplification is only successful from pure culture. Attempts to amplify VISA markers from mouse blood spiked

with 1e2 or 1e4 CFU/mL *S. aureus* failed with extensive nonspecific amplification. This suggests at least one culturing step may be necessary to isolate *S. aureus* for DNA extraction and PCR. However, this culturing step could be kept short (~1 hr), as in previous studies, or a method could be developed to selectively amplify bacterial DNA from the clinical sample. We must also realize that the success of the assay depends on the chance of finding a mutant in a clinical sample. The sample must contain at least enough bacteria that one bacterium carries the resistance mutation, whether metagenomic DNA or isolated bacterial genomic DNA serves as PCR template.

Future work will focus on the construction of a VISA mutation database for comprehensive mutation detection in unknown samples. As stated before, this will collect lab-evolved, clinical, and predicted VISA mutations in one database. This database should make detection possible across diverse strains. The critical control we must resolve is sequencing errors that overlap with mutations in this database. We will assess this by comparing mutations called from aligning N384 parental sequencing reads against the N384 reference genome with this database.

Materials and Methods

Strain selection and media

Either a MRSA parent strain (N384) and its lab-selected VISA mutant (N384-3) (Su et al., unpublished) or clinical VSSA (EUH15) or VISA (107) strains were used for all downstream experiments. N384-3 was selected from N384 through stepwise evolution in vancomycin up to 8 μg/mL. Strains and associated metadata (vancomycin MICs) are listed in Table 1. All strains were grown on tryptic soy agar (TSA) or in tryptic soy broth (TSB) at 37°C with 225 rpm agitation for liquid culture.

Preparation of mutant/parent mixtures (cell and DNA) for evaluating limit of detection

In order to evaluate the limit of mutation detection, mutant and parent sequence were mixed together in five ratios from 100% N384-3 mutant to 100% N384 parent (100% N384-3, 90% N384-3; 10% N384, 50% N384-3; 50% N384, 10% N384-3; 90% N384, 100% N384). Either turbid

bacterial cultures or amplicon DNA volume were mixed in these ratios (total amplicon volume of 100 µL; 1 mL culture volume; 3e9 CFU/mL). DNA was extracted from cell mixtures with a modified Qiagen DNeasy Blood and Tissue Mini Kit. Cultures were pre-treated for 30 minutes with 0.2 mg/mL lysostaphin and 1 mg/mL lysozyme at 37°C to cleave *S. aureus* cell walls and then DNA was extracted following manufacturer's directions.

Amplifying VISA-associated loci

VISA amplicons represented regions from the USA300 FPR3757 genome containing VISA-associated genes or clusters of such genes (e.g., the *walRK* two-component system) and 1000 bp adjacent sequence on either side (Table 2). These regions were amplified through PCR from N384 parent/mutant DNA mixture templates. All regions were amplified using NEB Q5 High-Fidelity 2X Master Mix and the following PCR program: 98°C for 30 seconds (initial denaturation); 98°C for 10 seconds, 65°C for 30 seconds (annealing), and 72°C for 3 minutes 30 seconds (35 cycles; extension for up to 6.5 kbp sequence); 72°C for 2 minutes, final extension; and then hold at 4°C.

Amplicon nanopore sequencing

VISA amplicons were pooled to a total of 1 µg for sequencing (equal quantities of each amplicon). VISA amplicon libraries were generated using the 1D ligation sequencing kit (SQK-LSK109; Oxford Nanopore) modified for amplicon sequencing. 0.2 pmol of amplicon DNA was ligated to adapters to bias ligation reactions toward individual amplicons. Libraries were sequenced on Flongle FLO-FLG001 flow cells (Oxford Nanopore). Read data (event-level, FAST5 format) was collected on a PC using MinKNOW software (Oxford Nanopore).

Basecalling, alignment, and variant calling

For *de novo* basecalling, bonito (version 0.3.6) (373) was used with nucleotide output stored as a FASTA file. For detecting nucleotide changes compared to a reference (mapping), bonito-basecalled reads were aligned against the USA300 FPR3757 reference using BWA

(version 0.7.17; parameter -x ont2d) (370). Variants were called with bcftools (mpileup/call on the resulting read alignment with consensus and multiallelic variant calling) and medaka (default parameters, reads against USA300 FPR3757 reference), with results saved in a VCF format file. Variants were further filtered down to biallelic sites with bcftools view. The consensus sequence was determined using samtools mpileup (-B option selected to cancel base alignment quality assessment) and bcftools (-c option for consensus determination). Variants were compared to the known N384-3 VISA mutations (*walK* 645C>G) in the case of N384 parent/mutant mixtures and known, previously identified VISA-associated mutations in the case of EUH15 and 107 test strains.

In silico read simulation and mutation frequency analysis

In silico amplicon sequence mixtures (300,000 reads in each case; 100% N384-3, through 100% N384 in 10% increments of each) were simulated using NanoSim (374) using default parameters from the wild-type (N384) and VISA mutant (N384-3) amplicon sequences concatenated into a single FASTA sequence (predicted amplicons from primers in Table 2). Amplicon reads were simulated with an error profile trained from bonito-basecalled N384 parent amplicon (sequenced from amplicon DNA rather than cell mixture series). A total of 300,000 reads were simulated in each case. These read mixtures were then used together with cell and DNA mixture reads to determine how VISA mutation frequency correlated between variant calling (observed mutation frequency) and VISA proportion (predicted VISA mutation frequency) and determine the detection threshold without use of medaka. In order to address these questions, simulated and real (cell or DNA) mixtures were aligned against the USA300 FPR3757 (N384) reference using BWA (version 0.7.17; parameter -x ont2d) (370) and analyzed for variant counts with samtools (371) mpileup (-B option selected to cancel base alignment quality assessment). Mpileup output was further processed into a table with counts of every base, positive strand matches (.), negative strand matches (,), and overall mapped read coverage at each amplicon

site. Overall non-reference nucleotide frequency was calculated for every mapped amplicon nucleotide to generate an overall distribution with a mean and standard deviation. For each proportion mutant, the detected mutation (*walK* 645C>G at USA300 FPR3757 position 26311) proportion relative to mapped coverage and the Z-score. The detection limit was calculated as the extrapolated observed mutation proportion for which the Z-score exceeded two (two standard deviations above the mean to give two-tailed 95% confidence).

Constructing a VISA mutation database

A database of VISA mutation was constructed from three levels of information: 1) lab-evolved VISA mutations, 2) VISA-associated mutations in clinical strains, 3) predicted VISA-causing mutations detected in our database of 40,000+ *S. aureus* genomes and 4) simulated mutations in VISA-associated genes expected to cause VISA. This database was constructed by 1) collecting mutations identified in our laboratory VISA evolution studies (Su et al., unpublished), 2) those previously identified in clinical strains in our lab or others, 3) finding nonsense or missense mutations relative to the N315 reference in our Staphopia database, and 4) performing a saturating mutagenesis of the genes in our amplicons and identifying all nonsense or missense mutations predicted to confer VISA.

Acknowledgements

We thank members of the lab for constructive criticism. Timothy D. Read was supported by NIH R21 AI 138079-02, while Abraham G. Moller was supported by the NSF GRFP. We thank NSF for providing GPU resources via the XSEDE supercomputing consortium to conduct high accuracy basecalling with bonito.

Figures

Figure 1: Overview of the VISA amplicon sequencing process, from sample isolation to SNP calling and comparison to known VISA mutations. Steps 1-5 represent sample isolation, DNA extraction, PCR, nanopore sequencing, and variant calling, respectively. Two VISA mutations are illustrated in red and green (steps 2, 3, and 5) while their parental strain is in blue.

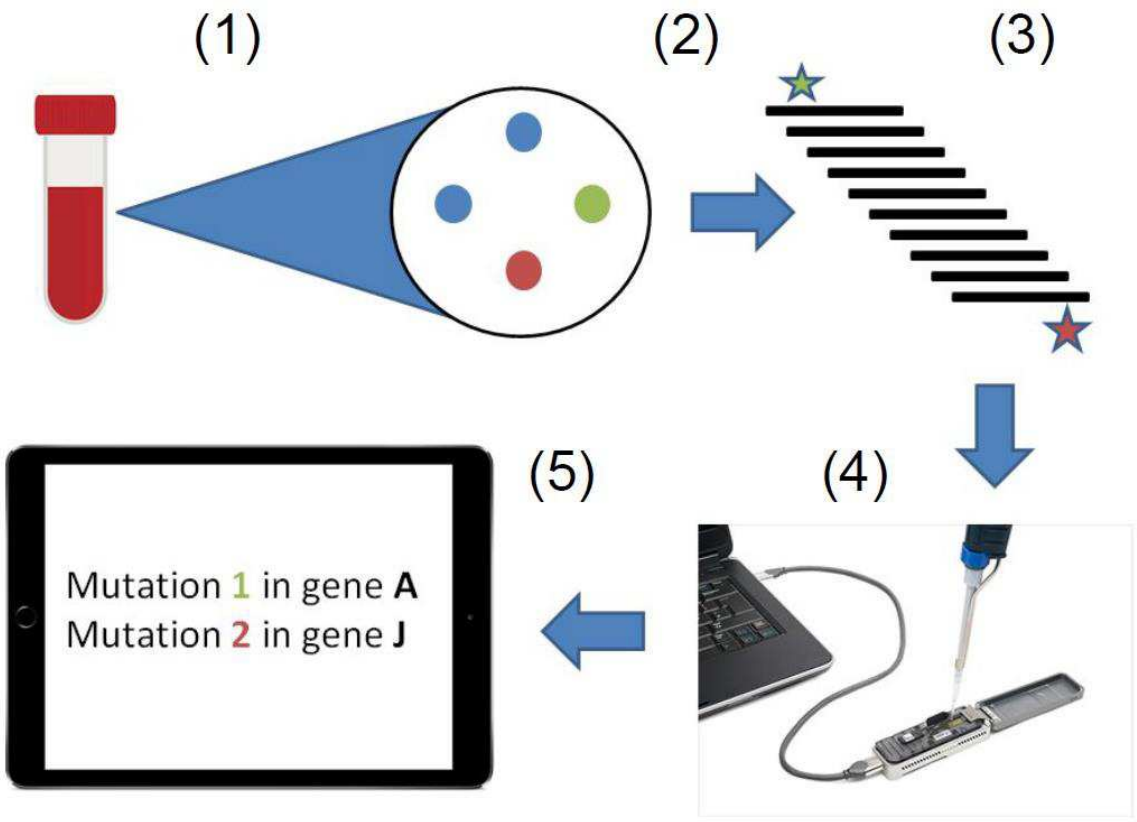


Figure 2: Evidence that our 10 VISA regions can be amplified from N384-3 mutant DNA extracted from pure culture. From left to right, the amplicons are *walRK*, *rpoB* (old primers, excluded from Table 2), *graRS*, *vraFG*, *stpI*, *vraRS*, *agrA*, *sarA*, *clpP*, *ccpA*, *prsA*, and *yvqF*. The amplicon to the right of the ladder is *rpoB* (new primers) in the second panel.

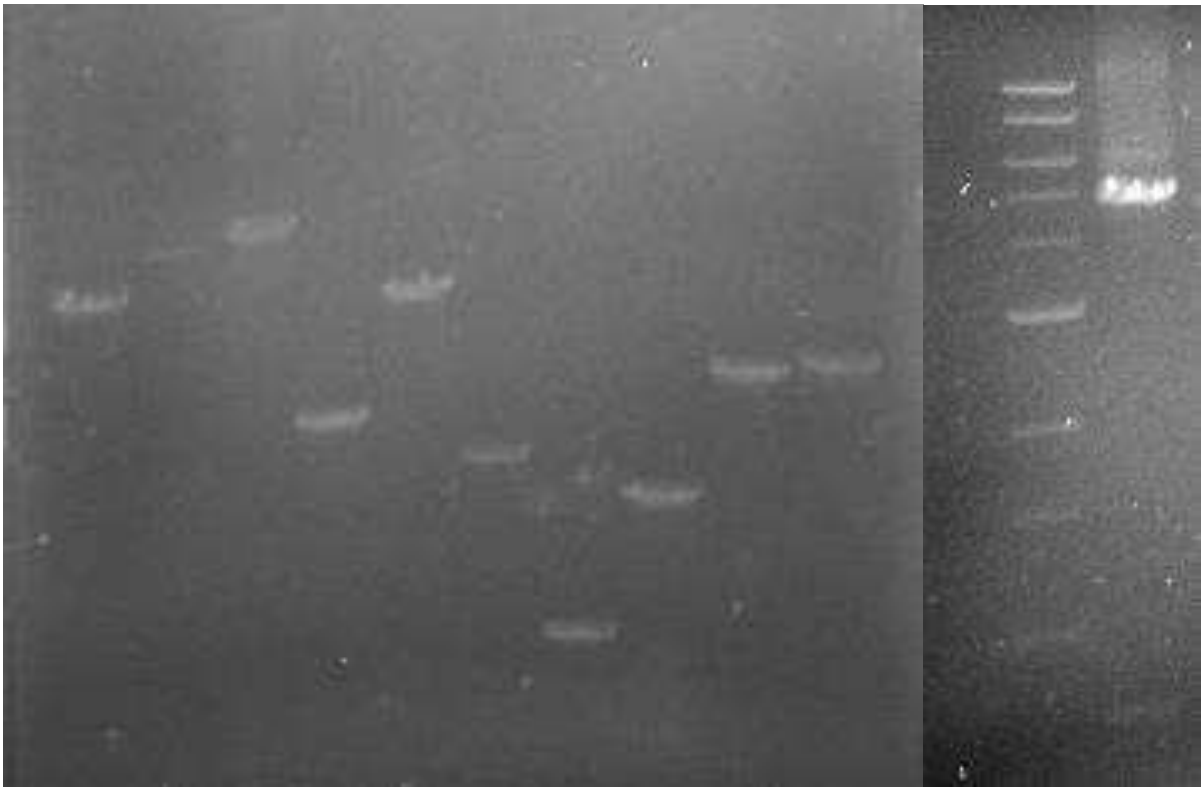


Figure 3: Gene coverage analyses for simulation (blue), DNA (amplicon; green), and cell (culture; red) parent/mutant mixtures. A) Coverage for each of the ten amplicons visualized as a violin plot; coverage for each amplicon is compared amongst mixture sets with a non-parametric WIlcoxon test. B) Coverage for each mixture set (simulation, DNA, or cell) visualized as a violin plot; coverage is compared between mixture sets with a non-parametric Wilcoxon test. C) Coverage against amplicon length for each mixture set presented with correlation and p-value.

