Reviewed Preprint
v1 • February 19, 2026
Not revised

For correspondence: tread@emory.edu

* Co-last authors

Competing interests: No competing interests declared

Funding: See page 22

Reviewing editor: Ben S Cooper, University of Oxford, United Kingdom

Transition of *Staphylococcus aureus* tetracycline resistance plasmid pT181 from independent multicopy replicon to predominantly integrated chromosomal element over 65 years

Megan A Phillips[1], Robert A Petit[2], Daniel B Weissman[3,*], Timothy D Read[4,*]

[1]Population Biology, Ecology, and Evolution Program, Graduate Division of Biological and Biomedical Sciences, Laney Graduate School, Emory University, Atlanta, United States • [2]Wyoming Public Health Laboratory, Cheyenne, United States • [3]Department of Physics, Emory University, Atlanta, United States • [4]Division of Infectious Diseases, Department of Medicine, Emory University, Atlanta, United States

eLife Assessment

Using genome databases, the authors performed **solid** bioinformatic analyses to trace the genomic history of the clinically relevant Staphylococcus aureus tetracycline resistance plasmid pT181 over the last seven decades. They discovered that this element has transitioned from a multicopy plasmid to a chromosomally integrated element, and the work represents a **valuable** demonstration of the use of publicly available data to investigate plasmid biology and inform clinical epidemiology. This work will appeal to researchers interested in staphylococcal evolution and plasmid biology.

https://doi.org/10.7554/eLife.109393.1.sa3

Abstract

Mobile genetic elements (MGEs), including plasmids, phages and genome islands, are major sources of bacterial genetic diversity. The small plasmid pT181 confers tetracycline resistance in bacterial pathogen *Staphylococcus aureus* via an efflux pump, TetK. pT181 was one of the earliest sequenced *S. aureus* plasmids, and has been isolated in both clinical and livestock-associated strains for decades, both as an independent replicon and integrated in the chromosome as part of staphylococcal cassette chromosome *mec* (SCC*mec*). Bacterial genome analysis tools and high-quality sequences with metadata are publicly available, but these resources remain underleveraged for examining historical data, especially when studying the spread of MGEs across a species and over time. Using publicly available reads and metadata, we explored the evolution of pT181 over almost seven decades of samples to identify temporal trends in sequence evolution, copy number changes, and spread across *S. aureus* and beyond. pT181 was prevalent across *S. aureus* (found in 9.5% of 83,366 genomes tested), with a conserved sequence outside of three hypervariable regions. The history of pT181 since 1954 is characterized by spread across strains, significant variation in plasmid copy number of the independent replicon, and increasing frequency of integration of the plasmid into the *S. aureus* chromosome. We have identified multiple chromosomal integration locations of the plasmid, including outside of the previously characterized SCC*mec*. We find that pT181 has been transferred across staphylococcaceae and into a Gram-negative species. The repeated integration of pT181 into the chromosome may indicate co-evolution of the plasmid and the host, potentially to facilitate increased antibiotic resistance.

Introduction

Plasmids are extrachromosomal DNA molecules capable of autonomous replication that play a major role in horizontal gene transfer (HGT) in many bacterial species. Plasmids have been the most important vectors for the spread of antibiotic resistance genes (ARGs) within bacterial pathogen species since the industrial production of the drugs, starting in the middle of the twentieth century [1,2]. There has been a great deal of research interest in understanding the barriers to plasmid-mediated HGT. The question of how plasmids themselves evolve is equally important and now can be addressed through analysis of huge public genome sequence databases. Here, we performed a genome data-driven exploration of historical evolution and horizontal spread of the small tetracycline-resistance plasmid pT181 in the important human pathogen *Staphylococcus aureus*.

S. aureus is a Gram positive bacterium that asymptomatically colonizes the nares and other body sites of a third of the human population, but is also able to cause a diverse set of infections ranging from mild to life-threatening. *S. aureus* can be found in both community and hospital settings, and is also a significant pathogen of livestock. In a recent paper [3], Raghuram et al. (2024) found that an average nucleotide identity (ANI) of 99.5% could be used as a robust threshold to cluster *S. aureus* into 145 distinct strains (numbered consecutively with a “S99.5_” prefix, e.g. S99.5_1, S99.5_2, etc.). This strain threshold is consistent with the gaps in genetic diversity between strains found in other species [4]. ANI defined strains, defined based on core genome identity, offer more natural boundaries or subspecies groups than traditional allele-based sequence type of clonal complexes such as MLST [5,6]. In this study, we tracked the transfer of pT181 between the strains defined by Raghuram et al. (2024).

S. aureus has undergone waves of evolution of resistance to antibiotics since the introduction of the drugs in the 1940s [7], with the most well-studied event being the acquisition of the *mecA* gene on the mobile SCC*mec* elements conferring resistance to second-generation beta-lactams and creating methicillin-resistant *S. aureus* (MRSA) strains. While much attention is given to beta-lactam resistance, tetracyclines have also been a clinically important treatment option [2,8]. Tetracycline antibiotics work by binding the ribosome to prevent protein synthesis, arresting growth [8]. While there is evidence for ancient human use of tetracyclines in Sudan [9], the first modern antibiotic in the tetracycline class, aureomycin, was isolated from *Streptomyces aureofaciens* in 1948 [1] and entered clinical use soon after. Second-generation tetracyclines were developed in subsequent decades, with doxycycline being introduced in 1967 [2]. Doxycycline is still widely used in humans for a variety of conditions such as Lyme disease, acne, and STIs, and tetracycline is commonly used in animal agriculture. In recent years, tetracycline-class antibiotics have been used frequently in the doxycycline post-exposure prophylaxis (doxy-PEP) regimen to reduce STIs [10], and this use has been linked to increased resistance in off-target bacteria [11]. Resistance to tetracyclines has emerged through two primary mechanisms: ribosomal protection and efflux pumps[8]. In a similar timeline to MRSA, tetracycline resistance in *S. aureus* emerged in clinical practice soon after the first clinical use of the drug in the 1940s [12–14].

pT181 has a circular, 4.4 kbp genome encoding three genes: *repC*, *tetK* and *pre*, and an *oriC* region [15]. It does not encode the molecular machinery to transfer itself into a new recipient strain and instead must rely on mobilization by a co-resident conjugative plasmid or transduction by phages [16–18]. pT181 confers tetracycline resistance to its host bacterium through the product of the *tetK* gene, which encodes a tetracycline efflux pump. Rolling-circle replication is controlled by the product of *repC* interacting with *oriC*, while *pre* facilitates mobilization [18,19]. pT181 was the first *S. aureus* tetracycline-resistance plasmid to be sequenced [15] and the plasmid was also significant for its use for multiple seminal studies on DNA replication control [19–21]. pT181 has been called by several names in the literature; pT181 is the most common [22], but it has also been called pSK52 [14], pK34-7-1 [23], and rms7 [24].

pT181 is an ideal subject for genome-driven evolutionary studies because of its small size and ubiquity. In this study, we used a public dataset of over 80,000 *S. aureus* sequences to explore the evolution and spread of this plasmid. We were particularly interested in trends in plasmid copy

number (average ratio of plasmid replicons to chromosomes in the cell; PCN). PCN is an important but understudied aspect of plasmid evolution, which may have tradeoffs with horizontal transfer potential and levels of antibiotic resistance. PCN is regulated by various genetic mechanisms, depending on the plasmid, and can range from a single copy to hundreds per cell [18,21]. High copy number plasmids can impose a fitness burden on the host due to the metabolic cost of replication and transcription [25–28], while low copy number plasmids risk not being inherited by daughter cells without segregation mechanisms encoded on the plasmid [29]. Larger plasmids often have genes ensuring proper segregation, whereas smaller, high copy number plasmids such as pT181 rely on random diffusion for distribution during cell division [18,29]. pT181 copy number has been previously found to be controlled by ncRNAs binding to repC to suppress plasmid replication [30,31]. It has been known to integrate into SCC*mec* [32–34] and is part of some larger staphylococcal plasmids [35–37].

In this study, we screened a large collection of raw *S. aureus* Illumina whole genome sequencing read files for pT181. We developed a novel workflow to use read depth from FASTQ to estimate PCN and found evidence that pT181 copy number has declined within *S. aureus* since the first resistant strains in the mid-twentieth century. This decline was driven in part by integration of the pT181 isolate into other genetic structures. We identified repeated pT181 integration into the chromosome, including events outside of the previously-characterized region of SCC*mec*.

Results

Early pT181 was found exclusively in a single strain, detected later in diverse *S. aureus* genetic backgrounds

Of 83,366 publicly available *S. aureus* Illumina raw data genome projects used in this study (Raghuram et al. 2024), 7,927 (9.5%) contained pT181, with all isolates having BLASTN [38] hit coverage >95% of the plasmid (see Methods) and <5 mutations in coding sequences. There was a reliable date of isolation attached to metadata from 42,522 (51%) of the 83,366 genomes, with the earliest sample from 1884 (SRX2619650) and the latest in our set in 2022. 2,460 isolates contained pT181 and had a valid date, strain assignment, CC assignment, and SCC*mec* typing data (Fig. S2).

The early pT181 genomes were almost all S99.5_2 [3], which are in MLST CC8 [6]. The earliest dated genome with pT181 was a methicillin-sensitive *S. aureus* (MSSA) isolate collected in 1954. pT181 was exclusively found in S99.5_2 genomes until 1980, when it was detected in S99.5_27 (CC8) integrated into the chromosome as part of type III SCC*mec*. By 2022 (the latest date in our collection), pT181 had been detected in 44 strains. While the number of pT181 genomes collected per year after 1954 increased, overall pT181 prevalence has decreased over time (Fig. 1A) [logistic regression, $p < 2e\text{-}16$]. The vast majority of high-quality *S. aureus* samples collected before 1995 were from S99.5_2 (225 out of 242 samples) (Fig. 1B). The subset of high-quality pT181-positive samples collected before 1995 were similarly almost all from S99.5_2 (182 out of 184 samples). Therefore, with this set of genomes, we cannot know if the pT181 *S. aureus* origin was in S99.5_2, or if it was simply detected there first.

We also identified pT181 plasmids in 61 genome assemblies of 14 other species (File S1). The majority were from *Staphylococcaceae*, but there was also a plasmid found in *Pseudomonas aeruginosa* with plasmid coverage that exceeded the length thresholds for inclusion (see Methods). The *P. aeruginosa* strain was collected from tracheal swabs [23], which are frequently taken from cystic fibrosis patients, in whom co-infections with *P. aeruginosa* and *S. aureus* are common. However, HGT between these gram-negative and gram-positive species are rarely documented.

pT181 sequence variation is concentrated in hotspots outside the coding sequence

Overall, there was a low level of sequence variation in pT181 *S. aureus* and non-*S. aureus* plasmids in this study. 519 sites out of 4,439 sites (12%) on the plasmid had more than one allele in our set of 7,988 sequences (Fig. 2). 664 unique mutations were identified with Snippy [39], and 656 (99%) of

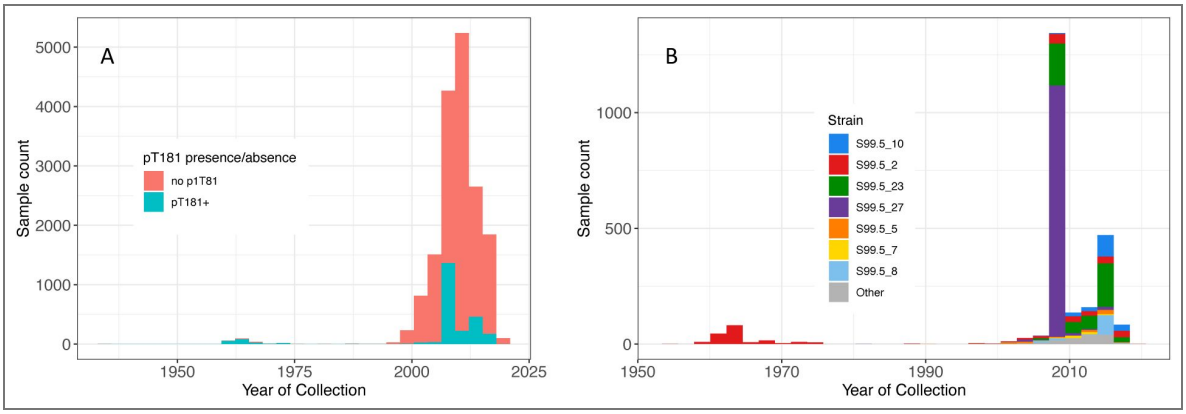


Fig. 1.

A) Presence/absence of pT181 in high quality *S. aureus* shotgun genome data over time (n = 16,928). pT181+ isolates comprise less of the sample set over time, though counts of collected samples increase. B) Strain composition of pT181-containing samples has diversified since pT181's initial detection in S99.5_2.

these are rare, occurring in <5% (<400/7,988) of samples. This variation was concentrated in three regions. The phylogenies of both the whole plasmid (Fig. S1A) and CDS (Fig. S1B) had a starlike structure with poorly supported branches. The mutation rate in pT181 CDS was slightly lower than former estimates of mutation rates in core chromosomal genes (pT181 = 1.15 mutations/Mb/year, chromosome = 1.2 - 4.4 mutations/Mb/year [40–44] when calculated via linear regression of SNP accumulation over time. While the relationship between time and mutation accumulation was significant (linear regression, p = 0.0001324), time since sample collection was a poor predictor of CDS SNP count (linear regression, adjusted R^2 = 0.006) (Fig. 3). As a result, we are unable to approximate dates for undated samples based on plasmid sequences. This regression excluded the intergenic regions and insertion mutations in *rep*, which had higher levels of variation. There was no evidence of recombination with a pairwise homoplasy index (PHI) test [45] of all (n = 7,988) aligned plasmid sequences (p = 0.39), as expected given that the low level of sequence diversity greatly reduced the power of the test. We concluded that pT181 entered the *S. aureus* population since the divergence of the strains [44], given that most pT181 plasmids have identical coding sequences (CDS) (Fig. 3).

