ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Optical Mapping: Detecting Genomic Resistance Cassettes in MRSA

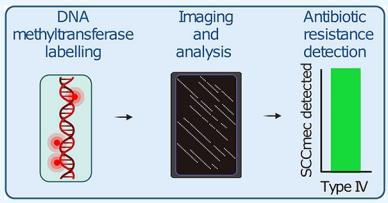
Elizabete Ruppeka-Rupeika, Sergey Abakumov, Mattias Engelbrecht, Xiong Chen, Debora do Carmo Linhares, Arno Bouwens, Volker Leen, and Johan Hofkens*

Cite This: *ACS Omega* 2024, 9, 8862−8873

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Methicillin-resistant *Staphylococcus aureus* (MRSA) is a multidrug-resistant bacterium with a global presence in healthcare facilities as well as community settings. The resistance of MRSA to beta-lactam antibiotics can be attributed to a mobile genetic element called the staphylococcal cassette chromosome *mec* (SCC*mec*), ranging from 23 to 68 kilobase pairs in length. The mec gene complex contained in SCC*mec* allows MRSA to survive in the presence of penicillin and other beta-lactam antibiotics. We demonstrate that optical mapping (OM) is able to identify the bacterium as *S. aureus*, followed by an investigation of the presence of kilobase pair range SCC*mec* elements by examining the associated OM-generated barcode patterns. By employing OM as an alternative to traditional DNA sequencing, we showcase its potential for the detection of complex genetic elements such as SCC*mec* in MRSA. This approach holds promise for enhancing our understanding of antibiotic resistance mechanisms and facilitating the development of targeted interventions against MRSA infections.



INTRODUCTION

Staphylococcus aureus, a Gram-positive bacterium, typically resides in the human body as a commensal organism but retains its capacity for transitioning into an opportunistic pathogen, provided the necessary conditions are met. *S. aureus* is commonly found in the anterior nares or nostrils, with approximately 20−30% of healthy individuals being persistent nasal carriers and at least 30% carrying *S. aureus* intermittently.[1−3] Due to its ease of transmission and ability to survive on many types of surfaces, *S. aureus* can readily cause infections upon contact with minor skin wounds, leading to a high potential for infection.

Of particular concern are infections with *S. aureus* strains that have acquired resistance to antibiotics, with considerable risk for lethal outcomes, especially in healthcare and healthcare-associated settings. A prime example is methicillin-resistant *Staphylococcus aureus* (MRSA), a group of *S. aureus* strains that exhibit resistance to methicillin, a beta-lactam antibiotic structurally similar to penicillin, with resistance capacity extending into a range of beta-lactam antibiotics. In the worst-case scenarios, MRSA infections progress to bacteremia and sepsis. In the absence of effective treatment options, the outcomes are likely fatal.[4] The global spread of MRSA contributes to the broader threat posed by antimicrobial superbugs, which are bacterial strains resistant to multiple antibiotics.[5]

Although antibiotics are generally effective in treating *S. aureus* infections, this bacterium is prone to acquiring mutations and horizontally transferred mobile genetic elements (MGEs) that confer antibiotic resistance. The first methicillin-resistant strains of *S. aureus* were identified as early as 1961 in a hospital setting in the United Kingdom, a mere 2 years after the introduction of methicillin to combat penicillin-resistant *S. aureus*.[6] One notable MGE is the Staphylococcal Cassette Chromosome *mec* (SCC*mec*), which is located on the chromosome and can be horizontally shared among strains.[7] SCC*mec* serves as a platform for incorporating resistance genes and other supporting gene complexes, including instances of genes that have jumped across species, such as Vancomycin resistance most likely acquired from *Enterococcus*.[8]

There are at least 13 types of SCC*mec*, with I−VI being the most common.[9] Many of these cassettes are well described in the literature in three loosely overlapping categories—healthcare-associated (HA)(for example, cassettes I−III), community-associated (CA) (cassette IV), and livestock-associated SCC*mec* types.[10−16] The SCC*mec* I−IV are 23−68 kbp in length and, as expected, exhibit overlaps at the *mec* and cassette chromosome recombinase (*ccr*) gene complex regions, described in detail by Hiramatsu and colleagues (2013).[17] The ccr genes contribute to excision and site-specific integration of SCCmec into the *S. aureus* genome. The gene that allows for drug resistance is *mecA*, encoding penicillin-binding protein 2a (PBP2a). PBP2a is a peptidoglycan transpeptidase able to catalyze cell wall biosynthesis in the presence of penicillin and other β-lactam antibiotics, thanks to its low affinity for these compounds.[18] The active transcription of *mecA* is controlled by

Received: August 26, 2023
Revised: January 15, 2024
Accepted: January 22, 2024
Published: February 13, 2024

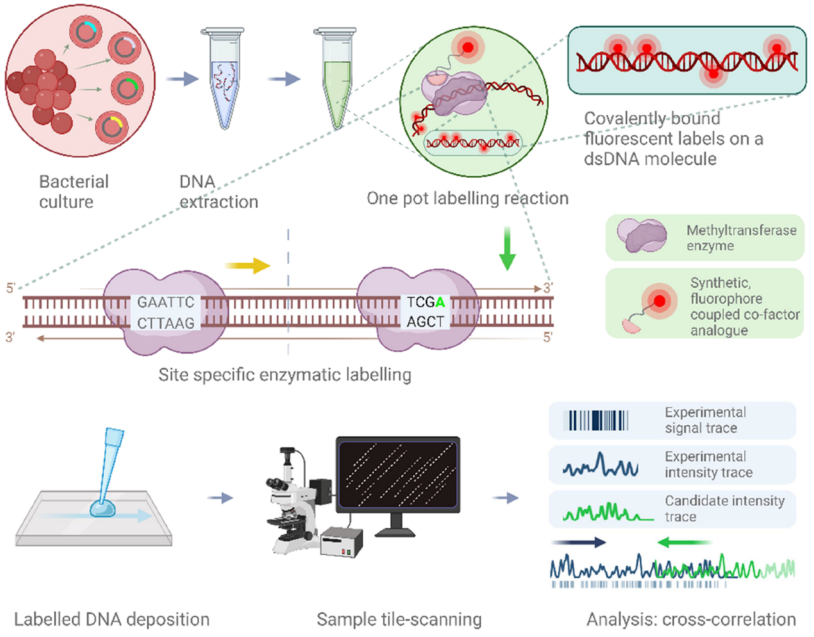


Figure 1. Fluorocode workflow. Above, from left: bacteria carrying different SCC*mec* cassettes. First, high molecular weight (HMW) DNA extraction is performed. Then, a one-pot labeling reaction follows, where the M.Taq methyltransferase delivers a fluorophore to its specific recognition site. The labeling reaction results in covalently bound labels at the M.Taq recognition sites on the double-stranded DNA molecule. Below, from left: labeled DNA is deposited using the rolling droplet technique, and the linearized DNA is tile-scanned using a widefield fluorescence microscope. Subsequently, the tile-scans are segmented to select the DNA traces, which are recorded as a signal barcode. An intensity trace is generated from the barcode and is cross-correlated against an in silico-generated candidate intensity trace from a database entry.

other genes in the *mec* gene complex, which is sensitive to the presence of β-lactam antibiotics, which will disrupt the native PBPs.[19]

Effectively, this means that there are two gene complexes represented on each SCC*mec* along with numerous non-essential inserts and transposons carrying genes for resistance to other drugs and heavy metals like mercury and cadmium.[19] The numerous inserts, transposons, and repeats can typically be found in the "junkyard" or J regions, of which there are three, J1, J2, and J3.[20] Generally, the genomic structure of a SCC*mec* can be considered in the following way: Right chromosomal junction, J1 region, mec gene complex, J2 region, ccr gene complex, J3 region, and the left chromosomal junction.[20] Each of these regions may contain homologues of the mec and ccr genes, the combination of which will determine the SCC*mec* type. Differences and unique elements in J1 and J3 regions are used to determine SCC*mec* subtypes. In Figure S1, a very rough annotation of the genomic regions of an alignment of the SCC*mec* Types I–IV can be consulted, with more advanced annotation available elsewhere.[21] The *S. aureus* genome is about 2.8 Megabase pairs in size, so the SCC*mec* typically constitutes 1–2% of the total genome.

In clinical settings, MRSA is typically identified in a number of ways: (1) culturing with antibiotics on agar and measuring the distance between colonies and the antibiotic pill after overnight culture, (2) running a PCR for the PBP2a gene together with other identifying genes, or (3) sequencing the sample. The main limitation of the most routinely used methods, the culturing and PCR, is that only a few targets can be multiplexed in practice (typically max 10), while Fluorocode OM allows for a large-scale screening. The main limitation for Fluorocode OM is the size of the genomic elements targeted—as will be duly explained in later sections.

