Science of the Total Environment 895 (2023) 165078

Contents lists available at ScienceDirect

Science of the Total Environment

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

Searching for links between environmental and clinical *mecA*$^+$ *Staphylococcus aureus*: A comparative genomics study

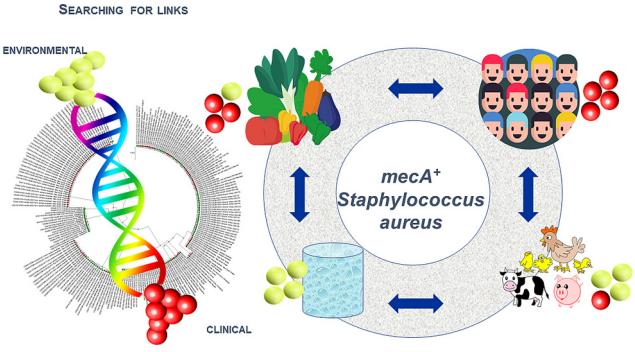
Jaqueline Rocha [a], Vanessa Silva [b,c,d,e], Patrícia Poeta [b,e], João Botelho [f], Célia M. Manaia [a,*]

[a] Universidade Católica Portuguesa, CBQF - Centro de Biotecnologia e Química Fina – Laboratório Associado, Escola Superior de Biotecnologia, Rua Diogo Botelho 1327, 4169-005 Porto, Portugal
[b] Microbiology and Antibiotic Resistance Team (MicroART), Department of Veterinary Sciences, University of Trás-os-Montes and Alto Douro (UTAD), 5000-801 Vila Real, Portugal
[c] CECAV – Veterinary and Animal Research Centre, University of Trás-os-Montes and Alto Douro (UTAD), 5000-801 Vila Real, Portugal
[d] Associate Laboratory for Animal and Veterinary Science (AL4AnimalS), University of Trás-os-Montes and Alto Douro (UTAD), 5000-801 Vila Real, Portugal
[e] Associated Laboratory for Green Chemistry (LAQV-REQUIMTE), University NOVA of Lisboa, Portugal
[f] Centro de Biotecnología y Genómica de Plantas (CBGP), Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA-CSIC), Madrid, Spain

HIGHLIGHTS

- Genomes of environmental and clinical *mecA*+ *Staphylococcus aureus* were compared.
- The same clonal lineages were observed in humans and in the environment.
- Virulence and antibiotic resistance genes were significantly different in both groups.
- Some genes may be associated with the environmental or clinical habitat.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Abasiofiok Mark Ibekwe

Keywords:
mecA$^+$ *Staphylococcus aureus*
Water
Food
Animals
Humans
Comparative genomics

ABSTRACT

Staphylococcus aureus integrate the list of highly virulent and antibiotic resistant pathogens, mainly due to the *mecA* gene, associated with methicillin resistance. Given the ubiquity of this species, the aim of this study was to investigate whether closely related *mecA*$^+$ *S. aureus* found in the environment can be also thrive as clinical isolates and if the respective accessory genome may suggest bacterial adaptation. The genomes of environmental (water, animal facilities, food products, $n = 111$) isolates were compared with closely related genomes of clinical origin (human patients, $n = 103$). These genomes, available in the public database NCBI, were analysed for phylogeny, accessory genome, and presence of selected clinically relevant genes ($n = 104$). The genomes of environmental isolates belonged to 18 multi-locus sequence types (MLSTs), 11 of which also included clinical genomes, a result confirmed based on core-genome analysis. Genes significantly ($p \leq 0.05$) more frequent among environmental genomes were related with resistance to β-lactams (*blaI*, *blaPCI*), aminoglycosides (*ant(6)-Ia*), macrolides (*mph(C)*, *erm(B)*), enterotoxins (*seg*, *sei*, *sem*, *sen*, *seo*, *seu*) and serine protease functions (*splB*), among others. Genes significantly more frequent among clinical genomes were associated with resistance to macrolides (*erm(C)*), phenicols (*fexA*), fosfomycin (*murA*), the leucocidin virulence gene (*lukS-PV*), and serine protease functions (*splA*, *splE*). It is suggested that *mecA*$^+$ *S. aureus* can be exchanged between clinical and environmental settings, with accessory traits (particularly antibiotic resistance,

* Corresponding author at: Universidade Católica Portuguesa, Escola Superior de Biotecnologia, Rua de Diogo Botelho 1327, 4169-005 Porto, Portugal.
E-mail address: cmanaia@ucp.pt (C.M. Manaia).

http://dx.doi.org/10.1016/j.scitotenv.2023.165078
Received 17 April 2023; Received in revised form 18 June 2023; Accepted 20 June 2023
Available online 24 June 2023

virulence and stress response) possibly being associated with the habitat. The interplay between phylogeny and accessory genome is an interesting contribution to better understanding the ecology and evolution of *mecA*$^+$ *S. aureus*.

1. Introduction

Some ubiquitous antibiotic resistant bacteria take advantage of the ability to live in the environment as an opportunity for dissemination and to reach environment-human interfaces, e.g. food-products or recreational areas. The species *Staphylococcus aureus* is a good example of this, as it has a wide distribution and includes opportunistic pathogens associated with conditions such as endocarditis, bacteremia, osteomyelitis or skin infections (Tong et al., 2015). The genetic diversity of *S. aureus* was originally studied with methods such as pulsed-field gel electrophoresis (PFGE), multi-locus sequence typing (MLST), or *Staphylococcus aureus* protein A (*spa*) (Enright et al., 2000; Harmsen et al., 2003). More recently, other approaches, particularly those based on genome sequencing, have provided important insights into intraspecies diversity. Single nucleotide polymorphism (SNP) analysis is a valuable genotyping method with higher resolution capacity than MLST (Huygens et al., 2004; Roisin et al., 2016). In addition, genome sequence-based analyses have the potential to elucidate aspects of ecology and adaptation that go far beyond the epidemiological assessment.