While the majority of the pT181 sequence was highly conserved, three regions of increased variability were present on the plasmid: 1) the intergenic region between *repC* and *tetK*, (coordinates 814 to 817 on accession DRX100326), 2) the intergenic region between *tetK* and *pre*, and (coordinates 2460 to 2466 on accession DRX100326) 3) in *oriC* and early in *repC* (coordinates 4133 to 4313 on accession DRX100326). Mutations immediately before, and at the 5' end of *tetK* are found exclusively in single-copy isolates (see later section on PCN). The spikes coincide with known plasmid cointegration sites [36,37,46].

pT181 in *S. aureus* was highly similar to pT181 found in other species. The *S. aureus* and non-*S. aureus* plasmids were all in a highly similar cluster (5 or fewer CDS mutations relative to the 1954 isolate), with the exception of *Staphylococcus pseudintermedius* (CP128253.1), which had 12 CDS mutations (8 SNPs, 4 complex mutations). Only 5 non-*S. aureus* isolates contained a mutation exclusive to non-*S. aureus* isolates, but these were at low prevalence, with a SNP present in 2 isolates and a complex mutation present in 3 isolates. These results indicate that the non-*S. aureus* pT181 sequences did not constitute an outgroup to the *S. aureus* pT181 sequences.

pT181 is most frequently found at a single copy per cell

We noted a skewed distribution of PCN in gold-quality *S. aureus* pT181 isolates, with a wide range of PCN estimates (Fig. 4A) and a strong peak of pT181 PCN = 1 (Fig. 4B), consistent with chromosomal copy number.

pT181 can be present on the chromosome as part of type III staphylococcal Cassette Chromosome mec (SCC*mec*), which confers both methicillin and tetracycline resistance [32,33]. We found that the average PCN of type III SCCmec strains was 1.16 with a range of 0.71 to 1.43 (see Methods), i.e. matching the expected distribution given SCC*mec*'s position near the origin. Based on this we designated any sample with pT181 PCN < 2 as single-copy. Of 2,460 genomes with valid date, strain assignment, CC assignment, and SCC*mec* typing data, 1,863 (76%) were single-copy (PCN 0.147 - 1.97; mean = 1.06, sd = 0.186). We suspect that these single-copy pT181 were integrated into the chromosome or a low copy number plasmid. Instances of integration into the chromosome in other types of SCC*mec* have been previously documented [34,46,47]. We found that 1,131/1,863 single-copy pT181 (61%) were present in SCC*mec* type III isolates, 695/1,863 (37%) were in other MRSA strains, and 37/1,863 (2%) were in MSSA, i.e. lacking a SCC*mec* cassette.

pT181 has inserted into multiple chromosome sites

There were 19 isolates with pT181 present on the same contig as a known chromosomal gene in the SRA-assembled genomes, comprising 6 unique insertion sites as defined by directly flanking genes. All had pT181 present in a single copy number. The isolates were distributed across multiple strains: 5 S99.5_5 (CC5), 2 S99.5_2 (CC8), 8 S99.5_23 (CC398), 1 S99.5_10 (CC1), and 3 with unassigned strain (CC1 & unassigned). One isolate was MSSA, 3 were SSC*mec* type IV, and 15 were SCC*mec* type V. To our knowledge, this is the first documented example of chromosomal pT181 in

Fig. 2. pT181 is characterized by hotspots of sequence variation.

Hypervariable regions have a large number of alleles present. Each hypervariable region is characterized by a set of mutation types (binwidth = 10 bp).

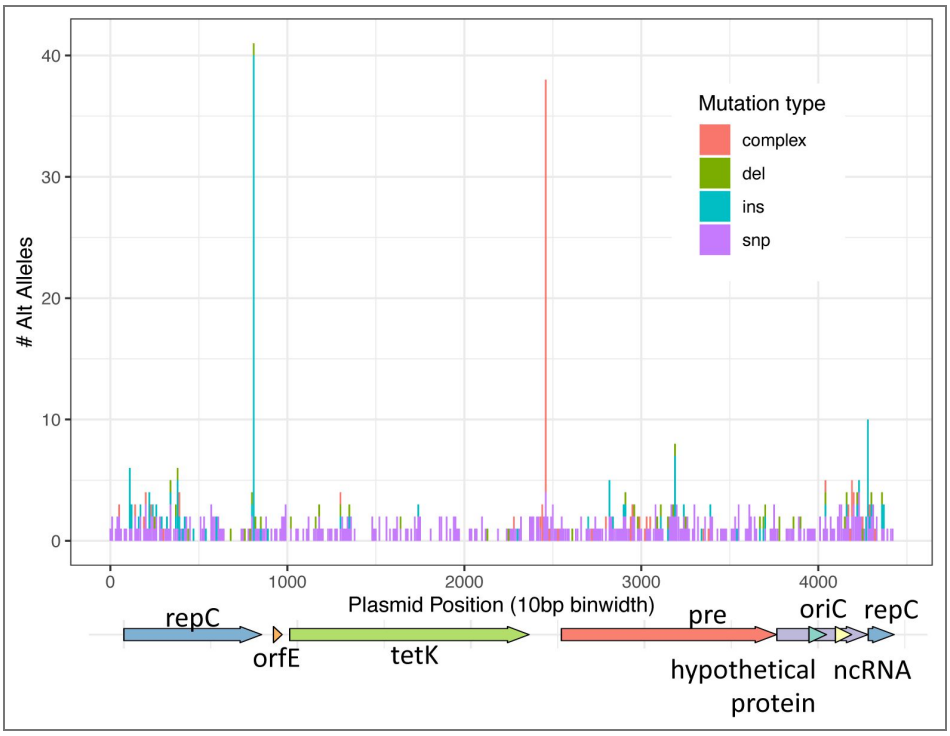
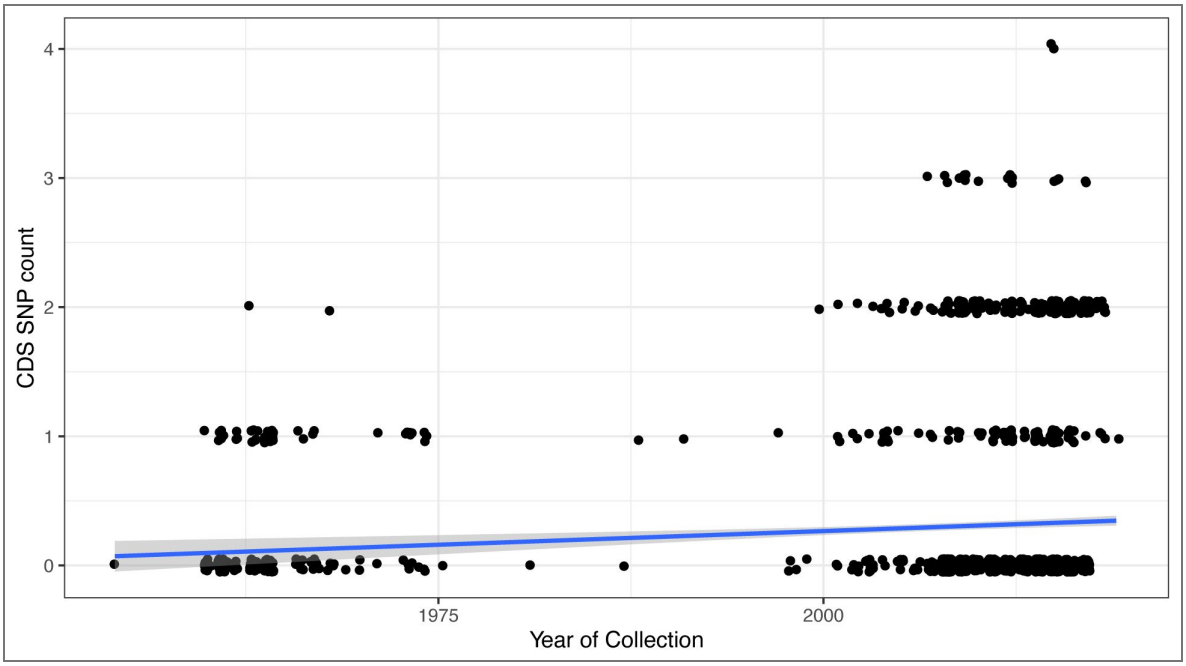


Figure 3. pT181 CDS SNP accumulation over time relative to the 1954 isolate, calculated from Snippy output.



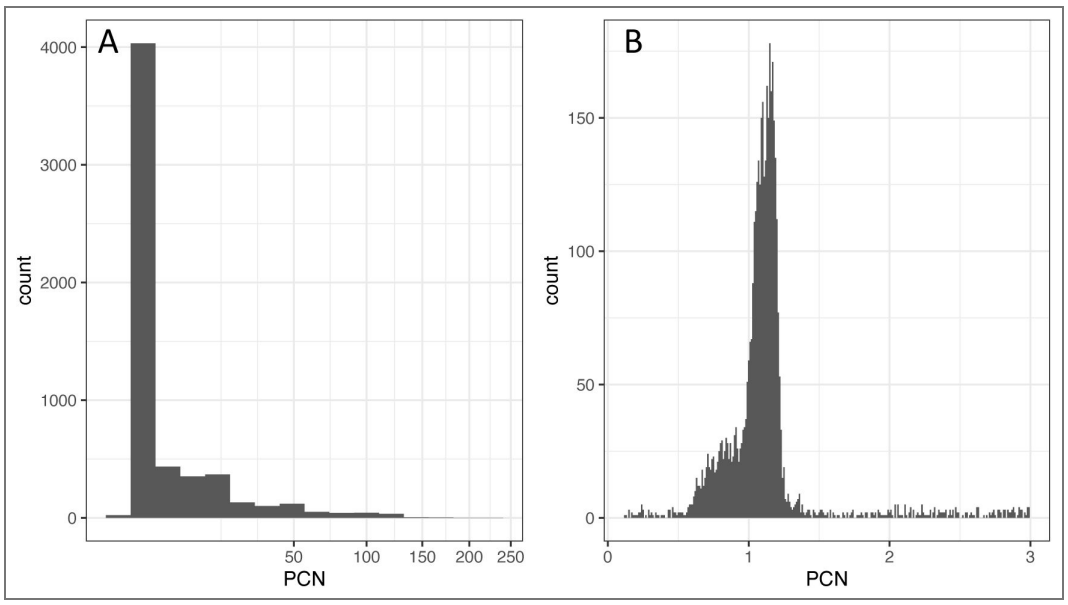


Fig. 4.

A) Overall PCN distribution for gold-ranked, pT181+ isolates. The majority of pT181 isolates in our dataset are single-copy. Given that pT181 lacks segregation mechanisms, this suggests that the plasmid has become integrated into the chromosome. B) Subset of PCN distribution, enlarged for 0-3 copies/cell.

MSSA. The integrated pT181 were flanked by insertion sequences (IS): IS6, ISSau6, or IS257 transposases, or by one transposase and a contig break. Integrated pT181 was present in multiple configurations and positions in a genome, and there were no nearby chromosomal genes common to all integration sites, other than some form of flanking transposase (Fig. 5). Some pT181 integrated with *repC* positioned nearest the chromosomal flanking genes, and other isolates *tetK* nearest, due to variation in the site of plasmid cointegration. Some *S. aureus* strains had multiple different insertion sites, indicating repeated integration following introduction into a strain. There were plasmids with identical sequences that were both single and multicopy, lending further support to our finding that the integrations have happened repeatedly over the course of pT181 history.

In 337 *S. aureus* complete genomes [3], we found that 14 (4.2%) contained pT181. 7 of these were integrated into the chromosome, 4 were integrated into larger plasmids (total plasmid length: 21-44kbp), and 3 were free in the cytoplasm. The integration of pT181 into larger plasmids may result in low PCN, as pT181 could assume the copy number of the larger element. The chromosomally integrated copies of pT181 occurred in positions 55-88kbp in these genomes, near SCC*mec*, with both positive and negative sense orientations. The region in which pT181 integrated into the chromosome in these genomes is populated primarily by rare and strain-diffuse genes [3] (Fig. S3). The presence of pT181 in larger plasmids, which tend to be lower copy number [50], as well as integration into the chromosome, indicate additional routes to evolve low copy number beyond regulation of free plasmid.

pT181 PCN has declined over time

There was an overall tendency for declining pT181 PCN from the 1950s to the end of sampling. This is the result of two trends: (1) increase in single-copy isolates over time and (2) lower PCN over time in multicopy isolates.

Over the period for which we have dated samples, pT181 was increasingly found in single-copy form (Fig. 6A) (linear regression, adj-R^2 = 0.31, p = 8.719e-05). Pre-1995 samples were mostly multicopy (8.2% single-copy), while 81% of samples from 1995 or later were single-copy. Single-copy pT181 was not evenly distributed across the strains ($p < 2 \times 10^{-16}$, Chi-squared test statistic = 2052), with some strains being mostly or entirely single-copy (Fig 6B). S99.5_27, all isolates of which are type III SCC*mec*, constitutes a substantial portion of single copy isolates (1,131/1,863). However, the increase in single-copy isolates over time is still significant even with this strain excluded (linear regression, $p < 2 \times 10^{-16}$, adj-R^2 = 0.1472).

We used a Gamma log-link generalized linear mixed effects model (GLMM) to assess the impact of date of collection and strain background on PCN in multicopy isolates. Strains and BioProjects containing at least 5 multicopy isolates were selected; to be used, a sample must belong to a strain and bioproject combination containing at least 4 other samples (n = 455). BioProject was assigned as a random effect within the fixed effects of strain and year, to account for possible technical variation between projects, while recognizing that Bioprojects frequently contain one or few strains and are typically collected in a narrow time frame. There has been a decline of PCN over time in these multicopy isolates ($p = 3.30 \times 10^{-12}$). No strains were significantly different from our reference, S99.5_8 (mean PCN = 17.4, SD = 4.35) (treatment vs. control test, emmeans package, p > 0.05, BH multitest correction). The decline in multicopy PCN over time indicates that possible evolution of the plasmid, chromosome, or both, toward lower PCN.

Identical plasmids can have variable plasmid copy numbers

pT181 copy number control mechanisms are located on the plasmid, which regulates its own copy number [19]. However, in our set of isolates, we find that identical pT181 isolates can have a wide range of copy numbers (Fig. 8). This suggests that there are factors other than the previously-identified copy number control mechanisms on the plasmid which influence pT181 copy number, such as host genetic background or environment.

Fig. 5. There were 6 non-type III SCC*mec* configurations of accessory genes directly surrounding the integration site of pT181 contigs across multiple strains in the SRA isolates.

This shows that the plasmid has integrated repeatedly into the chromosome, given that *S. aureus* is highly syntenic and accessory genes tend to cluster in specific regions of the genome [3,48,49].

