AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE

Species-Scale Genomic Analysis of *Staphylococcus aureus* Genes Influencing Phage Host Range and Their Relationships to Virulence and Antibiotic Resistance Genes

Abraham G. Moller,[a] Robert A. Petit III,[a] Timothy D. Read[a,b]

[a]Division of Infectious Diseases, Department of Medicine, School of Medicine, Emory University, Atlanta, Georgia, USA
[b]Department of Human Genetics, School of Medicine, Emory University, Atlanta, Georgia, USA

ABSTRACT Phage therapy has been proposed as a possible alternative treatment for infections caused by the ubiquitous bacterial pathogen *Staphylococcus aureus*. However, successful therapy requires understanding the genetic basis of host range—the subset of strains in a species that could be killed by a particular phage. We searched diverse sets of *S. aureus* public genome sequences against a database of genes suggested from prior studies to influence host range to look for patterns of variation across the species. We found that genes encoding biosynthesis of molecules that were targets of *S. aureus* phage adsorption to the outer surface of the cell were the most conserved in the pangenome. Putative phage resistance genes that were core components of the pangenome genes had similar nucleotide diversity, ratio of nonsynonymous to synonymous substitutions, and functionality (measured by delta-bitscore) to other core genes. However, phage resistance genes that were not part of the core genome were significantly less consistent with the core genome phylogeny than all noncore genes in this set, suggesting more frequent movement between strains by horizontal gene transfer. Only superinfection immunity genes encoded by temperate phages inserted in the genome correlated with experimentally determined temperate phage resistance. Taken together, these results suggested that, while phage adsorption genes are heavily conserved in the *S. aureus* species, HGT may play a significant role in strain-specific evolution of host range patterns.

IMPORTANCE *Staphylococcus aureus* is a widespread, hospital- and community-acquired pathogen that is commonly antibiotic resistant. It causes diverse diseases affecting both the skin and internal organs. Its ubiquity, antibiotic resistance, and disease burden make new therapies urgent, such as phage therapy, in which viruses specific to infecting bacteria clear infection. *S. aureus* phage host range not only determines whether phage therapy will be successful by killing bacteria but also horizontal gene transfer through transduction of host genetic material by phages. In this work, we comprehensively reviewed existing literature to build a list of *S. aureus* phage resistance genes and searched our database of almost 43,000 *S. aureus* genomes for these genes to understand their patterns of evolution, finding that prophages' superinfection immunity correlates best with phage resistance and HGT. These findings improved our understanding of the relationship between known phage resistance genes and phage host range in the species.

KEYWORDS *Staphylococcus aureus*, bacteriophage, bioinformatics, evolution, genomics, horizontal gene transfer, pangenome, phage host range, phage resistance, transduction

New treatments are needed for *Staphylococcus aureus* infections because of its high prevalence, increasing antibiotic resistance, diverse pathologies, and a lack of available vaccines. Phage therapy—clearing infecting bacteria with bacteriophages—has been proposed as a possible alternative treatment. Potential phage therapy advantages over antibiotics include reduced toxicity because of the specificity of the virus for its host and

Editor Michela Gambino, University of Copenhagen

Address correspondence to Timothy D. Read, tread@emory.edu.

The authors declare no conflict of interest.

Received 27 August 2021
Accepted 21 December 2021
Published 18 January 2022

the high diversity of natural phages available to be isolated for treatment, lessening chances of developing complete resistance (1, 2). 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.

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* (3). The *Siphoviridae* are both temperate and virulent, while the *Myo-* and *Podoviridae* are virulent (3). 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 (4–6).

Bacteriophage life cycles have been proposed to consist of five phages: attachment (adsorption), uptake, biosynthesis, assembly, and lysis (7, 8). From extensive literature review, we found that reported *Staphylococcus* phage resistance mechanisms only acted at the adsorption, biosynthesis, and assembly infection stages of the phage life cycle (4). Adsorption resistance mechanisms include phage receptor alteration, removal, or occlusion by large surface proteins or polysaccharides (capsule) (5, 9–14). Biosynthesis resistance mechanisms involve halting infection through metabolic arrest (abortive infection) or phage DNA degradation by adaptive (CRISPR) or innate (restriction-modification) immunity (15–19). 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 (20–25).

In this work, we searched 40,000+ annotated *S. aureus* genomes for genes known to influence phage resistance and thus potentially influence host range. We wanted to investigate the relationships between predicted phage resistance, empirically determined phage resistance, and other genes frequently transferred between *S. aureus* by horizontal gene transfer (HGT). We developed scoring metrics for resistance determinant diversity, abundance (frequency among strains), functionality, and concordance, as well as overall predicted phage resistance per strain, and accessory gene content as a measure of horizontal gene transfer. 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 among 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 genes influencing phage resistance in *S. aureus* genomes. We curated a list of 331 genes located on bacterial chromosomes and plasmids (Table S1) implicated in previous studies to influence phage resistance. Genes were included based on three criteria: (i) they were directly shown to influence phage resistance in a *S. aureus* strain in laboratory-based studies, 2) *S. aureus* phage resistant mutants had mutations in the gene, 3) they were shown to be associated with phage resistance in other bacterial species. We found that the overwhelming majority of genes "positively" influenced resistance to phages (306/331), i.e., they had the potential to increase the bacterial cell survival. A smaller number (25/331) were of potentially negative effect, i.e., bacteria may become more sensitive to phages. Each gene was classified as interfering with either the adsorption (87/331), biosynthesis (235/331) or assembly (8/331) portion of the phage life cycle.

We first determined whether each gene was core or noncore within the *S. aureus* pangenome based on a search of the Staphopia database containing nearly 43,000 *S. aureus* genomes (26). We defined conservation as the number of strains in our curated set found to encode a gene. We found that the most conserved (core) genes in the

