UNDERSTANDING THE SPECIFIC MOLECULAR MECHANISMS THAT GOVERN RNA MEDIATED SIGNALLING IN *Staphylococcus aureus*

A THESIS TO BE SUBMITTED TO

THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY

FOR THE AWARD OF THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY

SAVITHA DATTATRI NADIG

UNDER THE GUIDANCE OF

PROF. H.S. SUBRAMANYA

Director, *IBAB and Biocon Chair*
Institute of Bioinformatics and Applied Biotechnology
BioTech Park, Electronic City Phase 1
BENGALURU 560010

PROF. B. GOPAL

Co-Guide and Mentor, (DST WOS-A scheme)
Indian Institute of Science
BENGALURU 560012

AUGUST 2021

THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY

Private University Established in Karnataka by ACT 35 of 2013
BENGALURU - 560064

DECLARATION BY THE CANDIDATE

I declare that this thesis entitled "**Understanding the specific molecular mechanisms that govern RNA mediated signalling in *Staphylococcus aureus***" submitted for the award of Doctor of Philosophy to THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY, Bengaluru, is my original work, conducted under the supervision of my guide **Prof. H.S. Subramanya** and co-guide, **Prof. B. Gopal**. I also wish to inform that no part of the research has been submitted for a degree or examination at any university. References, help and material obtained from other sources have been duly acknowledged

I hereby confirm the originality of the work and that there is no plagiarism in any part of the dissertation.

Signature of the Candidate

Place: Bengaluru

Date: 27.08.2021

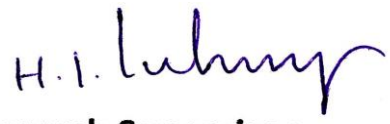
Name of candidate: Savitha Dattatri Nadig
Registration Number: 11117020539
(October, 2017)

THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY

Private University Established in Karnataka by ACT 35 of 2013
BENGALURU - 560064

CERTIFICATE

This is to certify that the work incorporated in this thesis **"Understanding the specific molecular mechanisms that govern RNA mediated signalling in *Staphylococcus aureus*"** submitted by **Savitha Dattatri Nadig** was carried out under my supervision. No part of this thesis has been submitted for a degree or examination at any university. References, help and material obtained from other sources have been duly acknowledged. I hereby confirm the originality of the work and that there is no plagiarism in any part of the dissertation.



Research Supervisor
Date: 27.08.2021

Prof. H. S. Subramanya

Director IBAB and Biocon Chair
Institute of Bioinformatics and Applied Biotechnology
Biotech Park, Electronic City Phase I
Bengaluru 560010,
India

Acknowledgements

I would like to express my sincere gratitude to my research supervisor, Prof. H. S. Subramanya, for guiding me through my research.

I am grateful to my co-supervisor, Prof. B. Gopal for mentoring me all through my journey up to this stage. My sincere gratitude to him for providing me an opportunity to continue my research in his lab and accomplish my dream. I will always be indebted to him for the time he provided in improving my scientific writing and presentation skills.

My sincere respect and gratitude to Prof. MRN Murthy, his guidance and support have been very useful. It has been a wonderful experience to read his short, informative, scientific articles in Kannada.

My heartfelt gratitude to Dr. Gayathri Arakere for introducing me to Staphylococcal research during my stint at Sir Dorabji Tata Centre for Research in Tropical Diseases (SDTC, IISc) and help me sharpen my skillset in this area of research. She has been a phenomenal force in my career and continues to be a great mentor.

I am thankful to Department of Science and Technology (DST, WOS-A scheme) for providing me the grant and fellowship to carry out my research work at Molecular Biophysics Unit, Indian Institute of Science.

I would like to thank my collaborators, Dr. Siddarth Junjuwala, BSSE, IISc for his support, scientific discussions, and suggestions. I am thankful to Preeti for being such a good friend and a collaborator.

I thank my other collaborators, Dr. Shruthi and Sneha from Institute of Bioinformatics and Applied Biotechnology (IBAB) for all the insightful work on Genome sequencing.

I thank Mrs. Sunita of Proteomics facility for collecting data on MALDI and LC-MS/MS spectra.

I am grateful to past members of lab 301, MBU- Anil, Arvind, Rishi, Kapil, Radhika, Prathima, Sharmista, Manoj, Amit and Prashanth for all their scientific suggestions. Special gratitude to Anuja for being a good friend and making me comfortable in the lab during the initial days. I will cherish all the moments we spent together in the lab.

I would like to acknowledge all the present members - Ashish, Twinkle, Ankur, Nikhil, Sushma and Nishank for making lab 301, a great place to work and for their helpful scientific suggestions. Special thanks to Vandana and Debashree, for always being there to help me out in lab work. I appreciate the help from Mirudula in image data processing and analysis.

I would like to specifically thank my friend Sushma, who has been a very supportive co-researcher since SDTC days. I also like to thank all my friends Seema, Varshini, Mukunda, Prasanna, Sandhya for all the encouragement and I extend my gratitude to my friends in Manasagangotri Microbes group.

I am forever indebted to my parents and parents-in-laws for being my pillar of strength and supporting me. I express my gratitude to my brother, Anil for his constant encouragement, my sister-in-law, Amrutha for being a great friend and their little ones, Advaith and Aishwarya for making weekends happier.

I would like to acknowledge the role of my little ones, Vedanth and Vaibhavi in completion of this thesis. Their understanding and good wishes have always motivated me to speed up my pace at work.

I owe my success to Ram, for all his patience, constant support and belief in my hard work.

Dedicated to my family

Table of Contents

List of Figures

List of Tables

List of Acronyms

ACME	Arginine Catabolite Mobile Genetic Element
Agr	Accessory gene regulator
AHL	N acyl-homoserine lactones
AIP	Autoinducing peptide
AIs	Autoinducers
ATc	Anhydrotetracycline
Atl	Autolysin
Bap	Biofilm associated protein
BAM	Binary Alignment Map
CA-MRSA	Community-associated MRSA
CC	Clonal Complex
CRISPR	Clustered regularly interspaced short palindromic repeats
Dcm	DNA adenine methyl transferase
Da	Dalton
EDTA	Ethylenediaminetetra acetic acid
ETA	Exfoliatin toxin A
ETB	Exfoliatin toxin B
FDR	False Discovery Rate
FISH	Fluorescent in situ Hybridisation
FLAPs	Fluorescent light-up aptamers
FnbA	Fibronectin binding protein A
FnbB	Fibronectin binding protein B
FPs	Fluorescent Proteins
GFP	Green fluorescent protein
GO	Gene Ontology
HA-MRSA	Hospital-associated MRSA
Hla	α-haemolysin
Hlb	β-haemolysin
Hld	δ-haemolysin
HSL	Homoserine lactones
Ica	Intracellular adherence factors
LB	Luria Bertani

LC-MS	Liquid chromatography -mass spectrometry
MGE	Mobile Genetic Element
MgrA	Master regulatory protein
min	Minute
ml	Millilitre
MLST	Multi-locus Sequence Typing
mM	Millimolar
MRSA	Methicillin resistant *Staphylococcus aureus*
ms	milli second
MSSA	Methicillin sensitive *Staphylococcus aureus*
MSCRAMM	Microbial surface components recognizing adhesive matrix molecules
MST	Minimum Spanning Tree
NGS	Next-Generation Sequencing
ORF	Open Reading Frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Poly-ethylene glycol
PFGE	Pulsed Field Gel Eletrophoresis
PIA	Polysaccharide intercellular adhesion
PSM	Phenol soluble modulins
PVL	Panton-Valentine Leukocidin
qRT-PCR	Quantitative Real Time Polymerase Chain Reaction
QS	Quorum Sensing
RBP	RNA Binding Protein
RNA	Ribonucleic acid
Rot	Repressor of Toxins
RPKM	Reads Per Kilobase Million
rpm	Revolutions per minute
SAM	Sequence Alignment Map
SaPI	Staphylococcal Pathogenesity Island
SasG	Surface protein G
SBA	Sheep blood agar
SCC	Staphylococcal cassette Chromosome

Scn	Staphylococcal inhibitor of complement
SCV	Small Colony Variant
Seb	Enterotoxin B
SD	Shine-Dalgarno
SEA	Staphylococcal enterotoxin A
SEB	Staphylococcal enterotoxin B
SELEX	Systemic evolution of ligands by exponential enrichment
Spa	Surface Protein A
sRNA	Small Regulatory RNA
ST	Sequence Type
TCS	Two component signalling system
TE	Tris- Ethylenediaminetetra acetic acid
TO	Thiazole Orange
TPM	Transcript Per Million
TSB	Tryptic soy broth
THB	Todd-Hewitt broth
Tsst	Toxic shock syndrome toxin
WGS	Whole Genome sequencing
µg	Microgram
µl	Microlitre
µM	Micromolar
B. subtilis	*Bacillus subtilis*
E. coli	*Escherichia coli*
P. aeruginosa	*Pseudomonas aeruginosa*
S. aureus	*Staphylococcus aureus*
S. epidermidis	*Staphylococcus epidermidis*
S. pneumonia	*Streptococcus pneumonia*
V. fischeri	*Vibrio fischeri*
V. cholera	*Vibrio cholera*

Synopsis

Staphylococcus aureus is a leading cause of both hospital and community associated infections. This Gram-positive pathogen rapidly acquires resistance to multiple frontline antibiotics, thus severely affecting efforts to combat this pathogen. Antimicrobial resistance in *S. aureus* is brought about by diverse mechanisms. These include acquisition of mobile elements like pathogenicity islands (SaPIs), chromosome cassettes (SCC), transposons, conjugative plasmids and prophages contributing to the genome variation in different staphylococcal species.

A distinctive feature of *S. aureus* is a phenotypic switch between a persistent and a virulent phenotype. *S. aureus* uses a biphasic approach to cause a disease. The dormant cells, generally referred as persister phenotypes, arise at low cell density upon expression of protein factors and other adhesins promoting colonization and biofilm formation. At high cell density these bacteria initiate the secretion of a variety of toxins giving rise to virulent phenotype. This switch is considered to be a key factor in *S. aureus* pathogenesis.

As in other bacteria, the quorum sensing mechanism plays a significant role in modulating bacterial infection. The dominant quorum sensing (QS) mechanism in *S. aureus* is the Accessory gene regulator (agr). This mechanism is based on a two-component system-AgrC is the histidine kinase that also recognizes the Autoinducing peptide (AIP). AgrA is the response regulator. An important feature of the quorum response is the expression of a non-coding RNA, RNAIII, and δ-hemolysin, a pore forming toxin. The goal of the studies reported in this thesis was to understand the role played by signaling RNA, in particular RNAIII, in the phenotypic switch. We proposed to evaluate this mechanism in the context of a specific, characterized *S. aureus* strain. Towards this goal, we chose a specific *S. aureus* strain, LVP-7, from our in-house collection to initiate the study. The thesis thus has studies and analysis to understand the LVP-7 genotype and associated features, the expression profile and aspects that could help in delineating the role of the non-coding RNAIII from other sRNAs. Finally, we report the development of a fluorescence imaging assay to monitor the cellular level and, potentially, conformational features of RNAIII.

A brief introduction to research goals and the underlying biology is provided in chapter 1. Where appropriate, the relevant aspects are summarized in different chapters of this thesis. Chapter 2 describes the genomic features of LVP-7. The draft genome assembly revealed important genetic components of LVP-7, a CA-MRSA belonging to ST88 clonal lineage. The genome carries important virulence encoding prophages like PhiNM3 and ΦSa2958PVL and mobile genetic elements like SCCmec type V and SaPI3. Some of virulent toxins carried by LVP-7 are PVL (a pore forming toxin) and the seb toxin. The agr locus of this ST88 clonal background belongs to type III agr group. Interestingly, this ST88 clonal lineage has remained as a singleton and might emerge as an epidemic clone. Hence the association of virulence factors, antibiotic resistance and agr type with specific clonal lineages help to assess the severity of an infection caused by *S. aureus*.

The expression profile of LVP-7 under logarithmic growth phase was evaluated by RNA sequencing. The transcriptome analysis identified several abundant transcripts in the LVP-7 strain, which are known to play an important role in pathogenesis and virulence of *S. aureus* and the data was validated using qRT-PCR. Using Gene Ontology (GO) enrichment analysis, several genes corresponding to pathogenesis like the adhesion protein spa was identified. Other transcripts that were over expressed include genes like *sodA* and *sodM* that govern response to oxidative stress. Comparative Gene Ontology enrichment analysis of LVP-7 with another epidemiologically important strain, USA300 showed the percentage enrichment of GO terms ‘pathogenesis’ and ‘response to stress’ in LVP-7. This analysis is described in chapter 3.

RNA aptamers that specifically recognize fluorescent ligands provide an elegant tool for conformational analysis. The utility of an RNA aptamer (RNA Mango) to investigate the conformational features and functional aspects of RNAIII is described in chapter 4.

Together, the studies described in this thesis provide insights into the significant role of RNAIII in the *S. aureus* phenotypic switch. More importantly, the genome and RNA-sequencing data and analysis provide a strong framework for research to understand the molecular features that govern the physiology of this

infectious pathogen. These studies thus serve to highlight the role of RNA mediated signal transduction that is crucial for this human pathogen as it switches from a benign persister to an aggressive form releasing pore forming toxins.

List of Publications

1. **Savitha Nadig**, Sneha Murthy, Vandanashree Muralidharan, Hosahalli S Subramanya, Balasubramanian Gopal, Shruthi Vembar. Draft Genome Sequence of the Community-Associated *Staphylococcus aureus* Sequence Type 88 Strain LVP-7, Isolated from an Ocular Infection. Microbiol Resour Announc. 2021 Feb 18;10(7):e00077-21. doi: 10.1128/MRA.00077-21. PMID: 33602729; PMCID: PMC7892662.

2. **Savitha Nadig**, Vandanashree Muralidharan, Shruthi Vembar, Hosahalli S Subramanya, Balasubramanian Gopal. A transcriptional rationale for the pathogenecity of Staphylococcus aureus LVP-7. (Manuscript under preparation)

Chapter 1

Introduction

1.1 Bacteria and bacterial phenotype

The ability of bacteria to thrive in everchanging environmental conditions depends on its genetic content. The survival strategies involving different molecular mechanisms are often used only by a subset of the bacterial population and this heterogeneity potentially helps few bacteria in a cohort survive under different environmental stresses. This adaptation is also known to increase the fitness cost of the species which uses a positive feedback mechanism in which a secreted molecule for example, a quorum sensing signalling molecule, activates its production across the entire population. This is a common feature of adaptive phenotypes (Smits WK et al., 2006). This adaptation to environmental fluctuations is often referred to as an adaptive response and several complex regulatory networks are triggered that help the bacterium to survive in less conducive environments. Various mechanisms help in this microbial adaptation, one of them being the phenotypic switch, involving physiological, behavioural and genetic changes inside the bacterium (Sousa AM et al., 2011). Horizontal transfer of mobile genetic elements (MGEs) and point mutations on the chromosome also lead to short term genetic changes which help the bacteria to acquire resistance to various antimicrobials.

An interesting feature of phenotypic switching is that the microorganism can switch between two states depending on the environmental stimuli and this mechanism is referred to as ON/OFF mechanism. This phase variation occurs in a small population at a much higher rates. This is a common mechanism described in various Gram-positive bacteria such as *Staphylococcus aureus* and *Pseudomonas aeruginosa* and Gram-negative bacteria like *Escherichia coli*, *Salmonella* spp. and *Neisseria* spp. and fungi like *Candida albicans*, *Cryptococcus neoformans* and other *Candida species* (Fries BC et al., 2002). Hence, phenotypic switching plays a pivotal role in microbial survival, pathogenicity and persistence in both pathogenic and non-pathogenic microbes (Sousa AM et al., 2011). This switching causes microbial heterogeneity and dynamic populations producing phenotypic variants that are considered to be fitter since they are better adapted to the new environmental cues.

The most visible feature of phenotypic variants is seen in the colony morphology and are termed as small colony variants (SCVs). Bacteria tend to live in communities

embedded in an exopolysaccharide matrix referred as biofilms. Complex processes (physiological and metabolic processes) help the bacteria to switch between planktonic to biofilm mode of growth and specific changes are observed in protein regulation. Several mechanisms like enhanced virulence and resistance to antimicrobials are activated in a specific biofilm forming phenotype (Drenkard E 2003, Sauer K et al., 2002 and Mah TFC et al., 2001).

Another striking feature of bacteria is their ability to be persistent even under several stressful conditions. When a bacteria culture is exposed to lethal dose of antibiotic, a small fraction of the population can still survive when most of the cells are killed. These are referred to as persister cells. Persister cells are not resistant and do not carry antibiotic resistant genes in the genome. They are generally dormant cells, *i.e.*, they are non-dividing and remain metabolically inactive. Persister cells have the ability to get back to the normal growth cycle and divide into a new population once the lethal conditions are removed. Hence, persisters are responsible for the increased antimicrobial resistance and play an important role in a biofilm community as they can prevent the clearance of bacteria (Lewis K, 2005). Bacteria evade host immune defences by escaping the innate and acquired immune mechanisms thereby leading to colonization and infection. Phenotypic switching plays a role in this mechanism and affects the host- pathogen interaction. The sub-population thus generated are known as Small Colony Variants (SCVs). SCVs possess different antigens which the host immune response may fail to recognise (Van der Woude MW, 2011). The enhanced virulence of SCVs plays an important role in opposing the host defence mechanisms.

Bacterial communication within a community occurs through the release of a chemical signalling molecule upon sensing population density and this process is known as quorum sensing (QS). Quorum sensing is a population density sensing cell-cell communication system which triggers changes in bacterial behaviour and this involves the release of small hormone like molecules termed as pheromones or autoinducers (AIs) (Waters CM et al., 2005). During cell growth AIs are produced and released into the environment. Upon reaching a threshold concentration, the QS system is activated (Figure 1.1). Beyond the threshold concentration, the signal molecule interacts with a receptor protein providing a coordinated change in gene expression in the population (Kareb O et al., 2020).

Both Gram-negative and Gram-positive bacteria produce different types of AIs like N-acyl-homoserine lactones (AHL) and cyclic oligopeptides with a thiolactone structure. It is believed that for interspecies communication, LuxS/AI-2 are produced both by Gram-positive and Gram-negative bacteria (Abisado RG et al., 2018, Bassler BL, 1999).

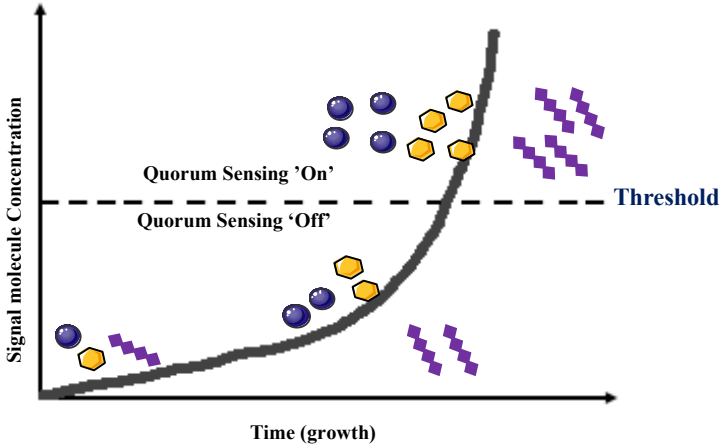


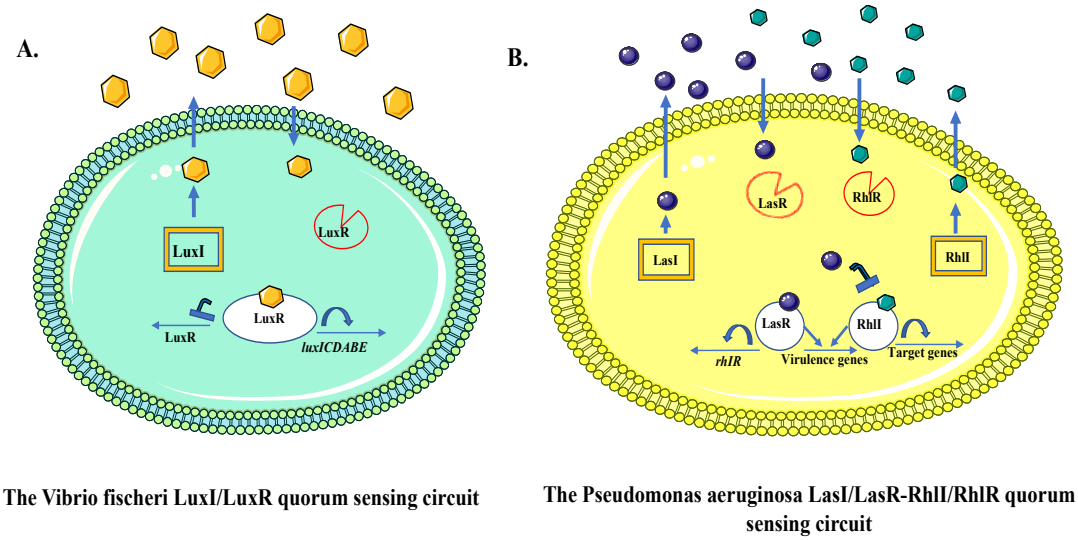
Figure 1.1: Schematic representation of quorum sensing effect triggered by autoinducers

Beyond a threshold concentration of autoinducers (represented in yellow, blue and purple colours), the quorum sensing system switches 'on' or 'off'. (Figure adapted and redrawn from Kareb O et al., 2020).

1.2 Quorum Sensing (QS) in Gram-negative bacteria

In the early 1960s, AHL based quorum sensing was first described in bioluminescent marine proteobacteria, *Vibrio fischeri*. This bacterium lives in symbiotic association with many eukaryotic hosts like pinecone fish, the Hawaiian bobtail squid, *Euprymna scolopes* and *Monocentris japonica*. In this symbiotic association, the eukaryotic host supplies a rich nutrient for the survival of *V. fischeri* and the pure culture of this microbe is confined to a specialized light organ of the eukaryotic host providing light to it. This bioluminescence helps during predator evasion by masking the shadows of pinecone fish and the squid mask their shadows. The *Vibrio fischeri* LuxI/LuxR quorum sensing circuit consists of five luciferase

structural genes (*luxCDABE)* and two regulatory genes (*luxR* and *luxI*) necessary for quorum sensing controlled light emission. Details of the mechanism are depicted in Figure 1.2.



The Vibrio fischeri LuxI/LuxR quorum sensing circuit

The Pseudomonas aeruginosa LasI/LasR-RhlI/RhlR quorum sensing circuit

Figure 1.2: Schematic representation of quorum sensing circuit Gram-negative bacteria

A. Quorum sensing circuit in *Vibrio fischeri* consists of five luciferase structural genes *luxCDABE* and two regulatory genes *luxR* and *luxI*. The LuxI protein represented in square box is responsible for synthesis of the HSL autoinducer- N-(3-oxohexanonyl)-homoserine lactone represented in orange coloured hexagons. The concentration of the autoinducer increases both intra and extracellularly at a high cell density and at a critical autoinducer concentration, it binds to the LuxR protein (circle) activating the downstream genes of *luxCDABE* operon. **B**. The *Pseudomonas aeruginosa* LasI/LasR-RhlI/RhlR quorum sensing circuit. The homoserine lactone signaling molecule N-(3-oxododecanonyl)-homoserine represented in purple balls, and the RhlI protein synthesizes N-(butryl)-homoserine lactone represented in orange pentagons. Figure Adapted and modified from (Miller MB et al., 2001)

The luciferase enzyme required for light production in *V. fischeri* consists of five luciferase genes which are encoded by the *luxICDABE* operon. One of the autoinducer regulatory protein is LuxI and acts in the production of an HSL, *N*-(3-oxohexanoyl)-homoserine lactone (Engebrecht J et al., 1983, Melissa MB et al., 2001). LuxR, another regulatory protein helps in binding to the autoinducer to activate the

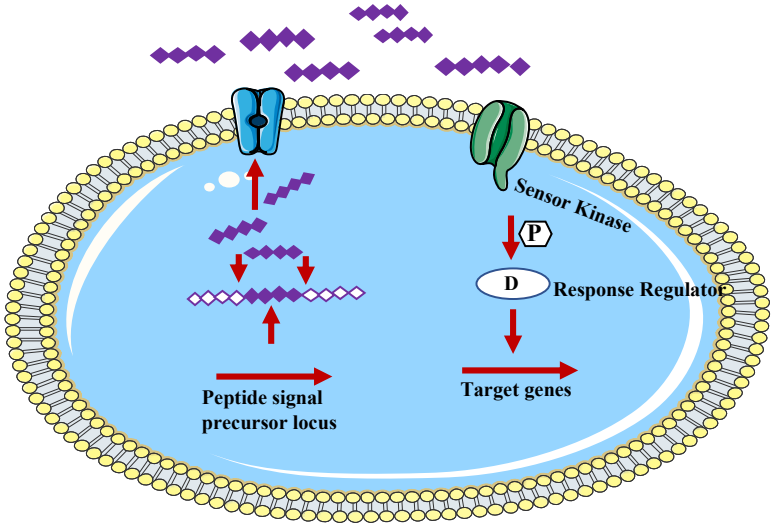
transcription of the *luxICDABE* operon. The transcription of *luxICDABE* operon is low at low cell densities, and correspondingly a low level of autoinducer is produced (via *luxI*). As the genes encoding luciferase are located downstream of luxI, a very low level light is produced (Engebrecht J et al., 1983, Miller MB et al., 2001). At a high cell density of *V. fischeri*, the autoinducer accumulates and binds to the cytoplasmic LuxR protein. Interaction of LuxR with the auto- inducer unmasks the LuxR DNA binding domain, allowing LuxR to bind the *luxICDABE* promoter and activate its transcription. This action results in an exponential increase in both autoinducer production and light emission. The LuxR- HSL complex also acts to negatively regulate expression of *luxR*. This negative feedback loop is a compensatory mechanism that decreases *luxICDABE* expression in response to the positive feedback circuit (Miller MB, et al., 2001).

Another well studied QS architecture involving more than one signal-receptor is in *Pseudomonas aeruginosa*. This opportunistic pathogen has two complete LuxR-I-type circuits (LasR-I and RhlR-I). Both the autoinducer synthases, LasI and RhlI catalyse the formation of HSL autoinducers namely *N*- (3-oxododecanoyl)-homoserine lactone and *N*-(butryl)-homoserine lactone. These two sequential regulatory circuits control the expression of many of the virulence factors in *P. aeruginosa*. The establishment of infection in the host occurs mainly due the activation of the promoter elements by HSL inducer upon binding to LasR, leading to the secretion of various virulence factors. LasR also induces the activation of the second QS circuit RhlR. Binding of RhlR to the autoinducer produced by RhlI induces the expression of the downstream target genes which are also under the control of LasI/LasR system (Passador L et al., 1993).

1.3 Quorum Sensing (QS) in Gram-positive bacteria:

A large number of cellular processes are regulated in response to cell population density via QS in Gram-positive bacteria. In contrast to Gram-negative bacteria, Gram-positive bacteria use secreted peptides as autoinducers. The concentration of the autoinducer peptide (AIPs) increases as the cell density increases. A two component sensor kinase senses these AIPs at their threshold concentration and this initiates a series of phosphorylation events and finally result in the phosphorylation of the

response regulator. The response regulator is activated upon phosphorylation and binds the target promoter DNA. This mechanism thus leads to the transcription of the QS controlled target genes. This process governs the virulence and pathogenesis in several Gram-positive pathogens like *Staphylococcus aureus* (Bassler BL. 1999 and Miller MB et al., 2001). The general mechanism of QS system in Gram-positive bacteria is represented in Figure 1.3.



A general model for peptide-mediated quorum sensing in Gram-positive bacteria

Figure 1.3: Schematic representation of quorum sensing circuit Gram-positive bacteria

The Gram-positive quorum sensing consists of a peptide signal precursor that gets translated into a precursor protein (purple and white diamonds) which gets cleaved (red arrows) to produce an autoinducer peptide signal (purple diamonds). The peptide is transported out of the cell via an ABC transporter (blue protein complex). At a threshold concentration of the peptide signal, it is sensed by the histidine sensor kinase leading to the phosphorylation of a conserved histidine residue. Subsequently the phosphoryl group is transferred to the response regulator leading to the phosphorylation of the aspartate residue (D). The phosphorylated response regulator activates the transcription of the target genes. Adapted and modified from (Melissa B et al., 2001).

1.4 Staphylococcus aureus

In 1882, a Scottish surgeon Sir Alexander Ogston isolated *Staphylococcus aureus* from surgical wounds (Ogston A, 1881) and he classified it as *Staphylococcus* (from the Greek word staphylos-'grape' and kokkos-'berry' or 'seed') (Ogston A, 1882). *Staphylococcus aureus* is a Gram-positive, sphere shaped, non-motile bacteria with a golden yellow pigmentation. This pigmentation of the colonies is due to a membrane bound carotenoid called staphyloxanthin which protects *S. aureus* from reactive oxygen species and phagocytic killing (Clauditz A, 2006). It is a common commensal of humans. It causes a wide range of infections starting from minor skin infections to more severe infections like meningitis, endocarditis and necrotizing pneumonia. It is one of the leading pathogens causing ophthalmic infections.

1.4.1 Staphylococcal Pathogenicity and Virulence:

Bacteria have to adjust their gene expression and physiology accordingly to adapt to different environmental stresses like temperature, pH, oxidative and antimicrobial stress. This is absolutely necessary for a pathogen, since it is in continuous contact with the host during infection. *S. aureus* has developed a complex regulatory system to regulate the production and release of virulence factors. As the regulatory machinery and virulence factors are not required for normal growth, they are known as accessory elements. These accessory factors contribute to the pathogenesis of *S. aureus* and are required to establish infection in the host. Accessory elements are mostly carried by mobile genetic elements (MGEs) like phages, plasmids, staphylococcal chromosome elements (SCC*mec*) and pathogenicity islands. Acquisition of mobile genetic elements by different staphylococcal strains is considered as the main contributor of strain-to-strain variation. Bacteriophage transduction is known to be the frequent mode of gene transfer between different staphylococcal strains and they majorly carry Panton-Valentine leukocidin (PVL) and immune evasion factors as virulence factors. The pore-forming toxins (PFT) are polypeptides that include the bi-component leukocidins (PVL, LukED, LukGH/AB and γ-hemolysin), β-hemolysin and phenol soluble modulins (PSMs). All these toxins have the ability to lyse leukocytes and the red blood cells. Other virulence factors carried by bacteriophages include cell wall-anchored virulence determinant SasX, exfoliate toxin A, staphylococcal inhibitor of complement (scn), staphylokinase (sak), enterotoxins encoded by sea, selk2, and sep. The Pathogenicity islands, another mobile genetic

element, encodes many super antigens like toxic shock toxin (tsst), enterotoxin B (seb), enterotoxin-like protein Q (selQ). Several toxins are also plasmid encoded like enterotoxin D, enterotoxin-like protein SER, and SEJ including exfoliate toxin B. Many of these virulence factors are responsible for more acute infections like endocarditis and osteomyelitis.

The most critical pathogenicity factor in *S. aureus* infections is the formation of biofilm which helps the bacteria colonize abiotic or biotic surfaces. Three-dimensional structure of biofilm helps to protect the bacteria against antimicrobial chemotherapy. Biofilm formation in bacteria seems to be the best way of protecting from environmental stress and hence bacteria needs to quickly program itself to activate the formation of biofilm. Biofilm formation in bacteria mainly involves three major steps - adhesion(attachment), proliferation (maturation) and detachment (dispersion) (Liu Y et al., 2020 Figure 1.4). During initial attachment, the surface proteins of the bacteria bind to host matrix proteins like fibrinogen and fibronectin. Further the biofilm-forming cells proliferate and get embedded within a protective extracellular matrix mainly composed of extracellular polysaccharide intercellular adhesion (PIA) protein, teichoic acid and DNA. This extracellular matrix protects the bacteria from different antibiotics or the host defence mechanism. Finally the detachment leads to the dissemination of the staphylococcal infection. Several factors involved in the biofilm process are further explained in the subsequent sections.

The initial adherence has been attributed to bacterial surface proteins like autolysin, Atl and another protein Bap is involved in the attachment to polystyrene surfaces leading to the formation of biofilms. After successful adhesion, many host factors serve as specific receptors for colonization. Biofilm accumulation in staphylococci can be mediated by either polysaccharide or protein factors. Polysaccharide intracellular adhesin (PIA) is produced by *icaADBC* operon encoded genes (Heilmann C, et al., 1996). Extracellular DNA has also been identified to contribute to biofilm development in *S. aureus*. The negatively charged DNA molecule interacts with the positive charged PIA upon its release and acts as a glue in biofilm formation. In protein mediated biofilm formation, surface protein G (SasG) is known to play an important role. Detachment of biofilm may lead to the dissemination of an infection and thus colonization of new sites. The *S. aureus* nuclease gene (nuc) is known to contribute to biofilm detachment (Heilmann C, et al., 2009). Another strategy involves the production and release of small peptides called phenol-soluble modulins

(PSMs). PSMs have been classified into two types depending on their size- α-type peptides (approximately 20 amino acid) and β-type peptides (40-45 amino acid) in length. PSMs are known to have a surfactant-like effect which might help in the biofilm detachment. (Yao Y, et al., 2005).

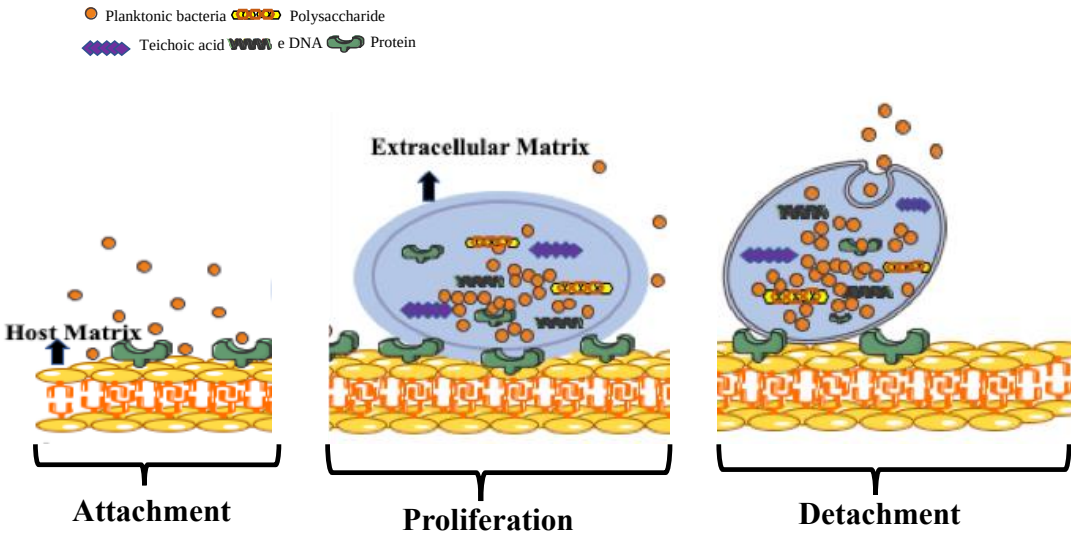


Figure 1.4: Model representation of Staphylococcal biofilm formation, stages and factors involved

Initial attachment involves attachment to host matrix. In this proliferation step, biofilm-forming cells get embedded within a protective extracellular matrix. Finally the detachment leads to the dissemination of the staphylococcal infection. (Figure adapted and remade from Liu Y et al., 2020)

Details of the different regulatory systems and their role is shown in Table 1.1.

Table 1.1: Regulatory systems and its role in *S. aureus* virulence

Regulatory system	Role
agr	Quorum sensing signaling, release of AIPs as a signal to carry out cell-to-cell communication. Activation of agrA leads to the expression of many exotoxins and exoenzymes.
SaeRS	Exo-proteins are induced including many virulence factors
SrrAB	Staphylococcal respiratory response-TCS. Induces the expression of *ica* operon (biofilm formation), repression of agr, TSST-1 and spa.
ArlRS	Autolysis regulated locus-TCS, induces the expression of MgrA. Repression of *agrA* and autolysins
SarA	Cytoplasmic regulator, induces the expression of exo-proteins and represses the expression of *spa*

Rot	Cytoplasmic regulator of toxins and extracellular proteases
MgrA	Cytoplasmic regulator, induces the expression of efflux pumps and capsule expression. Represses surface proteins
SigB	Stationary phase sigma factor, inhibits *agr* activity

TCS: Two component system, TSST-toxic shock syndrome toxin

1.4.2 Immune evasion by *Staphylococcus aureus*

The ability of *Staphylococcus aureus* to invade different host cell immune systems like non-professional phagocytes and professional phagocytes thereby surviving in the host environment leads to recurrent infections (Moldovan A et al., 2019). This pathogen uses alternative strategies to survive phagocytosis in different cell types of the host. For example, in non-professional phagocytes, *S. aureus* either escapes the endosome followed by cytoplasmic replicates or replicates within autophagosomes. Professional phagocytes possess a limited capacity to kill *S. aureus.* These bacteria can thus replicate within the host cells, lyse cells and persist in a continuous phagocytosis, host cell death and bacterial release cycle. Bacteria need to dynamically adapt to the intracellular environment and to the environmental changes during the transition from acute to chronic infection (Garzoni C et al., 2011). Bacteria change to SCV-like (Small Colony Variant) phenotype during this process. SCV's typically show a slow growth, altered morphological, metabolic changes and antibiotic profiles. Tuchscherr and co-workers first reported that SCVs come up in chronic infection models when *S. aureus* is internalized in non-professional phagocytes and survive intracellularly (Tuchscherr et al., 2011). As growth of SCVs intracellularly increases due to its longer stay, there is a reduction of host cell inflammatory response. Hence this study suggests that this mode of phenotype switching is an essential feature of the *S. aureus* infection and can explain an underlying cause of chronic and relapsing infections (Tuchscherr et al., 2011). The mechanism of *S. aureus* adaptation and bypassing of host defence system is represented in Figure 1.5.

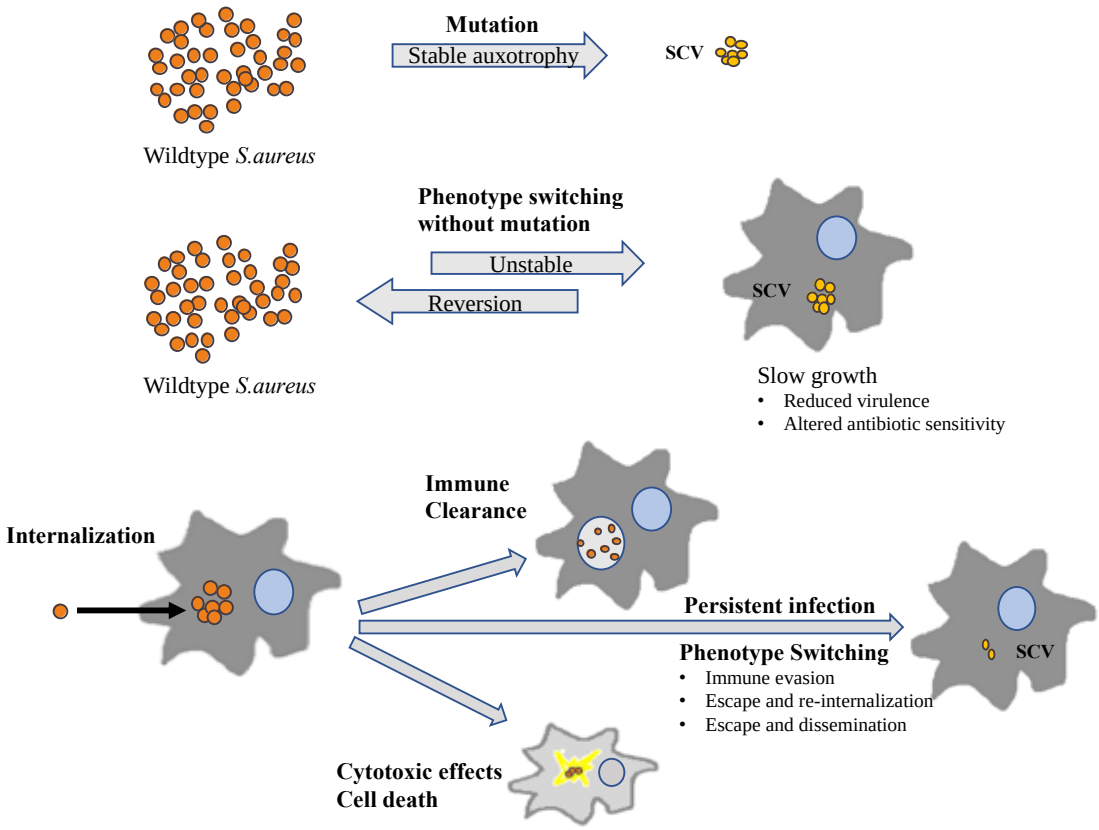


Figure 1.5: Immune evasion by *S. aureus*

Formation of SCVs helps to evade the host immune system resulting in intracellular phenotype switching. **A**. Represents the formation of stable SCVs by various routes like generation of auxotrophic mutants by variety of mutations. **B**. Without mutations, unstable SCVs with low virulence and less antibiotic sensitivity are generated during phenotype switching. These unstable SCVs can rapidly revert to the wild type form with increased virulence and rapid growth **C**. *S.aureus* internalization into host cell lead to many possibilities. *S.aureus* can be cleared by host immune responses by phagocytosis (Immune clearance) or by phenotype switching bacteria can cause persistent infection. This can help bacteria to disseminate and cause the disease. Figure adapted and remade from Garzoni C et al., 2011.

1.4.3 Signal transduction in *Staphylococcus aureus*

The ability of pathogenic bacteria to successfully infect the host depends on their ability to sense, respond and adapt to various environmental stimuli. The bacterial cell wall acts as a physical barrier imparting protection against environmental cues and

constitutes an important sensory interface for numerous sensing systems. Signal transduction includes Two-Component Systems (TCSs) and alternative σ factors.

