JCI insight

A population-level strain genotyping method to study pathogen strain dynamics in human infections

Sarah J. Morgan,[1] Samantha L. Durfey,[1] Sumedha Ravishankar,[1] Peter Jorth,[2] Wendy Ni,[1] Duncan T. Skerrett,[1] Moira L. Aitken,[3] Edward F. McKone,[4] Stephen J. Salipante,[5] Matthew C. Radey,[1] and Pradeep K. Singh[1,3]

[1]Department of Microbiology, University of Washington School of Medicine, Seattle, Washington, USA. [2]Department of Pathology and Laboratory Medicine, Cedars-Sinai Medical Center, Los Angeles, California, USA. [3]Department of Medicine, University of Washington School of Medicine, Seattle, Washington, USA. [4]St. Vincent's University Hospital, Dublin, Ireland. [5]Department of Laboratory Medicine and Pathology, University of Washington School of Medicine, Seattle, Washington, USA.

A hallmark of chronic bacterial infections is the long-term persistence of 1 or more pathogen species at the compromised site. Repeated detection of the same bacterial species can suggest that a single strain or lineage is continually present. However, infection with multiple strains of a given species, strain acquisition and loss, and changes in strain relative abundance can occur. Detecting strain-level changes and their effects on disease is challenging because most methods require labor-intensive isolate-by-isolate analyses, and thus, only a few cells from large infecting populations can be examined. Here, we present a population-level method for enumerating and measuring the relative abundance of strains called population multi-locus sequence typing (PopMLST). The method exploits PCR amplification of strain-identifying polymorphic loci, next-generation sequencing to measure allelic variants, and informatic methods to determine whether variants arise from sequencing errors or low-abundance strains. These features enable PopMLST to simultaneously interrogate hundreds of bacterial cells that are cultured en masse from patient samples or are present in DNA directly extracted from clinical specimens without ex vivo culture. This method could be used to detect epidemic or super-infecting strains, facilitate understanding of strain dynamics during chronic infections, and enable studies that link strain changes to clinical outcomes.

Introduction

Serial culturing of chronic infection sites often repeatedly yields the same pathogen species. For instance, chronic wounds can consistently grow *Staphylococcus* and *Pseudomonas* species (1), patients with urinary tract anomalies can be persistently infected by *Escherichia coli* (2), and chronically infected sinuses can recurrently yield the same anaerobes (1–3). The chronic infections that people with cystic fibrosis (CF) have are a prime example, as the same pathogen species are frequently cultured from patients' lung secretions for long periods. Some species, like *Pseudomonas aeruginosa* (*P. aeruginosa*) and *Staphylococcus aureus* (*S. aureus*) can be highly abundant in the lungs of individual patients for decades or even for life (4–9).

Repeated detection of the same bacterial species over time can imply that a single strain or lineage is continually present. However, even though most strain-level genotyping studies examine very few isolates from each infection, studies on chronic wound, urinary tract, ear, gastrointestinal, and lung infections suggest more complexity. For example, strain-level genotyping methods have shown that close to a third of people with CF and *S. aureus* lung infections simultaneously harbor more than 1 *S. aureus* strain (4, 5, 10, 11). Likewise, some studies have shown that up to 40% of people with CF are simultaneously infected by 2 or more *P. aeruginosa* strains (12–14), although other work has suggested a lower frequency of multi-strain infections (7, 15–19). In addition, strain relative abundance can change over time, and strains can be gained or lost in individual patients (13, 18, 20). Notorious examples are *P. aeruginosa* epidemic strains that can infect and eventually become dominant in already colonized patients and worsen disease (21–23).

Identifying infecting strains is important for several reasons. First, strains of the same species can differ markedly in traits like the capacity for injury, transmissibility, and resistance to antibiotics (19, 22–27). Thus,

the presence of multiple strains or changes in strain relative abundance could have clinical consequences. Second, strain abundance changes could provide information about the status of host defenses, treatment efficacy, or pathogen functioning. For example, strains may recede when host defenses or treatments to which they are susceptible intensify or when deleterious mutations arise. Likewise, new strain acquisition could indicate that host conditions have become more permissive, and analysis of succeeding strains could increase understanding of bacterial functions important in vivo. Third, sensitive methods for strain detection could reveal outbreaks and lapses in infection control procedures. Finally, early detection of new strains could spur eradication attempts, which may be more successful soon after strains are acquired (28–32).

Established methods for strain-level identification, such as pulse-field gel electrophoresis (PFGE), multi-locus sequence typing (MLST), and whole-genome sequencing (WGS), must generally be performed on 1 cultured isolate at a time (16, 33–36). Because pathogen populations can be extremely large and colonies from different strains may look identical (37–41), analyzing a few colonies per sample could miss multi-strain infections and strain acquisition and loss events. Newer methods using amplification of species-specific variable regions (42, 43) are not easily adaptable to multiple pathogens, and shotgun sequencing of clinical samples (41, 44) can be limited if nontarget DNA (e.g., host or other bacterial DNA) is abundant. To address these limitations, we developed PopMLST (population MLST), a method to enumerate and measure the relative abundance of strains present in pools of hundreds of cultured isolates, or in DNA directly extracted from clinical samples.

Results

Overview. In conventional MLST, bacterial colonies are isolated in pure culture, Sanger sequencing is used to identify allelic variation in MLST loci within conserved housekeeping genes (7 loci in the case of *S. aureus* and *P. aeruginosa*), and loci allele types are determined by comparison to a database (45–47). Because a single clone is analyzed, the loci are known to be linked, in that they originate from the same bacterial isolate. Thus, loci allele identities can be combined to define the MLST type of a pure culture isolate.

In contrast, the goal of PopMLST is to enumerate the pathogen strains and measure strain relative abundance in samples that could contain multiple strains, even when the infection site contains a vast excess of nontarget (e.g., human) DNA. To achieve this, PopMLST uses PCR to amplify MLST loci from complex samples and next-generation sequencing to measure allele relative abundance. The PCR primers act as probes to find conserved sequences flanking MLST loci (even when the targeted species is rare) and as vectors to amplify the strain-discriminating MLST loci. Amplicons are Illumina sequenced, and bioinformatic tools are used to distinguish rare variants from errors, group "like" sequences, and measure their relative abundance (Figure 1). A drawback to this approach is that PCR amplification and Illumina sequencing from complex mixtures are more error prone than Sanger sequencing of individual clones. Errors could be confused for low-abundance variant strains, particularly since different MLST loci can differ only at a few positions (45–47).

We addressed this problem in several ways. First, we used high-fidelity polymerases and as few PCR cycles as possible to reduce errors and PCR chimeras. Second, we adapted the DADA2 analysis pipeline for use on MLST loci amplicons. DADA2 (48) is designed for 16S rRNA amplicon sequencing and uses statistical methods to distinguish sequencing errors from low-abundance variants (Figure 1B) (43, 48). Third, we developed bioinformatic methods to adaptively trim the lower quality ends of the reverse read generated by Illumina sequencing to facilitate accurate read merging (see Methods and Figure 1B). Fourth, we amplified each MLST locus in triplicate and pooled the data to reduce random, preferential amplification of templates (i.e., "jackpot" amplifications) (49). Fifth, we omitted a GC repeat–rich *P. aeruginosa* MLST loci (*aro*) that was challenging to sequence with Illumina chemistries (Supplemental Figure 1, A–C; supplemental material available online with this article; https://doi.org/10.1172/jci.insight.152472DS1) (50, 51). Control analyses (Supplemental Figure 1, D and E) showed that omission of *aro* had only a minor effect on strain discrimination. Together, these approaches mitigate, but do not fully eliminate, the effects of PCR and sequencing errors.

Data interpretation. Although the PCR and Illumina sequencing used in PopMLST enable analysis of mixtures containing multiple strains and excess nontarget DNA, information from MLST loci is unlinked because many isolates are analyzed en masse and allele sequences for each locus are derived from separate PCRs. This issue does not generally limit PopMLST's ability to enumerate and measure strain relative abundance, which can be determined by examining the loci with the highest number of alleles represented.