In the species *S. aureus*, the core genome genes represent approximately 75 % of the total genome (Lindsay and Holden, 2004; Planet et al., 2017). The accessory genome, which represents approximately 25 % of the genome, is mostly composed of genes associated with mobile genetic elements, such as phage related sequences, transposons, insertion sequences, plasmids, and gene cassettes (Turner et al., 2019). While phage-related sequences have been reported to harbor genes encoding virulence factors, such as enterotoxins, Panton-Valentine leucocidin (*lukPV*), and others, transposons, insertion sequences and plasmids have been associated with antibiotic resistance genes (e.g. to penicillin, *blaZ*; erythromycin, *ermC*; tetracycline, *tetK* and *tetL*; trimethoprim, *dfrA* and *dfrK*) (Malachowa and Deleo, 2010; Tong et al., 2012). Other genes observed in the accessory genome are reported to be associated with surface proteins or structures involved in host-bacterial interaction or regulatory functions (e.g. capsule, fibronectin, coagulase, serin rich adhesins), thus contributing to the opportunistic nature of *S. aureus* (Lindsay, 2010).

The first cases of methicillin resistance were reported in hospitalized patients in the 1960´s and have spread to the community since the 1990s (Tenover et al., 2006). Methicillin resistant *S. aureus,* i.e. any strain that is resistant to methicillin, oxacillin or cefoxitin, often harbors the *mecA* gene, which can be acquired by horizontal transfer of the staphylococcal cassette chromosome mec (*SCCmec*) (Katayama et al., 2000; Alcock et al., 2023). The strains harboring *SCCmec* were mostly associated with MLST clonal complexes (CC), clustered into lineages CC1, CC5, CC8, CC22, CC30 and CC45, which are widely distributed in clinical and community settings (Cockfield et al., 2007; Lindsay, 2010).

Investigating the spread of antibiotic resistant bacteria across the One Health continuum of humans, animals and the environment has been identified as a priority, not only to better understand the routes of transmission to humans, but also the existence of niches that may serve as preferential reservoir of pathogens and antibiotic resistance (ECDC, 2022). Genomics is a valuable approach to address these questions, as it can help unravel the diversity of phylogenetic lineages often observed in One Health and elucidate the accessory genome features that predominate in each compartment. This study aimed to test two hypotheses: i) that genetically closely related lineages of *mecA*$^+$ *S. aureus* can occur either in humans or in the environment; ii) that thriving in the environment or in humans may be associated with distinct features in *mecA*$^+$ *S. aureus*. To test these hypotheses, genomes of environmental and clinical *mecA*$^+$ *S. aureus* isolates, representing phylogenetically closely related lineages and with origin in different geographic areas were compared based on core and accessory genome features.

2. Materials and methods

2.1. Genome selection

The genomes selected for the study belonged to strains of environmental and clinical origins, were phylogenetically closely related, and were reported in distinct geographical regions. These criteria permitted the selection of 111 environmental and 103 clinical *mecA*$^+$ non-repetitive genomes. The genomes were selected based on a tiered approach that started with the search of *Staphylococcus aureus* genomes at the National Centre for Biotechnology Information (NCBI) public database Isolates Browser (16/12/2020). A total of 9990 genomes whose source was identified as soil, water, animals, food, or other were retrieved in the first query, being classified as environmental for the purpose of this study. This search was narrowed based on the three following criteria: 1) selection of a single genome from a set where multiple shared antimicrobial resistance and stress-related gene profiles with the same annotation, indicated the same origin (source and country) and were submitted by the same author at the same date; 2) presence of the gene *mecA* (*mecA*$^+$ *S. aureus*), and 3) the source was described as animal facility, food or water, while other described (soil, food industry surfaces or plants) or undescribed environmental sources were not included given the limited potential exposure to humans. The group of genomes selected based on those criteria (n = 111) were organized according to the Single Nucleotide Polymorphism (SNP) clusters available at the NCBI. The resultant SNP clusters were used to query the NCBI for genomes of clinical *mecA*$^+$ *S. aureus* isolates (n = 103). The workflow of *mecA*$^+$ *S. aureus* genomes selection (111 environmental and 103 clinical) (Table S1) is depicted in Fig. S1.

2.2. Phylogenetic analysis and comparative genomics

The Multi-Locus Sequence Type (MLST) based on 7 housekeeping genes (*arcC*, *aroE*, *glpF*, *yqiL*, *pta*, *gmk*, *tpi*) was determined using the MLST 2.0 (https://cge.food.dtu.dk/services/MLST/). The gene sequences were extracted from each genome, concatenated, and aligned using MEGA7 (Kumar et al., 2016) performing a total of 3187 bp. A phylogenetic tree was determined based on the Neighbor-Joining method using a 1000 bootstrap (Fig. S2). The staphylococcal protein A (*spa)* typing was determined using the spaTyper 1.0 (https://cge.food.dtu.dk/services/spaTyper/) and the staphylococcal cassette chromosome mec (*SCCmec*) typing was determined using the SCCmecFinder 1.2 (https://cge.cbs.dtu.dk/services/SCCmecFinder/).

The genomes were annotated using PROKKA version 1.14.6 (Seemann, 2014) and a pangenome analysis was performed based on the deduced amino acid sequences, using the software Roary version 3.13.0 with a similarity of 95 % (Page et al., 2015). The pangenome analysis resulted in dendrograms based on genes' presence and absence that were obtained using the roary_plots (Fig. S2). The annotation of genes that, based on Roary comparisons, were detected as exclusive of clinical or of environmental genomes, was manually confirmed. Briefly, the sequences determined as being exclusive of clinical or environmental genomes (based on Roary output, which uses PROKKA annotation), were extracted and submitted to NCBI database to confirm the annotation. This procedure reduced the number of exclusive genes detected (clinical, n = 139 to n = 60; environmental, n = 259 to n = 143). The core genes (present in ≥99 % of the genomes) were aligned using MAFFT and concatenated using Roary. The Gblock tool was used to obtain the sequences that were present in all the analysed genomes. This analysis resulted in alignments of 802,783 bp (1939 genes) for subsequent phylogenetic analysis. A phylogenetic analysis was performed using the Neighbor-Joining method and a bootstrap of 100

Table 1
Closely related lineages of clinical (C) and environmental (E) genomes.

