AMERICAN SOCIETY FOR MICROBIOLOGY
Journals.ASM.org

Genomic surveillance of Methicillin Resistant *Staphylococcus aureus* in Saudi Arabia

Mohammed S. Alarawi[1,32,33], Musaad Altammami [4], Mohammed Abutarboush[4], Maxat Kulmanov [1,2,29,30,32], Dalal M. Alkuraithy [7], Senay Kafkas[1,32], Robert Radley[1], Marwa Abdelhakim[1,29,32], Hind Masfer Abdullah Aldakhil[1], Reema A. Bawazeer[12,14], Mohammed A. Alolayan[4], Basel M. Alnafjan[4], Abdulaziz A. Huraysi[4], Amani Almaabadi [4], Bandar A Suliman[5,6], Areej G Aljohani[24], Hassan A. Hemeg[6], Turki S Abujamel[8,9], Anwar M. Hashem [8,9] Ibrahim A Al-Zahrani[9], Mohammed S Abdoh[10], Haya I Hobani[8], Rakan F Felemban[11], Wafaa A Alhazmi[9], Pei-Ying Hong[3], Majed F. Alghoribi[12,13,14,15], Sameera Aljohani[12,13,15], Hanan Balkhy[17], Abdulrahman Alswaji[12,14], Maha Alzayer[12,14], Bassam Alalwan[15], Mai M. Kaaki[18], Sharif M. Hala[13,14,16,31], Omniya Ahmad Fallatah[13,14,16], Wesam Ahmad Bahitham[13,14,16], Samer Yahya Zakri[13,14,16], Mohammad A Alshehri[13,14,16], Nader Kameli[19], Abdullah Algaissi[19], Edrous Alamer[19], Abdulaziz Alhazmi[19], Amjad A.Shajri[20], Majid Ahmed Darraj[21], Bandar Kameli[22], O. O. Sufyani[23], Badreldin Rahama[20], Abrar A. Bakr[1], Fahad M. Alhoshani[4], Azzam A. Alquait[4], Ali Somily[26], Ahmed M. Albarrag[26] Lamia Alosaimi[27], Sumayh A. Aldakeel[27] Fayez S. Bahwerth [28], Essam J. Alyamani[4], Takashi Gojobori [1,32,33], Satoru Miyazaki[25], Mohammed B. Al-Fageeh[4], Robert Hoehndorf [1,2,29,30,32,33]

[1]Computational Bioscience Research Center, King Abdullah University of Science and Technology, KAUST

[2]SDAIA–KAUST Center of Excellence in Data Science and Artificial Intelligence, King Abdullah University of Science and Technology, KAUST

[3]Environmental Science and Engineering Program, Division of Biological and Environmental Science and Engineering, KAUST

[4]Wellness and Preventive Medicine Institute, Health Sector, King Abdulaziz City for Science and Technology (KACST), Riyadh, Saudi Arabia

[5]College of Applied Medical Sciences, Taibah University, Madinah, Saudi Arabia
[6]Department of Medical Laboratory Technology, College of Applied Sciences, Taibah University, Madinah, Saudi Arabia
[7]Department of biological sciences, college of science, University of Jeddah, Jeddah, Saudi Arabia.
[8]Vaccines and Immunotherapy Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[9]Department of Medical Laboratory Technology, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[10]Epidemiology department, public health Administration, king Abdullah Medical Complex, Jeddah 23816, Saudi Arabia
[11]Alnoor Specialist Hospital, Ministry of Health, Makkah, Saudi Arabia
[12]Infectious Diseases Research Department, King Abdullah International Medical Research Center (KAIMRC), Riyadh, Saudi Arabia.
[13]King Abdullah International Medical Research Center (KAIMRC)
[14]King Saud bin Abdulaziz University-Health Sciences Ministry of National Guard-Health Affairs (MNGHA)
[15]Department of Pathology and Laboratory Medicine, King Abdulaziz Medical City (KAMC),
[16]Ministry of National Guard Health Affairs (MNGHA), Riyadh, Saudi Arabia.
[17]World Health Organization, Geneva, Switzerland
[18]Medical Laboratory, King Abdulaziz Medical City (KAMC), Ministry of National Guard Health Affairs, Jeddah, Saudi Arabia
[19]Emerging and Epidemic Infectious Diseases Research Unit, Medical Research Center, Jazan University, Jazan 45142, Saudi Arabia
[20]Department of Medical Laboratories Technology, College of Applied Medical Sciences, Jazan University, Jazan, Saudi Arabia
[21]Department of Medicine,Faculty of Medicine, Jazan University, Jazan, Saudi Arabia
[22]Regional Laboratory & Central Blood Bank Jazan Health
[23]Saudi Public Health Authority, Vector-Borne Diseases Laboratory, Jazan 45142, Saudi Arabia

[24]BndrGene Medical Lab, Madinah, Saudi Arabia
[25]Faculty of Pharmaceutical Sciences, Department of Pharmacy, Tokyo University of Science, Noda, Chiba, Japan
[26]Department of pathology, college of Medicine, King Saud University and King Saud University Medical City, Riyadh, Saudi Arabia
[27]The National Center for Genomic Technology (NCGT), Life Science and Environment Research Institute, King Abdulaziz City for Science and Technology (KACST), Riyadh, Saudi Arabia
[28]Medical Microbiology Laboratory, Hera General Hospital, Makkah healthcare cluster, Makkah, Saudi Arabia.
[29]KAUST Center of Excellence for Smart Health (KCSH), King Abdullah University of Science and Technology, 4700 KAUST, Thuwal 23955, Saudi Arabia
[30]KAUST Center of Excellence for Generative AI, King Abdullah University of Science and Technology, 4700 KAUST, Thuwal 23955, Saudi Arabia
[31]Biothreat Response Department, Public Health Laboratory, the Saudi Public Health Authority
[32]Computer, Electrical and Mathematical Sciences & Engineering (CEMSE) Division, King Abdullah University of Science and Technology, King Abdullah University of Science and Technology, 4700 KAUST, Thuwal, Saudi Arabia.
[33]Biological and Environmental Sciences & Engineering (BESE) Division, King Abdullah University of Science and Technology, King Abdullah University of Science and Technology, 4700 KAUST, Thuwal, Saudi Arabia

[*]Address correspondence to: Mohammed Alarawi (mohammed.alarawi@kaust.edu.sa) and Robert Hoehndorf (robert.hoehndorf@kaust.edu.sa).

ABSTRACT

Methicillin-resistant Staphylococcus aureus (MRSA) surveillance in regions with mass gatherings presents unique challenges for public health systems. Saudi Arabia, hosting millions of pilgrims annually, provides a distinctive setting for studying how human mobility shapes bacterial populations, yet comprehensive genomic surveillance data from this region remains limited. Here, we present an integrated analysis of *S. aureus* isolates collected across seven Saudi Arabian regions, combining whole-genome sequencing with extensive antimicrobial susceptibility testing and standardized metadata capture. Our analysis revealed striking differences between pilgrimage and non-pilgrimage cities. Pilgrimage cities showed significantly higher genetic diversity and antimicrobial resistance rates, harboring numerous international strains including recognized clones from various geographic origins. While genomic prediction of antimicrobial resistance showed high accuracy for some antibiotics, particularly beta-lactams, it performed poorly for others, highlighting the ongoing need for phenotypic testing in clinical settings. Our findings demonstrate how mass gatherings shape bacterial population structures and emphasize the importance of integrated surveillance approaches in regions with significant international travel.

Introduction

Staphylococcus aureus, a gram-positive bacterium characterized by its grape-like clustering morphology, is both a common human and animal commensal organism and a significant pathogen that has emerged as a leading cause of hospital-acquired infections. Notably, methicillin-resistant *S. aureus* (MRSA) accounts for approximately 25-50% of these infections. MRSA emerged shortly after methicillin's introduction in 1958, following the acquisition of a mobile genetic element known as the staphylococcal chromosomal cassette mec (SCCmec). This cassette harbors the mec*A* gene, which encodes PBP2A, an alternative penicillin-binding protein involved in cell-wall synthesis. The modified protein exhibits reduced affinity to β-lactam antibiotics, conferring broad-spectrum resistance to this antibiotic class [1, 2].

