



Evolutionary drivers of the rise and fall of a dominant hybrid MRSA, ST239

Jacqueline Gill

Linacre College,
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2019

Abstract

Staphylococcus aureus infections pose a serious challenge in healthcare due to the emergence of resistance to multiple antibiotics. ST239 is a multidrug-resistant MRSA strain that has been responsible for epidemics in healthcare settings across the globe since the late 1970s. However, in recent years there have been reports of strains that have successfully replaced ST239 as the dominant strain in clinical environments.

In 2004, Robinson and Enright hypothesised that ST239 is a natural hybrid that formed through a single large chromosomal replacement event between two distantly-related *S. aureus* strains, where 20% of the ancestral ST8-like genome was replaced with the homologous region from an ST30-like ancestor. In this thesis, I use a diverse range of ST239, ST8 and ST30 sequences from publicly-available sequencing repositories to make genomic comparisons between ST239 and its hypothesised ancestors. I confirm that the evolution of ST239 is consistent with the Robinson-Enright model, and investigate how the acquired and backbone regions of ST239 have adapted to evolutionary pressures in the same genome. There is evidence of weaker selective constraints on the acquired region of ST239 compared to the genomic backbone, as well as evidence of parallel evolution in antimicrobial resistance and virulence related genes, which suggests that antibiotics and the host immune response are important evolutionary pressures faced by ST239. This is consistent with the close relationship between ST239 and the *SCCmec*-III element, which encodes multiple antimicrobial resistance genes. In over 40,000 *S. aureus* sequences, *SCCmec*-III is strongly linked to ST239, with only a few rare exceptions. In contrast with previous estimates, the most recent common ancestor of the ST239 sequences is dated several years prior to the

clinical introduction of methicillin, which suggests that use of earlier antimicrobials is more likely to have driven the global emergence of ST239.

Finally, I develop a method to calculate the relative competitive ability of a collection of 15 *S. aureus* isolates in co-culture, including ST239, ST8 and ST30 isolates, through whole-genome sequencing. Competitive ability and growth rate are not correlated, suggesting that these two fitness measures are independent. The ST239 isolates have greater antimicrobial resistance and overall lower fitness, however, fitness and antimicrobial resistance are not correlated. This may indicate that, although ST239 has been a globally successful multidrug-resistant epidemic MRSA strain, the extreme genetic transfer events in its early evolution have had an overall negative impact on fitness, contributing to its eventual decline.

Acknowledgments

During my time in Oxford, I have learned more than I ever thought I would know, and have achieved far more than I ever dreamed I could achieve. Along the way, I have been so lucky to get to know so many brilliant and kind people, who have each shaped my experience in their own way, and who as a result have shaped me into the person I have become.

I would like to thank my supervisor, Professor Craig MacLean, who laughed away my worries that I would not have enough results to fill this thesis, and always allowed me the space to pursue my own interests and ideas. Without his encouragement, and what I consider to be a very optimistic view of my abilities, I would not have had the belief in myself to undertake this DPhil project, in an area that I had never before studied, with absolutely no previous experience in bioinformatics or evolutionary microbiology. I'd also like to acknowledge my co-supervisor Dr Danny Wilson, who tried to help me understand some pretty complex ideas, even though I was probably the least qualified person he had ever tried to explain them to.

I would also like to thank all of the members of the MacLean group (the “MacLeanies”), past and present, for their continued encouragement, kindness and acceptance. Particularly, Dr Jess Hedge, who is the kindest bioinformatics mentor I could ever have wished for. A significant portion of this thesis would have been impossible without the never-ending efforts of Dr Natalia Kapel, who fought tirelessly for every single purchase order I somehow messed up, and never once turned me away, no matter how big or small my problem was.

I would also like to acknowledge all of the kind staff and members of the Department of

Zoology, the Doctoral Training Centre and Linacre College, who were helpful and supportive through every step of my journey. I am particularly grateful to Professor Gail Preston for her consistent positive presence throughout my DPhil and her endless efforts supporting the DTC, and for the BBSRC for funding my DPhil project.

My time in Oxford would not have been the same without the friends I have made along the way. Especially those who helped to make the Oxford Hands-On Science society such a huge success, and gave me some of the most enjoyable and rewarding experiences during my time in Oxford. Generations of my family have made their living on that market place in Chesterfield, and I can honestly say that standing there and demonstrating our science experiments with all of you has been my proudest achievement.

I would like to thank my wonderful family, who listened to all of my convoluted ramblings about my experiences during this DPhil, never begrudgingly. Through every single venture I have embarked upon, I have been so lucky to have always had the unwavering support and encouragement of each and every one of you. Not once did any of you ever question whether I could actually do this, even though I questioned it myself every day. Without you all, I never would have dreamed of coming this far.

And the biggest thank you of all goes to my amazing husband-to-be, Sam, who put up with me through all of the highs and lows, the breakthroughs and the tears, and even through it all, decided he wanted to spend the rest of his life with me. Sam, I will be forever grateful that you took every step of this journey with me, and with the submission of this thesis, I look forward to beginning a new chapter of our lives together.

Contributions

All of the work in this thesis is my own, supervised by Professor R. C. MacLean and with contributions from Dr D. Wilson.

Isolate ST8F was kindly donated by Leon Peto from the Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital.

DNA sequencing was carried out by the Oxford Genomics Centre at The Wellcome Centre for Human Genetics, University of Oxford.

Contents

Abstract	iv
Acknowledgments	vi
Contributions	viii
Contents	ix
List of abbreviations	xv
Chapter 1: Introduction	1
1. Introduction.....	2
1.1 <i>Staphylococcus aureus</i>	2
1.1.1 Methicillin-resistant and multidrug-resistant <i>S. aureus</i>	3
1.1.1.1 Methicillin resistance.....	5
1.1.1.2 Multi-drug resistance.....	8
1.1.1.3 Vancomycin resistance.....	9
1.2 MRSA lineages.....	10
1.2.1 MRSA ST239.....	11
1.2.1.1 The Robinson-Enright model.....	12
1.2.2 MRSA ST8.....	14
1.2.3 MRSA ST30.....	15
1.2.4 MRSA ST22.....	16
1.3 Fitness costs of horizontal gene transfer and antimicrobial resistance.....	16
1.3.1 Horizontal gene transfer.....	16
1.3.2 Antimicrobial resistance.....	19
1.3.3 SCC <i>mec</i>	19
1.3.4 Compensatory evolution.....	20
2. Thesis objectives.....	21
Chapter 2: Investigating the epidemiology of ST239	23
Abstract.....	24
1. Introduction.....	25
1.1 Aims and objectives.....	25
1.2 Bacterial DNA sequence databases.....	25
1.2.1 Staphopia.....	26
1.2.2 <i>S. aureus</i> PubMLST.....	27
1.2.3 Realising the power of bacterial DNA sequence databases.....	28
1.3 Chapter summary.....	29
2. Methods.....	31
2.1 ST239 literature review.....	31
2.2 ST239, ST8 and ST30 sequence identification.....	31
2.3 Geographic distributions.....	32
2.4 Mapping and phylogeny.....	32
2.5 SCC <i>mec</i> type.....	32

2.6 Experimental isolates.....	33
2.6.1 DNA extraction.....	33
2.6.2 Mapping and assembly.....	34
2.6.3 SCCmec.....	35
3. Results.....	37
3.1. Evidence of the rise and fall of ST239.....	37
3.2 ST239, ST8 and ST30 sequences.....	40
3.2.1 Staphopia.....	40
3.2.2 Geographic distribution of ST239, ST8 and ST30 sequences.....	41
3.2.3 Forming diverse data-sets.....	44
3.3 SCCmec type.....	49
3.3.1 SCCmec prevalence in ST239.....	49
3.3.2 SCCmec-III prevalence in <i>S. aureus</i>	51
3.4 Experimental isolates.....	53
3.4.1 <i>De novo</i> assembly.....	55
3.4.2 Identification of SCCmec type.....	57
3.4.3 SCCmec-III variability.....	58
4. Discussion.....	60
Chapter 3: Testing the Robinson-Enright model of ST239	65
Abstract.....	66
1. Introduction.....	67
1.1 Aims and objectives.....	67
1.2 The hybridisation of ST239.....	67
1.3 Hybrid pathogenic bacteria.....	68
1.4 Chapter summary.....	71
2. Methods.....	73
2.1 Genetic analysis of ST239.....	73
2.1.1 Artemis comparison tool.....	73
2.1.2 Variable regions.....	73
2.1.3 Hybrid boundaries.....	73
2.2 Structural comparison of ST239, ST8 and ST30 complete genome sequences...	74
2.3 Extracting the hybrid regions and SCCmec region from the ST239 genomes....	74
2.4 Identifying the most likely ST239 donor.....	75
2.4.1 Genetic distance between the homologous regions of ST239, ST8 and ST30.....	75
2.4.2 Finding the closest known ancestor of the acquired region.....	75
3. Results.....	77
3.1 Genetic analysis of ST239.....	77
3.1.1 Regions of variability in ST239 genomes.....	77
3.1.2 Identification of the hybrid regions in ST239.....	80
3.2 Structural comparison of ST239, ST8 and ST30 complete genome sequences...	84
3.3 Extracting the genetic regions of ST239.....	88
3.3.1 Hybrid regions.....	88
3.3.2 SCCmec-III.....	88
3.4 Identifying the most likely ST239 donor.....	91
3.4.1 Genetic distance between the homologous regions of ST239, ST8 and ST30.....	91

3.4.2 Finding the closest known ancestor of the acquired region.....	93
4. Discussion.....	100

Chapter 4: Investigating evolution in the ST239 genome 103

Abstract.....	104
1. Introduction.....	105
1.1 Aims and objectives.....	105
1.2 The emergence of ST239.....	105
1.3 The origin of the SCCmec element.....	109
1.4 Selection pressures of HGT events.....	110
1.4.1 Positive selection and parallelism.....	111
1.4.2 The McDonald-Kreitman test for selection.....	112
1.5 Chapter summary.....	113
2. Methods.....	115
2.1. Linear regression and BactDating.....	115
2.2 BEAST.....	115
2.3 McDonald-Kreitman test.....	116
2.4 Parallelism in ST239.....	117
2.5 Tajima's D.....	117
2.6 Purine-pyrimidine and transition-transversion substitution proportions.....	117
2.7 Intergenic regions.....	118
3. Results.....	119
3.1 Dating the time to MRCA of ST239.....	119
3.1.1 Linear regression and BactDating.....	119
3.1.2 BEAST.....	120
3.2 Dating the time to MRCA of SCCmec-III.....	122
3.2.1 Linear regression.....	122
3.2.2 BEAST.....	123
3.3 Adaptation of the acquired region and the backbone region in ST239.....	125
3.3.1 Substitution rates of the genetic regions of ST239.....	125
3.3.1.1 Linear regression.....	125
3.3.1.2 BEAST.....	126
3.3.2 McDonald-Kreitman test.....	127
3.3.3 Parallelism in ST239.....	128
3.3.4 Tajima's D.....	134
3.3.5 Purine - pyrimidine and transition - transversion substitution proportions.....	134
3.3.6 Consideration of intergenic regions.....	136
4. Discussion.....	137

Chapter 5: Quantifying the competitive ability of multiple *S. aureus* strains in co-culture using whole-genome sequencing 143

Abstract.....	144
1. Introduction.....	145
1.1 Aims and objectives.....	145
1.2 Bacteria as competitors.....	145
1.3 <i>S. aureus</i> as a competitor.....	146
1.4 Chapter summary.....	147

2. Methods.....	149
2.1 Bacterial cultures.....	149
2.2 Creating a pipeline to calculate competitive ability.....	149
2.2.1 Identification of unique variant sites.....	149
2.2.2 Identification of unique genes.....	150
2.2.3 Competition experiments.....	151
2.2.4 DNA extraction and sequencing.....	152
2.2.5 Mapping.....	153
2.2.6 Base calling quality control.....	153
2.3 Estimating competitive ability from unique variant sites.....	153
2.4 Estimating competitive ability from unique genes.....	155
3. Results.....	157
3.1 Creating a pipeline to calculate competitive ability.....	157
3.1.1 Identification of unique variant sites.....	157
3.1.2 Identification of unique genes.....	158
3.1.3 Competition experiments.....	159
3.1.4 Base calling quality control.....	160
3.2 Estimating competitive ability from unique variant sites.....	161
3.2.1 Estimating competitive ability from unique variant sites in antibiotics	161
3.2.2 Estimating competitive ability from unique variant sites in growth-	
promoting media.....	164
3.3 Estimating competitive ability from unique genes.....	166
3.3.1 Estimating competitive ability from unique genes in antibiotics.....	166
3.3.2 Estimating competitive ability from unique genes in growth-	
promoting media.....	168
3.3.3 Over-represented unique genes.....	171
4. Discussion.....	173
Chapter 6: Estimating the fitness effects of ST239 hybridisation	177
Abstract.....	178
1. Introduction.....	179
1.1 Aims and objectives.....	179
1.2 Fitness effects of HGT and AMR.....	179
1.3 Fitness of ST239.....	180
1.4 Culturing <i>S. aureus</i> in growth-promoting media.....	181
1.5 AMR prediction.....	182
1.6 Chapter summary.....	184
2. Methods.....	186
2.1 Bacterial cultures.....	186
2.2 Growth rate.....	186
2.3 Competitive ability.....	186
2.4 Modelling effects.....	186
2.5 Phenotypic AMR profiles.....	187
2.6 Genetic AMR profiles.....	187
3. Results.....	188
3.1 Fitness.....	188
3.1.1 Growth rate.....	188
3.1.2 Modelling the effects on growth rate.....	190

3.1.3 Modelling the effects on competitive ability.....	193
3.1.4 Competitive ability and growth rate as independent fitness measures..	196
3.2 AMR.....	196
3.2.1 Phenotypic AMR profiles.....	197
3.2.2 Genetic AMR profiles.....	198
3.2.3 Correlating phenotypic and genotypic AMR profiles.....	201
3.2.4 Modelling the effects of AMR on fitness.....	205
4. Discussion.....	207
4.1 Fitness effects.....	207
4.2 AMR.....	208
4.3 Conclusion.....	209
Chapter 7: Conclusions	211
1. Conclusions.....	212
1.1 Conclusions.....	212
1.2 Future work.....	216
2. Concluding remarks.....	219
References	221
Appendices	251
A Supplementary figures.....	252
A.1 Genome structure comparisons.....	252
A.2 <i>SCCmec</i>	255
A.3 OD.....	256
A.4 Correlation of fitness and AMR.....	258
B Supplementary tables.....	259
B.1 Sequence metadata and mapping quality.....	259
B.2 Supplementary references.....	280
B.3 MLST profiles.....	282
B.4 Gene presence and function.....	283
B.5 McDonald-Kreitman and parallelism.....	294
B.6 Growth rate.....	296
B.7 Statistical models.....	297

List of abbreviations

- ACME – arginine catabolic mobile element
- AMR – antimicrobial resistance
- ANOVA – analysis of variance
- API – application programming interface
- BEI – Biodefense and Emerging Infections Research Resources Repository
- BHI – brain heart infusion
- BIGSI - Bitsliced genomic signature index
- bp – base pair
- CC – clonal complex
- CDC – Centers for Disease Control
- CDS – coding sequence
- CFP – cyan fluorescent protein
- CFU – colony forming units
- CI – confidence interval
- CRE – carbapenem-resistant *Enterobacteriaceae*
- df – degrees of freedom
- DNA – deoxyribonucleic acid
- dsDNA – double stranded DNA
- DSMZ – Deutsche Sammlung von Mikroorganismen und Zellkulturen
- EDTA – ethylenediaminetetraacetic acid
- EMBL-EBI – European Molecular Biology Laboratory – European Bioinformatics Institute
- EMRSA – epidemic MRSA
- ENA – European Nucleotide Archive
- GFP – green fluorescent protein
- GTR – generalised time-reversible model
- HGT – horizontal gene transfer
- HPD – highest posterior density
- ICU – intensive care unit
- MCMC – Markov chain Monte Carlo
- MGE – mobile genetic element
- MH2 – Mueller-Hinton broth
- MIC – minimum inhibitory concentration
- MLST – multi-locus sequence type
- MRCA - most recent common ancestor
- MRSA – methicillin resistant *S. aureus*
- MSSA – methicillin sensitive *S. aureus*
- NCBI – National Center for Biotechnology Information
- NICE – National Institute for Health and Care Excellence
- PBP – penicillin binding protein
- PBS – phosphate buffered saline
- PCR – polymerase chain reaction
- PFGE – Pulsed-field gel electrophoresis
- PHE – Public Health England
- PS – porcine serum
- qPCR – quantitative PCR
- RNA – ribonucleic acid
- RPM – revolutions per minute
- SCCHg – *Staphylococcal* cassette chromosome mercury
- SCCmec – *Staphylococcal* cassette chromosome *mec*
- SNP – single nucleotide polymorphism
- SRA – short read archive
- ST – sequence type
- Tris – tris(hydroxymethyl)aminomethane
- TSA – tryptic soy agar
- TSB – tryptic soy broth
- VCF – variant call format
- VISA – vancomycin intermediate *S. aureus*
- VRSA – vancomycin resistant *S. aureus*
- WGS – whole genome sequencing
- XML – extensible markup language
- YFP – yellow fluorescent protein

Chapter 1

Introduction

1. Introduction

1.1 *Staphylococcus aureus*

Staphylococcus aureus is a common human pathogen, and a leading cause of healthcare-associated infections worldwide, causing potentially fatal bacteraemia, pneumonia, and skin and soft tissue infections (Ippolito *et al.*, 2010). Livestock-associated *S. aureus* infections are also a major concern. For example, *S. aureus* is the biggest cause of mastitis in dairy cattle leading to depressed milk production, and is estimated to cause annual global economic losses of \$533 billion (Shaheen *et al.*, 2016). Additionally, *S. aureus* is a common human commensal that is typically carried in the nares of around 27% of healthy individuals, as well as on skin and other body sites (Wertheim *et al.*, 2005).

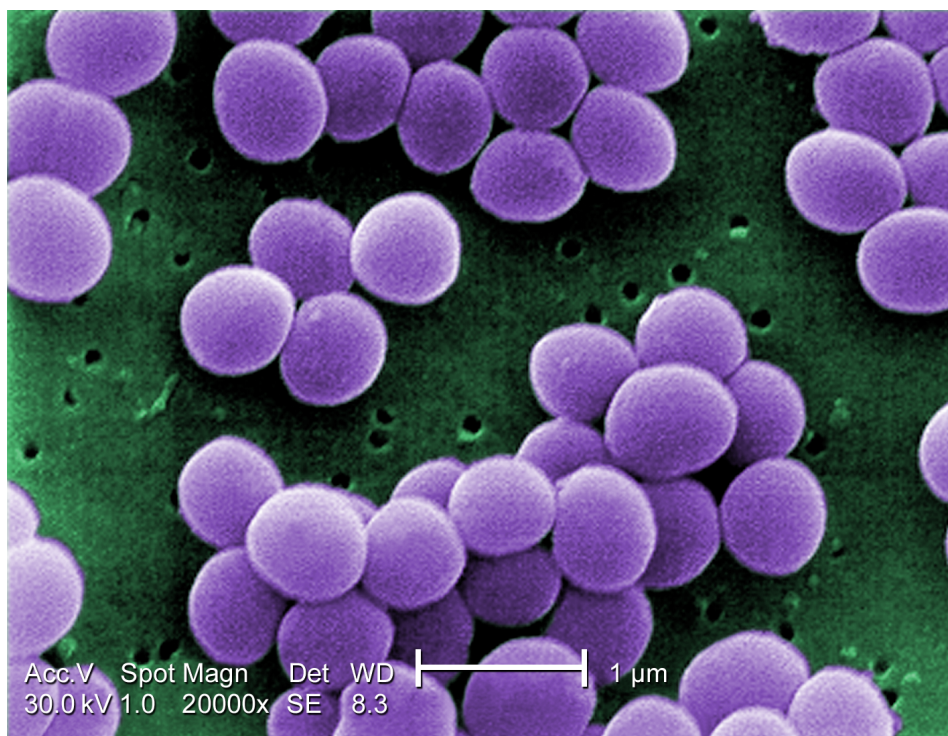


Figure 1.1. False-colour scanning electron micrograph of *S. aureus* colonies under 20,000x magnification (image from Centers for Disease Control and Prevention Public Health Image Library, 2001).

Bacteria from the genus *Staphylococcus* were first identified as pathogenic in 1880 by Sir Alexander Ogston, a Scottish surgeon who isolated bacteria from the pus of a surgical abscess in a knee joint (Licitra, 2013). The genus was named due to its distinctive shape (Figure 1.1), *staphyle* from the Greek for “bunch of grapes” and *kokkos* from the Greek for “berry”.

In 1884, German physician Friedrich Julius Rosenbach was able to isolate and distinguish two different species of *Staphylococcus* based on the colour of the colonies, which he named *aureus*, from the Latin for “gold”, and *albus*, from the Latin for “white”. *S. albus* was later renamed *S. epidermidis* to reflect its ubiquity on human skin (Licitra, 2013).

1.1.1 Methicillin-resistant and multidrug-resistant *S. aureus*

S. aureus is a highly successful, Gram-positive, human and animal pathogen that is described as a “high priority” for the development of new antibiotics by the World Health Organisation (World Health Organisation, 2017). It has been able to adapt to the antibiotic era by evolving resistance to all antibiotics (Foster, 2017).

Although the antimicrobial properties of *Penicillium* mould had been recorded long before Sir Alexander Fleming’s infamous serendipitous discovery in 1928 (Fleming, 1929), for example by Ernest Duchesne in the 1890s (Duckett, 1999), it was Fleming’s work that prompted the war-efforts of researchers such as Florey and Chain, which led to the introduction of penicillin into clinics in 1941, as the first antibiotic to treat infection (Abraham, 1941). *S. aureus* that was resistant to penicillin was described just one year later (Rammelkamp and Maxon, 1942). By the end of the 1940s, over half of staphylococcal infections in a large hospital in the UK were resistant to penicillin (Barber and Rozwadowska-Dowzenko, 1948), and by the late

Chapter 1

1950s several cases had been reported of *S. aureus* that were resistant to most antibiotics (Jessen *et al.*, 1969).

To combat this, new antibiotics were developed including methicillin, a second-generation penicillin antibiotic, which was released onto the market in October 1960. Six months later, methicillin-resistant *S. aureus* (MRSA) was reported in a hospital in South England (Jevons, 1961). By 1967, multidrug-resistant MRSA had been reported in Denmark, France, Switzerland, the UK and the USA (Benner, 1968), as well as in Australia (Rountree and Beard, 1968) and Asia (Pal and Ray, 1964). Rapid dissemination of MRSA around the globe followed (Grundmann *et al.*, 2006).

However, in the 1970s, reports of MRSA began to decrease in Europe. No single reason for this decline has been identified, although changes in streptomycin and tetracycline prescribing, the introduction of more rigorous infection-control policies, and the significant fitness costs of highly multidrug-resistant MRSA strains, may have contributed (Grundmann *et al.*, 2006; Nielsen *et al.*, 2012). Unfortunately, a dramatic increase in MRSA infections in Europe and the USA followed in the early 1980s (Stryjewski and Corey, 2014). A similar rise was subsequently observed in the Asia-Pacific region (Bell *et al.*, 2002) and in Latin America (Guzman-Blanco *et al.*, 2009).

More recently, reports indicate that the rate of MRSA infection has remained static or has begun to decrease (Stryjewski and Corey, 2014). In 2017, the Centers for Disease Control recorded 119,000 cases of *S. aureus* infections in the USA, resulting in almost 20,000 deaths (Kavanagh, 2019). This reflected an overall decrease in hospital-acquired infections of 74% between 2005 – 2016 (Kourtis *et al.*, 2019). Similarly, decreasing rates of MRSA infections

have been noted around the world (Dai *et al.*, 2018; Nikolaras *et al.*, 2019). However, MRSA still poses a significant public-health threat due to multi-drug resistance, and morbidity and mortality remain persistently high (Turner *et al.*, 2019). Although several candidates for an MRSA vaccine have been identified, none have successfully passed clinical trials (Turner *et al.*, 2019; Pfizer, 2018).

MRSA has typically been considered as a healthcare-associated pathogen, that occurs in patients with established risk-factors (Naimi *et al.*, 2003). However, since the 1990s, cases of community-acquired MRSA have been rising dramatically. These infections often occur in previously healthy individuals and in younger patients, and are typically associated with skin and soft tissue infections (Kong *et al.*, 2016). Although these community-associated infections are typically susceptible to non- β -lactam antibiotics, multi-drug resistant strains have been described, and reports of multi-drug resistant community-associated MRSA infections arising from healthcare settings are becoming much more common (Chambers and DeLeo, 2009; Chua *et al.*, 2011; Kong *et al.*, 2016). Several dominant community-associated MRSA strains have emerged internationally, which are typically prevalent in distinct geographic locations, although they are more rarely identified outside of these regions. For example, since the 2000s USA300 (ST8) has been the dominant community-associated MRSA strain in North America. In Western Europe, ST80 is a more common community-associated MRSA, and in Australia, ST93 and the Southwest Pacific clone of ST30 are both dominant MRSA strains in the community (David and Daum, 2010).

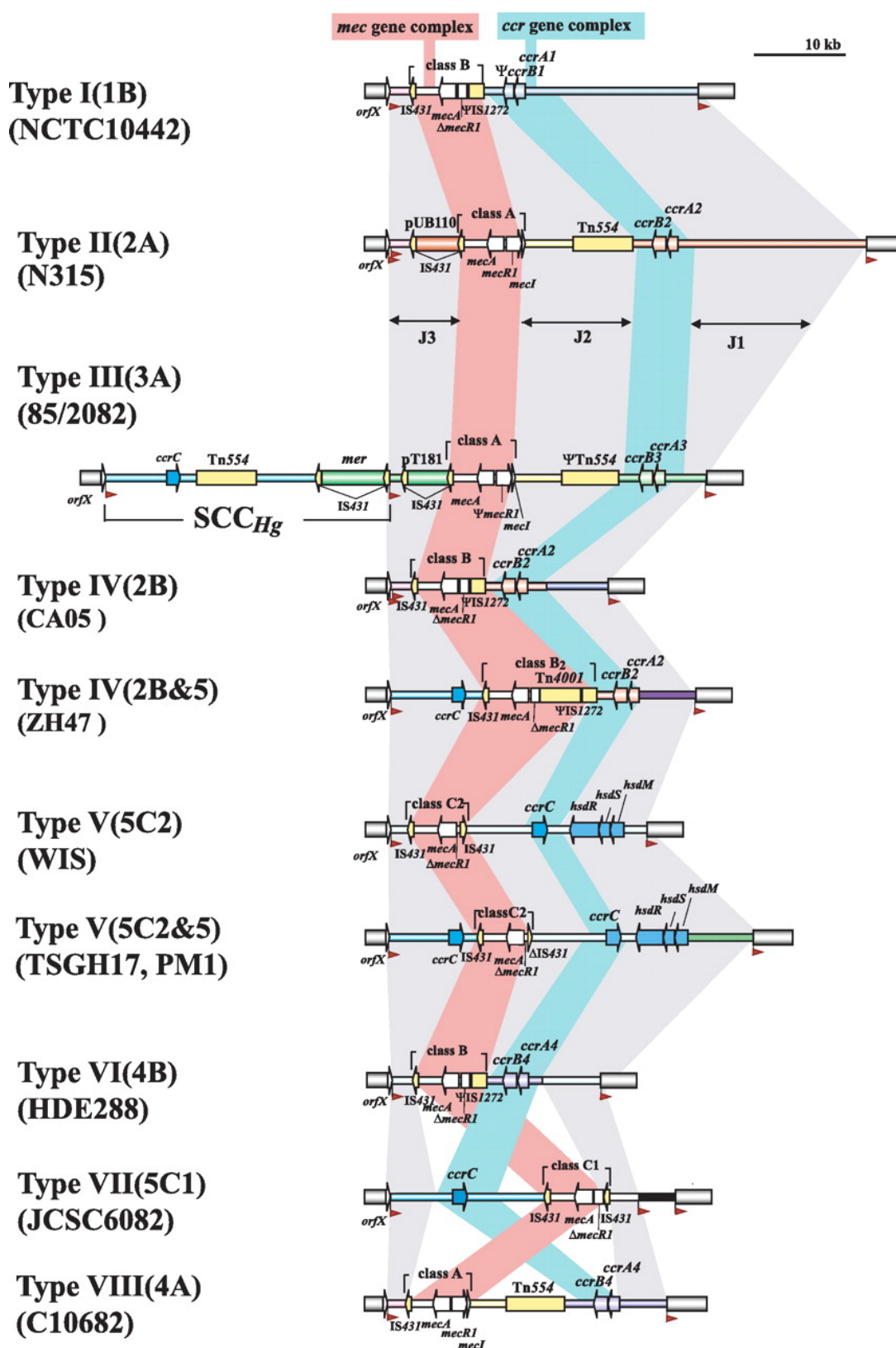
1.1.1.1 Methicillin resistance

Penicillin resistance in *S. aureus* is primarily conferred by the *blaZ* gene, which can be carried by plasmids or encoded on the chromosome (Olsen *et al.*, 2006). The *blaZ* gene encodes a β -

lactamase enzyme that inactivates penicillin through hydrolysis of the β -lactam ring (Rolinson *et al.*, 1960).

Methicillin, a second generation β -lactam alternative to penicillin, is effective against β -lactamase producing penicillin-resistant *S. aureus*. Resistance to methicillin is conferred through the expression of penicillin-binding protein PBP2a, which is typically encoded by the *mecA* gene and is induced by β -lactam binding to the *mecR1* regulator (Olsen *et al.*, 2006). PBP2a has a lower binding affinity to β -lactams than the endogenous *S. aureus* PBP, and therefore decreases susceptibility to β -lactams including penicillin and methicillin (Stryjewski and Corey, 2014). Resistance to methicillin is encoded in the mobile genetic element (MGE) Staphylococcal Chromosome Cassette *SCCmec*, which typically carries both the *mec* operon as well as site-specific recombinase genes *ccrAB/ccrC* (Liu *et al.*, 2016). The CcrAB/CcrC recombinases accurately excise the *SCCmec* element, which can then be acquired by methicillin-susceptible *S. aureus* (MSSA) strains and integrated into the chromosome at the *attB* insertion site near to the origin of replication, leading to the spread of methicillin resistance (Liu *et al.*, 2016). Horizontal transfer of *mec* into distantly-related lineages of *S. aureus* has been highly significant in the evolution of MRSA (Enright *et al.*, 2002), and several major MRSA clones have arisen separately from successful epidemic MSSA strains on multiple occasions (Stryjewski and Corey, 2014).

The first *SCCmec* element, *SCCmec-I*, was identified in 2000 (Katayama *et al.*, 2000) from the MRSA strain N315, which was isolated in Japan in the early 1980s (Okonogi *et al.*, 1989). Subsequently, as of October 2019, 14 different classes of *SCCmec* have been identified in *S. aureus* ranging from *SCCmec-I* to *SCCmec-XIV* (Urushibara *et al.*, 2019), based on the



nature of the *mec* and *ccr* gene complexes (Figure 1.2). *SCCmec* elements are highly variable, and are further sub-typed by the variable J-region segments (Liu *et al.*, 2016). For example, there are 79 different sub-types of *SCCmec*-IV (Liu *et al.*, 2010) and 39 different sub-types of *SCCmec*-III (Monecke *et al.*, 2018).

In addition to methicillin resistance, 27% of all known *S. aureus* AMR genes are carried on the *SCCmec* element, including resistance to β -lactams, aminoglycosides and macrolides (John *et al.*, 2019). Additionally, heavy-metal resistance genes can also be carried within the *SCCmec* element. For example, *SCCmec*-III is one of the largest *SCCmec* elements at >60 kb in length and carries *SCCmec* in tandem with *SCCHg*, which encodes mercury resistance (Ito, 2001). *SCCmec*-III also carries multiple antibiotic resistance genes through integrated plasmids in the J-regions, encoding resistances such as tetracycline, erythromycin and spectinomycin, as well as cadmium resistance (Lakhundi and Zhang, 2018).

1.1.1.2 Multi-drug resistance

MRSA infections pose a serious challenge in healthcare due to increasing resistance to multiple antibiotics (Lakhundi and Zhang, 2018).

Early MRSA infections isolated in the 1960s and early 1970s typically remained susceptible to non- β -lactam antibiotics, with the exception of some tetracycline-, streptomycin-, and erythromycin-resistant infections (Grubb, 1998). However, in the late 1970s, strains of MRSA that were resistant to multiple non- β -lactam antibiotics began to emerge. These strains were able to spread over several continents, causing serious, primarily hospital-associated, multi-drug resistant MRSA infections around the globe. These strains are described as epidemic MRSAs (EMRSAs).

Resistance determinants to all classes of antibiotics have been identified in *S. aureus* (Foster, 2017). Many of these can be carried on MGEs, such as aminoglycoside resistance, which is typically conveyed through inactivation of aminoglycosides by aminoglycoside-modification enzymes (Khosravi *et al.*, 2017). Alternatively, resistance can be conveyed through the expulsion of antibiotics via increased expression of multi-drug efflux pumps, such as *mepA*, which is known to be associated with resistance to several classes of antibiotics (Kaatz *et al.*, 2005). Finally, resistance can occur as a result of point mutations, such resistance to rifampicin, which occurs as a consequence of mutations in the *rpoB* gene (Aubry-Damon *et al.*, 1998).

1.1.1.3 Vancomycin resistance

Vancomycin is a glycopeptide antibiotic that was first approved for use in humans in 1958. The antibiotic interferes with peptidoglycan synthesis in Gram-positive bacteria, inhibiting cell wall synthesis (McGuinness *et al.*, 2017).

In the late 1980s, vancomycin began to be commonly used for the treatment of MRSA infections. The first *S. aureus* clinical isolate with reduced susceptibility to glycopeptides, known as vancomycin-intermediate *S. aureus* (VISA), was isolated from a patient in Japan in 1996 (Hiramatsu *et al.*, 1997). The first completely vancomycin-resistant *S. aureus* (VRSA) was reported in 2002 in the USA (CDC, 2002), and several further examples of VRSA have since been reported (McGuinness *et al.*, 2017).

Vancomycin resistance in *S. aureus* is conferred by the *vanA* operon, which is encoded on transposon Tn1546, which originated from a vancomycin-resistant *Enterococci* (VRE)

conjugative plasmid (Arthur, 1993). The *van* operon is comprised of *vanA*, *vanH*, *vanX*, *vanY*, and *vanZ* genes, and the two-component sensor-regulator system genes *vanS* and *vanR*, which activate transcription of the operon in the presence of vancomycin. Several genes and mutations are known to contribute to the development of VISA, including the two-component regulatory systems *graRS* and *walKR*, and the DNA-dependent RNA polymerase β -subunit *rpoB* (McGuinness *et al.*, 2017).

Despite increasing reports of VRSA, vancomycin remains one of the main antibiotics for the treatment of invasive MRSA infections, and hence there are significant concerns related to the decreasing susceptibility of MRSA to this antibiotic (Choo and Chambers, 2016).

1.2 MRSA lineages

The first *S. aureus* whole-genome sequences were published in 2001, and demonstrated a complex mixture of genes in the *S. aureus* genome, alongside substantial evidence of lateral gene transfer (Kuroda *et al.*, 2001). Subsequently, whole-genome sequencing has been able to capture the diversity of tens of thousands of *S. aureus* genomes (Petit *et al.*, 2018). There are many *S. aureus* lineages, which are highly diverse, however isolates belonging to the same lineage are typically well conserved (Lindsay, 2014). In 2000, a multi-locus sequence type (MLST) scheme was developed to describe the different *S. aureus* lineages (Enright *et al.*, 2000). The MLST typing scheme is based on the different allelic profiles of seven housekeeping genes; carbamate kinase (*arcC*), shikimate dehydrogenase (*aroE*), glycerol kinase (*glp*), guanylate kinase (*gmk*), phosphate acetyltransferase (*pta*), triosephosphate isomerase (*tpi*), and acetyl coenzyme A acetyltransferase (*yqiL*). Closely related MLST strains are categorised into clonal complexes (CCs). Over 70% of known *S. aureus* genomes can be

categorised into just seven predominant CCs; CC5, CC8, CC15, CC22, CC30, CC45, and CC398 (Petit *et al.*, 2018). The most prevalent *S. aureus* ST in online sequencing repositories is ST22, with over 6,000 ST22 sequences, which are mostly resistant to methicillin (Petit *et al.*, 2018).

1.2.1 MRSA ST239

ST239 is a major CC8 EMRSA strain that has been the causative agent of multiple epidemics in healthcare settings around the globe (Baines *et al.*, 2015). Several major EMRSA strains have been assigned to ST239, including EMRSA-1, EMRSA-4, EMRSA-7, and EMRSA-9 (Ko *et al.*, 2005). Although community-associated and commensal ST239 have been recorded, ST239 is primarily associated with bacteraemia or vascular devices, and is therefore typically a healthcare-associated MRSA strain (Song *et al.*, 2011; Holden *et al.*, 2010).

Several studies have estimated the initial emergence date of ST239 to around the mid-to-late 1960s (Harris *et al.*, 2010; Gray *et al.*, 2010; Didelot *et al.*, 2019). The first ST239 isolates were collected in the late 1970s in Australia (Pavillard *et al.*, 1982). This was followed by the first EMRSA, an ST239 strain known as EMRSA-1, which emerged in the UK in 1981 (Cookson and Phillips, 1988). Between 1987 – 1988, over 40% of MRSA isolates collected in England and Wales were ST239 EMRSA-1 (Kerr *et al.*, 1990) and ST239 became highly prevalent throughout Europe (Lakhundi and Zhang, 2018).

In Brazil between 1992 – 1994, ST239 was responsible for 70% of MRSA infections (Teixeira *et al.*, 1995). The ST239 Brazilian strain spread across Latin America to Argentina, Colombia, Ecuador, Chile, Uruguay and Paraguay (Rodriguez-Noriega and Seas, 2010). A collection of MRSA isolates from the Asian Bacterial Bank that were isolated between 1998

and 2003 suggested that ST239 (or ST241, which is a single-locus variant of ST239) was the dominant MRSA clones in China, India, Indonesia, the Philippines, Saudi Arabia, Singapore, Sri Lanka, Taiwan, Thailand, and Vietnam (Ko *et al.*, 2005). Several outbreaks of ST239 have also occurred in Australia, and in 2011, ST239 accounted for 49.3% of all hospital-associated MRSA in Australia (Coombs *et al.*, 2013). Reports of ST239 in the USA and Africa have been relatively rare (Wang *et al.*, 2012; Abdulgader *et al.*, 2015). More recently, reports from around the globe suggest that levels of ST239 MRSA have been decreasing dramatically (Amorim *et al.*, 2007; Conceicao *et al.*, 2007; Hu *et al.*, 2018).

1.2.1.1 The Robinson-Enright model

In 2004, Robinson and Enright published their hypothesis that ST239 is a hybrid bacteria, and that its genome had formed through a single large chromosomal replacement event (Robinson and Enright, 2004). A range of 41 genes across the ST239 genome were sequenced, and roughly half of these were closely related to the MRSA strain ST30. The remaining genes were closely related to the MRSA strain ST8. Further sequencing of the hybrid boundaries led Robinson and Enright to hypothesise that ST239 had formed through a natural hybridisation event, where ~557 kb spanning the origin of replication (*oriC*) was acquired from the ST30-like parent, and the remaining ~2,220 kb originated from the ST8-like parent (Figure 1.3) (Robinson and Enright, 2004).

In 2010, Holden published the first ST239 whole-genome sequence, from a strain that was isolated from an ICU in a London hospital in 2003, called ST239 TW20 (Edgeworth *et al.*, 2007; Holden *et al.*, 2010). Holden compared the ~635 kb ST30-like region of the TW20 genome to the CC30 MRSA252 complete genome sequence, and confirmed that this region

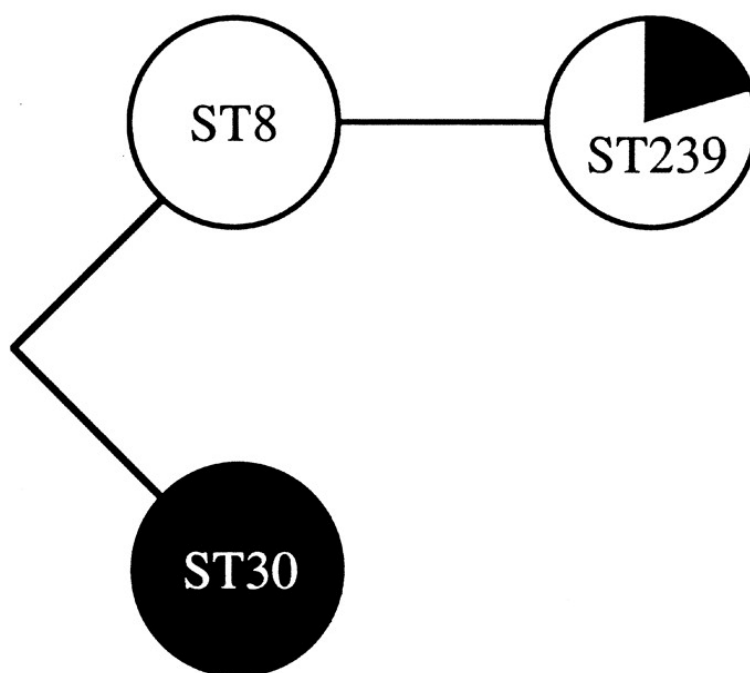


Figure 1.3. The Robinson-Enright model for the hypothesised evolutionary relationships of STs involved in the chromosomal replacement that formed the ST239 hybrid. The approximate contributions of each of the ST8 and ST30 parent chromosomes to the ST239 hybrid chromosome are shown with black and white sectors. Image from Robinson and Enright, 2004.

was more closely related to CC30 by a shift of roughly 1% in DNA percentage identity, compared to the ~2.5 Mb ST8-like region of TW20. This supported the Robinson-Enright hypothesis. Castillo-Ramirez (2011) was able to explicitly state the location of the genetic boundary between the ST30-like and the ST8-like regions in the TW20 genome, characterising the ST30-like region as a continuous 622,898 bp region spanning the origin of replication, located between bases 2,848,037 – 3,043,210 and 1 – 427,725, and the ST8-like region as a 2,420,310 bp region located between bases 427,726 – 2,848,036 (Castillo-Ramirez *et al.*, 2011).

Although Robinson and Enright could not identify the mechanism of the genetic exchange that formed the ST239 hybrid, they concluded that conjugative plasmids, transposons and

phage-mediated exchange were unlikely mechanisms because the exchanged region contains essential genetic material rather than accessory genetic material that is typically carried on MGEs (Robinson and Enright, 2004). Lindsay (2014) proposed that ST239 was more likely to have been formed by a natural transformation event via bacterially encoded competence proteins, which have been shown to be capable of transferring very large regions of DNA in *S. aureus* that would not fit in transducing phage heads (Lindsay, 2014; Morikawa *et al.*, 2012). Alternatively, the phage-mediated transduction of large regions of bacterial chromosomal DNA has been reported. For example, P1 phage can transfer up to 100 kb of *E. coli* DNA (Thomason *et al.*, 2007), and the G jumbo phage in *Bacillus subtilis*, which has one of the largest phage capsids known, can contain over 600 kb of DNA in its head (Hua *et al.*, 2017). However, there is no evidence of jumbo phage in the ST239 genome.

Robinson and Enright also noted that ST239 contains SCCmec-III, which is not typically carried by either the ST8 or ST30 parent (Robinson and Enright, 2004). ST8 and ST30 most commonly carry SCCmec-IV, however they can carry other SCCmec types, or none at all. Therefore, Robinson and Enright further hypothesised that SCCmec-III may have been acquired by ST239 in a separate HGT event, either before or after the hybridisation occurred.

1.2.2 MRSA ST8

The major hypothesised ST239 ancestor, ST8, has been responsible for several healthcare-associated epidemics, including EMRSA-2. This EMRSA emerged in the UK in the late 1980s (Richardson and Marples, 1988) and typically carried SCCmec-IV (Enright, 2003).

In the early 2000s, an aggressive ST8 community-associated MRSA strain emerged in the USA, known as USA300, also carrying SCCmec-IV (CDC, 2001; Planet, 2017). By the mid-

2000s, USA300 was the major cause of all skin and soft-tissue infections in the USA (Talan *et al.*, 2011), as well as a significant cause of invasive infections, such as bacteraemia (Seybold *et al.*, 2006), thus increasing rates of MRSA infection and significantly adding to the overall burden of disease (Planet, 2017).

Both community-associated and healthcare-associated USA300 remain a major healthcare challenge in the USA, however the expansion of the USA300 strain is beginning to slow (Planet, 2017; Turner *et al.*, 2019). Despite multiple reported USA300 infections in countries around the globe through numerous separate introduction events (Toleman *et al.*, 2016; Glaser *et al.*, 2016), USA300 has not become established outside of North America and northern South America (Planet 2017).

1.2.3 MRSA ST30

ST30 is the hypothesised minor ancestor of ST239. Between the mid-1990s and early 2000s, the dominant community-associated MRSA strain in New Zealand was an ST30 strain known as the South West Pacific strain, which typically carries SCC*mec*-IV (Williamson *et al.*, 2014). It has become one of the primary community-associated MRSA strains in countries in Oceania (Williamson *et al.*, 2014), Asia (Hsu *et al.*, 2006; Isobe *et al.*, 2012), South America (Fernandez *et al.*, 2013), Europe (Bartels *et al.*, 2010) and the Middle East (Mediavilla *et al.*, 2012).

The South West Pacific strain was first isolated in the community in Auckland, New Zealand in 1992, from individuals who had contact with Western Samoa (Mitchell *et al.*, 1996). Although it has since been replaced as the dominant New Zealand community-associated

strain by ST5-IV, which has additional fusidic acid resistance, the South West Pacific strain has remained responsible for a significant proportion of community-associated and healthcare-associated MRSA infections in New Zealand (Williamson *et al.*, 2014).

ST30 is hypothesised to have diverged from a recent common CC30 ancestor with the healthcare-associated EMRSA-16 strain, which is an ST36 strain that typically carries SCCmec-II (DeLeo *et al.*, 2011). EMRSA-16 was first reported in 1991 in the UK (Cox *et al.*, 1995), and has been one of the most successful EMRSA clones of the antibiotic era.

1.2.4 MRSA ST22

In recent years, MRSA ST22 has begun to replace ST239 and other prevalent healthcare-associated MRSA strains around the world (Hsu *et al.*, 2005a; Amorim *et al.*, 2007; Aires-de-Sousa *et al.*, 2008; Witte *et al.*, 2008). ST22 includes EMRSA-15, which carries an SCCmec-IV element, which is the most frequently identified strain of MRSA in Europe (Grundmann *et al.*, 2010). EMRSA-15 was initially isolated in the UK in 1991 (Richardson and Reith, 1993), and subsequently has been identified in countries around the globe, including as a cause of community-associated infection (Holden *et al.*, 2013). An analysis of 193 ST22 isolates from around the world by Holden (2013) estimated the date of emergence of the pandemic ST22 population to England in the 1980s, which originated from a community-associated MSSA population (Holden *et al.*, 2013).

1.3 Fitness costs of horizontal gene transfer and antimicrobial resistance

1.3.1 Horizontal gene transfer

Horizontal gene transfer (HGT) is the movement of genetic material across strain and species

boundaries, and can confer novel, beneficial phenotypes to the recipient (Baltrus, 2013). HGT is therefore an important evolutionary force, particularly in bacteria, where mechanisms of HGT include conjugation, transduction, and transformation (Gyles and Boerlin, 2013). HGT events contribute substantially to pan-genome variation within species, and are therefore key to the successful adaptation of bacterial species to new niches (Brockhurst *et al.*, 2019).

The drivers and mechanisms of HGT are complex and scientific understanding of important HGT processes is still incomplete. Recent observations of novel HGT processes in *S. aureus* include hyper-mobile regions of the *S. aureus* genome that are due to the presence of lateral-transducing *S. aureus* prophages (Chen *et al.*, 2018), and the presence of auto-transducing phage, which are able to acquire beneficial genes from competing strains within the same environment (Haaber *et al.*, 2016).

MGEs can be transferred at high frequency between individuals via HGT. In *S. aureus* genomes, MGEs make up 15 – 20% of the genome, and are highly diverse compared to the *S. aureus* core genome. These MGEs often encode genes involved in AMR, virulence and host-adaptation, and include plasmids, transposons, bacteriophage, *S. aureus* pathogenicity islands, and the SCCmec element (Lindsay, 2014). The hypothesised large chromosomal replacement event that formed the ST239 hybrid is an extreme example of a HGT event, where 20% of the ST30 genome was transferred horizontally to replace the homologous region in ST8 (Robinson and Enright, 2004). Similar large HGT have resulted in other hybrid *S. aureus* lineages, including ST42 (Robinson and Enright, 2004), ST34 (Thomas *et al.*, 2012) and ST2249 (Nimmo *et al.*, 2014), ST582 (Didelot and Wilson, 2015) and ST71 (Spoor *et al.*, 2015), as well as hybrid lineages of other pathogenic bacterial species, such as *Klebsiella*

pneumonia ST258 (Chen *et al.*, 2014), *Campylobacter coli* ST1150 (Sheppard *et al.*, 2013) and *Streptococcus agalactiae* (Brochet *et al.*, 2008).

Although HGT is often associated with the acquisition of beneficial phenotypic traits, it also commonly results in a fitness cost to the recipient, which can manifest as reduced growth rate or reduced competitive ability. For example, recently acquired genetic regions often function inefficiently within new genomic backgrounds, potentially disrupting existing regulatory and physiological networks (Baltrus, 2013). Therefore, HGT can cause disruption to interactions within complex cellular networks, as genes in the transferred region have not co-evolved with the pathways in the recipient genome. Additionally, the bio-synthetic cost of expressing and maintaining additional DNA, epistatic interactions between chromosomal and MGE genes, and the expression of toxic gene products, can also negatively impact recipient fitness (Brockhurst *et al.*, 2019). Overall, the fitness costs associated with the acquisition of MGEs are complex, as MGEs are able to replicate via both vertical and horizontal transmission, and therefore have fitness interests that do not necessarily align with the vertically-transmissible recipient chromosome (Brockhurst *et al.*, 2019). Finally, MGEs that act as genetic parasites can encode no beneficial accessory genes, and yet are still able to persist in the population (Brown *et al.*, 2013).

Robinson and Enright suggested that negative effects caused by the mixing of large portions of unrelated genomes may have resulted in ST239 being relatively unfit compared to its hypothesised parent strains (Robinson and Enright, 2004). This negative impact may translate into decreased growth rate, for example, the doubling time of ST239 is roughly 1.4x longer than that of ST30 and ST8 (Okuma *et al.*, 2002).

1.3.2 Antimicrobial resistance

Antimicrobial resistance (AMR) can arise through *de novo* mutations, for example, leading to increases in drug efflux or alterations to the antibiotic target site, or via HGT, which may, for example, encode drug modification enzymes (Andersson and Hughes, 2010). Experimental techniques to study AMR have typically focused on the evolution of resistance via chromosomal mutations, however, clinical microbial sequencing studies have shown that the acquisition of MGEs is the main mechanism behind AMR acquisition, which poses the greatest public health threat (MacLean and San Millan, 2019).

AMR is also often associated with a fitness cost (Andersson and Levin, 1999). This is because antibiotics typically target important cellular functions, such as cell wall synthesis, regulation of chromosome super-coiling, RNA transcription and protein synthesis, and therefore chromosomal AMR mutations also impact these vital processes (Andersson and Hughes, 2010). The fitness costs of AMR can be highly variable, as they are influenced by a wide variety of factors (Vogwill and MacLean, 2014).

1.3.3 SCC*mec*

In ST239, carriage of the large SCC*mec*-III element may impose an additional fitness cost, due to the maintenance of the large genetic element and accommodation of the alternative PBP2a into the staphylococcal cell wall synthesis complex (Ender *et al.*, 2004), as well as fitness costs imposed by other AMR genes carried by the SCC*mec*-III element. However, there is evidence that smaller SCC*mec* elements, such as those typically associated with community-acquired MRSA, carry a lesser fitness cost. Lee (2007) established that although the larger SCC*mec* elements, such as SCC*mec*-I, do impose a fitness cost to their host that

manifests as a reduced growth rate, the smaller SCCmec-IV element carries no fitness cost (Lee *et al.*, 2007). Knight (2012) observed a similar fitness cost of the SCCmec-II element that was not detected in *S. aureus* isolates that carried SCCmec-IV (Knight *et al.*, 2013). Lee (2007) suggested that the MecA protein encoded in SCCmec-IV, which has an altered ribosome binding site that produces less PBP2a, may contribute to the reduced fitness cost of SCCmec-IV. MRSA strains that carry smaller SCCmec elements, such as SCCmec-IV and SCCmec-V, are therefore more able to persist under non-antibiotic selective pressures, such as in the community. As these strains move into healthcare settings, they have been seen to out-compete less fit well-established healthcare-associated strains that are associated with larger SCCmec elements in clinical settings (Hsu *et al.*, 2007a; Knight *et al.*, 2012).

1.3.4 Compensatory evolution

Over time, compensatory evolution can reduce the fitness costs associated with HGT or AMR. Compensatory adaptations can effectively restore the efficiency of affected functions, or substitute the affected function with an alternative function. This has many consequences for the acquisition and maintenance of AMR, for example, by stabilising resistant populations in the absence of antibiotics without loss of resistance (Andersson and Hughes, 2010), and allowing AMR plasmids to persist in populations (San Millan and MacLean, 2017). This has massive implications for the reversal of AMR in healthcare.

2. Thesis objectives

In this thesis, I aim to explore the evolutionary drivers that have contributed to the formation, global emergence, and eventual decline of the hybrid MRSA ST239.

In Chapter 2, I aim to describe the global distribution of recorded ST239 isolates, and establish how prevalent the *SCCmec*-III element is within ST239 and the wider *S. aureus* population. Additionally, I will collect a diverse range of ST239, ST8 and ST30 sequences from publicly available sequencing repositories, as well as experimental isolates. These diverse data-sets and isolates will be the basis of the further chapters in this thesis.

In Chapter 3, I aim to analyse the separate genetic regions that comprise the ST239 genome, using the increased diversity of ST239 sequences in publicly available sequencing repositories, in order to test whether their evolution is consistent with the Robinson-Enright model. I also aim to identify the closest known ancestor of the true donor of the acquired region of the ST239 genome.

In Chapter 4, I will estimate the time to MRCA of the ST239 genomes, as an estimate of the date of the hybridisation event that formed ST239. I also aim to establish whether the *SCCmec*-III element was acquired prior to, or subsequently to, the hybridisation. Additionally, I will investigate the effect of compensatory adaptation in both the acquired and backbone regions, and explore the contribution of selection to the evolution of the ST239 genome.

Chapter 1

In Chapter 5, I aim to develop a method to calculate the competitive ability of the *S. aureus* isolates in co-culture, which can be used as a measure of bacterial fitness in Chapter 6. This method will apply whole-genome sequencing to determine the depth of coverage at variant sites and genes that are unique to individual isolates in a range of growth-limiting and growth-promoting conditions.

In Chapter 6, I aim to assess the fitness and AMR levels of the *S. aureus* isolates, to investigate the hypothesis that the large chromosomal replacement event that formed ST239 is associated with decreased fitness.

Chapter 2

Investigating the epidemiology of ST239

Abstract

For several decades, ST239 has been a dominant MRSA strain in healthcare settings around the globe, and has been the causative agent of major epidemics. In this chapter, I find that, although epidemiological data surrounding ST239 is sparse, there are records of ST239 in the scientific literature and in online sequencing repositories that describe ST239 in multiple countries on all inhabited continents since the late 1970s. This is evidence that ST239 MRSA is extremely widespread and has had a major impact in healthcare settings around the world. However, there is also evidence of MRSA strains that have successfully replaced ST239 as the dominant MRSA in clinical environments, most commonly ST22, which contains a small *SCCmec*-IV element that encodes methicillin resistance but has no fitness cost. ST239 is closely linked to the large ~60 kb *SCCmec*-III element, which was present in 94% of ST239 sequences in the Staphopia database. No alternative *SCCmec* elements were found in any ST239 sequence. *SCCmec*-III was identified in three *S. aureus* isolates that are not closely related to ST239 (ST5, ST21 and ST101), suggesting that there may be some limited mobility of this large genetic element. ST239, ST8 and ST30 isolates are highly represented in publicly-available online bacterial sequencing repositories, which were used to create a geographically and temporally diverse data-set of 96 ST239, 57 ST30 and 111 ST8 sequences, alongside 15 ST239, ST8 and CC30 isolates that were obtained and sequenced, to be used in further chapters of this thesis. By combining individual reports of ST239 from around the world to create a detailed overview, trends in *SCCmec* type prevalence and strain-replacement have been identified, which may help to anticipate the behaviour of current and future ST239 epidemics.

1. Introduction

1.1 Aims and objectives

In this chapter, I aim to describe the distribution of recorded ST239 isolates around the world, and characterise the prevalence of the large, AMR-linked SCC*mec*-III element within ST239 and within the wider *S. aureus* population. Additionally, I aim to collate a diverse data-set of ST239, ST8 and ST30 sequences and isolates. Where possible, these will be linked to both geographic location and isolation date metadata. These diverse data-sets and isolates will be used in future chapters to make robust evolutionary comparisons between ST239 and the closest relatives of its hypothesised parent strains.

1.2 Bacterial DNA sequence databases

During the last few decades, the number of publicly-available bacterial DNA sequences in online repositories, such as the National Center for Biotechnology Information (NCBI) and the European Nucleotide Archive (ENA), has been increasing. *S. aureus* sequences make up a large number of the bacterial DNA sequences in these repositories. For example, the number of *S. aureus* Illumina shotgun genome sequences in the NCBI Short Read Archive (NCBI SRA) and the ENA was almost 50,000, as of March 2018 (Petit *et al.*, 2018). However, the ability to search these databases for specific criteria or metadata on a large scale is very limited. Therefore, the development and utilisation of methods to easily query the available bacterial genomes could have far-reaching applications, including real-time genomic epidemiology and surveillance. One recently developed method is the Bitsliced Genomic Signature Index (BIGSI) (Bradley *et al.*, 2017). BIGSI is a DNA search engine for microbial data containing 447,833 bacterial and viral raw-read and assembled genomes, from the entire

bacterial and viral WGS content of the ENA (as of December 2016), and can be used to search SNPs, genes or MGEs, at ultra-fast rates.

1.2.1 Staphopia

Staphopia is a *S. aureus*-specific sequence analysis pipeline, database and application programming interface, that was published in 2018 (Petit *et al.*, 2018). The Staphopia library contains over 40,000 *S. aureus* sequences (Figure 2.1) from the ENA database, primarily from Illumina sequencing, that are searchable for a number of different attributes.

The authors identified several major challenges faced during the development of the Staphopia pipeline (Petit *et al.*, 2018). These included difficulties in tracing publications linked to public genome data, and problems collecting metadata from public sequencing projects. The authors found that this can lead to a vast amount of missing data, as only a limited number of metadata fields are required upon submission of sequencing data to public

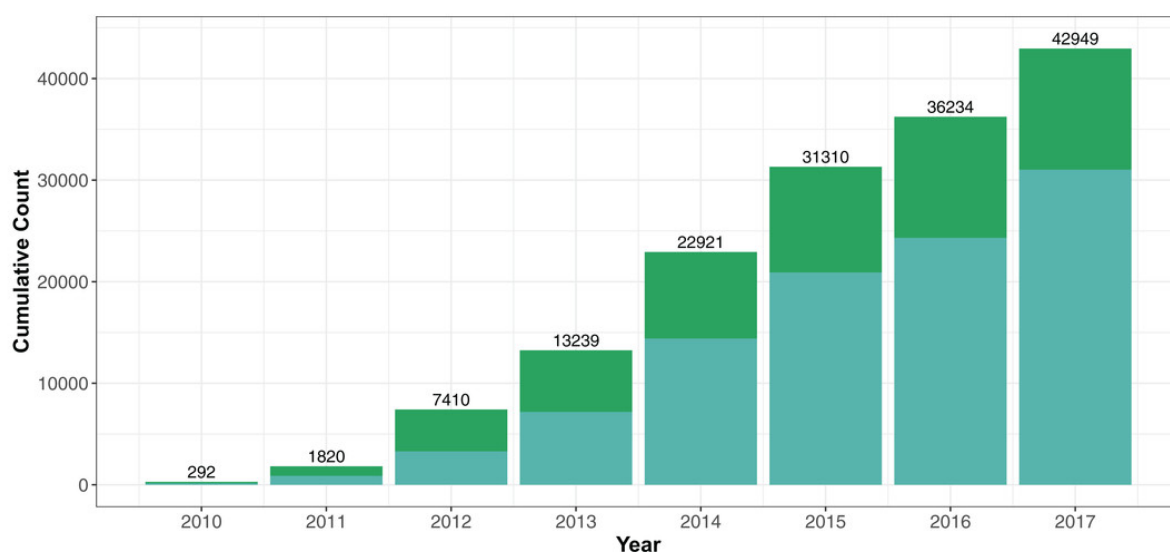


Figure 2.1. Cumulative submissions of *S. aureus* genome projects 2010 – 2017. Green = published, blue = unpublished. Image from Petit *et al.*, 2018.

repositories, and suggested that journals should enforce machine-readable standards for metadata associated with deposited strains. Within the Staphopia database, only 40% of sequences were associated with a collection date and around 35% were associated with a geographic location, and a strong bias was identified towards strains from Europe and North America.

1.2.2 *S. aureus* PubMLST

The MLST scheme is used to type microbial species, based on the DNA sequence of several housekeeping genes that are expressed by all sub-types of a species (Enright *et al.*, 2000). PubMLST (Jolley *et al.*, 2018) contains a collection of open-access, curated databases of over 100 bacterial species and genera, as well as several additional databases, including eukaryotes. The *S. aureus* PubMLST database contains at least one isolate for every known MLST allelic profile. As of August 2019, the *S. aureus* PubMLST database contained 5,204 different MLST profiles.

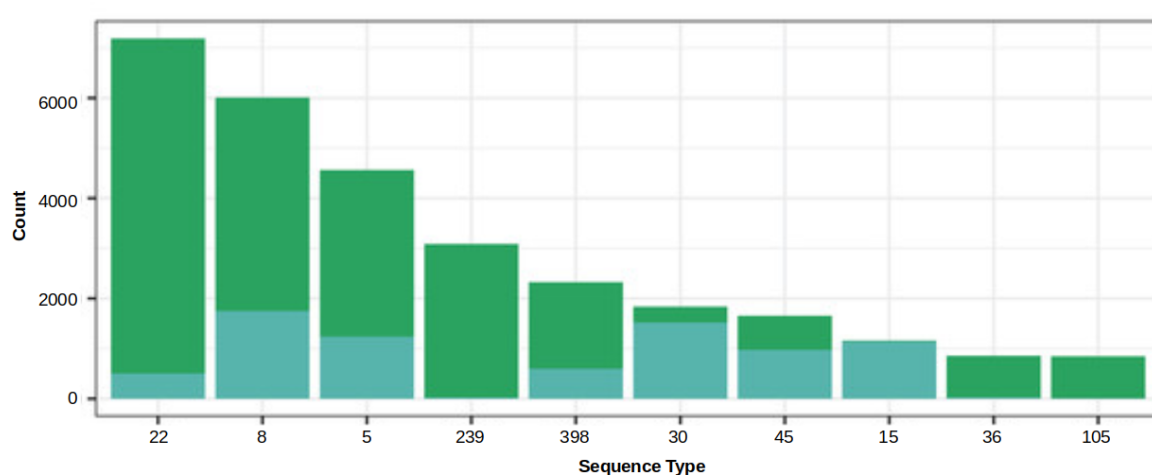


Figure 2.2. The number of records in the Staphopia database for the top ten STs. Green = methicillin resistant. Blue = methicillin susceptible. Image from Petit *et al.*, 2018.

However, many of the PubMLST records are not linked to sequencing data. Of the *S. aureus* genomes that have been sequenced to-date, these are heavily over-represented by a small number of common STs. For example, the Staphopia database contains only 1,090 different MLST profiles from all 40,000 *S. aureus* sequences (Petit *et al.*, 2018). The representation of STs is highly biased towards the top 10 STs, which make up 70% of all genomes in the Staphopia database (Figure 2.2).

1.2.3 Realising the power of bacterial DNA sequence databases

The full power of using bacterial DNA database search engines such as BIGSI and Staphopia for bacterial genome mining is yet to be realised. Outside of the initial Staphopia study, which analysed MLST, SCCmec type and prevalence of AMR genes in *S. aureus* (Petit *et al.*, 2018), the Staphopia database has been utilised for genome mining just once (as of August 2019), to identify genetic variation within the *S. aureus* *isdA* gene (Bennet *et al.*, 2019), which is a heme-iron transporter that plays a key role in the survival of *S. aureus* and other micro-organisms during human infection, and may be a target for the development of novel therapeutics.

The authors of BIGSI demonstrated its use to rapidly find AMR genes, quantify the prevalence of AMR genes, and to determine the host-range of thousands of plasmids (Bradley *et al.*, 2019). Additional studies have used the BIGSI database to investigate the prevalence of a rare genomic island conferring copper resistance in *Salmonella Typhimurium* (Branchu *et al.*, 2019); to identify the prevalence of the plasmid-encoded *qacA4* gene in *S. epidermidis* that confers resistance to the topical antiseptic chlorhexidine gluconate (Addetia *et al.*, 2019);

and to show that the 23S region is well-conserved within bacterial isolates (Beale *et al.*, 2019).

Bacterial DNA sequence databases are only just becoming practical for scientific research. As the number of sequenced bacterial genomes continues to increase, databases including further and higher-quality genomes will enable researchers to make even more powerful and robust conclusions about diverse pathogen data-sets.

1.3 Chapter summary

An extensive review of the prevalence of ST239 around the world was carried out, combining records from the scientific literature and DNA sequencing repositories. This showed that, since the late 1970s, ST239 MRSA has been recorded in multiple countries across Europe, Asia, Oceania, South America, North America and Africa. However, there is evidence that the prevalence of ST239 is now in decline, and there are several examples where the dominant ST239 MRSA has been replaced by a different MRSA strain in clinical settings. The most commonly reported strain that has replaced ST239 was ST22, which has decreased resistance to antimicrobials, higher fitness, and contains the smaller SCC*mec*-IV element.

The diversity of ST239, ST8 and ST30 sequences represented in the Staphopia database was investigated. The majority of the isolates were not associated with metadata, and those that were associated with metadata were frequently from a small number of large-scale studies. A total of 264 geographically and temporally diverse sequences were selected, including 96 ST239, 111 ST8 and 57 ST30 sequences, and these were mapped to their respective high-quality reference sequences for further analysis in Chapters 3 and 4.

Chapter 2

Of all 3,178 ST239 sequences in the Staphopia database, 93.7% were identified as containing SCC*mec*-III. An accurate SCC*mec* type could not be identified in the remaining ST239 sequences. This suggests a strong relationship between the ST239 genome and the SCC*mec*-III element. Within the wider *S. aureus* population, three sequences, corresponding to ST22, ST5 and ST101, that were not closely related to ST239, were found to contain SCC*mec*-III, which suggests that there is some limited mobility of the large SCC*mec*-III element. Although a single ST8 sequence containing SCC*mec*-III was identified, none of the ST30 isolates contained SCC*mec*-III. This may suggest that the SCC*mec*-III was not present within the ST30-like donor, and may have entered into ST239 separately to the hybridisation event.

Additionally, a total of fifteen ST239, ST8 and CC30 isolates were obtained, whole-genome sequenced, and their MLST and SCC*mec* type was verified. All of the ST239 isolates contained SCC*mec*-III, however a large amount of variability between the SCC*mec*-III elements was observed.

2. Methods

2.1 ST239 literature review

An extensive literature review of the prevalence of ST239 around the world was carried out. Countries where ST239 was documented as the dominant MRSA strain in at least one hospital, and where ST239 was described as being out-competed by another strain in at least one hospital, were recorded, as well as articles that reported the presence of ST239 in a country. The ST of any MRSA strain that had replaced ST239 as the dominant strain was also recorded.

2.2 ST239, ST8 and ST30 sequence identification

Sequences with an MLST profile corresponding to ST239, ST30 and ST8 (Table 2.1) were identified within the Staphopia database (Jolley *et al.*, 2018; Petit *et al.*, 2018). The isolation date and geographic location were extracted from the metadata files. Additional ST239, ST30 and ST8 sequences with corresponding isolation date and location metadata were identified from the NCBI database using MegaBLAST (Morgulis *et al.*, 2008). Further ST239, ST30 and ST8 sequences with corresponding isolation date and location metadata were also identified through a literature search.

	Housekeeping gene						
ST	<i>arcC</i>	<i>aroE</i>	<i>glpF</i>	<i>gmk</i>	<i>pta</i>	<i>tpi</i>	<i>yqiL</i>
239	2	3	1	1	4	4	3
8	3	3	1	1	4	4	3
30	2	2	2	2	6	3	2

Table 2.1. MLST alleles for ST239, ST8 and ST30.

2.3 Geographic distributions

All maps of ST239, ST8 and ST30 presence were created using MapCHART (MapChart, 2019). Location data was displayed by country only, and was collected from the metadata in Section 2.2.

2.4 Mapping and phylogeny

A total of 96 ST239, 111 ST8 and 57 ST30 sequences were downloaded from EMBL-EBI (Madeira *et al.*, 2019) (Appendix B.1.1 – B.1.3). Where the sequences were downloaded as assemblies or complete genomes, raw sequence reads were simulated with dwgsim v0.1.11 (Homer, 2011). Illumina reads were trimmed using Trimmomatic (Bolger *et al.*, 2014), and bwa v0.7.15 and SAMtools v1.3.1 (Li and Durbin, 2009) were used to map the ST239, ST8 and ST30 sequences to the reference sequences with accession numbers FN433596, CP007690 and LN626917, respectively. Consensus sequences were generated for each isolate, and combined to make a multifasta alignment. Genes were annotated with Prokka v1.13 (Seeman, 2014). The pan-genome was calculated and the core genes extracted using Roary v3.12.0 (Page *et al.*, 2015), and maximum-likelihood phylogenies were estimated with RaxML v8.2.9 (Stamatakis, 2014), using a GTR model with gamma correction for among site rate variation, which was replicated for 100 bootstraps. All trees were viewed in Figtree v1.4.4 (Rambaut, 2010).

2.5 SCCmec type

The SCCmec type of all ST239 sequences in the Staphopia database was predicted using the Staphopia API (Petit *et al.*, 2018). For ST239 sequences where no SCCmec type was accurately identified, the Staphopia SCCmec primer record was analysed for the presence of

any SCCmec-related genes, such as *mecA* or *ccrA*. Where no SCCmec-related genes were present, the SCCmec type was recorded as “no evidence of SCCmec”.

The Staphopia database was mined for all sequences predicted to contain SCCmec-III. The ST of these sequences was recorded, and the MLST of these sequences was identified using PubMLST (Jolley *et al.*, 2018). The MLST was confirmed using ARIBA v2.13.3 (Hunt *et al.*, 2017) and the SCCmec type was confirmed using SCCmecFinder (Kaya *et al.*, 2018).

2.6 Experimental isolates

2.6.1 DNA extraction

The *S. aureus* isolates in Table 2.2 were obtained from BEI Resources, Manassas Virginia, USA; the Leibniz Institute DSMZ, Braunschweig, Germany; Public Health England, London, UK; or the John Radcliffe Hospital, Oxford, UK. Stocks of all isolates were mixed with 50% glycerol and stored at -80°C.

For DNA extraction, the bacteria from cryostock were streaked on TSA, and incubated at 37°C for 24 hours. Single colonies were picked and incubated for 24 hours in 3 mL TSB, at 37°C with 225 RPM. Roughly 1 mL of each culture was digested in enzymatic lysis buffer containing 25 – 30 mg/mL lysozyme, 0.1 mg/mL lysostaphin, 20 mM TrisCl, 2 mM EDTA and 1.2% Triton-X100 for 2 hours. DNA was purified using the Qiagen DNeasy Blood and Tissue kit (QIAGEN, 2006), and following the protocol for purification of bacterial or yeast DNA with enzymatic lysis, using QIAcube. The concentration of genomic DNA in each sample was then measured using a Promega QuantiFluor dsDNA System, and samples were diluted to 10 – 50 ng/μL in DNase-free water before sequencing was carried out using an

ST	Name	BEI Accession	Isolation Year	Isolation Location	Predicated MRSA/ MSSA	Source
ST239A	BR15	NR-45889	1998	Brazil	MRSA	BEI
ST239B	Br5 / NRS56	NR-45890	1999	Brazil	MRSA	BEI
ST239C	TW20	-	2003	UK	MRSA	PHE
ST239D	SR2852	NR-50508	2011	USA	MRSA	BEI
ST8A	NCTC 8325	-	1950	UK	MSSA	DSMZ
ST8B	RN1	NR-45904	1960	UK	-	BEI
ST8C	A970230 / NRS172	NR-45971	1997	France	MRSA	BEI
ST8D	GA-355 / NRS685	NR-46214	2006	USA	MRSA	BEI
ST8E	880 (BR-VRSA) / SAMN02732485 / JICL01	NR-49120	2012	Brazil	MRSA	BEI
ST8F	JRH	-	2016	UK	-	JRH
ST30A	HIP12899 / NRS484	NR-46081	1996	USA	MRSA	BEI
ST30B	HT20020058 / SRR2910957	NR-46024	2002	France	MSSA	BEI
ST30C	160013 / NRS283	NR-46067	2002	UK	MRSA	BEI
ST30D	TN-82 / NRS740	NR-46269	2005	USA	MRSA	BEI
ST30E	CA-409 / NRS651	NR-46180	2006	USA	MRSA	BEI

Table 2.2. Metadata of the experimental isolates used in this study.

Illumina HiSeq4000 system with 150 bp paired-end reads by the Oxford Genomic Centre, Wellcome Trust Centre for Human Genetics, University of Oxford, UK.

2.6.2 Mapping and assembly

The ST239, ST8 and CC30 sequences were mapped to FN433596, CP007690 and LN626917, respectively, as in Section 2.4. Additionally, these sequence reads were *de novo* assembled into contigs using SPAdes v3.13.0 (Bankevich *et al.*, 2012) and the contigs were ordered against the FN433596, CP007690 or LN626917 reference sequences using abacas v1.3.1 (Assefa *et al.*, 2009) and annotated with Prokka v1.3 (Seeman, 2014). Maximum-likelihood

core genome phylogenies were estimated as in Section 2.4. MLST was determined using PubMLST (Jolley *et al.*, 2018).

The raw reads from the ST239C contaminant were uploaded to the PathogenWatch web application (Centre for Genomic Pathogen Surveillance, 2018), and the predicted species and ST of the contaminant was recorded.

2.6.3 SCCmec

SCCmec type was determined by PCR using the primers in Table 2.3, which were obtained from Thermo Fisher Scientific, Loughborough, UK. These were then identified in the assemblies of the isolates from Section 2.6.2. PCR amplification was carried out in a 50 µL reaction volume, consisting of 20 µL combined 100 µM primers, 5 µL template DNA and 25 µL GoTaq® DNA polymerase. The thermal cycle was programmed for 4 min at 94°C for initial denaturation, followed by 30 cycles of 30 s at 94°C for denaturation, 30 s at 53°C for annealing, 60 s at 72°C for extension, and 4 min at 72°C for the final extension. PCR products were examined by gel electrophoresis at 75 V for 45 – 85 min in a 1.5% (w/v) agarose gel in 1x TAE buffer with 5 µL SYBR™ Safe DNA gel stain. The marker used was a 100 bp ladder.

The boundaries of the SCCmec-III element were identified in the TW20 reference genome using BLAST against the reference SCCmec-III element with the accession number AB037671 (Ito *et al.*, 2001), and confirmed using SCCmecFinder (Kaya *et al.*, 2018). The SCCmec-III element was then extracted from the ST239A – D sequences, and visualised against the TW20 SCCmec-III element using Artemis Comparison Tool v13.0.0 (Carver *et al.*, 2005). Gene function for genes that could be annotated by Prokka in Section 2.5.2 were

SCC <i>mec</i> type	DNA sequence	Reverse complement of DNA sequence
I	1. TTCGAGTTGCTGATGAAGAAGG	2. CCTTCTTCATCAGCAACTCGAA
	3. ATTTACCACAAGGACTACCAGC	4. GCTGGTAGTCCTTGTGGTAAAT
II	1. AATCATCTGCCATTGGTGATGC	2. GCATCACCAATGGCAGATGATT
	3. CGAATGAAGTGAAAGAAAGTGG	4. CCACTTTCTTTCACTTCATTCTG
II,III	1. ATCAAGACTTGCATTCAGGC	2. GCCTGAATGCAAGTCTTGAT
	3. GCGGTTTCAATTCACCTTGTC	4. GACAAGTGAATTGAAACCGC
I,II,IV	1. CATCCTATGATAGCTTGGTC	2. GACCAAGCTATCATAGGATG
	3. CTAAATCATAGCCATGACCG	4. CGGTCATGGCTATGATTTAG
III	1. GTGATTGTTTCGAGATATGTGG	2. CCACATATCTCGAACAATCAC
	3. CGCTTTATCTGTATCTATCGC	4. GCGATAGATACAGATAAAGCG
III	1. TTCTTAAGTACACGCTGAATCG	2. CGATTCAGCGTGTACTTAAGAA
	3. GTCACAGTAATTCCATCAATGC	4. GCATTGATGGAATTACTGTGAC
<i>mecA</i>	1. TCCAGATTACAACCTCACCAGG	2. CCTGGTGAAGTTGTAATCTGGA
	3. CCACTTCATATCTTGTAACG	4. CGTTACAAGATATGAAGTGG

Table 2.3. PCR primers for determining SCC*mec* type (adapted from Oliveira and de Lencastre, 2002).

identified in UniProt (The UniProt Consortium, 2018), and MegaBLAST (Morgulis *et al.*, 2008) was used to identify genes where no annotation was found with Prokka.

3. Results

3.1. Evidence of the rise and fall of ST239

To investigate how widespread ST239 has become around the globe, literature reports of ST239 prevalence, as well as ST239 isolates from online sequencing repositories, were identified (Table 2.4a – 2.4b). Records of ST239 prevalence ranged from single cases of ST239 infections, to large-scale surveillance of epidemics that have resulted in ST239 becoming the dominant MRSA lineage in multiple different hospitals across several countries. Multiple incidences of ST239 MRSA were recorded in Europe, Asia, Oceania, South America, North America and Africa, with the earliest isolates being identified in the late 1970s. Since routine, global, detailed surveillance data of MRSA is not available, it is likely that the prevalence of ST239 has been severely under-reported, and that a vast amount of data has not been recorded. Therefore, gaps in the data do not necessarily mean that ST239 was absent. However, there is evidence that ST239 has been the dominant MRSA in a wide range of different countries over many years, which shows that ST239 MRSA is extremely widespread, and has had a major impact in healthcare settings across the globe.

Several examples where the dominant ST239 MRSA has been replaced by a different MRSA lineage were also identified. These replacement events may be evidence of ST239 being out-competed by strains that are more fit in the clinical environment. The most commonly reported MRSA strain that has replaced ST239 was ST22, which was recorded in India, Australia, Portugal, Kuwait, and on multiple occasions in Singapore. ST22 typically contains SCC*mec*-IV, which is a small SCC*mec* element that has been shown to have no fitness cost and carry few AMR genes (Lee *et al.*, 2007). Several of the studies showed that clinical ST22-

Year of isolation	North America, South America, Oceania and Asia																											
	Canada	USA	Argentina	Brazil	Chile	Ecuador	Paraguay	Peru	Trinidad and Tobago	Uruguay	Australia	New Zealand	China	India	Indonesia	Iran	Kuwait	Laos	Malaysia	Pakistan	Philippines	Saudi Arabia	Singapore	South Korea	Sri Lanka	Taiwan	Thailand	Vietnam
1979																												
1980																												
1981																												
1982																												
1983																												
1984																												
1985																												
1986																												
1987																												
1988																												
1989																												
1990																												
1991																												
1992																												
1993																												
1994																												
1995																												
1996																												
1997																												
1998																												
1999																												
2000																												
2001																												
2002																												
2003																												
2004																												
2005																												
2006																												
2007																												
2008																												
2009																												
2010																												
2011																												
2012																												
2013																												
2014																												
2015																												
2016																												
2017																												

Year of isolation	Europe and Africa																																
	Croatia	Czechia	Denmark	France	Finland	Germany	Greece	Hungary	Ireland	Italy	Lithuania	Malta	Netherlands	Norway	Poland	Portugal	Romania	Russia	Spain	Switzerland	Turkey	UK	Algeria	Egypt	Gambia	Ghana	Kenya	Morocco	Niger	Nigeria	Senegal	South Africa	Tunisia
1979																																	
1980																																	
1981																																	
1982																																	
1983																																	
1984																																	
1985																																	
1986																																	
1987																																	
1988																																	
1989																																	
1990																																	
1991																																	
1992																																	
1993																																	
1994																																	
1995																																	
1996																																	
1997																																	
1998																																	
1999																																	
2000																																	
2001																																	
2002																																	
2003																																	
2004																																	
2005																																	
2006																																	
2007																																	
2008																																	
2009																																	
2010																																	
2011																																	
2012																																	
2013																																	
2014																																	
2015																																	
2016																																	
2017																																	

Table 2.4a – 2.4b. Reports of ST239 prevalence around the world. Key: yellow, ST239 is a dominant MRSA strain in at least one hospital; green, at least one isolate of ST239 was recorded in this country; red, ST239 has been or is being out-competed by another strain in at least one hospital. The dominant strain that has replaced ST239-III is recorded as ^W ST36-II; ^F ST228-I/ST5-II; ^O ST22-IV; ^J ST22-IV/ST772-V; ^Z ST5-II; ^L ST59-IV; ⁸ ST80-IV; ^M ST772-V. Reference list in Appendix B.2.1.

IV isolates are less resistant to antimicrobials than ST239-III (Hsu *et al.*, 2005b; Hsu *et al.*, 2007a; Amorim *et al.*, 2007; D'Souza *et al.*, 2010), but clinical ST22-IV isolates have been shown to have higher fitness than ST239-III in *in vitro* growth rate comparison experiments (Knight *et al.*, 2012). This suggests that the advantages of high multi-drug resistance that ST239 has benefited from in the past are no longer able to compete with the fitness advantages of ST22-IV in these clinical environments.

In Australia, ST239 has persisted as one of the major healthcare-associated MRSA strains for decades. Although there was a comprehensive report suggesting that ST22-IV is beginning to overtake ST239 as the dominant MRSA strain in several hospitals across Australia (Coombs *et al.*, 2013), other evidence supports that ST239 was still the dominant MRSA strain in at least one hospital in Australia in 2013 (Abbott *et al.*, 2015). In China, studies have reported that ST239 was a dominant MRSA strain for over a decade (Aires de Sousa *et al.*, 2003b; Xu *et al.*, 2009). However, this has been replaced in multiple hospitals across China over several years by ST59-IV (Li *et al.*, 2018) and ST5-II (Dal *et al.*, 2018).

Overall this data suggests that, whilst ST239 has been a major healthcare-associated MRSA strain around the globe, its prevalence is now in decline, as several other dominant MRSA strains are beginning to emerge.

3.2 ST239, ST8 and ST30 sequences

3.2.1 Staphopia

ST239, ST8 and ST30 sequences were all represented in the top 10 STs within the Staphopia database (Table 2.5). Some of these sequences were associated with metadata, such as an

ST	Total number of sequences	Total number of sequences with date of isolation	Total number of sequences with location of isolation	Total number of sequences with both date and location of isolation
ST239	3,178	386 (12.1%)	313 (9.8%)	313 (9.8%)
ST8	6,280	3,886 (61.9%)	3,706 (59.0%)	3,397 (54.1%)
ST30	1,896	455 (24.0%)	467 (24.6%)	442 (23.3%)

Table 2.5. Number of ST239, ST8 and ST30 sequences associated with metadata in the Staphopia database.

isolation date and location, however the majority were not associated with any metadata. Of those that were associated with a location and date of isolation, many were closely related sequences from the same study, such as isolates that were collected over a set time-period from one particular hospital. For example, 276 of the 329 ST30 sequences associated with geographic location metadata from the UK were collected from health-care workers, environmental locations, and patients in an ICU and a high-dependency unit at the Royal Sussex County Hospital, Brighton, England, between 31st October 2011 and 23rd December 2012 (Price *et al.*, 2017).

3.2.2 Geographic distribution of ST239, ST8 and ST30 sequences

ST239 and ST8 belong to CC8, and ST30 belongs to CC30. CC8 and CC30 are two of the five major EMRSA CCs, which also include CC5, CC45 and CC22. These EMRSAs are well adapted to the hospital environment, are established in several hospitals within a country, or have spread internationally (Enright *et al.*, 2002), and their high frequency within healthcare settings is reflected by their prevalence within the Staphopia database. Within the Staphopia and NCBI databases, ST239 sequences were identified that had been isolated from 34 countries across all six inhabited continents. ST8 and ST30 sequences were identified from 25 countries and 22 countries, respectively (Figures 2.3a – 2.3c) (Appendix B.1.4).

A

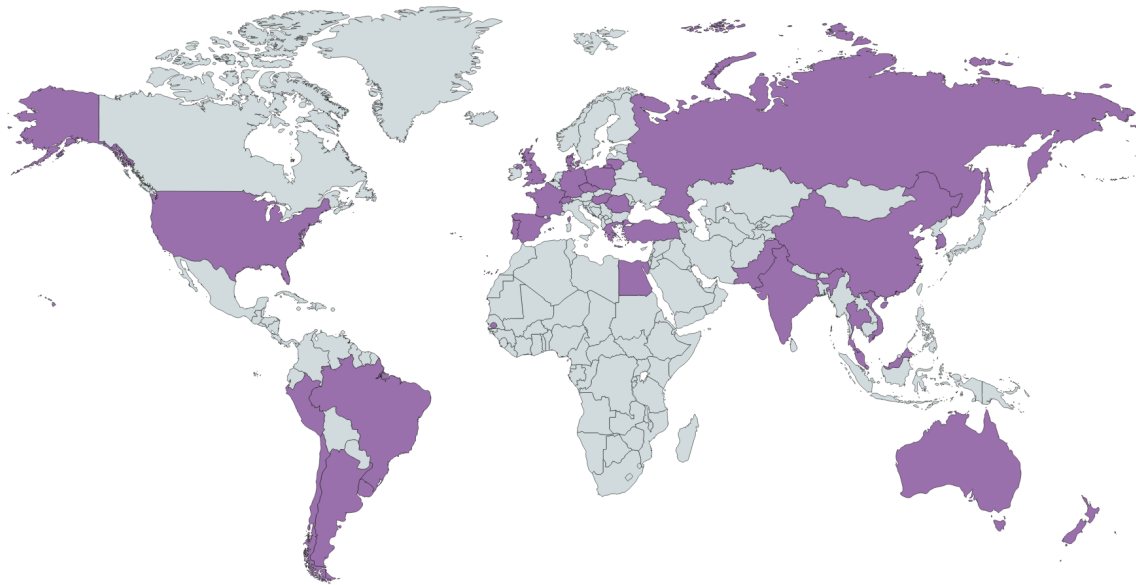


Figure 2.3a. 34 countries where ST239 has been isolated and recorded in online sequence databases. Purple, an ST239 sequence was present from this country.

B

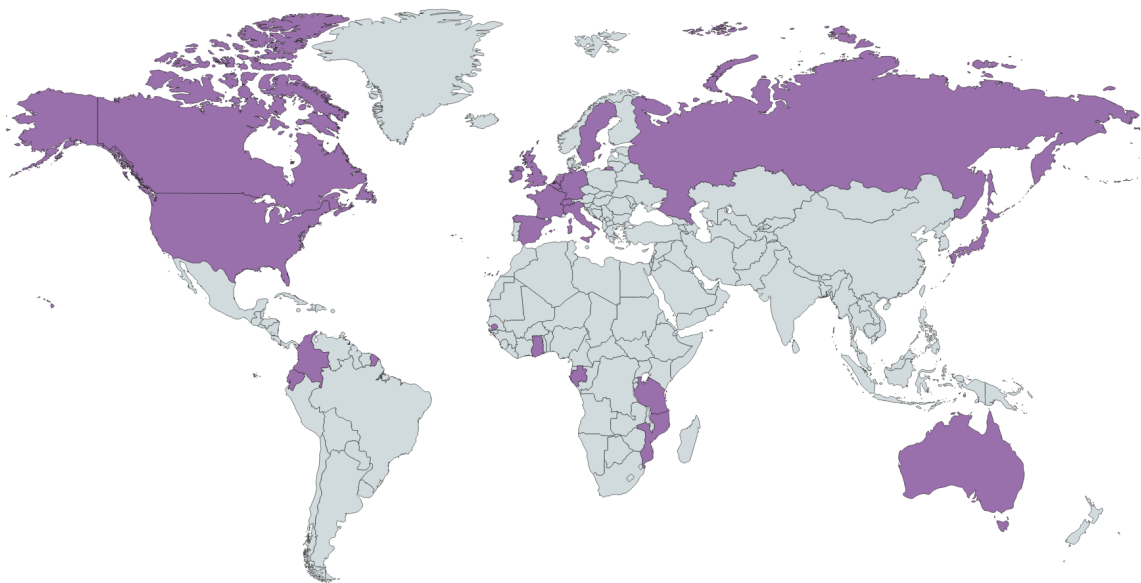


Figure 2.3b. 25 countries where ST8 has been isolated and recorded in online sequence databases. Purple, an ST8 sequence was present from this country.

C

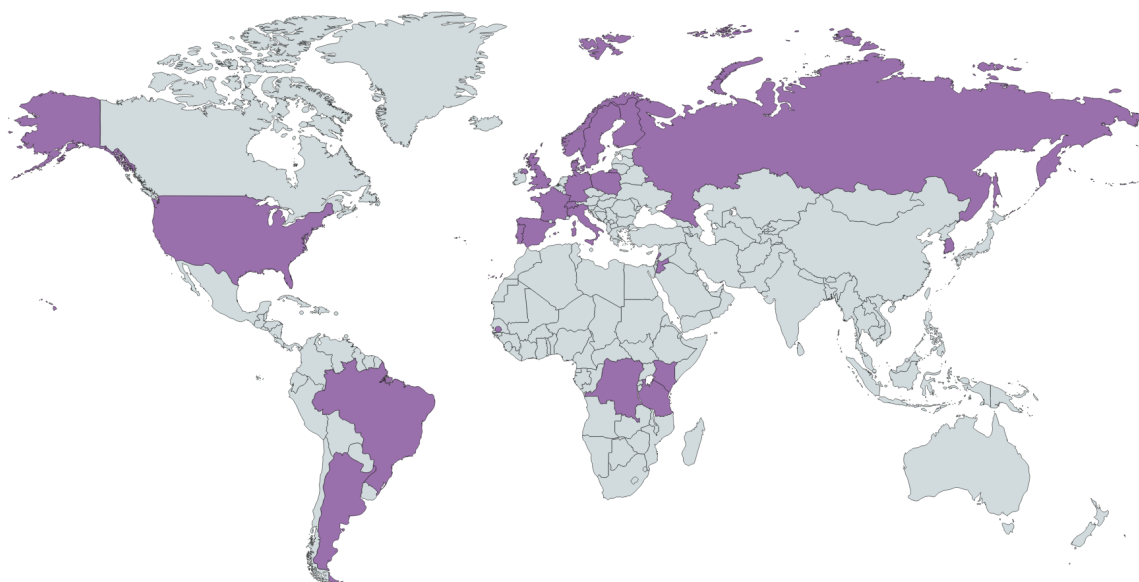


Figure 2.3c. 22 countries where ST30 has been isolated and recorded in online sequence databases. Purple, an ST30 sequence was present from this country.