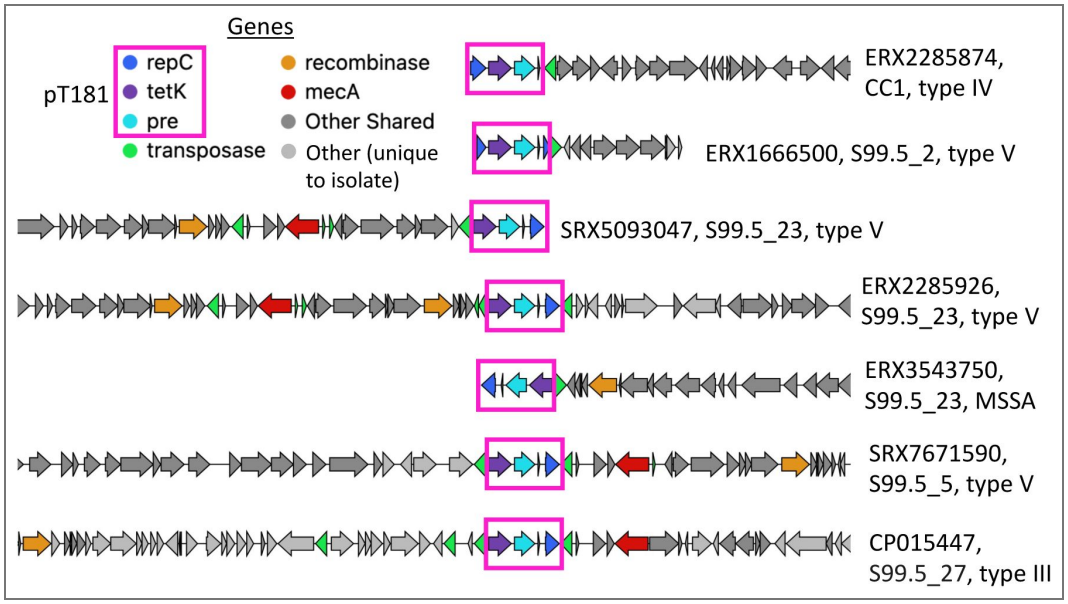


Fig. 6. Proportion of pT181 isolates present at single-copy number has increased since initial detection and varies across strains.

A) Proportion of single-copy pT181 increased over time. B) Single vs. multicopy pT181 across strains, for strains containing >= 2 pT181+ isolates in (n = 2,460) set. Strains are generally dominated by either single or multicopy pT181.

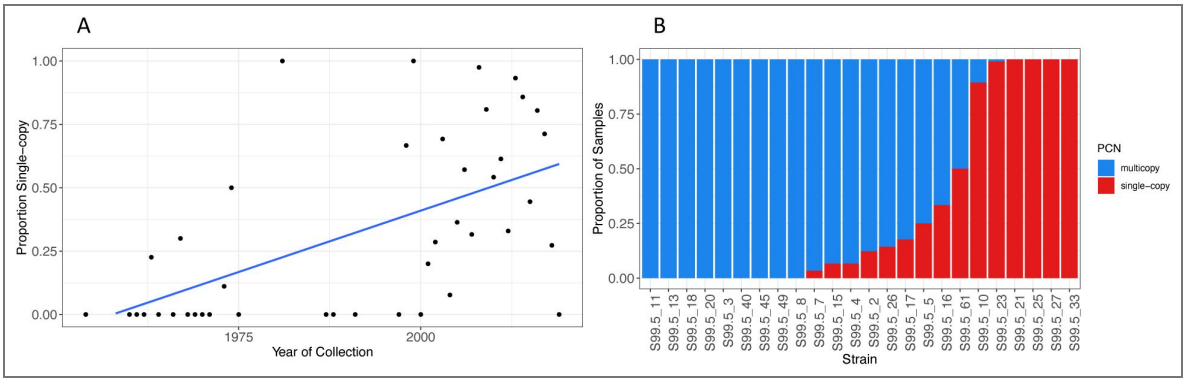


Fig. 7. PCN of multicopy isolates has declined over time.

Strains' PCN are not significantly different from reference strain, S99.5_8 (GLMM, treatment vs. control posthoc, BH multitest correction).

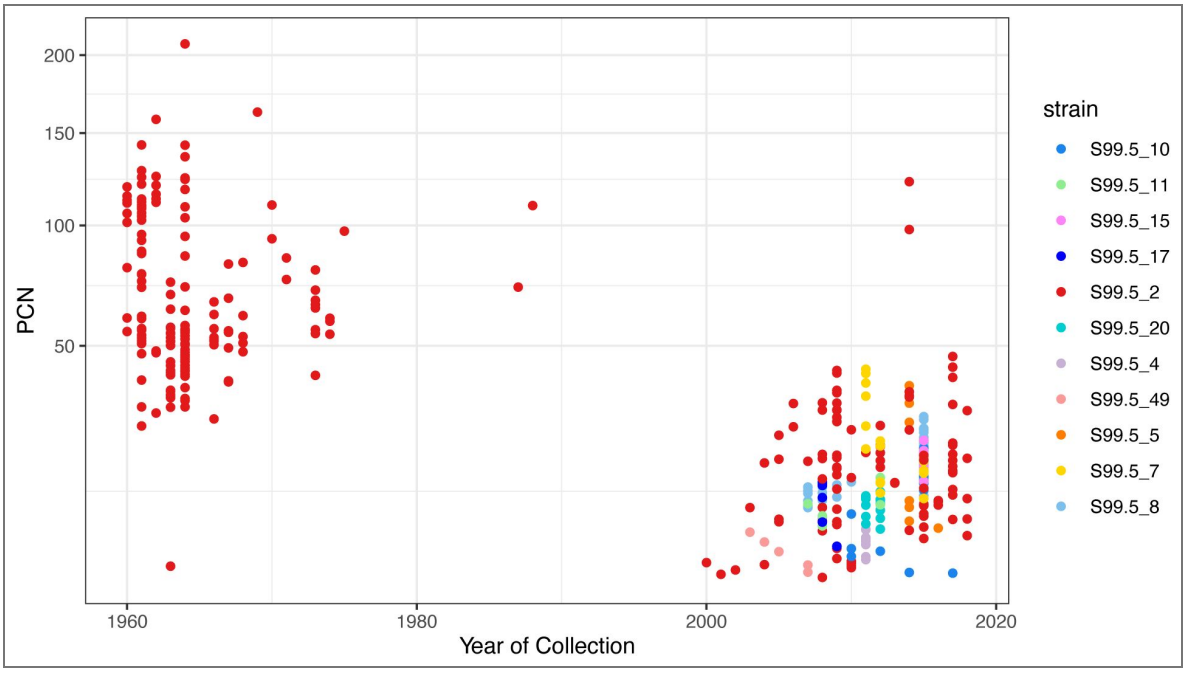
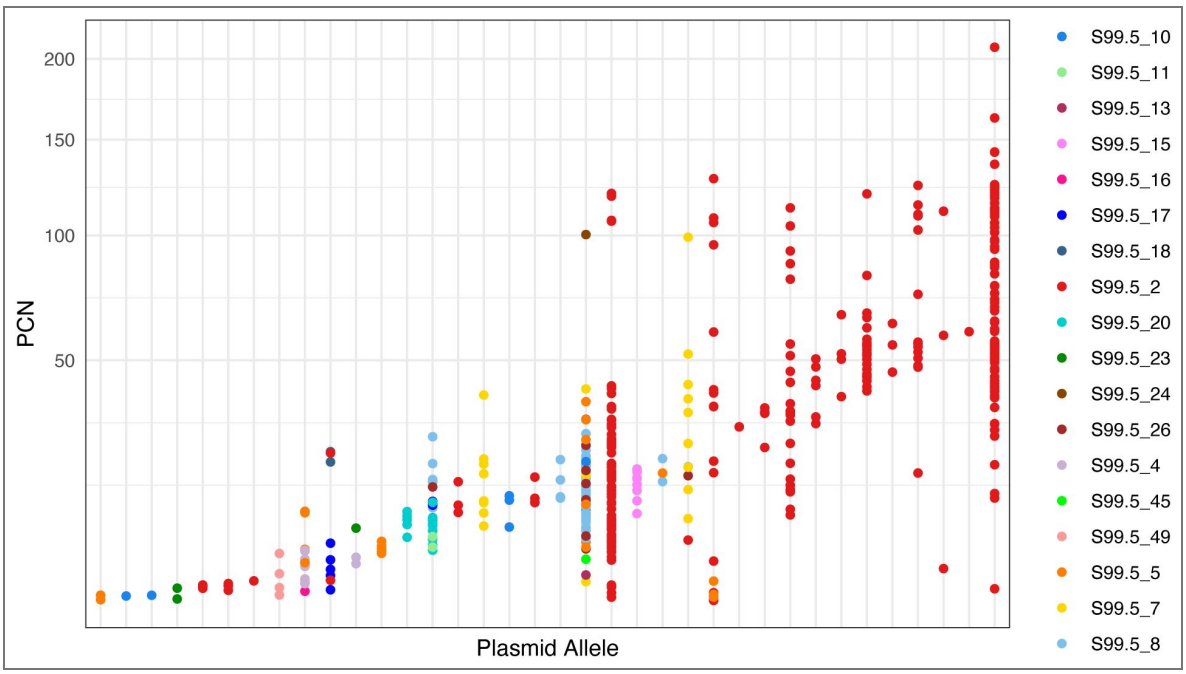


Fig. 8. Multicopy PCN varies within identical plasmid sequences and strains, indicating that there are environmental or genetic effects occurring at the substrain level which impact pT181 PCN.

Plot shows all unique plasmid sequences (across all sites) occurring at least twice as multicopy isolates in the (n = 2,460) dataset.



Discussion

This is the first study to use the thousands of available public genome sequences to look at variation of PCN of a single plasmid. We estimated the PCN of multicopy pT181 in high-quality isolates to be 2 to 208 (n = 2,460) with a median of 19 copies/cell (Q1 = 11 copies/cell, Q3 = 47 copies/cell) (Fig. 4A). Our median PCN agrees with experimental studies using qPCR, relative resistance to antibiotics [21], and fluorescence densitometry [51]. Initial PCN estimates for wild-type pT181 were 22 +/- 3 copies/cell for isolates collected pre-1977 [19,52,53]. Read depth has been previously used for ploidy estimations in cancer and yeast [54,55], but we have shown here that it can be accurately employed to estimate copy number of MGEs or other contigs in bacterial genomes, given sufficient knowledge of the genome such that a set of gene or genes can be used as the known single copy regions to compare the unknown region(s). The genomic-based PCN method can be easily applied to thousands of already-collected samples, without the need for specialized qPCR assay development. In addition, these methods are not culture dependent and could be deployed in metagenomic samples given sufficient quantity of the host in a metagenome sample. The disadvantage is that we do not have control of growth conditions that might influence plasmid copy, so we cannot account for all factors that may contribute to the measured copy number. However, *S. aureus* is usually grown overnight to stationary phase in rich media before genomic DNA extraction, and we therefore expect reasonable consistency in the genomes used for this study.

Previous research into the control of pT181 copy number has discovered the mechanisms present on the plasmid itself [19]. However, it is necessary to investigate contributors to copy number variation beyond the plasmid sequence, given our findings that identical plasmids can result in different multicopy PCN. At a population level, we find that the strains vary in their average copy number, but this difference was not statistically significant following post-hoc testing. Ideally, to compare the effect of strain background, a single large Bioproject with contemporaneously collected samples across strains would be used. The variation in multicopy PCN between strains indicates that there may be additional genetic contributors on the chromosome, or interaction between the chromosome and plasmid could contribute to the variable PCN of multicopy pT181. Integration into a wide array of larger multicopy plasmids may also contribute to the large variation in multicopy PCN we see in this study. This warrants further investigation into what impact genetic background or environment have on pT181 copy number, as genetic background likely plays a role in determining PCN for multicopy isolates, and variation could occur below the strain level.

While it was previously known that pT181 could integrate into the chromosome as part of SCC*mec* elements, it is a new finding from this study that the plasmid is at approximately chromosomal copy number (i.e. single-copy) in the majority (76%) of *S. aureus* isolates. The majority of single-copy pT181 is found in SCC*mec* type III isolates, with pT181 integrated into SCC*mec*. We also identify integration into other SCC*mec* types, as previously described in the literature [34]. Integration into the chromosome in MSSA isolates is previously uncharacterized in the literature to our knowledge. Integrated pT181 in our dataset are flanked by insertion sequences (IS), or by an IS and a contig break. IS elements have been previously documented to insert in these regions of pT181 [36,37], and these may account in part for the spikes of mutations occurring in these regions in our analysis. IS6 and IS6-like transposases, which flank integrated pT181 in our data set, have been known to integrate multicopy plasmids into the chromosome; IS257 has been previously found to integrate pT181 into larger plasmids and into the chromosome as part of SCC*mec* [32,33,36,56–60].

One major finding from this study is that PCN overall has declined over the second half of the twentieth century, driven by both pT181 integration and lower PCN of multi-copy pT181. PCN is likely under selection to reduce the fitness burden on *S. aureus* cells that comes from replicating larger amounts of plasmid DNA. An analogous argument for lower fitness penalty driving the transition of *S. aureus* from larger SCC*mec* type I to smaller type IV has been put forward [7]. Adaptation could arise through IS-mediated integration into the chromosome, or large plasmids, and/or mutations on the plasmid or chromosome replication control circuits to lower PCN. In some scenarios, high PCN may be advantageous as cells with higher PCN of pT181 have higher

tetracycline MIC [57] and previous studies have identified gene dosing as a mechanism to increase MIC [61–64]. The high PCN in early S99.5_2 isolates may have been a fitness defect that was permitted because of the benefit of conferred tetracycline resistance, and the decline over time was adaptive for long-term maintenance. pT181's decline in PCN brings it from median multicopy PCN of 55.7 copies/cell pre-1995 to 14.5 copies/cell post-1995 (n = 2,460), a decline from roughly 9% to approximately 2% of the chromosome's size. The post-1995 median is in line with recent research on plasmid copy number across taxa, which found that plasmids occupy roughly 2.5% of the chromosome's size [65]. The higher PCN state was likely not advantageous long-term and declined over time. There may also be genetic adaptations maintaining high levels of *tetK* transcription despite lower PCN. Simpson et al. (2000) [57] identified the appropriation of an IS257-derived hybrid promoter, which had higher activity than the native pT181 promoter, rendering the integrated copies more tetracycline resistant than the multicopy isolates. In that way, bacteria with integrated pT181 could have been selected for both because of their increased tetracycline resistance due to the hybrid promoter and reduced the cost of maintaining a multicopy plasmid. High levels of *tetK* transcript have been shown to be toxic in *E. coli*[66], so there may be a limit to benefits from high levels of transcription with respect to both resources used and impacts on the cell. An alternate but not mutually-exclusive hypothesis is that it was advantageous for *S. aureus* to integrate the plasmid to allow co-residence of another advantageous, but incompatible cytoplasmic plasmid.