Genome group and origin			Identical		
SNP cluster	MLST	Number of genomes / country	SCCmec cassette type	Spa type	Origin
28	1	1 E / China 1 C / China	IVg(2B)	t114	E: *n* = 1, China C: n = 1, China
14	5	6 E / USA 2 C / USA	IVa(2B)	t002	E: n = 2, USA C: n = 1, USA
1	8	1 E / Spain 29 C / Canada, USA, UK	IVa(2B)	t008	E: n = 1, Spain C: n = 13, USA, Canada, UK
12		1 E / Austria 1 C / Austria	n.d.	n.d	n.d
2	9	4 E / Germany, Czech Republic, Poland 3 C / Netherlands	n.d.	n.d	n.d
7	39	1 E / USA 4C /USA, China	II(2A)	t007	E: n = 1, USA C: n = 4, USA, China
34	93	3 E / Australia 39 C / Australia	IV(2B)	t202	E: n = 1, Australia C: n = 2, Australia
34			IVa(2B)	t202	E: n = 2, Australia C: *n* = 37, Australia
17	239	1 E / China 3 C /China	III(3A)	t969	E: n = 1, China C: n = 1, China
6	338	1 E / China 1 C / China	Vb(5C2&5)	t437	E: n = 1, China C: n = 1, China
20	398	3 E / Australia 10 C / Australia	Vc(5C2&5)	t011	E: n = 1, Australia C: *n* = 10, Australia
29		1 E / Poland 3 C / Denmark, Czech Republic	IVa(2B)	t899	E: n = 1, Poland C: n = 2, Denmark, Czech Republic
8	612	1 E / Australia 6 C / Australia	IVd(2B)	t064	E: n = 1, Australia C: n = 4, Australia
9	1535	1 E / Saudi Arabia 2 C / Saudi Arabia	V(5C2)	t328	E: n = 1, Saudi Arabia C: n = 2, Saudi Arabia

n.d. not detected.

replicates using MEGA7. In addition, the chewBBACA software (Silva et al., 2018) was used to determine the core genome based on gene-by-gene typing schemes and the software PHYLOViZ (Francisco et al., 2012) was used to perform goeBURST analysis and to visualize the relationship among the core genome gene-by-gene typing schemes and the origin of isolation, ST, *SCCmec* and *spa* type of the isolates. In summary, the core genome was determined based on two approaches, Roary and chewBBACA. The first made the comparison of the 214 genomes to determine the core genome, and the second determined the core genome based on alignments with a reference strain. Both led to slightly distinct number of genes. Each genome was screened for the presence of clinically relevant genes, including antibiotic resistance and stress-related genes, available and selected from the Isolates browser database, and VirulenceFinder 2.0 (https://cge.cbs.dtu.dk/services/VirulenceFinder/). These data were organized in a presence/absence (1/0) matrix, examined based on the simple matching similarity index using the software Primer & Permanova v6 (Primer-e, New Zealand), and expressed as a dendrogram obtained with UPGMA algorithm. Additionally, the Roary tool was used to establish a matrix of gene presence/absence, needed as input to the Scoary pipeline (Brynildsrud et al., 2016) which was used to infer possible association between genes presence and 1) clinical or environmental origins, 2) geographic origin, and 3) ST.

2.3. Statistical analysis

Statistically significant differences on the prevalence of antibiotic resistance, stress-response and virulence encoding genes between the environmental and clinical genomes was assessed using the Fisher's exact (p-value ≤ 0.05).

3. Results

3.1. Genome sequence availability and selection

Considering the hypotheses of the study, the comparison of closely related *S. aureus* of clinical and environmental origin, required phylogenetic proximity between isolates of both groups, as well as reliable metadata about genomes origin. These requirements were challenging, since ~85 % of the 9990 genomes of environmental isolates were of unknow sources, simply described as "other environmental samples" (Fig. S1). Among the environmental genomes that had reliable origin information, the selection of those with the *mecA* gene, led to a group of 111 genomes, from food ($n = 88$), animal facilities ($n = 21$) and water (n = 2) (Table S1) that according to the NCBI database were distributed by 49 SNP clusters. Thirty six of the 49 SNP clusters identified included more than one genome, while 13 were represented by a single genome (Table S1). To assure that genetically closely related strains were available for comparative analysis, the selection of genomes of clinical isolates was conditioned by the assortment of a balanced number of lineages that integrated SNP clusters defined for the environmental genomes. The 103 *mecA*$^+$ genomes of clinical isolates, selected from a total 8712, integrated 13 out of the 49 SNP clusters of genomes of environmental isolates (Table 1). The genomes integrating the 49 SNP clusters were reported in 13 countries (North America, Asia, Africa, Europe, Australia), being the environmental mostly from USA ($n = 72$) and China (=19); and the clinical mostly from USA ($n = 29$) and Australia ($n = 54$). Each of the 13 SNP clusters that integrated both environmental and clinical genomes was in general dominated by the later (Table 1).

