NIVERSITY OF PÉCS

Doctoral School of Biology and Sportbiology

Deciphering the genetic architecture of methicillin-resistant *Staphylococcus aureus* clinical isolates using whole-genome sequencing

Ph.D. Thesis

Naorem Romen Singh

Supervisor:

Dr. habil. Csaba Fekete, Ph.D.,
Department of General and Environmental Microbiology,

PÉCS, 2022

UNIVERSITY OF PÉCS

1. Introduction

Staphylococcus aureus is a major human and animal pathogen, leading to cause skin and soft tissue infections to life-threatening infections (Tenover et al., 2019). *S. aureus* acquires an arsenal of antibiotic resistance genes (ARGs) and virulence factors-encoding genes (VFGs) *via* horizontal gene transfer (HGT) and recombination. This HGT event enabled the emergence of hypervirulent and multi-drug resistant (MDR) strains and led to challenging issues in antibiotic therapy (Chan et al., 2011). Consequently, the morbidity and mortality rates caused by *S. aureus* infections have a substantial impact on health concerns. Thus, the prevention and control of MDR *S. aureus* infections have become the main concern in the public health sectors of the European countries (Köck et al., 2010). This strongly suggests that there is a continuous need to search for additional drug or vaccine targets in their genomes that would improve protection and long-lasting prevention (Perumal et al., 2007).

The polyphasic characterization of *S. aureus* is essential for delineating the occurrence of an epidemic and monitoring the transmission of the organism (Du et al., 2011). In this modern era, the whole-genome sequencing technology has brought with it the possibility of investigating pathogen outbreaks, defining the different lineages, and identifying the prophages among the *S. aureus* strains and their diversity (Köser et al., 2012). Further, this technology provides an excellent opportunity to accelerate the discovery process of drug target, or vaccine candidate proteins through subtractive genome-based and reverse vaccinology (RV) approaches (Rappuoli, 2000). These approaches enable the prediction of drug and vaccine target proteins for the pathogens that are challenging to culture in the laboratory.

The number of comprehensive studies addressing *S. aureus* infections by concurrently exploring the epidemiology, and genome comparison are currently low, thus limiting global surveillance and epidemiological monitoring. This study is to understand the molecular basis of drug resistance, pathogenesis, niche-specific difference, evolutionary relationship of *S. aureus* clinical strains, prophages diversity, and identification of potential drug target and vaccine candidate proteins of *S. aureus* strains. This knowledge would allow the extrapolation of basic principles to improve diagnosis, infection control, and treatment strategies of staphylococcal disease.

2 Aim of the study

Characterization of methicillin-resistant *S. aureus* (MRSA) clinical isolates through genomic approach, and an *in-silico* identification of potential drug and vaccine candidates against methicillin-resistant *S. aureus*.

Research activities were carried out to fulfill the following key scientific objectives:

1. Characterization of MRSA clinical isolates through the polyphasic approach (phenotypic and genotypic characterizations).
2. Genome-wide comparison of MRSA clinical isolates.
3. Comparative analysis of prophages carried by human and animal-associated *S. aureus* strains spreading across the European regions.
4. Identification of potential drug targets and vaccine candidates against MRSA strains through subtractive genomics-based and reverse vaccinology approaches.

3 Materials and Methods

3.1 Collection of the isolates

In this study, a total of 35 *S. aureus* clinical isolates were used. Among these isolates, 21 isolates belonged to Hungarian strains, and 14 isolates belonged to German strains.

3.2 Polyphasic characterization of *S. aureus* clinical isolates

The phenotypic characterization of *S. aureus* clinical isolates was performed using biochemical tests (catalase, coagulase, urease, DNase production, and mannitol fermentation test) (Collee et al., 1996), and susceptible to oxacillin (1μg), cefoxitin (30μg), and erythromycin (15μg) was determined according to the Clinical and Laboratory Standards Institute (CLSI) guidance (CLSI, 2014). Biofilm formation was performed as previously described (Rahimi et al., 2016).

The genotypic characterization of *S. aureus* clinical isolates was performed using the PCR amplification method to detect the presence of *S. aureus* species-specific sequence (Martineau et al., 1998), *mecA* (Strommenger et al., 2003), and *pvl* toxin (Karahan et al. 2007) genes. Molecular typing of *SCCmec,* (Staphylococcus Chromosomal Cassette) (Zhang et al., 2005), *coa* (coagulase) (Khoshkharam-Roodmajani et al., 2014), and *spa* (Staphylococcus protein A) (Harmsen et al., 2003) were performed. The PCR amplicon of *coa* gene was further digested using the *HaeIII*

restriction enzyme and banding patterns were analyzed, and the amplified *spa* gene was sequenced using ABI PRISM 310 Genetic Analyzer (Applied Biosystems, USA). The *spa* sequence types were assigned using spaTyper. The discriminatory power (DP) of *coa-*, and *spa* typing were calculated based on Simpson's index. The detection of biofilm-encoding genes such as *icaA, icaB, icaD, icaR, fnbA, cna, clfA, clfB,* and *ebps* was performed.