All three STs were represented at least once on each inhabited continent (with the exception of ST30, which was not identified in Oceania), which reflects their ability to spread internationally. The ST239 isolates had a much higher representation in Asian countries compared to the ST8 and ST30 isolates, for which only a single ST8 isolate from Japan was identified, and a single ST30 isolate from South Korea. The high prevalence of ST239 in Asia is consistent with the literature, with studies suggesting that, in the mid-2000s, ST239 was the predominant MRSA strain in Asia, causing up to 90% of hospital-acquired MRSA within a region accounting for over 60% of the world's population (Feil *et al.*, 2008). The high number of ST239 isolates in the Staphopia database that were associated with geographic location data from Asia ($n = 172$), is primarily due to a single study that included 106 *S. aureus* isolates from Singapore between 2000 – 2010 (Wellcome Sanger Institute, 2011). The 77 isolates from Australia that were associated with geographic location metadata, which contribute to a high prevalence of ST239 in Oceania ($n = 78$) compared to ST8 ($n = 14$) and ST30 ($n = 0$), were also from a single study, spanning 32 years (Baines *et al.*, 2015). There

has been a very low representation of ST239 in South America in these databases, even though ST239 was first described in Brazil, and is a known Brazilian epidemic strain (Teixeira *et al.*, 1995), which may reflect a lack of sequencing of isolates from this region.

Additionally, ST8 is a highly prevalent strain in both hospital- and community-acquired MRSA infections in the USA (Strauss *et al.*, 2017), and the majority of ST8 sequences were from isolates from the USA. The 3,284 ST8 isolates that were sequenced in the USA were collected from many different states, years and from 31 individual studies. Thirteen different studies sequenced ST8 isolates from the UK, however, 52% of these sequences were from the same study at the Royal Sussex County Hospital, Brighton, England, between 31st October 2011 and 23rd December 2012 (Price *et al.*, 2017).

CC30 has been prevalent in healthcare settings in the UK (Enright *et al.*, 2002), and is very well represented in the UK in the Staphopia database. The prevalence of ST30 on all other continents within the Staphopia database is at a low level.

3.2.3 Forming diverse data-sets

A total of 96 high-quality temporally and geographically diverse ST239 sequences were selected from a range of different studies, and mapped to the ST239 complete reference sequence with NCBI accession number FN433596, with an average depth of 120.0x, ranging from 40.3x to 503.1x (Appendix B.1.1). The average coverage was 95.5%, ranging from 90.7% – 100%, and the average SNP divergence from the reference was 4,159, ranging from 0 – 11,802. A total of 111 diverse ST8 sequences were mapped to the ST8 complete reference sequence with NCBI accession number CP007690, with an average depth of 117.3x, ranging

from 17.9x to 686.2x (Appendix B.1.2). The average coverage was 97.8%, ranging from 91.4 – 100%, and the average SNP divergence from the reference was 812, ranging from 0 – 2,360. Additionally, 57 diverse ST30 sequences were mapped to the ST30 complete reference sequence with NCBI accession number LN626917, with an average coverage of 113.9x, ranging from 7.7x to 289.4x (Appendix B.1.3). The average coverage was 96.1%, ranging from 92.1% – 100%, and the average SNP divergence from the reference was 2,983, ranging from 0 – 4,382. The lower range of SNP divergence within the ST8 sequences suggests a lower diversity of isolates compared to the ST239 and ST30 data-sets, even though the ST8 data-set contains the highest number of isolates.

The pan-genome of the 96 ST239 isolates consisted of a total of 3,337 genes. Of these, 1,889 were present in 99 – 100% of all strains, and are defined here as “core genes”, resulting in a core genome length of 1,754,805 bp. These core genes ranged from 104 bp – 7,436 bp in length, with an average of 928 bp per gene, including 683 core genes that were >1000 bp in length. This is a narrower core genome than that observed by Harris (2010), who analysed 43 ST239 sequences from a global isolate collection alongside 20 ST239 sequences from a single hospital in northeast Thailand that were isolated between 2006 – 2007 during an outbreak in an ICU. The core genome identified by Harris consisted of 1,492 core genes, but only included genes that were >1,000 bp in length. Therefore, the ST239 sequences collected in this study represent a greater diversity of ST239 isolates than those included in the study by Harris (Harris *et al.*, 2010).

Gene Type	Percentage of strains	No. genes in ST239 data-set	No. genes in ST8 data-set	No. genes in ST30 data-set	No. genes in combined ST239, ST8 and ST30 data-set
Core genes	99% – 100%	1,889	2,253	2,217	1,980
Soft core genes	95% – 98.9%	521	135	147	86
Shell genes	15% – 94.9%	542	328	297	1,213
Cloud genes	0% – 14.9%	385	221	393	1,059
Total genes	0% – 100%	3,337	2,937	3,054	4,338

Table 2.6. Breakdown of the number of genes in the pan-genome of the ST239, ST8 and ST30 sequences.

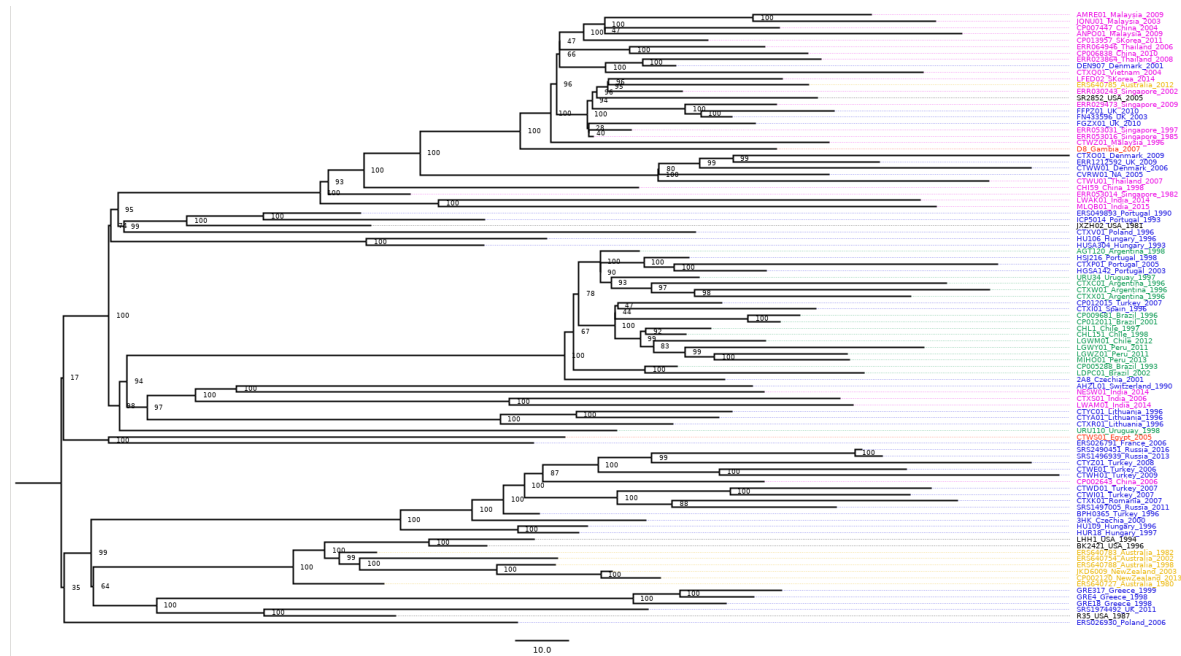
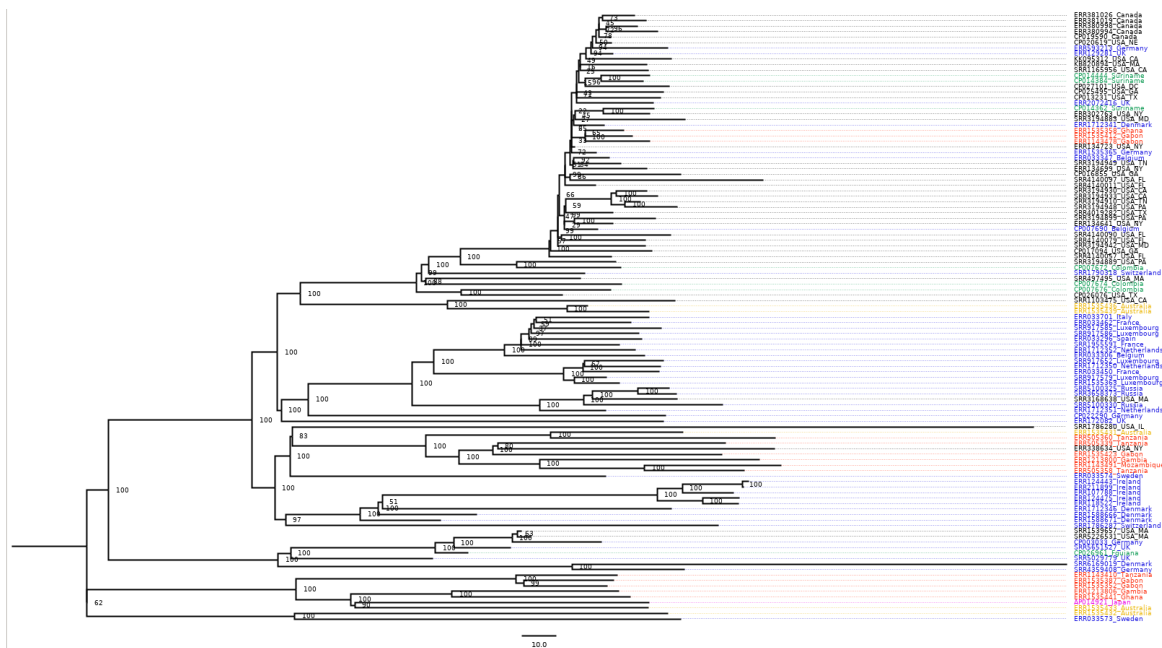


Figure 2.4. Out-group rooted maximum-likelihood phylogeny of the 96 ST239 sequences. Units in SNPs/Mb. Blue, Europe; Pink, Asia; Orange, Oceania; Green, South America; Black, North America; Red, Africa.

A



B

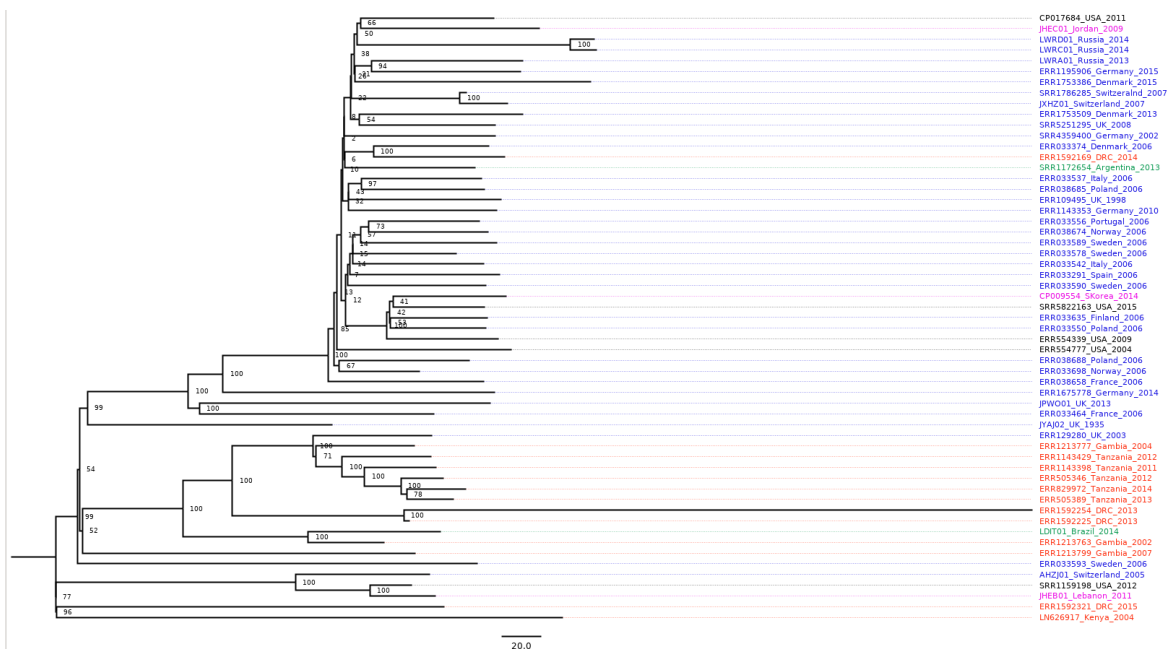


Figure 2.5. Out-group rooted maximum-likelihood phylogeny of a) the 111 ST8 sequences, b) the 57 ST30 sequences. Units in SNPs/Mb. Blue, Europe; Pink, Asia; Orange, Oceania; Green, South America; Black, North America; Red, Africa.

Chapter 2

The pan-genome of the ST8 isolates was made up of 2,937 genes, of which 2,253 were core, and the pan-genome of the ST30 isolates was made up of 3,054 genes, of which 2,217 were core (Table 2.6). Overall, 4,338 unique genes were identified across all 264 ST239, ST8 and ST30 sequences. The number of core genes, shared by >99% of isolates from all three STs was 1,980.

The maximum-likelihood phylogeny for the ST239 data-set had strong bootstrap support (Figure 2.4), and the long branch lengths were suggestive of a diverse set of isolates. The population structure is similar to that observed by Harris (2010), and by Castillo-Ramirez (2012), who identified strong geographical clustering of ST239 sequences on continental, national and city scales (Castillo-Ramirez, *et al.*, 2012). For example, Harris observed a tightly clustered South American clade, which may reflect a recent expansion of a single variant across the continent. The South American ST239 isolates in this study were consistent with this observation, with a single exception. Additionally, Harris noted clustering of isolates from Thailand and China into a single diverse Asian clade. In this study, isolates from a greater number of Asian countries were also included, and the majority of the Asian isolates were clustered, however there were several exceptions, including three isolates from India and one from China. Additionally, in this study a number of isolates from Oceania and North America were also included. The isolates from Oceania formed a tightly clustered clade, with one exception. These were more closely related to European and North American isolates than the Asian isolates, which is in contrast with the North American isolates, which are dispersed throughout the tree.

Harris observed evidence of geographical clustering within the European isolates, however, although Castillo-Ramirez observed clustering of the South American and Asian isolates, they found that the European isolates were dispersed throughout the phylogeny. Castillo-Ramirez found evidence of multiple independent introductions from Europe that were apparent in both the core and accessory ST239 genome. The European isolates in this study do show some evidence of geographical clustering, however, these are more diverse and spread throughout the tree, which is more similar to Castillo-Ramirez (Castillo-Ramirez, *et al.*, 2012).

The maximum-likelihood phylogeny for the ST8 and ST30 data-sets also had strong bootstrap support overall (Figure 2.5a and 2.5b), although the support for some of the ST30 sequences was quite low.

3.3 SCCmec type

3.3.1 SCCmec prevalence in ST239

ST239 is commonly resistant to methicillin due to the presence of the SCCmec-III element, which has been identified in ST239 isolates from around the globe (Monecke *et al.*, 2018). The Staphopia database can be mined for the presence of different SCCmec types in *S. aureus* sequences. The number of ST239 sequences containing each SCCmec type is recorded in

	No. ST239 sequences in Staphopia	Percentage of total ST239s
Evidence of SCCmec-III	2,979	93.7%
Some evidence of an SCCmec	108	6.3%
No evidence of SCCmec	91	
Total	3,178	100%

Table 2.7. SCCmec prevalence in ST239 sequences in Staphopia.

Table 2.7. Of the 3,178 ST239 sequences in the Staphopia database, 2,979 (93.7%) were identified as containing the SCC*mec*-III element. In the remaining 199 ST239 sequences, an SCC*mec* type could not be accurately determined. Of these, 108 sequences showed some evidence of the presence of an SCC*mec* element, for example, the presence of *ccrA* or *mecA* genes, which are commonly found in SCC*mec* elements (IWG-SCC, 2009), therefore it is unlikely that the SCC*mec* element is absent. It is possible that these isolates contain an unpublished SCC*mec* element, however, it is also possible that the SCC*mec* region of these sequences was poorly sequenced, and hence the SCC*mec* type could not be identified.

The remaining 91 ST239 sequences contained no genetic evidence of the presence of an SCC*mec* element. It is possible that the SCC*mec* element in these sequences was also poorly sequenced, and hence no SCC*mec* genes could be identified. It is also possible that these sequences do not contain any SCC*mec* element, having lost the SCC*mec* element over time.

One example of a single ST239 isolate that may not contain an SCC*mec* element was identified by Luedicke (2010). Although several SCC*mec* genes were not identified in this ST239 isolate using a DNA array, including *mecA* and *ccr* genes, the SCC*mec*-III linked mercury resistance genes *merA*, *merB*, *merR*, and *merT* were detected. Luedicke was also unable to detect the *spa* gene region in this isolate for *spa*-typing, and unfortunately, DNA sequencing of this isolate or phenotypic AMR susceptibility-testing was not carried out to confirm whether this ST239 isolate is a true MSSA isolate (Luedicke *et al.*, 2010). Monecke (2018) suggested that this isolate may be an example of an ST239 that has lost the SCC*mec*-III element (Monecke *et al.*, 2018).

Additionally, there are a few examples in the literature of ST239 isolates that may contain an alternative *SCCmec* type, including *SCCmec*-IV (Ramdani-Bouguessa *et al.*, 2006) and *SCCmec*-V (Breurec *et al.*, 2011). Abdulgader (2015) suggested that these may be examples of replacement of *SCCmec*-III in the ST239 genome (Abdulgader *et al.*, 2015). However, both of these isolates were typed using PFGE or PCR methods, and therefore no whole-genome sequencing data from these isolates can be accessed to determine whether these were incorrectly typed, and these ST239-IV and ST239-V isolates are not represented in the Staphopia database. Therefore, if these alternative *SCCmec* types are able to replace *SCCmec*-III in ST239, it is probably very infrequent.

3.3.2 *SCCmec*-III prevalence in *S. aureus*

A total of 3,152 sequences containing *SCCmec*-III were identified in the Staphopia database (Table 2.8). Of these, 2,979 (94%) were identified as ST239. Although the *SCCmec*-III element is associated with the hypothesised ST30-donated region of the ST239 genome, there was no evidence of *SCCmec*-III in any of the 1,896 ST30 sequences. However, a single ST8 isolate was identified as containing *SCCmec*-III; Staphopia ID 1196 (Tgen North, 2017) is a human isolate that was identified as the only *SCCmec*-III ST8 isolate in a large scale study of 497 CC8 genomes (Bowers *et al.*, 2018).

In total, 173 *S. aureus* sequences contained *SCCmec*-III that were not identified as ST239. These included 30 unique STs (Appendix B.3.1), and 47 sequences where the ST could not be identified, for example due to an undefined ST, poor sequencing quality of the housekeeping genes required for MLST identification, or contamination with several different MLSTs.

	No. SCC<i>mec</i>-III sequences in Staphopia	Percentage of total SCC<i>mec</i>-III sequences
Total	3,152	100%
ST239	2,979	94.5%
Shares 5 – 6 MLST genes with ST239	123	3.9%
Shares 3 – 4 MLST genes with ST239	0	0%
Shares 1 – 2 MLST genes with ST239	3	0.10%
Un-typable	47	1.5%

Table 2.8. SCC*mec*-III prevalence in Staphopia.

The majority of the non-ST239 sequences that were associated with an MLST were closely related to ST239 (Table 2.8), sharing five or more MLST alleles with ST239, suggesting that they are descended from a recent shared common ancestor with ST239. However, three of the non-ST239 STs that contain SCC*mec*-III shared only one or two MLST genes with ST239, and are therefore much less closely related to ST239. These included two sequences that were part of a pre-publication release by the Wellcome Trust Sanger; an ST101 isolate (Sanger, 2012) and an ST21 that was isolated from a human blood sample in England in 2011 (Sanger, 2011). Additionally, an ST5 isolate containing SCC*mec*-III was identified, which was isolated in 1995 from a human source. The ST5 isolate was previously recorded as the only SCC*mec*-III containing ST5 in a study into the evolution of a selection of an international collection of MRSA isolates (Robinson and Enright, 2003). The presence of SCC*mec*-III in three unrelated non-ST239-like isolates suggests that the SCC*mec*-III element can be mobile and has moved between *S. aureus* isolates. However although SCC*mec*-III has moved into epidemic isolates ST21 and ST5 on rare occasions, it has not become prevalent.

None of the sequences that contained SCC*mec*-III shared more than one MLST housekeeping gene with ST30. This suggests that the ST30-like common ancestor of the acquired region of ST239 may not have contained SCC*mec*-III.

3.4 Experimental isolates

Fifteen *S. aureus* isolates were obtained (Table 2.9). A range of dates and geographic locations were selected, including some MRSA and some MSSA, in order to try to find the most varied strains possible. In some cases, the SCC*mec* type had been previously characterised. Due to the low number of available ST30 isolates, some ST36 isolates were selected, which are very

ID	ST	Year of Isolation	Location of Isolation	Predicted MRSA/ MSSA	Predicted SCC<i>mec</i> Type	Source
ST239A	239	1998	Brazil	MRSA	N/A	BEI
ST239B	239	1999	Brazil	MRSA	N/A	BEI
ST239C	239	2003	UK	MRSA	N/A	PHE
ST239D	239	2011	USA	MRSA	N/A	BEI
ST8A	8	1950	UK	MSSA	None	DSMZ
ST8B	8	1960	UK	N/A	N/A	BEI
ST8C	8	1997	France	MRSA	N/A	BEI
ST8D	8	2006	USA	MRSA	IV	BEI
ST8E	8	2012	Brazil	MRSA	IV	BEI
ST8F	8	2016	UK	N/A	N/A	JRH
ST30A	30	1996	USA	MRSA	N/A	BEI
ST30B	30	2002	France	MSSA	None	BEI
ST30C	36	2002	UK	MRSA	N/A	BEI
ST30D	36	2005	USA	MRSA	II	BEI
ST30E	36	2006	USA	MRSA	II	BEI

Table 2.9. Metadata of the 15 experimental strains.

Chapter 2

closely related to ST30, differing by only one MLST allele and both sharing the same *arcC* type 2 housekeeping allele as ST239.

All DNA were extracted successfully and whole genome sequenced. All of the sequences mapped successfully to their respective references (Appendix B.1.5), with a coverage ranging from 93% – 99% and a depth from 273.1x – 449.3x, with the exception of ST239C and ST239E. The low coverage and depth of the ST239C sequence suggested that the isolate had become contaminated prior to sequencing. PathogenWatch analysis (Centre for Genomic Pathogen Surveillance, 2018) reported that the contaminant was *E. coli* ST1. All future experiments were run with the ST239C original stock, which was shown by PCR to be free of

ID	Predicted MLST	MLST Gene Type							Genetic MLST
		<i>arcC</i>	<i>aroE</i>	<i>glpF</i>	<i>gmk</i>	<i>pta</i>	<i>tpi</i>	<i>yqiL</i>	
239A	239	2	3	1	1	4	4	3	239
239B	239	2	3	1	1	4	4	3	239
239C	239	2	3	1	1	4	4	3	239
239D	239	2	3	1	1	4	4	3	239
30A	30	2	2	2	2	6	3	2	30
30B	30	2	2	2	2	3	2	2	N/A
30C	36	2	2	2	2	3	3	2	36
30D	36	2	2	2	2	3	3	2	36
30E	36	2	2	2	2	3	3	2	36
8A	8	3	3	1	1	4	4	3	8
8B	8	3	3	1	1	4	4	3	8
8C	8	3	3	1	1	4	4	3	8
8D	8	3	3	1	1	4	4	3	8
8E	8	3	3	1	1	4	4	3	8
8F	239	3	3	1	1	4	4	3	8

Table 2.10. MLST profiles of the 15 experimental isolates.

the contamination. The low coverage for ST239E suggested that this isolate may not be an ST239, as originally reported.

The MLST profiles of the isolates were determined (Table 2.10). ST30B was determined to have a novel MLST type, closely related to ST30 and ST36. ST239E was determined by MLST to be an ST8, and was therefore renamed as ST8F for the rest of the analysis.

3.4.1 De novo assembly

The sequences assembled successfully (Appendix B.1.6, A.1.1 – A.1.3), with a number of contigs ranging from 39 – 168, N50 ranging from 83,606 – 671,206, and a genome length ranging from 2,795,514 – 3,051,179 bp.

The pan-genome of all of the sequences contained a total of 4,928 genes, including 2,080 core genes that were shared between all isolates (Table 2.11). A maximum-likelihood phylogeny (Figure 2.6) showed the relatedness of the isolates, based on the core genes. The maximum SNP distance between the ST239 sequences was 501 SNPs, which was between ST239A and ST239D, and the minimum SNP distance was 83 SNPs, which was between ST239A and ST239B. The maximum SNP distance between the ST8 isolates was 574 SNPs between ST8E and ST8F, and the minimum SNP distance was 61 SNPs between ST8A and ST8B. The maximum SNP distance between the ST30 isolates was 735 SNPs between ST30A and ST30B, and the minimum SNP distance was 122 SNPs between ST30C and ST30D.

The combined phylogeny shows clustering of the ST239 isolates, ST8 isolates, and CC30 isolates. The ST8 isolates are more closely related to the ST239 isolates, which is to be

Gene Type	Frequency	Number of genes shared between all isolates	Number of genes shared between all ST239s	Number of genes shared between all ST8s	Number of genes shared between all ST30s
Core genes	99% – 100%	2,080	2,539	2,400	2,451
Soft core genes	95% – 98.9%	0	0	0	0
Shell genes	15% – 94.9%	1,324	726	1,348	772
Cloud genes	0% – 14.9%	1,524	0	0	0
Total genes	0% – 100%	4,928	3,265	3,748	3,223

Table 2.11. Breakdown of the number of genes in the pan-genome across all fifteen experimental isolates.

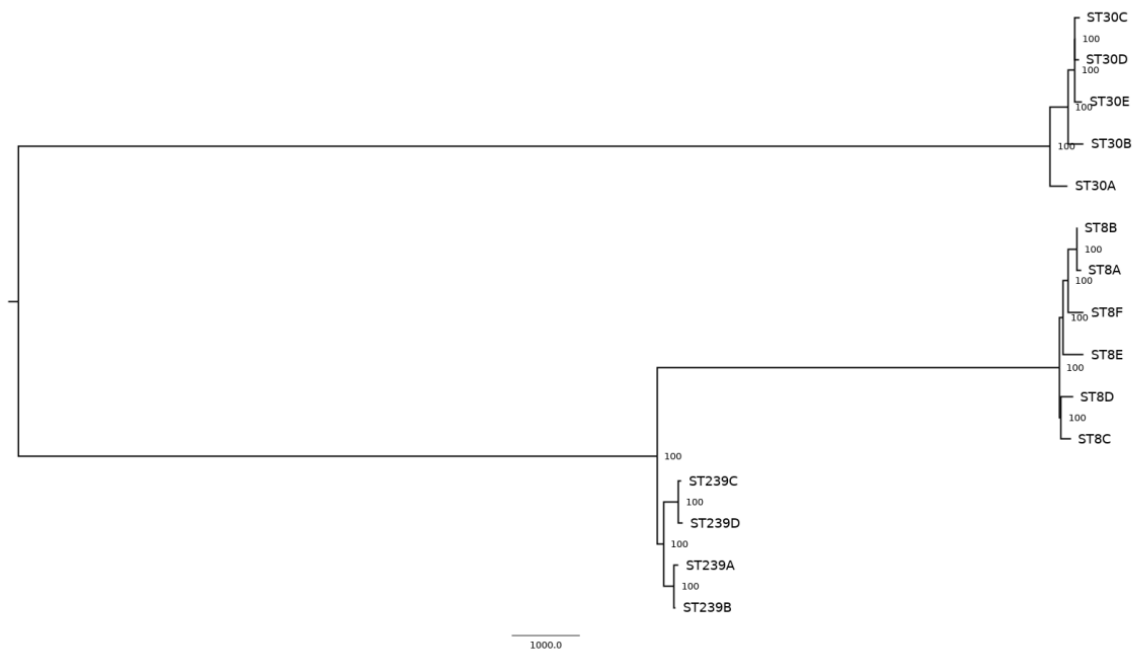


Figure 2.6. Mid-point rooted phylogeny of all experimental isolates. Units in total number of SNPs.

expected, as ST8 is predicted to be the donor of the large genetic region of ST239. The CC30 isolates were more closely related to the ST239 isolates than the ST8 isolates, which is also fitting with the hypothesis that ST30 is the donor of the smaller acquired region in ST239.

3.4.2 Identification of SCCmec type

The SCCmec type of the experimental isolates was determined (Table 2.12) by PCR (Appendix A.2.1; Appendix B.4.1.1 – B.4.1.3). No SCCmec bands were observed in the MSSA isolates. The SCCmec type was also determined genetically, and agreed with the PCR-determined typing.

All four of the ST239 isolates contained SCCmec-III, which is consistent with the high proportion of ST239 isolates that contain SCCmec-III in online sequence databases (Section 3.3.1). Three of the CC30 isolates contained SCCmec-II, and one CC30 contained SCCmec-IV, as well as three of the ST8 isolates. The remaining three ST8 isolates and one CC30 isolate were MSSA, and did not contain an SCCmec element.

ID	SCCmec type	ID	SCCmec type	ID	SCCmec type
ST239A	III	ST8A	MSSA	ST30A	IV
ST239B	III	ST8B	MSSA	ST30B	MSSA
ST239C	III	ST8C	IV	ST30C	II
ST239D	III	ST8D	IV	ST30D	II
		ST8E	IV	ST30E	II
		ST8F	MSSA		

Table 2.12. SCCmec type of the experimental isolates.

3.4.3 SCCmec-III variability

SCCmec-III is known to be highly variable (Monecke *et al.*, 2018), and this may contribute to difficulties in defining SCCmec-III type within a *S. aureus* genome (Section 3.3). Although all four ST239 isolates contained SCCmec-III (Section 3.4.2), there was a large amount of genetic variability between the SCCmec-III elements (Appendix B.4.2). Many of the genes that were shared between all of the ST239 SCCmec-III sequences have functions related to MGEs, metabolism, and AMR and heavy-metal resistance, including resistance to methicillin, streptomycin and erythromycin. However, 33 genes were unique to individual SCCmec-III sequences, and also include several MGE-related genes, metabolism genes, and some heavy metal resistance-related genes.

In addition to genetic variability caused by the presence and absence of genes, there were also multiple occurrences of gene duplication in the SCCmec-III elements of ST239A, B, and C, but not in ST239D (Table 2.13). Of these, the duplicate genes of *ant1* and *xerC* were located adjacent to one another on the MGE, and the *PROKKA_00056* integrase and *pspE* in ST239C are also located very close to one another along the MGE. These are therefore likely to be the result of gene duplications. The *hin* DNA invertase is involved in the inversion of genetic material, which may explain why multiple copies of the same *hin* gene are located distantly from one another on the MGE in ST239A, B and C. The presence of DNA manipulation genes, such as those encoding DNA invertase and recombinase enzymes, within the SCCmec-III element lends further evidence for the large amount of genetic recombination detected within ST239 SCCmec-III elements in Chapter 3.

Gene	Gene function	ST	Number of genes	Adjacent?
<i>ant1</i>	Streptomycin 3"-adenylyltransferase	ST239A	2	Yes
		ST239B	2	Yes
<i>hin</i>	DNA-invertase	ST239A	2	No
		ST239B	2	No
		ST239C	2	No
<i>PROKKA_00056</i>	Integrase core domain protein	ST239C	4	Close (separated by 1, and 3 and 6 genes)
<i>pspE</i>	Thiosulfate sulfurtransferase	ST239C	2	Close (separated by 1 gene)
<i>xerC</i>	Tyrosine recombinase	ST239A	2	Yes
		ST239B	2	Yes
		ST239C	3	Yes/No

Table 2.13. Duplicate genes in the SCC*mec*-III elements.

4. Discussion

The Staphopia database contains the genetic records of several thousand ST239, ST8 and ST30 *S. aureus* sequences. However, the majority of these are not associated with metadata such as date and location of isolation, or source, such as human blood, animal or environmental. Even where location or date are included, these are often limited to the country of isolation rather than a city or hospital, and a year rather than the month or day of isolation. Additionally, sequences that are associated with metadata are frequently from a small number of large-scale studies, and therefore, some dates and locations are heavily over-represented, such as during epidemics. Although ST239, ST8 and ST30 were represented around the globe, many geographic locations were not covered. This may be due to the non-presence of these isolates in these countries, however it is far more likely that studies to isolate and sequence a diverse range of healthcare-associated *S. aureus* isolates have not been carried out in these countries. Finally, the increasing quality of sequencing data over time may lead to potential batch effects, as improvements in sequencing quality mean that isolates that were sequenced further in the past may be of worse quality. These biases towards sampling, and lack of appropriate metadata for the majority of sequenced isolates, mean that conclusions drawn from these data are not truly representative of the global state of *S. aureus* epidemiology.

Taking into account the inherent biases within these online sequencing repositories, a total of 264 geographically and temporally diverse sequences were selected, including 96 ST239 isolates, 111 ST8 isolates and 57 ST30 isolates, and these were mapped to their respective high-quality reference sequences for further analysis in Chapters 3 and 4. Their dates of

isolation ranged from 1980 – 2016 (ST239), 1905 – 2016 (ST8), and 1935 – 2015 (ST30), and the isolates originated from countries across all six inhabited continents (with the exception of ST30, which was not sequenced from any country in Oceania).

Several studies have drawn conclusions surrounding the evolution of ST239, which were based on a single data-set by Harris that was published in 2010 as one of the first ever large-scale bacterial whole-genome sequencing projects (Harris *et al.*, 2010). However, this data-set contained only 45 ST239 isolates from a global isolate collection, and 20 from a single hospital in northeast Thailand that were isolated between 2006 – 2007 during an outbreak in an ICU. Harris employed similar methods to their initial sequencing analysis as those used in this study, for example, both studies mapped the ST239 sequences to the high-quality ST239 reference sequence, TW20. However, the data-set that has been prepared in this study captures a greater diversity of ST239 isolates.

The Staphopia database was mined for the presence of the *SCCmec*-III element within 3,178 ST239 sequences. 93.7% of the ST239 sequences were identified as containing *SCCmec*-III, and in the remaining sequences, an accurate *SCCmec* type could not be identified. This may be due to the loss of *SCCmec*, or simply due to poor sequencing quality of the highly variable *SCCmec*-III element. However, there is clearly a strong link between the ST239 genome and the *SCCmec*-III element, as no alternative *SCCmec* types were identified within the ST239 genomes. This may suggest that *SCCmec*-III cannot frequently be lost from the ST239 genome, for example due to a very low rate of excision, or may be evidence of a reliance of ST239 on the *SCCmec*-III element. This could be related to the high number of AMR genes

carried on the SCC*mec*-III element, which may therefore be key to the success of ST239 in clinical environments with high antibiotic usage.

The increased virulence of ST239 was initially observed by Holden (2010), who found that ST239 was four times more likely to result in bacteraemia compared to other epidemic strains, and also identified the presence of a number of virulence-associated genes (Holden *et al.*, 2010). This increased virulence has been linked to the success of ST239 as a healthcare-associated MRSA strain, including carriage of the prophage-mediated virulence determinant *sasX* (Li *et al.*, 2012), and the ACME-II variant (Espedido *et al.*, 2012). The *sasX* gene encodes a staphylococcal surface protein, which is a cell wall-associated protein that promotes nasal colonization, immune evasion and virulence, and is strongly associated with the ST239 genome (Li *et al.*, 2012), although it is found less commonly in other invasive MRSA strains (Kong *et al.*, 2018; Bakthavatchalam *et al.*, 2019; De Backer *et al.*, 2019). The arginine catabolic mobile element ACME contains the *arc* gene cluster *arcCBDAR*, which encodes a complete arginine deiminase pathway to breakdown L-arginine. It is associated with SCC*mec* elements and is linked to increased virulence and colonisation in *S. aureus*. The ACME-I variant is closely linked to the growth and survival of the prevalent community-associated ST8 USA300 strain (Diep *et al.*, 2008), and the ACME-II variant has been linked to the success of ST239 in healthcare settings (Espedido *et al.*, 2012).

Within the wider *S. aureus* population, three sequences that were not closely related to ST239 were identified that also contained SCC*mec*-III. These were ST21, ST5 and ST101. The detection of SCC*mec*-III in isolates that are not closely related to ST239 suggests that there is some mobility of the SCC*mec*-III element within *S. aureus*, although this is limited as the

SCC*mec*-III element has not become more prevalent in the *S. aureus* population, even though it has appeared within epidemic strains on rare occasions. This suggests that, although there is likely some benefit to the presence of SCC*mec*-III in ST239, this has not been the case for other epidemic MRSA lineages.

Although a high number of ST30 sequences are included in the Staphopia database, no ST30 isolates that contain SCC*mec*-III were identified, despite SCC*mec*-III being encoded within the hypothesised ST30-donated region of the ST239 genome. This may suggest that SCC*mec*-III may not have been present within the ST30-like ancestor of ST239, and may therefore have entered into ST239 separately to the hybridisation event.

A total of fifteen ST239, ST8 and ST30 isolates were obtained, whole-genome sequenced, and their MLST and SCC*mec* type was verified. These isolates will form the basis of Chapters 5 and 6. The ST239 isolates all contained SCC*mec*-III, but a large amount of variability between the SCC*mec*-III elements was observed. This variability within the SCC*mec*-III elements may contribute to difficulties in genetic typing of SCC*mec*-III elements.

ST239 MRSA has been reported as a dominant EMRSA strain in healthcare settings around the world since the late 1970s. However, there is evidence that ST239 is now being replaced by more fit MRSA strains, such as ST22. These strains do not encode the large SCC*mec*-III element that is strongly associated with ST239, which contains a high number of AMR genes. There is evidence that some of these strains may be more treatable with a wider range of antimicrobials than ST239, and may therefore contribute to the global trend towards the decreasing morbidity and mortality of MRSA infections. Over the next decade, the number of

Chapter 2

publicly-available *S. aureus* sequences in online sequence databases such as Staphopia will continue to increase, as DNA sequencing becomes cheaper, faster and more easily accessible. Significant improvements in the collection of metadata will hopefully enable large-scale global epidemiological analyses of MRSA, and other important infectious diseases, that will provide a basis to make robust conclusions and predictions about the behaviour of pathogens around the world.

Chapter 3

Testing the Robinson-Enright model of ST239

Abstract

In 2004, Robinson and Enright first hypothesised that ST239 was a hybrid bacteria that had formed through a single large chromosomal replacement event between an ST8-like isolate and an ST30-like isolate. However, their hypothesis was based on just 41 gene sequences from three ST239 isolates, two ST30 isolates and two ST8 isolates. Over the past decade, the vast increase in the number of bacterial whole-genome sequences in publicly available online sequence repositories has led to a much greater number of ST239, ST8 and ST30 sequences that can be accessed via DNA search engines. The incorporation of a wide range of ST239, ST8 and ST30 whole-genome sequences supports the Robinson-Enright model of the ST239 genome. ST30 of phage type 80/81 is the closest known relative of the acquired region of ST239, which suggests that the ST239 donor lineage may have been closely related to a healthcare-associated MSSA ST30 from this historic lineage. Both genetic regions of ST239 have subsequently evolved primarily through vertical inheritance, however the *SCCmec-III* element may not have evolved vertically alongside the rest of the ST239 genome. This suggests that since the initial large chromosomal replacement event that formed the ST239 genome, the long-term evolution of ST239 has not been driven by further large structural changes, but has primarily been driven by smaller genetic changes.

1. Introduction

1.1 Aims and objectives

In this chapter, I aim to utilise the increased diversity of *S. aureus* DNA sequence information in publicly available DNA sequence repositories to test the Robinson-Enright model of the ST239 hybrid with increased power. I aim to determine whether ST30 is the closest known ancestor of the true donor lineage, and to identify whether the subsequent evolution of ST239 has been impacted by further recombination events, regions of genetic variation, or large structural changes.

1.2 The hybridisation of ST239

In 2004, Robinson and Enright first proposed that ST239 was formed through a single large chromosomal replacement event, which they hypothesised after sequencing 41 genes across the ST239 genome (Robinson and Enright, 2004). They found that 22 of these genes,

Strain	Sequence type (ST)	SA0318	SA2339
		22445556689 3555688948341151440 4567637991503958496	3445566777788999 44581112837013526569 05257238934657427199
COL	ST250	CTCTGTCAGTAAATAGTAT	GGCGCGTATGAAACCTGTTC
E3008	ST8	CTCTGTCAGTAAATAGTAT	GGCGCGTATGAAACCGGTTC
Nott110	ST8	CTCTGTCAGTAAATAGTAT	GGCGCGTATGAAACCGGTTC
FFP200	ST239	TGGCACTTTCGGTGGGAT	GGCGCGTATAGACTAGGCCT
EMRSA11	ST239	TGGCACTTTCGGTGGGTAT	GGCGTACGTAGACTAGGCCT
Fin75541	ST239	TGGCACTTTCGGTGGGTAT	GGCGTACGTAGACTAGGCCT
E1410	ST30	TGGCACTTTCGGTGGTTCG	AATATACGTAGACTAGGCCT
Btn1260	ST30	TGGCACTTTCGGTGGTTCG	AATATACGTAGACTAGGCCT
MRSA252	ST36	TGGCACTTTCGGTGGTTCG	AATATACGTAGACTAGGCCT

Figure 3.1. Polymorphic sites spanning the left and right junctions of the ST239 mosaic chromosome. Loci are numbered in accordance with the finished genome sequence of strain N315 (Kuroda *et al.*, 2001). Parent nucleotides are shown with black and white backgrounds. A grey background indicates unique nucleotides. Image from Robinson and Enright, 2004.

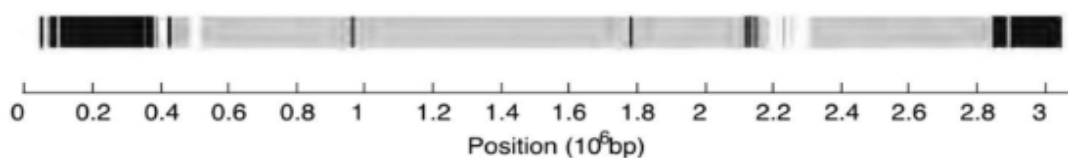


Figure 3.2. Alignment-free local homology analysis across the genome of *S. aureus* TW20. Regions inferred to have been transferred from CC30 (MRSA252) are represented in black, while regions homologous between TW20 and ST8 (TCH15156) are shown in grey. Image from Cong *et al.*, 2016.

asymmetrically orientated around the origin of replication, were closely related to ST30, and the remaining 19 genes were closely related to ST8. Further sequencing of the genetic boundaries of the hybridisation (Figure 3.1) led them to suggest that ST239 had acquired ~557 kb spanning the origin of replication (*oriC*) from an ST30-like parent, and ~2,220 kb spanning the terminus of replication (*terC*) from an ST8-like parent.

The Robinson-Enright model was later supported by Holden (2010), who published the first ST239 whole-genome sequence for the ST239 isolate TW20 (Holden *et al.*, 2010), and Castillo-Ramirez (2011), who explicitly stated the location of the hybridisation boundaries at base positions 427,726 and 2,848,036 along the TW20 genome (Castillo-Ramirez *et al.*, 2011). Subsequently, Cong (2016) visually represented the homologous regions of the TW20 complete genome compared with the ST8 TCH15156 and CC30 MRSA252 genomes (Figure 3.2) (Cong *et al.*, 2016).

1.3 Hybrid pathogenic bacteria

Although the identification of hybrid MRSA strains is rare, additional hybrid MRSA strains that are likely to have arisen from large-scale chromosomal recombination events have been characterised, including ST34, ST42, and ST2249. In these cases, the hypothesised large chromosomal replacement events that formed these hybrid MRSA genomes have involved

similar regions of the *S. aureus* genome, from an ST30-like donor. The ST42 genome has a primary ST10/ST145-like parent, and a single large chromosomal replacement making up ~8% of its genome originating from an ST39-like parent. Similarly, the ST34 genome consists primarily of an ST10/ST145-like parent, with a single chromosomal replacement making up ~4.4% of its genome originating from an ST30-like parent (Robinson and Enright, 2004; Thomas *et al.*, 2012). ST2249 shares 35.3% of its genome with an ST239-like parent, and 64.7% of its genome with an ST45-like parent (Nimmo *et al.*, 2014). Additionally, ST71 is a bovine-associated hybrid MRSA. Its major parent is ST97, however multiple successive recombination events have contributed to a region containing ~15% of the ST71 genome, which spans a similar region to the other hybrid *S. aureus* lineages (Spoor *et al.*, 2015). Finally, the genome of the hybrid ST582 MRSA consists primarily of an ST15 parent, with a single chromosomal replacement from an ST20-like parent making up ~10% of its genome (Didelot and Wilson, 2015). However, in this case, the region of DNA that has been horizontally transferred from the minor parent is a different region than the other *S. aureus* hybrid genomes, which does not span the origin of replication (Aanensen *et al.*, 2016). Although they are poorly understood, large-scale chromosomal replacement events have contributed to the evolution of *S. aureus* lineages on several occasions, however, only ST239 has emerged as a dominant international hybrid EMRSA strain.

There are further examples of highly successful pathogenic bacteria with composite genomes that may have also been formed by large chromosomal replacement events, which suggests that *S. aureus* is not the only pathogen that can undergo these large genetic transfer events. For example, *Klebsiella pneumonia* ST258 is a major cause of carbapenem-resistant *Enterobacteriaceae* (CRE) infections in healthcare facilities worldwide. The ST258 genome

Chapter 3

consists of 20% (1.1 Mb) from a *K. pneumonia* ST442-like parent, and 80% (4.2 Mb) from a *K. pneumonia* ST11-like parent (Chen *et al.*, 2014). The success of this hybrid lineage may have been driven by immune response and vaccine evasion, due to the transferred *cps* locus for capsule polysaccharide, which resulted from this hybridisation event (Croucher and Klugman, 2014). During the formation of several prevalent *K. pneumonia* lineages, very large recombination events, >100 kb in size, have occurred in two defined regions of the genome, which suggests that large chromosomal replacement events have had an important role in the emergence and diversification of these lineages (Comandatore *et al.*, 2019). This is further evidenced by the fact that, in pathogenic *K. pneumonia* lineages, although macro-chromosomal replacement events occur more rarely than micro-chromosomal replacement events, larger chromosomal replacement events have been the more major drivers of diversification (Mostowy *et al.*, 2014). Similar large chromosomal replacement events of >100 kb have also been identified in the formation of epidemic *Clostridium difficile* strains (He *et al.*, 2012).

Successive recombination events can also contribute to the formation of hybrid pathogens. For example, *Campylobacter coli* ST1150, which is a major cause of human gastroenteritis, has progressively acquired up to 23% of its genome through successive recombination events from the distantly-related *C. jejuni* (Sheppard *et al.*, 2013). Additionally, chromosomes that are made up of a mosaic of large chromosomal fragments from several different donors have also been identified. For example, *Streptococcus agalactiae* is the leading cause of neonatal infections in the developed world. Its chromosome is made up of a mosaic of large chromosomal fragments from several different donors, suggesting that multiple large DNA exchanges have occurred during its evolution (Brochet *et al.*, 2008). The *S. agalactiae*

genome is still evolving through conjugative events, frequently exchanging ~35 kb regions of DNA that are responsible for surface-antigen biosynthesis, which has a major impact on *S. agalactiae* vaccine success (Bellais *et al.*, 2012). Using *in vitro* experiments, Brochet (2012) were able to transfer up to 334 kb of the *S. agalactiae* genome through conjugation from multiple sites in the genome.

1.4 Chapter summary

In this chapter, the increased availability of bacterial DNA sequences in online sequence repositories was used to investigate the Robinson-Enright model of ST239; that ST239 is a natural hybrid that formed through a single large chromosomal replacement event between an ST8-like ancestor and an ST30-like ancestor.

Analysis of the core genes that are shared between a total of 264 ST239, ST8 and ST30 isolates (Chapter 2), confirmed that the regions of similarity between the STs are consistent with a single large chromosomal replacement event, and that the hybrid boundaries are consistent with those proposed by Castillo-Ramirez (2011). Large regions of genetic variation that were identified in the ST239 genome have been attributed to known variable regions such as phage, transposons and *SCCmec*. No large structural changes, such as large regions of genetic deletions/insertions or rearrangements, have occurred within the ST239 genomes compared to the ST8 and ST30 genomes, although some large inversions have occurred in some genomes.

The *SCCmec*-III element within the ST239 genomes was found to be variable, and not consistent with vertical evolution of this large MGE in ST239. However, the acquired region

Chapter 3

and backbone region did show evidence of the vertical evolution of the ST239 genome since the formative hybridisation event.

Comparing the genetic distances of the two regions of ST239 with the homologous regions in ST8 and ST30, showed that the ST8 sequences were more closely related to the ST8-like region of ST239 than the ST30 sequences to the ST30-like region of ST239. Following this, five ST30 isolates were determined to be the closest known relatives of the ST30-like region, from an analysis of all available bacterial DNA sequences. These five ST30 sequences were globally diverse, phage type 80/81 isolates from 1955 – 1968 and were all absent of *SCC_{mec}*, which suggests that the ST239 donor may have been closely related to a healthcare-associated MSSA ST30 from this historic lineage.

2. Methods

2.1 Genetic analysis of ST239

2.1.1 Artemis comparison tool

In Chapter 2 Section 2.4, the core genome of 96 ST239, 111 ST8, and 57 ST30 sequences was calculated. The 2,178 TW20 core genes were extracted and compared with the reference TW20 whole-genome sequence (accession number FN433596) using Artemis Comparison Tool v13.0.0 (Carver *et al.*, 2005).

2.1.2 Variable regions

ClonalFrameML v1.0-20 (Didelot and Wilson, 2015) was used to estimate regions of recombination in the 96 ST239 sequences that were mapped to the TW20 reference in Chapter 2 Section 2.4. Gene function for genes that could be annotated by Prokka (Seeman, 2014) was identified in UniProt (The UniProt Consortium, 2018), and MegaBLAST (Morgulis *et al.*, 2008) was used to identify genes where no annotation was found with Prokka. All graphs were plotted in R v3.4.4.

2.1.3 Hybrid boundaries

EMBOSS v6.6.0.0 (Rice *et al.*, 2010) was used to generate a consensus sequence from the core gene alignment of the ST239 core gene sequences, generated in Chapter 2 Section 2.4. This was repeated to produce an ST8 consensus sequence, and an ST30 consensus sequence, from the ST8 and ST30 core gene alignments that were generated in Chapter 2 Section 2.4. The percentage identity of each ST8 core gene consensus sequence, and each ST30 core gene consensus sequence, was calculated against each ST239 core gene consensus sequence using

MegaBLAST (Morgulis *et al.*, 2008). All graphs were plotted in R v3.4.4. ST239 genes that were more similar to the ST30 consensus sequence were labelled as “ST30-like”, and genes that were more similar to the ST8 consensus sequence were labelled as “ST8-like”. Gene function was assigned as in Section 2.1.2.

2.2 Structural comparison of ST239, ST8 and ST30 complete genome sequences

Complete genomes were downloaded from NCBI for ST239 (accession numbers CP005288, CP012015, CP012011, CP007447, CP006838, CP002120, CP013957), ST8 (accession numbers CP007690, CP019590, CP007672, CP026961, CP022290, AP014921, CP014362, CP003033, CP016855, CP026076), and ST30 (accession numbers LN626917, CP017684, CP009554). These were compared with the ST239 TW20 complete genome sequence (accession number FN433596) using Artemis Comparison Tool v13.0.0 (Carver *et al.*, 2005). SCCmec type was identified using SCCmecFinder (Kaya *et al.*, 2018).

2.3 Extracting the hybrid regions and SCCmec region from the ST239 genomes

The ST8-like genes were extracted from the ST239 sequences that were mapped to TW20 in Chapter 2 Section 2.4, using the hybrid boundaries that were identified in Section 2.1.3.

The boundaries of the SCCmec-III element were identified in the TW20 reference genome using BLAST against the reference SCCmec-III element with the accession number AB037671 (Ito *et al.*, 2001), and confirmed using SCCmecFinder (Kaya *et al.*, 2018). The corresponding SCCmec-III element genes were extracted from the ST239 sequences. The ST30-like genes, minus genes that were present in the SCCmec-III element, were also extracted.

Recombination was masked using Gubbins v1.4.5 (Croucher *et al.*, 2015), and maximum likelihood phylogenies were estimated using RaxML v8.2.9 (Stamatakis, 2014), using a GTR model with gamma correction for among site rate variation, which was replicated for 100 bootstraps. All trees were viewed in Figtree v1.4.4 (Rambaut, 2010). Tanglegram alignments were generated in dendroscope v3.5.9 (Huson and Scornavacca, 2012).

2.4 Identifying the most likely ST239 donor

2.4.1 Genetic distance between the homologous regions of ST239, ST8 and ST30

As in Section 2.3, the genetic region that was homologous to the ST8-like region was extracted from the ST8 sequence alignment from Chapter 2 Section 2.4, and the genetic region that was homologous to the ST30-like region was extracted from the ST30 sequence alignment from Chapter 2 Section 2.4. These alignments were combined with the corresponding genetic regions of the 96 ST239 sequences, generated in Section 2.3. Recombination was masked and maximum-likelihood phylogenies were estimated and viewed as in Section 2.3. Pairwise SNP distances were calculated using snp-dists v0.6.3 (Seemann, 2007).

2.4.2 Finding the closest known ancestor of the acquired region

EMBOSS v6.6.0.0 (Rice *et al.*, 2010) was used to generate consensus sequences for the 316 core genes from the acquired region of the ST239s, that was extracted in Section 2.3. These consensus sequences were queried for similar sequences in BIGSI (Bradley *et al.*, 2019) using a k-mer threshold of 99. This allowed for an average divergence of around 10 SNPs per gene. The most closely related sequences were identified by ST using the Staphopia API (Petit *et al.*, 2018). The MLST was double checked using Ariba v2.13.3 (Hunt *et al.*, 2017), and 28

Chapter 3

sequences were removed due to uncertain MLST typing, which indicated contamination. All graphs were plotted in R v3.4.4.

Sequences that shared at least 99% k-mer identity with over 200 of the ST239 core genes from the acquired region, and shared two or fewer MLST alleles with ST239, were downloaded from the EBI database. Only a selection of the twelve most closely related ST30, ST36 and ST39 sequences were included. These 143 sequences were mapped to the ST239 reference sequence (accession number FN433596), as in Chapter 2 Section 2.4. Four sequences were removed due to poor mapping quality (>25% gaps), leaving 139 non-ST239-like sequences. The acquired region was extracted (minus the *SCCmec* element), as in Section 2.3. These 139 sequences were combined into a multifasta alignment with the acquired regions from the 96 ST239 sequences from Section 2.3 and the 57 ST30 sequences from Section 2.4.1, and all variant sites extracted using snp-sites v2.3.3 (Page *et al.*, 2016). RAxML was used to estimate a maximum-likelihood phylogeny of all 235 sequences, rooted to the corresponding region of ST8 CP007690, as in Section 2.3. All trees were viewed in Figtree v1.4.4 (Rambaut, 2010).

The five sequences from the most closely related clade were downloaded from the EMBL-EBI database (Madeira *et al.*, 2019.), and the *SCCmec* type was identified using *SCCmecFinder* (Kaya *et al.*, 2018).

3. Results

3.1 Genetic analysis of ST239

3.1.1 Regions of variability in ST239 genomes

The 2,178 core genes that were identified in Chapter 2 Section 2.4, that were shared between all 96 ST239, 111 ST8 and 57 ST30 sequences, were aligned against the reference TW20 complete genome sequence (Figure 3.3). These were well distributed across the genome, which means that further analysis using these genes is not biased towards any particular region of the genome. Therefore, these core genes can be used to make valid conclusions about both the acquired and backbone regions of the ST239 genome. However, some small regions were lacking core genes, which indicates some regions of variability.

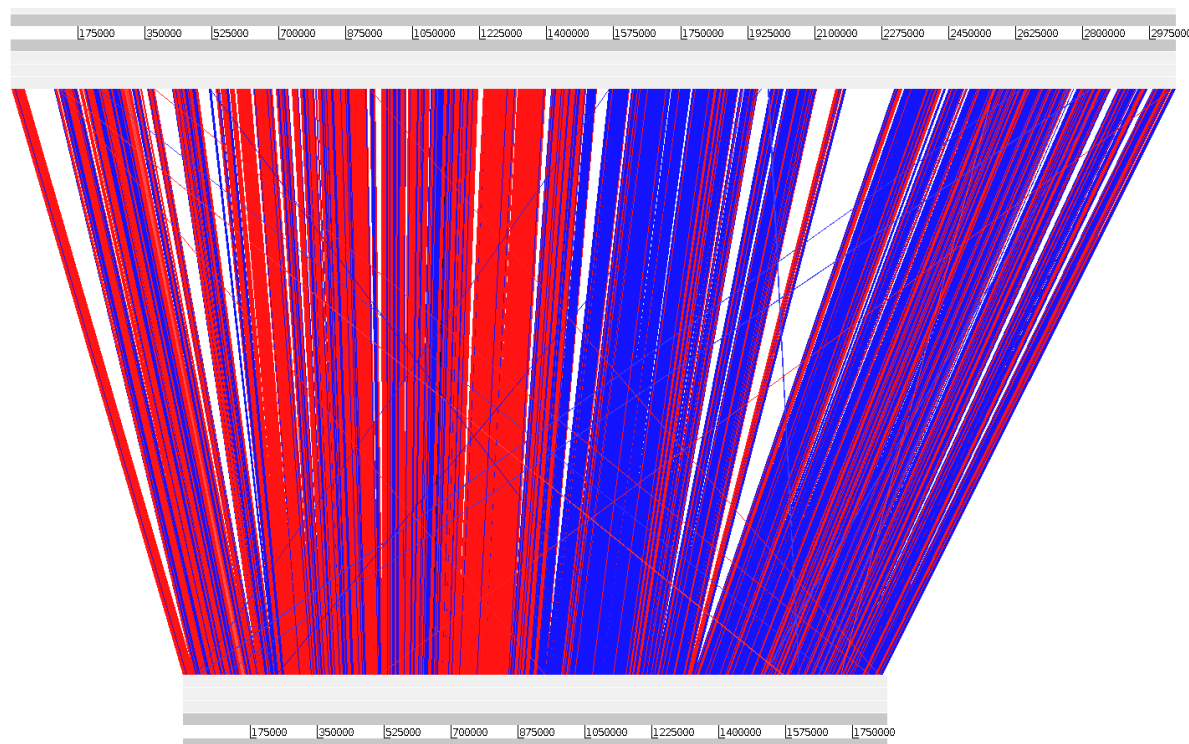


Figure 3.3. Artemis Comparison Tool alignment of the reference TW20 whole-genome sequence (top) against the core genes (bottom). Blue, inverted genes; red, forward genes.

ClonalFrameML (Didelot and Wilson, 2015) was used to identify regions of recombination that were present in the 96 ST239 genomes (Figure 3.4 and 3.5). Four major regions of recombination were predicted, numbered 1 – 4, which corresponded to base positions in the TW20 reference genome of 45,000 – 95,000, 375,000 – 420,000, 1,775,000 – 1,785,000, and 2,105,000 – 2,155,000. These regions of recombination were not present in the internal control (TW20 mapped to itself), or in the isolates that are most closely related to TW20, as TW20 was used as the reference. The four major regions of recombination corresponded to the largest gaps in core-gene presence along the TW20 genome in Figure 3.3.

The first variable region corresponds with the SCC*mec*-III element. The region contains genes that are typically found within the SCC*mec*-III element such as, methicillin resistance, transposase genes, heavy metal resistance, and recombinase genes (Monecke *et al.*, 2018). The fact that these were identified through the recombination analysis shows that the non-TW20-like isolates do not contain identical SCC*mec*-III elements to TW20, which suggests that there is variability within the SCC*mec*-III elements of ST239. This is consistent with other studies into the evolutionary history of SCC*mec* elements, including SCC*mec*-III (John *et al.*, 2019 ; Monecke *et al.*, 2018).

The second and fourth variable regions contain genes encoding primarily phage proteins, which may be present in the ST239 TW20 reference genome, but not within the majority of the ST239 sequences; ϕ Sa1(TW20) and ϕ Sa3(TW20) (Holden *et al.*, 2010). A third phage, ϕ SP β (TW20), was identified in the TW20 genome by Holden (2010). This phage is present in most of the 96 ST239 sequences, and is located close to the ϕ Sa3(TW20) phage on the TW20 genome. This conserved phage is hypothesised to contain a CDS that may have a role in

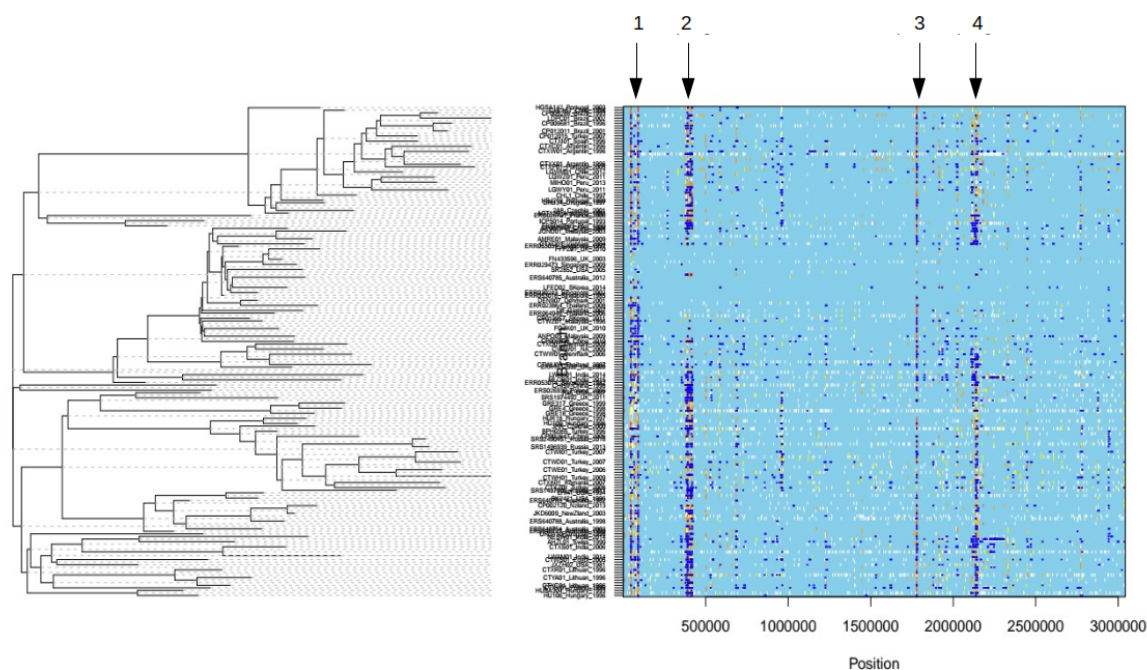


Figure 3.4. ST239 maximum-likelihood phylogeny displayed next to predicted recombination plot against the TW20 whole genome. Regions of high variability are labelled 1, 2, 3 and 4. Dark blue regions represent potential sites of recombination. Sites that are non polymorphic for a given branch are shown in light blue. Polymorphic sites are shown in a colour indicating their level of homoplasy; white is no homoplasy and the range from yellow to red is increasing degrees of homoplasy. Produced using ClonalFrameML (Didelot and Wilson, 2015).

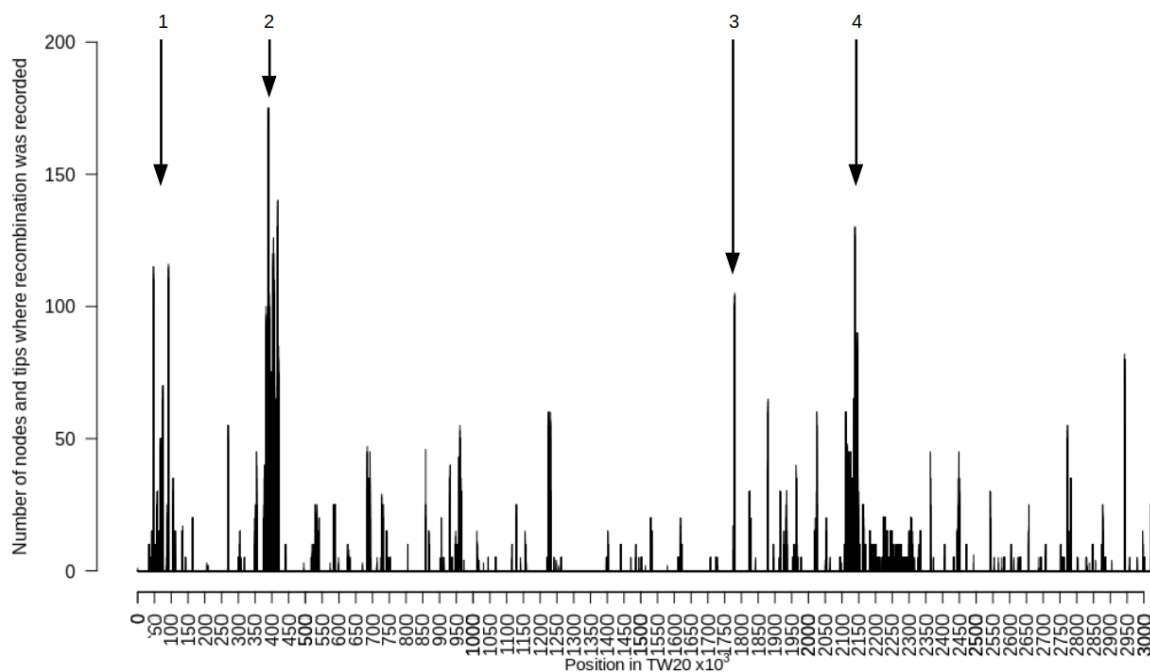


Figure 3.5. Amount of estimated recombination recorded by ClonalFrameML at each position along the TW20 genome. Regions of high recombination are labelled 1, 2, 3 and 4.

promoting the persistence of TW20 in hospital settings (Holden *et al.*, 2010). The third variable region is made up of Tn554 transposon genes (Holden *et al.*, 2010). A complete list of genes identified in each region is in Appendix B.4.3.1 – B.4.3.4. All four of these variable regions were included in a group of loci within the ST239 TW20 genome that were classified by Castillo-Ramirez (2011) as “extra-chromosomal elements” that had a likely horizontal origin.

Within the ST239 genome, Castillo-Ramirez found that recombination affected 3 – 10% of genes, excluding genes that are associated with MGEs (Castillo-Ramirez, *et al.*, 2012). The major functions of these genes related to cellular processes, macro-molecule metabolism, small molecule metabolism, cell envelope and regulatory functions, as well as some genes with unknown function. The regions of recombination identified here were consistent with those identified by Castillo-Ramirez (2012), and no further large regions of variation between the ST239 genomes were identified. This suggests that, since the global expansion of ST239, no additional major large recombination events have occurred in separate lineages of ST239, and that ST239 has been evolving primarily through vertical inheritance. This suggests that the global population of ST239 has occurred from a single ancestor with no major additional recombination events.

3.1.2 Identification of the hybrid regions in ST239

The BLAST DNA percentage identity of the ST8 and ST30 consensus sequences, compared to the ST239 consensus sequence showed a clear discrepancy between the two genetic regions (Figure 3.6), which was consistent with the shift of roughly 1% DNA percentage identity, that was observed by Holden (2010) between the genetic regions of TW20 and a CC30 and ST8

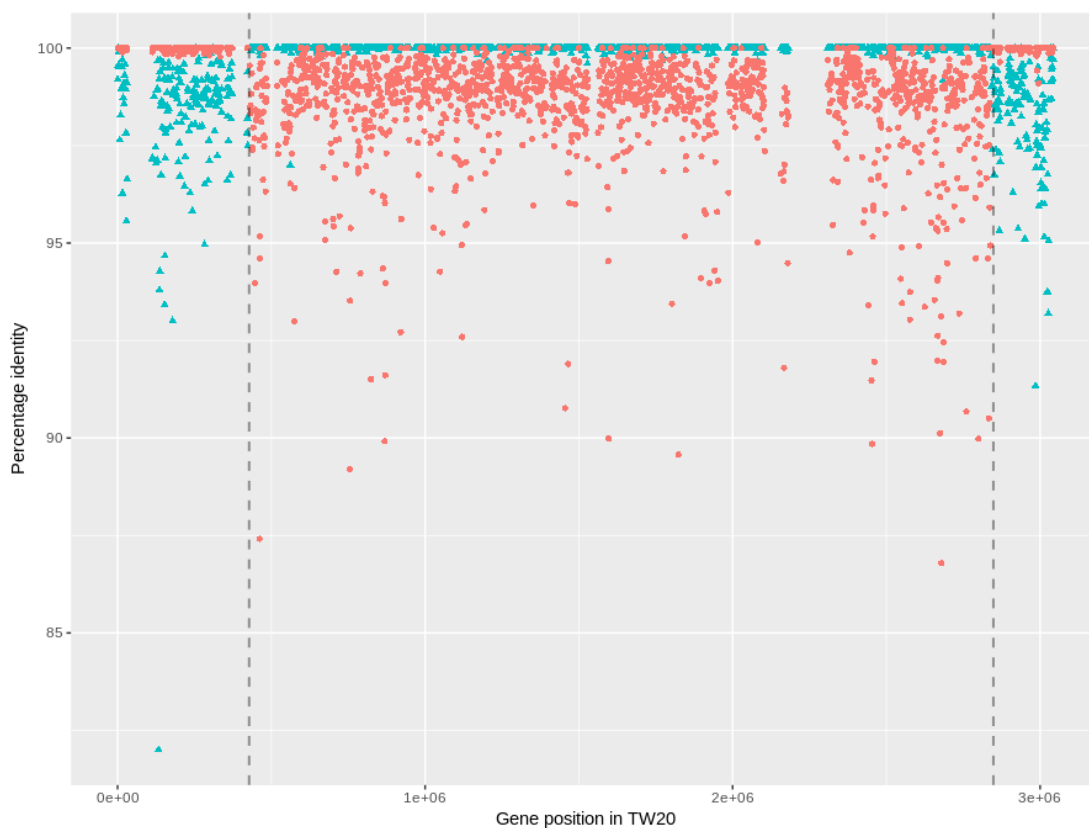


Figure 3.6. BLAST DNA percentage identity of the ST8 and ST30 consensus core genes, compared to the ST239 consensus core genes. Each data-point represents a single gene. The gene positions correspond to the genomic position within the TW20 reference sequence. Blue triangle, ST8; red circle, ST30. The hybrid boundaries estimated by Castillo-Ramirez (2011) are highlighted with vertical dashed grey lines.

isolate (Holden *et al.*, 2010). None of the core genes shared less than 82% identity with both the ST8 and ST30 consensus sequences. The ST8 genes shared much greater homology with ST239 within the central region of the TW20 genome, which is hypothesised to be the ST8-donated region. The ST30 genes shared much greater homology with ST239 within the two outer regions of the TW20 genome, which are hypothesised to be the ST30-donated regions. This supports a single chromosomal transfer event (Robinson and Enright, 2004), rather than multiple smaller HGT events. Although there is a wide coverage of genes that are core to ST239, ST8 and ST30 across the entire TW20 genome length, several gaps in coverage were present, which correspond to the same regions that were identified in Section 3.1.1.

Chapter 3

The hybridisation boundaries were characterised at the positions along the TW20 genome where the percentage identity switched from being most similar to ST8 to being most similar to ST30. To confirm the location of the hybrid boundaries between these core genes, the percentage identity and query cover of the ten core genes closest to the hybridisation boundaries were calculated (Table 3.1 and 3.2). This showed that hybridisation boundary A occurs within 423,903 – 429,719 bp, between the *SATW20_03930* gene on the ST30 side and the *cmtB* gene on the ST8 side. Hybridisation boundary B occurs within 2,841,155 – 2,842,225 bp, between the *ydfJ* gene on the ST8 side and the *rocA* gene on the ST30 side.

Gene	Gene function	Base position in TW20	ST8 percentage identity	ST30 percentage identity	Likely donor
<i>ylbJ</i>	Sporulation integral membrane protein	372,741 – 374,102	98.825	100	ST30
<i>SATW20_03900</i>	NADH-dependent flavin oxidoreductase	421,013 – 422,227	97.821	100	ST30
<i>limB_1</i>	Putative bacterial luciferase family protein	423,536 – 422,535	98.204	100	ST30
<i>SATW20_03920</i>	Glycine cleavage system protein H	423,902 – 423,570	99.399	100	ST30
<i>SATW20_03930</i>	Protein-ADP-ribose hydrolase	424,703 – 423,903	97.503	100	ST30
<i>cmtB</i>	Putative PTS transport system IIA component	429,719 – 430,162	100	99.550	ST8
<i>licR</i>	Putative PTS multi-domain regulator	430,167 – 432,122	100	98.466	ST8
<i>slyA_1</i>	MarR family transcriptional regulator	432,753 – 432,334	100	98.333	ST8
<i>mepA</i>	Multidrug export protein	434,215 – 432,860	100	98.525	ST8
<i>glpT</i>	Putative glycerol-3-phosphate transporter	434,842 – 436,200	99.926	98.823	ST8

Table 3.1. Core gene similarities between ST239, ST8 and ST30 at the genes on either side of the first hybridisation boundary.

Gene	Gene function	Base position in TW20	ST8 percentage identity	ST30 percentage identity	Likely donor
<i>clpL</i>	Putative ATP-dependent protease ATP-binding, subunit ClpL	2,835,986 – 2,833,881	100	90.508	ST8
<i>SATW20_26_710</i>	FeoB-associated Cys-rich membrane protein	2,836,055 – 2,836,225	100	95.906	ST8
<i>feoB_1</i>	Putative ferrous iron transport protein B	2,836,238 – 2,838,232	99.950	98.596	ST8
<i>SATW20_26_730</i>	Ferrous iron transport protein A	2,838,244 – 2,838,471	100	100	ST8
<i>ydfJ</i>	Putative transport protein	2,838,687 – 2,841,155	99.797	94.937	ST8
<i>rocA</i>	Aldehyde dehydrogenase family protein	2,842,225 – 2,843,769	99.029	100	ST30
<i>SATW20_26_780</i>	Hypothetical protein	2,844,988 – 2,844,797	97.396	100	ST30
<i>copA</i>	Putative copper importing ATPase A	2,847,636 – 2,845,228	99.045	100	ST30
<i>copZ</i>	Putative heavy-metal-associated protein	2,848,387 – 2,848,181	99.517	100	ST30
<i>ldhD_2</i>	D-isomer specific 2-hydroxyacid dehydrogenase	2,848,477 – 2,849,505	99.223	99.903	ST30

Table 3.2. Core gene similarities between ST239, ST8 and ST30 at the genes on either side of the second hybridisation boundary.

These boundaries are consistent with those estimated by Castillo-Ramirez (2011), and by Cong (2016).

Genes were allocated as “ST30-like” or “ST8-like”, compared to their position in the TW20 genome. Out of a total of 575 genes identified in the ST30-like region of the TW20 genome, 316 (55%) were classified as core genes that were present in >99% of all 96 ST239, 111 ST8 and 57 ST30 sequences. Out of a total of 2,382 genes identified in the ST8-like region of TW20, 1,659 (70%) were classified as core genes. None of the core genes were from within the ~60kb *SCCmec*-III region, which may account for the 15% difference between the ST30

and ST8 region core genes, as the *SCCmec*-III element is within the hypothesised ST30-donated region. Five genes were removed from the analysis because it could not be determined whether they originated from the ST30-like or ST8-like region, as they sat within the hybrid boundaries and had 100% identity to both the homologous ST30 and ST8 gene sequences.

3.2 Structural comparison of ST239, ST8 and ST30 complete genome sequences

Seven ST239, ten ST8 and three ST30 complete genomes were directly compared with the ST239 TW20 complete genome sequence (Figure 3.7a – 3.7c, Appendix B.1.7). Overall, the structure of the complete ST239 genomes aligned well with the TW20 reference genome sequence, which confirms that TW20 is a suitable reference to map ST239 sequences against. There was also a large amount of similarity between the structures of the ST8 and ST30 complete genome sequences with the TW20 reference genome sequence.

Many of the ST239 complete genomes, and all of the ST8 and ST30 complete genomes, had a large deletion around 2,275,000, which corresponds to the fourth variable region phage ϕ Sa3(TW20) (Section 3.1.1), suggesting that this phage is absent from other genomes. Additionally, the majority of genomes have a large deletion before 175,000, which corresponds to the *SCCmec*-III element. The variation in how much sequence is absent around the *SCCmec* area is due to the presence of different amounts of shared genes within the different *SCCmec* types.

All of the ST8 and ST30 complete genome sequences were also missing regions close to the second hybrid boundary, at 2,873,692 – 2,921,985, which correspond to a Tn552 transposon

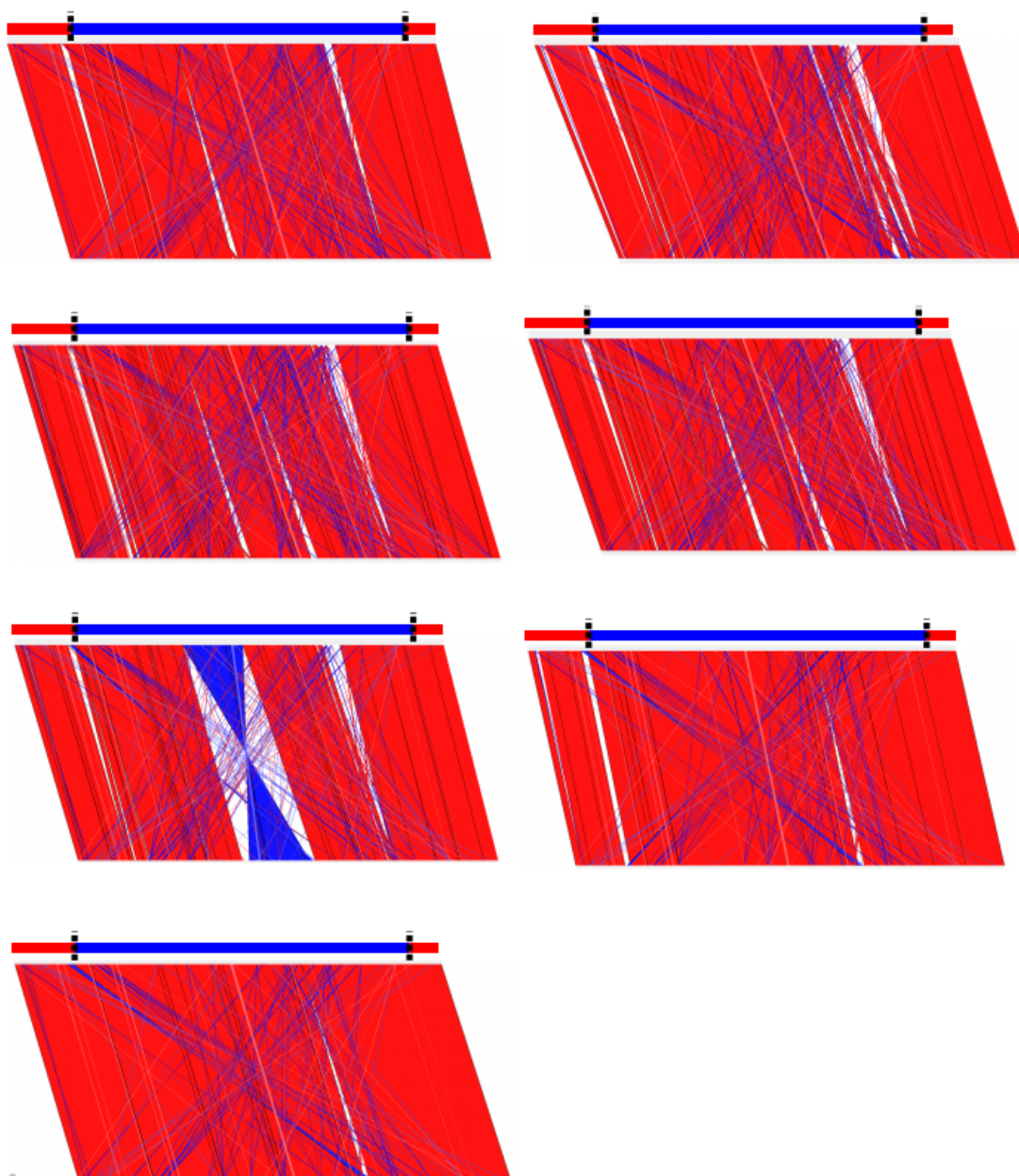


Figure 3.7a. Artemis comparison of the TW20 complete genome (top) against ST239s CP006838, CP002120, CP012011, CP012015, CP005288, CP007447, CP013957 (bottom). Red, forward matches; blue, inversions. The bar at the top of each map represents the ST239 genome, including the hybrid boundaries.

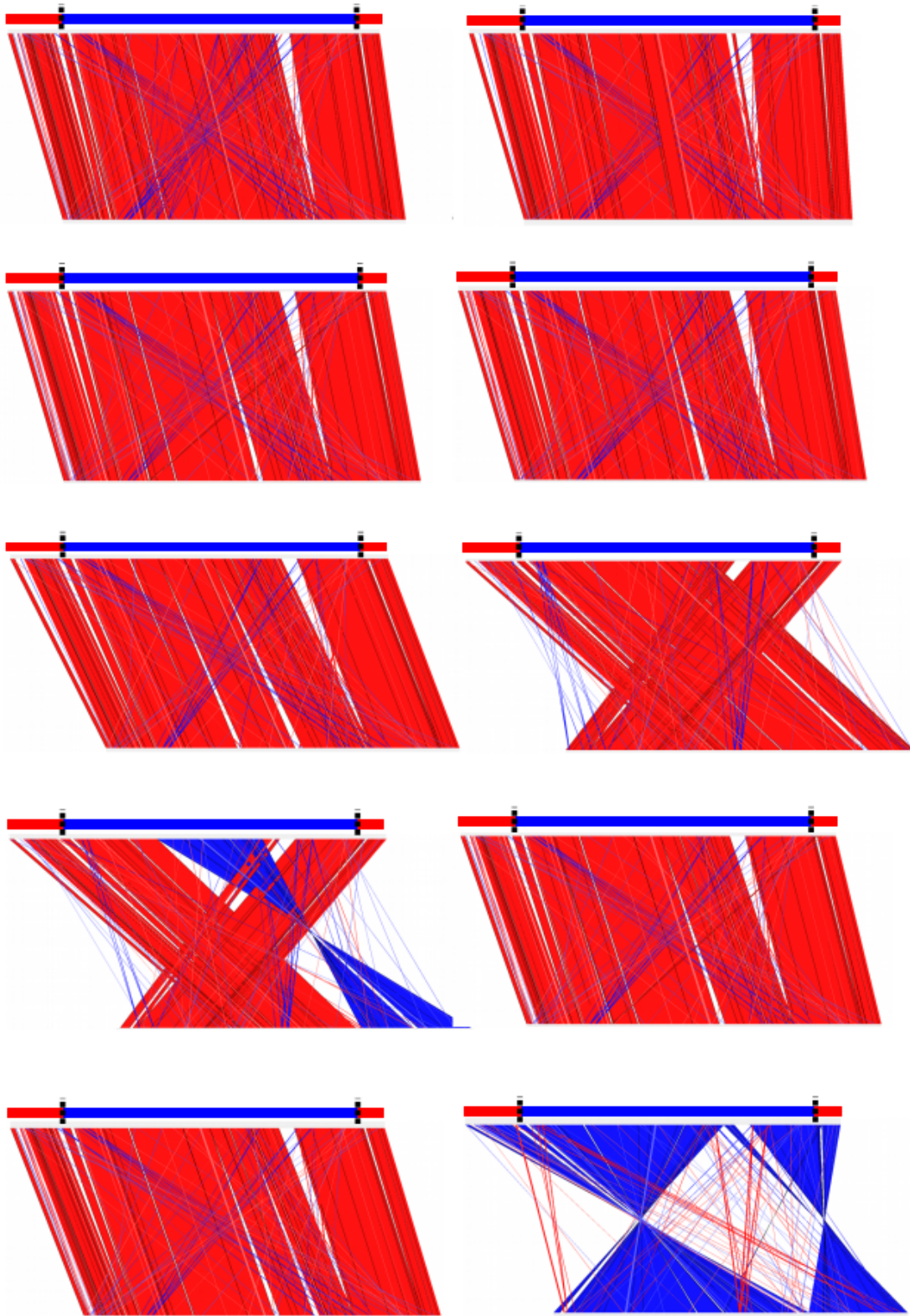


Figure 3.7b. Artemis comparison of the TW20 complete genome (top) against ST8s AP014921, CP0003033, CP007690, CP007672, CP022290, CP026076, CP026961, CP019590, CP014362, CP016855 (bottom). Red, forward matches; blue, inversions. The bar at the top of each map represents the ST239 genome, including the hybrid boundaries.

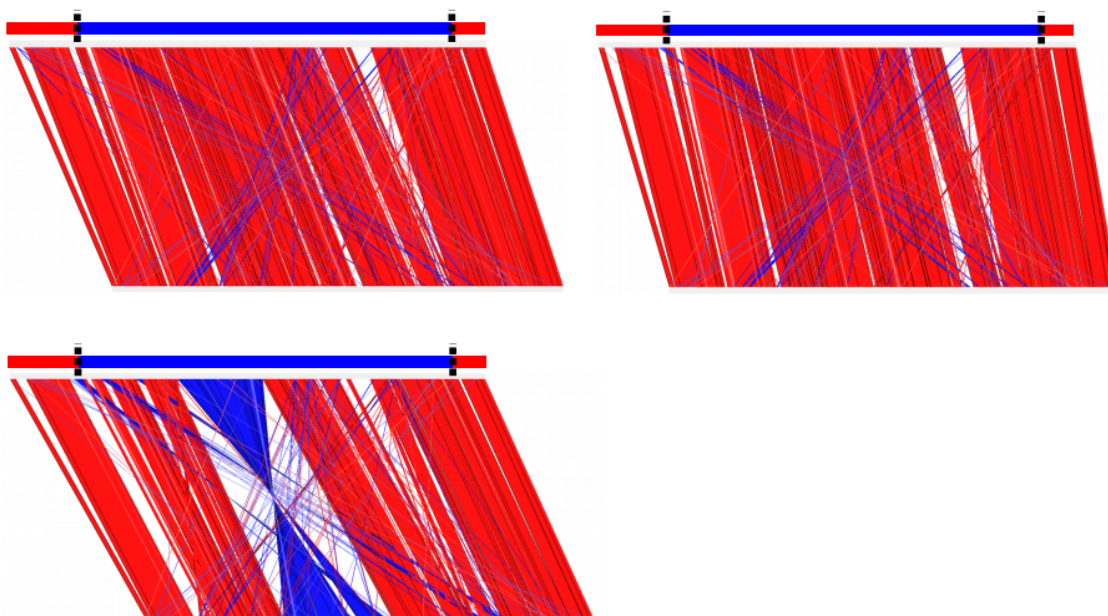


Figure 3.7c. Artemis comparison of the TW20 complete genome (top) against ST30s CP009554, CP017684, LN626917 (bottom). Red, forward matches; blue, inversions. The bar at the top of each map represents the ST239 genome, including the hybrid boundaries.

Accession	ST	Inversion	Start position in TW20	End position in TW20
LN626917	ST30	Yes	1,075,904	1,617,222
CP025288	ST239	Yes	1,193,109	1,619,588
CP026961	ST8	Yes	1,193,109	2,043,494
CP026076	ST8	No	2,020,282	-
CP016855	ST8	Yes	1	2,095,131
CP016855	ST8	Yes	2,095,131	3,043,210

Table 3.3. Large inversions and rearrangements in complete genomes, compared to TW20.

containing β -lactamase genes (Holden *et al.*, 2010), which is present within all of the ST239 complete genomes, as well as in all 96 of the ST239 genomes from Section 3.1.1. Therefore, this was most likely acquired by ST239 in a separate event to the hybridisation event.

Five of the sequences had very large inversions, or large rearrangements, compared to the TW20 genome (Table 3.3). Three sequences had an inversion in a similar region of the genome, ranging from 1,075,904 – 2,043,494. A different rearrangement occurred in two of the ST239 genomes around 2,000,000. These may be linked to the nearby transposase.

3.3 Extracting the genetic regions of ST239

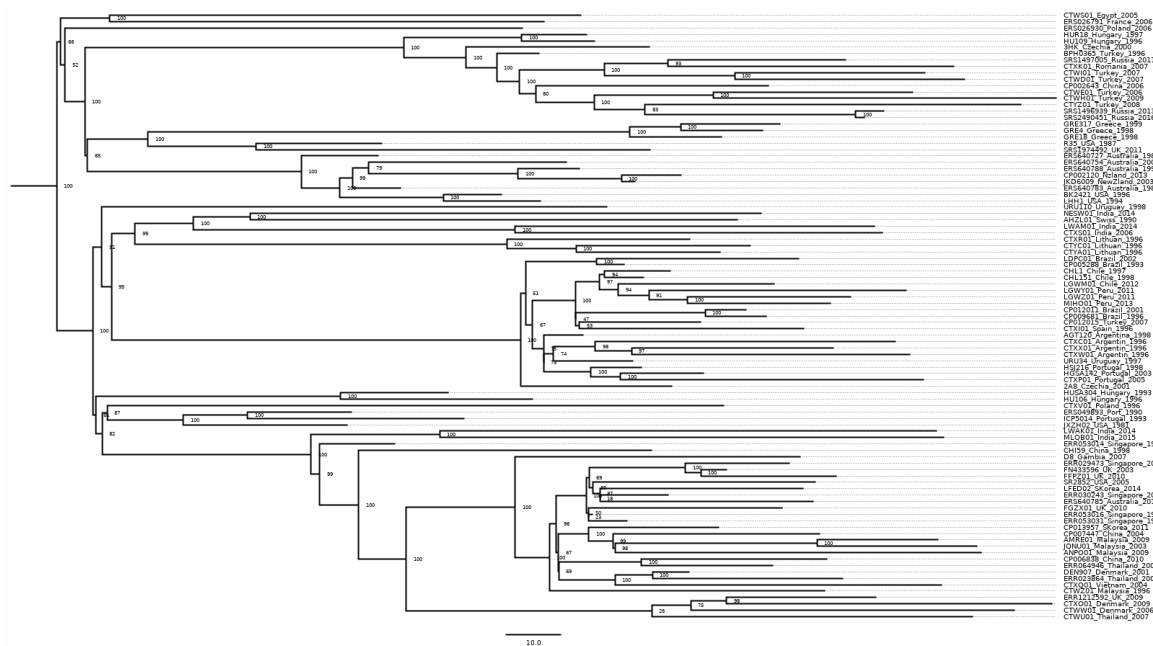
3.3.1 Hybrid regions

Maximum-likelihood phylogenies were estimated for the two genetic regions from the 96 ST239 sequences collected in Chapter 2 (Figure 3.8). The bootstrap support was relatively robust for each phylogeny, and the strong alignment of the phylogenies of both regions lends further support to the vertical evolution of both regions of the ST239 genome (Figure 3.9).

3.3.2 SCC*mec*-III

A maximum-likelihood phylogeny was also estimated for the SCC*mec*-III region from the 96 ST239 sequences collected in Chapter 2 (Figure 3.10). Several of the nodes had very low bootstrap support, particularly at the basal nodes. This may be causative of the large discrepancies between the SCC*mec*-III and the ST30-like region phylogenies, and may be further evidence that the SCC*mec*-III element has not been evolving entirely by vertical inheritance, alongside the rest of the ST239 genome. (Figure 3.11). However, overall the SCC*mec*-III phylogeny and the ST30-like region phylogeny were relatively well aligned.

