Data in Brief 64 (2026) 112453

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Draft genome data analysis and pathogenicity profiling of *Staphylococcus aureus* strain IHS3A with antibiotic resistance genes isolated from a hospital in Jordan

Saqr Abushattal[a], Sulaiman M. Alnaimat[a,*], Nidal Odat[b], Mahmoud Abushattal[c]

[a] *Department of Medical Analysis, Faculty of Allied Medical Sciences, Al-Hussein Bin Talal University, Ma'an, Jordan*
[b] *Department of Applied Biological Sciences, Faculty of Science, Al-Balqa Applied University, Salt, Jordan*
[c] *Department of Genetic Engineering and Biotechnology, Faculty of Allied Medical Sciences, Middle East University, Amman, Jordan*

ARTICLE INFO

Article history:
Received 2 August 2025
Revised 19 December 2025
Accepted 2 January 2026
Available online 8 January 2026

Dataset link: Staphylococcus aureus strain IHS3A, whole genome shotgun sequencing project (Original data)

Keywords:
Multi-drug resistance
Bacterial genomics
Antimicrobial resistance
Hospital pathogens
Jordanian isolates
Resistance genes
Genome analysis

ABSTRACT

This dataset provides a comprehensive genomic and pathogenicity profiling of *Staphylococcus aureus* strain IHS3A, a methicillin-resistant (MRSA) clinical isolate obtained from a healthcare worker in a teaching hospital in Jordan, Middle East. Whole genome sequencing was performed using the Illumina NextSeq 2000 platform, followed by high-quality de novo assembly using SPAdes. The genome spans 2821,373 bp across 90 contigs, with a GC content of 32.78%, and demonstrates high-quality metrics, including 99.67% completeness and minimal contamination (0.08%). The genome analysis identified 2611 predicted protein-coding sequences. Multilocus sequence typing (MLST) assigned the isolate to ST10647, SCC*mec* typing revealed type IVc (2B), and spa typing identified t131. The dataset includes comprehensive annotations of key antimicrobial resistance genes, such as *mecA* (methicillin resistance), *blaZ* (penicillin resistance), and *lmrS* (macrolide efflux), as well as virulence factors related to adherence (e.g., *atl, clfA*), immune evasion (e.g., *scn, adsA*), secretion systems (e.g., *esaA, esaB*), and toxins (e.g., *hla, lukF-PV, tsst*). Secondary metabolite biosynthetic gene clusters, such as staphylofer-

* Corresponding author.
E-mail address: s_alnaimat@ahu.edu.jo (S.M. Alnaimat).

https://doi.org/10.1016/j.dib.2026.112453

rin B and staphylopine, were identified. The genome also encodes a diverse carbohydrate-active enzyme (CAZyme) profile. These genomic data are valuable for further research on MRSA evolution, resistance mechanisms, and virulence factors in Jordan and the Middle East. The genome data have been deposited in the NCBI database under the accession number JBPPGA000000000, with a direct URL to data: https://www.ncbi.nlm.nih.gov/nuccore/JBPPGA000000000.1. Bioproject: PRJNA1283614, Biosample: SAMN49700843.

Specifications Table

Subject	Microbiology: Bacteriology.
Specific subject area	Bacterial genomics; Antimicrobial resistance; Virulence profiling.
Type of data	Table, Chart, Graph, Figure
Data collection	Strain IHS3A was isolated from a skin swab on MSA and identified via catalase/coagulase tests. DNA was extracted using the G-spin Total DNA Extraction Mini Kit (iNtRON, Korea); quality was assessed with the Nabi UV/Vis Nano Spectrophotometer (MicroDigital, Korea). Sequencing was conducted on the Illumina NextSeq 2000 (EzBiome, USA). Genome assembly used SPAdes v3.13.0; annotation via PGAP and BV-BRC; functional/genomic analysis used MiGA, EzBiome Genome-ID, and Protologger.
Data source location	The bacterial sample was collected at the Jordan University Teaching Hospital in Amman, Jordan (32.007623, 35.874776))
Data accessibility	Repository name: NCBI GenBank Data identification number: (The draft genome sequence of *S. aureus* strain IHS3A was deposited in NCBI GenBank under accession JBPPGA000000000.) Direct URL to data: https://www.ncbi.nlm.nih.gov/nuccore/JBPPGA000000000.1 Bioproject: PRJNA1283614 Biosample: SAMN49700843
Related research article	https://doi.org/10.18502/ijm.v16i5.16791 https://doi.org/10.29333/ejgm/16775

1. Value of the Data

- The high-quality draft genome sequence of methicillin-resistant *S. aureus* subsp. *aureus* (MRSA) strain IHS3A, isolated from a healthcare worker in a Jordanian teaching hospital, represents a novel sequence type (ST10647) as confirmed by the PubMLST database (isolate ID 51,931). MLST assigned the isolate to ST10647 (allelic profile: *arcC*-484, *aroE*-3, *glpF*-1, *gmk*-14, *pta*-11, *tpi*-51, *yqiL*-10), while rMLST classified it as rST4549. Core-genome MLST (cgMLST) identified the isolate as cgST29201, confirming its distinct genomic clustering within the *S. aureus* population.
- The strain carries SCC*mec* type IVc (2B) and belongs to spa type t131, both characteristic of community-acquired MRSA lineages. These combined genotypic features establish IHS3A as a key regional reference for understanding the genomic diversity and evolution of MRSA strains in the Middle East.
- The assembly demonstrates excellent quality metrics, including 99.67 % completeness, minimal contamination (0.08 %), and an N50 value of 218,515 bp, with a total genome size of 2.82 Mb, comprising 90 contigs, and 2611 predicted proteins. These data provide a solid basis for downstream comparative and functional genomics analyses.
- The coexistence of multiple high-priority antimicrobial resistance genes (e.g., *mecA, blaZ, ermC, tet*(38), *lmrS, norA*) and a broad array of virulence factors (including *lukF-PV, icaA, esaA*, and the capsule gene cluster *cap8*) enhances understanding of this strain's adaptation, resistance mechanisms, and pathogenic potential.

