

Shotgun-Metagenomics Based Prediction Of Antibiotic Resistance And Virulence Determinants In *Staphylococcus Aureus* From Prosthetic Joint Tissue On Blood Culture Bottles

Adriana Maria Sanabria

UiT – The Arctic University of Norway

Jessin Janice

UiT – The Arctic University of Norway

Erik Hjerde

UiT – The Arctic University of Norway

Gunnar Skov Simonsen

UiT – The Arctic University of Norway

Anne-Merethe Hanssen (✉ anne-merethe.hanssen@uit.no)

UiT – The Arctic University of Norway

Research Article

Keywords: antimicrobial, virulence, bioinformatic, potential pathogen

Posted Date: June 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-595378/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Shotgun-metagenomics may give valuable clinical information beyond the detection of potential pathogen(s). Identification of antimicrobial resistance (AMR), virulence genes and typing directly from clinical samples has been limited due to challenges arising from incomplete genome coverage. We assessed the performance of shotgun-metagenomics on positive blood culture bottles (n = 19) with prosthetic joint tissue for typing and prediction of AMR and virulence profiles in *Staphylococcus aureus*. We used different approaches to determine if sequence data from reads provides more information than from assembled contigs. Only 0.18% of total reads was derived from human DNA. Shotgun-metagenomics results and conventional method results were consistent in detecting *S. aureus* in all samples. AMR and known prosthetic joint infection virulence genes were predicted from *S. aureus*. Mean coverage depth, when predicting AMR genes was 209x. Resistance phenotypes could be explained by genes predicted in the sample in most of the cases. The choice of bioinformatic data analysis had a significant impact on the results. Read-based analysis was more accurate for pathogen identification, while contigs seemed better for AMR profiling. Our study demonstrates high genome coverage and potential for typing and prediction of AMR and virulence profiles in *S. aureus* from shotgun-metagenomics data.

Introduction

Staphylococcus aureus is an important opportunistic pathogen considered as the most common cause of prosthetic joint infections (PJIs)¹⁻⁵. The emergence and spread of resistance pose an increasing threat to public health, in particular, methicillin-resistant *S. aureus* (MRSA)⁶. The success of *S. aureus* as a pathogen is in part due to its ability to develop resistance to a wide variety of antimicrobial compounds. Additionally, *S. aureus* can adapt to a biofilm mode of growth whereby infections become persistent and recurrent, particularly in association with prosthetic implants⁷.

Microbiological diagnosis of PJI is challenging. A variety of tools are available for facilitating the diagnosis of PJI, including emerging technologies such as metagenomic approaches⁸. The use of shotgun-metagenomics (SMg) for the analysis of clinical specimens has emerged as a promising approach for pathogen identification, antimicrobial resistance (AMR) identification and outbreak investigation in clinical microbiology laboratories. This approach has been used for the analysis of different types of clinical specimens, including samples related to PJI, e.g. synovial fluid^{9,10}, sonication fluid¹¹⁻¹⁴ and tissue¹⁵, mainly for the identification of pathogens. In a previous study, we showed that SMg performed directly on positive blood culture bottles (BCBs) inoculated with prosthetic joint tissue (PJT), is a convenient method to identify potential pathogens causing PJI¹⁶. However, beyond the identification of pathogens, SMg provides unlimited access to other clinically relevant genomic features such as antibiotic resistance, virulence genes profiles and strain-level typing^{17,18}.

Currently, SMg is considered in its infancy for pathogen characterization, including inference of antibiotic susceptibility¹⁹. Challenges arise due to the diversity of drug resistance mechanisms, multidrug resistance, and incomplete genome coverage, leading to insufficient sequence reads for detection of ARGs²⁰. However, there are some studies that show the potential of SMg for the detection of ARGs^{21–23} by comparing the genotype against the phenotype, or generating AMR profiles from SMg assemblies and comparing them with whole genome sequences (WGS) from isolates. The use of SMg on samples from bone and joint infections has been used where they could predict antibiotic susceptibility in 94.1% (monomicrobial) and 76.5 % (polymicrobial) of the cases¹⁵. However, in these studies, the main obstacle has been a high background of genetic material mainly derived from the host, which generates very few bacterial reads. Similarly, when using SMg data for subtyping bacteria, one of the main challenges is missing loci. This problem arises when coverage is too low to guarantee the presence of a read containing a given sequence in the targeted genome²⁴.

We previously showed that SMg on BCBs with PJT resulted in acceptable high number of bacterial reads, genome coverage and genome sequencing depth¹⁶. Here, we wanted to assess the potential of SMg for the identification and typing of the most common cause of PJI, *S. aureus*, and the prediction of virulence and AMR directly from clinical samples.

Results

Sequencing data

SMg sequencing from the 19 samples resulted in a mean number of 3,949,678 reads per clinical sample (range 2,608,766–8,086,037) and 5,942,038 reads in the sample spiked with *S. aureus*, (Supplementary Table S1). Samples contained a lower proportion of reads classified either as human or horse or PhiX, while 98 % of the reads did not map to any of the reference sequences used for the alignment (Supplementary Table S2 and Figure S1). After data preprocessing, a mean number of 3,700,731 reads remained for further taxonomical classification (Supplementary Table S3). Kraken taxonomically classified a mean proportion of 98.36 % reads, with 95.74 % bacterial reads (Fig. 1 and Supplementary Table S4). Assembly with metaSPAdes yielded a mean number of 232 contigs (range 134–378), with a mean total size of 3.1Mb (range 2.6Mb – 4.8Mb) in the clinical samples and 213 contigs for a total length of 2.7Mb in the sample spiked with *S. aureus* (Supplementary Table S5). The total number of base pairs was higher in polymicrobial samples than in monomicrobial ones (4.8Mb vs 2.7Mb, respectively, t-test and P-value < 0.0001). The mean of the “maximum contig size” was 262,574 bp (median 264,931 bp, maximum 425,306 bp) in the clinical samples and 218,856 bp in the spiked sample, and no significant difference was observed between polymicrobial and monomicrobial samples (184,461bp and 283,495 bp, respectively, t-test and P-value = 0.149). Binning with MaxBin in polymicrobial samples grouped a mean number of 43 contigs in the bin assigned to *S. aureus* (range 39–51) with a mean total of 2.6Mb (range 2.6Mb – 2.7Mb) and a mean maximum length 297,040 bp (234,035 bp – 381,826 bp).

Identification of *S. aureus* by SMg

The taxonomical classification was performed on both the sequence reads and assembled contigs. When the taxonomical classification from the reads identified multiple highly abundant species (polymicrobial samples), contigs were grouped by species into bins, and then used for taxonomical classification. Relative abundance from the most abundant species in the samples varied depending on the selected approach. The bin classified as *S. aureus* was used for downstream analyses for pathogen characterization.

S. aureus was identified in all the samples by SMg. *S. aureus* (19/19) and *S. agalactiae* (4/19) was identified from both the reads and contigs. *S. aureus* was the most abundant species identified by SMg, with exception of samples 7 and 9, where *S. agalactiae* was more abundant (86.4 % and 8.8 %, from the reads, respectively) (Table 1 and Supplementary Figure S1).

Table 1

Bacteria identified in the clinical samples and in the positive control (PC) by MALDI-TOF from blood culture bottles (BCBs) and shotgun-metagenomics (SMg) by the reads, contigs and bins approaches.