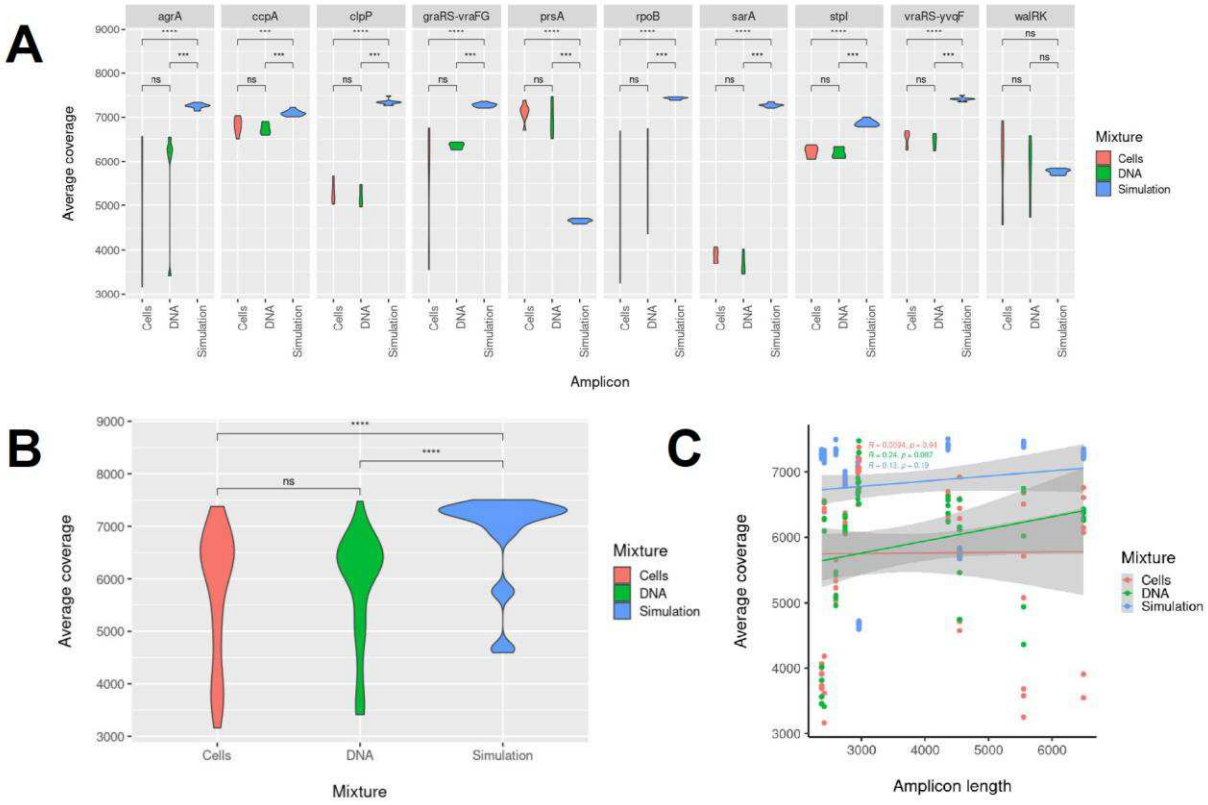
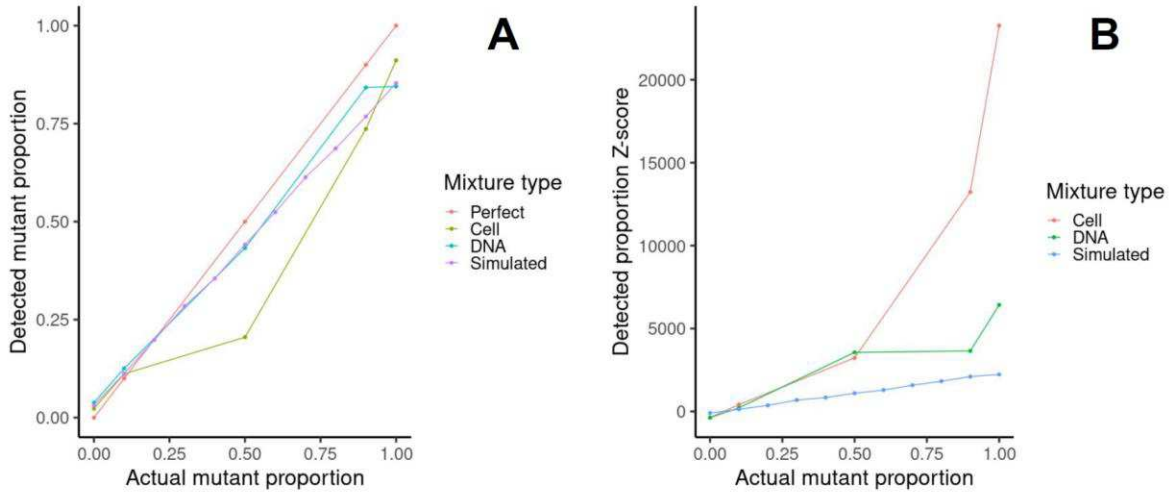


Figure 4: Evaluating limit of detection over a range of mutant/parent mixes (cultures, amplicon DNA, and simulated sequence reads). A) Introduced mutant proportion (x-axis) vs. detected mutation proportion (y-axis) for culture (olive-green), amplicon DNA (blue-green), or simulated (purple) sequence read mixtures along with the diagonal expected if the introduced mutant proportion matched detected mutation proportion (perfect; red). B) Introduced mutant proportion (x-axis) vs. detected mutation Z-score (y-axis) for cell (red), amplicon DNA (green), or simulated (blue) sequence read mixtures.



Tables

Table 1: Strains used in this study

Strain	Description	Reference(s)
N384 (NRS384)	Community-associated USA300 MRSA strain (CC8)	Su et al., unpublished
N384-3	Laboratory selected VISA mutant in N384 background; mutation *walK* 645C>G	Su et al., unpublished
EUH15	VSSA clinical isolate (CC5); vancomycin MIC is 1 µg/mL	(165)
107	VISA clinical isolate (CC5); vancomycin MIC is 4 µg/mL	(165)

Table 2: Primers used for amplifying each of 10 regions containing genes likely to contain VISA-conferring mutations

ID 1	Primer 1 sequence	Tm 1	ID 2	Primer 2 sequence	Tm 2
walRK F	GATTCCCGTCGAGACCGTAC	67	walRK R	CGCTTCATCTTCGGACAGGT	67
rpoB F	GCGAATTGCCGATGTTGGTT	67	rpoB R	TGCAACGAATTGACCTGGGT	67
graRS-vraFG F	GCTTTGAAGTTGACTGCCGG	67	graRS-vraFG R	AGCAGCACGATCCAGATTGA	66
stpI F	AAGCTTACACGCCGCAAAAG	67	stpI R	CTGATGATGAGCAGGCCCAT	67
vraRS-yvqF F	CATATGGCAGTATCGCGGGT	67	vraRS-yvqF R	ATGGGCTTTTGAAACGAGCG	66
agrA F	GAAGATGACATGCCTGGCCT	68	agrA R	TTGATACAACTGGGGCAGGG	67
sarA F	GCGGTGGCAATTCGTTCATT	67	sarA R	TCGGGCAAATGTATCGAGCA	67
clpP F	CGACATTGCGGGATTCTCTG	66	clpP R	TTGTCATCGGTCGTTTCGGT	67
ccpA F	CTCAACCTGGTCGAGCAAGT	67	ccpA R	TTGCACTTAGTGATGCGGGT	67
prsA F	ATCCCCACTTTCGCGTTTCA	67	prsA R	GCACCTTTATCACCGGCAGA	68

Table 3: Mutation calls (whether *walK* 645C>G was called) by bwa/bcftools and medaka for simulated, amplicon DNA, and culture mixtures

Proportion N384/N384-3	Simulated (NanoSim)	Amplicon DNA	Culture
100% N384	no/no	no/no	no/no
90% N384 10% N384-3	no/no	yes/no	yes/no
80% N384 20% N384-3	no/no		
70% N384 30% N384-3	no/no		
60% N384 40% N384-3	no/no		
50% N384 50% N384-3	yes/no	yes/yes	yes/no
40% N384 60% N384-3	yes/yes		
30% N384 70% N384-3	yes/yes		
20% N384 80% N384-3	yes/yes		
10% N384 90% N384-3	yes/yes	yes/yes	yes/yes
100% N384-3	yes/yes	yes/yes	yes/yes

Table 4: Mutation calls by bwa/bcftools and medaka for test strains EUH15 and 107. Mutations and reference coordinates are given relative to USA300 FPR3757.

Strain	bwa/bcftools	medaka
107 (VISA)	WalK A243T (26392) GraR E15K (719060) VraG T217I (718728) VraS A314V (2026309)	WalK A243T (26392) GraR E15K (719060) VraS A314V (2026309)
EUH15 (VSSA)	none	none

Chapter 6: Conclusions

My dissertation sought to understand the host basis of phage host range in *Staphylococcus aureus* with the ultimate goal of improving phage therapy against *S. aureus* infections. Increasing antibiotic resistance, high prevalence, and failure to develop vaccines makes alternative *S. aureus* therapies such as phage therapy crucial. However, while *S. aureus* phages have long been known to have broad host ranges (129), the mechanism behind the resistant exceptions remains unknown. This paradox justified comprehensive, population-wide studies to understand determinants of phage host range in *S. aureus* on a species rather than strain-wide level. In my dissertation, I not only reviewed the literature to identify what is known about host range and resistance factors in the species but also used population genomics to discover new host range factors and examine the evolution of known factors in the species. I also developed a sequence-based assay with the future potential to rapidly determine phage susceptibility of clinical isolates.

Summary of dissertation studies

In the second chapter of my dissertation, I reviewed the literature regarding host determinants of phage host range in *S. aureus* (53). I found that host resistance factors have been identified and characterized at three stages of the infection cycle (adsorption, biosynthesis, and assembly) but not the uptake nor lysis stages. I also hypothesized that these factors determine host range in a phylogenetically hierarchical manner given their respective conservation in the species. Upon completing this review, I proceeded to undertake a prospective study to discover new phage host range determinants in the species and a retrospective study to examine the evolutionary patterns of known host range determinants in the species, which represent the following two thesis chapters.

In chapter 3, I performed a genome-wide association study (GWAS) on diverse, host range phenotyped, genome-sequenced *S. aureus* strains to identify new determinants of phage host range in the species (166). *S. aureus* phages tested varied in host range, with the broadest being virulent *Myoviridae*/*Siphoviridae*, and host range lacked phylogenetic bias. GWAS identified novel core genes involved in host range phenotype as well as known (*tarP*) (110) host range determinants. Molecular genetic techniques (backcrossing and complementation) confirmed a subset of novel determinants (*trpA*, *phoR*, *isdB*, *sodM*, *fmtC*, and *relA*) did have causal associations with the phenotype. Significant determinants only partially explained the phenotype in predictive modeling and prediction accuracy was inversely proportional to phenotype complexity.

The phage host range GWAS could be improved by increasing strain scope and considering more phages. “Only” 263 genome-sequenced strains were tested, though this selection was quite diverse (at least 10 clonal complexes). The power of the GWAS would benefit from both more strains covering common CCs (e.g., CC5 and CC8) and more strains covering rare, diverse CCs (e.g., CC707). In order to test 263 strains with 8 phages I developed a high-throughput lab assay. However, this was still a tremendous effort in the laboratory. For this reason, it is likely that to extend GWAS, automation of phenotypic assays will be needed. Because there appear to be multiple, independent, metabolic causes (e.g., alterations to amino acid metabolism, nucleic acid biosynthesis, or translation) of virulent phage (especially *Myoviridae*, as receptor genes strongly explained the tested *Podoviridae* phage’s host range), our GWAS would benefit from more strains to capture rare strain-specific variants still possibly constrained by clade. We still have not considered the epistatic effects of strain background on emergence of rare resistant strains. In addition, only eight phages were phenotyped, though these represent the three morphological classes of *S. aureus* phages (*Siphoviridae*, *Myoviridae*, and *Podoviridae*). However, it has been observed that phages can encode host range factors such as anti-restriction proteins (227), and phage polymorphism can directly influence host range through elimination of

restriction (375) and CRISPR target sites (376), for example. It thus would be beneficial to test more representatives of each class to capture phage-specific host range factors as well as the consequences for host range determination on the host side as well. A phage immune to restriction, for example, might expose novel resistance mechanisms in the strains that remain resistant to it.

In chapter 4, I examined the evolutionary patterns of the known *S. aureus* phage resistance genes and their impacts on previously measured phage resistance and horizontal gene transfer. I performed a survey of 331 curated phage resistance or host range genes in *S. aureus* using our database of 43,000 annotated genome sequences covering the species. This curated list, which is based on the results of work in chapters 2 and 3, is a valuable open resource for future research efforts on phage resistance. I found adsorption genes to be the most conserved in the database, followed by assembly and biosynthesis genes. Core phage resistance genes were as functional and diverse as all core genes, and core phage resistance genes were under negative selection like core genes overall. Only superinfection immunity correlated with (temperate) phage resistance phenotype, accessory genome content, and number of virulence genes, while there was no relationship between overall phage resistance gene presence/absence and virulent phage resistance phenotype. All genes exhibited some level of phylogenetic signal, but it was weakest amongst the assembly genes.

The bioinformatic survey could be further improved through further examination of rare defense systems, possible separate treatment of neutral or phage sensitivity genes, and extending the same analyses to genes or mutations found in the host range GWAS. The survey results suggest phage resistance gene presence or absence does not predict measured phage resistance nor horizontal gene transfer, with the exception of superinfection immunity. Two alternative approaches could respond to this paradoxical result. The first is ignoring genes expected to always have neutral (e.g., for biosynthesizing type 5 and 8 capsule, which don't correlate directly with phage resistance in prior work) or sensitivity (e.g., for biosynthesizing wall

teichoic acid, which is the phage receptor, and thus promotes phage infection) effects and instead focusing on less conserved, often mobile, phage defense systems. The second is to repeat the bioinformatic survey instead on genes and mutations implicated in host range from the previous GWAS, with the limitation that host range factors would not be expected to also explain transduction efficiency or level of horizontal gene transfer.

In my final thesis chapter, I developed a nanopore assay to detect phenotypes based on mutations called in sequenced amplicons, which could be applied to a variety of contexts, such as detecting phage resistance point mutations. The assay detected vancomycin-intermediate *S. aureus* based on sequencing amplicons containing 16 genes often implicated in such resistance, and then calling causative mutations in the sequenced amplicons. I found that I could sequence full amplicons with high (>1000x) coverage using nanopore sequencing. I also found that it is possible to discriminate a SNP from its parental sequence in separate pools or even the same pool. I then calibrated the assay with SNP/parent mixtures to determine the limit of detection (SNP fraction). I found that the limit of detection was close to minor allele frequency average (~5-7%) regardless of the type of mixture (cultures, DNA amplicons, or simulated reads). I also found that I could detect VISA-causing mutations in clinical strains and discriminate clinical VSSA from VISA strains.