- The pathogenicity profiling confirms the presence of numerous adherence, enzyme, immune evasion, secretion system, and toxin genes, supporting the strain's capacity to cause invasive infections and evade host immune defenses.
- The genome also encodes secondary metabolite biosynthetic gene clusters (e.g., for staphyloferrin B, staphylopine, aureumisimine), and a diverse CAZyme profile, providing insights into metabolic versatility and ecological fitness.
- The data are valuable for comparative genomics, epidemiological tracking, and antimicrobial resistance monitoring of emerging hospital-associated MRSA lineages. They also serve as a resource for developing diagnostic tools, resistance prediction algorithms, and infection control strategies relevant to public health.

2. Background

S. aureus is an opportunistic Gram-positive bacterium that causes a wide range of human diseases, from mild to severe skin infections to life-threatening conditions, especially involving antibiotic-resistant strains. Among the most serious life-threatening conditions are those associated with sepsis, endocarditis, and pneumonia [1]. Tackling such diseases is challenging due to the rising prevalence of diverse antibiotic resistance genes, which are spread via mobile genetic elements such as SCC*mec*, plasmids, and transposons, thereby undermining the effectiveness of antibiotic treatments against many *S. aureus* strains [2]. These genes and mobile genetic elements facilitate the horizontal transfer of antibiotic resistance genes in *S. aureus*, enabling rapid adaptation to antimicrobial agents and the emergence of multidrug-resistant strains [3]. SCC*mec* carries the *mecA* gene, which confers methicillin resistance, while plasmids and transposons disseminate additional resistance genes such as *blaZ, erm*, and *tetM*, exacerbating treatment challenges in both community and hospital settings [4,5]. In our study, we provide the draft genome sequence data of *S. aureus* strain IHS3A, isolated in Jordan.

3. Data Description

3.1. Genomic features and phylogenetic analysis

The assembled genome of *S. aureus* strain IHS3A has a total length of 2821,373 bp (Fig. 1), distributed across 90 contigs. The GC content of the genome is 32.78 %, and the longest sequence measures 413,872 bp. The N50 value of the genome assembly is 218,515 bp, indicating the quality of the assembly (Table 1). Contig NODE 25 (2.55 kb) carries the macrolide-resistance gene *ermC* together with a complete *repL*(pDLK1) replicon (100 % identity and 100 % coverage), a combination that is characteristic of small autonomous *S. aureus* plasmids. The genome quality analysis shows 99.67 % completeness and minimal contamination at 0.08 %. The total number of predicted coding genes is 2690, along with 58 tRNA genes, 3 rRNA genes, and 83 pseudo genes. The consistency of the assembly is also evaluated, with fine consistency recorded at 99.4 %. Gene prediction data shows a total of 2611 predicted proteins, with an average protein length of 301.37 amino acids and a coding density of 84.23 %. Functional analysis identified 190 transporters, 26 secretion genes, 727 unique enzymes. In terms of essential genes, 105 out of 106 essential genes were found, with 4 genes having multiple copies. One essential gene, *glyQ,* was missing. The UBCG recovery rate is 100 %, with all 92 core genes successfully recovered (Table 1). MLST assigned the isolate to ST10647 (*arcC*-484, *aroE*-3, *glpF*-1, *gmk*-14, *pta*-11, *tpi*-51, *yqiL*-10), rMLST to rST4549, and cgMLST to cgST29201. The genome also harbors SCC*mec* type IVc (2B) and spa type t131. The CAZyme analysis reveals that strain IHS3A harbors a total of 132 enzymes, distributed across five families: 5 Carbohydrate Esterases, 27 Carbohydrate-Binding Modules, 44 Glycoside Hydrolases, 55 Glycoside Transferases, and 1 Polysaccharide Lyase (Fig. 2). Subsystem analysis revealed the distribution of genes across various subsystems: 541 genes as-