This approach is effective because even though strains sometimes share MLST alleles (and PopMLST will report the sum of the shared loci's relative abundance in these cases), the large number of alleles for each locus (e.g., *S. aureus* MLST loci have 484–892 distinct alleles, and *P. aeruginosa* MLST loci have 137–278 distinct
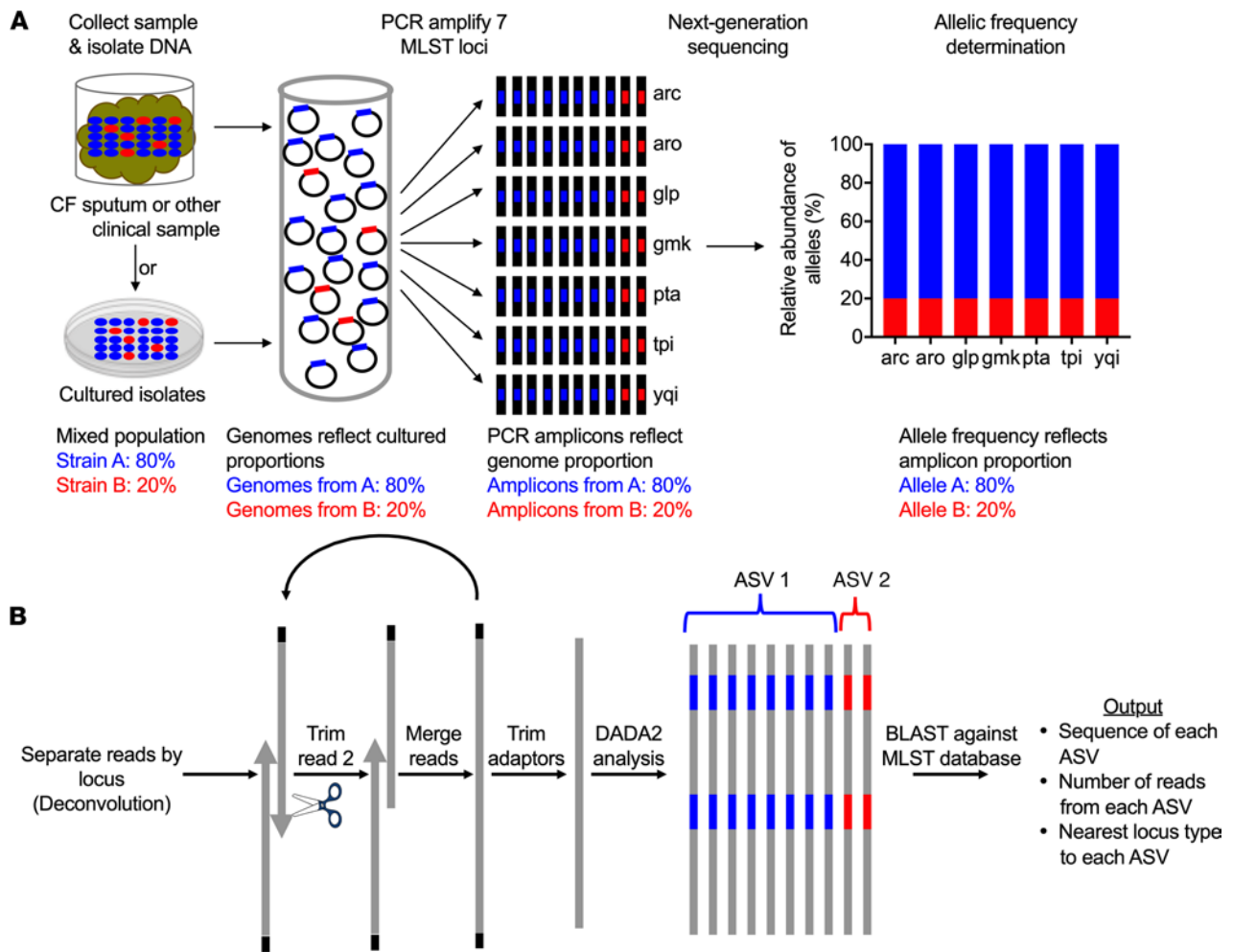
Figure 1. PopMLST methods. (**A**) PopMLST can be performed on clinical specimens (without culturing) or cultured isolates. MLST loci are PCR amplified in separate reactions, amplicons Illumina sequenced, and the relative abundance of reads representing MLST loci measured. (**B**) Bioinformatic analysis deconvolutes reads using permissive alignment to assign them to MLST loci, iteratively tests reverse read trimming lengths to optimize merging of paired-end reads, removes adaptors, identifies amplicon sequence variants (ASVs) using DADA2, and uses BLAST to identify the closest MLST locus type.

alleles) makes it unlikely that strains would have identical alleles at enough MLST loci to prevent strain enumeration. However, in mixed populations containing many strains, the likelihood that multiple strains share alleles increases, which could cause PopMLST to underestimate the number of strains present in the population.

The MLST types of strains within mixtures can also often be determined from PopMLST data. When a limited number of strains coexist, inference can determine which MLST alleles originate from the same strain, as linked alleles will be detected at a similar relative abundance. For example, if PopMLST finds that each loci contains 3 alleles at a relative abundance of ~70%:~25%:~5%, it is likely that the alleles identified at 70% relative abundance belong to one strain, alleles at 25% come from a second strain, and alleles at 5% come from a third strain. When many strains are present, if strains coexist at similar relative abundances, or if strains happen to share several alleles, inference can fail. If knowledge of the specific MLST types is important, conventional MLST can be performed on a few cultured colonies to determine which alleles are linked to one another to guide analysis of population-level data generated by PopMLST.

Once it is determined which MLST loci types likely belong to the same isolates (by inference, or by conventional MLST on single isolates), the relative abundance of each MLST type in the sample is calculated by averaging the relative abundance of all loci that differentiate the 2 strains. Averaging MLST loci abundances dampens the effect of error that could occur in individual loci measurements.

PopMLST identifies single strains after in vivo diversification. As an initial test of the method, we performed PopMLST on pure cultures containing single strains of *S. aureus* and *P. aeruginosa* and found that more than 99% of reads correctly reported a single MLST type in each of 21 independent experiments (Table 1).

Table 1. PopMLST correctly identifies single *S. aureus* and *P. aeruginosa* isolates as a single MLST type

Sample	Percentage of reads mapping to single loci type (SEM)
S. aureus	
NCTC8325	100% (0)
NCTC8325 (repeat)	100% (0)
MN8	100% (0)
MN8 (repeat)	100% (0)
Newman	100% (0)
CF isolate	100% (0)
CF isolate	100% (0)
CF isolate	100% (0)
CF isolate	100% (0)
CF isolate	100% (0)
P. aeruginosa	
PA14	100% (0)
PA14 (repeat)	99.96% (0.0272)[A]
LES	100% (0)
LES (repeat)	100% (0)
PAO1	100% (0)
PAO1 (repeat)	100% (0)
CF isolate	100% (0)
CF isolate	100% (0)
CF isolate	100% (0)
CF isolate	100% (0)
CF isolate	99.6% (0.006)[B]
CF isolate	100% (0)
Wound isolate	100% (0)
Wound isolate	100% (0)

[A]Three of the 6 loci indicated the presence of a second loci type at less than 1% likely due to sequencing error. [B]Four of the 6 loci indicated the presence of a second loci type at less than 1% likely due to sequencing error.

In CF and other chronic infections, strains genetically diversify during infection (10, 19, 25–27, 52, 53), and within-strain genetic diversity could be mistaken for strain differences. Thus, we tested PopMLST on pools of 90–96 clonally related *P. aeruginosa* isolates collected from different lung regions, from 3 people with CF undergoing lung transplantation. WGS showed that isolates from each patient were clonally related to one another but had genetically diversified via in vivo evolution (19). Importantly, 2 of the 3 collections exhibited hypermutator phenotypes due to mutations in either *mutL* or *mutS* mismatch repair genes, and WGS showed that the mutator populations contained far higher levels of genetic variation than the nonmutator population (19). Core genomes of 96 isolates from the patient who was not a hypermutator contained a total 328 SNP differences, and the 96 isolates from patients with hypermutator lineages contained 3169 and 1653 SNP differences (19).

Despite this extensive evolved diversity, PopMLST correctly identified each of the populations as containing a single MLST type (<0.01% of reads erroneously reported a second MLST allele) (Figure 2 and Supplemental Table 1). These data suggest that the measures used to mitigate PCR amplification and sequencing errors are effective for pure culture isolates and diversified clonally related populations.