Our study identified a wide range of PCN of pT181, and there are many factors that could have contributed to this. Integration into the chromosome has driven the transition to a predominantly single-copy lifestyle; integration into larger plasmids has likely contributed to the diversity of multicopy pT181 PCNs. While strains in our study were found to vary in PCN of multicopy pT181, these differences were not found to be significant (GLMM, BH posthoc, all p-values > 0.05). PCN is known to be controlled by pT181 antisense ncRNA binding to *repC* which modulates plasmid replication [30,53,67] when free in the cytoplasm. Variation that affects plasmid *repC* expression and/or *oriC* binding can affect the average number of copies per cell [20,21,35,67]. Chromosomal loci may influence pT181 PCN through mechanisms such as RNA degradation, but the strain level variation of this phenotype has not been an area of intense research, and it is likely that there are more genes involved in plasmid regulation. Further research on the genetic basis of variation of pT181 PCN and variation in relative levels of *tetK* production will help explain how this plasmid-bacterial symbiosis has adapted to the challenge of large scale antibiotic usage.

In our study, pT181 was initially detected as a high PCN plasmid in S99.5_2, with appearance in other genetic backgrounds occurring decades later. Over time, pT181 PCN declined, driven in part because of increased frequency of chromosomal integration of the plasmid. Extensive spread across *S. aureus* has occurred, as well as spillover into other species. Chromosomal restriction-modification and CRISPR systems appeared to have not prevented the horizontal transfer of pT181 across *S. aureus*. This is consistent with historical reports of early detection of multicopy pT181, followed by the tetracycline resistance phenotype becoming chromosomally encoded in subsequent decades [68]. However, we are limited in our ability to estimate the initial origins of pT181 in *S. aureus* and the timeline of transmission events due to uneven sampling and low sequence diversification. Other lineages may have contained pT181 at the same time or earlier than S99.5_2, but were undetected as these strains were not frozen and later sequenced when the era of large scale "next-generation" Illumina sequencing began in the 2010s [69].

The first recorded detection of tetracycline resistance via pT181 dates to 1953 [14], very soon after the modern discovery of tetracyclines [1]. It is likely that antibiotic use was the major selection pressure increasing the incidence of pT181+ isolates since the 1950s. Our finding that pT181 CDS were almost identical suggests recent expansion from a single introduction. S99.5_2, a subset of CC8, constituted the first wave of MRSA in the 1960s [13,70], and was therefore likely sampled at high frequency relative to its true abundance within the species. While samples were collected in the pre-antibiotic era, sampling began in earnest after antibiotics were deployed clinically, with *S. aureus* samples increasing in number in the 1960s [71]. Further conclusions about the patterns and timeline of spread taken from analysis of public genomes will need to be performed with more comprehensive sample sets, controlled for sampling bias. There may be collections of strains still

available to help refine our knowledge of the genomics of this era. Even the more recent genomes are disproportionately taken from human samples in hospital settings in the global north, failing to represent the full diversity of the species. The likely presence of pT181 in *S. aureus* before the 1950s may have reflected selection to defend against tetracyclines produced as natural products by other bacteria, such as Streptomyces [1]. Environmental conditions besides tetracycline resistance could also have lead to selection for pT181-containing isolates, such as heavy-metal and pH stressors, as *tetK* is able to efflux heavy metals and acts as a proton antiporter [72,73].

By using pT181 as a model, this project uses new approaches for leveraging historical data and determining plasmid copy number. Our work raises questions around MGE evolution and lifestyles, as this single plasmid's lifestyle has dramatically diversified since it was first detected in the 1950s and 1960s. This work adds to the knowledge base that suggests that MGEs are not as constrained to a single lifestyle, structure, and location. Going forward, our work has possible applications for others studying MGEs and AMR. We also show the importance of depositing sequence data from before the current era into public databases. The immediate post-WWII era was marked by a sudden, extreme chemical warfare against the most common bacterial pathogens. It is important to understand how they evolved to survive this challenge because it likely influences trade-offs in fitness in current clinical strains. Any existing unsequenced cultures of *S. aureus* collected before 1995 need to be preserved and sequenced, as they are vital historical records that can help us understand this important period in the evolution of this pathogen.

Methods

pT181 genomics

S. aureus isolates containing pT181 were identified by BLASTN (v2.12.0+) [38] in the Staphopia v2 (n = 83,366) [3] dataset of publicly available data assembled from SRA with SKESA [74,75] in Bactopia (v1.7.0) [76]. Samples were considered to contain pT181 if the sum of hit coverage exceeded 4.4 kbp (Fig. 9). Identical scripts and thresholds were used to identify pT181 in the complete genomes. Non-*S. aureus* isolates containing pT181 were identified using BLASTN [38,77] in the nucleotide database using the NCBI web interface (*https://blast.ncbi.nlm.nih.gov/*). Searches were performed on 2023-11-05 with the default parameters. We identified non-*S. aureus* sequences with >20% identity and >50% coverage with pT181. We then applied the same thresholds, as for the *S. aureus* genomes, ensuring that the sum of all pT181 coverage was >4.4 kbp. The CDS phylogeny of pT181 was created with IQ-TREE2 (v2.1.4-beta) [78], using a multisequence alignment generated by mapping to isolate DRX100326 with ska [79] and partitioning the CDS regions. Metadata of samples used in this study was downloaded from SRA for *S. aureus* isolates and from GenBank for the non-*S. aureus* isolates. Manual processing of *S. aureus* collection dates for downstream analyses was performed using RStudio (v1.3.1056) [80] with R (v4.0.2) [81] and tidyverse (v1.3.1) [82].

We reduced the set of plasmid samples to 2,460 *S. aureus* pT181 sequences by selecting only those with a valid date, typed SCC*mec* (see "SCCmec typing"), gold rank assignment [76], and a genetic strain assignment [3].

SCC*mec* typing

SCC*mec* typing was performed using sccmec (v1.2.0) [83]. The sccmec tool compares assembled genomes to a database of known SCC*mec* elements using BLASTN. The database includes specific SCC*mec* targets (*mec*, *ccr* complex) as well as full cassettes for 15 SCC*mec* types as well as 20 subtypes. Based on the presence of SCC*mec* targets and cassette coverage, sccmec will assign a type and subtype.

pT181 sequence alignment and SNP calling

As pT181 is a circular replicon, it was necessary to rotate (i.e., resetting the circular plasmid to a consistent start point) the plasmid so that all of the sequences in the files would start at the same position. The start position of BLASTN (v2.12.0+) [38] hits to the reference plasmid were used to identify an identical start position. Samtools (v1.9 with htslib v1.9) [84]; faidx was used to extract

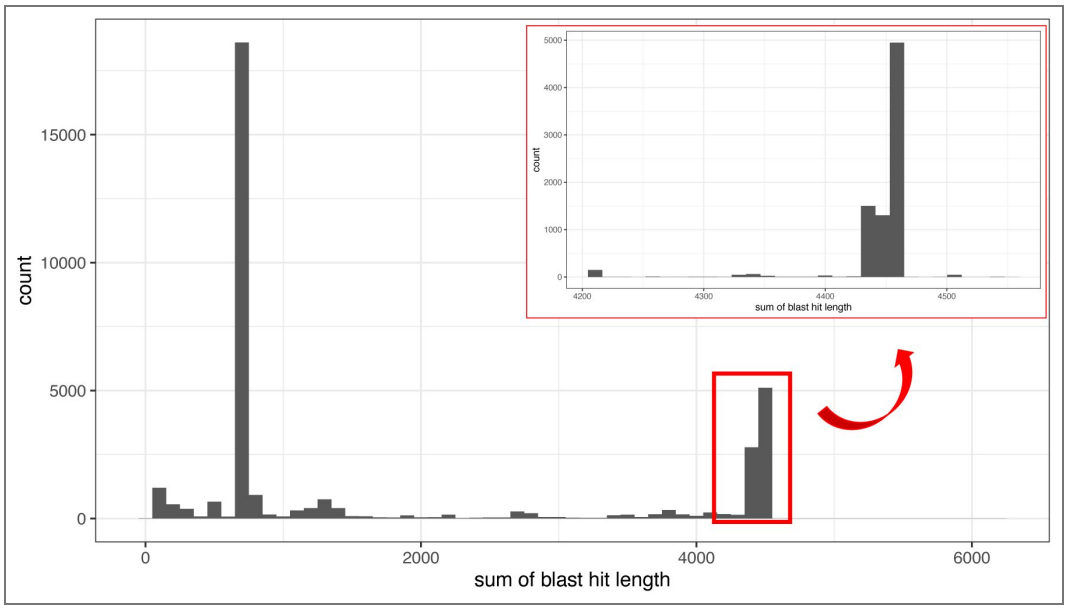


Fig. 9. Distribution of sum of blastn hits to screen for pT181 presence.

Primary peak was for sum of blastn hits distribution was 4425-4475 bp and the threshold for inclusion of isolates was set at 4400 bp.

these positions. Picard (v3.1.1) [85] normalize was used to generate fasta files with equal line lengths. SNPs, small indels, and complex mutations were identified using Snippy (v4.6.0) [39]. Mutation spikes were not due to our rotation of the plasmid for alignment. The mutations that occur in the spike positions occur regardless of where the start of the contig is located (Fig. S4). G. Gene positions were identified with bakta (v1.8.2) [86] annotation and gene plot (Fig. 2) was visualized with gggenes[87].

Linkage analysis

Linkage analysis was performed via a Phi test [45] run with Phipack (v1.1) [88], on the ska (v1.0) [79] alignment of all 7,988 pT181 sequences.

Plasmid copy number

Illumina reads from each isolate were mapped with BWA (v0.7.17) [89]; to each isolate's own assembled genome (Fig. 10). Read depth across the genome was calculated with samtools depth at all sites (v1.9, with using htslib 1.9) [84]. Single-copy chromosomal contigs were identified by the presence of known single copy chromosomal genes, identified by BLASTN (v2.12.0+)[38]. To ensure an accurate ratio, only assemblies ranked as "gold" by Bactopia [76] were used for the PCN analysis. PIRATE (v.1.0.5) [90] gene families were used to identify single copy chromosomal genes. We utilized gene families found in >95% of genomes at exactly 1 copy per genome, with fewer than five fission loci.

The cutoff for single-copy pT181 (PCN < 2) was validated using the distribution of coverage for contigs containing lawful genes for gold-ranked isolates (Fig 11). The distribution of chromosomal copy number (CCN) ranges was centered on 1, but values range from less than 0.5 to greater than 2. To determine the copy number of genes located near the origin, including SCC*mec*, we calculated average read depth of contigs containing one of the four genes (*rpmH, dnaA, dnaN,* and *recF*) located near the origin of replication [91]. The distribution of CCN for those genes was slightly higher (mean = 1.15, sd = 0.11). This distribution had a similar center to the distribution of pT181 for SCC*mec* type III, which has the pT181 plasmid integrated into the cassette (mean = 1.16, sd = 0.057).

We assessed the coefficient of variation, or the standard deviation of read depth divided by mean read depth for each contig for a test set of 10 isolates (Fig. 12). Core chromosomal and pT181 contigs are not significantly different in their coefficient of variation (Wilcoxon rank sum, p = 0.28). This analysis convinced us that read mapping variation across pT181 was not significantly different from chromosomal read mapping, allowing us to use this process for copy number analysis.

PCN as a function of time and strain

PCN over time was assessed using a Gamma (log link) generalized linear mixed effect model (GLMM) with R (v4.4.2) [81] using the package glmmTMB (v1.1.11) [92]. Posthoc testing was performed with the emmeans package (v1.11.1) [93]. Formula for glmmTMB call is as follows: PCN ~ numYear + strainFac + (1 | strainFac:BioProject) + (1 | numYear:BioProject). numYear refers to year of collection as a continuous variable and strainFac refers to strain as a factor.

Identifying pT181 novel chromosome insertion sites

Using the core gene blast hits used to identify known single copy genes, as well as searches for pT181, we looked for contigs that contained both pT181 and core genes. 19/7,927 *S. aureus* pT181 isolates were found to have core genes on the same contig as pT181. We then used Prodigal [94] annotations generated with Bactopia (v1.7.0) [76] to identify the genes directly surrounding the insertion site of the plasmid. Genes were visualized using clinker (v0.0.31) [95]. The majority of pT181-containing contigs were the same canonical length of the plasmid, approximately 4,440 kbp. We are likely unable to identify genes on other contigs of most putatively integrated pT181, due to repetitive sequences making it difficult to assemble in those regions.

Fig. 10. Cartoon of read mapping and relative read depth measurements.

The relative read depth of the plasmid provides an estimate of its copy number.

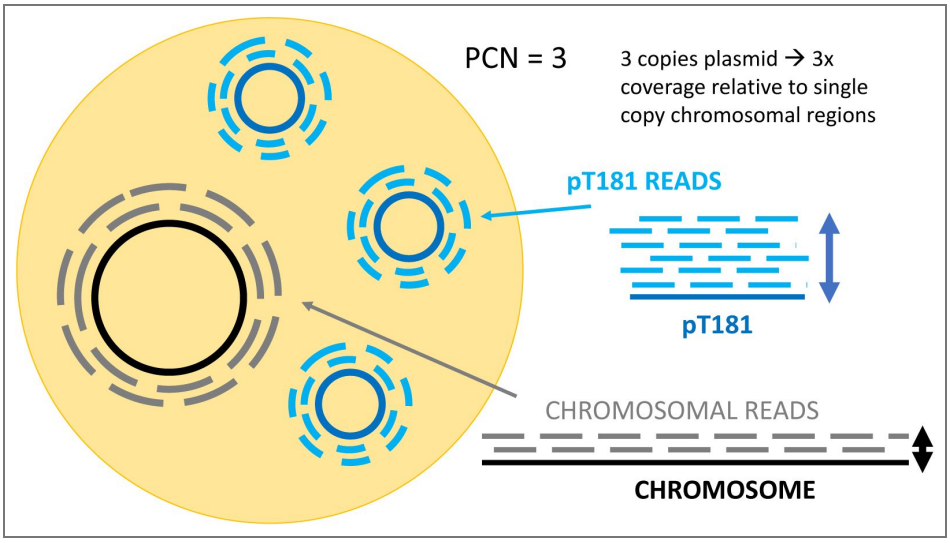


Fig. 11. Distribution of chromosomal copy number (CCN), colored by position relative to origin of replication.

