Investigating the DNA methylation landscape of *Staphylococcus aureus*
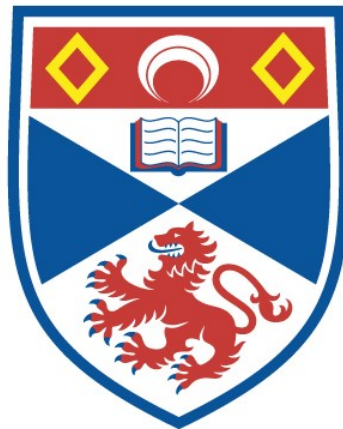
Rebecca Mekler

A thesis submitted for the degree of PhD
at the
University of St Andrews

2022

Dedicated to Manykó
Edit mama and Tóni papa,
and to my family.

Candidate's declaration

I, Rebecca Mekler, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 60,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in November 2016.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date   24/02/2021   Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date   24/02/2021   Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Rebecca Mekler, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date   24/02/2021   Signature of candidate

Date   24/02/2021   Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Rebecca Mekler, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date   24/02/2021   Signature of candidate

ACKNOWLEDGEMENTS

ABSTRACT

Staphylococcus aureus is a leading causative agent of healthcare-associated infections. One aspect of the organism that remains unknown is the methylome, specifically that of the whole genome. In prokaryotes, methylation is facilitated by methyltransferases, usually part of the organism's Restriction Modification system (RM). It is well established that RM are involved in cellular defense but have also been attributed to have secondary regulatory functions in host physiology and virulence by modulating gene expression through DNA methylation in numerous bacterial species.

In *S. aureus* the main RM present are Type I RM Sau1, which potential epigenetic role has not yet been studied. Using PacBio SMRT sequencing this study investigates the variability and distribution of *sau1* DNA binding specificity unit (*hsdS*) alleles and explores the frequency of whole genome 6mA methylation within the species using a historically and phylogenetically variable collection of *S. aureu*s isolates part of the NCTC3000 project. The results revealed lineage specific methylation patterns randomly distributed throughout the chromosome, but preferential methylation of the coding sequence and the core genome. Between the 24 represented STs, the detailed protein structure of 40 different HsdS homologs were characterised and matched to corresponding 6mA target recognition sequences, greatly augmenting the current knowledge of Sau1 methylation signatures.

Differential methylation was also investigated in novel ST622 hybrid strains as a natural experiment (variable methylation signatures across an identical sequence region between chimeric and closely related ST45 and ST22 donor strains) effectively looking at the effect of large-scale recombination on whole genome methylation using RNA-Sequencing. Mutagenesis of *hsdS* and further transcriptomic studies revealed that deletion of 6mA methylation by Sau1 in a set of isogenic mutants in multiple sequence backgrounds causes a pleotropic shift in expression of metabolic genes. This is not likely due to an epigenetic regulatory mechanism, but rather and induced global stress response.

TABLE OF CONTENTS

ABREVIATIONS

4mC	N4-methyl-cytosine
5mC	N5-methyl-cytosine
6mA	N6-methyl-adenine
AA	amino acid
ACME	arginine catabolic mobile element
ATP	adenosine triphosphate
BHI	Brain Heart Infusion
BLAST	Basic Local Alignment Search Tool
BLASTn	BLAST – protein
BLASTx	BLAST – translated nucleotide
bp	base-pair
CA	Community-associated
CC	clonal complex
CDS	coding sequence
ChIP-seq	chromatin immunoprecipitation sequencing
CPM	count-per-million
CR	conserved region
CV	core variable
D	domain
DE	Differential expression
DEG	Differentially expressed genes
DNA	Deoxyribonucleic acid
DR	down regulation/ed
EU	European
FDR	False discovery rate
FPKM	fragments per kilobase of transcript per million mapped reads
GO	gene ontology – biological or molecular pathway
h	hour
HA	Hospital-associated
HGAP	Hierarchical Genome Assembly Process
HGT	Horizontal gene transfer
INT	intergenic region
IPD	impulse duration
IS	insertion sequence
kb	kilo-base
KO	Knock out mutant
LA	Livestock-associated
logFC	log-fold-change
M/MTase	methyltransferase
Mb	million-base pairs
MGE	Mobile genetic element
MLST	multi-locus sequence typing
mRNA	Messenger RNA

MRSA	Methicillin-Resistant *Staphylococcus aureus*
MSSA	Methicillin-Sensitive *Staphylococcus aureus*
ncRNA	non-coding RNA
NCTC	National Culture Type Collection
NGS	next-generation sequencing
nm	nanometre(s)
OD_{600}	Optical density at 600nm
PacBio	Pacific Biosciences
PCR	Polymerase Chain Reaction
PDB	protein data bank
PI	pathogenicity island
R/REase	restriction endonuclease
RM	Restriction-Modification
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing – next generation, real time RNA quantification
RPKM	reads per kilobase of transcript per million mapped reads
S	specificity unit
SaPI	Staphylococcal pathogenicity island
SCC*mec*	Staphylococcal Chromosomal Cassette
SMRT	single molecule real-time
SNP	Single nucleotide polymorphism
spp	species
ST	sequence type
TI RM	Type I Restriction-Modification
TII RM	Type II Restriction-Modification
TIII RM	Type III Restriction-Modification
TIV RM	Type IV Restriction-Modification
TMM	trimmed mean of M-Values
TN	transposon
TRD	Target recognition domain
tRNA	transfer RNA
TRS	Target recognition sequence
TSA	Tryptic Soy Agar
TSB	Tryptic Soy Broth
UK	United Kingdom
UR	up regulation/ed
USA	United States of America
UTR	untranslated region
vSa	genomic island
WGA	whole genome amplification
WGS	whole genome sequencing
WHO	World Health Organisation

LIST OF FIGURES

1. INTRODUCTION

2. METHODS

3. Species Wide Characterisation of Restriction-Modification Systems and TI RM Sau1 6mA Methylation in *Staphylococcus aureus*

4. The Effect of Large-Scale Chromosomal Replacement on Whole Genome Methylation and Gene Expression Profiles in *Staphylococcus aureus*

5. Functional Impact of Sau1 Facilitated 6mA DNA Methylation in *Staphylococcus aureus*

8. APPENDIX

LIST OF TABLES

4. The Effect of Large-Scale Chromosomal Replacement on Whole Genome Methylation and Gene Expression Profiles in *Staphylococcus aureus*

5. Functional Impact of Sau1 Facilitated 6mA DNA Methylation in *Staphylococcus aureus*

8. APPENDIX

1. INTRODUCTION

1.1 *STAPHYLOCOCCUS AUREUS*

1.1.1 Diversity of *S. aureus* Population

Despite substantial increase in the prevention and control of infectious disease globally since the Millennium Development Goals, bacterial and viral infections remain a worldwide public health problem and are still one of the leading causes of morbidity and mortality worldwide (Cohen, 2000). Globalization has led to increased migration and advancements in technology have contributed to the adaptation of these microorganisms, causing evolving modern infections, the re-emergence of previously controlled disease, and the advent of antimicrobial resistance (Nigam, Gupta, & Sharma, 2014; Santoro, Simone, & Timen, 2015). Improved access and availability to drug treatments have given rise to multi-resistance 'superbugs', which pose a new challenge for researchers. The spread of these pathogens, especially in hospital-associated settings, has been steadily increasing since the World Health Organization (WHO) called attention to the clinical and socioeconomically problems caused these microbes (Cohen, 2000; Dye, 2014; Grundmann et al., 2006).

Staphylococcus aureus is the leading causative agent of healthcare-associated infections. It has become a global public health threat, especially disease caused by antimicrobial resistant, methicillin-resistant *Staphylococcus aureus* (MRSA), which is a high priority pathogen for research and development of new antibiotics published by the WHO in February, 2017 (Bradley, 2002; Klein et al., 2007; WHO, 2017). *S. aureus* is a Gram-positive bacterium, fundamentally a commensal microorganism asymptomatically colonizing about 30% of the human population carried as part of the normal skin microbiota, the nasal cavity and the mucous membranes (Naber, 2009). Apart from the asymptomatic carriage, *S. aureus* is capable of causing a wide variety of disease (Choo & Chambers, 2016; Miller & Cho, 2011; Naber, 2009) ranging from acute skin conditions, to severe life threatening invasive disease including joint and surgical site infection, pneumonia, septicemia and various toxicoses (Bradley, 2002; Feng et al., 2013; Grumann et al., 2014; Khan, Ahmad, & Mehboob, 2015). Most of these infections are healthcare-acquired, nosocomial infections and are avoidable (Hennekinne et al., 2012; Kinnevey et al., 2013; Richards et al., 1999). Efficient disease management and infection prevention caused by *S. aureus* is particularly important for immunosuppressed and high-risk, vulnerable patient groups including children and the elderly (Hiramatsu et al., 2001; Klein et al., 2009; Wong et al., 2016).

Although most human infections caused by *S. aureus* are induced in hospital-associated (HA) settings, the epidemiology, population and evolutionary dynamics of particular resistant *S. aureus* strains are quickly changing, now present as community-acquired MRSA (CA-MRSA) disease in the general population and as a food-born pathogen: livestock-associated MRSA (LA-MRSA) (Deleo, Otto, Kreiswirth, & Chambers, 2010; Fitzgerald & Holden, 2016; Prosperi et al., 2013). Staphylococci also have the inherent ability to form biofilms on abiotic and biotic surface enabling it to survive in a wide variety of environments (Doulgeraki, Di Ciccio, Ianieri, & Nychas, 2017; Golding et al., 2010; Mediavilla, Chen, Mathema, & Kreiswirth, 2012). Contact with the bacterial biofilms contaminated objects and surfaces add to the increased transmission of *S. aureus* disease, other than direct contact (skin-to-skin or droplet) with an infected individual/animal (Lozano et al., 2016; Tokajian, 2012). Infections caused by methicillin-susceptible *S. aureus* strains (MSSA) or resistant strains are primarily associated with a number of successful clones, and occur usually as an epidemic wave, being endemic in healthcare, community, agricultural and food industry settings (Aires-de-Sousa, 2016; Bal et al., 2016; Patel et al., 2015; Shambat et al., 2012). MRSA has become a major problem worldwide, and the emergence and spread of multi-drug resistant strains and vancomycin resistance *S. aureus* (VRSA) is becoming an increasing infectious disease burden especially in Asia and the Americas (Anderson et al., 2014; Lin et al., 2016; Otter & French, 2008). The constant change in distribution patterns of clones with increased virulence and resistances have significantly impacted the need for appropriate surveillance, prevention and management of the infections caused by *S. aureus* (Bosi et al., 2016; Fitzgerald & Holden, 2016; Klein et al., 2009; Mediavilla et al., 2012).

1.1.2 Diversity of *S. aureus* Population

Several phenotypic (phage typing, serotyping, multi locus enzyme electrophoresis, immunoblotting) and molecular genotypic typing tools (single or multi locus sequence typing (S/M LST), SCC*mec* typing (Kinnevey et al., 2013), *spa* typing, pulsed-field gel electrophoresis) have been used to study the differentiation between closely related lineages of *S. aureus* (Feil et al., 2003; Saunders & Holmes, 2007; Suzuki et al., 2009). The most commonly used is MLST including 7 core housekeeping genes (*arcC, aroE, glpF, gmk, pta, tpi, yqiL*) screened for any polymorphisms within these loci in result producing an allelic profile for each isolate, known as a sequence type (ST) (Enright, 2008; Joseph et al., 2016; Larsen et al., 2017; Stephens et al., 2006).

The population of *S. aureus* consist of several dominant lineages, 10 of which are commonly seen in most human nasal carriage worldwide, most notably CC5, CC8, CC22, CC30, CC45 and globally distributed sequence types including CC1, CC25, CC59 and CC121 (Feil et al., 2003). Specific geographic regions have dominant strains, as well as locally spread clones, which are the major cause of endemic outbreaks in hospital and community settings as seen in Figure 1.1 (Fitzgerald & Holden, 2016; Joseph et al., 2016).
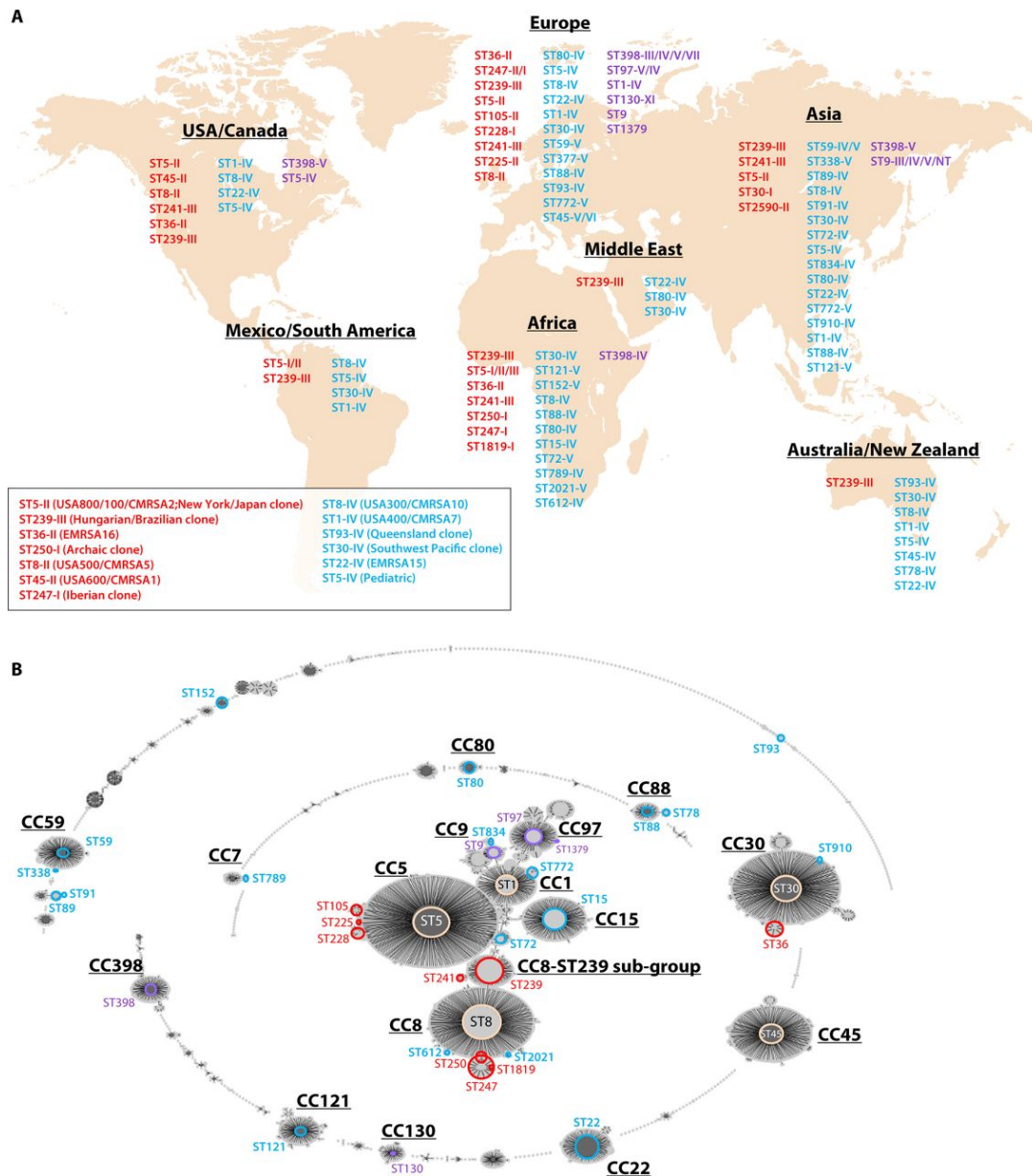
Figure 1.1 | Global MRSA population summary
ST types divided according to HA-MRSA (red), CA-MRSA (blue) and LA-MRSA (purple). **A.** MRSA population structure showing major clones reported for various geographic locations with accompanying SCCmec types, with traditional or alternative names for epidemic strains highlighted in the box on the left. **B**. Evolutionary relationships of MRSA lineages represented by eBURST analysis, highlighting clonal complexes which each of the highlighted STs in panel A belong to. Adapted from Lakhundi & Zhang, 2018.

1.1.2.1 Successful Lineages

Methicillin resistant *Staphylococcus aureus* (MRSA) emerged in early 1941 (ST250 background) with the acquisition of type I SCC*mec* element, but was first reported in 1960 one year after the introduction of methicillin beta lactam drug in 1959 in the UK (Harkins et al., 2017). The resistant clone quickly spread from the UK to Denmark throughout continental Europe and the present distribution of methicillin-resistant *Staphylococcus* is global. There are a handful of lineages which are more prominently seen, with some geographically condensed, but other epidemiologically 'successful' MRSA clones are seen in high frequency worldwide (Aanensen et al., 2016; Holden et al., 2010; Holden et al., 2013; Lindsay, 2010).

1.1.2.2 Hospital Associated MRSA (HA-MRSA)

Epidemic-MRSA-15 (EMRSA-15) is one of the most successful clones of ST 22 (Type IV SCC*mec*) HA-MRSA that spread rapidly in hospitals throughout the southeast region of England in 1991. By 2000 the level of UK *S. aureus* bacteraemia had increased from 2% to 40% and MRSA outbreaks were reported in various different European countries, New Zealand and Singapore (Holden et al., 2013). Strains from ST22 and close relatives (CC22) still remain the most frequent HA-MRSA lineage found across Europe along with clones belonging to CC30 (EU & Americas), CC8 (EU, Americas, Asia) with variants from CC239, CC5 (Italy, Americas, Korea, Japan), and CC45 (Germany, USA) being the dominant epidemic clonal complexes – Figure 1.1 A (red) (Choo & Chambers, 2016; Cockfield, Pathak, Edgeworth, & Lindsay, 2007; Deurenberg & Stobberingh, 2008; Harris et al., 2010; Lindsay, 2010; Patel et al., 2015).

1.1.2.2.1 Community Acquired MRSA (CA-MRSA)

CA-MRSA has evolved independently of hospital-acquired clones (Figure 1.1 – B - blue) and outbreaks are between persons in close contact including schools, prisons, and sports clubs but are also present in healthcare facilities. In the USA, USA300 (CC8) and USA400 (CC1) have been the main lineages whilst in Europe strains belonging to CC80; in Asia CC59 and in the South West Pacific strains belonging to CC30 cause the majority of infections (Deleo et al., 2010; Deurenberg & Stobberingh, 2008; Ellington et al., 2015; Lindsay, 2010; Mediavilla et al., 2012; Otter & French, 2008; Patel et al., 2015).

1.1.2.2.2 Livestock Associated MRSA (LA-MRSA)

MRSA does not only affect human hosts but causes severe disease in livestock leading to substantial animal morbidity, mortality and economic losses. The predominant lineage associated with infection of pigs is CC398 mostly localized to Denmark, Belgium and the Netherlands, whilst ST9 clones dominate in Asia (Lindsay, 2010). These LA-MRSA strains have descended from human MSSA isolates, and have acquired tetracycline and methicillin resistance genes after host jump – an epidemiological paradigm shift (Bal et al., 2016). A number of LA-MRSA strains (CC1, CC5, CC97, CC121, CC130 and ST425 Figure 1.1 - purple) have since been reported in livestock and there is a cross species risk (human – livestock) of LA-MRSA in direct occupational contact with animals and food products contaminated through animal faeces (Bal et al., 2016; Golding et al., 2010).

1.1.2.2.3 Methicillin-Sensitive *S. aureus* (MSSA)

There are also several methicillin sensitive *Staphylococcus aureus* (MSSA) linages that are endemic worldwide which are from a different genetic background to MRSA clones including CC1, CC5, CC7, CC8, CC9, CC12, CC15, CC22, CC25, CC30, CC45, CC51, CC59 and CC101. These clones are found within hospitals and communities causing substantial diseases in neonatal intensive care units, bacteraemia and local inflammation of surgery sites and prosthetic joint replacements (Deurenberg & Stobberingh, 2008; Graham et al., 2002; Nienaber et al., 2011). The relative success of MRSA and endemic MSSA can be directly correlated with the ability of the species to adapt to various environments, as well as the selective advantage, which the acquisition and expression of exogenous resistance/virulence genes harboured, by mobile elements present within a *S. aureus* population (Fitzgerald & Holden, 2016; Shore et al., 2011).

1.1.3 The *S. aureus* Genome

The capacity and success of *S. aureus* clones to colonies different anatomical and environmental niches in an assortment of host species is promoted by their genetic variability (Fitzgerald & Holden, 2016; Jodi A. Lindsay, 2008; Shambat et al., 2012). Bacterial genome sequencing has provided insight into not just the genetic architecture of an organism, but has also allowed the analysis of the genomic relationship between different strains (Lindsay & Holden, 2006; Lindsay, 2010; Prax, Lee, & Bertram, 2013). Comparative genomics studies the variation of genetic material within a population of bacteria. This information is correlated with the expressed phenotype of the given *S. aureus* clones, investigating the evolution of the strain and may also highlight factors that are preferentially associated with more predominant clonal lineages (Lindsay, 2010; Lindsay & Holden, 2004). To understand the diversity within a *S. aureus* population, it is important to know the features of the genomes, which are used to categorize the variability between isolates (Chen et al., 2013; Ebruke et al., 2016).

S. aureus have a circular chromosome that is approximately 2.8 Mb in size and contains ~2700 protein-coding sequences (CDSs), and regulatory or structural RNAs (Fournier, 2008; Holden & Lindsay, 2008; Lindsay & Holden, 2004). Previous WGS and microarray studies have revealed that the *S. aureus* genome can be classified into conserved core and variable accessory genome regions (Fitzgerald et al., 2001, Lindsay & Holden 2004).

1.1.3.1 Conserved Core Genome

A bacterial species can be described by its pan-genome, which consists of the 'core' backbone complement of the genome, containing a conserved pool of genes present in all the bacterial genomes for the given species (Joseph et al., 2016; Segerman, Mathee, & Rolain, 2012; Van Tonder et al., 2014). Over 75% of the CDSs (ranging from 2592-2748 CDS (Lindsay & Holden, 2004)) that belong to the core genome and are highly conserved across all strains (Fitzgerald et al., 2001, Boissy 2011). These include genes involved in essential metabolic functions for the growth and survival of the cell, shared across 95% of the species (Lindsay & Holden, 2004). Although the components of the core genome may be conserved, small scale sequence variations can arise through mutation or homologous recombination. Homologous recombination (HR) has been a driver in the microevolution of bacterial genomes through the homogenisation of core genomes leading to the formation of interrelated population structures (Gonzalez-Torres et al., 2019). Successful

clones form clusters of bacterial strains within dominant lineages (Bosi et al., 2016; Feng et al., 2013; Shukla et al., 2012).

Beyond the core component of the genome some CDSs may be less preserved and include genes and regulons characterized as 'core variable' (CV), which comprise of various non-essential genes with common species associated functions such as toxins, surface binding proteins, capsule biosynthesis genes and exoenzymes (Lindsay, 2010; Lindsay & Holden, 2004). These coding sequences are categorised by assigned function/functional class according to the predicted structure and homology of these genes with those of other species (Rouli, Merhej, Fournier, & Raoult, 2015; Sun, Jiang, Wu, & Zhou, 2013). One of these core variable regions includes the staphylococcal genomic islands (vSa) alpha and beta (vSaα and vSaβ), encoding a cluster of exotoxins (*set*) and lipoprotein (*lpl*), as well as super-antigen like serine proteases (*spl*), the pantone-valentine leucocidin cluster (*lukDE*), bacteriocins (*bsa* cluster) and enterotoxins (*se* cluster) respectively. Although these elements contain intact or remnant recombinases, they do not carry replication genes of their own, most likely arising originally through horizontal transfer and spread between *S. aureus* via transduction. The islands are quite stable within the *S. aureus* genome, with low frequency variation seen as allelic forms among lineages differentiated on the basis of genetic content and the distribution, including possible pseudogenes indicating recombination and repeats in these regions. Much of this genomic stability and control of lineage differentiation is attributed to carriage of Type I Restriction-Modification (RM) Sau1 methyltransferase (*hsdM*) and specificity unit (*hsdS*) on both of these genomic islands. (Lindsay, 2008).

1.1.3.2 Variable Accessory Genome

On a pan-genome level, the accessory genome of *S. aureus* contributes 20-25% of the genome, introducing a more variable superset of genes different between strains, giving each isolate their own customized genomic repertoire (Lindsay, 2010; Lindsay & Holden, 2004; Segerman et al., 2012; Shukla et al., 2012). The accessory genome is usually more heterogeneous (than the core genome), characterized to carry a range of genes which encode for non-essential functions proteins including many involved in antimicrobial, antiseptic and metal resistance and virulence factors seen in Table 1.1 (Chatterjee et al., 2011; Gill et al., 2005; Gomes, Vinga, Zavolan, & Lencastre, 2005; Laabei et al., 2014; McCarthy & Lindsay, 2010; McCarthy et al., 2012; Spanu et al., 2011).

Table 1.1 | Major Mobile Genetic Elements in *Staphylococcus aureus* Reference Strains

MGE	MRSA252	MSSA476	MW2	COL	Mu50	N315
SCC						
SCC*mec* type I				*mecA*		
SCC*mec* type II	*mecA*				*mecA*	*mecA*
SCC*mec*-type III			*mecA*			
SCC*mec*-type IV						
SCC476		*far 1*				
Transposons						
Tn*544*	*ermA, spc*				*ermA, spc*	*ermA, spc*
Tn*552*	*blaZ*					
Tn*5801*					*tetM*	
Tn*916*-like						
Bacteriophage						
ΦSa1						
ΦSa2			*lukSF-PV*			
ΦSa3	*sea, sak, chp*	*sea, sak, seg2, sek2*	*sea, sak, seg2, sek2*		*sea, sak*	*sep, sak, chp*
ΦSa4						
ΦSa5						
Genomic Islands						
νSaα	*set (9), lpl (6)*	*set (11) lpl (5)*	*set (11) lpl (5)*	*set (7), lpl (6)*	*set (9), lpl (9)*	*set (10), lpl (9)*
νSaβ	*hysA, spl (5), exotoxin (6)*	*spl (4), lukDE, bsa*	*spl (4), lukDE, bsa*	*spl (5), lukDE, bsa*	*lukDE, spl (5), exotoxin (6)*	*spl (5), lukDE, exotoxin (6)*
Pathogenicity Islands						
SaPI1						
SaPI2				*seb, ear, seq,*	*se1, sec3, tat*	*se1, sec3,*
SaPI3			*ear, sel2, sec4*		*fhuD*	
SaPI4						
SaPI5						
Plasmids						
I	*ble, kan (pUB110)*			*tet (pT181)*	*ble, kan (pUB110)*	
II	*cadAC, arsBC (integrated)*	*blaZ, cadD (pSAS)*	*blaZ, cadD (pWW2)*			*ble, kan (pUB110)*
III					*aacA-aphD, qacA*	

Adapted from Lindsay & Holden, 2004

Many of these determinants are encoded on mobile (or formerly mobile) genetic elements (MGEs). MGEs are segments of DNA that are readily transposable carrying their own insertion mechanisms allowing the free transfer and integration of these sequences into the host *S. aureus* bacterial chromosome or circulating plasmids (Feil, 2004; Hiramatsu et al., 2001; Kuroda et al., 2001; Lindsay, 2010; McCarthy et al., 2011). The main MGEs found in *S. aureus* include staphylococcal cassette chromosome (SCC), *S. aureus* pathogenicity islands (SaPI), genomic islands (*v*Sa), plasmids, prophages, and smaller elements like transposons (Tn) and insertion sequences (IS) as detailed in Table 1.1. These elements can be horizontally transferred between strains adding to the mobility of virulence/resistant factors and intra-species genome plasticity (Alibayov et al., 2014; Lowy, 2003; Malachowa & Deleo, 2010; Ramsay et al., 2016; Stanczak-Mrozek et al., 2015; Strommenger et al., 2014).

1.1.3.2.1 Plasmids

Plasmids are small, autonomously replicating DNA molecules, which can replicate independent of the host. Wild type *S. aureus* typically carry one or more plasmids per cell and are usually classified according to the size and the mechanism of replication associated. Group I plasmids are small (1.3-4.6 kb) multi rolling-copy (*rep* gene for replication) plasmids (10-55 copies per cell) carrying single virulence or resistance determinants, group II plasmids are slightly larger (11.5-46 kb) but fewer (4-6 per cell) (Młynarczyk et al., 1998). Group II plasmids may carry several resistance determinants – mostly encoding penicillinase and aminoglycoside/trimethoprim resistance genes (e.g. *pSK 1, pIP630*), and undergo theta replication. Group III plasmids are large (30-60 kb) conjugative, multi-resistance plasmids accompanied by *tra* genes for conjugative horizontal gene transfer (Lindsay, 2008; Berg et al., 1998; Malachowa & DeLeo, 2010, Alibayov et al., 2014). Although *S. aureus* has a naturally low competence to acquire foreign genetic elements or readily transform, most of the intra-species acquisition of plasmids occur through conjugation. These integrated elements often carry antibiotic resistance genes providing advantage to the host under selective pressures due to antibiotic exposure (Alibayov et al., 2014). Several strains have integrated plasmids in their chromosome or the SCC*mec* element, often carrying vital antibiotic resistance genes (Lindsay, 2008).

A whole range of antibiotic resistance genes have been described on *S. aureus* plasmids: penicillins (*blaZ, mecA, femA*) tetracycline (*tetKML*), erythromycin (*ermABC*), kanamycin

(*kan, aphD*), gentamicin (*aacA*), bleomycin (*ble*), fosfomycin (*fosB*), fusidic acid (*fusBC*), linomycin (*linA*), chloramphenicol (*cat, fexA*), macrolides (*mphBM*), vancomycin (*vanA -* transferred through enterococcal transposon) and various heavy metals, disinfectants and additional virulence factors seen in Table 1.1. Although many of the plasmids present in staphylococci are cryptic, it is presumed that they might provide a selective advantage to the host cell, but they also may be lost within a population through random selection (Malachowa & DeLeo, 2010, Feng et al., 2008).

1.1.3.2.2 Bacteriophages

Bacteriophages are small viruses (approximately 45 kb), which insert into the host bacterial chromosome resulting in prophage. There are over 70 *S. aureus* prophages, which can be grouped into 11 serotypes (A-H and J-L). Most phage seen in *S. aureus* are temperate prophages (serotype A, B, & F), which integrate into the staphylococcal genome, but have the ability to lyse the host cell during stress conditions through induced excision and replication of their DNA to release the viral progeny into the extracellular environment (Deghorain & Van Melderen, 2012). Bacteriophages are the most widespread and most variable mobile element within the species, with some lineages carrying up to 4 phage integrated into the staphylococcal chromosome, each with their own integrase (*int*) for site-specific integration at *attP* sites (Lindsay et al., 2006). Some of these phages also carry excision genes (*xis*) as well as lytic-lysogenic switch (*cI, pro*) regulators, which control the expression of these pathways for these elements, and essentially their transfer (Lindsay, 2008). These genes and the corresponding insertion sites are used to classify these phage families.

Bacteriophages are typically lost and acquired within a bacterial population through induction of lytic state of integrated phage. However, antibiotics among other stresses may increase the frequency of the lytic states and subsequent release of infectious phage progeny. Apart from the obvious lytic consequences prophage carriage, following positive lysogenic conversion (expression) within the lytic cycle, these viruses are advantageous to the host as they also carry virulence and toxin genes. Some of these include enterotoxins (*SEs = sea, selk2, selq, selp*), chemotaxis inhibitory protein (*chp*), staphylockinase (sak), staphylococcal inhibitor of complement (*scn*), Panton-Valentine leukocidin (*PV-luk)*, and exfoliative toxin AB (*etAB*) which have a strong epidemiological association and roles in food poisoning, haemolytic pneumonia and soft skin conditions respectively (Alibayov et al., 2014; Xia & Wolz, 2014).

The distribution of the types of phage within a strain of *S. aureus* gives insight into evolving rearrangement or static nature or the phage family (φ3 phage are more constant possibly due to their larger in size), and subsequently their functional necessity. The frequent and flexible acquisition of prophages via transduction introduce a tremendous amount of variation between *S. aureus* strains, in very short time periods, allowing these organisms to adapt to induced stimuli or change of habitat (Malachowa & DeLeo, 2010; Alibayov et al., 2014). Phage also vary structurally in their repetitive, short mosaic sequence composition which allow frequent rearrangement of these small fragments (Lindsay, 2008)

1.1.3.2.3 Staphylococcus aureus Pathogenicity Islands (SaPI)

Staphylococcus aureus pathogenicity islands (SaPIs) are discrete 14-17 kb genetic sequences, which show a very high level of conservation, which are present throughout the *S. aureus* species. A large proportion of these MGEs carry virulence genes, in particular enterotoxins (*SE*) and super antigens (*tst1*) associated with food poisoning (SaPischik-awa11 and SaPIno10) and toxic shock syndrome. These islands are transferred at a high frequency through horizontal gene transfer and are integrated in single orientation at one of six att_c (8', 9', 18' 19', 44', and 49') sites via a site-specific helper phage integrases (Alibayov et al., 2014). They are then mobilized following infection by lytic phage or other induced stresses. SaPI are constructed similar to phage, containing repressor, integrase (*int*) and terminase (*terS*) genes but also harbours core genes, which regulate the life cycle and replication (*rep*) as well as the successful interaction of the islands with their helper prophages (*pif*) (Malachowa & Deleo, 2010; Sato'o et al., 2013).

The excision of the elements depends on the host *recA/lexA* induction pathway, helper prophage SOS-induced excision. After entry, the SaPI detect and attach to a specific chromosomal att_c sites. This in turn induces the excision and replication of helper phage which aid in the subsequent life cycle of the island (Alibayov et al., 2014; Dearborn & Dokland, 2012; Novick, 2003). The nomenclature, distribution and mosaic structure of SaPI vary between lineages similarly to phage; they have been named and numbered according to the origin of isolation – SaPI1MW2, SaPI1, SaPI2, SaPI3 or from bovine mastitis: SaPIbov1-2,5. (Alibayov et al., 2014; Lindsay, 2008; Malachowa & Deleo, 2010).

1.1.3.2.4 Transposons (Tn) & Insertion Sequences (IS)

S. aureus accessory genome is also broadened by transposons, which are discrete DNA sequences encoding their own transposes, therefore enabling independent replication of these elements from the host DNA which they have inserted themselves into (Młynarczyk et al., 1998). These MGEs can come in a single or multiple copies and unlike plasmids and SaPIs they do not have site-specific integration requirements so are extensively distributed among the staphylococcal genome. Transposons can be horizontally transferred via conjugation (linked with *tra* genes) or generalized transduction interspecies wide (Malachowa & Deleo, 2010). These elements therefore bring a great source of differentiation between species as they can randomly integrate into the chromosome effectively inducing both genotypic and phenotypic changes quite easily. They often encode virulence or resistant factors such as vancomycin resistance gene *vanA* and resistance genes to penicillin (*mec, bla*), erythromycin (*erm*) and tetracycline (*tet*) have been shown as seen in Table 1.1. This not only provides selective advantages across the *S. aureus* species, but also exemplifies further variation within the gram-positive bacterial population through successful transfer of enterococcal and streptococcal conductive transposons, Tn*4001* and Tn*918* respectively (Flannagan et al., 2003; Lindsay, 2008).

Insertion sequences (IS) are often paired with transposons, as they are similar sequences but do not encode their own transposases. These elements can form hybrid pairs with composite transposons (e.g. IS*256* & IS*257* with Tn*4001* & Tn*4003*) to mediate resistance to gentamicin (*gmr*), kanamycin (*kmr*) and tobramycin (*tmr*) (Alibayov et al., 2014; Lyon, Gillespie, & Skurray, 1987; Rouch, Byrne, Kong, & Skurray, 1987). IS insert randomly, into CDSs, promoters or regulatory sequences having polar effects on the transcription of the surrounding genes, potentially facilitating quick genomic evolution through transposition. These sequences are correlated to specific *S. aureus* lineages, and are rarely transferred between strains, although present in multiple copies within a host cell (Malachowa & Deleo, 2010).

1.1.3.2.5 Staphylococcal Cassette Chromosome (SCC)

The staphylococcal cassette chromosome (SCC) is a large DNA fragment, which is inserted into the *S. aureus* chromosome at a specific attachment site (*attB*$_{SCC}$) on the 3' of the *orfX* gene via unique site-specific recombinases designates as the cassette chromosome recombinases (*ccr*). The *ccr* gene cluster (*ccrAB* or *ccrC*) is responsible for

the mobility of the element encoding excision and integration enzymes for horizontal gene transfer (HGT). Along with the recombinases, the SCC encodes various antibiotic resistance (*mecA, ermA, aad9, spc*) and virulence determinants (*aac-aphD, copA, pls*) (Malachova & Deleo, 2010; Lina et al., 2006; Ito et al., 2009; Chontrakool et al., 2006).

The SCC can be classified into SCC*mec* or non-SCC*mec* groups. Strains containing the SCC*mec* elements produce an additional penicillin-binding protein (PBP2a) encoded by methicillin resistance gene, *mecA*. This supplementary PBP has a very low affinity for most β-lactam antibiotics, conferring resistance to most semi-synthetic penicillins. The expression of *mec* genes *(mecA, mecB* or *mecC)* is controlled through *mecR1* (transmembrane signal transducer protein) and *mecI*, which encodes a repressor protein. These three genes construct the *mecA* regulon - *mecI-mecR1-mecA* (Kaya et al., 2018; Malachova & Deleo, 2010; Lakhundi & Zhang, 2018).

SCC*mec* elements can be characterized into Types I to VIII, and SCC*mec* complex A-E depending on the *mecA* regulon complex and *ccr* allotypes. Class A-C are most commonly seen, where Class A elements contain the full *mecA* operon, but in Class B and C the regulon is disrupted by an insertion sequence, (IS) *IS1272-ΔmecR1-mecA* and *IS431-ΔmecR1-mecA* respectively. The elements can also be further differentiated into various subgroups depending on the variation of resistance & virulence genes present in the J 'junkyard' region of the region. These additional factors are often encoded on plasmids, IS or transposons which can incorporate within the MGE carrying resistance determinants like fucidic acid resistance *far1* gene (transposon mediated; SCC$_{476}$) or *mer* operon involved in ion transport (transposon mediated; SCC$_{mercury}$). Interestingly CA-MRSA strains tend to have a smaller SCC*mec* elements (T IV, V, VII) while HA-MRSA isolates have larger regions with more resistance factors SCC*mec* I, II, III, IV, VIII) (Malachova & Deleo, 2010; Holden & Lindsay, 2008; Ito et al., 2009; Feng et al., 2008).

1.1.3.3 Mechanisms of Genetic Variability

There are two main mechanisms by which variation can be generated within the *S. aureus* genome: mutations and horizontal transfer (Holden & Lindsay, 2008; Segerman et al., 2012). Firstly, mutations such as single nucleotide polymorphisms (SNPs) drive genome diversity creating phenotypic differences between strains, not just through subtle translationally disruptive nonsense or frameshift mutations. These SNPs may arise due to point mutations or through homologous recombination within the genome sequences.

Small point mutations are seen to occur 15-fold more frequently than larger scale insertions or deletions in a polypeptide sequence, but these genetic changes can also generate phenotypic effects (Feil, 2004; Feil et al., 2003; Holden et al., 2013). The changes induced by SNPs may be functionally ineffective due to redundancy in the third base pair position of the protein synthesis mechanism (synonymous substitutions) (Bosi et al., 2016; Joseph et al., 2016; Lindsay, 2010; Tokajian, 2012; Van Tonder et al., 2014). Within bacterial species, conserved genome polymorphisms are widely distributed throughout a particular repertoire of genes, which are often associate to be lineage specific as in *S. aureus*. (Holden et al., 2013; McCarthy & Lindsay, 2013). Changes (mutation, selection, insertion/deletion) to particular genes and genetic regions which show considerable variability, drive the genomic evolution of different lineages within the *S. aureus* (Roe et al., 2016; Rouli et al., 2015; Stephens et al., 2006; Ye et al., 2014).

The second mechanism by which bacterial genomes evolve is via horizontal transfer of genetic information. The transfer of DNA into and between bacteria is facilitated through mechanisms of transformation (uptake of genetic material from environment), conjugation (direct transfer between organisms) and transduction (movement of genome segments via bacteriophage) (Lindsay, 2008; Malachowa & Deleo, 2010). Conjugation is frequently seen in *S. aureus*, and the mobilisation machinery required for HGT is well characterised with numerous conjugative transposons, plasmids and bacteriophages identified. Horizontal transfer is naturally frequent between strains, and is predominantly constrained to DNA exchange from other *S. aureus*, due to the species specific phage (Jones et al., 2015). Unlike most Gram-positive bacteria, *S. aureus* is also not readily transformable due to restriction barriers, in the form of Restriction-Modification systems, present in the organism, nor does it carry the necessary genes for competence, an essential phase in bacterial transformation process (Lindsay, 2008; McCarthy & Lindsay, 2010).

Although the main mechanism by which genetic variation is introduced within a bacterial population remains differentiation by influx of novel genetic material via HGT, transfer of genetic fragments via recombination, drive the microevolution in within bacterial species. Homologous recombination (HR) plays an important role in *S. aureus* genome evolution, through the homogenisation of core genomes leading to the formation of interrelated population structures (Gonzalez-Torres et al., 2019). Successful clones form clusters of bacterial strains within dominant lineages. These lineages continually evolve independently of each other through mutation and horizontal gene transfer of mostly accessory genes facilitated by the transfer and loss of mobile genetic elements.

Large-scale chromosomal replacements within the natural populations of pathogenic bacteria, specifically clonal species is like *S. aureus*, are 15-fold less likely than genotypes diversifying by point mutation (Feil et al., 2003). Everitt et al., (2014) have shown that chromosomal regions flanking MGEs (hotspots flanking ICE*6013*, SCC, SaPIs, vSaα) and a ~750 kb sequence region spanning the origin or replication (*oriC*) have elevated recombination rates. Although most recombination prone areas flank MGEs, core genome transfers (CGT) in the form of large-scale recombination which occur through a mobile genetic elmeent-independent mechanism remain a paradox (Wilson et al., 2014; Everitt et al., 2014). These chromosomal replacements are often associated with significant fitness disadvantages (Vogan and Higgs., 2011), yet there are several successful chimeric *S. aureus* strains including ST239, ST71, ST34 and ST42. The most successful hybrid MRSA ST239 (TW20) stain, made of a CC8 backbone (80%) with a ~557 kb CC30 donor sequence spanning the origin of replication, still remains one of the most prevalent lineages of *S. aureus* globally (Robinson & Enright, 2004; Holden et al., 2010).

Large-scale recombination events have been studied in other emerging epidemic clones of other bacterial species including *S. agalactiae* (Crochet et al., 2008) and *S. pneumoniae* (Mostowy et al., 2014; Cowley et al., 2018), *K. pneumoniae* (Chen et al., 2014), and *E. faecium* (de Been et al., 2013).

1.2 PROKARYOTIC EPIGENETICS: THE UNKNOWN

One aspect of the bacterial genome that remains poorly explored is the methylome, and the role that it plays in epigenetic regulation. Epigenetics refers to changes in phenotype or gene expression cause by mechanisms other than changes in genetic code, which is identical in every cell (Casadesús & Low, 2006; Willbanks et al., 2016). These changes are usually reversible, allowing cells to maximize the utilization of their existing gene pool, without permanently altering their contents. While the study of these mechanisms and their fundamental importance in the regulation of gene expression, genetic modification, and involvement in disease is well established for eukaryotes, little is known about the impact of epigenetic processes for prokaryotic organisms, and in particular *S. aureus* (Aravind et al., 2014; Blow et al., 2016; Sarkar et al., 2016).

The main epigenetic modification for both eukaryotes and prokaryotes is DNA methylation. DNA methylation is the addition of a methyl group (CH_3), predominantly in consecutively occurring as DNA modification is 5-methylcytosine (5mC) in eukaryotes and N6-methladenine (6mA) most prevalent in prokaryotes; both 6mA and 5mC modifications are present in unicellular eukaryotes (Clark et al., 2013; Luo, Andres Blanco, Lieberman Greer, He, & Shi, 2015). Insertion of a methyl group alters the appearance and structure of the DNA sequence which can change the protein-DNA interaction of transcription machinery and other regulatory proteins within the cell (Sánchez-Romero, Cota, & Casadesús, 2015; Suzuki, 2012).

1.2.1 DNA Methylation Signatures

Epigenetic regulation is a response to dynamic changes in an organism's environment. Prokaryotes respond to subtle differences in temperature, osmolarity, nutrient availability and pH through regulation of gene expression, which enhances their adaptability (Bird, 2002; Roberts et al., 2003; Xiaodong, Cheng & Blumenthal, 2003). The main epigenetic signal in bacteria is DNA methylation of nucleotide bases: N6-methyladenine (6mA), N4 methylcytosine (4mC) and C5 methylcytosine (5mC) illustrated in Figure 1.2 (Cheng, Xiaodong & Roberts, 2001; Suzuki, 2012). Several types of bacteria and phage also involve the hydroxymethylation (hm) or glycosyl-bydroxymethylation (ghm) cytosine nucelotides in the form of 5hmC and ghmC (further discussed under Type IV RM systems) (Suzuki, 2012).
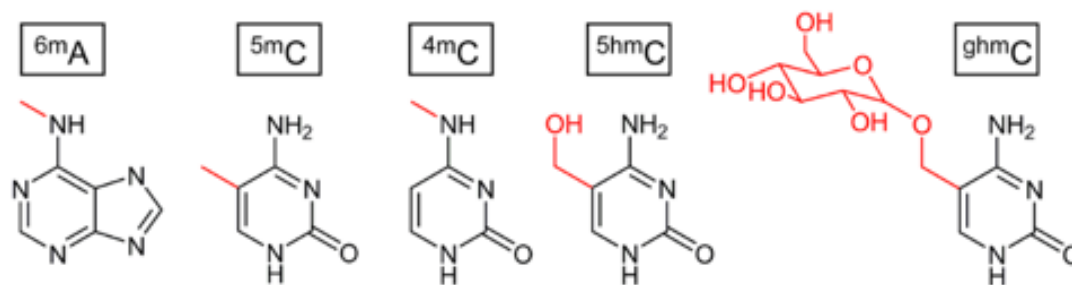
Figure 1.2 | Methylated nucleotide bases. Adapted from Suzuki, 2013.

The formation of 6mA, 5mC and 4mC is catalysed by DNA methyltransferases (MTases) which belong to a group of enzymes which catalyse the transfer of an activated methyl group from cofactor S-adenosyl-L-methionine (SAM also known as AdoMet) to a specific DNA sequence called the target recognition sequence (TRS) (Loenen & Raleigh, 2014). Methyl groups are attached to the exocyclic amino group of adenine and cytosine bases through a process called 'base flipping', during which the bases are rotated out of the DNA helix and the DNA substrate is methylate (Cheng, Xiaodong & Roberts, 2001). The addition of the methyl group to these bases does not affect the Watson Crick double helix pairing properties of the tagged adenine and cytosine nucleotides (Bierne, Hamon & Cossart, 2012). Methyl groups can be recognised by DNA binding proteins of various functions. Blow et al., 2016; S. KL et al., 2011; Kumar & Rao, 2012).

1.2.2 Methylation Machineries

In prokaryotes, most methyltransferases have been described as part of the restriction modification (RM) systems but can also exists as solitary or 'orphan' DNA methyltransferases which were probably derived from modification enzymes from ancestral RM systems (Gormley, Watson, & Halford, 2005; Kobayashi, 2005; Sánchez-Romero et al., 2015)

1.2.2.1 Restriction Modification Systems

Restriction-Modification systems (RM) have evolved as a defence mechanism by bacteria to combat invading pathogenic DNA (phage or viruses). RM systems are composed of a species-specific combination of restriction endonuclease enzymes (R/REase) which cleave foreign DNA, as well as methyltransferase (M/MTase) or modification enzymes which modify host DNA. REases recognise methylated nucleotides at distinct TRS within the organism's own genome, using these to selectively distinguish the endogenous DNA from exogenous pathogenic DNA. When restriction proteins encounter foreign DNA, usually in the form of invading phage or plasmids, that are not methylated within the specific TRS, they are cleaved by said REase at internal/non-terminal phosphodiester bonds within double stranded nucleotide chains of foreign DNA, protecting the host from infection (Gormley et al., 2005; Oliveira, Touchon, & Rocha, 2016; Roberts et al., 2003). This process is called restriction, by which the bacteria protect themselves from foreign exogenous genetic material entering the cell through horizontal gene transfer, invasion or other forms of predation (Ershova et al., 2015; Wilson & Murray, 1991). For each restriction endonuclease, there is usually a cognate modification methyltransferase enzyme which methylates the bacteria's own DNA at the previously mentioned sequence motifs and therefore protects the host DNA from cleavage (Kobayashi, 2001; Loenen & Raleigh, 2014; Roberts et al., 2003).

Although the functions of restriction-modification systems are similar across all species, their functional and structural organisation varies. There are three methylating RM system types (TI-TIII), characterised based on the REase and MTase complex composition, recognition structure or specificity subunits, location of cleavage as summarised in Figure 1.3. (Roberts et al., 2003). There are also Type IV restriction endonuclease which have methyl-directed activity, hydrolysing only methylated DNA (Loenen & Raleigh, 2014; Xiaodong, Cheng & Blumenthal, 2003).
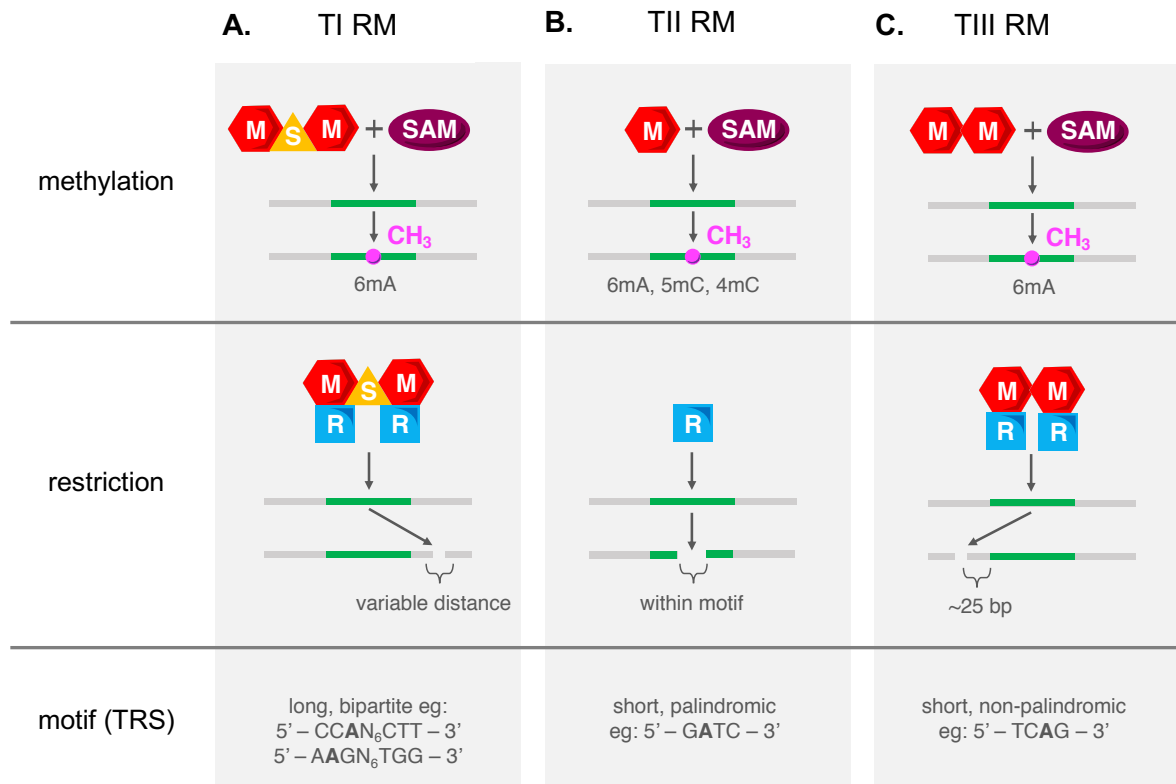
Figure 1.3 | Types of Restriction Modification (RM) Systems
Each RM system is composed of a methyltransferase (M, MTase – red), a restriction endonuclease (R, REase– blue) forming complexes of variable number ofsubunits for methylation and restriction activity at differing recognition sites (motifs) throughout a genome. Modification by all systems is S-adenosyl-L-methionine (SAM) cofactor dependent, transferring an activated methyl group from SAM to target A or C nucleotide located within the specific target recognition sites (TRS). **A.** TI RM systems are composed of an M, R and DNA binding specificity unit (S) which form a hetero-oligomeric complex for modification (M_2S) and restriction (R_2M_2S). The S subunit recognises long bipartite sequences, that is, two sub-motifs of a specific sequence separated by a fixed number of nonspecific nucleotide bases. Adenine bases within the 5' of both the forward and reverse subsequence are methylated, producing a 6mA signature. Cleavage of DNA by the recognition complex occur up to thousands of bases away from a non-methylated TRS. **B.** TII RM systems are usually composed of a single orthodox M and R (can also be homodimer) which modify and cleave DNA at short palindromic motifs. Methylation can occur on adenine (6mA) and cystine (5mC, 4mC) nucleotides. **C.** TIII RM are composed of a homodimer M complex which recognise short, non-palindromic sequences, and methylates adenine bases (6mA). DNA cleavage occurs roughly 25bp from target sequences facilitated by hetero oligomeric restriction complex (M_2R_2). Adapted from Beaurelier et al., 2019 – made with Servier Medical Art.

1.2.2.1.1 Type I Restriction-Modification Systems

Type I RM systems contain a characteristic multi-subunit protein consisting of two restriction (R), two modification (M) and one specificity (S) subunit forming a complex categorised into TIA-C, and less commonly found ID (*S. enterica, E. coli, K. pneumoniae*) based on sequence homology and genetic complementation (Murray, 2000; Gormley et al., 2005; Titheradge et al., 2001). The genes involved in the RM systems are termed as 'host specificity for DNA' (*hsd*) and the corresponding endonuclease (R), methyltransferase (M) and specificity (S) genes *hsdR*, *hsdM*, and *hsdS* respectively. Type IA and type IB genes are encoded on the bacterial chromosome, and genes encoded by Type IC RM families are predominantly on plasmids (Murray, 2000). The formation of a trimer M_2S functions as the active methyltransferase unit (MTase) whilst R_2M_2S heteropentamer holoenzyme structure is necessary for restriction (REase) activity in TI RM systems as seen in Figure 1.3 A (Dryden et al., 2001, Wilson & Murray, 1991).

TI RM system target recognition sites (TRS) are asymmetric bipartite sequences, comprising of two distinctly separated 3-5 bp half sites divided by a nonspecific degenerative sequence of typically 5-8 bp on the forward and reverse strand, eg.:

5'-**A**ACN$_6$GTGC-3' / 5'-GC**A**CN$_6$GTT-3'

(Loenen et al., 2014; Erschova et al., 2015). The S subunit is a dedicated DNA binding protein comprised of two target recognition domains (TRD) held apart by conserved alpha helices as see in in Figure 1.4. The HsdS TRDs recognize and bind the bipartite nucleotide sequence motifs to anchor protein complexes for both restriction and modification activities.

TRD1   5' …**AAC**NNNNNN**GTGC**…   TRD2
        3' …**TTG**NNNNNN**CAGC**…

B   TRD1   CR1   TRD2   CR2
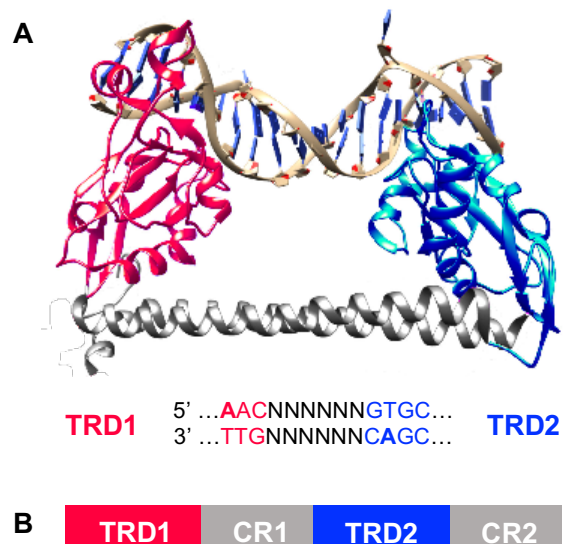
Figure 1.4 | Model of TI RM HsdS (from M.EcoKI – PDB:2Y7H) bound to DNA.

HsdS made up of two target recognition domains (TRD1:red, TRD2:blue) held apart by two conserved helices (gray: CR1, CR2). **A.** The HsdS protein binding double stranded DNA sequence with two globular TRDs at corresponding half string sequences on reverse and forward DNA strand. **B.** Protein domain organisation of HsdS – TRD1 – CR1 – TRD2 – CR2.
Modified from Loenen et al., 2014**.**

When M$_2$S complex is bound at a specificity TRS, each methyltransferase protein catalyzes the methylation reaction of an adenine nucleotide within each half sequences of the bound TRS with cofactor SAM acting as the methyl group donor. Like other methyltransferases, TI HsdM flip the target adenine nucleotide residue out of the DNA helix to facilitated methyl transfer (Su et al., 2005). Type IA MTase are maintenance methyltransferases, which prefer to methylate hemimethylated DNA substrates in which only one of the bipartite sequences are methylated. TIB methyltransferases prefer to methylate unmodified double strands of the DNA (Loenen et al., 2014).

The REase enzyme complex cleaves dsDNA thousands of base pairs away from unmethylated target recognition site (Loenen et al., 2014, Blumenthal & Cheng, 2002). DNA is cleaved by introducing a double strand break typically in a distant or random location downstream. This nuclease reaction is ATP and Mg^{2+} dependent, hydrolysing a substantial amount of both cofactors before the cleavage event, and cofactor SAM acts as an allosteric effector (Roberts et al., 2003).

Some examples of TI systems are: EcoKI (*E. coli* - Roer et al., 2015), StySBLI (*Salmonella spp*. - Kasarjian et al., 2004) KpnBI (*K. pneumoniae* - Chin et al., 2004), Sau1 (*S. aureus* - Waldron & Lindsay, 2006)

1.2.2.1.2 Type II Restriction-Modification Systems

Thus far, all genes coding for restriction enzymes for T II RM systems have been located adjacent to the gene coding for methyltransferase, either in a head-to-head, tail-to-tail or tail-to-head orientation; some may be fused into a single composite gene (Gormley et al., 2005, Roberts et al., 2003). In TII RM systems, the REase is present as a single monomer but in some cases will form an R$_2$ homodimeric complex for endonuclease function (requiring cofactor Mg^{2+}) cleaving ssDNA or dsDNA in variable fragmentation patterns. The methylation subunits act as single proteins and require SAM as a substrate for methylation (Figure 1.3 B) (Gormley et al., 2005, Vipond & Halford, 1995; Halford, 2001). The MTase and REase function independently from each other and can recognise TRS which may be asymmetric, palindromic, non-palindromic or bipartite. Both cleavage events and methylation happen within the recognition sequence or in near proximity thereof. TII methyltransferases modify a variety of nucleotide bases including 6mA, 5mC and 4mC (Pigoud & Jeltsch, 2001). Thus, there are numerous subdivisions of Type II RM systems including TII A-G, D-G, M, P, S, T, categorised based on the MTase target, the REase

enzyme structure, and the defined fragmentation pattern (staggered, blunt double stranded cut) of DNA cleavage usually within the target recognition sequence further defined by Roberts et al., 2003.

Some examples of TII systems are: HindIII (*H. influenzae* - Tang et al., 2000), EcoRI (*E. coli* - O'conner & Humphreys, 1982), BamHI (*E.coli* and *B. subtilis* - Ives, Nathan & Brooks, 1992), AhdI (*A.hydrophilia* - Streeter et al., 2004).

1.2.2.1.3 TIII Restriction-Modification Systems

Type III restriction modification systems are not as well studied due to occurring more seldom across bacterial species. They consist of two subunits usually coded for by a *mod* (modification) and *res* (restriction) gene located within the same operon. The methyltransferase can function as dimer M_2, (Figure 1.3 C) but can also form a heterodimer complex with two R, M_2R_2 to form a bifunctional MTase/REase enzyme. TII MTase methylate adenine bases (6mA) using SAM as a methyl group donor, but only on one strand of the duplex DNA helix at a time, which adequately interferes with the restriction activity on the host DNA (Rao et al., 2013).

The REase will only be active when associated with M subunits, which provides the sequence specificity for both enzyme complexes. The two enzymatic activities therefore compete for the asymmetric DNA recognition sites, but two of such 5-6 bp sites must be next to each other in opposite orientation for restriction to take place. If there is a single recognition sequence or the fragments are in the same symmetric orientation, restriction will not occur. The cleavage reaction also requires cofactor Mg^{2+}, ATP cleaving dsDNA 25-30 base pairs downstream from the recognition site (3') (Rao et al., 2000, Gromley et al., 2003). For the restriction activity to happen, two molecules of the RM enzyme complex must bind to two recognition sites with opposite orientation, where they undergo unidirectional translocation until the two enzymes face each other; cleavage upon protein-protein interaction between the two molecules (Gromely et al., 2003).

Some examples of TIII systems are: StyLTI (*S. typhimurium* - Backer & Colson, 1991), HinfIII (*H. influenzae* - Yuen & Hamilton, 1984), BceSI (*B. cereus* - Xu et al., 2012), NgoAXP (*N. gonorrhoeae* - Adamcyk-Poplawska et al., 2009)

1.2.2.1.4 Type IV Restriction Endonuclease

Type IV restriction modification systems are not true RM systems, as they only have restriction activity and most often consist of only a single REase. These restriction enzymes function in methylation dependent manner hydrolysing only unmodified DNA at distinct TRS. The enzymes are essentially solitary restriction enzymes which are fused to a methyltransferase similar to those in Type II RM systems also having asymmetrical recognition sequences which shift cleavage positions. However, these restriction enzymes are mostly non-specific and have more variable cleavage sites (Suzuki, 2013). In addition to acting on methylated bases 6mA, 5mC, 4mC, several Types of IV restriction system involve activities on hydroxymethylated (hm) or glycosyl-bydroxymethylated (ghm). One of these nucleobases is 5-hydroxymethylcytosine (5hmC) which is usually present in bacteriophage, but have been incorporated during phage DNA replication, as well as the further glycosylated product β-glucosyl-5-hydroxymethylcytosine (ghmC) (Roberts et al., 2003; Blumenthal & Cheng, 2002).

The TIV REase family of enzymes is greatly diverse and only a few have been characterised in detail, one of which is EcoKMcrBC. This system is comprised of two subunits - McrB responsible for DNA binding and hydrolysis of GTP and McrC which holds the catalytic domain (nuclease moiety) of the system responsible for DNA cleavage respectively. DNA is translocated and cleaved by the McrBC complex which requires GTP hydrolysis as well as Mg^{2+} cofactor ions, resulting in a double stranded break 30-35 bp away from the modified site (Loenen & Raleigh, 2014). Other well characterised TIV restriction enzymes including Mrr (*E. coli* - Bourgess et al., 2017) and SauUSI (*S. aureus* - Xu et al., 2011).

1.2.2.2 Solitary 'Orphan' DNA Methyltransferase

Some methylases do not belong to an RM system but carry out the same function. These solitary methylases do not form MTase complexes but add methyl groups to newly synthesized daughter strands subsequent to replication on their own. Usually, the parent strand is used as a template for the location specific methylation of both DNA strands. Some sequences may only have one, hemi-methylated strand which acts as a signal for various DNA interacting proteins for example: SecA, which recognises GATC sites near the origins of replication (*oriC*) at initiation of replication (Jost & Saluz, 1993).

One of the most in depth studied solitary DNA methyltransferases, is Dam which recognises 5′GATC3′ motifs in the DNA of gamma-proteobacteria and methylates the adenosine nucleotide moiety of the short recognition sequences. The cell cycle regulated (CcrM) methyltransferase is another important orphan modification enzyme which was first identified in *C.crescentus* and usually found in most alpha-proteobactereia. These modifying proteins methylate the adenine of the 5′GANTC3′ (where N is a variable nucleotide) motif sites (Kumar & Rao, 2013). Dam and CcrM are of independent evolutionary origin but have been both studied to investigate the control transcription in a DNA methylation-dependent fashion (Sanchez-Romero, Cota & Casadesus, 2015).

1.2.3 Methods off DNA Methylation Detection

Currently over 4,000 RM enzymes with over 400 different specificities from +5000 bacterial and archaeal genomes have been characterized and deposited in the REBASE database (Beualeurier et al., 2019, Roberts, Vincze, Posfai, & Macelis, 2015; Vasu & Nagaraja, 2013). Traditional RM discovery methods were limited to mostly identifying TII and solitary methyltransferases modifying cystine residues. With the advent of novel third-generation sequencing technologies, the full methylome landscapes of bacteria have been characterised detecting 6mA, 5mC and 4mC signatures and their target recognition sequences (Beaulauerier et al., 2019).

1.2.3.1 Legacy Methods

Most of the methodologies developed for DNA methylation detection have been mainly dedicated to characterising 5mC in eukaryotes, as the biological significance of cytosine modifications within mammalian cells and their involvement in human disease have been recognised for over half a century (Razin & Riggs, 1980; Robertson, 2005; Lebedev & Sazhenova; 2008; Waggoner, 2007). These approaches rely on affinity enrichment of methylated DNA fragments, digestion of methyl cytosine restriction enzymes, or chemical conversion of methylated cytosine residues using sodium bisulphite (Zhang et al., 2014; Bock, 2012; Hirst & Marra, 2010; Laird, 2010). These methods are limited to only detecting modified 4mC and 5mC residues, and cannot detect adenine methylation, which is the most prevalent epigenetic signature in bacteria (Casadesus & Low, 2006).

Adenine methylation has been studied within a known sequence context by specific methyltransferases such as Dam and CcrM, using restriction analysis by restriction

endonuclease enzymes DpnII and Mbol (Dam) and HinfL(CcrM) which cleave DNA at unmethylated GATC and GANTC sites (Lacks & Greenberg, 1977; Zweiger, Marcynski & Shapiro, 1994). Restriction enzyme-based studies were limited to known methylation motifs (mostly of TII RM systems with precise restriction activity at methylated target sequences) with known or partially characterised matches of known methyltransferases and restriction enzymes. Therefore, this approach did not lend itself for *de novo* discovery of prokaryotic methylation motifs (Beaulaurier et al., 2019).

Studying epigenetic regulation within prokaryotes relies on the ability to identify modified nucleotides within short methylation motifs recognised by solitary or RM methyltransferase. Several early sequencing methods using modified traces in dye-terminator Sanger sequencing were developed for direct detection of DNA methylation at unknown sites (methylation motifs) (Rao & Buckler-White, 1998; Bart et al., 2005; Wood et al., 2007; Li et al., 2007). These methods relied on variations in the amplitude of fluorescent peaks in sequencing trace for methylated nucleotides having the potential to detect 4mC, 5mC and 6mA (Broadbent et al., 2007; Bart et al., 2001). However, due to technical limitations often including subtle peak signature and low throughput of Sanger sequencing, these methods were not adopted for wider use (Korlach & Turner, 2012).

Single cell bisulphite sequencing has become a gold standard for Illumina short read sequencing providing high quality methylation data for modified cytosine bases within bacterial genomes (Gouil & Keniry, 2019) and novel long read sequencing techniques offer the possibility to study 6mA, 5mC and 4mC modifications in a single cell and microbiome setting (Blow et al., 2016; Tourancheau et al., 2020)

1.2.3.2 Next Generation Methods - Direct Detection Using Long-Read Sequencing

Recent advances in 'third-generation' sequencing technologies have enabled the comprehensive study of a wide range of prokaryotic nucleotide modifications through long-read single-molecule real time (SMRT) sequencing. These technologies include the well-established PacBio SMRT sequencing approach (Fang et al., 2012; Beaulaurier et al., 2015; Blow et al., 2016; Flusbery et al., 2010; Murray et al., 2012; Schadt et al., 2013) and emerging Oxford Nanopore Technologies sequencing method (Jain et al., 2016; Clarke et al., 2009; Manrao et al., 2013; Manrao et al., 2011; Laszlo et al., 2014; Laszlo et al., 2013; Deamer et al., 2016).

1.2.3.2.1 Established Method - Direct Detection using PacBio SMRT Sequencing

SMRT sequencing is commercially available with RSII and Sequel instruments manufactured by Pacific Biosciences. PacBio SMRT sequencing was the first third-generation sequencing approach which has allowed the characterisation of prokaryotic methylomes. The sequencing-by-synthesis technology is based on real time monitoring of the incorporation of fluorescently tagged nucleotides during replication, synthesised along a single stranded circular template strain, called the SMRTbell (Figure 1.5 A). Each incorporated flourophore-phospholinked deoxynucleotide triphosphate (dNTP) is held shortly at the immobilised polymerase active site, at the bottom of the SMRT cell, as it associates with the template DNA strand (Figure 1.5 B). During this short duration of time, the conjugated fluorophore is excited through a light pulse which identifies the base (different colours for A, T, G and C), and the fluorescent signal is emitted within the reaction cell are recorded. The dye-linker-pyrophosphate product of base incorporation is cleaved at the triphosphate chain from the associated nucleotide which ends the fluoresce pulse, and the polymerase translocates to the subsequent position on the SMRTbell, initiating the next fluorescent pulse (Rhoads & Au, 2015). The timing of these pulses, corresponding to the incorporated bases allows the analysis of methylated bases and the kinetics of the DNA synthesis. This is evident through the increased inter-pulse durations (IPD) for modified bases, as they slow processing by the polymerase seen in the successive fluorescence signal data collected (Figure 1.5 C).

Over the past decade, SMRT sequencing has become an established analysis tool (PacBio SMRT Motif and Modification Analysis) for the detection of mostly 6mA and 4mC and methylation motifs with high confidence as they produce strong kinetic signatures. With the introduction of this technology, there was a huge increase (n = >300) in the characterisation of various RM systems. In particular SMRT technology has great importance in the discovery of 6mA MTase specificities belonging to TI and TIII RM which have been challenging to study with previous restriction enzyme digest methods, as the restriction endonuclease complexes cleave DNA variable distances from the methylated target motif site. However, the direct detection of 5mC and 5hmC is subtle, hence needing enrichment by glycosylation or TET-conversion to 5-carbosylcytosine to produce larger kinetic effects (Pacific Biosciences – white paper; Clark et al., 2013). Nevertheless, a multitude of prokaryotic methylomes have been characterised with this technique and has greatly expanded our knowledge of methyltransferase (Beaularelier et al., 2019).
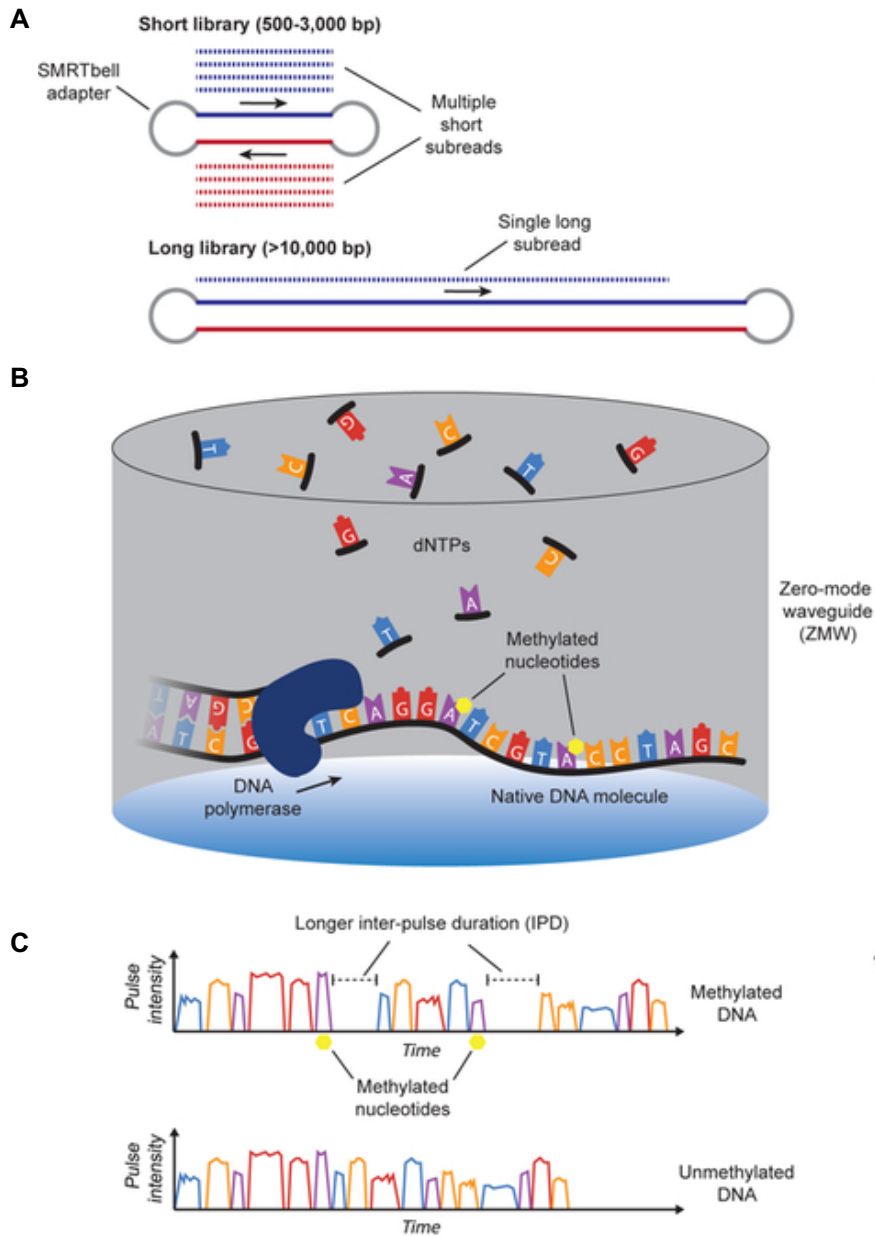
Figure 1.5 | Third generation SMRT sequencing for detection of DNA methylation
A. sequencing libraries for SMRT sequencing by PacBio involve double-stranded DNA fragments flanked by ligated SMRTbell hairpin adapters, permitting DNA polymerase to process both strands of the template DNA for both short and long inters libraries. **B.** SMRT sequencing via a sequence-by-synthesis approach in which a DNA polymerase is immobilised on the bottom of a zero-mode-waveguide (ZMW) nanophotonic sequencing chamber and uses a strand from the native sequencing library as a template for the read, sequentially incorporating fluorescently phospholinked deoxyribo-nucleosides triphosphate (dNTPs). Each incorporated dNTP is held at the polymerase active site for a short duration and emits a fluorescent pulse in the corresponding colour channel. C. The order of pulses informs the read sequence, with pauses between pulses called inter-pulse durations (IPDs) signify the presence of a modified base in the template DNA. Adapted from Beaulaurier et al., 2019)

1.2.3.2.2 Emerging Method - Direct Detection using Nanopore Sequencing

In addition to SMRT-based technologies, nanopore-based sequencing has shown promising capabilities in detecting methylation modified bases. Nanopore sequencing is commercially available through Oxford Nanopore Technologies (ONT) sequencing platform (Dreamer et al., 2016). This technology measures the variation in ionic current of a negatively charged single-stranded nucleic acids (ssDNA) processed through a lipid membrane embedded biological nanopore which has a voltage applied to it (Figure 1.6 B). There are multiple protocols for library preparation involving ligation of adaptors to the ends of double stranded DNA (dsDNA) fragments, recognise and couple to motor proteins which rachet the ssDNA strands through the nanopore at a fixed rate during sequencing (Figure 1.6 A and B) (Jain et al., 2016). Sensors monitor the ionic current during this process and detect fluctuations as a function of 4-6 nucleotides occupying the constricted nanopore channel at a given moment (Figure 1.6 C).   The detected currents are subsequently analysed by a recursive neural network to build the corresponding sequence of a bases within the read (Ip et al., 2015; Laszlo et al., 2014; de Lanoy et al., 2017).
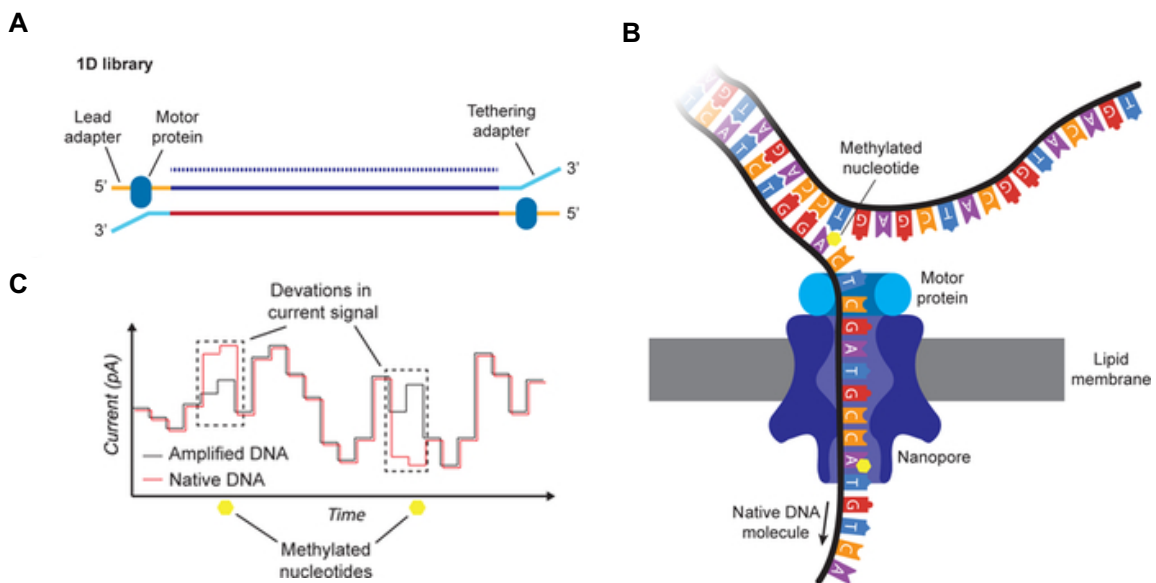
Figure 1.6 | Oxford Nanopore Technologies (ONT) next generation sequencing and real time methylation detection. A. ONT 1D library prep involves adapters being ligated onto both ends of DNA fragments – in the form of lead (5') and tethering (3') adaptors. **B.** A processive motor protein captures the lead adapter on a strand to co-locate molecules near the engineered biological nanopore (embedded in lipid membrane) sequencing of a single DNA. A voltage potential applied across the membrane and the single DNA strand (ssDNA) is processed through the nanopore. C. ionic current flowing through the nanopore depends on a precise number of nucleotides (5, or 4) which occupy the constriction point within the pore. Methylated nucleotides within the processed ssDNA create distinct current patterns enabling us to distinguish between modified and unmodified bases. Adapted from Beaulaurier et al., 2019)

Theoretically this approach allows the detection of all different types of base modifications, but developments of this technology have mainly been focused on eukaryotic applications in detecting 5mC and 5hmC until recently (Simpson et al., 2017). Rand et al., have developed a variable hidden Markov model (HMM) tool called signalAlign to identify variable methylation events within prokaryotic genomes demonstrating the feasibility of nanopore-based methylation detection (Rand et al., 2017). Paired with a hierarchical Dirichlet process (HDP) and trained with the correct training population, showed promising sensitivity in distinguishing between methyl-cytosine read-level 6mA detection in bacterial genomes (McIntyre et al., 2019). Several other tools with differing algorithms have been demonstrated to detect 6mA and 5mC including Deepsignal (Ni et al., 2019), DeepMod (Liu et al., 2019) Megalodon (Megalodon GitHub), and most recently Guppy (nanopore sequencing data analysis). However, these model-based approaches remain limited in their ability to *de novo* identify correlating modification sequence motifs (Stoiber et al., 2019).

1.2.4 The Role of DNA Methylation in Bacteria

As in eukaryotes, post-replicative DNA methylation in bacteria have been shown to play a role in the epigenetic control of various cellular functions. DNA methylation acts as a cellular regulatory signal recognised by proteins involved in cellular defence, replication, DNA repair, and transposase activity. Modification of DNA at regulatory regions also plays a role in control of gene expression via different epigenetic mechanisms.

1.2.4.1 DNA Methylation Dependent Host Functions

Methylation as part of a RM system acts as a positive feedback mechanism within the host cell. Methylation of the host chromosome at distinct target motif sites, is recognised by both the MTase and REase unit. Firstly, this protects the cell from cleaving self-DNA and stabilises the host chromosome (Furuta et al., 2014). Loss of methylation status may induce large scale genome rearrangements following accidental RM mediated lethality in the host or result in cell death (Bickle & Krüger, 1993). Secondly, RM mediated methylation plays and important role in cellular defence as exogenous non-self DNA with variable methylation signatures and unmethylated target sites are recognised and cleaved by the REase (Vasu & Nagaraja, 2013). This protects the host from infection but also controls horizontal gene transfer of various mobile elements (Corvahla et al., 2010; Kuroda, 2001;

Lindsay, 2010). The interplay between DNA modification therefore maintains the species identity and genetic isolation, modulating the rate of their evolution (Jeltsch, 2003; Arber 2000).

Hemi-methylated DNA can act as a signal for various cellular processes involving different methylases (Loebner-Olensen et al., 2005; Casadesus & Low, 2006). The most well studied such modification unit is Dam, facilitating 6mA methylation at 5'-GATC-'3 recognition sites in various species. Hemi-methylated GATC sites play an important role in chromosomal replication control in *E. coli*. SeqA protein, which negatively regulates the initiation of DNA replication at the origin of replication (*oriC*) by preferentially binding hemi-methylated DNA target sites (GATC) around the *oriC*, hindering replication initiation by blocking DnaA binding (Bogan et al., 1997; Campbell & Kleckner, 1990; Han et al., 2003; Molina & Starstad, 2004; Taghbalout et al., 2000; Casadesus & Low, 2006; Lobner-Olensen et al., 2005). Hemi-methlyated DNA *in E. coli* also has a part in DNA repair, which methyl-directed mismatch repair protein MutH binds to cleave non-methylated daughter DNA strands generated by semiconservative replication. By cutting the unmethylated strand, only the methylated parental strand can be used as a template for repair associated DNA synthesis (Au, Welsch, and Modrich, 1992; Bakker & Smith, 1989; Baghwat & Lieb, 2002; Schlagman et al., 1986). Further bacterial functions are impacted by DNA methylation acting as an epigenetic regulator of gene expression.

1.2.4.2 Epigenetic Mechanisms of Gene Expression Regulation

Prokaryotic DNA methylation and the modifying MTase units can induce changes in gene expression resulting in differential phenotypes. These epigenetic mechanisms include interference of DNA binding proteins through methylation at specific regulatory sites and competitive binding between MTase and transcription factors (TF) or by modulating the MTase machinery through phase variation detailed in Figure 1.7 (Beaulaurier et al., 2019).

Methylation of specific target motifs that lie within promoter regions of a given gene can hinder the binding of transcription factors, thereby repressing local transcription of the downstream genes or operons (Figure 1.7 A). The transcription of bacterial loci can be affected by 6mA modifications directly altering interactions with regulatory proteins by: 1) direct alteration of DNA-protein interactions of regulatory TF either by direct steric effect (methylation blocking active site) or 2) indirect effect on DNA structure (changes in DNA curvature) (Diekmann, 1987; Blyn et al., 1989; Polaczek et al., 1998; Low et al., 2001).
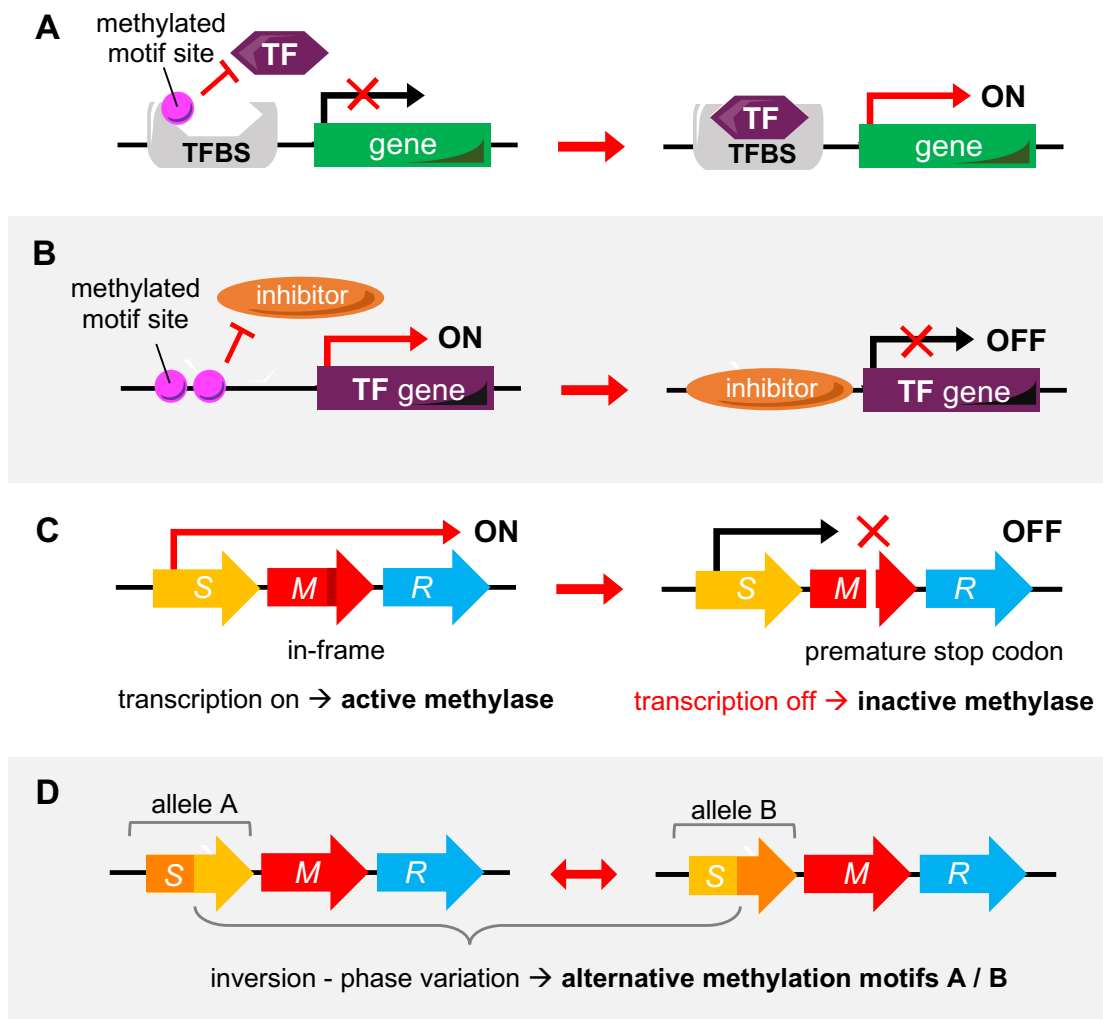
Figure 1.7 | Prokaryotic epigenetic mechanisms
A. site specific competitive binding mechanism in which methylation motifs that sit within the upstream regulatory promoter can affect the expression of the downstream gene. Methylated bases within this region can interfere with regulatory proteins like transcription factors (TF) from binding the transcription factor binding sites (TFBS) preventing the transcription of the downstream gene. **B.** if a methylation event occurs within the promoter region of a TF encoding gene with promiscuous DNA binding specificity, local methylation may hinder transcription inhibition resulting in a transcriptional cascade. Regulatory proteins bind non-methylated promoter sequences with highest affinity forming DNA methylation patterns (DMP) overlapping methylation motif targets inhibiting their methylation. Undermethylation of TRS motif clusters are relevant as they may lead to downstream blockage or decrease in gene expression if inhibitors bind to unmodified TF promoters **C.** MTase ON/OFF switch – phase variable methyltransferases can induce a genome wide changes in methylation status due to spontaneous or reversible mutations inducing premature stop codons in the gene encoding the M subunit. Cells with inactive or non-functional forms of methylases lack methylation, but also restriction activity in TI and TIII RM which need full length M subunits for REase complex formation and endonuclease activity. This may lead to clonal expansion of a given bacterial species with divergent methylation expression and variable accessory genome due to lack of host defence with the loss of restriction activity. **D.** Other genetic rearrangements like inversions, specifically in genes encoding specificity subunits (in TI and TII RM) can produce multiple S alleles, resulting in variable expression due to multiple motif targets for methylation. Adapted from Beaulaurier et al., 2019 – created with Servier Medical Art)

Methylation within consensus promoter regions and alteration of DNA interaction with the transcription machinery inhibits the transcription of some proteins including *trpR*, (tryptophan repressor controlling amino acid metabolism in *E. coli* - Peterson et al., 1985, Marinus, 1985), transposases (limiting transposition of Tn*903*, Tn*5*, Tn*10*, IS*10* and IS*50* - Roberts et al., 1985; Dodson & Berg., 1989; Haniford et al., 1989), Agn43 (outer membrane protein - regulating autoaggregation in liquid by blocking binding of OxyR (oxidative stress response regulator), *traJ* encoded transcription factor active in conjugative virulence plasmid pSLT in *S. enterica* contributing to the repression of mating by blocking leucine-responsive regulatory protein, Lrp transcription factor (Camacho & Casadesus, 2005). In exceptional cases DNA methylation of promoter regions can also enhance transcription of some regulators like DnaA (Braun & Wright, 1986).

Methylation motifs within promoter regions may also induce a transcriptional shift by local competition between MTase and other regulatory proteins. The canonical example of this competition model is 6mA methylation at GATC sites by Dam methyltransferase in *E. coli* (Casadesús & Low, 2006). Methylation motif clusters directly upstream and within promoter region of the pyelophritis-accociated pili (*pap*) operon *papBA* (containing 6 Lrp TF binding sites with, sites 1-3 containing 2 GATC motifs at site 2, and sites 4-6 (upstream of sites 1-3) containing GATC in site 5) locally slow the processivity of Dam methyltransferase, increasing the time of DNA binding proteins Lrp (leucine-responsive regulatory protein) have access to their target sites (Woude, Braaten & Low, 1996; Adhikari & Curtis, 2016). Differential methylation (non / hemi / full) of Lrp binding sites, affects at which site Lrp is bound, with binding at sites 1-3 inhibiting RNA polymerase binding at the promoter preventing the transcription of *papBA*, creating a phase-variable, reversible transcriptional switch, resulting in Pap pili (ON state) or without Pap pili (OFF state) within a bacterial population (Wion & Casadesus, 2006).

Several studies using microarray analysis have been performed to study the role of 6mA methylation in global gene regulation through alteration of transcription of global regulatory proteins or through perturbation of DNA MTases (Adhikari & Curtis, 2016). An example of this cell cycle regulation by CcrM methylase in *C. cresentes*, which regulates itself by self-methylation at GAnTC motifs within the promoter region of *ccrM,* inhibiting transcription and therefore methylation activity (Stephens et al., 1995). 6mA methylation by CrrM is essential for chromosomal replication, regulating the transcription of three unstable regulators DnaA, GcrA, and CtrA. The transcriptional cascade triggered by these regulators depends on the methylation state of the promoter sequences, activating the

sequential transcription (hemimethylation) as well as and repression of each of these proteins (Collier et al., 2007). Phase variable methylation of regulatory GAnTC motifs can inhibit the activity of regulatory proteins due to fully methylated motifs instead of preferential unmethylated or hemi-methylated sites, as in the case of GcrA failing to activate the transcription of CtrA, and CtrA failing to repress the transcription of GcrA (Figure 1.7 B). This may lead to completely altered transcriptional profiles as these global regulators control the transcription of over 100 genes in *Caulobacter* (Laub et al., 2002; Fioravanti et al., 2013, Murray et al., 2013, Gonzalez et al., 2014).

Differential methylation through alteration of MTase activity (Figure 1.7 C) or MTase specificity (Figure 1.7 D) can also lead to variable transcription profiles and epigenetic heterogeneity. Hypervariable mutations within the regulatory and coding sequence regions encoding MTase machinery will result in differential methylation and impact the global expression patterns from one cell to another (Henderson et al., 1999; Shinkhanta et al., 2005; Atack et al., 2018). Phase variable MTases have been observed in a wide range of bacteria within TIIG RM system in *C. jejuni* (Anjum et al., 2016), and TIII *mod* genes in *H. pylori* (De Vries et al., 2002; Srikhanta et al., 2011), *H. influenzae* (Zaleski et al., 2005; Fox et al., 2007), *P. haemolytica* (Ryan et al., 1999) and *Neisseria* species (Sein et al., 2011; Gawthorne et al., 2012). Phase variable motif specificity by inverting TI *hsdS* loci have also been shown in a host of species including *S. pneumoniae* (Tettelin et al., 2001), *M. pulmonis* (Dybvig et al., 1998), *L. monocytogenes* (Furuta et al., 2014), *S. suis* (Willemse et al., 2001; Li et al., 2016) and *Lactobacillus* species (O'sullivan et al., 2000, Claesson et al., 2006) among others. The resulting global differential methylation states have been linked to a number of virulence phenotypes including biofilm formation, host immune evasion and antibiotic resistance (De Ste Croix et al., 2017, Phillips et al., 2019).

Although technological advances like SMRT sequencing have vastly contributed to our ability to detect methylation events and discover methylation motifs, more comprehensive studies mapping methylation as a function of distinct MTases and characterising the precise mechanisms of gene expression modulation via whole genome transcriptome (RNASeq) and mutagenesis studies is necessary to fully quantify and understand variable methylomes and altered bacterial phenotypes.

1.3 STUDY AIMS & MAIN QUESTIONS

DNA methylation, in particular 6-methyl adenine modification (6mA), is known to act as an epigenetic regulatory signal for protein-DNA interactions involved in replication, cellular defence, DNA repair, cell cycle regulation, gene expression cascades and virulence in various prokaryotic species. With the advent of next generation methodologies like PacBio SMRT sequencing, increasing numbers of 6mA signatures and corresponding methyltransferases (predominantly solitary or Type I Restriction Modification methylation units) have been identified. Multiple mechanisms of 6mA epigenetic control have been described (site specific competitive binding at promoters, transcriptional cascade hinderance, phase variability) leading to not just differential gene expression, but also alternative methylation motifs throughout a population of variable bacterial species.

Sau1 is the main Restriction-Modification (RM) system in *Staphylococcus aureus*, a TI RM with a 6mA signature. The role of this system in host defence has been well established, but not much is known about the variability of Sau1 within *S. aureus* lineages, distribution of 6mA modifications through the whole genome, or the role of 6mA methylation signature as a potential secondary and/or epigenetic regulator within the species.

Hence, the first aim of this study was to characterise the Sau1 system throughout the *S. aureus* species focusing on the DNA binding specificity units (*hsdS*) and resolving each corresponding methylation motif / target recognition sequence (TRS) predicted by PacBio SMRT sequencing. How variable is Sau1 HsdS (pertaining to DNA binding target recognition domains (TRDs)) and 6mA methylation motifs (TRS)? Are *hsdS* alleles linage specific? Is there suggestion for phase variability of *hsdS*? Is there any evidence for differential methylation attributed to Sau1 within a strain in regard to *hsdS* alleles?

The second aim of this study was to characterise the methylation frequency of the whole genome to investigate any 6mA methylation biases, hot/cold spots in various genetic locations to gain a wholistic picture of the 6mA landscape throughout the species. Are there parts of the genome that are more densely methylated including: the origin or replication, the coding sequence, the intergenic region, or mobile elements? Are there differences in methylation frequency for different HsdS giving insight to a potential functional bias of a specific system (HsdS_α or HsdS_β)?

The third aim of this study was to investigate the role of Sau1 6mA methylation in gene expression, and potential epigenetic regulation. Does differential 6mA methylation have a differential gene expression effect? Do largescale genome rearrangements affect DNA methylation signatures? Does deletion of any or all *hsdS* and resulting loss of 6mA methylation signature have an effect on gene expression? Does Sau1 have an epigenetic regulatory role in *S. aureus*?

2. METHODS

2.1 CHEMICAL REAGENTS AND KITS

All reagents used were of analytical grade. A list of chemical solutions, buffers antibiotics and reagents used in Polymerase Chain Reaction (PCR) and Gel Electrophoresis can be found in Table 2.1 detailing supplier and working concentrations.

Table 2.1 | Buffers, Antibiotics, Enzymes, PCR Reagents, Electrophoresis Reagents

Buffers	Supplier	Composition
Lysis Buffer	Millipore, UK	20mM Tris HCl
	Sigma-Aldrich, UK	2mM EDTA
	Sigma-Aldrich, UK	1% Triton X-100
Tris-acetate-EDTA (TAE)	Millipore, UK	2M Tris HCl
	Sigma-Aldrich, UK	0.10M EDTA
	Sigma-Aldrich, UK	1M acetate
Electroporation Buffer	Sigma-Aldrich, UK	10% glycerol
	Sigma-Aldrich, UK	0.5M Sucrose
		ddH$_2$O
Antibiotics	**Supplier**	**Concentration / Solvent (S)**
Chloramphenicol (Cm)	Sigma-Aldrich, UK	10 µg/ml (S - 70% EtOH)
Anhydrotetracycline (ATC)	Sigma-Aldrich, UK	1 µg/ml (S - ddH$_2$O)
Ampicillin (AMP)	Sigma-Aldrich, UK	100 µg/ml (S - ddH$_2$O)
Polymerase Chain Reaction	**Supplier**	**Concentration**
dNTPs	Thermo Fisher, UK	10 mM
Forward Primer (1:10)	Biomers, Germany	10 pmol/µl
Reverse Primer (1:10)	Biomers, Germany	10 pmol/µl
Phusion Hot Start II High Fidelity DNA Polymerase	Thermo Fisher, UK	2 U/µl
Phire II DNA Polymerase	Thermo Fisher, UK	2 U/µl
Reaction Buffer x6	Thermo Fisher, UK	
Gel Electrophoresis	**Supplier**	**Concentration**
Agarose	Bioline GmBH, Germany	1% (in 1x TAE)
Tris-acetate-EDTA (TAE)	see under 'Buffers'	1%
SYBR SAFE DNA Gel Stain	Invitrogen, UK	X10,000
TriTrack DNA Loading Dye (6X)	Thermo Fisher, UK	6X
GeneRuler 1kb DNA Ladder	Thermo Fisher, UK	
Enzymes	**Supplier**	**Concentration**
KpnI+ 10x Buffer KpnI	Thermo Fisher, UK	10 U/µl
SacI + 10x Buffer SacI	Thermo Fisher, UK	10 U/µl
DpnI + 10x Buffer Tango	Thermo Fisher, UK	10 U/µl
10x FastDigest Green Buffer	Thermo Fisher, UK	10 U/µl
Lysostaphin	Sigma-Aldrich, UK	5 mg/µl
T4 DNA Ligase + Buffer x10	Thermo Fisher,UK	5 U/ µl
DNA and RNA Extraction Kits	**Supplier**	
MasterPure™ Gram + DNA Purification Kit	Epicentre, Lucigen, UK	
RiboPure RNA Purification Kit	Ambion, Invitrogen, Thermo Fisher, UK	
NucleoSpin Microbial DNA Kit	Macherey-Nagel, Germany	
DNA Clean and Concentrator-5	Zymo Research, UK	
PlasmidPlus Midi Kit	Quaigen, UK	
GeneJet Plasmid Miniprep Kit	Thermo Fisher, UK	

2.2 LABORATORY EQUIPMENT AND CONSUMABLES

A list of equipment and consumables for the study can be found in Table 2.2.

Table 2.2 | Kits, Equipment and Consumables

Laboratory Equipment	Supplier
Fast-Prep-24 Tissue and Cell Homogeniser	MP-Biomedicals, UK
Precellys 24 Tissue Homogeniser	Bertin Instruments, UK
Labnet Z 233 M2 Microcentrifuge	Hermle, Germany
Eppendorf 5424 Microcentrifuge	Eppendorf, Germany
Heraeus Biofuge Stratos Centrifuge	Thermo Fisher, UK
Heraeus Fresco 21 Centrifuge	Thermo Fisher, UK
Eppendorf Thermomixer Compact	Eppendorf, Germany
Qubit 3.0 Fluorometer	Illumina, UK
Horizontal Electrophoresis Chamber Midi	Labortechnik, Germany
Consort Power Supplies	Labortechnik, Germany
BioDoc Analyse (gel documentation)	Biometra
Pulse Controller and Gene Pulser (Electroporator)	Biorad
PCR Thermocycler T3	Biometra

Consumables	Supplier
1mm electroporation cuvette	Sarstedt, UK
2 ml microcentrifuge tubes	Eppendorf, UK
1.50 ml microcentrifuge tubes	Axygen, UK
0.20 ml PCR tubes	Thermo Fisher, UK
2 ml screw cap cryo tubes	Thermofisher, UK
50 ml conical Falcon tubes	Greiner Bio-One, UK
15 ml conical Falcon tubes	Greiner Bio-One, UK
Disposable inoculation loops	Greiner Bio-One, UK

2.3 GROWTH MEDIA

A list of growth media and nutrient composition for the study can be found in Table 2.3. All *S. aureus* strains for sequencing were grown in Brain Heart Infusion (BHI) media or on BHI supplemented with 1.50% bacteriological agar (Oxoid, UK). All *S. aureus* and *E. coli* strains for mutagenesis studies were grown in Tryptic Soy Broth (TSB) or on Luria-Bertani (LB) supplemented with 1.50% bacteriological agar. For confirmation of haemolytic activity colonies were grown on blood agar, TSA with 5% sheep blood. Strain stocks were taken from overnight cultures (15h) inoculated into 25 ml BHI at 37ºC agitated at 160 RPM. 750 µl of these overnight cultures were added to equal volume 80% heat sterilised glycerol solution, incubated at room temperature for 30 minutes and stored at -80ºC.

Table 2.3 | Growth Media and Composition

Media	Supplier	Composition
Tryptic Soy Broth (TSB)	Oxoid, UK	17 g/L pancreatic digest of casein 3 g/L enzymatic digest of soya bean 5 g/L sodium chloride 2.50 g/L dipotassium hydrogen phosphate 2.50 g/L glucose
Blood Agar – 5% Sheep blood (TSA)	Thermo Fisher, UK	5% sheep blood 17 g/L pancreatic digest of casein 3 g/L enzymatic digest of soya bean 5 g/L sodium chloride 2.5 g/L dipotassium hydrogen phosphate 2.5 g/L glucose
Brain Heart Infusion (BHI)	Oxoid, UK	5 g/L beef heart infusion solids 12.5 g/L brain infusion solids 2.5 g/L disodium phosphate 2 g/L glucose 10 g/L proteose peptone 2.5 g/L disodium phosphate
Luria-Bertani (LB)	Roth, UK	10 g/L peptone 5 g/L yeast extract 5 g/L NaCl
Bacteriological Agar European-Agar	Oxoid, UK Bectron-Dickinson, UK	15 g/L 15 g/L

2.4 DNA EXTRACTION (FOR SEQUENCING)

The **MasterPure ™ Gram Positive DNA Purification Kit** (Epicentre, Lucigen, UK) was used to extract genomic DNA. A single bacterial colony was grown overnight in 25 ml of BHI in a 250 ml conical flask with a filter cap overnight at 37ºC, 160 RPM. A 750 µl aliquot of overnight culture (late log phase growth) was pelleted by centrifugation at 13,000 x G for 3 minutes. The supernatant was discarded, and the pellet was resuspended in 150 µl of cell wall lysis buffer (147.50 µl Tris-EDTA buffer (TE) (Epicentre, Lucigen, UK), 1 µl Ready-Lyse™ Lysozyme Solution (~ 30,000 U/µl) (Epicentre, Lucigen, UK), and 1.50 µl lysostaphin (50 µl/ml) (Sigma Aldrich, UK)) and incubated for 30 minutes at 37ºC. After the initial cell wall lysis step, 149 µl of Gram-Positive-Lysis solution and 1 µl Protinase K (Epicentre, Lucigen, UK) was added to each sample and vigorously vortexed for 10 seconds. Subsequently the samples were incubated at 65ºC for 15 minutes, vortexing each sample for 5 seconds in 5-minute intervals. The samples were then incubated at room temperature for 5 minutes, after which they were placed on ice for 5 minutes. To precipitate the DNA, 175 µl MPC Protein Precipitation reagent (Epicentre, Lucigen, UK) was added to each lysed sample and vortexed for 10 seconds. To remove cell debris, samples were centrifuged at 4°C for 10 minutes at 13,000 x G. The supernatant was transferred into a clean Eppendorf tube and 1 µl of RNase A (5 µg/µl) (Epicentre, Lucigen, UK) was added, mixed thoroughly and incubated at 37ºC for 30 minutes. Next, 500 µl of isopropanol was added to the supernatant and well mixed by inverting 40 times. The DNA was then pelleted by centrifugation at 4°C for 10 minutes at 13,000 x G. The isopropanol was discarded by pipetting, and the DNA pellets were subsequently washed with 500 µl 70% Ethanol. The DNA was pelleted again at 4°C for 3 minutes at 13,000 x G, and the ethanol was carefully removed. The DNA pellets were left to air dry for 20 minutes at room temperature and subsequently resuspended in 45 µl of TE buffer. The DNA was eluted overnight at 4ºC.

2.4.1 Quantification Of DNA

The concentration and quality of the extracted genomic DNA was quantified using the Qubit dsDNA broad range assay kit (Life Technologies) according to the manufacturers protocol. The purity of the DNA extracted was estimated by measuring the 260/280 nm ABS ratio using the Nanodrop. The DNA was then subjected to Quality Control at the Wellcome Sanger Institute prior to PacBio SMRT Sequencing library preparation.

2.5 RNA EXTRACTION (FOR SEQUENCING)

The RiboPure RNA Purification Kit (Ambion, Invitrogen, Thermofisher, UK) was used for RNA extraction. Prior to the extraction, 1.50 ml of fresh bacterial overnight culture was pelleted (cultured in 25 mL BHI in a 250mL baffled flask 37ºC, 160 RPM in shaking incubator sampled at late log phase growth – preliminary growth curves determined), the supernatant removed, and the cells were resuspended in 750 µl RNA*later* Stabilization Solution (Invitrogen, Thermofisher, UK) and stored at 4ºC for up to a week. For the extraction, the cells form the RNAlater suspension were collected by centrifugation for 1 minute at 4°C and 13,000 x G. The supernatant was then removed and 350 µl of RNAWiz was added to resuspend the cell pellets, vortexing the mixture vigorously for 10 seconds. The suspension was then transferred into 0.50 ml screw cap tubes containing 250 µl ice cold Zirconia Beads. To lyse the cells, the samples were beat 4 times with a benchtop tissue homogenizer (Precellys 24 Tissue Homogeniser, Bertin Instruments, FR) at 6000RAM x 45 seconds placing the samples on ice between each run. The cell debris was removed by centrifugation at 4°C for 3 minutes at 13,000 x G. The bacterial lysate was carefully removed and transferred into a clean Eppendorf tube, and 0.2 volumes of chloroform was added. The samples were mixed well by shaking for 30 seconds and were left to incubate at room temperature for 10 minutes. The lysates were then spun down again at 4°C for 3 minutes at 13,000 x G, which after the aqueous phase containing partially purified RNA was recovered. Next, 0.50 volumes of 100% Ethanol were added to the aqueous phase, mixed thoroughly, transferred onto a Filter Cartridge within a collection tube, and spun down for 1 minute at 4°C and13,000 x G until all the ethanol was through the filter. The filters were then washed with 700 µl of Wash Solution 1, centrifuged for 1 minute at 4°C and 13,000 x G, then washed with 500 µl Wash Solution 2/3, and subsequently spun down for 1 minute at 4°C and 13,000 x G. The wash step with 500 µl Wash Solution 2/3 was repeated and the filters were centrifuged for 3 minutes at 4°C and 13,000 x G to get rid of an excess liquid. The RNA was then eluted by applying 30 µl of Elution Solution preheated to 95ºC to the center of the filter. After a 2-minute incubation time at room temperature, the filters were centrifuged for 1 minute to elute the RNA. The elution step was repeated with another 30 µl, incubated for 2 minutes at RT, and spun down to give better RNA yield in 60 µl total volume. The RNA was then treated with 4 µl DNAse I (with 1/9 volume 10X DNase Buffer), incubated for 30 minutes at 37ºC to digest any residual genomic DNA present. 0.20 volume of DNase Deactivation Reagent was subsequently added and incubated at room temperature for 2 minutes, after which the samples were quickly spun down for 1 minute at 4°C and 13,000 x G to pellet the inactivation reagent. Purified RNA solution was recovered into a new RNase-free tube.

2.5.1 Quantification of RNA

The concentration and purity of the extracted RNA was estimated by measuring the 260/280 nm ABS ratio and 230 nm reading using the Nanodrop. The RNA was then subjected to Quality Control at the Wellcome Sanger Institute prior to RNA-Seq library preparation.

2.6 BACTERIAL ISOLATES

2.6.1 *S. aureus* – Historic NCTC Collection (Chapter 3)

A collection of 108 historic *S. aureus* isolates from the National Collection of Type Cultures (NCTC) of Public Health England was sequenced in collaboration with The Wellcome Sanger Institute and Pacific Biosciences. Data for this project was kindly supplied by Dr. Simon Harris (Pathogen Variation Group, The Wellcome Trust Sanger Institute), who is a co-supervisor on this project. All sequenced isolates used in this study are detailed in Table 2.4. and the annotated genomes can be accessed through the Wellcome Sanger Institute (www.sanger.ac.uk/resources/downloads/bacteria/nctc/). Each isolate was sequenced using Pacific Bioscience's Single Molecule Real Time (SMRT) sequencing technology. 108 isolates were used in this study which were curated according to the quality of sequencing. An additional known 12 reference strains from varying ST types were added to the study. The assemblies of each reference strain and their plasmids can be accessed from National Centre for Biotechnology Information (NCBI) Genome Database detailed in Table 2.5.

Table 2.4 | NCTC *Staphylococcus aureus* strains in Historic *S. aureus* Study

Strains	Accession	Runs	Sequence Type (ST)	Clonal Complex (CC)
NCTC13297	ERS798856	40677_B01	1	1
NCTC7415	ERS811733	41004_F01	5	5
NCTC10656	ERS825168	41315_B02	5	5
NCTC4136	ERS825154	41255_B01	8	8
NCTC10652	ERS825165	41315_G01	8	8
NCTC8325	ERS980038	43024_E01	8	8
NCTC9369	ERS806206	40740_G01	8	8
NCTC13758	ERS1066616	44738_C01	8	8
NCTC13136	ERS798851	40657_D01	8	8
NCTC13139	ERS798854	40657_G01	8	8
NCTC13140	ERS798855	40677_A01	8	8
NCTC12232	ERS846858	41665_H01	8	8
NCTC12233	ERS798847	40574_H02	8	8
NCTC13394	ERS798860	40677_F01	8	8
NCTC13395	ERS811722	40871_F01	8	8
NCTC13812	ERS1247821	49508_D02	8	8
NCTC13141	ERS846859	41665_A02	8	8
NCTC13133	ERS654930	35910_B02	8	8
NCTC11939	ERS798843	40415_E02	239	8
NCTC11940	ERS921426	42197_E02	239	8
NCTC13135	ERS798850	40657_C01	239	8
NCTC13134	ERS1178934	46837_D02	239	8
NCTC13626	ERS659566	35910_E01	239	8
NCTC12981	ERS646605	34347_A01	243	8
NCTC13132	ERS836413	41559_A01	247	8
NCTC10654	ERS825166	41315_H01	250	8
NCTC10442	ERS836417	41559_E01	250	8
NCTC10443	ERS846849	41594_D01	250	8
NCTC10657	ERS846851	41594_F01	250	8
NCTC13138	ERS798853	40657_F01	250	8
NCTC8004	ERS806200	40740_C01	254	8
NCTC8178	ERS807423	40871_B01	254	8
NCTC10804	ERS846853	41767_G01	254	8
NCTC10833	ERS654923	35891_F01	254	8
NCTC10988	ERS798845	40415_G02	254	8
NCTC10703	ERS846852	41594_G01	3526	8
NCTC6136	ERS811731	41004_D01	9	9
NCTC8723	ERS904739	42042_D02	9	9
NCTC8725	ERS811742	41004_F02	9	9
NCTC8726	ERS825157	41315_B01	9	9
NCTC8765	ERS807420	40853_G01	9	9
NCTC5657	ERS806209	40757_C01	10	10
NCTC6137	ERS806225	40853_A01	10	10
NCTC6507	ERS807415	40853_B01	10	10
NCTC7972	ERS806188	40574_G02	10	10
NCTC10655	ERS825167	41315_A02	10	10
NCTC13616	ERS659565	35910_F01	22	22
NCTC13142	ERS798841	40415_C02	22	22
NCTC6134	ERS806215	40798_D01	25	25
NCTC8317	ERS544010	33763_A01	25	25
NCTC2669	ERS812507	40105_H02	30	30
NCTC8530	ERS806204	40740_E01	30	30
NCTC5655	ERS806208	40757_B01	30	30
NCTC5656	ERS825156	41255_D01	30	30
NCTC6571	ERS798862	40677_H01	30	30
NCTC7361	ERS807421	40853_H01	30	30
NCTC7445	ERS811737	41004_A02	30	30
NCTC7446	ERS811738	41004_B02	30	30

NCTC8507	ERS807424	40871_C01	30	30
NCTC11561	ERS825171	41315_E02	30	30
NCTC11962	ERS846861	41665_C02	30	30
NCTC11965	ERS836410	41556_G01	30	30
NCTC13299	ERS798858	40677_D01	30	30
NCTC13811	ERS1247820	49386_H02	30	30
NCTC13143	ERS846860	41665_B02	30	30
NCTC13277	ERS654931	35910_A02	30	30
NCTC11963	ERS836409	41556_F01	36	30
NCTC13373	ERS654933	35910_G01	36	30
NCTC3750	ERS811726	40961_E01	121	51
NCTC8531	ERS1018578	43295_D02	121	51
NCTC7414	ERS811732	41004_E01	121	51
NCTC7791	ERS807422	40871_A01	121	51
NCTC13298	ERS798857	40677_C01	121	51
NCTC13434	ERS798861	40677_G01	121	51
NCTC13435	ERS445051	27294_E01	80	80
NCTC7121	ERS798864	40677_B02	97	97
NCTC8399	ERS811740	41004_D02	97	97
NCTC9547	ERS811724	40961_C01	97	97
NCTC9551	ERS825159	41255_G01	97	97
NCTC9552	ERS950459	42545_H02	97	97
NCTC9752	ERS825152	41236_D02	97	97
NCTC10344	ERS825153	41236_E02	97	97
NCTC10345	ERS819824	41236_G01	97	97
NCTC3761	ERS811727	40961_F01	464	97
NCTC4163	ERS825155	41255_C01	464	97
NCTC4137	ERS811728	40961_G01	464	97
NCTC5658	ERS836414	41559_B01	464	97
NCTC10788	ERS636089	35473_F01	464	97
NCTC13841	ERS1295514	50450_A01	464	97
NCTC1803	ERS1031245	43874_B01	133	133
NCTC7988	ERS806190	40798_F01	133	133
NCTC9555	ERS825160	41255_H01	133	133
NCTC7712	ERS811739	41004_C02	136	133
NCTC12880	ERS798849	40657_B01	151	151
NCTC5663	ERS811729	41004_B01	350	350
NCTC7485	ERS807416	40853_C01	351	350
NCTC9546	ERS806207	40757_A01	692	385
NCTC9556	ERS950460	42552_A01	692	385
NCTC9611	ERS950452	42545_A02	692	385
NCTC9612	ERS825161	41315_C01	692	385
NCTC9613	ERS825162	41315_D01	692	385
NCTC9614	ERS825151	41236_C02	692	385
NCTC10399	ERS950461	42552_B01	707	707
NCTC6966	ERS798863	40677_A02	890	890
NCTC7856	ERS798867	40677_E02	890	890
NCTC6135	ERS811730	41004_C01	1021	1021
NCTC13137	ERS798852	40657_E01	1148	1148
NCTC10649	ERS1043807	43941_A01	1254	1254

Table 2.5 | Reference Genomes in Historic *S. aureus* Study

Genome	Accession	ST	CC	Plasmid	Plasmid Accession	Reference
MW2	BA000033	1	1	–	–	Baba et al., 2002
MSSA476	BX571857	1	1	pSAS	BX571858	Holden et al., 2004
Mu50	BA000017	5	5	VRSAp	AP003367	Kuroda et al., 2001
N315	BA000018	5	5	pN315	AP003139	Kuroda et al., 2001
NCTC8325	CP000253	8	8	–	–	Gillaspy et al., 2006
JE2 (JH9)	CP000703	105	5	pSJH901	CP000704	Mwangi et al., 2007
EMRSA-15	HE681097	22	22	–	–	Holden et al., 2013
MRSA252	BX571856	36	30	–	–	Holden et al., 2004
RF122	AJ938182	151	151	–	–	Herron-Olson et al., 2007
COL	CP000046	250	8	pT181	CP000045	Gill et al., 2005
ST398	AM990992	398	398	pS0385-1	AM990993	Schijffelen et al., 2010
				pS0385-2	AM990994	
				pS0385-3	AM990995	
TW20	FN433596	239	8	pTW20_1	FN433597	Holden et al., 2010
				pTW20_2	FN433598	

*ST = sequence type, CC = clonal complex

2.6.2 *S. aureus* – Singapore Collection (Chapter 4)

All bacterial strains in the Singapore Collection are methicillin-resistant *S. aureus* (MRSA) and were provided by Prof Li-Yang Hsu (National University of Singapore). They are part of a cross-sectional study (Chow et al., 2017) from a network of an acute hospital and five closely affiliated intermediate (n=2) and long-term care (n=3) facilities in Singapore. The study took place over 3 years, collecting nasal, axillary and groin swabs from all study subjects over a six-week period (June-July) each year. The sampling was random, swabbing 999 patients who stayed >48h in the acute hospital and all residence of the intermediate and long-term care home were included. In total 1552 isolates were sequenced using Illumina HiSeq technology at the Wellcome Sanger Institute (2014, n=385; 2015, n=597; 2016, n=570). Table 2.6 details isolates included in this study.

Table 2.6 | Singapore Collection *S. aureus* strains in Singapore Study

Strains	Accession	Runs	ST	CC
CD141496	ERS737577	58366_B01	622	22
CD150713	ERS1077854	58366_D01	622	22
CD150916	ERS1077679	58366_C01	622	22
CD140866	ERS737492	58366_A01	22	22
CD140400	ERS737319	58275_D01	22	22
CD140638	ERS737395	58275_E01	22	22
CD140657	ERS737400	58275_B01	45	45
CD140901	ERS737478	58275_A01	45	45
CD140392	ERS737297	58275_C01	45	45

*ST = sequence type, CC = clonal complex

Reference strains used for comparative genomic analysis detailed in Table 2.7.

Table 2.7 | Reference Genomes for Singapore Study

Genome	Accession	ST	CC	Plasmid	Plasmid Accession	Reference
EMRSA-15	HE681097	22	22	–	–	Holden et al., 2013
CA-347	CP006044	45	45	unnamed	CP006045	Stegger et al., 2013

*ST = sequence type, CC = clonal complex

2.6.3 *S. aureus* and *E. coli* – Mutagenesis (Chapter 5)

E. coli strains were used for propagation of deletion vector pIMAY before transformation into *S. aureus* as detailed in Table 2.8. The strains and vectors were constructed as described in Monk et al., 2012, detailed in section 2.9 Molecular Mutagenesis, and were kindly supplied by the Heilbronner group. *S. aureus* isolates from the Singapore collection were used to create knockout mutants for different *sau1hsdS* genes within the ST45 and ST22/622 background detailed in Table 2.9

Table 2.8 | List of Isolates used in Singapore Mutagenesis Study

E. coli Isolate	*S. aureus* Isolate	ST	Construct	KO *hsdS*
IM93	CD141496	622/22	RM1	ΔhsdS_α
IM93	CD140293	45	RM2	ΔhsdS_α
IM93	CD140293	45	RM3	ΔhsdS_β
IM01B	CD150713	622/22	RM4	ΔhsdS_X
IM01B	CD150713	622/22	RM5	ΔhsdS_S
IM01B	CD150713	622/22	RM6	ΔhsdS_α

Table 2.9 | Mutant Collection *S. aureus* strains Singapore Mutagenesis Study

Strains	Accession	Runs	ST	CC	Parent Strain
RM2_C1A2			45	45	CD140293
RM3_A4			45	45	CD140293
RM2+RM3_C1B5			45	45	CD140293
RM1_C1B6			622	22/45	CD141496
RM4+RM5_C1C3			622	22/45	CD150713
RM4+RM6_C3CD8			622	22/45	CD150713
RM5+6_K45			622	22/45	CD150713
RM5_C1C7			622	22/45	CD150713
RM6_C2B2			622	22/45	CD150713

*ST = sequence type, CC = clonal complex

2.7 SEQUENCING

2.7.1 DNA Sequencing

The samples NCTC strains (Chapter 3), and Singapore strains (Chapter 4) were subjected for DNA sequencing with the third generation PacBio Sequel II SMRT sequence (Pacific Biosciences, Menlo Park, CA) at the Wellcome Sanger Institute, Hinxton Cambridge. Four *ΔhsdS* strains in Chapter 5 were subjected to DNA sequencing with PacBio Sequel I SMRT sequencing through GeneWiz NGS, New Jersey.

2.7.2 RNA Sequencing

All RNA sequencing was carried out at the Wellcome Sanger Institute, Hinxton Cambridge. The experimental design was the same for the two sets of samples: for each isolate of interests, 3 biological replicates were taken to give a more accurate and reliable result and isolate any source of random biological variation within the experiment. Two technical replicates were also created to exclude any variability throughout the RNA sequencing.

Both RNA experiments were sequenced using Illumina HiSeq 4000 RNA-Seq in a strand specific manner on 2 lanes (1 pool of 14 samples, and another pool of 13 samples) with a read length of 75 bp. However, there were some changes in the preparation of sequencing libraries. The 2017 WT Singapore strains (Chapter 4) were prepared using Illumina TruSeq stranded RNA kit, the 2020 *ΔhsdS* strains (Chapter 5) with NEB Ultra II stranded kit. The libraries were Ribozero depleted, removing all ribosomal RNA (rRNA) prior to sequencing. The two sets were also sequenced using different libraries the 2017 WT isolate sequencing carried out with Illumina-C Library PCR whilst the 2020 mutant isolates were sequenced with a novel Liber PCR Bespoke approach and were multiplexed.

