

Genomic divergence events in methicillin resistant *Staphylococcus aureus* (MRSA): an *in silico* approach to investigate the evolutionary epidemiology of *S. aureus*

Peter Money

A thesis submitted in partial fulfilment of the requirements of Kingston University for the degree of
Doctor of Philosophy

June 2016

Abstract

Staphylococci commonly form a part of the normal micro-flora for humans and a wide variety of animals but also represent a group of clinically significant pathogens. However, it has also been recognised that for bacteria to become pathogens, they must have the ability to accomplish infection and consequently cause disease in the host. This is known as pathogenicity and virulence is a degree of the pathogenicity observed. Virulence factors are proteins made through the expression of virulence factor genes produced by a pathogen which contribute to the pathogenicity, and literature research identified a variety of virulence factor genes. The predominant pathogenic staphylococcal species is *Staphylococcus aureus*, with methicillin-resistant *Staphylococcus aureus* (MRSA) recognised as significant in both veterinary and medical contexts. The role of the *S. aureus* in human disease is clear; however, the organism is also well recognised as a cause of animal disease with animal specific strains such as ST398 becoming increasingly common. The approach taken to determine how similar or different the *S. aureus* strains are at genomic level as well as investigate the virulence factors that may be under positive selection was undertaken by employing the wealth of nucleotide sequences data available for a number methicillin-resistant and methicillin-sensitive *S. aureus* strains. This helped to establish if there was genomic variation between strains and investigate the potential relationship of these alterations in host/ niche preferences. In total, 53 strains of *S. aureus* were investigated; including, three novel isolates of *S. aureus* from human patients that were provided by Prof. Valerie Edward-Jones (Manchester Metropolitan University) and 3 novel isolates from various farm animals (cattle, chicken & swine) provided by Dr Patrick Butaye (The Veterinary & Agrochemical Research Centre, Belgium). The novel isolates were sequenced with Ion Torrent Next-Generation sequencing methodologies. Literature research has classified the strains of *S. aureus* as either hospital-associated (HA), community-associated (CA) or livestock-associated (LA); these classifications were further investigated. Virulence factor gene sequences, if present, were isolated from the *S. aureus* strains and a comparative analysis was carried out; this includes initial MAUVE analysis and more detailed maximum-likelihood phylogenetic analysis, using

MEGA. A bioinformatics software called DnaSP was used for the analysis of polymorphism from aligned virulence factor nucleotide sequences and the dN/dS (synonymous vs nonsynonymous mutations) ratio (ω) was investigated to infer the evolutionary mechanisms that may be acting on the genes. Haplotype networks were also created to investigate the *S. aureus* population to determine the diversity of the population, determine any geographical relationship and further investigate the classification HA, CA & LA that has been given in literature. This study concluded that the presence of functional Pantone-Valentine Leukocidin (PVL) a bi-component virulence factor encoded by the *lukS-PV* and *lukF-PV* genes, typically found in highly virulent CA *S. aureus* strains, were detected in these virulent strains, but surprisingly also in the novel livestock-associated strains. This study also concluded that the population of *S. aureus* was very diverse and a geographical relationship was not observed. It was further concluded that the selective pressures on individual virulence factor genes causes *S. aureus* to evolve, with a number of genes recognised as undergoing positive selection; *lukS-PV*, *lekF-PV*, *chp*, *sspA*, *vwb*, *lip*, *lip1*, *nucA*, *hysA*, *hysA1*, *aur*, *fnb*, *ebpS*, *bbp*, *entE*, *seb* and *tst*. It was also suggested that the 3 classifications denoted in the literature may possibly be artificial and at a molecular level the strains are in fact very similar and do not represent genuine divisions. This has also been recently published by Bal *et al.* (2016) and this thesis supports their hypothesis; they describe the blurring of each of the 3 epidemiological groups (HA, CA & LA), demonstrating the limited relevance of this classification.

Dedication

This PhD thesis is dedicated to my wife, Elisa, who has been a constant source of support and encouragement during the challenges of research and life. I am truly thankful for having you in my life. This work is also dedicated to my parents, Edgar and Parveen Money, for their unconditional love and whose good examples have taught me to work hard.

Acknowledgments

There are a number of people without whom this thesis may not have been written, and to whom I am greatly indebted.

First and foremost, I would like to sincerely thank my supervisor, Professor Mark Fielder, for his guidance and support throughout this study, and especially for his confidence in me. Secondly, a special thanks also goes to my co-supervisor, Dr Gary Forster-Wilkins, for his valuable advice, and constructive criticism. I am also very grateful to Dr Scott Lawton for his insightful discussions and suggestions. I would also like to thank Kingston University, I recognise that this research would not have been possible without their financial assistance.

This project is also in memory of Meesha Patel who started this project. A dedicated student and a real budding scientist who but never got to fully bloom. I have heard many wonderful stories of you.

Abstract	II
Dedication	IV
Acknowledgments	V
Content	VI
Content of tables	XI
Content of figures	XIII
List of abbreviations	XXIII
1.0 Introduction	1
1.1 Discovery of <i>Staphylococcus aureus</i>	2
1.2 Staphylococci	4
1.3 Treatment of <i>Staphylococcus aureus</i> infection	5
1.3.1 Pre-antibiotic era	5
1.3.2 The era of antibiotics	5
1.3.3 Structure of the bacterial cell wall	6
1.3.4 Penicillin resistance	6
1.3.5 Methicillin resistance	7
1.3.6 Treatment option for <i>Staphylococcus aureus</i>	9
1.4 <i>Staphylococcus aureus</i> genome	11
1.5 Virulence factors	13
1.5.1 Genes	13
1.5.1.1 mecA gene and penicillin-binding protein-2	14
1.5.2 Adhesion genes	14
1.5.2.1 Fibronectin-binding protein	14
1.5.2.2 Collagen-binding protein	14
1.5.2.3 Elastin-binding protein	15
1.5.2.4 Bone-sialoprotein protein	15
1.5.2.5 Clumping factor	16
1.5.2.6 Extracellular fibrinogen-binding protein	16
1.5.2.7 Extracellular adherence protein	16
1.5.3 Virulence factors involved in persistence	17
1.5.3.1 Delta-aminolevulinic acid dehydratase	17
1.5.3.2 Bacterial efflux pumps	17
1.5.4 Immune evasion genes	18
1.5.4.1 Staphylokinase	18
1.5.4.2 Staphylococcal complement inhibitor	19

1.5.4.3 Chemotaxis inhibitory protein	20
1.5.4.4 IgG binding protein	20
1.5.5 Enzymes and other secreted proteins	20
1.5.5.1 Serine Protease, cysteine protease precursor, cysteine protease	20
1.5.5.2 Aureolysin	21
1.5.5.3 Hyaluronate lysate	21
1.5.5.4 Staphylococcal coagulase	22
1.5.5.5 von Willebrand factor binding protein	22
1.5.5.6 Lipase	22
1.5.5.7 Staphylococcal nuclease	23
1.5.5.8 Thermonuclease	23
1.5.5.9 1-Phosphatidylinositol phosphodiesterase	24
1.5.6 Cytotoxins	24
1.5.6.1 Alpha hemolysin	24
1.5.6.2 Beta hemolysin	25
1.5.6.3 Gamma hemolysin and Panton-Valentine leuckocidin	25
1.5.7 Enterotoxins	26
1.5.7.1 Staphylococcal enterotoxins	26
1.5.7.2 Toxic shock syndrome toxin-1	27
1.6 Clinical significance & epidemiology	27
1.7 Symptoms and clinical presentations	33
1.7.1 Cellulitis	35
1.7.2 Impetigo	35
1.7.3 Folliculitis	35
1.7.4 Toxic shock syndrome	36
1.7.5 Scalded skin syndrome	36
1.8 MRSA evolution	38
1.8.1 Hospital-associated MRSA	38
1.8.2 Community-associated MRSA	38
1.8.3 Livestock-associated MRSA	40
1.9 Next generation sequencing	41
1.10 Bioinformatics and its use in the study of <i>S. aureus</i>	42
1.11 Aims and objectives	44
2.0 Materials and Methods	45
2.1 Bacterial strains	46
2.2 Storage of isolates	47
2.3 Resuscitation of isolates prior to use	47
2.4 Isolation of bacterial genomic DNA	48
2.5 Spectrophotometric analysis	50
2.6 Gel electrophoresis	50
2.7 Ion Torrent: Next Generation Sequencing	51
2.7.1 Ion Xpress™ Plus gDNA fragment library preparation	51
2.7.1.1 Fragmentation of the DNA	52

2.7.1.2	Clean up using Agencourt® AMPure® XP reagents	52
2.7.1.3	Bio-analyser	53
2.7.1.4	Ligate adapters	54
2.7.1.5	E-Gel® SizeSelect™ agarose electrophoresis	55
2.7.1.6	Amplify and purify libraries	56
2.7.1.7	Quantify and pool barcoded libraries	57
2.7.2	Ion PGM™ Template OT2 400 Kit	58
2.7.2.1	Configuring the machine	58
2.7.2.2	Loading Ion PGM OneTouch Plus Reaction Filter	61
2.7.2.3	Recover the Ion Sphere particles	63
2.7.2.4	Enrich the template-positive Ion PGM Template	64
2.7.3	Ion PGM™ Sequencing 400 Kit	65
2.7.3.1	Chip check	67
2.7.3.2	Loading the chip	68
2.7.3.3	Ion Torrent	69
2.8	CLC Bio	70
2.8.1	Uploading files to CLC Bio	70
2.8.2	Read mapping	71
2.8.3	Virulence factor database	72
2.8.4	Extracting virulence factors	73
2.9	GenBank	74
2.9.1	Acquiring the bacterial genome	74
2.9.2	<i>S. aureus</i> literature review	76
2.9.3	Virulence factor literature review	76
2.9.4	Acquiring the virulence factor genes	75
2.10	MAUVE	79
2.11	BLAST	82
2.12	MEGA	85
2.12.1	Load sequence	85
2.12.2	Constructing a neighbour-joining tree	87
2.12.3	Constructing a maximum-parsimony tree	89
2.12.4	Constructing a maximum-likelihood tree	91
2.13	Muscle	94
2.14	Phylogeny.fr	96
2.15	DnaSP	97
2.15.1	Load sequence alignment	98
2.15.2	Generate polymorphism data	102
2.15.3	Generate polymorphic sites	103
2.15.4	Generate synonymous and nonsynonymous substitutions	104
2.15.5	Define datasets	105
2.15.6	DNA divergence	106
2.15.7	Generate haplotype data	107
2.15.8	BioEdit	109
2.15.9	Generating a NEXUS file	110

2.16 PopART	110
3.0 Strain selection, strain classification and virulence factor extraction	110
3.1 Introduction	111
3.2 Aims	112
3.3 Materials and methods	113
3.4 Results	114
3.4.1 Strain selection	114
3.4.2 Strain classification	114
3.4.3 MAUVE analysis	126
3.4.4 Novel bacterial classification	133
3.4.5 Bacterial genomic DNA isolation	135
3.4.6 Spectrophotometric analysis	135
3.4.7 Gel electrophoresis	136
3.4.8 Next generation sequencing	137
3.4.8.1 Fragmentation of the DNA	137
3.4.8.2 Ligation to adapters	140
3.4.8.3 Chip loading	143
3.4.8.4 Ion Torrent output	145
3.4.8.5 Genome assembly	145
3.4.9 Selecting and acquiring virulence factors	145
3.4.10 Extracting virulence factors from published <i>S. aureus</i> strains	148
3.4.11 Extracting virulence factors from novel <i>S. aureus</i> strains	148
3.5 Discussion	155
4.0 Phylogenetic analysis of <i>S. aureus</i> strains	170
4.1 Introduction	171
4.2 Aims	172
4.3 Materials and methods	173
4.4 Results	174
4.3.1 Aligning the <i>S. aureus</i> sequences	174
4.3.2 Phylogeny.fr analysis	175
4.3.3 MEGA analysis	177
4.5 Discussion	187
5.0 Investigating positive selection in <i>S. aureus</i> and divergence between populations	192
5.1 Introduction	192
5.2 Aims	194
5.3 Materials and Methods	194
5.4 Results	195
5.4.1 Haplotype and nucleotide diversity	195
5.4.2 Positive selection	197
5.4.3 Divergence between populations	207
5.5 Discussion	209

6.0 Population genetics of <i>S. aureus</i>	213
6.1 Introduction	214
6.2 Aims	216
6.3 Materials and methods	218
6.4 Results	218
6.4.1 Quantifying DNA polymorphism in the <i>S. aureus</i> population	218
6.4.2 Investigating the geographical relationship of the <i>S. aureus</i> population	220
6.4.3 Investigating the relationship of the <i>S. aureus</i> population	223
6.4.4 Sliding window analysis of synonymous and nonsynonymous mutations	227
6.4.5 Pairwise comparisons of <i>S. aureus</i> virulence factor genes	230
6.5 Discussion	233
7.0 Discussion and future work	235
7.1 Discussion	236
7.2 Future work	241
8.0 Bibliography	243
9.0 Appendix	267

List of Tables:

1.1	Clinical symptoms and diseases associated with <i>Staphylococcus aureus</i> in different hosts	34
2.1	Table 2.1: The animal and human <i>Staphylococcus aureus</i> strains that will be used for further studies along with its date of arrival at Kingston University.	47
2.2	The thermal cycler was programed to run the sequence shown in the table for this stage.	57
2.3	The thermal cycler was programed to run the sequence shown in the table, cycling for a maximum of 5 times.	55
3.1	Table identifying the the <i>S. aureus</i> strain acquired from GenBank, its corresponding NCBI/ GI accession numbers, the length of the genome in base pairs, classification and its corresponding reference in the literature.	120
3.2	Table identifying the the <i>S. aureus</i> strain acquired from the collaborators indicating its host, and the classification it has been designated.	140
3.3a	The results of the spectrophotometric analysis for the six <i>S. aureus</i> samples; which include the average concentration and the A260/ A280 readings.	141
3.3b	Showing the assigned barcode adaptors, the pool number and output of the Bioanalyser®; which includes the size of the isolate in base pairs, its molarity and the aligned migration time.	146
3.4a	Showing the success rate of Pool One and Pool Two chip loading; which is represented by % usable reads, % loaded / empty wells, % enrichment and % clonal/ polyclonal.	150
3.4b	This table shows the results of the ion torrent run which includes the total number of contigs as well as the average length of the contigs. It also shows the % of mapped reads including the average length of the mapped reads.	152
3.5	The <i>S. aureus</i> virulence factors and their respective genes.	154
3.6	A representation of the presence of virulence factor genes present in published hospital-associated MRSA strains. The MRSA strains are given across the top of the figure and the genes along the side. The black colour indicates that a gene is present in the <i>S. aureus</i> and the white indicates that it is not.	158
3.7	A representation of the presence of virulence factor genes present in published community-associated MRSA strains. The MRSA strains are given across the top of the figure and the genes along the side. The coloured block indicates that a gene is present in the <i>S. aureus</i> and the white indicates that it is not present.	159

3.8	A representation of the presence of virulence factor genes present in published livestock-associated MRSA strains. The MRSA strains are given across the top of the figure and the genes along the side. The coloured block indicates that the gene is present in the <i>S. aureus</i> strain and the white indicates that it is not.	160
3.9	A representation of the presence of virulence factor genes present in the Novel MRSA strains. The MRSA strains are given across the top of the figure and the genes along the side. The coloured block indicates that the gene is present in the <i>S. aureus</i> strain and the white indicates that it is not.	161
5.1	Table showing the virulence factor genes that are under positive selection and the statistical significance tests. Data was generated using DnaSP. The complete data generate dis shown in Appendix 1.	198
6.1	A table showing the <i>Staphylococcus aureus</i> strains that were used in this study along with the countries that that they were isolated from. The information was extracted from Table 3.1.	221
6.2	A table showing the statistical tests of neutrality and the goodness-of-fit indexes for population expansion for <i>S. aureus</i> viruelnce factors.	232

List of Figures

1.1a	Figure 1.1a: Taken and adapted from Lowey., 2003. Mechanisms of penicillin resistance.	7
1.1b	Figure 1.1b: A flow cycle for the evolution of drug resistance in <i>S. aureus</i> (Adapted from WHO-1, 2012). This diagram shows that every time that a new antibiotic is discovered to combat the <i>S. aureus</i> ; drug resistance will follow. Notably, Teixobactin is the first of a new class of antibiotic that has been discovered, in January 2015. The question mark for '2015 >?' represents that it takes years for new drugs to be developed, but only a short amount of time for antibiotic resistance to be detected (Ventola., 2015).	10
1.2	A genetic map of the Genome of <i>S. aureus</i> MW2.	12
1.3	Taken from the Annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net); showing the percentage (%) of <i>S. aureus</i> isolates with resistance to methicillin (MRSA), by countries in 2009 (ECDC-1, 2010).	29
1.4	Taken from the Annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net); showing the percentage (%) of <i>S. aureus</i> isolates with resistance to methicillin (MRSA), by countries in 2009 (ECDC-2, 2015).	29
1.5	Figure 1.5: A graph representing the number of death certificates mentioning <i>S. aureus</i> : by Methicillin Resistant <i>Staphylococcus aureus</i> (MRSA) in England & Wales. The data was adapted from the Office for National Statistics (ONS, 2013).	30
1.6	Figure 1.6: Age specific mortality rates for deaths mentioning MRSA by sex and age group between 2008 and 2012, in England and Wales; rates per million population. The data was adapted from the Office for National Statistics (ONS, 2013).	31
1.7	Figure 1.7: A graph representing the annual MRSA counts from April 2007 to March 2016. The data was adapted from the Health Protection Agency which is now a part of the Public Health England (HPA-1, 2016).	32
2.1	An example of the E-Gel® SizeSelect™ gel that was used.	56

