INPLASY

INPLASY202660087

doi: 10.37766/inplasy2026.6.0087

Received: 18 June 2026

Published: 18 June 2026

# Genome-based models for the prediction of Staphylococcus aureus virulence phenotypes: a protocol for a scoping review

Collinson, SL; McCann, KM; Saha, SK; Athan, E; Botheras, CL.

**Corresponding author:**
Samuel Collinson

sam.collinson@deakin.edu.au

**Author Affiliation:**
CIIDIR, IMPACT, School of Medicine, Deakin University, Geelong, Australia.

**ADMINISTRATIVE INFORMATION**

**Support -** Australian RTP Scholarship.

**Review Stage at time of this submission -** Formal screening of search results against eligibility criteria.

**Conflicts of interest -** None declared.

**INPLASY registration number:** This research was supported by an Australian Government Research Training Program (RTP) Scholarship (doi.org/10.82133/C42F-K220).

**Amendments -** This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 18 June 2026 and was last updated on 18 June 2026.

## INTRODUCTION

**Review question / Objective** The overarching aim of this scoping review is to descriptively map existing evidence on genome-based predictive models for the prediction of clinically relevant virulence phenotypes in S aureus infections. Specifically, this review will aim to:
i. To identify the scope of clinically relevant S. aureus virulence phenotypes that genome-based models have been used to predict, including but not limited to adhesion, toxicity, antibiotic resistance, clot formation, and biofilm formation.
ii. To map what type of genetic data (e.g. aligned genes, k-mers and/or single-nucleotide polymorphisms) these models have been designed for, and what sequencing technologies have been used to collect this genetic data (e.g. short-read sequencing, long-read sequencing, adaptive sampling).
iii. To examine the breadth, feasibility, and accuracy of these models.
iv. To assess whether these models have informed the development of rapid genetic diagnostic tests or lead to other changes in treatment or clinical protocols.

**Background** Staphylococcus (S.) aureus is one of the deadliest bacterial pathogens worldwide. When S. aureus is disseminated into the bloodstream it is known as S. aureus bacteraemia (SAB). SAB has high mortality rates of up to 30% with approximately 300,000 deaths per year (1, 2). Complications from a wide range of metastatic infections including osteomyelitis, septic arthritis, endocarditis (3, 4) and various device-associated infections (5-7) are one of the driving factors of high mortality rates. Several bacterial virulence phenotypes effect the risk of complications and mortality, such as antimicrobial resistance, toxin production and biofilm production (3, 6, 8).

Due to the high genetic diversity and severity observed among SAB infections, identifying S. aureus virulence features quickly and accurately is vital to improve SAB outcomes (1, 6, 9). Rapid molecular testing techniques, such as polymerase chain reaction (PCR) amplification testing, have seen a surge in popularity over the past 20 years in diagnostics due to their speed and effectiveness, with a majority of the clinical use of these test being focussed on the identification of methicillin resistant S. aureus based on the presence of resistance-encoding genes such as femA, mecA or mecC (10-12). Apart from methicillin resistance, many other virulence phenotypes can be identified by rapid genetic testing, such as the production of Panton-Valentine leucocidin (13, 14) and toxic-shock syndrome toxin (15). However, many other virulence phenotypes are the result of complex interplay between a large number of genes that are too vast for traditional PCR testing. For example, biofilm formation can be initiated and mediated by various cell wall-anchored adhesive proteins (16), while the extracellular polymeric component that gives biofilm their mass varies between strains and can be composed of glucosamine polymer (17), extracellular DNA (18) and/or fibrillar protein (19). As a result of this complexity, there is currently no clinically relevant molecular diagnosis technique for these complex phenotypes such as biofilm production (9). Similarly, vancomycin resistance in S. aureus can be mediated by genes within the horizontally transferred vanA / vanB genes (20), or through mutations in various chromosomal enzymes and regulatory systems which control cell-wall thickness and structure (21). Because of these additional mutations which concur vancomycin resistance, many phenotypically vancomycin resistant S. aureus lack vanA / vanB (22, 23), impeding traditional PCR testing and highlighting the need to further understand the genetic basis of vancomycin resistance.