Quorum sensing (QS) is a signalling process that allows cell-to-cell communication in bacteria (discussed in detail in section1.2 and 1.3). QS in *S. aureus* is regulated by accessory gene regulator (Agr*)* comprising of *agr* locus. This locus consists of two divergent transcripts, RNAII and RNAIII. The transcription of these mRNAs is initiated from two promoters known as P2 and P3. The RNAII transcript consists of *agrBDCA* proteins where in AgrB and AgrD acts as a cell density sensing cassette and is responsible for the synthesis, maturation and secretion of the autoinducing peptides (AIPs) while AgrA and AgrC are the response regulator and kinase in the two component system. AgrD undergoes proteolysis by AgrB, a membrane protease embedded in the cytoplasmic membrane. AgrD is a 46-amino acid precursor peptide which is processed by the membrane protease AgrB to form the signalling peptide. This step results in the release of autoinducing peptide (AIP- 7 to 9 amino acids). A threshold concentration of AIPs in the extracellular environment is sensed by the membrane kinase AgrC, a 46kDa membrane kinase belonging to the class 10 receptor-histidine protein kinase (HPK) family. Binding of the AIP to AgrC causes the autophosphorylation of the histidine kinase domain of AgrC. The phosphoryl group is then transferred to the response regulator AgrA resulting in the dimerization of the response regulator. AgrA is a 27kDa response regulator and dimerization leads to enhanced expression of the regulatory RNA, RNAIII. RNAIII an effector molecule of the Agr quorum sensing pathway. RNAIII is a bifunctional RNA having both coding and non-coding regions. The open reading frame (ORF) in the 5' end coding region of RNAIII encodes δ-hemolysin- a 26-amino acid cytolytic peptide conferring haemolytic activity to the bacterium and acts as an important virulent factor in the pathogenesis of *S. aureus* (Novick RP et al., 1993, Figure 1.6A). While AgrA is mostly conserved, sequence variations are seen in AgrB, AgrC and AgrD and in *S. aureus.* The allelic variations in the *agr* regions B, D and C lead to at least four *agr* specificity groups. It was previously shown that activation of specific a cognate receptor by an AIP is highly sequence specific (Novick RP. 2003, Figure 1.6B). AIP, the main *agr* autoinducing peptide has a five membered thiolactone ring containing 7 to 9 amino acids. The AIP sequences of *agr* I,II,III and IV groups in *S.aureus* are YSTCDFTM, GVNACSSLF, YINCDFLL, and YSTCYFTM, respectively (Yarwood JM et al., 2003) and these are

represented in Figure 1.6 B. AgrB dependent proteolysis of the AgrD precursor has been identified as an important step for AIP production (Wang B et al., 2015). RNAIII, a 514nt bifunctional RNA is considered as the main intracellular effector molecule of the *agr* system and plays an important role in the virulence of *S. aureus* by controlling the switch between the expression of surface proteins and synthesis of exotoxins.

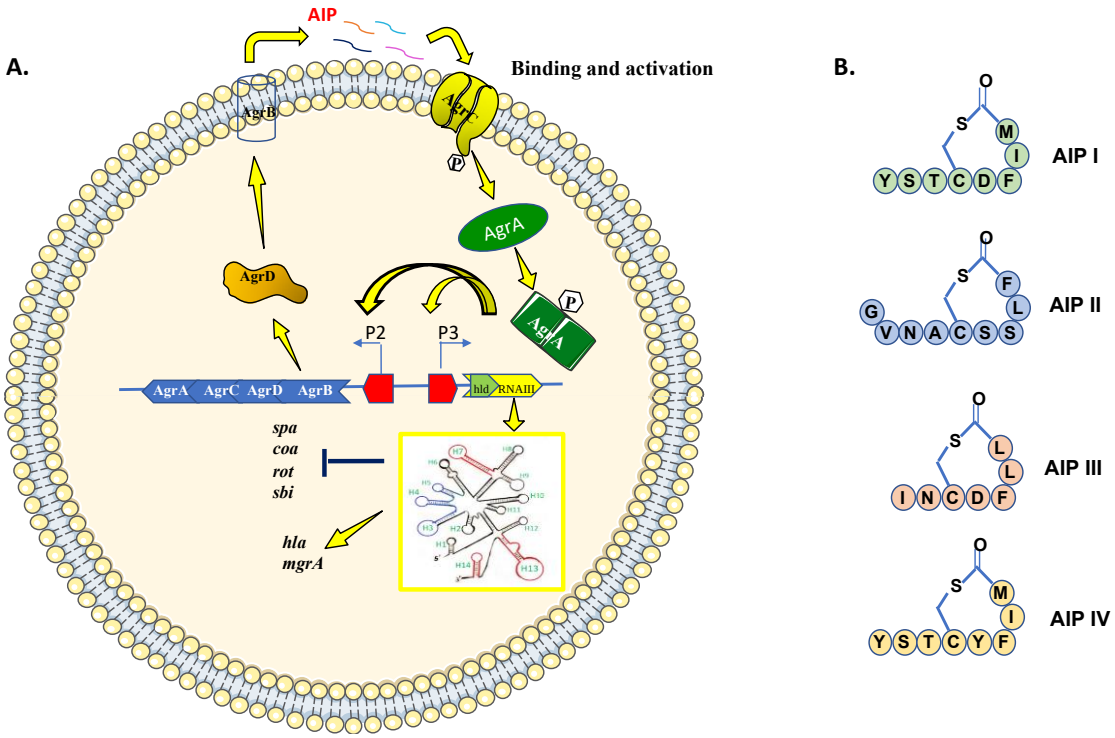


Figure 1.6: Schematic representation of Agr quorum sensing system in *S. aureus*

A. Agr operon (*agrBDCA*) is driven by the P2 promoter while the P3 promoter transcribes bifunctional RNAIII. The structure of RNAIII is shown in the inset box. Genes upregulated by RNAIII are shown by an arrow and genes down regulated by RNAIII are shown by a bar. **B**. Four different Autoinducing peptides (AIPs) containing a thiolactone ring structure. Secondary structure of RNAIII is made from the template taken from Benito Y et al., 2000.

1.4.4 RNA mediated process in *Staphylococcus aureus*

The severity of infections caused by *S. aureus* (both hospital and community associated infections) varies substantially. It depends on several parameters corresponding to the host and the bacterium. The expression of virulence factors is governed by multi-layered regulatory steps which include transcription, translation and

mRNA degradation. Numerous genome wide approaches have revealed that a variety of trans-acting factors (σ factors, two component systems (TCSs), RNA binding proteins and regulatory RNAs govern these regulatory mechanisms. The rapid adaptive nature of *S. aureus* to different environmental cues comes from these complex and dynamic networks which helps the bacterium to adapt its metabolism and fine tune the synthesis of virulence factors. In case of *S. aureus*, high genetic variability is observed at the genome level and has led to the evolution of strains carrying different genotypes. Some of these regulators vary among the different strains due to genomic rearrangements and the acquisition of mobile genetic elements. The genome content of the bacterium and the condition of the host plays an important role in causing disease. In prokaryotes, about 10-20% of the genes code for RNA's which might have regulatory functions. Because of their small size (from 80 to 500 nucleotides), these RNA regulators are called as small regulatory RNAs (sRNAs). sRNAs are a heterologous group and can modulate the gene expression with respect to various physiological changes. Substantial efforts have been made to decipher the function and mechanism by which sRNA's identify their primary targets and participate in more complex regulatory mechanisms as they function in various physiological responses.

1.4.5 Small regulatory RNAs (sRNAs) in *S. aureus* virulence and pathogenesis

The *S. aureus* genome (2.8M bp) with low GC content (32.8%), encodes a huge diversity of small regulatory RNAs (sRNAs), affecting its RNome. About 256 sRNAs have been reported from various studies which include computer software predictions coupled to expression analysis, cloning and sequencing analysis. (Felden B et al., 2011). These regulatory sRNAs can be divided into four main classes:

1. The trans-acting sRNAs interact with target mRNAs by base pairing and act as antisense RNA. These sRNAs act by two different mechanisms. In the first, these trans-acting RNAs and the target mRNA are transcribed in opposite directions even though they are encoded on the same DNA locus. In the second case, the sRNA and the target mRNAs are encoded on different DNA loci and share partial base pairings. Some of these sRNAs are bona fide non-coding RNAs (ncRNAs) and some may have an open reading frame (ORF).
2. The second class includes RNAs that are part of ribonucleoprotein complexes. sRNAs sequester regulatory proteins there by affecting their activity.

3. The third class includes CRISPR RNAs, a unique family of small processed RNAs.
4. The fourth class comprises of cis-acting regulatory RNAs which respond to trans-acting agents like metabolite concentration (riboswitches) and environmental cues (thermosensors)

Regulatory RNA's are known to play an important role as they carry out several regulatory tasks. Regulatory RNA's can directly sense environmental conditions like change in temperature, pH, nutrient availability and antimicrobial stresses through non-coding regions (regulatory regions) that lie upstream of the coding region in the same transcript. This leads to an altered transcriptional read-through or translation initiation of downstream coding sequence. Since pathogenic bacteria survive in extreme environmental conditions, a rapid regulatory system or a circuit is required to maintain or control the bacterial virulence. Hence regulatory RNAs are more suitable for controlling bacterial virulence. Many parameters of regulatory RNAs make them more suitable than proteins in controlling gene expression. These include faster control of gene expression compared to proteins, acting at the transcriptional level and thereby reducing the energy cost for the bacterium and rapid clearance of regulatory RNAs compared to proteins. They dictate and overcome the effects of transcription since they act at the post transcriptional level, thereby modifying the expressed mRNAs (Romilly C et al., 2014).

sRNAs act as regulators of virulence genes and are part of signaling pathways in both Gram-negative and Gram-positive bacteria. Most of the sRNAs are found in the core genome while some of them are located in plasmids and pathogenicity islands. The features of sRNAs in *S. aureus* differ from those of enteric bacteria, specifically in *E. coli* and *B. subtilis* where RNase E is required in the sRNA dependent mRNA turnover, and Hfq, an RNA chaperone, is required for the activity of most of the trans-encoded sRNAs. However, *S. aureus* does not have an RNase E, but contains RNases J1, J2 and Y functional homologs. In *S. aureus*, sRNA expression appears to be strain specific and these expression profiles of sRNA might contribute to the diversity of phenotypes (Bohn C et al., 2010; Geissmann T et al., 2009; Pichon C et al., 2005). Toe-printing studies have indicated that several sRNAs carry a conserved and unpaired UCCC (C-rich) sequence motif that helps to bind the Shine-Dalgarno sequence of the target mRNA preventing the 30s ribosomal binding (Geissmann T et al., 2009).

The alternative transcription factor σ^B is activated under different stress conditions and is involved in many regulatory networks in *S. aureus* controlling virulence gene expression, biofilm formation and modulation of antibiotic resistance. Several research groups demonstrated that many sRNAs could be part of the σ^B regulon and by binding to the promoter upstream of the target gene the RNAs could regulate the expression of many genes. One sRNA, *rsaA*, is under the control of a σ^B dependent promoter (Geissmann T et al., 2009, Panel II, Figure 1.7A). In another study, RsaA was shown to repress the synthesis of the global transcriptional regulator MgrA. It forms an imperfect duplex with Shine-Dalgarno sequence and a loop-loop interaction within the coding region of the target mRNA (Panel II, Figure 1.7A). This results in the enhanced production of biofilm and decreased synthesis of capsule formation as revealed by several strain backgrounds (Romilly C et al., 2014). The bi-functional RNAIII with an open reading frame (ORF) encoding a 26aa PSM δ-haemolysin and possessing a regulatory region plays an important role in the gene regulation (Benito Y et al., 2000). RNAIII is characterized by a 14- stem-loop structure in which the C-rich motif sequences in H7, H13, and H14 hairpin loops are shown to interact with several target mRNAs (Felden B et al., 2011, Panel I, Figure 1.7). These hairpin loops of RNAIII act separately or co-ordinately to repress the synthesis of early expressed virulence factors like repressor of toxins-Rot, coagulase and protein A. Rot inhibits the transcription of *hla* and *spa* genes (Panel II, Figure 1.7B). In general, RNAIII functions as an antisense RNA that interacts with target mRNAs. These complexes result in the repression of translation initiation and finally RNase III is recruited for rapid mRNA degradation. A recent report demonstrated the interaction of a 14-nt stretch of the 3′ RNAIII domain with a 14-nt sequence of the *mgrA* upstream region (UTR) by base pairing (Gupta RK et al., 2015). After this initial interaction, the 28-nt 5′ RNAIII base pairs with the 33-nt sequence of *mgrA* UTR. The stabilization of *mgrA* mRNA occurs through this complex (interaction between RNAIII and *mgrA* mRNA), thus inhibiting or blocking the degradation by one or more RNases (Gupta RK et al., 2015, Panel II, Figure 1.7C). Some of the sRNAs carrying UCCC-motif are listed in Table 1.2.

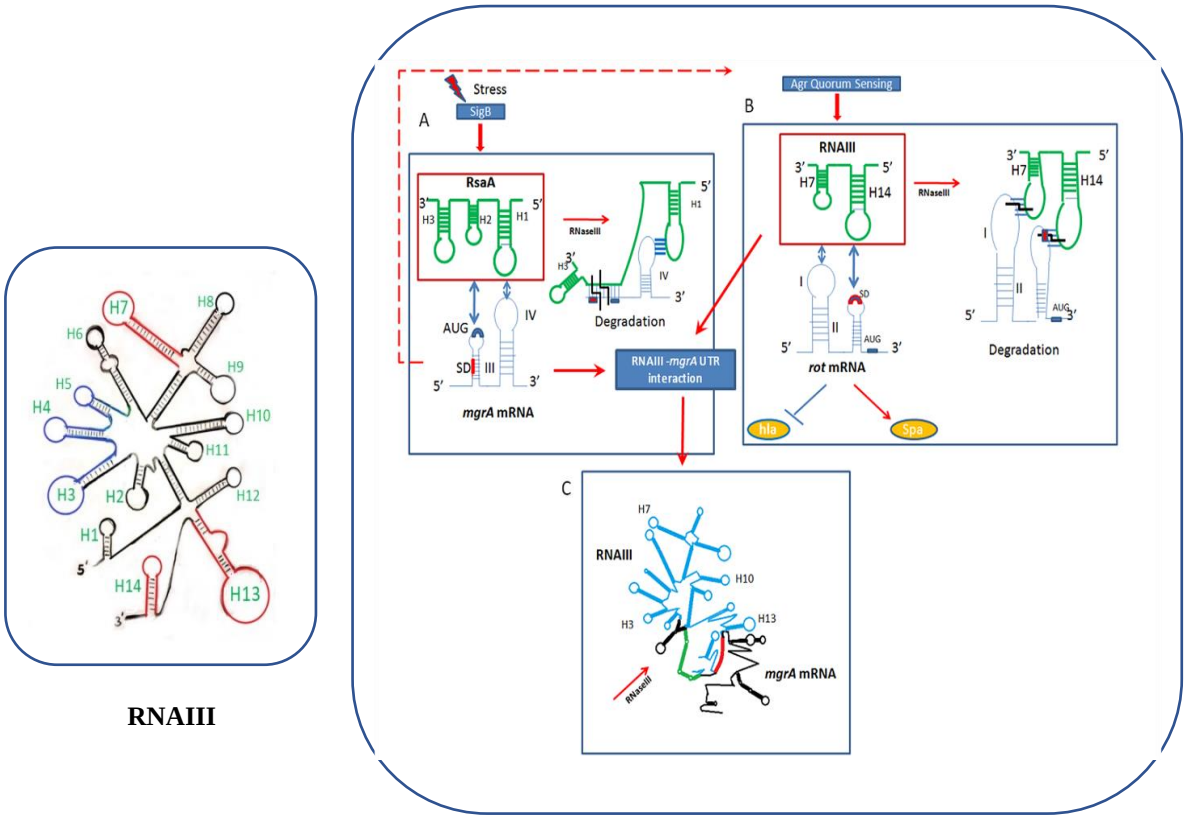


Figure 1.7: Regulatory RNAs and their interaction

Secondary structure of RNAIII shown in Panel I is redrawn from the template taken from Benito Y et al., 2000. Panel II **A**. σ^B -dependent non-coding sRNA, RsaA binds to *mgrA* mRNA (Master or global regulatory RNA) and inhibits translation by preventing the 30S sub unit binding, and recruit's RNase III to induce simultaneous degradation of both RNAs. **B**. RNAIII, an effector of *agr* quorum sensing system down regulates the *rot* mRNA (repressor of toxins). Repression of *rot* mRNA by RNAIII results in the indirect transcriptional regulation of many genes, including activation of alpha-toxin (*hla*) and repression of protein A (*Spa*), both of which are directly regulated by RNAIII. **C**. RNAIII and *mgrA* UTR interaction. The two RNAs are likely to interact initially at the high affinity region (in red) followed by the low affinity region (in green). As a result of this interaction, the mgrA mRNA is stabilized likely by being less vulnerable to RNase degradation. Arrows and bars represent activation and repression respectively.

Table 1.2: List of sRNAs carrying UCCC-motif

Name	UCCC motif	Validation	Reference

Srn_0335	+	NB	*Mäder Uet al., 2016.* *Carroll RK et al., 2016.*
RsaG	+	NB,RT,5′	*Geissmann T et al., 2009.*
RsaA	+	NB,RT, 5′ 3′	*Geissmann T et al., 2009* *Abu-Qatouseh et al., 2010.* *Bohn C et al., 2010.*
ptRsaD	+	NB,RT	*Geissmann T et al., 2009.*
RsaH	+	NB,RT, 5′ 3′	*Geissmann T et al., 2009.* *Bohn C et al., 2010.*
RsaE	+	NB,RT, 5′ 3′	*Geissmann T et al., 2009.* *Bohn C et al., 2010.*
SprX2	+	NB, 5’	*Bohn C et al., 2010.* *Eyraud A et al., 2014.*
SprD	+	NB	*Abu-Qatouseh et al., 2010.* *Pichon and Felden, 2005.*
SprX1	+	NB, 5′	*Bohn C et al., 2010.* *Eyraud A et al., 2014.*
RNAIII	+	NB, 5′3′	*Carroll RK et al., 2016.* *Abu-Qatouseh et al., 2010.* *Pichon and Felden, 2005.* *Novick RP et al., 1993.*

Validation methods used by various reports-Northern blot, RNA extremity mapping, or RT-qPCR. **NB**-Northern blot, **PE**-primer extension, **RACE**-random amplification of cDNA ends, **RT-PCR**- real-time PCR.

1.4.6 Ribonucleases and their role in *S. aureus* pathogenesis

Many ribonucleases are known to be associated with sRNA dependent regulation. RNase III belongs to a highly conserved, ubiquitous class of endoribonucleases among bacteria and it is a double standard RNA (dsRNA) specific enzyme generating short RNA duplexes with 2 nt 3’overhang. *S. aureus* RNase III is a Mg^{+2} dependent endoribonuclease consisting of a catalytic and a dsRNA binding domain and functions as a homodimer (Condon C et al., 2011). A variety of RNA structures such as imperfect duplexes, loop-loop interactions and helices interrupted by bulged residues are recognised and cleaved by *S. aureus* RNase III (Chevalier C et al., 2008). In *S. aureus*, RNase III acts as a co-factor of RNAIII, a major effector molecule of the quorum sensing system. Other than RNAIII, several RNA targets have been identified (Lioliou E et al., 2012) based on *in vivo* immunoprecipitation of wild type strain and a catalytically inactive mutant of RNase III. Interestingly, RNase III has been shown to cleave sRNAs like RsaA in the bulged loop of 5’ hairpin structure at two sites.

RsaA is more stable in the *rnc* mutant strain when compared to the wild type. Another study demonstrated that synthesis of extracellular toxins was decreased in the *rnc* gene inactivated NCTC 8325-4 strain due to the destabilization of RNAIII and sec Y2 mRNA (Liu Y et al., 2011). Thus, many studies suggest that RNase III could be a global regulator of RNA-dependent regulation in *S. aureus*. RNase III known to repress the synthesis of virulence factors in *S. aureus*. More detailed description of sRNAs and their role in *S. aureus* virulence is provided in the next section.

Many enzymes other than RNase III are also believed to be associated with sRNA regulation. In *E. coli*, the sRNA dependent regulation requires both endoribonuclease RNase E and phosphorolytic exoribonuclease PNPase (Massé E et al., 2003). These two enzymes are known to be part of the multiprotein-enzymatic complex, also referred to as the degradosome. Other enzymes like RhlB, an RNA helicase and enolase, a glycolytic enzyme, are also considered to be part of this degradosome. In case of Gram-positive bacteria like *B. subtilis* and *S. aureus*, several enzymes were proposed to be part of the multiprotein enzymatic complex (Marcaida MJ et al., 2006). There was no homolog identified for RNase E in Gram positive bacteria. Several strategies were used to identify the interactions between these enzymatic partners (limited to DEAD-box RNA helicase CshA, RNase J1, RNase J2, RNase Y and PNPase). Although the degradosome like component was conserved in Gram positive bacteria, the interaction network varied between the components. A bacterial two-hybrid analysis was performed to show the interaction between the components (Roux CM et al., 2011). RNase Y, the membrane bound endoribonuclease, cleaves unpaired regions of RNA and is considered to be part of degradosome- like complex along with RNase J1/J2, PNPase, DEAD box RNA helicase, enolase and phosphofructokinase in *S. aureus*. Transcriptome analysis revealed the effect of RNase Y on global gene regulation as it processes and stabilizes the immature transcripts of the global virulence regulator *SaePQRS* (Marincola G et al., 2012). Apart from its major role in virulence genes processing, it is also involved in the turnover of two sRNAs, RsaA and Sau63 and it was shown that the half-life of these two sRNAs increased in the *rny* mutant when compared to the wild type strain (Figure 1.8). The DEAD-box RNA helicase, CshA, belongs to the DEAD-box protein family with typical RNA-dependent ATPase activity and it is known to play a role in the stability of *agrA* mRNA (Oun et al., 2013, Figure 1.8)

In *S. epidermidis*, the exonuclease and endonuclease activity in RNase J1 and RNase J2 respectively was demonstrated and interestingly the expression levels of these two enzymes varied across staphylococcal strains (Raj R et al., 2020). In case of *B. Subtilis*, both RNase J1 and RNase J2 have dual activity of an endoribonuclease and 5'-3' exoribonuclease activity (Mathy N et al., 2007).

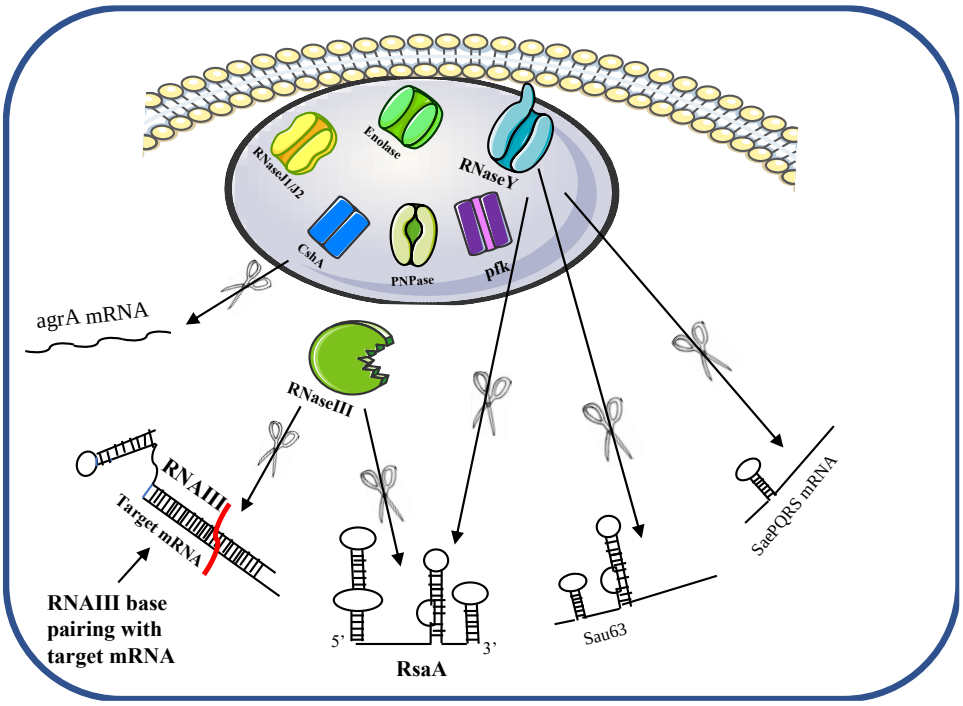


Figure 1.8: Role of RNases in *S. aureus* pathogenesis

This figure represents the interaction between RNA-binding proteins (RBPs) and RNAs in *S. aureus.* The so-called RNA degradosome complex (grey-circle) comprises of RNase Y, RNase J1/J2, the DEAD-box helicase, CshA, enolase, phosphofructokinase (pfk) and PNPase .The degradation of *agrA* (*agrBCDA* operon) mRNA is mediated by CshA. Several mRNAs like *saePQRS* and ncRNAs like RsaA and sau63 are degraded by RNase Y. Double stranded RNA specific RNase III is not yet demonstrated to be part of the degradosome like complex. RNase III is one of the major enzymes playing a role in RNA degradation and maturation. RNase III is involved in the rapid degradation of RNAIII repressed target mRNAs (RNAIII-target mRNA duplexes are cleaved by RNase III). Figure adapted and modified from (Tomasini A et al., 2014).

More details on the relevant components involved in sRNA processing and regulation are stated, where necessary, in the subsequent chapters of this thesis.

Chapter 2

The *Staphylococcus aureus* LVP-7 strain as a model system to understand staphylococcal virulence and pathogenesis

2.1 Introduction

Staphylococcus aureus can infect both healthy individuals and hospital patients with effects ranging from superficial skin infections to life-threatening pneumonia, septicaemia, and endocarditis (Feng YC et al., 2008). This pathogen is also reported to cause severe ocular infections, often resulting in destructive vision loss. This is one of the most common ophthalmic pathogens recovered from conjunctivitis (Behlau I et al., 2008 and Johnson WL., 2021). Despite the notable morbidity linked with ocular infections, not much is known about a specific strain type that may have a preference in causing an eye disease. Similarly, there is very little data on the virulence factors affecting the pathogenicity in this specific biological niche. *S. aureus* has a high genotypic variation and different strains have been grouped into various clonal lineages. A brief outline on the various molecular methods applied in genotyping of *S. aureus* is explained in subsequent sections of this chapter.

The Staphylococcal cassette chromosome *mec* typing (SCC*mec*) helps to classify MRSA strains based on the mobile genetic element they carry on the chromosome. The SCC*mec* cassette is diverse in structural organization and genetic content with a size that varies from 21Kb to 67Kb. Eleven different SCC*mec* types have been reported till date (IWG-SCC, 2009 or http://www.sccmec.org/) and as a representation, five different types of SCC*mec* elements are depicted in Figure 2.1. Pulsed-field gel electrophoresis (PFGE) is a method of choice to investigate the epidemiological out breaks (Tenover FC et al., 1995). Multi-locus sequence typing (MLST) can be applied to understand the long-term genetic changes in the strain (Enright MC et al., 2000, http://saureus.mlst. net). The staphylococcal protein A (spa) typing can also be used to discriminate between the strains by performing a simple PCR and sequencing single locus of *spa* gene (Frenay HM et al., 1996, http://spaserver. ridom.de).

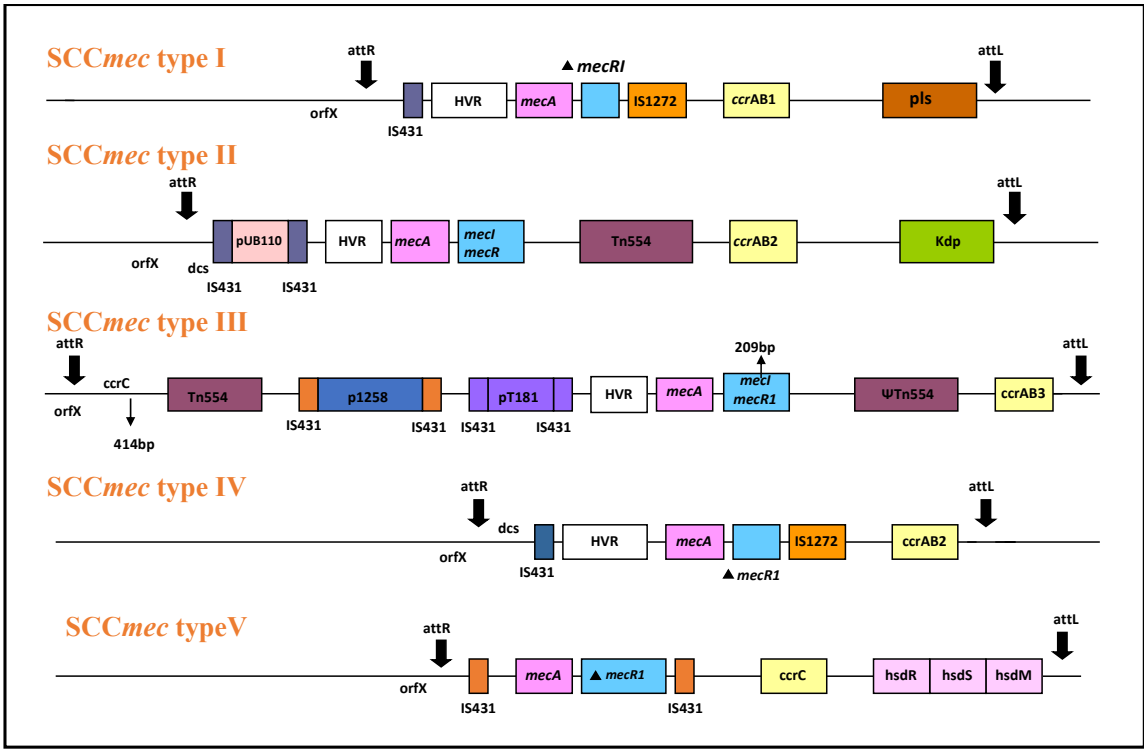


Figure 2.1: Schematic representation of five different types of SCC*mec* elements in *S. aureus*

Integration of SCC*mec* element is sequence specific- attBSCC (bacterial chromosome attachment site-attR-aatL) and aatBSCC is found downstream of an open reading frame (ORF) and designated as orfX, ccr- SCC*mec* carries specific recombinases for the movement of the cassette belonging to invertase or resolvase family. Each of the cassette consists of mec gene complex (mecA, mecI and mecR). HVR-hypervariable region is seen in type I-IV SCC*mec* element. A variety of insertion sequences (ISs-IS1431 and IS1272), plasmids (pT181 and p1258) and transposons (Tn554) are found in different SCC*mec* elements In case of SCC*mec* I, *pls*- plasmin sensitive surface protein, *kdp*-operon involved in ATP-dependent potassium transport across the bacterial cell membrane are present. *hsd*, type I restriction-modification system is found on SCC*mec* cassette type V. (Figure adapted and modified from Hanssen AM et al., 2006).

All the above mentioned methodologies help us to understand the molecular characteristics of human pathogens but this is often not sufficient in the investigation of an outbreak and transmission. Next-generation sequencing (NGS) technology can sequence the entire genome in a very short period. The data generated from this can be used to extract information on virulence, resistance pattern, MLST and Spa type

(Deurenberg RH et al., 2017). In a recent report, bacterial whole genome sequencing (WGS) was used to compare the sequence features of 163 clinical ocular *S. aureus* isolates with 116 fully sequenced *S. aureus* from non-ocular sources. This helped to define virulence factors that might play a role in ocular infections (Johnson WL et al., 2021). In this study, the majority of the ocular strains carried type II SCC*mec* (Staphylococcal Cassette Chromosome, Figure 2.1) element and belonged to ST5 (sequence type) and ST8 clonal lineages. Type II SCC*mec* elements are generally carried by hospital-associated methicillin resistant *Staphylococcus aureus* (HA-MRSA). Previous epidemiological studies from India have demonstrated that community-associated methicillin resistant *S. aureus* (CA-MRSA) and methicillin sensitive *S. aureus* (MSSA) strains belong to different clonal lineages (ST22, ST772, ST672, ST120, ST45 and ST30) containing SCC*mec* type IV or V cassettes. (D'Souza N et al., 2010, Nadig S et al., 2012 and Shambhat S et al., 2012). Earlier studies reported that ST772 clonal lineage causing severe and non-severe eye infections and ST1 and ST88 as emerging clone (Nadig S et al., 2012). The sequence type 88 (ST88) lineage is known to be more prevalent in India (Nadig S et al., 2012 and Chakrakodi B et al., 2014), Africa (Ghebremedhin B et al., 2009, Kpeli G et al., 2017), and China (Liu Y et al., 2009). Several ST88 MRSA isolates causing conjunctivitis have been reported from Nigeria and known to carry type IV SCC*mec* cassette along with a Panton–Valentine leucocidin (PVL) gene (Ghebremedhin et al., 2009). Thus, important elements like pathogenicity islands (SaPIs), chromosome cassettes (SCC), transposons, conjugative plasmids and prophages contribute to the genome variation in different staphylococcal species.

The focus of our study was to establish a *S. aureus* model system in the laboratory to understand phenotypic switching (discussed in chapter 1) in *S. aureus*, hence we were interested in understanding the genomic and proteomic features of LVP-7 strain isolated from an orbital abscess. In this context, we obtained the whole genome sequence of LVP-7 to understand its genome features. This chapter also covers the methodologies we used for draft genome assembly. The draft genome sequence of LVP-7 has been submitted to the GenBank database (GenBank accession ID JADRJK000000000). The associated BioProject and BioSample accession numbers are PRJNA679674 and SAMN16843707, respectively. The raw reads from the Illumina sequencing have been submitted to the Sequence Read Archive (SRA) and are available under the accession number PRJNA679674. Below shown is the schematic

representation of the overview of Next generation Sequencing (NGS) and its applications (Figure 2.2).

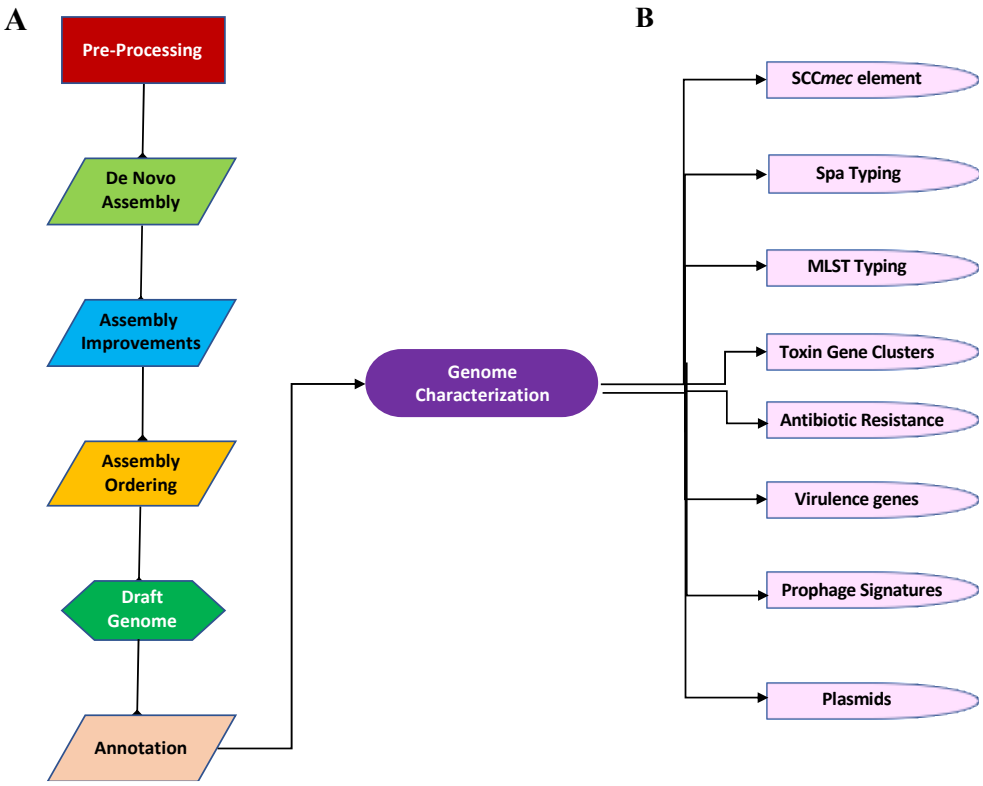


Figure 2.2: Overview of the Next Generation Sequencing (NGS) and its applications to study genome characterization

A detailed pipeline is given on the left side (A) of the flowchart. Down-stream analysis for genome characterization is given in the right side (B) of the flowchart.

2.2 Glossary of scores and Indices

Some of the important terms used in NGS are read length, Phred score and FastQC tool. The **read length** relates to the number of base pairs sequenced from a DNA fragment and **Phred score** is a quality score measuring the quality of the identified nucleobases generated by automated DNA sequencing. **Scaffold and contig** –Scaffold is a reconstructed small portion of the genome obtained from end-sequenced whole-genome shotgun clones. They are nothing but ordered contigs. A contig is a contiguous length of genomic sequence in which the order of bases is known to a high

confidence level (Figure 2.3). The sequence quality is assessed by **FastQC** tool developed by Simon Andrews of Babraham Bioinformatics (Andrews S, 2010). This tool provides an overview of basic quality control for raw next generation sequencing data. The following are the definitions of the modules or analysis that were performed on a sequencing data.

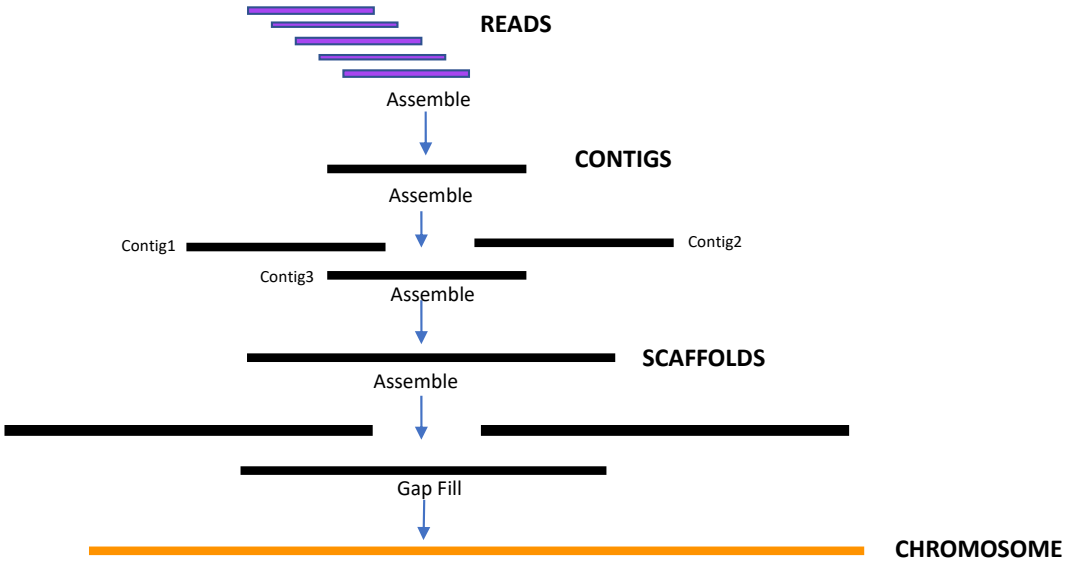


Figure 2.3: Schematic representation of the terms explained in glossary section

(Figure adapted and redrawn from www.thesequencingcenter.com)

2.2.1 Per Base Sequence Quality Scores

Per base sequence quality score measure gives a range of quality values measured as Phred scores across all bases at each position in the FastQC file.

This refers to the step where subset of sequences with presence of low quality scores are identified. Usually the subset of sequences that may be poorly imaged (on the edge of the field view etc) will have universally poor quality scores. Ideally, such sequences with poor quality must be only a small percentage of the total sequences.

2.2.2 Per Base Sequence Content

Per Base Sequence Content plots position of each of the four DNA bases and proportion of the each base position in a file.

2.2.3 Per Sequence GC Content

In this module the GC content is measured and compared with the normal distribution model of GC content.

2.2.4 Per Base N Content

This module plots the percentage of base calls at each position for which a nucleotide (N) was called.

2.2.5 Sequence Length Distribution

This module plots a graph representing how the fragment sizes are distributed in the analysed file. Uniform length sequence fragments are generated by most high throughput sequencers but others may generate sequence fragments of varying lengths. Pipelines are available which will trim sequences that may have resulted in poor quality base calls even within the uniform libraries.

2.2.6 Per tile sequence quality

This quality plot indicates the quality scores from each tile across all of the bases to check for a loss in quality associated with only one part of the flow cell. This plot is generated only in an Illumina library which retains its original sequence identifiers.

2.2.7 Adapter Content

The important sequences which one would want to analyse in their sequencing are the adapter sequences that are added during gDNA library preparation. This step is useful in identifying the adapter content in the library and trimming the same.

2.3 Materials and Methods

2.3.1 Background of the clinical sample

The LVP-7 strain was isolated from an orbital abscess of a patient from LV Prasad eye institute, Bhubaneswar, India (Nadig S et al., 2012). This isolate is a part of a repository maintained in this laboratory.

2.3.2 Bacterial growth

For genomic DNA (gDNA) extraction, LVP-7 glycerol stock stored at -80°C was streaked on chromogenic agar medium (chromAgar, bioMérieux, Marcy-L'Etoile, France). A single colony was picked and grown overnight in BHI (Brain heart infusion) broth in an aerobic condition.

2.3.3 Chromosomal DNA Extraction

Genomic DNA was prepared using the phenol:chloroform method (Arakere G et al., 2005). Briefly, cells grown overnight in an aerobic condition in BHI broth at 37°C were centrifuged at 7,000 rpm for 10 min. The pellet was suspended in TE buffer (10 mM Tris-HCl, pH 8.0, 1mM ethylenediaminetetraacetic acid pH 8.0) containing 1500 U of achromopeptidase (Sigma Aldrich). The cell suspension was incubated at 37°C for 30 minutes. An equal volume of phenol:chloroform (1:1) mixture was then added to the lysed cells and the suspension was mixed vigorously in a vortex mixer. After centrifugation at 10,000 rpm for 10 min, the aqueous phase was transferred to a fresh tube. The DNA was precipitated by addition of 30 µl of 3 M sodium acetate and three volumes of cold 99% ethanol. The mixture was kept at -20°C overnight. The DNA pellet was washed twice with cold 99% ethanol, air dried, and suspended in 500 µl of TE buffer.

2.3.4 Library preparation and Sequencing

Quality of genomic DNA (gDNA) was analysed by Agilent Tape station 2200 (Agilent Technologies). 200ng of gDNA was taken and proceeded for library preparation using NEBNext DNA Ultra II Library Prep Kit (NEB.Inc) as per the

manufacturer's instructions. Final library was quantitated by Qubit 2.0 fluorometer (Thermo Fisher Scientific) and its profile was analysed by Agilent Tape station 2200. Sequencing was done using v3 chemistry in an Illumina HiSeq 2500 instrument (2 × 100-bp paired-end format).

A total of 9,561,330 read pairs were demultiplexed to fastq format using bcl2fastq v2.20.0.422. The quality of the fastq files was ascertained using FastQC v0.11.7 (Andrews, 2010). Adapter content and low-quality reads were removed using TrimGalore (Bolger AM et al., 2014). *De novo* assembly was performed using SPAdes v3.14.1 (Bankevich A et al., 2012) and assembly quality assessed using QUAST v4.5 (Gurevich A et al., 2013). Gap filling, ordering of contigs and optimal scaffolding was done using RagTag (Alonge M et al., 2019) and final re-ordering was performed using *S. aureus* M013 genome as reference (Huang T-W et al., 2012). Following sections gives the detail outline of the entire pipeline used in the study and schematic representation of the work flow is given in the left hand side (A) of Figure 2.2.