2.2	Layout of the IonOneTouch™ 2 Instrument. Image taken from the manufacturer manual.	58
2.3	An enhanced view of the IonOneTouch™ 2 built in centrifuge with the replaceable recovery tubes (x2) and recovery router (x1) shown <i>in situ</i> . Image taken from the manufacturer manual.	59
2.4	The disposable tubing in the pinch valve with the handle on the heat block closed. Image taken from the manufacturer manual.	59
2.5	The disposable injector being inserted into the injector hub at the base of the router. Image taken from the manufacturer manual.	60
2.6	An example of the Ion PGM OneTouch Plus Reaction Filter Assembly that was used. Image taken from the manufacturer manual.	61
2.7	A representation of how the Ion PGM OneTouch Plus Reaction Filter was inverted, concluding with the 3 ports of the filter facing down. Image taken from the manufacturer manual.	62
2.8	Figure 2.8: A representation of the 8-well strip that was used.	63
2.9	The layout of the Ion PGM™ instrument with the reagent and wash bottle. Image taken from the manufacturer manual.	65
2.10	A picture representing the - Ion 316™ Chip v2 that was used. Image taken from the manufacturer manual.	68
2.11	The CLC Genomics Workbench platform with the novel MRSA genomes uploaded.	70
2.12	An output of the the NCTC 8325 reference strain that has been mapped to a gene (SAOUHSC_02861) and the consensus sequence of the MRSA OP100 isolates displayed.	71
2.13	CLC Genomics workbench software showing the uploaded virulence factor genes.	72
2.14	The search box of the software package with the ‘alpha-hemolysin’ gene as the query gene, and an output of the annotated consensus sequence that corresponds to this gene. This consensus sequence can be extracted for further analysis.	73
2.15	A dendrogram showing the <i>Staphylococcus aureus</i> genomes that have been mapped.	75
2.16	An organism overview showing the Genome assembly and annotation report of <i>Staphylococcus aureus subsp. aureus</i> ED133.	76

2.17	A results page output of the GenBank® database showing the presence of the alpha-hemolysin gene in both <i>Staphylococcus aureus</i> and <i>Cyanidioschyzon merolae</i> species.	78
2.18	An output of the alpha-hemolysin (SAOUHSC-01121) gene that is present in <i>Staphylococcus aureus subsp. aureus</i> NCTC 8325.	78
2.19	An alignment of six <i>Staphylococcus aureus</i> strains (04-02981, OBA02176, 11819-17, 502A, 55-2053 and 6850) generated with Mauve 2.1 using the progressive Mauve algorithm.	80
2.20	An output showing the Mauve toolbar with the various tool available.	80
2.21	An output showing Mauve search tools interface that will also display the corresponding position of the genome.	81
2.22	An output showing the software's functionality to zoom in all the way down to the base level of the DNA.	81
2.23a	The Complete genome of <i>Staphylococcus aureus subsp. aureus</i> ED133 and the virulence factor gene of alpha-hemolysin, also from <i>Staphylococcus aureus</i> was acquired through GenBank.	83
2.23b	After hitting 'BLAST', the description box is presented which highlights the region of the <i>Staphylococcus aureus subsp. aureus</i> ED133 that corresponds to the alpha-hemolysin gene.	83
2.23c	This box is then presented which shows the query sequence (alpha-hemolysin) aligned with the corresponding section of the subject strain (<i>S. aureus</i> ED133) genome sequence.	84
2.23d	The region of the query genome was then entered ('1184334 - 1183375') into the 'selected region' view. The FASTA output shown is the region of the ED133 genome that coincides with the coding region of the alpha-hemolysin gene.	84
2.24	An output showing the home screen of MEGA (version 6.0.6).	85
2.25	An output of MEGA showing the options needed to be selected for a DNA alignment.	86
2.26	An output showing the Alignment Explorer screen with the aligned <i>S. aureus</i> sequences.	86
2.27	An output of the MEGA home screen with the aligned sequences now ready to be analysed phylogenetically.	87
2.28	An output showing the Analysis Preferences in order to construct a Neighbour Joining Tree.	88
2.29	An output of the progress screen.	88

2.30	An output showing the Analysis Preferences in order to construct a Maximum Parsimony Tree.	89
2.31	An output of the Tree Explorer window which shows that the alpha-hemolysin gene has produced 97 trees.	90
2.32	An output showing the Tree Options windows to compute a consensus tree.	90
2.33	An output of the Analysis preference for finding the best DNA Model for a particular virulence factor.	91
2.34	An output of a table showing the Maximum likelihood fits of 24 different nucleotide substitutions models.	92
2.35	An output showing the Analysis Preferences in order to construct a Maximum Likelihood Tree.	92
2.36	An output showing the MUSCLE tool homepage with the alpha-hemolysin virulence factor pasted in FASTA format.	94
2.37	An output of the aligned sequence results shown on the MUSCLE homepage.	95
2.38	An output showing a set of alpha-hemolysin FASTA sequences pasted in to the phylogeny.fr query box.	96
2.39a	An output of the DnaSP program showing the home screen.	98
2.39b	An output of the Data Information for this virulence factor.	98
2.39c	An output showing the Nucleotide Sequence Format for alpha-haemolysin.	99
2.39d	An output showing the 'Protein Coding Regions' that need to be selected for this data set.	100
2.39e	An output showing that the 'Current Genetic Code' option need to be set to 'Nuclear Universal'.	100
2.39f	An output showing the updated information of the current dataset (alpha-hemolysin).	101
2.40	An overview screen showing the Polymorphism Data for the dataset (alpha-hemolysin).	102
2.41	An overview screen showing the Polymorphic Sites for the dataset (alpha-hemolysin).	103
2.42	An overview screen showing the options that need to be selected to generate synonymous and non-synonymous substitutions.	104

2.43	The output file generated in DnaSP showing the synonymous and non-synonymous mutations for this dataset (alpha-hemolysin).	104
2.44	An overview showing the 'define data set' screen with four data sets; all sequences, hospital associated, community associated and livestock associated.	105
2.45	An overview screen showing the options that need to be selected to generate the DNA divergence data.	106
2.46	An overview screen showing the results for the DNA divergence between populations data.	107
2.47	An overview screen showing the options that need to be selected to generate the Haplotype data.	107
2.48	An overview screen showing the results for the Haplotype data.	108
2.49	An overview screen of BioEDIT showing the sequence alignment file for alpha-hemolysin.	109
2.50	An example of a NEXUS file showing the haplotype data for alpha-hemolysin.	110
2.51	An overview screen showing the edited NEXUS file for alpha-hemolysin.	111
2.52	An overview screen showing the PopART home screen with the toolbar and loaded NEXUS files.	112
3.1	An output of MAUVE representing the 47-published Methicillin-Resistant <i>S. aureus</i> genomes from Table 3.1.	132
3.2	An output of MAUVE representing the 22-published Hospital-associated <i>S. aureus</i> Methicillin-Resistant genomes taken from Table 3.1.	133
3.3	An output of MAUVE representing the 16-published Community-associated <i>S. aureus</i> genomes from Table 3.1.	134
3.4	An output of MAUVE representing the 7-published Livestock-associated <i>S. aureus</i> genomes from Table 3.1.	135
3.5	An output of MAUVE comparing a Hospital-associated (MRSA252), Community-associated (USA300_FPR3757) and Livestock-associated (ST3980) <i>S. aureus</i> strains.	135
3.6	An image of the gel electrophoresis results, the bands from left to right show; the ladder, MRSA OP100, B51, IP70, B53, S62 and AS3. The gel was run in a FlowGen Gel Electrophoresis system at a voltage of 200V for 3 hours and viewed using a Molecular Imager® Gel Doc™ XR System.	142

3.7a	An electropherogram of the high sensitivity ladder, showing the lower and upper markers.	139
3.7b-g	The results show the following: OP100 peaks at 481 bp; B51 peaks at 479 bp; IP70 peaks at 483 bp; B53 peaks at 478 bp; S62 peaks at 481 bp and AS3 peaks at 488 bp.	144
3.8a	Electropherogram showing the fluorescence intensity (FU) versus the size/migration time for the ladder markers.	148
3.8b-g	Figures 3.8b-g: Shows the fluorescence intensity (FU) versus the size/migration time for the six samples.	148
3.9a	A representation of the loading density for pool two which includes MRSA OP100, B51 & IP70, on to the Ion 316™ v2 Chip. This heat map shows that the analysis has achieved an average loading density of 71%. As shown on the indicator bar; red shows a good loading density and blue indicates a poor loading density on the specific area of the chip.	149
3.9b	A representation of the loading density for pool two which includes MRSA B53, S62 & AS3, on to the Ion 316™ v2 Chip. This heat map shows that the analysis has achieved an average loading density of 72%. As shown on the indicator bar; red shows a good loading density and blue indicates a poor loading density on the specific area of the chip.	150
3.10	A line graph representing the the occurrence of virulence factor genes in the <i>S. aureus</i> strains. The list of virulence factor genes is on the \mathcal{X} axis and the frequency representing the number of times the gene is present in an organism is on the \mathcal{y} axis.	157
3.11	A clustered column bar graph representing the frequency of virulence factors amongst the hospital-associated, community-associated and livestock-associated <i>S. aureus</i> strains.	162
4.1a	A phylogenetic tree for the virulence factor alpha-hemolysin (<i>hla</i>) produced using phylogeny.fr.	176
4.1b	A maximum-likelihood tree for the virulence factor alpha-hemolysin (<i>hla</i>) produced using MEGA.	176
4.2	A bar graph representing the mean nodal support values (achieved from 500 bootstrap replications) from maximum-likelihood trees that were produced for the <i>S. aureus</i> virulence factor genes. The error bars represent the standard deviation of the nodal support values.	179
4.3	A Maximum Likelihood phylogenetic analysis, based on the General Time Reversible (GTR-G) evolution model. The relationship of hospital-associated, community-associated and livestock-associated <i>S. aureus</i> strains for the <i>PBP2</i> gene which encodes Penicillin-binding protein 2. The numbers at the nodes are	

	bootstrap percent probabilities values based on 500 replications. Coloured boxes are consistent with strain classifications.	182
4.4	A Maximum Likelihood phylogenetic analysis, based on the Tamura Nei (TN93+G) model of evolution. The relationship of hospital-associated, community-associated and livestock-associated <i>S. aureus</i> strains for the <i>lukF-PV</i> gene which encodes Panton-Valentine leucocidin subunit F. The numbers at the nodes are bootstrap percent probabilities values based on 500 replications. Coloured boxes are consistent with strain classifications.	183
4.5	Figure 4.5: A Maximum Likelihood phylogenetic analysis, based on the Tamura 3-parameter (T92). The relationship of hospital-associated, community-associated and livestock-associated <i>S. aureus</i> strains for the <i>vwb</i> gene which encodes von Willebrand factor-binding protein. The numbers at the nodes are bootstrap percent probabilities values based on 500 replications Coloured boxes are consistent with strain classifications.	184
4.6	A Maximum Likelihood phylogenetic, based on the Hasegawa-Kisaino-Yano parameter (HKY). The relationship of hospital-associated, community-associated and livestock-associated <i>S. aureus</i> strains for the <i>seb</i> gene which encodes enterotoxin B. The numbers at the nodes are bootstrap percent probabilities values based on 500 replications Coloured boxes are consistent with strain classifications.	185
4.7	A Maximum Likelihood phylogenetic analysis, based on the General Time Reversible (GTR-G) parameter. The relationship of hospital-associated, community-associated and livestock-associated <i>S. aureus</i> strains for the <i>clf</i> gene which encodes clumping factor. The numbers at the nodes are bootstrap percent probabilities values based on 500 replications. Coloured boxes are consistent with strain classifications.	186
5.1	A graph representing the nucleotide diversity (π) of the <i>S. aureus</i> virulence factor genes. The error bars represent the standard deviation of the nucleotide diversity. Data was generated using DnaSP. The complete data generates is shown in Appendix 1.	196
5.2a	A graph showing the dN/dS ratio (ω) of all of the virulence factor genes in all of the strains. Complete data is available in Appendix 1.	201
5.2b	A graph showing the dN/dS ratio (ω) of the virulence factor that are present in the livestock-associated strains. Complete data is available in Appendix 1.	202
5.3a	A regression analysis showing the observed nucleotide diversity (π) vs the length of the virulence factor gene. The trend line is shown. (Slope: 0.000024; Intercept 0.017501; r^2 0.292441; DoF 42; $P=0.000151$ therefore $P<0.001$).	204
5.3b	A regression analysis showing the observed dN/dS ratio (ω) vs the length of the virulence factor gene. The trend line is shown. (Slope: -0.000127; Intercept 2.271417; r^2 0.003483; DoF 42; $P=0.704$).	204

5.4a	A regression analysis showing the observed nucleotide diversity (π) vs the length of the virulence factor gene in livestock-associated strains. The trend line is shown. (Slope: 0.00003; Intercept 0.018278; r^2 0.248114; DoF 34; $P=0.001991$ therefore $P<0.001$).	206
5.4b	A regression analysis showing the observed nucleotide diversity (π) vs the length of the virulence factor gene in livestock-associated strains. The trend line is shown. (Slope: 0.000414; Intercept 0.951858; r^2 0.061562; DoF 34; $P=0.145$).	206
5.5	A graph representing the pairwise nucleotide differences observed between two populations (D_x & D_y). This is used to quantify divergence within the populations.	208
6.1a-c	Figure 6.1a is a graph representing the relationship between the number of variable sites and nucleotide diversity per site of the <i>S. aureus</i> population. Figure 6.1b is a graph representing the relationship between the number of haplotypes and the nucleotide diversity. Figure 6.1c is a graph representing the relationship between the number of variable sites and haplotype diversity per site of the <i>S. aureus</i> populations.	219
6.2a-b	Are haplotype network representing the <i>SCIN</i> & <i>PBP2</i> virulence factor genes, created using the TCS methodology. The areas of the circle are proportionate to the number of individual strains harbouring the specific haplotype and the different colours represent the country of origin of the isolate. The numbers in brackets represent the mutational steps between the haplotypes and the small black circles indicate missing haplotypes.	222
6.3a	A haplotype network that was created for the Staphylocoagulase (<i>coa</i>) virulence factor gene, created using the TCS methodology. The areas of the circle are proportionate to the number of individual strains harbouring the specific haplotype and the different colours represent the country of origin of the isolate. The numbers in brackets represent the mutational steps between the haplotypes and the small black circles indicate missing haplotypes.	224
6.3b	A haplotype network that was created for the Thermonuclease (<i>nucA</i>) virulence factor gene, created using the TCS methodology. The areas of the circle are proportionate to the number of individual strains harbouring the specific haplotype and the different colours represent the country of origin of the isolate. The numbers in brackets represent the mutational steps between the haplotypes and the small black circles indicate missing haplotypes.	225
6.3c	A haplotype network that was created for the Thermonuclease (<i>nucA</i>) virulence factor gene, created using the TCS methodology. The areas of the circle are proportionate to the number of individual strains harbouring the specific haplotype and the different colours represent the country of origin of the isolate. The numbers in brackets represent the mutational steps between the haplotypes and the small black circles indicate missing haplotypes.	225

- 6.3d A haplotype network that was created for the enterotoxin B (*seb*) virulence factor gene, created using the TCS methodology. The areas of the circle are proportionate to the number of individual strains harbouring the specific haplotype and the different colours represent the country of origin of the isolate. The numbers in brackets represent the mutational steps between the haplotypes and the small black circles indicate missing haplotypes. 226
- 6.3e Figures 6.3e: A haplotype network that was created for the Panton-Valentine toxin, subunit F (*lukF-PV*) virulence factor gene, created using the TCS methodology. The areas of the circle are proportionate to the number of individual strains harbouring the specific haplotype and the different colours represent the country of origin of the isolate. The numbers in brackets represent the mutational steps between the haplotypes and the small black circles indicate missing haplotypes. 226
- 6.4a-e Sliding window analysis showing the ratio of synonymous and nonsynonymous polymorphism observed in a number of *S. aureus* virulence factor genes: a) *lukF-PV*, b) *nucA*, c) *vwb*, d) *coa* and d) *seb*. The nonsynonymous mutations are highlighted in red and the synonymous or silent mutations are highlighted in blue. 229
- 6.5a-e Pairwise distributio analysis for a number of *S. aureus* virulence factor genes: a) *lukF-PV*, b) *nucA*, c) *vwb*, d) *coa* and d) *seb*. All analysis are shoing a multimodal distributon and as such are not significant. 231