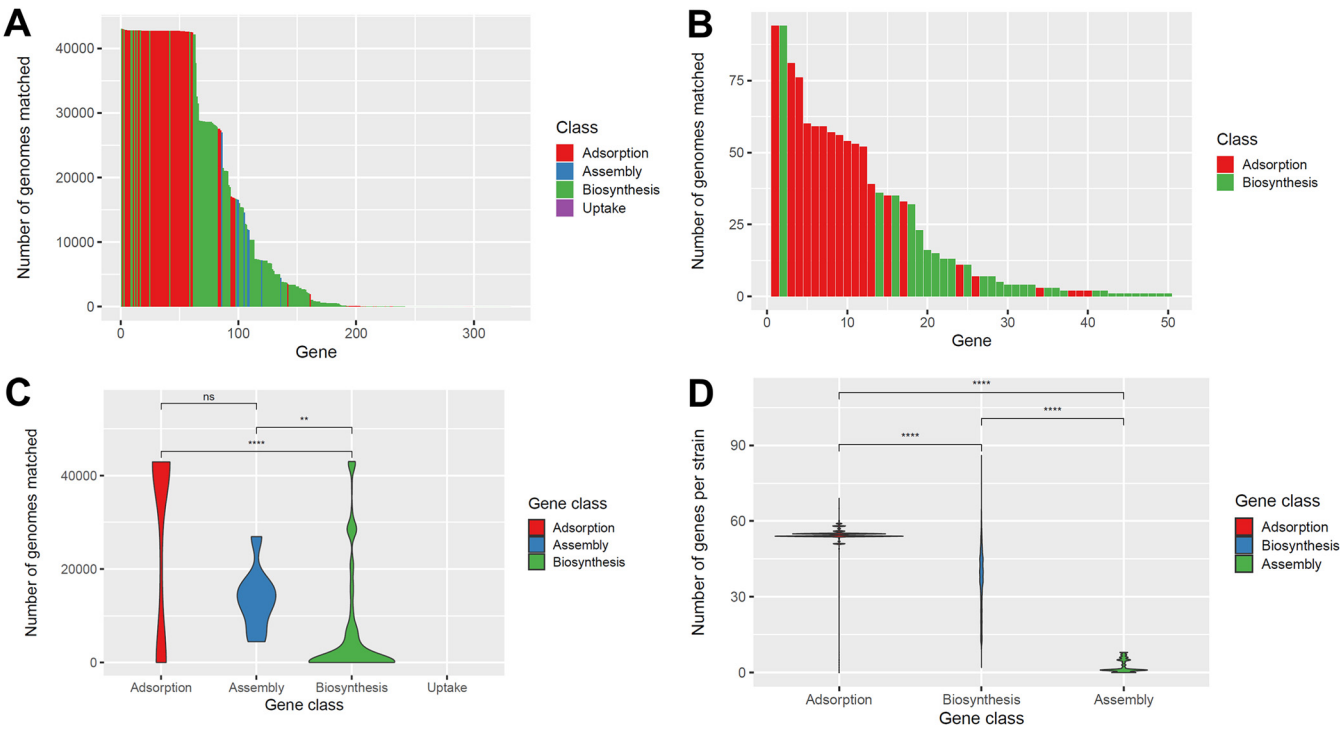

FIG 1 Conservation of examined phage resistance genes in the species based on a search of 40,000+ *S. aureus* genomes. We used BLAST to search for our set of 331 curated phage resistance genes in the Staphopia database. (A) Conservation of each gene (*y* axis) ranked from most to least conserved on the *x* axis. Genes are colored by class—adsorption in red, assembly in blue, and biosynthesis in green. (B) An inset of A showing genes with 100 or fewer strain matches is included in the upper right hand corner. (C) Distributions of conservation for each considered category visualized as violin plots. (D) Distributions of gene count per strain 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).

database were mainly adsorption genes (Fig. 1A; 51/63 core genes present in over 80% of Staphopia genomes). The least conserved genes (less than 100 strain matches) were either biosynthesis or adsorption genes (Fig. 1B). Both adsorption and assembly resistance genes were significantly ($P = 6.7e\text{-}14$ and $P = 0.0028$, respectively) more conserved than biosynthesis resistance genes (here referred to by resistance category—e.g., adsorption) but adsorption genes did not significantly differ in conservation from assembly genes (Fig. 1C) based on non-parametric Wilcoxon tests. Conversely, no assembly resistance genes were core genes (Fig. 1A). On the other hand, looking at the distribution of genes per strain, the fewest assembly genes were encoded per strain, followed by biosynthesis, and then adsorption (Fig. 1D and $P < 2.22e\text{-}16$ for all comparisons). We expected this result given that our gene set includes some 87 adsorption, 235 biosynthesis, and 8 assembly genes, and that most adsorption genes are core, while biosynthesis are not.

The process of collating putative phage resistance genes and matching them against a large number of genomes revealed the presence of a number of systems that were not previously described in *S. aureus*. Rare systems newly discovered in the species (*S. aureus*), listed with number of strains matching and taxon of original gene, included *abiR*, encoding an abortive infection system that inhibits phage DNA replication (27) (13 strains); *ietS*, encoding the serine protease toxin of the *IetAS* toxin-antitoxin system (28), (94 strains), and *ptuA*, encoding the ATPase component of the Septu phage defense system (29) (504 strains). Systems found in many more strains but newly reported here included Ec67/retron-TOPRIM (15372 strains), retron-TIR (15992 strains), and *avs1a*/MBL + protease-STAND (42123 strains) (Table S1). Retrons are RNA-DNA chimeric molecules synthesized by reverse transcriptases that induce cell death upon sensing infection (30, 31), while Avs1a is an AVAST (antiviral ATPases/NTPases of the STAND superfamily) protein (32). Although these genes are common in the species, their function was only recently discovered (30, 32). We further examined the genomic contexts of

noncore genes across 535 *S. aureus* complete genomes. We found these seven noncore genes (Fig. S1A-G) to be present near genes encoded by transposons as well as the chromosomally integrated mobile genetic element SCC*mec* (Fig. S1C) and restriction-modification genes (Fig. S1A, C, and D). This suggests these genes are readily transferred horizontally and associated with regions of genomic plasticity or previously known phage resistance determinants.

Diversity, functionality, and strength of selection of core phage resistance genes was similar to core genes in general. We next assessed the diversity and functionality of core phage resistance genes in the 380 genome Staphopia nonredundant diversity (NRD) set (26). The NRD set includes one randomly selected representative genome for every known *S. aureus* sequence type. We asked whether core 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) (33), 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 would also 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 genes and corresponding phage resistance gene subsets (P = 0.24, 0.92, not determined, and 0.12, respectively, from non-parametric Wilcoxon tests) (Fig. 2A to D). We further confirmed that codon usage did not differ between core and core phage resistance genes through comparison of codon adaptation index distributions for each set (Fig. 2E and P = 0.24), which suggests these gene sets have similar evolutionary histories.

Noncore phage resistance genes have greater phylogenetic signal than other noncore genes. Phylogenetic signal is the tendency of a trait to follow the pattern of an organismal phylogenetic tree (34). The phylogenetic signal may reveal features about patterns of acquisition and uptake of a noncore gene. We asked whether noncore phage resistance genes were unusual in their phylogenetic signal compared to other noncore 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). 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. Homoplasy—the independent loss or gain of a trait in separate lineages during evolution—is inversely proportional to consistency index and phylogenetic signal. To examine these relationships, we plotted the number of NRD strains encoding the gene of interest against the number of changes necessary for gene presence/absence to be consistent with the phylogeny, both for actual cases and the average of 999 gene presence-absence permutations per gene (Fig. 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. We found assembly genes had the highest changes among any group considered (Fig. 3), though noncore adsorption and biosynthesis genes also were significantly less consistent with the phylogeny than noncore genes in general (Fig. 3B). As the permuted presence-absence data suggest, gene frequency distribution, or how many genomes encoded a gene on average, could have an effect on the results, because by random chance intermediate frequency genes needed the largest number of presence/absence changes on the tree to be consistent with the phylogeny (Fig. 3A). We thus compared frequencies of genes in the data set (Fig. S2). Because noncore genes in general were far less frequent than any other category (P = 0.0046, 0.00013, and 2.2e-5, for adsorption, biosynthesis, and assembly genes relative to noncore genes, respectively), we would expect them to have fewer changes and thus higher phylogenetic signal (Fig. 3A) than intermediate frequency genes solely based on frequency in the genome set.