3.2. Phylogenomic analysis: MLST and core genome

The analysis of deduced amino acid sequences in the 214 genomes revealed a total of 8119 gene sequences. Among these, 1601 sequences were identified as core genome genes (present in ≥99 % of the strains), 220 as soft core (present in ≥95 % and <99 % strains), 1365 as shell (present in ≥15 % and <95 % strains) and 4933 the cloud (present in <15 % strains). The phylogenetic analysis based on the genes present in 100 % of the genomes (802,783 bp) (Fig. 1) confirmed the relationships determined based on the MLST analysis (Fig. S2). In summary, the phylogenetic analysis based on the core genome genes grouped the STs into six major groups i) ST8, ST612 and ST239; ii) ST1535, ST1, ST9, ST840 and

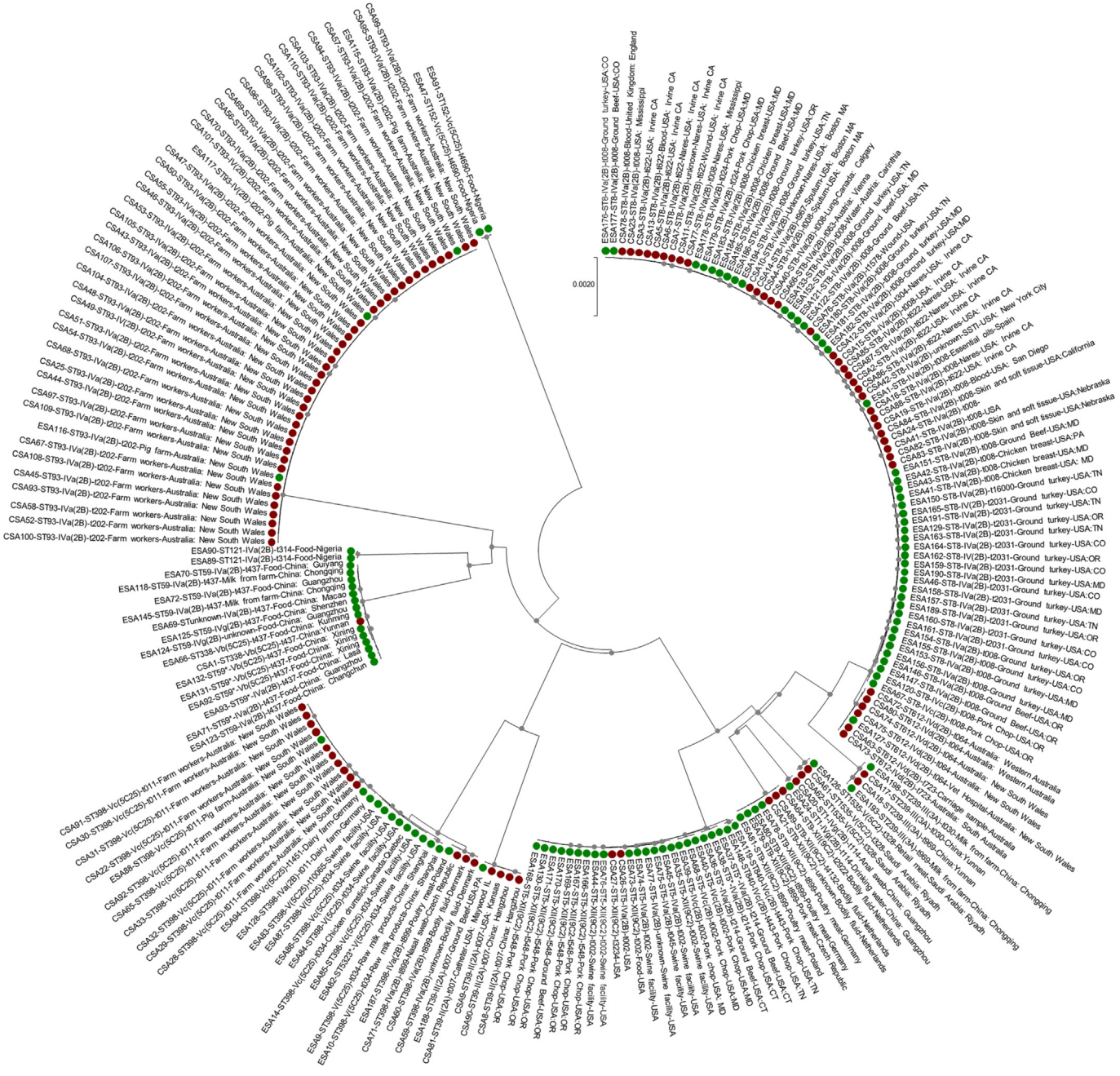


Fig. 1. Genome-based phylogenetic analysis of the 214 *mecA*$^{+}$ isolates core-genomes. Neighbor-Joining Tree based on 802,783 bp with a bootstrap of 100 replicates. Green and red symbols refer to environmental and clinical isolates, respectively. Grey circles in the nodes indicate values of bootstrap above 70 %.

ST5; iii) ST39, ST398; iv) ST59, ST338, ST59 and ST121; v) ST93; and vi) ST152. The most notorious difference between core genome- and MLST based analysis was the clustering of ST121 genomes with ST59 based or with ST39, respectively.

The genomes of clinical isolates belonged to 11 STs, all also represented by environmental isolates. Most clinical genomes were affiliated to ST93 (39/103, 38 %), ST8 (30/103, 29 %), and ST398 (13/100, 13 %) (Fig. S2, Table S1). Among the genomes of environmental isolates, ST8 (n = 43 food; n = 1 water; 44/111, 40 %) and ST398 (n = 4 food; n = 8 animal facilities; 12/111, 11 %) were also the most frequent, together with ST5 (n = 13 food; n = 8 animal facilities; 21/111, 19 %). Some environmental genomes (n = 18) belonged to STs in which clinical isolates were not represented, such as ST59 (n = 12, food, China), ST121 (n = 2, food, Nigeria), ST152 (n = 2, food, Nigeria) or ST840 (n = 2, food, USA) (Table S1). The most common *SCCmec* cassette was the *SCCmec* type IVa (2B), present in 132 of the 214 isolates (Table S1). This *SCCmec* cassette was more frequent among clinical isolates (n = 71/103), compared to the environmental isolates, found mainly in food products and animal facilities (n = 61/111) (Table S1). In general, the *SCCmec* cassettes were conserved within each and among distinct SNP clusters (e.g. cluster 1 and cluster 3) (Table S1). However, more than one *SCCmec* cassette was observed in the SNP clusters 14, 34 and 43 (Table S2). Also, most SNP clusters (40/49) yielded a single *spa* type, although exceptions were observed (clusters 1, 2, 8, 12, 13, 14, 17, 25 and 29, Table S2). The analysis of the SNP cluster, MLST, *SCCmec* cassette and spa type suggested close relationships between environmental and clinical genomes (Table 1). In most cases, these close relationships corresponded to isolates from the same country, although transnational occurrences were observed for ST8, ST39 and ST 398. However, the geographic distribution of the isolates was not homogeneous, since ST5 and ST8 genomes were reported mostly in USA,