Genome-based models are predictive models, including machine-learning models (e.g. random forest, gradient boosting, mixed-effects linear model) (24, 25) and phylogeny-based models (e.g. Scoary (26)), which utilise genomic and genetic data to infer phenotypic features based on the complex interplay across multiple different genes. The first whole S. aureus genome was assembled in 2001 (27), and since then, advances in clinically-applicable sequencing technologies have enabled the generation of whole-genome sequences within 36 hours of overnight SAB blood cultures (28), potentially allowing for the analysis of whole S. aureus genomes as a viable diagnostic alternative to PCR-based approaches and traditional microbiological characterisation (9).

**Rationale** Due to advances in the speed, accuracy and cost-effectiveness of whole genome sequencing technologies, genome-based models may soon become a highly effective diagnostics tool (28). Additionally, analysing more complex S. aureus phenotypes based off data from whole genome sequencing would allow for the identification of the underlying genetic mechanisms of S. aureus disease; the identification of potential targets for innovative treatments and the identification of novel S. aureus phenotypes (25, 29).

Several genome-based modes for predicting S. aureus phenotypic virulence features have seen success in research settings in recent years, such as models for predicting antibiotic resistance (25, 29, 30) or differentiating between S. aureus infection versus carriage (24). For example, a study by Wang and colleagues (31) utilised a support-vector machine-learning model to identify multiple virulence genes that significantly enhanced resistance to a variety of antibiotics, including map, lpl2, sdrE, essG and ebh, highlighting the ability of genome-based models to predict complex phenotypic relationships. Work by Rishishwar and colleagues (30) to differentiate vancomycin-susceptible and intermediate vancomycin-resistant S. aureus identified a set of 11 genes that could reliably differentiate vancomycin resistance with an accuracy of 84%, however this study was limited by its sample size of 24 isolates. Another study (24) compared the ability of various genome-based models to differentiate infective S. aureus and commensal S. aureus using a collection of 512 unique isolates. Their strongest model, a mixed-effects linear model, identified a set of 10 genes that achieved a classification accuracy of 77%.

Despite the numerous genome-based models employed in research settings, their ability to translate into clinical practise needs to be better understood. Additionally, many different genome-based models have been utilised in S. aureus research (and microbiology in general) (24, 25, 29), making comparisons between studies difficult. Thus, the scoping review will aim to descriptively map evidence around the use of genome-based models for the prediction of clinically relevant virulence phenotypes in S. aureus by examining the range of virulence phenotypes that successful models have been designed for, comparing the types and sources of genetic data used in each model, the observed accuracies (and other model evaluation statistics) of different type of genome-based models, and investigating whether these models have made an impact on the development

of rapid genetic diagnostic tests or lead to other changes in treatment or clinical protocols for S. aureus infection. The findings of this scoping review will help guide the future development and implementation of genome-based models for S. aureus virulence phenotype prediction by providing valuable data to researchers on the potential strengths and weaknesses of specific modelling techniques and sequencing techniques in the context of implementing these models within clinical settings.

## METHODS

**Strategy of data synthesis** A systematic search of study title and abstracts across the four electronic databases PubMed, Web of Science Core Collection, EMBASE and the Bielefeld Academic Search Engine (BASE) will be performed to identify relevant peer reviewed studies. A set of five seed articles (25, 30, 32-34) was used to validate and guide the development of this search strategy. Database-specific search strategies will be developed according to the following keyword search strategy: "Staphylococcus aureus" AND ("machine-learning" OR "predictive model" OR "gen* model" OR "gen* based" OR "gen* sequence-based" OR "algorithm") AND (GWAS OR "genome wide" OR "gen* predictors" OR "virulence genes" OR "whole genome" OR "gen* markers" OR WGS OR SNP OR "k-mer") AND (phenotyp* OR infecti* OR pathogen* OR virulen* OR toxi* OR resistan* OR susceptib*). Extracted studies will be restricted to studies published from 1 January 2001 to 15 June 2026 of which English full text is available.