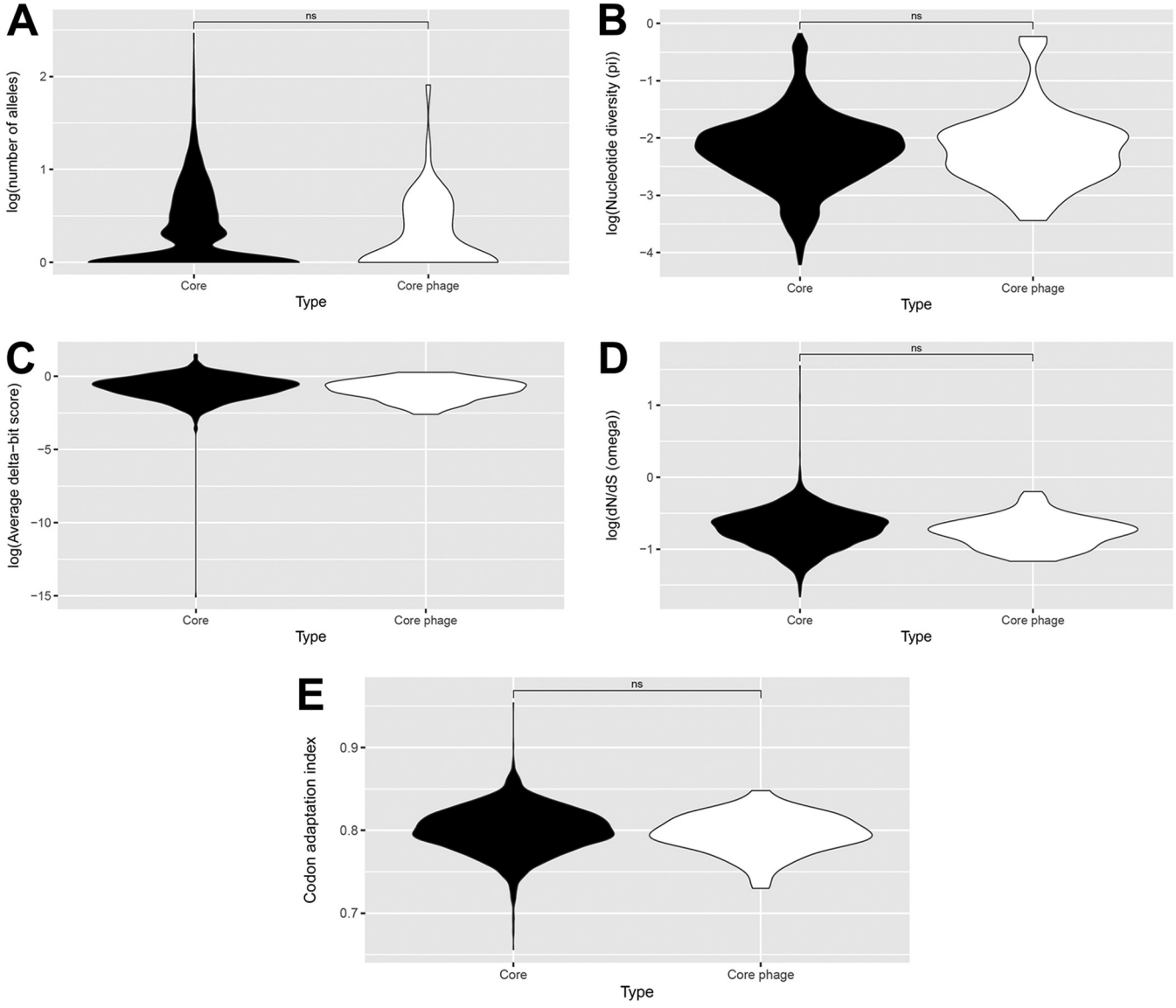

FIG 2 Core phage resistance genes do not differ from core genes overall in terms of diversity, functionality, selection, or codon usage. Core genes were those found in 80% or more of the NRD set. Diversity (A, B) was measured as the number of alleles or translated nucleotide diversity (π). Functionality (C) was measured by delta-bit score, selection (D) was measured through the dN/dS metric, and codon usage (E) was measured through codon adaptation index. 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). Delta-bit scores did not provide enough data points for a Wilcoxon or *t* test to be conducted.

Taken together, these results indicate all noncore phage resistance genes had lesser phylogenetic signal than other noncore genes with assembly genes having lowest signals. Gene in the assembly category are frequently found on mobile SaPIs, which are frequently exchanged by HGT and lost through deletion. Unlike assembly genes, some noncore adsorption and biosynthesis genes approach complete consistency with the tree, but the proportion is lower than other noncore genes. However, we cannot 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.

We further examined the relationship between phylogeny and phage resistance genes by assessing the differences between clonal complexes (CC) in their numbers of noncore phage genes. For all gene categories examined, there was statistically significant variation ($P < 1e\text{-}5$ in all cases) as determined by an analysis of variance (ANOVA) test (Fig. 4A). We found significantly higher ($P < 2.22e\text{-}16$) noncore phage resistance gene

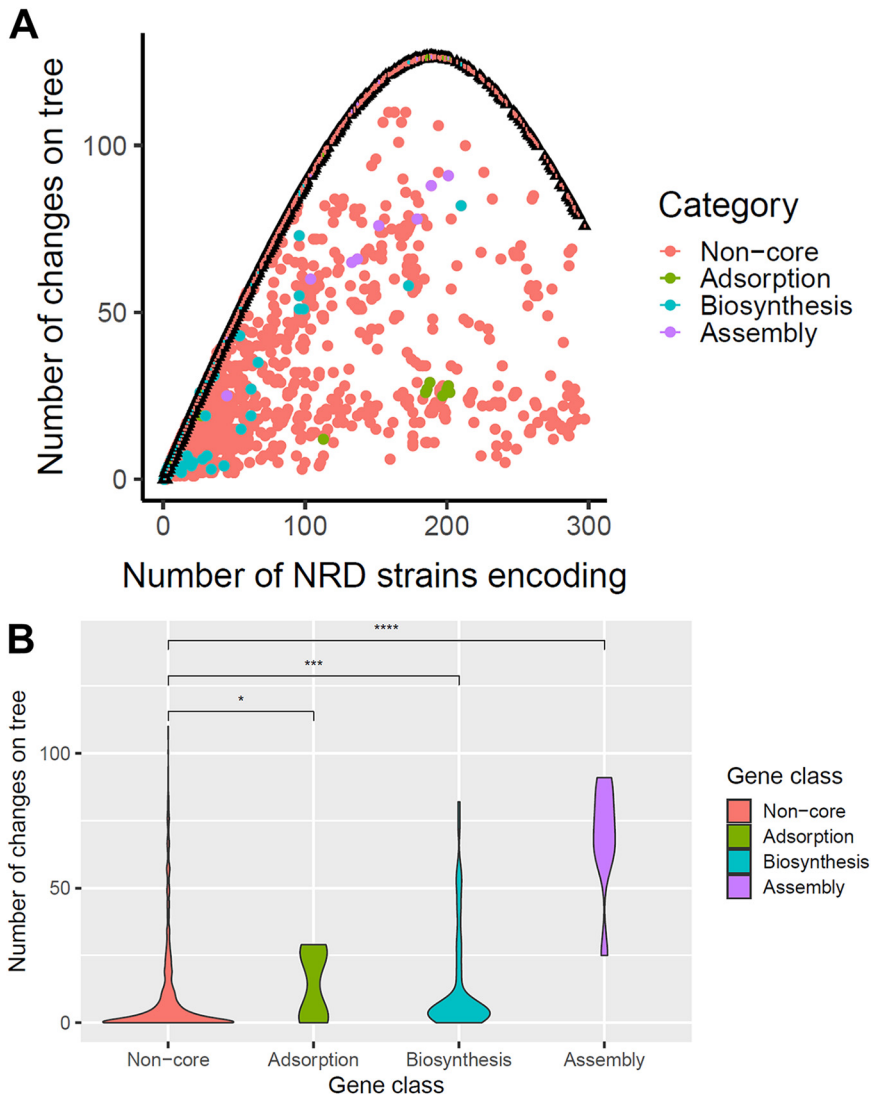

FIG 3 All noncore phage resistance genes have phylogenetic signal, but assembly genes have the least phylogenetic signal among all categories. (A) The relationship between the number of changes necessary to make gene presence/absence patterns consistent with the phylogeny (*y* axis) and the number of strains encoding each gene (*x* axis). Average change numbers after 999 permutations of gene presence/absence on the tree are shown as triangles with thick black borders and error bars (1 standard error above and below the mean), while change numbers for actual gene presence/absence are shown as circles. Noncore 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 (noncore, adsorption, biosynthesis, and assembly) visualized as violin plots. Group differences were tested for significance using the nonparametric 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).

concordance within CCs relative to between CCs (Fig. 4B). This suggests clonal complex is associated with both accessory genome and numbers of phage resistance genes.