The use of optical mapping (OM) to identify plasmids carrying resistance genes was explored in a recent study by the Westerlund lab.[22−25] Here, we investigate whether OM can be used to detect MRSA, distinguish it from MSSA, and identify the SCC*mec* types. The advantage of using OM for MRSA identification over the standard methods listed above is that it can detect many more targets (e.g., other resistance MGEs, or even other bacterial strains) and pick up structural variation in the MGEs and elsewhere in bacterial genomes. The implementation of OM used here is termed Fluorocode OM. Fluorocode OM is a high-resolution method relying on methyltransferase-mediated labeling of DNA, permitting recognition of the MRSA resistance cassettes, as small as 26 kbp in length. Fluorocode OM was shown to identify bacteriophage genomes and subsequently individual bacterial cultures, such as the bacteriophages lambda and T7 as well as the chromosomes and plasmids of bacteria *Escherichia coli* and *Vibrio campbelii* (formerly *Harveyi*).[26,27] In silico work by our group has produced interesting results regarding the identification of closely related strains, as well as the identification of relatively shorter genomic elements such as the SCC*mec*.[27] The bacteriophage genomes are between 40 and 70 kbp, while bacterial genomes are typically 1–4 Megabase pairs in length. It is of note that the scale of genomic elements to consider is different in OM compared to sequencing where the typical lengths are, for example, reads of 150 base pairs or amplicons of up to 1000 base pairs for PCR. Typically, OM methods will look at elements that are tens of kilobase pairs, not unlike third-generation sequencing approaches, such as the Oxford Nanopore and Pacific Biosciences SMRT sequencing.

The Fluorocode methodology is described in detail in Bouwens *et al.* and D'Huys *et al.*[26,27] The method relies on the natural functionality of the methyltransferase enzyme M.Taq,

https://doi.org/10.1021/acsomega.3c05902
ACS Omega 2024, 9, 8862−8873

Table 1. Summary Information about the *S. aureus* Strains and SCC*mec* Used in This Work[a]

strain	NCBI accession code	length kbp	SCC*mec*	SCCmec type	country and year of first record	TCGA sltes/kbp (average)	ref
T0131	CP002643.1	2913.9	A8037671	III	China, 2006	2.4	42
N315	NC_002745	2814.8	D86934	II	Japan, 1982	2.4	43
USA300	NC_007793.1	2872.8	A8063172	IV	USA, 2000	2.4	44
11819-97	CP003194.1	2846.5	AB063172	IV	Denmark, 1997	2.4	45
NCTC10442	NZ_UNCH01000003	2853.1	AB033763	I	UK, 1960	2.4	46
ATCC 43300		2946.9	D86934	II	USA, 1996	2.4	47
ATCC 12600		2782.7	none		UK, 1935	2.4	48
NCTC 8325	NC_007795.1	2821.4	none		UK, 1943	2.4	49
SCCmec type							
Type I	AB033763	39.3				2.3	21
Type II	D86934	58.2				2.1	21
Type III	AB037671	68.3				2.4	21
Type IV	AB063172	26.1				2.5	21

[a]The publicly available NCBI accession code (if available), genome length in kilobase pairs (kbp), the SCC*mec* present by accession code and type, year, and country of the first record, average number of 5′-TCG**A**-3′ sites present per kbp, and the reference to the strain information in the final column to the right.

originally isolated from the thermophilic bacterium *Thermus aquaticus*.[28] In native environments, this enzyme methylates DNA, with S-adenosylmethionine (SAM) serving as the methyl donor. The sites recognized for methylation by M.Taq are 5′-TCG**A**-3′, transferring a methyl group to the sixth nitrogen of the adenosine. In the case of Fluorocode, we provide a synthetic SAM analog carrying a fluorescent dye instead of a methyl group.[29] The enzyme transfers the dye with the PEG linker to the sixth position on the adenosine in the 5′-TCG**A**-3′ sites, resulting in DNA with covalently bound fluorophores at the respective sites (Figures 1 and S2). During the labeling step, the enzyme will capture the cofactor in its active pocket and slide along the DNA, searching for the recognition site. Once the enzyme encounters a 5′-TCG**A**-3′ site, it will flip out the Adenosine from the double helix and capture it in the appropriate binding pocket.[30,31] The proximity of the cofactor and the labeling target catalyzes the transfer of the linker with the dye from the cofactor to the 6-position of the adenosine in a trans-alkylation reaction. Subsequently, the adenosine is flipped back into the double helix.[32] Research by other groups has shown that off-target labeling by M.Taq is unlikely due to the sequence-specific docking of the methyltransferase enzyme.[30] Any (unlikely) off-target labels will impact the significance score threshold, designed to filter out poorly matched molecules, which are typically characterized by low matching or cross-correlation scores.

The labeled DNA is linearized. Individual DNA strands are deposited on a microscopy cover slide in a fashion that leaves them stretched out as straight lines and even overstretches the DNA molecules to approximately 1.7 times the length undisturbed.[26] Next, these microscopy cover slides are imaged with a wide-field fluorescence microscope. Large tile-scans are produced: a single field of view from such a tile-scan with a magnified single labeled DNA trace in a separate window in blue is shown in Figure S3.[26,33] The DNA is labeled uniformly across the genome at the 5′-TCGA-3′ sites present in the sequence, with labeling efficiency well ≥80% and to completion, as routinely tested via restriction assays.[32] The efficiency is likely even higher, since previous efforts by our group determined the efficiency of a predecessor method to be at 70% due to the use of a two-step labeling mechanism.[34] The conclusion was that the labeling efficiency was lowered due to the inefficient second step of the labeling chemistry. Since then, we have developed direct labeling cofactors, which considerably improve on the labeling chemistry efficiency of the previous iterations of Fluorocode.[29] The amount of enzyme and cofactor used in the wet-lab labeling reaction is optimized to avoid under-labeling samples. The resulting readout is a linear string of signals in a genome-specific pattern, resembling a barcode when imaged with a fluorescence microscope. Such barcodes constitute a unique shorthand for the genome in question. The experimentally obtained signal traces are compared to a database of in silico-generated traces.[26,27] The database is curated to contain a set of reference species of interest, in this case, the SCC*mec* elements I–IV. The "bars" of the barcode are preprocessed: each trace is locally normalized within a sliding window of 10 kb to avoid biassing by local changes in label density.[26,35] The intensity traces of the measured labeled DNA traces are then aligned to each reference trace by means of cross-correlation, and the maximum alignments are subjected to significance testing, as described in detail previously.[26] The signal or recognition site density per kilobase pair (kbp) of the genomic sequence is of relevance for successful identification by Fluorocode, with a minimum average recognition site density of 1 site per kbp. All the SCC*mec* cassettes carry between 2 and 2.5 5′-TCG**A**-3′ sites per kbp, making them a suitable target for Fluorocode detection (Table 1).

We have previously demonstrated by means of simulation that Fluorocode has the theoretical capacity to identify the exact SCC*mec* present in an *S. aureus* sample.[27] In this work, we showcase how Fluorocode performs in relation to real samples from cultures of *S. aureus*. Subsequently, we show the identification of MRSA DNA in complex mixtures of bacteria. Together, the experimental work shown in this work constitutes a proof of concept, clearly showing the potential of this technology, as we work toward fulfilling key requirements for a move toward clinical utility. Finally, this proof of concept shows the capacity of Fluorocode OM to capture genomic elements larger than genes but smaller than whole plasmids for genomic research.