List of abbreviations

α	Alpha
β	Beta
BHI	Brain-heart infusion agar
bp	Base pairs
BLAST	Basic Local Alignment Search Tool
CA	Community associated
CDS	Coding sequence
CHIPS	Chemotaxis Inhibitory Protein
dN	Non-synonymous
dS	Synonymous
DNA	Deoxyribonucleic acid
Dxy	Divergence
Eap	Extracellular adherence protein
EASRS-Net	European Antimicrobial Resistance Surveillance Network
ECM	Extracellular matrix
Efb	Extracellular fibrinogen-binding protein
Fg	Fibrinogen
g	Gram
gDNA	Genomic DNA
<i>h</i>	Haplotype diversity
HA	Hospital associated
HAI	Healthcare associated infection
Hd	Haplotype diversity
HLA	Alpha hemolysin
HLB	Beta-hemolysin
ICU	Intensive care unit
Indel	Insertions or deletions
LA	Livestock associated
LCB	Locally Collinear Blocks
MEGA	Molecular evolutionary genetics analysis
MGE	Mobile genetic element

mg	Milligram
ml	Minute
ML	Maximum-likelihood tree
MRSA	Methicillin-resistant <i>S. aureus</i>
MSSA	Methicillin-sensitive <i>S. aureus</i>
MSCRAMM	Microbial surface components recognising adhesive matrix molecules
MP	Maximum-parsimony tree
N	No. of Sequences
NAG	N-acetylglucosamine
NAM	N-acetylmuramic acid
NCBI	National centre for biotechnology information
NJ	Neighbour-joining tree
ONS	The Office of National Statistics
PBP	Penicillin binding proteins
PGM	Personal genome machine
PHE	Public Health England
PVL	Panton-Valentine leuckocidin
PCR	Polymerase Chain Reaction
RNA	Ribonucleic acid
s	Polymorphic or variable sites
SaPI	Staphylococcal pathogenicity island
SCC	Staphylococcal Cassette Chromosome
SCI	Staphylococcal Complement Inhibitor
SCV	Small colony variants
SNP	Single nucleotide polymorphism
SPI	<i>Salmonella</i> pathogenicity islands
SSSS	Staphylococcal Scalded skin syndrome
SSTI	Skin and Soft Tissue Infections
TSS	toxic shock syndrome
UTIs	Urinary tract infections
V	Volt
VISA	Vancomycin-intermediate <i>S. aureus</i>
VRSA	Vancomycin-resistant <i>S. aureus</i>

vWbp	von Willebrand factor-binding protein
WGS	Whole-genome sequencing
ω	Relative rate of non-synonymous to synonymous substitutions
π	Nucleotide diversity per site
μ	Microlitre

1.0 Introduction

1.1 Discovery of *Staphylococcus aureus*

In the late 1860's, a British surgeon influenced by Louis Pasteur's work on infectious diseases, promoted the idea of sterile surgery while working at the Glasgow Royal Infirmary. Joseph Lister successfully introduced carbolic acid, now known as phenol, to sterilise surgical instruments and clean wounds which directly led to the reduction of wound infections and thus making surgery safer (Smith, 2012; Lister, 1867). Joseph Lister observed, in 1867, that when he applied his antiseptic techniques towards the general cleanliness of the hospital wards where he was working at the time, not even a single case of pyraemia, hospital gangrene or erysipelas (skin infection) occurred during the 9 months of this trial. Previous to this, the hospital ward was rife with infections. This helped Lister show the importance of using carbolic acid as an antiseptic agent and for the general cleanliness of the hospital to his peers (Lister, 1867). In 1878, Staphylococci were seen in pus by Robert Koch and they were then cultured by Pasteur in 1880. However, it was Alexander Ogston who made the initial detailed studies. Ogston, A Scottish surgeon who was working at the the Aberdeen Royal Infirmary opened the wound of one of his patients and drew some pus and examined it under the microscope (Newsom, 2008).

“My delight may be conceived when there were revealed to me beautiful tangles, tufts and chains of round organisms in great numbers, which stood out clear and distinct among the pus cells and debris, all stained with aniline-violet solution ...” (Wilson, 1987; Newsom, 2008).

Ogston then decided to see if abscess would occur in the testing animals if he injected pus from acute abscess into guinea pigs, wild mice and white mice. The animals developed similar pathological features, a new abscess was formed, followed by signs of septicaemia (Wilson, 1987; Newsom, 2008).

“The animals refused food, sat cowering in a retired place in their case, were listless and apathetic, their coat was disordered and sometimes wet, their eyes were kept closed save when startled, and the mice showed purulent conjunctivitis and gluing together of the eyelids described by Koch in his experiments on septicaemia”
(Wilson, 1987; Newsom, 2008).

If, however, he pre-treated the pus with carbolic acid or heat before injecting it into the animals, abscess

formation was avoided and signs of infection were not present. Ogston named the clustered micrococci “staphylococci” from the Greek word “staphyle”, which had the meaning; bunch of grapes (Newsom, 2008; Ogston, 1881). A short time later, in 1884, Rosenbach, a German scientist, isolated two strains of staphylococci, which he named for the pigmented colour of their colonies; *Staphylococcus aureus*, from the Latin ‘aurum’ for gold and *Staphylococcus albus* (now called *Staphylococcus epidermis*) from the Latin ‘albus’, meaning white (Cowan *et al.*, 1954; Wilson, 1987).

1.2 Staphylococci

Staphylococcus aureus is a Gram-positive bacterium, member of the Firmicutes phylum (Frank *et al.*, 2010). To date, 36 species and 9 subspecies of the genus of *Staphylococcus* have been identified (Kazmierczak *et al.*, 2014). These bacteria are non-motile, non-spore forming facultative anaerobes (Plata *et al.*, 2009) with diameters of approximately 1 µm. When observed under the microscope they resemble grape-like-clusters, since cell division takes place in more than one plane (Goss & Muhlebach, 2011; Plata *et al.*, 2009; Pinchuck *et al.*, 2012;). *S. aureus* is also catalase-positive, oxidase negative and require complex nutritional requirements, such as many acids and vitamins for growth. *S. aureus* is very tolerant of high concentrations of salt, up to 1.7 molar (Plata *et al.*, 2009). Coagulase is an enzyme produced by several bacteria that enables the conversion of fibrinogen to fibrin and is used to distinguish between different types of *Staphylococcus* isolates. The majority of these staphylococci are coagulase-negative, with the exception of: *S. aureus*, *Staphylococcus intermedius*, *Staphylococci delphini*, *Staphylococcus schleiferi* subs. *coagulans* and some strains of *Staphylococcus hyicus* (Kwok *et al.*, 1999).

S. aureus is frequently part of the human micro-flora (Sousa & Lenastre, 2004). In a healthy population, approximately 20% of individuals are colonised with *S. aureus* and are thus called persistent carriers. Another 30% of the individuals are intermittently colonised for example the organism may or may not be isolated at different intervals, and the remaining 50% of the population appear not to be susceptible to this bacterium and therefore called non-carriers (Kluytmans *et al.*, 1997, Heidi *et al.*, 2009). In humans, *S. aureus* colonisation occurs most commonly in the nose with a reported 40 - 44% of individuals displaying this trait (Kluytmans *et al.*, 1997, Williams, 1963), more specifically arising in the anterior nares (the nostrils). This has shown to be the case because there are no human defences present to resist against the bacterium before they colonise (Heidi *et al.*, 2009). It has been observed that when the nares are treated for the carriage of *S. aureus*, the organism will also disappear from other parts of the body (Kluytmans *et al.*, 1997). Other body sites such as the perineum and throat may also be colonised (Heidi *et al.*, 2009). If this bacterium is present anywhere else in the body such as the gastrointestinal tract or gains entry via an open wound or catheterisation, it is known as an

opportunistic pathogen and can cause disease especially if the immune system is compromised (Sousa & Lencastre, 2004).

1.3 Treatment of *Staphylococcus aureus* Infections

1.3.1 Pre-antibiotic era

In the pre-antibiotic era, patients with *S. aureus* infection were associated with a mortality rate that exceeded 80% with over 70% developing metastatic infections (Lowy, 2003). This situation changed markedly after the introduction of Penicillin in the early 1940s and consequently improved the prognosis of patients with *S. aureus* infections. However, it was as early as 1942 that *S. aureus* strains resistant to Penicillin had been detected in St Mary's Hospital in London and subsequently in the community (Lowy, 2003).

1.3.2 The era of antibiotics

If *S. aureus* is to be treated effectively, appropriate antibiotics must be administered. In 1928, Alexander Fleming observed that when a culture plate of bacteria became contaminated with *Penicillium notatum* mould, it was noted that this fungus appeared to inhibit the growth of bacteria; the mould was also non-toxic when it was injected into mice and rabbits (Fleming, 1929; Rolinson, 1998). It was not until 1940 that this antibacterial substance was isolated in pure culture by Florey and Chain and their colleagues in Oxford, UK and subsequently named Penicillin (Rolinson, 1998). A short time later, it was tested on mice that were infected with a lethal streptococcal infection; Penicillin was injected subcutaneously and was considered a highly effective treatment (Rolinson, 1998). Penicillin was consequently considered a bactericidal agent that now belongs to a large group of antibiotics known as beta-lactams (β -lactams); these antibiotics have been used to treat most Gram-positive bacterial infections, including *S. aureus* and some Gram-negative infections (Martin *et al.*, 2010). Other members of the β -lactam antibiotic family include Cephalosporins, Carbapenems and Monobactams that are characterised by a common β -lactam ring (James & Gurk-Turner, 2001). Penicillins contain a bicyclic nucleus, which includes the β -lactam ring and a thiazolidine ring; in Cephalosporins the thiazolidine ring is replaced by a dihydrothiazine ring; Carbapenems also contain a bicyclic nucleus with a β -lactam ring; the only exception is the Monobactam

antibiotic, as it contains a β -lactam ring but no bicyclic ring structure (James & Gurk-Turner, 2001).

1.3.3 Structure of the bacterial cell wall

β -lactams target the proteins involved in the synthesis of the cell wall (Martin *et al.*, 2010). Peptidoglycan is a macromolecule present in the cell wall of Gram-positive bacteria, such as *S. aureus* and is essential for maintaining the size, shape, mechanical and structural integrity of the bacteria (Wilke *et al.*, 2005). This macromolecule is composed of glycan chains cross-linked with peptides (Wilke *et al.*, 2005; Meroueh *et al.*, 2006). The glycan portion contains alternating units of N- acetylmuramic acid (NAM) and N-acetylglucosamine (NAG), with the latter sugar in this disaccharide being modified by a characteristic pentapeptide which varies amongst the Gram-positive and Gram-negative bacterium but always terminates in two D-alanine residues (Wilke *et al.*, 2005; Meroueh *et al.*, 2006). The peptides from adjacent glycan strands are cross linked and form the characteristic structure of peptidoglycan; this reaction is called transglycolation and is catalysed by transpeptidase enzymes, also called penicillin binding proteins (PBPs) (Wilke *et al.*, 2005). Although the individual peptidoglycan units are produced inside the cell, transglycosylation occurs outside the cytoplasmic membrane (Wilke *et al.*, 2005).

1.3.4 Penicillin resistance

Resistance of Staphylococcal infections to Penicillin is mediated by the *blaZ* genes that encode β -lactamase, an enzyme that is synthesised when the bacteria are exposed to β -lactam antibiotics (Figure 1.1a). This enzyme hydrolyses the β -lactam ring of the antibiotic, rendering the drug inactive (Zhang *et al.*, 2001; Lowey, 2003). Two adjacent regulatory genes; *blaR1*, (the anti-repressor) and *blaI* (the repressor) control the *blaZ* gene. Signalling pathways responsible for β -lactamase synthesis requires sequential cleavage of the regulatory proteins; BlaR1 and BlaI (Zhang *et al.*, 2001; Lowey, 2003). After exposure to β -lactam antibiotics, BlaR1, a transmembrane sensory transducer cleaves itself, this cleaved protein then functions as a protease that cleaves the repressor BlaI and allows expression of the *blaZ* genes and synthesis of the β -lactamase (Gregory *et al.*, 1997; Zhang *et al.*, 2001).

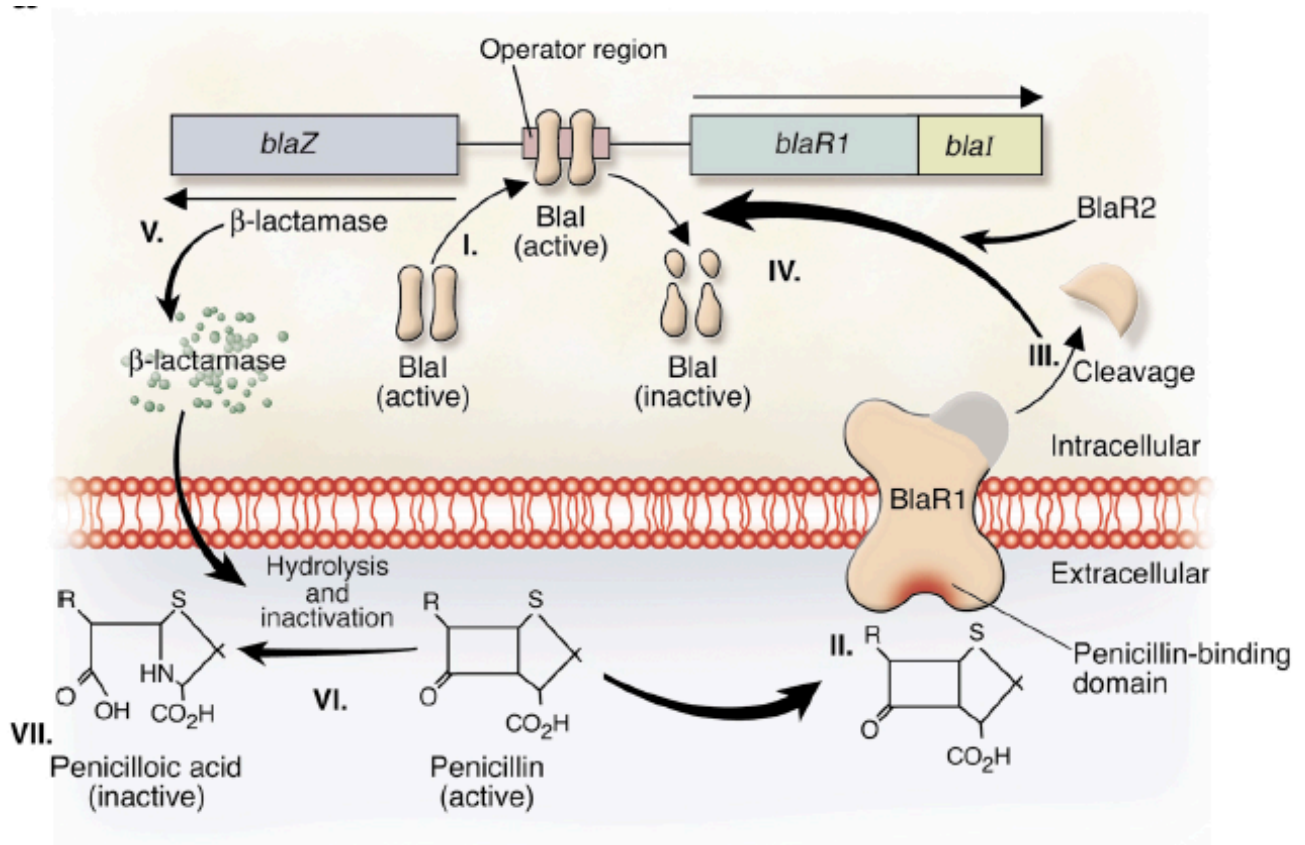


Figure 1.1a: Taken from Lowey., 2003. (I) B1aI, a DNA binding protein binds to the operator region which suppresses RNA transcription from both *blaZ* and *blaR1*; in the absence of penicillin, β -lactamase I suppressed at lower levels. (II) The binding of penicillin to the transmembrane sensor-transducer B1aR1 stimulates B1aR1 autocatalytic activation. (III – IV) Active B1aR1 cleaves (via second protein, BlaR2) B1aI into inactive fragments allowing the transcription of the *blaZ* genes to commence. These genes allow the synthesis of β -lactamase to occur. (V) β -lactamase hydrolyses the β -lactam ring of penicillin. (VI) it therefore renders the penicillin inactive and penicillin resistance occurs (Lowey., 2003).