Further research work on the VISA nanopore amplicon assay would include read downsampling to see how few reads it takes to call a mutation and a VISA database to comprehensively detect VISA in unknown strains. We would select read subsets of progressively lower numbers until we could no longer call the VISA mutation in order to determine the minimum coverage, and thus better understand the minimum time necessary for mutation detection. We also need to build a VISA mutation database to make mutation detection comprehensive and effective against diverse clinical isolates. This database will be built from 1) lab-identified VISA mutations, 2) clinical VISA-associated mutations, 3) all nonsynonymous mutations in VISA-

associated genes found in Staphopia, and 4) all possible nonsynonymous mutations in VISA-associated genes expected to confer a VISA phenotype.

Future work would tailor this nanopore assay to detecting phage host range determinants instead, but our GWAS suggests many more strains must be tested to identify loci most commonly associated with resistance. Nonetheless, despite some issues with higher error rates than Illumina sequencing, nanopore sequencing has great potential for both detecting antibiotic resistance and personalizing phage therapy. Already studies have shown that nanopore sequencing can detect antibiotic resistance gene presence or absence very quickly (160), and ours indicates we can call resistance-causing mutations from such data. If we can identify the set of host loci responsible for phage host range determination, we could develop a comparable host range nanopore amplicon assay to rapidly determine which phages best killed a strain causing infection, thus personalizing phage therapy and improving its efficacy.

Future research directions building off the research performed in this dissertation

There are numerous future projects that would answer outstanding questions about *S. aureus* phages and their host ranges that are a natural extension of this thesis. The ecology of *S. aureus* phages remains largely unexplored, as their natural reservoirs and population dynamics are unknown. For these reasons I suggest that the following projects (explained in greater detail in the following paragraphs) would be the “next steps”. 1) a laboratory phage discovery project; 2) phage survey in existing sequenced metagenomes to better understand where *S. aureus* phages are found and why lytic phages have not killed off *S. aureus*, given that resistance is so rare; 3) prophages also remain poorly characterized despite their roles in horizontal gene transfer, so I propose a bioinformatic survey of prophage functionality and diversity with our comprehensive *S. aureus* genome database; 4) two further GWAS studies to resolve the differences between host range and transduction determinants, followed by construction of predictive models for host

range and transduction phenotypes once we have a larger strain sample size; and 5) additional studies to evolve phages with expanded host ranges and examine tradeoffs between bacterial within-host evolution and phage resistance.

The small number of phages used in the GWAS inspires the laboratory phage discovery project, which is intended to answer questions about *S. aureus* phage ecology (where are *S. aureus* phages present? Why can we find *S. aureus* lytic phages if corresponding resistance evolution is so difficult?). The phage host GWAS only relied on eight phages belonging to the three major classes of *S. aureus* phages. New phages would thus be isolated by enrichment on a permissive *S. aureus* host such as the restriction-free, prophage-cured RN4220 (189). Phages would be isolated from samples previously reported to be rich in *S. aureus* phages, such as sewage, human skin swabs, nasal swabs, and farm animals (chicken excrement and cow manure; chicken, pig, and cow swabs).

A related phage discovery project (study #2) would be bioinformatic instead - a metagenomic survey of *S. aureus* phages and corresponding resistance/host range factors. The goal of this work is to determine the diversity of *S. aureus* phages present in the natural world (like study #1) and to understand the relationship between *S. aureus* phage abundance and phage resistance in natural environments. A preliminary study I performed examining phage matches to metagenomes based on k-mers suggest lytic phages are very rare. This suggests that it will be necessary to focus on the temperate phages, which will be complicated by integrated prophages. We would need to develop a method to differentiate integrated prophages from free temperate phage based on either identifying reads that overlap the phage integration (*attB*/*P*) site or comparing levels of read mapping coverage between reference prophages and surrounding genome. In the latter case, we would expect excess coverage to represent free phage, while in the former case, we would compare coverage of the “closed” *attB*/*P* site (free phage) or genome-prophage boundaries (integrated phage) to total phage coverage to determine free phage levels.

We would expect free phage abundance to be inversely proportional to the level of phage resistance expected in the metagenome based on phage resistance genes.

While the previous two projects serve to identify new *S. aureus* phages in the environment, we also need to explore the diversity of prophages in *S. aureus* given their implication in transduction and other consequences of their induction (e.g., killing the host). We would thus conduct a bioinformatic survey of the prophage diversity spectrum in *S. aureus* species using the Staphopia database of 40,000+ annotated genomes, as indeed not much is known (only a small set of strains has been explored, containing 0-4 prophages). It would be worth systematically categorizing complete and cryptic phages in *S. aureus,* given findings in the general phage resistance bioinformatic survey and the potential roles in transduction. We would classify all detected prophages most likely by integrase type. We would also use the delta-bit score approach (i.e., mapping detected and reference genes against profile Hidden Markov models and calculating the difference as the delta-bit score, a proxy for functionality of the detected gene) and possibly just the number of prophage-encoded genes to assess the level of genome degradation in putatively non-functional cryptic prophages. We would finally evaluate the levels of accessory genome associated with prophages and compare accessory genome content amongst particular prophage classes.

The next two projects would be genome-wide association studies on horizontal gene transfer and transduction as measured in the lab, as we expect these determinants to be different from those responsible for phage host range. These would putatively be different from host range determinants found in the previous GWAS because transduction requires 1) donor infection and 2) recipient abortive infection followed by survival while host range inclusion requires infection and killing of the infected strain. We would thus perform a GWAS on accessory genome related determinants, especially in core and in intermediate-frequency (near core) genes. This would serve to identify potential host determinants of transduction, but could also just be confounded by

strains with higher transducing phage exposure over time or higher likelihood to exist in a transduction-favorable location.

We would also conduct a second GWAS on experimental transduction frequency in addition to the previous accessory genome GWAS. This would especially be difficult because transduction assays rely on mutual, complementary antibiotic resistance markers to find transductants. A nanopore transduction assay could improve transductant detection by removing the need for markers, but it would also be far more expensive. Autotransduction (the ability of a strain's own prophage to infect a second strain, take up bacterial DNA, and transduce the original lysogen) (40) and/or transposon screening (or TnSeq) could be alternative approaches requiring fewer markers. Nonetheless, both this GWAS and the previous study are necessary to differentiate host range from transduction determinants. Depending on how transduction is performed, the study may rediscover previous findings of lateral transduction, in which pac-type prophages enhance transduction frequency, but in a broader population context.

If these two and the host range GWAS cover enough strains to saturate host range and transduction determinants, we could develop predictive models for transduction and phage host range. These models could rely on random forests, gradient-boosted decision trees, and neural networks to predict both phenotypes based on a comprehensive set of core and intermediate-frequency determinants. In order for this to have high power, many more strains must be tested. Currently, at least for phage host range, not enough strains have been represented to develop models explaining more than 75% of the variation in the phenotype.

The final proposed projects would cover tradeoffs with within-host evolution and phage evolution to alter host range. Within-host evolution effects on phage resistance (trade-offs or collateral resistance) may be examined as part of collaboration of the Read lab with Dr. Michael David at the University of Pennsylvania to examine within-host evolution of MRSA in different patient subjects. We may simply examine phage host range via spot or high-throughput assays and resistance through the high-throughput or efficiency of plating (EOP) assays for descendant

strains after different stages of evolution. We would expect phage resistance to increase depending on localization in the human body, as strains would likely overproduce surface protein A to evade antibodies, which consequently would block the phage receptor from the surface. Maintenance of wall teichoic acid (WTA), on the other hand, due to its roles in nasal colonization (261) and innate immune resistance (49), would be expected to maintain phage sensitivity.

In the final proposed project, I would evolve phages with broader host ranges or better lytic properties via Applemans protocol. As has been done with *Pseudomonas aeruginosa* (377, 378), multiple *S. aureus* phages could be mixed in a cocktail and tested amongst sensitive and resistant strains. The cocktail would be diluted in a 10-fold series and co-cultured with different strains in a 96-well plate. The minimum phage dilution killing each strain and the previous dilution (not killing the strain) would be collected for all tested strains and mixed together to propagate the new phage mixture on a mixture of strains. Then the successive lysate would be diluted again and co-cultured as previous for up to 30 transfers. At each step, the lysate would also be titered on sensitive bacterial lawns to select new, mutant, recombinant phages with altered host ranges (and possibly dynamic properties - burst size and latent period). We would then isolate evolved phages from each evolutionary step, evaluate their new host ranges, and sequence their genomes to detect mutations and/or recombination.

Because most *S. aureus* phages already have broad host ranges, we may see instead enhanced lytic properties to overcome the few resistant strains that exist. On the other hand, we may also see signatures of phage evasion of defense systems such as mutated or recombined restriction sites. We may also be able to use these results to ask questions about how evolution affects the phage specialist-generalist continuum. We may expect generalist phages to have broader host ranges and weaker dynamic lytic properties while specialist phages would have narrower host ranges but stronger dynamic properties amongst the few strains they did kill.

Summary

Overall, this dissertation presents many new insights regarding phage-host interactions in *S. aureus*, especially regarding the paradox of broad host range. It has been exciting to work on *S. aureus* phages because so many questions in the field still remain unanswered and virtually nothing had been addressed on the population-level. I have been able to quickly and inexpensively develop techniques to evaluate host range at a large scale (many strains) as well as use genomics to answer such questions at nucleotide and species-level scales. There are most certainly many more questions to be addressed in the field as the second half of this section describes. Certainly the most important future efforts for the field are predicting host range in therapeutic contexts, standardizing lysate preparation, and better understanding how phages shape species evolution through transduction and phage resistance selection.

Chapter 7: Bibliography

1. Kluytmans J, Belkum A van, Verbrugh H. 1997. Nasal carriage of Staphylococcus aureus: epidemiology, underlying mechanisms, and associated risks. Clin Microbiol Rev 10:505–520.
2. Al EK et. Hospitalizations and Deaths Caused by Methicillin-Resistant Staphylococcus aureus, United States, 1999–2005 - Volume 13, Number 12—December 2007 - Emerging Infectious Disease journal - CDC https://doi.org/10.3201/eid1312.070629.
3. Foster TJ. 2017. Antibiotic resistance in Staphylococcus aureus. Current status and future prospects. FEMS Microbiology Reviews 41:430–449.
4. Bagnoli F, Bertholet S, Grandi G. 2012. Inferring Reasons for the Failure of Staphylococcus aureus Vaccines in Clinical Trials. Front Cell Infect Microbiol 2.
5. Breed RS. 1956. Staphylococcus pyogenes Rosenbach. International Journal of Systematic and Evolutionary Microbiology, 6:35–42.
6. Rosenbach AJF. 1884. Mikro-organismen bei den Wund-Infections-Krankheiten des Menschen. J.F. Bergmann.
7. Licitra G. 2013. Etymologia: Staphylococcus. Emerg Infect Dis 19:1553.
8. Orenstein A. The Discovery and Naming of Staphylococcus aureus 2.
9. Foster T. 1996. Staphylococcus, p. . *In* Baron, S (ed.), Medical Microbiology, 4th ed. University of Texas Medical Branch at Galveston, Galveston (TX).
10. Kaito C, Sekimizu K. 2007. Colony Spreading in Staphylococcus aureus. J Bacteriol 189:2553–2557.
11. Pollitt EJG, Diggle SP. 2017. Defining motility in the Staphylococci. Cell Mol Life Sci 74:2943–2958.
12. Proctor RA, Kriegeskorte A, Kahl BC, Becker K, Löffler B, Peters G. 2014. Staphylococcus aureus Small Colony Variants (SCVs): a road map for the metabolic

pathways involved in persistent infections. Front Cell Infect Microbiol 4.

13. Zhang H, Zheng Y, Gao H, Xu P, Wang M, Li A, Miao M, Xie X, Deng Y, Zhou H, Du H. 2016. Identification and Characterization of Staphylococcus aureus Strains with an Incomplete Hemolytic Phenotype. Front Cell Infect Microbiol 6.
14. Shields P, Tsang AY. 2006. Mannitol Salt Agar Plates Protocols. Text.
15. Matias VRF, Beveridge TJ. 2006. Native Cell Wall Organization Shown by Cryo-Electron Microscopy Confirms the Existence of a Periplasmic Space in Staphylococcus aureus. Journal of Bacteriology 188:1011–1021.
16. Dmitriev BA, Toukach FV, Holst O, Rietschel ET, Ehlers S. 2004. Tertiary Structure of Staphylococcus aureus Cell Wall Murein. J Bacteriol 186:7141–7148.
17. Mazmanian SK, Liu G, Ton-That H, Schneewind O. 1999. Staphylococcus aureus Sortase, an Enzyme that Anchors Surface Proteins to the Cell Wall. Science 285:760–763.
18. Xia G, Kohler T, Peschel A. 2010. The wall teichoic acid and lipoteichoic acid polymers of Staphylococcus aureus. International Journal of Medical Microbiology 300:148–154.
19. Cafiso V, Bertuccio T, Purrello S, Campanile F, Mammina C, Sartor A, Raglio A, Stefani S. 2014. dltA overexpression: A strain-independent keystone of daptomycin resistance in methicillin-resistant Staphylococcus aureus. International Journal of Antimicrobial Agents 43:26–31.
20. Xia G, Maier L, Sanchez-Carballo P, Li M, Otto M, Holst O, Peschel A. 2010. Glycosylation of wall teichoic acid in Staphylococcus aureus by TarM. J Biol Chem 285:13405–13415.
21. O'Riordan K, Lee JC. 2004. Staphylococcus aureus Capsular Polysaccharides. Clin Microbiol Rev 17:218–234.
22. Wilkinson BJ, Holmes KM. 1979. Staphylococcus aureus cell surface: capsule as a barrier to bacteriophage adsorption. Infect Immun 23:549–552.

23. Enright MC, Day NPJ, Davies CE, Peacock SJ, Spratt BG. 2000. Multilocus Sequence Typing for Characterization of Methicillin-Resistant and Methicillin-Susceptible Clones ofStaphylococcus aureus. Journal of Clinical Microbiology 38:1008–1015.

24. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res 3.

25. Planet PJ, Narechania A, Chen L, Mathema B, Boundy S, Archer G, Kreiswirth B. 2017. Architecture of a Species: Phylogenomics of Staphylococcus aureus. Trends in Microbiology 25:153–166.

26. Turner NA, Sharma-Kuinkel BK, Maskarinec SA, Eichenberger EM, Shah PP, Carugati M, Holland TL, Fowler VG. 2019. Methicillin-resistant Staphylococcus aureus : an overview of basic and clinical research. Nature Reviews Microbiology 1.