The global impact of MRSA has expanded significantly beyond healthcare settings, establishing distinct reservoirs within communities and livestock populations [3]. Understanding the epidemiological distribution of MRSA lineages and their corresponding resistance phenotypes across communities, regions, and countries has become increasingly crucial. This knowledge directly informs the development of targeted mitigation strategies, encompassing evidence-based policies for antibiotic stewardship, systematic screening protocols, and comprehensive livestock management practices [4]. The clinical significance of MRSA is particularly evident in its role as a major pathogen in post-surgical complications, community-onset infections, and foodborne illness outbreaks [3].

Saudi Arabia presents a unique context for studying MRSA transmission and resistance patterns. As the largest country on the Arabian Peninsula by both area and population, it holds profound religious significance as the custodian of Islam's two holiest sites. This unique position draws millions of Muslim pilgrims annually, creating massive gatherings that may serve as potential vectors for microbial transmission and the introduction of novel clones and antimicrobial resistance patterns. The epidemiology of MRSA in Saudi Arabia has been documented through various investigations [5, 6, 7, 8], although most studies have been limited to individual healthcare facilities or specific geographical regions. These investigations range from comprehensive phenotypic characterization of MRSA isolates' antimicrobial susceptibility profiles [7, 9, 10] to more detailed molecular analyses of strain characteristics [6].

Despite extensive research on MRSA in Saudi Arabia, there remains a significant gap in comprehensive datasets that integrate both genotypic and phenotypic information across multiple geographical regions. Such integrated data is valuable in elucidating molecular mechanisms of antimicrobial resistance and identifying novel therapeutic targets in addition to transmission burden. Moreover, combined genotype–phenotype datasets enable the development of machine learning approaches for predicting drug resistance patterns and pathogenicity profiles, while also informing the design of targeted studies aligned with the One Health approach.

To address this knowledge gap, we established a collection of 686 *S. aureus* isolates from diverse regions across Saudi Arabia, encompassing clinical specimens, community screening samples, and wastewater isolates. We conducted whole-genome sequencing on the isolates and performed extensive drug resistance phenotyping screening, creating a novel geospatial genotype–phenotype dataset. Our analysis reveals high genetic diversity, including several previously unidentified sequence types distributed across different regions. These novel sequence types exhibit elevated patterns of drug resistance. Significantly, we observed higher genetic diversity and increased levels of drug resistance in cities associated with mass gatherings (Jeddah, Makkah, Madinah), suggesting that mass gathering may serve as drivers of antimicrobial resistance evolution and transmission. These findings provide crucial insights for developing targeted intervention strategies and underscore the importance of surveillance in regions experiencing regular mass gatherings.

Materials and Methods

Sample Selection and Sourcing

Our study includes a total of 686 *S. aureus* isolates collected from seven distinct regions across Saudi Arabia. The majority of samples were obtained from tertiary care hospitals, with the Western region contributing the largest proportion: Jeddah (275 isolates), Madinah (102 isolates), and Makkah (36 isolates). Additional regional collections include Riyadh in the Central region (158 isolates), Hail in the North (64 isolates), AlHasaa in the East (33 isolates), and Jazan in the South (18 isolates). To broaden the scope of our surveillance, we supplemented the hospital-sourced isolates with 60 samples collected through a community screening of healthy individuals in Jeddah city and 5 environmental wastewater samples from the same region. A detailed breakdown of sample distribution is provided in Supplementary Table 1. Sample selection within each hospital followed a random sampling approach, and we included both methicillin-resistant (MRSA) and methicillin-susceptible *S. aureus* (MSSA) isolates to ensure comprehensive representation of circulating strains.

Bacterial Identification and Susceptibility Testing

All isolates were initially enriched on tryptic soy agar (TSA) supplemented with 5% NaCl (Sigma-Aldrich, Germany) to selectively culture *S. aureus*. Single colonies were used for subsequent bacterial identification, susceptibility testing, and glycerol stock archival. Throughout the identification process, the ATCC 29213 strain served as an internal reference control. Phenotypic drug resistance profiling was performed using the VITEK® 2 System with GP ID cards for identification and AST-P580 cards for antimicrobial susceptibility testing (bioMérieux, France) Supplementary Table 2 shows the antimicrobial agents and their concentrations of the AST-P580 card. All test results were digitally recorded and stored for analysis.

DNA Isolation

Genomic DNA isolation was performed using single colonies obtained either directly from Petri dishes or from enriched broth cultures following identification. For each sample, 2 mL of overnight culture was pelleted for DNA extraction. Concurrently, we prepared archived samples by combining 500 µL molecular grade glycerol (Thermofisher, USA) with 500 µL of the culture for future investigations. Genomic DNA was isolated using the automated KingFisher MagMax DNA isolation kit protocol (Thermofisher, USA). DNA quality and quantity were assessed using both a nano-drop instrument (Thermofisher, USA) and Qubit dsDNA BR kit (Thermofisher, USA).

Library Preparation and DNA Sequencing

Genomic DNA libraries were constructed using 100 ng of input DNA with the QIASeq FX DNA library kit (Qiagen, Germany) following the manufacturer's protocol. Library quality was assessed using an Agilent Bioanalyzer system (Agilent, USA). Whole-genome sequencing was performed on the Illumina NovaSeq 6000 platform (Illumina, USA) using SP flow-cells, generating 150-base paired-end reads with average insert sizes ranging from 350 to 450 bp, and 50-100X sequence depth.

Genomic Analysis

We performed initial quality control and preprocessing of raw sequencing reads using TrimGalore v0.4.4 for adapter trimming [11]. We screened for taxonomic identification and contamination using Kraken v2.0.8 beta [12] and Mash v2.3 [13]. We assembled genomes using SKESA v2.4.0 [14] and called variants using Snippy v4.6.0 against the reference genome NC_007795.1.

We annotated the genomes using prokka v1.14.6 [15] and constructed the pangenome using Roary v3.13.0 [16]. We built a neighbor-joining phylogenetic tree using FastTree2 2.1.10 [17] based on the core genome alignment. We determined multilocus sequence types (MLST) using ABRicate v1.0.1 with the MLST tool.

We employed multiple approaches for antimicrobial resistance analysis. We identified drug resistance genes using both the CARD v5.1.1 [18] and ResFinder v4.0 [19] databases via ABRicate v1.0.1 [20]. We also applied DeepARG v1.0.2 [21], a deep learning approach, to enhance resistance gene detection. We identified virulence factors using the Virulence Factor Database (VFDB) [22] through ABRicate v1.0.1.

We characterized methicillin resistance mechanisms by screening for the mec*A* gene using the staphopia-Sccmec v1.0.0 typing tool [23]. We complemented this analysis using a local database containing whole cassette sequences (mec*A*, mec*B*, mec*C*), cassette recombinase genes (CCR) genes, and insertion sequences for each SCC*mec* type. We performed additional mec*A* cassette detection using minimap2 v2.24-r1122 [24] with a local database. We determined SCC*mec* type assignments based on cassette element arrangements following the workflow at https://github.com/cdnstp/SCCmec_CLA. Finally, we conducted Protein A gene polymorphism (spa) typing using spaTyper v4.6.1 [25] with its associated database.