PopMLST accurately measures pathogen strains in experimental mixtures. A key assumption of the PopMLST approach is that the relative abundance of MLST loci present in samples is maintained through DNA extraction, amplification, sequencing, and enumeration steps (Figure 1). We therefore began testing PopMLST's ability to detect multiple strains using defined mixtures of purified DNA from different strains. PopMLST identified the expected ratios (within 2-fold) of mixtures containing 2 *S. aureus* or *P. aeruginosa* strains over a wide relative abundance range (Figure 3). Replicate experiments using different sequencing runs and different MLST types produced similar results (Figure 3 and Supplemental Figures 2 and 3). Linear regression of data from the experimental mixtures indicated close agreement between expected results and average MLST allele loci measurements ($R^2 = 0.9916$ for *S. aureus* and $R^2 = 0.9901$ for *P. aeruginosa*), with slopes approximating 1 (*S. aureus*: 1.017 [95% CI: 0.9872–1.047]; *P. aeruginosa*: 0.9806 [95% CI: 0.9454–1.016]).
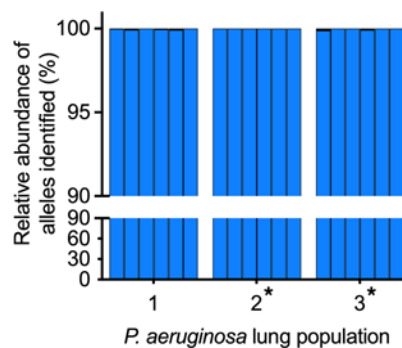
Figure 2. PopMLST correctly identifies genetically diversified, clonally related *P. aeruginosa* as a single MLST type. PopMLST was performed on pools of 90–96 clonally related *P. aeruginosa* isolates collected from different lung regions, from 3 CF patients undergoing lung transplantation. Plot shows the relative abundance of each MLST allele (from pool) that matches the known MLST sequence (determined by WGS; see Supplemental Table 1). The 6 bars for each sample show the relative abundance of *acs*, *gua*, *mut*, *nuo*, *pps*, and *trp* loci (in order). Black bars indicate any additional MLST loci types detected, which in all cases were less than 0.2%. Asterisk indicates hypermutable populations due to *mutS* (population 2) or *mutL* (population 3) mutations.

We also tested PopMLST's ability to measure strain relative abundance in 3- and 4-strain mixtures (Figure 4). PopMLST measurements of strain relative abundance in mixtures containing equal ratios of 3 *S. aureus* strains reported an average strain relative abundance of 33.7% (SEM 7.9%), and measurements of equal-ratio 4-strain *S. aureus* mixtures reported average strain relative abundance of 24.9% (SEM 5.57%) (Figure 4A). In equal-ratio *P. aeruginosa* 3-strain mixtures, PopMLST reported average strain relative abundance of 33.2% (SEM 9.5%). In equal-ratio 4-strain *P. aeruginosa* mixtures, PopMLST reported average strain relative abundance of 24.6% (SEM 6.0%) (Figure 4B). We have not tested mixtures containing more than 4 strains.

Despite use of triplicate PCR reactions, a single locus type was occasionally detected at higher-than-expected abundance (Figure 3 and Supplemental Figures 2 and 3). These findings are likely due to PCR bias (indicated by ‡ in Figure 3 and Supplemental Figures 2 and 3) or jackpot amplifications (indicated by # in Supplemental Figure 3). PCR bias and jackpots also occur in 16S rRNA gene measurements (54), which likewise use amplicon sequencing. However, because PopMLST averages data from 6 or 7 independently amplified loci (unlike 16S sequencing, which relies on a single locus), the effect of error in any given locus is dampened. Furthermore, loci that appear to be outliers can be interpreted in the context of others to estimate strain relative abundance. Because of these advantages, PCR bias and jackpot amplifications had little effect on strain relative abundance measurements in the experimental mixtures we tested (Figure 3, B and D, and Supplemental Figures 2 and 3).

PopMLST has a low frequency of false-positive strain calls. Errors inherent to PCR and Illumina sequencing could cause PopMLST to artifactually report strains that are not present. We examined control experiments containing between 1 and 4 strains of known composition (n = 38 for *P. aeruginosa* and n = 41 for *S. aureus*) to examine the effect of using different abundance thresholds to make strain presence and absence calls.

As shown in Table 2, using the criterion that a single variant locus be present at ≥1% relative abundance falsely registered the presence of a new strain in 9/49 (18%) of control experiments with *P. aeruginosa* and 7/41 (17%) of control experiments with *S. aureus*. Using the criterion that 2 or more loci be present at ≥1%, or raising the relative abundance threshold for a single allele to ≥4%, produced accurate calls in all 49 *P. aeruginosa* and all 41 *S. aureus* experiments. We conclude that using a threshold of a single variant locus at greater than 4% relative abundance or 2 variant loci at greater than 1% relative abundance results in a low likelihood of erroneously interpreting sequencing error as strain presence. Moreover, because the false-positive calls we detected tended to be sequencing run specific (Supplemental Table 2), their impact on PopMLST's accuracy could be decreased by repeated sequencing of samples.

PopMLST can detect specific MLST types with high sensitivity. In certain settings, clinicians and researchers need to detect specific strains with known MLST types. Examples include superinfections with virulent *P. aeruginosa* epidemic strains in people with CF already colonized by *P. aeruginosa* or infection control surveillance during outbreaks. Theoretically, known MLST types should be detectable with higher sensitivity than unknown types, as it is extremely unlikely that the chance occurrence of errors would report the presence of the specific MLST loci of interest.
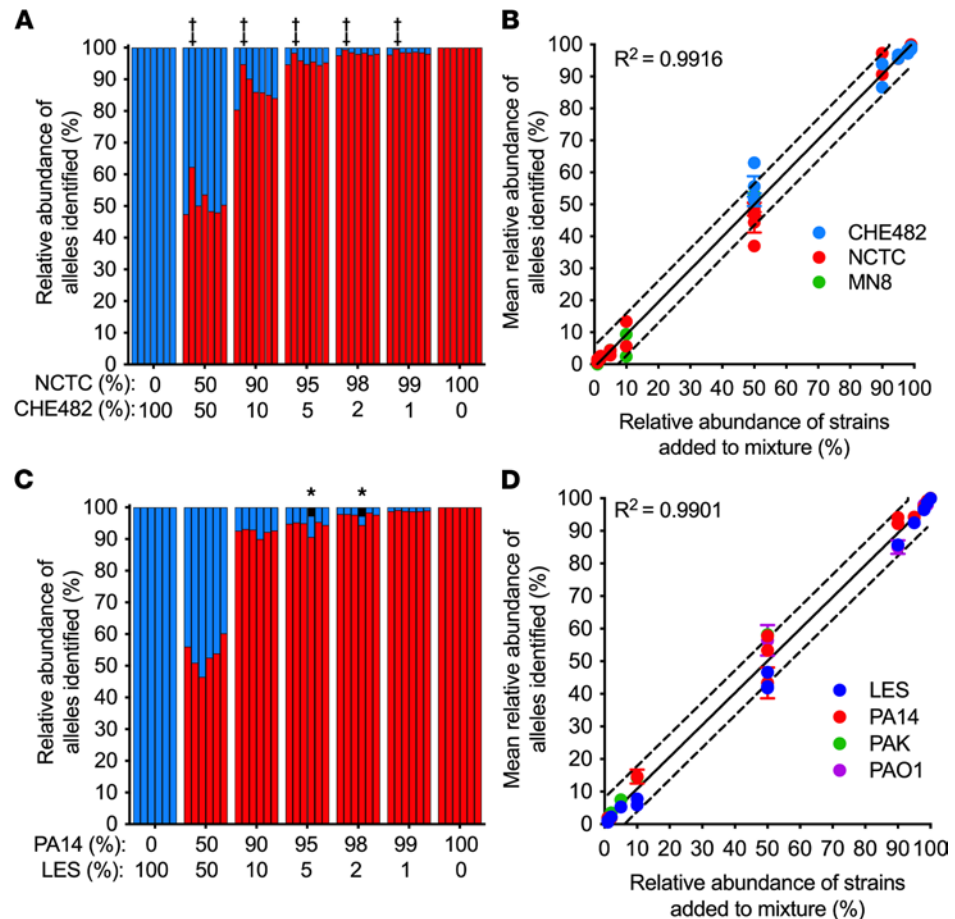
Figure 3. PopMLST measures strain relative abundance. (**A** and **C**) DNA from 2 *S. aureus* (Sa) (**A**) and *P. aeruginosa* (Pa) (**C**) strains with different MLST types were mixed at indicated ratios. Red bars indicate reads that PopMLST called as alleles from Sa NCTC8325 (**A**) and Pa PA14 (**C**); blue bars indicate reads called as alleles from Sa CHE482 (**A**) and Pa LES (**C**). (**B** and **D**) Mean loci relative abundance and SEM corresponding to the indicated MLST type from 21 independent 2-strain Sa mixtures (**B**), including NCTC8325 (red) with CHE482 (blue) or MN8 (green), and 20 independent 2-strain Pa mixtures (**D**), including PA14 (red) and LES (blue) or PAK (green), or PAO1 (purple) and LES (blue). Data for individual allele measurement can be found in Supplemental Figures 2 and 3. Some error bars (SEM) were smaller than symbols; solid line indicates expected result, dashed lines indicate ±10%. Bars in **A** show relative abundance of *arc*, *aro*, *glp*, *gmk*, *pta*, *tpi*, and *yqi* (in order). Bars in **C** show relative abundance of *acs*, *gua*, *mut*, *nuo*, *pps*, and *trp* (in order). MLST alleles identified but not present in the mixtures (likely sequencing error) are indicated in black, and those detected at more than 1% are indicated with *. ‡ indicates PCR bias as evidenced by 1 allele being consistently under- or overrepresented.