2.3.5 Pre-processing and Quality Check

This step involved a thorough quality check of the data and quality improvement measures. Primarily, the pre-processing operations included a quality check (FastQC v0.11.7), removal of adapter content (used during library preparation), quality thresholding (to remove any bad quality reads, Phred Score < 28) were ensured for LVP-7 genome sequence. The adapter content and low quality reads were removed using Trim Galore (Bolger AM et al., 2014)

2.3.6 *De novo* assembly of LVP-7

For the *de novo* assembly of LVP-7, the SPAdes v3.14.1 (Bankevich A et al., 2012) sequence assembler was used to get long contiguous pieces of sequence like scaffolds, contigs, chromosomes etc. Advantages of the assembler is to take a large number of DNA sequences and assembling them together to create the best representation of the original genome. Assembly quality improvement statistics using QUAST v4.5 were checked after SPAdes run (Gurevich et al., 2013).

2.3.6.1 Assembly Improvement and Statistics

The scaffolds obtained as SPAdes v3.14.1 output was further refined to fill the gaps, re-order the contigs using Ragtag v1.0.1 (earlier known as RaGOO) (Alonge M et al., 2019).

2.3.6.2 Reference based Assembly and Ordering

In order to achieve an optimal scaffolding, RagTag v1.0.1 was used for reference-guided genome assembly improvement as the final step to achieve the draft genome of LVP-7. For this purpose, *S. aureus* M013 genome was used as a reference. *S. aureus* strain, M013, a CA-MRSA (clone from Taiwan) strain belonging to ST59, carrying SCC*mec* type V (Huang TW et al., 2012). RagTag tool uses features such as mis-assembly correction and scaffolding, while employing the principles and mechanism of RaGOO.

2.3.7 LVP-7 Genome Annotation

Genome annotation is the process of identifying and labelling all the relevant features on a genome sequence. It includes coordinates of predicted coding regions and their putative products. For genome annotation of LVP-7, a rapid annotation tool called PROKKA v1.14.6 software tool was used against a reference genome (*S. aureus* M013) (Seemann T et al., 2014). This tool is widely used for the rapid annotation of prokaryotic genomes. It employs efficient tools and repositories such as UniProt, Refseq, Pfam and TIGRFAM databases.

2.3.8 Clonal lineages and relationship with *agr* type in Indian *S. aureus* strains

In order to understand the distribution of different clonal complexes (CCs) and relate CCs with *agr* type among the Indian *S. aureus* strains, a minimum spanning tree and a hierarchical clustering analysis was performed.

The profile data for *arcC, aroE, glpF, gmk, pta, tpi*, and *yqiL* loci from Multi-locus sequence typing (MLST) method and auxiliary data containing the details of methicillin resistance pattern, clonal complex, and epidemiology for *S. aureus* Indian isolates were used as inputs for PHYLOViZ Online tool (Ribero-Goncalves B et al.,

2016). All the data used for the analysis were obtained from PubMLST (last update: 13th July 2021, Jolley KA et al., 2018) and few STs were taken from previously published reports (Shambhat S et al., 2012). The PHYLOViZ Online tool uses the goeBURST (Francisco AP et al., 2009) algorithm to generate the minimum spanning tree (MST). The nodes in MST represent the sequence type (ST). The distance between two STs is calculated using Hamming Distance, i.e. the number of differences between the profiles. With the aid of the Tree-cut off algorithm, that splits the MST by removing links above a certain value, those links were removed between the STs which had a distance greater than 4. The nodes in the MST representing the ST were coloured based on epidemiology and were grouped together based on their clonal complexes.

For hierarchical clustering analysis to understand the relationship between clonal lineages and agr type, an input file with profile data obtained for 20 selected STs (PubMLST) was generated for analysis in PHYLOViZ 2.0a (Nascimento M et al., 2017). The average linked method UPGMA (unweighted pair group method with arithmetic mean) on the Hamming distance matrix for hierarchical clustering was used. In addition, the information on clonal complex (data obtained from PubMLST), and *agr* type (Nadig S et al., 2012, Shambhat S et al., 2012, Liu Y et al., 2009, Kashif A et al., 2019) for each STs were incorporated.

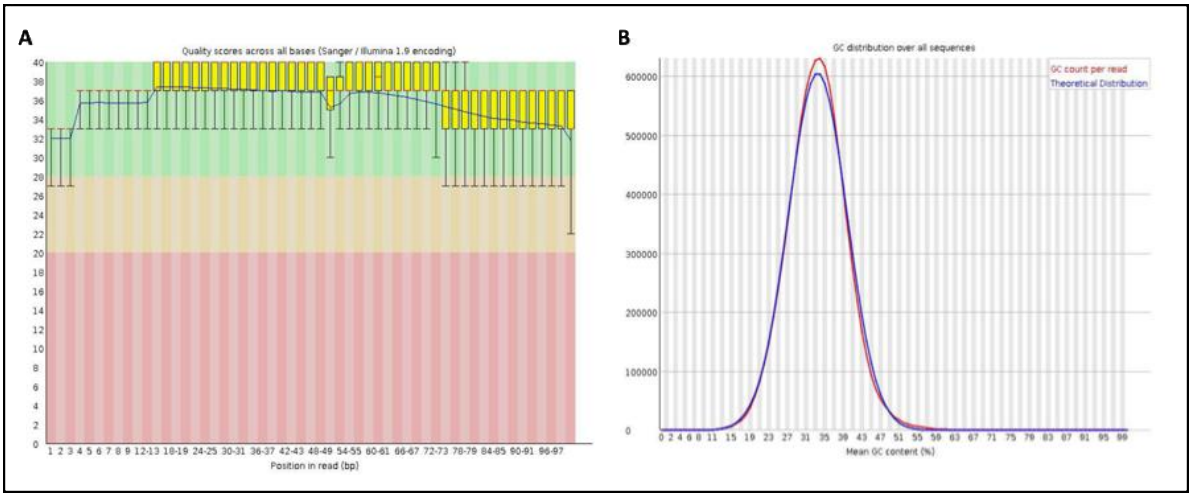
2.3.9 Relationship between antimicrobial resistance to methicillin and clonal lineages in *S. aureus*

In order to understand the distribution of methicillin resistance or susceptible pattern Indian *S. aureus* strains among various MLST, distance matrix was generated by selecting certain STs (nodes) with either a methicillin resistant/ susceptible isolate from the MST in PHYLOViZ Online tool (Ribero-Goncalves B et al., 2016) . The MST was generated as mentioned in the previous paragraph, where the nodes representing STs were colored based on the methicillin resistance pattern.

2.4 Results

2.4.1 Features of the *de novo* assembly and assembly statistics

De novo assembly involves assembling the genome from a large number of DNA fragments (long or short), without using a reference genome. The draft *de novo* genome assembly of LVP-7 was obtained and reported to the NCBI database. This genome harbours a Panton-Valentine leukocidin phage, a type V staphylococcal cassette chromosome *mec* element, the δ-hemolysin-converting Newman phage ΦNM3, and the pathogenicity island SaPI3, encoding the super antigen enterotoxin B. The sequencing raw reads passed the quality threshold (phred score > 28), and the associated quality parameters from the FastQC report are represented as plots in Figure 2.4. The base calls of this assembly reads lie in the high-quality range as shown in the green region in Figure 2.4A. The average quality per read achieved was the phred score of 37 (Figure 2.4B). The average GC content of the genome is 32.7% (Figure 2.4D).



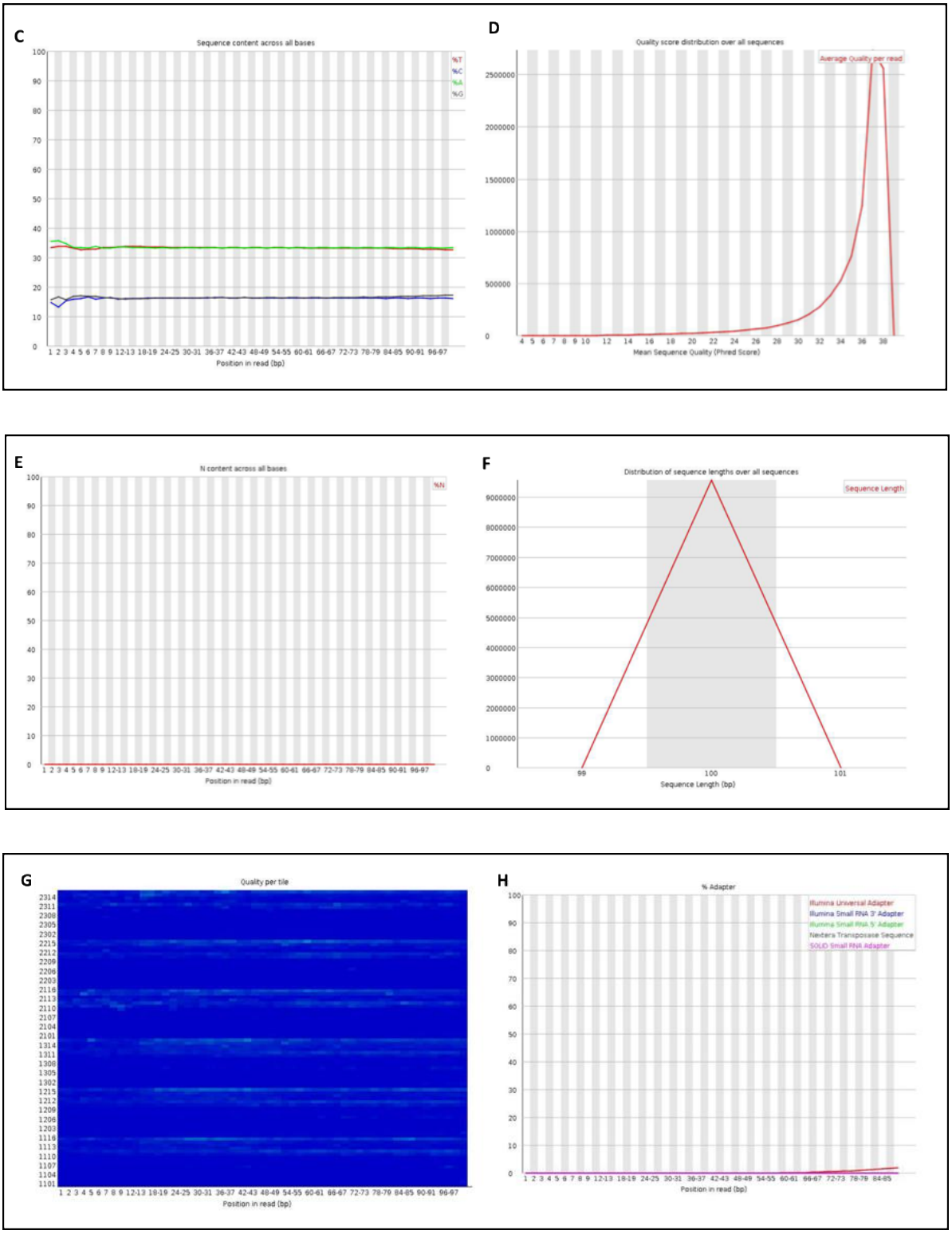


Figure 2.4: Representation of quality check graphs for draft genome of LVP-7

A. Per Base Quality scores. Higher the scores, the base call is more confident. The yellow box represents the inter-quartile range (25-75%), central red line and blue line indicate the median value and mean quality respectively. The upper and lower whiskers of the box plot represent the 10% and 90% points. The coloured background red, orange

and green represent poor, reasonable, good quality calls respectively. **B. Per Sequence Quality Scores.** The average per read quality was of phred score 37. **C**. **Per Base Sequence Content.** The percentage of the bases A and T, G and C were between 30 and 40, 10 and 20 respectively. **D**. **Per sequence GC content.** The mean GC content was reported to be 32.73%. **E**. **Per Base N Content.** The base calls were confident without Ns. **F. Sequence Length Distribution.** The distribution of fragment sizes was uniform indicated as a single peak. **G. Per tile sequence quality.** The quality scores obtained from each flow cell tile across all the bases were of good quality (indicated as blue colour). **H. Adapter Content.** The raw reads were clean without any overrepresented sequences (adapter dimers).

2.4.2 Assembly and annotation:

A *de novo* assembly of these raw reads was done using SPAdes v3.14.1 to get the assembly in the form of scaffolds / contigs. Nearly, 117 contigs were obtained and the largest contig size was 2,777,888 bp and all the contigs were further processed for Gap filling, ordering of contigs, and optimal scaffolding using RagTag. The scaffold assembly generated had good statistics with respect to features like N50, the largest contig length, No. of Ns, scaffold length etc. These results are tabulated in Table 2.1.

Table 2.1: Features of the outputs from SPAdes and RagTag (reference genome-M013) files

Category	SPAdes Scaffolds	RagTag_M013
# contigs (>= 0 bp)	117	69
# contigs (>= 1000 bp)	45	12
# contigs (>= 5000 bp)	35	6
# contigs (>= 10000 bp)	32	2
# contigs (>= 25000 bp)	25	2
# contigs (>= 50000 bp)	15	1
Total length (>= 0 bp)	2853672	2858759
Total length (>= 1000 bp)	2835718	2847191
Total length (>= 5000 bp)	2804566	2832860
Total length (>= 10000 bp)	2785280	2804297
Total length (>= 25000 bp)	2653233	2804297
Total length (>= 50000 bp)	2264955	2777888

# contigs	53	16
Largest contig	385195	2777888
Total length	2841689	2850087
GC %	32.72	32.73

The gap-filled contigs were finally re-ordered using the reference genome of *S. aureus* M013. The annotation of this draft genome was done using PROKKA v1.14.6 to highlight 2,722 protein coding genes (Table 2.2).

Table 2.2: Genome Assembly Statistics and annotation of the Staphylococcus aureus ST88 strain LVP-7

Feature	Value
Draft *de novo* assembly statistics	
No. of contigs	69
No. of contigs >500 bp	16
Largest contig size (bp)	2,777,888
Genome size (bp)	2,858,759
G+C content (%)	32.73
N_{50} (bp)[a]	2,777,888
No. of *N*s per 100 kbp	217.54
Genome annotation features	
No. of ORFs[b]	2,722
No. of mRNAs and rRNAs	2,668
No. of tRNAs	53
No. of tmRNA[c]	1
Positive strand (bp)	1,323
Negative strand (bp)	1,399

[a]N_{50}, shortest contig length at 50% of the total genome length.
[b]ORFs, open reading frames.
[c]tmRNAs, transfer-messenger RNAs

After annotating using the PROKKA v1.14.6 tool, downstream analyses were performed using SCC*mec*Finder, SPAtyper v0.1.0, and TA finder to identify important genome features. The NAuRA-curated enterotoxin database was used to predict toxin gene clusters. ResFinder and PathogenFinder were used to identify the antibiotic resistance and virulence gene clusters, respectively. Prophage Hunter and PhiSpy

helped to identify prophage gene signatures. Default parameters were used for all software unless otherwise specified. Detailed results of the downstream analysis is explained further in the chapter.

2.4.3 Important genomic features of ST88, LVP-7 strain

The following section describes the important genome features of LVP-7 and its relevance will be further discussed.

2.4.3.1 Type V SCCmec cassette in Staphylococcus aureus

LVP-7 strain carried a type V SCC*mec* (5C2) element and was identified using SCCmecfinder. The type V SCC*mec* element is estimated to be 27,624 bp and does not carry any antibiotic resistance genes besides *mecA*. The *mecA* gene is flanked by two insertional sequences, IS431. Unlike the extant SCC*mec* types, it carries a set of foreign genes encoding a type I restriction-modification system composed of *hsdR*, *hsdS*, and *hsdM* that might play a role in the stabilization of the element on the chromosome (Ito T et al., 2004). Schematic representation of typical type V SCC*mec* cassette is shown in Figure 2.5.

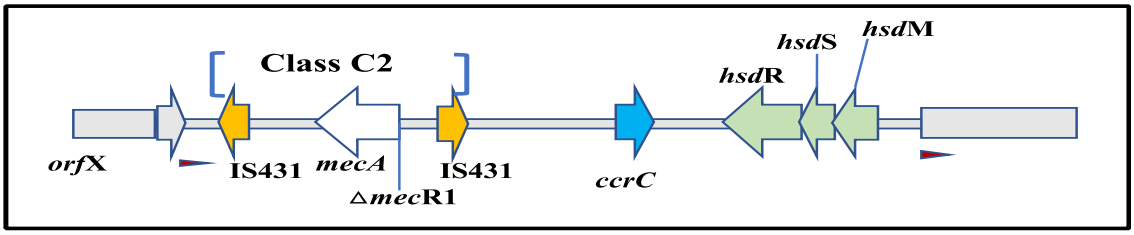


Figure 2.5: Classical representation of SCC*mec* type V cassette

MecA gene (white arrow) is flanked by two insertional sequences (IS431- yellow arrow). The restriction modification system genes (*hsd*M, *hsd*S and *hsd*R-green arrow).

2.4.3.2 Staphylococcal Pathogenicity island 3 (SaPI3) in LVP-7 strain

LVP-7 carried a mobile pathogenicity island 3 (SaPI3) on its genome. The pathogenicity island SaPI3 in this ST88 strain harbours the *seb* enterotoxin (Figure 2.6). Typically, SaPI3 varies in size between 15 to 17 kb and SaPI3 has a size of about 15.6 kb (Novick RP et al., 2007). They carry a conserved set of genes like the integrase, the Rep protein (usually annotated as helicase-like or primase-like), the terminase small subunit homolog and an integrase gene flanked by two open reading frames (ORFs)

encoding transcriptional regulatory proteins. These specific ORFs are unique to SaPI3 and no orthologues are observed in the staphylococcal genome (Novick RP et al., 2007). A recent study demonstrated the role of *seb* toxin in an isogenic mutant in a representative ST59 isolate and assessed its virulence potential in mouse infection models (Bae JS et al., 2020). The authors demonstrated the significant role of *seb* toxin in the hypervirulence of ST59, a predominant Asian CA-MRSA lineage in mouse infection models. They detected a significant contribution of *seb* toxin to cause a systemic ST59 infection that was associated with a cytokine storm. Interestingly, full length SaPI3 carrying *seb* gene was also identified in the LVP-7 genome.

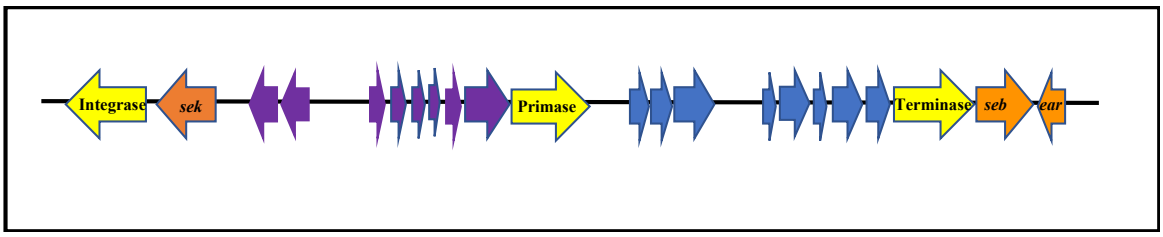


Figure 2.6: Schematic representation of pathogenicity island-3 (SaPI3) in LVP-7 strain

Integrase, Primase and Terminase (represented in yellow arrow), *sek*, and *seb* - enterotoxins, *ear* -*E. coli* ampicillin resistance (orange arrow), purple and blue arrow-genes of unknown function involved in the SaPI life cycle. (Figure adapted and modified from Bae T et al., 2006).

2.4.3.3 Phages identified in LVP-7 strain

S. aureus phages are known to carry diverse virulence factors. The LVP-7 strain carried a PhiNM3 prophage and Panton–Valentine leukocidin (PVL) Phage ΦSa2958PVL on its genome. A few details about the characteristics of these two phages is given below.

PhiNM3-ΦNM3 prophage, hlb-converting is thought to play important roles during *S. aureus* infections of human hosts and known to carry innate immune modulatory genes like sea, *sak*, *chp* and *scn*. (Bae T et al., 2006).

Panton–Valentine leukocidin (PVL) Phage ΦSa2958PVL - The ΦSa2958 is a PVL phage carrying two components of the PVL toxin (LukF-PV and LukS-PV). Both the

components act synergistically and cause damage on cell membranes by forming pores, resulting in the lysis of polymorphonuclear leukocytes and macrophages (Ma XX et al., 2008). The ΦSa2958PVL phage is 46,046 bp in length and it was identified in clonal lineages ST5 and ST30 carrying type II and type IV SCC*mec* elements respectively.

2.4.3.4 Virulence genes in LVP-7 strain

Similar to many of CA-MRSA strains, the LVP-7 strain also contains many virulence factors (Figure 2.7). The virulence factors identified in the LVP-7 genome are the secreted super antigens (*sea, seb, seg and sek*), Panton-Valentine leukocidins (*pvl*) and the exfoliative toxin type A (*eta*).

2.4.3.5 Multi locus sequence type (MLST) and Spa type of LVP-7

The LVP-7 strain belonged to ST88 with an allelic profile pattern 22-1-14-23-12-4-31 identified using MLST 2.0 (Larsen MV et al., 2012) and Spa type t2526 with spa repeat units 07-12-21-17-13-13-13-34-33-13 identified using spa Typer 1.0 (Bartels MD et al., 2014) from the draft genome of LVP-7. The seven allelic loci of the seven housekeeping genes (*arcC, aroE, glpF, gmk, pta* and *yqil*) and for spa repeat units on the surface of the Spa protein were identified.

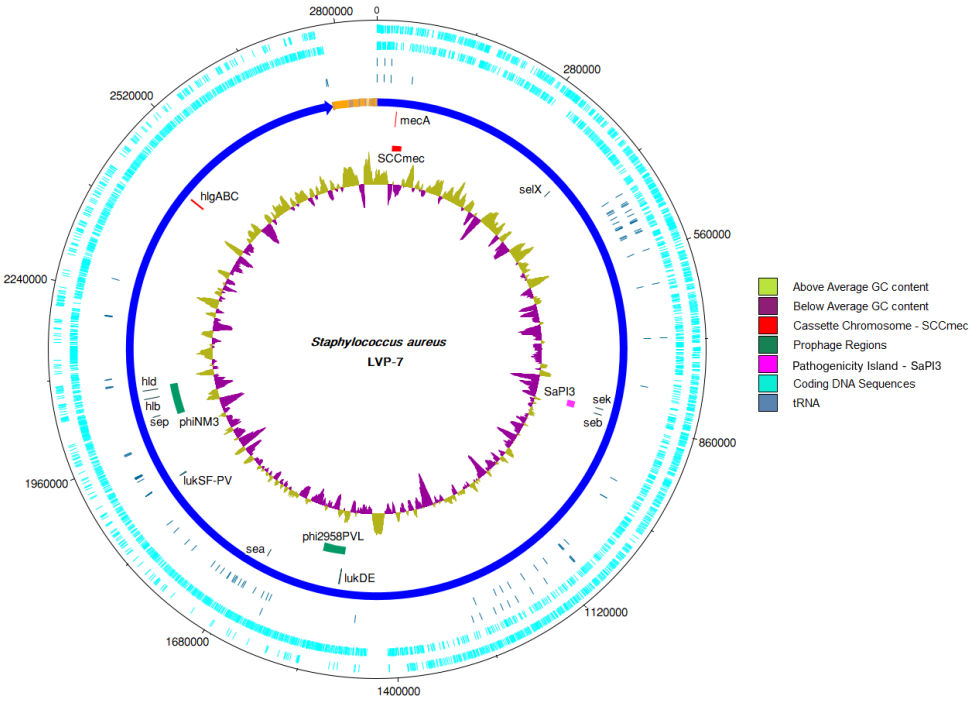


Figure 2.7: Major genetic elements in *Staphylococcus aureus* ST88 strain LVP-7

The inner most circular track (track 1) represents GC content. Track 2 displays select mobile genetic elements including SCC*mec*, the pathogenicity island SaPI3 and the Panton–Valentine leucocidin (PVL) and phiNM3 prophages. Track 3 displays select virulence genes. While track 4 represents major (blue) and minor (orange) contigs, tracks 5 and 6 show the location of tRNAs. The outer tracks (7 and 8) represent coding sequences. This representation was made using DNAPlotter (Carver T et al., 2009).

In an effort to understand the clonal distribution of different Indian *S. aureus* and branching of ST88, a minimum spanning tree was generated. Subsequently, in order to draw a relationship between clonal lineages, methicillin resistance and *agr* groups, a distance matrix and hierarchical clustering were generated.

2.4.3.6 Clonal distribution of Indian *S. aureus* strains

In an effort to understand the clonal distribution of Indian *S. aureus* strains, particularly focussing on ST88, we carried out minimum spanning tree analysis of various clonal lineages. The tree was generated using PHYLOViZ Online tool for the reported STs from India (Figure 2.8). As MLST is considered a powerful tool to understand phylogeny, thereby defining the taxonomic structure and evolutionary significance of diverse bacteria, a relatively detailed picture of global dissemination of

S. aureus can be obtained by MLST (Enright MC et al., 2000). In this context, a genotype (ST) which gave rise to each clonal complex (CC) is also known as clonal ancestor. Clonal complexes are groups in which each isolate shows identity in five or more of the seven loci with at least another isolate. Further, clonal complexes give rise to single locus variants (SLVs) which are assumed to be the direct descendants of ancestor clone differing at only one locus (Day NP et al., 2001). A majority of the Indian isolates belong to clonal complexes CC1, CC5, CC22, CC8, CC15, CC30, CC45, and CC97 (Shambhat S et al., 2012). The information on sequence type (STs) of most of Indian *S. aureus* and their clonal lineages is represented in Figure 2.8. Previous epidemiological studies reported that the majority of HA-MRSA belonged to ST239 (CC8) and other genotyping methods (PFGE, Spa and SCC*mec*) revealed that Indian isolates were genetically and clonally like Hungarian and Brazilian MRSA isolates (Arakere G et al., 2005, Figure 2.8). The presence of epidemic methicillin resistance *S. aureus* (EMRSA)-15 belonging to ST22 (CC22) clonal lineage and its variants in healthy and diseased individuals was also reported from India (Nadig S et al., 2010). These isolates were PVL positive and susceptible to non-β-lactam antibiotics, a characteristic feature of CA-MRSA. Similarly, the ST88 strain (LVP-7) carried a type V SCC*mec* element and PVL phage. CA-MRSA belonging to ST-22-MRSA-IV and ST772-MRSA-V have slowly replaced ST-239-HA-MRSA-III by infiltrating into healthcare environments in India (D'souza N et al., 2010). In another study, clinical samples collected from inpatients from various Indian cities as well as nasal carriage by rural and urban healthy volunteers revealed that ST22 and ST772 were dominant MRSA clones (Shambat S et al., 2012). The same study reported the emergence of ST1208 (CC8) and ST672 as emerging lineages. The majority of the isolates from various eye infections belonged to major clonal lineage, ST772 (CC1). Other STs like 672, 121, 72 and ST88 were also reported (Nadig S et al., 2012). In Figure 2.8, ST88 is also highlighted to show that it remains as a singleton, whereas ST2066 (change in allelic loci of *yqil* gene was observed) and ST3241(change in allelic loci of *pta* gene was observed) were the single locus variants (SLV) of ST88.

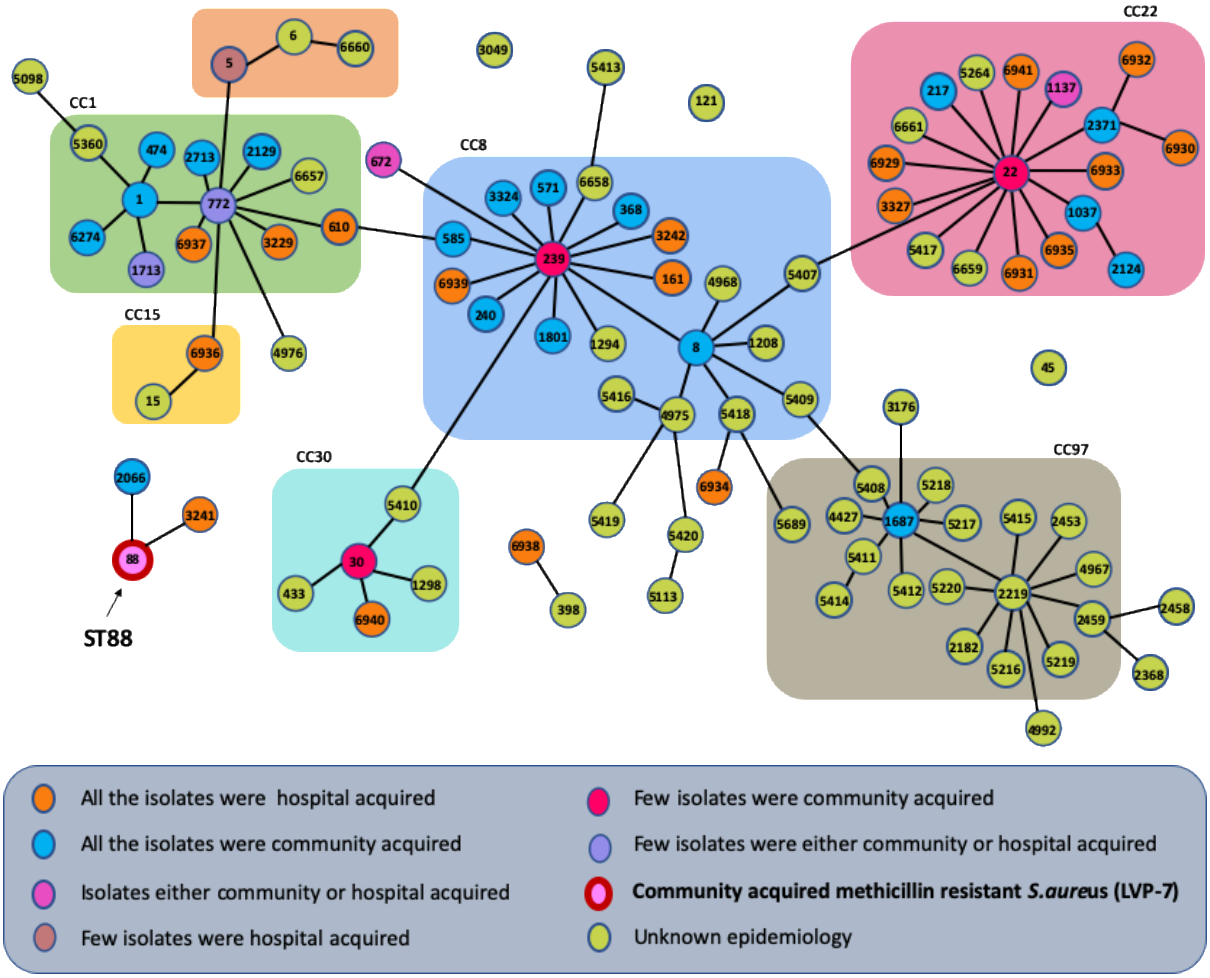


Figure 2.8: Minimum Spanning Tree analysis for Indian *S. aureus* strains

Minimum spanning tree analysis (STs taken from PubMLST database) for Indian *S. aureus* isolates are represented in the figure shows several clonal complexes (CC1, CC5, CC22, CC8, CC15, CC30, CC45 and CC97). The nodes represent ST and the nodes were linked if the difference between the MLST profiles is less than 5. The figure was generated using PHYLOViZ Online tool. The node color represents the epidemiology and the nodes were grouped based on their clonal complexes (CC) which is also represented in the figure. ST88 (LVP-7) has remained as a singleton (light pink node with red outline). ST2066 and ST3241 are single locus variants (SLV) of ST88.

2.4.3.7 Relationship between clonal lineages and methicillin resistance in *S. aureus*

As most of the virulence and antimicrobial resistance genes in *S. aureus* are located on the mobile genetic elements (MGEs) such as SCCs (staphylococcal chromosome cassette), plasmids, bacteriophages, transposons and pathogenicity

islands, it was intriguing to draw a relationship between clonal lineages and methicillin resistance. We generated a distance matrix to understand the distribution of methicillin resistant or susceptible Indian *S. aureus* strains among various multi-locus sequence types (MLST) (Figure 2.9). Interestingly, the MGEs make up 15-20% of the genome (Haaber J et al., 2017). Antimicrobial resistance is acquired by *S. aureus* mainly due to horizontal transfer and plasmids are known to play an important role this transfer (Jensen SO et al., 2009) and methicillin resistance is carried by *mecA* gene which codes for penicillin binding protein (PBP2A) and shows low affinity for β-lactam antibiotics. Several studies have reported that methicillin susceptible *S. aureus* (MSSA) acquires SCC*mec* by horizontal transfer thus converting MSSA to MRSA (Hiramatsu K et al., 2001 and Robinson DA et al., 2004). This analysis demonstrates that most of the strains belonging to ST88 showed resistance to methicillin including LVP-7.

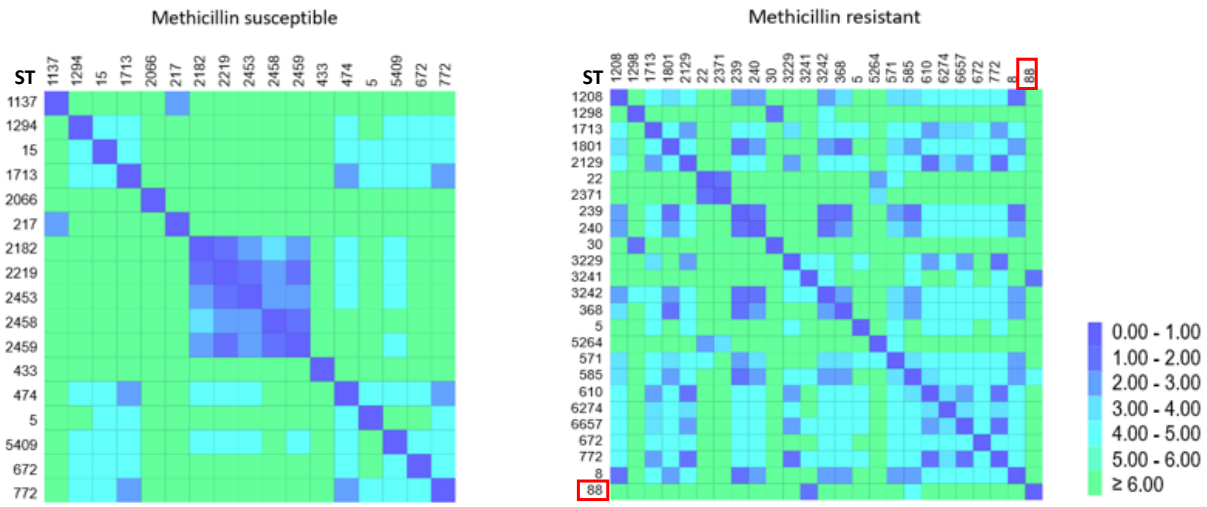


Figure 2.9: Distance matrix constructed for Indian *S. aureus* MLST profiles

Multi-locus sequence type (MLST) data for Indian strains was collected from PubMLST database (one or more isolate in each ST is either methicillin susceptible or resistant). The distance matrix was constructed using PHYLOViZ Online tool. Matrix cells were ordered according to the STs and color represent distances calculated through pairwise comparisons of profiles. ST88 is highlighted in red as most of the strains in this clonal lineage were methicillin resistant.

2.4.3.8 Relationship between clonal lineages and *agr* types

In order to correlate relationships between the accessory gene regulator (*agr)* type and clonal lineages in Indian *S. aureus* strains, a hierarchical clustering analysis was performed using PHYLOViZ 2.0a based on UPGMA (unweighted pair group method with arithmetic mean). This analysis was performed to highlight the *agr* loci with the ST88 clonal lineage. Since the *agr* locus carries the variable region of the core genome, it is strongly linked with clonal lineages. Polymorphism in the *agr* locus of *S. aureus* can be classified into four *agr* groups, I, II, III and IV (Ji G et al., 1997). Several studies have demonstrated the association among certain *agr* groups, clonal complexes, types of diseases and corresponding virulence factors. Jarraud and colleagues speculated the preferential association between *agr* alleles and a particular genetic background (Jarraud S et al., 2002). This association might activate the virulence factors more efficiently. They found that endocarditis is mainly caused by *agr* groups I and II isolates while *agr* group III causes toxic shock syndrome (TSS) which is associated with the expression of exotoxin gene, such as TSST-1. Exfoliate toxins causing generalised exfoliate syndromes were closely related to *agr* group IV isolates. Hierarchical clustering analysis performed using PHYLOViZ 2.0a reveals the relationship between *agr* and clonal complexes (Figure 2.10). Several studies reported that clonal lineages like CC8, CC22, CC45 harboured *agr* I whereas CC5 and CC15 isolates carried *agr* II (Holtfreter S et al., 2007). A majority of the strains in CC30 and some isolates in CC1 carried *agr* III, while *agr* IV occurred only in CC121 (Jarraud S et al., 2002 and Yan Xu et al., 2021). The *agr* locus of most of the ST88 strains (singletons) including LVP-7 is part of *agr* III group (Figure 2.10).

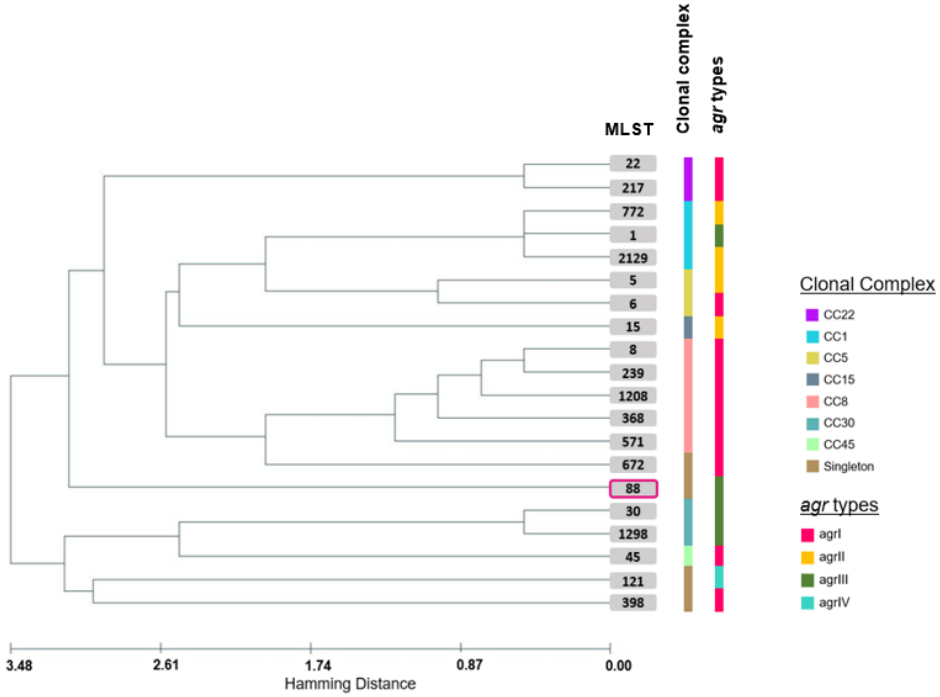


Figure 2.10: Evolutionary significance of Indian *S. aureus* isolates

Hierarchical clustering analysis performed using PHYLOViZ 2.0a based on UPGMA (unweighted pair group method with arithmetic mean) method (data taken from PubMLST). ST and clonal complex distribution among different *agr* groups is also shown. The abbreviations are MLST- multi-locus sequence typing, CC-clonal complex, *agr*-accessory gene regulator, ST-sequence type. ST88 is highlighted in red as most the strains reported in this lineage carried *agr* type III system.

2.5 Discussion

This study focussed on sequencing a *S. aureus*, CA-MRSA strain LVP-7 isolated from an eye infection. The draft genome assembly of LVP-7 using Illumina HiSeq had a longest contig length of 2,777,888bp in a total genome size of 2,858,759bp. The G+C content is 32.73% with 2,722 protein coding genes, 53 tRNAs, 2,668 mRNAs and rRNAs (Table 2.2). LVP-7 carries a type V SCC*mec* element and belonged to ST88 clonal lineage similar to other eye isolates reported from Africa and China except in the SCC*mec* element (Figure 2.5, Ghebremedhin B et al., 2009, Kpeli G et al., 2017 and Liu Y et al., 2009). Most of the ST88 strains reported earlier carried a type IV SCC*mec* element. The LVP-7 strain belonged to the spa type t2526 and diverse spa types have

been reported in this ST88 clonal lineage (Ghebremedhin B et al., 2009). The enterotoxins are considered as super antigens (SAgs), as they have the ability to bind to major histocompatibility complex II, resulting in widespread, non-specific T-cell activation (Spaulding AR et al., 2013). Similarly, the presence of enterotoxins *selv, sed, sej,* associated with SaPIs in ocular isolates were reported by Johnson WL et al., 2021. Interestingly, LVP-7 carried an enterotoxin *seb* along with *sea, seg* and *sek* (Figure 2.6). The pathogenicity island SaPI3 carrying the *seb* toxin gene is of particular interest as it has been demonstrated to be toxic to corneal epithelial cells and elicit the cytokine expression in an in-vitro model (Thakur A et al., 1997). Similarly, a recent report demonstrated the role of *seb* toxin in superantigenicity of a ST59, CA-MRSA strain. The higher activity of *seb* was observed in spleen in a mouse model leading to systemic infection (Bae JS et al., 2020). LVP-7 also carried a hlb-converting phage phiNM3 which is known to play important role in *S. aureus* infection. Another prophage carried by LVP-7 is the PVL phage, ΦSa2958PVL, known to lyse polymorphonuclear leukocytes and macrophages during *S. aureus* infection in the host (Figure 2.7, Ma XX et al., 2008). The virulence factors encoded by these prophage may allow the pathogenesis of varied staphylococcal diseases. A minimum spanning tree analysis demonstrated that ST88 clonal lineage remained as a singleton and has given rise to two single locus variants (SLVs) and might emerge as a parental clone with more SLVs and double locus variants (DLVs) adding to the group (Figure 2.8). Also the Indian strains belonging to ST88 were mostly resistant to methicillin (Figure 2.9) Interestingly, the *agr* locus of this ST88 clonal background belongs to type III *agr* (Figure 2.10). The *S. aureus agr* quorum sensing gene cluster upregulates the production of secreted toxins and down regulates the production of cell-associated virulence factors (Ji G et al., 1997) . Hence the association of virulence factors, antibiotic resistance and *agr* type with specific clonal lineages may help to assess the severity of an infection caused by this *S. aureus strain.*

Chapter 3

Transcriptome analysis of LVP-7 - A community associated methicillin resistant *Staphylococcus aureus* (CA-MRSA) strain

3.1 Introduction

Staphylococcus aureus is an opportunistic pathogen causing both nosocomial and community-acquired infections. This pathogen is known to cause serious ocular infections and is one of the most common pathogens recovered from conjunctivitis (Behlau I et al., 2008). Earlier reports showed USA300, an epidemiologically important isolate was a leading cause of severe ophthalmic manifestations (Rutar T et al., 2006). In recent years, high-throughput sequencing of RNA converted to cDNA (RNA-sequencing) has helped researchers gather information on regulators of gene expression and the phenotypic properties of diverse *S. aureus* strains. Several factors involved in different stages of biofilm formation in both hospital associated methicillin resistant *S. aureus* (HA-MRSA) and community associated methicillin resistant *S. aureus* (CA-MRSA) were identified by employing transcriptomic profiling (Tomlinson BR et al., 2021). In general, biofilm formers have a high tendency to exhibit resistance to antimicrobials (in the form of small colony variants) and it has been reported that *S. aureus* isolated from hospitalized patients have a high degree of biofilm forming ability (Neopane P et al., 2018). Several regulatory mechanisms are known to play an important role in the *S. aureus* virulence and one regulatory system which helps in the phenotypic switching of *S. aureus* is the *agr* quorum sensing (QS) system. Hence, an understanding of the transcriptome of an isolate helps to extract information on the functional elements present in the genome (Wang Z et al., 2009). A transcriptome study of USA300, a well-studied community associated methicillin resistant *S. aureus* (CA-MRSA) revealed the differential sRNA expression in physiologically relevant conditions. The annotation files from this work are useful for various comparative transcriptome analysis (Carroll RK et al., 2016).