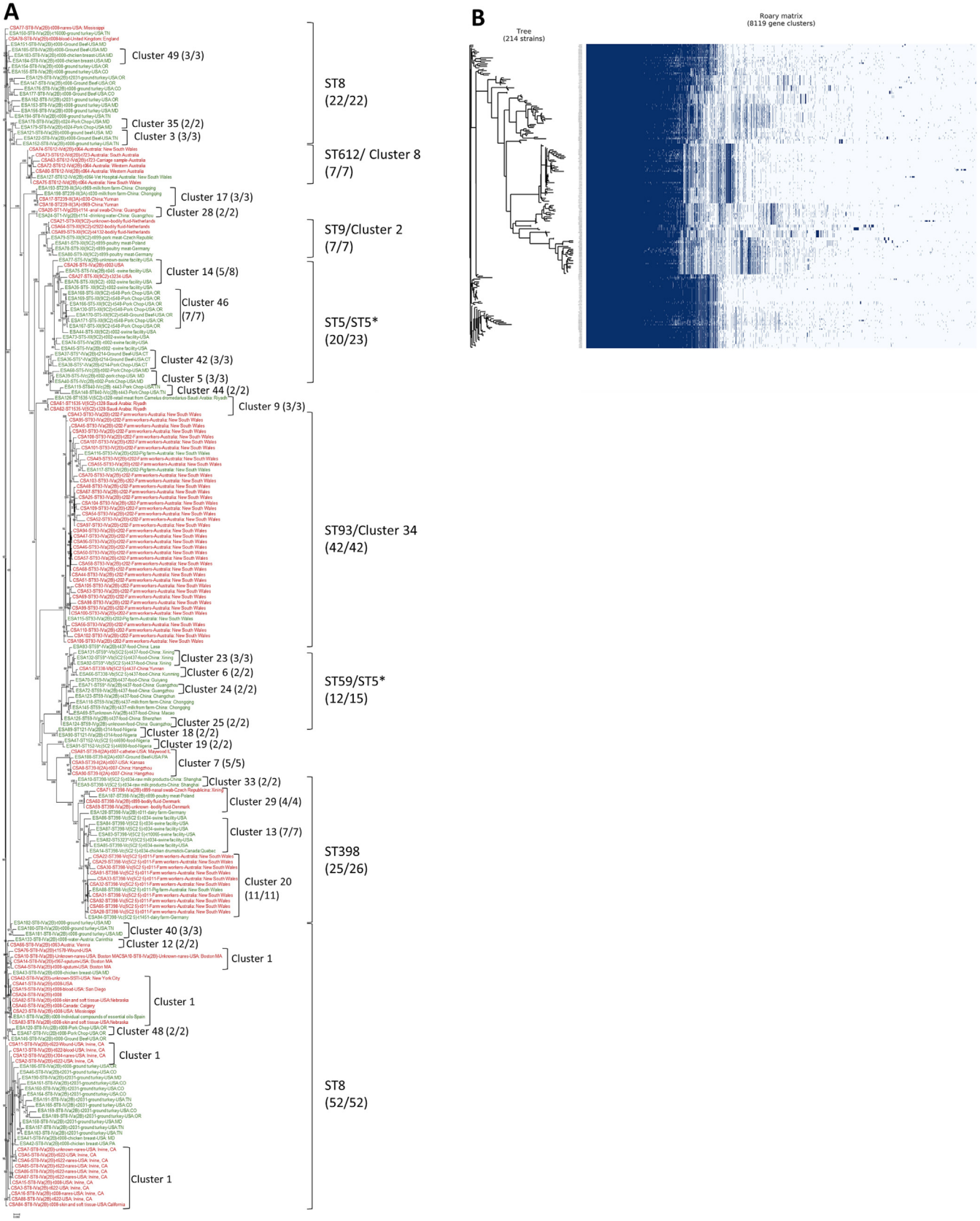


Fig. 2. Analysis of the core and accessory genome based on 8119 amino acid sequences demonstrating the A) UPGMA dendrogram representing the degree of similarity of the 214 *mecA*$^{+}$ genomes based on the amino acid sequences presence or absence and the B) the UPGMA dendrogram and the corresponding heatmap presenting the amino acid sequences presence and absence.

ST93 and ST398 genomes were exclusively or most prevalent among Australian isolates (Fig. 1). The eBURST analysis (based on the allele scheme of the genes of the core genome) revealed that most of the clinical genomes that were close to animal facility isolates, belonged to the ST5, ST93 and ST398 (Fig. 3-A and B).

3.3. Core plus accessory genome analysis

Phylogenetic relationships based on core genome genes were not fully reproduced by genomes clustering based on the presence/absence of 8119 deduced amino acid sequences (Figs. 2, S3). Based on this analysis, ST8 was split into three major subgroups, ST612 and ST239 formed a cluster that subclustered also ST1, ST9, ST5 and ST840, while ST1535, ST93, ST59, ST338, ST121, ST152, ST39 and ST398 were grouped in a common cluster. However, contrary to our initial hypothesis, this reorganization did not segregate environmental from clinical isolates' genomes (Fig. 2). Based on Scoary analysis, the pangenome composition of environmental isolates was not significantly different from that of clinical isolates (Table S3). However, the genome comparison of clinical and environmental genomes belonging to the same SNP, hence genetically closely related, resulted in the identification of exclusive gene sequences observed in either clinical or environmental groups of isolates (Fig. S4). The genes observed only in clinical genomes were related with genes mobilization (insertion sequences, transposases, recombinases, aminoglycoside resistance, among others (Table S4). The genes observed only in environmental genomes were related with adhesins, toxins, streptomycin resistance, and phages (Table S4). The comparison of the genomes of clinical, and animal facilities and food products (the ones that co-clustered with clinical genomes), showed that these groups differed in 1014 and 352 genes, respectively (Table S3).