**Eligibility criteria** We will use the PRISMA-ScR checklist (35) and the 2024 Joanne-Briggs Institute Manual for Evidence Synthesis (36) to guide the development of this scoping review protocol and identify eligible studies.

Eligible studies will be identified according to the eligibility criteria outlined below:
i. Condition: The determination of S. aureus virulence phenotypes (e.g. antibiotic resistance, toxicity, etc.) based primarily on genetic data.
ii. Population: S. aureus clinical samples isolated from patients with a S. aureus disease, which have had their DNA sequencing data analysed by genome-based models, including machine-learning techniques (e.g. random forest, gradient boosting, mixed linear model) and phylogeny-based modelling techniques (e.g. Scoary).
iii. Intervention: Any genome-based models used to predict virulence phenotypes in S. aureus. The predictive elements can be any type of genetic data (e.g. aligned genes, single nucleotide polymorphisms, k-mers, gene copy numbers) gathered using any sort of genetic sequencing technology.
iv. Study design: Retrospective studies, diagnostic accuracy studies, randomised controlled trials, cohort studies, case-control studies, cross-sectional studies, case series.
v. Availability: Full-text, English articles are available.
vi. Time: Studies published from 1 January 2001 to 15 June 2026.

Studies will be excluded if they meet any of the following exclusion criteria
i. Models utilising data exclusively from S. aureus isolates not causing disease.
ii. Models utilising data from genetically modified S. aureus.
iii. Models utilising isolates from non-human sources.
iv. Studies where the virulence phenotype was measured in a polymicrobial infection.
v. Studies that only used models to infer phylogenetic relationships without predicting virulence.
vi. Studies that used models that incorporated data from multiple bacterial species and had no S. aureus exclusive model.
vii. Studies of which full-text or English text is unavailable for.

**Source of evidence screening and selection** Study screening will be conducted in a two-level process consisting first of title and abstract screening, then a second level of full-text screening. For level one screening, two reviewers (SC and CB) will independently screen the title and abstract of each study against the eligibility criteria. Studies will be excluded from this review if a single eligibility criterion is failed. Discordance will be discussed between the reviewers, with a third adjudicating reviewer (KM) used if agreement cannot be reached. To verify the effectiveness and rigor of this studies reliability criteria, this first level of screening will be pilot tested on a random sample of 25 titles and abstracts by two reviewers (SC and CB), followed by a team meeting to discuss discrepancies until a consensus among all reviewers of 80% or more is achieved.For level two screening, all studies that passed level 1 screening will have their full text screened independently by two reviewers (SC and CB) against the eligibility criteria. Studies will be excluded from this review if a single eligibility criterion is failed. Discordance will be discussed between the reviewers, with a third adjudicating reviewer (KM) used if agreement cannot be reached.

**Data management** Identified studies will be imported into Covidence (https://www.covidence.org). Removal of duplicates and two-level study screening will be conducted by reviewers using Covidence's built-in tools. A data extraction form will be developed using Covidence's data extraction tool will be used to systematically extract data from each study. A PRISMA-ScR flow-chart will be constructed to maintain transparency around the number of included and excluded studies at each stage of study selection.

**Reporting results / Analysis of the evidence** A data collection form developed a priori in Covidence will be used to collect:
i. Study data: authorship, date published and study design.
ii. Isolate collection data: sample size, country or countries of origin, infection type, sub-population type ie strain specific studies, sequencing platform / technology used.
iii. Model data: number of target genes / significant k-mers / significant SNPs included in models, most significant predictive genes, model outcome phenotype/s, model type, software version, model evaluation statistics (sensitivity, specificity, accuracy, receiver-operating characteristic, positive predictive value, negative predictive value, etc.).

Once data collection is complete, study data will be tabulated and presented descriptively in terms of:
i. The breadth of models used across studies, along with their accuracy (proportion of isolates that were classified into the correct virulence phenotypes by models) and other model evaluation statistics.
ii. The type of genetic data utilised as predictors in each study (aligned genes, k-mers or SNPs) and how genetic data was sequenced and processed (short-read, long-read sequencing, adaptive sampling, etc.).
iii. The breadth and scope of S. aureus phenotypes that genome-based models have been designed for. The levels of phenotype were models designed for (binary, categorical, continuous).
iv. The impact (measured in field weight citation index) of each study. Study citations will be searched for guidelines, patents, animal trials and clinical trials to determine if they have made any meaningful contribution to clinical practice.