2.8 BIOINFORMATIC ANALYSIS

To conduct this work, most sequence analysis pipelines, and downstream analysis tools were accessed remotely by courtesy of the Pathogen Informatics Group at the Wellcome Sanger Institute (WSI), Hinxton Cambridge. All of the software developed by Pathogen Informatics at the Wellcome Sanger Institute is freely available for download from GitHub under an open-source license, GNU GPL 3.

2.8.1 PacBio Assembly and Annotation – WSI Automated Pipelines

PacBio reads generated with the RSII in h5 format are first converted to BAM format using the SMRTlink pipeline (version 5.0.1.9585). The subreads BAM file is then converted to FASTQ format using SAMtools (version 1.6 - https://github.com/samtools) (Li et al., 2009) containing uncorrected reads which did not pass the PacBio QC. The FASTQ file is then run through CANU (version 1.6 - https://github.com/marbl/canu - Koren et al., 2017) which creates a FASTQ file with corrected reads.

The BAM file of uncorrected reads is run through the PacBio SMRTLink *de novo* assembly pipeline, which uses HGAP (version 4.0 - Chin et al., 2013), to produce an assembled genome. The assembly is circularised using Circlator (version 1.5.3 - https://github.com/sanger-pathogens/circlator) (Hunt et al., 2015) with the FASTQ file of corrected reads from CANU. The circularised file is then run through Quiver (https://github.com/PacificBiosciences/GenomicConsensus) which removes any error which may have been introduced during the PacBio SMRTlink resequencing pipeline. The corrected reads are mapped back to the final assembly using minimap2 (version 2.6 - https://github.com/lh3/minimap2) (Li, 2017) and statistics are generated using SAMtools. Each assembly is automatically annotated using PROKKA (version 1.5 - https://github.com/tseemann/prokka) (Seeman, 2014) and a genus-specific database from RefSeq (Pruitt et al., 2012). The annotated and assembled genomes were visualised using Artemis (https://github.com/sanger-pathogens/Artemis) (Carver et al., 2012).

2.8.2 DNA Modification and Methylation Detection from PacBio Data

The PacBio SMRTlink Bio-PacbioMethylation pipeline (https://github.com/sanger-pathogens/Bio-PacbioMethylation/blob/master/README.md) was run at the Wellcome Sanger Institute. The pipeline, running the RS_Modification_and_Motif_Analysis protocol, was used for the the final assembly with the BAM file of uncorrected reads. The output files used for the analysis in this study is the motif_summary.xlm, motifs.gff, modifications.gff detailing the methylation motifs, positions and average matches within each genome. The methylation motifs for each reference strain were attained from REBASE genomes database (http://tools.neb.com/genomes/) (Roberts et al., 2015). The DNA modification motifs and their genomic locations were visualised with Artemis (https://github.com/sanger-pathogens/Artemis) (Carver et al., 2012).

2.8.3 Multi-Locus-Sequence-Typing (MLST)

Sequence types were determined using MLSTcheck (https://github.com/sanger-pathogens/mlst_check) used to compare the assembled genomes against the MLST database for *S. aureus* (https://pubmlst.org/saureus/) (Page, Taylor & Keane, 2016).

2.8.4 RNA-Seq Expression Pipeline – WSI Automated Pipelines

The RNA-Seq pipeline at the WSI was used to map and compute gene expression values for the isolates in this study. RNA sequence reads were mapped against reference genome CD140400 (ERS737319) using BWA (version 0.7.12 - https://github.com/lh3/bwa) (Li & Durbin, 2010) to produce a BAM file. BWA was used to index the reference and the reads were aligned using default parameters with the quality threshold for read trimming set at 15 (q=15) and maximum insert size 75 bp set as the maximum requested fragment size of the sequencing library.

Gene expression values were computed from the read alignments to the coding sequencing to generate the number of reads mapping and RPKM (reads per kilobase per million) through the Bio-RNASeq pipeline (https://github.com/sanger-pathogens/Bio-RNASeq). Only reads with a mapping quality score of 10 were included in the count. The output BAM files, BAM coverage plots of the reads mapped for each isolate were visualised in Artemis (https://github.com/sanger-pathogens/Artemis) (Carver et al., 2012), and the expression.csv containing the count data used for DE experiments.

2.8.5   Phylogenetic Reconstructions

Phylogenetic reconstructions were created using maximum likelihood (ML) analysis on the core genome for each library carried out with FastTree (http://www.microbesonline.org/fasttree/#Install) (Price, Degal, & Arkin, 2009). The WGS phylogeny for the NCTC collection was rooted by reference strain MSSA476 (BX571857). The resulting trees were visualised in FigTree (version 1.4.4 - https://github.com/rambaut/figtree/releases) and iTOL (https://itol.embl.de) (Letunic and Bork, 2019) for the addition of metadata and annotations.

2.8.6   Pangenome Analysis

Pangenome analysis was conducted to analyse the differences between the composition of core and accessory genes within each lineage. The pan genome was determined using Roary (Page et al., 2015) (https://github.com/sanger-pathogens/roary), using a BLASTp percentage identity of 95% and a core definition of 98% (Page et al., 2015). The output files were visualised with phandango (https://jameshadfield.github.io/phandango/#/) (Hadfield et al., 2017).

2.8.7   Comparative Genomic Analysis

2.8.7.1   Artemis Comparison Tool (ACT)

Artemis Comparison Tool (ACT) (https://github.com/sanger-pathogens/Artemis) (Carver et al., 2008) was used to visualise common features between the staphylococcal strains within each collection. This tool was also used to allow visualisation of mobile genetic elements (MGE) within each isolate.

2.8.7.2   Blast Ring Image Generator (BRIG)

Blast Ring Image Generator (BRIG) (http://brig.sourceforge.net/) (Alikhan et al., 2011) was used to visualise whole genome sequence homology between staphylococcal strains with the Singapore collection.

2.8.8   Identification of Mobile Genetic Element (MGE)

MGEs including, staphylococcal cassette chromosome (SCC) elements, plasmids, genomic island (*v*Sa), *S. aureus* pathogenicity islands (SaPI), transposons, and phage, were identified through visualisation in Artemis (https://github.com/sanger-pathogens/Artemis) (Carver et al., 2012) and Artemis Comparison Tool (ACT) (https://github.com/sanger-pathogens/Artemis) (Carver et al., 2008), and each accessory region manually annotated in comparison to the reference genomes. Nucleotide BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to search unknown plasmid types.

Prophage Hunter (http://pro-hunter.bgi.com) (Song et al., 2019) was used to identify unknown prophage elements.

SCC*mec*Finder (https://cge.cbs.dtu.dk/services/SCCmecFinder) (Kaya et al., 2018) was used to identify SCC*mec* elements which did not match the reference ST isolates and could not be determined from literature.

2.8.9   Identification and Characterization of RM Elements

Previously characterised RM elements of *S. aureus* reference strains were found on REBASE, from which FASTA protein sequences were extracted using Artemis, to create a protein sequence library for BLASTp. The library was queried using BLASTp against each isolate in the collection. The distribution and location of each element was then found within each sequence and extracted as a FASTA file. The presence or absence of each type of RM system and the combination of elements were visually represented with iTOL v4 (https://itol.embl.de) (Letunic & Bork, 2019).

2.8.10 Characterization Sau1 HsdS and RM TI Methylation Motifs

Protein sequences were extracted for each HsdS specificity element as FASTA files, to investigate their sequence similarity using SeaView (version 4.7 - http://doua.prabi.fr/software/seaview) (Gouy, Guindon & Gascuel, 2009). Previously identified Target recognition domain (TRD) protein sequences by Cooper et al., 2017 were added to the clustered protein sequences to categorise TRD and the corresponding bipartite target recognition sequence (TRS). These were compared to the RM Motif group_tags generated by the PacBio SMRT analysis, to be able to identify novel and

known TRD:TRS matches. Novel TRDs and TRSs were noted and matched with the process of elimination and analysis of the protein sequence similarities between already characterised *sau1hsdS* and motifs.

2.8.11 Structural Modelling and Characterization Sau1 HsdS

The protein sequences for each Sau1 HsdS specificity element were extracted in FASTA format. These were used to create three-dimensional protein models using SWISS-MODEL (https://swissmodel.expasy.org) (Waterhouse et al., 2018),

Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) (Kelley, et al., 2015) Both software's 'automated mode' was used as default, searching for the closest interactive template. The top-ranking template identified for target-template alignment for each *S. aureus* sau1 HsdS was PDB: 1YF2.1/1YF2.2 HsdS from *Methanocaldococcus jannshii*, (Kim et al., 2005). The models created were superimposed onto 1YF2.2 which was used to identify the DNA binding loops for each protein structure and binding predictions were also created for each model with I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Yang & Zhang, 2015).

2.8.12 TRS Frequency

The TRS frequencies were investigated with Artemis by searching for the known TRS for both forward and reverse DNA strands, to distinguish modification events in coding/intergenic regions and core genome/accessory genome. The number of occurrences for each genomic region were calculated on a per kb basis to normalise the sum of matches. To investigate hyper or hypomethylated areas within the chromosome, the location of each TRS was extracted from the modifications.gff and motifs.gff results from the RS_Modification_and_Motif_Analysis and binned into 10,000 bp sliding windows. The motifs.gff files were also put though PACMAN (https://bugfri.unibe.ch) (Falquet & Loetsher, 2015) for comparison and graphical view of each motif. The absolute number of TRS present in the genomes visualized with Artemis were cross referenced with the methylation events from PacBio restriction and Modification SRMT analysis.

2.8.13 TRS Frequency Statistical Analysis

The frequency of TRS were normalized by dividing the sum of TRS matches by the corresponding number of base pairs in the region of interest. When comparing the relative frequency of matches between two regions/systems, a 2-tailed paired T.TEST was conducted to validate that the data sets were significantly different to one another. The relative frequencies were then divided by each other to yield a ratio per isolate. The gross absolute average was taken for the ratio, giving a percentage difference between the two data sets. The standard deviation of the ratios were also calculated to show the distribution of the data with R (**R** Core Team (2013). **R**: A language and environment for statistical computing. **R** Foundation for Statistical Computing, Vienna, Austria. URL http://www.**R**-project.org/).

2.8.14 Differential Expression (DE) Analysis

Differential Expression Analysis was conducted using Trinity of the mapped RNA-Seq data was conducted with Trinity (Haas, et al., 2013) (https://github.com/trinityrnaseq/trinityrnaseq/wiki/Trinity-Differential-Expression) run within R. EdgeR (McCarthy, Chen, & Smyth, 2010) (https://bioconductor.org/packages/release/bioc/html/edgeR.html) was used as the method for identifying DE features as the data set comprised of 3 tiered multi factor design (technical + biological replicates, within CC, between CC) for differential expression. Alongside EdgeR, DESeq2 (Love, Huber and Anders, 2014) and limma-voom (Law et al., 2014) were also trialled for DE analysis. Along with the default results figures, the count data was also visualised in an interactive web-based differential expression platform WebMeV (http://mev.tm4.org/#/welcome), and IDEAMex (Jimenez-Jacinto et al., 2019) http://www.uusmb.unam.mx/ideamex/) to run transcript QC, differential gene expression and to visualise results In a detailed overview. The transcripts levels were normalised using the Trimmed Mean of M-values (TMM) method, to standardise the distribution of count values according to the sequencing yield (sequencing depth, gene lengths, RNA composition (number of genes expressed, highly expressed genes, contamination)) of each sample. Batch corrections were performed on count data for comparison in Chapter 5 with limma – and subsequent differential expression was run in EdgeR.

removeBatchEffect function (https://rdrr.io/bioc/limma/man/removeBatchEffect.html)

2.8.15 mRNA Structure, Promoter and Transcriptional Start/Stop Site Predictions

The structure of the single stranded mRNA was predicted using RNAfold Server (http://rna.tbi.univie.ac.at) to analyse methylation interference within the 5' (leader) and 3' (trailer) untranslated region (UTR) flanking the transcriptional start and stop sites respectively. The mapped RNA transcript level BAM files as well as the coverage plots from the RNA-Seq Expression were visualised in RNA to locate transcript start and end sites. These were compared with Bacterial RNA maps from *S. aureus* strains MW2 (Saenz-Lahoya et al., 2019) and NCTC8325 (Ruiz de los Mozos et al., 2013; Lasa, et al 2011) (http://rnamaps.unavarra.es) and *S. aureus* NCTC325 Expression Data Browser (http://genome.jouy.inra.fr/cgi-bin/aeb/index.py) (Mäder et al., 2016). The promoters, ribosomal binding sites and start/termination sites were predicted using BPROM (http://www.softberry.com/berry.phtml?topic=bpromandgroup=programsandsubgroup=gfindb) (Solovyev & Salamov, 2011), PRODORIC (http://www.prodoric.de/vfp) (Münch et al., 2005) DOOR (reference HO 5096 0412) (Mao et al., 2009) (http://161.117.81.224/DOOR2/index.php) and Pepper (http://genome2d.molgenrug.nl) (Jong et al., 2012). Promoters, UTRs, Shine-Dalgarno ribosomal binding sequences, transcriptional and translational start/stop sites were visualised and annotated in SnapGene 5.0 software (from GSK Biotech; www.snapgene.com).

2.9 MOLECULAR BIOLOGY – MUTAGENESIS

All experiments were carried out in the laboratory of Prof. Andreas Peschel, under Dr. Simon Heilbronner at his team at the University of Tübingen, Germany.

2.9.1 Cloning

2.9.1.1 Insert Preparation

The construction of the inserts and vector, were generously prepared by Darya Belikova with Overlap Extension PCR Method for which the primers were designed in St Andrews, detailed in Table 2.10. The cloning method and pIMAY plasmid selected are detailed in Monk et al., 2012.

Table 2.10 | Primer Sequences for Overlap Extension PCR for Deletion of *sau1hsdS*

Isolate	ST	Construct	*hsdS*		Primer Sequence *
CD141496	622/22	RM1	*ΔhsdS_α*	A	ttgtgcGGTACCtaccacatggcgtcttattc
				B	catctttcaacaccccaagttctttcag
				C	cttggggtgttgaaagatgtaagcatttgagcacatctatcaattaag
				D	cttgagGAGCTCtcgcttgtgatctaacctctaa
CD140293	45	RM2	*ΔhsdS_α*	A	cagtcgGGTACCtactcccacatggtgtattattc
				B	catctttcaacaccccaagttcttttagg
				C	gaacttggggtgttgaaagatgtaaatctgctgaagtttattttataggaaatg
				D	GagctaGAGCTCtccatactttctgcttttactggc
CD140293	45	RM3	*ΔhsdS_β*	A	AtggcaGGTACCgttgtgctacctcatggtgtc
				B	catctttcaacaccccaagttcttttagg
				C	cttggggtgttgaaagatgtaatagtattaaatatgatatttagttcagcatag
				D	gatgcaGAGCTCaagaaaagtcaagaagaagctag
CD150713	622/22	RM4	*ΔhsdS_X*	A	gatagaGGTACCcatcatagaagttccagacg
				B	catttttcaacactcctagttctttgag
				C	gaactaggagtgttgaaaaatgtaattcttataaagttctattatg
				D	cctgagttcagtggtaaagGAGCTCgttatgc
CD150713	622/22	RM5	*ΔhsdS_S*	A	cctgagttcagtggtaaagGAGCTCgttatgc
				B	cataacctaatccctccaatgacttacg
				C	cattggagggattaggttatgtagtgaagatgatttagaacaggttgcac
				D	ctaactGAGCTCgatgatacccgtcttcaatacc
CD150713	622/22	RM6	*ΔhsdS_α*	A	ttgtgcGGTACCtaccacatggcgtcttattc
				B	catctttcaacaccccaagttctttcag
				C	cttggggtgttgaaagatgtaagcatttgagcacatctatcaattaag
				D	cttgagGAGCTCtcgcttgtgatctaacctctaa

* Upper case letters for restriction site for KpnI (fragment A) of SacI (fragment D)

2.9.1.2 Overlap Extension PCR

In overlap extension PCR mediated deletion mutagenesis Figure 2.1, PCR products, flanking the gene of interest, are prepared using a nonchimeric and chimeric primer pairs, A and B or C and D respectively. The PCR products (AB and CD) from the primary PCR reaction are used as the template for ligation PCR containing the outermost A-D primer pair detailed in Lee et al., 2010.
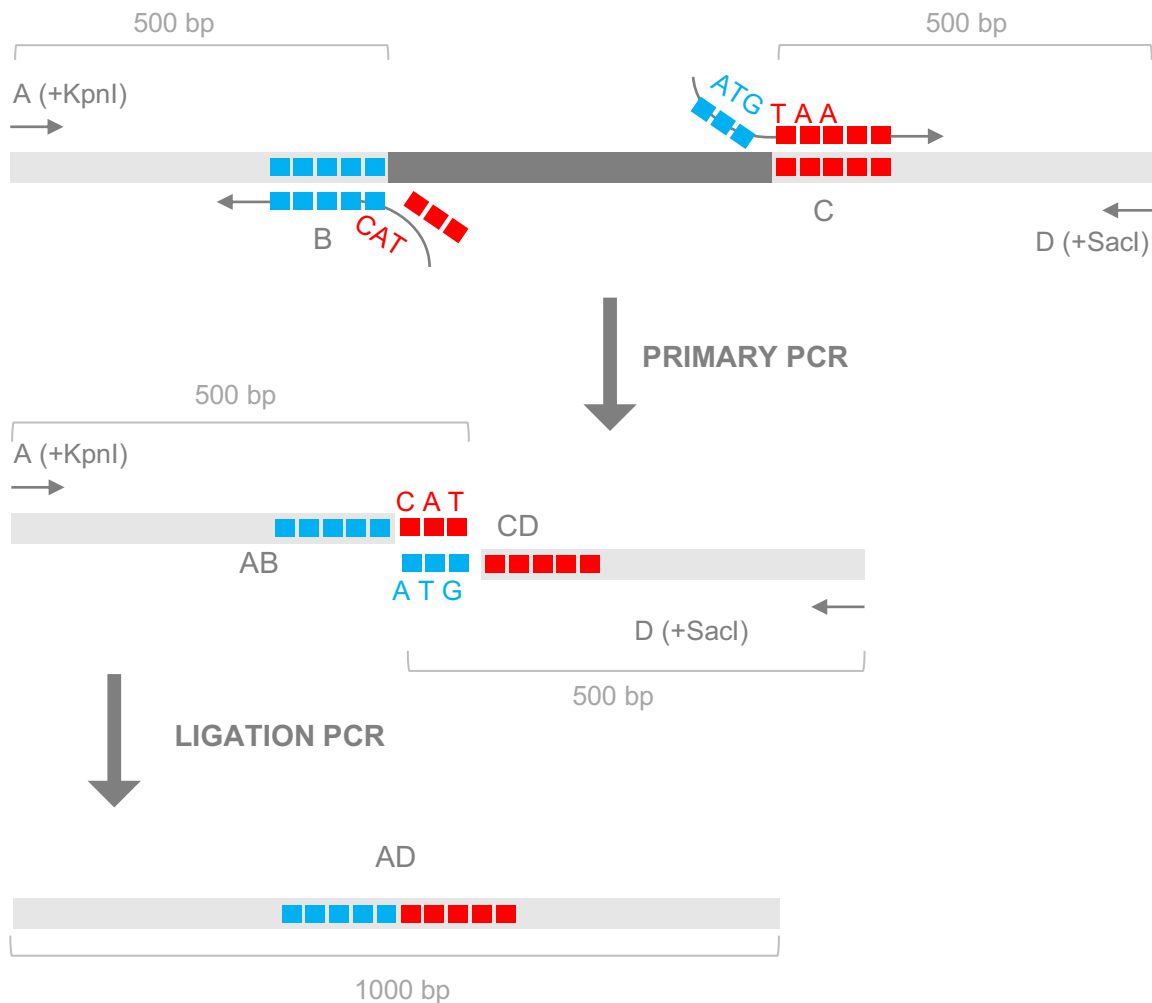
Figure 2.1 | Overlap Extension PCR – Deletion Mutagenesis
For deletion mutagenesis, two pairs of chimeric and non-chimeric primers, A - B and C - D are designed to flank the gene to be deleted. Both primers A and D contain a restriction site for a distinct restriction enzyme – A: KpnI and D: SacI – for subsequent restriction digestion prior to ligation with vector plasmid. Primer B contains the first 3 start codons (ATG) of the gene in reverse compliment orientation (CAT), whilst primer C contains around 15-20 nucleotides of primer B encompassing the start site (ATG), attached to the stop codon (TAA) and a further 25 nucleotides downstream of the gene of interest. This creates sticky ends for the primary PCR products AB and CD which overlap at the overhanging chimeric sequences during the ligation PCR using the purified AB CD fragments as the template DNA and primer A and D. As a result of the second PCR, the 2 x 500 bp long AB and CD fragments are ligated to produce a product 1000 bp long. Adapted from Lee et al., 2010.

Genomic DNA was extracted from the isolates of interest using the NucleoSpin Microbial DNA Kit (Macherey-Nagel, Germany) as per the user manual. The concentration of the genomic DNA was measured with Nanodrop 2000 (Thermofisher, UK). For insert preparation, the template DNA from the isolates were diluted so that the final concentration of DNA in a 20 µl PCR reaction was 100 ng/µl. The PCR reaction mixture and volumes of reagents are detailed in Table 2.11.

Table 2.11 | Overlap Extension PCR Reaction Setup

PRIMARY PCR	
20 µl PCR Reaction	Volumes x 1
dNTPs (10 mM)	0.40 µl
Forward Primer (1:10)	1 µl
Reverse Primer (1:10)	1 µl
Phusion Polymerase	0.40 µl
Loading Buffer x6	4 µl
ddH$_2$0	Depends on DNA [ ]
DNA template	Calc. V for 50ng
LIGATION PCR	
50 µl PCR Reaction	Volumes x 1
dNTPs (10 mM)	1 µl
Forward Primer (1:10)	1 µl
Reverse Primer (1:10)	1 µl
Phusion Polymerase	0.50 µl
Loading Buffer x6	10 µl
ddH$_2$0	31.5 µl
Template AB (PCR product)	1 µl (1:20 dilution)
Template CD (PCR product)	1 µl (1:20 dilution)

For the primary PCR reaction fragments A-B and C-D were amplified in pairs using PCR programme 1 detailed in Table 2.12 (adjusted to the temperatures for the primer pairs). The PCR products were subsequently mixed with 4 µl Tri-Track Loading dye and loaded onto a gel with 5 µl of 1 kb Ladder and separated (small gel protocol – 120V, 400 milli amp, 50 Watt for 35 minutes) and visualized under UV light. To make the 1% agarose gel, 0.7 g of agarose powder was suspended in 70 ml 1 x TAE buffer in a 250 ml Erlenmeyer flask and heated in microwave until fully dissolved (2-5 minutes), with additional 7 µl of SYBR Safe dye for visualization in UV transilluminator. The expected length of the amplified fragments AB and CD were 500 bp, which was validated through Sanger sequencing.

Next the AB and CD fragments were used as the DNA template for the second round of PCR for ligating together the two sequence fragments. Primers for A and D were added to the PCR reaction detailed in Table 2.10 using the same PCR protocol (Table 2.12) as previous but increasing the extension time to fit 1 kb PCR product (30 seconds). The PCR

product was purified using the Zymo Research DNA Clean and Concentrator-5 Kit as per the manufacturer's manual.

Table 2.12 | PCR Programme Setup - Phusion Polymerase

Cycle Steps	2-step protocol		3-step protocol		
	Temp.	Time	Temp.	Time	Cycles
Initial Denaturation	98°C	30 s	98°C	30 s	1
Denaturation Annealing	98°C	5-10 s	98°C X °C	5-10s 10-30s	25-35
Extension	72°C	30 s/kb	72°C	30 s/kb	
Final Extension	72°C / 4°C	5-10 min hold	72°C / 4°C	5-10 min hold	1

2.9.1.3 Vector Preparation

The vector used for the mutagenesis studies was pIMAY (Figure 2.2 – Addgene plasmid #68939; http://n2t.net/addgene:68939; PRID: Addgene_68939) detailed in Monk et al., 2012. This plasmid is a low-copy and uses temperature sensitive (*repBCAD*) replicon allowing effective integration at 37ºC. The plasmid also carries a tetracycline-inducible antisense *secY* region (anti-*secY*) which prevents the growth of bacterial cells which have retained the integrated plasmid, thus selecting for cells that have lost the plasmid. The plasmid also carries the *cat* gene coding for chloramphenicol (Cm) resistance which ensures selection throughout the transformation and chromosomal integration of the plasmid.
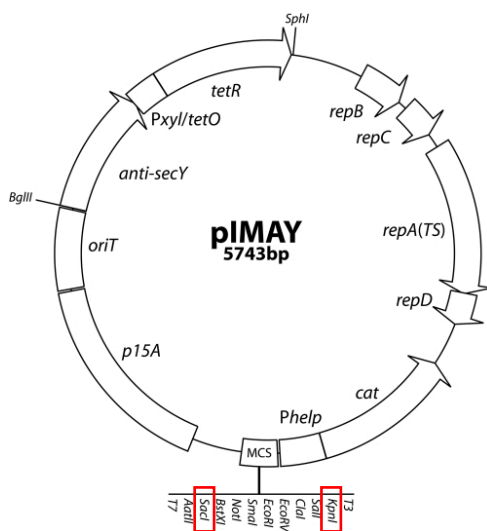
Figure 2.2 | Genetic Map of pIMAY
Backbone of *E. coli* / staphylococcal temperature sensitive plasmid for allelic exchange. This plasmid is low copy, with an *E. coli* origin of replication (p15A), a pBluescript multiple cloning site (MCS) with restriction enzyme sites (SacI and KpnI – used in this mutagenesis study). The plasmid also contained an *E. coli* origin of transfer for conjugation (*oriT*) and highly expressed *Phelp-cat* gene for chloramphenicol resistance.The plasmid carries a temperature sensitive replicon for Gram positive bacteria (*repBCAD*) and a tetracycline-inducible antisense *secY* region (*anti-secY*) for counter selecting bacteria which have lost the plasmid after integration. Sourced from Addgene.

2.9.1.4 Isolation of Vector Plasmid

The pIMAY plasmid was isolated from the stock DH5alpha *E. coli* strain. 100 ml of TSB containing 10 µg/µl Cm was inoculated with a loopful of DH5alpha *E. coli* containing pIMAY from agar stock. The culture was grown overnight at 37ºC at 160 RPM and the full volume of the culture was pelleted in 2 x 50 ml falcon tubes for 10 minutes at 6000 xG at 4ºC. The pIMAY plasmid was extracted from the bacterial cells with Qaigen Midi Prep kit as per manufacturer's instructions. The plasmid DNA was eluted in 30 µl of ddH$_2$O and the concentration measured with Nanodrop.

2.9.1.5 Restriction Digestion of Plasmid and Constructs

In order to insert the deletion construct into the plasmid (within the MCS) both the AD PCR fragment and pIMAY were restriction digested with two non-compatible end enzymes (KpnI and SacI). In separate reactions (Table 2.13) 500 ng of the AD PCR product, and 1000 ng of pIMAY plasmid were incubated with KpnI and SacI for 1 hour at 37ºC. This ensured the linearization of the plasmid (linear 5743 bp) with cleavage points at KpnI / SacI which are now the ends of the sequence.

Table 2.13 | Restriction Digestion Reaction Set Up

AD PCR PRODUCT (Construct)	
Reaction V (20 µl)	Volumes x 1
FastDigest Green Buffer	10 µl
KpnI	1 µl
SacI	1 µl
AD PCR Product (500ng)	Calculate
ddH$_2$O	Calculate
piMAY (Vector)	
Reaction V (20.0 µl)	Volumes x 1
FastDigest Green Buffer	10 µl
KpnI	1 µl
SacI	1 µl
piMAY DNA (1000ng)	Calculate
ddH$_2$O	Calculate

The AD PCR fragment was also digested at the restriction enzyme sequences, KpnI at the A end, and SacI at the D end. Half of the digest reaction for pIMAY (10 µl) was mixed with 2 µl of Loading Buffer and run on a 1% agarose gel on a small gel protocol (120V, 400 milli amp, 50 Watt for 35 minutes) to confirm the linearization of the plasmid. The remaining 10 µl of digested pIMAY was purified using the Zymo Research DNA Clean and Concentrator kit, as well as the 20 µl AD fragment digestion. The concentrations of the purified digests were then measured by Nanodrop to calculate ligation volumes.

2.9.1.6 Plasmid and Construct Ligation

To anneal the digested insert fragment AD and the linearized pIMAY vector, a ligation reaction was set up as detailed in Table 2.14 and left to incubate at RT for 1 hour. Subsequently, 10 µl of the ligate was mixed with 4 µl TriTrack DNA Loading Dye and run on a 1% gel on a small gel protocol (120V, 400 milli amp, 50 Watt for 35 minutes) to confirm the ligation of the vector and the insert resulting in a product ~7500 bp in length.

Table 2.14 | Ligation Reaction Set Up

Reaction V (20.0 µl)	Volumes x 1
T4 DNA Ligase Buffer x10	10 µl
T4 DNA Ligase	0.5 µl
Insert (3:1 ratio to vector)	3X
pIMAY (100ng)	X
ddH$_2$O	Calculate

2.9.2 Transformation into *E. coli*

2.9.2.1 Transformation of Deletion Vector into *E. coli* (IM93B and IM01B)

Aliquots of 70 µl competent *E. coli* IM93B (RM1, RM2, RM3) and IM01B (RM4, RM5, RM6) which were previously prepared by Darya Belikova were used to transform the deletion vectors. The *E. coli* cells were thawed on ice for 10 minutes after which 10 µl of the ligation product vectors were added to the cells, and further incubated on ice for 20 minutes. The *E. coli*-vector mixture was then heat shocked at 42ºC for 90 seconds and immediately re-suspending with 500 µl of room temperature TSB, and placed into the shaking incubator horizontally for 1 hour at 37ºC. After incubation, 100 µl of the neat culture and the pellet with backflow were plated onto TSA containing 10 µg/µl chloramphenicol (TSA Cm10) and incubated overnight at 37ºC. The colonies that grew on the Cm10 infused media suggest successful transformation. Three colonies per construct (RM1, RM2, RM3, RM4, RM5, RM6) were picked with a loop, were subsequently dotted onto a TSA Cm10 plate and incubated overnight and save as a 'safety plate' for plasmid extraction. The cells which remained on the inoculation loop were re-suspended in 25 ml of TSB and incubated overnight. The plasmid DNA from each transformed vector was extracted from the overnight bacterial culture as per the manual of the Thermo Fisher GeneJet kit. 1000 ng worth of the plasmid DNA for each construct was re-digested with KpnI and SacI to validate that the deletion insert was integrated into the vector. 10 µl digested products were run on a 1% agarose gel as described under 'Restriction Digestion and Plasmid'. For each successful plasmid, a 1000 bp AD fragment and a 5.8 kb fragment for the vector plasmid

were resolved. The remaining plasmid DNA for successfully transformed clones were purified with the Zymo Research DNA Clean and Concentrator kit and sent for Sanger sequencing with the respective A and D primer pairs to validate the full knock out of *hsdS*.

2.9.2.2 Plasmid Midi-Prep and Extraction

Subsequent to the constructs having been confirmed within the vectors, a larger yield of plasmids is necessary to transform into *S. aureus*. Colonies from the 'safety plates' from the previous step were taken to inoculate 100 ml of TSB Cm10 and left to grow overnight at 37ºC at 160 RPM. The next day, glycerol stocks were created using 700 µl of the *E. coli* strains carrying the deletion vectors and the rest of the 100 µl culture was subjected to plasmid extraction with the Qaigen Midi Kit as per the manufacturer's manual. The concentration for the eluted pIMAY constructs was measured by Nanodrop, and 50 ng of each RM construct was sent for Sanger sequencing with vector specific primers to validate the inserts within each vector.

pIMAY-F TACATGTCAAGAATAAACTGCCAAAGC / pIMAY-R AATACCTGTGACGGAAGATCACTTCG)

2.9.3 Transformation into *S. aureus*

2.9.3.1 Competent Cells

Prior to transformation of the extracted deletion vectors into the *S. aureus* strains, competent cells were made. 20 ml of TSB was inoculated with one *S. aureus* colony (CD141496 for RM1, CD140293 for RM2 and RM3, and CD150713 for RM4, RM5, and RM6) and incubated overnight at 37ºC at 160 RPM. The next morning, the OD_{600} of each bacterial culture was measured and the volume of culture needed to inoculate 100 ml of fresh media to reach a 0.5 OD_{600} was calculated (~5 ml). 100 ml of pre-warmed TSB was inoculated for each strain to 0.5 OD_{600} and was incubated for 45-55 minutes at 37ºC at 160RPM until the culture reached 0.6 OD_{600}. The cultures were transferred into 2 x 50 ml falcon tubes and incubated on ice for 15 minutes. Subsequently the cells were pelleted at 5000 x G at 4ºC for 10 minutes in a fixed angle rotor centrifuge. The supernatants were discarded and the pellets were resuspended with 1 ml of ice cold ddH_2O. The two resuspended pellets per strain were combined and washed with 50 ml of ice cold ddH_2O. The cells were centrifuged for 10 minutes at 5000 x G at 4ºC. The supernatants were discarded, the pellets resuspended in 50 ml of ice cold ddH_2O and centrifuged for 10 minutes at 5000 x G at 4ºC. The supernatants were discarded, the pellets resuspended in

50 ml of ice cold 10% glycerol solution and centrifuged for 10 minutes at 5000xG at 4ºC. The pellets were finally re-suspended in 500 µl of ice cold 10% glycerol solution. 70 µl aliquots of competent cells were prepared and stored at -80ºC.

2.9.3.2   Transformation by Electroporation

A single 70 µl aliquot of competent *S. aureus* cells was transformed using electropermeabilization. The aliquot of cells was thawed on ice for 5 minutes and subsequently incubated at room temperature for additional 5 minutes. The cells were pelleted at room temperature at 5000 x G for 5 minutes. The supernatant was discarded, and the cells were re-suspended in 80 µl electroporation buffer (0.50M sucrose dissolved in 10% glycerol solution). Next, 5 ng of plasmid DNA (< 20 µl) was added to the cells and were transferred into a 1mm electroporation cuvette (Geneflow, UK). The plasmid and cell suspension was electroporated (2.1 kV, 100Ω, 25 uF) and immediately re-suspended in 950 µl pre-warmed recover medium (TSB with 0.50M sucrose at 37ºC). The 1 ml mixture and transferred into a 15 ml falcon tube and incubated with agitation at 30ºC at 160 RPM for 2 hours. Post-incubation, 100 of the neat culture and the pellet with the backwash, were plated onto TSA Cm10 plates and incubated at 30ºC for 2 days. This step ensures that the plasmid is transformed into the *S. aureus* strains for allelic exchange.

2.9.3.3   Integration of pIMAY into *S. aureus*

As pIMAY is a temperature sensitive plasmid, it must be forced to integrated into the staphylococcal chromosome at 37ºC. Three colonies (C1, C2, C3) from the trans-formation TSA plates grown at 30ºC were picked with a loop and resuspended in 150 µl phosphate-buffered saline (PBS). The cell suspension was vortexed for 5 seconds and a dilution series of x10^{-1}, x10^{-2}, x10^{-3} was made. 100 µl of the neat bacterial solution and for the 3 further dilutions were plated onto TSA Cm10 and incubated at 37ºC. In this step the temperature sensitive Gram-positive replicon (*repBCAD*) is induced, and the plasmid starts to replicate within the cell to subsequently integrate into the chromosome. After overnight incubation, 10 colonies for C1, C2, and C3 were dotted on Cm10 TSA plates to save and the rest were used for colony PCR to check if the plasmid was still replicating within the cell or if it was integrated successfully into the staphylococcal chromosome. The safety plates were incubated overnight at 37ºC.

For Colony PCR, a single colony of was re-suspended in 50 µl of lysis buffer in a 1.5 Eppendorf tube (20mM Tris, 2mn EDTA, 1% Triton X-100) and an additional 5 µl of 5 µg/µl lysostaphin and mixed well by vortexing for 10 seconds. The resuspended cells were incubated on a heat block for 30 minutes at 37ºC at 350 RPM after which the cells were pelleted at 10,000 x G for 5 minutes. The supernatant was used as the DNA template for the PCR reactions as detailed in Table 2.15.

Table 2.15 | Colony PCR Reaction Setup

PRIMARY PCR	
20 µl PCR Reaction	Volumes x 1
dNTPs (10 mM)	0.40 µl
Forward Primer (1:10)	1 µl
Reverse Primer (1:10)	1 µl
Phire II Polymerase	0.40 µl
Loading Buffer x6	4 µl
ddH$_2$0	12.20 µl
Colony supernatant	1 µl

Vector specific MCS primers (pIMAY-F: TACATGTCAAGAATAAACTGCCAAAGC / pIMAY-R: AATACCTGTGACGGAAGATCAC TTCG) were used as the forward and reverse primers to check for actively replicating plasmid. The PCR products were mixed with 4 µl TriTrack DNA Loading Dye alongside a 10 kb Ladder and run on a 1% gel on a small gel protocol (120V, 400 milli amp, 50 Watt for 35 minutes). If there were no bands visible on the gel, the plasmid was integrated into the host chromosome. Next, 4 colonies for each of the 3 colonies (C1, C2, C3) were patched onto a new Cm10 plate, and the same loop was used to inoculate 10 ml of TSB for 2 days at 30ºC at 160 RPM (C1 A-D, C2 A-D, C3 A-D). These cultures were passaged 3 (P3) times in 5 ml TSB in culture tubes for 2 days at a time at 30ºC at 160 RPM.

2.9.3.4 Excision of pIMAY from *S. aureus*

To recover colonies that   have   successfully mutagenized   and   lost   the   plasmid vector, tetracycline-inducible reporter   gene   expression   *anti-SecY* was stimulated. TetR represses the expression of antisense SecY protein, which is the main part of the transmembrane subunit of the TII secretory pathway. Expression of the SecY antisense RNA (*anti-SecY*) inhibits the growth of cells which are maintaining the plasmid. Hence, the cells which do not grow on tetracycline infused media have lost the plasmid and can be screened for insert orientation. Consequently, a dilution series of x 10^{-5} and x   10^{-6} overnight   cultures (post P3),   were   plated   onto TSA with 1 µg/ml anhydrotetracycline (ATC) and were incubated for 2 days at 30ºC.

2.9.3.5 Validation of Excised Plasmid

To validate the successful deletion of *sau1hsdS*, the colonies growing on TSA ATC were counter selected for on TSA Cm10 and normal TSA plates. The plasmid is chloramphenicol resistant, hence will grow on TSA Cm10 plates, whilst colonies which have lost the plasmid will be susceptible and will not grow. For each ATC TSA dilution plate, 4 large dominant colonies were patched onto normal TSA and subsequently TSA Cm10 plates and incubated at 37ºC overnight. The colonies which did not grow on TSA Cm10, were marked on the normal TSA plates and were screened for correct insert orientation through colony PCR detailed in Figure 2.3.
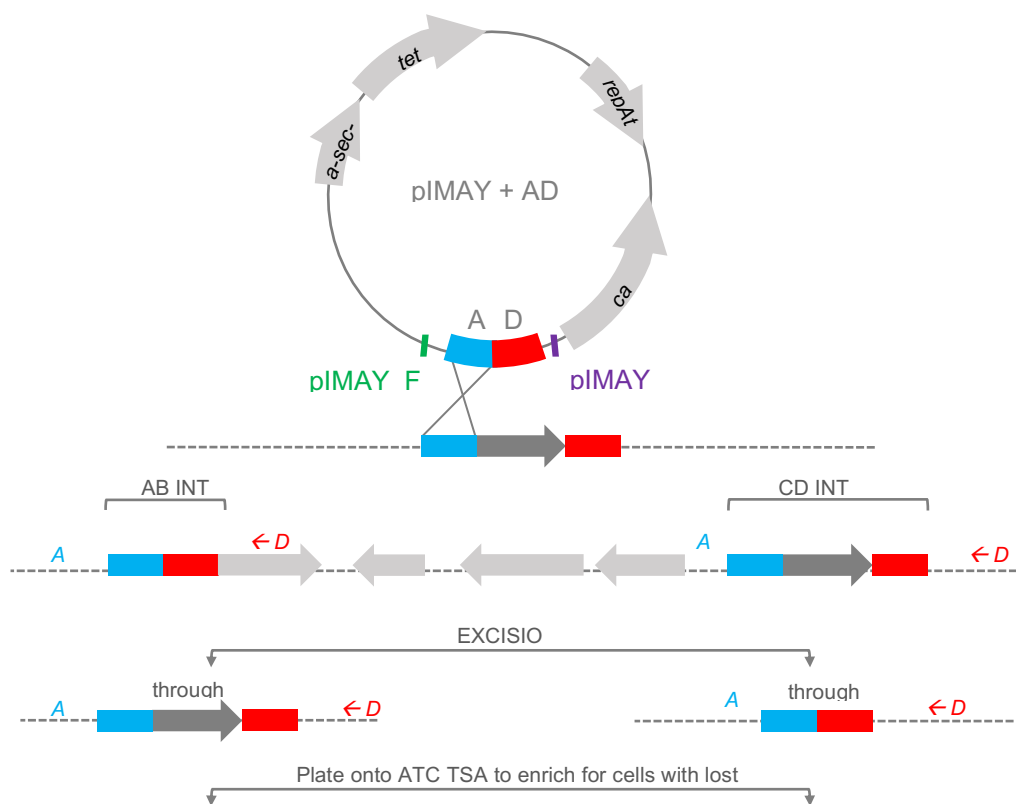
Figure 2.3 | Allelic Exchange with pIMAY
The deletion plasmids were isolated from *E. coli* IM93B or IM01B (depending on RM construct) and subsequently transformed into *S. aureus* at 30ºC for two days. Single-crossover (SCO) for integration was induced by growth in the presence of Cm10 at 37ºC. Colony PCR was conducted to validate the loss of replicating plasmid using MCS primers pIMAY_F and pIMAY_R – no product indicating no active replication. The negative clones were then screened for orientation of plasmid integration with chromosomal and up/downstream (AB INT: A UP/ D DOWN or CD INT: A UP/ D DOWN) via PCR cloning primers (Table 2.16). Clones from either integration orientation are grown at 30ºC in Cm10 TSB to stimulate growth in several passages. These cultures are then plated onto ATC TSA plates to excise the plasmid. Expression of the SecY antisense RNA (*anti-SecY*) inhibits the growth of cells which are maintaining the plasmid. Excision via the AB side will result in recreation of the WT locus, whilst excision through CD will yield the mutant gene. To validate successful mutants, the colonies are then plated on Cm10 and normal TSA plates. The colonies which only grow on the antibiotic free media, have definitely lost the plasmid and can be propagated for gDNA extraction and further PCR validation with the original AD primers for a 1000 bp PCR product. Altered from Monk et al., 2012.

2.9.3.6 Mutant Screening

To validate that the construct was integrated, and the vector was knocked out, a pair of chromosomal primers for up and downstream the to-be-deleted *hsdS* gene were designed detailed in Table 2.16. Colony PCR was performed and if the screened colony was successfully mutagenized the resulting band from the PCR product should have been ~1000 bp indicating amplification of only the AD fragment. If the band is larger ~2500 bp then the isolate still contained the gene of interest and the two flanking fragments (WT allele) as seen in Figure 2.3.

Table 2.16 | Primer Sequences for Up and Downstream s*au1hsdS* genes

Isolate	ST	Construct	hsdS	U/D	Primer Sequence *
CD141496	622/22	RM1	*ΔhsdS α*	U	TCAACACATGGTACATTACCTAG
				D	CCTCCTATATTTTTCAGATCAAAAC
CD140293	45	RM2	*ΔhsdS α*	U	AAGCAAGCCATAGCAGAATATG
				D	GTAGAAATTAATTACATCCATCGT
CD140293	45	RM3	*ΔhsdS β*	U	CAACTATGCATGGATTGAACATATG
				D	CTAATACTGCATTATGATTAAATAATTGTGG
CD150713	622/22	RM4	*ΔhsdS orfX*	U	CAACACATGGTACATTACCTAG
				D	CTTGTTCCTGTATAATAATCTTTC
CD150713	622/22	RM5	*ΔhsdS SCC*	U	CTTCGCCTTTATTCAACACATGGTAC
				D	GCGATGTTTTTGGACTATTTAGTTAAAGA
CD150713	622/22	RM6	*ΔhsdS α*	U	TCAACACATGGTACATTACCTAG
				D	CCTCCTATATTTTTCAGATCAAAAC

Successful mutant colonies were secured on TSA and TSA Blood Agar (incubated overnight at 37ºC) to check for haemolytic activity. A single colony from the TSA plate was inoculated into 20 µl TSB and cultured overnight at 37ºC at 160 RPM. The overnight culture was used to extract genomic DNA, cultures for double knockouts and glycerol stocks stored at -80ºC. Double knockouts were achieved by transforming deletion constructs into successful single knockout mutant competent cells detailed in Table 2.17.

Table 2.17 | Mutant Strains and *sau1hsdS* Knockout List

Strain	ST	Construct	Knockout (KO)	Functional *sau1hsdS*	Mutant Strain
CD141494	ST22/622	*RM1*	*ΔhsdS α*	*none*	**RM1_C1B6**
CD140293	ST45	*RM2*	*ΔhsdS α*	*sau1hsdS_β*	**RM2_C1A2**
CD140293	ST45	*RM3*	*ΔhsdS β*	*sau1hsdS_α*	**RM3_A4***
CD150713	ST22/622	*RM4*	*ΔhsdS orfX*	*sau1hsdS_S + sau1hsdS_α*	RM4_C2A2*
CD150713	ST22/622	*RM5*	*ΔhsdS SCC*	*sau1hsdS_X + sau1hsdS_α*	**RM5_C1C7**
CD150713	ST22/622	*RM6*	*ΔhsdS α*	*sau1hsdS_S + sau1hsdS_X*	**RM6_C2B2***
RM3_A4*	ST45	*ΔRM3 + RM2*	*ΔΔhsdS α+β*	*none*	**RM2+RM3_C1B5**
RM4_C2A2*	ST22/622	*ΔRM4 + RM6*	*ΔΔhsdS orfX+α*	*sau1hsdS_S*	**RM5+6_K45**
RM4_C2A2*	ST22/622	*ΔRM4 + RM5*	*ΔΔhsdS orfX+SCC*	*sau1hsdS_α*	**RM4+RM5_C1C3**
RM6_C2B2*	ST22/622	*ΔRM6 + RM5*	*ΔΔhsdS α+SCC*	*sau1hsdS_X*	**RM4+RM6_C3CD8**

*Mutant strains used for double knockout (2KO; *ΔΔ*), mutant strains in bold were sequenced

2.9.3.7 Mutant Validation

To validate that the mutant clones were successful, the genomic DNA was extracted from the isolates of interest using the NucleoSpin Microbial DNA Kit (Macherey-Nagel, Germany) as per the user manual. The concentration of the genomic DNA was measured with Nanodrop 2000 (Thermofisher, UK). For mutant validation, the gDNA from the isolates were diluted so that the final concentration of DNA in a 20 µl PCR reaction was 100ng/µl. The original A and D primers (Table 2.10) were used as the forward and reverse primer, with the same PCR reaction mixture detailed in Table 2.15. Double KO isolates were tested for both constructs (RM). The PCR products were subsequently mixed with 4 µl Tri-Track Loading dye and loaded onto a gel with 5 µl of 1 kb Ladder and separated (small gel protocol – 120V, 400 milli amp, 50 Watt for 35 minutes) and visualized under UV light. The mutant isolates will ultimately be PacBio sequenced to truly validate the deletion of each *sau1hsdS* of interest.

2.9.3.8 ST622-2015 Typing Primers

Rapid PCR-based assay for detection of ST622-2015 strains using two non-redundant primer pairs (Table 2.18) based on variable genes – *nikB* (ST22 background) and *crtN* (ST45 background).

Table 2.18 | Primer Sequences for ST622-2015 Detection

Gene	ST	A/B	Primer Sequence
crtN	22	A	GCAGATCAATTAATTGAGCAGTACATTGAT
		B	GCGTAAAGATGGTTTCCGATATAATATGC
nikB	45	A	GGTAGTCTAATAGGTGGTACTGTAGTG
		B	CGGTCACTACCTTCTGATATCAATGG

3. SPECIES WIDE CHARACTERISATION OF RESTRICTION-MODIFICATION SYSTEMS & TI RM SAU1 METHYLATION IN *S. AUREUS*

3.1 INTRODUCTION

To date, all four types (TI-IV) of restriction-modification systems have been found in sequenced *Staphylococcus aureus* genomes (Sadykov, 2015). The main RM system remains TI Sau1, which has been shown to actively block intra and inter species horizontal gene transfer (HGT) of MGEs containing resistance determinants (Waldron & Lindsay, 2006). The Sau1 modification systems (core *hsdMS*) within each genomic islands (vSa) have been linked to the stabilisation of these islands within the *S. aureus* genome (Kuroda et al., 2001). The amino acid identity of all HsdM within vSa range from 99-100% within the species (Baba et al., 2002). However, the HsdS protein encoded within the same operon exist in allelic forms with amino acid identity lower than 66%, with the main AA differences associated with the TRD domains (Baba, 2002, Kläui, Boss & Braber, 2019). Differing 'core' *hsdS* alleles have been linked to distinct allelic vSaα and vSaβ forms found within different lineages *of S. aureus* (Baba et al., 2002; Baba, 2008; Kläui, Boss & Braber, 2019).

Although we have some knowledge about the different polymorphic *hsdS*, the structural detail of sequence variations introduced within the DNA binding TRD domains of each HsdS is not well studied. Additionally, the *S. aureus* methylome, including the diversity of Sau1 HsdS target recognition sequences (TRS), the distribution of 6mA spanning the genome, and potential biases of methylation between M_2S modification complexes remains mostly unknown. With the advent of PacBio SMRT sequencing, the TRS of each Sau1 HsdS protein as well as the position of 6mA can be predicted, giving unprecedented information about *S. aureus* TI RM 6mA methylation, linking structure to function: TRD:TRS. Recently, Cooper et al., (2017) began the characterisation of Sau1 HsdS proteins detailing their TRDs and TRS for a representative population of *S. aureus* allowing an initial insight into the variability of these TI system and components.

3.2 AIMS & OBJECTIVES

This study encompasses an extensive, phylogenetically and historically divergent collection of *S. aureus*, representing globally successful strains from various clonal complexes as part of the National Culture Type Collection from Public Health England. The collection of PacBio SMRT genome sequences were used to:

1. Characterize the variability and distribution of RM systems within each lineage.
2. Investigate the variability and distribution of TI RM elements, in specific the diversity attributed to the *hsdS* specificity unit by protein homology analysis.
3. Characterize the target recognition domain (TRD) elements of each *hsdS* variant through secondary structure analysis and investigate the intra and inter-lineage TRD homology.
4. Predict the protein-DNA interactions, TRD:TRS, through binding pocket and protein structure modeling of each HsdS.
5. Use PacBio Modification and Motif analysis to characterize the repertoire of 6mA TI methylation motifs to match structural and functional attributes of these systems and characterize the overall methylation profile of the species.
6. Explore any methylation motif trends between different Sau1 M_2S TRS pertaining to their genetic location which may show distinct functional differences between Sau1 units.
7. Calculate the frequency of 6mA methylation motifs and their location throughout genomes to give insight into:
   a. the overall methylation landscape of *S. aureus*
   b. methylation biases in various genomic regions
   c. methylation biases between various Sau1 units

This largescale characterisation study will give an overall insight into the role of 6mA methylation in *S. aureus* as an epigenetic regulator within this species.

3.4 ORIGINS OF COLLECTION

The isolates utilised in this study were derived from the National Culture Type Collection (NCTC3000) *Staphylococcus aureus* collection at Public Health England (PHE). 108 isolates from the NCTC3000 and an additional 12 references strains (with respective plasmids) were selected, showing large phylogenetic diversity. All NCTC genome sequences and annotations are available from the NCTC3000 project from the Wellcome Sanger Institute (Sanger; https://www.sanger.ac.uk) and the reference strains (n=12) and plasmids are available at European Bioinformatics Institute (EMBL-EBI; https://www.ebi.ac.uk). A complete list of the bacterial strain information and DNA sources has been provided in Table 2.4 and Table 2.5 in Methods.

3.5 RESULTS

3.5.1 Broad-Scale Phylogenetic Overview

To characterise the population structure of the collection and construct a phylogeny, the sequences of the 108 NCTC isolates, augmented with 12 reference strains were mapped to the core genome of CC1 MSSA476 (Genbank: BX571857) (Holden et al., 2004). Analysis of core SNPs between the strains resolved the isolates into 15 clonal complexes (CC) encompassing 37 sequence types (STs) as defined by multi-locus sequences typing (MLST) seen in Figure 3.1.
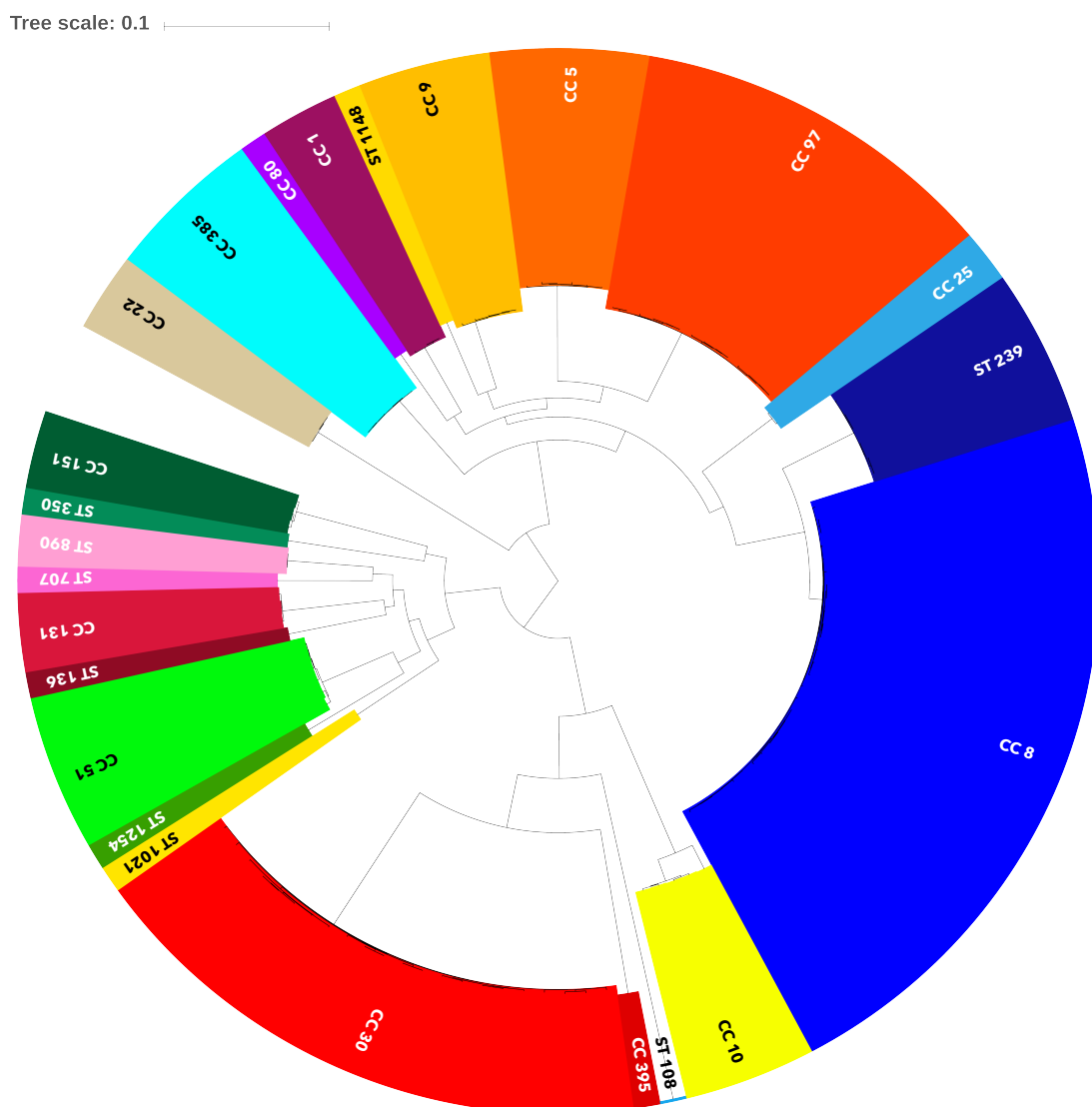
Figure 3.1 | Phylogenetic relationship of NCTC collection of *S. aureus*

120 isolates resolved into lineages highlighted and named according to CC or ST of isolates encompassing 15 CC and 9 singleton ST groups – phylogenetic reconstructions were created using maximum likelihood (ML) analysis on the core genome SNP tree rooted to MSSA476 (A: BX571857) – visualised with iTOL.

Genotypic antimicrobial resistance analysis determined that 79% (n=83/108) were methicillin-susceptible *S. aureus* (MSSA) and 21% (n=23/108) were methicillin resistant S. aureus (MRSA) clones. Out of the added reference strains MSSA476, RF122 and NCTC08325 represented methicillin susceptible strains whilst 9 other added strains of different STs represented methicillin resistant strains for comparison. The majority of the isolates fall into CCs associated with healthcare-associated MRSA (HA-MRSA), community-associated MRSA (CA-MRSA) or major endemic MSSA clones: CC8 (n=34), CC30 (n=20), CC97 (n=15), CC293 (n=7), CC51 (n=7), CC385 (n=6), CC9 (n=4), CC5 (n=5), CC10 (n=5) and CC22 (n=3).

3.5.2 Restriction-Modification Systems

With the population structure of the NCTC isolates defined, the restriction-modification systems present in the collection were characterised. *S. aureus* is known to be highly clonal (Feil et al., 2003), with isolates having a conserved core genome. Most of the variation between strains is seen within the accessory genome, determined by the composition of MGEs within a specific isolate, mainly dictated through horizontal gene transfer (HGT) or recombination. The flexibility of HGT affects the flux of these elements which are in part controlled by restriction-modification (RM) systems. To help understand the variability in accessory genome content and sequences type diversification, BLAST was used to locate and characterise RM elements of various types of systems.

Throughout the historic collection of *S. aureus,* a variety of RM system types including TI, TII and TIV elements (Figure 3.2) were found. In particular, isolates consistently possess elements from at least two of the three types of RM systems, a TI and TII/TIV unit.
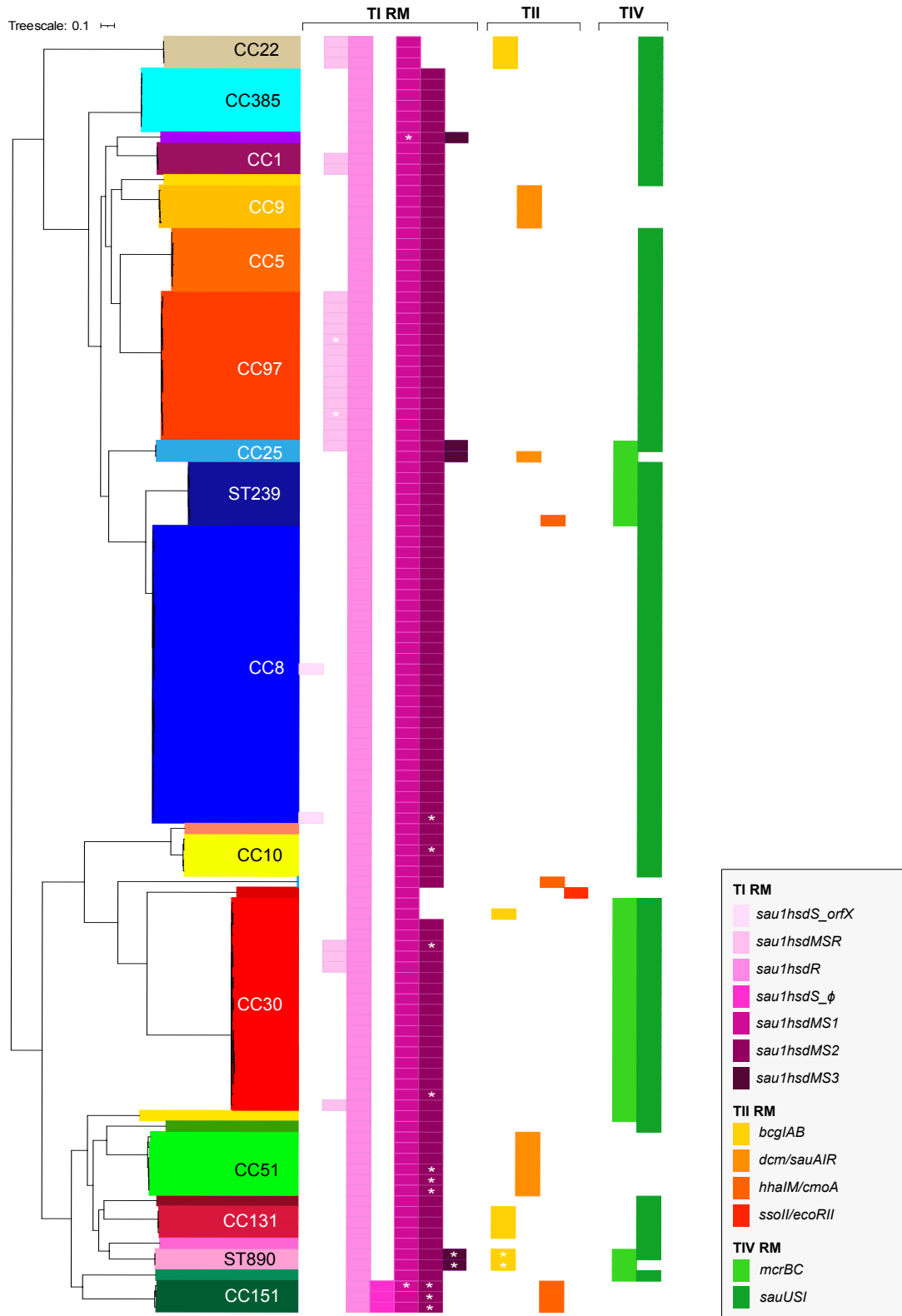
Figure 3.2 | Restriction-Modification Systems within NCTC collection of *S. aureus*
RM Systems found in 120 isolates shown as a vertical ML likelihood core genome SNP tree on the left. Isolates within the same clonal complex (CC) or sequence type (ST) are highlighted in different colours. Three different types of RM systems were present throughout the collection: TI RM represented by the pink colours (*sau1*), TII RM systems represented by the orange colours (*bcgIAB, dcm/sauAIR, hhaIM/cmoAm*, and *ssoII/ecoRII*), and TIV RM systems represented in green (*mcrBC, sauUSI*). Asterisk (*) marks a gene or locus with either a nonsense mutation, frameshift mutation, fragmentation or truncated CDS.

95

3.5.2.1 TI Restriction-Modification Elements

The TI system in *S. aureus* is Sau1 consisting of a 'core' restriction endonuclease (*sau1hsdR - R*) and usually two methyltransferase (*hsdM-M*) and DNA binding specificity CDSs (*hsdS-S*) encoded as an operon *sau1hsdMS* in each genomic island vSα and genomic island vSβ. Sau1HsdMS forms a M_2S complex to methylate double stranded DNA at a specific 3-5 bp bipartite target recognition sequence (TRS) which the specificity unit (*hsdS*) recognises. The R_2M_2S complex is formed for restriction activity, promiscuously cleaving DNA often thousands of base pairs away from the *hsdS* recognised TRS. Within this study, all 120 isolates contained the core restriction gene and *sau1hsdMS1* located in the vSaα region, while only 114/120 isolates contained the second set of *sau1hsdMS2* (reveres orientation) located in vSaβ region. Comparative genomic analysis revealed a further three *sau1* elements within the collection of isolates associated with MGEs. Two of these *sau1* are associated with the *orfX*_SCC junction, *sau1hsdS_orfX* and *sau1hsdRSM*, a single specificity gene *sau1hsdS_ϕ* carried on a possible integrated phage with similarity to ϕ12, as well as *sau1hsdMS3*, another phage element, inserted around 2.5 Mb of the *S. aureus* genome detailed in Figure 3.3 and Table 3.1.
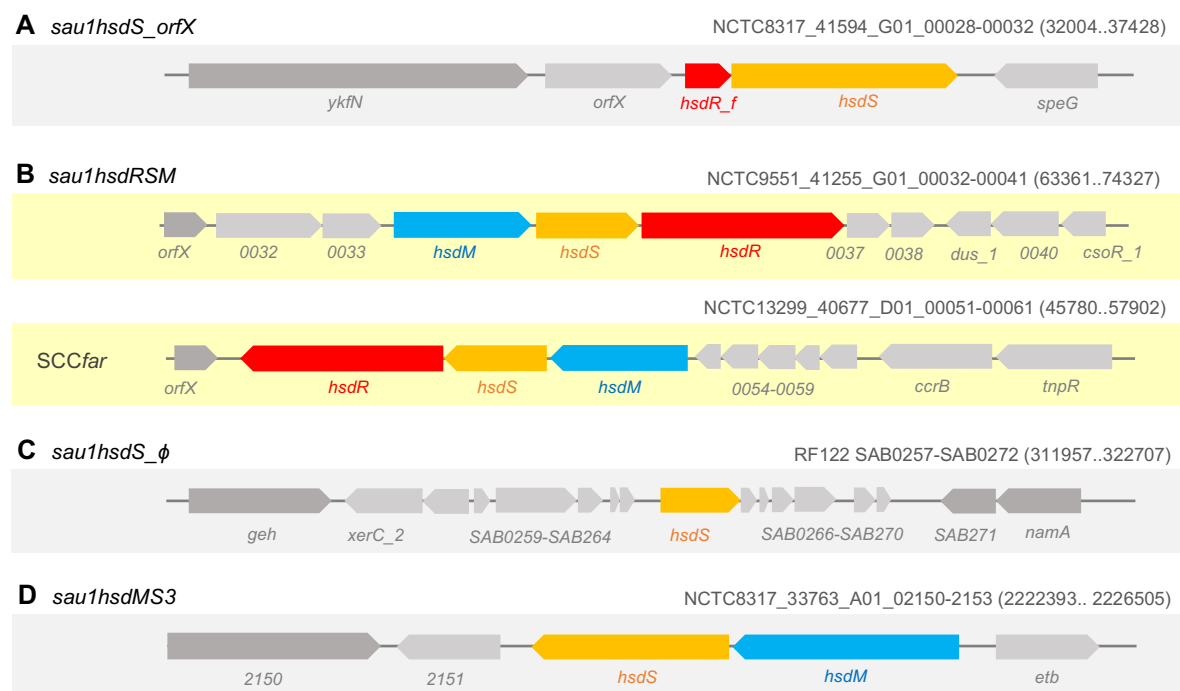
Figure 3.3 | Genetic location and organisation of accessory *sau1* elements
A. *sau1hsdS_orfX* inserted downstream of the *orfX* with a *hsdR* fragment and functional *hsdS*. **B**. *sau1hsdRSM* operon inserted downstream of the *orfX*. Two different positional inserts were characterised: NCTC9551 (upper) *sau1hsdMSR* in most MSSA, or NCTC13299 for isolates with an SCC*far* (lower). MRSA isolates not depicted as the *sau1hsdRMS* is part of the SCC*mec*. C. *sau1hsdS_ϕ* in reference strain RF122 **D.** *sau1hsdMS3* inserted around 2.2-2.5 Mb in the chromosome.

The first 'accessory' *sau1* is inserted directly downstream of the *orfX* and upstream of the core *speG* (Figure 3.3 A) denoted *sau1hsdS_orfX* (HsdS_X). The *orfX* gene harbours the site-specific *attB* insertion site for SCC elements and other MGEs within its C-terminus. This gene is conserved among all staphylococci (Boundy et al., 2013). The *hsdS* is coupled to a small fragment of a non-functional *hsdR* gene. It is unclear which *hsdM* it uses to form a complex for methyltransferase activity. This *hsdS* insert can be shifted downstream of the *orfX,* if the SCC*mec* inserts at the site-specific site, but is always located upstream from the *speG* gene if an extra insertion event occurs. Only 2/120 isolates (both CC8) (Table 3.1) carried this *hsdS* gene which is a potential remnant from a recombination event within the MGE insertion region downstream of the *orfX*.

The second *hsdS* associated with the SCC element, is part of a 3 gene operon, *sau1hsdMSR*, coding for an entire *sau1* system (Figure 3.3 B). This system is also inserted within either directly after the *orfX*, as a part of the SCC*far*, (Wong et al., 2010), directly after the SCC*mec and/*or other transposable elements (plasmids, transposon, ICE). This RM unit is usually flanked by 2/3 hypothetical proteins upstream of the *hsdM* and 2 membrane proteins followed by *dus_1* (tRNA-dihydrouridine synthase), an uncharacterised gene - sulphite exporter TauE/SafE product, and *csoR_1* (copper-sensitive operon repressor). *sau1hsdMSR* was found in 20/120 isolates (CC 97 (n=12/14), CC 1 (n=2/3), CC22 (n=3/3), and some of CC 30 (n=3/18)) (Table 3.1). This full operon of the restriction and modification elements seem to be lineage specific, depending on the mobile element inserted in the region as seen for the ST22, ST1 and ST30 isolates. Both *sau1hsdMSR* and *sau1hsdS_orfX* insert within the SCC*mec* region. It is likely that these two systems are related, but in some isolates the full operon has been lost. This warrants differences in functionality, regarding complex formation with other *hsdM* for *sau1hsdS_orfX*, and potential differences in functionality.

The third accessory *hsdS*, *sau1hsdS_ϕ* was found only in 3 isolates (Table 3.1). The specificity unit is associated with a potential integrated phage remnant with similarity to ϕ12 around 320,000 bp. It is a single specificity unit, flanked by various phage protein coding genes (Figure 3.3 C), with no remnants of other *sau1* elements in contrast to *sau1hsdS_orfX*. This mobile element seems to be lineage specific to only CC151. The last *hsdS* element, is also associated with a prophage remnant found at approximately 2.5 Mb downstream of the *lac* operon denoted *sau1hsdMS3* (Figure 3.3 D). These genes were present in 3 isolates belonging to CC80 (n=1) and CC25 (n=2/2) detailed in Table 3.1.

Table 3.1 | TI RM Accessory *sau1* elements within historic *S. aureus* collection

Isolate	ST	SCC	hsdM (identifier)	hsdS (identifier)	hsdR (identifier)
sau1hsdS_orfX					
NCTC13394	8	Type II		40677_F01_00062	
NCTC10703	3526			41594_G01_00031	
sau1hsdMSR					
EMRSA15	22	Type IV	SAEMRSA15_00450	SAEMRSA15_00451	SAEMRSA15_00452
NCTC13142	22	Type IV	40415_C02_00054	40415_C02_00055	40415_C02_00056
NCTC13616	22	Type IV	35910_F01_00054	35910_F01_00055	35910_F01_00056
MSSA476	1	SCC*far*	SAS0027	SAS0026	SAS0025
NCTC13297	1	SCC*far*	40677_B01_00032	40677_B01_00031	40677_B01_00030
NCTC5658	464		41559_B01_00032	41559_B01_00033	41559_B01_00034
NCTC3761	464		40961_F01_00031	40961_F01_00032	40961_F01_00033
NCTC4137	464		40961_G01_00031	40961_G01_00032	40961_G01_00033
NCTC4163	464		41255_C01_00032	41255_C01_00033	41255_C01_00034
NCTC13841	464		50450_A01_00032	50450_A01_00033	50450_A01_00034
NCTC8399	97		41004_D02_00058	41004_D02_00059	41004_D02_00060
NCTC9547	97		40961_C01_00033	40961_C01_00034	40961_C01_00035
NCTC9552	97		42545_H02_00034	42545_H02_00035	42545_H02_00036
NCTC9551	97		41255_G01_00034	41255_G01_00035	41255_G01_00036
NCTC7121	97		40677_B02_00107	40677_B02_00108	40677_B02_00109
NCTC10344	97		41236_E02_00108	41236_E02_00109	41236_E02_00110
NCTC10345	97	SCC*far-like*	41236_G01_00128	41236_G01_00129	41236_G01_00130
NCTC13299	30	SCC*far-like*	40677_D01_00054	40677_D01_00053	40677_D01_00052
NCTC6571	30	SCC*far-like*	40677_H01_00031	40677_H01_00030	40677_H01_00029
NCTC7361	30	SCC*far-like*	40853_H01_00846	40853_H01_00847	40853_H01_00848
sau1hsdS_φ					
RF122	151			SAB0265	
NCTC12880	151			49657_B01_01995	
NCTC7485	151			40853_C01_00279	
sau1hsdMS3					
NCTC13435	80	Type II	27294_E01_02276	27294_E01_02275	
NCTC6134	25		40798_D01_01968	40798_D01_01967	
NCTC8317	25		33763_A01_02152	33763_A01_02151	

3.5.2.2 TII Restriction-Modification Elements

There are three distinct TII RM system types in the NCTC collection: TIIG - *bcgIAB* (RMS – 6mA), TIIC - *dcm/sau3Air* (RM), and *hhaIM/cmoA* (RM) having RM activity specifically at 5 and 4-methyl-cytosine (5mC, 4mC) nucleotide position close to 4-7 bp recognition sites. All TII systems present were lineage specific, associate with the presence of an MGE or MGE remnants (prophage (φ42 lysogen, transposon, SaPI (Veiga and Pinho, 2009; van Wamel et al., 2006)) detailed in Table 3.2. A total of 26 isolates contained a TII system of which: 9 isolates carried *bcgIAB lineage specific for* ST22*,* ST133 *and* ST890*,* 11 isolates carried *dcm/sau3Air – lineage specific for* ST9 *and* ST121, 5 isolates carried *hhaIM/cmoA*. Reference strain ST398 had a different inserted *E. coli* derived TII system *ecoRII*/*ssoII*.

Table 3.2 | TII RM Elements within the historic *S. aureus* collection

Isolate	ST	Modification	*Restriction*
TIIG bcgIAB_6mA		*bcgIB (R+M)*	*bcgIA (S)*
NCTC13142	22	40415_C02_01446	40415_C02_01447
NCTC13616	22	35910_F01_01424	35910_F01_01425
EMRSA15	22	SAEMRSA1513490	SAEMRSA1513491
NCTC11963	36	41556_F01_01454	41556_F01_01455
NCTC7988	133	40798_F01_01929	40798_F01_01930
NCTC1803	133	43874_B01_01610	43874_B01_01611
NCTC9555	133	41255_H01_01937	41255_H01_01938
NCTC6966	890	40677_A02_00797	40677_A02_00799
NCTC7856	890	40677_E02_00775	40677_E02_00777
TII dcm/sau3AIR_5mC		*dcm*	*sau3AIR (AdoMet binding)*
NCTC8765	9	40853_G01_02509	40853_G01_02510
NCTC6136	9	41004_D01_02575	41004_D01_02576
NCTC8725	9	41004_F02_02503	41004_F02_02504
NCTC8723	9	42042_D02_02521	42042_D02_02522
NCTC8317	25	33763_A01_02428	33763_A01_02427
NCTC13434	121	40677_G01_02521	40677_G01_02522
NCTC13298	121	40677_C01_02618	40677_C01_02619
NCTC8531	121	43295_D02_02558	43295_D02_02559
NCTC7791	121	40871_A01_01308	40871_A01_01307
NCTC3750	121	40961_E01_01195	40961_E01_01196
NCTC7414	121	41004_E01_02693	41004_E01_02694
TII hhaIM/cmoA (prophage) 5mC		*haIM*	*cmoA*
NCTC13626	239	35910_E01_02176	35910_E01_02177
NCTC8726	9	41315_B01_02436	41315_B01_02437
RF122	151	SAB2369	SAB2370
NCTC12880	151	40657_B01_01709	40657_B01_01708
NCTC7485	351	40853_C01_02419	40853_C01_02420
TII ssoII/ecoRII		*ecoRII*	*ssoII*
ST398	398	SAPIG2546	SAPIG2545

3.5.2.3 TVI Restriction-Modification Elements

TIV RM only have restriction activity, but they cleave non-specifically for 6mA, 4mC, 5mC and hydroxymethylated or glycosyl-hydroxymethylated nucleotides as well. Most isolates (n=103/120) also contained a TIV restriction element either as *sauUSI (annotated as srmB in most genomes)* (n=101/120) or *mcrBC* (n=32/120) which cleave methylated DNA with low specificity. Lineage specificity can be seen among isolates containing *mcrBC* which is associated with the SCC element insertion region clearly seen for CC30, ST890 ST239 as detailed in Table 3.3.

Table 3.3 | TIV RM *mcrBC* elements within the historic *S. aureus* collection

Isolate	ST	Unit A	*Unit B*
mcrBC		*mcrC*	*mcrB*
NCTC6134	25	40798_D01_02334	40798_D01_02333
NCTC8317	25	33763_A01_00031	33763_A01_00030
NCTC13135	239	40657_C01_00097	40657_C01_00096
NCTC11940	239	42197_E02_00030	42197_E02_00029
NCTC11939	239	40415_E02_00170	40415_E02_00169
NCTC13134	239	46837_D02_00107	46837_D02_00106
TW20	239	SATW20_00990	SATW20_00980
NCTC13626	239	35910_E01_00106	35910_E01_00105
NCTC5656	30	41255_D01_00657	41255_D01_00656
NCTC11963	36	41556_F01_00051	41556_F01_00050
NCTC13373	36	35910_G01_00168	35910_G01_00167
NCTC12981	243	34347_A01_00127	34347_A01_00126
NCTC13299	30	40677_D01_00082	40677_D01_00081
NCTC6571	30	40677_H01_00059	40677_H01_00058
NCTC7361	30	40853_H01_00818	40853_H01_00819
NCTC7446	30	41004_B02_00067	41004_B02_00066
NCTC7445	30	41004_A02_00036	41004_A02_00035
NCTC8507	30	40871_C01_00082	40871_C01_00081
NCTC8530	30	40740_E01_00036	40740_E01_00035
NCTC2669	30	40105_H02_00068	40105_H02_00067
NCTC13811	30	49386_H02_02666	49386_H02_02665
NCTC11965	30	41556_G01_00036	41556_G01_00035
NCTC13143	30	41665_B02_02446	41665_B02_02445
MRSA252	36	SAR0088	SAR0087
NCTC13277	30	35910_A02_00089	35910_A02_00088
NCTC11962	30	41665_C02_01220	41665_C02_01221
NCTC11561	30	41315_E02_00037	41315_E02_00036
NCTC5655	30	40757_B01_00038	40757_B01_00037
NCTC6135	1021	41004_C01_00114	41004_C01_00113
NCTC6966	890	40677_A02_00053	40677_A02_00052
NCTC7856	890	40677_E02_00030	40677_E02_00029
NCTC5663	350	41004_B01_00030	41004_B01_00029

3.5.3 *S. aureus* TI Sau1 6ma Motifs and HsdS Structure

Having established the combination of RM complexes within each strain, the 6mA methylation signatures were investigated with PacBio SMRT sequencing and matching matched to the respective categorised Sau1 HsdS.

In staphylococci the main methylation signature is 6mA facilitated by Sau1 M_2S modification complex, which is the main focus of this study. A single 6mA methylation motif consists of two target recognition sequence (TRS) strings, one on the forward and one on the reverse DNA strand (5' → 3') as represented by AGGNNNNNGAT / ATCNNNNNCCT, in Figure 3.4. Each TRS string is made of two bipartite sequences separated by a spacer region (N), in which the 5' sequence contains the target adenine for modification.
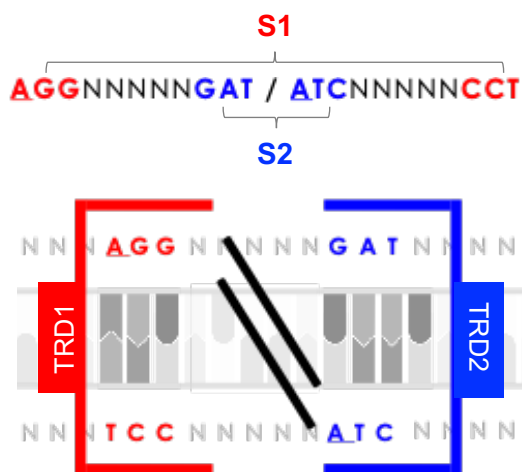
Figure 3.4 | 6mA Methylation Motif
Bipartite 6mA TRS sequences (on forward and reverse DNA strand) bound by *sau1* HsdS **TRD1** at **S1** and **TRD2** at **S2** of HsdS. Black diagonal lines indicate conserved helical structure of HsdS protein holding the two TRS apart.

HsdS proteins recognise DNA at target sequences, and bind the nucleotide strings on both DNA strands, with their two target recognition domains denoted TRD1 and TRD2. Both TRDs bind both strands of DNA at complementary base sequences denoted S1 and S2, recognised by TRD1 and TRD2 respectively. The two DNA binding domains are held apart by a conserved helical structure. The distance and torsion angle at which TRDS are bound to the helix has been linked to the length of the spacer region within each TRS as described by Loenen et al., 2014.

PacBio SMRT sequencing allows the detection of modified nucleotide bases on a single base level as well as providing a putative methylation motif and location where they occur within a whole genome. SMRT analysis and protein structure predictions were used to investigate the methylation signatures of each isolate within the NCTC collection.

3.5.3.1   PacBio SMRT Methylation Analysis – *sau1* 6mA TRS

PacBio SMRT sequencing technology allows the detection of modified nucleotide bases within DNA reads by investigating variations within polymerase kinetics of single DNA base incorporation in the form of inter-pulse durations (IPDs). IPDs are defined as the time duration between two successive base incorporation events by the DNA polymerase. The IPD is altered by the speed of base incorporation by the DNA polymerase, which in the presence of a modified base in the DNA template is slowed (higher IPD). Sequence motifs associated with DNA modifications are predicted, to categorise the versatility of modified bases by varying active restriction-modification complexes present in a given isolate.

PacBio SMRT RS Modifications and Motif Analysis was run on 120 PacBio sequenced *S. aureus* strains. A collection of 42 different 6mA, with highlighted S1/S2 TRS detailed in Table 3.4. The IPD ratios for the modified adenine bases within each methylation TRS (collated from all isolates with recognition sequence of interest – raw predictions detailed under Supplementary Table 8.1 in the Appendix) ranged between 3.778-8.049 (median IPD: 5.087 (calculated from raw data per described by Sanchéz-Busó et al., 2019), with a methylated/detected TRS ratio of 97.30% (median: 98.10% - range: 80.30-100.00%). The mean IPD value decreases if the proportion of modified bases per detected motif decreases calculated as the fraction of molecules that carry the modification. In 28/42 predicted 6mA motifs the forward, and 12/42 the reverse strands had higher methylated TRS / detected TRS ratio and higher IPD ratio, indicating lesser modification efficacy for one of the strands of DNA at any given time. This could also be an artifact of methylation prediction itself.

There were 9 isolates in which only half of a methylation TRS was reported for TRS #8 (n=1), #16 (n=3), and #26 (n=5) highlighted in green (Table 3.4). Within these isolates, the number of motifs detected, and methylated stay correlated to those reported in other isolates with the same, double stranded motif string. To investigate whether the half string motifs are examples of solely hemi-methylated 6mA TRS, a closer look was taken at the raw call data of the of the DNA polymerisation kinetics for base incorporation. Using scripts developed by Leonor Sánchez Busó (Wellcome Sanger Institute - https://github.com/leosanbu/MethylationProject/tree/master/scripts/visualize_IPDs) the raw IPD data for each single string motif were plotted to explore each methylation event for the given TRS prediction on double stranded DNA, reporting back the modified base on both the forward (red) and complement strand (blue) show in Figure 3.5.

Table 3.4 | PacBio SMRT Modification and Motif Summary

TRS #	ST/CC	Strand	Methylation motif (F/R 5') / TRS String	mean Modified / Detected TRS Ratio	mean IPD Ratio Min	mean IPD Ratio Max
1	CC8 CC5		AGGNNNNNGAT/ATCNNNNNCCT	0.998	5.218	7.049
		F	AGGNNNNNGAT	0.999	6.399	7.049
		R	ATCNNNNNCCT	0.998	5.218	5.431
2	CC30		CAGNNNNNRAAT/ATTYNNNNNCTG	0.983	5.412	8.049
		F	CAGNNNNNRAAT	0.987	7.583	8.049
		R	ATTYNNNNNCTG	0.982	5.412	5.896
3	CC131		CAGNNNNNRTGA/TCAYNNNNNCTG	0.966	4.603	5.875
		F	CAGNNNNNRTGA	0.970	5.627	5.875
		R	TCAYNNNNNCTG	0.961	4.603	4.761
4	ST198 ST398		ACCNNNNNRTGA/TCAYNNNNNGGT	0.972	4.350	4.597
		F	ACCNNNNNRTGA	1.000		4.597
		R	TCAYNNNNNGGT	0.943		4.350
5	CC10		GACNNNNNNTAG/CTANNNNNNGTC	0.951	4.313	4.756
		F	GACNNNNNNTAG	0.981	4.563	4.756
		R	CTANNNNNNGTC	0.932	4.313	4.422
6	CC51		GGANNNNNNCCT/AGGNNNNNNTCC	0.985	4.689	5.969
		F	GGANNNNNNCCT	0.983	4.689	5.006
		R	AGGNNNNNNTCC	0.990	5.673	5.969
7	ST350		GAGNNNNNNGAT/ATCNNNNNNCTC	0.995	5.073	5.992
		F	GAGNNNNNNGAT	0.995		5.992
		R	ATCNNNNNNCTC	0.996		5.073
8	CC385		GACNNNNNNTGG/CCANNNNNNGTC	0.950	3.994	4.547
		F	<mark>GACNNNNNNTGG</mark>	0.979	4.311	4.547
		R	CCANNNNNNGTC	0.945	3.994	4.487
9	CC22		AGGNNNNNNTGAR/YTCANNNNNNCCT	0.989	5.309	6.471
		F	AGGNNNNNNTGAR	0.989	5.547	6.471
		R	YTCANNNNNNCCT	0.993	5.309	5.749
10	CC97 CC80		GACNNNNNNTTYG/CRAANNNNNNGTC	0.940	4.084	5.531
		F	GACNNNNNNTTYG	0.991	4.999	5.531
		R	CRAANNNNNNGTC	0.930	4.084	4.593
11	CC51		GACNNNNNNTAYG/CRTANNNNNNGTC	0.969	4.220	4.823
		F	GACNNNNNNTAYG	0.988	4.619	4.823
		R	CRTANNNNNNGTC	0.956	4.220	4.500
12	CC890		GGANNNNNNNRTGA/TCAYNNNNNNNTCC	0.928	3.778	4.940
		F	GGANNNNNNNRTGA	0.896	3.778	3.817
		R	TCAYNNNNNNNTCC	0.969	4.909	4.940
13	ST1254		GAGNNNNNNNRTTC/GAAYNNNNNNNCTC	0.975	4.817	6.364
		F	GAGNNNNNNNRTTC	0.989		6.364
		R	GAAYNNNNNNNCTC	0.962		4.817
14	CC30		GGANNNNNNNTCG/CGANNNNNNNTCC	0.943	4.264	5.264
		F	GGANNNNNNNTCG	0.944	4.468	5.264
		R	CGANNNNNNNTCC	0.940	4.264	4.951
15	ST1148		GGANNNNNNNTGC/GCANNNNNNNTCC	0.974	4.307	5.166
		F	GGANNNNNNNTGC	0.958		4.307
		R	GCANNNNNNNTCC	0.989		5.166
16	CC10 CC131		GGANNNNNNNTTRG/CYAANNNNNNNTCC	0.957	4.765	5.375
		F	GGANNNNNNNTTRG	0.949	4.765	5.001
		R	<mark>CYAANNNNNNNTCC</mark>	0.968	5.187	5.375
17	CC22 CC97		GAAGNNNNNTAC/GTANNNNNCTTC	0.994	6.044	7.248
		F	GAAGNNNNNTAC	0.992		7.248
		R	GTANNNNNCTTC	0.996		6.044
18	CC25		CCAYNNNNNGAT/ATCNNNNNRTGG	0.958	4.246	5.077
		F	CCAYNNNNNGAT	0.944	4.246	4.748
		R	ATCNNNNNRTGG	0.994	5.034	5.077

#	Group	F/R	Sequence			
19	CC30		GWAGNNNNNGAT/ATCNNNNNCTWC	0.989	5.178	6.945
		F	GWAGNNNNNGAT	0.989	5.654	5.828
		R	ATCNNNNNCTWC	0.989	5.178	6.945
20	ST707		AGGNNNNNRTGG/CCAYNNNNNCCT	0.976	4.818	5.362
		F	AGGNNNNNRTGG	0.968		5.362
		R	CCAYNNNNNCCT	0.984		4.818
21	ST1254		GAAANNNNNCCT/AGGNNNNNTTTC	0.987	5.086	6.484
		F	GAAANNNNNCCT	0.985		5.086
		R	AGGNNNNNTTTC	0.988		6.484
22	CC1 CC80		CCAYNNNNNTTAA/TTAANNNNNRTGG	0.974	4.933	5.118
		F	CCAYNNNNNTTAA	0.985	5.063	5.118
		R	TTAANNNNNRTGG	0.962	4.933	4.970
23	CC5, CC25		CCAYNNNNNVGTA/TACBNNNNNRTGG	0.958	4.394	5.449
		F	CCAYNNNNNVGTA	0.950	4.394	4.843
		R	TACBNNNNNRTGG	0.993	4.745	5.449
24	ST498		CCAYNNNNNRTTT/AAAYNNNNNRTGG	0.961	4.386	5.212
		F	CCAYNNNNNRTTT	0.983		5.212
		R	AAAYNNNNNRTGG	0.939		4.386
25	CC890		GWAGNNNNNRTKC/GMAYNNNNNCTWC	0.815	4.625	6.036
		F	GWAGNNNNNRTKC	0.824	5.988	6.036
		R	GMAYNNNNNCTWC	0.804	4.625	4.666
26	CC1 CC8		CCAYNNNNNNTGT/ACANNNNNNRTGG	0.941	4.240	5.159
		F	ACANNNNNNRTGG	0.941	4.376	5.159
		R	CCAYNNNNNNTGT	0.932	4.240	4.733
27	ST350		GAAGNNNNNNTGT/ACANNNNNNCTTC	0.990	5.018	5.620
		F	GAAGNNNNNNTGT	0.996		5.620
		R	ACANNNNNNCTTC	0.985		5.018
28	CC8		TAAGNNNNNNTTC/GAANNNNNNCTTA	0.994	5.048	6.861
		F	TAAGNNNNNNTTC	0.998	6.802	6.861
		R	GAANNNNNNCTTA	0.988	5.048	5.122
29	ST1148 CC97		CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.962	4.422	5.054
		F	CCAYNNNNNNRTC	0.977	4.636	5.054
		R	GAYNNNNNNRTGG	0.958	4.422	4.832
30	CC9		TCTANNNNNNNTTAA/TTAANNNNNNNTAGA	0.993	5.480	6.308
		F	TCTANNNNNNNTTAA	0.991	5.936	6.308
		R	TTAANNNNNNNTAGA	0.994	5.480	5.880
31	CC9		GAAGNNNNNNTTRG/CYAANNNNNNCTTC	0.989	4.883	6.886
		F	GAAGNNNNNNTTRG	1.000	6.474	6.886
		R	CYAANNNNNNCTTC	0.981	4.883	5.184
32	CC97 ST498		CCAYNNNNNNTTYG/CRAANNNNNNRTGG	0.968	4.314	5.143
		F	CCAYNNNNNNTTYG	0.986	4.891	5.143
		R	CRAANNNNNNRTGG	0.950	4.314	#######
33	CC1 CC25		TCTANNNNNNRTTC/GAAYNNNNNNNTAGA	0.983	4.862	5.864
		F	GAAYNNNNNNNTAGA	0.973	4.862	5.293
		R	TCTANNNNNNNRTTC	0.987	5.431	5.864
34	CC151		CAAGNNNNNNNTARC/GYTANNNNNNNCTTG	0.990	5.160	6.061
		F	CAAGNNNNNNNTARC	1.000	6.042	6.061
		R	GYTANNNNNNNCTTG	0.985	5.160	5.288
35	ST198		CAACNNNNNNNTAYG/CRTANNNNNNNGTTG	0.959	4.385	4.954
		F	CAACNNNNNNNTAYG	1.000		4.954
		R	CRTANNNNNNNGTTG	0.917		4.385
36	CC385		CCAYNNNNNNNTAAA/TTTANNNNNNNRTGG	0.953	4.638	5.242
		F	CCAYNNNNNNNTAAA	0.944	4.638	4.774
		R	TTTANNNNNNNRTGG	0.964	5.097	5.242
37	ST707		CAAYNNNNNCTTC/GAAGNNNNNRTTG	0.990	4.542	6.121
		F	CAAYNNNNNCTTC	0.983		4.542
		R	GAAGNNNNNRTTG	0.998		6.121
38	ST1021		GARANNNNNNYTCC/GGARNNNNNNTYTC	0.976	5.004	5.217
		F	GARANNNNNNYTCC	0.971		5.004
		R	GGARNNNNNNTYTC	0.981		5.217

39	ST1021		GARANNNNNNNRTYC/GRAYNNNNNNNTYTC	0.979	4.742	4.995
		F	GARANNNNNNNRTYC	0.977		4.742
		R	GRAYNNNNNNNTYTC	0.982		4.995
40	CC1		GNNGANNNNNNNRTTA/TAAYNNNNNNNNTCNNC	0.981	4.987	5.515
		F	GNNGANNNNNNNRTTA	0.988		5.515
		R	TAAYNNNNNNNNTCNNC	0.974		4.987
41	CC22 CC30		YACNNNNNTGG/CCANNNNNGTR	0.987	4.945	5.706
		F	YACNNNNNTGG	0.987	4.945	5.494
		R	CCANNNNNGTR	0.989	5.255	5.706
42	CC131		CGANNNNNNNTAC/GTANNNNNNNTCG	0.978	4.849	5.445
		F	CGANNNNNNNTAC	0.982	4.849	5.008
		R	GTANNNNNNNTCG	0.978	5.281	5.445

PacBio SMRT Analysis results from motif_summary.csv output, detailing the double stranded DNA 6mA TRS string, as well as the two single TRS strings on the forward and reverse DNA strand separated with (/).The modified adenine nucleotide is underlined. Each double stranded TRS (grouped TRS) details the S1 (red) and S2 (blue) sequences which HsdS TRDs bind. The mean of the methylated / detected TRS are shown as well as the IPD ratio concatenated from all isolates with the given motif of interest. All represented motifs had a minimum coverage of 100 (deemed to be the cuttoff for lower quality methylation prediction).
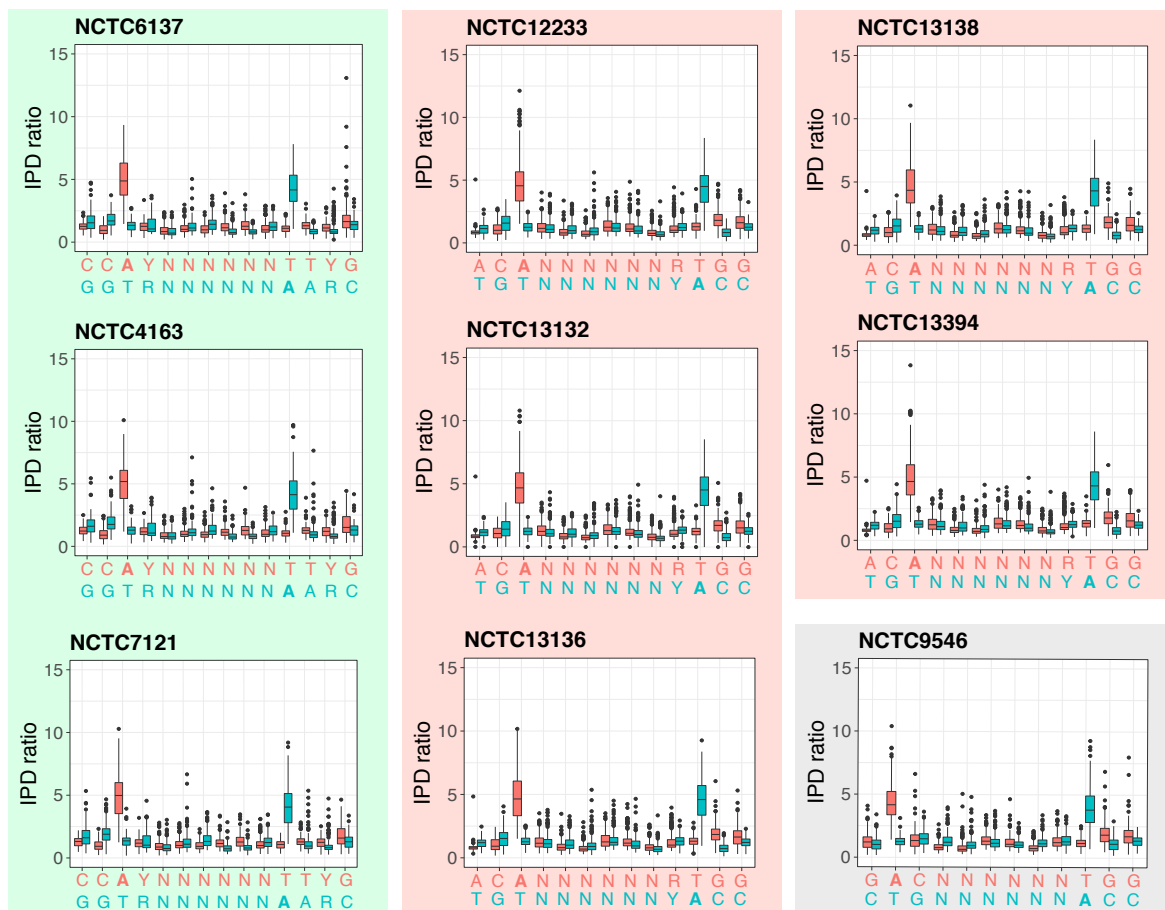
Figure 3.5 | IPD ratio of single 6mA methylation motifs of TI RM HsdS from SMRT Restriction and Modification analysis in 9 isolates of *S. aureus* NCTC collection.

Per-base distribution of IPD ratios are shown by box plots representing values in the forward (red) and reverse (blue) strand of DNA with the accompanying predicted methylation motifs. Pink box: CC8, green box: CC10 (NCTC6137) and CC97, grey box: CC385.

The IPD box plots showed the presence of modified A bases on the complement strand for each of the half strings presented, with equal meanIPD ratios ranging from 4.89-6.65 IPD, whilst unmodified bases had a median IPD score of 0.836 (range 0.022-3.113 IPD). The presence of the modification on the complement strand thus rules out the presence of hemi-methylated 6mA motifs in any of the isolates and suggests that the single motifs were an artefact of errors in the methylation base call potential due to low quality sequence assembly.

Thirteen motifs (#1, #4, #10, #11, #16, #17, #22, #23, #26, #29, #33, #34, #41 – Table 3.4 highlighted in grey) were represented across multiple linages, suggesting possible conservation of HsdS proteins across lineages. HsdS encoded within the genomic islands (vSa), have been linked to the allelic form associated with the islands themselves (Baba et al., 2002), potential explaining the homogeneity in reported 6mA TRS. Some methylation motifs may also belong to the 'accessory' genome associated HsdS, which may be more readily transferred between lineages of the species via HGT. In *S. aureus*, it is not just the Sau1 system but also the TIIG BcgIAB system (found within 9 strains) which have a 6mA epigenetic signature. To further characterise the diversity of each HsdS and to match the methylation motif (TRS) to the right Sau1 HsdS (dependent on location of CDS within genome), the protein sequence of each specificity unit was analysed next.

3.5.3.2 Characterisation of TRDs with corresponding *sau1* 6mA TRS

Within the previous section, it was established that methylation target motifs vary throughout the *S. aureus* species, mainly following clonal specificity. Differing 'core' *hsdS* alleles have been linked to distinct allelic vSaα and vSaβ forms found within different lineages *of S. aureus* (Baba et al., 2002; Baba, 2008; Kläui, Boss & Braber, 2019). Within this study, the conservation of *hsdS* within different genomic locations and whether lineage specificity is present not just for the 'core' HsdS_α, HsdS_α β, but also three 'accessory' HsdS_X, HsdS_S, HsdS_E was investigated; HsdS_ϕ may be non-functional as it did not have a corresponding motif. The protein sequences of each HsdS were analysed to be able to match the methylation motifs (TRS) to a specific HsdS binding domain (TRD). By knowing which HsdS each methylation motif belongs to, there is potential to investigate any structural or functional bias for either one of these specificity units.

The protein sequences of a collection 251 *hsdS* within the NCTC isolates were examined. As the HsdS protein holds two globular domains, the protein sequences were broken in the middle-conserved sequence, to be able to identify where each TRD begins and ends. Each 'half' HsdS sequence was multiple aligned to cluster AA sequences with analogous sequence similarity. These sequence collections were compared to 'half' HsdS amino acid strings previously resolved and assigned to HsdS proteins by Cooper et al., (2017). They classified TRD targets (S1/S2) for isolates belonging to 17 ST/CC types, assigning identifying letters and TRS target sequences (S) which each TRD binds (e.g.: A_CCAY). Exploiting the previously identified database of Sau1 TRS S:TRD (Cooper et al., 2017), we were able to resolve the protein sequences of HsdS within the NCTC collection isolates detailed in Table 3.5. Using Table 3.5 as a key for the TRD classification, the TRD composition of each HsdS in relation to the genetic location of the *hsdS* genes were also visualised in Figure 3.6. This allowed the analysis of diversity of TRD:TRS S1/S2 and lineage specificity of certain HsdS in context of the maximum likelihood phylogeny tree (core genome).

This analysis only includes functional HsdS which were matched to a given Sau1 6mA methylation motif. HsdS_ϕ (coded for by *sau1hsdS_ϕ* carried on CC151 isolates) seems to be non-functional as all three isolates which carried this specificity gene, only produced on single methylation motif (Table 3.4 #34) which was matched to HsdS_α. All further analysis denoting 'accessory' motifs and proteins only include those for HsdS_O, HsdS_S and HsdS_E.

Table 3.5 | *Sau1* HsdS Specificity Unit TRD & TRS

# TRS	TRS (S1 + S2)	TRD1	TRD2	ST/CC
HsdS_α - *sau1hsdSα* (*sau1hsdMS1*)				
20	AGGNNNNNNRTGG/CCAYNNNNNCCT	NT1*G_GAAG	NT2*U_CCAY	707
9	AAGNNNNNNNTGAR/YTCANNNNNNNCCT	B_AGG	I_YTCA	22
1	AGGNNNNNGAT/ATCNNNNNCCT	**B_AGG**	**D_ATC**	**8, 5**
19	GWAGNNNNNGAT/ATCNNNNNCTWC	C_GWAG	D_ATC	30
7	GAGNNNNNNGAT/ATCNNNNNNCTC	e*_GAG	D_ATC	350
18	CCAYNNNNNGAT/ATCNNNNNRTGG	A_CCAY	D_ATC	25
36	CCAYNNNNNNTAAA/TTTANNNNNNNRTGG	A_CCAY	L_TTTA	385
29	CCAYNNNNNNNRTC/GAYNNNNNNNRTGG	**A_CCAY**	**c*_GAY**	**1148, 97**
24	CCAYNNNNNNRTTT/AAAYNNNNNRTGG	A_CCAY	NT2*S_AAAY	498
22	CCAYNNNNNNTTAA/TTAANNNNNNRTGG	**A_CCAY**	**F_TTAA**	**1, 136, 151**
30	TCTANNNNNNNTTAA/TTAANNNNNNNTAGA	X_TCTA	F_TTAA	9
4	ACCNNNNNNRTGA/TCAYNNNNNNGGT	**N_ACC**	**E_TCAY**	**198, 398**
3	CAGNNNNNNRTGA/TCAYNNNNNCTG	M_CAG	E_TCAY	**131**
12	GGANNNNNNNRTGA/TCAYNNNNNNNTCC	J_GGA	E_TCAY	890
6	GGANNNNNNNCCT/AGGNNNNNNNTCC	J_GGA	P_AGG	51
21	GAAANNNNNCCT/AGGNNNNNTTTC	NT1*H_GAAA	P_AGG	1254
5	GACNNNNNNNTAG/CTANNNNNNNGTC	Z_GAC	Y_CTA	10
38	GARANNNNNNYTCC/GGARNNNNNNNTYTC	R_GARA	NT2*R_GGAR	1021
34	CAAGNNNNNNNTARC/GYTANNNNNNNCTTG	T_CAAG	NT2*M_GYTA	151
HsdS_β - *sau1hsdSβ* (*sau1hsdMS2*)				
35	CAACNNNNNNNTAYG/CRTANNNNNNNGTTG	O_CAAC	N_3_CRTA	198
11	GACNNNNNNNTAYG/CRTANNNNNNNGTC	**Z_GAC**	**N_3_CRTA**	**51, 136**
8	GACNNNNNNNTGG/CCANNNNNNNGTC	Z_GAC	NT2*N_CCA	385
10	GACNNNNNNNTTYG/CRAANNNNNNNGTC	**Z_GAC**	**W_CRAA**	**97, 80**
32	CCAYNNNNNNNTTYG/CRAANNNNNNNRTGG	**A_CCAY**	**W_CRAA**	**97, 498**
23	CCAYNNNNNVGTA/TACBNNNNNRTGG	**A_CCAY**	**H_TACB**	**5, 25**
37	CAAYNNNNNCTTC/GAAGNNNNNRTTG	A_CCAY	NT2*V_GAAG	707
26	CCAYNNNNNNTGT/ACANNNNNNNRTGG	**A_CCAY**	**G_ACA**	**1, 8, 151**
27	GAAGNNNNNNNTGT/ACANNNNNNNCTTC	NT1*G_GAAG	G_ACA	350
25	GWAGNNNNNRTKC/GMAYNNNNNCTWC	NT1*G_GAAG	S_GCA	890
31	GAAGNNNNNNNTTRG/CYAANNNNNNNCTTC	NT1*G_GAAG	d*_CYAA	9
16	GGANNNNNNNTTRG/CYAANNNNNNNNTCC	**J_GGA**	**d*_CYAA**	**131, 10**
14	GGANNNNNNNTCG/CGANNNNNNNNTCC	J_GGA	K_CGA	30
15	GGANNNNNNNTGC/GCANNNNNNNNTCC	J_GGA	S_GCA	1148
39	GARANNNNNNNRTYC/GRAYNNNNNNNTYTC	R_GARA	f*_GAAY	1021
13	GAGNNNNNNNRTTC/GAAYNNNNNNNCTC	e*_GAG	f*_GAAY	1254
HsdS_X - *sauhsdS_orfX*				
28	TAAGNNNNNNNTTC/GAANNNNNNNCTTA	NT1*A_TAAG	a*_GAA	8
HsdS_S - *sau1hsds_SCC* (*sau1hsdRMS*)				
17	GAAGNNNNNTAC/GTANNNNNCTTC	**NT1*X_GAAG**	**NT2*O_GTA**	**22, 97**
40	GNNGANNNNNNNRTTA/TAAYNNNNNNNNTCNNC	NT1*C_GNNGA	NT2*P_TAAY	1
2	CAGNNNNNRAAT/ATTYNNNNNCTG	NT1*E_CAG	NT2*Q_ATTY	30
HsdS_E - *sau1hsdS_E* (*sau1hsdMS3*)				
33	TCTANNNNNNRTTC/GAAYNNNNNNNTAGA	**X_TCTA**	**f*_GAAY**	**80, 25**

Bipartite TRS sequences bind by each HsdS. TRS target sequences, S1 and S2 were matched to TRD1 and TRD2; underlined A nucleotides signified the modified base. Each TRD protein sequences was assigned letter and target sequence (to be modified) of each resolved HsdS (coloured boxes with e.g.: J_GG**A** - as per Cooper et al., 2017) composition in relation to the genomic location of the specificity unit coding gene (*sau1hsdS*) within the different *S. aureus* sequence backgrounds. The motifs are ordered according to similarities within TRD composition / TRS sequence to highlight like TRD segments between isolates from different ST types. **Bold**: differing ST types sharing TRS.
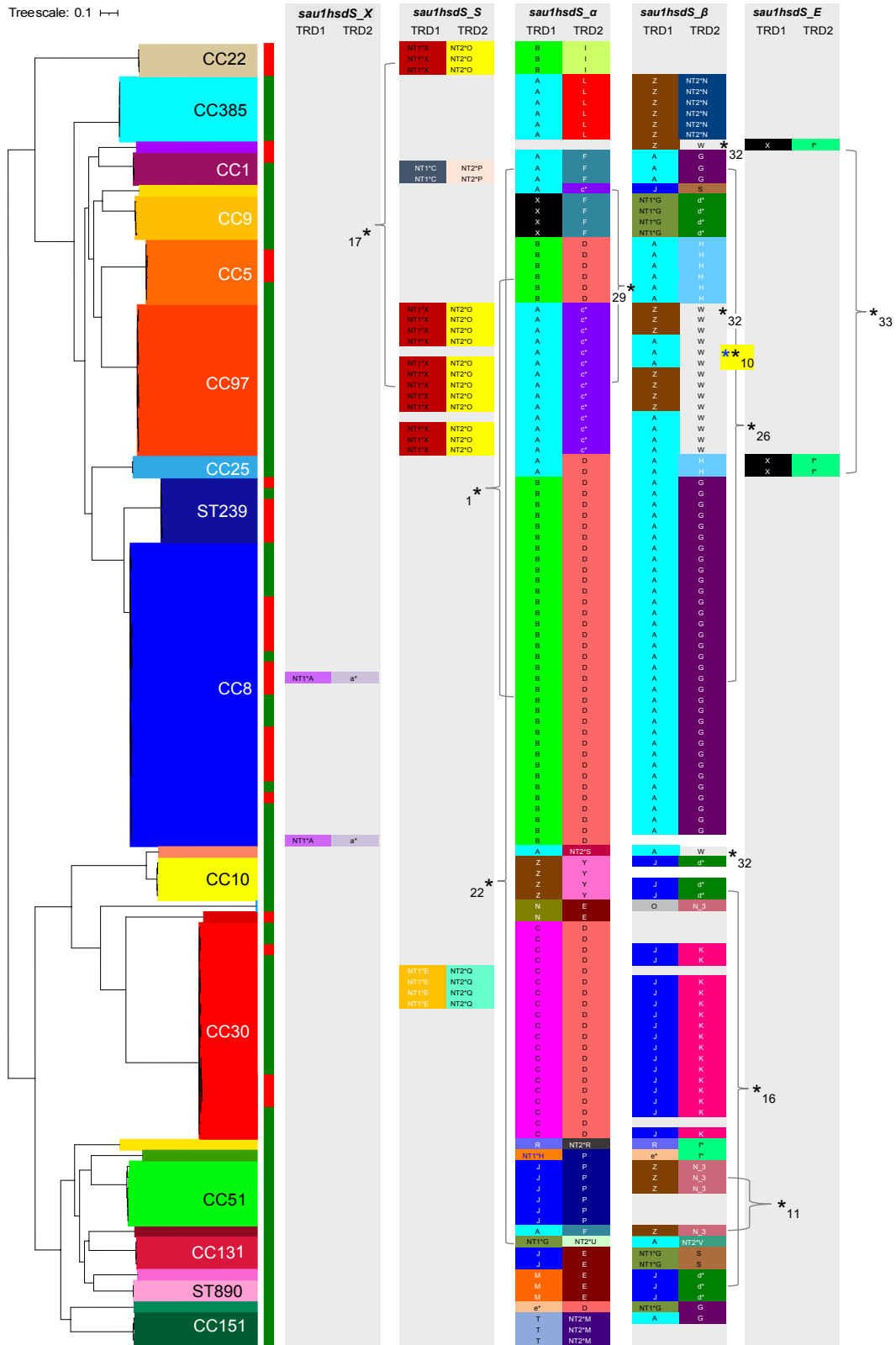
Figure 3.6 | TRD composition of 5 different Sau1 HsdS in context of ML tree for the NCTC collection
The green/red column next to phylogenetic tree indicates MSSA (green) or MRSA (red) isolates to put *sau1hsdS_S* into context. Each grey column represents a different HsdS, organised according to the genomic location of the coding *hsdS* gene → *sau1hsdS_X, sau1hsdS_S, sau1hsdS_α, sau1hsdS_β, sau1hsdS_E* (core system elements in bold). Use Table – for key regarding TRD colours (colourful blocks). * marks within lineage HsdS divergence. Annotations include given number identifier of TRS corresponding to: HsdS shared between lineages (★), HsdS dissimilarity – between lineage differentiation (★)

The corresponding TRS:TRD results revealed 5 tiered trends for Sau1 HsdS:

1) There are 40 unique HsdS (differing TRD combinations) with no duplication within differing genomic location associated with the corresponding coding *hsdS*.

2) There is evidence of **within lineage, allelic heterogeneity** of HsdS (multiple TRS and *hsdS* alleles for CC97 *sau1hsdS_β*).

3) There is evidence of **between lineage similarity** of the HsdS protein sequence (TRD1+TRD2) with same Sau1 TRS.

4) There is evidence of **between lineage similarity** of certain HsdS TRDs domains, overrepresented:
   a. within a protein domain position (TRD1 or TRD2)
   b. within a certain HsdS associated with a genomic location
   c. between HsdS associated with different genomic locations

5) There are examples of **multi-recognition TRS S1/S2 and TRD**
   a. multiple TRDs recognize the same TRS S1/S2 sequence
   b. multiple TRS S1/S2 can be recognized by the same TRD

3.5.3.2.1 40 Unique HsdS Sequences / HsdS Allelic Heterogeneity

A database of 252 HsdS protein sequences resolved into **18 different TRD1 and 28 distinct TRD2 protein domains** (Table 3.5 / Figure 3.6), in varying combinations for the 40 different methylation TRS group tags previously predicted for the collection of isolates. Five novel TRD1 and 9 novel TRD domains were characterised denoted as NT1* and NT2*, augmenting the database of described *sau1* 6mA specificities. Each of the 5 *sau1hsdS* regions constituted individual combinations of TRDs, resulting in no repetition of the same TRS signature between regions. The protein sequences identified for the N-terminus (TRD1) and the C-terminus (TRD2) HsdS protein domains also had no overlap, meaning no evidence of phase variation within the *hsdS* with no TRD2 domain switching into a TRD1 position, and *vice versa*. As we have previously characterised the number of Sau1 HsdS for each isolate, we were able to deduce that TRS #41 and #42 (Table 3.4) were attributed to BcgIAB system and were not investigated further.

A collection of 120 HsdS_α proteins from *sau1hsdMS1* in vSaα resolved in 19 specific TRS with the most variable in composition and combinations of TRDs: 12 TRD1 and 13 TRD2. From Figure 3.6 we can see that the HsdS (TRD1+TRD2) clustered within the major clonal complexes. Regarding the *sau1hsdMS2* in vSaβ, 105 HsdS_β proteins were

resolved into 16 specific TRS (TRD1 n=7, TRD2 n=10). The characterised HsdS clustered according to different lineages, with one single example of allelic variation of HsdS within CC97 (marked by ⭐ in Figure 3.6). The sequence heterogeneity within CC97 is due to variation of the TRD1 position, switching between A_CCAY and Z_GAC with a stable TRD2 (W_CRAA), potentially by mechanism of recombination at flanking DNA repeats on either end of each TRD as proposed by Furuta et al., 2011. The 'accessory' TI elements associated with the mobile genetic elements, represented 5 methylation motifs recognised by the HsdS. 21 HsdS_S proteins resolved into 3 TRS (TRD1 n=3, TRD2=3) coded for by *sau1hsdS_SCC.* Both HsdS_O and HsdS_E had one TRS match each.

3.5.3.2.2 HsdS Sequence Similarity Between Lineages

There were 12 TRS shared between lineages as pointed out in the previous section which can be visualised in the context of TRD domains of the given HsdS on an isolate basis marked by ★ in Figure 3.6 and highlighted in bold in Table 3.5. In total, 10 of these TRS were associated with the 'core' HsdS (4 TRS: HsdS_α (#1, #4, #22, #29), 6 TRS: HsdS_β (#10, #11, #16, #23, #26, #32)), and 2 TRS, were recognised by 'accessory' HsdS (HsdS_S: #17, HsdS_E: #33). The 'accessory' HsdS are shared between distantly related lineages and are easily spread through the acquisition of a mobile element as either an SCC type element (HsdS_S) or phage (HsdS_E).

No two lineages had the same set of HsdS_α and HsdS_β, there being homology between only one of the two 'core' vSa harboured *hsdS* between isolates of differing sequence type. This could indicate common ancestral genomic islands between lineages, and subsequent differentiation of the extremely stable vSa via postulated horizontal gene transfer (Baba et al., 2008) affecting the clonal expansion of *S. aureus* strains.

3.5.3.2.3 HsdS TRD Domain Sequences Similarity Between Lineages

By looking at the individual domains of each protein, TRDs (accounting for one half of a bipartite TRS) shared between HsdS belonging to different sequence backgrounds including B_AGG, D_ATC, A_CCAY, J_GGA, Z_GAC, NT1*G, E_TCAY, P_AGG among others were characterised (Table 3.5). Regarding the presences of TRD1 and TRD2 domains within different STs, there are more TRD1 domains which are shared between different sequence backgrounds than TRD2 domains. For example, 9 methylation motifs

(5 TRS: HsdS_α, 4 TRS: HsdS_β) have CC<u>A</u>Y as the forward (TRD1) target sequence, corresponding to TRD1 **A** (A_CCAY), present in isolates belonging to 12 different sequence backgrounds (CC97, CC10, CC5, CC25, CC1, CC8, ST239, ST151, CC385, ST1148, ST707, ST136). The characterised TRD1 domains are more commonly shared between ST types (a greater number of represented TRD1 than TRD2 domain, with 1.5x (TRD2 28:18 TRD1) as many types of characterised TRD2 domains throughout the library of HsdS. This may indicate that the C-terminus TRD2 domain is more readily altered via mutation or mobilised via recombination at the 3' end of the *hsdS* CDS. We can attribute increased differentiation of 6mA methylation motifs within the *S. aureus* species to introduction of variability within the TRD2 protein domain.

HsdS_α proteins (n=120) resolved into 19 specific TRS counting almost equal number TRD diversity within both positions, 12 TRD1 and 13 TRD2, illustrated in Table 3.5 and Figure 3.6. The most common TRD1 domain within the HsdS_α is A_CCAY within the TRD1 position recognising 5 TRS (#36, #29, #24, #22, #18) shared among CC25, CC385, CC97 (including ST464), ST1148, CC10, CC1, CC151 and ST136. J_GCA (#12, #6) and B_AGG (#9, #1) were also shared between two HsdS, whilst 9 other TRD1 were only represented once in a distinct lineage. There were also several TRD2, D_ATC E_TCAY, and P_AGG, which were overrepresented across lineages, recognising 4 (#1, #19, #7, #18), 3 (#4, #3, #12) and 2 (#21, #6) TRS respectively. There were 9/12 'unshared' TRD1 and 10/13 'unshared' TRD2, resulting in a higher number of possible TRD combinations and HsdS sequence differentiation.

HsdS_β proteins (n=105) resolved into 16 TRS made of a combination of 7 TRD1 and 10 TRD2 domains. The most common TRD1 domains within HsdS_β were C_CCAY recognising 4 TRS and Z_GAC, NT1*G_GAAG and J_GGA all recognising 3 TRS each, leaving only 3/7 'unshared' TRD1 domains (Table 3.5). There were 4 TRD2 domains N3_CRTA, W_CRAA, G_ACA, d*CYAA and f*GAAY which were shared between 2 TRS each, leaving only 4/10 'unshared' domains. The TRD landscape of HsdS_β seems less variable than that of HsdS_α, indicated by fewer types of characterised TRD domains and fewer unique (unshared) TRD domains in both the TRD1 (HsdS_β (3/7 TRS)**:** HsdS_α (9/12)), and TRD2 (HsdS_β (4/10 TRS)**:** HsdS_α (10/13)) position.

3.5.3.2.4 HsdS Sequences Similarity Between Genomic Locations

On a global scale, 6 TRD domains were shared between HsdS_α, HsdS_β and HsdS_E: TRD1: A_CCAY, Z_GAC, X_TCTA, e*GAG, J_GGA, R_GARA, G_GAAG and TRD2: f*GAAY as illustrated in Table 3.5. The TRD domains were conserved among different HsdS within the same domain, TRD1 or TRD2. HsdS_E was the only protein sequence which shared TRD domains from both HsdS_α and HsdS_β, X and f* respectively. Four TRD1 (A, Z, J, and e*) were shared between HsdS_α and HsdS_β. TRD1 A_CC**A**Y is shared by HsdS_α (*sau1hsdS_α)* and HsdS_β (*sau1hsdS_β)* within the same lineages (CC25, CC1, CC97, ST151), potentially giving insight to the evolution of these non-mobile genetic elements, the differentiation of *hsdS* gene within different vSa.

The sequence alignments also highlighted the dissimilarity of protein sequences and more potentially more distant relatedness of 'accessory' HsdS_X and HsdS_S and the rest of the 'core' HsdS sequences. BLAST-P analysis of the protein sequences in question showed 100-97% sequence similarity to other coagulase negative *Staphylococcus* species, likely to have been horizontally transferred.

HsdS_X – CC8 (NT1*A+A*)	99.5% SeqID	*S. epidermidis*
HsdS_S – CC30 (NT1*E+NT2*Q)	99.7% SeqID	*S. hominis*
HsdS_S – CC1 (NT1*C+NT2*P)	98.2% SeqID	*S. gallinarum*
HsdS_S – CC22/CC97 (NT1*X+NT2*O)	97.5% SeqID	*S. sciuri*

3.5.3.2.5 Multi-recognition TRS S1/S2 and TRD

Some TRS S1/S2 DNA sequence can be recognised by multiple TRDs (Table 3.6 and Figure 3.7). These TRDs usually reside within different positions (TRD1/TRD2), falling into two categories: **a)** both domains having virtually identical sequence (eg: TRS AGG recognised by TRD1 B, and TRD2 P (subject to likely recombination between target domains) or **b)** domains having much lower sequence similarity (eg: TRS GAY recognised by TRD1 U and TRD2 c* (~36% similarity)) as previously described by Cooper et al., 2017. This study highlights several other TRS which are recognised by multiple TRDs, including AGG (P, B, NT1*G – Figure 3.7 A) GAAG (NT1*G, NT2*V, NT1*X – Figure 3.7 B), GWAG (NT1*G, C – Figure 3.7 C), CAG (NT1*E, M – Figure 3.7 D) and CCAY (NT1*U, A, Figure

3.7 E). All of the listed TRD:TRS multiple recognition events fell into category **b**, with low sequence similarity between compared TRD domains, except for P and B.