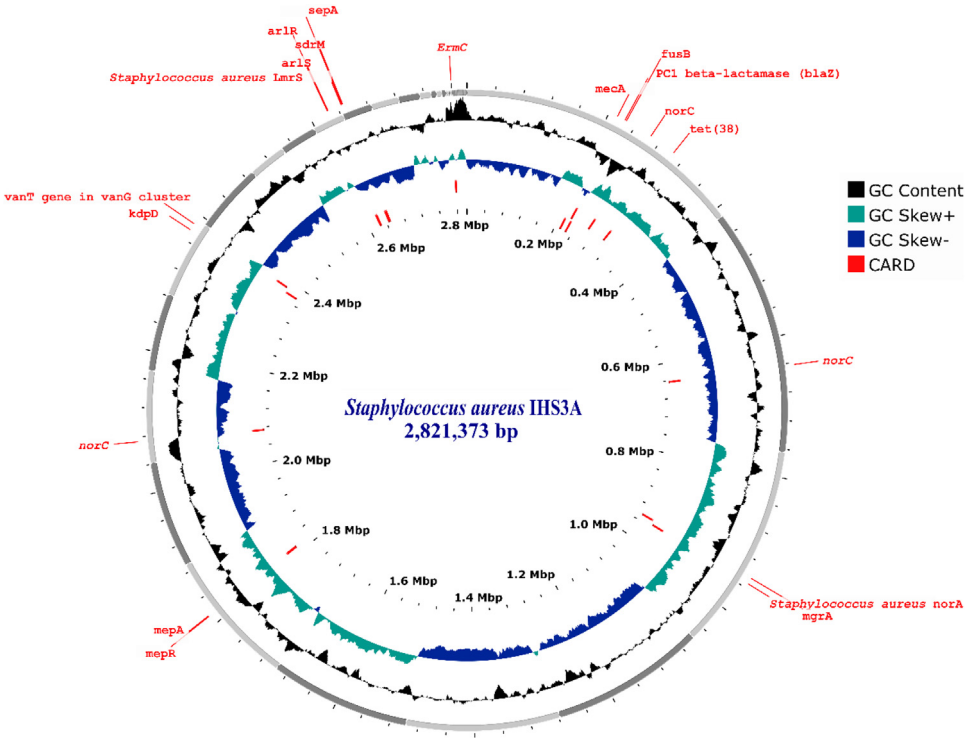

Fig. 1. Circular genome map of *S. aureus* IHS3A (2821,373 bp) generated using Proksee [6]. Concentric tracks from outermost to innermost show: (1) CARD resistance gene results (red), highlighting key resistance genes such as *mecA*, *norC*, *tet*(38), and *LmrS*, (2) GC content variation (black), (3) GC skew analysis (green/blue), (4) annotated functional regions including antibiotic resistance genes (*mecA*, *mepA*, *fusB*, *norC*) and virulence markers (*sdrM*, *arlS*), and (5) contig boundaries.

sociated with Metabolism (83 subsystems), 214 genes involved in Protein Processing (40 subsystems), 146 genes related to Stress Response, Defense, Virulence (33 subsystems), 179 genes involved in Energy (26 subsystems), 100 genes related to DNA Processing (20 subsystems), 89 genes in Cellular Processes (15 subsystems), 53 genes in RNA Processing (13 subsystems), and 63 genes associated with Membrane Transport (11 subsystems). Additional subsystems include 16 genes related to Regulation and Cell Signaling (4 subsystems), 4 genes in Miscellaneous (3 subsystems), and 7 genes in the Cell Envelope (2 subsystems) (Fig. 3).

Average Nucleotide Identity (ANI) analysis identified the input genome of strain IHS3A as being a strain of *S. aureus*, with an ANI value of 98.9513 % to *S. aureus* strain DSM 20,231 (GCF_001027105.1).

3.2. Pathogenicity

Virulence gene detection in *S. aureus* strain IHS3A based on VFDB analysis categorized the virulence factors into adherence, enzymes, immune evasion, secretion systems, and toxins (Table 2). Adherence factors include *atl*, *efb*, *fnbA*, *fnbB*, *icaA*, *icaB*, *icaC*, *icaD*, *icaR*, *sdrC*, *sdrD*, *sdrE*, and *spa*. Enzyme factors include *sspB*, *sspC*, *sspA*, *splA*, *splB*, *splC*, *splD*, *splF*, and *sak*. Immune evasion factors include *adsA*, *scn*, and *sbi*. Secretion system genes include *esaA*, *esaB*, *essB*, *essC*, *esxA*, *esxB*, and *esxD*. Toxin genes include *hlgA*, *hlgB*, *hlgC*, *lukD*, *lukF-PV*, and *lukS-PV*. The correspond-

Table 1
Integrated genomic profile of strain IHS3A based on multi-source data: The Microbial Genomes Atlas (MiGA) webserver [7]. Bacterial Bioinformatics Resource Center (BV-BRC) [8]. NCBI Prokaryotic Genome Annotation Pipeline (PGAP) [9], functional analysis via Galaxy Protologger [10].

Category	Feature	Value	Source(s)
Genome Assembly	Contigs	90	BV-BRC
	Total length	2821,373 bp	BV-BRC
	N50	218,515	MiGA, BV-BRC
	Longest sequence	413,872 bp	MiGA
	$G + C$ content	32.78	BV-BRC
	G-C skew	2.1726 %	MiGA
	A-T skew	1.6291 %	MiGA
Genome Quality	Completeness	99.67 %	Protologger
	Contamination	0.08 %	Protologger
	Coarse Consistency	100.0 %	BV-BRC
	Fine Consistency	99.4 %	BV-BRC
Gene Prediction	Predicted proteins	2611	MiGA
	Average protein length	301.3661 aa	MiGA
	Coding density	84.2324 %	MiGA
	CDS	2773	NCBI PGAP
	tRNA	58	BV-BRC
	rRNA	3	BV-BRC
	Genes (total)	2834	NCBI PGAP
	Genes (coding)	2690	NCBI PGAP
	Genes (RNA)	61	NCBI PGAP
	Pseudo Genes (total)	83	NCBI PGAP
Essential Genes	Essential genes found	105/106	MiGA
	Multiple Copies	4	MiGA
	Missing Genes	1 (*glyQ*)	MiGA
	UBCG* Recovery	100.0 % (92/92)	EzBiome Genome-ID
Functional Analysis	Number of transporters	190	Protologger
	Number of secretion genes	26	Protologger
	Number of unique enzymes	727	Protologger
	CAZymes	131	Protologger

*UBCG: Universal Bacterial Core Gene.

ing GenBank Locus Tags and Protein IDs, only for genes present in the GenBank annotation, are included in Table 2.

Table 3 presents the antibiotic resistance genes identified in strain IHS3A based on the CARD analysis. The identified resistance genes include *kdpD* (aminoglycoside efflux), *mecA* (methicillin resistance via antibiotic target replacement), *LmrS* (macrolide efflux), *arlS* (fluoroquinolone efflux), *norC* (fluoroquinolone efflux), *mepA* (glycylcycline efflux), and *tet(38)* (tetracycline efflux). Additional resistance genes such as *sdrM, norA*, and *blaZ* (penicillin inactivation) were also detected. The analysis also found genes like *ErmC* (macrolide target alteration), *fusB* (fusidane target protection), and *sepA* (disinfectant efflux). Genes involved in fluoroquinolone resistance, such as *arlR, mgrA*, and *vanT* (glycopeptide target alteration) were identified as well.