https://doi.org/10.1021/acsomega.3c05902
ACS Omega 2024, 9, 8862−8873

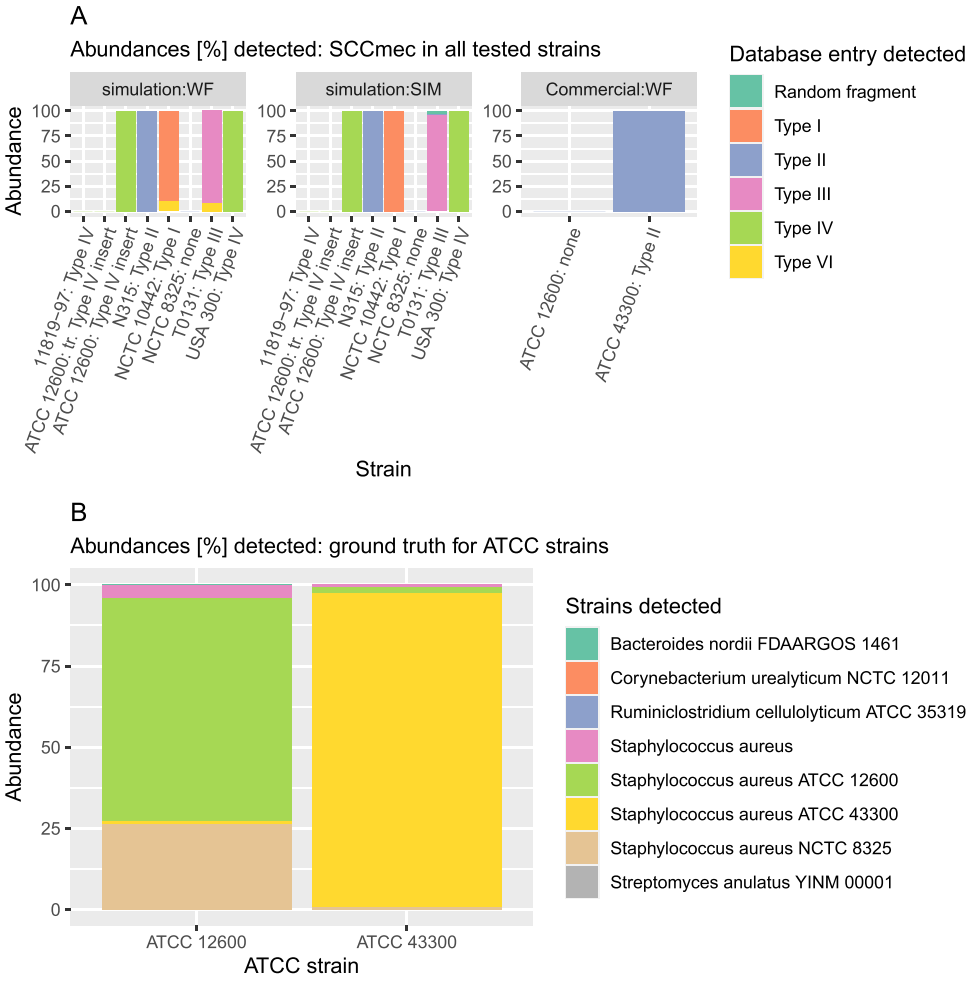


Figure 2. (A) The associated SCCmec type is indicated next to the name of the strain. The legend on the right indicates color representation of SCCmec in the graph. The three panels illustrate (in order from left) the simulations in Widefield and SIM systems, following the results from wet lab experiments with ATCC samples imaged with a WF system. The stacked bar charts document the abundance of the SCCmec type detected. The SCCmec was detected in all the cases except 11819-97, where we hypothesize that the SCCmec is truncated to a size below the limit of detection by Fluorocode OM. To test this hypothesis, we simulated the MSSA strain ATCC 12600 to contain the SCCmec Type IV cut up into three pieces (6, 8, and 12 kbp) randomly inserted into the genome. As expected, in this case, no SCCmec was detected neither with WF nor SIM, contributing to a positive conclusion for the truncation hypothesis, as these results coincide with the results of the strain 11819-97. The simulations with SR-SIM microscopy show that most false-positive matches are eliminated, while with WF alone, there are about 12.5% of false positives for NCTC 10442 and T0131. While eliminating FPs is preferable, the loss of time using SR-SIM outweighed the gain in TP abundance, as at >85% abundance of a single SCCmec, we consider the match of good quality. The WF results show that experimentally, Fuorocode OM detects the absence and presence of SCCmec correctly in both cases for ATCC 12600 and ATCC 43300, respectively. No FPs were observed. (B) The experimental data sets were tested for ground truth detection as a control. The strain ATCC 12600 was matched to itself at a little less than 75% and exhibits an average nucleotide identity (ANI) of 99.83% with NCTC 8325, the second most abundant matching species at about 25%. The status of this match as an FP is not obvious, given the sequence similarity between the two strains. The third strain showing up with about 5% matches is the *S. aureus* genome that is provided by the commercial bacterial mixture used for the other experiment. In the meantime, the strain ATCC 43300 has an ANI of 97.45% with NCTC 8325 and 97.55% with ATCC 12600, which turned out to be sufficiently different to tell the strains apart with the Fluorocode OM workflow presenting ∼1% of FPs each.

RESULTS

All the strains were subjected to the same alignment tests using two separate bioinformatic tools, one specially designed to detect and identify MRSA resistance elements and another to reveal the presence of closely related sequence fragments despite inversions and fragmentation across sets of query sequences. First, the top hit of the SCC*mec* Finder (SmF) package was selected (Figure 2).[36] All the strains were then aligned with their respective SmF top hit cassette using the Mauve alignment package.[37] The SmF and Mauve are described in detail in the sections discussing the steps to define the ground truth.

Performance Evaluation: Simulation. An extensive in silico study was conducted to investigate the theoretical boundaries of applying Fluorocode OM to the identification of MRSA. In silico-processed MRSA and MSSA strains were analyzed against the SCC*mec* types I−IV as the most frequently encountered in human populations in both CA and HA-MRSA strains based on the International Working Group on the Staphylococcal Cassette Chromosome elements (IWG-SCC). The simulation pipeline starts by generating images of labeled, randomly fragmented genomic DNA of the organisms of choice, in this case, five fully sequenced MRSA strains and one MSSA strain. The main parameters of interest

https://doi.org/10.1021/acsomega.3c05902
ACS Omega 2024, 9, 8862−8873

include molecule length in kilobase pairs, labeling efficiency, and level of background noise, and, of course, the parameters of the optical systems. Off-target labeling is not an issue for M.Taq in our experience, and this is corroborated by recent investigations showing that this enzyme is especially faithful due to the amino acid functionality within the active pocket.[31]

Next, we constructed a database of the SCC*mec* I−IV as well as 1000 artificially scrambled random DNA sequences, 80 kbp in length (Materials and Methods: Simulations). A database containing a sufficiently high amount of random fragments is a strategy to evaluate the chance of random matching and, therefore, false positives being scored as significant. To assess the impact and potential advantage of higher optical resolution of the microscopy modalities used on the capacity of the Fluorocode OM readout precision, the simulations are repeated with the theoretical parameters of two imaging systems: first, a lower resolution system—diffraction-limited widefield system—and second, a higher resolution (referred to as a super-resolution) structured illumination microscopy system (SR-SIM).

The results for SCCmec types I−IV identification in simulations of five MRSA and one MSSA strain can be consulted in the first two charts from the left in Figure 2, panel A. From these results, we conclude a positive evaluation to continue with the wet lab experiments, as four MRSA strains out of five are exactly determined with their respective cassette present. The detected SCC*mec* lengths range from 26 to 68 kbp (see Table 1 for SCC*mec* and MRSA/MSSA strain information). The MSSA strain serving as the negative control shows no matches with any of the cassettes. The simulation results for both SIM and WF can be consulted in panel A in Figure 2.

One MRSA strain (11819-97) does not exhibit any matches. Originally, the hypothesis was that optical mapping would encounter issues when dealing with sequences shorter than any of the SCC*mec*, at a threshold below 20 kbp. By extension, should the cassettes in the genome that is being mapped be truncated and split up into shorter fragments, Fluorocode OM would struggle to detect the cassettes altogether. Since the results of the other simulated MRSA strains indicate that in general, SCC*mec* detection and identification go very well, the most likely scenario would be that the cassette in 11819-97 is split apart into shorter fragments. This hypothesis has been tested by aligning the strain assembly available from NCBI and the expected SCC*mec* Type IV. It became apparent that indeed the cassette was truncated to fragments of approximately 10 kbp and less (Mauve readout in Figure S4, Panel A).

To further corroborate this line of reasoning, we generated a simulated MSSA data set of ATCC 12600 and incorporated a full Type IV SCC*mec* as well as an ATCC 12600 data set with randomly distributed Type IV cassette split in three pieces at random, with each piece at approximately 8 kbp in size (6, 12, and 8 kb). We repeated the simulation for both WF and SIM imaging systems. The results are available along with the rest of the simulation results for WF and SIM in Figure 2, panel A. These results indeed agree with our hypothesis, as neither WF nor SIM were able to detect the SCC*mec* when it was truncated, while both detected the full 26 kilobase pair SCC*mec* when it was present as an unfragmented genomic structure.

A priori one may expect that an SR-SIM approach would yield better results, meaning that the chance of true positive detection would be higher and false positives would be lower than the regular WF approach. For the WF simulations, some type IV cassettes are identified in NCTC 10442, although only Type I should be found. These false positives do not occur in the SIM results. A similar improvement is observed for T0131. Indeed, the majority of the already few false positives perceived with a WF system were removed upon using SR-SIM; however, the supposedly truncated strain was detected with neither method. These results revealed that the additional time investment for imaging and analysis with an SR-SIM system is not justified, since ≥85% of the WF matches were true positives. Therefore, a WF system was selected for the imaging step of the wet lab workflow.

Performance Evaluation: Wet Lab Experiments. The wet lab experiments constitute a good test for the preliminary results obtained with simulations as well as provide an indication of the viability of the method for the use of MRSA typing. To test the Fluorocode workflow for the task of MRSA typing, two ATCC *S. aureus* strains were investigated—a methicillin-resistant strain (ATCC 43300) and a methicillin-susceptible (ATCC 12600) strain. The strains were cultured under standard conditions and extracted using a commercial high molecular weight DNA extraction kit. The DNA samples were labeled, purified, and deposited, as previously described.[26,27] The imaging and analysis workflows are described in detail in the Materials and Methods section.

The ATCC strains are well-characterized *S. aureus* strains, with full genomic sequences available from ATCC. We evaluated the optical mapping readout against the analysis for two databases, both the SCC*mec* database described previously (Figure 2, panel A) as well as an integrated database including 3700 bacterial species, with the two ATCC sequences included as the ground truth (GT) (Figure 2, panel B). Here, we expected to see that only the GT would be picked up, with the caveat that very similar *S. aureus* strains with high average nucleotide identity (ANI) would also likely show up. Panel B of Figure 2 shows that even when the ANI is above 97%, *S. aureus* strains can reasonably be differentiated, while at ≥99% ANI, we are reaching the limits of the method for differentiating between two strains. The MSSA strain NCTC 8325 constitutes around 26% of the ATCC 12600 data set, with the ANI between these two strains at 99.83%. Meanwhile, there are practically no matches with NCTC 8325 from the ATCC 43300 data set, with the ANI between the two strains at 97.45%.