Sample ID	Patient No.	Microorganism(s) Identified				
		BCBs (MALDI-TOF)	Shotgun-metagenomics			
			Reads	Contigs	Bins	
No.	Taxonomy	No.	Taxonomy	No.	Taxonomy	
1	1	<i>S. aureus</i>	<i>S. aureus</i> (99.9 %)	<i>S. aureus</i> (100 %)		
2	2	<i>S. aureus</i>	<i>S. aureus</i> (99.9 %)	<i>S. aureus</i> (100 %)		
3	3	<i>S. aureus</i>	<i>S. aureus</i> (99.9 %)	<i>S. aureus</i> (100 %)		
4	4	<i>S. aureus</i>	<i>S. aureus</i> (99.3 %)	<i>S. aureus</i> (98.6 %)		
5	5	<i>S. aureus</i>	<i>S. aureus</i> (99.9 %)	<i>S. aureus</i> (100 %)		
6	6	<i>S. aureus</i>	<i>S. aureus</i> (90.9 %)	<i>S. aureus</i> (60.6 %)	1	<i>S. aureus</i> (100 %)
			<i>S. agalactiae</i> (8.5 %)	<i>S. agalactiae</i> (39.4 %)	2	<i>S. agalactiae</i> (100 %)
7	7	<i>S. aureus</i> <i>S. agalactiae</i>	<i>S. aureus</i> (10.7 %)	<i>S. aureus</i> (62.4 %)	1	<i>S. agalactiae</i> (97 %)
			<i>S. agalactiae</i> (86.4 %)	<i>S. agalactiae</i> (37.6 %)	2	<i>S. aureus</i> (2 %)
8	8	<i>S. aureus</i>	<i>S. aureus</i> (97.6 %)	<i>S. aureus</i> (51.5 %)	1	<i>S. aureus</i> (100 %)
			<i>S. agalactiae</i> (2.2 %)	<i>S. agalactiae</i> (48.2 %)	2	<i>S. agalactiae</i> (100 %)
9	9	<i>S. aureus</i>	<i>S. agalactiae</i> (85.8 %)	<i>S. agalactiae</i> (39.8 %)	1	<i>S. agalactiae</i> (94.8 %)
			<i>S. aureus</i> (11.4 %)	<i>S. aureus</i> (60.3 %)	2	<i>S. aureus</i> (100 %)
10	10	<i>S. aureus</i>	<i>S. aureus</i> (99.8 %)	<i>S. aureus</i> (100 %)		
11	11	<i>S. aureus</i>	<i>S. aureus</i> (98.2 %)	<i>S. aureus</i> (98.3 %)		

12	12	<i>S. aureus</i>	<i>S. aureus</i> (99.8 %)	<i>S. aureus</i> (100 %)
13	13	<i>S. aureus</i>	<i>S. aureus</i> (99.4 %)	<i>S. aureus</i> (100 %)
14	14	<i>S. aureus</i>	<i>S. aureus</i> (99.6 %)	<i>S. aureus</i> (100 %)
15	15	<i>S. aureus</i>	<i>S. aureus</i> (99.4 %)	<i>S. aureus</i> (100 %)
16	16	<i>S. aureus</i>	<i>S. aureus</i> (99.9 %)	<i>S. aureus</i> (99.0 %)
17	17	<i>S. aureus</i>	<i>S. aureus</i> (89.8 %)	<i>S. aureus</i> (99.2 %)
18	2	<i>S. aureus</i>	<i>S. aureus</i> (89.7 %)	<i>S. aureus</i> (97.6 %)
19	1	<i>S. aureus</i>	<i>S. aureus</i> (99.9 %)	<i>S. aureus</i> (84.7 %)
PC	NA	<i>S. aureus</i>	<i>S. aureus</i> (99.9 %)	<i>S. aureus</i> (95.1 %)

***S. aureus* antibiotic resistance determinants by SMg**

The presence of antibiotic resistance genes (ARGs) was determined by SMg from the reads and contigs for all the samples, and also from the bins classified as *S. aureus* in polymicrobial samples (Sample 6, 7, 8 and 9). The presence of ARGs found in the reads, contigs and bins using the Bacterial Antimicrobial Resistance Reference Gene Database from the National Center for Biotechnology Information (NCBI) as reference, were determined and compared with the results obtained by the phenotypic antimicrobial susceptibility testing (AST) (Table 2, Supplementary Table S6 and S7).

Table 2

Antibiotic resistance genes (ARGs) detected in this study using different approaches (reads, contigs and bins) with the NCBI Bacterial Antimicrobial Resistance Reference Gene Database and conventional antibiotic susceptibility testing (AST).

Sample No.	Conventional antibiotic susceptibility test	ARGs detected from shotgun-metagenomics		
		Reads	Contigs	Bins
1	Penicillin	<i>blaI, blaPC1, blaR1, tet38</i>	<i>fosB, blaPC1, blaR1, blaI, tet38</i>	
2	Penicillin	<i>blaZ, tet38</i>	<i>blaZ, blaR1, blaI, tet38</i>	
3		<i>tet38</i>	<i>tet38</i>	
4	Penicillin		<i>fosB, blaPC1, blaR1, blaI, tet38</i>	
5		<i>tet38</i>	<i>tet38</i>	
6	Penicillin		<i>tet38</i>	<i>tet38</i>
7	Penicillin	<i>tetM, tet38</i>	<i>fosB, tetM, tet38</i>	<i>fosB, tet38</i>
8		<i>tet38, tetM</i>	<i>tet38</i>	
9	Penicillin	<i>tet38, tetM</i>	<i>fosB, tetM, tet38</i>	<i>fosB, tet38</i>
10			<i>tet38</i>	
11			<i>fosB, tet38, blaZ, blaR1, blaI</i>	
12	Penicillin Fusidic acid		<i>fosB, blaR1, blaI, tet38, fusA, fusE</i>	
13			<i>blaZ, blaR1, blaI, tet38</i>	
14	Penicillin	<i>blaZ</i>	<i>fosB, blaZ, blaR1, blaI, tet38</i>	
15	Penicillin	<i>blaZ, blaR1, blaPC1, tet38</i>	<i>blaI, blaR1, blaPC1, tet38</i>	
16	Penicillin Fusidic acid	<i>blaZ, tet38</i>	<i>fosB, blaZ, blaR1, blaI, tet38, fusA, fusE</i>	
17			<i>tet38</i>	
18	Penicillin Fusidic acid		<i>blaZ, blaR1, blaI, tet38</i>	

Sample No.	Conventional antibiotic susceptibility test	ARGs detected from shotgun-metagenomics		
		Reads	Contigs	Bins
19	Penicillin Fusidic acid	<i>blaI, blaR1, tet38</i>	<i>fosB, blaI, blaR1, blaPC1, tet38</i>	
PC		<i>tet38</i>	<i>fosB, tet38</i>	

The mean coverage depth when predicting ARGs from the reads was 209x (Supplementary Table S6). We applied a threshold of a minimum 20x coverage depth, 90% sequence identity and 90 % sequence coverage for determining the presence of ARGs. In total, we were able to identify six different resistance genes in *S. aureus* (*tet38*, *blaZ*, *blaPC1*, *blaR1*, *blaI* and *fosB*) in the 20 samples (including the spiked sample) by SMg. A higher number of ARGs were detected in the samples using the contigs approach compared with the reads approach (64 and 25 predictions in total, respectively).

The gene *tet38* encoding the chromosomally encoded efflux pump of *S. aureus* was detected across all samples. The other genes detected, corresponded mainly to the *S. aureus* beta-lactamase operon; *blaZ* (35 %), *blaPC1* (25 %), *blaR1* (55 %) and *blaI* (55 %), and the fosfomycin resistance gene *fosB* (50 %). In the polymicrobial samples, the gene *tetM* was also detected, but it was not identified in the *S. aureus* bins.