Contigs are considered near the origin if they are on the same contig as at least one of the 4 genes (*rpmH, dnaA, dnaN,* and *recF*) known to be close to the origin of replication [91].

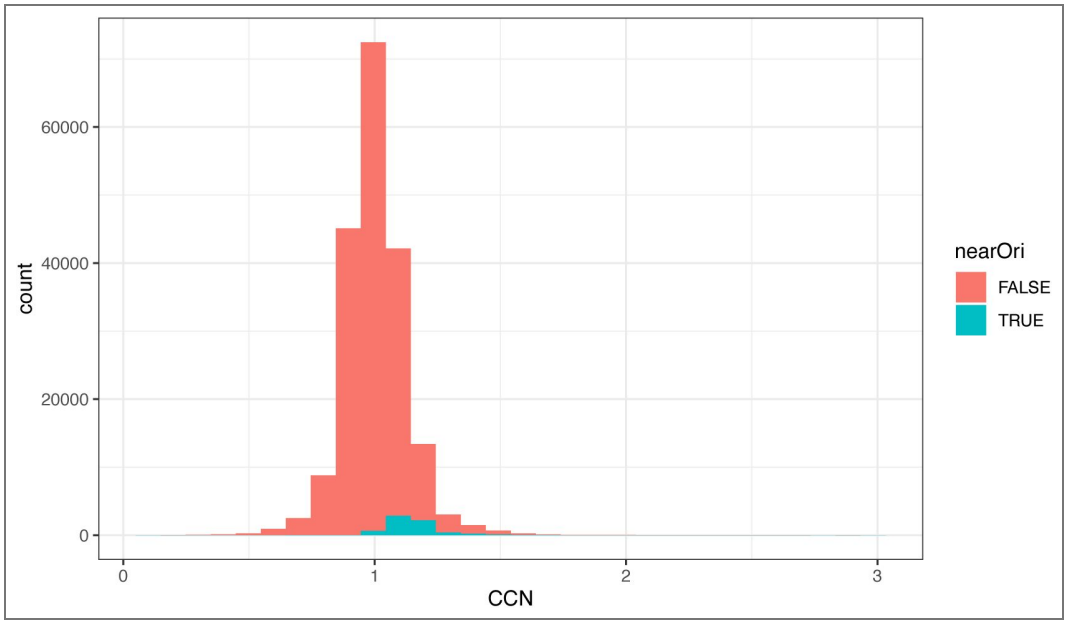


Fig. 12. **Coefficients of variation of read depth (standard deviation divided by mean) for chromosomal contigs, pT181, and other contigs follow a similar distribution across 10 test gold-quality isolates.**

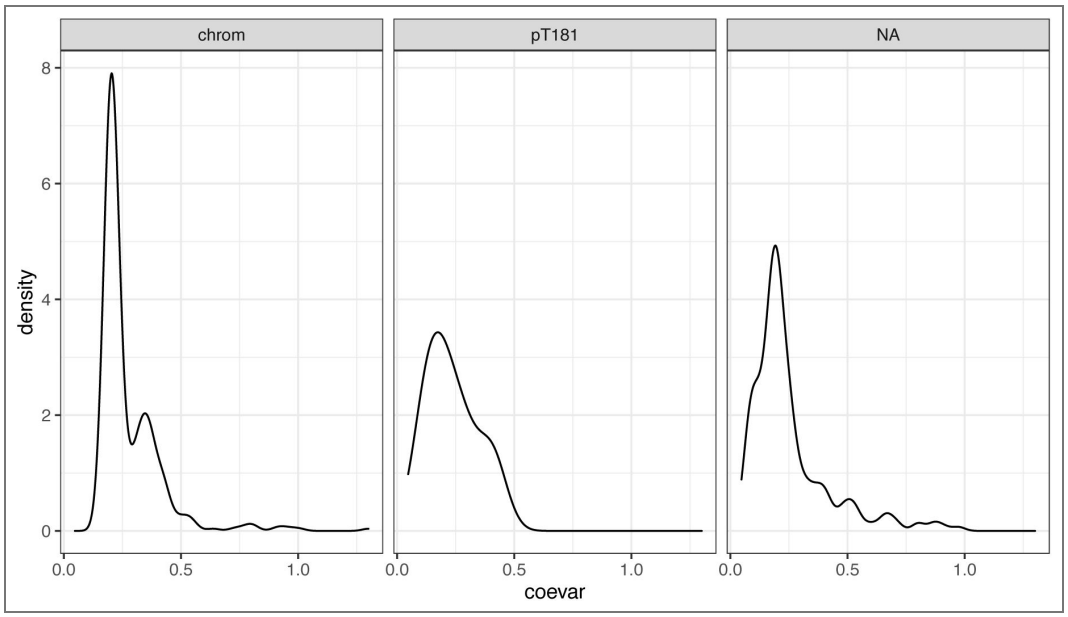
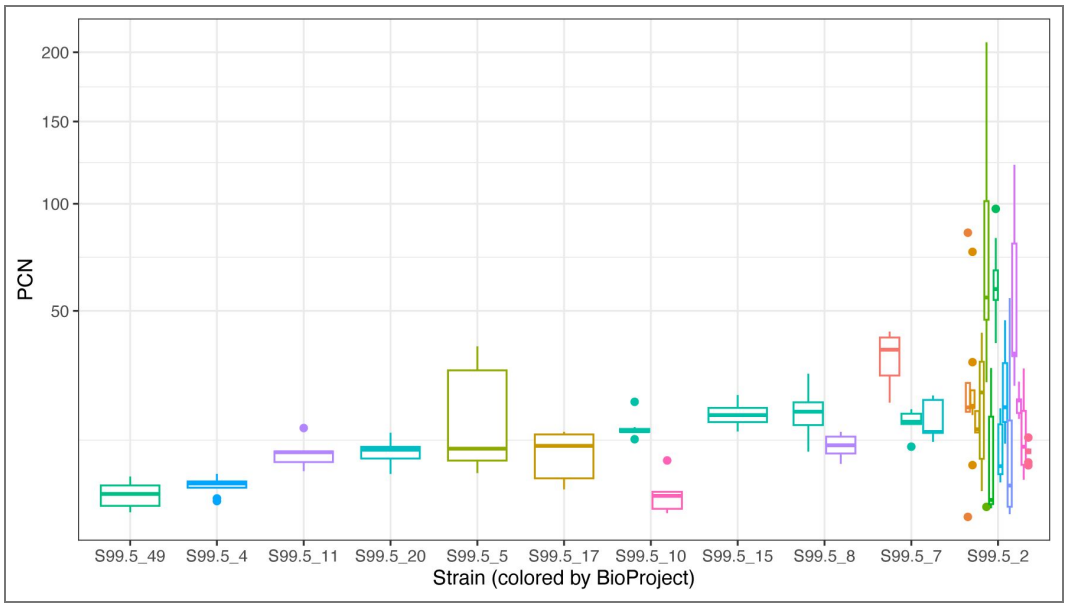


Fig. 13. **Strains vary in median copy number, with an effect of BioProject.**

Each BioProject has a different boxplot. Data shown for strain x BioProject combinations containing at least 5 isolates.



Binning samples by time

There were no samples in the n = 2,460 set collected 1992-1996. This lull in high quality samples formed a natural distinction between these two groups, given the difference in sample composition and plasmid frequency in pre- and post-1995 pT181 isolates.

Supplementary Figures

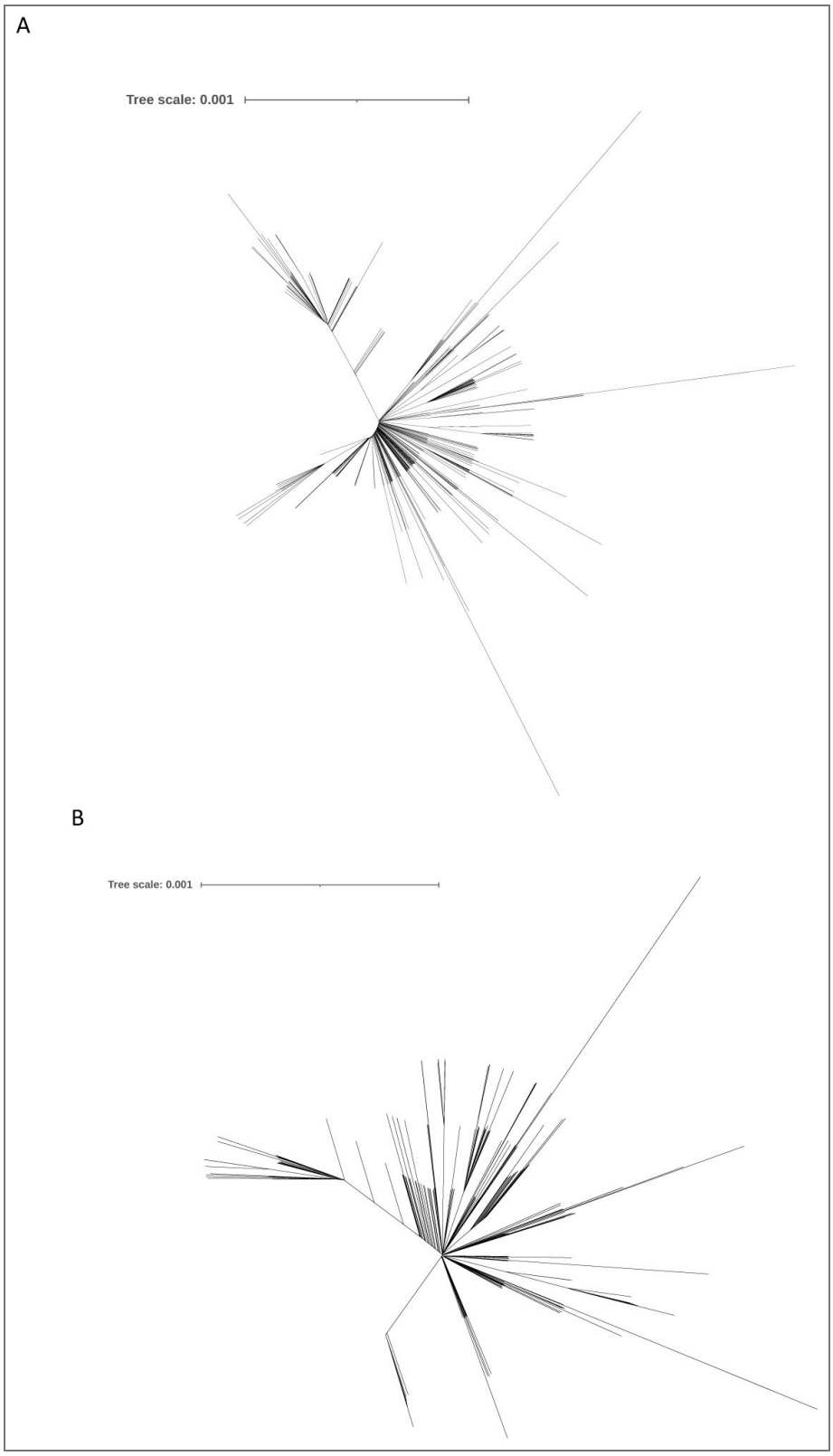


Fig. S1. **Unrooted phylogenetic trees for pT181 A) whole plasmid sequence and B) CDS.**

Fig. S2. Diagram of sample filtering steps

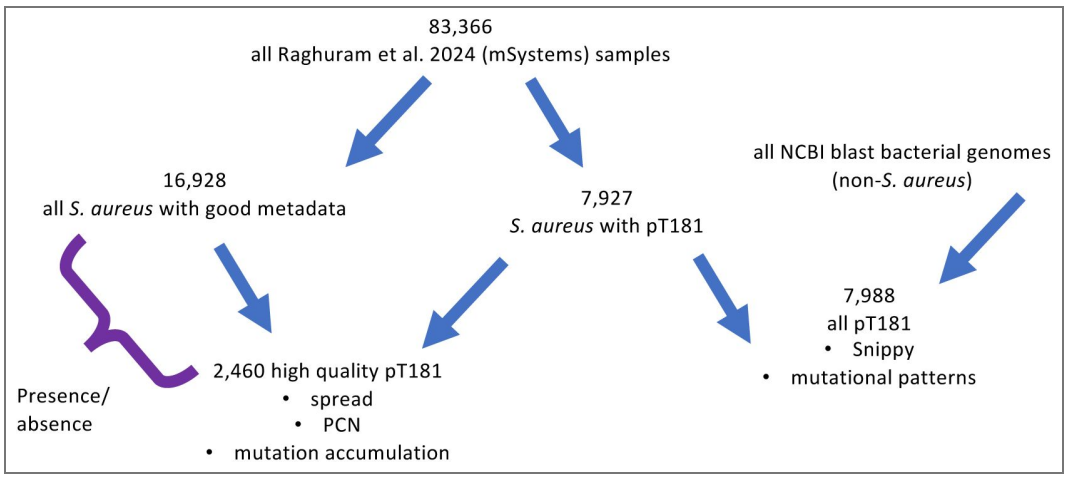
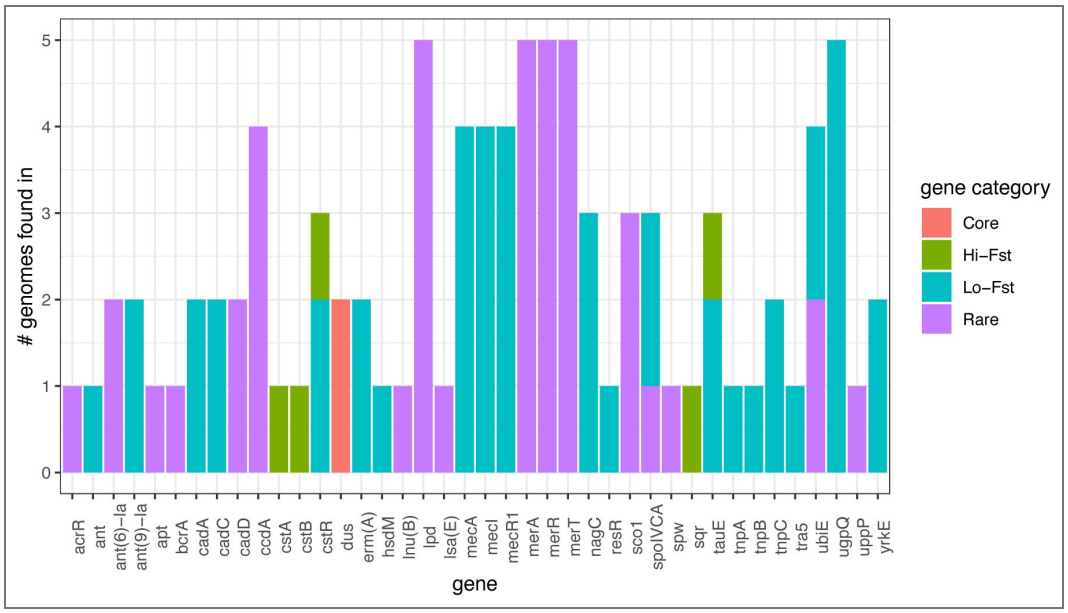


Fig. S3. pT181 integrates into a region of the chromosome consisting primarily of strain diffuse and rare genes[3].