3.2.1 Data setting for polyphasic approach

Individual result of the applied techniques was converted into the unweighted binary code (0, 1), and Jacquard's similarity index was generated and visualized according to the Neighbour-Joining (NJ) clustering method using Past 3.x (Hammer et al., 2001). Besides the binary data was used to perform a logistic Principal Component Analysis (PCA) in R software (RStudio Team, 2020).

3.3 Whole-genome sequencing

Based on the dendrogram and PCA plot generated from polyphasic characterization data showed that the strains originated from the same geographical region were found in the close group (SA G6 and SA G8; SA H27 and SA H32) while the other two strains originated from different geographical regions *viz*., SA G5 (German strain) and SA H29 (Hungarian strain) were found in the same group. Based on this information, these 6 *S. aureus* strains were chosen for in-depth comparative genome levels study to better understand the genomic differences among the strains.

Whole-genome sequencing was performed using Ion PGM Hi-Q View Sequencing Kit on Ion Torrent Personal Genome Machine (PGM) (Thermo Fisher Scientific Inc. USA). *In-silico* trimming of adapter and barcode sequences and data analysis were performed using Torrent Suite 5.4.0 (Thermo Fisher Scientific Inc., USA) and the trimmed paired-end reads were assembled by *de novo* assembler SPAdes 3.7.1 software (Nurk et al., 2013). Gene annotation of the genome assemblies was performed *via* the fully automated RAST (Rapid Annotation using Subsystem Technology) pipeline.

3.3.1 *In-silico* characterization of genome assemblies

In-silico epidemiologic characterization of genome assemblies was performed using a web-based server provided by the Center for Genomic Epidemiology (*CGE Server*, n.d.). *In-silico* mining of candidate ARGs and VFGs were performed using CARD

(Comprehensive Antibiotic Resistance Database) version 3.0.8 (Alcock et al., 2020), and a comprehensive set of *S. aureus* VFGs was analyzed using VFDB (Virulence Factor Database) in VFanalyzer (Liu et al., 2019). Secondary metabolite biosynthesis gene clusters and the detection of genes encoding bacteriocins were analyzed using antiSMASH 5.0 (Blin et al., 2019).

3.3.2 Comparative genome analysis

EDGAR 2.0 software framework was used to calculate, ANI, pan-genome, core-genome, and singletons (Blom et al., 2016). The functional annotation was performed using EggNOG (Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups) mapper 5.0 database (Huerta-Cepas et al., 2019) and RAST server-based SEED viewer (Overbeek et al., 2014).

3.3.3 Phylogenetic analysis

The genome assemblies of the isolates were used for a phylogenomic analysis using TYGS (Type/Strain Genome Server) (Meier-Kolthoff & Göker, 2019) engaging with genomes of closely related strains of *S. aureus*.

3.4 Comparative genomic analyses of *S. aureus* prophages

A total of 60 whole genomes of *S. aureus* strains reported to cause human and animal infections across the European regions were used in this study. Of these 60 whole genome sequences of *S. aureus* strains, 54 were retrieved from the NCBI database and additional 6 genome assemblies of *S. aureus* were current study genomes. The *S. aureus* strains used in this study originated from Austria (n=7), Denmark (n=5), France (n=12), Germany (n=11), Hungary (n=3), Italy (n=9), Netherlands (n=11), and Spain (n=2). The genome sequences were analyzed for *SCCmec* types and MLST using a web-based server provided by the Center for Genomic Epidemiology.

PHAge Search Tool Enhanced Release (PHASTER) algorithm was used to identify prophage sequences from 60 *S. aureus* genomes (Arndt et al., 2016). Further, the intact prophage genomes were re-annotated using prokka 1.14 (Seemann, 2014). A total of 65 intact prophage nucleotide sequences were subjected to Multiple Sequence Alignment using Fast Fourier Transform (MAFFT) version v7.475 (Katoh et al., 2018), and further generate cluster tree on SplitsTree4 software (Huson & Bryant, 2006). The identified intact prophage sequences were analyzed for ARGs and VFGs using CARD (Alcock et al., 2020) and VirulenceFinder-2.0 Server (*CGE Server*, n.d),

respectively. The intact prophage sequences were used for a pan-genome comparison using the TBLASTX and prophage phiH14-1 as a seed genome in Gview server (https://server.gview.ca/). Furthermore, the prophage sequences belonging to each cluster or clade were analyzed for core and accessory genomes using Spine and AGEnt version 0.3.1 webserver (Ozer et al., 2014).