Gene Ontology (GO) is a comprehensive method to classify genes into specific classes based on Molecular function (MF), Cellular component (CC) and Biological processes (BP). The main goal of the GO project is to develop ontologies leading to controlled and structured communication describing the ‘terms’ and ‘classes’ outlining the functions gene products and also the terms involved with each gene in performing a particular function (Ashburner M et al., 2000, Figure 3.1). Thus, GO annotation is considered to describe the functional properties of genes and biological process in which they are involved (Xijin Ge S et al., 2020).

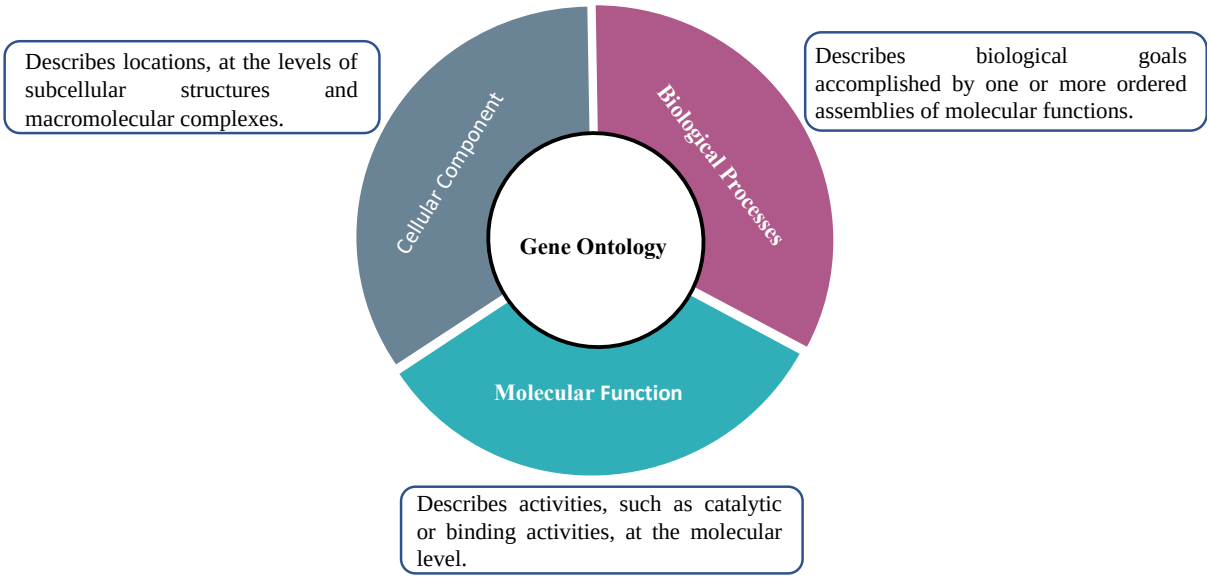


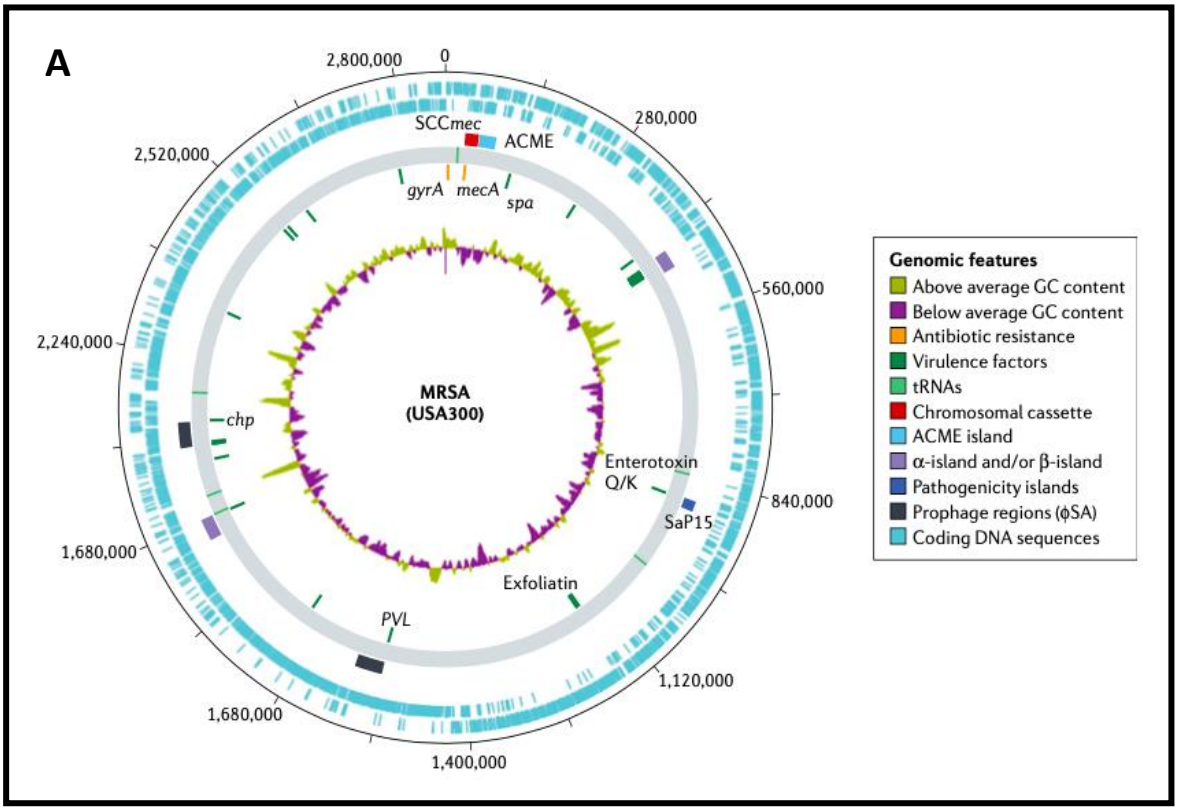
Figure 3.1: Pictorial representation of various Functional categories of Gene Ontology

Gene Ontology (GO) incorporates information that includes gene product features and biological sequences. While Molecular function (MF, cyan colour) describes activities, such as catalytic or binding activities, at the molecular level, Biological Process (BP, purple colour) describes biological goals accomplished by one or more ordered assemblies of molecular functions. Cellular Component (CC, grey colour) describes locations, at the levels of subcellular structures and macromolecular complexes.

In the analysis described in this chapter, we chose the transcriptome data set of USA300, a well-studied CA-MRSA strain from the GEO database to compare the different GO terms with the LVP-7 strain. We performed Gene Ontology analysis of the transcriptome to determine the classes of housekeeping and virulence genes expressed by this CA-MRSA strain. A detailed description of the virulence factors of both CA-MRSAs, LVP-7 and USA300 are shown in Figure 3.2. The next section briefly highlights some of the important virulence factors present in both strains.

3.1.1 Virulence factors in LVP-7 and USA300 CA-MRSA strains

Several distinguishing features of USA300 make it an epidemiologically important clonal lineage. The genome encodes a type IV SCC*mec* element, Panton-Valentine leukocidin (PVL) gene and the ACME (Arginine catabolite mobile element), a 30.9kb element that is known to help in the fitness of the strain (David MZ et al., 2010). The genome sequence of multidrug resistant USA300 strain, FPR3757, belonging to the clonal lineage 8 displayed interesting genetic features (Diep BA et al., 2006) In a rat model study, Montgomery and colleagues (2008) demonstrated that USA300 had higher transcription levels of the global regulators agr, sarA and saeRS including PVL and alpha-toxin (hla) when compared to other CA-MRSA strains (Montgomery CP et al., 2008). Hence, some of these features play a role in increasing the virulence and pathogenesis of USA300 strain (Figure 3.2A). The draft genome assembly of the LVP-7 strain provided a basis to understand the expression of important virulence factors carried by this S. aureus strain. Like USA300, the LVP-7 genome also carries a type V SCCmec element apart from the pathogenicity island SaPI3 (Figure 3.2B, Nadig at el., 2021).



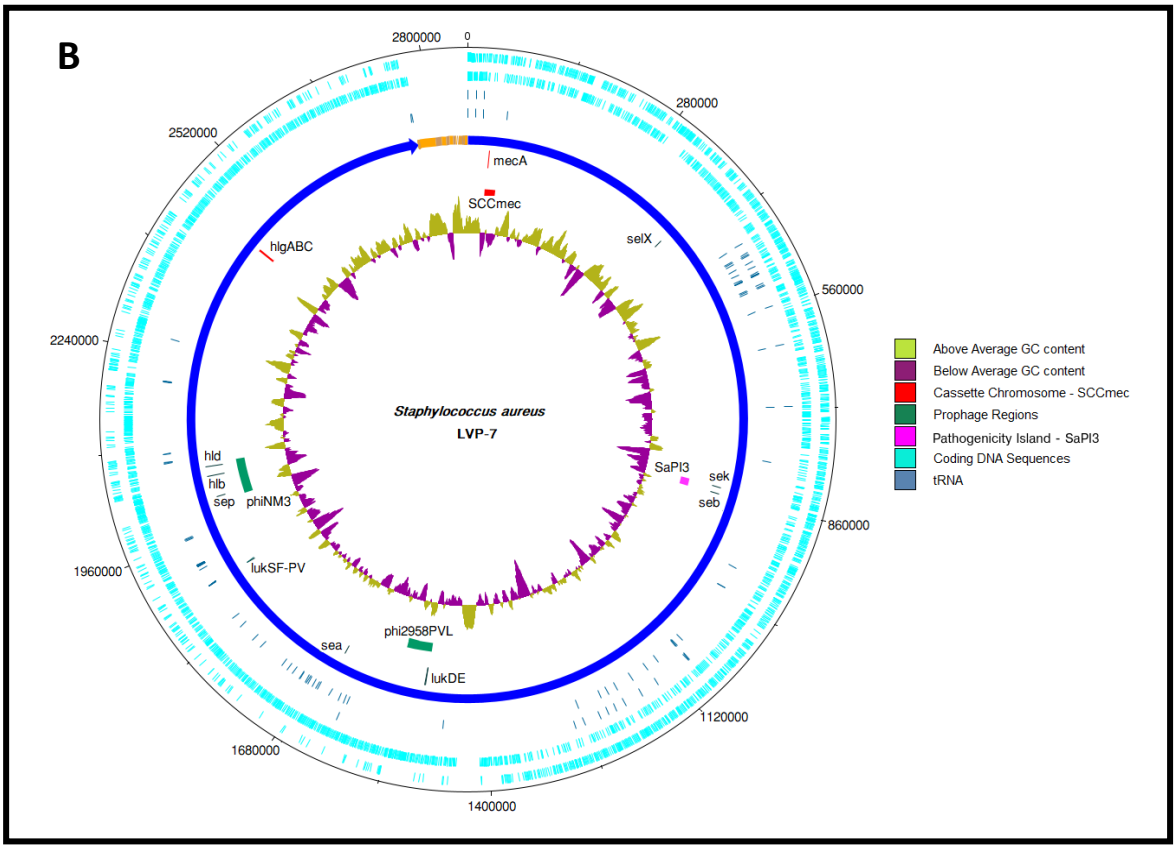


Figure 3.2: Major genetic elements present in CA-MRSA strains

A and B represent the genomic maps of ST8 strain USA300 (FPR3757) and ST88 strain LVP-7. A. Track 1- the GC content, track 2- antibiotic resistance genes in orange and virulence factors in green are shown. Track 3- location of tRNAs. Track 4- select mobile genetic elements, with chromosomal cassettes in red, various pathogenicity islands in shades of blue through violet and prophages in black. The outer two tracks (5 and 6) have the coding sequences in blue. B. Track 1- the GC content. Track 2- mobile genetic elements, including the staphylococcal cassette chromosome (SCCmec) element, the pathogenicity island SaPI3, and the Panton-Valentine leukocidin (PVL) and phiNM3 prophages. Track 3- select virulence genes. Track 4- the major (blue) and minor (orange) contigs, tracks 5 and 6- the location of the tRNAs. The outer tracks (7 and 8)- coding sequences. Figures adapted from (Nadig S et al., 2021) and (Turner NA et al., 2019).

In an effort to understand the transcriptome of ST88 LVP-7 strain, we performed RNA-sequencing analysis of LVP-7 grown at the exponential (logarithmic) phase. The data was obtained using the high-throughput RNA-sequencing technology commercially available in the Illumina platform. We subsequently compared this data

with that of the published transcriptome data set of USA300 (Carroll RK et al., 2016). Interestingly, earlier reports have demonstrated that USA300 has the propensity to form biofilm and this characteristic feature might have also played a possible role in its highly epidemic behaviour (Vanhommerig E, et al., 2014). Similarly, in our study, we have observed from a quantitative biofilm assay that even LVP-7 has the capacity to form biofilms (data shown in chapter 4). Hence, both USA300 and LVP-7 share some of the common genetic features and it was intriguing to understand the differences in transcriptome of both the strains. In this aspect, a comparative analysis of Gene Ontology enriched processes and network analysis of GO terms in both LVP-7 and USA300 was performed. Homology mapping of some of the protein coding genes in LVP-7 led to the identification of genes involved in various biological processes (BP) including pathogenesis.

3.2 Materials and Methods:

3.2.1 Growth condition and total RNA extraction

A single colony of *S. aureus* LVP-7 was grown in tryptic soya broth (TSB) and incubated at 37°C overnight. The next day, 1% inoculum was reinoculated onto fresh 10 ml of TSB and grown till to 0.5 O.D at 600 nm (approximately for 5 hrs) was reached. For RNA-seq analysis, total RNA was extracted from the exponential growth phase culture using the RNeasy mini kit (Qiagen, Inc). Initially, cells were lysed in buffer A (30 mM Tris pH 8.0, 1mM EDTA) containing 100 µg of lysozyme, 100 µg of Lysostaphin (Sigma-Aldrich, Inc) and 200 µg of Proteinase K (Qiagen) for 10 mins at room temperature. RNA was then extracted as per the manufacturer's instructions. Finally the RNA was eluted in 40 µl of RNase-free water and concentration was measured using the Qubit 2.0 fluorimeter (Thermo Fisher Scientific).

3.2.2 cDNA Library preparation and RNA Sequencing

Sequencing libraries were prepared from 200 ng of RNA using NEBNext® Ultra™ II RNA Library Prep Kit for Illumina® (New England Biolabs) as

per the manufacturer's instructions. The ribosomal RNA depletion was done using NEBNext® rRNA Depletion Kit (Bacteria, New England Biolabs) before the library preparation. The final Library was quantitated using the Qubit 2.0 fluorometer (Thermo Fisher Scientific) and its profile was analysed by Agilent Tape station 2200. Sequencing was performed using v3 chemistry in an Illumina HiSeq 2500 instrument (2 x 100-bp paired-end format). Image files captured by the Illumina sequencer were demultiplexed and converted to fastq format using bcl2fastq (Andrews S et al., 2010). The resulting paired-end fastq files contained 102,175,322 reads with a sequence length of 100 nucleotides and were analysed. The pipeline used for the transcriptome analysis is described below (Figure 3.3).

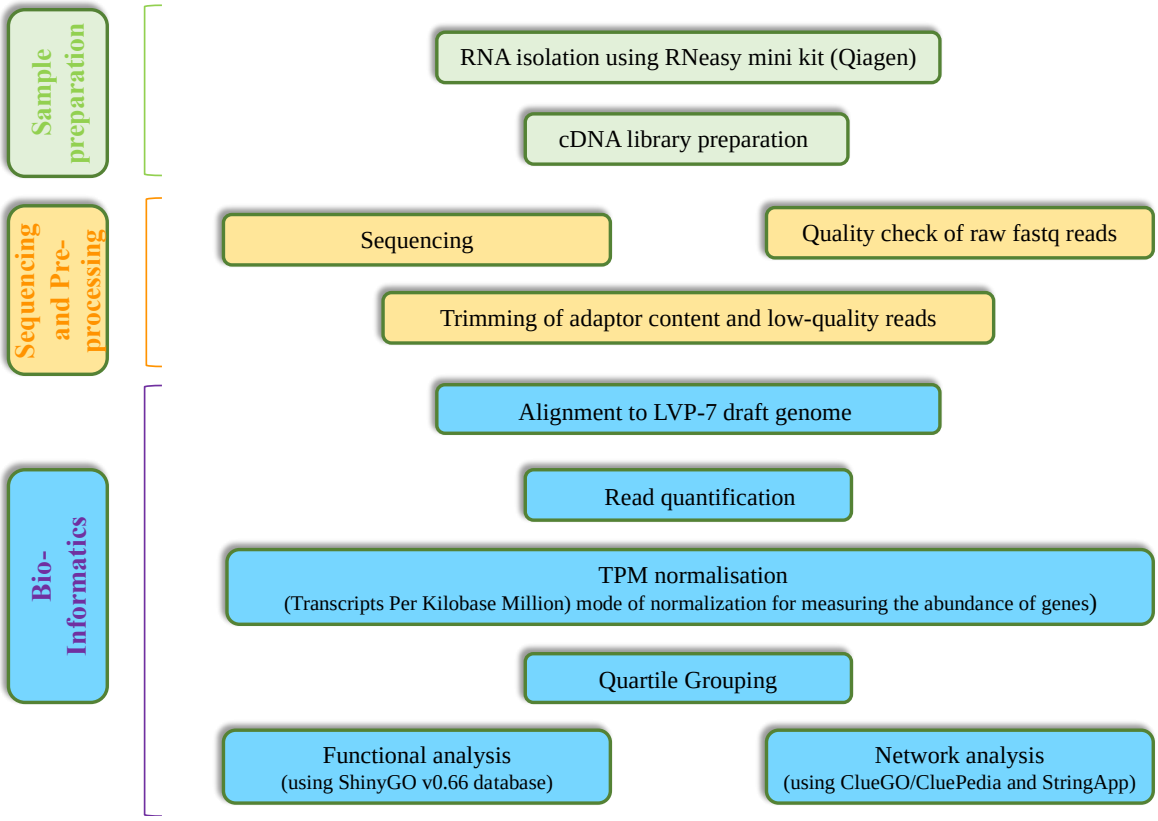


Figure 3.3: Chart representation of the pipeline used for the transcriptome analysis of LVP-7

Three major steps involved in RNA-sequencing. Detailed internal processes of each step are represented in the chart.

3.2.3 Pre-processing and Quality Check of RNA-seq data

The pre-processing operations included a quality check of the fastq file using FastQC (Andrews S et al., 2010), adapter trimming using TrimGalore (Bolger AM et al., 2014), and quality thresholding (removing reads with a Phred Score < 28). This resulted in filtering of 2.1% of the sequenced reads.

3.2.4 Alignment of RNA-seq reads to the LVP-7 draft genome

Alignment of the trimmed fastq files to the LVP-7 draft genome (GenBank Accession ID: JADRJK000000000, Nadig S et al., 2021) was performed using STAR version 2.6.0. (Dobin A et al., 2013). The alignment rate was found to be 97.98%, corresponding to 98,689,719 uniquely mapped reads. The resulting SAM files were converted to BAM files (compressed binary version of SAM files) using SAMtools (Li H et al., 2009).

3.2.5 Quantification of Aligned reads

To determine the number of raw read counts that mapped to genes annotated for the LVP-7 strain, the annotation.gff3 file associated with the assembly (GenBank Accession ID-GCA_015887435.1) and the BAM alignment files were analysed using htseq-count version 0.11.2 (Anders S et al., 2015). The resulting read count table was further processed as described below.

3.2.6 TPM normalisation of Aligned reads

TPM (Transcripts Per Kilobase Million) is a popular mode of normalisation for RNA-Sequencing data and is particularly relevant to comparing intra-sample expressions. TPM was calculated as follows using a custom R-based script based on the formula below (Zhao et al., 2020):

$$\mathrm{TPM} = 10^6 * \frac{\text{reads mapped to transcript} / \text{transcript length}}{\mathrm{Sum}(\text{reads mapped to transcript} / \text{transcript length})}$$

The steps in the script included:

1. For each transcript in the gene model, the number (raw count) of reads estimated by htseq-count was divided by the transcript's length, giving a normalised transcript-level expression.
2. The sum of ALL normalised transcript expression values was divided by 1,000,000 to create a scaling factor.
3. Each transcript's normalised expression was divided by the scaling factor, which resulted in the TPM value (Appendix VI).

3.2.7 Quartile Grouping of expressed transcripts

The genes were ranked based on highest to lowest TPM values and divided into quartiles (quarters of nearly equal size), with Q1 containing the most highly expressed genes (top 25%) and Q4 containing genes with the lowest expression (bottom 25%). To visualise the distribution of TPM within each quartile, a box-and-whisker plot was graphed using GraphPad Prism. TPM values were extracted for the genes of interest and the quartile to which they belong was estimated (Table 3.1).

3.2.8 RNA isolation, cDNA synthesis, and qRT- PCR analysis

Total RNA samples were extracted from the logarithmic growth phase of LVP-7 using RNeasy mini kit (Qiagen). Initially, cells were lysed in 30mM Tris pH 8.0, 1mM EDTA containing 100 µg of lysozyme (Sigma-Aldrich), 200 µg of Proteinase K (Qiagen), 100 µg of Lysostaphin (Sigma-Aldrich) and further steps were followed as suggested by the manufacturer and finally eluted in 40 µl of RNase-free water. cDNA was prepared using QuantiTect® Reverse Transcription kit (Qiagen) according to the manufacturer's instructions. Real Time PCR was carried out using cDNA (20 ng) using SYBR green super mix (Bio-Rad) in Bio-Rad iQ5TM thermocycler. 16sRNA gene was amplified as an endogenous control. The relative levels of expression of some the selected genes from RNA-seq data were determined by using the $2^{-\Delta CT}$ method (Livak KJ et al., 2001). The primers used to amplify the target genes are listed in Appendix Table 6.1.

3.2.9 Gene Ontology (GO) analysis for identification of various enriched functional pathways

For the genes belonging to each quartile (Q1, Q2, Q3 and Q4) in LVP-7, we performed GO enrichment analysis for the Biological Process (BP), Cellular Component (CC) and Molecular Function (MF) using ShinyGO v0.66 (Xijin Ge S et al., 2020), with the default false discovery rate (FDR) threshold (p- value ≤ 0.05). The top five significantly enriched GO terms for each category and for each quartile were extracted. All the three categories based on significance (FDR value) for LVP-7 in each quartile were represented in the form of bar graphs (Figure 3.4). Further, to compare GO enrichment analysis between the LVP-7 and USA-300 transcriptome, a published data set of USA300 was chosen which had similar growth parameters for the comparative study (Carroll RK et al., 2016, GEO accession number GSE74936). For this comparative study, the transcript expression levels in USA300 were reported as RPKM (reads per kilobase transcripts per million reads mapped) by Carroll group. These were normalised to TPM using the formula mentioned in section 2.6, following which GO analysis was performed. After performing the GO enrichment analysis of both LVP-7 and USA300, the percentage enrichment of top 500 GO terms based on FDR value was calculated for all the four quartiles of LVP-7 and USA300.

3.2.10 Network analysis using homology mapping

In addition, to understand the relationship between different GO terms for gene functional category biological process (BP), the STRING 11.0b database and the ClueGO/CluePedia- Cytoscape plugin tools were used (Szklarczyk D, et al., 2017, Shannon P et al., 2003, Bindea G et al., 2009 and Bindea G et al., 2013). The STRING database was used to identify the enriched GO terms under the functional category BP for LVP-7 strain. The minimal confidence score used for the STRING database was 0.4. Further to understand the interaction between the obtained GO terms ClueGO/CluePedia pre-selected functional analysis was performed. Staphylococcus aureus subsp. aureus USA300-TCH1516 (451516, available in ClueGO tool) and GO_BiologicalProcess-GOA_07.03.2016_12h13 were the organism and the ontology respectively, used in ClueGO v2.5.8 for analysis. A functionally grouped network was generated with terms as nodes linked based on their kappa score level (≥ 0.1). A two-

side hypergeometric test (a statistical test) was used for the analysis. The term with the smallest p-value was referred to as the most significant term in a cluster and was highlighted. The homology mapping method was used to generate the Gene Ontology terms and protein-protein interaction network as described by Subramanian and co-workers (Subramanian D et al., 2019) in order to obtain the interaction information on protein-coding genes between *S. aureus* strains LVP-7 and USA300-TCH1516. Further, the protein-protein interaction network of the genes involved in the pathways involved in pathogenesis and stress response in LVP-7 were generated using StringApp (Doncheva NT et al., 2018) with a minimal confidence score of 0.4.

3.3 Results

The RNA-sequencing data of *S. aureus* LVP-7 grown till the logarithmic growth phase enabled the measurement of transcript abundance for 2,720 annotated genes (6.5 Appendix VI).

3.3.1 Estimation of transcript abundance of important virulence genes in LVP-7

As discussed earlier, TPM values were used to divide the genes into different quartiles in LVP-7. The most enriched genes (top 25%) were grouped into quartile 1. The majority of the genes involved in *agr* quorum sensing pathway in *S. aureus* were identified in quartile 1. Genes responsible for biofilm formation and regulating the stress response were also abundant in quartile 1 (Figure 3.4 and Table 3.1). Staphylococcal enterotoxins like *seb* and *sek* were found in quartile 2.

3.3.1.1 Genes involved in pathogenesis and stress response

LVP-7 is a clinical sample isolated from an eye infection. In this context, the genes related to pathogenesis and stress response were of particular interest. A comprehensive list of the virulence genes in LVP-7 alongside PROKKA annotation, transcript abundance and the quartile each gene maps to, is compiled in Appendix Table 1. As biofilms help in the pathogenesis of *S. aureus*, it was interesting to understand the abundance of the transcripts corresponding to this mechanism in LVP-7. The genes

responsible for the adhesion or attachment or colonization like *spa* (TPM value-37414.065) and the *agr* QS genes like *agrC*, *agrD* and *agrB* were present in quartile 1. The MSCRAMMs (microbial surface components recognizing adhesive matrix molecules) family of adhesion protein genes *clfA* and *clfB* could be located in quartile 1 and quartile 2 respectively. Similarly, the *S. aureus* superoxide dismutases like *sodA* (TPM value-3116.94) and *sodM* (TPM values-214.846) were enriched in quartile 1 and 2 respectively. *sodA* is known to express in the early phases of biofilm formation (Ballal A et al., 2009). Similarly, *katA* which is known to confer resistance to H_2O_2 stress was amongst the top 25% highly expressed genes(Cosgrove et al., 2007). MgrA, a global regulator protein known to act as a DNA-binding oxidant sensor in *S. aureus* (Chen PR et al., 2006) and was noted to lie in quartile 1 (Figure 3.4 and Table 3.1). Similarly, the enterotoxin *seb* was found to be enriched in quartile 2 and the role of *seb* toxin in superantigenicity causing systemic infection in a mouse model has been demonstrated in a CA-MRSA, ST59 clonal lineage (Bae JS et al., 2020). Interestingly, the abundant expression of phenol soluble modulin *psmA3* (*psmα3*) transcript is observed in quartile 1 and it is noted that the virulence potential of MRSA is increased when there is high expression of PSMαs (Wang R et al., 2007).

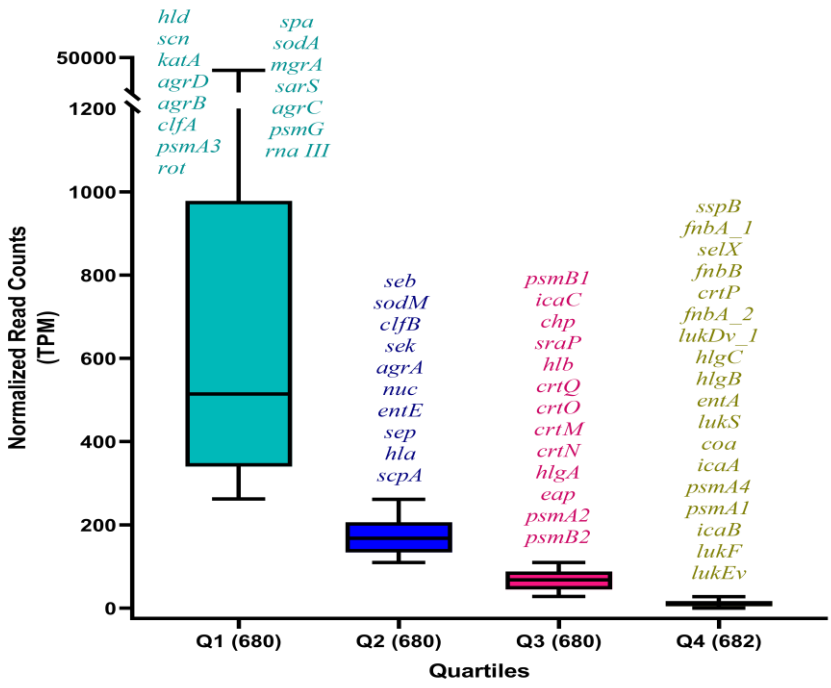


Figure 3.4: RNA-sequencing analysis was used to estimate the expression of transcripts encoding key virulence factors in the LVP-7 strain

The distribution of TPM values for ~680 genes in each quartile was visualised using a box-and-whisker plot. 56 genes (listed in Appendix VI) encoding virulence factors or factors linked to *S. aureus* pathogenesis were mapped. These genes are compiled in the corresponding quartile.

3.3.1.2 Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis

To validate findings from transcriptome analysis, the expression level of a few genes were quantified by qRT-PCR. This set included genes in the *agr* operon (*agrB*, *agrD*, *agrC* and *agrA*), the highly expressed staphylococcal protein A (*spa*) and some genes that have been noted to enhance the virulence in *S. aureus* such as like *sodA*, *seb* and *psmA3* were checked for the expression in qRT-PCR. In this analysis, 16sRNA served as an internal control. From this analysis, the expression level of *spa* was identified to be higher compared to other genes- a finding that is consistent with the RNA-sequencing data (Figure 3.5).

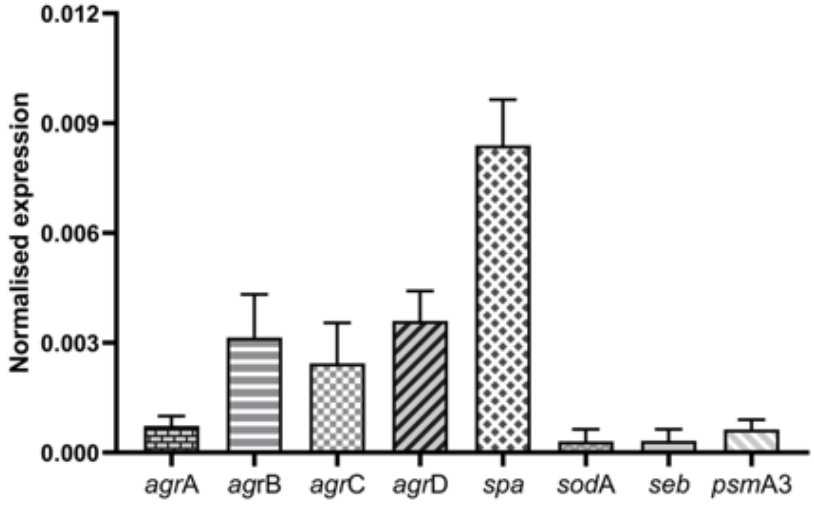


Figure 3.5: The expression levels of representative genes in logarithmic growth phase of LVP-7 strain.

The data shown represent an average of six measurements (two biological replicates).

Table 3.1: Summary of the expression of 56 genes encoding virulence factors or factors linked to *S. aureus* pathogenesis

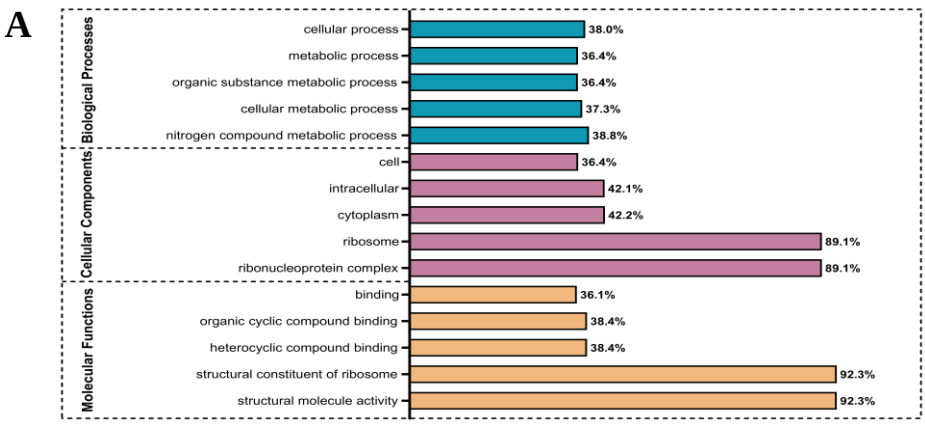
Gene Name	General Description	PROKKA Annotation	PROKKA ID	Start	End	Length	Raw Counts	TPM	Quartile
spa	staphylococcal protein A		SALVP7_00066	82521	83849	1328	6205476	37414.065	I
sodA	superoxide dismutase		SALVP7_01491	1606402	1607001	599	233120	3116.094	I
mgrA	L-glyceraldehyde 3-phosphate reductase		SALVP7_00614	686555	686998	443	65945	1191.890	I
sarS	HTH-type transcriptional regulator SarS		SALVP7_00067	84270	85022	752	89323	951.050	I
rot	virulence factor regulator protein		SALVP7_01685	1832283	1832684	401	45638	911.255	I
agrC	GHKL domain-containing protein		SALVP7_01954	2063958	2065250	1292	61039	682.578	I
psmG	Putative tryptophan 2,3-dioxygenase		SALVP7_01951	2062856	2062990	134	10705	639.646	I
rnaIII	Small regulatory RNAIII	psmG	SALVP7_01951	2062856	2062990	134	10705	639.646	I
hld	delta-hemolysin	psmG	SALVP7_01951	2062856	2062990	134	10705	639.646	I
scn	complement inhibitor SCIN-A	Scn_3	SALVP7_01878	2004586	2004936	350	25609	585.845	I
katA	catalase KatA		SALVP7_01220	1307703	1309220	1517	80770	426.307	I
agrD	cyclic lactone autoinducer peptide		SALVP7_01953	2063793	2063933	140	63967	396.416	I
agrB	Accessory gene regulator protein B		SALVP7_01952	2063172	2063789	617	6179	353.385	I
clfA	MSCRAMM family adhesin clumping factor ClfA		SALVP7_00732	809152	812067	2915	122231	335.738	I
psmA3	20S proteasome subunit alpha-3		SALVP7_01804	1925800	1926945	1145	41162	287.838	I
seb	Staphyloccal enterotoxin		SALVP7_00784	845241	846041	800	23118	231.376	II
sodM	superoxide dismutase		SALVP7_00087	107901	108500	599	16073	214.846	II
clfB	MSCRAMM family adhesin clumping factor ClfB		SALVP7_02543	2680134	2680589	455	66781	210.761	II
sek	Staphylococcal enterotoxin		SALVP7_00762	833293	834021	728	15511	170.595	II
agrA	Accessory gene regulator protein A		SALVP7_01955	2065269	2065985	716	28901	155.933	II
nuc	EDTA-resistant nuclease		SALVP7_00736	816362	817048	686	12725	148.522	II
entE	(2,3-dihydroxybenzoyl)adenylate synthase EntE		SALVP7_01884	2009442	2010224	782	12801	131.067	II
sep	enterotoxin P	entE	SALVP7_01884	2009442	2010224	782	12801	131.067	II
hla	Alpha-hemolysin	hly	SALVP7_01040	1112460	1113419	959	15237	127.215	II
scpA	methylmalonyl-CoA mutase		SALVP7_01433	1553070	1553825	755	10761	114.120	II
psmB1	proteasome subunit beta type 1		SALVP7_01053	1124545	1124679	134	1475	88.134	III
icaC	)lysaccharide intercellular adhesin biosynthesis/export protein IcaC		SALVP7_02584	2734906	2735958	1052	6921	87.682	III
chp	chemotaxis-inhibiting protein CHIPS		SALVP7_01879	2005619	2006068	449	4819	85.935	III
sraP	LPXTG cell wall anchor domain-containing protein		SALVP7_02569	2718564	2724113	5549	3551	66.430	III
hlb	sphingomyelin phosphodiesterase	hlb_1	SALVP7_01877	2003211	2003411	200	1607	64.334	III
crtQ	zeta-carotene desaturase		SALVP7_02478	2610990	2612117	1127	3854	62.089	III
crtO	hypothetical protein	crtQ	SALVP7_02478	2610990	2612117	1127	3854	62.089	III
crtM	presqualene diphosphate synthase HpnD		SALVP7_02477	2610090	2610953	863	10856	58.219	III
crtN	dehydrosqualene desaturase		SALVP7_02476	2608570	2610078	1508	7810	55.486	III
hlgA	Gamma-hemolysin component A		SALVP7_02339	2450691	2451656	965	4130	33.846	III
eap	uncharacterized protein	map_1	SALVP7_00865	927906	928340	434	1765	32.562	III
psmA2	proteasome subunit alpha type 2		SALVP7_01048	1119856	1121412	1556	6288	32.356	III
psmB2	proteasome subunit beta type 2		SALVP7_01054	1124736	1124870	134	537	32.087	III
sspB	ClpXP protease specificity-enhancing factor		SALVP7_00930	994320	995501	1181	4010	27.186	IV
fnbA_1	Fibronectin-binding protein A	FnbA_1	SALVP7_02417	2538541	2539152	611	4533	26.727	IV
selX	toxin		SALVP7_00324	378801	379412	611	1787	23.418	IV
fnbB	fibronectin-binding protein FnbB	gtaB	SALVP7_02416	2537495	2538361	866	1420	17.737	IV
crtP	hypothetical protein		SALVP7_02479	2612123	2613616	1493	1537	16.045	IV
fnbA_2	Fibronectin-binding protein A	FnbA_2	SALVP7_02418	2542281	2542922	641	2958	15.251	IV
lukDv_1	Leucotoxin LukDv		SALVP7_01365	1500505	1501482	977	1739	14.252	IV
hlgC	leukocidin s subunit	hlgC_2	SALVP7_02340	2452223	2453170	947	800	13.716	IV
hlgB	leukocidin f subunit		SALVP7_02341	2453172	2454149	977	961	11.119	IV
entA	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase EntA		SALVP7_01538	1650594	1651403	809	1001	9.907	IV
lukS	LukS protein	HlgC_1	SALVP7_01366	1501484	1502422	938	1103	9.415	IV
coa	staphylocoagulase		SALVP7_00175	220096	222168	2072	2154	8.324	IV
icaA	N-acetylglucosaminyltransferase		SALVP7_02581	2732543	2733781	1238	891	8.181	IV
psmA4	20S proteasome subunit alpha-4	wbpI	SALVP7_00108	132954	134078	1124	1057	7.530	IV
psmA1	Proteasome subunit alpha type-1	ssl4_1	SALVP7_00354	407094	407789	695	610	7.028	IV
icaB	intercellular adhesin biosynthesis polysaccharide N-deacetylase		SALVP7_02583	2734047	2734919	872	1595	6.245	IV
lukF	LukF protein	lukDv_2	SALVP7_01739	1882912	1883895	983	431	3.511	IV
lukEv	leukotoxin LukE		SALVP7_01740	1883897	1884817	920	349	3.037	IV

White colour represents quartile I and quartile III, pink colour represents quartile II and purple colour represents quartile IV.

3.3.2 Gene Ontology (GO) Analysis

3.3.2.1 Gene Ontology (GO) functional enrichment analysis of the LVP-7 transcriptome

To gain an insight into the biological roles of the genes expressed in the logarithmic growth phase of *S. aureus* LVP-7, Gene Ontology enrichment analysis was performed with a false discovery rate (FDR) threshold (p value ≤ 0.05) and the top 500 GO terms were selected for further analysis. As shown in Figure 3.6A and Figure 3.6C, the top three enriched terms in the BP category were ‘cellular process’, ‘metabolic process’ and ‘organic substance metabolic process’ in quartiles 1 and 3. However, in quartile 2, the three GO terms in BP were ‘cellular process’, ‘cellular metabolic process’ and ‘metabolic process’. Quartile 4 had ‘organic acid metabolic process’, ‘carboxylic acid metabolic process’ and ‘oxoacid metabolic process’ as the top enriched process. This data is compiled in Figure 3.6B and Figure 3.6D. In CC category, the top three GO terms were ‘cell’, ‘intracellular’ and ‘cytoplasm’ for all the 3 quartiles except quartile 4 (Figure 3.6). In quartile 4, the top three GO terms were ‘extracellular region’, ‘cell wall’ and ‘external encapsulating structure’. In the MF category, ‘binding’, ‘organic cyclic compound binding’, and ‘heterocyclic compound binding’ were the top three GO terms in quartile 1 (Figure 3.6A). For quartile 2 and 3 it was ‘binding’, ‘ion binding’, ‘ and ‘catalytic activity’ (Figure 3.6B and Figure 3.6C). In quartile 4, ‘catalytic activity’, ‘co-factor binding’ and ‘lyase activity’ were the top three GO terms (Figure 3.6D).