27. Lindsay JA. 2014. Staphylococcus aureus genomics and the impact of horizontal gene transfer. International Journal of Medical Microbiology 304:103–109.

28. Morikawa K, Takemura AJ, Inose Y, Tsai M, Thi LTN, Ohta T, Msadek T. 2012. Expression of a Cryptic Secondary Sigma Factor Gene Unveils Natural Competence for DNA Transformation in Staphylococcus aureus. PLOS Pathogens 8:e1003003.

29. Noble WC, Virani Z, Cree RG. 1992. Co-transfer of vancomycin and other resistance genes from Enterococcus faecalis NCTC 12201 to Staphylococcus aureus. FEMS Microbiol Lett 72:195–198.

30. 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.

31. 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. N Engl J Med 348:1342–1347.

32. O'Brien FG, Ramsay JP, Monecke S, Coombs GW, Robinson OJ, Htet Z, Alshaikh FAM, Grubb WB. 2015. Staphylococcus aureus plasmids without mobilization genes are mobilized by a novel conjugative plasmid from community isolates. Journal of Antimicrobial Chemotherapy 70:649–652.

33. Rossi F, Diaz L, Wollam A, Panesso D, Zhou Y, Rincon S, Narechania A, Xing G, Di Gioia TSR, Doi A, Tran TT, Reyes J, Munita JM, Carvajal LP, Hernandez-Roldan A, Brandão D, van der Heijden IM, Murray BE, Planet PJ, Weinstock GM, Arias CA. 2014. Transferable Vancomycin Resistance in a Community-Associated MRSA Lineage. New England Journal of Medicine 370:1524–1531.

34. Sansevere EA, Robinson DA. 2017. Staphylococci on ICE: Overlooked agents of horizontal gene transfer. Mobile Genetic Elements 7:1–10.

35. Botelho AMN, Cerqueira e Costa MO, Moustafa AM, Beltrame CO, Ferreira FA, Côrtes MF, Costa BSS, Silva DNS, Bandeira PT, Lima NCB, Souza RC, Almeida LGP de, Vasconcelos ATR, Narechania A, Ryan C, O'Brien K, Kolokotronis S-O, Planet PJ, Nicolás MF, Figueiredo AMS. 2019. Local Diversification of Methicillin- Resistant Staphylococcus aureus ST239 in South America After Its Rapid Worldwide Dissemination. Front Microbiol 10.

36. Novick RP, Ram G. 2017. Staphylococcal pathogenicity islands – movers and shakers in the genomic firmament. Curr Opin Microbiol 38:197–204.

37. Chen J, Quiles-Puchalt N, Chiang YN, Bacigalupe R, Fillol-Salom A, Chee MSJ, Fitzgerald JR, Penadés JR. 2018. Genome hypermobility by lateral transduction. Science 362:207–212.

38. Olson ME. 2016. Bacteriophage Transduction in Staphylococcus aureus, p. 69–74. *In* Bose, JL (ed.), The Genetic Manipulation of Staphylococci: Methods and Protocols.

Springer, New York, NY.

39. Xia G, Wolz C. 2014. Phages of Staphylococcus aureus and their impact on host evolution. Infection, Genetics and Evolution 21:593–601.
40. Haaber J, Leisner JJ, Cohn MT, Catalan-Moreno A, Nielsen JB, Westh H, Penadés JR, Ingmer H. 2016. Bacterial viruses enable their host to acquire antibiotic resistance genes from neighbouring cells. 1. Nature Communications 7:13333.
41. Hiramatsu K, Katayama Y, Matsuo M, Sasaki T, Morimoto Y, Sekiguchi A, Baba T. 2014. Multi-drug-resistant Staphylococcus aureus and future chemotherapy. Journal of Infection and Chemotherapy 20:593–601.
42. Harkins CP, Pichon B, Doumith M, Parkhill J, Westh H, Tomasz A, de Lencastre H, Bentley SD, Kearns AM, Holden MTG. 2017. Methicillin-resistant Staphylococcus aureus emerged long before the introduction of methicillin into clinical practice. Genome Biol 18.
43. Vitko NP, Richardson AR. 2013. Laboratory Maintenance of Methicillin-Resistant Staphylococcus aureus (MRSA). Curr Protoc Microbiol 0 9:Unit-9C.2.
44. Jensen SO, Lyon BR. 2009. Genetics of antimicrobial resistance in Staphylococcus aureus. Future Microbiology 4:565–582.
45. McGuinness WA, Malachowa N, DeLeo FR. 2017. Vancomycin Resistance in Staphylococcus aureus. Yale J Biol Med 90:269–281.
46. 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. Clin Microbiol Rev 23:99–139.
47. Bayer AS, Mishra NN, Cheung AL, Rubio A, Yang S-J. 2016. Dysregulation of mprF and dltABCD expression among daptomycin-non-susceptible MRSA clinical isolates. J Antimicrob Chemother 71:2100–2104.
48. Mechler L, Bonetti E-J, Reichert S, Flötenmeyer M, Schrenzel J, Bertram R, François P,

Götz F. 2016. Daptomycin Tolerance in the Staphylococcus aureus pitA6 Mutant Is Due to Upregulation of the dlt Operon. Antimicrob Agents Chemother 60:2684–2691.

49. Peschel A, Otto M, Jack RW, Kalbacher H, Jung G, Götz F. 1999. Inactivation of the dlt Operon inStaphylococcus aureus Confers Sensitivity to Defensins, Protegrins, and Other Antimicrobial Peptides. J Biol Chem 274:8405–8410.
50. Eliopoulos GM, Huovinen P. 2001. Resistance to Trimethoprim-Sulfamethoxazole. Clin Infect Dis 32:1608–1614.
51. O'Neill AJ, Huovinen T, Fishwick CWG, Chopra I. 2006. Molecular Genetic and Structural Modeling Studies of Staphylococcus aureus RNA Polymerase and the Fitness of Rifampin Resistance Genotypes in Relation to Clinical Prevalence. Antimicrobial Agents and Chemotherapy 50:298–309.
52. Cui L, Isii T, Fukuda M, Ochiai T, Neoh H, 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.
53. Moller AG, Lindsay JA, Read TD. 2019. Determinants of Phage Host Range in Staphylococcus Species. Appl Environ Microbiol 85:e00209-19.
54. Hendrix RW. 2002. Bacteriophages: Evolution of the Majority. Theoretical Population Biology 61:471–480.
55. Abedon ST. 2006. The Bacteriophages. Oxford University Press, USA.
56. Bertozzi Silva J, Storms Z, Sauvageau D. 2016. Host receptors for bacteriophage adsorption. FEMS Microbiology Letters 363.
57. Heller K, Braun V. 1982. Polymannose O-antigens of Escherichia coli, the binding sites for the reversible adsorption of bacteriophage T5+ via the L-shaped tail fibers. J Virol 41:222–227.
58. Mondigler M, Holz T, Heller KJ. 1996. Identification of the Receptor-Binding Regions of

pb5 Proteins of Bacteriophages T5 and BF23. Virology 219:19–28.

59. Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. Nat Rev Microbiol 8:317–327.

60. Seed KD. 2015. Battling Phages: How Bacteria Defend against Viral Attack. PLOS Pathogens 11:e1004847.

61. Moak M, Molineux IJ. 2004. Peptidoglycan hydrolytic activities associated with bacteriophage virions. Molecular Microbiology 51:1169–1183.

62. Rodríguez-Rubio L, Martínez B, Donovan DM, Rodríguez A, García P. 2013. Bacteriophage virion-associated peptidoglycan hydrolases: potential new enzybiotics. Critical Reviews in Microbiology 39:427–434.

63. Latka A, Maciejewska B, Majkowska-Skrobek G, Briers Y, Drulis-Kawa Z. 2017. Bacteriophage-encoded virion-associated enzymes to overcome the carbohydrate barriers during the infection process. Appl Microbiol Biotechnol 101:3103–3119.

64. Kimura K, Itoh Y. 2003. Characterization of Poly-γ-Glutamate Hydrolase Encoded by a Bacteriophage Genome: Possible Role in Phage Infection of Bacillus subtilis Encapsulated with Poly-γ-Glutamate. Appl Environ Microbiol 69:2491–2497.

65. Mamberti S, Prati P, Cremaschi P, Seppi C, Morelli CF, Galizzi A, Fabbi M, Calvio C. 2015. γ-PGA Hydrolases of Phage Origin in Bacillus subtilis and Other Microbial Genomes. PLoS One 10.

66. Ozaki T, Abe N, Kimura K, Suzuki A, Kaneko J. 2017. Genomic analysis of Bacillus subtilis lytic bacteriophage ϕNIT1 capable of obstructing natto fermentation carrying genes for the capsule-lytic soluble enzymes poly-γ-glutamate hydrolase and levanase. Bioscience, Biotechnology, and Biochemistry 81:135–146.

67. Inamdar MM, Gelbart WM, Phillips R. 2006. Dynamics of DNA Ejection from Bacteriophage. Biophysical Journal 91:411–420.

68. Mahony J, McGrath S, Fitzgerald GF, van Sinderen D. 2008. Identification and

Characterization of Lactococcal-Prophage-Carried Superinfection Exclusion Genes. Appl Environ Microbiol 74:6206–6215.

69. Weigel C, Seitz H. 2006. Bacteriophage replication modules. FEMS Microbiology Reviews 30:321–381.

70. Wilson GG, Murray NE. 1991. Restriction and Modification Systems. Annual Review of Genetics 25:585–627.

71. Lopatina A, Tal N, Sorek R. 2020. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. Annu Rev Virol 7:371–384.

72. Berngruber TW, Weissing FJ, Gandon S. 2010. Inhibition of Superinfection and the Evolution of Viral Latency. J Virol 84:10200–10208.

73. Hyman P, Abedon ST. 2010. Chapter 7 - Bacteriophage Host Range and Bacterial Resistance, p. 217–248. *In* Advances in Applied Microbiology. Academic Press.

74. Barrangou R, Marraffini LA. 2014. CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity. Molecular Cell 54:234–244.

75. Horvath P, Barrangou R. 2010. CRISPR/Cas, the Immune System of Bacteria and Archaea. Science 327:167–170.

76. Sorek R, Kunin V, Hugenholtz P. 2008. CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Micro 6:181–186.

77. Aksyuk AA, Rossmann MG. 2011. Bacteriophage Assembly. Viruses 3:172–203.

78. Penadés JR, Chen J, Quiles-Puchalt N, Carpena N, Novick RP. 2015. Bacteriophage-mediated spread of bacterial virulence genes. Current Opinion in Microbiology 23:171–178.

79. Novick RP, Christie GE, Penadés JR. 2010. The phage-related chromosomal islands of Gram-positive bacteria. Nat Rev Microbiol 8:541–551.

80. Cahill J, Young R. 2019. Chapter Two - Phage Lysis: Multiple Genes for Multiple Barriers, p. 33–70. *In* Kielian, M, Mettenleiter, TC, Roossinck, MJ (eds.), Advances in Virus

Research. Academic Press.

81. Kongari R, Rajaure M, Cahill J, Rasche E, Mijalis E, Berry J, Young R. 2018. Phage spanins: diversity, topological dynamics and gene convergence. BMC Bioinformatics 19:326.

82. Casjens SR, Hendrix RW. 2015. Bacteriophage lambda: early pioneer and still relevant. Virology 0:310–330.

83. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. 2017. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. 7. The ISME Journal 11:1511–1520.

84. Little JW. 2005. Lysogeny, Prophage Induction, and Lysogenic Conversion, p. 37–54. *In* Phages. John Wiley & Sons, Ltd.

85. Moineau S, Durmaz E, Pandian S, Klaenhammer TR. 1993. Differentiation of Two Abortive Mechanisms by Using Monoclonal Antibodies Directed toward Lactococcal Bacteriophage Capsid Proteins. Appl Environ Microbiol 59:208–212.

86. Middelboe M. 2016. One-step growth experiments (bacteriophages) https://doi.org/10.17504/protocols.io.dpw5pd.

87. Willey JM, Sherwood L, Prescott LM, Woolverton CJ. 2008. Prescott, Harley, and Klein's Microbiology. McGraw-Hill Higher Education.

88. Lenski RE. 1988. Dynamics of Interactions between Bacteria and Virulent Bacteriophage, p. 1–44. *In* Marshall, KC (ed.), Advances in Microbial Ecology. Springer US, Boston, MA.

89. Sulakvelidze A, Alavidze Z, Morris JG. 2001. Bacteriophage Therapy. Antimicrob Agents Chemother 45:649–659.

90. Stone R. 2002. Stalin's Forgotten Cure. Science 298:728–731.

91. Bradbury J. 2004. "My enemy's enemy is my friend." The Lancet 363:624–625.

92. Kutter E, De Vos D, Gvasalia G, Alavidze Z, Gogokhia L, Kuhl S, Abedon ST. 2010. Phage Therapy in Clinical Practice: Treatment of Human Infections. Current

Pharmaceutical Biotechnology 11:69–86.

93. Kutateladze M, Adamia R. 2010. Bacteriophages as potential new therapeutics to replace or supplement antibiotics. Trends in Biotechnology 28:591–595.

94. Altamirano FLG, Barr JJ. 2019. Phage Therapy in the Postantibiotic Era. Clinical Microbiology Reviews 32.

95. Lu TK, Koeris MS. 2011. The next generation of bacteriophage therapy. Current Opinion in Microbiology 14:524–531.

96. Nobrega FL, Costa AR, Kluskens LD, Azeredo J. 2015. Revisiting phage therapy: new applications for old resources. Trends in Microbiology 23:185–191.

97. Ly-Chatain MH. 2014. The factors affecting effectiveness of treatment in phages therapy. Front Microbiol 5.

98. Gutiérrez D, Rodríguez-Rubio L, Fernández L, Martínez B, Rodríguez A, García P. 2017. Applicability of commercial phage-based products against Listeria monocytogenes for improvement of food safety in Spanish dry-cured ham and food contact surfaces. Food Control 73:1474–1482.

99. Kutter EM, Kuhl SJ, Abedon ST. 2015. Re-establishing a place for phage therapy in western medicine. Future Microbiology 10:685–688.

100. Loc-Carrillo C, Abedon ST. 2011. Pros and cons of phage therapy. Bacteriophage 1:111–114.

101. Deghorain M, Van Melderen L. 2012. The Staphylococci Phages Family: An Overview. Viruses 4:3316–3335.

102. Alves DR, Gaudion A, Bean JE, Esteban PP, Arnot TC, Harper DR, Kot W, Hansen LH, Enright MC, Jenkins ATA. 2014. Combined Use of Bacteriophage K and a Novel Bacteriophage To Reduce Staphylococcus aureus Biofilm Formation. Appl Environ Microbiol 80:6694–6703.