Resistant phenotypes were observed only for penicillin in 12/19 samples (63.5 %), and for fusidic acid in 2/19 (10.5 %) samples (Table 2). The penicillin resistant phenotype could be explained by the presence of members of the *bla*-operon in most of the samples. However, in samples 6, 7 and 9, the penicillin resistance phenotype could not be explained by the genotypic profile generated from the reads, contigs or bin. Samples presenting a fusidic acid resistance phenotype were checked additionally for the presence of *fusA* or *fusE* genes using the CARD database.

Virulence factors

Overall, a total of 73 genes coding for virulence factors (VFs) were found by SMg in *S. aureus* (Supplementary Table S8) and a mean of 55 virulence genes (range, 50–62) were detected per sample (Fig. 2a). Toxins, adhesins and immune evasion molecules were among the genes predicted. Genes encoding 40 virulence factors were present in all the samples, e.g., the *cap8* capsule genes (*cap8A-G* and *cap8L-P*), iron sequestration operon *isdA-isdG*, and exotoxins *hla*, *hld*, *hlgA-C*, among others. Additionally, five samples contained the toxic shock syndrome toxin 1 (*tsst-1*) gene.

Virulence genes recognized as belonging to *S. agalactiae* were removed from the analysis of polymicrobial samples. Several virulence genes known or proposed to play a role for the pathogenicity of *S. aureus* in PJI were identified from the metagenomes (Fig. 2b and Table 3).

Table 3

Prevalence of virulence genes known or proposed to play a role in *S. aureus* pathogenicity in PJI predicted from SMg in this study.

Virulence gene	Product	%
<i>aur</i>	Zinc metalloproteinase aureolysin	50
<i>clfA</i>	Clumping factor A fibrinogen-binding protein	25
<i>cna</i>	Collagen adhesin precursor	10
<i>fnbA</i>	Fibronectin-binding protein A	25
<i>hld</i>	Delta-hemolysin	100
<i>hlgA</i>	Gamma-hemolysin chain II precursor	100
<i>hlgB</i>	Beta-hemolysin	100
<i>hlgC</i>	Gamma-hemolysin component C	100
<i>sdrD</i>	Ser-Asp rich fibrinogen-binding bone sialoprotein-binding protein	25
<i>sdrE</i>	Ser-Asp rich fibrinogen-binding bone sialoprotein-binding protein	25
<i>spa</i>	Immunoglobulin G binding protein A precursor	60
<i>sspA</i>	Serine protease; V8 protease; glutamyl endopeptidase	100
<i>sspB</i>	Staphopain cysteine proteinase SspB	100
<i>sspC</i>	Staphostatin B	100
<i>hly/hla</i>	Alpha-hemolysin precursor	100

MLST and cgMLST analysis

The *S. aureus* Multilocus sequence types (ST) were identified for all the samples at both core genome and whole genome (core genome + accessory genome genes) level (Fig. 3 and Supplementary Table S9). Typing from SMg data showed that *S. aureus* in our samples are of different lineages. *S. aureus* in the samples represent six different clonal complexes (CCs), and they belonged to different STs (ST45 (30%), ST121 (15%), ST30 (15%), ST22 (10%), ST5 (10%), ST15 (5%), ST243 (5%), ST7 (5%) and ST97 (5%)). With respect to polymicrobial samples 7 and 9, typing analyses from both the contigs and the bins classified *S. aureus* as belonging to the same sequence type (ST5). CCs could not be identified for six samples (23%). Samples belonging to the same patient (sample 1 and 18; sample 2 and 19) did not cluster together and did not belong to the same ST (Fig. 3 and Supplementary Table S9). The *S. aureus* isolates analyzed here represent eight phylogenetically diverse STs. cgMLST subdivided the samples into 13 different complex types (CT, 21308–21321) (Supplementary Table S9).

Discussion

Here, we assessed the performance of SMg for the prediction of ARGs, virulence gene determinants and typing of *S. aureus* from clinical PJT samples on BCBs. We investigated if there was a difference in outcome from analyzing sequencing data from reads and contigs, and our data analysis followed different analytical approaches in order to identify the procedures that may give the most relevant and accurate information from our SMg data.

It is established that it is possible to analyze sequence data without assembly, but most analyses can be improved by constructing longer contiguous sequences (contigs) through assembly processes²⁵. According to our results, the selection of approach to some degree depends on what type of information you require from the data. For instance, if the aim is pathogen identification, taxonomy from the reads is sufficient while if AMR is the focus, the contigs approach will provide a more comprehensive resistome resolution.

For identification of potential pathogen(s) causing PJI, we found that the relative abundance of the taxa was influenced by the approach used to analyze the data. This was evidenced by the differences in abundance obtained, especially in the polymicrobial samples. We observed that the proportion of contigs classified into a certain taxonomical level is also influenced by the genome size. Determining the taxa present and the relative level or number of cells of one taxon vs another (e.g., polymicrobial samples) in a clinical sample is important for identifying the pathogen(s) causing the infection. Our results suggest that analyzing reads provides a more trustworthy representation of the species in a clinical sample than analyzing contigs. Moreover, it is considered that the reads will describe more accurately the proper distribution of species in the sample since each read belongs to one species and their length size is uniformly distributed. In addition, it could also be errors in joining contigs from two closely related species. Specificity is lost when working with contigs as the quality of the assembly will depend strongly on the length and quality of the reads, sometimes misrepresenting the original sample²⁶.

S. agalactiae has been reported as one of the most common pathogens found together with *S. aureus* in polymicrobial PJIs²⁷. *S. agalactiae* was detected by SMg in four of the samples, but not by the laboratory method (BCBs). This could suggest an increased sensitivity of SMg to detect bacteria, or alternatively a contamination problem in the laboratory workflow.

S. aureus has been studied extensively with a special focus on resistance and virulence. In this study, only resistance to penicillin (63.5 %) and fusidic acid (10.5 %) were observed by AST. No MRSA isolates were detected, neither by phenotypical nor genotypical testing. This result is not surprising, since Norway has a very low prevalence of MRSA²⁸.

Although SMg is a promising approach, the *in silico* translation from genotype into phenotype relies on the knowledge about the genomic resistance determinants²⁹. Our SMg analysis allowed the prediction of the genotypical resistance profile from *S. aureus* present in the samples. The ARG *tet38* was detected in all samples. However, tetracycline resistant phenotypes were not observed, and the presence of *tet38* is not enough to produce a resistant phenotype. The *tet38* determinant is an inherent, chromosomally

encoded efflux pump in *S. aureus* and resistance to tetracyclines is often associated with plasmid-mediated genes encoding active efflux pumps or proteins that protect ribosomes from drug action³⁰.

In three of the metagenomic samples with phenotypic penicillin resistance, the *blaZ* gene was absent (25 %). This may be explained by the variable location of the *blaZ* gene on a plasmid or integrated into the chromosome. Isolates with chromosomally integrated *blaZ* are likely to have average coverage in the sequencing reads, while in the isolates with plasmid-carried copies, the coverage could vary from very high (if multiple copies are carried) or very low coverage on that region because of poor mapping to the reference³¹.

Samples with phenotype-genotype disagreement were found to be polymicrobial by SMg (samples 6, 7 and 9). In samples 7 and 9, *S. aureus* was not the most abundant species present in the sample (< 12 %) which may affect the prediction of ARGs by a lower genome coverage. Bacterial isolates with plasmid-encoded copies may have very high (if multiple copies are carried) or very low (because of poor mapping to the reference) coverage in that region³². As a result, these regions may be rejected as low genome coverage when predicted from the reads, or as poor quality by the assembly tools when predicted from the contigs and bins since they fall outside the coverage level of the rest of the genome. This problem may be overcome in the future with long-read sequencing methods such as nanopore sequencing or alternative methods for *de novo* assembly.