Noncore 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 concordance" in each category (e.g., noncore adsorption, biosynthesis, or assembly genes). We calculated genomic concordance as the average number of subject genes in a particular category encoded by strains containing the query gene. 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. We found that all phage resistance genes (noncore, all

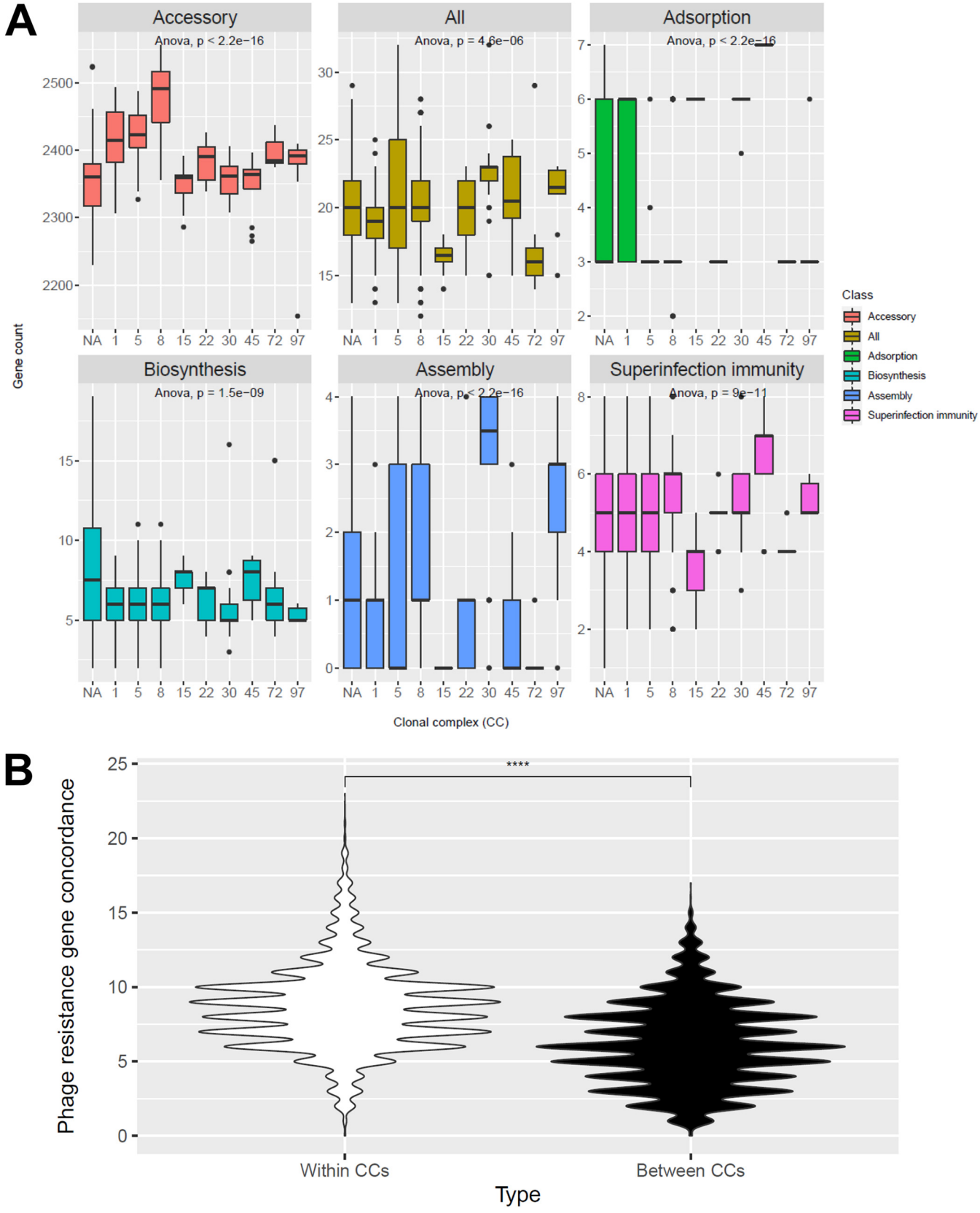

FIG 4 Relationship between clonal complex (CC) and accessory genome and noncore phage resistance gene content. We used BLAST to search for our set of noncore phage resistance genes and accessory genes in the set of 535 complete *S. aureus* genomes in Staphopia. We then visualized the distributions of genes for each CC as boxplots (A). Strains without a defined CC are listed as NA. Analysis of variance (ANOVA) assessed significant overall differences, with significance values posted on each facet. We also compared concordances of noncore phage resistance gene presence between CCs and within CCs, showing more such genes were shared within CCs than between CCs (B).

phage resistance, biosynthesis; P = 2.4e-5, 0.00034, and 0.00136, respectively) or biosynthesis genes (noncore, all phage resistance, adsorption, biosynthesis, assembly; P = 6.3e-6, 0.00119, 0.03988, 1.7e-5, and 0.00335, respectively) were most often significantly different (non-parametric Wilcoxon test) from noncore genes (Fig. S3) in subject genomic concordance distribution (subject listed here in parentheses). In each phage resistance gene category, when query and subject were the same class, query genomic concordance was most significantly different (P = 2.5e-9 for adsorption, 1.7e-5 for biosynthesis, and 8.2e-5 for assembly) from noncore gene query concordance and increased relative to noncore gene query concordance in all cases (Fig. S3). This supported our hypothesis that genes with similar categories were encoded together, as adsorption, biosynthesis, and assembly genes were encoded together significantly more often than similar noncore genes at random.

Superinfection immunity had weak correlation with empirically determined phage resistance and accessory genome content. We next asked whether phage resistance gene presence could explain two relevant phenotypes—empirical resistance to infection against 8 phages in 259 diverse *S. aureus* strains determined by high-throughput laboratory turbidity assays that measure the ability of bacteria to survive phage challenge (35), and variation in the total number of accessory genes (as a proxy for levels of horizontal gene transfer) 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) (36) correlated positively with temperate (p11 – R^2=0.051/P = 3.1e-4, 0.043/9.7e-4, and 0.021/0.022, respectively, and p0040 – R^2=0.04/P = 1.3e-3, 0.04/1.3e-3, and 0.016/0.046, respectively) phage resistance (Fig. 5). This result confirmed the hypothesis that prophages confer temperate phage resistance through superinfection immunity, although the effect size was low. We also saw that total accessory genome content was positively correlated with superinfection immunity genes (R^2=0.26/P < 2.2e-16; Fig. 6) while all phage resistance genes had only a weak positive correlation (R^2=0.0068/P = 0.056). 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 noncore adsorption, nonsuperinfection immunity biosynthesis, or assembly gene counts (Fig. S4), suggesting none of these factors prevented acquisition of superinfection immunity genes through lysogeny. We did note that adsorption gene count correlated slightly with decreased accessory genome content (Fig. 6; R^2=0.19/P < 2.2e-16). All relationships were confirmed to be robust through a permutation test—all regressions between permuted accessory genome content and respective categories had P values between 0.12 and 0.75.

To account for quantitation effects of multigene systems, we compressed matches from gene to system (counting a "system" like restriction-modification to be present if a single system gene was present, such as *hsdR*). We also accounted for the small number of genes that act to reduce phage resistance by multiplying gene presence by a resistance effect factor (1 if gene presence conferred resistance, −1 if gene presence conferred sensitivity). We largely found the same patterns as we did without these corrections (Supplemental Fig. S5-S6). Unlike the previous accessory genome analysis, however, we found a positive correlation between biosynthesis gene count, number of systems, or effect and accessory genome content (Supplemental Fig. S6A-C; R^2=0.19/P < 2.2e-16, 0.097/1.9e-13, and 0.19/<2.2e-16, respectively) when we considered data from the gene presence-absence matrix.