3.3. Biosynthetic gene cluster (BGC) identification

The antiSMASH analysis of strain IHS3A identified several secondary metabolite regions. These include Region 1.1 (terpene), Region 1.2 (NI-siderophore), Region 1.3 (NRPS), Region 2.1 (opine-like metallophore), Region 2.2 (T3PKS), Region 4.1 (lanthipeptide-class-I), Region 8.1 (terpene-precursor), and Region 16.1 (NI-siderophore). Each region is associated with specific secondary metabolite biosynthesis, with some regions showing high similarity to known clusters such as Staphyloferrin B, Aureumisimine A, Staphylopine, and Hyicin 3682. Regions with similarity to known clusters in MIBiG 3.1 are illustrated in Fig. 4.

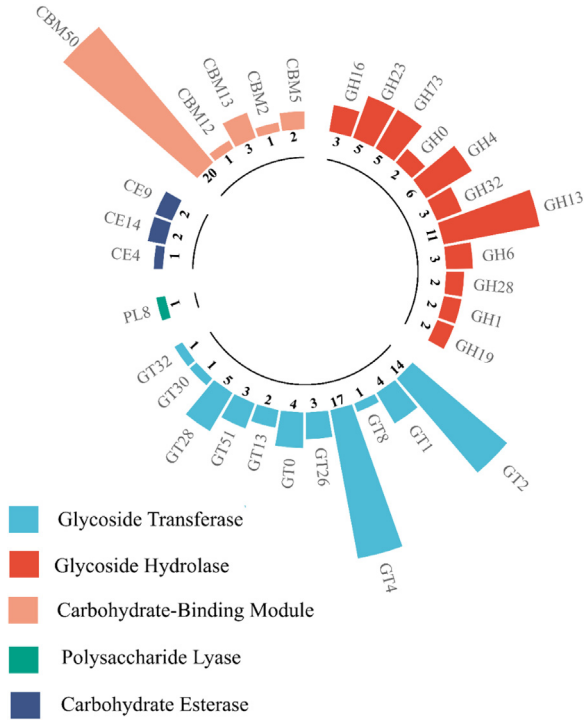

Fig. 2. Circular bar plot of CAZyme families identified in *S. aureus* strain IHS3A, created using SRplot [11]. The chart displays the distribution of various CAZyme families, including Glycoside Hydrolases (GH), Glycoside Transferases (GT), Carbohydrate-Binding Modules (CBM), Carbohydrate Esterases (CE), and Polysaccharide Lyases (PL). The numbers adjacent to each family represent the count of genes belonging to each category, with color-coded segments indicating the enzyme type.

4. Experimental Design, Materials and Methods

4.1. Bacterial isolation and identification

The strain IHS3A was isolated from a skin swab collected from a healthcare worker at the Jordan University Teaching Hospital, which includes physicians, nurses, and other staff members. The sample was collected as part of a broader study conducted between September and mid-November 2021, with informed consent obtained from the participant. The study was approved by the Research Ethics Committee at the University of Jordan Hospital (approval date: 14 September 2021, reference number: 285/2021). The sample was cultured on Mannitol Salt Agar (MSA) and incubated at 37 °C for 24–48 h under standard laboratory conditions. Presumptive colonies showing characteristic MRSA morphology were subcultured on Blood Agar, then on Nutrient Agar to obtain a pure culture. Biochemical tests, including catalase and coagulase, were conducted for phenotypic identification. The confirmed pure isolate was stored in Nutrient Broth supplemented with 80 % glycerol at −80 °C for long-term preservation.

4.2. Whole genome sequencing, assembly, annotation, and feature analysis

Genomic DNA from strain IHS3A was extracted using the G-spin Total DNA Extraction Mini Kit (iNtRON Biotechnology, Suwon, Korea), following the standard protocol for Gram-positive

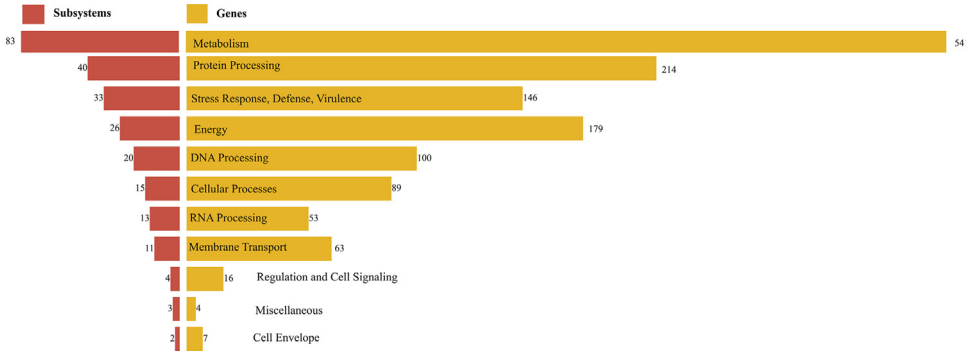

Fig. 3. Butterfly bar chart illustrating the distribution of genes in *S. aureus* strain IHS3A across various subsystems classified by biological functions, created using SRplot [11]. The chart highlights the number of genes associated with each subsystem, including Protein Processing, Stress Response, Defense, Virulence, Energy, DNA Processing, Cellular Processes, RNA Processing, Membrane Transport, Regulation and Cell Signaling, Miscellaneous, and Cell Envelope.