Prediction of ARGs was done at the reads, contigs and bin level. The total number of different genes detected (7 ARGs) was influenced by the parameters selected to report a gene as present, as these parameters constitute a trade-off between specificity and depth³³. We used strict parameters, and only ARGs that had ≥ 90 % similarity and coverage to that of the reference were reported from the reads, contigs and bins. In addition, at the read level, only ARGs with at least 20x coverage depth were considered as present. The selection of ARGs using stringent cutoffs (≥ 90 % per read or contig) can increase the probability of targeting genes that are actually functional³³. We consider that the high genome depth coverage (> 200x) is an advantage in our SMg approach.

Most of the tools developed for identifying ARGs from metagenomic reads can detect acquired ARGs, but are not able to identify point mutations associated with AMR³⁴. The focus here was on acquired resistance since we have used the tools Groot or ABRicate with the NCBI Bacterial Antimicrobial Resistance Reference Gene Database. In our study, more genes were detected at the contig level than at the read level. The detection of more ARGs from the contig-based approach may be explained by the fact that it is easier to reach a high coverage of the gene (90 %) from contigs (which are longer) than by reads that are shorter in length. We consider that prediction of ARGs at the contig level is the best approach when looking for ARGs. However, it is important to highlight that both approaches are valid for certain purposes, and both have their limitations. Read-based prediction of ARGs provides an advantage when dealing with metagenomic samples, as ARGs in less abundant organisms can be predicted despite low coverage, which may be missed by assembly-based methods owing to incomplete or poor assemblies³⁵. Detecting ARGs from reads is more prone to false positive results because of sequencing errors present in

single reads or from DNA contamination from other bacteria. A previous study comparing ARG detection from reads and contigs suggested using both approaches when the contig coverage is set to a high percentage, since it is possible that ARGs may be separated into different contigs when the coverage depth is too low during the assembly process²¹. The use of long-read sequencing can overcome this problem.

Strategies for predicting AMR phenotypes in polymicrobial samples present an interesting challenge³⁶. We tested the binning approach for the prediction of ARGs in *S. aureus* and we found that it gives similar results as prediction of ARGs in monomicrobial samples for most of the genes with exception of the *blaZ* gene. This approach allowed us to separate the contigs belonging to *S. agalactiae* from the contigs belonging to *S. aureus* and predict the *S. aureus* resistance profile, even though there were not many contigs (mean, 44,3 contigs; range 39–51). This means that AMR genotype predictions could be made from contigs that are binned in a metagenomic assembly, even when they belong to a species that is not in a high abundance. In this study, the binning strategy was no further evaluated. Conclusions from bins should be made with precaution since the binning process can lead to incorrect assumptions due to misbinning (the wrong assignment of a genome fragment from one organism to another), namely if the abundance of the species is very similar, which may lead to neglect of specific determinants³⁷.

In our study, we identified some VFs that are known or proposed to play a role in *S. aureus* PJI, e.g. genes involved in colonization and attachment of host tissue or implanted biomaterials such as the adhesins *clfA* and *fnbA* that encode the fibrinogen and fibronectin-binding proteins, respectively³⁸.

MLST was used for strain level typing including a contig-based approach, which means that we had sufficient depth for assembly from the metagenome. wgMLST demonstrated that *S. aureus* in the samples consisted of several lineages. Our MLST results were in accordance with results from the population-based Tromsø Staph and Skin Study, showing that CC30, CC45 and CC15 were the most common CCs in MSSA³⁹. Additionally, the most common *S. aureus* lineages in PJI reported in a recent study, were CC30, CC45, CC5, CC15 and CC22. *S. aureus* in our study belonged to the same CCs^{27,39}.

Our study has several limitations. First, the sample size was small (n = 19), and we only analyzed a limited number of polymicrobial samples. Second, *S. aureus* isolates were not whole genome sequenced for comparison and confirmation of ARG and VF profiles. Third, no clinical data about the patients were obtained making it difficult to classify the samples as true PJI. Fourth, we have used short-read sequencing which makes detection of ARGs on mobile genetic elements difficult. Long-read sequencing may overcome this limitation. Fifth, we only predicted acquired resistance, which made it difficult explaining the disagreement found between the phenotype and the genotype for penicillin resistance. Errors in sensitivity and specificity of ARG prediction can have different consequences for PJI treatment. False negatives (phenotypically resistant and SMg-susceptible) can lead to inadequate treatment of a resistant infection, increasing morbidity and mortality, whereas false positives (phenotypically susceptible and SMg-resistant) may lead to inappropriate antibiotic use and increase the risk of resistance development⁴⁰.

In conclusion, this study showed that SMg from BCBs inoculated with PJT, allowed the identification of potential PJI pathogens and strain-level typing of *S. aureus*. We obtained *S. aureus* ARGs and virulence gene profiles from both monomicrobial and polymicrobial samples. However, the use of this approach for the detection of AMR to help guide clinical treatment needs to be further elucidated, due to the disagreement between the AMR phenotype and genotype. We conclude that the approach chosen for analyzing SMg data (reads, contigs or metagenomic assembled genomes) will have a key impact on the results. Precise AMR prediction is required for mainstream adoption of SMg into the clinical microbiology laboratory. Thus, several improvements are needed for AMR prediction using SMg, including a better understanding of the mechanisms underlying AMR and the procedures (including workflows, tools and databases) that may give the most relevant and accurate information when analyzing SMg data.

Methods

Ethics statement

This study was performed in accordance with the ethical guidelines established by UiT - The Arctic University of Norway. The project has been evaluated by the Regional Committee for Medical and Health Research Ethics, REC North, Norway (document no. 2016/1247/REK nord), concluding that ethical approval was not required. According to the Norwegian guidelines, informed consent of the patients is not needed and there were not ethical issues to consider due to the use of anonymous clinical samples and the development of methodological procedures.

Sample collection

Nineteen positive BCBs inoculated with PJT from 17 individual patients with suspicion of PJI were used in this study. Clinical samples were selected on the basis of being positive for *S. aureus* by the BCB method¹⁶, either monomicrobial (n = 18) or polymicrobial (n = 1). Samples were collected continuously over a 28-month period (August 2017–December 2019). Samples were anonymized and de-identified. All samples were taken from aerobic bottles (Bact/Alert® FA plus bottles, bioMérieux, Marcy l'Etoile, France).

Five of the 19 clinical samples in this study, in addition to one positive control (spiked sample, BCB inoculated with tissue spiked with *S. aureus* ATCC 2592) were obtained from a sample collection in a previous study^{16,41}. For further details on the BCB sample preparation method, see Sanabria et al., 2019⁴¹. Bacterial identification was performed using matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF® MS Bruker Daltonics - microflex™). An overview of all the samples included in this study, is shown in Table 1.

Antibiotic susceptibility testing

Antibiotic susceptibility testing (AST) of *S. aureus* isolates was performed by disc diffusion test according to EUCAST guidelines and the breakpoint table v.10.0 (2020)⁴². The antibiotics tested were penicillin (10 µg), trimethoprim-sulfamethoxazole (25µg), ceftioxin (30µg), fusidic acid (10µg),

clindamycin (2µg), erythromycin (15µg), linezolid (10µg), tetracycline (30µg), norfloxacin (10µg), ciprofloxacin (5µg), gentamicin (10µg) and rifampicin (5µg) (Oxoid, Basingstoke, UK).

DNA preparation

DNA was extracted and processed as previously described¹⁶. In short, all samples were pre-treated using MoLYsis™ Basic5 kit (Molzym, Bremen, Germany) to deplete human and horse DNA before DNA extraction using the QIAamp BiOstic Bacteremia DNA Kit (Qiagen, Hilden, Germany). Total DNA concentration was measured using a Qubit dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA, USA) and DNA quality was determined by Nanodrop.