No genes present in positions 55-90kbp in complete genomes with chromosomally integrated pT181 are universal to this region in all isolates.



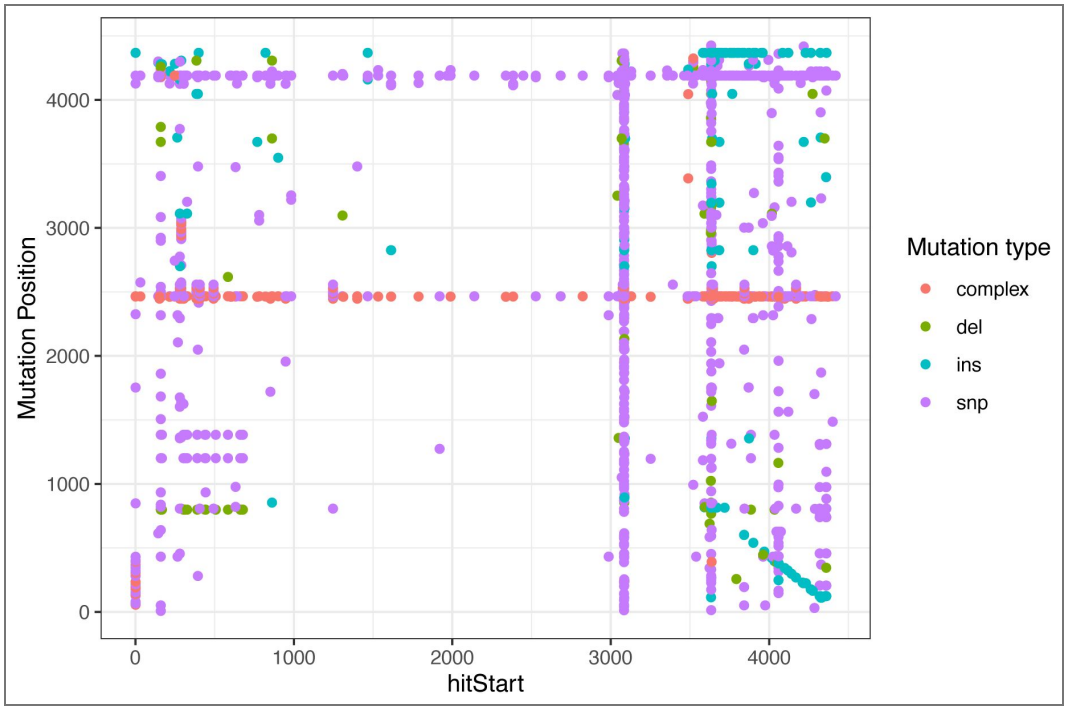


Fig. S4. Pre-rotation start position of plasmid contig does not influence position of mutations detected in pT181.

X-axis, hitStart, indicates start position of isolates genome relative to blast query (i.e., subject start / "sstart" for retrieval with blastn). Y-axis, mutation position, indicates where in the isolate mutations occurred, following their rotation to have equal start points. The presence of the mutational spikes across a range of hitStart positions indicates that it is not the rotation positions that are generating these spikes; they are biological in origin and not due to technical error.

Data availability

All genomes used in this study are publicly available. The 61 non-*S. aureus* accessions are included as a supplementary file. The summary file containing *S. aureus* SRA accessions is available on Zenodo at *https://zenodo.org/records/10471309* (staphopia_v2_all_report.txt). The 337 complete *S. aureus* genome accessions with PIRATE gene annotations are available at *https://github.com/VishnuRaghuram94/SCAPE/blob/main/timsfigsData.zip* (dnaA_start_coords.gfs). Scripts used in analysis and creating figures will be made available at *https://github.com/phillma2/pT181* following submission for publication.

Acknowledgements

M.A.P. was supported by funding from NSF 2146260, NIH awards AI158452, AI153116, and T32 AI138952 (Infectious Disease Across Scales Training Program). R.A.P. III & T.D.R. were supported by the Office of Advanced Molecular Detection, Centers for Disease Control and Prevention cooperative agreement number CK22-2204 through contract 40500-050-23234506 from the Georgia Department of Public Health. D.B.W. was supported by funding from the Sloan Foundation (Sloan Research Fellowship FG-2021-16667) and NSF 2146260. T.D.R. was supported by funding from NIH awards AI158452 and AI153116.

We thank Vishnu Raghuram and Ananya Saha for discussions about the manuscript.

Additional files

Suplemental File 1

Additional information

Funding

Funder	Grant reference number	Author
National Science Foundation (NSF)	NSF CAREER 2146260	Daniel Weissman
HHS | NIH | National Institute of Allergy and Infectious Diseases (NIAID)	AI158452	Timothy Read
HHS | NIH | National Institute of Allergy and Infectious Diseases (NIAID)	AI153116	Timothy Read
HHS | NIH | National Institute of Allergy and Infectious Diseases (NIAID)	T32 AI138952	Megan A Phillips
Alfred P. Sloan Foundation (APSF)	Sloan Research Fellowship FG-2021-16667	Daniel Weissman
HHS | Centers for Disease Control and Prevention (CDC)	CK22-2204 through contract 40500-050-23234506 from the Georgia Department of Public Health	Timothy Read

Author ORCID iDs

Megan A Phillips: https://orcid.org/0000-0002-0781-3325
Robert A Petit: https://orcid.org/0000-0002-1350-9426
Daniel B Weissman: https://orcid.org/0000-0002-7799-1573
Timothy D Read: https://orcid.org/0000-0001-8966-9680

References

(1) **Duggar B. M** (1948) Aureomycin: a product of the continuing search for new antibiotics. *Ann. N. Y. Acad. Sci* **51**:177-181 https://doi.org/10.1111/j.1749-6632.1948.tb27262.x

(2) **Pearson J. C.,** Gillett E., Gadri N. D., Dionne B (2025) Tetracyclines, the Old and the New: A Narrative Review. *CMI Commun* **2**:105059 https://doi.org/10.1016/j.cmicom.2025.105059

(3) **Raghuram V.,** Petit R. A., Karol Z., Mehta R., Weissman D. B., Read T. D (2024) Average Nucleotide Identity-Based *Staphylococcus Aureus* Strain Grouping Allows Identification of Strain-Specific Genes in the Pangenome. *mSystems* **9**:e00143-24 https://doi.org/10.1128/msystems.00143-24

(4) **Rodrigues-R L. M.,** Conrad R. E., Feistel D. J., Viver T., Rossello-Mora R., Konstantinidis K. T. (2022) A Natural Definition for a Bacterial Strain and Clonal Complex. *bioRxiv*

(5) **Maiden M. C. J.,** Bygraves J. A., Feil E., Morelli G., Russell J. E., Urwin R., Zhang Q., Zhou J., Zurth K., Caugant D. A., *et al.* (1998) Multilocus Sequence Typing: A Portable Approach to the Identification of Clones within Populations of Pathogenic Microorganisms. *Proc. Natl. Acad. Sci* **95**:3140-3145 https://doi.org/10.1073/pnas.95.6.3140

(6) **Enright M. C.,** Day N. P. J., Davies C. E., Peacock S. J., Spratt B. G (2000) Multilocus Sequence Typing for Characterization of Methicillin-Resistant and Methicillin-Susceptible Clones of Staphylococcus Aureus. *J. Clin. Microbiol* **38**:1008-1015 https://doi.org/10.1128/JCM.38.3.1008-1015.2000

(7) **Chambers H. F.,** DeLeo F. R (2009) Waves of Resistance: Staphylococcus Aureus in the Antibiotic Era. *Nat. Rev. Microbiol* **7**:629-641 https://doi.org/10.1038/nrmicro2200

(8) **Chopra I.,** Roberts M (2001) Tetracycline Antibiotics: Mode of Action, Applications, Molecular Biology, and Epidemiology of Bacterial Resistance. *Microbiol. Mol. Biol. Rev* **65**:232-260 https://doi.org/10.1128/MMBR.65.2.232-260.2001

(9) **Bassett E. J.,** Keith M. S., Armelagos G. J., Martin D. L., Villanueva A. R (1980) Tetracycline-Labeled Human Bone from Ancient Sudanese Nubia (A.D. 350). *Science* **209**:1532-1534 https://doi.org/10.1126/science.7001623

(10) **Bachmann L. H.,** Barbee L. A., Chan P., Reno H., Workowski K. A., Hoover K., Mermin J., Mena L (2024) CDC Clinical Guidelines on the Use of Doxycycline Postexposure Prophylaxis for Bacterial Sexually Transmitted Infection Prevention, United States, 2024. *MMWR Recomm Rep* **73**:1-8 https://doi.org/10.15585/mmwr.rr7302a1

(11) **Soge O. O.,** Thibault C. S., Cannon C. A., McLaughlin S. E., Menza T. W., Dombrowski J. C., Fang F. C., Golden M. R (2025) Potential Impact of Doxycycline Post-Exposure Prophylaxis on Tetracycline Resistance in *Neisseria Gonorrhoeae* and Colonization With Tetracycline-Resistant *Staphylococcus Aureus* and Group A *Streptococcus*. *Clin. Infect. Dis* ciaf089 https://doi.org/10.1093/cid/ciaf089

(12) **Harkins C. P.,** Pichon B., Doumith M., Parkhill J., Westh H., Tomasz A., de Lencastre H., Bentley S. D., Kearns A. M., Holden M. T. G (2017) Methicillin-Resistant Staphylococcus Aureus Emerged Long before the Introduction of Methicillin into Clinical Practice. *Genome Biol* **18**:1-11 https://doi.org/10.1186/s13059-017-1252-9

(13) **Jevons M. P** (1961) "Celbenin" - Resistant Staphylococci. *BMJ* **1**:124-125 https://doi.org/10.1136/bmj.1.5219.124-a

(14) **Gillespie M. T.,** May J. W., Skurray R. A (1985) Antibiotic Resistance in Staphylococcus Aureus Isolated at an Australian Hospital between 1946 and 1981. *J. Med. Microbiol* **19**:137-147 https://doi.org/10.1099/00222615-19-2-137

(15) **Khan S. A.,** Novick R. P (1983) Complete Nucleotide Sequence of pT181, a Tetracycline-Resistance Plasmid from Staphylococcus Aureus. *Plasmid* **10**:251-259 https://doi.org/10.1016/0147-619X(83)90039-2

(16) **Novick R. P.,** Projan S. J., Rosenblum W., Edelman I (1984) Staphylococcal Plasmid Cointegrates Are Formed by Host- and Phage-Mediated General Rec Systems That Act on Short Regions of Homology. *Mol. Gen. Genet. MGG* **195**:374-377 https://doi.org/10.1007/BF00332777

(17) **Novick R. P.** (1989) Staphylococcal plasmids and their replication. *Annu. Rev. Microbiol* **43**:537-563 https://doi.org/10.1146/annurev.mi.43.100189.002541

(18) **Kwong S. M.,** Ramsay J. P., Jensen S. O., Firth N (2017) Replication of Staphylococcal Resistance Plasmids. *Front. Microbiol* **8**:1-16 https://doi.org/10.3389/fmicb.2017.02279

(19) **Novick R. P.,** Adler G. K., Projan S. J., Carleton S., Highlander S. K., Gruss A., Khan S. A., Iordanescu S (1984) Control of pT181 Replication I. The pT181 Copy Control Function Acts by Inhibiting the Synthesis of a Replication Protein. *EMBO J* **3**:2399-2405 https://doi.org/10.1002/j.1460-2075.1984.tb02146.x

(20) **Carleton S.,** Projan S. J., Highlander S. K., Moghazeh S. M., Novick R. P (1984) Control of pT181 Replication II. Mutational Analysis. *EMBO J* **3**:2407-2414 https://doi.org/10.1002/j.1460-2075.1984.tb02147.x

(21) **Wang P.-Z.,** Henriquez V. B., Projan S. J., Iordanescu S., Novick R. P (1991) The Effect of Plasmid Copy Number Mutations on pT181 Replication Initiator Protein Expression. *Plasmid* **25**:198-207 https://doi.org/10.1016/0147-619X(91)90013-M

(22) **Novick R** (1976) Plasmid-Protein Relaxation Complexes in Staphylococcus Aureus. *J. Bacteriol* **127**:1177-1187 https://doi.org/10.1128/jb.127.3.1177-1187.1976

(23) **Taiaroa G.,** Samuelsen Ø., Kristensen T., Økstad O. A. L., Heikal A (2018) Complete Genome Sequence of Pseudomonas Aeruginosa K34-7, a Carbapenem-Resistant Isolate of the High-Risk Sequence Type 233. *Microbiol. Resour. Announc* **7** https://doi.org/10.1128/MRA.00886-18

(24) **Inoue M.,** Mitsuhashi S (1976) Recombination between Phage S1 and the TC-Resistant Gene on Staphylococcus Aureus Plasmid. *Virology* **72**:322-329 https://doi.org/10.1016/0042-6822(76)90161-6