103. Hsieh S-E, Lo H-H, Chen S-T, Lee M-C, Tseng Y-H. 2011. Wide Host Range and Strong

Lytic Activity of Staphylococcus aureus Lytic Phage Stau2. Appl Environ Microbiol 77:756–761.

104. Son J-S, Lee S-J, Jun SY, Yoon SJ, Kang SH, Paik HR, Kang JO, Choi Y-J. 2010. Antibacterial and biofilm removal activity of a podoviridae Staphylococcus aureus bacteriophage SAP-2 and a derived recombinant cell-wall-degrading enzyme. Appl Microbiol Biotechnol 86:1439–1449.
105. Cha Y, Chun J, Son B, Ryu S. 2019. Characterization and Genome Analysis of Staphylococcus aureus Podovirus CSA13 and Its Anti-Biofilm Capacity. Viruses 11:54.
106. Li X, Hu T, Wei J, He Y, Abdalla AE, Wang G, Li Y, Teng T. 2021. Characterization of a Novel Bacteriophage Henu2 and Evaluation of the Synergistic Antibacterial Activity of Phage-Antibiotics. Antibiotics (Basel) 10.
107. Zschach H, Larsen MV, Hasman H, Westh H, Nielsen M, Międzybrodzki R, Jończyk-Matysiak E, Weber-Dąbrowska B, Górski A. 2018. Use of a Regression Model to Study Host-Genomic Determinants of Phage Susceptibility in MRSA. Antibiotics (Basel) 7.
108. Winstel V, Sanchez-Carballo P, Holst O, Xia G, Peschel A. 2014. Biosynthesis of the Unique Wall Teichoic Acid of Staphylococcus aureus Lineage ST395. mBio 5:e00869-14.
109. Li X, Gerlach D, Du X, Larsen J, Stegger M, Kühner P, Peschel A, Xia G, Winstel V. 2015. An accessory wall teichoic acid glycosyltransferase protects Staphylococcus aureus from the lytic activity of Podoviridae. Sci Rep 5.
110. Gerlach D, Guo Y, Castro CD, Kim S-H, Schlatterer K, Xu F-F, Pereira C, Seeberger PH, Ali S, Codée J, Sirisarn W, Schulte B, Wolz C, Larsen J, Molinaro A, Lee BL, Xia G, Stehle T, Peschel A. 2018. Methicillin-resistant Staphylococcus aureus alters cell wall glycosylation to evade immunity. Nature 1.
111. Nordström K, Forsgren A. 1974. Effect of Protein A on Adsorption of Bacteriophages to Staphylococcus aureus. J Virol 14:198–202.
112. Waldron DE, Lindsay JA. 2006. Sau1: a Novel Lineage-Specific Type I Restriction-

Modification System That Blocks Horizontal Gene Transfer into Staphylococcus aureus and between S. aureus Isolates of Different Lineages. J Bacteriol 188:5578–5585.

113. Roberts GA, Houston PJ, White JH, Chen K, Stephanou AS, Cooper LP, Dryden DTF, Lindsay JA. 2013. Impact of target site distribution for Type I restriction enzymes on the evolution of methicillin-resistant Staphylococcus aureus (MRSA) populations. Nucleic Acids Res 41:7472–7484.

114. Dempsey RM. 2005. Sau42I, a BcgI-like restriction-modification system encoded by the Staphylococcus aureus quadruple-converting phage 42. Microbiology 151:1301–1311.

115. Xu S, Corvaglia AR, Chan S-H, Zheng Y, Linder P. 2011. A type IV modification-dependent restriction enzyme SauUSI from Staphylococcus aureus subsp. aureus USA300. Nucleic Acids Res 39:5597–5610.

116. Depardieu F, Didier J-P, Bernheim A, Sherlock A, Molina H, Duclos B, Bikard D. 2016. A Eukaryotic-like Serine/Threonine Kinase Protects Staphylococci against Phages. Cell Host & Microbe 20:471–481.

117. Damle PK, Wall EA, Spilman MS, Dearborn AD, Ram G, Novick RP, Dokland T, Christie GE. 2012. The roles of SaPI1 proteins gp7 (CpmA) and gp6 (CpmB) in capsid size determination and helper phage interference. Virology 432:277–282.

118. Ram G, Chen J, Kumar K, Ross HF, Ubeda C, Damle PK, Lane KD, Penadés JR, Christie GE, Novick RP. 2012. Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. PNAS 109:16300–16305.

119. Ram G, Chen J, Ross HF, Novick RP. 2014. Precisely modulated pathogenicity island interference with late phage gene transcription. Proc Natl Acad Sci U S A 111:14536–14541.

120. Winstel V, Liang C, Sanchez-Carballo P, Steglich M, Munar M, Bröker BM, Penadés JR, Nübel U, Holst O, Dandekar T, Peschel A, Xia G. 2013. Wall teichoic acid structure

governs horizontal gene transfer between major bacterial pathogens. Nature Communications 4:1–9.

121. Kaneko J, Narita-Yamada S, Wakabayashi Y, Kamio Y. 2009. Identification of ORF636 in Phage φSLT Carrying Panton-Valentine Leukocidin Genes, Acting as an Adhesion Protein for a Poly(Glycerophosphate) Chain of Lipoteichoic Acid on the Cell Surface of Staphylococcus aureus. J Bacteriol 191:4674–4680.
122. Archibald AR, Glassey K, Green RS, Lang WK. 1989. Cell wall composition and surface properties in Bacillus subtilis: anomalous effect of incubation temperature on the phage-binding properties of bacteria containing varied amounts of teichoic acid. J Gen Microbiol 135:667–673.
123. São-José C, Baptista C, Santos MA. 2004. Bacillus subtilis Operon Encoding a Membrane Receptor for Bacteriophage SPP1. J Bacteriol 186:8337–8346.
124. Levner MH. 1972. Replication of viral DNA in SPO1-infected Bacillus subtilis: II. DNA maturation during abortive infection. Virology 48:417–424.
125. Ito J, Spizizen J. 1971. Abortive Infection of Sporulating Bacillus subtilis 168 by φ2 Bacteriophage. J Virol 7:515–523.
126. Pritikin WB, Reiter H. 1969. Abortive Infection of Bacillus subtilis Bacteriophage PBS1 in the Presence of Actinomycin D. J Virol 3:578–585.
127. Johnson CM, Harden MM, Grossman AD. 2020. An integrative and conjugative element encodes an abortive infection system to protect host cells from predation by a bacteriophage. bioRxiv 2020.12.13.422588.
128. Kaźmierczak Z, Górski A, Dąbrowska K. 2014. Facing Antibiotic Resistance: Staphylococcus aureus Phages as a Medical Tool. 7. Viruses 6:2551–2570.
129. Eaton MD, Bayne-Jones S. 1934. BACTERIOPHAGE THERAPY: REVIEW OF THE PRINCIPLES AND RESULTS OF THE USE OF BACTERIOPHAGE IN THE TREATMENT OF INFECTIONS. JAMA 103:1847–1853.

130. Matsuzaki S, Yasuda M, Nishikawa H, Kuroda M, Ujihara T, Shuin T, Shen Y, Jin Z, Fujimoto S, Nasimuzzaman MD, Wakiguchi H, Sugihara S, Sugiura T, Koda S, Muraoka A, Imai S. 2003. Experimental Protection of Mice against Lethal Staphylococcus aureus Infection by Novel Bacteriophage ϕMR11. J Infect Dis 187:613–624.

131. Wills QF, Kerrigan C, Soothill JS. 2005. Experimental Bacteriophage Protection against Staphylococcus aureus Abscesses in a Rabbit Model. Antimicrobial Agents and Chemotherapy 49:1220–1221.

132. Verstappen KM, Tulinski P, Duim B, Fluit AC, Carney J, Nes A van, Wagenaar JA. 2016. The Effectiveness of Bacteriophages against Methicillin-Resistant Staphylococcus aureus ST398 Nasal Colonization in Pigs. PLOS ONE 11:e0160242.

133. Deamer D, Akeson M, Branton D. 2016. Three decades of nanopore sequencing. 5. Nature Biotechnology 34:518–524.

134. Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17:1–11.

135. Karamitros T, Magiorkinis G. 2018. Multiplexed Targeted Sequencing for Oxford Nanopore MinION: A Detailed Library Preparation Procedure, p. 43–51. *In* Head, SR, Ordoukhanian, P, Salomon, DR (eds.), Next Generation Sequencing: Methods and Protocols. Springer, New York, NY.

136. Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol 20:1–10.

137. Silvestre-Ryan J, Holmes I. 2021. Pair consensus decoding improves accuracy of neural network basecallers for nanopore sequencing. Genome Biology 22:38.

138. Edwards HS, Krishnakumar R, Sinha A, Bird SW, Patel KD, Bartsch MS. 2019. Real-Time Selective Sequencing with RUBRIC: Read Until with Basecall and Reference-Informed Criteria. 1. Scientific Reports 9:11475.

139. Loose M, Malla S, Stout M. 2016. Real-time selective sequencing using nanopore

technology. Nat Meth 13:751–754.

140. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736.

141. Rhoads A, Au KF. 2015. PacBio Sequencing and Its Applications. Genomics, Proteomics & Bioinformatics 13:278–289.

142. Mardis ER. 2017. DNA sequencing technologies: 2006–2016. 2. Nature Protocols 12:213–218.

143. Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A. 2015. The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. Front Genet 6.

144. Lee H, Gurtowski J, Yoo S, Nattestad M, Marcus S, Goodwin S, McCombie WR, Schatz MC. 2016. Third-generation sequencing and the future of genomics. bioRxiv 048603.

145. Oxford Nanopore launches Flongle, for rapid, smaller DNA/RNA sequencing tests in any environment. Oxford Nanopore Technologies.

146. Quick J, Quinlan AR, Loman NJ. 2014. A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer. GigaScience 3:22.

147. Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, Wain J, O'Grady J. 2015. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nature Biotechnology 33:296–300.

148. Szabó M, Nagy T, Wilk T, Farkas T, Hegyi A, Olasz F, Kiss J. 2016. Characterization of Two Multidrug-Resistant IncA/C Plasmids from the 1960s by Using the MinION Sequencer Device. Antimicrob Agents Chemother 60:6780–6786.

149. Faria NR, Sabino EC, Nunes MRT, Alcantara LCJ, Loman NJ, Pybus OG. 2016. Mobile real-time surveillance of Zika virus in Brazil. Genome Medicine 8:97.

150. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G,

Robles-Sikisaka R, Rogers TF, Beutler NA, Burton DR, Lewis-Ximenez LL, Jesus JG de, Giovanetti M, Hill S, Black A, Bedford T, Carroll MW, Nunes M, Alcantara LC, Sabino EC, Baylis SA, Faria N, Loose M, Simpson JT, Pybus OG, Andersen KG, Loman NJ. 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. bioRxiv 098913.

151. 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. 2017. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. J Antimicrob Chemother 72:104–114.
152. Votintseva AA, Bradley P, Pankhurst L, Elias C del O, 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. J Clin Microbiol JCM.02483-16.
153. Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC, McCombie WR. 2015. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome Res 25:1750–1756.
154. Stockton JD, Nieto T, Wroe E, Poles A, Inston N, Briggs D, Beggs AD. 2020. Rapid, highly accurate and cost-effective open-source simultaneous complete HLA typing and phasing of class I and II alleles using nanopore sequencing. HLA 96:163–178.
155. Girgis HS, DuPai CD, Lund J, Reeder J, Guillory J, Durinck S, Liang Y, Kaminker J, Smith PA, Skippington E. 2021. Single-molecule nanopore sequencing reveals extreme target copy number heterogeneity in arylomycin-resistant mutants. PNAS 118.
156. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, Suppes R, Feinstein D, Zanotti S, Taiberg L, Gurka D, Kumar A, Cheang M. 2006. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*. Critical Care Medicine 34:1589–1596.

157. Lemon JK, Khil PP, Frank KM, Dekker JP. 2017. Rapid Nanopore Sequencing of Plasmids and Resistance Gene Detection in Clinical Isolates. J Clin Microbiol 55:3530–3543.

158. Peter S, Bosio M, Gross C, Bezdan D, Gutierrez J, Oberhettinger P, Liese J, Vogel W, Dörfel D, Berger L, Marschal M, Willmann M, Gut I, Gut M, Autenrieth I, Ossowski S. 2020. Tracking of Antibiotic Resistance Transfer and Rapid Plasmid Evolution in a Hospital Setting by Nanopore Sequencing. mSphere 5.

159. Prussing C, Snavely EA, Singh N, Lapierre P, Lasek-Nesselquist E, Mitchell K, Haas W, Owsiak R, Nazarian E, Musser KA. 2020. Nanopore MinION Sequencing Reveals Possible Transfer of blaKPC–2 Plasmid Across Bacterial Species in Two Healthcare Facilities. Front Microbiol 11.

160. Břinda K, Callendrello A, Ma KC, MacFadden DR, Charalampous T, Lee RS, Cowley L, Wadsworth CB, Grad YH, Kucherov G, O'Grady J, Baym M, Hanage WP. 2020. Rapid inference of antibiotic resistance and susceptibility by genomic neighbour typing. 3. Nature Microbiology 5:455–464.

161. Sanderson ND, Swann J, Barker L, Kavanagh J, Hoosdally S, Crook D, Street TL, Eyre DW. 2020. High precision Neisseria gonorrhoeae variant and antimicrobial resistance calling from metagenomic Nanopore sequencing. Genome Res 30:1354–1363.

162. Ohno A, Umezawa K, Asai S, Kryukov K, Nakagawa S, Miyachi H, Imanishi T. 2021. Rapid profiling of drug-resistant bacteria using DNA-binding dyes and a nanopore-based DNA sequencer. 1. Scientific Reports 11:3436.

163. Niu H, Zhang W, Wei L, Liu M, Liu H, Zhao C, Zhang P, Liao Q, Liu Y, Yuan Q, Wu S, Kang M, Geng J. 2019. Rapid Nanopore Assay for Carbapenem-Resistant Klebsiella pneumoniae. Front Microbiol 10.

164. Su M, Satola SW, Read TD. 2019. Genome-Based Prediction of Bacterial Antibiotic Resistance. Journal of Clinical Microbiology 57.

165. Alam MT, Petit RA, Crispell EK, Thornton TA, Conneely KN, Jiang Y, Satola SW, Read TD. 2014. Dissecting Vancomycin-Intermediate Resistance in Staphylococcus aureus Using Genome-Wide Association. Genome Biol Evol 6:1174–1185.

166. Moller AG, Winston K, Ji S, Wang J, Davis MNH, Solís-Lemus CR, Read TD. 2021. Genes Influencing Phage Host Range in Staphylococcus aureus on a Species-Wide Scale. mSphere 6.