Reproducibility and Implementation of FAIR Data Principles

We implemented our entire analysis workflow using the Common Workflow Language (CWL) [26] to ensure computational reproducibility. The two main analysis

workflows are shown in Supplementary Figures 1 and 2 . We executed all analyses on a local Arvados system [27], with complete workflow definitions and execution environments preserved. To promote Findability, Accessibility, Interoperability, and Reusability (FAIR), we have made all source code, workflow definitions, and configuration files freely available at https://github.com/bio-ontology-research-group/mrsa-sequences. The repository includes detailed documentation of computational requirements, software versions, and usage instructions. Additionally, we have deposited raw sequencing data in the NCBI Sequence Read Archive (SRA) under accession number PRJEB59751. Furthermore, we prepared the metadata of our samples as well as the measured antimicrobial resistance phenotypes using the Resource Description Framework (RDF) [28] and make it available in a public repository [29] and on a Github repository. Supplementary Figure 3 describes the RDF data model.

Results

Genetic diversity and geographic distribution of *S. aureus* in Saudi Arabia

Geolocation	Total samples	Unique CCs	Unique ST	Sccmec types rate%	PVL rate %	TSST rate %
Alhasa (East)	33	8	15	IVa 45.5; V 30.3; IVc 12.1; MSSA 6.1; VI 6.1	24.2	6
Hail (North)	64	9	19	MSSA 34.4; IVa 28.1; V 23.4; VI 6.3; IVc 4.7; IIIa 3.1	6.3	26
Jazan (South)	18	4	6	V 61.1; IVa 27.8; IVd 11.1	22.2	0
Jeddah (West)	275	9	43	V 45.1; IVa 32.4; MSSA 10.2; IVc/IVd 3.6; VI 2.9; Undetermined 0.7; VI/Vc/IVh/IIIa 0.4	13.5	8
Madina (West)	102	9	26	V 34.3; IVa 33.3; IVc 13.7; VI 6.9; MSSA 3.9; IVg 2.0; I 2.0; Others 1.0	32.4	13
Makkah (West)	36	8	14	V 41.7; IVa 38.9; MSSA 13.9; Undetermined 2.8; IVc 2.8	13.9	8
Riyadh (Central)	158	9	29	V 41.8; IVa 19.6; MSSA 18.4; IVc 8.9; IVd 4.4; VI 3.2; IVb 1.9; IVg/IVh 0.6	19.0	7

TABLE 1 Summary of MRSA statistics across different geolocations, including total counts, unique clonal complexes (CCs), unique sequence types (STs),

SCC*mec* types rate, Panton-Valentine leukocidin (PVL) rate, and toxic shock syndrome toxin (TSST) rate.

We collected 686 clinical samples of *S. aureus* from multiple regions in Saudi Arabia. We sequenced the samples, performed quality control, and confirmed them as *S aureus* through whole-genome sequencing and phenotypic testing. The 686 isolates consist of methicillin-resistant (MRSA) and methicillin-susceptible (MSSA) strains, with MSSA proportions varying significantly by region: from 3.9% in Madinah to 34.4% in Hail (Table **1**).

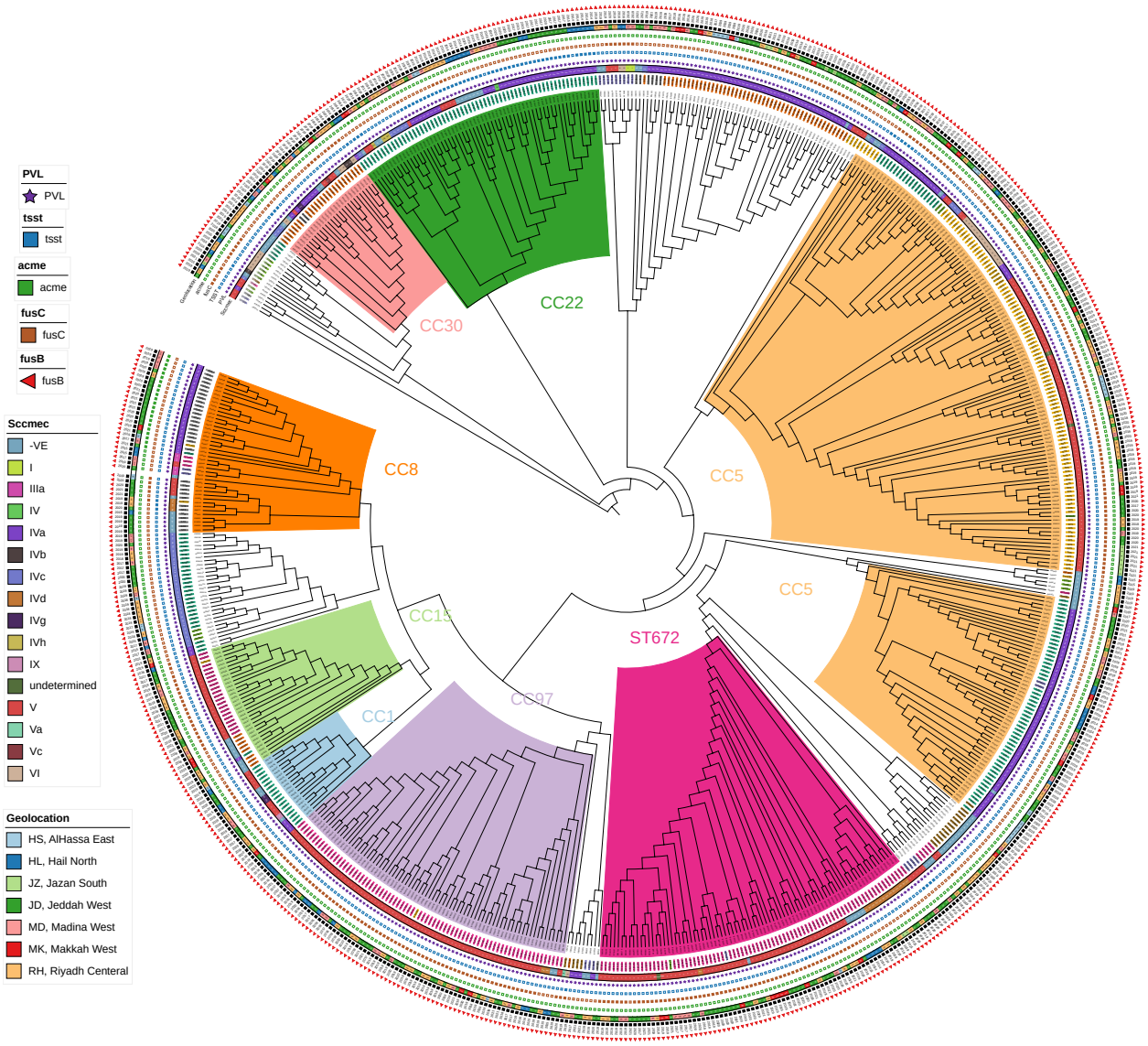


FIG 1 Maximum likelihood phylogenetic tree based on core genome SNPs of *S. aureus* isolates from Saudi Arabia. The tree highlights the distribution of major sequence types (STs) and their associated SCCmec types across different geographical regions, with CC5 emerging as the dominant clonal complex.

Molecular typing through PubMLST [30] classification was performed to understand the evolutionary relationships and population structure of these isolates, as this knowledge is crucial for tracking transmission patterns and implementing targeted infection control measures. The classification revealed nine major clonal complexes, with CC5 (26.38%), CC97 (12.39%), and CC22 (9.48%) being the most prevalent (Supplementary Figures 4 and 5 . 31.92% of isolates belonged to previously unassigned clonal complexes. Our analysis also identified 14 novel sequence types (Supplementary Table 3 . Figure **1** provides an overview of the molecular diversity and relatedness of the samples included in our study.

We observed distinct patterns of strain distribution aligned with pilgrimage routes in major cities. Cities along the Hajj and Umrah routes exhibited high genetic diversity: Jeddah (9 CCs, 43 STs), Madinah (9 CCs, 26 STs), and Makkah (8 CCs, 14 STs), likely reflecting the international convergence of travelers. In contrast, regions with less international traffic showed more limited diversity, particularly Jazan (4 CCs, 6 STs). Analysis of the SCC*mec* revealed region-specific patterns, with SCC*mec*V ranging from 23.4% in Hail to 61.1% in Jazan, and SCC*mec*IVa ranging from 19.6% in Riyadh to 45.5% in AlHasa (Table **1** and Supplementary Figure 6.