Along with some TRD recognising the same TRS, multiple TRS S1/S2 were recognised by the same TRD including 4 examples: S_GCA recognising TRS S2: GCA and GMAY, A_CCAY recognising TRS S1: CCAY and CAAY, f*_GAAY recognising TRS GRAY and GAAY and NT1*G_GAAG recognising TRS S1 GWAG, AGG, and GAAG as detailed in Table 3.6. Although most of this variation is introduced with degenerate nucleotides (M: A or C, Y: C or T, R: A or G) within the predicted TRS methylation targets, TRD CCAY and NT1*G_GAAG matched predicted TRS with definite nucleotide changes: CCAY vs CAAY and GAAG vs AGG respectively. This indicates a promiscuity in the TRS recognised by some TRDS.

Table 3.6 | TRD Recognising Multiple TRS

# TRD + target	TRS (S1 + S2)	# TRS
S_GCA		
GCA	GGA (N)₇ GCA	15
GMAY	GWAG (N)₅ GMAY	25
NT1*G_GAAG		
GWAG	GWAG (N)₅ GMAY	25
GAAG	GAAG (N)₆ ACA	27
AGG	AGG (N)₅ CCAY	20
f*_GAAY		
GRAY	GARA (N)₆ GRAY	39
GAAY	GAG (N)₆ GAAY	13
A_CCAY		
CCAY	CCAY (N)₅ TACB	23
CAAY	CAAY (N)₅ GAAG	37

Both multiple TRS being matched the same TRD and multiple TRDs matching the same TRS highlight that there is still significant lack of detail when predicting TRS:TRD matches just from the amino acid sequence alignments. Proteins have complex quaternary structure and geometry, which can shift with a single amino acid change, especially within the DNA binding active site of HsdS. Therefore, a more detailed analysis of the HsdS structure and probable binding pockets may offer additional detail to the determinants of *sau1* HsdS TRS targets.

A. AAG

```
               1
A_B_AGG        MSNTQKKNVP ELRFPGFEGE WEEKQLGDLT DRVIRKNKNL ESKK-PLTIS GQLGLIDQ-- ---TEYFSKS
A_P_AGG        ---------- ----GNDYPD WEEKELGEVA DRVIRKNKNF ESKK-PLTIS GQLGLIDQ-- ---TEYFSKS
B_NT1_G_GAAG   MSNTQKKNVP ELRFPGFEDE WEEKKLGELC EFNNGINAKK EQYGMGRKFI NVLDILNNNF ITYESIIGKV

               71
A_B_AGG        VSSKNLENYT LIKNGEFAYN KSYSNGYPLG AIKRLTRYDS GVLSSLYICF SIKSEMSKDF MEAYFDSTHW
A_P_AGG        VSSKNLENYT LIKNGEFAYN KSYSNGYPLG AIKRLTRYDS GVLSSLYICF SIKSEMSKDF MEAYFDSTHW
B_NT1_G_GAAG   SVPENVEKNN KVEFGDLVFL RSSETREDVG LCNVYLD--- -KNYALYGGF IIRGKKVSDY NPIFLKEALN

               141
A_B_AGG        Y---REVSGI AVEGARNHGL LNVSVNDFFT ILIKYPSL*
A_P_AGG        Y---REVSGI AVEGARNHGL LNISVNDFFT ILIKYPSL*
B_NT1_G_GAAG   IPKKRYEIGS AAGG---STR FNVSQDILRK INVKFPPI*
```

B. GAAG

```
               1
B_NT1_G_GAAG   MSNTQKKNVP ELRFPGFEDE WEEKKLGELC EFNNGINAKK EQYGMGRKFI NVLDILNNNF ITYESIIGKV
S_NT1_X_GAAG   ---------- --------ME FETFNLTDLY TISSGLSKNR KYFGTGTPFL TFKDVFDNLI LP-NEFSGQV
B_NT2_V_GAAG   ---------- ----SEDYPN WEEKQLGEVA DRVIRKNKN- --LESKKPL- TISGQLG--L IDQTEYFSKS

               71
B_NT1_G_GAAG   SVPENVEKNN KVEFGDLVFL RSSETREDVG LCNVYLDK-N YA----LYGG FIIRGKKVS- ---DYNPIFL
S_NT1_X_GAAG   ITEEKEREKY AVKKGDLFLT RTSEKQNELG ISAVALKDYK NA----TFNG FTKRLRPNKY CENKLLPVFA
B_NT2_V_GAAG   VSSKNLGNYT LIKNGEFAYN KSYSNGYPLG AIK-RLTRYE SGLLSSLYIC FSIK------ --SEMSKDFM

               141
B_NT1_G_GAAG   KEALNIPKKR YEIGSAAGGS TRF----NVS QDILRKINVK FPPIK*---- ----------
S_NT1_X_GAAG   AFYFRSNNFR NQVNSMSIMS TRA----SLN NEMISKLKIT IPSLQNQMKI SHILLALLI*
B_NT2_V_GAAG   EAYFDSTHWY REVSGIAVEG ARNHGLLNIS VNDFFNILTK YPSL*----- ----------
```

C. GWAG

```
               1
A_C_GWAG       MSNTQTKNVP ELRFPGFEGE WEEKKVGELL EFKNGLNKGK EYFGSGSSIV NFKDVFNNRS LNTNNLTGKV
B_NT1_G_GAAG   --------VP ELRFPGFEGE WEEKKLGEFC EFNNGINAKK EQYGMGRKFI NVLDILNNNF ITYENIIGKV

               71
A_C_GWAG       NVNSKELKNY SVEKGDVFFT RTSEVIGEIG YPSVILNDPE NTVFSGFVLR GRPKSGIDLI NNNFKRYVFF
B_NT1_G_GAAG   SVPENVEKNN KVEFGDLVFL RSSETREDVG LCNVYLD-KN YALYGGFIIR GKKVSDYNPI FLK---EALN

               141
A_C_GWAG       TNSFRKEMIT KSSMTTRALT SGSAINKMKV IYPVSAKEQR KIGDFFSKLD RQIEL*
B_NT1_G_GAAG   IPKKRYEIGS AAGGSTRFNV SQDILRKINV KFPP-IEEQQ KIGDLFSKLD RQIEL*
```

D. CAG

```
               1
A_M_CAG        MSNTQTKNVP ELRFPGFEGE WEEKKLGDLG LFQKSYSFSR AKEGNGKTKH IHYGDIHS-- ----KFKTVL
S_NT*1_E       MLTRKMKDSG IKWIGEIPED WEIRKLK--- -----YTLEE RNEKNN---P IITDNILSLS VERGIFPYAE

               71
A_M_CAG        DSDGNIPNII EKAVFELIQK GDIVFADASE DYSDLGKAVM IDFKPNSLIS GLHTHLFRPL ----NNAISN
S_NT*1_E       KTGGGNKSKS DLTAYKVAHP NDIVINSMNI LAGAVG---- -----LSRYT GVVSPVYYTL YTTSEEINIT

               141
A_M_CAG        FLIFYTKTLS YKKFIRQQGT GISVLGISKK SLLNLNVLIP RSELEQQ--- ----KVGKF- FSKLDRQIEL
S_NT*1_E       YYYYLFRTRE FQRSLLGLGN GIMMRESSTG KLNTIRMRIP MDKLAGLLLP LPPRRVQDLI VKSLAKDIKV

               211
A_M_CAG        *--------- ---
S_NT*1_E       VNKLINQTEQ SI*
```

E. CCAY

```
               1
A_NT2*U_CCAY   ---------- ----GNDYSD WEEKKMGEST TMF-SGGTPQ STNTRYYKGD IPFIRSGEIS KTYTELK---
A_A_CCAY       MSNTQKKNVP ELRFPEFEGE WEEKKLGNLT TKIGSGKTPK GGSENYTNKG IPFLRSQNIR NGKLNLNDLV

               71
A_NT2*U_CCAY   -INEEALNNS SAKLVEVGDL LYALYGATSG EVAISKING- --AINQAVLC IRTNE--SVE FLLNYLFFSK
A_A_CCAY       YISKDIDDEM KNSRTYYGDV LLNITGASIG RTAINSIVET HANLNQHVCI IRLKKEYYYN FFGQYLLSRK

               141
A_NT2*U_CCAY   NK--ILNTFI QGGQGGNLSAN IIKNLIIETP TIIEQRKIAD FFCNIDNSIE IQGRKLEFLK QRKQSLLQKM
A_A_CCAY       GKRKIFLAQS GGSREGLNFK EIANLKIFTP TIF*------ ---------- ---------- ----------

               211
A_NT2*U_CCAY   FV*
A_A_CCAY       ---
```

Figure 3.7 | Multiple sequence alignment for multiple TRDs recognising same TRS
A. AGG, **B.** GAAG, **C.** GWAG, **D.** CAG, **E.** CCAY. Yellow blocks indicate sequence homology on individual amino acid level within the TRD domain region of each sequence segment visualised in SeaView.

3.5.4 *S. aureus* TI Sau1 HsdS Protein

3.5.4.1 Sau1 HsdS Protein Structure & DNA Binding Loops

To understand the specificity of TI methylation TRS to each HsdS, the structure and possible trends seen in the AA composition and length of both TRDs, the conserved alpha helices, and the predicted DNA binding loops of each HsdS protein were analysed. For each of the 40 TRS motifs uncovered with the PacBio SMRT Methylation analysis software, the amino acids for a representative HsdS protein were extracted. These were aligned with a multiple sequence aligner and models for each protein were created using a template 1YF2.1/1YF2.2 HsdS from *Methanocaldococcus jannschii* (Kim et al., 2005) within SWISS-MODEL (ExPASy Web Sever) and visualised in UCSF Chimera. The 1YF2.1A template guided models had a template-target sequence identity ranging from 16.80% - 27.78% (mean: 21.23%), 29-33% sequence identity, and 92-98% coverage. The binding domains for each TRD were found using binding predictions, the 3D model structures, the predicted binding domains for 1YF2 as described by Kim et al (2015). An example of a modelled HsdS (CC385 HsdS_α – silver) superimposed onto PDB 1YF2.1.A (navy), with annotated secondary structure and DNA binding domains is illustrated in Figure 3.8. Summary of domain lengths and annotated binding loop predictions for HsdS_α (+ HsdS_E) and HsdS_β protein sequences found in Table 3.7 and Table 3.8 respectively. The amino acid sequences of HsdS_O and HsdS_S differed from the 'core' HsdS, and although best mapped to 1YF2.1/1YF2.2, the predicated secondary structure models were of low confidence. Due to this, the secondary structure of these proteins was studied in less detail and their DNA binding loops were not predicted.

The monomeric structure of Sau1 'core' HsdS follows 5 continuous structural motifs as previously determined for HsdS: short, conserved region at the N-terminal (CR1) – TRD1 – conserved helix 1 (CR2) – TRD2 – distal conserved region containing second conserved helix structure (CR3) at C terminal as seen in Figure 3.8. Both globular TRDs (highlighted in faint red and blue) comprise of mostly beta ribbons separated by alpha helices in generally a 3xβ – α – β – α – 2xβ – α – 5xβ – 3xα – β secondary structure followed by 39AA alpha helixes within the two conserved region domains. Each TRD contains 4 DNA binding loops (TRD1: A-D; TRD2 E-H). Both Loop A and E (at the proximal end of TRD1/TRD2) are located between β4 and β5 usually as part of helical structure. Binding loops B and F reside between β9 – β10, loop C and G between β11 – β12, and D and G between α6 – α7).
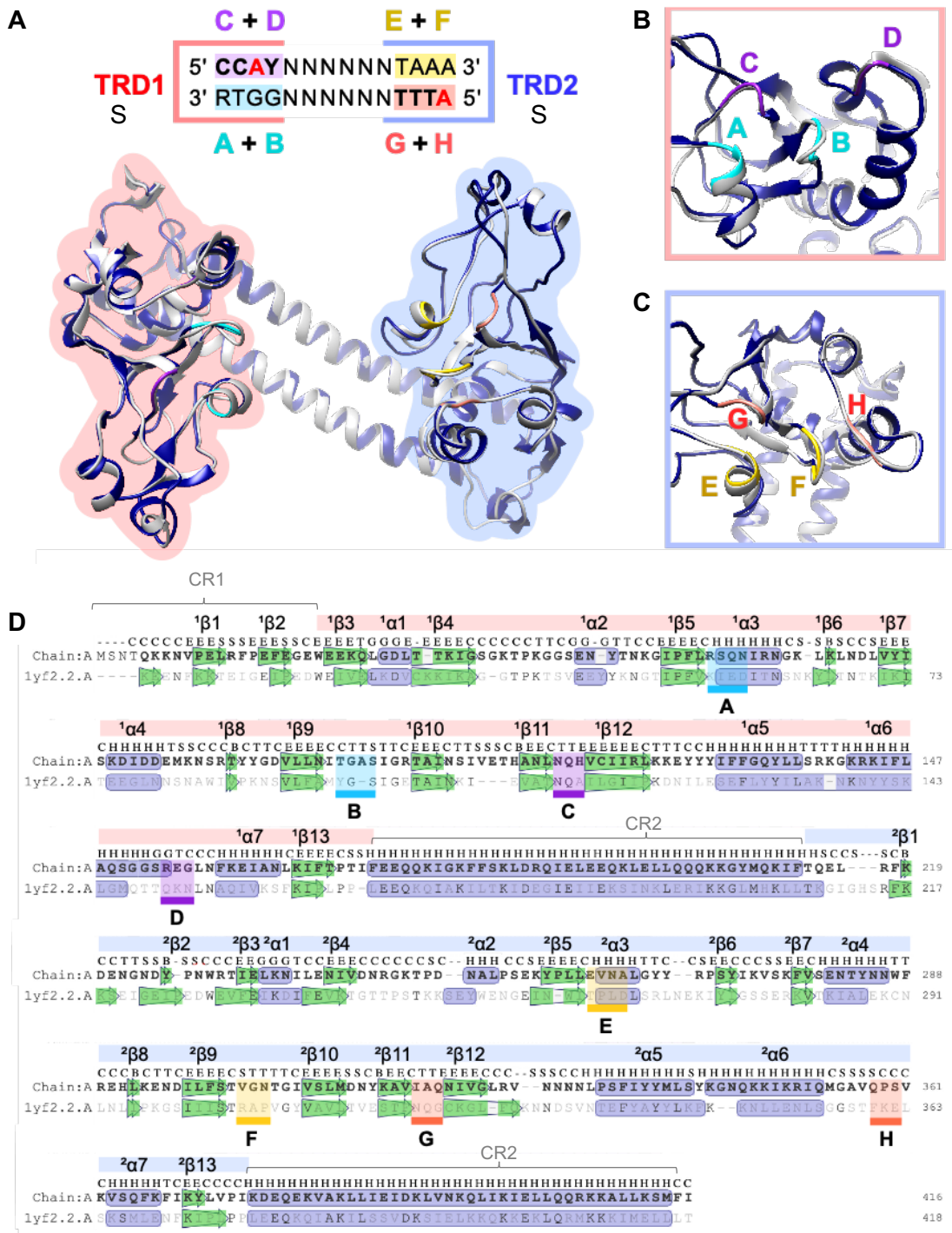
Figure 3.8 | HsdS Protein Structure Model

Template (1YF2.1A) – Target (CC385 Sau1 HsdS_ α (NCTC9611_42545_A02) Alignment

A. Sau1 HsdS_α (NCTC9611_42545_A02) protein model (grey) superimposed onto *M. jannschii* HsdS 1YF2.1A PDB crystal structure (navy). Two globular TRD domains are highlighted in red (TRD1) and blue (TRD2), with a simplified schematic of these structures binding a bipartite methylation motif. Coloured binding loops for *M. jannschii* TRD1 (aquamarine: A+B, purple: C+D) and TRD2 (gold: E+F, salmon: G+H), magnified in **B** and **C**. **D.** Amino acid sequence alignment of template (*M. jannschii*) and target (*S. aureus*) HsdS with secondary structure annotated with beta sheets (green arrows) and alpha helices (periwinkle). Amino acids in annotated *M. jannschii* binding pockets (aquamarine: A+B, purple: C+D, gold: E+F, salmon: G+H) were used to predict *S. aureus* HsdS binding loops corresponding to the nucleotide string within each modified target sequence (S1/S2) showing in A.

117

Table 3.7 | HsdS Domain Lengths (Target Recognition Domains (TRD), Conserved Regions (CR)

ST	Representative Strain CDS	TRS #	TRS Motif (5' / 3')	S1	N	S2	CR1 AA	TRD1 AA	CR2 AA	TRD2 AA	CR3 AA	TOTAL AA
HsdS_α												
CC30	NCTC7361_40853_H01_00504	19	GWAGNNNNNNGAT/ATCNNNNNNCTWC	4	5	3	20	157	52	142	39	410
ST350	NCTC5663_41004_B01_00437	7	GAGNNNNNNNGAT/ATCNNNNNNNCTC	3	6	3	20	141	52	142	39	394
ST1021	NCTC6135_41004_C01_00438	38	GARANNNNNNNYTCC/GGARNNNNNNNTYTC	4	6	4	20	143	52	125	39	379
ST198, ST398	NCTC8726_41315_B01_00373	4	ACCNNNNNRTGA/TCAYNNNNNGGT	3	5	4	20	149	52	132	39	392
ST1254	NCTC10649_43941_A01_00400	21	GAAANNNNNCCT/AGGNNNNNTTTC	4	5	3	20	151	52	152	39	414
ST1148, CC97	NCTC13137_40657_E01_00372	29	CCAYNNNNNNNRTC/GAYNNNNNNNRTGG	4	6	3	20	154	52	147	39	412
CC385	NCTC9611_42545_A02_00455	36	CCAYNNNNNNNTAAA/TTTANNNNNNNRTGG	4	6	4	20	154	52	151	39	416
CC890	NCTC6966_40677_A02_00381	12	GGANNNNNNNRTGA/TCAYNNNNNNNTCC	3	6	4	20	123	52	132	39	366
CC8, CC5	NCTC13394_40677_F01_00430	1	AGGNNNNNGAT/ATCNNNNNCCT	3	5	3	20	150	52	142	39	403
CC25	NCTC8317_33763_A01_00390	18	CCAYNNNNNGAT/ATCNNNNNRTGG	4	5	3	20	154	52	142	39	407
CC51	NCTC13298_40677_C01_00412	6	GGANNNNNNNCCT/AGGNNNNNNNTCC	3	6	3	20	123	52	152	39	386
CC9	NCTC8765_40853_G01_00357	30	TCTANNNNNNNTTAA/TTAANNNNNNNTAGA	4	6	4	20	143	52	154	39	408
CC22	NCTC12035_35910_F02_0420	9	AGGNNNNNNNTGAR/YTCANNNNNNNCCT	3	6	4	20	150	52	158	39	419
CC10	NCTC7972_40574_G02_01518	5	GACNNNNNNNTAG/CTANNNNNNNGTC	3	6	4	20	142	52	147	39	400
ST707	NCTC10399_42552_B01_00637	20	AGGNNNNNNRTGG/CCAYNNNNNCCT	3	5	4	20	152	52	143	39	406
CC1, CC80	NCTC13297_40677_B01_00391	22	CCAYNNNNNTTAA/TTAANNNNNRTGG	4	5	4	20	154	52	154	39	419
CC151	NCTC7485_40853_C01_00396	34	CAAGNNNNNNTARC/GYTANNNNNNCTTG	4	6	4	20	150	52	148	39	409
CC131	NCTC7988_40798_F01_448_00448	3	CAGNNNNNRTGA/TCAYNNNNNCTG	3	5	4	20	154	52	132	39	397
ST498	NCTC6137_40853_A01_00348	24	CCAYNNNNNRTTT/AAAYNNNNNRTGG	4	5	4	20	154	52	135	39	400
HsdS_β												
ST350	NCTC5663_41004_B01_01824	27	GAAGNNNNNNNTGT/ACANNNNNNNCTTC	4	6	3	11	144	51	135	39	389
CC97, ST498	NCTC13841_50450_A01_01749	32	CCAYNNNNNNNTTYG/CRAANNNNNNNRTGG	4	6	4	20	154	51	150	39	414
ST1254	NCTC10649_43941_A01_01828	13	GAGNNNNNNRTTC/GAAYNNNNNNCTC	3	6	4	20	141	51	153	39	404
CC51	NCTC13298_40677_C01_01876	11	GACNNNNNNTAYG/CRTANNNNNNGTC	3	6	4	20	142	51	144	39	396
ST1021	NCTC6135_41004_C01_01910	39	GARANNNNNNRTYC/GRAYNNNNNNNTYTC	4	6	4	20	143	51	153	39	406
CC10, CC131	NCTC7972_40574_G02_00112	16	GGANNNNNNNTTRG/CYAANNNNNNNTCC	3	7	4	20	123	51	159	39	392
CC385	NCTC9611_42545_A02_01834	8	GACNNNNNNTGG/CCANNNNNNGTC	3	6	3	20	142	51	150	39	402

Clonal Complex	CDS ID	TRS	Sequence	S1	N	S2		TRD1		TRD2		Total
CC97, CC80	NCTC9551_41255_G01_01804	10	GACNNNNNNNTTYG/CRAANNNNNNNGTC	3	6	4	20	142	51	150	39	402
CC30	NCTC7361_40853_H01_01828	14	GGANNNNNNNTCG/CGANNNNNNNNTCC	3	7	3	20	123	51	151	39	384
CC1, CC8	NCTC13297_40677_B01_01762	26	CCAYNNNNNNNTGT/ACANNNNNNNRTGG	4	6	3	20	154	51	135	39	399
CC9	NCTC8765_40853_G01_01775	31	GAAGNNNNNNNTTRG/CYAANNNNNNNCTTC	4	6	4	20	152	51	159	39	421
CC890	NCTC6966_40677_A02_01877	25	GWAGNNNNNRTKC/GMAYNNNNNCTWC	4	5	4	20	152	51	139	39	401
ST707	NCTC10399_42552_B01_02009	37	CAAYNNNNNCTTC/GAAGNNNNNRTTG	4	5	4	20	154	51	145	39	409
ST1148	NCTC13137_40657_E01_01721	15	GGANNNNNNNTGC/GCANNNNNNNNTCC	3	7	3	20	123	51	139	39	372
ST198	NCTC8726_41315_B01_01730	35	CAACNNNNNNNTAYG/CRTANNNNNNNGTTG	4	6	4	20	146	51	144	39	400
CC5, CC25	NCTC7415_41004_F01_01825	23	CCAYNNNNNVGTA/TACBNNNNNRTGG	4	5	4	20	154	51	145	39	409
HsdS_X				S1	N	S2						
CC8	NCTC13394_40677_F01_00062	28	TAAGNNNNNNNTTC/GAANNNNNNNCTTA	4	6	3	20	142	40	158	39	400
HsdS_S				S1	N	S2						
CC1	NCTC13297_40677_B01_00031	40	GNNGANNNNNNNRTTA/TAAYNNNNNNNTCNNC	5	7	4	20	146	45	176	48	436
CC22, CC97	NCTC13552_40415_D02_00067	17	GAAGNNNNNTAC/GTANNNNNCTTC	4	5	3	20	137	52	142	59	410
CC30	NCTC7361_40853_H01_00847	2	CAGNNNNNRAAT/ATTYNNNNNCTG	3	5	4	20	154	45	159	48	445
HsdS_E				S1	N	S2						
CC1, CC25	NCTC8317_33763_A01_02151	33	TCTANNNNNNNRTTC/GAAYNNNNNNNTAGA	4	6	4	20	154	52	135	39	406

Legend: Representative HsdS (with CDS ID) recognising 40 different TRS. The length of the bipartite sequences, S1 and S2, which TRD1 and TRD2 bind are detailed as well as the spacer region (N). These were correlated to the amino acid length of TRD1 and TRD2 and conserved regions (containing alpha helix).

Table 3.8 | HsdS TRD Binding Loops - Amino Acid Sequence Strings and Position (TRD1:A-D=S1, TRD2:E:G=S2)

HsdS_α

CC30 NCTC7361_40853_H01_00504

GWAGNNNNNNGAT/ATCNNNNNCTWC

Color	Seq	BL	Seq	AA	BL	Seq	AA
C	CTWC / GWAG	A	NFKD	51	B	RTSE	92
		C	TVF	112	D	TRAL	156
D	GAT / ATC	E	TINS	257	F	MRMW	294
		G	SPA	311	H	SDTW	350

ST350 NCTC5663_41004_B01_00437

GAGNNNNNNNGAT/ATCNNNNNNCTC

Color	Seq	BL	Seq	AA	BL	Seq	AA
e*	CTC / GAG	A	KKLT	44	B	NCAF	87
		C	PSF	107	D	AKR	142
D	GAT / ATC	E	TINS	241	F	MRMW	278
		G	SPA	295	H	SDTW	334

ST1021 NCTC6135_41004_C01_00438

GARANNNNNNNYTCC/GGARNNNNNNNTYTC

Color	Seq	BL	Seq	AA	BL	Seq	AA
R	TYTC / GARA	A	SVEN	51	B	IGD	87
		C	YVS	105	D	PKK	144
NT2*R	YTCC / GGAR	E	GTGG	247	F	KGT	270
		G	DTL	287	H	VPS	321

ST198, ST398 NCTC8726_41315_B01_00373

ACCNNNNNNRTGA/TCAYNNNNNNGGT

Color	Seq	BL	Seq	AA	BL	Seq	AA
N	GGT / ACC	A	SQKD	56	B	NYAP	90
		C	SPL	109	D	DRFS	148
E	RTGA / TCAY	E	VRSP	255	F	DGVG	278
		G	HQR	296	H	VDS	334

HsdS_β

ST350 NCTC5663_41004_B01_01824

GAAGNNNNNNNTGT/ACANNNNNNNCTTC

Color	Seq	BL	Seq	AA	BL	Seq	AA
NT1*G	CTTC / GAAG	A	NVLD	43	B	RSSE	84
		C	ALY	103	D	TRFN	145
G	TGT / ACA	E	GATG	252	F	DGAK	276
		G	NNH	297	H	PAK	331

CC97, ST498 NCTC13841_50450_A01_01749

CCAYNNNNNNNTTYG/CRAANNNNNNNRTGG

Color	Seq	BL	Seq	AA	BL	Seq	AA
A	RTGG / CCAY	A	RSQN	55	B	TGAS	95
		C	NQH	115	D	REG	154
W	TTYG / CRAA	E	RISD	265	F	ASS	304
		G	HDGF	321	H	QVN	354

ST1254 NCTC10649_43941_A01_01828

GAGNNNNNNNRTTC/GAAYNNNNNNNCTC

Color	Seq	BL	Seq	AA	BL	Seq	AA
e*	CTC / GAG	A	KKLT	58	B	LNCA	86
		C	SPS	106	D	AKR	142
f*	TTC / GAAY	E	RITD	247	F	TGA	285
		G	AGF	307	H	QPG	346

CC51 NCTC13298_40677_C01_01876

GACNNNNNNNTAYG/CRTANNNNNNNGTC

Color	Seq	BL	Seq	AA	BL	Seq	AA
Z	GTC / GAC	A	IELD	47	B	LRPY	86
		C	CSS	101	D	MPR	143
NOVEL	TAYG / CRTA	E	KIGN	249	F	SGS	286
		G	QDSN	304	H	IKR	338

ST1254 — NCTC10649_43941_A01_00400

GA**A**ANNNNNCCT/A**G**GNNNNNTTTC

NT1*H	TTTC
P	GAAA
	CCT
	AGG

BL	Seq	AA	BL	Seq	AA
A	NLND	51	B	SVKP	92
C	VVF	111	D	NTN	152
E	SGQL	252	F	KSYS	291
G	YSSL	313	H	RNHG	352

ST1021 — NCTC6135_41004_C01_01910

GAR**A**NNNNNNNRTYC/GR**A**YNNNNNNNTYTC

R	TYC
f*	GARA
	TYTC
	GAAY

BL	Seq	AA	BL	Seq	AA
A	SVEN	51	B	IGD	87
C	YVS	105	D	PKK	144
E	RITD	249	F	TGA	287
G	AGF	309	H	QPG	348

ST1148, CC97 — NCTC13137_40657_E01_00372

CC**A**YNNNNNNNRTC/G**A**YNNNNNNNRTGG

A	RTGG
c*	CCAY
	RTC
	GAY

BL	Seq	AA	BL	Seq	AA
A	RSQN	55	B	TGAS	95
C	NQH	115	D	REG	154
E	HCKG	261	F	QNF	294
G	SSD	314	H	SKR	354

CC10, CC131 — NCTC7972_40574_G02_00112

GG**A**NNNNNNNTTRG/CYA**A**NNNNNNNTCC

J	TCC
d*	GGA
	TTRG
	CYAA

BL	Seq	AA	BL	Seq	AA
A	GTGG	50	B	AVGI	69
C	TVDT	87	D	VPS	124
E	QRNA	238	F	SGT	275
G	NQA	294	H	ITN	333

CC385 — NCTC9611_42545_A02_00455

CC**A**YNNNNNNTAAA/TTT**A**NNNNNNNRTGG

A	RTGG
L	CCAY
	RTGG
	TTTA

BL	Seq	AA	BL	Seq	AA
A	RSQN	56	B	TGAS	95
C	NQH	115	D	REG	154
E	EVNA	261	F	VGN	302
G	IAQ	318	H	QPS	358

CC385 — NCTC9611_42545_A02_01834

GA**C**NNNNNNNTGG/CC**A**NNNNNNNGTC

Z	GTC
NT2*N	GAC
	TGG
	CCA

BL	Seq	AA	BL	Seq	AA
A	IELD	51	B	LRPY	86
C	CSS	101	D	MPR	143
E	SPIE	250	F	RAG	290
G	QGF	307	H	TFS	343

CC890 — NCTC6966_40677_A02_00381

GG**A**NNNNNNNRTGA/TC**A**YNNNNNNNTCC

J	TCC
E	GGA
	RTGA
	TCAY

BL	Seq	AA	BL	Seq	AA
A	GTGG	50	B	VGIG	69
C	TVDT	87	D	VPS	124
E	VRSP	229	F	DGVG	252
G	HQR	270	H	VDS	308

CC97, CC80 — NCTC9551_41255_G01_01804

GA**C**NNNNNNNTTYG/CRA**A**NNNNNNNGTC

Z	GTC
W	GAC
	TTYG
	CRAA

BL	Seq	AA	BL	Seq	AA
A	IELD	51	B	LRPY	86
C	CSS	101	D	MPR	143
E	RISD	253	F	AAS	292
G	HDGF	308	H	QVN	344

CC8, CC5   NCTC13394_40677_F01_00430

AGGNNNNNNGAT/ATCNNNNNCCT

B — CCT / AGG
D — GAT / ATC

BL	Seq	AA	BL	Seq	AA
A	SGQL	48	B	KSYS	88
C	YSSL	108	D	RNHG	148
E	TINS	250	F	MRMW	287
G	SPA	304	H	SDTW	344

CC30   NCTC7361_40853_H01_01828

GGANNNNNNNTCG/CGANNNNNNNNTCC

J — TCG / GGA
K — TCC / CGA

BL	Seq	AA	BL	Seq	AA
A	GTGG	50	B	AVGI	69
C	TVDT	87	D	VPS	124
E	QSSD	232	F	RVG	273
G	QDF	290	H	IKG	325

CC25   NCTC8317_33763_A01_00390

CCAYNNNNNNGAT/ATCNNNNNRTGG

A — RTGG / CCAY
D — GAT / ATC

BL	Seq	AA	BL	Seq	AA
A	RSQN	55	B	TGAS	95
C	NQH	115	D	REG	154
E	TINS	245	F	MRMW	291
G	SPA	308	H	SDTW	348

CC1, CC8   NCTC13297_40677_B01_01762

CCAYNNNNNNTGT/ACANNNNNNRTGG

A — RTGG / CCAY
G — TGT / ACA

BL	Seq	AA	BL	Seq	AA
A	RSQN	55	B	TGAS	103
C	NQH	115	D	REG	154
E	GATG	262	F	DGAK	286
G	NNH	307	H	PAK	341

CC51   NCTC13298_40677_C01_00412

GGANNNNNNNCCT/AGGNNNNNNNTCC

J — TCC / GGA
P — CCT / AGG

BL	Seq	AA	BL	Seq	AA
A	GTGG	50	B	AVGI	67
C	TVDT	87	D	VPS	124
E	SGQL	226	F	KSYS	263
G	YSSL	296	H	RNHG	328

CC9   NCTC8765_40853_G01_01775

GAAGNNNNNNNTTRG/CYAANNNNNNNCTTC

NT1*G — CTTC / GAAG
d* — TTRG / CYAA

BL	Seq	AA	BL	Seq	AA
A	NVLD	51	B	RSSE	92
C	ALY	111	D	TRFN	153
E	QRNA	267	F	SGT	304
G	NQA	323	H	ITN	362

CC9   NCTC8765_40853_G01_00357

TCTANNNNNNNTTAA/TTAANNNNNNNTAGA

X — TAGA / TCTA
F — TTAA / TTAA

BL	Seq	AA	BL	Seq	AA
A	ITNE	55	B	EDIL	82
C	NVF	98	D	TIP	143
E	SIGS	249	F	DKTK	288
G	YNQR	310	H	QVY	350

CC890   NCTC6966_40677_A02_01877

GWAGNNNNNRTKC/GMAYNNNNNCTWC

NT1*G — CTWC / GAAG
S — RTKC / GCA

BL	Seq	AA	BL	Seq	AA
A	NVLD	51	B	RSSE	92
C	ALY	111	D	TRFN	153
E	QGNA	260	F	RAP	293
G	NAC	309	H	FES	343

CC22 — NCTC12035_35910_F02_0420

AAGNNNNNNNTGAR/YTCANNNNNNNCCT

		BL	Seq	AA	BL	Seq	AA
B	CCT / AGG	A	SGQL	48	B	KSYS	88
		C	SSLY	108	D	RNHG	148
I	TGAR / YTCA	E	ALSA	259	F	EAP	303
		G	SQR	321	H	AKG	360

ST707 — NCTC10399_42552_B01_02009

CAAYNNNNNNCTTC/GAAGNNNNNNRTTG

		BL	Seq	AA	BL	Seq	AA
A	RTGG / CCAY	A	RSQN	47	B	TGAS	87
		C	NQH	107	D	REG	146
NT2*V	CTTC / GAAG	E	ISGL	250	F	KSYS	284
		G	SSLY	308	H	RNHG	346

CC10 — NCTC7972_40574_G02_01518

GACNNNNNNNTAG/CTANNNNNNNGTC

		BL	Seq	AA	BL	Seq	AA
Z	GTC / GAC	A	IELD	47	B	LRPY	86
		C	CSS	102	D	MPR	143
Y	TAG / CTA	E	TPTD	251	F	CIAS	288
		G	NQQ	305	H	QIV	343

ST1148 — NCTC13137_40657_E01_01721

GGANNNNNNNTGC/GCANNNNNNNTCC

		BL	Seq	AA	BL	Seq	AA
J	TGC / GGA	A	GTGG	50	B	AVGI	69
		C	TVDT	87	D	VPS	124
S	TCC / GCA	E	QGNA	231	F	RAP	264
		G	NAC	280	H	FES	314

ST707 — NCTC10399_42552_B01_00637

AGGNNNNNRTGG/CCAYNNNNNCCT

		BL	Seq	AA	BL	Seq	AA
NT1*G	CCT / AGG	A	NVLD	51	B	RSSE	92
		C	YALY	111	D	TRFN	153
NT2*U	RTGG / CCAY	E	RSGE	260	F	YGAT	296
		G	NQA	313	H	QGN	348

ST198 — NCTC8726_41315_B01_01730

CAACNNNNNNNTAYG/CRTANNNNNNNGTTG

		BL	Seq	AA	BL	Seq	AA
O	GTTG / CAAC	A	TTGD	54	B	GQGK	94
		C	NQA	112	D	QKN	147
NOVEL	TAYG / CRTA	E	KIGN	253	F	SGS	290
		G	QDSN	307	H	IKR	342

CC1, CC80 — NCTC13297_40677_B01_00391

CCAYNNNNNTTAA/TTAANNNNNRTGG

		BL	Seq	AA	BL	Seq	AA
A	RTGG / CCAY	A	RSQN	55	B	TGAS	95
		C	NQH	115	D	REG	154
F	TTAA / TTAA	E	SIGS	260	F	DKTK	299
		G	YNQR	321	H	QVY	361

CC5, CC25 — NCTC7415_41004_F01_01825

CCAYNNNNNVGTA/TACBNNNNNRTGG

		BL	Seq	AA	BL	Seq	AA
A	RTGG / CCAY	A	RSQN	55	B	TGAS	95
		C	NQH	115	D	REG	154
H	GTA / TACB	E	SRQG	257	F	RSD	294
		G	SKY	314	H	QLV	351

CC151 NCTC7485_40853_C01_00396

CAAGNNNNNNNTARC/GYTANNNNNNNCTTG

		BL	Seq	AA	BL	Seq	AA
T	CTTG CAAG	A	TTDL	55	B	GFNQ	96
		C	NQA	113	D	DPN	151
NT2*M	TARC GYTA	E	SKNL	256	F	IGT	296
		G	KNV	313	H	QKF	351

CC131 NCTC7988_40798_F01_448_00448

CAGNNNNNRTGA/TCAYNNNNNCTG

		BL	Seq	AA	BL	Seq	AA
M	CTG CAG	A	HYGD	52	B	ASED	92
		C	SGLH	114	D	VLG	154
E	RTGA TCAY	E	VRSP	261	F	GVGV	283
		G	HQR	301	H	VDS	339

ST498 NCTC6137_40853_A01_00348

CCAYNNNNNRTTT/AAAYNNNNNRTGG

		BL	Seq	AA	BL	Seq	AA
A	RTGG CCAY	A	RSQN	55	B	TGAS	95
		C	NQH	115	D	REG	154
NT2*S	RTTT AAAY	E	SSQT	265	F	VSYR	298
		G	NVC	309	H	NPK	342

HsdS_E

CC1, CC25 NCTC8317_33763_A01_02151

TCTANNNNNNNRTTC/GAAYNNNNNNNTAGA

		BL	Seq	AA	BL	Seq	AA
X	TAGA TCTA	A	ITNE	55	B	EDIL	82
		C	NVF	88	D	TIP	143
f*	TTC GAAY	E	RITD	249	F	TGA	287
		G	AGF	309	H	QPG	348

Overall, the average number of amino acids in the conserved regions stayed constant between HsdS_α, HsdS β and HsdS_E (CR1 – 20 AA, CR2: 51/52 AA, CR3: 39 AA). There was only one homologue of HsdS_X, whist 3 differing homologues of HsdS_S each variable in domain sizes as seen in Table 3.7. Two of HsdS_S proteins showed comparable protein conservation within the outlined conserved regions (CR1:20AA, CR2:45AA, CR2:48AA), with variable TRD lengths. These proteins were on average 25 amino acids longer than the core HsdS_α (+ HsdS_E) and HsdS_β. There was no correlation between lengths of conserved regions or TRDs in relation to the length of spacer (N) nucleotides within a TRS, hence this is solely affected by the structural conformation and angles at which the DNA sequence is bound within each active site.

Figure 3.9 summaries the differences between the average amino acid lengths for each characterised protein domain between the two 'core' HsdS (and HsdS_E homogenous to the HsdS_α structure). There were slight variations between the TRD lengths and spacing between the DNA binding loops were noted for the specificity units of HsdS_α and HsdS_β. Firstly, the lengths of TRD1 and TRD2. The average length of Sau1 HsdS_α was 148 AAs whilst that of HsdS_β was slightly smaller in the average and median calculations at 144 AAs. TRD2 for HsdS_α was 144 AA in length whilst HsdS_β was 147 AAs. The distances between the 4 binding sites were also slightly variable for both TRDs which is greatly dependent on the number of AAs within each binding loop (BL). On average the predicted active binding sequences varied from 3-5 AAs but most being 4 AAs in length. The amino acid composition for all collated binding loop AA sequences was investigated using ExPASy ProtParam tool. Of the 1007 amino acids present, the composition showed: 12.6% serine (S), 10.8% glycine (G), 9.0% asparagine (N), 7.2% threonine (T), 6.7% alanine (A), 6.4% arginine (R) and glycine (Q), 5.5% valine (V) and aspartic acid. The binding loops for HsdS_α and HsdS_β followed the same AA composition: S: 14.3% and 12%, G: 10.2% and 12%, T: 7.2% and 7.0% respectively. There was no correlation between certain amino acids preferentially binding a given DNA nucleotide.

The binding loops for each methylation motifs were unique (Table 3.8), with exception of a few TRDs previously mentioned, B and P recognising TRS AGG. There was some conservation of amino acid strings within binding loops for several TRS including similar nucleotide composition:

TRS: CAAG / CAAC / CCAY / CYAA   Binding Loop C/G: NQA
TRS: CAAC / CCAY   Binding Loop D/H: QKN / QG
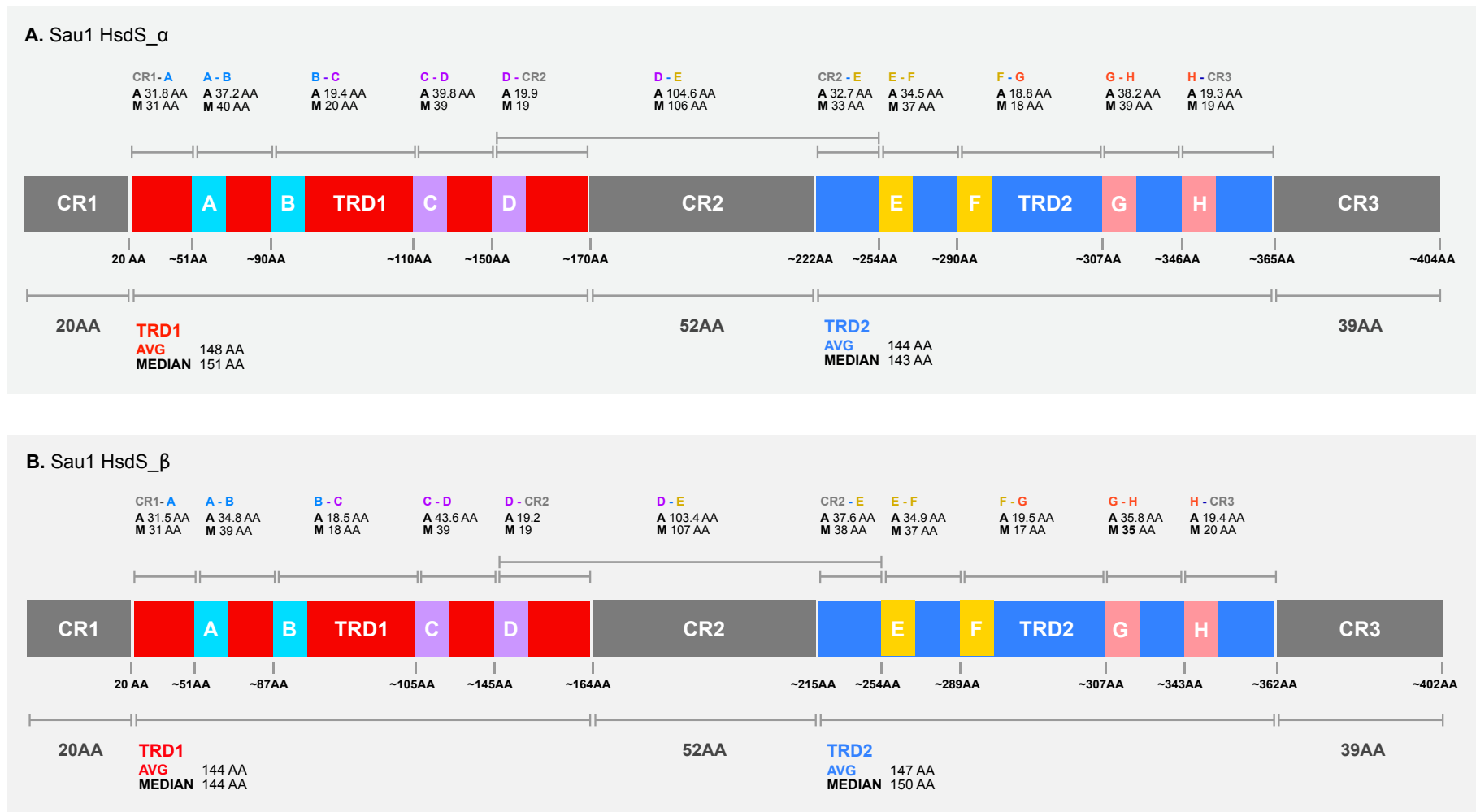
Figure 3.9 | HsdS_α and HsdS_β Protein Domain Map
A. Sau1 HsdS_α and **B.** HsdS_β protein domains including conserved regions (CR: dark grey), TRD1 (red) and binding loops A+B (aquamarine) and C+D (periwinkle), and TRD2 (blue) with binding loop E+F(yellow) and G+H (salmon).

In summary, PacBio Restriction and Modification analysis coupled with in-depth protein sequence analysis of HsdS has allowed the assignment of each predicted 6mA methylation target recognition sequence to an individual specificity protein. This was important as it allowed the characterisation of the variability of *S. aureus* Sau1 HsdS on a global scale, as well as within the collection of proteins associated to the different genomic locations of the coding sequences: HsdS_S_α, HsdS_β, HsdS_S, HsdS_X and HsdS_E. Categorisation of the PacBio predicted 6mA motifs allowed the investigation of the *S. aureus* 6mA methylation landscape next, focusing on the frequency and distribution of detected TRS motifs throughout the genome, recognised by the 5 different HsdS units.

3.5.5 TI Sau1 6mA Methylation Landscape – NCTC Collection

PacBio SMRT Sequencing technology has allowed not just the prediction of Sau1 6mA methylation motifs, but also gives positional information of each 6mA through direct detection of kinetic changes generate by modifications in the DNA sequence. This allows the precise analysis of the distribution of predicted and detected modified adenine bases throughout the genome. Using this combined approach, the 6mA methylome within each of the 120 NCTC *S. aureus* genome sequences was characterised, investigating the variability, the distribution and frequency of TRS within different regions of the genome, facilitated by either of the two 'core' HsdS (HsdS_α, HsdS_β) and 'accessory' HsdS_S, HsdS_X and HsdS_E matched to their corresponding TRS in the previous section.

3.5.5.1 Whole Genome TRS Distribution - NCTC Collection

The absolute number of detected TRS depends on the length of a whole genome sequence, which vary isolate by isolate due to the flexible accessory genome content (variable number and lengths of mobile genetic elements). Specific motif patterns were searched to locate each TRS motif reported by PacBio SMRT Modification and Motif Analysis, and cross reference the number of TRS matches between both approaches. The average number of detected TRS (methylated (mean: 97.3%, median: 98.2%) and unmethylated (mean: 2.7%, median: 1.8%) varied for each of the 40 TRS detected by PacBio SMRT Analysis as seen in Figure 3.10. These were categorised according to genomic location of the coding CDS, as one of the main questions being investigated was the potential functional differences in methylation frequency associated with HsdS_α (red) or HsdS_β (green) in particular. HsdS_S, HsdS_X and HsdS_E, were grouped together (purple), to represent the TRS matches for the 'accessory' HsdS, present in only 1/5 of the NCTC isolates.
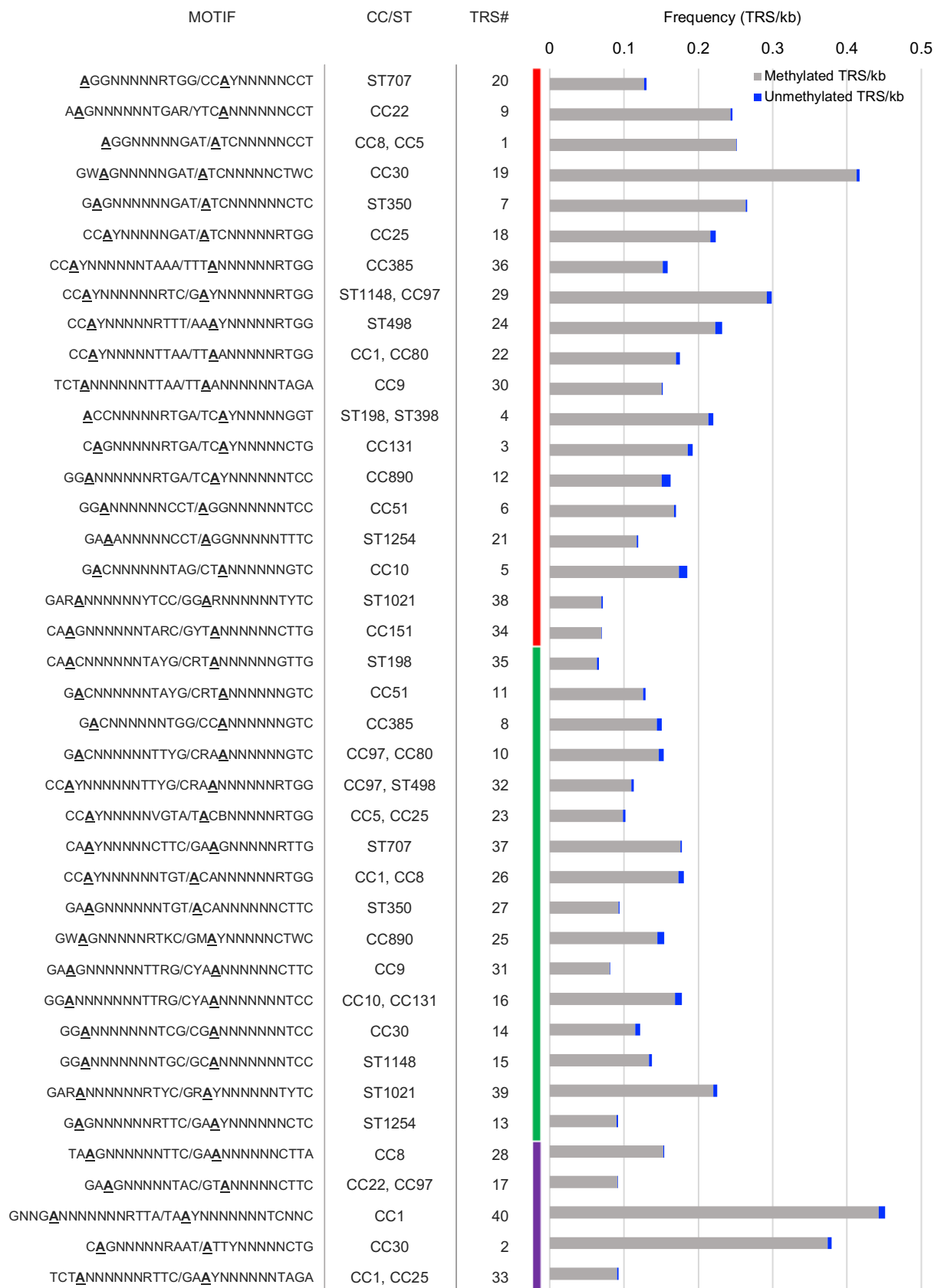
Figure 3.10 | Average number of TRS within NCTC *S. aureus* genomes
Frequency of TRS detected with modified adenine bases (grey) and unmodified adenine bases (blue) within a given TRS. 40 Sau1 TRS motifs resulting from PabBio Modification and Motif Analysis. Motifs were matched to varying HsdS: **red**: HsdS_α, **green**: HsdS_β, **purple**: *OTHER (*HsdS_S, HsdS_X and HsdS_E**).

The average frequency of predicted TRS within lineages showed that motifs recognised by HsdS_α (mean: 548 TRS, 0.197 TRS/kb) were on average 31.6% greater than those recognised by HsdS_β (mean 371 TRS, 0.135 TRS/kb). The accessory HsdS (658 TRS) had 18.8% and 73.6% higher TRS densities than HsdS_α and HsdS_β respectively. Although the relative frequency of TRS in a 10,000 bp region gives a good quantitative overview of the methylation potential of each TRS, it does not give any information on their distribution throughout the genome. To investigate the potential of hypermethylated areas of the genome, and difference in TRS match scattering per system, each motif was visualised in Artemis genome browser. Subsequently, frequency plots were created with positional information of TRS within 16 isolates from different sequence backgrounds. As all plots were very similar, one isolate map with 3 and one with 2 active HsdS are presented in Figure 3.11 (TRS/10 kb). No hypermethylated areas were detected, and the overall distribution of TRS matches is evenly spread, for the total TRS and TRS recognised by a single HsdS.

As each isolate's genome varied in size due to variable MGE composition (detailed later), the TRS matches for 6mA methylation motifs found within individual isolates were normalised according to genome size, expressed as a frequency (TRS/kb). The frequency of 6mA TRS motifs within a single isolate depends on the specific TRS of each HsdS (as represented in Figure 3.12), the number of functional specificity proteins present in the genome and the sequence type of the isolates. Isolates with 3 methylation motifs generally showed higher total TRS frequency than isolates which only have 1 methylated motif (dependent on frequency distinct recognition pattern of the given TRS). 72% (86/120) of the isolates had methylation signatures for two different 6mA motifs, 20% (25/120) having matches for 3 motifs, and 9 isolates (8%) only had one motif for 6mA modification (Figure 3.12 A). The total TRS frequency of each isolate was further broken down into the match frequencies for the individual HsdS which the isolates possess: HsdS_α, HsdS_β, and/or accessory HsdS (Figure 3.12 B). This was to investigate if there were any global trends or biases in methylation frequency by either of the two core systems, and to also examine whether the distribution of TRS/kb matches for HsdS_α and HsdS_β remained preserved within a lineage. From Figure 3.12 B, we can clearly see that in all but 5 isolates (CC80 n=1, CC131 n=3, ST1021 n=1), have equal or higher TRS frequency recognised by HsdS_α ((A) red) than HsdS_β ((B) green). As each lineage had differential TRS frequency, the TRS signatures of isolates within 6 of the main represented CCs (CC8, CC30, CC5, CC9, CC97, and CC385) were studied in detail, to exemplify differential methylation by the **two core HsdS** throughout *S. aureus* as seen in Figure 3.13.

A

Numb[...]

B

10,000 [...]regions[...]

16
14
12
10
8
6
4
2
0

Mobile Genetic Elements

- ■ prophage element
- ■ genomic island
- ■ SaPI
- ■ SCC element
- ■ plasmid
- ■ transposon/ICE element
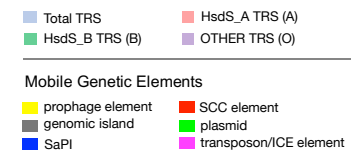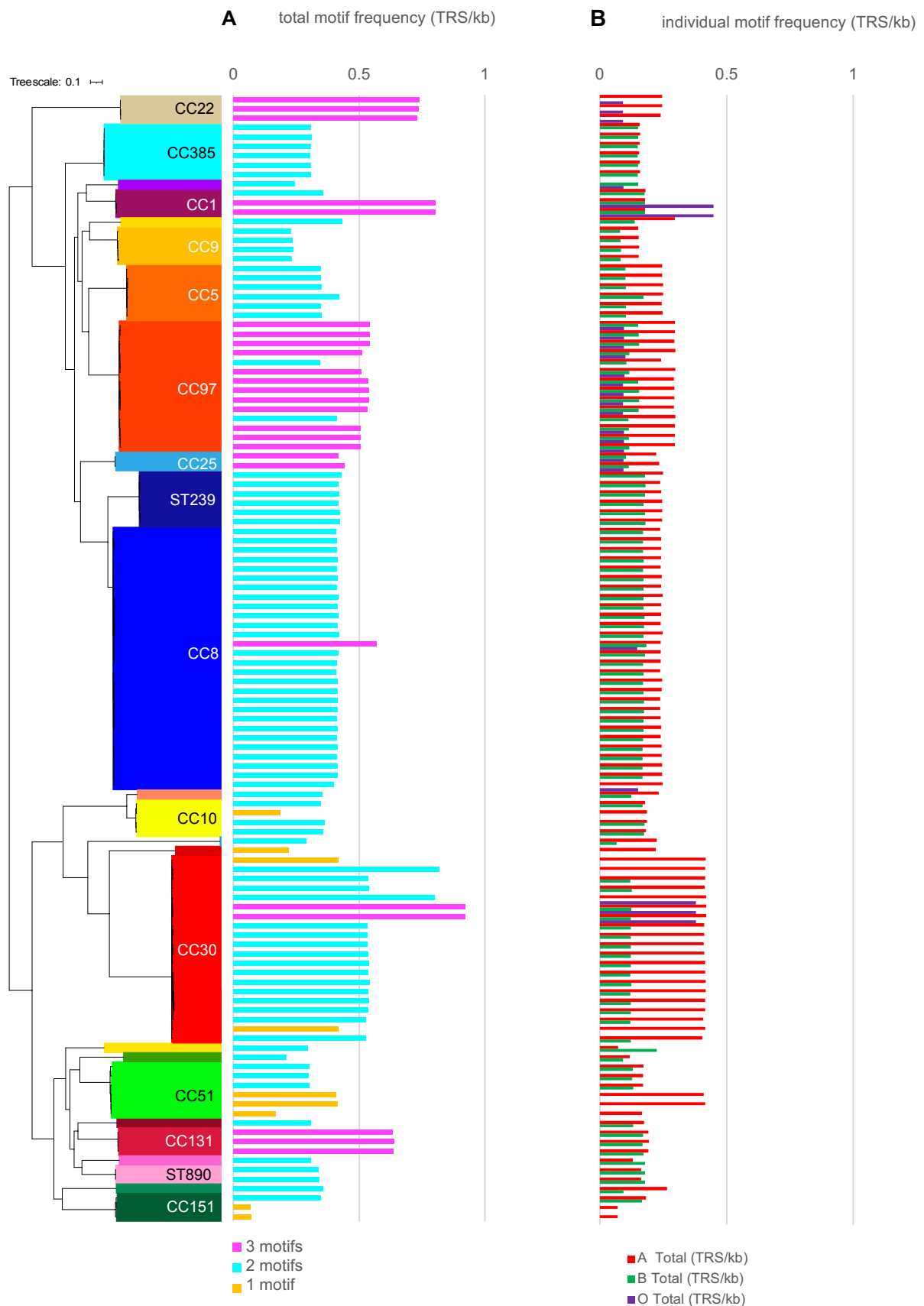
Figure 3.12 | Frequency Difference of Motif Matches in the Whole Genome for NCTC Isolates
A. Total motif frequency (TRS/kb) depending on number of motifs present (pink: 3 motifs, blue: 2 motifs, yellow: 1 motif). **B.** Individual motif frequency (TRS/kb) corresponding to the *hsdS* responsible for TRS binding (red: *sau1hsdMS1* (HsdS_α – A Total); green: *sau1hsdMS2* (HsdS_β – B Total); **purple**: Other (O Total - 3 accessory *sau1hsdS* - *sau1hsdS_SCC* (HsdS_S), *sau1hsdS_orfX* (HsdS_X), *sau1hsdS_E* (HsdS_E)).
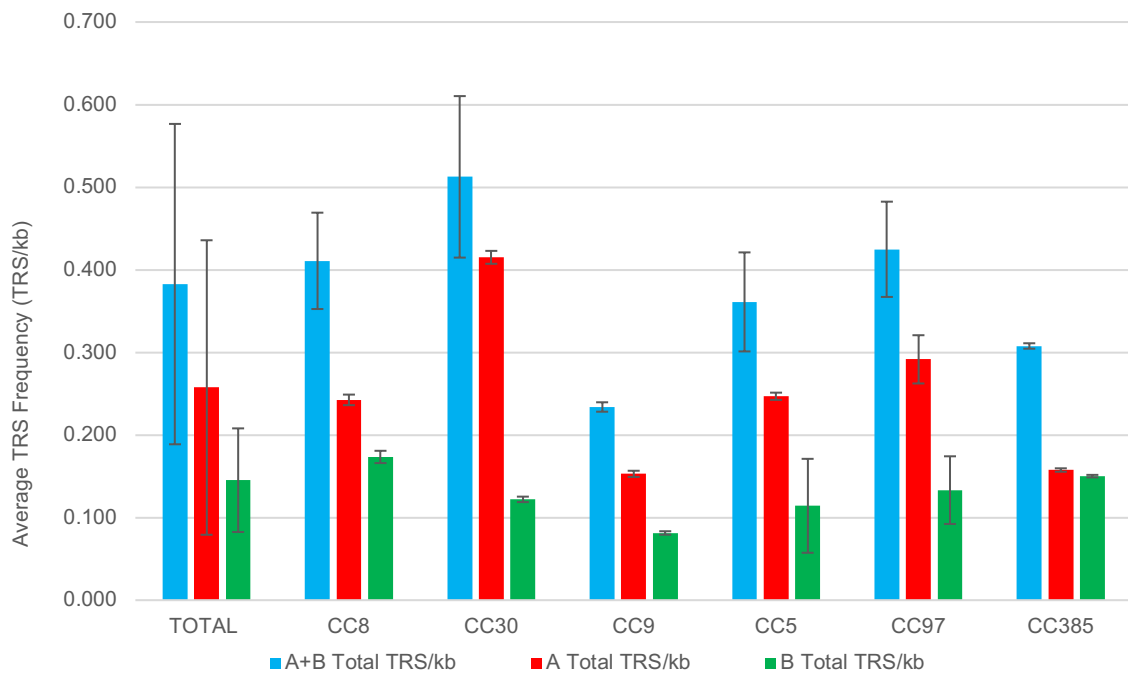
Figure 3.13 | Average TRS Frequency (TRS/kb) for HsdS_α and HsdS_β.
Detailed comparison of methylation patterns within top 6 CCs – CC8, CC30, CC9, CC5, CC97 and CC385.
Blue: Total TRS/kb density for both HsdS_α and HsdS_β, red: TRS match density for HsdS_α, green: TRS match density for HsdS_β. HsdS_α methylates more densly than the HsdS_ β in all lineages.

On the total whole genome level, the average TRS/kb resulting from both systems (A+B) was 0.383 TRS/kb (σ 0.097 TRS/kb), with variations throughout isolates in the main CCs (Figure 3.13). Overall, there were 1.86-fold higher motif matches from HsdS_α (AVG TRS/kb: 0.257, σ 0.097 TRS/kb) than HsdS_β (AVG TRS/kb: 0.146 TRS/kb, σ 0.031 TRS/kb). This trend continues throughout the represented lineages, with CC97 over double, and CC30 over triple the amount of HsdS_α matches than HsdS_β. Isolates from both CC385 and CC10 (see Figure 3.12 B) had an almost 1:1 distribution of both methylation motifs by both core HsdS. There was no distinguishable characteristic found between the TRS recognised by either HsdS_α and HsdS_β regarding length and type of nucleotides making up the S1/S2, or length of spacer (N) nucleotide strings. The combination of TRS for each lineage were also not correlated with either of the two HsdS recognising longer or shorter S1/S2 or separated by longer or shorter degenerate sequences (N).

To gain further insight on the methylation potential of each HsdS, the TRS signatures were investigated in context of different genomic regions including the coding sequence (CDS), intergenic regions (IGR) and the accessory (ACC) vs core genome.

3.5.5.2 Methylation Within Coding Sequences (CDS) and Intergenic Regions (IGR)

DNA methylation as an epigenetic signal regulates specific DNA-protein interactions, hindering protein binding at DNA sequences at methylated target sites (Casadesus and Low, 2006). Important functional units like transcription factor binding sites (within promoter regions), integrase binding sites, small open reading frames, transposons and pseudogene or sRNA-encoding genes are carried within the intergenic region (IGR) of the genome (Srindhar et al., 2011). An important aspect to investigate was whether there is preferential methylation for the intergenic regions or the coding sequences (CDS), which would inform the potential of Sau1 as an epigenetic regulatory mechanism.

Prior to evaluating the methylation frequency within each genomic region, we calculated the proportion of CDS to IGR within each isolate to discount for any discrepancies, in the usually conserved and neutral IGR (Thorpe et al., 2017). The relative proportion of the CDS and IGR were constant in the genomes within the collection. For the coding portion of the genome, the average size was 2.39 Mb (σ 70.63 kb) per isolate, composed of 2675.82 CDSs (σ 113.21) and an average GC content of 33.35% (σ 0.40%). For intergenic regions, the average size was 0.48 kb (σ 20.90 kb) with an average GC content of 32.11% (σ 0.33%) as seen in Figure 3.14.
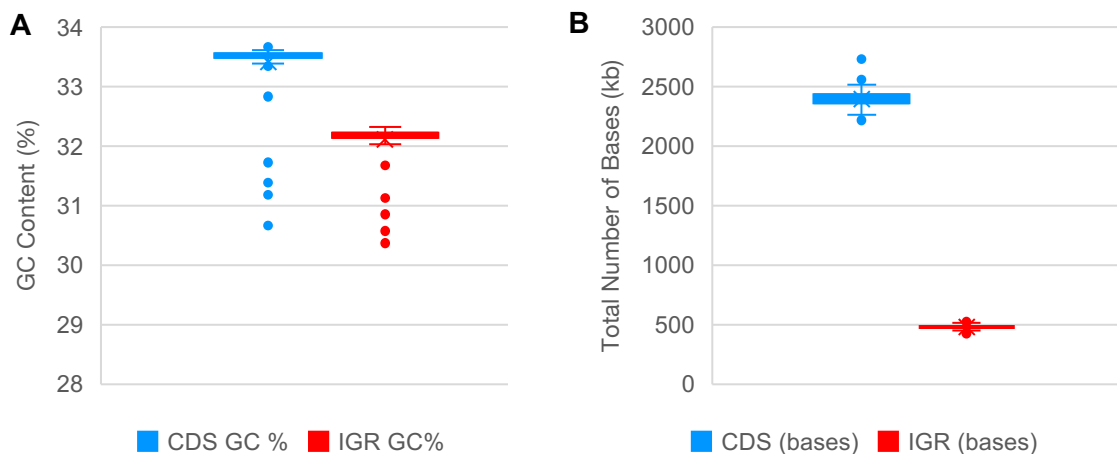
Figure 3.14 | Box-plots showing the distribution of 120 *S. aureus.*
A. GC% content with CDS and IGR regions of the genome, **B.** CDS and IGR lengths.

The average TRS frequency within the IGR and CDS of all isolates (TOTAL) and more detailed data for isolates within 6 CCs is illustrated in Figure 3.15. Across the 6 main CCs, the average motif frequency in the coding sequence for both modification units were 0.405 TRS/kb (σ 0.113 TRS/kb), 1.754-fold higher than in the IGR region (0.265 TRS/KB, σ 0.098

TRS/kb). This pattern is true for HsdS_α with a 1.723 CDS:IGR ratio and HsdS_β with double the amount of average TRS/kb in the CDS as seen in Figure 3.15 (TOTAL).
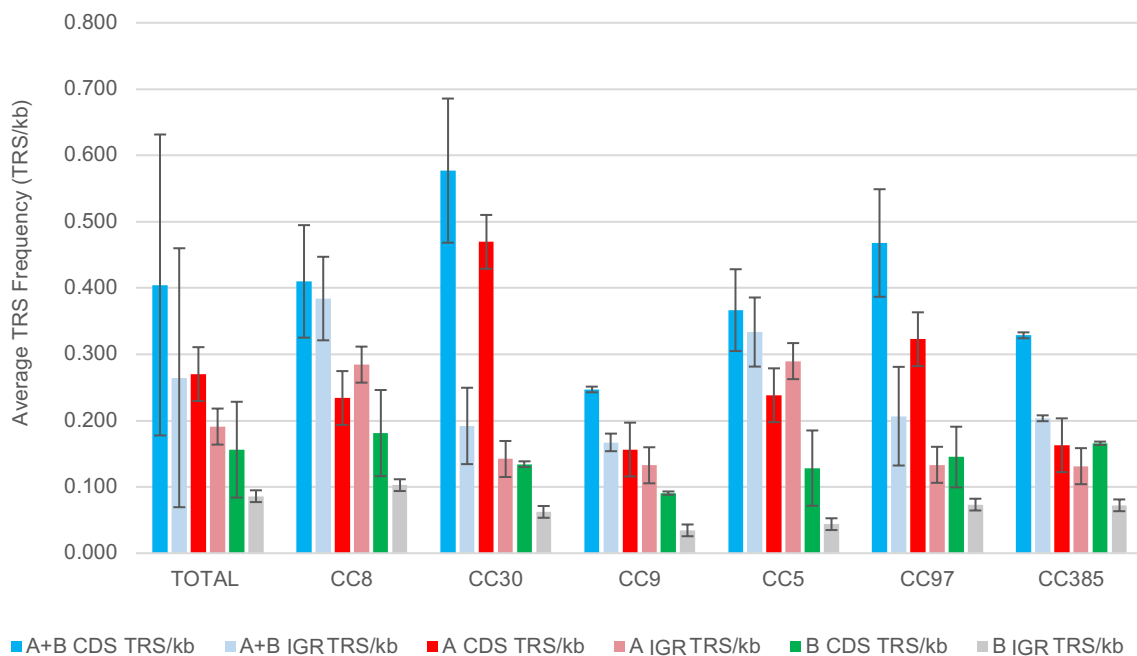
Figure 3.15 | Average TRS Frequency (TRS/kb) for Coding Sequence (CDS) and Intergenic Region (IGR) for HsdS_α and HsdS_β. Detailed comparison of methylation patterns within top 6 CCs – CC8, CC30, CC9, CC5, CC97 and CC385. Blues: Total TRS/kb density for both HsdS_α and HsdS_β within the CDS (sky blue) and IGR (light blue), reds: TRS match density for HsdS_α within the CDS (cherry red) and IGR (salmon), greens: TRS match density for HsdS_β within the CDS (grass green) and IGR (pastel green). The CDS is more densly methylated than the IGR in all lineages.

Although the overall average motif frequencies were higher in the CDS than IGR by HsdS_α, isolates in CC8 and CC5 had slightly increased methylation densities within the IGR, whilst isolates in CC30 and CC97 had +2.612-3.320-fold higher HsdS_α motifs in the CDS. There were no lineages in which the IGR had higher TRS frequency than the CDS recognised by HsdS_β. This is important as it may indicate that within some lineages, HsdS_α preferentially methylates the intergenic region, potentially resulting in disproportional epigenetic regulatory effect between different strains *of S. aureus*.

The *S. aureus* genome does not only differ in coding and non-coding regions, but also varies in composition, due to the presence of variable mobile genetic elements which are horizontally transferred within the species (Lindsay et al., 2006). Although Sau1 has been confirmed to block horizontal gene transfer (Waldron & Lindsay, 2006), some mobile elements have evolved anti-restriction strategies, including alteration of DNA sequence to

remove recognition sites, transient occlusion of restriction sites, and subversion or total inhibition of host RM activity by varying mechanism for successful chromosomal integration or propagation (Tock and Dryden, 2005). Consequently, one main importance of this study was to explore the distribution of 6mA TRS across the MGEs within *S. aureus*, to inform whether some species-specific mobile elements like the SCC elements, SaPI, prophage, *v*Sa, plasmids or transposons were lesser or more densely methylated, giving insight to the relative stability of these elements within the *S. aureus* genome. MGEs also carry antimicrobial, antiseptic and metal resistance genes along with numerous virulence determinants. The selective regulation of these virulence and resistance genes may give advantages to the host.

3.5.5.3 Methylation Within Core (CORE) & Accessory (ACC) Genome

The accessory genome of *S. aureus* makes up about 25% of the total genome (Lindsay et al., 2005). The mobile pan-genome content within the species, as well as within individual lineages, has been shown to be significantly heterologous, in comparison to the highly conserved core genome (Lindsay and Holden; 2004, Bosi et al., 2016; Jamrozy et al., 2016). To investigate the distribution of 6mA Sau1 TRS between the variable accessory genome and the core genome (including the CDS and IGR regions), firstly the accessory genome content within each isolate was defined.

Each isolate, regardless of lineage, had a unique accessory genome as illustrated by Figure 3.16. The average accessory genome size was 171.73 kb (median: 157.00 kb, σ 80.23 kb) per isolate. The accessory genomes were categorised according to 5 types of staphylococcal MGEs: transposons/ICE elements (n=67/120, GC: 30.75%), SCC elements (n=46/120, GC: 32.01%), SaPI (90/120, GC: 30.75%), prophage (n=120/120, GC: 32.84%) and plasmids (n=79/120, GC: 30.22%). The median sequence length comprised by each element (taking into account multiple numbers of some MGEs within one genome) within isolates were: SaPI – 15.45 kb (σ 6.39 kb), transposon/ICE – 17.77 kb (σ 30.19 kb), prophage – 87.48 (σ 51.96 kb), plasmids – 31.73 kb (σ 25.13 kb), and 31.24 kb (σ 18.34 kb) for SCC elements as detailed in Figure 3.17. The median values were used to best represent each MGE total length as there were several outlier isolates, which caried increased numbers of MGEs, mostly accounted for by plasmids (CC90 n=1, CC30 n=1) and prophage (ST239 n=2, ST97 n=2, CC8=3), present within their genome (Figure 3.16 and Figure 3.19).
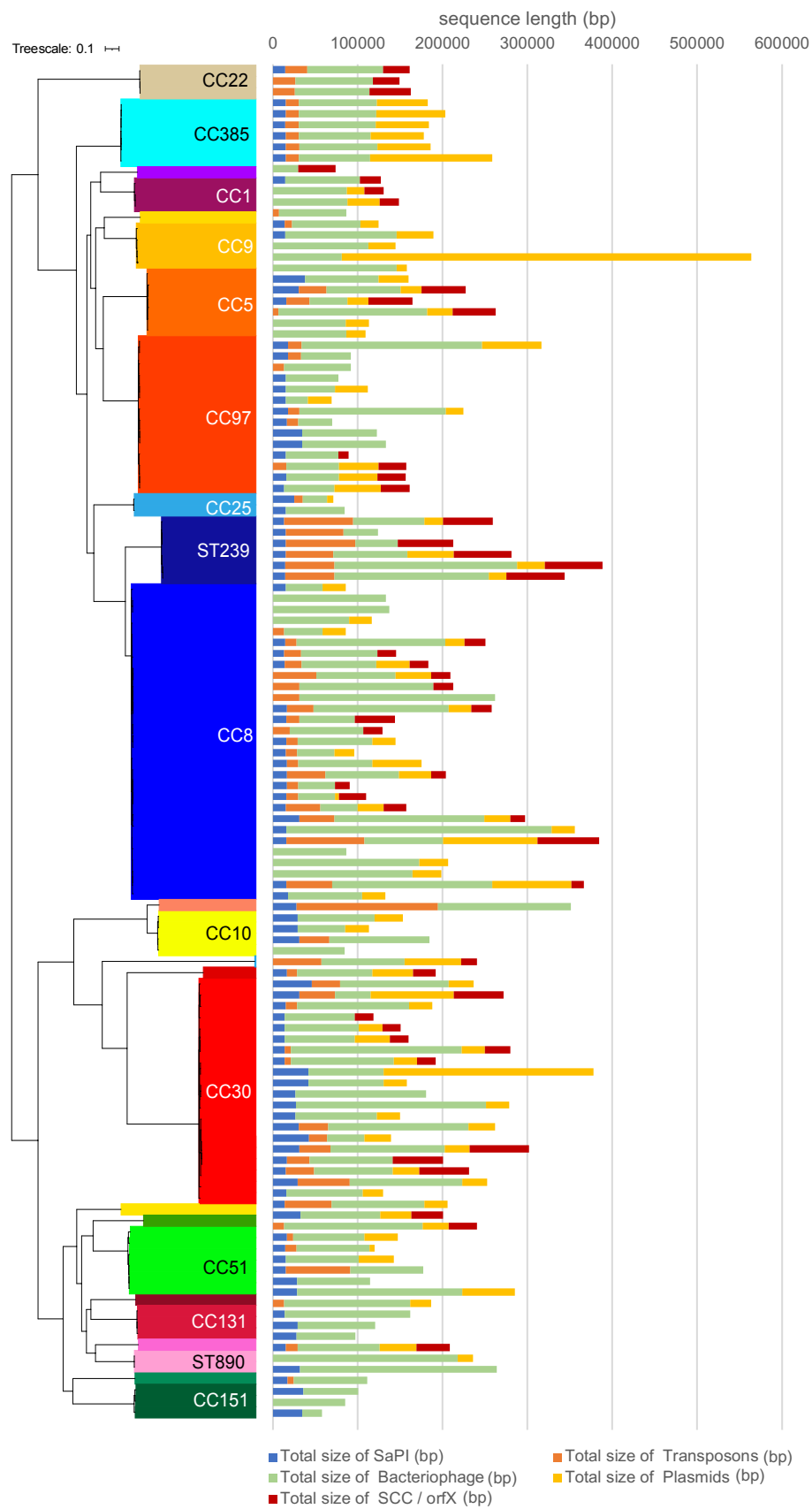
Figure 3.16 | Accessory genome size and distribution of MGEs within whole genome of NCTC Isolates.
SaPI (blue), prophage (green), transposons (orange), SCC/*orfX* inserts (red), plasmids (yellow).
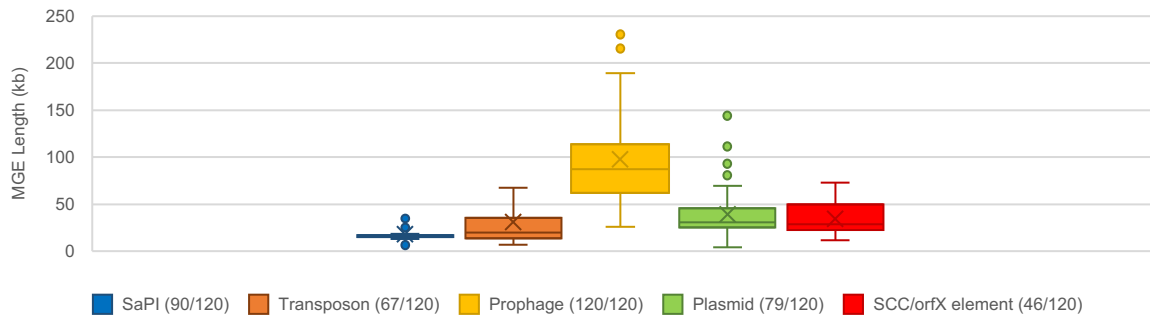
Figure 3.17 | Mean aggregate genome length comprised by MGEs in NCTC Isolates.
Size (kb) of: total sum of MGEs (grey), SaPI (blue: n= 90/120), prophage (green: n=120/120), transposons (orange: n=67/120), SCC/*orfX* inserts (red: n=46/120), plasmids (yellow: n=79/120).

The global comparison of TRS matches within the accessory genome and the core genome of isolates belonging to 6 main CCs represented within the collection was investigated as seen in Figure 3.18. The relative TRS/kb for the core genome follows the previously seen clonal distribution. The mean motif frequency for the core genome (0.387 TRS/KB, σ 0.099 TRS/kb) was 1.35-fold greater than in the accessory genome (0.317 TRS/KB, σ 0.133 TRS/kb). HsdS_α and HsdS_β have higher methylation match densities (1.26-fold and 1.965-fold) for core genome compared to the accessory genome. HsdS_α matched both the core (1.829-fold) and the ACC (3.14-fold) more frequently than HsdS_β. The mean TRS/kb by HsdS_α for isolates within CC8 and CC9 showed marginally higher methylation frequencies in the accessory genome. This may be due to a higher density of methylation of a specific type of mobile element. Consequently, we investigated the methylation landscape within the five major types of MGEs was analysed (Figure 3.19).
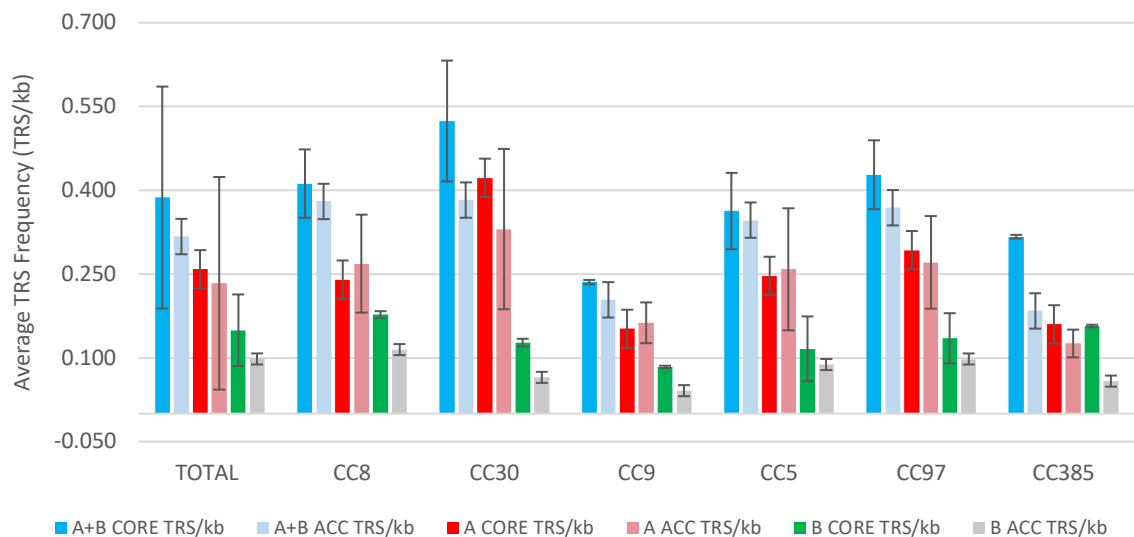
Figure 3.18 | Average TRS/kb Frequency in Core/Accessory (ACC) genome for HsdS_α and HsdS_β.
Detailed comparison of methylation patterns within top 6 CCs – CC8, CC30, CC9, CC5, CC97 and CC385. Blues: Total TRS/kb density for both HsdS_α and HsdS_β within the CORE (sky blue) and ACC (light blue), reds: TRS match density for HsdS_α within the CORE (cherry red) and ACC (salmon), greens: TRS match density for HsdS_β within the CORE (grass green) and ACC (pastel green). The CDS is more densly methylated than the IGR in all lineages.
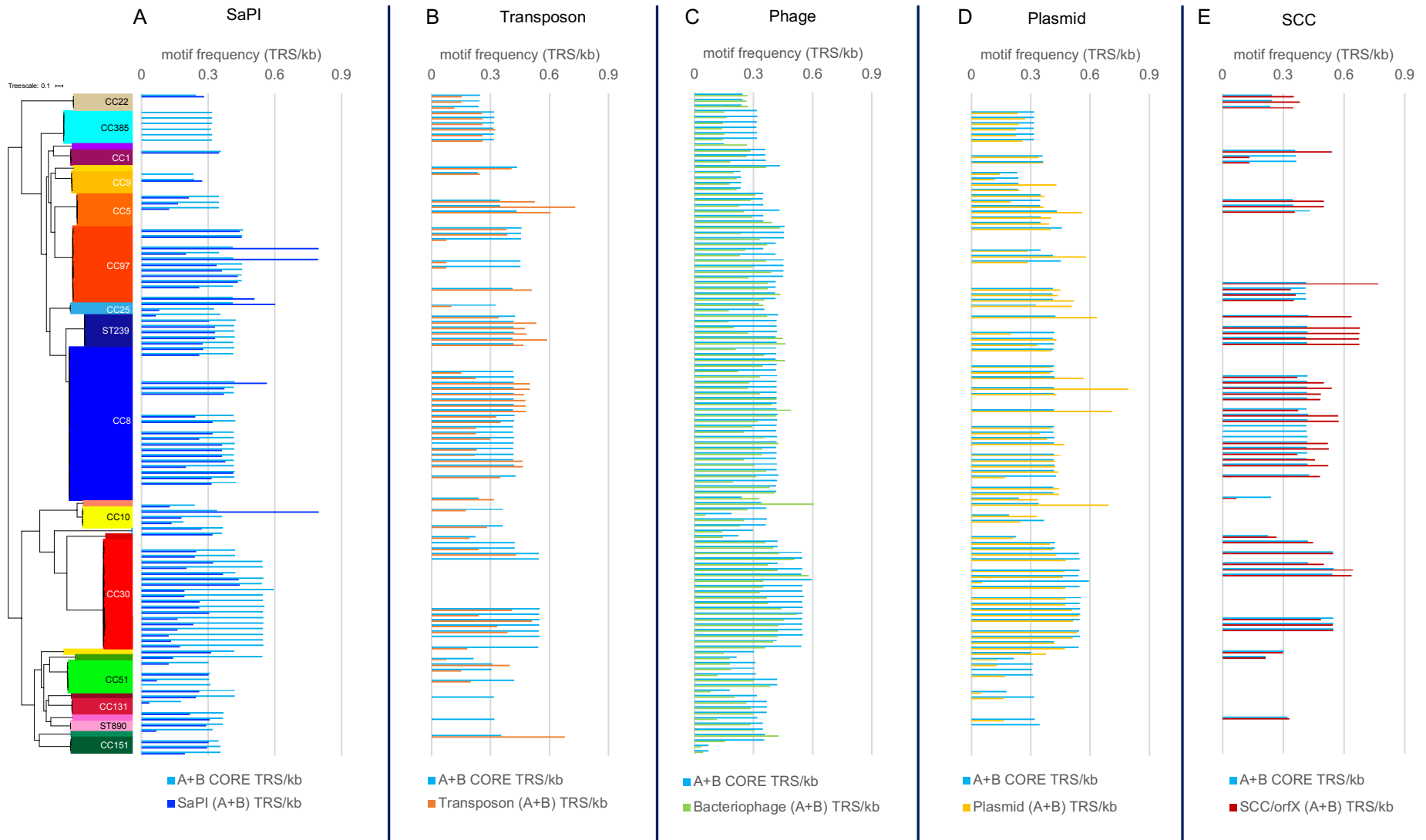
Figure 3.19 | Motif Frequency within MGEs (SaPI, Transposons, Prophage, Plasmids, SCC/*orfX* element inserts) in comparison to the motif frequency within the core genome of 120 historic *S. aureus* isolates. A. Average TRS frequency within SaPI for HsdS_α and HsdS_β. **B.** Average TRS frequency within Transposons for HsdS_α and HsdS_β. **C.** Average TRS frequency within Bacteriophage for HsdS_α and HsdS_β. **D.** Average TRS frequency within Plasmids for HsdS_α and HsdS_β. **E.** Average TRS frequency within SCC/*orfX* inserted elements for HsdS_α and HsdS_β.

The TRS/kb for the core genome seems to average out to match the total TRS/kb within SaPI (Figure 3.19 A), transposon (Figure 3.19 B) and prophage (Figure 3.19 C), although it is clear that the methylation of each MGE varies within each isolate, with moderate conservation within represented lineages. Smaller elements like SaPI and transposon may not carry methylation motifs as a whole, escaping Sau1 restriction activity, potentially increasing the ease at which these elements are horizontal transferred within the species. The accessory genome of each isolate even within a lineage varies significantly, hence methylation trends / biases were more difficult to deduce for each MGE compared to the methylation density of the core genome.

3.5.5.4 TRS Frequency Within *S. aureus* Pathogenicity Islands (SaPI)

As show in in Figure 3.19 A, 75% (n=90/120,) NCTC isolates carried at least one SaPI, with 54 isolates containing 2, and 26 isolates carrying 3 islands respectively. Eight isolates contained only a single SaPIs within CC385 (6), CC9 (1), and CC51 (1), and none were methylated. The average TRS/kb for both HsdS_α and HsdS_β was 0.265 TRS/kb (σ 0.168 TRS/kb), with the average methylation in the core genome being 1.532-fold higher than SaPI elements. Within some isolates (ex: CC97), the motif density was higher than the average core methylation density due to two isolates carrying multiple SaPI with significantly higher TRS/kb matches. Neither the *sau1hsdMS* in either genomic islands or the accessory related Sau1 methylation elements had higher methylation densities within the SaPI than the core genome (CORE:SaPI - HsdS_α (1.469 – exp CC97); HsdS_β (1.764 – exp CC9). There were almost twice (1.894) as many TRS/kb matches for HsdS_α than HsdS_β, with far more isolates not having many HsdS_β motif matches, increasing the difference between the two methylation units as for SaPI in CC30, CC5 and CC97.

3.5.5.5 TRS Frequency Within Transposons

As shown in Figure 3.19 B, 56% (n=79/120) of the isolates carried at least 1 transposon sequence, mainly Tn*554* (*erm* – erythromycin resistance, *spc* – spectionmycin resistance), Tn*552* (*blaZ* – penicillin resistance), Tn*916*-like and Tn*5801* (*tetM* – tetracycline resistance). Five isolates in CC30 (2), CC8, ST707, and ST136, contained transposons with TRS matches which were not methylated. Overall, the core genome was 1.28x more densely methylated than transposons heavily skewed by the very low mean level of motif

matches to HsdS_β in transposons (0.074 TRS/kb, σ 0.068 TRS/kb) compared to over twice (2.180-fold) number of TRS in the core genome (0.149 TRS/kb, σ 0.032 TRS/kb) for this protein. Overall, there were approximately equal (A CORE: A Transposon, 1:1.013) methylation matches in the core genome versus transposons for HsdS_α. Motif matches for HsdS_α were 3.689 more frequent in transposons compared to HsdS_β. This average ratio is influenced heavily by very low frequency of methylation matches to HsdS_β throughout transposons in CC30 and CC97.

3.5.5.6 TRS Frequency Within Prophage

All 120 isolates contained at least one phage (Figure 3.19 C), with over 75% of the isolates containing ≥2 viral sequences. The distribution of methylation for this particular MGE shows some conservation within the recognised frequency of motifs by HsdS_α, within CC8, CC30, CC97, CC22. As for SaPI and Transposons, the average methylation frequency within the core genome (0.387 TRS/kb, σ 0.099 TRS) is 1.285-fold higher in comparison to prophage (0.301 TRS/kb, σ 0.112 TRS/kb). The average core genome matches are also 1.198 times more frequent (except CC9) than the HsdS_α and 1.611 times higher than the methylation frequency to HsdS_β across prophage for all other CCs. In total, there was a 1.718 higher methylation match density to HsdS_α in bacteriophage compared to HsdS_β, as some phage did not have TRS matches for the later unit.

3.5.5.7 TRS Frequency Within Plasmids

Figure 3.19 D details 66% (n=79/120) of isolates contained at least 1 plasmid. Plasmids within 8 isolates CC51 (2) ST890 (1), CC8 (3), CC25 (1), ST1254 (1) had no methylation matches. Overall, the methylation frequency for plasmids follows the same clonal distribution as for the core genome, with instances of higher TRS/kb in plasmid elements for individual isolates spanning a variety of CCs. This could be influenced by the length of the plasmid DNA which is potentially methylated. The average methylation frequency within the core genome (0.387 TRS/kb, σ 0.099 TRS/kb) was nearly equal (1:1.065) to the TRS/kb within plasmid (0.369 TRS/kb, σ 0.160 TRS/kb). Isolates within CC 385 and CC30 were the only strains with slightly lower methylation density within the MGE compared to the core genome. This is influenced by the higher proportion of plasmid TRS/kb presented in throughout the collection (CORE:plasmid 1:0.965), and specifically CC8 (CORE:plasmid

1:0.735), CC5 (CORE:plasmid 1:0.816), and CC385 (CORE:plasmid 1:0.904) corresponding to HsdS_α. HsdS_α had slightly lower TRS/kb than the matches by the same system in the core genome (1:0.965). HsdS_β had lower TRS matches within plasmids than in the core genome (1.427-fold higher TRS/kb) for isolates in all CCs.

3.5.5.8   TRS Frequency Within SCC/*orfX* Inserted Elements

Within this study, 38% (n=46/120) of isolates contained an SCC element/*orfX* insert, of which 70% were SCC*mec* elements (n=32/46: CC30 (n=3/3); CC80 (n=1/1), CC1 (n=1/3), CC5 (n=3/6); ST239 (n=5/6), CC8 (n=14); CC395 (n=1/1); CC30 (4/20) detailed in Figure 3.19 E. There were 3 CC8 isolates carried a SCC*mec* IVc element which did not have any target recognition sites. The SCC elements were the only MGEs which were more frequently methylated than the core genome, as a total by both units (SCC:CORE - 1:0.866), by HsdS_α (SCC:CORE - 1:0.831) and also by HsdS_β (SCC: ORE - 0.965). Overall, there was 1.330-fold more methylation motif density matched to HsdS_α than HsdS_β

3.6 DISCUSSION

This historically and phylogenetically diverse collection of *S. aureus* isolates has allowed the characterisation of the overall species-wide Restriction-Modification diversity, and Sau1 6mA methylation landscape, with unprecedented detail. PacBio sequencing of this largescale collection allowed the investigation of the diversity of Sau1 6mA recognition motifs as well as the frequency of adenine modification throughout different regions of the *S. aureus* genome. Within this work, the protein structure of 40 HsdS alleles was studied in depth, predicting DNA binding loops for each specificity unit, to collate a comprehensive database of TRDs and corresponding PacBio derived TRS motifs. Frequency analysis of 6mA methylation throughout each isolate genome allowed the exploration of methylation biases and trends throughout the *S. aureus* chromosome, pertaining not only to different genome regions (CDS, IGR, core, accessory) but also the 6 different *sau1hsdS* loci uncovered in this study.

3.6.1 Multiple Restriction-Modification within *S. aureus* genome

This study characterised the RM landscape within *S. aureus*, augmenting the number of categorised RM elements within a large collection of isolates, giving a unique, holistic snapshot of the Restriction-Modification activity within the species. Every isolate (n=120) in this study carried a TI *sau1* system and an additional supplemental TIV (Figure 3.3) restriction endonuclease (5mC and 4mC activity) or a TII RM system (Table 3.2) with 6mA and 5mC modification/restriction activity (Figure 3.2). Over 80% of bacterial strains contain multiple RM systems (Vasu & Nagaraja, 2013), which have undergone extensive inter and intraspecies HGT (Furuta & Kobayashi, 2011). Paradoxically, these mobile systems (often carried on MGEs) act as increased levels of 'innate immunity' within the host cell against invasion of exogenous DNA (Makarova et al., 2014).

Most NCTC strains possess a TI (*sau1hsdR*) and TIV (*sauUSI*) endonuclease posing a significant barrier to genetic manipulation, yet inactivation of either of these restriction proteins is not sufficient to generate *S. aureus* strains capable of completely accepting foreign DNA more efficiently (Monk et al., 2012; Xu et al., 2012, Veiga & Pinho, 2009; Roberts et al., 2003). Thus, isolates carrying multiple RM types will have increased protection from invading DNA, due to restriction and methylation activity on additional RM substrate specificities, 5mC, 4mC, and 6mA (Oliveira et al., 2014; Corvagha et al., 2010). Consequently, although *sau1* is the most abundant RM system within *S. aureus*, it can be

proposed that the control of exchanging foreign DNA within the species, does not solely rely on TI RM *sau1*, but is broadened by TII RM activity and TIV restriction as exemplified in this study. *Sau1* is also credited to be have defined the clonal landscape of *S. aureus* (Waldron & Lindsay, 2006), nonetheless additional mobile RM systems have also influenced the more recent evolution of lineages, their accessory genome content (Oliveira et al., 2016), and greatly contribute to the ability for strains to rapidly adapt to their environment.

3.6.2 Sau1 TI Restriction-Modification of *S. aureus*

The results of this study concur that *sau1* TI RM systems are widespread and 'define' a predominantly clonal landscape of *S. aureus* (Figure 3.6), especially as they are part of the genomic islands, an important genomic location of variable lineage specific determinants (Baba et al., 2002; Waldron & Lindsay, 2006; Kläui, Boss & Braber, 2019). The *sau1* modification elements (core *hsdMS* loci) within each vSa have been linked to the stabilisation of these islands within the *S. aureus* genome (Kuroda et al., 2001), and may be the key factor in the slow evolutionary change within these elements, acting as the main barriers against horizontal transfer and genome diversification. Juxtaposing the slow evolutionary change incurred by the *sau1hsdMS* within the genomic islands (Roberts et al., 2013) are the four additional 'accessory' genomes associated *sau1* elements (Figure 3.3 and Table 3.1) characterised in this study, which are readily transferred on mobile genetic elements like SCC elements (SCC*mec*; SCC*far*; *orfX* inserted phage, ICE element or transposon) or plasmids via HGT. The acquisition of either an additional accessory orphan (*sau1hsdS_orfX, sau1hsdS_ϕ*), modification unit (*sau1hsdMS3*) or full *sau1* system (*sau1hsdRMS*) by some strains may be to potentially alleviate the cost of a lost, truncated or inactive core *sau1hsdMS*, as the case with ST22 isolates, with only one functioning 'core' *sau1hsdMS* in vSaα, with an additional s*au1hsdRMS* inserted within the SCC*mec* (Figure 3.2, Figure 3.6).

Along with the high level of stability noted to the location of the 'core' *hsdMS* loci, the amino acid homology of the coded HsdM show 99-100% sequence identity within both genomic islands (Baba et al., 2002), as well as the methyltransferase coded for by *sau1hsdMS3* and *sau1hsdRMS* systems uncovered in this study. HsdS proteins encoded within *sau1hsdMS* operon (and solitary *sau1hsdS*) exist in allelic forms with AA identity lower than 66%, due to variability within TRD domains (Baba, 2008; Kläui, Boss & Braber, 2019). Thus, the variability within the HsdS target recognition is which contributes to the diversity in methylation and restriction specificity of each Sau1 system defining each dominant

lineage of *S. aureus* (Feil et al., 2003; Waldron & Lindsay, 2006; Lee et al., 2019, Blackeway et al., 2014, Titheradge et al., 2001).

3.6.3 Sau1 HsdS in *S. aureus* – Diversity and Structure

One of the most important elements of this study was the opportunity to use PacBio sequenced isolates to not only identify the methylation motifs within a strain, but also use the modification predictions to identify the specificity of each HsdS protein within the NCTC collection. The detailed protein structure of 40 HsdS homologs (Figure 3.6) were characterised and matched to their corresponding 6mA specificities. The protein sequences resolved into 18 TRD1 and 28 TRD2 domains, uncovering 6 novel TRD1 and 10 novel TRD2 protein domains for the Sau1 HsdS. Variants of HsdS and the relative conservation of the proteins stayed consistent within ST groups as described by Lee et al., 2019. The findings of this study support this lineage specificity for Sau1, as identical *hsdS* alleles were found between lineages within not just the previously studied *sau1hsdMS_α, sau1hsdMS_β* (Baba, 2008; Kläui, Boss & Braber, 2019)*,* but also the additional 'accessory' genome associated *sau1* elements (Figure 3.6 and Table 3.5), which strongly correlated to the allelic forms of genomic islands and MGEs within each ST. There was no evidence of phase variation as no duplicate *hsdS* gene was found in a single locus, suggesting no 'classical' epigenetic switch or differential methylation as seen in other bacteria (Bayliss 2009; Srikhanta *et al.*, 2011; Casadesús & Low, 2013; Atack 2015; Seib *et al.*, 2015; Anjum *et al.*, 2016). However, there was one example of allelic heterogeneity of HsdS_ β TRD1 domain (**TRD1 A**+TRD2 W→ **TRD1 Z**+ TRD2 W – Table 3.5 / Figure 3.6) within CC97 lineage, resulting in differential methylation motifs This target domain deviation results in differential methylation motifs (TRS#10 GAC (N_6) CRAA → TRS#32 CCAY (N_6) CRAA) within the same lineage. Sullivan et al (2019) also found one single instance of recombination for within the TRD1 position of HsdS_*α* within CC5 isolates (USA100) changing the typical **TRD1 B** + TRD2 D, to **TRD1 A** + TRD2 D. Roberts et al., suggest that the recombinant *hsdS* have evolved through extensive recombination of the two 'core' *sau1hsdMS* in repeated occurrence, with potential for reversible chromosomal inversions (CI) at multiple repeats within the conserved regions of the *hsdS* as exemplified by Guérillot et al., 2019. These discrete TRD switches illustrate the evolution of *hsdS* within the species and the potential for differential methylation within the species.

Another significant aim of this study was to identify trends for the TRD composition and conservation of domains throughout the characterised HsdS variants. Firstly, the protein

sequences for HsdS_α being more variable than those in HsdS_β, suggesting differences between the relative evolution and recombination between the two genomic islands (Figure 3.6 and Table 3.5). Everitt et al., 2014 suggest that vSaα is a hotspot for recombination, potentially giving rise to more heterologous HsdS within vSaα (19 HsdS variants, 12 TRD1: 13 TRD2) rather than vSaβ (16 HsdS variants, 7 TRD1: 10 TRD2). Secondly, conservation of TRD1 domains between HsdS_α and HsdS_β, sharing 6 TRD1 (A, NT1*G, e*, J, R, Z, and X), whilst there was no sharing of TRD2 domains (Table 3.5). The conservation of TRD1 between the HsdS_α and HsdS_β between lineages could be attributed to possible duplication of the same *sau1hsdMS*, and to avoid redundancy, a single TRD variation in the flexible TRD2 position (more readily recombinogenic between the mosaic structure - central and C-terminus conserved region) is introduced (Roberts et al., 2013; Guéillot et al., 2019; Attack et al., 2020). Taking all of this into account, some isolates within the collection also only carried one working HsdS, with a single 6mA methylation target, which raises the question of potential functional redundancy between the two core TI Sau1 systems.

Conducting a detailed characterisation of Sau1 HsdS also entailed predicting the proposed DNA binding loops of each TRD (Table 3.8) with modelling approaches using known TI RM crystal structures as a reference (Kim et al., 2005, Obraska et al., 2006; Gao et al., 2001; Liu et al., 2017). This approach aimed to further our understanding of the structure and molecular function of HsdS, by characterisation of the secondary structure homology of each conserved region (Figure 3.9), globular domain and proposed binding loops (Table 3.8). The proposed positional information of the AA strings which make up each binding loop matched to the corresponding nucleotide strings within the TRS they bind, coupled with a machine learning approach, may potentially alleviate the need for costly methylation detection techniques including whole genome methylation profiling, bisulphite conversion, differential enzymatic cleavage and affinity capture (Kurdyukov & Bullock, 2016). Deep learning models have already been successfully used for regression of genome-wide DNA methylation eukaryotes (Tian et al., 2019; Jurmeister et al., 2019; Zhang et al., 2015) and have even been specifically designed for N6-adenine methylation (Basith et al., 2019; Khanh Le et, 2019; Lv et al 2019). A collaborative machine learning modelling project was begun with Dr. Ognjen Arandjelovic's group at the University of St Andrews, but further adjustments to the methodology need to be made. Ultimately, the 3D crystal structure of each individual TI HsdS, HsdM and HsdR elements, as well as the heterodimer complexes M_2S and R_2M_2S need to be resolved first, for full confirmation of interacting protein regions.

3.6.4 Sau1 Facilitated 6mA Methylation in *S. aureus*

This study is unique in itself as it is the first largescale study of targeted PacBio SMRT sequencing to study the 6mA methylation landscape within the *S. aureus* species. Previously, Roberts et al., (2013) reported 6mA methylation was randomly dispersed throughout the *S. aureus* genome, but there was no information on the relative frequency at which the chromosome was methylated, or methylated by M_2S, namely which specific HsdS recognised each target motif. Within this study each predicted methylation motif (set of TRS – Figure 3.10) was matched to the *sau1hsdMS* system responsible for modifying at the determined sequence string, building a comprehensive database of HsdS structure (TRD) and function (TRS).

The relative frequency of TRS matches for 40 6mA methylation motifs (each corresponding to a different HsdS) were investigated throughout the whole genomes of NCTC isolates (Table 3.4, Figure 3.10) to give a snapshot of motif densities as per motif. Methylation motifs recognised by HsdS_α seem more frequently distributed throughout the genome than HsdS_β, although no correlation could be drawn between HsdS carrying prevalent TRD types (A, B, J, E, P, NT1*G, Z – Table 3.5, Figure 3.10) and higher/lower TRS frequency within a genome. There was also no evidence of hot/cold spots for methylation within any of the 120 *S. aureus* strains, concurring with the findings of Furuta et al., 2014 (Figure 3.11).

As previously established, HsdS variants followed a clonal distribution, hence there was no significant difference in the methylation density within a lineage regarding the core Sau1 modification complexes. The overall methylation within a lineage was only altered by additional methylation from supplementary 'accessory' HsdS, or methylation differences introduced with increased accessory genome size (additional phage or plasmids) on an individual isolate basis. Notably, there were differences in frequency of detected methylation motifs by distinctive methylation complexes (formed with HsdS_α, HsdS_β, HsdS_X, HsdS_S, and HsdS_E) within different regions of the genome (Figure 3.12, Figure 3.13, Figure 3.14, Figure 3.15, Figure 3.18, Figure 3.19). Isolates with increased number of active HsdS proteins had overall higher methylation frequencies (Figure 3.12A) but was heavily influenced by the specificity of the given HsdS within a lineage. When focusing on the two 'core' Sau1 *hsdMS*, strikingly, HsdS_α had 1.86x higher methylation frequency than HsdS_β, throughout the whole genome (Figure 3.12 B, Figure 3.13), and preferentially methylating CDS over IGR (Figure 3.15), and the core rather than accessory genome (Figure 3.18, Figure 3.19). Although the methylation distribution stays random,

the preferential methylation by the modification complex formed with HsdS_α indicates a clear bias and functional difference between the two core *sau1hsdMS*.

Methylation by either of the two core Sau1 modification units exposed a bias towards methylation of the CDS rather than the intergenic region (Figure 3.15). The likelihood of TI Sau1 having a role in transcriptional regulation by DNA adenine methylation is therefore lessened. Lower TRS frequencies within the intergenic region, with base modification located outside of the promoter or regulatory region of CDSs, would indicate low probability of modulating the binding of transcriptional factors or the RNA polymerase as classically seen in the literature (Sanchez-Romero, Cota & Casadesus, 2015).

It has been suggested by Roberts et al., (2013) that mobile genetic elements such as plasmids, have evolved to have decreased TRS sites allowing restriction escape. In concurrence with this, the results from our study show higher TRS frequencies in the core genome compared to the mobile accessory genome and 3/5 main mobile element types found within *S. aureus* (Figure 3.18– **Error! Reference source not found.**). Out of the 5 main mobile elements present in the, the average TRS/kb within SaPI, transposons, prophage was lower or equal to (plasmids) that of the core genome, with only SCC elements being more densely methylated (Figure 3.19). Smaller MGEs like SaPI and transposon showed some non-methylated TRS, mainly lacking TRS matches to HsdS_β. SaPI, transposon and bacteriophage all have mosaic structures with repeat regions, which might influence the density at which their sequence is methylated (Novick & Ram, 2016; McCarthy et al., 2012).

Mobile elements have also evolved anti-restriction strategies, including alteration of DNA sequence to remove recognition sites, transient occlusion of restriction sites, and subversion or total inhibition of host RM activity by varying mechanism for successful chromosomal integration or propagation (Tock & Dryden, 2005). Some mobile elements carry their own RM systems and out-compete the host Sau1 RM activity via propagation of their own RM system (Vasu & Valakunja, 2013). This could be the likely function of the 4 HsdS characterised to be associated with the accessory genome characterised throughout this chapter. This suggests that some MGEs which have established themselves within the *S. aureus* genome, have methylation protection by the host Sau1 methyltransferase, and have successfully bypassed not only the Sau1 restriction endonuclease, but SauUSI (TIV) and in some cases TII 5mC specific restriction units.

Consequently, RM systems may function not just as a barrier to MGE infection, but also acts as a stabiliser of the already established MGEs within the cell (Oliveira et al., 2016).

3.6.5 Limitations and Future Work

One key limitation of this study was only using the absolute number of TRS detected within each genome to give an overview of Sau1 6mA methylation within *S. aureus*, as several methods and attempts to parse the positional information of unmethylated TRS within the raw PacBio motifs.gff and modifications.gff files were unsuccessful. Additionally, none of the existing methylation analysis tools could be used to distinguish detect vs methylated TRS, nor locate the exact position of the ~1.8% (median) unmethylated bases within genomes (mean: 2.7%, range: 9.70 – 100.00%). The average number of unmethylated nucleotides (calculated from all modified vs all detected TRS) equated ~15 nucleotide bases per genome. The exact positions of unmethylated adenine residues would have given insight into hemi-methylation of some TRS sites which may have an epigenetic regulatory role yet undiscovered. There were also some quality issues with some of the PacBio SMRT sequencing, forcing the removal of 40 NCTC *S. aureus* isolates from the collection, and also variance in quality and efficacy of 6mA methylation calling for some of the assembled genomes. Additional analysis of the raw IPD data was necessary to determine modification artifacts and validate double stranded methylation for 9 isolates from various lineages (Table 3.4 and Figure 3.5). Re-sequencing of some isolates with the updated PacBio sequencing platforms may be beneficial, especially of isolates belonging to lesser represented STs.

Although the NCTC collection included a phylogenetically diverse set of isolates, several dominant STs were overrepresented like CC8, CC30, CC97 with many represented STs only containing a single isolate ST1148, ST1254, ST1021, ST151, ST707, ST80, ST489 among others. This collection is also mainly historical isolates dating back over 80 years, with limited metadata, heavily biased towards human samples, mainly from Europe and the USA. Conducting these analyses on a larger cohort of international strains, with at least 5 isolates from each representative ST, including LA, HA, CA, MRSA and MSSA isolates would greatly augment the current knowledge of the diversity of Sau1 within *S. aureus*.

Another obvious limitation of this study was the lack of crystalised *S. aureus* Sau1 proteins, hence all modelling within this study was template guided (Figure 3.8) basing the secondary structure on 1YF2.1/1YF2.2 HsdS from *Methanocaldococcus jannschii* (Kim et

al., 2005). The 3D structure of at least one Sau1 HsdS would be necessary to validate the structural predictions, especially for the active site of each TRD. Confirmation is also necessary for the accurate development of a machine learning model for prediction of Sau1 modification throughout whole genome sequences.

Perhaps the most important further study is investigation of the potential gene regulatory role of Sau1 6mA throughout the whole genome, which was explored in subsequent chapters of this thesis.

4. THE EFFECT OF LARGE-SCALE CHROMOSOMAL REPLACEMENT ON WHOLE GENOME 6mA METHYLATION AND GENE EXPRESSION PROFILES IN *S. AUREUS*

4.1 INTRODUCTION

The genetic repertoires of prokaryotic species are continuously diversifying, which is vital for their adaptation to changing environments (Blow et al., 2016). The main mechanism of genetic variation in bacteria is through mutation or genetic exchange via horizontal gene transfer. Recombination and large-scale chromosomal rearrangements are far less likely than genotypic diversification within the natural populations of pathogenic bacteria, specifically clonal species like *S. aureus* (Feil et al., 2003). The mechanisms and importance of recombination in the long-term evolution of the species is still undetermined, but Everitt et al., (2014) have identified that the chromosomal regions flanking MGEs (hotspots flanking ICE*6013*, SCC*mec*, SaPIs, νSaα) and a ~750 kb region spanning the origin or replication (*oriC*) have elevated recombination rates. Core genome transfers (CGT) which occur through an MGE-independent mechanism, also remain a paradox, especially in seemingly untransformable bacteria like *S. aureus* (Didelot & Wilson et al., 2015). Chromosomal replacements are often associated with fitness disadvantages (Vogan & Higgs, 2011), yet one such large-scale recombination event resulted in the extremely successful hybrid MRSA, ST239 strain, persisting as one of the most prevalent genotypes of *S. aureus* globally (Robinson & Enright, 2004; Holden et al., 2010).

In 2014, a novel hybrid HA-MRSA strain, ST622, was discovered during a 3-year (2014-2016) cross-sectional carriage study assessing major strains circulating closely affiliated intermediate and long-term care facilities in Singapore (Chow et al., 2017). A single isolate (CD141496) of the chimeric strain was sequenced and preliminary WGS and phylogenetic analysis (Matthew Holden - unpublished) resolved ST622 into clonal complex 22. The novel strain is a hybrid of the two most prevalent HA-MRSA lineages in Singapore, ST22 - the 2.5 Mb backbone, and a 366 kb genome segment replaced along the origin and terminus of replication with sequence of ST45 origin as seen in Figure 4.1). Phylogenetic reconstruction of
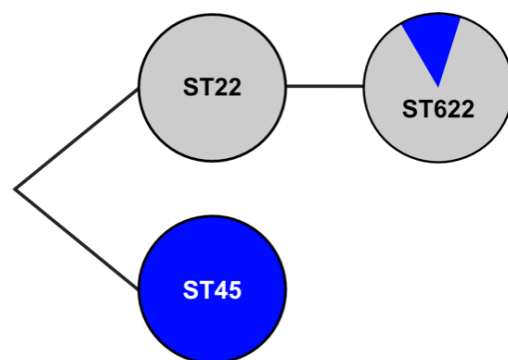
Figure 4.1 | ST622 Genome Replacement
Evolutionary relationship between ST22, ST45 and ST622 lineages. The novel chimeric strain has a 2.5 Mb ST22 background with a 366 kb chromosomal replacement of ST45 genome origin along the origin and terminus of replication.

the two varying sequence fragments of isolate CD141496 (ST622-2014) within the respective ST22 and ST45 backgrounds were used to investigate the origin of the chimeric strain and to identify the nearest ancestral isolates seen in Figure 4.2, for downstream comparative analyses. The ST22 backbone clusters within a subcluster on a single short branch (Figure 4.2 A) whilst the ST45 fragment also clusters within a larger group of isolates on a very long branch (Figure 4.2 B) due to further genetic differentiation in the form of recombination across the origin of replication.

A. ST22 FRAGMENT   **B.** ST45 FRAGMENT
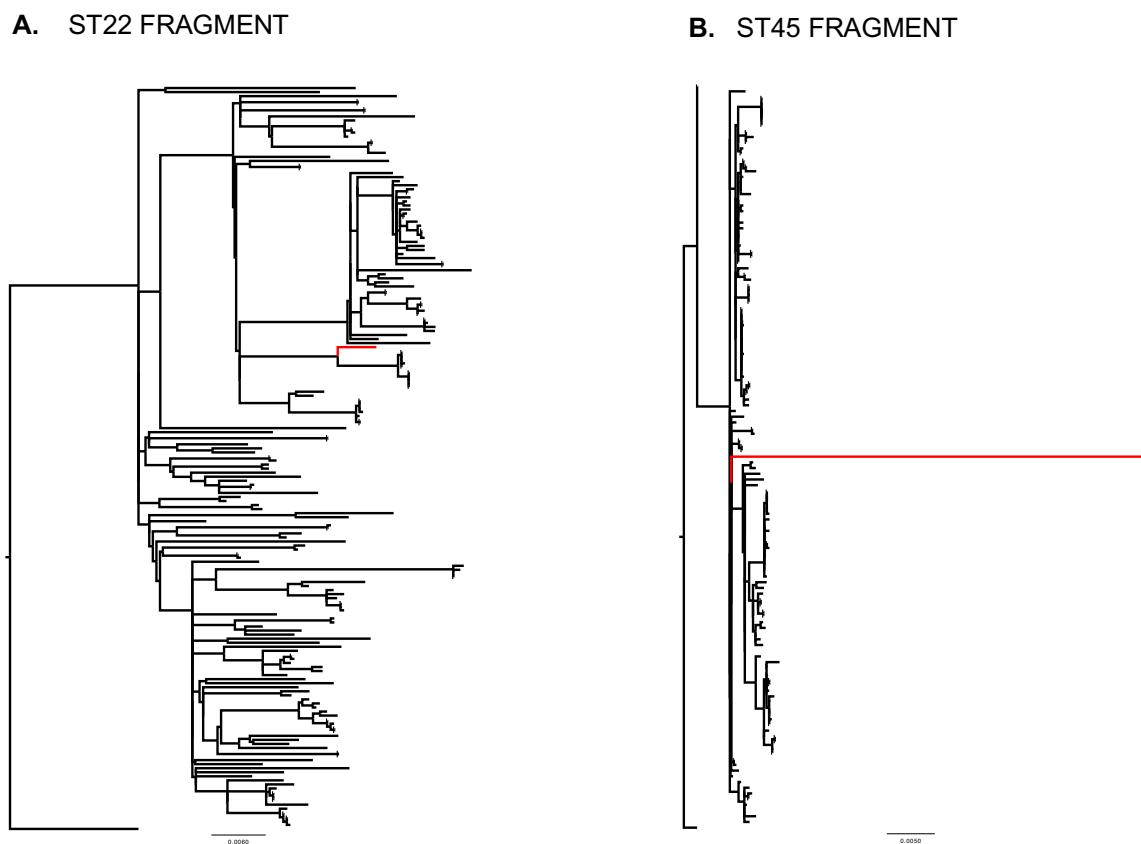
Figure 4.2 | ST622 – CD141496 phylogenetic reconstruction of core genome (A), ST22 backbone (B) and ST45 insert in the context of ST22 and ST45 strains respectively.
ML likelihood reconstruction of both the ST22 and ST45 fragments which make up the ST622 isolate CD141496 were clustered among other isolates within the respective sequence background to identify the closest ancestor from which the large-scale rearrangements originated throughout the Singaporean collection. **A.** Phylogenetic reconstruction of 2.5Mb ST22 backbone fragment of ST622 (red branch) clustering within the same genetic fragment in a population of ST22 strains rooted with reference strain EMRSA15 (**HE681097**). The fragment clusters closely within a subcluster indicating closely related genetic identity. **B.** Phylogenetic reconstruction of 366 kb ST45 fragment of ST622 (red branch) clustering within a population of the same genetic fragment in a population of ST45 background isolates rooted with reference strain CA347 (**CP006044**). The ST45 donor fragment from the ST22 hybrid strain sits on a long branch within the ST45 fragment population, indicating some genetic changes potentially due recombination which the fragment when inserted into the chimeric genome.

During subsequent years of data collection, all ST622 isolates sequenced were a second variant of the chimeric strain, denoted ST622-2015, present in increased numbers: 16 isolates in 2015 and 26 isolates in 2016 (unpublished). ST622-2015 variant regained a 134 kb sequence fragment of the ST22 background downstream of the origin including the SCC*mec* region as seen in Figure 4.3. It is unclear by what means this further chromosomal shuffle occurred, and why the variant persisted within the circulating Singaporean MRSA population.
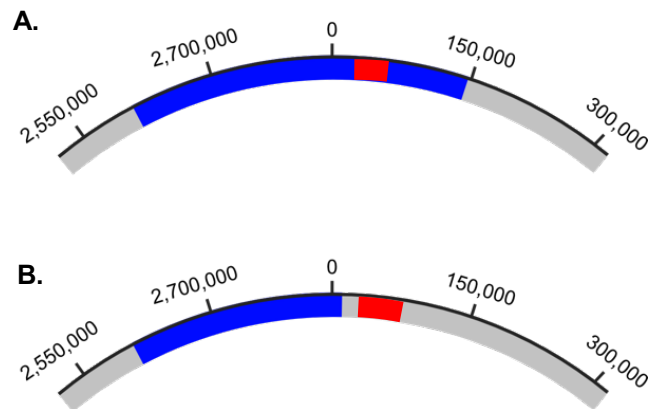
Figure 4.3 | ST622-2014 and ST622-2015 Variant Chromosomal Replacement Maps

A. ST622-2014 variant with 366 kb core genome replacement from ST45 donor into ST22 backbone, spanning the termination and origin of replication – SCC*mec* in red. **B.** ST622-2015 variant with 232 kb genome replacement (ST45 insert into ST22 chromosomal background) throughout the termination and origin of replication - SCC*mec* in red.

Until now, little is known about the mechanisms of genetic exchange resulting such large-scale recombination events in *S. aureus*, and further evolution of differing variants of hybrid strains, especially *in vitro* (Robinson & Enright, 2004; Holden et al., 2010). Even less is known about what biological impact the introduction of large chromosomal replacements have within a new strain. Not only are these strains differentially methylated genomic DNA segments, but also have the potential to introduce new Restriction-Modification system elements into new sequence backgrounds. This may lead to the complete alteration of the existing methylation landscape within the given lineage, potentially also affecting epigenetic gene expression, DNA replication and repair, conjugal transfer and further essential functions (Casadesús & Low, 2006).

This collection of isolates is a snapshot of a very unusual, natural large-scale recombination event. The core genome transfer of the 232 kb ST45 origin fragment in the ST622 isolates allows for the study of the impact of differential methylation patterns (ST22 or ST45 patterns) on an identical sequence region within the novel chimeric ST622 and closely related parent strains. WGS, methylation and transcriptomic analysis of the ST622 strains and selected donor strains, would allow the investigation of how largescale genome rearrangements affect the 6mA methylation and gene expression profiles of the hybrid MRSA strains.

4.2 AIMS & OBJECTIVES

This study encompasses the two ST622 variants and a representative group of phylogenetically closely related natural comparators from ST45 and ST22 MRSA strain backgrounds circulating intermediate and long-term care facilities in Singapore.

The collection of isolates was sequenced with PacBio SMRT technology enabling:
1) Comparative genomic analysis of chimeric ST622 and parent strains (ST45, ST22).
2) Characterization of the RM systems within each lineage.
3) Characterization of 6mA methylation motifs (TRSs) with PacBio SMRT Motif and Modification analysis and corresponding TI Sau1 HsdS (TRDs) using protein homology and secondary structure analysis.
4) Calculation of the overall methylation landscape of isolates and methylation differences resulting from the large-scale recombination in the hybrid ST622 isolates.

The set of isolates were also RNA-Sequenced enabling:
1) The investigation of whole genome transcriptomic profiles of ST622, ST22 and ST45 isolates, as well as the differences between the novel hybrid strain and parent lineages
2) Differential expression (DE) analysis focusing on the identical sequence region of ST45 origin within the ST622 isolates and the ST45 isolates, to investigate potential gene expression effects as a result of differential methylation, and pinpoint specific genes which may be under epigenetic control.

4.3 ORIGINS OF COLLECTION

The isolates used in this study are part of a cross-sectional study (Singapore Collection) investigating the carriage of *Staphylococcus aureus* in interconnected acute (n=2), intermediate-term and long-term (n=3) healthcare facilities Singapore, led by Angela Chow and Dr. Li-Yang Hsu (National University of Singapore) and Professor Matthew Holden. The WGS study data for strains isolated in 2014 can be accessed through accession number PRJEB9390 from the European Nucleotide Archive (http://www.ebi.ac.uk/ena/). WGS data for isolates from 2015 and 2016 are unpublished. A complete list of bacterial strains and DNA sources has been provided in Table 2.6 and Table 2.7 in Methods.

4.4 RESULTS

The isolates included in this study were determined by Professor Matthew Holden through phylogenetic reconstruction of Illumina HiSeq sequenced isolates (Wellcome Sanger Institute) from the Singapore Collection. The collection of strains are representative MRSA strains from the two ST622 variants, ST622-2014 (n=1) and ST622-2015 (n=2) as well as phylogenetically most closely related donor strains, ST45 (n=3) and ST22 (n=3) of the two sequence fragments making up the chimera. The donor strains were included as natural comparators to be able to investigate the genetic, transcriptomic and methylomic changes which mixing ST22 and ST45 genetic components result in the hybrid ST622 strains. The differing sequence fragments and position of chromosomal replacements within the ST622 strains were previously determined via recombination analysis, which were further characterised in the results below.