1.3.5 Methicillin resistance

Methicillin, a semi-synthetic derivative of Penicillin, was introduced in 1959 and overcame the problem of Penicillin-resistance (Taubes, 2008). Normally, β -lactam antibiotics such as methicillin successfully inhibit the bacterial transpeptidase enzymes, also called PBPs by mimicking the D-alanine-D-alanine substrate to bind to the active site of the transpeptidase and then react covalently with the enzyme (Wilke *et al.*, 2005). During this time this enzyme will not be able to carry out its normal function of peptidoglycan assembly (Wilke *et al.*, 2005; Fisher & Mobashery, 2010). This results in weakly cross-linked peptidoglycan, which hinders the mechanical strength of the bacterial cell wall, eventually weakening it and causing the cell to become deformed and

consequently lyse (Wilke *et al.*, 2005; Fisher & Mobashery, 2010). However, MRSA was subsequently observed in 1961 with some reports of resistance reported within a few months of the introduction of Methicillin in hospitals located in England (Jevons, 1961; Taubes, 2008). Since then, it has become an important pathogen in both health-care setting and communities around the world (Taubes, 2008). A study performed by Cosgrove *et al.*, (2003) between 1980 and 2000 concluded that the therapeutic outcomes of infection due to MRSA are far worse than the outcome of those that result from methicillin-sensitive strains (MSSA) and that MRSA bacteraemia is associated with a higher mortality rate. Normally, *S. aureus* has the ability to produce four penicillin-binding proteins (PBPs) involved in cell wall synthesis, in addition to these PBPs, a novel PBP was discovered in MRSA termed PBP2a which is encoded by the presence of the methicillin resistance gene (*mecA*) (Weilders *et al.*, 2001). As discussed earlier, PBPs are enzymes that catalyse the transpeptidase reaction, required for the formation of the cell wall (Wilke *et al.*, 2005). PBP2a substitutes for other PBPs because of its reduced binding affinity for β -lactam antibiotics; this enables the bacteria to survive exposure to high concentrations of these agents and therefore resistance to methicillin occurs (Rice, 2012). The *mecA* gene, present in all strains of MRSA, is carried on a 21 – to – 60-kb DNA cassette; the Staphylococcal Cassette Chromosome (*SCCmec*) and is required for methicillin resistance to occur (Ito & Hiramatsu, 1998; Katayama, Ito & Hiramatsu, 2000). The *SCCmec*, is a large genetic mobile element that differs in genetic composition and also contains genetic structures such as *Tn554*, pUB110 and pT181 which all encode resistance to non- β -lactam antibiotics (Deurenberg *et al.*, 2006; Ito & Hiramatsu, 1998). For instance, pT181 codes for tetracycline resistance, and *Tn554* is responsible for lincosamide and streptogramin resistance (Deurenberg *et al.*, 2006). Kreiswirth *et al* (1993) proposed that all MRSA strains were descended from a single *S. aureus* strain, however; more recent studies by Fitzgerald *et al* (2001), show that the *mecA* gene has been horizontally transferred into *S. aureus* numerous times, concluding that methicillin resistant strains have evolved from multiple independent strains and not just one single ancestral strain (Fitzgerald *et al.*, 2001).

1.3.6 Treatment options for *S. aureus* infections

The choice of treatment against MRSA has been limited, with the antibacterial therapy of choice being

the antibiotic vancomycin, a glycopeptide (Smith *et al.*, 1999). MRSA strains with reduced susceptibility to vancomycin have been reported in clinical specimens since the late 1990s; this is possibly due to the misuse and/ or the over-prescription of this antibiotic which has resulted in Vancomycin-resistant *S. aureus* (VRSA) (Appelbaum, 2006). This is a real concern, if VRSA becomes prevalent globally, and new, effective antibiotics are not discovered to treat this threat, it could result in a high mortality rate and it threatens to return the world to the pre-antibiotic area. Administration of vancomycin should therefore be restricted and methods for detecting resistant bacteria must be improved (Smith *et al.*, 1999; Chang *et al.*, 2003).

A recent report to determine the global priority list of antibiotic resistant bacteria to guide research, discovery and development of new antibiotics by was published by The World Health Organisation (WHO) (WHO-2., 2017). This report categorises pathogens as either Priority 1, stating that there is a critical need; priority 2, stating that there is a high need; or priority 3, stating that there is a medium need. This report determined that *S. aureus* that are either Methicillin-resistant or Vancomycin-intermediate resistant are a priority 2 (WHO-2., 2017).

In 2015, a new class of antibiotic was discovered called Teixobactin (Ventola., 2015). This antibiotic is produced by *Eleftheria terrae*, a previously impossible-to-culture microbe; but with new techniques this was made possible (Ventola., 2015). Figure 1.1 is an overview flow diagram of some of the antibiotics used to treat *S. aureus* infections taken from the WHO website (http://www.who.int/patientsafety/events/05/apsd/EURO_Initiatives_on_blood_safety_Hafner.pdf) (WHO-1, 2012). It shows when *S. aureus* became resistant to penicillin, methicillin, the emergence of vancomycin intermediate strains (VISA) and finally the vancomycin resistant strains (VRSA).

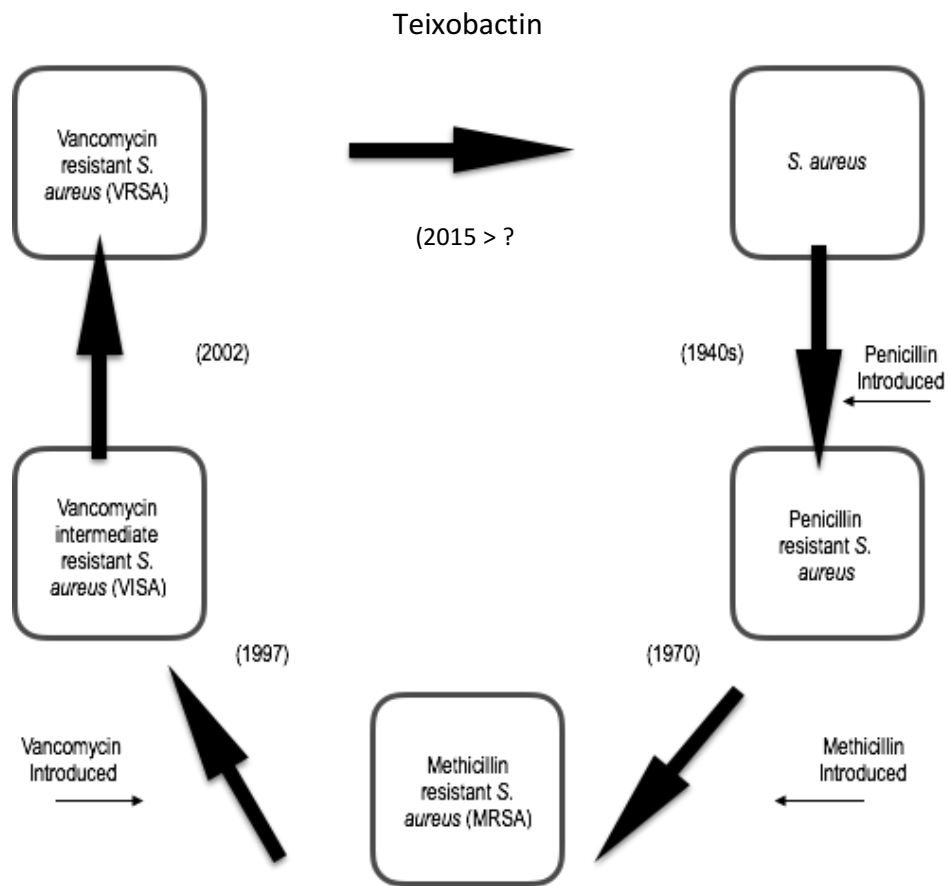
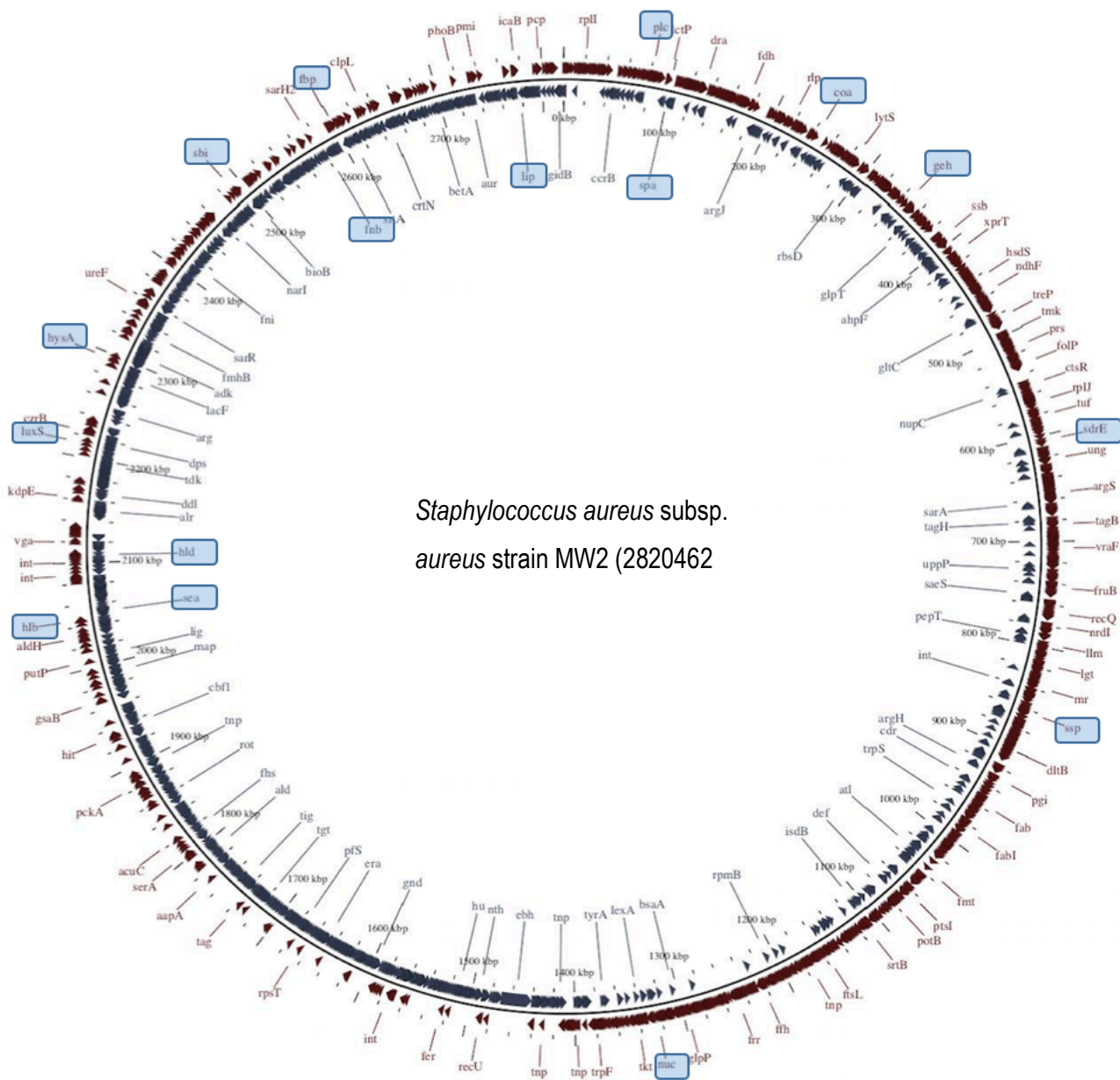


Figure 1.1b: A flow cycle for the evolution of drug resistance in *S. aureus* (Adapted from WHO-1, 2012). This diagram shows that every time that a new antibiotic is discovered to combat the *S. aureus*; drug resistance will follow. Notably, Teixobactin is the first of a new class of antibiotic that has been discovered, in January 2015. The question mark for '2015 >?' represents that it takes years for new drugs to be developed, but only a short amount of time for antibiotic resistance to be detected (Ventola., 2015).

1.4 Staphylococcus aureus genome

The first genomes of *S. aureus* to be published were strains N315 and Mu50 in 2001 (Baba *et al.*, 2002). Figure 1.2 shows the genome of *S. aureus* MW2, a highly virulent MRSA that was linked to four paediatric deaths in Minnesota and North Dakota, in 1998; the genome of MW2 is 2820462 base pairs (bp) long (Baba *et al.*, 2002). GView was used by the author to produce Figure 1.2; it is a software package that has the capabilities to pan-and-zoom bacterial genomes (Petaku *et al.*, 2010). This figure also only shows a snapshot of the genes that are present; the more you zoom, the greater the content that you can see.

In order for bacteria such as *S. aureus* to become pathogens, they must have the ability to accomplish infection and consequently cause disease in the host. This is known as pathogenicity and virulence is a degree of the pathogenicity observed. Virulence factors are encoded by genes of the pathogen which contribute to the pathogenicity. The locations of numerous genes including the virulence factor genes of interest are shown on the genetic map (Figure 1.2) of *S. aureus* MW2, highlighted in blue.



Staphylococcus aureus subsp.
aureus strain MW2 (2820462)

Positive strand
Negative strand

Figure 1.2. genome of This genome pairs long.

This map was me using GV Genomic V Stuart-Edward Domselaar, 2

The location genes including genes are shown positive strand of the genes shown investigated instead the genes blue were investigated with others represented

It is important only chromosomes investigated in

1.5 Virulence Factors

In this study, bioinformatics methods will be employed to investigate the arsenal of virulence factors produced by *S. aureus* and the role they play, including some of those highlighted in Figure 1.2 will be investigated. The genes needed for methicillin resistance, adhesion factors; needed for host attachment, persistence factors; needed for biofilm accumulation and intracellular persistence, immune evasions factors; needed for evading/ or destroying host defences that include the innate and complement immune systems, enzymes and other secreted proteins; that include a range of functions such as tissue invasion and destruction, cytotoxins; that include the lysis of erythrocytes and neutrophils and enterotoxins; that involve toxin mediated disease or sepsis such as food poisoning, toxic shock syndrome and scalded skin syndrome.

This study only investigated chromosomal virulence factor genes and not plasmid virulence genes. A plasmid is a small, circular, double-stranded DNA molecule that is distinct from a cell's chromosomal DNA (Solar *et al.*, 1998). Plasmids have the potential to carry resistance genes and virulence genes and are spread through *S. aureus* populations by horizontal gene transfer (HGT) mechanisms (McCarthy & Lindsay., 2012).

1.5.1 Virulence associated genes

1.5.1.1 *mecA* gene and Penicillin binding protein 2 (PBP2a)

The *mecA* gene is part of a 21- to 60-kB Staphylococcal Cassette Chromosome (SCC), a mobile genetic mobile element which may differ in genetic composition (Ito & Hiramatsu, 1998). *mecA* is 2010 bp long and PBP2a consists of 670 amino acids (Ito & Hiramatsu, 1998). PBP2a substitutes for other PBPs because of its reduced binding affinity for β -lactam antibiotics and PBP2a differs from other PBPs in that its active site blocks binding of all β -lactam antibiotics but allows the transpeptidation reactions to proceed (Lim & Strynadka, 2002). PBP2a permits transpeptidase

activity to allow cell wall synthesis at concentrations of β -lactams which would otherwise inhibit transpeptidation (Lim & Strynadka, 2002; Rice, 2012).

1.5.2 Adhesion genes

Adhesion is the first step in colonisation and the subsequent infection of the host. In order to initiate infection, adhesins which are bacterial proteins, adhere to components of the host extracellular matrix (ECM). Adherence is mediated by surface protein adhesions called MSCRAMM (microbial surface components recognising adhesive matrix molecules) family, which in most cases are covalently anchored to the cell wall peptidoglycan (Foster & Hook, 1998).

1.5.2.1 Fibronectin-binding protein

Fibronectin binding protein belongs to the MSCRAMMs family of proteins. The majority of *S. aureus* strains express two binding proteins, FnbpA and FnbpB which are encoded by closely linked genes termed *fnbpA* and *fnbpB* (Wann *et al.*, 2000). These closely linked genes and the proteins they encode, facilitate the attachment of *S. aureus* to human mucosal epithelium (Foster & Hook, 1998). Studies by Mongodin *et al* (2002) have shown that FnBPs are involved in the adherence and colonisation to the human airway epithelium with further research by Shinji *et al* (2011) suggesting that fibronectin-binding protein are also important for not only *in vitro* and *in vivo* adhesion to cells but may also have a secondary role of internalisation by cells.

1.5.2.2 Collagen-binding protein

Collagen is targeted by the collagen-binding adhesin which is encoded by the *cna* gene, this adhesin also belongs to the MSCRAMMs family of proteins (Foster & Hook, 1998). *cna* is not present in all strains but where it is present the collagen binding adhesin binds to collagen substrates and

collagenous tissue with a high degree of specificity and affinity to humans which is an important step in pathogenesis of osteomyelitis and septic arthritis (Patti *et al.*, 1992; Rhem *et al.*, 2000).