Metagenomic Sequencing

Sequencing libraries were prepared as previously described¹⁶, using the ThruPLEX® DNA-seq Kit (Rubicon Genomics, USA) following the manufacturer's instructions. Approximately 100 ng of DNA was used as input for library preparation from the clinical and spiked samples. The sequencing process was performed at the Norwegian Sequencing Centre, Oslo, using a MiSeq sequencer (Illumina Inc., San Diego, CA, USA) with v2 chemistry and 500 cycles for 250 bp paired-end sequencing. Samples were multiplexed with four samples per lane.

Bioinformatic data analysis

The bioinformatic analysis followed in this study is summarized in Fig. 4. The quality of the raw reads in fastq format was assessed using FastQC software v0.11.8 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Optical duplicates were removed using the program Clumpify v38.36 from BBTools suite (<https://jgi.doe.gov/data-and-tools/bbtools/>) with default parameters. Adapter sequences were trimmed off and the poor-quality reads were removed using BBDuk of BBTools suite. The minimal read length and Phred score were set to 50 nucleotides and 20, respectively. In order to filter out known sources of contaminant host DNA, the reads mapped against the reference genomes of human GRCh38.p13 (GCF_000001405.39), horse (GCF_002863925.1) and the PhiX phage (Escherichia virus phiX174, GCF_000819615.1) aligning with Bowtie2 in the tool FastQ Screen v0.13.0⁴³. Unmapped reads were used in subsequent analyses.

Data analyses in this study followed different approaches: using the preprocessed reads directly and using assembled contigs or bins for monomicrobial and polymicrobial samples, respectively. Reads were assembled into contigs using metaSPAdes⁴⁴ from SPAdes v.3.14.0⁴⁵ and the resultant contigs were annotated using Prokka v.1.13 (<http://github.com/tseemann/prokka>). Contigs can be grouped by species into discrete units, referred to as bins, which were predicted using the tool MaxBin v.2.2.7⁴⁶, for recovering the *S. aureus* genome from the metagenomic datasets in polymicrobial samples.

Bacterial species were identified using Kraken v1.1.1⁴⁷ and the 8GB DustMasked MiniKraken database (as of Oct. 18, 2017) with default parameters. Re-estimation of abundance was done using Bracken⁴⁸.

The detection of antimicrobial resistance genes (ARGs) from the reads was determined using the tool Groot v.1.0.2⁴⁹ (<https://github.com/will-rowe/groot>). ARGs and virulence genes from the assembled contigs and bins were detected using ABRicate v0.8 (<https://github.com/tseemann/abricate>). For the detection of ARGs, two databases were used: the NCBI Bacterial Antimicrobial Resistance Reference Gene Database (BioProject accession number PRJNA313047) and CARD⁵⁰ databases (5,386 genes and 2,631 genes respectively, as of April 24th, 2020). For detection of virulence genes, the virulence factor database (VFDB) was used⁵¹. The thresholds used for determining the presence of ARG genes and VFs were set as 90 % identity and 90 % sequence coverage. Additionally, for ARGs prediction from the reads, a coverage depth of at least 20x was considered to report an ARG as present.

Typing

The assembled contigs and the bins were imported into SeqSphere+ software v.6.0.2 (Ridom GmbH, Münster, Germany) for a gene-by-gene allele calling comparison using the *S. aureus* species-specific scheme within SeqSphere+ for a cgMLST scheme for comparison of the 1,816 core loci in *S. aureus*, and an accessory typing scheme (wgMLST) with 706 accessory loci. Loci that flagged as failed (i.e., found but bearing frameshifts, or a differing consensus sequence, or having too-low coverage) were considered absent. Phylogenetic trees were constructed in SeqSphere+ using a minimum-spanning tree; missing values were pairwise ignored. The cluster-alert distance was set at a default of 24 allelic differences²⁷.

Statistical analysis

Descriptive statistics for categorical variables were based on percentages and frequencies, while continuous variables were based on means, standard deviations (SDs), medians and interquartile ranges (IQRs). In addition, Wilcoxon rank sum test was used to evaluate if the differences between the DNA sample preparation methods were statistically significant. The differences were considered statistically significant with p values < 0.05. Data were analyzed utilizing GraphPad Prism software, version 8.3.0 (GraphPad Software Inc., CA, US). Classification results from the metagenomics experiments were explored using the Pavian R package version 0.8.4⁵² by using their data tables, heatmaps and Sankey flow diagrams.

Abbreviations

SMg, Shotgun-metagenomics; BCBs, Blood culture bottles; PJI, Prosthetic joint infection; PJT, Prosthetic joint tissue; AMR, antimicrobial resistance; ARGs, antibiotic resistance genes; MRSA, methicillin-resistant *S. aureus*; WGS, whole genome sequencing; PC, positive control; AST, antimicrobial susceptibility testing; VFs, virulence factors; MLST, Multi locus sequence typing; cgMLST, Core-genome multi locus sequence typing; wgMLST, whole-genome multi locus sequence typing; ATCC, American type culture collection; MALDI-TOF, Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; ENA, European Nucleotide Archive.

Declarations

Data availability

The preprocessed reads generated for this study for each sample can be found in the European Nucleotide Archive (ENA) repository (www.ebi.ac.uk/ena) under the project number PRJEB43858.

Acknowledgements

We thank the Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, especially Lisbeth Engen Mortensen for help with the Antibiotic susceptibility testing (AST).

Author contributions

AS, GSS, and AMH initiated and planned the study design. AS performed the experiments and the bioinformatic analysis. AS, GSS and AMH assisted in scientific and technical design of experiments. JJ and EH contributed to the bioinformatic analysis. AS, GSS and AMH analyzed and interpreted the results. AS prepared the first version of the manuscript. All authors reviewed the manuscript, gave inputs and approved the submitted version.

Additional information

Competing Interests

The authors declare no competing interests. The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This work was supported by grants from joint “Miljøstøtte” financed by Strategisk-HN05–14 (Helse Nord RFH) and Faculty of Health Sciences A20389 (2014–2017), and the National Graduate School in Infection Biology and Antimicrobials (Grant No. 249062). The publication charges for this article have been funded by a grant from the publication fund of UiT – The Arctic University of Norway. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Li, Z.-L. *et al.* Identifying Common Pathogens in Periprosthetic Joint Infection and Testing Drug-resistance Rate for Different Antibiotics: A Prospective, Single Center Study in Beijing. *Orthop. Surg.* **10**, 235–240 (2018).
2. Tande, A. J. & Patel, R. Prosthetic Joint Infection. *Clin. Microbiol. Rev.* **27**, 302–345 (2014).