4.4.1 Characterisation of Singapore Collection Isolates

4.4.1.1 Multi-Locus Sequence Typing (MLST)

Multi-locus sequence typing analysis resolved the Singapore isolates (n=9) into 3 different STs, ST22, ST45 and ST622 as seen in Table 4.1. The ST622 hybrid strains resembled MLST type 22 (ST22), with the chimeric differing only in one of the seven housekeeping genes – *arcC*. The ST622 alleles for *arcC* were identical to that of multi-locus sequence type 45 (ST45), with four single nucleotide polymorphisms (SNPs) in comparison to the ST22 isolates at positions: 156bp (C > T), 333bp (T > A), 486bp (T > A) and 787bp (C > A). The *arcC* allele change within the ST622 is directly related to the recombination event which encompasses the location where the gene is found within the genome.

Table 4.1 | MLST Profiles of Singapore Isolates

Name	ST	*arcC*	*aroE*	*glpF*	*gmk*	*pta*	*tpi*	*yqiL*
EMRSA15	22	**7**	6	1	5	8	8	6
CD140400	22	**7**	6	1	5	8	8	6
CD140638	22	**7**	6	1	5	8	8	6
CD140866	22	**7**	6	1	5	8	8	6
CD150713	622	**10**	6	1	5	8	8	6
CD150916	622	**10**	6	1	5	8	8	6
CD141496	622	**10**	6	1	5	8	8	6
CD140392	45	**10**	14	8	6	10	3	2
CD140901	45	**10**	14	8	6	10	3	2
CD140657	45	**10**	14	8	6	10	3	2
CA-347	45	**10**	14	8	6	10	3	2

4.4.1.2 Genetic Relatedness

The group of ST22 and ST45 donor isolates included in this study were selected according to the phylogenetic relatedness of each fragment to those within the chimeric ST622-2014 strain in terms of sequence similarity as well as gene content. The whole genome similarity between each isolate of the same sequence background were characterised by investigating the core genome SNP differences between core genome alignments generated for each isolate. The ST22 and ST622 isolates were aligned to references strain EMRSA15, whilst the ST45 isolates to CA-347. Mobile genetic elements and regions of highly variable sequence regions were excluded from the alignment, including the 134 kb extended recombined region within the ST622-2014 and ST622-2015 isolates. The number of SNPs between pairwise alignment comparisons within each group of isolates can be seen in Table 4.2.

Table 4.2 | Core Genome SNP Sites between Isolates by ST

ST	Isolate 1	Isolate 2	SNPs	% ID	Aligned bases
ST45	CD140657	CD140392	51	99.99799285	2540910
	CD140657	CD140901	58	99.99977115	2534376
	CD140392	CD140901	82	99.99676618	2535703
ST22	CD140866	CD140400	138	99.99484262	2675779
	CD140866	CD140638	224	99.99163014	2676268
	CD140400	CD140638	342	99.98723647	2679509
ST622	CD141496	CD150916	71	99.99693409	2315784
	CD141496	CD150713	69	99.99702072	2315995
	CD150916	CD150713	3	99.99987048	2316314

The core genome of the ST45 isolates were ~99.998% identical to each other with at most 82 SNP changes separating two isolates, CD140392 vs CD140901. The core genome of the ST22 isolates also showed ~99.991% similarity, with slightly higher number of SNP differences between isolates, with at most 342 SNPs, CD140400 vs CD140638, potentially impacted by the higher number of aligned bases included in the analysis than for the ST45 or ST622 pairwise comparisons. The core genome of the two ST622-2015 isolates (CD150713 vs CD150916) were genetically identical, separated from each other by 3 SNPs, whilst they differed by 69 and 71 SNPs from the ST622-2014 variant.

4.4.1.3 Comparative Genomic Analysis

To gain a better understanding of the molecular difference within the collection of isolates, and what area of the genome the large-scale recombination occurred within the ST622 isolates, *in silico* comparative analysis of the assemblies was conducted using BLAST. The whole genome alignments of the 9 Singaporean isolates and additional reference strains CA-347 (ST45) and EMRSA15 (ST22) are shown in Figure 4.4.
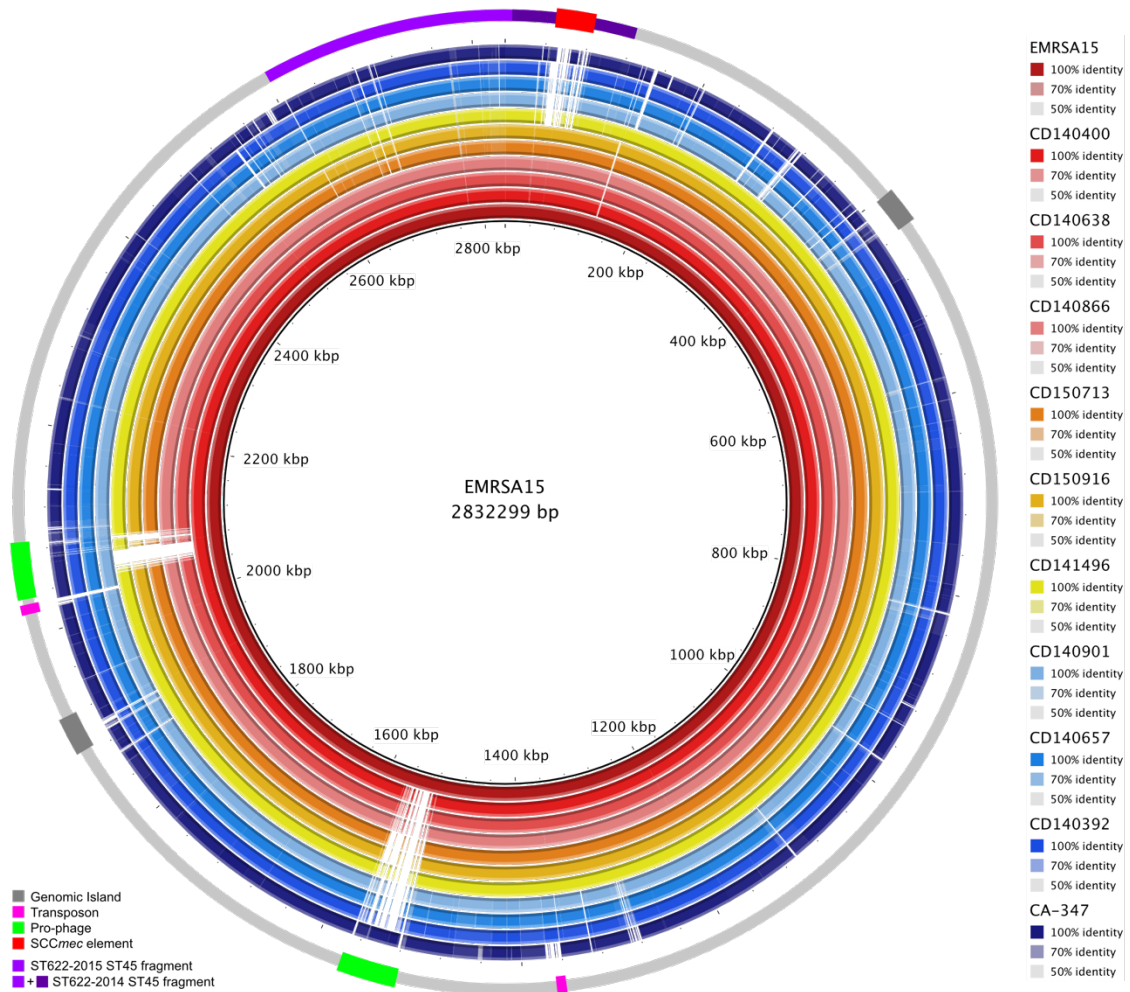
Figure 4.4 | Genomic organisation of Singapore Isolates.
Genome alignments of 9 Singapore isolates (reds: ST22, yellows: ST622, blues: ST45) and reference genomes EMRSA15 and CA-347, where EMRSA15 (HE681097) was used as a reference for mapping represented by the innermost ring in dark red. Isolates from the inside ring (dark red) towards the outside (dark blue): EMRSA15 (HE681097), CD140400, CD140638, CD140866, CD150713, CD150916, CD141496, CD140901, CD140657, CD140392, CA-347 (CP006044)). The outermost ring (grey) shows the mobile genetic elements (MGEs) harboured by the reference strain EMRSA15 including genomic islands (grey), transposons (fuchsia), SCC*mec* (red) and prophage (lime), as well as the position of recombined sequence region of ST45 origin within the ST622-2014 (purple + aubergine) and ST622-2015 variant (purple). Map created with Blast Ring Image Generator (BRIG).

It can be clearly seen that the core chromosome alignments for all isolates (excluding additional plasmids), regardless which sequence type, show >99% sequence identity over the core genome regions, with the only major sequence variation encompassing the MGEs (SCC*mec*, and prophages). This can be explained to be due to the highly clonal nature of *S. aureus,* with 75% of the genome being conserved throughout the species, with mainly all genetic distinction associated with the vastly variable accessory genome composition. This is important as it shows that the recombinant sequence region (ST45 origin) which is present in hybrid ST622 strains encompasses a core genome region, which is homologous to that of ST22.

The total number of bases, GC content, number of CDS and the length of the coding, intergenic, core genome and accessory genome can for each isolate can be seen in Table 4.3. The only major difference between the isolates was the overall size of the assembled genomes, mainly due to the variable size of the accessory genome (ACC). The core genome (CORE) for each isolate stayed around ~2.7 Mb.

Table 4.3 | Whole Genome Lengths and GC content within Singapore Isolates

Isolate	ST	Total Bases (kb)	Total GC Content	Total CDS features	Bases in CDS (kb)	CDS GC Content	Bases in IGR (kb)	IGR GC %	CORE Genome Size (kb)	ACC Size (kb)	ACC GC %
CD140400	22	2882.33	32.81	2661.00	2403.09	33.46	479.25	32.17	2686.86	195.47	30.19
CD140638	22	2846.33	32.78	2611.00	2374.70	33.44	471.63	32.13	2686.26	160.07	30.97
CD140866	22	2873.99	32.82	2662.00	2399.03	33.45	474.96	32.20	2695.11	178.89	30.76
CD150713	622	2804.16	32.76	2543.00	2332.00	33.41	472.16	32.12	2686.86	117.30	29.67
CD150916	622	2808.94	32.74	2537.00	2334.60	33.40	474.34	32.09	2683.07	125.87	29.66
CD141496	622	2884.74	32.73	2660.00	2404.38	33.37	480.36	32.10	2670.91	213.83	30.28
CD140392	45	3014.35	32.73	2837.00	2514.39	33.37	499.96	32.10	2667.95	346.40	30.38
CD140901	45	2931.65	32.85	2707.00	2445.68	33.50	485.97	32.21	2698.99	232.66	30.25
CD140657	45	2922.83	32.83	2693.00	2437.79	33.48	485.05	32.19	2668.67	254.16	30.35

IGR: Intergenic Region, CORE: core genome, ACC: accessory genome

Overall, the ST622 strains isolated in 2015 (CD150713, CD150916) had the smallest accessory genome (117 kb, 126 kb), followed by the ST22 isolates (160-196 kb), and followed by 2014 isolated ST622, CD141496 (214 kb). The ST45 isolates had the largest accessory genome (233-254 kb). The main difference within the accessory genome of the 9 isolates is the composition, type and number of MGEs highlighted in Table 4.5, bringing in a wide range of antimicrobial resistance determinants giving each isolate a unique resistance genotype (Table 4.4).

Table 4.5 | Major Mobile Genetic Elements Harboured within Singapore Isolates

ST22	CD140400		CD140638		CD140866	
MGE Type	genomic region	bp	genomic region	bp	genomic region	bp
SCC*mec* (TIVh)	34117..68500	34383	33840..69966	36126	33919..71183	37264
SaPI3 (*ear, sec-bov, sel*)	425108..439041	13933	426163..440178	14015	427378..441312	13934
ICE element (ICE6013)	1433687..1448593	14906	1429491..1444882	15391	1438510..1452714	14204
Prophage (*StauST398-5*)	1222501..1266082	43581	1218027..1261608	43581	1226851..1270188	43337
Prophage (*φSa3*)	2063597..2108473	44876				
Phage (non-integrated)					2849975..2873951	23976
Transposon (Tn*552*)	2042360..2050867	8507	2037884..2046391	8507	2046432..2054939	8507
Plasmid II (*blaZ,cadAC, arsBC,qacA*)	2852081..2882298	30217	2803737..2831601	27864		
Plasmid III (qacC)			2831602..2846183	14580	2812278..2849945	37667

ST622	CD150713 (2015)		CD150916 (2015)		CD141496 (2014)	
MGE Type	genomic region	bp	genomic region	bp	genomic region	bp
SCC*mec* (TIVh or *TII)	33814..70507	36693	35785..72459	36674	32952..59983	27031
CRISPR element					60012..79911	19899
SaPI3 (*ear, sec-bov,sel*)	426745..440353	13608	428728..442906	14178	420385..434481	14096
SaPI1 (*seb,ear,seq,sek*)					870415..884835	14420
ICE element (ICE6013)	1392232..1405844	13612	1393664..1410514	16850	1443167..1457506	14339
Prophage (TEM123)					892231..934918	42687
Transposon (Tn*552*)	2000118..2008543	8425	2000660..2010463	9803	2051361..2059491	8130
Plasmid II (*blaZ,cadAC, arsBC,qacA*)	2759115. 2804153	44959	2760505..2808865	48360	2849693..2869781	34991
Plasmid (tra-genes, *soj*)					2811458..2849692	38234

ST45	CD140392		CD140901		CD140657	
MGE Type	genomic region	bp	genomic region	bp	genomic region	bp
SCC*mec* (*TII)	33873..71336	37463	1372771..1399030	26259	33608..80056	46448
CRISPR element			1399029..1414501	15472	80332..97125	16793
SaPI3 (*fhuD*)	2094488..2110527	16039	533557..549192	15635	2116939..2132250	15311
ICE element (ICE*6013*)	2291632..2306133	14501	730520..744777	14257	2313447..2327949	14502
Prophage (3 AJ-2017)	868144..911725	43581	2210410..2251480	41070	892564..933390	40826
Prophage (*φL54a-like*)	1339905..1365528	25623	2679752..2706241	26489	1362548..1388144	25596
Prophage (*φSa3*)	2044539..2088524	43985	483253..526264	43011	2066553..2110536	43983
Phage (non-integrated)	2994939..3014319	19360				
Plasmid II (integrated)	2563687..2614962	51275	1003177..1053642	50465	2584649..2635032	50383
Plasmid (tra-complex)	2900131..2948894	48763				
Plasmid (tra-complex)	2948973..2988016	39043				
Plasmid fragment (rep)	2988033..2994839	6806				

Table 4.4 | Resistance Genotypes in Singapore Isolates

Resistance	Gene	CD140400	CD140638	CD140866	CD150713	CD150916	CD141496	CD140392	CD140901	CD140657
Aminoglycosides	*aacA*							●	●	●
	aacA-aphD		●	●	●		●	●	●	●
Macrolides	*ermC*	●	●	●	●	●	●		●	●
Mupirocin	*ileS-2*	●	●			●	●	●	●	●
Methicillin	*mecA*	●	●	●	●	●	●	●	●	●
Penicillin	*blaZ*	●	●	●	●	●	●	●	●	●
Antiseptics	*qacA*		●	●	●	●	●	●	●	●
	qacC	●	●	●	●	●		●		●
Tetracyclines	*tetK*	●	●				●	●	●	●
Fluoroquinolones	*grlA - S80F*	●	●	●	●	●	●	●	●	●
	gyrA - S84L	●	●	●	●	●	●	●	●	●

Key: white block indicates absent gene / modification, red box indicates presence of resistant determinant allele

4.4.1.3.1 Staphylococcal Cassette Chromosome (SCC*mec*) Element

The ST22 isolates and the ST622-2015 isolates CD150713 and CD150916 (isolated in 2015) contained an SCC*mec* element of Type IVh (from the ST22 sequence background). This SCC*mec* element has a class B *mec* gene complex and a type 2 *ccrAB* cluster as seen in Figure 4.5 A. Although ST45 reference strain CA-347 (CP006044) contained an SCC*mec* Type II element (class A *mec* complex, *ccrAB* type 2 – Figure 4.5 B), the staphylococcal cassette chromosome within the Singaporean ST45 isolates and ST622 isolate CD141496 (isolated in 2014), resembled a Type V (5C2&5) SCC*mec*, with 2 *ccrC1* (allele 8, allele 2), *mec* class C2, similar as found in reference strain PM1 (Chlebowicz et al., 2009, ASM30889) (Figure 4.5 C). The non-essential J-region (junkyard) are also different in the two SCC*mec* types, including variable insertion sequences, transposons and inserted CRISPR-*Cas* systems. The full set of genes for both types of SCC*mec* elements can be found in the Appendix under Supplementary Table 8.3, Table 8.4 and Table 8.5.
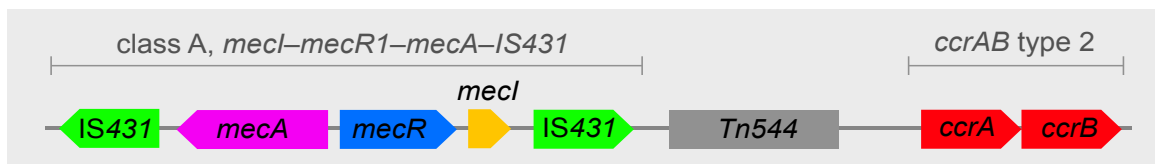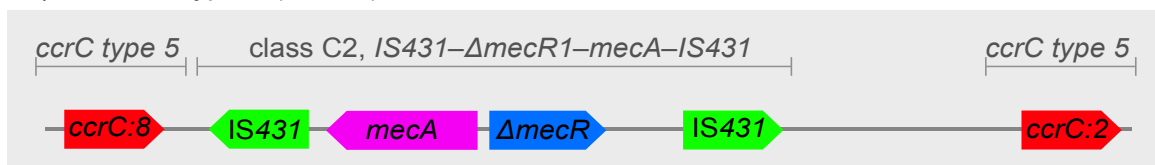
Figure 4.5 | SCC*mec* Types within Singapore Isolates
SCC*mec* types IVh (**A**) type II (**B**) and type V (**C**) from reference strains EMRSA-15 (GE681097), CA-347 (CP006044) and PM1 (ASM30889). **A**. SCC*mec* T IVh holds a *mecA* penicillin binding protein 2 and a fragmented *mecR*, where the penicillin-binding domain is deleted. The *mec*-gene complex is flanked by IS*431* downstream the *mecA* and a fragmented IS1272 downstream the truncated *mecR* gene, followed by a type 2 *ccrAB* complex and junkyard region. **B**. SCC*mec* T II holds a class A *mec*- gene complex including *mecI*, encoding repressor, *mecR1* encoding inducer for *mecA* penicillin binding protein 2. The *mec*-complex in this type of SCC*mec* element is flanked by two IS*431*, and a Tn*544* transposon is also inserted prior to the type 2 *ccrAB* complex, followed by the junkyard region. **C.** SCC*mec* T V (5C2&5) carries 2 *ccrC1* type 5 recombinases, *ccrC1*:8 upstream, and *ccrC1*:2 downstream the IS*431* elements flanking the SCC*mec* complex (*mecA* class C2 and truncated *mecR* gene).

4.4.1.3.2 Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR)

The CRISPR-Cas system presented in this study was of Type III-A previously characterised by Cao et al (2013), only found in the 2 of the ST45 isolates (CD140901 and CD140657) and ST622-2014 isolate CD141496; no ST22 carried a CRISPR loci. The gene cluster found CD140657 and CD141496 contained 3 *cas* genes and 6 *csm* genes whereas isolate CD140901 was missing genes *cas1, cas2* and *csm1*. The *cas1-cas2* complex is recognised as the information processing module, involved in the acquisition of spacer regions for the CRISPR-Cas system (Makarova & Koonin, 2015). This suggests that the CRISPR locus within isolate CD140901 is truncated, yet still functional as it still holds the *csm* effector complex which interfere with either DNA or RNA target sequences (Samai et al., 2015).

4.4.1.3.3 Transposons, IS Elements and ICE Elements

Other than the Tn*544* and IS*431mec* carried as part of the SCC*mec* element, none of the ST45 isolates carried an additional transposable element in the form of Tn*552*, which both ST22 and ST622 isolates carried. Smaller IS elements were not investigated in this study. All isolates carried a larger ICE*6013*, an integrative conjugative element.

4.4.1.3.4 *Staphylococcus aureus* Pathogenicity Island (SaPI)

The ST22 and ST622 isolates contain a SaPI3 (bovine) inserted at the *attS* site prior to νSaα and carry enterotoxin genes: *ear, sec-bov*, and *sel*. CD141496 ST622-2014 contained an additional SaPI1, integrated at *attS* site downstream *metNPQ* operon, carrying enterotoxins: *seb, ear, seq*, and *sek*. The ST45 isolates carried a SaPI3 similar to that carried in Mu50, integrated downstream of *groESL* heat shock protein coding operon. This pathogenicity island carries gene *fhuD*, ferrichrome transport permease, along with uncharacterised pathogenicity island genes, inserting downstream of prophage ϕSa3.

4.4.1.3.5 Prophage

Whilst all ST22 isolates had at least 1 prophage (StauST398-5) and the ST45 isolates at least 3 each (ϕSa3 (*scn, sak, chp*), 3 AJ-2017, ϕL54a-like); out of the ST622 isolates, the only isolate which contained a prophage (TEM*123*) was the 2014 ST622, CD141496.

4.4.1.3.6 Plasmids

All ST45 isolates carried an integrated plasmid 300 kb upstream of the terminus of replication. Of the ST45 strains, only isolate CD140392, carried additional plasmids, both containing the plasmid conjugation transfer *tra*-gene complex linked to *repA* replication initiator. These plasmids did not carry any other virulence determinants. ST622 and ST22 isolates contained Plasmid II with a selection of virulence and resistance genes including: *blaZ* (beta-lactamase resistance), *cadAC* (cadmium resistance), *arsBC* (arsenic resistance), *qacA/qacC* (quaternary ammonium compound resistance). Interestingly the only ST622 isolate that contained a plasmid carrying *tra* genes was CD141496 (ST622-2014).

Overall, the core genomes of the 9 isolates within the Singapore collection are homologous. The main genomic differences are due to variability within the accessory genomes of the isolates, attributed to differences in the number and composition of MGEs. To better understand the evolution and success of the ST622 strains and the genetic variability between the strains isolated from different years, a more stringent comparative analysis was needed.

4.4.2 Characterisation of ST622

As the core genome homology between the isolates was established, the minimum sequence identity cut-off was lowered to 98% to characterise the recombinant sequence region within the ST622 isolates. Artemis Comparison Tool (ACT) was used to illustrate the sequence differences when compared to ST22 reference EMRSA15 as shown in Figure 4.6.
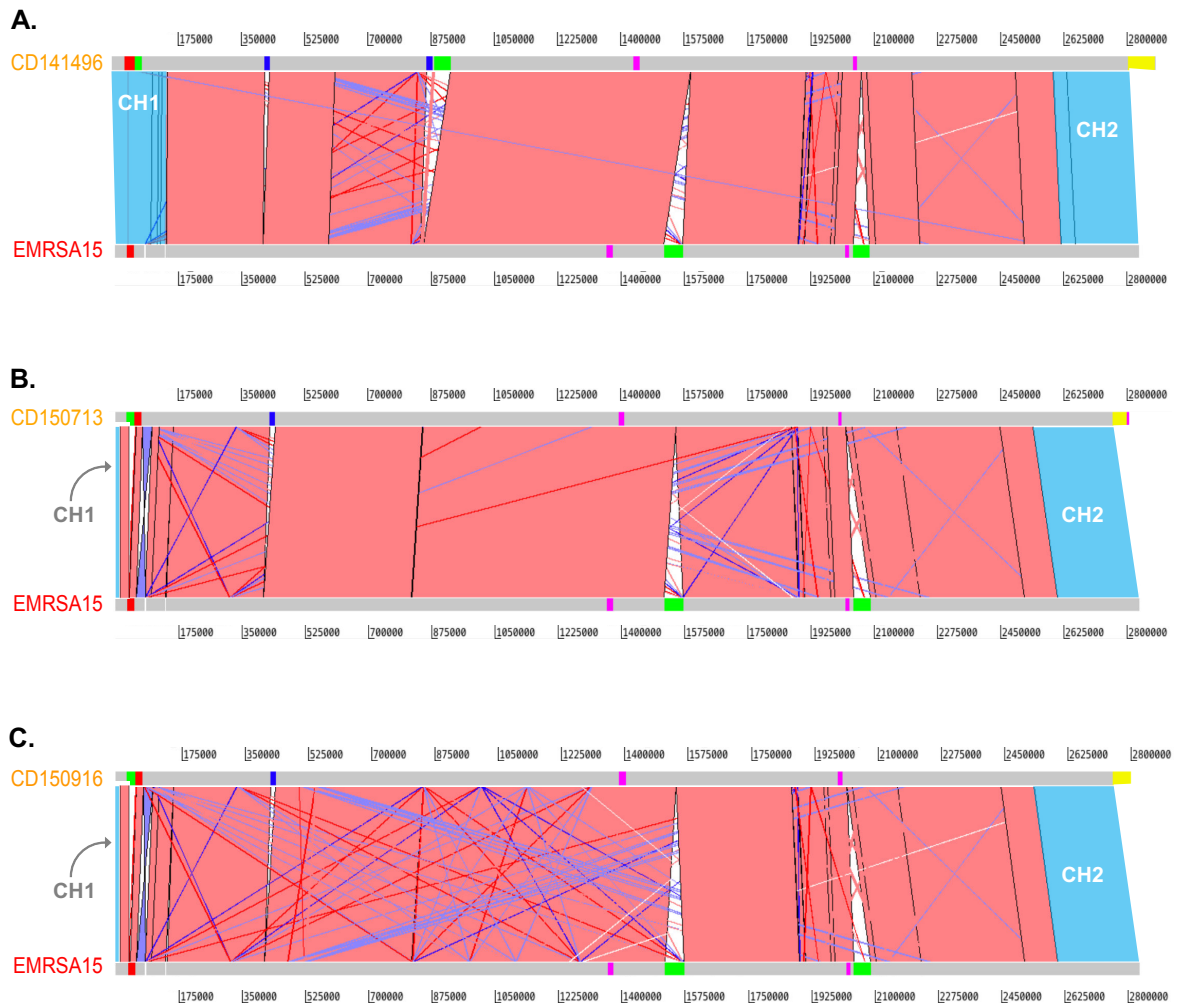
Figure 4.6 | Recombinant Sequence Region in ST622 Isolates
Genome alignments of ST622 isolates **A.** CD141496 (ST622-2014), and ST62-20105 variants **B.** CD150713 and **C.** CD150916 against ST22 reference strain EMRSA15 (GE681097). The red bars show >98% sequence identity between the strains with the dark blue strings representing inversions and the light blue bars the recombinant region of ST45 sequence origin within these strains which still show 97% identity compared to the reference strain sequence. The grey bars show the genome length with detailed positions of mobile genetic elements (MGEs) harboured within each strain including pathogenicity islands (blue), transposons (fuchsia), SCC*mec* (red) prophage (lime), and plasmids (yellow). As the genomes are visualised horizontally starting from the origin of replication, the chimeric fragments are also divided into two segments, one starting from the origin of replication denoted **CH1** (134 kb longer within the ST622-2014 isolates than the 2015 variant as seen in **A.**) and the second starting around 2.6Mb towards the terminus of replication denoted **CH2**.

The recombinant DNA sequence encompasses both the origin of replication and the terminus within the ST622 strains, and the collection of isolates include two different variants of the chimeric strain. CD141496, denoted variant ST622-2014 was isolated from Tan Tock Seng Hospital in 2014 as the single chimeric ST622 strain among 385 *S. aureus* samples. ST622-2014 has a 2.5 Mb ST22 backbone, with a ~376 kb (13%) ST45 DNA fragment replacement spanning both the origin of replication and termination (chimeric fragment 2 (CH2): *dedA → oriT* – 216 kb*; chimeric fragment 1 (CH1): oriC → capD_2* – 156Kb as seen in Figure 4.6 A). The recombination event encompasses the SCC*mec* element, switching the ST22 TIVh SCC*mec* to a ST45 TV (5C2&5) type SCC*mec* element. The second ST622 variant represented by strains CD150713 and CD150916, denoted ST622-2015, were collected in the subsequent year. Both isolates contained a smaller ST45 DNA fragment (~232 kb – 8.25%; (CH2: *dedA → oriT* – 220Kb*; CH1: oriC → hutH* – 12 kb as seen in Figure 4.6 B and C), regaining a 134 kb core genome segment of an ST22 origin. The relative size and CDS count for each isolate are shown in Table 4.6.

Table 4.6 | Size of Recombinant Sequence Region and CDS Count in ST622 isolates

Isolate	Chimera Segment 1 (CH1)*		Chimera Segment 2 (CH2)**		Total Chimera Segment	
	CH1 Size (bp)	# Genes in CH1	CH2 Size (bp)	# Genes in CH2	Total CH Size (bp)*	# Genes in Total CH
CD141496	156901	141	221733	206	378634	347
CD150916	11892	8	219952	195	231844	203
CD150713	11892	8	220330	195	232222	203

* from the origin of replication to *hutH* (ST622-2015) or *capD_1* (ST622-2014)

** from 2.5 Mb towards the terminus of replication

The genetic composition of the 232 kb recombined region within both ST622 strains are identical, and furthermore, they are homologous within the ST45 and ST22 strains presented in this collection with only a few CDS differences. The full list locus tags of CDS within the recombinant sequence and equivalent genetic regions within the ST22 and ST45 isolates can be found in the Appendix - Supplementary Table 8.2 and Table 8.3.

There were a handful of gene differences between the ST622 variants, and their parent STs highlighted in Table 4.7 and Table 4.8. Most of these genes were putative cytosolic proteins, transposases, hypothetical proteins and a couple proteins with metabolic activity (ST22): acetyltransferase, alcohol dehydrogenase, *mpr* - serine peptidase. One interesting difference is the absence and presence of cell-wall-anchored surface protein *sasD* (CH1 - ST45 background) and putative *sasK* (CH2 - ST22 background).

Table 4.7 | Variable Genes within Chimeric Region 1 (CH1) Starting from the Origin of Replication (without SCC*mec*)

Gene/Product	CD140392 Locus	CD140901 Locus	CD140657 Locus	CD141496 Locus	CD150713 Locus	CD150916 Locus	CD140400 Locus	CD140638 Locus	CD140866 Locus
staphylococcal tandem lipoprotein	58275_C01_00089	58275_A01_01437		58366_B01_00096					
putative cytosolic protein	58275_C01_00090	58275_A01_01438	58275_B01_00114	58366_B01_00097					
putative cytosolic protein	58275_C01_00091	58275_A01_01439	58275_B01_00115	58366_B01_00098					
putative cytosolic protein	58275_C01_00092	58275_A01_01440	58275_B01_00116	58366_B01_00099					
hypothetical protein: FIG01107877					58366_D01_00117	58366_C01_00119	58275_D01_00117	58275_E01_00116	58366_A01_00116
sasD - surface protein	58275_C01_00122	58275_A01_01470	58275_B01_00146	58366_B01_00129					
transposase IS*256*		58275_A01_01472							
transposase IS*605*/IS*200*					58366_D01_00139	58366_C01_00140	58275_D01_00138	58275_E01_00137	58366_A01_00137

Table 4.8 | Variable Genes within Chimeric Region 2 (CH2) Towards the Origin of Termination (without plasmids)

Gene/Product	CD140392 Locus	CD140901 Locus	CD140657 Locus	CD141496 Locus	CD150713 Locus	CD150916 Locus	CD140400 Locus	CD140638 Locus	CD140866 Locus
yvmA - drug transporter							58275_D01_02569	58275_E01_02497	58366_A01_02511
yvnA - transcriptional repressor							58275_D01_02570	58275_E01_02498	58366_A01_02512
transposase IS*1272*							58275_D01_02596	58275_E01_02524	58366_A01_02538
transposase IS*1272*							58275_D01_02597	58275_E01_02525	58366_A01_02539
transposase IS*1272*							58275_D01_02598	58275_E01_02526	58366_A01_02540
transposase							58275_D01_02599	58275_E01_02527	58366_A01_02541
acetyltransferase							58275_D01_02630	58275_E01_02558	58366_A01_02572
alcohol dehydrogenase							58275_D01_02643	58275_E01_02571	58366_A01_02585
cell wall anchored protein (*sasK*)							58275_D01_02657	58275_E01_02585	58366_A01_02599
mpr - serine peptidase							58275_D01_02658	58275_E01_02586	58366_A01_02600
transposase IS*605*/IS*200*							58275_D01_02664	58275_E01_02592	58366_A01_02606
hypothetical protein	58275_C01_02729						58275_D01_02686	58275_E01_02614	58366_A01_02628
hypothetical protein	58275_C01_02768	58275_A01_01295	58275_B01_02784	58366_B01_02671	58366_D01_02580	58366_C01_02581			
hypothetical protein				58366_B01_02684	58366_D01_02593	58366_C01_02594			
hypothetical protein							58275_D01_02748	58275_E01_02676	58366_A01_02690

4.4.3 Restriction-Modification Systems Within Singapore Isolates

One of the main questions within this study focuses on whether or not a large-scale genetic rearrangement has an effect on the 6mA methylation landscape and ultimately the resulting gene expression profile of the hybrid strains. As that the genetic landscapes of the chimeric variants have been described in the previous section, the next steps were to investigate the methylome of each isolate.

DNA methylation within prokaryotes is facilitated through methyltransferase enzymes as part of a restriction modification (RM) system or a solidary methylase. These enzymes modify adenine and cytosine nucleotides within specific sequence patterns called methylation motifs. As seen in the previous chapter, on a species-wide level, *S. aureus* carries a variety of RM system types including TI, TII, and TIV, as well as solitary methyltransferases on some occasion. To help locate and characterise the variability between RM systems within the Singapore collection, a BLAST search was run with a database of *S. aureus* RM units collected from Rebase.

The Singapore isolates held units from all three types of RM systems, detailed in Figure 4.7. For a more streamlined representation of the relatedness between isolates (including reference strain EMRSA-15 (GE681097), CA-347 (CP006044), a maximum likelihood phylogeny was generated and visualised with iTOL.
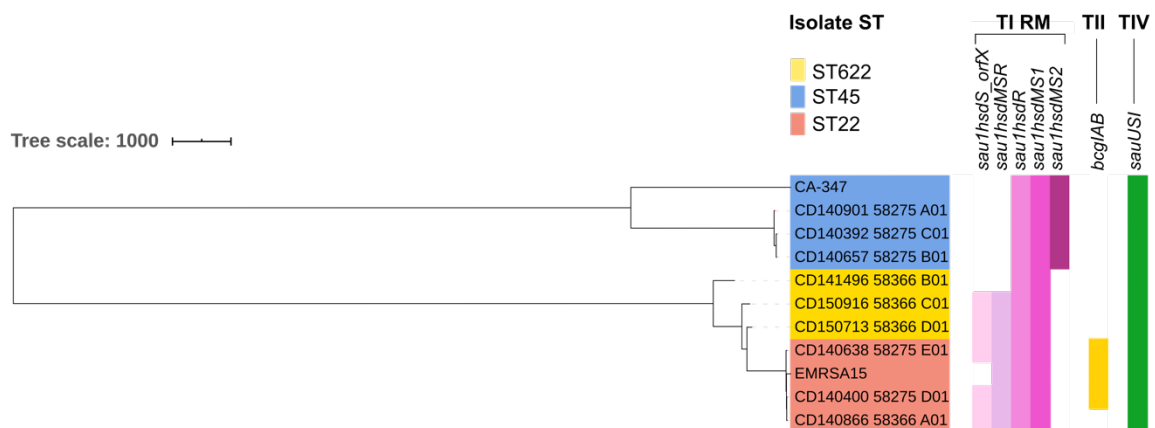
Figure 4.7 | Restriction Modification System Types within Singapore Isolates
Maximum-likelihood phylogenetic tree with annotated presence-or-absence of restriction modification system units represented by coloured columns for each isolate. TI RM *sau1* represented by the pink colours including two accessory *sau1* systems: *sau1hsdS_orfX* and *sau1hsdMSR*, and 'core' *sau1* units including: *hsdR*, and two sets of *sau1hsdMS*, one in each genomic island (*sau1hsdMS1* (vSaα) and *sau1hsdMS2* (vSaβ). TII RM *bcgIAB* system represented by the gold colour, and TIV RM *sauUSI* (*srmB*) restriction endonuclease in green.

4.4.3.1 TI RM System Elements

The Singaporean isolates have a *sau1* TI RM system with 6-methyl-adenine (6mA) restriction and modification activity. In the previous chapter, several other 'accessory' *sau1* elements were described, including SCC*mec* associated *hsdS_orfX* and *sau1hsdMSR*, both present within this collection.

The ST45 isolates harboured the classical *sau1hsdMS1* (vSaα) and *sau1hsdMS2* (vSaβ), whilst the ST22 and ST622 isolates only carry *sau1hsdMS1* (vSaα). ST45 CD140657 also had an *hsdR* fragment and *hsdS* fragment inserted upstream of the CRISPR-Cas system; neither code for functioning proteins. The ST22 background has a truncated *hsdS* remnant within (vSaβ) but this element is non-functional. Both the ST22 and ST622-2015 (CD150713 and CD150916) carried SCC*mec* associated accessory associated *hsdS_orfX* and *sau1hsdMSR*. The 2014 ST622 (CD141494) variant only carried one functional *hsdMS* unit (*sau1hsdMS1* (vSaα)), as the recombinant sequence, from ST45 donor, stretches downstream of the SCC*mec* element. The locus tags for each CDS coding for each *sau1* unit can be found in the Appendix under Supplementary Table 8.6).

4.4.3.2 TII & TIV RM System Elements

Although all 3 Singaporean ST22 isolates, carried prophage StauST398-5, only 2 carried TIIG *bcgIAB*. This system has 6mA activity with a fused RM unit (*bcgIB*) and a DNA binding specificity unit (*bcgIA*). All Singaporean isolates also contained TIV restriction element *sauUSI* (annotated *srmB*). This restriction endonuclease is promiscuous, being non-specific for 6mA, 4mC, 5mC and even hydroxy-methylated or glycosyl-hydroxy-methylated cleave sites. The locus tags for each CDS coding for each *bcgIAB* unit and *sauUSI* unit can be found in the Appendix under Supplementary Table 8.7).

4.4.4 *S. aureus* TI Sau1 6mA Methylation Motifs

4.4.4.1 TI RM *sau1* 6mA Methylation Signatures – Corresponding TRD and TRS

DNA methylation occurs within specific nucleotide sequence motifs. PacBio SMRT sequencing technology has the ability to detect and identify modified/methylated nucleotides and the sequence pattern which they are found in, constituting a methylation motif. The TRS making up each methylation motif are determined by the variable TRDs of each HsdS. To be able to match the predicted 6mA methylation motifs to a specific HsdS, the protein sequences of each specificity unit were analysed. The amino acid homology was compared to the augmented database of TRDs described in Chapter 3, to verify the TRS matched to the complementary TRDs of each specificity unit.

The characterised TRDs, with their matched TRS can be seen in Table 4.9. The ST22 and ST622-2015 variants had 3 functioning *hsdS*, (1 core, 2 accessory) producing 3 methylation motifs. The TRDs and combinations seen for the ST22 and ST622 were identical to the ST22 isolates previously characterised within the NCTC Collection investigated in Chapter 3. The ST622-2014 variant only produced 1 methylation motif, for the single *hsdS* coded in the vSaα. The ST45 isolates produced 2 methylated motifs, matched to *sau1hsdS*α and *sau1hsdS*β. The TRD combinations for the two proteins coded within the ST45 isolates matched those identified by Cooper et al., (2017).

Table 4.9 | HsdS Specificity Unit TRD & TRS for Singapore Isolates

HsdS	ST	TRD1	TRD2	TRS (→ **5'** - 3') F	TRS (←**5'** - 3') R
HsdS_α - *sau1hsdSα*	22, 622	B	I	**A**GG (N)$_6$ TGAR	**YTCA** (N)$_6$ CCT
	45	C	L	**GWAG** (N)$_6$ TAAA	**TTTA** (N)$_6$ CTWC
HsdS_β - *sau1hsdSβ*	45	W	J	**CRAA** (N)$_7$ TCC	**GGA** (N)$_7$ TTYG
HsdS_X - *sauhsdS_orfX*	22, 622-2015	NT1*A	a*	**TAAG** (N)$_6$ TTC	**GAA** (N)$_6$ CTTA
HsdS_S - *sau1hsdS_SCC*	22, 622-2015	NT1*X	NT2*O	**GAAG** (N)$_5$ TAC	**GTA** (N)$_5$ CTTC

After each specificity unit was matched to the corresponding methylation motif, the frequency at which they appeared throughout the host genomes, and the potential differences in whole genome methylation resulting from the largescale genome rearrangement was analysed.

4.4.5 TI Sau1 6mA Methylation Landscape – Singapore Isolates

4.4.5.1 Whole Genome Methylation Overview – Singapore Collection

One of the main questions of this chapter is whether the large-scale recombination events within the ST622 isolates had an impact on the global methylation of the whole genome. Overall, 98.75% of the detected methylation motifs within the 9 isolates were methylated (calculated from RAW data found in Supplementary Table 8.8 in the Appendix). Summary statistics of the average number of motifs for each half string TRS found throughout the genome of the isolates are detailed in Table 4.10.

Table 4.10 | Average Motif Numbers for 6mA Methylation Motifs for Singapore Isolates

Average Number of Motifs	Methylated	Detected	Methylated/Detected	Mean IPD Ratio Range
ST22 Isolates				
GAAGNNNNNTAC	261	261	1.000	6.461-6.946
GTANNNNNCTTC	260	261	0.996	5.426-5.846
TAAGNNNNNNTTC	434	434	0.999	6.307-6.775
GAANNNNNNCTTA	424	434	0.978	4.854-5.143
YTCANNNNNNCCT	682	691	0.987	5.366-5.689
AGGNNNNNNTGAR	679	691	0.983	5.247-5.670
Average/Isolate*	1370	1386	0.989	4.854-6.775
ST622 Isolates				
TAAGNNNNNNTTC	427	427	1.000	6.501-6.630
GAANNNNNNCTTA	421	427	0.987	4.835
GAAGNNNNNTAC	259	259	1.000	6.561-6.775
GTANNNNNCTTC	258	259	0.996	5.665-5.675
YTCANNNNNNCCT	680	688	0.989	5.456-5.665
AGGNNNNNNTGAR	677	688	0.984	5.303-5.559
Average/Isolate**	1359	1371	0.990	4.835-6.775
ST45 Isolates				
TTTANNNNNNCTWC	562	570	0.986	5.710-5.844
GWAGNNNNNNTAAA	561	570	0.984	6.366-6.664
GGANNNNNNNTTYG	372	381	0.976	4.961-5.180
CRAANNNNNNNTCC	371	380	0.978	5.037-5.042
Average/Isolate*	950	968	0.982	4.961-6.664

* Averages were calculated from RAW DATA (Supplementary Table 8.8),
** Average calculated without CD141496 which only had one methylation motif (YTCAN$_6$AGG)

The ST22 carried three functioning *hsdS*, resulting in 3 methylation motifs, with an average sum of 1386 motifs detected for each isolate of which only 1.15% (16 motifs) were unmethylated. Similarly, the ST622-2015 isolates also carried the same three *hsdS* units, resulting in the same methylation signature and average sum of motifs detected (n=1371) and methylated (n=1359 – 0.81% unmethylated (11 motifs). CD141496 (ST622-2014) on carried *sau1hsdSα*, with one accompanying methylation motif $YTCAN_6AGG$, which was detected 692 times throughout the genome, methylated at 98.55% leaving 10 motifs unmethylated. The ST45 isolates carried the two core *hsdMS* systems resulting in two methylation motifs, with an average of 968 motifs of which 98.14% (950) were modified. The average IPD Ratio range for all motifs was between 4.835-6.775 agreeing with the median IPD ratio calculated for the species in the previous chapter (Table 3.4).

Each methylation motif was detected at a different rate throughout the genomes. To gain a better understanding of the differential detected motif frequency produced by the variable *hsdS* within each isolate, the frequency of detected double stranded methylation signatures were plotted against the total detected motif frequency of each isolate (Figure 4.8). The highest total TRS frequency was seen for the ST22 and ST622-2015 isolates containing three *sau1* methylation systems with an average total TRS match frequency of 0.485 TRS/kb (± 0.003 TRS/kb; ±13 TRS matches), followed by the ST45s with 0.322 total TRS/kb (± 0.002 TRS/kb; ± 15 TRS matches), and lastly the ST622-2014 isolate with 0.240 total TRS/kb as it only carried one functioning *hsdS*.

The average TRS match frequency for the motif produced by HsdS_α (A) for the ST22 and ST622 isolates is 0.242 TRS/kb (± 0.002 TRS/kb; ± 5 TRS matches) 20% higher than the A motif found within ST45, 0.193 TRS/kb (± 0.0004 TRS/kb; ± 11 TRS matches). Only the ST45 carried a HsdS_β (B), having TRS binding sites detected 0.129 kb (± 0.002 TRS/kb; ± 5 TRS matches). Only the ST22 and ST622-2015 variant isolates contained motifs for HsdS_SCC (S) and HsdS_*orfX* (X) with average match frequencies of 0.0915 TRS/kb (±0.001 TRS/kb; ± 3 TRS matches) and 0.152 TRS/kb (±0.002 TRS/kb ± 6 TRS matches).
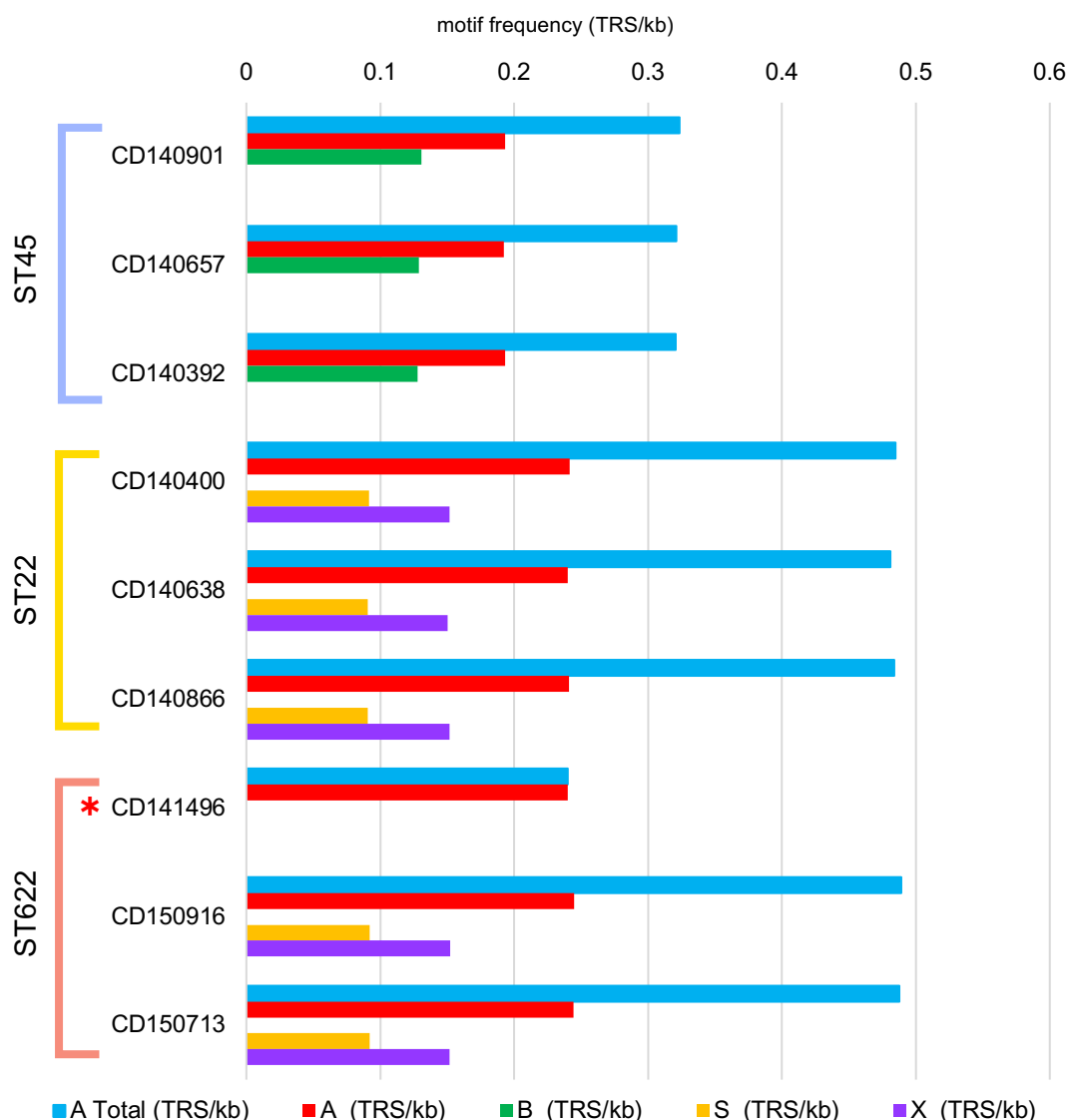
Figure 4.8 | Average TRS Frequency (TRS/kb) within genome of Singaporean isolates.
Grouped according to ST type – ST45 (top), ST22 (middle), and ST622 (bottom). The total frequency of *sau1* 6mA motifs throughout each genome are represented in sky blue. TRS only recognised by HsdS_α (A) are shown in red, motifs recognised by HsdS_β (B) in green, motifs recognised by HsdS_SCC (S) in gold, and motifs recognised by HsdS_*orfX*(X) are shown in purple. ST622-2014 strain marked by red asterisk (✱).

To distinguish between the slight variation seen within the total number of TRS detected throughout the whole genome of each isolate, the methylation motif frequency produced by each HsdS was investigated within the coding sequence (CDS) and intergenic region (IGR) seen in Figure 4.9. Potential changes to the number and position of methylation motifs within the intergenic region, specifically within the promoter regions of genes could have latent epigenetic regulatory effects.
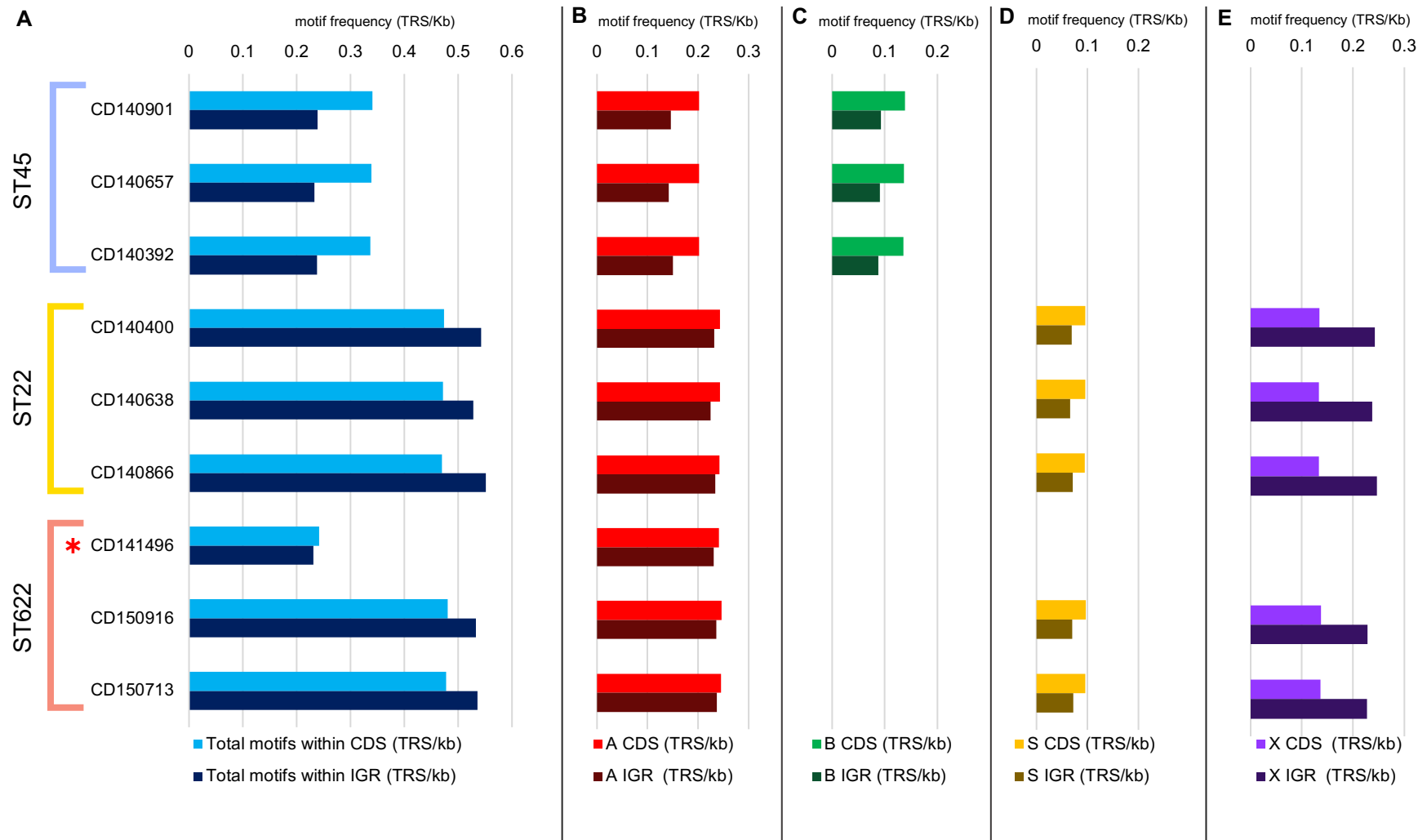
Figure 4.9 | Average TRS Frequency (TRS/kb) within the Coding Sequence (CDS) and Intergenic Region (IGR).
ST622-2014 strain marked by red asterisk (✱). **A**. TRS Frequency within the CDS and IGR for the Total detected motifs within an isolate (blue), **B**. TRS Frequency of HsdS_α (A) motifs (red), **C**. TRS Frequency of HsdS_β (B) motifs (green), **D**. TRS Frequency of HsdS_SCC (S) motifs (gold), **E**. TRS Frequency of HsdS_*orfX* (X) motifs (purple).

The average total number of motif matches within the CDS and IGR for the ST22 and ST622-2015 isolates were 0.475 TRS/kb (±0.004 TRS/kb) with ± 9 motif variation between each strain (Figure 4.9 A). The density at which motifs were found within the intergenic region for the same strains, was 11.3% higher than the CDS region, with an average of 0.538 TRS/kb (±0.009 TRS/kb = ±5 detected motifs). The increased total motif matches within the IGR region of the ST22 background isolates are greatly influenced by the 3:2 CDS to IGR motif ratio for the TRS corresponding to accessory HsdS_*orfX* (X) (Figure 4.9E), as motif matches for the CDS region are slightly higher than those found in the IGR for both HsdS_α (A) and HsdS_SCC (S) (Figure 4.9 B and D). The average total number of motif matches within the ST45 isolates were 30% higher in the CDS (0.338 ± 0.002 TRS/kb) than IGR (0.237 ± 0.003 TRS/kb). The same trend can be seen for the average motif matches corresponding to HsdS_α (A) and HsdS_β (B) seen in Figure 4.9 B and C.

It appears that the introduction of the ST45 DNA segment into the chimeric ST622 strains does not alter the overall methylation frequency of any *sau1* systems. The frequency of methylation by the HsdS_α present in both ST622 variants, as well as HsdS_X and HsdS_S present in only the ST622-2015 variant, were approximately identical to the ST22 isolates.

4.4.5.2 *In situ* Analysis - Global Sequence Methylation – Singapore Collection

Does the introduction of DNA sequence from a different lineage have an effect on the methylation potential of Sau1 through by altering the nucleotide base sequence? In the case of the ST622 isolates, the introduction of the ST45 sequence segment into the ST22 backbone may have introduced distinct polymorphism, especially important considering the intergenic regions harbouring transcriptional regulatory regions. SNPs within promoters, around the transcriptional start and termination sites alter the nucleotide sequence and may impact the presence of methylation motifs (Shell et al., 2013). Loss of 6mA signatures within these regulatory regions may result in loss of epigenetic control of certain genes as explained in Figure 4.10. To gain a better understanding of the variability of methylation signatures, an *in situ* experiment was conducted investigating all *sau1* methylation events across all Singapore isolate, running each of the 5 motifs present in the study within all 3 sequence backgrounds. This approach would ideally show any 6mA modification differences which sequence variation of each ST background may induce within the mostly conserved core genome of *S. aureus*.
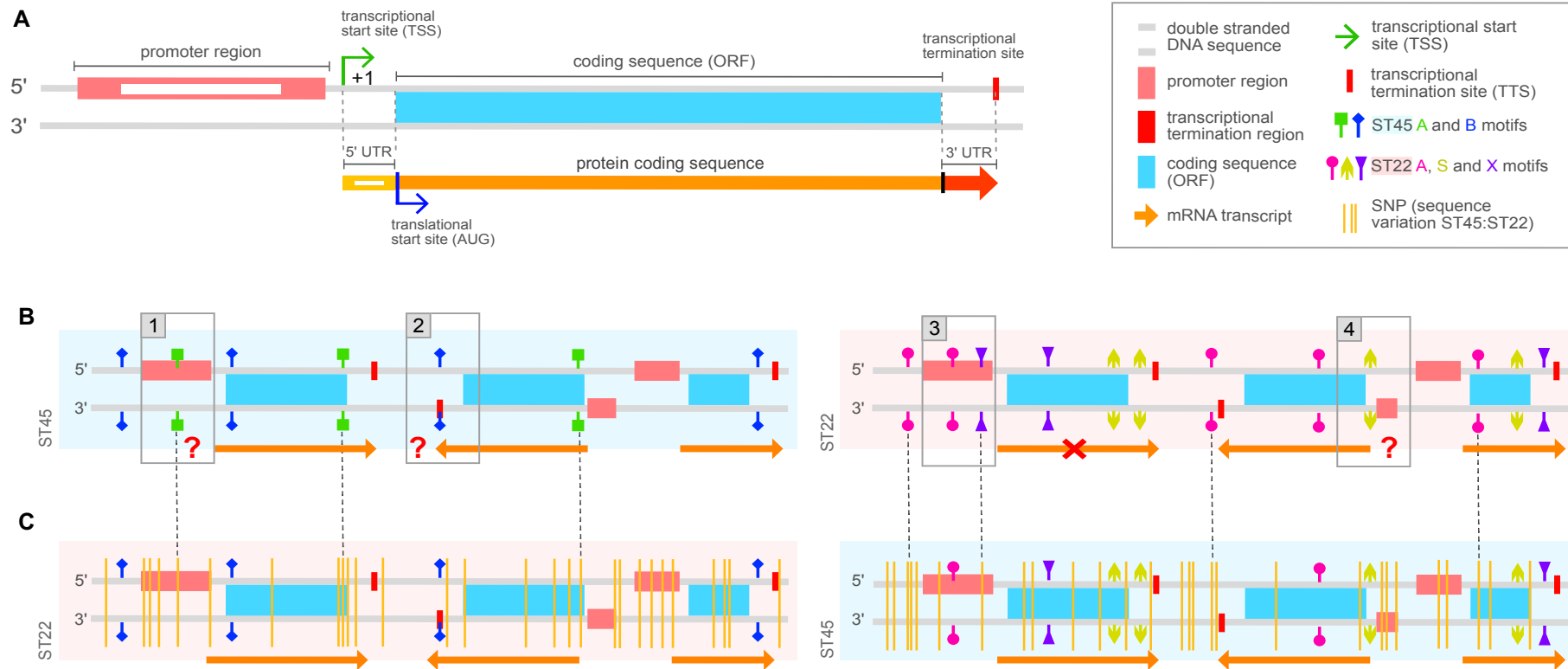
Figure 4.10 | Prokaryotic transcription and potential hinderances due to 6mA methylation. **A.** Bacterial transcription mechanism, on double stranded DNA coding from the 5' to 3' end, or forward strand. The coding strand holds the promoter (salmon) for the downstream gene. The promoter region contains the RNA polymerase binding sites (white rectangle), followed by the transcriptional start site (green arrow) from which transcript for the coding ORF (sky blue) in question is transcribed. Downstream of the CDS lies the transcriptional termination regions which induce either intrinsic termination (rho-independent) or rho-dependent termination of DNA transcription. The resulting mRNA transcript (orange) is headed by the 5' untranslated region (UTR) holding the ribosomal binding sites for translation of the protein coding region and terminating with the 3' UTR region which regulate the mRNA stability, localisation and translation (Mayr, 2017). **B**. *Sau1* 6mA methylation throughout ST45 (light blue – left) background and ST22 (pink – right) background. Both sequence backgrounds have differing HsdS and TRS motifs, hence the position and frequency at which 6mA occur throughout the genome is variable as portrayed by the blue (HsdS_B) and green (HsdS_A) methylation events within the ST45 sequence background and the magenta (HsdS_A), lime (HsdS_S) and purple (HsdS_X) methylation events in ST22/622 sequence backgrounds. Potential epigenetic regulatory effects on translation in for form of 1. 6mA of the promoter potentially hindering/delaying binding of RNA polymerase; 2. 6mA methylation of transcriptional termination region could potentially hinder/delay detaching of RNA polymerase, decreasing mRNA concentrations; 3. Full blockage of RNA polymerase binding to the promoter region, obstructing transcription; 4. 6mA methylation in close proximity to the transcriptional start site (TSS) could impede start of transcription by RNA polymerase. **C**. Left: ST45 motifs investigated within an ST22 sequence background. Gold lines signify single nucleotide polymorphisms (SNPs) to represent the sequence variation between the two sequence types (as compared with the sequence represented above in the left-hand side of part B. The SNPs alter the nucleotide sequence, hence at some localised positions may alter the sequence at which the HsdS of the *Sau1* methylation complex would bind to. This results in loss of methylation in some positions as seen with the absent HsdS_B (green) methylation. Right: ST22 motifs investigated within the ST45 sequence background. As for the left side, SNPs within an alternative sequence background will cause disruption within nucleotide order, so that HsdS cannot recognise the TRS, subsequently fewer adenines are modified.

174

The results for this investigation are summarised in Table 4.11, variations highlighted in yellow. To normalise the data set, the accessory genome regions of each isolate were excluded from this investigation as these were variable. From previous analysis presented in Chapter 3, it is certain that each MGE is differentially methylated when compared to the core genome. Overall, when running the ST45 TRS pattern for the three ST types, both the absolute motif numbers detected and the normalised TRS/kb for the Total ST45 6mA (core HsdS_A & HsdS_B) were very similar for the core genome of each isolate (0.325 ± 0.001 TRS/kb) highlighted in yellow. Although both HsdS_A motif and HsdS_B were present at equal frequencies through the three STs, 0.192 ± 0.001 TRS/kb and 0.133 ± 0.001 TRS/kb respectively, when breaking the core genome matches down into the motifs detected in the CDS and IGR region, there were slight differences. The more significant variance is in the CORE IGR matches, where for both motifs, the ST45 isolates had slightly lower motif frequencies (HsdS_A: 0.256± 0.004; HsdS_B: 0.119 ± 0.003) than the ST22 (HsdS_A: 0.275± 0.002; HsdS_B: 0.127 ± 0.001) and ST622 (HsdS_A: 0.278± 0.005; HsdS_B: 0.129 ± 0.001) isolates.

Overall, the total ST22 motifs detected within the isolates differed by ±30 motifs over the entire genome, with the ST45 isolates having the lowest average CORE matches of 0.330 ± 0.001 TRS/kb, followed by the ST22 isolates (0.337 ± 0.001 TRS/kb) and ST622 isolates (0.339 ± 0.001 TRS/kb). The same trend can be seen for the total HsdS_A and HsdS_S motif individually, showing variation due to the lower motif matches within the CORE IGR region of the ST45 isolates. However, the ST45 isolates have a slightly higher motif frequency in the CORE IGR for the HsdS_X associated motif (0.290 ± 0.002 TRS/kb) than the ST 22 (0.285 ± 0.001) or ST622 (0.277 ± 0.002 TRS/kb) isolates.

Although there are slight differences in motif frequency between STs for all HsdS motifs within the Singapore study, none were significant (STDEV all > 0.001). The variation seen in the motif matches within the intergenic region of each genome could be due to changes or switches between the IGR as suggested by Thorpe et al., 2018. No region of the core genome had substantial sequence differentiation between the three STs, hence the overall methylation frequencies were not impacted considerably.

Table 4.11 | Detected ST45 and ST22 Motifs and TRS/kb Motifs within All Singapore Isolates

ST	ST45			ST22			ST622				
Isolate	CD140901	CD140657	CD140392	CD140400	CD140638	CD140866	CD141496*	CD150916	CD150713	MEAN	STDEVA
TOTAL ST45 6mA MOTIFS IN											
WHOLE	946	938	967	920	913	934	924	904	904	927.778	20.729
WHOLE (TRS/kb)	0.323	0.321	0.321	0.319	0.321	0.325	0.32	0.322	0.322	0.322	0.002
CORE	881	864	868	876	874	878	861	871	877	872.222	6.741
CORE (TRS/kb)	0.326	0.324	0.325	0.326	0.325	0.326	0.322	0.325	0.326	0.325	0.001
ACC	65	74	99	44	39	56	63	33	27	55.556	22.65
ACC (TRS/kb)	0.279	0.291	0.286	0.225	0.244	0.313	0.295	0.262	0.23	0.269	0.031
ST45 HsdS_A motif (GWAGNNNNNNTAAA/TTTANNNNNNCTWC) in											
WHOLE	563	562	582	543	538	555	553	534	535	551.667	15.89
WHOLE (TRS/kb)	0.192	0.192	0.193	0.188	0.189	0.193	0.192	0.19	0.191	0.191	0.002
ACC	47	53	70	26	23	37	43	17	16	36.889	18.162
ACC (TRS/kb)	0.202	0.209	0.202	0.133	0.144	0.207	0.201	0.135	0.136	0.174	0.036
CORE	516	509	512	517	515	518	510	517	519	514.778	3.598
CORE (TRS/kb)	0.191	0.191	0.192	0.192	0.192	0.192	0.191	0.193	0.193	0.192	0.001
CORE CDS	400	396	396	393	392	395	388	391	393	393.778	3.456
CORE CDS (TRS/kb)	0.178	0.178	0.178	0.176	0.175	0.176	0.174	0.175	0.176	0.176	0.002
CORE IGR	116	113	116	124	123	123	122	126	126	121	4.77
CORE IGR (TRS/kb)	0.256	0.252	0.261	0.276	0.277	0.273	0.273	0.281	0.281	0.270	0.011
ST45 HsdS_B motif (CRAANNNNNNNTCC/GGANNNNNNNTTYG) in											
WHOLE	383	376	385	377	375	379	371	370	369	376.111	5.6
WHOLE (TRS/kb)	0.131	0.129	0.128	0.131	0.132	0.132	0.129	0.132	0.132	0.13	0.002
ACC	18	21	29	18	16	19	20	16	11	18.667	4.848
ACC (TRS/kb)	0.077	0.083	0.084	0.092	0.1	0.106	0.094	0.127	0.094	0.095	0.015
CORE	365	355	356	359	359	360	351	354	358	357.444	4.035
CORE (TRS/kb)	0.135	0.133	0.133	0.134	0.134	0.134	0.131	0.132	0.133	0.133	0.001
CORE CDS	310	303	303	302	302	303	294	296	300	301.444	4.586
CORE CDS (TRS/kb)	0.138	0.136	0.136	0.135	0.135	0.135	0.132	0.132	0.134	0.135	0.002
CORE IGR	55	52	53	57	57	57	57	58	58	56	2.179
CORE IGR (TRS/kb)	0.122	0.116	0.119	0.127	0.128	0.127	0.128	0.129	0.129	0.125	0.005
TOTAL ST22 6mA MOTIFS IN											
WHOLE	1385	1377	1407	1398	1369	1390	1402	1367	1368	1384.778	15.409
WHOLE (TRS/kb)	0.472	0.471	0.467	0.485	0.481	0.484	0.486	0.487	0.488	0.48	0.008
CORE	893	880	877	907	902	909	905	908	913	899.333	13.067
CORE (TRS/kb)	0.331	0.330	0.329	0.338	0.336	0.337	0.339	0.338	0.340	0.335	0.004
ACC	59	63	89	53	39	45	54	34	30	51.778	17.922
ACC (TRS/kb)	0.254	0.248	0.257	0.271	0.244	0.252	0.253	0.27	0.256	0.256	0.009

ST22 HsdS_A motif (YTCANNNNNCCT/AGGNNNNNNTGAR) in											
WHOLE	684	679	692	696	683	693	692	685	685	687.667	5.701
WHOLE (TRS/kb)	0.233	0.232	0.23	0.241	0.24	0.241	0.24	0.244	0.244	0.238	0.005
ACC	37	40	55	36	26	32	34	22	20	33.556	10.584
ACC (TRS/kb)	0.159	0.157	0.159	0.184	0.162	0.179	0.159	0.175	0.171	0.167	0.01
CORE	647	639	637	660	657	661	658	663	665	654.111	10.458
CORE (TRS/kb)	0.240	0.239	0.239	0.246	0.245	0.245	0.246	0.247	0.248	0.244	0.004
CORE CDS	470	467	466	474	474	475	474	471	476	471.889	3.586
CORE CDS (TRS/kb)	0.209	0.210	0.210	0.212	0.211	0.212	0.213	0.211	0.213	0.211	0.001
CORE IGR	177	172	171	186	183	186	184	192	189	182.222	7.345
CORE IGR (TRS/kb)	0.391	0.384	0.384	0.414	0.412	0.413	0.412	0.428	0.422	0.407	0.016

ST22 HsdS_S motif (GAAGNNNNNTAC/GTANNNNNCTTC) in											
WHOLE	268	264	274	264	258	261	267	257	258	263.444	5.615
WHOLE (TRS/kb)	0.099	0.099	0.103	0.098	0.096	0.097	0.1	0.096	0.096	0.098	0.002
ACC	22	23	34	17	13	13	20	12	10	18.222	7.513
ACC (TRS/kb)	0.095	0.09	0.098	0.087	0.081	0.073	0.094	0.095	0.085	0.089	0.008
CORE	246	241	240	247	245	248	247	245	248	245.222	2.906
CORE (TRS/kb)	0.091	0.09	0.09	0.092	0.091	0.092	0.092	0.091	0.092	0.091	0.001
CORE CDS	217	214	214	214	214	214	214	214	214	214.333	1.000
CORE CDS (TRS/kb)	0.097	0.096	0.096	0.096	0.095	0.095	0.096	0.096	0.096	0.096	0.000
CORE IGR	29	27	26	33	31	34	33	31	34	30.889	2.977
CORE IGR (TRS/kb)	0.064	0.06	0.058	0.074	0.07	0.076	0.074	0.069	0.076	0.069	0.007

ST45 HsdS_X motif (TAAGNNNNNNTTC/GAANNNNNNCTTA) in											
WHOLE	433	434	441	438	428	436	443	425	425	433.667	6.595
WHOLE (TRS/kb)	0.148	0.148	0.146	0.152	0.15	0.152	0.154	0.151	0.152	0.15	0.002
ACC	41	45	52	33	22	28	38	20	18	33	11.906
ACC (TRS/kb)	0.176	0.177	0.15	0.169	0.137	0.157	0.178	0.159	0.153	0.162	0.014
CORE	392	389	389	405	406	408	405	405	407	400.667	8.109
CORE (TRS/kb)	0.145	0.146	0.146	0.151	0.151	0.151	0.152	0.151	0.151	0.149	0.003
CORE CDS	260	260	260	277	279	280	282	281	282	273.444	10.199
CORE CDS (TRS/kb)	0.116	0.117	0.117	0.124	0.124	0.125	0.127	0.126	0.126	0.122	0.004
CORE IGR	132	129	129	128	127	128	123	124	125	127.222	2.819
CORE IGR (TRS/kb)	0.292	0.288	0.29	0.285	0.286	0.284	0.275	0.276	0.279	0.284	0.006

*ST622-2014 isolate

4.4.5.3 Chimeric Sequence Methylation – Singapore Collection

To discriminate between the potential methylation changes caused by the core genome recombination within the ST622 isolates, the motif match frequency of each HsdS was investigated with chimeric regions (CH1 and CH2). The two ST622 variants were both included, but the ST622-2015 chimeric region was used from comparison with the identical sequence region within the ST22 and ST45 isolates. This was done to analyse the methylation signature for the native methylation motifs of each isolate, as well as those originating from the opposite ST background (methylation landscape of ST45 isolates with ST22 motif repertoire, and vice versa). The variation of motif frequencies within different regions of the genome, including the chimeric genome region (CH) versus whole genome backbone minus the CH region (BACK), and coding sequence (CDS) versus intergenic region (IGR) seen in Figure 4.11. The ST622-2014 motif frequency results differ from the rest of the collection as the analyses use the original larger recombinant ST45 fragment as previously shown in Table 4.6. This was only included to see if there were any significant motif differences within the larger chimeric region by HsdS_α in comparison to the other isolates and sequence backgrounds but was not investigated in detail.
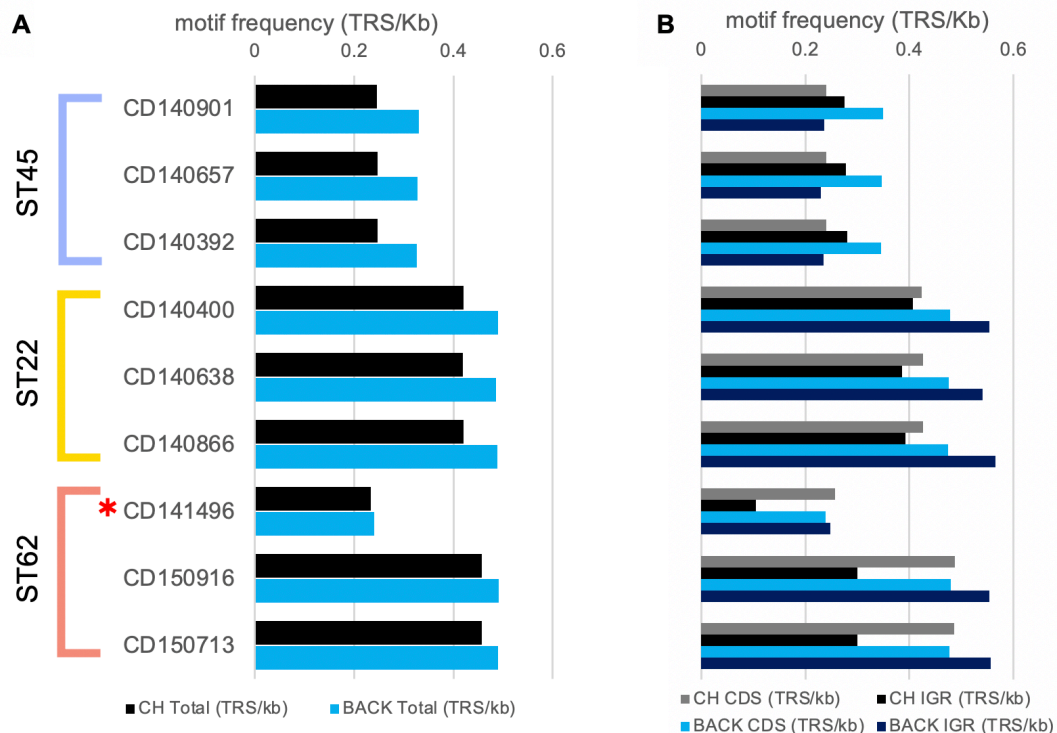
Figure 4.11 | Average TRS Frequency (TRS/kb) within the Chimera genome region (CH) and the core genome backbone minus the CH region (BACK). ST622-2014 strain marked by red asterisk (✱). **A**. Total TRS Frequency within CH (black) and BACK (blue) of each isolate **B**. TRS Frequency within the coding sequence (CDS - lighter colour) and intergenic region (IGR – darker colour) for the CH (pink) and BACK (blue) sequence regions within each isolate.

Overall, within all three ST backgrounds, the total 6mA TRS frequency within the chimeric region (CH), or equivalent position, was lower in comparison to the backbone sequence (BACK) (Figure 4.11 A). On average there were 33% more motif matches within the background genome (0.328 ± 0.002 TRS/kb – 894 motif matches) than the chimera region (0.247 ± 0.0003 TRS/kb – 57 motif matches) within the ST45 isolates. There were 16% and 7.5% higher total matches BACK:CH within the ST22 (BACK: 0.489 TRS/kb (~1286 motif matches) vs CH: 0.420 TRS/kb (~102 motif matches)) and ST622-2015 (BACK: 0.491 TRS/kb (~1265 motif matches) vs CH: 0.457 TRS/kb (~106 matches)) isolates respectively.

For this part of the study, the comparison of the motif frequencies found within the replaced chimeric region within the ST622-2015 isolates and the equivalent genome region within the ST22 isolates was imperative. The same three *sau1hsdS* (HsdS_A, HsdS_S, HsdS_X) were found in both ST background and the same methylation motifs were detected. Generally, the ST622-2015 variant isolates 1.12% higher total number of motif matches than those of ST22 (ST22: 0.489 ± 0.002 TRS/kb, ST622-2015: 0.491 ± 0.001TRS/kb). As shown by the blue bars in Figure 4.11 A, the total *sau1* 6mA motif frequency within the ST22 originated background sequences (BACK) for the ST22 and ST622-2015 isolates were almost identical, differing by only 0.50%, with the slight increase in motif frequencies in the ST22 isolates. When looking at the methylation frequency within the chimeric region (CH – black in Figure 4.11 A), the ST622-2015 strains have 8.5% higher motif match rates than the ST22 isolates (ST622-2015: 0.457 ± 0.001TRS/kb, ST22: 0.420 ± 0.009TRS/kb).

The higher overall methylation frequencies in the ST622-2015 isolates were a result of solely the increase number of motifs found within the chimeric ST45 originated region within the hybrid strain in comparison to the ST22 isolates. The ST622-2015 isolates have a total of 106 motifs (0.457 TRS/kb) whilst the ST22 isolates have 99 motif matches for the equivalent 231,200 bp CH region. To gain a better understanding of where the extra motifs lie, the total motif frequency between the CDS and the IGR region for both the CH and BACK region were analysed (Figure 4.11 B). Overall, the chimera region (greys) had a 14.5% higher motif frequency within the intergenic region (0.278 ± 0.003 TRS/kb) than within the coding region (0.247 ± 0.0003 TRS/kb) for the ST45 isolates. Contrasting this, the core genome backbone (0.347 ± 0.002 TRS/kb) of the ST45 isolates were 42% more frequently methylated than the intergenic region (0.233 ±0.003 TRS/kb).

The exact opposite trends are seen for the motif frequency within the chimeric region, the background sequence region, and their constituents for the ST22 and ST622-2015 isolates. Both ST types had a 15% higher frequency of motifs within the intergenic region (0.554 ± 0.007 TRS/kb) rather than the CDS (0.477 ± 0.002 TRS/kb) for the sequence background. The prominent difference between the motif frequencies for the two STs, is the frequency of matches within the chimeric region. For both, the matches within the CDS region is higher than that in the IGR, but the resulting TRS frequencies shift the ratio between the CDS:IGR matches. The ST622-2015 isolates have a 14% higher TRS frequency (0.487 ± 0.0004 TRS/kb – 95 motif matches) within the CDS than the ST22 isolates (0.425 ± 0.001 TRS/kb – 85 motif matches), whilst the ST22 have a 31% higher motif frequency (0.395 ± 0.011 TRS/kb – 15 motif matches) within the IGR than the ST622-2015 (0.300 ± 0.0002 TRS/kb – 11 motif matches).

To be able to distinguish between the variation seen within the chimeric sequence region matches, and potential functional difference of *sau1* systems, the number of matches attributed to each specific HsdS were analysed. The total motif frequency within the chimera and the background sequence for HsdS_α are shown in Figure 4.12 A. The core background region for the ST45 isolates is 26% more densely methylated (0.196 ± 0.0004 TRS/kb) than the chimera region (0.156 ± 0.0001 TRS/kb). The same trend is reflected for the matches within the CDS and IGR (Figure 4.12 B) for each sequence region as for the total motifs in the previous section; CH 25% higher frequency in IGR than CDS (0.149 ± 0.0002 TRS/kb), BACK: 32% higher frequency in CDS than IGR (0.142 ± 0.004 TRS/kb).

The HsdS_α motif ratio within the CH and BACK for the ST22 isolates is almost 1:1, with 2% higher methylation frequency within the chimeric region, at 0.245 TRS/kb (± 0.003 TRS/kb). Although the methylation match frequency for the background sequence region is almost identical for the ST22 and ST622-2015 isolates (0.241 ± 0.0007 TRS/kb and 0.242 ± 0.0002 TRS/kb; 0.7% higher TRS/kb in the ST622-2015), the chimeric isolates had 10% higher motif frequency within the CH region (0.272 ± 0.0003 TRS/kb).
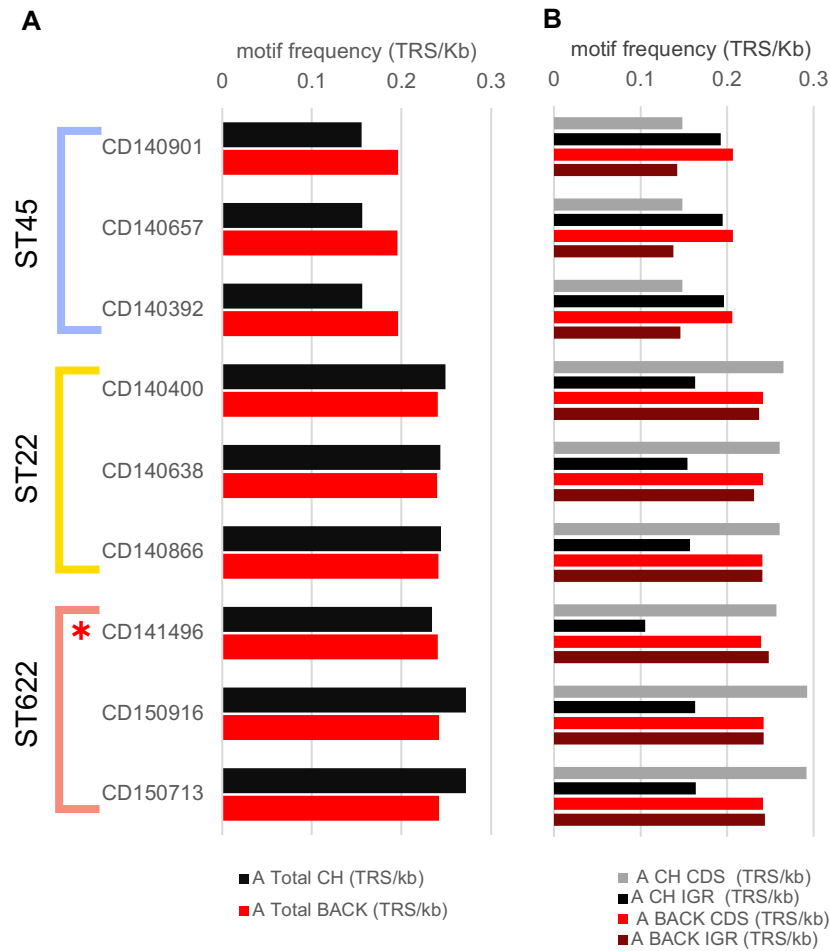
Figure 4.12 | Average TRS Frequency (TRS/kb) for motifs associated with HsdS_α. Matches wiithin **A.** the Chimera (CH - black) and the core genome (BACK - red). **B.** TRS Frequency within the coding sequence (CDS - lighter colour) and intergenic region (IGR – darker colour). ST622-2014 strain marked by red asterisk (✱).
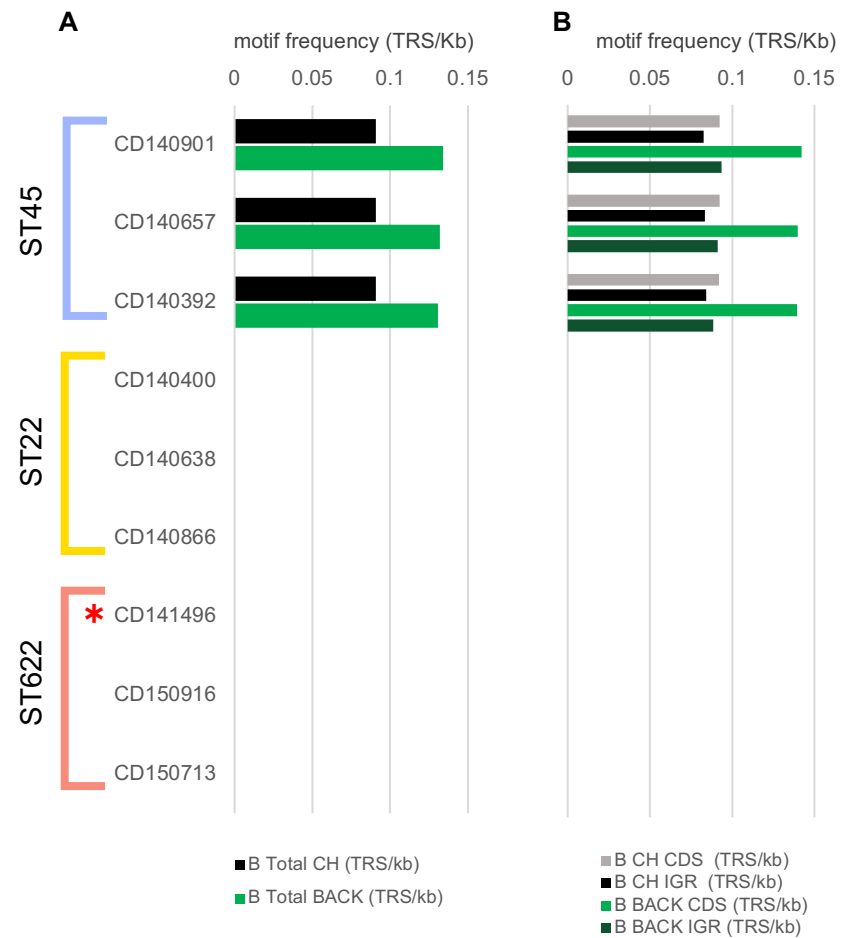
Figure 4.13 | Average TRS Frequency (TRS/kb) for motifs associated with HsdS_β. Matches within **A.** the Chimera (CH - black) and the core genome (BACK-green). **B**. TRS Frequency within the coding sequence (CDS - lighter colour) and intergenic region (IGR – darker colour). ST622-2014 strain marked by red asterisk (✱).
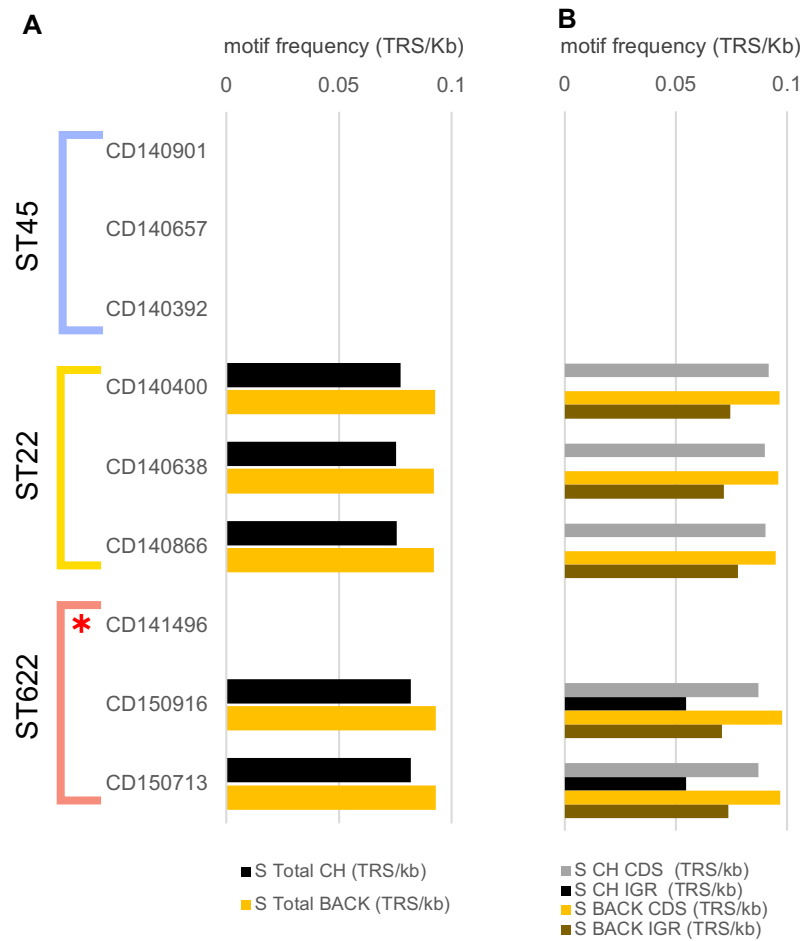
Figure 4.15 | Average TRS Frequency (TRS/kb) for motifs associated with HsdS_S, Motifs within **A.** the Chimera (CH - black) and the core genome (BACK-gold). **B.** TRS Frequency within the coding sequence (CDS - lighter colour) and intergenic region (IGR – darker colour). ST622-2014 strain marked by red asterisk (*).
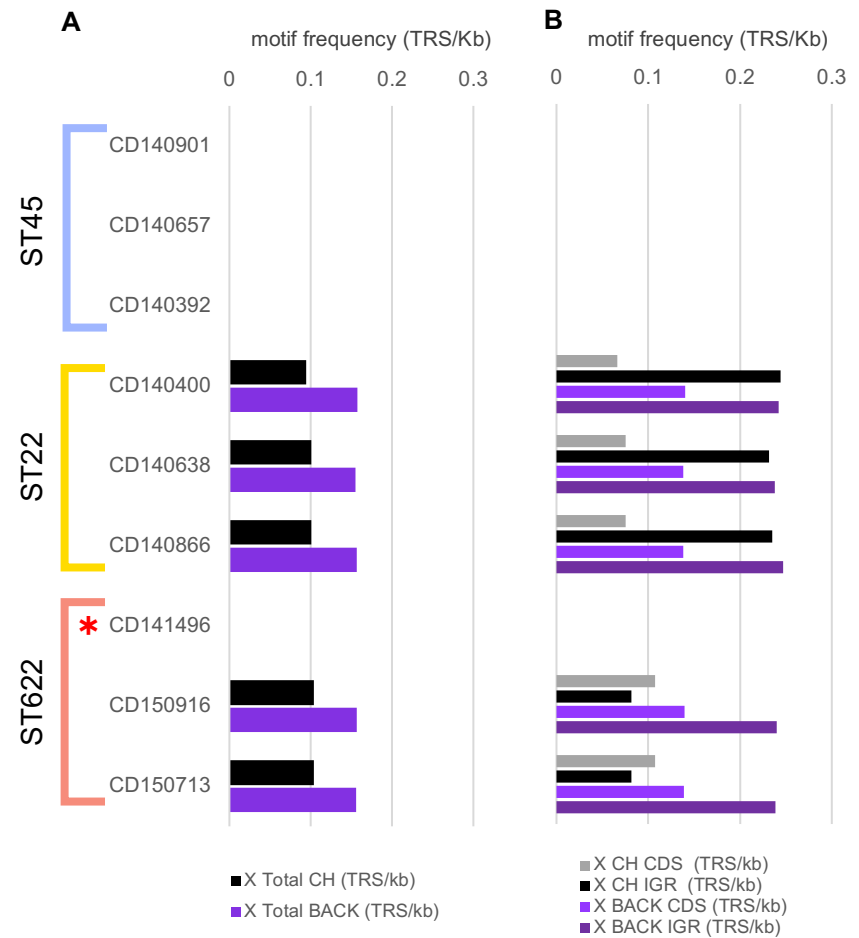
Figure 4.14 | Average TRS Frequency (TRS/kb) for motifs associated with HsdS_X. Motifs within **A.** the Chimera (CH - black) and the core genome (BACK-violet). **B.** TRS Frequency within the coding sequence (CDS - lighter colour) and intergenic region (IGR – darker colour). ST622-2014 strain marked by red asterisk (*).

This increase in HsdS_α motif frequency can be directly attributed to higher number of matches within the CDS within the CH region as seen in Figure 4.12 B. The ST622-2015 isolates have 11% higher motif frequency within the CDS than the ST22 (0.262 ± 0.003 TRS/kb). Generally, the CDS region was more densely methylated than the IGR within the CH region for both ST types, whilst the CDS:IGR ratio for the background genome was almost 1:1; the ST22 have a 2% increased methylation frequency within the CDS than the IGR (0.236 ± 0.005 TRS/kb). Along with an alpha system the ST45 isolates also have methylation signatures associated with HsdS_β (Figure 4.13). Overall, the core background genome was 31% more densely methylated than the sequence equivalent to the recombined chimeric region (0.091 ± 0.0002 TRS/kb). Both the CH and the BACK had higher motif frequencies within the CDS than the IGR region seen in Figure 4.13 B.

Only the ST22 and ST622-2015 isolates contained two additional accessory Sau1 systems. The overall motif signatures for both the HsdS_S and HsdS_X follow the same trend: higher methylation density within the BACK than the CH region seen in Figure 4.15 A and Figure 4.14 A. HsdS_S had the lowest TRS/kb frequency out of all 4 systems. The motifs attributed to HsdS_S were 20% higher within the BACK (0.092 ± 0.0004 TRS/kb) than the CH for ST22 isolates, whilst on 12% higher in the BACK (0.093 ± 0.0002 TRS/kb) than CH within the ST622-2015 isolates. The slightly higher frequency within the CH of the ST622-2015 can be attributed to 2 motif matches within the IGR for this region, which are fully absent within the ST22 as seen in Figure 4.15 B.

Regarding the motifs recognised by HsdS_X, there were 36% and 34% more matches within the background sequence rather than the chimera, in both the ST22 (0.099 ± 0.004 TRS/kb) and ST622-2015 (0.104 ± 0.0001 TRS/kb) respectively (Figure 4.14 A). The differences within overall motif frequency within the two different sequence regions can be clearly attributed to the 60% decreased motif matches within the intergenic region, and 32% increase within the CDS of the CH of the ST622-2015 isolates compared to the ST22 isolates seen in Figure 4.14 B.

4.4.5.4 6mA TRS Matches with Chimeric Region in ST622 Isolates (CH1 + CH2)

The region of recombination within the ST622 isolates is homologous throughout the ST types, and as the previous section concludes there are subtle changes in the total number of 6mA TRS found within differing regions (CDS or IGR) of the chimeric sequence. To deduce specific positions of each motif, the differences we may see between the ST622 and ST22 methylation landscape, and the potential gene expression effects these may have, each methylation motif was compared gene by gene within the CH region. Table 4.12 shows the number of methylation motifs found within the 203 CDS within the chimeric region as well as 200 bp upstream (promoter (PROM) and 5'UTR (marked <)) and downstream (3'UTR and TTS (marked >)) between ST45, ST622-2015 and ST22.

When comparing the methylation landscape of the recombinant chimeric sequence region in all 3 ST types, 102/203 CDS (including ± 200 bp regulatory regions (REG)) contained 6mA modified bases. As the methylation specificities for the ST45 and ST22/622 differed, it was expected that the location of the modified 6mA would be differential. Nevertheless, modification motifs from all ST types were found within 23 CDS regions (navy). 62 methylation events were shared between ST22 and ST622-2015 isolates (yellow), with 37 common between the two STs but not shared by ST45. Eighteen genes and their regulatory regions were uniquely methylated within the ST45 sequence (sky blue), whilst 9 and 11 CDS/REG were uniquely methylated within ST22 (red) and ST622 (gold) sequence backgrounds of the CH region respectively.

The ST45 isolates contained 57 TRS motifs within 40 CDS within the 203 gene region, and 10 motifs within the IGR (3 motifs within PROM, 2 motifs in 3'UTR, 5 motifs in non-coding region). The ST622-2015 isolates had overall 106 TRS matches within the chimeric region, 14 IGR (5 motifs within PROM, 7 in 3'UTR, 2 motifs in non-coding region) and 70 CDS matches. ST ST22 isolates had on average 100 TRS matches, 15 in the IGR (8 motifs in PROM, 7 motifs in 3'UTR) and 68 CDS matches.

Although overall it seems as though the ST622-2015 isolates have acquired an additional 6 motifs in comparison to the ST22 isolates, in reality there were 28 motif event differences between the two STs (highlighted in bold). There are 8 REG locations (< = 3 motifs, > = 4 motifs) and 13 CDS locations (17 TRS in total) in which the ST622-2015 isolates gain a motif, and 9 REG (< = 5 motifs, > = 4 motifs) and 10 CDS locations where the ST22 isolates gain a motif. Hence, there is potential for differential methylation between ST22/ST622.

Table 4.12 | Methylation Motifs within CDS and INT Regions as per Group Tag

ST45	ST622	ST22	Tag	CDS TAG (CD140400)	Gene / protein coded
Chimeric region 1 (CH1 - from origin of replication)					
				1	*dnaA*
	•	•	yellow	2	*dnaN*
	•	•	yellow	3	conserved protein
	••••	••••	yellow	4	*recF*
•	•	•	navy ≠	5	*gyrB*
•	••	••	navy ≠	6	*gyrA*
			white	7	*nnrD*
	• >	•	yellow ≠	8	**hutH**

ST45	ST622	ST22	Tag	CDS TAG (CD140400)	Gene / protein coded
Chimeric region 2 (CH2 - towards terminus)					
<<<	•	•	navy ≠	2571	*ecsA_3*
				2572	membrane protein
	•••	••	yellow ≠	**2573**	***fbp***
	>	>		2574	membrane protein
	>	>		2575	phosphoesterase
	•	••	yellow ≠	**2576**	***mhq0_2***
				2577	transcriptional regulator
				2578	acetyltransferase
				2579	*catE_2*
				2580	NADPH reductase
•	•	• >	navy ≠	**2581**	***ldhD_1***
	•	•	yellow	2582	*supH*
•			light blue	2583	*ybbL_2*
•	•	•	navy	2584	exported permease
	•	•	yellow	2585	*srtA*
				2586	*yncA*
				2587	hypothetical protein
				2588	*sdhA_2*
•			light blue	2589	*sdhB*
	<		orange	**2590**	***pfoR_2***
				2591	hypothetical protein
				2592	*yicL*
•	•	••	navy ≠	**2593**	***mlhB_2***
	< •	< •	yellow	2594	thioredoxin
				2595	thioesterase
				2596	transposase IS*1271*
				2597	transposase IS*1271*
				2598	transposase IS*1271*
				2599	transposase
				2600	*glcB1*
				2601	*pox5*
				2602	*cidB*
				2603	*cidA*
•			light blue	2604	*cidR*
				2605	putative cytosolic protein
				2606	*ssA2_4*
				2607	*mvaA*
		•	red	**2608**	***mvaS***
				2609	*ogt*
•	•	•	navy	2610	*clpL*
				2611	virus attachement protein
•	•	•	navy	2612	*feoB_1*
				2613	*feoA*
	•	•	yellow	2614	*mmpL8*
				2615	*tra*

S1	S2	S3	Color		Gene	Name
					2616	*rocA*
					2617	acetyltransferase
					2618	exported protein
•	•	•	dark blue		2619	*copA*
					2620	*coZ*
					2621	*ldhD_2*
					2622	*dapL*
	••	••	yellow		2623	*crtN*
					2624	*crtM*
•			blue		2625	*crtQ*
	•		orange		**2626**	**crtP**
	<				**2627**	**acetyltransferase precursor**
		NP			− (ST45/ST622)	*ssaA2_5*
••	•	•	dark blue	≠	2629	*oatA_2*
NP	NP	<	red		**2630**	**acetyltransferase**
>			blue		2631	*isaA*
	• >		orange		**2632**	**membrane protein**
<			blue		2633	T Reg
					2634	hypothetical protein
	•		orange		**2635**	**ynzC**
					2636	glyoxalase
	•	•	yellow		2637	*azoB*
					2638	hypothetical protein
		•	red		**2639**	**acrR**
					2640	*cpnA*
					2641	decarboxylase
					2642	lipase
•	••	••	dark blue	≠	2644	*cobW*
					2645	*feoB_2*
•			blue		2646	*czc0*
>					2647	*esaC_2*
	•		orange		**2648**	**hypothetical protein**
					2649	TVII secretion effector
•	•	•	dark blue		2650	fructosamine kinase
					2651	*pyrD*
	•	•	yellow		2652	membrane protein
					2653	adenine hydrolase
					2654	*phnb*
					2655	transcriptional regulator
••	•	•	dark blue	≠	2656	*cocE*
NP	NP	•	red		**2657**	**cell wall protein**
NP	NP	••	red		**2658**	**mpr (serine protease)**
					2659	*panD*
	•	•	yellow		2660	*panC*
					2661	*panB*
					2662	*panE_2*
	<		orange		**2663**	**aldC_2**
••			blue		2666	*ldh2*
•					2667	*pheP*
	•	•	yellow		2668	*puuE/gabT*
					2669	hypothetical protein
	••	••	yellow		2670	*fda*
	•	•			2671	*mqo2*
					2672	hypothetical protein
•			blue		2673	*bclA*
					2674	antibiotic biosynthesis protein
					2675	putative cytosolic protein
					2676	*betA*

Col1	Col2	Col3	Col4	ID	Gene
	●	●	yellow	2677	*gbsA*
				2678	transcriptional regulator
				2679	*betT*
				2680	*betP*
< ●	●●	●●	≠ navy	2681	*nrdD*
●	●	●		2682	*citN*
●			blue	2683	*sirC_2*
●				2684	*cysJ*
●	● >	●	navy	**2685**	***bsaA_2***
	●	<	≠ yellow	**2686**	**permease**
NP	NP	●	red	**2687**	**hypothetical protein**
				2688	*lolD_2*
				2689	*graS*
				2690	*yvcP*
				2692	*phoB*
				2693	hypothetical protein
				2694	transcriptional regulator
				2695	hydrolase
				2696	*clfB_1*
●			blue	2697	*arcR*
				2698	*arcC*
				2699	*arcD_2*
				2700	*arcB*
	●●●	●●●	yellow	2701	*arcA*
				2702	*argR_2*
>	< ●●	●	≠ navy	**2703**	***aur***
				2704	*isaB*
	>	red	**2705**	**low complexity protein**
●				2706	*licR_4*
●	●	●●	≠ navy	**2707**	***manP***
●	●	●		2708	*pmi*
●	●	●●	≠ navy	**2709**	**phage infection protein**
	● >	● >	yellow	2710	amidase
	>	>	yellow	2711	amidase
	●	●		2712	hypothetical protein
●●			blue	2713	glycosyltransferase
				2714	*gftA*
●			blue	2715	*secA*
	●	●	yellow	2716	*asp3*
●				2717	*asp2*
	●●	●●	yellow	2718	*asp1*
				2719	*secY2*
				2720	*sraP*
	●	●	yellow	2721	flavin reductase
	<	<		2722	lipoprotein
				2723	lipoprotein
				2724	hypothetical protein
				2725	hypothetical protein
				2726	nitrilotriacetate monoxidase
				2727	*msrA_2*
				2728	*capC_2*
	●	orange	**2729**	***epsD***
	●	●	yellow	2730	*cap8A*
				2731	*icaR*
				2732	*icaA*
●	●	●	navy	2733	*icaD*
				2734	*icaB*
>			blue	2735	*icaC*
>	●	●	≠ navy	2736	*lipA_3*

187

				2737	*hisI*
•	•	•	■	2738	*hisF*
				2739	*hisA*
				2740	*hisH*
	•	•	■ (yellow)	2741	*hisB*
				2742	*hisC_2*
•	•	•	■	2743	*hisD*
				2744	*hisG*
•			■ (cyan)	2745	phosphoribosyltranferase
				2746	lipoprotein
	•	•	■ (yellow)	2747	hypothetical protein
	••	••		2749	lactonase
				2750	sulfurtransferase
				2751	*pcp*
	•		■ (orange)	**2752**	**hypothetical protein**
	•			**2753**	**PadR T Reg**
•	• >		≠ (purple)	**2754**	**dibB protein**
••••	•••••	••••	≠	**2755**	*cna*
	••••			**2756**	***citT/ttdT***
	•	• >>	≠ (yellow)	**2757**	***rarD***
•	•	•	■	2758	DNA binding protein
				2759	*nixA*
				2760	*nhoA*
				2761	membrane protein
				2762	*vraD*
				2763	permease
		<	■ (red)	**2764**	**hypothetical protein**
		<		**2766**	***immR - T Reg***
				2767	membrane protein
				2768	permease
				2771	*parB_2*
•			■ (cyan)	2772	*gidB*
	••	< •	≠ (yellow)	**2773**	***mnmG***
	•••	••	≠	**2774**	***mnmE***
				2775	*rnpA*
				2776	*rpmH*

•	*6mA motif within CDS*
<	*6mA motif upstream of CDS in PROM*
>	*6mA motif downstream of CDS in 3'UTR*
< or > (black)	*6mA motif outside of 200bp IGR cutoff*
≠	*difference between 6mA numbers*
NP	*Gene NOT PRESENT in genome*
■ (navy)	*6mA present in all 3 STs (ST22, ST622, ST45)*
■ (yellow)	*6mA present in both ST22 and ST622, not ST45*
■ (purple)	*6mA present in both ST45 and ST622, not ST22*
■ (cyan)	*6mA present in only ST45*
■ (red)	*6mA present in only ST622*
■ (orange)	*6mA present in only ST22*

4.4.6 Transcriptomic & Differential Expression Analysis

To gain an understanding of the potential epigenetic regulation facilitated through Sau1 6mA methylation, the complete transcriptome of the three STs within the Singapore collection were sequenced at single base resulting with RNASeq (Illumina-C HiSeq 4000). One of the key variables in this experiment was keeping the growth conditions for each bacterial strain identical (37ºC, 160RPM in 250 ml baffled flasks with filter caps grown overnight until each culture reached 0.6 OD_{600}) so that the gene expression profiles show the same growth cycles in the same experimental conditions. For accurate detection of differentially expressed genes, 3 biological replicates, for each of the three isolates within an ST were included. The libraries were also sequenced twice resulting in a technical replicate for each sample prepared, totalling 54 assets. The sequencing pipeline at the Wellcome Sanger Institute has automatic QC (FASTQC) which all assets passed, and the average sequencing coverage was 1147.69x. The raw RNA sequence reads were mapped against reference genome CD140400 (ST22) via Wellcome Sanger Institute internal Prokaryotic RNASeq Expression Analysis. BWA was used to index the reference and the reads are aligned using default parameters. Gene expression values were computed from the read alignments to the coding sequence to generate a number of reads mapped and RPKM (reads per kilobase per million), with only reads with mapping quality score of 10 were included in the count. The assembled transcript output files used in downstream analysis included the corrected bam files for the mapped reads for visualisation in Artemis, and the expression.csv containing the count data used for differential gene expression experiments. The mapped RPKM count data was parsed to only include CDS features, and excluded any ORFs pertaining to MGEs, so that only the core genome (including the recombined region) was used in the analysis.

To get a global overview of the transcriptomic landscape for the three represented ST types, the assembled transcripts were investigated in four consecutive analysis steps with edgeR statistical methods including: 1) Transcript Quantification, 2) QC of Samples and Replicates, 3) Differential Gene Expression, 4) Promoter Analysis.

Several differentially expression (DE) analysis methods for RNASeq data were investigated including DESeq2 and EdgeR. These methods are benchmarked 'gold-standards' and were selected as they can do multiple comparisons, rather than just pairwise comparisons which was essential for the multi-layered experimental design.

IDEAMex, MeV and Trinity were used to run transcript QC, differential gene expression and to visualise results in a detailed overview.

It is also important to note that all downstream results include both variants of ST622. The set of analyses presented were conducted both with the inclusion and with the exclusion of ST622-2014 isolate CD141496 with the ST622-2015 isolates (CD150713 and CD150916) and showed no significant DE gene differences. As previously characterised, the major sequence difference between the two variants is the larger CH1 recombined region of ST45 origin, spanning the SCC*mec*. The focus of this study was investigated the differences in CORE genome expression patterns as a result of differential 6mA methylation, and thus MGEs and variable sequence regions were removed from the transcriptomic analyses, therefore most of differing sequence region between the two ST622 variants was also removed. As there were no significant in overall gene composition or DE results, the ST622-2014 isolate and replicates were included in this study to increase the sample size and statistical power of the downstream analyses.

4.4.6.1 Transcript Quantification & Filtering

The overall transcript count levels per isolate was quantified to distinguish between number and density at which genes are expressed. A Count Per Million (CPM) plot was generated, showing the number of genes within each sample and the number of counts detected for each, 0, 1, 2, 5, or 10 CPM (Figure 4.16). There are genes in each isolate which had a CPM of 0 (neon green), meaning they are not expressed. These were filtered out as they are most likely genes which are not conserved between the 3 STs.
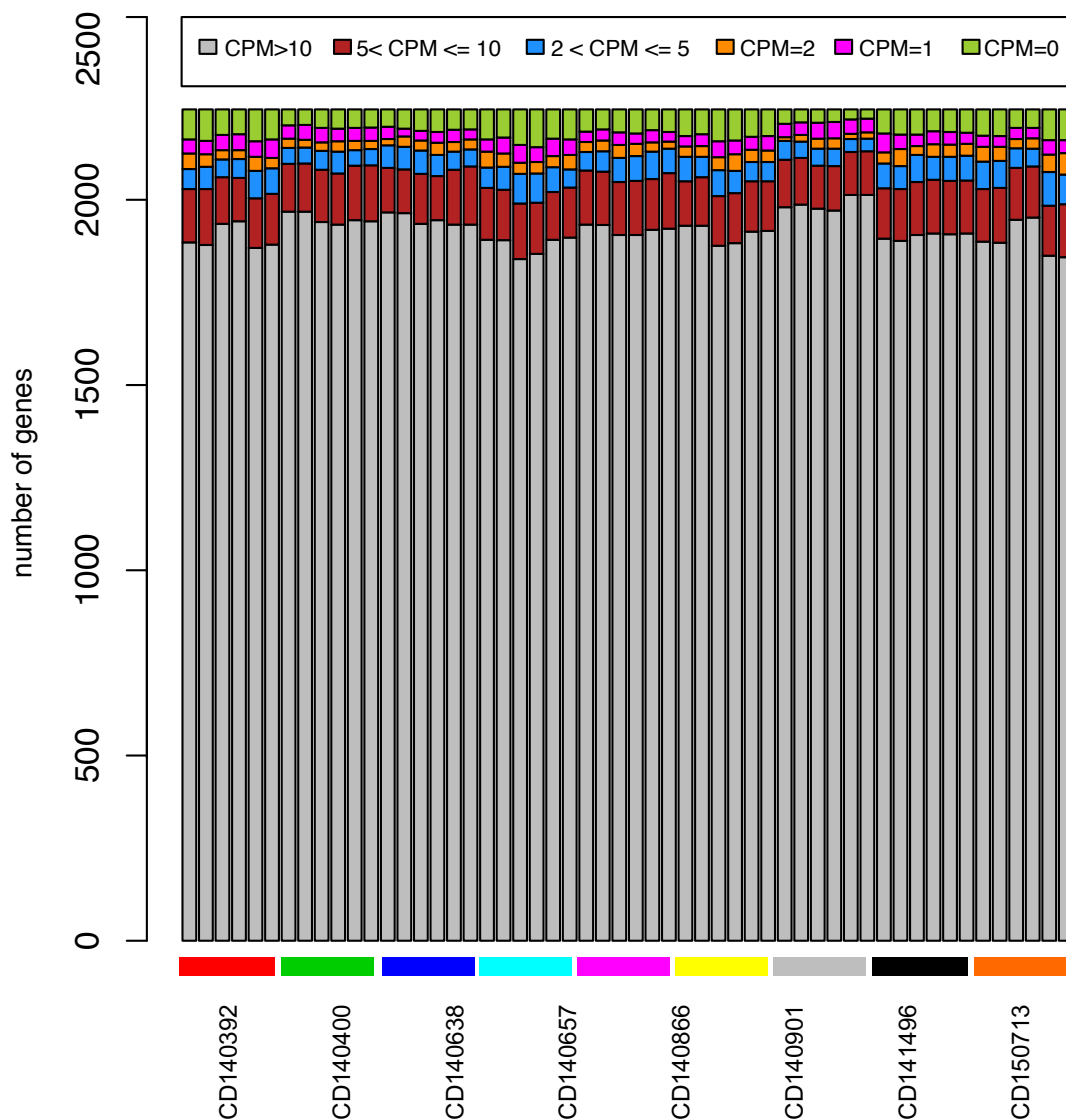
Figure 4.16 | Counts Per Million (CPM) Plot showing the number of genes within each sample, and the transcript counts for each: 0, 1, 2, 5, or 10 < CPM. 2000+ genes within each isolate have a CPM of 5 <, indicating that the genes are expressed at high levels. CPM is calculated by normalising the read counts by the total counts per sample. ST45 isolates: red (CD140392), aqua (CD140657), yellow (CD140901); ST22 isolates: green (CD140400), blue (140638), magenta (CD140866); ST622 isolates: grey (CD141496), black (CD150713), orange (CD150916).

The count and density distributions of raw log-intensities for each sample can be seen in Figure 4.17 A. Although the raw count values distribution was not highly skewed (median not too different between samples), there are slight sequencing yield differences between samples seen in Figure 4.17 A. To exclude biases introduced throughout sample preparation or sequencing process, the Trimmed Mean of M-values (TMM) method was used to normalise the distribution of count values according to the sequencing yield (sequencing depth, gene lengths, RNA composition (number of genes expressed, highly expressed genes, contamination)) of each sample (Figure 4.17 B).
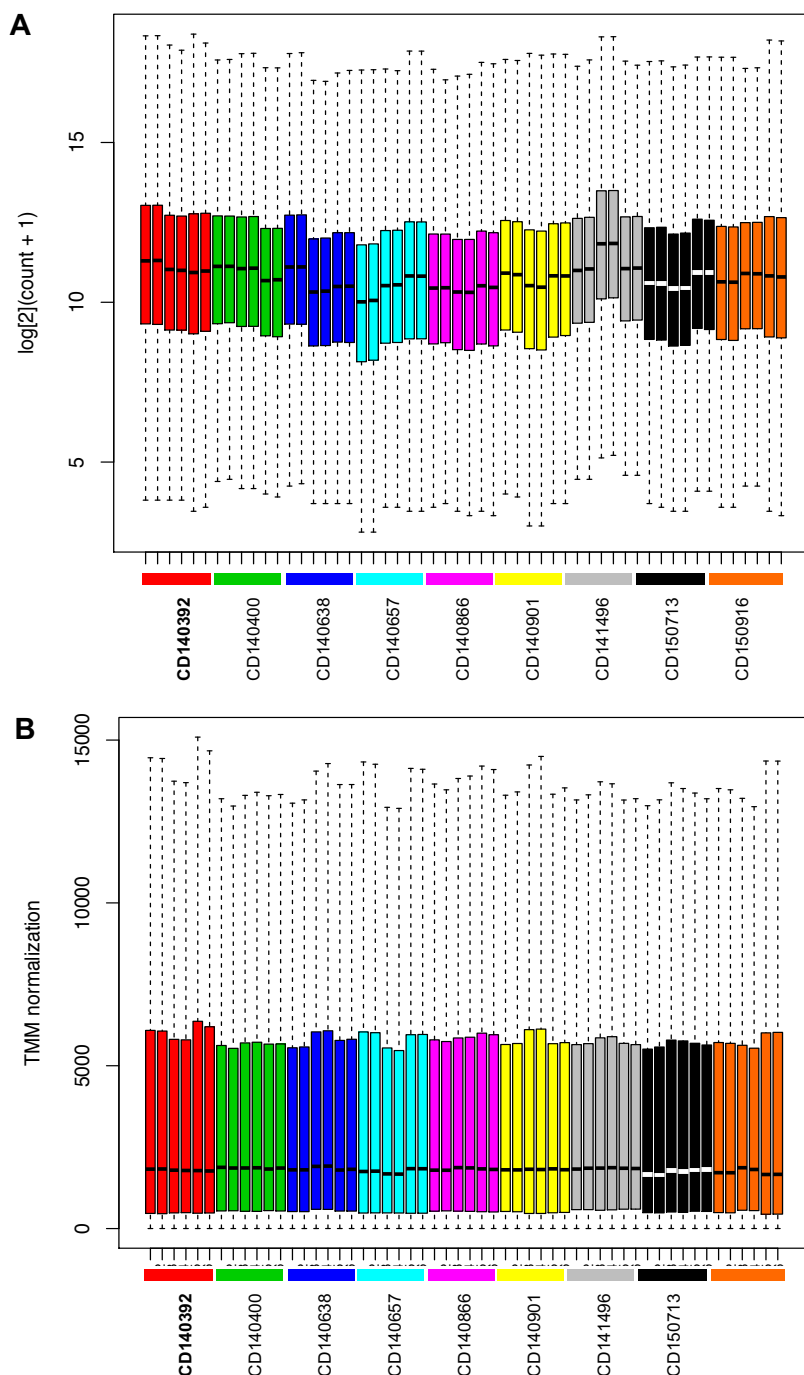
Figure 4.17 | Transcript Count Distribution Boxplots.

A. Count distribution per sample (log2)

B. TMM normalised count distribution per sample – normalised according to the sequencing yield of each sample.

ST45 isolates
red (CD140392)
aqua (CD140657)
yellow (CD140901)

ST22 isolates
green (CD140400)
blue (CD140638)
magenta (CD140866)

ST622 isolates
grey (CD141496)
black (CD150713)
orange (CD150916)

4.4.6.2 QC of Samples and Replicates

Biological and technical replicates were included for accurate detection of differential expressed transcripts. To visualise the proximities (similarities or differences) between each replicate, the normalised count data was analysed by mutli-dimensional scaling (MDS) illustrated in Figure 4.18. Similar expression profiles cluster closely together, whilst outlying replicates may inform whether our experimental condition represents the major source of variation in the data. In our data set, the ST45 isolates form one distinct cluster on the left, and the ST22 and ST622 isolates form a second cluster, although more dispersed. This indicates that on a global scale, the gene expression levels for the core genome of the ST622 isolates resemble that of the ST22, rather than having a unique signature. It is interesting to note that the 2014 ST622 variant (grey – CD141496), clusters more closely to the ST22 isolates, rather than the ST622-2015 isolates, although containing a much larger ST45 sequence fragment. Although most biological replicates for each isolate cluster moderately tightly together, CD140657 (aqua), CD140668 (blue), CD150713 (black) and CD150916 (orange) are more dispersed, indicating higher experimental variation.
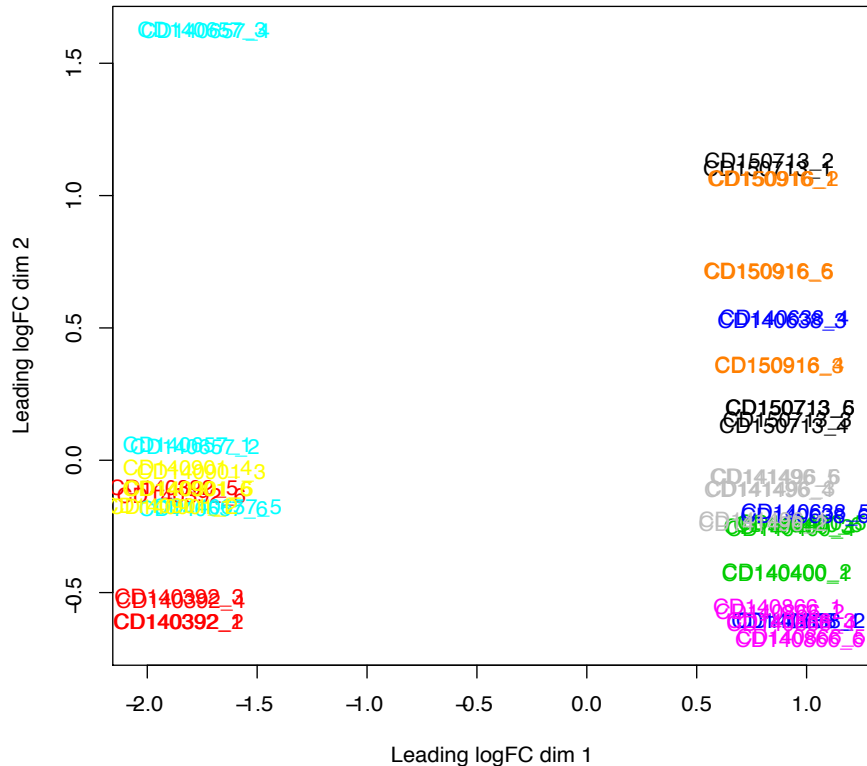
Figure 4.18 | Multidimensional Scaling of TMM normalised Count Data.
Expression Data for all core genes. ST45 isolates: red (CD140392), aqua (CD140657), yellow (CD140901); ST22 isolates: green (CD140400), blue (CD140638), magenta (CD140866); ST622 isolates: grey (CD141496), black (CD150713), orange (CD150916).