3.5 Identification and characterization of potential vaccine and drug target candidates by reverse vaccinology

The genome sequences of 16 MRSA strains were retrieved from the NCBI database (***http://www.ncbi.nlm.nih.gov/genbank/***) among these, 6 genome sequences were from this present study. The identification of conserved proteins (core-proteome) between the 16 genomes was analyzed in the genome set using reciprocal best BLAST hits of all CDS in EDGAR 2.0 software framework (Blom et al., 2016).

3.5.1 Identification of conserved proteins of MRSA and subtractive genomics

Paralog sequences from the *S. aureus* core-proteome were removed using CD-HIT (Li & Godzik, 2006) with a sequence identity cut-off of 0.7 (70%), and further analyzed using GEPTOP 2.0 with an essentiality score cut-off of 0.24 (Wen et al., 2019). The essential non-paralogous protein sequences were subjected to BLASTp against the *Homo sapiens* genome (host) using default parameters, and Vaxign 2 tool (He et al., 2010). The resultant sequences showing non-host homologous sequences with no hit found were selected for subsequent analysis.

3.5.2 Prediction of subcellular location and 3D modelling of cytoplasmic protein structures

The non-host homologous protein sequences were used for the prediction of subcellular location using Vaxign 2 tool (He et al., 2010). The cytoplasmic (CYT) protein sequences were submitted to MHOLline 2.0 server (Rossi et al., 2020) to model a three-dimensional (3D) structure.

3.5.3 Ligand libraries, virtual screening, and molecular docking analysis

The ligands, drug-like molecules mentioned in the article (Vilela et al., 2019) were obtained from the ZINC 15 database, and an in-house library was constructed. The docking was performed using the MolDock Optimizer search algorithm with default parameters in Molegro Virtual Docker (MVD) 6.0 (Bitencourt-Ferreira & de Azevedo,

2019) and visualized using PyMOL 2.4.1 (Schrödinger & DeLano, 2020), and Discovery Studio Visualizer 2020 (BIOVIA, 2020), respectively.

3.5.4 Reverse vaccinology approach for prediction of putative vaccine candidates

The non-CYT proteins were submitted to the Vaxign 2.0 tool (He et al., 2010) to identify the potential vaccine candidate proteins. The antigenicity of vaccine candidate proteins was predicted using the VaxiJen v2.0 server with a threshold of 0.4 (Doytchinova & Flower, 2007).

4 Results

4.1 Polyphasic characterization

In the present study, phenotypic and genotypic characterization confirmed that the 35 *S. aureus* clinical isolates were *S. aureus* strains and showed a high prevalence of *mecA* genes (94.28%) positive strains. The distribution of *SCCmec* types showed that 36.36% of strains were CA-MRSA and 54.54% of strains were HA-MRSA. The quantitative test for biofilm production revealed that among the 33 *mecA* gene-positive strains, 87.87% of strains produced biofilm. Among 33 *mecA* gene-positive strains, 15 *S. aureus* isolates were selected based on non- (26.7%), moderate- (40%), and strong- (33.3%) biofilm-forming abilities for molecular typing and discrimination. The *coa* gene typing showed the discriminatory index (DI) of 0.8381. *coa*-RFLP analysis revealed 11 different patterns, revealing the DI of 0.9619. Typing of the *spa* gene revealed DI of 0.9429. Analysis of the *spa* gene revealed 12 known *spa* types. The presence of *fnbA* and *fnbB* genes were detected in 73.3% and 66.6% strains, respectively. Genes associated with biofilm-forming ability *viz*., *cna*, *clfA*, *clfB*, and *ebps* were found present in 53.33%, 80%, 73.3%, and 86.6% strains, respectively.

Although, the individual test has the advantage of being cost-effective, but often cannot differentiate among the strains. To avoid a misleading conclusion, the data from all applied methods were coupled to perform cluster and PCA analysis. The analysis results suggested that Hungarian strains (SA H27 & SA H32) belonging to the same *SCCmec*-IV have shared the same cluster, however, these strains were isolated from the different sites of infections (nostrils and trachea) and showed different antibiotic resistance patterns and biofilm-forming abilities. Germany and Hungarian strains (SA G5 & SA H29) from different geographical locations clustered together. However, the

strains collected from Germany *viz*., SA G6 & SA G8 belonged to different *SCCmec* types and had similar antibiotic resistance patterns and biofilm-forming profiles, but these strains were isolated from the different sites of infections (skin and other body sites) and were not clustered in the same group.