1.5.2.3 Elastin-binding protein

Elastin is a major component of the ECM and plays a crucial role in maintaining structural integrity of tissues which gives it the property of reversible elasticity, it is present in abundance in the lung, skin and major blood vessels (Nakakido, Yoshikazu & Tsumoto, 2007). Elastin-binding protein which is encoded by the *ebpS* gene has been shown to bind both soluble elastin and tropoelastin and *S. aureus* can therefore adhere to and colonise the host tissue (Downer *et al.*, 2002).

1.5.2.4 Bone-sialoprotein binding protein

Bone-sialoprotein (Bsp) is a glycoprotein of the bone and dentine extracellular matrix (Tung *et al.*, 2000). The identification of *S. aureus* strains that cause osteomyelitis has consequently led to the discovery that they interact with bone-sialoprotein (Tung *et al.*, 1989). *S. aureus* bind to bone-sialoprotein via a bone-sialoprotein-binding protein (Bbp) which belongs to the Sdr family of proteins (Tung *et al.*, 2000; Perrson, Johansson & Ryden, 2009). The Sdr family of proteins, SdrC, SdrD and SdrE are a collection of surface proteins with no ligand specificity that recognise adhesive matrix molecules and are consequently members of the MSCRAMM family that are encoded respectively by the *sdrC*, *sdrD* and *sdrE* genes (Sabat *et al.*, 2006; Vasquez *et al.*, 2011). Bbp, also a member of the MSCRAMM family is a 97-kDa protein initially isolated from *S. aureus* strain 024 and has an A and B domain that is 76% and 96% homologous to SdrE domains (Perrson *et al.*, 2009).

1.5.2.5 Clumping factor

Fibrinogen (Fg) plays a major role in blood clotting, fibrinogenesis, the inflammatory response and wound healing, it is a large 340-kDa plasma glycoprotein composed of two sets of three polypeptides chains A α , B β and γ that are linked by disulphide bonds that are essential for maintaining normal homeostasis (Palma *et al.*, 2001; Mosesson, 2005). Clumping factors A and B (ClfA and ClfB) belong to the bifunctional MACRAMM family of proteins and bind to different sites on fibrinogen; clfA binds to the χ -chain and *clfB* binds to the α -chain (Wertheim *et al.*, 2008). Interactions between the ClfB and the α -chain is also important in platelet activation and aggregation (Xiang *et al.*, 2012).

1.5.2.6 Extracellular fibrinogen-binding protein

Extracellular fibrinogen-binding protein (Efb) is a 15.8-kDa extracellular protein that has two repeated regions and consequently has two binding sites for fibrinogen located at both the N-terminal and C-terminal regions that form a divalent interaction with fibrinogen (Fg) resulting in an Efb-Fg complex (Palma *et al.*, 2001). Efb is usually produced in large quantities during the post exponential phase of bacterial growth and is associated with a more severe infection (Palma *et al.*, 2001). However, studies by Shannon *et al.*, (2006) concluded that immunisation with Efb produces a protective antibody response to Efb from *S. aureus* which is important as it resulted in the vaccinated animals developing less severe infection than the unvaccinated controls.

1.5.2.7 Extracellular adherence protein

Extracellular adherence protein (Eap) is a 72-kDa protein consisting of 689 amino acids, which includes a 110-amino-acid sequence that is repeated six times. Within each domain there is a subdomain that consists of 31 residues that has high homology to the N-terminal β -chain of many major histocompatibility complex (MHC) class II molecule and lacks the LPXTG motif which is

essential for membrane anchoring (Harraghy *et al.*, 2003). Eap has the ability to bind to numerous plasma proteins including fibronectin, fibrinogen and prothrombin (Hagggar *et al.*, 2003). Eap has also been shown to play a crucial role in the complex internalisation process of *S. aureus* into non-professional phagocytes and can also cause agglutination (Hagggar *et al.*, 2003).

1.5.3 Virulence Factors Involved in Persistence

Once attachment has been established by *S. aureus* to the host, other factors that cause relapsing infections are readily available to adapt different environmental niches of the host.

1.5.3.1 Delta-aminolevulinic acid dehydratase

Staphylococcus aureus can be isolated from the chronically infected airways of patients who suffer from cystic fibrosis, and such *S. aureus* isolates are termed small-colony variants (SCV), that are characterised by the slow growing morphological variants associated with persistence and antibiotic resistant infections (Besier *et al.*, 2007). SCV's that are isolated from patients with chronic or relapsing infections often indicate that they are auxotrophic for compounds, such as hemin that are involved in the synthesis of the electron transport chain, suggesting that the SCV phenotype is associated with genetic changes that consequently impair electron transport (Sifri *et al.*, 2006). The *hemB* gene is part of the family of genes encoding enzymes of the porphyrin biosynthetic pathway and specifically codes for the enzyme porphobilinogen synthase. This enzyme is responsible for the conversion of delta-aminoevulinic acid to porphobilinogen of the electron transport system (Kafala & Sasarman, 1994). Studies by von Eiff *et al* (1997) have concluded that a defect in the electron transport system allows for *S. aureus* SCVs to persist intracellularly as well as resist aminoglycoside antibiotics. Stable mutants of the electron transport chain were generated by interrupting the *hemB* gene, and *in vitro* studies showed that the mutant showed typical characteristics of SCVs, which included slow growth and persistence within the cultured endothelial cells (von Eiff *et al.*, 1997).

Studies by Vaudaux *et al.* (2002) have also suggested that there is an increased expression of clumping factor and fibronectin-binding protein by *hemB* mutants.

1.5.3.2 Bacterial efflux pumps

Bacterial efflux pumps are responsible for expelling a broad range of substances such as toxins and antibiotics out of the cell. An over expression of multi drug efflux pumps have also been found to be associated with drug resistance because they lower the intracellular concentration of antibiotics, hindering the effect of the antibiotic (Naruni *et al.*, 2002). A multi drug efflux gene, termed staphylococcal efflux pump gene (*sepA*) was investigated in this study, it is 474 bps and encodes the drug efflux pump that has four predicted transmembrane segments (Narui *et al.*, 2002).

1.5.4 Immune Evasion genes

Once persistence has been achieved by *S.aureus*, the organism has developed methods to achieve host immune evasion.

1.5.4.1 Staphylokinase

Staphylokinase, is secreted by many *S. aureus* strains that infect humans (Kwiecinski *et al.*, 2013). The highly conserved Staphylokinase gene (*sak*) is carried on a bacteriophage that also contains other important genes, such as complementary inhibitory factors and enterotoxins (Peetermans *et al.*, 2014). The *sak* gene is investigated further in the results chapters. Studies have suggested that three natural variants of the *sak* gene have been identified with only 4 nucleotide differences found in the coding regions of these variants (Kim *et al.*, 1997). Staphylokinase is synthesised as a polypeptide precursor of 163 amino acids which includes a 27 amino acid signal peptide in a single polypeptide chain without sulphide bridges and is itself not an enzyme, instead it forms a 1:1 stoichiometric complex with plasminogen that in turn activates other plasminogen

molecules, promoting fibrinolysis (Kim *et al.*, 1997). Staphylokinase has a dual role; studies have suggested that increased production promotes the establishment of *S. aureus* skin infections and increased bacterial penetration through skin barriers whilst surprisingly decreasing the severity of the disease (Kwienski *et al.*, 2013). Once penetration has occurred through the skin barriers, staphylokinase activates plasminogen molecules to plasmin, which is an enzyme that degrades fibrin, a protein component of the extracellular matrix (ECM). This consequently results in the evasion of the host tissue with *S. aureus* (Kwienski *et al.*, 2013). This contributes to the infection by spreading the bacterium.

1.5.4.2 Staphylococcal complement inhibitor (SCIN)

The complement system is part of the human immune system that has the ability to enhance antibodies to phagocytose cells which defend and clear against any invading pathogens, which include *S. aureus* (Rooijackers *et al.*, 2007). *S. aureus* has developed factors that can evade and counteract the hosts complement immune response by secreting a unique Staphylococcal Complement Inhibitor (SCIN) which enables it to survive and persist in the host (Jongerijs *et al.*, 2009). The SCIN gene (*scn*) is an 85 amino acids (aa) residue that is part of the immune evasion cluster (IEC-1) along with Chemotaxis Inhibitory Protein (CHIPS) on a bacteriophage that also encodes staphylokinase and enterotoxin A (van Wamel *et al.*, 2006; Jongerijs *et al.*, 2007). SCIN, an innate immune modulator of *S. aureus* is a C3 convertase inhibitor of the complement system, on the surface of the bacterium that blocks the formation of C3b deposition and C5a generation and therefore blocks all downstream effector functions; this in turn reduces or stops phagocytosis by human neutrophils following opsonisation and ensures the survival of the pathogen in the host (van Wamel *et al.*, 2006; Rooijackers *et al.*, 2007).

1.5.4.3 Chemotaxis inhibitory protein (CHIPS)

The CHIPS gene (*chp*) is also located in the immune evasion cluster and is found in in >60% of all clinical isolates of *S. aureus* (de Haas *et al.*, 2004; van Wamel *et al.*, 2006). CHIPS is a 14.1-kDa protein that is secreted by *S. aureus* to specifically impair the response of neutrophils and monocytes (de Haas *et al.*, 2004). It attenuates the responses of both the C5a receptor as well as the formylated peptide receptor of human neutrophils (van Wamel *et al.*, 2006).

1.5.4.4 IgG binding protein

Staphylococcus aureus also secretes a 436-residue protein called Staphylococcal Immunoglobulin-binding protein (Burman *et al.*, 2008). Sbi (Second binding protein of immunoglobulin), has two IgG-binding domains (D1 and D2) at its N-terminus that have a similar sequence to Protein A (*spa*) which was thought to be the only IgG-binding protein in *S. aureus* (Smith *et al.*, 2011). Sbi also has two independently folded domains, D3 and D4 that bind to complement factor C3, with the C-terminal domain which is rich in the amino acids tyrosine and threonine, likely to be involved in attaching the protein to the cell envelope (Smith *et al.*, 2011). The *sbi* gene which was selected for further studies shows levels of high conservation in both nucleotide and protein sequences and the size of the gene can range from 1311 to 1314 nucleotides, it is dependent on the human strain isolated (Atkins *et al.*, 2008). Sbi helps the *S. aureus* to avoid the hosts innate immune defence, this ensures the survival of the bacteria in the host (Atkins *et al.*, 2011).

1.5.5 Enzymes and other secreted proteins

1.5.5.1 Serine Protease (*sspA*), cysteine protease precursor (*sspB*), cysteine protease (*sspC*)

Serine protease (*ssp*) was one of the first secreted enzymes of *S. aureus* to be purified and characterised in detail; it is a member of the glutamyl endopeptidase family of enzymes (Drapeau *et al.*, 1972). The serine protease structural gene *sspA* is followed closely by an open reading frame

(ORF) encoding a cysteine protease, designated *sspB*. The *sspA* and *sspB* proteases are transcribed as an operon, which also includes a third open reading frame *sspC*, of unknown function (Rice *et al.*, 2001). *sspA* is required for the maturation of *sspB*, which encodes a 40.6-kDa cysteine protease and plays a crucial role in controlling autolytic activity (Rice *et al.*, 2001). A secreted *sspA* may also promote invasion by degrading the bacterial cell surface fibronectin-binding protein (FnBP), a member of the MSCRAMM family (McGavin *et al.*, 1997). *sspA*, *sspB* and *sspC* were selected for further investigation in this thesis.

1.5.5.2 Aureolysin

The complement system is part of the innate immune system and is the first defence against bacteria, once activated it attracts neutrophils to the site of infections and opsonises bacteria to facilitate phagocytosis (Laarman *et al.*, 2011). *S. aureus* has successfully developed ways to evade the human complement system; Metalloprotease aureolysin is a potent complement inhibitor, it functions by inhibiting the central complement protein C3 by cleaving it, preventing phagocytosis of the bacteria, contributing to the *S. aureus* survival in the host (Laarman *et al.*, 2011).

1.5.5.3 Hyaluronate lyase

Hyaluronate lyase is an extracellular enzyme produced by approximately 91.2% of *S. aureus* strains (Makris *et al.*, 2004). The genes that encode this enzyme are *hysA*. *HysA1* and *HysA2* and were subsequently further investigated in this thesis. Studies suggest that it is capable of degrading acidic mucopolysaccharide hyaluronic acid, a major component of the human and animal connective tissue and has a crucial role in the early stages of subcutaneous infections (Makris *et al.*, 2004).

1.5.5.4 Staphylococcal Coagulase (Coa)

The gene that encodes coagulase is *coa* (Shopsin *et al.*, 2000). Coagulase (Coa) is a protein of approximately 670 amino acids secreted by *S. aureus* that promotes coagulation by activating prothrombin, which in turn, converts fibrinogen into fibrin and promotes the clotting of plasma or blood (McAdow *et al.*, 2012). The N-terminal end of Coa associates with the pro site of prothrombin that completes an active site that is otherwise only formed in thrombin (McAdow *et al.*, 2012). This gene was further investigated in this thesis.

1.5.5.5 von Willebrand factor binding protein (vWbp)

Staphylococcus aureus also secretes a secondary coagulase exoenzyme called von Willebrand factor-binding protein (vWbp) of approximately 482 amino acids that has an N-terminal homologous to coagulase (Bjerketorp *et al.*, 2004). The gene that encodes von Willebrand binding-protein is *vwb* (Bjerketorp *et al.*, 2002). The N-terminal end of vWbp associates with the pro site of prothrombin that completes an active site that is otherwise only formed in thrombin (McAdow *et al.*, 2012).

1.5.5.6 Lipase

Nine different lipase genes have been identified from six *Staphylococcus* species which secrete lipase exoenzymes that degrade or digest lipids (Rosenstein & Gotz, 2000). One of the lipases is encoded by the *geh* gene (Rosenstein & Gotz, 2000). The lipase is secreted as a 82-kDa pro lipase protein that comprises of 295 amino acids with full enzymatic activity that is subsequently processed into a mature 45- to 46-kDa protein (Rolof & Normak, 1992). When the body is fighting a bacterial infection, the number of granulocytes and specifically neutrophils in the body will increase (Molne *et al.*, 2000). Staphylococcal lipase will in turn interact with human granulocytes and cause aggregation (Rolof *et al.*, 1992). Studies involving anti-lipase serum have hypothesised that lipase

may also have a secondary role in the process of biofilm formation in *S. aureus* (Xiong *et al.*, 2009). More recent studies have strengthened this hypothesis; Hu *et al* (2012) concluded that the deletion of the lipase coding gene reduced biofilm formation.

1.5.5.7 Staphylococcal nuclease

Staphylococcus aureus produces an exoenzyme called Staphylococcal nuclease that can hydrolyse both DNA and RNA in the host resulting in tissue destruction and promotes the further spreading of the disease (Hu *et al.*, 2013). Staphylococcal nuclease is encoded by the *nuc* gene (Tang *et al.*, 2008), this gene was further investigated in this thesis. Nucleases can be characterised as either sugar specific nucleases which include ribonuclease and deoxyribonuclease, or the sugar non-specific nucleases which includes the Staphylococcal nuclease, which is characterised by its ability to hydrolyse both DNA and RNA, without exhibiting base preference (Ragarajan & Shankar, 2001). Staphylococcal nuclease functions by hydrolysing the phosphodiester bonds on the DNA or RNA to yield 3' mono-nucleotides (Tang *et al.*, 2008).

1.5.5.8 Thermonuclease

The second nuclease of *S. aureus* is called thermo-nuclease, and it was annotated on the whole genome sequencing data of methicillin-resistant *S. aureus*, as N315 and Mu50 (Kuroda *et al.*, 2001). More recently, it has been made clear that thermo-nuclease (TNase) is encoded by the *nucA* gene (Tang *et al.*, 2008). It is a heat-stable nuclease with both endo- and exo-nucleolytic properties that can hydrolyse both DNA and RNA and has been used as a specific PCR diagnostic test to identify *S. aureus* contamination in blood cultures (Kaplan, 2003).

1.5.5.9 1-phosphatidylinositol phosphodiesterase (plc)

S. aureus produces an exoenzyme called 1-phosphatidylinositol phosphodiesterase (plc) which was further investigated in this thesis, that was part of the group of molecules identified in CA-MRSA culture media that may be involved in bacteria-host interaction and also act as a membrane damaging phospholipase, however it is not well characterised (Burlak *et al.*, 2007; Suzuki *et al.*, 2012). However, the phosphatidylinositol-specific phospholipase C (PI-PLC-) of *Listeria monocytogenes* promotes adhesion to epithelial cells and mediates the escape of pathogens from phagosomes which in turn allows it to proliferate and grow in the host cells cytosol and consequently spread to neighbouring cells (Wei *et al.*, 2005). Studies have also shown that *plc* in *Bacillus subtilis* strains plays an accessory role that enhances bacterial binding to epithelial cells prior to its internalisation (Krawczyk-Balska & Bielecki, 2005).