Several internationally recognized strains were observed, particularly in pilgrimage cities. The Bengal Bay clone (CC1-ST772-PVL positive) [31] and CC30-ST1482 [32] were detected primarily in Makkah and Jeddah, regions that host large numbers of international pilgrims. The South Pacific Clone (CC30-ST30) [33] was found across most cities.

Phenotypic antimicrobial resistance

Antibiotic	Riyadh	Jeddah	Makkah	Medina	Hail	Alhasa
Benzylpenicillin	98.3	97.4	100.0	97.4	95.6	88.9
Clindamycin	19.3	23.1	18.9	25.9	20.6	19.4
Erythromycin	27.8	25.6	37.8	37.9	25.0	19.4
FusidicAcid	70.5	70.9	62.2	60.9	35.3	46.9
Gentamicin	18.2	31.6	0.0	22.4	20.6	13.9
Levofloxacin	27.3	23.9	40.5	45.7	14.7	25.0
Linezolid	0.0	0.0	0.0	0.0	0.0	0.0
Moxifloxacin	21.6	19.3	33.3	41.5	13.4	25.0
Nitrofurantoin	0.0	0.9	0.0	1.7	0.0	0.0
Oxacillin	81.8	90.6	86.5	96.6	60.3	93.9
Rifampicin	0.0	0.9	0.0	1.7	0.0	0.0
Teicoplanin	0.0	0.0	0.0	0.0	0.0	0.0
Tetracycline	20.5	26.5	10.8	20.9	7.4	6.2
Tigecycline	0.0	0.0	0.0	0.0	0.0	0.0
Tobramycin	26.7	32.5	8.1	27.8	20.6	13.9
Vancomycin	0.0	0.0	0.0	0.0	0.0	0.0

TABLE 2 Antibiotic resistance rates by sampling location in percent.

We performed antimicrobial susceptibility testing on all isolates (Table 2). We observed the highest resistance rates for benzylpenicillin across all regions, ranging from 88.9% in AlHasa to 100% in Makkah. Oxacillin resistance, mediated by *mecA*, showed considerable variation, from 60.3% in Hail to 96.6% in Madinah. Fusidic acid emerged as the next most prevalent resistance, with particularly high rates in Riyadh (70.5%) and Jeddah (70.9%) but notably lower in Hail (35.3%). Fluoroquinolone resistance varied significantly by region, with levofloxacin resistance ranging from 14.7% in Hail to 45.7% in Madinah, and similar patterns for moxifloxacin (13.4% to 41.5%). Erythromycin resistance rates ranged from 19.4% in AlHasa to 37.9% in Madinah. We observed variable resistance to aminoglycosides, with tobramycin resistance ranging from 8.1% in Makkah

to 32.5% in Jeddah, while gentamicin resistance showed similar patterns. Tetracycline resistance rates varied considerably, from 6.2% in AlHasa to 26.5% in Jeddah.

Several antibiotics maintained complete effectiveness across all regions. We observed no resistance to linezolid, tigecycline, or teicoplanin. Vancomycin, often considered a last-resort antibiotic [34], showed nearly complete effectiveness, with only a single borderline case identified in a sample from Madinah with updated MIC break point. Similarly, nitrofurantoin and rifampicin maintained high effectiveness, with resistance rates not exceeding 1.7% in any region.

We also compared overall differences of antimicribial resistance in locations associated with mass gatherings to other cities. Pilgrimage-associated cities (Madinah, Makkah, and Jeddah) showed significantly higher overall resistance rates compared to other locations (33.9% vs 27.6%, $p = 0.047$, one-tailed t-test).

Genotypic identification of antimicrobial resistance

Antibiotic	CARD	Resfinder	DeepARG	Concordance
Benzylpenicillin	.983/.980/.982	.994/.572/.726	.983/.980/.982	High CARD-DeepARG
Cefoxitin	1.0/.004/.007	.984/.950/.967	-	Low
Oxacillin	.858/.976/.913	-	.858/.976/.913	High CARD-DeepARG
Gentamicin	.172/.937/.291	.752/.793/.772	.172/.937/.291	Low
Tobramycin	.220/.950/.357	.785/.807/.796	.221/.950/.358	Low CARD-DeepARG
Erythromycin	.283/1.0/.441	.776/.889/.828	.584/.895/.707	Moderate
Clindamycin	.446/.861/.588	.704/.826/.760	.378/.861/.525	Moderate
Tetracycline	.167/.962/.285	.853/.771/.810	.491/.800/.609	Low
Fusidic Acid	.937/.925/.931	.940/.845/.890	1.0/.005/.010	High CARD-Resfinder

TABLE 3 Concordance of antimicrobial resistance predictions with the observed resistance patterns using three tools (CARD Database, Resfinder 4.0, and DeepARG). Metrics shown as Precision/Recall/F-score. Dash (-) indicates that no predictions could be made by the tool for this antimicrobial agent.

We compared phenotypic resistance results to genetic predictions using three approaches: CARD [35], ResFinder [19], and the deep learning tool DeepARG [21].

Genetic predictions showed varying levels of concordance with phenotypic resistance across different antibiotic classes (Table 3).

Beta-lactam predictions demonstrated high accuracy, particularly for benzylpenicillin, where both CARD and DeepARG achieved identical high performance (precision 0.983, recall 0.980), while ResFinder showed higher precision (0.994) but lower recall (0.572). Oxacillin predictions showed strong agreement between CARD and DeepARG (precision 0.858, recall 0.976), while ResFinder did not provide predictions for this antibiotic. For fusidic acid, both CARD and ResFinder achieved accurate results (precision 0.937 and 0.940 respectively), while DeepARG could only make few predictions.

However, we observed substantial discrepancies for other antibiotic classes. Aminoglycoside resistance predictions varied markedly across tools. For gentamicin and tobramycin, CARD and DeepARG showed low precision (0.172–0.221) but high recall (0.937), while ResFinder maintained more balanced performance (precision 0.752–0.785, recall 0.793–0.807). Similarly, predictions for macrolides and lincosamides showed moderate concordance: erythromycin and clindamycin predictions shows only low concordance with phenotypic observations.

To test for the presence of known and potentially novel mechanisms of antimicrobial resistance, we constructed a pangenome (Supplementary Figure 7 and performed a pangenome-wide association study using Scoary [36] (Supplementary Table 4) . This bacterial GWAS identified both established resistance mechanisms and some novel associations. For established mechanisms, our findings confirmed several canonical resistance determinants for beta-lactam, cefoxitin, oxacillin, aminoglycoside, macrolide and lincosamide, clindamycin, and tetracycline resistance.

Additional non-canonical associations were identified for various resistance phenotypes. Fluoroquinolone resistance showed associations with *hsdM* and *entC2*. Benzylpenicillin resistance associated with *fcl_1*. Fusidic acid resistance showed associations with several transport-related genes including *natA*, *macB*, and *dppB*, this association might contribute to the resistance fitness.

Virulence and pathogenicity

Toxin profiles exhibited marked regional variation (see Figure **1** and Supplementary Table 5 . We detected Panton-Valentine leukocidin (PVL) in 17% of isolates. PVL-positive strains were more prevalent in Madinah (32.4%) and AlHasa (24.2%), with the lowest rate in Hail (6%).

Toxic shock syndrome toxin (TSST) was found in 10% of isolates and TSST rates showed an inverse geographic distribution pattern, ranging from 0% in Jazan to 26% in Hail, with intermediate rates in pilgrimage cities: Jeddah (8%), Makkah (8%), and Madinah (13%). Western cities (Jeddah, Madina, and Makkah) harbored the majority of Novel Arginine Catabolic Mobile Element (ACME)-positive isolates.