To gain a better understanding of how each replicate corresponds to each other within each ST type (condition), hierarchical clustering was conducted to identify potential outliers. Figure 4.19 displays the sample-to-sample correlation of gene expression for all pairwise combinations of samples within the dataset within each ST.
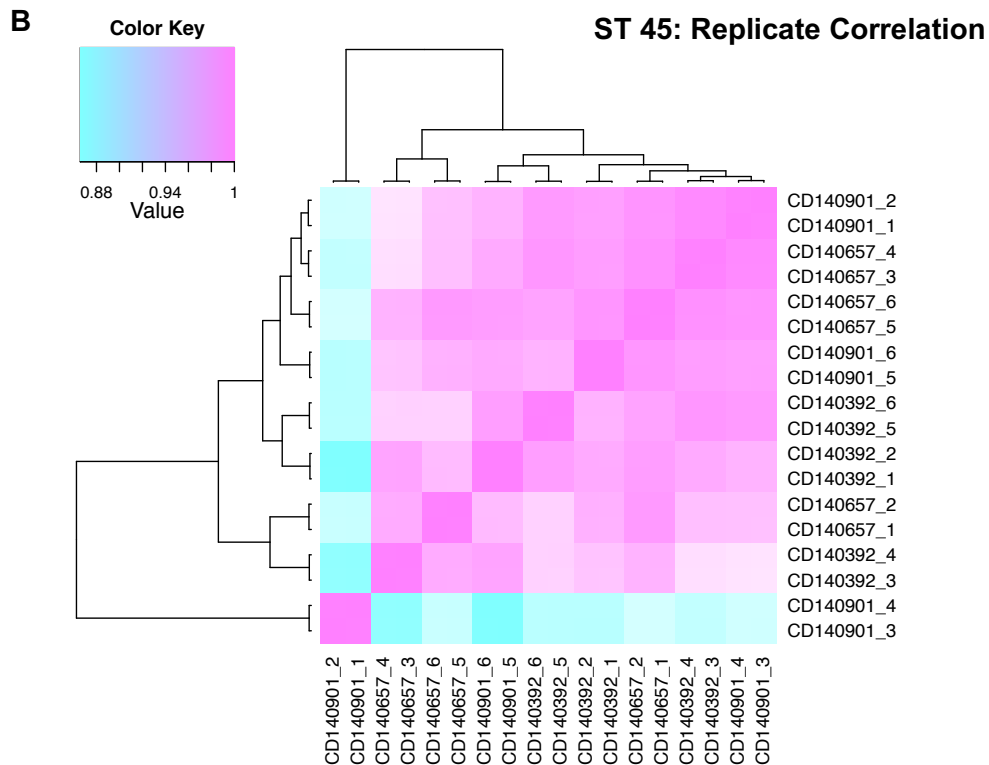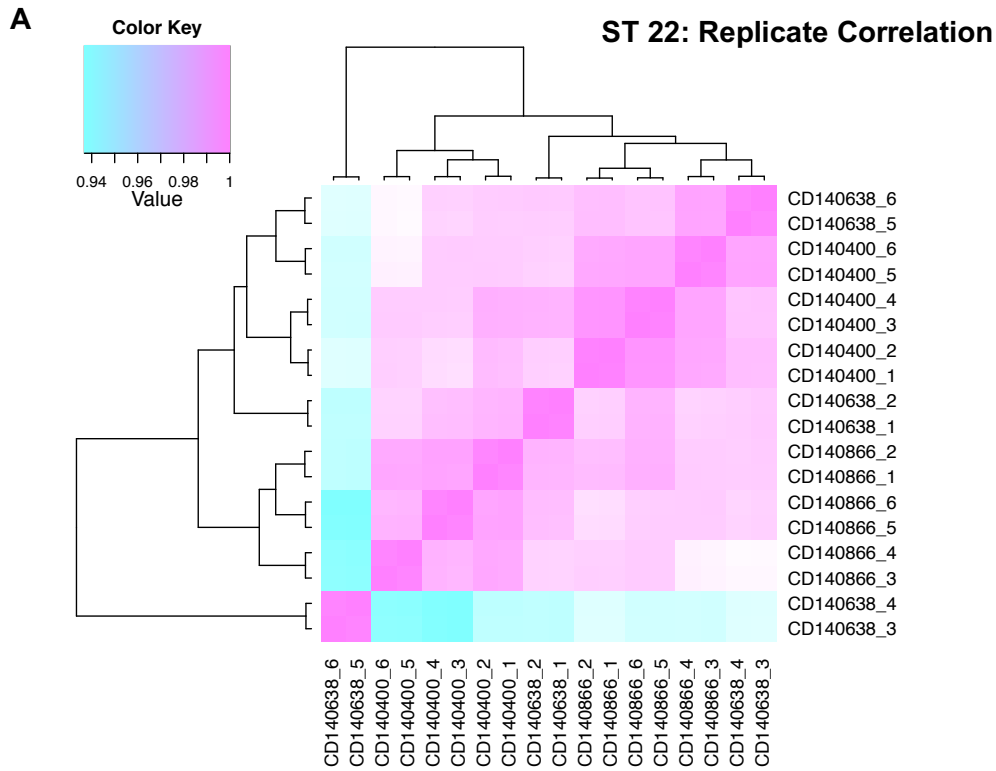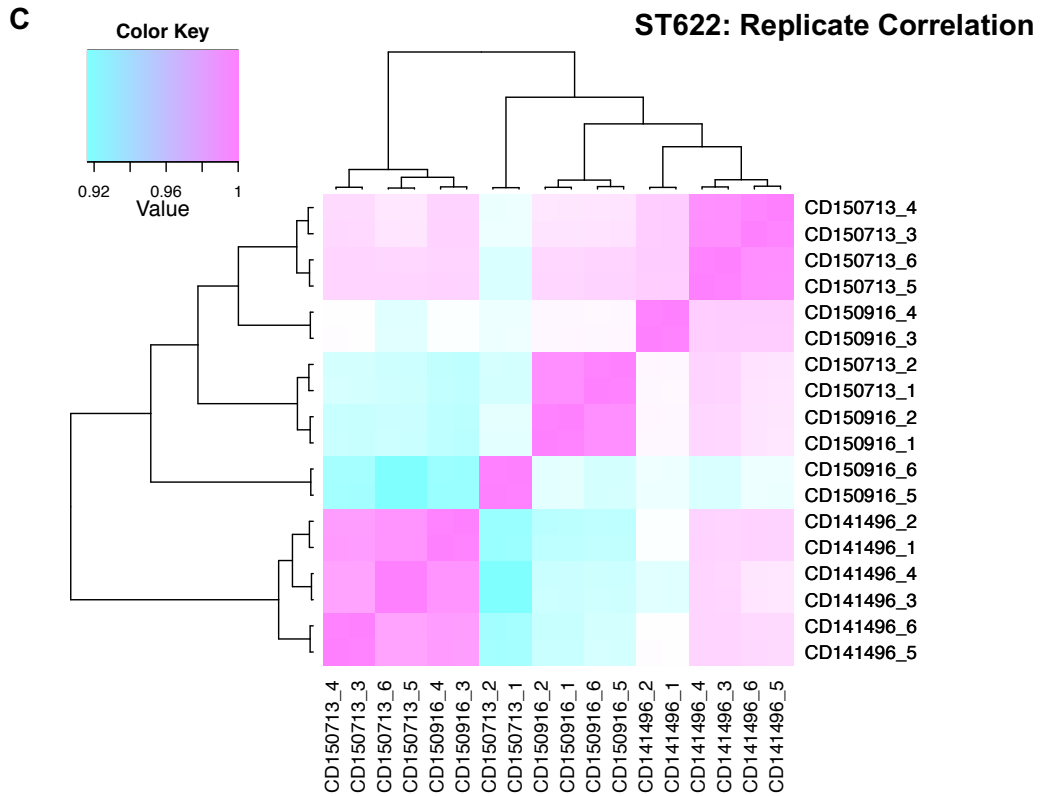
Figure 4.19 | Hierarchical Clustering Tree of Gene Expression for 3 *S. aureus* STs.
A. Replicated Correlation of ST22 samples. **B**. Replicate Correlation of ST45 samples. **C**. Replicate Correlation of ST622 isolates.

None of the clustering comparisons had a Pearson's-correlation coefficient (PCC) lower than 0.8, indicating no significant differences between replicates / outliers. The ST45 (PCC: 0.88) and ST22 replicates (PCC: 0.94) clustered according to each biological replicate making 9 distinct clusters, all clustering closely together with 1 biological replicate seeming to be more different to the rest of the dataset (ST45: CD140901_1 and CD140901_2; ST22: CD140638_6 and CD140638_5). The ST622 replicates clustered more tightly for each represented isolate (PCC: 0.92), with correlation patterns for the two ST622 variant types.

4.4.6.3 Overall Differential Expression Analysis

As the first step to differential gene expression analysis, the transcripts of the most differentially expressed genes were clustered according to their expression profile across samples (FDR and P-value > 0.001; ±2 LogFC). The TMM-normalised reads are log2 transformed, mean centered and clustered according to expression profile to produce a heatmap (Figure 4.20). The results for the ST622 group include both the 2014 and 2015 variant of the chimeric strain. Downstream analysis where ST622 was only represented by the ST622-2015 variant, or with the inclusion of the 2014 chimeric variant had the same results with not significant gene differences between comparisons
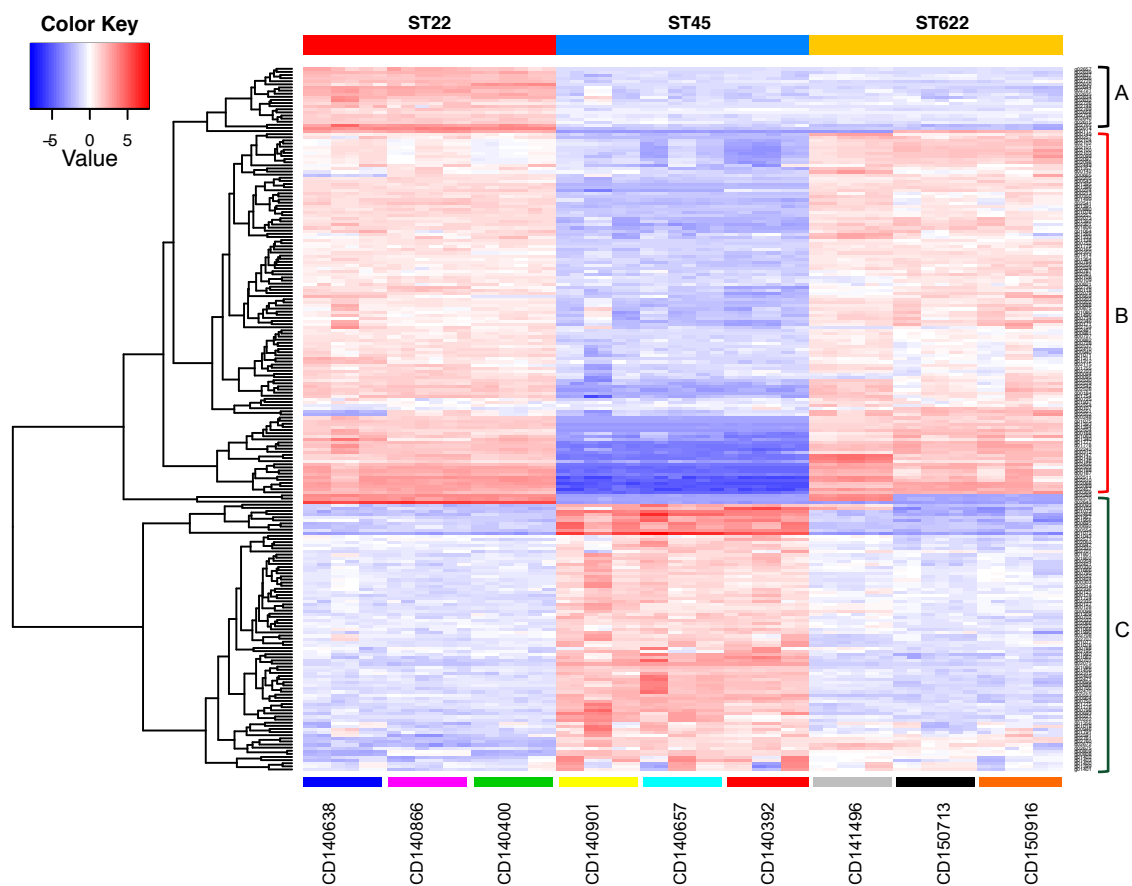
Figure 4.20 | Hierarchical Clustering of DE Genes vs ST group.
Overall 227 genes (rows) with differential expression between 3 ST represented by the columns (ST22-red; ST45-blue; ST622-gold). Expression values (RPKM) are log2 transformed then median-centred by gene (right most column). Genes are hierarchically clustered according to expression profile clustering in 3 groups (A: n=30 genes, B: n=111, C: n=85). Heat map values were calculated by subtracting each gene media log2 (RPKM) value from the log (RPKM) value of each sample – upregulated genes are red whilst downregulated genes are blue (edgeR > 4, FDR< 0.001, Pvalue < 0.001). Each replicate of each sample is also annotated on the bottom of the gene matrix (ST45 isolates: red (CD140392), aqua (CD140657), yellow (CD140901); ST22 isolates: green (CD140400), blue (CD140638), magenta (CD140866); ST622 isolates: grey (CD141496), black (CD150713), orange (CD150916).

Overall, 227 genes were differentially expressed between the three ST groups. From Figure 4.20 we can see that there are 3 distinct gene clusters which are up or down-regulated between ST types (Cluster A: 30 genes, Cluster B: 111 genes, Cluster C: 85 genes). The differential expression profile patterns for ST22 and ST622 look very similar, with upregulated patterns for genes in Cluster B and downregulated for Cluster C in comparison to ST45. The expression profile for genes within Cluster A, are similar for ST622 and ST45, both being downregulated in comparison to ST22. Pairwise comparisons were conducted to identify differentially expressed genes between STs detailed in Figure 4.21.
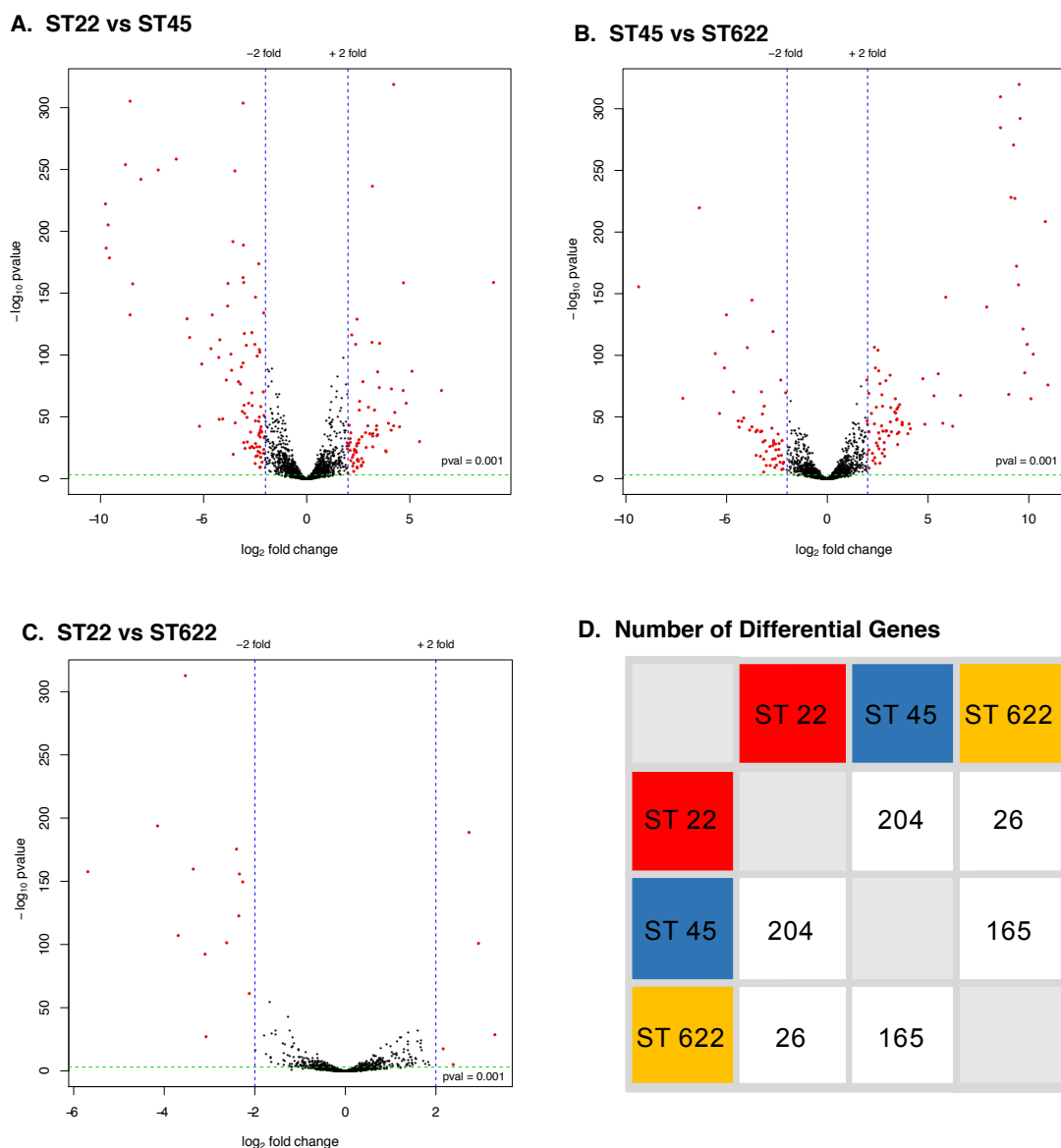
Figure 4.21 | Pairwise Comparisons of Transcript Expression. Volcano plots highlighting all significant DE genes (red). Genes are called significant if they have an FDR of <0.001 green dashed line) and a log fold change of ±2 (blue dashed lines). Volcano Plot highlighting significant DE genes between **A.** ST22 and ST45, **B.** ST45 and ST22, and **C.** ST622vsST22. **D.** DE gene matrix between STs: ST22vsST45 n=204, ST45vsST622 n=165, ST22vsST622 n=26.

The volcano plots above (Figure 4.21) show significance versus fold change in differentially expressed genes between STs. 204 genes (ST45 UP n=83, ST45 DOWN n= 121) were DE between ST22 and ST45 isolates and 165 genes (ST45 UP n=89, ST45 DOWN n=76) were DE between ST622 and ST45. 145 DE genes were shared between ST22 and ST622 isolates when compared to ST45 isolates. As visualised from the expression data (Figure 4.20), the transcriptional profiles of the ST22 and ST622 were highly similar. Overall, 107 CDS were shared between pairwise DE comparison of ST45 vs ST622 and ST45 vs ST22, of which 59 annotated genes were analysed with GO Enrichment to characterise the molecular functions of the DE genes between lineages (Figure 4.22).

A. ST45 ↓ (ST22/ST622 ↑)
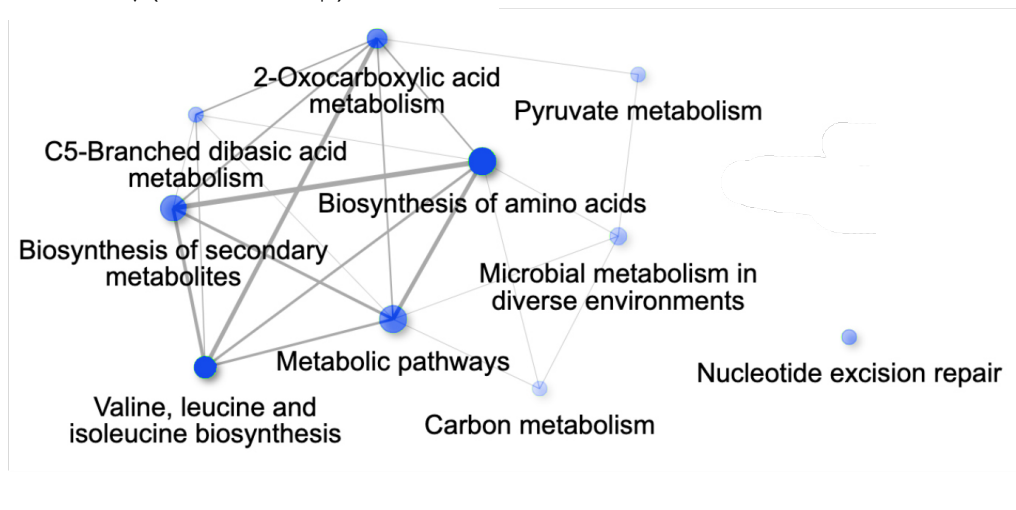
B. ST45 ↑ (ST22/ST622 ↓)
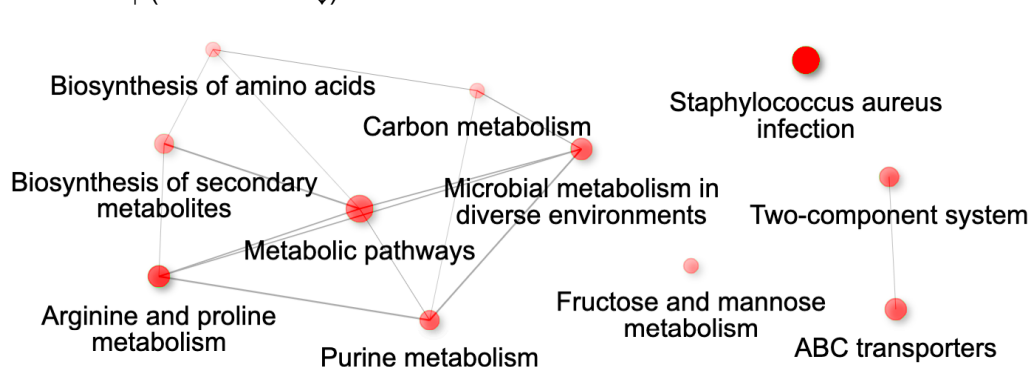
Figure 4.22 | GO Enrichment Analysis - KEGG Functional Pathways of DE genes between ST45 and ST22/ST622 isolates. A. Downregulated genes (n=28) in ST45 in comparison to ST22/ST622 isolates in major functional categories encompassing mostly metabolic and biosynthesis of amino acids (valine, leucine, isoleucine) and metabolites, **B.** Upregulated genes (n=31) in ST45 in comparison to ST22/ST622 isolates, mainly involved in metabolic functions and biosynthesis of amino acids and metabolites. Also see upregulated sugar metabolism (fructose and mannose), ABC transporters and two component systems as well as *Staphylococcus aureus* infection related genes.

Most DE genes were categorised into 'biosynthesis of metabolic pathways' clusters by KEGG Pathways database analysis, but genes in several molecular function families were distinctly upregulated within the ST45 isolates when compared to ST22/622 including purine metabolism, arginine and proline metabolism, fructose and mannose metabolism, ABC-transporters and two component systems as well as *Staphylococcus aureus* infection. Pyruvate metabolism, nucleotide excision repair, as well as valine and leucine biosynthesis were specifically downregulated within the ST45 isolates when compared to ST22/622.

Only 26 core genes (ST22 UP n=15, ST22 DOWN n=11) were differentially expressed between the ST22 and ST622 isolates, of which 18 were situated in the recombined genome region.

4.4.6.4 Chimeric Region - Differential Expression Analysis

The focus of the transcriptomic analysis was to characterise the transcriptomic profile and differentially expressed genes within the recombined chimeric region of the ST622, and the equivalent identical segments within the ST22, but more importantly the ST45 donor strains. As the 232 kb recombined sequence fragments within the ST622 and ST45 strains are identical, their comparison provides an ideal natural experiment to investigate whether differential *sau1* 6mA methylation has an effect on gene regulation, and ultimately epigenetics. Therefore, the levels of expression of genes within the chimeric sequence fragment and the equivalent in both ST22 and ST45 were compared between the 3 sequence backgrounds in a pairwise DE analysis seen in Figure 4.23. The gene lists with up and down regulated (UR, DR) genes, logFC, p-value and FDR scores can be found under Table 8.9 in the Appendix.

Differential expression analysis of pairwise comparison of the gene expression profiles of the chimeric region of ST22 vs ST45 isolates revealed 25 gene expression differences, 10 genes UR (logFC: mean 2.659, max 3.695, min 2.046), and 15 genes DR (logFC: mean 3.169, max -2.057, min -5.693,) in ST45 compared to ST22 as seen in Figure 4.22 A. Between ST622 and ST22, 18 genes were differentially expressed, 4 genes DR (logFC: mean -2.847, max -2.306, min logFC: -3.411), and 14 genes UR (logFC: mean 3.290, max 5.733, min 2.118) in ST22 compared to ST622 as seen in Figure 4.23 B. As marked in orange, 17 of the DE genes of varying functions and same expression patterns shared between the ST22 vs either the ST45 or ST622 isolates.
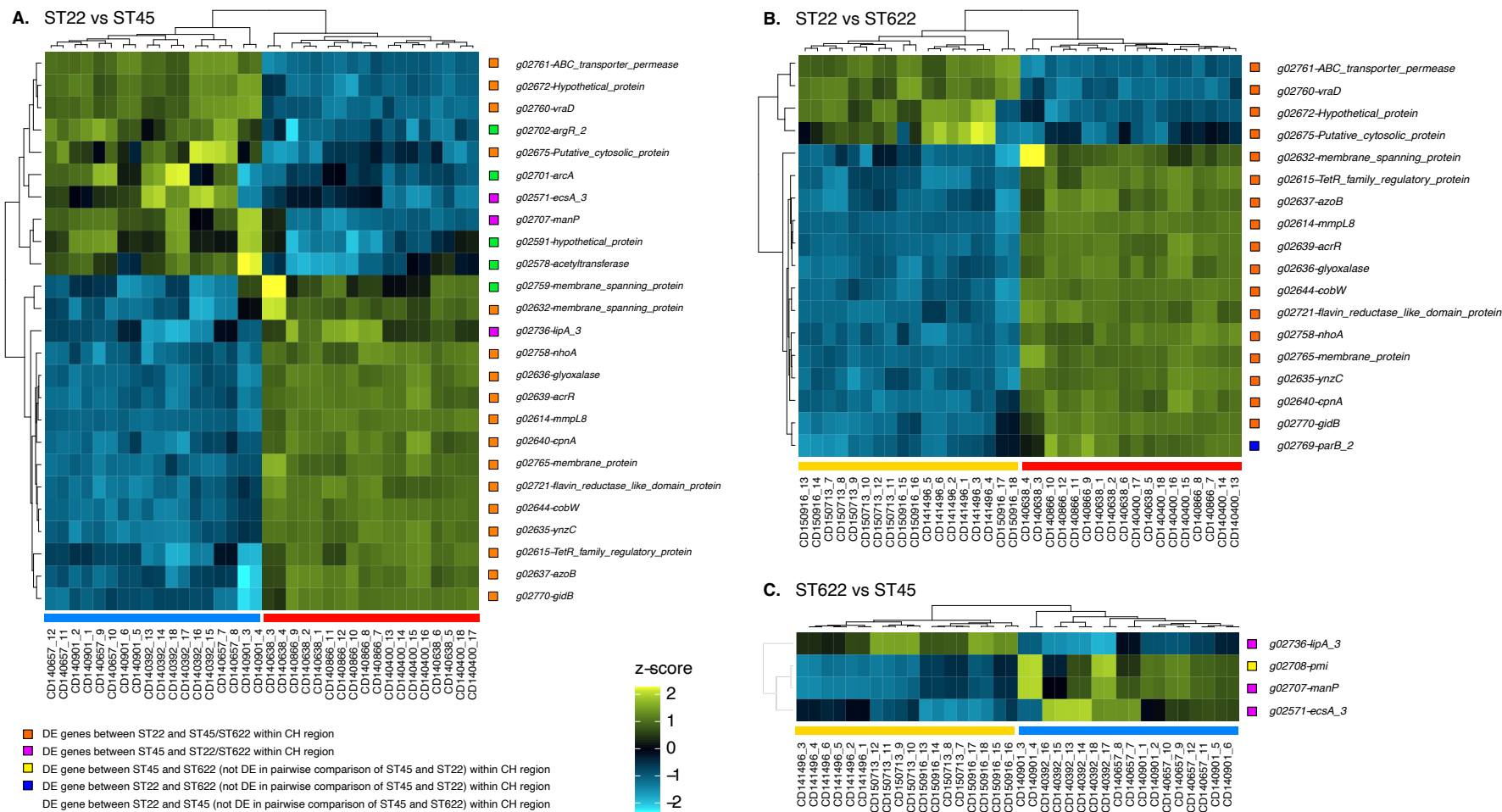
Figure 4.23 | Differentially Expressed Genes within the Chimeric Region for 3 ST types (Z score transformed expression values).
Z-score for the expression of a gene is obtained by subtracting the mean of all transcripts (for said gene) and dividing that by the standard deviation. The represented Z score of a gene is how many standard deviations the expression level for said gene is from the mean of all transcripts (how different it is from the overall observation). Scale or blue to yellow equals lower to higher gene expression. Coloured column footers indicate samples within ST22 (red), ST45 (blue) and ST622 (gold). DE genes from pairwise comparison of **A**. ST22 and ST45 (n=25). **B**. ST22 and ST622 (n=18), **C**. ST622 and ST45 (n=4). DE genes present across comparisons are marked with colour boxes (orange: ST22vsST45/ST622; magenta: ST622/ST22 vs ST45, yellow: ST45vsST622, blue: ST22vsST622, green: ST22vsST45.

It is unlikely that the set of DE genes for the chimeric region between the ST22/ST622 and ST22/ST45 isolates are a result of differential methylation, as the methylation signatures of ST22 and ST622 are nearly identical. Although differences in positions of 6mA modifications were recorded between the ST22 and ST622 chimeric region (Table 4.12), 2 TRS were within the 17 genes described, with only 1 TRS being in a regulatory region potential altering transcription (02762_hypothetical protein). Rather it may be influenced by sequence divergences (Figure 8.1 Appendix) introduced between the sequence backgrounds.

The motivation of this study was the comparison of identical sequences from the same ST background with differing methylation motifs, represented by the natural comparators: ST45 and the recombined hybrid ST622. The transcriptional profile of the CH region of the ST622 to the equivalent ST45 sequence are highly similar in expression levels, differing only in 4 genes (DR in ST45 vs ST622: g02736_*lipA*_3 (triacylglycerol lipase), UR in ST45 vs ST622: g02571_*ecsA_3* (ABC transporter*),* g02707_*manP* (sugar phosphotransferase), g02708_*pmi* (mannose-6-phosphate isomerase) which were differentially expressed between the two ST types (Figure 4.23 C). Three of the DE genes (g02736_*lipA*_3, g02571_*ecsA_3* and g02707_*manP* – Figure 4.23 A) between the ST622 and ST45 isolates were also differentially expressed between ST22 and ST45. As the sequence background of the chimeric sequence region between ST45 and ST622 are homologous, rather than sequence divergence, any differences in gene expression for the 4 distinct genes may likely be due to direct effects of differential methylation pattern. To be able to distinguish the cause of the differential expression patterns for the selected genes, detailed transcriptional analysis and promoter predictions were conducted.

4.4.6.4.1 Promoter, UTR and Transcriptional Start and Termination Site (TSS & TTS)
Prediction of 4 DE Genes between ST45 and ST622

To be able to distinguish whether the differentially expressed genes within the chimeric region of the ST622 and identical sequence region within the ST45 isolates were induced by sequence divergence or were the result of epigenetic regulation by differential methylation, the sequence identity and mRNA transcriptional sequence regions (promoter, UTRs, TSS and TTS) were analysed. Mutations within regulatory regions including the transcription factor (TF) binding site within the promoter, or SNPs within the transcriptional start or termination site could result in obstruction of transcription as the polymerase cannot attach, hindrance of transcription due to lack of transcriptional initiation at TSS, or truncated or elongated mRNA transcript due to premature stop or lack of transcriptional termination. Similarly, DNA methylation within these regulatory regions could delay or hinder transcription by interfering with the DNA polymerase enzyme.

To exclude sequence divergence, the sequence of the 4 genes were compared for the ST45 and ST622 isolates. The nucleotide sequence homology of g02736_*lipA_3,* g02707_*manP* and g02708_*pmi* including the flanking intergenic regions of said genes was 100% identity between the ST45 and ST622 isolates. The sequence homology for *ecsA_3* (including downstream CDS (transporter)) was 100% between ST622-2015 and ST45. As *ecsA_3* is around the recombinant break of the CH region, there were slight sequence variation (6 SNPs, 1 SNP within the 5'UTR, and 5 SNPs within 696 bp CDS, translating to 4 AA changes within the 231 AA protein sequence at 70 (N→S), 140 (E→K), 223 (D→E), 227 (K→R)) between the 2014 ST622 variant and the ST622 variant and the ST45/ST622-2015 isolates with 99.2% sequence identity (regardless which variant, *ecsA_3* was differentially expressed between the ST622 and the ST45 isolates). The identical (or near identical) sequence suggests that the differential expression patterns are not associated with sequence variation for the flagged genes.

Consequently, the promoter regions, polymerase binding sites, the mRNA transcript lengths including the 5' and 3' untranslated regions (UTRs) and transcriptional termination regions (TTR) for each of the 4 DE genes were mapped (Figure 4.24). The 6mA methylation motifs recognised by all HsdS from the ST45 and ST622 were also mapped to investigate whether regions involved in site-specific competition with other DNA binding proteins at regions linked to transcriptional/translational regulation facilitating epigenetic control of gene expression.

As seen from Figure 4.24 the location of 6mA modified bases and TRS motifs recognised by ST45 (green blocks) and ST622 (aqua blocks) HsdS differed when mapped onto the DNA sequence of each of the 4 differentially expressed genes. Although the same sequence region is differentially methylated, none of the methylation events were located on or within close proximity (<100 bp) of either promoter TSS or TTR indicating no hinderance/delay in transcriptional initiation or termination respectively. There were no methylation motifs within the predicted mRNA 3' or 5' untranslated regions, suggesting 6mA methylation within UTRs is not functionally relevant and are not involved in downstream post-transcriptional regulation. This indicates that 6mA methylation facilitated by the Sau1 system does not have a role in differential gene expression by site-specific competitive binding of transcriptional regulatory regions.

Each of the 4 DE genes were methylated within the CDS region, containing either a ST622 motif or both ST622 and ST45 TRS. However, to date there is no evidence that methylation of the coding sequence at either adenine or cystine nucleotides has any effect on gene expression in the prokaryotic genomes.
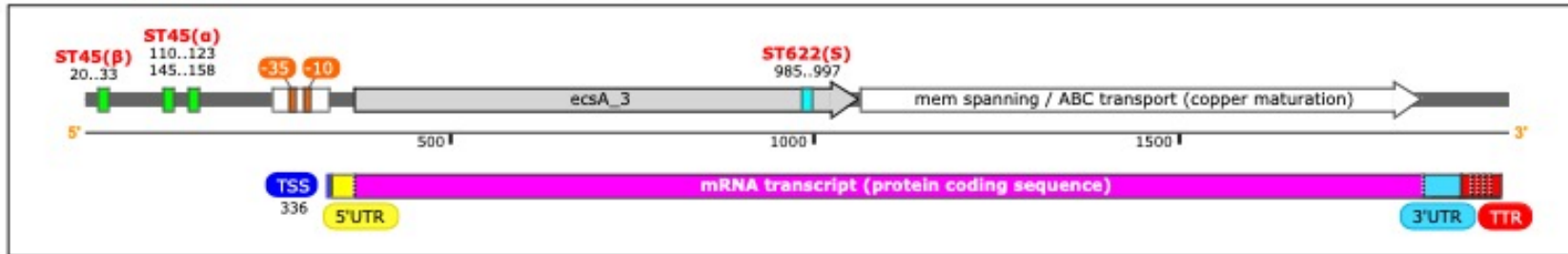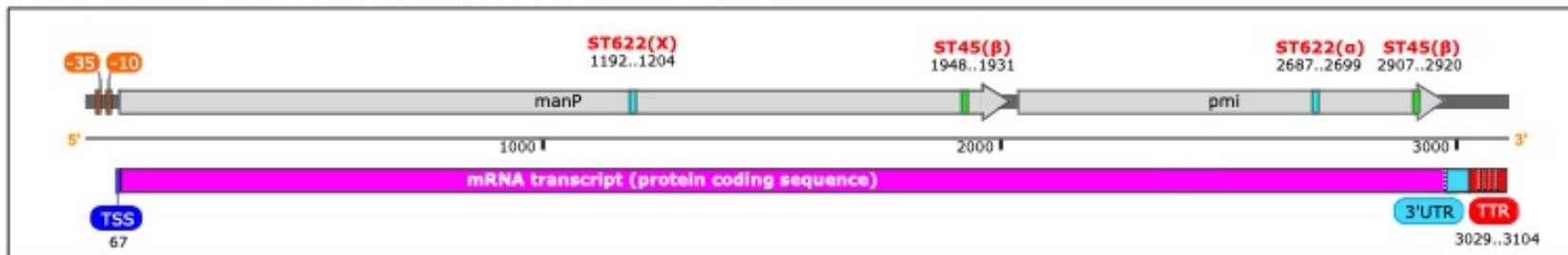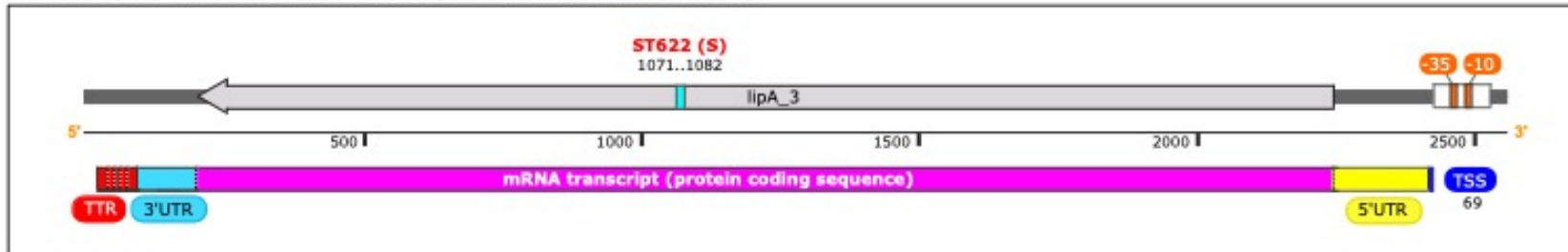
Figure 4.24 | Schematic of DNA sequences detailed with 6mA TRS locations and transcribed mRNA components for differentially expressed genes between ST45 and ST622 isolates (ecsA_3, manP, pmi, lipA_3). DNA sequences with the CDS (grey) of each of the 3 ORFs and promoter regions including the -35 and -10 binding sites (orange) for RNA polymerase sigma factor SigA, and 6mA DNA modifications from ST45 backgrounds (neon green) and ST622 backgrounds (aqua). The transcribed mRNA sequences are shown underneath each DNA sequence headed by the transcriptional start site (TSS – royal blue), in some cases by a the 5' untranslated region (UTR - yellow), preceding the translated CDS known as the protein coding sequence (magenta), followed by the 3' UTR (sky blue) before the intrinsic (rho-independent) transcriptional termination region (TTR) forming a hairpin structure (stem loops marked in red) on the elongating transcript disrupting the mRNA-DNA RNA polymerase complex. A. *ecsA_3*, B. *manP* and *pmi*, C. *lipA_3*. All DNA sequences are from CD140392 with the corresponding CDS tag and sequence region detailed in the title of each segment.

4.5 DISCUSSION

Largescale genome recombination events are rarely seen in clonal species like *S. aureus,* and little is known about the biological impact of the introduction of chromosomal replacements have on chimeric strains (Everitt et al., 2014). This study aimed to compare the genome, methylome and transcriptome of novel chimeric ST622 strain, a hybrid of the two most prevalent HA-MRSA lineages in Singapore, ST22 and ST45 (Chow et al., 2017). The curated collection of ST622 variants and a representative group of phylogenetically closely related natural comparators from 'parental' ST45 and ST22 MRSA strain backgrounds, provided a unique opportunity to study how largescale genome rearrangements affect the methylation landscape within *S. aureus*. Additionally, it also allowed the exploration of the potential for differential methylation and resulting possible epigenetic regulation induced by *sau1* TI RM within *S. aureus.*

4.5.1 Large-scale Core Genome Rearrangement of ST622

Large-scale chromosomal replacements are seen rarely in clonal species like *S. aureus,* where most allelic divergence is attributed to mutation-driven variation of the core genome rather than recombination (Feil et al., 2003; Lindsay & Holden, 2004, Everitt et al., 2014). Nevertheless, several highly successful, endemic MRSA strains have arisen including ST239, ST71, ST34, ST42, CC398, and ST2249, in all of which the recombinant DNA segments (or most of) span the origin and terminus of replication (Robinson & Enright, 2004; Holden et al., 2010; Smyth et al., 2010; Budd et al., 2015; Spoor et al., 2015, Thomas et al., 2012; Nimmo et al., 2015). Similarly, the recombinant region of the two variants of novel ST622 chimeric strain characterised within this study, also encompassed both the *oriC* and terminus of replication (Figure 4.6), which have been linked to greater levels of homoplasy, along with elevated recombination rates close to mobile elements of their insertion sites (Everitt et al., 2014).

Although the mechanisms by which these large-scale recombination events occur is still largely unknown, it is most likely mobile element driven (Everitt et al., 2014), with cryptic transformation (Morikawa et al., 2012), generalised phage transduction and conjugative transfer having been hypothesised as candidate transfer mechanisms for CGT in *S. aureus* (Spoor et al., 2015; Lindsay & Holden, 2006). In other species including *V. cholera*

(Hochhut et al., 2000, Nonaka et al 2018), *S. agalactiae* (Crochet et al., 2008), *C. difficile* (Brouwer et al., 2013) and *E. faecalis* (Flannagan & Clewell, 1991, Whittle et al., 2006), conjugative transposons have been attributed to mobilising large sequence regions of chromosomal DNA.

Within this study, two variants of the ST622 strain were investigated. Only one single example of the ST622-2014 variant was collected during sampling period, being the first and only chimeric strain isolated in 2014. This variant contained a 366 kb ST45 genome segment spanning from 2.6 Mb towards the terminus of replication stretching downstream of the SCC*mec* (acquiring the ST45 Type V SCC*mec* switching from the ST22 Type IVh) as seen in Figure 4.3 and Figure 4.5. During subsequent sampling years, no other example of the ST622-2014 variant was collected, but 16 and 26 samples of a second variant (ST622-2015) were sequenced in 2015 and 2016 respectively. The ST622-2015 variant contained a shorter, 233 kb recombinant DNA segment, with the recombination boundary around 2.5 Mb being identical to that of the ST622-2014. The second variant regained a 134 kb sequence fragment of the ST22 background downstream of the *oriT* encompassing the SCC*mec* region (Figure 4.6). This poses an interesting evolutionary question regarding the origin and the mechanism by which the new lineage has arisen, and why the ST622-2015 variant persisted within the circulating Singaporean MRSA population. It is uncertain if the two ST622 variants result from separate, and/or multiple recombination events (ST622-2014 further chromosomal shuffle succeeded by the 2015 variant), but extensive recombination and gene content variability has been attributed to exactly the staphylococcal chromosome cassette (Everitt et al., 2014). The SCC-associated recombination hotspot may have driven recurrent core genome rearrangements within the ST622 in a yet unknown mechanism. It is unlikely that 134 kb of ST22 core genome was regained through conjugative elements (phage, plasmid, transposon) as large-scale core genetic material would have needed to be carried and excised from the precise loci of the recombined segments, on multiple occasion (Robinson & Enright, 2004). However, lytic phage may have played an important role in the selective pressure contributing to the succession (becoming the prevalent variant) of the ST622-2014 by the ST622-2015, infecting and killing the host by circumventing the less complex RM barriers of the original variant (Samson et al., 2013; Safari et al., 2020; Hyman & Abedon, 2010; Orzechowska & Mohammed, 2019). Taking all this into account, it may be assumed that the ST622-2014 is likely the progenitor of the ST622-2015 variant, but further evolutionary studies are needed to confirm this and characterise the exact relationship between the two variants.

4.5.2 Comparative Genomic Analysis – ST622, ST22 and ST45

To explore the genetic variability of the ST622 strains and fully characterise the recombinant region, comparative genomic analysis with both the ST22 and ST45 parental strains were conducted. These studies revealed that the ST622 core genome, including the recombinant DNA segment, was 98% identical to both of the donor strains (Figure 4.4, Figure 4.6). In contrast to other hybrid *S. aureus* strains like ST239 and ST71 (Holden et al., 2010; Spoor et al., 2015), the newly introduced recombinant DNA segment did not result in the significant loss/gain of additional resistance and virulent determinants, other than the SCC*mec* switch from ST22 TIVh to ST45 TV (5C2&5) in the ST622-2014 variant, subsequently swapped back in the assumed successor ST622-2015 strain. There were no largescale gene insertions/deletions or duplication of genes, and no significant deviation in the uniform sequential arrangement of the conservative *S. aureus* core genome (Figure 4.6, Table 4.7, Table 4.8). Most of the genetic variability introduced between the 9 isolates is attributed to accessory genome, seen in the differential MGE content of each strain (Table 4.5, Figure 4.5, Figure 4.6). This was important to confirm as it allowed the downstream characterisation and comparison of both the differential methylome and transcriptome between these strains, as the identical homologous sequence regions could be compared.

4.5.3 6mA Sau1 Methylation Landscape – Singapore Collection

To date, no examples of differential methylation as a result of large-scale genome rearrangements have been reported. Exploiting the unusual natural recombination event within ST622, the main focus of the study was to analyse the effects CGT on 6mA methylation and potential epigenetic modulation of gene expression by TI RM Sau1. The Sau1 system landscape of ST622 was identical to ST22 with one core *sau1hsdMS* located in *ν*Saα in all variants, and two additional accessory TI RM – *sau1hsdS_orfX* (identical to previously characterised specificity element in CC8 NCTC isolates in Chapter 3 - Table 3.5) and *sau1hsdRMS* inserted within TV SCC*mec* only present in ST622-2015. Distinct 6mA TRS were predicted for the ST22-like (including ST622) and ST45 isolates using PacBio SMRT analysis, with >98% methylated/detected 6mA motifs correlating to the high modification ratio of Sau1 presented in the previous chapter (Table 4.9, Table 4.10).

On the whole genome level, there were no significant changes to the overall TRS detected between the ST22 and ST622 isolates (Table 4.10). There were no significant changes to the total methylation density, or the frequency of motif matches within the CDS/INT regions

by any of the three characterised Sau1 modification units between the parental ST22 and ST622 (Figure 4.8, Figure 4.9). *In situ* analysis of ST45 methylation motifs within the ST22/ST622 backgrounds and vice versa, revealed no significant difference in methylation potential throughout the core/accessory genome nor in the CDS/intergenic of any genome region, regardless the sequence background or methylation motif (Table 4.11). This is important as it shows that the introduction of recombinant DNA, hypothetically, does not induce significant changes within the total frequency of motif matches, indicating no differential methylation in terms of overall motif matches throughout the whole genome.

Comparative analysis of the methylation frequency within the chimeric sequence (CH) region and the remaining backbone (BACK) chromosomal sequence, revealed that region of recombinant DNA in the ST622, and the equivalent genome region in the ST22 and ST45 isolates are methylated at a lower frequency that the backbone sequence which on average had 33% more frequent matches over the three ST types (Figure 4.11). When specifically looking at the ST22 and ST622, the backbone (BACK) contained 16% and 7.5% higher TRS matches than the chimeric region (CH). The recombinant DNA region had 8.5% higher motif match rates than the equivalent sequence region within the ST22, equating to ~7 additional motifs across the 232 Kb string (Figure 4.11). The additional motifs were linked to increased methylation of the CDS solely attributed to HsdS_α matches, as matches from both HsdS_S and HsdS_X were decreased compared to the equal sequence in ST22 (Figure 4.11, Figure 4.12, Figure 4.13, Figure 4.15, Figure 4.14). Upon closer investigation of the methylation motif differences at the specific genomic positions across the recombinant genome region, it was revealed that although on average the ST622-2015 isolates only acquired 6 additional methylation motifs in comparison to the ST22, in reality there were 28 motif event differences between the two STs (Table 4.12), of which 18 were in regulatory regions of CDS. This is significant, as it indicates that the ~2% sequence divergence introduced within the chimeric region within the ST622 background incurs methylation differences at defined locations, with 2/3 lying in regulatory regions which may lead to potential changes in gene expression (between ST622 / ST22 isolates).

Synonymous SNPs introducing base rearrangements within target recognition sites located within the regulatory regions, specifically the untranslated regions (5' and 3' UTR) of a CDS, may alter the methylation status of the given region (Figure 4.10). This in turn may alter the binding potential of transcription factors to promoter, or potential steric interference at transcriptional start and stop sites, possibly modifying the transcription of

the given locus (Tourancheua et al., 2020; Weyand & Low, 2000; Polaczek et al., 1998; Casadesus, 2016; Low et al., 2001). This has been established to be particularly significant in the context of transcription of regulatory proteins (Shell et al., 2013), including transcription factors altering regulatory cascades (*rpoN* – *C. jejuni* - Ghatak et al., 2020) or altering the activity of activator (*traJ* – *S. enterica* – Camacho & Casadesus, 2005) and repressor genes (Gammaproteobacteria – *lrp, oxyR* – Casadeuss & Low, 2006; Wion & Casadesus, 2006; Cota et al., 2012).

4.5.4 The Role of Differential 6mA on Gene Expression

The main aim, and the most unique part of this study was the investigation of the potential epigenetic effect of Sau1 facilitated differential 6mA methylation within hybrid ST622 strains. The introduction of recombinant DNA from ST45 background into the ST22 backbone allowed the comparison of ST22 and ST45 methylation signatures throughout a large core genome segment from a single sequence background. Whole genome RNA sequencing of isolates from all three STs, and subsequent differential gene expression analysis of genes within chimeric sequence region, revealed little evidence for TI RM 6mA methylation having a transcriptional regulatory role. Four genes with metabolic functions (*lipA_3, ecsA_3, manP* and *pmi*) were differentially (± 2.7 logFC) expressed between the ST622 and ST45 isolates throughout the identical segment of DNA (Figure 4.23). However, cross-referenced with the motif analysis spanning the recombined region (Table 4.12), no methylation motifs, from neither ST45 nor ST622 signatures, were found within the regulatory regions of any of the given genes, indicating that direct moderation of gene expression through overlapping 6mA methylation and transcription factor binding motifs is unlikely.

6-methyladenine modification by non-phase variable TI RM systems in *E. coli* (Fang et al., 2012) *M. tuberculosis* (Shell et al., 2013), *S. pyogenes* (Nye et al., 2019), *B. subtillis* (Nye et al., 2020) and Mycoplasma species (Lluch-Senar et al., 2013) have been shown to elicit site-specific competitive binding within promoter regions outcompeting other DNA-binding proteins (repressors/initiators and TF), resulting in phenotypic changes and modulation of virulence. Although the direct effects on promoter binding were not observed for Sau1, 6mA methylation within *S. aureus* may have an indirect regulatory effect, in which DNA modifications in distal regulatory regions or modification of transcriptional regulators may cause differential expression of a cascade of genes. This has been described for loss of

TI RM 6mA (not differential methylation) within *S. pyogenes* (Nye et al., 2019) and *P. aeruginosa* (Dobrenze et al., 2017) where indirect methylation of transcriptional regulators Mpa and PrrF1, perturb the downstream regulation of virulence and iron metabolism associated genes within the respective species. This epigenetic regulatory mechanism has also been extensively studied for orphan methyltransferases Dam *in E. coli* (Camacho and Casadesus, 2002) and CcrM in *C. crescentus* (Gonzalez et al., 2014; Zhou et al., 2015).

4.5.5 Limitations and Future Work

One point of consideration for this study is the comparison of gene expression data within the 223 kb ST45 derived DNA, in two different strain backgrounds. Although only 4 genes were statistically differentially expressed between ST622-2015 and ST45 isolates, the conclusion that these were due to indirect or direct effects of differential methylation should be done with caution. While the core genome segment may be the identical sequence, the transcriptional regulatory network within each strain is unique, as is the metabolic and physiological state at any given moment regardless if growth phases are kept the same (Capra et al., 2017; Osmundson, Dewell & Darst, 2013; Chaves-Moreno et al., 2016). In essence this experiment was a naturally created ad-hoc opportunity to preliminarily investigate the regulatory potential of differential Sau1 6mA modifications throughout an homologous piece of *S. aureus* core genome.

Evidently, creation of knockout *hsdS* alleles to study the transcriptomic effects of loss of 6mA methylation throughout the *S. aureus* would greatly benefit this study to validate the marginal gene expression effects potentially attributed to differential methylation. Creation of isogenic mutants with deletion of either of the two core *hsdS*, in the genomic islands of ST45, as well as the accessory *hsdS, hsdS_X* and *hsdS_S* (ST22/ST622 background) could potentially shed light on functional difference between methylation systems (explored in the subsequent chapter).

Although this study did not address the phenotypic characterisation of the novel ST622 strain, conducting growth experiments to characterise the physiological diversity of the hybrid strain compared to the ST45 and ST22 parent strains would be beneficial to uncover the virulence of the strain. Further carriage and surveillance studies would be valuable to analyse the prevalence and epidemiology of the chimeric strain throughout intermediate

and long-term care facilities in Singapore. To determine whether ST622-2015 has established itself as the new dominant HA-MRSA strain, a rapid PCR-based assay (Supplementary Figure 8.2, Methods Table 2.18) was developed involving a pair of nonredundant primer pairs based on two variable genes, *nikB* within the ST22 backbone and *crtN* within the imported ST45 sequence region.

5. FUNCTIONAL IMPACT OF SAU1 FASCILITATED 6mA DNA METHYLATION IN *S. AUREUS*

5.1 INTRODUCTION

To date, several non-phase variable TI Restriction Modification associated with 6mA DNA methylation, *S. pyogenes* (Nye et al., 2019), *M. tuberculosis* (Shell et al., 2013) and *P. aeruginosa* (Dobreneze et al., 2017) with *hsdRMS* systems have been speculatively linked to having epigenetic regulatory effects. Although Sullivan et al., 2019, discuss potential (epi)genetic changes as a result of differential methylation correlated to partial recombination within *S. aureus hsdS*, the impact of methylation changes and mechanism of suspected Sau1 epiregulation remain unclear (Sullivan et al., 2019). To gain a comprehensive understanding of the potential secondary regulatory function of TI RM systems within *S. aureus*, further downstream analyses investigating the functional impacts of Sau1 6mA DNA methylation throughout the whole genome are necessary. In the case of the TI Sau1 system, the most simple and efficient experimental approach is defined, single-gene mutagenesis of the *hsdS* unit, to compare the of methylation capacity (PacBio SMRT Modification and Motif Analysis) and transcriptome (RNA-Seq) of mutant versus wild-type parental strains. Without the specificity unit, there is neither methylation nor restriction activity within the cell as firstly, the M_2S and R_2M_2S complexes will not be able to form due to the lack of specificity unit. Secondly, due to the lack of DNA binding ability, neither complex will be able to carry out their function (Loenen et al., 2014).

The Singapore Collection introduced in the Chapter 4 includes strains with single, double and triple Sau1 *hsdS* units, which were ideal for the creation of mutant *S. aureus* strains to study the functional impact of loss of 6mA throughout the whole genome.

5.2 AIMS & OBJECTIVES

The aim of this chapter is to determine if 6mA methylation facilitated by Sau1 has an epigenetic gene regulatory role in *S. aureus*. Using the isolates presented in the previous chapter, this study aimed to

1. Create isogenic mutant strains by allelic replacement mutagenesis of the TI RM *hsdS* gene (*ΔhsdS*) in two different sequence backgrounds – ST622, and ST45.
    a. Creation of mutants fully devoid of 6mA methylation to study the effect of complete 6mA loss in within differing *S. aureus* backgrounds.
    b. Creation of *ΔhsdS* strains deleting either of the two core HsdS (HsdS_α or HsdS_β) to investigate potential biases in function.
    c. Creation of sequential *ΔhsdS* mutants to investigate the differential functional impact of the accessory HsdS (HsdS_X and HsdS_S).

2. Validate the genetic deletion of *hsdS* and loss of methylation of the corresponding TI methylation TRS using PacBio SMRT sequencing technology and Methylation and Motif Analysis.

3. Investigate transcriptomic changes resulting from loss of Sau1 6mA methylation between *ΔhsdS* mutants and WT strains using RNA-Sequencing and differential gene expression analysis.

5.3 ORIGINS OF COLLECTION

Several wild-type strains including CD141496 (ST622-2014), CD150713 (ST622-2015) and CD140392 (ST45) from the Singapore collection (detailed in the Chapter 4) were used as parent strains to create *ΔhsdS* mutants. The mutagenesis studies were carried out at the University of Tübingen within the Heilbronner working group within the laboratory of Prof Andreas Peschel. Eight of the nine mutants could be created in the 5-week exchange, except double knockout mutant *ΔΔhsdS orfX+α (*RM4+RM1*)* which was generously finished by Darya Belikova in the Heilbonner group. All mutant constructs were pre-prepared by Darya Belikova.

RNA sequencing was conducted at the Wellcome Sanger Institute. PacBio SMRT sequencing and methylation analysis were to have been also run by the Wellcome Sanger Institute but due to technical problems these samples were not processed for whole genome sequencing. Instead, a subset of 4 mutant strains including *ΔhsdS*α (RM2), *ΔhsdSβ* (RM3) and *ΔhsdSα+*β (RM2+3) from the ST45 background (parent strain CD140392), and *ΔhsdS*α (RM1) from the ST622-2014 background (parent strain CD141496), were sequenced by GeneWiz in their NGS facility in New Jersey – the sequencing data for these strains are not in the public domain. A complete list of bacterial strains and DNA sources has been provided in Table 2.8 and Table 2.9 in Methods section.

5.4 RESULTS

To be able to investigate the full mechanistic and functional importance of *sau1* methylation within the *S. aureus*, each *hsdS* gene within three different sequence backgrounds (ST45, ST622-2015, and ST622-2014) were deleted. Single and double *ΔhsdS* mutants were created to investigate the potential functional redundancy of either core *hsdMS* system within the ST45 isolates, and the accessory *hsdMS* systems within the ST622-2015 sequence background. Two mutant isolates of differing backgrounds were created with no working *hsdS*, resulting in no DNA binding ability, leaving the Sau1 RM system dysfunctional as shown in Figure 5.1.
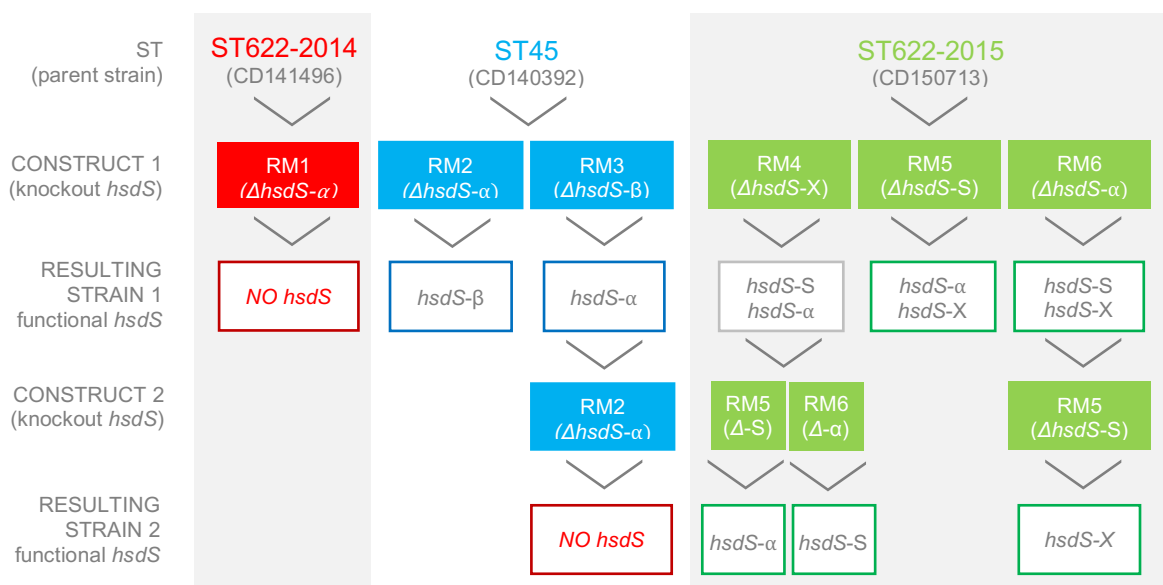
Figure 5.1 | Schematic of Mutant Isolate Experimental Design
Mutant isolates were created using deletion constructions (RM1-RM6) in three different parental strains each belonging to a different ST (ST45 – CD140392 (blue); ST622-2014 – CD141496 (red); ST622-2015 – CD150713 (green)).

The nomenclature used to describe mutant isolates according to the deletion construct (RM) as the following:

RM1	*ΔhsdS_α*	= no functional *hsdS*	ST622-2014
RM2	*ΔhsdS_α*	= functional *hsdS_ β*	ST45
RM3	*ΔhsdS_β*	= functional *hsdS_ α*	ST45
RM2+3	*ΔhsdS_α + ΔhsdS_β*	= no functional *hsdS*	ST45
RM4	*ΔhsdS_X*	= functional *hsdS_S* and *hsdS_α*	ST622-2015
RM5	*ΔhsdS_S*	= functional *hsdS_X* and *hsdS_α*	ST622-2015
RM6	*ΔhsdS_α*	= functional *hsdS_S* and *hsdS_X*	ST622-2015
RM4+5	*ΔhsdS_X* and *ΔhsdS_S*	= functional *hsdS_ α*	ST622-2015
RM4+6	*ΔhsdS_X* and *ΔhsdS_α*	= functional *hsdS_ S*	ST622-2015
RM6+5	*ΔhsdS_α* and *ΔhsdS_S*	= functional *hsdS_ X*	ST622-2015

5.4.1 Mutant Validation

Mutants were constructed via molecular cloning experiments using overlap-extension deletion constructs of each *hsdS* gene and were introduced into the host cells via temperature sensitive cloning vector pIMAY. The absence of the deleted genes was firstly verified using PCR primers and colony PCR, followed by Sanger sequencing of genomic DNA as described in Methods under section 2.9.3.7 and seen in Supplementary Figure 8.3 detailed in the Appendix. Secondly, the mutant isolates were subjected to whole genome sequencing and methylation analysis via PacBio SMRT analysis to validate the absence of the deleted specificity genes. Due to the delay in sequencing time and COVID-19, only 4 out of the 9 mutant strains (RM1, RM2, RM3, RM2+3) were sent off for PacBio SMRT whole genome sequencing. WGS confirmed the mutant genotypes, as seen in Figure 5.2 showing the absence of the *hsdS* genes of interest within each mutant isolate compared to the WT strains. The deletion of single *hsdS* with RM1 (Δ*hsdS*_α - Figure 5.2 D), RM2 (Δ*hsdS*_α - Figure 5.2 A) and RM3 (Δ*hsdS*_β Figure 5.2 B), as well double knockout of *hsdS* within RM2+3 (Δ*hsdS*_α and Δ*hsdS*_β - Figure 5.2 C) were successful. PacBio SMRT Methylation and Motif analysis also confirmed the loss of methylation motifs corresponding to the deleted *hsdS*, with no 6mA methylation in RM1 and RM2+3, and single 6mA motifs for RM2 and RM3 corresponding to functional *hsdS*_β and *hsdS*_α. This allowed the functional characterisation of either core *hsdS* (*hsdS*_α or *hsdS*_β) within the ST45 background as well as characterisation of isolates with no functional Sau1 RM system within an ST45 (RM2+3) and ST622 background (RM1).

Although not all isolates could be genome sequenced, each isolate was RNA-sequenced and the sequence reads were mapped to the reference WT parent genome to compare differences in mRNA transcript levels and coverage on a gene by gene level. From the PCR validation and the RNA-Seq data we were able to infer the deletion of the *hsdS* genes of interest from the absence of mRNA transcript for the given CDS. Figure 5.3 illustrates stacked read-alignment data (.bam) of the mRNA transcript reads of each mutant and WT strain mapped to the whole genome sequence of the WT parent strains. All 9 mutant strains (ST622-2014: RM1 (*hsdS*_α - Figure 5.3 A), ST45: RM2 (*hsdS*_α), RM3 (*hsdS*_β), RM2+3 (*hsdS*_α+β), ST622-2015: RM5 (*hsdS*_S), RM6 (*hsdS*_α), RM4+5 (*hsdS*_X+S), RM4+6 (*hsdS*_X+α), RM5+6 (*hsdS*_S+α)) had correctly mutagenized *hsdS* genes, which were fully deleted as seen from the lack of RNA transcripts mapped to the whole genome assembly for the given CDS.
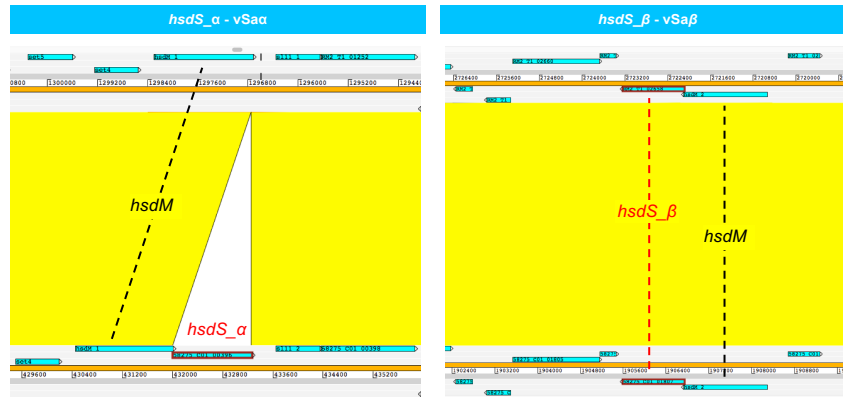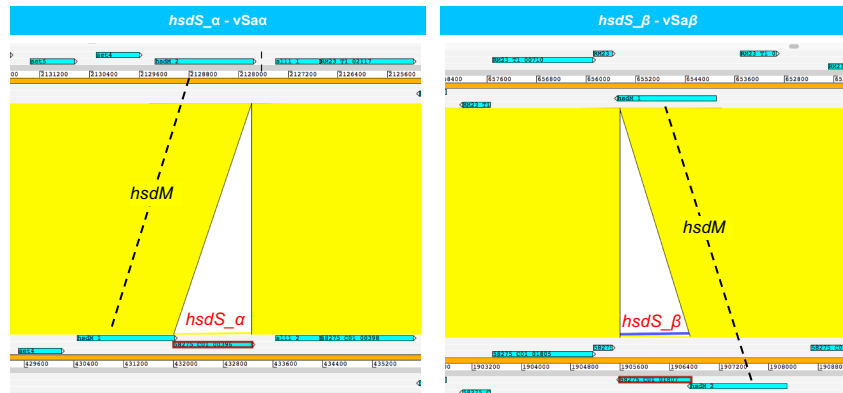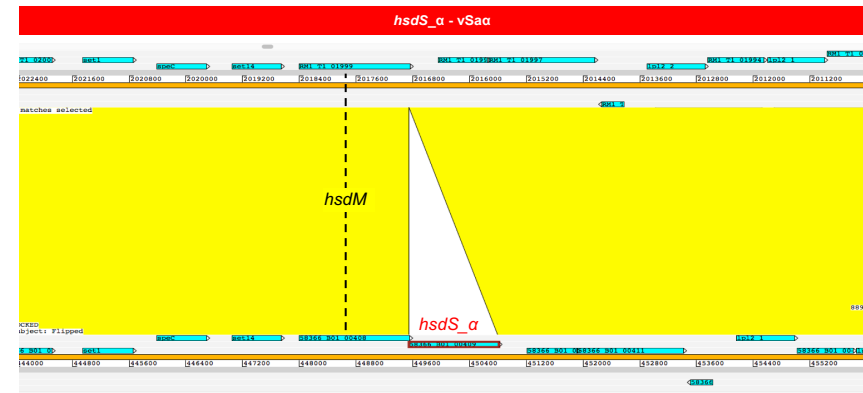
Figure 5.2 | Absence of *ΔhsdS* CDS compared to wild type (WT) reference strains visualised in ACT (Artemis Comparison Tool). WT strains (ST45 – CD140392 for blue headers; ST622-2014 – CD141496 for red header) are seen as the bottom and mutant strains files are the top genome panels. Aquamarine blocks indicate CDS along each genome panel, yellow block colour indicates 100% sequence identity, whilst white triangle blocks indicate absence of CDS between the two compared genomes. **A**. RM2 strain with *ΔhsdS_α,* showing clear deletion of *hsdS_α,* and retention of *hsdS_β* in *v*Saβ **B.** RM3 with *ΔhsdS_β* showing clear deletion of *hsdS_β,* and retention of *hsdS α* in *v*Saα **C.** *RM2+3 double knockout ΔΔhsdS α+β showing deletion of both hsdS*, **D**. RM1 with *ΔhsdS α* showing clear deletion of *hsdS_α* in *v*Saα.

218
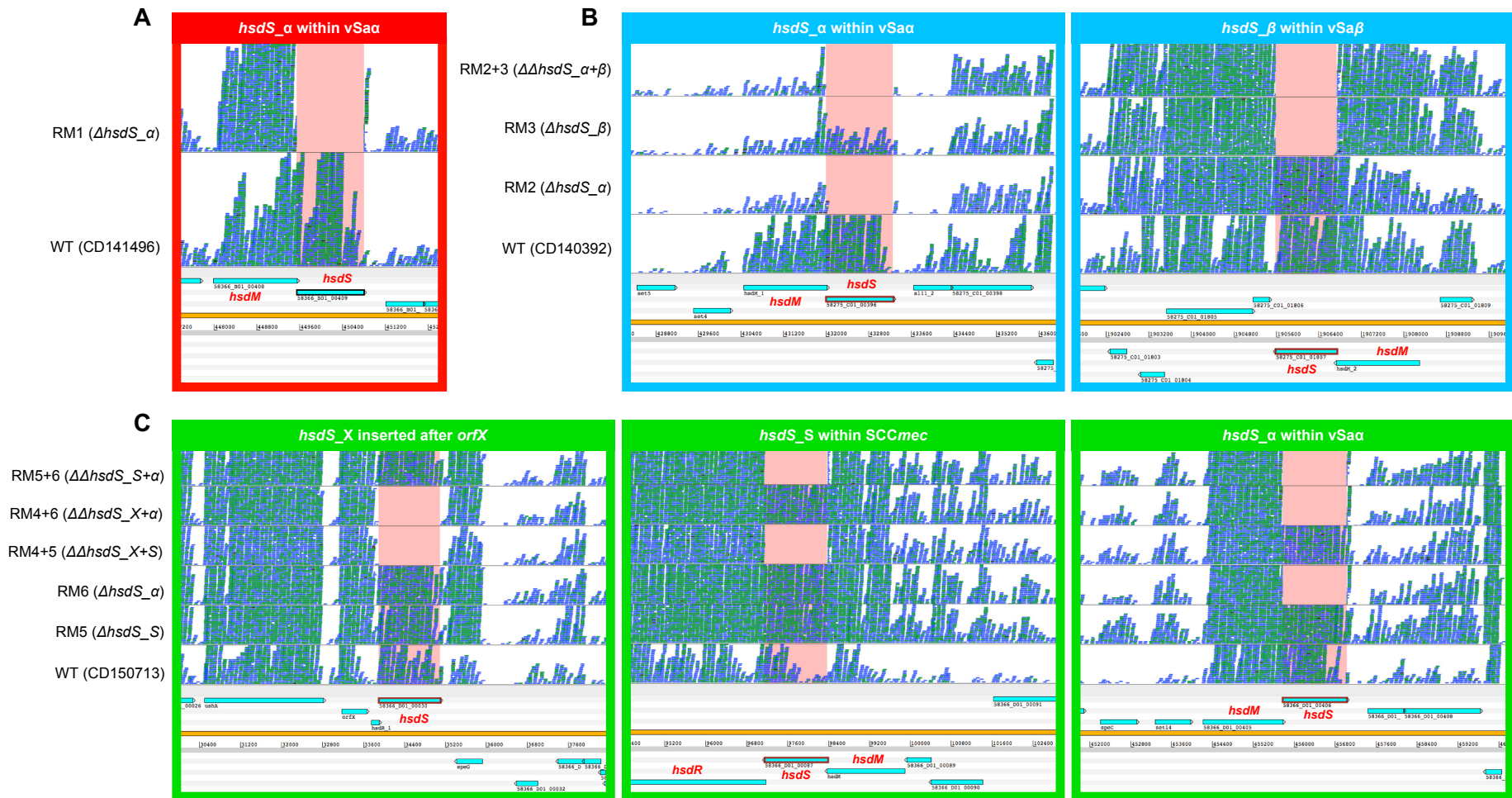
Figure 5.3 | Absence of RNA transcript reads for *ΔhsdS* genes compared to wild type (WT) reference strains visualisation of read-alignment data in Artemis. WT parent strains are seen in the first window (bottom) above the reference genomes. Green/blue stacked blocks are mRNA transcript reads aligned to the whole genome assembly. The pink block highlights the sequences region of the *hsdS* coding sequence; deleted genes have no RNA reads resulting in a clear block in the alignment. **A.** Aligned mRNA transcript levels for *hsdM* and *hsdS* genes within the νSaα for the ST622-2014 WT stain CD141496 and *ΔhsdS α (RM1)*, **B.** aligned *mRNA transcript levels for hsdM and hsdS for both the νSaβ (right) and νSaβ (left) for ST45 WT strain CD140392 (bottom window) and 3 mutant strains (bottom to top) ΔhsdS α (RM2), ΔhsdS β (RM3), double knockout ΔΔhsdS α+β (RM2+RM3),* **C.** aligned mRNA transcript reads for ST622-2015 *hsdS* (and *hsdM*) for the *orfX* insert (left), the SCCmec region (middle) and νSaα (left). CD150713 was used as a parent strain to create mutants including: *ΔhsdS S (RM5), ΔhsdS α (RM6)*, *ΔΔhsdS orfX+SCC (RM4+RM5), ΔΔhsdS orfX+α (RM4+RM6), and ΔΔhsdS α+SCC (RM6+RM5).*

5.4.2   Differential Expression and Methylation Analysis

To gain a deeper understanding of the functional importance and potential epigenetic regulation facilitated through Sau1 6mA, the complete transcriptomes of all mutant strains were sequenced with RNA-Seq. Transcripts were assembled and mapped to WT strains, and differential expression (DE) analysis was conducted. The following results compare each RM mutant RNA-Seq transcripts to give insight into the epigenetic potential of 6mA methylation within *S. aureus*, and any functional variance between differential methylation patterns pertaining to either core, or accessory HsdMS complex. The bacterial strains for both the WT and RM experiments were grown in identical conditions (overnight growth, 37ºC, 160 RPM in TSB rich medium), harvested at late-log phase growth according to predetermined timeframes with preliminary growth assessment (see Figure 8.4 in Appendix for growth curves of WT vs RM mutants) and RNA sequenced at the Wellcome Sanger Institute using Illumina-C HiSeq 4000.

The methylation patterns of WT strains and mutant strains were also compared using PacBio Methylation and Motif Analysis and visualised in Artemis, to further investigate the methylation states of genes highlighted in the DE analysis.

5.4.2.1   WT CD141496 vs RM1 (*ΔhsdS_*α)

Wild-type strain CD141496 (CC622-2014) only contained one functional *hsdS*, carried in *v*Saα. By creating an *ΔhsdS_*α (RM1), both the 6mA methylation and restriction activity by the single operational Sau1 RM was removed entirely. Hence the main focus of this experiment was to investigate the functional effect of removing 6mA methylation (and restriction activity) in *S. aureus* by comparing the transcriptome of WT and RM1 mutant strain.

5.4.2.1.1 DE Analysis - WT CD141496 vs RM1 *(ΔhsdS_*α)

The RNA-Seq transcript data was first normalised according to sequencing yield of each sample for both data sets (WT and RM1) to standardise the count data used for downstream differential gene expression analysis. The TMM normalised transcript data show a slightly wider distribution of count levels, indicated by higher upper quartile and whisker, for the WT strain (CD141496) than RM1 (*ΔhsdS_*α) mutant as seen in Figure 5.4 A. This indicates slightly higher overall transcript yields for the WT strains over the RM1 mutant samples which had lower quantities of raw RNA material as a substrate for RNA-Seq needing multiple rounds of PCR amplification prior to sequencing.

To evaluate the variation among RNA-Seq samples, a multidimensional scaling plot was created to visualise dissimilarities (distances) between samples. Figure 5.4 B shows the WT (green) and RM1 strains (red) cluster in two distinct groups along dimension 1 (logFC -2/2). Biological replicate RM1_1 and technical replicate RM1_2 show dispersion along dimension 2 (logFC -0.2/0.6) compared to rest of the RM1 replicates, indicating small levels of transcript level differences. The increase in distance between the RM1_1/RM1_2 and the rest of the clustered RM1 replicates was found to be +2 fold higher transcript levels for 21 randomly distributed genes with differing functions (Appendix - Supplementary Table 8.10). The genes were included in the downstream analysis as they represent biological variation between samples.
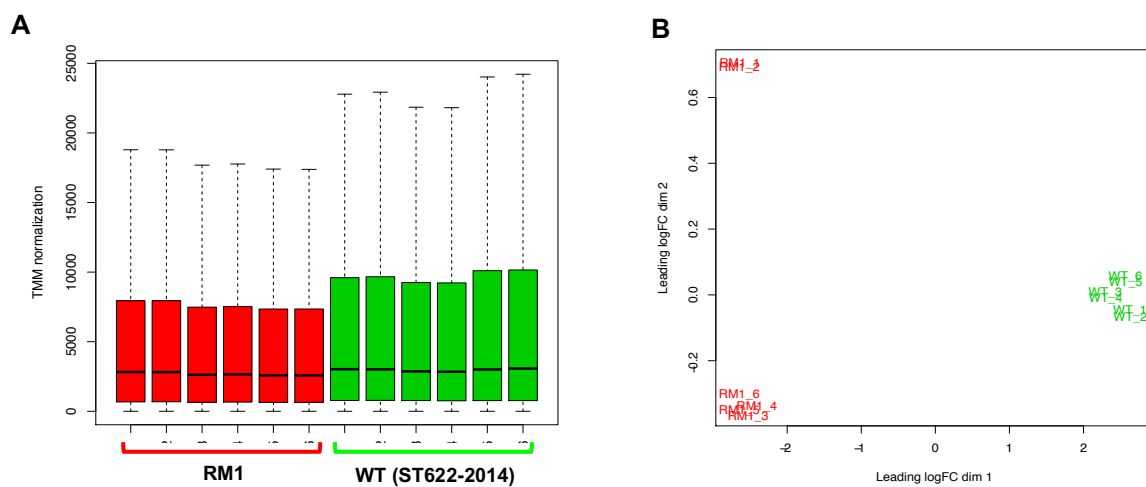
Figure 5.4 | Normalised Transcript Count Data ST622-2014.
A. Transcript count distribution boxplot of TMM normalised count distribution per sample – normalised according to sequencing yield of each sample. Distribution of WT count numbers are wider compared to RM1 seen in higher upper quartile and whisker values. **B**. Multidimensional Scaling of TMM normalised count data – WT and mutant strains cluster in two distinct groups along dimension 1 (logFC -2/2) with RM1_1 and RM1_2 (top left corner) showing variability along dimension 2 (logFC -0.2/0.6) compared to rest of the replicates of this strain.

Differential expression analysis revealed 585 genes, 22% of all CDS within this strain, were differentially expressed (n=2660 - including hypothetical/conserved genes) between the WT and RM1 mutant strain (Figure 5.5). This suggests that Sau1 facilitated methylation (and restriction) may be involved in a wide range of biological and cellular processes. Out of the 585 differentially expressed genes, 193 genes could be assigned a function, with which gene set enrichment (functional enrichment) analysis was conducted. Genes involved in various functional categories including mainly metabolic pathways, ribosome function, ABC transporters, two-component systems, secondary metabolite biosynthesis among others were over-represented in the DE gene set as seen in Figure 5.6.
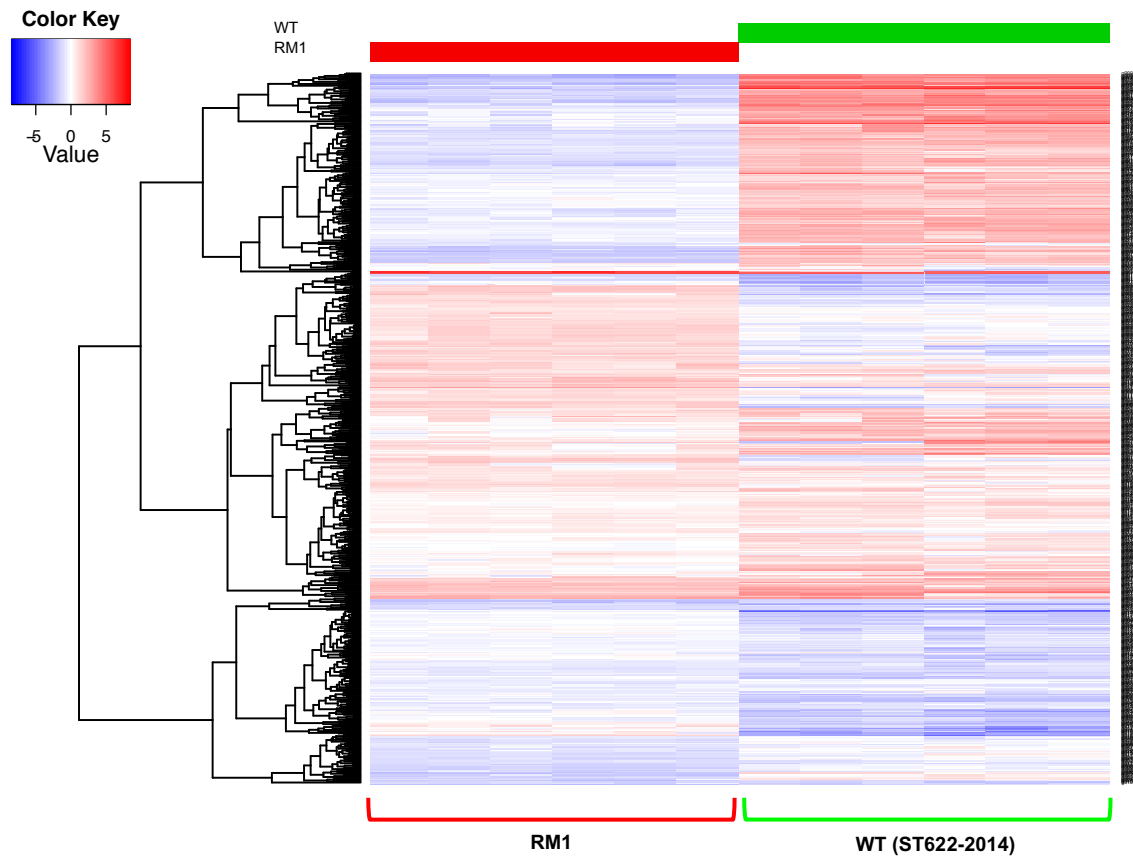
Figure 5.5 | Hierarchical Clustering of DE Genes vs ST622-2014 Samples.
Global comparison of log2 transformed centre-clustered expression profile for each gene (right column) between ST622-2014 WT sample CD141496 (green header) and mutant strain RM1 (red header). Three distinct clusters of genes with similar expression profiles within the WT replicates and RM1 mutant replicates. Heat map values were calculated by subtracting each gene medium log2 (RPKM) value from the log (RPKM) value of each sample – unregulated genes are red whilst down regulated genes are blue (edgeR > 4, FDR < 0.001, P-value < 0.001).
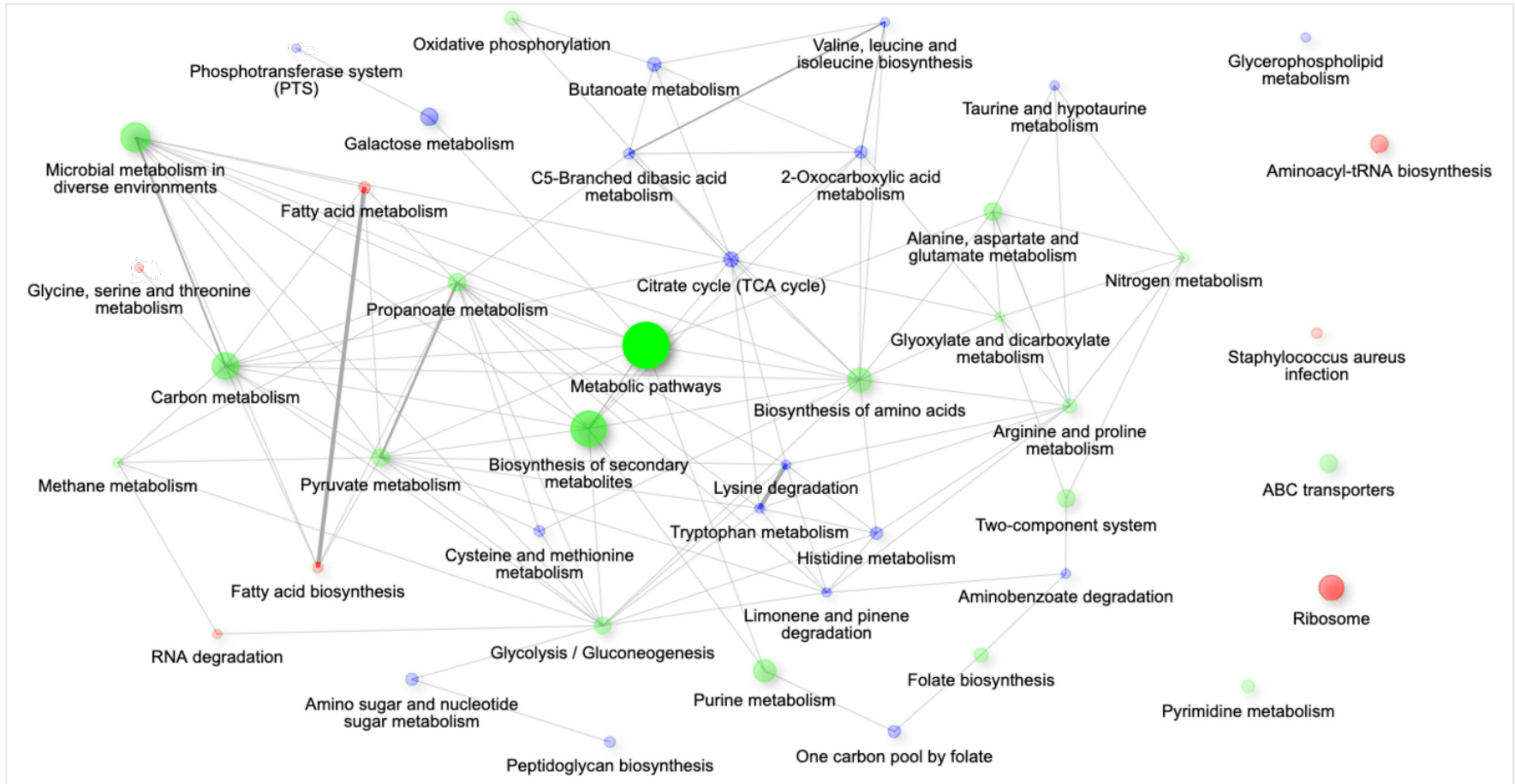
Figure 5.6 | KEGG Enrichment Plot Network Pathway of top 50 functional categories of annotated genes between RM1 (*ΔhsdS_α*) and WT strain CD141496. The enrichment of ShinyGo genes at a significance level of p=0.05, using a gene list of differentially expressed genes. Larger and more vivid bubbles indicate increased number of genes and lower Enrichment False Discovery Rates (FDR) within the category represented. **Green** bubbles indicate pathway in which genes are both up and downregulated within RM1 compared to WT (metabolic pathways, biosynthesis of secondary metabolites, biosynthesis of amino acids, ABC transport and oxidative phosphorylation, carbon metabolism), whilst **red** bubbles indicate upregulation in RM1 (ribosome functions, *Staphylococcus aureus* infection, fatty acid metabolism/biosynthesis, RNA degradation, glycine, serine and threonine metabolism, and aminoacyl-tRNA biosynthesis), and **blue** bubbles indicate downregulation in RM1 compared to WT strain (amino acid metabolism/degradation (various), sugar metabolism (amino and nucleotide sugar, galactose), citrate cycle, amino acid biosynthesis (various), peptidoglycan biosynthesis, one carbon pool by folate, metabolite degradation (various – aminobenzoate, limonene and pinene, lysine, glycerophospholipid metabolism).

Of the DE genes, 372 were downregulated, and 213 were upregulated in RM1 in comparison to the WT strain. Many up and down regulated genes belonged to the same functional pathway including various metabolic pathways, biosynthesis of secondary metabolites, biosynthesis of amino acids, two-component systems, oxidative phosphorylation, glycolysis and ABC transport (Figure 5.6). Genes involved in ribosomal function, fatty acid metabolism/biosynthesis, glycine, serine and threonine metabolism, RNA degradation and aminoacyl-tRNA biosynthesis, as well as *Staphylococcus aureus* infection were upregulated in RM1 compared to WT strain. Genes involved in the biosynthesis and metabolism of various amino acids, sugars, the citrate cycle (TCA cycle), peptidoglycan biosynthesis, and glycerophospholipid metabolism were downregulated in RM1 vs WT isolate.

The deletion of the single functional *hsdS* within the CD141496 background, consequently deleting all 6mA methylation and restriction activity, impacts a wide variety of metabolic and cellular processes. This could either be the result of pleotropic effect of knocking out Sau1 activity, or rather a consequence of differential methylation throughout the two representative genomes.

5.4.2.1.2 Methylation Analysis of DE Genes - WT CD141496 vs RM1 *(ΔhsdS_α)*

To understand whether the potential role of loss of 6mA methylation plays in the differential expression of the above characterised 585 genes between RM1 and WT parent strain, the methylation status of the regulatory regions as well as the coding sequences of each DE genes was investigated within the WT strains. As previously described in Chapter 4, no phase variation within the *hsdS* or *hsdM* units were found within the representative isolates, and there was little evidence that differential methylation of promoter regions had any epigenetic effect. However, deletion of 6mA methylation on a whole genome level in RM1 may give deeper mechanistic insights to epigenetic regulation and the role of 6mA in *S. aureus* gene expression.

WT strain CD141496 has one functional *hsdS* situated in *v*Saα*,* coding for HsdS_α, which recognises methylation motif **A**GG(N_6)YTC**A** present in 692 sites throughout the genome. 493 (71.24%) methylation motifs were within CDS; overall only 18.53% of the 2660 CDS within the WT strain were methylated. Out of the 493 methylated CDS recognised, only 97 genes (19.68%) were identified to be differentially expressed - 35 upregulated and 62

downregulated genes in RM1. No methylation motifs were found within promoter regions (+200 bp from TSS) or 5' and 3' UTRs. One example of methylation within the intergenic region between *atpF* and *atpE* was found, which Mäder et al., (2016) have found to be 'strictly intracistronic', so would likely not impact the transcription of either genes.

Only 16.58% (97/585) of DE genes between the WT and full *sau1hsdS* knockout contained a 6mA methylation site. None of the methylation events were within regulatory regions, or in regions previously shown to affect gene expression. Therefore, it is unlikely that the differentially expressed genes seen between the WT and *ΔhsdS_α* were due to epigenetic regulation but are rather the result of a global regulatory response on general central metabolism and protein synthesis.

To gain more insight on not just the effect of full 6mA deletion throughout the *S. aureus* genome, but also capture and better understand the role of differing *hsdS* within a strain, differential expression and methylation analysis was conducted on a set of mutants from different sequence backgrounds including ST45 (RM2, RM3, RM2+3) and ST622-2015 (RM5, RM6, RM4+5, RM4+6, RM5+6).

5.4.2.2   WT CD140392 vs RM2 (*ΔhsdS_*α), RM3 (*ΔhsdS_*β), RM2+3 (*ΔΔhsdS_*α+β)

WT strain CD140392 is a ST45 isolate containing two functional *hsdS*, one in each genomic island. Three separate mutant strains were created: *ΔhsdS_*α (RM2), *ΔhsdS_*β (RM3) and *ΔΔhsdS_*α+β (RM2+3) to investigate the functional differences between methylation facilitated by either core HsdS-HsdM complex, and the effect of full deletion of 6mA methylation within the ST45 strain background.

5.4.2.2.1 DE Analysis - ST45 – CD140392 vs RM2, RM3 and RM2+3

As for the previous data set, the TMM normalised transcript data for the ST45 background show a slightly wider distribution of count levels for WT strain (CD140392 - aquamarine) than RM2 (red), RM3 (blue), RM2_3 (green) as seen in Figure 5.7 A. The samples included in this study cluster in two distinct groups along dimension 1 (logFC -1/3) of the MDS plot (Figure 5.7 B) with the WT biological (+technical) replicates being dispersed across dimension 2 (logFC -1.5-0.5). This indicates specific count differences introduced within each biological replicate of CD140392 previously discussed in the QC of Samples and Replicates in Chapter 4.
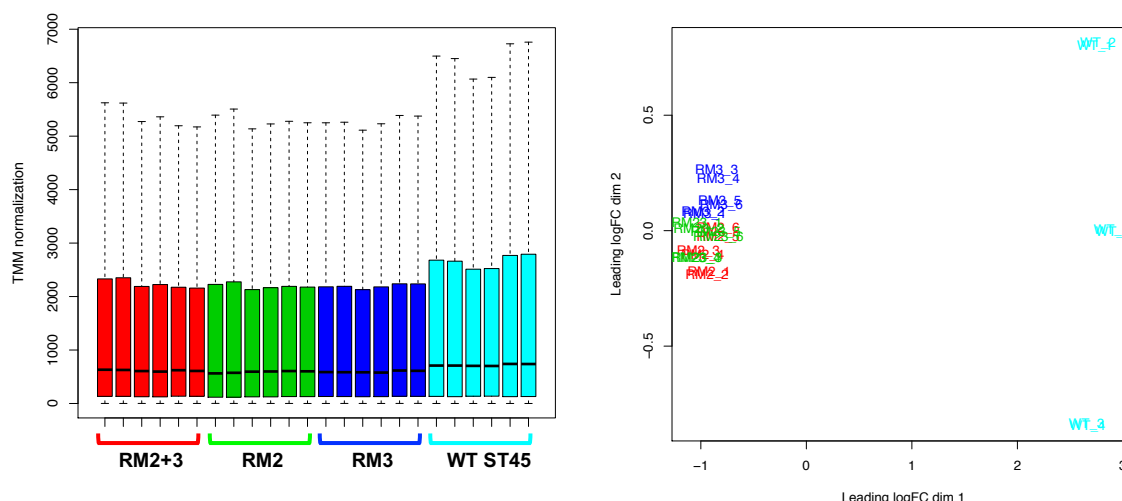
Figure 5.7 | Normalised Transcript Count Data ST45.
A. Transcript count distribution boxplot of TMM normalised count distribution per sample – normalised according to sequencing yield of each sample. Distribution of WT (aquamarine: CD140392) count numbers are wider compared to RM mutants (red: RM2+3, green: RM2, blue: RM3) seen in higher upper quartile and whisker values. **B**. Multidimensional Scaling of TMM normalised count data – WT and mutant strains cluster in two distinct groups along dimension 1 (logFC -1/3). WT replicates have with slightly variable count distribution along dimension 2 (logFC -0.5-0.5).

Differential expression analysis of pairwise comparisons of the RM mutants and the WT strain showed +625 (RM2 vs WT: 627, RM3 vs WT: 638, RM2+3 vs WT: 640) DE genes between each of the coupled strains as seen in **Error! Reference source not found.**. Pairwise DE analysis also revealed minimal expression changes (1-3 gene differences) between RM mutant strains.

As the number of differentially expressed genes between each mutant and WT strain was so similar and the number of DE gene differences between each mutant was so little, the DEG lists for each pairwise analysis was investigated in a global comparison of expression profiles for each gene, to explore the similarity of the repertoire of DE genes between each isolate. The represented DE genes formed two distinct gene clusters with similar expression profiles within the WT replicates and the mutant replicates as seen in Figure 5.8. The transcript levels, expression patterns and number of DE genes when compared to the WT strains for all ST45 RM mutants looked almost identical.

Table 5.1 | ST45 Differentially Expressed Gene Count Matrix

	RM2_3 ($\Delta hsdS\_\alpha+\beta$)	RM2 ($\Delta hsdS\_\alpha$)	RM3 ($\Delta hsdS\_\beta$)	WT (CD140392)
RM2_3 ($\Delta hsdS\_\alpha+\beta$)	0	1	2	640
RM2 ($\Delta hsdS\_\alpha$)	1	0	3	627
RM3 ($\Delta hsdS\_\beta$)	2	3	0	638
WT (CD140392)	640	627	638	0

Pairwise comparison of the number of differently expressed genes between mutant and WT strain. Green colouring indicates minimum gene changes (n=0) whilst red indicates maximum gene changes (n=640).

226
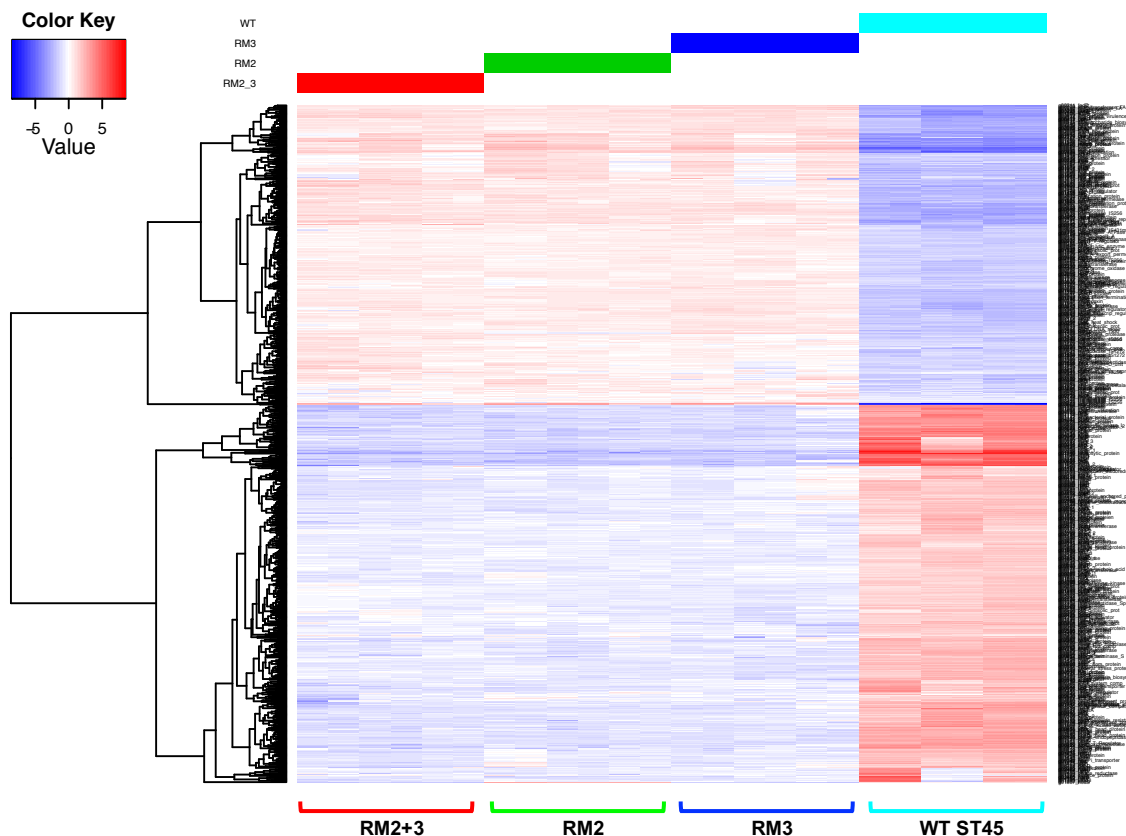
Figure 5.8 | Hierarchical Clustering of DE Genes vs ST45 Samples.
Global comparison of log2 transformed centre-clustered expression profile for each gene (right column) between ST45 WT sample CD140392 (aquamarine header) and mutant strain RM2+3 (red), RM2 (green), RM3 (blue Two distinct clusters of genes with similar expression profiles within the WT replicates and RM mutant replicates. Heat map values were calculated by subtracting each gene medium log2 (RPKM) value from the log (RPKM) value of each sample – unregulated genes are red whilst down regulated genes are blue (edgeR > 4, FDR < 0.001, P-value < 0.001).

499 DE genes (281 downregulated, 218 upregulated in RM mutants vs WT) were shared between each mutant vs WT pairwise comparison regardless which *hsdS* was deleted. RM2 had 42 (29 down – 13 upregulated genes), RM3 had 33 (13 down – 20 upregulated genes) and RM2+3 had 42 (17 down – 25 upregulated genes) uniquely differentially expressed against the WT strain (detailed in Supplementary Table 8.11). Similar to the ST622-2014 RM1 mutant vs CD141496 WT, ~22% of the total CDS (2837) were differentially regulated between the CC45 RM2, RM3 and RM2+3 mutants and WT strain CD140392. Functional analysis also revealed that the DE genes reported in the ST45 background were also DE in the ST622-2014 background reporting functional enrichment within mostly the same (with several additional) biological process (KEGG) as seen in Figure 5.9.
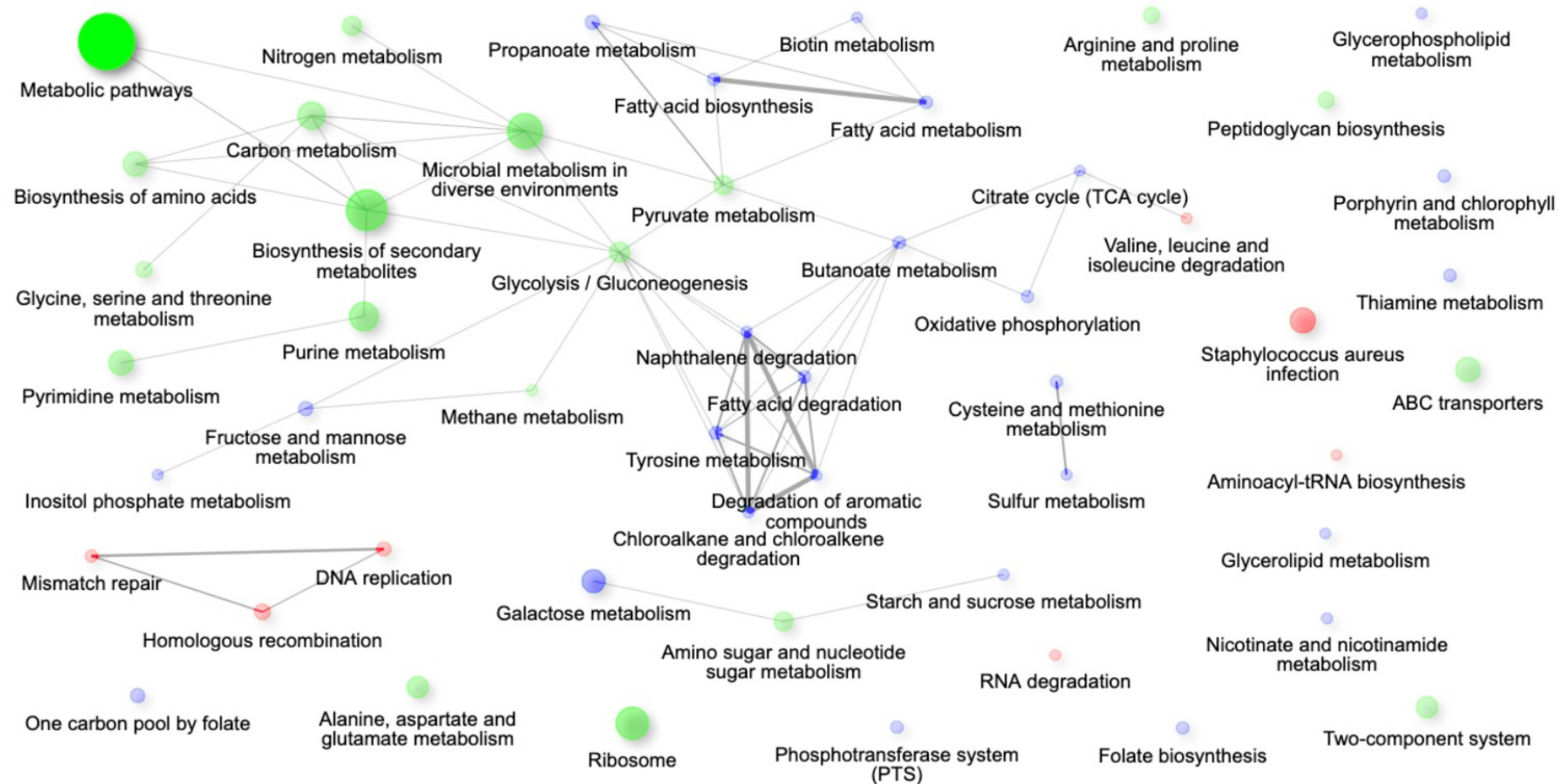
Figure 5.9 | KEGG Enrichment Plot Network Pathway of top 50 functional categories between ST45 RM mutants (RM2, RM3, RM2+3) and WT strain CD140392. The enrichment of ShinyGo genes at a significance level of p=0.05, using a gene list of differentially expressed genes. Larger and more vivid bubbles indicate increased number of genes and lower Enrichment False Discovery Rates (FDR) within the category represented. **Green** bubbles indicate pathway in which genes are both up and downregulated within CC45 RM strains compared to WT (metabolic pathways, biosynthesis of secondary metabolites, biosynthesis of amino acids, ABC transport, two component systems, glycolysis oxidative phosphorylation, and ribosomal function), whilst **red** bubbles indicate upregulation in CC45 RM mutants (*Staphylococcus aureus* infection, RNA degradation, and aminoacyl-tRNA biosynthesis, valine, leucine and isoleucine degradation, mismatch repair, DNA replication and homologous recombination), and **blue** bubbles indicate downregulation in RM compared to WT strain (amino acid metabolism/degradation (various), sugar metabolism, citrate cycle, amino acid biosynthesis (various), peptidoglycan biosynthesis and one carbon pool by folate).

When comparing the ST45 full methylation knockout (RM2+3) and the ST622-2014 full methylation knockout (RM1), the only major functional process differences were the upregulation of DNA replication associated functions (mismatch repair, homologous recombination) and both the up and down regulation of genes with ribosome associated functions within the RM2+3 vs WT (Figure 5.6, Figure 5.9). The overall similarity between enriched functions in both strain backgrounds strongly suggests that the full deletion of *sau1* specificity units generates a global regulatory response on general central metabolism and protein synthesis. This effect is seen not just in the full *sau1* mutants, but also in single knockouts RM2 and RM3, each with a functioning *hsdS*. Hence this suggests that the differential gene expression profiles between mutants and WT strains are not linked to differential methylation but are rather the result of a different pleotropic phenomena, potentially the accumulation of increased concentrations of methyl donor substrate SAM, or a global stress effect in response to perturbing the flux of essential metabolic functions.

5.4.2.2.2 DE Analysis - CC45 – Mutant vs Mutant - RM2, RM3 and RM2+3

Pairwise DE comparisons were also conducted between the RM mutants as previously seen in **Error! Reference source not found.** and Figure 5.8. Three genes were differentially expressed between RM2 (Δ*hsdS_α*) and RM3 (Δ*hsdS_ β*), namely *hsdS*_α (ID: 00396) was downregulated in RM2 vs RM3, *hsdS*_β (01807) was upregulated in RM2 vs RM3, and a lipoprotein coding CDS (g01188) was marginally upregulated (logFC: 2.07) in RM2 vs RM3. As expected, only *hsdS*_β (ID: 01807) was upregulated between RM2 and RM2+3, and *hsdS*_α (ID: 00396) was upregulated and an additional acetyltransferase coding CDS (ID: 02548) was marginally downregulated (logFC: -2.05) in RM3 vs RM2+3. Full DE gene lists detailed in Supplementary Table 8.12 in the Appendix.

The lack of differentially expressed genes between mutant strains indicates that deletion of *hsdS* within the mutant strain backgrounds did not cause major genotypic differences, and the loss of methylation due to loss of *hsdS* has no differential transcriptomic effect. To fully investigate this hypothesis, methylation analysis of the differentially expressed genes and the overall whole genome of mutant strains and the WT was conducted.

5.4.2.2.3 Methylation Analysis of DE Genes - ST45 – CD140392 vs RM2, RM3, RM2+3

To further investigate the role which loss of 6mA methylation plays in differential gene expression, the methylation events within each CC45 RM mutant and the WT parental strain genome were analysed in relation to the DE gene lists from the previous analysis. The methylation status of the three mutant strains were validated with PacBio SMRT sequencing and subsequent Methylation and Motif Analysis.

RM2 (Δ*hsdS_α*)	functional *hsdS_ β*	CR**A**A (N)$_7$ GG**A**
RM3 (Δ*hsdS_ β*)	functional *hsdS_ α*	GW**A**G (N)$_6$ TTT**A**
RM2+3 (ΔΔ*hsdS_α + β)*	no functional *hsdS*	none

The WT ST45 strain, CD140392, carrying two functional *hsdS* in each genomic island contains a total of 967 6mA methylation events – 582 recognised by HsdS_α and 385 recognised by HsdS_ β. The WT strain contains 2837 CDS, of which 755 (26.61%) overlap a methylation event, 443 (76.11%) and 312 (81.04%) methylation events overlapping CDS recognised by HsdS_α and HsdS_ β respectively. Only 56 of the reported 627 DE genes (8.93%) between RM2 vs WT, whilst 94/638 DE genes (14.73%) were methylated between RM3 and WT. None of the reported 649 DE genes were methylated between RM2+3 and the WT strain. The consistent number of DE genes seen in the pairwise RM and WT suggests that there is no correlation between the number of differentially expressed genes and the number of active *hsdS* present within the genome of an RM mutant in the ST45 background. There was also no evidence of significant gene expression for RM2 (Δ*hsdS_α*) or RM3 (Δ*hsdS_ β*) vs the WT strain, suggesting that there is no functional bias towards either of the two core *hsdS*. Therefore, it can be assumed that the loss of methylation by one or all Sau1 units due to deletion of variable *hsdS* in the ST45 background, has no effect on gene expression, and therefore Sau1 6mA methylation is not likely to play a role in epigenetic regulation.

Although the deletion of 'core' *hsdS* within ST45 showed little evidence of Sau1 6mA having epigenetic potential, the partial and full deletion of methylation did result in what seems to be a global DE response. To gain further insight into the role which Sau1 facilitated 6mA methylation plays within *S. aureus*, including the 'non-core' accessory genome associated *hsdS* units, further mutagenesis studies were conducted.

5.4.2.3 WT CD150713 vs RM5 (Δ*hsdS_S*), RM6 (Δ*hsdS_α*), RM4+5 (ΔΔ*hsdS_X+S*), RM4+6 (ΔΔ*hsdS_X+α*) and RM5+6 (ΔΔ*hsdS_S+α*)

The parental ST622-2015 WT strain CD150713, contained three functional *hsdS*, one inserted downstream the *orfX*, *hsdS_X*, one within the SCCmec, *hsdS_S*, and one in *ν*Saα, *hsdS_α*. Five differing knockout mutant strains were created from the above mentioned WT strain: RM5 (Δ*hsdS_S*), RM6 (Δ*hsdS_α*), RM4+5 (ΔΔ*hsdS_X+S*), RM4+6 (ΔΔ*hsdS_X+α*) and RM5+6 (ΔΔ*hsdS_S+α*) to investigate the functional differences between methylation facilitated by the non-core Sau1 systems.

5.4.2.3.1 DE Analysis – ST622-2015 - CD150713 vs RM5, RM6, RM4+5, RM4+6, RM5+6

As for the previous data sets, the TMM normalised transcript data for the ST622-2015 background show a slightly wider distribution of count levels for WT strain (CD150713) than RM5, RM6, RM4+5, RM4+6, and RM5+6 mutants as seen in Figure 5.10 A. The samples included in this study cluster in two distinct groups along dimension 1 (logFC -2/2) of the MDS plot (Figure 5.10 B) with mutants RM5, RM5+6, and RM6 with slightly variable count distribution along dimension 2 (logFC -1.5-0.5) compared to RM4+6 and RM4+5. This indicates specific count differences introduced by the RM4 mutation step, as only these two RM mutants have a deleted *hsdS_X;* RM4 stain used for subsequent creation of double knockout mutants.
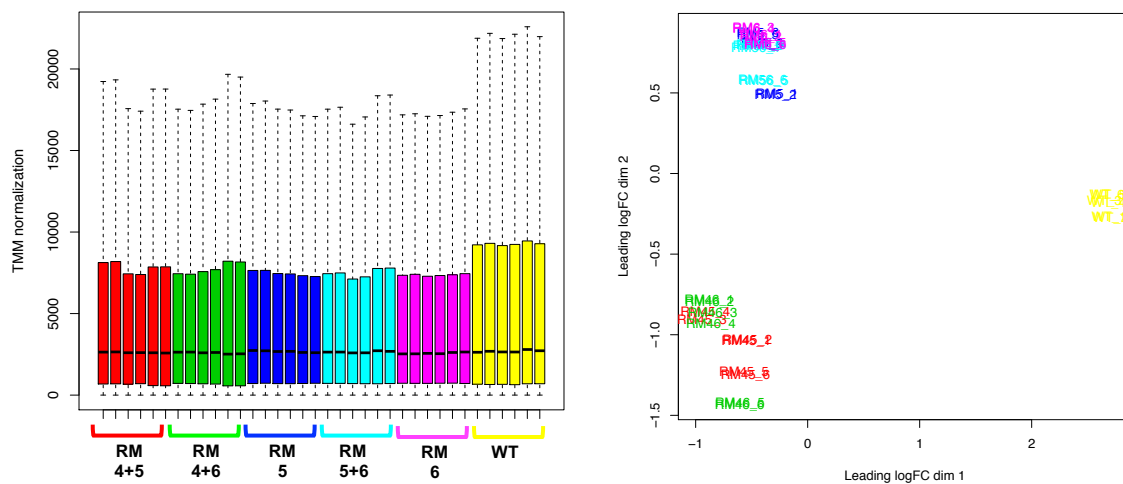
**Figure 5.10 | Normalised Transcript Count Data ST622-2015.**
A. Transcript count distribution boxplot of TMM normalised count distribution per sample – normalised according to sequencing yield of each sample. Distribution of WT (yellow: CD150713) count numbers are wider compared to RM mutants (red: RM4+5, green: RM4+6, blue: RM5, aquamarine: RM5+6, magenta: RM6) seen in higher upper quartile and whisker values. B. Multidimensional Scaling of TMM normalised count data – WT and mutant strains cluster in two distinct groups along dimension 1 (logFC -2/2) with mutants RM5, RM5+6, and RM6 with slightly variable count distribution along dimension 2 (logFC -1.5-0.5) compared to RM4+6 and RM4+5.

Differential expression analysis of pairwise comparisons of RM mutants and WT strain showed +450 DE genes (RM4+5: 456 genes, RM4+6: 478 genes, RM5+6: 502 genes, RM5: 489 genes, RM6: 528 genes) as seen in **Table 5.2**. On average about 42.5% of DE genes were upregulated in the RM mutants (RM 4+5: 187 genes, RM 4+6: 201 genes, RM 5+6: 215 genes, RM5: 211 genes, RM6: 230 genes) vs the WT strain. A global comparison of DE genes for all 5 mutants in this sequence background showed 5 distinct clusters of genes with similar expression profiles for the RM strains in comparison to the WT strain as seen in Figure 5.11. RM5 had 7 genes (3 down – 4 upregulated genes), RM6 had 30 genes (17 down – 13 upregulated genes) RM4+5 had 9 genes (2 down – 7 upregulated genes), RM4+6 had 6 genes

(6 upregulated genes), and RM5+6 had 5 genes (2 down – 3 upregulated genes) uniquely differentially expressed against the WT strain (detailed in Supplementary Table 8.14).

Table 5.2 | ST622-2015 Differentially Expressed Gene Count Matrix.

	RM4+5 (hsdS_X+S)	RM4+6 (hsdS_X+A)	RM5+6 (hsdS_S+A)	RM5 (hsdS_S)	RM6 (hsdS_A)	WT (CD150713)
RM4+5 (hsdS_X+S)	0	2	21	21	29	456
RM4+6 (hsdS_X+A)	2	0	21	21	21	478
RM5+6 (hsdS_S+A)	21	21	0	1	1	502
RM5 (hsdS_S)	21	21	1	0	2	489
RM6 (hsdS_A)	29	21	1	2	0	528
WT (CD150713)	456	478	502	489	528	0

Pairwise comparison of the number of differentially expressed genes between mutant and WT strain. Green colouring indicates minimum gene changes (n=0) whilst red indicates maximum gene changes (n=528).
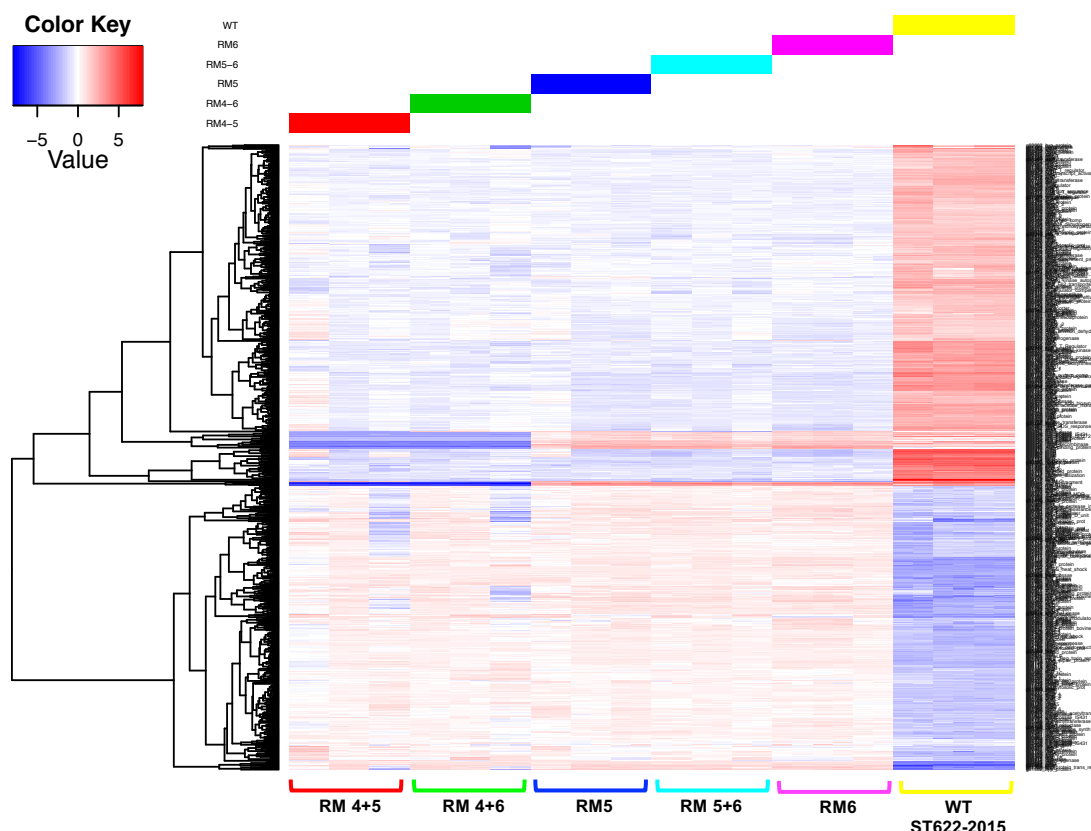
Figure 5.11 | Hierarchical Clustering of DE Genes vs ST622-2014 Samples.
Global comparison of log2 transformed centre-clustered expression profile for each gene (right column) between ST622-2015 WT sample CD150713 (yellow header) and mutant strain RM4+5 (red), RM4+6 (green), RM5 (blue), RM5+6 (aquamarine), RM6 (magenta). Five distinct clusters of genes with similar expression profiles within the WT replicates and RM mutant replicates. Heat map values were calculated by subtracting each gene medium log2 (RPKM) value from the log (RPKM) value of each sample – unregulated genes are red whilst down regulated genes are blue (edgeR > 4, FDR < 0.001, P-value < 0.001).

Over 400+ of the DE genes between the WT ST622-2015 isolate and the *ΔhsdS* daughter strains were shared among the knockout mutants. Thus, as previously seen in the ST45 RM (RM2, RM3 and RM2+3) vs WT comparison, it can be hypothesized that most of the genes differentially expressed between any of the ST622-2015 mutants and WT CD150713 were the same, having very similar transcriptomic profiles resulting in plausibly the same genotypic and phenotypic effects. The DE gene lists were therefore further investigated to confirm similarity with functional enrichment analysis.

As for the ST622-2014 and ST45 mutants vs WT comparisons, the functional categories associated with the differentially expressed genes between the ST622-2015 WT strain and the mutant *ΔhsdS* daughter isolates were namely biosynthesis and metabolism pathways. Figure 5.12 shows upregulation of genes with *Staphylococcus aureus* infection, DNA replication (mismatch repair) and downstream RNA transcription and translation (ribosome related functions, t-RNA biosynthesis, RNA degradation) as well as certain amino acid and sugar metabolism pathways with the mutants when compared to the WT strain. As seen in the previous sections, genes involved in the biosynthesis and metabolism of various amino acids, sugars, the citrate cycle (TCA cycle), peptidoglycan biosynthesis, fatty acid metabolism, and glycerophospholipid metabolism were downregulated in the ST622-2015 RM mutants vs parent isolate CD150713. Identical gene regulatory effects can be seen in in the pairwise comparisons of the WT strain and both the single (RM5, RM6) and double knockouts (RM4+5, RM4+6, RM5+6) with only a handful of unique genes DE between WT and RM (Supplementary Table 8.14 in the Appendix). Hence, the global response on general metabolism and protein synthesis is likely not due to differential methylation but are rather the result of a different pleotropic phenomena, as previously discussed for the ST45 mutants.
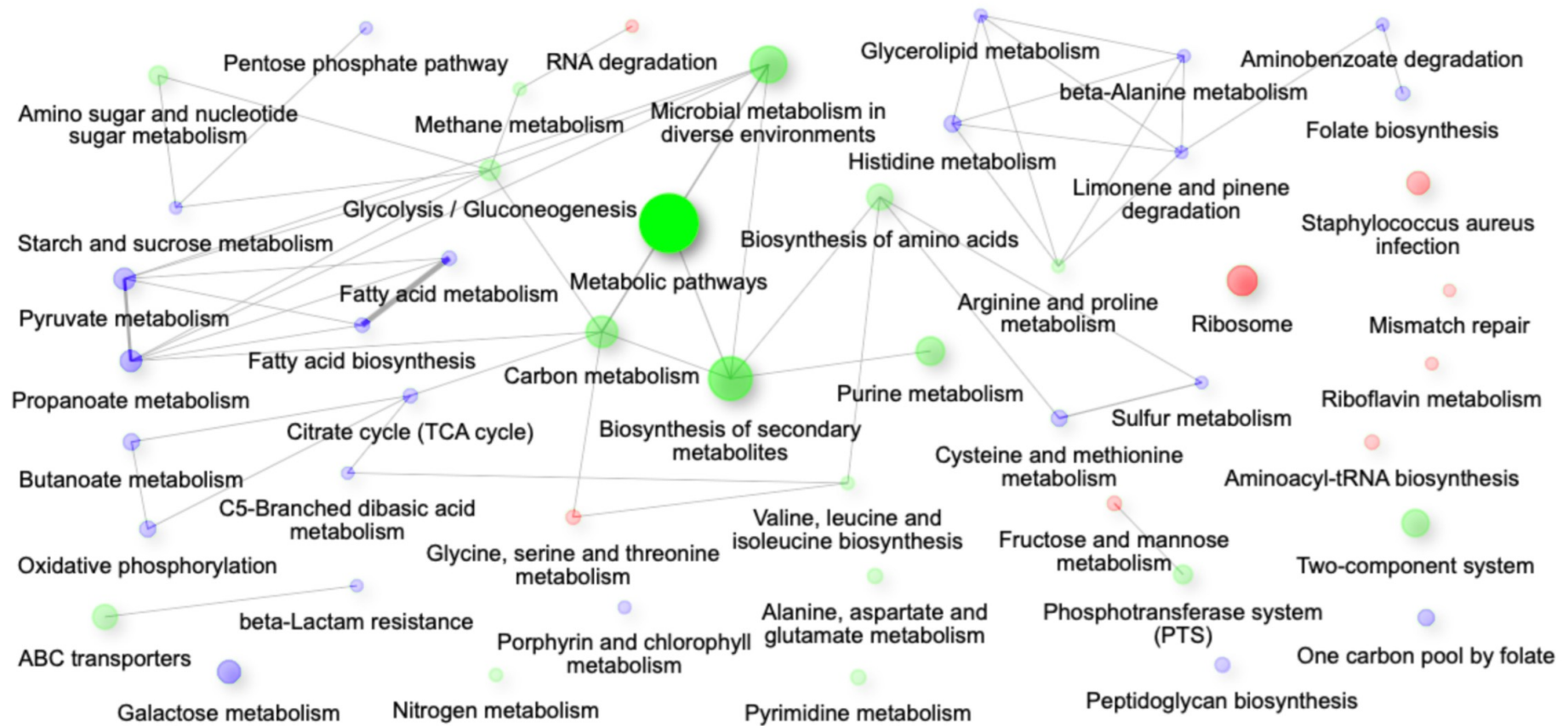
Figure 5.12 | KEGG Enrichment Plot Network Pathway of top 50 functional categories between ST622-2015 RM mutants (RM5, RM6, RM4+5, RM4+6, RM5+6) and WT strain CD150713. The enrichment of ShinyGo genes at a significance level of p=0.05, using a gene list of differentially expressed genes. Larger and more vivid bubbles indicate increased number of genes and lower Enrichment False Discovery Rates (FDR) within the category represented. **Green** bubbles indicate pathway in which genes are both up and downregulated within RM strains compared to WT (metabolic pathways, biosynthesis of secondary metabolites, biosynthesis of amino acids, ABC transport, two component systems, glycolysis phosphotransferase system, amino sugar and nucleotide sugar metabolism), whilst **red** bubbles indicate upregulation in ST622-2015 RM mutants (*Staphylococcus aureus* infection, ribosomal function, and aminoacyl-tRNA biosynthesis, glycine, serine, and threonine metabolism, mismatch repair, fructose and mannose metabolism, and riboflavin metabolism), and **blue** bubbles indicate downregulation in RM compared to WT strain (amino acid metabolism/degradation (various), sugar metabolism, citrate cycle, amino acid biosynthesis (various), peptidoglycan biosynthesis and one carbon pool by folate, fatty acid metabolism and synthesis, porphyrin and chlorophyll metabolism, beta-lactam resistance).