1.5.6 **Cytotoxins**

1.5.6.1 Alpha Hemolysin

Staphylococcus aureus secretes a pore-forming cytotoxin called alpha-hemolysin (Bhakdi & Tranum-Jenson, 2009). It is pore forming because it has the ability to form a 2-nm heptameric pore in the plasma membrane of susceptible cells (Ragle & Waldenburg, 2009). The alpha, beta and delta toxins are also the primary cause of erythrocyte lysis that induce the release of cytokines and chemokines such as IL-6, IL1 β , IL-1 α , IL-8, TNF- α , KC and MIP-2 (Burnside *et al.*, 2010). The first 26 residues of the secreted alpha-hemolysin protein is cleaved under secretion and the mature protein comprises of 239 residues, that have a molecular weight of 33-kDa (Dinges, Orwin & Schlievert, 2000). The alpha-hemolysin also has no cysteine amino acids and is instead compromised of primarily β -sheets (65%), with α -helical structures making up a much smaller percentage (10%) of the secondary structure (Dinges, Orwin & Schlievert, 2000). The alpha-hemolysin is encoded by the *hla* gene and is important for *S. aureus* pneumonia (Burnside *et al.*,

2010). Studies have also concluded that immunisation with anti-alpha-hemolysin monoclonal antibodies mediate protection against *S. aureus* pneumonia (Ragle & Waldenburg, 2009).

1.5.6.2 Beta Hemolysin

Beta-hemolysin (HLB) is well known for its haemolytic activity of erythrocytes (Katayama *et al.*, 2012). This gene was selected for further study in this thesis. HLB is an exotoxin with a molecular weight of 35-kDa that is encoded by the beta-hemolysin gene, termed *hlyB*, that is chromosomally located on a 4-kb DNA fragment (Dinges, Orwin & Schlievert, 2000). Recent studies have also suggested that HLB may have a dual role and that it also plays an important role in skin colonisation; the study constructed two *S. aureus* MW2 strains *in vitro*, both with and without the *hlyB* gene. Colonisation tests concluded that colonisation efficiency was 50-fold greater with the *hlyB* gene than without the *hlyB* gene (Katayama *et al.*, 2012).

1.5.6.3 Gamma Hemolysin and Panton-Valentine Leuckocidin

Staphylococcus aureus also produces two other cytotoxins called gamma-hemolysin and Panton-Valentine leuckocidin (PVL) toxin; both function as bi-component toxins and play a crucial role in the disruption and subsequent lysis of erythrocytes and leukocytes (Gouaux, Hobaugh & Song, 1997; Dinges, Orwin & Schlievert, 2000). Each of these toxins are made up of two non-associated secreted proteins, referred to as S and F components (for slow- and fast-eluting proteins in an ion exchange column) (Dinges, Orwin & Schlievert, 2000). Gamma-hemolysin is present in virtually all human strains (99.3%) whilst the PV leuckocidin is not as widely distributed (Spaan *et al.*, 2014). The genes for gamma-hemolysin are transcribed from a single locus, located on a 5.4-kb *ScaI*-digested chromosomal fragment; the open reading frames are named *hlyA*, *hlyC* and *hlyB* (Table 3.5) (Dinges, Orwin & Schlievert, 2000). *hlyC* and *hlyB* have molecular weights of 34-kDa and 32.5-kDa, respectively, and are transcribed on a single mRNA, whilst *hlyA* has a molecular weight of

32-kDa, and is expressed separately (Dinges, Orwin & Schlievert, 2000). The genes encoding PV-Leukocidin are named *lukS*-PV and *lukF*-PV, are expressed as an operon and the *lukS*-PV gene encodes a 312-amino acid polypeptide, including a 28 residue signal sequence that is cleaved to produce the mature protein with a molecular weight of 32-kDa (Dinges, Orwin & Schlievert, 2000). The *lukS* gene encodes the S-component of Staphylococcal leukocidin and contains 857 bp (Rahman *et al.*, 1991).

1.5.7 Enterotoxins

1.5.7.1 Staphylococcal Enterotoxins

Staphylococcal Enterotoxins are a family of nine major serological types of heat stable enterotoxins, SEA through SEE and SEG through to SEJ, that cause sporadic food-poisoning outbreaks in both the hospital and community environment (Balaban & Rasooly, 2000; Ortego *et al.*, 2010). Staphylococcal food poisoning is characterised by a short incubation period, approximately 2- 6 hours and is followed by symptom's such as nausea, vomiting, abdominal pain and diarrhoea, although not lethal, the elderly are more susceptible to morbidity and mortality (Balaban & Rasooly, 2000).

The enterotoxins are bi-functional; they act as potent gastrointestinal toxins and also as super antigens that stimulate non-specific T cell proliferation; both functions are located in two separate domains that target different host tissues (Balaban & Rasooly, 2000; Ortego *et al.*, 2010). However, the mechanisms are not very well understood (Rasooly & Hernlem, 2014). The enterotoxins all share common phylogenetic, structure, function and sequence homology (Rasooly & Hernlem, 2014). Enterotoxin A (SEA) is, however, the most common toxin implicated in staphylococcal food poisoning (Rasooly & Hernlem, 2014). SEA is carried by the *entA* gene on a bacteriophage and is composed of 771 base pairs, it also encodes a enterotoxin A precursor of 257

amino acids (Balaban & Rasooly, 2000). Studies have suggested that the amount of enterotoxin needed to cause food poisoning is very small, the emetic dose in primate assays was approximately 5-20 µg per animal (Balaban & Rasooly, 2000).

1.5.7.2 Toxic shock syndrome toxin-1 (TSST-1)

Approximately 20% of all *S. aureus* isolates also possess the gene encoding TSST-1 (*tst*) (Table 3.5), which is harboured by a family of mobile staphylococcus pathogenicity islands designated SaPI, that is not carried by all *S. aureus* strains (Al Laham *et al.*, 2014). It is carried by *S. aureus* strains RN4282, N315, Mu50, RN3984 and RF122 in mobile pathogenicity islands termed SaPI1, SaPI_n1, SaPI_m1, SaPI₂ and SaPI_{bov}1, respectively (Andrey *et al.*, 2015). TSST-1 is a potent super antigen and is the most common cause of toxic shock syndrome (tss), it is a life threatening multi system disorder characterised by a high fever, hypotension, rash and skin desquamation (peeling) upon recovery (Prasad *et al.*, 1997).

1.6 Clinical Significance & Epidemiology

It is evident that methicillin resistant *S. aureus* infections pose a significant health burden around the globe. Figures 1.3 and 1.4 are heat maps taken from the annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net). Figures 1.3 and 1.4 show that the incidence of MRSA is changing around Europe. An improvement has been observed in the UK; in 2009, 28% of the *S. aureus* isolates were resistant to methicillin, in 2014, only 11% of the 3551 isolates were resistant to methicillin (ECDC-1, 2010; ECDC-2, 2015). A slight decrease in the number of MRSA reported has also been observed in Spain where the number of *S. aureus* isolates resistant to methicillin decreased from 26% to 22% from 2009 to 2014. Other countries have remained consistently high. Portugal reported that in 2009, 53% of the isolates were methicillin resistant; in 2014, 47% of the *S. aureus* isolates were methicillin resistant (ECDC-1, 2010; ECDC-2, 2015).

It is also interesting to note that the surveillance of *S. aureus* has also improved. Slovakia did not report at all in 2009; however, in 2014, 28% of the 640 *S. aureus* isolates reported were MRSA. Romania reported only 48 *S. aureus* isolates in 2009 of which 38% were MRSA. This increased to 399, in 2014 of which 54% were methicillin resistant. The number of laboratories contributing to the EARS-Net surveillance has also been increasing. Portugal had 20 laboratories in 2009, this increased to 53 in 2014. The same trend is shown in Italy, where the number of laboratories doubled from 23 in 2009 to 46 in 2014 (ECDC-1, 2010; ECDC-2, 2015). In contrast, countries like Sweden and The Netherlands have remained consistently low. In 2009, The Netherlands reported 1035 isolates of *S. aureus*, of which <1% were resistant to methicillin. The same trend continued in 2014 even though the number of *S. aureus* isolates reported was 2580. The same trend was observed in Sweden; in 2009 and 2014, 2460 and 3501 isolates of *S. aureus* were reported. Nevertheless, the number of samples resistant to methicillin was <1% (ECDC, 2010 & 2015).

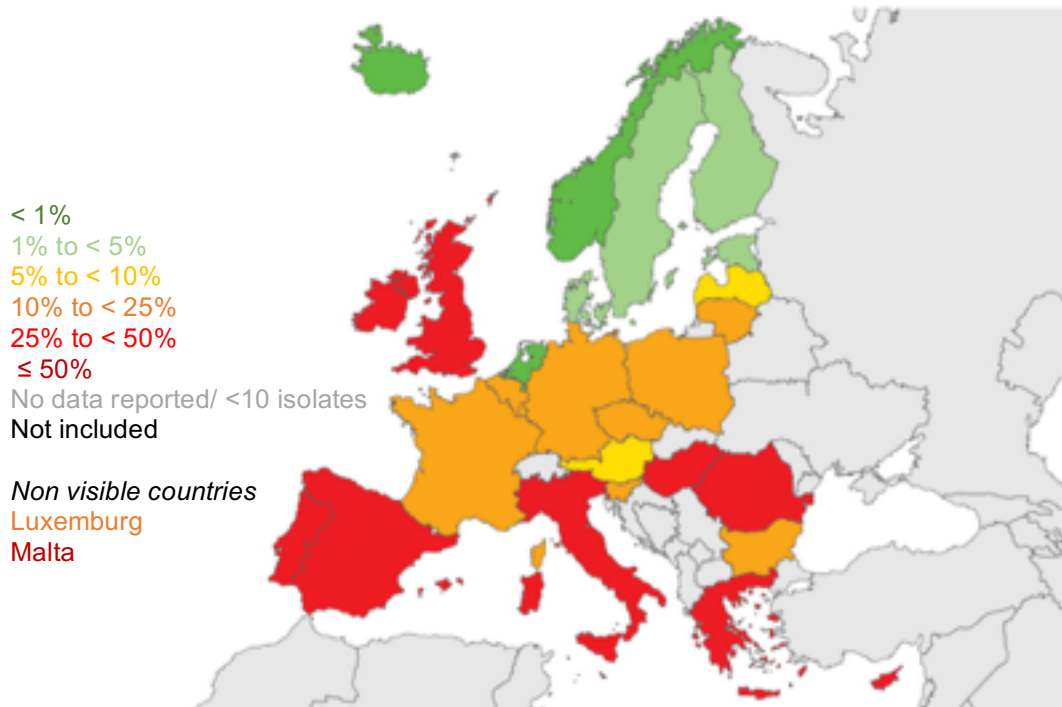


Figure 1.3: taken from the Annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net); showing the percentage (%) of *S. aureus* isolates with resistance to methicillin (MRSA), by countries in 2009 (ECDC-1, 2010).

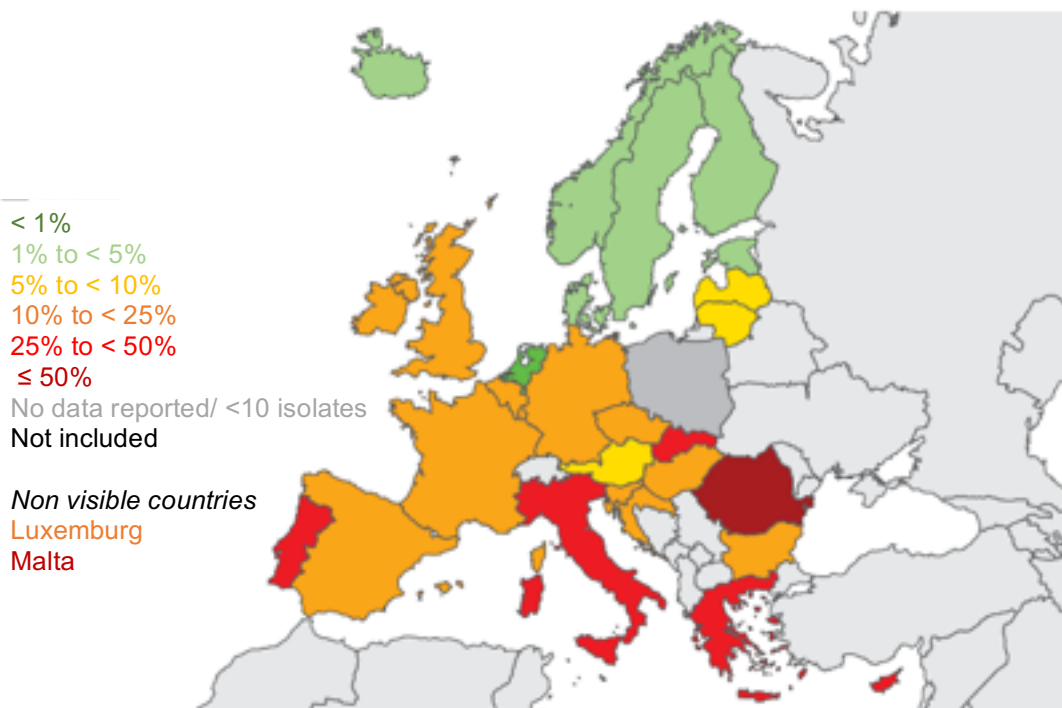


Figure 1.4: taken from the Annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net); showing the percentage (%) of *S. aureus* isolates with resistance to methicillin (MRSA), by countries in 2014 (ECDC-2, 2015).

The Office of National Statistics in England and Wales has also published the number of reported deaths caused by *S. aureus*. Figure 1.5 is a graph showing the number of death certificates mentioning MRSA between 1993 and 2012. The number of deaths from MRSA between 1993 and 2006 has risen from 430 in 1993, peaking in 2005 (2,099) and 2006 (2,150) and gradually decreasing to 557 in 2012 (ONS-1, 2013); this is nevertheless a cause for concern. The graph is also showing the number of death certificates mentioning *S. aureus* that are either not specified as resistant or strains that are resistant to antibiotics (MRSA). Interestingly, between 1993 – 2000 the majority of death certificates mentioning *S. aureus* were not specified as resistant, this could be due to the lack of testing for antibiotic resistance. It is not a surprise that the majority of strains that cause deaths between 2000 – 2009 are resistant (MRSA), with 2010 being the anomaly to this trend.

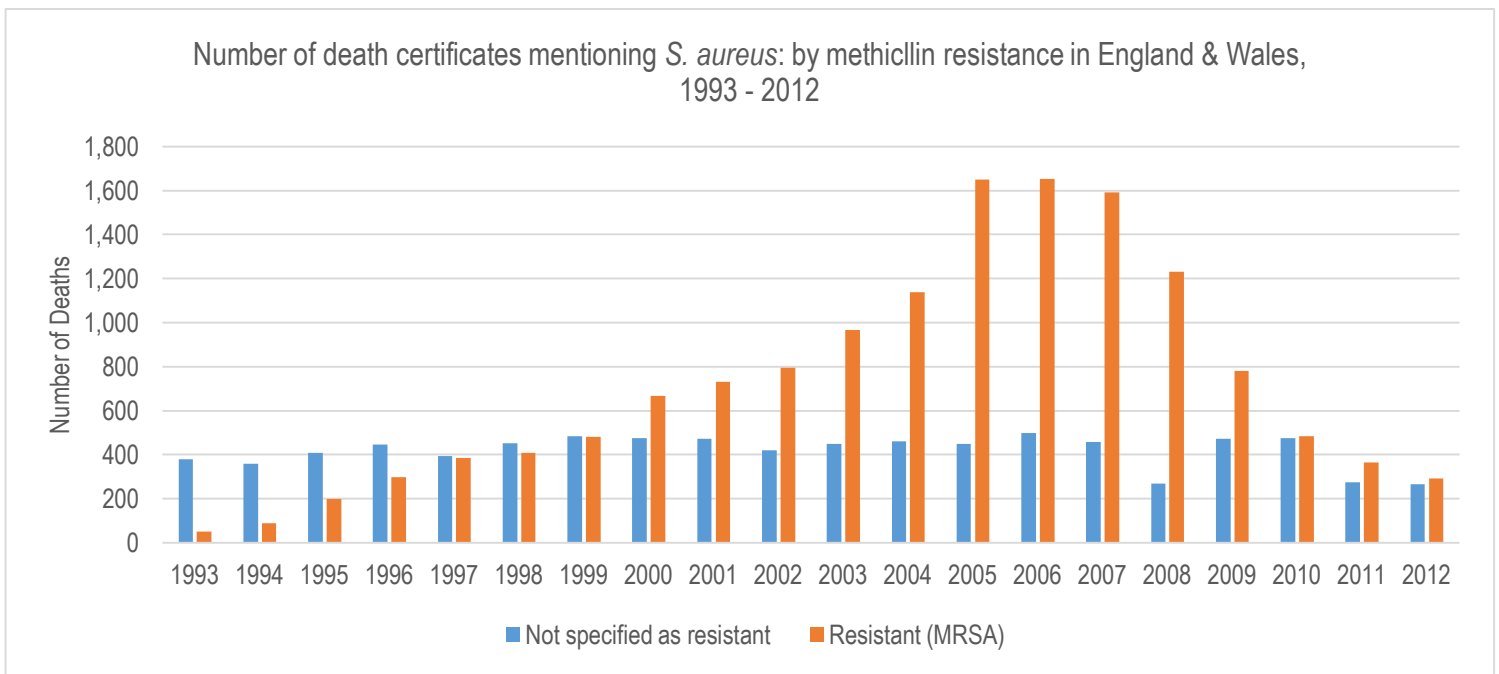


Figure 1.5: A graph representing the number of death certificates mentioning *S. aureus*: by Methicillin Resistant *Staphylococcus aureus* (MRSA) in England & Wales. The data was adapted from the Office for National Statistics (ONS, 2013).