To test this, we measured PopMLST's sensitivity to detect targeted low-abundance MLST alleles in complex mixtures. As shown in Tables 3 and 4, targeted low-abundance alleles were detected in all experiments when present at 5% relative abundance or greater and in almost all experiments when present at 2%, 1%, and 0.1% relative abundance. These findings suggest that PopMLST could be used for early detection of known strains with high transmissibility or virulence or to investigate efficacy of infection control measures.

PopMLST works in the presence of excess human or nontarget bacterial DNA. Clinical samples can contain vast amounts of human and nontarget bacterial DNA. For example, despite high pathogen density (*P. aeruginosa* can reach 10^8–10^9 CFU/mL in CF sputum), 95%–99% of CF sputum DNA is human (55), and DNA from other pathogens or oral bacteria can also be highly abundant.

We investigated the effects of contaminating DNA on PopMLST in 2 ways. First, we performed PCR on human and nontarget bacterial DNA (including closely related species) using *S. aureus* and *P. aeruginosa* PopMLST primers. Amplicon yields and the number of reads mapping to *S. aureus* and *P. aeruginosa* MLST loci in these experiments were similar to no-template controls (Figure 5, A–D). Second, we tested the ability of PopMLST to detect strains in the presence of 95% human DNA, and we found that the vast excess of human DNA did not compromise detection, even when strains were present at as low as 1% relative abundance (Figure 5, E and F).
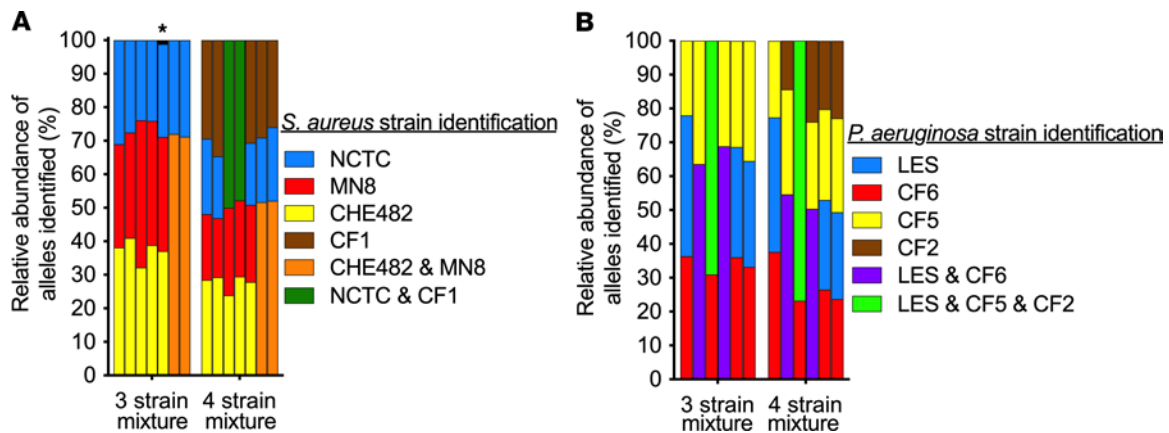
Figure 4. PopMLST can identify at least 4 unique MLST types in a mixture. Equimolar mixtures of DNA from 3 or 4 strains were analyzed by PopMLST. Unique MLST loci alleles of the 4 strains present in the mixtures are shown as blue, red, yellow, and brown. Alleles shared between strains added to the mixture cannot be assigned to a particular strain and are therefore colored as follows: alleles common to strains indicated with yellow and red are indicated with orange; alleles common to strains indicated with blue and brown are indicated with green (**A**); alleles common to strains indicated with red and blue are indicated with purple; and alleles common to strains indicated with blue and yellow and brown are indicated with green (**B**). Bars in **A** show relative abundance of Sa MLST alleles *arc*, *aro*, *glp*, *gmk*, *pta*, *tpi*, and *yqi* (in order). Bars in **B** show relative abundance of Pa MLST alleles *acs*, *gua*, *mut*, *nuo*, *pps*, and *trp* (in order). MLST alleles identified but not present in the mixtures (likely sequencing error) are indicated in black, and those detected at more than 1% are indicated with *.

PopMLST measures strain abundance in clinical samples. Encouraging results with experimental strain mixtures led us to perform proof-of-principle tests of PopMLST on clinical samples. In the first test, we cultured sputum from 7 *S. aureus*-infected CF patients and performed PopMLST on DNA prepared from about 100 colonies that grew from each sample (scraped en masse from culture plates). PopMLST reported that 3 of 7 samples contained 2 *S. aureus* MLST types (Figure 6A, Supplemental Figure 4, and Supplemental Table 3). We verified the presence of 2 strains in these samples by Sanger sequencing a distinguishing MLST locus in 20–30 individual colonies from each sample and found MLST types at a similar relative abundance as determined by PopMLST (R^2 = 0.9247; slope = 0.9296 [95% CI: 0.5612–1.298]) (Figure 6, A and B).

Second, we tested PopMLST's ability to resolve strains in CF sputum by mixing sputum samples that contained a single strain as determined by PopMLST (>99% of reads from each loci reporting a single allele) at varying ratios. When samples were combined, MLST types were identified in proportions close to the expected ratios (Figure 6C and Supplemental Table 1).

Finally, we analyzed sputum samples from 5 patients who were all known to harbor 2 *S. aureus* strains each. We performed PopMLST on DNA prepared directly from sputum and from several hundred *S. aureus* isolates from each sample scraped en masse from culture plates. As shown in Figure 6D, PopMLST detected the same MLST loci types in sputum DNA and culture scrapes from all 5 patients. Conventional MLST of representative isolates confirmed the dominant and secondary sequence types PopMLST detected (Figure 6D and Supplemental Table 3). In patients 11 and 12, strain relative abundance differed in sputum as compared with the culture scrapes (*P* < 0.001 by multiple *t* test), and in patient 9, a third MLST loci type was detected in sputum that was not present in ex vivo culture. These findings could be due to differential growth capacity of strains in ex vivo culture conditions.

Discussion

Infecting bacteria can exhibit heritable diversity at the species, strain, and intra-strain level. While recent findings and new methods have accelerated work on species diversity (56–64) and diversification within individual strains (10, 19, 25–27, 52, 53), studies of strain-level diversity have lagged. A major factor limiting progress is that established methods to detect strains must generally be performed on one cultured isolate at a time (16, 33–36). Thus, nondominant strains are difficult to detect.

PopMLST addresses this problem because it can estimate the relative abundance of strains in pools of tens to hundreds of isolates that have been cultured en masse from infected sites and can be used on DNA extracted directly from clinical specimens without prior culture. Other advantages include technical approaches to minimize PCR and sequencing error, its robustness when human or other bacterial DNA is in vast excess (Figure 5), its ability to detect targeted strains at low relative abundance (Tables 3 and 4), and its ability to detect

Table 2. Frequency of false-positive calls depending on criteria used to call MLST type presence

Abundance of loci used to call MLST type presence	P. aeruginosa		S. aureus	
	One locus used to call MLST type presence # experiments with false positive (%)	Two loci used to call MLST type presence # experiments with false positive (%)	One locus used to call MLST type presence # experiments with false positive (%)	Two loci used to call MLST type presence # experiments with false positive (%)
5%	0/38 (0%)	0/38 (0%)	0/41 (0%)	0/41 (0%)
4%	0/38 (0%)	0/38 (0%)	0/41 (0%)	0/41 (0%)
3%	0/38 (0%)	0/38 (0%)	2/41 (5%)	0/41 (0%)
2%	3/38 (8%)	0/38 (0%)	5/41 (12%)	0/41 (0%)
1%	4/38 (11%)	0/38 (0%)	7/41 (17%)	0/41 (0%)
0.50%	5/38 (13%)	1/38 (3%)	7/41 (17%)	0/41 (0%)
0.25%	6/38 (16%)	1/38 (3%)	8/41 (20%)	0/41 (0%)
0.10%	9/38 (23%)	2/38 (5%)	10/41 (24%)	2/41 (5%)
0.01%	14/38 (37%)	7/38 (18%)	10/41 (24%)	2/41 (5%)
0%	14/38 (37%)	7/38 (18%)	10/41 (24%)	2/41 (5%)

Accuracy is shown as a function of the abundance and number of loci used to make calls.

strains with ex vivo growth defects when the method is used on DNA prepared directly from clinical specimens. Furthermore, the method is accurate even when intra-strain genetic diversity has evolved (Figure 2 and Supplemental Table 1), likely because MLST loci are within conserved "housekeeping" genes that may be less variable than elements involved in Spa typing (6), Random Amplified Polymorphic DNA, or PFGE (16, 33). Finally, MLST databases are in widespread use and exist for over 100 bacterial species (47), so PopMLST can easily be adapted for use with many organisms with results being comparable between laboratories.