**Language restriction** This scoping review will only include studies of which their full-text is available in English.

**Country(ies) involved** Australia.

**Search strategies**
PubMed search
"Staphylococcus aureus"[Title/Abstract] AND ("machine learning"[Title/Abstract] OR "machine-learning"[Title/Abstract] OR "predictive model"[Title/Abstract] OR "geno* model"[Title/Abstract] OR "geno*-based"[Title/Abstract] OR "geno* based"[Title/Abstract] OR "geno* sequence-based"[Title/Abstract] OR "gene* model"[Title/Abstract] OR "gene*-based"[Title/Abstract] OR "gene* based"[Title/Abstract] OR "gene* sequence-based"[Title/Abstract] OR "algorithm"[Title/Abstract]) AND (GWAS[Title/Abstract] OR "genome-wide"[Title/Abstract] OR "genome wide"[Title/Abstract] OR "geno* predictors"[Title/Abstract] OR "gene* predictors"[Title/Abstract] OR "virulence genes" [Title/Abstract] OR "whole genome"[Title/Abstract] OR "gene* markers"[Title/Abstract] OR "geno* markers"[Title/Abstract] OR "whole-genome"[Title/Abstract] OR WGS[Title/Abstract] OR SNP[Title/Abstract] OR "k-mer"[Title/Abstract] OR kmer[Title/Abstract] OR "k mer"[Title/Abstract]) AND (phenotyp*[Title/Abstract] OR infecti*[Title/Abstract] OR pathogen*[Title/Abstract] OR virulen*[Title/Abstract] OR toxi*[Title/Abstract] OR resistan*[Title/Abstract] OR susceptib*[Title/Abstract]) AND 2001:2026[dp] AND (medline[sb] OR "pubmed pmc"[sb])
Web of Science Core Collection search [Timespan: 2001-01-01 to 2026-06-15 (Publication Date)]
((TI=("Staphylococcus aureus")) OR AB=("Staphylococcus aureus")) AND ((TI=("machine learning" OR "machine-learning" OR "predictive model" OR "geno* model" OR "geno*-based" OR "geno* based" OR "geno* sequence-based" OR "gene* model" OR "gene*-based" OR "gene* based" OR "gene* sequence-based" OR "algorithm")) OR AB=("machine learning" OR "machine-learning" OR "predictive model" OR "geno* model" OR "geno*-based" OR "geno* based" OR "geno* sequence-based" OR "gene* model" OR "gene*-based" OR "gene* based" OR "gene* sequence-based" OR "algorithm")) AND ((TI=(GWAS OR "genome-wide" OR "genome wide" OR "geno* predictors" OR "gene* predictors" OR "virulence genes" OR "geno* markers" OR "gene* markers" OR "whole genome" OR "whole-genome" OR WGS OR SNP OR "k-mer" OR "kmer" OR "k mer")) OR AB=(GWAS OR "genome-wide" OR "genome wide" OR "geno* predictors" OR "gene* predictors" OR "virulence genes" OR "geno* markers" OR "gene* markers" OR "whole genome" OR "whole-genome" OR WGS OR SNP OR "k-mer" OR kmer OR "k mer")) AND ((TI=(phenotyp* OR infecti* OR pathogen* OR virulen* OR toxi* OR