167. Kong EF, Johnson JK, Jabra-Rizk MA. 2016. Community-Associated Methicillin-Resistant Staphylococcus aureus: An Enemy amidst Us. PLOS Pathogens 12:e1005837.

168. Wang I-N, Smith DL, Young R. 2000. Holins: The Protein Clocks of Bacteriophage Infections. Annual Review of Microbiology 54:799–825.

169. Young R, Bläsi U. 1995. Holins: form and function in bacteriophage lysis. FEMS Microbiology Reviews 17:195–205.

170. Loessner MJ. 2005. Bacteriophage endolysins — current state of research and applications. Current Opinion in Microbiology 8:480–487.

171. Pirisi A. 2000. Phage therapy—advantages over antibiotics? The Lancet 356:1418.

172. Jault P, Leclerc T, Jennes S, Pirnay JP, Que Y-A, Resch G, Rousseau AF, Ravat F, Carsin H, Le Floch R, Schaal JV, Soler C, Fevre C, Arnaud I, Bretaudeau L, Gabard J. 2018. Efficacy and tolerability of a cocktail of bacteriophages to treat burn wounds infected by Pseudomonas aeruginosa (PhagoBurn): a randomised, controlled, double-blind phase 1/2 trial. The Lancet Infectious Diseases https://doi.org/10.1016/S1473-3099(18)30482-1.

173. Synnott AJ, Kuang Y, Kurimoto M, Yamamichi K, Iwano H, Tanji Y. 2009. Isolation from Sewage Influent and Characterization of Novel Staphylococcus aureus Bacteriophages with Wide Host Ranges and Potent Lytic Capabilities. Appl Environ Microbiol 75:4483–4490.

174. O'Flaherty S, Ross RP, Meaney W, Fitzgerald GF, Elbreki MF, Coffey A. 2005. Potential

of the Polyvalent Anti-Staphylococcus Bacteriophage K for Control of Antibiotic-Resistant Staphylococci from Hospitals. Appl Environ Microbiol 71:1836–1842.

175. Gutiérrez D, Briers Y, Rodríguez-Rubio L, Martínez B, Rodríguez A, Lavigne R, García P. 2015. Role of the Pre-neck Appendage Protein (Dpo7) from Phage vB_SepiS-phiIPLA7 as an Anti-biofilm Agent in Staphylococcal Species. Front Microbiol 6.
176. Mccarthy AJ, Witney AA, Lindsay JA. 2012. Staphylococcus aureus Temperate Bacteriophage: Carriage and Horizontal Gene Transfer is Lineage Associated. Front Cell Infect Microbiol 2.
177. Carroll JD, Cafferkey MT, Coleman DC. 1993. Serotype F double- and triple-converting phage insertionally inactivate the Staphylococcus aureusβ-toxin determinant by a common molecular mechanism. FEMS Microbiology Letters 106:147–155.
178. Iandolo JJ, Worrell V, Groicher KH, Qian Y, Tian R, Kenton S, Dorman A, Ji H, Lin S, Loh P, Qi S, Zhu H, Roe BA. 2002. Comparative analysis of the genomes of the temperate bacteriophages φ11, φ12 and φ13 of Staphylococcusaureus 8325. Gene 289:109–118.
179. Hyman P, Abedon ST. 2010. Bacteriophage Host Range and Bacterial Resistance. Advances in Applied Microbiology 70:217–248.
180. Allison GE, Klaenhammer TR. 1998. Phage Resistance Mechanisms in Lactic Acid Bacteria. International Dairy Journal 8:207–226.
181. Xia G, Corrigan RM, Winstel V, Goerke C, Gründling A, Peschel A. 2011. Wall Teichoic Acid-Dependent Adsorption of Staphylococcal Siphovirus and Myovirus. J Bacteriol 193:4006–4009.
182. Brown S, Santa Maria JP, Walker S. 2013. Wall Teichoic Acids of Gram-Positive Bacteria. Annu Rev Microbiol 67.
183. Shaw DRD, Mirelman D, Chatterjee AN, Park JT. 1970. Ribitol Teichoic Acid Synthesis in Bacteriophage-resistant Mutants of Staphylococcus aureus H. J Biol Chem 245:5101–5106.

184. Shaw DRD, Chatterjee AN. 1971. O-Acetyl Groups as a Component of the Bacteriophage Receptor on Staphylococcus aureus Cell Walls. J Bacteriol 108:584–585.

185. Chatterjee AN. 1969. Use of Bacteriophage-resistant Mutants to Study the Nature of the Bacteriophage Receptor Site of Staphylococcus aureus. J Bacteriol 98:519–527.

186. Chatterjee AN, Mirelman D, Singer HJ, Park JT. 1969. Properties of a Novel Pleiotropic Bacteriophage-Resistant Mutant of Staphylococcus aureus H1. J Bacteriol 100:846–853.

187. NCTC 3000 Project.

188. Capparelli R, Nocerino N, Lanzetta R, Silipo A, Amoresano A, Giangrande C, Becker K, Blaiotta G, Evidente A, Cimmino A, Iannaccone M, Parlato M, Medaglia C, Roperto S, Roperto F, Ramunno L, Iannelli D. 2010. Bacteriophage-Resistant Staphylococcus aureus Mutant Confers Broad Immunity against Staphylococcal Infection in Mice. PLOS ONE 5:e11720.

189. Nair D, Memmi G, Hernandez D, Bard J, Beaume M, Gill S, Francois P, Cheung AL. 2011. Whole-Genome Sequencing of Staphylococcus aureus Strain RN4220, a Key Laboratory Strain Used in Virulence Research, Identifies Mutations That Affect Not Only Virulence Factors but Also the Fitness of the Strain. Journal of Bacteriology 193:2332–2335.

190. Sobhanifar S, Worrall LJ, Gruninger RJ, Wasney GA, Blaukopf M, Baumann L, Lameignere E, Solomonson M, Brown ED, Withers SG, Strynadka NCJ. 2015. Structure and mechanism of Staphylococcus aureus TarM, the wall teichoic acid α-glycosyltransferase. Proc Natl Acad Sci USA 112:E576-585.

191. Sobhanifar S, Worrall LJ, King DT, Wasney GA, Baumann L, Gale RT, Nosella M, Brown ED, Withers SG, Strynadka NCJ. 2016. Structure and Mechanism of Staphylococcus aureus TarS, the Wall Teichoic Acid β-glycosyltransferase Involved in Methicillin Resistance. PLOS Pathogens 12:e1006067.

192. Li X, Koç C, Kühner P, Stierhof Y-D, Krismer B, Enright MC, Penadés JR, Wolz C, Stehle

T, Cambillau C, Peschel A, Xia G. 2016. An essential role for the baseplate protein Gp45 in phage adsorption to Staphylococcus aureus. Sci Rep 6.

193. Brown S, Xia G, Luhachack LG, Campbell J, Meredith TC, Chen C, Winstel V, Gekeler C, Irazoqui JE, Peschel A, Walker S. 2012. Methicillin resistance in Staphylococcus aureus requires glycosylated wall teichoic acids. PNAS 109:18909–18914.
194. Qian Z, Yin Y, Zhang Y, Lu L, Li Y, Jiang Y. 2006. Genomic characterization of ribitol teichoic acid synthesis in Staphylococcus aureus: genes, genomic organization and gene duplication. BMC Genomics 7:74.
195. Cerca N, Oliveira R, Azeredo J. Susceptibility of Staphylococcus epidermidis planktonic cells and biofilms to the lytic action of staphylococcus bacteriophage K. Letters in Applied Microbiology 45:313–317.
196. Azam AH, Hoshiga F, Takeuchi I, Miyanaga K, Tanji Y. 2018. Analysis of phage resistance in Staphylococcus aureus SA003 reveals different binding mechanisms for the closely related Twort-like phages ϕSA012 and ϕSA039. Appl Microbiol Biotechnol 1–15.
197. Baptista C, Santos MA, São-José C. 2008. Phage SPP1 reversible adsorption to Bacillus subtilis cell wall teichoic acids accelerates virus recognition of membrane receptor YueB. J Bacteriol 190:4989–4996.
198. Duerkop BA, Huo W, Bhardwaj P, Palmer KL, Hooper LV. 2016. Molecular Basis for Lytic Bacteriophage Resistance in Enterococci. mBio 7:e01304-16.
199. Ton-That H, Liu G, Mazmanian SK, Faull KF, Schneewind O. 1999. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. PNAS 96:12424–12429.
200. Baba T, Bae T, Schneewind O, Takeuchi F, Hiramatsu K. 2008. Genome Sequence of Staphylococcus aureus Strain Newman and Comparative Analysis of Staphylococcal Genomes: Polymorphism and Evolution of Two Major Pathogenicity Islands. Journal of Bacteriology 190:300–310.

201. Bae T, Baba T, Hiramatsu K, Schneewind O. 2006. Prophages of Staphylococcus aureus Newman and their contribution to virulence. Molecular Microbiology 62:1035–1047.

202. Sutra L, Audurier A, Poutrel B. 1988. Relationship between capsular types 5 and 8 and phage types in Staphylococcus aureus isolates from cow, goat and ewe milk. FEMS Microbiol Lett 55:83–85.

203. Sompolinsky D, Samra Z, Karakawa WW, Vann WF, Schneerson R, Malik Z. 1985. Encapsulation and capsular types in isolates of Staphylococcus aureus from different sources and relationship to phage types. J Clin Microbiol 22:828–834.

204. Ohshima Y, Schumacher-Perdreau F, Peters G, Pulverer G. 1988. The role of capsule as a barrier to bacteriophage adsorption in an encapsulated Staphylococcus simulans strain. Med Microbiol Immunol 177:229–233.

205. O'Gara JP. 2007. ica and beyond: biofilm mechanisms and regulation in Staphylococcus epidermidis and Staphylococcus aureus. FEMS Microbiol Lett 270:179–188.

206. Götz F. 2002. Staphylococcus and biofilms. Molecular Microbiology 43:1367–1378.

207. McCourt J, O'Halloran DP, McCarthy H, O'Gara JP, Geoghegan JA. 2014. Fibronectin-binding proteins are required for biofilm formation by community-associated methicillin-resistant Staphylococcus aureus strain LAC. FEMS Microbiol Lett 353:157–164.

208. O'Neill E, Pozzi C, Houston P, Humphreys H, Robinson DA, Loughman A, Foster TJ, O'Gara JP. 2008. A Novel Staphylococcus aureus Biofilm Phenotype Mediated by the Fibronectin-Binding Proteins, FnBPA and FnBPB. Journal of Bacteriology 190:3835–3850.

209. Geoghegan JA, Corrigan RM, Gruszka DT, Speziale P, O'Gara JP, Potts JR, Foster TJ. 2010. Role of Surface Protein SasG in Biofilm Formation by Staphylococcus aureus. Journal of Bacteriology 192:5663–5673.

210. Kelly D, McAuliffe O, Ross RP, Coffey A. Prevention of Staphylococcus aureus biofilm formation and reduction in established biofilm density using a combination of phage K

and modified derivatives. Letters in Applied Microbiology 54:286–291.

211. Curtin JJ, Donlan RM. 2006. Using Bacteriophages To Reduce Formation of Catheter-Associated Biofilms by Staphylococcus epidermidis. Antimicrob Agents Chemother 50:1268–1275.

212. Gutiérrez D, Martínez B, Rodríguez A, García P. 2012. Genomic characterization of two Staphylococcus epidermidis bacteriophages with anti-biofilm potential. BMC Genomics 13:228.

213. Resch A, Fehrenbacher B, Eisele K, Schaller M, Götz F. Phage release from biofilm and planktonic Staphylococcus aureus cells. FEMS Microbiology Letters 252:89–96.

214. Kameyama L, Fernandez L, Calderon J, Ortiz-Rojas A, Patterson TA. 1999. Characterization of Wild Lambdoid Bacteriophages: Detection of a Wide Distribution of Phage Immunity Groups and Identification of a Nus-Dependent, Nonlambdoid Phage Group. Virology 263:100–111.

215. 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.

216. Sadykov MR. 2014. Restriction–Modification Systems as a Barrier for Genetic Manipulation of Staphylococcus aureus, p. 9–23. *In* Bose, JL (ed.), The Genetic Manipulation of Staphylococci. Springer New York, New York, NY.

217. 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:e00277-11.

218. Monk IR, Foster TJ. 2012. Genetic manipulation of Staphylococci – Breaking through the barrier. Front Cell Infect Microbiol 2.

219. Costa SK, Donegan NP, Corvaglia A-R, François P, Cheung AL. 2017. Bypassing the Restriction System To Improve Transformation of Staphylococcus epidermidis. J Bacteriol

199.

220. Roberts RJ, Vincze T, Posfai J, Macelis D. 2015. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 43:D298–D299.

221. Cooper LP, Roberts GA, White JH, Luyten YA, Bower EKM, Morgan RD, Roberts RJ, Lindsay JA, Dryden DTF. 2017. DNA target recognition domains in the Type I restriction and modification systems of Staphylococcus aureus. Nucleic Acids Res 45:3395–3406.

222. Stegger M, Lindsay JA, Moodley A, Skov R, Broens EM, Guardabassi L. 2011. Rapid PCR Detection of Staphylococcus aureus Clonal Complex 398 by Targeting the Restriction-Modification System Carrying sau1-hsdS1. Journal of Clinical Microbiology 49:732–734.

223. Veiga H, Pinho MG. 2009. Inactivation of the SauI Type I Restriction-Modification System Is Not Sufficient To Generate Staphylococcus aureus Strains Capable of Efficiently Accepting Foreign DNA. Appl Environ Microbiol 75:3034–3038.

224. Seeber S, Kessler C, Götz F. 1990. Cloning, expression and characterization of the Sau3AI restriction and modification genes in Staphylococcus carnosus TM300. Gene 94:37–43.

225. Sussenbach JS, Steenbergh PH, Rost JA, van Leeuwen WJ, van Embden JD. 1978. A second site-specific restriction endonuclease from Staphylococcus aureus. Nucleic Acids Res 5:1153–1163.

226. Lebenka Ai, Rachkus I. 1989. [DNA-methylase Sau 3A: isolation and various properties]. Biokhimiia 54:1009–1014.

227. Tock MR, Dryden DT. 2005. The biology of restriction and anti-restriction. Current Opinion in Microbiology 8:466–472.

228. O'Flaherty S, Coffey A, Edwards R, Meaney W, Fitzgerald GF, Ross RP. 2004. Genome of Staphylococcal Phage K: a New Lineage of Myoviridae Infecting Gram-Positive

Bacteria with a Low G+C Content. Journal of Bacteriology 186:2862–2871.