PopMLST also has limitations. First, several circumstances can limit PopMLST's ability to enumerate and identify strains. These include the presence of unrelated strains having the same MLST sequence type, *P. aeruginosa* strain mixtures in which *aro* is the only distinguishing locus (see Supplemental Figure 1), strain mixtures in which most MLST alleles are shared between strains, or cases where evolved diversity substantially alters MLST loci. However, these circumstances will be relatively rare because many MLST types have been identified for most pathogens. For example, about 3500 *P. aeruginosa* and about 5500 *S. aureus* MLST types have been described to date (47), so strain-distinguishing power is generally robust. Moreover, MLST loci are generally conserved, and our control experiments showed PopMLST accurately identified mutator strains that had genetically diversified in vivo (Figure 2).

Second, while PCR and Illumina sequencing enable the method to be used on complex mixtures containing multiple strains and abundant nontarget DNA, these techniques are subject to errors and biases. We reduce, but cannot entirely eliminate, the effect of these problems using replicate PCR, adaptive trimming, and statistical methods.

Third, PopMLST does not identify which MLST loci are linked in individual isolates. Although this limitation does not compromise strain enumeration and relative abundance measurements under most circumstances,

Table 3. PopMLST's sensitivity for detecting known *S. aureus* MLST types

% low-abundance MLST type	MLST loci						
	arc # times loci detected (range)	*aro* # times loci detected (range)	*glp* # times loci detected (range)	*gmk* # times loci detected (range)	*pta* # times loci detected (range)	*tpi* # times loci detected (range)	*yqi* # times loci detected (range)
10%	4/4 (19.5–2.8)	4/4 (6.0–1.8)	4/4 (10.6–1.9)	4/4 (14.0–2.8)	4/4 (14.1–2.3)	4/4 (14.9–2.4)	4/4 (16.0–2.3)
5%	4/4 (5.3–3.8)	4/4 (3.2–1.6)	4/4 (6.1–1.8)	4/4 (5.2–2.7)	4/4 (4.6–3.0)	4/4 (5.5–3.4)	4/4 (4.7–3.9)
2%	4/4 (2.4–1.7)	4/4 (1.2–0.6)	4/4 (3.4–0.7)	4/4 (11.1–2.0)	3/4 (1.7–0)	4/4 (2.3–1.4)	3/4 (1.9–0)
1%	3/4 (2.2–0)	4/4 (1.1–0.1)	3/4 (1.5–0)	4/4 (1.5–0.2)	2/4 (1.4–0)	3/4 (1.5–0)	3/4 (2.0–0)
0.1%	2/3 (0.4–0)	3/3 (0.3–0.1)	2/3 (0.3–0)	3/3 (1.5–0.2)	0/3 (0–0)	3/3 (0.3–0.2)	1/3 (0.2–0)

Each replicate experiment was performed 3 or 4 independent times. The number of replicate experiments analyzed is indicated by the denominator.

Table 4. PopMLST's sensitivity for detecting known *P. aeruginosa* MLST types

% low-abundance MLST type	MLST loci					
	acs # times loci detected (range)	*gua* # times loci detected (range)	*mut* # times loci detected (range)	*nuo* # times loci detected (range)	*pps* # times loci detected (range)	*trp* # times loci detected (range)
10%	4/4 (13.5–6.2)	3/3 (24.8–4.1)	4/4 (14.9–6.9)	3/3 (13.3–8.2)	4/4 (14.8–3.9)	4/4 (15.2–6.0)
5%	3/3 (7.0–4.6)	2/2 (6.6–4.8)	3/3 (7.7–2.4)	1/2 (6.6–1)	3/3 (7.6–4.6)	3/3 (7.7–2.8)
2%	3/3 (3.18–1.7)	2/2 (2.3–2.1)	3/3 (3.2–1.4)	2/2 (2.8–1.9)	3/3 (4.2–1.7)	3/3 (3.5–2.0)
1%	4/4 (1.6–0.7)	3/3 (4.0–0.4)	4/4 (2.1–0.7)	3/3 (1.6–1.1)	4/4 (2.7–0.5)	4/4 (1.8–0.7)
0.1%	3/3 (0.3–0.1)	2/3 (0.5–0)	3/3 (0.1–0.06)	1/3 (0.4–0)	3/3 (0.3–0.08)	3/3 (0.2–0.08)

Each replicate experiment was performed 3 or 4 independent times. The number of replicate experiments analyzed is indicated by the denominator. Two of the strains used for in 1 replicate experiment (for the 10%, 5%, 2%, and 1% low-abundance MLST type) shared alleles for the gua and nuo loci. Because the 2 strains could not be differentiated at these loci, they were eliminated from analysis for these loci, thereby decreasing the denominator only at the gua and nuo loci.

it can complicate identification of the strain types present in complex mixtures. Finally, we have not yet performed head-to-head tests comparing PopMLST's resolution with other available methods for strain enumeration.

Despite these limitations, we think that PopMLST could provide valuable strain-level information in several settings. One key area is in infection control, as PopMLST's ability to detect strains of interest at low relative abundance is advantageous, particularly if the strain being tracked belongs to a commonly encountered species. PopMLST performed serially on clinical samples could also detect episodes of strain gain or loss in individual patients. Strain changes could herald important variation in the host environment (e.g., changes in host defenses or treatments that select for new strains) or mutational changes in existing strains that compromise strain persistence in vivo. Additionally, PopMLST can detect strains with ex vivo growth defects.

PopMLST could also be used for early detection of superinfections with highly virulent or antibiotic-resistant strains. A notorious example is the Liverpool epidemic strain of *P. aeruginosa*, which can displace existing *P. aeruginosa* strains that infect the lungs of people with CF and cause increased morbidity and treatment resistance (65). Since the discovery of the Liverpool strain, other epidemic strains have been identified at disparate locations worldwide (24). Early detection of such strains is difficult in already colonized patients because clinical microbiology analyses only report the species present, and detecting epidemic strains when they are at low relative abundance would require tests on tens or hundreds of cultured isolates using conventional colony-by-colony assays (like MLST).

Finally, PopMLST could be used to investigate strain changes not associated with epidemic strains. Patients with chronic bacterial infections frequently experience highly variable disease manifestations, antibiotic responses, and rates of progression. The acquisition of different strains of a given species or changes in strain relative abundance could account for some of this variation (66). PopMLST will enable new hypotheses that explore the effects of strain-level diversity on human infection to be tested.

Methods

Patient samples. *S. aureus* was isolated after sputolysin-diluted sputum was cultured on Mannitol Salt Agar (Difco). Populations were scraped from plates containing more than 100 colonies by flooding the plate with 2 mL of LB and using an L-spreader to resuspend the bacteria. *P. aeruginosa* was isolated after sputolysin-diluted sputum was cultured on MacConkey (Difco). All cultures were stored at –80°C in 15% glycerol prior to analysis. Sputum samples analyzed directly were diluted with sputolysin as above and stored at –80°C until DNA isolation could be performed.