Core virulence factors displayed widespread but distinctive distribution patterns across our isolates. Adhesion factors showed varying prevalence: clumping factors (clfA, clfB) appeared in over 95% of isolates, while the collagen binding protein (cna) occurred exclusively in CC5. The majority of isolates (>90%) carried immune evasion genes (sak, sbi, scn) and biofilm formation genes (icaABCD cluster). Among toxin genes, sea dominated with 30% prevalence, followed by other enterotoxins (seb, sec, sed, seh).

Discussion

Impact of mass gatherings on MRSA diversity and resistance patterns

Our study describes the genetic epidemiology of *S. aureus* across Saudi Arabia, a country with over 32 million inhabitants of which 41% are immigrants [37]. As of 2023, Saudi Arabia hosts over 2 million pilgrims for Hajj and 27 million pilgrims for Umrah annually [38]. While previous work has investigated the effect of mass gatherings and pilgrimage on antimicrobial resistance in Saudi Arabia [39], our integrated genotype–phenotype dataset reveals several distinct patterns associated with mass gathering activities.

The impact of mass gatherings on MRSA epidemiology is particularly evident in cities along the Hajj and Umrah routes. Jeddah, Makkah, and Madina showed significantly higher genetic diversity and unique strain patterns compared to other regions, with Jeddah, the main entry point for pilgrims, exhibiting the highest number of novel sequence types. This pattern aligns with observations from other Gulf regions, where mass gatherings coincide with diverse strain distributions [40, 41]. The presence of international strains, such as the Bengal Bay clone and South Pacific Clone, predominantly in these cities further supports the role of pilgrimages in strain dissemination [42].

Our findings suggest that mass gatherings influence not only genetic diversity but also antimicrobial resistance patterns. Pilgrimage cities showed significantly higher overall resistance rates compared to other locations, suggesting these sites may serve as hotspots for resistance transmission. This is particularly evident in the distribution of SCC*mec* types, which suggests a shift from hospital-associated to community-associated MRSA, reflecting global trends [43, 44]. The concentrated presence of ACME-positive isolates in western pilgrimage cities further supports the role of international travel in strain dissemination.

The successful persistence of these strains in pilgrimage-associated regions appears to be facilitated by multiple factors. The high prevalence of adaptive features, including biofilm formation genes and immune evasion clusters, alongside various resistance mechanisms, may enable these strains to establish themselves in these high-traffic areas. These findings extend previous work on the impact of mass gatherings on bacterial populations [10], demonstrating how such events can facilitate the concurrent spread of both antimicrobial resistance determinants and virulence factors.

This unique demographic pattern and its influence on MRSA evolution and clonal expansion [3] presents both challenges and opportunities for public health interventions. Our findings suggest that pilgrimage cities may serve as sentinel sites for emerging strains and resistance patterns, potentially allowing early detection of concerning variants before they achieve wider distribution.

Limitations of Genomic Prediction for Antimicrobial Resistance

Our findings highlight the complex relationship between genotype and phenotype in antimicrobial resistance prediction. While genomic prediction tools showed high precision for certain antibiotics (over 0.98 for beta-lactams and fusidic acid), they performed poorly for other drug classes, particularly aminoglycosides. This variable performance reflects both the complexity of resistance mechanisms and the current limitations of prediction methods.

The GWAS analysis revealed both expected and unexpected genetic associations. While we confirmed established resistance determinants — such as *mecA_1* and *mecR1* for beta-lactams, and *ermC* and *msr(A)* for macrolide resistance — we also identified several unexpected associations requiring careful interpretation. For instance, fluoroquinolone resistance showed associations with *hsdM* and *entC2*, likely representing indirect relationships rather than causative mechanisms. Similarly, the association of fusidic acid resistance with transport-related genes (*natA*, *macB*, *dppB*) rather than expected target site mutations suggests potential limitations in our GWAS approach for detecting point mutations.

These discrepancies between genetic markers and phenotypic resistance highlight several challenges. First, resistance mechanisms may involve complex genetic interactions not captured by current prediction tools. Second, environmental factors and gene expression regulation may contribute to resistance phenotypes independently of genetic markers. Third, the convergence of diverse international strains in our study setting may facilitate the exchange of resistance determinants through mobile genetic elements, complicating genotype–phenotype relationships.

Given these limitations, we would suggest maintaining phenotypic testing alongside genomic surveillance for effective clinical decision-making [45]. Such a dual approach is especially crucial in regions experiencing high levels of international travel and strain diversity, where rapid evolution and transmission of resistance mechanisms may occur. Future improvements in prediction accuracy will likely require integration of multiple

data types, including transcriptomics and proteomics, to better capture the complex determinants of antimicrobial resistance.

Data Availability and Reproducibility

A key contribution of our study is the comprehensive implementation of FAIR (Findable, Accessible, Interoperable, and Reusable) principles [46] in data sharing. We structured our metadata using the Resource Description Framework (RDF) [28], creating a semantically rich representation of both sequence data and phenotypic measurements. Our data model (Supplementary Figure 3) incorporates standard ontologies including the NCBI Taxonomy [47] for species annotation, the ChEBI ontology [48] for antimicrobial agents, the NCI Thesaurus [49] and the GENEPIO ontology [50] for health status descriptions, and the Unit Ontology (UO) [51] for standardized units, enhancing interoperability with existing datasets.

This structured approach to data sharing serves multiple purposes. First, it enables direct integration with other surveillance datasets through standardized terminology and relationships. The use of controlled vocabularies for key metadata elements — such as specimen sources, host characteristics, and antimicrobial susceptibility measurements — facilitates automated data integration and comparative analyses across different studies and geographical regions.

Second, our dataset is particularly valuable for developing and validating machine learning approaches for antimicrobial resistance prediction. By providing paired genomic and phenotypic data in a standardized format, along with detailed host and environmental metadata, we enable researchers to develop prediction models that can account for contextual factors beyond purely genetic determinants, and may be more adapted to the complex demographics found among pilgrims. The comprehensive antimicrobial susceptibility data, including both MIC values and categorical interpretations, further provides rich training and validation data for such models.

The complete computational reproducibility of our analysis, implemented through a workflow in the Common Workflow Language (CWL) [52], ensures that other researchers

can not only access our data but also reproduce and build upon our analytical methods. By making our workflows, source code, and execution environments publicly available, we facilitate easier adaptation of our methods to new datasets and research questions.

Our implementation of FAIR principles extends beyond technical accessibility to practical reusability. The standardized metadata format allows researchers to integrate our findings with their local surveillance data, potentially revealing new patterns of MRSA transmission and evolution. Furthermore, our dataset can serve as a baseline for tracking changes in MRSA populations, particularly in the context of mass gatherings and international travel, where standardized data collection and sharing are crucial for effective surveillance.

Future Work

Our findings point to several key directions for future surveillance strategies, particularly focusing on early detection at pilgrimage entry points. The implementation of systematic sampling could be strategically enhanced by targeting major travel hubs. Specifically, wastewater surveillance at airports serving pilgrimage routes, particularly the international airports in Jeddah and Madinah, could provide early warning signals of emerging strains. Aircraft wastewater sampling from pilgrimage flights, which has proven effective for COVID-19 surveillance [53], could be adapted for MRSA monitoring, offering insights into strain importation patterns before pilgrims arrive at their destinations.

Such a targeted approach could be complemented by environmental sampling at key congregation points along pilgrimage routes. Strategic sampling points could include ablution facilities in major mosques and shared sanitation facilities in pilgrim accommodations, providing a comprehensive picture of strain circulation during mass gatherings. Integration of these environmental samples with clinical surveillance would create a more complete understanding of MRSA transmission dynamics during mass gathering.

The application of long-read sequencing technologies to these samples could resolve complex genetic elements that are challenging to characterize with short-read data alone. This would be particularly valuable for understanding the structure of SCCmec cassettes and other mobile genetic elements that may be exchanged during mass gatherings.