A



B

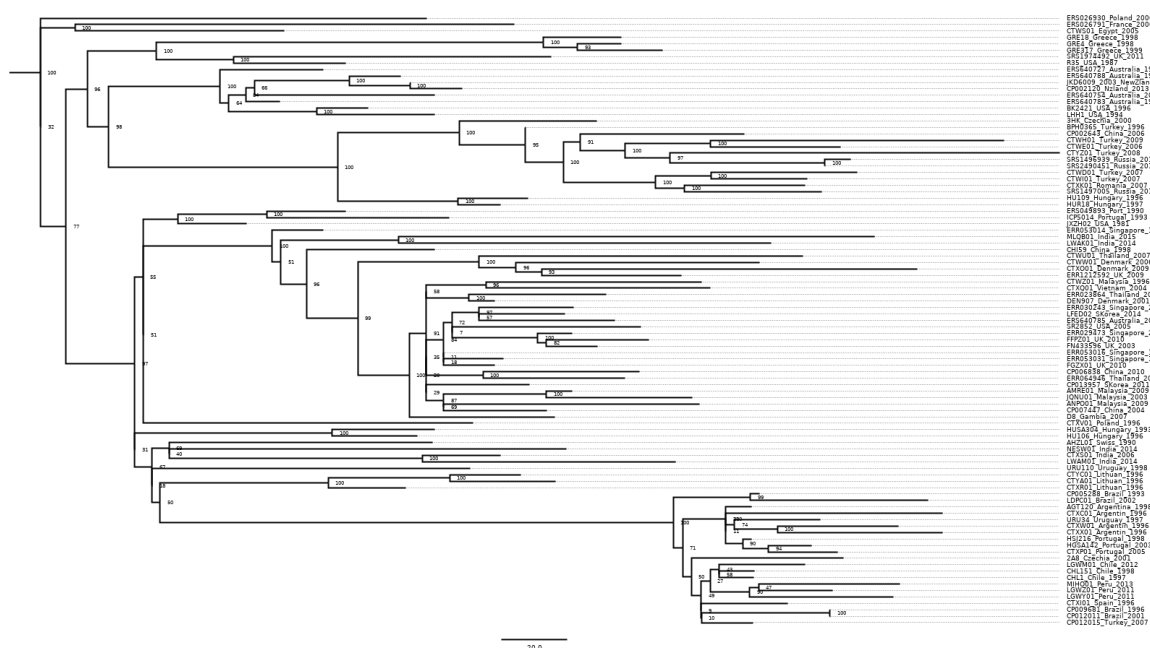


Figure 3.8. Outgroup rooted maximum-likelihood phylogeny of a) the ST8-like region of ST239, and b) the ST30-like region of ST239. Units in SNPs/Mb.

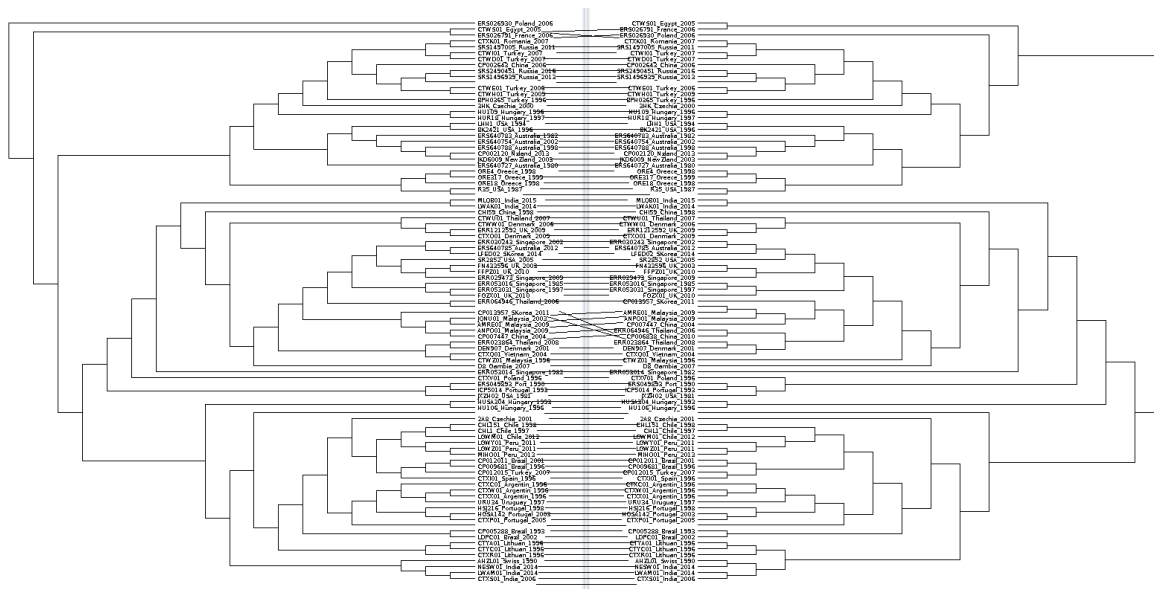


Figure 3.9. Dendroscope tanglegram of the ST30-like region phylogeny (left) and the ST8-like region phylogeny (right) of the ST239 sequences. Sequences from the same isolate are connected with a light grey line.

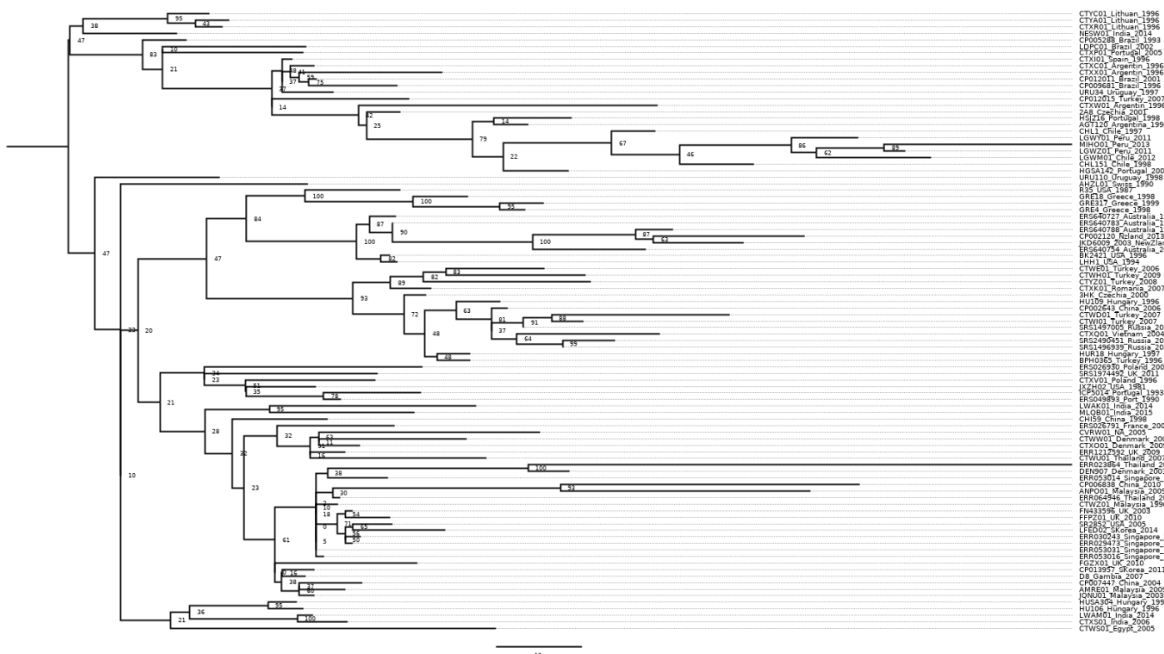


Figure 3.10. Midpoint rooted maximum-likelihood phylogeny of the SCCmec region of ST239. Units in total number of SNPs.

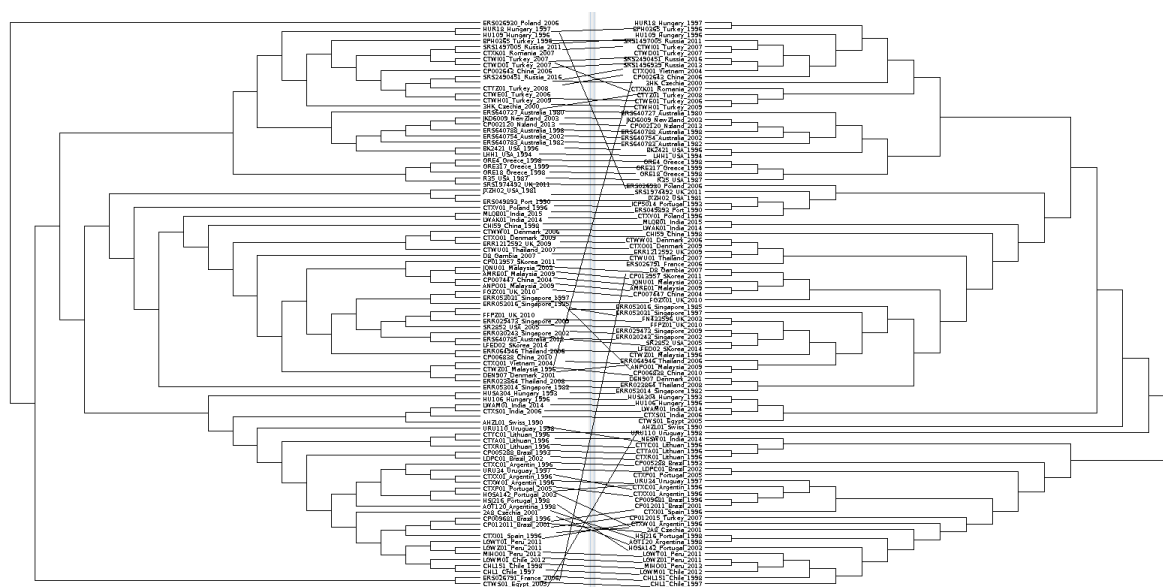


Figure 3.11. Dendroscope tanglegram comparing the ST30-like region phylogeny (left) with the SCCmec-III region phylogeny (right) of the ST239 sequences. Sequences from the same isolate are connected with a light grey line.

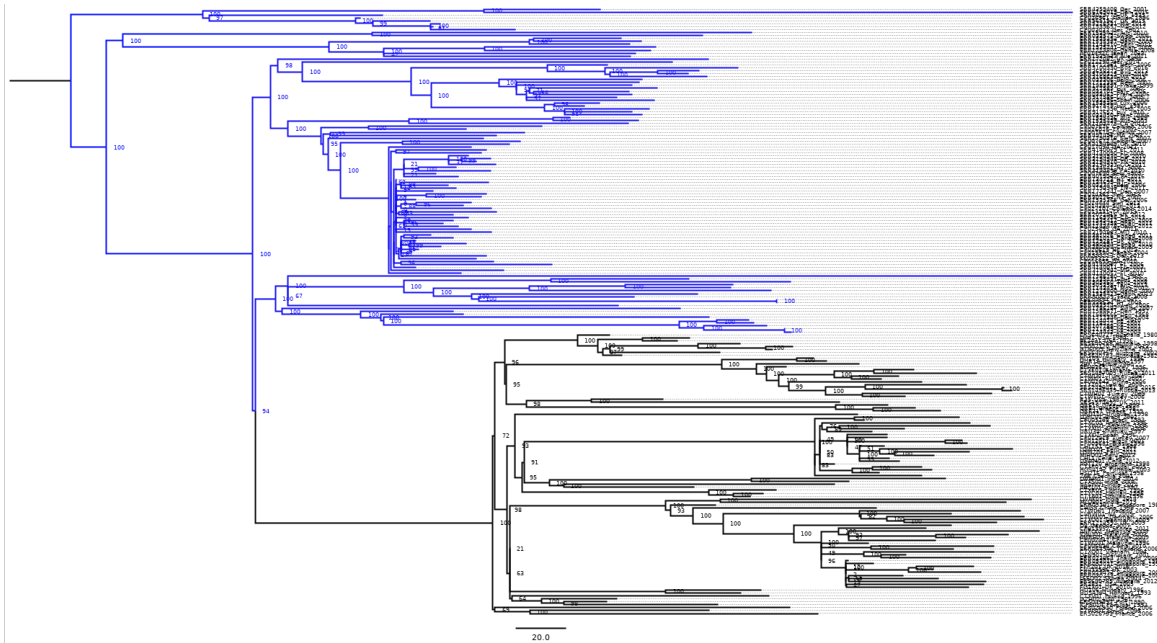
3.4 Identifying the most likely ST239 donor

3.4.1 Genetic distance between the homologous regions of ST239, ST8 and ST30

In order to determine how closely related the ST239 sequences are to the homologous regions in the ST8 and ST30 sequences, maximum-likelihood phylogenies were estimated, combining ST239 with the hypothesised donor lineages for both genetic regions (Figure 3.12). In both phylogenies, the two STs were separated into individual clades.

The pairwise SNP distances were calculated between ST239 and ST8 or ST30 for the corresponding genetic regions (Table 3.4). The average SNP distance between the ST239 and ST8 sequences in the hypothesised ST8-donated region was 1.88×10^{-4} SNPs/site, ranging from 1.01×10^{-4} – 3.65×10^{-4} SNPs/site. The average SNP distance between the ST239 and ST30 sequences in the hypothesised ST30-donated region was 3.26×10^{-4} SNPs/site, ranging from 2.18×10^{-4} – 4.62×10^{-4} SNPs/site. This suggests that the ST8 sequences may be more

A



B

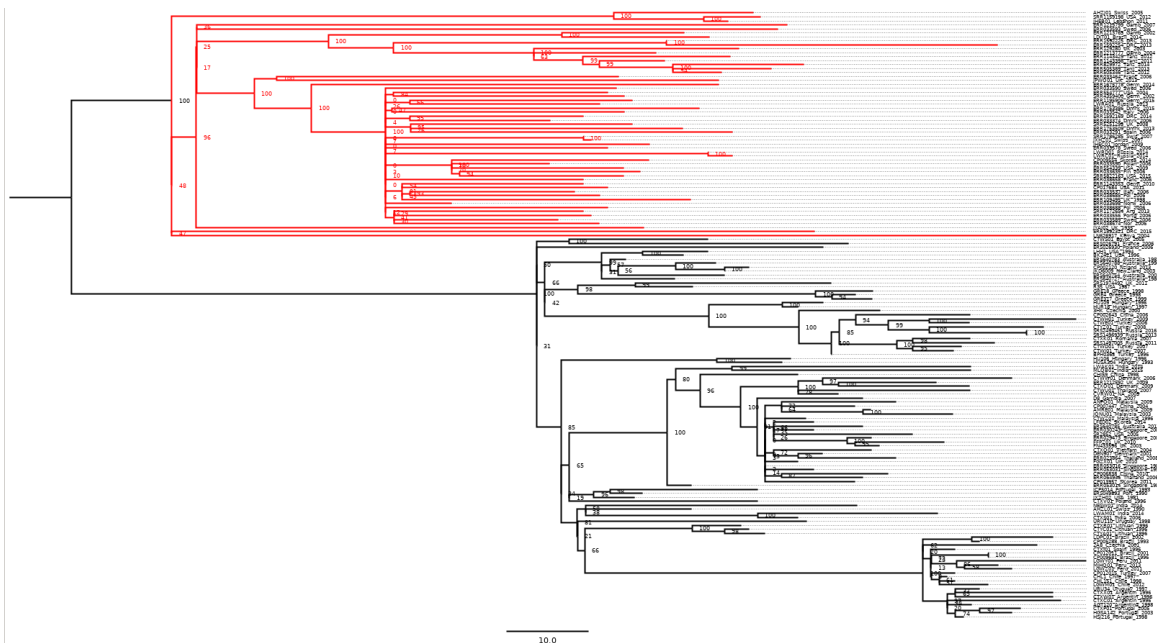


Figure 3.12. Midpoint-rooted maximum-likelihood phylogeny of a) the ST8-like region from ST239 and ST8 sequences, and b) the ST30-like region from ST239 and ST30 sequences. Blue, ST8; red, ST30; black, ST239. Units in SNPs/Mb.

Region	ST	No. sequences	Alignment length (bp)	SNP distance/site (SNPs/site) [total SNP distance (SNPs)]		
				Minimum	Maximum	Mean
ST8 region	ST8	111	2,415,589	1.24×10^{-6} [3]	3.38×10^{-4} [817]	9.02×10^{-5} [218]
	ST239	96	2,415,589	2.07×10^{-6} [5]	2.03×10^{-4} [491]	1.17×10^{-4} [282]
	ST8 and ST239	207	2,415,589	1.01×10^{-4} [244]	3.65×10^{-4} [881]	1.88×10^{-4} [455]
ST30 region	ST30	57	563,349	1.78×10^{-6} [1]	4.14×10^{-4} [233]	1.85×10^{-4} [105]
	ST239	96	563,349	7.10×10^{-6} [4]	2.47×10^{-4} [139]	1.35×10^{-4} [76]
	ST30 and ST239	153	563,349	2.18×10^{-4} [123]	4.62×10^{-4} [260]	3.26×10^{-4} [184]

Table 3.4. Pairwise SNP distances between ST239, ST8 and ST30 in the homologous regions of ST239.

closely related to the true ancestor of the homologous region of ST239, and that the ST30 sequences may be less closely related to the true donor of the acquired region of ST239.

One group of ST8 isolates was more closely related to the homologous region of ST239, which consisted of 20 diverse isolates from the USA, Sweden, Switzerland, Denmark, Ireland, Australia, Tanzania, Mozambique, the Gambia and Gabon that were isolated between 1957 – 2013. This suggests that the ancestor of the major ST239 region may have been closely related to this international ST8 strain.

3.4.2 Finding the closest known ancestor of the acquired region

In 2004, Robinson and Enright hypothesised that ST30 was the closest ancestor of the ST239 acquired-region donor. However, this conclusion was based on the partial sequencing of only 15 genes from 220 *S. aureus* isolates. In recent years, the number of publicly available *S.*

Chapter 3

aureus DNA sequences has increased massively (Chapter 2). In order to determine whether a closer relative of the ST239 acquired region is present in online sequence databases, the consensus 316 core gene sequences from the ST239 acquired region (Section 3.1.2) were queried for similar sequences in BIGSI, a DNA search engine for microbial data containing 447,833 bacterial and viral raw-read and assembled genomes from the entire bacterial and viral whole-genome sequence content of EMBL-ENA, as of December 2016 (Bradley *et al.*, 2017). The number of gene sequences that shared at least 99% k-mer identity for each gene is displayed in Figure 3.13.

A total of 39,424 unique accession numbers were identified that shared at least 99% k-mer identity with at least one core gene from the acquired region of the ST239 consensus sequence. The majority of these accession numbers ($n = 33,845$) shared this similarity with 20 genes or fewer. A total of 5,396 unique accession numbers shared at least 99% k-mer identity with over 100 of the 316 core genes (Figure 3.14), with the most similar accession number sharing 292 out of the 316 genes (92.4%).

A total of 2,792 unique accession numbers shared at least 99% k-mer identity with over 260 (>82.3%) of the ST239 acquired region core genes. These accession numbers were attributed to 20 unique *S. aureus* STs (Table 3.5), and included 2,707 ST239 sequences. The remaining 85 sequences that were not ST239 all shared six out of the seven MLST housekeeping alleles with ST239. The similarity of these sequences to the whole of the ST239 genome, including both the acquired and backbone regions, suggests that these STs are close relatives of ST239, which have diverged since the hybridisation event.

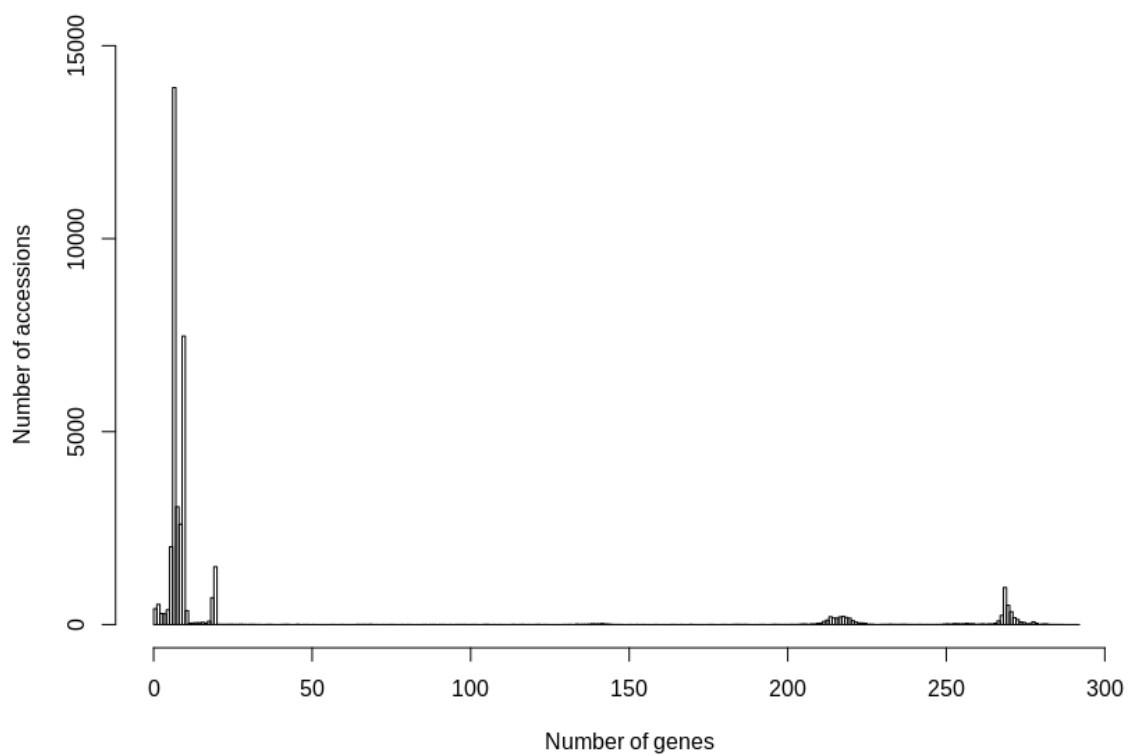


Figure 3.13. Number of sequences identified in BIGSI that share at least 99% k-mer identity with each ST239 core gene from the acquired region.

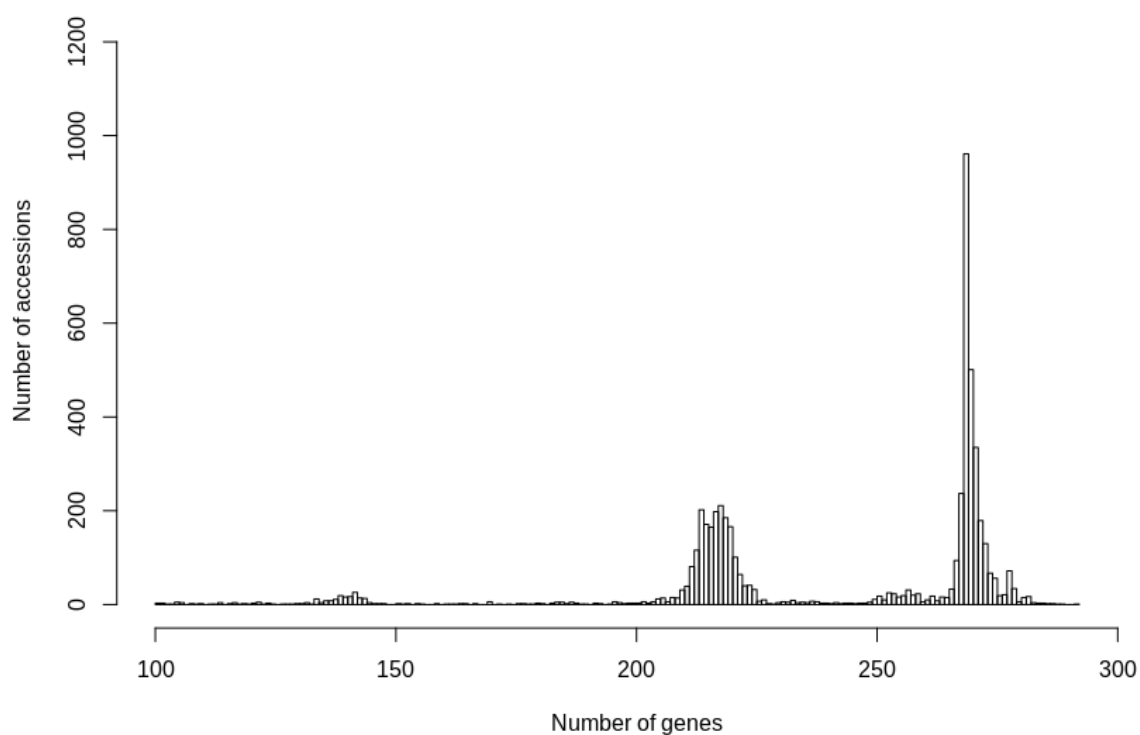


Figure 3.14. Number of sequences identified in BIGSI that share at least 99% k-mer identity with over 100 ST239 core genes from the acquired region.

ST	No. isolates	ST	No. isolates	ST	No. isolates
239	2,707	1971	2	1283	1
757	36	3349	2	1441	1
241	24	176	1	1545	1
368	3	242	1	1849	1
1282	3	247	1	3587	1
1463	2	343	1	4359	1
1531	2	372	1		

Table 3.5. ST of the isolates that share at least 99% k-mer identity with over 260 ST239 core genes from the acquired region.

A total of 2,260 unique accession numbers shared at least 99% k-mer identity with between 200 – 259 (63.3 – 82.0%) of the ST239 core genes. These were attributed to 80 unique STs (Table 3.6), including 1,233 ST30 and 575 ST36 sequences and 54 sequences where the ST could not be identified (recorded as ST0). The high number of ST30 and ST36 sequences is representative of the high number of these isolates in online sequencing databases, compared to other less well-studied STs (Petit *et al.*, 2018). Additionally, 210 ST239 sequences were also identified. These may be ST239 sequences with lower quality mapping than those previously identified, or sequences that have had more time to diverge from the consensus.

Of the 78 non-ST239 STs that were identified, six shared six MLST housekeeping alleles with ST239, suggesting that these are close relatives of ST239, which have diverged since the hybridisation event. The remaining 72 non-ST239 STs shared two or fewer MLST housekeeping alleles with ST239, suggesting that these may be close relatives of the acquired region of ST239, which diverged prior to the hybridisation event. Of these 72 STs, 68 shared five or more MLST alleles with ST30, suggesting that these may all be close relatives of ST30.

ST	No isolates	Total no. isolates in Staphopia	ST	No isolates	Total no. isolates in Staphopia	ST	No isolates	Total no. isolates in Staphopia	ST	No isolates	Total no. isolates in Staphopia
30	1233	1896	3622	2	2	1829	1	3	3012	1	1
36	575	868	3809	2	2	2047	1	1	3321	1	1
239	210	3178	31	1	1	2421	1	1	3428	1	1
39	61	136	37	1	1	2444	1	1	3430	1	1
0	54	-	174	1	2	2450	1	1	3438	1	1
241	11	56	237	1	1	2586	1	1	3488	1	1
1472	7	7	433	1	1	2858	1	1	3505	1	2
3620	7	7	579	1	1	2864	1	1	3530	1	1
2374	6	6	605	1	1	2879	1	1	3605	1	1
2994	5	9	878	1	1	2891	1	1	3638	1	1
346	4	5	1104	1	1	2896	1	2	3648	1	1
3458	4	9	1282	1	5	2897	1	1	3661	1	1
977	3	3	1329	1	2	2898	1	1	3671	1	1
1580	3	4	1390	1	1	2901	1	1	3672	1	1
2938	3	4	1456	1	3	2917	1	1	3674	1	1
3423	3	109	1482	1	1	2936	1	1	3692	1	1
3452	3	3	1530	1	2	2937	1	1	3749	1	1
243	2	18	1536	1	1	2939	1	1	3825	1	1
1163	2	9	1565	1	1	2995	1	1	3859	1	1
3437	2	3	1567	1	1	2996	1	1	3932	1	1

Table 3.6. ST of the isolates that share at least 99% k-mer identity with between 200-259 ST239 core genes from the acquired region.

A selection of 139 sequences from these 72 non-ST239-like STs were mapped to the TW20 reference genome, and the genetic regions that were homologous to the ST30-like region were extracted and aligned, alongside the 96 ST239 sequences and 57 ST30 sequences from Chapter 2. A total of 3,696 variant sites were identified, and a maximum-likelihood phylogeny was estimated and is displayed in Figure 3.15. The ST239 clade is clearly distinct from the

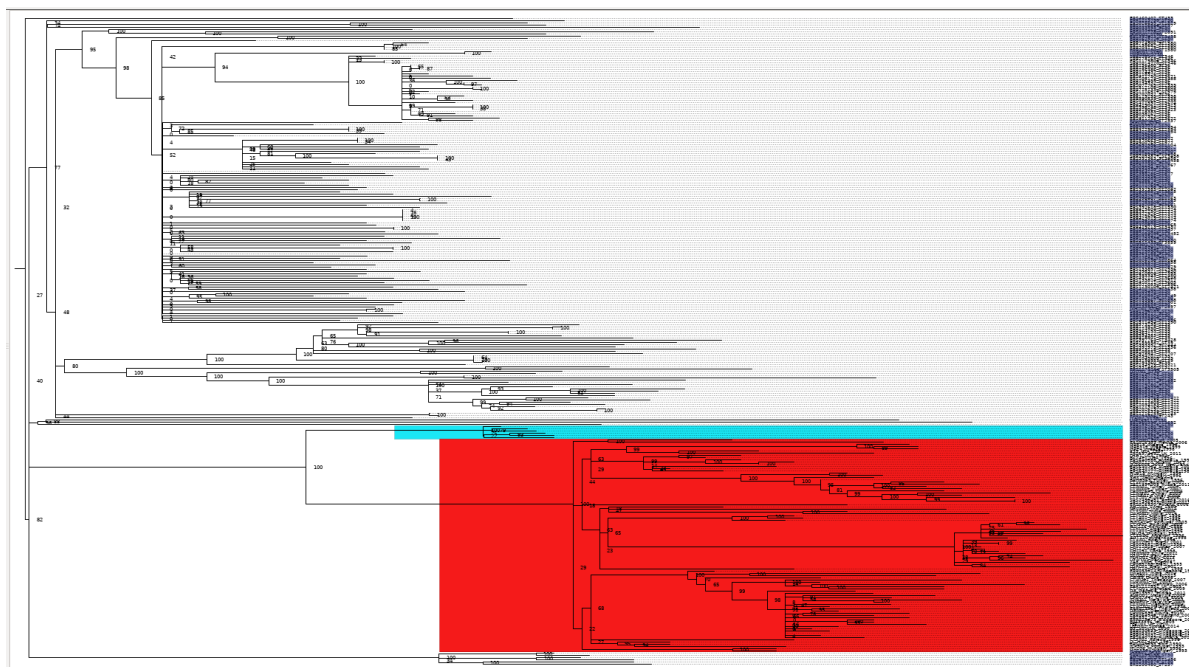


Figure 3.15. Outgroup-rooted maximum likelihood phylogeny of the acquired region of the 139 non-ST239 isolates from the BIGSI analysis, 96 ST239 and 57 ST30 sequences. The ST239 clade is highlighted in red, and the most closely related non-ST239 clade is highlighted in cyan. The accession numbers of all ST30 isolates are highlighted in blue.

other STs. The most closely related clade contains five ST30 isolates, with accession numbers SRR016400, SRR016388, SRR016398, SRR016399 and SRR016780. Although there is low bootstrap support for many of the branches, particularly within the non-ST239 clades, there is high support that these five ST30 isolates share the most recent common ancestor with the ST239 clade.

These five ST30 isolates were all sequenced as part of the same USA study by the Broad Institute in 2009 (Broad Institute, 2009). SRR016399 (55/2798) is a clinical isolate from the UK in 1955, SRR016398 (55/3224) was also isolated in the UK in 1955, SRR016780 (M1016) was isolated in Australia in 1962, SRR016388 (65-1119) was isolated in the USA 1965, and SRR016400 (68-81) was isolated from blood in the USA in 1968 (Robinson *et al.*, 2005; van Tonder *et al.*, 2014). These isolates are all from the penicillin-resistant *S. aureus*

clone known as phage type 80/81, which caused serious healthcare-associated and community-associated infections worldwide and was largely eliminated in the 1960s (Robinson *et al.*, 2005). None of these sequences contains an SCC*mec* element or *mecA* gene, and therefore it is likely that these isolates are MSSA. This suggests that the donor of the ST239 acquired region may have been closely related to a healthcare-associated MSSA ST30 of phage type 80/81.

4. Discussion

The increased availability of ST239, ST8 and ST30 *S. aureus* DNA sequences in publicly available online sequence repositories has enabled a more powerful investigation into the ST239 hybrid genome than Robinson and Enright were able to carry out in 2004. A total of 1,973 core genes that were shared between a diverse data-set of ST239, ST8 and ST30 sequences from Chapter 2, were assigned as ST8-like or ST30-like. The position along the ST239 TW20 reference genome where the genes switched from being ST8-like to ST30-like was consistent with the hybrid boundaries that were proposed by Castillo-Ramirez (Castillo-Ramirez *et al.*, 2011), and were consistent with the single large chromosomal replacement event hypothesised by Robinson and Enright (Robinson and Enright, 2004). It is likely that similar hybridisation events in bacteria are common, but not frequently observed; only a few hybrid lineages emerge successfully as “hopeful monsters” (Croucher and Klugman, 2014).

Four potential regions of recombination were identified in the 96 diverse ST239 genomes from Chapter 2. These included two phage regions, a transposon, and the SCCmec-III element. Castillo-Ramirez (2011) also suggested that only very low levels of recombination had occurred within the core genes shared by the 63 ST239 isolates from the Harris (2010) study (Castillo-Ramirez *et al.*, 2011). This is unlike some other pathogenic hybrid species, which have formed through several successive recombination events, or through a mosaic of large chromosomal fragments from several different donors (Sheppard *et al.*, 2013; Brochet *et al.*, 2008).

Evidence of recombination in the *SCCmec*-III region, as well as a poorly supported maximum-likelihood phylogeny, suggests that *SCCmec*-III may not have evolved entirely through vertical inheritance alongside the rest of the ST239 genome. This suggests mobility, either of the entire *SCCmec*-III element, or of smaller genetic regions within the *SCCmec*-III element. This is consistent with the high variability of *SCCmec* elements (IWG-SCC, 2009), which have been shown to have high levels of laterally transferable genes (John, 2019). There are at least 38 different sub-types of *SCCmec*-III, due to the loss and acquisition of accessory, purposeless or redundant genes (Monecke *et al.*, 2018). Even during the short period of time that ST239 has been evolving, the *SCCmec*-III element has undergone such a large amount of recombination that it is difficult to study using conventional evolutionary methods.

Robinson and Enright were able to identify ST30 as the most closely related ST to the smaller hybrid region of the ST239 genome in 2004, from the partial sequences of a selection of 220 MRSA isolates (Robinson and Enright, 2004). Using the BIGSI DNA search engine, it was possible to query nearly half a million viral and bacterial genome sequences, to identify the most closely related sequences to the acquired region of ST239, in order to find the closest known ancestor. There were a number of potential candidates, which were primarily ST30-like. The isolates that were identified as sharing the closest MRCA with the ST239 acquired region was a group of five ST30 isolates that were isolated between 1955 – 1968 from the UK, USA and Australia. These were all phage type 80/81 and were all absent of *SCCmec*, which suggests that the ST239 donor may have been related to a healthcare-associated MSSA ST30 from this historic lineage.

Chapter 3

In conclusion, this chapter has provided strong support for the Robinson-Enright model of ST239 and identified the closest known ancestors of the donor lineage. The long-term evolution of the ST239 hybrid genome has not been driven primarily by further large structural changes to its genome, but has more likely been driven by smaller genetic changes, which will be further investigated in Chapter 4.

Chapter 4

Investigating evolution in the ST239 genome

Abstract

ST239 is a recently emerged healthcare-associated EMRSA. Its hybrid origin provides a unique opportunity to investigate how the genetic regions from two divergent parent lineages have adapted to evolutionary pressures in the same genome. In this chapter, the time to MRCA of a diverse data-set of 96 ST239 sequences was estimated as between 1934.7 – 1943.3, which is earlier than previous estimates. This is prior to the clinical introduction of methicillin, and suggests that the use of earlier antimicrobials, such as penicillin, tetracycline and streptomycin, are more likely to have driven the emergence of ST239. Dating of the *SCCmec*-III element, which is associated with resistance to these antimicrobials in addition to methicillin resistance, suggested that *SCCmec*-III was acquired at a similar time as ST239 MRCA. Evidence of weaker selective constraints on the acquired region of the ST239 genome, compared to the backbone region, was observed through the increased rate of substitution in this region and evidence of weaker purifying selection, which may explain why the large genetic transfer event involving this region was possible during the formation of ST239. Parallel evolution in multiple genes that have functions relating to AMR and virulence suggests that these are some of the main evolutionary pressures that have been faced by ST239. These findings provide some insight into the evolutionary drivers of the global emergence of ST239, through adaptation to evolutionary pressures, such as the use of antimicrobials and host immune response, and weaker selective constraints to the acquired genetic region.

1. Introduction

1.1 Aims and objectives

In this chapter, I aim to estimate the time to MRCA of the diverse data-set of ST239 sequences from Chapter 2. I also aim to estimate whether the acquisition of the *SCCmec-III* element that is closely linked to the ST239 genome occurred prior to, or subsequently to, the hybridisation event. Finally, I aim to investigate the rate of evolution of the different genetic regions of ST239, and the contribution of selection to the evolution of the ST239 genome.

1.2 The emergence of ST239

Since the first ST239 whole-genome sequence was published in 2010 (Holden *et al.*, 2010), several studies have estimated a range of different times to MRCA of ST239, using a variety of different approaches. These estimates have ranged from 1965.7 – 1974.4, to as broad as 1885 – 1973, and primarily focus on isolates from a single study by Harris in 2010 (Harris *et al.*, 2010).

The first ST239 phylogeny was estimated by Harris in 2010 (Figure 4.1), and included 43 ST239 isolates from a global collection isolated between 1982 – 2003, and an additional 20 isolates that were collected over a 7-month period from a hospital in Thailand between 2006 – 2007. A maximum-likelihood phylogeny was estimated from the 4,310 SNPs in the 2,597,610 bp core genome, and a linear regression estimate of time to MRCA was calculated as mid-to-late 1960s (Harris *et al.*, 2010).

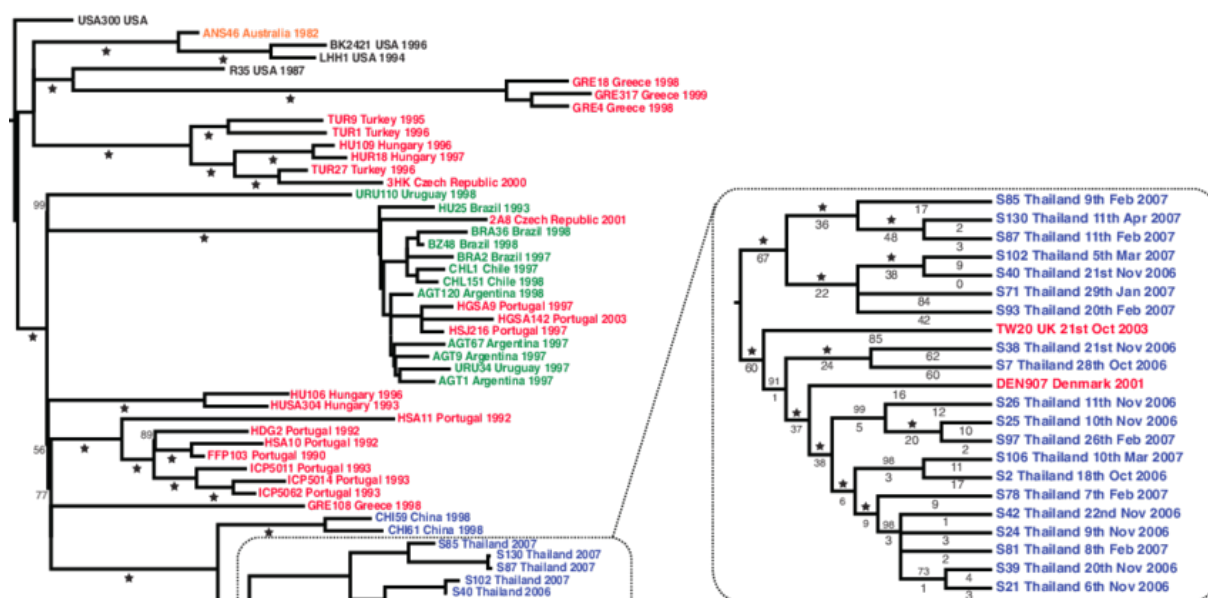


Figure 4.1. ST239 phylogeny estimated by Harris (2010). Image from Harris *et al.*, 2010.

Following the Harris study, a Bayesian molecular clock analysis by Gray (2010) (Figure 4.2) used the same 4,310 SNPs from the 63 ST239 isolates sequenced by Harris (2010), to estimate a time to MRCA ranging from 1957 – 1972, with largely overlapping 95% HPD intervals from similar models (Gray *et al.*, 2010).

Using an alternative collection of 111 diverse ST239 isolates from 29 countries, isolated over 34 years, Smyth (2010) constructed a most parsimonious haplotype tree (Figure 4.3) from 43 variable nucleotide sites. From this, Smyth was able to date the time to MRCA of the ST239 sequences as mid-20th Century, which was a similar time to estimates from the Harris collection (Smyth *et al.*, 2010).

In 2012, Castillo-Ramirez supplemented the 63 Harris (2010) ST239 genomes with an additional 31 ST239 sequences from a global collection and 71 ST239 sequences from three cities in Turkey (Castillo-Ramirez, *et al.*, 2012). The authors estimated an emergence date of

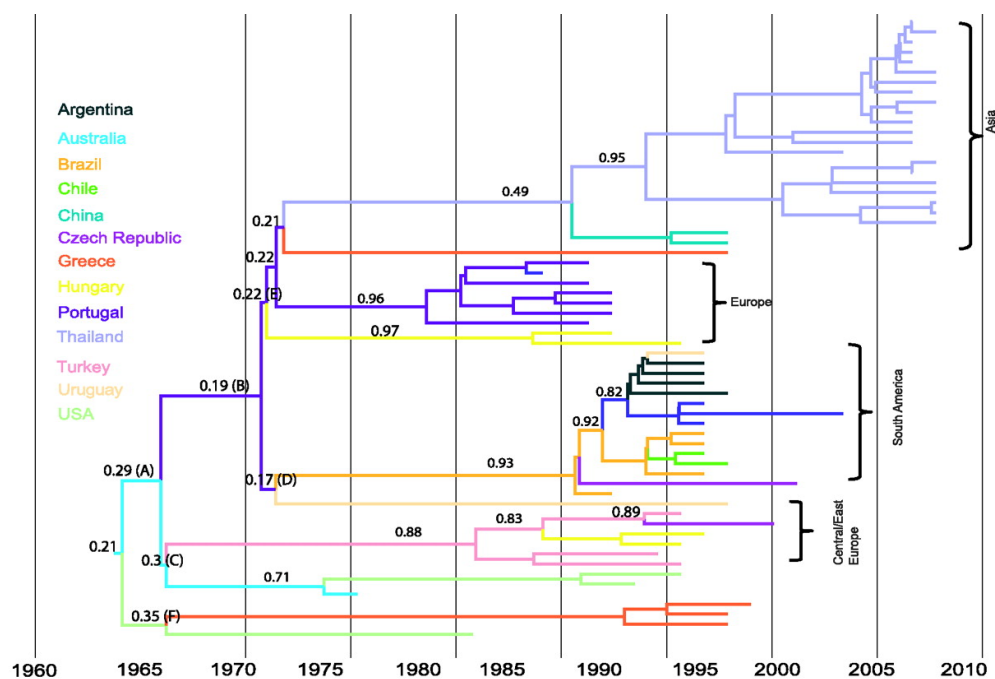


Figure 4.2. ST239 phylogeny estimated by Gray (2010). Image from Gray *et al.*, 2010.

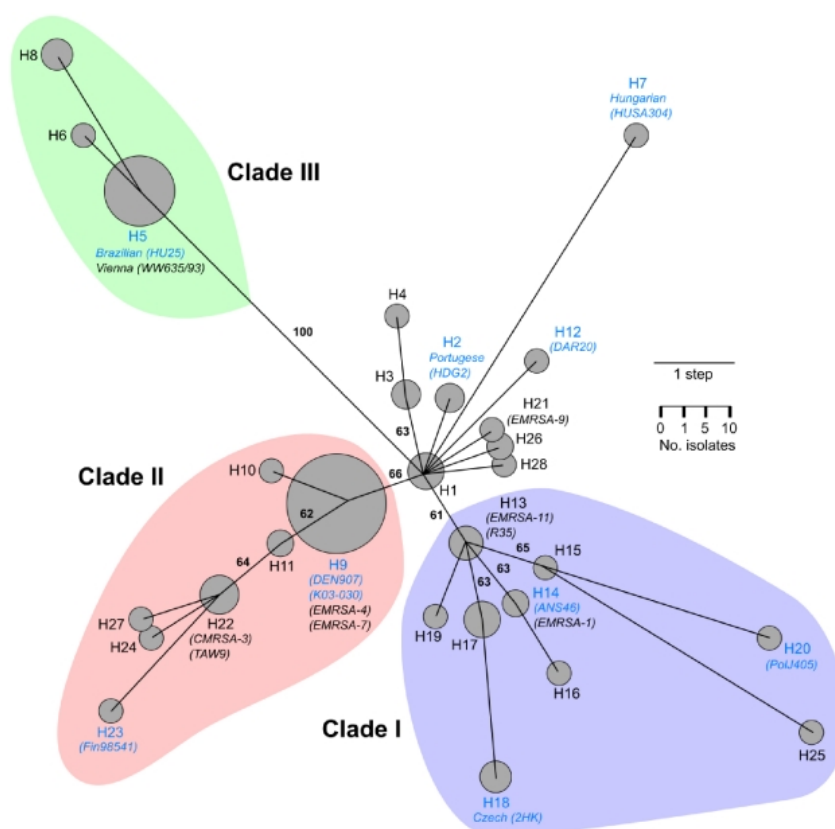


Figure 4.3. ST239 haplotype tree estimated by Smyth (2010). Image from Smyth *et al.*, 2010.

ST239 between late 1960 – early 1973, which is consistent with the estimate by Harris (2010). Due to the high diversity of European isolates and the prevalence of European isolates in basal groups of the phylogeny, Castillo-Ramirez estimated a European origin of ST239.

Baines (2015) identified two separate introduction events of ST239 MRSA into Australia. In this study, Baines estimated the time to MRCA of 87 ST239 isolates that were collected from healthcare institutions in Australia and New Zealand between 1980 – 2012, and 35 of the diverse international ST239 isolates sequenced by Harris (2010). To estimate the time to MRCA, Baines used BEAST (Drummond *et al.*, 2012), a software package for Bayesian phylogeny inference that uses MCMC algorithms for evolutionary analyses, to construct a maximum clade credibility tree (Figure 4.4) from 2,716 core genome SNPs, which represented a decreased diversity compared to the Harris (2010) study. The time to MRCA was estimated with 95% HPD intervals as 1885 – 1973, and the 95% HPD interval rate of nucleotide substitutions was 1.2×10^{-6} – 2.0×10^{-6} substitutions per site per year (Baines *et al.*, 2015).

In 2018, Didelot developed the software package BactDating, a fast Bayesian methodology to construct dated phylogenies from bacterial genomes. To demonstrate the application of BactDating, the time to MRCA of the 63 ST239 sequences from the Harris collection was estimated (Didelot, 2019). The BactDating estimate of the time to MRCA was 1965.7 – 1974.4, which was also consistent with similar estimates of the same data-set by Gray (2010) and Harris (2010).

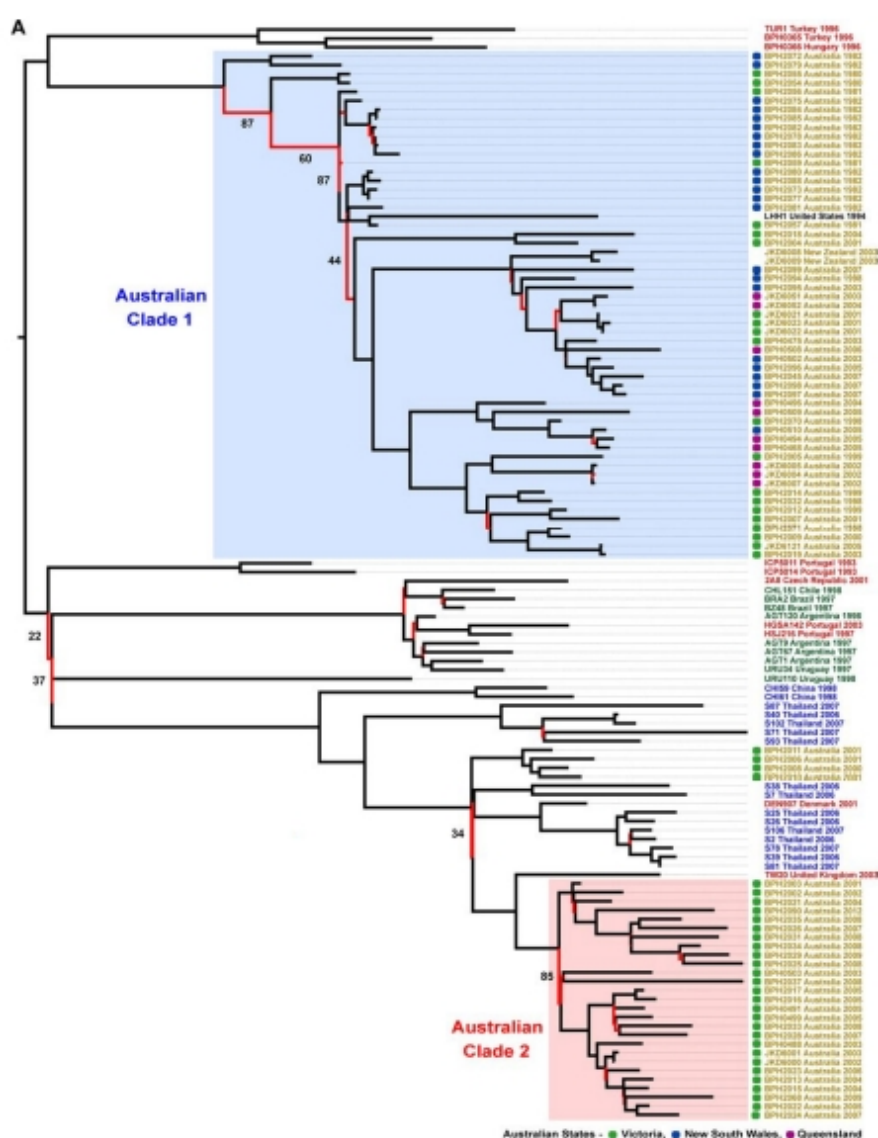


Figure 4.4. ST239 phylogeny estimated by Baines (2015). Image from Baines *et al.*, 2015.

1.3 The origin of the SCC_{mec} element

In 2000, the first *SCCmec* element, *SCCmec-I*, was identified (Katayama *et al.*, 2000) from the MRSA strain N315, which was isolated in Japan in the early 1980s (Okonogi *et al.*, 1989). Harkins (2017) attempted to reconstruct the evolutionary history of *SCCmec-I* elements from 122 early MRSA isolates that were linked with isolation date and location metadata. The study used a linear regression analysis based on 39 core variant sites with a maximum SNP

distance of eight, to estimate the MRCA of the SCCmec-I element as 1941, which is 14 years prior to the first therapeutic use of methicillin. Harkins suggest that the initial spread of the SCCmec-I element in *S. aureus* was therefore more likely to have been driven by the use of earlier antimicrobials. Harkins also suggested that the SCCmec element most likely originated in coagulase-negative *Staphylococci*, such as *S. sciuri*, before it was taken up by *S. aureus* (Harkins *et al.*, 2017; Couto *et al.*, 1996). In an analysis of 152 *S. aureus* SCCmec elements, John (2019) showed that the structure of the SCCmec phylogeny was significantly different than the *S. aureus* core genome phylogeny, and suggested that resistance genes within the SCCmec element were likely laterally transferred from non-*Staphylococcus* species (John *et al.*, 2019).

SCCmec-III was identified in 2001, from MRSA strain 85/2082, which was isolated in New Zealand in 1985 (Ito *et al.*, 2001). Rolo (2017) identified an SCCmec-III-like ancestor in *S. sciuri*. Rolo hypothesised that an ancestor of the ST239 SCCmec-III originated in *S. sciuri* through a recombination event in a human host or a human-created environment, and was subsequently transferred to *S. aureus* (Rolo *et al.*, 2017).

1.4 Selection pressures of HGT events

Compensatory adaptation and positive evolution play a significant role in the success of HGT events. For example, positive selection acts to promote horizontally-transferred traits, and compensatory adaptation acts to eliminate the cost of the horizontally-transferred element. These two evolutionary forces therefore act to reinforce each other's effects (San Millan *et al.*, 2014). Compensatory adaptation can occur in both the chromosome and in the horizontally acquired region. For example, to stabilise a plasmid in a bacterial population, coevolution

occurs between both the bacterial chromosome and the plasmid (Harrison and Brockhurst, 2012).

Castillo-Ramirez (2011) suggested that the ST239 population may still be in the initial stages of diversification, due to the recent hybridisation event, which may result in high levels of selection to remove deleterious alleles from the ST239 genome (Castillo-Ramirez, 2011). By analysing the acquired and backbone regions of the 63 ST239 isolates that were sequenced by Harris (2010), Castillo-Ramirez found evidence of a high level of non-synonymous mutations in both the acquired and backbone regions, which is indicative of strong selection pressures acting on the regions that originated from each genetic background (Castillo-Ramirez *et al.*, 2011). It is possible that different selection pressures may act on the different genetic regions of ST239 as they adapt to the same genome. Smyth (2010) estimated a similar rate of evolution for both of the ST239 genetic regions, however, this was based on a linear regression analysis from only 43 variable nucleotide sites (Smyth *et al.*, 2010).

1.4.1 Positive selection and parallelism

One way to identify genes that are under positive selection is to test for parallel evolution, also known as convergent evolution. Parallel evolution results in similar genetic changes in independent species (Stern, 2013). Experimental evolution, genetic studies and genetic-association studies have provided evidence for the significance of parallelism in the evolution of single-celled and multi-cellular organisms through natural selection, as it is unlikely that a large number of genetic similarities would arise through purely random chance (Stern, 2013).

Parallel evolution is an example of the repeatability of evolutionary change. For example, in twelve identical populations of *E. coli* that were evolved in identical environments for 12,000 generations, there was evidence of extensive parallelism at the level of genes, although only rarely at the nucleotide level (Woods *et al.*, 2006).

1.4.2 The McDonald-Kreitman test for selection

The McDonald-Kreitman test is a simple statistical test to infer how selection contributes to patterns of molecular evolution, by determining whether the difference in the amino-acid sequences across species has evolved due to the accumulation of neutral mutations by random drift, due to the fixation of adaptive mutations by selection, or a mixture of the two (McDonald and Kreitman, 1991). McDonald and Kreitman suggested that if the ratio of non-synonymous to synonymous differences that are fixed between species is the same as the ratio of non-synonymous to synonymous polymorphisms within a species, then this is due to the accumulation of neutral mutations. Alternatively, if there are more fixed replacement differences between species, then this is likely due to positive selection through the adaptive fixation of selectively advantageous mutations. The McDonald-Kreitman net neutrality index (Figure 4.5) is used to calculate the ratio between these values (Rand and Kann, 1996). If $N = 1$, this suggests that evolution has occurred due to the accumulation of neutral mutations by random drift. Values where $N > 1$ are indicative of negative selection because natural selection is favoring the purifying selection, and the weeding out of deleterious alleles (i.e. levels of divergence between species at non-synonymous sites are low, given the relative abundance of non-synonymous polymorphisms). Values where $N < 1$ are indicative of positive selection, because natural selection favors a specific phenotype over other phenotypes, and the favored

$$N = \frac{P_n/P_s}{D_n/D_s}$$

Figure 4.5. Equation for the McDonald-Kreitman net neutrality index. Net neutrality index (N); number of non-synonymous polymorphisms (P_n); number of synonymous polymorphisms (P_s); number of non-synonymous substitutions (D_n); number of synonymous substitutions (D_s).

phenotype begins to go to fixation in the species as the allele frequency for that phenotype increases.

1.5 Chapter summary

In this chapter, two complementary bayesian phylogeny inference methods were used to estimate the time to MRCA of the 96 diverse ST239 sequences from Chapter 2; BEAST and BactDating. Both methods had similar accuracy, and estimated a similar time to MRCA within 1.5 years, ranging from 1934.7 – 1943.3. This estimate is much earlier than previous estimates of the ST239 time to MRCA, which primarily used a less diverse collection of isolates with a smaller number of variant sites. This suggests that the formation and subsequent global emergence of the ST239 lineage was not driven by the clinical introduction of methicillin in 1959, but was more likely driven by the earlier clinical use of antibiotics such as tetracycline and streptomycin, resistance determinants of which are also common to the SCC*mec*-III element that is highly prevalent in ST239. Although there were larger confidence intervals, the time to MRCA of the SCC*mec*-III element from the 96 ST239 sequences was dated to a similar time as the MRCA of the ST239 whole genome sequences. This suggests that the acquisition of the SCC*mec*-III element, and the hybridisation event that formed ST239, may have occurred at a similar time.

Chapter 4

To investigate the evolutionary consequences of the hybridisation event that formed ST239, the molecular evolution of ST239 was investigated. The substitution rate of the acquired region was greater than the substitution rate of the backbone region, which may be evidence of stronger positive selection or weaker selective constraints in the acquired region. Overall, purifying selection was determined to be the dominant selection pressure in both genetic regions of the ST239 genome. However, there was weaker evidence of purifying selection in the acquired region. Several highly variable genes in both the acquired and backbone regions displayed evidence of positive selection. The functions of many of these genes were related to AMR and adaptation to host immunity, suggesting that these are some of the major evolutionary pressures faced by ST239. Collectively, these results suggest that the acquired region has undergone accelerated evolution in the ST239 genome, in part due to selection pressures related to AMR and virulence. Additionally, the weaker selective constraints that act on the acquired region may explain why large genetic transfer events involving this region have been observed in ST239 and in other *S. aureus* hybrid genomes.

2. Methods

2.1 Linear regression and BactDating

Linear regression was carried out on the phylogenies that were estimated for the 96 ST239 whole-genome sequences, backbone region sequences, acquired region sequences, and the SCC*mec*-III elements, from Chapter 3, using the BactDating R package (Didelot *et al.*, 2018) in R v3.4.4. Mixed gamma, strict gamma and relaxed gamma evolutionary models were run for 10,000,000 MCMC to estimate a time to MRCA of the ST239 whole-genome sequences. All ESS values were over 100. All graphs were plotted in R v3.4.4.

2.2 BEAST

A total of 6,819 variant sites were extracted from the alignment of the 96 ST239 whole-genome sequences from Chapter 2 Section 2.4, using snp-sites v2.3.3 (Page *et al.*, 2016). BEAST analysis was carried out using BEAST v1.10.4 (Drummond *et al.*, 2012), using the GTR model. The XML file was edited to reflect the number of unchanging sites in the original alignments. All combinations of Strict, Relaxed, Constant and Exponential molecular clock models were run. For each model, three 300,000,000 MCMC chains were run and combined, with path sampling/stepping-stone sampling every 100 steps for a chain length of 3,000,000. In all cases, the Bayes Factor showed no significant difference in the likelihood of the different models. The burnin was set at 10%, and runs were combined using LogCombiner, with a re-sample size of 10,000. TreeAnnotator was used to estimate the maximum clade credibility tree, which was displayed in figtree v1.4.4 (Rambaut, 2010). The MRCA and evolutionary rates were estimated with 95% HPD intervals. All graphs were plotted in R v3.4.4.

This was repeated for 5,353 variant sites from the 96 ST239 backbone region sequences, 1,449 variant sites from the ST239 acquired region sequences, and 54 variant sites from the ST239 SCC*mec*-III elements, from Chapter 3 Section 2.3.

2.3 McDonald-Kreitman test

An alignment of the 1,980 core genes from all 96 ST239 sequences was created from Chapter 2 Section 2.4, and each sequence was also converted into an amino acid sequence. A consensus sequence was created from the nucleotide alignment using EMBOSS v6.6.0.0 (EMBOSS, 2000), and this consensus sequence was also converted into an amino acid sequence. The location of variable sites within the 96 ST239 core gene and amino acid sequences that differed from the DNA and amino acid consensus alignments were extracted using SNP-sites (Page *et al.*, 2016). The number of fixed synonymous and fixed non-synonymous SNPs was calculated for core genes within the whole ST239 genome, core genes within the backbone region, and core genes within the acquired region.

This analysis was repeated for the alignment of the 1,980 core genes from all 111 ST8 sequences, and for the alignment of the 1,980 core genes from all 57 ST30 sequences, from Chapter 2 Section 2.3, for genes within the whole genome, and genes that were homologous to the acquired region and backbone region.

The McDonald-Kreitman net neutrality index was calculated as in Figure 4.5 Section 1.4.2.

2.4 Parallelism in ST239

The predicted Poisson distribution of substitutions within the acquired and backbone regions of the 96 ST239 sequences was calculated in R v3.4.4. The true and predicted distributions of substitutions in each gene in each genetic region were plotted in R v3.4.4. Two-sample Kolmogorov-Smirnov and Chi-squared tests were carried out in R v3.4.4. Gene names and functions were identified in AureoWiki (Fuchs *et al.*, 2018). A literature search of the non-synonymous mutations in the *walk* and *rpoB* genes was carried out. The number of synonymous and non-synonymous mutations in genes that were classed as highly variable were also calculated from the homologous genes in the ST30 and ST8 sequences, as in Section 2.3.

2.5 Tajima's D

The backbone and acquired regions (excluding the SCC*mec*-III element) were extracted from the 96 ST239 sequences that were aligned to the FN433596 ST239 reference genome in Chapter 2, Section 2.4, using the hybrid boundaries identified in Chapter 3, Section 2.1.3. Recombination was masked using Gubbins v1.4.5 (Croucher *et al.*, 2015), and a consensus sequence was generated using EMBOSS v6.6.0.0 (EMBOSS, 2000). All variant sites were extracted using snp-sites v2.3.3, and a vcf file was generated (Page *et al.*, 2016). Tajima's D was calculated using the StrataG R package (Archer and Schneiders, 2016) in R v3.6.3.

2.6 Purine-pyrimidine and transition-transversion substitution proportions

All variants from the consensus sequence of the 96 ST239 sequence alignment in Section 2.5 were assigned as either a transition (A → G; C → T; G → A; T → C) or a transversion (A →

Chapter 4

C; A → T; C → A; C → G; G → C; G → T; T → A; T → G). Chi-squared tests were carried out in R v3.4.4.

2.7 Intergenic regions

From the 96 ST239 sequences that were mapped to the FN433596 reference sequence in Chapter 2, Section 2.4, the number of core intergenic region variant sites that were shared between all 96 sequences was calculated using PIGGY v1.5 (Thorpe, *et al.*, 2018), and all variant sites were extracted using snp-sites v2.3.3 (Page *et al.*, 2016).

3. Results

3.1 Dating the time to MRCA of ST239

Linear regression, and two bayesian phylogeny inference methods (BEAST and BactDating), were used to estimate the time to MRCA of the 96 temporally and geographically diverse ST239 isolates from Chapter 2.

3.1.1 Linear regression and BactDating

Initially, a linear regression was used to determine whether there was any temporal signal in the ST239 data-set, and provide a rough estimate of the time to MRCA (Figure 4.6). There was significant temporal signal in the data ($p < 0.0001$), and the root-to-tip rate was estimated as 2.46×10^{-6} SNPs/site/year (95% CI $2.29 \times 10^{-6} - 2.64 \times 10^{-6}$ SNPs/site/year) with a time to MRCA of 1950.9.

Three evolutionary models were run with the BactDating R package (Didelot *et al.*, 2018) to estimate a more probable time to MRCA. The estimated dates of the three models converged between of 1936.2 – 1943.0 (Table 4.1). This is prior to the estimated time to MRCA using linear regression, and earlier than most previous estimates of the ST239 MRCA (Harris *et al.*, 2010; Gray *et al.*, 2010; Smyth *et al.*, 2010; Didelot, 2019).

Estimated time to MRCA	BactDating evolutionary model (95% HPD)			Consensus
	Mixed gamma	Relaxed gamma	Strict gamma	
	1906.0 – 1949.8	1903.1 – 1947.3	1936.2 – 1943.0	1936.2 – 1943.0

Table 4.1. ST239 BactDating estimates of time to MRCA.

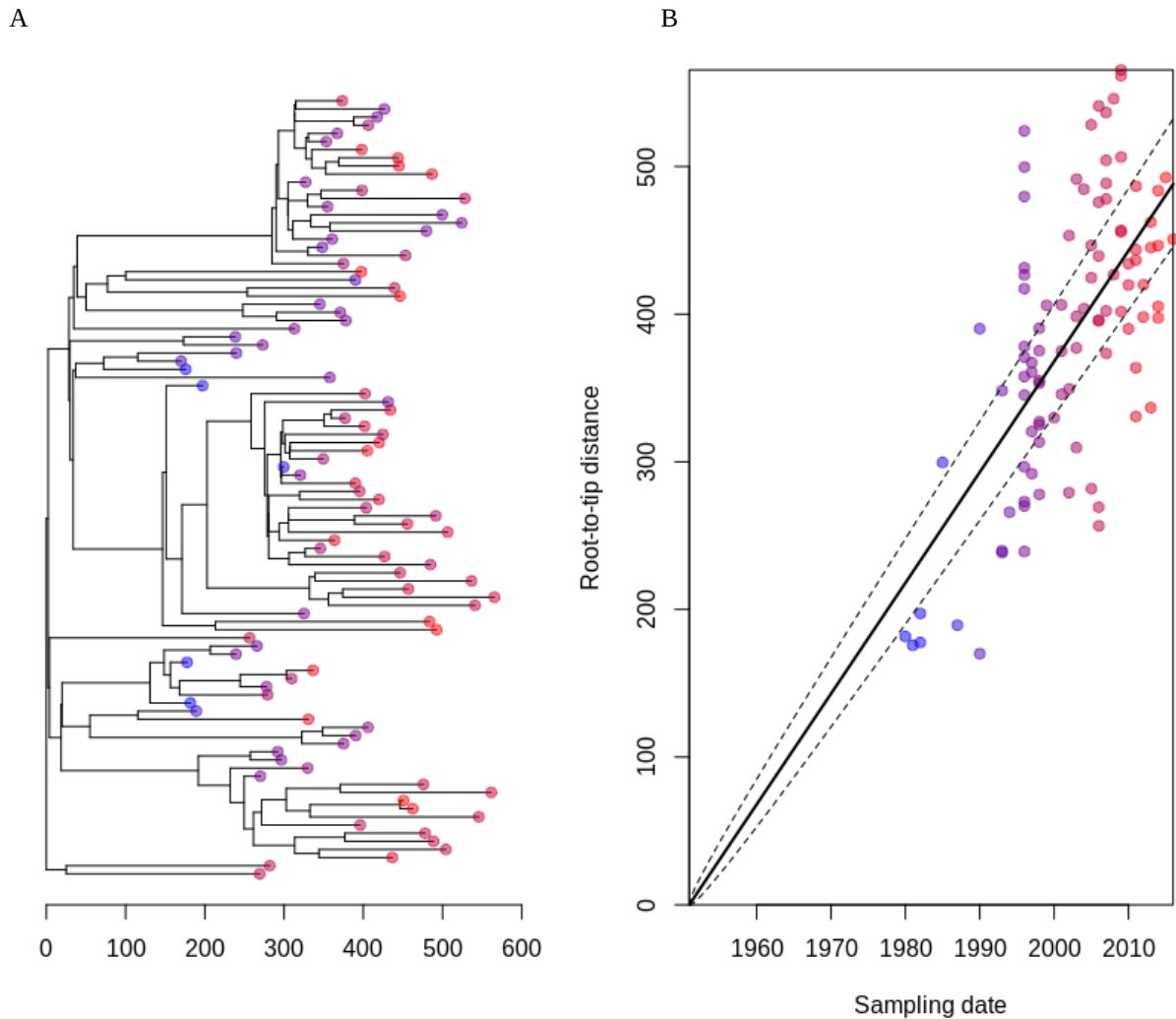


Figure 4.6. A) Maximum-likelihood phylogeny of ST239 (units in total number of SNPs). B) Root-to-tip distance plot against date of isolation (years). The solid black line is the linear regression line of best fit, and the dotted black lines represent the 95% CI.

3.1.2 BEAST

Four evolutionary models were run using BEAST (Drummond *et al.*, 2012) to estimate a time to MRCA of the ST239 sequences (Table 4.2, Figure 4.7), which converged between 1934.7 – 1943.3 (95% HPD), which was consistent with the time to MRCA estimated using BactDating.

	BEAST evolutionary model (95% HPD)				Consensus
Substitution rate (95% HPD) ($\times 10^{-6}$ SNPs/site/ year)	Strict constant	Strict exponential	Relaxed constant	Relaxed exponential	
	1.20 (1.13 - 1.28)	1.20 (1.13 - 1.28)	1.25 (0.99 - 1.53)	1.29 (1.05 - 1.55)	1.13 – 1.28
MRCA	1940.1 (1934.7 - 1945.4)	1940.1 (1934.6 - 1945.3)	1929.1 (1899.3 - 1953.0)	1946.5 (1931.5 - 1959.9)	1934.7 - 1943.3

Table 4.2. ST239 BEAST estimates of time to MRCA.

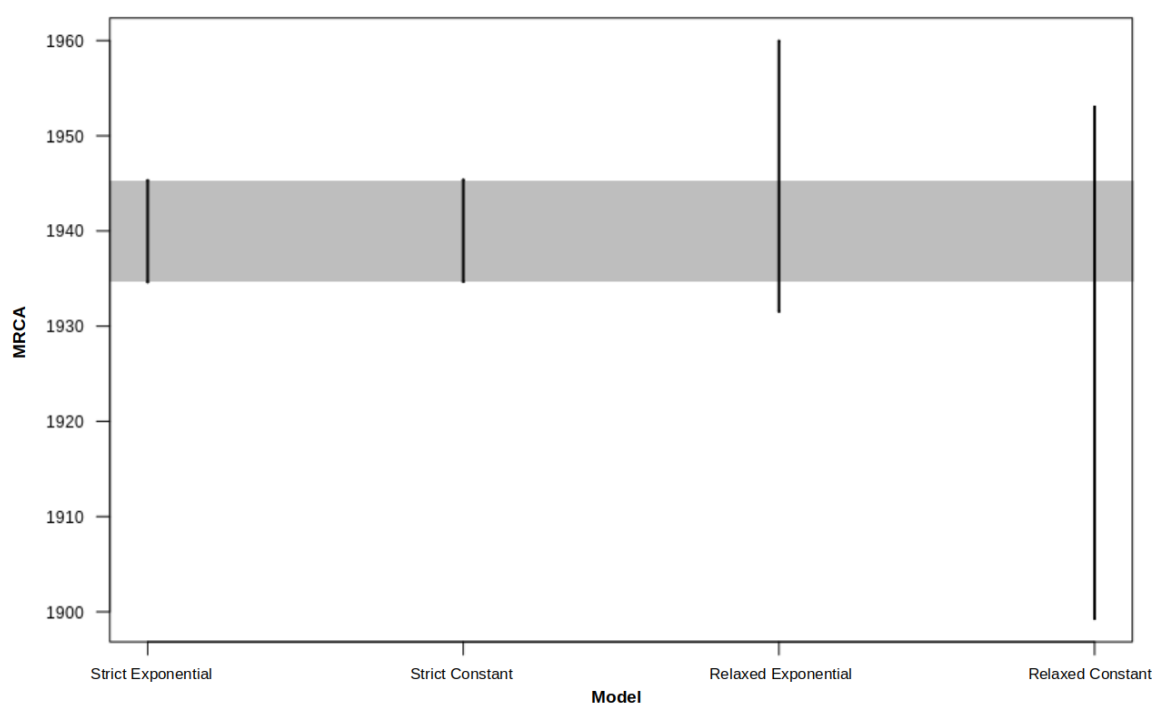


Figure 4.7. BEAST estimates of time to MRCA of ST239. The 95% HPD interval of each model is displayed by vertical black bars, and the consensus (where the MRCA of all models overlaps) is in grey.

The consensus estimated substitution rate was $1.13 \times 10^{-6} - 1.28 \times 10^{-6}$ SNPs/site/year (95% HPD) (Table 4.2). This is on the slower side of the ST239 substitution rate estimated by Harris (2010) of $1.20 \times 10^{-6} - 2.52 \times 10^{-6}$ SNPs/site/year. The ST239 substitution rate is also similar to the substitution rates of other EMRSAs, such as ST22 ($1.2 \times 10^{-6} - 1.4 \times 10^{-6}$ SNPs/site/year; Holden *et al.*, 2013), and ST225 ($1.2 \times 10^{-6} - 2.9 \times 10^{-6}$ SNPs/site/year; Nubel *et al.*, 2010), including for collections of ST8 and CC30 sequences, which have been estimated as $1.02 \times 10^{-6} - 1.52 \times 10^{-6}$ SNPs/site/year and $1.05 \times 10^{-6} - 1.78 \times 10^{-6}$ SNPs/site/year, respectively (Alam *et al.*, 2015; Mcadam *et al.*, 2012).

3.2 Dating the time to MRCA of SCCmec-III

The SCCmec-III element is closely linked to the ST239 genome (Chapter 2 Section 3.3), and may have been acquired by ST239 subsequently to the hybridisation event, or may have been picked up by the ST30-like donor prior to the hybridisation. To determine a more accurate date of SCCmec-III acquisition by ST239, the SCCmec-III element of the 96 diverse ST239 sequences was extracted, and the time to MRCA was estimated using linear regression and BEAST.

3.2.1 Linear regression

Initially, a linear regression was used to determine whether there was any temporal signal in the SCCmec-III region that was isolated from the ST239 data-set, and provide a rough estimate of the time to MRCA (Figure 4.8). Even though there was substantial evidence of HGT in the SCCmec-III element (Chapter 3), the linear regression estimated significant temporal signal in the data from the core genes ($p = 0.00370$). The root-to-tip rate was

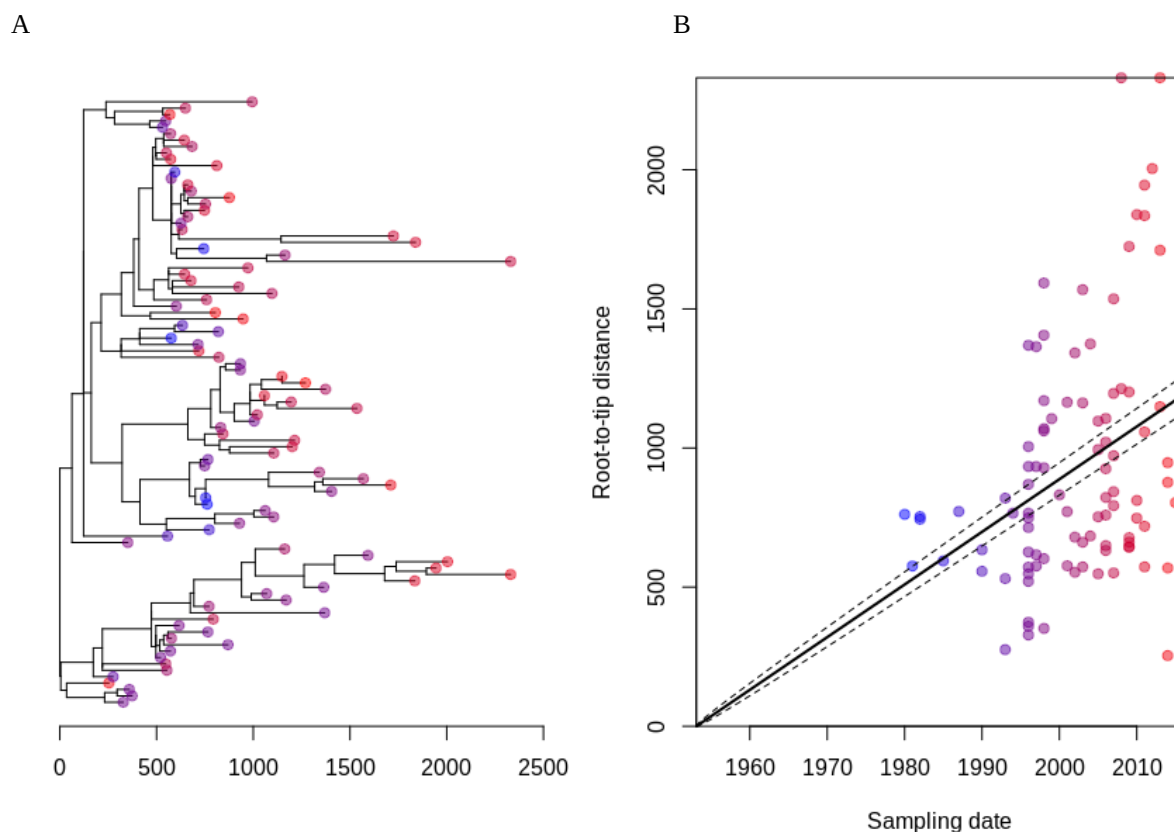


Figure 4.8. A) Maximum-likelihood phylogeny of SCCmec-III (units in SNPs/Mb). B) Root-to-tip distance plot against date of isolation (years). The solid black line is the linear regression line of best fit, and the dotted black lines represent the 95% CI.

estimated as 0.184×10^{-6} SNPs/site/year (95% CI $0.181 \times 10^{-6} - 0.198 \times 10^{-6}$ SNPs/site/year), and the time to MRCA was estimated as 1953.1, which is similar to the time to MRCA of the whole ST239 genome that was estimated using linear regression (Section 3.1.1).

3.2.2 BEAST

Four evolutionary models were run using BEAST (Drummond *et al.*, 2012) to estimate a time to MRCA of the SCCmec-III sequences (Table 4.3, Figure 4.9), which converged between 1939.1 – 1967.6 (95% HPD). This overlaps with the estimated time to MRCA for the ST239 whole-genome sequences of 1934.7 – 1943.3. Therefore, the most parsimonious interpretation is that the SCCmec-III element was acquired around the same time as the hybridisation event that formed ST239.

	BEAST evolutionary model (95% HPD)				Consensus
Substitution rate (95% HPD) ($\times 10^{-6}$ SNPs/site/year)	Strict constant	Strict exponential	Relaxed constant	Relaxed exponential	
	1.13 (0.649 – 1.64)	1.23 (0.743 – 1.75)	1.58 (0.606 – 2.50)	1.37 (0.751 – 2.06)	0.75 – 1.64
MRCA	1935.0 (1892.7 – 1967.6)	1959.7 (1939.1 – 1976.4)	1901.5191 (1791.9 – 1968.3)	1960.2 (1938.1 – 1977.4)	1939.1 – 1967.6

Table 4.3. SCCmec-III BEAST estimates of time to MRCA.

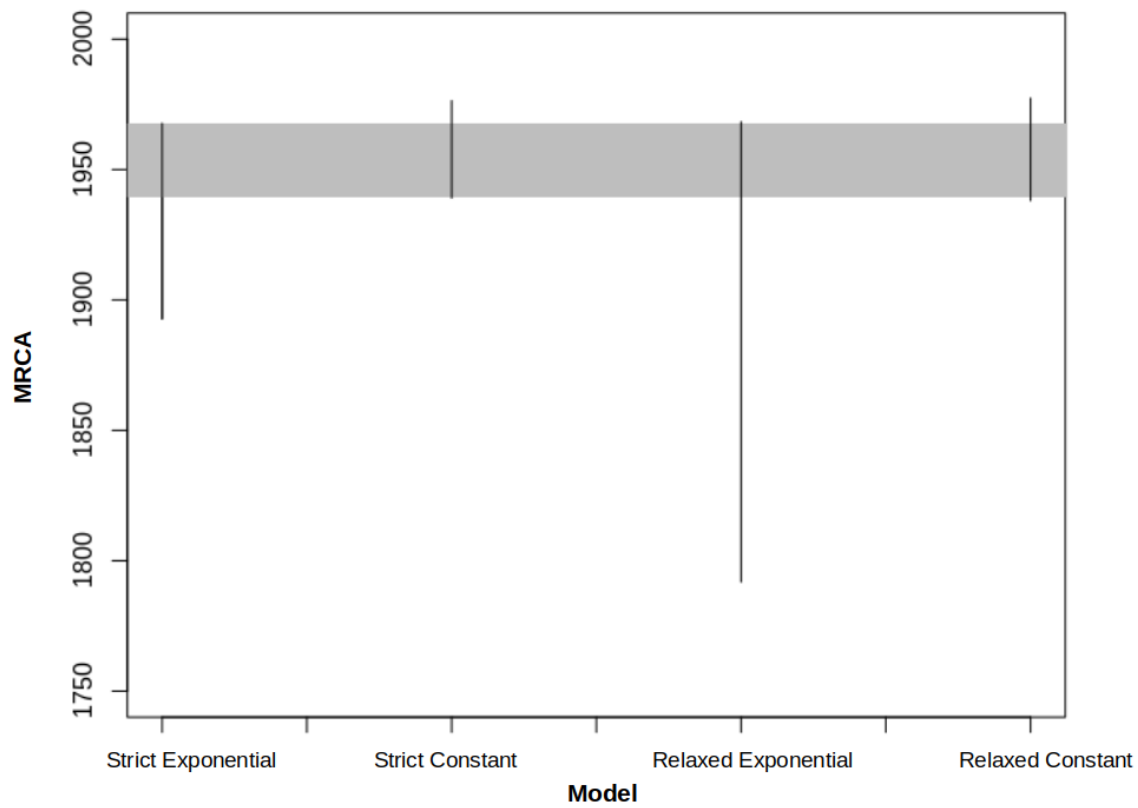


Figure 4.9. BEAST estimates of time to MRCA of SCCmec-III from ST239. The 95% HPD interval of each model is displayed by vertical black bars, and the consensus (where the MRCA of all models overlaps) is in grey.

The consensus estimated substitution rate was $0.751 \times 10^{-6} - 1.64 \times 10^{-6}$ SNPs/site/year (95% HPD) (Table 4.3), which is similar to the substitution rate of the ST239 whole genome (Section 3.1.2).

3.3 Adaptation of the acquired region and the backbone region in ST239

Evidence of compensatory adaptation in response to the large genetic transfer event that formed ST239 was investigated, in both the acquired region and the backbone region of the ST239 genome.

3.3.1 Substitution rates of the genetic regions of ST239

To investigate the hypothesis that the acquired region may have been evolving at a faster rate than the backbone region in ST239 due to compensatory adaptation that occurred subsequently to the hybridisation event, the two hybrid regions were extracted from the 96 ST239 sequences, and the substitution rate was estimated using linear regression and BEAST (Drummond *et al.*, 2012).

3.3.1.1 Linear regression

Initially, a linear regression was used to determine whether there was any temporal signal in the acquired region and backbone region of the ST239 data-set. There was significant temporal signal in both regions ($p < 0.0001$). For the backbone region, the linear regression estimated a root-to-tip rate of 2.61×10^{-6} SNPs/site/year (95% CI $2.40 \times 10^{-6} - 2.82 \times 10^{-6}$ SNPs/site/year). For the acquired region, the linear regression estimated a root-to-tip rate of 6.05×10^{-6} SNPs/site/year (95% CI $5.42 \times 10^{-6} - 6.68 \times 10^{-6}$ SNPs/site/year). This is suggestive of a faster substitution rate within the acquired region.

3.3.1.2 BEAST

Four evolutionary models were run using BEAST (Drummond *et al.*, 2012). For the backbone region, the consensus estimated substitution rate was $1.13 \times 10^{-6} - 1.29 \times 10^{-6}$ SNPs/site/year (95% HPD) (Table 4.4, Figure 4.10). This is consistent with the ST239 whole-genome consensus substitution rate of $1.13 \times 10^{-6} - 1.28 \times 10^{-6}$ SNPs/site/year (Section 3.1.2). For the acquired region, the consensus estimated substitution rate was $1.32 \times 10^{-6} - 1.68 \times 10^{-6}$ SNPs/site/year (95% HPD) (Table 4.4). This is further evidence of a faster substitution rate within the acquired region of 2.3% – 49.0% compared to the backbone region.

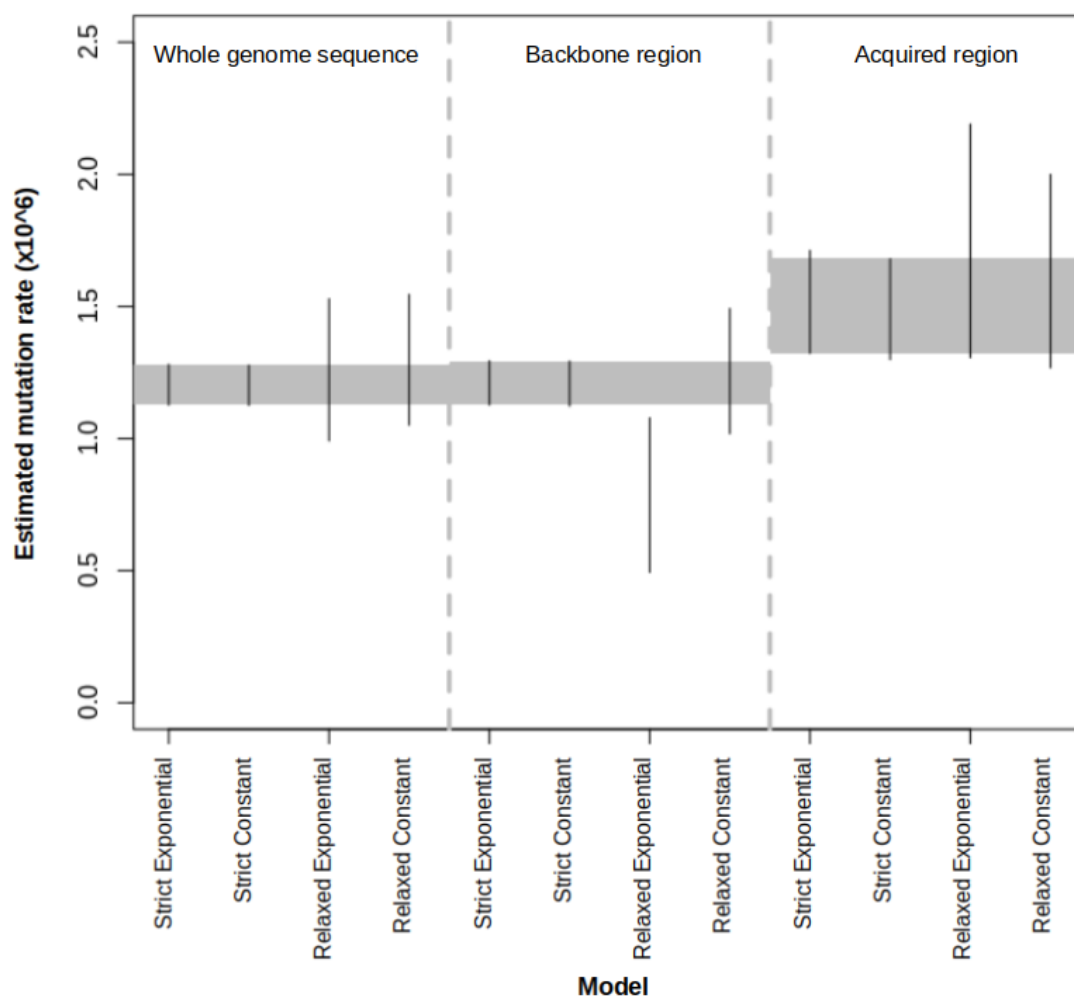


Figure 4.10. BEAST estimates of substitution rate of the ST239 whole-genome sequences whole genome, backbone region, and acquired region. The 95% HPD interval of each model is displayed by vertical black bars, and the consensus (where the substitution rates of all models for each region overlap) is in grey.

	Region	BEAST evolutionary model (95% HPD)				Consensus
		Strict constant	Strict exponential	Relaxed constant	Relaxed exponential	
Substitution rate (95% HPD) ($\times 10^{-6}$ SNPs/site/year)	ST8-like	1.21 (1.13 – 1.29)	1.21 (1.12 – 1.29)	0.749 (0.494 – 1.08)	1.25 (1.02 – 1.50)	1.13 – 1.29
	ST30-like	1.52 (1.32 – 1.71)	1.48 (1.30 – 1.68)	1.72 (1.31 – 2.19)	1.63 (1.27 – 2.00)	1.32 – 1.68

Table 4.4. ST239 BEAST estimates of substitution rate within the acquired and backbone regions.

3.3.2 McDonald-Kreitman test

To test for evidence of selection pressures that may be linked to the increased substitution rate of the acquired region of the ST239 genome, the 96 ST239 and 111 ST8 sequences, and the 96 ST239 and 57 ST30 sequences, were compared using the McDonald-Kreitman test.

The 316 core genes from the “ST30-donated” region, and the 1,659 core genes from the “ST8-donated” from Chapter 3 Section 3.1.2, were concatenated into two “mega-gene” sequences, and the total number of fixed synonymous and fixed non-synonymous substitutions within the ST239 sequences, and the total number of synonymous polymorphisms and non-synonymous polymorphisms between the ST239 sequences and the

STs	Region	Dn	Ds	Dn/Ds	Pn	Ps	Pn/Ps	N	Selection Pressure
ST239 and ST30	ST30-like	40	32	1.25	1406	740	1.90	1.52	Negative
ST239 and ST8	ST8-like	42	50	0.84	6080	3508	1.73	2.06	Negative

Table 4.5. McDonald-Kreitman net neutrality index for the two genetic regions of ST239. Fixed non-synonymous substitutions (Dn); fixed synonymous substitutions (Ds); non-synonymous polymorphisms (Pn); synonymous polymorphisms (Ps); McDonald-Kreitman net neutrality index (N).

ST30 or ST8 sequences, were recorded (Table 4.5; Appendix B.4.6.1). The McDonald-Kreitman net neutrality index was calculated for both genetic regions, and both regions showed evidence of purifying negative selection ($N > 1$), however, less evidence of purifying selection was observed in the acquired region.

3.3.3 Parallelism in ST239

Although purifying negative selection was determined to be the dominant selection pressure in both genetic regions of the ST239 sequences, it is likely that some positive selection may also have occurred. To test for evidence of positive selection, parallel evolution was investigated in the acquired and backbone regions of the 96 ST239 genome sequences.

The predicted Poisson distribution and the true distribution of substitutions in each core gene within each region of the ST239 sequences is displayed in Figure 4.11. A two-sample Kolmogorov-Smirnov test found that the predicted and observed distributions within the backbone region were significantly different ($D = 0.571$, $p = 0.00210$), and that the predicted and observed distributions within the acquired region were significantly different ($D = 0.571$, $p = 0.00210$). This suggests that there is evidence of parallel evolution in both genetic regions. A two-sample Kolmogorov-Smirnov test found that the true distributions were significantly different between the two hybrid regions ($D = 0.39286$, $p = 0.02656$) (Figure 4.12; Appendix B.5.2).