DNA isolation. *P. aeruginosa* DNA was isolated from 100 μL of resuspended culture using the DNeasy Blood and Tissue kit (QIAGEN) using the protocol for Gram-negative bacteria. Due to the difficulty of lysing Gram-positive bacteria, all *S. aureus* and sputum samples were extracted using methods with increased lysis efficacy. DNA extraction of *S. aureus* and sputum samples was performed using the DNEasy PowerSoil Pro Kit (QIAGEN) with the following modifications: samples were incubated with 2.9 mg lysozyme and 0.14 mg lysostaphin prior to lysis with 0.1 mm beads using a bead beater (Mini-Beadbeater-16; Biospec), a method validated previously by our laboratory for efficient *P. aeruginosa* and *S. aureus* lysis from sputum (41), or using the DNeasy PowerSoil Pro Kit for the Qiacube automated extraction system, which results in comparable lysis of
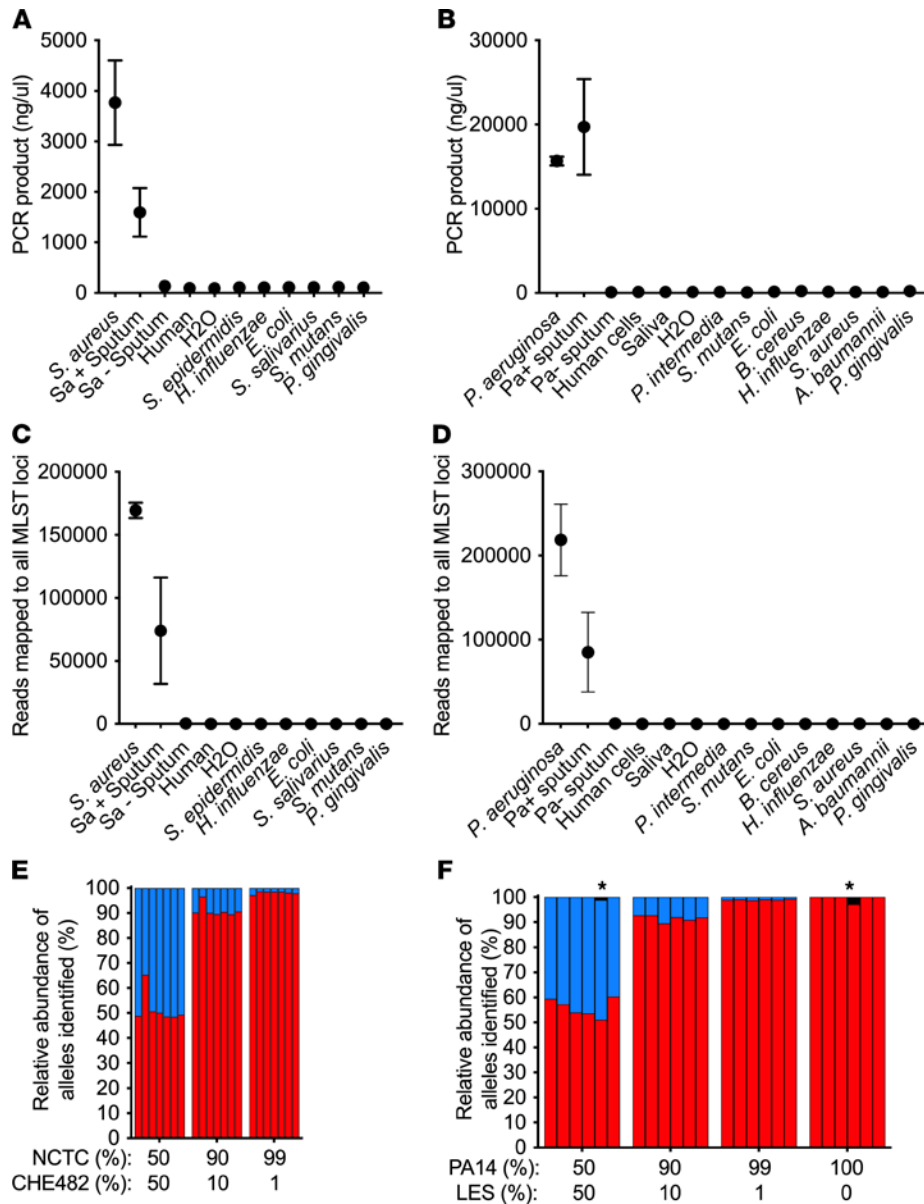
Figure 5. Heterologous DNA does not interfere with PopMLST. (**A** and **B**) The average concentration of PCR product from the 7 or 6 amplified loci after PCR using PopMLST primers for Sa (**A**) and Pa (**B**) on DNA from the indicated sources. "Human" indicates DNA extracted from tissue culture cells; "H2O" indicates ultrapure water; "Sa+ sputum" and "Pa+ sputum" indicate sputum from 3 patients with CF who were culture positive for Sa and Pa, respectively; and "Sa- sputum" and "Pa- sputum" indicate sputum from a CF patient who was culture negative for Sa and Pa. (**C** and **D**) The sum of sequence reads produced by PopMLST that mapped to 7 Sa (**C**) or 6 Pa (**D**) MLST loci shown for samples containing target and nontarget DNA from PCR reactions in **A** and **B**. Mixtures of Sa NCTC8325/CHE482 were used as positive control in **A** and **C**, and mixtures of PA14/PAO1 were used in **B** and **D**. For samples with negligible PCR amplification, more than 2 times volume of sample was used for Illumina sequencing than was used for other samples. The average and SEM of 3 separate samples are shown for Sa, Sa+ sputum, Pa, and Pa+ sputum in **A**–**D**. (**E** and **F**) 95% human DNA from tissue culture cells was added to the same mixtures of 2 control strains from Figure 2C. Bars in **E** show relative abundance of *arc*, *aro*, *glp*, *gmk*, *pta*, *tpi*, and *yqi* matching the MLST type of NCTC8325 (red) or CHE482 (blue). Bars in **F** show relative abundance of *acs*, *gua*, *mut*, *nuo*, *pps*, and *trp* (in order) matching the MLST type of PA14 (red) or LES (blue). Asterisk indicates the presence of an unexpected loci type (black), likely due to sequencing error.

S. aureus and *P. aeruginosa* from sputum in our laboratory. The lysis method used depended on the availability of the extraction method at the time of sample collection. All samples were eluted in Buffer EB (QIAGEN) regardless of extraction method.

Control mixtures. Control strains (Supplemental Tables 4 and 5) were streaked from freezer stocks and grown overnight prior to DNA isolation. *S. aureus* and *P. aeruginosa* strains were from the Singh lab as well as donated by Matthew Parsek (University of Washington, Seattle, Washington, USA) and Lucas Hoffman (University of Washington, Seattle, Washington, USA). Cultured HELA or HEK293 cells (donated by Joseph Mougous,
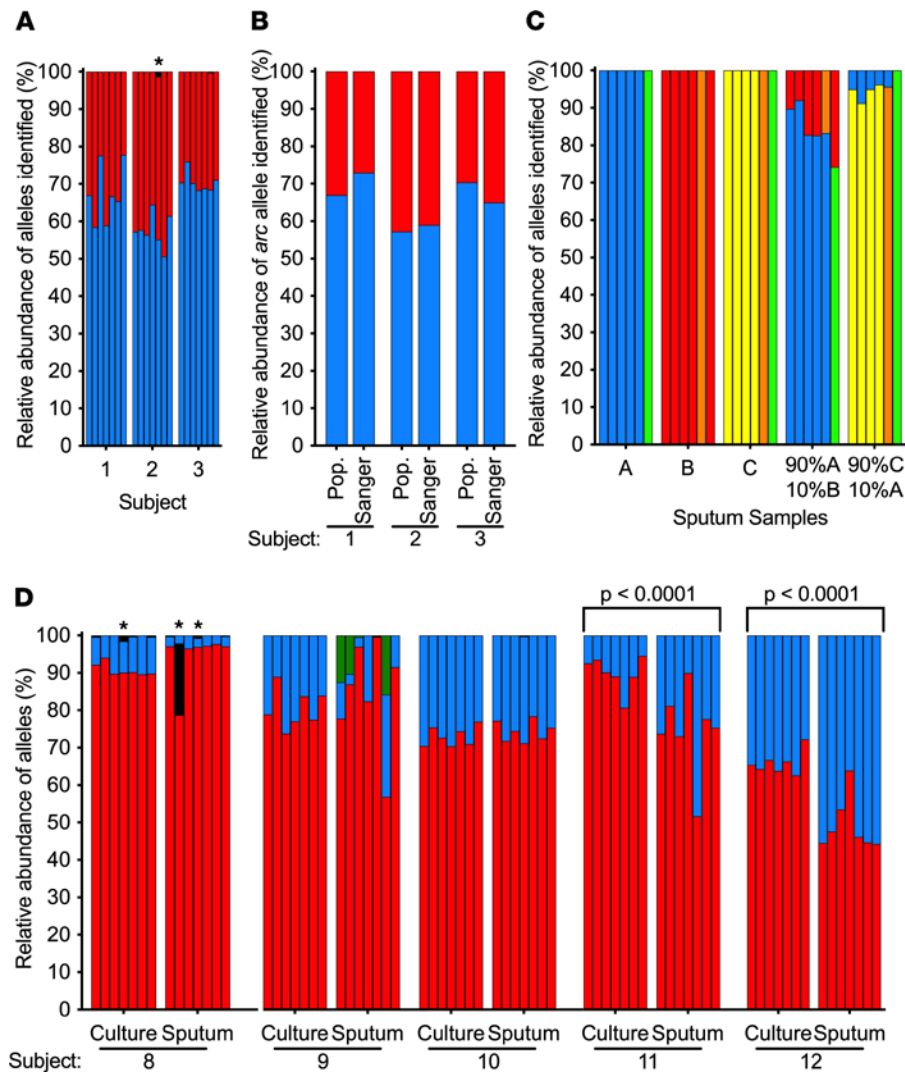
Figure 6. PopMLST on clinical samples. (A and **B)** PopMLST results from 3 of 7 patients with CF in whom PopMLST detected 2 Sa strains (see Supplemental Figure 4 for the 4 of 7 patients in whom 1 strain was detected). (**A**) PopMLST results from DNA pooled from 90–95 cultured Sa isolates. Blue and red bars indicate different MLST loci alleles for each patient (Supplemental Table 3). Bars for each sample show relative abundance of *arc*, *aro*, *glp*, *gmk*, *pta*, *tpi*, and *yqi*. A third *pta* allele (black bar and indicated with *) differed at a single nucleotide from the allele indicated in red, likely representing sequencing error or mutation. (**B**) Relative abundance of *arc* locus as measured by PopMLST (Pop) and by individually Sanger sequencing (Sanger) 20–30 isolates from each sample from **A**. (**C**) PopMLST performed directly on DNA isolated from 3 sputum samples, A–C, and from mixtures of these samples. Red, blue, and yellow bars indicate the abundance of MLST alleles corresponding to samples A–C, respectively; green bars indicate an allele shared between samples A and C; and orange bars indicate an allele shared between samples B and C (see Supplemental Table 1). Control experiments examining more than 100 Pa isolates cultured from sputum samples A–C showed each contained a single Pa MLST type. (**D**) PopMLST of DNA pooled from more than 100 Sa cultured colonies (culture) and directly from sputum (sputum). Red and blue indicate different MLST types, which were confirmed by Sanger sequencing individual isolates (Supplemental Table 1). Subject 9's sputum contained 3 loci with an additional allele, likely indicating a third MLST type (green) that was not detected in culture. Indicated significant differences were determined by the multiple *t* test of minor allele abundance. Black bar indicated with * indicates the presence of a third allele, likely due to mutation or sequencing error.