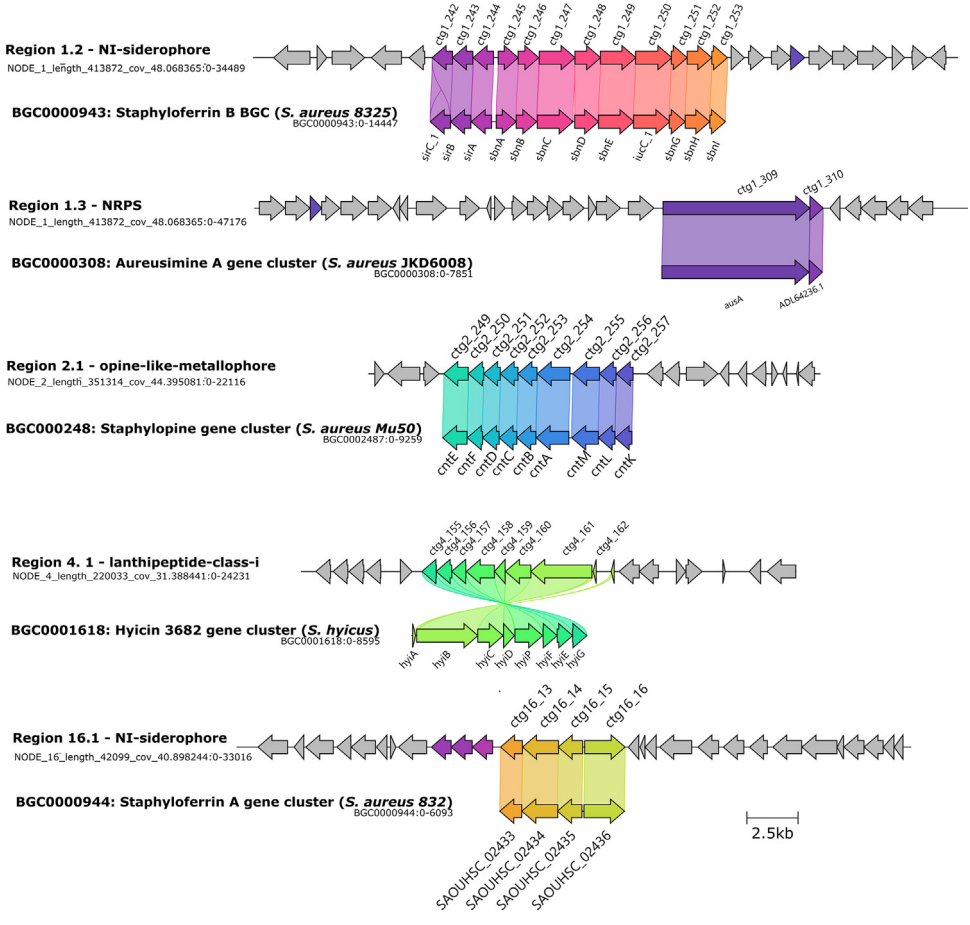

Fig. 4. Biosynthetic Gene Clusters (BGCs) identified strain IHS3A using antiSMASH 8.0 [14], highlighting regions with similarity to known clusters in MIBiG 3.1 [15]. Cluster comparisons were performed using Clinker [16]. All gene names are shown except for BGC0000944 (Staphyloferrin A), where only locus tags are indicated.

Table 2
Virulence factors and related genes in *S. aureus* strain IHS3A based on VFDB analysis [12]. The table lists key virulence factors categorized by adherence, enzymes, immune evasion, secretion systems, and toxins, along with the associated genes, GenBank Locus Tags, and Protein IDs. Only genes present in the GenBank annotation are shown; pseudogenes and unannotated genes are excluded.

VF class	Virulence factors	Related genes	GenBank Locus Tag	Protein ID
Adherence	Autolysin	*atl*	ACSEL7_08010	MGK9014278.1
	Fibrinogen binding protein	*efb*	ACSEL7_11,870	MGK9015016.1
	Fibronectin binding proteins	*fnbA*	ACSEL7_03315	MGK9013365.1
		fnbB	ACSEL7_03305	MGK9013364.1
	Intercellular adhesin	*icaA*	ACSEL7_00580	MGK9012841.1
		icaB	ACSEL7_00590	MGK9012843.1
		icaC	ACSEL7_00595	MGK9012844.1
		icaD	ACSEL7_00585	MGK9012842.1
		icaR	ACSEL7_00575	MGK9012840.1
	Ser-Asp rich fibrinogen-binding proteins	*sdrC*	ACSEL7_03850	MGK9013470.1
		sdrD	ACSEL7_03855	MGK9013471.1
		sdrE	ACSEL7_03860	MGK9013472.1
	Staphylococcal protein A	*spa*	ACSEL7_01250	MGK9012965.1
Enzyme	Cysteine protease	*sspB*	ACSEL7_07980	MGK9014272.1
		sspC	ACSEL7_07975	MGK9014271.1
	Serine V8 protease	*sspA*	ACSEL7_07985	MGK9014273.1
	Serine protease	*splA*	ACSEL7_05870	MGK9013867.1
		splB	ACSEL7_05865	MGK9013866.1
		splC	ACSEL7_05860	MGK9013865.1
		splD	ACSEL7_05845	MGK9013863.1
		splF	ACSEL7_05840	MGK9013862.1
	Staphylokinase	*sak*	ACSEL7_10,685	MGK9014792.1
Immune evasion	AdsA	*adsA*	ACSEL7_00930	MGK9012908.1
	SCIN	*scn*	ACSEL7_10,695	MGK9014793.1
	Sbi	*sbi*	ACSEL7_02910	MGK9013288.1
Secretion system	Type VII secretion system	*esaA*	ACSEL7_08435	MGK9014359.1
		esaB	ACSEL7_08445	MGK9014360.1
		essB	ACSEL7_08450	MGK9014361.1
		essC	ACSEL7_08455	MGK9014362.1
		esxA	ACSEL7_08430	MGK9014358.1
		esxB	ACSEL7_08465	MGK9014364.1
		esxD	ACSEL7_08475	MGK9014366.1
	Gamma hemolysin	*hlgA*	ACSEL7_02915	MGK9013289.1
		hlgB	ACSEL7_02925	MGK9013291.1
		hlgC	ACSEL7_02920	MGK9013290.1
	Leukotoxin D	*lukD*	ACSEL7_05925	MGK9013878.1
	Panton-Valentine leukocidin	*lukF-PV*	ACSEL7_10,195	MGK9014695.1
		lukS-PV	ACSEL7_10,200	MGK9014696.1