(25) **Baltrus D. A** (2013) Exploring the Costs of Horizontal Gene Transfer. *Trends Ecol. Evol* **28**:489-495 https://doi.org/10.1016/j.tree.2013.04.002

(26) **Vogwill T.,** MacLean R. C (2015) The Genetic Basis of the Fitness Costs of Antimicrobial Resistance: A Meta-analysis Approach. *Evol Appl* **8**:284-295 https://doi.org/10.1111/eva.12202

(27) **San Millan A.,** MacLean R. C. (2017) Fitness Costs of Plasmids: A Limit to Plasmid Transmission. *Microbiol. Spectr* **5** https://doi.org/10.1128/microbiolspec.MTBP-0016-2017

(28) **Andersson D. I.,** Levin B. R (1999) The Biological Cost of Antibiotic Resistance. *Curr. Opin. Microbiol* **2**:489-493 https://doi.org/10.1016/s1369-5274(99)00005-3

(29) **Summers D. K** (1991) The Kinetics of Plasmid Loss. *Trends Biotechnol* **9**:273-278 https://doi.org/10.1016/0167-7799(91)90089-Z

(30) **Kumar C. C.,** Novick R. P (1985) Plasmid pT181 Replication Is Regulated by Two Countertranscripts. *Proc. Natl. Acad. Sci* **82**:638-642 https://doi.org/10.1073/pnas.82.3.638

(31) **Novick R. P.,** Iordanescu S., Projan S. J., Kornblum J., Edelman I (1989) pT181 Plasmid Replication Is Regulated by a Countertranscript-Driven Transcriptional Attenuator. *Cell* **59**:395-404 https://doi.org/10.1016/0092-8674(89)90300-0

(32) **Katayama Y.,** Ito T., Hiramatsu K (2000) A New Class of Genetic Element, Staphylococcus Cassette Chromosome *Mec*, Encodes Methicillin Resistance in *Staphylococcus Aureus*. *Antimicrob Agents Chemother* **44**:1549-1555 https://doi.org/10.1128/AAC.44.6.1549-1555.2000

(33) **Ito T.,** Katayama Y., Asada K., Mori N., Tsutsumimoto K., Tiensasitorn C., Hiramatsu K (2001) Structural Comparison of Three Types of Staphylococcal Cassette Chromosome *Mec* Integrated in the Chromosome in Methicillin-Resistant *Staphylococcus Aureus*. *Antimicrob Agents Chemother* **45**:1323-1336 https://doi.org/10.1128/AAC.45.5.1323-1336.2001

(34) **Côrtes M. F.,** Botelho A. M. N., Almeida L. G. P., Souza R. C., Cunha O. D. L., Nicolás M. F., Vasconcelos A. T. R., Figueiredo A. M. S (2018) Community-Acquired Methicillin-Resistant Staphylococcus Aureus from ST1 Lineage Harboring a New SCCmec IV Subtype (SCCmec IVm) Containing the tetK Gene. *Infect. Drug Resist* **11**:2583-2592 https://doi.org/10.2147/IDR.S175079

(35) **Needham C.,** Rahman M., Dyke K. G. H., Noble W. C (1994) An Investigation of Plasmids from Staphylococcus Aureus That Mediate Resistance to Mupirocin and Tetracycline. *Microbiology* **140**:2577-2583 https://doi.org/10.1099/00221287-140-10-2577

(36) **Stewart P. R.,** Dubin D. T., Chikramane S. G., Inglis B., Matthews P. R., Poston S. M (1994) IS257 and Small Plasmid Insertions in the Mec Region of the Chromosome of Staphylococcus Aureus. *Plasmid* **31**:12-20 https://doi.org/10.1006/plas.1994.1002

(37) **Werckenthin C.,** Schwarz S., Roberts M. C (1996) Integration of pT181-like Tetracycline Resistance Plasmids into Large Staphylococcal Plasmids Involves IS257. *Antimicrob Agents Chemother* **40**:2542-2544 https://doi.org/10.1128/AAC.40.11.2542

(38) **Camacho C.,** Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T. L. (2009) BLAST+: Architecture and Applications. *BMC Bioinformatics* **10**:421 https://doi.org/10.1186/1471-2105-10-421

(39) **Seemann T.** (no date) Snippy. https://github.com/tseemann/snippy

(40) **Duchêne S.,** Holt K. E., Weill F.-X., Le Hello S., Hawkey J., Edwards D. J., Fourment M., Holmes E. C (2016) Genome-Scale Rates of Evolutionary Change in Bacteria. *Microb*. *Genomics* **2** https://doi.org/10.1099/mgen.0.000094

(41) **Gibson B.,** Wilson D. J., Feil E., Eyre-Walker A (2018) The Distribution of Bacterial Doubling Times in the Wild. *Proc. R. Soc. B Biol. Sci* **285**:20180789 https://doi.org/10.1098/rspb.2018.0789

(42) **Young B. C.,** Golubchik T., Batty E. M., Fung R., Larner-Svensson H., Votintseva A. A., Miller R. R., Godwin H., Knox K., Everitt R. G., *et al.* (2012) Evolutionary Dynamics of Staphylococcus Aureus during Progression from Carriage to Disease. *Proc. Natl. Acad. Sci* **109**:4550-4555 https://doi.org/10.1073/pnas.1113219109

(43) **Holden M. T. G.,** Hsu L.-Y., Kurt K., Weinert L. A., Mather A. E., Harris S. R., Strommenger B., Layer F., Witte W., De Lencastre H., *et al.* (2013) A Genomic Portrait of the Emergence, Evolution, and Global Spread of a Methicillin-Resistant *Staphylococcus Aureus* Pandemic. *Genome Res* **23**:653-664 https://doi.org/10.1101/gr.147710.112

(44) **Yebra G.,** Harling-Lee J. D., Lycett S., Aarestrup F. M., Larsen G., Cavaco L. M., Seo K. S., Abraham S., Norris J. M., Schmidt T., *et al.* (2022) Multiclonal Human Origin and Global Expansion of an Endemic Bacterial Pathogen of Livestock. *Proc. Natl. Acad. Sci* **119**:e2211217119 https://doi.org/10.1073/pnas.2211217119

(45) **Bruen T. C.,** Philippe H., Bryant D (2006) A Simple and Robust Statistical Test for Detecting the Presence of Recombination. *Genetics* **172**:2665-2681 https://doi.org/10.1534/genetics.105.048975

(46) **Gillespie M. T.,** May J. W., Skurray R. A (1986) Detection of an Integrated Tetracycline Resistance Plasmid in the Chromosome of Methicillin-Resistant Staphylococcus Aureus. *Microbiology* **132**:1723-1728 https://doi.org/10.1099/00221287-132-6-1723

(47) **Gillespie M. T.,** Lyon B. R., Loo L. S. L., Matthews P. R., Stewart P. R., Skurray R. A (1987) Homologous Direct Repeat Sequences Associated with Mercury, Methicillin, Tetracycline and Trimethoprim Resistance Determinants in *Staphylococcus Aureus*. *FEMS Microbiol Lett* **43**:165-171 https://doi.org/10.1111/j.1574-6968.1987.tb02117.x

(48) **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. *J. Bacteriol* **190**:300-310 https://doi.org/10.1128/JB.01000-07

(49) **Everitt R. G.,** Didelot X., Batty E. M., Miller R. R., Knox K., Young B. C., Bowden R., Auton A., Votintseva A., Larner-Svensson H., *et al.* (2014) Mobile Elements Drive Recombination Hotspots in the Core Genome of Staphylococcus Aureus. *Nat. Commun* **5**:3956 https://doi.org/10.1038/ncomms4956

(50) **Maddamsetti R.,** Wilson M. L., Son H.-I., Zhou Z., Lu J., You L (2024) Scaling Laws of Plasmids across the Microbial Tree of Life. *bioRxiv* https://doi.org/10.1101/2024.10.04.616653

(51) **Projan S.,** Carleton S., Novick R. (1983) Determination of Plasmid Copy Number by Fluorescence Densitometry. *Plasmid* **9**:182-190 https://doi.org/10.1016/0147-619x(83)90019-7

(52) **Iordanescu S** (1976) *Arch Roum Pathol Exp Microbiol* **35**:207-215

(53) **Novick R. P.,** Adler G. K., Majumder S., Khan S. A., Carleton S., Rosenblum W. D., Iordanescu S (1982) Coding Sequence for the pT181 repC Product: A Plasmid-Coded Protein Uniquely Required for Replication. *Proc. Natl. Acad. Sci. U. S. A* **79**:4108-4112 https://doi.org/10.1073/pnas.79.13.4108

(54) **Todd R. T.,** Selmecki A. (2023) Copy Number Variation and Allele Ratio Analysis in *Candida Albicans* Using Whole Genome Sequencing Data. *Antifungal Drug Resistance*

(55) **Kendall J.,** Krasnitz A (2014) Computational Methods for DNA Copy-Number Analysis of Tumors. *Methods Mol Biol* **1176**:243-259 https://doi.org/10.1007/978-1-4939-0992-6_20

(56) **Needham C.,** Noble W. C., Dyke K. G. H (1995) The Staphylococcal Insertion Sequence IS257Is Active. *Plasmid* **34**:198-205 https://doi.org/10.1006/plas.1995.0005

(57) **Simpson A. E.,** Skurray R. A., Firth N (2000) An IS *257* -Derived Hybrid Promoter Directs Transcription of a *tetA* (K) Tetracycline Resistance Gene in the *Staphylococcus Aureus* Chromosomal *Mec* Region. *J. Bacteriol* **182**:3345-3352 https://doi.org/10.1128/JB.182.12.3345-3352.2000

(58) **Varani A.,** He S., Siguier P., Ross K., Chandler M (2021) The IS6 Family, a Clinically Important Group of Insertion Sequences Including IS26. *Mob DNA* **12**:11 https://doi.org/10.1186/s13100-021-00239-x

(59) **Partridge S. R.,** Kwong S. M., Firth N., Jensen S. O (2018) Mobile Genetic Elements Associated with Antimicrobial Resistance. *Clin. Microbiol. Rev* **31** https://doi.org/10.1128/CMR.00088-17

(60) **Rouch D. A.,** Skurray R. A (1989) IS257 from Staphylococcus Aureus: Member of an Insertion Sequence Superfamily Prevalent among Gram-Positive and Gram-Negative Bacteria. *Gene* **76**:195-205 https://doi.org/10.1016/0378-1119(89)90160-1

(61) **Shannon K.,** Williams H., King A., Phillips I (1990) Hyperproduction of TEM-1 Beta-Lactamase in Clinical Isolates of *Escherichia Coli* Serotype O15. *FEMS Microbiol Lett* **67**:319-323 https://doi.org/10.1111/j.1574-6968.1990.tb04040.x

(62) **San Millan A.,** Santos-Lopez A., Ortega-Huedo R., Bernabe-Balas C., Kennedy S. P, Gonzalez-Zorn B. (2015) Small-Plasmid-Mediated Antibiotic Resistance Is Enhanced by Increases in Plasmid Copy Number and Bacterial Fitness. *Antimicrob Agents Chemother* **59**:3335-3341 https://doi.org/10.1128/AAC.00235-15

(63) **Dimitriu T.,** Matthews A. C., Buckling A (2021) Increased Copy Number Couples the Evolution of Plasmid Horizontal Transmission and Plasmid-Encoded Antibiotic Resistance. *Proc. Natl. Acad. Sci* **118**:e2107818118 https://doi.org/10.1073/pnas.2107818118

(64) **French G. L.,** Shannon K. P., Simmons N (1996) Hospital Outbreak of Klebsiella Pneumoniae Resistant to Broad-Spectrum Cephalosporins and Beta-Lactam-Beta-Lactamase Inhibitor Combinations by Hyperproduction of SHV-5 Beta-Lactamase. *J. Clin. Microbiol* **34**:358-363 https://doi.org/10.1128/jcm.34.2.358-363.1996

(65) **Ramiro-Martínez P.,** De Quinto I., Lanza V. F., Gama J. A., Rodríguez-Beltrán J (2025) Universal Rules Govern Plasmid Copy Number. *Nat. Commun* **16**:6022 https://doi.org/10.1038/s41467-025-61202-5

(66) **Guay G. G.,** Rothstein D. M (1993) Expression of the tetK Gene from Staphylococcus Aureus in Escherichia Coli: Comparison of Substrate Specificities of TetA(B), TetA(C), and TetK Efflux Proteins. *Antimicrob Agents Chemother* **37**:191-198 https://doi.org/10.1128/AAC.37.2.191

(67) **Novick R. P.,** Adler G. K., Projan S. J., Carleton S., Highlander S. K., Gruss A., Khan S. A., Iordanescu S (1984) Control of pT181 Replication I. The pT181 Copy Control Function Acts by Inhibiting the Synthesis of a Replication Protein. *EMBO J* **3**:2399-2405 https://doi.org/10.1002/j.1460-2075.1984.tb02146.x

(68) **Gillespie M. T.,** May J. W., Skurray R. A (1984) Antibiotic Susceptibilities and Plasmid Profiles of Nosocomial Methicillin-Resistant Staphylococcus Aureus: A Retrospective Study. *J. Med. Microbiol* **17**:295-310 https://doi.org/10.1099/00222615-17-3-295

(69) **Harris S. R.,** Feil E. J., Holden M. T. G., Quail M. A., Nickerson E. K., Chantratita N., Gardete S., Tavares A., Day N., Lindsay J. A., *et al.* (2010) Evolution of MRSA During Hospital Transmission and Intercontinental Spread. *Science* **327**:469-474 https://doi.org/10.1126/science.1182395

(70) **Crisóstomo M. I.,** Westh H., Tomasz A., Chung M., Oliveira D. C., de Lencastre H (2001) The Evolution of Methicillin Resistance in Staphylococcus Aureus: Similarity of Genetic Backgrounds in Historically Early Methicillin-Susceptible and -Resistant Isolates and Contemporary Epidemic Clones. *Proc. Natl. Acad. Sci. U. S. A* **98**:9865-9870 https://doi.org/10.1073/pnas.161272898

(71) **Benvenga V.,** Cuénod A., Purushothaman S., Dasen G., Weisser M., Bassetti S., Roloff T., Siegemund M., Heininger U., Bielicki J., *et al.* (2024) Historic Methicillin-Resistant Staphylococcus Aureus: Expanding Current Knowledge Using Molecular Epidemiological Characterization of a Swiss Legacy Collection. *Genome Med* **16**:23 https://doi.org/10.1186/s13073-024-01292-w

(72) **Yamaguchi A.,** Shiina Y., Fujihira E., Sawai T., Noguchi N., Sasatsu M (1995) The Tetracycline Efflux Protein Encoded by the Tet (K) Gene from Staphylococcus Aureus Is a Metal-Tetracycline/H + Antiporter. *FEBS Lett* **365**:193-197 https://doi.org/10.1016/0014-5793(95)00455-I

(73) **Paulsen I. T.,** Brown M. H., Skurray R. A (1996) Proton-Dependent Multidrug Efflux Systems. *Microbiol. Rev* **60**:575-608 https://doi.org/10.1128/mr.60.4.575-608.1996

(74) **Souvorov A.,** Agarwala R., Lipman D. J (2018) SKESA: Strategic k-Mer Extension for Scrupulous Assemblies. *Genome Biol* **19**:153 https://doi.org/10.1186/s13059-018-1540-z

(75) **Souvorov A.,** Agarwala R (2021) SAUTE: Sequence Assembly Using Target Enrichment. *BMC Bioinformatics* **22**:375 https://doi.org/10.1186/s12859-021-04174-9

(76) **Petit R. A.,** Read T. D (2020) Bactopia: A Flexible Pipeline for Complete Analysis of Bacterial Genomes. *mSystems* **5** https://doi.org/10.1128/mSystems.00190-20

(77) **Altschul S. F.,** Gish W., Miller W., Myers E. W., Lipman D. J (1990) Basic Local Alignment Search Tool. *J. Mol. Biol* **215**:403-410 https://doi.org/10.1016/S0022-2836(05)80360-2

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

(79) **Harris S** (2018) SKA: Split Kmer Analysis Toolkit for Bacterial Genomic Epidemiology. *bioRxiv* https://doi.org/10.1101/453142

(80) **Posit Team** (no date) RStudio: Integrated Development Environment for R. Posit Software.