It was particularly notable that the acquired region contained a lower proportion of genes containing 0 or 1 substitutions, which is consistent with the higher substitution rate of the acquired region estimated in Section 3.3.1. Of the genes that had less than nine substitutions in total, genes in the acquired region had a higher ratio of non-synonymous to synonymous

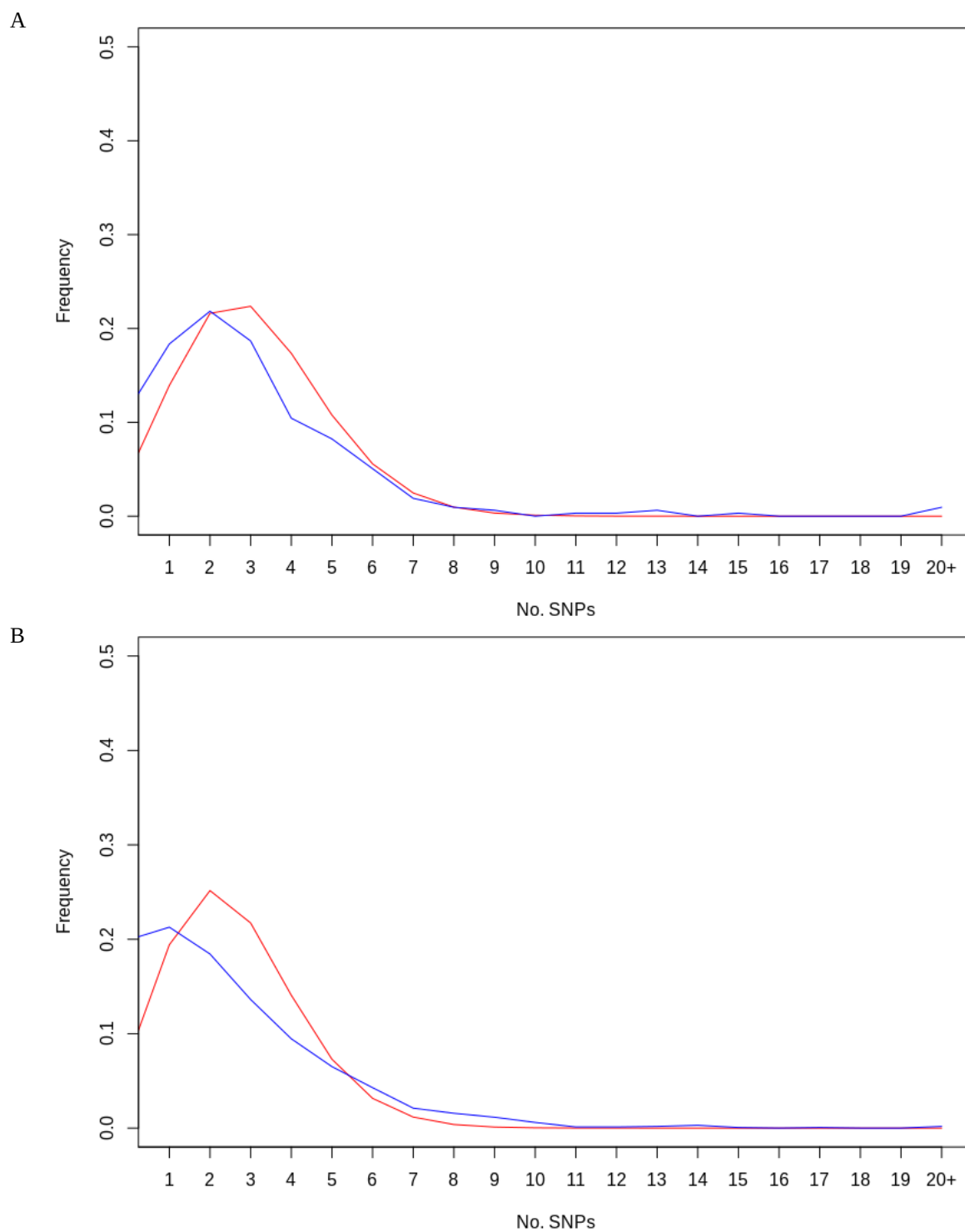


Figure 4.11. Distribution of fixed substitutions across a) the acquired region of ST239 and b) the backbone region of ST239. Red, predicted Poisson distribution; blue, true distribution.

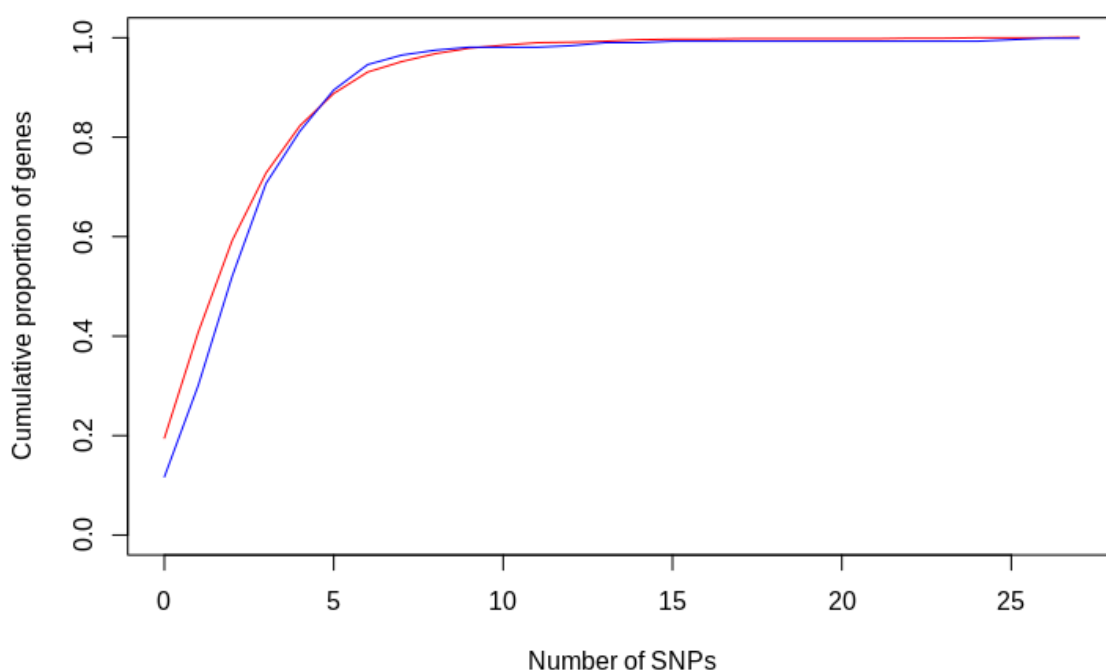


Figure 4.12. Distribution of fixed substitutions across the acquired region and the backbone region of ST239. Red, backbone region; blue, acquired region.

substitutions, compared to the backbone region (Appendix B.5.3). However, this difference was not significant (Chi-squared test, $X^2 = 0.4858$, $p = 0.4858$).

In both the acquired and backbone region, several genes contained a higher total number of substitutions than expected (Table 4.6 and Table 4.7), which may be evidence of positive selection in these genes. Six highly variable genes were from the acquired region, making up 1.90% of the total genes from this region. An additional 20 highly variable genes were from the backbone region, making up 1.21% of the total genes from this region. There was no significant difference between the proportion of highly variable genes from each genetic region (Chi-squared test, $X^2 = 0.9818$, $p = 0.3218$). Of the total substitutions in the acquired region, 154 (15.75%) were found in these six highly variable genes, and of the total

Gene name	Total SNPs	Dn	Ds	Dn/Ds	Function	General Function	Dn/Ds	
							in ST8	in ST30
<i>walk</i>	13	12	1	12.00	Cell wall metabolism sensor histidine kinase WalK. Regulates genes involved in autolysis, biofilm formation and cell wall metabolism (Ji <i>et al.</i> , 2016), and linked to vancomycin resistance (Jasen <i>et al.</i> , 2007)	AMR, met	0.50	0.50
<i>spa</i>	62	48	14	3.43	Peptidoglycan-binding protein LysM. Virulence factor that enables host immune response evasion. Highly variable (Votintseva <i>et al.</i> , 2014)	Vir	2.53	3.63
<i>gsiA</i>	13	10	3	3.33	Peptide ABC transporter ATP-binding protein. Down-regulation linked to antimicrobial peptide resistance (Kubicek-Sutherland <i>et al.</i> , 2016)	AMR	5.00	2.00
<i>ausA</i>	26	18	8	2.25	Hypothetical aureusimine synthesis protein – potential virulence factor (Wilson <i>et al.</i> , 2013)	Vir	2.29	2.63
<i>manP</i>	15	7	8	0.88	Phospho-transferase system mannose transporter subunit IIABC (Kotrba <i>et al.</i> , 2001)	Met	0.60	4.00
<i>tarL</i>	25	7	18	0.39	Teichoic acid biosynthesis protein (make up the cell wall structure) (Pereira <i>et al.</i> , 2008)	Met	1.00	0.75

Table 4.6. Highly variable genes within the acquired region of the ST239 sequences. AMR = antimicrobial resistance; Met = metabolism; Vir = virulence.

substitutions in the backbone region, 284 (6.61%) were found in these 20 highly variable genes. This was significantly different (Chi-squared, $X^2 = 86.75$, $p < 0.00001$).

An outlier in the acquired region contained the highest number of substitutions ($n = 62$), which is over twice as many substitutions as the next-highest substitution-containing gene. This outlier was the *spa* gene, which is a known highly-variable gene that frequently undergoes rearrangements and is used for sub-typing different *S. aureus* STs (Votintseva *et al.*, 2014). After excluding the *spa* outlier from the acquired region, five highly variable genes were found in the acquired region, including 92 (10.0%) of the total substitutions in the

Gene name	Total SNPs	Dn	Ds	Dn/Ds	Function	General Function	Dn/Ds	
							in ST8	in ST30
SAUPA N00268 9000	11	11	0	>	Putative RNase adaptor protein RapZ	Met	0.50	2.00
<i>frp</i>	12	12	0	>	NAD(P)H-flavin oxidoreductase. Cell-wall protein linked to iron-restricted growth conditions (Morrissey, 2002)	Met	>	>
<i>parC</i>	14	13	1	13.00	DNA topoisomerase IV subunit A. Fluoroquinolone resistance.	AMR	3.33	1.00
<i>mprF</i>	10	9	1	9.00	Phosphatidylglycerol lysyltransferase, linked to methicillin resistance	AMR	1.25	3.00
<i>pknB</i>	10	9	1	9.00	Serine/threonine-protein kinase	Met	0.33	1.00
SAOUH SC_011 30	13	11	2	5.50	YfcC family arginine/ornithine APC Transporter	Met	2.00	1.00
<i>pycA</i>	11	9	2	4.50	Putative pyruvate carboxyl transferase	Met	0.50	1.80
<i>glyA</i>	11	9	2	4.50	Serine hydroxymethyltransferase	Met	1.33	0.67
<i>fdhA</i>	13	10	3	3.33	Formate dehydrogenase. Contains FeS cluster (Mäder <i>et al.</i> , 2016)	Met	4.50	2.00
<i>rpoB</i>	22	16	6	2.67	DNA-directed RNA polymerase subunit β . Rifampicin resistance (Aubry-Damon <i>et al.</i> , 1998)	AMR	1.33	1.00
<i>ponA</i>	24	16	8	2.00	PBP2, cell wall biosynthesis. β -lactam-resistance (Leski and Tomasz, 2005)	AMR	1.00	1.00
<i>sucA</i>	14	9	5	1.80	2-oxoglutarate dehydrogenase E1 component	Met	1.50	3.00
<i>dinG</i>	14	9	5	1.80	ATP-dependent helicase	Met	1.40	1.00
<i>dnaK</i>	14	9	5	1.80	Molecular chaperone	Met	1.00	2.67
<i>pbpB</i>	15	8	7	1.14	Penicillin binding protein 2, related to β -lactam resistance	AMR	0.83	1.20
<i>atl</i>	17	9	8	1.13	Bifunctional autolysin (virulence factor)	Vir	0.57	1.22
<i>hsdS1</i>	14	7	7	1.00	Restriction endonuclease subunit S	Met	2.00	0.50
<i>sdrH</i>	27	8	19	0.42	Hypothetical Serine-Aspartate Repeat protein. Highly variable (Xue <i>et al.</i> , 2011)	Met	0.75	0.19
<i>lpdA</i>	9	1	8	0.13	Dihydrolipoamide dehydrogenase	Met	6.00	0.00
<i>ybbH</i>	9	1	8	0.13	MurR/RpiR family transcriptional regulator	Met	0.00	0.50

Table 4.7. Highly variable genes within the backbone region of the ST239 sequences. AMR = antimicrobial resistance; Met = metabolism; Vir = virulence.

acquired region. This difference was still significant (Chi-squared, $X^2 = 13.111$, $p = 0.000294$). This suggests that the highly variable genes from the acquired region contain more substitutions than in the backbone region.

The functions of the 26 highly variable genes include multiple AMR-related functions, such as β -lactam, rifampicin, fluoroquinolone, vancomycin and antimicrobial peptide resistance, suggesting that AMR has been one of the major evolutionary pressures faced by ST239. Additionally, several virulence related genes also contained a high number of substitutions, suggesting adaptation of ST239 to host immunity. This is consistent with a study by Baines (2015), where two Australian ST239 clades were observed that had evolved different AMR and virulence profiles independently through mutations in highly conserved essential regulatory genes. Baines suggested that the hospital environment had favoured the accumulation of mutations in essential MRSA genes that increase resistance to antimicrobials, attenuate virulence, and promote persistence in the healthcare environment, and identified a high number of mutations in the *rpoB* and *walk* genes that were linked to decreased susceptibility to the glycopeptide vancomycin (Baines *et al.*, 2015). Both *rpoB* and *walk* were also identified as highly variable genes in this study (Appendix B.5.4.1 – B.5.4.2).

In the majority of cases, the highly variable genes had a high ratio of non-synonymous to synonymous substitutions. Therefore, the structure and functions of these genes may be impacted by these substitutions, suggesting that these are evidence of adaptation through positive selection. In a few examples, such as *tarL*, *lpdA* and *ybbH*, there was a greater number of synonymous substitutions, and therefore it is unclear why these genes are so variable in this data-set, however, it is possible that these genes are hypermutable or that there

Chapter 4

is positive selection acting on synonymous mutations in these genes (Lebeuf-Taylor *et al.*, 2019).

In the majority of cases, genes that were highly variable in ST239 were also highly variable within the ST8 or ST30 data-sets, suggesting that ST8 and ST30 have also faced similar selection pressures, including adaptation to host immunity (e.g. *spa* and *ausA*) and AMR (e.g. *gslA*, *mprF* and *rpoB*). There are a few examples where genes that were highly variable in ST239 were less variable in ST8 or ST30, such as *walk*, *pknB*, *ponA*, *SAUPAN002689000*, and *SAOUHSC_01130*, which are involved in metabolism and resistance to vancomycin and β -lactams. Therefore, this suggests that there may be some differences in the selection pressures facing ST239 compared to the ST8 and ST30 lineages.

3.3.4 Tajima's D

Overall, the ST239 sequences showed evidence of a high proportion of rare alleles (Tajima's $D = -2.483$, $p = 5.618 \times 10^{-4}$) (Korneliussen, *et al.*, 2013). This is suggestive of a recent expansion of the ST239 population, which is consistent with the structure of the ST239 phylogeny in Chapter 2, Section 3.2.3. Both the acquired and the backbone region had similar negative Tajima's D values (Tajima's $D = -2.942$, $p = 1.252 \times 10^{-7}$ and Tajima's $D = -2.842$, $p = 3.423 \times 10^{-6}$, respectively), suggesting that both regions have been impacted equally by the recent population expansion event.

3.3.5 Purine-pyrimidine and transition-transversion substitution proportions

The average guanine and cytosine (GC) content of genomes varies between bacterial species, and the *S. aureus* genome typically has a GC content of roughly 33% (Muto and Osawa,

1987; Petit *et al.*, 2018). The GC content of both the acquired region and backbone region of the ST239 reference sequence TW20 were similar (33.5% and 32.6%, respectively), despite the presence of the large SCCmec-III MGE in the acquired region. Overall, both regions of the ST239 genome had a positive GC $\rightarrow$ AT/AT $\rightarrow$ GC ratio, which is consistent with the typical bias towards GC $\rightarrow$ AT substitutions in bacterial species, regardless of genomic GC content (Hershberg and Petrov, 2010). However, the GC $\rightarrow$ AT/AT $\rightarrow$ GC ratio in the acquired region was significantly greater than that of the backbone region ($X^2 = 4.784$, $p = 0.0287$) (Table 4.8), which is a further indicator of the weaker selective pressures acting on the acquired region.

Transition substitutions generally occur more frequently than transversion substitutions, and adaptation can be influenced by this mutational bias (Stoltzfus and McCandlish, 2017). The proportion of transitions to transversions observed in both regions of the ST239 genome (Table 4.8) was consistent with those of other pathogens (Foster *et al.*, 2015; Dettman *et al.*, 2016; Kucukyildirim *et al.*, 2016). However, there was no significant difference between the transition-transversion proportions within the acquired region compared to the backbone region ($X^2 = 1.912$, $p = 0.1668$).

	Total number of variant sites	
	Acquired region	Backbone region
GC $\rightarrow$ AT	795	2,710
AT $\rightarrow$ GC	474	1,864
GC $\rightarrow$ AT / AT $\rightarrow$ GC	1,677	1,454
No. transitions	945	3,578
No. transversions	510	1,772
Transitions/transversions	1.853	2.019

Table 4.8. Purine-pyrimidine and transition-transversion substitution ratios within the ST239 genome.

3.3.6 Consideration of intergenic regions

Non-translated intergenic regions compose 10 – 15% of bacterial genomes, and contain many regulatory elements with key functions, and there is evidence of the action of selective pressures on these regions (Molina and van Nimwegen, 2008; Thorpe, *et al*, 2017). However, studies of selection pressures on bacterial genomes do not typically include non-translated intergenic regions. One limitation of this study has therefore been the focus on selective pressures that act on only the core and accessory genome, without considering non-translated intergenic regions. The core intergenic region shared between all 96 ST239 sequences consists of 340,643 bp, making up 11.2% of the ST239 genome, and includes 1,093 additional variant sites. The inclusion of core intergenic region variant sites in this study would therefore increase the number of core SNPs included in this study by 16.0%.

4. Discussion

Epidemic ST239 was first described in 1981 in the UK (Cookson and Phillips, 1988). However, estimates of the emergence date of ST239, and the formative hybridisation event, range from between 1965.7 – 1974.4 (Didelot, 2019), to as broad as 1885 – 1973 (Baines *et al.*, 2015). In this study, the time to MRCA was estimated from the 96 temporally and geographically diverse ST239 sequences from Chapter 2, using two complementary bayesian phylogeny inference methods. The MRCA was estimated as between 1934.7 – 1943.3 (overlapping 95% HPD intervals from similar models), which is earlier than previous estimates (Harris *et al.*, 2010; Gray *et al.*, 2010; Smyth *et al.*, 2010; Didelot, 2019). The time to MRCA of the SCC*mec*-III element from the same 96 ST239 sequences was estimated to a similar time.

This earlier estimated MRCA of ST239 suggests that the emergence of ST239 was not driven by the clinical introduction of methicillin in 1959 (Knox, 1960). This is similar to estimates of the date of introduction of the first SCC*mec*-I element into *S. aureus*, which was estimated by Harkins (2017) as 1941. This was 14 years prior to the clinical introduction of methicillin. Harkins hypothesised that the introduction of penicillin into clinics in the early 1940s was the more likely driver of the emergence of MRSA, rather than methicillin, and also identified increased resistance determinants to additional antibiotics in archaic *S. aureus* isolates, such as for tetracycline and streptomycin, that were used in higher concentrations and frequency in the 1950s and 1960s (Harkins *et al.*, 2017). Both of these resistances are known to be carried by SCC*mec*-III elements, and were associated with the SCC*mec*-III elements in this study. Therefore, it is likely that clinical use of antibiotics such as penicillin, tetracycline and

streptomycin had a greater influence on the emergence and global expansion of ST239 than methicillin.

The estimates of the time to MRCA of the ST239 isolates in this study were based on a total of 6,819 SNPs from 96 diverse sequences. This study was able to capture a greater diversity, and utilise more genetic information, than previous studies that estimated the MRCA of ST239 sequences. For example, studies that analysed the Harris (2010) ST239 isolates were based on 4,310 SNPs from 45 isolates from a global isolate collection, and 20 isolates from a single hospital in northeast Thailand that were isolated between 2006 – 2007 during an outbreak in an ICU (Harris *et al.*, 2010). The earlier estimated MRCA of the ST239 isolates in this study may therefore be reflective of the greater diversity of the ST239 isolates used, which may be more representative of the true population structure of ST239, and therefore closer to the true date of the hybridisation event that formed ST239.

In this study, linear regression was used to determine evidence of temporal signal in each data-set, and estimate a rough time to MRCA. The estimated time to MRCA was not consistent with bayesian phylogeny inference methods, as the method for estimating the time to MRCA is more crude (Drummond *et al.*, 2003). Both BactDating and BEAST bayesian phylogeny inference methods were able to estimate a similar time to MRCA, within 1.5 years. Therefore, the BactDating methodology is able to infer dated bacterial phylogenies much faster and with a similar accuracy to BEAST (Didelot *et al.*, 2018).

The time to MRCA of the SCC*mec*-III elements was estimated as 1939.1 – 1967.6 (overlapping 95% HPD intervals from similar models), which overlapped with the time to

MRCA that was estimated for the whole ST239-genome. This may be further evidence that SCC*mec*-III may have been a key driver of the emergence of the ST239 hybrid prior to clinical methicillin usage. Unfortunately, the estimate of the time to MRCA of the SCC*mec*-III element had wide CIs, which is most likely due to the smaller number of variant sites that were included in the SCC*mec*-III analysis. A large proportion of genetic information was removed from the SCC*mec*-III analyses due to evidence of recombination in these regions (Chapter 3), and therefore the temporal analysis is based on just 54 variant sites. However, there was evidence of temporal signal, and this is still an improvement on the reconstructed evolutionary history of the SCC*mec*-I element, which was based on just 39 variant sites (Harkins *et al.*, 2017).

The substitution rate of the ST239 isolates in this study was estimated as $1.13 \times 10^{-6} - 1.28 \times 10^{-6}$ SNPs/site/year, which is on the slower side of the ST239 substitution rate estimated by Harris (2010) of $1.20 \times 10^{-6} - 2.52 \times 10^{-6}$ SNPs/site/year. The substitution rate of the acquired region was found to be greater than the substitution rate of the backbone region, which may be evidence of stronger positive selection or weaker selective constraints in the acquired region. It is possible that the weaker selective constraints that act on this region were key to the success of the hybridisation event that formed ST239, and allowed for such a large genetic transfer event involving this region of the genome. This may also explain why large genetic transfer events involving this region of the *S. aureus* genome have been involved in the formation of other *S. aureus* hybrids (Robinson and Enright, 2004; Thomas *et al.*, 2012; Nimmo *et al.*, 2014).

Chapter 4

The McDonald-Kreitman analysis showed evidence that purifying negative selection was the dominant selection pressure in both the acquired and backbone regions. However there was less evidence of purifying selection in the acquired region, and there was a lower proportion of genes containing 0 or 1 substitutions in the acquired region. This is further evidence of weaker purifying selection in the acquired region compared to the backbone region.

Many of the highly variable genes in both genetic regions displayed strong evidence of positive selection, with a high ratio of non-synonymous to synonymous substitutions. The functions of many of these highly variable genes were related to AMR and virulence, suggesting that these are two major evolutionary pressures faced by ST239. However, not all non-synonymous mutations will result in a phenotypic change. For example, only five of the 16 non-synonymous mutations in the *rpoB* gene are known to be causative of rifampicin resistance. Less is known about *rpoB* mutations that do not cause rifampicin resistance, but at least one of the *rpoB* mutations that was observed in the ST239 data-set, V135L, is a known rifampicin-sensitive *rpoB* mutation. It is possible that substitutions such as these reflect compensatory adaptation to offset the fitness costs of AMR substitutions, however, these substitutions may not necessarily be related to the adaptation of ST239 to evolutionary pressures.

Finally, one drawback of using clinical data for evolution studies, compared to controlled lab-based experimental evolution studies, is that the evolutionary comparisons drawn between the ST239, ST8 and ST30 genomes may be confounded by the different evolutionary distances between these three lineages. The ST30 sequences are more distantly related to the homologous ST239 acquired region, than the ST8 sequences are to the homologous ST239

backbone (Chapter 2 Section 3.4.1). This may result in a higher number of polymorphisms per gene in the acquired region, which may confound the McDonald-Kreitman analysis. Therefore, stronger signals of purifying selection in the analysis of the acquired region may be expected, because there has been more evolutionary time for selection to purge very weakly deleterious mutations in this region. However, in this study, less evidence of purifying selection was observed in the acquired region, which is further evidence of weaker purifying selection acting on the acquired region.

The large chromosomal replacement event that formed ST239 has driven the distinct evolutionary trajectories of the genetic regions that originated from two divergent ancestors within the same genome. An accelerated rate of evolution in the acquired region is linked to evidence of weaker selective constraints acting on this region of the genome. There is evidence of adaptation in the ST239 genome due to subsequent major evolutionary drivers, including AMR and host-immune response, which may have been key to the success of ST239 in healthcare settings around the globe. By analysing sequences from an increased diversity of ST239 isolates, an earlier time to MRCA was predicted than previous estimates, which may provide clues to the factors that drove the hybridisation and global expansion of ST239, which occurred prior to the clinical introduction of methicillin. Combined with evidence of the acquisition of the *SCCmec*-III element early on in its evolutionary history, this suggests that the formation and global expansion of ST239 was more likely driven by the use of earlier antimicrobials.

Chapter 5

Quantifying the competitive ability of multiple *S. aureus* strains in co-culture using whole-genome sequencing

Abstract

In their natural environment, bacteria often face competition for resources from a variety of competitors in the presence of additional complex factors. Healthcare-associated bacteria in particular are exposed to highly stressful environments, including high levels of antibiotics, the host immune system, and the presence of other pathogens. Techniques to study bacterial competition *in vitro* often focus on pairwise bacterial interactions and can be difficult to perform on clinical isolates. However, whole-genome sequencing provides a platform to study larger scale competition experiments between several different strains in co-culture. In this chapter, I aim to develop a method to estimate the competitive ability of 15 ST239, ST8 and CC30 *S. aureus* isolates in co-culture, by calculating the ratio of sequencing coverage at variant sites that are unique to each of the isolates. By comparing the sequencing coverage at a wide range of loci across the isolate genomes before and after competition, it is possible to estimate the success of each isolate in competition within a range of different growth-promoting and growth-inhibiting conditions. This method could be applied in the future with a greater range of isolates and increased depth of sequencing coverage for larger scale competition studies.

1. Introduction

1.1 Aims and objectives

In this chapter, I aim to develop a method to determine the competitive ability of several *S. aureus* isolates in co-culture using whole genome sequencing, which can be used as a measure of bacterial fitness in Chapter 6.

1.2 Bacteria as competitors

In the natural world, bacteria live in complex communities, colonising extremely diverse ecological niches, and rapidly adapting to changing environments. These environments are defined by a number of complex abiotic factors, including chemical, spatial and nutritional factors, which impact microbial growth (Stubbendieck and Straight, 2016). In addition to overcoming environmental factors, bacteria must co-operate, or more likely compete, with one another to survive and as such they have evolved a range of mechanisms to battle against competitors, including complex biological, chemical and mechanical weapons (Foster and Bell, 2012; Granato *et al.*, 2019). In microbial systems, competition can include growth limitation, inhibition due to exploitation and interference, or a variety of different mechanisms that may enhance the competitive fitness of a bacterial species, such as antibiotic synthesis, motility, sporulation, predatory functions and biofilm formation (Stubbendieck and Straight, 2016).

Culture-dependent pairwise interaction experiments are a common method of studying bacterial competitions *in vitro*. For example, competition on an agar plate can be visualised through changes in visible phenotypes, such as developmental defects, growth inhibition,

lysis, motility, and pigment production (Stubbendieck *et al.*, 2016). Methods to quantify morphologically distinct bacterial species in competition, such as *S. aureus* and *P. aeruginosa*, include counting CFUs plated on agar, estimating cell count using flow cytometry, and direct measurement of bacterial count using microscopy (Magalhaes *et al.*, 2019). In order to differentiate between strains in pairwise competitions, strains can be genetically labelled with fluorescent constructs including GFP, YFP and CFP (Keymer *et al.*, 2008), which can be monitored in intra-species competitions using fluorescence microscopy (Stannek *et al.*, 2014). Other labels include genes for AMR or visible reactions with dyes, or selective mutations such as Ara⁺/Ara⁻, which can be used to distinguish between strains through counting CFUs on selective media after co-culturing (Elena and Lenski, 2003; Wiser and Lenski, 2015). Alternatively, qPCR can be used to quantify the presence of 16S RNA compared to a reference RNA, and make direct comparisons between the competitive ability of two different species (Magalhaes *et al.*, 2019).

The number of strains that can be included in each competition using many of these methods is limited. However, techniques such as qPCR can be used to quantitatively assess the competitive ability of a mixture of numerous strains simultaneously (Yoon *et al.*, 2011). This requires that each genome contains a unique DNA bar-code that can be used to distinguish between the strains, which can be introduced experimentally or can be naturally occurring (Nadkarni *et al.*, 2002).

1.3 *S. aureus* as a competitor

In order to establish an infection, healthcare associated *S. aureus* must compete against other human pathogens, other healthcare associated *S. aureus* strains, and the host natural microbial

flora, in hostile environmental conditions that may include the host immune response and high concentrations of antimicrobials. For example, the lung microbiome of patients with cystic fibrosis is dominated by *S. aureus* at an early age. However, by their late teens, *S. aureus* is typically out-competed by *P. aeruginosa*, which is associated with a worse prognosis and general decline in lung function (Harrison, 2007). Antagonism occurs as *P. aeruginosa* is able to lyse *S. aureus* cells in cystic fibrosis sputum, increasing the competitive ability of *P. aeruginosa in vitro* (Palmer *et al.*, 2005).

S. aureus is also a common human commensal, colonising the nares of about 27% of people (Wertheim *et al.*, 2005). In this niche, *S. aureus* must compete against other species, such as *S. epidermidis* and *Corynebacteria* (Uehara *et al.*, 2000), as well as other strains of *S. aureus* (Dall’Antonia *et al.*, 2005). The competitive ability of *S. aureus* in the nares can have significant impacts on the host, as *S. aureus* carriers can become infected by the endogenous *S. aureus* that colonise their nasal mucosa (Von Eiff, 2001), making carriers three times more likely to develop bloodstream infection when undergoing surgery (Wertheim *et al.*, 2004).

1.4 Chapter summary

The relative competitive ability of 15 ST239, ST8 and CC30 *S. aureus* isolates, in co-culture in several different growth conditions, was estimated using a novel whole-genome sequencing approach. This method will be used as a measure of bacterial fitness in Chapter 6.

The ratio of coverage at variant sites and genes that are unique to each single isolate was used to calculate the competitive ability of the isolates relative to one another in each growth condition. Competitive ability was calculated from unique variant site coverage, and from

Chapter 5

unique gene coverage, which were found to be correlated. Each competition experiment was highly reproducible, which allowed for differences in the relative competitive ability of the strains to be estimated. Growth of isolates that were susceptible to antibiotics could not be detected after competition in these antibiotics.

Although there were limitations of this method, such as a low number of unique variant sites and genes, a high proportion of MGE-associated unique genes, and a conservative threshold for detection, this method was able to identify differences in competitive ability between the 15 isolates of *S. aureus* in a variety of conditions.

2. Methods

2.1 Bacterial cultures

Unless otherwise stated, cryostocks of fifteen isolates from Chapter 2 Section 2.6.1 were streaked on TSA, and incubated at 37°C for 24 hours. Single colonies were incubated for 24 hours in 3 mL TSB at 37°C with 225 RPM shaking.

2.2 Creating a pipeline to calculate competitive ability

2.2.1 Identification of unique variant sites

A consensus sequence was created from the core gene sequences of the experimental isolates from Chapter 2 Section 2.6.2, using EMBOSS v6.6.0.0 (EMBOSS, 2000). Variant sites from each core genome sequence, compared to the consensus sequence, were extracted using snp-sites (Page *et al.*, 2016). Unique variants are described as variant sites that were found in only a single isolate (Figure 5.1).

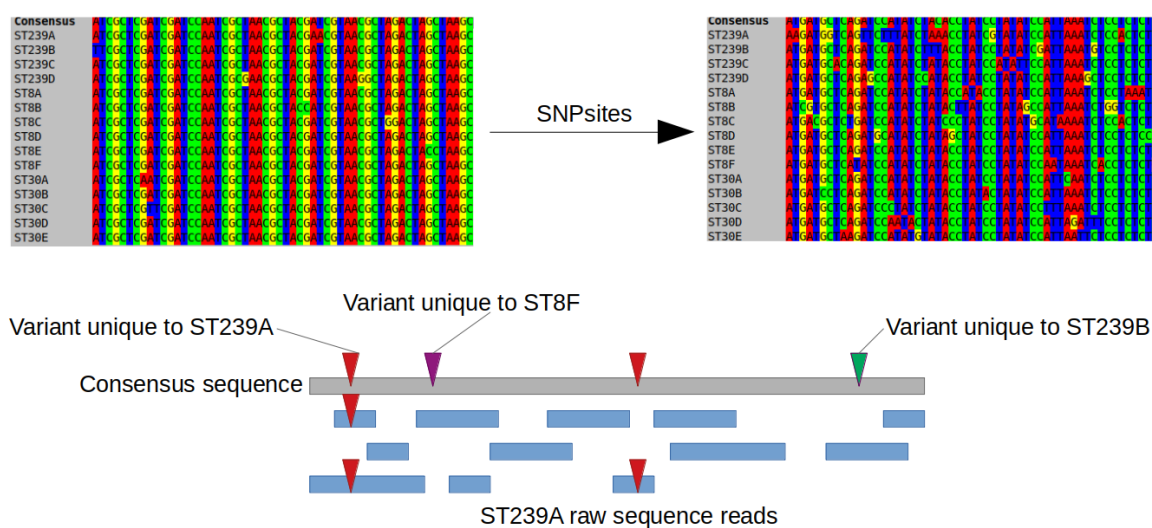


Figure 5.1. Identification of unique variant sites.

To verify the validity of each unique variant site, the DNA sequences for each individual isolate (Chapter 2 Section 2.6.2) were mapped to the consensus sequence using bwa v0.7.15 and SAMtools v1.3.1 (Li and Durbin, 2009) (Appendix B.1.8). The number of reads supporting each unique variant site was extracted from the resulting vcf file. Only unique variant sites that could be used to accurately identify a single isolate, and that were supported with 4 or more reads, were included as true unique variant sites.

The consensus sequence of each core gene was aligned to the ST239 TW20 reference sequence (accession number FN433596) using MegaBLAST (Morgulis *et al.*, 2008), and the locations of each gene containing at least one unique variant site were recorded. All graphs were plotted in R v3.4.4.

2.2.2 Identification of unique genes

Roary (Page *et al.*, 2015) was used to extract the genes that were unique to a single isolate from the accessory genome of the isolates from Chapter 2 Section 2.6.2. In total, 1,120 candidate unique genes were identified (Appendix B.1.9). Of these, 690 genes that were under 300 bp in length were discounted, as these were likely due to the presence of pseudo-genes. The remaining 430 candidate unique genes were concatenated to form a single sequence containing only the remaining candidate unique genes (Figure 5.2).

To test the validity of these unique genes, the raw sequence reads for each isolate were mapped to the concatenated candidate unique gene sequence using bwa v0.7.15 and SAMtools v1.3.1 (Li and Durbin, 2009) (Appendix B.1.10), and the number of reads supporting each nucleotide in each gene along the sequence was recorded. Only genes that

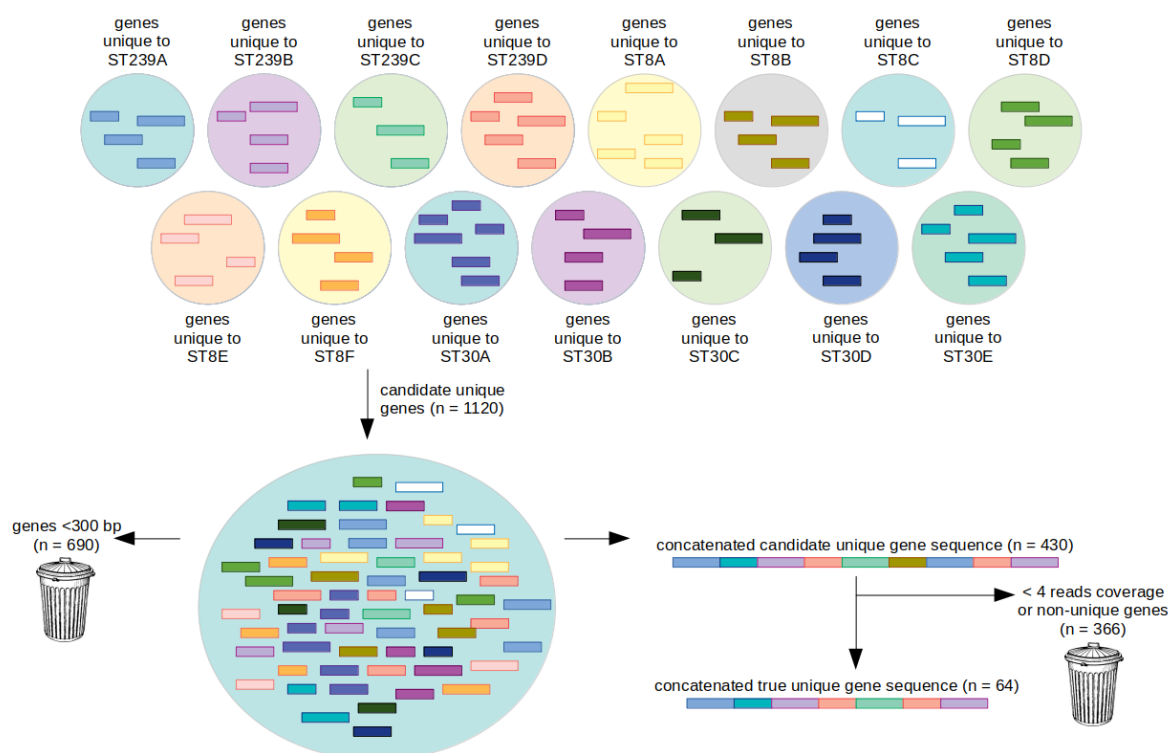


Figure 5.2. Identification of unique genes.

could be used to accurately identify a single isolate, and that had a maximum coverage of 4 or more reads were included as true unique genes. The remaining 64 true unique genes (Appendix B.4.4) were concatenated into a single sequence.

2.2.3 Competition experiments

All isolates were grown from cryostock as in Section 2.1. One mL of each culture was combined into a single mixture and mixed thoroughly. One mL of the mixture was reserved for DNA extraction (Section 2.2.4). Six mL of the mixture was pelleted and washed three times in PBS, and separated into six 1 mL aliquots. These were diluted 50x in either TSB, BHI, PS, TSB with 4 µg/mL oxacillin, TSB with 2 µg/mL norfloxacin, or TSB with 8 µg/mL trimethoprim and 152 µg/mL sulfamethoxazole. A total of 3 mL of each mixed culture was

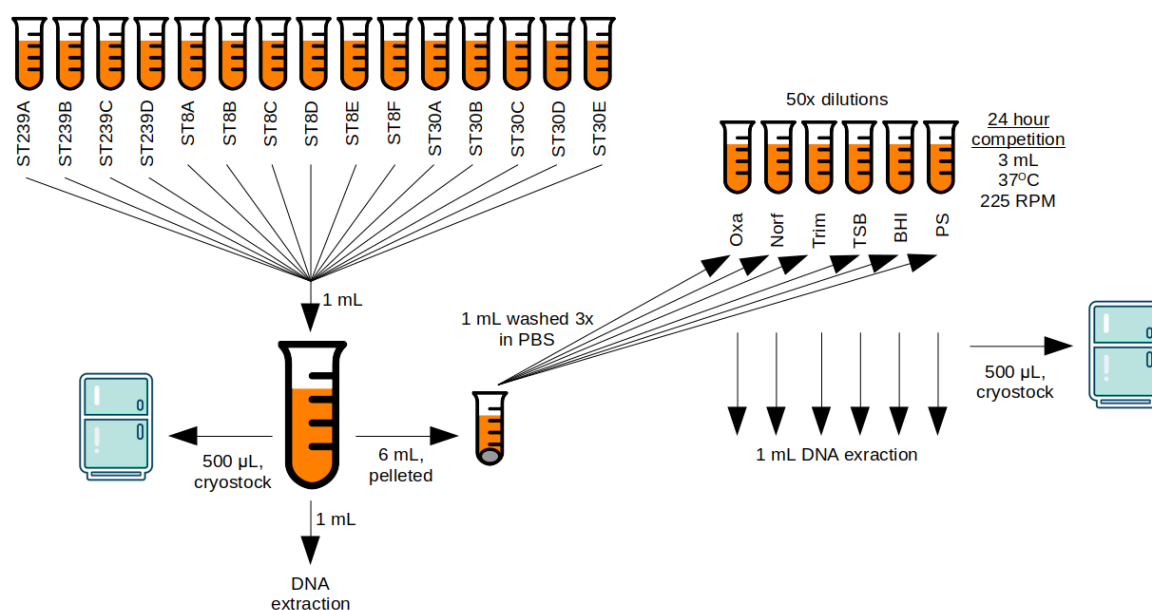


Figure 5.3. Competition experimental protocol. Oxa = 4 µg/mL oxacillin, Norf = 2 µg/mL quinolone, Trim = 8 µg/mL trimethoprim with 152 µg/mL sulfamethoxazole.

incubated at 37°C with 225 RPM shaking for 24 hours. After 24 hours, 1 mL of each solution was reserved for DNA extraction (Section 2.2.4) (Figure 5.3). All competition experiments were repeated in triplicate.

A total of 500 µL from all mixtures before and after each competition were mixed with 50% glycerol and stored at -80°C.

2.2.4 DNA extraction and sequencing

Roughly 1 mL of each culture was digested in enzymatic lysis buffer containing 25 – 30 mg/mL lysozyme, 0.1 mg/mL lysostaphin, 20 mM TrisCl, 2 mM EDTA and 1.2% Triton-X100 for 2 hours. DNA was purified using the Qiagen DNeasy Blood and Tissue kit (QIAGEN, 2006), and following the protocol for purification of bacterial or yeast DNA with enzymatic lysis, using QIAcube. The concentration of genomic DNA in each sample was then measured

using a Promega QuantiFluor dsDNA System, and samples were diluted to 10 – 50 ng/μL in DNase-free water before sequencing was carried out using an Illumina HiSeq4000 system with 150 bp paired-end reads by the Oxford Genomic Centre, Wellcome Trust Centre for Human Genetics, University of Oxford, UK.

2.2.5 Mapping

The DNA sequences were mapped to the core gene consensus sequence (Section 2.2.1), and the concatenated true unique genes sequence (Section 2.2.2) using bwa v0.7.15 and SAMtools v1.3.1 (Li and Durbin, 2009).

2.2.6 Base calling quality control

The number of reads supporting each unique variant site allele (either the reference allele or the variant allele) was determined from the mapping output (Section 2.2.5) from each competition in BHI (Section 2.2.3). The number of reads supporting incorrect alleles (i.e. a base that is neither the reference base nor the unique variant base, and is therefore likely to be a sequencing error) was determined for each position. A cut-off of four supporting reads was selected as the number of supporting reads required to call the reference allele and/or variant allele at any position. Any base at any position that was supported by three or fewer reads was discarded for all further analyses.

2.3 Estimating competitive ability from unique variant sites

The number of reads supporting each unique variant site allele (for both the reference allele and the variant allele) was determined from the mapping of the raw sequence reads from each

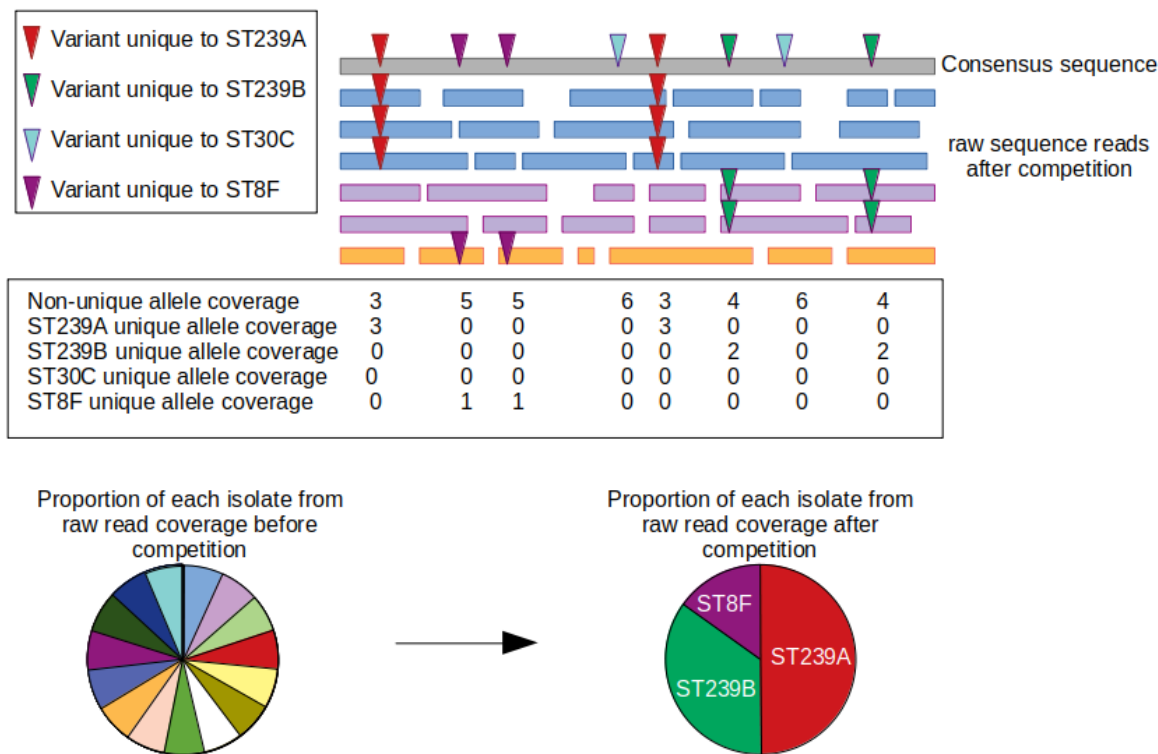


Figure 5.4. Simplified example schematic of how competitive ability was estimated from unique variant sites.

competition experiment (Figure 5.4) (Appendix B.1.11 – B.1.12). Any sites that were supported with a total of three reads or less, for both the variant allele and the consensus allele, were removed from the analysis (Section 2.2.6). Positions that had variant bases that were supported by three or less reads were recorded as having no variant bases, as this may be due to sequencing error. Therefore, only sites that had a variant coverage with four or more supporting reads were included as true variant sites (Appendix B.1.13 – B.1.14).

For each competition experiment, the number of supporting reads for each unique variant site was recorded. From this, the average coverage of all variants that were unique to each isolate was determined, before and after being exposed to the competition conditions for 24 hours, to

$$M = 4 \div \frac{N_z + N'_z}{V}$$

Figure 5.5. Calculation for minimum detection limit. M = minimum detection limit for isolate Z ; N_z = total number of reads in support of isolate Z at sites unique to isolate Z ; N'_z = total number of reads in support of non- Z isolates at sites unique to isolate Z ; V = number of variant sites that are unique to isolate Z .

determine how the proportion of each isolate changed during each competition. The limit of detection for each isolate was also calculated, as in Figure 5.5. Any isolate with an average coverage that was lower than the limit of detection was called at the limit of detection for that isolate. For each competition experiment, the ratio of the corrected average variant site coverage for all isolates summed to roughly 1, ranging from 0.93 – 1.25 (Appendix B.1.15). Graphs of the log difference between the proportion of each isolate in each growth condition compared to the proportion of each isolate in the initial mixture were produced in R v3.4.4. Statistical analysis was carried out with JMP® data analysis software (SAS Institute Inc., 2019).

2.4 Estimating competitive ability from unique genes

The number of reads supporting each base in the concatenated unique gene alignment was determined from the mapping of the raw sequence reads from each competition experiment. The maximum number of reads supporting a position for each gene was determined (Appendix B.1.16 – B.1.17). Any genes that were supported with a maximum number of reads less than four were recorded as being supported by zero reads, as this may be due to mapping error (Section 2.2.6). The average of the maximum coverage for each unique gene for each isolate was calculated. Any isolate that had an average maximum unique gene coverage of less than four was recorded at the minimum threshold of four reads.

Chapter 5

The maximum coverage of any unique variant site (Section 2.3) was 122 reads, therefore genes where the maximum coverage was over 122 reads were investigated. Genes that had a high coverage in both the original mixture and the competition mixture were likely to have been due to multi-copy genes, for example genes located on multi-copy plasmids or lysogenic phage.

Gene function for genes that could be annotated by Prokka were identified in UniProt (The UniProt Consortium, 2018), and MegaBLAST (Morgulis *et al.*, 2008) was used to identify genes where no annotation was found with Prokka. All graphs were produced in R v3.4.4.

3. Results

3.1 Creating a pipeline to calculate competitive ability

Fifteen ST239, ST8 and CC30 *S. aureus* isolates were used to develop a method to quantitatively measure the competitive ability of several bacterial isolates in co-culture, through whole genome sequencing.

3.1.1 Identification of unique variant sites

Variant sites that were unique to each of the 15 sequences were identified, to be used as markers to determine the relative competitive fitness of each isolate in competition.

The core gene sequences of each isolate (Chapter 2 Section 3.4.1) were compared, and variant sites that were unique to one isolate were recorded. As an internal control, the DNA sequences from each isolate were mapped to the core gene consensus sequence with an average depth of 374.3x, ranging from 98.0x – 487.8x. The average coverage was 99.88%, ranging from 99.86% – 99.91% (Appendix B.1.8).

ST	Number of unique variant sites	ST	Number of unique variant sites	ST	Number of unique variant sites
ST239A	34	ST8A	55	ST30A	222
ST239B	13	ST8B	3	ST30B	116
ST239C	35	ST8C	133	ST30C	57
ST239D	61	ST8D	156	ST30D	55
		ST8E	259	ST30E	82
		ST8F	184		

Table 5.1. Number of variant sites that were unique to each of the 15 isolates.

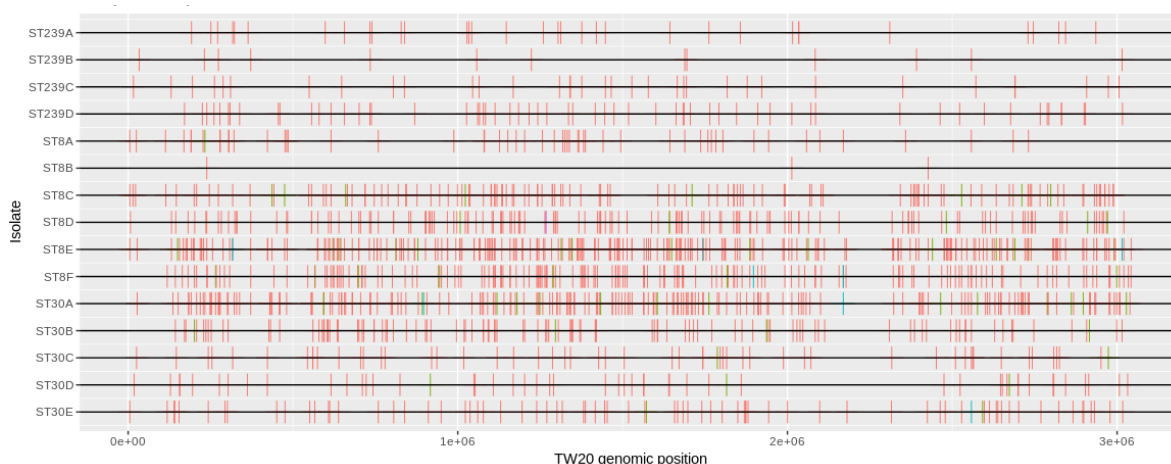


Figure 5.6. Genomic position of the unique variant sites compared to the ST239 TW20 reference genome. Red, one variant site per gene; green, two variant sites per gene; blue, three variant sites per gene; magenta, four variant sites per gene.

Of the 2,080 core genes that were shared between the sequences of all 15 isolates, 1,315 genes contained one variant site that was unique to a single isolate, 61 genes contained 2 variant sites that were unique to a single isolate, 8 genes contained 3 variant sites that were unique to a single isolate, and 1 gene contained 4 variant sites that were unique to a single isolate. A combined total of 1,465 variant sites were identified that were unique to one isolate and could be used to accurately identify each isolate genome, ranging from 3 variant sites that were unique to ST8B to 259 variant sites that were unique to ST8E (Table 5.1). These variant sites were well dispersed across the whole *S. aureus* TW20 reference genome length (Figure 5.6).

3.1.2 Identification of unique genes

Genes that were unique to each of the fifteen isolates were also identified as markers to determine the relative competitive fitness of each isolate in competition.

As an internal control, the DNA sequences from each isolate (Chapter 2) were mapped to the concatenated unique genes sequence with an average depth of 408.4x, ranging from 98.7x –

Isolate	Number of true unique genes	Isolate	Number of true unique genes	Isolate	Number of true unique genes
ST239A	7	ST8A	0	ST30A	2
ST239B	7	ST8B	0	ST30B	4
ST239C	0	ST8C	4	ST30C	0
ST239D	3	ST8D	0	ST30D	0
		ST8E	0	ST30E	12
		ST8F	25		

Table 5.2. Number of genes that were unique to each isolate.

539.7x. The average coverage was 25.1%, ranging from 17.00% – 32.00% (Appendix B.1.10).

From an initial list of 1,120 candidate unique genes, a final selection of 64 true unique genes (Appendix B.4.4), that were >300 bp in length and could be used to accurately identify each isolate genome, was selected (Table 5.2). These included several plasmid-linked genes, and genes that encoded uncharacterised hypothetical proteins, as well as phage-related genes and AMR genes. No unique genes could be identified within the ST239C, ST8A, ST8B, ST8D, ST8E, ST30C and ST30D genomes.

3.1.3 Competition experiments

To investigate how successful this method could be to determine the competitive ability of each isolate in a mixture, all fifteen isolates were combined, and underwent 24 hour competitions in oxacillin, quinolone, and trimethoprim/sulfamethoxazole. These antibiotics were selected as a high number of the isolates had clinical levels of resistance to the β -lactam oxacillin, trimethoprim/sulfamethoxazole combination, and the quinolone norfloxacin (Chapter 6, Section 3.2.1), and because none of the genes that confer resistance to these

antibiotics (*mecA*, *dfrG*, *dfrC*, *gyrA* and *parC*) (Chapter 6, Section 3.2.2) encode extracellular resistance enzymes that may be protective of susceptible isolates in co-culture (Frost, 2018). The competitions were also carried out in TSB, BHI and PS media, to be analysed in Chapter 6 Section 3.1.3.

The raw sequence reads from each competition were mapped to the core gene consensus sequence with an average depth of 184.6x, ranging from 142.4x – 211.4x. The average coverage was 99.97%, ranging from 99.92% – 99.98% (Appendix B.1.11). The raw read sequences from each competition were also mapped to the concatenated true unique genes sequence with an average depth of 110.7x, ranging from 80.8x – 145.4x. The average coverage was 80.5%, ranging from 67.0% – 87.0% (Appendix B.1.12).

3.1.4 Base calling quality control

To ensure the accuracy of the variant calls, several different cut-offs were investigated for the minimum number of reads that were required to accurately support the presence of the variant allele at each site. An “incorrect allele” was defined as a base that corresponds to neither the reference allele nor the unique variant allele, and is therefore likely to be the result of a sequencing error. This may underestimate the number of sequencing errors, as some reference alleles may be incorrectly called as the unique variant allele, and vice versa. Therefore, the minimum number of supporting reads required to call a variant was selected conservatively, as four reads. This cut-off value is also consistent with Illumina sequencing guidelines, where the default cut-off used to call SNP positions is three (Illumina Inc., 2009).

Number of supporting reads	Total number of variants called	Number of correct variants called	Percentage of correct variants called	Number of incorrect variants called	Percentage of incorrect variants called
1	510	98	19%	412	81%
1 – 2	631	160	25%	471	75%
1 – 3	700	217	31%	483	69%
2+	1,303	1,225	94%	78	6%
3+	1,181	1,163	98.5%	18	1.5%
4+	1,113	1,107	99.5%	6	0.5%

Table 5.3. Number of incorrect variants called for different minimum numbers of supporting reads, in the first replicate of the competition in BHI.

The number of incorrect bases that were called for each variant site from the first replicate of the competition in BHI are displayed in Table 5.3. A cut-off of four supporting reads resulted in just 0.5% of variants called incorrectly, compared to 69% of variants that were called incorrectly at sites where there were three or fewer supporting reads.

3.2 Estimating competitive ability from unique variant sites

The phenotypic resistance or susceptibility of each of the 15 *S. aureus* isolates to clinical levels of oxacillin, quinolone and trimethoprim/sulfamethoxazole was measured in Chapter 6, Section 3.2.1, which indicated the ability of each isolate to grow in co-culture in the presence of these antibiotics. Coverage of variant sites that were unique to each isolate should therefore be able to detect the growth of resistant isolates, and not be able to detect the growth of susceptible isolates, in each antibiotic condition.

3.2.1 Estimating competitive ability from unique variant sites in antibiotics

The coverage of each unique variant site was calculated from the raw sequence reads from each mixture, before and after each competition. All isolates could be detected above the minimum detection limit in the initial mixture prior to the competitions in antibiotics.

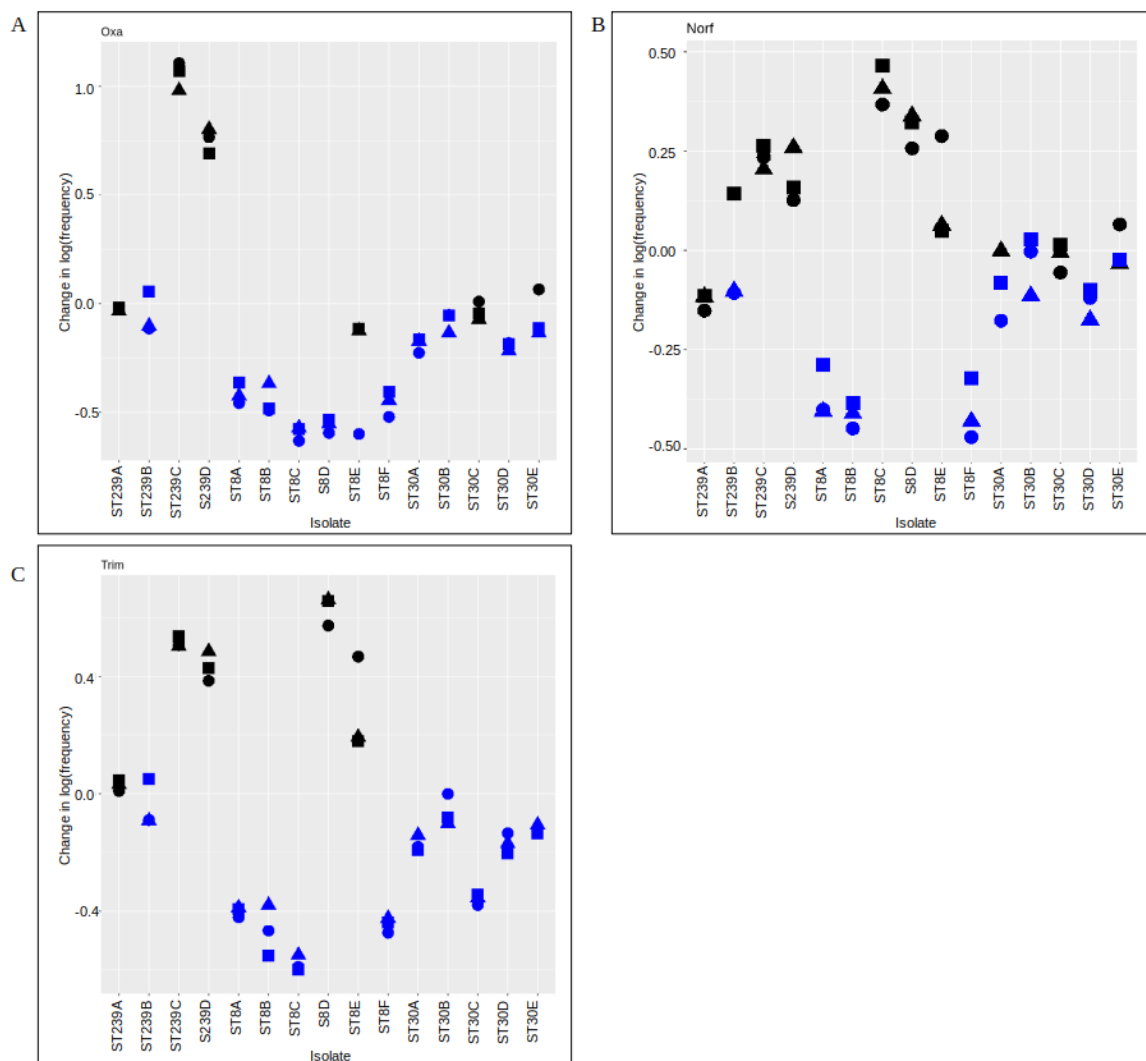


Figure 5.7. Competitive ability in a) 4 $\mu\text{g/mL}$ oxacillin b) 2 $\mu\text{g/mL}$ quinolone and c) 8/152 $\mu\text{g/mL}$ trimethoprim/sulfamethoxazole. Isolates where competitive ability was below the minimum detection limit are coloured in blue. Circle, replicate 1; triangle, replicate 2; square, replicate 3.

The log difference in the proportion of every isolate in competition in each antibiotic, compared to the initial mixture, was calculated (Appendix B.1.13) to assess the competitive ability of each isolate in these conditions (Figure 5.7). Where an isolate could not be detected above the minimum detection limit, the coverage for that isolate was recorded as the minimum detection limit. Therefore, the true competitive ability of the isolates that were only present at very low levels is likely to be much lower than this estimate. A high level of

similarity was observed in the three replicates of each competition. Therefore, the outcome of the competitions are consistent and reproducible.

ST239C had the greatest competitive ability in oxacillin, followed by ST239D. All other isolates either remained at a similar level, or decreased during the competition. In quinolone, several isolates had a strong competitive ability, including ST239C, ST239D, ST8C and ST8D. ST239C, ST239D and ST8D, along with ST8E, also had a strong competitive ability in trimethoprim/sulfamethoxazole.

	Oxacillin				Quinolone				Trim/Sulf			
Isolate	Phenotypic resistance	Replicate			Phenotypic resistance	Replicate			Phenotypic resistance	Replicate		
		1	2	3			1	2		3		1
ST239A	R	/	/	/	R	/	/	/	R	/	/	/
ST239B	R	N	N	N	R	N	N	N	R	N	N	N
ST239C	R	/	/	/	R	/	/	/	R	/	/	/
ST239D	R	/	/	/	R	/	/	/	R	/	/	/
ST8A	S	N	N	N	S	N	N	N	S	N	N	N
ST8B	S	N	N	N	S	N	N	N	S	N	N	N
ST8C	R	N	N	N	R	/	/	/	S	N	N	N
ST8D	R	N	N	N	R	/	/	/	R	/	/	/
ST8E	R	N	/	/	R	/	/	/	R	/	/	/
ST8F	S	N	N	N	S	N	N	N	S	N	N	N
ST30A	R	N	N	N	S	N	/	N	S	N	N	N
ST30B	S	N	N	N	S	N	N	N	S	N	N	N
ST30C	R	/	/	/	R	/	/	/	S	N	N	N
ST30D	R	N	N	N	R	N	N	N	S	N	N	N
ST30E	R	/	N	N	R	/	/	N	R	N	N	N

Table 5.4. Correlation between phenotypic resistance to oxacillin (oxa), quinolone (quin) and trimethoprim/sulfamethoxazole (trim) from Chapter 6, Section 3.2.1, and competitive ability. Orange (R), phenotypically resistant; green (S), phenotypically susceptible; orange (/), growth above the minimum threshold was detected in competition; white (N), no growth above the minimum threshold was detected in competition.

None of the isolates that were susceptible to the antibiotics tested in Chapter 6 Section 3.2.1 had a positive competitive ability. All of the isolates that were susceptible to the antibiotics could not be detected above the minimum detection limit after competition (Table 5.4), with the exception of a single replicate of ST30A in quinolone, where a very low level of ST30A was detected above the minimum threshold, but this was not reproduced in the replicates. Quinolone resistance can be conferred through a single point mutation in *gyrA* or *parC*, and therefore, it is possible that resistance arose in ST30A during this replicate of the competition, that enabled its growth in these conditions.

3.2.2 Estimating competitive ability from unique variant sites in growth-promoting media

The 15 *S. aureus* isolates also underwent competition in co-culture in the absence of antibiotics in TSB, BHI and PS media. The coverage of each unique variant site was calculated from the raw sequence reads from each mixture, before and after each competition, and the competitive ability of each isolate was calculated.

All isolates could be detected above the minimum detection limit in the initial mixture prior to the competition. The log difference in the proportion of every isolate in each growth condition compared to the initial mixture was calculated (Appendix B.1.14), to assess the competitive ability of each isolate in those conditions (Figure 5.8).

ST8F had the highest competitive ability in TSB, alongside ST30A, ST30B, ST8C and ST8D. High competitive ability was also attributed to these same five isolates after competition in BHI. In both TSB and BHI competitions, the levels of all four ST239 isolates decreased, with

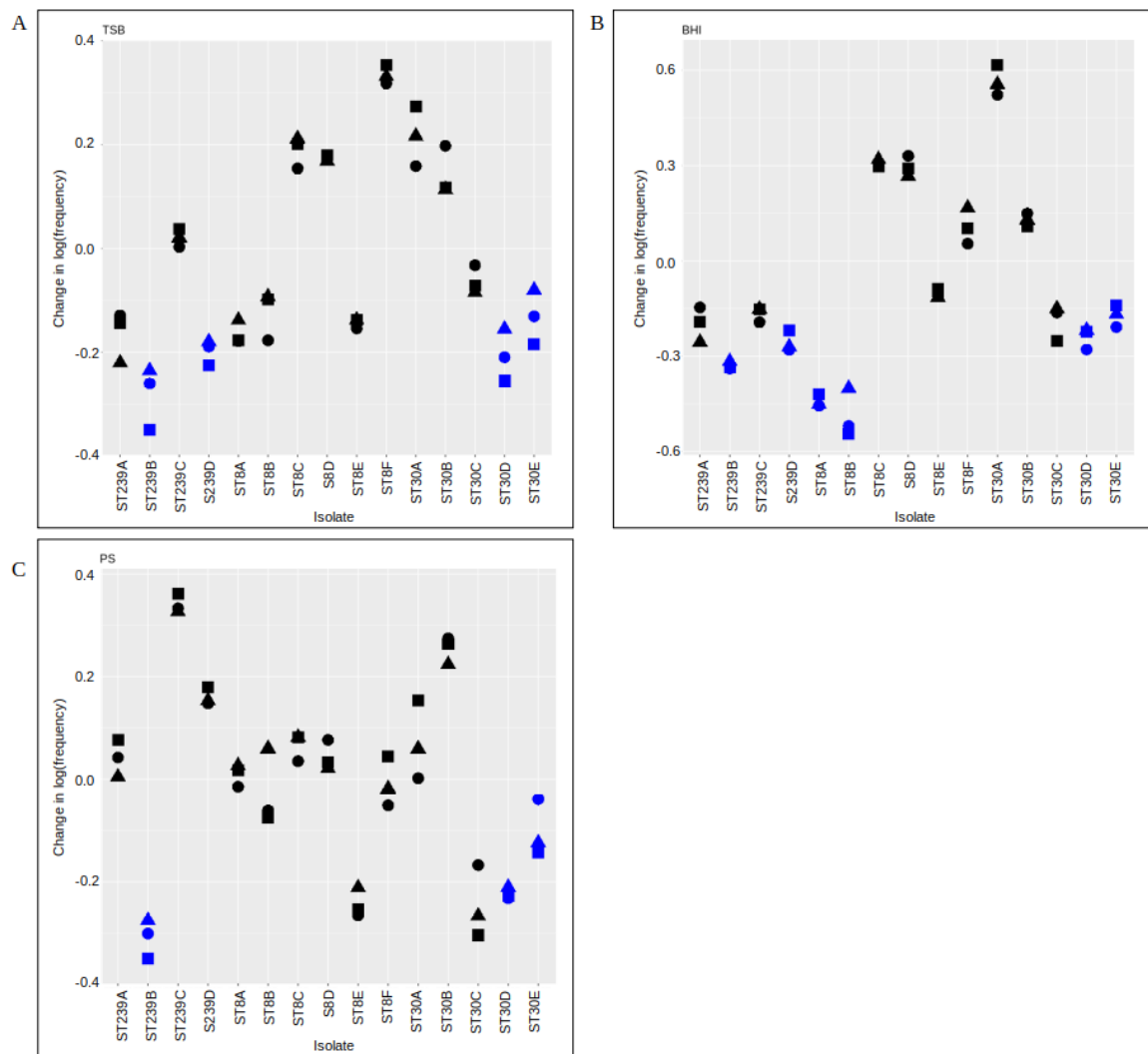


Figure 5.8. Competitive ability in a) TSB b) BHI and c) PS. Isolates where competitive ability was below the minimum detection limit are coloured in blue. Circle, replicate 1; triangle, replicate 2; square, replicate 3.

the exception of ST239C, where a slight increase was observed in TSB. In PS, ST239C had the greatest competitive ability, alongside ST239D and ST30B.

Several isolates were not detected above the minimum detection threshold, which shows that they grew particularly poorly in these competitions. ST239B, ST30E and ST30D were not recorded above the limit of detection in any media. ST239D was not detected above the limit

of detection in TSB or BHI, although it had a positive competitive ability in PS. ST8A and ST8B could not be detected above the limit of detection in BHI.

3.3 Estimating competitive ability from unique genes

A total of 64 unique genes were identified in the 15 *S. aureus* isolates in Section 2.2.2, although not all isolates contained any unique genes. Where possible, the unique gene coverage of each isolate was compared with the unique variant site coverage from Section 3.2.

3.3.1 Estimating competitive ability from unique genes in antibiotics

For the isolates where unique genes could be identified (Section 2.2.2), the maximum coverage of each unique gene was calculated from the raw sequence reads from each mixture, before and after each competition. The log difference in the proportion of every isolate in competition in each antibiotic, compared to the initial mixture, was calculated (Appendix B.1.16) to assess the competitive ability of each isolate in these conditions (Figure 5.9). Where an isolate could not be detected above the minimum detection limit, the coverage for that isolate was recorded as the minimum detection limit.

ST239D consistently had a high competitive ability in all three antibiotics, which is consistent with the unique variant analysis. In trimethoprim/sulfamethoxazole, ST239A also had a high competitive ability, but the remaining isolates could not be detected above the minimum threshold. This is also consistent with the unique variant analysis. Similarly, in oxacillin and norfloxacn, the isolates that had a coverage of below the minimum threshold of detection

corresponded with isolates that had a presence below the minimum threshold of detection in the unique variant sites analyses.

The average unique variant site coverage (Section 3.2.1) and the average maximum unique gene coverage were correlated ($R^2 = 0.2166$, $p = 0.01271$) (Figure 5.10).

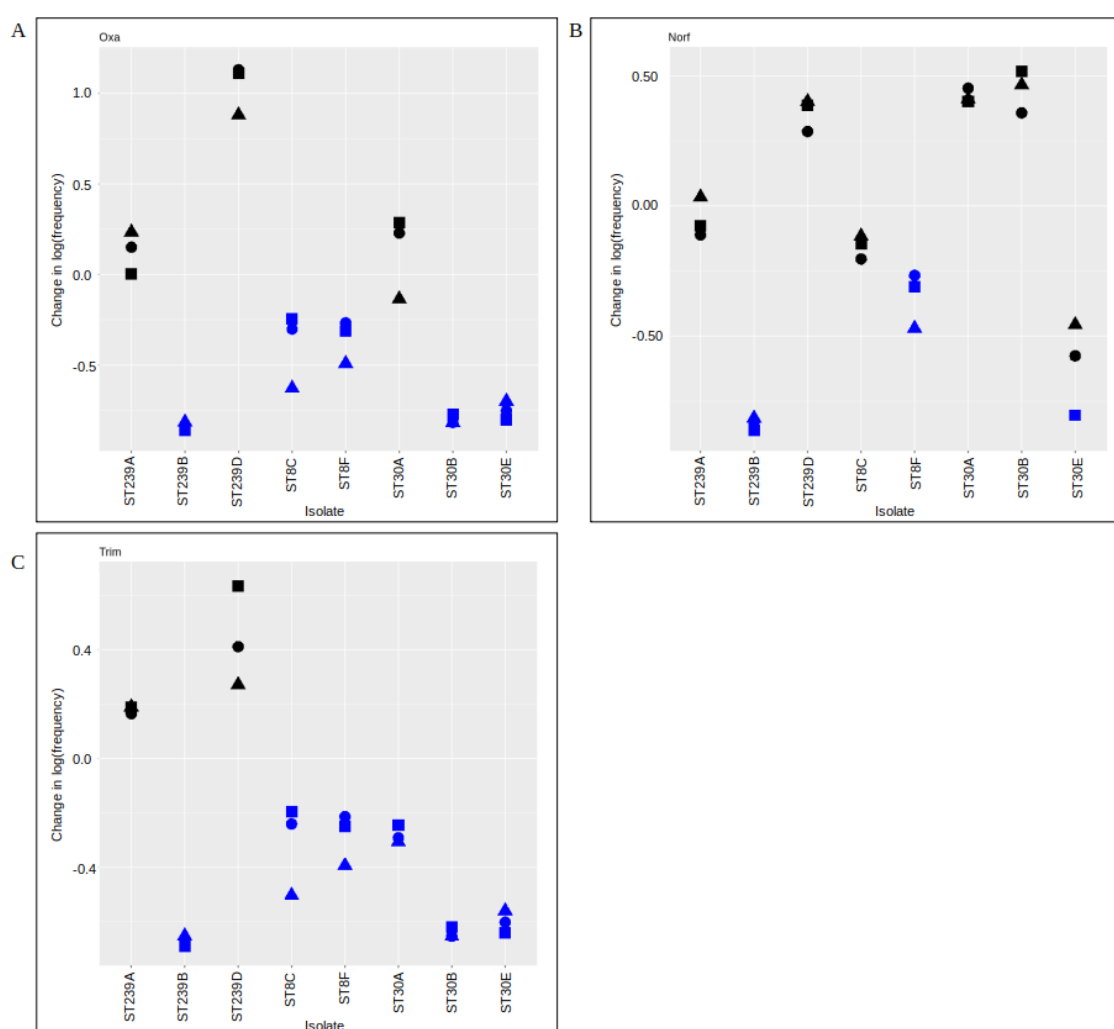


Figure 5.9. Competitive ability in a) 4 µg/mL oxacillin b) 2 µg/mL quinolone and c) 8/152 µg/mL trimethoprim/sulfamethoxazole. Isolates where competitive ability was below the minimum detection limit are coloured in blue. Circle, replicate 1; triangle, replicate 2; square, replicate 3.

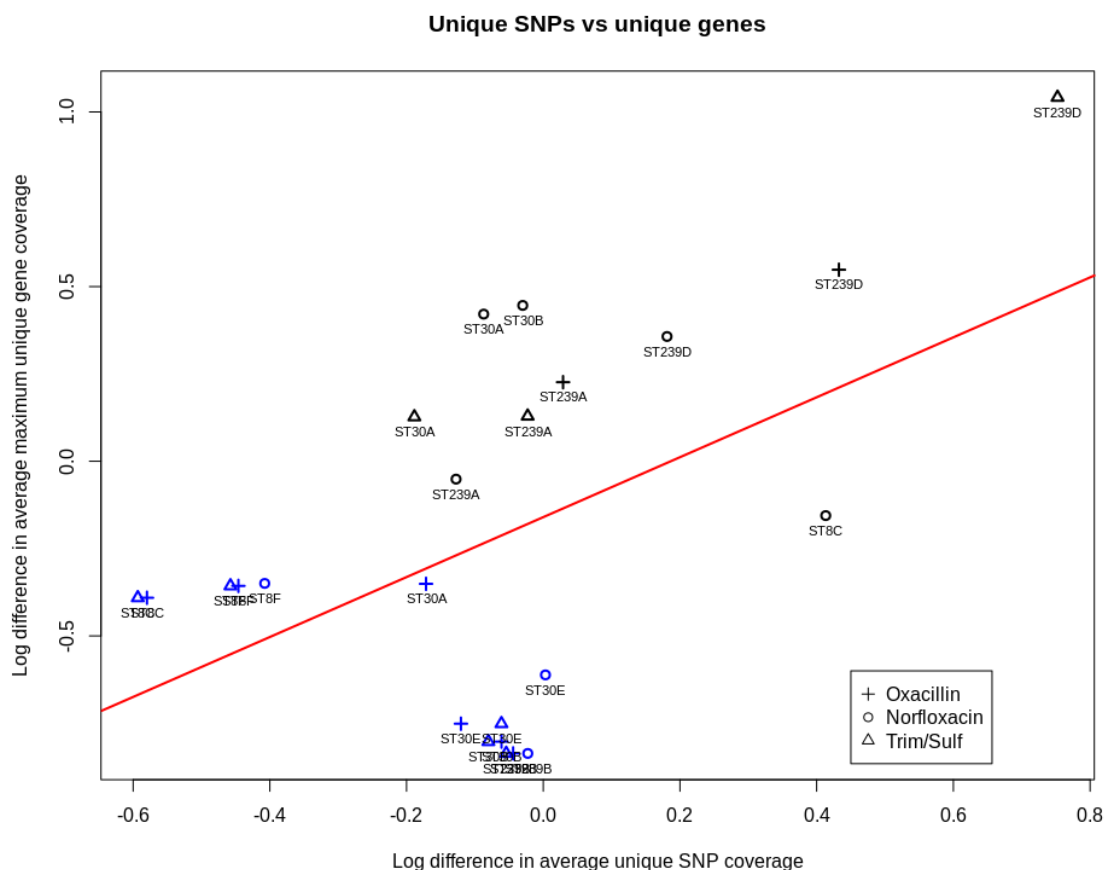


Figure 5.10. Average maximum unique gene coverage and average unique variant site coverage in oxacillin, quinolone and trimethoprim/sulfamethoxazole. Isolates where competitive ability was below the minimum detection limit are coloured in blue.

3.3.2 Estimating competitive ability from unique genes in growth-promoting media

For the isolates where unique genes could be identified (Section 2.2.2), the maximum coverage of each unique gene was calculated from the raw sequence reads from each mixture, before and after each competition. The log difference in the proportion of every isolate in competition in each media, compared to the initial mixture, was calculated (Appendix B.1.17) to assess the competitive ability of each isolate in these conditions (Figure 5.11). Where an isolate could not be detected above the minimum detection limit, the coverage for that isolate was recorded as the minimum detection limit.

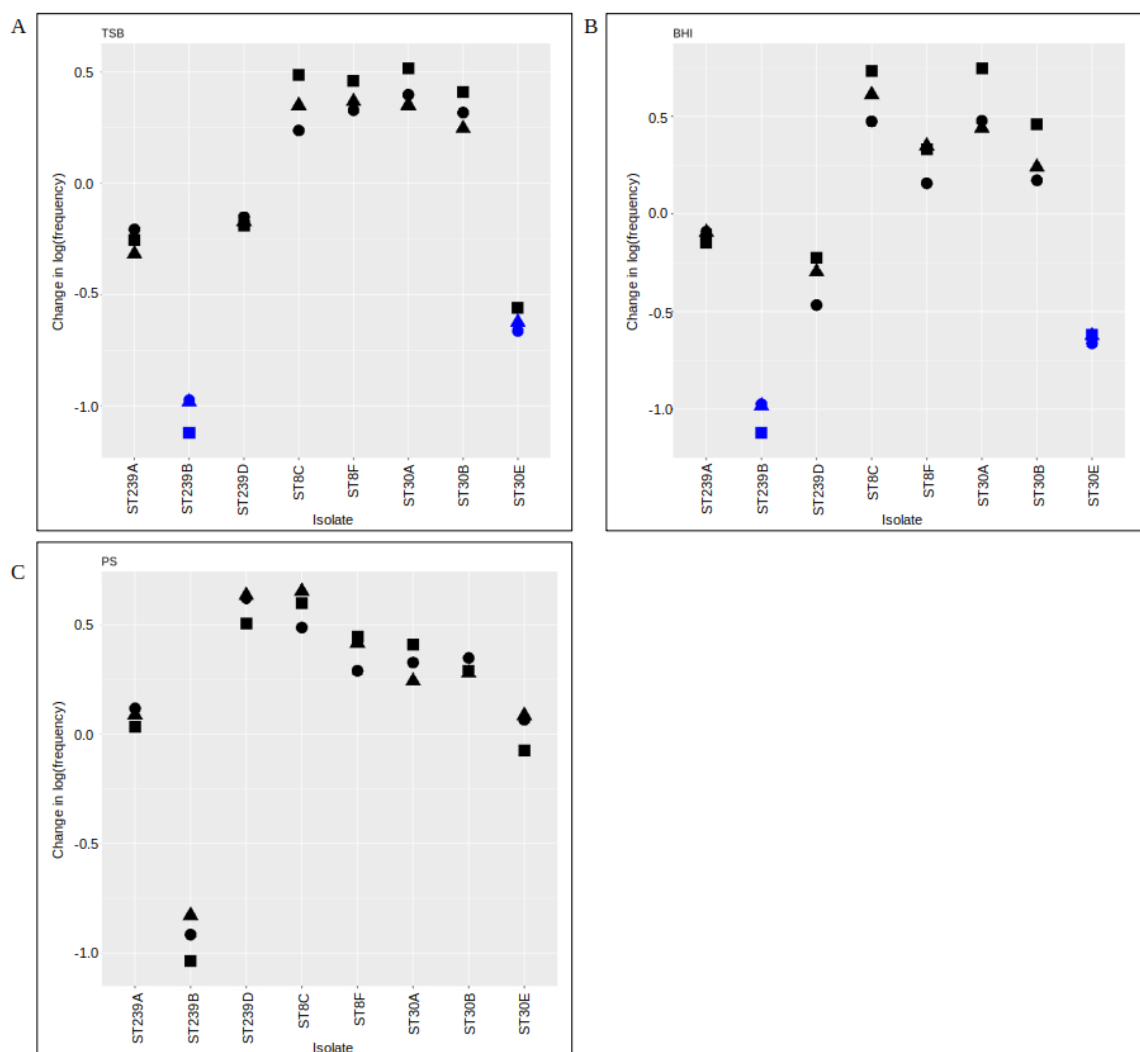


Figure 5.11. Competitive ability in a) TSB b) BHI and c) PS. Isolates where competitive ability was below the minimum detection limit are coloured in blue. Circle, replicate 1; triangle, replicate 2; square, replicate 3.

The genes that were unique to ST239B could not be detected above the minimum threshold in TSB or BHI, which is consistent with the results of the unique variant sites analysis. Low levels of ST239B were detected in PS, but this still was the isolate with the lowest competitive ability in the mixture. Similarly, ST30E could not be detected in TSB or BHI, with the exception of one replicate from the TSB competition, where a very low level of ST30E was detected above the minimum threshold. However, in PS, low levels of ST30E were detected, although it did not have a positive competitive ability.

Chapter 5

In all conditions, ST8C, ST8F, ST30A and ST30B had the highest competitive ability, alongside ST239D in PS. The highest average maximum coverage in all conditions was for the unique genes in ST239A, including in the initial mixture prior to the competition. This may suggest that the genes that were unique to ST239A are from a multi-copy plasmid, as ST239A was not determined by the unique variant sites analysis to be over-represented in the initial mixture prior to the competition.

There was a correlation ($R^2 = 0.7082$, $p = 1.56 \times 10^{-7}$) between the average maximum unique gene coverage and the average unique variant site coverage (Figure 5.12). Additionally, the

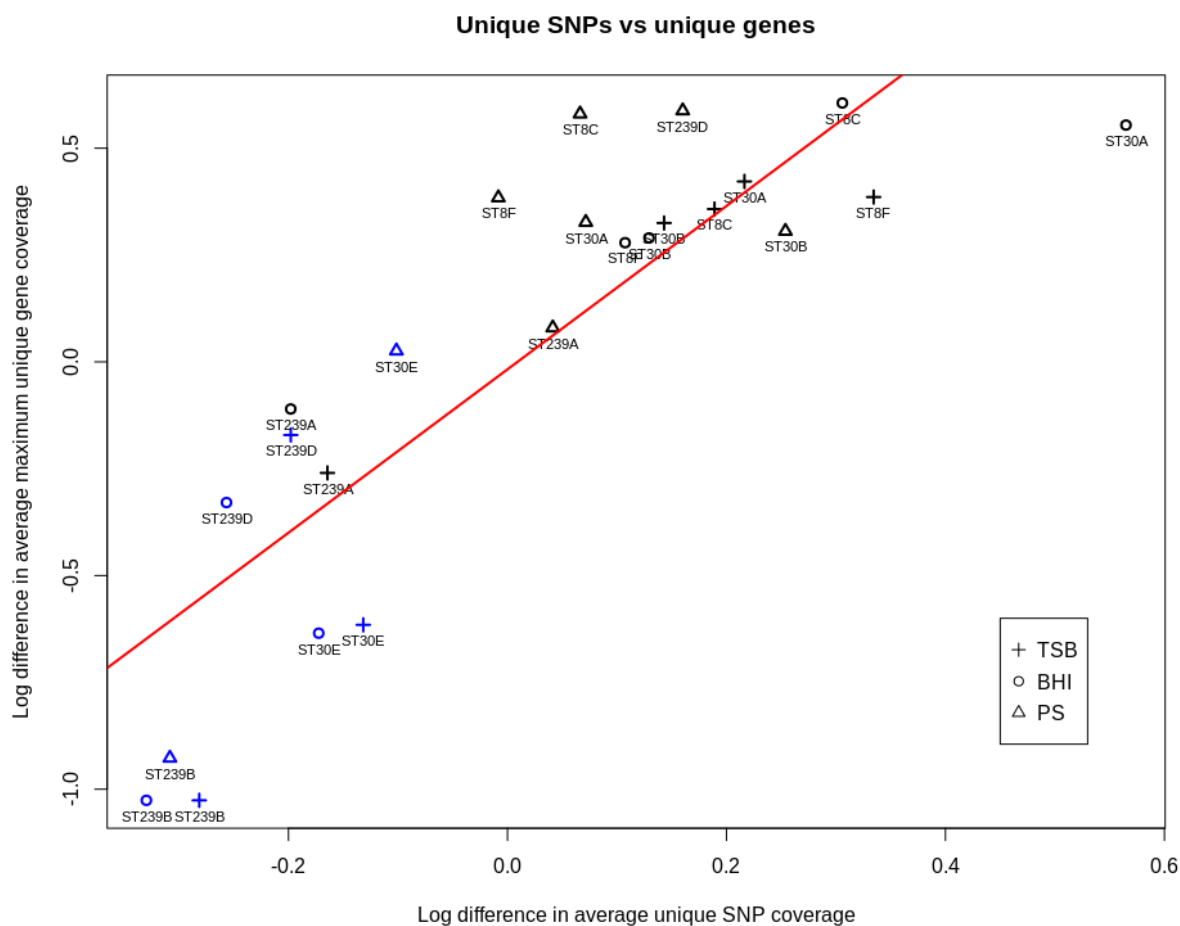


Figure 5.12. Average maximum unique gene coverage and the average variant sites coverage in TSB, BHI and PS. Isolates where competitive ability was below the minimum detection limit are coloured in blue.

unique genes that had a coverage of below the minimum threshold of detection corresponded to isolates that had a presence below the minimum threshold of detection in the unique variant sites analysis.

3.3.3 Over-represented unique genes

Some of the 15 *S. aureus* isolates used in this study were closely related, including several isolates from the same ST. Therefore, few genes that were unique to each isolate could be identified, and many of these were phage-related, or may be encoded within multi-copy plasmids. Unique genes that are linked to MGEs may have a higher or lower coverage than genes that are encoded within the chromosome, as the levels of DNA in the competition mixture that encode these genes are not constrained by vertical inheritance. To investigate whether the coverage of MGE-linked unique genes in the competition experiments was higher than expected, the unique genes with the highest maximum coverage were further examined.

The maximum coverage of any unique variant site was 122 reads (Section 2.3), therefore the gene function of the six genes with the highest maximum coverage (>122 reads) were further investigated (Table 5.5). Each gene was over-represented in more than one competition, and not always consistent between replicates from the same experiment. These genes were primarily associated with plasmids, which may suggest some increase in plasmid copy number during some competitions.

Additionally, in all three replicates of the competition experiments in BHI, a high coverage of a *lexA*-like gene was detected. This gene was unique to ST239A, however the six other ST239A-specific genes were not over-represented in the BHI competitions. Low levels of ST239A were detected in BHI, but ST239A had a negative competitive ability determined by

Isolate	Gene ID	Mixture	Maximum coverage	Gene function
ST239D	<i>cadC</i>	PS 2	133	Cadmium resistance transcriptional regulatory protein CadC (associated with plasmids in <i>S. aureus</i>)
ST239D		Trim 6	145	
ST239D		Oxa 4	286	
ST239D		Oxa 5	230	
ST239D		Oxa 6	287	
ST239D	<i>CIPDHBFC_2946</i>	PS 1	134	Hypothetical protein (associated with plasmids in <i>S. aureus</i>)
ST239D		PS 2	140	
ST239D		Oxa 4	240	
ST239D		Oxa 5	143	
ST239D		Oxa 6	233	
ST239D	<i>qacR</i>	Oxa 4	245	HTH-type transcriptional regulator QacR (associated with plasmids in <i>S. aureus</i>)
ST239D		Oxa 5	129	
ST239D		Oxa 6	206	
ST8C	<i>JGFFFGPO_2741</i>	PS 1	128	Hypothetical protein (associated with plasmids in <i>S. aureus</i>)
ST8C		BHI 3	137	
ST8C	<i>JGFFFGPO_2743</i>	BHI 3	140	Hypothetical protein (associated with plasmids in <i>S. aureus</i>)
ST8C		PS 2	171	
ST239A	<i>lexA</i>	BHI 1	151	LexA repressor (regulates SOS response)
ST239A		BHI 2	160	
ST239A		BHI 3	178	

Table 5.5. Gene function of the unique genes with the highest maximum coverage.

both unique gene and unique variant site analyses. Therefore, it is likely that this is an example of a phage-encoded *lexA* homologue (Fornelos *et al.*, 2018), that had lysed from its host and hence resulted in a high sequencing coverage. This is further evidence that unique gene coverage may not be a reliable indicator of competitive ability, compared to unique variant site coverage.

4. Discussion

In their natural environment, bacteria must compete with a whole multitude of factors, including a range of other bacterial competitors, as well as additional environmental factors (Stubbendieck and Straight, 2016). Therefore, studying single species in culture, or pairwise competition experiments, are far from an accurate representation of the natural bacterial environment. In this study, I was able to make comparisons between the competitive ability of fifteen *S. aureus* isolates in co-culture in a variety of different conditions, using a whole-genome sequencing approach. This approach will be used as a measure of bacterial fitness in Chapter 6.

In the initial experiments, all 15 *S. aureus* isolates were grown in co-culture in the presence of antibiotics, and only isolates that were resistant to these antibiotics were detected at the end of the competitions. Each competition experiment was highly reproducible, which allowed for differences in the relative competitive ability of the strains to be estimated.