4.2 Characterization through genomic approach

The genomes of *S. aureus* isolates generated reads per sample cover more than 98% of the reference genome (ASM1342v1) with an average depth of 152.4X. Among the *S. aureus* isolates, SA G8 has the largest genome size (28633393 bp) with a high % GC content (32.81%). The numbers of protein-coding sequences (CDSs) in the *S. aureus* strains varied from 2630 (SA H27) to 2743 (SA H29).

4.2.1 *In-silico* characterization of genomic assemblies

The Sequence Types (STs) of Hungarian isolates belonged to CC5, and CC22 suggests its relationship with CA-MRSA, while Germany isolates SA G6, and SA G8 belonged to CC8 which is typical of HA-MRSA.

A putative plasmid (p1G6) of the SA G6 genome has 30.97% sequence coverage with plasmids pTW20_1 (FN433597.1). Two putative plasmids, p1H29 and p2H29 of the SA H29 genome constitute *rep20* and *rep7C* type genes. Identified plasmid, p2H29 has *blaZ* (beta-lactamase), *blaR*, and *mecI* genes that conferred resistance to penicillin. Then, the SA H32 genome also consists of a putative plasmid (p2H32) which has 71.32% sequence coverage with plasmids AR_0472 (NZ_CP029648.1). It carried an erythromycin resistance gene (*emrC*).

Six study genomes of *S. aureus* strains shared 63.33% (19/30) antibiotic resistance and associated genes. This *in-silico* identification and antibiotic susceptibility test results were correlated with beta-lactam and erythromycin-resistant analysis. The auto-inducing peptide-II gene was identified in SA G6 and SA G8 genomes while the auto-inducing peptide-I gene was identified in SA G5, SA H27, SA H29, and SA H32 genomes. However, lanthipeptide A (gallidermin) was present in the SA H29 genome. The clinically isolated SA G6 carried an array of antibiotic and heavy-metal resistance genes, this finding suggests that SA G6 could be a multi-drug resistant strain. The VFGs of *S. aureus* genomes against the VFDB revealed that 40.98% (50/122) of genes were shared in all the strains of *S. aureus*. The genome of SA G8 isolate has occupied 3.40% of VFGs against its CDS, whereas the genome of SA H32 isolate has 2.97% of

VFGs against its CDS. In this study, SA G8, and SA H28 genomes present 77.27% (17/22) of adherence-associated genes with additional genes of clumping factor A (*clfA*), clumping factor B (*clfB*), and fibronectin-binding proteins (*fnbB*). The highest number of toxin genes were identified in the genome of SA G5 *i.e.*, 52. 94% (27/51).

4.2.2 Comparative genome analysis

The genome comparative analysis based on core genome ANI mean matrices result indicated that all the *S. aureus* strains showed >96.69% sequence identities. Among the core genomes of *S. aureus* isolates, the SA H27 genome showed the highest sequence identities (97.3 to 99.9%), however, the SA G5 genome has 96.7 to 97.2% sequence identities to genomes of *S. aureus* isolates. The functional annotations using SEED subsystem and eggNOG revealed that *S. aureus* isolates required amino acids than carbohydrates as the energy source and suggests that these isolates adapted to grow in a protein-rich medium than carbohydrates. The comparison of six study *S. aureus* genomes generated a pan-genome size of 3691 genes, of which 2202 (59.7%) genes were core-genome, 721 (19.5%) genes were dispensable, and 768 (20.8%) genes were singletons. The highest (107 CDSs), and singletons (4) were observed in SA G5, and SA H27, respectively. The pan-genome analysis of *S. aureus* strains showed a growing and open pan-genome model. The phylogenomic revealed that the strains with ST225 (Germany), ST228 (Germany, Switzerland), ST105 (USA), and ST5 (Japan) were grouped in the same CC5 cluster, and a different cluster was noticed among the UK origin ST22 (CC22). This analysis revealed that the German and Hungarian isolates are genetically diverse and showed variation among them due to the gain or loss of MGEs. The event of MGEs transfer was observed in ST5, ST225, and ST228.

4.3 Comparative genomic analyses of *S. aureus* prophages

To study prophages diversity, genome sequences of *S. aureus* strains associated with human (*Homo sapiens*) infections (n=34), bovine (*Bos tourus*) infections (n=22), and dog (*Canis lupus familiaris*) infections (n=4) were retrieved from the NCBI database.

The MLST analysis result revealed that *S. aureus* strains associated with the animal infections have ST398 (CC398) and ST151 (CC151), while the *S. aureus* strains associated with human infections have ST398 (CC398), ST8 (CC8), ST5 (CC5), *etc*. The *S. aureus* with CC398 (ST398) and CC151 (ST151) are the most identified clone types of bovines in the European regions (Fluit, 2012). It was reported that these CCs

are an emerging zoonotic agent (Springer et al., 2009). Prophages of CC398 and CC151 strains associated with animals were selected to compare with prophages of CC398 or other CCs strains associated with human infections.