Figure 1.6 is a graph representing the number of death certificates mentioning MRSA by sex and age group from 2008 to 2012 (ONS-1, 2013). It is interesting to note that males are more

susceptible to MRSA then females in all age groups. The Office for National Statistics has also concluded that the morbidity rate due to MRSA is greater in the elderly (ONS-1, 2011). This is the trend that can be seen here with males 75 and over far more likely to have death certificates mentioning MRSA.

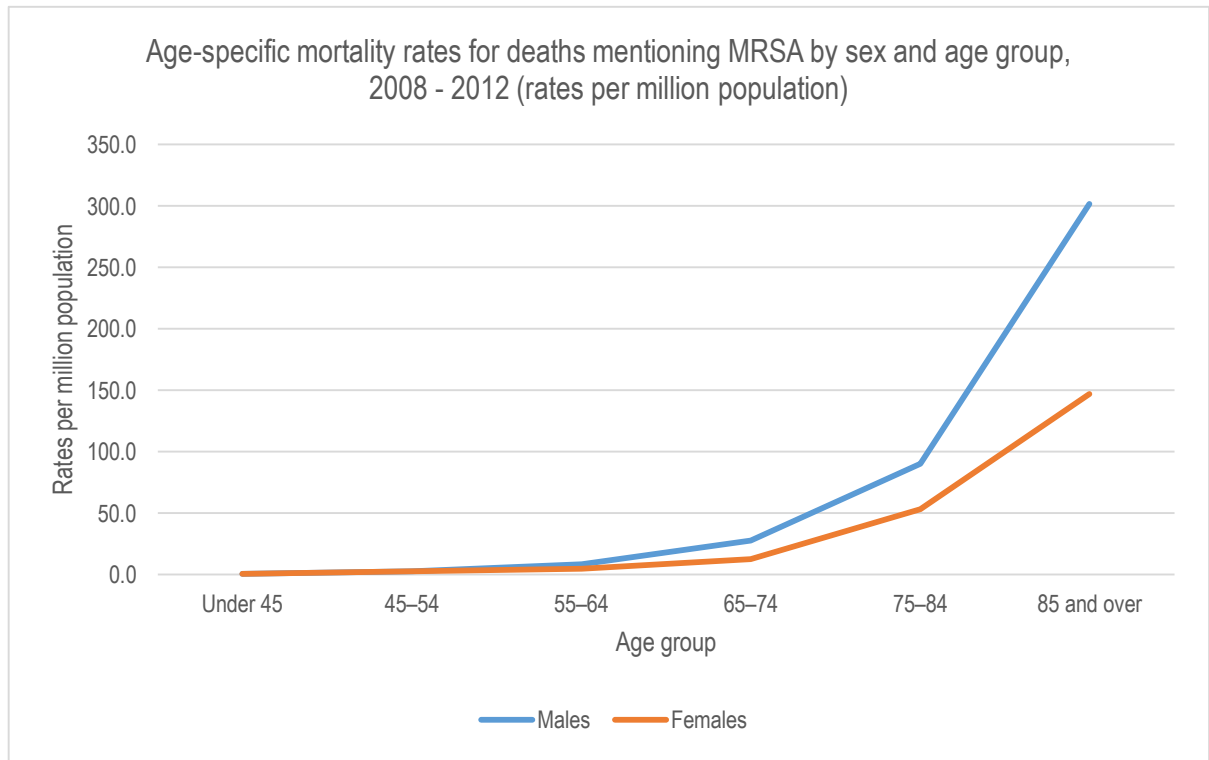


Figure 1.6: Age specific mortality rates for deaths mentioning MRSA by sex and age group between 2008 and 2012, in England and Wales; rates per million population. The data was adapted from the Office for National Statistics (ONS, 2013).

Public Health England (PHE), formerly The Health Protection Agency have also published mandatory surveillance data from NHS hospitals in England & Wales. Figure 1.7 is a graph representing the annual MRSA counts from April 2007 to March 2016; the overall bacteremia has

gradually reduced from 4,451 (between April 2008 – March 2009) to 1,898 (April 2009 – March 2010) to a decreased rate of 823 (between April 2015 – March 2016) (HPA-1, 2011).

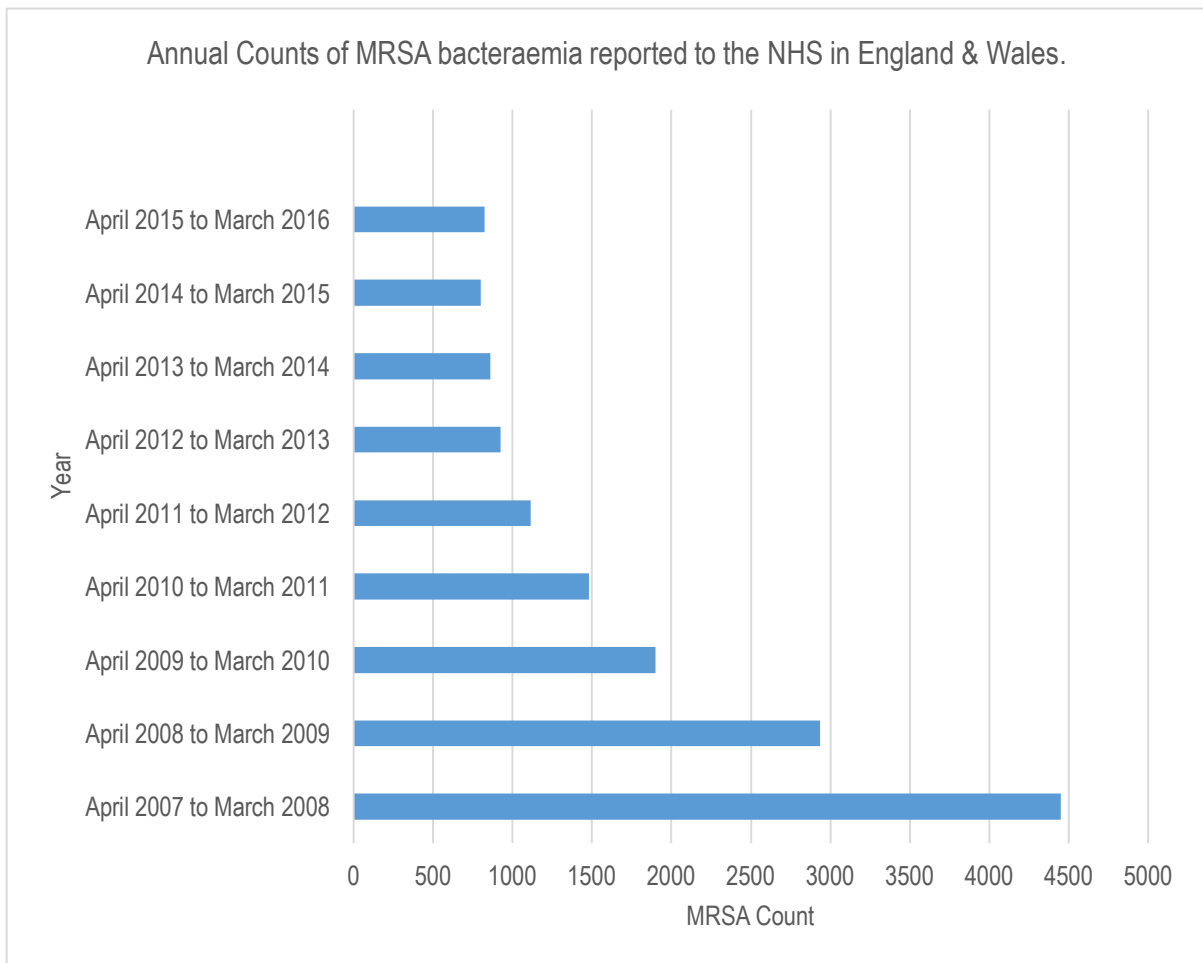


Figure 1.7: A graph representing the annual MRSA counts from April 2007 to March 2016. The data was adapted from the Health Protection Agency which is now a part of the Public Health England (HPA-1, 2016).

1.7 Symptoms and Clinical Presentation

S. aureus is a major pathogen that can cause a diverse range of clinical conditions and diseases in humans. This can range from superficial infections of the skin such as boils and styes to more severe soft tissue infections (Casey *et al.*, 2013). A population-based study of methicillin-resistant *Staphylococcus aureus* skin infection in Pennsylvania from 2001 – 2010 identified 78216

Skin and Soft Tissue Infections (SSTI). Among the SSTI cases the three most common diagnoses were cellulitis (62%), carbuncle (17%) and impetigo (12%) (Casey *et al.*, 2013). Infections with this organism are not just restricted to humans; the bacteria have the ability to invade and cause harm to a diverse range of animal species which include commercially important species such as cattle, sheep, swine and poultry, as well as companion animals and horses. *S. aureus* can cause an assortment of clinical symptoms and diseases ranging from superficial to potential life threatening illnesses, as can be seen in Table 1.1.

Host	Clinical symptoms/ diseases associated with <i>S. aureus</i> infections	References
Human	<ul style="list-style-type: none"> • Blood stream infections • Toxic shock syndrome • Scalded skin syndrome • Impetigo • Folliculitis • Cellulitis 	<p>Amaral <i>et al.</i>, 2005</p> <p>DeVries <i>et al.</i>, 2011</p> <p>Shi <i>et al.</i>, 2011</p> <p>Pereira, 2012</p> <p>Nichols & Florman, 2001</p> <p>Khawcharoenporn <i>et al.</i>, 2010</p>
Horses	<ul style="list-style-type: none"> • Wound infection, infected joints and suspected infections of various organ systems • Septic arthritis/ Pneumonia 	<p>Cuny <i>et al.</i>, 2006</p> <p>Weese <i>et al.</i>, 2006</p>
Cattle	<ul style="list-style-type: none"> • Mastitis • Purulent dermatitis 	<p>Haran <i>et al.</i>, 2012</p> <p>Grinberg <i>et al.</i>, 2004</p>
Sheep	<ul style="list-style-type: none"> • Mastitis • Pyemia 	<p>Chu <i>et al.</i>, 2012</p>
Companion animals	<ul style="list-style-type: none"> • Skin and soft tissue infections • Pleuropneumonia 	<p>Pantosti, 2012</p> <p>Baptiste <i>et al.</i>, 2005</p>
Swine	<ul style="list-style-type: none"> • Exudative epidermitis • Infections of the urogenital tract, uterus & mammary glands. • Septicaemia 	<p>van Duijkeren <i>et al.</i>, 2007</p> <p>Crombe <i>et al.</i>, 2013</p>
Rabbit	<ul style="list-style-type: none"> • Mastitis • Gangrene mastitis • Dermal lesions • Subcutaneous abscess 	<p>Hamed & Ahmed, 2013</p> <p>Vautor <i>et al.</i>, 2009</p>
Poultry	<ul style="list-style-type: none"> • Bacterial chondronecrosis with osteomyelitis (BCO) • Septicaemia 	<p>McNamee <i>et al.</i>, 2010</p>

Table 1.1: Clinical symptoms and diseases associated with *S. aureus* in different hosts.

1.7.1 Cellulitis

Cellulitis is one of the most common out-patient infections presented to physicians (Khawcharoenporn *et al.*, 2010). It is an acute infection of the skin and underlying soft tissue and is usually present amongst the elderly or immunocompromised (Nichols & Florman, 2001). It presents as tender red but firm areas of the skin and has the potential to lead to necrotizing fasciitis and gangrene (Nichols & Florman, 2001).

1.7.2 Impetigo

Impetigo is the most common bacterial infection of the skin caused by *S. aureus*. There are two types of impetigo; bullous impetigo and non-bullous impetigo (Pereira, 2012). Non-bullous impetigo represents 70% of all cases of impetigo and is recognised by the development of a serous, yellow-brown exudate which dries into a golden crust (Nichols & Florman, 2001). Non-bullous impetigo generally occurs in adults and children, although rarely in children under 2 years of age (Pereira, 2012). Bullous impetigo is caused by exfoliative toxin A and is most common among children aged two to five years (Pereira, 2012).

1.7.3 Folliculitis

Folliculitis is a pyoderma that arises within a hair follicle and usually occur in the moist areas of the body that undergo friction and perspiration (Nichols & Florman, 2001). The incidence of folliculitis is higher among patients who receive treatment with corticosteroids or cytotoxic agents (Nichols & Florman, 2001).

1.7.4 Toxic Shock Syndrome

In addition to this, *S. aureus* also has the ability to cause toxinoses such as scalded skin syndrome and toxic shock syndrome by releasing super antigens into the blood stream (Sousa & Lencastre, 2004).

Methicillin resistant *S. aureus* can excrete toxins that can potentially cause toxic shock syndrome (TSS), a rare life threatening illness that is defined by clinical and laboratory evidence of fever, rash, skin desquamation, hypotension and multiple organ failure (Prasad *et al.*, 1997). A population based surveillance study conducted from 2000 – 2006 in the Minneapolis – St. Paul area (population 2,642,056), has concluded that the average annual incidence of toxic shock syndrome is 0.52 per 100,000 of the population, the rate was slightly higher in menstrual TSS (0.69) with women aged 13 – 24 years having the highest incidence at 1.41 (DeVries *et al.*, 2011).

1.7.5 Scalded Skin Syndrome

Staphylococcal Scalded skin syndrome (SSSS) is caused by MRSA that produce exfoliative toxins. It is a life threatening infection for neonates in neonatal units and can develop in extremely low-birth-weight neonates (Shi *et al.*, 2011). It usually affects infants and children, but has also been reported in adults associated with immunosuppression and/or renal failure resulting in a high mortality rate (Ito *et al.*, 2002). A retrospective study conducted on 399 children under the age of 1-year old (177 girls/ 222 boys) hospitalised in the Czech Republic from 1994 – 2009 concluded that the average incidence of SSSS was 25.11 cases per 100,000 with 1994 resulting in the highest incidence of 53.47 per 100,000 children. (Lipovy *et al.*, 2011).

S. aureus can also cause urinary tract infections (UTIs), difficult to treat infections of the blood, cardiovascular infections and in severe life threatening cases, it can cause; osteomyelitis, meningitis, brain abscess and endocarditis (Amaral *et al.*, 2005; Klein *et al.*, 2007; Sheng-Yun *et al.*,

2011). This bacterium is also a major cause of food poisoning; it achieves this by releasing enterotoxins, specifically Enterotoxin C, into the food (Aydin *et al.*, 2011). Recently developed economic simulation models to help quantify the costs associated with MRSA infection and the outcomes are alarming (Lee *et al.*, 2013). In the USA, depending on the MRSA incidence, this bacterium alone can cause an annual economic burden of \$478 million – \$2.2 billion on third-party payers and \$1.4 billion – \$13.8 billion on society (Lee *et al.*, 2013). This clearly demonstrates the importance of *S. aureus* as a pathogen

1.8 MRSA Evolution

1.8.1 Hospital-associated MRSA

It was once recognised that MRSA only occurred in health-care settings such as inpatient hospital admission, haemodialysis units or same-day surgery, and was therefore defined as a health care-associated infection (HAI), with patients who were generally older and immunocompromised being most at risk (Sydnor & Perl, 2011). MRSA found in the health-care setting were therefore termed health-care associated MRSA/ hospital-associated MRSA (HA-MRSA) (Fey *et al.*, 2003). HAIs can have a devastating impact in the hospital environment by increasing morbidity, mortality, length of stay and dramatically increasing the cost (Sydnor & Perl, 2011). The US Centre for Disease Control and prevention (CDC) have provided definitions of HA-MRSA and Community-associated (CA) -MRSA. “HA-MRSA is identified >48h after admission of a healthcare facility, or if an individual has previously had a history of MRSA infection or colonisation, admission to a healthcare facility, dialysis, surgery or insertion or indwelling devices in the past year” (Bal *et al.*, 2016).