University of Washington, Seattle, Washington, USA) were pelleted and DNA was isolated as above. Isolated DNA was quantified by Qubit and mixed at ratios described in the figures. MLST types of control strains were based on the MLST database and confirmed by MLST typing of single isolates if necessary (Supplemental Tables 4 and 5). PAO1-lacZ:PA14 mixtures were premixed at designated ratios and plated on LB+xGal to confirm the ratio. Growth from the plate was scraped and subjected to DNA isolation as above.

PopMLST amplification and sequencing methods. A total of 5 ng/μL DNA from cultured bacteria, 20 ng/μL of sputum DNA, or 20 ng/μL of bacterial DNA mixed with human DNA was amplified by PCR using published MLST primers for *S. aureus* and *P. aeruginosa* (45, 46) with Illumina adaptors on the 5′ ends to enable next-generation sequencing of MLST loci (Supplemental Table 6). PCR amplification of each of the 7 MLST loci

was performed in triplicate to reduce chances of random PCR bias using reagents listed in Supplemental Table 7. Triplicate reactions were pooled after PCR, and amplified DNA was visualized by agarose gel electrophoresis and quantified by Pico green (Thermo Fisher Scientific). After cleaning with Ampure beads (Beckman Coulter), the 7 MLST loci for each sample were pooled in equimolar amounts and barcoded with Illumina Nextera XT indexes. PCR amplification, indexing, and cleanup was performed as described in the 16S Metagenomic Sequencing Library Preparation guide (Illumina). Indexed MLST loci were cleaned with Ampure beads (Beckman Coulter) prior to pooling and sequenced on the Illumina MiSeq to produce 2 × 300 bp paired-end reads.

Bioinformatic analysis. Methods outlined below for PopMLST are available at https://github.com/marade/PopMLST (commit ID 0de7f83). Because DADA2 is designed for 1 amplicon locus (typically 16S) at a time, reads were deconvolved based on their locus-specific primer sequence using Python tre, with approximate matching to MLST loci allowing for up to a 25% mismatch (https://github.com/laurikari/tre/; commit ID 6092368) before analysis. Due to a large number of paired-end reads failing to merge for some loci, we developed a dynamic read-trimming method using a binary search algorithm, which iteratively trims the 3′ end of the reverse read and retests merging until it maximizes the number of merged reads of the correct size. Trimmed reads for all MLST loci, except *yqi*, were merged using VSEARCH 2.13.4 fastq_mergepairs. Two base pairs of the sequence in the *yqi* locus beyond the 3′ ends of reads 1 and 2 (due to the length of this amplicon) were artificially supplied (these bases are conserved according to the MLST database; ref. 47). *yqi* reads were joined using VSEARCH 2.13.4 fastq_join. Merged reads, with their adaptors trimmed using Cutadapt 2.3, were then processed using the remaining standard DADA2 pipeline steps to generate ASVs for each locus.

To determine the identity and quantify the relative abundance of each MLST locus, the ASVs were queried against the PubMLST database (https://pubmlst.org/saureus/ and https://pubmlst.org/paeruginosa/) (47) for the appropriate species using BLAST+ BLASTN (67). The matching sequence with the highest identity and longest length (less than or equal to the maximum locus length present in the database) was used to label each ASV by locus type, with less than 100% identity matches being marked as potentially novel alleles. The resulting output table includes each MLST loci type identified, the ASV, and the number of reads assigned to each type, much like a classic 16S operational taxonomic unit table.

Statistics. The abundance of MLST types was reported as the average relative abundance of 7 corresponding *S. aureus* MLST loci or 6 corresponding *P. aeruginosa* MLST loci, and standard error of the mean was reported. When strain type(s) were known (Figures 2 and 3 and Table 1), the relative abundances of the alleles matching that strain type were used. When strain types were unknown, alleles with similar relative abundances were assumed to be from the same strain. Linear regression of the average relative abundance was used to determine the accuracy of measuring strain abundance (Figure 3, B and D). R^2 values were reported in text. Multiple *t* tests were used to compare relative abundance of individual loci from sputum with those from bacterial culture (Figure 6D). *P* values more than 0.05 were determined to be not significant and were not reported. All statistical analyses were performed in GraphPad Prism.

Study approval. Sputum samples were collected in accordance with the University of Washington Institutional Review Board (approved protocol numbers 06-4469 and STUDY00011983) and the Research Ethics Committee at St. Vincent's University Hospital, Dublin, Ireland (RS20-048). Patients provided written informed consent prior to collection of samples.

Author contributions

SJM, SLD, and PKS conceived the study; SJM, SLD, and MCR designed the methodology; MCR developed the software; SJM, SLD, and WN validated data; PKS, SJM, SLD, SR, and DTS investigated; PKS, MLA, SJS, PJ, and EFM provided resources; SJM, SLD, and PKS wrote the manuscript; PKS supervised; and PKS acquired funding. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

Address correspondence to: Pradeep Singh, UW Box #357735, University of Washington, Seattle, Washington 98195, USA. Phone: 206.221.7151; Email: singhpr@uw.edu.

1. Gjodsbol K, et al. Multiple bacterial species reside in chronic wounds: a longitudinal study. *Int Wound J*. 2006;3(3):225–231.

2. McGeachie J. Recurrent infection of the urinary tract: reinfection or recrudescence? *Br Med J*. 1966;1(5493):952–954.

3. Frederick JB, Braude AI. Anaerobic infection of the paranasal sinuses. *N Engl J Med*. 1974;290(3):135–137.

4. Branger C, et al. Persistence of Staphylococcus aureus strains among cystic fibrosis patients over extended periods of time. *J Med Microbiol*. 1996;45(4):294–301.

5. Kahl BC, et al. Population dynamics of persistent Staphylococcus aureus isolated from the airways of cystic fibrosis patients during a 6-year prospective study. *J Clin Microbiol*. 2003;41(9):4424–4427.

6. Hirschhausen N, et al. Extended Staphylococcus aureus persistence in cystic fibrosis is associated with bacterial adaptation. *Int J Med Microbiol*. 2013;303(8):685–692.

7. Struelens MJ, et al. Genome macrorestriction analysis of diversity and variability of Pseudomonas aeruginosa strains infecting cystic fibrosis patients. *J Clin Microbiol*. 1993;31(9):2320–2326.

8. Burns JL, et al. Longitudinal assessment of Pseudomonas aeruginosa in young children with cystic fibrosis. *J Infect Dis*. 2001;183(3):444–452.

9. Smith EE, et al. Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. *Proc Natl Acad Sci U S A*. 2006;103(22):8487–8492.

10. Goerke C, et al. High phenotypic diversity in infecting but not in colonizing Staphylococcus aureus populations. *Environ Microbiol*. 2007;9(12):3134–3142.

11. Long DR, et al. Polyclonality, shared strains, and convergent evolution in chronic cystic fibrosis Staphylococcus aureus airway infection. *Am J Respir Crit Care Med*. 2021;203(9):1127–1137.

12. Jelsbak L, et al. Molecular epidemiology and dynamics of Pseudomonas aeruginosa populations in lungs of cystic fibrosis patients. *Infect Immun*. 2007;75(5):2214–2224.

13. Jain M, et al. Type III secretion phenotypes of Pseudomonas aeruginosa strains change during infection of individuals with cystic fibrosis. *J Clin Microbiol*. 2004;42(11):5229–5237.

14. Anthony M, et al. Genetic analysis of Pseudomonas aeruginosa isolates from the sputa of Australian adult cystic fibrosis patients. *J Clin Microbiol*. 2002;40(8):2772–2778.

15. Syrmis MW, et al. Rapid genotyping of Pseudomonas aeruginosa isolates harboured by adult and paediatric patients with cystic fibrosis using repetitive-element-based PCR assays. *J Med Microbiol*. 2004;53(pt 11):1089–1096.