PHASTER identified 65 intact/complete prophages, among these prophages, 57 prophages were extracted from MRSA strains, while the 8 analyzed prophages were extracted from MSSA strains in which 4 prophages were from 4 MSSA strains associated with human infections, and another 4 prophages were extracted from 4 MSSA strains associated with bovine and dog infections. The genome sizes of intact/complete prophages were in the range of approximately 24.8 to 87.8 kb, and the GC content varied between 32.16 and 35.38%. These complete prophages belonged to the *Siphoviridae* family and had temperate lifestyles. The highest number (132) of coding sequence (CDS) was found in phiG4-3 (*S. aureus* SA G6) and the lowest number (29) of CDS was found in phiH7-2 (*S. aureus* strain H7).

The phylogenetic analysis of 65 intact prophage sequences generated the prophages into 7 different clades. The prophages carried by human and animal-associated *S. aureus* strains linked with different infected sites and different geographical locations were dispersed randomly in all clades. There was no correlation between the clade, the host, the site of infection, or geographical locations. The phylogenetic analyses result showed that the prophages carried by MSSA strains showed high proximity among them and were found clustered in subclades. This similar clustering pattern and prophages sequences similarities are favored by the identical *SCCmec* types, STs, or CCs of their host strains. This study suggested that the frequency of phage-mediated HGT is higher between the *S. aureus* strains with the identical *SCCmec* type, STs, or CCs.

The clade/subclade-wise prophages analysis revealed that the prophages carried by *S. aureus* associated with human or bovine infections have relatively high genome sizes in comparison with prophages of *S. aureus* associated with dogs. It was found that the prophage sequences from a different host spectrum showed 100% sequence similarity, however, their host strains owned the identical *SCCmec* Iva, clonal lineage ST398. Remarkably, this showed the evidence of phage-mediated horizontal gene transfer between the *S. aureus* associated with bovine infections and the *S. aureus* associated with dog infection.

The prophages carried by animal-associated *S. aureus* have 50.34% more VFGs than the prophages harbored by human-associated *S. aureus*. The immune evasion

cluster (IEC) was identified in both prophages harbored by human and animal-associated *S. aureus* strains. The prophages carried by human-associated *S. aureus* strains of different serotypes obtained from different geographical locations scattered themselves in all the clades, suggesting that these phages have a wide distribution across the European regions.

4.3 Identification and characterization of potential drug and vaccine target candidates by reverse vaccinology

4.3.1 Identification of conserved proteins of MRSA and subtractive genomics

A set of 1719 CDSs shared by all the strains was identified and these sequences were extracted in form of protein sequences. Out of these protein sequences, 1678 non-redundant protein sequences were retrieved by using CD-HIT. Essential core proteins of 278 were identified by GEPTOP 2.0 server. Further comparison of the core-proteins of essential protein sequences to the human host proteome resulted in a set of 98 targets as essential non-host homologous.

4.3.2 Prediction of subcellular location and 3D modelling of cytoplasmic protein structures

The subcellular localization of 98 essential non-homologous proteins predicted that 78 proteins were localized in CYT regions, 2 proteins were in SEC, 6 proteins were in potentially PSE, and 12 proteins in MEM. The analysis of proteins localized in CYT regions using MHOLline 2.0 predicted 19 proteins (9 very high quality, 8 high, and 2 good) as potential candidates for drug targets. Based on the druggability score, 8 drug target proteins *viz.,* biotin protein ligase (BPL), thymidine kinase (TMK), UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-L-lysine ligase (MurE), Phosphate acetyltransferase (Pta), HPr kinase/phosphorylase, UTP-glucose-1-phosphate uridylyltransferase (UGPase), Fatty acid/phospholipid synthesis (PlsX), and Pantothenate synthetase (PanC) were selected for the docking process.

4.3.3 Virtual screening and molecular docking analysis

The molecular docking analysis suggested that the ZINC4235426 compound is a promising drug molecule. This compound was found in good interactions along with lower MolDock scores with four drug target proteins such as BPL, TMK, MurE, HPr kinase/phosphorylase, and PanC. Also, this drug molecule satisfies Lipinski's rule of

five with zero violations and had no PAIN alerts. Also, using the reverse vaccinology approach, 4 putative antigenic proteins were identified, among these, PrsA and EssA proteins were found to be more promising vaccine candidates. These identified proteins can be used for further rational drug or vaccine design to identify novel therapeutic agents for the treatment of MDR staphylococcal infections.