bacteria. DNA quality and concentration were assessed using the Nabi UV/Vis Nano Spectrophotometer (MicroDigital, South Korea). Sequencing was performed on the Illumina NextSeq 2000 platform (paired-end 150 bp, 15 million reads per sample) at EzBiome Inc. (Gaithersburg, Maryland, USA). Raw read quality was evaluated with MultiQC v1.11 [17], and *de novo* assembly was carried out using SPAdes v3.13.0 [18].

Genome annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) [9] and the Bacterial Bioinformatics Resource Center (BV-BRC) [8]. Additional genomic features were analyzed using Microbial Genomes Atlas (MiGA) [7], EzBiome Genome-ID [19], and Galaxy Protologger [10]. Plasmid-associated sequences were identified using PlasmidFinder v2.1.6 (https://usegalaxy.eu/) with default parameters. A circular genome map was generated using the Proksee web server [6]. Carbohydrate-active enzymes were identified using the CAZy database via Protologger [10], and results were visualized with SRplot [11]. Secondary metabolite biosynthetic gene clusters were predicted using antiSMASH v8.01 [14] with default parameters. Clinker was used for comparative cluster analysis [16].

Table 3
Antibiotic resistance genes in strain IHS3A identified by the CARD analysis [13].

Gene	Contig	Identity %	Bitscore	Resistance Mechanism	Drug Class	Length (bp)
kdpD	11	99 %	1799.6	antibiotic efflux	aminoglycoside antibiotic	2730
mecA	1	99.7 %	1340.1	antibiotic target replacement	penam	2007
lmrS	16	99.4 %	951.8	antibiotic efflux	macrolide antibiotic	1446
arlS	15	99.6 %	899.6	antibiotic efflux	fluoroquinolone antibiotic	1356
norC	1	98.7 %	899.8	antibiotic efflux	fluoroquinolone antibiotic	1389
mepA	7	99.6 %	882.9	antibiotic efflux	glycylcycline	1356
tet(38)	1	99.8 %	871.7	antibiotic efflux	tetracycline antibiotic	1359
sdrM	16	99.8 %	869.4	antibiotic efflux	fluoroquinolone antibiotic	1344
norA	3	100 %	765.4	antibiotic efflux	fluoroquinolone antibiotic	1167
norC	9	70.6 %	644.0	antibiotic efflux	fluoroquinolone antibiotic	1521
norC	2	59.2 %	533.1	antibiotic efflux	fluoroquinolone antibiotic	1401
blaZ	1	95 %	511.5	antibiotic inactivation	penam	846
ermC	25	99.6 %	491.5	antibiotic target alteration	macrolide antibiotic	735
arlR	15	100 %	442.6	antibiotic efflux	fluoroquinolone antibiotic	660
fusB	1	100 %	430.3	antibiotic target protection	fusidane antibiotic	642
sepA	16	96.2 %	302.4	antibiotic efflux	disinfecting agents	474
mgrA	3	100 %	301.2	antibiotic efflux	fluoroquinolone antibiotic	444
mepR	7	100 %	285.8	antibiotic efflux	glycylcycline	420
vanT	11	33.3 %	183.0	antibiotic target alteration	glycopeptide antibiotic	1149

4.3. Genome-Based taxonomy and molecular typing

The taxonomic status of strain IHS3A was established using Protologger [10], which identified the closest related strains. The sequence type (ST) and allelic profiles were determined using the PubMLST *S. aureus* database (https://pubmlst.org/bigsdb?db=pubmlst_saureus_seqdef). The *mecA* gene and SCC*mec* element were identified using SCCmecFinder v2.0 (https://cge.food.dtu.dk/services/SCCmecFinder/), and spa typing was performed with spaTyper v1.0 (https://cge.food.dtu.dk/services/spaTyper/).

4.4. Identification of pathogenicity and resistance markers

Antibiotic resistance genes were identified using the Comprehensive Antibiotic Resistance Database (CARD) [13]. Virulence factors were detected using the Virulence Factor Database (VFDB) [12].

Limitations

The dataset is based on a single MRSA isolate (IHS3A) from a healthcare worker at a teaching hospital in Jordan, and is the only isolate subjected to full genome sequencing in this broader study. As such, it may not reflect the full genomic diversity of MRSA strains in the region. While the genome assembly is of high quality (99.67 % completeness, 0.08 % contamination), it remains a draft with 90 contigs. The large number of contigs and the shorter lengths of some of these contigs may limit the completeness and continuity of the genome assembly. Future work should include sequencing additional isolates to better capture the diversity of MRSA strains for comparative analyses

Ethics Statement

This study involved the isolation of *S. aureus* strain IHS3A from a skin swab collected from a healthcare worker after obtaining informed consent. The study was approved by the Research Ethics Committee at the University of Jordan Hospital on 14 September 2021 (Reference No 285/2021). No, the current work does not involve human subjects, animal experiments, or any

data collected from social media platforms. The authors have read and follow the ethical requirements for publication in Data in Brief.

CRediT Author Statement

Saqr Abushattal: Conceptualization, Methodology, Data curation, Writing - Reviewing and Editing, Software. Sulaiman M. Alnaimat: Methodology, Data curation, Writing - Reviewing and Editing, Visualization, Software. Nidal Odat: Investigation, Writing - Reviewing and Editing. Mahmoud Abushattal: Investigation, Writing - Reviewing and Editing.