Just like in the simulation work, the MSSA data set served as the negative control for the wet lab pipeline processed samples for SCC*mec* detection. No SCC*mec* hits were expected for the MSSA. The ATCC 43300 MRSA strain carries a full Type II SCC*mec*, additionally confirmed by aligning the Type I−IV SCC*mec* sequences against the ATCC 43300 genome sequence. The alignment readout can be consulted in Figure S4, panel B.

To discuss OM in terms comparable to sequencing, we examined the genome coverage achieved in the optical mapping runs. In sequencing, coverage is a metric quantifying the number of times the same nucleobase is *read,* or how many reads on average are matched to every genomic location, informing of the confidence one may have in the final sequencing readout. The situation is comparable in Fluorocode OM, where the signal traces obtained from single imaged DNA molecules are considered as reads, which are likewise matched along the whole candidate genome in search of a significant match at a particular genomic location. Just like with

https://doi.org/10.1021/acsomega.3c05902
ACS Omega 2024, 9, 8862−8873

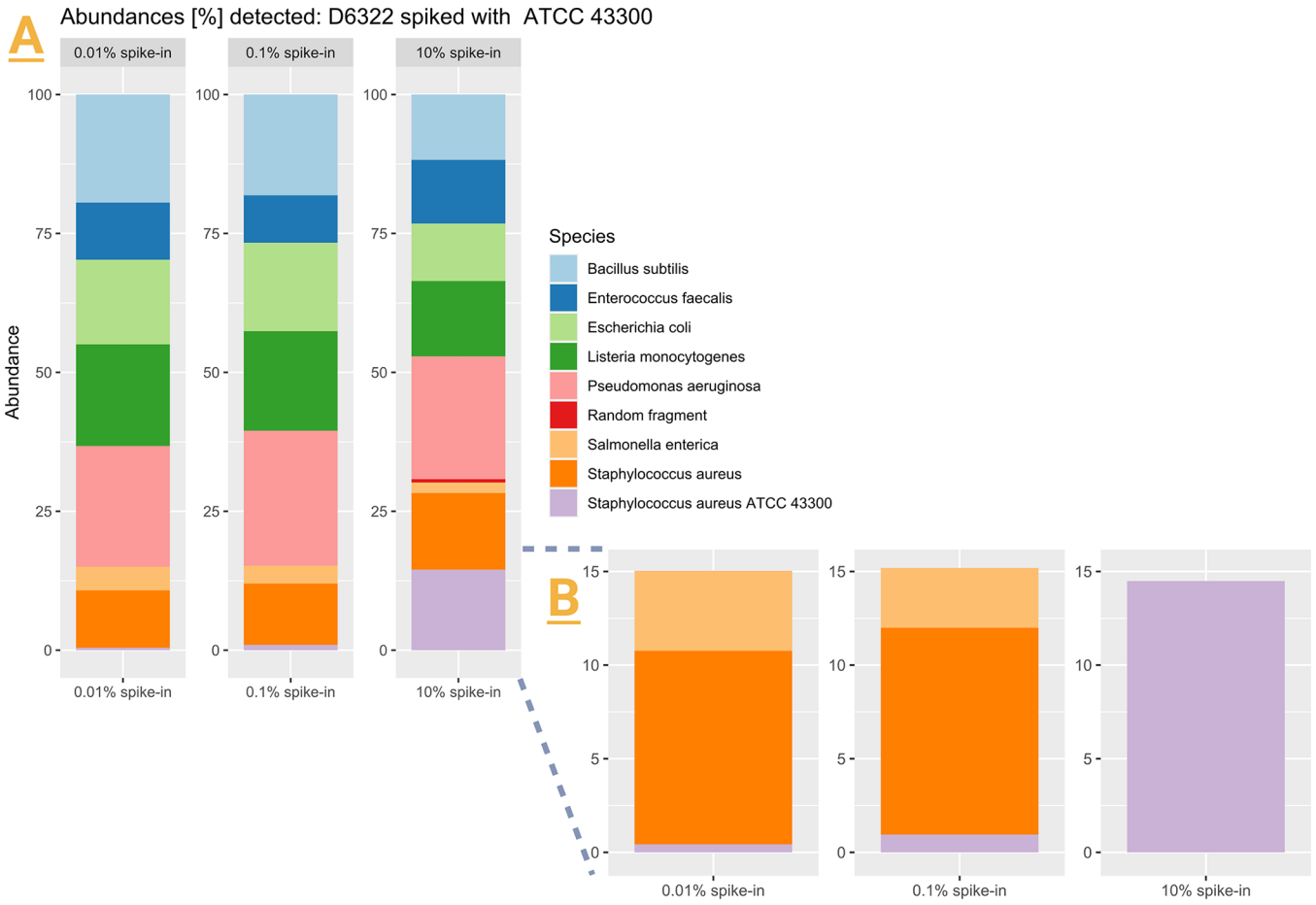


Figure 3. (A) Three samples of seven bacterial genomic DNA in equal proportions spiked with 0.01, 0.1, and 10% MRSA (ATCC 43300) DNA, respectively. All three samples show detection of the MRSA strain separately from *S. aureus* originally present in the mixture. (B) "zoom in" of the stacked bar. The exact percentage of the ATCC 43300 detection is 0.42, 0.94, and 14.48% for the 0.01, 0.1, and 10% spike-ins, respectively.

sequencing, Fluorocode OM can determine the minimum and average coverage across the genome—visualization of Fluorocode OM coverage achieved in the wet lab data sets is shown in Figure S5 (for full sample coverage) and Figure S6 (for the coverage of the individual genome of a spike-in strain added to a mixed species sample).

For the simulated data sets, the sampling and as a consequence the coverage was designed to be even across the genome, while in the wet lab experiments, such uniform sampling has to be established empirically. An *a priori* advantage of optical mapping compared to 16S rRNA sequencing and shotgun sequencing is the fact that no polymerase chain reaction (PCR) amplification is used, avoiding bias toward certain genomic regions introduced by both local polymerase enzyme affinity and primer design.[38] The evaluation of the coverage of the cultured ATCC strains corroborated this expectation. The global coverage achieved by Fluorocode was satisfactory, covering the full genome several times. However, we do observe local variations or "spikes". We hypothesize these to be at least partially an effect of intrinsic local properties of the double-stranded DNA as well as native enzymatic activity, where some nucleobase combinations will be more prone to double-strand breakage.[39−41]

The SCC*mec* typically constitutes ∼1−2% of the total genome size, making it important to achieve sufficient coverage for Fluorocode to perceive such small elements. The ATCC 43300 case illustrates that Fluorocode is sensitive enough and does not actually require the coverage of 100x used for the simulations, as the cassette is already detected at 3−5× coverage when investigating a pure culture sample.

Performance Evaluation: MRSA-Spiked gDNA Mixture. The next step was to investigate how well Fluorocode is able to discern an MRSA strain in a mixture of bacterial genomes. We used commercial genomic DNA (gDNA) mixtures (ZymoBIOMICS D6322) containing seven different bacterial strains including a methicillin-susceptible *S. aureus* strain. Three mixtures were prepared "spiked" with 10, 0.1, and 0.01% ATCC MRSA strain gDNA (Figure 3, Panel A). MRSA was detected in all cases, including the 0.01% MRSA-spiked sample (Figure 3, panel B). The ANI between MSSA present in the gDNA mixture originally and the ATCC 43300 MRSA strain is 97.51%. This experiment provided a suitable test case to check whether Fluorocode OM can detect MRSA in the presence of other bacteria and, especially, other *S. aureus* strains.

No SCC*mec* were detected in the spiked samples. To understand why this may be the case, we plotted the coverage of the individual strain ATCC 43300 from the three mixed samples, which is shown in Figure S6. In all three cases, the coverage of the MRSA strain is below a full one time (1×). While this was apparently enough to detect the strain as such among other bacterial strains and even a highly similar other *S. aureus* strain, the coverage was too low to detect the SCC*mec*. The data sets with individual ATCC strains already show that a