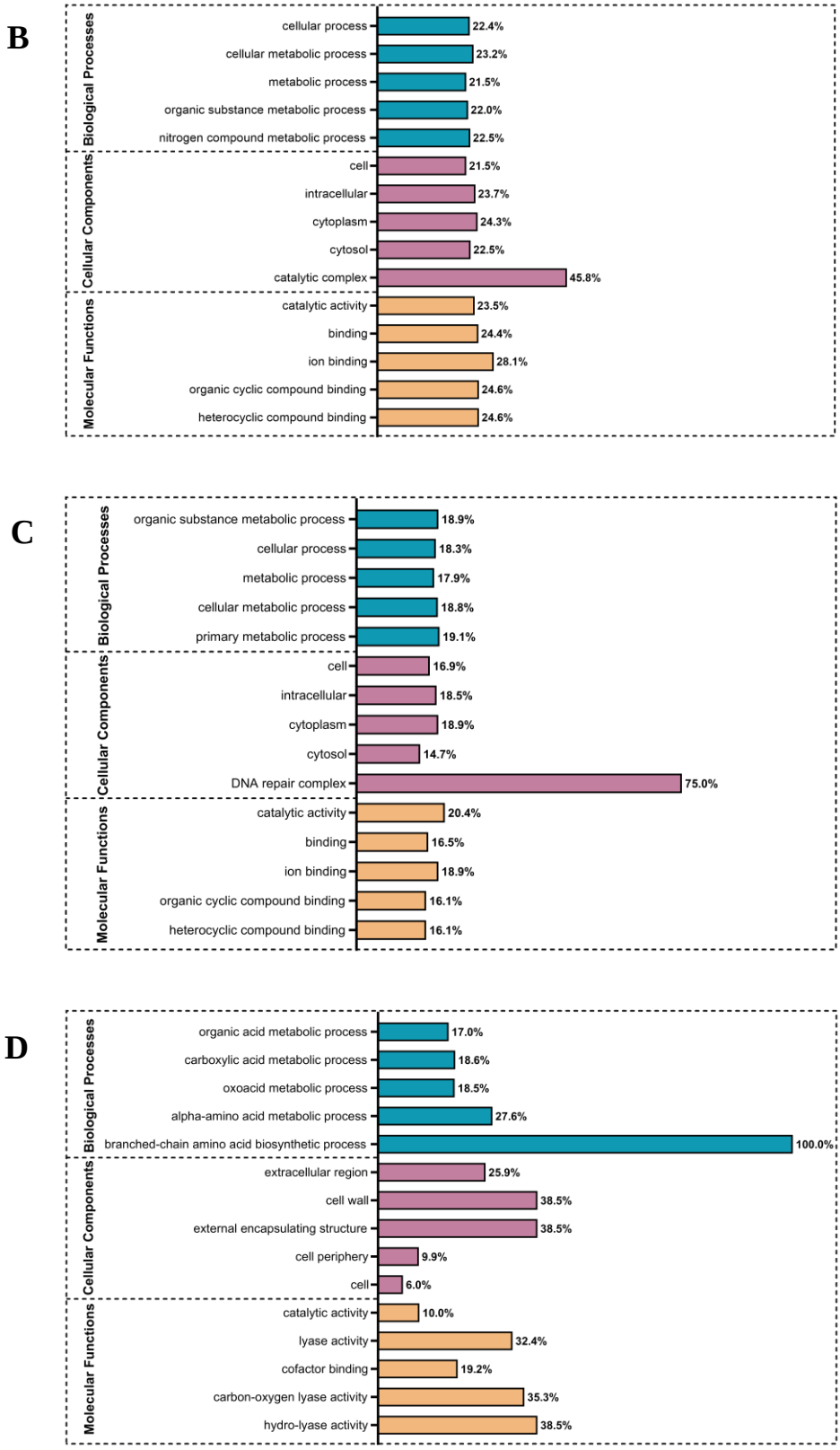


Figure 3.6: Bar graph representation of Gene Ontology (GO) functional enrichment analysis for LVP-7

GO enrichment analysis performed using ShinyGO v0.66. Top five GO terms for each category is represented in the above figure. A represents quartile 1, B- quartile 2, C-

quartile 3 and D- quartile 4. Cyan colour represents the Biological Process (BP), Pink colour represents Cellular Component (CC) and yellow colour represents Molecular Function (MF).

3.3.2.2 Comparative analysis between LVP-7 and USA300 based on GO functional enrichment analysis

The total number of genes identified in LVP-7 is 2,720 and in USA300 it is 2,936 (Carroll RK et al., 2016, GEO accession number GSE74936). After performing the quartile grouping and GO functional enrichment analysis for the USA300 transcriptome, percentage enrichment of GO various terms were calculated for each strain. A heatmap showing the comparative analysis of GO terms between the two CA-MRSA strains is shown in Figure 3.7. As discussed earlier, the important virulence factors or genes present in LVP-7 in various pathways like biofilm formation, responsible for the pathogenesis of *S. aureus* are highly abundant in quartile 1. Correspondingly, the percentage enrichment for the GO term 'pathogenesis' in LVP-7 were distributed into quartiles 1, 3 and 4 (36.1%, 14.8% and 11.5% respectively). However, in USA300 it was only 24.6% enrichment in quartile 1. The GO term 'response to oxidative stress', and its broader parent term for 'response to stress', had a total percentage enrichment or abundance of 77.8% and 82.9% (with 26.8% in quartile 1 alone) in case of LVP-7. In USA300, both the parental term 'response to stress' and 'response to oxidative stress' were enriched only in quartile 1 with about 22% and 44.4% enrichment, respectively. Similar differences were observed in 'homeostatic process' in both LVP-7 and USA300 (Figure 3.7). In the CC category, the GO term 'ribonucleoprotein complex' was significantly enriched in both LVP-7 and USA300 in quartile 1.

Process	GO term	LVP-7				USA300			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Biological Processes	electron transport chain	60.0%	0.0%	0.0%	0.0%	50.0%	0.0%	0.0%	0.0%
	translation	76.5%	14.1%	0.0%	0.0%	65.9%	0.0%	0.0%	0.0%
	ion transport	29.2%	0.0%	0.0%	12.5%	22.9%	0.0%	0.0%	10.4%
	metabolic process	36.4%	21.5%	17.9%	7.7%	29.2%	18.5%	0.0%	8.5%
	biosynthetic process	42.6%	21.8%	17.8%	6.4%	34.7%	22.7%	0.0%	8.6%
	gene expression	60.1%	19.0%	10.7%	0.0%	45.2%	18.5%	0.0%	0.0%
	RNA processing	27.0%	32.4%	29.7%	0.0%	0.0%	21.6%	0.0%	0.0%
	response to oxidative stress	77.8%	0.0%	0.0%	0.0%	44.4%	0.0%	0.0%	0.0%
	response to stress	26.8%	29.3%	26.8%	0.0%	22.0%	0.0%	0.0%	0.0%
	pathogenesis	36.1%	0.0%	14.8%	11.5%	24.6%	0.0%	0.0%	0.0%
	establishment of localization	23.0%	0.0%	0.0%	12.2%	18.9%	0.0%	0.0%	0.0%
	homeostatic process	77.8%	0.0%	0.0%	0.0%	44.4%	0.0%	0.0%	0.0%
	cell division	52.4%	38.1%	0.0%	0.0%	33.3%	38.1%	0.0%	0.0%
	cell communication	0.0%	28.6%	0.0%	0.0%	28.6%	0.0%	0.0%	0.0%
	transmembrane transport	22.8%	15.8%	0.0%	0.0%	22.8%	0.0%	0.0%	0.0%
	pigment biosynthetic process	0.0%	31.3%	25.0%	0.0%	0.0%	25.0%	0.0%	0.0%
Cellular Components	catalytic complex	25.0%	45.8%	0.0%	0.0%	37.5%	25.0%	0.0%	0.0%
	intracellular	42.1%	23.7%	18.5%	0.0%	35.5%	21.8%	11.3%	7.4%
	ribonucleoprotein complex	89.1%	0.0%	0.0%	0.0%	76.4%	0.0%	0.0%	0.0%
	plasma membrane	25.2%	15.1%	11.5%	0.0%	17.3%	0.0%	0.0%	0.0%
	cell periphery	24.3%	14.5%	11.8%	9.9%	16.4%	0.0%	0.0%	7.2%
Molecular Functions	drug binding	25.2%	28.6%	18.4%	0.0%	27.2%	21.1%	13.6%	8.8%
	ion binding	26.7%	28.1%	18.9%	7.8%	24.8%	19.6%	13.3%	10.4%
	ATP binding	25.0%	29.4%	19.9%	0.0%	27.2%	22.8%	14.7%	8.1%
	nucleic acid binding	52.3%	23.2%	13.9%	0.0%	39.7%	17.9%	11.3%	0.0%
	tRNA binding	66.7%	0.0%	0.0%	0.0%	60.0%	0.0%	0.0%	0.0%
	oxidoreductase activity	34.3%	15.7%	18.6%	0.0%	20.0%	0.0%	0.0%	0.0%
	endonuclease activity	0.0%	0.0%	41.2%	0.0%	0.0%	0.0%	0.0%	23.5%
	catalytic activity	24.4%	23.5%	20.4%	10.0%	20.6%	18.6%	12.2%	10.4%
	lyase activity	0.0%	21.6%	18.9%	32.4%	0.0%	24.3%	0.0%	32.4%
	exonuclease activity	0.0%	54.5%	0.0%	0.0%	0.0%	36.4%	0.0%	0.0%
	kinase activity	30.2%	18.6%	16.3%	0.0%	25.6%	18.6%	0.0%	0.0%
	transferase activity	25.2%	27.9%	21.1%	0.0%	17.7%	26.5%	8.8%	9.5%

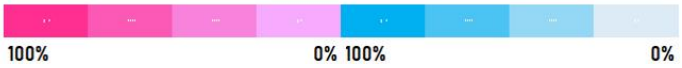


Figure 3.7: Gene Ontology enrichment comparison between LVP-7 and USA300

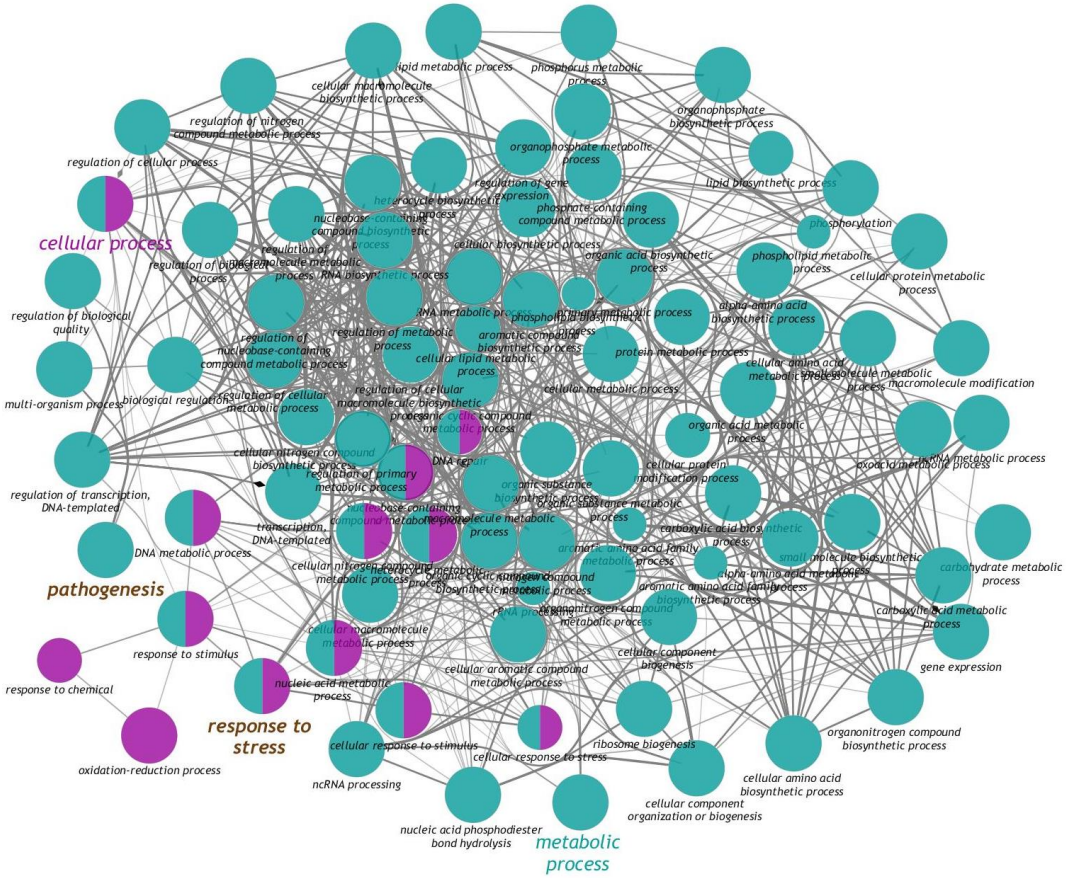
The percentage of enrichment of various processes like biological processes, molecular function and cellular components between two LVP-7 and USA300 is represented in the form of a heatmap. Pink and blue colours represent strains LVP-7 and USA300 respectively. Q1-Q4- Quartile grouping according to the TPM values.

3.3.2.3 Interaction of various GO terms and protein-protein interaction network analysis using homology mapping

A network analysis was performed using the Cytoscape application-ClueGO/ CluePedia and STRING database to identify the interaction between various GO terms of BP category (Szklarczyk D et al., 2017, Shannon P et al., 2003, Bindea G et al., 2009

and Bindea G et al., 2013). Since the GO annotation files for LVP-7 were not available, the GO annotation files of USA300-TCH1516 (section 3.2.9), another well studied CA-MRSA, were used for the downstream analysis. After merging the common terms, two clusters were obtained and the more significant term in each cluster is highlighted- 'cellular process' (cluster 1) and 'metabolic process' (cluster 2). The GO term 'pathogenesis' belongs to cluster 2 whereas the GO term 'response to stress' belongs to both clusters 1 and 2 (Figure 3.8A). Further, to understand the protein-protein interaction (PPI) between the genes involved in the GO terms 'pathogenesis' and 'response to stress', a PPI was generated using StringApp. Except for two proteins Sle1 under the GO term 'pathogenesis' and ScdA under the GO term 'response to stress', all other proteins interact with each other. Besides the proteins such as ClfA, Spa and Eno involved in 'pathogenesis' also interact with the proteins involved in 'response to stress' category (Figure 3.8B).

A.



B.

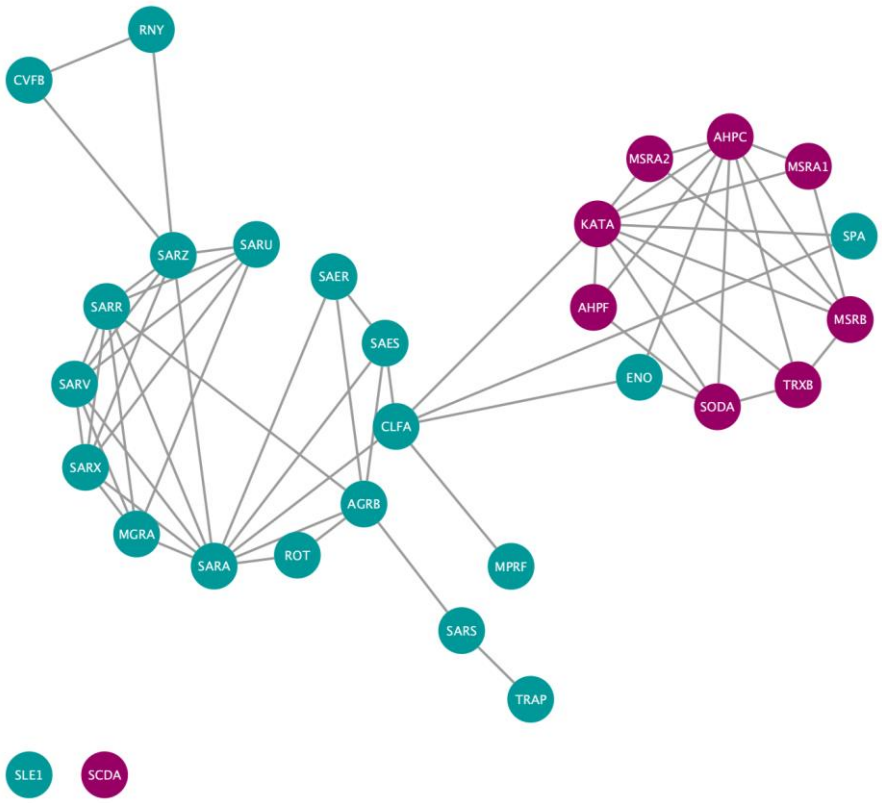


Figure 3.8: Network analysis for GO enriched biological process

A. Representation of the gene distribution with respect to biological process (BP). GO terms are represented as nodes, the node size shows the term enrichment whereas the edges represent term-term interactions. Functionally grouped network with terms as nodes linked based on the basis of their kappa score levels (≥0.1), where the label of the most significant term per group is highlighted. Functionally related groups partially overlap. This network was generated using the pre-selected functional analysis in Cytoscape ClueGO/CluePedia application. B. Genes involved in 'pathogenesis' (cyan color) and 'response to stress' (dark red color) are represented. This network is based on known and predicted interactions from the StringApp, with minimal confidence score of 0.4.

3.4 Discussion

Global transcriptome analysis can provide valuable insights into bacterial physiology and adaptation to environmental conditions. RNA-sequencing-based transcriptomic analysis of LVP-7 (a clinical isolate from an orbital abscess) growing in the logarithmic growth phase enabled the measurement of transcript abundance for 2,720 annotated genes. Liu Q and co-workers investigated the differentially expressed genes (DEGs) in *S. epidermidis* isolated from the conjunctival sac of healthy individuals versus postoperative endophthalmitis (an eye infection) using Illumina high-throughput RNA-sequencing technology (Liu Q et al., 2020) . In this study the GO enrichment analysis revealed that several terms in BP like 'homeo-static process', 'cellular homeostasis', and 'cell redox homeostasis' were significantly enriched. Similarly, seven differentially expressed genes involved in SaeRS two-component system were enriched. Genes coding for histidine kinase and for *agr* B were also seen significantly enriched in postoperative endophthalmitis strains. Hence, using GO enrichment analysis authors were able to elucidate genes involved in the pathogenesis and identified potential therapeutic targets of *S. epidermidis* related to endophthalmitis. Not surprisingly, the RNA-seq data of LVP-7 showed enhanced gene expression of a family of adhesion proteins involved in colonization, (spa, clfA and clfB) and also genes from *agr* two-component system (Figure 3.4). Adhesion proteins (MSCRAMMs) are the key factors required for establishment, invasion and persistence within the host. It was revealed that the ability to express higher levels of MSCRAMMs and low levels of leukotoxins might significantly contribute towards the successful invasion and persistence of *S. aureus* in chronic inflammatory conditions (Kanangat. S et al, 2007). The same study showed that *S.aureus* exposed to Intraleukin-1beta (IL-1b) had higher expression of MSCRAMMs genes. As several regulatory systems play a pivotal role in pathogenic bacterial adaptation and survival in hostile environments, the essential two-component system comprising of a membrane bound histidine kinase and a response regulator assist in bacterial adaptation, virulence and antibacterial resistance. In case of *S. aureus*, the *agr* QS mechanism plays an important role in switch between the persistent to virulent phenotype (Novick RP, 2006). Hence LVP-7 strain might also rely on these regulatory mechanisms to cause an eye infection.

The results of enrichment analysis in LVP-7 strain, indicated that the terms 'cellular process', 'cell' and 'binding' as the top most enriched terms in BP, CC and MF category respectively (Figure 3.6). Similarly in homology mapping for protein-protein interaction (PPI), most of the genes involved in 'pathogenesis' and 'response to stress' interacted with each other except Spa, ClfA and Eno (Figure 3.8). Also *S. aureus* is known to use multiple mechanisms to survive the host oxidative burst. In the case of host-pathogen interactions, neutrophils, the frontline of defence in the human host engulf bacteria by phagocytosis and killing through the activity of NADPH oxidase and myeloperoxidase (MPO). The generation of reactive oxygen species (ROS) and reactive nitrogen species (RNS) leads to oxidative burst forming the initial neutrophil extracellular trap (NETs) formation and thus clearing the bacteria (Parker H et al., 2012). In the transcriptome profile of LVP-7, two superoxide dismutases, *sodA* and *sodM* that superoxide anion to H_2O_2 and O_2 were seen to be over expressed. Phenol soluble modulins (PSMs) are recognised as determining factors for biofilm formation in *S. aureus* and their co-operative interaction (psmα1-α4 and psmβ1, β2) results in fast and efficient biofilm formation (Zaman M et al., 2020). Intriguingly, we found psmA3 (psmα3) transcript abundance in quartile 1, Figure 3.4, Table 3.1). Also, enhanced expression of the *katA* gene was observed which helps in conferring resistance to H_2O_2 stress during host-pathogen interactions. Staphylococcal enterotoxins like *seb* and *sek* were found in quartile 2 (Figure 3.6, Ballal A et al., 2009 and Cosgrove et al. 2007). The role of the *seb* toxin in the hypervirulence of ST59 clonal lineage has been demonstrated in mouse infection models (Bae JS et al., 2020). By performing Gene Ontology functional enrichment analysis and comparative study between LVP-7 and USA300, we were able to compare the transcript abundance corresponding to various GO terms. Comparative GO enrichment analysis between LVP-7 and USA300 showed that the percentage abundance of genes involved in 'response to oxidative stress' were 77.8% and 44.4% respectively (Figure 3.7). This observation suggests a dominant role for genes governing the response to oxidative stress in the case of LVP-7. Whether this finding is corroborated with other features specific to eye infection remains to be seen. Investigation of the role of other differentially-regulated molecules is likely to provide new insights into the pathogenesis of *S. aureus* infection.

All TPM files for LVP-7 are given in Appendix VI.
Primer list for qRT-PCR is given in Appendix Table 6.1.

Codes used in transcriptome analysis of LVP-7 is given in Appendix 6.2.1

Chapter 4

In vivo visualization of RNAIII – an effector molecule of the Accessory Gene Regulator (agr) quorum sensing system

4.1 Introduction

Small regulatory RNAs (sRNAs) play an important role in regulating gene expression in bacteria. The role of sRNAs in various physiological responses have been extensively examined in different organisms. These studies led to the identification of the bonafide targets of sRNAs thus decoding their role and involvement in complex regulatory networks (Vogel J et al., 2007 and Waters LS et al., 2009). The Accessory Gene Regulator (*agr*) is a characterized quorum sensing mechanism in *Staphylococcus aureus*. A distinguishing feature of this system is the role played by a regulatory RNA, RNAIII, in the quorum response. All components of the *agr* mechanism, including the effector RNA, RNAIII, are coded by genes in the *agr* operon. The key constituents of the *agr* quorum sensing mechanism are AgrC- a sensor kinase, a response regulator-AgrA, a membrane protease also referred to as the cell density sensing cassette-AgrB and AgrD that encodes a precursor peptide also known as auto-inducing peptide, (AIP). The quorum sensing mechanism senses population density through the signalling peptide molecule (AIP). AIP binding to AgrC leads to the phosphorylation of AgrA. Native AgrA regulates the expression of genes in the *agr* operon, including *agrA*. These genes are located downstream of the P2 promoter element. Phosphorylated AgrA, on the other hand, selectively enhances expression of genes downstream of the P3 promoter. This includes RNAIII and α-hemolysin (Novick RP et al., 1993). The non-coding RNA, RNAIII, regulates the expression of many surface proteins and synthesis of secreted proteins thereby regulating the phenotypic switch in *S. aureus* from a persister to a virulent phenotype (Figure 4.1A). *S. aureus* strains differ in the *agr*-type. This refers to changes in the sequence of AgrD encoding four variant AIPs. (AIPI- IV, Figure 4.1B). It was also noted that more complex and temporal factors play a role in activating *agr in vivo* while the activation occurs during the late exponential or early stationary growth phases *in vitro* (Novick RP et al., 2008, Kavanaugh JS et al., 2016 and Balasubramanian D et al., 2017).

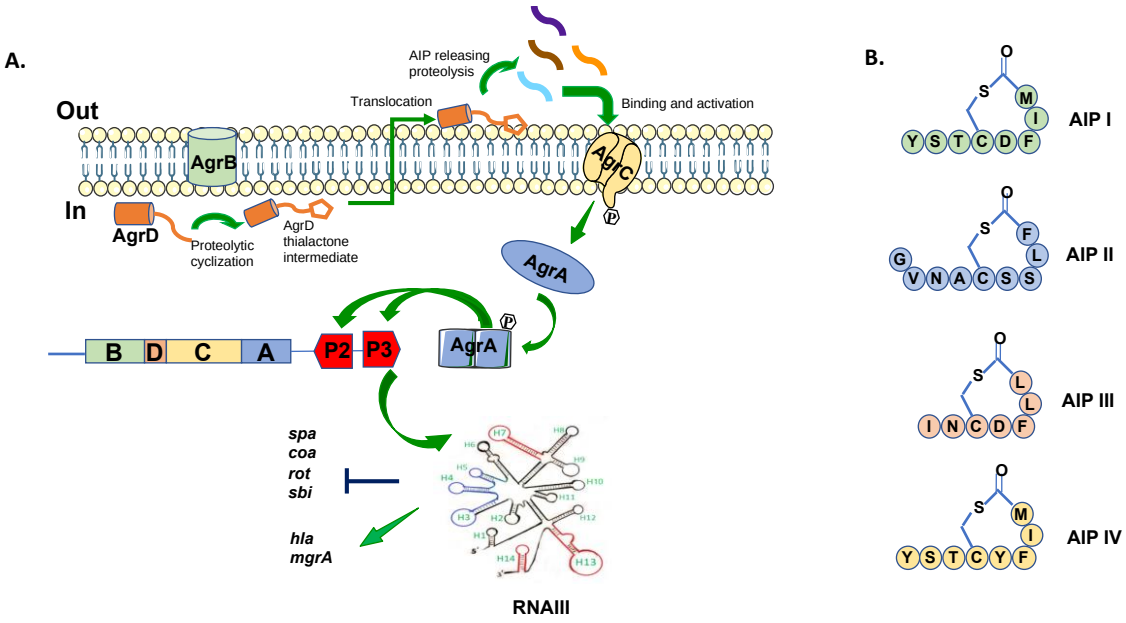


Figure 4.1: Accessory gene regulatory (agr) circuit in *S. aureus*

A. The *S. aureus* cell. Agr locus containing two divergent promoters P2 and P3 (red arrow). P2 drives the *agrBDCA* operon with gene products AgrA, AgrB, AgrC and AgrD. In AIP biosynthesis, AgrD undergoes proteolytic cyclization to form the AgrD thiolactone intermediate. This gets translocated to release the mature autoinducing peptide (AIP). AIP binding to AgrC activates the two-component system (AgrA and AgrC). Phosphorylated AgrC transfers the phosphoryl group to AgrA. Further phosphorylated AgrA binds and activates P2 and P3 promoters resulting in a positive feedback loop. P3 activates RNAIII **B.** Structure of the four AIPs. Lengths vary from 7-9 amino acid containing a five residue thiolactone ring**.**

Different cellular imaging techniques have been developed over the past decade to understand the structural features and temporal association of biomolecules, both *in vitro* and *in vivo*. Fluorescent proteins, like GFP isolated from the jellyfish, *Aequorea victoria*, established a new route to biomolecular imaging. Indeed, multiple diverse application of this strategy have led to novel insights that were hitherto unclear (Enterina JR et al., 2015). Recent efforts have focussed more on the development of more powerful tools for imaging biomolecules (König J et al., 2012). The conventional method of imaging RNA has been fluorescent *in situ* hybridization (FISH). Short exogeneous RNA probes complementary to the target RNA sequence are used in FISH. However, the methodology involved in FISH, like fixing the cells and washing pose severe limitations in evaluating dynamic interactions. (Femino AM et al., 1998). In the

case of the MS2-GFP system, mRNA fluorescence was achieved by tagging the target with a MS2 aptamer (24 nucleotides). Aptamers are small molecules which will bind to RNA mimics of FPs (Ellington AD et al., 1990). These fluorogenic dyes are cell permeable small molecules and bind to the aptamers there by increasing their fluorescence by as much as 5000 times (Ouellet J, 2016 and Bouhedda F et al., 2017). These aptamers are widely known as fluorescent light-up aptamers (FLAPs). FLAPs are significantly small RNA sequences that can bind to cell-permeable, non-toxic, small-molecule fluorogens and enhance their fluorescence upon binding. While the aptamer-GFP fusion was used in a few studies, a limitation in this case is the background fluorescence of unbound MS2-GFP that affects the noise-to-signal ratio (Fusco D et al., 2003). A schematic representation of FLAP is shown in Figure 4.2.

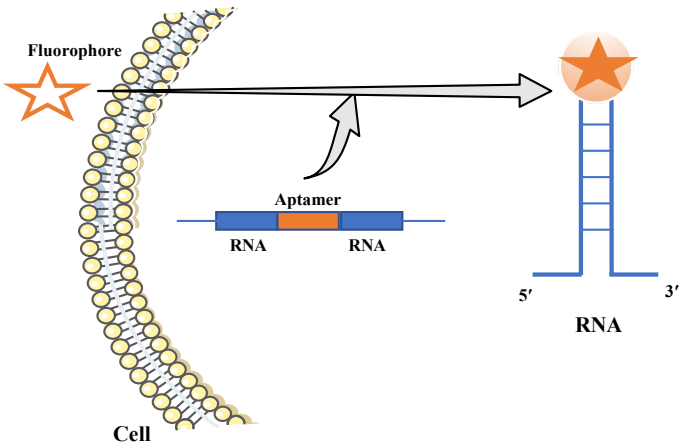


Figure 4.2: Schematic overview of FLAP

A DNA carrying the genetic information of an ROI and aptamer is transferred into a cell. Cellular transcriptional machinery transcribes the RNA. A small molecule fluorogen binds to the aptamer and leads to a drastic increase in the fluorescence.

Multiple strategies have been employed for the design of fluorogenic aptamers. For example, *in vitro* selection of fluorogenic aptamers was performed from a large pool of random RNA sequences through a process termed systematic evolution of ligands by exponential enrichment (SELEX) (Tuerk C et al., 1990). Other strategies to select aptamers are geared to more cell like conditions wherein the focus was to enhance folding of the RNA aptamer and hence increasing the fluorescence brightness (Gotrik M et al., 2018). Many aptamers have been developed till date and each one of them uses diverse classes of fluorophores displaying the strong fluorescence enhancement.

Commonly employed aptamers include the malachite green aptamer, Spinach, Broccoli, Corn, Chili, Mango etc. A list of some RNA aptamers and their corresponding fluorophores are given in Table 4.1.

Table 4.1: Overview of spectral characteristics and fluorogen-aptamer affinities

Aptamer	Fluorogen	K_D [nM]	λ_{ex} [nM]	λ_{em} [nM]	Ref.
MGA	MG	117	630	650	Bebendure JR *et al.*, 2003
Spinach	DFHBI	537	469	501	Paige JS *et al.*, 2011
Broccoli	DFHBI-1T	360	472	507	Filonov GS *et al.*, 2014
Corn	DFHO	70	505	545	Song W *et al.*, 2017
Mango	TO1-Biotin	3	510	535	Dolgosheina EV *et al.*, 2014

K_D- Binding affinity, λ_{ex}- Excitation, λ_{em}- Emission, MG-Malachite Green, DFHBI-3,5-difluoro-4-hydroxybenzylidene imidazolinone, DMHBI- 3,5-dimethoxy-4-hydroxybenzylidene imidazolinone, DFHO- 3,5-difluoro-4-hydroxybenzylidene imidazolinone-2-oxime, TO-Thiazole Orange.

4.1.1 RNA mango aptamer

The RNA Mango aptamer is small sequence of 29 nucleotides with high binding affinity to a thiazole orange-pegylated biotin (TO1-3PEG-Biotin) fluorophore. The TO1-Biotin fluorophores possess several characteristics that makes it an ideal fluorophore which include the small size, lack of toxicity, plasma and nuclear membrane permeability (Dologosheina et al., 2014). The RNA mango aptamer has a G-quadruplex structure and the fluorophore stacks on top of this structure. Upon fluorophore binding, the RNA Mango aptamer prevents free rotation between the benzothiazole and quinoline rings of TO1-3PEG-Biotin. The overall structure of RNA Mango aptamer with TO1-3PEG-Biotin is shown in Figure 4.3. (Dologosheina et al., 2014).

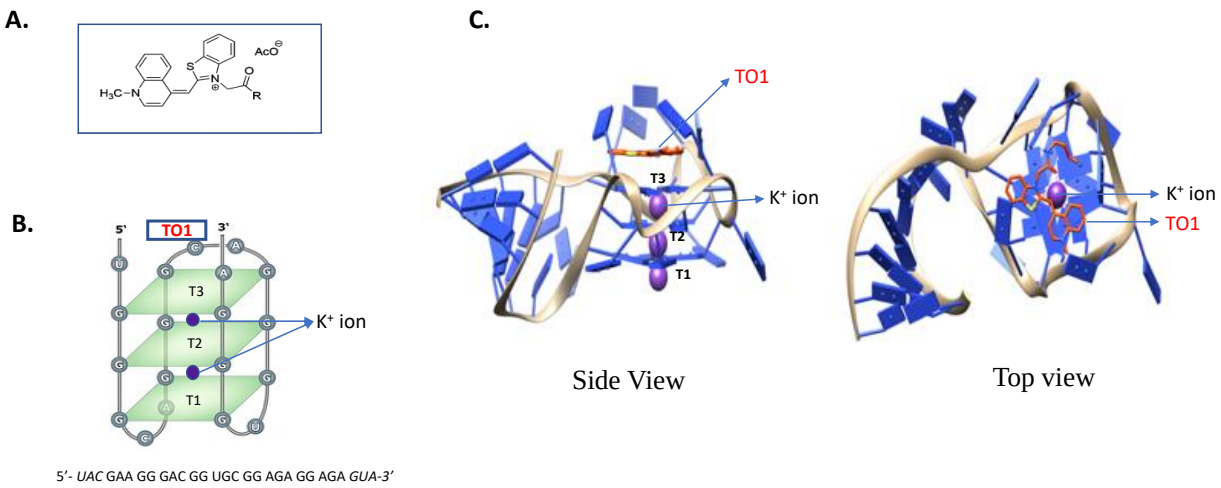


Figure 4.3: Schematic representation of Fluorophore, secondary structure of RNA Mango and crystal structure of RNA Mango complex with fluorophore

A. Chemical structure of TO1 fluorophore (R = Biotin-PEG_3-amine). **B.** Secondary structure of RNA Mango (G-quadruplex structure) in complex with TO1-Biotin. Blue circles represents K^+ ions. Potassium helps in the stabilization of G-quadruplex structure. TO1-3PEG-Biotin binds on top of the G-quadruplex structure. RNA Mango sequence is given below. **C.** Structural illustration of RNA Mango-TO1 complex generated using Chimera (Pettersen EF et al., 2004). The three tiers of the G-quadruplex structure of RNA mango are shown as T1, T2 and T3. The purple spheres represent K^+ ions. (Trachman RJ et al., 2017. PDB ID-6C63)

In general, sRNAs regulate gene expression by base pairing with the mRNA targets (Boisset S et al., 2007, Chabelskaya S et al., 2014). These interactions mediate transcriptional regulation thus altering the transcriptional profile. A pictorial representation of this feature is shown in Figure 4.4.

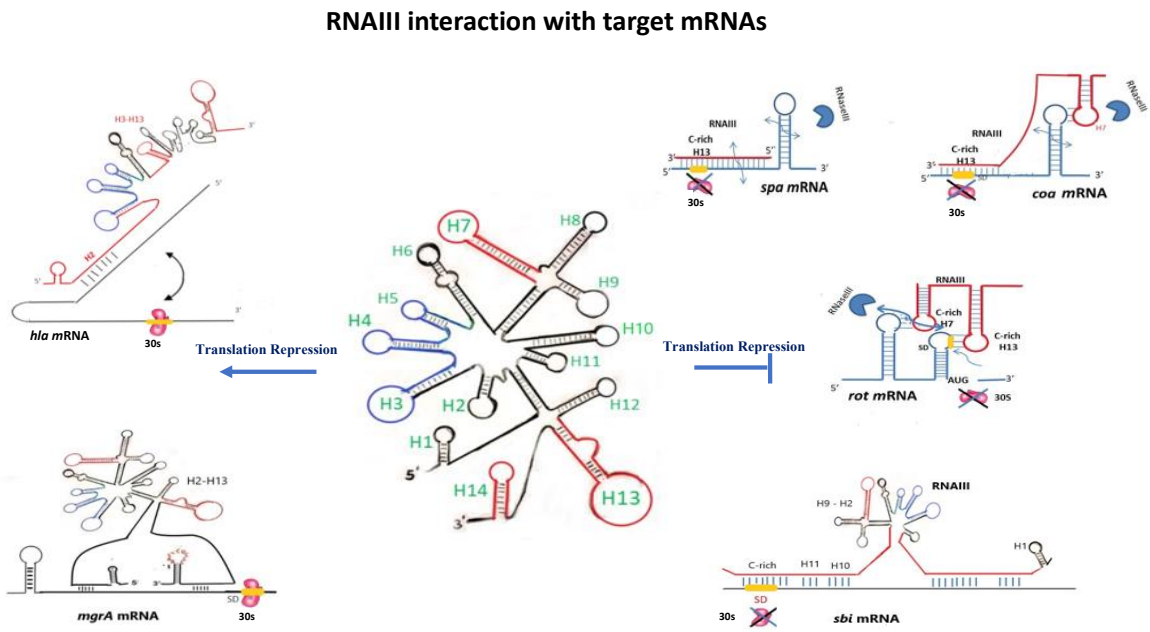


Figure 4.4: RNAIII interaction with target mRNAs and its mechanism of action

Schematic representation of RNAIII structure (adapted and redrawn from Benito Y et al., 2000). Various topologies of RNAIII with its mRNA targets are shown in the figure. Various C-rich motifs or the seed sequences bind to the SD (Shine-Dalgarno) sequence of mRNA targets (yellow). This would prevent the binding of 30S ribosomal RNA (red) to SD sequence and in many cases the mRNA is rapidly degraded by the double strand specific endoribonuclease III (RNaseIII, blue). The *hld* gene encoding δ-hemolysin is given in blue in the structure. For *hla* mRNA, it forms an inhibitory structure sequestering the SD sequence when it interacts with RNAIII. For *mgrA* mRNA, binding of RNAIII stabilizes the mRNA. Right hand side panel represents the translation repression of many virulence factors like protein A (*spa*), coagulase (*coa*), repressor of toxins (*rot*) and immunoglobulin-binding protein (*sbi*) by RNAIII hairpin loops H7 and H13 in 3′ domain. These hairpin loops co-ordinately or separately act to repress the target mRNAs (Boisset S et al., 2007). In case of *sbi* mRNA both 5′ and 3′ domains of RNAIII are involved in the interaction (Chabelskaya S et al., 2014). Left hand side panel represents the translation activation of *hla* (codes for alpha toxin) and *mgrA* mRNA (codes for global regulator protein). Interaction between *hla* mRNA and RNAIII helps the SD sequence to ease the translation of *hla* mRNA (Morfeldt E et al., 1995). In case of *mgrA* mRNA interaction with RNAIII stabilizes the complex thereby occluding the degradation by RNases (Gupta RK et al., 2015).

As is evident from the observations compiled in Figure 4.4, the bifunctional RNAIII (514 nt) regulates gene expression in *S. aureus*. RNAIII has 14 stem loops with both coding and non-coding regions. This ORF also encodes for δ-hemolysin, a pore forming toxin. Expression of RNAIII was noted to trigger a switch between the expression of surface proteins and synthesis of excreted toxins (Novick RP et al., 2008, Tomasini A et al., 2014). As shown in Figure 4.4, RNAIII (loops H7, H13, and H14) contains unpaired and accessible UCCC stretches (C-rich motifs). These are located in the conserved hairpin regions and are predicted to interact with the mRNA translation initiation signals. RNase III-dependent cleavages are observed at the site of interaction (Romilly C et al., 2014).

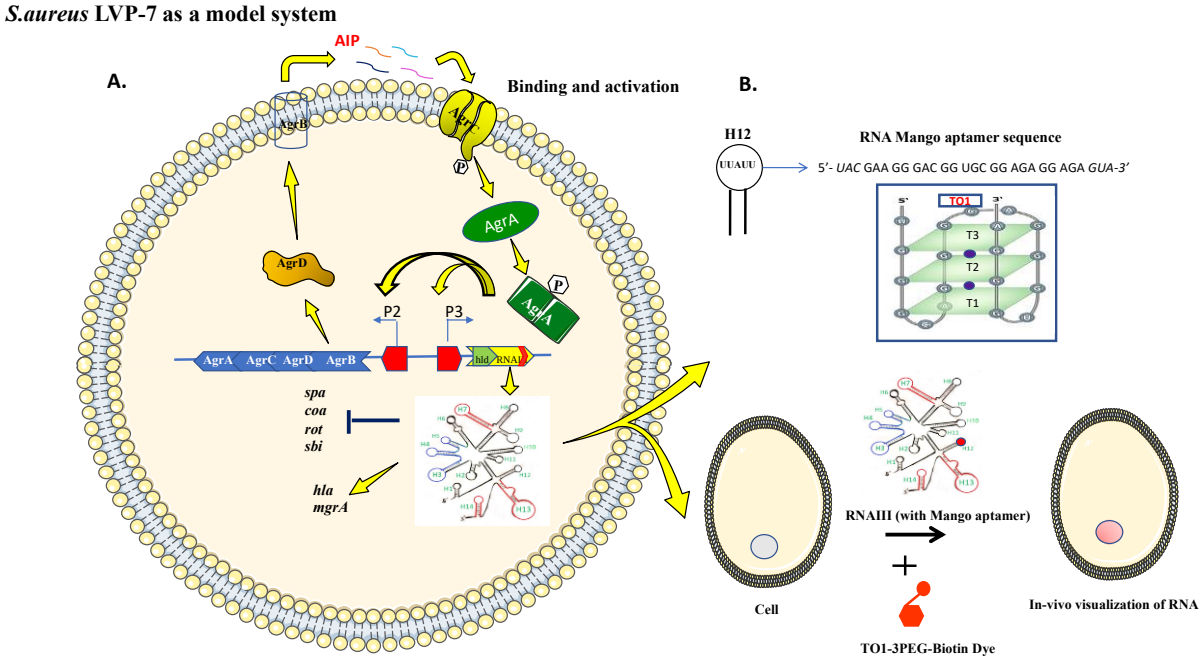


Figure 4.5: Schematic representation of the workflow for *in vivo* visualization of RNAIII in the model organism, *S. aureus* LVP-7

A. Agr quorum sensing system is depicted. Autoinducing peptide (AIP) mediated signal transduction results in virulence regulation through the activation of RNAIII (shown in inset box). Genes activated by RNAIII is shown by yellow arrow. Blue line depicts genes down-regulated by RNAIII. **B.** Top panel represents the schematics of RNA Mango aptamer sequence and its secondary structure (G-quadruplex structure). Dye (TO1-3PEG-Biotin) binding on top of the G-quadruplex structure. Hairpin loop H12 of RNAIII is replaced with RNA Mango aptamer sequence. Down panel represents the steps involved in *in vivo*

visualization of aptamer tagged RNAIII in LVP-7 strain. (RNAIII structure adapted and redrawn from Benito Y et al., 2000)

In this study, we employed a RNA Mango aptamer strategy to visualize RNAIII in *S. aureus* (Dologosheina et al., 2014). The RNA Mango aptamer was used to replace specific hairpin segments in RNAIII. In particular, we chose two hairpin loops of RNAIII- H12 and H3. With several interaction studies on RNAIII with its target mRNAs, the hairpin loop H12 was shown to not participate in interactions with target mRNAs. On the other hand, hairpin loops H7, H12, H13 and H14 participate in translation repression of many target mRNAs by loop-loop interaction (Boisset S et al., 2007, Chabelskaya S et al., 2014, Figure 4.4). The 5′ domain of RNAIII helps in the translation activation of *hla* mRNA and positively regulates δ-hemolysin (*hld*) production (Morfeldt E et al., 1995). δ-hemolysin derives from translation of RNAIII and is considered to be one of the important secreted virulence factor in *S. aureus*. In effect, since *hld* is regulated by the *agr* operon, it can be used as an indicator of *agr* function (Novick RP, 2000). It was envisaged that monitoring the fluorescence *in vivo* by replacing two specific stem loops (H12 and H3) would allow us to understand structure-function features of RNAIII *in vivo*. The *agr* locus in LVP-7 strain belongs to type III *agr* group (Nadig S et al., 2012). This work also discusses on the mechanism of activation of *agr* operon by autoinducing peptide (AIPIII). AIP induced activation and inhibition or agonist and antagonist stimuli activity is observed to affect the virulence gene regulation in *S. aureus* and thus plays a role in bacterial virulence and colonization (Ji G et al., 1997, Srivastava SK et al., 2014). The results of this structure-function analysis is described in the latter sections of this chapter.