5. Summary/Research finding

- PCR-RFLP method and suggested that *spa* typing method has better than other typing methods.
- The characterization of MRSA through a polyphasic approach could determine the relatedness among geographically diverse MRSA strains.
- The data generated from the WGS confirmed the diversity of MRSA strains among the same CCs.
- The genomic data mining approach has clearly stated that the biofilm-forming ability of MRSA was not correlated with the presence of biofilm-forming genes, also the genetic constituents have no information regarding the infection sites.
- The core-genome and accessory genome revealed that *S. aureus* isolates required amino acids than carbohydrates as the energy source and suggests that these isolates adapted to grow in a protein-rich medium than carbohydrates.
- The mining of prophage signals from *S. aureus* genomic data suggested that the *S. aureus* strains with the same *SCCmec* type, and clonal complex favored the harboring of similar prophage sequences and suggested that the frequency of phage mediated HGT is higher between them.
- The IEC is highly human-specific, however, this computational study finding revealed that the presence of IEC could not differentiate between the phages of human and animal-associates *S. aureus.*
- This study proposed that prophages of animal-associated *S. aureus* could be more pathogenic than prophages of human-associated *S. aureus*.
- Remarkably, this study showed evidence of phage-mediated HGT between the *S. aureus* associated with bovine (cow) infections and the *S. aureus* associated with dog infections.
- This study showed that the *S. aureus* phages are dispersed among the several *S. aureus* serotypes and around the European regions.

- The molecule docking analysis result suggested that the ZINC04235426 compound is the most confident drug molecule.
- The reverse vaccinology approach revealed foldase protein (PrsA), and ESAT-6 secretion machinery protein (EssA) are the most promising vaccine candidates.

6. References

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., et al., (2020). CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, *48*(D1), D517–D525. https://doi.org/10.1093/nar/gkz935

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, *44*(W1), W16–W21. https://doi.org/10.1093/nar/gkw387

Aziz, R. K., Bartels, D., Best, A., et al., (2008). The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*, *9*. https://doi.org/10.1186/1471-2164-9-75

Bitencourt-Ferreira, G., & de Azevedo, W. F. (2019). Molegro virtual docker for docking. In *Methods in Molecular Biology* (Vol. 2053, pp. 149–167). Humana Press Inc. https://doi.org/10.1007/978-1-4939-9752-7_10

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., & Weber, T. (2019). AntiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, *47*(W1). https://doi.org/10.1093/nar/gkz310

Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., & Goesmann, A. (2016). EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Research*, *44*(W1). https://doi.org/10.1093/nar/gkw255

CGE Server. (n.d.). Retrieved May 8, 2019, from https://cge.cbs.dtu.dk/services/

Chan, C. X., Beiko, R. G., & Ragan, M. A. (2011). Lateral Transfer of genes and gene fragments in Staphylococcus extends beyond mobile elements. *Journal of Bacteriology*, *193*(15). https://doi.org/10.1128/JB.01524-10

CLSI. (2014). *Performance Standards for Antimicrobial Susceptibility Testing; Twenty-Fourth Informational Supplement* (9th ed., Vol. 34).

Collee, J., Fraser, A., & Marmion, B. P. (1996). *Mackie and McCartney Practical Medical Microbiology* (14th ed.). Elsevier Health Sciences.

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, *8*. https://doi.org/10.1186/1471-2105-8-4

Du, J., Chen, C., Ding, B., et al., (2011). Molecular Characterization and Antimicrobial Susceptibility of Nasal Staphylococcus aureus Isolates from a Chinese Medical College Campus. *PLoS ONE*, *6*(11), e27328. https://doi.org/10.1371/journal.pone.0027328

Fluit, A. C. (2012). Livestock-associated Staphylococcus aureus. *Clinical Microbiology and Infection*, *18*(8), 735–744. https://doi.org/10.1111/j.1469-0691.2012.03846.x

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. *Journal of Clinical Microbiology*, *41*(12). https://doi.org/10.1128/JCM.41.12.5442-5448.2003

He, Y., Xiang, Z., & Mobley, H. L. T. (2010). Vaxign: The first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *Journal of Biomedicine and Biotechnology*, *2010*. https://doi.org/10.1155/2010/297505

Huerta-Cepas, J., Szklarczyk, D., Heller, D., et al., (2019). EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1). https://doi.org/10.1093/nar/gky1085

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. In *Molecular Biology and Evolution* (Vol. 23, Issue 2, pp. 254–267). Mol Biol Evol. https://doi.org/10.1093/molbev/msj030

Katoh, K., Rozewicki, J., & Yamada, K. D. (2018). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4). https://doi.org/10.1093/bib/bbx108