5.4.2.3.2 DE Analysis - ST622-2015 - Mutant vs Mutant - RM5, RM6, RM4+5, RM4+6, RM5+6

The pairwise DE comparisons between mutants resulted in minimal gene differences as previously seen in Table 5.2. The only gene differences between RM4+5 vs RM4+6 were two *hsdS* genes: *hsdS*_S (ID: 00087 – RM5 construct) upregulated in RM4+6 and *hsdS*_α (ID: 00406 – RM6 construct) upregulated in RM4+5. This was expected as the only functional *hsdS* within RM4+6 is *hsdS*_S, and the only functional *hsdS* within RM4+5 is *hsdS*_α. The only differentially expressed gene between RM5 and RM5+6 was *hsdS*_α (ID: 00406 – RM6 construct), upregulated in RM5 in comparison to RM5+6, as *hsdS*_α is deleted in the double knockout mutant. Similarly, the only DE gene difference between RM6 vs RM5+6 was *hsdS*_S (ID: 00087 – RM5 construct), upregulated in RM6, as this *hsdS*_S was deleted in the double knockout RM5+6. Between RM5 and RM6, *hsdS*_α (ID: 00406 – RM6 construct) was upregulated in RM5 and *hsdS*_S (ID: 00087 – RM5 construct) was upregulated in RM6 when pairwise compared.

A slightly larger cluster of genes (<30 genes) were flagged as differentially expressed within the double knockout mutants RM 4+5 and RM 4+6 when compared to the rest of the mutant isolates. 21 DE genes were recorded between RM4+5 compared to RM5/RM6/RM5+6 as well as RM4+6 vs RM5/RM5+6. Along with *hsdS*_X (ID: 00030 – RM4 construct) a cluster of 20 genes (Figure 5.13) within the SCC*mec* (ID: 00039-00059) were not expressed or expressed at a minimal level within RM4+5 and RM4+6 samples. The lack of expression of these genes could potentially be loss of function due to transformation of *hsdS*_X creating RM4, which was used as the parent strain to construct RM4+5 and RM4+6 double knockout mutants. 29 genes were differentially expressed between RM6 and RM4+5, including the deleted *hsdS* genes (*hsdS*_X (ID: 00030) and *hsdS*_α (ID: 00406)), the SCC*mec* gene cluster (n= 20) with the addition of and 5 genes from the purine cluster (*purFMNHD* – ID: 00890-894) which were upregulated marginally within RM4+6 (logFC: ~2.24 (2.12-2.30) – Supplementary Figure 8.4). The difference in expression profile of these genes can be clearly seen in the middle clustering of Figure 5.11, which account for the variable dispersion of counts seen in the MDS (Figure 5.10) above. Full DE gene lists detailed in Supplementary Table 8.13 in the Appendix.

The lack of differentially expressed genes between mutant strains indicates that differential methylation as a result of *hsdS* deletion has no significant effect on transcription. To fully investigate this hypothesis, methylation analysis of the differentially expressed genes and the overall whole genome of mutant strains and the WT was conducted.
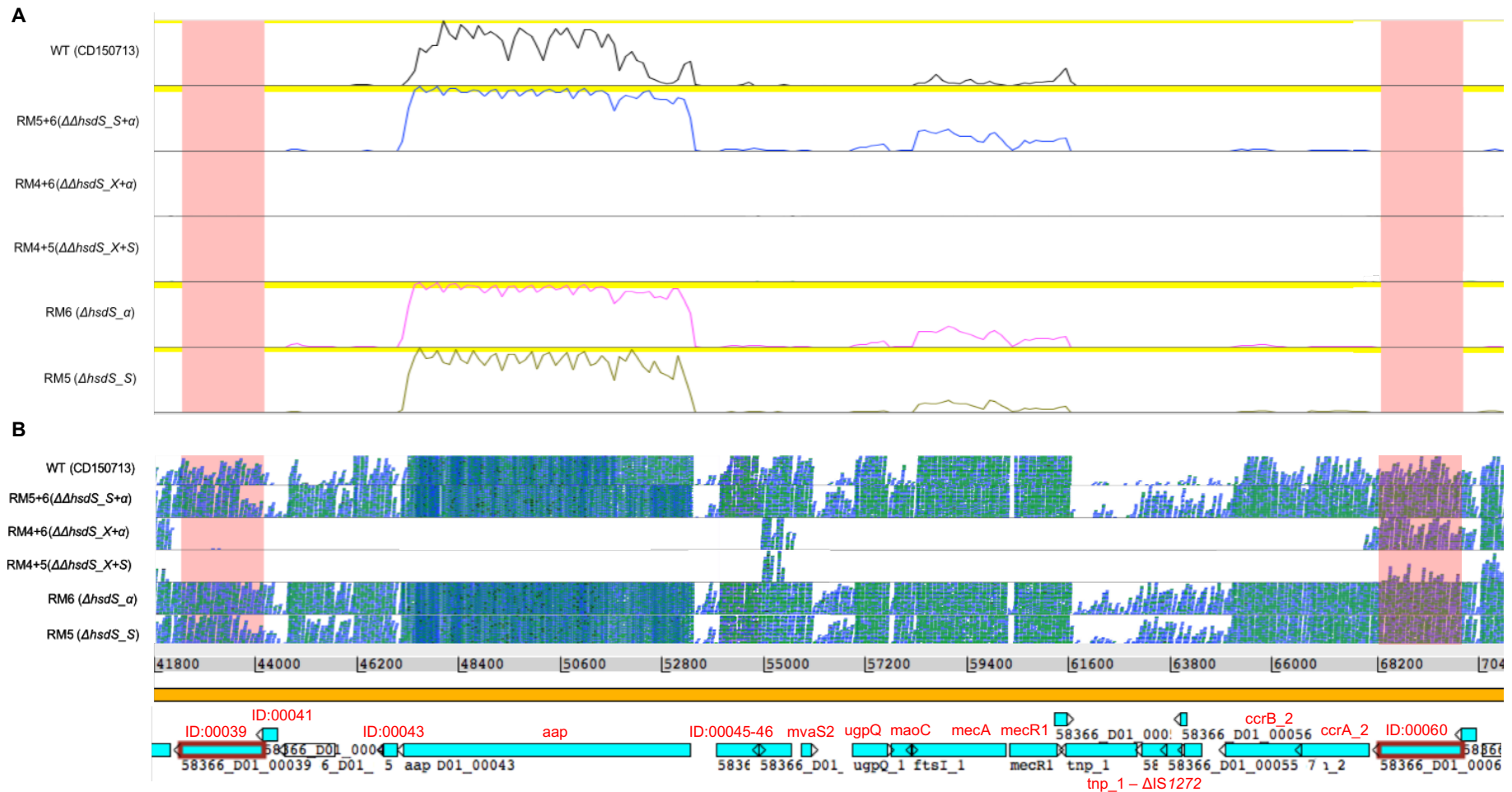
Figure 5.13 | RNA transcript levels for gene cluster 00039-00059 – ST622 SCC*mec*.
Artemis visualisation of RNASeq generated transcript reads aligned to WT reference CD150713 genome for each of the 5 generated mutants within this lineage (WT, RM5+6, RM4+6, RM4+5, RM6, RM5). The transcript levels of each RM mutant are represented in stacked across the SCC*mec* containing 20 gene cluster (CDS in aquamarine) which were not expressed in mutant isolates transformed with RM4 KO construct deleting *hsdS_X*. **A.** Scaled transcript coverage view with separate plots for each isolate (top down: WT CD150713, RM5+6, RM4+6, RM4+5, RM6, RM5) at windows set at 10,000 RPKM to see relative transcript counts throughout the region of interest. There is no coverage for this sequence reigon for RM4+6 and RM4+5. **B.** Stacked read alignment view (paired reads in blue, multiple reads in green and single reads in black) to show lack of transcripts for the given genes for RM4+6 and RM4+5; stacked transcripts are visible for CDS prior to ID:00039 (left) as well as ID:00060, suggesting that the whole SCCmec element is either not expressed or may have been lost – this could be validated with genome sequencing.

5.4.2.3.3 Methylation Analysis of DE Genes – ST622-2015 – CD141496 vs RM5, RM6, RM4+5, RM4+6 and RM5+6

To further investigate the role which loss of 6mA methylation plays in differential gene expression, the methylation events within each ST622-2015 RM mutant and the WT parental strain genome (CD150713) were analysed in relation to the DE gene lists from the previous analysis. From the PCR results and lack of expression data (RNASeq) for the mutagenized genes it was concluded that the Δ *hsdS* genes were truly deleted. Hence it was inferred that with the deletion of the gene we also remove the methylation activity pertaining to the distinct HsdS-HsdM complexes. Due to some technical issues regarding PacBio sequencing at the Wellcome Sanger Institute, the methylation status of the full deletion of *hsdS* genes through whole genome sequencing could not be confirmed.

RM5 (Δ*hsdS_S*)	functional *hsdS_α* + *hsdS_X*	**A**GG (N$_6$) YTC**A**	TA**A**G (N$_6$) GA**A**
RM6 (Δ*hsdS_α*)	functional *hsdS_S* + *hsdS_X*	GA**A**G (N$_5$) GT**A**	TA**A**G (N$_6$) GA**A**
RM4+5 (ΔΔ*hsdS_X*+*S*)	functional *hsdS_α*	**A**GG (N$_6$) YTC**A**	
RM4+6 (ΔΔ*hsdS_X*+ *α*)	functional *hsdS_S*	GA**A**G (N$_5$) GT**A**	
RM5+6 (ΔΔ*hsdS_S*+ *α*)	functional *hsdS_X*	TA**A**G (N$_6$) GA**A**	

The WT ST622-2015 strain, CD150713, carries three functional *hsdS* contained a total of 1368 6mA methylation events – 425 TRS (31.07%) recognised by *hsdS_X*, 258 TRS (18.86%) recognised by *hsdS_S* and 685 TRS (50.07%) recognised by *hsdS_α*. The WT strain contained 2543 CDS, of which 982 (38.62% or all CDS) overlap a methylation event. Overall, around 71.78% of all *sau1* 6mA TRS (68.24% - 290/425 TRS recognised by *hsdS_X*, 81.78% - 211/259 TRS recognised by *hsdS_S*, and 70.22% - 481/685 TRS recognised by *hsdS_α*) were located within a CDS (+ very few TRS within 200bp of the 5' and 3' UTR). The location of each TRS recognised by each *hsdS* within the mutant strains and the WT remain the same.

Within the *ΔhsdS* mutant strains, the number of methylated CDS was conditional on which *hsdS* were still functional within the strain. Irrespective of this, the number of DE genes between each RM mutant and the ST622-2015 WT averaged around 19.36% (17.93-20.76%; 456-528 genes) of the total CDS as seen in Table 5.3.

Table 5.3 | Number of DE genes (DE-CDS) containing 6mA motif (ST622-2015)

Number of differentially expressed genes (DE-CDS) between wild-type and mutant isolates

	RM5	RM6	RM4+5	RM4+6	RM5+6
TOTAL	498	528	456	478	502
% of CDS*	19.58	20.76	17.93	18.81	19.74

Number of differentially expressed genes (DE-CDS) in MUTANT isolates with 6mA motifs

	RM5	RM6	RM4+5	RM4+6	RM5+6
hsdS_X	45	48			40
hsdS_S		92		39	
hsdS_A	126		83		
hsdS_B	–	–	–	–	–
TOTAL	171	140	83	39	40
% of DE-CDS*	34.34	26.52	18.2	8.16	7.97

*calculated using the number of CDS within the WT strain CD150713

Moreover, the number of DE genes (and respective 5' and 3' UTR regions) flagged between the WT and mutant strain which contained a 6mA TRS varied between 7.97% in RM5+6 to 34.34% in RM5, due to the manipulation to which and how many functional *hsdS* each RM strain contained (Table 5.3). The variability and low proportion of methylated DE genes within the RM mutants paired with the consistent similarity in identity and number of DE genes between the mutants and the WT strain, suggest that there is no link between *sau1* methylation and differential gene regulation in *S. aureus*. There was also no evidence of significant gene expression differences linked any of the ST622-2015 RM strains, suggesting that there is no functional bias between any of the three *hsdS*, including no difference between the 'core' *hsdS_α* and the 'accessory' genome related *hsdS_S* and *hsdS_X* function. Therefore, it can be assumed that the loss of methylation by one or two Sau1 units due to deletion of variable *hsdS* in the ST622-2015 background, has no effect on gene expression, and therefore Sau1 6mA methylation is not likely to play a role in epigenetic regulation.

5.4.2.4 Differential Expression of Global Regulators and Transcription Factors Between WT and RM Mutant Strains

As previously determined in the functional enrichment analyses of the differentially expressed genes between WT and RM mutants, most of the DE genes were shared among the three sequence backgrounds, having the same expression patterns, belonging to the same molecular function categories (Figure 5.6, Figure 5.9 and Figure 5.12). This suggests that the deletion of any, some or all *hsdS*, induces a pleotropic regulatory effect

of core metabolic functions within the cell. To understand the mechanism by which this may occur, the gene expression levels of global regulators and transcription factors between WT and RM mutants were investigated. Out of ~135 TFs, sigma factors and global regulators described by Ibarra et al, (2013), 28 regulators were consistently up/down regulated within the RM vs WT strains highlighted in Table 5.4.

Out of the 28 regulators, 14 were uniformly up/down regulated by all RM mutants throughout the three STs. These included a cluster of stress and heat shock response regulators: DnaK, CtsR, HrcA, GroEL, GroES all upregulated in the RM mutants vs the WT strains, clearly indicating a complex regulatory response to stress potentially induced through the perturbation/deletion of Sau1 activity within the cell. Global virulence regulator Rot and zinc homeostasis regulator Zur were also upregulated in the mutants vs WT whilst. various metabolic regulators including PhoP (inorganic phosphate import), ArsR (arsenical resistance) and CcpA (carbon catabolism) were downregulated in RM vs WT, as well as SigB housekeeping sigma factor and competence regulating TF.

Nine regulators were uniformly up (n=6) / downregulated (=3) in only the ST45 mutants (RM2, RM3 and RM2+3), with most upregulated genes in the RM vs WT strains being involved with oxidative and metal ion stress (Fur, PerR, SarZ, MntR) along with SigA the 'housekeeping' transcriptional initiator. Three regulators were uniquely regulated in ST622 backgrounds: FabR regulating lipid biosynthesis was upregulated in RM vs WT, whilst FeoA (iron homeostasis) and NreC (oxygen response) were downregulated. Two regulators, TreR (trehalose metabolism) and LexA (SOS response) were DE between WT and RM for all three sequence backgrounds, but produced a differing regulatory profile, upregulated in RM vs ST45 WT, and downregulated in RM compared with ST622 WT strains. This indicates that along with the set of 14 equably expressed global regulators, each group of mutants also has individual regulatory cascades which are ST specific. Further investigation is necessary to determine the complexity of regulatory network responsible for the pleotropic metabolic effect seen across RM vs WT strains.

Table 5.4 | Differentially Expressed Regulatory Genes between WT and RM Mutant Strains

Regulator	Function	ST45			ST622-2014	ST622-2015				
		WT vs RM2	WT vs RM3	WT vs RM2+3	WT vs RM1	WT vs RM5	WT vs RM6	WT vs RM4+5	WT vs RM4+6	WT vs RM5+6
DnaK	Stress and heat shock response	-2.692	-3.286	-2.958	-3.020	-3.636	-3.835	-3.346	-3.723	-3.661
CtsR	Stress and heat shock response	-2.563	-2.797	-2.294	-2.309	-3.017	-3.116	-2.704	-2.797	-2.981
HrcA	Negative regulator - heat shock response	-1.839	-2.244	-1.921	-2.870	-3.612	-4.095	-3.153	-3.363	-3.599
GroEL	Stress and heat shock response	-2.650	-2.829	-2.469	-2.603	-2.526	-2.606	-2.397	-2.624	-2.555
GroES	Stress and heat shock response	-1.571	-1.554	-1.590	-2.294	-3.362	-3.385	-3.090	-3.362	-3.350
Rot	Global regulator - virulence, quorum sensing	-3.635	-3.574	-3.962	-1.776	-2.178	-2.278	-2.268	-2.151	-2.166
Zur	Oxidative stress (zinc homeostasis)	-3.359	-3.134	-3.362	-2.621	-2.550	-2.931	-2.415	-2.516	-2.399
PhoP	2C* PhoPR – inorganic phosphate import	3.612	3.488	3.141	3.702	3.045	3.414	3.104	3.321	3.169
SigB	Housekeeping transcription	2.265	2.196	2.064	3.114	2.674	2.729	2.311	2.412	2.493
ArsR	Arsenical resistance (repressor)	4.526	4.921	4.706	2.483	2.428	2.341	2.677	2.466	2.521
ybaK/EbsC	TF – edit incorrectly charged AAs	3.993	4.287	3.853	2.794	3.477	3.572	3.597	3.769	3.442
SrrA	2C* SrrAB (toxic shock syndrome)	1.943	2.303	2.278	2.671	2.461	2.320	2.226	2.172	2.449
SAUSA300_0954**	TF – competence (not ComK)	3.476	3.185	3.792	2.017	2.274	1.865	2.559	1.839	2.445
CcpA	Carbon metabolism, virulence	2.771	3.108	2.583	2.563	2.100	2.087	1.854	1.889	2.025
GapR	Glycolysis regulator protein	2.150	2.895	3.606	0.150	0.646	0.297	0.714	0.816	0.783
HU	Virulence	2.215	1.817	2.113	-1.119	-0.471	-1.002	-1.356	-1.116	-1.245
CzrA	Negative regulator - oxidative stress	2.298	2.154	2.496	-1.161	-1.454	-1.294	-1.461	-1.218	-1.331
Fur	Oxidative stress (ferric ions)	-3.511	-3.066	-3.557	-1.210	-1.344	-0.924	-1.242	-0.742	-1.149
PerR	Oxidative stress (multi-gene repressor)	-2.781	-2.827	-2.471	-0.835	-0.108	-0.240	-0.539	-0.136	-0.371
SigA	Housekeeping transcription	-3.236	-3.317	-3.380	-0.326	-0.050	-0.239	-0.441	-0.261	-0.304
SarZ	Oxidative stress, metabolic switch, virulence	-2.962	-3.111	-3.037	0.351	0.455	0.241	-0.175	0.050	-0.049
MntR	Global regulator – oxidative stress (manganese)	-4.642	-4.869	-4.989	-0.224	-0.589	-0.408	-0.788	-0.642	-0.534
IcaR	Intracellular adhesin production / biofilm	-2.114	-1.843	-2.737	0.682	-0.311	-0.861	-0.315	-0.461	-0.526
TreR (*treR_1*)	**Trehalose utilisation (repressor)**	-2.414	-2.261	-2.577	-1.714	2.847	2.805	2.708	2.782	2.861
LexA	**Repressor - DNA damage (SOS response)**	-2.106	-2.290	-2.560	3.456	3.560	3.823	2.708	3.074	3.338
NreC	Oxygen response – nitrate/nitrite reduction	-0.434	-0.673	-0.492	2.077	2.714	2.773	2.459	2.556	2.772
FeoA	Iron homeostasis	0.628	0.473	0.787	2.881	2.933	2.549	2.482	2.938	2.785
FapR	Membrane lipid homeostasis – lipid biosynthesis	0.381	0.671	0.324	-2.011	-2.347	-2.329	-2.129	-1.917	-2.182

2C = two-component system; **gene ID of TF as identified by Ibarra et al., 2013*

5.5 DISCUSSION

In this study isogenic mutant strains were successfully created isogenic mutant strains in multiple *S. aureus* lineages by allelic replacement and mutagenesis of Sau1 *hsdS* to subsequently study the effects this deletion on whole genome methylation and cellular functions. Transcriptomic changes between the set of *ΔhsdS* mutant strains, as well as their WT parental strain has provided an opportunity to investigate the regulatory effect of the sequential loss of 6mA methylation.

5.5.1 Effect of loss of 6mA Methylation on *S. aureus* gene expression

RNA sequencing was used to exploration the differential gene expression effects of loss of 6mA DNA methylation in a subset of 6mA deficient strains and the WT strains. It is important to acknowledge that the RNA-Seq data compared for the WT and the mutant strains were conducted under the same experimental design conditions but were from two different RNA sequencing batches.

The transcriptomic analysis evaluating the gene expression effect of loss of Sau1 facilitated 6mA methylation yielded contrasting results. The comparison of WT *S. aureus* strains from 3 different lineages with mutagenized *ΔhsdS* daughter stains revealed a large cohort of differential expressed genes (**Error! Reference source not found.**, **Table 5.2**, Figure 5.5;). Functional analysis of the DE genes revealed that most were involved in core metabolic and biosynthetic pathways (Figure 5.6, Figure 5.9, Figure 5.12), with many shared between all samples regardless of ST type including: glycolysis, biosynthesis of amino acids and secondary metabolites with distinct downregulation of galactose metabolism, citrate cycle (TCA), glycolipid and butanoate metabolism and upregulation of ribosomal functions, genes involved in *Staphylococcus aureus* infection, RNA degradation, mismatch repair and amino-acyl tRNA synthesis in RM mutants vs WT. However, there were differences between up/down regulated pathways within each lineage pertaining to specific amino acid and fatty acid biosynthesis, metabolism and degradation as well as sugar metabolism among others. It should be noted that each *ΔhsdS* sample within a given lineage shared the almost identical expression profile, sharing the great majority of differentially expressed genes when pairwise compared to the WT data (ST45 – Figure 5.8; ST622-2015 – Figure 5.11; ST622-2014 – Figure 5.5). Hence, the deletion and loss

of 6mA modification from one, some or all *hsdS* alleles, may have a global pleiotropic regulatory effect.

In some cases, hundreds or even thousands of gene expression differences have been attributed to the hinderance of a single 6mA DNA methyltransferase unit (mostly phase-varions) in *E. coli* (Fang et al., 2012) *H. pylori* (Beaulaurier et al., 2014*), M. genitalium and M. pneumoniae* (Lluch-Senar et al., 2013) *S. pneumoniae* (Manso et al., 2014), *C. crescentus* (Bo Zhou et al., 2015). Loss of 6mA methylation by TIII EcoGIII (*E. coli*) was also shown to cause a largescale transcriptomic effect, with 38% of the 5131 annotated genes within *E. coli* strain C227-11 being differentially expressed in comparison to *ΔhsdRM* strains (Fang et al., 2012). Similarly, deletion of 6mA modifying TIIG RM systems in *B. burgdorferi* resulted in global changes in gene expression, with 25-39% of annotated genes being differentially expressed between WT and ΔRM strains (Caselli et al., 2018). Generally, these studies have found that only a small fragment (<10%) of differentially expressed genes were found to have methylated adenine bases within their promoter regions. This study found that ~4% of DE genes between WT and *ΔhsdS* strains with active methylation systems (strains with multiple *hsdS)* contained a modified adenine within 200 bp of the CDS; DE genes between WT and full *ΔhsdS* knockouts had no methylation throughout the whole genome. This suggests that the proposed 'local competition' (Fang et al., 2012) model of MTase/Modification complexes (M_2S) competing with other DNA binding proteins at the promoter of a CDS, does not apply to most DE genes, and may not apply to Sau1 entirely. Studies have also found that not all DE genes were methylated within the coding regions, suggesting DNA methylation has a selective regulatory effect in which modification of a select cohort of promoters and CDS results in a cascade of gene expression regulation (Nye et al., 2019; Fang et al., 2012; Kahramanoglau et al., 2012). This correlates with the WT vs mutant DE results of this study, where modification of DE genes ranged from 50% (WT ST622 – 3 active HsdS) to 0% (no functional HsdS - ST622-2015 RM1 (*ΔhsdS_α)*, ST45 RM2+3 (*ΔΔhsdS_α+ β*). However, it is unlikely that Sau1 differential methylation or methylation within regulatory regions has any effect on gene expression as strains fully devoid of Sau1 6mA, RM2+3 in ST45 background, have the same expression profile against the WT as single *ΔhsdS* allele knockouts RM2 (*ΔhsdS_α)* and RM3 (*ΔhsdS_β)* (Figure 5.8). Additionally, mutant strains within a lineage (both in ST45 and ST622-2015) have no DE genes between them, other than the delete *hsdS* alleles (with exception of non-expressed gene cluster (20 CDS) introduced in RM4 during transformation) (**Error! Reference source not found.**, **Table 5.2**, Supplementary Table 8.13).

The high number of differentially expressed genes between the WT and mutant strains are potentially the result of a general stress response as the natural metabolic homeostasis and flux of substrates may have been perturbed through the deletion of the DNA binding specificity units. This can be clearly seen in the uniform upregulation of ubiquitous stress and heat shock response regulators DnaK, CtsR, HrcA and GroES, GroES (Figure 5.4) in the RM mutant vs the WT strains in all three sequence backgrounds (Singh et al., 2007; Fleury et al., 2009; Roncarati & Scarlato, 2017; Anderson et al., 2006; Chastanet et al., 2003). It is uncertain why or by what mechanism this stress response is induced, but perhaps the lack of Sau1 6mA methylation is recognised by global sensor (unknown feedback mechanism) which stimulates this complex regulatory reaction. Alternatively, the lack of Sau1 activity – both host DNA methylation and exogenous DNA restriction – leaves the host DNA vulnerable inducing this stress response cascade to protect itself from self/exogenous DNA damage (Kobayashi, 2004). Although growth rates and phases were preliminarily determined using growth experiments prior to experimentation (Appendix Figure 8.4), potential errors in sampling and biological variability between experiments may have captured RM mutants at early stationary phase. Hence, HsdS and/or DNA methylation may also play a part in later metabolic / global regulatory response or involved in stringent response (entering stationary phase) indicated by upregulation of stress response proteins in RM mutants. The large-scale gene expression shift of many general cellular processes between the RM mutants and WT strains will have also been influenced by the decreased expression of sigma factor B and the downstream transcriptional effects. SigB has been shown to be part of a complex regulatory network in *S. aureus* influencing the transcription of around 200 genes involved in various cellular function, including intermediate metabolism, membrane transport and cell composition (Bischoff et al., 2004; Pané-Farré et al., 2006; Fleury et al., 2009).

As 6mA methylation by Sau1 is S-adenosylmethionine (SAM) substrate dependent, it was postulated that the potential accumulation of the methyl donor substrate, may have downstream metabolic effects as SAM is used not just in nucleotide (DNA and RNA) methylation but a range of metabolic and biosynthetic processes within a cell (Schoenfelder et al., 2013; Paveen & Cornell, 2011; Martin & McMillan 2002; Lu, 2000). SAM is synthesized from ATP and methionine by SAM synthetase, encoded by *metK* (Saint-Girons et al., 1988). SAM is known to act as a strong feedback inhibitor of MetK in various bacterial species, and also acts as a transcriptional corepressor (SAM-binding S-box riboswitch) for other genes within the active methyl cycle (Usuda and Kurahashi, 2005; Eustáwuio et al., 2008; Fuchs et al., 2006; Epshtein et al., 2003; Alvarez et al., 1994;

Posnick & Samson, 1999). In staphylococci the methionine biosynthesis operon (*metICFE-mdh* operon) is controlled by a hierarchical network involving an initiator (Cod-Y) tRNA-specific T-box riboswitch, as unlike other bacilli, they lack a methionine salvage and polyamine synthesis pathway (Rodinov et al., 2004). This suggests that the synthesis of not just methionine, but ubiquitous methyl donor SAM, is more stringent in *Staphylococcus* species and heavily relies on recycling of SAM for these processes (Schoenfelder et al., 2013; Rodinov et al., 2004; Wencker et al., 2021). It is therefore unlikely that SAM is overproduced or over accumulated in the cell due to decrease/loss of DNA methylation and no study to date have characterised either of these events in *S. aureus*.

The lack of DE variation between strains with multiple (RM5, RM6), single (RM2, RM3 RM4+5, RM4+6, RM5+6) or no functioning *hsdS* (RM2+3, RM1), suggests that TI RM Sau1 is non-essential, and furthermore demonstrate no differences in functionality or necessity between either of the core *hsdS* (*hsdS*_α and *hsdS*_β) nor the accessory *hsdS* (*hsdS*_X and *hsdS*_S), other than DNA binding for RM activity. This indicates that TI RM Sau1 may not likely have a distinct secondary regulatory function in *S. aureus*, and is exclusively involved in host defence, which is well documented (Waldron & Lindsay, 2006, McCarthy & Lindsay, 2013; McCarthy et al., 2012; Jones et al., 2015; Monk & Foster, 2012; Chen et al., 2016; Roberts et al., 2013; Cooper et al., 2017, Sullivan et al., 2019). This is in line with the results recently published by Meherashahi and Chen (2021) evaluating the potential regulatory role of DNA methylation mediated by archetypal TI RM systems in *E. coli* in a very extensive study, finding zero evidence of any epigenetic regulation of TI 6mA DNA methylation in three unrelated strains. Their broad-scale study concluded deletion or switching of TI methylation systems *ΔhsdSMR* – EcoUTI, EcoKI, EcoCFTI, as well as single *hsdS* allelic replacements in 3 different uropathogenic strains UTI189, MG1655 and CFT073, had no effect on gene expression of any growth phenotypes in a screen of 1190 conditions (Meharashahi & Chen, 2021).

5.5.2 Study Limitations

One evident limitation to this study was the lack of genomic validation of all mutant strains due to delays imposed by COVID-19. Sequencing of ST622-2015 RM mutants with PacBio SMRT sequencing would be necessary to validate the success of *ΔhsdS* allelic replacements, as well as loss of methylation with SMRT Motif and Modification analysis. Validation of the differentially expression results using microarray or RT-qPCR of specific

DE genes between WT and mutant strains would greatly benefit this study to better understand the role of Sau1 6mA, and in particular *hsdS*.

5.5.2.1 Experimental Design

A definite point of uncertainty of this study was the differential gene expression analysis between RNA-Seq data between the WT and *ΔhsdS* isolates. Although the pairwise DE analysis between WT and mutant strains suggest the deletion of any, some or all *hsdS* alleles and consequential loss of 6mA DNA methylation, has a general effect on central metabolism and protein synthesis, the pairwise comparisons of sequential mutant strains within a lineage show no differential expression changes. The lack of transcriptomic changes between mutant strains, descending from the same parent strain, suggests loss of Sau1 6mA within various lineage did not result in genotypic differences. Hence it is tentative to attribute Sau1 methylation to having a regulatory role in *S. aureus*. However, validation of the presented gene expression data through additional resequencing (with WT strains included in sequencing round) or RT-qPCR and additional growth studies to evaluate physiological impact of loss of methylation would add confidence to this study.

DNA methylation is a dynamic process and may vary within different growth environments and stress conditions (eg: hypoxia, high/low sodium, variable carbon source, temperature, nitrogen source, pH, amino acids) as well as differing growth phases (lag phase, log phase (early vs late), stationary, decline phase). It is important to note that all bacterial cultures within this study were grown under standardised optimal laboratory conditions (for comparable RNA-Seq results), and therefore the potential dynamic nature of 6mA DNA methylation in response to variable stress and growth conditions (persisting in non-laboratory environments) and the impact thereof could not be captured in this study. Further investigating DNA methylation under different phenotypic conditions would greatly benefit our knowledge of the potential variability in methylation by Sau1 due to external stress and stimuli (as mentioned below in Future Work).

5.5.2.1.1 Bacterial Growth Phase - Culturing and Sampling

Each sample was cultured in 25mL of BHI broth in baffled 250mL culture flasks with filter caps, incubated overnight at 37ºC, shaken at 160 RPM until they reached end of log-phase growth. Growth curves were generated for both the WT and RM mutants prior to experimentation to determine the appropriate sampling point at the optimal growth phase (Figure 8.4 of the Appendix) and to also investigate any significant differences in growth rate or other growth characteristics which could be highlighted with optical density

technique. There were no significant changes in growth in rich TSB media for any of the RM mutants in comparison to the WT parental strain in the 16-hour sampling period until stationary phase. Within early stationary phase the RM mutants seem to plateau and OD measurements start to decline slightly faster than for the WT strains, particularly for ST622-2015 (CD150713) variant. During stationary phase bacterial cells remain metabolically active, but cease to grow, and essentially the cells induce a stress response as their growth environment starts to become nutrient deficient. Various physiological, morphological, and gene expression changes have been studied within bacterial species as they enter stationary phase (smaller spherical cells with rigid cell envelope, membrane fluidity reduction, activation of stringent response (CodY), alternative sigma factor activation, activation of stationary phase promoters (growth stage limited) among others) (Aldea et al., 1993; Jaishankar & Srivastava, 2017). Therefore, slight error in sampling timing between end of log phase growth and early stationary phase could result in differential RNA expression profiles (Weiss, Borach and Shaw, 2016; Resch et al, 2005; Klumpp and Hwa, 2014; Klummp, Zhang and Hua, 2009).

5.5.2.1.2 Sample Preparation & Sequencing

Although the RNA-Seq transcript levels follow the same pattern, the WT strains had a differing mRNA signal when compared to the mutant strains, which seemed to have higher transcript levels and high GC content (WT – 32% GC, RM mutants – 36%). The GC-content effects on RNA-Seq read counts have been shown to substantially bias differential expression analysis, potentially due to additional PCR cycles (x35) of mutant RNA samples which were below the minimum sequencing concentration (Risso et al., 2011, Parekh et al., 2016). There were additional differences between the library prep kits with the WT strains being prepared with Illumina TruSeq stranded RNA kit whilst the *ΔhsdS* strains were prepared using NEB Ultra II stranded RNA kit. The two sets were also sequenced using different libraries the 2017 WT isolate sequencing carried out with Illumina-C Library PCR whilst the 2020 mutant isolates were sequenced with a novel Limber PCR Bespoke approach and were multiplexed. The differences in library preparation may have introduced some systematic error and technical issues regarding sequencing resolution (Sarantopoulou et al., 2019; Manga et al., 2016; Williams et al., 2014; Robles et al., 2012).

5.5.2.1.3 RNA-Seq Data Analysis and Differential Expression Analysis

Another limitation of this study was the lack of consensus for comprehensive differential expression analysis, especially for expression data normalisation and 'batch effect' corrections (Papiez et al., 2018; Fei and Yu, 2020; Soneson & Robinson, 2018; Chen et al., 2020; Tran et al., 2020). Determining cut-off thresholds without over-fitting or over-correction of data to yield statistically significant results is still based individual experiments, and variability in estimated gene expression among commonly used RNA-Seq pipelines was a key consideration in this work, hence running DE analysis through numerous analysis tools. Batch effect filtering was run in limma as it was the only tool which did not cause hyper-normalisation of input data, but even so the DE pairwise comparisons of mutant vs WT strains in high numbers of differentially expressed genes. It would be interesting to re-sequence some WT and RM mutant strains with RNA-Seq as a pilot study to evaluate the accuracy of the batch effect approach used in this study.

5.5.2.2 Future Work

Apart from sequencing and methylation validation of the indicated RM mutants, supplementary phenotypic experiments using Biolog Phenotype MicroArray, would greatly augment this study to investigate physiological effects resulting from the differential gene expression between WT and mutant strains. Furthermore, investigating DNA methylation under variable phenotypic conditions would greatly benefit our knowledge of the potential variability in methylation by Sau1 due to external stress or stimuli. Metabolomic profiling of these strains would highlight specific substrates/metabolites changes induced through the deletion of *hsdS,* and potentially pinpoint regulatory networks they may be linked to. Further studies reconstructing Sau1 6mA methylation and restriction activity in RM1 and RM2+3 mutants (no functional *hsdS*) and subsequently transcriptomic profiling of each strain would also be interesting to investigate, to explore if reintroduction of *hsdS* has the reverse gene expression effects noted in this study.

Although the lack of transcriptomic changes between mutant strains with differential methylation indicates 6mA does not have a direct epigenetic regulatory role in *S. aureus*, further investigation of methylation dependent indirect regulation should be considered. Detailed mapping of 6mA modifications overlapping with transcription factor binding sites (SigA, SigB and other regulator motifs (Mäder et al, 2016; Shell et al, 2013, Nye et al,

2020, Chimer-oms et al, 2019) could characterise the complex transcriptional regulatory network potentially affected by Sau1 producing the pleiotropic metabolic shift seen in this study.

6. GENERAL DISCUSSION & FUTURE DIRECTIONS

6.1 Summary of Key Findings

6.1.1 Chapter 3 – Species Wide Characterisation of RM Systems and TI RM Sau1 6mA Methylation in *Staphylococcus aureus*

This study characterised the Restriction-Modification landscape of *S. aureus* in a historically and phylogenetically divergent set of isolates from Public Health England's National Culture Type Collection. Comparative genomic analysis highlighted the presence of multiple RM systems with differing nucleotide specificities within any given isolate (TI Sau1 (6mA) + TII RM (5mC, 6mA) or TV R/RM (6mA, 5mC, 4mC), emphasising the complexity of RM activity within *S. aureus*, contributing to the increased control of HGT previously noted for the 'untransformable' species (Figure 3.2). The diverse NCTC collection was used to investigate the distribution and variability of both 'core' *sau1hsdSM* homologs in *v*Sa and additionally discovered four further 'accessory' MGE related *sau1hsdS_orfX* and *sau1hsdRMS* linked to the *orfX*/SCC*mec* insertion sites (most likely transferred from coagulase negative *staphylococci*), and two phage associated units, *sau1hsdS_ɸ* and *sau1hsdMS3* (Figure 3.3). Between the 24 represented STs, the detailed protein structure of 40 different HsdS homologs were characterised and matched to their corresponding 6mA target recognition sequences (motifs) predicted with PacBio SMRT Motif and Modification analysis (Table 3.4). These specificity proteins resolved into 18 TRD1 and 28 TRD2 domains, uncovering 6 novel TRD1 and 10 novel TRD2 protein domains for the Sau1 HsdS augmenting our knowledge of specific HsdS and their recognition targets (Table 3.5). Variants of HsdS and the relative conservation of HsdS stayed consistent within lineages, correlating to the allelic forms of genomic islands and MGE they were associated to (Figure 3.6). A single HsdS_β in CC97 exhibited recombination within the TRD1 domain, exemplifying the potential for differential methylation by heterologous specificity proteins within *S. aureus*. No evidence for phase-variable TI RM *hsdS* (or *hsdM*) were found.

Frequency analysis of TRS motifs matched in any given strain of *S. aureus* revealed 6mA methylation within the species is randomly distributed, with no hemi-methylation and no hyper/hypomethylated areas of the genome detected (Figure 3.11). However, the results suggest that there is methylation bias towards the coding sequence (Figure 3.12, Figure 3.15), and the core rather than accessory genome (Figure 3.18). Between the two 'core' HsdS, TRS recognised by HsdS_α were 31.6% higher than those for HsdS_β, although this is highly lineage specific (Figure 3.13), potentially indicating a functional difference between the two systems.

6.1.2 Chapter 4 – The Effect of Large-Scale Chromosomal Replacement on Whole Genome Methylation and Gene Expression Profiles *in Staphylococcus aureus*

Large-scale genome recombination events are rarely seen clonal species like *S. aureus,* and with little known about the biological impact of the introduction of chromosomal replacements have on chimeric strains (Everitt et al., 2014). Singaporean hybrid HA-MRSA strain ST622 provided a unique opportunity to study the effects of core genome transfers specifically on the methylation landscape of *S. aureus* to investigate the epigenetic potential of differential methylation.

Comparative genomic analysis revealed that the chromosomal exchange did not introduce/switch Sau1 *hsdS* alleles within the novel ST622, which retained the ST22 methylation signature, carrying 3 functional specificity units: *sau1hsdS_α*, *sau1hsdS_X*, and *sau1hsdS_S (*except ST622-2014 variant which only carried *sau1hsdS_α* due to recombinant sequence stretching over the SCC*mec* losing both accessory *hsdS* with the introduction of ST45 TV SCC*mec* – Figure 4.7*).* In contrast, ST45 isolates carried the two usual 'core' *hsdS* in either *v*Sa. PacBio SMRT sequencing was used to predict the 6mA methylation motifs for the 3 ST types, revealing 5 heterologous TRS, matched to each HsdS (Table 4.9). Analysis of each methylation motif density within the correct sequence background (ex: ST45 motifs in ST45 background), as well as mismatched ST backgrounds (ex: ST45 motifs in ST22/ST622 background), revealed that altering the *S. aureus* sequence background does not have a significant effect on methylation potential, regarding the total number and frequency of TRS matches within any given genic region (CDS, INT, core, accessory – Table 4.11).

The largescale core genome recombination encompassing the *oriC* and terminus of replication has minimal effect (2% shift) on the overall TRS detected between ST22 and ST622 daughter cells which share the same methylation signatures (Figure 4.8, Figure 4.9, Figure Figure 4.11- Figure 4.14). Conversely, detailed analysis of the chimeric region (CH) within the hybrid strains and equivalent sequence region between the ST22 revealed 28 instances of differential methylation (gain/loss of 6mA) of which 2/3 were in regulatory regions rather than CDS (Table 4.12). However, transcriptomic analysis suggests that these methylation changes have no direct effect on gene expression (Figure 4.23).

The recombinant sequence region within the hybrid ST622 provided an isogenic sequence segment (ST45 origin) to investigate the potential gene expression effect of clear

differential methylation by ST45 HsdS_α/HsdS_β and ST22 HsdS_α, HsdS_X, and HsdS_S. RNA-Seq and differential expression analysis of the resulting transcript data comparing solely the CH sequence region, highlighted 4 DE between ST45 and ST622 isolates, all with variable metabolic functions (Figure 4.23). None of the candidate genes had methylation motifs within their regulatory region (Figure 4.24), and thus direct moderation of gene expression through overlapping 6mA methylation and transcription factor binding motifs is unlikely. Although the direct effects on promoter binding within the represented DE genes could not be concluded for Sau1, 6mA methylation within *S. aureus* may have an indirect regulatory effect, in which DNA modifications in distal regulatory regions or modification of transcriptional regulators may cause differential expression of a cascade of genes.

6.1.3 Chapter 5 – Functional Impact of Sau1 Facilitated 6mA DNA Methylation in *Staphylococcus aureus*

Targeted allelic replacement mutagenesis of Sau1 *hsdS* was used to investigate whether Sau1 has a definitive role in modulating gene expression via direct or indirect mechanisms through 6mA methylation. A collection of isogenic mutant strains was created from the ST45 and ST622 (both variants) Singapore HA-MRSA isolates, to compare the transcriptomic effects of deleting differing *hsdS* (Δ*sau1hsdS_α*, Δ*sau1hsdS_β*, Δ*sau1hsdS_X*, Δ*sau1hsdS_S)* or loss of complete 6mA methylation through sequential knockout of all functional *hsdS* within a given strain (Figure 5.2 and Figure 5.3).

RNA-Seq and differential expression analysis of each Δ*hsdS* compared to parent WT strains suggests a pleotropic regulatory effect of core metabolic and biosynthetic genes (Figure 5.5, Figure 5.8, Figure 5.11). However, the pairwise comparison of mutant strains revealed identical expression profiles for Δ*hsdS* within a given lineage, indicating that loss of some or all 6mA methylation throughout the genome does not result in transcriptomic changes, and Sau1 does not likely have an epigenetic regulatory role *in S. aureus*. This also led to the conclusion that there are no functional differences between the differing HsdS within the host. Further analyses need to be conducted to complete this study and explore the mechanism by which this pleotropic core regulatory effect occurred as well as the definite secondary functional role of Sau1 within *S. aureus*.

6.2 Role of TI RM Sau1 6mA in *S. aureus*

The role of restriction modification systems in host defence was first identified over 60 years ago with TI RM EcoKI in *E. coli* (Loenen, 2003). Subsequently this archetypal defence system was found in various other bacterial species, along a multitude of other RM types, assumed to have similar roles in control of horizontal gene transfer (Oliveira et al., 2016; Thomas & Nielsen, 2005; Blow et al., 2016). However, it soon became apparent that DNA methylation could alter transcription, eliciting regulatory control of genes involved in metabolism, virulence and cell physiology (Vasu & Nagaraja, 2013; Casadesus, 2006; Heusipp, Fälker & Schmidt, 2007; Cohen et al., 2016; Casadesus & Low, 2006). Most of the epigenetic regulatory effects have been attributed to orphan methyltransferases or phase-variable TI and TIII RM systems but DNA methylation by few non-phase variable TI systems have also shown to affect gene expression (Blow et al., 2016, Wion & Casadesus, 2006; Seib et al., 2020; Oliveira & Fang, 2021; Manso et al., 2014; Nye et al., 2019, Dobrenez et al., 2017, Furuta et al., 2014, Mehershahi & Chen, 2021). The proposed epigenetic effect in non-phase variable TI RM have only been attributed to indirect methylation of transcriptional regulators which perturb the downstream regulation of virulence and iron metabolism associated genes as exemplified by differential modulation of Mga (multiple gene regulator of GAS) in *S. pyogenes* (Nye et al., 2019) and PrrF1 (small regulatory RNA) in *P. aeruginosa* (Dobrenze et al., 2017) through loss of 6mA methylation.

The role of Sau1 6mA methylation and restriction activity has been extensively characterised in *Staphylococcus aureus* (Lindsay, 2006; Waldron & Lindsay, 2006; McCarthy et al., 2012; Chen et al., 2016; Monk & Foster, 2012; Jones et al., 2015; Monk et al., 2015). This study aimed to investigate the potential secondary regulatory functions of Sau1 associated 6mA methylation and the mechanism of epigenetic control it may elicit. It was established that Sau1 was not a phase-varion, prompting no reversible switch between active and inactive form of genes through differential methylation. However, one example of a distinct specificity changes through recombination of HsdS TRDs (CC97) confirmed the potential for differential methylation switches within a lineage as previously demonstrated in CC5 by Sullivan et al., (2019). Transcriptomic analysis found little evidence for direct modulation of expression within the regulatory region of DE genes through the loss of methylation in knockout mutants and by variable methylation signatures in an isogenic sequence segment. No specific 6mA linked effector (regulator/repressor) responsible for indirect downstream regulatory effects was uncovered.

Analogous lack of direct regulatory effects linked to 6mA methylation were recently seen for multiple TI RM systems in four differing strains of *E. coli*, giving rise to the concept of 'regulation avoidance' (Mehershahi & Chen, 2021), perhaps also applicable for *S. aureus* Sau1. This model defines a condition in which changing the methylation state throughout the genome (loss / gain of methylation at same sites) have no significant impact on gene expression or phenotype, which may also be seen for Sau1 in *S. aureus* (Mehershahi & Chen, 2021). This lack of regulatory function is suspectedly linked to longer, bipartite target recognition sequences confined to TI RM which also elicit the fewest methylation motifs compared to DNA modifications by other RM types. Nonetheless, this is not a universal phenomenon as observed by the clear regulatory effect of phase variable RMs (De Ste Croix et al., 2017).

However, disruption of TI RM activity by deletion of DNA binding specific units has an indirect regulatory effect (by an unknown mechanism) seen through the differential expression of a large cohort of core metabolic genes between *ΔhsdS* and WT strains. Although this is strongly contrasted by the lack of differentially expressed genes between mutant isolates, further verification and phenotypic studies would determine the role of Sau1 6mA within *S. aureus*.

6.3 Future Perspectives

To gain a better understanding of the proposed indirect, global regulatory effect induced by the loss of 6mA, further verification of the differential expression results from Chapter 5 should be considered. Firstly, targeted quantitative RT-PCR on a selection of DE genes highlighted between *ΔhsdS* and WT strains within each lineage should be conducted using target-specific primers to validate the altered expression profiles between samples. Alternatively, resequencing of a select number of *ΔhsdS*, for example RM1, RM2 and RM2+3 created within the ST45 background which were also validated to have lost 6mA using PacBio SMRT sequencing, as a verification pilot study - including the WT strain within the same sequencing run should be conducted.

This study would also benefit from conducting phenotypic studies on both WT and mutant strains to gain insight into the physiological changes which loss of 6mA methylation may have induced not just in the one experimental growth condition as per RNA-Seq. This

could be done in the most comprehensive manner using Biolog Phenotype MicroArray technology to correlate genotypes with phenotypes (Biolog, 2020).

Another beneficial study would be extensive study of methylation of promoter regions, especially those of the 135 transcription factors and sigma factors identified within *S. aureus* (Ibarra et al., 2013). Furthermore, mapping binding motifs of Sigma A, Sigma B (mapped by Mäder et al., 2016) and any other TF with predicted DNA binding sites, to see any overlaps with 6mA methylation motifs within multiple lineages would be interesting to investigate. This may shed light on any 6mA induced regulatory changes and how that translates to the overall modulation of the transcriptional regulatory network, previously investigated in *B. subtilis* (Nye et al., 2020), *M. tuberculosis* (Chiners-Oms et al., 2019; Shell et al., 2013).

Further investigation of the complete methylation landscape of *S. aureus* would be important to gain a comprehensive view of the potential regulatory functions of DNA methylation within the species. It is not only Sau1 which has RM activity with 6mA specificity within *S. aureus*, but also TIIG RM *bcgIAB*. Within this body of work, two different homologs of the BcgIAB systems were found, previously also described in Chapter 3 (Table 3.4 - TRS #41 (CC22, CC30), #42 (CC131)). Perhaps further investigation of methylation by this fused MTase/REase would uncover potential secondary functions as seen for TIIG Cj0031 in *C. jejuni* linked to broad-scale gene expression changes in a wide range of cellular functions (Anjum et al., 2016). This could also be further translated to investigating the role of 5mC signatures in *S. aureus* facilitated by the TII *dcm/sau3AIR, hhaIM/cmoA*, and *ssoII/ecoRII* characterised throughout the NCTC collection (Table 3.2). The frequency of 5mC methylation and modification motifs could be investigated directly through Nanopore sequencing (McIntyre et al., 2020), or PacBio SMRT sequencing with TET conversion for accurate 5mC detection.

7. REFERENCES

Aanensen, D. M., Feil, E. J., Holden, M. T. G., Dordel, J., Yeats, C. A., Fedosejev, A., … Kearns, A. (2016). Whole-genome sequencing for routine pathogen surveillance in public health: A population snapshot of invasive Staphylococcus aureus in Europe. *MBio*, *7*(3), e00444-16–15. https://doi.org/10.1128/mBio.00444-16

Adamczyk-Popławska, M., Kondrzycka, A., Urbanek, K., & Piekarowicz, A. (2003). Tetra-amino-acid tandem repeats are involved in HsdS complementation in type IC restriction - Modification systems. *Microbiology*, *149*(11), 3311–3319. https://doi.org/10.1099/mic.0.26497-0

Adamczyk-Poplawska, M., Lower, M., & Piekarowicz, A. (2009). Characterization of the NgoAXP: Phase-variable type III restriction-modification system in Neisseria gonorrhoeae: Phase-variable. *FEMS Microbiology Letters*, *300*(1), 25–35. https://doi.org/10.1111/j.1574-6968.2009.01760.x

Adamczyk-Poplawska, M., Lower, M., & Piekarowicz, A. (2011). Deletion of one nucleotide within the homonucleotide tract present in the hsdS gene alters the DNA sequence specificity of type I restriction-modification system NgoAV. *Journal of Bacteriology*, *193*(23), 6750–6759. https://doi.org/10.1128/JB.05672-11

Adhikari, S., & Curtis, P. D. (2016). DNA methyltransferases and epigenetic regulation in bacteriaa. *FEMS Microbiology Reviews*, *40*(5), 575–591. https://doi.org/10.1093/femsre/fuw023

Aires-de-Sousa, M. (2017). Methicillin-resistant Staphylococcus aureus among animals: current overview. *Clinical Microbiology and Infection*, *23*(6), 373–380. https://doi.org/10.1016/j.cmi.2016.11.002

Alibayov, B., Zdeňková, K., Purkrtová, S., Demnerová, K., & Karpíšková, R. (2014). Detection of some phenotypic and genotypic characteristics of Staphylococcus aureus isolated from food items in the Czech Republic. *Annals of Microbiology*, *64*(4), 1587–1596. https://doi.org/10.1007/s13213-014-0802-6

Aliberti, S., Reyes, L. F., Faverio, P., Sotgiu, G., Dore, S., Rodriguez, A. H., … Labra, L. (2016). Global initiative for meticillin-resistant Staphylococcus aureus pneumonia (GLIMP): an international, observational cohort study. *The Lancet Infectious Diseases*, *16*(12), 1364–1376. https://doi.org/10.1016/S1473-3099(16)30267-5

Alikhan, N. F., Petty, N. K., Ben Zakour, N. L., & Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genomics*, *12*(1), 402. https://doi.org/10.1186/1471-2164-12-402

Alvarez, L., Mingorance, J., Pajares, M. A., & Mato, J. M. (1994). Expression of rat liver S-adenosylmethionine synthetase in Escherichia coli results in two active oligomeric forms. *Biochemical Journal*, *301*(2), 557–561. https://doi.org/10.1042/bj3010557

Anderson, D. J., Harris, S. R., Godofsky, E., Toriscelli, T., Rude, T. H., Elder, K., … Peacock, S. J. (2014). Whole genome sequencing of a methicillin-resistant Staphylococcus aureus pseudo-outbreak in a professional football team. *Open Forum Infectious Diseases*, *1*(3). https://doi.org/10.1093/ofid/ofu096

Anderson, K. L., Roberts, C., Disz, T., Vonstein, V., Hwang, K., Overbeek, R., … Dunman, P. M. (2006). Characterization of the Staphylococcus aureus heat shock, cold shock, stringent, and SOS responses and their effects on log-phase mRNA turnover. *Journal of Bacteriology*, *188*(19), 6739–6756. https://doi.org/10.1128/JB.00609-06

Anjum, A., Brathwaite, K. J., Aidley, J., Connerton, P. L., Cummings, N. J., Parkhill, J., … Bayliss, C. D. (2016). Phase variation of a Type IIG restriction-modification enzyme alters site-specific methylation patterns and gene expression in Campylobacter jejuni strain NCTC11168. *Nucleic Acids Research*, *44*(10), 4581–4594. https://doi.org/10.1093/nar/gkw019

Aravind, L., Maxwell Burroughs, A., Zhang, D., & Iyer, L. M. (2014). Protein and DNA modifications: Evolutionary imprints of bacterial biochemical diversification and geochemistry on the provenance of eukaryotic epigenetics. *Cold Spring Harbor Perspectives in Biology*, *6*(7), a016063–a016063. https://doi.org/10.1101/cshperspect.a016063

Arber, W. (2000). Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiology Reviews*, *24*(1), 1–7. https://doi.org/10.1111/j.1574-6976.2000.tb00529.x

Atack, J. M., Guo, C., Litfin, T., Yang, L., Blackall, P. J., Zhou, Y., & Jennings, M. P. (2020). Systematic Analysis of REBASE Identifies Numerous Type I Restriction-Modification Systems with Duplicated, Distinct hsdS Specificity Genes That Can Switch System Specificity by Recombination . *MSystems*, *5*(4). https://doi.org/10.1128/msystems.00497-20

Atack, J. M., Tan, A., Bakaletz, L. O., Jennings, M. P., & Seib, K. L. (2018). Phasevarions of Bacterial Pathogens: Methylomics Sheds New Light on Old Enemies. *Trends in Microbiology*, *26*(8), 715–726. https://doi.org/10.1016/j.tim.2018.01.008

Au, K. G., Welsh, K., & Modrich, P. (1992). Initiation of methyl-directed mismatch repair. *The Journal of Biological Chemistry*, *267*(17), 12142–12148.

Baba, T., Bae, T., Schneewind, O., Takeuchi, F., & Hiramatsu, K. (2008). Genome sequence of Staphylococcus aureus strain newman and comparative analysis of staphylococcal genomes: Polymorphism and evolution of two major pathogenicity islands. *Journal of Bacteriology*, *190*(1), 300–310. https://doi.org/10.1128/JB.01000-07

Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K. I., Oguchi, A., … Hiramatsu, K. (2002). Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet*, *359*(9320), 1819–1827. https://doi.org/10.1016/S0140-6736(02)08713-5

Bakker, A., & Smith, D. W. (1989). Methylation of GATC sites is required for precise timing between rounds of DNA replication in Escherichia coli. *Journal of Bacteriology*, *171*(10), 5738 LP – 5742. https://doi.org/10.1128/jb.171.10.5738-5742.1989

Bal, A. M., Coombs, G. W., Holden, M. T. G., Lindsay, J. A., Nimmo, G. R., Tattevin, P., & Skov, R. L. (2016). Genomic insights into the emergence and spread of international clones of healthcare-, community- and livestock-associated meticillin-resistant Staphylococcus aureus: Blurring of the traditional definitions. *Journal of Global Antimicrobial Resistance*, *6*, 95–101. https://doi.org/10.1016/j.jgar.2016.04.004

Bart, A., Pannekoek, Y., Dankert, J., & Van Der Ende, A. (2001). NmeSI restriction-modification system identified by representational difference analysis of a hypervirulent Neisseria meningitidis strain. *Infection and Immunity*, *69*(3), 1816–1820. https://doi.org/10.1128/IAI.69.3.1816-1820.2001

Bart, A., van Passel, M. W. J., van Amsterdam, K., & van der Ende, A. (2005). Direct detection of methylation in genomic DNA. *Nucleic Acids Research*, *33*(14), 1–6. https://doi.org/10.1093/nar/gni121

Basith, S., Manavalan, B., Shin, T. H., & Lee, G. (2019). SDM6A: A Web-Based Integrative Machine-Learning Framework for Predicting 6mA Sites in the Rice Genome. *Molecular Therapy - Nucleic Acids*, *18*, 131–141. https://doi.org/10.1016/j.omtn.2019.08.011

Beaulaurier, J., Schadt, E. E., & Fang, G. (2019). Deciphering bacterial epigenomes using modern sequencing technologies. *Nature Reviews Genetics*, *20*(3), 157–172. https://doi.org/10.1038/s41576-018-0081-3

Beaulaurier, J., Zhang, X. S., Zhu, S., Sebra, R., Rosenbluh, C., Deikus, G., … Fang, G. (2015). Single molecule-level detection and long read-based phasing of epigenetic variations in

bacterial methylomes. *Nature Communications*, *6*(1), 7438. https://doi.org/10.1038/ncomms8438

Bendall, M. L., Luong, K., Wetmore, K. M., Blow, M., Korlach, J., Deutschbauer, A., & Malmstrom, R. R. (2013). Exploring the roles of DNA methylation in the metal-reducing bacterium shewanella oneidensis MR-1. *Journal of Bacteriology*, *195*(21), 4966–4974. https://doi.org/10.1128/JB.00935-13

Berg, T., Firth, N., Apisiridej, S., Hettiaratchi, A., Leelaporn, A., & Skurray, R. A. (1998). Complete nucleotide sequence of pSK41: Evolution of staphylococcal conjugative multiresistance plasmids. *Journal of Bacteriology*, *180*(17), 4350–4359. https://doi.org/10.1128/jb.180.17.4350-4359.1998

Bhagwat, A. S., & Lieb, M. (2002). Cooperation and competition in mismatch repair: very short-patch repair and methyl-directed mismatch repair in Escherichia coli. *Molecular Microbiology*, *44*(6), 1421–1428. https://doi.org/https://doi.org/10.1046/j.1365-2958.2002.02989.x

Bickle, T. A., & Kruger, D. H. (1993). Biology of DNA restriction. *Microbiological Reviews*, *57*(2), 434–450. https://doi.org/10.1128/mmbr.57.2.434-450.1993

Bierne, H., Hamon, M., & Cossart, P. (2012). Epigenetics and bacterial infections. *Cold Spring Harbor Perspectives in Medicine*, *2*(12), a010272–a010272. https://doi.org/10.1101/cshperspect.a010272

Biolog. (2020). Phenotype Micrarrays for Microbial Cells.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes and Development*, *16*(1), 6–21. https://doi.org/10.1101/gad.947102

Bischoff, M., Dunman, P., Kormanec, J., Macapagal, D., Murphy, E., Mounts, W., … Projan, S. (2004). Microarray-based analysis of the Staphylococcus aureus σB regulon. *Journal of Bacteriology*, *186*(13), 4085–4099. https://doi.org/10.1128/JB.186.13.4085-4099.2004

Blair, J. M. A., Richmond, G. E., Bailey, A. M., Ivens, A., & Piddock, L. J. V. (2013). Choice of Bacterial Growth Medium Alters the Transcriptome and Phenotype of Salmonella enterica Serovar Typhimurium. *PLoS ONE*, *8*(5), e63912. https://doi.org/10.1371/journal.pone.0063912

Blakeway, L. V., Power, P. M., Jen, F. E. C., Worboys, S. R., Boitano, M., Clark, T. A., … Seib, K. L. (2014). ModM DNA methyltransferase methylome analysis reveals a potential role for Moraxella catarrhalis phasevarions in otitis media. *FASEB Journal*, *28*(12), 5197–5207. https://doi.org/10.1096/fj.14-256578

Blow, M. J., Clark, T. A., Daum, C. G., Deutschbauer, A. M., Fomenkov, A., Fries, R., … Roberts, R. J. (2016). The Epigenomic Landscape of Prokaryotes. *PLoS Genetics*, *12*(2), e1005854-28. https://doi.org/10.1371/journal.pgen.1005854

Blumenthal, R. M., & Cheng, X. (2002). *Restriction-Modification Systems*.

Blyn, L. B., Braaten, B. A., White-Ziegler, C. A., Rolfson, D. H., & Low, D. A. (1989). Phase-variation of pyelonephritis-associated pili in Escherichia coli: evidence for transcriptional regulation. *The EMBO Journal*, *8*(2), 613–620.

Bock, C. (2012). Analysing and interpreting DNA methylation data. *Nature Reviews. Genetics*, Vol. 13, pp. 705–719. https://doi.org/10.1038/nrg3273

Bogan, J. A., & Helmstetter, C. E. (1997). DNA sequestration and transcription in the oriC region of Escherichia coli. *Molecular Microbiology*, *26*(5), 889–896.

https://doi.org/https://doi.org/10.1046/j.1365-2958.1997.6221989.x

Boissy, R., Ahmed, A., Janto, B., Earl, J., Hall, B. G., Hogg, J. S., … Hu, F. Z. (2011). Comparative supragenomic analyses among the pathogens Staphylococcus aureus, Streptococcus pneumoniae, and Haemophilus influenzae Using a modification of the finite supragenome model. *BMC Genomics*, *12*(1), 187. https://doi.org/10.1186/1471-2164-12-187

Bosi, E., Monk, J. M., Aziz, R. K., Fondi, M., Nizet, V., & Palsson, B. (2016). Comparative genome-scale modelling of Staphylococcus aureus strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(26), E3801–E3809. https://doi.org/10.1073/pnas.1523199113

Bourges, A. C., Torres Montaguth, O. E., Ghosh, A., Tadesse, W. M., Declerck, N., Aertsen, A., & Royer, C. A. (2017). High pressure activation of the Mrr restriction endonuclease in Escherichia coli involves tetramer dissociation. *Nucleic Acids Research*, *45*(9), 5323–5332. https://doi.org/10.1093/nar/gkx192

Braaten, B. A., Blyn, L. B., Skinner, B. S., & Low, D. A. (1991). Evidence for a methylation-blocking factor (mbf) locus involved in pap pilus expression and phase variation in Escherichia coli. *Journal of Bacteriology*, *173*(5), 1789–1800. https://doi.org/10.1128/jb.173.5.1789-1800.1991

Bradley, S. F. (2002). Staphylococcus aureus infections and antibiotic resistance in older adults. *Clinical Infectious Diseases*, *34*(2), 211–216. https://doi.org/10.1086/338150

Braun, R. E., & Wright, A. (1986). DNA methylation differentially enhances the expression of one of the two E. coli dnaA promoters in vivo and in vitro. *MGG Molecular & General Genetics*, *202*(2), 246–250. https://doi.org/10.1007/BF00331644

Broadbent, S. E., Balbontin, R., Casadesus, J., Marinus, M. G., & Van Der Woude, M. (2007). YhdJ, a nonessential CcrM-like DNA methyltransferase of Escherichia coli and Salmonella enterica. *Journal of Bacteriology*, *189*(11), 4325–4327. https://doi.org/10.1128/JB.01854-06

Brochet, M., Rusniok, C., Couvé, E., Dramsi, S., Poyart, C., Trieu-Cuot, P., … Glaser, P. (2008). Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of Streptococcus agalactiae. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(41), 15961–15966. https://doi.org/10.1073/pnas.0803654105

Brockman, K. L., Jurcisek, J. A., Atack, J. M., Srikhanta, Y. N., Jennings, M. P., & Bakaletz, L. O. (2016). ModA2 phasevarion switching in nontypeable haemophilus influenzae increases the severity of experimental otitis media. *Journal of Infectious Diseases*, *214*(5), 817–824. https://doi.org/10.1093/infdis/jiw243

Brouwer, M. S. M., Roberts, A. P., Hussain, H., Williams, R. J., Allan, E., & Mullany, P. (2013). Horizontal gene transfer converts non-toxigenic Clostridium difficile strains into toxin producers. *Nature Communications*, *4*(1), 1–6. https://doi.org/10.1038/ncomms3601

Budd, K. E., McCoy, F., Monecke, S., Cormican, P., Mitchell, J., & Keane, O. M. (2015). Extensive Genomic Diversity among Bovine-Adapted Staphylococcus aureus: Evidence for a Genomic Rearrangement within CC97. *PLOS ONE*, *10*(8), e0134592. Retrieved from https://doi.org/10.1371/journal.pone.0134592

Burroughs, A. M., Zhang, D., Schäffer, D. E., Iyer, L. M., & Aravind, L. (2015). Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Research*, *43*(22), 10633–10654. https://doi.org/10.1093/nar/gkv1267

Camacho, E. M., & Casadesús, J. (2002). Conjugal transfer of the virulence plasmid of Salmonella enterica is regulated by the leucine-responsive regulatory protein and DNA adenine

methylation. *Molecular Microbiology*, *44*(6), 1589–1598. https://doi.org/10.1046/j.1365-2958.2002.02981.x

Camacho, E. M., & Casadesús, J. (2005). Regulation of traJ transcription in the Salmonella virulence plasmid by strand-specific DNA adenine hemimethylation. *Molecular Microbiology*, *57*(6), 1700–1718. https://doi.org/10.1111/j.1365-2958.2005.04788.x

Campbell, J. L., & Kleckner, N. (1990). E. coli oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell*, *62*(5), 967–979. https://doi.org/10.1016/0092-8674(90)90271-F

Capra, E., Cremonesi, P., Pietrelli, A., Puccio, S., Luini, M., Stella, A., & Castiglioni, B. (2017). Genomic and transcriptomic comparison between Staphylococcus aureus strains associated with high and low within herd prevalence of intra-mammary infection. *BMC Microbiology*, *17*(1), 1–16. https://doi.org/10.1186/s12866-017-0931-8

Carver, T., Berriman, M., Tivey, A., Patel, C., Böhme, U., Barrell, B. G., … Rajandream, M.-A. (2008). Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics (Oxford, England)*, *24*(23), 2672–2676. https://doi.org/10.1093/bioinformatics/btn529

Carver, T., Harris, S. R., Berriman, M., Parkhill, J., & McQuillan, J. A. (2012). Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*, *28*(4), 464–469. https://doi.org/10.1093/bioinformatics/btr703

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G., & Parkhill, J. (2005). ACT: The Artemis comparison tool. *Bioinformatics*, *21*(16), 3422–3423. https://doi.org/10.1093/bioinformatics/bti553

Casadesús, J. (2016). Bacterial DNA Methylation and Methylomes. *Advances in Experimental Medicine and Biology*, *945*, 35–61. https://doi.org/10.1007/978-3-319-43624-1_3

Casadesús, J., & Low, D. (2006). Epigenetic Gene Regulation in the Bacterial World. *Microbiology and Molecular Biology Reviews*, *70*(3), 830–856. https://doi.org/10.1128/mmbr.00016-06

Casadesús, J., & Low, D. A. (2013). Programmed heterogeneity: Epigenetic mechanisms in bacteria. *Journal of Biological Chemistry*, *288*(20), 13929–13935. https://doi.org/10.1074/jbc.R113.472274

Casselli, T., Tourand, Y., Scheidegger, A., Arnold, W. K., Proulx, A., Stevenson, B., & Brissette, C. A. (2018). DNA methylation by restriction modification systems affects the global transcriptome profile in Borrelia burgdorferi. *Journal of Bacteriology*, *200*(24), 395–413. https://doi.org/10.1128/JB.00395-18

Chastanet, A., Fert, J., & Msadek, T. (2003). Comparative genomics reveal novel heat shock regulatory mechanisms in Staphylococcus aureus and other Gram-positive bacteria. *Molecular Microbiology*, *47*(4), 1061–1073. https://doi.org/https://doi.org/10.1046/j.1365-2958.2003.03355.x

Chatterjee, S. S., Chen, L., Joo, H. S., Cheung, G. Y. C., Kreiswirth, B. N., & Otto, M. (2011). Distribution and regulation of the mobile genetic element-encoded phenol-soluble modulin PSM-mec in methicillin-resistant Staphylococcus aureus. *PLoS ONE*, *6*(12), 1–6. https://doi.org/10.1371/journal.pone.0028781

Chaves-Moreno, D., Wos-Oxley, M. L., Jáuregui, R., Medina, E., Oxley, A. P., & Pieper, D. H. (2016). Exploring the transcriptome of Staphylococcus aureus in its natural niche. *Scientific Reports*, *6*(1), 1–11. https://doi.org/10.1038/srep33174

Chen, K., Reuter, M., Sanghvi, B., Roberts, G. A., Cooper, L. P., Tilling, M., … Dryden, D. T. F. (2014). ArdA proteins from different mobile genetic elements can bind to the EcoKI Type i DNA methyltransferase of E. coli K12. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1844*(3), 505–511. https://doi.org/10.1016/j.bbapap.2013.12.008

Chen, K., Stephanou, A. S., Roberts, G. A., White, J. H., Cooper, L. P., Houston, P. J., … Dryden, D. T. F. (2016). The type I restriction enzymes as barriers to horizontal gene transfer: Determination of the DNA target sequences recognised by livestock-associated methicillin-resistant Staphylococcus aureus clonal complexes 133/st771 and 398. In *Advances in Experimental Medicine and Biology* (Vol. 915, pp. 81–97). https://doi.org/10.1007/978-3-319-32189-9_7

Chen, W., Zhang, S., Williams, J., Ju, B., Shaner, B., Easton, J., … Chen, X. (2020). A comparison of methods accounting for batch effects in differential expression analysis of UMI count based single cell RNA sequencing. *Computational and Structural Biotechnology Journal*, *18*, 861–873. https://doi.org/10.1016/j.csbj.2020.03.026

Chen, X., Wang, W. K., Han, L. Z., Liu, Y., Zhang, H., Tang, J., … Ni, Y. X. (2013). Epidemiological and Genetic Diversity of Staphylococcus aureus Causing Bloodstream Infection in Shanghai, 2009-2011. *PLoS ONE*, *8*(9), 2009–2011. https://doi.org/10.1371/journal.pone.0072811

Cheng, X., & Roberts, R. J. (2001). AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Research*, *29*(18), 3784–3795. https://doi.org/10.1093/nar/29.18.3784

Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., … Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*, *10*(6), 563–569. https://doi.org/10.1038/nmeth.2474

Chin, V., Valinluck, V., Magaki, S., & Ryu, J. (2004). KpnBI is the prototype of a new family (IE) of bacterial type I restriction-modification system. *Nucleic Acids Research*, *32*(18). https://doi.org/10.1093/nar/gnh134

Chiner-Oms, Á., Berney, M., Boinett, C., González-Candelas, F., Young, D. B., Gagneux, S., … Comas, I. (2019). Genome-wide mutational biases fuel transcriptional diversity in the Mycobacterium tuberculosis complex. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-11948-6

Choo, E. J., & Chambers, H. F. (2016). Treatment of methicillin-resistant Staphylococcus aureus bacteremia. *Infection and Chemotherapy*, *48*(4), 267–273. https://doi.org/10.3947/ic.2016.48.4.267

Chow, A., Lim, V. W., Khan, A., Pettigrew, K., Lye, D. C. B., Kanagasabai, K., … Holden, M. T. G. (2017). MRSA transmission dynamics among interconnected acute, intermediate-term, and long-term healthcare facilities in Singapore. *Clinical Infectious Diseases*, *64*(suppl_2), S76–S81. https://doi.org/10.1093/cid/cix072

Claesson, M. J., Li, Y., Leahy, S., Canchaya, C., Van Pijkeren, J. P., Cerdeño-Tárraga, A. M., … O'Toole, P. W. (2006). Multireplicon genome architecture of Lactobacillus salivarius. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(17), 6718–6723. https://doi.org/10.1073/pnas.0511060103

Clark, T. A., Lu, X., Luong, K., Dai, Q., Boitano, M., Turner, S. W., … Korlach, J. (2013). Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. *BMC Biology*, *11*(1), 4. https://doi.org/10.1186/1741-7007-11-4

Clarke, J., Wu, H. C., Jayasinghe, L., Patel, A., Reid, S., & Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology*, *4*(4),

265–270. https://doi.org/10.1038/nnano.2009.12

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

Cohen, M. L. (2000). Changing patterns of infectious disease : Article : Nature. *Nature*, *406*(6797), 762–767. Retrieved from http://www.nature.com/doifinder/10.1038/35021206%5Cnpapers3://publication/doi/10.1038/35021206

Collier, J., McAdams, H. H., & Shapiro, L. (2007). A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(43), 17111–17116. https://doi.org/10.1073/pnas.0708112104

Cooper, L. P., Roberts, G. A., White, J. H., Luyten, Y. A., Bower, E. K. M., Morgan, R. D., … Dryden, D. T. F. (2017). DNA target recognition domains in the Type i restriction and modification systems of Staphylococcus aureus. *Nucleic Acids Research*, *45*(6), 3395–3406. https://doi.org/10.1093/nar/gkx067

Corvaglia, A. R., François, P., Hernandez, D., Perron, K., Linder, P., & Schrenzel, J. (2010). A type III-like restriction endonuclease functions as a major barrier to horizontal gene transfer in clinical Staphylococcus aureus strains. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(26), 11954–11958. https://doi.org/10.1073/pnas.1000489107

Cota, I., Blanc-Potard, A. B., & Casadesús, J. (2012). STM2209-STM2208 (opvAB): a phase variation locus of Salmonella enterica involved in control of O-antigen chain length. *PloS One*, *7*(5), e36863. https://doi.org/10.1371/journal.pone.0036863

Cowley, L. A., Petersen, F. C., Junges, R., Jimson D. Jimenez, M., Morrison, D. A., & Hanage, W. P. (2018). Evolution via recombination: Cell-to-cell contact facilitates larger recombination events in Streptococcus pneumoniae. *PLoS Genetics*, *14*(6), e1007410. https://doi.org/10.1371/journal.pgen.1007410

De Backer, O., & Colson, C. (1991). Identification of the recognition sequence for the M· StyLTI methyltransferase of Salmonella typhimurium LT7: an asymmetric site typical of type-III enzymes. *Gene*, *97*(1), 103–107. https://doi.org/10.1016/0378-1119(91)90015-4

De Been, M., Van Schaik, W., Cheng, L., Corander, J., & Willems, R. J. (2013). Recent recombination events in the core genome are associated with adaptive evolution in Enterococcus faecium. *Genome Biology and Evolution*, *5*(8), 1524–1535. https://doi.org/10.1093/gbe/evt111

de Jong, A., Pietersma, H., Cordes, M., Kuipers, O. P., & Kok, J. (2012). PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC Genomics*, *13*(1), 299. https://doi.org/10.1186/1471-2164-13-299

de Lannoy, C., de Ridder, D., & Risse, J. (2017). The long reads ahead: De novo genome assembly using the MinION. *F1000Research*, Vol. 6. https://doi.org/10.12688/f1000research.12012.2

de Oliveira, T. L. R., Cavalcante, F. S., Chamon, R. C., Ferreira, R. B. R., & dos Santos, K. R. N. (2019). Genetic mutations in the quinolone resistance-determining region are related to changes in the epidemiological profile of methicillin-resistant Staphylococcus aureus isolates. *Journal of Global Antimicrobial Resistance*, *19*, 236–240. https://doi.org/10.1016/j.jgar.2019.05.026

De Ste Croix, M., Chen, K. Y., Vacca, I., Manso, A. S., Johnston, C., Polard, P., … Oggioni, M. R.

(2019). Recombination of the Phase-Variable spnIII Locus Is Independent of All Known Pneumococcal Site-Specific Recombinases. *Journal of Bacteriology*, *201*(15). https://doi.org/10.1128/JB.00233-19

de Vries, N., Duinsbergen, D., Kuipers, E. J., Pot, R. G. J., Wiesenekker, P., Penn, C. W., … Kusters, J. G. (2002). Transcriptional phase variation of a type III restriction-modification system in Helicobacter pylori. *Journal of Bacteriology*, *184*(23), 6615–6623. https://doi.org/10.1128/jb.184.23.6615-6624.2002

Deamer, D., Akeson, M., & Branton, D. (2016, May 6). Three decades of nanopore sequencing. *Nature Biotechnology*, Vol. 34, pp. 518–524. https://doi.org/10.1038/nbt.3423

Dearborn, A. D., & Dokland, T. (2012). Mobilization of pathogenicity islands by Staphylococcus aureus strain Newman bacteriophages. *Bacteriophage*, *2*(2), 70–78. https://doi.org/10.4161/bact.20632

Deghorain, M., & Van Melderen, L. (2012). The staphylococci phages family: An overview. *Viruses*, *4*(12), 3316–3335. https://doi.org/10.3390/v4123316

Deurenberg, R. H., & Stobberingh, E. E. (2008). The evolution of Staphylococcus aureus. *Infection, Genetics and Evolution*, *8*(6), 747–763. https://doi.org/10.1016/j.meegid.2008.07.007

Deurenberg, R. H., Vink, C., Oudhuis, G. J., Mooij, J. E., Driessen, C., Coppens, G., … Stobberingh, E. E. (2005). Different Clonal Complexes of Methicillin-Resistant. *Antimicrobial Agents and Chemotherapy*, *49*(10), 4263–4271. https://doi.org/10.1128/AAC.49.10.4263

Diekmann, S. (1987). DNA methylation can enhance or induce DNA curvature. In *The EMBO journal* (Vol. 6). https://doi.org/10.1002/j.1460-2075.1987.tb02769.x

Doberenz, S., Eckweiler, D., Reichert, O., Jensen, V., Bunk, B., Spröer, C., … Doberenz, C. S. (2017). Identification of a Pseudomonas aeruginosa PAO1 DNA Methyltransferase, Its Targets, and Physiological Roles Identification of a Pseudomonas aeruginosa PAO1 DNA methyltransferase, its targets, and physiological roles Downloaded from. *MBio*, *8*(1). https://doi.org/10.1128/mBio.02312-16

Dodson, K. W., & Berg, D. E. (1989). Factors affecting transposition activity of IS50 and Tn5 ends. *Gene*, *76*(2), 207–213. https://doi.org/10.1016/0378-1119(89)90161-3

Doulgeraki, A. I., Di Ciccio, P., Ianieri, A., & Nychas, G. J. E. (2017). Methicillin-resistant food-related Staphylococcus aureus: a review of current knowledge and biofilm formation for future studies and applications. *Research in Microbiology*, *168*(1), 1–15. https://doi.org/10.1016/j.resmic.2016.08.001

Dryden, D. T. F., Murray, N. E., & Rao, D. N. (2001). Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Research*, *29*(18), 3728–3741. https://doi.org/10.1093/nar/29.18.3728

Dybvig, K., Sitaraman, R., & French, C. T. (1998). A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(23), 13923–13928. https://doi.org/10.1073/pnas.95.23.13923

Dybvig, K., & Yu, H. (1994). Regulation of a restriction and modification system via DNA inversion in Mycoplasma pulmonis. *Molecular Microbiology*, *12*(4), 547–560. https://doi.org/10.1111/j.1365-2958.1994.tb01041.x

Dye, C. (2014). After 2015: Infectious diseases in a new era of health and development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1645),

20130426–20130426. https://doi.org/10.1098/rstb.2013.0426

Ebruke, C., Dione, M. M., Walter, B., Worwui, A., Adegbola, R. A., Roca, A., & Antonio, M. (2016). High genetic diversity of Staphylococcus aureus strains colonising the nasopharynx of Gambian villagers before widespread use of pneumococcal conjugate vaccines. *BMC Microbiology*, *16*(1), 38. https://doi.org/10.1186/s12866-016-0661-3

Ellington, M. J., Reuter, S., Harris, S. R., Holden, M. T. G., Cartwright, E. J., Greaves, D., … Peacock, S. J. (2015). Emergent and evolving antimicrobial resistance cassettes in community-associated fusidic acid and meticillin-resistant Staphylococcus aureus. *International Journal of Antimicrobial Agents*, *45*(5), 477–484. https://doi.org/10.1016/j.ijantimicag.2015.01.009

Enright, M. C. (2008). The Population Structure of Staphylococcus aureus. In J. A. Lindsay (Ed.), *Staphylococcus: molecluar genetics* (pp. 29–44). Caister Academic Press.

Epshtein, V., Mironov, A. S., & Nudler, E. (2003). The riboswitch-mediated control of sulfur metabolism in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(9), 5052–5056. https://doi.org/10.1073/pnas.0531307100

Ershova, A., Rusinov, I., Vasiliev, M., Spirin, S., & Karyagina, A. (2016). Restriction-Modification systems interplay causes avoidance of GATC site in prokaryotic genomes. *Journal of Bioinformatics and Computational Biology*, *14*(2), 1641003–1641019. https://doi.org/10.1142/S0219720016410031

Eustáquio, A. S., Härle, J., Noel, J. P., & Moore, B. S. (2008). S-adenosyl-L-methionine hydrolase (adenosine-forming), a conserved bacterial and archeal protein related to SAM-dependent halogenases. *ChemBioChem*, *9*(14), 2215–2219. https://doi.org/10.1002/cbic.200800341

Everitt, R. G., Didelot, X., Batty, E. M., Miller, R. R., Knox, K., Young, B. C., … Wilson, D. J. (2014). Mobile elements drive recombination hotspots in the core genome of Staphylococcus aureus. *Nature Communications*, *5*, 1–9. https://doi.org/10.1038/ncomms4956

Falquet, L., & Loetscher, A. (2015). PACMAN: PacBio Methylation Analyzer. *EMBnet.Journal*, *21*(A), 807. https://doi.org/10.14806/ej.21.a.807

Fang, G., Munera, D., Friedman, D. I., Mandlik, A., Chao, M. C., Banerjee, O., … Schadt, E. E. (2012). Genome-wide mapping of methylated adenine residues in pathogenic Escherichia coli using single-molecule real-time sequencing. *Nature Biotechnology*, *30*(12), 1232–1239. https://doi.org/10.1038/nbt.2432

Fei, T., & Yu, T. (2020). ScBatch: Batch-effect correction of RNA-seq data through sample distance matrix adjustment. *Bioinformatics*, *36*(10), 3115–3123. https://doi.org/10.1093/bioinformatics/btaa097

Feil, E. J., Cooper, J. E., Grundmann, H., Robinson, D. A., Enright, M. C., Berendt, T., … Day, N. P. J. (2003). How clonal is Staphylococcus aureus? *Journal of Bacteriology*, *185*(11), 3307–3316. https://doi.org/10.1128/JB.185.11.3307-3316.2003

Feil, E. J., Nickerson, E. K., Chantratita, N., Wuthiekanun, V., Srisomang, P., Cousins, R., … Peacock, S. J. (2008). Rapid detection of the pandemic methicillin-resistant Staphylococcus aureus clone ST 239, a dominant strain in Asian hospitals. *Journal of Clinical Microbiology*, *46*(4), 1520–1522. https://doi.org/10.1128/JCM.02238-07

Feng, J., Michalik, S., Varming, A. N., Andersen, J. H., Albrecht, D., Jelsbak, L., … Frees, D. (2013). Trapping and proteomic identification of cellular substrates of the ClpP protease in staphylococcus aureus. *Journal of Proteome Research*, *12*(2), 547–558. https://doi.org/10.1021/pr300394r

Feng, M. H., Cui, J. C., Nakane, A., & Hu, D. L. (2013). Vaccination with plasmid DNA encoding a mutant toxic shock syndrome toxin-1 ameliorates toxin-induced lethal shock in mice. *Tohoku Journal of Experimental Medicine*, *231*(1), 1–8. https://doi.org/10.1620/tjem.231.1

Feng, Y., Chen, C. J., Su, L. H., Hu, S., Yu, J., & Chiu, C. H. (2008). Evolution and pathogenesis of Staphylococcus aureus: Lessons learned from genotyping and comparative genomics. *FEMS Microbiology Reviews*, *32*(1), 23–37. https://doi.org/10.1111/j.1574-6976.2007.00086.x

Fioravanti, A., Fumeaux, C., Mohapatra, S. S., Bompard, C., Brilli, M., Frandi, A., … Biondi, E. G. (2013). DNA Binding of the Cell Cycle Transcriptional Regulator GcrA Depends on N6-Adenosine Methylation in Caulobacter crescentus and Other Alphaproteobacteria. *PLoS Genetics*, *9*(5), e1003541. https://doi.org/10.1371/journal.pgen.1003541

Fitzgerald, J. Ross, & Holden, M. T. G. (2016). Genomics of Natural Populations of Staphylococcus aureus. *Annual Review of Microbiology*, *70*(1), 459–478. https://doi.org/10.1146/annurev-micro-102215-095547

Fitzgerald, J R, Sturdevant, D. E., Mackie, S. M., Gill, S. R., & Musser, J. M. (2001). Evolutionary genomics of Staphylococcus aureus: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(15), 8821–8826. https://doi.org/10.1073/pnas.161098098

Flannagan, S. E., & Clewell, D. B. (1991). Conjugative transfer of Tn916 in Enterococcus faecalis: trans Activation of homologous transposons. *Journal of Bacteriology*, *173*(22), 7136–7141. https://doi.org/10.1128/jb.173.22.7136-7141.1991

Flannagan, Susan E., Chow, J. W., Donabedian, S. M., Brown, W. J., Perri, M. B., Zervos, M. J., … Clewell, D. B. (2003). Plasmid Content of a Vancomycin-Resistant Enterococcus faecalis Isolate from a Patient also Colonized by Staphylococcus aureus with a VanA Phenotype. *Antimicrobial Agents and Chemotherapy*, *47*(12), 3954–3959. https://doi.org/10.1128/AAC.47.12.3954-3959.2003

Fleury, B., Kelley, W. L., Lew, D., Götz, F., Proctor, R. A., & Vaudaux, P. (2009). Transcriptomic and metabolic responses of Staphylococcus aureus exposed to supra-physiological temperatures. *BMC Microbiology*, *9*(1), 76. https://doi.org/10.1186/1471-2180-9-76

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., … Turner, S. W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods*, *7*(6), 461–465. https://doi.org/10.1038/nmeth.1459

Fournier, B. (2008). Global Regulators of Staphylococcus aureus Virulence Genes. In J. A. Lindsay (Ed.), *Staphylococcus: molecluar genetics* (pp. 131–184).

Fox, K. L., Dowideit, S. J., Erwin, A. L., Srikhanta, Y. N., Smith, A. L., & Jennings, M. P. (2007). Haemophilus influenzae phasevarions have evolved from type III DNA restriction systems into epigenetic regulators of gene expression. *Nucleic Acids Research*, *35*(15), 5242–5252. https://doi.org/10.1093/nar/gkm571

Fu, Y., Wu, P. H., Beane, T., Zamore, P. D., & Weng, Z. (2018). Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genomics*, *19*(1), 531. https://doi.org/10.1186/s12864-018-4933-1

Fuchs, R. T., Grundy, F. J., & Henkin, T. M. (2006). The SMK box is a new SAM-binding RNA for translational regulation of SAM synthetase. *Nature Structural and Molecular Biology*, *13*(3), 226–233. https://doi.org/10.1038/nsmb1059

Furuta, Y., & Kobayashi, I. (2013). Restriction-Modification Systems as Mobile Epigenetic

Elements. In A. P. Roberts & P. Mullany (Eds.), *Bacterial Integrative Mobile Genetic Elements* (pp. 85–103). Retrieved from http://www.landesbioscience.com/pdf/05Roberts_Kobayashi.pdf

Furuta, Y., Namba-Fukuyo, H., Shibata, T. F., Nishiyama, T., Shigenobu, S., Suzuki, Y., … Kobayashi, I. (2014). Methylome Diversification through Changes in DNA Methyltransferase Sequence Specificity. *PLoS Genetics*, *10*(4), e1004272–e1004272. https://doi.org/10.1371/journal.pgen.1004272

Gao, P., Tang, Q., An, X. M., Yan, X. X., & Liang, D. C. (2011). Structure of HsdS subunit from Thermoanaerobacter tengcongensis sheds lights on mechanism of dynamic opening and closing of type i methyltransferase. *PLoS ONE*, *6*(3), e17346. https://doi.org/10.1371/journal.pone.0017346

Gawthorne, J. A., Beatson, S. A., Srikhanta, Y. N., Fox, K. L., & Jennings, M. P. (2012). Origin of the Diversity in DNA Recognition Domains in Phasevarion Associated modA Genes of Pathogenic Neisseria and Haemophilus influenzae. *PLOS ONE*, *7*(3), e32337. Retrieved from https://doi.org/10.1371/journal.pone.0032337

Ghatak, S., Armstrong, C. M., Reed, S., & He, Y. (2020). Comparative Methylome Analysis of Campylobacter jejuni Strain YH002 Reveals a Putative Novel Motif and Diverse Epigenetic Regulations of Virulence Genes. *Frontiers in Microbiology*, *11*, 610395. https://doi.org/10.3389/fmicb.2020.610395

Gill, S. R., Fouts, D. E., Archer, G. L., Mongodin, E. F., DeBoy, R. T., Ravel, J., … Fraser, C. M. (2005). Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant Staphylococcus aureus strain and a biofilm-producing methicillin-resistant Staphylococcus epidermidis strain. *Journal of Bacteriology*, *187*(7), 2426–2438. https://doi.org/10.1128/JB.187.7.2426-2438.2005

Golding, G. R., Bryden, L., Levett, P. N., McDonald, R. R., Wong, A., Wylie, J., … Mulvey, M. R. (2010). Livestock-associated methicillin-resistant Staphylococcus aureus sequence type 398 in humans, Canada. *Emerging Infectious Diseases*, *16*(4), 587–594. https://doi.org/10.3201/eid1604.091435

Gomes, A. R., Vinga, S., Zavolan, M., & De Lencastre, H. (2005). Analysis of the genetic variability of virulence-related loci in epidemic clones of methicillin-resistant Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy*, *49*(1), 366–379. https://doi.org/10.1128/AAC.49.1.366-379.2005

González-Torres, P., Rodríguez-Mateos, F., Antón, J., & Gabaldón, T. (2019). Impact of homologous recombination on the evolution of prokaryotic core genomes. *MBio*, *10*(1). https://doi.org/10.1128/mBio.02494-18

Gonzalez, D., Kozdon, J. B., McAdams, H. H., Shapiro, L., & Collier, J. (2014). The functions of DNA methylation by CcrM in Caulobacter crescentus: A global approach. *Nucleic Acids Research*, *42*(6), 3720–3735. https://doi.org/10.1093/nar/gkt1352

Gormley, N. A., Watson, M. A., & Halford, S. E. (2005). Bacterial Restriction-Modification Systems. In *Encyclopedia of Life Sciences* (pp. 1–11). https://doi.org/10.1038/npg.els.0003897

Gouil, Q., & Keniry, A. (2019). Latest techniques to study DNA methylation. *Essays in Biochemistry*, *63*(6), 639–648. https://doi.org/10.1042/EBC20190027

Gouy, M., Guindon, S., & Gascuel, O. (2010). Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, *27*(2), 221–224. https://doi.org/10.1093/molbev/msp259

Graham, P. L., Morel, A.-S., Zhou, J., Wu, F., Della-Latta, P., Rubenstein, D., & Saiman, L. (2002). Epidemiology of Methicillin-Susceptible Staphylococcus Aureus in the Neonatal Intensive Care Unit . *Infection Control & Hospital Epidemiology*, *23*(11), 677–682. https://doi.org/10.1086/501993

Grumann, D., Nübel, U., & Bröker, B. M. (2014). Staphylococcus aureus toxins - Their functions and genetics. *Infection, Genetics and Evolution*, *21*(C), 583–592. https://doi.org/10.1016/j.meegid.2013.03.013

Grundmann, H., Aires-de-Sousa, M., Boyce, J., & Tiemersma, E. (2006). Emergence and resurgence of meticillin-resistant Staphylococcus aureus as a public-health threat. *Lancet*, *368*(9538), 874–885. https://doi.org/10.1016/S0140-6736(06)68853-3

Guérillot, R., Kostoulias, X., Donovan, L., Li, L., Carter, G. P., Hachani, A., … Howden, B. P. (2019). Unstable chromosome rearrangements in Staphylococcus aureus cause phenotype switching associated with persistent infections. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(40), 20135–20140. https://doi.org/10.1073/pnas.1904861116

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Halford, S. E. (2001). Hopping, jumping and looping by restriction enzymes. *Biochemical Society Transactions*, *29*(4), 363–373. https://doi.org/10.1042/bst0290363

Han, J. S., Kang, S., Lee, H., Kim, H. K., & Hwang, D. S. (2003). Sequential Binding of SeqA to Paired Hemi-methylated GATC Sequences Mediates Formation of Higher Order Complexes *. *Journal of Biological Chemistry*, *278*(37), 34983–34989. https://doi.org/10.1074/jbc.M304923200

Haniford, D. B., Chelouche, A. R., & Kleckner, N. (1989). A specific class of IS10 transposase mutants are blocked for target site interactions and promote formation of an excised transposon fragment. *Cell*, *59*(2), 385–394. https://doi.org/10.1016/0092-8674(89)90299-7

Harkins, C. P., Pichon, B., Doumith, M., Parkhill, J., Westh, H., Tomasz, A., … Holden, M. T. G. (2017). Methicillin-resistant Staphylococcus aureus emerged long before the introduction of methicillin into clinical practice. *Genome Biology*, *18*(1), 130. https://doi.org/10.1186/s13059-017-1252-9

Harris, S. R., Feil, E. J., Holden, M. T. G., Quail, M. A., Nickerson, E. K., Chantratita, N., … Bentley, S. D. (2010). Evolution of MRSA during hospital transmission and intercontinental spread. *Science*, *327*(5964), 469–474. https://doi.org/10.1126/science.1182395

Hau, S. J., Mou, K. T., Bayles, D. O., & Brockmeier, S. L. (2019). Transcriptomic differences noted in Glaesserella parasuis between growth in broth and on agar. *PLoS ONE*, *14*(8). https://doi.org/10.1371/journal.pone.0220365

Henderson, I. R., & Owen, P. (1999). The major phase-variable outer membrane protein of Escherichia coli structurally resembles the immunoglobulin A1 protease class of exported protein and is regulated by a novel mechanism involving Dam and oxyR. *Journal of Bacteriology*, *181*(7), 2132–2141. https://doi.org/10.1128/JB.181.7.2132-2141.1999

Hennekinne, J. A., De Buyser, M. L., & Dragacci, S. (2012). Staphylococcus aureus and its food poisoning toxins: Characterization and outbreak investigation. *FEMS Microbiology Reviews*, *36*(4), 815–836. https://doi.org/10.1111/j.1574-6976.2011.00311.x

Heusipp, G., Fälker, S., & Alexander Schmidt, M. (2007, February 23). DNA adenine methylation

and bacterial pathogenesis. *International Journal of Medical Microbiology*, Vol. 297, pp. 1–7. https://doi.org/10.1016/j.ijmm.2006.10.002

Hiramatsu, K., Cui, L., Kuroda, M., & Ito, T. (2001). The emergence and evolution of methicillin-resistant Staphylococcus aureus. *Trends in Microbiology*, *9*(10), 486–493. https://doi.org/10.1016/S0966-842X(01)02175-8

Hirst, M., & Marra, M. A. (2010). Next generation sequencing based approaches to epigenomics. *Briefings in Functional Genomics*, *9*(5–6), 455–465. https://doi.org/10.1093/bfgp/elq035

Hochhut, B., Marrero, J., & Waldor, M. K. (2000). Mobilization of plasmids and chromosomal DNA mediated by the SXT element, a constin found in Vibrio cholerae O139. *Journal of Bacteriology*, *182*(7), 2043–2047. https://doi.org/10.1128/JB.182.7.2043-2047.2000

Holden, M. T. G., Hsu, L. Y., Kurt, K., Weinert, L. A., Mather, A. E., Harris, S. R., … Nübel, U. (2013). A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic. *Genome Research*, *23*(4), 653–664. https://doi.org/10.1101/gr.147710.112

Holden, M. T. G., & Lindsay, J. A. (2008). Whole Genomes: Sequence, Microarray and Systems Biology. In J. A. Lindsay (Ed.), *Staphylococcus Molecular Genetics* (1st ed., pp. 1–28). Caister Academic Press.

Holden, M. T. G., Lindsay, J. A., Corton, C., Quail, M. A., Cockfield, J. D., Pathak, S., … Edgeworth, J. D. (2010). Genome sequence of a recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant of methicillin-resistant Staphylococcus aureus, sequence type 239 (TW). *Journal of Bacteriology*, *192*(3), 888–892. https://doi.org/10.1128/JB.01255-09

Horii, T., Suzuki, Y., Monji, A., Morita, M., Muramatsu, H., Kondo, Y., … Maekawa, M. (2003). Detection of mutations in quinolone resistance-determining regions in levofloxacin- and methicillin-resistant Staphylococcus aureus: Effects of the mutations on fluoroquinolone MICs. *Diagnostic Microbiology and Infectious Disease*, *46*(2), 139–145. https://doi.org/10.1016/S0732-8893(03)00037-3

Hsu, L. Y., Harris, S. R., Chlebowicz, M. A., Lindsay, J. A., Koh, T. H., Krishnan, P., … Holden, M. T. G. (2015). Evolutionary dynamics of methicillin-resistant Staphylococcus aureus within a healthcare system. *Genome Biology*, *16*(1), 1–13. https://doi.org/10.1186/s13059-015-0643-z

Hunt, M., Silva, N. De, Otto, T. D., Parkhill, J., Keane, J. A., & Harris, S. R. (2015). Circlator: Automated circularization of genome assemblies using long sequencing reads. *Genome Biology*, *16*(1), 294. https://doi.org/10.1186/s13059-015-0849-0

Hyman, P., & Abedon, S. T. (2010). Bacteriophage Host Range and Bacterial Resistance. *Advances in Applied Microbiology*, Vol. 70, pp. 217–248. https://doi.org/10.1016/S0065-2164(10)70007-1

Ibarra, J. A., Pérez-Rueda, E., Carroll, R. K., & Shaw, L. N. (2013). Global analysis of transcriptional regulators in Staphylococcus aureus. *BMC Genomics*, *14*(1), 126. https://doi.org/10.1186/1471-2164-14-126

Ip, C. L. C., Loose, M., Tyson, J. R., de Cesare, M., Brown, B. L., Jain, M., … Olsen, H. E. (2015). MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Research*, *4*. https://doi.org/10.12688/f1000research.7201.1

Ives, C. L., Nathan, P. D., & Brooks, J. E. (1992). Regulation of the BamHI restriction-modification system by a small intergenic open reading frame, bamHIC, in both Escherichia coli and Bacillus subtilis. *Journal of Bacteriology*, *174*(22), 7194–7201.

https://doi.org/10.1128/jb.174.22.7194-7201.1992

J. Page, A., Taylor, B., & A. Keane, J. (2016). Multilocus sequence typing by blast from de novo assemblies against PubMLST. *The Journal of Open Source Software*, *1*(8), 118. https://doi.org/10.21105/joss.00118

Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*, *17*(1). https://doi.org/10.1186/s13059-016-1103-0

Jamrozy, D. M., Harris, S. R., Mohamed, N., Peacock, S. J., Tan, C. Y., Parkhill, J., … Holden, M. T. G. (2016). Pan-genomic perspective on the evolution of the staphylococcus aureus usa300 epidemic. *Microbial Genomics*, *2*(5). https://doi.org/10.1099/mgen.0.000058

Jeltsch, A. (2003). Maintenance of species identity and controlling speciation of bacteria: A new function for restriction/modification systems? *Gene*, *317*(1–2), 13–16. https://doi.org/10.1016/S0378-1119(03)00652-8

Jiménez-Jacinto, V., Sanchez-Flores, A., & Vega-Alvarado, L. (2019). Integrative Differential Expression Analysis for Multiple EXperiments (IDEAMEX): A Web Server Tool for Integrated RNA-Seq Data Analysis   . *Frontiers in Genetics* , Vol. 10, p. 279. Retrieved from https://www.frontiersin.org/article/10.3389/fgene.2019.00279

Jones, M. J., Donegan, N. P., Mikheyeva, I. V., & Cheung, A. L. (2015). Improving transformation of Staphylococcus aureus belonging to the CC1, CC5 and CC8 clonal complexes. *PLoS ONE*, *10*(3), e0119487-14. https://doi.org/10.1371/journal.pone.0119487

Joseph, S. J., Li, B., Petit, R. A., Qin, Z. S., Darrow, L., & Read, T. D. (2016). The single-species metagenome: Subtyping Staphylococcus aureus core genome sequences from shotgun metagenomic data. *PeerJ*, *2016*(10). https://doi.org/10.7717/peerj.2571

Jurmeister, P., Bockmayr, M., Seegerer, P., Bockmayr, T., Treue, D., Montavon, G., … Klauschen, F. (2019). Machine learning analysis of DNA methylation profiles distinguishes primary lung squamous cell carcinomas from head and neck metastases. *Science Translational Medicine*, *11*(509). https://doi.org/10.1126/scitranslmed.aaw8513

Kahramanoglou, C., Prieto, A. I., Khedkar, S., Haase, B., Gupta, A., Benes, V., … Seshasayee, A. S. N. (2012). Genomics of DNA cytosine methylation in Escherichia coli reveals its role in stationary phase transcription. *Nature Communications*, *3*(1), 886. https://doi.org/10.1038/ncomms1878

Kasarjian, J. K. A., Hidaka, M., Horiuchi, T., Iida, M., & Ryu, J. (2004). The recognition and modification sites for the bacterial type I restriction systems KpnAI, StySEAI, StySENI and StySGI. *Nucleic Acids Research*, *32*(10). https://doi.org/10.1093/nar/gnh079

Kaya, H., Hasman, H., Larsen, J., Stegger, M., Johannesen, T. B., Allesøe, R. L., … Larsen, A. R. (2018). SCCmecFinder, a Web-Based Tool for Typing of Staphylococcal Cassette Chromosome mec in Staphylococcus aureus Using Whole-Genome Sequence Data. *MSphere*, *3*(1). https://doi.org/10.1128/msphere.00612-17

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*(6), 845–858. https://doi.org/10.1038/nprot.2015.053

Khan, H. A., Ahmad, A., & Mehboob, R. (2015). Nosocomial infections and their control strategies. *Asian Pacific Journal of Tropical Biomedicine*, *5*(7), 509–514. https://doi.org/10.1016/j.apjtb.2015.05.001

Kim, J. S., DeGiovanni, A., Jancarik, J., Adams, P. D., Yokota, H., Kim, R., & Kim, S. H. (2005). Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(9), 3248–3253. https://doi.org/10.1073/pnas.0409851102

Kinnevey, P. M., Shore, A. C., Brennan, G. I., Sullivan, D. J., Ehricht, R., Monecke, S., … Coleman, D. C. (2013). Emergence of sequence type 779 methicillin-resistant Staphylococcus aureus harboring a novel pseudo staphylococcal cassette chromosome mec (SCCmec)-SCC-SCCCRISPR composite element in Irish Hospitals. *Antimicrobial Agents and Chemotherapy*, *57*(1), 524–531. https://doi.org/10.1128/AAC.01689-12

Kläui, A. J., Boss, R., & Graber, H. U. (2019). Characterization and comparative analysis of the staphylococcus aureus Genomic Island vSaβ: An in silico approach. *Journal of Bacteriology*, *201*(22). https://doi.org/10.1128/JB.00777-18

Klein, E., Smith, D. L., & Laxminarayan, R. (2007). Hospitalizations and deaths caused by methicillin-resistant Staphylococcus aureus, United States, 1999-2005. *Emerging Infectious Diseases*, *13*(12), 1840–1846. https://doi.org/10.3201/eid1312.070629

Klein, E., Smith, D. L., & Laxminarayan, R. (2009). Community-associated methicillin-resistant Staphylococcus aureus in outpatients, United States, 1999-2006. *Emerging Infectious Diseases*, *15*(12), 1925–1930. https://doi.org/10.3201/eid1512.081341

Klumpp, S., & Hwa, T. (2014). Bacterial growth: Global effects on gene expression, growth feedback and proteome partition. *Current Opinion in Biotechnology*, Vol. 28, pp. 96–102. https://doi.org/10.1016/j.copbio.2014.01.001

Klumpp, S., Zhang, Z., & Hwa, T. (2009). Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell*, *139*(7), 1366–1375. https://doi.org/10.1016/j.cell.2009.12.001

Kobayashi, I. (2001). Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Research*, *29*(18), 3742–3756. https://doi.org/10.1093/nar/29.18.3742

Kobayashi, I. (2004). Restriction-Modification Systems as Minimal Forms of Life. In A. P. (Ed.) (Ed.), *Nucleic Acids and Molecular Biology* (Vol. 14, pp. 19–62). https://doi.org/10.1007/978-3-642-18851-0_2

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/gr.215087.116

Korlach, J., & Turner, S. W. (2012, June). Going beyond five bases in DNA sequencing. *Current Opinion in Structural Biology*, Vol. 22, pp. 251–261. https://doi.org/10.1016/j.sbi.2012.04.002

Kumar, R., & Rao, D. N. (2013). Role of DNA methyltransferases in epigenetic regulation in bacteria. In *Sub-Cellular Biochemistry* (Vol. 61, pp. 81–102). https://doi.org/10.1007/978-94-007-4525-4_4

Kurdyukov, S., & Bullock, M. (2016). DNA methylation analysis: Choosing the right method. *Biology*, *5*(1), 3–21. https://doi.org/10.3390/biology5010003

Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., … Hiramatsu, K. (2001). Whole genome sequencing of meticillin-resistant Staphylococcus aureus. *Lancet*, *357*(9264), 1225–1240. https://doi.org/10.1016/S0140-6736(00)04403-2

Laabei, M., Recker, M., Rudkin, J. K., Aldeljawi, M., Gulay, Z., Sloan, T. J., … Massey, R. C. (2014). Predicting the virulence of MRSA from its genome sequence. *Genome Research*, *24*(5), 839–

849. https://doi.org/10.1101/gr.165415.113

Laabei, M., Uhlemann, A. C., Lowy, F. D., Austin, E. D., Yokoyama, M., Ouadi, K., … Massey, R. C. (2015). Evolutionary Trade-Offs Underlie the Multi-faceted Virulence of Staphylococcus aureus. *PLoS Biology*, *13*(9), e1002229. https://doi.org/10.1371/journal.pbio.1002229

Lacks, S., & Greenberg, B. (1977). Complementary specificity of restriction endonucleases of Diplococcus pneumoniae with respect to DNA methylation. *Journal of Molecular Biology*, *114*(1), 153–168. https://doi.org/10.1016/0022-2836(77)90289-3

Laird, P. W. (2010, March 2). Principles and challenges of genome-wide DNA methylation analysis. *Nature Reviews Genetics*, Vol. 11, pp. 191–203. https://doi.org/10.1038/nrg2732

Larsen, J., Andersen, P. S., Winstel, V., & Peschel, A. (2017). Staphylococcus aureus CC395 harbours a novel composite staphylococcal cassette chromosome mec element. *Journal of Antimicrobial Chemotherapy*, *72*(4), 1002–1005. https://doi.org/10.1093/jac/dkw544

Lasa, I., Toledo-Arana, A., Dobin, A., Villanueva, M., De Los Mozos, I. R., Vergara-Irigaray, M., … Gingeras, T. R. (2011). Genome-wide antisense transcription drives mRNA processing in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 20172–20177. https://doi.org/10.1073/pnas.1113521108

Laszlo, A. H., Derrington, I. M., Brinkerhoff, H., Langford, K. W., Nova, I. C., Samson, J. M., … Gundlach, J. H. (2013). Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(47), 18904–18909. https://doi.org/10.1073/pnas.1310240110

Laszlo, A. H., Derrington, I. M., Ross, B. C., Brinkerhoff, H., Adey, A., Nova, I. C., … Gundlach, J. H. (2014). Decoding long nanopore sequencing reads of natural DNA. *Nature Biotechnology*, *32*(8), 829–833. https://doi.org/10.1038/nbt.2950

Laub, M. T., Chen, S. L., Shapiro, L., & McAdams, H. H. (2002). Genes directly controlled by CtrA, a master regulator of the Caulobacter cell cycle. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(7), 4632–4637. https://doi.org/10.1073/pnas.062065699

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*(2), R29. https://doi.org/10.1186/gb-2014-15-2-r29

Le, N. Q. K. (2019). iN6-methylat (5-step): identifying DNA N 6-methyladenine sites in rice genome using continuous bag of nucleobases via Chou's 5-step rule. *Molecular Genetics and Genomics*, *294*(5), 1173–1182. https://doi.org/10.1007/s00438-019-01570-y

Lebedev, I. N., & Sazhenova, E. A. (2008). Epimutations of imprinting genes in the human genome: classification, causes, association with hereditary pathology. *Genetika*, *44*(10), 1356–1373. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19062532

Lee, J., Shin, M. K., Ryu, D. K., Kim, S., & Ryu, W. S. (2010). Insertion and deletion mutagenesis by overlap extension PCR. In *Methods in Molecular Biology* (Vol. 634, pp. 137–146). https://doi.org/10.1007/978-1-60761-652-8_10

Lees, J. A., Kremer, P. H. C., Manso, A. S., Croucher, N. J., Ferwerda, B., Serón, M. V., … Bentley, S. D. (2017). Large scale genomic analysis shows no evidence for pathogen adaptation between the blood and cerebrospinal fluid niches during bacterial meningitis. *Microbial Genomics*, *3*(1), mgen000103. https://doi.org/10.1099/mgen.0.000103

Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, *44*(W1), W242–W245. https://doi.org/10.1093/nar/gkw290

Letunic, I., & Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, *47*(W1), W256–W259. https://doi.org/10.1093/nar/gkz239

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, Jing, Li, J. W., Feng, Z., Wang, J., An, H., Liu, Y., … Zhang, J. R. (2016). Epigenetic Switch Driven by DNA Inversions Dictates Phase Variation in Streptococcus pneumoniae. *PLoS Pathogens*, *12*(7). https://doi.org/10.1371/journal.ppat.1005762

Li, Jun, Yan, H., Wang, K., Tan, W., & Zhou, X. (2007). Hairpin fluorescence DNA probe for real-time monitoring of DNA methylation. *Analytical Chemistry*, *79*(3), 1050–1056. https://doi.org/10.1021/ac061694i

Li, X. Z., & Nikaido, H. (2009). Efflux-mediated drug resistance in bacteria: An update. *Drugs*, *69*(12), 1555–1623. https://doi.org/10.2165/11317030-000000000-00000

Lin, D., Ou, Q., Lin, J., Peng, Y., & Yao, Z. (2017). A meta-analysis of the rates of Staphylococcus aureus and methicillin-resistant S aureus contamination on the surfaces of environmental objects that health care workers frequently touch. *American Journal of Infection Control*, *45*(4), 421–429. https://doi.org/10.1016/j.ajic.2016.11.004

Lindsay, J. A. (Ed.). (2008). *Staphylococcus: molecular genetics* (1st ed.). Norfolk: Caister Academic Press.

Lindsay, J. A. (2010). International Journal of Medical Microbiology Genomic variation and evolution of Staphylococcus aureus. *International Jl of Medical Microbiology*, *300*(2–3), 98–103. Retrieved from http://dx.doi.org/10.1016/j.ijmm.2009.08.013

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

Lindsay, J. A., Moore, C. E., Day, N. P., Peacock, S. J., Witney, A. A., Stabler, R. A., … Hinds, J. (2006). Microarrays reveal that each of the ten dominant lineages of Staphylococcus aureus has a unique combination of surface-associated and regulatory genes. *Journal of Bacteriology*, *188*(2), 669–676. https://doi.org/10.1128/JB.188.2.669-676.2006

Liu, L., Zhang, Y., Jiang, D., Du, S., Deng, Z., Wang, L., & Chen, S. (2019). Recent Advances in the Genomic Profiling of Bacterial Epigenetic Modifications. *Biotechnology Journal*, *14*(1). https://doi.org/10.1002/biot.201800001

Liu, Y. P., Tang, Q., Zhang, J. Z., Tian, L. F., Gao, P., & Yan, X. X. (2017). Structural basis underlying complex assembly and conformational transition of the type I R-M system. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(42), 11151–11156. https://doi.org/10.1073/pnas.1711754114

Lluch-Senar, M., Luong, K., Lloréns-Rico, V., Delgado, J., Fang, G., Spittle, K., … Serrano, L. (2013). Comprehensive Methylome Characterization of Mycoplasma genitalium and Mycoplasma pneumoniae at Single-Base Resolution. *PLoS Genetics*, *9*(1), 1003191. https://doi.org/10.1371/journal.pgen.1003191

Løbner-Olesen, A., Skovgaard, O., & Marinus, M. G. (2005, April 1). Dam methylation: Coordinating cellular processes. *Current Opinion in Microbiology*, Vol. 8, pp. 154–160. https://doi.org/10.1016/j.mib.2005.02.009

Loenen, W. A.M., Daniel, A. S., Braymer, H. D., & Murray, N. E. (1987). Organization and sequence of the hsd genes of Escherichia coli K-12. *Journal of Molecular Biology*, *198*(2), 159–170. https://doi.org/10.1016/0022-2836(87)90303-2

Loenen, Wil A.M. (2003). Tracking EcoKI and DNA fifty years on: A golden story full of surprises. *Nucleic Acids Research*, *31*(24), 7059–7069. https://doi.org/10.1093/nar/gkg944

Loenen, Wil A.M., Dryden, D. T. F., Raleigh, E. A., & Wilson, G. G. (2014). Type i restriction enzymes and their relatives. *Nucleic Acids Research*, *42*(1), 20–44. https://doi.org/10.1093/nar/gkt847

Loenen, Wil A.M., & Raleigh, E. A. (2014). The other face of restriction: Modification-dependent enzymes. *Nucleic Acids Research*, *42*(1), 56–69. https://doi.org/10.1093/nar/gkt747

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Low, D. A., Weyand, N. J., & Mahan, M. J. (2001). Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infection and Immunity*, *69*(12), 7197–7204. https://doi.org/10.1128/IAI.69.12.7197-7204.2001

Low, David A., & Casadesús, J. (2008, April). Clocks and switches: bacterial gene regulation by DNA adenine methylation. *Current Opinion in Microbiology*, Vol. 11, pp. 106–112. https://doi.org/10.1016/j.mib.2008.02.012

Lowy, F. D. (2003). Antimicrobial resistance: The example of Staphylococcus aureus. *Journal of Clinical Investigation*, *111*(9), 1265–1273. https://doi.org/10.1172/JCI18535

Lozano, C., Gharsa, H., Ben Slama, K., Zarazaga, M., & Torres, C. (2016). Staphylococcus aureus in Animals and Food: Methicillin Resistance, Prevalence and Population Structure. A Review in the African Continent. *Microorganisms*, *4*(1), 12. https://doi.org/10.3390/microorganisms4010012

Lu, S. C. (2000). S-Adenosylmethionine. *International Journal of Biochemistry and Cell Biology*, *32*(4), 391–395. https://doi.org/10.1016/S1357-2725(99)00139-9

Luo, G. Z., Blanco, M. A., Greer, E. L., He, C., & Shi, Y. (2015). DNA N6-methyladenine: A new epigenetic mark in eukaryotes? *Nature Reviews Molecular Cell Biology*, *16*(12), 705–710. https://doi.org/10.1038/nrm4076

Lv, H., Dao, F. Y., Guan, Z. X., Zhang, D., Tan, J. X., Zhang, Y., … Lin, H. (2019). iDNA6mA-Rice: A Computational Tool for Detecting N6-Methyladenine Sites in Rice. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.00793

Lyon, B. R., Gillespie, M. T., & Skurray, R. A. (1987). Detection and characterization of IS256, and insertion sequence in Staphylococcus aureus. *Journal of General Microbiology*, *133*(11), 3031–3038. https://doi.org/10.1099/00221287-133-11-3031

Makarova, K. S., & Koonin, E. V. (2015). Annotation and classification of CRISPR-Cas systems. *Methods in Molecular Biology*, *1311*, 47–75. https://doi.org/10.1007/978-1-4939-2687-9_4

Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2013). Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Research*, *41*(8), 4360–4377.

https://doi.org/10.1093/nar/gkt157

Malachowa, N., & Deleo, F. R. (2010). Mobile genetic elements of Staphylococcus aureus. *Cellular and Molecular Life Sciences*, *67*(18), 3057–3071. https://doi.org/10.1007/s00018-010-0389-4

Manga, P., Klingeman, D. M., Lu, T.-Y. S., Mehlhorn, T. L., Pelletier, D. A., Hauser, L. J., … Brown, S. D. (2016). Replicates, Read Numbers, and Other Important Experimental Design Considerations for Microbial RNA-seq Identified Using Bacillus thuringiensis Datasets. *Frontiers in Microbiology*, *7*(MAY), 794. https://doi.org/10.3389/fmicb.2016.00794

Manrao, E. A., Derrington, I. M., Laszlo, A. H., Langford, K. W., Hopper, M. K., Gillgren, N., … Gundlach, J. H. (2012). Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nature Biotechnology*, *30*(4), 349–353. https://doi.org/10.1038/nbt.2171

Manrao, E. A., Derrington, I. M., Pavlenok, M., Niederweis, M., & Gundlach, J. H. (2011). Nucleotide discrimination with DNA immobilized in the MSPA nanopore. *PLoS ONE*, *6*(10). https://doi.org/10.1371/journal.pone.0025723

Manso, A. S., Chai, M. H., Atack, J. M., Furi, L., De Ste Croix, M., Haigh, R., … Oggioni, M. R. (2014). A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nature Communications*, *5*, 5055. https://doi.org/10.1038/ncomms6055

Marinus, M. G. (1985). DNA methylation influences trpR promoter activity in Escherichia coli K-12. *Molecular and General Genetics MGG*, *200*(1), 185–186. https://doi.org/10.1007/BF00383334

Martin, J. L., & McMillan, F. M. (2002, December 1). SAM (dependent) I AM: The S-adenosylmethionine-dependent methyltransferase fold. *Current Opinion in Structural Biology*, Vol. 12, pp. 783–793. https://doi.org/10.1016/S0959-440X(02)00391-3

McCarthy, A. J., & Lindsay, J. A. (2012). The distribution of plasmids that carry virulence and resistance genes in Staphylococcus aureus is lineage associated. *BMC Microbiology*, *12*(1), 104. https://doi.org/10.1186/1471-2180-12-104

McCarthy, A. J., & Lindsay, J. A. (2013). Staphylococcus aureus innate immune evasion is lineage-specific: A bioinfomatics study. *Infection, Genetics and Evolution*, *19*(C), 7–14. https://doi.org/10.1016/j.meegid.2013.06.012

McCarthy, A. J., Witney, A. A., Gould, K. A., Moodley, A., Guardabassi, L., Voss, A., … Lindsay, J. A. (2011). The distribution of mobile genetic elements (MGEs) in MRSA CC398 Is associated with both host and country. *Genome Biology and Evolution*, *3*(1), 1164–1174. https://doi.org/10.1093/gbe/evr092

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, *40*(10), 4288–4297. https://doi.org/10.1093/nar/gks042

McIntyre, A. B. R., Alexander, N., Grigorev, K., Bezdan, D., Sichtig, H., Chiu, C. Y., & Mason, C. E. (2019). Single-molecule sequencing detection of N6-methyladenine in microbial reference materials. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-08289-9

Mediavilla, J. R., Chen, L., Mathema, B., & Kreiswirth, B. N. (2012). Global epidemiology of community-associated methicillin resistant Staphylococcus aureus (CA-MRSA). *Current Opinion in Microbiology*, *15*(5), 588–595. https://doi.org/10.1016/j.mib.2012.08.003

Mehershahi, K. S., & Chen, S. L. (2021). Methylation by multiple type I restriction modification systems avoids influencing gene 1 regulation in uropathogenic Escherichia coli 2 3 4. *BioRxiv*, 2021.01.08.425850. https://doi.org/10.1101/2021.01.08.425850

Miller, L. S., & Cho, J. S. (2011). Immunity against Staphylococcus aureus cutaneous infections. *Nature Reviews Immunology*, *11*(8), 505–518. https://doi.org/10.1038/nri3010

Młynarczyk, A., Młynarczyk, G., & Jeljaszewicz, J. (1998). The genome of Staphylococcus aureus: A review. *Zentralblatt Fur Bakteriologie*, *287*(4), 277–314. https://doi.org/10.1016/S0934-8840(98)80165-5

Molina, F., & Skarstad, K. (2004). Replication fork and SeqA focus distributions in Escherichia coli suggest a replication hyperstructure dependent on nucleotide metabolism. *Molecular Microbiology*, *52*(6), 1597–1612. https://doi.org/https://doi.org/10.1111/j.1365-2958.2004.04097.x

Monk, I. R., & Foster, T. J. (2012). Genetic manipulation of Staphylococci-breaking through the barrier. *Frontiers in Cellular and Infection Microbiology*, *2*(49), 49. https://doi.org/10.3389/fcimb.2012.00049

Monk, I. R., Shah, I. M., Xu, M., Tan, M. W., & Foster, T. J. (2012). Transforming the untransformable: Application of direct transformation to manipulate genetically Staphylococcus aureus and Staphylococcus epidermidis. *MBio*, *3*(2). https://doi.org/10.1128/mBio.00277-11

Monk, I. R., Tree, J. J., Howden, B. P., Stinear, T. P., & Foster, T. J. (2015). Complete bypass of restriction systems for major staphylococcus aureus lineages. *MBio*, *6*(3), 1–12. https://doi.org/10.1128/mBio.00308-15

Morikawa, K., Takemura, A. J., Inose, Y., Tsai, M., Nguyen Thi, L. T., Ohta, T., & Msadek, T. (2012). Expression of a Cryptic Secondary Sigma Factor Gene Unveils Natural Competence for DNA Transformation in Staphylococcus aureus. *PLoS Pathogens*, *8*(11), e1003003. https://doi.org/10.1371/journal.ppat.1003003

Mostowy, R., Croucher, N. J., Hanage, W. P., Harris, S. R., Bentley, S., & Fraser, C. (2014). Heterogeneity in the Frequency and Characteristics of Homologous Recombination in Pneumococcal Evolution. *PLoS Genetics*, *10*(5), e1004300. https://doi.org/10.1371/journal.pgen.1004300

Münch, R., Hiller, K., Grote, A., Scheer, M., Klein, J., Schobert, M., & Jahn, D. (2005). Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. *Bioinformatics (Oxford, England)*, *21*(22), 4187–4189. https://doi.org/10.1093/bioinformatics/bti635

Murray, Noreen E., & Murray, N. E. (2000). Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and Weigle). *Microbiology and Molecular Biology Reviews*, *64*(2), 1–23. https://doi.org/10.1128/mmbr.64.2.412-434.2000

Murray, I. A., Clark, T. A., Morgan, R. D., Boitano, M., Anton, B. P., Luong, K., … Roberts, R. J. (2012). The methylomes of six bacteria. *Nucleic Acids Research*, *40*(22), 11450–11462. https://doi.org/10.1093/nar/gks891

Murray, S. M., Panis, G., Fumeaux, C., Viollier, P. H., & Howard, M. (2013). Computational and Genetic Reduction of a Cell Cycle to Its Simplest, Primordial Components. *PLoS Biology*, *11*(12). https://doi.org/10.1371/journal.pbio.1001749

Nabera, C. K. (2009). Staphylococcus aureus bacteremia: Epidemiology, pathophysiology, and management strategies. *Clinical Infectious Diseases*, *48*(SUPPL. 4), S231–S237. https://doi.org/10.1086/598189

Ni, P., Huang, N., Zhang, Z., Wang, D. P., Liang, F., Miao, Y., … Wang, J. (2019). DeepSignal: Detecting DNA methylation state from Nanopore sequencing reads using deep-learning.