Relationship between noncore phage resistance genes and horizontal gene transfer, antibiotic resistance, and virulence. Phages are a major conduit for gene acquisition in *S. aureus* and therefore it is possible that strains with greater phage resistance may have reduced uptake of horizontally transferred virulence and antibiotic resistance genes. We asked whether phage resistance acted as a barrier to antibiotic resistance and virulence gene acquisition by 1) correlating counts of phage resistance genes with noncore antibiotic resistance and virulence gene counts 2) calculating phylogenetic concordance between phage resistance genes of each category and noncore genes or subsets of antibiotic resistance or virulence genes, and 3) calculating genomic concordance between the same sets as 2). We found that

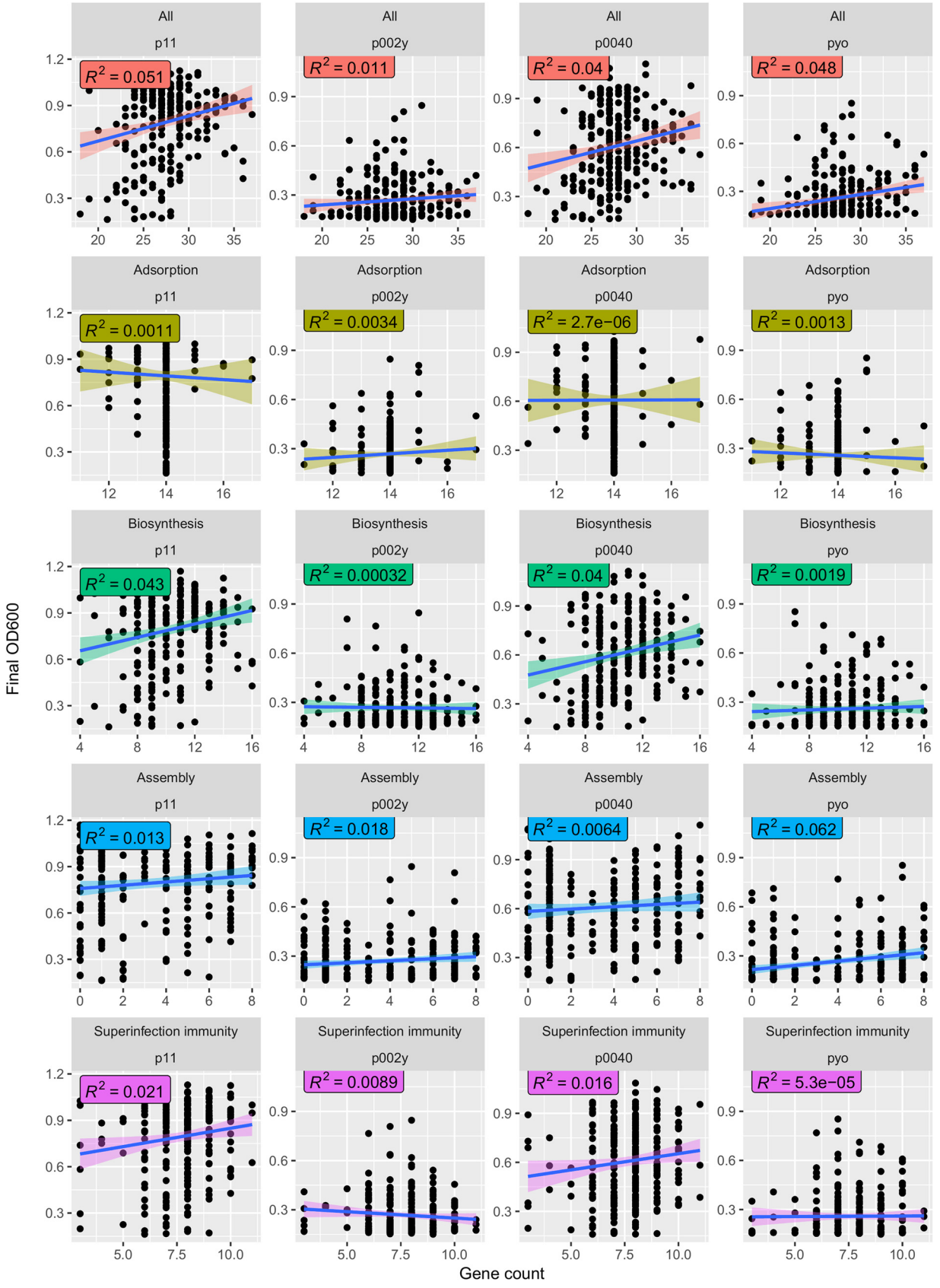

FIG 5 Superinfection immunity but neither adsorption nor assembly genes correlates with empirical temperate phage resistance. We used BLAST to search for our set of noncore phage resistance genes in the set of 263 *S. aureus* genomes from our recent phage host range study.

(Continued on next page)

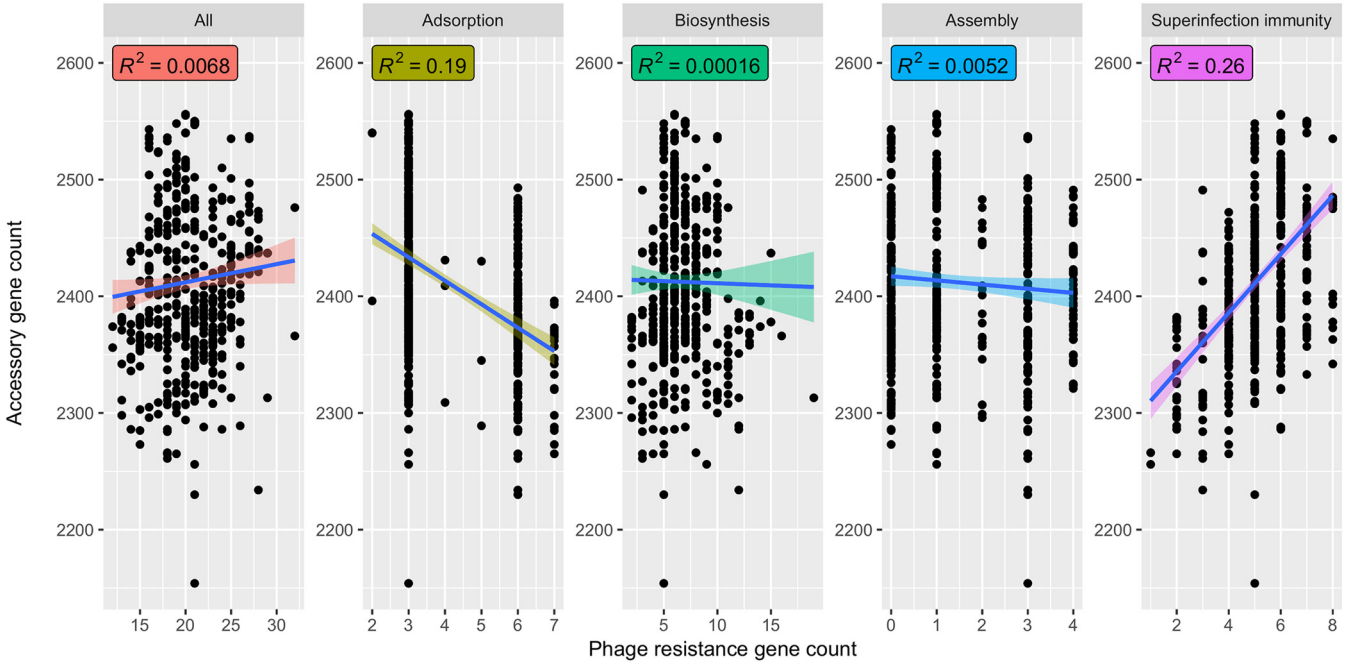

FIG 6 Superinfection immunity but neither adsorption nor assembly genes correlates with accessory genome content. We used BLAST to search for our set of noncore phage resistance genes in the set of 535 complete *S. aureus* genomes in Staphopia. We then plotted the number of matches to all noncore phage resistance, adsorption, biosynthesis, assembly, or superinfection immunity genes (*x* axis) against accessory genome content (*y* axis) and calculated correlations (R^2) between each.

total noncore phage resistance and biosynthesis genes positively correlated (R^2=0.072, P = 2.9e-10; R^2=0.099, P = 8.4e-14) with the noncore antibiotic resistance genes, but all categories positively correlated (R^2=0.18/P < 2.2e-16; 0.43/<2.2e-16; 0.047/3.6e-07; 0.15 < 2.2e-16; 0.023/4.4e-4 for all, adsorption, biosynthesis, assembly, and superinfection immunity genes, respectively) with noncore virulence genes (Fig. S7). All relationships were confirmed to be robust through a permutation test—all regressions between permuted noncore antibiotic resistance/virulence gene content and respective categories had P values between 0.1 and 0.93. In terms of phylogenetic concordance, assembly genes were co-encoded with antibiotic resistance and virulence genes in more strains than noncore genes on average, while adsorption genes were co-encoded with virulence genes in far fewer strains than noncore genes on average (P = 0.00290, 0.00274, and 0.00256, respectively, Wilcoxon test, Fig. S8). Regarding genomic concordance, more antibiotic resistance genes were encoded by adsorption gene-encoding strains than noncore gene-encoding strains, while more virulence genes were encoded for all phage resistance and biosynthesis gene-encoding strains than noncore gene-encoding strains (P = 0.00373, 0.00025, and 4.8e-5, respectively, Wilcoxon test, Fig. S3). These phylogenetic and genomic concordance results thus indicate noncore antibiotic resistance genes tended to be encoded together with adsorption genes, while noncore virulence genes tended to be encoded together with biosynthesis genes, although effect sizes are weak.