Data Availability

Staphylococcus aureus strain IHS3A, whole genome shotgun sequencing project (Original data) (NCBI).

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to thank Dr. Esraa H. Al-Nsour from the Department of Basic Pharmaceutical Sciences, Faculty of Pharmacy, Isra University, Amman, Jordan, for her assistance in sample collection.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S.Y. Tong, J.S. Davis, E. Eichenberger, T.L. Holland, V.G. Fowler Jr., Staphylococcus aureus infections: epidemiology, pathophysiology, clinical manifestations, and management, Clin. Microbiol. Rev. 28 (3) (2015) 603–661, doi:10.1128/cmr.00134-14.

[2] D. Dewi, H.O. Khalifa, H. Khandar, J. Hisatsune, S. Kutuno, L. Yu, W. Hayashi, S. Kayama, C.E. Mason, M. Sugai, H. Suzuki, T. Matsumoto, Detection and genetic characterization of multidrug-resistant staphylococci isolated from public areas in an international airport, Sci Rep 14 (1) (2024) 27738, doi:10.1038/s41598-024-79447-3.

[3] N.A. Turner, B.K. Sharma-Kuinkel, S.A. Maskarinec, E.M. Eichenberger, P.P. Shah, M. Carugati, T.L. Holland, V.G. Fowler Jr., Methicillin-resistant Staphylococcus aureus: an overview of basic and clinical research, Nat. Rev. Microbiol. 17 (4) (2019) 203–218, doi:10.1038/s41579-018-0147-4.

[4] N. Malachowa, F.R. DeLeo, Mobile genetic elements of Staphylococcus aureus, Cell. Mol. Life Sci. 67 (18) (2010) 3057–3071, doi:10.1007/s00018-010-0389-4.

[5] T. Ito, Y. Katayama, K. Asada, N. Mori, K. Tsutsumimoto, C. Tiensasitorn, K. Hiramatsu, Structural comparison of three types of staphylococcal cassette chromosome mec integrated in the chromosome in methicillin-resistant Staphylococcus aureus, Antimicrob. Agents Chemother. 45 (5) (2001) 1323–1336, doi:10.1128/aac.45.5.1323-1336.2001.

[6] J.R. Grant, E. Enns, E. Marinier, A. Mandal, E.K. Herman, C.Y. Chen, M. Graham, G. Van Domselaar, P. Stothard, Proksee: in-depth characterization and visualization of bacterial genomes, Nucleic Acids Res. 51 (W1) (2023) W484-w492, doi:10.1093/nar/gkad326.

[7] R.L. Rodriguez, S. Gunturu, W.T. Harvey, R. Rosselló-Mora, J.M. Tiedje, J.R. Cole, K.T. Konstantinidis, The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of archaea and Bacteria at the whole genome level, Nucleic Acids Res. 46 (W1) (2018) W282-w288, doi:10.1093/nar/gky467.

[8] R.D. Olson, R. Assaf, T. Brettin, N. Conrad, C. Cucinell, J.J. Davis, D.M. Dempsey, A. Dickerman, E.M. Dietrich, R.W. Kenyon, M. Kuscuoglu, E.J. Lefkowitz, J. Lu, D. Machi, C. Macken, C. Mao, A. Niewiadomska, M. Nguyen, G.J. Olsen, J.C. Overbeek, B. Parrello, V. Parrello, J.S. Porter, G.D. Pusch, M. Shukla, I. Singh, L. Stewart, G. Tan, C. Thomas, M. VanOeffelen, V. Vonstein, Z.S. Wallace, A.S. Warren, A.R. Wattam, F. Xia, H. Yoo, Y. Zhang, C.M. Zmasek, R.H. Scheuermann, R.L. Stevens, Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR, Nucleic Acids Res. 51 (D1) (2023) D678-d689, doi:10.1093/nar/gkac1003.

[9] T. Tatusova, M. DiCuccio, A. Badretdin, V. Chetvernin, E.P. Nawrocki, L. Zaslavsky, A. Lomsadze, K.D. Pruitt, M. Borodovsky, J. Ostell, NCBI prokaryotic genome annotation pipeline, Nucleic Acids Res. 44 (14) (2016) 6614–6624, doi:10.1093/nar/gkw569.

[10] T.C.A. Hitch, T. Riedel, A. Oren, J. Overmann, T.D. Lawley, T. Clavel, Automated analysis of genomic sequences facilitates high-throughput and comprehensive description of bacteria, ISME Commun. 1 (1) (2021) 16, doi:10.1038/s43705-021-00017-z.

[11] D. Tang, M. Chen, X. Huang, G. Zhang, L. Zeng, G. Zhang, S. Wu, Y. Wang, SRplot: a free online platform for data visualization and graphing, PloS one 18 (11) (2023) e0294236, doi:10.1371/journal.pone.0294236.

[12] B. Liu, D. Zheng, S. Zhou, L. Chen, J. Yang, VFDB 2022: a general classification scheme for bacterial virulence factors, Nucleic Acids Res. 50 (D1) (2022) D912-d917, doi:10.1093/nar/gkab1107.