3.4. Antibiotic resistance, virulence and stress related genes

Clinical and environmental genomes differed on the frequency of genes associated with antibiotic resistance ($n = 57$ in 58 searched), virulence ($n = 21$ in 31 searched) and stress related ($n = 1$ in 15 searched). The environmental genomes presented higher frequency of genes related with resistance to β-lactams (*blaI*, *blaPCI*), aminoglycosides (*ant(6)-Ia*), macrolides (*mph(C), erm(B)*), enterotoxin (*seg*, *sei*, *sem*, *sen*, *seo*, *seu*) and serine protease functions (*splB*), among others. In contrast, clinical genomes yielded higher percentage of genes related with resistance to macrolides (*erm(C)*), phenicols (*fexA*), fosfomycin (*murA*) or with leucocidin virulence gene (*lukS-PV*) and serine protease functions (*splA*, *splE*) were more frequent in the clinical isolates ($p \leq 0.05$) (Table S5). When the genomes were compared based on the presence/absence of antibiotic resistance, virulence and stress related genes, the clusters differed from those observed based on phylogenetic proximity or pangenome resemblance (Fig. 3). Considering a cutoff of 80 %, three major groups were formed including by i) ST9 and ST5; ii) ST8, ST1535, ST5, ST840, ST612 and ST239; and iii) ST338, ST59, ST93 and ST398. This distribution suggests that the potential clinically relevant genes are associated with the pangenome composition, although structured by the phylogenetic lineage (Fig. 3).

4. Discussion

The environmental dimensions of antibiotic resistance have received much attention from the scientific community (Baquero, 2021). The major argument behind such interest is that antibiotic resistant bacteria and their genes can move freely across distinct One Health sectors – humans, animals, and the environment (Hernando-Amado et al., 2019). However, it is still unclear if some bacterial features may favor a ubiquitous lifestyle that permits living either in the environment or in an animal or human body. It is also yet poorly understood if the environment or the human body may shape the adaptation towards distinct features. These questions have been addressed in previous studies, which suggest the role of the environment on shaping bacterial features and their evolution (Christaki et al., 2020; Liu et al., 2022; Rocha et al., 2022a, 2022b; Wright, 2019). This study aimed to understand if *mecA*$^+$ *S. aureus* occurring in natural environments can also colonize humans, which would be suggested by the observation of closely related lineages in both compartments. In addition, we aimed to understand if the environment or the human body might be associated with distinct features in *mecA*$^+$ *S. aureus*. Despite the limited availability of environmental genomes in public databases, constrained by insufficient or inadequate metadata, the results confirmed our initial hypotheses. Closely related lineages were observed in the environment and in humans, and both types of compartment may select different features in *mecA*$^+$ *S. aureus*. Among the 49 SNPs of genomes of environmental isolates that could be retrieved from NCBI, only 13 also included genomes of clinical isolates. Indeed, this may have resulted from an artifact of selecting the genomes based on SNPs, as it has been discussed that this typing method may be inaccurate to establish relationships between genomes that originate from a wide range of sources (26). In such a case, this would suggest that closely related lineages thriving in distinct habitats would display distinct SNPs. Nevertheless, genomes with identical SNP, ST, *SCCmec* cassette type and Spa type were observed in humans and environmental samples, and even in distinct countries. These observations suggest that the interchangeability of *mecA*$^+$ *S. aureus* between humans and the environment may occur regularly, as it was observed for 10 distinct STs, some of which are not among the most common.

The analysis based on core genome genes and on concatenated MLST sequences yielded equivalent results suggesting the reliability of the later method and justifying that part of the discussion relies on STs designations. Most of the *mecA*$^+$ *S. aureus* genomes analysed (161/214) belonged to ST5, ST8, ST398, and ST93. The first three, correspond to lineages representing about 90 % of known *S. aureus* genomes and are affiliated to four clonal complexes (CC5, CC8, CC398 and CC30), closely related with the respective STs (Turner et al., 2019). In our study, all ST93 genomes were from Australia and, although mostly were clinical ($n = 39/42$), it was suggested a clonal relationship with the three environmental genomes (animal facilities). This observation is agreement with the previous description of ST93 as community acquired methicillin resistant *S. aureus* associated with Australian indigenous populations (Turner et al., 2019). Other STs only observed in environmental isolates (ST59, ST121, ST152 and ST840) were probably overlooked based on SNPs search for clinical isolates, as their association to humans has been documented (https://pathogen.watch/1 March 2023).

Considering the hypothesis that the habitat may influence the genomic profile, it would be expected that genomes of the same ST but thriving under different environmental conditions would yield distinct genes. According to Liu et al. (2022), the ST is a reliable measure for the distribution of bacterial genomes and phylogenetic assessments. However, the same authors argue that when the genomes under analysis are from diverse sources, the pan-genome analysis may give better inferences about evolutionary distances (Liu et al., 2022). Although we observed a redistribution of the genomes when compared the phylogenetic analyses with the presence/absence dendrograms based on core plus accessory or on antibiotic resistance, virulence and stress related genes, in any case it separated clinical and environmental genomes. However, clinical genomes differed from closely related environmental lineages on the presence of genes related with functions such insertion sequences, transposases, recombinases or aminoglycoside resistance, while the environmental genomes differed on the presence of genes related with adhesins, toxins, streptomycin resistance or phages. Moreover, the targeted analysis of antibiotic resistance, virulence and stress related genes revealed features that were more associated to clinical or to environmental genomes. Specifically, genes encoding resistance against macrolides (*erm(C)*), phenicols (*fexA*), fosfomycin (*murA*) or virulence factors - leucocidin (*lukS-PV*) and serine protease functions (*splA*, *splE*) were more frequent in the clinical isolates. Also, genes encoding resistance to β-lactams (*blaI*, *blaPCI*), aminoglycosides (*ant(6)-Ia*) and macrolides (*mph(C), erm(B)*) or virulence factors such as enterotoxin (*seg*, *sei*, *sem*, *sen*, *seo*, *seu*) and serine protease functions (*splB*) were more frequent in the environmental isolates. These results suggest that some