For each condition, the relative competitive ability of the isolates was calculated using both unique variant site coverage and unique gene coverage. These two approaches were found to be correlated in all conditions. This is in spite of the fact that many of the unique genes were plasmid or phage-related, and may therefore have variable copy numbers. Several high-coverage plasmid-related unique genes, as well as a *lexA*-like gene that was potentially phage-related, were observed at the end of some competitions.

Chapter 5

Although unique variant sites were identified in the genomes of all 15 *S. aureus* isolates, ranging from 3 – 259 unique variant sites per isolate, unique genes could not be identified for all isolates. Unique genes were only identified in eight of the 15 isolates, ranging from 2 – 25 unique genes per isolate. The low number of unique genes identified in this collection of isolates, combined with the high number of MGE-associated unique genes that were identified, suggests that unique gene coverage may be a less accurate measure of competitive ability than unique variant site coverage in this study. Unique gene coverage may be a more useful measure of competitive ability between more diverse isolates that have larger accessory genomes, such as between different species or genus. However, as the isolates get more diverse, their shared core genome size decreases, and therefore the number of unique variant sites would likely also decrease. Therefore, the usefulness of unique variant sites or unique genes to calculate competitive ability depends on the similarity between the isolates.

A unique variant coverage with a conservative cut-off of four supporting reads per site was used as a measure of competitive ability, as this produced a low number of false positive results, with just 0.5% of variant sites called incorrectly. Although this conservative cut-off value was selected to reduce the rate of false positive variant calls, it also increases the risk of false negatives, and therefore decreases the power to detect low levels of each of the isolates, which is a major limitation of this study. Sequencing technology with a greater sequencing depth than the 184.6x depth that was the average in this study, would increase the power to detect these low level isolates.

There are several benefits of the technique developed in this chapter, compared to other more common techniques to measure the proportion of isolates in competition, such as qPCR. For

example, depending on the genetic variability between isolates, this method can be used on a much larger scale than qPCR, where there is a limit to how many different fluorescent probes it would be feasible to use in a single competition experiment. Additionally, the method developed in this chapter can be used to quantify a mixture of very closely related isolates, by measuring the ratio of coverage of unique variant sites from across the whole genome. For example, in the mixed isolates in this Chapter, very few unique genes were identified, due to the close genetic relationship between isolates, which would hinder the use of techniques such as qPCR. Finally, this technique is much less laborious, as it does not involve a large number of qPCR assays, and does not require the addition of genetic labels to isolates prior to competition.

Recent advances in the field of genomics have revolutionized the study of microbial evolution, for example to facilitate the study of tens of thousands of generations in large-scale evolution experiments (Good *et al.*, 2017), and to successfully identify resistance determinants in large-scale genome-wide association studies (Earle *et al.*, 2016). However, further developments are also necessary to decrease the cost, laboriousness and error of estimating relative fitness from large scale microbial phenomics experiments (Zackrisson *et al.*, 2016). As DNA sequencing becomes cheaper, faster and more readily available, the method developed in this chapter, which uses unique variant site coverage and unique gene coverage to quantify competitive ability, may be used for larger scale competition and phenotyping assays. Based on their genome sequences, isolates should be selected for mixing to optimise the number of unique variant sites and unique genes per isolate, and high-coverage DNA sequencing should be carried out, in order to address some of the limitations of this method.

Chapter 6

Estimating the fitness effects of the ST239 hybridisation

Abstract

HGT and AMR often impose fitness costs on bacteria. In its recent evolutionary history, ST239 has undergone two large HGT events; a large chromosomal replacement event where ~20% of the ST8 genome was replaced by a homologous region of the ST30 genome, and the acquisition of a large ~60 kb *SCCmec*-III element that contains a high number of AMR genes. Despite the fitness costs that may have been imposed due to these extreme genetic transfer events, ST239 has been a highly prevalent and successful MRSA strain in healthcare settings around the globe. A total of 15 ST239, ST8 and CC30 isolates were investigated for genetic and phenotypic AMR, and their relative fitness was measured through growth rate and competitive ability experiments across three different culture media with different growth-promoting properties. Phenotypic and genotypic AMR were well-correlated for the majority of antibiotics, and the ST239 isolates had greater levels of AMR than the ST8 and CC30 isolates. Competitive ability and maximum growth rate were not found to be correlated, suggesting that these two fitness measures are independent, and measure different aspects of fitness. The ST239 isolates had overall lower fitness compared to both ST8 and CC30. This may indicate that although ST239 has become an extremely successful global EMRSA strain, the recent hybridisation event that formed ST239 has had an overall negative impact on its fitness. This suggests that large HGT events that occur between bacteria in nature can have measurable negative impacts on fitness, which, in the case of ST239, can continue to burden hybrid bacteria many years after the HGT event.

1. Introduction

1.1 Aims and objectives

In this chapter, I aim to compare the relative fitness and levels of AMR of a collection of 15 ST239, ST8 and CC30 isolates, to investigate whether the large chromosomal replacement event that formed ST239 is associated with a measurable fitness cost.

1.2 Fitness effects of HGT and AMR

In microbes, HGT facilitates the acquisition of genetic regions of DNA that can lead to a range of different physiological effects, including novel phenotypes. HGT may incur positive or negative fitness effects on the host, for example horizontally transferred MGEs can act as parasites or mutualists (Rankin *et al.*, 2011). Common methods of quantifying bacterial fitness *in vitro* include measuring the maximum growth rate (V_{max}) of a bacterial culture growing on its own, quantification of the minimum inhibitory concentration (MIC) of a stressful compound such as an antibiotic, or a competition assay of one strain against another to measure relative fitness (Wiser and Lenski, 2015).

HGT is a common mechanism of acquisition of AMR phenotypes in susceptible bacteria, for example MGEs such as plasmids can carry AMR genes but may also carry additional genes with functions unrelated to AMR. Alternatively, AMR can arise via chromosomal mutations. Because antibiotics typically target important cellular functions (e.g. cell wall synthesis, regulation of chromosome super-coiling, RNA transcription and protein synthesis), AMR often carries a biological fitness cost that reduces the ability of the resistant lineage to survive and reproduce (Andersson and Hughes, 2010). The fitness costs of AMR are highly variable,

and can result in reduced growth rate, competitive ability, or virulence (Vogwill and MacLean, 2015).

The drivers behind the acquisition of AMR phenotypes through different evolutionary routes are complex. However, evolving AMR through the acquisition of MGEs such as plasmids, has been shown to carry a lower fitness cost than evolving resistance via *de novo* mutation (Vogwill and MacLean, 2015). Additionally, the size of horizontally transferred genetic regions is not necessarily correlated with fitness cost, however, few studies have focused on estimation of the fitness cost of acquiring non-plasmid MGEs.

Compensatory mutations can counterbalance the decreased fitness incurred from the acquisition of MGEs and AMR. This compensatory adaptation can occur rapidly in serial passage experiments, resulting in resistant bacterial populations that are as fit as susceptible clones in the absence of antibiotics, and has also been observed in resistant/susceptible isogenic clinical isolates (Andersson and Hughes, 2010).

1.3 Fitness of ST239

During the early 1990s – late 2000s, it was estimated that ST239 was responsible for up to 90% of hospital-acquired MRSA infections within a region accounting for over 60% of the world's population (Feil *et al.*, 2008). This suggests that, despite its hybrid origin, ST239 was a relatively fit bacteria in healthcare settings at this time, as it was able to out-compete many well-established healthcare-associated strains to become the dominant MRSA in this niche (Ko *et al.*, 2005). Over the past decade, several healthcare-associated MRSA strains have been observed to have out-competed ST239 in clinical settings, including ST22 (Hsu *et al.*,

2015), ST228 and ST5 (Conceicao *et al.*, 2007) (Chapter 2). The decline of ST239 may be related to the emergence of competitors with higher fitness in changing conditions.

The replacement of historical dominant healthcare-associated multi-drug resistant MRSA strains has been driven by similar pressures. For example, the disappearance of the first international multi-drug resistant MRSA from Denmark, which was a phage complex 83A strain, has been attributed to this strain being out-competed by fitter strains after a reduction in antibiotic usage (Nielsen *et al.*, 2012).

1.4 Culturing *S. aureus* in growth-promoting media

In order to study infectious disease agents, *in vivo* studies can provide the closest model of the natural pathogen environment. However, alternative methods, such as *in vitro* techniques and mathematical and computer modeling, should always be used in place or in support of *in vivo* experiments whenever possible (Colby *et al.*, 2017). One obstacle of culturing microbes *in vitro* can be choice of growth conditions, including growth medium, which can have significant effects on bacterial culturability (Davis *et al.*, 2005).

Tryptic soy broth (TSB) is a complex media, which includes enzymatically-digested plant and dairy-sourced proteins and provides nitrogen, vitamins, minerals, and sugars to promote the growth of micro-organisms, and is particularly recommended for isolation of microbes from blood, including *S. aureus* (Sigma-Aldrich, 2018). Brain heart infusion (BHI) is also a highly nutritious growth medium, intended for the culture of fastidious micro-organisms, including *Staphylococcus*. The nutrients, such as carbon, nitrogen, amino acids, essential growth factors and vitamins, are sourced from an infusion of bovine brain and heart (Thermo Fisher

Scientific, 2001), which may be more representative of the nutrients available to clinical *S. aureus*.

Porcine models of *S. aureus* sepsis are strong candidates for large animal models of human *S. aureus* infection (Soerensen *et al.*, 2013). *S. aureus* can cause infection in pigs, and there is evidence for the exchange of commensal *S. aureus* between pigs and pig farmers, suggesting that both humans and pigs can host similar *S. aureus* strains (Armand-Lefever *et al.*, 2005). Growth media containing serum mimics *in vivo* conditions in both pH and protein concentration, and therefore, porcine serum (PS) can be a good model for *S. aureus* growth experiments. Serum contains peptides, toxins, sugars, complement factors, and other cellular elements that can be cytotoxic (Thermo Fisher Scientific, 2002), and serum components such as complement factors can erratically inhibit bacterial growth (Zeitlinger *et al.*, 2011). The limits of using a PS model for *S. aureus* sepsis include that serum does not contain blood cells, has no adaptive immunity, and the presence of different antimicrobial peptides due to differences in innate immunity between pigs and humans (Wang, 2014).

1.5 AMR prediction

Antimicrobial susceptibility testing is an essential microbiological technique that can help clinicians to make informed decisions about appropriate antimicrobial prescribing. Standard antimicrobial susceptibility tests range from simple disk diffusion methods to automated systems that can rapidly measure MICs for large sample sizes of bacteria (Reller *et al.*, 2009). However, there are several drawbacks to these methods that can include time consuming and impractical bacterial culture requirements, alongside the use of large amounts of consumables including expensive antibiotics.

As of 2019, there were at least 47 freely accessible bioinformatics resources for the detection of AMR determinants from genetic sequencing data (Hendriksen *et al.*, 2019). Using these bioinformatics resources, several studies have been able to accurately identify AMR determinants associated with phenotypic resistance above the clinical break-point for several antimicrobials, with over 87% sensitivity (true positive) and over 98% specificity (true negative) (Hendriksen *et al.*, 2019). Bioinformatics approaches to AMR prediction may substantially decrease the amount of time it takes to determine the antimicrobial susceptibility of a bacterial infection, leading to improved treatment. For example, Bradley (2015) estimated that a sequencing-based analysis using Mykrobe, a web-based *S. aureus* AMR prediction tool that was designed for use by healthcare workers with no bioinformatics experience, could

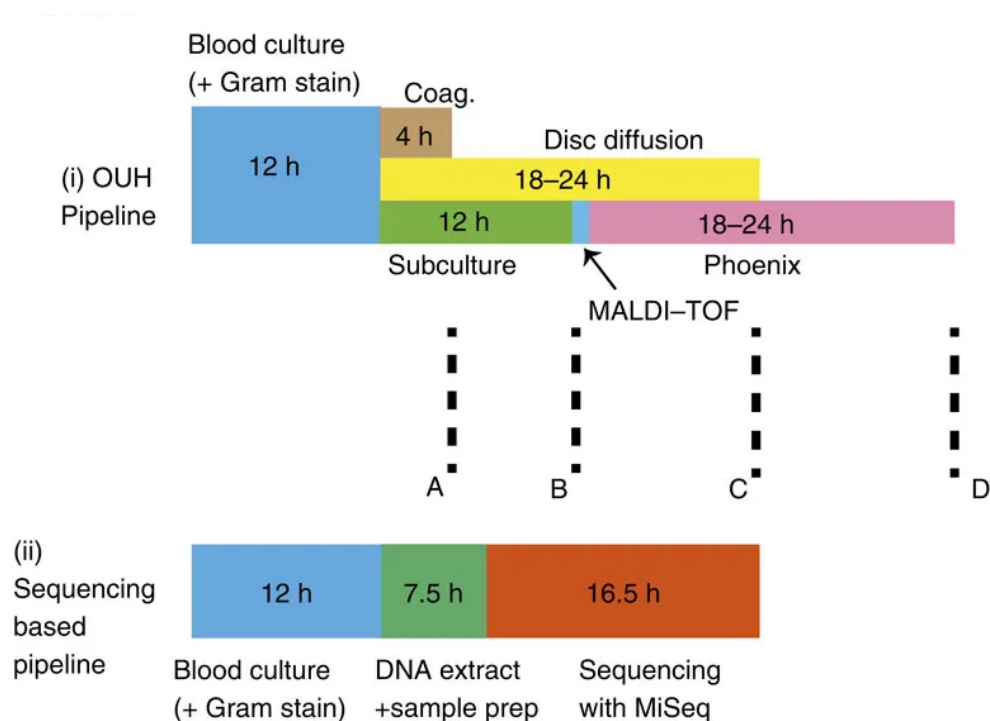


Figure 6.1. Workflow to predict AMR from *S. aureus* isolates using i) a culture-based approach at Oxford University Hospitals and ii) an estimated sequencing-based pipeline using Mykrobe. A = presumptive identification of species; B = species confirmation; C = first-line susceptibility to five antimicrobials; D = extended susceptibility. Image from Bradley *et al.*, 2015.

provide a full antimicrobial susceptibility report up to 12 hours before a standard culture-based approach (Figure 6.1) (Bradely *et al.*, 2015).

Further advances in next generation sequencing platforms are now able to sequence whole bacterial genomes much faster, and continue to improve (Quainoo *et al.*, 2017). Costs of DNA sequencing have been decreasing rapidly over the past decade (National Human Genome Research Institute, 2019), and culture-independent sequencing platforms, such as Oxford Nanopore, are now able to provide AMR susceptibility predictions within hours (Charalampous *et al.*, 2019). Therefore, it is likely that sequencing-based workflows in clinical settings could become even faster and more practical.

1.6 Chapter summary

To compare the relative fitness of all 15 ST239, ST8 and CC30 isolates, growth rate and competitive ability were measured in three different media with different growth-promoting properties; TSB, BHI and PS. Competitive ability was calculated using whole-genome sequencing analysis of co-cultures of all 15 isolates in Chapter 5. Growth rate of the isolates was measured in individual growth conditions.

Competitive ability and growth rate were not found to be correlated, suggesting that these two measurements are independent, and measure different aspects of fitness. Overall, the growth rate of the ST239 isolates was lower than the ST8 isolates, and not significantly different than the CC30 isolates. The competitive ability of the ST239 isolates was also lower overall than both the ST8 and CC30 isolates.

The phenotypic AMR profile for each of the 15 ST239, ST8 and CC30 isolates, was determined using a panel of 17 clinically relevant antibiotics from 13 antibiotic families. The ST239 isolates were phenotypically resistant to more antibiotic families than the ST8 and CC30 isolates, and the ST239 genomes also contained a greater number of AMR genes, which encode resistance to a greater number of antibiotic families. Phenotypic and genotypic resistance were well-correlated for the majority of antibiotic resistances.

There was no correlation between the fitness and AMR measures, suggesting that AMR is not the main driver of the fitness differences between these isolates. Therefore, the lower overall fitness of the ST239 isolates is more likely to have been driven by the large genetic transfer events in its recent evolutionary history.

2. Methods

2.1 Bacterial cultures

Unless otherwise stated, cryostocks of fifteen isolates from Chapter 2 Section 2.6.1 were streaked on TSA, and incubated at 37°C for 24 hours. Single colonies were incubated for 24 hours in 3 mL TSB at 37°C with 225 RPM shaking.

2.2 Growth rate

All isolates were grown from cryostock as in Section 2.1. 1 mL of each culture was pelleted and washed three times in PBS, then diluted 50x in either TSB, BHI or PS. 100 µL of diluted culture was added to each well of a 96 well plate, in triplicate, and incubated at 37°C with 225 RPM shaking. OD₅₉₅ was measured every 20 min over 24 hours, or until growth rate had reached plateau. This was repeated five times for each isolate.

Growth curves were generated and analysed with Growthcurver (Sprouffske and Wagner, 2016). All plots were generated in R v3.4.4.

2.3 Competitive ability

Competitive ability of the 15 isolates in TSB, BHI and PS media was estimated in Chapter 5.

2.4 Modelling effects

Statistical analysis was carried out in JMP® data analysis software (SAS Institute Inc., 2019). The statistical model used was a standard least squares model, with emphasis on effect leverage. Variables were either growth rate or competitive ability, and both nested and crossed

effects were considered, as described in the resulting ANOVA tables. All plots were generated in R v3.4.4.

2.5 Phenotypic AMR profiles

Each isolate was grown as in Section 2.1, using MH2 media. Cultures were diluted 100x in MH2 broth (to a density of $\sim 1 \times 10^6$ CFUs/mL), according to the MicroNaut evaluation protocol for MicroNaut-S MRSA/GP (MERLIN Diagnostika GmbH, 2018). 100 μ L was added to each well of a MicroNaut-S MRSA/GP plate and incubated for 18 hours at 37°C with 225 RPM shaking, after which OD₅₉₅ was measured. AMR breakpoints were assessed according to EUCAST v4.0 (The European Committee on Antimicrobial Susceptibility Testing, 2014). This was repeated in triplicate for each isolate.

2.6 Genetic AMR profiles

AMR genes were identified in each of the 15 sequences generated Chapter 2 Section 2.6.2, using ARIBA v3 (Hunt *et al.*, 2017) with the CARD v3.0.7 reference database (Jia *et al.*, 2017). Locations of the genes compared to the reference ST239 TW20 genome (accession number FN433596) were visualised in Artemis Comparison Tool v13.0.0 (Carver *et al.*, 2005).

3. Results

3.1 Fitness

3.1.1 Growth rate

To test the hypothesis that ST239 has lower fitness due to the hybridisation and increased AMR, fitness tests were carried out on all 15 ST239, ST8 and ST30 isolates in three media with different growth-promoting properties; TSB, BHI and PS.

As an estimate of the fitness of each of the isolates, the maximum growth rate of five replicates from all fifteen isolates was calculated from their growth curves in TSB, BHI and PS (Figure 6.2, Appendix B.6.1). The maximum growth rate ranged from 0.39 – 1.55 mODh⁻¹ in TSB, 0.37 – 1.03 mODh⁻¹ in BHI, and 0.00 – 0.44 mODh⁻¹ in PS.

The maximum OD was also calculated from the growth curves of five replicates from all fifteen isolates in TSB, BHI and PS (Appendix B.6.1; Appendix A.3.1). The maximum OD ranged from 0.39 – 1.55 in TSB, 0.69 – 1.04 in BHI, and 0.00 – 0.27 in PS. A positive correlation was observed between the maximum growth rate and maximum OD in TSB ($R^2 = 0.556$, $p = 0.000859$) and PS ($R^2 = 0.363$, $p = 0.0103$), but there was no significant correlation in BHI ($p = 0.719$) (Appendix A.3.2).

Several of the ST8 isolates grew at a particularly fast rate in TSB; ST8C, ST8D, ST8E and ST8F all had high maximum growth rates, and grew to a high OD. ST8A and ST8F had high growth rates in PS, and ST8A and ST8B had high growth rates in BHI. In BHI, ST8F had the lowest average maximum growth rate of all isolates, and also had a low maximum OD.

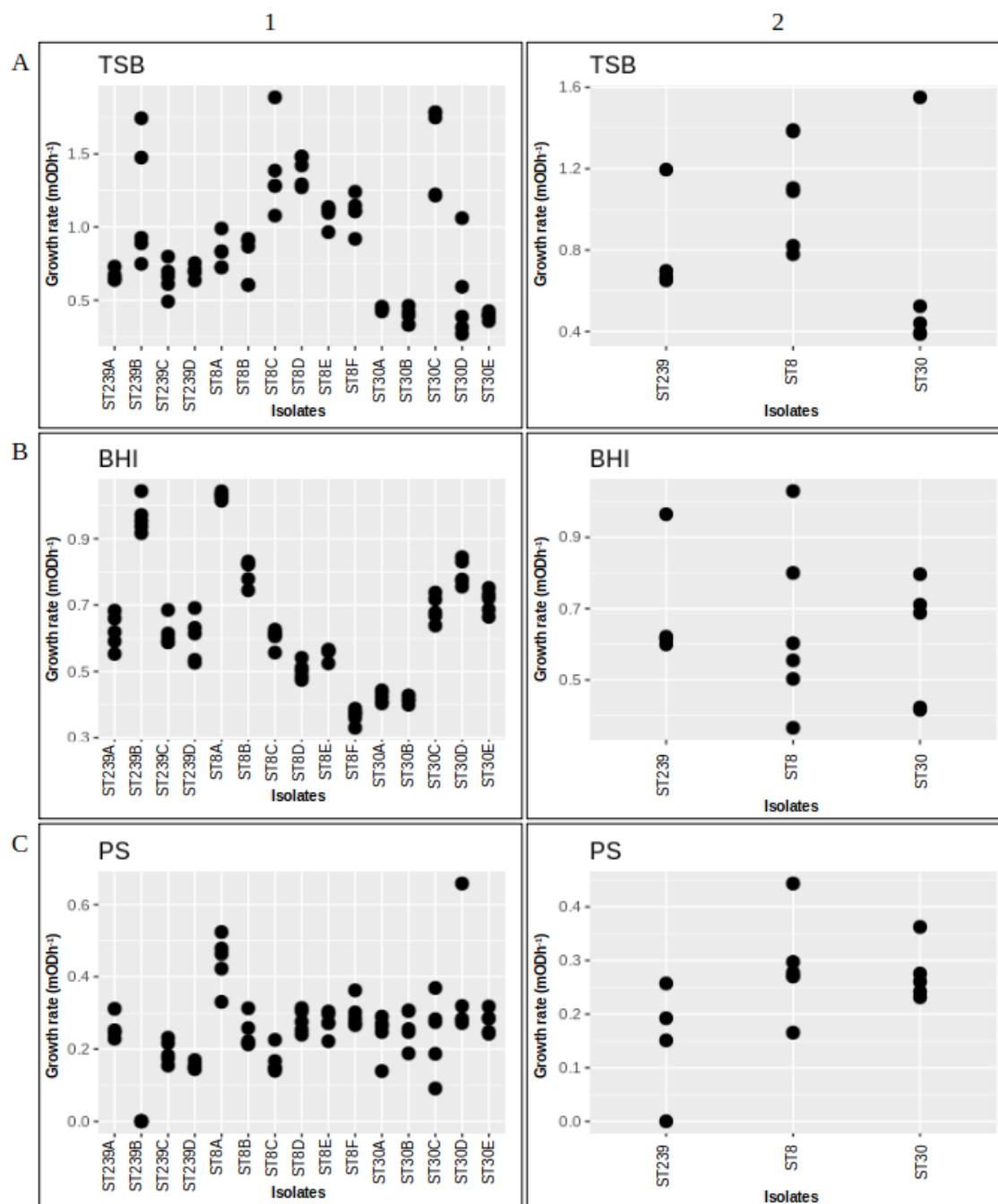


Figure 6.2. Maximum growth rate vs isolate in a) TSB, b) BHI, c) PS. 1) Each point represents one repeat for each isolate and 2) each point represents the average of all repeats for each isolate.

The growth of several of the CC30 isolates was particularly poor in TSB; ST30A, ST30B, ST30D and ST30E had the lowest average maximum growth rates and the lowest average maximum OD of all isolates. These isolates also had a low maximum growth rate and maximum OD in BHI. However, ST30C had the highest average maximum growth rate in

Chapter 6

TSB, and also reached a high maximum OD, and had the highest maximum OD in BHI. Conversely, ST30C reached a low maximum OD in PS, whereas ST30A, ST30B and ST30D had the highest maximum OD of all isolates in PS.

In PS, ST239B was the only isolate where no detectable growth was observed. However, this isolate had a high growth rate and maximum OD in both TSB and BHI. ST239D also grew poorly in PS, with both a low maximum growth rate and a low maximum OD. ST239A reached the highest average maximum OD in TSB.

Overall, for each isolate, growth ability in a particular medium was not a strong predictor of growth rate in the other conditions.

3.1.2 Modelling the effects on growth rate

To test the hypothesis that ST239 has lower fitness due to the hybridisation and increased AMR, the growth rate data from all 15 isolates was fitted to a statistical model. Initially, a full model was constructed (Appendix B.7.1) and after removing non-significant effects, yielded a reduced model table (Table 6.1). ST, media, media by ST interactions, and isolate were found

Effect	df	Sum of squares	Mean squares	F-value	p-value
ST	2	1.163551	0.581776	11.8932	0.00001
Media	2	14.238509	7.119255	145.5380	0.00000
Media x ST	4	1.19894	0.479973	9.8120	0.00000
Isolate[ST]	12	3.133390	0.261116	5.3380	0.00000
Error	204	9.979031	0.04892	-	-

Table 6.1. Reduced ANOVA table of significant growth rate effects.

to be significant effects (Appendix B.7.2 – B.7.4). Plate number was not found to be significant ($p > 0.05$), and was therefore not included in the reduced model.

Overall, the average growth rate of the ST8 isolates ($0.6507 - 0.6973 \text{ mODh}^{-1}$) was higher than the ST239 and CC30 isolates ($0.5194 - 0.5766$ and $0.4885 - 0.5395 \text{ mODh}^{-1}$, respectively) (Figure 6.3). However, there was no significant difference between the ST239 and CC30 growth rates. This suggests that ST239 has a lower fitness than ST8, even though the two strains share ~80% of their genome with a recent common ancestor. It is not clear whether this fitness cost is caused by the presence of the horizontally acquired *SCCmec*-III element and acquired region, or additional factors affecting ST239.

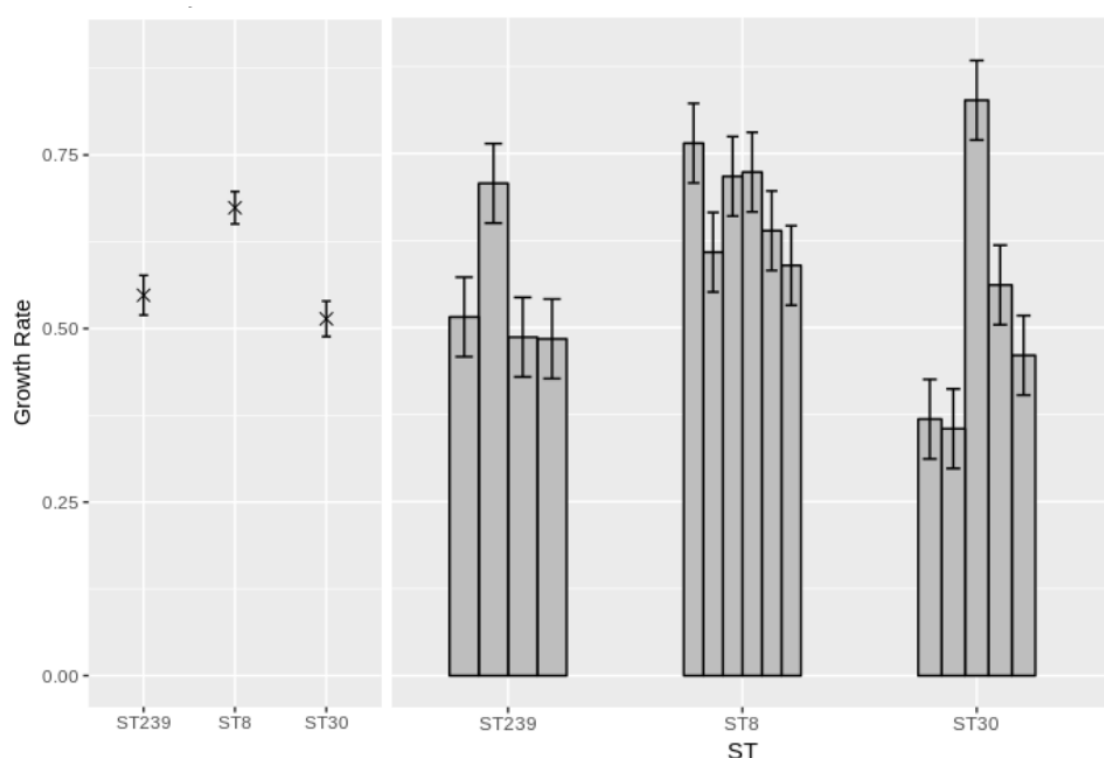


Figure 6.3. Least squares mean growth rate by ST, for ST239A – D, ST8A – F and ST30A – E.

The effect of isolate was also significant, which suggests that there is intra-ST variation in fitness, in addition to inter-ST variation. The largest variation occurred between the CC30 isolates, which included the isolate with the fastest average maximum growth rate (ST30C) and the lowest average maximum growth rate (ST30B), ranging from 0.355 – 0.826 mODh⁻¹.

Overall, the average growth rate of the isolates was slower in PS (0.2111 – 0.2629 mODh⁻¹) compared to BHI and TSB (0.6241 – 0.6759 and 0.8231 – 0.8749 mODh⁻¹, respectively), and the isolates grew fastest in TSB (Figure 6.4). This is most likely due to the growth promoting factors present in TSB and BHI, such as nitrogen, vitamins, minerals, sugars, amino acids and essential growth factors (Sigma-Aldrich, 2018; Thermo Fisher Scientific, 2001), and the

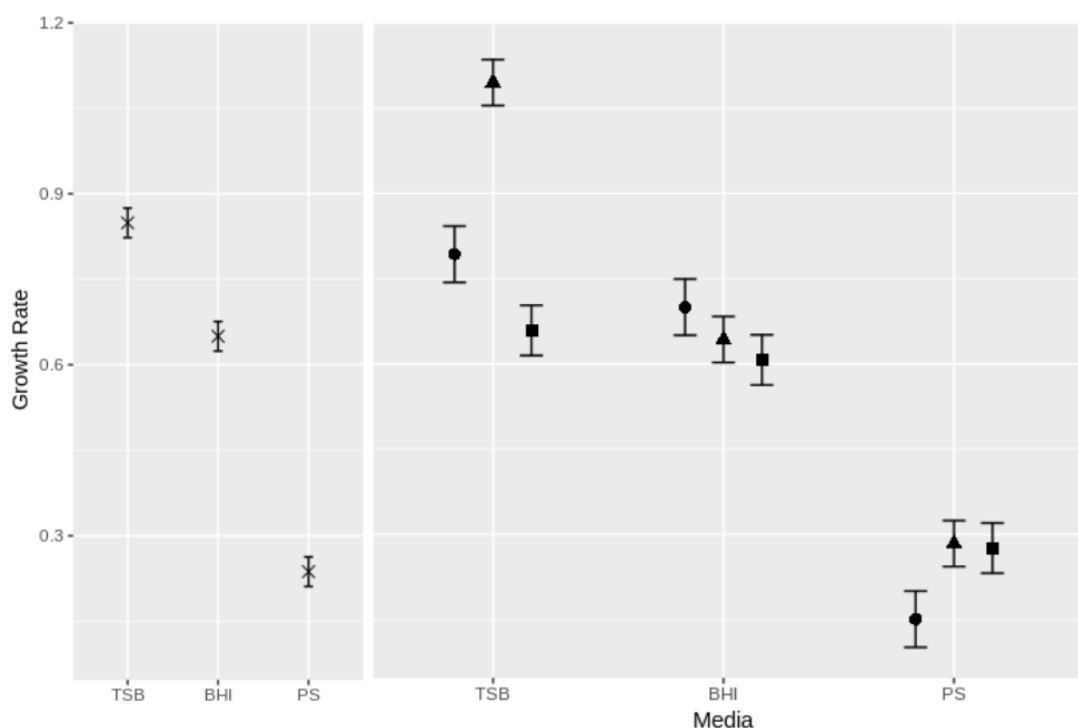


Figure 6.4. Least squares mean growth rate for each media condition; TSB, BHI and PS. Circle, ST239; triangle, ST8; square, CC30.

cytotoxic factors present in PS, such as antimicrobial peptides, toxins and complement factors (Thermo Fisher Scientific, 2002).

In TSB, the ST8 isolates had a significantly higher growth rate ($1.0546 - 1.1354 \text{ mODh}^{-1}$) than the ST239 isolates ($0.7435 - 0.8425 \text{ mODh}^{-1}$), and the CC30 isolates ($0.6148 - 0.7032 \text{ mODh}^{-1}$) had a significantly lower growth rate than the ST239 isolates. In BHI, there was no significant difference between the average growth rate of the ST239, ST8 and ST30 isolates ($0.6505 - 0.7495$, $0.6026 - 0.6834$ and $0.5628 - 0.6512 \text{ mODh}^{-1}$, respectively). In PS, ST239 ($0.1015 - 0.2005 \text{ mODh}^{-1}$) had a significantly lower growth rate than both CC30 and ST8 isolates ($0.2318 - 0.3202$ and $0.2436 - 0.3244 \text{ mODh}^{-1}$, respectively).

3.1.3 Modelling the effects on competitive ability

To further test the hypothesis that ST239 has lower fitness due to the hybridisation and increased AMR, the competitive ability of all 15 isolates in TSB, BHI and PS media (Chapter 5 Section 3.2.2) was fitted to a statistical model. Initially, a full model was constructed (Appendix B.7.5), and after removing non-significant effects, yielded a reduced model table (Table 6.2). Isolate, media by ST interactions, ST and media were found to be significant

Effect	df	Sum of squares	Mean squares	F-value	p-value
ST	2	0.2605196	0.1302598	8.2380	0.00046
Media	2	0.1618169	0.0809085	5.1169	0.00745
Media x ST	4	0.5758495	0.1439624	9.1046	0.00000
Isolate[ST]	12	4.2435429	0.3536286	22.3645	0.00000
Error	114	1.8025746	0.015812	-	-

Table 6.2. Reduced ANOVA table of significant competitive ability effects.

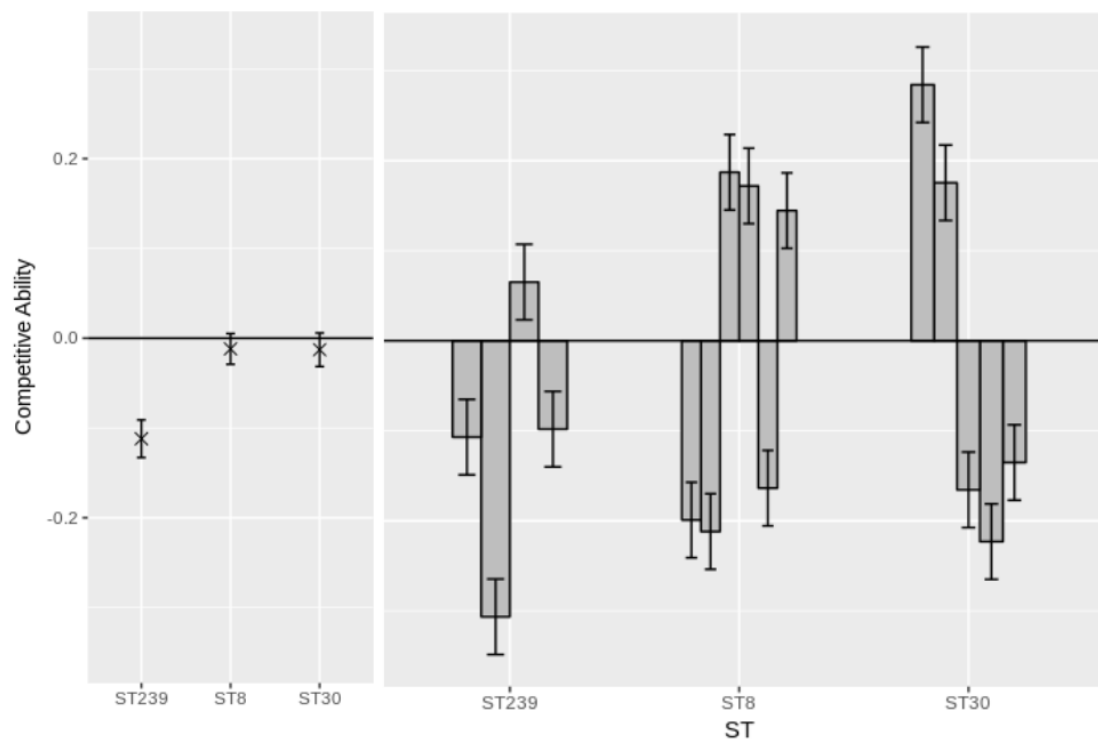


Figure 6.5. Least squares mean competitive ability by ST, for ST239A – D, ST8A – F and ST30A – E.

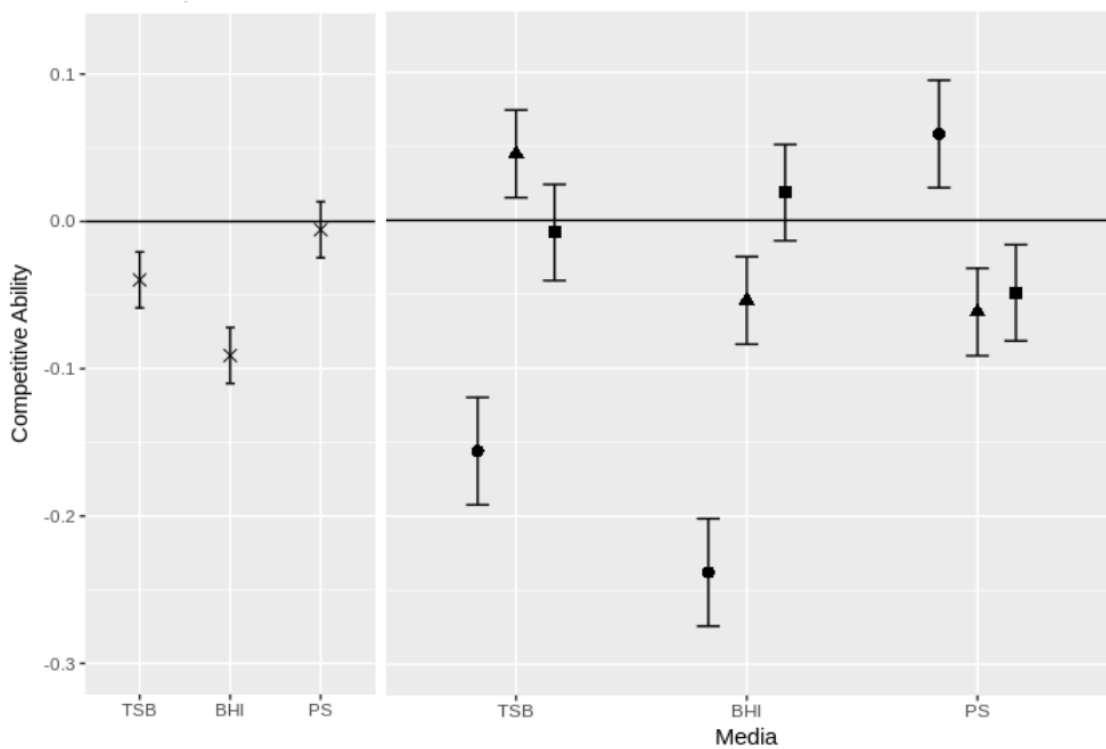


Figure 6.6. Least squares mean competitive ability for each media condition; TSB, BHI and PS. Circle, ST239; triangle, ST8; square, CC30.

effects (Appendix B.7.6 – B.7.8). Replicate number was not found to be a significant effect ($p > 0.05$), and was therefore not included in the reduced model.

The effect of ST was significant (Figure 6.5). Overall, the ST239 isolates had a lower competitive ability ($-0.133 - -0.091$) than the ST8 and CC30 isolates ($-0.0290 - 0.0052$ and $-0.0315 - 0.0059$, respectively), which suggests that the fitness of ST239 is lower than both ST8 and CC30. The effect of isolate was also significant, which indicates that there is high intra-ST variation, as well as inter-ST variation. The isolate with the lowest competitive ability was ST239B (-0.3065), which could not be detected above the minimum detection limit in any of the three media, and the isolate with the highest competitive ability was ST30A (0.2843).

Overall, the isolates had a lower competitive ability in BHI ($-0.1100 - -0.072$), followed by TSB ($-0.0587 - -0.0207$), and the overall highest competitive ability was in PS media ($-0.0246 - 0.0134$) (Figure 6.6). This reflects that less isolates had a negative competitive ability in PS, and less isolates were below the limit of detection after competition in PS, and therefore more isolates were able to grow during the PS competition. In both TSB and BHI, ST8 ($0.0153 - 0.00745$ and $-0.0838 - -0.0246$, respectively) and CC30 ($-0.0407 - 0.0242$ and $-0.0138 - 0.0512$, respectively) had a significantly higher competitive ability than ST239 ($-0.1923 - -0.1197$ and $-0.2743 - -0.2017$, respectively). However, in PS, the competitive ability of ST239 ($0.0220 - 0.0946$) was significantly higher than either CC30 or ST8 ($-0.0815 - -0.0165$ and $-0.0916 - -0.0324$, respectively). This shows that ST239 can have greater fitness than its hypothesised parent lineages under certain conditions, which may be a closer model of the conditions that ST239 has adapted to.

3.1.4 Competitive ability and growth rate as independent fitness measures

An overall competitive ability and an overall maximum growth rate were estimated for each isolate from the ANOVA analyses (Section 3.1.2 and 3.1.3). There was no correlation between competitive ability and maximum growth rate ($p = 0.158$) (Figure 6.7), suggesting that these two measurements are independent, and measure different aspects of fitness.

3.2 AMR

To investigate whether the lower fitness observed in ST239 compared to ST8 and CC30 (Section 3.1) may be related to increased levels of AMR, the ST239, ST8 and CC30 isolates were investigated for phenotypic resistance to clinically relevant antibiotics, and for genetic markers of AMR.

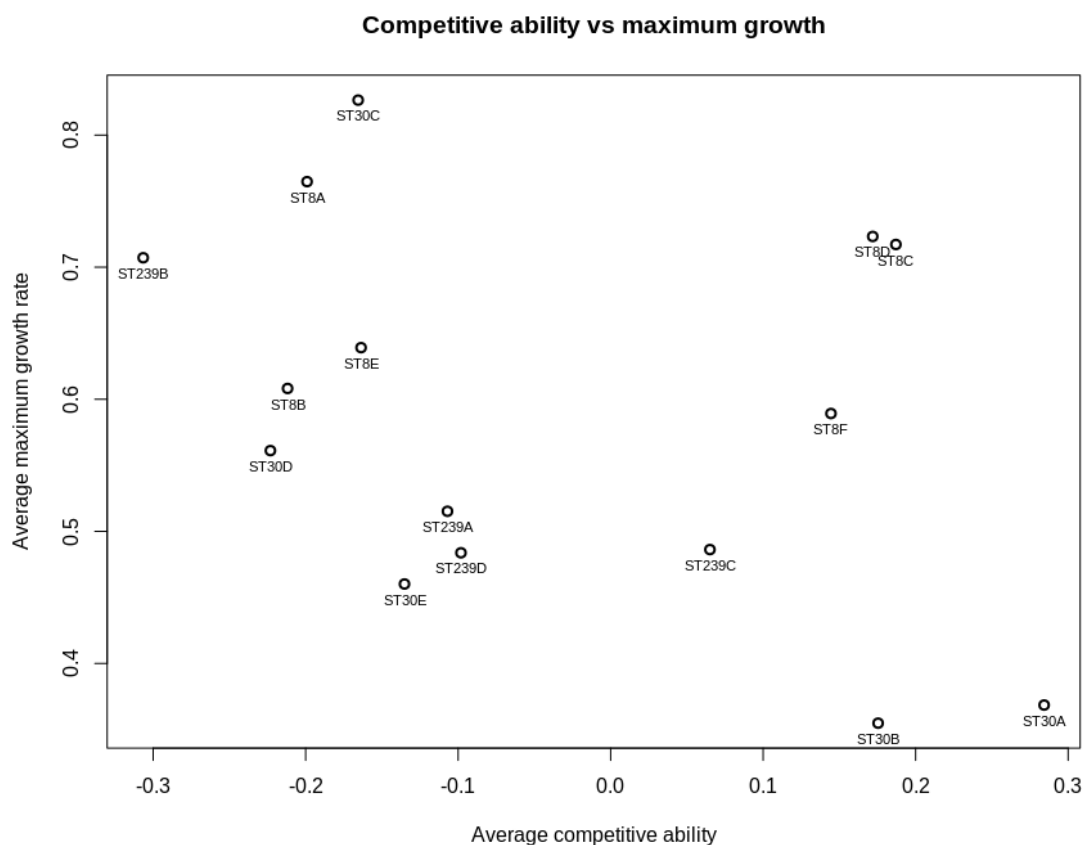


Figure 6.7. Correlation between competitive ability and maximum growth rate.

3.2.1 Phenotypic AMR profiles

Phenotypic resistance to 17 clinically relevant antibiotics from 13 antibiotic families was determined using MicroNaut-S MRSA/GP plates (MERLIN Diagnostika GmbH, 2018) (Table 6.3, Figure 6.8). Isolates ranged from having resistance to 0 – 8 of the families of antibiotics tested.

Isolate	Aminoglycoside (gentamycin)	β -lactam (ampicillin)	β -lactam (oxacillin)	β -lactam (penicillin)	Carboxylic Acid (mupirocin)	Cephalosporin (ceftarolin)	Fosfomycin	Fusidic Acid	Glycopeptide (teicoplanin)	Glycopeptide (vancomycin)	Glycylcycline (tigecycline)	Macrolide (erythromycin)	Oxazolidinone (linezolid)	Quinolone (moxifloxacin)	Rfamycin (rifampicin)	Streptogramin (synercid)	Diaminopyrimidine (trim/sulf)	Total antibiotic family resistances
ST239A	R	R	R	R	R	S	R	S	S	S	S	R	S	R	R	S	R	8
ST239B	R	R	R	R	S	S	R	S	S	S	S	R	S	R	R	S	R	7
ST239C	R	R	R	R	S	S	S	S	S	S	S	R	S	R	S	S	R	5
ST239D	R	R	R	R	S	S	S	S	S	S	S	R	S	R	S	S	R	5
ST8A	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	0
ST8B	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	0
ST8C	S	R	R	R	S	S	S	S	S	S	S	S	S	R	S	S	S	2
ST8D	S	R	R	R	S	S	S	S	S	S	S	R	S	R	S	S	R	4
ST8E	R	R	R	R	S	S	S	S	R	R	S	R	S	R	S	S	R	6
ST8F	S	R	S	R	S	S	S	R	S	S	S	S	S	S	S	S	S	2
ST30A	S	R	R	R	S	S	S	S	S	S	S	S	S	S	S	S	S	1
ST30B	S	R	S	R	S	S	S	S	S	S	S	S	S	S	S	S	S	1
ST30C	S	R	R	R	S	S	S	S	S	S	S	R	S	R	S	S	S	3
ST30D	S	R	R	R	S	S	S	S	S	S	S	R	S	R	S	S	S	3
ST30E	R	R	R	R	S	S	S	S	S	S	S	R	S	R	S	S	R	5

Table 6.3. Phenotypic antibiotic resistance profiles. Orange (R), resistant; green (S), susceptible.

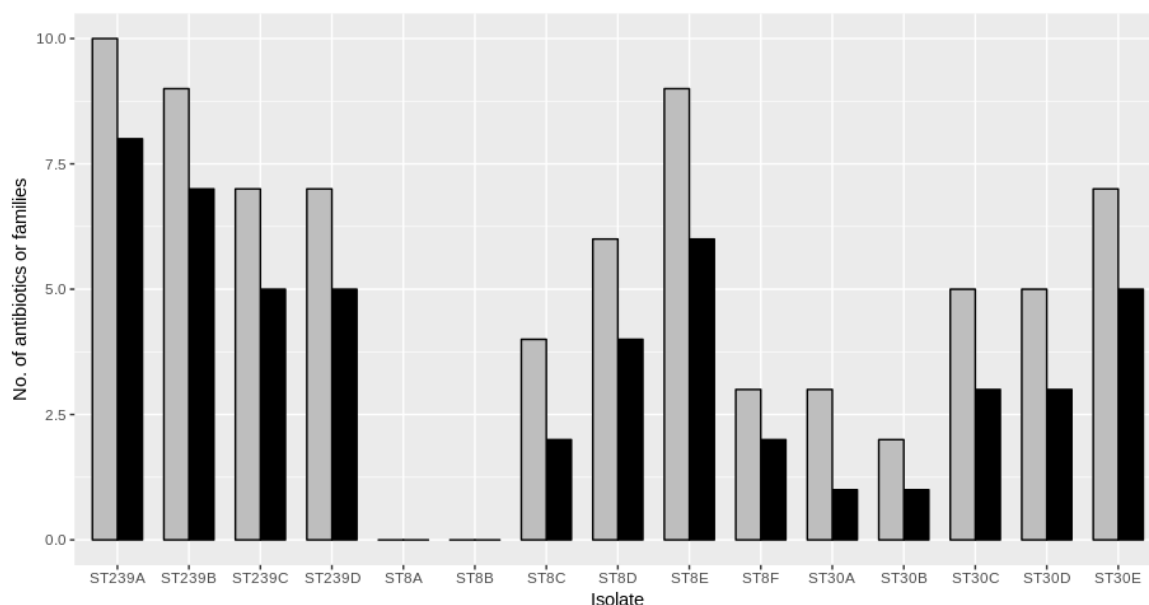


Figure 6.8. Number of AMR phenotypes for each isolate. Grey, total number of AMR phenotypes; black, total number of AMR families.

The ST239 isolates were resistant to significantly more antibiotic families than the ST8 (*t*-test, $t = 3.35$, $df = 6.27$, $p = 0.0144$) and CC30 isolates (*t*-test, $t = 2.79$, $df = 5.37$, $p = 0.0354$). All of the ST239 isolates had clinical levels of resistance to gentamycin, 1st, 2nd and 3rd generation β -lactams, erythromycin, moxifloxacin and the trimethoprim/sulfamethoxazole combination. Two of the ST8 isolates were susceptible to all 17 antibiotics tested. All isolates were susceptible to ceftarolin, tigecyclin, linezolid and streptogramin.

3.2.2 Genetic AMR profiles

The presence of AMR genes and mutations in the different hybrid regions of the genomes was identified (Appendix B.4.5). The total number of AMR genes was significantly higher in the ST239 genomes than in the ST8 genomes (*t*-test, $t = 3.78$, $df = 7.16$, $p = 0.00662$) and CC30 genomes (*t*-test, $t = 3.93$, $df = 6.63$, $p = 0.00633$) (Figure 6.9). This is consistent with the higher number of AMR phenotypes observed in the ST239 isolates in Section 3.2.1.

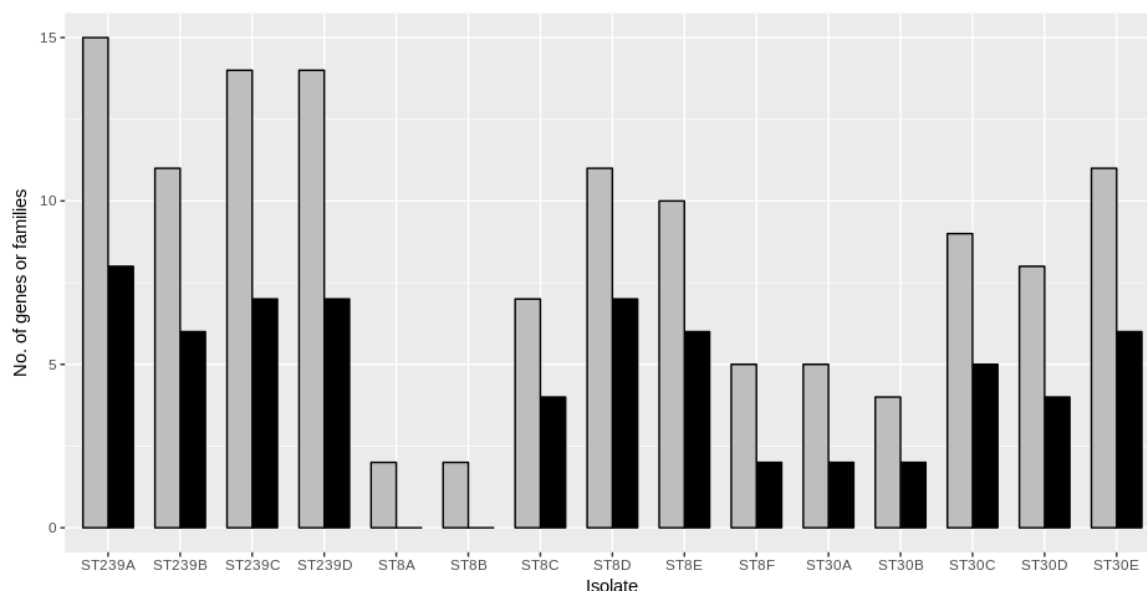


Figure 6.9. Number of AMR genes identified for each isolate. Grey, total number of AMR genes; black, total number of AMR families.

In addition to genes that encode resistance to specific classes of antibiotics, four multi-drug resistance efflux pumps were identified within the genomes, encoded by *qac*, *mepA*, *tet(K)* and *tet(38)*. The *mepA* and *tet(38)* efflux pump genes were identified in all 15 of the isolates.

The *mepA* efflux pump has been linked with resistance to tetracycline, glycylicycline, acridine dye and fluoroquinolone antibiotics, and reduced susceptibility to tigecycline, and is recorded in 99.2% of all *S. aureus* NCBI whole-genome shotgun assemblies (CARD, 2019). The *tet(38)* efflux-pump has been linked with resistance to over 20 different antibiotic classes, and is present in 99.4% of *S. aureus* NCBI whole-genome shotgun assemblies. The efflux pump encoded by *tet(K)* has a similar resistance profile to the *tet(38)* efflux pump, but was identified only in the ST239A, ST239C, ST239D and ST8F genomes. It is present in 10% of *S. aureus* NCBI whole-genome shotgun assemblies. Finally, the ST239C and ST239D isolates also contained the *qac* multi-drug efflux pump, which has been linked to resistance to over 20

different antibiotic classes, and is present in 2.2% of *S. aureus* NCBI whole-genome sequences (CARD, 2019).

Excluding these multi-drug efflux pumps, the total number of antibiotic families (Appendix B.4.6) where resistance genes were identified in a single isolate ranged from 0 – 8 (Figure 6.9), with the ST239 isolates encoding resistance genes to a significantly greater number of

Isolate	Antibiotic Family											Total resistances
	Aminoglycoside	Diaminopyrimidine	Fluoroquinolone	Fosfomycin	Fusidic acid	Glycopeptide	Macrolide/Lincosamide/Streptogramin	Mupirocin	Nucleoside	Penam	Rifamycin	
ST239A	R	S	R	R	S	S	R	R	R	R	R	8
ST239B	R	S	R	R	S	S	R	S	R	R	S	6
ST239C	R	R	R	S	S	S	R	R	R	R	S	7
ST239D	R	R	R	R	S	S	R	S	R	R	S	7
ST8A	S	S	S	S	S	S	S	S	S	S	S	0
ST8B	S	S	S	S	S	S	S	S	S	S	S	0
ST8C	R	S	R	R	S	S	S	S	S	R	S	4
ST8D	R	R	R	R	S	S	R	S	R	R	S	7
ST8E	R	R	R	S	S	R	R	S	S	R	S	6
ST8F	S	S	S	S	R	S	S	S	S	R	S	2
ST30A	S	S	S	R	S	S	S	S	S	R	S	2
ST30B	S	S	S	R	S	S	S	S	S	R	S	2
ST30C	R	S	R	R	S	S	R	S	S	R	S	5
ST30D	S	S	R	R	S	S	R	S	S	R	S	4
ST30E	R	R	R	R	S	S	R	S	S	R	S	6

Table 6.4. Genetic AMR profiles. Orange (R), resistance gene identified; green (S), no resistance gene identified.

antibiotic families compared to the ST8 isolates (t -test, $t = 2.97$, $df = 6.05$, $p = 0.0246$) and the CC30 isolates (t -test, $t = 3.56$, $df = 5.83$, $p = 0.0125$) (Table 6.4).

3.2.3 Correlating phenotypic and genotypic AMR profiles

Phenotypic and genotypic resistance were well correlated for the majority of antibiotics tested; 93.8% of phenotypic and genotypic resistance/susceptibility profiles were in agreement (Table 6.5). Only nucleoside resistance determinants were identified genotypically where no representative antibiotic was tested for phenotypically, and therefore nucleoside resistance is not included in Table 6.5. All isolates were susceptible to cephalosporin, glycylcycline, oxazolidinone and streptogramin antibiotics, and no evidence of resistance was identified in the isolate genomes for these antibiotic classes.

Only ST8A and ST8B were susceptible to all antibiotics tested, and they also did not contain any specific AMR genes, although they did both contain the genes for the *tet(38)* and *mepA* multi-drug efflux pumps. All isolates, with the exception of ST8A and ST8B, were resistant to the β -lactam antibiotics penicillin and ampicillin, which is encoded by *mecA* in the *SCCmec* element, or *blaZ*. ST8F and ST30B, which do not encode *mecA* but do encode *blaZ*, did not have phenotypic resistance to oxacillin, whereas all isolates that encoded *mecA* were also resistant to oxacillin.

NICE guidelines (The National Institute for Health and Care Excellence, 2019) for the treatment of MRSA-associated septicaemia and severe skin and soft-tissue infections recommend treatment with a glycopeptide, such as vancomycin or teicoplanin. All isolates, with the exception of ST8E, were susceptible to vancomycin and teicoplanin, and

Isolate	Aminoglycoside (gentamycin)	β -lactam (penicillin, ampicillin or oxacillin)	β -lactam (oxacillin)	Carboxylic Acid (mupirocin)	Cephalosporin (ceftarolin)	Fosfomycin	Fusidic Acid	Glycopeptide (vancomycin or teicoplanin)	Glycylcycline (tigecycline)	Macrolide (erythromycin)	Oxazolidinone (linezolid)	Quinolone (moxifloxacin)	Rfamycin (rifampicin)	Streptogramin (synercid)	Diaminopyrimidine (trimethoprim/sulfamethoxazole)
ST239A	R	R	R	R	S	R	S	S	S	R	S	R	R	S	R/S
ST239B	R	R	R	S	S	R	S	S	S	R	S	R	R/S	S	R/S
ST239C	R	R	R	S	S	S	S	S	S	R	S	R	S	S	R
ST239D	R	R	R	S	S	S/R	S	S	S	R	S	R	S	S	R
ST8A	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
ST8B	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
ST8C	S/R	R	R	S	S	S/R	S	S	S	S	S	R	S	S	S
ST8D	S/R	R	R	S	S	S/R	S	S	S	R	S	R	S	S	R
ST8E	R	R	R	S	S	S	S	R	S	R	S	R	S	S	R
ST8F	S	R	S	S	S	S	R	S	S	S	S	S	S	S	S
ST30A	S	R	R	S	S	S/R	S	S	S	S	S	S	S	S	S
ST30B	S	R	S	S	S	S/R	S	S	S	S	S	S	S	S	S
ST30C	S/R	R	R	S	S	S/R	S	S	S	R	S	R	S	S	S
ST30D	S	R	R	S	S	S/R	S	S	S	R	S	R	S	S	S
ST30E	R	R	R	S	S	S/R	S	S	S	R	S	R	S	S	R

Table 6.5. Correlation between phenotypic and genotypic resistance. Orange (R), both phenotypically and genotypically resistant; green (S), both phenotypically and genotypically susceptible; yellow (S/R), phenotypically susceptible and genotypically resistant; blue (R/S), phenotypically resistant and genotypically susceptible.

glycopeptide resistance determinants were only detected in ST8E, which were encoded in a vancomycin resistance operon (containing *vanA*, *vanH*, *vanR*, *vanS*, *vanX* and *vanZ*). Vancomycin resistance in MRSA isolates is relatively rare, and the *vanA* gene cluster is only present in 0.2% of *S. aureus* NCBI whole-genome sequences (CARD, 2019). The AMR profile of ST8E is more similar to the ST239 isolates, as it has a higher number of AMR genes and resistance to a greater number of AMR families than the other ST8 isolates.

NICE guidelines (The National Institute for Health and Care Excellence, 2019) also state that for severe skin and soft tissue infections where a glycopeptide is unsuitable, linezolid may be used, or tigecycline and daptomycin for complicated skin and soft-tissue infections. All isolates tested were phenotypically susceptible to linezolid and tigecycline, as well as ceftarolin and synecid. Although daptomycin resistance was not tested for phenotypically, several genes are known to be associated with daptomycin resistance in *S. aureus* (Tran *et al.*, 2015), and none of these were identified in these isolates.

The mutations in DNA topoisomerase (*gyrA* and *parC*) conferring quinolone resistance did correspond with moxifloxacin resistance. Some isolates had resistance caused by a single mutation in *gyrA* or *parC*, and some had several mutations in both genes, which may correspond to an increase in quinolone resistance (Bagel *et al.*, 1999). None of these mutations in *gyrA* or *parC* were found in susceptible isolates. Similarly for erythromycin resistance, isolates with phenotypic resistance contained either *ermA* or *ermC*, which encode methyltransferases that alter that ribosomal drug binding site (Leclercq, 2002), and neither of these resistance genes were found in the susceptible isolates.

Several discrepancies between phenotypic and genotypic AMR profiles were observed. For example, ST239B was found to be phenotypically resistant to rifampicin, however no point mutations in the *rpoB* gene that encode resistance were detected. It is possible that resistance to rifampicin emerged in ST239B during the AMR assay as it was exposed to rifampicin, which occurred subsequently to DNA sequencing. Spontaneous mutations in *rpoB* that confer resistance to rifampicin are a measure of hypermutation (Galan *et al.*, 2004).

Additionally, ST239A and ST239B were both phenotypically resistant to trimethoprim/sulfamethoxazole, however neither encoded any of the dihydrofolate reductase genes that were present in the other diaminopyrimidine-resistant isolates. Diaminopyrimidine resistance typically arises through the acquisition of an alternative dihydropteroate synthase, either through a plasmid or integron (Brolund *et al.*, 2010). Both *dfrC* and *dfrG* were identified in other diaminopyrimidine-resistant isolates in this study, and several other *dfr* family genes are recorded in the CARD database (CARD). Although it is possible that ST239A and ST239B encode an unknown determinant for diaminopyrimidine resistance, it is more likely that the phenotypic resistance is due to a low-level plasmid that conveys resistance, or due to errors in phenotyping.

All isolates that encoded the aminoglycoside modification enzyme *ACC(6')-Ie-APH(2'')-Ia* were phenotypically resistant to gentamycin, which was the only aminoglycoside antibiotic that was tested for phenotypically (Rouch *et al.*, 1987). However, all isolates that only encoded aminoglycoside modification enzymes that were not *AAC(6')-Ie-APH(2'')-Ia* were found to be phenotypically susceptible to gentamycin. Aminoglycoside modification enzymes are specific to certain aminoglycosides (Ramirez and Tolmasky, 2011), for example *APH(3')-*

IIIa is specific to kanamycin resistance (Trieu-Cuot and Courvalin, 1983), and ANT(4')-Ib is specific to amikacin and tobramycin (Sabtcheva *et al.*, 2003).

Mutations associated with fosfomycin resistance were encoded by a mutation in the fosfomycin-binding site of *murA* that was identified in ten of the genomes. However, only two of the isolates, ST239A and ST239B, were determined to be phenotypically resistant to fosfomycin. Therefore, eight isolates had *murA* mutations but were not phenotypically resistant to fosfomycin. None of these eight isolates were able to grow on even the lowest levels of fosfomycin that were included in the phenotypic assay (Section 3.2.1). Only three mutations were identified in *murA* in the isolates in this study. All of the ST30 isolates were phenotypically susceptible to fosfomycin and contained two *murA* mutations, D278E and E291D. Individually, these mutations have been shown to increase the fosfomycin MIC range to 1 – 2 mg/L (Fu *et al.*, 2015), which is lower than the concentrations of fosfomycin that were phenotypically tested using MicroNaut plate assays, and lower than the fosfomycin clinical break-point of 64 mg/L (The European Committee on Antimicrobial Susceptibility Testing, 2014). A G257D mutation was identified in both the phenotypically resistant and the phenotypically susceptible isolates. This mutation has been found to convey a fosfomycin MIC rate of between 2 – >1024 mg/L (Fu *et al.*, 2015), and therefore the levels of resistance conferred by this mutation may not always exceed those included in the MicroNaut plate assays.

3.2.4 Modelling the effects of AMR on fitness

To test the hypothesis that the lower fitness of ST239 (Section 3.1), is due the higher levels of AMR (Section 3.2.1 and Section 3.2.2), the AMR level of all 15 isolates, and the fitness of

Model	Effect	df	Sum of squares	Mean squares	F-value	p-value
Competitive ability vs genotypic AMR	ST	2	0.09415503	0.0470775	1.8714	0.20972
	No. AMR genes	1	0.00290800	0.0029080	0.1156	0.741966
	No. AMR genes x ST	2	0.20731605	0.1036580	4.1206	0.05365
	Error	9	0.22640258	0.025156	-	-
Competitive ability vs phenotypic AMR	ST	2	0.03335565	0.0166778	0.5351	0.60312
	No. AMR phenotypes	1	0.08350440	0.0835044	2.6794	0.13608
	No. AMR phenotypes x ST	2	0.11821063	0.0591053	1.8965	0.20546
	Error	9	0.28049170	0.03116	-	-
Growth rate vs genotypic AMR	ST	2	0.04550721	0.0227536	1.5220	0.26955
	No. AMR genes	1	0.00208898	0.0020890	0.1397	0.71720
	No. AMR genes x ST	2	0.06816742	0.0340837	2.2798	0.15811
	Error	9	0.13455084	0.014950	-	-
Growth rate vs phenotypic AMR	ST	2	0.04534532	0.0226727	1.1403	0.36191
	No. AMR phenotypes	1	0.02000731	0.0200073	1.0063	0.34201
	No. AMR phenotypes x ST	2	0.02114421	0.0105721	0.5317	0.60497
	Error	9	0.17894683	0.019883	-	-

Table 6.6. ANOVA tables of the effects of AMR on growth rate and competitive ability.

each isolate, were fitted to a statistical model. There was no correlation between growth rate and number of phenotypic AMR families or AMR genes, and there was no correlation between competitive ability and number of phenotypic AMR families or AMR genes ($p > 0.05$) (Table 6.6, Appendix A.4.1 – A.4.2). This suggests that high levels of AMR may not be the main driver of the lower fitness of ST239.

4. Discussion

4.1 Fitness effects

ST239 has been a highly prevalent and successful MRSA strain in healthcare settings around the globe, despite the extreme genetic transfer events that have occurred in its recent evolutionary history, and typically high levels of AMR, which may impose large fitness costs.

In this study, the relative fitness of ST239, ST8 and CC30 isolates was compared using two independent fitness measures, and the overall fitness of ST239 was found to be lower than ST8 and CC30. This is despite ~80% of the ST239 genome that is shared with a recent common ST8-like ancestor. This is evidence of a negative fitness effect on ST239, that may be linked to the recent large chromosomal replacement event that formed ST239.

Growth rate and competitive ability of the isolates were not found to be correlated. This suggests that these two measurements are independent, and measure different aspects of fitness. The relationship between isolates in co-culture is complex, and therefore, for a microorganism to have a strong competitive ability requires traits beyond a high individual growth rate in that environment. For example, Abreu (2019) observed that *Escherichia coli* and *Pseudomonas* isolates with lower growth rates had higher competitive ability at low dilutions (Abreu *et al.*, 2019).

The media in which the isolates were either mono-cultured or co-cultured also had a significant effect on the fitness of the isolates. A major limitation of this study, and many studies that use standard lab culture methods, is that the controlled lab environment is unlikely

to be representative of the natural environment that clinical microorganisms have adapted to, and therefore the fitness of the isolates in this study may not necessarily reflect their fitness in the clinical environment. Experiments that are more reflective of the typical environment of ST239 may include the use of animal models, antimicrobials, or co-culture including a range of human commensal and pathogenic bacterial species that ST239 may be more likely to come into contact with in a healthcare setting.

Further limitations of this study include conservative limits of detection in the competition experiments, which resulted in the least fit isolates being over-represented in these analyses. Because the ST239 isolates had a lower competitive ability overall, more of the ST239 isolates were measured below the minimum detection threshold. It is therefore likely that the competitive ability, and therefore fitness, of ST239 may be even lower than calculated in this study.

Finally, there was a large fitness difference detected between isolates from within the same ST. The inclusion of a larger number of diverse isolates from each ST, would be more representative, and enable more robust conclusions to be made about the fitness differences between these STs.

4.2 AMR

To investigate whether the lower fitness of ST239 may be related to increased levels of AMR, rather than the effects of the extreme genetic transfer events, the AMR profiles of the 15 ST239, ST8 and CC30 isolates was also investigated. The ST239 isolates displayed an overall higher level of AMR, both phenotypically and genotypically, compared to the ST8 and CC30

isolates. Increased levels of AMR are typically associated with a high fitness cost (Vogwill and MacLean, 2015), however, no correlation was observed in this data-set between increased AMR and decreased fitness.

The low number of isolates from each ST was a major limitation of this study, and the isolates may be biased towards virulent, multi-drug resistant isolates due to sampling bias, as they were mainly selected from international repositories. Although the number of ST239 isolates in this data-set was limited, the least fit ST239 isolates did not have higher levels of AMR. This further supports that AMR may not be the driving factor behind the lower fitness of ST239.

The accuracy of the AMR determinants called by ARIBA has been verified in a number of studies, for example Lauener (2019) found that there was high congruence of over 99% between phenotypic resistance and SNPs in the target genes called with ARIBA (Lauener *et al.*, 2019). In this chapter, the majority of genotypic resistances predicted using ARIBA was consistent with the phenotypic resistances determined using the MicroNaut assays, which further supports the accuracy of ARIBA. Where there were discrepancies between the AMR phenotype and the AMR genotype that was initially predicted using ARIBA, these could mostly be justified after further investigation of these genes.

4.3 Conclusion

Overall, ST239 had lower fitness compared to both ST8 and ST30, despite sharing the majority of its genome with a recent common ST8-like ancestor. Although ST239 also had higher AMR, this was not found to be the main driver of the fitness cost. Therefore, this

fitness cost is more likely the result of the recent extreme genetic transfer events that formed ST239. The lower fitness attributed to ST239 compared to its parent lineages shows that large HGT events that occur between bacteria in nature can have measurable negative impacts on the fitness of these isolates, and that in the case of ST239, these negative fitness effects can still be detected after many years.

The decreased fitness and increased AMR of ST239 has also been observed in comparison with other successful MRSA lineages, including ST59-IV and ST22-IV, which have both been able to overtake ST239 in clinical settings in recent years (Li, *et al.*, 2018; Knight *et al.*, 2012). Therefore, it is likely that properties that have benefitted ST239 in the past, including high levels of AMR, are no longer able to out-compete some of these emerging higher-fitness strains in clinical settings.

Chapter 7

Conclusions

1. Conclusions

1.1 Conclusions

The rise of multidrug-resistant pathogens is one of the most urgent global public health threats. Each year, 700,000 people die from AMR infections, and the World Health Organisation estimates that AMR could cause up to 10 million deaths per year by 2050 and contribute to a catastrophic global financial crisis (IACG, 2019). The ability of bacteria to rapidly and effectively evolve resistance to all known antibiotics is threatening to plunge the world into a post-antibiotic era, where common infections and injuries are no longer treatable (World Health Organisation, 2014). The drivers and mechanisms of bacterial evolution, which have resulted in the emergence and persistence of multidrug-resistant pathogens, are extremely complex and not thoroughly understood (Holmes *et al.*, 2016). The aim of this thesis has been to provide an insight into the drivers that have contributed to the formation, global emergence, and eventual decline of a dominant healthcare-associated MRSA strain, ST239. The relatively recent formation of this prevalent hybrid lineage, which resulted from a single large chromosomal replacement event between two divergent *S. aureus* lineages (Robinson and Enright, 2004), has provided a unique opportunity to investigate the evolutionary consequences of extreme horizontal genetic transfer events, which in the case of ST239, can generate powerful and pervasive multidrug-resistant pathogens.

For decades, ST239 has persisted as a dominant MRSA strain in healthcare settings around the globe, as reviewed in Chapter 2. During the early 1990s – late 2000s, it was estimated that ST239 was responsible for up to 90% of hospital-acquired MRSA infections within a region accounting for over 60% of the world's population (Feil *et al.*, 2008). The fact that ST239 has

been able to spread into healthcare settings on an international scale indicates that it is extremely well adapted to the healthcare environment. This is further evidenced by high levels of AMR and virulence that are associated with this MRSA strain, which have provided ST239 with additional tools to succeed in this niche.

The evolution of ST239 was found to be consistent with the Robinson-Enright model, supporting the hypothesis that ST239 formed through a single large chromosomal transfer event that involved both an ST8-like and an ST30-like ancestor. From an analysis of all available bacterial DNA sequences in Chapter 3, a group of five historical ST30 isolates were identified as the closest known relatives of the acquired region of the ST239 genome. These five ST30 sequences (Robinson *et al.*, 2005) were from globally diverse phage type 80/81 isolates from between 1955 – 1968, and were all lacking *SCCmec*, which suggests that the donor of the minor region of the ST239 genome may have been closely related to an MSSA of phage type 80/81 from this historic healthcare-associated lineage.

Evidence of weaker selective constraints on the acquired region of ST239, compared to the backbone region, were observed in Chapter 4. This may explain why the large genetic transfer event involving this region of the *S. aureus* genome was possible, and why the formation of other hybrid *S. aureus* strains has also involved this genetic region (Robinson and Enright, 2004; Thomas *et al.*, 2012; Nimmo *et al.*, 2014). Evidence of parallel evolution in multiple genes related to adaptation to host immune response and AMR suggest that these are two of the major evolutionary pressures that have been faced by ST239, which is consistent with the persistence of ST239 as a healthcare-associated pathogen.

Chapter 7

In Chapter 4, the date of the hybridisation event that formed ST239 was estimated as many years before methicillin was introduced into clinics, which is earlier than previous estimates using less diverse ST239 data-sets. Therefore, the use of earlier antimicrobials such as penicillin may have been the most likely drivers of the international spread of ST239. During this time, the antibiotic era was only just beginning, and the landscape that was populated by healthcare-associated pathogens was changing dramatically to include high concentrations of antimicrobials. It is possible that in this setting, the advantages of the ST239 genome were able to outweigh the fitness costs of its unique hybrid ancestry, which enabled the newly formed ST239 hybrid to seize this opportunity to emerge as a dominant strain, and subsequently spread around the world.

In Chapter 2, the ST239 genome was found to be strongly linked to the large *SCCmec*-III element, although *SCCmec*-III was hardly ever seen outside of this one lineage. In Chapter 4, it was estimated that the *SCCmec*-III element was acquired by ST239 at around the same time as the large chromosomal transfer event, and it is therefore likely that there is some advantage of *SCCmec*-III that is unique to the ST239 lineage, which may have contributed to its success. The fact that ST239 has not been able to consistently acquire any alternative *SCCmec*-type signifies the importance of this genetic element to the ST239 lineage. The *SCCmec* element is a hot-spot for recombination and has been highly variable, even over the relatively short period of time that ST239 has existed. It is therefore possible that the *SCCmec*-III element was able to pick up other AMR genes that conferred resistance to alternative antibiotics as they were introduced into clinics, such as tetracycline and streptomycin, which further contributed to the success of ST239. Additionally, the increased virulence of ST239 has also been linked to its success as a healthcare-associated MRSA strain, including carriage of the

prophage-mediated virulence determinant *sasX* (Li *et al.*, 2012), and the ACME-II variant (Espedido *et al.*, 2012).

Competition experiments in mono-culture and co-culture were used to test the hypothesis that ST239 has lower fitness due to the fitness costs of the extreme genetic transfer events in its recent evolutionary history. Using a novel method to measure relative bacterial fitness in competition that was developed in Chapter 5, in Chapter 6 the ST239 isolates in this study were found to have overall lower fitness than the parent lineages in mono-culture and in co-culture. ST239 is less fit than ST8, with which it shares ~80% of its genome with a recent common ancestor, and it was determined that higher levels of AMR were probably not the main driver of this fitness cost. Therefore, it is likely that the decreased fitness of ST239 is a consequence of the large HGT events that formed ST239. Although it was not possible to attribute this fitness cost to either the acquisition of the *SCCmec*-III element or the large chromosomal replacement event, these two examples of extreme genetic transfer events seem to be strongly linked, and therefore it is likely that both events have contributed to the decreased fitness of ST239.

There have been several examples in healthcare settings of different *S. aureus* isolates that have replaced ST239 as the dominant MRSA strain. The most common example is ST22, which typically carries the small *SCCmec*-IV element, is fitter than ST239 *in vitro* (Knight *et al.*, 2012), and is less resistant to antimicrobials (Amorim *et al.*, 2007). ST22-IV is more similar to conventional community-acquired MRSA strains, which have been moving into healthcare settings (Uhlemann *et al.*, 2014). As the line between healthcare and community-associated MRSA infections blurs, it is possible that ST239 has been unable to compete with

MRSA strains that are associated with higher fitness and smaller *SCCmec* elements (Lee *et al.*, 2007), or in the community where there are different evolutionary pressures that ST239 is less well adapted to. This inability of ST239 to adapt to changing niches with new competitors, alongside the fitness burden of its hybrid origin and the strong link between the ST239 genome and the *SCCmec*-III element that initially contributed to its success, may therefore be driving the downfall of the ST239 lineage.

1.2 Future work

A large amount of the work in this thesis was carried out using publicly available online bacterial DNA sequence databases and repositories. As the quality of DNA sequencing and number of available bacterial sequences continues to increase dramatically, researchers can use these tools to sort, filter and search thousands of bacterial genomes extremely quickly, to make robust conclusions about large and diverse pathogen data-sets. However, in order for researchers to be able to make practical use of this wealth of data, more consistent collection and inclusion of searchable metadata and supporting information is necessary (Schmedes *et al.*, 2015). Additionally, efforts should be made to increase the diversity of sequences included in these databases and reduce sampling bias, following on from similar efforts for eukaryotes (Wellcome Sanger Institute, 2019).

Although it is hard to accurately quantify the extent of HGT in bacteria, its effect on the evolution of life is so apparent, particularly in prokaryotes, that it has been suggested that the traditional concept of the “tree of life” is closer to a “rhizome of life” or “web of life” (Raoult, 2010; Vos *et al.*, 2015). The work in this thesis is just a single example of an application of using publicly available bacterial DNA sequence databases to investigate a single hybrid

MRSA strain, and focussed on just three prevalent STs from *S. aureus*. However, applications such as BIGSI (Bradley *et al.*, 2019) can facilitate the analysis of diverse microbial genomes, enabling researchers to make inferences about the origins of both intra- and inter-species HGT events for a wide range of species.

Evidence of hybrid bacteria that were formed through large chromosomal replacement events, such as ST239, challenge the concept that HGT involves only relatively short sections of DNA. Large HGT events in bacteria are thought to be rare, however there are already several other examples of successful pathogenic hybrids in the literature (Brochet *et al.*, 2008; Sheppard *et al.*, 2013; Chen *et al.*, 2014). Further comparative studies of a wider range of diverse bacterial genomes may find that successful hybrid bacteria are more common than we know. It is likely that many hybridisation events in bacteria are unsuccessful and thus are never observed, and that only a few are able to successfully emerge as “hopeful monsters”. However, it is important to appreciate the effects of this type of saltational evolution, as it has a considerable potential to generate new public health threats suddenly and unpredictably (Croucher and Klugman, 2014).

Although hybrid bacteria experience fitness costs due to the large HGT events that define them, the example of ST239 shows that these hybridisation events can be important, and possibly key, to the success of these bacteria. The mechanisms that underpinned the formation of the hybrid ST239 genome have not yet been identified, and it may be that large chromosomal replacement events between bacteria occur via fundamentally different processes than typical HGT events. Only recently have mechanisms that may facilitate large horizontal transfer events in bacteria begun to be understood (Morikawa *et al.*, 2012; Hua *et*

al., 2017). However, it is likely that further *in vitro* studies to investigate large HGT events in microbes are key to uncovering how hybrid bacteria such as ST239 have formed. Understanding these mechanisms, as well as the implications that they have on bacterial fitness and subsequent evolution, would provide insights into how these events can lead to extremely successful and highly multi-drug resistant pathogens, such as ST239.

2. Concluding remarks

“Thus strangely are our souls constructed, and by such slight ligaments are we bound to prosperity or ruin.” — Mary Shelley, Frankenstein.

My thesis has provided an insight into the rise and fall of a dominant hybrid MRSA ST239, from an epidemiological, genomic and evolutionary perspective. ST239 has been a highly successful EMRSA around the world for several decades, is well adapted to its healthcare niche, and has high levels of AMR and virulence. However, there is evidence that the pervasiveness of ST239 is beginning to wane. Key factors that are related to the downfall of this dominant hybrid lineage are likely to include the large *SCCmec*-III element that carries multi-drug resistance and has remained closely linked to the ST239 genome, the fitness costs of the formative horizontal genetic transfer events that have continued to burden ST239, and an inability to adapt to a changing healthcare landscape in the presence of new competitors.

Microbiology, epidemiology and genomics provide powerful tools to investigate the agents of devastating infectious diseases. As the problem of multi-drug resistance becomes more and more urgent, and we teeter on the brink of a post-antibiotic era, it is more important than ever that we utilise these tools to improve our understanding of the drivers of the rise and fall of dominant pathogenic bacterial lineages, so that we can work towards the informed prevention and eradication of multi-drug resistant super-bugs.