Khoshkharam-Roodmajani, H., Sarvari, J., Bazargani, A., Kandekar-Ghahraman, M. R., Nazari-Alam, A., & Motamedifar, M. (2014). Molecular typing of methicillin-resistant and methicillin-susceptible Staphylococcus aureus isolates from Shiraz teaching hospitals by PCR-RFLP of coagulase gene. *Iranian Journal of Microbiology*, *6*(4), 246–252. http://ijm.tums.ac.ir

Köck, R., Becker, K., Cookson, B., van Gemert-Pijnen, J. E., et al., (2010). Methicillin-resistant Staphylococcus aureus (MRSA): Burden of disease and control challenges in Europe. In *Eurosurveillance* (Vol. 15, Issue 41, pp. 1–9). European Centre for Disease Prevention and Control (ECDC). https://doi.org/10.2807/ese.15.41.19688-en

Köser, C. U., Ellington, M. J., Cartwright, E. J. P., Gillespie, S. H., Brown, N. M., Farrington, M., Holden, M. T. G., Dougan, G., Bentley, S. D., Parkhill, J., & Peacock, S. J. (2012). Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology. *PLoS Pathogens*, *8*(8), e1002824. https://doi.org/10.1371/journal.ppat.1002824

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Martineau, F., Picard, F. J., Roy, P. H., Ouellette, M., & Bergeron, M. G. (1998). Species-specific and ubiquitous-DNA-based assays for rapid identification of Staphylococcus aureus. *Journal of Clinical Microbiology*, *36*(3). https://doi.org/10.1128/jcm.36.3.618-623.1998

McNair, K., Bailey, B. A., & Edwards, R. A. (2012). PHACTS, a computational approach to classifying the lifestyle of phages. *Bioinformatics*, *28*(5). https://doi.org/10.1093/bioinformatics/bts014

Meier-Kolthoff, J. P., & Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10210-3

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A. A., et al., (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *Journal of Computational Biology*, *20*(10). https://doi.org/10.1089/cmb.2013.0084

Overbeek, R., Olson, R., Pusch, G. D., et al., (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, *42*(D1). https://doi.org/10.1093/nar/gkt1226

Ozer, E. A., Allen, J. P., & Hauser, A. R. (2014). Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools Spine and AGEnt. *BMC Genomics*, *15*(1), 1–17. https://doi.org/10.1186/1471-2164-15-737

Perumal, D., Lim, C. S., Sakharkar, K. R., & Sakharkar, M. K. (2007). Differential genome analyses of metabolic enzymes in pseudomonas aeruginosa for drug target identification. *In Silico Biology*, *7*(4–5), 453–465. https://europepmc.org/article/med/18391237

Rahimi, F., Katouli, M., & Karimi, S. (2016). Biofilm production among methicillin resistant Staphylococcus aureus strains isolated from catheterized patients with urinary tract infection. *Microbial Pathogenesis*, *98*, 69–76. https://doi.org/10.1016/j.micpath.2016.06.031

Rappuoli, R. (2000). Reverse vaccinology. In *Current Opinion in Microbiology* (Vol. 3, Issue 5, pp. 445–450). Current Biology Ltd. https://doi.org/10.1016/S1369-5274(00)00119-3

Rossi, A. D., Oliveira, P. H. E., Siqueira, D. G., Reis, V. C. C., Dardenne, L. E., & Goliatt, P. V. Z. C. (2020). MHOLline 2.0: Workflow for automatic large-scale modeling and analysis of proteins. *Revista Mundi Engenharia, Tecnologia e Gestão (ISSN: 2525-4782)*, *5*(6). https://doi.org/10.21575/25254782rmetg2020vol5n61325

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14). https://doi.org/10.1093/bioinformatics/btu153

spa Type Finder/Identifier. (n.d.). Retrieved February 8, 2019, from http://spatyper.fortinbras.us/

Springer, B., Orendi, U., Much, P., Höger, G., Ruppitsch, W., Krziwanek, K., Metz-Gercek, S., & Mittermayer, H. (2009). Methicillin-resistant Staphylococcus aureus: A

new zoonotic agent? In *Wiener Klinische Wochenschrift* (Vol. 121, Issues 3–4, pp. 86–90). Springer. https://doi.org/10.1007/s00508-008-1126-y

Strommenger, B., Kettlitz, C., Werner, G., & Witte, W. (2003). Multiplex PCR assay for simultaneous detection of nine clinically relevant antibiotic resistance genes in Staphylococcus aureus. *Journal of Clinical Microbiology*, *41*(9). https://doi.org/10.1128/JCM.41.9.4089-4094.2003