229. Iida S, Streiff MB, Bickle TA, Arber W. 1987. Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA− phages. Virology 157:156–166.

230. Kruger DH, Barcak GJ, Reuter M, Smith HO. 1988. EcoRII can be activated to cleave refractory DNA recognition sites. Nucleic Acids Res 16:3997–4008.

231. Zabeau M, Friedman S, Van Montagu M, Schell J. 1980. The ral gene of phage λ. Molec Gen Genet 179:63–73.

232. Studier FW, Movva NR. 1976. SAMase gene of bacteriophage T3 is responsible for overcoming host restriction. Journal of Virology 19:136–145.

233. Zavilgelsky GB, Kotova VYu, Rastorguev SM. 2008. Comparative analysis of anti-restriction activities of ArdA (ColIb-P9) and Ocr (T7) proteins. Biochemistry Moscow 73:906.

234. McMahon SA, Roberts GA, Johnson KA, Cooper LP, Liu H, White JH, Carter LG, Sanghvi B, Oke M, Walkinshaw MD, Blakely GW, Naismith JH, Dryden DTF. 2009. Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance. Nucleic Acids Res 37:4887–4897.

235. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science 315:1709–1712.

236. Li Q, Xie X, Yin K, Tang Y, Zhou X, Chen Y, Xia J, Hu Y, Ingmer H, Li Y, Jiao X. 2016. Characterization of CRISPR-Cas system in clinical Staphylococcus epidermidis strains revealed its potential association with bacterial infection sites. Microbiological Research 193:103–110.

237. Yang S, Liu J, Shao F, Wang P, Duan G, Yang H. 2015. Analysis of the features of 45 identified CRISPR loci in 32 Staphylococcus aureus. Biochemical and Biophysical

Research Communications 464:894–900.

238. Carte J, Wang R, Li H, Terns RM, Terns MP. 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev 22:3489–3496.

239. Jiang F, Doudna JA. 2015. The structural biology of CRISPR-Cas systems. Current Opinion in Structural Biology 30:100–111.

240. Gasiunas G, Barrangou R, Horvath P, Siksnys V. 2012. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. PNAS 109:E2579–E2586.

241. Kinnevey PM, Shore AC, Brennan GI, Sullivan DJ, Ehricht R, Monecke S, Slickers P, Coleman DC. 2013. Emergence of Sequence Type 779 Methicillin-Resistant Staphylococcus aureus Harboring a Novel Pseudo Staphylococcal Cassette Chromosome mec (SCCmec)-SCC-SCCCRISPR Composite Element in Irish Hospitals. Antimicrob Agents Chemother 57:524–531.

242. Gill SR, Fouts DE, Archer GL, Mongodin EF, DeBoy RT, Ravel J, Paulsen IT, Kolonay JF, Brinkac L, Beanan M, Dodson RJ, Daugherty SC, Madupu R, Angiuoli SV, Durkin AS, Haft DH, Vamathevan J, Khouri H, Utterback T, Lee C, Dimitrov G, Jiang L, Qin H, Weidman J, Tran K, Kang K, Hance IR, Nelson KE, Fraser CM. 2005. Insights on Evolution of Virulence and Resistance from the Complete Genome Analysis of an Early Methicillin-Resistant Staphylococcus aureus Strain and a Biofilm-Producing Methicillin-Resistant Staphylococcus epidermidis Strain. J Bacteriol 187:2426–2438.

243. Marraffini LA, Sontheimer EJ. 2008. CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. Science 322:1843–1845.

244. Bondy-Denomy J, Garcia B, Strum S, Du M, Rollins MF, Hidalgo-Reyes Y, Wiedenheft B, Maxwell KL, Davidson AR. 2015. Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. Nature 526:136–139.

245. Pawluk A, Staals RHJ, Taylor C, Watson BNJ, Saha S, Fineran PC, Maxwell KL,

Davidson AR. 2016. Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. Nature Microbiology 1:16085.

246. Maxwell KL. 2016. Phages Fight Back: Inactivation of the CRISPR-Cas Bacterial Immune System by Anti-CRISPR Proteins. PLOS Pathogens 12:e1005282.
247. Rauch BJ, Silvis MR, Hultquist JF, Waters CS, McGregor MJ, Krogan NJ, Bondy-Denomy J. 2017. Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. Cell 168:150-158.e10.
248. Lindsay JA, Ruzin A, Ross HF, Kurepina N, Novick RP. 1998. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in Staphylococcus aureus. Molecular Microbiology 29:527–543.
249. Ruzin A, Lindsay J, Novick RP. 2001. Molecular genetics of SaPI1 – a mobile pathogenicity island in Staphylococcus aureus. Molecular Microbiology 41:365–377.
250. Úbeda Carles, Maiques Elisa, Barry Peter, Matthews Avery, Tormo María Ángeles, Lasa Íñigo, Novick Richard P., Penadés José R. 2007. SaPI mutations affecting replication and transfer and enabling autonomous replication in the absence of helper phage. Molecular Microbiology 67:493–503.
251. Poliakov A, Chang JR, Spilman MS, Damle PK, Christie GE, Mobley JA, Dokland T. 2008. Capsid size determination by Staphylococcus aureus pathogenicity island SaPI1 involves specific incorporation of SaPI1 proteins into procapsids. J Mol Biol 380:465–475.
252. Frígols B, Quiles-Puchalt N, Mir-Sanchis I, Donderis J, Elena SF, Buckling A, Novick RP, Marina A, Penadés JR. 2015. Virus Satellites Drive Viral Evolution and Ecology. PLoS Genet 11.
253. Atilano ML, Pereira PM, Yates J, Reed P, Veiga H, Pinho MG, Filipe SR. 2010. Teichoic acids are temporal and spatial regulators of peptidoglycan cross-linking in Staphylococcus aureus. PNAS 107:18991–18996.
254. Sun X, Göhler A, Heller KJ, Neve H. 2006. The ltp gene of temperate Streptococcus thermophilus phage TP-J34 confers superinfection exclusion to Streptococcus

thermophilus and Lactococcus lactis. Virology 350:146–157.

255. Chopin M-C, Chopin A, Bidnenko E. 2005. Phage abortive infection in lactococci: variations on a theme. Current Opinion in Microbiology 8:473–479.

256. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R. 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. Science eaar4120.

257. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, Afik S, Ofir G, Sorek R. 2015. BREX is a novel phage resistance system widespread in microbial genomes. The EMBO Journal 34:169–183.

258. Barrangou R, van der Oost J. 2015. Bacteriophage exclusion, a new defense system. EMBO J 34:134–135.

259. Ofir G, Melamed S, Sberro H, Mukamel Z, Silverman S, Yaakov G, Doron S, Sorek R. 2018. DISARM is a widespread bacterial defence system with broad anti-phage activities. Nat Microbiol 3:90–98.

260. Petit III RA, Read TD. 2018. Staphylococcus aureus viewed from the perspective of 40,000+ genomes. PeerJ 6:e5261.

261. Weidenmaier C, Kokai-Kun JF, Kristian SA, Chanturiya T, Kalbacher H, Gross M, Nicholson G, Neumeister B, Mond JJ, Peschel A. 2004. Role of teichoic acids in Staphylococcus aureus nasal colonization, a major risk factor in nosocomial infections. Nature Medicine 10:243–245.

262. D'Elia MA, Henderson JA, Beveridge TJ, Heinrichs DE, Brown ED. 2009. The N-Acetylmannosamine Transferase Catalyzes the First Committed Step of Teichoic Acid Assembly in Bacillus subtilis and Staphylococcus aureus. J Bacteriol 191:4030–4034.

263. Winstel V, Kühner P, Salomon F, Larsen J, Skov R, Hoffmann W, Peschel A, Weidenmaier C. 2015. Wall Teichoic Acid Glycosylation Governs Staphylococcus aureus Nasal Colonization. mBio 6:e00632-15.

264. Schlag M, Biswas R, Krismer B, Kohler T, Zoll S, Yu W, Schwarz H, Peschel A, Götz F. 2010. Role of staphylococcal wall teichoic acid in targeting the major autolysin Atl. Molecular Microbiology 75:864–873.

265. Biswas R, Martinez RE, Göhring N, Schlag M, Josten M, Xia G, Hegler F, Gekeler C, Gleske A-K, Götz F, Sahl H-G, Kappler A, Peschel A. 2012. Proton-binding capacity of Staphylococcus aureus wall teichoic acid and its role in controlling autolysin activity. PLoS ONE 7:e41415.

266. Bera A, Biswas R, Herbert S, Kulauzovic E, Weidenmaier C, Peschel A, Götz F. 2007. Influence of Wall Teichoic Acid on Lysozyme Resistance in Staphylococcus aureus. J Bacteriol 189:280–283.

267. Kohler T, Weidenmaier C, Peschel A. 2009. Wall Teichoic Acid Protects Staphylococcus aureus against Antimicrobial Fatty Acids from Human Skin. J Bacteriol 191:4482–4484.

268. Holland LM, Conlon B, O'Gara JP. 2011. Mutation of tagO reveals an essential role for wall teichoic acids in Staphylococcus epidermidis biofilm development. Microbiology 157:408–418.

269. Yu L, Wang S, Guo Z, Liu H, Sun D, Yan G, Hu D, Du C, Feng X, Han W, Gu J, Sun C, Lei L. 2018. A guard-killer phage cocktail effectively lyses the host and inhibits the development of phage-resistant strains of Escherichia coli. Appl Microbiol Biotechnol 102:971–983.

270. Gu J, Liu X, Li Y, Han W, Lei L, Yang Y, Zhao H, Gao Y, Song J, Lu R, Sun C, Feng X. 2012. A Method for Generation Phage Cocktail with Great Therapeutic Potential. PLOS ONE 7:e31698.

271. McCarthy AJ, Loeffler A, Witney AA, Gould KA, Lloyd DH, Lindsay JA. 2014. Extensive Horizontal Gene Transfer during Staphylococcus aureus Co-colonization In Vivo. Genome Biol Evol 6:2697–2708.

272. Azam AH, Tanji Y. 2019. Peculiarities of Staphylococcus aureus phages and their

possible application in phage therapy. Appl Microbiol Biotechnol 103:4279–4289.

273. Takeuchi I, Osada K, Azam AH, Asakawa H, Miyanaga K, Tanji Y. 2016. The Presence of Two Receptor-Binding Proteins Contributes to the Wide Host Range of Staphylococcal Twort-Like Phages. Appl Environ Microbiol 82:5763–5774.
274. Cao L, Gao C-H, Zhu J, Zhao L, Wu Q, Li M, Sun B. 2016. Identification and functional study of type III-A CRISPR-Cas systems in clinical isolates of Staphylococcus aureus. International Journal of Medical Microbiology 306:686–696.
275. Zhao X, Yu Z, Xu Z. 2018. Study the Features of 57 Confirmed CRISPR Loci in 38 Strains of Staphylococcus aureus. Front Microbiol 9.
276. Christie GE, Dokland T. 2012. Pirates of the Caudovirales. Virology 434:210–221.
277. Abatángelo V, Peressutti Bacci N, Boncompain CA, Amadio AF, Carrasco S, Suárez CA, Morbidoni HR. 2017. Broad-range lytic bacteriophages that kill Staphylococcus aureus local field strains. PLoS ONE 12:e0181671.
278. Iwano H, Inoue Y, Takasago T, Kobayashi H, Furusawa T, Taniguchi K, Fujiki J, Yokota H, Usui M, Tanji Y, Hagiwara K, Higuchi H, Tamura Y. 2018. Bacteriophage ΦSA012 Has a Broad Host Range against Staphylococcus aureus and Effective Lytic Capacity in a Mouse Mastitis Model. Biology (Basel) 7.
279. Peng C, Hanawa T, Azam AH, LeBlanc C, Ung P, Matsuda T, Onishi H, Miyanaga K, Tanji Y. 2019. Silviavirus phage ϕMR003 displays a broad host range against methicillin-resistant Staphylococcus aureus of human origin. Appl Microbiol Biotechnol 103:7751–7765.
280. 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.
281. Krausz KL, Bose JL. 2016. Bacteriophage Transduction in Staphylococcus aureus: Broth-Based Method. Methods Mol Biol 1373:63–68.

282. Su M, Lyles JT, Iii RAP, 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.

283. Bernardy EE, Petit RA, Moller AG, Blumenthal JA, McAdam AJ, Priebe GP, Chande AT, Rishishwar L, Jordan IK, Read TD, Goldberg JB. 2019. Whole-Genome Sequences of Staphylococcus aureus Isolates from Cystic Fibrosis Lung Infections. Microbiol Resour Announc 8:e01564-18.

284. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer.

285. 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.

286. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. 1. Nature Communications 9:5114.

287. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology 19:455–477.

288. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

289. 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.

290. 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 Res 43:e15–e15.

291. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32:268–274.

292. Seemann T. 2020. tseemann/mlst. Perl.

293. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259.

294. Münkemüller T, Lavergne S, Bzeznik B, Dray S, Jombart T, Schiffers K, Thuiller W. 2012. How to measure and test phylogenetic signal. Methods in Ecology and Evolution 3:743–756.

295. Jombart T, Balloux F, Dray S. 2010. adephylo: new tools for investigating the phylogenetic signal in biological traits. Bioinformatics 26:1907–1909.

296. Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223.

297. 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. Nat Commun 7.

298. 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.

299. 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. Microb Genom 2.

300. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. 2018. Phandango: an interactive viewer for bacterial population genomics. Bioinformatics 34:292–293.

301. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM.

2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6:80–92.

302. 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.
303. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. 2015. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43:D447–D452.
304. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res 45:D183–D189.
305. Wardenburg JB, Williams WA, Missiakas D. 2006. Host defenses against Staphylococcus aureus infection require recognition of bacterial lipoproteins. PNAS 103:13831–13836.
306. 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.
307. Löfblom J, Kronqvist N, Uhlén M, Ståhl S, Wernérus H. 2007. Optimization of electroporation-mediated transformation: Staphylococcus carnosus as model organism. Journal of Applied Microbiology 102:736–747.
308. Chen T, He T. xgboost: eXtreme Gradient Boosting 4.
309. Kingma DP, Ba J. 2017. Adam: A Method for Stochastic Optimization. arXiv:14126980 [cs].
310. Shannon CE. 1948. A mathematical theory of communication. The Bell System Technical

Journal 27:379–423.