Consequences of type I restriction-modification specificity gene (hsdS) allelism on horizontal gene transfer, antibiotic resistance, and virulence. Type I restriction systems are well known as a significant barrier to HGT in *S. aureus* (37, 38). Although core genes, the barrier is dependent on DNA binding specificity of their hsdS DNA binding domain allele. We analyzed type I restriction-modification specificity gene (*hsdS*) alleles for their relationships with HGT, given their correlation with clonal complex and role in restricting

FIG 5 Legend (Continued)
We then plotted the number of matches to all noncore phage resistance, adsorption, biosynthesis, assembly, or superinfection immunity genes (*x* axis) against previously measured phage resistance phenotypes (OD_{600} or turbidity after coculture; *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).

transduction between clonal complexes based on restriction specificity (37). We calculated phylogenetic and genomic concordances between detected NRD *hsdS* alleles (81 total) and all noncore genes or respective antibiotic resistance and virulence gene subsets. In all cases, *hsdS* allele phylogenetic concordance was significantly less ($P <$ 2e-16, P = 3.2e-16, 2.1e-15, <2e-16, <2e-16, <2e-16, <2e-16, and 1.8e-12, Wilcoxon test) than that of noncore genes generally (Fig. S8). However, only for subject noncore, adsorption, biosynthesis, and antibiotic resistance genes were *hsdS* allele genomic concordance significantly different (P = 7.3e-7, 0.00012, 0.02711, and 6.4e-8, respectively, Wilcoxon test) than that of noncore genes generally (Fig. S3). 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. This is consistent with the hypothesis that phylogeny (CC)-associated *hsdS* alleles restrict horizontal gene transfer.

DISCUSSION

In this work, we curated genes potentially influencing phage resistance in *S. aureus* and mapped them against the thousands of genomes in the Staphopia database. We then used these results to ask questions about the relation of such genes to empirically determined phage resistance and horizontally transferred antibiotic and virulence genes. 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), retrons, and Lamassu systems, among others. This study is the first to note the recently discovered Septu, AVAST, and retron phage defense systems (32) in *S. aureus*. Assembly genes examined include all three main mechanisms known in SaPIs- capsid remodeling, packaging interference, and helper phage late gene repression.

We noted four major findings from these analyses. First, we found that core phage resistance genes do not significantly differ in diversity, functionality, and selection from other core genes, refuting a possible arms race hypothesis in the evolution of host receptors and phage. This suggested a different evolutionary dynamic to that seen in *E. coli*, where phages specific to strains in that species often exist in an arms race dynamic with the outer-surface of the bacterium, in which phage and host coevolve to outcompete each other (39, 40). 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 phage resistance genes have similar high functionality (low delta-bit score), low diversity, and negative selection to 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 (41–43), methicillin resistance (44), nasal colonization (45), and antimicrobial peptide resistance (46), among other roles. Alternatively, our strain set may not capture transient strains resistant through core gene mutation (especially unstable mutations as in phase variation) (47), or noncore genes may instead be undergoing arms race dynamics through frequent gain or loss (48).

A second finding was that noncore phage resistance genes had less phylogenetic signal (measured using consistency index) than other noncore genes. Frequent loss or HGT of these genes or a preponderance of intermediate frequency phage resistance genes would explain this finding. Of the classes of resistance genes, those in the assembly category had the lowest consistency. This may be expected as genes found on *Staphylococcus aureus* pathogenicity islands (SaPIs) are in this category and frequent HGT of SaPIs through transduction assisted by helper phages may lead to inconsistencies with the core genome phylogeny.

Third, we found that noncore phage resistance genes within the same categories are encoded together (e.g., adsorption) by measuring genomic concordance. Thus, if a

strain encoded a noncore adsorption gene, for example, it was more likely to encode a gene of the same category than a strain lacking this gene. 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 (29). As was previously noted (29, 32), we also found that noncore recently discovered phage resistance systems were sometimes located near restriction-modification genes. This is likely due to as yet 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). This was a relatively simplistic analysis that may be confounded to an extent by phylogenetic effects, but it bears future exploration. An example would be determining how many genes are part of phage defense islands (29) in the same genomic region to further address this clustering.

The fourth finding was that superinfection immunity among noncore 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 (49).

After discovery in this analysis of some species-wide patterns of phage resistance, the future challenge is to connect them to phenotypes and evolutionary processes. Somewhat surprisingly, simple models of phage resistance gene presence did not predict experimentally determined phage resistance nor measures of HGT such as virulence and antibiotic resistance accessory gene numbers. There are several possible reasons why the situation may be more complicated than expected. We cannot 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) (50, 51). It is possible that existing studies characterizing phage resistance genes in *S. aureus* do not reflect growth conditions (e.g., rich medium versus 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 in some cases. It may be that cryptic phage resistance loci not yet characterized are important in predicting each phenotype. It is possible that complete resistance to virulent phage infection is extremely rare and instead subtle metabolic or surface protein changes are 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 (48). 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 concordance between *hsdS* alleles and various noncore genes of interest, for example, finding significantly lower concordance than for noncore genes in general, and a bimodal distribution with a subset of alleles having lower and higher concordance. 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.

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

Curated list of *S. aureus* phage resistance genes. The list of genes known to influence phage resistance was collated from extensive literature review. We previously searched the literature exhaustively for host genes identified through laboratory evolution or molecular genetics to act at three stages of phage infection (adsorption, biosynthesis, and assembly) (4). Such genes represented well characterized mechanisms of phage resistance both in *S. aureus* and other bacteria. Adsorption genes included those implicated in phage receptor biosynthesis (wall teichoic acid) and capsule biosynthesis, biosynthesis genes included restriction-modification, abortive infection, superinfection immunity, and CRISPR/Cas genes, and assembly genes included capsid remodeling, phage packaging interference, and late phage life cycle repressor genes. 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 reported *S. aureus* homologs (to avoid redundancy). An example of the final criterion was the inclusion of the *Lactococcus lactis* superinfection exclusion (*sie*) uptake resistance gene (52), which is not reported in *S. aureus* nor contains a known *S. aureus* homolog. 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. We acknowledge our list does not include genes that lack a demonstrated role in phage resistance but may nonetheless influence it (e.g., metabolism genes). It also fails to include genes that do directly affect phage resistance but have yet to be discovered in *S. aureus* or other bacterial species.

Determining phage resistance gene conservation in the *S. aureus* species. Nucleotide sequences of the curated phage resistance gene list were matched against 42,949 *S. aureus* genomes (the Staphopia database) (26) using a BLASTN (53, 54) 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 (conservation) and the number of unique matches per strain to determine the number of genes encoded by each stain. 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 nonparametric Wilcoxon tests.