resistan* OR susceptib*)) OR AB=(phenotyp* OR infecti* OR pathogen* OR virulen* OR toxi* OR resistan* OR susceptib*))
EMBASE search [broad search mode]
'staphylococcus aureus':ti,ab,kw AND ('machine learning':ti,ab,kw OR 'machine-learning':ti,ab,kw OR 'predictive model':ti,ab,kw OR 'geno* model':ti,ab,kw OR 'geno*-based':ti,ab,kw OR 'geno* based':ti,ab,kw OR 'geno* sequence-based':ti,ab,kw OR 'gene* model':ti,ab,kw OR 'gene*-based':ti,ab,kw OR 'gene* based':ti,ab,kw OR 'gene* sequence-based':ti,ab,kw OR 'algorithm':ti,ab,kw) AND (gwas:ti,ab,kw OR 'genome-wide':ti,ab,kw OR 'genome wide':ti,ab,kw OR 'gen* predictors':ti,ab,kw OR 'gen* markers':ti,ab,kw OR 'virulence genes':ti,ab,kw OR 'whole genome':ti,ab,kw OR 'whole-genome':ti,ab,kw OR wgs:ti,ab,kw OR snp:ti,ab,kw OR 'k-mer':ti,ab,kw OR 'kmer':ti,ab,kw OR 'k mer':ti,ab,kw) AND ('phenotyp*':ti,ab,kw OR 'infecti*':ti,ab,kw OR 'pathogen*':ti,ab,kw OR 'virulen*':ti,ab,kw OR 'toxi*':ti,ab,kw OR 'resistan*':ti,ab,kw OR 'susceptib*':ti,ab,kw) AND (2001:py OR 2002:py OR 2003:py OR 2004:py OR 2005:py OR 2006:py OR 2007:py OR 2008:py OR 2009:py OR 2010:py OR 2011:py OR 2012:py OR 2013:py OR 2014:py OR 2015:py OR 2016:py OR 2017:py OR 2018:py OR 2019:py OR 2020:py OR 2021:py OR 2022:py OR 2023:py OR 2024:py OR 2025:py OR 2026:py)
BASE search
“Staphylococcus aureus” (“machine learning” “gen* model” “gen* based” “predictive model” “gen* sequence-based” “algorithm”) (GWAS “genome wide” “gen* predictors” “virulence genes” “gen* markers” “whole genome” WGS SNP “k*mer”) (phenotyp* infecti* pathogen* virulen* toxi* resistan* susceptib*) year:[2001 TO 2026] lang:en

**Keywords** Staphylococcus aureus; Scoping Review; Genome-Based Models; Machine Learning; Phenotypic Prediction; Virulence Factors.

**Dissemination plans** This scoping review's findings will be disseminated through presentations and seminars at Deakin University (and affiliate organisations) and through publication in an open-access peer-reviewed journal. The results of this study will also form a portion of the PhD thesis of author SC.

## Contributions of each author

Author 1 - Samuel Collinson.
Email: sam.collinson@deakin.edu.au
Author 2 - Kirsty McCann.
Email: kirsty.mccann@deakin.edu.au
Author 3 - Sajal Saha.
Email: sajal.saha@deakin.edu.au
Author 4 - Eugene Athan.
Email: e.athan@deakin.edu.au
Author 5 - Carly Botheras.
Email: carly.botheras@deakin.edu.au