References

References

- Aanensen, D.M., Feil, E.J., Holden, M.T.G., Dordel, J., Yeats, C.A., Fedosejev, A., Goater, R., Castillo-Ramírez, S., Corander, J., Colijn, C., Chlebowicz, M.A., Schouls, L., Heck, M., Pluister, G., Ruimy, R., Kahlmeter, G., Åhman, J., Matuschek, E., Friedrich, A.W., Parkhill, J., Bentley, S.D., Spratt, B.G., Grundmann, H., 2016. Whole-Genome Sequencing for Routine Pathogen Surveillance in Public Health: a Population Snapshot of Invasive *Staphylococcus aureus* in Europe. *mBio* 7. <https://doi.org/10.1128/mbio.00444-16>
- Abbott, I.J., Jenney, A.W.J., Jeremiah, C.J., Mirčeta, M., Kandiah, J.P., Holt, D.C., Tong, S.Y.C., Spelman, D.W., 2015. Reduced in Vitro Activity of Ceftaroline by Etest among Clonal Complex 239 Methicillin-Resistant *Staphylococcus aureus* Clinical Strains from Australia. *Antimicrobial Agents and Chemotherapy* 59, 7837–7841. <https://doi.org/10.1128/aac.02015-15>
- Abdulgader, S.M., Shittu, A.O., Nicol, M.P., Kaba, M., 2015. Molecular Epidemiology of Methicillin-Resistant *Staphylococcus aureus* in Africa: a Systematic Review. *Frontiers in Microbiology* 6. <https://doi.org/10.3389/fmicb.2015.00348>
- Abimanyu, N., Murugesan, S., Krishnan, P., 2012. Emergence of Methicillin-Resistant *Staphylococcus aureus* ST239 with High-Level Mupirocin and Inducible Clindamycin Resistance in a Tertiary Care Center in Chennai, South India. *Journal of Clinical Microbiology* 50, 3412–3413. <https://doi.org/10.1128/jcm.01663-12>
- Abraham, E.P., Chain, E., Fletcher, C.M., Gardner, A.D., Heatley, N.G., Jennings, M.A., Florey, H.W., 1941. Further Observations on Penicillin. *The Lancet* 238, 177–189. [https://doi.org/10.1016/s0140-6736\(00\)72122-2](https://doi.org/10.1016/s0140-6736(00)72122-2)
- Abreu, C.I., Friedman, J., Andersen Woltz, V.L., Gore, J., 2019. Mortality Causes Universal Changes in Microbial Community Composition. *Nature Communications* 10. <https://doi.org/10.1038/s41467-019-09925-0>
- Addetia, A., Greninger, A.L., Adler, A., Yuan, S., Makhous, N., Qin, X., Zerr, D.M., 2019. A Novel, Widespread *AacA* Allele Results in Reduced Chlorhexidine Susceptibility in *Staphylococcus Epidermidis*. *Antimicrobial Agents and Chemotherapy* 63. <https://doi.org/10.1128/aac.02607-18>
- Aiken, A.M., Mutuku, I.M., Sabat, A.J., Akkerboom, V., Mwangi, J., Scott, J.A.G., Morpeth, S.C., Friedrich, A.W., Grundmann, H., 2014. Carriage of *Staphylococcus aureus* in Thika Level 5 Hospital, Kenya: a Cross-Sectional Study. *Antimicrobial Resistance and Infection Control* 3, 22. <https://doi.org/10.1186/2047-2994-3-22>
- Aires de Sousa (a), M., Bartzavali, C., Spiliopoulou, I., Santos Sanches, I., Crisostomo, M.I., de Lencastre, H., 2003. Two International Methicillin-Resistant *Staphylococcus aureus* Clones Endemic in a University Hospital in Patras, Greece. *Journal of Clinical Microbiology* 41, 2027–2032. <https://doi.org/10.1128/jcm.41.5.2027-2032.2003>
- Aires-de-Sousa, M., Correia, B., de Lencastre, H., 2008. Changing Patterns in Frequency of Recovery of Five Methicillin-Resistant *Staphylococcus aureus* Clones in Portuguese Hospitals: Surveillance over a 16-Year Period. *Journal of Clinical Microbiology* 46, 2912–2917. <https://doi.org/10.1128/jcm.00692-08>
- Aires de Sousa (b), M., Crisostomo, M.I., Santos Sanches, I., Wu, J.S., Fuzhong, J., Tomasz, A., de Lencastre, H., 2003. Frequent Recovery of a Single Clonal Type of Multidrug-Resistant *Staphylococcus aureus* from Patients in Two Hospitals in Taiwan and China. *Journal of Clinical Microbiology* 41, 159–163. <https://doi.org/10.1128/jcm.41.1.159-163.2003>
- Aires de Sousa, M., Miragaia, M., Santos Sanches, I., Avila, S., Adamson, I., Casagrande, S.T., Brandileone, M.C.C., Palacio, R., Dell'Acqua, L., Hortal, M., Camou, T., Rossi, A., Velazquez-Meza, M.E., Echaniz-Aviles, G., Solorzano-Santos, F., Heitmann, I., de Lencastre, H., 2001. Three-Year Assessment of Methicillin-Resistant *Staphylococcus aureus* Clones in Latin America from 1996 to

1998. *Journal of Clinical Microbiology* 39, 2197–2205. <https://doi.org/10.1128/jcm.39.6.2197-2205.2001>
- Akpaka, P.E., Kisson, S., Rutherford, C., Swanston, W.H., Jayaratne, P., 2007. Molecular Epidemiology of Methicillin-Resistant *Staphylococcus aureus* Isolates from Regional Hospitals in Trinidad and Tobago. *International Journal of Infectious Diseases* 11, 544–548. <https://doi.org/10.1016/j.ijid.2007.03.004>
- Alam, M.T., Read, T.D., Petit, R.A., III, Boyle-Vavra, S., Miller, L.G., Eells, S.J., Daum, R.S., David, M.Z., 2015. Transmission and Microevolution of USA300 MRSA in U.S. Households: Evidence from Whole-Genome Sequencing. *mBio* 6. <https://doi.org/10.1128/mbio.00054-15>
- Albrecht, N., Jatzwauk, L., Slickers, P., Ehricht, R., Monecke, S., 2011. Clonal Replacement of Epidemic Methicillin-Resistant *Staphylococcus aureus* Strains in a German University Hospital Over a Period of Eleven Years. *PLoS ONE* 6, e28189. <https://doi.org/10.1371/journal.pone.0028189>
- Alp, E., Klaassen, C.H.W., Doganay, M., Altoparlak, U., Aydin, K., Engin, A., Kuzucu, C., Ozakin, C., Ozinel, M.A., Turhan, O., Voss, A., 2009. MRSA Genotypes in Turkey: Persistence Over 10 years of a Single Clone of ST239. *Journal of Infection* 58, 433–438. <https://doi.org/10.1016/j.jinf.2009.04.006>
- Amaral, M.M., Coelho, L.R., Flores, R.P., Souza, R.R., Silva, M.C., Teixeira, L.A., Ferreira, B.T., Figueiredo, A.M.S., 2005. The Predominant Variant of the Brazilian Epidemic Clonal Complex of Methicillin Resistant *Staphylococcus aureus* has an Enhanced Ability to Produce Biofilm and to Adhere to and Invade Airway Epithelial Cells. *The Journal of Infectious Diseases* 192, 801–810. <https://doi.org/10.1086/432515>
- Amorim, M.L., Aires de Sousa, M., Sanches, I.S., Sá-Leão, R., Cabeda, J.M., Amorim, J.M., de Lencastre, H., 2002. Clonal and Antibiotic Resistance Profiles of Methicillin-Resistant *Staphylococcus aureus* (MRSA) from a Portuguese Hospital Over Time. *Microbial Drug Resistance* 8, 301–309. <https://doi.org/10.1089/10766290260469561>
- Amorim, M.L., Faria, N.A., Oliveira, D.C., Vasconcelos, C., Cabeda, J.C., Mendes, A.C., Calado, E., Castro, A.P., Ramos, M.H., Amorim, J.M., de Lencastre, H., 2007. Changes in the Clonal Nature and Antibiotic Resistance Profiles of Methicillin-Resistant *Staphylococcus aureus* Isolates Associated with Spread of the EMRSA-15 Clone in a Tertiary Care Portuguese Hospital. *Journal of Clinical Microbiology* 45, 2881–2888. <https://doi.org/10.1128/jcm.00603-07>
- Andersson, D.I., Hughes, D., 2010. Antibiotic Resistance and its Cost: is it Possible to Reverse Resistance? *Nature Reviews Microbiology* 8, 260–271. <https://doi.org/10.1038/nrmicro2319>
- Andersson, D.I., Levin, B.R., 1999. The Biological Cost of Antibiotic Resistance. *Current Opinion in Microbiology* 2, 489–493. [https://doi.org/10.1016/s1369-5274\(99\)00005-3](https://doi.org/10.1016/s1369-5274(99)00005-3)
- Archer, F.I., Adams, P.E., Schneiders, B.B., 2016. stratag: Anrpackage for manipulating, summarizing and analysing population genetic data. *Molecular Ecology Resources* 17, 5–11. <https://doi.org/10.1111/1755-0998.12559>
- Armand-Lefevre, L., Ruimy, R., Andremon, A., 2005. Clonal Comparison of *Staphylococcus aureus* Isolates from Healthy Pig Farmers, Human Controls, and Pigs. *Emerging Infectious Diseases* 11, 711–714. <https://doi.org/10.3201/eid1105.040866>
- Assefa, S., Keane, T.M., Otto, T.D., Newbold, C., Berriman, M., 2009. ABACAS: Algorithm-Based Automatic Contiguation of Assembled Sequences. *Bioinformatics* 25, 1968–1969. <https://doi.org/10.1093/bioinformatics/btp347>
- Aubry-Damon, H., Soussy, C.-J., Courvalin, P., 1998. Characterization of Mutations in the *rpoB* Gene That Confer Rifampin Resistance in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy* 42, 2590–2594. <https://doi.org/10.1128/aac.42.10.2590>

References

- Bagel, S., Hüllen, V., Wiedemann, B., Heisig, P., 1999. Impact of *gyrA* and *parC* Mutations on Quinolone Resistance, Doubling Time, and Supercoiling Degree of *Escherichia coli*. *Antimicrobial Agents and Chemotherapy* 43, 868–875. <https://doi.org/10.1128/aac.43.4.868>
- Baines, S.L., Holt, K.E., Schultz, M.B., Seemann, T., Howden, B.O., Jensen, S.O., van Hal, S.J., Coombs, G.W., Firth, N., Powell, D.R., Stinear, T.P., Howden, B.P., 2015. Convergent Adaptation in the Dominant Global Hospital Clone ST239 of Methicillin-Resistant *Staphylococcus aureus*. *mBio* 6. <https://doi.org/10.1128/mbio.00080-15>
- Bakthavatchalam, Y.D., Dwarakanathan, H.T, Munusamy, E., Jennifer, L., Veeraraghavan, B., 2019. A Distinct Geographic Variant of *sasX* in Methicillin-Resistant *Staphylococcus aureus* ST239 and ST368 Lineage from South India. *Microbial Drug Resistance* 25, 413–420. <https://doi.org/10.1089/mdr.2018.0292>
- Baltrus, D.A., 2013. Exploring the costs of horizontal gene transfer. *Trends in Ecology & Evolution* 28, 489–495. <https://doi.org/10.1016/j.tree.2013.04.002>
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a New Genome Assembly Algorithm and its Applications to Single-Cell Sequencing. *Journal of Computational Biology* 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Baranovich, T., Zaraket, H., Shabana, I.I., Nevzorova, V., Turcutyucov, V., Suzuki, H., 2010. Molecular Characterization and Susceptibility of Methicillin-Resistant and Methicillin-Susceptible *Staphylococcus aureus* Isolates from Hospitals and the Community in Vladivostok, Russia. *Clinical Microbiology and Infection* 16, 575–582. <https://doi.org/10.1111/j.1469-0691.2009.02891.x>
- Barber, M., Rozwadowska-Dowzenko, M., 1948. Infection by Penicillin-Resistant *Staphylococci*. *The Lancet* 252, 641–644. [https://doi.org/10.1016/s0140-6736\(48\)92166-7](https://doi.org/10.1016/s0140-6736(48)92166-7)
- Bartels, M.D., Kristoffersen, K., Boye, K., Westh, H., 2010. Rise and Subsequent Decline of Community-Associated Methicillin Resistant *Staphylococcus aureus* ST30-IVc in Copenhagen, Denmark through an effective search and destroy policy. *Clinical Microbiology and Infection* 16, 78–83. <https://doi.org/10.1111/j.1469-0691.2009.02829.x>
- Beale, M.A., Marks, M., Sahi, S.K., Tantalo, L.C., Nori, A.V., French, P., Lukehart, S.A., Marra, C.M., Thomson, N.R., 2019. Genomic Epidemiology of Syphilis Reveals Independent Emergence of Macrolide Resistance Across Multiple Circulating Lineages. *Nature Communications* 10. <https://doi.org/10.1038/s41467-019-11216-7>
- Bekkhoucha, S.N., Cady, A., Gautier, P., Itim, F., Donnio, P.-Y., 2008. A Portrait of *Staphylococcus aureus* from the Other Side of the Mediterranean Sea: Molecular Characteristics of Isolates from Western Algeria. *European Journal of Clinical Microbiology & Infectious Diseases* 28, 553–555. <https://doi.org/10.1007/s10096-008-0660-x>
- Bell, J.M., Turnidge, J.D., 2002. High Prevalence of Oxacillin-Resistant *Staphylococcus aureus* Isolates from Hospitalized Patients in Asia-Pacific and South Africa: Results from SENTRY Antimicrobial Surveillance Program, 1998-1999. *Antimicrobial Agents and Chemotherapy* 46, 879–881. <https://doi.org/10.1128/aac.46.3.880-882.2002>
- Bellais, S., Six, A., Fouet, A., Longo, M., Dmytruk, N., Glaser, P., Trieu-Cuot, P., Poyart, C., 2012. Capsular Switching in Group B *Streptococcus* CC17 Hypervirulent Clone: a Future Challenge for Polysaccharide Vaccine Development. *Journal of Infectious Diseases* 206, 1745–1752. <https://doi.org/10.1093/infdis/jis605>
- Benner, E.J, Kayser, F.H., 1968. Growing Clinical Significance of Methicillin-Resistant *Staphylococcus aureus*. *The Lancet* 292, 741–744. [https://doi.org/10.1016/s0140-6736\(68\)90947-1](https://doi.org/10.1016/s0140-6736(68)90947-1)

- Bennett, M.R., Bombardi, R.G., Kose, N., Parrish, E.H., Nagel, M.B., Petit, R.A., III, Read, T.D., Schey, K.L., Thomsen, I.P., Skaar, E.P., Crowe, J.E., Jr, 2018. Human mAbs to *Staphylococcus aureus* IsdA Provide Protection Through Both Heme-Blocking and Fc-Mediated Mechanisms. *The Journal of Infectious Diseases* 219, 1264–1273. <https://doi.org/10.1093/infdis/jiy635>
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a Flexible Trimmer For Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Boswihi, S.S., Udo, E.E., Al-Sweih, N., 2016. Shifts in the Clonal Distribution of Methicillin-Resistant *Staphylococcus aureus* in Kuwait Hospitals: 1992-2010. *PLOS ONE* 11, e0162744. <https://doi.org/10.1371/journal.pone.0162744>
- Botelho, A.M.N., Costa, M.O.C., Beltrame, C.O., Ferreira, F.A., Côrtes, M.F., Bandeira, P.T., Lima, N.C.B., Souza, R.C., Almeida, L.G.P., Vasconcelos, A.T.R., Nicolás, M.F., Figueiredo, A.M.S., 2016. Complete Genome Sequence of An Agr-Dysfunctional Variant of the ST239 Lineage of the Methicillin-Resistant *Staphylococcus aureus* Strain GV69 from Brazil. *Standards in Genomic Sciences* 11. <https://doi.org/10.1186/s40793-016-0154-x>
- Bowers, J.R., Driebe, E.M., Albrecht, V., McDougal, L.K., Granade, M., Roe, C.C., Lemmer, D., Rasheed, J.K., Engelthaler, D.M., Keim, P., Limbago, B.M., 2018. Improved Subtyping of *Staphylococcus aureus* Clonal Complex 8 Strains Based on Whole-Genome Phylogenetic Analysis. *mSphere* 3. <https://doi.org/10.1128/msphere.00464-17>
- Bradley, P., den Bakker, H.C., Rocha, E.P.C., McVean, G., Iqbal, Z., 2019. Ultrafast Search of All Deposited Bacterial and Viral Genomic Data. *Nature Biotechnology* 37, 152–159. <https://doi.org/10.1038/s41587-018-0010-1>
- Bradley, P., Gordon, N.C., Walker, T.M., Dunn, L., Heys, S., Huang, B., Earle, S., Pankhurst, L.J., Anson, L., de Cesare, M., Piazza, P., Votintseva, A.A., Golubchik, T., Wilson, D.J., Wyllie, D.H., Diel, R., Niemann, S., Feuerriegel, S., Kohl, T.A., Ismail, N., Omar, S.V., Smith, E.G., Buck, D., McVean, G., Walker, A.S., Peto, T.E.A., Crook, D.W., Iqbal, Z., 2015. Rapid Antibiotic-Resistance Predictions from Genome Sequence Data For *Staphylococcus aureus* and *Mycobacterium Tuberculosis*. *Nature Communications* 6. <https://doi.org/10.1038/ncomms10063>
- Branchu, P., Charity, O.J., Bawn, M., Thilliez, G., Dallman, T.J., Petrovska, L., Kingsley, R.A., 2019. SGI-4 in Monophasic *Salmonella Typhimurium* ST34 is a Novel ICE That Enhances Resistance to Copper. *Frontiers in Microbiology* 10. <https://doi.org/10.3389/fmicb.2019.01118>
- Breurec, S., Zriouil, S.B., Fall, C., Boisier, P., Brisse, S., Djibo, S., Etienne, J., Fonkoua, M.C., Perrier-Gros-Claude, J.D., Pouillot, R., Ramarokoto, C.E., Randrianirina, F., Tall, A., Thiberge, J.M., Laurent, F., Garin, B., 2011. Epidemiology of Methicillin-Resistant *Staphylococcus aureus* Lineages in Five Major African Towns: Emergence and Spread of Atypical Clones. *Clinical Microbiology and Infection* 17, 160–165. <https://doi.org/10.1111/j.1469-0691.2010.03219.x>
- Brochet, M., Rusniok, C., Couve, E., Dramsi, S., Poyart, C., Trieu-Cuot, P., Kunst, F., Glaser, P., 2008. Shaping a Bacterial Genome by Large Chromosomal Replacements, the Evolutionary History of *Streptococcus Agalactiae*. *Proceedings of the National Academy of Sciences* 105, 15961–15966. <https://doi.org/10.1073/pnas.0803654105>
- Brockhurst, M.A., Harrison, E., Hall, J.P.J., Richards, T., McNally, A., MacLean, C., 2019. The Ecology and Evolution of Pangenomes. *Current Biology* 29, R1094–R1103. <https://doi.org/10.1016/j.cub.2019.08.012>
- Brolund, A., Sundqvist, M., Kahlmeter, G., Grape, M., 2010. Molecular Characterisation of Trimethoprim Resistance in *Escherichia coli* and *Klebsiella Pneumoniae* During a Two Year Intervention on Trimethoprim Use. *PLoS ONE* 5, e9233. <https://doi.org/10.1371/journal.pone.0009233>

References

- Brown, C.J., Sen, D., Yano, H., Bauer, M.L., Rogers, L.M., Van der Auwera, G.A., Top, E.M., 2013. Diverse Broad-Host-Range Plasmids from Freshwater Carry Few Accessory Genes. *Applied and Environmental Microbiology* 79, 7684–7695. <https://doi.org/10.1128/aem.02252-13>
- Budimir, A., Deurenberg, R.H., Bošnjak, Z., Stobberingh, E.E., Cetkovic, H., Kalenic, S., 2010. A Variant of the Southern German Clone of Methicillin-Resistant *Staphylococcus aureus* is Predominant in Croatia. *Clinical Microbiology and Infection* 16, 1077–1083. <https://doi.org/10.1111/j.1469-0691.2009.03042.x>
- Cameron, D.R., Mortin, L.I., Rubio, A., Mylonakis, E., Moellering, R.C., Jr, Eliopoulos, G.M., Peleg, A.Y., 2015. Impact of Daptomycin Resistance on *Staphylococcus aureus* Virulence. *Virulence* 6, 127–131. <https://doi.org/10.1080/21505594.2015.1011532>
- Campanile, F., Cafiso, V., Cascone, C., Giannino, V., Di Marco, O., Stefani, S., 2001. Clonal Diffusion and Evolution of Meca and Tn554 Polymorphisms in Methicillin-Resistant *Staphylococcus aureus* in Italy. *Le Infezioni in Medicina* 9, 30–38.
- CARD, 2019, the Comprehensive Antibiotic Resistance Database, CARD, accessed Aug 2019 <<https://card.mcmaster.ca/home>>
- Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.-A., Barrell, B.G., Parkhill, J., 2005. ACT: the Artemis Comparison Tool. *Bioinformatics* 21, 3422–3423. <https://doi.org/10.1093/bioinformatics/bti553>
- Castillo-Ramírez, S., Corander, J., Martinen, P., Aldeljawi, M., Hanage, W.P., Westh, H., Boye, K., Gulay, Z., Bentley, S.D., Parkhill, J., Holden, M.T., Feil, E.J., 2012. Phylogeographic variation in recombination rates within a global clone of methicillin-resistant *Staphylococcus aureus*. *Genome Biology* 13, R126. <https://doi.org/10.1186/gb-2012-13-12-r126>
- Castillo-Ramírez, S., Harris, S.R., Holden, M.T.G., He, M., Parkhill, J., Bentley, S.D., Feil, E.J., 2011. The Impact of Recombination on dN/dS within Recently Emerged Bacterial Clones. *PLoS Pathogens* 7, e1002129. <https://doi.org/10.1371/journal.ppat.1002129>
- CDC, 2001, Methicillin-Resistant *Staphylococcus aureus* Skin or Soft Tissue Infections in a State Prison --- Mississippi, 2000, accessed December 2019. <<https://www.ncbi.nlm.nih.gov/pubmed/11699844>>
- CDC, 2002, *Staphylococcus aureus* Resistant to Vancomycin --- United States, 2002, accessed December 2019. <<https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5126a1.htm>>
- CDC, 2019, Antibiotic resistance threats in the United States, 2019, accessed December 2019. <<https://www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf>>
- Centers for Disease Control and Prevention Public Health Image Library #11157, 2001, accessed December 2019. <https://phil.cdc.gov/details.aspx?pid=11157>
- Centre for Genomic Pathogen Surveillance, 2018, Pathogenwatch, Wellcome Sanger Institute, accessed Aug 2018 <<https://pathogen.watch/>>
- Cha, H.Y., Moon, D.C., Choi, C.H., Oh, J.Y., Jeong, Y.S., Lee, Y.C., Seol, S.Y., Cho, D.T., Chang, H.-H., Kim, S.-W., Lee, J.C., 2005. Prevalence of the ST239 Clone of Methicillin-Resistant *Staphylococcus aureus* and Differences in Antimicrobial Susceptibilities of ST239 and ST5 Clones Identified in a Korean Hospital. *Journal of Clinical Microbiology* 43, 3610–3614. <https://doi.org/10.1128/jcm.43.8.3610-3614.2005>
- Chambers, H.F., DeLeo, F.R., 2009. Waves of resistance: *Staphylococcus aureus* in the antibiotic era. *Nature Reviews Microbiology* 7, 629–641. <https://doi.org/10.1038/nrmicro2200>

- Charalampous, T., Kay, G.L., Richardson, H., Aydin, A., Baldan, R., Jeanes, C., Rae, D., Grundy, S., Turner, D.J., Wain, J., Leggett, R.M., Livermore, D.M., O'Grady, J., 2019. Nanopore Metagenomics Enables Rapid Clinical Diagnosis of Bacterial Lower Respiratory Infection. *Nature Biotechnology* 37, 783–792. <https://doi.org/10.1038/s41587-019-0156-5>
- Chen, F.-J., Lauderdale, T.-L., Wang, L.-S., Huang, I.-W., 2013. Complete Genome Sequence of *Staphylococcus aureus* Z172, a Vancomycin-Intermediate and Daptomycin-Nonsusceptible Methicillin-Resistant Strain Isolated in Taiwan. *Genome Announcements* 5, 1(6). <https://doi.org/10.1128/genomea.01011-13>
- Chen, J., Quiles-Puchalt, N., Chiang, Y.N., Bacigalupe, R., Fillol-Salom, A., Chee, M.S.J., Fitzgerald, J.R., Penadés, J.R., 2018. Genome Hypermobility by Lateral Transduction. *Science* 362, 207–212. <https://doi.org/10.1126/science.aat5867>
- Chen, L., Mathema, B., Pitout, J.D.D., DeLeo, F.R., Kreiswirth, B.N., 2014. Epidemic *Klebsiella pneumoniae* ST258 is a Hybrid Strain. *mBio* 5. <https://doi.org/10.1128/mbio.01355-14>
- Chen, L., Mediavilla, J.R., Smyth, D.S., Chavda, K.D., Ionescu, R., Roberts, R.B., Robinson, D.A., Kreiswirth, B.N., 2010. Identification of a Novel Transposon (Tn6072) and a Truncated *Staphylococcal* Cassette Chromosome *mec* Element in Methicillin-Resistant *Staphylococcus aureus* ST239. *Antimicrobial Agents and Chemotherapy* 54, 3347–3354. <https://doi.org/10.1128/aac.00001-10>
- Chiang, Y.N., Penadés, J.R., Chen, J., 2019. Genetic Transduction by Phages and Chromosomal Islands: the New and Noncanonical. *PLOS Pathogens* 15, e1007878. <https://doi.org/10.1371/journal.ppat.1007878>
- Chongtrakool, P., Ito, T., Ma, X.X., Kondo, Y., Trakulsomboon, S., Tiensasitorn, C., Jamklang, M., Chavalit, T., Song, J.-H., Hiramatsu, K., 2006. *Staphylococcal* Cassette Chromosome *mec* (SCCmec) Typing of Methicillin-Resistant *Staphylococcus aureus* Strains Isolated in 11 Asian Countries: a Proposal for a New Nomenclature for SCCmec Elements. *Antimicrobial Agents and Chemotherapy* 50, 1001–1012. <https://doi.org/10.1128/aac.50.3.1001-1012.2006>
- Choo, E.J., Chambers, H.F., 2016. Treatment of Methicillin-Resistant *Staphylococcus aureus* Bacteremia. *Infection & Chemotherapy* 48, 267. <https://doi.org/10.3947/ic.2016.48.4.267>
- Chua, K., Laurent, F., Coombs, G., Grayson, M.L., Howden, B.P., 2011. Not Community-Associated Methicillin-Resistant *Staphylococcus aureus* (CA-MRSA)! a Clinician's Guide to Community MRSA - its Evolving Antimicrobial Resistance and Implications for Therapy. *Clinical Infectious Diseases* 52, 99–114. <https://doi.org/10.1093/cid/ciq067>
- Cirlan, M., Saad, M., Coman, G., Bilal, N., Elbashier, A., Kreft, D., Snijders, S., Vanleeuwen, W., Vanbelkum, A., 2005. International Spread of Major Clones of Methicillin Resistant *Staphylococcus aureus*: Nosocomial Endemicity of Multi Locus Sequence Type 239 in Saudi Arabia and Romania. *Infection, Genetics and Evolution* 5, 335–339. <https://doi.org/10.1016/j.meegid.2004.09.005>
- Clausen, P.T.L.C., Aarestrup, F.M., Lund, O., 2018. Rapid and Precise Alignment of Raw Reads Against Redundant Databases with KMA. *BMC Bioinformatics* 19. <https://doi.org/10.1186/s12859-018-2336-6>
- Coimbra, M.V.D.S., Figueiredo, A.M.S., Famiglietti, A., Predari, S.C., Teixeira, L.A., Vay, C., Castello, L., Klan, L., Ramos, R.L.B., 2000. Spread of the Brazilian Epidemic Clone of a Multiresistant MRSA in Two Cities in Argentina. *Journal of Medical Microbiology* 49, 187–192. <https://doi.org/10.1099/0022-1317-49-2-187>
- Colby, L.A., Quenee, L.E., Zitzow, L.A., 2017. Considerations for Infectious Disease Research Studies Using Animals. *Comparative Medicine* 67, 222–231.

References

- Collins, J., Rudkin, J., Recker, M., Pozzi, C., O’Gara, J.P., Massey, R.C., 2010. Offsetting Virulence and Antibiotic Resistance Costs by MRSA. *The ISME Journal* 4, 577–584. <https://doi.org/10.1038/ismej.2009.151>
- Comandatore, F., Sasser, D., Bayliss, S.C., Scaltriti, E., Gaiarsa, S., Cao, X., Gales, A.C., Saito, R., Pongolini, S., Brisse, S., Feil, E.J., Bandi, C., 2019. Gene Composition as a Potential Barrier to Large Recombinations in the Bacterial Pathogen *Klebsiella pneumoniae*. *Genome Biology and Evolution* 11, 3240–3251. <https://doi.org/10.1093/gbe/evz236>
- Conceição, T., Aires-de-Sousa, M., Füzi, M., Tóth, Á., Pászti, J., Ungvári, E., van Leeuwen, W.B., van Belkum, A., Grundmann, H., de Lencastre, H., 2007. Replacement of Methicillin-Resistant *Staphylococcus aureus* Clones in Hungary Over Time: a 10-Year Surveillance Study. *Clinical Microbiology and Infection* 13, 971–979. <https://doi.org/10.1111/j.1469-0691.2007.01794.x>
- Cong, Y., Chan, Y., Ragan, M.A., 2016. A Novel Alignment-Free Method For Detection of Lateral Genetic Transfer Based on TF-IDF. *Scientific Reports* 6, 30308. <https://doi.org/10.1038/srep30308>
- Cookson, B.D., Phillips, I., 1988. Epidemic Methicillin-Resistant *Staphylococcus aureus*. *Journal of Antimicrobial Chemotherapy* 21, 57–65. https://doi.org/10.1093/jac/21.suppl_c.57
- Coombs, G.W., Pearson, J.C., Nimmo, G.R., Collignon, P.J., Bell, J.M., McLaws, M.-L., Christiansen, K.J., Turnidge, J.D., 2013. Antimicrobial Susceptibility of *Staphylococcus aureus* and Molecular Epidemiology of Methicillin-Resistant *S. Aureus* Isolated from Australian Hospital Inpatients: Report from the Australian Group on Antimicrobial Resistance 2011 *Staphylococcus aureus* Surveillance Programme. *Journal of Global Antimicrobial Resistance* 1, 149–156. <https://doi.org/10.1016/j.jgar.2013.04.005>
- Coppens, J., Xavier, B.B., Vanjari, L., De Backer, S., Lammens, C., Vemu, L., Goossens, H., Malhotra-Kumar, S., 2018. Novel Composite SCCmec Type III Element in ST239 MRSA Isolated from An Indian Hospital. *Journal of Antimicrobial Chemotherapy* 74, 264–266. <https://doi.org/10.1093/jac/dky399>
- Costa, M.O.C., Beltrame, C.O., Ferreira, F.A., Botelho, A.M.N., Lima, N.C.B., Souza, R.C., de Almeida, L.G.P., Vasconcelos, A.T.R., Nicolas, M.F., Figueiredo, A.M.S., 2013. Complete Genome Sequence of a Variant of the Methicillin-Resistant *Staphylococcus aureus* ST239 Lineage, Strain BMB9393, Displaying Superior Ability To Accumulate ica-Independent Biofilm. *Genome Announcements* 1, 8(4). <https://doi.org/10.1128/genomea.00576-13>
- Couto, I., De Lencastre, H., Severina, E., Kloos, W., Webster, J.A., Hubner, R.J., Sanches, I.S., Tomasz, A., 1996. Ubiquitous Presence of a *mecA* Homologue in Natural Isolates of *Staphylococcus sciuri*. *Microbial Drug Resistance* 2, 377–391. <https://doi.org/10.1089/mdr.1996.2.377>
- Cox, R.A., Conquest, C., Mallaghan, C., Marples, R.R., 1995. A Major Outbreak of Methicillin-Resistant *Staphylococcus aureus* Caused by a New Phage-Type (EMRSA-16). *Journal of Hospital Infection* 29, 87–106. [https://doi.org/10.1016/0195-6701\(95\)90191-4](https://doi.org/10.1016/0195-6701(95)90191-4)
- Croucher, N.J., Klugman, K.P., 2014. The Emergence of Bacterial “Hopeful Monsters.” *mBio* 5. <https://doi.org/10.1128/mbio.01550-14>
- Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D., Parkhill, J., Harris, S.R., 2014. Rapid Phylogenetic Analysis of Large Samples of Recombinant Bacterial Whole Genome Sequences Using Gubbins. *Nucleic Acids Research* 43, e15–e15. <https://doi.org/10.1093/nar/gku1196>
- Cuevas, O., Cercenado, E., Bouza, E., Castellares, C., Trincado, P., Cabrera, R., Vindel, A., 2007. Molecular Epidemiology of Methicillin-Resistant *Staphylococcus aureus* in Spain: a Multicentre Prevalence Study (2002). *Clinical Microbiology and Infection* 13, 250–256. <https://doi.org/10.1111/j.1469-0691.2006.01672.x>

- D'Souza, N., Rodrigues, C., Mehta, A., 2010. Molecular Characterization of Methicillin-Resistant *Staphylococcus aureus* with Emergence of Epidemic Clones of Sequence Type (ST) 22 and ST 772 in Mumbai, India. *Journal of Clinical Microbiology* 48, 1806–1811. <https://doi.org/10.1128/jcm.01867-09>
- Dai, Y., Liu, J., Guo, W., Meng, H., Huang, Q., He, L., Gao, Q., Lv, H., Liu, Y., Wang, Y., Wang, H., Liu, Q., Li, M., 2019. Decreasing Methicillin-Resistant *Staphylococcus aureus* (MRSA) Infections is Attributable To the Disappearance of Predominant MRSA ST239 Clones, Shanghai, 2008–2017. *Emerging Microbes & Infections* 8, 471–478. <https://doi.org/10.1080/22221751.2019.1595161>
- Dall'Antonia, M., Coen, P., Wilks, M., Whiley, A., Millar, M., 2005. Competition Between Methicillin-Sensitive and -Resistant in the Anterior Nares. *Journal of Hospital Infection* 61, 62–67. <https://doi.org/10.1016/j.jhin.2005.01.008>
- David, M.Z., Daum, R.S., 2010. Community-Associated Methicillin-Resistant *Staphylococcus aureus*: Epidemiology and Clinical Consequences of an Emerging Epidemic. *Clinical Microbiology Reviews* 23, 616–687. <https://doi.org/10.1128/cmr.00081-09>
- Davis, K.E.R., Joseph, S.J., Janssen, P.H., 2005. Effects of Growth Medium, Inoculum Size, and Incubation Time on Culturability and Isolation of Soil Bacteria. *Applied and Environmental Microbiology* 71, 826–834. <https://doi.org/10.1128/aem.71.2.826-834.2005>
- De Backer, S., Xavier, B.B., Vanjari, L., Coppens, J., Lammens, C., Vemu, L., Carevic, B., Hryniewicz, W., Jorens, P., Kumar-Singh, S., Lee, A., Harbarth, S., Schrenzel, J., Tacconelli, E., Goossens, H., Malhotra-Kumar, S., 2019. Remarkable Geographical Variations Between India and Europe in Carriage of the Staphylococcal Surface Protein-Encoding SasX/SesI and in the Population Structure of Methicillin-Resistant *Staphylococcus aureus* Belonging To Clonal Complex 8. *Clinical Microbiology and Infection* 25, 628.e1–628.e7. <https://doi.org/10.1016/j.cmi.2018.07.024>
- DeLeo, F.R., Kennedy, A.D., Chen, L., Wardenburg, J.B., Kobayashi, S.D., Mathema, B., Braughton, K.R., Whitney, A.R., Villaruz, A.E., Martens, C.A., Porcella, S.F., McGavin, M.J., Otto, M., Musser, J.M., Kreiswirth, B.N., 2011. Molecular Differentiation of Historic Phage-Type 80/81 and Contemporary Epidemic *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences* 108, 18091–18096. <https://doi.org/10.1073/pnas.1111084108>
- Dettman, J.R., Sztepanacz, J.L., Kassen, R., 2016. The properties of spontaneous mutations in the opportunistic pathogen *Pseudomonas aeruginosa*. *BMC Genomics* 17. <https://doi.org/10.1186/s12864-015-2244-3>
- Didelot X., 2019, Application of BactDating on *S. Aureus* ST239, Didelot X., accessed Aug 2019 <<https://xavierdidelot.github.io/BactDating/articles/Staph.html>>
- Didelot, X., Croucher, N.J., Bentley, S.D., Harris, S.R., Wilson, D.J., 2018. Bayesian Inference of Ancestral Dates on Bacterial Phylogenetic Trees. *Nucleic Acids Research* 46, e134–e134. <https://doi.org/10.1093/nar/gky783>
- Didelot, X., Wilson, D.J., 2015. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLOS Computational Biology* 11, e1004041. <https://doi.org/10.1371/journal.pcbi.1004041>
- Diep, B.A., Stone, G.G., Basuino, L., Graber, C.J., Miller, A., Etages, S. des, Jones, A., Palazzolo Ballance, A.M., Perdreau Remington, F., Sensabaugh, G.F., DeLeo, F.R., Chambers, H.F., 2008. The Arginine Catabolic Mobile Element and Staphylococcal Chromosomal Cassette Mec Linkage: Convergence of Virulence and Resistance in the USA300 Clone of Methicillin Resistant *Staphylococcus aureus*. *The Journal of Infectious Diseases* 197, 1523–1530. <https://doi.org/10.1086/587907>

References

- Drummond, A., Pybus, O.G., Rambaut, A., 2003. Inference of Viral Evolutionary Rates from Molecular Sequences, in: *Advances in Parasitology* 54. Elsevier, pp. 331–358. [https://doi.org/10.1016/s0065-308x\(03\)54008-8](https://doi.org/10.1016/s0065-308x(03)54008-8)
- Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29, 1969–1973. <https://doi.org/10.1093/molbev/mss075>
- Duckett, S., 1999. Ernest Duchesne and the concept of fungal antibiotic therapy. *The Lancet* 354, 2068–2071. [https://doi.org/10.1016/s0140-6736\(99\)03162-1](https://doi.org/10.1016/s0140-6736(99)03162-1)
- Earle, S.G., Wu, C.-H., Charlesworth, J., Stoesser, N., Gordon, N.C., Walker, T.M., Spencer, C.C.A., Iqbal, Z., Clifton, D.A., Hopkins, K.L., Woodford, N., Smith, E.G., Ismail, N., Llewelyn, M.J., Peto, T.E., Crook, D.W., McVean, G., Walker, A.S., Wilson, D.J., 2016. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nature Microbiology* 1, 16041. <https://doi.org/10.1038/nmicrobiol.2016.41>
- Edgeworth, J.D., Yadegarfar, G., Pathak, S., Batra, R., Cockfield, J.D., Wyncoll, D., Beale, R., Lindsay, J.A., 2007. An Outbreak in an Intensive Care Unit of a Strain of Methicillin-Resistant *Staphylococcus aureus* Sequence Type 239 Associated with an Increased Rate of Vascular Access Device--Related Bacteremia. *Clinical Infectious Diseases* 44, 493–501. <https://doi.org/10.1086/511034>
- Egyir, B., Guardabassi, L., Sørum, M., Nielsen, S.S., Kolekang, A., Frimpong, E., Addo, K.K., Newman, M.J., Larsen, A.R., 2014. Molecular Epidemiology and Antimicrobial Susceptibility of Clinical *Staphylococcus aureus* from Healthcare Institutions in Ghana. *PLoS ONE* 9, e89716. <https://doi.org/10.1371/journal.pone.0089716>
- Elena, S.F., Lenski, R.E., 2003. Evolution Experiments with Microorganisms: the Dynamics and Genetic Bases of Adaptation. *Nature Reviews Genetics* 4, 457–469. <https://doi.org/10.1038/nrg1088>
- Ender, M., McCallum, N., Adhikari, R., Berger-Bachi, B., 2004. Fitness Cost of SCCmec and Methicillin Resistance Levels in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy* 48, 2295–2297. <https://doi.org/10.1128/aac.48.6.2295-2297.2004>
- Enright, M., 2003. The Evolution of a Resistant Pathogen – the Case of MRSA. *Current Opinion in Pharmacology* 3, 474–479. [https://doi.org/10.1016/s1471-4892\(03\)00109-7](https://doi.org/10.1016/s1471-4892(03)00109-7)
- Enright, M.C., Day, N.P., Davies, C.E., Peacock, S.J., Spratt, B.G., 2000. Multilocus Sequence Typing For Characterization of Methicillin-Resistant and Methicillin-Susceptible Clones of *Staphylococcus aureus*. *Journal of Clinical Microbiology* 38, 1008–1015.
- Enright, M.C., Robinson, D.A., Randle, G., Feil, E.J., Grundmann, H., Spratt, B.G., 2002. The Evolutionary History of Methicillin-Resistant *Staphylococcus aureus* (MRSA). *Proceedings of the National Academy of Sciences* 99, 7687–7692. <https://doi.org/10.1073/pnas.122108599>
- Espedido, B.A., Steen, J.A., Barbagiannakos, T., Mercer, J., Paterson, D.L., Grimmond, S.M., Cooper, M.A., Gosbell, I.B., van Hal, S.J., Jensen, S.O., 2012. Carriage of An ACME II Variant May Have Contributed To Methicillin-Resistant *Staphylococcus aureus* Sequence Type 239-Like Strain Replacement in Liverpool Hospital, Sydney, Australia. *Antimicrobial Agents and Chemotherapy* 56, 3380–3383. <https://doi.org/10.1128/aac.00013-12>
- Essa, Z.I., Essack, S.Y., Connolly, C., 2009. *Staphylococcus aureus* from Public Hospitals in Kwazulu—Natal, South Africa—Infection Detection and Strain—Typing. *Southern African Journal of Epidemiology and Infection* 24, 4–7. <https://doi.org/10.1080/10158782.2009.11441332>
- The European Committee on Antimicrobial Susceptibility Testing, 2014, Breakpoint tables for interpretation of MICs and zone diameters version 4.0, accessed Aug 2019 <<http://www.eucast.org>>

- Feil, E.J., Nickerson, E.K., Chantratita, N., Wuthiekanun, V., Srisomang, P., Cousins, R., Pan, W., Zhang, G., Xu, B., Day, N.P.J., 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, 1520–1522. <https://doi.org/10.1128/jcm.02238-07>
- Fernandez, S., de Vedia, L., Lopez Furst, M.J., Gardella, N., Di Gregorio, S., Ganaha, M.C., Prieto, S., Carbone, E., Lista, N., Rotrying, F., Stryjewski, M.E., Mollerach, M., 2013. Methicillin-resistant *Staphylococcus aureus* ST30-SCCmec IVc Clone As the Major Cause of Community-Acquired Invasive Infections in Argentina. *Infection, Genetics and Evolution* 14, 401–405. <https://doi.org/10.1016/j.meegid.2012.12.018>
- Fleming, A., 1929. on the Antibacterial Action of Cultures of a *Penicillium*, with Special Reference To Their Use in the Isolation of *B. influenzae*. *British journal of experimental pathology* 10, 226–236.
- Fornelos, N., Browning, D.F., Pavlin, A., Podlesek, Z., Hodnik, V., Salas, M., Butala, M., 2018. Lytic Gene Expression in the Temperate Bacteriophage GIL01 is Activated by a Phage-Encoded Lexa Homologue. *Nucleic Acids Research* 46, 9432–9443. <https://doi.org/10.1093/nar/gky646>
- Foster, K.R., Bell, T., 2012. Competition, Not Cooperation, Dominates Interactions among Culturable Microbial Species. *Current Biology* 22, 1845–1850. <https://doi.org/10.1016/j.cub.2012.08.005>
- Foster, P.L., Lee, H., Popodi, E., Townes, J.P., Tang, H., 2015. Determinants of spontaneous mutation in the bacterium *Escherichia coli* as revealed by whole-genome sequencing. *Proceedings of the National Academy of Sciences* 112, E5990–E5999. <https://doi.org/10.1073/pnas.1512136112>
- Foster, T.J., 2017. Antibiotic Resistance in *Staphylococcus aureus*. *Current Status and Future Prospects*. *FEMS Microbiology Reviews* 41, 430–449. <https://doi.org/10.1093/femsre/fux007>
- Frost, I., Smith, W.P.J., Mitri, S., Millan, A.S., Davit, Y., Osborne, J.M., Pitt-Francis, J.M., MacLean, R.C., Foster, K.R., 2018. Cooperation, Competition and Antibiotic Resistance in Bacterial Colonies. *The ISME Journal* 12, 1582–1593. <https://doi.org/10.1038/s41396-018-0090-4>
- Fu, Z., Ma, Y., Chen, C., Guo, Y., Hu, F., Liu, Y., Xu, X., Wang, M., 2016. Prevalence of Fosfomycin Resistance and Mutations in *murA*, *glpT*, and *uhpT* in Methicillin-Resistant *Staphylococcus aureus* Strains Isolated from Blood and Cerebrospinal Fluid Samples. *Frontiers in Microbiology* 6, 1544. <https://doi.org/10.3389/fmicb.2015.01544>
- Fuchs, S., Mehlan, H., Bernhardt, J., Hennig, A., Michalik, S., Surmann, K., Pané-Farré, J., Giese, A., Weiss, S., Backert, L., Herbig, A., Nieselt, K., Hecker, M., Völker, U., Mäder, U., 2018. Aureo Wiki—the Repository of the *Staphylococcus aureus* Research and Annotation Community. *International Journal of Medical Microbiology* 308, 558–568. <https://doi.org/10.1016/j.ijmm.2017.11.011>
- Galan, J.-C., Tato, M., Baquero, M.-R., Turrientes, C., Baquero, F., Martinez, J.-L., 2004. Fosfomycin and Rifampin Disk Diffusion Tests for Detection of *Escherichia coli* Mutator Strains. *Journal of Clinical Microbiology* 42, 4310–4312. <https://doi.org/10.1128/jcm.42.9.4310-4312.2004>
- Ghaznavi-Rad, E., Shamsudin, M.N., Sekawi, Z., Khoon, L.Y., Aziz, M.N., Hamat, R.A., Othman, N., Chong, P.P., van Belkum, A., Ghasemzadeh-Moghaddam, H., Neela, V., 2010. Predominance and Emergence of Clones of Hospital-Acquired Methicillin-Resistant *Staphylococcus aureus* in Malaysia. *Journal of Clinical Microbiology* 48, 867–872. <https://doi.org/10.1128/jcm.01112-09>
- Glaser, P., Martins-Simões, P., Villain, A., Barbier, M., Tristan, A., Bouchier, C., Ma, L., Bes, M., Laurent, F., Guillemot, D., Wirth, T., Vandenesch, F., 2016. Demography and Intercontinental Spread of the USA300 Community-Acquired Methicillin-Resistant *Staphylococcus aureus* Lineage. *mBio* 7, e02183-15. <https://doi.org/10.1128/mbio.02183-15>
- Good, B.H., McDonald, M.J., Barrick, J.E., Lenski, R.E., Desai, M.M., 2017. The Dynamics of Molecular Evolution Over 60,000 Generations. *Nature* 551, 45–50. <https://doi.org/10.1038/nature24287>

References

- Goudarzi, M., Bahramian, M., Satarzadeh Tabrizi, M., Udo, E.E., Figueiredo, A.M.S., Fazeli, M., Goudarzi, H., 2017. Genetic Diversity of Methicillin Resistant *Staphylococcus aureus* Strains Isolated from Burn Patients in Iran: ST239-SCC mec III/t037 Emerges As the Major Clone. *Microbial Pathogenesis* 105, 1–7. <https://doi.org/10.1016/j.micpath.2017.02.004>
- Granato, E.T., Meiller-Legrand, T.A., Foster, K.R., 2019. The Evolution and Ecology of Bacterial Warfare. *Current Biology* 29, R521–R537. <https://doi.org/10.1016/j.cub.2019.04.024>
- Gray, R.R., Tatem, A.J., Johnson, J.A., Alekseyenko, A.V., Pybus, O.G., Suchard, M.A., Salemi, M., 2010. Testing Spatiotemporal Hypothesis of Bacterial Evolution Using Methicillin-Resistant *Staphylococcus aureus* ST239 Genome-wide Data within a Bayesian Framework. *Molecular Biology and Evolution* 28, 1593–1603. <https://doi.org/10.1093/molbev/msq319>
- Grubb, W.B., 1998. Genetics of MRSA. *Reviews in Medical Microbiology* 9.
- Grundmann, H., Aanensen, D.M., van den Wijngaard, C.C., Spratt, B.G., Harmsen, D., Friedrich, A.W., 2010. Geographic Distribution of *Staphylococcus aureus* Causing Invasive Infections in Europe: A Molecular-Epidemiological Analysis. *PLoS Medicine* 7, e1000215. <https://doi.org/10.1371/journal.pmed.1000215>
- Grundmann, H., Aires-de-Sousa, M., Boyce, J., Tiemersma, E., 2006. Emergence and Resurgence of Meticillin-Resistant *Staphylococcus aureus* As a Public-Health Threat. *The Lancet* 368, 874–885. [https://doi.org/10.1016/s0140-6736\(06\)68853-3](https://doi.org/10.1016/s0140-6736(06)68853-3)
- Guzmán-Blanco, M., Mejía, C., Isturiz, R., Alvarez, C., Bavestrello, L., Gotuzzo, E., Labarca, J., Luna, C.M., Rodríguez-Noriega, E., Salles, M.J.C., Zurita, J., Seas, C., 2009. Epidemiology of Meticillin-Resistant *Staphylococcus aureus* (MRSA) in Latin America. *International Journal of Antimicrobial Agents* 34, 304–308. <https://doi.org/10.1016/j.ijantimicag.2009.06.005>
- Gyles, C., Boerlin, P., 2013. Horizontally Transferred Genetic Elements and Their Role in Pathogenesis of Bacterial Disease. *Veterinary Pathology* 51, 328–340. <https://doi.org/10.1177/0300985813511131>
- Haaber, J., Leisner, J.J., Cohn, M.T., Catalan-Moreno, A., Nielsen, J.B., Westh, H., Penadés, J.R., Ingmer, H., 2016. Bacterial Viruses Enable Their Host To Acquire Antibiotic Resistance Genes from Neighbouring Cells. *Nature Communications* 7. <https://doi.org/10.1038/ncomms13333>
- Harkins, C.P., Pichon, B., Doumith, M., Parkhill, J., Westh, H., Tomasz, A., de Lencastre, H., Bentley, S.D., Kearns, A.M., Holden, M.T.G., 2017. Methicillin-Resistant *Staphylococcus aureus* Emerged Long Before the Introduction of Methicillin Into Clinical Practice. *Genome Biology* 18. <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., Gardete, S., Tavares, A., Day, N., Lindsay, J.A., Edgeworth, J.D., de Lencastre, H., Parkhill, J., Peacock, S.J., Bentley, S.D., 2010. Evolution of MRSA During Hospital Transmission and Intercontinental Spread. *Science* 327, 469–474. <https://doi.org/10.1126/science.1182395>
- Harrison, E., Brockhurst, M.A., 2012. Plasmid-Mediated Horizontal Gene Transfer is a Coevolutionary Process. *Trends in Microbiology* 20, 262–267. <https://doi.org/10.1016/j.tim.2012.04.003>
- Harrison, F., 2007. Microbial Ecology of the Cystic Fibrosis Lung. *Microbiology* 153, 917–923. <https://doi.org/10.1099/mic.0.2006/004077-0>
- He, M., Miyajima, F., Roberts, P., Ellison, L., Pickard, D.J., Martin, M.J., Connor, T.R., Harris, S.R., Fairley, D., Bamford, K.B., D’Arc, S., Brazier, J., Brown, D., Coia, J.E., Douce, G., Gerding, D., Kim, H.J., Koh, T.H., Kato, H., Senoh, M., Louie, T., Michell, S., Butt, E., Peacock, S.J., Brown, N.M., Riley, T., Songer, G., Wilcox, M., Pirmohamed, M., Kuijper, E., Hawkey, P., Wren, B.W., Dougan, G.,

- Parkhill, J., Lawley, T.D., 2012. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nature Genetics* 45, 109–113. <https://doi.org/10.1038/ng.2478>
- Hendriksen, R.S., Bortolaia, V., Tate, H., Tyson, G.H., Aarestrup, F.M., McDermott, P.F., 2019. Using Genomics to Track Global Antimicrobial Resistance. *Frontiers in Public Health* 7, 242. <https://doi.org/10.3389/fpubh.2019.00242>
- Hershberg, R., Petrov, D.A., 2010. Evidence That Mutation Is Universally Biased towards AT in Bacteria. *PLoS Genetics* 6, e1001115. <https://doi.org/10.1371/journal.pgen.1001115>
- Hiramatsu, K., 1997. Methicillin-Resistant *Staphylococcus aureus* Clinical Strain with Reduced Vancomycin Susceptibility. *Journal of Antimicrobial Chemotherapy* 40, 135–136. <https://doi.org/10.1093/jac/40.1.135>
- Holden, M.T.G., Hsu, L.-Y., Kurt, K., Weinert, L.A., Mather, A.E., Harris, S.R., Strommenger, B., Layer, F., Witte, W., de Lencastre, H., Skov, R., Westh, H., Zemlickova, H., Coombs, G., Kearns, A.M., Hill, R.L.R., Edgeworth, J., Gould, I., Gant, V., Cooke, J., Edwards, G.F., McAdam, P.R., Templeton, K.E., McCann, A., Zhou, Z., Castillo-Ramirez, S., Feil, E.J., Hudson, L.O., Enright, M.C., Balloux, F., Aanensen, D.M., Spratt, B.G., Fitzgerald, J.R., Parkhill, J., Achtman, M., Bentley, S.D., Nubel, U., 2013. A Genomic Portrait of the Emergence, Evolution, and Global Spread of a Methicillin-Resistant *Staphylococcus aureus* Pandemic. *Genome Research* 23, 653–664. <https://doi.org/10.1101/gr.147710.112>
- Holden, M.T.G., Lindsay, J.A., Corton, C., Quail, M.A., Cockfield, J.D., Pathak, S., Batra, R., Parkhill, J., Bentley, S.D., 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, 888–892. <https://doi.org/10.1128/jb.01255-09>
- Holmes, A.H., Moore, L.S.P., Sundsfjord, A., Steinbakk, M., Regmi, S., Karkey, A., Guerin, P.J., Piddock, L.J.V., 2016. Understanding the Mechanisms and Drivers of Antimicrobial Resistance. *The Lancet* 387, 176–187. [https://doi.org/10.1016/s0140-6736\(15\)00473-0](https://doi.org/10.1016/s0140-6736(15)00473-0)
- Homer N., 2011, DWGSIM, accessed June 2018 <<https://github.com/nh13/DWGSIM/wiki>>
- Hon, P.Y., Chan, K.S., Holden, M.T., Harris, S.R., Tan, T.Y., Zu, Y.-B., Krishnan, P., Oon, L.L., Koh, T.H., Hsu, L.Y., 2013. Arginine Catabolic Mobile Element in Methicillin-Resistant *Staphylococcus aureus* (MRSA) Clonal Group ST239-MRSA-III Isolates in Singapore: Implications for PCR-Based Screening Tests. *Antimicrobial Agents and Chemotherapy* 57, 1563–1564. <https://doi.org/10.1128/aac.02518-12>
- Howden, B.P., Seemann, T., Harrison, P.F., McEvoy, C.R., Stanton, J.-A.L., Rand, C.J., Mason, C.W., Jensen, S.O., Firth, N., Davies, J.K., Johnson, P.D.R., Stinear, T.P., 2010. Complete Genome Sequence of *Staphylococcus aureus* Strain JKD6008, an ST239 Clone of Methicillin-Resistant *Staphylococcus aureus* with Intermediate-Level Vancomycin Resistance. *Journal of Bacteriology* 192, 5848–5849. <https://doi.org/10.1128/jb.00951-10>
- Hsu, L.-Y., Harris, S.R., Chlebowicz, M.A., Lindsay, J.A., Koh, T.-H., Krishnan, P., Tan, T.-Y., Hon, P.-Y., Grubb, W.B., Bentley, S.D., Parkhill, J., Peacock, S.J., Holden, M.T., 2015. Evolutionary Dynamics of Methicillin-Resistant *Staphylococcus aureus* Within a Healthcare System. *Genome Biology* 16, 81. <https://doi.org/10.1186/s13059-015-0643-z>
- Hsu (a), L.-Y., Koh, T.-H., Singh, K., Kang, M.-L., Kurup, A., Tan, B.-H., 2005. Dissemination of Multi Susceptible Methicillin-Resistant *Staphylococcus aureus* in Singapore. *Journal of Clinical Microbiology* 43, 2923–2925. <https://doi.org/10.1128/jcm.43.6.2923-2925.2005>

References

- Hsu (b), L.-Y., Koh, T.-H., Tan, B.-H., 2005. Health Care-Associated Methicillin-Resistant *Staphylococcus aureus* is Evolving. *Clinical Infectious Diseases* 40, 1860–1861. <https://doi.org/10.1086/430448>
- Hsu, L.-Y., Koh, Y.-L., Chlebicka, N.L., Tan, T.-Y., Krishnan, P., Lin, R.T.-P., Tee, N., Barkham, T., Koh, T.-H., 2006. Establishment of ST30 As the Predominant Clonal Type among Community-Associated Methicillin-Resistant *Staphylococcus aureus* Isolates in Singapore. *Journal of Clinical Microbiology* 44, 1090–1093. <https://doi.org/10.1128/jcm.44.3.1090-1093.2006>
- Hsu (a), L.-Y., Loomba-Chlebicka, N., Koh, T.H., Kang, M.L., Tan, B.H., & Tambyah, P.A., 2007. EMRSA-15 Bacteremia is Not Associated with a Worse Outcome Compared with Bacteremia Caused by Multidrug-Resistant MRSA. *International journal of biomedical science* 3, 97–103.
- Hsu (b), L.-Y., Loomba-Chlebicka, N., Koh, Y.-L., Tan, T.-Y., Krishnan, P., Lin, R.T.-P., Tee, N.W.-S., Fisher, D.A., Koh, T.-H., 2007. Evolving EMRSA-15 Epidemic in Singapore Hospitals. *Journal of Medical Microbiology* 56, 376–379. <https://doi.org/10.1099/jmm.0.46950-0>
- Hu, F., Zhu, D., Wang, F., Wang, M., 2018. Current Status and Trends of Antibacterial Resistance in China. *Clinical Infectious Diseases* 67, S128–S134. <https://doi.org/10.1093/cid/ciy657>
- Hu, J., Zhang, X., Liu, X., Chen, C., Sun, B., 2014. Mechanism of Reduced Vancomycin Susceptibility Conferred by WalK Mutation in Community-Acquired Methicillin-Resistant *Staphylococcus aureus* Strain MW2. *Antimicrobial Agents and Chemotherapy* 59, 1352–1355. <https://doi.org/10.1128/aac.04290-14>
- Hua, J., Huet, A., Lopez, C.A., Toropova, K., Pope, W.H., Duda, R.L., Hendrix, R.W., Conway, J.F., 2017. Capsids and Genomes of Jumbo-Sized Bacteriophages Reveal the Evolutionary Reach of the HK97 Fold. *mBio* 8(5), e01579-17. <https://doi.org/10.1128/mbio.01579-17>
- Humphreys, H., Carroll, J.D., Keane, C.T., Cafferkey, M.T., Pomeroy, H.M., Coleman, D.C., 1990. Importation of Methicillin-Resistant *Staphylococcus aureus* from Baghdad To Dublin and Subsequent Nosocomial Spread. *Journal of Hospital Infection* 15, 127–135. [https://doi.org/10.1016/0195-6701\(90\)90121-4](https://doi.org/10.1016/0195-6701(90)90121-4)
- Hunt, M., Mather, A.E., Sánchez-Busó, L., Page, A.J., Parkhill, J., Keane, J.A., Harris, S.R., 2017. ARIBA: Rapid Antimicrobial Resistance Genotyping Directly from Sequencing Reads. *Microbial Genomics* 3(10), e000131. <https://doi.org/10.1099/mgen.0.000131>
- Huson, D.H., Scornavacca, C., 2012. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Systematic Biology* 61, 1061–1067. <https://doi.org/10.1093/sysbio/sys062>
- IACG, 2019, No Time To Wait: Securing the Future from Drug-Resistant Infections, Interagency Coordination Group on Antimicrobial Resistance, accessed Jan 2020
<https://www.who.int/antimicrobial-resistance/interagency-coordination-group/IACG_final_report_EN.pdf?ua=1>
- Illumina Inc., 2009, Calling Sequence SNPs, Illumina Inc., accessed Aug 2019
<https://www.illumina.com/Documents/products/technotes/technote_snp_caller_sequencing.pdf>
- International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC), 2009. Classification of Staphylococcal Cassette Chromosome mec (SCCmec): Guidelines for Reporting Novel SCCmec Elements. *Antimicrobial Agents and Chemotherapy* 53, 4961–4967. <https://doi.org/10.1128/aac.00579-09>
- Ippolito, G., Leone, S., Lauria, F.N., Nicastrì, E., Wenzel, R.P., 2010. Methicillin-Resistant *Staphylococcus aureus*: the Superbug. *International Journal of Infectious Diseases* 14, S7–S11. <https://doi.org/10.1016/j.ijid.2010.05.003>

- Isobe, H., Takano, T., Nishiyama, A., Hung, W.-C., Kuniyuki, S., Shibuya, Y., Reva, I., Yabe, S., Iwao, Y., Higuchi, W., Khokhlova, O.E., Okubo, T., Yamamoto, T., 2012. Evolution and Virulence of Pantone-Valentine Leukocidin-Positive ST30 Methicillin-Resistant *Staphylococcus aureus* in the Past 30 Years in Japan. *Biomedical Research* 33, 97–109. <https://doi.org/10.2220/biomedres.33.97>
- Ito, T., Katayama, Y., Asada, K., Mori, N., Tsutsumimoto, K., Tiensasitorn, C., Hiramatsu, K., 2001. Structural Comparison of Three Types of Staphylococcal Cassette Chromosome *mec* Integrated in the Chromosome in Methicillin-Resistant *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy* 45, 1323–1336. <https://doi.org/10.1128/aac.45.5.1323-1336.2001>
- Ito, T., Ma, X.X., Takeuchi, F., Okuma, K., Yuzawa, H., Hiramatsu, K., 2004. Novel Type V Staphylococcal Cassette Chromosome *mec* Driven by a Novel Cassette Chromosome Recombinase, *ccrC*. *Antimicrobial Agents and Chemotherapy* 48, 2637–2651. <https://doi.org/10.1128/aac.48.7.2637-2651.2004>
- Iwao, Y., Khokhlova, O.E., Takano, T., Hung, W.C., Isobe, H., Peryanova, O.V., Salmina, A.B., Yamamoto, T., 2012. Fatal Pneumonia in HIV-infected Patients from a Novel ST239 Methicillin-Resistant *Staphylococcus aureus* Carrying the Toxic Shock Syndrome Toxin-1 Gene in Krasnoyarsk, Siberian Russia. *Japanese Journal of Infectious Disease* 65, 184–186.
- Jaillard, M., Lima, L., Tournoud, M., Mahé, P., van Belkum, A., Lacroix, V., Jacob, L., 2018. A Fast and Agnostic Method For Bacterial Genome-Wide Association Studies: Bridging the Gap Between K-Mers and Genetic Events. *PLOS Genetics* 14, e1007758. <https://doi.org/10.1371/journal.pgen.1007758>
- Jamil, B., Gawlik, D., Syed, M.A., Shah, A.A., Abbasi, S.A., Müller, E., Reißig, A., Ehricht, R., Monecke, S., 2017. Hospital-Acquired Methicillin-Resistant *Staphylococcus aureus* (MRSA) from Pakistan: Molecular Characterisation by Microarray Technology. *European Journal of Clinical Microbiology & Infectious Diseases* 37, 691–700. <https://doi.org/10.1007/s10096-017-3161-y>
- Jansen van Rensburg, M.J., Eliya Madikane, V., Whitelaw, A., Chachage, M., Haffeejee, S., Gay Elisha, B., 2011. The Dominant Methicillin Resistant *Staphylococcus aureus* Clone from Hospitals in Cape Town Has An Unusual Genotype: ST612. *Clinical Microbiology and Infection* 17, 785–792. <https://doi.org/10.1111/j.1469-0691.2010.03373.x>
- Jansen, A., Türck, M., Szekat, C., Nagel, M., Clever, I., Bierbaum, G., 2007. Role of Insertion Elements and *Yycfg* in the Development of Decreased Susceptibility To Vancomycin in *Staphylococcus aureus*. *International Journal of Medical Microbiology* 297, 205–215. <https://doi.org/10.1016/j.ijmm.2007.02.002>
- Jessen, O., Rosendal, K., Bülow, P., Faber, V., Eriksen, K.R., 1969. Changing Staphylococci and Staphylococcal Infections. *New England Journal of Medicine* 281, 627–635. <https://doi.org/10.1056/nejm196909182811201>
- Jevons, M.P., 1961. “Celbenin” - resistant Staphylococci. *British Medical Journal*, 1, 124–125.
- Ji, Q., Chen, P.J., Qin, G., Deng, X., Hao, Z., Wawrzak, Z., Yeo, W.-S., Quang, J.W., Cho, H., Luo, G.-Z., Weng, X., You, Q., Luan, C.-H., Yang, X., Bae, T., Yu, K., Jiang, H., He, C., 2016. Structure and Mechanism of the Essential Two-Component Signal-Transduction System WalKR in *Staphylococcus aureus*. *Nature Communications* 7. <https://doi.org/10.1038/ncomms11000>
- Jia, B., Raphenya, A.R., Alcock, B., Waglechner, N., Guo, P., Tsang, K.K., Lago, B.A., Dave, B.M., Pereira, S., Sharma, A.N., Doshi, S., Courtot, M., Lo, R., Williams, L.E., Frye, J.G., Elsayegh, T., Sardar, D., Westman, E.L., Pawlowski, A.C., Johnson, T.A., Brinkman, F.S.L., Wright, G.D., McArthur, A.G., 2016. CARD 2017: Expansion and Model-Centric Curation of the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research* 45, D566–D573. <https://doi.org/10.1093/nar/gkw1004>

References

- Joensen, K.G., Scheutz, F., Lund, O., Hasman, H., Kaas, R.S., Nielsen, E.M., Aarestrup, F.M., 2014. Real-Time Whole-Genome Sequencing for Routine Typing, Surveillance, and Outbreak Detection of Verotoxigenic *Escherichia coli*. *Journal of Clinical Microbiology* 52, 1501–1510. <https://doi.org/10.1128/jcm.03617-13>
- John, J., George, S., Nori, S.R.C., Nelson-Sathi, S., 2019. Phylogenomic Analysis Reveals the Evolutionary Route of Resistant Genes in *Staphylococcus aureus*. *Genome Biology and Evolution* 11, 2917–2926. <https://doi.org/10.1093/gbe/evz213>
- Jolley, K.A., Bray, J.E., Maiden, M.C.J., 2018. Open-Access Bacterial Population Genomics: BIGSdb Software, the PubMLST.org website and Their Applications. *Wellcome Open Research* 3, 124. <https://doi.org/10.12688/wellcomeopenres.14826.1>
- Jolley, K.A., Maiden, M.C., 2010. BIGSdb: Scalable Analysis of Bacterial Genome Variation At the Population Level. *BMC Bioinformatics* 11. <https://doi.org/10.1186/1471-2105-11-595>
- Kaatz, G.W., McAleese, F., Seo, S.M., 2005. Multidrug Resistance in *Staphylococcus aureus* Due to Overexpression of a Novel Multidrug and Toxin Extrusion (MATE) Transport Protein. *Antimicrobial Agents and Chemotherapy* 49, 1857–1864. <https://doi.org/10.1128/aac.49.5.1857-1864.2005>
- Kallen, A.J., Driscoll, T.J., Thornton, S., Olson, P.E., Wallace, M.R., 2000. Increase in Community-Acquired Methicillin-Resistant *Staphylococcus aureus* at a Naval Medical Center. *Infection Control & Hospital Epidemiology* 21, 223–226. <https://doi.org/10.1086/501750>
- Katayama, Y., Ito, T., Hiramatsu, K., 2000. A New Class of Genetic Element, *Staphylococcus* Cassette Chromosome *mec*, Encodes Methicillin Resistance in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy* 44, 1549–1555. <https://doi.org/10.1128/aac.44.6.1549-1555.2000>
- Kavanagh, K.T., 2019. Control of MSSA and MRSA in the United States: Protocols, Policies, Risk Adjustment and Excuses. *Antimicrobial Resistance & Infection Control* 8, 103. <https://doi.org/10.1186/s13756-019-0550-2>
- Kaya, H., Hasman, H., Larsen, J., Stegger, M., Johannesen, T.B., Allesøe, R.L., Lemvig, C.K., Aarestrup, F.M., Lund, O., 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. <https://doi.org/10.1128/msphere.00612-17>
- Kerr, S., Kerr, G.E., Mackintosh, C.A., Marples, R.R., 1990. A Survey of Methicillin-Resistant *Staphylococcus aureus* Affecting Patients in England and Wales. *Journal of Hospital Infection* 16, 35–48. [https://doi.org/10.1016/0195-6701\(90\)90047-r](https://doi.org/10.1016/0195-6701(90)90047-r)
- Keymer, J.E., Galajda, P., Lambert, G., Liao, D., Austin, R.H., 2008. Computation of Mutual Fitness by Competing Bacteria. *Proceedings of the National Academy of Sciences* 105, 20269–20273. <https://doi.org/10.1073/pnas.0810792105>
- Khosravi, A.D., Jenabi, A., Montazeri, E.A., 2017. Distribution of Genes Encoding Resistance To Aminoglycoside Modifying Enzymes in Methicillin-Resistant *Staphylococcus aureus* (MRSA) Strains. *The Kaohsiung Journal of Medical Sciences* 33, 587–593. <https://doi.org/10.1016/j.kjms.2017.08.001>
- Kirby, W.M.M., 1944. Extraction of a Highly Potent Penicillin Inactivator from Penicillin Resistant *Staphylococci*. *Science* 99, 452–453. <https://doi.org/10.1126/science.99.2579.452>
- Knight, G.M., Budd, E.L., Lindsay, J.A., 2013. Large mobile genetic elements carrying resistance genes that do not confer a fitness burden in healthcare-associated methicillin-resistant *Staphylococcus aureus*. *Microbiology* 159, 1661–1672. <https://doi.org/10.1099/mic.0.068551-0>
- Knight, G.M., Budd, E.L., Whitney, L., Thornley, A., Al-Ghusein, H., Planche, T., Lindsay, J.A., 2012. Shift in Dominant Hospital-Associated Methicillin-Resistant *Staphylococcus aureus* (HA-MRSA)

- Clones Over Time. *Journal of Antimicrobial Chemotherapy* 67, 2514–2522. <https://doi.org/10.1093/jac/dks245>
- Knox, R., 1960. A New Penicillin (BRL 1241) Active Against Penicillin-resistant Staphylococci. *BMJ* 2, 690–693. <https://doi.org/10.1136/bmj.2.5200.690>
- Ko, D.-S., Seong, W.-J., Kim, D., Kim, E.-K., Kim, N.-H., Lee, C.-Y., Kim, J.-H., Kwon, H.-J., 2018. Molecular Prophage Typing of Staphylococcus aureus Isolates from Bovine Mastitis. *Journal of Veterinary Science* 19, 771. <https://doi.org/10.4142/jvs.2018.19.6.771>
- Ko, K.S., Lee, J.-Y., Suh, J.Y., Oh, W.S., Peck, K.R., Lee, N.Y., Song, J.-H., 2005. Distribution of Major Genotypes among Methicillin-Resistant Staphylococcus aureus Clones in Asian Countries. *Journal of Clinical Microbiology* 43, 421–426. <https://doi.org/10.1128/jcm.43.1.421-426.2005>
- Kong, E.F., Johnson, J.K., Jabra-Rizk, M.A., 2016. Community-Associated Methicillin-Resistant Staphylococcus aureus: An Enemy amidst Us. *PLOS Pathogens* 12, e1005837. <https://doi.org/10.1371/journal.ppat.1005837>
- Kong, H., Fang, L., Jiang, R., Tong, J., 2018. Distribution of SasX, Pvl, and QacA/B Genes in Epidemic Methicillin-Resistant Staphylococcus aureus Strains Isolated from East China. *Infection and Drug Resistance* 11, 55–59. <https://doi.org/10.2147/idr.s153399>
- Korneliussen, T.S., Moltke, I., Albrechtsen, A., Nielsen, R., 2013. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics* 14. <https://doi.org/10.1186/1471-2105-14-289>
- Kotrba, P., Inui, M., Yukawa, H., 2001. Bacterial Phosphotransferase System (PTS) in Carbohydrate Uptake and Control of Carbon Metabolism. *Journal of Bioscience and Bioengineering* 92, 502–517. [https://doi.org/10.1016/s1389-1723\(01\)80308-x](https://doi.org/10.1016/s1389-1723(01)80308-x)
- Kourtis, A.P., Hatfield, K., Baggs, J., Mu, Y., See, I., Epton, E., Nadle, J., Kainer, M.A., Dumyati, G., Petit, S., Ray, S.M., Ham, D., Capers, C., Ewing, H., Coffin, N., McDonald, L.C., Jernigan, J., Cardo, D., 2019. Vital Signs: Epidemiology and Recent Trends in Methicillin-Resistant and in Methicillin-Susceptible Staphylococcus aureus Bloodstream Infections — United States. *MMWR. Morbidity and Mortality Weekly Report* 68, 214–219. <https://doi.org/10.15585/mmwr.mm6809e1>
- Kubicek-Sutherland, J.Z., Lofton, H., Vestergaard, M., Hjort, K., Ingmer, H., Andersson, D.I., 2016. Antimicrobial Peptide Exposure Selects For Staphylococcus aureus Resistance To Human Defence Peptides. *Journal of Antimicrobial Chemotherapy* 72, 115–127. <https://doi.org/10.1093/jac/dkw381>
- Kucukyildirim, S., Long, H., Sung, W., Miller, S.F., Doak, T.G., Lynch, M., 2016. The Rate and Spectrum of Spontaneous Mutations in Mycobacterium smegmatis, a Bacterium Naturally Devoid of the Postreplicative Mismatch Repair Pathway. *Genes|Genomes|Genetics* 6, 2157–2163. <https://doi.org/10.1534/g3.116.030130>
- Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., Takahashi, N.K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H., Hiramatsu, K., 2001. Whole Genome Sequencing of Methicillin-Resistant Staphylococcus aureus. *The Lancet* 357, 1225–1240. [https://doi.org/10.1016/s0140-6736\(00\)04403-2](https://doi.org/10.1016/s0140-6736(00)04403-2)
- Lakhundi, S., Zhang, K., 2018. Methicillin-Resistant Staphylococcus aureus: Molecular Characterization, Evolution, and Epidemiology. *Clinical Microbiology Reviews* 31(4), e00020-18. <https://doi.org/10.1128/cmr.00020-18>

References

- Lauener, F.N., Imkamp, F., Lehours, P., Buissonnière, A., Benejat, L., Zbinden, R., Keller, P.M., Wagner, K., 2019. Genetic Determinants and Prediction of Antibiotic Resistance Phenotypes in *Helicobacter pylori*. *Journal of Clinical Medicine* 8, 53. <https://doi.org/10.3390/jcm8010053>
- Lebeuf-Taylor, E., McCloskey, N., Bailey, S.F., Hinz, A., Kassen, R., 2019. The Distribution of Fitness Effects among Synonymous Mutations in a Gene Under Directional Selection. *eLife* 8. <https://doi.org/10.7554/elife.45952>
- Leclercq, R., 2002. Mechanisms of Resistance to Macrolides and Lincosamides: Nature of the Resistance Elements and Their Clinical Implications. *Clinical Infectious Diseases* 34, 482–492. <https://doi.org/10.1086/324626>
- Lee, L., Teh, L., Zainuddin, Z., Salleh, M., 2014. The Genome Sequence of a Type ST239 Methicillin-Resistant *Staphylococcus aureus* Isolate from a Malaysian Hospital. *Standards in Genomic Sciences* 9, 933–939. <https://doi.org/10.4056/sigs.3887716>
- Lee, S.M., Ender, M., Adhikari, R., Smith, J.M.B., Berger-Bachi, B., Cook, G.M., 2007. Fitness Cost of Staphylococcal Cassette Chromosome *mec* in Methicillin-Resistant *Staphylococcus aureus* by Way of Continuous Culture. *Antimicrobial Agents and Chemotherapy* 51, 1497–1499. <https://doi.org/10.1128/aac.01239-06>
- Lencastre, H. de, Severina, E.P., Milch, H., Thege, M.K., Tomasz, A., 1997. Wide Geographic Distribution of a Unique Methicillin-Resistant *Staphylococcus aureus* Clone in Hungarian Hospitals. *Clinical Microbiology and Infection* 3, 289–296. <https://doi.org/10.1111/j.1469-0691.1997.tb00616.x>
- Leski, T.A., Tomasz, A., 2005. Role of Penicillin-Binding Protein 2 (PBP2) in the Antibiotic Susceptibility and Cell Wall Cross-Linking of *Staphylococcus aureus*: Evidence for the Cooperative Functioning of PBP2, PBP4, and PBP2A. *Journal of Bacteriology* 187, 1815–1824. <https://doi.org/10.1128/jb.187.5.1815-1824.2005>
- Li, H., Durbin, R., 2009. Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, M., Du, X., Villaruz, A.E., Diep, B.A., Wang, D., Song, Y., Tian, Y., Hu, J., Yu, F., Lu, Y., Otto, M., 2012. MRSA Epidemic Linked To a Quickly Spreading Colonization and Virulence Determinant. *Nature Medicine* 18, 816–819. <https://doi.org/10.1038/nm.2692>
- Li, S., Sun, S., Yang, C., Chen, H., Yin, Y., Li, H., Zhao, C., Wang, H., 2018. The Changing Pattern of Population Structure of *Staphylococcus aureus* from Bacteremia in China from 2013 to 2016: ST239-030-MRSA Replaced by ST59-t437. *Frontiers in Microbiology* 9, 332. <https://doi.org/10.3389/fmicb.2018.00332>
- Li, Y., Cao, B., Zhang, Y., Zhou, J., Yang, B., Wang, L., 2011. Complete Genome Sequence of *Staphylococcus aureus* T0131, an ST239-MRSA-SCCmec Type III Clone Isolated in China. *Journal of Bacteriology* 193, 3411–3412. <https://doi.org/10.1128/jb.05135-11>
- Licitra, G., 2013. Etymologia: *Staphylococcus*. *Emerging Infectious Diseases* 19. <https://doi.org/10.3201/eid1909.et1909>
- Lindsay, J.A., 2014. *Staphylococcus aureus* Genomics and the Impact of Horizontal Gene Transfer. *International Journal of Medical Microbiology* 304, 103–109. <https://doi.org/10.1016/j.ijmm.2013.11.010>
- Liu, J., Chen, D., Peters, B.M., Li, L., Li, B., Xu, Z., Shirliff, M.E., 2016. Staphylococcal Chromosomal Cassettes *mec* (SCCmec): a Mobile Genetic Element in Methicillin-Resistant *Staphylococcus aureus*. *Microbial Pathogenesis* 101, 56–67. <https://doi.org/10.1016/j.micpath.2016.10.028>

- Liu, Y., Kong, F., Xiao, M., Wang, Q., O'Sullivan, M., Sintchenko, V., Ma, L., Gilbert, G.L., 2010. Unexpected Diversity of Staphylococcal Cassette Chromosome *mec* Type IV in Methicillin-Resistant *Staphylococcus aureus* Strains. *Journal of Clinical Microbiology* 48, 3628–3634. <https://doi.org/10.1128/jcm.00351-10>
- Luedicke, C., Slickers, P., Ehricht, R., Monecke, S., 2010. Molecular Fingerprinting of *Staphylococcus aureus* from Bone and Joint Infections. *European Journal of Clinical Microbiology & Infectious Diseases* 29, 457–463. <https://doi.org/10.1007/s10096-010-0884-4>
- MacLean, R.C., San Millan, A., 2019. The Evolution of Antibiotic Resistance. *Science* 365, 1082–1083. <https://doi.org/10.1126/science.aax3879>
- Mäder, U., Nicolas, P., Depke, M., Pané-Farré, J., Debarbouille, M., van der Kooi-Pol, M.M., Guérin, C., Dérozier, S., Hiron, A., Jarmer, H., Leduc, A., Michalik, S., Reilman, E., Schaffer, M., Schmidt, F., Bessières, P., Noirot, P., Hecker, M., Msadek, T., Völker, U., van Dijk, J.M., 2016. *Staphylococcus aureus* Transcriptome Architecture: from Laboratory to Infection-Mimicking Conditions. *PLOS Genetics* 12, e1005962. <https://doi.org/10.1371/journal.pgen.1005962>
- Madeira, F., Park, Y. mi, Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., Lopez, R., 2019. The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019. *Nucleic Acids Research* 47, W636–W641. <https://doi.org/10.1093/nar/gkz268>
- MapChart, 2019, MapChart, accessed August 2019 <<https://mapchart.net/world.html>>
- Mariem, B.J.-J., Ito, T., Zhang, M., Jin, J., Li, S., Ilhem, B.-B.B., Adnan, H., Han, X., Hiramatsu, K., 2013. Molecular Characterization of Methicillin-Resistant Pantone-Valentine Leukocidin Positive *Staphylococcus aureus* Clones Disseminating in Tunisian Hospitals and in the Community. *BMC Microbiology* 13, 2. <https://doi.org/10.1186/1471-2180-13-2>
- Mayor, L., Ortellado, J., Menacho, C., Lird, G., Courtier, C., Gardon, C., Meugnier, H., Bes, M., Vandenesch, F., Etienne, J., 2007. Molecular Characterization of Methicillin-Resistant *Staphylococcus aureus* Isolates Collected in Asuncion, Paraguay. *Journal of Clinical Microbiology* 45, 2298–2300. <https://doi.org/10.1128/jcm.00040-07>
- McAdam, P.R., Templeton, K.E., Edwards, G.F., Holden, M.T.G., Feil, E.J., Aanensen, D.M., Bargawi, H.J.A., Spratt, B.G., Bentley, S.D., Parkhill, J., Enright, M.C., Holmes, A., Girvan, E.K., Godfrey, P.A., Feldgarden, M., Kearns, A.M., Rambaut, A., Robinson, D.A., Fitzgerald, J.R., 2012. Molecular Tracing of the Emergence, Adaptation, and Transmission of Hospital-Associated Methicillin-Resistant *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences* 109, 9107–9112. <https://doi.org/10.1073/pnas.1202869109>
- McDonald, J.H., Kreitman, M., 1991. Adaptive Protein Evolution At the *Adh* Locus in *Drosophila*. *Nature* 351, 652–654. <https://doi.org/10.1038/351652a0>
- McGuinness, W.A., Malachowa, N., DeLeo, F.R., 2017. Vancomycin Resistance in *Staphylococcus aureus*. *The Yale journal of biology and medicine* 90, 269–281.
- 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, 588–595. <https://doi.org/10.1016/j.mib.2012.08.003>
- Melter, O., Santos Sanches, I., Schindler, J., Aires de Sousa, M., Mato, R., Kovárova, V., de Lencastre, H. 1999. Methicillin-Resistant *Staphylococcus aureus* Clonal Types in the Czech Republic. *Journal of clinical microbiology* 37, 2798–2803.
- MERLIN Diagnostika GmbH, 2018, MICRONAUT-S, Bruker, accessed Aug 2019 <https://www.merlin-diagnostika.de/fileadmin/mediapool/downloads/Produkte/Micronaut_Spezialplatten/1858846_MICRONAUT_S_brochure_03_2018_ebook.pdf>

References

- Mitchell, J.M., MacCulloch, D., Morris, A.J., 1996. MRSA in the community. *New Zealand Medical Journal* 109, 411.
- Molina, N., van Nimwegen, E., 2007. Universal patterns of purifying selection at noncoding positions in bacteria. *Genome Research* 18, 148–160. <https://doi.org/10.1101/gr.6759507>
- Monecke, S., Müller, E., Dorneanu, O.S., Vremeră, T., Ehricht, R., 2014. Molecular Typing of MRSA and of Clinical *Staphylococcus aureus* Isolates from Iași, Romania. *PLoS ONE* 9, e97833. <https://doi.org/10.1371/journal.pone.0097833>
- Monecke, S., Nitschke, H., Slickers, P., Ehricht, R., Swanston, W., Manjunath, M., Roberts, R., Akpaka, P.E., 2011. Molecular Epidemiology and Characterisation of MRSA Isolates from Trinidad and Tobago. *European Journal of Clinical Microbiology & Infectious Diseases* 31, 1497–1500. <https://doi.org/10.1007/s10096-011-1469-6>
- Monecke, S., Skakni, L., Hasan, R., Ruppelt, A., Ghazal, S.S., Hakawi, A., Slickers, P., Ehricht, R., 2012. Characterisation of MRSA Strains Isolated from Patients in a Hospital in Riyadh, Kingdom of Saudi Arabia. *BMC Microbiology* 12, 146. <https://doi.org/10.1186/1471-2180-12-146>
- Monecke, S., Slickers, P., Gawlik, D., Müller, E., Reissig, A., Ruppelt-Lorz, A., Akpaka, P.E., Bandt, D., Bes, M., Boswihi, S.S., Coleman, D.C., Coombs, G.W., Dorneanu, O.S., Gostev, V.V., Ip, M., Jamil, B., Jatzwauk, L., Narvaez, M., Roberts, R., Senok, A., Shore, A.C., Sidorenko, S.V., Skakni, L., Somily, A.M., Syed, M.A., Thürmer, A., Udo, E.E., Vremeră, T., Zurita, J., Ehricht, R., 2018. Molecular Typing of ST239-MRSA-III from Diverse Geographic Locations and the Evolution of the SCCmec III Element During its Intercontinental Spread. *Frontiers in Microbiology* 9. <https://doi.org/10.3389/fmicb.2018.01436>
- Monecke, S., Stieber, B., Roberts, R., Akpaka, P.E., Slickers, P., Ehricht, R., 2014. Population Structure of *Staphylococcus aureus* from Trinidad & Tobago. *PLoS ONE* 9, e89120. <https://doi.org/10.1371/journal.pone.0089120>
- Moodley, A., Oosthuysen, W.F., Dusé, A.G., Marais, E., 2010. Molecular Characterization of Clinical Methicillin-Resistant *Staphylococcus aureus* Isolates in South Africa. *Journal of Clinical Microbiology* 48, 4608–4611. <https://doi.org/10.1128/jcm.01704-10>
- Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T.L., Agarwala, R., Schäffer, A.A., 2008. Database Indexing For Production MegaBLAST Searches. *Bioinformatics* 24, 1757–1764. <https://doi.org/10.1093/bioinformatics/btn322>
- 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, e1003003. <https://doi.org/10.1371/journal.ppat.1003003>
- Morrissey, J.A., 2002. Conservation, Surface Exposure, and in Vivo Expression of the Frp Family of Iron-Regulated Cell Wall Proteins in *Staphylococcus aureus*. *Infection and Immunity* 70, 2399–2407. <https://doi.org/10.1128/iai.70.5.2399-2407.2002>
- 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, e1004300. <https://doi.org/10.1371/journal.pgen.1004300>
- Murugan, S., Neetu, T.P., 2016. Genotyping of Methicillin Resistant *Staphylococcus aureus* from Tertiary Care Hospitals in Coimbatore, South India. *Journal of Global Infectious Diseases* 8, 68. <https://doi.org/10.4103/0974-777x.182119>
- Muto, A., Osawa, S., 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proceedings of the National Academy of Sciences* 84, 166–169. <https://doi.org/10.1073/pnas.84.1.166>

- Nadkarni, M.A., Hunter, N., Jacques, N.A., Martin, F.E., 2002. Determination of Bacterial Load by Real-Time PCR Using a Broad-Range (Universal) Probe and Primers Set. *Microbiology* 148, 257–266. <https://doi.org/10.1099/00221287-148-1-257>
- Naimi, T.S., 2003. Comparison of Community- and Health Care–Associated Methicillin-Resistant *Staphylococcus aureus* Infection. *JAMA* 290, 2976. <https://doi.org/10.1001/jama.290.22.2976>
- National Human Genome Research Institute, 2019, DNA Sequencing Costs: Data, National Institute of Health, accessed Aug 2019 <<https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>>
- Neela, V., Ghasemzadeh Moghaddam, H., van Belkum, A., Horst-Kreft, D., Mariana, N.S., Ghaznavi Rad, E., 2009. First Report on Methicillin-Resistant *Staphylococcus aureus* of Spa type T037, Sequence type 239, SCCmec type III/IIIA in Malaysia. *European Journal of Clinical Microbiology & Infectious Diseases* 29. <https://doi.org/10.1007/s10096-009-0813-6>
- Nielsen, K.L., Pedersen, T.M., Udekwi, K.I., Petersen, A., Skov, R.L., Hansen, L.H., Hughes, D., Frimodt-Møller, N., 2012. Fitness Cost: a Bacteriological Explanation For the Demise of the First International Methicillin-Resistant *Staphylococcus aureus* Epidemic. *Journal of Antimicrobial Chemotherapy* 67, 1325–1332. <https://doi.org/10.1093/jac/dks051>
- Nikolaras, G.P., Papaparaskevas, J., Samarkos, M., Tzouvelekis, L.S., Psychogiou, M., Pavlopoulou, I., Goukos, D., Polonyfi, K., Pantazatou, A., Deliolanis, I., Smilakou, S., Daikos, G.L., 2019. Changes in the Rates and Population Structure of Methicillin-Resistant *Staphylococcus aureus* (MRSA) from Bloodstream Infections: a Single-Centre Experience (2000–2015). *Journal of Global Antimicrobial Resistance* 17, 117–122. <https://doi.org/10.1016/j.jgar.2018.11.023>
- Nimmo, G.R., Steen, J.A., Monecke, S., Ehricht, R., Slickers, P., Thomas, J.C., Appleton, S., Goering, R.V., Robinson, D.A., 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, 444–450. <https://doi.org/10.1016/j.cmi.2014.12.018>
- Nübel, U., Dordel, J., Kurt, K., Strommenger, B., Westh, H., Shukla, S.K., Žemličková, H., Leblois, R., Wirth, T., Jombart, T., Balloux, F., Witte, W., 2010. A Timescale for Evolution, Population Expansion, and Spatial Spread of an Emerging Clone of Methicillin-Resistant *Staphylococcus aureus*. *PLoS Pathogens* 6, e1000855. <https://doi.org/10.1371/journal.ppat.1000855>
- Okonogi, K., Noji, Y., Kondo, M., Imada, A., Yokota, T., 1989. Emergence of Methicillin-Resistant Clones from Cephamycin-Resistant *Staphylococcus aureus*. *Journal of Antimicrobial Chemotherapy* 24, 637–645. <https://doi.org/10.1093/jac/24.5.637>
- Okuma, K., Iwakawa, K., Turnidge, J.D., Grubb, W.B., Bell, J.M., O'Brien, F.G., Coombs, G.W., Pearman, J.W., Tenover, F.C., Kapi, M., Tiensasitorn, C., Ito, T., Hiramatsu, K., 2002. Dissemination of New Methicillin-Resistant *Staphylococcus aureus* Clones in the Community. *Journal of Clinical Microbiology* 40, 4289–4294. <https://doi.org/10.1128/jcm.40.11.4289-4294.2002>
- Oliveira, D.C., Lencastre, H. d., 2002. Multiplex PCR Strategy for Rapid Identification of Structural Types and Variants of the *mec* Element in Methicillin-Resistant *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy* 46, 2155–2161. <https://doi.org/10.1128/aac.46.7.2155-2161.2002>
- Olsen, J.E., Christensen, H., Aarestrup, F.M., 2006. Diversity and Evolution of *bla_Z* from *Staphylococcus aureus* and Coagulase-Negative *Staphylococci*. *Journal of Antimicrobial Chemotherapy* 57, 450–460. <https://doi.org/10.1093/jac/dki492>
- O'Neill, A.J., Huovinen, T., Fishwick, C.W.G., Chopra, I., 2005. Molecular Genetic and Structural Modeling Studies of *Staphylococcus aureus* RNA Polymerase and the Fitness of Rifampin Resistance Genotypes in Relation to Clinical Prevalence. *Antimicrobial Agents and Chemotherapy* 50, 298–309. <https://doi.org/10.1128/aac.50.1.298-309.2006>

References

- Oosthuysen, W.F., Orth, H., Lombard, C.J., Sinha, B., Wasserman, E., 2014. Population structure analyses of *Staphylococcus aureus* at Tygerberg Hospital, South Africa, Reveals a Diverse Population, a High Prevalence of Pantón–Valentine Leukocidin Genes, and Unique Local Methicillin-Resistant *S. Aureus* Clones. *Clinical Microbiology and Infection* 20, 652–659. <https://doi.org/10.1111/1469-0691.12452>
- Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: Rapid Large-Scale Prokaryote Pan Genome Analysis. *Bioinformatics* 31, 3691–3693. <https://doi.org/10.1093/bioinformatics/btv421>
- Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A., Harris, S.R., 2016. SNP-Sites: Rapid Efficient Extraction of SNPs from Multi-FASTA Alignments. *Microbial Genomics* 2(4), e000056. <https://doi.org/10.1099/mgen.0.000056>
- Pal, S.C., Ray, B.G., 1964 Methicillin-Resistant *Staphylococci*. *Journal of the Indian Medical Association* 42, 512–517.
- Palmer, K.L., Mashburn, L.M., Singh, P.K., Whiteley, M., 2005. Cystic Fibrosis Sputum Supports Growth and Cues Key Aspects of *Pseudomonas aeruginosa* Physiology. *Journal of Bacteriology* 187, 5267–5277. <https://doi.org/10.1128/jb.187.15.5267-5277.2005>
- Pavillard, R., Harvey, K., Douglas, D., Hewstone, A., Andrew, J., Collopy, B., Asche, V., Carson, P., Davidson, A., Gilbert, G., Hospital, A., Spicer, J., Hospital, A., Tosolini, F., 1982. Epidemic of Hospital Acquired Infection Due To Methicillin Resistant *Staphylococcus aureus* in Major Victorian Hospitals. *Medical Journal of Australia* 1, 451–454. <https://doi.org/10.5694/j.1326-5377.1982.tb132413.x>
- Peck, K.R., Baek, J.Y., Song, J.-H., Ko, K.S., 2009. Comparison of Genotypes and Enterotoxin Genes Between *Staphylococcus aureus* Isolates from Blood and Nasal Colonizers in a Korean Hospital. *Journal of Korean Medical Science* 24, 585. <https://doi.org/10.3346/jkms.2009.24.4.585>
- Peng, H., Hu, Q., Shang, W., Yuan, J., Zhang, X., Liu, H., Zheng, Y., Hu, Z., Yang, Y., Tan, L., Li, S., Hu, X., Li, M., Rao, X., 2016. WalK(S221P), a Naturally Occurring Mutation, Confers Vancomycin Resistance in VISA strain XN108. *Journal of Antimicrobial Chemotherapy* 4, 1006–1013. <https://doi.org/10.1093/jac/dkw518>
- Pereira, M.P., D’Elia, M.A., Troczynska, J., Brown, E.D., 2008. Duplication of Teichoic Acid Biosynthetic Genes in *Staphylococcus aureus* Leads to Functionally Redundant Poly(Ribitol Phosphate) Polymerases. *Journal of Bacteriology* 190, 5642–5649. <https://doi.org/10.1128/jb.00526-08>
- Petit, R.A., III, Read, T.D., 2018. *Staphylococcus aureus* Viewed from the Perspective of 40,000+ Genomes. *PeerJ* 6, e5261. <https://doi.org/10.7717/peerj.5261>
- Planet, P.J., 2017. Life After USA300: the Rise and Fall of a Superbug. *The Journal of Infectious Diseases* 215, S71–S77. <https://doi.org/10.1093/infdis/jiw444>
- Potel, C., Álvarez, M., Álvarez, P., Otero, I., Fluiters, E., 2007. Evolution, Antimicrobial Susceptibility and Assignment To International Clones of Methicillin-Resistant *Staphylococcus aureus* Isolated Over a 9-Year Period in Two Spanish Hospitals. *Clinical Microbiology and Infection* 13, 728–730. <https://doi.org/10.1111/j.1469-0691.2007.01728.x>
- Price, J.R., Cole, K., Bexley, A., Kostiou, V., Eyre, D.W., Golubchik, T., Wilson, D.J., Crook, D.W., Walker, A.S., Peto, T.E.A., Llewelyn, M.J., Paul, J., 2017. Transmission of *Staphylococcus aureus* Between Health-Care Workers, the Environment, and Patients in An Intensive Care Unit: a Longitudinal Cohort Study Based on Whole-Genome Sequencing. *The Lancet Infectious Diseases* 17, 207–214. [https://doi.org/10.1016/s1473-3099\(16\)30413-3](https://doi.org/10.1016/s1473-3099(16)30413-3)
- QIAGEN, 2006, DNeasy® Blood & Tissue Handbook, QIAGEN, accessed Dec 2019 <<http://www.bea.ki.se/documents/EN-DNeasy20handbook.pdf>>

- Quainoo, S., Coolen, J.P.M., van Hijum, S.A.F.T., Huynen, M.A., Melchers, W.J.G., van Schaik, W., Wertheim, H.F.L., 2017. Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. *Clinical Microbiology Reviews* 30, 1015–1063. <https://doi.org/10.1128/cmr.00016-17>
- Raji, A., Ojemhen, O., Umejiburu, U., Ogunleye, A., Blanc, D.S., Basset, P., 2013. High Genetic Diversity of *Staphylococcus aureus* in a Tertiary Care Hospital in Southwest Nigeria. *Diagnostic Microbiology and Infectious Disease* 77, 367–369. <https://doi.org/10.1016/j.diagmicrobio.2013.08.030>
- Ramdani-Bouguessa, N., Bes, M., Meugnier, H., Forey, F., Reverdy, M.-E., Lina, G., Vandenesch, F., Tazir, M., Etienne, J., 2006. Detection of Methicillin-Resistant *Staphylococcus aureus* Strains Resistant to Multiple Antibiotics and Carrying the Panton-Valentine Leukocidin Genes in an Algiers Hospital. *Antimicrobial Agents and Chemotherapy* 50, 1083–1085. <https://doi.org/10.1128/aac.50.3.1083-1085.2006>
- Ramirez, M.S., Tolmasky, M.E., 2010. Aminoglycoside Modifying Enzymes. *Drug Resistance Updates* 13, 151–171. <https://doi.org/10.1016/j.drug.2010.08.003>
- Rammelkamp, C.H., Maxon, T., 1942. Resistance of *Staphylococcus aureus* To the Action of Penicillin. *Experimental Biology and Medicine* 51, 386–389. <https://doi.org/10.3181/00379727-51-13986>
- Rand, D.M., Kann, L.M., 1996. Excess Amino Acid Polymorphism in Mitochondrial DNA: Contrasts among Genes from *Drosophila*, Mice, and Humans. *Molecular Biology and Evolution* 13, 735–748. <https://doi.org/10.1093/oxfordjournals.molbev.a025634>
- Rankin, D.J., Rocha, E.P.C., Brown, S.P., 2010. What Traits Are Carried on Mobile Genetic Elements, and Why? *Heredity* 106, 1–10. <https://doi.org/10.1038/hdy.2010.24>
- Raoult, D., 2010. The Post-Darwinist Rhizome of Life. *The Lancet* 375, 104–105. [https://doi.org/10.1016/s0140-6736\(09\)61958-9](https://doi.org/10.1016/s0140-6736(09)61958-9)
- Reller, L.B., Weinstein, M., Jorgensen, J.H., Ferraro, M.J., 2009. Antimicrobial Susceptibility Testing: a Review of General Principles and Contemporary Practices. *Clinical Infectious Diseases* 49, 1749–1755. <https://doi.org/10.1086/647952>
- Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics* 16, 276–277. [https://doi.org/10.1016/s0168-9525\(00\)00204-2](https://doi.org/10.1016/s0168-9525(00)00204-2)
- Richardson, J.F., Marples, R.R., 1988. Another Strain of Methicillin-Resistant *Staphylococcus aureus* Epidemic in London. *The Lancet* 332, 748–749. [https://doi.org/10.1016/s0140-6736\(88\)90224-3](https://doi.org/10.1016/s0140-6736(88)90224-3)
- Richardson, J.F., Reith, S., 1993. Characterization of a strain of methicillin-resistant *Staphylococcus aureus* (EMRSA-15) by conventional and molecular methods. *Journal of Hospital Infection* 25, 45–52. [https://doi.org/10.1016/0195-6701\(93\)90007-m](https://doi.org/10.1016/0195-6701(93)90007-m)
- Robinson, D.A., Enright, M.C., 2003. Evolutionary Models of the Emergence of Methicillin-Resistant *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy* 47, 3926–3934. <https://doi.org/10.1128/aac.47.12.3926-3934.2003>
- Robinson, D.A., Enright, M.C., 2004. Evolution of *Staphylococcus aureus* by Large Chromosomal Replacements. *Journal of Bacteriology* 186, 1060–1064. <https://doi.org/10.1128/jb.186.4.1060-1064.2004>
- Robinson, D.A., Kearns, A.M., Holmes, A., Morrison, D., Grundmann, H., Edwards, G., O'Brien, F.G., Tenover, F.C., McDougal, L.K., Monk, A.B., Enright, M.C., 2005. Re-Emergence of Early Pandemic *Staphylococcus aureus* As a Community-Acquired Methicillin-Resistant Clone. *The Lancet* 365, 1256–1258. [https://doi.org/10.1016/s0140-6736\(05\)74814-5](https://doi.org/10.1016/s0140-6736(05)74814-5)