Visualizing genomic contexts of noncore recently discovered phage resistance systems. In addition to evaluating conservation of phage resistance genes, we examined the genomic surroundings of the most recently discovered phage resistance systems in our set (mainly retrons and cyclic oligonucleotide-based antiphage signaling systems—CBASS) (32, 55). We first constructed the pangenome of the Staphopia completed genome set (535 genomes) with PIRATE (56). We then matched phage resistance gene sequences (Table S1) from the two previously cited papers (32, 55) against noncore (present in 80% or less of the genomes) PIRATE pangenome clusters and identified up to the first five genomes containing this noncore gene subset (seven genes total). Five pseudogene or CDS sequences up or downstream of the noncore phage resistance gene of interest were selected from the respective genome GFF file for each genome and each noncore phage resistance gene of interest. Noncore genes and their genomic contexts obtained as described were then visualized using the R package gggenes (57).

Determining core phage resistance gene diversity, functionality, and selection. The pangenome of the Staphopia nonredundant (NRD) set (380 strains representing each sequence type) (26) was constructed using PIRATE (56) run with default parameters. Core genes were those unique PIRATE gene clusters only present in 80% of NRD genomes (304) or more. We focused gene diversity, functionality, and selection studies on these core phage resistance genes, comparing these sets against corresponding total core genes (excluding 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 (58). We evaluated functionality using delta-bit score (33) and measured selection by calculating dN/dS for each gene with the package Hypothesis Testing using Phylogenies (HyPhy) (59). 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 gene. Amino acid diversity (π) was calculated from translated PIRATE gene cluster nucleotide alignments. dN/dS was calculated using individual gene phylogenetic trees inferred from nucleotide alignments with IQ-TREE (60, 61). dN/

dS was calculated using single-likelihood ancestor counting (SLAC), which uses maximum likelihood (ML) and counting approaches to infer dN and dS per site given a codon alignment and corresponding phylogeny (59, 62, 63). SLAC first predicts the most likely ancestral sequence at each node of the phylogeny with ML and then counts nonsynonymous and synonymous changes per site in a manner similar to the Suzuki-Gojobori method (63, 64). For delta-bit score analysis, for each tested gene, we calculated average delta-bit scores for all strains encoding protein sequences that matched a corresponding HMM. Proteins were matched to Pfam family HMMs using HMMER (65). We also compared codon usage between core phage resistance gene protein sequences and core gene protein sequences using the codon adaptation index (cai) tool of the EMBOSS suite (66).

Evaluating phylogenetic associations with noncore phage resistance genes. We evaluated phylogenetic associations with phage resistance genes by 1) determining homoplasy for each noncore phage resistance gene and 2) correlating clonal complex (CC) with noncore phage resistance gene count. Noncore 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 (67) 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 (68)-recombination corrected PIRATE core genome alignment. Consistency index, or the consistency between a character among 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 nonparametric 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 noncore 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. We also determined concordances of noncore phage resistance gene presence/absence (number of genes co-encoded by two different strains) among all complete genome strains that had an identified CC to determine whether phage resistance genes were more shared within CCs than between different CCs.

Noncore resistance gene correlation analyses with empirical phage resistance and accessory genome content. We examined the relationship between noncore phage resistance genes and 1) experimentally measured phage resistance phenotypes, 2) accessory genome content, and 3) 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 (35) and genomes of all completely assembled *S. aureus* genomes (26) 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 (69), whereas virulence genes searched were the Virulence Factor Database (VFDB) set (70). As defined for the NRD set, noncore phage resistance, antibiotic resistance, and virulence genes were defined as those unique PIRATE gene cluster matches present in 80% of the respective genome set or less (complete or GWAS). We then used BLAST to match these three sets of noncore 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 noncore antibiotic resistance or virulence gene matches normalized by accessory genome content. Linear regressions were performed on each distribution to assess correlations between these variables. Accessory genome content and noncore antibiotic resistance or virulence regressions were compared with those performed on permuted data (999 permutations of accessory genome content or normalized antibiotic resistance and virulence gene counts relative to all other data) to evaluate robustness of the relationships observed.

In addition, we performed two subsequent analyses in which we converted phage resistance gene matches to the number of systems and net phage resistance effects. These analyses relied on the PIRATE gene presence-absence matrices (complete or GWAS) rather than BLAST matches to genes of interest. Number of systems per strain was counted as the number of subclasses (Table S1) for which there was a match to at least one gene. Net phage resistance effect on the other hand was determined for each strain by multiplying the phage resistance gene presence-absence matrix by a vector containing the resistance effect direction (Table S1) of the presence of each phage resistance gene (+1 for more resistance and −1 for more sensitivity if a gene is present). Number of systems or net phage resistance effect was then plotted against empirically measured phage resistance (GWAS set), accessory genome content, or noncore antibiotic resistance or virulence gene matches. Linear regressions were performed on each distribution to assess correlations between these variables.

Calculating noncore phage resistance gene phylogenetic and genomic concordance. In addition to assessing correlations between noncore phage resistance gene counts and accessory genome content on a per-strain level, we also evaluated strain and gene level concordance between noncore 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 data set. 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 values. 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 noncore genes (phage resistance genes from the complete list present in less than 80% of Staphopia genomes) were considered for further analysis. Genomic concordance per gene was calculated as the average number of genes in a subject category (e.g., adsorption) encoded by strains encoding the query gene of interest, while phylogenetic concordance per gene was calculated as the total number of genes in a subject category encoded by all strains encoding the query gene of interest divided by the number of genes in that subject category. Genomic concordance measures how many genes of a certain type are co-encoded with a gene of interest on average, while phylogenetic concordance measures how many strains co-encode a gene and all those of a certain type on average. Genomic and phylogenetic concordance were compared between groups with violin plots and significant differences assessed with nonparametric Wilcoxon tests. For the *hsdS* gene, this analysis was repeated with the 81 alleles detected in the NRD set PIRATE pangenome. Genomic and phylogenetic concordance distributions were compared with violin plots and significant differences assessed using nonparametric Wilcoxon tests.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.2 MB.

FIG S2, PDF file, 0.02 MB.

FIG S3, PDF file, 0.2 MB.

FIG S4, PDF file, 0.02 MB.

FIG S5, PDF file, 0.4 MB.

FIG S6, PDF file, 0.5 MB.

FIG S7, PDF file, 0.3 MB.

FIG S8, PDF file, 0.2 MB.

TABLE S1, XLSX file, 0.03 MB.

ACKNOWLEDGMENTS

We thank members of the lab—Michelle Su, Emily Wissel, Robert Petit, and Ashley Alexander—for constructive criticism. Timothy D. Read was supported by United States National Institutes of Health (NIH) R21 AI 138079-02, while Abraham G. Moller was supported by the United States National Science Foundation (NSF) Graduate Research Fellowship Program (GRFP).