Bioinformatics, *35*(22), 4586–4595. https://doi.org/10.1093/bioinformatics/btz276

Nienaber, J. J. C., Sharma Kuinkel, B. K., Clarke-Pearson, M., Lamlertthon, S., Park, L., Rude, T. H., … Fowler, V. G. (2011). Methicillin-susceptible staphylococcus aureus endocarditis isolates Are associated with clonal complex 30 genotype and a distinct repertoire of enterotoxins and adhesins. *Journal of Infectious Diseases*, *204*(5), 704–713. https://doi.org/10.1093/infdis/jir389

Nigam, A., Gupta, D., & Sharma, A. (2014). Treatment of infectious disease: Beyond antibiotics. *Microbiological Research*, *169*(9–10), 643–651. https://doi.org/10.1016/j.micres.2014.02.009

Nimmo, G. R., Steen, J. A., Monecke, S., Ehricht, R., Slickers, P., Thomas, J. C., … Coombs, G. W. (2015). ST2249-MRSA-III: A second major recombinant methicillin-resistant Staphylococcus aureus clone causing healthcare infection in the 1970s. *Clinical Microbiology and Infection*, *21*(5), 444–450. https://doi.org/10.1016/j.cmi.2014.12.018

Nonaka, L., Yamamoto, T., Maruyama, F., Hirose, Y., Onishi, Y., Kobayashi, T., … Yano, H. (2018). Interplay of a non-conjugative integrative element and a conjugative plasmid in the spread of antibiotic resistance via suicidal plasmid transfer from an aquaculture Vibrio isolate. *PLoS ONE*, *13*(6), e0198613. https://doi.org/10.1371/journal.pone.0198613

Novick, R. P. (2003). Mobile genetic elements and bacterial toxinoses: The superantigen-encoding pathogenicity islands of Staphylococcus aureus. *Plasmid*, *49*(2), 93–105. https://doi.org/10.1016/S0147-619X(02)00157-9

Novick, R. P., & Ram, G. (2016). The Floating (Pathogenicity) Island: A Genomic Dessert. *Trends in Genetics*, *32*(2), 114–126. https://doi.org/10.1016/j.tig.2015.11.005

Novick, R. P., & Ram, G. (2017, August 1). Staphylococcal pathogenicity islands — movers and shakers in the genomic firmament. *Current Opinion in Microbiology*, Vol. 38, pp. 197–204. https://doi.org/10.1016/j.mib.2017.08.001

Nye, T. M., Jacob, K. M., Holleyid, E. K., Nevarez, J. M., Dawidid, S., Simmons, L. A., … Watson, M. E. (2019). DNA methylation from a type I restriction modification system influences gene expression and virulence in streptococcus pyogenes. *PLoS Pathogens*, *15*(6). https://doi.org/10.1371/journal.ppat.1007841

Nye, T. M., van Gijtenbeek, L. A., Stevens, A. G., Schroeder, J. W., Randall, J. R., Matthews, L. A., & Simmons, L. A. (2020). Methyltransferase DnmA is responsible for genome-wide N6-methyladenosine modifications at non-palindromic recognition sites in Bacillus subtilis. *Nucleic Acids Research*, *48*(10), 5332–5348. https://doi.org/10.1093/nar/gkaa266

O'Connor, C. D., & Humphreys, G. O. (1982). Expression of the EcoRI restriction-modification system and the construction of positive-selection cloning vectors. *Gene*, *20*(2), 219–229. https://doi.org/10.1016/0378-1119(82)90041-5

O'sullivan, D., Twomey, D. P., Coffey, A., Hill, C., Fitzgerald, G. F., & Ross, R. P. (2000). Navel type I restriction specificities through domain shuffling of HsdS subunits in Lactococcus lactis. *Molecular Microbiology*, *36*(4), 866–875. https://doi.org/10.1046/j.1365-2958.2000.01901.x

Obarska, A., Blundell, A., Feder, M., Vejsadová, Š., Šišáková, E., Weiserová, M., … Firman, K. (2006). Structural model for the multisubunit Type IC restriction-modification DNA methyltransferase M.EcoR124I in complex with DNA. *Nucleic Acids Research*, *34*(7), 1992–2005. https://doi.org/10.1093/nar/gkl132

Oliveira, P. H., & Fang, G. (2021, January 1). Conserved DNA Methyltransferases: A Window into Fundamental Mechanisms of Epigenetic Regulation in Bacteria. *Trends in Microbiology*, Vol. 29, pp. 28–40. https://doi.org/10.1016/j.tim.2020.04.007

Oliveira, P. H., Touchon, M., & Rocha, E. P. C. (2014). The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Research*, *42*(16), 10618–10631. https://doi.org/10.1093/nar/gku734

Oliveira, P. H., Touchon, M., & Rocha, E. P. C. (2016). Regulation of genetic flux between bacteria by restriction-modification systems. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(20), 5658–5663. https://doi.org/10.1073/pnas.1603257113

Orzechowska, B., & Mohammed, M. (2019). The War between Bacteria and Bacteriophages. In *Growing and Handling of Bacterial Cultures.* https://doi.org/10.5772/intechopen.87247

Osmundson, J., Dewell, S., & Darst, S. A. (2013). RNA-Seq Reveals Differential Gene Expression in Staphylococcus aureus with Single-Nucleotide Resolution. *PLoS ONE*, *8*(10), e76572. https://doi.org/10.1371/journal.pone.0076572

Osmundson, J., Montero-Diez, C., Westblade, L. F., Hochschild, A., & Darst, S. A. (2012). Promoter-specific transcription inhibition in staphylococcus aureus by a phage protein. *Cell*, *151*(5), 1005–1016. https://doi.org/10.1016/j.cell.2012.10.034

Otter, J. A., & French, G. L. (2008). The emergence of community-associated methicillin-resistant Staphylococcus aureus at a London teaching hospital, 2000-2006. *Clinical Microbiology and Infection*, *14*(7), 670–676. https://doi.org/10.1111/j.1469-0691.2008.02017.x

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

Pané-Farré, J., Jonas, B., Förstner, K., Engelmann, S., & Hecker, M. (2006). The σB regulon in Staphylococcus aureus and its regulation. *International Journal of Medical Microbiology*, *296*(4), 237–258. https://doi.org/https://doi.org/10.1016/j.ijmm.2005.11.011

Papanek, B., O'Dell, K. B., Manga, P., Giannone, R. J., Klingeman, D. M., Hettich, R. L., … Guss, A. M. (2018). Transcriptomic and proteomic changes from medium supplementation and strain evolution in high-yielding Clostridium thermocellum strains. *Journal of Industrial Microbiology & Biotechnology*, *45*(11), 1007–1015. https://doi.org/10.1007/s10295-018-2073-x

Papiez, A., Marczyk, M., Polanska, J., & Polanski, A. (2019). BatchI: Batch effect Identification in high-Throughput screening data using a dynamic programming algorithm. *Bioinformatics*, *35*(11), 1885–1892. https://doi.org/10.1093/bioinformatics/bty900

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., & Hellmann, I. (2016). The impact of amplification on differential expression analyses by RNA-seq. *Scientific Reports*, *6*(1), 1–11. https://doi.org/10.1038/srep25533

Parveen, N., & Cornell, K. A. (2011, January). Methylthioadenosine/S-adenosylhomocysteine nucleosidase, a critical enzyme for bacterial metabolism. *Molecular Microbiology*, Vol. 79, pp. 7–20. https://doi.org/10.1111/j.1365-2958.2010.07455.x

Patel, D., Ellington, M. J., Hope, R., Reynolds, R., Arnold, C., & Desai, M. (2015). Identification of genetic variation exclusive to specific lineages associated with Staphylococcus aureus bacteraemia. *Journal of Hospital Infection*, *91*(2), 136–145. https://doi.org/10.1016/j.jhin.2015.07.003

Peterson, K. R., Wertman, K. F., Mount, D. W., & Marinus, M. G. (1985). Viability of Escherichia coli K-12 DNA adenine methylase (dam) mutants requires increased expression of specific genes in the SOS regulon. *MGG Molecular & General Genetics*, *201*(1), 14–19. https://doi.org/10.1007/BF00397979

Pfeifer, G. P. (1994). DNA Methylation: Molecular Biology and Biological Significance. J. P. Jost , H. P. Saluz . In *The Quarterly Review of Biology* (Vol. 69). https://doi.org/10.1086/418565

Phillips, Z. N., Husna, A. U., Jennings, M. P., Seib, K. L., & Atack, J. M. (2019). Phasevarions of bacterial pathogens-phase-variable epigenetic regulators evolving from restriction-modification systems. *Microbiology (United Kingdom)*, *165*(9), 917–928. https://doi.org/10.1099/mic.0.000805

Pingoud, A., & Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Research*, *29*(18), 3705–3727. https://doi.org/10.1093/nar/29.18.3705

Plongthongkum, N., Diep, D. H., & Zhang, K. (2014, October 11). Advances in the profiling of DNA modifications: Cytosine methylation and beyond. *Nature Reviews Genetics*, Vol. 15, pp. 647–661. https://doi.org/10.1038/nrg3772

Polaczek, P., Kwan, K., & Campbell, J. L. (1998). GATC motifs may alter the conformation of DNA depending on sequence context and N6-adenine methylation status: Possible implications for DNA-protein recognition. *Molecular and General Genetics*, *258*(5), 488–493. https://doi.org/10.1007/s004380050759

Posnick, L. M., & Samson, L. D. (1999). Influence of S-adenosylmethionine pool size on spontaneous mutation, dam methylation, and cell growth of Escherichia coli. *Journal of Bacteriology*, *181*(21), 6756–6762. https://doi.org/10.1128/jb.181.21.6756-6762.1999

Poudel, S., Tsunemoto, H., Seif, Y., Sastry, A. V., Szubin, R., Xu, S., … Palsson, B. O. (2020). Revealing 29 sets of independently modulated genes in Staphylococcus aureus, their regulators, and role in key physiological response. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(29), 17228–17239. https://doi.org/10.1073/pnas.2008413117

Prax, M., Lee, C. Y., & Bertram, R. (2013). An update on the molecular genetics toolbox for staphylococci. *Microbiology (United Kingdom)*, *159*(PART3), 421–435. https://doi.org/10.1099/mic.0.061705-0

Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, *26*(7), 1641–1650. https://doi.org/10.1093/molbev/msp077

Prosperi, M., Veras, N., Azarian, T., Rathore, M., Nolan, D., Rand, K., … Salemi, M. (2013). Molecular epidemiology of community-associated Methicillin-resistant Staphylococcus aureus in the genomic era: A cross-sectional study. *Scientific Reports*, *3*, 1–8. https://doi.org/10.1038/srep01902

Pruitt, K. D., Tatusova, T., Brown, G. R., & Maglott, D. R. (2012). NCBI Reference Sequences (RefSeq): Current status, new features and genome annotation policy. *Nucleic Acids Research*, *40*(D1), D130–D135. https://doi.org/10.1093/nar/gkr1079

Ramsay, J. P., Kwong, S. M., Murphy, R. J. T., Yui Eto, K., Price, K. J., Nguyen, Q. T., … Firth, N. (2016). An updated view of plasmid conjugation and mobilization in Staphylococcus . *Mobile Genetic Elements*, *6*(4), e1208317. https://doi.org/10.1080/2159256x.2016.1208317

Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M., & Paten, B. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nature Methods*, *14*(4), 411–413. https://doi.org/10.1038/nmeth.4189

Rao, B. S., & Buckler-White, A. (1998). Direct visualization of site-specific and strand-specific DNA methylation patterns in automated DNA sequencing data. In *Nucleic Acids Research* (Vol. 26). https://doi.org/10.1093/nar/26.10.2505

Rao, D. N., Dryden, D. T. F., & Bheemanaik, S. (2014). Type III restriction-modification enzymes: A historical perspective. *Nucleic Acids Research*, *42*(1), 45–55. https://doi.org/10.1093/nar/gkt616

Rao, D. N., Saha, S., & Krishnamurthy, V. (2000). ATP-dependent restriction enzymes. *Progress in Nucleic Acid Research and Molecular Biology*, *64*, 1–63. https://doi.org/10.1016/s0079-6603(00)64001-1

Razin, A, Cedar, H., & Riggs, A. (2012). DNA Methylation - biochemistry and biological significance. In *Springer Science* (Vol. 4). Retrieved from https://books.google.co.uk/books?id=c6zeBwAAQBAJ&pg=PA75&lpg=PA75&dq=hsds+structure&source=bl&ots=lAmQn_EG1h&sig=oeDWiBhkZ1RbRQlSM66p6ZrEbJs&hl=en&sa=X&ved=0ahUKEwi-zYOihbzaAhXlAsAKHRMbD_s4ChDoAQgrMAE#v=onepage&q=hsds structure&f=false

Razin, Aharon, & Riggs, A. D. (1980). DNA Methylation and gene function. *Science*, *210*(4470), 604–610. https://doi.org/10.1126/science.6254144

Resch, A., Rosenstein, R., Nerz, C., & Götz, F. (2005). Differential gene expression profiling of Staphylococcus aureus cultivated under biofilm and planktonic conditions. *Applied and Environmental Microbiology*, *71*(5), 2663–2676. https://doi.org/10.1128/AEM.71.5.2663-2676.2005

Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics, Proteomics and Bioinformatics*, *13*(5), 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Richards, M. J., Edwards, J. R., Culver, D. H., & Gaynes, R. P. (1999). Nosocomial infections in medical intensive care units in the United States. *Critical Care Medicine*, *27*(5), 887–892. https://doi.org/10.1097/00003246-199905000-00020

Risso, D., Schwartz, K., Sherlock, G., & Dudoit, S. (2011). GC-Content Normalization for RNA-Seq Data. *BMC Bioinformatics*, *12*(1), 480. https://doi.org/10.1186/1471-2105-12-480

Roberts, D., Hoopes, B. C., McClure, W. R., & Kleckner, N. (1985). IS10 transposition is regulated by DNA adenine methylation. *Cell*, *43*(1), 117–130. https://doi.org/10.1016/0092-8674(85)90017-0

Roberts, G. A., Chen, K., Cooper, L. P., White, J. H., Blakely, G. W., & Dryden, D. T. F. (2012). Removal of a frameshift between the hsdM and hsdS genes of the EcoKI Type IA DNA restriction and modification system produces a new type of system and links the different families of Type i systems. *Nucleic Acids Research*, *40*(21), 10916–10924. https://doi.org/10.1093/nar/gks876

Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., … Xu, S. Y. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Research*, *31*(7), 1805–1812. https://doi.org/10.1093/nar/gkg274

Roberts, R. J., Carneiro, M. O., & Schatz, M. C. (2017). Erratum to: The advantages of SMRT sequencing. *Genome Biology*, *18*(1), 1–4. https://doi.org/10.1186/s13059-017-1295-y

Roberts, R. J., Vincze, T., Posfai, J., & Macelis, D. (2015). REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Research*, *43*(D1), D298–D299. https://doi.org/10.1093/nar/gku1046

Robertson, K. D. (2005, August). DNA methylation and human disease. *Nature Reviews Genetics*, Vol. 6, pp. 597–610. https://doi.org/10.1038/nrg1655

Robinson, D. A., & Enright, M. C. (2004). Evolution of Staphylococcus aureus by large chromosomal

replacements. *Journal of Bacteriology*, *186*(4), 1060–1064. https://doi.org/10.1128/jb.186.4.1060-1064.2004

Robinson, M., & McCarthy…, D. (2010). edgeR: differential expression analysis of digital gene expression data. In *Bioconductor.Fhcrc.Org*. Retrieved from http://bioconductor.fhcrc.org/packages/2.6/bioc/vignettes/edgeR/inst/doc/edgeR.pdf

Robles, J. A., Qureshi, S. E., Stephen, S. J., Wilson, S. R., Burden, C. J., & Taylor, J. M. (2012). Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics*, *13*(1), 484. https://doi.org/10.1186/1471-2164-13-484

Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., & Gelfand, M. S. (2004). Comparative genomics of the methionine metabolism in Gram-positive bacteria: A variety of regulatory systems. *Nucleic Acids Research*, *32*(11), 3340–3353. https://doi.org/10.1093/nar/gkh659

Roe, C. C., Horn, K. S., Driebe, E. M., Bowers, J., Terriquez, J. A., Keim, P., & Engelthaler, D. M. (2016). Whole genome SNP typing to investigate methicillin-resistant Staphylococcus aureus carriage in a health-care provider as the source of multiple surgical site infections. *Hereditas*, *153*, 11. https://doi.org/10.1186/s41065-016-0017-x

Roer, L., Aarestrup, F. M., & Hasman, H. (2015). The EcoKI type I restriction-modification system in Escherichia coli affects but is not an absolute barrier for conjugation. *Journal of Bacteriology*, *197*(2), 337–342. https://doi.org/10.1128/JB.02418-14

Roncarati, D., & Scarlato, V. (2017). Regulation of heat-shock genes in bacteria: from signal sensing to gene expression output. *FEMS Microbiology Reviews*, *41*(4), 549–574. https://doi.org/10.1093/femsre/fux015

Rouch, D. A., Byrne, M. E., Kong, Y. C., & Skurray, R. A. (1987). The aacA-aphD gentamicin and kanamycin resistance determinant of Tn4001 from Staphylococcus aureus: Expression and nucleotide sequence analysis. *Journal of General Microbiology*, *133*(11), 3039–3052. https://doi.org/10.1099/00221287-133-11-3039

Rouli, L., Merhej, V., Fournier, P. E., & Raoult, D. (2015). The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes and New Infections*, *7*, 72–85. https://doi.org/10.1016/j.nmni.2015.06.005

Ruiz de los Mozos, I., Vergara-Irigaray, M., Segura, V., Villanueva, M., Bitarte, N., Saramago, M., … Toledo-Arana, A. (2013). Base Pairing Interaction between 5′- and 3′-UTRs Controls icaR mRNA Translation in Staphylococcus aureus. *PLoS Genetics*, *9*(12), e1004001. https://doi.org/10.1371/journal.pgen.1004001

Rusinov, I., Ershova, A., Karyagina, A., Spirin, S., & Alexeevski, A. (2015). Lifespan of restriction-modification systems critically affects avoidance of their recognition sites in host genomes. *BMC Genomics*, *16*(1), 1–15. https://doi.org/10.1186/s12864-015-2288-4

Ryan, K. A., & Lo, R. Y. C. (1999). Characterization of a CACAG pentanucleotide repeat in Pasteurella haemolytica and its possible role in modulation of a novel type III restriction-modification system. *Nucleic Acids Research*, *27*(6), 1505–1511. https://doi.org/10.1093/nar/27.6.1505

Sadykov, M. R. (2015). Restriction–Modification Systems as a Barrier for Genetic Manipulation of Staphylococcus aureus BT - Epigenetics Protocols. In *Epigenetics Protocols* (Vol. 1373, pp. 9–23). Retrieved from http://link.springer.com/10.1007/7651_2014_180%5Cnpapers3://publication/doi/10.1007/7651_2014_180

Sáenz-Lahoya, S., Bitarte, N., García, B., Burgui, S., Vergara-Irigaray, M., Valle, J., … Lasa, I. (2019). Noncontiguous operon is a genetic organization for coordinating bacterial gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(5), 1733–1738. https://doi.org/10.1073/pnas.1812746116

Safari, F., Sharifi, M., Farajnia, S., Akbari, B., Karimi Baba Ahmadi, M., Negahdaripour, M., & Ghasemi, Y. (2020, February 17). The interaction of phages and bacteria: the co-evolutionary arms race. *Critical Reviews in Biotechnology*, Vol. 40, pp. 119–137. https://doi.org/10.1080/07388551.2019.1674774

Saint-Girons, I., Parsot, C., Zakin, M. M., Bărzu, O., Cohen, G. N., & Weissbach, H. (1988). Methionine biosynthesis in enterobacteriaceae: Biochemical, regulatory, and evolutionary aspect. *Critical Reviews in Biochemistry and Molecular Biology*, *23*(S1). https://doi.org/10.3109/10409238809083374

Samai, P., Pyenson, N., Jiang, W., Goldberg, G. W., Hatoum-Aslan, A., & Marraffini, L. A. (2015). Co-transcriptional DNA and RNA cleavage during type III CRISPR-cas immunity. *Cell*, *161*(5), 1164–1174. https://doi.org/10.1016/j.cell.2015.04.027

Samson, J. E., Magadán, A. H., Sabri, M., & Moineau, S. (2013, October). Revenge of the phages: Defeating bacterial defences. *Nature Reviews Microbiology*, Vol. 11, pp. 675–687. https://doi.org/10.1038/nrmicro3096

Sánchez-Busó, L., Golparian, D., Parkhill, J., Unemo, M., & Harris, S. R. (2019). Genetic variation regulates the activation and specificity of Restriction-Modification systems in Neisseria gonorrhoeae. *Scientific Reports*, *9*(1), 14685. https://doi.org/10.1038/s41598-019-51102-2

Sánchez-Romero, M. A., Cota, I., & Casadesús, J. (2015). DNA methylation in bacteria: From the methyl group to the methylome. *Current Opinion in Microbiology*, *25*, 9–16. https://doi.org/10.1016/j.mib.2015.03.004

Santoro, A., Simone, B., & Timen, A. (2015). Health trends of communicable diseases. In *A Systematic Review of Key Issues in Public Health* (pp. 5–18). https://doi.org/10.1007/978-3-319-13620-2_2

Sarantopoulou, D., Tang, S. Y., Ricciotti, E., Lahens, N. F., Lekkas, D., Schug, J., … Grant, G. R. (2019). Comparative evaluation of RNA-Seq library preparation methods for strand-specificity and low input. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-019-49889-1

Sato'o, Y., Omoe, K., Ono, H. K., Nakane, A., & Hu, D. L. (2013). A novel comprehensive analysis method for Staphylococcus aureus pathogenicity islands. *Microbiology and Immunology*, *57*(2), 91–99. https://doi.org/10.1111/1348-0421.12007

Saunders, N. A., & Holmes, A. (2007). Multilocus Sequence Typing (MLST) of Staphylococcus aureus. *Methods in Molecular Biology (Clifton, N.J.)*, *391*(1), 71–85. https://doi.org/10.1007/978-1-59745-468-1_6

Schadt, E. E., Banerjee, O., Fang, G., Feng, Z., Wong, W. H., Zhang, X., … Kasarskis, A. (2013). Modeling kinetic rate variation in third generation DNA sequencing data to detect putative modifications to DNA bases. *Genome Research*, *23*(1), 129–141. https://doi.org/10.1101/gr.136739.111

Schlagman, S. L., Hattman, S., & Marinus, M. G. (1986). Direct role of the Escherichia coli Dam DNA methyltransferase in methylation-directed mismatch repair. *Journal of Bacteriology*, *165*(3), 896–900. https://doi.org/10.1128/jb.165.3.896-900.1986

Schoenfelder, S. M. K., Marincola, G., Geiger, T., Goerke, C., Wolz, C., & Ziebuhr, W. (2013). Methionine Biosynthesis in Staphylococcus aureus Is Tightly Controlled by a Hierarchical

Network Involving an Initiator tRNA-Specific T-box Riboswitch. *PLoS Pathogens*, *9*(9). https://doi.org/10.1371/journal.ppat.1003606

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

Segerman, B. (2012). The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Frontiers in Cellular and Infection Microbiology*, *2*, 116. https://doi.org/10.3389/fcimb.2012.00116

Seib, K. L., Pigozzi, E., Muzzi, A., Gawthorne, J. A., Delany, I., Jennings, M. P., & Rappuoli, R. (2011). A novel epigenetic regulator associated with the hypervirulent Neisseria meningitidis clonal complex 41/44 . *The FASEB Journal*, *25*(10), 3622–3633. https://doi.org/10.1096/fj.11-183590

Seib, K. L., Srikhanta, Y. N., Atack, J. M., & Jennings, M. P. (2020, September 8). Epigenetic Regulation of Virulence and Immunoevasion by Phase-Variable Restriction-Modification Systems in Bacterial Pathogens. *Annual Review of Microbiology*, Vol. 74, pp. 655–671. https://doi.org/10.1146/annurev-micro-090817-062346

Shambat, S., Nadig, S., Prabhakara, S., Bes, M., Etienne, J., & Arakere, G. (2012). Clonal complexes and virulence factors of Staphylococcus aureus from several cities in India. *BMC Microbiology*, *12*(1471-2180 (Electronic)), 64. https://doi.org/10.1186/1471-2180-12-64

Shell, S. S., Prestwich, E. G., Baek, S. H., Shah, R. R., Sassetti, C. M., Dedon, P. C., & Fortune, S. M. (2013). DNA Methylation Impacts Gene Expression and Ensures Hypoxic Survival of Mycobacterium tuberculosis. *PLoS Pathogens*, *9*(7). https://doi.org/10.1371/journal.ppat.1003419

Shore, A. C., Rossney, A. S., Brennan, O. M., Kinnevey, P. M., Humphreys, H., Sullivan, D. J., … Coleman, D. C. (2011). Characterization of a novel arginine catabolic mobile element (ACME) and staphylococcal chromosomal cassette mec composite island with significant homology to Staphylococcus epidermidis ACME type II in methicillin-resistant Staphylococcus aureus genotype . *Antimicrobial Agents and Chemotherapy*, *55*(5), 1896–1905. https://doi.org/10.1128/AAC.01756-10

Shukla, S. K., Pantrangi, M., Stahl, B., Briska, A. M., Stemper, M. E., Wagner, T. K., … Dykes, C. W. (2012). Comparative whole-genome mapping to determine Staphylococcus aureus genome size, virulence motifs, and clonality. *Journal of Clinical Microbiology*, *50*(11), 3526–3533. https://doi.org/10.1128/JCM.01168-12

Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., & Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods*, *14*(4), 407–410. https://doi.org/10.1038/nmeth.4184

Smyth, D. S., McDougal, L. K., Gran, F. W., Manoharan, A., Enright, M. C., Song, J. H., … Robinson, D. A. (2010). Population structure of a hybrid clonal group of methicillin-resistant Staphylococcus aureus, ST239-MRSA-III. *PLoS ONE*, *5*(1), e8582. https://doi.org/10.1371/journal.pone.0008582

Smyth, D. S., & Robinson, D. A. (2009). Integrative and sequence characteristics of a novel genetic element, ICE6013, in Staphylococcus aureus. *Journal of Bacteriology*, *191*(19), 5964–5975. https://doi.org/10.1128/JB.00352-09

Solovyev, V. (2011). *V. Solovyev, A Salamov (2011) Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies (Ed. R.W. Li), Nova Science Publishers, p.61-78.*

Soneson, C., & Robinson, M. D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. *Nature Methods*, *15*(4), 255–261. https://doi.org/10.1038/nmeth.4612

Song, W., Sun, H. X., Zhang, C., Cheng, L., Peng, Y., Deng, Z., … Xiao, M. (2019). Prophage Hunter: an integrative hunting tool for active prophages. *Nucleic Acids Research*, *47*(W1), W74–W80. https://doi.org/10.1093/nar/gkz380

Spoor, L. E., Richardson, E., Richards, A. C., Wilson, G. J., Mendonca, C., Gupta, R. K., … Ross Fitzgerald, J. (2015). Recombination-mediated remodelling of host-pathogen interactions during Staphylococcus aureus niche adaptation. *Microbial Genomics*, *1*(4), e000036–e000036. https://doi.org/10.1099/mgen.0.000036

Sridhar, J., Sabarinathan, R., Balan, S. S., Rafi, Z. A., Gunasekaran, P., & Sekar, K. (2011). Junker: An Intergenic Explorer for Bacterial Genomes. *Genomics, Proteomics and Bioinformatics*, *9*(4–5), 179–182. https://doi.org/10.1016/S1672-0229(11)60021-1

Srikhanta, Y. N., Dowideit, S. J., Edwards, J. L., Falsetta, M. L., Wu, H. J., Harrison, O. B., … Jennings, M. P. (2009). Phasevarions mediate random switching of gene expression in pathogenic Neisseria. *PLoS Pathogens*, *5*(4), e1000400. https://doi.org/10.1371/journal.ppat.1000400

Srikhanta, Y. N., Gorrell, R. J., Steen, J. A., Gawthorne, J. A., Kwok, T., Grimmond, S. M., … Jennings, M. P. (2011). Phasevarion mediated epigenetic gene regulation in Helicobacter pylori. *PLoS ONE*, *6*(12), e27569. https://doi.org/10.1371/journal.pone.0027569

Srikhanta, Y. N., Maguire, T. L., Stacey, K. J., Grimmond, S. M., & Jennings, M. P. (2005). The phasevarion: A genetic system controlling coordinated, random switching of expression of multiple genes. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(15), 5547–5551. https://doi.org/10.1073/pnas.0501169102

Stanczak-Mrozek, K. I., Manne, A., Knight, G. M., Gould, K., Witney, A. A., & Lindsay, J. A. (2015). Within-host diversity of MRSA antimicrobial resistances. *Journal of Antimicrobial Chemotherapy*, *70*(8), 2191–2198. https://doi.org/10.1093/jac/dkv119

Stephens, A. J., Huygens, F., Inman-Bamber, J., Price, E. P., Nimmo, G. R., Schooneveldt, J., … Giffard, P. M. (2006). Methicillin-resistant Staphylococcus aureus genotyping using a small set of polymorphisms. *Journal of Medical Microbiology*, *55*(1), 43–51. https://doi.org/10.1099/jmm.0.46157-0

Stoiber, M., Quick, J., Egan, R., Eun Lee, J., Celniker, S., Neely, R., … Brown, J. (2016). De novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing. *BioRxiv*, 094672. https://doi.org/10.1101/094672

Streeter, S. D., Papapanagiotou, I., McGeehan, J. E., & Kneale, G. G. (2004). DNA footprinting and biophysical characterization of the controller protein C.AhdI suggests the basis of a genetic switch. *Nucleic Acids Research*, *32*(21), 6445–6453. https://doi.org/10.1093/nar/gkh975

Strommenger, B., Braulke, C., Heuck, D., Schmidt, C., Pasemann, B., Nübel, U., & Witte, W. (2008). spa typing of Staphylococcus aureus as a frontline tool in epidemiological typing. *Journal of Clinical Microbiology*, *46*(2), 574–581. https://doi.org/10.1128/JCM.01599-07

Strommenger, Birgit, Bartels, M. D., Kurt, K., Layer, F., Rohde, S. M., Boye, K., … Nübel, U. (2014). Evolution of methicillin-resistant Staphylococcus aureus towards increasing resistance. *Journal of Antimicrobial Chemotherapy*, *69*(3), 616–622. https://doi.org/10.1093/jac/dkt413

Su, T. J., Tock, M. R., Egelhaaf, S. U., Poon, W. C. K., & Dryden, D. T. F. (2005). DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping. *Nucleic Acids Research*, *33*(10), 3235–3244. https://doi.org/10.1093/nar/gki618

Sun, D. L., Jiang, X., Wu, Q. L., & Zhou, N. Y. (2013). Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Applied and Environmental Microbiology*, *79*(19), 5962–5969. https://doi.org/10.1128/AEM.01282-13

Suzuki, H. (2012). Host-Mimicking Strategies in DNA Methylation for Improved Bacterial Transformation. In *Methylation - From DNA, RNA and Histones to Diseases and Treatment* (pp. 1–19). https://doi.org/10.5772/51691

Suzuki, M., Matsumoto, M., Takahashi, M., Hayakawa, Y., & Minagawa, H. (2009). Identification of the clonal complexes of Staphylococcus aureus strains by determination of the conservation patterns of small genomic islets. *Journal of Applied Microbiology*, *107*(4), 1367–1374. https://doi.org/10.1111/j.1365-2672.2009.04321.x

Taghbalout, A., Landoulsi, A., Kern, R., Yamazoe, M., Hiraga, S., Holland, B., … Malki, A. (2000). Competition between the replication initiator DnaA and the sequestration factor SeqA for binding to the hemimethylated chromosomal origin of E. coli in vitro. *Genes to Cells*, *5*(11), 873–884. https://doi.org/https://doi.org/10.1046/j.1365-2443.2000.00380.x

Tang, D., Ando, S., Takasaki, Y., & Tadano, J. (2000). Mutational analyses of restriction endonuclease - HindIII mutant E86K with higher activity and altered specificity. In *Protein Engineering* (Vol. 13). https://doi.org/10.1093/protein/13.4.283

Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., … Fraser, C. M. (2001). Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. *Science*, *293*(5529), 498–506. https://doi.org/10.1126/science.1061217

Tewatia, D. (2007). PDF processed with CutePDF evaluation edition www.CutePDF.com. *October*, *24*(October), 149–154. https://doi.org/10.1016/S0140

Thomas, C. M., & Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology*, *3*(9), 711–721. https://doi.org/10.1038/nrmicro1234

Thomas, J. C., Godfrey, P. A., Feldgarden, M., & Robinson, D. A. (2012). Draft genome sequences of Staphylococcus aureus sequence type 34 (ST34) and ST42 Hybrids. *Journal of Bacteriology*, Vol. 194, pp. 2740–2741. https://doi.org/10.1128/JB.00248-12

Thorpe, H. A., Bayliss, S. C., Hurst, L. D., & Feil, E. J. (2017). Comparative analyses of selection operating on nontranslated intergenic regions of diverse bacterial species. *Genetics*, *206*(1), 363–376. https://doi.org/10.1534/genetics.116.195784

Tian, Q., Zou, J., Tang, J., Fang, Y., Yu, Z., & Fan, S. (2019). MRCNN: A deep learning model for regression of genome-wide DNA methylation. *BMC Genomics*, *20*(S2), 192. https://doi.org/10.1186/s12864-019-5488-5

Titheradge, A. J. B., King, J., Ryu, J., & Murray, N. E. (2001). Families of restriction enzymes: An analysis prompted by molecular and genetic data for type ID restriction and modification systems. *Nucleic Acids Research*, *29*(20), 4195–4205. https://doi.org/10.1093/nar/29.20.4195

Tock, M. R., & Dryden, D. T. F. (2005, August 1). The biology of restriction and anti-restriction. *Current Opinion in Microbiology*, Vol. 8, pp. 466–472. https://doi.org/10.1016/j.mib.2005.06.003

Tokajian, S. (2012). Molecular Typing of Staphylococcus aureus: Understanding and Controlling Epidemic Spread. *Journal of Forensic Research*, *03*(04). https://doi.org/10.4172/2157-7145.1000e104

Tourancheau, A., Mead, E., Zhang, X.-S., & Fang, G. (2020). Discovering and exploiting multiple

types of DNA methylation from individual bacteria and microbiome using nanopore sequencing. *BioRxiv*, 2020.02.18.954636. https://doi.org/10.1101/2020.02.18.954636

Tran, H. T. N., Ang, K. S., Chevrier, M., Zhang, X., Lee, N. Y. S., Goh, M., & Chen, J. (2020). A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biology*, *21*(1), 12. https://doi.org/10.1186/s13059-019-1850-9

Usuda, Y., & Kurahashi, O. (2005). Effects of deregulation of methionine biosynthesis on methionine excretion in Escherichia coli. *Applied and Environmental Microbiology*, *71*(6), 3228–3234. https://doi.org/10.1128/AEM.71.6.3228-3234.2005

Van Der Woude, M., Braaten, B., & Low, D. (1996). Epigenetic phase variation of the pap operon in Escherichia coli. *Trends in Microbiology*, *4*(1), 5–9. https://doi.org/10.1016/0966-842X(96)81498-3

van Tonder, A. J., Mistry, S., Bray, J. E., Hill, D. M. C., Cody, A. J., Farmer, C. L., … Brueggemann, A. B. (2014). Defining the Estimated Core Genome of Bacterial Populations Using a Bayesian Decision Model. *PLoS Computational Biology*, *10*(8), e1003788. https://doi.org/10.1371/journal.pcbi.1003788

Vasu, K., & Nagaraja, V. (2013). Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. *Microbiology and Molecular Biology Reviews*, *77*(1), 53–72. https://doi.org/10.1128/mmbr.00044-12

Vasu, Kommireddy, Nagamalleswari, E., & Nagaraja, V. (2012). Promiscuous restriction is a cellular defense strategy that confers fitness advantage to bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(20). https://doi.org/10.1073/pnas.1119226109

Vasu, Kommireddy, Rao, D. N., & Nagaraja, V. (2019). Restriction-modification systems. In R. E. Y. Uldis N. Streips (Ed.), *Encyclopedia of Microbiology* (pp. 102–109). https://doi.org/10.1016/B978-0-12-809633-8.90789-1

Veiga, H., & Pinho, M. G. (2009). Inactivation of the saul type i restriction-modification system is not sufficient to generate Staphylococcus aureus strains capable of efficiently accepting foreign DNA. *Applied and Environmental Microbiology*, *75*(10), 3034–3038. https://doi.org/10.1128/AEM.01862-08

Vipond, I. B., & Halford, S. E. (1995). Specific DNA Recognition by EcoRV Restriction Endonuclease Induced by Calcium Ions. *Biochemistry*, *34*(4), 1113–1119. https://doi.org/10.1021/bi00004a002

Vogan, A. A., & Higgs, P. G. (2011). The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. *Biology Direct*, *6*, 1. https://doi.org/10.1186/1745-6150-6-1

Waggoner, D. (2007). Mechanisms of Disease: Epigenesis. *Seminars in Pediatric Neurology*, *14*(1), 7–14. https://doi.org/10.1016/j.spen.2006.11.004

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., … Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Weiss, A., Broach, W. H., & Shaw, L. N. (2016). Characterizing the transcriptional adaptation of Staphylococcus aureus to stationary phase growth. *Pathogens and Disease*, *74*(5), ftw046. https://doi.org/10.1093/femspd/ftw046

Wencker, F. D. R., Marincola, G., Schoenfelder, S. M. K., Maaß, S., Orte Becher, D. ¨, & Ziebuhr,

W. (2021). Another layer of complexity in Staphylococcus aureus methionine biosynthesis control: unusual RNase III-driven T-box riboswitch cleavage determines met operon mRNA stability and decay. *Nucleic Acids Research*, (1). https://doi.org/10.1093/nar/gkaa1277

Weyand, N. J., & Low, D. A. (2000). Regulation of pap phase variation. Lrp is sufficient for the establishment of the phase off pap DNA methylation pattern and repression of pap transcription in vitro. *Journal of Biological Chemistry*, *275*(5), 3192–3200. https://doi.org/10.1074/jbc.275.5.3192

Whittle, G., Hamburger, N., Shoemaker, N. B., & Salyers, A. A. (2006). A Bacteroides conjugative transposon, CTnERL, can transfer a portion of itself by conjugation without excising from the chromosome. *Journal of Bacteriology*, *188*(3), 1169–1174. https://doi.org/10.1128/JB.188.3.1169-1174.2006

WHO. (2017). Who Publishes List of Bacteria for Which New Antibiotics Are Urgently Needed. *Saudi Medical Journal*, Vol. 38, pp. 444–445. Retrieved from http://www.who.int/mediacentre/news/releases/2017/bacteria-antibiotics-needed/en/

Willbanks, A., Leary, M., Greenshields, M., Tyminski, C., Heerboth, S., Lapinska, K., … Sarkar, S. (2016). The evolution of epigenetics: From prokaryotes to humans and its biological consequences. *Genetics and Epigenetics*, *1*(8), 25–36. https://doi.org/10.4137/GeG.s31863

Willemse, N., & Schultsz, C. (2016). Distribution of type I restriction-modification systems in streptococcus suis: An outlook. *Pathogens*, *5*(4), 62. https://doi.org/10.3390/pathogens5040062

Williams, A. G., Thomas, S., Wyman, S. K., & Holloway, A. K. (2014). RNA-seq Data: Challenges in and Recommendations for Experimental Design and Analysis. *Current Protocols in Human Genetics*, *83*(1), 11.13.1-11.13.20. https://doi.org/10.1002/0471142905.hg1113s83

Wilson, G. G., & Murray, N. E. (1991). Restriction and modification systems. *Annual Review of Genetics*, *25*(1), 585–627. https://doi.org/10.1146/annurev.ge.25.120191.003101

Wion, D., & Casadesús, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nature Reviews. Microbiology*, *4*(3), 183–192. https://doi.org/10.1038/nrmicro1350

Wong, J. G., Chen, M. I., Win, M. K., Ng, P. Y., & Chow, A. (2016). Length of stay an important mediator of hospital-acquired methicillin-resistant Staphylococcus aureus. *Epidemiology and Infection*, *144*(6), 1248–1256. https://doi.org/10.1017/S0950268815002733

Wood, R. J., Maynard-Smith, M. D., Robinson, V. L., Oyston, P. C. F., Titball, R. W., & Roach, P. L. (2007). Kinetic analysis of Yersinia pestis DNA adenine methyltransferase activity using a hemimethylated molecular break light oligonucleotide. *PLoS ONE*, *2*(8), e801. https://doi.org/10.1371/journal.pone.0000801

Xia, G., & Wolz, C. (2014). Phages of Staphylococcus aureus and their impact on host evolution. *Infection, Genetics and Evolution*, *21*, 593–601. https://doi.org/10.1016/j.meegid.2013.04.022

Xu, B., Zhang, P., Li, W., Liu, R., Tang, J., & Fan, H. (2017). hsdS, belonging to the type i restriction-modification system, contributes to the streptococcus suis serotype 2 survival ability in phagocytes. *Frontiers in Microbiology*, *8*(AUG), 1524. https://doi.org/10.3389/fmicb.2017.01524

Xu, S. Y., Corvaglia, A. R., Chan, S. H., Zheng, Y., & Linder, P. (2011). A type IV modification-dependent restriction enzyme SauUSI from Staphylococcus aureus subsp. aureus USA300. *Nucleic Acids Research*, *39*(13), 5597–5610. https://doi.org/10.1093/nar/gkr098

Xu, S. yong, Nugent, R. L., Ksmkttil, J., Fomenkov, A., Gupta, Y., Aggarwal, A., … Morgan, R. (2012). Characterization of type II and III restriction-modification systems from Bacillus cereus strains ATCC 10987 and ATCC 14579. *Journal of Bacteriology*, *194*(1), 49–60. https://doi.org/10.1128/JB.06248-11

Yang, J., & Zhang, Y. (2015). I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Research*, *43*(W1), W174–W181. https://doi.org/10.1093/nar/gkv342

Ye, Z., Vasco, D. A., Carter, T. C., Brilliant, M. H., Schrodi, S. J., & Shukla, S. K. (2014). Genome wide association study of SNP-, gene-, and pathway-based approaches to identify genes influencing susceptibility to Staphylococcus aureus infections. *Frontiers in Genetics*, *5*(MAY), 1–1. https://doi.org/10.3389/fgene.2014.00125

Yuan, R., & Hamilton, D. L. (1984). *Type I and Type III Restriction-Modification Enzymes*. https://doi.org/10.1007/978-1-4613-8519-6_2

Zaleski, P., Wojciechowski, M., & Piekarowicz, A. (2005). The role of Dam methylation in phase variation of Haemophilus influenzae genes involved in defence against phage infection. *Microbiology (Reading, England)*, *151*(Pt 10), 3361–3369. https://doi.org/10.1099/mic.0.28184-0

Zhang, W., Spector, T. D., Deloukas, P., Bell, J. T., & Engelhardt, B. E. (2015). Predicting genome-wide DNA methylation using methylation marks, genomic position, and DNA regulatory elements. *Genome Biology*, *16*(1). https://doi.org/10.1186/s13059-015-0581-9

Zhou, B., Schrader, J. M., Kalogeraki, V. S., Abeliuk, E., Dinh, C. B., Pham, J. Q., … Shapiro, L. (2015). The Global Regulatory Architecture of Transcription during the Caulobacter Cell Cycle. *PLoS Genetics*, *11*(1), e1004831. https://doi.org/10.1371/journal.pgen.1004831

Zweiger, G., Marczynski, G., & Shapiro, L. (1994). A Caulobacter DNA methyltransferase that functions only in the predivisional cell. *Journal of Molecular Biology*, *235*(2), 472–485. https://doi.org/10.1006/jmbi.1994.1007

8. APPENDIX

8.1 CHAPTER 3 APPENDIX

Table 8.1 | Restriction_and_Modification Analysis 6mA Motif Results for *S. aureus* NCTC Collection

Isolate	ST	motifString	partnerMotifString	groupTag	Methylated / Detected	Methylated Motif	Detected Motif	Mean Score	Mean IPD Ratio	Mean Coverage
NCTC10344	97	GAAGNNNNNTAC	GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	1.0000	268	268	201.89	6.67	132.01
		GTANNNNNCTTC	GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	0.9963	267	268	187.90	5.63	137.59
		CCAYNNNNNNTTYG	CRAANNNNNNRTGG	CCAYNNNNNNTTYG/CRAANNNNNNRTGG	0.9817	321	327	179.95	4.92	137.51
		CRAANNNNNNRTGG	CCAYNNNNNNTTYG	CCAYNNNNNNTTYG/CRAANNNNNNRTGG	0.9450	309	327	156.27	4.32	133.68
		CCAYNNNNNNRTC	GAYNNNNNNRTGG	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9749	817	838	176.98	4.64	136.55
		GAYNNNNNNRTGG	CCAYNNNNNNRTC	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9690	812	838	172.83	4.42	136.74
NCTC10345	97	GAAGNNNNNTAC	GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	0.9963	268	269	233.45	6.81	152.22
		GTANNNNNCTTC	GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	0.9926	267	269	223.14	5.88	156.09
		CCAYNNNNNNNTTYG	CRAANNNNNNNRTGG	CCAYNNNNNNNTTYG/CRAANNNNNNNRTGG	0.9909	326	329	207.66	5.01	155.72
		CRAANNNNNNNRTGG	CCAYNNNNNNNTTYG	CCAYNNNNNNNTTYG/CRAANNNNNNNRTGG	0.9544	314	329	182.55	4.45	152.19
		CCAYNNNNNNRTC	GAYNNNNNNRTGG	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9727	821	844	205.74	4.84	156.53
		GAYNNNNNNRTGG	CCAYNNNNNNRTC	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9645	814	844	202.11	4.60	156.20
NCTC10399	707	GAAGNNNNNRTTG	CAAYNNNNNCTTC	GAAGNNNNNRTTG/CAAYNNNNNCTTC	0.9981	518	519	212.99	6.12	146.17
		CAAYNNNNNCTTC	GAAGNNNNNRTTG	GAAGNNNNNRTTG/CAAYNNNNNCTTC	0.9827	510	519	206.22	4.54	147.42
		CCAYNNNNNCCT	AGGNNNNNRTGG	CCAYNNNNNCCT/AGGNNNNNRTGG	0.9842	374	380	201.67	4.82	152.44
		AGGNNNNNRTGG	CCAYNNNNNCCT	CCAYNNNNNCCT/AGGNNNNNRTGG	0.9684	368	380	201.83	5.36	150.07
NCTC10442	250	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9957	701	704	220.70	5.35	155.08
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9957	701	704	221.84	6.35	154.74
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9641	483	501	190.62	4.68	159.14
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9561	479	501	195.62	4.42	159.56
NCTC10443	250	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	1.0000	672	672	197.02	5.30	136.85
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9970	670	672	198.49	6.23	136.71
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9695	476	491	170.93	4.67	140.25
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9511	467	491	176.93	4.42	139.99
NCTC10649	1254	GAGNNNNNNRTTC	GAAYNNNNNNCTC	GAGNNNNNNRTTC/GAAYNNNNNNCTC	0.9887	262	265	96.37	6.36	60.40
		GAAYNNNNNNCTC	GAGNNNNNNRTTC	GAGNNNNNNRTTC/GAAYNNNNNNCTC	0.9623	255	265	91.06	4.82	59.53
		AGGNNNNNTTTC	GAAANNNNNCCT	AGGNNNNNTTTC/GAAANNNNNCCT	0.9883	338	342	100.08	6.48	61.11
		GAAANNNNNCCT	AGGNNNNNTTTC	AGGNNNNNTTTC/GAAANNNNNCCT	0.9854	337	342	84.61	5.09	60.74
NCTC10652	8	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	1.0000	683	683	211.61	5.54	143.91
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9941	679	683	212.87	6.67	143.72
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9671	470	486	180.00	4.91	143.61
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9588	466	486	185.58	4.61	144.15
NCTC10654	250	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9958	716	719	188.60	5.60	126.03
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9958	716	719	190.68	6.66	125.98

		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9699	484	499	160.01	4.92	125.49
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9519	475	499	168.07	4.73	125.79
NCTC10655	10	GACNNNNNNTAG	CTANNNNNNGTC	GACNNNNNNTAG/CTANNNNNNGTC	0.9924	520	524	498.80	4.76	424.26
		CTANNNNNNGTC	GACNNNNNNTAG	GACNNNNNNTAG/CTANNNNNNGTC	0.9542	500	524	403.15	4.31	423.51
		CYAANNNNNNNTCC	GGANNNNNNNTTRG	CYAANNNNNNNTCC/GGANNNNNNNTTRG	0.9900	496	501	487.33	5.29	408.87
		GGANNNNNNNTTRG	CYAANNNNNNNTCC	CYAANNNNNNNTCC/GGANNNNNNNTTRG	0.9741	488	501	468.47	4.85	414.23
NCTC10656	5	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9956	678	681	83.44	5.49	49.47
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9897	674	681	82.38	6.58	49.56
		TACBNNNNNRTGG	CCAYNNNNNVGTA	TACBNNNNNRTGG/CCAYNNNNNVGTA	0.9862	285	289	79.60	4.89	50.84
		CCAYNNNNNVGTA	TACBNNNNNRTGG	TACBNNNNNRTGG/CCAYNNNNNVGTA	0.9170	265	289	77.48	4.53	50.42
NCTC10657	254	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9985	679	680	213.32	5.21	150.31
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9956	677	680	215.33	6.21	149.84
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9650	469	486	178.62	4.54	150.71
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9527	463	486	186.81	4.39	150.86
NCTC10703	3526	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9987	760	761	206.09	5.24	145.38
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9961	758	761	204.13	6.26	144.62
		TAAGNNNNNNTTC	GAANNNNNNCTTA	TAAGNNNNNNTTC/GAANNNNNNCTTA	0.9957	463	465	213.88	6.80	141.11
		GAANNNNNNCTTA	TAAGNNNNNNTTC	TAAGNNNNNNTTC/GAANNNNNNCTTA	0.9871	459	465	188.17	5.05	142.89
NCTC10804	254	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9986	689	690	100.31	5.42	61.04
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9971	688	690	97.93	6.56	60.93
		TACBNNNNNRTGG	CCAYNNNNNVGTA	TACBNNNNNRTGG/CCAYNNNNNVGTA	0.9965	287	288	95.25	4.83	64.16
		CCAYNNNNNVGTA	TACBNNNNNRTGG	TACBNNNNNRTGG/CCAYNNNNNVGTA	0.9340	269	288	92.02	4.52	63.66
NCTC10833	254	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9903	711	718	103.01	5.11	64.77
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9889	710	718	108.05	6.97	64.48
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9878	487	493	99.58	5.02	65.38
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9797	483	493	99.44	4.60	66.43
NCTC10988	254	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9332	754	808	100.78	5.14	63.48
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9295	751	808	99.34	6.24	63.08
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.8839	495	560	92.15	4.58	65.04
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.8839	495	560	95.38	4.39	64.99
NCTC11561	30	HATCNNNNNCTWC		HATCNNNNNCTWC	0.9962	1057	1061	96.52	5.29	60.95
		GWAGNNNNNGAT		GWAGNNNNNGAT	0.9839	1160	1179	93.94	5.80	60.02
NCTC11939	239	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9807	710	724	94.14	5.17	58.54
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9738	705	724	92.92	6.31	57.75
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9268	494	533	84.74	4.60	58.54
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9212	491	533	88.15	4.40	59.10
NCTC11940	239	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	1.0000	671	671	190.73	5.48	127.22
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9955	668	671	187.26	6.47	126.08
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9763	495	507	165.30	4.81	128.59
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9625	488	507	167.24	4.51	128.61
NCTC11962	30	ATCNNNNNCTWC	GWAGNNNNNGAT	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9942	1198	1205	165.18	4.96	120.09
		GWAGNNNNNGAT	ATCNNNNNCTWC	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9876	1190	1205	165.51	5.40	118.40

		CGANNNNNNNTCC	GGANNNNNNNTCG	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9746	345	354	142.34	4.14	121.16
		GGANNNNNNNTCG	CGANNNNNNNTCC	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9605	340	354	155.18	4.36	121.17
NCTC11963	30	CCANNNNNGTR	YACNNNNNTGG	CCANNNNNGTR/YACNNNNNTGG	0.9813717	1159	1181	208.73	5.31	159.00
		YACNNNNNTGG	CCANNNNNGTR	CCANNNNNGTR/YACNNNNNTGG	0.977138	1154	1181	199.43	4.95	159.64
		ATCNNNNNCTWC	GWAGNNNNNGAT	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9812092	1201	1224	222.42	5.37	158.82
		GWAGNNNNNGAT	ATCNNNNNCTWC	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9754902	1194	1224	223.69	5.79	156.59
NCTC11965	30	ATCNNNNNCTWC	GWAGNNNNNGAT	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9983	1182	1184	149.41	5.39	100.67
		GWAGNNNNNGAT	ATCNNNNNCTWC	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9916	1174	1184	147.39	5.81	99.02
		CGANNNNNNNTCC	GGANNNNNNNTCG	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9829	345	351	125.04	4.43	102.31
		GGANNNNNNNTCG	CGANNNNNNNTCC	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9801	344	351	138.28	4.67	102.30
NCTC12232	8	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9958	709	712	141.38	5.43	91.90
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9916	706	712	139.92	6.37	91.56
		ACANNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNRTGG/CCAYNNNNNNTGT	0.9697	480	495	123.46	4.74	92.60
		CCAYNNNNNNTGT	ACANNNNNRTGG	ACANNNNNRTGG/CCAYNNNNNNTGT	0.9556	473	495	125.17	4.50	92.35
NCTC12233	8	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	1.0000	686	686	118.92	5.22	75.95
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9956	683	686	118.12	6.42	76.01
		ACANNNNNRTGG		ACANNNNNRTGG	0.9612	471	490	108.20	4.68	78.44
NCTC12880	151	CAAGNNNNNNTARC	GYTANNNNNNCTTG	CAAGNNNNNNTARC/GYTANNNNNNCTTG	1.0000	197	197	199.41	6.04	136.18
		GYTANNNNNNCTTG	CAAGNNNNNNTARC	CAAGNNNNNNTARC/GYTANNNNNNCTTG	0.9797	193	197	184.17	5.16	139.85
NCTC12981	243	ATCNNNNNCTWC	GWAGNNNNNGAT	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9813717	1159	1181	101.011	5.24495	63.04918
		GWAGNNNNNGAT	ATCNNNNNCTWC	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9669771	1142	1181	97.4063	5.74844	61.99212
		CGANNNNNNNTCC	GGANNNNNNNTCG	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9387187	337	359	88.2938	4.49042	63.24036
		GGANNNNNNNTCG	CGANNNNNNNTCC	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9387187	337	359	92.3739	4.61065	63.90801
NCTC13132	247	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9903	712	719	198.46	5.49	135.15
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9861	709	719	196.40	6.41	134.72
		CCAYNNNNNNTGT		CCAYNNNNNNTGT	0.9553	492	515	173.14	4.53	134.47
NCTC13133	8	AGGNNNNNGAT	ATCNNNNNCCT	AGGNNNNNGAT/ATCNNNNNCCT	1.0000	736	736	118.40	7.05	72.39
		ATCNNNNNCCT	AGGNNNNNGAT	AGGNNNNNGAT/ATCNNNNNCCT	1.0000	736	736	113.60	5.30	72.86
		ACANNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNRTGG/CCAYNNNNNNTGT	0.9981	512	513	106.41	5.16	71.49
		CCAYNNNNNNTGT	ACANNNNNRTGG	ACANNNNNRTGG/CCAYNNNNNNTGT	0.9786	502	513	106.00	4.66	72.04
NCTC13134	239	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	1.0000	723	723	178.70	5.30	119.78
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9959	720	723	175.81	6.14	119.83
		ACANNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNRTGG/CCAYNNNNNNTGT	0.9666	492	509	152.76	4.49	121.27
		CCAYNNNNNNTGT	ACANNNNNRTGG	ACANNNNNRTGG/CCAYNNNNNNTGT	0.9548	486	509	158.20	4.42	121.25
NCTC13135	239	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9802	742	757	175.28	5.29	118.66
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9762	739	757	174.64	6.34	118.00
		ACANNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNRTGG/CCAYNNNNNNTGT	0.9576	519	542	154.25	4.67	121.52
		CCAYNNNNNNTGT	ACANNNNNRTGG	ACANNNNNRTGG/CCAYNNNNNNTGT	0.9502	515	542	156.46	4.36	121.30
NCTC13136	8	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	1.0000	669	669	163.73	5.44	107.17
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9985	668	669	160.93	6.51	106.25
		CCAYNNNNNNTGT		CCAYNNNNNNTGT	0.9623	460	478	146.09	4.53	108.25

NCTC13137	1148	GCANNNNNNTCC	GGANNNNNNTGC	GCANNNNNNTCC/GGANNNNNNTGC		0.9894	375	379	196.22	5.17	148.24
		GGANNNNNNTGC	GCANNNNNNTCC	GCANNNNNNTCC/GGANNNNNNTGC		0.9578	363	379	185.02	4.31	148.14
		CCAYNNNNNNRTC	GAYNNNNNNRTGG	CCAYNNNNNNRTC/GAYNNNNNNRTGG		0.9804	802	818	192.51	4.69	147.55
		GAYNNNNNNRTGG	CCAYNNNNNNRTC	CCAYNNNNNNRTC/GAYNNNNNNRTGG		0.9707	794	818	191.15	4.54	148.12
NCTC13138	250	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT		1.0000	690	690	199.28	5.21	138.59
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT		0.9986	689	690	198.15	6.24	138.32
		CCAYNNNNNNTGT		CCAYNNNNNNTGT		0.9554	471	493	173.57	4.35	139.16
NCTC13139	8	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT		0.9931	719	724	190.34	5.16	133.78
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT		0.9917	718	724	189.98	6.23	132.38
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT		0.9632	498	517	163.88	4.57	136.46
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT		0.9439	488	517	171.69	4.41	137.51
NCTC13140	8	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT		1.0000	714	714	197.68	5.40	134.59
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT		1.0000	714	714	198.65	6.38	134.61
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT		0.9674	504	521	172.03	4.80	136.77
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT		0.9539	497	521	175.73	4.45	136.93
NCTC13141	8	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT		0.9973	736	738	209.17	5.39	146.30
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT		0.9946	734	738	210.49	6.35	145.41
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT		0.9719	518	533	178.16	4.75	146.32
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT		0.9493	506	533	185.79	4.46	146.98
NCTC13142	22	GAAGNNNNTAC	GTANNNNCTTC	GAAGNNNNTAC/GTANNNNCTTC		1.0000	260	260	98.22	6.55	59.51
		GTANNNNCTTC	GAAGNNNNTAC	GAAGNNNNTAC/GTANNNNCTTC		0.9923	258	260	95.62	5.57	62.45
		YTCANNNNNNCCT	AGGNNNNNNTGAR	YTCANNNNNNCCT/AGGNNNNNNTGAR		0.9872	692	701	93.97	5.31	61.76
		AGGNNNNNNTGAR	YTCANNNNNNCCT	YTCANNNNNNCCT/AGGNNNNNNTGAR		0.9800	687	701	89.88	5.55	60.26
NCTC13143	36	ATCNNNNNCTWC	GWAGNNNNNGAT	ATCNNNNNCTWC/GWAGNNNNNGAT		0.9927	1224	1233	208.16	5.19	150.35
		GWAGNNNNNGAT	ATCNNNNNCTWC	ATCNNNNNCTWC/GWAGNNNNNGAT		0.9862	1216	1233	208.89	5.56	148.23
		CGANNNNNNNTCC	GGANNNNNNNTCG	CGANNNNNNNTCC/GGANNNNNNNTCG		0.9831	348	354	177.63	4.21	154.94
		GGANNNNNNNTCG	CGANNNNNNNTCC	CGANNNNNNNTCC/GGANNNNNNNTCG		0.9746	345	354	193.42	4.41	154.39
NCTC13277	30	ATCNNNNNCTWC	GWAGNNNNNGAT	ATCNNNNNCTWC/GWAGNNNNNGAT		1.0000	1208	1208	120.30	5.45	78.58
		GWAGNNNNNGAT	ATCNNNNNCTWC	ATCNNNNNCTWC/GWAGNNNNNGAT		1.0000	1208	1208	118.53	6.97	77.40
		GGANNNNNNNTCG	CGANNNNNNNTCC	GGANNNNNNNTCG/CGANNNNNNNTCC		1.0000	352	352	114.79	5.26	79.21
		CGANNNNNNNTCC	GGANNNNNNNTCG	GGANNNNNNNTCG/CGANNNNNNNTCC		0.9943	350	352	109.51	4.81	79.56
NCTC13297	1	GNNGANNNNNNRTTA	TAAYNNNNNNNTCNNC	GNNGANNNNNNRTTA/TAAYNNNNNNNTCNNC		0.9882	1260	1275	185.35	5.52	128.78
		TAAYNNNNNNNTCNNC	GNNGANNNNNNNRTTA	GNNGANNNNNNNRTTA/TAAYNNNNNNNTCNNC		0.9741	1242	1275	183.07	4.99	126.31
		CCAYNNNNNTTAA	TTAANNNNNRTGG	CCAYNNNNNTTAA/TTAANNNNNRTGG		0.9841	496	504	181.79	5.06	128.77
		TTAANNNNNRTGG	CCAYNNNNNTTAA	CCAYNNNNNTTAA/TTAANNNNNRTGG		0.9623	485	504	164.57	4.93	128.55
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT		0.9663	487	504	167.21	4.85	129.29
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT		0.9544	481	504	167.56	4.49	129.27
NCTC13298	121	AGGNNNNNNTCC	GGANNNNNCCT	AGGNNNNNNTCC/GGANNNNNCCT		0.9919	492	496	179.54	5.76	126.27
		GGANNNNNCCT	AGGNNNNNNTCC	AGGNNNNNNTCC/GGANNNNNCCT		0.9778	485	496	173.96	4.69	128.73
		GACNNNNNNTAYG	CRTANNNNNNGTC	GACNNNNNNTAYG/CRTANNNNNNGTC		0.9865	366	371	179.18	4.62	133.19
		CRTANNNNNNGTC	GACNNNNNNTAYG	GACNNNNNNTAYG/CRTANNNNNNGTC		0.9569	355	371	151.35	4.25	132.92
NCTC13299	30	CAGNNNNNRAAT	ATTYNNNNNCTG	CAGNNNNNRAAT/ATTYNNNNNCTG		0.9972	1064	1067	243.05	7.83	150.96
		ATTYNNNNNCTG	CAGNNNNNRAAT	CAGNNNNNRAAT/ATTYNNNNNCTG		0.9953	1062	1067	216.21	5.59	148.24

ID	N	Seq1	Seq2	Combined					
		ATCNNNNNCTWC	GWAGNNNNNGAT	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9958	1178	1183	211.10	5.29	149.65
		GWAGNNNNNGAT	ATCNNNNNCTWC	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9907	1172	1183	213.81	5.72	147.53
NCTC13373	39	GWAGNNNNNGAT	ATCNNNNNCTWC	GWAGNNNNNGAT/ATCNNNNNCTWC	0.9984	1230	1232	84.73	6.95	51.93
		ATCNNNNNCTWC	GWAGNNNNNGAT	GWAGNNNNNGAT/ATCNNNNNCTWC	0.9976	1229	1232	87.64	5.18	54.62
		CGANNNNNNNTCC	GGANNNNNNNTCG	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9944	352	354	79.07	4.81	54.52
		GGANNNNNNNTCG	CGANNNNNNNTCC	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9774	346	354	81.27	4.99	53.79
NCTC13394	8	TAAGNNNNNNTTC	GAANNNNNNCTTA	TAAGNNNNNNTTC/GAANNNNNNCTTA	1.0000	434	434	210.99	6.86	136.15
		GAANNNNNNCTTA	TAAGNNNNNNTTC	TAAGNNNNNNTTC/GAANNNNNNCTTA	0.9885	429	434	187.86	5.12	138.91
		AGGNNNNNGAT	ATCNNNNNCCT	AGGNNNNNGAT/ATCNNNNNCCT	0.9986	701	702	205.52	6.40	138.73
		ATCNNNNNCCT	AGGNNNNNGAT	AGGNNNNNGAT/ATCNNNNNCCT	0.9986	701	702	206.22	5.43	139.35
		CCAYNNNNNTGT		CCAYNNNNNTGT	0.9537	515	540	181.73	4.47	141.23
NCTC13395	8	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	1.0000	687	687	101.02	5.46	60.90
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9898	680	687	98.75	6.51	60.73
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9587	488	509	91.88	4.82	61.87
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9411	479	509	94.78	4.67	61.76
NCTC13434	121	AGGNNNNNNTCC	GGANNNNNNCCT	AGGNNNNNNTCC/GGANNNNNNCCT	0.9919	492	496	204.60	5.97	141.91
		GGANNNNNNCCT	AGGNNNNNNTCC	AGGNNNNNNTCC/GGANNNNNNCCT	0.9879	490	496	197.66	4.91	143.90
		GACNNNNNNTAYG	CRTANNNNNNGTC	GACNNNNNNTAYG/CRTANNNNNNGTC	0.9892	368	372	197.46	4.81	143.68
		CRTANNNNNNGTC	GACNNNNNNTAYG	GACNNNNNNTAYG/CRTANNNNNNGTC	0.9624	358	372	166.52	4.43	142.66
NCTC13435	80	GACNNNNNNTTYG	CRAANNNNNNGTC	GACNNNNNNTTYG/CRAANNNNNNGTC	0.9976	423	424	130.91	5.53	88.64
		CRAANNNNNNGTC	GACNNNNNNTTYG	GACNNNNNNTTYG/CRAANNNNNNGTC	0.9764	414	424	107.44	4.59	86.43
		TCTANNNNNNRTTC	GAAYNNNNNNNTAGA	TCTANNNNNNRTTC/GAAYNNNNNNNTAGA	0.9923	258	260	125.74	5.74	87.63
		GAAYNNNNNNNTAGA	TCTANNNNNNRTTC	TCTANNNNNNRTTC/GAAYNNNNNNNTAGA	0.9808	255	260	120.82	5.03	83.78
NCTC13616	22	GAAGNNNNNTAC	GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	1.0000	259	259	253.14	6.86	169.41
		GTANNNNNCTTC	GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	0.9961	258	259	238.24	5.78	175.05
		AGGNNNNNNTGAR	YTCANNNNNNCCT	AGGNNNNNNTGAR/YTCANNNNNNCCT	0.9986	696	697	226.96	5.92	171.42
		YTCANNNNNNCCT	AGGNNNNNNTGAR	AGGNNNNNNTGAR/YTCANNNNNNCCT	0.9986	696	697	233.71	5.56	176.04
NCTC13626	239	AGGNNNNNGAT	ATCNNNNNCCT	AGGNNNNNGAT/ATCNNNNNCCT	0.9973	750	752	225.27	6.48	157.70
		ATCNNNNNCCT	AGGNNNNNGAT	AGGNNNNNGAT/ATCNNNNNCCT	0.9960	749	752	219.50	5.22	157.94
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9819	541	551	190.35	4.68	160.33
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9710	535	551	195.07	4.40	160.60
NCTC13758	8	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	1.0000	666	666	197.68	5.34	136.76
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9955	663	666	199.88	6.31	136.52
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9646	463	480	168.28	4.62	138.13
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9458	454	480	174.93	4.46	137.80
NCTC13811	30	ATCNNNNNCTWC	GWAGNNNNNGAT	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9959	1225	1230	141.93	5.30	93.46
		GWAGNNNNNGAT	ATCNNNNNCTWC	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9886	1216	1230	138.49	5.68	92.42
		CGANNNNNNNTCC	GGANNNNNNNTCG	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9832	351	357	120.85	4.30	95.34
		GGANNNNNNNTCG	CGANNNNNNNTCC	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9748	348	357	128.37	4.59	94.14
NCTC13812	8	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9972	714	716	186.87	5.44	125.95
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9958	713	716	186.94	6.39	125.51
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9645	489	507	163.62	4.71	129.26
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9428	478	507	169.22	4.53	129.56

ID	N	Motif 1	Motif 2	Combined	Score	Count 1	Count 2	Val 1	Val 2	Val 3
NCTC13841	464	GAAGNNNNNTAC	GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	1.0000	270	270	203.83	6.39	133.91
		GTANNNNNCTTC	GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	0.9963	269	270	192.56	5.59	137.68
		CCAYNNNNNNTTYG	CRAANNNNNNRTGG	CCAYNNNNNNTTYG/CRAANNNNNNRTGG	0.9909	325	328	184.80	4.89	139.13
		CRAANNNNNNRTGG	CCAYNNNNNNTTYG	CCAYNNNNNNTTYG/CRAANNNNNNRTGG	0.9543	313	328	158.85	4.31	135.73
		CCAYNNNNNNRTC	GAYNNNNNNRTGG	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9808	817	833	183.50	4.72	138.28
		GAYNNNNNNRTGG	CCAYNNNNNNRTC	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9652	804	833	178.66	4.48	137.69
NCTC1803	133	CYAANNNNNNNTCC	GGANNNNNNNTTRG	CYAANNNNNNNTCC/GGANNNNNNNTTRG	0.9791	469	479	134.61	5.19	92.53
		GGANNNNNNNTTRG	CYAANNNNNNNTCC	CYAANNNNNNNTCC/GGANNNNNNNTTRG	0.9499	455	479	131.16	4.90	92.42
		CAGNNNNNRTGA	TCAYNNNNNCTG	CAGNNNNNRTGA/TCAYNNNNNCTG	0.9619	530	551	134.71	5.63	91.20
		TCAYNNNNNCTG	CAGNNNNNRTGA	CAGNNNNNRTGA/TCAYNNNNNCTG	0.9546	526	551	124.01	4.60	90.50
NCTC2669	30	HATCNNNNNCTWC		HATCNNNNNCTWC	0.9981	1060	1062	192.79	5.28	136.31
		GWAGNNNNNGAT		GWAGNNNNNGAT	0.9941	1170	1177	192.99	5.69	134.92
		GGANNNNNNNTCG	CGANNNNNNNTCC	GGANNNNNNNTCG/CGANNNNNNNTCC	0.9828	343	349	180.78	4.60	138.83
		CGANNNNNNNTCC	GGANNNNNNNTCG	GGANNNNNNNTCG/CGANNNNNNNTCC	0.9799	342	349	168.82	4.42	140.15
NCTC3750	121	GGANNNNNCCT		GGANNNNNCCT	0.9839	489	497	203.33	4.81	149.70
		AGGNNNNNNWCC		AGGNNNNNNWCC	0.7259	874	1204	166.20	4.89	147.69
NCTC3761	464	GAAGNNNNNTAC	GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	1.0000	262	262	206.07	6.84	133.84
		GTANNNNNCTTC	GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	0.9962	261	262	197.45	5.87	138.58
		GACNNNNNNTTYG	CRAANNNNNNGTC	GACNNNNNNTTYG/CRAANNNNNNGTC	0.9953	420	422	199.14	5.06	140.32
		CRAANNNNNNGTC	GACNNNNNNTTYG	GACNNNNNNTTYG/CRAANNNNNNGTC	0.9218	389	422	156.52	4.21	135.19
		CCAYNNNNNNRTC	GAYNNNNNNRTGG	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9830	809	823	189.09	4.85	139.77
		GAYNNNNNNRTGG	CCAYNNNNNNRTC	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9721	800	823	184.90	4.68	139.54
NCTC4136	8	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	1.0000	668	668	224.77	5.47	156.65
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9970	666	668	224.24	6.48	155.35
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9769	466	477	188.76	4.74	155.95
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9623	459	477	195.02	4.49	155.41
NCTC4137	464	GAAGNNNNNTAC	GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	1.0000	260	260	196.18	6.89	123.97
		GTANNNNNCTTC	GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	0.9962	259	260	185.77	5.96	128.41
		GACNNNNNNTTYG	CRAANNNNNNGTC	GACNNNNNNTTYG/CRAANNNNNNGTC	0.9953	422	424	179.98	5.07	128.01
		CRAANNNNNNGTC	GACNNNNNNTTYG	GACNNNNNNTTYG/CRAANNNNNNGTC	0.9340	396	424	142.54	4.16	122.95
		CCAYNNNNNNRTC	GAYNNNNNNRTGG	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9790	794	811	173.89	4.89	127.64
		GAYNNNNNNRTGG	CCAYNNNNNNRTC	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9766	792	811	170.90	4.60	129.34
NCTC4163	464									
		GAAGNNNNNTAC	GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	1.0000	267	267	135.40	6.82	84.06
		GTANNNNNCTTC	GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	0.9963	266	267	130.46	5.98	84.69
		CCAYNNNNNNTTYG		CCAYNNNNNNTTYG	0.9875	316	320	131.92	5.14	87.82
		CCAYNNNNNNRTC	GAYNNNNNNRTGG	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9793	803	820	128.76	4.96	86.60
		GAYNNNNNNRTGG	CCAYNNNNNNRTC	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9659	792	820	123.05	4.70	86.19
NCTC5655	30	ATCNNNNNCTWC	GWAGNNNNNGAT	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9740	1163	1194	116.82	5.12	79.12
		GWAGNNNNNGAT	ATCNNNNNCTWC	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9598	1146	1194	115.20	5.59	78.03
		CGANNNNNNNTCC	GGANNNNNNNTCG	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9389	338	360	102.46	4.22	82.10
		GGANNNNNNNTCG	CGANNNNNNNTCC	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9389	338	360	110.98	4.43	81.59
NCTC5656	30	ATCNNNNNCTWC	GWAGNNNNNGAT	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9893	1197	1210	131.05	5.31	85.69
		GWAGNNNNNGAT	ATCNNNNNCTWC	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9793	1185	1210	127.18	5.69	84.60

NCTC5657	10	GACNNNNNNTAG	CTANNNNNNGTC	GACNNNNNNTAG/CTANNNNNNGTC	0.9695	509	525	69.69	4.72	45.46
		CTANNNNNNGTC	GACNNNNNNTAG	GACNNNNNNTAG/CTANNNNNNGTC	0.9105	478	525	65.98	4.39	45.14
		CYAANNNNNNNTCC	GGANNNNNNNTTRG	CYAANNNNNNNTCC/GGANNNNNNNTTRG	0.9538	475	498	68.87	5.30	43.42
		GGANNNNNNNTTRG	CYAANNNNNNNTCC	CYAANNNNNNNTCC/GGANNNNNNNTTRG	0.9237	460	498	69.93	5.00	44.50
NCTC5658	464	GAAGNNNNNTAC	GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	0.9963	270	271	218.36	6.58	143.10
		GTANNNNNCTTC	GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	0.9889	268	271	202.19	5.63	149.00
		GACNNNNNNTTYG	CRAANNNNNNNGTC	GACNNNNNNTTYG/CRAANNNNNNNGTC	0.9907	426	430	202.02	5.01	146.60
		CRAANNNNNNNGTC	GACNNNNNNTTYG	GACNNNNNNTTYG/CRAANNNNNNNGTC	0.9349	402	430	159.72	4.10	142.48
		CCAYNNNNNNRTC	GAYNNNNNNRTGG	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9775	826	845	193.00	4.77	147.86
		GAYNNNNNNRTGG	CCAYNNNNNNRTC	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9763	825	845	188.85	4.55	148.51
NCTC5663	350	GAAGNNNNNNTGT	ACANNNNNNCTTC	GAAGNNNNNNTGT/ACANNNNNNCTTC	0.9962	262	263	190.68	5.62	133.94
		ACANNNNNNCTTC	GAAGNNNNNNTGT	GAAGNNNNNNTGT/ACANNNNNNCTTC	0.9848	259	263	176.39	5.02	137.35
		ATCNNNNNNCTC	GAGNNNNNNGAT	ATCNNNNNNCTC/GAGNNNNNNGAT	0.9960	744	747	190.93	5.07	135.77
		GAGNNNNNNGAT	ATCNNNNNNCTC	ATCNNNNNNCTC/GAGNNNNNNGAT	0.9946	743	747	195.57	5.99	136.36
NCTC6134	25	TACBNNNNNRTGG	CCAYNNNNNVGTA	TACBNNNNNRTGG/CCAYNNNNNVGTA	0.9966	290	291	97.11	4.91	66.99
		CCAYNNNNNVGTA	TACBNNNNNRTGG	TACBNNNNNRTGG/CCAYNNNNNVGTA	0.9210	268	291	103.18	4.77	67.71
		ATCNNNNNNRTGG	CCAYNNNNNGAT	ATCNNNNNNRTGG/CCAYNNNNNGAT	0.9874	626	634	93.25	5.03	58.23
		CCAYNNNNNGAT	ATCNNNNNNRTGG	ATCNNNNNNRTGG/CCAYNNNNNGAT	0.9164	581	634	85.12	4.25	59.78
		TCTANNNNNNNRTTC	GAAYNNNNNNNTAGA	TCTANNNNNNNRTTC/GAAYNNNNNNNTAGA	0.9735	257	264	92.04	5.43	60.56
		GAAYNNNNNNNTAGA	TCTANNNNNNNRTTC	TCTANNNNNNNRTTC/GAAYNNNNNNNTAGA	0.9659	255	264	94.97	4.86	59.40
NCTC6135	1021	GRAYNNNNNNNTYTC	GARANNNNNNNRTYC	GRAYNNNNNNNTYTC/GARANNNNNNNRTYC	0.9816	639	651	196.23	5.00	134.99
		GARANNNNNNNRTYC	GRAYNNNNNNNTYTC	GRAYNNNNNNNTYTC/GARANNNNNNNRTYC	0.9770	636	651	172.83	4.74	134.74
		GGARNNNNNNNTYTC	GARANNNNNNNYTCC	GGARNNNNNNNTYTC/GARANNNNNNNYTCC	0.9807	203	207	185.17	5.22	134.87
		GARANNNNNNNYTCC	GGARNNNNNNNTYTC	GGARNNNNNNNTYTC/GARANNNNNNNYTCC	0.9710	201	207	172.01	5.00	132.91
NCTC6136	9	GAAGNNNNNNNTTRG	CYAANNNNNNNCTTC	GAAGNNNNNNNTTRG/CYAANNNNNNNCTTC	1.0000	231	231	209.79	6.61	136.69
		CYAANNNNNNNCTTC	GAAGNNNNNNNTTRG	GAAGNNNNNNNTTRG/CYAANNNNNNNCTTC	0.9827	227	231	197.27	5.11	138.19
		TCTANNNNNNNTTAA	TTAANNNNNNNTAGA	TCTANNNNNNNTTAA/TTAANNNNNNNTAGA	0.9931	433	436	202.74	6.11	138.85
		TTAANNNNNNNTAGA	TCTANNNNNNNTTAA	TCTANNNNNNNTTAA/TTAANNNNNNNTAGA	0.9931	433	436	191.66	5.73	138.59
NCTC6137	10	CCAYNNNNNRTTT	AAAYNNNNNRTGG	CCAYNNNNNRTTT/AAAYNNNNNRTGG	0.9831	640	651	170.71	5.21	119.42
		AAAYNNNNNRTGG	CCAYNNNNNRTTT	CCAYNNNNNRTTT/AAAYNNNNNRTGG	0.9386	611	651	143.49	4.39	114.10
		CCAYNNNNNYTTYG		CCAYNNNNNYTTYG	0.9821	165	168	176.52	5.35	119.47
NCTC6507	10	GACNNNNNNTAG	CTANNNNNNGTC	GACNNNNNNTAG/CTANNNNNNGTC	0.9888	531	537	132.12	4.56	94.44
		CTANNNNNNGTC	GACNNNNNNTAG	GACNNNNNNTAG/CTANNNNNNGTC	0.9367	503	537	119.88	4.35	94.15
NCTC6571	30	ATCNNNNNCTWC	GWAGNNNNNGAT	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9983	1196	1198	221.51	5.30	157.63
		GWAGNNNNNGAT	ATCNNNNNCTWC	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9925	1189	1198	221.56	5.67	155.49
		CAGNNNNNRAAT	ATTYNNNNNCTG	CAGNNNNNRAAT/ATTYNNNNNCTG	0.9981	1079	1081	256.07	7.93	160.96
		ATTYNNNNNCTG	CAGNNNNNRAAT	CAGNNNNNRAAT/ATTYNNNNNCTG	0.9963	1077	1081	229.99	5.63	159.00
		GGANNNNNNNTCG	CGANNNNNNNTCC	GGANNNNNNNTCG/CGANNNNNNNTCC	0.9887	349	353	208.75	4.56	160.54
		CGANNNNNNNTCC	GGANNNNNNNTCG	GGANNNNNNNTCG/CGANNNNNNNTCC	0.9773	345	353	193.30	4.42	162.02
NCTC6966	890	TCAYNNNNNNTCC	GGANNNNNNRTGA	TCAYNNNNNNTCC/GGANNNNNNRTGA	0.9719	450	463	104.68	4.94	70.62
		GGANNNNNNRTGA	TCAYNNNNNNTCC	TCAYNNNNNNTCC/GGANNNNNNRTGA	0.9093	421	463	86.37	3.82	69.92
		GWAGNNNNNRTKC	GMAYNNNNNCTWC	GWAGNNNNNRTKC/GMAYNNNNNCTWC	0.8225	417	507	104.92	6.04	67.35

		GMAYNNNNNCTWC	GWAGNNNNNRTKC	GWAGNNNNNRTKC/GMAYNNNNNCTWC	0.8028	407	507	100.67	4.67	68.39
NCTC7121	97	GAAGNNNNNTAC	GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	1.0000	268	268	123.73	6.53	76.62
		GTANNNNNCTTC	GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	0.9963	267	268	121.61	5.71	80.43
		CCAYNNNNNNTTYG		CCAYNNNNNNTTYG	0.9817	321	327	118.39	5.06	79.85
		CCAYNNNNNNRTC	GAYNNNNNNRTGG	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9761	817	837	115.98	4.76	79.79
		GAYNNNNNNRTGG	CCAYNNNNNNRTC	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9582	802	837	111.40	4.52	79.73
NCTC7361	30	CAGNNNNNRAAT	ATTYNNNNNCTG	CAGNNNNNRAAT/ATTYNNNNNCTG	0.9954	1086	1091	248.81	7.82	155.96
		ATTYNNNNNCTG	CAGNNNNNRAAT	CAGNNNNNRAAT/ATTYNNNNNCTG	0.9908	1081	1091	222.29	5.61	154.75
		ATCNNNNNCTWC	GWAGNNNNNGAT	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9950	1201	1207	216.60	5.26	155.19
		GWAGNNNNNGAT	ATCNNNNNCTWC	ATCNNNNNCTWC/GWAGNNNNNGAT	0.9892	1194	1207	215.04	5.68	152.81
		CGANNNNNNNTCC	GGANNNNNNNTCG	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9858	348	353	181.76	4.35	155.29
		GGANNNNNNNTCG	CGANNNNNNNTCC	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9858	348	353	200.70	4.55	155.41
NCTC7414	121	AGGNNNNNNTCC	GGANNNNNNCCT	AGGNNNNNNTCC/GGANNNNNNCCT	0.9919	491	495	148.17	5.97	98.18
		GGANNNNNNCCT	AGGNNNNNNTCC	AGGNNNNNNTCC/GGANNNNNNCCT	0.9838	487	495	145.50	5.01	99.20
		AGGVNNNNYACC		AGGVNNNNYACC	0.6298	148	235	93.70	4.53	99.28
NCTC7415	5	TACBNNNNNRTGG	CCAYNNNNNVGTA	TACBNNNNNRTGG/CCAYNNNNNVGTA	1.0000	288	288	202.32	4.75	152.88
		CCAYNNNNNVGTA	TACBNNNNNRTGG	TACBNNNNNRTGG/CCAYNNNNNVGTA	0.9479	273	288	186.22	4.39	151.55
		ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9957	692	695	217.59	5.37	150.24
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9942	691	695	217.50	6.39	150.06
NCTC7445	30	HATCNNNNNCTWC		HATCNNNNNCTWC	0.9991	1059	1060	192.88	5.41	133.56
		GWAGNNNNNGAT		GWAGNNNNNGAT	0.9932	1169	1177	193.04	5.83	132.40
		CGANNNNNNNTCC	GGANNNNNNNTCG	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9886	346	350	168.27	4.47	137.63
		GGANNNNNNNTCG	CGANNNNNNNTCC	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9771	342	350	183.58	4.72	137.54
NCTC7446	30	HATCNNNNNCTWC		HATCNNNNNCTWC	0.9962	1056	1060	128.37	5.24	86.63
		GWAGNNNNNGAT		GWAGNNNNNGAT	0.9907	1166	1177	126.73	5.65	86.42
		CGANNNNNNNTCC	GGANNNNNNNTCG	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9771	341	349	111.89	4.27	88.58
		GGANNNNNNNTCG	CGANNNNNNNTCC	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9742	340	349	118.25	4.55	87.85
NCTC7485	351	CAAGNNNNNNTARC	GYTANNNNNNCTTG	CAAGNNNNNNTARC/GYTANNNNNNCTTG	1.0000	190	190	225.58	6.06	150.38
		GYTANNNNNNCTTG	CAAGNNNNNNTARC	CAAGNNNNNNTARC/GYTANNNNNNCTTG	0.9895	188	190	207.78	5.29	155.28
NCTC7712	136	GACNNNNNNTAYG	CRTANNNNNNGTC	GACNNNNNNTAYG/CRTANNNNNNGTC	0.9894	375	379	201.48	4.82	145.31
		CRTANNNNNNGTC	GACNNNNNNTAYG	GACNNNNNNTAYG/CRTANNNNNNGTC	0.9551	362	379	175.01	4.50	145.81
		CCAYNNNNNTTAA	TTAANNNNNRTGG	CCAYNNNNNTTAA/TTAANNNNNRTGG	0.9860	493	500	203.75	5.12	146.12
		TTAANNNNNRTGG	CCAYNNNNNTTAA	CCAYNNNNNTTAA/TTAANNNNNRTGG	0.9620	481	500	187.73	4.97	147.78
NCTC7791	121	GGANNNNNNCCT	GGANNNNNNCCT		0.9879276	491	497	102.477	4.98297	68.18941
		AGGNNNNNNWCC	AGGNNNNNNWCC		0.6777963	812	1198	91.9741	5.38036	67.42857
NCTC7856	351	TCAYNNNNNNTCC	GGANNNNNNRTGA	TCAYNNNNNNTCC/GGANNNNNNRTGA	0.9654	447	463	105.99	4.91	72.65
		GGANNNNNNRTGA	TCAYNNNNNNTCC	TCAYNNNNNNTCC/GGANNNNNNRTGA	0.8834	409	463	90.70	3.78	73.12
		GWAGNNNNNRTKC	GMAYNNNNNCTWC	GWAGNNNNNRTKC/GMAYNNNNNCTWC	0.8264	419	507	112.07	5.99	72.25
		GMAYNNNNNCTWC	GWAGNNNNNRTKC	GWAGNNNNNRTKC/GMAYNNNNNCTWC	0.8047	408	507	107.49	4.63	73.22
NCTC7972	10	GACNNNNNNTAG	CTANNNNNNGTC	GACNNNNNNTAG/CTANNNNNNGTC	0.9694	539	556	163.16	4.74	116.34
		CTANNNNNNGTC	GACNNNNNNTAG	GACNNNNNNTAG/CTANNNNNNGTC	0.9119	507	556	146.42	4.42	116.81
		CYAANNNNNNNTCC	GGANNNNNNNTTRG	CYAANNNNNNNTCC/GGANNNNNNNTTRG	0.9451	499	528	164.20	5.26	114.40

		GGANNNNNNNTTRG	CYAANNNNNNNTCC	CYAANNNNNNNTCC/GGANNNNNNNTTRG	0.9318	492	528	157.04	4.77	115.36
NCTC7988	131	CYAANNNNNNNTCC	GGANNNNNNNTTRG	CYAANNNNNNNTCC/GGANNNNNNNTTRG	0.9898	483	488	173.40	5.37	120.05
		GGANNNNNNNTTRG	CYAANNNNNNNTCC	CYAANNNNNNNTCC/GGANNNNNNNTTRG	0.9672	472	488	167.14	4.94	120.51
		CAGNNNNNRTGA	TCAYNNNNNCTG	CAGNNNNNRTGA/TCAYNNNNNCTG	0.9780	534	546	172.80	5.70	121.11
		TCAYNNNNNCTG	CAGNNNNNRTGA	CAGNNNNNRTGA/TCAYNNNNNCTG	0.9670	528	546	159.22	4.63	121.29
NCTC8004	254	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9972	708	710	208.85	5.33	143.55
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9930	705	710	211.35	6.38	143.89
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9633	473	491	177.28	4.61	146.91
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9450	464	491	183.46	4.45	146.40
NCTC8178	254	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	0.9958	713	716	94.02	5.42	55.99
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9902	709	716	91.24	6.45	55.81
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9573	471	492	82.89	4.79	55.31
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9329	459	492	86.79	4.58	55.91
NCTC8317	25	TCTANNNNNNRTTC	GAAYNNNNNNNTAGA	TCTANNNNNNRTTC/GAAYNNNNNNNTAGA	1.0000	257	257	83.33	5.86	52.85
		GAAYNNNNNNNTAGA	TCTANNNNNNRTTC	TCTANNNNNNRTTC/GAAYNNNNNNNTAGA	0.9767	251	257	77.87	5.29	49.80
		ATCNNNNNRTGG	CCAYNNNNNGAT	ATCNNNNNRTGG/CCAYNNNNNGAT	1.0000	615	615	83.69	5.08	52.15
		CCAYNNNNNGAT	ATCNNNNNRTGG	ATCNNNNNRTGG/CCAYNNNNNGAT	0.9707	597	615	82.71	4.64	52.72
		TACBNNNNNRTGG	CCAYNNNNNVGTA	TACBNNNNNRTGG/CCAYNNNNNVGTA	1.0000	284	284	83.92	5.45	53.53
		CCAYNNNNNVGTA	TACBNNNNNRTGG	TACBNNNNNRTGG/CCAYNNNNNVGTA	0.9824	279	284	85.61	4.84	53.78
NCTC8325	8	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	1.0000	685	685	190.15	4.96	137.22
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9956	682	685	191.26	5.95	136.70
		ACANNNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9566	463	484	161.57	4.38	138.30
		CCAYNNNNNNTGT	ACANNNNNNRTGG	ACANNNNNNRTGG/CCAYNNNNNNTGT	0.9421	456	484	168.43	4.24	138.49
NCTC8399	97	GAAGNNNNNTAC	GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	1.0000	268	268	209.68	6.98	134.58
		GTANNNNNCTTC	GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	0.9963	267	268	199.65	6.00	139.29
		GACNNNNNNTTYG	CRAANNNNNNGTC	GACNNNNNNTTYG/CRAANNNNNNGTC	0.9955	440	442	200.50	5.14	140.67
		CRAANNNNNNGTC	GACNNNNNNTTYG	GACNNNNNNTTYG/CRAANNNNNNGTC	0.9231	408	442	158.00	4.20	136.95
		CCAYNNNNNNRTC	GAYNNNNNNRTGG	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9848	841	854	186.34	4.85	139.64
		GAYNNNNNNRTGG	CCAYNNNNNNRTC	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9731	831	854	184.87	4.64	139.67
NCTC8507	30	HATCNNNNNCTWC		HATCNNNNNCTWC	0.9953	1054	1059	95.83	5.29	60.04
		GWAGNNNNNGAT		GWAGNNNNNGAT	0.9805	1157	1180	92.69	5.67	59.37
		CGANNNNNNNTCC	GGANNNNNNNTCG	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9773	344	352	84.47	4.33	61.39
		GGANNNNNNNTCG	CGANNNNNNNTCC	CGANNNNNNNTCC/GGANNNNNNNTCG	0.9659	340	352	90.66	4.58	62.18
NCTC8530	30	HATCNNNNNCTWC		HATCNNNNNCTWC	0.9972	1063	1066	184.78	5.22	129.60
		GWAGNNNNNGAT		GWAGNNNNNGAT	0.9916	1176	1186	184.36	5.61	127.39
		GGANNNNNNNTCG	CGANNNNNNNTCC	GGANNNNNNNTCG/CGANNNNNNNTCC	0.9859	349	354	172.26	4.47	134.96
		CGANNNNNNNTCC	GGANNNNNNNTCG	GGANNNNNNNTCG/CGANNNNNNNTCC	0.9718	344	354	162.38	4.26	135.65
NCTC8531	121	AGGNNNNNNTCC	GGANNNNNNCCT	AGGNNNNNNTCC/GGANNNNNNCCT	0.9877	482	488	194.47	5.67	138.36
		GGANNNNNNCCT	AGGNNNNNNTCC	AGGNNNNNNTCC/GGANNNNNNCCT	0.9775	477	488	187.23	4.71	139.95
		GACNNNNNNTAYG	CRTANNNNNNGTC	GACNNNNNNTAYG/CRTANNNNNNGTC	0.9867	370	375	196.11	4.67	143.53
		CRTANNNNNNGTC	GACNNNNNNTAYG	GACNNNNNNTAYG/CRTANNNNNNGTC	0.9493	356	375	162.59	4.22	143.16
NCTC8723	9	GAAGNNNNNNTTRG	CYAANNNNNNCTTC	GAAGNNNNNNTTRG/CYAANNNNNNCTTC	1.0000	231	231	200.33	6.47	133.09
		CYAANNNNNNCTTC	GAAGNNNNNNTTRG	GAAGNNNNNNTTRG/CYAANNNNNNCTTC	0.9784	226	231	184.12	4.88	135.19

		TTAANNNNNNTAGA	TCTANNNNNNTTAA	TTAANNNNNNTAGA/TCTANNNNNNTTAA	0.9930	427	430	181.66	5.48	136.70
		TCTANNNNNNTTAA	TTAANNNNNNTAGA	TTAANNNNNNTAGA/TCTANNNNNNTTAA	0.9907	426	430	193.12	5.94	136.50
NCTC8725	9	GAAGNNNNNNTTRG	CYAANNNNNNCTTC	GAAGNNNNNNTTRG/CYAANNNNNNCTTC	1.0000	232	232	201.45	6.89	131.40
		CYAANNNNNNCTTC	GAAGNNNNNNTTRG	GAAGNNNNNNTTRG/CYAANNNNNNCTTC	0.9828	228	232	188.85	5.18	131.77
		TTAANNNNNNTAGA	TCTANNNNNNTTAA	TTAANNNNNNTAGA/TCTANNNNNNTTAA	0.9977	433	434	187.60	5.83	135.21
		TCTANNNNNNTTAA	TTAANNNNNNTAGA	TTAANNNNNNTAGA/TCTANNNNNNTTAA	0.9908	430	434	198.80	6.25	134.74
NCTC8726	198	CAACNNNNNNTAYG	CRTANNNNNNGTTG	CAACNNNNNNTAYG/CRTANNNNNNGTTG	1.0000	181	181	188.86	4.95	122.22
		CRTANNNNNNGTTG	CAACNNNNNNTAYG	CAACNNNNNNTAYG/CRTANNNNNNGTTG	0.9171	166	181	147.52	4.38	123.80
		ACCNNNNNRTGA	TCAYNNNNNGGT	ACCNNNNNRTGA/TCAYNNNNNGGT	1.0000	619	619	177.15	4.60	123.81
		TCAYNNNNNGGT	ACCNNNNNRTGA	ACCNNNNNRTGA/TCAYNNNNNGGT	0.9435	584	619	154.60	4.35	124.79
NCTC8765	9	GAAGNNNNNNTTRG	CYAANNNNNNCTTC	GAAGNNNNNNTTRG/CYAANNNNNNCTTC	1.0000	227	227	169.77	6.74	109.37
		CYAANNNNNNCTTC	GAAGNNNNNNTTRG	GAAGNNNNNNTTRG/CYAANNNNNNCTTC	0.9824	223	227	162.26	5.14	110.09
		TTAANNNNNNTAGA	TCTANNNNNNTTAA	TTAANNNNNNTAGA/TCTANNNNNNTTAA	0.9906	421	425	158.71	5.88	110.02
		TCTANNNNNNTTAA	TTAANNNNNNTAGA	TTAANNNNNNTAGA/TCTANNNNNNTTAA	0.9882	420	425	168.16	6.31	110.90
NCTC9369	8	ATCNNNNNCCT	AGGNNNNNGAT	ATCNNNNNCCT/AGGNNNNNGAT	1.0000	683	683	219.00	5.38	152.32
		AGGNNNNNGAT	ATCNNNNNCCT	ATCNNNNNCCT/AGGNNNNNGAT	0.9956	680	683	222.58	6.44	152.11
		ACANNNNNRTGG	CCAYNNNNNNTGT	ACANNNNNRTGG/CCAYNNNNNNTGT	0.9650	469	486	187.06	4.71	154.64
		CCAYNNNNNNTGT	ACANNNNNRTGG	ACANNNNNRTGG/CCAYNNNNNNTGT	0.9506	462	486	194.58	4.50	155.04
NCTC9546	692	GACNNNNNNTGG		GACNNNNNNTGG	0.9747	424	435	122.93	4.47	85.40
		TTTANNNNNRTGG	CCAYNNNNNNNTAAA	TTTANNNNNRTGG/CCAYNNNNNNNTAAA	0.9498	435	458	121.29	5.13	86.20
		CCAYNNNNNNNTAAA	TTTANNNNNRTGG	TTTANNNNNRTGG/CCAYNNNNNNNTAAA	0.9279	425	458	126.33	4.72	85.31
NCTC9547	97	GAAGNNNNTAC	GTANNNNCTTC	GAAGNNNNTAC/GTANNNNCTTC	1.0000	256	256	239.88	6.65	157.23
		GTANNNNCTTC	GAAGNNNNTAC	GAAGNNNNTAC/GTANNNNCTTC	0.9961	255	256	229.71	5.86	161.35
		GACNNNNNNTTYG	CRAANNNNNNGTC	GACNNNNNNTTYG/CRAANNNNNNGTC	0.9953	425	427	223.87	5.00	162.99
		CRAANNNNNNGTC	GACNNNNNNTTYG	GACNNNNNNTTYG/CRAANNNNNNGTC	0.9344	399	427	176.44	4.08	160.55
		GAYNNNNNNRTGG	CCAYNNNNNNRTC	GAYNNNNNNRTGG/CCAYNNNNNNRTC	0.9804	799	815	207.65	4.59	161.14
		CCAYNNNNNNRTC	GAYNNNNNNRTGG	GAYNNNNNNRTGG/CCAYNNNNNNRTC	0.9791	798	815	211.15	4.82	161.53
NCTC9551	97	GTANNNNCTTC	GAAGNNNNTAC	GTANNNNCTTC/GAAGNNNNTAC	0.9961	258	259	74.64	6.04	44.61
		GAAGNNNNTAC	GTANNNNCTTC	GTANNNNCTTC/GAAGNNNNTAC	0.9923	257	259	76.20	7.25	43.42
		GACNNNNNNTTYG	CRAANNNNNNGTC	GACNNNNNNTTYG/CRAANNNNNNGTC	0.9840	430	437	70.63	5.24	44.50
		CRAANNNNNNGTC	GACNNNNNNTTYG	GACNNNNNNTTYG/CRAANNNNNNGTC	0.8833	386	437	61.47	4.40	42.91
		CCAYNNNNNNRTC	GAYNNNNNNRTGG	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9689	809	835	75.59	5.05	44.91
		GAYNNNNNNRTGG	CCAYNNNNNNRTC	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9389	784	835	69.73	4.83	45.03
NCTC9552	97	GAAGNNNNTAC	GTANNNNCTTC	GAAGNNNNTAC/GTANNNNCTTC	1.0000	257	257	150.24	7.05	95.70
		GTANNNNCTTC	GAAGNNNNTAC	GAAGNNNNTAC/GTANNNNCTTC	0.9961	256	257	144.28	6.07	98.05
		GACNNNNNNTTYG	CRAANNNNNNGTC	GACNNNNNNTTYG/CRAANNNNNNGTC	0.9954	430	432	143.86	5.21	98.80
		CRAANNNNNNGTC	GACNNNNNNTTYG	GACNNNNNNTTYG/CRAANNNNNNGTC	0.9306	402	432	116.98	4.25	96.11
		CCAYNNNNNNRTC	GAYNNNNNNRTGG	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9794	809	826	139.31	4.93	99.02
		GAYNNNNNNRTGG	CCAYNNNNNNRTC	CCAYNNNNNNRTC/GAYNNNNNNRTGG	0.9637	796	826	135.50	4.70	98.81
NCTC9555	133	CYAANNNNNNNTCC	GGANNNNNNNTTRG	CYAANNNNNNNTCC/GGANNNNNNNTTRG	0.9898	484	489	126.98	5.37	85.59
		GGANNNNNNNTTRG	CYAANNNNNNNTCC	CYAANNNNNNNTCC/GGANNNNNNNTTRG	0.9673	473	489	123.80	5.00	86.15
		CAGNNNNNRTGA	TCAYNNNNNCTG	CAGNNNNNRTGA/TCAYNNNNNCTG	0.9725	530	545	131.68	5.88	87.93
		TCAYNNNNNCTG	CAGNNNNNRTGA	CAGNNNNNRTGA/TCAYNNNNNCTG	0.9651	526	545	122.94	4.76	88.17

NCTC9556	692	GACNNNNNNTGG	CCANNNNNNGTC	GACNNNNNNTGG/CCANNNNNNGTC	0.9770	424	434	173.00	4.49	127.33
		CCANNNNNNGTC	GACNNNNNNTGG	GACNNNNNNTGG/CCANNNNNNGTC	0.9309	404	434	152.83	4.11	127.80
		TTTANNNNNNRTGG	CCAYNNNNNNTAAA	TTTANNNNNNRTGG/CCAYNNNNNNTAAA	0.9756	440	451	170.59	5.20	128.31
		CCAYNNNNNNTAAA	TTTANNNNNNRTGG	TTTANNNNNNRTGG/CCAYNNNNNNTAAA	0.9601	433	451	174.24	4.71	127.92
NCTC9611	692	GACNNNNNNTGG	CCANNNNNNGTC	GACNNNNNNTGG/CCANNNNNNGTC	0.9768	421	431	171.54	4.55	124.20
		CCANNNNNNGTC	GACNNNNNNTGG	GACNNNNNNTGG/CCANNNNNNGTC	0.9258	399	431	146.35	4.09	124.09
		TTTANNNNNNRTGG	CCAYNNNNNNTAAA	TTTANNNNNNRTGG/CCAYNNNNNNTAAA	0.9711	437	450	165.56	5.19	125.83
		CCAYNNNNNNTAAA	TTTANNNNNNRTGG	TTTANNNNNNRTGG/CCAYNNNNNNTAAA	0.9511	428	450	170.18	4.73	125.28
NCTC9612	692	GACNNNNNNTGG	CCANNNNNNGTC	GACNNNNNNTGG/CCANNNNNNGTC	0.9859	421	427	200.87	4.51	150.31
		CCANNNNNNGTC	GACNNNNNNTGG	GACNNNNNNTGG/CCANNNNNNGTC	0.9344	399	427	174.63	4.14	150.10
		TTTANNNNNNRTGG	CCAYNNNNNNTAAA	TTTANNNNNNRTGG/CCAYNNNNNNTAAA	0.9777	439	449	195.51	5.24	151.47
		CCAYNNNNNNTAAA	TTTANNNNNNRTGG	TTTANNNNNNRTGG/CCAYNNNNNNTAAA	0.9599	431	449	202.28	4.77	151.32
NCTC9613	692	GACNNNNNNTGG	CCANNNNNNGTC	GACNNNNNNTGG/CCANNNNNNGTC	0.9772	428	438	105.61	4.31	77.54
		CCANNNNNNGTC	GACNNNNNNTGG	GACNNNNNNTGG/CCANNNNNNGTC	0.9132	400	438	100.11	4.12	77.53
		TTTANNNNNNRTGG	CCAYNNNNNNTAAA	TTTANNNNNNRTGG/CCAYNNNNNNTAAA	0.9698	449	463	111.30	5.16	82.05
		CCAYNNNNNNTAAA	TTTANNNNNNRTGG	TTTANNNNNNRTGG/CCAYNNNNNNTAAA	0.9395	435	463	118.80	4.76	83.22
NCTC9614	692	TTTANNNNNNRTGG	CCAYNNNNNNTAAA	TTTANNNNNNRTGG/CCAYNNNNNNTAAA	0.9733	438	450	160.99	5.10	123.80
		CCAYNNNNNNTAAA	TTTANNNNNNRTGG	TTTANNNNNNRTGG/CCAYNNNNNNTAAA	0.9444	425	450	165.19	4.64	123.04
		GACNNNNNNTGG	CCANNNNNNGTC	GACNNNNNNTGG/CCANNNNNNGTC	0.9721	418	430	169.47	4.42	124.58
		CCANNNNNNGTC	GACNNNNNNTGG	GACNNNNNNTGG/CCANNNNNNGTC	0.9302	400	430	144.26	3.99	125.02
NCTC9752	97	CCAYNNNNNNNRTC	GAYNNNNNNNRTGG	CCAYNNNNNNNRTC/GAYNNNNNNNRTGG	0.9815	797	812	197.96	4.72	154.23
		GAYNNNNNNNRTGG	CCAYNNNNNNNRTC	CCAYNNNNNNNRTC/GAYNNNNNNNRTGG	0.9729	790	812	197.26	4.45	155.12
		CCAYNNNNNNNTTYG	CRAANNNNNNNRTGG	CCAYNNNNNNNTTYG/CRAANNNNNNNRTGG	0.9772	300	307	203.71	4.91	154.94
		CRAANNNNNNNRTGG	CCAYNNNNNNNTTYG	CCAYNNNNNNNTTYG/CRAANNNNNNNRTGG	0.9479	291	307	171.11	4.43	148.45

8.2 CHAPTER 4 APPENDIX

Table 8.2 | Chimeric Region 1 (CH1) starting from the origin of replication (without SCC*mec*)

Gene/Product	CD140392 Locus	CD140901 Locus	CD140657 Locus	CD141496 Locus	CD150713 Locus	CD150916 Locus	CD140400 Locus	CD140638 Locus	CD140866 Locus
dnaA	58275_C01_00001	58275_A01_01345	58275_B01_00001	58366_B01_00001	58366_D01_00001	58366_C01_00004	58275_D01_00001	58275_E01_00001	58366_A01_00001
dnaN	58275_C01_00002	58275_A01_01346	58275_B01_00002	58366_B01_00002	58366_D01_00002	58366_C01_00005	58275_D01_00002	58275_E01_00002	58366_A01_00002
YaaA	58275_C01_00003	58275_A01_01347	58275_B01_00003	58366_B01_00003	58366_D01_00003	58366_C01_00006	58275_D01_00003	58275_E01_00003	58366_A01_00003
recF	58275_C01_00004	58275_A01_01348	58275_B01_00004	58366_B01_00004	58366_D01_00004	58366_C01_00007	58275_D01_00004	58275_E01_00004	58366_A01_00004
gyrB	58275_C01_00005	58275_A01_01349	58275_B01_00005	58366_B01_00005	58366_D01_00005	58366_C01_00008	58275_D01_00005	58275_E01_00005	58366_A01_00005
gryA	58275_C01_00006	58275_A01_01350	58275_B01_00006	58366_B01_00006	58366_D01_00006	58366_C01_00009	58275_D01_00006	58275_E01_00006	58366_A01_00006
nnrD	58275_C01_00007	58275_A01_01351	58275_B01_00007	58366_B01_00007	58366_D01_00007	58366_C01_00010	58275_D01_00007	58275_E01_00007	58366_A01_00007
hutH	58275_C01_00008	58275_A01_01352	58275_B01_00008	58366_B01_00008	58366_D01_00008	58366_C01_00011	58275_D01_00008	58275_E01_00008	58366_A01_00008
ncRNA	58275_C01_00009	58275_A01_01353	58275_B01_00009	58366_B01_00009	58366_D01_00009	58366_C01_00012	58275_D01_00009	58275_E01_00009	58366_A01_00009
serS	58275_C01_00010	58275_A01_01354	58275_B01_00010	58366_B01_00010	58366_D01_00010	58366_C01_00013	58275_D01_00010	58275_E01_00010	58366_A01_00010
yagZ	58275_C01_00011	58275_A01_01355	58275_B01_00011	58366_B01_00011	58366_D01_00011	58366_C01_00014	58275_D01_00011	58275_E01_00011	58366_A01_00011
AA transporter	58275_C01_00012	58275_A01_01356	58275_B01_00012	58366_B01_00012	58366_D01_00012	58366_C01_00015	58275_D01_00012	58275_E01_00012	58366_A01_00012
ncRNA	58275_C01_00013	58275_A01_01357	58275_B01_00013	58366_B01_00013	58366_D01_00013	58366_C01_00016	58275_D01_00013	58275_E01_00013	58366_A01_00013
metX	58275_C01_00014	58275_A01_01358	58275_B01_00014	58366_B01_00014	58366_D01_00014	58366_C01_00017	58275_D01_00014	58275_E01_00014	58366_A01_00014
hypothetical protein	58275_C01_00015	58275_A01_01359	58275_B01_00015	58366_B01_00015	58366_D01_00015	58366_C01_00018	58275_D01_00015	58275_E01_00015	58366_A01_00015
phosphoesterase	58275_C01_00016	58275_A01_01360	58275_B01_00016	58366_B01_00016	58366_D01_00016	58366_C01_00019	58275_D01_00016	58275_E01_00016	58366_A01_00016
rplI	58275_C01_00017	58275_A01_01361	58275_B01_00017	58366_B01_00017	58366_D01_00017	58366_C01_00020	58275_D01_00017	58275_E01_00017	58366_A01_00017
dnaC	58275_C01_00018	58275_A01_01362	58275_B01_00018	58366_B01_00018	58366_D01_00018	58366_C01_00021	58275_D01_00018	58275_E01_00018	58366_A01_00018
purA	58275_C01_00019	58275_A01_01363	58275_B01_00019	58366_B01_00019	58366_D01_00019	58366_C01_00022	58275_D01_00019	58275_E01_00019	58366_A01_00019
tRNA	58275_C01_00020	58275_A01_01364	58275_B01_00020	58366_B01_00020	58366_D01_00020	58366_C01_00023	58275_D01_00020	58275_E01_00020	58366_A01_00020
tRNA	58275_C01_00021	58275_A01_01365	58275_B01_00021	58366_B01_00021	58366_D01_00021	58366_C01_00024	58275_D01_00021	58275_E01_00021	58366_A01_00021
vicR	58275_C01_00022	58275_A01_01366	58275_B01_00022	58366_B01_00022	58366_D01_00022	58366_C01_00025	58275_D01_00022	58275_E01_00022	58366_A01_00022
walK	58275_C01_00023	58275_A01_01367	58275_B01_00023	58366_B01_00023	58366_D01_00023	58366_C01_00026	58275_D01_00023	58275_E01_00023	58366_A01_00023
yycH	58275_C01_00024	58275_A01_01368	58275_B01_00024	58366_B01_00024	58366_D01_00024	58366_C01_00027	58275_D01_00024	58275_E01_00024	58366_A01_00024
yycI	58275_C01_00025	58275_A01_01369	58275_B01_00025	58366_B01_00025	58366_D01_00025	58366_C01_00028	58275_D01_00025	58275_E01_00025	58366_A01_00025
ZnZn-dependent hydrolase	58275_C01_00026	58275_A01_01370	58275_B01_00026	58366_B01_00026	58366_D01_00026	58366_C01_00029	58275_D01_00026	58275_E01_00026	58366_A01_00026
ykfN_1/ushA	58275_C01_00027	58275_A01_01371	58275_B01_00027	58366_B01_00027	58366_D01_00027	58366_C01_00030	58275_D01_00027	58275_E01_00027	58366_A01_00027
orfX	58275_C01_00028	58275_A01_01372	58275_B01_00028	58366_B01_00028	58366_D01_00028	58366_C01_00031	58275_D01_00028	58275_E01_00028	58366_A01_00028

SCC*mec*									
plc	58275_C01_00086	58275_A01_01434	58275_B01_00111	58366_B01_00093	58366_D01_00093	58366_C01_00095	58275_D01_00093	58275_E01_00092	58366_A01_00092
tandem lipoprotein	58275_C01_00087	58275_A01_01435	58275_B01_00112	58366_B01_00094	58366_D01_00094	58366_C01_00096	58275_D01_00094	58275_E01_00093	58366_A01_00093
tandem lipoprotein	58275_C01_00088	58275_A01_01436	58275_B01_00113	58366_B01_00095	58366_D01_00095	58366_C01_00097	58275_D01_00095	58275_E01_00094	58366_A01_00094
tandem lipoprotein	58275_C01_00089	58275_A01_01437		58366_B01_00096					
putative cytosolic protein	58275_C01_00090	58275_A01_01438	58275_B01_00114	58366_B01_00097					
putative cytosolic protein	58275_C01_00091	58275_A01_01439	58275_B01_00115	58366_B01_00098					
putative cytosolic protein	58275_C01_00092	58275_A01_01440	58275_B01_00116	58366_B01_00099					
btr	58275_C01_00093	58275_A01_01441	58275_B01_00117	58366_B01_00100	58366_D01_00096	58366_C01_00098	58275_D01_00096	58275_E01_00095	58366_A01_00095
yxeP_1	58275_C01_00094	58275_A01_01442	58275_B01_00118	58366_B01_00101	58366_D01_00097	58366_C01_00099	58275_D01_00097	58275_E01_00096	58366_A01_00096
norB_1	58275_C01_00095	58275_A01_01443	58275_B01_00119	58366_B01_00102	58366_D01_00098	58366_C01_00100	58275_D01_00098	58275_E01_00097	58366_A01_00097
Na/Pi cotransporter	58275_C01_00096	58275_A01_01444	58275_B01_00120	58366_B01_00103	58366_D01_00099	58366_C01_00101	58275_D01_00099	58275_E01_00098	58366_A01_00098
myosin	58275_C01_00097	58275_A01_01445	58275_B01_00121	58366_B01_00104	58366_D01_00100	58366_C01_00102	58275_D01_00100	58275_E01_00099	58366_A01_00099
hypothetical protein	58275_C01_00098	58275_A01_01446	58275_B01_00122	58366_B01_00105	58366_D01_00101	58366_C01_00103	58275_D01_00101	58275_E01_00100	58366_A01_00100
lctP_1	58275_C01_00099	58275_A01_01447	58275_B01_00123	58366_B01_00106	58366_D01_00102	58366_C01_00104	58275_D01_00102	58275_E01_00101	58366_A01_00101
spa	58275_C01_00100	58275_A01_01448	58275_B01_00124	58366_B01_00107	58366_D01_00103	58366_C01_00105	58275_D01_00103	58275_E01_00102	58366_A01_00102
sarS	58275_C01_00101	58275_A01_01449	58275_B01_00125	58366_B01_00108	58366_D01_00104	58366_C01_00106	58275_D01_00104	58275_E01_00103	58366_A01_00103
sirC_1	58275_C01_00102	58275_A01_01450	58275_B01_00126	58366_B01_00109	58366_D01_00105	58366_C01_00107	58275_D01_00105	58275_E01_00104	58366_A01_00104
sirB	58275_C01_00103	58275_A01_01451	58275_B01_00127	58366_B01_00110	58366_D01_00106	58366_C01_00108	58275_D01_00106	58275_E01_00105	58366_A01_00105
sirA	58275_C01_00104	58275_A01_01452	58275_B01_00128	58366_B01_00111	58366_D01_00107	58366_C01_00109	58275_D01_00107	58275_E01_00106	58366_A01_00106
sbnA	58275_C01_00105	58275_A01_01453	58275_B01_00129	58366_B01_00112	58366_D01_00108	58366_C01_00110	58275_D01_00108	58275_E01_00107	58366_A01_00107
sbnB	58275_C01_00106	58275_A01_01454	58275_B01_00130	58366_B01_00113	58366_D01_00109	58366_C01_00111	58275_D01_00109	58275_E01_00108	58366_A01_00108
sbnC	58275_C01_00107	58275_A01_01455	58275_B01_00131	58366_B01_00114	58366_D01_00110	58366_C01_00112	58275_D01_00110	58275_E01_00109	58366_A01_00109
sbnD	58275_C01_00108	58275_A01_01456	58275_B01_00132	58366_B01_00115	58366_D01_00111	58366_C01_00113	58275_D01_00111	58275_E01_00110	58366_A01_00110
sbnE	58275_C01_00109	58275_A01_01457	58275_B01_00133	58366_B01_00116	58366_D01_00112	58366_C01_00114	58275_D01_00112	58275_E01_00111	58366_A01_00111
iucC_1	58275_C01_00110	58275_A01_01458	58275_B01_00134	58366_B01_00117	58366_D01_00113	58366_C01_00115	58275_D01_00113	58275_E01_00112	58366_A01_00112
sbnG	58275_C01_00111	58275_A01_01459	58275_B01_00135	58366_B01_00118	58366_D01_00114	58366_C01_00116	58275_D01_00114	58275_E01_00113	58366_A01_00113
sbnH	58275_C01_00112	58275_A01_01460	58275_B01_00136	58366_B01_00119	58366_D01_00115	58366_C01_00117	58275_D01_00115	58275_E01_00114	58366_A01_00114
sbnI	58275_C01_00113	58275_A01_01461	58275_B01_00137	58366_B01_00120	58366_D01_00116	58366_C01_00118	58275_D01_00116	58275_E01_00115	58366_A01_00115
hypothetical protein					58366_D01_00117	58366_C01_00119	58275_D01_00117	58275_E01_00116	58366_A01_00116
hypothetical protein	58275_C01_00114	58275_A01_01462	58275_B01_00138	58366_B01_00121	58366_D01_00118	58366_C01_00120	58275_D01_00118	58275_E01_00117	58366_A01_00117
butA	58275_C01_00115	58275_A01_01463	58275_B01_00139	58366_B01_00122	58366_D01_00119	58366_C01_00121	58275_D01_00119	58275_E01_00118	58366_A01_00118
galE	58275_C01_00116	58275_A01_01464	58275_B01_00140	58366_B01_00123	58366_D01_00120	58366_C01_00122	58275_D01_00120	58275_E01_00119	58366_A01_00119
wcaJ	58275_C01_00117	58275_A01_01465	58275_B01_00141	58366_B01_00124	58366_D01_00121	58366_C01_00123	58275_D01_00121	58275_E01_00120	58366_A01_00120
pimB_1	58275_C01_00118	58275_A01_01466	58275_B01_00142	58366_B01_00125	58366_D01_00122	58366_C01_00124	58275_D01_00122	58275_E01_00121	58366_A01_00121