https://doi.org/10.1021/acsomega.3c05902
ACS Omega 2024, 9, 8862−8873

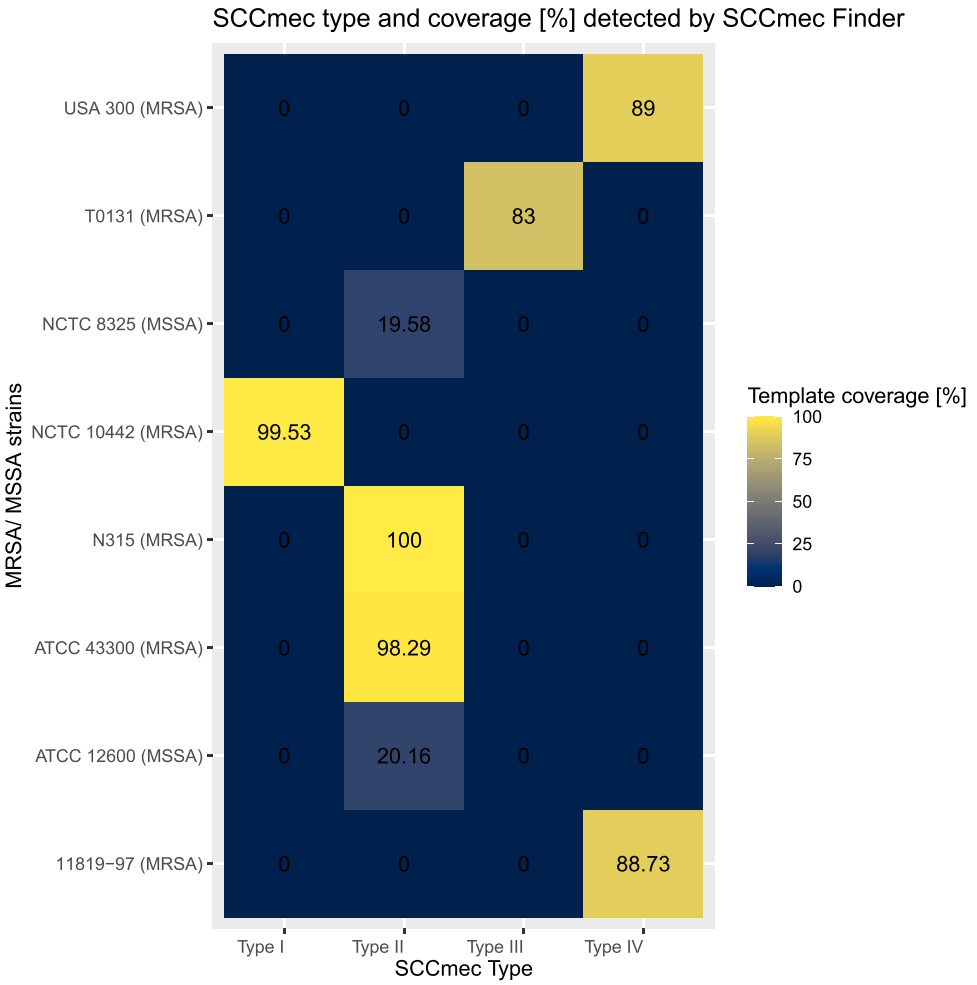


Figure 4. Readout summary from the SCCmecFinder pipeline. All the strains investigated in this work are presented on the *y* axis, with MRSA or MSSA status indicated in the brackets. The k-mer template coverage in percentage is one of the metrics of SmF for determining a SCCmec-type presence. The lowest reliable threshold for template coverage proposed by the Larsen group is 50%, and all the MRSA strains investigated in this work exhibit well above 75% of k-mer template coverage.

coverage of at least 3−4× of the SCC*mec* carrying strain is one of the requirements for successful SCC*mec* detection with Fluorocode OM.

The amount of classified DNA was 449 Mbp for the 0.01% spike-in, 403 Mbp for the 0.1% spike-in, and only 14 Mbp for the 10% spike-in, with other workflow parameters being equal. Figure 3 panel A shows the detected abundances of individual bacterial species from the mixture and the two *S. aureus* strains. Figure 3 panel B shows a close-up of the detected abundances of the MRSA strain ATCC 43300. The data set of the 10% spike-in is a whole order of magnitude smaller than the other two data sets. However, since the spiked amount constitutes a relatively large proportion of the sample, we were nevertheless able to achieve sufficient coverage to detect the ATCC 43300 strain. In the meantime, while the data set for the 0.01% spike-in is a lot bigger, the proportion of the ATCC 43300 is 1000 times less than the 10% spike-in. As a consequence, while we were able to detect the presence of the strain as such, Figure S6 reveals that the coverage we did achieve of this strain is very low.

SCC*mec* Finder: Defining the Ground Truth. To ensure external validation of the Fluorocode readout, we matched our results to those obtained with a bioinformatic tool specifically designed to detect and type SCC*mec* types present in sequenced *S. aureus* genomes, the SCC*mec* Finder (SmF).[36] Here, a template coverage of at least 50% is suggested to yield a reliable SCC*mec* type prediction. All the strains used in both the Fluorocode simulation and wet lab data acquisition were processed with the SmF tool, and the SCC*mec* with the highest k-mer template coverage was chosen for further reference in optical mapping as the ground truth (full readout shown in Figure 4). It is noteworthy that the MSSA strains also showed some template coverage, but well below the reliability threshold. This highlights the ambiguity issues that can arise in SCC*mec* detection and identification, despite the use of advanced, sequencing-based tools.

Mauve Alignment: Determining the Genomic Arrangement of the Ground Truth. Mauve is a genome alignment package. The Mauve aligner is tailored to genomes that have a close evolutionary relationship while carrying strain-defining differences such as inversions, horizontal transfers, and other structural rearrangements. Mauve is particularly effective at identifying collinear clusters of matches or matches with the same chromosomal coordinates, where the exact order of genomic elements such as genes is repeated. Mauve is designed to detect the presence of such regions across two or more sequences and does so with higher precision than the more widely used alignment algorithms,

https://doi.org/10.1021/acsomega.3c05902
ACS Omega 2024, 9, 8862−8873

such as the Basic Local Alignment Tool (BLAST). BLAST is useful for checking when there is a full SCC*mec* present, while Mauve is more useful when the cassettes carry minor structural changes. In this case, Mauve proved to be a visually more accessible tool in the cases of inversions of full or truncated cassettes, permitting a close examination in order to evaluate the Fluorocode OM findings.

Every strain subjected to Fluorocode OM workflow was aligned with the ground truth SCC*mec* as defined by the SmF using the Mauve package, both for simulations and wet lab work pipelines. Figure S4, panel A shows that the type IV SCC*mec* detected by the SmF in the MRSA strain 11819-97 is truncated according to the Mauve alignment. This result corroborates our hypothesis that Fluorocode was unable to detect this cassette in the simulated data set because the 26 kbp cassette is present in a truncated form. In the meantime, the Mauve alignment of the MRSA strain ATCC 43300 shows a full type II SCC*mec* present in the strain genome; the alignment is shown in panel B of Figure S4.

DISCUSSION

To effectively validate the Fluorocode workflow for the detection of kbp range unique fragments, we set out to benchmark the MRSA and MSSA strains of interest to two separate bioinformatic pipelines, SmF and Mauve.[36,37] With SmF, the exact SCC*mec* was determined. As described earlier, the SCC*mec*'s share several gene complexes and functional regions, resulting in homologous regions across all the cassettes investigated. To validate the optical mapping result, it was important to determine the exact SCC*mec* type that we may expect to perceive. We have illustrated that OM can successfully detect that the strains are MRSA and correctly determine the SCC*mec* cassettes in simulation, with one exception, strain 11819-97. Investigation of the sequence assembly using the Mauve alignment yields the conclusion that the fragmentation of the cassette renders the fragments too short for reliable identification by Fluorocode. Using increased optical resolution (SIM) improved the specificity of detection for the other strains by reducing the false positive detection, while the type IV cassette in 11819-97 remains undetected. This case is indicative of the limit of detection of unique kbp elements by the Fluorocode method, where fluorescence traces of sequences below 20 kbp do not contain sufficient genomic information to be assigned to their genomic origin with confidence.

With positively convincing simulation results, we prepared data sets with the two known strains: ATCC 43300 and ATCC 12600 as positive and negative MRSA controls, respectively. From the Fluorocode analysis readout (Figure 2), we observed that both the MSSA and MRSA were correctly determined. Moreover, the exact cassette present was determined, and no false positives were observed. An important consideration is that the coverage achieved for the ATCC MSSA and MRSA genomes was found to be less evenly spread than suggested by the in silico-generated data sets (Figure S5). Nevertheless, while there are several regions that exhibit especially high coverage for both data sets ("peaks") in Figure S5, one may observe that both genomes have optical maps matched to the full length of the genome several times over, ensuring complete coverage. This fulfills the requirement for the detection of a relatively smaller genomic element, such as an SCC*mec*. While the simulations represent a proof of concept for the ability of Fluorocode to identify MRSA, they also represent an ideal case in terms of genome coverage and data set size. The ATCC strain data sets illustrate that, while experimental results are much more varied despite all the preparatory steps being equal, Fluorocode is resilient against an uneven and relatively low coverage and data set size and will detect the SCC*mec* regardless.

The experimental readouts of the spiked gDNA mixtures illustrate the capacity of Fluorocode to detect and resolve highly similar strains, as determined by ANI scores, despite relatively low coverage and low relative abundance of the strain of interest. Since no cassettes were detected in these samples, coverage was examined, revealing that none of the spiked samples achieved full coverage across the ATCC 43300 genome. The task was further complicated by the mixture containing another *S. aureus* strain altogether. Regardless, Fluorocode can detect the MRSA strain amidst all of the other present bacterial genomes, with a 100% recall for every species present. These samples show that in cases where a genome of interest may be present in a low proportion, the sample size requirement is higher. When the genomes of interest are abundantly present, the genome size may be lower, as illustrated by the reproducibility of composition detection across the three spiked data sets. The 10% spike-in shows altered relative abundances, as it accommodates an eighth component of the mixture, close in relative abundance to the other bacterial species.