Finally, establishing a real-time surveillance system that combines this strategic sampling with rapid genomic and phenotypic analysis would enable faster response to emerging resistant strains. Such a system could serve as a model for monitoring other pathogens in similar mass gathering settings, contributing to global pathogen surveillance efforts and early warning systems for emerging antimicrobial resistance.

Conclusion

Our study demonstrates how mass gatherings and regional factors jointly shape bacterial population structures and antimicrobial resistance patterns at a national scale. While pilgrimage cities exemplify the intersection of global human mobility with pathogen evolution, the distinct patterns we observed in non-pilgrimage regions highlight the importance of local healthcare practices and potential livestock–human transmission routes in shaping MRSA populations. This comprehensive view, spanning both high-traffic pilgrimage sites and regional healthcare settings, provides evidence for a framework to understand pathogen dynamics in complex environments where international, community, and agricultural factors converge. Through standardized surveillance approaches and data sharing, our work establishes a foundation for monitoring antimicrobial resistance in the context of both global human mobility and regional One Health challenges, with implications for public health strategies across the Middle East and beyond.

Data Availability Statement

Sequencing data is available on the Sequence Read Archive (SRA) as a BioProject under accession number PRJEB59751 The phenotypes for each samples are available on Zenodo under DOI 10.5281/zenodo.14250852 [29].

ETHICS APPROVAL

This work approved by the Institutional Biosafety and Bioethics Committee (IBEC) of King Abdullah University of Science and Technology under approval number 18IBEC14.

FUNDING

This work was sponsored by King Abdulaziz City for Science and Technology under award number ETSC&KACST-KAUST-2018-05-27-01.

CONFLICTS OF INTEREST

The authors declare that no conflicts of interest exist.

References

[1] **Jevons MP**. Jan 1961. "Celbenin" - Resistant Staphylococci. *BMJ* 1 (5219):124–125. doi: 10/bg6d3h.

[2] **Katayama Y, Ito T, Hiramatsu K**. Jun 2000. A New Class of Genetic Element, Staphylococcus Cassette Chromosome mec , Encodes Methicillin Resistance in Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy* 44 (6):1549–1555. doi: 10.1128/aac.44.6.1549-1555.2000.

[3] **Lakhundi S, Zhang K**. Oct 2018. Methicillin-Resistant Staphylococcus Aureus: Molecular Characterization, Evolution, and Epidemiology. *Clinical Microbiology Reviews* 31 (4). doi:10.1128/CMR.00020-18.

[4] **Algammal AM, Hetta HF, Elkelish A, Alkhalifah DHH, Hozzein WN, Batiha GES, El Nahhas N, Mabrok MA**. Sep 2020. Methicillin-Resistant Staphylococcus aureus (MRSA): One Health Perspective Approach to the Bacterium Epidemiology, Virulence Factors, Antibiotic-Resistance, and Zoonotic Impact. *Infection and Drug Resistance* Volume 13:3255–3265. doi:10.2147/idr.s272733.

[5] **Aljeldah MM**. Mar 2020. Prevalence of Methicillin-Resistant Staphylococcus Aureus (MRSA) in Saudi Arabia: A Systematic Review. *Journal of Pure and Applied Microbiology* 14 (1):37–46. doi:10/gk35qd.

[6] **Senok A, Ehricht R, Monecke S, Al-Saedan R, Somily A**. 2016. Molecular characterization of methicillin-resistant Staphylococcus aureus in nosocomial infections in a tertiary-care facility: emergence of new clonal complexes in Saudi Arabia. *New microbes and new infections* 14:13–18.

[7] **Al Yousef S, Taha E**. 2016. Methicillin-Resistant Staphylococcus Aureus in Saudi Arabia: Genotypes distribution review. *Saudi Journal of Medicine and Medical Sciences* 4 (1):2. doi:10.4103/1658-631x.170880.

[8] **Balkhy H, Memish Z, Almuneef M, Cunningham G, Francis C, Fong K, Nazeer Z, Tannous E**. 2007. Methicillin-resistant Staphylococcus aureus: a 5-year review of surveillance data in a tertiary care hospital in Saudi Arabia. *Infection Control & Hospital Epidemiology* 28 (8):976–982.

[9] **Alkuraythi DM, Alkhulaifi MM, Binjomah AZ, Alarwi M, Aldakhil HM, Mujallad MI, Alharbi SA, Alshomrani M, Alshahrani SM, Gojobori T, et al.**. 2023. Clonal Flux and Spread of Staphylococcus aureus Isolated from Meat and Its Genetic Relatedness to Staphylococcus aureus Isolated from Patients in Saudi Arabia. *Microorganisms* 11 (12):2926.

[10] **Ahmed B, Mashat BH**. 2014. Prevalence of classical enterotoxin genes in Staphylococcus aureus isolated from food handlers in Makkah city kitchens. *Asian J. Sci. Tech* 5 (11):727–731.

[11] **Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Virk B, Dalle-Pezze P, Wingett S, Saadeh H, Ahlfors H**. 2015. Trim Galore. *Trim Galore* .

[12] **Wood DE, Lu J, Langmead B**. Nov 2019. Improved metagenomic analysis with Kraken 2. *Genome Biology* 20 (1). doi:10.1186/s13059-019-1891-0.

[13] **Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM**. Jun 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology* 17 (1). doi:10.1186/s13059-016-0997-x.

[14] **Souvorov A, Agarwala R, Lipman DJ**. Oct 2018. SKESA: strategic k-mer extension for scrupulous assemblies. *Genome Biology* 19 (1). doi:10.1186/s13059-018-1540-z.

[15] **Seemann T**. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30 (14):2068–2069.

[16] **Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J**. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31 (22):3691–3693.

[17] **Price MN, Dehal PS, Arkin AP**. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one* 5 (3):e9490.

[18] **Alcock BP, Raphenya AR, Lau TT, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen ALV, Cheng AA, Liu S, et al.**. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic acids research* 48 (D1):D517–D525.

[19] **Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AF, et al.**. 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. *Journal of Antimicrobial Chemotherapy* 75 (12):3491–3500.

[20] **Torsten S**. 2020. Abricate. GitHub https://github.com/tseemann/abricate.

[21] **Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, Zhang L**. 2018. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* 6:1–15.

[22] **Chen L, Zheng D, Liu B, Yang J, Jin Q**. Nov 2015. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Research* 44 (D1):D694–D697. doi:10.1093/nar/gkv1239.

[23] **Petit III RA, Read TD**. 2020. Bactopia: a flexible pipeline for complete analysis of bacterial genomes. *Msystems* 5 (4):10–1128.

[24] **Li H**. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34 (18):3094–3100.

[25] **Sanchez-Herrero JF, Mjsull**. 2020. spaTyper: Staphylococcal protein A (spa) characterization pipeline. Zenodo doi:10.5281/ZENODO.4063625.

[26] **Crusoe MR, Abeln S, Iosup A, Amstutz P, Chilton J, Tijanić N, Ménager H, Soiland-Reyes S, Gavrilović B, Goble C, Community TC**. May 2022. Methods included: standardizing computational reuse and portability with the Common Workflow Language. *Communications of the ACM* 65 (6):54–63. doi:10.1145/3486897.

[27] **The Arvados Authors, Amstutz P, César N, Clegg T, Di Pentima L, Kutyła D, Li J, Smith S, Vandewege W, Wait Zaranek A, Wait Zaranek S**. Arvados doi: 10.5281/zenodo.6382942.

[28] **Klyne G, Carroll JJ**. 2004. Resource Description Framework (RDF): Concepts and Abstract Syntax http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/. Retrieved 2015-03-15.

[29] **Maxat K**. 2024. Metadata of Methicillin Resistant Staphylococcus Aureus (MRSA) sequences in Saudi Arabia. Zenodo doi:10.5281/ZENODO.14250852.