[13] B.P. Alcock, W. Huynh, R. Chalil, K.W. Smith, A.R. Raphenya, M.A. Wlodarski, A. Edalatmand, A. Petkau, S.A. Syed, K.K. Tsang, S.J.C. Baker, M. Dave, M.C. McCarthy, K.M. Mukiri, J.A. Nasir, B. Golbon, H. Imtiaz, X. Jiang, K. Kaur, M. Kwong, Z.C. Liang, K.C. Niu, P. Shan, J.Y.J. Yang, K.L. Gray, G.R. Hoad, B. Jia, T. Bhando, L.A. Carfrae, M.A. Farha, S. French, R. Gordzevich, K. Rachwalski, M.M. Tu, E. Bordeleau, D. Dooley, E. Griffiths, H.L. Zubyk, E.D. Brown, F. Maguire, R.G. Beiko, W.W.L. Hsiao, F.S.L. Brinkman, G. Van Domselaar, A.G. McArthur, CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database, Nucleic Acids Res. 51 (D1) (2023) D690-d699, doi:10.1093/nar/gkac920.

[14] K. Blin, S. Shaw, L. Vader, J. Szenei, Z.L. Reitz, H.E. Augustijn, J.D.D. Cediel-Becerra, V. de Crécy-Lagard, R.A. Koetsier, S.E. Williams, P. Cruz-Morales, S. Wongwas, A.E. Segurado Luchsinger, F. Biermann, A. Korenskaia, M.M. Zdouc, D. Meijer, B.R. Terlouw, J.J.J. van der Hooft, N. Ziemert, E.J.N. Helfrich, J. Masschelein, C. Corre, M.G. Chevrette, G.P. van Wezel, M.H. Medema, T. Weber, antiSMASH 8.0: extended gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation, Nucleic Acids Res. 53 (W1) (2025) W32–w38, doi:10.1093/nar/gkaf334.

[15] M.H. Medema, R. Kottmann, P. Yilmaz, M. Cummings, J.B. Biggins, K. Blin, I. de Bruijn, Y.H. Chooi, J. Claesen, R.C. Coates, P. Cruz-Morales, S. Duddela, S. Düsterhus, D.J. Edwards, D.P. Fewer, N. Garg, C. Geiger, J.P. Gomez-Escribano, A. Greule, M. Hadjithomas, A.S. Haines, E.J. Helfrich, M.L. Hillwig, K. Ishida, A.C. Jones, C.S. Jones, K. Jungmann, C. Kegler, H.U. Kim, P. Kötter, D. Krug, J. Masschelein, A.V. Melnik, S.M. Mantovani, E.A. Monroe, M. Moore, N. Moss, H.W. Nützmann, G. Pan, A. Pati, D. Petras, F.J. Reen, F. Rosconi, Z. Rui, Z. Tian, N.J. Tobias, Y. Tsunematsu, P. Wiemann, E. Wyckoff, X. Yan, G. Yim, F. Yu, Y. Xie, B. Aigle, A.K. Apel, C.J. Balibar, E.P. Balskus, F. Barona-Gómez, A. Bechthold, H.B. Bode, R. Borriss, S.F. Brady, A.A. Brakhage, P. Caffrey, Y.Q. Cheng, J. Clardy, R.J. Cox, R. De Mot, S. Donadio, M.S. Donia, W.A. van der Donk, P.C. Dorrestein, S. Doyle, A.J. Driessen, M. Ehling-Schulz, K.D. Entian, M.A. Fischbach, L. Gerwick, W.H. Gerwick, H. Gross, B. Gust, C. Hertweck, M. Höfte, S.E. Jensen, J. Ju, L. Katz, L. Kaysser, J.L. Klassen, N.P. Keller, J. Kormanec, O.P. Kuipers, T. Kuzuyama, N.C. Kyrpides, H.J. Kwon, S. Lautru, R. Lavigne, C.Y. Lee, B. Linquan, X. Liu, W. Liu, A. Luzhetskyy, T. Mahmud, Y. Mast, C. Méndez, M. Metsä-Ketelä, J. Micklefield, D.A. Mitchell, B.S. Moore, L.M. Moreira, R. Müller, B.A. Neilan, M. Nett, J. Nielsen, F. O'Gara, H. Oikawa, A. Osbourn, M.S. Osburne, B. Ostash, S.M. Payne, J.L. Pernodet, M. Petricek, J. Piel, O. Ploux, J.M. Raaijmakers, J.A. Salas, E.K. Schmitt, B. Scott, R.F. Seipke, B. Shen, D.H. Sherman, K. Sivonen, M.J. Smanski, M. Sosio, E. Stegmann, R.D. Süssmuth, K. Tahlan, C.M. Thomas, Y. Tang, A.W. Truman, M. Viaud, J.D. Walton, C.T. Walsh, T. Weber, G.P. van Wezel, B. Wilkinson, J.M. Willey, W. Wohlleben, G.D. Wright, N. Ziemert, C. Zhang, S.B. Zotchev, R. Breitling, E. Takano, F.O. Glöckner, Minimum information about a biosynthetic gene cluster, Nat. Chem. Biol. 11 (9) (2015) 625–631, doi:10.1038/nchembio.1890.

[16] C.L.M. Gilchrist, Y.H. Chooi, clinker & clustermap.Js: automatic generation of gene cluster comparison figures, Bioinformatics 37 (16) (2021) 2473–2475, doi:10.1093/bioinformatics/btab007.

[17] P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report, Bioinformatics 32 (19) (2016) 3047–3048, doi:10.1093/bioinformatics/btw354.

[18] A. Prjibelski, D. Antipov, D. Meleshko, A. Lapidus, A. Korobeynikov, Using SPAdes De Novo Assembler, Current Protocols Bioinform. 70 (1) (2020) e102, doi:10.1002/cpbi.102.

[19] M. Chalita, Y.O. Kim, S. Park, H.S. Oh, J.H. Cho, J. Moon, N. Baek, C. Moon, K. Lee, J. Yang, G.G. Nam, Y. Jung, S.I. Na, M.J. Bailey, J. Chun, EzBioCloud: a genome-driven database and platform for microbiome identification and discovery, Int. J. Syst. Evol. Microbiol. 74 (6) (2024), doi:10.1099/ijsem.0.006421.