3. Tsai, J.-C., Sheng, W.-H., Lo, W.-Y., Jiang, C.-C. & Chang, S.-C. Clinical characteristics, microbiology, and outcomes of prosthetic joint infection in Taiwan. *J. Microbiol. Immunol. Infect.* **48**, 198–204 (2015).
4. Peel, T. N., Buising, K. L. & Choong, P. F. M. Prosthetic joint infection: challenges of diagnosis and treatment. *ANZ J. Surg.* **81**, 32–39 (2011).
5. Benito, N. *et al.* Time trends in the aetiology of prosthetic joint infections: a multicentre cohort study. *Clin. Microbiol. Infect.* **22**, 732.e1-732.e8 (2016).
6. Goudarzi, M., Fazeli, M., Goudarzi, H., Azad, M. & Seyedjavadi, S. S. Spa Typing of Staphylococcus aureus Strains Isolated From Clinical Specimens of Patients With Nosocomial Infections in Tehran, Iran. *Jundishapur J. Microbiol.* **9**, (2016).
7. Otto, M. Staphylococcal biofilms. *Curr. Top. Microbiol. Immunol.* **322**, 207–228 (2008).
8. Vasoo, S. Improving the Diagnosis of Orthopedic Implant-Associated Infections: Optimizing the Use of Tools Already in the Box. *J. Clin. Microbiol.* **56**, e01379-18 (2018).
9. Ivy, M. I. *et al.* Direct Detection and Identification of Prosthetic Joint Infection Pathogens in Synovial Fluid by Metagenomic Shotgun Sequencing. *J. Clin. Microbiol.* **56**, (2018).
10. Wang, C. *et al.* Preliminary assessment of nanopore-based metagenomic sequencing for the diagnosis of prosthetic joint infection. *Int. J. Infect. Dis.* **97**, 54–59 (2020).
11. Zhang, C. *et al.* Value of mNGS in sonication fluid for the diagnosis of periprosthetic joint infection. *Arthroplasty* **1**, 9 (2019).
12. Street, T. L. *et al.* Molecular diagnosis of orthopedic-device-related infection directly from sonication fluid by metagenomic sequencing. *J. Clin. Microbiol.* **55**, 2334–2347 (2017).
13. Thoendel, M. J. *et al.* Identification of Prosthetic Joint Infection Pathogens Using a Shotgun Metagenomics Approach. *Clin. Infect. Dis.* **67**, 1333–1338 (2018).
14. Sanderson, N. D. *et al.* Real-time analysis of nanopore-based metagenomic sequencing from infected orthopaedic devices. *BMC Genomics* **19**, 714 (2018).
15. Ruppé, E. *et al.* Clinical metagenomics of bone and joint infections: a proof of concept study. *Sci. Rep.* **7**, 7718 (2017).
16. Sanabria, A. *et al.* Shotgun-Metagenomics on Positive Blood Culture Bottles Inoculated With Prosthetic Joint Tissue: A Proof of Concept Study. *Front. Microbiol.* **11**, 1687 (2020).
17. Jacob, J. J., Veeraraghavan, B. & Vasudevan, K. Metagenomic next-generation sequencing in clinical microbiology. *Indian J. Med. Microbiol.* **37**, 133–140 (2019).

18. Wilson, M. R. *et al.* Clinical Metagenomic Sequencing for Diagnosis of Meningitis and Encephalitis. *N. Engl. J. Med.* **380**, 2327–2340 (2019).
19. Ellington, M. J. *et al.* The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clin. Microbiol. Infect.* **23**, 2–22 (2017).
20. Watts, G. S. & Hurwitz, B. L. Metagenomic Next-Generation Sequencing in Clinical Microbiology. *Clin. Microbiol. Newsl.* **42**, 53–59 (2020).
21. Couto, N. *et al.* Critical steps in clinical shotgun metagenomics for the concomitant detection and typing of microbial pathogens. *Sci. Rep.* **8**, (2018).
22. Hasman, H. *et al.* Rapid Whole-Genome Sequencing for Detection and Characterization of Microorganisms Directly from Clinical Samples. *J. Clin. Microbiol.* **52**, 139 LP – 146 (2014).
23. Watts, G. S. *et al.* Identification and quantitation of clinically relevant microbes in patient samples: Comparison of three k-mer based classifiers for speed, accuracy, and sensitivity. *PLOS Comput. Biol.* **15**, e1006863 (2019).
24. Joseph, S. J. *et al.* The single-species metagenome: subtyping *Staphylococcus aureus* core genome sequences from shotgun metagenomic data. *PeerJ* **4**, e2571–e2571 (2016).
25. Sangwan, N., Xia, F. & Gilbert, J. A. Recovering complete and draft population genomes from metagenome datasets. *Microbiome* **4**, 8 (2016).
26. Rodríguez-Brazzarola, P., Pérez-Wohlfeil, E., Díaz-del-Pino, S., Holthausen, R. & Trelles, O. Analyzing the Differences Between Reads and Contigs When Performing a Taxonomic Assignment Comparison in Metagenomics BT - Bioinformatics and Biomedical Engineering. in (eds. Rojas, I. & Ortuño, F.) 450–460 (Springer International Publishing, 2018).
27. Wildeman, P. *et al.* Genomic characterization and outcome of prosthetic joint infections caused by *Staphylococcus aureus*. *Sci. Rep.* **10**, 5938 (2020).
28. 2018, N.-V. *NORM/NORM-VET 2018. Usage of Antimicrobial Agents and Occurrence of Antimicrobial Resistance in Norway.* (2018). doi:ISSN:1502-2307
29. Ruppé, E., Cherkaoui, A., Lazarevic, V., Emonet, S. & Schrenzel, J. Establishing Genotype-to-Phenotype Relationships in Bacteria Causing Hospital-Acquired Pneumonia: A Prelude to the Application of Clinical Metagenomics. *Antibiot. (Basel, Switzerland)* **6**, 30 (2017).
30. Chen, C. & Hooper, D. C. Effect of *Staphylococcus aureus* Tet38 native efflux pump on in vivo response to tetracycline in a murine subcutaneous abscess model. *J. Antimicrob. Chemother.* (2018). doi:10.1093/jac/dkx432

31. Olsen, J. E., Christensen, H. & Aarestrup, F. M. Diversity and evolution of bla_Z from *Staphylococcus aureus* and coagulase-negative staphylococci. *J. Antimicrob. Chemother.* **57**, 450–460 (2006).
32. Gordon, N. C. *et al.* Prediction of *Staphylococcus aureus* Antimicrobial Resistance by Whole-Genome Sequencing. *J. Clin. Microbiol.* **52**, 1182 LP – 1191 (2014).
33. Lal Gupta, C., Kumar Tiwari, R. & Cytryn, E. Platforms for elucidating antibiotic resistance in single genomes and complex metagenomes. *Environ. Int.* **138**, 105667 (2020).
34. Uelze, L. *et al.* Typing methods based on whole genome sequencing data. *One Heal. Outlook* **2**, 3 (2020).
35. Boolchandani, M., D'Souza, A. W. & Dantas, G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat. Rev. Genet.* **20**, 356–370 (2019).
36. Nguyen, M., Olson, R., Shukla, M., VanOeffelen, M. & Davis, J. J. Predicting Antimicrobial Resistance Using Conserved Genes. *bioRxiv* 2020.04.29.068254 (2020). doi:10.1101/2020.04.29.068254
37. Chen, L. X., Anantharaman, K., Shaiber, A., Murat Eren, A. & Banfield, J. F. Accurate and complete genomes from metagenomes. *Genome Research* (2020). doi:10.1101/gr.258640.119
38. Greene, C. *et al.* Adhesion properties of mutants of *Staphylococcus aureus* defective in fibronectin-binding proteins and studies on the expression of *fnb* genes. *Mol. Microbiol.* **17**, 1143–1152 (1995).
39. Sangvik, M. *et al.* Age- and Gender-Associated *Staphylococcus aureus* spa Types Found among Nasal Carriers in a General Population: the Tromso Staph and Skin Study. *J. Clin. Microbiol.* **49**, 4213–4218 (2011).
40. Su, M., Satola, S. W. & Read, T. D. Genome-Based Prediction of Bacterial Antibiotic Resistance. *J. Clin. Microbiol.* **57**, (2018).
41. Sanabria, A. *et al.* Culturing periprosthetic tissue in BacT/Alert® Virtuo blood culture system leads to improved and faster detection of prosthetic joint infections. *BMC Infect. Dis.* **19**, 607 (2019).
42. EUCAST. Testing Breakpoint tables for interpretation of MICs and zone diameters. https://www.eucast.org/Ast_of_Bacteria/ (2020).
43. Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* **7**, 1338 (2018).
44. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
45. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

46. Wu, Y.-W., Tang, Y.-H., Tringe, S. G., Simmons, B. A. & Singer, S. W. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2**, 26 (2014).
47. Venter, J. C. Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science* (80-). **304**, 66–74 (2004).
48. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104 (2017).
49. Rowe, W. P. M. & Winn, M. D. Indexed variation graphs for efficient and accurate resistome profiling. *Bioinformatics* **34**, 3601–3608 (2018).
50. Jia, B. *et al.* CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).
51. Chen, L. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* **33**, D325–D328 (2004).
52. Breitwieser, F. P. & Salzberg, S. L. Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics* **36**, 1303–1304 (2020).