## References

1. Bai AD, Lo CKL, Komorowski AS, Suresh M, Guo K, Garg A, et al. Staphylococcus aureus bacteraemia mortality: a systematic review and meta-analysis. Clinical Microbiology and Infection. 2022;28(8):1076-84.
2. Ikuta KS, Swetschinski LR, Robles Aguilar G, Sharara F, Mestrovic T, Gray AP, et al. Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2022;400(10369):2221-48.
3. Alonso B, Pérez-Granda MJ, Latorre MC, Sánchez-Carrillo C, Bouza E, Muñoz P, et al. Production of biofilm by Staphylococcus aureus: Association with infective endocarditis? Enferm Infecc Microbiol Clin (Engl Ed). 2022;40(8):418-22.
4. Lerche CJ, Schwartz F, Theut M, Fosbøl EL, Iversen K, Bundgaard H, et al. Anti-biofilm Approach in Infective Endocarditis Exposes New Treatment Strategies for Improved Outcome. Frontiers in Cell and Developmental Biology. 2021;Volume 9 - 2021.
5. Gutiérrez-Gutiérrez B, Gallego-Mesa B, Kaasch AJ, Riediger M, Rieg S, Trigo M, et al. Identification and validation of clinical phenotypes in Staphylococcus aureus bloodstream infection and their association with mortality (FEN-AUREUS study). EClinicalMedicine. 2025;83:103240.
6. Russell CD, Berry K, Cooper G, Sim W, Lee RS, Gan TY, et al. Distinct Clinical Endpoints of Staphylococcus aureus Bacteraemia Complicate Assessment of Outcome. Clinical Infectious Diseases. 2024;79(3):604-11.
7. Tong SYC, Fowler VG, Jr., Skalla L, Holland TL. Management of Staphylococcus aureus Bacteremia: A Review. Jama. 2025;334(9):798-808.
8. Blechman SE, Wright ES. Vancomycin-resistant Staphylococcus aureus (VRSA) can overcome the cost of antibiotic resistance and may threaten vancomycin's clinical durability. PLoS Pathog. 2024;20(8):e1012422.
9. Høiby N, Moser C, Oliver A, Williams C, Ramage G, Borghi E, et al. To update or not to update the ESCMID guidelines for the diagnosis and treatment of biofilm infections – That is the question! The opinion of the ESGB board. Biofilm. 2023;6:100135.
10. Peri AM, Ling W, Furuya-Kanamori L, Harris PNA, Paterson DL. Performance of BioFire Blood Culture Identification 2 Panel (BCID2) for the detection of bloodstream pathogens and their associated resistance markers: a systematic