4.2 Materials and Methods

4.2.1 Cloning of *rnaIII* gene into *E.coli*-*S. aureus* shuttle vector (pRMC2)

LVP-7 gDNA was used to amplify the *rnaIII* gene using the primers: RNAIII KpnIF_Fw-
5’-GGGTACCTAACTAGATCACAGAGATGTGATGGAAAATAGTTGATGA-3’
RNAIII_EcoR1_Rv- 5’ TGTGAATTCGCCGCGAGCTTGGGAGGGGCTC 3’

between *KpnI* and *EcoR1* restriction sites using Taq DNA polymerase (NEB. Inc). The PCR cycle was set up in Bio-Rad thermocycler using the following conditions: Initial denaturation at 95°C for 30sec and 30 cycles with denaturation at 95°C for 30sec, annealing at 60°C for 40 sec, extension at 72°C for 1 min and final extension at 72°C for 5 min. PCR products were purified using PCR purification kit (Favorgen. Inc). PCR product and the vector-pRMC2 (Corrigan RM et al., 2009) were double digested with *KpnI* and *EcoR1* and further ligated using T_4 DNA ligase enzyme (NEB. Inc). The ligated product were transformed into TOP10 E.coli competent cells and plated on to LB agar plates containing 100µg of ampicillin (Sigma-Aldrich. Inc) and incubated at 37ºC for 16 h. Colony PCR was performed and PCR positive colonies were further processed for plasmid extraction. Positive clone was confirmed by sanger sequencing.

4.2.2 Cloning of RNA Mango aptamer into RNAIII loops

RNAIII hairpin H12 loop nucleotides 388-392 (sequence-5′UUAUU) and hairpin H3 loop nucleotides 89-98 (sequence- 5′AAUGGCACAA) were replaced with the RNA Mango aptamer sequence-5′UACGAAGGGACGGUGCGGAGAGGAGAGUA3′ (Dolgosheina EV et al., 2014). To prepare pRMC2_RNAIII_hairpin H12 construct, we employed overlap extension PCR method using Phusion polymerase (NEB). Briefly, PCR products, AB and CD regions were amplified separately using A/B and C/D primers. The PCR conditions used were as follows -Initial denaturation at 95°C for 3 min and 30 cycles with denaturation at 95°C for 30 sec, annealing at 55°C for 30 sec, extension at 72°C for 1 min and final extension at 72°C for 5 min. PCR products were purified using PCR purification kit (Favorgen. Inc). Further, the AB and CD products were mixed (100 ng each) and subjected to 2 cycles of PCR at 55°C annealing temperature. Then, 250 nM of A and D primers were added and 30 cycles PCR with an annealing temperature of 60°C was performed to finally achieve the product of interest (AD). The product AD and vector pRMC2 were digested with *KpnI* and *EcoR1*. Further protocol for cloning was followed as given in section 4.2.1. The schematic representation of the same is shown in Figure 4.6. The primers designed for the study are given in Appendix Table 6.2. Further these constructs were electroporated into LVP-7 wild type strain.

4.2.3 Electroporation of RM12 and RM3 constructs into LVP-7

Detailed methodology for electrocompetent cells preparation and electroporation is mentioned in Appendix III. The electrocompetent cells of LVP-7 was prepared as mentioned in Löfblom J et al., 2007. Briefly, 50µl of electrocompetent cells were mixed with 5µg of plasmid DNA (pRMC2_hairpin H12 replaced with RNA Mango aptamer and pRMC2_hairpin H3 replaced with RNA Mango aptamer) and electroporation was performed. The cells were immediately mixed with Tryptic Soya Broth (TSB, Himedia) containing 500 mM sucrose and incubated at 37°C for 2 h. Plated the cells on TSB plate containing 10 µg/ml chloramphenicol. The RNAIII constructs replaced with RNA Mango aptamer were further referred as below unless otherwise specified.

RM12 construct: pRMC2_hairpin H12 replaced with RNA Mango aptamer transformed in LVP-7

RM3 construct: pRMC2_hairpin H3 replaced with RNA Mango aptamer transformed in LVP-7

4.2.4 Growth Curve analysis of RM12 and RM3 constructs along with LVP-7 strain

Two different constructs RM12 and RM3 were inoculated into TSB containing 10 µg/ml chloramphenicol. Untransformed LVP-7 was inoculated onto TSB. O.D 600 nm of the overnight grown cells were measured. Cells from same O.D at 600 nm from overnight grown culture (1%) was re-inoculated into fresh TSB and absorbance at 600 nm was recorded at 3, 5, 7, 9 and 12^{th} hour time points in a microplate reader (SpectraMax M3). The growth curve was plotted using GraphPad Prism v9.1.0.

4.2.5 Hemolysin Assay

δ-hemolysin assay was performed with LVP-7 untransformed, RM12 and RM3 constructs. The protocol for the same was adapted from Sakoulas et al., (2002). *S. aureus* cultures grown in TSB media were streaked perpendicular to RN4220 strain on sheep blood agar plate (SBA, Himedia). The plate was incubated at 37°C for 24 h and then at 4°C for 24 h to trigger a cold shock response. Demonstration of this assay is shown in Figure 4.10

4.2.6 Biofilm assay

Biofilm formation by RM12 and RM3 constructs was assessed with colorimetric method based on crystal violet. Biofilm assay for RM12 and RM3 constructs along with LVP-7 untransformed strain was performed in 96-well Nunclon micro titre plates (Nunc), using Brain Heart Infusion broth containing 0.25% glucose. *S.aureus* strain SH1000 and *S.* epidermidis strain ATCC12228 as negative control. Briefly, bacteria was grown in BHI-0.25% glucose broth, at 37ºC for 18-24h. Cultures were diluted 1:100 in BHI -0.25% glucose broth and 200 μl was inoculated into each well. The micro titre plate was incubated at 37 ^{o}C for 24 h. Without disturbing the attached cells, the supernatant was removed from each well carefully and biofilms was washed gently twice with 1X PBS, and then dried. Finally, plate was stained with 0.1% crystal violet, gently washed twice and destaining solution (50% ethanol and 50% glacial acetic acid) was added and the absorbance at 490 nm of released stain was measured with a micro plate reader (Jain A et al., 2009). This quantitative test results are shown in Figure 4.11.

4.3 Fluorescence studies

4.3.1 Live cell FACS counting of RM12 and RM3 constructs

LVP-7 strain was transformed with 5 µg of both pRMC2_RM12 plasmid and pRMC2_RM3 plasmids separately and plated on TSA containing 10 µg/ml chloramphenicol. Single colonies picked from these plates were labelled as RM12 and RM3 (refer section 4.2.3) and grown overnight in TSB medium containing 10 µg/ml chloramphenicol (TSB+cm). Overnight culture is diluted 50-fold into fresh TSB+cm and grown to the OD_{600} of 0.2. 500 ul of this culture was removed, incubated with 5 µM TO1-3PEG-Biotin (Applied Biological Materials Inc.) and induced with 200 ng of anhydrotetracycline (ATc) at the same time. Aliquots of growing cells expressing RNAIII was removed at specified times and counted for mean fluorescence using flowcytometry. Bacterial samples were pelleted by centrifugation at 4500 rpm for 10 min and washed once with 1XPBS after which the samples were fixed using 3% paraformaldehyde (PFA) for 30 min at room temperature in dark. Next, PFA was

quenched using 1 ml of PBS, washed and finally suspended in 300 µl of PBS. The fluorescence intensity was collected by exciting the samples at 488 nm laser on a flow cytometer (BD GBD FACSCelesta™ Flow Cytometer, BD Biosciences, US). Data was analyzed using Flow Jo (Tree Star, Inc. Ashland, OR, USA).

4.3.2 Live cell imaging of RM12 and RM3 constructs using Spinning disc confocal microscopy (SDCM)

LVP-7 WT, RM12 and RM3 constructs were grown for up to an O.D 600 nm of 0.2 and induced with 200 ng of anhydrotetracycline (ATc). Post induction, the cells were further grown for 90 min. 500 µl of cells were centrifuged and washed with 1XPBS. The cells are mixed with 100 µM of TO1-PEG-Biotin fluorophore. This mixture was added on to Poly-L-lysine 0.1% (w/v) treated glass slide and coverslip. The images were captured in SD confocal microscopy using 488 nm filter (Yokogawa CSU-X1 spinning disk module, an Andor iXon 897 EMCCD camera and an Oko Lab stage top incubator). The microscope was controlled using Nikon NIS Elements software or Micromanager. Different fields (bright field and fluorescence at 488nM) were captured at 30P (15 mW) laser power for intensity plots. Photobleaching movie was captured at laser power-70P (35 mW) for 3min. Images were further processed using ImageJ software.

4.3.3 Monitoring AIP induced activation of RNAIII expression

To study AIP-induced activation of *agr* QS, a supernatant of *agr type* III strain LVP-7 was prepared. Briefly, the cells were grown overnight at 37^{o}C in TSB up to 1.34 O.D at 600 nm. The culture supernatants were subsequently passed through 0.2-µm filters as described earlier (Balaban N et al., 2011). LVP-7 untransformed and RM12 construct were induced with AIPIII. AIP induction was done at the same time point of ATc induction for RM12 construct. Post induction, the cells were further incubated at 37^{o}C for about 90 min and proceeded further for monitoring the fluorescence by live-cell imaging and 500 µl of same O.D cells were stored in RNA protect reagent (Qiagen) for RNA isolation. To confirm the AIPIII peptide mass, the prepared supernatant of

AIPIII was analysed by LC-MS/MS (Esi Q-TOF, Impact HD, Bruker Daltonics, Germany). Analysis of the mass spectra data was done as given in Kalkum et al., 2003. The data is represented in Figure 4.15.

4.3.4 RNA isolation, cDNA synthesis and qRT- PCR analysis for AIPIII induced RM12 construct:

LVP-7 WT, RM12 and RM3 constructs were grown for up to an O.D 600 nm of 0.2 and induced with 200 ng of anhydrotetracycline (ATc). At the same time point the cells were induced with culture supernatant of AIPIII in 1:1 ratio. Post induction, the cells were further grown for 90 min (O.D 600 nm of 0.3-0.4). At this point cells can be frozen and stored at -80 ^{o}C. Total RNA samples were extracted using RNeasy mini kit (Qiagen). Initially, cells were lysed in 30mM Tris pH 8.0, 1mM EDTA containing 100 µg of lysozyme (Sigma-Aldrich), 200 µg of Proteinase K (Qiagen), 100µg of Lysostaphin (Sigma-Aldrich) and further steps were followed as suggested by the manufacturer and finally eluted in 40µl of RNase- free water. cDNA was prepared using QuantiTect® Reverse Transcription kit (Qiagen) according to the manufacturer's instructions. Real Time PCR was carried out using cDNA (20ng) to quantitate the levels of RNAIII using SYBR green super mix (Bio-Rad) in Bio-Rad iQ5^{TM} thermo cycler. 16sRNA gene was amplified as an endogenous control. The relative levels of expression of genes were determined by using the $2^{-\Delta CT}$ method (Livak KJ et al., 2001). The primers used to amplify the target genes are listed in Appendix Table 6.1

4.4 Results

4.4.1 Cloning of RNAIII into the pRMC2 shuttle vector

The successful ligation of *rnaIII* gene into pRMC2 vector was confirmed by sequencing. The map of the plasmid DNA construct is represented in Figure 4.5A and all the agarose gel snapshots that confirm the cloning steps are shown in Figure 4.5B.

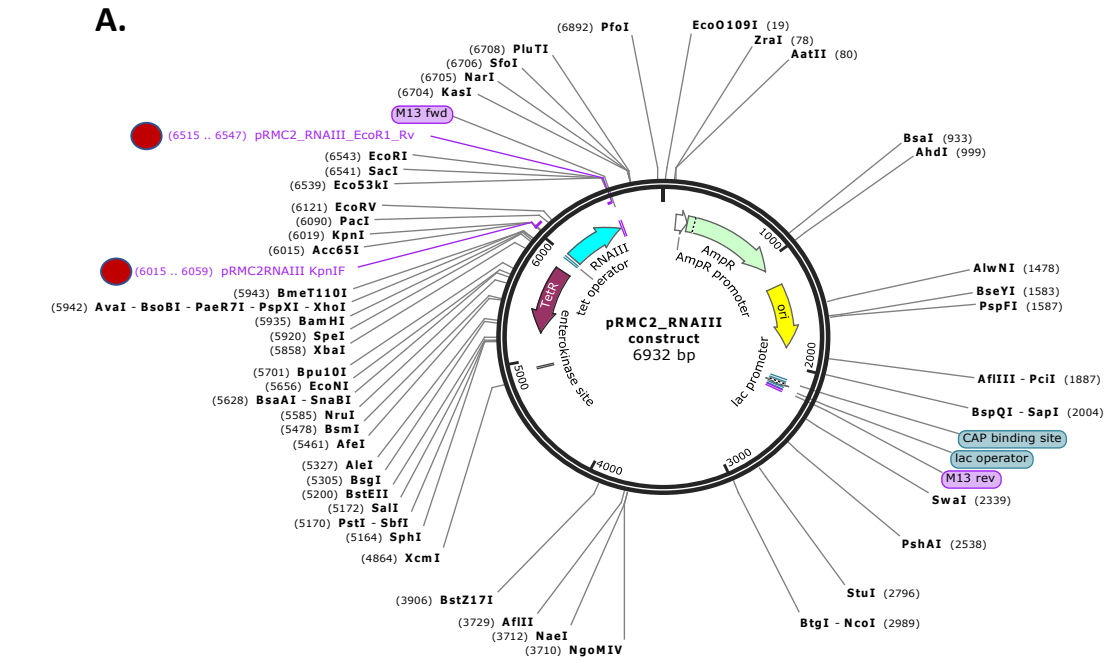


Figure 4.6: Cloning of RNAIII into pRMC2 vector

A. Map of pRMC2_RNAIII construct generated using snap gene tool. RNAIII was cloned in between *KpnI* and *EcoR1* restriction sites (shown in red circle) **B.** Agarose gel pictures of PCR amplification and cloning of RNAIII.

4.4.2 Insertion of the RNA Mango aptamer into RNAIII at positions H3 and H12

Overlap extension PCR: The pRMC2_RNAIII_hairpin H12 and pRMC2_RNAIII_hairpin H3 constructs were successfully generated using overlap-

extension PCR method using Phusion polymerase (NEB). A schematic representation of the same is shown below.

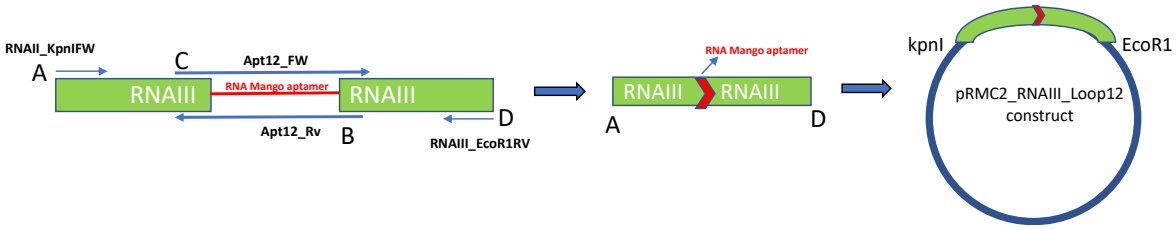
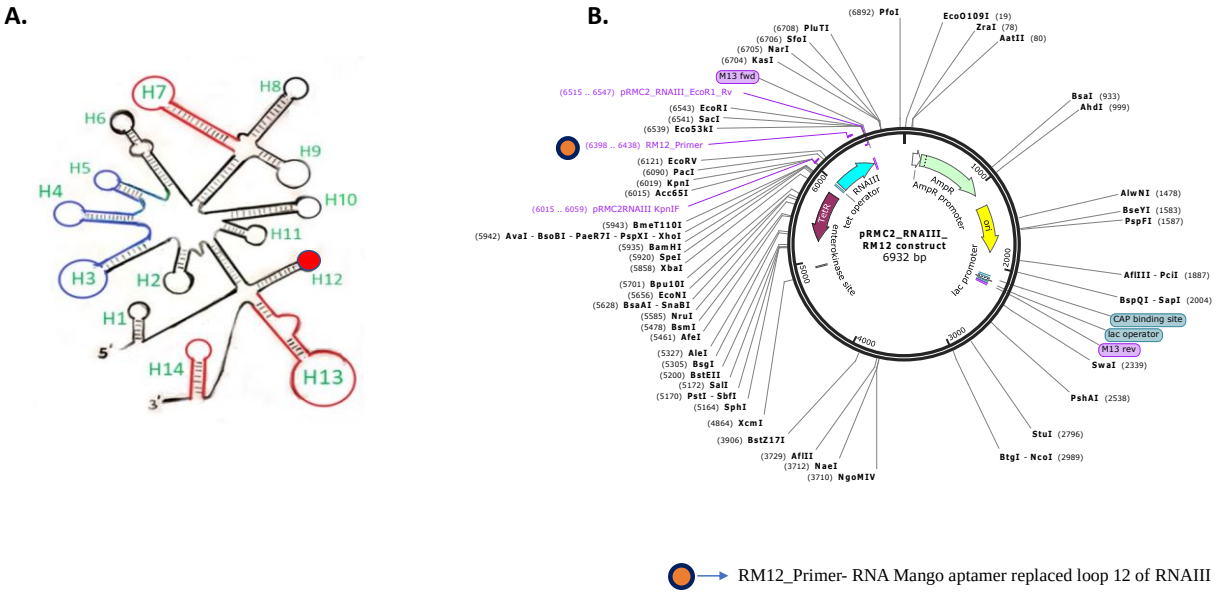


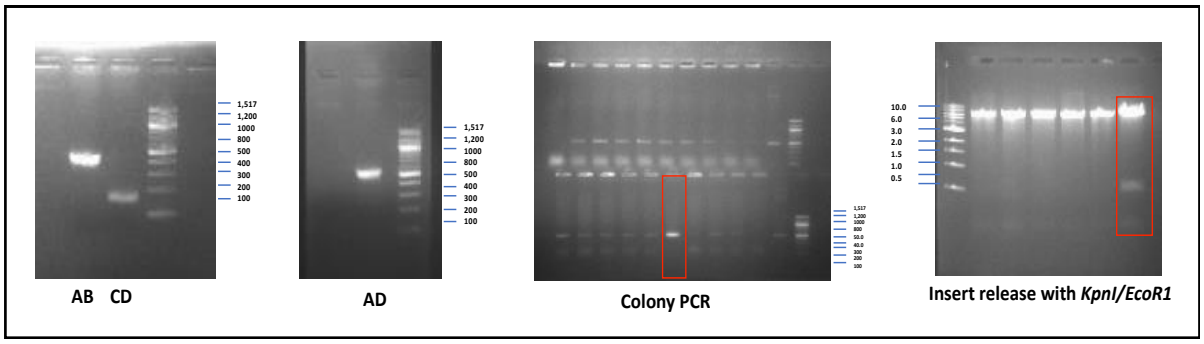
Figure 4.7: Schematic representation of Overlap extension PCR

Overlapping primers C and B are shown as long blue arrow. Primers A and D have the restriction sites *kpnI* and *EcoR1* respectively. Small red arrow represents the RNA Mango aptamer sequence.

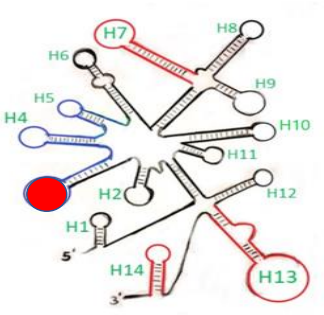
To tag the RNA Mango aptamer to the RNAIII hairpin loops we chose two loops, hairpin H12 and hairpin H3. As stated in the introduction, the H12 loop is considered to play the role of a scaffold and is not known to interact with any mRNA element. The H3 loop, on the other hand, codes for δ-hemolysin (*hld*) and carries the start codon for *hld* mRNA (Figure 4.4).



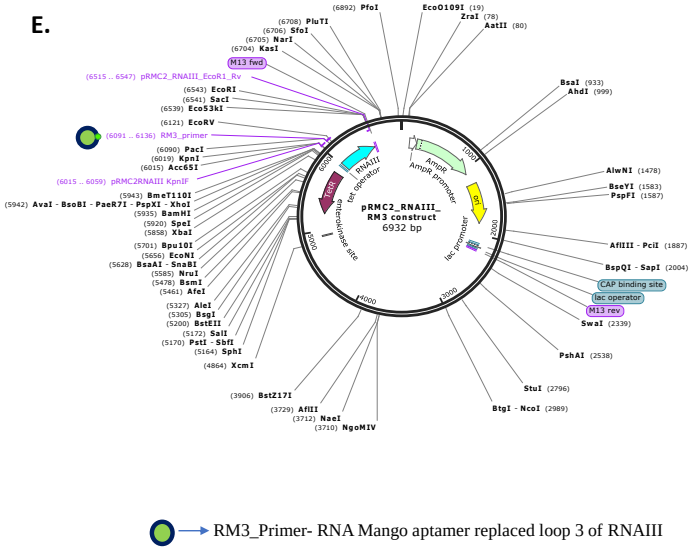
C.



D.



E.



F.

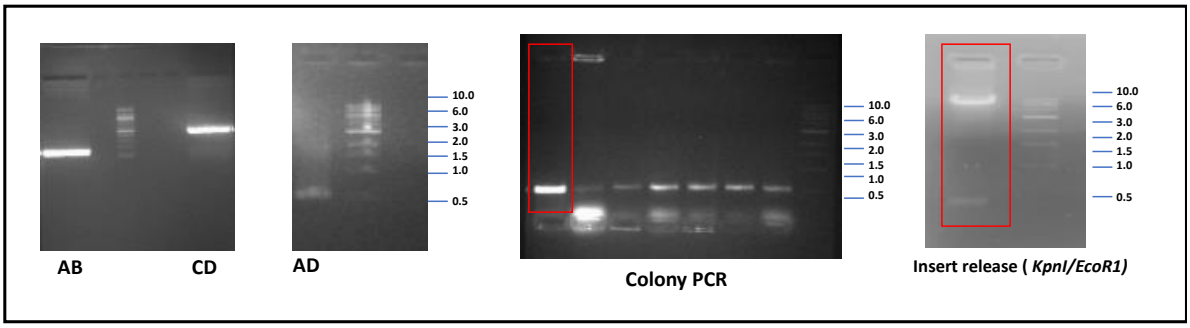


Figure 4.8: PCR amplification and cloning of the aptamer constructs

A. Secondary structure of RNAIII representing the hairpin H12 replaced with RNA Mango aptamer. **B.** Map of pRMC2_hairpin H12 replaced with RNA mango aptamer, **C.** Agarose gel pictures of PCR amplification and cloning of RM12 construct. **D.** Secondary structure of RNAIII representing the loop hairpin H3 replaced with RNA Mango aptamer. **E.** Map of pRMC2_hairpin H3 replaced with RNA Mango aptamer.

F. Gel pictures of PCR amplification and cloning of RM3 construct. These constructs were subsequently transformed into the *S. aureus* LVP-7 strain. In the subsequent sections, the RM12 construct refers to pRMC2_hairpin H12 replaced with RNA mango aptamer whereas the RM3 construct refers to pRMC2_hairpin H3 replaced with RNA Mango aptamer.

4.4.3 Introduction of mutant RNAIII does not significantly affect bacterial growth

A growth curve (Figure 4.9) was superposing data from the RM12, RM3 and the un-transformed *S. aureus* LVP-7 strain revealed that introduction of the plasmids did not significantly affect the viability of the bacterial cell. For this analysis, the Optical Density at 600 nm was monitored at different time points using a microplate reader. As observed in Figure 4.9, the LVP-7 untransformed and RM12 showed normal growth pattern. In comparison, RM3 had a slow growth.

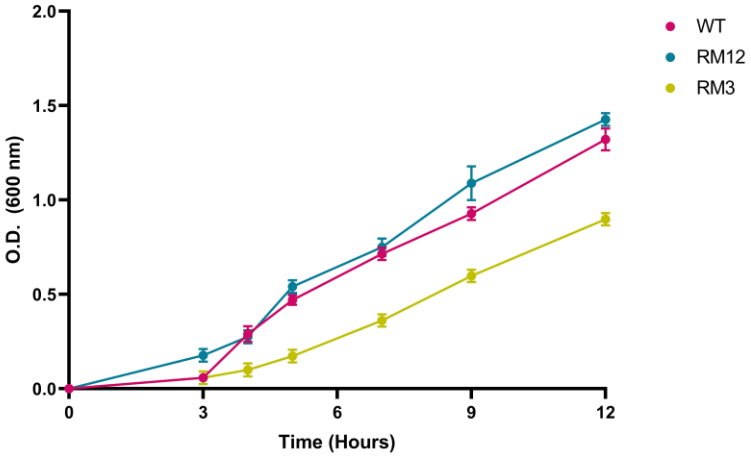


Figure 4.9: Growth curve analysis

Growth curve analysis at different time points was performed for RM12 and RM3 constructs along with LVP-7 untransformed strain (WT). The data shown represent an average of six measurements (two biological replicates). WT- LVP-7, RM12-pRMC2_hairpin H12 replaced with RNA mango aptamer, transformed in LVP-7, RM3-pRMC2_hairpin H3 replaced with RNA Mango aptamer, transformed in LVP-7.

4.4.4 Hemolysin Assay for RNAIII induction (virulent state)

Induction of the *agr* operon by the quorum sensing mechanism can be directly monitored by the production of δ-hemolysin, a pore forming toxin. It is known that both β- and δ-hemolysin collectively act to lyse sheep red blood cells (Sakoulas G et al., 2002). For this study, the *S. aureus* strain, RN4220, a β-hemolysin producer was used as a control in the hemolysin assay. Both LVP-7 (untransformed) and the RM12 strain producing δ-hemolysin resulted in an enhanced zone for hemolysis. This was significantly lower in the case of the RM3 construct. This finding suggests that δ-hemolysin production is significantly lower in RM3 wherein the H3 loop in RNAIII was replaced by the RNA Mango aptamer.

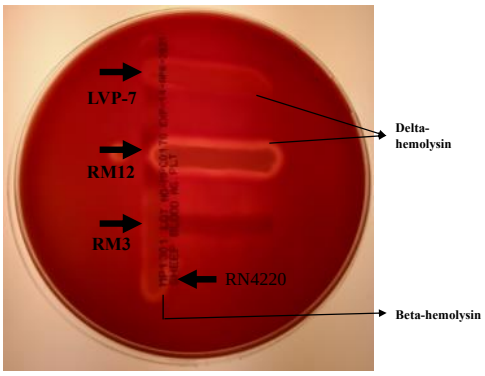


Figure 4.10: Hemolysin assay

δ-hemolysin assay was performed on sheep blood agar plate. Perpendicular to RN4220 strain, test strains (LVP-7, RM12 and RM3) were streaked. Enhanced zone of clearance was observed in LVP-7 (untransformed) and RM12 construct. In RM3 construct, zone of clearance was not observed indicating that *hld* production was perturbed.

4.4.5 Biofilm Assay to estimate the persistent state

The ability of the LVP-7 untransformed strain, RM12 and RM3 constructs to form biofilms (a characteristic of the persistent state) was analysed. In this assay, *S. aureus* (SH-1000) served as a control to evaluate high biofilm formation whereas *S. epidermidis* ATCC12228 served as a negative control for non-biofilm producer. As shown in Figure 4.11A and Figure 4.11B, the RM3 construct shows high biofilm formation similar to SH-1000. On the other hand, the biofilm formed in the case of the

RM12 construct is comparable to the wild type LVP-7 untransformed strain (a significance value of $p \leq 0.05\%$ with SH-1000).

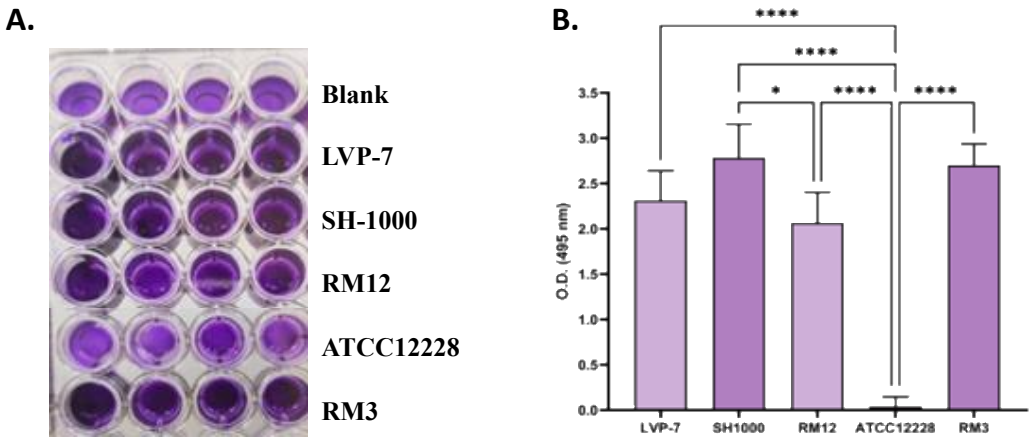


Figure 4.11: Biofilm Formation evaluated by the crystal violet method

A. Qualitative analysis of biofilm performed in microtiter plate. **B**. The colour intensity was measured at O.D 495 nm. *S. aureus* (SH1000) and *S. epidermidis* (ATCC12228) were used as positive and negative controls respectively. Statistical analysis was performed using One-way Anova test.

4.4.6 Transcriptional regulation of RNAIII leads to fluorescence activation by AIP stimuli

In order to assess the transcriptional regulation of RNAIII by AIP stimuli, both fluorescence based imaging and qRT-PCRs were performed. As a basic step, the fluorescence intensity of both the RNAIII constructs (RM12 and RM3) induced with anhydrotetracycline (ATc) to drive RNAIII expression along with untransformed LVP-7 were analysed using flowcytometry (Figure 4.12) and fluorescence imaging (Figure 4.13). The image series were captured until photobleaching occurred (Figure 4.14). Further, AIP induced activation assay was performed with RM12 construct. Presence of AIPIII in supernatants was confirmed by mass spectroscopy and AIPIII induced RNAIII expression was analysed by live-cell imaging of LVP-7 using fluorescence microscopy. These results are represented in figure (Figure 4.15A, Figure 4.15B and Figure 4.15C).

4.4.6.1 Live cell FACS counting of RM12 and RM3 constructs

The fluorescence intensity corresponding to RNAIII expression was measured using flowcytometry. As expected, in the case of RM12 induced with anhydrotetracycline (ATc) we note the highest fluorescence intensity when compared to the uninduced RM12 and LVP-7 untransformed strain. The fluorescence intensity of RM3 induced with ATc was similar to the uninduced sample, suggesting that the TO1-3PEG-Biotin fluorophore did not bind to the RNA Mango aptamer in RNAIII. It appears likely that perturbing the *hld* start codon site might affect the folding of the RNAIII. Further experiments would be needed to evaluate this hypothesis.

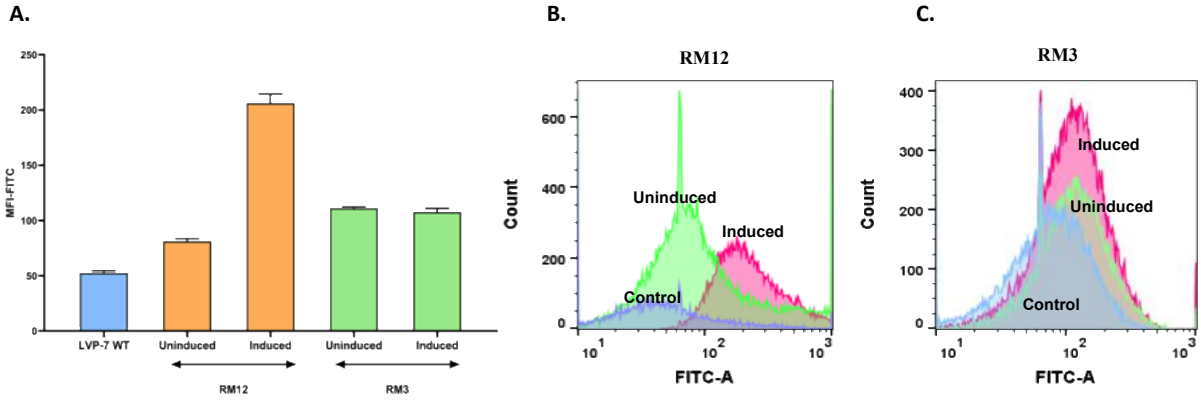


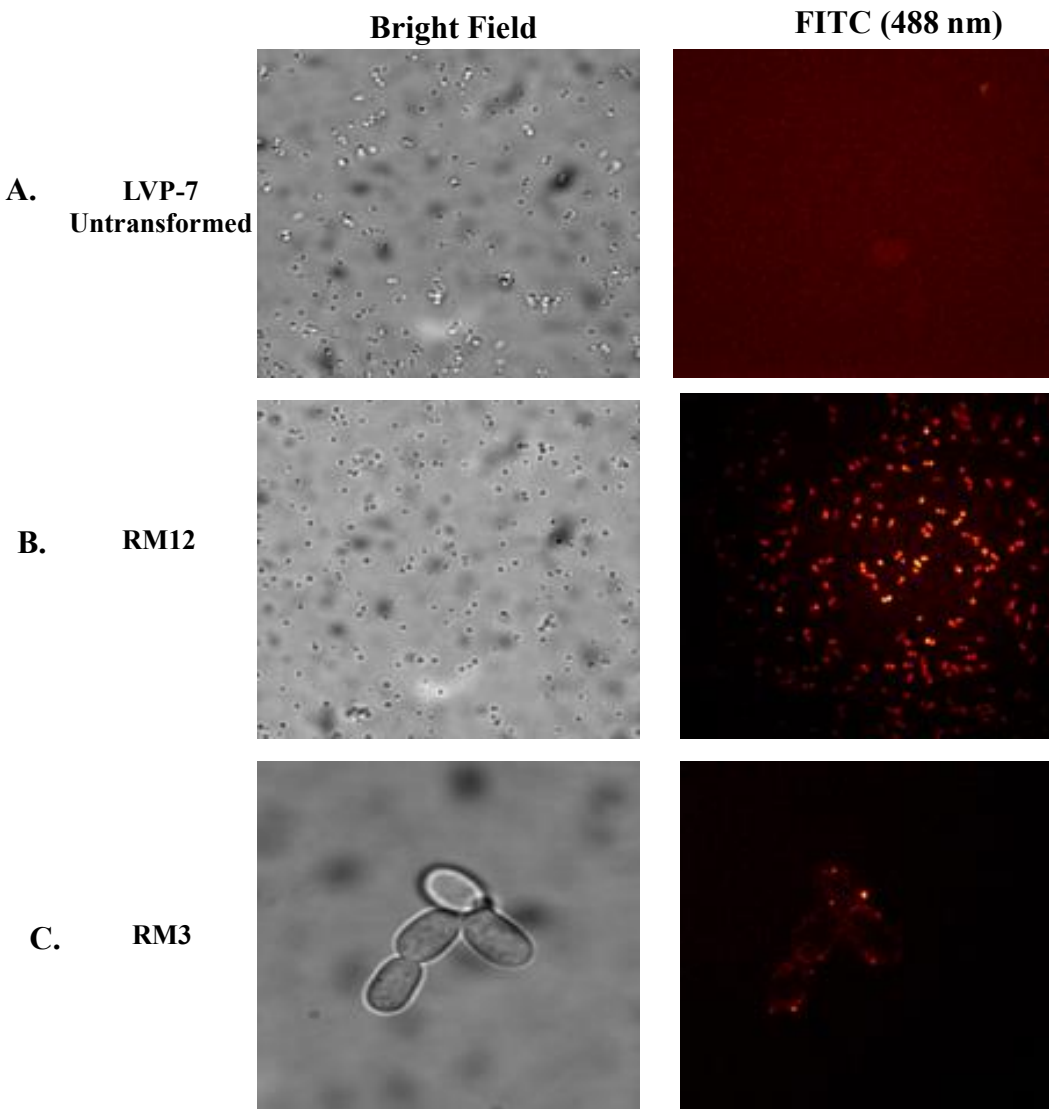
Figure 4.12: Fluorescence intensity measurements in RM12 and RM3 constructs

A. Median fluorescence intensity (MFI) of RM12 and RM3 constructs measured using flowcytometry. **B** and **C**. Corresponding representative histograms of RM12 and RM3 constructs.

4.4.6.2 Live cell imaging of RM12 and RM3 constructs using Spinning disc confocal microscopy (SDCM)

Live-cell imaging of the RM12 and RM3 constructs were performed using spinning disc confocal microscopy. The images of both constructs in bright field and in FITC (488 nm) filter was used for imaging the cells. The LVP-7 untransformed strain was used as a control for autofluorescence demonstrating that TO1-3PEG-Biotin binds specifically to RNA Mango aptamer replaced into RNAIII loop. Different fields (bright field and FITC 488 nM filter) were captured at 30P (15 mW) laser power for the intensity plots and the final images were processed in Image J software. The RM12

construct induced with 200 ng of anhydrotetracycline (ATc) showed brighter fluorescence indicating the expression of RNAIII driven by Tet promoter in pRMC2 vector. An interesting observation is that the RM3 construct differs in its morphology when induced with 200 ng of ATc, Indeed, the fluorescence was very low almost similar to the background fluorescence. The RM12 construct was alone considered for further analysis to evaluate the lifetime of RNAIII *in vivo*.



D.

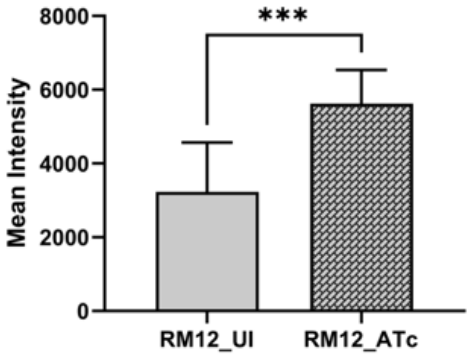


Figure 4.13: Live cell imaging of RM12 and RM3 constructs

A. LVP-7 untransformed, cells were observed only in bright field. **B.** In RM12 construct, cells observed in bright field and fluorescent cells seen at 488 nm FITC filter. **C.** In RM3 construct, morphology difference observed in bright field and less brighter cells seen at 488 nm FITC filter. Different fields for both bright field and FITC (488 nM) filter were captured at 30P (15 mW) and the images were processed in Image J software. The background fluorescence was subtracted before plotting the fluorescence intensity. **D.** Bar graph representation of fluorescence intensity of RM12, RM12_uninduced and RM12 induced with ATc.

4.4.6.3 Photobleaching effect of the RNA Mango aptamer-TO1-3PEG-Biotin complex in RM12 construct

Photobleaching of RNA Mango aptamer bound with TO1-3PEG-Biotin in RM12 construct was observed for 3 mins at 488 nm. Different frames were processed and the mean intensity was calculated to observe the photobleaching effect.

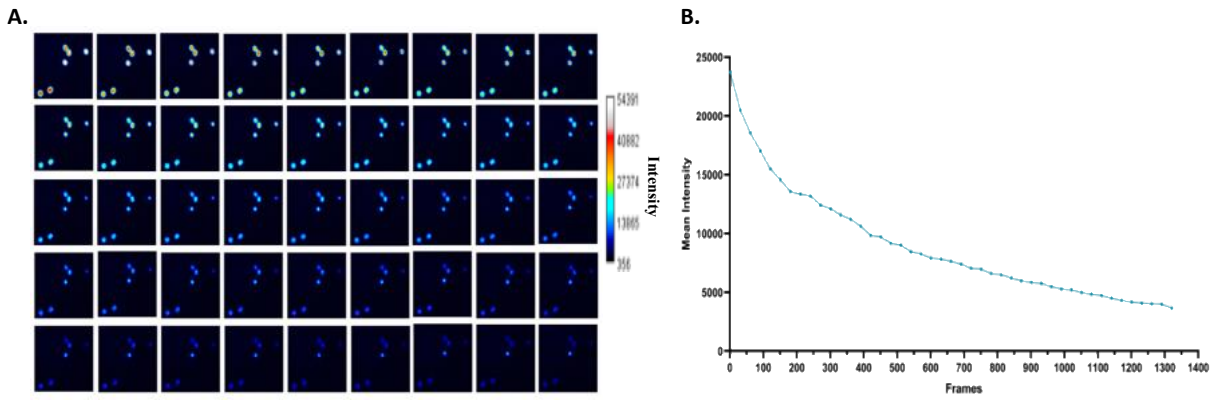


Figure 4.14: Fluorescence microscopy for the detection of RNA Mango aptamer complex with TO1-3-PEG-Biotin in RM12 construct

A typical image of photobleaching of RNA Mango-TO1-3PEG-Biotin complex inside *S. aureus* cell LVP-7 (RM12 construct). **B.** For each image series, the focus intensity was computed to plot an intensity trace. Images processed using ImageJ software. Graph plotted using GraphPad Prism v9.1.0.