Tenover, F. C., Tickler, I. A., Le, V. M., Dewell, S., Mendes, R. E., & Goering, R. v. (2019). Updating molecular diagnostics for detecting methicillin- susceptible and methicillin-resistant staphylococcus aureus Isolates in Blood Culture Bottles. *Journal of Clinical Microbiology*, *57*(11). https://doi.org/10.1128/JCM.01195-19

Wen, Q. F., Liu, S., Dong, C., Guo, H. X., Gao, Y. Z., & Guo, F. B. (2019). Geptop 2.0: An updated, more precise, and faster Geptop server for identification of prokaryotic essential genes. *Frontiers in Microbiology*, *10*(JUN). https://doi.org/10.3389/fmicb.2019.01236

Zhang, K., McClure, J. A., Elsayed, S., Louie, T., & Conly, J. M. (2005). Novel multiplex PCR assay for characterization and concomitant subtyping of staphylococcal cassette chromosome mec types I to V in methicillin-resistant Staphylococcus aureus. *Journal of Clinical Microbiology*, *43*(10), 5026–5033. https://doi.org/10.1128/JCM.43.10.5026-5033.2005

7. List of publications

Publications related to thesis topic:

1. **Naorem, R. S**., Urban, P., Goswami, G., & Fekete, C. (2020). Characterization of methicillin-resistant *Staphylococcus aureus* through genomics approach. *3 Biotech*, *10*(9), 401. doi:10.1007/s13205-020-02387-y. **(Q2; IF: 2.406; 2020)**

2. **Naorem, R. S.**, Blom, J., & Fekete, C. (2021). Genome-wide comparison of four MRSA clinical isolates from Germany and Hungary. *PeerJ, 9*, e10185. doi:10.7717/peerj.10185. **(Q1; IF: 2.98; 2020)**

3. **Naorem, R. S.**, Goswami, G., Gyorgy, S., & Fekete, C. (2021). Comparative analysis of prophages carried by human and animal-associated *Staphylococcus aureus* strains spreading across the European regions. *Scientific reports 11*, 18994. https://doi.org/10.1038/s41598-021-98432-8. **(Q1; IF: 4.379; 2020)**

4. **Naorem, R. S.,** Pangabam, B., Bora, S., Goswami, G., Barooah, M., Hazarika, D., & Fekete, C. (2022). Identification of Putative Vaccine and Drug Targets against the Methicillin-Resistant *Staphylococcus aureus* by Reverse Vaccinology and Subtractive Genomics Approaches. *Molecules*, 27(7), 2083. https://doi.org/10.3390/molecules27072083. **(Q1; IF: 4.411; 2020)**

Publications outside to thesis topic:

1. Bouchelaghem, S., Das, S., Naorem, R. S., Czuni, L., Papp, G., & Kocsis, M. (2022). Evaluation of Total Phenolic and Flavonoid Contents, Antibacterial and Antibiofilm Activities of Hungarian Propolis Ethanolic Extract against *Staphylococcus aureus*. *Molecules, 27(2)*, 574. http://dx.doi.org/10.3390/molecules27020574. **(Q1; IF: 4.411; 2020)**

Conference proceedings related to thesis topic:

1. **Naorem, R. S.,** Urban, P., & Fekete, C. (2019). Comparative analysis of (pro)phages carried by human and livestock-associated *Staphylococcus aureus* strains spreading across the European regions. *In: XVI. János Szentágothai Multidisciplinary Conference and Student Competition*, p. 251, ISBN: 9789634293446.
2. **Naorem, R. S**., Urban, P., Czuni, L., Gazdag, Z., Papp, G., & Fekete, C. (2018). Genome-based comparative analysis of Methicillin-resistant *Staphylococcus aureus* strains. *In: Bódog, Ferenc (eds.). 7th Interdisciplinary Doctoral Conference*, 2018, p. 183, ISBN: 9789634292104

Conference proceedings outside to thesis topic:

1. Bouchelaghem, S., Das, S., Czuni, L., Gazdag, Z., Fekete, C., Kőszegi, T., **Naorem, R, S.,** Papp, G. (2019). Antimicrobial activities of Hungarian propolis alone and in combination with antibiotics and its antibiofilm activity on *Staphylococcus aureus*. *In: XVI. János Szentágothai Multidisciplinary Conference and Student Competition*, p. 16, ISBN: 9789634293446
2. Urbán, P., Kovács-Valasek, A., Schönhardt, K., Czuni, L., **Singh Romen N**, Papp, G., Gazdag, Z., Manczinger, L., Kredics, L., Fekete, C. (2017). NGS based peptaibol synthase prediction in the green mould disease causing *Trichoderma pleuroti*. *In: VI. Magyar Mikológiai Konferencia*, 56(1), 155-157.