Fluorocode OM can also be considered for researching structural variations of genomic regions across related genomes. The main limiting factor that surfaced in the present work is that once genomic fragments become too short, it becomes difficult for the Fluorocode analysis algorithm to match the experimental trace with any library trace. Therefore, the detection of structural variation in the shorter genomic elements such as the shorter SCC*mec* will be difficult. There are several potential solutions here, including increased optical resolution, such as using SIM. However, SIM alone will not be enough, as we have already demonstrated in the present work. An additional strategy to improve the detection capacity of Fluorocode is introducing multicolor labeling, where the same number of recognition sites would be read out in at least two color channels, allowing for the improvement of resolution through the sample preparation instead of altering the imaging systems. In the same vein, multitarget labeling is another strategy that may be used to improve the resolution, where an enzyme with a different recognition site and spectrally separated fluorophore would be used. Together, these methods would allow us to tackle the length limitation Fluorocode OM encounters, as concluded from the in silico and experimental data shown in this work. An in-depth analysis of structural variation is indeed possible with Fluorocode OM. To this end, we intend to develop methods to construct a *de novo* map from experimental "Fluorocoded" molecules. Comparing the *de novo* map to a reference map will then allow us to identify the structural variants.

CONCLUSIONS

This work constitutes a proof of concept for the use of Fluorocode OM for detection of genomic elements, SCC*mec* of MRSA in particular, an idea that was previously examined by our group in silico.[27] With previous findings by other optical mapping methods showing the detection of plasmids at 66 kbp, we have shown that OM as applied by Fluorocode is capable of detecting elements as short as 26 kbp.[22] Additionally, we have

https://doi.org/10.1021/acsomega.3c05902
ACS Omega 2024, 9, 8862−8873

shown that Fluorocode OM is able to pick up a MRSA strain along with various other bacteria and even another *S. aureus* strain from a complex mixture, thereby further emphasizing the value of this approach for simultaneous screening of a sample for a wide variety of targets besides the MRSA-resistant element.

These results are encouraging for further investigation into the potential of Fluorocode OM for pathogen detection and identification. The next steps include validation in a clinical setting with patient isolates. This goes together with the goal of optimizing cultivation-free workflows for pathogen detection, sourcing the DNA from skin swabs, for example. In this work, we have already shown that as little as 100 ng of DNA is sufficient for sample identification for both individual cultures and mixed samples.

MATERIALS AND METHODS

Samples, Culturing, and DNA Extraction: *S. aureus*. All *S. aureus* cultures were stored in glycerol at −80 °C. Before extraction, the strains were cultured on a Luria−Bertani (LB) agar plate overnight at +37 °C and subsequently cultured overnight in a liquid LB medium overnight at +37 °C in a shaking incubator. DNA of all strains was extracted with the Circulomics CBB Nanobind kit (now Nanobind HT CBB kit from PacBio), using Lysostaphin (Sigma-Aldrich), according to the manufacturer's instructions. The MRSA strain was extracted in the same way, except prior to starting the kit, the MRSA culture was vortexed and subsequently centrifuged with 100% ethanol (99.8%+ for analysis, Acros Organics).

Spiked Genomic DNA Mixtures. The gDNA mixture (ZymoBIOMICS, D6322) was spiked with 10, 0.1, and 0.01% ATCC 43300 DNA based on stock DNA concentration. 100 ng (Qubit) of each D6322/MRSA sample was labeled.

YOYO1 Labeling. Before the OM workflow, the DNA quality was assessed using Oxazole Yellow Homodimer (YOYO1) labeling. 100−300 ng of the extracted DNA was pipetted with a cut pipet tip in a solution with 10 μM YOYO1 dye (produced in-house), and 2-ethanesulfonic acid (MES hydrate, Thermo Fisher dissolved in Milli-Q water to the final concentration 5.6 Mol with pH) was added in proportion 1:10 to the final experimental volume. Milli-Q water was added as needed to achieve the full sample volume (typically 10−20 μL). The DNA was linearized using the rolling droplet deposition method, followed by widefield imaging.[23]

Optical Mapping. The extracted HMW DNA was labeled as described previously in a one-pot reaction with the enzyme and synthetic fluorescent cofactor and subsequent purification of excess dye.[26,27] As little as 100 ng of HMW DNA was incubated with 40-fold excess to recognition sites of the methyltransferase enzyme (VIB, Ghent) at 60 °C for 60 min, followed by enzymatic quenching of the methyltransferase activity by incubation with Proteinase K (NEB) for 1 h at 50 °C. The mixture was purified from excess dye in a 2% agarose plug, as described previously, followed by dialysis on Millipore membranes in two 45 min steps (0.1 μm, MF-Millipore membrane, Merck).[26]

Imaging. The linearized labeled DNA samples were imaged by tile-scanning on an Axio Imager Z2 (Zeiss)—a widefield epi-fluorescence setup with the following parameters: Hamamatsu Orca Flash (ORCA-Flash 4.0 V3) camera with the camera pixel size at 6.5 μm and a 63x oil immersion objective with 1.4 NA. A 100 mW 561 nm excitation laser (Toptica iChrome MLE) was used for methyltransferase-labeled samples with an integration time of 500 ms per field of view. A 480 nm excitation was used for imaging the YOYO1-labeled samples.

Analysis. The acquired tile-scans are processed in the following manner. First, the tile-scans are segmented, where the signal traces or optical maps are selected and filtered for length; traces below 26−30 kbp are discarded. The experimentally acquired optical maps or experimental traces are converted to intensity traces to be compared against a database of candidate traces, which are prepared in the following way. The database was constructed by selecting reference sequences from NCBI RefSeq and GenBank with complete assemblies. The bacterial genomes were filtered for the two following parameters: (1) the specimens were originally isolated from the human gut; (2) the genomes contain the recognition site 5′-TCGA-3′ at least once every kilobase pair. The filtered database contained a total of 3765 reference sequences. This database was used to analyze the pure ATCC MRSA and MSSA data sets. To analyze the gDNA mixed sample, a database was prepared, containing the sequences of the bacteria present in the original gDNA mixture as obtained from the manufacturer (obtained in FASTA format from ZymoBIOMICS), the sequence of the ATCC 43300 (obtained in FASTA format from ATCC), as well as 1000 randomized genomes of the size 4 Mbp with the recognition site density at 3.5 sites/kbp.

For analysis of the SCC*mec* presence, a special database was prepared including SCC*mec* I−VI and 1000 randomly generated 80 kbp fragments, with recognition site density at 3 sites/kbp to mimic a typical recognition site density of bacterial genomes. Each database entry is processed to generate an *in silico*-labeled genome based on the optical parameters of the system, with which the experimental traces were obtained. Each experimental trace is cross-correlated to each candidate trace and a cross-correlation maximum is obtained. Finally, a significance score is computed, based on which a decision for acceptance or rejection of the alignment is made. Only unambiguous matches are accepted.

Simulations. Five MRSA strains and one MSSA strain were selected for simulations to both repeat the previous simulation (N315 and T0131) as well as simulate different SCC*mec* types and genomic arrangements, namely, NCTC 10442, 11819-97, USA 300, and NCTC 8325 as the MSSA control strain. Once the exact cassette type was identified using SmF, the Mauve alignment package was used to pinpoint the SCC*mec* genomic arrangement in the strain genome. The data sets were simulated at 10,000 traces of 40 kbp for WF and 20 kbp for SIM simulations, achieving about 300 Mbp per strain, with the theoretical coverage of the *S. aureus* genome that is typically around 2.7−8 Mbp, approaching a 100x. The strains were evenly sampled to achieve a reliably high minimal coverage at all genomic locations. Nucleotide sequences are converted into optical maps. First, all 5′-TCGA-3′ labeling sites within the genome are precisely located. Next, the sites are randomly selected for labels with a labeling efficiency ranging between 75 and 95% per simulated trace. There is no single set point for efficiency; as per our experience, a minimal efficiency of 75% is required to achieve matching of the optical map to the database entry. The higher limit of efficiency at 95% is assumed based on two factors: the restriction assays, where we estimate the enzyme amount required for full protection (and in principle, this means a 100% efficiency) as well as our observations with the highest cross-correlation

https://doi.org/10.1021/acsomega.3c05902
ACS Omega 2024, 9, 8862−8873

scores for true positive matches, which come close to 100% label coincidence between the experimental map and the *in silico*-generated map of the ground truth genome. Next, the simulated labeled traces are converted into optical maps by convolution with the point-spread function (PSF) of the microscope. The PSF was approximated using a Gaussian distribution with a full width at half-maximum (fwhm) of 2.38 pixels. The fwhm can, in turn, be derived from the numerical aperture (NA) of the objective (1.4), the virtual pixel size of the microscope (103 nm), and the emission wavelength (590 nm) of the fluorophore used for our labeling. The simulation of SIM fragments was performed with a pixel size that is half the value that is available on our optical system and the same fwhm to approximate a two-fold increase in resolution.

To simulate random genome traces, 1000 optical maps 80 kilobase pairs in length were generated. The signal patterns were constructed by placing labels uniformly throughout the fragment so that their total density would be equal to 3 labels/kbp, a value typical for a bacterial genome. These were then converted into SIM and WF optical maps, respectively.

Bacterial Strains and SCC*mec* Types. The various strains and SCC*mec* investigated are shown in Table 1. The ATCC strains 43300 and 12600 were commercially obtained, and the sequence files were equally obtained from ATCC upon the purchase of the strains. Consequently, the sequence files are unavailable from a public repository and have no NCBI accession code.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c05902.