[30] **Jolley KA, Bray JE, Maiden MCJ**. Sep 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Research* 3:124. doi:10.12688/wellcomeopenres.14826.1.

[31] **Nadig S, Velusamy N, Lalitha P, Kar S, Sharma S, Arakere G**. 2012. Staphylococcus aureus eye infections in two Indian hospitals: emergence of ST772 as a major clone. *Clinical ophthalmology* p 165–173.

[32] **Bakthavatchalam YD, Vasudevan K, Rao S, Varughese S, Rupali P, Gina M, Zervos M, Peter JV, Veeraraghavan B**. 2021. Genomic portrait of community-associated methicillin-resistant Staphylococcus aureus ST772-SCCmec V lineage from India. *Gene Reports* 24:101235.

[33] **Udo E, O'brien F, Al-Sweih N, Noronha B, Matthew B, Grubb W**. 2008. Genetic lineages of community-associated methicillin-resistant Staphylococcus aureus in Kuwait hospitals. *Journal of clinical microbiology* 46 (10):3514–3516.

[34] **Boneca IG, Chiosis G**. Jun 2003. Vancomycin resistance: occurrence, mechanisms and strategies to combat it. *Expert Opinion on Therapeutic Targets* 7 (3):311–328. doi: 10.1517/14728222.7.3.311.

[35] **Alcock BP, Raphenya AR, Lau TT, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen ALV, Cheng AA, Liu S, et al.**. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic acids research* 48 (D1):D517–D525.

[36] **Brynildsrud O, Bohlin J, Scheffer L, Eldholm V**. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome biology* 17:1–9.

[37] **Authority GS**. 2022. SaudCensus: Population Summary Report https://portal.saudicensus.sa/static-assets/media/content/20230531_GASTAT_Population_Report.pdf.

[38] **Authority GS**. 2023. Umrah Statistics 2023 [Accessed 23-11-2024].

[39] **Haseeb A, Saleem Z, Faidah HS, Saati AA, AlQarni A, Iqbal MS, Alghamdi S, Elrggal ME, AlGethamy M, Radwan RM, Mahrous AJ, Abuhussain SSA, Khayyat SM, Ibrahim K, Godman B, Sheikh A**. Aug 2023. Threat of Antimicrobial Resistance among Pilgrims with Infectious Diseases during Hajj: Lessons Learnt from COVID-19 Pandemic. *Antibiotics* 12 (8):1299. doi:10.3390/antibiotics12081299.

[40] **Senok A, Nassar R, Celiloglu H, Nabi A, Alfaresi M, Weber S, Rizvi I, Müller E, Reissig A, Gawlik D, et al.**. 2020. Genotyping of methicillin resistant Staphylococcus aureus from the United Arab Emirates. *Scientific reports* 10 (1):18551.

[41] **Boswihi SS, Udo EE, AlFouzan W**. 2020. Antibiotic resistance and typing of the methicillin-resistant Staphylococcus aureus clones in Kuwait hospitals, 2016–2017. *BMC microbiology* 20:1–9.

[42] **Blomfeldt A, Larssen K, Moghen A, Gabrielsen C, Elstrøm P, Aamot H, Jørgensen S**. 2017. Emerging multidrug-resistant Bengal Bay clone ST772-MRSA-V in Norway: molecular epidemiology 2004–2014. *European Journal of Clinical Microbiology & Infectious Diseases* 36:1911–1921.

[43] **David MZ, Cadilla A, Boyle-Vavra S, Daum RS**. 2014. Replacement of HA-MRSA by CA-MRSA infections at an academic medical center in the midwestern United States, 2004-5 to 2008. *PLoS One* 9 (4):e92760.

[44] **Mediavilla JR, Chen L, Mathema B, Kreiswirth BN**. 2012. Global epidemiology of community-associated methicillin resistant Staphylococcus aureus (CA-MRSA). *Current opinion in microbiology* 15 (5):588–595.

[45] **Bhattacharyya RP, Bandyopadhyay N, Ma P, Son SS, Liu J, He LL, Wu L, Khafizov R, Boykin R, Cerqueira GC, et al.**. 2019. Simultaneous detection of genotype and phenotype enables rapid and accurate antibiotic susceptibility determination. *Nature medicine* 25 (12):1858–1864.

[46] **Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, et al.**. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3.

[47] **Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, Leipe D, Mcveigh R, O'Neill K, Robbertse B, Sharma S, Soussov V, Sullivan JP, Sun L, Turner S, Karsch-Mizrachi I**. Jan 2020. NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* 2020. doi:10.1093/database/baaa062.

[48] **Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C**. Oct 2015. ChEBI in 2016: Improved services

and an expanding collection of metabolites. *Nucleic Acids Research* 44 (D1):D1214–D1219. doi:10.1093/nar/gkv1031.

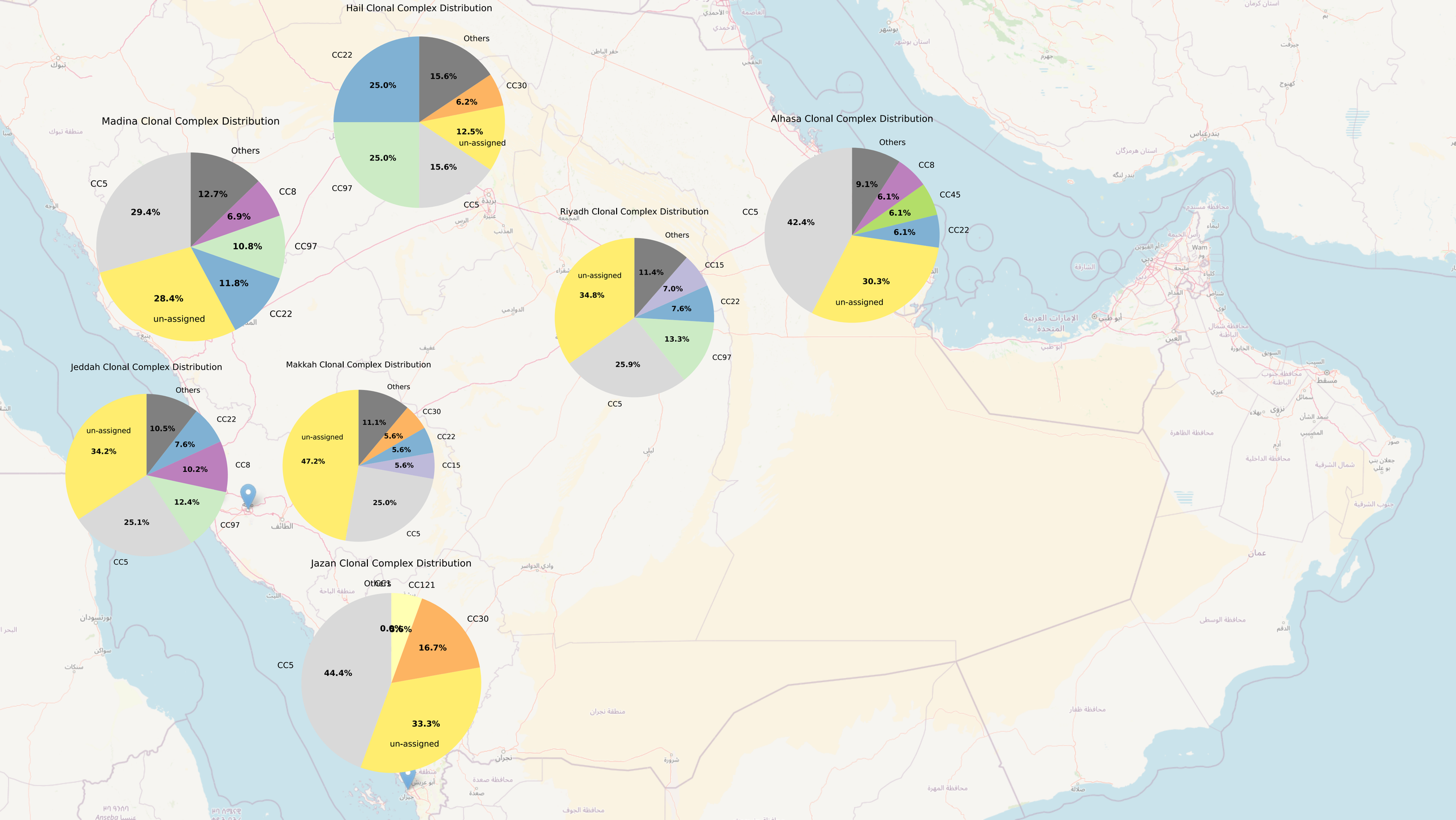
[49] **Golbeck J, Fragoso G, Hartel F, Hendler J, Oberthaler J, Parsia B**. Dec 2003. The National Cancer Institute's Thésaurus and Ontology. *Journal of Web Semantics* 1 (1):75–80. doi:10.1016/j.websem.2003.07.007.