References

- Rodríguez-Noriega, E., Seas, C., 2010. The Changing Pattern of Methicillin-Resistant *Staphylococcus aureus* Clones in Latin America: Implications For Clinical Practice in the Region. *Brazilian Journal of Infectious Diseases* 14, 87–96. <https://doi.org/10.1590/s1413-86702010000800004>
- Rolinson, G.N., Batchelor, F.R., Stevens, S., Wood, J.C., Chain, E.B., 1960. Bacteriological Studies on a New Penicillin—Brl.1241. *The Lancet* 276, 564–567. [https://doi.org/10.1016/s0140-6736\(60\)91642-1](https://doi.org/10.1016/s0140-6736(60)91642-1)
- Rolo, J., Worning, P., Nielsen, J.B., Bowden, R., Bouchami, O., Damborg, P., Guardabassi, L., Perreten, V., Tomasz, A., Westh, H., de Lencastre, H., Miragaia, M., 2017. Evolutionary Origin of the Staphylococcal Cassette Chromosome *mec* (SCC*mec*). *Antimicrobial Agents and Chemotherapy* 61(6), e02302-16. <https://doi.org/10.1128/aac.02302-16>
- 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. *Microbiology* 133, 3039–3052. <https://doi.org/10.1099/00221287-133-11-3039>
- Rountree, P.M., Beard, M.A., 1968. Hospital Strains of *Staphylococcus aureus*, with Particular Reference To Methicillin-Resistant Strains. *The medical journal of Australia* 2, 1163–1168.
- Sabtcheva, S., Galimand, M., Gerbaud, G., Courvalin, P., Lambert, T., 2003. Aminoglycoside Resistance Gene *ant(4')-IIb* of *Pseudomonas aeruginosa* BM4492, a Clinical Isolate from Bulgaria. *Antimicrobial Agents and Chemotherapy* 47, 1584–1588. <https://doi.org/10.1128/aac.47.5.1584-1588.2003>
- San Millan, A., MacLean, R.C., 2017. Fitness Costs of Plasmids: a Limit to Plasmid Transmission. *Microbiology Spectrum* 5. <https://doi.org/10.1128/microbiolspec.mtbp-0016-2017>
- San Millan, A., Peña-Miller, R., Toll-Riera, M., Halbert, Z.V., McLean, A.R., Cooper, B.S., MacLean, R.C., 2014. Positive Selection and Compensatory Adaptation Interact To Stabilize Non-Transmissible Plasmids. *Nature Communications* 5. <https://doi.org/10.1038/ncomms6208>
- SAS Institute Inc., 2019, JMP®, SAS Institute Inc, accessed Aug 2019 <<https://www.jmp.com>>
- Schmedes, S.E., King, J.L., Budowle, B., 2015. Correcting Inconsistencies and Errors in Bacterial Genome Metadata Using an Automated Curation Tool in Excel (AutoCurE). *Frontiers in Bioengineering and Biotechnology* 3. <https://doi.org/10.3389/fbioe.2015.00138>
- Scicluna, E.A., Shore, A.C., Thürmer, A., Ehricht, R., Slickers, P., Borg, M.A., Coleman, D.C., Monecke, S., 2009. Characterisation of MRSA from Malta and the Description of a Maltese Epidemic MRSA Strain. *European Journal of Clinical Microbiology & Infectious Diseases* 29, 163–170. <https://doi.org/10.1007/s10096-009-0834-1>
- Seemann, T., 2007, *snp-dists*, Seeman, T., accessed Dec 2019 <<https://github.com/tseemann/snp-dists>>
- Seemann, T., 2014. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* 30, 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Senok, A., Ehricht, R., Monecke, S., Al-Saedan, R., Somily, A., 2016. Molecular Characterization of Methicillin-Resistant *Staphylococcus aureus* in Nosocomial Infections in a Tertiary-Care Facility: Emergence of New Clonal Complexes in Saudi Arabia. *New Microbes and New Infections* 14, 13–18. <https://doi.org/10.1016/j.nmni.2016.07.009>
- Seybold, U., Kourbatova, E.V., Johnson, J.G., Halvosa, S.J., Wang, Y.F., King, M.D., Ray, S.M., Blumberg, H.M., 2006. Emergence of Community-Associated Methicillin-Resistant *Staphylococcus aureus* USA300 Genotype as a Major Cause of Health Care--Associated Blood Stream Infections. *Clinical Infectious Diseases* 42, 647–656. <https://doi.org/10.1086/499815>

- Shaheen, M., Tantary, H.A., Nabi, S.U., 2015. A Treatise on Bovine Mastitis: Disease and Disease Economics, Etiological Basis, Risk Factors, Impact on Human Health, Therapeutic Management, Prevention and Control Strategy. *Advances in Dairy Research* 4, 1. <https://doi.org/10.4172/2329-888x.1000150>
- Shahsavan, S., Jabalameli, L., Maleknejad, P., Aligholi, M., Imaneini, H., Jabalameli, F., Halimi, S., Taherikalani, M., Khoramian, B., Eslampour, M., Feizabadi, M., Emaneini, M., 2011. Molecular Analysis and Antimicrobial Susceptibility of Methicillin Resistant *Staphylococcus aureus* in One of the Hospitals of Tehran University of Medical Sciences: High Prevalence of Sequence Type 239 (ST239) Clone. *Acta Microbiologica et Immunologica Hungarica* 58, 31–39. <https://doi.org/10.1556/amicr.58.2011.1.4>
- Shang, W., Hu, Q., Yuan, W., Cheng, H., Yang, J., Hu, Z., Yuan, J., Zhang, X., Peng, H., Yang, Y., Hu, X., Li, M., Zhu, J., Rao, X., 2016. Comparative Fitness and Determinants for the Characteristic Drug Resistance of ST239-MRSA-III-t030 and ST239-MRSA-III-t037 Strains Isolated in China. *Microbial Drug Resistance* 22, 185–192. <https://doi.org/10.1089/mdr.2015.0226>
- Sheppard, S.K., Didelot, X., Jolley, K.A., Darling, A.E., Pascoe, B., Meric, G., Kelly, D.J., Cody, A., Colles, F.M., Strachan, N.J.C., Ogden, I.D., Forbes, K., French, N.P., Carter, P., Miller, W.G., McCarthy, N.D., Owen, R., Litrup, E., Egholm, M., Affourtit, J.P., Bentley, S.D., Parkhill, J., Maiden, M.C.J., Falush, D., 2012. Progressive Genome Wide Introgression in Agricultural *Campylobacter Coli*. *Molecular Ecology* 22, 1051–1064. <https://doi.org/10.1111/mec.12162>
- Shittu, A., Nubel, U., Udo, E., Lin, J., Gaogakwe, S., 2009. Characterization of Meticillin-Resistant *Staphylococcus aureus* Isolates from Hospitals in Kwazulu-Natal Province, Republic of South Africa. *Journal of Medical Microbiology* 58, 1219–1226. <https://doi.org/10.1099/jmm.0.011452-0>
- Sigma-Aldrich, 2018, 22092 Tryptic Soy Broth, Sigma-Aldrich, accessed Aug 2019
<<https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma-Aldrich/Datasheet/1/22092dat.pdf>>
- Simor, A.E., Ofner Agostini, M., Bryce, E., McGeer, A., Paton, S., Mulvey, M.R., 2002. Laboratory Characterization of Methicillin Resistant *Staphylococcus aureus* in Canadian Hospitals: Results of 5 Years of National Surveillance, 1995–1999. *The Journal of Infectious Diseases* 186, 652–660. <https://doi.org/10.1086/342292>
- Smyth, D.S., McDougal, L.K., Gran, F.W., Manoharan, A., Enright, M.C., Song, J.-H., de Lencastre, H., Robinson, D.A., 2010. Population Structure of a Hybrid Clonal Group of Methicillin-Resistant *Staphylococcus aureus*, ST239-MRSA-III. *PLoS ONE* 5, e8582. <https://doi.org/10.1371/journal.pone.0008582>
- Soares, M.J.D.S., Figueiredo, A.M.S., Carvalho, M.C.D.S., Teixeira, L.A., Ferreira-Carvalho, B.T., Nunes, M.D.R., 2001. Analysis of Different Molecular Methods For Typing Methicillin-Resistant *Staphylococcus aureus* Isolates Belonging To the Brazilian Epidemic Clone. *Journal of Medical Microbiology* 50, 732–742. <https://doi.org/10.1099/0022-1317-50-8-732>
- Soerensen, K.E., Olsen, H.G., Skovgaard, K., Wiinberg, B., Nielsen, O.L., Leifsson, P.S., Jensen, H.E., Kristensen, A.T., Iburg, T.M., 2013. Disseminated Intravascular Coagulation in a Novel Porcine Model of Severe *Staphylococcus aureus* Sepsis Fulfills Human Clinical Criteria. *Journal of Comparative Pathology* 149, 463–474. <https://doi.org/10.1016/j.jcpa.2013.04.003>
- Song, J.-H., Hsueh, P.-R., Chung, D.R., Ko, K.S., Kang, C.-I., Peck, K.R., Yeom, J.-S., Kim, S.-W., Chang, H.-H., Kim, Y.-S., Jung, S.-I., Son, J.S., So, T.M., Lalitha, M.K., Yang, Y., Huang, S.-G., Wang, H., Lu, Q., Carlos, C.C., Perera, J.A., Chiu, C.-H., Liu, J.-W., Chongthaleong, A., Thamlikitkul, V., Van, P.H., Song, J.-H., Chung, D.R., Yeom, J.-S., Lee, H., Kim, S.-W., Chang, H.-H., Kim, Y.-S., Jung, S.-I., Son, J.S., So, T.M.K., Thamlikitkul, V., Chongthaleong, A., Hsueh, P.-R., Chiu, C.-H., Liu, D.J.-W., Lalitha, M.K., Mathai, D., Perera, J., Hung Van, P., Van Ngoc, T., Carlos, C.C., 2011. Spread

References

- of Methicillin-Resistant *Staphylococcus aureus* Between the Community and the Hospitals in Asian Countries: An ANSORP Study. *Journal of Antimicrobial Chemotherapy* 66, 1061–1069.
<https://doi.org/10.1093/jac/dkr024>
- Spoor, L.E., Richardson, E., Richards, A.C., Wilson, G.J., Mendonca, C., Gupta, R.K., McAdam, P.R., Nutbeam-Tuffs, S., Black, N.S., O’Gara, J.P., Lee, C.Y., Corander, J., Ross Fitzgerald, J., 2015. Recombination-mediated remodelling of host–pathogen interactions during *Staphylococcus aureus* niche adaptation. *Microbial Genomics* 1. <https://doi.org/10.1099/mgen.0.000036>
- Stamatakis, A., 2014. RAxML Version 8: a Tool For Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* 30, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stannek, L., Egelkamp, R., Gunka, K., Commichau, F.M., 2014. Monitoring Intraspecies Competition in a Bacterial Cell Population by Cocultivation of Fluorescently Labelled Strains. *Journal of Visualized Experiments*. <https://doi.org/10.3791/51196>
- Stern, D.L., 2013. The Genetic Causes of Convergent Evolution. *Nature Reviews Genetics* 14, 751–764. <https://doi.org/10.1038/nrg3483>
- Stoltzfus, A., McCandlish, D.M., 2017. Mutational Biases Influence Parallel Adaptation. *Molecular Biology and Evolution* 34, 2163–2172. <https://doi.org/10.1093/molbev/msx180>
- Strauß, L., Stegger, M., Akpaka, P.E., Alabi, A., Breurec, S., Coombs, G., Egyir, B., Larsen, A.R., Laurent, F., Monecke, S., Peters, G., Skov, R., Strommenger, B., Vandenesch, F., Schaumburg, F., Mellmann, A., 2017. Origin, Evolution, and Global Transmission of Community-Acquired *Staphylococcus aureus* ST8. *Proceedings of the National Academy of Sciences* 114, E10596–E10604. <https://doi.org/10.1073/pnas.1702472114>
- Stryjewski, M.E., Corey, G.R., 2013. Methicillin-Resistant *Staphylococcus aureus*: An Evolving Pathogen. *Clinical Infectious Diseases* 58, S10–S19. <https://doi.org/10.1093/cid/cit613>
- Stubbendieck, R.M., Straight, P.D., 2016. Multifaceted Interfaces of Bacterial Competition. *Journal of Bacteriology* 198, 2145–2155. <https://doi.org/10.1128/jb.00275-16>
- Stubbendieck, R.M., Vargas-Bautista, C., Straight, P.D., 2016. Bacterial Communities: Interactions to Scale. *Frontiers in Microbiology* 7. <https://doi.org/10.3389/fmicb.2016.01234>
- Talan, D.A., Krishnadasan, A., Gorwitz, R.J., Fosheim, G.E., Limbago, B., Albrecht, V., Moran, G.J., 2011. Comparison of *Staphylococcus aureus* from Skin and Soft-Tissue Infections in US Emergency Department Patients, 2004 and 2008. *Clinical Infectious Diseases* 53, 144–149. <https://doi.org/10.1093/cid/cir308>
- Teixeira, L.A., Resende, C.A., Ormonde, L.R., Rosenbaum, R., Figueiredo, A.M., de Lencastre, H., Tomasz, A., 1995. Geographic Spread of Epidemic Multiresistant *Staphylococcus aureus* Clone in Brazil. *Journal of clinical microbiology* 33, 2400–2404.
- Tekeli, A., Ocal, D.N., Ozmen, B.B., Karahan, Z.C., Dolapci, I., 2016. Molecular Characterization of Methicillin-Resistant *Staphylococcus aureus* Bloodstream Isolates in a Turkish University Hospital Between 2002 and 2012. *Microbial Drug Resistance* 22, 564–569. <https://doi.org/10.1089/mdr.2015.0116>
- Teo, J., Tan, T.Y., Hon, P.Y., Lee, W., Koh, T.H., Krishnan, P., Hsu, L.Y., 2012. ST22 and ST239 MRSA Duopoly in Singaporean Hospitals: 2006–2010. *Epidemiology and Infection* 141, 153–157. <https://doi.org/10.1017/s0950268812000337>
- Tgen North, 2017, Rapid typing of *Staphylococcus aureus* CC8 strains, NCBI, accessed Aug 2019 <https://www.ncbi.nlm.nih.gov/sra?LinkName=biosample_sra&from_uid=6321027>

- The National Institute for Health and Care Excellence, 2019, MRSA Management, British National Formulary, accessed Aug 2019 <<https://bnf.nice.org.uk/treatment-summary/mrsa.html>>
- The UniProt Consortium., 2018. UniProt: a Worldwide Hub of Protein Knowledge. *Nucleic Acids Research* 47, D506–D515. <https://doi.org/10.1093/nar/gky1049>
- Thermo Fisher Scientific, 2001, Dehydrated Culture Media Brain Heart Infusion Broth, Thermo Fisher Scientific Inc., accessed Aug 2019, <http://www.oxoid.com/UK/blue/prod_detail/prod_detail.asp?pr=CM1135&c=UK&lang=EN&org=&print=Y>
- Thermo Fisher Scientific, 2002, Porcine Serum, Thermo Fisher Scientific, accessed Aug 2019 <<https://www.thermofisher.com/uk/en/home/life-science/cell-culture/mammalian-cell-culture/fbs/other-sera/porcine-serum.html>>
- 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* 194, 2740–2741. <https://doi.org/10.1128/jb.00248-12>
- Thomason, L.C., Costantino, N., Court, D.L., 2007. *E. coli* Genome Manipulation by P1 Transduction. *Current Protocols in Molecular Biology* 1, 1.17. <https://doi.org/10.1002/0471142727.mb0117s79>
- 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, 363–376. <https://doi.org/10.1534/genetics.116.195784>
- Thorpe, H.A., Bayliss, S.C., Sheppard, S.K., Feil, E.J., 2018. Piggy: a rapid, large-scale pan-genome analysis tool for intergenic regions in bacteria. *GigaScience* 7. <https://doi.org/10.1093/gigascience/giy015>
- Toleman, M.S., Reuter, S., Coll, F., Harrison, E.M., Blane, B., Brown, N.M., Török, M.E., Parkhill, J., Peacock, S.J., 2016. Systematic Surveillance Detects Multiple Silent Introductions and Household Transmission of Methicillin-Resistant *Staphylococcus aureus* USA300 in the East of England. *Journal of Infectious Diseases* 214, 447–453. <https://doi.org/10.1093/infdis/jiw166>
- Tong, S.Y.C., Holden, M.T.G., Nickerson, E.K., Cooper, B.S., Köser, C.U., Cori, A., Jombart, T., Cauchemez, S., Fraser, C., Wuthiekanun, V., Thaipadungpanit, J., Hongsuwan, M., Day, N.P., Limmathurotsakul, D., Parkhill, J., Peacock, S.J., 2014. Genome Sequencing Defines Phylogeny and Spread of Methicillin-Resistant *Staphylococcus aureus* in a High Transmission Setting. *Genome Research* 25, 111–118. <https://doi.org/10.1101/gr.174730.114>
- Tran, T.T., Munita, J.M., Arias, C.A., 2015. Mechanisms of Drug Resistance: Daptomycin Resistance. *Annals of the New York Academy of Sciences* 1354, 32–53. <https://doi.org/10.1111/nyas.12948>
- Trieu-Cuot, P., Courvalin, P., 1983. Nucleotide Sequence of the *Streptococcus Faecalis* Plasmid Gene Encoding the 3'5'-Aminoglycoside Phosphotransferase Type III. *Gene* 23, 331–341. [https://doi.org/10.1016/0378-1119\(83\)90022-7](https://doi.org/10.1016/0378-1119(83)90022-7)
- Turner, N.A., Sharma-Kuinkel, B.K., Maskarinec, S.A., Eichenberger, E.M., Shah, P.P., Carugati, M., Holland, T.L., Fowler, V.G., 2019. Methicillin-Resistant *Staphylococcus aureus*: An Overview of Basic and Clinical Research. *Nature Reviews Microbiology* 17, 203–218. <https://doi.org/10.1038/s41579-018-0147-4>
- Uehara, Y., Nakama, H., Agematsu, K., Uchida, M., Kawakami, Y., Abdul Fattah, A.S.M., Maruchi, N., 2000. Bacterial Interference among Nasal Inhabitants: Eradication of *Staphylococcus aureus* from Nasal Cavities by Artificial Implantation of *Corynebacterium* sp. *Journal of Hospital Infection* 44, 127–133. <https://doi.org/10.1053/jhin.1999.0680>

References

- Uhlemann, A.-C., Otto, M., Lowy, F.D., DeLeo, F.R., 2014. Evolution of Community- and Healthcare-Associated Methicillin-Resistant *Staphylococcus aureus*. *Infection, Genetics and Evolution* 21, 563–574. <https://doi.org/10.1016/j.meegid.2013.04.030>
- Urushibara, N., Aung, M.S., Kawaguchiya, M., Kobayashi, N., 2019. Novel Staphylococcal Cassette Chromosome Mec (SCCmec) Type XIV (5A) and a Truncated SCCmec Element in SCC Composite Islands Carrying *speG* in ST5 MRSA in Japan. *Journal of Antimicrobial Chemotherapy* 75(1), 45–50. <https://doi.org/10.1093/jac/dkz406>
- van Tonder, A.J., Mistry, S., Bray, J.E., Hill, D.M.C., Cody, A.J., Farmer, C.L., Klugman, K.P., von Gottberg, A., Bentley, S.D., Parkhill, J., Jolley, K.A., Maiden, M.C.J., Brueggemann, A.B., 2014. Defining the Estimated Core Genome of Bacterial Populations Using a Bayesian Decision Model. *PLoS Computational Biology* 10, e1003788. <https://doi.org/10.1371/journal.pcbi.1003788>
- Vivoni, A.M., Diep, B.A., de Gouveia Magalhaes, A.C., Santos, K.R.N., Riley, L.W., Sensabaugh, G.F., Moreira, B.M., 2006. Clonal Composition of *Staphylococcus aureus* Isolates At a Brazilian University Hospital: Identification of International Circulating Lineages. *Journal of Clinical Microbiology* 44, 1686–1691. <https://doi.org/10.1128/jcm.44.5.1686-1691.2006>
- Vogwill, T., MacLean, R.C., 2014. The Genetic Basis of the Fitness Costs of Antimicrobial Resistance: a Meta-Analysis Approach. *Evolutionary Applications* 8, 284–295. <https://doi.org/10.1111/eva.12202>
- von Eiff, C., Becker, K., Machka, K., Stammer, H., Peters, G., 2001. Nasal Carriage As a Source of *Staphylococcus aureus* Bacteremia. *New England Journal of Medicine* 344, 11–16. <https://doi.org/10.1056/nejm200101043440102>
- Vos, M., Hesselman, M.C., te Beek, T.A., van Passel, M.W.J., Eyre-Walker, A., 2015. Rates of Lateral Gene Transfer in Prokaryotes: High But Why? *Trends in Microbiology* 23, 598–605. <https://doi.org/10.1016/j.tim.2015.07.006>
- Votintseva, A.A., Fung, R., Miller, R.R., Knox, K., Godwin, H., Wyllie, D.H., Bowden, R., Crook, D.W., Walker, A.S., 2014. Prevalence of *Staphylococcus aureus* Protein a (*spa*) Mutants in the Community and Hospitals in Oxfordshire. *BMC Microbiology* 14, 63. <https://doi.org/10.1186/1471-2180-14-63>
- Wang, G., 2014. Human Antimicrobial Peptides and Proteins. *Pharmaceuticals* 7, 545–594. <https://doi.org/10.3390/ph7050545>
- Wang, S.-H., Khan, Y., Hines, L., Mediavilla, J.R., Zhang, L., Chen, L., Hoet, A., Bannerman, T., Pancholi, P., Robinson, D.A., Kreiswirth, B.N., Stevenson, K.B., 2012. Methicillin-Resistant *Staphylococcus aureus* Sequence Type 239-III, Ohio, USA, 2007–2009. *Emerging Infectious Diseases* 18, 1557–1565. <https://doi.org/10.3201/eid1810.120468>
- Wang, W.-Y., Chiueh, T.-S., Sun, J.-R., Tsao, S.-M., Lu, J.-J., 2012. Molecular Typing and Phenotype Characterization of Methicillin-Resistant *Staphylococcus aureus* Isolates from Blood in Taiwan. *PLoS ONE* 7, e30394. <https://doi.org/10.1371/journal.pone.0030394>
- Wang, Z., Zhou, H., Wang, H., Chen, H., Leung, K.K., Tsui, S., Ip, M., 2015. Data From: Comparative Genomics of Methicillin-Resistant *Staphylococcus aureus* ST239: Distinct Geographical Variants in Beijing and Hong Kong. <https://doi.org/10.5061/DRYAD.12773>
- Wellcome Sanger Institute, 2011. Major lineages of healthcare-associated methicillin-resistant *Staphylococcus aureus* in Singapore, EMBL-EBI, accessed Aug 2019
<<https://www.ebi.ac.uk/ena/data/view/PRJEB2295>>
- Wellcome Sanger Institute, 2019. Darwin Tree of Life, Wellcome Sanger Institute, accessed Dec 2019
<<https://www.darwintreeoflife.org/>>

- Wertheim, H.F., Melles, D.C., Vos, M.C., van Leeuwen, W., van Belkum, A., Verbrugh, H.A., Nouwen, J.L., 2005. The Role of Nasal Carriage in *Staphylococcus aureus* Infections. *The Lancet Infectious Diseases* 5, 751–762. [https://doi.org/10.1016/s1473-3099\(05\)70295-4](https://doi.org/10.1016/s1473-3099(05)70295-4)
- Wertheim, H.F., Vos, M.C., Ott, A., van Belkum, A., Voss, A., Kluytmans, J.A., van Keulen, P.H., Vandenbroucke-Grauls, C.M., Meester, M.H., Verbrugh, H.A., 2004. Risk and Outcome of Nosocomial *Staphylococcus aureus* Bacteraemia in Nasal Carriers Versus Non-Carriers. *The Lancet* 364, 703–705. [https://doi.org/10.1016/s0140-6736\(04\)16897-9](https://doi.org/10.1016/s0140-6736(04)16897-9)
- Wichelhaus, T.A., 2001. Differential Effect of RpoB Mutations on Antibacterial Activities of Rifampicin and KRM-1648 Against *Staphylococcus aureus*. *Journal of Antimicrobial Chemotherapy* 47, 153–156. <https://doi.org/10.1093/jac/47.2.153>
- Williamson, D.A., Coombs, G.W., Nimmo, G.R., 2014. *Staphylococcus aureus* ‘Down Under’: Contemporary Epidemiology of *S. aureus* in Australia, New Zealand, and the South West Pacific. *Clinical Microbiology and Infection* 20, 597–604. <https://doi.org/10.1111/1469-0691.12702>
- Wilson, D.J., Shi, C., Teitelbaum, A.M., Gulick, A.M., Aldrich, C.C., 2013. Characterization of AusA: a Dimodular Nonribosomal Peptide Synthetase Responsible for the Production of aureusimine Pyrazinones. *Biochemistry* 52, 926–937. <https://doi.org/10.1021/bi301330q>
- Wiser, M.J., Lenski, R.E., 2015. A Comparison of Methods to Measure Fitness in *Escherichia coli*. *PLOS ONE* 10, e0126210. <https://doi.org/10.1371/journal.pone.0126210>
- Wiser, M.J., Ribick, N., Lenski, R.E., 2013. Long-Term Dynamics of Adaptation in Asexual Populations. *Science* 342, 1364–1367. <https://doi.org/10.1126/science.1243357>
- Witte, W., Cuny, C., Klare, I., Nübel, U., Strommenger, B., Werner, G., 2008. Emergence and spread of antibiotic-resistant Gram-positive bacterial pathogens. *International Journal of Medical Microbiology* 298, 365–377. <https://doi.org/10.1016/j.ijmm.2007.10.005>
- Woods, R., Schneider, D., Winkworth, C.L., Riley, M.A., Lenski, R.E., 2006. Tests of Parallel Molecular Evolution in a Long-Term Experiment with *Escherichia coli*. *Proceedings of the National Academy of Sciences* 103, 9107–9112. <https://doi.org/10.1073/pnas.0602917103>
- World Health Organisation, 2014, Antimicrobial Resistance: Global Report on Surveillance 2014, World Health Organisation, accessed Jan 2020
<<https://www.who.int/drugresistance/documents/surveillancereport/en/>>
- World Health Organisation., 2017. Global Priority List of Antibiotic-Resistant Bacteria, accessed December 2019. <<https://www.who.int/medicines/publications/global-priority-list-antibiotic-resistant-bacteria/en/>>
- Xu, B.L., Zhang, G., Ye, H.F., Feil, E.J., Chen, G.R., Zhou, X.M., Zhan, X.M., Chen, S.M., Pan, W.B., 2009. Predominance of the Hungarian Clone (ST 239-III) among Hospital-Acquired Methicillin-Resistant *Staphylococcus aureus* Isolates Recovered Throughout Mainland China. *Journal of Hospital Infection* 71, 245–255. <https://doi.org/10.1016/j.jhin.2008.10.029>
- Xue, H., Lu, H., Zhao, X., 2011. Sequence Diversities of Serine-Aspartate Repeat Genes among *Staphylococcus aureus* Isolates from Different Hosts Presumably by Horizontal Gene Transfer. *PLoS ONE* 6, e20332. <https://doi.org/10.1371/journal.pone.0020332>
- Yamamoto, T., Takano, T., Higuchi, W., Iwao, Y., Singur, O., Reva, I., Otsuka, Y., Nakayashiki, T., Mori, H., Reva, G., Kuznetsov, V., Potapov, V., 2012. Comparative Genomics and Drug Resistance of a Geographic Variant of ST239 Methicillin-Resistant *Staphylococcus aureus* Emerged in Russia. *PLoS ONE* 7, e29187. <https://doi.org/10.1371/journal.pone.0029187>
- Yeap, A.D., Woods, K., Dance, D.A.B., Pichon, B., Rattanavong, S., Davong, V., Phetsouvanh, R., Newton, P.N., Shetty, N., Kearns, A.M., 2017. Molecular Epidemiology of *Staphylococcus aureus*

References

- Skin and Soft Tissue Infections in the Lao People's Democratic Republic. *The American Journal of Tropical Medicine and Hygiene* 97, 423–428. <https://doi.org/10.4269/ajtmh.16-0746>
- Yoon, H., Gros, P., Heffron, F., 2010. Quantitative PCR-Based Competitive Index for High-Throughput Screening of Salmonella Virulence Factors. *Infection and Immunity* 79, 360–368. <https://doi.org/10.1128/iai.00873-10>
- Zackrisson, M., Hallin, J., Ottosson, L.-G., Dahl, P., Fernandez-Parada, E., Ländström, E., Fernandez-Ricaud, L., Kaferle, P., Skyman, A., Stenberg, S., Omholt, S., Petrovič, U., Warringer, J., Blomberg, A., 2016. Scan-o-matic: High-Resolution Microbial Phenomics at a Massive Scale. *G3; Genes|Genomes|Genetics* 6, 3003–3014. <https://doi.org/10.1534/g3.116.032342>
- Zafar, A., Stone, M., Ibrahim, S., Parveen, Z., Hasan, Z., Khan, E., Hasan, R., Wain, J., Bamford, K., 2010. Prevalent Genotypes of Methicillin-Resistant Staphylococcus aureus: Report from Pakistan. *Journal of Medical Microbiology* 60, 56–62. <https://doi.org/10.1099/jmm.0.022707-0>
- Zeitlinger, M.A., Derendorf, H., Mouton, J.W., Cars, O., Craig, W.A., Andes, D., Theuretzbacher, U., 2011. Protein Binding: Do We Ever Learn? *Antimicrobial Agents and Chemotherapy* 55, 3067–3074. <https://doi.org/10.1128/aac.01433-10>
- Zhang, K., McClure, J.-A., Elsayed, S., Conly, J.M., 2008. Novel Staphylococcal Cassette Chromosome mec Type, Tentatively Designated Type VIII, Harboring Class a mec and Type 4 ccr Gene Complexes in a Canadian Epidemic Strain of Methicillin-Resistant Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy* 53, 531–540. <https://doi.org/10.1128/aac.01118-08>
- Zhang, X., Hu, Q., Yuan, W., Shang, W., Cheng, H., Yuan, J., Zhu, J., Hu, Z., Li, S., Chen, W., Hu, X., Rao, X., 2013. First Report of a Sequence Type 239 Vancomycin-Intermediate Staphylococcus aureus Isolate in Mainland China. *Diagnostic Microbiology and Infectious Disease* 77, 64–68. <https://doi.org/10.1016/j.diagmicrobio.2013.06.008>
- Zurita, J., Barba, P., Ortega-Paredes, D., Mora, M., Rivadeneira, S., 2016. Local Circulating Clones of Staphylococcus aureus in Ecuador. *The Brazilian Journal of Infectious Diseases* 20, 525–533. <https://doi.org/10.1016/j.bjid.2016.08.006>

Appendices

A. Supplementary Figures

A.1 Genome structure comparisons

A.1.1 TW20 complete genome vs ST239 assemblies

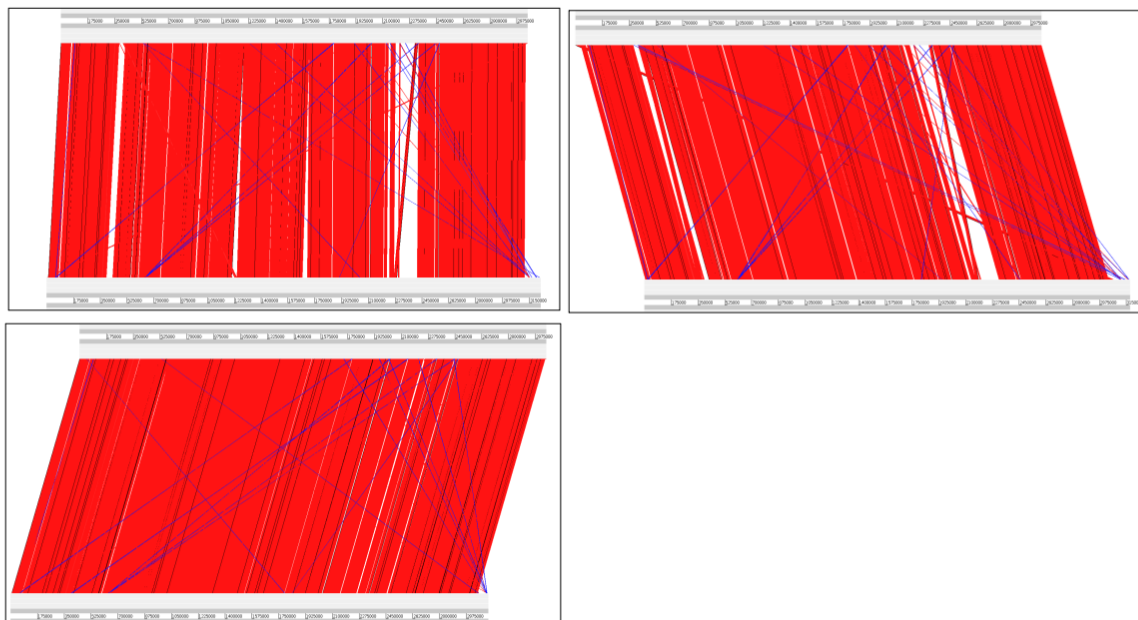


Figure A.1.1. Artemis comparison of the TW20 complete genome (top) against ST239A, ST239B and ST239D assemblies (bottom). Red, forward matches; blue, inversions.

A.1.2 CP007690 complete genome vs ST8 assemblies

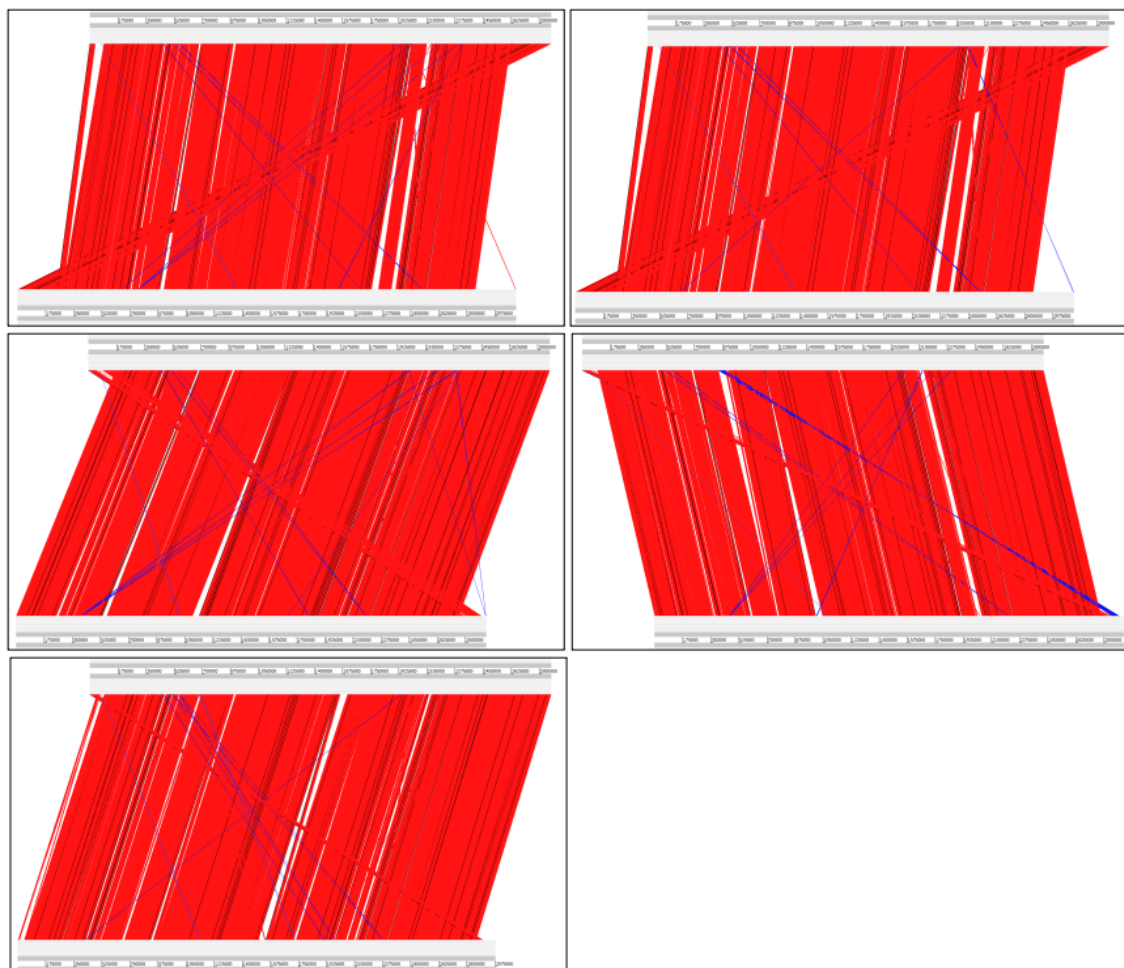


Figure A.1.2. Artemis comparison of the TW20 complete genome (top) against ST8A, ST8B and ST8C, ST8D and ST8E assemblies (bottom). Red, forward matches; blue, inversions.

A.1.3 LN626917 complete genome vs ST30 assemblies

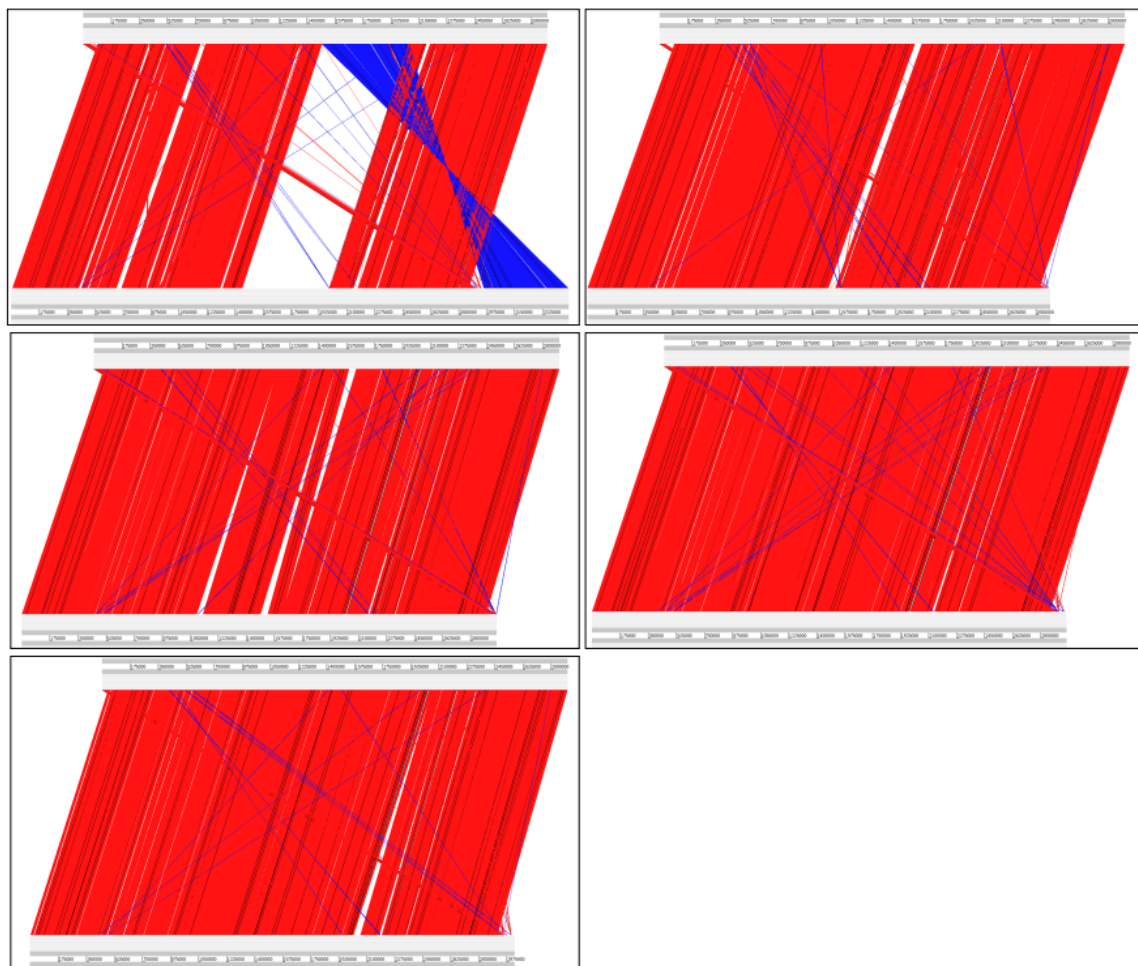


Figure A.1.3. Artemis comparison of the TW20 complete genome (top) against ST30A, ST30B, ST30C, ST30D and ST30E assemblies (bottom). Red, forward matches; blue, inversions.

A.2 SCCmec

A.2.1 SCCmec type identification

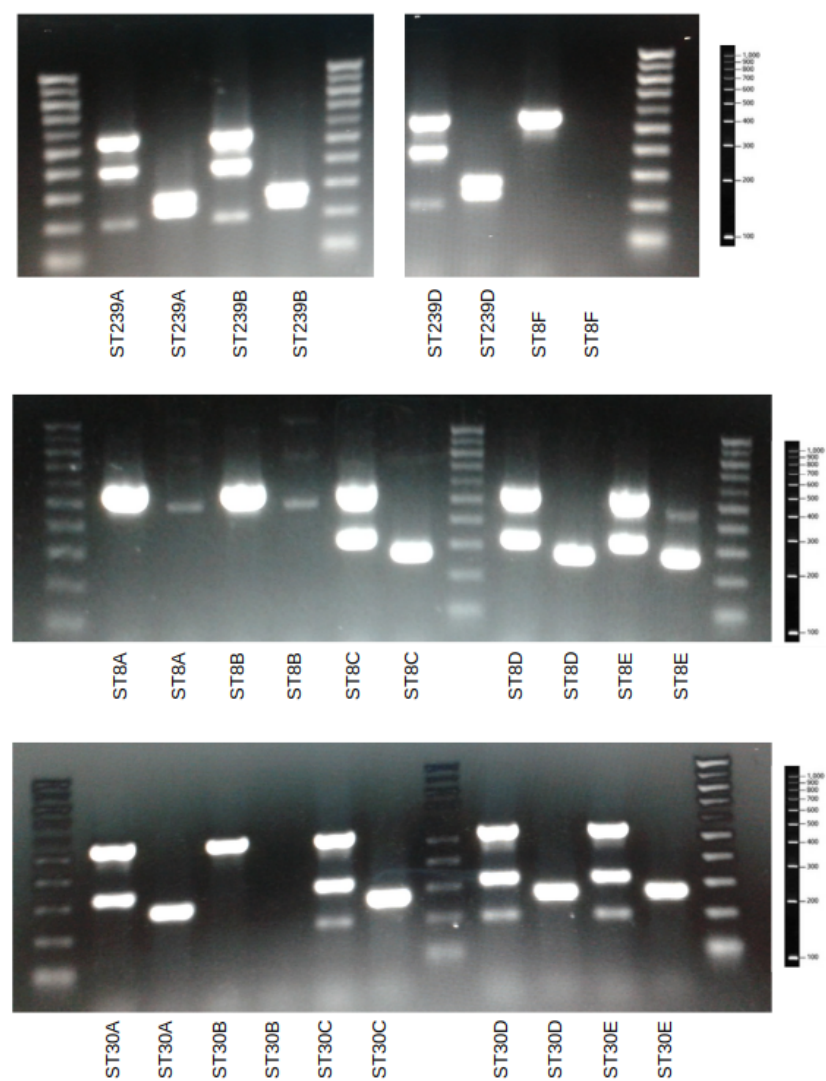


Figure A.2.1. PCR gels of SCCmec type of experimental isolates. Corresponding PCR product table in Appendix B.4.1.

A.3 OD

A.3.1 Maximum OD in TSB, BHI and PS

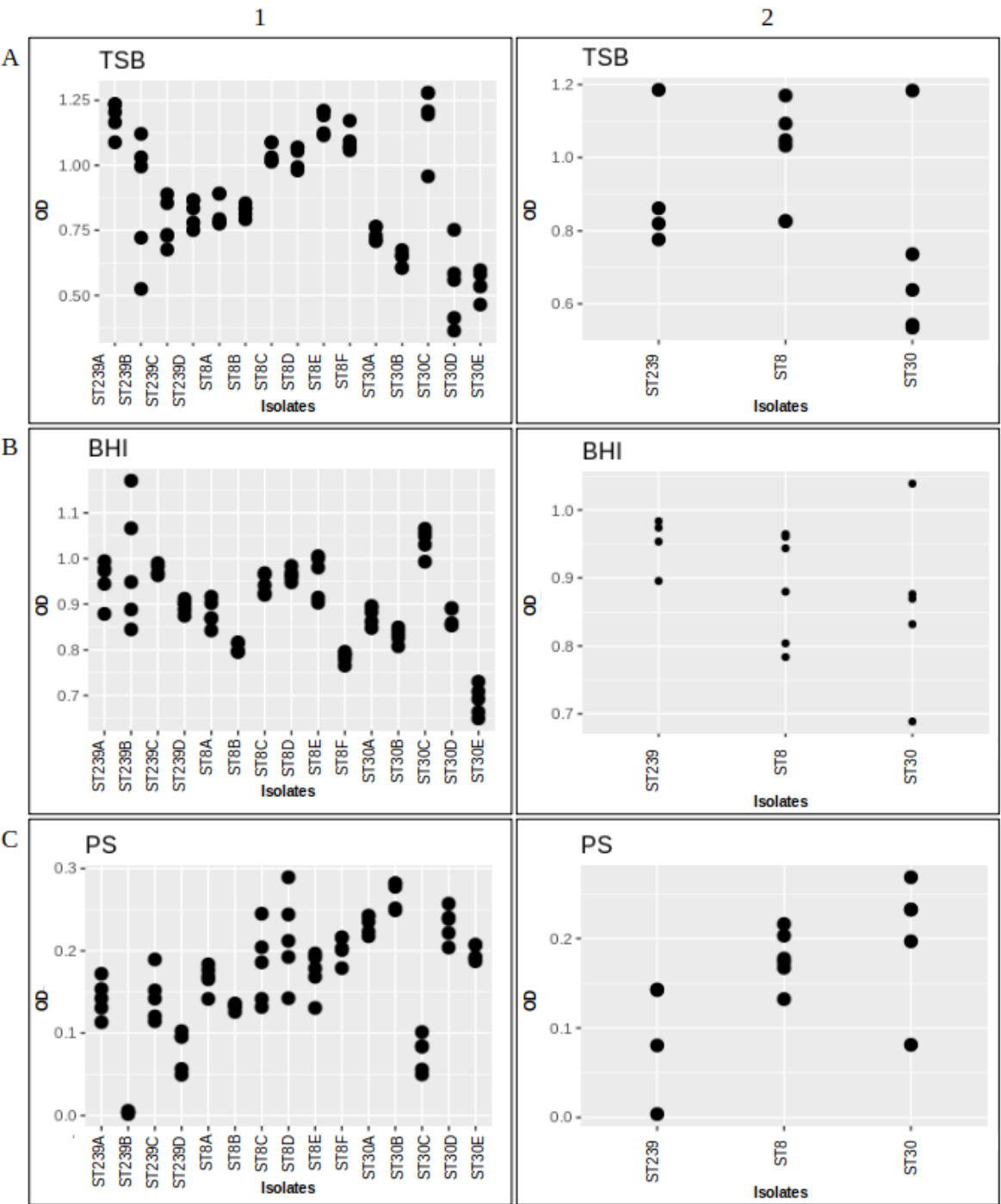


Figure A.3.1. Maximum OD vs isolate in a) TSB, b) BHI, C) PS. 1) Each point represents one repeat of the experiment, 2) Each point represents the average of all repeats for each isolate.

A.3.2 Correlation between maximum OD and maximum growth rate

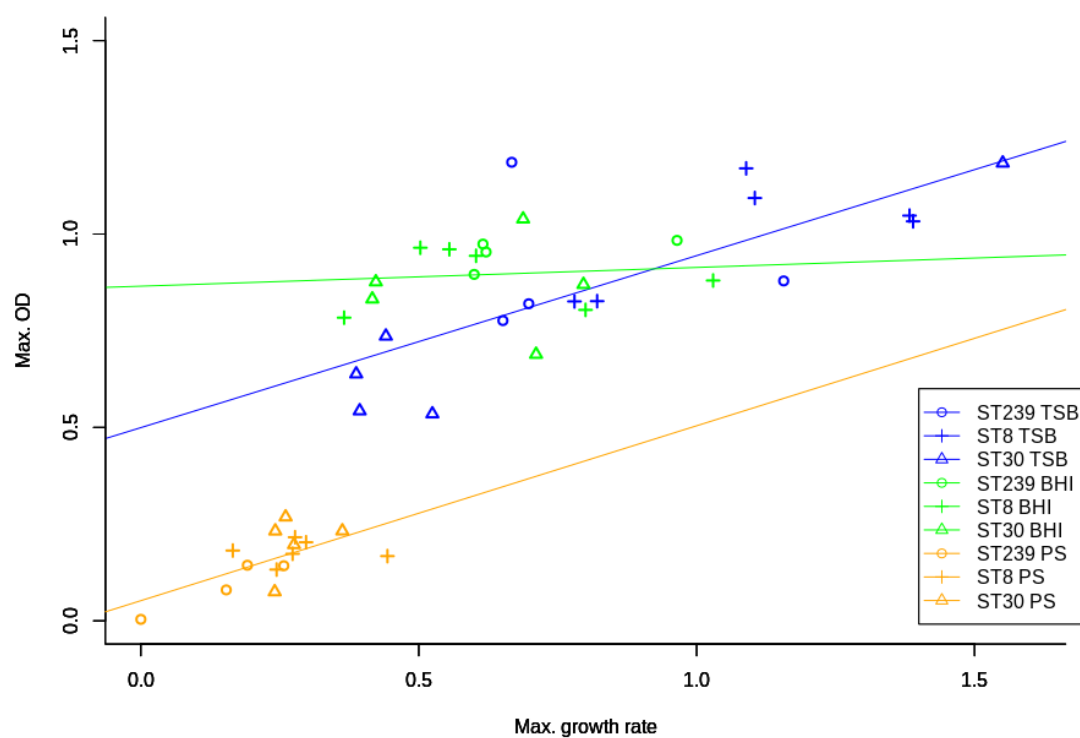


Figure A.3.2. Maximum OD vs maximum growth rate in TSB, BHI and PS.

A.4 Correlation of fitness and AMR

A.4.1 Growth rate vs AMR

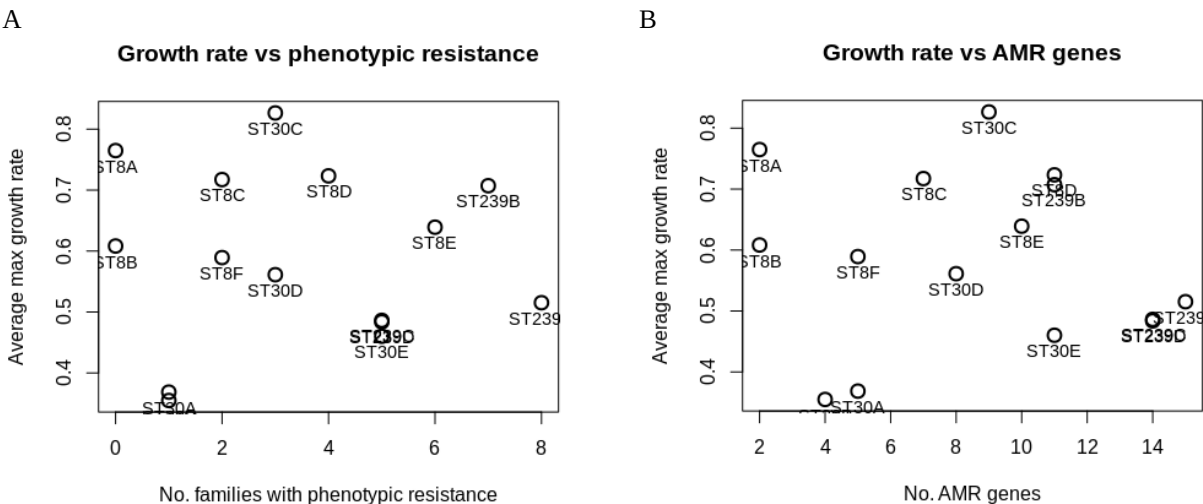


Figure A.4.1. Correlation between growth rate and a) number of phenotypic resistance families and b) number of AMR genes.

A.4.2 Competitive ability vs AMR

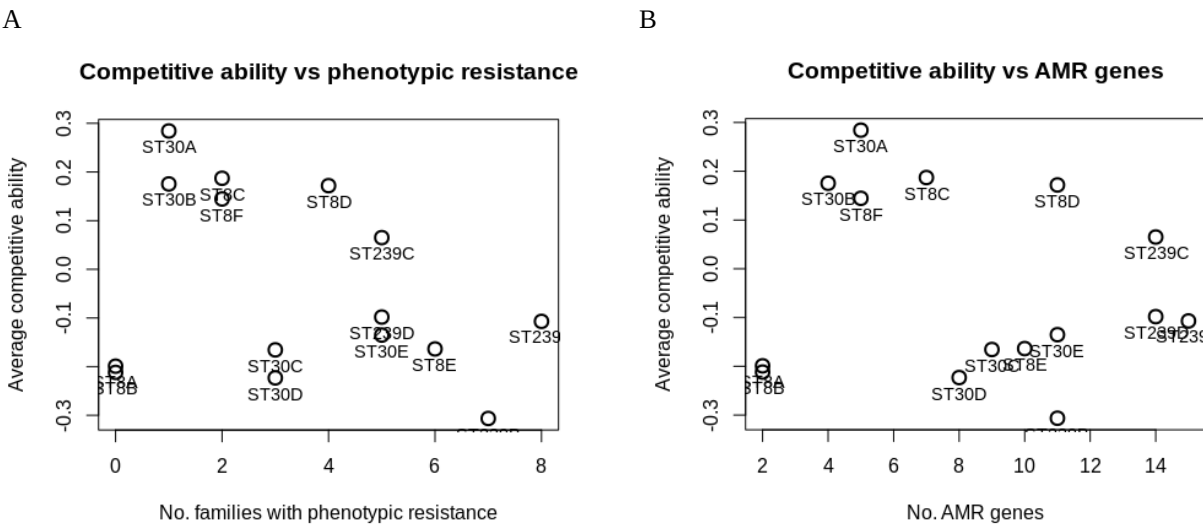


Figure A.4.2. Correlation between competitive ability and a) number of phenotypic resistance families and b) number of AMR genes.

B. Supplementary Tables

B.1 Sequence metadata and mapping quality

B.1.1 Mapping quality of ST239 sequences

Sequence ID	Country of isolation	Continent of isolation	Date of isolation	Sequencing database	Percentage mapping coverage to reference	Average depth	SNP divergence from reference
CTXC01	Argentina	South America	1996	Assembly	93.1	109.037	5814
CTXW01	Argentina	South America	1996	Assembly	92.8	109.734	5418
CTXX01	Argentina	South America	1996	Assembly	92.8	108.935	4785
AGT120	Argentina	South America	1998	Short read	94.4	94.5344	3760
ERS640727	Australia	Oceania	1980	Short read	93.7	56.8902	2710
ERS640783	Australia	Oceania	1982	Short read	94.4	81.2537	2689
ERS640788	Australia	Oceania	1998	Short read	94.4	210.931	4834
ERS640754	Australia	Oceania	2002	Short read	94.3	52.6838	4084
ERS640785	Australia	Oceania	2012	Short read	97.9	71.0821	658
CP005288	Brazil	South America	1993	Assembly	94.7	100.179	6981
CP009681	Brazil	South America	1996	Assembly	94.4	101.059	6435
CP012011	Brazil	South America	2001	Assembly	94.4	101.054	6471
LDPC01	Brazil	South America	2002	Assembly	94.5	136.225	7480
CHL1	Chile	South America	1997	Short read	94.3	93.0198	3766
CHL151	Chile	South America	1998	Short read	93.6	89.5231	3873
LGWM01	Chile	South America	2012	Assembly	92.4	107.912	5172
CHI59	China	Asia	1998	Short read	98.7	84.2657	3260
CP007447	China	Asia	2004	Assembly	98.8	100.149	2138
CP002643	China	Asia	2006	Assembly	93.4	100.741	4334
CP006838	China	Asia	2010	Assembly	97.6	99.9395	4883
3HK	Czechia	Europe	2000	Short read	94.6	93.4518	4097
2A8	Czechia	Europe	2001	Short read	94.0	80.8089	3479
DEN907	Denmark	Europe	2001	Short read	99.2	66.5416	665
CTWW01	Denmark	Europe	2006	Assembly	98.7	102.703	3530
CTXO01	Denmark	Europe	2009	Assembly	93.8	110.7	2966
CTWS01	Egypt	Africa	2005	Assembly	94.4	109.12	5880
ERS026791	France	Europe	2006	Short read	93.6	66.2764	4009
D8	Gambia	Africa	2007	Short read	100.0	73.2409	1071
GRE108	Greece	Europe	1998	Short read	90.7	76.3975	60322
GRE18	Greece	Europe	1998	Short read	95.0	86.2281	4830
GRE4	Greece	Europe	1998	Short read	94.1	98.057	5042
GRE317	Greece	Europe	1999	Short read	94.7	99.5089	4993
HUSA304	Hungary	Europe	1993	Short read	95.1	89.9131	3699
HU106	Hungary	Europe	1996	Short read	95.1	96.4601	3712
HU109	Hungary	Europe	1996	Short read	94.8	109.919	5122
HUR18	Hungary	Europe	1997	Short read	94.9	40.2685	4666
CTXS01	India	Asia	2006	Assembly	96.9	105.366	6834
LWAK01	India	Asia	2014	Assembly	97.8	102.538	6279
LWAM01	India	Asia	2014	Assembly	96.8	103.785	6648
NESW01	India	Asia	2014	Assembly	97.7	102.261	11676
MLQB01	India	Asia	2015	Assembly	97.6	105.225	6528
CTXR01	Lithuania	Europe	1996	Assembly	93.9	111.44	5180
CTYA01	Lithuania	Europe	1996	Assembly	94.6	106.112	5444

Appendices

CTYC01	Lithuania	Europe	1996	Assembly	94.6	112.297	5438
CTWZ01	Malaysia	Asia	1996	Assembly	99.3	101.73	1629
JQNU01	Malaysia	Asia	2003	Assembly	97.9	103.704	2310
AMRE01	Malaysia	Asia	2009	Assembly	97.6	105.18	1899
ANPO01	Malaysia	Asia	2009	Assembly	91.7	111.526	2127
JKD6009	New Zealand	Oceania	2003	Short read	93.6	449.374	4714
CP002120	New Zealand	Oceania	2013	Assembly	93.6	100.346	5811
LGWY01	Peru	South America	2011	Assembly	93.5	107.548	6284
LGWZ01	Peru	South America	2011	Assembly	93.5	105.885	5747
MIHO01	Peru	South America	2013	Assembly	92.4	110.194	6203
CTXV01	Poland	Europe	1996	Assembly	96.8	105.139	7264
ERS026930	Poland	Europe	2006	Short read	98.8	503.097	4873
ERS049893	Portugal	Europe	1990	Short read	92.8	101.187	3125
ICP5014	Portugal	Europe	1993	Short read	92.9	92.6949	3433
HSJ216	Portugal	Europe	1998	Short read	94.3	102.569	3798
HGSA142	Portugal	Europe	2003	Short read	94.7	88.728	4168
CTXP01	Portugal	Europe	2005	Assembly	92.7	109.508	4737
CTXK01	Romania	Europe	2007	Assembly	93.8	107.855	6710
SRS1497005	Russia	Europe	2011	Short read	94.3	93.0183	4700
SRS1496939	Russia	Europe	2013	Short read	93.9	99.2495	4154
SRS2490451	Russia	Europe	2016	Short read	94.0	168.066	4280
ERR053014	Singapore	Asia	1982	Short read	99.8	395.229	1252
ERR053016	Singapore	Asia	1985	Short read	100.0	309.058	310
ERR053031	Singapore	Asia	1997	Short read	100.0	468.551	370
ERR030243	Singapore	Asia	2002	Short read	100.0	99.2949	499
ERR029473	Singapore	Asia	2009	Short read	100.0	126.337	334
LFED02	Sout Korea	Asia	2014	Assembly	92.0	109.145	994
CP013957	South Korea	Asia	2011	Assembly	99.9	100.928	1092
CTXI01	Spain	Europe	1996	Assembly	93.4	107.662	6603
AHZL01	Switzerland	Europe	1990	Assembly	97.8	103.664	11802
ERR064946	Thailand	Asia	2006	Short read	98.1	100.248	948
CTWU01	Thailand	Asia	2007	Assembly	97.4	103.897	4380
ERR023864	Thailand	Asia	2008	Short read	99.8	154.808	1143
BPH0365	Turkey	Europe	1996	Short read	93.7	79.3955	3329
CTWE01	Turkey	Europe	2006	Assembly	93.6	107.63	6223
CP012015	Turkey	Europe	2007	Assembly	94.4	100.423	6360
CTWD01	Turkey	Europe	2007	Assembly	93.6	108.231	6328
CTWI01	Turkey	Europe	2007	Assembly	93.7	107.944	6148
CTYZ01	Turkey	Europe	2008	Assembly	92.7	109.397	6286
CTWH01	Turkey	Europe	2009	Assembly	93.4	106.977	6457
FN433596	UK	Europe	2003	Assembly	100.0	99.9745	0
ERR1212592	UK	Europe	2009	Short read	93.7	197.316	2029
FFPZ01	UK	Europe	2010	Assembly	99.9	100.595	648
FGZX01	UK	Europe	2010	Assembly	99.6	101.241	1194
SRS1974492	UK	Europe	2011	Short read	94.3	138.302	4477
URU34	Uruguay	South America	1997	Short read	94.2	80.3989	3529
URU110	Uruguay	South America	1998	Short read	94.5	89.2389	3047
JXZH02	USA	North America	1981	Assembly	94.3	127.474	4795
R35	USA	North America	1987	Short read	94.8	92.5076	4102
LHH1	USA	North America	1994	Short read	95.6	91.9445	4436
BK2421	USA	North America	1996	Short read	95.7	102.385	4027
SR2852	USA	North America	2005	Short read	100.0	181.149	754
CTXQ01	Vietnam	Asia	2004	Assembly	98.1	103.187	1949

Table B.1.1. Mapping quality of ST239 sequences.

B.1.2 Mapping quality of ST8 sequences

Sequence ID	Country of isolation	Continent of isolation	Date of isolation	Sequencing database	Percentage mapping coverage to reference	Average depth	SNP divergence from reference
ERR1535436	Australia	Oceania	1989	Short Read	96.5	146.637	1058
ERR1535439	Australia	Oceania	2003	Short Read	95.6	160.896	787
ERR1535431	Australia	Oceania	2008	Short Read	97.3	183.375	1499
ERR1535432	Australia	Oceania	2010	Short Read	96.6	181.923	1868
ERR1535433	Australia	Oceania	2011	Short Read	95.9	166.387	1573
CP007690	Belgium	Europe	1905	Assembled	100.0	99.9717	0
ERR033306	Belgium	Europe	2006	Short Read	97.2	96.7154	1599
ERR033347	Belgium	Europe	2006	Short Read	100.0	53.8818	75
CP019590	Canada	North America	2004	Assembled	100.0	100.178	74
ERR381019	Canada	North America	2008	Short Read	99.6	84.568	121
ERR380998	Canada	North America	2009	Short Read	99.9	165.294	84
ERR380994	Canada	North America	2010	Short Read	100.0	109.807	108
ERR381026	Canada	North America	2011	Short Read	99.7	121.604	157
CP007676	Colombia	South America	2006	Assembled	98.6	100.131	648
CP007672	Colombia	South America	2007	Assembled	97.8	100.22	430
CP007674	Colombia	South America	2007	Assembled	98.5	100.229	322
ERR1588671	Denmark	Europe	1957	Short Read	96.2	84.3795	1001
ERR1588666	Denmark	Europe	1968	Short Read	95.3	37.7338	830
ERR1712346	Denmark	Europe	2004	Short Read	97.0	93.5244	1215
ERR1712341	Denmark	Europe	2007	Short Read	100.0	102.415	69
SRR6169019	Denmark	Europe	2016	Short Read	96.3	58.9474	2061
SRR1955591	France	Europe	1999	Short Read	97.5	366.121	1695
ERR033450	France	Europe	2006	Short Read	97.0	20.9944	1447
ERR033462	France	Europe	2006	Short Read	97.2	35.2216	1509
CP026961	French Guiana	South America	1996	Assembled	95.7	100.437	2034
ERR1535412	Gabon	Africa	2005	Short Read	100.0	98.5555	95
ERR1535387	Gabon	Africa	2009	Short Read	95.1	111.846	1256
ERR1535352	Gabon	Africa	2011	Short Read	95.1	96.1442	1328
ERR1143478	Gabon	Africa	2012	Short Read	100.0	127.343	114
ERR1535423	Gabon	Africa	2013	Short Read	97.2	124.005	1450
ERR1213800	Gambia	Africa	2007	Short Read	95.3	26.9248	1119
ERR1213806	Gambia	Africa	2008	Short Read	95.7	29.2768	1334
SRR4359408	Germany	Europe	2001	Short Read	95.9	102.533	1702
CP003033	Germany	Europe	2002	Assembled	93.5	99.9751	984
ERR1535365	Germany	Europe	2006	Short Read	100.0	139.186	63
ERR593213	Germany	Europe	2013	Short Read	99.8	111.994	567
CP022290	Germany	Europe	2014	Assembled	97.1	100.137	2360
ERR1535441	Ghana	Africa	2005	Short Read	95.5	98.0067	1301
ERR1535358	Ghana	Africa	2011	Short Read	100.0	75.2317	84
ERR107788	Ireland	Europe	2001	Short Read	96.4	54.4136	1464
ERR118522	Ireland	Europe	2001	Short Read	95.1	94.8211	1298
ERR124443	Ireland	Europe	2004	Short Read	97.1	30.6759	1700
ERR211899	Ireland	Europe	2004	Short Read	97.3	127.205	1926
ERR124475	Ireland	Europe	2010	Short Read	94.8	17.9351	1201
ERR033701	Italy	Europe	2006	Short Read	97.8	126.635	1662
AP014921	Japan	Asia	2013	Assembled	95.5	99.9793	2018
SRR917579	Luxembourg	Europe	1997	Short Read	97.4	98.3276	1543
SRR917586	Luxembourg	Europe	2005	Short Read	97.3	112.442	1580
ERR1535363	Luxembourg	Europe	2006	Short Read	97.3	96.1561	1469
SRR917585	Luxembourg	Europe	2007	Short Read	97.5	230.219	1742

Appendices

SRR917652	Luxembourg	Europe	2010	Short Read	98.1	67.0159	1454
ERR1143491	Mozambique	Africa	2012	Short Read	95.8	110.83	1208
ERR1712350	Netherlands	Europe	2005	Short Read	96.4	107.226	1092
ERR1712351	Netherlands	Europe	2006	Short Read	97.6	99.3519	1408
ERR1712352	Netherlands	Europe	2007	Short Read	97.1	118.603	1388
SRR3658373	Russia	Europe	2013	Short Read	97.7	114.129	1374
SRR5100325	Russia	Europe	2016	Short Read	97.3	94.5864	1195
SRR5100330	Russia	Europe	2016	Short Read	97.0	71.3461	1265
ERR033296	Spain	Europe	2006	Short Read	97.2	110.849	1551
CP014362	Suriname	South America	2013	Assembled	99.6	100.079	128
CP014384	Suriname	South America	2013	Assembled	100.0	100.027	106
CP014444	Suriname	South America	2013	Assembled	100.0	100.047	108
ERR033573	Sweden	Europe	2006	Short Read	95.1	49.8074	1606
ERR033574	Sweden	Europe	2006	Short Read	95.0	54.1259	1211
SRR1786287	Switzerland	Europe	2007	Short Read	96.2	92.3943	1376
SRR1790318	Switzerland	Europe	2007	Short Read	96.4	147.861	435
ERR505339	Tanzania	Africa	2008	Short Read	96.4	138.886	1377
ERR505360	Tanzania	Africa	2008	Short Read	96.0	122.037	1466
ERR1143410	Tanzania	Africa	2011	Short Read	97.2	101.164	1539
ERR505358	Tanzania	Africa	2013	Short Read	96.8	116.885	1401
SRR5029779	UK	Europe	1935	Short Read	94.3	118.226	815
ERR172082	UK	Europe	1998	Short Read	95.9	104.943	1111
ERR129281	UK	Europe	2003	Short Read	99.8	162.976	75
ERR2072416	UK	Europe	2012	Short Read	100.0	686.167	117
SRR5651527	UK	Europe	2015	Short Read	91.4	28.2703	691
SRR1103475	USA, California	North America	2011	Short Read	97.1	140.871	1382
KK095312	USA, California	North America	2012	Assembled	99.5	100.033	127
SRR1165956	USA, California	North America	2013	Short Read	100.0	118.757	114
SRR3194930	USA, Colorado	North America	2010	Short Read	100.0	82.2425	97
SRR3194933	USA, Colorado	North America	2011	Short Read	100.0	119.898	106
CP027101	USA, DC	North America	2014	Assembled	99.4	100.323	156
SRR4140011	USA, Florida	North America	2003	Short Read	99.7	62.6281	63
SRR4140097	USA, Florida	North America	2006	Short Read	98.1	70.1798	209
SRR4140090	USA, Florida	North America	2008	Short Read	98.7	63.4958	156
SRR4140057	USA, Florida	North America	2010	Short Read	98.9	461.837	131
SRR4140079	USA, Florida	North America	2011	Short Read	100.0	438.763	123
CP016855	USA, Georgia	North America	2010	Assembled	98.3	100.231	323
CP017094	USA, Georgia	North America	2011	Assembled	99.2	100.22	133
CP025495	USA, Georgia	North America	2011	Assembled	99.5	100.181	146
SRR1786280	USA, Illinois	North America	1965	Short Read	94.1	113.393	917
SRR3194883	USA, Maryland	North America	2010	Short Read	99.5	108.074	139
SRR3194942	USA, Maryland	North America	2011	Short Read	99.9	93.4211	115
SRR497495	USA, Massachusetts	North America	2003	Short Read	98.6	121.444	415
KB820894	USA, Massachusetts	North America	2007	Assembled	100.0	99.9815	108
SRR1539657	USA, Massachusetts	North America	2012	Short Read	93.6	123.678	752
SRR5226531	USA, Massachusetts	North America	2013	Short Read	93.6	31.7962	741
SRR3168638	USA, Massachusetts	North America	2015	Short Read	97.4	157.061	1502
CP020619	USA, Nebraska	North America	2013	Assembled	99.8	100.176	76
ERR134641	USA, New York	North America	2006	Short Read	100.0	145.906	68
ERR302763	USA, New York	North America	2007	Short Read	100.0	167.417	108
ERR338634	USA, New York	North America	2008	Short Read	99.3	145.727	1319
ERR134699	USA, New York	North America	2010	Short Read	100.0	176.86	114
ERR134723	USA, New York	North America	2011	Short Read	100.0	231.72	37
SRR3194889	USA, Oregon	North America	2010	Short Read	98.4	102.386	316
SRR3194948	USA, Oregon	North America	2011	Short Read	99.7	39.2622	122
SRR3194899	USA, Pennsylvania	North America	2010	Short Read	100.0	113.523	103

SRR3194910	USA, Tennessee	North America	2010	Short Read	100.0	87.6728	93
SRR3194949	USA, Tennessee	North America	2011	Short Read	99.9	102.947	106
CP026076	USA, Texas	North America	2002	Assembled	97.6	100.201	290
CP013231	USA, Texas	North America	2013	Assembled	100.0	100.847	117
SRR4019282	USA, Texas	North America	2016	Short Read	97.7	72.5952	154

Table B.1.2. Mapping quality of ST8 sequences.

B.1.3 Mapping quality of ST30 sequences

Sequence ID	Country of isolation	Continent of isolation	Date of isolation	Sequencing database	Percentage mapping coverage to reference	Average depth	SNP divergence from reference
SRR1172654	Argentina	South America	2013	Short Read	97.0	193.057	3476
LDIT01	Brazil	South America	2014	Assembly	95.0	99.7948	2195
ERR033374	Denmark	Europe	2006	Short Read	95.8	116.531	2751
ERR1753509	Denmark	Europe	2013	Short Read	95.9	120.618	2713
ERR1753386	Denmark	Europe	2015	Short Read	95.7	31.0651	2483
ERR1592225	DRC	Africa	2013	Short Read	95.6	121.05	2384
ERR1592254	DRC	Africa	2013	Short Read	92.1	7.74494	2745
ERR1592169	DRC	Africa	2014	Short Read	94.8	123.771	2340
ERR1592321	DRC	Africa	2015	Short Read	95.0	126.799	2177
ERR033635	Finland	Europe	2006	Short Read	95.9	38.4598	3111
ERR033464	France	Europe	2006	Short Read	96.3	70.4038	3131
ERR038658	France	Europe	2006	Short Read	97.0	119.093	3441
SRR4359400	Germany	Europe	2002	Short Read	95.9	192.496	3004
ERR1143353	Germany	Europe	2010	Short Read	95.4	150.51	2069
ERR1675778	Germany	Europe	2014	Short Read	96.1	97.4517	2463
ERR1195906	Germany	Europe	2015	Short Read	95.6	143.945	2494
ERR033537	Italy	Europe	2006	Short Read	97.3	79.1786	3710
ERR033542	Italy	Europe	2006	Short Read	95.0	123.826	2337
JHEC01	Jordan	Asia	2009	Assembly	95.7	98.4373	3313
LN626917	Kenya	Africa	2004	Assembly	100.0	99.9717	0
JHEB01	Lebanon	Asia	2011	Assembly	96.0	98.1932	3834
ERR033698	Norway	Europe	2006	Short Read	96.6	87.9975	2888
ERR038674	Norway	Europe	2006	Short Read	97.5	225.336	3530
ERR033550	Poland	Europe	2006	Short Read	96.9	91.9431	4008
ERR038685	Poland	Europe	2006	Short Read	97.0	62.4093	3480
ERR038688	Poland	Europe	2006	Short Read	97.5	58.4481	3163
ERR033556	Portugal	Europe	2006	Short Read	97.5	163.204	4277
LWRA01	Russia	Europe	2013	Assembly	95.7	97.7277	4382
LWRC01	Russia	Europe	2014	Assembly	94.2	98.1535	2634
LWRD01	Russia	Europe	2014	Assembly	94.2	98.138	2577
CP009554	South Korea	Asia	2014	Assembly	96.3	100.05	4070
ERR033291	Spain	Europe	2006	Short Read	96.9	127.839	3518
ERR033578	Sweden	Europe	2006	Short Read	97.5	72.2829	3989
ERR033589	Sweden	Europe	2006	Short Read	96.5	90.625	3181
ERR033590	Sweden	Europe	2006	Short Read	96.3	84.4088	3068
ERR033593	Sweden	Europe	2006	Short Read	95.5	97.5144	2766
AHZJ01	Switzerland	Europe	2005	Assembly	93.8	98.3829	4039
JXHZ01	Switzerland	Europe	2007	Assembly	95.9	98.1402	3862
SRR1786285	Switzerland	Europe	2007	Short Read	96.1	122.44	2595
ERR1143398	Tanzania	Africa	2011	Short Read	95.6	109.485	1923
ERR1143429	Tanzania	Africa	2012	Short Read	95.7	117.893	2019

Appendices

ERR505346	Tanzania	Africa	2012	Short Read	97.9	128.562	3657
ERR505389	Tanzania	Africa	2013	Short Read	96.8	143.581	2873
ERR829972	Tanzania	Africa	2014	Short Read	97.2	182.67	3746
ERR1213763	The Gambia	Africa	2002	Short Read	95.9	55.1392	2193
ERR1213777	The Gambia	Africa	2004	Short Read	96.6	96.9302	2570
ERR1213799	The Gambia	Africa	2007	Short Read	95.7	89.7567	2314
JYAJ02	UK	Europe	1935	Assembly	96.3	99.8234	3401
ERR109495	UK	Europe	1998	Short Read	96.5	289.434	3279
ERR129280	UK	Europe	2003	Short Read	96.3	138.655	2717
SRR5251295	UK	Europe	2008	Short Read	97.1	154.612	3482
JPWO01	UK	Europe	2013	Assembly	95.5	98.8324	3522
ERR554777	USA	North America	2004	Short Read	95.2	144.628	2514
ERR554339	USA	North America	2009	Short Read	96.6	142.045	3474
CP017684	USA	North America	2011	Assembly	95.5	99.4066	3321
SRR1159198	USA	North America	2012	Short Read	96.3	169.795	2926
SRR5822163	USA	North America	2015	Short Read	95.8	104.722	1924

Table B.1.3. Mapping quality of ST30 sequences.

B.1.4 Geographical distribution of sequences from Staphopia

Country	ST			Country	ST		
	ST239	ST8	ST30		ST239	ST8	ST30
Africa				Europe			
Gabon	0	26	0	Belgium	0	12	0
Gambia	1	2	3	Denmark	4	34	11
Ghana	0	2	0	Finland	0	0	2
Mozambique	0	1	0	France	2	8	6
Tanzania	0	14	7	Germany	1	49	23
Asia				Greece	3	0	0
China	6	0	0	Hungary	4	0	0
Malaysia	6	0	0	Ireland	0	5	0
Singapore	133	0	0	Italy	0	1	2
Thailand	27	0	0	Luxembourg	0	11	0
North America				Netherlands	0	3	0
Canada	0	41	0	Norway	0	0	2
USA	4	3284	71	Poland	2	0	3
Oceania				Portugal	11	0	1
Australia	77	14	0	Russia	9	13	0
New Zealand	1	0	0	Spain	1	1	1
South America				Sweden	0	2	4
Argentina	5	0	1	Switzerland	0	2	1
Brazil	4	0	0	Turkey	3	0	0
Chile	2	0	0	UK	7	181	329

Table B.14. Geographical distribution of ST239, ST8 and ST30 sequences from Staphopia.

B.1.5 Mapping quality of experimental isolates

Isolate ID	Sequence ID	No. bases	Coverage	Depth
ST8A	WTCHG_556541_209108	2724962	0.95	286.639
ST8B	WTCHG_556541_210120	2723696	0.95	359.47
ST8C	WTCHG_556541_211132	2768249	0.96	320.091
ST8D	WTCHG_556541_212144	2756819	0.96	348.294
ST8E	WTCHG_556541_213156	2744359	0.96	338.962
ST30A	WTCHG_556541_214168	2749838	0.95	413.705
ST30B	WTCHG_556541_215180	2767691	0.95	382.294
ST30C	WTCHG_556541_216192	2820682	0.97	449.324
ST30D	WTCHG_556541_217109	2859770	0.99	423.678
ST30E	WTCHG_556541_218121	2862792	0.99	346.826
ST239A	WTCHG_556541_219133	2835252	0.93	319.971
ST239B	WTCHG_556541_220145	2820827	0.93	273.092
ST239C	WTCHG_556541_221157	4676	0.00	1.02331
ST239D	WTCHG_556541_222169	3007581	0.99	335.903
ST8F	WTCHG_556541_223181	2648510	0.87	378.069

Table B.1.5. Mapping quality of the experimental isolates, sequenced at Oxford Genomic Centre.

B.1.6 Assembly quality of experimental sequences

Isolate ID	No Contigs	N50	Percentage coverage to reference
ST239A	118	83,606	97.68
ST239B	109	95,236	96.36
ST239D	75	172,149	100.26
ST30A	83	224,504	97.13
ST30B	95	115,786	96.97
ST30C	54	150,818	98.41
ST30D	168	170,782	100.14
ST30E	113	170,845	102.20
ST8A	28	253,428	97.32
ST8B	28	186,241	97.34
ST8C	33	304,265	100.02
ST8D	54	304,381	100.18
ST8E	42	143,363	99.39
ST8F	39	671206	98.41

Table B.1.6. Assembly quality of experimental sequences.

B.1.7 Assembly quality of experimental sequences

ST	NCBI accession	Date	Location	SCCmec type
ST239	CP005288	1993	Brazil	III(3A)
ST239	CP012015	1997	Brazil	III(3A)
ST239	CP012011	2001	Brazil	III(3A)
ST239	CP007447	2004	China	III(3A)
ST239	CP006838	2010	China	III(3A)
ST239	CP002120	2013	New Zealand	III(3A)
ST239	CP013957	2011	South Korea	III(3A)
ST8	CP007690	1905	Belgium	IVa(2B)
ST8	CP019590	2004	Canada	IVa(2B)
ST8	CP007672	2007	Colombia	IVc(2B)
ST8	CP026961	1996	French Guiana	MSSA
ST8	CP022290	2014	Germany	MSSA
ST8	AP014921	2013	Japan	IV(2B)
ST8	CP014362	2013	Suriname	IVa(2B)
ST8	CP003033	2014	Columbia, USA	MSSA
ST8	CP016855	2010	Georgia, USA	IVa(2B)
ST8	CP026076	2002	Texas, USA	MSSA
ST30	LN626917	2004	Kenya	MSSA
ST30	CP017684	2011	USA	MSSA
ST30	CP009554	2014	South Korea	MSSA

Table B.1.7. Metadata of the ST239, ST8 and ST30 complete genome sequences from Figure 3.7a – c in Chapter 3.

B.1.8 Mapping isolates to the core consensus sequence

Isolate ID	Total no. bases	Average coverage	Average depth
ST239A	1944728	0.9991	346.884
ST239B	1944134	0.9988	299.388
ST239C	1944362	0.9989	97.9761
ST239D	1944666	0.9990	372.378
ST8A	1944195	0.9988	321.524
ST8B	1944097	0.9987	398.386
ST8C	1944653	0.9990	355.918
ST8D	1944164	0.9988	382.246
ST8E	1944481	0.9989	378.444
ST8F	1944597	0.9990	444.416
ST30A	1944197	0.9988	460.699
ST30B	1944135	0.9988	418.115
ST30C	1943803	0.9986	487.744
ST30D	1943856	0.9986	468.946
ST30E	1943774	0.9986	381.85

Table B.1.8. Mapping quality of the experimental sequences mapped to the core consensus sequence.

B.1.9 Identification of unique genes

Isolate ID	Number of candidate unique genes	Number of candidate unique genes >300bp	Number of true unique genes
ST239A	34	20	7
ST239B	32	19	7
ST239C	28	21	0
ST239D	19	15	3
ST8A	4	2	0
ST8B	3	2	0
ST8C	61	37	4
ST8D	66	41	0
ST8E	69	50	0
ST8F	536	68	25
ST30A	110	35	2
ST30B	34	20	4
ST30C	16	9	0
ST30D	49	45	0
ST30E	59	46	12
TOTAL	1120	430	64

Table B.1.9. Number of initial candidate unique genes and true unique genes identified for each isolate.

B.1.10 Mapping isolates to the concatenated unique gene sequence

Isolate ID	Total no. bases	Average coverage	Average depth
ST239A	131696	0.30	508.948
ST239B	132796	0.30	346.05
ST239C	105503	0.24	98.7328
ST239D	119578	0.27	384.077
ST8A	77619	0.17	330.448
ST8B	78179	0.18	403.442
ST8C	114743	0.26	367.073
ST8D	113280	0.25	490.305
ST8E	119315	0.27	482.894
ST8F	111400	0.25	452.507
ST30A	101983	0.23	411.373
ST30B	90426	0.20	495.794
ST30C	100496	0.23	539.68
ST30D	129453	0.29	396.571
ST30E	141879	0.32	418.509

Table B.1.10. Mapping quality of the experimental sequences mapped to the concatenated unique gene sequence.

B.1.11 Mapping competition to core consensus sequence

Media	Replicate	Total no. bases	Average coverage	Average depth
Mix	1	1946184	0.9998	179.956
Mix	2	1946177	0.9998	196.033
Mix	3	1946180	0.9998	168.124
Mix	4	1946180	0.9998	210.282
Mix	5	1946173	0.9998	185.291
Mix	6	1946177	0.9998	199.012
Oxa	1	1946167	0.9998	176.363
Oxa	2	1946170	0.9998	173.237
Oxa	3	1946181	0.9998	147.312
Norf	1	1946179	0.9998	189.826
Norf	2	1946180	0.9998	208.592
Norf	3	1946175	0.9998	195.399
Trim	1	1945160	0.9992	172.913
Trim	2	1945182	0.9992	181.156
Trim	3	1945479	0.9994	177.161
TSB	1	1946180	0.9998	173.403
TSB	2	1946173	0.9998	142.353
TSB	3	1946182	0.9998	209.456
BHI	1	1946175	0.9998	211.364
BHI	2	1946204	0.9998	192.808
BHI	3	1946180	0.9998	202.339
PS	1	1946184	0.9998	171.027
PS	2	1946259	0.9998	187.73
PS	3	1946182	0.9998	179.026

Table B.1.11. Mapping quality of the competition experiments mapped to the core consensus sequence. Mix, before competition; Oxa, 24 hour competition in 4 µg/mL oxacillin; Norf, 24 hour competition in 2 µg/mL norfloxacin; trim, 24 hour competition in 152 µg/mL trimethoprim/sulfamethoxazole; TSB, 24 hour competition in TSB; BHI, 24 hour competition in BHI; PS, 24 hour competition in PS.

B.1.12 Mapping competition to concatenated unique gene sequence

Media	Replicate	Total no. bases	Average coverage	Average depth
Mix	1	384626	0.87	124.044
Mix	2	374207	0.84	116.216
Mix	3	384508	0.87	124.241
Mix	4	384317	0.87	102.754
Mix	5	384287	0.87	113.67
Mix	6	373634	0.84	102.619
Oxa	1	309060	0.70	87.2476
Oxa	2	309264	0.70	127.722
Oxa	3	299367	0.67	80.8432
Norf	1	371823	0.84	119.766
Norf	2	373342	0.84	109.532
Norf	3	349781	0.79	114.608
Trim	1	316454	0.71	98.5692
Trim	2	316737	0.71	100.683
Trim	3	301885	0.68	105.864
TSB	1	372564	0.84	101.02
TSB	2	360135	0.81	89.9891

TSB	3	372441	0.84	117.873
BHI	1	366077	0.82	110.834
BHI	2	357081	0.80	135.49
BHI	3	370826	0.84	111.964
PS	1	381180	0.86	122.618
PS	2	370951	0.84	93.7115
PS	3	381935	0.86	145.397

Table B.1.12. Mapping quality of the competition experiments mapped to the concatenated unique gene sequence. Mix, before competition; Oxa, 24 hour competition in 4 µg/mL oxacillin; Norf, 24 hour competition in 2 µg/mL norfloxacin; trim, 24 hour competition in 152 µg/mL trimethoprim/sulfamethoxazole; TSB, 24 hour competition in TSB; BHI, 24 hour competition in BHI; PS, 24 hour competition in PS.