> Roughly annotated Mauve alignment of SCCmec Types I−IV, biochemistry of the labeling reaction by the methyltransferase M.Taq and the synthetic SAM analog carrying Rhodamine B dye, a typical single field of view from a tile-scan, Mauve alignment of the MRSA strain 11819−97 and Type IV SCCmec, which is the cassette present as determined by SCCmec Finder, Mauve alignment of ATCC 43300 and Type II SCCmec, which is the ground truth as determined by SCCmec Finder, coverage achieved by Fluorocode OM of the two data sets: ATCC 12600 [above] and ATCC 43300 [below], and coverage achieved by Fluorocode OM of the spiked-in *S. aureus* strain ATCC 43300 in the three data sets of bacterial mixtures: 10% spike-in [above], 0.1% spike-in [middle], and 0.01% spike-in [below] (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Johan Hofkens − *Chemistry, KU Leuven Faculty of Science, Leuven, Flanders 3001, Belgium; Max Planck Institute for Polymer Research, Mainz 55128 Rheinland-Pfalz, Germany;* orcid.org/0000-0002-9101-0567; Email: johan.hofkens@kuleuven.be

Authors

Elizabete Ruppeka-Rupeika − *Chemistry, KU Leuven Faculty of Science, Leuven, Flanders 3001, Belgium;* orcid.org/0000-0002-7093-8147

Sergey Abakumov − *Chemistry, KU Leuven Faculty of Science, Leuven, Flanders 3001, Belgium*

Mattias Engelbrecht − *Perseus Biomics B.V., Tienen 3300, Belgium*

Xiong Chen − *Chemistry, KU Leuven Faculty of Science, Leuven, Flanders 3001, Belgium*

Debora do Carmo Linhares − *Chemistry, KU Leuven Faculty of Science, Leuven, Flanders 3001, Belgium;* orcid.org/0000-0002-9055-9187

Arno Bouwens − *Perseus Biomics B.V., Tienen 3300, Belgium*

Volker Leen − *Perseus Biomics B.V., Tienen 3300, Belgium*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c05902

Notes

The authors declare the following competing financial interest(s): Arno Bouwens and Mattias Engelbrecht are employees of Perseus Biomics. Johan Hofkens and Volker Leen are co-founders of Perseus Biomics.

Figures were made using Biorender 2023, R and R Studio, Mauve and InkScape.

■ ACKNOWLEDGMENTS

J.H. acknowledges financial support from the FWO (Fonds voor Wetenschappelijk Onderzoek, grant number G0C1821N), the Flemish Government through long-term structural funding Methusalem (CASAS2, Meth/15/04), and the Max Planck Institute through an MPI fellowship. The authors would like to thank Prof Katrien Lagrou and Prof Jacques Schrenzel for their advice and help.

■ REFERENCES

(1) Kluytmans, J.; van Belkum, A.; Verbrugh, H. Nasal carriage of Staphylococcus aureus: Epidemiology, underlying mechanisms, and associated risks. *Clin. Microbiol. Rev.* **1997**, *10* (3), 505−520.

(2) Wertheim, H. F.; Melles, D. C.; Vos, M. C.; van Leeuwen, W.; van Belkum, A.; Verbrugh, H. A.; Nouwen, J. L. The role of nasal carriage in Staphylococcus aureus infections. *Lancet Infectious Diseases* **2005**, *5* (12), 751−762.

(3) Jeon, Y. J.; Gil, C. H.; Won, J.; Jo, A.; Kim, H. J. Symbiotic microbiome Staphylococcus aureus from human nasal mucus modulates IL-33-mediated type 2 immune responses in allergic nasal mucosa. *BMC Microbiology* **2020**, *20* (1), 301.

(4) Siddiqui, A. H.; Koirala, J. *Methicillin-Resistant Staphylococcus aureus*; StatPearls Publishing. In StatPearls, 2023. http://www.ncbi.nlm.nih.gov/books/NBK482221/.

(5) Murray, C. J. L.; Ikuta, K. S.; Sharara, F.; Swetschinski, L.; Robles Aguilar, G.; Gray, A.; Han, C.; Bisignano, C.; Rao, P.; Wool, E.; Johnson, S. C.; Browne, A. J.; Chipeta, M. G.; Fell, F.; Hackett, S.; Haines-Woodhouse, G.; Kashef Hamadani, B. H.; Kumaran, E. A. P.; McManigal, B.; Naghavi, M. Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *Lancet* **2022**, *399* (10325), 629−655.

(6) Jevons, M. P. Celbenin"—Resistant Staphylococci. *British Medical Journal* **1961**, *1* (5219), 124−125.

(7) Katayama, Y.; Ito, T.; Hiramatsu, K. A New Class of Genetic Element, Staphylococcus Cassette Chromosome mec, Encodes Methicillin Resistance in Staphylococcus aureus. *Antimicrob. Agents Chemother.* **2000**, *44* (6), 1549−1555.

(8) Cong, Y.; Yang, S.; Rao, X. Vancomycin resistant Staphylococcus aureus infections: A review of case updating and clinical features. *Journal of Advanced Research* **2020**, *21*, 169−176.

(9) Baig, S.; Johannesen, T. B.; Overballe-Petersen, S.; Larsen, J.; Larsen, A. R.; Stegger, M. Novel SCC mec type XIII (9A) identified