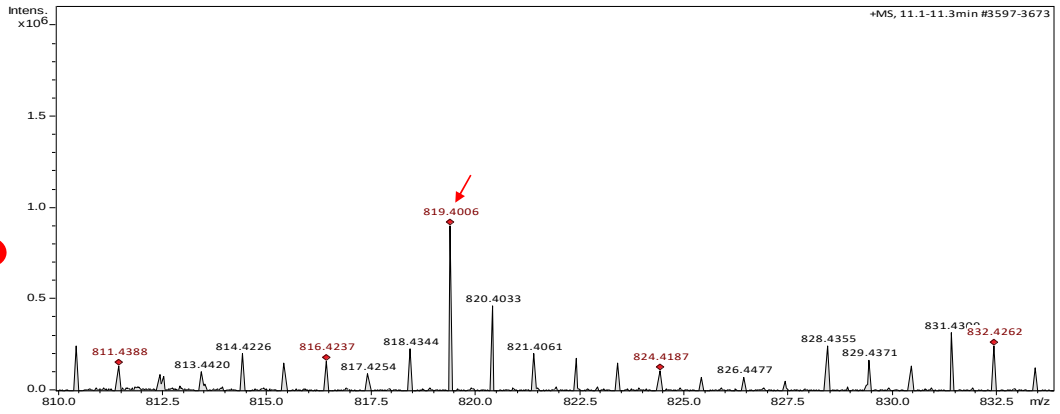
4.4.6.4 Effect of AIP stimuli on transcriptional regulation of RNAIII

In order to understand the AIP induced activation of RNAIII, mutant construct of RNAIII (RM12) was induced with 1:1 ratio of culture supernatant of AIPIII . As LVP-7 strain carries a type III *agr* locus, AIPIII containing supernatants were prepared from LVP-7 strain. The presence of AIPIII peptide was confirmed by LC-MS/MS (Figure 4.15A) and the mass spectra of supernatant containing AIPIII peptide matched with the earlier published data (Kalkum M et al., 2003). The fluorescence images in Figure 4.15B clearly indicates that cells induced with AIPIII have brighter fluorescence. Same is depicted in Figure 4.15B and photobleaching effect is plotted for the AIPIII induced RNAIII construct in Figure 4.15B right panel.

A.

MS

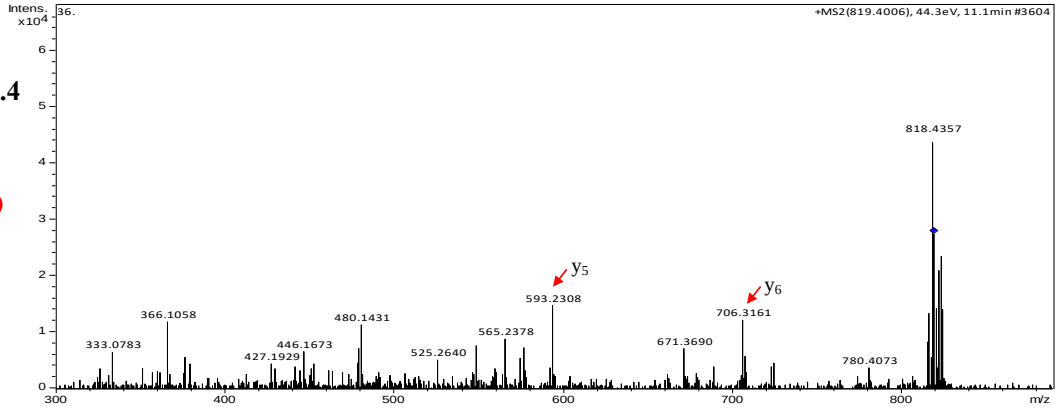
INCDFLL



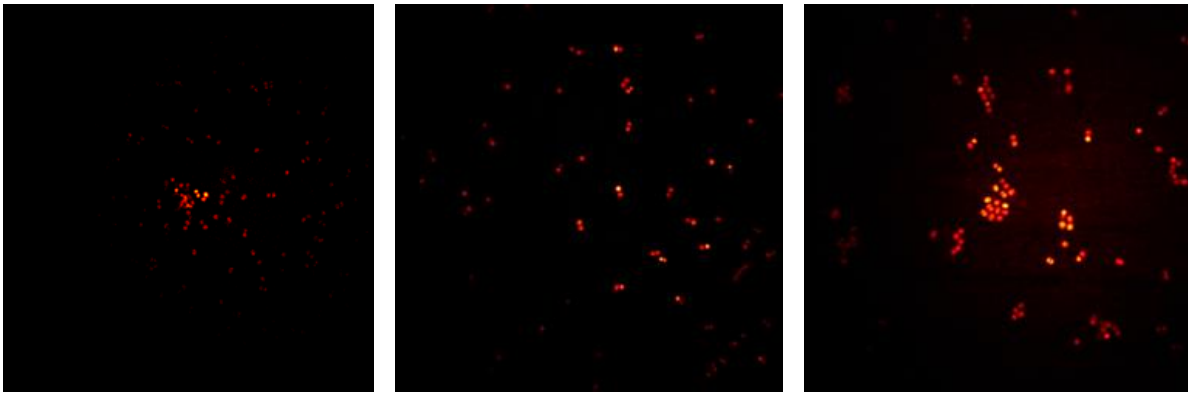
MS/MS: 819.4

y_6 y_5

INCDFLL



B.



RM12_ UI **RM12 _ATc** **RM12_ATc_AIPIII**

C.

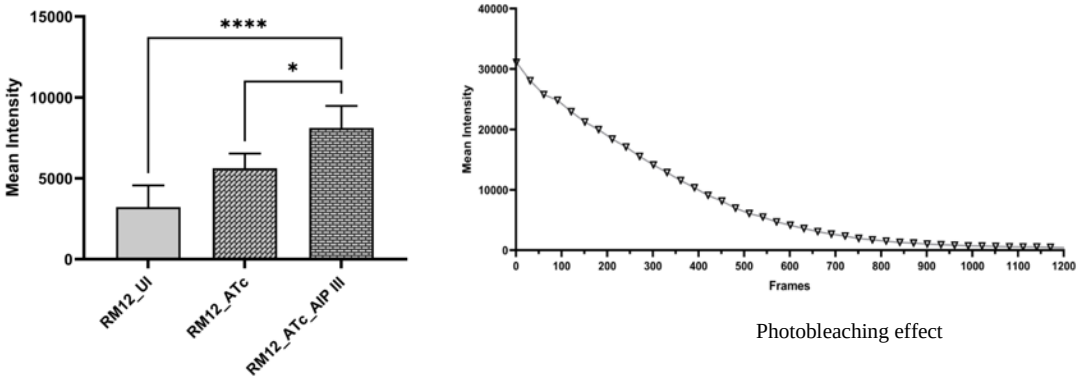


Figure 4.15: AIPIII induced activation of RNAIII expression

A. LC-MS/MS profile of AIPIII from culture supernatants. LC-MS and MS/MS performed in positive mode. LC-MS peak profile depicts mass of AIPIII peptide (819.4 Da), m/z values of y^6 and y^5 ions are shown in MS/MS peak profile. **B.** RNAIII expression was monitored through live cell imaging of LVP-7 strain. To compare the fluorescence intensities, uninduced, anhydrotetracycline (ATc) induced and both ATc and AIPIII induced samples of RM12 construct were used. **C.** Bar graph representation of the intensity plots of RM12 construct-UI (Uninduced), RM12-Induced with ATc, RM12- Induced with ATc and AIPIII.

4.4.6.5 Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis

To validate the findings from fluorescence assay, the AIP induced RNAIII activation was analysed by qRT-PCR. The expression levels of *agr* operon genes like *agrA, agrB, agrC agrD* and *rnaIII* was determined using both uninduced and induced cells (RM12). In this analysis, 16sRNA served as an internal control. The normalised values were plotted and same is depicted in Figure 4.16. As expected AIPIII induced RM12 construct showed higher expression of RNAIII (Figure 4.16). Thus, qRT-PCR data corroborated with the fluorescence data.

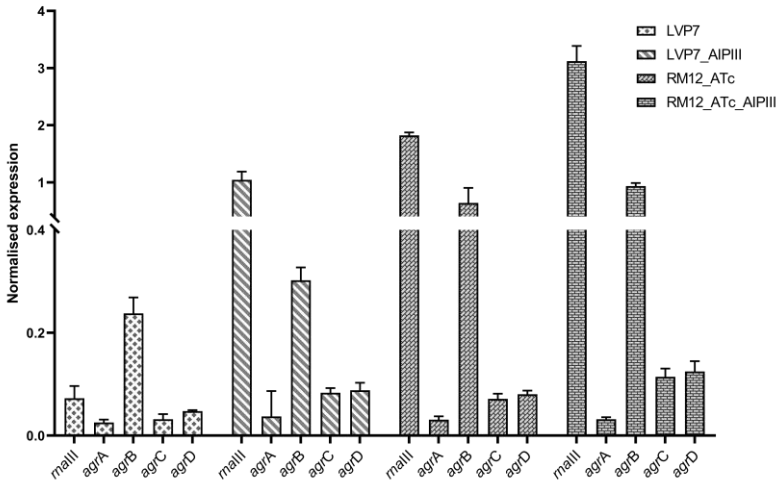


Figure 4.16: The effect of agonist stimuli on the transcription level of genes in the agr operon

Both LVP-7 untransformed and RM12 construct were induced with AIPIII to analyse the stimulus activation of the *agr* quorum sensing (QS) system. Normalised expression using 16sRNA as internal control is calculated. The levels of *agrA* remains similar in both uninduced and induced cells whereas *rnaIII* gene levels increased post induction. The data shown represent an average of six measurements (two biological replicates).

Put together, the fluorescence based assay and qRT-PCR data indicates the AIP induced transcriptional regulation of RNAIII and the distinct roles of different stem loops in this non-coding RNA.

4.5 Discussion

This study demonstrates the feasibility of using the RNA Mango aptamer to evaluate the *in vivo* conformational state of the non-coding RNA, RNAIII. The basis for this study was the finding that the RNA Mango aptamer in complex with TO1-3PEG-Biotin shows a brighter fluorescence (Dolgosheina EV et al., 2014). In a previous study, a novel array of Mango aptamers were used for RNA imaging in both live and fixed cells. This study also demonstrated that tagging of both coding and non-coding RNA by Mango aptamer does not affect the cellular localisation (Cawte DA et al., 2020).

RNAIII, a 514 nt sRNA modulates the gene expression in *S.aureus* by helping bacteria to switch between persistent form to a virulent phenotype (Novick RP et al.,

2008). Towards this goal we have presented a preliminary data on two RNAIII hairpin loops, H12 (non-interacting loop) and H3 (codes for δ-hemolysin) which were replaced by the RNA Mango aptamer in the recombinant RNAIII. An interesting finding was that replacing a functionally relevant loop (the RM3 construct) leads to a reduction in growth rate (Figure 4.9). A relevant report in this context is the finding by Balaban and Novick RP who reported that post RNAIII transcription, there is a delay of 1 hr for *hld* synthesis due to unknown mechanisms (Balaban N et al., 1995). This observation led these authors to hypothesize that *hld* expression is an indicator for the role of RNAIII dependent regulation. Another related observation from our study is that both RM12 and RM3 show high biofilm formation (Figure 4.11). Intriguingly, the RM3 construct also appeared to have a different cell morphology (Figure 4.12). The fluorescence based *in vivo* imaging indicated the AIP induced activation of RNAIII expression and the data was validated using qRT-PCR (Figure 4.15 and Figure 4.16). Indeed, AIP induced activation and inhibition is observed to affect the virulence gene regulation in *S. aureus* and thus plays a role in bacterial virulence and colonization (Ji G et al., 1997, Srivastava, SK et al., 2014). These observations are summarized in Figure 4.17.

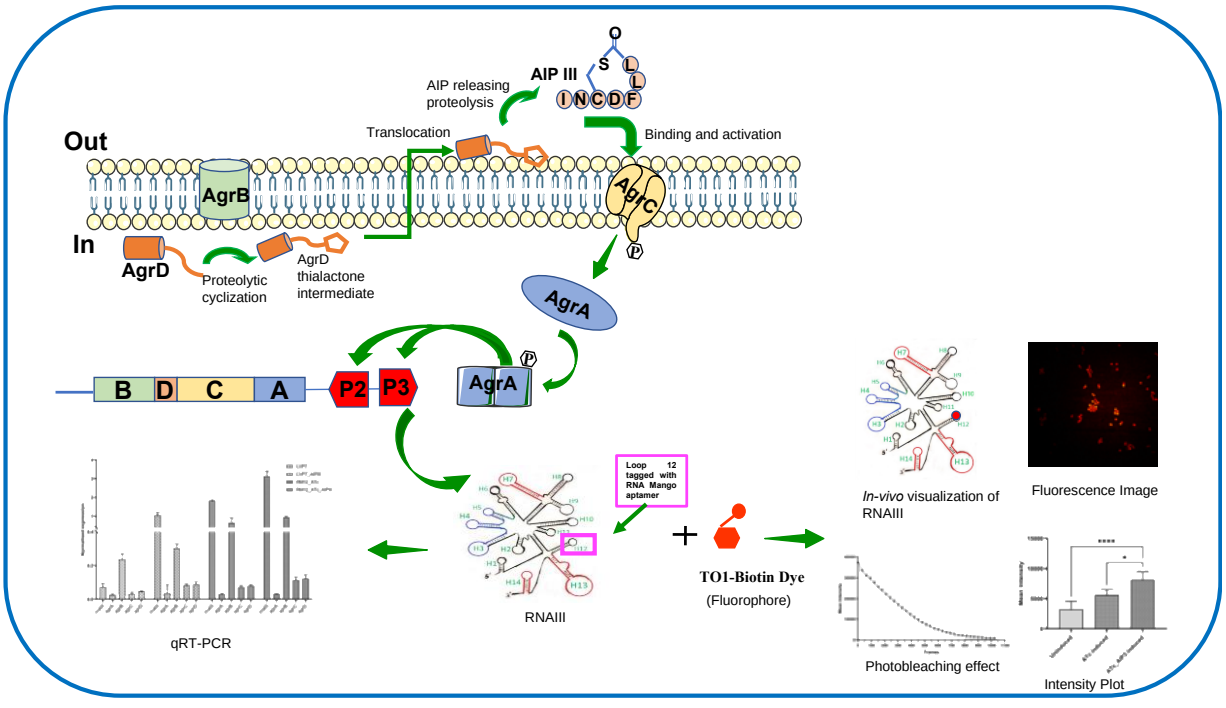


Figure 4.17: The role of RNAIII in quorum sensing mechanism

A schematic representation of the *agr* quorum sensing (QS) system is represented on top panel of the figure. The entire experimental results are depicted in the bottom panel

of the figure. The cell represents LVP-7 *S. aureus*. External QS stimulus AIPIII binds and activates AgrC (histidine kinase). Upon autophosphorylation (indicated by P) of AgrC, the phosphoryl group is then transferred to AgrA (response regulator) resulting in the dimerization of AgrA. This further autoactivates the *agr* operon by P2 promoter (red box). RNAIII is regulated by P3 promoter (red box). The loop 12 (H12) of RNAIII tagged with RNA Mango aptamer is shown in pink box. Aptamer binds to fluorophore (TO1-3PEG-Biotin) with a high affinity and AIPIII induced fluorescence activation upon RNAIII expression is visualised *in-vivo* in LVP-7 (fluorescence image and intensity plot). Photobleaching effect is depicted as an intensity plot. Fluorescence data was corroborated with qRT-PCR analysis (represented in the left panel). Secondary structure of RNAIII is adapted from Benito et al.,and redrawn (Benito et al., 2000).

Together, these studies reveal the potential of systematic hairpin loop replacement to understand structure-function relationships in the case of the non-coding RNA, RNAIII. This study could thus form the basis for a more detailed analysis by perturbing other interacting hairpin loops of RNAIII with an aptamer. This aspect would perhaps be aided by both knock out and complementation of RNAIII to understand the dynamics of this important effector molecule in phenotypic switching in *S.aureus.*

4.6 Statistical analysis

For data involving comparisons between multiple groups in case of biofilm analysis the normality test Shapiro-Wilk test followed by ordinary one-way ANOVA (parametric) and Tukey test was used for statistical comparisons. For mean intensity comparison between RM12 construct Uninduced and ATc induced, the Mann- Whitney Test (non-parametric), two tailed with confidence level 95% was used. For all other imaging analysis the normality test Shapiro-Wilk test followed by one-way ANOVA (non-parametric) and Dunn's test were employed. The significance levels are represented as * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$,**** $P \leq 0.0001$.

Chapter 5

Conclusions and scope for further studies

The work described in this thesis highlights the crucial role of RNA mediated intracellular signalling in pathogenic bacteria. A key ingredient to enable this analysis was a genotypically and phenotypically characterized model *S. aureus* strain. After multiple rounds of evaluation, we honed in on a particular strain *S. aureus* LVP-7 from our in-house collection. While the *agr* typing for this strain was performed earlier and specific molecular aspects had been evaluated in different context(s), the genomic features had not been reported thus far. This aspect was addressed in the first phase of this study.

The genomic features of community associated methicillin resistant *S. aureus*, LVP-7 exhibited an array of virulence factors. The draft genome assembly revealed several important genetic components of LVP-7. This train carries virulence encoding prophages like PhiNM3 and ΦSa2958PVL alongside mobile genetic elements like SCC*mec* and SaPI3 carrying *seb* toxin. Interestingly, it belongs to a clonal lineage ST88 which has remained a singleton. The *agr* locus of this ST88 clonal background belongs to type III *agr* group and most of the strains with the ST88 background are resistant to β-lactam antibiotics. Hence this association of virulence factors, antibiotic resistance and *agr* type with specific clonal lineages could potentially help to assess the severity of an infection caused by *S. aureus*. The enterotoxin *seb* on pathogenicity island SaPI3 is particularly noteworthy, as its superantigenicity has been reported in another clonal lineage, ST59. The draft genome of LVP-7 thus provided necessary data to facilitate this study (as well as those performed by other groups) to understand pathogenesis and virulence regulation in *S. aureus*.

Given that the mandate of the study was to understand the role of RNA mediated signaling, we also initiated an analysis of the RNome in *S. aureus* LVP-7. The first step towards this aspect of the study was RNA-sequencing using Illumina high-throughput RNA-sequencing technology. The expression profile of *S. aureus* LVP-7 was evaluated in the logarithmic growth phase. This transcriptome analysis identified several important abundant transcripts in the LVP-7 strain, which are known to play an important role in pathogenesis and virulence of *S. aureus.* Alongside, from Gene Ontology enrichment analysis, several genes corresponding to pathogenesis like the adhesion protein *spa* and genes like *sodA* and *sodM* that govern response to oxidative

stress were seen to have enhanced expression in *S. aureus* LVP-7. Comparative Gene Ontology enrichment analysis of LVP-7 with another epidemiologically important strain, USA300, showed higher enrichment in terms of genes corresponding to 'pathogenesis' and 'response to stress' sub-sets in LVP-7. Whether this finding is corroborated with other features specific to eye infection remains to be explored. Indeed, further investigation of the role of other differentially-regulated molecules is likely to provide new insights into the pathogenesis of *S. aureus* infection.

RNA structures can exhibit functionally relevant dynamic features. This feature has been well characterized across different RNA moieties, most notably, in RNA riboswitches. This aspect has been less explored in the case of long non-coding RNA. Towards the goal of understanding the conformational role of RNAIII and the functional repercussions, we used RNA aptamer insertions at specific positions of RNAIII to understand the relative roles of these segments. Two sites were chosen in RNAIII at the first instance and these are reported in chapter 4. While one site was functionally relevant, the other had a scaffolding role. A light up RNA aptamer was used to understand the cellular folding of RNAIII. Preliminary functional assays demonstrate the feasibility and utility of this approach. The intriguing features that have not been fully explored yet is the finding that mutation in the H3 hairpin loop affects the growth rate and morphology of LVP-7 strain. When this sequence-structure mapping of RNAIII is complete, alongside whole cell transcriptome and phenotype analysis, this data is likely to help understand the phenotypic switch of *S. aureus*.

Together, the work reported in this thesis establishes a *S. aureus* strain with well characterized genomic and phenotypic features. RNA-sequencing analysis appears to provide the baseline to evaluate sRNA mediated changes. This, in turn, would provide a basis to evaluate the role of a specific RNA, RNAIII, in a specific environmental context. The finding that synthetic aptamers could be effectively employed for non-coding RNA such as RNAIII in bacterial pathogens thus opens up the opportunity to understand the molecular basis for *S. aureus* pathogenesis.

6

Appendix

6.1 Appendix I

Table 6.1: List of qRT-PCR Primers

Primer Name	Primer Sequence (5' to 3')
MgrA RT F	TGCTCAAAGACAAGTTAATCGCT
MgrA RT R	CTGTACCAGTATCGAGTGCTAA
16S RT F	TGTCGTGAGATGTTGGG
16S RT R	CGATTCCAGCTTCATGT
agrA RT F	TGATAATCCTTATGAGGTGCTT
agrA RT R	CACTGTGACTCGTAACGAAAA
agrBRT F	TTTTGCAAGTACGATTAGGG
agrB RT R	ACACCAAAAAGATGAAGGTG
agrC RT F	TTTTACGTGCACAAGAAATG
agrC RT R	AATGCAACTCGAATGATAGG
agrD RT F	TCAACAAAGTTATTGAGCTTTT
agrD RT R	TTTTGGTACTTCAGCTTCGT
spa RT F	GCAAACGGCACTACTGCTGA
Spa RT R	CACCAGTTTCTGGTAATGCTTGA
psmA3 RT F	AGTGATGAGTTGTTGATCGT
psmA3 RTR	TCACATGGAATTCGTAGCAA
seb RT F	AGTGACTGCTCAAGAATTAGA
seb RT R	AGGCATCATGTCATACCAAA

6.2 Appendix II

6.2.1 Codes used in transcriptome analysis of LVP-7

6.2.1.1 FastQC and Trim Galore

```
fastqc *.gz or fastqc *.fastq
/TrimGalore-0.6.6/trim_galore LVP-7-
RNA_S6_L004_R1_001.fastq LVP-7-
RNA_S6_L004_R2_001.fastq -q 25 --paired --stringency
5
```

6.2.1.2 Genome Indexing

Creating a Reference genome index using STAR, for aligning to RNA-Seq data

```
STAR --runThreadN 1 --runMode genomeGenerate --
genomeDir star_out_prokka --genomeSAindexNbases 9
--genomeFastaFiles PROKKA_10152020.fna
```

Here, the input file is annotated nucleotide fasta from PROKKA. Essentially needs a **genome fasta** as input.

6.2.1.3 Alignment

Aligning to the genome which is the reference using STAR aligner, using the earlier created indexes

```
STAR  --runThreadN 8 --runMode alignReads --
limitBAMsortRAM 2147114089 --
sjdbGTFtagExonParentTranscript PROKKA_10152020_2.gff
--outFileNamePrefix LVP-7 --outSAMtype BAM Unsorted
SortedByCoordinate --genomeDir
/home/group_bioIT01/Sneha/IISC/Rna-Seq-
IISC/star_out_prokka --readFilesIn LVP-7-
RNA_S6_L004_R1_001_val_1.fq LVP-7-
RNA_S6_L004_R2_001_val_2.fq
```

Here, the input files are **trimmed fastq** files, along with index directory and **gff annotation** file.

6.2.1.4 Read Quantification

Raw read count is calculated using Htseq as follows

```
/home/group_bioIT01/miniconda3/bin/htseq-count   LVP-
7Aligned.sortedByCoord.out.bam
/home/group_bioIT01/Sneha/IISC/Rna-Seq-
IISC/PROKKA_10152020_2.gff --idattr=ID -t gene -f bam
> htseq_LVP-7.txt
```

Here, the input files are **bam alignment** output files as well as **gff annotation** files

6.2.1.5 TPM Normalisation

Following is the method for calculating TPM:

1. Divide the read counts by the length of each gene in kilobases. This gives reads per kilobase (RPK).
2. Count up all the RPK values in a sample and divide this number by 1,000,000. This is called “per million” scaling factor.
3. Divide the RPK values by the “per million” scaling factor. This gives the final TPM.

The R code used to calculate the same are given below:

In R console:

```
file_len<-
read.delim("gen_len.txt",header=F,sep="\t")
file_count<-
read.delim("count.txt",header=F,sep="\t")
colnames(file_len)<- c("GeneName","Len")
colnames(file_count)<- c("GeneName","Count")
```

```
file_count<-file_count[                    !grepl("__",
file_count$GeneName) ,]
#total_count<- sum(file_count$Count)
oneB<-1e6
finallist <- merge(file_len,file_count,by="GeneName")
finallist$div<-0
finallist$div<- (finallist$Count)/(finallist$Len)
total_count<- sum(finallist$div)
finallist$TPM<-0
finallist[,2:4]      <-      (sapply(finallist[,2:4],
as.double))
finallist$TPM<- (oneB*finallist$div)/(total_count)
write.table(finallist,file="tpm1.txt",sep="\t", col.names = T, row.names = F)
```

Gene length files can be obtained from gff annotation files and raw count output from htseq as input.

6.2.1.6 Quartile Binning

The genes can be ranked based on highest to lowest TPM values and divided into quartiles (quarters of more-or-less equal size), with Q1 containing the most highly expressed genes (top 25%) and Q4 containing genes with the lowest expression (bottom 25%).

6.3 Appendix III

Table 6.2: RNAIII and RNA Mango aptamer cloning primers

Primer Name	Primer Sequence (5' to 3')
pRMC2 RNAIII KpnI F	GGGTACCTAACTAGATCACAGAGATGTGATGGAAAATAGTTGATG A
pRMC2 RNAIII EcoR1 R	TGTGAATTCGCCGCGAGCTTGGGAGGGGCTC
Apt 12 F	AACAACATCGATTTATCATACGAAGGGACGGTGCGGAGAGGAGAG TATGATAAATAAAATTTTTT
Apt 12 R	AAAAAATTTTATTTATCATACTCTCCTCTCCGCACCGTCCCTTCGTA TGATAAATCGATGTTGTT
Apt 3 F	TAAGGAAGGAGTGATTTCTACGAAGGGACGGTGCGGAGAGGAGAG TAGATATCATTTCAACAATC
Apt 3 R	GATTGTTGAAATGATATCTACTCTCCTCTCCGCACCGTCCCTTCGTA GAAATCACTCCTTCCTTA
RNAIII RT NF	GTGATGGAAAATAGTTGATGAGTTGTTT
RNAIII RT NR	GAATTTGTTCACTGTGTCGATAATCC

6.3.1 Electrocompetent cell preparation:

1. Briefly, the overnight grown cells in Tryptic soya broth (TSB, HiMedia, Inc) at 37°C were diluted to 0.5 (O.D 600 nm) in 50 ml prewarmed TSB.
2. Cells were re incubated at 37°C for 30min. Measured the OD600nm after 30min (generally it will be between 0.8 to 0.9 O.D).
3. The cells were kept in an ice slurry and this was maintained throughout the preparation.
4. The cells were harvested in a centrifuge (used Kubota) at 3,900xg for 10 min at 4°C.
5. Supernatant was discarded and washed the pellet in 50 ml of ice cold autoclaved milliQ water twice.
6. Discarded the supernatant and cells were resuspended in ice cold 10% (w/v) glycerol. Made up to 10 ml
7. Harvested the cells again at 3,900xg for 10 min at 4°C and the pellet was resuspended in ice cold 10% (w/v) glycerol. Made up to 2 ml.
8. Again cells were harvested at 3,900xg for 10 min at 4°C and the pellet was resuspended in 250µl ice cold 10% (w/v) glycerol using a pipette.
9. Aliquoted 50µl of the cells and frozen at -80°C.

6.3.2 Electroporation:

1. The electrocompetent cells were thawed on ice for 5min and then placed at room temperature for 5 min.
2. The cells were centrifuged at 5,000xg for 1 min. Discarded the supernatant and the cells were resuspended in 50µl of 10% glycerol and 500 nM Sucrose.
3. Up to 5µg of plasmid DNA was gently mixed with the electrocompetent cells and added to 0.1 cm electroporation cuvette (Bio-Rad).
4. The cuvette was placed in an electroporator (Bio-Rad) and pulsed at 21kv/cm, and 25µF. Time constant of about 2.0-2.4 ms.
5. Immediately 1 ml of TSB containing 500 mM sucrose was added and resuspended using a pipette. Transferred to an Eppendorf and incubated at 37°C for 2 hours. The incubated cells were pelleted down and the pellet was dissolved in 100 µl of TSB containing 500 mM sucrose. Finally the cells were plated on to LB containing 10µg/ml chloramphenicol.
6. The plate was incubated at 37°C for 48 hours.

6.3.3 Slide preparation for Fluorescence microscopy:

1. Slides and coverslips were soaked in acetone for 15 min and then in 1 M KOH for 20 mins.
2. Coverslips were sonicated for 20 mins in 1M KOH.
3. Finally slide and coverslip were rinsed in MilliQ H_2O.
4. They were air dried and stored.

6.3.4 Allelic exchange using pIMAY vector:

In order to generate RNAIII deletion constructs in *S. aureus* LVP-7, following protocol was tried (Monk IR et al., 2012).

Upstream and downstream sequence of RNAIII were amplified separately using the primers A/B and C/D respectively. The PCR conditions used were as follows -Initial denaturation at 95°C for 3 min and 30 cycles with denaturation at 95°C for 30 sec, annealing at 55°C for 30 sec, extension at 72°C for 1 min and final extension at 72°C for 5 min. PCR products were purified using PCR purification kit (Favorgen. Inc). Further, the AB and CD products were mixed (100 ng each) and subjected to 2 cycles of PCR at 55°C annealing temperature. Then, 250 nM of A and D primers were added

and 30 cycles PCR with an annealing temperature of 60°C was performed to finally achieve the product of interest (AD) Further overlap extension PCR was performed to amplify AD products using A/D primers (Figure 6.1A and Figure 6.1B). A and D primers contain the endonuclease sites (*KpnI* and *EcoR1*). The cleaved products were ligated into pIMAY vector (Addgene) cut with the same enzymes. Successful ligation was not seen and further optimisation of the protocol is required.

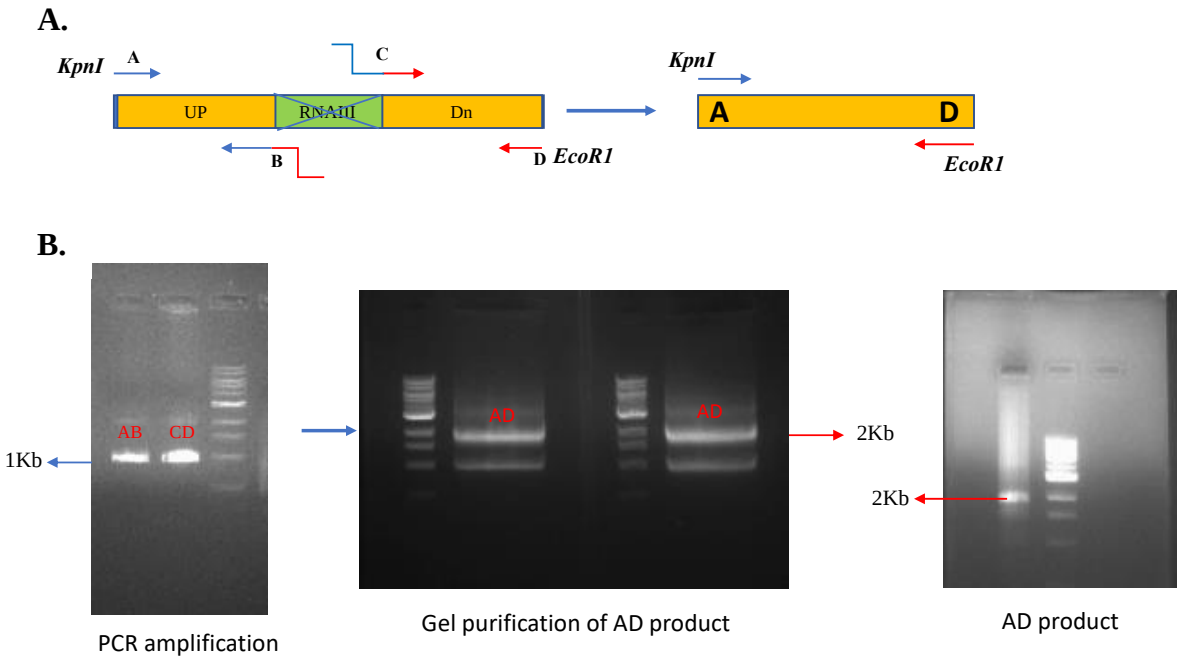


Figure 6.1: Gene Knock out strategy. 1A. Schematic representation of the overlap extension PCR. **1B.** Corresponding agarose gel figures of the PCR performed.

Primers used in this section:

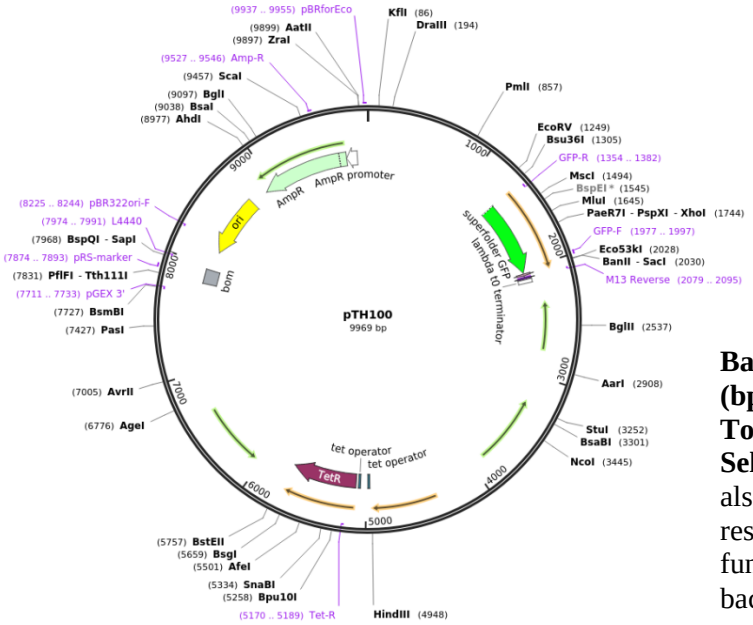
LVP7_RNAIIIKO_A_KpnI - 5′ GCGGTACCGAGATAGTTCTAAAAATGAAAC 3′
LVP7_RNAIIIKO_B - 5′ CTTTTTTATTGTAATAAAGAAGGGGAGAGTTAA 3′
LVP7_RNAIIIKO_C - 5′ TATTAAAACATGCTAAAAGCATTTATTTTCGATT 3′
LVP7_RNAIIIKO_D_EcoR1- 5′GCAATTCATAAAGAACATTAATGAAACTTGAACAAT 3′

Reference: Monk IR, Shah IM, Xu M, Tan MW, Foster TJ. Transforming the untransformable: application of direct transformation to manipulate genetically Staphylococcus aureus and Staphylococcus epidermidis. mBio. 2012 Mar 20;3(2):e00277-11. doi: 10.1128/mBio.00277-11. PMID: 22434850; PMCID: PMC3312211.

6.4 Appendix IV

6.4.1 To understand the Macrophage/Neutrophil killing of *S. aureus*/*S. epidermidis* using reporter plasmids

In order to monitor the killing of *S. aureus* or *S. epidermidis* by neutrophil or macrophages by flowcytometry, we generated strains of both the species with sGFP integrated into the chromosome. We used a sGFP carrying plasmid, pTH100 (Addgene). The methodology was followed as given in de Jong NW et, 2017.



Backbone size w/o insert (bp)8815
Total vector size (bp)9969
Selectable markers
also contains Chloramphenicol resistance gene, but only functional in Gram-postive bacteria.

Figure 6.2: Vector map of pTH100

The plasmid carries a chloramphenicol resistance gene.

6.4.2 Transformation and electroporation of pTH100 into *S. epidermidis* (ATCC12228) and *S. aureus* (LVP-7)

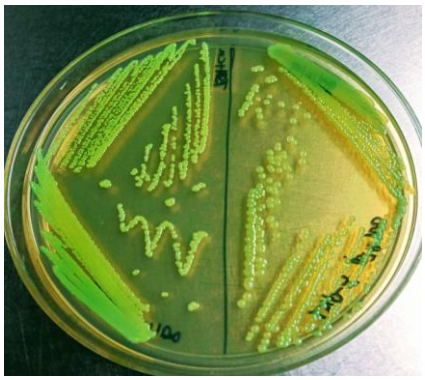
1. Plasmid pTH100 was transformed into *E. coli,* DH5α cells. Plated on to LB plate containing 10 µg/ml chloramphenicol.
2. Extracted the plasmid DNA from these cells and transformed into DC10B cells (Δdcm strain)
3. Plasmid extracted from DC10B (5 µg DNA) cells were electroporated into *S. aureus*, LVP-7 and *S. epidermidis*, ATCC12228 and plated on to Todd-Hewitt

agar media containing 10ug/ml chloramphenicol and incubated at 30°C for 24-48 hrs. A detailed chromosomal integration protocol is given below.

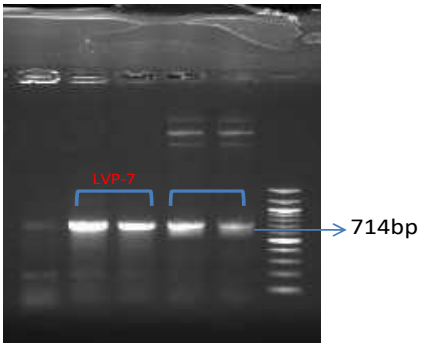
6.4.3 Chromosomal integration of fluorescent reporters into *S. epidermidis* and *S. aureus*

1. Post electroporation the Todd-Hewitt agar containing 10 μg/ml chloramphenicol plates were incubated at 30°C for 24-48 hrs.
2. The fluorescent colonies were picked and plated on a fresh plate (THA+cm) and incubated at 45°C overnight for the selection of integration of pTH100 into the chromosome.
3. Single recombinants (large colonies) were selected and re-streaked on to TH media without antibiotic and plates were incubated at 30°C overnight.
4. Passaging of the cultures were done during 7 hrs (during day time), re-diluted and grown for 16 hrs overnight. Approximately, 5-7 culture dilution cycles to allow for double cross over events to occur.
5. Further, the culture was plated onto TH agar containing 100 ng/ml of anhydrotetracycline (ATc) and incubated overnight at 37°C to select for double cross over mutants.
6. Single colonies were plated on to antibiotic containing TH plate (TH+AB) and without antibiotic plate (TH-AB). Plates were incubated at 37°C.
7. Non-fluorescent colonies from TH-AB plates were further analysed by PCR for integration.
8. Green colonies were further screened for the presence of *gfp* gene using forward primer (sGFP Fw) and reverse primer (sGFP Rv) by PCR.
9. Positive *gfp* PCR colony was processed for genomic DNA extraction.
10. Further genomic DNA was used as template for PCR with forward primer (NWMN29end Fw) and reverse primer (NWMN30end Rv) and confirmed for GFP integration into the chromosome
11. *S. aureus* and *S. epidermidis* cells containing chromosomally integrated sGFP were used for Macrophage and Neutrophil infection and killing assays. Monitored the fluorescence with flowcytometry.

A.



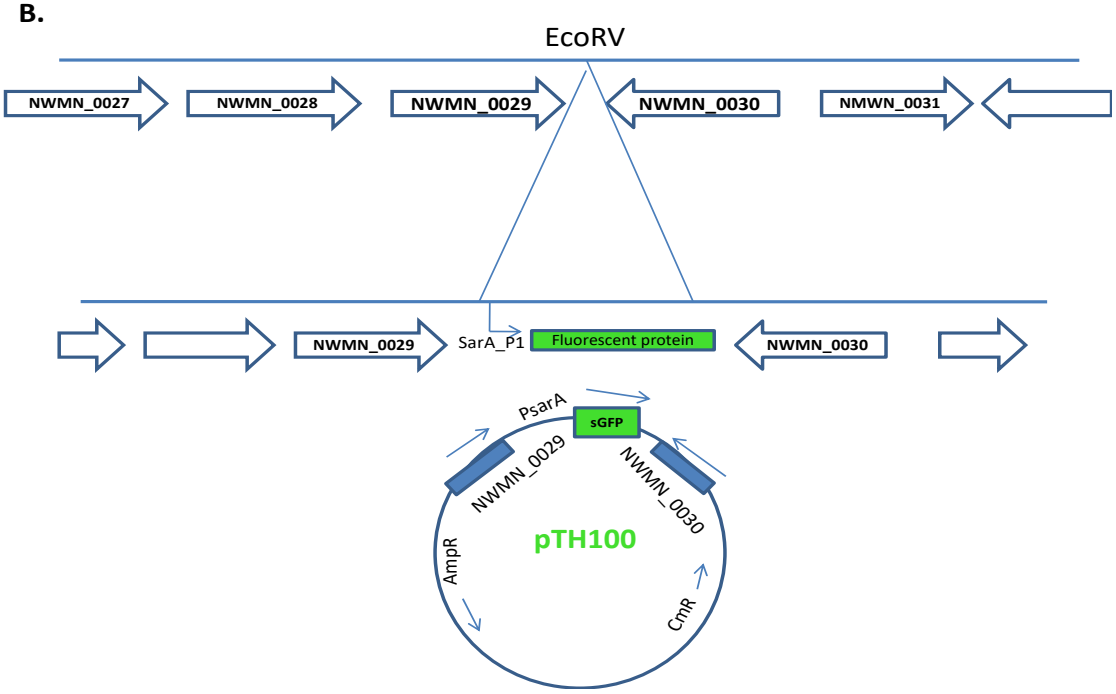
S. epidermidis ATCC 12228



S. aureus LVP-7

Confirmation of GFP by PCR

B.



C.