REFERENCES

1. Pirisi A. 2000. Phage therapy—advantages over antibiotics? Lancet 356:1418. https://doi.org/10.1016/S0140-6736(05)74059-9.
2. Nobrega FL, Costa AR, Kluskens LD, Azeredo J. 2015. Revisiting phage therapy: new applications for old resources. Trends Microbiol 23:185–191. https://doi.org/10.1016/j.tim.2015.01.006.
3. Xia G, Wolz C. 2014. Phages of Staphylococcus aureus and their impact on host evolution. Infect Genet Evol 21:593–601. https://doi.org/10.1016/j.meegid.2013.04.022.
4. Moller AG, Lindsay JA, Read TD. 2019. Determinants of phage host range in Staphylococcus species. Appl Environ Microbiol 85:e00209-19. https://doi.org/10.1128/AEM.00209-19.
5. 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 102:1–15.
6. 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. https://doi.org/10.1128/AEM.01385-16.
7. Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. Nat Rev Microbiol 8:317–327. https://doi.org/10.1038/nrmicro2315.
8. Hyman P, Abedon ST. 2010. Bacteriophage host range and bacterial resistance. Adv Appl Microbiol 70:217–248. https://doi.org/10.1016/S0065-2164(10)70007-1.
9. Wilkinson BJ, Holmes KM. 1979. Staphylococcus aureus cell surface: capsule as a barrier to bacteriophage adsorption. Infect Immun 23:549–552. https://doi.org/10.1128/iai.23.2.549-552.1979.
10. Nordström K, Forsgren A. 1974. Effect of protein A on adsorption of bacteriophages to Staphylococcus aureus. J Virol 14:198–202. https://doi.org/10.1128/JVI.14.2.198-202.1974.
11. 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 563:705–709. https://doi.org/10.1038/s41586-018-0730-x.
12. 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. https://doi.org/10.1128/JB.01412-10.
13. 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 5:17219. https://doi.org/10.1038/srep17219.
14. 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. https://doi.org/10.1128/jb.98.2.519-527.1969.
15. 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. https://doi.org/10.1016/j.chom.2016.08.010.
16. 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. https://doi.org/10.1128/AAC.01689-12.
17. 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. Int J Med Microbiol 306:686–696. https://doi.org/10.1016/j.ijmm.2016.08.005.
18. 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. Biochem Biophys Res Commun 464:894–900. https://doi.org/10.1016/j.bbrc.2015.07.062.
19. Zhao X, Yu Z, Xu Z. 2018. Study the features of 57 confirmed CRISPR loci in 38 strains of Staphylococcus aureus. Front Microbiol 9.
20. 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. https://doi.org/10.1016/j.virol.2012.05.026.
21. 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. Proc Natl Acad Sci U S A 109:16300–16305. https://doi.org/10.1073/pnas.1204615109.
22. Christie GE, Dokland T. 2012. Pirates of the caudovirales. Virology 434:210–221. https://doi.org/10.1016/j.virol.2012.10.028.
23. 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. https://doi.org/10.1073/pnas.1406749111.
24. Novick RP, Christie GE, Penadés JR. 2010. The phage-related chromosomal islands of Gram-positive bacteria. Nat Rev Microbiol 8:541–551. https://doi.org/10.1038/nrmicro2393.
25. 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. https://doi.org/10.1016/j.jmb.2008.04.065.
26. Petit RA, III, Read TD. 2018. Staphylococcus aureus viewed from the perspective of 40,000+ genomes. PeerJ 6:e5261. https://doi.org/10.7717/peerj.5261.
27. Twomey DP, De Urraza PJ, McKay LL, O'Sullivan DJ. 2000. Characterization of AbiR, a novel multicomponent abortive infection mechanism encoded by plasmid pKR223 of Lactococcus lactis subsp. lactis KR2. Appl Environ Microbiol 66:2647–2651. https://doi.org/10.1128/AEM.66.6.2647-2651.2000.
28. Yamamoto S, Kiyokawa K, Tanaka K, Moriguchi K, Suzuki K. 2009. Novel toxin-antitoxin system composed of serine protease and AAA-ATPase homologues determines the high level of stability and incompatibility of the tumor-inducing plasmid pTiC58. J Bacteriol 191:4656–4666. https://doi.org/10.1128/JB.00124-09.
29. 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 359. https://doi.org/10.1126/science.aar4120.
30. Millman A, Bernheim A, Stokar-Avihail A, Fedorenko T, Voichek M, Leavitt A, Oppenheimer-Shaanan Y, Sorek R. 2020. Bacterial retrons function in anti-phage defense. Cell 183:1551–1561.e12. https://doi.org/10.1016/j.cell.2020.09.065.
31. Otto G. 2021. A function for retrons. Nat Rev Microbiol 19:3–3. https://doi.org/10.1038/s41579-020-00488-2.
32. 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. https://doi.org/10.1126/science.aba0372.
33. 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. https://doi.org/10.1093/bioinformatics/btw518.
34. Kamilar JM, Cooper N. 2013. Phylogenetic signal in primate behaviour, ecology and life history. Philos Trans R Soc Lond B Biol Sci 368:20120341. https://doi.org/10.1098/rstb.2012.0341.
35. 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 6https://doi.org/10.1128/mSphere.01263-20.
36. Berngruber TW, Weissing FJ, Gandon S. 2010. Inhibition of superinfection and the evolution of viral latency. J Virol 84:10200–10208. https://doi.org/10.1128/JVI.00865-10.
37. 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. https://doi.org/10.1128/JB.00418-06.
38. 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. https://doi.org/10.1093/nar/gkx067.
39. Stern A, Sorek R. 2011. The phage-host arms-race: shaping the evolution of microbes. Bioessays 33:43–51. https://doi.org/10.1002/bies.201000071.
40. Weitz JS, Hartman H, Levin SA. 2005. Coevolutionary arms races between bacteria and bacteriophage. Proc Natl Acad Sci U S A 102:9535–9540. https://doi.org/10.1073/pnas.0504062102.
41. 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. Mol Microbiol 75:864–873. https://doi.org/10.1111/j.1365-2958.2009.07007.x.
42. Winstel V, Xia G, Peschel A. 2014. Pathways and roles of wall teichoic acid glycosylation in Staphylococcus aureus. Int J Med Microbiol 304:215–221. https://doi.org/10.1016/j.ijmm.2013.10.009.
43. 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 Pathog 15:e1008044. https://doi.org/10.1371/journal.ppat.1008044.
44. 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. Proc Natl Acad Sci U S A 109:18909–18914. https://doi.org/10.1073/pnas.1209126109.
45. 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. Nat Med 10:243–245. https://doi.org/10.1038/nm991.
46. 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. https://doi.org/10.1074/jbc.274.13.8405.
47. Berryhill BA, Huseby DL, McCall IC, Hughes D, Levin BR. 2021. Evaluating the potential efficacy and limitations of a phage for joint antibiotic and phage therapy of Staphylococcus aureus infections. Proc Natl Acad Sci U S A 118. https://doi.org/10.1073/pnas.2008007118.
48. 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.
49. 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. https://doi.org/10.1126/science.aat5867.
50. Tock MR, Dryden DT. 2005. The biology of restriction and anti-restriction. Curr Opin Microbiol 8:466–472. https://doi.org/10.1016/j.mib.2005.06.003.
51. Pawluk A, Davidson AR, Maxwell KL. 2018. Anti-CRISPR: discovery, mechanism and function. Nat Rev Microbiol 16:12–17. https://doi.org/10.1038/nrmicro.2017.120.
52. 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. https://doi.org/10.1128/AEM.01053-08.
53. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
54. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
55. Millman A, Melamed S, Amitai G, Sorek R. 2020. Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. Nat Microbiol 5:1608–1608. https://doi.org/10.1038/s41564-020-0777-y.
56. 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. https://doi.org/10.1093/gigascience/giz119.
57. Wilkins D, Kurtz Z. 2020. gggenes: draw Gene Arrow Maps in "ggplot2."

58. Lees J. 2019. Conservation of core genes in S. pneumoniae. John Lees.
59. Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679. https://doi.org/10.1093/bioinformatics/bti079.
60. 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. https://doi.org/10.1093/molbev/msu300.
61. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281.
62. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. 2018. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Mol Biol Evol 35:773–777. https://doi.org/10.1093/molbev/msx335.
63. Kosakovsky Pond SL, Frost SDW. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22:1208–1222. https://doi.org/10.1093/molbev/msi105.
64. Suzuki Y, Gojobori T. 1999. A method for detecting positive selection at single amino acid sites. Mol Biol Evol 16:1315–1328. https://doi.org/10.1093/oxfordjournals.molbev.a026042.
65. Eddy SR. 2009. A new generation of homology search tools based on probabilistic inference, p 205–211. In Genome Informatics 2009. PUBLISHED BY IMPERIAL COLLEGE PRESS AND DISTRIBUTED BY WORLD SCIENTIFIC PUBLISHING CO.
66. Rice P. 2000. EMBOSS: the European Molecular Biology Open Software Suite 2.
67. Crispell J, Balaz D, Gordon SV. 2019. HomoplasyFinder: a simple tool to identify homoplasies on a phylogeny. Microbial Genomics 5:e000245. https://doi.org/10.1099/mgen.0.000245.
68. 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. https://doi.org/10.1093/nar/gku1196.
69. 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. https://doi.org/10.1128/JCM.03117-13.
70. Liu B, Zheng D, Jin Q, Chen L, Yang J. 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res 47:D687–D692. https://doi.org/10.1093/nar/gky1080.