311. Wang H, Gill CJ, Lee SH, Mann P, Zuck P, Meredith TC, Murgolo N, She X, Kales S, Liang L, Liu J, Wu J, Santa Maria J, Su J, Pan J, Hailey J, Mcguinness D, Tan CM, Flattery A, Walker S, Black T, Roemer T. 2013. Discovery of Wall Teichoic Acid Inhibitors as Potential Anti-MRSA β-Lactam Combination Agents. Chemistry & Biology 20:272–284.
312. Proctor AR, Kloos WE. 1973. Tryptophan Biosynthetic Enzymes of Staphylococcus aureus. Journal of Bacteriology 114:169–177.
313. Kaur S, Harjai K, Chhibber S. 2012. Methicillin-Resistant Staphylococcus aureus Phage Plaque Size Enhancement Using Sublethal Concentrations of Antibiotics. Appl Environ Microbiol 78:8227–8233.
314. Kelliher JL, Radin JN, Kehl-Fie TE. 2018. PhoPR contributes to Staphylococcus aureus growth during phosphate starvation and pathogenesis in an environment-specific manner. Infect Immun IAI.00371-18.
315. Botella E, Devine SK, Hubner S, Salzberg LI, Gale RT, Brown ED, Link H, Sauer U, Codée JD, Noone D, Devine KM. 2014. PhoR autokinase activity is controlled by an intermediate in wall teichoic acid metabolism that is sensed by the intracellular PAS domain during the PhoPR-mediated phosphate limitation response of Bacillus subtilis. Molecular Microbiology 94:1242–1259.
316. Liu W, Eder S, Hulett FM. 1998. Analysis of Bacillus subtilis tagAB and tagDEF Expression during Phosphate Starvation Identifies a Repressor Role for PhoP∼P. J Bacteriol 180:753–758.
317. Myers CL, Li FKK, Koo B-M, El-Halfawy OM, French S, Gross CA, Strynadka NCJ, Brown ED. 2016. Identification of Two Phosphate Starvation-induced Wall Teichoic Acid Hydrolases Provides First Insights into the Degradative Pathway of a Key Bacterial Cell Wall Component. J Biol Chem 291:26066–26082.
318. Jorge AM, Schneider J, Unsleber S, Xia G, Mayer C, Peschel A. 2018. Staphylococcus

aureus counters phosphate limitation by scavenging wall teichoic acids from other staphylococci via the teichoicase GlpQ. J Biol Chem 293:14916–14924.

319. Ladjouzi R, Bizzini A, Lebreton F, Sauvageot N, Rincé A, Benachour A, Hartke A. 2013. Analysis of the tolerance of pathogenic enterococci and Staphylococcus aureus to cell wall active antibiotics. J Antimicrob Chemother 68:2083–2091.
320. Ladjouzi R, Bizzini A, Schaik W van, Zhang X, Rincé A, Benachour A, Hartke A. 2015. Loss of Antibiotic Tolerance in Sod-Deficient Mutants Is Dependent on the Energy Source and Arginine Catabolism in Enterococci. J Bacteriol 197:3283–3293.
321. Sacher JC, Flint A, Butcher J, Blasdel B, Reynolds HM, Lavigne R, Stintzi A, Szymanski CM. 2018. Transcriptomic Analysis of the Campylobacter jejuni Response to T4-Like Phage NCTC 12673 Infection. 6. Viruses 10:332.
322. Chatterjee A, Willett JLE, Nguyen UT, Monogue B, Palmer KL, Dunny GM, Duerkop BA. 2020. Parallel Genomics Uncover Novel Enterococcal-Bacteriophage Interactions. mBio 11.
323. Peschel A, Jack RW, Otto M, Collins LV, Staubitz P, Nicholson G, Kalbacher H, Nieuwenhuizen WF, Jung G, Tarkowski A, Kessel KPM van, Strijp JAG van. 2001. Staphylococcus aureus Resistance to Human Defensins and Evasion of Neutrophil Killing via the Novel Virulence Factor Mprf Is Based on Modification of Membrane Lipids with l-Lysine. Journal of Experimental Medicine 193:1067–1076.
324. Bayer AS, Schneider T, Sahl H-G. 2013. Mechanisms of daptomycin resistance in Staphylococcus aureus: role of the cell membrane and cell wall. Annals of the New York Academy of Sciences 1277:139–158.
325. Andrä J, Goldmann T, Ernst CM, Peschel A, Gutsmann T. 2011. Multiple peptide resistance factor (MprF)-mediated Resistance of Staphylococcus aureus against antimicrobial peptides coincides with a modulated peptide interaction with artificial membranes comprising lysyl-phosphatidylglycerol. J Biol Chem 286:18692–18700.

326. Young R. 2014. Phage lysis: three steps, three choices, one outcome. J Microbiol 52:243–258.

327. Catalão MJ, Gil F, Moniz-Pereira J, São-José C, Pimentel M. 2013. Diversity in bacterial lysis systems: bacteriophages show the way. FEMS Microbiol Rev 37:554–571.

328. Geiger T, Goerke C, Fritz M, Schäfer T, Ohlsen K, Liebeke M, Lalk M, Wolz C. 2010. Role of the (p)ppGpp Synthase RSH, a RelA/SpoT Homolog, in Stringent Response and Virulence of Staphylococcus aureus. Infect Immun 78:1873–1883.

329. Fernández L, González S, Campelo AB, Martínez B, Rodríguez A, García P. 2017. Low-level predation by lytic phage phiIPLA-RODI promotes biofilm formation and triggers the stringent response in Staphylococcus aureus. Sci Rep 7.

330. Gaca AO, Colomer-Winter C, Lemos JA. 2015. Many Means to a Common End: the Intricacies of (p)ppGpp Metabolism and Its Control of Bacterial Homeostasis. J Bacteriol 197:1146–1156.

331. Conlon BP, Rowe SE, Gandt AB, Nuxoll AS, Donegan NP, Zalis EA, Clair G, Adkins JN, Cheung AL, Lewis K. 2016. Persister formation in Staphylococcus aureus is associated with ATP depletion. Nat Microbiol 1.

332. Torres VJ, Pishchany G, Humayun M, Schneewind O, Skaar EP. 2006. Staphylococcus aureus IsdB Is a Hemoglobin Receptor Required for Heme Iron Utilization. Journal of Bacteriology 188:8421–8429.

333. Power RA, Parkhill J, de Oliveira T. 2017. Microbial genome-wide association studies: lessons from human GWAS. 1. Nature Reviews Genetics 18:41–50.

334. 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.

335. Mutalik VK, Adler BA, Rishi HS, Piya D, Zhong C, Koskella B, Kutter EM, Calendar R, Novichkov PS, Price MN, Deutschbauer AM, Arkin AP. 2020. High-throughput mapping of

the phage resistance landscape in E. coli. PLoS Biol 18.

336. Dickey J, Perrot V. 2019. Adjunct phage treatment enhances the effectiveness of low antibiotic concentration against Staphylococcus aureus biofilms in vitro. PLoS One 14.

337. Dalen R van, Diaz JSDLC, Rumpret M, Fuchsberger FF, Teijlingen NH van, Hanske J, Rademacher C, Geijtenbeek TBH, Strijp JAG van, Weidenmaier C, Peschel A, Kaplan DH, Sorge NM van. 2019. Langerhans Cells Sense Staphylococcus aureus Wall Teichoic Acid through Langerin To Induce Inflammatory Responses. mBio 10:e00330-19.

338. Wanner S, Schade J, Keinhörster D, Weller N, George SE, Kull L, Bauer J, Grau T, Winstel V, Stoy H, Kretschmer D, Kolata J, Wolz C, Bröker BM, Weidenmaier C. 2017. Wall teichoic acids mediate increased virulence in Staphylococcus aureus. Nature Microbiology 2:16257.

339. Everitt RG, Didelot X, Batty EM, Miller RR, Knox K, Young BC, Bowden R, Auton A, Votintseva A, Larner-Svensson H, Charlesworth J, Golubchik T, Ip CLC, Godwin H, Fung R, Peto TEA, Walker AS, Crook DW, Wilson DJ. 2014. Mobile elements drive recombination hotspots in the core genome of *Staphylococcus aureus*. Nature Communications 5:3956.

340. Narita S, Kaneko J, Chiba J, Piémont Y, Jarraud S, Etienne J, Kamio Y. 2001. Phage conversion of Panton-Valentine leukocidin in Staphylococcus aureus: molecular analysis of a PVL-converting phage, φSLT. Gene 268:195–206.

341. Grant SG, Jessee J, Bloom FR, Hanahan D. 1990. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. Proc Natl Acad Sci U S A 87:4645–4649.

342. Christie GE, Matthews AM, King DG, Lane KD, Olivarez NP, Tallent SM, Gill SR, Novick RP. 2010. The complete genomes of Staphylococcus aureus bacteriophages 80 and 80α—Implications for the specificity of SaPI mobilization. Virology 407:381–390.

343. Ajuebor J, Buttimer C, Arroyo-Moreno S, Chanishvili N, Gabriel EM, O'Mahony J,

McAuliffe O, Neve H, Franz C, Coffey A. 2018. Comparison of Staphylococcus Phage K with Close Phage Relatives Commonly Employed in Phage Therapeutics. Antibiotics (Basel) 7.

344. Berryhill BA, McCall IC, Huseby DL, Hughes D, Levin B. 2020. Joint antibiotic and phage therapy: addressing the limitations of a seemingly ideal phage for treating Staphylococcus aureus infections. bioRxiv 2020.04.24.060335.

345. Gao L, Altae-Tran H, Böhning F, Makarova KS, Segel M, Schmid-Burgk JL, Koob J, Wolf YI, Koonin EV, Zhang F. 2020. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. Science 369:1077–1084.

346. Stern A, Sorek R. 2011. The phage-host arms-race: Shaping the evolution of microbes. Bioessays 33:43–51.

347. Weitz JS, Hartman H, Levin SA. 2005. Coevolutionary arms races between bacteria and bacteriophage. PNAS 102:9535–9540.

348. Winstel V, Xia G, Peschel A. 2014. Pathways and roles of wall teichoic acid glycosylation in Staphylococcus aureus. International Journal of Medical Microbiology 304:215–221.

349. Jensen C, Bæk KT, Gallay C, Thalsø-Madsen I, Xu L, Jousselin A, Torrubia FR, Paulander W, Pereira AR, Veening J-W, Pinho MG, Frees D. 2019. The ClpX chaperone controls autolytic splitting of Staphylococcus aureus daughter cells, but is bypassed by β-lactam antibiotics or inhibitors of WTA biosynthesis. PLOS Pathogens 15:e1008044.

350. Pawluk A, Davidson AR, Maxwell KL. 2018. Anti-CRISPR: discovery, mechanism and function. Nature Reviews Microbiology 16:12–17.

351. Hussain FA, Dubert J, Elsherbini J, Murphy M, VanInsberghe D, Arevalo P, Kauffman K, Rodino-Janeiro BK, Polz M. 2021. Rapid evolutionary turnover of mobile genetic elements drives microbial resistance to viruses. bioRxiv 2021.03.26.437281.

352. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403–410.

353. Bayliss SC, Thorpe HA, Coyle NM, Sheppard SK, Feil EJ. 2019. PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. GigaScience 8.

354. 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.

355. Crispell J, Balaz D, Gordon SV. 2019. HomoplasyFinder: a simple tool to identify homoplasies on a phylogeny. Microbial Genomics, 5:e000245.

356. 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. J Clin Microbiol 52:1182–1191.

357. Liu B, Zheng D, Jin Q, Chen L, Yang J. 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Research 47:D687–D692.

358. Deresinski S. 2005. Methicillin-Resistant Staphylococcus aureus: An Evolutionary, Epidemiologic, and Therapeutic Odyssey. Clin Infect Dis 40:562–573.

359. Backo M, Gaenger E, Burkart A, Chai YL, Bayer AS. 1999. Treatment of Experimental Staphylococcal Endocarditis Due to a Strain with Reduced Susceptibility In Vitro to Vancomycin: Efficacy of Ampicillin-Sulbactam. Antimicrob Agents Chemother 43:2565–2568.

360. Howden BP, Ward PB, Charles PGP, Korman TM, Fuller A, du Cros P, Grabsch EA, Roberts SA, Robson J, Read K, Bak N, Hurley J, Johnson PDR, Morris AJ, Mayall BC, Grayson ML. 2004. Treatment Outcomes for Serious Infections Caused by Methicillin-Resistant Staphylococcus aureus with Reduced Vancomycin Susceptibility. Clin Infect Dis 38:521–528.

361. 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.

362. Charles PGP, Ward PB, Johnson PDR, Howden BP, Grayson ML. 2004. Clinical Features Associated with Bacteremia Due to Heterogeneous Vancomycin-Intermediate Staphylococcus aureus. Clin Infect Dis 38:448–451.
363. Hal SJ van, Wehrhahn MC, Barbagiannakos T, Mercer J, Chen D, Paterson DL, Gosbell IB. 2011. Performance of Various Testing Methodologies for Detection of Heteroresistant Vancomycin-Intermediate Staphylococcus aureus in Bloodstream Isolates. Journal of Clinical Microbiology 49:1489–1494.
364. Read TD, Satola SW. 2014. Using Genomics To Standardize Population Analysis Profile-Area under the Curve Ratio for Vancomycin-Intermediate Staphylococcus aureus. Journal of Clinical Microbiology 52:3824–3826.
365. 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.
366. Che Y, Xia Y, Liu L, Li A-D, Yang Y, Zhang T. 2019. Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. Microbiome 7:44.
367. Arango-Argoty GA, Dai D, Pruden A, Vikesland P, Heath LS, Zhang L. 2019. NanoARG: a web service for detecting and contextualizing antimicrobial resistance genes from nanopore-derived metagenomes. Microbiome 7:88.
368. 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 Res 45:e61–e61.

369. 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.

370. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760.

371. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. GigaScience 10.

372. 2021. nanoporetech/medaka. Python, Oxford Nanopore Technologies.

373. 2021. nanoporetech/bonito. Python, Oxford Nanopore Technologies.

374. Yang C, Chu J, Warren RL, Birol I. 2017. NanoSim: nanopore sequence read simulator based on statistical characterization. Gigascience 6.

375. Pantůček R, Rosypalová A, Doškař J, Kailerová J, Růžičková V, Borecká P, Snopková Š, Horváth R, Götz F, Rosypal S. 1998. The Polyvalent Staphylococcal Phage φ812:Its Host-Range Mutants and Related Phages. Virology 246:241–252.

376. Paez-Espino D, Sharon I, Morovic W, Stahl B, Thomas BC, Barrangou R, Banfield JF. 2015. CRISPR Immunity Drives Rapid Phage Genome Evolution in Streptococcus thermophilus. mBio 6:e00262-15.

377. Mapes AC, Trautner BW, Liao KS, Ramig RF. 2016. Development of expanded host range phage active on biofilms of multi-drug resistant Pseudomonas aeruginosa. Bacteriophage 6.

378. Burrowes BH, Molineux IJ, Fralick JA. 2019. Directed in Vitro Evolution of Therapeutic Bacteriophages: The Appelmans Protocol. Viruses 11:241.