in an ST152 methicillin-resistant Staphylococcus aureus. *Infection, Genetics and Evolution* **2018**, *61*, 74−76.
(10) Ito, T.; Katayama, Y.; Asada, K.; Mori, N.; Tsutsumimoto, K.; Tiensasitorn, C.; Hiramatsu, K. Structural Comparison of Three Types of Staphylococcal Cassette Chromosome mec Integrated in the Chromosome in Methicillin-Resistant Staphylococcus aureus. *Antimicrob. Agents Chemother.* **2001**, *45* (5), 1323−1336.
(11) Heusser, R.; Ender, M.; Berger-Bächi, B.; McCallum, N. Mosaic Staphylococcal Cassette Chromosome mec Containing Two Recombinase Loci and a New mec Complex, B2. *Antimicrob. Agents Chemother.* **2007**, *51* (1), 390−393.
(12) Hanssen, A.-M.; Ericson Sollid, J. U. SCCmec in staphylococci: Genes on the move. *FEMS Immunology & Medical Microbiology* **2006**, *46* (1), 8−20.
(13) Hiramatsu, K.; Katayama, Y.; Yuzawa, H.; Ito, T. Molecular genetics of methicillin-resistant Staphylococcus aureus. *International Journal of Medical Microbiology* **2002**, *292* (2), 67−74.
(14) Bal, A. M.; Coombs, G. W.; Holden, M. T. G.; Lindsay, J. A.; Nimmo, G. R.; Tattevin, P.; Skov, R. L. Genomic insights into the emergence and spread of international clones of healthcare-, community- and livestock-associated meticillin-resistant Staphylococcus aureus: Blurring of the traditional definitions. *Journal of Global Antimicrobial Resistance* **2016**, *6*, 95−101.
(15) Miller, L. G.; Diep, B. A. Colonization, Fomites, and Virulence: Rethinking the Pathogenesis of Community-Associated Methicillin-Resistant Staphylococcus aureus Infection. *Clinical Infectious Diseases* **2008**, *46* (5), 752−760.
(16) Okuma, K.; Iwakawa, K.; Turnidge, J. D.; Grubb, W. B.; Bell, J. M.; O'Brien, F. G.; Coombs, G. W.; Pearman, J. W.; Tenover, F. C.; Kapi, M.; Tiensasitorn, C.; Ito, T.; Hiramatsu, K. Dissemination of New Methicillin-Resistant Staphylococcus aureus Clones in the Community. *Journal of Clinical Microbiology* **2002**, *40* (11), 4289−4294.
(17) Hiramatsu, K.; Ito, T.; Tsubakishita, S.; Sasaki, T.; Takeuchi, F.; Morimoto, Y.; Katayama, Y.; Matsuo, M.; Kuwahara-Arai, K.; Hishinuma, T.; Baba, T. Genomic Basis for Methicillin Resistance in Staphylococcus aureus. *Infection & Chemotherapy* **2013**, *45* (2), 117−136.
(18) Kim, C.; Milheiriço, C.; Gardete, S.; Holmes, M. A.; Holden, M. T. G.; de Lencastre, H.; Tomasz, A. Properties of a Novel PBP2A Protein Homolog from Staphylococcus aureus Strain LGA251 and Its Contribution to the β-Lactam-resistant Phenotype. *J. Biol. Chem.* **2012**, *287* (44), 36854−36863.
(19) Deurenberg, R. H.; Stobberingh, E. E. The evolution of Staphylococcus aureus. *Infection, Genetics and Evolution* **2008**, *8* (6), 747−763.
(20) Uehara, Y. Current Status of Staphylococcal Cassette Chromosome mec (SCCmec). *Antibiotics* **2022**, *11* (1), 86.
(21) Teruyu, I. Classification of Staphylococcal Cassette Chromosome mec (SCCmec): Guidelines for Reporting Novel SCCmec Elements. *Antimicrob. Agents Chemother.* **2009**, *53* (12), 4961−4967.
(22) Nyblom, M.; Johnning, A.; Frykholm, K.; Wrande, M.; Müller, V.; Goyal, G.; Robertsson, M.; Dvirnas, A.; Sewunet, T.; Kk, S.; Ambjörnsson, T.; Giske, C. G.; Sandegren, L.; Kristiansson, E.; Westerlund, F. Strain-level bacterial typing directly from patient samples using optical DNA mapping. *Communications Medicine* **2023**, *3* (1), 1.
(23) Müller, V.; Nyblom, M.; Johnning, A.; Wrande, M.; Dvirnas, A.; Sriram, K. K.; Giske, C. G.; Ambjörnsson, T.; Sandegren, L.; Kristiansson, E.; Westerlund, E. Cultivation-Free Typing of Bacteria Using Optical DNA Mapping. *ACS Infect. Dis.* **2020**, *6* (5), 1076−1084.
(24) Bikkarolla, S. K.; Nordberg, V.; Rajer, F.; Müller, V.; Kabir, M. H.; Sriram, K. K.; Dvirnas, A.; Ambjörnsson, T.; Giske, C. G.; Navér, L.; Sandegren, L.; Westerlund, F. Optical DNA Mapping Combined with Cas9-Targeted Resistance Gene Identification for Rapid Tracking of Resistance Plasmids in a Neonatal Intensive Care Unit Outbreak. *MBio* **2019**, *10* (4), No. e00347.
(25) Müller, V.; Nyblom, M.; Johnning, A.; Wrande, M.; Dvirnas, A.; S, K. K.; Giske, C. G.; Ambjörnsson, T.; Sandegren, L.; Kristiansson, E.; Westerlund, F. Cultivation-Free Typing of Bacteria Using Optical DNA Mapping. *ACS Infect. Dis.* **2020**, *6* (5), 1076−1084.
(26) Bouwens, A.; Deen, J.; Vitale, R.; D'Huys, L.; Goyvaerts, V.; Descloux, A.; Borrenberghs, D.; Grussmayer, K.; Lukes, T.; Camacho, R.; Su, J.; Ruckebusch, C.; Lasser, T.; Van De Ville, D.; Hofkens, J.; Radenovic, A.; Frans Janssen, K. P. Identifying microbial species by single-molecule DNA optical mapping and resampling statistics. *NAR Genomics and Bioinformatics* **2020**, *2* (1), No. lqz007.
(27) D'Huys, L.; Vitale, R.; Ruppeka-Rupeika, E.; Goyvaerts, V.; Ruckebusch, C.; Hofkens, J. Assessing the Resolution of Methyltransferase-Mediated DNA Optical Mapping. *ACS Omega* **2021**, *6* (33), 21276−21283.
(28) McClelland, M. Purification and characterization of two new modification methylases: MClaI from Caryophanon latum L and MTaqI from Thermus aquaticus YTI. *Nucleic Acids Res.* **1981**, *9* (24), 6795−6804.
(29) Goyvaerts, V.; Snick, S. V.; D'Huys, L.; Vitale, R.; Lauer, M. H.; Wang, S.; Leen, V.; Dehaen, W.; Hofkens, J. Fluorescent SAM analogues for methyltransferase based DNA labeling. *Chem. Commun.* **2020**, *56* (22), 3317−3320.
(30) Goedecke, K.; Pignot, M.; Goody, R. S.; Scheidig, A. J.; Weinhold, E. Structure of the N6-adenine DNA methyltransferase M•TaqI in complex with DNA and a cofactor analog. *Nat. Struct. Biol.* **2001**, *8* (2), 121.
(31) Gade, M.; Gardner, J. M.; Jain, P.; Laurino, P. Nucleoside-Driven Specificity of DNA Methyltransferase**. *ChemBioChem* **2017**, *24*, No. e202300094.
(32) Lauer, M. H.; Vranken, C.; Deen, J.; Frederickx, W.; Vanderlinden, W.; Wand, N.; Leen, V.; Gehlen, M. H.; Hofkens, J.; Neely, R. K. Methyltransferase-directed covalent coupling of fluorophores to DNA. *Chemical Science* **2017**, *8* (5), 3804−3811.
(33) Deen, J.; Sempels, W.; De Dier, R.; Vermant, J.; Dedecker, P.; Hofkens, J.; Neely, R. K. Combing of Genomic DNA from Droplets Containing Picograms of Material. *ACS Nano* **2015**, *9* (1), 809−816.
(34) Vranken, C.; Deen, J.; Dirix, L.; Stakenborg, T.; Dehaen, W.; Leen, V.; Hofkens, J.; Neely, R. K. Super-resolution optical DNA Mapping via DNA methyltransferase-directed click chemistry. *Nucleic Acids Res.* **2014**, *42* (7), No. e50.
(35) *Local Normalization.*(n.d.). Retrieved October 29, 2023, from https://bigwww.epfl.ch/sage/soft/localnormalization/.
(36) Kaya, H.; Hasman, H.; Larsen, J.; Stegger, M.; Johannesen, T. B.; Allesøe, R. L.; Lemvigh, C. K.; Aarestrup, F. M.; Lund, O.; Larsen, A. R. SCCmecFinder, a Web-Based Tool for Typing of Staphylococcal Cassette Chromosome mec in Staphylococcus aureus Using Whole-Genome Sequence Data. *MSphere* **2018**, *3*, No. e00612.
(37) Darling, A. C. E.; Mau, B.; Blattner, F. R.; Perna, N. T. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res.* **2004**, *14* (7), 1394−1403.
(38) Pfeiffer, F.; Gröber, C.; Blank, M.; Händler, K.; Beyer, M.; Schultze, J. L.; Mayer, G. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Sci. Rep.* **2018**, *8* (1), 1.
(39) Zhou, Z.; Ma, M.-J. L.; Chan, R. W. Y.; Lam, W. K. J.; Peng, W.; Gai, W.; Hu, X.; Ding, S. C.; Ji, L.; Zhou, Q.; Cheung, P. P. H.; Yu, S. C. Y.; Teoh, J. Y. C.; Szeto, C.-C.; Wong, J.; Wong, V. W. S.; Wong, G. L. H.; Chan, S. L.; Hui, E. P.; Jiang, P. Fragmentation landscape of cell-free DNA revealed by deconvolutional analysis of end motifs. *Proc. Natl. Acad. Sci. U. S. A.* **2023**, *120* (17), No. e2220982120.
(40) Thys, R. G.; Lehman, C. E.; Pierce, L. C. T.; Wang, Y.-H. DNA Secondary Structure at Chromosomal Fragile Sites in Human Disease. *Curr. Genomics* **2015**, *16* (1), 60−70.
(41) Xie, K. T.; Wang, G.; Thompson, A. C.; Wucherpfennig, J. I.; Reimchen, T. E.; MacColl, A. D. C.; Schluter, D.; Bell, M. A.; Vasquez, K. M.; Kingsley, D. M. DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science* **2019**, *363* (6422), 81−84.

https://doi.org/10.1021/acsomega.3c05902
ACS Omega 2024, 9, 8862−8873

(42) Li, Y.; Cao, B.; Zhang, Y.; Zhou, J.; Yang, B.; Wang, L. Complete Genome Sequence of Staphylococcus aureus T0131, an ST239-MRSA-SCCmec Type III Clone Isolated in China. *J. Bacteriol.* **2011**, *193* (13), 3411−3412.
(43) Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Aoki, K.; Nagai, Y.; Lian, J.; Ito, T.; Kanamori, M.; Matsumaru, H.; Maruyama, A.; Murakami, H.; Hosoyama, A.; Mizutani-Ui, Y.; Takahashi, N. K.; Hiramatsu, K. Whole genome sequencing of meticillin-resistant Staphylococcus aureus. *Lancet* **2001**, *357* (9264), 1225−1240.
(44) Tenover, F. C.; Goering, R. V. Methicillin-resistant Staphylococcus aureus strain USA300: Origin and epidemiology. *J. Antimicrob. Chemother.* **2009**, *64* (3), 441−446.
(45) Stegger, M.; Price, L. B.; Larsen, A. R.; Gillece, J. D.; Waters, A. E.; Skov, R.; Andersen, P. S. Genome Sequence of Staphylococcus aureus Strain 11819−97, an ST80-IV European Community-Acquired Methicillin-Resistant Isolate. *J. Bacteriol.* **2012**, *194* (6), 1625−1626.
(46) Bacteria detail. *Culture Collections*. Retrieved October 19, 2023, from: https://www.culturecollections.org.uk/products/bacteria/detail.jsp?refId=NCTC+10442&collection=nctc.
(47) Staphylococcus aureus genome assembly ASM305244v1. *NCBI*. Retrieved October 25, 2023, from: https://www.ncbi.nlm.nih.gov/data-hub/assembly/GCF_003052445.1/.
(48) Bacteria detail. *Culture Collections*. Retrieved October 25, 2023, from: https://www.culturecollections.org.uk/products/bacteria/detail.jsp?refId=NCTC+8532&collection=nctc.
(49) Bacteria detail. *Culture Collections*. Retrieved October 23, 2023, from: https://www.culturecollections.org.uk/products/bacteria/detail.jsp?refId=NCTC+8325&collection=nctc.

https://doi.org/10.1021/acsomega.3c05902
ACS Omega 2024, 9, 8862−8873