Figures

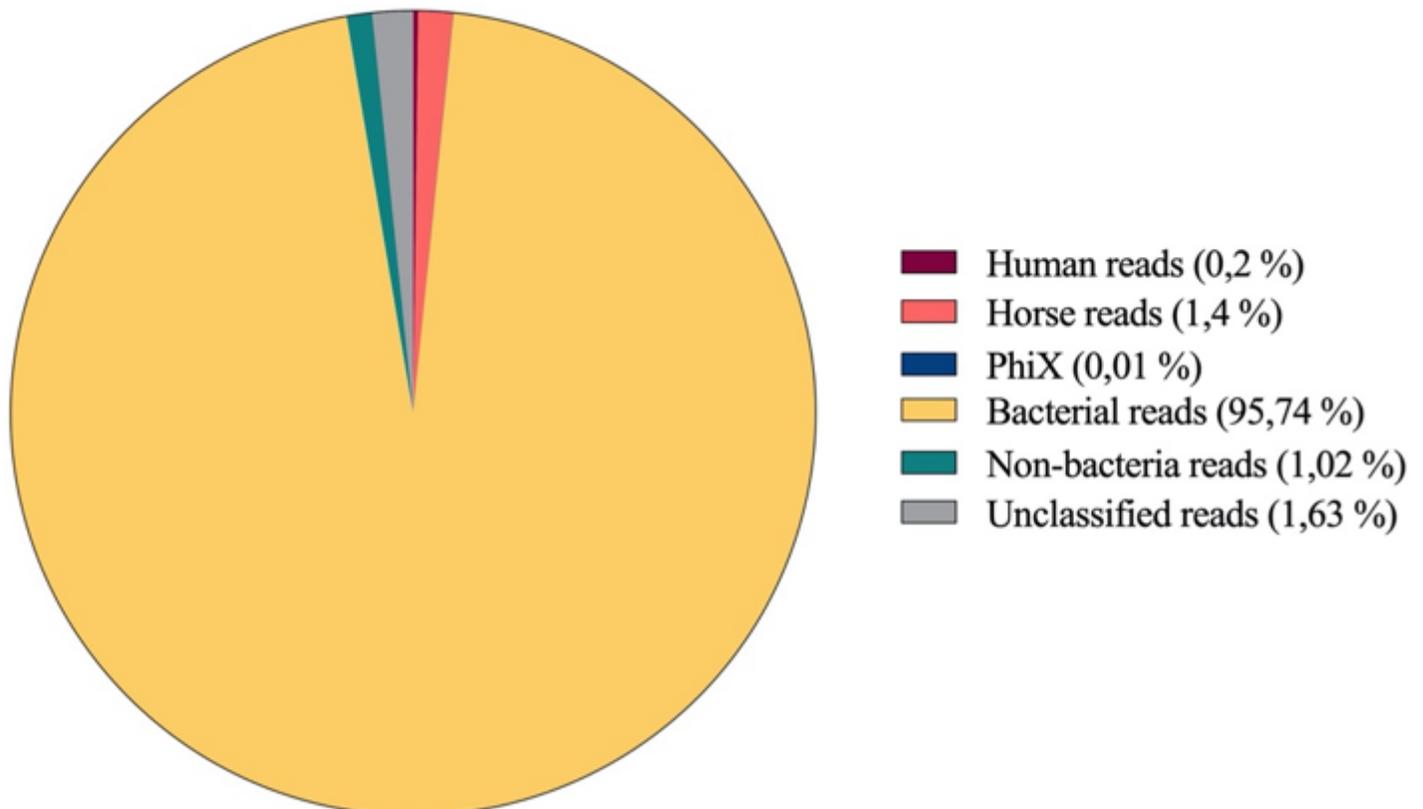


Figure 1

Proportion of reads taxonomically classified as human, horse, PhiX, bacteria, non-bacterial and unclassified.

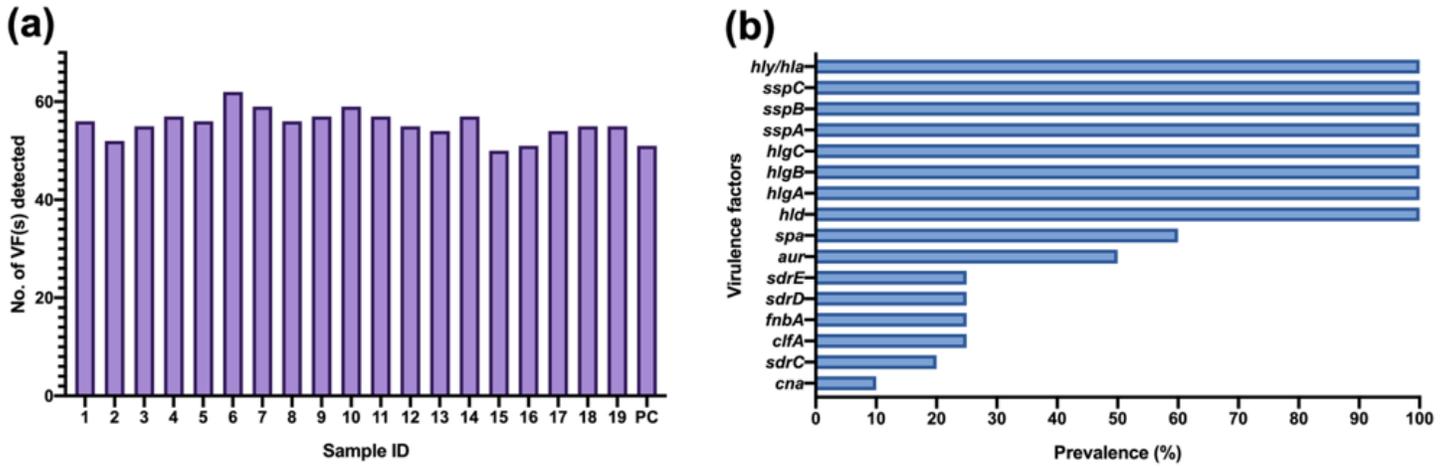


Figure 2

Virulence genes predicted by SMg from *S. aureus* in PJT samples in this study. (a) Number of virulence genes predicted per sample. (b) Prevalence of virulence genes associated with pathogenicity of *S. aureus* in PJI.