Fig. 3. Clustering analysis of 214 *mecA*$^{+}$ isolates based on selected clinically relevant features (genes encoding antibiotic resistance ($n = 58$), stress ($n = 15$), and virulence ($n = 31$)). A presence/absence table supported the Simple matching similarity index estimation and UPGMA clustering analysis. Green and red labels refer to environmental and clinical isolates, respectively. Each column next to the dendrogram refer to ST, *SCCmec* and *spa* type, respectively. Similar symbols in each column represent similar isolates typing.

genomic traits associated with fitness and defense may contribute to the propagation or persistence of *S. aureus* in the environment-human nexus, as has been demonstrated for some successful epidemic clones (Chen et al., 2022).

As has been discussed in the literature (Liu et al., 2022; Turner et al., 2019), the complex interplay between genetic adaptation and environmental factors may be determinant for shaping the ecology of *S. aureus*. Genomics has an important potential for elucidating the ecology and evolution of ubiquitous antibiotic resistant bacteria and the role of the environment. However, the availability of genomes and metadata, seriously limit the implementation of broad comparative genomic studies. The enrichment of public databases with genomes of environmental isolates supported by reliable metadata (type of sample, isolation procedure, known relevant phenotypic features, among other) will be an important contribute to better understand antibiotic resistance evolution and transmission.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.165078.

Funding

This work was funded by the R&D Project CAREBIO2: Comparative assessment of antimicrobial resistance in environmental biofilms through proteomics—towards innovative theranostic biomarkers, with references NORTE-01-0145-FEDER-030101 and PTDC/SAUINF/30101/2017, financed by the European Regional Development Fund (ERDF) through the Northern Regional Operational Program (NORTE 2020) and the Foundation for Science and Technology (FCT). This work was financially supported by FEDER through project RISK.AR "Assessing the risks associated with environmental antibiotic resistant bacteria: propagation and transmission to humans" (PTDC/CTA-AMB/28196/2017) – Programa Operacional Competitividade e Internacionalização. This work was also supported by the projects UIDB/CVT/00772/2020 and LA/P/0059/2020 funded by the Portuguese Foundation for Science and Technology (FCT).

CRediT authorship contribution statement

C.M. conceptualized the work; J.R. and C.M. were involved in the methodology; J.R., J.B. and C.M. validated the work and the results; J.R. performed the investigation work; J.R. and C.M. curated the data; J.R., J.B. and C.M. prepared the original draft of the manuscript; V.S. and P.P. reviewed the manuscript; C.M. supervised the work; C.M. administered the project and C.M. and P.P. obtained funding for the research work.

Data availability

The genomes used in this study are available at the NCBI database and the accession numbers of the genomes are indicated in Table S1.

Declaration of competing interest

We declare no competing or non-financial interests.

Acknowledgements

This work was supported by National Funds from FCT - Fundação para a Ciência e a Tecnologia through project UIDB/50016/2020.

References

Alcock, B.P., Huynh, W., Chalil, R., Smith, K.W., Raphenya, A.R., Wlodarski, M.A., Edalatmand, A., Petkau, A., Syed, S.A., Tsang, K.K., Baker, S.J.C., Dave, M., McCarthy, M.C., Mukiri, K.M., Nasir, J.A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., Kwong, M., Liang, Z.C., Niu, K.C., Shan, P., Yang, J.Y.J., Gray, K.L., Hoad, G.R., Jia, B., Bhando, T., Carfrae, L.A., Farha, M.A., French, S., Gordzevich, R., Rachwalski, K., Tu, M.M., Bordeleau, E., Dooley, D., Griffiths, E., Zubyk, H.L., Brown, E.D., Maguire, F., Beiko, R.G., Hsiao, W.W.L., Brinkman, F.S.L., Van Domselaar, G., McArthur, A.G., 2023. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the comprehensive antibiotic resistance database. Nucleic Acids Res. 51, D690–D699. https://doi.org/10.1093/nar/gkac920.

Baquero, F., 2021. Threats of antibiotic resistance: an obliged reappraisal. Int. Microbiol. 24, 499–506. https://doi.org/10.1007/s10123-021-00184-y.

Brynildsrud, O., Bohlin, J., Scheffer, L., Eldholm, V., 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol. 17 (1), 1–9. https://doi.org/10.1186/s13059-016-1108-8.

Chen, F., Yin, Y., Chen, H., Jin, L., Li, S., Wang, R., Wang, S., Wang, Q., Sun, S., Wang, H., 2022. Phenotypic and genomic comparison of *Staphylococcus aureus* highlight virulence and host adaptation favoring the success of epidemic clones. mSystems 7 (6), e0083122. https://doi.org/10.1128/msystems.00831-22.

Christaki, E., Marcou, M., Tofarides, A., 2020. Antimicrobial resistance in bacteria: mechanisms, evolution, and persistence. J. Mol. Evol. Vol. 88 (issue 1), 26–40 (Springer) https://doi.org/10.1007/s00239-019-09914-3 (Springer).

Cockfield, J.D., Pathak, S., Edgeworth, J.D., Lindsay, J.A., 2007. Rapid determination of hospital-acquired meticillin-resistant *Staphylococcus aureus* lineages. J. Med. Microbiol. 56 (5), 614–619. https://doi.org/10.1099/jmm.0.47074-0.

Enright, M.C., Day, N.P.J., Davies, C.E., Peacock, S.J., Spratt, B.G., 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J. Clin. Microbiol. Vol. 38 (issue 3).. https://journals.asm.org/journal/jcm.