O-antigen ligase	58275_C01_00119	58275_A01_01467	58275_B01_00143	58366_B01_00126	58366_D01_00123	58366_C01_00125	58275_D01_00123	58275_E01_00122	58366_A01_00122
rfbX	58275_C01_00120	58275_A01_01468	58275_B01_00144	58366_B01_00127	58366_D01_00124	58366_C01_00126	58275_D01_00124	58275_E01_00123	58366_A01_00123
sodM	58275_C01_00121	58275_A01_01469	58275_B01_00145	58366_B01_00128	58366_D01_00125	58366_C01_00127	58275_D01_00125	58275_E01_00124	58366_A01_00124
sasD	58275_C01_00122	58275_A01_01470	58275_B01_00146	58366_B01_00129					
treR_1	58275_C01_00123	58275_A01_01471	58275_B01_00147	58366_B01_00130	58366_D01_00126	58366_C01_00128	58275_D01_00126	58275_E01_00125	58366_A01_00125
transposase IS256		58275_A01_01472							
deoD_1	58275_C01_00124	58275_A01_01473	58275_B01_00148	58366_B01_00131	58366_D01_00127	58366_C01_00129	58275_D01_00127	58275_E01_00126	58366_A01_00126
norB_2	58275_C01_00125	58275_A01_01474	58275_B01_00149	58366_B01_00132	58366_D01_00128	58366_C01_00130	58275_D01_00128	58275_E01_00127	58366_A01_00127
dra_1	58275_C01_00126	58275_A01_01475	58275_B01_00150	58366_B01_00133	58366_D01_00129	58366_C01_00131	58275_D01_00129	58275_E01_00128	58366_A01_00128
deoB	58275_C01_00127	58275_A01_01476	58275_B01_00151	58366_B01_00134	58366_D01_00130	58366_C01_00132	58275_D01_00130	58275_E01_00129	58366_A01_00129
phnE_1	58275_C01_00128	58275_A01_01477	58275_B01_00152	58366_B01_00135	58366_D01_00131	58366_C01_00133	58275_D01_00131	58275_E01_00130	58366_A01_00130
phnE_2	58275_C01_00129	58275_A01_01478	58275_B01_00153	58366_B01_00136	58366_D01_00132	58366_C01_00134	58275_D01_00132	58275_E01_00131	58366_A01_00131
phnC	58275_C01_00130	58275_A01_01479	58275_B01_00154	58366_B01_00137	58366_D01_00133	58366_C01_00135	58275_D01_00133	58275_E01_00132	58366_A01_00132
phnD	58275_C01_00131	58275_A01_01480	58275_B01_00155	58366_B01_00138	58366_D01_00134	58366_C01_00136	58275_D01_00134	58275_E01_00133	58366_A01_00133
tRNA-helicase	58275_C01_00132	58275_A01_01481	58275_B01_00156	58366_B01_00139	58366_D01_00135	58366_C01_00137	58275_D01_00135	58275_E01_00134	58366_A01_00134
yfkN_1	58275_C01_00133	58275_A01_01482	58275_B01_00157	58366_B01_00140	58366_D01_00136	58366_C01_00138	58275_D01_00136	58275_E01_00135	58366_A01_00135
adhE	58275_C01_00134	58275_A01_01483	58275_B01_00158	58366_B01_00141	58366_D01_00138	58366_C01_00139	58275_D01_00137	58275_E01_00136	58366_A01_00136
transposase IS605/IS1200					58366_D01_00139	58366_C01_00140	58275_D01_00138	58275_E01_00137	58366_A01_00137
cap5A	*cap8A*	58275_C01_00135	58275_A01_01484	58275_B01_00159	58366_B01_00142	58366_D01_00140	58366_C01_00141	58275_D01_00139	58275_E01_00138	58366_A01_00138
cap5B	*cap8B*	58275_C01_00136	58275_A01_01485	58275_B01_00160	58366_B01_00143	58366_D01_00141	58366_C01_00142	58275_D01_00140	58275_E01_00139	58366_A01_00139
cap5C_1	*cap8C_1*	58275_C01_00137	58275_A01_01486	58275_B01_00161	58366_B01_00144	58366_D01_00142	58366_C01_00143	58275_D01_00141	58275_E01_00140	58366_A01_00140
cap5D_1	*cap8D_1*	58275_C01_00138	58275_A01_01487	58275_B01_00162	58366_B01_00145	58366_D01_00143	58366_C01_00144	58275_D01_00142	58275_E01_00141	58366_A01_00141

Table 8.3 | Chimeric Region 2 (CH2) towards the origin of termination (without plasmids)

Gene/Product	CD140392 Locus	CD140901 Locus	CD140657 Locus	CD141496 Locus	CD150713 Locus	CD150916 Locus	CD140400 Locus	CD140638 Locus	CD140866 Locus
dedA	58275_C01_02622	58275_A01_01150	58275_B01_02639	58366_B01_02526	58366_D01_02435	58366_C01_02436	58275_D01_02568	58275_E01_02496	58366_A01_02510
yvrmA							58275_D01_02569	58275_E01_02497	58366_A01_02511
yvnA							58275_D01_02570	58275_E01_02498	58366_A01_02512
ecsA_3	58275_C01_02623	58275_A01_01151	58275_B01_02640	58366_B01_02527	58366_D01_02436	58366_C01_02437	58275_D01_02571	58275_E01_02499	58366_A01_02513
copper transporter	58275_C01_02624	58275_A01_01152	58275_B01_02641	58366_B01_02528	58366_D01_02437	58366_C01_02439	58275_D01_02572	58275_E01_02500	58366_A01_02514
fbp	58275_C01_02625	58275_A01_01153	58275_B01_02642	58366_B01_02529	58366_D01_02438	58366_C01_02440	58275_D01_02573	58275_E01_02501	58366_A01_02515
membrane protein	58275_C01_02627	58275_A01_01155	58275_B01_02644	58366_B01_02531	58366_D01_02440	58366_C01_02441	58275_D01_02574	58275_E01_02502	58366_A01_02516
phospholipase	58275_C01_02628	58275_A01_01156	58275_B01_02645	58366_B01_02532	58366_D01_02441	58366_C01_02442	58275_D01_02575	58275_E01_02503	58366_A01_02517
mhqO_2	58275_C01_02629	58275_A01_01157	58275_B01_02646	58366_B01_02533	58366_D01_02442	58366_C01_02443	58275_D01_02576	58275_E01_02504	58366_A01_02518
regulator protein HpaR	58275_C01_02630	58275_A01_01158	58275_B01_02647	58366_B01_02534	58366_D01_02443	58366_C01_02444	58275_D01_02577	58275_E01_02505	58366_A01_02519
acetyltransferase	58275_C01_02631	58275_A01_01159	58275_B01_02648	58366_B01_02535	58366_D01_02444	58366_C01_02445	58275_D01_02578	58275_E01_02506	58366_A01_02520
catE_2	58275_C01_02632	58275_A01_01160	58275_B01_02649	58366_B01_02536	58366_D01_02445	58366_C01_02446	58275_D01_02579	58275_E01_02507	58366_A01_02521
quinone reductase	58275_C01_02633	58275_A01_01161	58275_B01_02650	58366_B01_02537	58366_D01_02446	58366_C01_02447	58275_D01_02580	58275_E01_02508	58366_A01_02522
ldhD_1	58275_C01_02634	58275_A01_01162	58275_B01_02651	58366_B01_02538	58366_D01_02447	58366_C01_02448	58275_D01_02581	58275_E01_02509	58366_A01_02523
supH	58275_C01_02635	58275_A01_01163	58275_B01_02652	58366_B01_02539	58366_D01_02448	58366_C01_02449	58275_D01_02582	58275_E01_02510	58366_A01_02524
ybbL_2	58275_C01_02636	58275_A01_01164	58275_B01_02653	58366_B01_02540	58366_D01_02449	58366_C01_02450	58275_D01_02583	58275_E01_02511	58366_A01_02525
export permease	58275_C01_02637	58275_A01_01165	58275_B01_02654	58366_B01_02541	58366_D01_02450	58366_C01_02451	58275_D01_02584	58275_E01_02512	58366_A01_02526
srtA	58275_C01_02638	58275_A01_01166	58275_B01_02655	58366_B01_02542	58366_D01_02451	58366_C01_02452	58275_D01_02585	58275_E01_02513	58366_A01_02527
yncA	58275_C01_02639	58275_A01_01167	58275_B01_02656	58366_B01_02543	58366_D01_02452	58366_C01_02453	58275_D01_02586	58275_E01_02514	58366_A01_02528
hypothetical protein	58275_C01_02640	58275_A01_01168	58275_B01_02657	58366_B01_02544	58366_D01_02453	58366_C01_02454	58275_D01_02587	58275_E01_02515	58366_A01_02529
sdhA_2	58275_C01_02641	58275_A01_01169	58275_B01_02658	58366_B01_02545	58366_D01_02454	58366_C01_02455	58275_D01_02588	58275_E01_02516	58366_A01_02530
sdhB	58275_C01_02642	58275_A01_01170	58275_B01_02659	58366_B01_02546	58366_D01_02455	58366_C01_02456	58275_D01_02589	58275_E01_02517	58366_A01_02531
pfoR_2	58275_C01_02643	58275_A01_01171	58275_B01_02660	58366_B01_02547	58366_D01_02456	58366_C01_02457	58275_D01_02590	58275_E01_02518	58366_A01_02532
hypothetical protein	58275_C01_02644	58275_A01_01172	58275_B01_02661	58366_B01_02548	58366_D01_02457	58366_C01_02458	58275_D01_02591	58275_E01_02519	58366_A01_02533
yicL	58275_C01_02645	58275_A01_01173	58275_B01_02662	58366_B01_02549	58366_D01_02458	58366_C01_02459	58275_D01_02592	58275_E01_02520	58366_A01_02534
mlhB_2	58275_C01_02646	58275_A01_01174	58275_B01_02663	58366_B01_02550	58366_D01_02459	58366_C01_02460	58275_D01_02593	58275_E01_02521	58366_A01_02535
thioredoxin	58275_C01_02647	58275_A01_01175	58275_B01_02664	58366_B01_02551	58366_D01_02460	58366_C01_02461	58275_D01_02594	58275_E01_02522	58366_A01_02536
thioesterase protein	58275_C01_02648	58275_A01_01176	58275_B01_02665	58366_B01_02552	58366_D01_02461	58366_C01_02462	58275_D01_02595	58275_E01_02523	58366_A01_02537
transposase IS1272							58275_D01_02596	58275_E01_02524	58366_A01_02538
transposase IS1272							58275_D01_02597	58275_E01_02525	58366_A01_02539
transposase IS1272							58275_D01_02598	58275_E01_02526	58366_A01_02540
transposase							58275_D01_02599	58275_E01_02527	58366_A01_02541
glcB_2	58275_C01_02649	58275_A01_01177	58275_B01_02666	58366_B01_02553	58366_D01_02462	58366_C01_02463	58275_D01_02600	58275_E01_02528	58366_A01_02542
pox5	58275_C01_02650	58275_A01_01178	58275_B01_02667	58366_B01_02554	58366_D01_02463	58366_C01_02464	58275_D01_02601	58275_E01_02529	58366_A01_02543

cidB	58275_C01_02651	58275_A01_01179	58275_B01_02668	58366_B01_02555	58366_D01_02464	58366_C01_02465	58275_D01_02602	58275_E01_02530	58366_A01_02544
cidA	58275_C01_02652	58275_A01_01180	58275_B01_02669	58366_B01_02556	58366_D01_02465	58366_C01_02466	58275_D01_02603	58275_E01_02531	58366_A01_02545
cidR	58275_C01_02653	58275_A01_01181	58275_B01_02670	58366_B01_02557	58366_D01_02466	58366_C01_02467	58275_D01_02604	58275_E01_02532	58366_A01_02546
cytosolic protein	58275_C01_02654	58275_A01_01182	58275_B01_02671	58366_B01_02558	58366_D01_02467	58366_C01_02468	58275_D01_02605	58275_E01_02533	58366_A01_02547
ssaA2_4	58275_C01_02655	58275_A01_01183	58275_B01_02672	58366_B01_02559	58366_D01_02468	58366_C01_02469	58275_D01_02606	58275_E01_02534	58366_A01_02548
mvaA	58275_C01_02656	58275_A01_01184	58275_B01_02673	58366_B01_02560	58366_D01_02469	58366_C01_02470	58275_D01_02607	58275_E01_02535	58366_A01_02549
mvaS	58275_C01_02657	58275_A01_01185	58275_B01_02674	58366_B01_02561	58366_D01_02470	58366_C01_02471	58275_D01_02608	58275_E01_02536	58366_A01_02550
ogt	58275_C01_02658	58275_A01_01186	58275_B01_02675	58366_B01_02562	58366_D01_02471	58366_C01_02472	58275_D01_02609	58275_E01_02537	58366_A01_02551
clpL	58275_C01_02659	58275_A01_01187	58275_B01_02676	58366_B01_02563	58366_D01_02472	58366_C01_02473	58275_D01_02610	58275_E01_02538	58366_A01_02552
virus attachment protein	58275_C01_02660	58275_A01_01188	58275_B01_02677	58366_B01_02564	58366_D01_02473	58366_C01_02474	58275_D01_02611	58275_E01_02539	58366_A01_02553
feoB_1	58275_C01_02661	58275_A01_01189	58275_B01_02678	58366_B01_02565	58366_D01_02474	58366_C01_02475	58275_D01_02612	58275_E01_02540	58366_A01_02554
feoA	58275_C01_02662	58275_A01_01190	58275_B01_02679	58366_B01_02566	58366_D01_02475	58366_C01_02476	58275_D01_02613	58275_E01_02541	58366_A01_02555
mmpL8	58275_C01_02663	58275_A01_01191	58275_B01_02680	58366_B01_02567	58366_D01_02476	58366_C01_02477	58275_D01_02614	58275_E01_02542	58366_A01_02556
regulator protein (TetR family)	58275_C01_02664	58275_A01_01192	58275_B01_02681	58366_B01_02568	58366_D01_02477	58366_C01_02478	58275_D01_02615	58275_E01_02543	58366_A01_02557
rocA	58275_C01_02665	58275_A01_01193	58275_B01_02682	58366_B01_02569	58366_D01_02478	58366_C01_02479	58275_D01_02616	58275_E01_02544	58366_A01_02558
acetyltransferase	58275_C01_02666	58275_A01_01194	58275_B01_02683	58366_B01_02570	58366_D01_02479	58366_C01_02480	58275_D01_02617	58275_E01_02545	58366_A01_02559
exported protein	58275_C01_02667	58275_A01_01195	58275_B01_02684	58366_B01_02571	58366_D01_02480	58366_C01_02481	58275_D01_02618	58275_E01_02546	58366_A01_02560
copA	58275_C01_02668	58275_A01_01196	58275_B01_02685	58366_B01_02572	58366_D01_02481	58366_C01_02482	58275_D01_02619	58275_E01_02547	58366_A01_02561
copZ	58275_C01_02669	58275_A01_01197	58275_B01_02686	58366_B01_02573	58366_D01_02482	58366_C01_02483	58275_D01_02620	58275_E01_02548	58366_A01_02562
ldhD_2	58275_C01_02670	58275_A01_01198	58275_B01_02687	58366_B01_02574	58366_D01_02483	58366_C01_02484	58275_D01_02621	58275_E01_02549	58366_A01_02563
dapL	58275_C01_02671	58275_A01_01199	58275_B01_02688	58366_B01_02575	58366_D01_02484	58366_C01_02485	58275_D01_02622	58275_E01_02550	58366_A01_02564
crtN	58275_C01_02672	58275_A01_01200	58275_B01_02689	58366_B01_02576	58366_D01_02485	58366_C01_02486	58275_D01_02623	58275_E01_02551	58366_A01_02565
crtM	58275_C01_02673	58275_A01_01201	58275_B01_02690	58366_B01_02577	58366_D01_02486	58366_C01_02487	58275_D01_02624	58275_E01_02552	58366_A01_02566
crtQ	58275_C01_02674	58275_A01_01202	58275_B01_02691	58366_B01_02578	58366_D01_02487	58366_C01_02488	58275_D01_02625	58275_E01_02553	58366_A01_02567
crtP	58275_C01_02675	58275_A01_01203	58275_B01_02692	58366_B01_02579	58366_D01_02488	58366_C01_02489	58275_D01_02626	58275_E01_02554	58366_A01_02568
Acetyltransferase precursor	58275_C01_02676	58275_A01_01204	58275_B01_02693	58366_B01_02580	58366_D01_02489	58366_C01_02490	58275_D01_02627	58275_E01_02555	58366_A01_02569
ssaA2_5	58275_C01_02677	58275_A01_01205	58275_B01_02694	58366_B01_02581	58366_D01_02490	58366_C01_02491	58275_D01_02628	58275_E01_02556	58366_A01_02570
oatA_2	58275_C01_02678	58275_A01_01206	58275_B01_02695	58366_B01_02582	58366_D01_02491	58366_C01_02492	58275_D01_02629	58275_E01_02557	58366_A01_02571
acetyltransferase							58275_D01_02630	58275_E01_02558	58366_A01_02572
isaA	58275_C01_02679	58275_A01_01207	58275_B01_02696	58366_B01_02583	58366_D01_02492	58366_C01_02493	58275_D01_02631	58275_E01_02559	58366_A01_02573
membrane protein	58275_C01_02680	58275_A01_01208	58275_B01_02697	58366_B01_02584	58366_D01_02493	58366_C01_02494	58275_D01_02632	58275_E01_02560	58366_A01_02574
Regulator protein (EnvR)	58275_C01_02681	58275_A01_01209	58275_B01_02698	58366_B01_02585	58366_D01_02494	58366_C01_02495	58275_D01_02633	58275_E01_02561	58366_A01_02575
hypothetical protein	58275_C01_02682	58275_A01_01210	58275_B01_02699	58366_B01_02586	58366_D01_02495	58366_C01_02496	58275_D01_02634	58275_E01_02562	58366_A01_02576
ynzC	58275_C01_02683	58275_A01_01211	58275_B01_02700	58366_B01_02587	58366_D01_02496	58366_C01_02497	58275_D01_02635	58275_E01_02563	58366_A01_02577
glyoxalase	58275_C01_02684	58275_A01_01212	58275_B01_02701	58366_B01_02588	58366_D01_02497	58366_C01_02498	58275_D01_02636	58275_E01_02564	58366_A01_02578
azoB	58275_C01_02685	58275_A01_01213	58275_B01_02702	58366_B01_02589	58366_D01_02498	58366_C01_02499	58275_D01_02637	58275_E01_02565	58366_A01_02579

hypothetical protein	58275_C01_02686	58275_A01_01214	58275_B01_02703	58366_B01_02590	58366_D01_02499	58366_C01_02500	58275_D01_02638	58275_E01_02566	58366_A01_02580
acrR	58275_C01_02687	58275_A01_01215	58275_B01_02704	58366_B01_02591	58366_D01_02500	58366_C01_02501	58275_D01_02639	58275_E01_02567	58366_A01_02581
cpnA	58275_C01_02688	58275_A01_01216	58275_B01_02705	58366_B01_02592	58366_D01_02501	58366_C01_02502	58275_D01_02640	58275_E01_02568	58366_A01_02582
acid decarboxylase	58275_C01_02689	58275_A01_01217	58275_B01_02706	58366_B01_02593	58366_D01_02502	58366_C01_02503	58275_D01_02641	58275_E01_02569	58366_A01_02583
esterase	58275_C01_02690	58275_A01_01218	58275_B01_02707	58366_B01_02594	58366_D01_02503	58366_C01_02504	58275_D01_02642	58275_E01_02570	58366_A01_02584
alcohol dehydrogenase							58275_D01_02643	58275_E01_02571	58366_A01_02585
cobW	58275_C01_02691	58275_A01_01219	58275_B01_02708	58366_B01_02595	58366_D01_02504	58366_C01_02505	58275_D01_02644	58275_E01_02572	58366_A01_02586
feoB_2	58275_C01_02692	58275_A01_01220	58275_B01_02709	58366_B01_02596	58366_D01_02505	58366_C01_02506	58275_D01_02645	58275_E01_02573	58366_A01_02587
czcO	58275_C01_02693	58275_A01_01221	58275_B01_02710	58366_B01_02597	58366_D01_02506	58366_C01_02507	58275_D01_02646	58275_E01_02574	58366_A01_02588
esaC	58275_C01_02694	58275_A01_01222	58275_B01_02711	58366_B01_02598	58366_D01_02507	58366_C01_02508	58275_D01_02647	58275_E01_02575	58366_A01_02589
Hypothetical protein	58275_C01_02695	58275_A01_01223	58275_B01_02712	58366_B01_02599	58366_D01_02508	58366_C01_02509	58275_D01_02648	58275_E01_02576	58366_A01_02590
T VII secretion effector	58275_C01_02696	58275_A01_01224	58275_B01_02713	58366_B01_02600	58366_D01_02509	58366_C01_02510	58275_D01_02649	58275_E01_02577	58366_A01_02591
fructosamine kinase	58275_C01_02697	58275_A01_01225	58275_B01_02714	58366_B01_02601	58366_D01_02510	58366_C01_02511	58275_D01_02650	58275_E01_02578	58366_A01_02592
pyrD	58275_C01_02698	58275_A01_01226	58275_B01_02715	58366_B01_02602	58366_D01_02511	58366_C01_02512	58275_D01_02651	58275_E01_02579	58366_A01_02593
membrane protein	58275_C01_02699	58275_A01_01227	58275_B01_02716	58366_B01_02603	58366_D01_02512	58366_C01_02513	58275_D01_02652	58275_E01_02580	58366_A01_02594
adenine alpha hydrolases	58275_C01_02700	58275_A01_01228	58275_B01_02717	58366_B01_02604	58366_D01_02513	58366_C01_02514	58275_D01_02653	58275_E01_02581	58366_A01_02595
Phnb	58275_C01_02701	58275_A01_01229	58275_B01_02718	58366_B01_02605	58366_D01_02514	58366_C01_02515	58275_D01_02654	58275_E01_02582	58366_A01_02596
regulatory protein	58275_C01_02702	58275_A01_01230	58275_B01_02719	58366_B01_02606	58366_D01_02515	58366_C01_02516	58275_D01_02655	58275_E01_02583	58366_A01_02597
cocE	58275_C01_02703	58275_A01_01231	58275_B01_02720	58366_B01_02607	58366_D01_02516	58366_C01_02517	58275_D01_02656	58275_E01_02584	58366_A01_02598
cell wall anchored protein (sasK)							58275_D01_02657	58275_E01_02585	58366_A01_02599
mpr							58275_D01_02658	58275_E01_02586	58366_A01_02600
panD	58275_C01_02704	58275_A01_01232	58275_B01_02721	58366_B01_02608	58366_D01_02517	58366_C01_02518	58275_D01_02659	58275_E01_02587	58366_A01_02601
panC	58275_C01_02705	58275_A01_01233	58275_B01_02722	58366_B01_02609	58366_D01_02518	58366_C01_02519	58275_D01_02660	58275_E01_02588	58366_A01_02602
panB	58275_C01_02706	58275_A01_01234	58275_B01_02723	58366_B01_02610	58366_D01_02519	58366_C01_02520	58275_D01_02661	58275_E01_02589	58366_A01_02603
panE_2	58275_C01_02707	58275_A01_01235	58275_B01_02724	58366_B01_02611	58366_D01_02520	58366_C01_02521	58275_D01_02662	58275_E01_02590	58366_A01_02604
aldC_2	58275_C01_02708	58275_A01_01236	58275_B01_02725	58366_B01_02612	58366_D01_02521	58366_C01_02522	58275_D01_02663	58275_E01_02591	58366_A01_02605
transposase							58275_D01_02664	58275_E01_02592	58366_A01_02606
ldh2	58275_C01_02709	58275_A01_01237	58275_B01_02726	58366_B01_02613	58366_D01_02522	58366_C01_02523	58275_D01_02666	58275_E01_02594	58366_A01_02608
pheP	58275_C01_02710	58275_A01_01238	58275_B01_02727	58366_B01_02614	58366_D01_02523	58366_C01_02524	58275_D01_02667	58275_E01_02595	58366_A01_02609
gabT/puuE	58275_C01_02711	58275_A01_01239	58275_B01_02728	58366_B01_02615	58366_D01_02524	58366_C01_02525	58275_D01_02668	58275_E01_02596	58366_A01_02610
Hypothetical protein	58275_C01_02712	58275_A01_01240	58275_B01_02729	58366_B01_02616	58366_D01_02525	58366_C01_02526	58275_D01_02669	58275_E01_02597	58366_A01_02611
fda	58275_C01_02713	58275_A01_01241	58275_B01_02730	58366_B01_02617	58366_D01_02526	58366_C01_02527	58275_D01_02670	58275_E01_02598	58366_A01_02612
mqo2	58275_C01_02714	58275_A01_01242	58275_B01_02731	58366_B01_02618	58366_D01_02527	58366_C01_02528	58275_D01_02671	58275_E01_02599	58366_A01_02613
Hypothetical protein	58275_C01_02715	58275_A01_01243	58275_B01_02732	58366_B01_02619	58366_D01_02528	58366_C01_02529	58275_D01_02672	58275_E01_02600	58366_A01_02614
bclA	58275_C01_02716	58275_A01_01244	58275_B01_02733	58366_B01_02620	58366_D01_02529	58366_C01_02530	58275_D01_02673	58275_E01_02601	58366_A01_02615

antibiotic monooxygenase	58275_C01_02717	58275_A01_01245	58275_B01_02734	58366_B01_02621	58366_D01_02530	58366_C01_02531	58275_D01_02674	58275_E01_02602	58366_A01_02616
Putative cytosolic protein	58275_C01_02718	58275_A01_01246	58275_B01_02735	58366_B01_02622	58366_D01_02531	58366_C01_02532	58275_D01_02675	58275_E01_02603	58366_A01_02617
betA	58275_C01_02719	58275_A01_01247	58275_B01_02736	58366_B01_02623	58366_D01_02532	58366_C01_02533	58275_D01_02676	58275_E01_02604	58366_A01_02618
gbsA	58275_C01_02720	58275_A01_01248	58275_B01_02737	58366_B01_02624	58366_D01_02533	58366_C01_02534	58275_D01_02677	58275_E01_02605	58366_A01_02619
regulatory protein (ArsR family)	58275_C01_02721	58275_A01_01249	58275_B01_02738	58366_B01_02625	58366_D01_02534	58366_C01_02535	58275_D01_02678	58275_E01_02606	58366_A01_02620
betT	58275_C01_02722	58275_A01_01250	58275_B01_02739	58366_B01_02626	58366_D01_02535	58366_C01_02536	58275_D01_02679	58275_E01_02607	58366_A01_02621
Ribonucleotide reductase	58275_C01_02723	58275_A01_01251	58275_B01_02740	58366_B01_02627	58366_D01_02536	58366_C01_02537	58275_D01_02680	58275_E01_02608	58366_A01_02622
nrdD	58275_C01_02724	58275_A01_01252	58275_B01_02741	58366_B01_02628	58366_D01_02537	58366_C01_02538	58275_D01_02681	58275_E01_02609	58366_A01_02623
citN	58275_C01_02725	58275_A01_01253	58275_B01_02742	58366_B01_02629	58366_D01_02538	58366_C01_02539	58275_D01_02682	58275_E01_02610	58366_A01_02624
sirC_2	58275_C01_02726	58275_A01_01254	58275_B01_02743	58366_B01_02630	58366_D01_02539	58366_C01_02540	58275_D01_02683	58275_E01_02611	58366_A01_02625
cysJ	58275_C01_02727	58275_A01_01255	58275_B01_02744	58366_B01_02631	58366_D01_02540	58366_C01_02541	58275_D01_02684	58275_E01_02612	58366_A01_02626
bsaA_2	58275_C01_02728	58275_A01_01256	58275_B01_02745	58366_B01_02632	58366_D01_02541	58366_C01_02542	58275_D01_02685	58275_E01_02613	58366_A01_02627
hypothetical protein	58275_C01_02729						58275_D01_02686	58275_E01_02614	58366_A01_02628
ABC transporter permease	58275_C01_02730	58275_A01_01257	58275_B01_02746	58366_B01_02633	58366_D01_02542	58366_C01_02543	58275_D01_02687	58275_E01_02615	58366_A01_02629
lolD_2	58275_C01_02731	58275_A01_01258	58275_B01_02747	58366_B01_02634	58366_D01_02543	58366_C01_02544	58275_D01_02688	58275_E01_02616	58366_A01_02630
graS	58275_C01_02732	58275_A01_01259	58275_B01_02748	58366_B01_02635	58366_D01_02544	58366_C01_02545	58275_D01_02689	58275_E01_02617	58366_A01_02631
yvcP	58275_C01_02733	58275_A01_01260	58275_B01_02749	58366_B01_02636	58366_D01_02545	58366_C01_02546	58275_D01_02690	58275_E01_02618	58366_A01_02632
membrane protein	58275_C01_02734	58275_A01_01261	58275_B01_02750	58366_B01_02637	58366_D01_02546	58366_C01_02547	58275_D01_02691	58275_E01_02619	58366_A01_02633
phoB	58275_C01_02735	58275_A01_01262	58275_B01_02751	58366_B01_02638	58366_D01_02547	58366_C01_02548	58275_D01_02692	58275_E01_02620	58366_A01_02634
hypothetical protein	58275_C01_02736	58275_A01_01263	58275_B01_02752	58366_B01_02639	58366_D01_02548	58366_C01_02549	58275_D01_02693	58275_E01_02621	58366_A01_02635
Regulator protein (SlyA)	58275_C01_02737	58275_A01_01264	58275_B01_02753	58366_B01_02640	58366_D01_02549	58366_C01_02550	58275_D01_02694	58275_E01_02622	58366_A01_02636
Tributyrin esterase	58275_C01_02738	58275_A01_01265	58275_B01_02754	58366_B01_02641	58366_D01_02550	58366_C01_02551	58275_D01_02695	58275_E01_02623	58366_A01_02637
clfB_1	58275_C01_02739	58275_A01_01266	58275_B01_02755	58366_B01_02642	58366_D01_02551	58366_C01_02552	58275_D01_02696	58275_E01_02624	58366_A01_02638
arcR	58275_C01_02740	58275_A01_01267	58275_B01_02756	58366_B01_02643	58366_D01_02552	58366_C01_02553	58275_D01_02697	58275_E01_02625	58366_A01_02639
arcC	58275_C01_02741	58275_A01_01268	58275_B01_02757	58366_B01_02644	58366_D01_02553	58366_C01_02554	58275_D01_02698	58275_E01_02626	58366_A01_02640
arcD_2	58275_C01_02742	58275_A01_01269	58275_B01_02758	58366_B01_02645	58366_D01_02554	58366_C01_02555	58275_D01_02699	58275_E01_02627	58366_A01_02641
arcB	58275_C01_02743	58275_A01_01270	58275_B01_02759	58366_B01_02646	58366_D01_02555	58366_C01_02556	58275_D01_02700	58275_E01_02628	58366_A01_02642
arcA	58275_C01_02744	58275_A01_01271	58275_B01_02760	58366_B01_02647	58366_D01_02556	58366_C01_02557	58275_D01_02701	58275_E01_02629	58366_A01_02643
argR_2	58275_C01_02745	58275_A01_01272	58275_B01_02761	58366_B01_02648	58366_D01_02557	58366_C01_02558	58275_D01_02702	58275_E01_02630	58366_A01_02644
aur	58275_C01_02746	58275_A01_01273	58275_B01_02762	58366_B01_02649	58366_D01_02558	58366_C01_02559	58275_D01_02703	58275_E01_02631	58366_A01_02645
isaB	58275_C01_02747	58275_A01_01274	58275_B01_02763	58366_B01_02650	58366_D01_02559	58366_C01_02560	58275_D01_02704	58275_E01_02632	58366_A01_02646
putative low-complexity protein	58275_C01_02748	58275_A01_01275	58275_B01_02764	58366_B01_02651	58366_D01_02560	58366_C01_02561	58275_D01_02705	58275_E01_02633	58366_A01_02647
licR_4	58275_C01_02749	58275_A01_01276	58275_B01_02765	58366_B01_02652	58366_D01_02561	58366_C01_02562	58275_D01_02706	58275_E01_02634	58366_A01_02648
manP	58275_C01_02750	58275_A01_01277	58275_B01_02766	58366_B01_02653	58366_D01_02562	58366_C01_02563	58275_D01_02707	58275_E01_02635	58366_A01_02649
pmi	58275_C01_02751	58275_A01_01278	58275_B01_02767	58366_B01_02654	58366_D01_02563	58366_C01_02564	58275_D01_02708	58275_E01_02636	58366_A01_02650

phage infection protein	58275_C01_02752	58275_A01_01279	58275_B01_02768	58366_B01_02655	58366_D01_02564	58366_C01_02565	58275_D01_02709	58275_E01_02637	58366_A01_02651
amidase	58275_C01_02753	58275_A01_01280	58275_B01_02769	58366_B01_02656	58366_D01_02565	58366_C01_02566	58275_D01_02710	58275_E01_02638	58366_A01_02652
amidase	58275_C01_02754	58275_A01_01281	58275_B01_02770	58366_B01_02657	58366_D01_02566	58366_C01_02567	58275_D01_02711	58275_E01_02639	58366_A01_02653
hypothetical protein	58275_C01_02755	58275_A01_01282	58275_B01_02771	58366_B01_02658	58366_D01_02567	58366_C01_02568	58275_D01_02712	58275_E01_02640	58366_A01_02654
GftB	58275_C01_02756	58275_A01_01283	58275_B01_02772	58366_B01_02659	58366_D01_02568	58366_C01_02569	58275_D01_02713	58275_E01_02641	58366_A01_02655
gftA	58275_C01_02757	58275_A01_01284	58275_B01_02773	58366_B01_02660	58366_D01_02569	58366_C01_02570	58275_D01_02714	58275_E01_02642	58366_A01_02656
secA	58275_C01_02758	58275_A01_01285	58275_B01_02774	58366_B01_02661	58366_D01_02570	58366_C01_02571	58275_D01_02715	58275_E01_02643	58366_A01_02657
asp3	58275_C01_02759	58275_A01_01286	58275_B01_02775	58366_B01_02662	58366_D01_02571	58366_C01_02572	58275_D01_02716	58275_E01_02644	58366_A01_02658
asp2	58275_C01_02760	58275_A01_01287	58275_B01_02776	58366_B01_02663	58366_D01_02572	58366_C01_02573	58275_D01_02717	58275_E01_02645	58366_A01_02659
asp1	58275_C01_02761	58275_A01_01288	58275_B01_02777	58366_B01_02664	58366_D01_02573	58366_C01_02574	58275_D01_02718	58275_E01_02646	58366_A01_02660
secY2	58275_C01_02762	58275_A01_01289	58275_B01_02778	58366_B01_02665	58366_D01_02574	58366_C01_02575	58275_D01_02719	58275_E01_02647	58366_A01_02661
sraP	58275_C01_02763	58275_A01_01290	58275_B01_02779	58366_B01_02666	58366_D01_02575	58366_C01_02576	58275_D01_02720	58275_E01_02648	58366_A01_02662
flavin reductase	58275_C01_02764	58275_A01_01291	58275_B01_02780	58366_B01_02667	58366_D01_02576	58366_C01_02577	58275_D01_02721	58275_E01_02649	58366_A01_02663
putative lipoprotein	58275_C01_02765	58275_A01_01292	58275_B01_02781	58366_B01_02668	58366_D01_02577	58366_C01_02578	58275_D01_02722	58275_E01_02650	58366_A01_02664
putative lipoprotein	58275_C01_02766	58275_A01_01293	58275_B01_02782	58366_B01_02669	58366_D01_02578	58366_C01_02579	58275_D01_02723	58275_E01_02651	58366_A01_02665
hypothetical protein	58275_C01_02767	58275_A01_01294	58275_B01_02783	58366_B01_02670	58366_D01_02579	58366_C01_02580	58275_D01_02724	58275_E01_02652	58366_A01_02666
hypothetical protein	58275_C01_02768	58275_A01_01295	58275_B01_02784	58366_B01_02671	58366_D01_02580	58366_C01_02581			
nitrilotriacetate oxygenase	58275_C01_02769	58275_A01_01296	58275_B01_02785	58366_B01_02672	58366_D01_02581	58366_C01_02582	58275_D01_02725	58275_E01_02653	58366_A01_02667
msrA_2	58275_C01_02770	58275_A01_01297	58275_B01_02786	58366_B01_02673	58366_D01_02582	58366_C01_02583	58275_D01_02726	58275_E01_02654	58366_A01_02668
acetyltransferase	58275_C01_02771	58275_A01_01298	58275_B01_02787	58366_B01_02674	58366_D01_02583	58366_C01_02584	58275_D01_02727	58275_E01_02655	58366_A01_02669
capC_2	58275_C01_02772	58275_A01_01299	58275_B01_02788	58366_B01_02675	58366_D01_02584	58366_C01_02585	58275_D01_02728	58275_E01_02656	58366_A01_02670
epsD	58275_C01_02773	58275_A01_01300	58275_B01_02789	58366_B01_02676	58366_D01_02585	58366_C01_02586	58275_D01_02729	58275_E01_02657	58366_A01_02671
cap8A	58275_C01_02774	58275_A01_01301	58275_B01_02790	58366_B01_02677	58366_D01_02586	58366_C01_02587	58275_D01_02730	58275_E01_02658	58366_A01_02672
icaR	58275_C01_02775	58275_A01_01302	58275_B01_02791	58366_B01_02678	58366_D01_02587	58366_C01_02588	58275_D01_02731	58275_E01_02659	58366_A01_02673
icaA	58275_C01_02776	58275_A01_01303	58275_B01_02792	58366_B01_02679	58366_D01_02588	58366_C01_02589	58275_D01_02732	58275_E01_02660	58366_A01_02674
icaD	58275_C01_02777	58275_A01_01304	58275_B01_02793	58366_B01_02680	58366_D01_02589	58366_C01_02590	58275_D01_02733	58275_E01_02661	58366_A01_02675
icaB	58275_C01_02778	58275_A01_01305	58275_B01_02794	58366_B01_02681	58366_D01_02590	58366_C01_02591	58275_D01_02734	58275_E01_02662	58366_A01_02676
icaC	58275_C01_02779	58275_A01_01306	58275_B01_02795	58366_B01_02682	58366_D01_02591	58366_C01_02592	58275_D01_02735	58275_E01_02663	58366_A01_02677
lipA_3	58275_C01_02780	58275_A01_01307	58275_B01_02796	58366_B01_02683	58366_D01_02592	58366_C01_02593	58275_D01_02736	58275_E01_02664	58366_A01_02678
hypothetical protein				58366_B01_02684	58366_D01_02593	58366_C01_02594			
hisI	58275_C01_02781	58275_A01_01308	58275_B01_02797	58366_B01_02685	58366_D01_02594	58366_C01_02595	58275_D01_02737	58275_E01_02665	58366_A01_02679
hisF	58275_C01_02782	58275_A01_01309	58275_B01_02798	58366_B01_02686	58366_D01_02595	58366_C01_02596	58275_D01_02738	58275_E01_02666	58366_A01_02680
hisA	58275_C01_02783	58275_A01_01310	58275_B01_02799	58366_B01_02687	58366_D01_02596	58366_C01_02597	58275_D01_02739	58275_E01_02667	58366_A01_02681
hisH	58275_C01_02784	58275_A01_01311	58275_B01_02800	58366_B01_02688	58366_D01_02597	58366_C01_02598	58275_D01_02740	58275_E01_02668	58366_A01_02682
hisB	58275_C01_02785	58275_A01_01312	58275_B01_02801	58366_B01_02689	58366_D01_02598	58366_C01_02599	58275_D01_02741	58275_E01_02669	58366_A01_02683
hisC_2	58275_C01_02786	58275_A01_01313	58275_B01_02802	58366_B01_02690	58366_D01_02599	58366_C01_02600	58275_D01_02742	58275_E01_02670	58366_A01_02684

hisD	58275_C01_02787	58275_A01_01314	58275_B01_02803	58366_B01_02691	58366_D01_02600	58366_C01_02601	58275_D01_02743	58275_E01_02671	58366_A01_02685
hisG	58275_C01_02788	58275_A01_01315	58275_B01_02804	58366_B01_02692	58366_D01_02601	58366_C01_02602	58275_D01_02744	58275_E01_02672	58366_A01_02686
ATP transferase	58275_C01_02789	58275_A01_01316	58275_B01_02805	58366_B01_02693	58366_D01_02602	58366_C01_02603	58275_D01_02745	58275_E01_02673	58366_A01_02687
lipoprotein	58275_C01_02790	58275_A01_01317	58275_B01_02806	58366_B01_02694	58366_D01_02603	58366_C01_02604	58275_D01_02746	58275_E01_02674	58366_A01_02688
YceI	58275_C01_02791	58275_A01_01318	58275_B01_02807	58366_B01_02695	58366_D01_02604	58366_C01_02605	58275_D01_02747	58275_E01_02675	58366_A01_02689
hypothetical protein							58275_D01_02748	58275_E01_02676	58366_A01_02690
lactonase drp35	58275_C01_02792	58275_A01_01319	58275_B01_02808	58366_B01_02696	58366_D01_02605	58366_C01_02606	58275_D01_02749	58275_E01_02677	58366_A01_02691
sulfurtransferase	58275_C01_02793	58275_A01_01320	58275_B01_02809	58366_B01_02697	58366_D01_02606	58366_C01_02607	58275_D01_02750	58275_E01_02678	58366_A01_02692
pcp	58275_C01_02794	58275_A01_01321	58275_B01_02810	58366_B01_02698	58366_D01_02607	58366_C01_02608	58275_D01_02751	58275_E01_02679	58366_A01_02693
hypothetical protein	58275_C01_02795	58275_A01_01322	58275_B01_02811	58366_B01_02699	58366_D01_02608	58366_C01_02609	58275_D01_02752	58275_E01_02680	58366_A01_02694
thermal regulator protein	58275_C01_02796	58275_A01_01323	58275_B01_02812	58366_B01_02700	58366_D01_02609	58366_C01_02610	58275_D01_02753	58275_E01_02681	58366_A01_02695
dinB superfamily protein	58275_C01_02797	58275_A01_01324	58275_B01_02813	58366_B01_02701	58366_D01_02610	58366_C01_02611	58275_D01_02754	58275_E01_02682	58366_A01_02696
cna	58275_C01_02798	58275_A01_01325	58275_B01_02814	58366_B01_02702	58366_D01_02611	58366_C01_02612	58275_D01_02755	58275_E01_02683	58366_A01_02697
ttdT	58275_C01_02799	58275_A01_01326	58275_B01_02815	58366_B01_02703	58366_D01_02612	58366_C01_02613	58275_D01_02756	58275_E01_02684	58366_A01_02698
rarD	58275_C01_02800	58275_A01_01327	58275_B01_02816	58366_B01_02704	58366_D01_02613	58366_C01_02614	58275_D01_02757	58275_E01_02685	58366_A01_02699
DNA	58275_C01_02801	58275_A01_01328	58275_B01_02817	58366_B01_02705	58366_D01_02614	58366_C01_02615	58275_D01_02758	58275_E01_02686	58366_A01_02700
nixA	58275_C01_02802	58275_A01_01329	58275_B01_02818	58366_B01_02706	58366_D01_02615	58366_C01_02616	58275_D01_02759	58275_E01_02687	58366_A01_02701
nhoA	58275_C01_02803	58275_A01_01330	58275_B01_02819	58366_B01_02707	58366_D01_02616	58366_C01_02617	58275_D01_02760	58275_E01_02688	58366_A01_02702
membrane spanning protein	58275_C01_02804	58275_A01_01331	58275_B01_02820	58366_B01_02708	58366_D01_02617	58366_C01_02618	58275_D01_02761	58275_E01_02689	58366_A01_02703
vraD	58275_C01_02805	58275_A01_01332	58275_B01_02821	58366_B01_02709	58366_D01_02618	58366_C01_02619	58275_D01_02762	58275_E01_02690	58366_A01_02704
ABC transporter permease	58275_C01_02806	58275_A01_01333	58275_B01_02822	58366_B01_02710	58366_D01_02619	58366_C01_02620	58275_D01_02763	58275_E01_02691	58366_A01_02705
Hypothetical protein	58275_C01_02807	58275_A01_01334	58275_B01_02823	58366_B01_02711	58366_D01_02620	58366_C01_02621	58275_D01_02764	58275_E01_02692	58366_A01_02706
ncRNA	58275_C01_02808	58275_A01_01335	58275_B01_02824	58366_B01_02712	58366_D01_02621	58366_C01_02622	58275_D01_02765	58275_E01_02693	58366_A01_02707
immR	58275_C01_02809	58275_A01_01336	58275_B01_02825	58366_B01_02713	58366_D01_02622	58366_C01_02623	58275_D01_02766	58275_E01_02694	58366_A01_02708
membrane protein	58275_C01_02810	58275_A01_01337	58275_B01_02826	58366_B01_02714	58366_D01_02623	58366_C01_02624	58275_D01_02767	58275_E01_02695	58366_A01_02709
Permease	58275_C01_02811	58275_A01_01338	58275_B01_02827	58366_B01_02715	58366_D01_02624	58366_C01_02625	58275_D01_02768	58275_E01_02696	58366_A01_02710
parB_2	58275_C01_02812	58275_A01_01339	58275_B01_02828	58366_B01_02716	58366_D01_02625	58366_C01_02626	58275_D01_02769	58275_E01_02697	58366_A01_02711
gidB	58275_C01_02813	58275_A01_01340	58275_B01_02829	58366_B01_02717	58366_D01_02626	58366_C01_02627	58275_D01_02770	58275_E01_02698	58366_A01_02712
mnmG	58275_C01_02814	58275_A01_01341	58275_B01_02830	58366_B01_02718	58366_D01_02627	58366_C01_02628	58275_D01_02771	58275_E01_02699	58366_A01_02713
mnmE	58275_C01_02815	58275_A01_01342	58275_B01_02831	58366_B01_02719	58366_D01_02628	58366_C01_02629	58275_D01_02772	58275_E01_02700	58366_A01_02714
rnpA	58275_C01_02816	58275_A01_01343	58275_B01_02832	58366_B01_02720	58366_D01_02629	58366_C01_02630	58275_D01_02773	58275_E01_02701	58366_A01_02715
rpmH	58275_C01_02817	58275_A01_01344	58275_B01_02833	58366_B01_02721	58366_D01_02630	58366_C01_02631	58275_D01_02774	58275_E01_02702	58366_A01_02716

Table 8.4 | Type V (5C2&5) SCC*mec* Genes for ST45 Isolates and CD141496 (ST622-2014)

Gene/Product	CD140392 Locus	CD140901 Locus	CD140657 Locus	CD141496 Locus
hypothetical protein	58275_C01_00029	58275_A01_01373	58275_B01_00029	58366_B01_00029
hypothetical protein	58275_C01_00030	58275_A01_01374	58275_B01_00030	58366_B01_00030
hypothetical protein	58275_C01_00031	58275_A01_01375	58275_B01_00031	58366_B01_00031
hypothetical protein				58366_B01_00032
DNA polymerase	58275_C01_00032	58275_A01_01376	58275_B01_00032	58366_B01_00033
hypothetical protein	58275_C01_00033	58275_A01_01377	58275_B01_00033	58366_B01_00034
putative primase	58275_C01_00034	58275_A01_01378	58275_B01_00034	58366_B01_00035
hypothetical protein	58275_C01_00035	58275_A01_01379	58275_B01_00035	58366_B01_00036
ccrC_1	58275_C01_00036	58275_A01_01380	58275_B01_00036	58366_B01_00037
hypothetical protein	58275_C01_00037	58275_A01_01381	58275_B01_00037	58366_B01_00038
hypothetical protein	58275_C01_00038	58275_A01_01382		
pyridoxal enzyme	58275_C01_00039	58275_A01_01383	58275_B01_00038	58366_B01_00039
hypothetical protein	58275_C01_00040	58275_A01_01384	58275_B01_00039	58366_B01_00040
transposase	58275_C01_00041	58275_A01_01385	58275_B01_00040	58366_B01_00041
mvaS2	58275_C01_00042	58275_A01_01386	58275_B01_00041	58366_B01_00042
ugpQ_1	58275_C01_00043	58275_A01_01387	58275_B01_00042	58366_B01_00043
maoC	58275_C01_00044	58275_A01_01388	58275_B01_00043	58366_B01_00044
mecA	58275_C01_00045	58275_A01_01389	58275_B01_00044	58366_B01_00045
transposase	58275_C01_00046	58275_A01_01390	58275_B01_00045	58366_B01_00046
phnB protein	58275_C01_00047	58275_A01_01391	58275_B01_00046	58366_B01_00047
regulatory protein	58275_C01_00048	58275_A01_01392	58275_B01_00047	58366_B01_00048
hypothetical protein	58275_C01_00049	58275_A01_01393	58275_B01_00048	58366_B01_00049
DNA polymerase/exonuclease	58275_C01_00050	58275_A01_01394	58275_B01_00049	58366_B01_00050
hypothetical protein	58275_C01_00051	58275_A01_01395	58275_B01_00050	58366_B01_00051
putative primase	58275_C01_00052	58275_A01_01396	58275_B01_00051	58366_B01_00052
ccrC_2	58275_C01_00053	58275_A01_01397	58275_B01_00052	58366_B01_00053
hypothetical protein	58275_C01_00054	58275_A01_01398	58275_B01_00053	58366_B01_00054
hypothetical protein	58275_C01_00055	58275_A01_01399	58275_B01_00054	58366_B01_00055
pyridoxal phosphate	58275_C01_00056	58275_A01_01400	58275_B01_00055	58366_B01_00056
hypothetical protein	58275_C01_00057	58275_A01_01401	58275_B01_00056	58366_B01_00057
transposase	58275_C01_00058	58275_A01_01402	58275_B01_00057	58366_B01_00058
tet	58275_C01_00059	58275_A01_01403	58275_B01_00058	58366_B01_00059
plasmid recomb enzyme	58275_C01_00060	58275_A01_01404	58275_B01_00059	58366_B01_00060
plasmid recomb enzyme	58275_C01_00061	58275_A01_01405	58275_B01_00060	
hypothetical protein	58275_C01_00062	58275_A01_01406	58275_B01_00061	58366_B01_00061
repC	58275_C01_00063	58275_A01_01407	58275_B01_00062	58366_B01_00062
transposase	58275_C01_00064	58275_A01_01408	58275_B01_00063	58366_B01_00063
topB_1			58275_B01_00064	
hypothetical protein			58275_B01_00065	
hypothetical protein			58275_B01_00066	
hypothetical protein			58275_B01_00067	
czrA_1			58275_B01_00068	
hypothetical protein			58275_B01_00069	
zosA			58275_B01_00070	
pksB			58275_B01_00071	
glpE			58275_B01_00072	
csoR_1			58275_B01_00073	
sulfite exporter			58275_B01_00074	
copB			58275_B01_00075	
ydhK			58275_B01_00076	
hypothetical protein			58275_B01_00077	
hypothetical protein			58275_B01_00078	
transposase			58275_B01_00079	
hsdR_1 FRAGMENT			58275_B01_00080	
hsdS_1			58275_B01_00081	58366_B01_00064
cas1			58275_B01_00082	58366_B01_00065
cas3			58275_B01_00083	58366_B01_00066
csm1			58275_B01_00084	58366_B01_00067
csm2		58275_A01_01409	58275_B01_00085	58366_B01_00068
csm3		58275_A01_01410	58275_B01_00086	58366_B01_00069
csm4		58275_A01_01411	58275_B01_00087	58366_B01_00070
csm5		58275_A01_01412	58275_B01_00088	58366_B01_00071
csm6		58275_A01_01413	58275_B01_00089	58366_B01_00072
cas6	58275_C01_00065	58275_A01_01414	58275_B01_00090	58366_B01_00073
transposase	58275_C01_00066	58275_A01_01415	58275_B01_00091	58366_B01_00074
IS ATP binding domain	58275_C01_00067	58275_A01_01416	58275_B01_00092	58366_B01_00075
transposase	58275_C01_00068	58275_A01_01417	58275_B01_00093	58366_B01_00076

Gene/Product				
transposase, integrase	58275_C01_00069	58275_A01_01418	58275_B01_00094	58366_B01_00077
hypothetical protein	58275_C01_00070	58275_A01_01419	58275_B01_00095	58366_B01_00078
proline rich protein	58275_C01_00071	58275_A01_01420	58275_B01_00096	58366_B01_00079
membrane protein	58275_C01_00072	58275_A01_01421	58275_B01_00097	58366_B01_00080
emrB_1	58275_C01_00073	58275_A01_01422	58275_B01_00098	58366_B01_00081
N-acetyltransferase	58275_C01_00074	58275_A01_01423	58275_B01_00099	58366_B01_00082
N-acetyltransferase	58275_C01_00075	58275_A01_01424	58275_B01_00100	58366_B01_00083
DNA binding protein	58275_C01_00076	58275_A01_01425	58275_B01_00101	58366_B01_00084
acetyltransferase	58275_C01_00077	58275_A01_01426	58275_B01_00102	58366_B01_00085
dus	58275_C01_00078	58275_A01_01427	58275_B01_00103	58366_B01_00086
transposase	58275_C01_00079			
TfoX	58275_C01_00080	58275_A01_01428	58275_B01_00104	58366_B01_00087
membrane protein	58275_C01_00081	58275_A01_01429	58275_B01_00106	58366_B01_00088
hydrolase	58275_C01_00082	58275_A01_01430	58275_B01_00107	58366_B01_00089
regulator protein (FabR)	58275_C01_00083	58275_A01_01431	58275_B01_00108	58366_B01_00090
ubiE	58275_C01_00084	58275_A01_01432	58275_B01_00109	58366_B01_00091
hypothetical protein	58275_C01_00085	58275_A01_01433	58275_B01_00110	58366_B01_00092

Table 8.5 | Type IVh SCCmec Genes for ST22 isolates, ST622-2015 isolates CD150916, CD150713

Gene/Product	CD140400 Locus	CD140638 Locus	CD140866 Locus	CD150916 Locus	CD150713 Locus
hsdR_1 FRAGMENT	58275_D01_00029	58275_E01_00029	58366_A01_00029	58366_C01_00031	58366_D01_00029
hsdS_1	58275_D01_00030	58275_E01_00030	58366_A01_00030	58366_C01_00032	58366_D01_00030
speG	58275_D01_00031	58275_E01_00031	58366_A01_00031	58366_C01_00033	58366_D01_00031
membrane protein	58275_D01_00032	58275_E01_00032	58366_A01_00032	58366_C01_00034	58366_D01_00032
hypothetical protein	58275_D01_00033	58275_E01_00033	58366_A01_00033	58366_C01_00035	58366_D01_00033
pyridoxal phosphate	58275_D01_00034	58275_E01_00034	58366_A01_00034	58366_C01_00036	58366_D01_00034
hypothetical protein	58275_D01_00035	58275_E01_00035	58366_A01_00035	58366_C01_00037	58366_D01_00035
hypothetical protein	58275_D01_00036	58275_E01_00036	58366_A01_00036	58366_C01_00038	58366_D01_00036
ccrB_1	58275_D01_00037	58275_E01_00037	58366_A01_00037	58366_C01_00039	58366_D01_00037
ccrA_1	58275_D01_00038	58275_E01_00038	58366_A01_00038	58366_C01_00040	58366_D01_00038
hypothetical protein	58275_D01_00039	58275_E01_00039	58366_A01_00039	58366_C01_00041	58366_D01_00039
hypothetical protein	58275_D01_00040	58275_E01_00040	58366_A01_00040	58366_C01_00042	58366_D01_00040
Putative_ATP	58275_D01_00041	58275_E01_00041	58366_A01_00041	58366_C01_00043	58366_D01_00041
rli23	58275_D01_00042	58275_E01_00042	58366_A01_00042	58366_C01_00044	58366_D01_00042
Transposase	58275_D01_00043	58275_E01_00043	58366_A01_00043	58366_C01_00045	58366_D01_00043
aap	58275_D01_00044	58275_E01_00044	58366_A01_00044	58366_C01_00046	58366_D01_00044
hypothetical protein	58275_D01_00045	58275_E01_00045	58366_A01_00045	58366_C01_00047	58366_D01_00045
transposase	58275_D01_00046	58275_E01_00046	58366_A01_00046	58366_C01_00048	58366_D01_00046
mvaS2	58275_D01_00047	58275_E01_00047	58366_A01_00047	58366_C01_00049	58366_D01_00047
ugpQ_1	58275_D01_00048	58275_E01_00048	58366_A01_00048	58366_C01_00050	58366_D01_00048
maoC	58275_D01_00049	58275_E01_00049	58366_A01_00049	58366_C01_00051	58366_D01_00049
mecA	58275_D01_00050	58275_E01_00050	58366_A01_00050	58366_C01_00052	58366_D01_00050
mecR1	58275_D01_00051	58275_E01_00051	58366_A01_00051	58366_C01_00053	58366_D01_00051
hsdR_2 FRAGMENT	58275_D01_00052	58275_E01_00052	58366_A01_00052	58366_C01_00054	58366_D01_00052
tnp_1	58275_D01_00053	58275_E01_00053	58366_A01_00053	58366_C01_00055	58366_D01_00053
tnp_2	58275_D01_00054	58275_E01_00054		58366_C01_00056	58366_D01_00054
hypothetical protein	58275_D01_00055	58275_E01_00055	58366_A01_00054	58366_C01_00057	58366_D01_00055
pyridoxal phosphate	58275_D01_00056	58275_E01_00056	58366_A01_00055	58366_C01_00058	58366_D01_00056
hypothetical protein	58275_D01_00057	58275_E01_00057	58366_A01_00056	58366_C01_00059	58366_D01_00057
hypothetical protein	58275_D01_00058	58275_E01_00058	58366_A01_00057	58366_C01_00060	58366_D01_00058
ccrB_2	58275_D01_00059	58275_E01_00059	58366_A01_00058	58366_C01_00061	58366_D01_00059
ccrA_2	58275_D01_00060	58275_E01_00060	58366_A01_00059	58366_C01_00062	58366_D01_00060
hypothetical protein	58275_D01_00061	58275_E01_00061	58366_A01_00060	58366_C01_00063	58366_D01_00061

hypothetical protein	58275_D01_00062	58275_E01_00062	58366_A01_00061	58366_C01_00064	58366_D01_00062
hypothetical protein	58275_D01_00063	58275_E01_00063	58366_A01_00062	58366_C01_00065	58366_D01_00063
hypothetical protein	58275_D01_00064	58275_E01_00064	58366_A01_00063	58366_C01_00066	58366_D01_00064
hypothetical protein	58275_D01_00065	58275_E01_00065	58366_A01_00064	58366_C01_00067	58366_D01_00065
hypothetical protein	58275_D01_00066	58275_E01_00066	58366_A01_00065	58366_C01_00068	58366_D01_00066
hypothetical protein	58275_D01_00067		58366_A01_00066	58366_C01_00069	58366_D01_00067
transposase	58275_D01_00068	58275_E01_00067	58366_A01_00091	58366_C01_00070	58366_D01_00068
transposase	58275_D01_00069	58275_E01_00068	58366_A01_00090	58366_C01_00071	58366_D01_00069
hypothetical protein	58275_D01_00070	58275_E01_00069	58366_A01_00089	58366_C01_00072	58366_D01_00070
DNA/RNA helicase	58275_D01_00071	58275_E01_00070	58366_A01_00088	58366_C01_00073	58366_D01_00071
membrane protein	58275_D01_00072	58275_E01_00071	58366_A01_00087	58366_C01_00074	58366_D01_00072
membrane protein	58275_D01_00073	58275_E01_00072	58366_A01_00086	58366_C01_00075	58366_D01_00073
yofA	58275_D01_00074	58275_E01_00073	58366_A01_00085	58366_C01_00076	58366_D01_00074
hypothetical protein	58275_D01_00075	58275_E01_00074	58366_A01_00084	58366_C01_00077	58366_D01_00075
gltR	58275_D01_00076	58275_E01_00075	58366_A01_00083	58366_C01_00078	58366_D01_00076
ywqN_1	58275_D01_00077	58275_E01_00076	58366_A01_00082	58366_C01_00079	58366_D01_00077
macrolide-efflux protein	58275_D01_00078	58275_E01_00077	58366_A01_00081	58366_C01_00080	58366_D01_00078
cytosolic protein	58275_D01_00079	58275_E01_00078	58366_A01_00080	58366_C01_00081	58366_D01_00079
dus	58275_D01_00080	58275_E01_00079	58366_A01_00079	58366_C01_00082	58366_D01_00080
regulatory protein	58275_D01_00081	58275_E01_00080	58366_A01_00078	58366_C01_00083	58366_D01_00081
emrB_1	58275_D01_00082	58275_E01_00081	58366_A01_00077	58366_C01_00084	58366_D01_00082
hypothetical protein	58275_D01_00083	58275_E01_00082	58366_A01_00076	58366_C01_00085	58366_D01_00083
hypothetical protein	58275_D01_00084	58275_E01_00083	58366_A01_00075	58366_C01_00086	58366_D01_00084
membrane protein	58275_D01_00085	58275_E01_00084	58366_A01_00074	58366_C01_00087	58366_D01_00085
hsdR_2	58275_D01_00086	58275_E01_00085	58366_A01_00073	58366_C01_00088	58366_D01_00086
hsdS_2	58275_D01_00087	58275_E01_00086	58366_A01_00072	58366_C01_00089	58366_D01_00087
hsdM_1	58275_D01_00088	58275_E01_00087	58366_A01_00071	58366_C01_00090	58366_D01_00088
hypothetical protein	58275_D01_00089	58275_E01_00088	58366_A01_00070	58366_C01_00091	58366_D01_00089
hypothetical protein	58275_D01_00090	58275_E01_00089	58366_A01_00069	58366_C01_00092	58366_D01_00090
transposase	58275_D01_00091	58275_E01_00090	58366_A01_00068	58366_C01_00093	58366_D01_00091
transposase	58275_D01_00092	58275_E01_00091	58366_A01_00067	58366_C01_00094	58366_D01_00092

Table 8.6 | TI RM Core *sau1* elements within Singapore *S. aureus* collection

Isolate	ST	sau1hsdS_orfX	sau1hsdMSR (hsdM)	sau1hsdMSR (hsdS)	sau1hsdMSR (hsdR)	sau1hsdR (core)	sau1hsdMS1 (hsdM)	sau1hsdMS1 (hsdS)	sau1hsdMS2 (hsdS)	sau1hsdMS2 (hsdM)
EMRSA15	22		SAEMRSA15_00450	SAEMRSA15_00451	SAEMRSA15_00452	SAEMRSA15_01600	SAEMRSA15_03610	SAEMRSA15_03611	SAEMRSA15_17240	
CD140400	22	58275_D01_00030	58275_D01_00088	58275_D01_00087	58275_D01_00086	58275_D01_00185	58275_D01_00405	58275_D01_00406	58275_D01_01823	
CD140638	22	58275_E01_00030	58275_E01_00087	58275_E01_00086	58275_E01_00085	58275_E01_00184	58275_E01_00404	58275_E01_00405	58275_E01_01820	
CD140866	22	58366_A01_00030	58366_A01_00071	58366_A01_00072	58366_A01_00073	58366_A01_00184	58366_A01_00405	58366_A01_00406	58366_A01_01834	
CD150713	622	58366_D01_00030	58366_D01_00088	58366_D01_00087	58366_D01_00086	58366_D01_00185	58366_D01_00405	58366_D01_00406	58275_D01_01823	
CD150916	622	58366_C01_00033	58366_C01_00090	58366_C01_00089	58366_C01_00088	58366_C01_00187	58366_C01_00407	58366_C01_00408	58366_C01_01761	
CD141496	622					58366_B01_00188	58366_B01_00408	58366_B01_00409	58366_B01_01848	
CD140392	45					58275_C01_00181	58275_C01_00395	58275_C01_00396	58275_C01_01807	58275_C01_01808
CD140901	45					58275_A01_01530	58275_A01_01744	58275_A01_01745	58275_A01_00336	58275_A01_00337
CD140657	45					58275_B01_00205	58275_B01_00418	58275_B01_00419	58275_B01_01824	58275_B01_01825
CA-347	45					CA347_205	CA347_430	CA347_431	CA347_78	CA347_79

* isoaltes marked with **RED** text have a truncated hsdS which is predicted non-funtional

Table 8.7 | TII & TIV RM elements within Singapore *S. aureus* collection

Isolate	ST	bcgIA	bcgIB	srmB
EMRSA15	22	SAEMRSA15_13500	SAEMRSA15_13490	SAEMRSA15_23880
CD140400	22	58275_D01_01212	58275_D01_01213	58275_D01_02544
CD140638	22	58275_E01_01209	58275_E01_01210	58275_E01_02473
CD140866	22			58366_A01_02487
CD150713	622			58366_D01_02412
CD150916	622			58366_C01_02413
CD141496	622			58366_B01_02501
CD140392	45			58275_C01_02600
CD140901	45			58275_A01_01128
CD140657	45			58275_B01_02617
CA-347	45			CA347_2568

Table 8.8 | Restriction_and_Modification Analysis 6mA Motif Results for S. aureus Singapore Collection

Isolate	ST	motifString	partnerMotifString	groupTag	Methylated / Detected	Methylated Motif	Detected Motif	Mean Score	Mean IPD Ratio	Mean Coverage
CD140400	22	GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	GTANNNNNCTTC	1.000	264	264	228.742	6.644	152.027
		GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	GAAGNNNNNTAC	0.996	263	264	215.981	5.813	156.871
		TAAGNNNNNNTTC	TAAGNNNNNNTTC/GAANNNNNNCTTA	GAANNNNNNCTTA	1.000	438	438	232.094	6.715	154.292
		GAANNNNNNCTTA	TAAGNNNNNNTTC/GAANNNNNNCTTA	TAAGNNNNNNTTC	0.984	431	438	199.977	4.933	156.271
		YTCANNNNNNCCT	YTCANNNNNNCCT/AGGNNNNNNTGAR	AGGNNNNNNTGAR	0.989	688	696	210.924	5.553	156.805
		AGGNNNNNNTGAR	YTCANNNNNNCCT/AGGNNNNNNTGAR	YTCANNNNNNCCT	0.984	685	696	197.952	5.330	154.155
CD140638	22	GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	GTANNNNNCTTC	1.000	258	258	184.992	6.946	118.837
		GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	GAAGNNNNNTAC	0.996	257	258	175.479	5.846	124.381
		TAAGNNNNNNTTC	TAAGNNNNNNTTC/GAANNNNNNCTTA	GAANNNNNNCTTA	1.000	428	428	185.407	6.775	119.217
		GAANNNNNNCTTA	TAAGNNNNNNTTC/GAANNNNNNCTTA	TAAGNNNNNNTTC	0.981	420	428	163.767	5.143	121.290
		YTCANNNNNNCCT	YTCANNNNNNCCT/AGGNNNNNNTGAR	AGGNNNNNNTGAR	0.988	675	683	174.904	5.689	123.359
		AGGNNNNNNTGAR	YTCANNNNNNCCT/AGGNNNNNNTGAR	YTCANNNNNNCCT	0.984	672	683	166.571	5.670	120.406
CD140866	22	GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	GTANNNNNCTTC	1.000	261	261	131.360	6.461	86.602
		GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	GAAGNNNNNTAC	0.996	260	261	126.138	5.426	89.981
		TAAGNNNNNNTTC	TAAGNNNNNNTTC/GAANNNNNNCTTA	GAANNNNNNCTTA	0.998	435	436	128.910	6.307	85.901
		GAANNNNNNCTTA	TAAGNNNNNNTTC/GAANNNNNNCTTA	TAAGNNNNNNTTC	0.968	422	436	114.841	4.854	88.033
		YTCANNNNNNCCT	YTCANNNNNNCCT/AGGNNNNNNTGAR	AGGNNNNNNTGAR	0.986	683	693	126.925	5.366	89.539
		AGGNNNNNNTGAR	YTCANNNNNNCCT/AGGNNNNNNTGAR	YTCANNNNNNCCT	0.980	679	693	120.027	5.247	87.589
CD150713	622	TAAGNNNNNNTTC	TAAGNNNNNNTTC/GAANNNNNNCTTA	GAANNNNNNCTTA	1.000	425	425	222.920	6.501	150.214
		GAANNNNNNCTTA	TAAGNNNNNNTTC/GAANNNNNNCTTA	TAAGNNNNNNTTC	0.984	418	425	192.148	4.835	151.856
		GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	GTANNNNNCTTC	1.000	258	258	218.167	6.561	145.295
		GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	GAAGNNNNNTAC	0.996	257	258	207.953	5.675	152.245
		YTCANNNNNNCCT	YTCANNNNNNCCT/AGGNNNNNNTGAR	AGGNNNNNNTGAR	0.987	676	685	206.460	5.456	155.913
		AGGNNNNNNTGAR	YTCANNNNNNCCT/AGGNNNNNNTGAR	YTCANNNNNNCCT	0.988	677	685	195.046	5.303	152.390
CD150916	622	TAAGNNNNNNTTC	TAAGNNNNNNTTC/GAANNNNNNCTTA	GAANNNNNNCTTA	1.000	428	428	217.488	6.630	145.030
		GAANNNNNNCTTA	TAAGNNNNNNTTC/GAANNNNNNCTTA	TAAGNNNNNNTTC	0.991	424	428	187.113	4.835	148.462
		GAAGNNNNNTAC	GAAGNNNNNTAC/GTANNNNNCTTC	GTANNNNNCTTC	1.000	259	259	220.170	6.775	145.297
		GTANNNNNCTTC	GAAGNNNNNTAC/GTANNNNNCTTC	GAAGNNNNNTAC	0.996	258	259	208.814	5.665	151.151
		YTCANNNNNNCCT	YTCANNNNNNCCT/AGGNNNNNNTGAR	AGGNNNNNNTGAR	0.991	681	687	200.784	5.497	149.698
		AGGNNNNNNTGAR	YTCANNNNNNCCT/AGGNNNNNNTGAR	YTCANNNNNNCCT	0.981	674	687	193.697	5.412	147.886

|										
CD141496	622	YTCANNNNNCCT	YTCANNNNNCCT/AGGNNNNNNTGAR	AGGNNNNNNTGAR	0.988	684	692	185.965	5.665	134.598
		AGGNNNNNNTGAR	YTCANNNNNCCT/AGGNNNNNNTGAR	YTCANNNNNCCT	0.984	681	692	174.687	5.559	131.185
CD140392	45	TTTANNNNNCTWC	TTTANNNNNCTWC/GWAGNNNNNNTAAA	GWAGNNNNNNTAAA	0.978	569	582	139.144	5.585	98.311
		GWAGNNNNNNTAAA	TTTANNNNNCTWC/GWAGNNNNNNTAAA	TTTANNNNNCTWC	0.976	568	582	149.162	6.366	97.141
		GGANNNNNNNTTYG	GGANNNNNNNTTYG		0.977	375	384	138.285	4.961	100.373
CD140901	45	TTTANNNNNCTWC	TTTANNNNNCTWC/GWAGNNNNNNTAAA	GWAGNNNNNNTAAA	0.988	559	566	145.419	5.710	100.971
		GWAGNNNNNNTAAA	TTTANNNNNCTWC/GWAGNNNNNNTAAA	TTTANNNNNCTWC	0.984	557	566	153.063	6.529	98.828
		GGANNNNNNNTTYG	GGANNNNNNNTTYG/CRAANNNNNNNTCC	CRAANNNNNNNTCC	0.974	373	383	145.193	5.180	103.775
		CRAANNNNNNNTCC	GGANNNNNNNTTYG/CRAANNNNNNNTCC	GGANNNNNNNTTYG	0.974	373	383	138.386	5.042	100.962
CD140657	45	GWAGNNNNNNTAAA	GWAGNNNNNNTAAA/TTTANNNNNCTWC	TTTANNNNNCTWC	0.993	558	562	178.369	6.664	115.792
		TTTANNNNNCTWC	GWAGNNNNNNTAAA/TTTANNNNNCTWC	GWAGNNNNNNTAAA	0.991	557	562	170.917	5.844	118.955
		GGANNNNNNNTTYG	GGANNNNNNNTTYG/CRAANNNNNNNTCC	CRAANNNNNNNTCC	0.979	368	376	169.003	5.144	120.351
		CRAANNNNNNNTCC	GGANNNNNNNTTYG/CRAANNNNNNNTCC	GGANNNNNNNTTYG	0.981	369	376	159.339	5.037	116.328

Table 8.9 | EdgeR Pairwise DE Comparisons of ST22, ST45 and ST622 Chimeric Region

ST22 vs ST45	logFC	logCPM	PValue	FDR		Regulation
g02614-mmpL8	-5.693	12.139	0	0		ST45 ↓
g02636-glyoxalase	-2.921	10.064	1.37E-174	1.33E-172		ST45 ↓
g02639-acrR	-3.306	10.034	3.60E-157	2.34E-155		ST45 ↓
g02761-ABC_transporter_permease	3.350	10.152	7.79E-155	3.80E-153		ST45 ↑
g02644-cobW	-3.678	9.508	7.42E-133	2.89E-131		ST45 ↓
g02765-membrane_protein	-4.480	9.056	2.97E-131	9.64E-130		ST45 ↓
g02758-nhoA	-2.331	8.435	1.70E-109	4.74E-108		ST45 ↓
g02770-gidB	-2.784	9.727	5.79E-93	1.41E-91		ST45 ↓
g02635-ynzC	-3.522	8.687	1.35E-92	2.93E-91		ST45 ↓
g02760-vraD	3.605	8.118	1.39E-89	2.72E-88		ST45 ↑
g02721-flavin_reductase	-3.468	9.225	4.16E-87	7.37E-86		ST45 ↓
g02672-Hypothetical_protein	3.695	9.388	2.54E-85	4.13E-84		ST45 ↑
g02640-cpnA	-2.105	9.046	1.32E-72	1.84E-71		ST45 ↓
g02637-azoB	-2.993	10.269	1.95E-64	2.53E-63		ST45 ↓
g02702-argR_2	2.046	5.737	3.20E-41	3.90E-40		ST45 ↑
g02615-TetR_family_regulatory_protein	-2.057	12.356	2.01E-40	2.31E-39		ST45 ↓
g02736-lipA_3	-2.686	18.028	7.48E-26	5.21E-25		ST45 ↓
g02707-manP	2.371	12.781	3.07E-24	1.99E-23		ST45 ↑
g02632-membrane_spanning_protein	-3.422	6.798	3.65E-23	2.22E-22		ST45 ↓
g02675-Putative_cytosolic_protein	2.453	4.677	1.55E-17	7.19E-17		ST45 ↑
g02701-arcA	2.380	8.704	1.00E-14	3.55E-14		ST45 ↑
g02571-ecsA_3	2.076	9.301	4.34E-13	1.46E-12		ST45 ↑
g02591-hypothetical_protein	2.546	8.667	7.54E-12	2.30E-11		ST45 ↑
g02578-acetyltransferase	2.067	9.518	2.45E-11	7.12E-11		ST45 ↑
g02759-membrane_spanning_protein	-2.091	12.612	6.93E-11	1.96E-10		ST45 ↓

ST22 vs ST622	logFC	logCPM	PValue	FDR		Regulation
g02614-mmpL8	5.733	8.106	2.48E-178	5.56E-175		ST22 ↑
g02615-TetR_family_regulatory_protein	2.338	8.276	7.08E-42	9.93E-40		ST22 ↑
g02632-membrane_spanning_protein	3.000	2.794	2.81E-59	6.29E-57		ST22 ↑
g02635-ynzC	3.331	4.670	9.36E-76	3.00E-73		ST22 ↑
g02636-glyoxalase	2.376	6.107	6.74E-43	1.08E-40		ST22 ↑
g02637-azoB	2.584	6.289	1.17E-49	2.02E-47		ST22 ↑
g02639-acrR	3.512	5.984	2.60E-83	1.17E-80		ST22 ↑
g02640-cpnA	2.356	4.969	5.10E-42	7.62E-40		ST22 ↑
g02644-cobW	4.097	5.449	2.07E-106	1.55E-103		ST22 ↑
g02672-Hypothetical_protein	-3.411	5.095	7.47E-79	2.79E-76		ST22 ↓
g02675-Putative_cytosolic_protein	-2.306	0.545	1.15E-29	1.17E-27		ST22 ↓
g02721-flavin_reductase	3.610	5.174	8.98E-87	5.04E-84		ST22 ↑
g02758-nhoA	2.260	4.413	7.51E-39	9.91E-37		ST22 ↑
g02760-vraD	-2.941	3.495	4.60E-60	1.15E-57		ST22 ↓
g02761-ABC_transporter_permease	-2.731	5.571	2.07E-54	4.22E-52		ST22 ↓
g02765-membrane_protein	5.633	4.986	1.88E-165	2.11E-162		ST22 ↑
g02769-parB_2	2.118	6.138	4.56E-35	5.68E-33		ST22 ↑
g02770-gidB	3.116	5.657	3.02E-68	8.46E-66		ST22 ↑

| ST622 vs ST45	logFC	logCPM	PValue	FDR		Regulation
g02707-manP	-2.659	12.811	7.12E-44	1.39E-41		ST622 ↓
g02708-pmi	-2.538	11.122	4.19E-32	4.09E-30		ST622 ↓
g02736-lipA_3	2.400	17.860	1.38E-25	8.95E-24		ST622 ↑
g02571-ecsA_3	-2.456	9.356	1.42E-14	4.60E-13		ST622 ↓

🟨 *Upregulated*

🟦 *Downregulated*
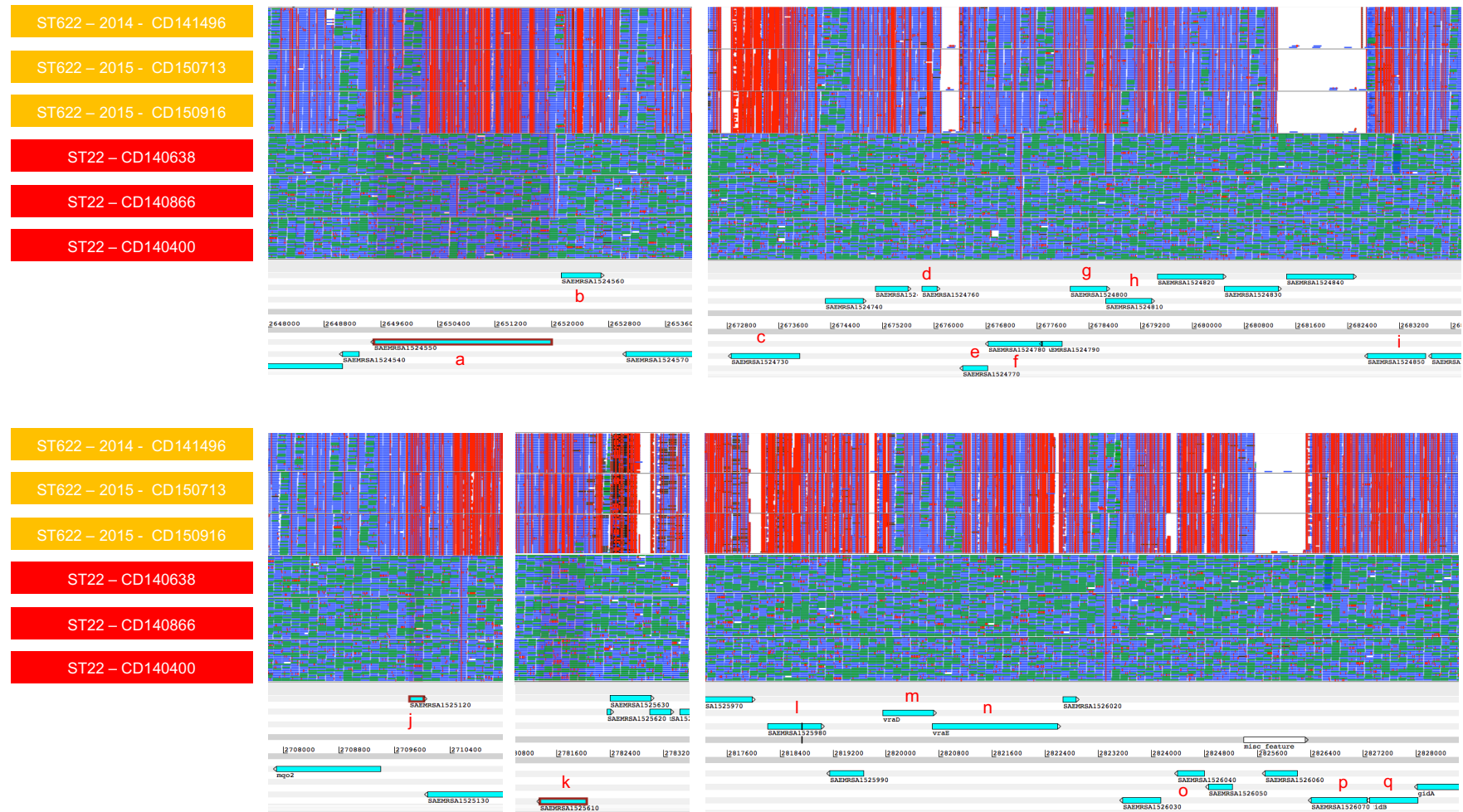
Figure 8.1 | Sequence Variation within DE genes between ST22 and ST622. To visualise the sequence comparisons, the alignments of ST22 and ST622 isolates were by mapped to reference strain EMRSA-15 (GE681097) using SMALT. BAM files visualised in Artemis with each row signifying 1 sequence alignment. The red vertical lines indicate single nucleotide polymorphisms (SNPs), whilst white gap regions indicate indels. Differentially expressed genes are marked by red lettering above the CDS feature: a) g02614-*mmpL8*; b) g02615-TetR_family_regulatory_protein; c) g02632-membrane_spanning_protein; d) g02635-*ynzC*; e) g02636-glyoxalase; f) g02637-*azoB*; g) g02639-*acrR*; h) g02640-*cpnA*; i) g02644-*cobW*; j) g02672-Hypothetical_protein; k) g02721-flavin_reductase_like_domain_protein; l) g02758-*nhoA*; m) g02760-*vraD*; n) g02761-ABC_transporter_permease; o) g02765-membrane_protein; p) g02769-*parB_2*; q) g02770-*gidB*.

317
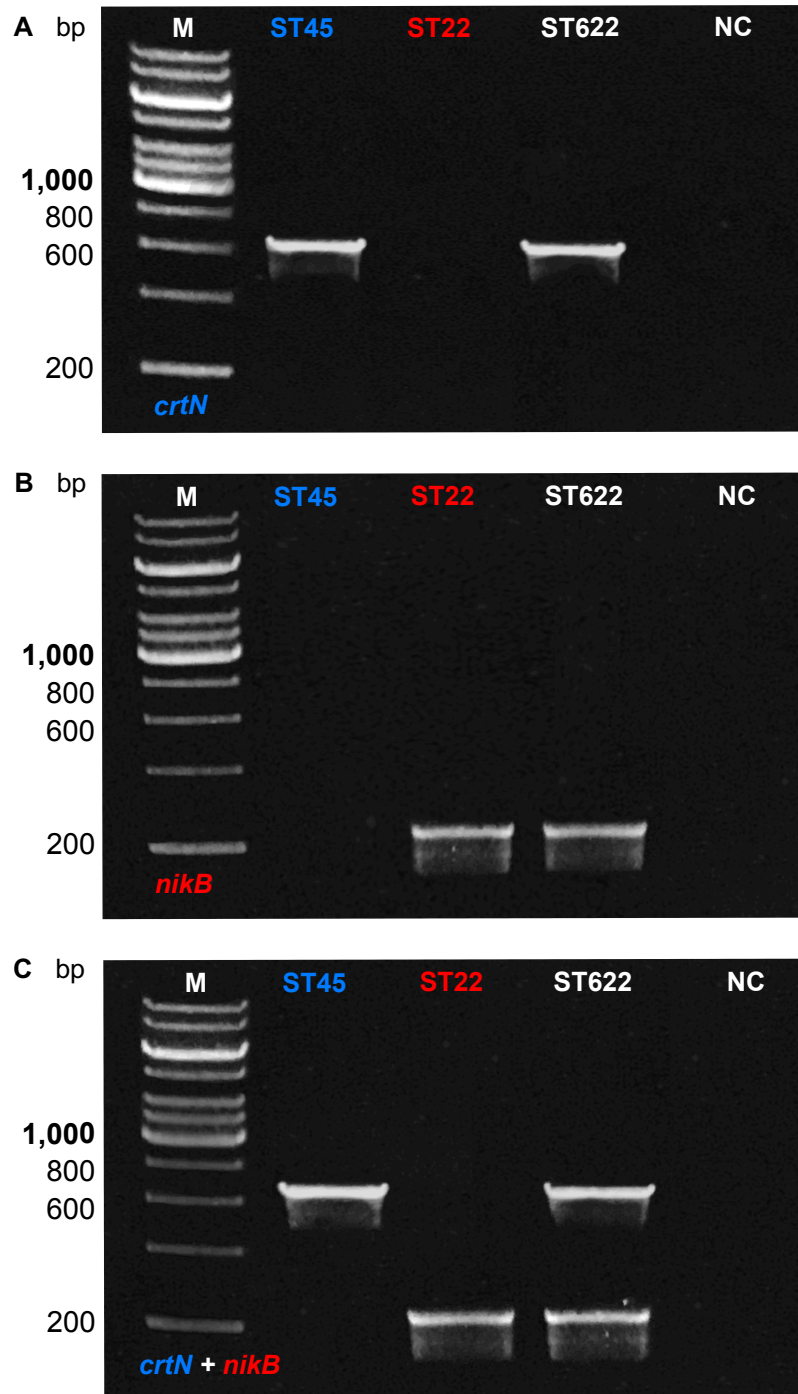
Figure 8.2 | Rapid PCR-test for identification of ST622-2015 strains – agarose gel electrophoresis visualisation of amplified PCR products for *crtN* and *nikB* target genes.

A. Amplification of genomic DNA with primer pair created for ST45 (blue) *crtN* allele (located within recombinant chimeric sequence region) will result in a 748bp band present within the ST45 (blue) and ST622 (white), but not ST22 (red). **B.** Amplification of genomic DNA with primer pair for ST22 allele *nikB* will result in a 257bp band in ST22 (red) and ST622 (white). **C**. Amplification with both primer sets will result in double bands within ST622 for both *crtN* and *nikB*, but will only amplify *crtN* in ST45 and *nikB* in ST22 due to divergence in sequence for the given homologs between the opposite stains. Primer pairs detailed in Table – in Methods.

8.3 CHAPTER 5 APPENDIX
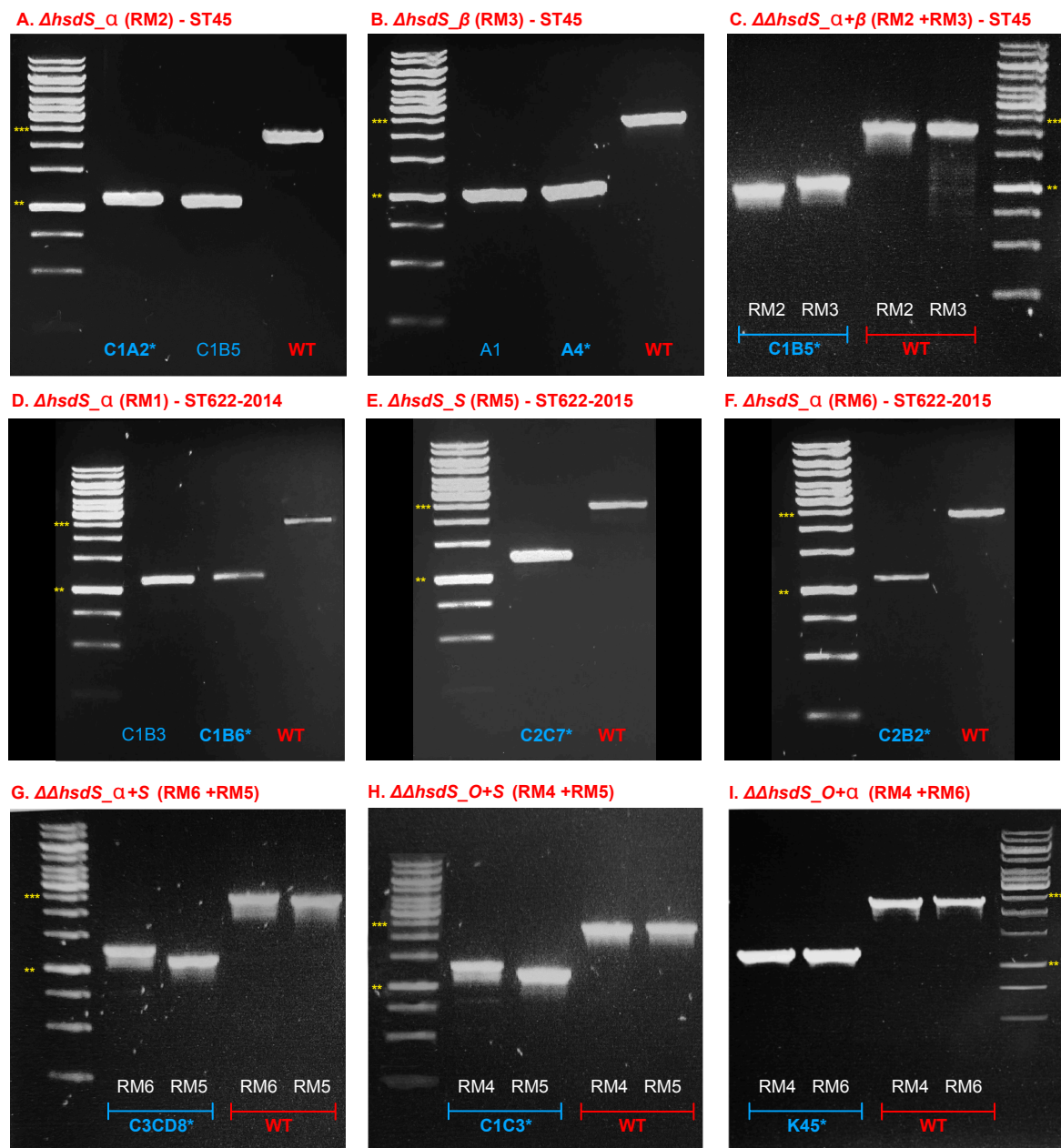
Figure 8.3 | Agarose gel electrophoresis visualization of PCR products validating the deletion of hsdS in 9 mutant *ΔhsdS* strains. Mutant construct primers A and D were used as a forward and reverse primer for each gene deletion (RM) – primers detailed in Table 2.9.7 in Methods. Double *hsdS* knockout mutants were test for both deletion constructions (RM) marked in white. Thermo Scientific GeneRuler 1kb DNA Ladder was used a guide to fragment sizes; 1000bp is market by 2 gold asterisks (**) and 2500bp marked by 3(***). Successful mutagenized isolates resulted in a ~1000bp amplified fragment (the length of primers A and D). gDNA from WT strains were included to show the presence of the hsdS gene of interest signified by the larger amplified DNA fragment (2500 bp). Isolates marked by bold font and an asterisk (*) were selected for sequencing and used to create competent cells for the creation of double *hsdS* deletion mutants. ST45 mutants were created form parent strain CD140392 to produce A. *ΔhsdS α (RM2)*, B. *ΔhsdS β (RM3), C. double knockout ΔΔhsdS α+β (RM2+RM3)* resulting in no functional Sau1 *hsdS.* ST622-2014 mutant used CD141496 as a parent strain creating D. *ΔhsdS α (RM1) rendering this isolate with no functional Sau1 hsdS.* CD150713 was used as a parent strain to create ST622-2015 mutants including: E. *ΔhsdS S (RM5)* F. *ΔhsdS α (RM6)* G. *ΔΔhsdS α+SCC* (RM6+RM5) H. *ΔΔhsdS orfX+SCC (RM4+RM5)* and I. *ΔΔhsdS orfX+α (RM4+RM6).*
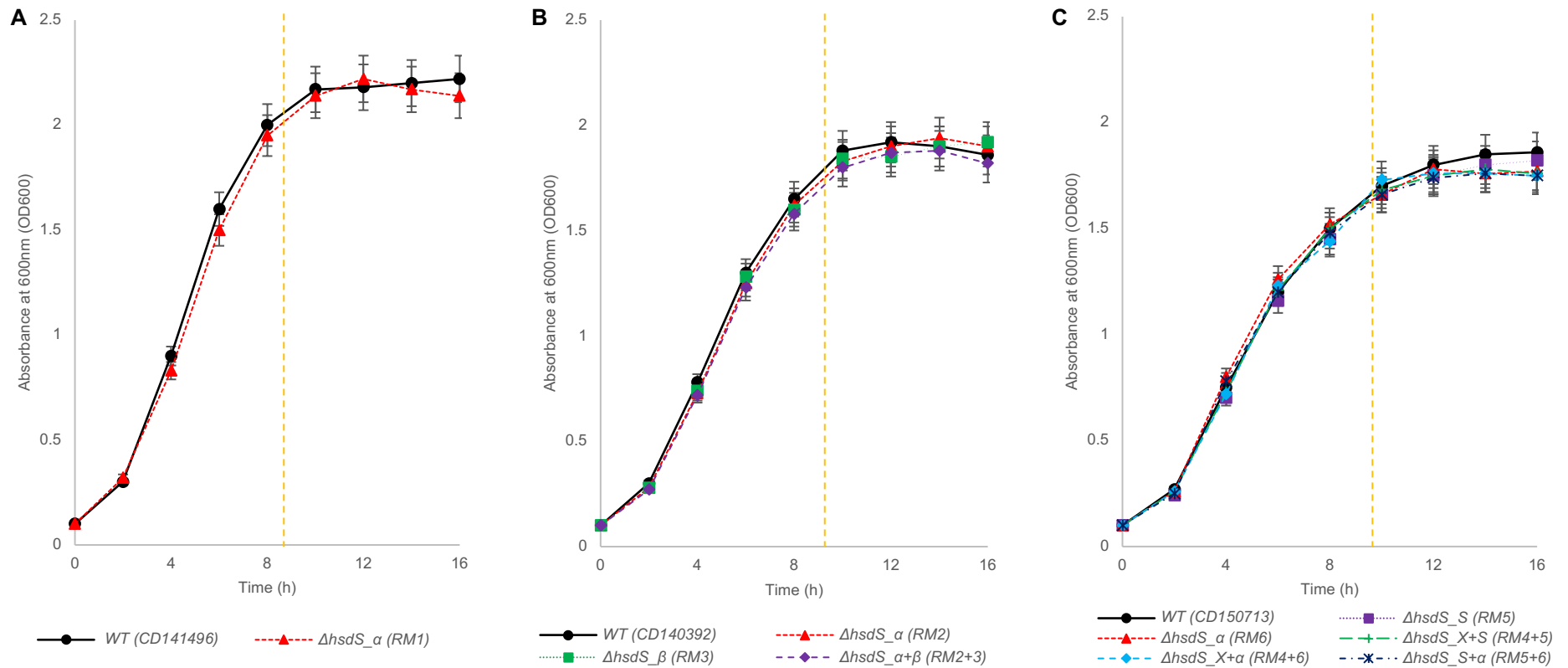
Figure 8.4 | Growth curves (optical density, OD = 600 nm) of WT vs RM mutant *S. aureus* strains in TSB (tryptic soy broth) rich media.
A. WT ST622-2014 variant CD141496 (black circle - full line) vs RM1 mutant *ΔhsdS*_α (red triangle). **B.** WT ST45 isolate CD140392 (black circle – full line) vs RM2 *ΔhsdS*_α (red triangle), RM3 *ΔhsdS*_β (green square), RM2+3 *ΔhsdS*_α+β (purple diamond). **C.** WT ST622-2015 variant CD150713 (black circle – full line) vs RM5 *ΔhsdS*_S (purple square), RM6 *ΔhsdS*_α (red triangle), RM4+5 *ΔhsdS*_X+S (green +), RM4+6 *ΔhsdS*_X+α (sky blue diamond), and RM5+6 *ΔhsdS*_S+α (navy blue x). Dashed orange lines indicate end of log phase growth at which samples were taken for DNA and RNA sequencing experiments. Each value was expressed using mean of triplicates for each strain with standard error of the mean (except WT - CD150713 and CD140392 which were only done in duplicate).

Table 8.10 | Differentially expressed genes between RM1_1/RM1_2 and RM cluster

Gene/ID	logFC	logCPM	PValue	FDR	Regulation
g00026_Zn_hydrolase	2.14	6.30	0	0	RM1 UP
g00358_XRE_Regulator	2.40	6.37	0	0	RM1 UP
g00748_ribosomal_subunit	2.01	11.91	0	0	RM1 UP
g00907_hyp_protein	4.83	7.64	0	0	RM1 UP
g00955_ydjZ	2.31	7.73	0	0	RM1 UP
g01108_hyp_protein	2.91	6.24	0	0	RM1 UP
g01343_hyp_protein	2.03	6.61	0	0	RM1 UP
g01362_memb_protein	2.09	7.24	0	0	RM1 UP
g01372_plsY	2.65	7.83	0	0	RM1 UP
g01441_cspA	3.37	9.87	0	0	RM1 UP
g01608_rpsU	2.41	9.31	0	0	RM1 UP
g02077_ydcD	2.11	8.36	0	0	RM1 UP
g02535_acetyltransferase	2.36	6.71	0	0	RM1 UP
g02558_put_cytosolic_prot	3.46	13.52	0	0	RM1 UP
g01371_iron_sulfur_biosynth	2.33	6.11	9.88E-324	5.73E-322	RM1 UP
g01772_peptidase	2.13	6.01	0.00E+00	8.03E-307	RM1 UP
g00668_sugar_transporter	2.22	5.55	2.12E-223	7.43E-222	RM1 UP
g00061_hyp_protein	2.27	4.44	5.15E-173	1.46E-171	RM1 UP
g01154_memb_protein	2.17	4.66	5.65E-159	1.45E-157	RM1 UP
g01632_memb_protein	2.12	3.39	2.39E-91	3.06E-90	RM1 UP
g01877_memb_protein	2.07	2.19	1.89E-47	1.28E-46	RM1 UP
g02789_tnsB	9.08	-0.28	1.24E-32	6.03E-32	RM1 UP

Table 8.11 | Uniquely DE genes - WT ST45 CD140392 vs RM2, RM3 and RM2+3

Downregulated in RM mutant vs WT

RM2	RM3	RM2+3
g00195_ycjS_1	g00318_efeO	g00268_esaB
g00205_ugpQ_2	g00657_hyp_protein	g00395_hsdM_1
g00284_hel	g00798_gcvH_2	g00565_ung
g00417_mccA	g01135_hyp_protein	g00575_mvaK1
g00619_mntR	g01136_hyp_protein	g00763_pgk
g00765_pgm_1	g01781_yvgN	g00784_phage_protein
g00776_acetyltransferase	g02103_metallopeptidase_SprT	g00830_phage_protein
g00781_thermonuclease	g02241_TIV_secretion_protein	g01073_potA
g01222_surface_protein	g02365_nagX	g01384_hydrolase
g01223_xlyA	g02542_blaI_1	g01511_malR_T_Regulator
g01392_phoU	g02566_put_cytosolic_prot	g01631_hisS
g01393_pstB3	g02596_put_cytosolic_prot	g01971_scn_3
g01394_pstA	g02841_traG_1	g02309_rplB
g01441_norB_5		g02313_rpsJ
g02254_lacG		g02317_acetyltransferase
g02255_lacE		g02573_yehR
g02256_lacF		g02714_mqo2
g02257_lacD		
g02404_yhaI		
g02452_narT		
g02459_nitrate_reductase		
g02460_narH		
g02461_narG		
g02462_nasF		
g02463_nasE		
g02464_nasD		
g02486_bioW		
g02763_sraP		
g02798_cna		

Upregulated in RM mutant vs WT

RM2	RM3	RM2+3
g00035_hyp_protein	g00034_phage_primase	g00026_Zn_hydrolase
g00885_memb_protein	g00079_transposase_IS256	g00277_lipoprotein
g00939_put_cytosolic_prot	g00105_sbnA	g00327_memb_protein
g01263_hyp_protein	g00150_mnaA_1	g00435_treA
g01285_xerC_2_integrase	g00431_transposase	g00541_dgk
g01587_DNA_polym_D_unit	g00943_lysR	g00583_memb_protein
g01668_clpX	g01574_hyp_protein	g00720_feuC
g01797_hyp_protein	g01575_memb_protein	g00800_primase
g01898_lipid_A_export_permease	g01583_hrcA	g01262_rny
g01919_put_staph_protein	g01717_soluable_hydrogenase	g01413_xpaC
g01939_His_repressor	g01738_put_cytosolic_prot	g01522_recN
g01973_chp	g01775_protease	g01964_memb_protein
g02220_htsB	g01776_protease	g02212_smrB
	g01808_hsdM	g02213_bmr3
	g01976_autolysin	g02230_transposase_IS1272
	g02005_phage_protein	g02380_memb_protein
	g02090_ilvH	g02389_memb_protein
	g02118_pot_transport_A	g02436_hyp_protein
	g02129_yedJ	g02443_PTS_system_comp

g02771_acetyltransferase	g02500_tcaB_2
	g02521_cadmium_efflux_repress
	g02532_hyp_protein
	g02549_transposase_IS256
	g02553_ohrR
	g02565_hyp_protein

Table 8.12 | DE Genes within ST45 Mutants (RM2, RM3, RM2+3)

RM2 vs RM3	logFC	logCPM	PValue	FDR	Regulation
01807_hsdS_2	-4.50	7.32	1.22E-249	3.19E-246	↑ RM2
00396_hsdS_1	5.73	3.69	2.15E-123	2.80E-120	↓ RM2
01188_lipoprotein	2.07	6.60	1.00E-79	8.72E-77	↓ RM2

RM2+3 vs RM2	logFC	logCPM	PValue	FDR	Regulation
01807_hsdS_2	4.65	7.38	2.94E-248	7.68E-245	↑ RM2

RM2+3 vs RM3	logFC	logCPM	PValue	FDR	Regulation
00396_hsdS_1	7.03	3.72	2.66E-120	6.93E-117	↑ RM3
02548_acetyltransferase	-2.05	5.15	1.88E-12	7.32E-10	↑ RM2+3

Table 8.13 | DE Genes within ST622-2015 Mutants (RM5, RM6, RM4+5, RM4+6, RM5+6)

RM5vsRM6	logFC	logCPM	PValue	FDR	Regulation
g00087_hsdS_2	4.47	5.51	5.40E-233	6.67E-230	↑ RM6
g00406_hsdS_3	-6.02	6.30	0	0	↑ RM5

RM5vsRM45	logFC	logCPM	PValue	FDR	RM45_1
g00030_hsdS_1	-10.00	4.74	1.78E-260	2.20E-257	↑ RM5
g00039_hyp_protein	-9.32	3.67	7.47E-110	1.84E-107	↑ RM5
g00040_hyp_protein	-6.99	0.13	2.10E-49	2.72E-47	↑ RM5
g00041_ATP-binding_protein	-11.58	4.66	3.04E-75	6.25E-73	↑ RM5
g00043_transposase	-10.25	1.40	1.65E-64	3.13E-62	↑ RM5
g00044_aap	-12.63	13.07	1.98E-56	3.05E-54	↑ RM5
g00045_hyp_protein	-9.96	5.14	1.21E-127	3.31E-125	↑ RM5
g00046_transposase_IS431	-5.58	3.07	3.71E-56	5.39E-54	↑ RM5
g00047_mvaS2	-8.63	2.42	1.88E-43	2.32E-41	↑ RM5
g00048_ugpQ_1	-10.00	6.04	2.11E-62	3.72E-60	↑ RM5
g00049_maoC	-10.63	3.01	2.06E-54	2.83E-52	↑ RM5
g00050_mecA_ftsI_1	-11.22	9.49	6.16E-188	2.53E-185	↑ RM5
g00051_mecR1	-11.22	7.32	1.99E-196	9.84E-194	↑ RM5
g00052_TI_RM_fragment	-11.79	6.11	5.25E-177	1.85E-174	↑ RM5
g00053_transposase_IS1272	-12.60	3.72	5.72E-137	1.76E-134	↑ RM5
g00054_hyp_protein	-10.44	1.58	9.45E-97	2.12E-94	↑ RM5
g00055_phos_dep_protein	-9.41	0.59	1.71E-56	2.82E-54	↑ RM5
g00057_hyp_protein	-9.33	0.53	2.16E-40	2.53E-38	↑ RM5
g00058_ccrB_2	-14.69	5.79	7.96E-248	6.55E-245	↑ RM5
g00059_hin_2_recombinase	-8.29	5.57	8.96E-276	2.21E-272	↑ RM5
g00087_hsdS_2	-4.72	5.50	3.10E-198	1.91E-195	↑ RM5

RM5vsRM46	logFC	logCPM	PValue	FDR	Regulation
g00030_hsdS_1	-10.15	4.75	1.96E-261	2.42E-258	↑ RM5
g00039_hyp_protein	-9.93	3.68	4.61E-108	1.26E-105	↑ RM5
g00040_hyp_protein	-9.00	0.13	1.47E-50	2.02E-48	↑ RM5
g00041_ATP-binding_protein	-13.64	4.67	2.29E-79	5.15E-77	↑ RM5
g00043_transposase	-10.34	1.42	1.94E-62	3.43E-60	↑ RM5
g00044_aap	-12.99	13.07	3.08E-23	3.63E-21	↑ RM5

	logFC	logCPM	PValue	FDR	Regulation
g00045_hyp_protein	-10.74	5.15	2.21E-128	6.81E-126	↑ RM5
g00046_transposase_IS431	-6.30	3.07	4.71E-57	7.76E-55	↑ RM5
g00047_mvaS2	-9.49	2.43	8.94E-46	1.16E-43	↑ RM5
g00048_ugpQ_1	-10.61	6.04	1.81E-76	3.73E-74	↑ RM5
g00049_maoC	-11.98	3.02	5.81E-54	8.43E-52	↑ RM5
g00050_mecA_ftsI_1	-12.83	9.49	4.62E-195	2.85E-192	↑ RM5
g00051_mecR1	-11.67	7.33	2.94E-170	1.21E-167	↑ RM5
g00052_TI_RM_fragment	-13.21	6.12	9.95E-184	4.91E-181	↑ RM5
g00053_transposase_IS1272	-11.46	3.73	2.22E-133	7.83E-131	↑ RM5
g00054_hyp_protein	-10.53	1.60	3.15E-95	7.79E-93	↑ RM5
g00055_phos_dep_protein	-9.49	0.60	1.50E-54	2.31E-52	↑ RM5
g00057_hyp_protein	-9.41	0.54	3.21E-39	3.97E-37	↑ RM5
g00058_ccrB_2	-12.11	5.80	7.31E-247	6.02E-244	↑ RM5
g00059_hin_2_recombinase	-8.49	5.58	3.20E-308	7.89E-305	↑ RM5
g00406_hsdS_3	-5.53	6.29	1.25E-74	2.38E-72	↑ RM5

RM5vsRM56	logFC	logCPM	PValue	FDR	Regulation
g00406_hsdS_3	-5.98	6.28	0	0	↑ RM5

RM6vsRM45	logFC	logCPM	PValue	FDR	Regulation
g00030_hsdS_1	-9.97	4.71	0	0	↑ RM6
g00031_speG	-2.04	4.56	2.83E-28	2.79E-26	↑ RM6
g00039_hyp_protein	-9.59	3.94	4.59E-259	9.90E-257	↑ RM6
g00040_hyp_protein	-7.31	0.42	9.30E-75	1.10E-72	↑ RM6
g00041_ATP-binding_protein	-12.30	5.36	1.40E-237	2.66E-235	↑ RM6
g00043_transposase	-10.43	1.60	7.03E-134	1.16E-131	↑ RM6
g00044_aap	-12.78	13.22	1.19E-67	1.34E-65	↑ RM6
g00045_hyp_protein	-10.12	5.28	0	0	↑ RM6
g00046_transposase_IS431	-5.73	3.17	1.12E-131	1.73E-129	↑ RM6
g00047_mvaS2	-8.86	2.69	6.37E-214	1.12E-211	↑ RM6
g00048_ugpQ_1	-10.21	6.20	0	0	↑ RM6
g00049_maoC	-11.03	3.42	1.57E-271	4.32E-269	↑ RM6
g00050_mecA_ftsI_1	-11.36	9.59	0	0	↑ RM6
g00051_mecR1	-11.47	7.58	0	0	↑ RM6
g00052_TI_RM_fragment	-11.95	6.29	0	0	↑ RM6
g00053_transposase_IS1272	-12.69	3.84	2.53E-287	7.81E-285	↑ RM6
g00054_hyp_protein	-10.17	1.34	3.82E-102	5.56E-100	↑ RM6
g00055_phos_dep_protein	-9.30	0.50	1.00E-75	1.24E-73	↑ RM6
g00056_hyp_protein	-8.54	-0.21	5.40E-54	5.81E-52	↑ RM6
g00057_hyp_protein	-9.33	0.53	1.78E-85	2.32E-83	↑ RM6
g00058_ccrB_2	-14.76	5.91	0	0	↑ RM6
g00059_hin_2_recombinase	-7.77	5.04	4.80E-259	9.90E-257	↑ RM6
g00087_hsdS_2	-4.72	5.49	1.58E-270	3.90E-268	↑ RM6
g00406_hsdS_3	5.72	5.93	4.85E-102	6.66E-100	↓ RM6
g00980_purF	2.12	7.87	6.95E-09	4.00E-07	↓ RM6
g00981_purM	2.26	7.46	3.34E-08	1.62E-06	↓ RM6
g00982_purN	2.30	6.67	2.48E-07	9.72E-06	↓ RM6
g00983_purH	2.23	8.37	1.24E-07	5.28E-06	↓ RM6
g00984_purD	2.29	8.35	2.42E-07	9.67E-06	↓ RM6

RM6vsRM46	logFC	logCPM	PValue	FDR	Regulation
g00030_hsdS_1	-10.14	4.72	0	0	↑ RM6
g00039_hyp_protein	-10.25	3.95	1.80E-279	4.05E-277	↑ RM6
g00040_hyp_protein	-9.30	0.42	1.94E-82	2.66E-80	↑ RM6
g00041_ATP-binding_protein	-14.30	5.37	9.55E-258	1.97E-255	↑ RM6
g00043_transposase	-10.53	1.62	4.70E-140	7.74E-138	↑ RM6
g00044_aap	-13.17	13.24	1.20E-22	1.29E-20	↑ RM6
g00045_hyp_protein	-10.90	5.30	0	0	↑ RM6
g00046_transposase_IS431	-6.48	3.18	4.67E-156	8.26E-154	↑ RM6
g00047_mvaS2	-9.80	2.71	5.46E-228	1.04E-225	↑ RM6
g00048_ugpQ_1	-10.88	6.21	0	0	↑ RM6
g00049_maoC	-12.36	3.44	2.07E-300	5.70E-298	↑ RM6
g00050_mecA_ftsI_1	-13.05	9.60	0	0	↑ RM6
g00051_mecR1	-12.00	7.59	0	0	↑ RM6
g00052_TI_RM_fragment	-13.40	6.31	0	0	↑ RM6

	logFC	logCPM	PValue	FDR	Regulation
g00053_transposase_IS1272	-11.59	3.86	4.31E-304	1.33E-301	↑ RM6
g00054_hyp_protein	-10.27	1.36	1.24E-102	1.92E-100	↑ RM6
g00055_phos_dep_protein	-9.39	0.51	1.24E-76	1.62E-74	↑ RM6
g00056_hyp_protein	-8.63	-0.21	5.32E-56	6.58E-54	↑ RM6
g00057_hyp_protein	-9.42	0.54	3.43E-91	4.98E-89	↑ RM6
g00058_ccrB_2	-12.23	5.92	0	0	↑ RM6
g00059_hin_2_recombinase	-7.98	5.05	1.67E-291	4.12E-289	↑ RM6

RM6vsRM56	logFC	logCPM	PValue	FDR	Regulation
g00087_hsdS_2	-4.49	5.52	0	0	↑ RM6

RM45vsRM46	logFC	logCPM	PValue	FDR	Regulation
g00087_hsdS_2	4.46	5.22	1.73E-86	4.23E-83	↑ RM4+6
g00406_hsdS_3	-5.21	5.92	6.43E-47	7.85E-44	↓ RM4+6

RM45vsRM56	logFC	logCPM	PValue	FDR	Regulation
g00030_hsdS_1	9.94	4.67	5.14E-263	4.22E-260	↓ RM4+5
g00039_hyp_protein	9.35	3.69	1.71E-170	6.01E-168	↓ RM4+5
g00040_hyp_protein	6.73	-0.13	2.73E-44	3.54E-42	↓ RM4+5
g00041_ATP-binding_protein	11.44	4.52	7.85E-79	1.38E-76	↓ RM4+5
g00043_transposase	10.15	1.33	2.80E-72	4.31E-70	↓ RM4+5
g00044_aap	12.68	13.11	1.29E-60	1.77E-58	↓ RM4+5
g00045_hyp_protein	9.90	5.06	5.62E-145	1.54E-142	↓ RM4+5
g00046_transposase_IS431	5.47	2.93	4.14E-75	6.79E-73	↓ RM4+5
g00047_mvaS2	8.55	2.37	4.11E-91	7.78E-89	↓ RM4+5
g00048_ugpQ_1	10.10	6.10	1.31E-158	4.05E-156	↓ RM4+5
g00049_maoC	10.65	3.05	1.74E-131	4.28E-129	↓ RM4+5
g00050_mecA_ftsI_1	11.21	9.46	4.18E-193	1.71E-190	↓ RM4+5
g00051_mecR1	11.30	7.39	4.41E-200	2.48E-197	↓ RM4+5
g00052_TI_RM_fragment	11.87	6.18	5.04E-200	2.48E-197	↓ RM4+5
g00053_transposase_IS1272	12.57	3.72	1.49E-130	3.33E-128	↓ RM4+5
g00054_hyp_protein	10.18	1.37	2.88E-71	4.17E-69	↓ RM4+5
g00055_phos_dep_protein	9.10	0.33	1.12E-42	1.31E-40	↓ RM4+5
g00057_hyp_protein	9.31	0.52	7.82E-43	9.63E-41	↓ RM4+5
g00058_ccrB_2	14.49	5.63	0	0	↓ RM4+5
g00059_hin_2_recombinase	8.04	5.31	0	0	↓ RM4+5
g00406_hsdS_3	-5.70	5.92	2.94E-103	6.04E-101	↑ RM4+5

RM46vs56	logFC	logCPM	PValue	FDR	Regulation
g00030_hsdS_1	10.10	4.68	1.35E-256	1.11E-253	↓ RM4+6
g00039_hyp_protein	9.99	3.71	6.70E-167	2.07E-164	↓ RM4+6
g00040_hyp_protein	8.70	-0.13	4.19E-47	5.74E-45	↓ RM4+6
g00041_ATP-binding_protein	13.46	4.53	6.33E-81	1.12E-78	↓ RM4+6
g00043_transposase	10.24	1.35	2.78E-70	4.28E-68	↓ RM4+6
g00044_aap	13.06	13.11	8.32E-23	9.33E-21	↓ RM4+6
g00045_hyp_protein	10.68	5.07	3.87E-142	1.06E-139	↓ RM4+6
g00046_transposase_IS431	6.20	2.94	6.32E-78	1.04E-75	↓ RM4+6
g00047_mvaS2	9.46	2.39	7.36E-90	1.51E-87	↓ RM4+6
g00048_ugpQ_1	10.75	6.11	6.94E-184	2.85E-181	↓ RM4+6
g00049_maoC	11.98	3.06	3.07E-126	7.57E-124	↓ RM4+6
g00050_mecA_ftsI_1	12.83	9.47	9.96E-207	6.15E-204	↓ RM4+6
g00051_mecR1	11.76	7.40	1.10E-167	3.90E-165	↓ RM4+6
g00052_TI_RM_fragment	13.28	6.19	1.61E-199	7.95E-197	↓ RM4+6
g00053_transposase_IS1272	11.46	3.73	7.69E-122	1.73E-119	↓ RM4+6
g00054_hyp_protein	10.27	1.39	1.01E-67	1.47E-65	↓ RM4+6
g00055_phos_dep_protein	9.19	0.35	5.80E-41	7.15E-39	↓ RM4+6
g00057_hyp_protein	9.39	0.53	1.22E-41	1.58E-39	↓ RM4+6
g00058_ccrB_2	11.96	5.64	0	0	↓ RM4+6
g00059_hin_2_recombinase	8.24	5.32	0	0	↓ RM4+6
g00087_hsdS_2	-4.27	5.26	1.22E-81	2.31E-79	↑ RM4+6

Genes marked in blue – hsdS KO genes; genes marked in red – RM4 construct 20 gene cluster which was not expressed (inactivation suspected to be a result of transformation)
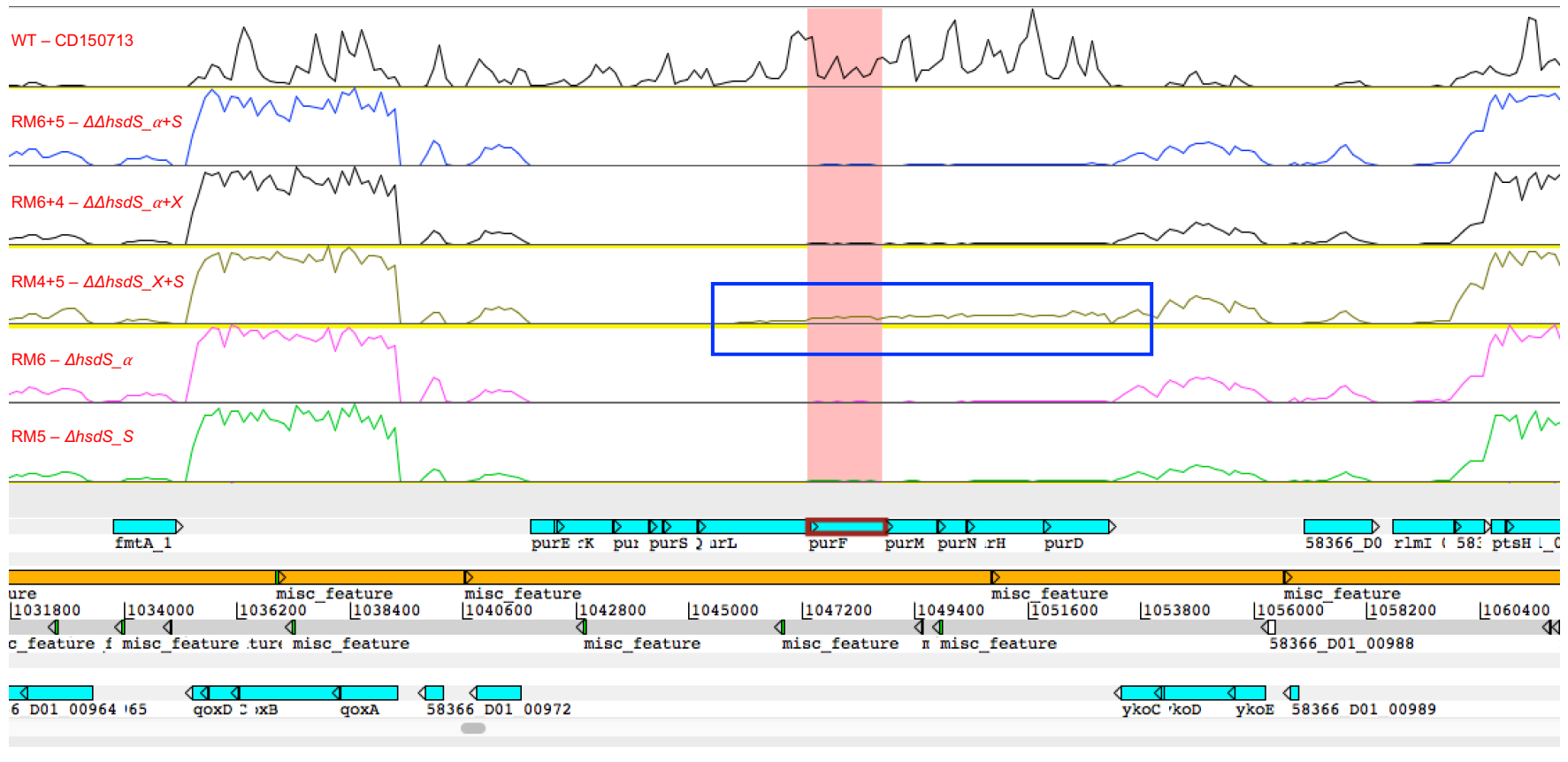
Figure 8.4 | RNA transcript levels for gene cluster purine cluster in ST622-2015.
Artemis visualisation of RNASeq generated transcript reads aligned to WT reference CD150713 genome for each RM mutant (top down: WT, RM5+6, RM4+6, RM4+5, RM6, RM5 - stacked windows – scaled at 10,000 RPKM) visualising genome region containing purine cluster (*purEKCSQLFMNHD*). *PurF* is highlighted in pink – *purFMNHD* were upregulated in RM4+5 in comparison to only RM6 (+2 logFC). The putative purine biosynthesis operon is transcriptionally linked and is regulated by *purR* repressor. Although the 10 genes belonging to the purine biosynthesis operon are transcriptionally linked, only half of the genes were upregulated, seen in the higher transcript levels in RM4+5 marked by the blue box.

Table 8.14 | Uniquely DE genes between WT ST622-2015 CD150713 vs RM5, RM6, RM4+5, RM4+6, RM5+6 mutant strains

Downregulated in RM mutant vs WT

RM5	RM6	RM4_5	RM4_6	RM5_6
g01277_citB	g00432_amiD	g01117_hyp_protein		g00823_nitrate_monoxygenase
g01949_agrC	g00619_mrpF	g01396_gpsB		g00832_ghrB_1
g02096_put_cytosolic_prot	g00751_uvrB_2			
	g00763_maeA			
	g00859_rocD2_2			
	g00892_oppA			
	g01023_cfiB_1			
	g01196_recA			
	g01518_memb_protein			
	g01603_Abrb_ammon_dehydroxylase			
	g01630_phoR			
	g01744_yokF			
	g02224_formate_dehydrogenase			
	g02434_garP			
	g02445_catE_2			
	g02485_crtN			
	g02487_crtQ			

Upregulated in RM mutant vs WT

RM5	RM6	RM4_5	RM4_6	RM5_6
g00554_sdrD	g00049_maoC	g00127_deoD_1	g00230_gatB_1	g00213_staphylocoagulase
g00946_memb_protein	g00375_phage_reg_protein	g00149_capK5	g00413_lipoprotein	g00355_T_regulator
g01194_pgsA	g00384_SaPI_protein_spore	g00835_dltB	g01620_thrS	g00369_xprT
g01811_transposase	g00989_hyp_protein	g00837_dltD	g02159_rpmC	
	g01073_hyp_protein	g00985_ykoC	g02578_lipoprotein	
	g01239_hyp_protein	g02131_aldC_1	g02649_repA_plasmid	
	g01405_dinG	g02490_ssaA2_5		
	g01594_hyp_protein			
	g01894_cysteine_protease_inhibitor			
	g02127_hyp_protein			
	g02238_hyp_protein			
	g02269_T_regulator_MDR			
	g02412_srmB_TIV_restriction			