(81) **R Core Team** (2020) R: A Language and Environment for Statistical Computing. https://www.R-project.org/

(82) **Wickham H.,** Averick M., Bryan J., Chang W., McGowan L., François R., Grolemund G., Hayes A., Henry L., Hester J., *et al.* (2019) Welcome to the Tidyverse. *J. Open Source Softw* **4**:1686 https://doi.org/10.21105/joss.01686

(83) **Petit Robert A,** Read T. D. (2024) Sccmec: A Tool for Typing SCCmec Cassettes in Assemblies. Github. https://github.com/rpetit3/sccmec

(84) **Li H.,** Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. (2009) Subgroup, 1000 Genome Project Data Processing. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **25**:2078-2079 https://doi.org/10.1093/bioinformatics/btp352

(85) **Broad Institute** (no date) Picard. https://broadinstitute.github.io/picard/

(86) **Schwengers O.,** Jelonek L., Dieckmann M. A., Beyvers S., Blom J., Goesmann A (2021) Bakta: Rapid and Standardized Annotation of Bacterial Genomes via Alignment-Free Sequence Identification. *Microb Genomics* **7** https://doi.org/10.1099/mgen.0.000685

(87) **Wilkins D.** (2023) Gggenes: Draw Gene Arrow Maps in "Ggplot2". https://wilkox.org/gggenes/

(88) **Bruen T.** (no date) PhiPack: PHI Test and Other Tests of Recombination. https://www.maths.otago.ac.nz/~dbryant/software.html

(89) **Li H.,** Durbin R (2010) Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* **26**:589-595 https://doi.org/10.1093/bioinformatics/btp698

(90) **Bayliss S. C.,** Thorpe H. A., Coyle N. M., Sheppard S. K., Feil E. J (2019) PIRATE: A Fast and Scalable Pangenomics Toolbox for Clustering Diverged Orthologues in Bacteria. *GigaScience* **8**:1-9 https://doi.org/10.1093/gigascience/giz119

(91) **Kuroda M.,** Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I., Cui L., Oguchi A., Aoki K., Nagai Y., *et al.* (2001) Whole Genome Sequencing of Meticillin-Resistant Staphylococcus Aureus. *The Lancet* **357**:1225-1240 https://doi.org/10.1016/S0140-6736(00)04403-2

(92) **Brooks M.,** Kristensen K., van Benthem K. J., Magnusson A., Berg C. W., Nielsen A., Skaug H. J., Mächler M., Bolker B. M. (2017) glmmTMB Balances Speed and Flexibility Among Packages for Zero-Inflated Generalized Linear Mixed Modeling. *R J* **9**:378 https://doi.org/10.32614/RJ-2017-066

(93) **Lenth Russell V.** (no date) Emmeans. https://rvlenth.github.io/emmeans/

(94) **Hyatt D.,** Chen G.-L., LoCascio P. F., Land M. L., Larimer F. W., Hauser L. J (2010) Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinformatics* **11**:119 https://doi.org/10.1186/1471-2105-11-119

(95) **Gilchrist C. L. M.,** Chooi Y.-H. Clinker, Clustermap Js (2021) Automatic Generation of Gene Cluster Comparison Figures. *Bioinformatics* **37**:2473-2475 https://doi.org/10.1093/bioinformatics/btab007

Peer reviews

Reviewer #1 (Public review):

The study provides a robust bioinformatic characterization of the evolution of pT181. My main criticism of the work is the lack of experimental validation for the hypotheses proposed by the authors.

Comments on the study:

(1) One potential reason for the decline in pT181 copy number over time may be a high cost associated with the multicopy state. In this sense, it would be interesting if the authors could use (or construct) isogenic strains differing only in the state of the plasmid (multicopy/integrated). With this system, the authors could measure the fitness of the strains in the presence and absence of tetracycline, and they could be able to understand the benefit associated with the plasmid transition. The authors discuss these ideas, but it would be nice to test them.

(2) It would be interesting to know the transfer frequencies of the multicopy mobilizable pT181 plasmid, compared to the transfer frequency of the plasmid integrated into the SSCmec element (which can be co-transferred, integrated in conjugative plasmids, or by transduction).

(3) One important limitation of the study that should be mentioned is that inferring pT181 PCN from whole genome data can be problematic. For example, some DNA extraction methods may underestimate the copy number of small plasmids because the small, circular plasmids are preferentially depleted during the process (see, for example, *https://www.nature.com/articles/srep28063*).

https://doi.org/10.7554/eLife.109393.1.sa2

Reviewer #2 (Public review):

Summary:

The authors performed bioinformatic analyses to trace the genomic history of the clinically relevant pT181 plasmid. Specifically, they:

(1) tracked the presence of pT181 across different S. aureus strain backgrounds through time. It was first found in one, later multiple strains, though this may reflect changes in sampling over time.

(2) estimated the mutation rate of the chromosome and plasmid.

(3) estimated the plasmid copy number of pT181, and found that it decreased over time. The latter was supported by two sets of statistical analyses, first showing that the number of single-copy isolates increased over time, and second, that the multicopy isolates demonstrated a lower PCN over time.

(4) reported the different integration sites at which pT181 integrated into the genome.

As a caveat, they mentioned that identical plasmid sequences have variable plasmid copy numbers across different genomes in their dataset.

Strengths:

This is a very solid, well-considered bioinformatic study on publicly available data. I greatly appreciate the thoughtful approach the authors have taken to their subject matter, neither over- nor underselling their results. It is a strength that the authors focussed on a single plasmid in a single bacterial species, as it allowed them to take into account unique knowledge about the biology of this system and really dive deep into the evolution of this specific plasmid. It makes for a compelling case study. At the same time, I think the introduction and discussion can be strengthened to demonstrate what lessons might be drawn from this case study for other plasmids.

Weaknesses:

The finding that the pT181 copy number declined over time is the most interesting claim of the paper to me, and not something that I have seen done before. While the authors have looked at some confounders in this analysis, I think this could be strengthened further in a revision.

For the flow of the storyline, I also think the estimation of mutation rates (starting L181) and integration into the chromosome (starting L255) could be moved to the supplement or a later position in the main text.

Clearly, the use of publicly available data prevents the authors from controlling the growth and sequencing conditions of the isolates. It is striking that they observe a clear signal in spite of this, but I would have loved to see more discussion of the metadata that came with the publicly available sequences and even more use of that metadata to control for confounding.

https://doi.org/10.7554/eLife.109393.1.sa1

Author response:

> ***eLife Assessment***
>
> *Using genome databases, the authors performed solid bioinformatic analyses to trace the genomic history of the clinically relevant Staphylococcus aureus tetracycline resistance plasmid pT181 over the last seven decades. They discovered that this element has transitioned from a multicopy plasmid to a chromosomally integrated element, and the work represents a valuable demonstration of the use of publicly available data to investigate plasmid biology and inform clinical epidemiology. This work will appeal to researchers interested in staphylococcal evolution and plasmid biology.*

Thank you, we agree with this overview. We also think this work is interesting to people interested in antimicrobial resistance and bacterial genome structure.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public review):***
>
> *The study provides a robust bioinformatic characterization of the evolution of pT181. My main criticism of the work is the lack of experimental validation for the hypotheses proposed by the authors.*
>
> *Comments on the study:*
>
> *(1) One potential reason for the decline in pT181 copy number over time may be a high cost associated with the multicopy state. In this sense, it would be interesting if the authors could use (or construct) isogenic strains differing only in the state of the plasmid (multicopy/integrated). With this system, the authors could measure the fitness of the strains in the presence and absence of tetracycline, and they could be able to understand the benefit associated with the plasmid transition. The authors discuss these ideas, but it would be nice to test them.*

We agree that the relative fitness of integrated versus multicopy plasmids is interesting and a costly multicopy state could explain the transition of independent pT181 replicons to chromosomal integration. This is a project we are exploring for a future study. However, we think that this additional experimental work goes beyond the scope of the paper.

> *(2) It would be interesting to know the transfer frequencies of the multicopy mobilizable pT181 plasmid, compared to the transfer frequency of the plasmid integrated into the SSCmec element (which can be co-transferred, integrated in conjugative plasmids, or by transduction).*

We agree with the reviewer that this is an interesting question. However, we think inferring these rates from natural sequence data is not feasible in this case given the low heterogeneity of the plasmid sequence. A laboratory-based experimental study could not address the real transfers we observe over the course of decades, as *in vitro S. aureus* transfer rates are often not good proxies for *in vivo* (McCarthy et al., 2014). In addition, we do not know what is moving the integrated plasmid. pT181 could be moved by a phage or plasmid, so we are uncertain what the correct experiment would be to explore this.

> *(3) One important limitation of the study that should be mentioned is that inferring pT181 PCN from whole genome data can be problematic. For example, some DNA extraction methods may underestimate the copy number of small plasmids because the small, circular plasmids are preferentially depleted during the process (see, for example, https://www.nature.com/articles/srep28063).*

We will investigate this issue further in the revisions. The kits used to extract DNA for the earlier-collected samples may possibly yield more plasmid DNA relative to the chromosome compared to newer ones on average; however, we think this is not driving the decline that we observe in multicopy pT181 copy number. Multiple BioProjects find the same result, where earlier samples have higher copy number compared to later samples. We expect extraction methods to be consistent within a BioProject, suggesting that this decline is genuine and not technical. In revisions, we intend to evaluate the effect of date of sequencing and additional metadata on copy number.

> ***Reviewer #2 (Public review):***
>
> *Summary:*

> *The authors performed bioinformatic analyses to trace the genomic history of the clinically relevant pT181 plasmid. Specifically, they:*
>
> *(1) Tracked the presence of pT181 across different S. aureus strain backgrounds through time. It was first found in one, later multiple strains, though this may reflect changes in sampling over time.*
>
> *(2) Estimated the mutation rate of the chromosome and plasmid.*
>
> *(3) Estimated the plasmid copy number of pT181, and found that it decreased over time. The latter was supported by two sets of statistical analyses, first showing that the number of single-copy isolates increased over time, and second, that the multicopy isolates demonstrated a lower PCN over time.*
>
> *(4) Reported the different integration sites at which pT181 integrated into the genome.*
>
> *As a caveat, they mentioned that identical plasmid sequences have variable plasmid copy numbers across different genomes in their dataset.*
>
> *Strengths:*
>
> *This is a very solid, well-considered bioinformatic study on publicly available data. I greatly appreciate the thoughtful approach the authors have taken to their subject matter, neither over- nor underselling their results. It is a strength that the authors focused on a single plasmid in a single bacterial species, as it allowed them to take into account unique knowledge about the biology of this system and really dive deep into the evolution of this specific plasmid. It makes for a compelling case study. At the same time, I think the introduction and discussion can be strengthened to demonstrate what lessons might be drawn from this case study for other plasmids.*
>
> *Weaknesses:*
>
> *The finding that the pT181 copy number declined over time is the most interesting claim of the paper to me, and not something that I have seen done before. While the authors have looked at some confounders in this analysis, I think this could be strengthened further in a revision.*

In the revisions, we will further explore the impact that technical variation could have in contributing to copy number variation and update our claims for the decline in copy number of the independent replicon over time and variation for the same plasmid sequence accordingly. Multiple BioProjects show earlier samples have higher copy number compared to later samples; we expect extraction methods to be consistent within a BioProject, supporting our initial findings that this decline over time is not due to technical variation.

> *For the flow of the storyline, I also think the estimation of mutation rates (starting L181) and integration into the chromosome (starting L255) could be moved to the supplement or a later position in the main text.*

We will revisit the text organization for flow and clarity of storyline.

> *Clearly, the use of publicly available data prevents the authors from controlling the growth and sequencing conditions of the isolates. It is striking that they observe a clear signal in spite of this, but I would have loved to see more discussion of the metadata that came with the publicly available sequences and even more use of that metadata to control for confounding.*

In revisions, we will further investigate possible contributors to the observed decline in copy number of multicopy pT181 over time. We have incorporated the date of sample collection

and BioProject in our analysis, but not the date of sequencing or extraction technique.

References

McCarthy, A. J., Loeffler, A., Witney, A. A., Gould, K. A., Lloyd, D. H., & Lindsay, J. A. (2014). Extensive horizontal gene transfer during Staphylococcus aureus co-colonization in vivo. Genome Biology and Evolution, 6(10), 2697–2708. *https://doi.org/10.1093/gbe/evu214*

https://doi.org/10.7554/eLife.109393.1.sa0