16. Mahenthiralingam E, et al. Random amplified polymorphic DNA typing of Pseudomonas aeruginosa isolates recovered from patients with cystic fibrosis. *J Clin Microbiol*. 1996;34(5):1129–1135.

17. Oliver A, et al. High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. *Science*. 2000;288(5469):1251–1254.

18. Hogardt M, et al. Stage-specific adaptation of hypermutable Pseudomonas aeruginosa isolates during chronic pulmonary infection in patients with cystic fibrosis. *J Infect Dis*. 2007;195(1):70–80.

19. Jorth P, et al. Regional isolation drives bacterial diversification within cystic fibrosis lungs. *Cell Host Microbe*. 2015;18(3):307–319.

20. Langhanki L, et al. In vivo competition and horizontal gene transfer among distinct Staphylococcus aureus lineages as major drivers for adaptational changes during long-term persistence in humans. *BMC Microbiol*. 2018;18(1):152.

21. McCallum SJ, et al. Superinfection with a transmissible strain of Pseudomonas aeruginosa in adults with cystic fibrosis chronically colonised by P aeruginosa. *Lancet*. 2001;358(9281):558–560.

22. Salunkhe P, et al. A cystic fibrosis epidemic strain of Pseudomonas aeruginosa displays enhanced virulence and antimicrobial resistance. *J Bacteriol*. 2005;187(14):4908–4920.

23. Parkins MD, et al. Twenty-five-year outbreak of Pseudomonas aeruginosa infecting individuals with cystic fibrosis: identification of the prairie epidemic strain. *J Clin Microbiol*. 2014;52(4):1127–1135.

24. Duong J, et al. Phenotypic and genotypic comparison of epidemic and non-epidemic strains of Pseudomonas aeruginosa from individuals with cystic fibrosis. *PLoS One*. 2015;10(11):e0143466.

25. Yagci S, et al. Prevalence and genetic diversity of Staphylococcus aureus small-colony variants in cystic fibrosis patients. *Clin Microbiol Infect*. 2013;19(1):77–84.

26. Lieberman TD, et al. Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nat Genet*. 2014;46(1):82–87.

27. Williams D, et al. Divergent, coexisting Pseudomonas aeruginosa lineages in chronic cystic fibrosis lung infections. *Am J Respir Crit Care Med*. 2015;191(7):775–785.

28. Allendorf FW, Lundquist LL. Introduction: population biology, evolution, and control of invasive species. *Conserv Biol*. 2003;17(1):24–30.

29. Frederiksen B, et al. Antibiotic treatment of initial colonization with Pseudomonas aeruginosa postpones chronic infection and prevents deterioration of pulmonary function in cystic fibrosis. *Pediatr Pulmonol*. 1997;23(5):330–335.

30. Hansen CR, et al. Early aggressive eradication therapy for intermittent Pseudomonas aeruginosa airway colonization in cystic fibrosis patients: 15 years experience. *J Cyst Fibros*. 2008;7(6):523–530.

31. Ratjen F, et al. Effect of inhaled tobramycin on early Pseudomonas aeruginosa colonisation in patients with cystic fibrosis. *Lancet*. 2001;358(9286):983–984.

32. Treggiari MM, et al. Approach to eradication of initial Pseudomonas aeruginosa infection in children with cystic fibrosis. *Pediatr Pulmonol*. 2007;42(9):751–756.

33. Tenover FC, et al. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol*. 1995;33(9):2233–2239.

34. van Belkum A, et al. Co-evolutionary aspects of human colonisation and infection by Staphylococcus aureus. *Infect Genet Evol*. 2009;9(1):32–47.

35. Kidd TJ, et al. Comparison of three molecular techniques for typing Pseudomonas aeruginosa isolates in sputum samples from patients with cystic fibrosis. *J Clin Microbiol*. 2011;49(1):263–268.

36. Waters V, et al. Comparison of three typing methods for Pseudomonas aeruginosa isolates from patients with cystic fibrosis. *Eur J Clin Microbiol Infect Dis*. 2012;31(12):3341–3350.

37. Gentili V, et al. Panbacterial real-time PCR to evaluate bacterial burden in chronic wounds treated with Cutimed™ Sorbact™. *Eur J Clin Microbiol Infect Dis*. 2012;31(7):1523–1529.

38. Yusuf E, et al. High bacterial load in negative pressure wound therapy (NPWT) foams used in the treatment of chronic wounds. *Wound Repair Regen*. 2013;21(5):677–681.

39. Rudkjøbing VB, et al. An exploratory study of microbial diversity in sinus infections of cystic fibrosis patients by molecular methods. *J Cyst Fibros*. 2014;13(6):645–652.

40. Fischer AJ, et al. Sustained coinfections with Staphylococcus aureus and Pseudomonas aeruginosa in cystic fibrosis. *Am J Respir Crit Care Med*. 2021;203(3):328–338.

41. Hisert KB, et al. Restoring cystic fibrosis transmembrane conductance regulator function reduces airway bacteria and inflammation in people with cystic fibrosis and chronic lung infections. *Am J Respir Crit Care Med*. 2017;195(12):1617–1628.

42. Shevchenko SG, et al. Escherichia coli clonobiome: assessing the strain diversity in feces and urine by deep amplicon sequencing. *Appl Environ Microbiol*. 2019;85(23):e01866–19.

43. Rosen MJ, et al. Microbial diversity. Fine-scale diversity and extensive recombination in a quasisexual bacterial population occupying a broad niche. *Science*. 2015;348(6238):1019–1023.

44. Gan GL, et al. Deconvoluting the diversity of within-host pathogen strains in a multi-locus sequence typing framework. *BMC Bioinformatics*. 2019;20(suppl 20):637.

45. Curran B, et al. Development of a multilocus sequence typing scheme for the opportunistic pathogen Pseudomonas aeruginosa. *J Clin Microbiol*. 2004;42(12):5644–5649.

46. Enright MC, et al. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of Staphylococcus aureus. *J Clin Microbiol*. 2000;38(3):1008–1015.

47. Jolley KA, et al. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res*. 2018;3:124.

48. Callahan BJ, et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13(7):581–583.

49. Polz MF, Cavanaugh CM. Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol*. 1998;64(10):3724–3730.

50. Nakamura K, et al. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res*. 2011;39(13):e90.

51. Schirmer M, et al. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics*. 2016;17:125.

52. McAdam PR, et al. Adaptive evolution of Staphylococcus aureus during chronic endobronchial infection of a cystic fibrosis patient. *PLoS One*. 2011;6(9):e24301.

53. Morgan S, et al. Bacterial fitness in chronic wounds appear to be mediated by the capacity for high-density growth, not virulence or biofilm functions. *PLoS Pathog*. 2019;15(3):e1007511.

54. Wen C, et al. Evaluation of the reproducibility of amplicon sequencing with Illumina MiSeq platform. *PLoS One*. 2017;12(4):e0176716.

55. Feigelman R, et al. Sputum DNA sequencing in cystic fibrosis: non-invasive access to the lung microbiome and to pathogen details. *Microbiome*. 2017;5(1):20.

56. Wolcott RD, et al. Analysis of the chronic wound microbiota of 2,963 patients by 16S rDNA pyrosequencing. *Wound Repair Regen*. 2016;24(1):163–174.

57. Loesche M, et al. Temporal stability in chronic wound microbiota is associated with poor healing. *J Invest Dermatol*. 2017;137(1):237–244.

58. Price LB, et al. Community analysis of chronic wound bacteria using 16S rRNA gene-based pyrosequencing: impact of diabetes and antibiotics on chronic wound microbiota. *PLoS One*. 2009;4(7):e6462.

59. Stokell JR, et al. Analysis of changes in diversity and abundance of the microbial community in a cystic fibrosis patient over a multiyear period. *J Clin Microbiol*. 2015;53(1):237–247.

60. Zhao J, et al. Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc Natl Acad Sci U S A*. 2012;109(15):5809–5814.

61. Klepac-Ceraj V, et al. Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and Pseudomonas aeruginosa. *Environ Microbiol*. 2010;12(5):1293–1303.

62. Cox MJ, et al. Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. *PLoS One*. 2010;5(6):e11044.

63. Coburn B, et al. Lung microbiota across age and disease stage in cystic fibrosis. *Sci Rep*. 2015;5:10241.

64. Whiteside SA, et al. The microbiome of the urinary tract--a role beyond infection. *Nat Rev Urol*. 2015;12(2):81–90.

65. Al-Aloul M, et al. Increased morbidity associated with chronic infection by an epidemic Pseudomonas aeruginosa strain in CF patients. *Thorax*. 2004;59(4):334–336.

66. McCallum SJ, et al. Spread of an epidemic Pseudomonas aeruginosa strain from a patient with cystic fibrosis (CF) to non-CF relatives. *Thorax*. 2002;57(6):559–560.

67. Camacho C, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10(1):421.