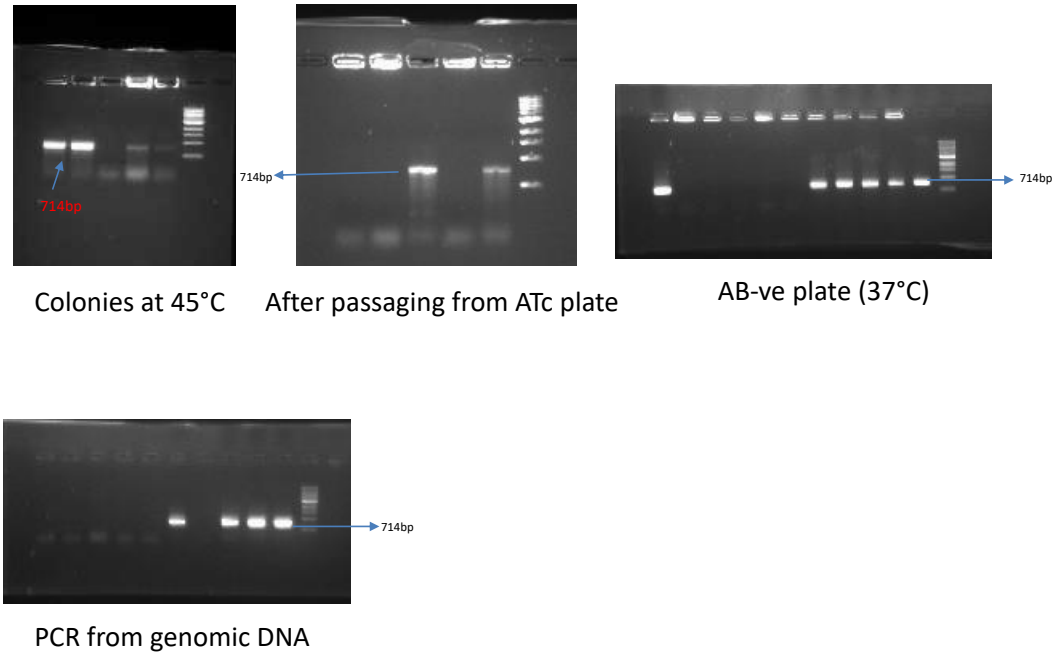


Figure 6.3: Chromosomal integration of fluorescent reporters into *S. epidermidis* and *S. aureus*

A. GFP positive colonies on THB plate. GFP PCR amplification **B.** Schematic representation of integration of super folder green fluorescent protein (sGFP). Genomic region sites represented in *S. aureus* strain Newman where the fluorescent constructs are integrated **C.** GFP PCRs at intermittent step of chromosomal integration.

Primers used in this study:
sGFP Fw: 5′ TGTCTGCAGCGATGAGCAAAGGAGAAGAACTTTTC 3′
sGFP Rv: 5′ TGTAAGCTTTTAGTGGTGGTGGTGGTGCG 3′
NWMN29end Fw: 5′ TATGTCACTTATCCTTTTGGAAA 3′
NWMN30end Rv: 5′ CATAATGTGTTGTAAACATTTTTTTTG 3′

Reference:

de Jong NW, van der Horst T, van Strijp JA, Nijland R. Fluorescent reporters for markerless genomic integration in Staphylococcus aureus. Sci Rep. 2017 Mar 7;7:43889. doi: 10.1038/srep43889. PMID: 28266573; PMCID: PMC5339689.

6.5 Appendix VI

The Table of all TPM is documented in the attached document.

7 References

1. Abu-Qatouseh L. F., Seggewiss J., Chinni S. V., Brosius J., Rozhdestvensky T. S., Peters G., et al. (2007). RNomics: experimental identification of novel small nonprotein-coding RNAs in *Staphylococcus aureus*. *Int. J. Med. Microbiol.* 297, 106–106. 10.1093/nar/gkj469

2. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol* 20:224.

3. Anders S , Pyl PT , Huber W . 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, Volume 31, Issue 2, 15 January 2015, Pages 166–169, https://doi.org/10.1093/bioinformatics/btu638.

4. Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

5. Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C, Wingett S. 2010. FastQC: a quality control tool for high throughput sequence data. *https://www.bioinformatics.babraham.ac.uk/projects/fastqc*.

6. Arakere G, Nadig S, Swedberg G, Macaden R, Amarnath SK, and Raghunath D. Genotyping of Methicillin-Resistant *Staphylococcus aureus* Strains from Two Hospitals in Bangalore, South India. *J. Clin. Microbiol*. 2005.

7. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nat Genet*. 2000 May;25(1):25-9. doi: 10.1038/75556. PMID: 10802651; PMCID: PMC3037419.

8. Babendure JR, Adams SR & Tsien RY. Aptamers switch on fluorescence of triphenylmethane dyes. *J. Am. Chem. Soc.* 125, 14716–14717 (2003).

9. Bae JS, Da F, Liu R, He L, Lv H, Fisher EL, Rajagopalan G, Li M, Cheung GYC, Otto M. 2020. Staphylococcal enterotoxin B contributes to *Staphylococcus*

aureus systemic infection. *J Infec.Dis*. doi: 10.1093/infdis/jiaa584. Online ahead of print.

10. Bae T, Baba T, Hiramatsu K and Schneewind O. Prophages of *Staphylococcus aureus* Newman and their contribution to virulence. 2006. *Molecular Microbiology*. doi:10.1111/j.1365-2958.2006.05441.

11. Balaban N, Novick RP. Translation of RNAIII, the *Staphylococcus aureus* agr regulatory RNA molecule, can be activated by a 3'-end deletion. *FEMS Microbiol Lett*. 1995 Nov 1;133(1-2):155-61. doi: 10.1111/j.1574-6968.1995.tb07877.x. PMID: 8566701.

12. Balaban N. Isolation of agr quorum sensing autoinducers. *Methods Mol Biol*. 2011;692:47-59. doi: 10.1007/978-1-60761-971-0_4. PMID: 21031303.

13. Balasubramanian D, Harper L, Shopsin B, Torres VJ. 2017. *Staphylococcus aureus* pathogenesis in diverse host environments. *Pathog Dis* 75. https://doi.org/10.1093/femspd/ftx005.

14. Ballal A, Manna AC. Regulation of superoxide dismutase (sod) genes by SarA in *Staphylococcus aureus*. *J Bacteriol* 2009;191:3301–10.

15. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455-477. doi:10.1089/cmb.2012.0021

16. Bartels MD, Petersen A, Worning P, Nielsen JB, Larner-Svensson H, Johansen HK, Andersen LP, Jarløv JO, Boye K, Larsen AR, Westh H. Comparing whole-genome sequencing with Sanger sequencing for spa typing of methicillin-resistant *Staphylococcus aureus*. *J. Clin. Microbiol*. 2014. 52(12): 4305-8.

17. Behlau I, Gilmore MS. 2008. Microbial biofilms in ophthalmology and infectious disease. *Arch Ophthamol*. **126**(11):1572–1581.

18. Benito Y, Kolb FA, Romby P, Lina G, Etienne J, et al. (2000) Probing the structure of RNAIII, the *Staphylococcus aureus* agr regulatory RNA, and identification of the RNA domain involved in repression of protein A expression. *RNA* 6: 668–679.

19. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics*. 2013 Mar

1;29(5):661-3. doi: 10.1093/bioinformatics/btt019. Epub 2013 Jan 16. PMID: 23325622; PMCID: PMC3582273.

20. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman W-H, Pagès F, Trajanoski Z, Galon J, ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, *Bioinformatics*, Volume 25, Issue 8, 15 April 2009, Pages 1091–1093, https://doi.org/10.1093/bioinformatics/btp101

21. Bohn C, Rigoulay C, Chabelskaya S, Sharma CM, Marchais A, Skorski P, Borezée-Durant E, Barbet R, Jacquet E, Jacq A, Gautheret D, Felden B, Vogel J, Bouloc P. Experimental discovery of small RNAs in *Staphylococcus aureus* reveals a riboregulator of central metabolism. *Nucleic Acids Res*. 2010 Oct;38(19):6620-36. doi: 10.1093/nar/gkq462. Epub 2010 May 28. PMID: 20511587; PMCID: PMC2965222.

22. Boisset S, Geissmann T, Huntzinger E, Fechter P, Bendridi N, et al. *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev.* 2007;21:1353–1366.

23. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* ;30(15):2114-2120. doi:10.1093/bioinformatics/btu170.

24. Bouhedda F, Autour A, Ryckelynck M, Bouhedda F,Autour A, Ryckelynck M (2017) Light-up RNA aptamers and their cognate fluorogens: from their development to their applications. *Int J Mol Sci* 19:44

25. Cawte, A.D., Unrau, P.J. & Rueda, D.S. Live cell imaging of single RNA molecules with fluorogenic Mango II arrays. *Nat Commun* 11**,** 1283 (2020). https://doi.org/10.1038/s41467-020-14932-7

26. Chabelskaya S, Bordeau V, Felden B. Dual RNA regulatory control of a *Staphylococcus aureus* virulence factor. *Nucleic Acids Res*. 2014 Apr;42(8):4847-58. doi: 10.1093/nar/gku119. Epub 2014 Feb 7. PMID: 24510101; PMCID: PMC4005698.

27. Corrigan RM, Foster TJ. An improved tetracycline-inducible expression vector for *Staphylococcus aureus*. *Plasmid*. 2009 Mar;61(2):126-9. doi: 10.1016/j.plasmid.2008.10.001. Epub 2008 Nov 25. PMID: 18996145.

28. Dolgosheina EV, CY Jeng, Panchapakesan SSS, Cojocaru R, Chen PSK, Wilson PD, Hawkins N, Wiggins PA and Unrau PJ. RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking. *ACS Chemical Biology* 2014 9 (10), 2412-2420 DOI: 10.1021/cb500499x

29. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25:119–120.

30. Condon C, Bechhofer DH. Regulated RNA stability in the Gram positives. *Curr Opin Microbiol.* 2011; 14:148-54; PMID:21334965; http://dx.doi.org/10.1016/j.mib.2011.01.010.

31. Carroll RK, Weiss A, Broach WH, Wiemels RE, Mogen AB, Rice KC, Shaw LN. 2016. Genome-wide annotation, identification, and global transcriptomic analysis of regulatory or small RNA gene expression in *Staphylococcus aureus*. *mBio* 7(1):e01990-15. doi:10.1128/mBio.01990-15

32. Chakrakodi B, Prabhakara S, Nagaraj S, Etienne J, Arakere G. 2014. High prevalence of ciprofloxacin resistance in community associated Staphylo- coccus aureus in a tertiary care Indian hospital. *Adv Microbiol* 4:133–141. https://doi.org/10.4236/aim.2014.42018.

33. Chen PR, Bae T, Williams WA et al. An oxidation-sensing mechanism is used by the global regulator MgrA in *Staphylococcus aureus*. *Nat Chem Biol* 2006 ;2:591–5.

34. Chevalier C, Huntzinger E, Fechter P, Boisset S, Vandenesch F, Romby P, et al. *Staphylococcus aureus* endoribonuclease III purification and properties. *Methods Enzymol* 2008; 447:309-27; PMID:19161850; http://dx.doi.org/10.1016/S0076-6879(08)02216-7.

35. Clauditz A, Resch A, Wieland KP, Peschel A, Gotz F. Staphyloxanthin plays a role in the fitness of *Staphylococcus aureus* and its ability to cope with oxidative stress. *Infect Immu*. 2006. Aug:74(8): 4950-3

36. Cosgrove K, Coutts G, Jonsson IM et al. Catalase (KatA) and alkyl hydroperoxide reductase (AhpC) have compensatory roles in peroxide stress resistance and are required for survival, persistence, and nasal colonization in *Staphylococcus aureus*. *J Bacteriol* 2007;189:1025–35.

37. David MZ and Daum RS. Community-Associated Methicillin-Resistant *Staphylococcus aureus*: Epidemiology and Clinical Consequences of an Emerging Epidemic. 2010. *CLINICAL MICROBIOLOGY REVIEWS*, 616–687. doi:10.1128/CMR.00081-09

38. Day NP, Moore CE, Enright MC, Berendt AR, Smith JM, Murphy MF, Peacock SJ, Spratt BG, Feil EJ. A link between virulence and ecological abundance in natural populations of *Staphylococcus aureus*. *Science*. 2001 Apr 6;292(5514):114-6. doi: 10.1126/science.1056495.

39. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, García-Cobos S, Kooistra-Smid AMD, Raangs EC, Rosema S, Veloo ACM, Zhou K, Friedrich AW, Rossen JWA, Application of next generation sequencing in clinical microbiology and infection prevention, *Journal of Biotechnology*, Volume 243,2017,Pages 16-24, https://doi.org/10.1016/j.jbiotec.2016.12.022.

40. Diep BA, Gill SR, Chang RF, Phan TH, Chen JH, Davidson MG, Lin F, Lin J, Carleton HA, Mongodin EF, Sensabaugh GF, Perdreau-Remington F. Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant *Staphylococcus aureus*. *Lancet*. 2006 Mar 4;367(9512):731-9. doi: 10.1016/S0140-6736(06)68231-7. PMID: 16517273.

41. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*,1;29(1):15-21.doi: 10.1093/bioinformatics/bts635. Epub 2012 Oct 25

42. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J Proteome Res*. 2019 Feb 1;18(2):623-632. doi: 10.1021/acs.jproteome.8b00702. Epub 2018 Dec 5. PMID: 30450911; PMCID: PMC6800166.

43. Drenkard E. Antimicrobial resistance of Pseudomonas aeruginosa biofilms. *Microbes Infect* 5(13): 1213-9, 2003.

44. D'souza N, Rodrigues C, Mehta A. Molecular characterization of methicillin resistant *Staphylococcus aureus* with emergence of epidemic clones of sequence type (ST) 22 and ST 772 in Mumbai, India. *J Clin Microbiol* 2010; 48:1806–11

45. Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* 346:818–822.

46. Engebrecht J, Nealson K, Silverman M. 1983. *Bacterial bioluminescence*: isolation and genetic analysis of functions from Vibrio fischeri. *Cell* 32:773–81

47. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG. 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol*. 38(3):1008-15.doi: 10.1128/JCM.38.3.1008-1015

48. Enterina JR, Wu L, and Campbell RE. (2015). Emerging fluorescent protein technologies. *Curr. Opin. Chem. Biol.* 27, 10–17. doi: 10.1016/j.cbpa.2015.05.001

49. Eyraud A, Tattevin P, Chabelskaya S, Felden B. (2014). A small RNA controls a protein regulator involved in antibiotic resistance in *Staphylococcus aureus*. *Nucleic Acids Res.* 42, 4892–4905. 10.1093/nar/gku149

50. Felden B, Vandenesch F, Bouloc P, Romby P (2011). The *Staphylococcus aureus* RNome and its commitment to virulence. *PLoS Pathog* 7(3):e1002006

51. Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of single RNA transcripts in situ. *Science*. 1998 Apr 24;280(5363):585-90. doi: 10.1126/science.280.5363.585. PMID: 9554849.

52. Feng Y, Chen CJ, Su LH, Hu S, Yu J, Chiu CH. Evolution and pathogenesis of *Staphylococcus aureus*: lessons learned from genotyping and comparative genomics. *FEMS Microbiol Rev*. 2008 Jan;32(1):23-37. doi: 10.1111/j.1574-6976.2007.00086.x. Epub 2007 Nov 5. PMID: 17983441.

53. Filonov GS, Moon JD, Svensen N & Jaffrey SR. Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J. Am. Chem. Soc.* 136, 16299–16308 (2014).

54. Francisco AP, Bugalho M, Ramirez M et al. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* 10, 152 (2009). https://doi.org/10.1186/1471-2105-10-152

55. Frenay HM, Bunschoten AE, Schouls LM, van Leeuwen WJ, VandenbrouckeGrauls CM, Verhoef J, et al. Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. *Eur J Clin Microbiol Infect Dis* 1996;15:60–4.

56. Fries BC, Goldman DL, Casadevall A. Phenotypic switching in Cryptococcus neoformans. *Microbes Infect* 4(13): 1345-52, 2002.

57. Fusco D, Accornero N, Lavoie B, Shenoy SM, Blanchard JM, Singer RH, et al. (2003). Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Curr. Biol.* 13, 161–167. doi: 10.1016/S0960- 9822(02)01436-7

58. Garzoni C, Kelley WL. Return of the Trojan horse: intracellular phenotype switching and immune evasion by *Staphylococcus aureus*. *EMBO Mol Med.* 2011 Mar;3(3):115-7. doi: 10.1002/emmm.201100123. Epub 2011 Feb 25. PMID: 21365763; PMCID: PMC3395108.

59. Geissmann T, Chevalier C, Cros M-J, Boisset S, Fechter P, Noirot C, Schrenzel J, François P, Vandenesch F, Gaspin C, Romby P, 2009. A search for small noncoding RNAs in *Staphylococcus aureus* reveals a conserved sequence motif for regulation. *Nucleic Acids Res.* 37, 7239–7257.

60. Ghebremedhin B, Olugbosi MO, Raji AM, Layer F, Bakare RA, Konig B, Konig W. 2009. Emergence of a community-associated methicillin-resist- ant *Staphylococcus aureus* strain with a unique resistance profile in South- west Nigeria. *J Clin Microbiol* 47:2975–2980. https://doi.org/10.1128/JCM .00648-09.

61. Gotrik M, Sekhon G, Saurabh S, Nakamoto M, Eisenstein M, Soh HT. Direct Selection of Fluorescence-Enhancing RNA Aptamers. *J Am Chem Soc*. 2018 Mar 14;140(10):3583-3591. doi: 10.1021/jacs.7b10724. Epub 2018 Mar 5. PMID: 29505267.

62. Gupta RK, Luong TT, & Lee CY (2015). RNAIII of the *Staphylococcus aureu*s agr system activates global regulator MgrA by stabilizing mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(45), 14036–14041. https://doi.org/10.1073/pnas.1509251112

63. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072-1075. doi:10.1093/bioinformatics/btt086

64. Haaber, J, Penadés, J.R, Ingmer, H. Transfer of antibiotic resistance in *Staphylococcus aureus*. *Trends Microbiol*. 2017, 25, 893–905.

65. Hanssen AM, Ericson Sollid JU. SCCmec in staphylococci: genes on the move. *FEMS Immunol Med Microbiol*. 2006 Feb;46(1):8-20. doi: 10.1111/j.1574-695X.2005.00009.x. PMID: 16420592.

66. Heilmann C, Schweitzer O, Gerke C, Vanittanakom N, Mack D and G€otz F. (1996) Molecular basis of intercellular adhesion in the biofilm-forming *Staphylococcus epidermidis*. *Mol. Microbiol*., 20, 1083–1091.

67. Heilmann C, and Gotz F. Cell-Cell Communication and Biofilm Formation in Gram-Positive bacteria. *Bacterial Signaling*, 2009. https://doi.org/10.1002/9783527629237.ch1

68. Hiramatsu K, Cui L, Kuroda M & Ito T (2001) The emergence and evolution of methicillin-resistant *Staphylococcus aureus*. *Trends Microbiol* 9: 486–493.

69. Holtfreter S, Grumann D, Schmudde M, Nguyen HTT, Eichler P, Strommenger B, Kopron K, Kolata J, Giedrys-Kalemba S, Steinmetz I, Witte W and Bröker BM . J Clin Microbiol. 2007 Aug; 45(8): 2669–2680 Clonal Distribution of Superantigen Genes in Clinical *Staphylococcus aureus* Isolates

70. Huang T-W, Chen F-J, Miu W-C, Liao T-L, Lin A-C, Huang I-W, Wu K-M, Tsai S-F, Chen Y-T, Lauderdale T-LY. 2012. Complete genome sequence of *Staphylococcus aureus* M013, a pvl-positive, ST59-SCCmec type V strain isolated in Taiwan. *J Bacteriol* 194:1256–1257. https://doi.org/10.1128/JB .06666-11.

71. International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC). 2009. Classification of staphylococcal

cassette chromosome mec (SCCmec): guidelines for reporting novel SCCmec elements. Antimicrob. Agents Chemother. 53:4961–4967.

72. Ito T, Ma XX, Takeuchi F, Okuma K, Yuzawa H, Hiramatsu K. 2004. Novel type V staphylococcal cassette chromosome mec driven by a novel cassette chromosome recombinase, ccrC. *Antimicrob Agents Chemother*. 48(7):2637-51.

73. Jain A, Agarwal A. Biofilm production, a marker of pathogenic potential of colonizing and commensal staphylococci. *J Microbiol Methods*. 2009 Jan;76(1):88-92. doi: 10.1016/j.mimet.2008.09.017. Epub 2008 Sep 30. PMID: 18851996.

74. Jarraud S, Mougel C, Thioulouse J, Lina G, Meugnier H, Forey F, Nesme X, Etienne J, Vandenesch F. Relationships between *Staphylococcus aureus* genetic background, virulence factors, agr groups (alleles), and human disease. *Infect Immun*. 2002 Feb;70(2):631-41. doi: 10.1128/IAI.70.2.631-641.2002. PMID: 11796592; PMCID: PMC127674.

75. Jensen SO, Lyon BR. Genetics of antimicrobial resistance in S*taphylococcus aureus*. *Future Microbiol*. 2009, 4, 565–582

76. Ji G, Beavis R, Novick RP. Bacterial interference caused by autoinducing peptide variants. Science. 1997 Jun 27;276(5321):2027-30. doi: 10.1126/science.276.5321.2027. PMID: 9197262.

77. Johnson WL, Sohn MB, Taffner S, Chatterjee P, Dunman PM, Pecora N, et al. (2021) Genomics of *Staphylococcus aureus* ocular isolates. PLoS ONE 16(5): e0250975. https://doi.org/10.1371/journal.pone.0250975Lewis K. Persister cells. *Annu Rev Microbiol* 64(357-72, 2010.

78. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res*. 2018 Sep 24;3:124. doi: 10.12688/wellcomeopenres.14826.1. PMID: 30345391; PMCID: PMC6192448.

79. Kalkum M, Lyon GJ, Chait BT. Detection of secreted peptides by using hypothesis-driven multistage mass spectrometry. *Proc Natl Acad Sci U S A*. 2003 Mar 4;100(5):2795-800. doi: 10.1073/pnas.0436605100. Epub 2003 Feb 18. PMID: 12591958; PMCID: PMC151420.

80. Kanangat S, Postlethwaite A, Cholera S, Williams L, Schaberg D. Modulation of virulence gene expression in Staphylococcus aureus by interleukin-1β: Novel implications in bacterial pathogenesis. Microbes and Infection. 2007; 9 (3): 408-415. DOI: 10.1016/j.micinf.2006.12.018 PMID: 17307379

81. Kareb O, Aïder M. Quorum Sensing Circuits in the Communicating Mechanisms of Bacteria and Its Implication in the Biosynthesis of Bacteriocins by Lactic Acid Bacteria: a Review. *Probiotics Antimicrob Proteins*. 2020 Mar;12(1):5-17. doi: 10.1007/s12602-019-09555-4. PMID: 31104210.

82. Kashif A, McClure J-A, Lakhundi S, Pham M, Chen S, Conly JM, Zhang K. *Staphylococcus aure*us ST398 Virulence Is Associated With Factors Carried on Prophage ϕSa3 JOURNAL=Frontiers in Microbiology. VOLUME=10 YEAR=2019PAGES=2219 URL=https://www.frontiersin.org/article/10.3389/fmicb.2019.02219

83. Kavanaugh JS, Horswill AR. 2016. Impact of environmental cues on staphylococcal quorum sensing and biofilm development. *J Biol Chem* 291:12556 –12564. https://doi.org/10.1074/jbc.R116.722710.

84. König J, Zarnack K, Luscombe NM, Ule J. Protein-RNA interactions: new genomic technologies and perspectives. *Nat Rev Genet.* 2012 Jan 18;13(2):77-83. doi: 10.1038/nrg3141. Erratum in: Nat Rev Genet. 2012 Mar;13(3):220. PMID: 22251872.

85. Kpeli G, Buultjens AH, Giulieri S, Owusu-Mireku E, Aboagye SY, Baines SL, Seemann T, Bulach D, da Silva AG, Monk IR, Howden BP, Pluschke G, Yeboah-Manu D and Stinear T. 2017. Genomic analysis of ST88 community acquired methicillin resistant *Staphylococcus aureus* in Ghana..*PeerJ*,

86. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM and Lund O. Multilocus Sequence Typing of Total Genome Sequenced Bacteria. *J. Clin. Micobiol*. 2012. 50(4): 1355-1361.

87. Lewis K. Persister cells and the riddle of biofilm survival. *Biochemistry (Mosc)*. 2005 Feb;70(2):267-74. doi: 10.1007/s10541-005-0111-6. PMID: 15807669.

88. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools, *Bioinformatics*, Volume 25, Issue 16, 15 August 2009, Pages 2078–2079, https://doi.org/10.1093/bioinformatics/btp352.

89. Lioliou E, Sharma CM, Caldelari I, Helfer A-C, Fechter P, Vandenesch F, Vogel J, Romby P, 2012. Global regulatory functions of the *Staphylococcus aureu*s endoribonuclease III in gene expression. *PLoS Genet*. 8, e1002782.

90. Liu Q, Chen N, Chen H *et al.* RNA-Seq analysis of differentially expressed genes of *Staphylococcus epidermidis* isolated from postoperative endophthalmitis and the healthy conjunctiva. *Sci Rep* **10,** 14234 (2020). https://doi.org/10.1038/s41598-020-71050-6

91. Liu Y, Wang H, Du N, Shen E, Chen H, Niu J, Ye H, Chen M. 2009. Molecular evidence for spread of two major methicillin-resistant Staphylococcus aur- eus clones with a unique geographic distribution in Chinese hospitals. Antimicrob Agents Chemother 53:512–518. https://doi.org/10.1128/AAC .00804-08.

92. Liu Y, Dong J, Wu N, Gao Y, Zhang X, Mu C, et al. (2011) The Production of Extracellular Proteins Is Regulated by Ribonuclease III via Two Different Pathways in *Staphylococcus aureus*. *PLoS ONE* 6(5): e20554. https://doi.org/10.1371/journal.pone.0020554

93. Liu Y, Zhang J, Ji Y. Environmental factors modulate biofilm formation by *Staphylococcus aureus.* https://doi.org/10.1177/0036850419898659

94. Livak KJ and Schmittgen TD. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\triangle\triangle C(T)}$ method. *Methods* 25, 402–408

95. Löfblom J, Kronqvist N, Uhlén M, Ståhl S, Wernérus H. Optimization of electroporation-mediated transformation: *Staphylococcus carnosus* as model organism. *J Appl Microbiol*. 2007 Mar;102(3):736-47. doi: 10.1111/j.1365-2672.2006.03127.x. PMID: 17309623.

96. Ma XX, Ito T, Kondo Y, Cho M, Yoshizawa Y, Kaneko J, Katai A, Higashiide M, Li S and Hiramatsu K. 2008. Two Different Panton-Valentine Leukocidin

Phage Lineages Predominate in Japan. J Clin Microbiol. 46(10): 3246–3258. doi: 10.1128/JCM.00136-08

97. Mah TFC, O'Toole GA. Mechanisms of biofilm resistance to antimicrobial agents. *Trends in Microbiology* 9(1): 34-39, 2001

98. Marcaida MJ, DePristo MA, Chandran V, Carpousis AJ, Luisi BF. The RNA degradosome: life in the fast lane of adaptive molecular evolution. *Trends Biochem Sci* 2006; 31:359-65; PMID:16766188; http://dx.doi. org/10.1016/j.tibs.2006.05.005.

99. Mäder U, Nicolas P, Depke M, Pane-Farre J, Debarbouille M, Van Der Kooi-Pol MM, et al. (2016). *Staphylococcus aureu*s transcriptome architecture: from laboratory to infection-mimicking conditions. *PLoS Genet*. 12:e1005962. 10.1371/journal.pgen.1005962

100. Marincola G, Schäfer T, Behler J, Bernhardt J, Ohlsen K, Goerke C, Wolz C. RNase Y of *Staphylococcus aureus* and its role in the activation of virulence genes. *Mol. Microbiol*, 2012. 85, 817–832.

101. Massé E, Escorcia FE, Gottesman S. Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. *Genes Dev* 2003; 17:2374- 83; PMID:12975324; http://dx.doi.org/10.1101/ gad.1127103.

102. Mathy N, Bénard L, Pellegrini O, Daou R, Wen T, Condon C. 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell* 2007; 129:681-92; PMID:17512403; http://dx.doi.org/10.1016/j.cell.2007.02.051.

103. Miller MB and Bassler BL. Quorum Sensing In Bacteria. *Annu. Rev. Microbiol*. 2001. 55:165–99

104. Moldovan A, Fraunholz MJ. In or out: Phagosomal escape of *Staphylococcus aureus*. *Cell Microbiol*. 2019 Mar;21(3):e12997. doi: 10.1111/cmi.12997. Epub 2019 Jan 18. PMID: 30576050.

105. Montgomery CP, Boyle-Vavra S, Adem PV, Lee JC, Husain AN, Clasen J and Daum RS. 2008. Comparison of virulence in community- associated methicillin-resistant *Staphylococcus aureus* pulsotypes USA300 and USA400 in a rat model of pneumonia. *J. Infect. Dis*. 198:561–570.

106. Morfeldt E, Taylor D, von Gabain A, Arvidson S. Activation of alpha-toxin translation in *Staphylococcus aureus* by the trans-encoded antisense RNA, RNAIII. *EMBO J*. 1995 Sep 15;14(18):4569-77. PMID: 7556100; PMCID: PMC394549.

107. Nadig S, Ramachandra Raju S, Arakere G. Epidemic meticillin-resistant *Staphylococcus aureus* (EMRSA-15) variants detected in healthy and diseased individuals in India. *J Med Microbiol*. 2010 Jul;59(Pt 7):815-821. doi: 10.1099/jmm.0.017632-0. Epub 2010 Mar 25. PMID: 20339016.

108. Nadig S, Veluswamy N, Lalitha P, Kar S, Sharma S, Arakere G. 2012. *Staphylococcus aureus* eye infections in two Indian hospitals: emergence of ST772 as a major clone. *Clin Ophthalmol*. 6: 165-173.

109. Nadig S, Murthy S, Vandanashree M, Subramanya HS, Gopal B, Vembar S. Draft Genome Sequence of the Community-Associated *Staphylococcus aureus* Sequence Type 88 Strain LVP-7, Isolated from an Ocular Infection. *Microbiol Resour Announc*. 2021 Feb 18;10(7):e00077-21. doi: 10.1128/MRA.00077-21. PMID: 33602729; PMCID: PMC7892662.

110. Nascimento M, Sousa A, Ramirez M, Francisco AP, Carriço JA, Vaz C, PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods, *Bioinformatics*, Volume 33, Issue 1, 1 January 2017, Pages 128–129

111. Neopane P, Nepal HP, Shrestha R, Uehara O and Abiko Y. (2018). In vitro biofilm formation by *Staphylococcus aureus* isolated from wounds of hospital-admitted patients and their association with antimicrobial resistance. *Int. J. Gen. Med.* 11, 25–32. doi: 10.2147/IJGM.S153268

112. Novick RP, Ross HF, Projan SJ, Kornblum J, Kreiswirth B, Moghazeh S. 1993. Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO J*. 12:3967–75

113. Novick RP. 2000. Pathogenicity factors and their regulation, p. 392–407. In Fischett VA, Novick RP, Ferretti JJ, Portnoy DA and Rood JI (ed.), Gram-positive pathogens. *American Society for Microbiology*, Washington, D.C.

114. Novick RP. 2006. Staphylococcal pathogenesis and pathogenicity factors: genetics and regulation, in Gram-positive pathogens, 2nd edition. *ASM Press*, Washington, DC, pp, 496-516.

115. Novick RP, Subedi A. The SaPIs: mobile pathogenicity islands of Staphylococcus. *Chem Immunol Allergy*. 2007; 93:42-57. doi: 10.1159/000100857. PMID: 17369699.

116. Novick RP and Geisinger E (2008) Quorum sensing in staphylococci. *Annu. Rev. Genet*. 42, 541–564.

117. Ogston A. Report upon Micro-Organisms in Surgical Diseases. *Br Med J*. 1881 Mar 12:1(1054): 369 b2-75.

118. Ogston A. Micrococcus Poisioning. *J Anat Physiol*. 1882 Jul;16:526:67

119. Ouellet J (2016) RNA fluorescence with light-up aptamers. *Front Chem* 4:29.

120. Oun S, Redder P, Didier J-P, François P, Corvaglia A-R, Buttazzoni E, Giraud C, Girard M, Schrenzel J, Linder P. The CshA DEAD-box RNA helicase is important for quorum sensing control in *Staphylococcus aureus. RNA Biol.*, 2013, 10, 157–165.

121. Paige JS, Wu KY & Jaffrey SR. RNA mimics of green fluorescent protein. *Science* 333, 642–646 (2011).

122. Parker H, Albrett AM, Kettle AJ et al. Myeloperoxidase associated with neutrophil extracellular traps is active and mediates bacterial killing in the presence of hydrogen peroxide. *J Leukocyte Biol* 2012 ;91:369–76.

123. Passador L, Cook JM, Gambello MJ, Rust L, Iglewski BH. 1993. Expression of Pseudomonas aeruginosa virulence genes requires cell-to-cell communication. *Science* 260:1127–30

124. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem*. 25, 1605– 1612

125. Pichon C, Felden B. Small RNA genes expressed from *Staphylococcus aureus* genomic and pathogenicity islands with specific expression among pathogenic strains. Proc Natl Acad Sci U S A. 2005 Oct 4;102(40):14249-54. doi:

10.1073/pnas.0503838102. Epub 2005 Sep 23. Erratum in: *Proc Natl Acad Sci U S A*. 2005 Nov 15;102(46):16905. PMID: 16183745; PMCID: PMC1242290.

126. Raj R, Nadig S, Patel T, Gopal B. Structural and biochemical characteristics of two Staphylococcus epidermidis RNase J paralogs RNase J1 and RNase J2. *J. Biol. Chem*. September 29, 2020, DOI 10.1074/jbc.RA120.014876

127. Ribeiro-Gonçalves B, Francisco AP, Vaz C, Ramirez M, Carriço JA. PHYLOViZ Online: web-based tool for visualization, phylogenetic inference, analysis and sharing of minimum spanning trees. *Nucleic Acids Res*. 2016 Jul 8;44(W1):W246-51. doi: 10.1093/nar/gkw359. Epub 2016 Apr 29. PMID: 27131357; PMCID: PMC4987911.

128. Robinson DA & Enright MC (2004) Multilocus sequence typing and the evolution of methicillin-resistant *Staphylococcus aureus*. *Clin Microbiol Infect* 10: 92–97.

129. Romilly C, Lays C, Tomasini A, Caldelari I, Benito Y, Hammann P, Geissmann T, Boisset S, Romby P, and Vandenesch F. A Non-Coding RNA Promotes Bacterial Persistence and Decreases Virulence by Regulating a Regulator in *Staphylococcus aureus*. *PLoS Pathog*. 2014 Mar; 10(3): e1003979

130. Roux CM, DeMuth JP, Dunman PM. Characterization of components of the *Staphylococcus aureus* mRNA degradosome holoenzyme-like complex. *J Bacteriol* 2011; 193:5520-6; PMID:21764917; http://dx.doi.org/10.1128/JB.05485-11

131. Rutar T, Chambers HF, Crawford JB, et al. Ophthalmic manifestations of infections caused by the USA 300 clone of community associated methicillin resistant *Staphylococcus aureu*s. *Ophthalmology*. 2006;**113**:1455–1462.

132. Sakoulas G, Eliopoulos GM, Moellering RC, Jr, Wennersten C, Venkataraman L, Novick RP, Gold HS. 2002. Accessory gene regulator (*agr*) locus in geographically diverse *Staphylococcus aureus* isolates with reduced susceptibility to vancomycin. *Antimicrob. Agents Chemother*. 46:1492–1502. 10.1128/AAC.46.5.1492-1502.2002

133. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

134. Shambat S, Nadig S, Prabhakara S, Bes M, Etienne J and Arakere G. 2012. Clonal complexes and virulence factors of *Staphylococcus aureus* from several cities in India. *BMC Microbiology*.

135. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003 Nov;13(11):2498-504. doi: 10.1101/gr.1239303. PMID: 14597658; PMCID: PMC403769.

136. Smits WK, Kuipers OP and Veening J-W. Phenotypic variation in bacteria: the role of feedback regulation. *Nature reviews* 2006, vol 4

137. Song W, Strack RL, Svensen N & Jaffrey SR. Plug-and-play fluorophores extend the spectral properties of Spinach. *J. Am. Chem. Soc.* 136, 1198–1201 (2014).

138. Sousa AM, Machado I, and Pereira MO. Phenotypic switching: an opportunity to bacteria thrive. Science against microbial pathogens: *communicating current research and technological advances*. A. Méndez-Vilas (Ed.) 2011

139. Spaulding AR, Salgado-Pabon W, Kohler PL, Horswill AR, Leung DY, Schlievert PM. Staphylococcal and streptococcal superantigen exotoxins. *Clin Microbiol Rev*. 2013; 26(3):422–47. https://doi.org/10. 1128/CMR.00104-12 PMID: 23824366

140. Srivastava SK, Rajasree K, Fasim A, Arakere G, & Gopal B (2014). Influence of the AgrC-AgrA complex on the response time of *Staphylococcus aureus* quorum sensing. *Journal of bacteriology*, *196*(15), 2876–2888. https://doi.org/10.1128/JB.01530-14

141. Subramanian D, Bhasuran B, Natarajan J, Genomic analysis of RNA-Seq and sRNA-Seq data identifies potential regulatory sRNAs and their functional roles in *Staphylococcus aureus*, *Genomics*, Volume 111, Issue 6, 2019, Pages 1431-1446, ISSN 0888-7543, https://doi.org/10.1016/j.ygeno.2018.09.016.

142. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, & von Mering C. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made

broadly accessible. *Nucleic acids research*, 45(D1), D362–D368. https://doi.org/10.1093/nar/gkw937

143. Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, et al. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* 1995;33:2233–9

144. Thakur A, Clegg A, Chauhan A, Willcox MD. Modulation of cytokine production from an EpiOcular corneal cell culture model in response to *Staphylococcus aureu*s superantigen. *Aust N Z J Ophthalmol*. 1997; 25 Suppl 1:S43–5. https://doi.org/10.1111/j.1442-9071.1997.tb01754.x PMID: 9267623

145. Tomasini A, François P, Howden BP, Fechter P, Romby P, Caldelari I. The importance of regulatory RNAs in S*taphylococcus aureus*. infection, *Genetics and Evolution*, 2014, 21:616-26

146. Tomlinson BR, Malof ME and Shaw LN. A global transcriptomic analysis of *Staphylococcus aureus* biofilm formation across diverse clonal lineages. *Microbial Genomics*, 2021.7:000598

147. Trachman RJ, Demeshkina NA, Lau M, Panchapakesan S, Jeng S, Unrau PJ & Ferré-D'Amaré AR. (2017). Structural basis for high-affinity fluorophore binding and activation by RNA Mango. *Nature chemical biology*, *13*(7), 807–813. https://doi.org/10.1038/nchembio.2392

148. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*. 1990 Aug 3;249(4968):505-10. doi: 10.1126/science.2200121. PMID: 2200121.

149. Tuchscherr L, Medina E, Hussain M, Völker W, Heitmann V, Niemann S, Holzinger D, Roth J, Proctor RA, Becker K, Peters G, Löffler B. *Staphylococcus aureus* phenotype switching: an effective bacterial strategy to escape host immune response and establish a chronic infection. *EMBO Mol Med*. 2011 Mar;3(3):129-41. doi: 10.1002/emmm.201000115. Epub 2011 Jan 26. PMID: 21268281; PMCID: PMC3395110.

150. Turner NA, Sharma-Kuinkel BK, Maskarinec SA *et al.* Methicillin-resistant *Staphylococcus aureus*: an overview of basic and clinical research. *Nat Rev Microbiol* **17,** 203–218 (2019). https://doi.org/10.1038/s41579-018-0147-4

151. van der Woude MW. Phase variation: how to create and coordinate population diversity. *Curr Opin Microbiol*. 2011 Apr;14(2):205-11. doi: 10.1016/j.mib.2011.01.002. Epub 2011 Feb 1. PMID: 21292543.

152. Vanhommerig E, Moons P, Pirici D, Lammens C, Hernalsteens J-P, et al. Comparison of biofilm formation between major clonal lineages of methicillin resistant *Staphylococcus aureus*. *PLoS One* 2014;9:e104561.

153. Vogel J, Wagner EG (2007) Target identification of small noncoding RNAs in bacteria. *Curr Opin Microbiol* 10:262–270

154. Wang B, Zhao A, Novick RP, and Muir TW. (2015). Key driving forces in the biosynthesis of autoinducing peptides required for staphylococcal virulence. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10679–10684. doi: 10.1073/pnas.15060 30112

155. Wang R, Braughton KR, Kretschmer D, Bach TH, Queck SY, Li M, Kennedy AD, Dorward DW, Klebanoff SJ, Peschel A, DeLeo FR, Otto M. 2007. Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA. *Nature Medicine* 13:1510–1514. DOI: https://doi.org/10.1038/nm1656, PMID: 17 994102

156. Wang Z, Gerstein M & Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 2009.**10**, 57–63. https://doi. org/10.1038/nrg2484

157. Waters CM and Bassler BL. Quorum sensing: cell-to-cell communication in bacteria. *Annu. Rev. Cell. Dev. Biol*., 21, 319-346, 2005.

158. Waters LS, Storz G (2009) Regulatory RNAs in bacteria. *Cell* 136:615–628

159. Xijin Ge S, Jung D, Yao R, ShinyGO: a graphical gene-set enrichment tool for animals and plants, *Bioinformatics*, Volume 36, Issue 8, 15 April 2020, Pages 2628–2629, https://doi.org/10.1093/bioinformatics/btz931

160. Yan Xu, SuYun Qian, KaiHu Yao, Fang Dong, WenQi Song, Chen Sun, Xin Yang, Jing Hui Zhen, XiQing Liu, Zhi Yong Lv, Xi Yang. *World Journal of Pediatrics* (2021) 17:180–188 Clinical and molecular characteristics

of *Staphylococcus aureus* isolated from Chinese children: association among the agr groups and genotypes, virulence genes and disease types

161. Yao Y, Sturdevant DE, and Otto M. (2005) Genomewide analysis of gene expression in Staphylococcus epidermidis biofilms: insights into the pathophysiology of S. epidermidis biofilms and the role of phenol-soluble modulins in formation of biofilms. *J. Infect. Dis.*, 191, 289–298.

162. Yarwood JM and Schlievert PM (2003). Quorum sensing in Staphylococcus infections. *J. Clin. Invest.* 112, 1620–1625. doi: 10.1172/JCI20442

163. Zaman M, Andreasen M. Cross-talk between individual phenol-soluble modulins in *Staphylococcus aureus* biofilm enables rapid and efficient amyloid formation. *Elife*. 2020 Dec 1;9:e59776. doi: 10.7554/eLife.59776. PMID: 33259287; PMCID: PMC7732344.

164. Zhao S, Ye Z and Stanton R. 2020. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA*. doi:10.1261/rna.074922.120