review and meta-analysis of diagnostic test accuracy studies. BMC Infect Dis. 2022;22(1):794.
11. Yetukuri J, Patel D, Bandali A, Giordano P, Roland R, Kessler J. Impact of rapid blood culture identification PCR panel on optimal antibiotic use in methicillin-susceptible Staphylococcus aureus bacteremia. Microbiol Spectr. 2024;12(12):e0062924.
12. Na SH, Kim C-J, Kim M, Park JS, Song K-H, Choe PG, et al. Impact of the multiplex polymerase chain reaction in culture-positive samples on appropriate antibiotic use in patients with staphylococcal bacteremia. Diagnostic Microbiology and Infectious Disease. 2016;84(4):353-7.
13. Ibrahim RA, Berhe N, Mekuria Z, Seyoum ET, Balada-Llasat JM, Abebe T, et al. Antimicrobial Resistance and Virulence Gene Profile of Clinical Staphylococcus aureus: A Multi-Center Study from Ethiopia. Infect Drug Resist. 2023;16:4835-44.
14. Adeyemi FM, Oyedara OO, Yusuf-Omoloye NA, Ajigbewu OH, Ndaji OL, Adegbite-Badmus MK, et al. Guardians of resistance and virulence: detection of mec, femA, Van, pvl, hlg and spa genes in methicillin and vancomycin-resistant Staphylococcus aureus from clinical and food samples in Southwestern Nigeria. BMC Microbiol. 2024;24(1):498.
15. Sharma P, Wang N, Chervin AS, Quinn CL, Stone JD, Kranz DM. A Multiplex Assay for Detection of Staphylococcal and Streptococcal Exotoxins. PLoS One. 2015;10(8):e0135986.
16. Foster TJ, Geoghegan JA, Ganesh VK, Höök M. Adhesion, invasion and evasion: the many functions of the surface proteins of Staphylococcus aureus. Nature Reviews Microbiology. 2014;12(1):49-62.
17. Bamneshin K, Poudineh M, Alibabaei RH, Amiri MRJ, Fateminasab ZS, Ghorbani Z, et al. Prevalence of icaADBC genes, and correlation with biofilms and antibiotic resistance in S. aureus: a systematic review and meta-analysis. Germs. 2024;14(4):387-401.
18. Kavanaugh JS, Flack CE, Lister J, Ricker EB, Ibberson CB, Jenul C, et al. Identification of Extracellular DNA-Binding Proteins in the Biofilm Matrix. mBio. 2019;10(3):10.1128/mbio.01137-19.
19. Yonemoto K, Chiba A, Sugimoto S, Sato C, Saito M, Kinjo Y, et al. Redundant and Distinct Roles of Secreted Protein Eap and Cell Wall-Anchored Protein SasG in Biofilm Formation and Pathogenicity of Staphylococcus aureus. Infection and Immunity. 2019;87(4):10.1128/iai.00894-18.
20. Zhu W, Clark Nancye C, McDougal Linda K, Hageman J, McDonald LC, Patel Jean B. Vancomycin-Resistant Staphylococcus aureus Isolates Associated with Inc18-Like vanA Plasmids in Michigan. Antimicrobial Agents and Chemotherapy. 2008;52(2):452-7.
21. Cui J, Zhang H, Mo Z, Yu M, Liang Z. Cell wall thickness and the molecular mechanism of heterogeneous vancomycin-intermediate Staphylococcus aureus. Letters in Applied Microbiology. 2021;72(5):604-9.
22. Bamigboye BT, Olowe OA, Taiwo SS. Phenotypic and Molecular Identification of Vancomycin Resistance in Clinical Staphylococcus Aureus Isolates in Osogbo, Nigeria. Eur J Microbiol Immunol (Bp). 2018;8(1):25-30.
23. Tawfeek CE, Khattab S, Elmaraghy N, Heiba AA, Nageeb WM. Reduced vancomycin susceptibility in Staphylococcus aureus clinical isolates: a spectrum of less investigated uncertainties. BMC Infect Dis. 2024;24(1):1218.
24. Chen J, Du W, Li Y, Zhou H, Ouyang D, Yao Z, et al. Genome-based model for differentiating between infection and carriage Staphylococcus aureus. Microbiol Spectr. 2024;12(10):e0049324.
25. Wang S, Zhao C, Yin Y, Chen F, Chen H, Wang H. A Practical Approach for Predicting Antimicrobial Phenotype Resistance in Staphylococcus aureus Through Machine Learning Analysis of Genome Data. Frontiers in Microbiology. 2022;Volume 13 - 2022.
26. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biology. 2016;17(1):238.
27. Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, et al. Whole genome sequencing of meticillin-resistant Staphylococcus aureus. The Lancet. 2001;357(9264):1225-40.
28. Köser CU, Holden MTG, Ellington MJ, Cartwright EJP, Brown NM, Ogilvy-Stuart AL, et al. Rapid Whole-Genome Sequencing for Investigation of a Neonatal MRSA Outbreak. New England Journal of Medicine. 2012;366(24):2267-75.
29. Nguyen M, Olson R, Shukla M, VanOeffelen M, Davis JJ. Predicting antimicrobial resistance using conserved genes. PLOS Computational Biology. 2020;16(10):e1008319.
30. Rishishwar L, Petit Robert A, Kraft Colleen S, Jordan IK. Genome Sequence-Based Discriminator for Vancomycin-Intermediate Staphylococcus aureus. Journal of Bacteriology. 2014;196(5):940-8.
31. Wang W, Baker M, Hu Y, Xu J, Yang D, Maciel-Guerra A, et al. Whole-Genome Sequencing and Machine Learning Analysis of Staphylococcus aureus from Multiple Heterogeneous Sources in China Reveals Common Genetic Traits of Antimicrobial Resistance. mSystems. 2021;6(3):10.1128/msystems.01185-20.
32. Sassi M, Bronsard J, Pascreau G, Emily M, Donnio P-Y, Revest M, et al. Forecasting

Staphylococcus aureus Infections Using Genome-Wide Association Studies, Machine Learning, and Transcriptomic Approaches. mSystems. 2022;7(4):e00378-22.
33. Liu Y, Wang X, Wu L, Shi Z, Zhu M, Ye L, et al. Prediction of antimicrobial resistance in Staphylococcus aureus with a machine learning classifier based on WGS data. Microbiol Spectr. 2025;13(9):e0006525.
34. Laabei M, Recker M, Rudkin JK, Aldeljawi M, Gulay Z, Sloan TJ, et al. Predicting the virulence of MRSA from its genome sequence. Genome Research. 2014;24(5):839-49.
35. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Annals of internal medicine. 2018;169(7):467-73.
36. Peters DJP, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Scoping Reviews (2020). In: Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. JBI Manual for Evidence Synthesis: JBI; 2024.