European Centre for Disease Prevention and Control, 2022. Antimicrobial Resistance Surveillance in Europe 2022–2020. https://doi.org/10.2900/112339.

Francisco, A.P., Vaz, C., Monteiro, P.T., Melo-Cristino, J., Ramirez, M., Carriço, J.A., 2012. SOFTWARE Open Access PHYLOViZ: Phylogenetic Inference and Data Visualization for Sequence-based Typing Methods. http://www.biomedcentral.com/1471-2105/13/87.

Harmsen, D., Claus, H., Witte, W., Rothgänger, J., Claus, H., Turnwald, D., Vogel, U., 2003. Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for spa repeat determination and database management. J. Clin. Microbiol. 41 (12), 5442–5448. https://doi.org/10.1128/JCM.41.12.5442-5448.2003.

Hernando-Amado, S., Coque, T.M., Baquero, F., Martínez, J.L., 2019. Defining and combating antibiotic resistance from one health and global health perspectives. Nat. Microbiol. 4, 1432–1442. https://doi.org/10.1038/s41564-019-0503-9.

Huygens, F., Stephens, A.J., Nimmo, G.R., Giffard, P.M., 2004. mecA locus diversity in methicillin-resistant *Staphylococcus aureus* isolates in Brisbane, Australia, and the development of a novel diagnostic procedure for the western samoan phage pattern clone. J. Clin. Microbiol. 42 (5), 1947–1955. https://doi.org/10.1128/JCM.42.5.1947-1955.2004.

Katayama, Y., Ito, T., Hiramatsu, K., 2000. A New Class of Genetic Element, Staphylococcus Cassette Chromosome Mec, Encodes Methicillin Resistance in *Staphylococcus aureus*. Vol. 44 (Issue 6).

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33 (7), 1870–1874. https://doi.org/10.1093/MOLBEV/MSW054.

Lindsay, J.A., 2010. Genomic variation and evolution of *Staphylococcus aureus*. Int. J. Med. Microbiol. Vol. 300 (issues 2–3), 98–103. https://doi.org/10.1016/j.ijmm.2009.08.013.

Lindsay, J.A., Holden, M.T.G., 2004. *Staphylococcus aureus*: superbug, super genome? Trends Microbiol. 12 (8), 378–385. https://doi.org/10.1016/j.tim.2004.06.004.

Liu, N., Liu, D., Li, K., Hu, S., He, Z., 2022. Pan-genome analysis of *Staphylococcus aureus* reveals key factors influencing genomic plasticity. Microbiol. Spectr. 10 (6), e0311722. https://doi.org/10.1128/spectrum.03117-22.

Malachowa, N., Deleo, F.R., 2010. Mobile genetic elements of *Staphylococcus aureus*. Cell. Mol. Life Sci. Vol. 67 (Issue 18), 3057–3071. https://doi.org/10.1007/s00018-010-0389-4.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31 (22), 3691–3693. https://doi.org/10.1093/bioinformatics/btv421.

Planet, P.J., Narechania, A., Chen, L., Mathema, B., Boundy, S., Archer, G., Kreiswirth, B., 2017 Feb. Architecture of a species: phylogenomics of *Staphylococcus aureus*. Trends Microbiol. 25 (2), 153–166. https://doi.org/10.1016/j.tim.2016.09.009.

Rocha, J., Ferreira, C., Mil-Homens, D., Busquets, A., Fialho, A.M., Henriques, I., Gomila, M., Manaia, C.M., 2022a. Third generation cephalosporin-resistant Klebsiella pneumoniae thriving in patients and in wastewater: what do they have in common? BMC Genomics 23 (1). https://doi.org/10.1186/s12864-021-08279-6.

Rocha, J., Henriques, I., Gomila, M., Manaia, C.M., 2022b. Common and distinctive genomic features of Klebsiella pneumoniae thriving in the natural environment or in clinical settings. Sci. Rep. 12 (1). https://doi.org/10.1038/s41598-022-14547-6.

Roisin, S., Gaudin, C., de Mendonça, R., Bellon, J., van Vaerenbergh, K., de Bruyne, K., Byl, B., Pouseele, H., Denis, O., Supply, P., 2016. Pan-genome multilocus sequence typing and outbreak-specific reference-based single nucleotide polymorphism analysis to resolve two concurrent *Staphylococcus aureus* outbreaks in neonatal services. Clin. Microbiol. Infect. 22 (6), 520–526. https://doi.org/10.1016/j.cmi.2016.01.024.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30 (14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Silva, M., Machado, M.P., Silva, D.N., Rossi, M., Moran-Gilad, J., Santos, S., Ramirez, M., Carriço, J.A., 2018. chewBBACA: a complete suite for gene-by-gene schema creation and strain identification. Microb. Genom. 4 (3). https://doi.org/10.1099/mgen.0.000166.

Tenover, F.C., McDougal, L.K., Goering, R.v., Killgore, G., Projan, S.J., Patel, J.B., Dunman, P.M., 2006. Characterization of a strain of community-associated methicillin-resistant *Staphylococcus aureus* widely disseminated in the United States. J. Clin. Microbiol. 44 (1), 108–118. https://doi.org/10.1128/JCM.44.1.108-118.2006.

Tong, A., Tong, S.Y.C., Zhang, Y., Lamlertthon, S., Sharma-Kuinkel, B.K., Rude, T., Ahn, S.H., Ruffin, F., Llorens, L., Tamarana, G., Biek, D., Critchley, I., Fowler, V.G., 2012. Panton-valentine leukocidin is not the primary determinant of outcome for Staphylococcus aureus skin infections: evaluation from the canvas studies. PLoS One 7 (5). https://doi.org/10.1371/journal.pone.0037212.

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

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

Wright, G.D., 2019. Environmental and clinical antibiotic resistomes, same only different. Curr. Opin. Microbiol. Vol. 51, 57–63 (Elsevier Ltd.) https://doi.org/10.1016/j.mib.2019.06.005 (Elsevier Ltd.).