B.1.13 Average variant site coverage for competitions in oxacillin, quinolone and trimethoprim/sulfamethoxazole

Isolate ID	Condition	Replicate	No variant sites	No reads supporting non-unique variants	No reads supporting unique variants	Total no reads	Minimum detection limit	Corrected average variant coverage	Log (difference)
ST239A	Mix	1	34	5142	537	5679	0.02395	0.09456	-
ST239B	Mix	1	13	1961	65	2026	0.02567	0.03208	-
ST239C	Mix	1	35	5134	213	5347	0.02618	0.03984	-
ST239D	Mix	1	61	8867	401	9268	0.02633	0.04327	-
ST8A	Mix	1	55	7838	570	8408	0.02617	0.06779	-
ST8B	Mix	1	3	425	35	460	0.02609	0.07609	-
ST8C	Mix	1	132	17771	2026	19797	0.02667	0.10234	-
ST8D	Mix	1	156	20934	2227	23161	0.02694	0.09615	-
ST8E	Mix	1	259	34650	3723	38373	0.02700	0.09702	-
ST8F	Mix	1	184	24646	2225	26871	0.02739	0.08280	-
ST30A	Mix	1	222	31138	1356	32494	0.02733	0.04173	-
ST30B	Mix	1	116	16855	481	17336	0.02677	0.02775	-
ST30C	Mix	1	57	8355	546	8901	0.02562	0.06134	-
ST30D	Mix	1	55	8078	298	8376	0.02627	0.03558	-
ST30E	Mix	1	82	11900	439	12339	0.02658	0.03558	-
ST239A	Mix	2	34	4073	468	4541	0.02995	0.10306	-
ST239B	Mix	2	13	1547	35	1582	0.03287	0.03287	-
ST239C	Mix	2	35	4239	203	4442	0.03152	0.04570	-
ST239D	Mix	2	61	7308	313	7621	0.03202	0.04107	-
ST8A	Mix	2	55	6442	457	6899	0.03189	0.06624	-
ST8B	Mix	2	3	389	28	417	0.02878	0.06715	-
ST8C	Mix	2	132	14610	1542	16152	0.03269	0.09547	-
ST8D	Mix	2	156	17420	1799	19219	0.03247	0.09361	-
ST8E	Mix	2	259	28563	3287	31850	0.03253	0.10320	-
ST8F	Mix	2	184	20513	1655	22168	0.03320	0.07466	-
ST30A	Mix	2	222	25716	1046	26762	0.03318	0.03909	-
ST30B	Mix	2	116	13666	509	14175	0.03273	0.03591	-
ST30C	Mix	2	57	6860	434	7294	0.03126	0.05950	-
ST30D	Mix	2	55	6762	279	7041	0.03125	0.03963	-
ST30E	Mix	2	82	9650	354	10004	0.03279	0.03539	-
ST239A	Mix	3	34	5560	556	6116	0.02224	0.0909	-
ST239B	Mix	3	13	2181	52	2233	0.02329	0.0233	-
ST239C	Mix	3	35	5533	232	5765	0.02428	0.0402	-

Appendices

ST239D	Mix	3	61	9574	468	10042	0.02430	0.0466	-
ST8A	Mix	3	55	8334	557	8891	0.02474	0.0627	-
ST8B	Mix	3	3	458	46	504	0.02381	0.0913	-
ST8C	Mix	3	132	19174	2219	21393	0.02468	0.1037	-
ST8D	Mix	3	156	22241	2384	24625	0.02534	0.0968	-
ST8E	Mix	3	259	36160	4621	40781	0.02540	0.1133	-
ST8F	Mix	3	184	26316	2078	28394	0.02592	0.0732	-
ST30A	Mix	3	222	33631	1457	35088	0.02531	0.0415	-
ST30B	Mix	3	116	18141	597	18738	0.02476	0.0319	-
ST30C	Mix	3	57	9164	531	9695	0.02352	0.0548	-
ST30D	Mix	3	55	8460	360	8820	0.02494	0.0408	-
ST30E	Mix	3	82	12499	471	12970	0.02529	0.0363	-
ST239A	Oxa	1	34	5596	559	6155	0.02210	0.09082	-0.0175
ST239B	Oxa	1	13	2113	0	2113	0.02461	0.02461	-0.1151
ST239C	Oxa	1	35	3028	3136	6164	0.02271	0.50876	1.1062
ST239D	Oxa	1	61	7502	2530	10032	0.02432	0.25219	0.7655
ST8A	Oxa	1	55	9314	0	9314	0.02362	0.02362	-0.4579
ST8B	Oxa	1	3	490	0	490	0.02449	0.02449	-0.4923
ST8C	Oxa	1	132	22102	0	22102	0.02389	0.02389	-0.6318
ST8D	Oxa	1	156	25568	7	25575	0.02440	0.02440	-0.5956
ST8E	Oxa	1	259	42410	134	42544	0.02435	0.02435	-0.6004
ST8F	Oxa	1	184	29551	0	29551	0.02491	0.02491	-0.5217
ST30A	Oxa	1	222	35848	0	35848	0.02477	0.02477	-0.2265
ST30B	Oxa	1	116	18880	0	18880	0.02458	0.02458	-0.0527
ST30C	Oxa	1	57	9390	628	10018	0.02276	0.06269	0.0095
ST30D	Oxa	1	55	9384	5	9389	0.02343	0.02343	-0.1814
ST30E	Oxa	1	82	12876	555	13431	0.02442	0.04132	0.0650
ST239A	Oxa	2	34	5162	544	5706	0.02383	0.09534	-0.0338
ST239B	Oxa	2	13	2007	0	2007	0.02591	0.02591	-0.1033
ST239C	Oxa	2	35	3170	2456	5626	0.02488	0.43654	0.9801
ST239D	Oxa	2	61	7129	2507	9636	0.02532	0.26017	0.8017
ST8A	Oxa	2	55	8834	0	8834	0.02490	0.02490	-0.4249
ST8B	Oxa	2	3	417	0	417	0.02878	0.02878	-0.3680
ST8C	Oxa	2	132	20612	0	20612	0.02562	0.02562	-0.5713
ST8D	Oxa	2	156	23820	9	23829	0.02619	0.02619	-0.5532
ST8E	Oxa	2	259	36678	3074	39752	0.02606	0.07733	-0.1253
ST8F	Oxa	2	184	27445	0	27445	0.02682	0.02682	-0.4446
ST30A	Oxa	2	222	33831	4	33835	0.02625	0.02625	-0.1729
ST30B	Oxa	2	116	17617	0	17617	0.02634	0.02634	-0.1346
ST30C	Oxa	2	57	8855	469	9324	0.02445	0.05030	-0.0730
ST30D	Oxa	2	55	9111	35	9146	0.02405	0.02405	-0.2169
ST30E	Oxa	2	82	12342	328	12670	0.02589	0.02589	-0.1358
ST239A	Oxa	3	34	4867	466	5333	0.02550	0.0874	-0.0171
ST239B	Oxa	3	13	1962	0	1962	0.02650	0.0265	0.0559
ST239C	Oxa	3	35	2817	2506	5323	0.02630	0.4708	1.0686
ST239D	Oxa	3	61	7051	2085	9136	0.02671	0.2282	0.6899
ST8A	Oxa	3	55	8089	0	8089	0.02720	0.0272	-0.3627
ST8B	Oxa	3	3	399	0	399	0.03008	0.0301	-0.4819
ST8C	Oxa	3	132	19177	0	19177	0.02753	0.0275	-0.5765
ST8D	Oxa	3	156	22055	9	22064	0.02828	0.0283	-0.5341
ST8E	Oxa	3	259	34248	3265	37513	0.02762	0.0870	-0.1147
ST8F	Oxa	3	184	25644	0	25644	0.02870	0.0287	-0.4066
ST30A	Oxa	3	222	31425	0	31425	0.02826	0.0283	-0.1663
ST30B	Oxa	3	116	16450	4	16454	0.02820	0.0282	-0.0535
ST30C	Oxa	3	57	8268	429	8697	0.02622	0.0493	-0.0459
ST30D	Oxa	3	55	8266	18	8284	0.02656	0.0266	-0.1858

ST30E	Oxa	3	82	11440	274	11714	0.02800	0.0280	-0.1128
ST239A	Norf	1	34	5181	370	5551	0.02450	0.06665	-0.1519
ST239B	Norf	1	13	2075	0	2075	0.02506	0.02506	-0.1073
ST239C	Norf	1	35	4806	353	5159	0.02714	0.06842	0.2349
ST239D	Norf	1	61	8808	541	9349	0.02610	0.05787	0.1263
ST8A	Norf	1	55	8173	0	8173	0.02692	0.02692	-0.4011
ST8B	Norf	1	3	443	0	443	0.02709	0.02709	-0.4485
ST8C	Norf	1	132	14742	4616	19358	0.02728	0.23845	0.3674
ST8D	Norf	1	156	18705	3930	22635	0.02757	0.17362	0.2567
ST8E	Norf	1	259	30402	7051	37453	0.02766	0.18826	0.2879
ST8F	Norf	1	184	25927	330	26257	0.02803	0.02803	-0.4704
ST30A	Norf	1	222	31266	722	31988	0.02776	0.02776	-0.1770
ST30B	Norf	1	116	16813	35	16848	0.02754	0.02754	-0.0033
ST30C	Norf	1	57	8190	467	8657	0.02634	0.05394	-0.0558
ST30D	Norf	1	55	8146	5	8151	0.02699	0.02699	-0.1200
ST30E	Norf	1	82	11334	489	11823	0.02774	0.04136	0.0654
ST239A	Norf	2	34	4958	424	5382	0.02527	0.07878	-0.1167
ST239B	Norf	2	13	2003	0	2003	0.02596	0.02596	-0.1025
ST239C	Norf	2	35	5095	402	5497	0.02547	0.07313	0.2042
ST239D	Norf	2	61	8470	682	9152	0.02666	0.07452	0.2587
ST8A	Norf	2	55	8443	0	8443	0.02606	0.02606	-0.4051
ST8B	Norf	2	3	460	0	460	0.02609	0.02609	-0.4106
ST8C	Norf	2	132	15181	4897	20078	0.02630	0.24390	0.4073
ST8D	Norf	2	156	18161	4659	22820	0.02734	0.20416	0.3386
ST8E	Norf	2	259	33851	4593	38444	0.02695	0.11947	0.0636
ST8F	Norf	2	184	26386	175	26561	0.02771	0.02771	-0.4305
ST30A	Norf	2	222	31445	1274	32719	0.02714	0.03894	-0.0017
ST30B	Norf	2	116	16792	28	16820	0.02759	0.02759	-0.1145
ST30C	Norf	2	57	8369	522	8891	0.02564	0.05871	-0.0058
ST30D	Norf	2	55	8286	24	8310	0.02647	0.02647	-0.1753
ST30E	Norf	2	82	11690	397	12087	0.02714	0.03285	-0.0323
ST239A	Norf	3	34	4218	317	4535	0.02999	0.0699	-0.1141
ST239B	Norf	3	13	1609	0	1609	0.03232	0.0323	0.1418
ST239C	Norf	3	35	4050	322	4372	0.03202	0.0737	0.2632
ST239D	Norf	3	61	6908	497	7405	0.03295	0.0671	0.1583
ST8A	Norf	3	55	6825	0	6825	0.03223	0.0322	-0.2894
ST8B	Norf	3	3	319	0	319	0.03762	0.0376	-0.3853
ST8C	Norf	3	132	11118	4824	15942	0.03312	0.3026	0.4651
ST8D	Norf	3	156	14622	3725	18347	0.03401	0.2030	0.3216
ST8E	Norf	3	259	27380	3982	31362	0.03303	0.1270	0.0496
ST8F	Norf	3	184	21079	12	21091	0.03490	0.0349	-0.3217
ST30A	Norf	3	222	25566	333	25899	0.03429	0.0343	-0.0828
ST30B	Norf	3	116	13641	0	13641	0.03402	0.0340	0.0277
ST30C	Norf	3	57	6863	411	7274	0.03134	0.0565	0.0133
ST30D	Norf	3	55	6757	8	6765	0.03252	0.0325	-0.0988
ST30E	Norf	3	82	9292	236	9528	0.03442	0.0344	-0.0233
ST239A	Trim	1	34	5051	540	5591	0.02432	0.0966	0.0093
ST239B	Trim	1	13	1994	0	1994	0.02608	0.0261	-0.0896
ST239C	Trim	1	35	4618	676	5294	0.02645	0.1277	0.5059
ST239D	Trim	1	61	8102	950	9052	0.02696	0.1050	0.3850
ST8A	Trim	1	55	8556	0	8556	0.02571	0.0257	-0.4212
ST8B	Trim	1	3	461	0	461	0.02603	0.0260	-0.4664
ST8C	Trim	1	132	20111	0	20111	0.02625	0.0263	-0.5901
ST8D	Trim	1	156	14713	8260	22973	0.02716	0.3596	0.5729
ST8E	Trim	1	259	27396	10882	38278	0.02707	0.2843	0.4669
ST8F	Trim	1	184	26429	0	26429	0.02785	0.0278	-0.4740

Appendices

ST30A	Trim	1	222	32242	0	32242	0.02754	0.0275	-0.1811
ST30B	Trim	1	116	16757	0	16757	0.02769	0.0277	-0.0008
ST30C	Trim	1	57	8889	0	8889	0.02565	0.0256	-0.3795
ST30D	Trim	1	55	8438	0	8438	0.02607	0.0261	-0.1346
ST30E	Trim	1	82	12162	0	12162	0.02697	0.0270	-0.1198
ST239A	Trim	2	34	4670	581	5251	0.02590	0.1107	0.0311
ST239B	Trim	2	13	1952	0	1952	0.02664	0.0266	-0.0919
ST239C	Trim	2	35	4507	766	5273	0.02655	0.1453	0.5023
ST239D	Trim	2	61	7844	1124	8968	0.02721	0.1253	0.4844
ST8A	Trim	2	55	8128	0	8128	0.02707	0.0271	-0.3882
ST8B	Trim	2	3	428	0	428	0.02804	0.0280	-0.3799
ST8C	Trim	2	132	19659	4	19663	0.02685	0.0269	-0.5501
ST8D	Trim	2	156	12760	9584	22344	0.02793	0.4289	0.6610
ST8E	Trim	2	259	31630	6061	37691	0.02749	0.1608	0.1926
ST8F	Trim	2	184	26160	0	26160	0.02813	0.0281	-0.4244
ST30A	Trim	2	222	31540	0	31540	0.02815	0.0282	-0.1418
ST30B	Trim	2	116	16307	5	16312	0.02845	0.0284	-0.1019
ST30C	Trim	2	57	8664	0	8664	0.02632	0.0263	-0.3546
ST30D	Trim	2	55	8214	4	8218	0.02677	0.0268	-0.1699
ST30E	Trim	2	82	11823	0	11823	0.02774	0.0277	-0.1064
ST239A	Trim	3	34	5091	573	5664	0.02401	0.1012	0.0466
ST239B	Trim	3	13	1995	0	1995	0.02607	0.0261	0.0493
ST239C	Trim	3	35	4755	761	5516	0.02538	0.1380	0.5357
ST239D	Trim	3	61	8418	1201	9619	0.02537	0.1249	0.4282
ST8A	Trim	3	55	8681	0	8681	0.02534	0.0253	-0.3941
ST8B	Trim	3	3	468	0	468	0.02564	0.0256	-0.5522
ST8C	Trim	3	132	20257	4	20261	0.02606	0.0261	-0.5991
ST8D	Trim	3	156	12918	10118	23036	0.02709	0.4392	0.6568
ST8E	Trim	3	259	32868	6767	39635	0.02614	0.1707	0.1780
ST8F	Trim	3	184	27644	0	27644	0.02662	0.0266	-0.4396
ST30A	Trim	3	222	33267	5	33272	0.02669	0.0267	-0.1915
ST30B	Trim	3	116	17528	0	17528	0.02647	0.0265	-0.0805
ST30C	Trim	3	57	9203	0	9203	0.02477	0.0248	-0.3443
ST30D	Trim	3	55	8640	0	8640	0.02546	0.0255	-0.2041
ST30E	Trim	3	82	12345	4	12349	0.02656	0.0266	-0.1350

Table B.1.13. Average variant site coverage for each isolate before and after competition in oxacillin, norfloxacin and trimethoprim/sulfamethoxazole. Mix, before competition; Oxa, 24 hour competition in 4 µg/mL oxacillin; Norf, 24 hour competition in 2 µg/mL norfloxacin; trim, 24 hour competition in 152 µg/mL trimethoprim/sulfamethoxazole.

B.1.14 Average variant site coverage for competitions in TSB, BHI and PS

Isolate ID	Condition	Replicate	No variant sites	No reads supporting non-unique variants	No reads supporting unique variants	Total no reads	Minimum detection limit	Corrected average variant coverage	Log (difference)
ST239A	Mix	1	34	6204	642	6846	0.01987	0.09378	-
ST239B	Mix	1	13	2438	118	2556	0.02034	0.04617	-
ST239C	Mix	1	35	6248	316	6564	0.02133	0.04814	-
ST239D	Mix	1	61	10904	461	11365	0.02147	0.04056	-
ST30A	Mix	1	222	37871	1608	39479	0.02249	0.04073	-
ST30B	Mix	1	116	20142	715	20857	0.02225	0.03428	-
ST30C	Mix	1	57	9952	628	10580	0.02155	0.05936	-

ST30D	Mix	1	55	9832	427	10259	0.02144	0.04162	-
ST30E	Mix	1	82	14280	534	14814	0.02214	0.03605	-
ST8A	Mix	1	55	9479	634	10113	0.02175	0.06269	-
ST8B	Mix	1	3	528	41	569	0.02109	0.07206	-
ST8C	Mix	1	132	21543	2529	24072	0.02193	0.10506	-
ST8D	Mix	1	156	25053	2471	27524	0.02267	0.08978	-
ST8E	Mix	1	259	41174	4444	45618	0.02271	0.09742	-
ST8F	Mix	1	184	28628	2981	31609	0.02328	0.09431	-
ST239A	Mix	2	34	6113	656	6769	0.02009	0.09691	-
ST239B	Mix	2	13	2238	105	2343	0.02219	0.04481	-
ST239C	Mix	2	35	5626	272	5898	0.02374	0.04612	-
ST239D	Mix	2	61	10291	448	10739	0.02272	0.04172	-
ST30A	Mix	2	222	36158	1514	37672	0.02357	0.04019	-
ST30B	Mix	2	116	18937	709	19646	0.02362	0.03609	-
ST30C	Mix	2	57	9585	564	10149	0.02247	0.05557	-
ST30D	Mix	2	55	9212	367	9579	0.02297	0.03831	-
ST30E	Mix	2	82	13068	466	13534	0.02424	0.03443	-
ST8A	Mix	2	55	8643	590	9233	0.02383	0.06390	-
ST8B	Mix	2	3	506	32	538	0.02230	0.05948	-
ST8C	Mix	2	132	20383	2284	22667	0.02329	0.10076	-
ST8D	Mix	2	156	23886	2503	26389	0.02365	0.09485	-
ST8E	Mix	2	259	38444	4670	43114	0.02403	0.10832	-
ST8F	Mix	2	184	27591	2514	30105	0.02445	0.08351	-
ST239A	Mix	3	34	6263	611	6874	0.01978	0.0889	-
ST239B	Mix	3	13	2377	126	2503	0.02078	0.0503	-
ST239C	Mix	3	35	5857	275	6132	0.02283	0.0449	-
ST239D	Mix	3	61	10562	438	11000	0.02218	0.0398	-
ST30A	Mix	3	222	37770	1658	39428	0.02252	0.0421	-
ST30B	Mix	3	116	19625	659	20284	0.02288	0.0325	-
ST30C	Mix	3	57	10341	540	10881	0.02095	0.0496	-
ST30D	Mix	3	55	9915	410	10325	0.02131	0.0397	-
ST30E	Mix	3	82	13560	488	14048	0.02335	0.0347	-
ST8A	Mix	3	55	9126	614	9740	0.02259	0.0630	-
ST8B	Mix	3	3	513	44	557	0.02154	0.0790	-
ST8C	Mix	3	132	21884	2376	24260	0.02176	0.0979	-
ST8D	Mix	3	156	25021	2565	27586	0.02262	0.0930	-
ST8E	Mix	3	259	40608	4409	45017	0.02301	0.0979	-
ST8F	Mix	3	184	29319	2771	32090	0.02294	0.0864	-
ST239A	TSB	1	34	5299	397	5696	0.02388	0.06970	-0.1289
ST239B	TSB	1	13	2049	0	2049	0.02538	0.02538	-0.2599
ST239C	TSB	1	35	5242	267	5509	0.02541	0.04847	0.0030
ST239D	TSB	1	61	9081	219	9300	0.02624	0.02624	-0.1891
ST8A	TSB	1	55	8141	353	8494	0.02590	0.04156	-0.1785
ST8B	TSB	1	3	417	21	438	0.02740	0.04795	-0.1769
ST8C	TSB	1	132	17196	3032	20228	0.02610	0.14989	0.1543
ST8D	TSB	1	156	20388	3204	23592	0.02645	0.13581	0.1798
ST8E	TSB	1	259	36247	2659	38906	0.02663	0.06834	-0.1540
ST8F	TSB	1	184	21750	5309	27059	0.02720	0.19620	0.3181
ST30A	TSB	1	222	30936	1929	32865	0.02702	0.05869	0.1587
ST30B	TSB	1	116	16449	940	17389	0.02668	0.05406	0.1978
ST30C	TSB	1	57	8642	504	9146	0.02493	0.05511	-0.0323
ST30D	TSB	1	55	8559	4	8563	0.02569	0.02569	-0.2095
ST30E	TSB	1	82	12266	27	12293	0.02668	0.02668	-0.1307
ST239A	TSB	2	34	4987	309	5296	0.02568	0.05835	-0.2203
ST239B	TSB	2	13	1995	0	1995	0.02607	0.02607	-0.2352
ST239C	TSB	2	35	4928	251	5179	0.02703	0.04846	0.0215

Appendices

ST239D	TSB	2	61	8619	222	8841	0.02760	0.02760	-0.1794
ST8A	TSB	2	55	7794	380	8174	0.02691	0.04649	-0.1381
ST8B	TSB	2	3	416	21	437	0.02746	0.04805	-0.0927
ST8C	TSB	2	132	16036	3149	19185	0.02752	0.16414	0.2119
ST8D	TSB	2	156	19313	3137	22450	0.02780	0.13973	0.1683
ST8E	TSB	2	259	34715	2972	37687	0.02749	0.07886	-0.1379
ST8F	TSB	2	184	21431	4679	26110	0.02819	0.17920	0.3316
ST30A	TSB	2	222	29332	2079	31411	0.02827	0.06619	0.2167
ST30B	TSB	2	116	15748	775	16523	0.02808	0.04690	0.1138
ST30C	TSB	2	57	8203	393	8596	0.02652	0.04572	-0.0847
ST30D	TSB	2	55	8184	24	8208	0.02680	0.02680	-0.1552
ST30E	TSB	2	82	11457	0	11457	0.02863	0.02863	-0.0801
ST239A	TSB	3	34	6191	422	6613	0.02057	0.0638	-0.1441
ST239B	TSB	3	13	2312	0	2312	0.02249	0.0225	-0.3494
ST239C	TSB	3	35	6038	311	6349	0.02205	0.0490	0.0380
ST239D	TSB	3	61	10918	265	11183	0.02182	0.0237	-0.2251
ST8A	TSB	3	55	9317	407	9724	0.02262	0.0419	-0.1771
ST8B	TSB	3	3	549	37	586	0.02048	0.0631	-0.0976
ST8C	TSB	3	132	19864	3657	23521	0.02245	0.1555	0.2010
ST8D	TSB	3	156	23510	3849	27359	0.02281	0.1407	0.1798
ST8E	TSB	3	259	42226	3241	45467	0.02279	0.0713	-0.1377
ST8F	TSB	3	184	25680	6220	31900	0.02307	0.1950	0.3535
ST30A	TSB	3	222	35168	3021	38189	0.02325	0.0791	0.2739
ST30B	TSB	3	116	19116	850	19966	0.02324	0.0426	0.1175
ST30C	TSB	3	57	10053	442	10495	0.02172	0.0421	-0.0712
ST30D	TSB	3	55	9940	17	9957	0.02210	0.0221	-0.2544
ST30E	TSB	3	82	14429	23	14452	0.02270	0.0227	-0.1843
ST239A	BHI	1	34	6373	457	6830	0.01991	0.06691	-0.1466
ST239B	BHI	1	13	2456	0	2456	0.02117	0.02117	-0.3386
ST239C	BHI	1	35	6246	199	6445	0.02172	0.03088	-0.1928
ST239D	BHI	1	61	11393	71	11464	0.02128	0.02128	-0.2801
ST8A	BHI	1	55	9920	106	10026	0.02194	0.02194	-0.4560
ST8B	BHI	1	3	545	6	551	0.02178	0.02178	-0.5196
ST8C	BHI	1	132	18842	5012	23854	0.02213	0.21011	0.3010
ST8D	BHI	1	156	22255	5293	27548	0.02265	0.19214	0.3304
ST8E	BHI	1	259	42274	3454	45728	0.02266	0.07553	-0.1105
ST8F	BHI	1	184	28590	3419	32009	0.02299	0.10681	0.0541
ST30A	BHI	1	222	33644	5279	38923	0.02281	0.13563	0.5224
ST30B	BHI	1	116	19994	1016	21010	0.02208	0.04836	0.1494
ST30C	BHI	1	57	10431	444	10875	0.02097	0.04083	-0.1625
ST30D	BHI	1	55	10040	0	10040	0.02191	0.02191	-0.2787
ST30E	BHI	1	82	14676	29	14705	0.02231	0.02231	-0.2084
ST239A	BHI	2	34	6087	346	6433	0.02114	0.05379	-0.2557
ST239B	BHI	2	13	2405	0	2405	0.02162	0.02162	-0.3165
ST239C	BHI	2	35	5980	201	6181	0.02265	0.03252	-0.1517
ST239D	BHI	2	61	10779	116	10895	0.02240	0.02240	-0.2701
ST8A	BHI	2	55	9668	66	9734	0.02260	0.02260	-0.4514
ST8B	BHI	2	3	508	0	508	0.02362	0.02362	-0.4011
ST8C	BHI	2	132	18039	4800	22839	0.02312	0.21017	0.3193
ST8D	BHI	2	156	21953	4665	26618	0.02344	0.17526	0.2666
ST8E	BHI	2	259	40293	3654	43947	0.02357	0.08315	-0.1148
ST8F	BHI	2	184	26761	3741	30502	0.02413	0.12265	0.1669
ST30A	BHI	2	222	31616	5342	36958	0.02403	0.14454	0.5559
ST30B	BHI	2	116	19140	980	20120	0.02306	0.04871	0.1302
ST30C	BHI	2	57	9652	395	10047	0.02269	0.03932	-0.1502
ST30D	BHI	2	55	9469	17	9486	0.02319	0.02319	-0.2180

ST30E	BHI	2	82	14019	13	14032	0.02338	0.02338	-0.1681
ST239A	BHI	3	34	5835	354	6189	0.02197	0.0572	-0.1915
ST239B	BHI	3	13	2227	0	2227	0.02335	0.0233	-0.3342
ST239C	BHI	3	35	5676	185	5861	0.02389	0.0316	-0.1526
ST239D	BHI	3	61	10102	56	10158	0.02402	0.0240	-0.2197
ST8A	BHI	3	55	9100	75	9175	0.02398	0.0240	-0.4191
ST8B	BHI	3	3	526	4	530	0.02264	0.0226	-0.5435
ST8C	BHI	3	132	17655	4252	21907	0.02410	0.1941	0.2972
ST8D	BHI	3	156	20684	4606	25290	0.02467	0.1821	0.2918
ST8E	BHI	3	259	38967	3380	42347	0.02446	0.0798	-0.0888
ST8F	BHI	3	184	26199	3207	29406	0.02503	0.1091	0.1013
ST30A	BHI	3	222	29063	6121	35184	0.02524	0.1740	0.6163
ST30B	BHI	3	116	18030	784	18814	0.02466	0.0417	0.1083
ST30C	BHI	3	57	9447	269	9716	0.02347	0.0277	-0.2530
ST30D	BHI	3	55	9262	0	9262	0.02375	0.0238	-0.2222
ST30E	BHI	3	82	13084	0	13084	0.02507	0.0251	-0.1407
ST239A	PS	1	34	5176	597	5773	0.02356	0.1034	0.0424
ST239B	PS	1	13	2250	5	2255	0.02306	0.0231	-0.3007
ST239C	PS	1	35	4940	571	5511	0.02540	0.1036	0.3329
ST239D	PS	1	61	9042	546	9588	0.02545	0.0570	0.1478
ST8A	PS	1	55	8161	526	8687	0.02533	0.0606	-0.0147
ST8B	PS	1	3	493	33	526	0.02281	0.0627	-0.0604
ST8C	PS	1	132	18286	2350	20636	0.02559	0.1139	0.0351
ST8D	PS	1	156	21319	2558	23877	0.02613	0.1071	0.0766
ST8E	PS	1	259	38192	2135	40327	0.02569	0.0529	-0.2652
ST8F	PS	1	184	25661	2350	28011	0.02628	0.0839	-0.0508
ST30A	PS	1	222	32658	1393	34051	0.02608	0.0409	0.0018
ST30B	PS	1	116	16451	1135	17586	0.02638	0.0645	0.2745
ST30C	PS	1	57	8953	377	9330	0.02444	0.0404	-0.1671
ST30D	PS	1	55	8902	115	9017	0.02440	0.0244	-0.2319
ST30E	PS	1	82	12118	414	12532	0.02617	0.0330	-0.0384
ST239A	PS	2	34	5374	584	5958	0.02283	0.0980	0.0049
ST239B	PS	2	13	2178	4	2182	0.02383	0.0238	-0.2748
ST239C	PS	2	35	5104	553	5657	0.02475	0.0978	0.3264
ST239D	PS	2	61	9173	579	9752	0.02502	0.0594	0.1534
ST8A	PS	2	55	8214	598	8812	0.02497	0.0679	0.0264
ST8B	PS	2	3	505	37	542	0.02214	0.0683	0.0600
ST8C	PS	2	132	18983	2631	21614	0.02443	0.1217	0.0820
ST8D	PS	2	156	22310	2466	24776	0.02519	0.0995	0.0208
ST8E	PS	2	259	38388	2738	41126	0.02519	0.0666	-0.2112
ST8F	PS	2	184	26359	2294	28653	0.02569	0.0801	-0.0181
ST30A	PS	2	222	32938	1593	34531	0.02572	0.0461	0.0596
ST30B	PS	2	116	16826	1082	17908	0.02591	0.0604	0.2237
ST30C	PS	2	57	9394	291	9685	0.02354	0.0301	-0.2663
ST30D	PS	2	55	9172	138	9310	0.02363	0.0236	-0.2104
ST30E	PS	2	82	12402	278	12680	0.02587	0.0259	-0.1236
ST239A	PS	3	34	5537	657	6194	0.02196	0.1061	0.0768
ST239B	PS	3	13	2308	0	2308	0.02253	0.0225	-0.3494
ST239C	PS	3	35	5347	615	5962	0.02348	0.1032	0.3614
ST239D	PS	3	61	9630	616	10246	0.02381	0.0601	0.1790
ST8A	PS	3	55	8580	601	9181	0.02396	0.0655	0.0169
ST8B	PS	3	3	491	35	526	0.02281	0.0665	-0.0748
ST8C	PS	3	132	19277	2585	21862	0.02415	0.1182	0.0818
ST8D	PS	3	156	22782	2541	25323	0.02464	0.1003	0.0328
ST8E	PS	3	259	40117	2319	42436	0.02441	0.0547	-0.2528
ST8F	PS	3	184	26304	2781	29085	0.02531	0.0956	0.0439

Appendices

ST30A	PS	3	222	33488	2132	35620	0.02493	0.0599	0.1531
ST30B	PS	3	116	17563	1113	18676	0.02484	0.0596	0.2634
ST30C	PS	3	57	9583	243	9826	0.02320	0.0247	-0.3028
ST30D	PS	3	55	9265	97	9362	0.02350	0.0235	-0.2277
ST30E	PS	3	82	12903	208	13111	0.02502	0.0250	-0.1424

Table B.1.14. Average variant site coverage for each isolate before and after competition in TSB, BHI and PS. Mix, before competition; TSB, 24 hour competition in TSB; BHI, 24 hour competition in BHI; PS, 24 hour competition in PS.

B.1.15 Sum of corrected average variant site coverage

Condition	Replicate	Sum of corrected average SNP coverage
Mix	1	0.93392
Mix	2	0.93255
Mix	3	0.9473
Mix	4	0.96201
Mix	5	0.94497
Mix	6	0.9397
Oxa	1	1.19883
Oxa	2	1.18043
Oxa	3	1.2021
Norf	1	1.07796
Norf	2	1.08434
Norf	3	1.172
Trim	1	1.239
Trim	2	1.2451
Trim	3	1.2338
TSB	1	1.02977
TSB	2	1.03119
TSB	3	1.0351
BHI	1	1.03759
BHI	2	1.04692
BHI	3	1.0401
PS	1	0.9714
PS	2	0.9692
PS	3	0.9854

Table B.1.15. Sum of corrected average variant site coverage. Mix, before competition; Oxa, 24 hour competition in 4 µg/mL oxacillin; Norf, 24 hour competition in 2 µg/mL norfloxacin; trim, 24 hour competition in 152 µg/mL trimethoprim/sulfamethoxazole; TSB, 24 hour competition in TSB; BHI, 24 hour competition in BHI; PS, 24 hour competition in PS.

B.1.16 Average maximum unique gene coverage for competitions in oxacillin, quinolone and trimethoprim/sulfamethoxazole

Isolate ID	Condition	Replicate	Minimum Detection Limit	Corrected average max unique gene coverage	Log(difference)
ST239A	Mix	1	4	120.43	-
ST239B	Mix	1	4	27.14	-
ST239D	Mix	1	4	19	-

ST8C	Mix	1	4	8	-
ST8F	Mix	1	4	7.4	-
ST30A	Mix	1	4	9.25	-
ST30B	Mix	1	4	26.25	-
ST30E	Mix	1	4	22.6	-
ST239A	Mix	2	4	122	-
ST239B	Mix	2	4	26.14	-
ST239D	Mix	2	4	22	-
ST8C	Mix	2	4	17	-
ST8F	Mix	2	4	12.4	-
ST30A	Mix	2	4	9.67	-
ST30B	Mix	2	4	26.25	-
ST30E	Mix	2	4	20.04	-
ST239A	Mix	3	4	123.14	-
ST239B	Mix	3	4	29.14	-
ST239D	Mix	3	4	18.67	-
ST8C	Mix	3	4	7	-
ST8F	Mix	3	4	8.2	-
ST30A	Mix	3	4	8.08	-
ST30B	Mix	3	4	23.75	-
ST30E	Mix	3	4	25.36	-
ST239A	Oxa	1	4	170.43	0.1508
ST239B	Oxa	1	4	4	-0.8315
ST239D	Oxa	1	4	257	1.1312
ST8C	Oxa	1	4	4	-0.3010
ST8F	Oxa	1	4	4	-0.2672
ST30A	Oxa	1	4	15.67	0.2289
ST30B	Oxa	1	4	4	-0.8171
ST30E	Oxa	1	4	4	-0.7520
ST239A	Oxa	2	4	209.29	0.2344
ST239B	Oxa	2	4	4	-0.8152
ST239D	Oxa	2	4	167.33	0.8812
ST8C	Oxa	2	4	4	-0.6284
ST8F	Oxa	2	4	4	-0.4914
ST30A	Oxa	2	4	7.08	-0.1354
ST30B	Oxa	2	4	4	-0.8171
ST30E	Oxa	2	4	4	-0.6998
ST239A	Oxa	3	4	123.71	0.0020
ST239B	Oxa	3	4	4	-0.8624
ST239D	Oxa	3	4	242	1.1127
ST8C	Oxa	3	4	4	-0.2430
ST8F	Oxa	3	4	4	-0.3118
ST30A	Oxa	3	4	15.67	0.2877
ST30B	Oxa	3	4	4	-0.7736
ST30E	Oxa	3	4	4	-0.8021
ST239A	Norf	1	4	93.14	-0.1116
ST239B	Norf	1	4	4	-0.8315
ST239D	Norf	1	4	36.67	0.2856
ST8C	Norf	1	4	5	-0.2041
ST8F	Norf	1	4	4	-0.2672
ST30A	Norf	1	4	26.17	0.4517
ST30B	Norf	1	4	59.75	0.3572
ST30E	Norf	1	4	6	-0.5760
ST239A	Norf	2	4	131.86	0.0338
ST239B	Norf	2	4	4	-0.8152
ST239D	Norf	2	4	55.33	0.4005

Appendices

ST8C	Norf	2	4	13	-0.1165
ST8F	Norf	2	4	4.2	-0.4702
ST30A	Norf	2	4	24.92	0.4111
ST30B	Norf	2	4	76.5	0.4645
ST30E	Norf	2	4	7	-0.4568
ST239A	Norf	3	4	103.43	-0.0758
ST239B	Norf	3	4	4	-0.8624
ST239D	Norf	3	4	45.33	0.3852
ST8C	Norf	3	4	5	-0.1461
ST8F	Norf	3	4	4	-0.3118
ST30A	Norf	3	4	20.33	0.4007
ST30B	Norf	3	4	78	0.5164
ST30E	Norf	3	4	4	-0.8021
ST239A	Trim	1	4	193.43	0.2058
ST239B	Trim	1	4	4	-0.8315
ST239D	Trim	1	4	62	0.5136
ST8C	Trim	1	4	4	-0.3010
ST8F	Trim	1	4	4	-0.2672
ST30A	Trim	1	4	4	-0.3641
ST30B	Trim	1	4	4	-0.8171
ST30E	Trim	1	4	4	-0.7520
ST239A	Trim	2	4	210.57	0.2370
ST239B	Trim	2	4	4	-0.8152
ST239D	Trim	2	4	48	0.3388
ST8C	Trim	2	4	4	-0.6284
ST8F	Trim	2	4	4	-0.4914
ST30A	Trim	2	4	4	-0.3834
ST30B	Trim	2	4	4	-0.8171
ST30E	Trim	2	4	4	-0.6998
ST239A	Trim	3	4	212.14	0.2362
ST239B	Trim	3	4	4	-0.8624
ST239D	Trim	3	4	115.67	0.7921
ST8C	Trim	3	4	4	-0.2430
ST8F	Trim	3	4	4	-0.3118
ST30A	Trim	3	4	4	-0.3054
ST30B	Trim	3	4	4	-0.7736
ST30E	Trim	3	4	4	-0.8021

Table B.1.16. Unique gene coverage before and after competition in oxacillin, quinolone and trimethoprim/sulfamethoxazole. Mix, before competition; Oxa, 24 hour competition in 4 µg/mL oxacillin; Norf, 24 hour competition in 2 µg/mL norfloxacin; trim, 24 hour competition in 152 µg/mL trimethoprim/sulfamethoxazole.

B.1.17 Average maximum unique gene coverage for competitions in TSB, BHI and PS

Isolate ID	Condition	Replicate	Minimum Detection Limit	Corrected average max unique gene coverage	Log(difference)
ST239A	Mix	1	4	157.86	-
ST239B	Mix	1	4	37.71	-
ST239D	Mix	1	4	28.33	-
ST8C	Mix	1	4	24.75	-
ST8F	Mix	1	4	26.32	-
ST30A	Mix	1	4	8	-
ST30B	Mix	1	4	7.8	-

ST30E	Mix	1	4	18.42	-
ST239A	Mix	2	4	183.29	-
ST239B	Mix	2	4	38.43	-
ST239D	Mix	2	4	29.67	-
ST8C	Mix	2	4	19.75	-
ST8F	Mix	2	4	20.56	-
ST30A	Mix	2	4	8	-
ST30B	Mix	2	4	8.4	-
ST30E	Mix	2	4	16.83	-
ST239A	Mix	3	4	212.43	-
ST239B	Mix	3	4	52.86	-
ST239D	Mix	3	4	34.67	-
ST8C	Mix	3	4	17.25	-
ST8F	Mix	3	4	20.48	-
ST30A	Mix	3	4	7	-
ST30B	Mix	3	4	7.6	-
ST30E	Mix	3	4	16.58	-
ST239A	TSB	1	4	98	-0.2070
ST239B	TSB	1	4	4	-0.9744
ST239D	TSB	1	4	20	-0.1512
ST8C	TSB	1	4	42.75	0.2374
ST8F	TSB	1	4	56.04	0.3282
ST30A	TSB	1	4	20	0.3979
ST30B	TSB	1	4	16.2	0.3174
ST30E	TSB	1	4	4	-0.6632
ST239A	TSB	2	4	88.29	-0.3172
ST239B	TSB	2	4	4	-0.9826
ST239D	TSB	2	4	20	-0.1713
ST8C	TSB	2	4	44.25	0.3503
ST8F	TSB	2	4	47.92	0.3675
ST30A	TSB	2	4	18	0.3522
ST30B	TSB	2	4	14.8	0.2460
ST30E	TSB	2	4	4	-0.6240
ST239A	TSB	3	4	118	-0.2553
ST239B	TSB	3	4	4	-1.1211
ST239D	TSB	3	4	22.33	-0.1911
ST8C	TSB	3	4	52.75	0.4854
ST8F	TSB	3	4	59.28	0.4616
ST30A	TSB	3	4	23	0.5166
ST30B	TSB	3	4	19.6	0.4114
ST30E	TSB	3	4	4.58	-0.5587
ST239A	BHI	1	4	128.14	-0.0906
ST239B	BHI	1	4	4	-0.9744
ST239D	BHI	1	4	9.67	-0.4668
ST8C	BHI	1	4	73.75	0.4742
ST8F	BHI	1	4	37.76	0.1567
ST30A	BHI	1	4	24	0.4771
ST30B	BHI	1	4	11.6	0.1724
ST30E	BHI	1	4	4	-0.6632
ST239A	BHI	2	4	147.43	-0.0946
ST239B	BHI	2	4	4	-0.9826
ST239D	BHI	2	4	15	-0.2962
ST8C	BHI	2	4	80.75	0.6116
ST8F	BHI	2	4	45.92	0.3490
ST30A	BHI	2	4	22	0.4393
ST30B	BHI	2	4	14.6	0.2401

Appendices

ST30E	BHI	2	4	4	-0.6240
ST239A	BHI	3	4	151.86	-0.1458
ST239B	BHI	3	4	4	-1.1211
ST239D	BHI	3	4	20.67	-0.2246
ST8C	BHI	3	4	93	0.7317
ST8F	BHI	3	4	43.88	0.3309
ST30A	BHI	3	4	39	0.7460
ST30B	BHI	3	4	21.8	0.4576
ST30E	BHI	3	4	4	-0.6175
ST239A	PS	1	4	207.14	0.1180
ST239B	PS	1	4	4.57	-0.9165
ST239D	PS	1	4	119	0.6233
ST8C	PS	1	4	76	0.4872
ST8F	PS	1	4	51.32	0.2900
ST30A	PS	1	4	17	0.3274
ST30B	PS	1	4	17.4	0.3485
ST30E	PS	1	4	21.42	0.0655
ST239A	PS	2	4	224.71	0.0885
ST239B	PS	2	4	5.71	-0.8280
ST239D	PS	2	4	128	0.6349
ST8C	PS	2	4	89	0.6538
ST8F	PS	2	4	53.72	0.4171
ST30A	PS	2	4	14	0.2430
ST30B	PS	2	4	16	0.2798
ST30E	PS	2	4	20.5	0.0857
ST239A	PS	3	4	229.29	0.0332
ST239B	PS	3	4	4.86	-1.0365
ST239D	PS	3	4	111	0.5054
ST8C	PS	3	4	68.75	0.6005
ST8F	PS	3	4	57.16	0.4458
ST30A	PS	3	4	18	0.4102
ST30B	PS	3	4	14.8	0.2894
ST30E	PS	3	4	14	-0.0735

Appendix B.1.17. Unique gene coverage before and after competition in TBS, BHI and PS. Mix, before competition; TSB, 24 hour competition in TSB; BHI, 24 hour competition in BHI; PS, 24 hour competition in PS.

B.2 Supplementary references

B.2.1 List of references

Algeria	Coppens <i>et al.</i> , 2018	Singapore
Bekkhoucha <i>et al.</i> , 2008	D'Souza <i>et al.</i> , 2010	Hon <i>et al.</i> , 2013
Ramdani-Bouguessa <i>et al.</i> , 2006	Murugan and Neetu, 2016	Hsu <i>et al.</i> , 2005a
Argentina	Indonesia	Hsu <i>et al.</i> , 2007b
Aires de Sousa <i>et al.</i> , 2001	Chongtrakool <i>et al.</i> , 2006	Hsu <i>et al.</i> , 2015
Coimbra <i>et al.</i> , 2000	Iran	Teo <i>et al.</i> , 2012
Australia	Goudarzi <i>et al.</i> , 2017	South Africa
Abbott <i>et al.</i> , 2015	Shahsavani <i>et al.</i> , 2011	Essa <i>et al.</i> , 2009
Baines <i>et al.</i> , 2015	Ireland	Jansen van Rensburg <i>et al.</i> , 2011
Coombs <i>et al.</i> , 2013	Humphreys <i>et al.</i> , 1990	Moodley <i>et al.</i> , 2010
Pavillard <i>et al.</i> , 1982	Italy	Oosthuysen <i>et al.</i> , 2014
Brazil	Campanile <i>et al.</i> , 2001	Shittu <i>et al.</i> , 2009

Aires de Sousa <i>et al.</i> , 2001	Kenya	South Korea
Amaral <i>et al.</i> , 2005	Aiken <i>et al.</i> , 2014	Ko <i>et al.</i> , 2005
Botelho <i>et al.</i> , 2016	Kuwait	Peck <i>et al.</i> , 2009
Costa <i>et al.</i> , 2013	Boswihi <i>et al.</i> , 2016	Spain
Soares <i>et al.</i> , 2001	Laos	Cuevas <i>et al.</i> , 2007
Teixeira <i>et al.</i> , 1995	Yeap <i>et al.</i> , 2017	Potel <i>et al.</i> , 2007
Vivoni <i>et al.</i> , 2006	Malaysia	Sri Lanka
Canada	Ghaznavi-Rad <i>et al.</i> , 2010	Chongtrakool <i>et al.</i> , 2006
Simor <i>et al.</i> , 2002	Lee <i>et al.</i> , 2014	Taiwan
Chile	Neela <i>et al.</i> , 2009	Chen <i>et al.</i> , 2013
Aires de Sousa <i>et al.</i> , 2001	Malta	Wang <i>et al.</i> , 2012
China	Scicluna <i>et al.</i> , 2009	Thailand
Aires de Sousa <i>et al.</i> , 2003b	New Zealand	Chongtrakool <i>et al.</i> , 2006
Chongtrakool <i>et al.</i> , 2006	Baines <i>et al.</i> , 2015	Feil <i>et al.</i> , 2008
Dai <i>et al.</i> , 2019	Howden <i>et al.</i> , 2010	Tong <i>et al.</i> , 2014
Li <i>et al.</i> , 2018	Niger	Trinidad and Tobago
Li <i>et al.</i> , 2011	Breurec <i>et al.</i> , 2011	Akpaka <i>et al.</i> , 2007
Shang <i>et al.</i> , 2016	Nigeria	Monecke <i>et al.</i> , 2011
Wang <i>et al.</i> , 2015	Raji <i>et al.</i> , 2013	Monecke <i>et al.</i> , 2014b
Xu <i>et al.</i> , 2009	Pakistan	Tunisia
Zhang <i>et al.</i> , 2013	Jamil <i>et al.</i> , 2017	Mariem <i>et al.</i> , 2013
Croatia	Zafar <i>et al.</i> , 2010	Turkey
Budimir <i>et al.</i> , 2010	Paraguay	Alp <i>et al.</i> , 2009
Czechia	Mayor <i>et al.</i> , 2007	Tekeli <i>et al.</i> , 2016
Melter <i>et al.</i> , 1999	Portugal	UK
Ecudaor	Aires de Sousa <i>et al.</i> , 1998	Cookson and Phillips 1988
Zurita <i>et al.</i> , 2016	Amorim <i>et al.</i> , 2002	Cox <i>et al.</i> , 1995
Hungary	Amorim <i>et al.</i> , 2007	Edgeworth <i>et al.</i> , 2007
Conceição <i>et al.</i> , 2007	Romania	Kerr <i>et al.</i> , 1990
de Lencastre <i>et al.</i> , 1997	Chen <i>et al.</i> , 2010	Uruguay
Germany	Monecke <i>et al.</i> , 2014a	Aires de Sousa <i>et al.</i> , 2001
Albrecht <i>et al.</i> , 2011	Russia	USA
Ghana	Baranovich <i>et al.</i> , 2010	Wang <i>et al.</i> , 2012
Egyir <i>et al.</i> , 2014	Iwao <i>et al.</i> , 2012	Vietnam
Greece	Yamamoto <i>et al.</i> , 2012	Chongtrakool <i>et al.</i> , 2006
Aires de Sousa <i>et al.</i> , 2003a	Saudi Arabia	ST239 Reviews
India	Chongtrakool <i>et al.</i> , 2006	Abdulgader, S.M., <i>et al.</i> , 2015
Abimanyu <i>et al.</i> , 2012	Cirlan <i>et al.</i> , 2005	Harris <i>et al.</i> , 2010.
Bakthavatchalam <i>et al.</i> , 2019	Monecke <i>et al.</i> , 2012	Monecke <i>et al.</i> , 2018
Chongtrakool <i>et al.</i> , 2006	Senok <i>et al.</i> , 2016	Smyth <i>et al.</i> , 2010

Table B.2.1. List of references used in Chapter 2, Table 2.4a – 2.4b.

B.3 MLST profiles

B.3.1 MLST profiles of the *Staphoplia* isolates that contain SCCmec-III

ST	No. Isolates	MLST allele							No MLST genes different to ST239
		<i>arcC</i>	<i>aroE</i>	<i>glpF</i>	<i>gmk</i>	<i>pta</i>	<i>tpi</i>	<i>yqiL</i>	
239	2979	2	3	1	1	4	4	3	0
241	50	2	3	1	1	4	4	30	1
NA	47	-	-	-	-	-	-	-	-
757	36	2	3	1	1	4	363	3	1
3620	7	2	3	490	1	4	4	3	1
368	3	2	3	1	1	4	65	3	1
1329	2	162	3	1	1	4	4	3	1
1463	2	2	3	1	1	4	4	422	1
1971	2	218	3	1	1	4	4	3	1
3622	2	2	3	1	263	4	4	3	1
5	1	1	4	1	4	12	1	10	6
8	1	3	3	1	1	4	4	3	1
22	1	7	6	1	5	8	8	6	6
101	1	3	1	14	15	11	19	3	5
174	1	2	3	378	1	4	4	3	1
176	1	2	3	1	71	4	4	3	1
242	1	2	3	1	1	4	4	420	1
250	1	3	3	1	1	4	4	16	2
254	1	3	32	1	1	4	4	3	2
343	1	2	85	1	1	4	4	3	1
372	1	2	3	1	1	4	73	30	2
1225	1	2	3	1	1	4	148	3	1
1283	1	2	3	1	137	4	4	3	1
1441	1	2	3	443	1	4	4	3	1
1531	1	2	3	1	1	4	4	424	1
1545	1	2	3	1	150	4	4	3	1
1849	1	2	491	1	1	4	4	3	1
2223	1	2	3	1	1	12	4	3	1
2246	1	2	323	1	1	4	4	3	1
2586	1	2	3	1	1	282	4	3	1
3349	1	376	3	1	1	4	4	3	1
4359	1	2	3	1	1	4	4	76	1

Table B.3.1. MLST profiles of the *Staphoplia* isolates that contain SCCmec-III.

B.4 Gene presence and function

B.4.1 SCCmec type identification

B.4.1.1 ST239 SCCmec type identification

	PCR product	PCR product size (bp)	SCCmec type	ST239A	ST239B	ST239C	ST239D
Master mix 1	arcCA/B	596	MSSA or MRSA	/	/	/	/
	SCCmec124A/B	342	I, II, IV				
	SCCmec23A/B	209	II, III	/	/	/	/
	SCCmec3C/D	414	III	/	/	/	/
Master mix 2	mecAA/B	200	I, II, III, IV, V	/	/	/	/
	SCCmec1A/B	495	I				
	SCCmec2A/B	284	II				
	SCCmec3A/B	243	III	/	/	/	/
SCCmec type	by PCR			III	III	III	III
	by DNA seq			III	III	III	III

Table B.4.1.1. SCCmec type identification table for ST239 isolates. The presence of a PCR product is indicated by “/”.

B.4.1.2 ST8 SCCmec type identification

	PCR product	PCR product size (bp)	SCCmec type	ST8A	ST8B	ST8C	ST8D	ST8E	ST8F
Master mix 1	arcCA/B	596	<i>S. aureus</i>	/	/	/	/	/	/
	SCCmec124A/B	342	I, II, IV			/	/	/	
	SCCmec23A/B	209	II, III						
	SCCmec3C/D	414	III						
Master mix 2	mecAA/B	200	MRSA			/	/	/	
	SCCmec1A/B	495	I						
	SCCmec2A/B	284	II						
	SCCmec3A/B	243	III						
SCCmec type	by PCR			MSSA	MSSA	IV	IV	IV	MSSA
	by DNA seq			MSSA	MSSA	IV	IV	IV	MSSA

Table B.4.1.2. SCCmec type identification table for ST8 isolates. The presence of a PCR product is indicated by “/”.

B.4.1.3 ST30 SCCmec type identification

	PCR product	PCR product size (bp)	SCCmec type	ST30A	ST30B	ST30C	ST30D	ST30E
Master mix 1	arcCA/B	596	MSSA or MRSA	/	/	/	/	/
	SCCmec124A/B	342	I, II, IV	/		/	/	/
	SCCmec23A/B	209	II, III			/	/	/
	SCCmec3C/D	414	III					
Master mix 2	mecAA/B	200	I, II, III, IV, V	/		/	/	/
	SCCmec1A/B	495	I					
	SCCmec2A/B	284	II					
	SCCmec3A/B	243	III					
SCCmec type	by PCR			IV	MSSA	II	II	II
	by DNA seq			IV	MSSA	II	II	II

Table B.4.1.3. SCCmec type identification table for ST30 isolates. The presence of a PCR product is indicated by “/”.

B.4.2 Gene presence in the SCCmec-III element of the ST239 isolates

Gene	Gene function	Number of copies			
		ST239A	ST239B	ST239C	ST239D
<i>xerC</i>	Tyrosine recombinase	2	2	3	1
<i>hin</i>	DNA-invertase	2	2	2	1
<i>ant1</i>	Streptomycin 3"-adenyltransferase	2	2	1	1
<i>pspE</i>	Thiosulfate sulfurtransferase	1	1	2	1
<i>cadA</i>	Cadmium-transporting ATPase	1	1	1	1
<i>cadC</i>	Cadmium resistance transcriptional regulatory protein	1	1	1	1
<i>ermA1</i>	rRNA adenine N-6-methyltransferase	1	1	1	1
<i>garB</i>	Glutathione amide reductase	1	1	1	1
<i>hsdR</i>	Type-1 restriction enzyme R protein	1	1	1	1
<i>mcrB</i>	5-methylcytosine-specific restriction enzyme B	1	1	1	1
<i>mecI</i>	Methicillin resistance regulatory protein	1	1	1	1
<i>mecR1</i>	Methicillin resistance protein	1	1	1	1
<i>merA</i>	Mercuric reductase	1	1	1	1
<i>merB</i>	Alkylmercury lyase	1	1	1	1
<i>merR1</i>	Mercuric resistance operon regulatory protein	1	1	1	1
<i>nagC</i>	N-acetylglucosamine repressor	1	1	1	1
<i>paaZ</i>	Bifunctional protein PaaZ	1	1	1	1
<i>pbp</i>	Beta-lactam-inducible penicillin-binding protein	1	1	1	1
<i>tnpR</i>	Transposon gamma-delta resolvase	1	1	1	1
<i>ugpQ</i>	Glycerophosphodiester phosphodiesterase, cytoplasmic	1	1	1	1
<i>ycgJ</i>	putative methyltransferase	1	1	1	1
<i>dsrE2</i>	Sulfur carrier protein	1	1		1
<i>gloC</i>	Hydroxyacylglutathione hydrolase	1	1		1
<i>ricR</i>	Copper-sensing transcriptional repressor	1	1		1
KMGABMOG_00101	ACP synthase	1	1		1
KMGABMOG_00102	ADP-ribosyltransferase	1	1		1
CIPDHBFC_00083	cadmium resistance protein CadD	1	1		1

KMGABMOG_00051	DNA recombinase	1	1		1
CIPDHBFC_00049	HTH domain-containing protein	1	1		1
CIPDHBFC_00071	hydroxymethylglutaryl-CoA synthase family protein	1	1		1
hsrA	Putative transport protein	1		1	1
rlmH	Ribosomal RNA large subunit methyltransferase H	1		1	1
CIPDHBFC_66	putative HTH-type transcriptional regulator			1	1
stoA	Sporulation thiol-disulfide oxidoreductase A precursor			1	1
mcrC	Methyl-coenzyme M reductase I operon protein C	1	1		
KMGABMOG_00068	Mercury transport protein	1	1		
KMGABMOG_00069	mercury transporter	1	1		
repN	Replication initiation protein	1			1
KMGABMOG_00044	DNA primase	1			
CEJCKJAB_00020	transposition regulatory protein TnpC_3		1		
PROKKA_00056	Integrase core domain protein			4	
csoR	Copper-sensing transcriptional repressor			1	
pksB	putative polyketide biosynthesis zinc-dependent hydrolase			1	
PROKKA_00103	ADP-ribosyltransferase exoenzyme			1	
PROKKA_00086	Cadmium resistance transporter			1	
PROKKA_00060	Cytochrome C biogenesis protein transmembrane region			1	
PROKKA_00031	DNA polymerase family A			1	
PROKKA_00070	Plasmid recombination enzyme			1	
PROKKA_00072	Replication initiation factor			1	
PROKKA_00085	Sulfite exporter TauE/SafE			1	
PROKKA_00036	Transposase IS200 like protein			1	
mggB	Mannosylglucosyl-3-phosphoglycerate phosphatase				1
pre	Plasmid recombination enzyme				1
CIPDHBFC_00060	cytochrome C biosynthesis protein				1

Table B.4.2. Gene presence in the SCC*mec*-III element of the ST239 isolates.

B.4.3 Gene function in the regions of recombination

B.4.3.1 Gene function within the first recombination region (SCC*mec*-III element)

Gene	Start position	End position	Gene function	Gene	Start position	End position	Gene function
<i>tnpA1</i>	45122	46207	transposase A 1	<i>repC</i>	72454	72537	replication initiation protein C (pseudogene)
<i>tnpB1</i>	46204	48096	transposase B 1	<i>repC</i>	72548	73408	replication initiation protein C (pseudogene)
<i>tnpC1</i>	48103	48480	transposase C 1	SATW20_00690	73497	74171	IS431 transposase
<i>ant1</i>	48631	49413	streptomycin 3"-adenylyltransferase 1	SATW20_00700	74429	74596	HMG-CoA synthase (partial)
<i>ermA1</i>	49539	50270	rRNA adenine N-6-methyltransferase ErmA1	SATW20_00710	74694	74903	putative membrane protein
SATW20_00440	50779	51441	SAM dependent methyltransferase	<i>ugpQ</i>	75513	76256	glycerophosphoryl diester phosphodiesterase
SATW20_00450	51989	52861	hypothetical protein	SATW20_00730	76353	76781	conserved hypothetical protein
SATW20_0	52909	53208	putative membrane	<i>mecA</i>	76827	78833	penicillin-binding protein 2

Appendices

0460			protein				prime
SATW20_00470	53224	53712	putative membrane protein	<i>mecR1</i>	79035	80690	methicillin resistance protein MecR1
SATW20_00480	53847	54374	putative membrane protein	<i>mecI</i>	80690	80890	methicillin resistance regulatory protein MecI (pseudogene)
SATW20_00490	54437	54853	hypothetical protein	<i>mecI</i>	80894	81058	methicillin resistance regulatory protein MecI (pseudogene)
SATW20_00500	55205	55306	hypothetical protein (fragment)	SATW20_00761	81146	81214	phenol-soluble modulins-mec
SATW20_00510	55320	55541	hypothetical protein	SATW20_00770	81534	82682	MarR family regulatory protein
SATW20_00520	55691	56257	putative exported protein	SATW20_00780	82796	83191	metallo-beta-lactamase superfamily protein (pseudogene)
SATW20_00530	56337	58550	type I restriction-modification system endonuclease (fragment)	SATW20_00780	83195	84130	metallo-beta-lactamase superfamily protein (pseudogene)
SATW20_00540	58612	59286	IS431 transposase	SATW20_00790	84162	85226	hypothetical protein
SATW20_00550	59569	60915	NADH-binding protein	SATW20_00800	85362	85622	conserved hypothetical protein
<i>merR</i>	61215	61622	mercuric resistance operon regulatory protein	SATW20_00810	85622	86266	putative membrane protein (fragment)
SATW20_00570	61639	61992	putative exported protein	SATW20_00820	86534	87193	cadmium resistance transporter
SATW20_00580	61989	62669	putative membrane protein	<i>cadA</i>	87232	89646	probable cadmium-transporting ATPase
<i>merT</i>	62757	63143	mercuric transport protein	<i>cadC</i>	89639	90004	putative cadmium efflux system accessory protein
<i>merA</i>	63201	64844	mercuric reductase	<i>tnpC</i>	90190	90567	transposase for transposon Tn554
<i>merB</i>	64926	65576	alkylmercury lyase	<i>tnpB</i>	90574	92403	transposase B from transposon Tn554
SATW20_00620	65771	66319	conserved hypothetical protein	SATW20_00870	92661	92984	DNA repair protein RadC (fragment)
SATW20_00630	66405	67079	IS431 transposase	SATW20_00880	92977	93183	conserved hypothetical protein
SATW20_00631	67512	67844	putative membrane protein	SATW20_00890	93185	93706	conserved hypothetical protein
SATW20_00640	67873	68115	putative DNA-binding protein (fragment)	SATW20_00900	93725	94036	conserved hypothetical protein
SATW20_00650	68247	68921	IS431 transposase	SATW20_00901	94121	94381	hypothetical protein
<i>tet</i>	69118	70497	tetracycline resistance protein	<i>ccrB</i>	94942	96570	site-specific recombinase CcrB
SATW20_00670	70686	71927	plasmid recombination enzyme				

Table B.4.3.1. Genes within the first recombination region correspond to the SCC*mec*-III element.

B.4.3.2 Gene function within the second recombination region (φSa1 phage)

Gene	Start position	End position	Gene function	Gene	Start position	End position	Gene function
<i>geh</i>	374471	375775	lipase precursor (pseudogene)	SATW20_03570	392401	392574	phage transcriptional activator RinB
<i>geh</i>	375779	376405	lipase precursor (pseudogene)	SATW20_03580	392575	392940	phage protein
SATW20_03260	376499	377704	phage integrase	<i>geh</i>	392941	393087	phage protein
SATW20_03270	377974	378906	phage protein	SATW20_03600	393111	393551	phage transcriptional activator RinA
SATW20_03280	378938	379663	phage protein	SATW20_03610	393862	394356	phage protein
SATW20_03290	379691	380365	phage protein	SATW20_03620	394349	395557	phage terminase
SATW20_03300	380382	380714	phage DNA-binding protein	SATW20_03630	395571	396989	phage portal protein
SATW20_03310	380978	381172	phage DNA-binding protein	SATW20_03640	396928	397911	phage minor head protein
SATW20_03320	381172	381939	phage protein	SATW20_03650	397913	398119	phage protein
SATW20_03330	381940	382164	phage protein	SATW20_03660	398224	398820	phage protein
SATW20_03340	382204	382422	phage protein	SATW20_03670	398841	399665	phage protein
SATW20_03350	382452	382715	putative DNA-binding protein	SATW20_03680	399682	400008	phage protein
SATW20_03360	382765	382956	phage protein	SATW20_03690	400008	400322	phage protein
SATW20_03370	382980	383240	phage protein	SATW20_03700	400315	400650	phage protein
SATW20_03380	383234	383512	phage protein	SATW20_03710	400637	401050	phage protein
SATW20_03390	383521	385464	phage protein	SATW20_03720	401063	401500	phage protein
SATW20_03400	385466	386386	phage protein	SATW20_03730	401487	402047	phage protein
SATW20_03410	386599	387084	phage protein	SATW20_03740	402109	402603	phage protein
SATW20_03420	387085	387555	single stranded DNA-binding protein	SATW20_03750	402624	402965	phage protein
SATW20_03430	387585	388478	putative phage regulatory protein	SATW20_03760	402968	405937	phage tape measure protein
SATW20_03440	388485	388703	phage protein	SATW20_03770	405952	406887	phage protein
SATW20_03450	388712	389116	putative phage endodeoxyribonuclease protein	SATW20_03780	406898	408781	phage protein
SATW20_03460	389129	389419	phage protein	SATW20_03790	408794	410692	phage protein
SATW20_03470	389501	389758	phage protein	SATW20_03800	410692	412515	phage protein
SATW20_03480	389755	390003	phage protein	SATW20_03810	412515	412892	phage protein
SATW20_03490	390016	390417	phage protein	SATW20_03820	412893	413075	hypothetical phage protein

Appendices

SATW20_03500	390414	390761	phage protein	SATW20_03830	413116	413415	hypothetical phage protein
SATW20_03510	390758	391066	phage protein	SATW20_03840	413552	415426	putative cell wall hydrolase
SATW20_03520	391059	391307	phage protein	SATW20_03850	415439	416677	hypothetical phage protein
SATW20_03530	391300	391836	phage 2-deoxyuridine 5-triphosphate nucleotidohydrolase protein	SATW20_03860	416682	417077	hypothetical phage protein
SATW20_03540	391873	392073	phage protein	<i>geh</i>	417133	417570	holin
SATW20_03550	392060	392179	phage protein (fragment)	<i>lytA</i>	417551	418996	autolysin
SATW20_03560	392172	392408	phage protein	<i>geh</i>	419746	419886	lipase precursor (pseudogene)

Table B.4.3.2. Genes within the second recombination region correspond to the ϕ Sa1 phage.

B.4.3.3 Gene function within the second recombination region (Tn554 transposon)

Gene	Start position	End position	Gene function	Gene	Start position	End position	Gene function
<i>radC</i>	1775045	1775191	DNA repair protein RadC (pseudogene)	<i>tnpC2</i>	1780660	1781745	transposase C 2
SATW20_16470	1775426	1776088	SAM dependent methyltransferase	<i>radC</i>	1781879	1782418	DNA repair protein RadC (pseudogene)
<i>ermA2</i>	1776597	1777328	rRNA adenine N-6-methyltransferase 2	SATW20_16530	1782415	1783122	type III leader peptidase family protein
<i>ant2</i>	1777454	1778236	3'-adenylyltransferase 2	<i>folC</i>	1783392	1784663	putative folylpolyglutamate synthase
SATW20_16500	1778387	1778764	Tn554 transposase C	<i>valS</i>	1784676	1787306	valyl-tRNA synthetase
<i>tnpB2</i>	1778771	1780663	transposase B 2				

Table B.4.3.3. Genes within the third recombination region correspond to the Tn554 transposon.

B.4.3.4 Gene function within the second recombination region (ϕ Sa3 phage)

Gene	Start position	End position	Gene function	Gene	Start position	End position	Gene function
SATW20_19300	2104222	2105508	conserved hypothetical protein	SATW20_19660	2135389	2135775	phage protein
<i>mapW</i>	2105832	2106137	MHC class II analog (pseudogene)	SATW20_19670	2135772	2135978	phage protein
<i>mapW</i>	2106136	2107887	MHC class II analog (pseudogene)	SATW20_19680	2136015	2136551	phage dUTPase
<i>h1b</i>	2108276	2108458	phospholipase C precursor	SATW20_19690	2136544	2136954	phage protein

			(pseudogene)				
<i>h1b</i>	2108845	2109024	phage protein	SATW20_19700	2136951	2137436	phage protein
SATW20_19340	2109048	2109356	phage protein	SATW20_19710	2137429	2137713	phage protein
SATW20_19350	2109335	2109595	phage protein	SATW20_19720	2137710	2138159	phage protein
<i>scn</i>	2109648	2109998	staphylococcal complement inhibitor SCIN	SATW20_19730	2138224	2138466	phage protein
SATW20_19370	2110508	2110843	amidase (fragment)	SATW20_19740	2138469	2138726	phage protein
<i>sak</i>	2111495	2111986	staphylokinase precursor	SATW20_19750	2138808	2139098	phage protein
SATW20_19390	2112177	2112932	autolysin	SATW20_19760	2139111	2139515	phage protein
SATW20_19400	2112944	2113198	holin	SATW20_19770	2139524	2139742	phage protein
<i>sea</i>	2113695	2114468	enterotoxin type A precursor	SATW20_19780	2139749	2140642	phage DNA replication protein
SATW20_19420	2114841	2115215	phage protein	SATW20_19790	2140672	2141142	putative single-strand DNA-binding protein
SATW20_19430	2115271	2115558	phage protein	SATW20_19800	2141143	2141628	phage protein
SATW20_19440	2115604	2115756	phage protein	SATW20_19810	2141841	2142761	RecT family phage protein
SATW20_19450	2115749	2119531	phage protein	SATW20_19820	2142763	2144706	phage protein
SATW20_19460	2119547	2121031	phage protein	SATW20_19830	2144715	2144993	phage protein
SATW20_19470	2121028	2125557	phage tail length tape measure protein	SATW20_19840	2144987	2145247	phage protein
SATW20_19480	2125614	2125748	phage protein	SATW20_19850	2145228	2145557	phage protein
SATW20_19490	2125802	2126152	phage protein	SATW20_19860	2145647	2145808	phage protein
SATW20_19500	2126202	2126471	hypothetical phage protein (pseudogene)	SATW20_19870	2145805	2146125	phage protein
SATW20_19500	2126471	2127112	hypothetical phage protein (pseudogene)	SATW20_19880	2146181	2146816	phage protein
SATW20_19510	2127113	2127520	phage protein	SATW20_19890	2146831	2146971	phage protein
SATW20_19520	2127517	2127921	hypothetical phage protein	SATW20_19900	2147002	2147199	phage protein
SATW20_19530	2127918	2128280	phage protein	SATW20_19910	2147215	2147964	phage antirepressor protein
SATW20_19540	2128264	2128548	phage protein	SATW20_19920	2148015	2148344	phage protein
SATW20_19550	2128538	2128822	phage protein	SATW20_19930	2148333	2148548	phage protein
SATW20_19560	2128842	2129987	phage protein	SATW20_19940	2148564	2148827	phage DNA-binding protein
SATW20_195	2130011	2130748	putative Clp	SATW20_199	2148960	2149673	putative phage

Appendices

70			protease	50			repressor protein
SATW20_19580	2130732	2131919	phage portal protein	SATW20_19960	2149689	2150621	putative phage exonuclease
SATW20_19590	2131935	2133596	phage terminase protein	SATW20_19970	2150627	2150968	phage protein
SATW20_19600	2133593	2133937	phage protein	SATW20_19980	2151172	2151354	phage protein
SATW20_19610	2133925	2134074	phage protein	<i>hly</i>	2151448	2151918	putative lipoprotein
SATW20_19620	2134068	2134367	phage HNH endonuclease protein	<i>int</i>	2151977	2153014	integrase
SATW20_19630	2134599	2135015	phage protein	<i>hly</i>	2153086	2153895	phospholipase C precursor (pseudogene)
SATW20_19640	2135043	2135243	phage protein	SATW20_20010	2154133	2155149	putative leukocidin F subunit
SATW20_19650	2135243	2135392	phage protein				

Table B.4.3.4. Genes within the fourth recombination region correspond to the ϕ Sa3 phage.

B.4.4 Predicted unique gene functions

Isolate	Gene ID	Gene function
ST239A	<i>KMGABMOG_02764</i>	Glycine/sarcosine N-methyltransferase (plasmid-linked)
	<i>KMGABMOG_02765</i>	Isoleucine--tRNA ligase (plasmid-linked)
	<i>KMGABMOG_02768</i>	Chloramphenicol acetyltransferase (plasmid-linked)
	<i>KMGABMOG_02769</i>	Rlx – mobilization protein (plasmid-linked)
	<i>KMGABMOG_02771</i>	Rlx – relaxation complex formation protein (plasmid-linked)
	<i>KMGABMOG_02772</i>	Jypothetical protein (plasmid-linked)
	<i>KMGABMOG_02844</i>	LexA repressor
ST239B	<i>CEJCKJAB_00001</i>	Hypothetical protein
	<i>CEJCKJAB_00002</i>	Poly(glycerol-phosphate) alpha-glucosyltransferase
	<i>CEJCKJAB_00003</i>	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase
	<i>CEJCKJAB_02784</i>	Phage protein
	<i>CEJCKJAB_02786</i>	Helix-turn-helix transcriptional regulator
	<i>CEJCKJAB_02790</i>	Chloramphenicol acetyltransferase
	<i>CEJCKJAB_02791</i>	Replication protein
ST239D	<i>CIPDHBFC_02943</i>	Cadmium resistance transcriptional regulatory protein CadC (plasmid-linked)
	<i>CIPDHBFC_02946</i>	Hypothetical protein (plasmid-linked)
	<i>CIPDHBFC_02948</i>	HTH-type transcriptional regulator QacR (plasmid-linked)
ST30A	<i>IOKCAHHF_00759</i>	Hypothetical protein
	<i>IOKCAHHF_00760</i>	Hypothetical protein
ST30B	<i>HBIBGFGD_01353</i>	Hypothetical phage protein
	<i>HBIBGFGD_01354</i>	Phosphatidylinositol kinase
	<i>HBIBGFGD_01373</i>	Hypothetical protein
	<i>HBIBGFGD_01408</i>	Hypothetical phage protein
ST30E	<i>BNDAHCID_02854</i>	Hypothetical protein (plasmid-linked)
	<i>BNDAHCID_02855</i>	Hypothetical protein (plasmid-linked)
	<i>BNDAHCID_02856</i>	TraD conjugation protein (plasmid-linked)
	<i>BNDAHCID_02857</i>	DUF87 domain-containing protein (plasmid-linked)
	<i>BNDAHCID_02858</i>	Hypothetical protein (plasmid-linked)
	<i>BNDAHCID_02859</i>	N-acetylmuramoyl-L-alanine amidase domain-containing protein
	<i>BNDAHCID_02860</i>	Conjugal transfer protein TraH (plasmid-linked)

	<i>BNDACID_02862</i>	Hypothetical protein (plasmid-linked)
	<i>BNDACID_02863</i>	Tpe VI secretion protein VirD4 (plasmid-linked)
	<i>BNDACID_02864</i>	Hypothetical protein (plasmid-linked)
	<i>BNDACID_02887</i>	putative replication initiation protein Rep(RC) (plasmid-linked)
	<i>BNDACID_02895</i>	DegV domain-containing protein (plasmid-linked)
ST8C	<i>JGFFFGPO_01385</i>	Hypothetical protein
	<i>JGFFFGPO_02665</i>	Abortive phage resistance protein
	<i>JGFFFGPO_02741</i>	Replication protein (plasmid-linked)
	<i>JGFFFGPO_02743</i>	Hypothetical protein (plasmid-linked)
ST8F	<i>IHLBGJB_01515</i>	Hypothetical protein
	<i>IHLBGJB_01516</i>	Hypothetical protein
	<i>IHLBGJB_02193</i>	PHP domain protein (plasmid-linked)
	<i>IHLBGJB_02195</i>	Immunodominant staphylococcal antigen B (plasmid-linked)
	<i>IHLBGJB_02200</i>	Hypothetical protein (plasmid-linked)
	<i>IHLBGJB_02212</i>	Transposase (plasmid-linked)
	<i>IHLBGJB_02214</i>	HTH-type transcriptional regulator NmtR (plasmid-linked)
	<i>IHLBGJB_02215</i>	DNA-invertase hin (plasmid-linked)
	<i>IHLBGJB_02284</i>	Hypothetical protein
	<i>IHLBGJB_02396</i>	Tyrosine recombinase XerC
	<i>IHLBGJB_02397</i>	Transposase
	<i>IHLBGJB_02398</i>	Yold-like family protein
	<i>IHLBGJB_02770</i>	EcoKI restriction-modification system protein HsdS
	<i>IHLBGJB_02772</i>	Hypothetical protein
	<i>IHLBGJB_02779</i>	Hypothetical protein
	<i>IHLBGJB_02780</i>	Hypothetical protein
	<i>IHLBGJB_02781</i>	FRG domain protein
	<i>IHLBGJB_02782</i>	Hypothetical protein
	<i>IHLBGJB_02784</i>	TIR domain protein
	<i>IHLBGJB_02787</i>	Hypothetical protein
	<i>IHLBGJB_02788</i>	Fibronectin-binding protein
	<i>IHLBGJB_02792</i>	Hypothetical protein
	<i>IHLBGJB_03198</i>	Membrane-associated protein TcaA
	<i>IHLBGJB_03199</i>	Putative multidrug ABC transporter ATP-binding protein YbhF (plasmid-linked)
	<i>IHLBGJB_03200</i>	Putative multidrug ABC transporter permease YbhR

Table B.4.4. Predicted unique gene functions.

B.4.5 Location of AMR genes

ST	AMR gene	Hybrid region	ST	AMR gene	Hybrid region	ST	AMR gene	Hybrid region
ST239A	<i>gyrA</i>	ST30	ST8A	<i>tet(38)</i>	ST30	ST30A	<i>tet(38)</i>	ST30
	<i>ermA</i>	ST30		<i>mepA</i>	ST30		<i>mepA</i>	ST8
	<i>tet(K)</i>	ST30	ST8B	<i>tet(38)</i>	ST30		<i>murA</i>	ST8
	<i>tet(38)</i>	ST30		<i>mepA</i>	ST8		<i>mecA</i>	NA
	<i>blaZ</i>	ST30	ST8C	<i>mepA</i>	ST8		<i>blaZ</i>	NA
	<i>mecA</i>	SCCmec		<i>murA</i>	ST8	ST30B	<i>tet(38)</i>	ST30
	<i>mepA</i>	SCCmec		<i>gyrA</i>	NA		<i>mepA</i>	ST8
	<i>rpoB</i>	SCCmec		<i>ANT(4')-Ib</i>	NA		<i>blaZ</i>	ST8
	<i>ileS</i>	ST8		<i>mecA</i>	NA		<i>murA</i>	ST8
	<i>parC</i>	ST8		<i>tet(38)</i>	NA	ST30C	<i>tet(38)</i>	ST30
	<i>AAC(6')-Ie-APH(2'')-Ia</i>	ST8		<i>blaZ</i>	NA		<i>mepA</i>	ST8
	<i>APH(3')-IIIa</i>	ST8	ST8D	<i>tet(38)</i>	ST30		<i>parC</i>	ST8

Appendices

	<i>SAT-4</i>	ST8		<i>dfrG</i>	ST30		<i>blaZ</i>	ST8
	<i>murA</i>	ST8		<i>mepA</i>	ST8		<i>murA</i>	ST8
	<i>mupA</i>	NA		<i>parC</i>	ST8		<i>gyrA</i>	NA
ST239B	<i>ermA</i>	ST30		<i>murA</i>	ST8		<i>ANT(4')-Ib</i>	NA
	<i>blaZ</i>	ST30		<i>gyrA</i>	NA		<i>mecA</i>	NA
	<i>gyrA</i>	ST30		<i>mecA</i>	NA		<i>ermA</i>	NA
	<i>tet(38)</i>	ST30		<i>blaZ</i>	NA		<i>tet(38)</i>	ST30
	<i>mecA</i>	SCCmec		<i>SAT-4</i>	NA		<i>mepA</i>	ST8
	<i>mepA</i>	ST8		<i>APH(3')-IIIa</i>	NA		<i>parC</i>	ST8
	<i>parC</i>	ST8		<i>mphC</i>	NA		<i>ermA</i>	ST8
	<i>AAC(6')-Ie-APH(2'')-Ia</i>	ST8		<i>tet(38)</i>	ST30	ST30D	<i>blaZ</i>	ST8
	<i>APH(3')-IIIa</i>	ST8		<i>mepA</i>	ST8		<i>murA</i>	ST8
	<i>SAT-4</i>	ST8		<i>dfrG</i>	ST8		<i>gyrA</i>	NA
	<i>murA</i>	ST8		<i>parC</i>	ST8		<i>mecA</i>	NA
ST239C	<i>gyrA</i>	ST30	ST8E	<i>gyrA</i>	NA	ST30E	<i>ANT(4')-Ib</i>	ST30
	<i>blaZ</i>	ST30		<i>mecA</i>	NA		<i>mecA</i>	ST30
	<i>tet(38)</i>	ST30		<i>ErmC</i>	NA		<i>tet(38)</i>	ST30
	<i>ermA</i>	SCCmec		<i>AAC(6')-Ie-APH(2'')-Ia</i>	NA		<i>mepA</i>	ST8
	<i>tet(K)</i>	SCCmec	ST8F	<i>vanA</i>	NA		<i>parC</i>	ST8
	<i>mecA</i>	SCCmec		<i>blaZ</i>	NA		<i>blaZ</i>	ST8
	<i>mepA</i>	ST8		<i>tet(38)</i>	ST30		<i>murA</i>	ST8
	<i>dfrG</i>	ST8		<i>mepA</i>	ST30		<i>gyrA</i>	NA
	<i>ileS</i>	ST8		<i>blaZ</i>	ST30		<i>ermA</i>	NA
	<i>parC</i>	ST8		<i>fusC</i>	ST8		<i>AAC(6')-Ie-APH(2'')-Ia</i>	NA
	<i>AAC(6')-Ie-APH(2'')-Ia</i>	ST8		<i>tet(K)</i>	NA		<i>dfrC</i>	NA
	<i>APH(3')-IIIa</i>	ST8						
	<i>SAT-4</i>	ST8						
	<i>qac</i>	NA						
	<i>gyrA</i>	ST30						
	<i>tet(38)</i>	ST30						
	<i>blaZ</i>	ST30						
ST239D	<i>ermA</i>	SCCmec						
	<i>tet(K)</i>	SCCmec						
	<i>mecA</i>	SCCmec						
	<i>mepA</i>	ST8						
	<i>dfrG</i>	ST8						
	<i>parC</i>	ST8						
	<i>AAC(6')-Ie-APH(2'')-Ia</i>	ST8						
	<i>APH(3')-IIIa</i>	ST8						
	<i>SAT-4</i>	ST8						
	<i>murA</i>	ST8						
	<i>qac</i>	NA						

Table B.4.5. Predicted location of the AMR genes in the ST8-like or ST30-like region of the experimental sequences.

B.4.6 AMR gene targets

Antibiotic	Family	Target	Resistance genes detected
Penicillin	Beta-lactam	Penicillin binding protein (inhibits peptidoglycan cross-linking in cell wall formation)	<i>blaZ</i> (beta-lactamase to break down beta-lactams); <i>mecA</i> (encoded PBP2 which does not bind beta-lactams)
Ampicillin	Beta-lactam (broad spectrum)		
Oxacillin	Beta-lactam (beta-lactamase resistant)		<i>mecA</i> (encoded PBP2 which does not bind beta-lactams)
Rifampicin	Rifamycin	RNA polymerase (inhibits RNA synthesis)	<i>rpoB_1</i> (mutation in the rifampicin binding site of RNA polymerase)
Vancomycin	Glycopeptide	D-Ala-D-Ala H-bond polymerisation in NAM and NAG (inhibits cell wall biosynthesis)	<i>vanA</i> , <i>vanHA</i> , <i>vanRA</i> , <i>vanXA</i> , <i>vanSA</i> , <i>vanZA</i> (alteration to the terminal amino acid residues of the NAM/NAG-peptide subunits)
Teicoplanin			
Fosfomycin	Fosfomycin	MurA (inhibits the committed step in peptidoglycan biosynthesis, inhibiting cell wall formation).	<i>murA_2</i> (mutation in fosfomycin binding site of murA)
Moxifloxacin	Fluoroquinolone	DNA replication (DNA gyrase inhibitor)	<i>gyrA_6</i> , <i>parC_1</i> (mutations in DNA gyrase so it is no longer affected by the antibiotic)
Fusidic Acid	Steroid Antibiotic	Ribosome EF-G (protein synthesis inhibition)	<i>fusC</i> (protection of the drug target site)
Gentamicin	Aminoglycoside	30S ribosomal subunit (protein synthesis inhibition)	<i>aac(6'')-aph(2'')</i> , <i>aph(3')-III</i> , <i>ant(4)-Ib</i> (aminoglycoside modifying enzymes)
Mupirocin	Carboxylic Acid Antibiotic	Antibiotic	<i>mupA</i> (a separate isoleucyl t-RNA synthetase gene that is not affected by mupirocin), <i>ileS_1</i> (point mutation in the isoleucyl t-RNA synthetase gene).
Erythromycin	Macrolide/ Lincosamide.Strep togramin	50S ribosomal subunit (protein synthesis inhibition)	<i>ermA</i> (methylates the ribosomal RNA target site so it no longer is affected by the antibiotic), <i>ErmC</i> , (methylates the ribosomal RNA target site so it no longer is affected by the antibiotic) <i>MphC</i> (phosphotransferase that inactivates the antibiotic)
Trimethoprim/ Sulfamethoxazole	Diamino-pyrimidine	Thymide synthesis via dihydrofolate reductase	<i>dfrG</i> (plasmid-encoded dihydrofolate reductase), <i>dfrC</i> (chromosome-encoded dihydrofolate reductase)
Nucleoside	Nucleoside	Protein synthesis	<i>SAT-4</i> (plasmid-encoded streptothricin acetyltransferase)

Table B.4.6. Antibiotic families and the corresponding resistance genes identified in the isolate sequences.

B.5 McDonald-Kreitman and parallelism

B.5.1 McDonald-Kreitman Net Neutrality Index

STs	Region	Dn	Ds	Dn/Ds	Pn	Ps	Pn/Ps	N	Selection pressure
ST239, ST8, ST30	ST30-like	1071	4146	0.258	2174	1474	1.475	5.710	>1 = Negative Selection
	ST8-like	6766	15235	0.444	6037	9147	0.660	1.486	>1 = Negative Selection
ST239, ST30	ST30-like	40	32	1.25	1406	740	1.900	1.520	>1 = Negative Selection
	ST8-like	4032	18147	0.222	6023	3939	1.529	6.882	>1 = Negative Selection
ST239, ST8	ST30-like	888	4331	0.205	1431	1040	1.376	6.711	>1 = Negative Selection
	ST8-like	42	50	0.840	6080	3508	1.733	2.063	>1 = Negative Selection

Table B.5.1. McDonald-Kreitman net neutrality index for the two genetic regions of ST239. Fixed non-synonymous substitutions (Dn); fixed synonymous substitutions (Ds); non-synonymous polymorphisms (Pn); synonymous polymorphisms (Ps); McDonald-Kreitman net neutrality index (N).

B.5.2 Distribution of fixed substitutions in ST239

Number of SNPs per gene	Number of genes containing X SNPs		Proportion of genes containing X SNPs	
	ST8-like region	ST30-like region	ST8-like region	ST30-like region
0	324	37	0.195	0.117
1	353	58	0.213	0.184
2	306	69	0.184	0.218
3	226	59	0.136	0.187
4	157	33	0.095	0.104
5	108	26	0.065	0.082
6	71	16	0.043	0.051
7	35	6	0.021	0.019
8	26	3	0.016	0.009
9	19	2	0.011	0.006
10	10	0	0.006	0.000
11	9	0	0.005	0.000
12	2	1	0.001	0.003
13	3	2	0.002	0.006
14	5	0	0.003	0.000
15	1	1	0.001	0.003
16	0	0	0.000	0.000
17	1	0	0.001	0.000
18	0	0	0.000	0.000
19	0	0	0.000	0.000
20	0	0	0.000	0.000
21	0	0	0.000	0.000
22	1	0	0.001	0.000
23	0	0	0.000	0.000
24	1	0	0.001	0.000

25	0	1	0.000	0.003
26	0	1	0.000	0.003
27	1	0	0.001	0.000
62	0	1	0.000	0.003
TOTAL	1659	316	1	1

Table B.5.2. True distribution of fixed substitutions across the acquired and backbone regions of the ST239 sequences.

B.5.3 Non-synonymous and synonymous substitutions in ST239

Region	Genes	Total genes	Total SNPs	Dn	Ds	Dn/Ds
ST30-like region	All	316	981	640	341	1.877
	Highly variable	6	154	102	52	1.962
	Non-highly variable	310	827	538	289	1.862
ST30-like region minus <i>spa</i> outlier	All	315	919	592	327	1.810
	Highly variable	5	92	54	38	1.421
	Non-highly variable	310	827	538	289	1.862
ST8-like region	All	1659	4298	2746	1552	1.769
	Highly variable	20	284	186	98	1.898
	Non-highly variable	1639	4014	2560	1454	1.761

Table B.5.3. Proportion of non-synonymous (Dn) and synonymous (Ds) substitutions in ST239.

B.5.4 Vancomycin resistance mutations in ST239

B.5.4.1 Mutations in *rpoB*

Consensus amino acid	Amino acid position in <i>RpoB</i>	Alternative amino acid	Rifampicin susceptibility (MIC µg/mL)	Reference
A	86	V	-	-
V	135	L	S	Ko <i>et al.</i> , 2018
T	229	I	-	-
F	461	Y	-	-
L	466	S	R (0.25)	O'Neill, <i>et al.</i> , 2005
D	471	E	R (0.125 – 0.5)	O'Neill, <i>et al.</i> , 2005; Wichelhaus, 2001
A	477	T	R (1024)	Wichelhaus, 2001
H	481	*,Y	R (512 - 1024)	Wichelhaus, 2001; O'Neill, <i>et al.</i> , 2005
S	529	L	R (128 – 512)	Wichelhaus, 2001
V	536	A	-	-
H	677	Q	-	-
H	986	Y	-	-
D	1012	G	-	-
I	1119	V	-	-
A	1174	T	-	-
P	1175	T	-	-

Table B.5.4.1. Non-synonymous resistance mutations in *rpoB* in ST239. R, resistant; S, susceptible.

B.5.4.2 Mutations in *walK*

Consensus amino acid	Amino acid position in <i>WalK</i>	Alternative amino acid	Vancomycin susceptibility (MIC µg/mL)	Reference
I	76	V	-	-
I	197	T	-	-
M	220	I	R (3)	Baines <i>et al.</i> , 2015
S	221	P	R (8)	Peng <i>et al.</i> , 2016
G	223	D	R (2)	Hu, <i>et al.</i> , 2014
T	226	M	-	-
R	263	C	Datpomycin R (4)	Cameron <i>et al.</i> , 2015
L	340	I	-	-
E	424	D	-	-
T	484	A	-	-
T	590	A	-	-
T	595	I	-	-

Table B.5.4.2. Non-synonymous mutations in *walK* in ST239. R, resistant; S, susceptible.

B.6 Growth rate

B.6.1 Growth rate

Isolate ID	Max. growth rate in TSB (mODh ⁻¹)	Max. OD in TSB (mOD)	Max. growth rate in BHI (mODh ⁻¹)	Max. OD in BHI (mOD)	Max. growth rate in PS (mODh ⁻¹)	Max. OD in PS (mOD)
ST239A	0.6673127	1.18552	0.6210978	0.9535461	0.2571664	0.142363
ST239B	1.156642	0.8787449	0.9647533	0.9835746	0.000	0.003733333
ST239C	0.6514542	0.7762167	0.6155266	0.9739809	0.1916588	0.1436573
ST239D	0.6979599	0.8197601	0.5997019	0.8956542	0.1536503	0.0799178
ST8A	0.8209837	0.8265013	1.029663	0.8798304	0.4436607	0.1672324
ST8B	0.7802226	0.8257646	0.8000994	0.8039103	0.2442413	0.1324631
ST8C	1.383074	1.047663	0.6031034	0.9436856	0.1653859	0.1817249
ST8D	1.38952	1.032838	0.5028648	0.9645712	0.2774309	0.216094
ST8E	1.089149	1.169782	0.5550694	0.9605367	0.2730356	0.1733057
ST8F	1.104703	1.093125	0.3656964	0.7837674	0.2972563	0.2030428
ST30A	0.4408317	0.7359695	0.4228023	0.8762588	0.241872	0.2322049
ST30B	0.3874646	0.6382172	0.4164292	0.8320057	0.2604669	0.2684529
ST30C	1.55082	1.183488	0.6877863	1.038819	0.2408207	0.0749451
ST30D	0.524502	0.534915	0.7963747	0.869563	0.3625332	0.2323706
ST30E	0.3937441	0.5427411	0.7111333	0.6889607	0.2755679	0.1968535

Table B.6.1. Average maximum growth rate and maximum OD of isolates in TSB, BHI and PS.

B.7 Statistical models

B.7.1 Full growth rate modeling

Effect	Degrees of freedom	Sum of squares	Mean squares	F-value	p-value
ST	2	0.8605699	0.4302855	2.6211	0.00026
Plate	18	0.6956621	0.038648	0.7743	0.72842
Media	2	5.8354736	2.917737	58.4593	0.00000
Isolate[ST]	12	2.80154214	0.233460	4.6776	0.00000
Media x ST	4	1.9538383	0.488466	9.7867	0.00000
Error	186	9.283369	0.049911	-	-

Table B.7.1. Full ANOVA table of growth rate effects.

B.7.2 Least squares means table of ST effect

ST	Least squares mean	Standard Error
ST8	0.674	0.0233
ST30	0.514	0.0255
ST239	0.548	0.0286

Table B.7.2. Least squares means table of ST effect on maximum growth rate.

B.7.3 Least squares means table of media effect

Media	Least squares mean	Standard Error
BHI	0.650	0.0259
PS	0.237	0.0259
TSB	0.849	0.0259

Table B.7.3. Least squares means table of media effect on maximum growth rate

B.7.4 Least squares means table of media-ST effect

Media	ST	Least squares mean	Standard error
TSB	ST239	0.793	0.0495
TSB	ST8	1.095	0.0404
TSB	ST30	0.659	0.0442
BHI	ST239	0.700	0.0495
BHI	ST8	0.643	0.0404
BHI	ST30	0.607	0.0442
PS	ST239	0.151	0.0495
PS	ST8	0.284	0.0404
PS	ST30	0.276	0.0442

Table B.7.4. Least squares means table of the effect of media-ST interactions on maximum growth rate.

B.7.5 Full growth rate modeling

Effect	Degrees of freedom	Sum of squares	Mean squares	F-value	p-value
ST	2	0.2605196	0.1302598	7.8562	0.00065
Isolate[ST]	12	4.2435429	0.3536286	21.3279	0.00000
Media	2	0.1618169	0.0809085	4.8797	0.00936
Media x ST	4	0.5758495	0.1439624	8.6826	0.00000
Replicate	2	0.0018381	0.0009191	0.0554	0.94611
Replicate x ST	4	0.0086542	0.0021635	0.1305	0.97097
Error	108	1.7906990	0.016581	-	-

Table B.7.5. Full ANOVA table of growth rate effects.

B.7.6 Least squares means table of ST effect

ST	Least squares mean	Standard Error
ST8	-0.0119	0.0171
ST30	-0.0128	0.0187
ST239	-0.112	0.0210

Table B.7.6. Least squares means table of ST effect on maximum growth rate.

B.7.7 Least squares means table of media effect

Media	Least squares mean	Standard Error
BHI	-0.0910	0.0190
PS	-0.00561	0.0190
TSB	-0.0397	0.0190

Table B.7.7. Least squares means table of media effect on maximum growth rate

B.7.8 Least squares means table of media-ST effect

Media	ST	Least squares mean	Standard error
TSB	ST239	-0.156	0.0363
TSB	ST8	0.0449	0.0296
TSB	ST30	-0.00827	0.0325
BHI	ST239	-0.238	0.0363
BHI	ST8	-0.0542	0.0296
BHI	ST30	0.0187	0.0325
PS	ST239	0.0583	0.0363
PS	ST8	-0.062	0.0296
PS	ST30	-0.0490	0.0325

Table B.7.8. Least squares means table of the effect of media-ST interactions on maximum growth rate.