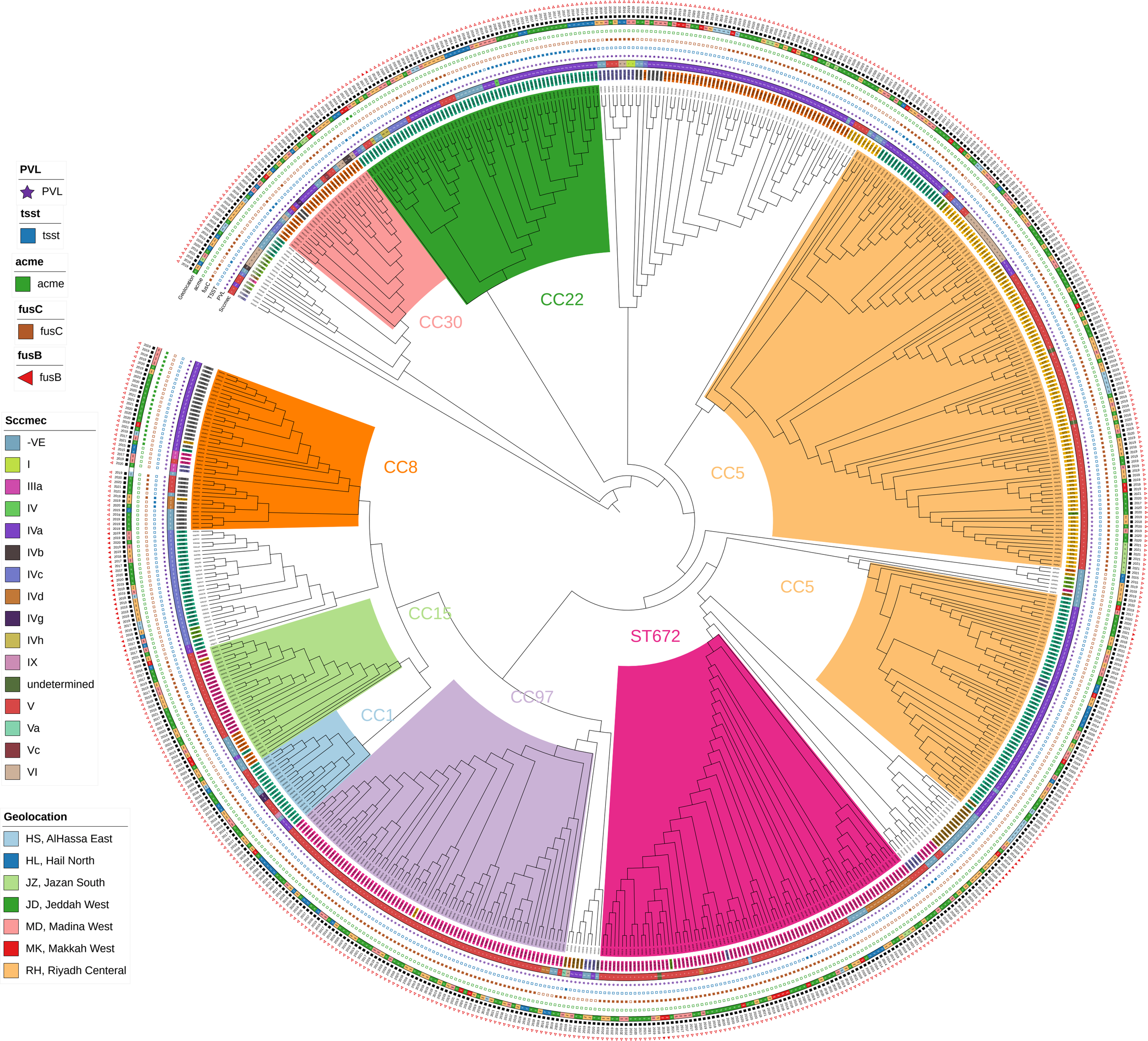
[50] **Griffiths E, Dooley D, Graham M, Van Domselaar G, Brinkman FSL, Hsiao WWL**. Jun 2017. Context Is Everything: Harmonization of Critical Food Microbiology Descriptors and Metadata for Improved Food Safety and Surveillance. *Frontiers in Microbiology* 8. doi:10.3389/fmicb.2017.01068.

[51] **Gkoutos GV, Schofield PN, Hoehndorf R**. Oct 2012. The Units Ontology: a tool for integrating units of measurement in science. *Database* 2012 (0):bas033–bas033. doi:10.1093/database/bas033.

[52] **Amstutz P, Crusoe MR, Tijanić N, Chapman B, Chilton J, Heuer M, Kartashov A, Leehr D, Ménager H, Nedeljkovich M, et al.**. 2016. Common workflow language, v1.0. .

[53] **Tay M, Lee B, Ismail MH, Yam J, Maliki D, Gin KYH, Chae SR, Ho ZJM, Teoh YL, Ng LC, et al.**. 2024. Usefulness of aircraft and airport wastewater for monitoring multiple pathogens including SARS-CoV-2 variants. *Journal of Travel Medicine* 31 (5):taae074.


Hail Clonal Complex Distribution
Others
15.6%
CC30
6.2%
12.5%
un-assigned
15.6%
CC5
25.0%
CC97
25.0%
CC22
Madina Clonal Complex Distribution
Others
12.7%
CC8
6.9%
10.8%
CC97
11.8%
CC22
28.4%
un-assigned
29.4%
CC5
Riyadh Clonal Complex Distribution
Others
11.4%
CC15
7.0%
CC22
7.6%
13.3%
CC97
25.9%
CC5
un-assigned
34.8%
Alhasa Clonal Complex Distribution
Others
9.1%
CC8
6.1%
CC45
6.1%
CC22
6.1%
30.3%
un-assigned
42.4%
CC5
Jeddah Clonal Complex Distribution
Others
10.5%
CC22
7.6%
10.2%
CC8
12.4%
CC97
25.1%
CC5
un-assigned
34.2%
Makkah Clonal Complex Distribution
Others
11.1%
CC30
5.6%
CC22
5.6%
CC15
5.6%
25.0%
CC5
un-assigned
47.2%
Jazan Clonal Complex Distribution
Others
CC121
CC30
16.7%
33.3%
un-assigned
44.4%
CC5


PVL
PVL
tsst
tsst
acme
acme
fusC
fusC
fusB
fusB
Sccmec
-VE
I
IIIa
IV
IVa
IVb
IVc
IVd
IVg
IVh
IX
undetermined
V
Va
Vc
VI
Geolocation
HS, AlHassa East
HL, Hail North
JZ, Jazan South
JD, Jeddah West
MD, Madina West
MK, Makkah West
RH, Riyadh Centeral
Geolocation
acme
fusC
TSST
PVL
Sccmec
CC30
CC22
CC8
CC5
CC5
CC15
ST672
CC97
CC1