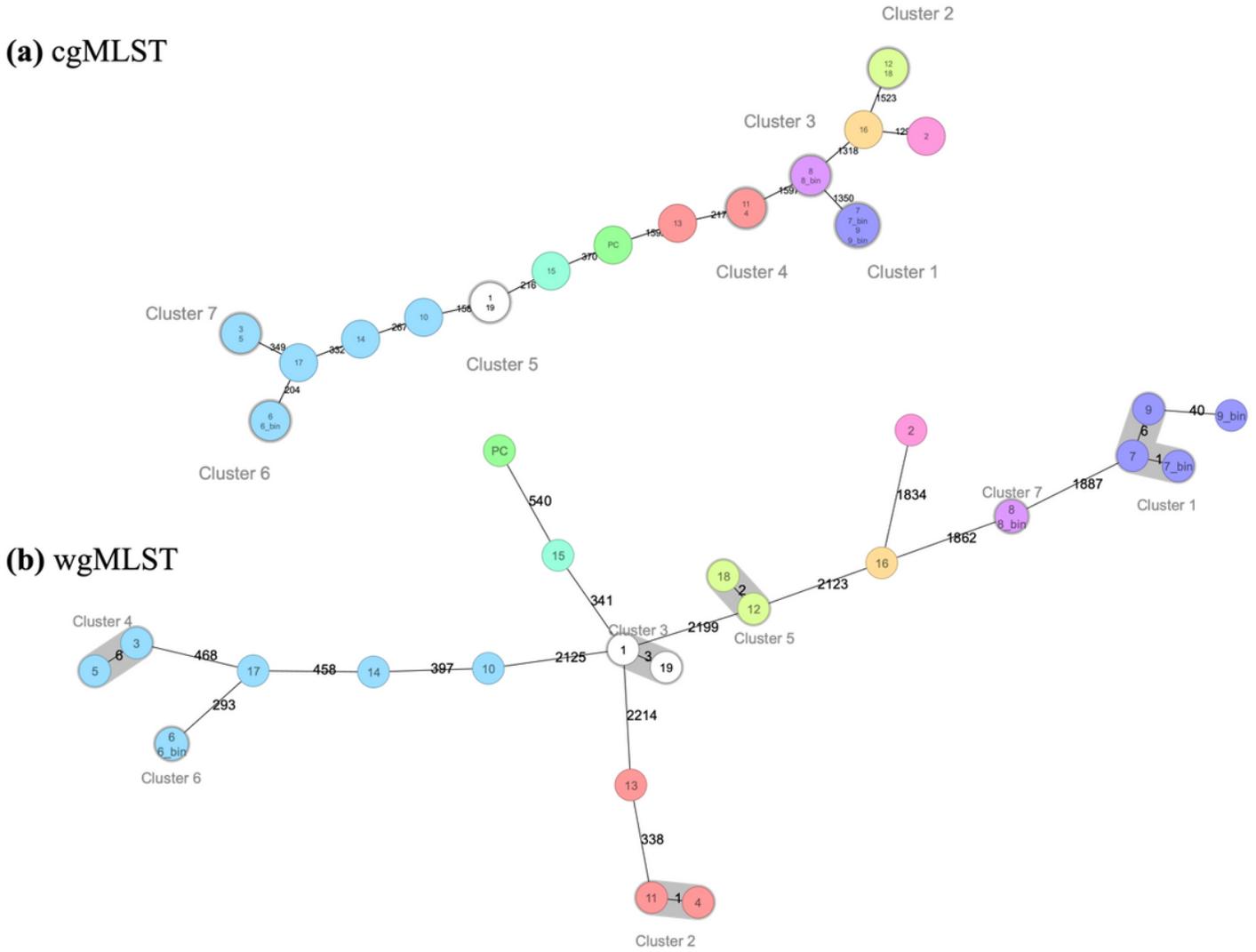


Figure 3

Minimum-spanning tree based on cgMLST (a) and wgMLST (b) allelic profiles of *S. aureus* genomes obtained from SMg. Color nodes according to sequence type. The number in the connecting lines illustrates the number of targeted genes with differing alleles.

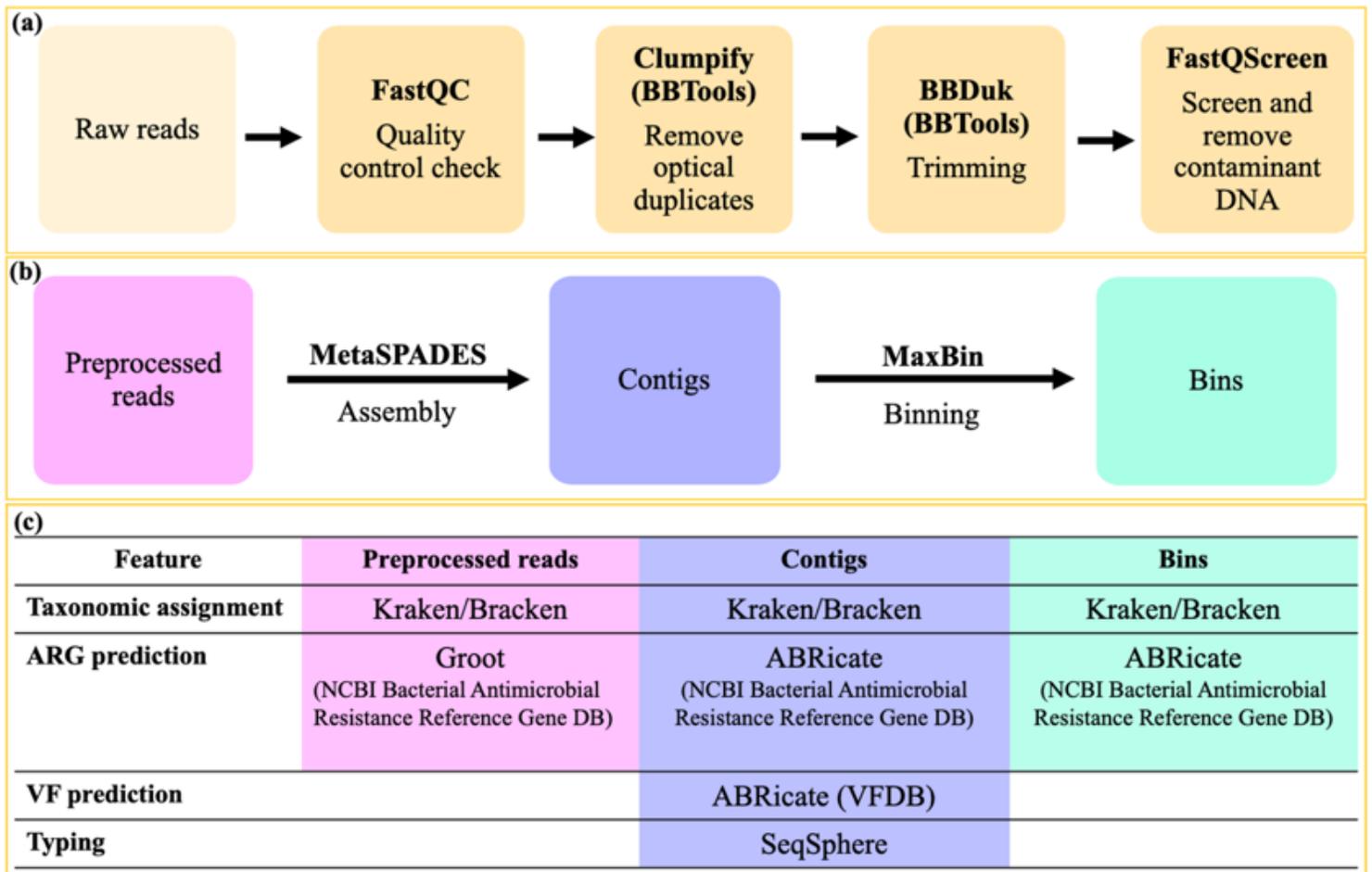


Figure 4

Workflow summarizing the bioinformatic analyses in this study, including (a) data preprocessing, (b) data analyses approaches and (c) data analyses and interpretation. ARG, antimicrobial resistance gene; VF, virulence factor; AMR, antimicrobial resistance.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTables.xlsx](#)
- [SupplementaryFigureS1.png](#)
- [Supplementarymateriallegends.docx](#)