1.8.2 Community-associated MRSA

It is now recognised that MRSA is not just restricted to the health-care setting. A shift in MRSA epidemiology occurred when cases of MRSA started appearing in the community in the 1990's and affected persons with no healthcare-related predisposing factors and were termed community-associated MRSA (CA-MRSA) (Frazee *et al.*, 2005). Initial reports of a new type of MRSA causing infections amongst healthy individuals was reported in the remote Aboriginal communities of Western Australia (Deurenberg & Stobberingh, 2008). Now, CA-MRSA have become increasingly endemic in many communities and are responsible for a large proportion of the disease burden and certain strains have the potential to cause fatal infections in otherwise healthy individuals (Baba *et al.*, 2002). CA-MRSA is now the most common cause of skin and soft tissue infections (SSTIs) presenting in

the American hospital emergency rooms (Chen *et al.*, 2011). In the 1990s, cases started occurring among children who had none of the common identifiable predisposing risk factors (Herold *et al.*, 1998). In one particular instance, between 1997 – 1999 four healthy children from a relatively small region in the USA died of clonally related MRSA strains (CDC, 1999). In these cases, the patients were initially treated with cephalosporin antibiotics. The delayed use of the correct antibiotic to which MRSA is susceptible may have contributed to the fatal outcomes, as a result of this and so that the correct therapies could be administered, patients are now routinely screened for MRSA cultures upon admission in America (CDC, 1999; Manian *et al.*, 2012). The CDC defines CA-MRSA as “being identified in the outpatient setting or within 48h following hospital admission in an individual with no medical history or MRSA infection or colonisation, admission to a healthcare facility, dialysis, surgery or insertion of indwelling devices in the past year” (Bal *et al.*, 2016).

An example of a CA-MRSA isolate is USA300. A cohort study by Kreisel *et al.* (2011), carried out between 1997 to 2008 in four hospitals in the USA compared sepsis in response to infection in 271 patients. Kreisel *et al.* (2011) concluded that 25% (67 strains of USA300 MRSA) of the patients with MRSA bacteraemia were USA300 MRSA and 75% (204 strains non-USA300 MRSA) were non-USA300 MRSA. USA300 MRSA were categorised because they possessed two unique virulence factors in the genome; the genes for PVL and the arginin catabolic mobile element (ACME) (Kreisel *et al.* (2011)). USA300 MRSA increased over time and patients with USA300 MRSA were more likely to have severe sepsis or septic shock in response to their infection than patients with non-USA300 MRSA (Kreisel *et al.*, 2011). The patients with USA300 MRSA bacteraemia were younger than patients with non-USA300 MRSA bacteraemia and were considered less likely to be hospitalised, had surgery or to acquired their infection nosocomially (Kreisel *et al.*, 2011). This study shows evidence that the community-associated USA300 MRSA strains may be more virulent than other strains of MRSA. A study by Diep *et al* (2006), showed that the major differences between HA-MRSA and CA-MRSA are that HA-MRSA strains harbour the SCC*mec* type II element, whereas CA-MRSA

strains harbour the SCC mec type IV element. Certain strains of CA-MRSA also harbour genes for the Panton-Valentine Leukocidin (PVL), a bi-component cytolytic toxin (1.5.6.3). It is also noted that PVL is not essential for CA-MRSA infection but may be partly responsible for the increased virulence (Diep *et al.*, 2006; Boyle-Vavra & Daum, 2007). This observation has led to a change in that the presence of PVL is no longer an indicator of a strain that is Community-acquired.

1.8.3 Livestock-associated MRSA

Initial reports of MRSA in farm food-producing animals were published in 1972; the bacteria were isolated at a low prevalence, from dairy cows with mastitis in Belgium, and sporadic cases were occasionally reported (Weese, 2010). It was not until 2004 that significant concerns about MRSA and food animals emerged (Weese, 2010). Reports of MRSA in livestock, and particularly in pigs and individuals who work with pigs emerged in The Netherlands (Arman-Lefevre *et al.*, 2005). This particular MRSA clone was not related to the historic HA-MRSA or even the CA-MRSA clones due to the presence of a novel DNA methylation enzyme that was not typeable by PFGE (pulse gel field electrophoresis), the gold standard for typing *S. aureus* strains; further typing concluded that this clone was ST398 (Sequence Type 398) (Deurenberg & Stobberingh, 2008).

A novel MRSA reservoir had therefore been identified, with reports of MRSA isolation from pigs and pig-farmers (Arman-Lefevre *et al.*, 2005). This has provided evidence of a live-stock associated reservoir and live-stock associated MRSA (LA-MRSA) are mainly associated with clonal complex (CC) 398; although CC398 does not have host specificity and can colonise or infect a variety of animals (Cuny *et al.*, 2013). However, because ST398 strains have been predominantly reported in pigs around the globe, it has been suggested that this strain emerged among pigs and was subsequently disseminated to other species (Weese, 2010). These animals were food-producing animals and were given heavy loads of antimicrobial agents and were therefore likely to be highly

colonised with resistant bacteria due to antimicrobial selective pressure; consequently, the farmers are at a particular risk of infection due to their close contact with these animals (Arman-Lefevre *et al.*, 2005). These clones have now been categorised as live-stock associated MRSA (LA-MRSA) and are predominately found in the Netherlands, Denmark, Belgium and surrounding countries but are rapidly disseminating around the world (Lindsay, 2010).

The US Centre for Disease Control and prevention (CDC) have provided definitions of HA-MRSA and CA-MRSA. The CDC defines CA-MRSA is identified in the outpatient setting or within 48h following hospital admission in an individual with no medical history or MRSA infection or colonisation, admission to a healthcare facility, dialysis, surgery or insertion of indwelling devices in the past year.

1.9 Next Generation Sequencing

In 1977, Frederick Sanger and his colleagues announced a new method for DNA sequencing which was later commonly called 'Sanger Sequencing' (Sanger, Nicklen & Coulson, 1977). This method was very labour intensive and expensive but nevertheless revolutionised the field of molecular biology. Over the years, DNA sequencing has been optimised to make it considerably cheaper and more efficient. The first whole genome to be sequenced was of the bacterium *Haemophilus influenzae*; the complete nucleotide sequence consisted of 1,830,137 base pairs (Fleischmann *et al.*, 1995). Since then, hundreds of microorganisms have been sequenced and can be publicly accessed using GenBank which was set up by the National Centre for Biotechnology and Information (NCBI) (Bansal, 2005). The revolution in DNA sequencing began 10 years later in 2005 with 454 Life Science, which was the first of the 2nd- generation techniques of DNA sequencing (Rothberg & Laemon, 2008). The dawn of this technology allowed DNA sequencing to be carried out in a fraction of the time and at a reduced cost (Rothberg & Laemon, 2008). Over the same time

period, competing companies; Illumina and Applied Biosystems have developed their own sequence platform and technologies; the Genome Analyser and SOLiD system, respectively Rothberg & Laemon, 2008).

In this study, the DNA sequencing will be undertaken using the Life Technologies Ion Torrent platform. This 3rd-generation technology is cost effective, not labour intensive and is comparatively rapid (Daum *et al.*, 2012). Ion Torrent technology uses a semiconductor platform that utilises a small chip to measure the release of hydrogen ions emitted during DNA polymerisation (Daum *et al.*, 2012).

1.10 Bioinformatics and its use in the study of *S. aureus*

Bioinformatics was developed in the 1980s. In the past, scientists would study one gene at a time. This was due to the limitation of computer technology; however, since the rapid evolution in computer technology and the ever increasing improvement of memory storage capabilities; it has been possible to store the extremely large information data sets of the organism's whole genome (Bansal, 2005). The advancement of internet has also allowed researchers to upload the complete sequence to open access databases (Bansal, 2005). The advantages of bioinformatics are that it has and will continue to provide invaluable information; for instance, improve laboratory diagnosis for infectious (and non-infectious) disease markers, help improve/ develop new drugs, help the production of cheaper yet effective vaccines, and help understand the mechanisms of virulent genes and how they perform at a cellular, molecular and systemic level (Bansal, 2005). One such area of bioinformatics, which is used in this thesis is the use of Basic Local Alignment Search Tool Nucleotide (BLASTn) which permits a queried sequence to be compared with known sequences whether it is from the same species or from a completely different genus (Altschul *et al.*, 1990; Altschul *et al.*, 1997; Wren & Parkhill, 2011). As previously described in Section 1.3.6, effective treatment options for drug resistant *S. aureus* are becoming less available. A computational analysis by Yadav *et al* (2012), has led to the identification of novel potential therapeutic drug. All the gene products involved

in metabolic pathways of community-associated MRSA in the KEGG database were searched against the human genome; and using the BLAST program seven targets were identified as potential therapeutic drug targets (Yadav *et al.*, 2012). The KEGG database consists of metabolic pathway maps for cellular and organism functions, complete genomes and the drug interactions associated with it (Kanehisa, 2002). These gene products have the potential to be future targets of any drugs that may be developed to combat *S. aureus* infections.

1.11 Aims and objectives

- This thesis will evaluate the the three classifications that have been defined in the literature (hospital-associated, community-associated and livestock-associated) for *S. aureus* by performing a comparative genomic analysis of the genomes of published and novel strains.
- Investigate the presence/ absence of virulence factor genes that are expressed on the genomes of these *S. aureus* strains and infer how this may contribute to the virulence of the strains.
- Investigate and attempt to establish a phylogenetic relationship between the *S. aureus* strains isolated from hospital-associated, community-associated and livestock-associated settings.
- Investigate the importance of positive selection in the evolution of the *S. aureus* species, by identifying the virulence factor genes that are showing this characteristic and determine how positive selection of virulence factor genes may affect the adaptation to the hosts.
- Investigate the geographical relationship of the *S. aureus* strains and determine the subsequent genetic diversity of the population.

2.0 Materials and Methods

2.1 Bacterial Strains

Isolates of *Staphylococcus aureus* from human and animal hosts were investigated in this study. Strains from various host species were kindly provided by Dr Patrick Butaye (Veterinary & Agrochemical Research Centre, Brussels, Belgium). These included: human (pig farmers, n=4), pig (surveillance study, n=5), cattle (mastitis, n=3), horse (various infections, n=3), chicken (surveillance study, n=4) and rat (surveillance study, n=3), and were collected in Belgium between 2008 and 2010. Isolates from human patients were kindly provided by Professor Valerie Edward-Jones (Manchester Metropolitan University, UK). The isolates included: outpatients (n=10), inpatients (n=10) and burns patients (n=10). The School of Healthcare Science, Manchester Metropolitan University specialises in infectious disease, antimicrobial resistance and alternative antimicrobial therapies. The Veterinary & Agrochemical Research Centre specialises in research that contributes to international public healthcare policy focusing on food production and animal health.

From this diverse range of isolates available for use (n=52), a mixed sample set of human and animal *S. aureus* strains (n=6) were used for further studies (Table 2.1), and the remaining samples (n=46) were frozen (-80° C) in vials. The strains that were used, included animal *S. aureus* strains B53, S62 and AS3, from cattle, chicken and swine, respectively and were all designated livestock-associated strains by The Veterinary & Agrochemical Research Centre. As well as human *S. aureus* strains, OP100, B51 and IP70 from an outpatient, inpatient and burns patient, respectively, that have all been designated as hospital-associated strains by the Manchester Metropolitan collaborators.

Isolate	Isolated from:	Institution	Kingston University date of arrival
OP100	Out patient	Manchester Metropolitan University, UK	12 th February 2015
B51	Burns patient		
IP70	In patient		
B53	Cattle	Veterinary & Agrochemical Research Centre, Belgium	2012
S62	Chicken		
AS3	Swine		

Table 2.1: The animal and human *Staphylococcus aureus* strains that will be used for further studies along with its date of arrival at Kingston University.

2.2 Storage of Isolates

The human Isolates from Manchester Metropolitan University were transported to Kingston University February 2015 on swabs in transport medium and in a temperature controlled box with ice. Upon arrival, all of the *S. aureus* isolates (n=30) were transferred to Brain-Heart Infusion agar (BHI agar) to generate overnight mass cultures. These cultures were then streaked on BHI agar to produce single colonies. All isolates were then transferred to 1 mL aliquots of Brain-heart infusion broth, containing 15% glycerol and frozen at -80° C, for long term storage. The Belgium cultures (*S. aureus* B53, S62 & AS3) were also sent to Kingston University on swab cultures and streaked on BHI agar.

2.3 Resuscitation of Isolates prior to use

To generate an overnight culture, isolates were removed from the -80°C storage and using aseptic technique; a swab was taken on the inside of the tube and swabbed onto the BHI agar. Plates were then incubated for 14 - 16 hours at 37° C producing a mass culture. The isolate was then aseptically inoculated onto to Brain-Heart Infusion agar or inoculated in Brain-Heart Infusion Broth and was incubated for a further 14 - 16 hours at 37° C.

2.4 Isolation of Bacterial Genomic DNA

Pure genomic DNA was isolated from six *Staphylococcus aureus* strains (OP100, B51, IP70, B53, S62 & AS3) using the Bacterial Genomic DNA Kit (Sigma-Aldrich GenElute™, Dorset England) following the manufactures instructions with minor modifications. Briefly, the isolation of genomic DNA was performed as follows.

The GenElute™ Bacterial Genomic DNA kit reagents were stored at room temperature; any reagents that formed a precipitate before use were warmed at 55°C in a preheated water bath (Grant Instruments, Cambridge, UK.) until the precipitate was dissolved. For each *S. aureus strain*, 200 µL of Lysozyme Solution was prepared by dissolving 2.115 x 10⁶ unit/mL Lysozyme in 1.5 mL of Gram-positive lysis buffer. This was further supplemented with a 250 units/mL of Lysostaphin; the solution was thoroughly mixed before use. Overnight cultures of the bacteria were grown in 5 mL of Brain Heart Infusion (BHI) broth and 1.5 mL was transferred to a sterile micro-centrifuge tube and pellet by centrifuging for 2 minutes at 12,000 x g, the culture medium subsequently completely removed and discarded. The pellet was then re-suspended in 200 µL of freshly prepared Lysozyme Solution, and incubated in a preheated water bath (37°C for 30 minutes) to achieve cell wall digestion. RNase A solution (20 µL) from the kit was added and incubated for a further 2 minutes at room temperature to ensure that RNA-free genomic DNA is achieved. Proteinase K powder (10 mg) was dissolved in 0.5 mL of water to obtain a 20 mg/mL stock solution and stored at room temperature. To lyse the cell, 20 µL of Proteinase K was added to each sample, followed by 200 µL of Lysis Solution C. The mixture was mixed thoroughly using a vortex (15 seconds) and incubated in a preheated water bath (55°C for 15 minutes). Column preparation solution (500 µL) was added to each pre-assembled GenElute MiniPrep Binding Column, seated in a 2mL collection tube and centrifuged (12,000 x g for 1 minute), the eluate was subsequently discarded. This process achieved maximum binding of DNA to the membrane of the GenElute binding column. Ethanol (95 – 100%/ 200 µL) was added to the lysate mixture and thoroughly

mixed with the use of a vortex (10 seconds), ensuring a homogenous solution. The entire contents were transferred to the prepared GenElute binding column using a wide bore pipette, to reduce shearing of the DNA and centrifuged (6500 x g for 1 minute). The collection tube containing the eluate was discarded and placed in a new 2 mL collection tube. Wash Solution 1 (500 μ L) was added to the GenElute binding column and centrifuged (6500 x g for 1 minute), the collection tube and eluate was discarded and the column placed in a new 2 mL collection tube. For the second wash, the Wash Solution Concentrate bottle (20 mL) was diluted with 95 – 100% ethanol (80 mL) and mixed by inverting. Wash solution Concentrate (500 μ L) was added to the column and centrifuged (16,000 x g for 3 minutes) to dry the column and further centrifuged (16,000 x g for 1 minute) to remove any residual ethanol, ensuring an ethanol-free column. The collection tube was discarded and placed in a new 2 mL collection tube. Molecular biology grade water (200 μ L) was transferred directly to the center of the GenElute column and incubated at room temperature for 5 minutes and subsequently centrifuged (16,000 x g for 1 minute) to elute the DNA. Using the same collection tube, a further 200 μ L of molecular grade water was transferred to the center of the column and incubate for a further 5 minutes at room temperature, centrifuged (12,000 x g for 1 minute) to elute the DNA. The eluate in the collection tube contains the RNA-free genomic DNA, approximately 400 μ L and was stored at 2 – 8 °C until the next stage.

2.5 Spectrophotometric analysis

A GE Healthcare NanoVue™ Plus Spectrophotometer was used to determine the concentration and purity check for the DNA samples. The spectrophotometer was programmed to read DNA samples. Using a Gilson® P100 pipette, 0.5 µL of molecular grade water was loaded directly on the sample plate. This was used as the reference sample, to ensure that the machine was blanked and readings were accurate. The loading plate was cleaned using a sterile disposable wipe and 0.5 µL of the RNA-free genomic DNA was loaded directly on to the loading plate. The average concentration (A260/ A280) and protein concentration (A260/ A230) readings of the samples were tabulated (Table 3.3).

2.6 Gel Electrophoresis

Isolated RNA-free genomic DNA was analysed by Gel Electrophoresis. A batch of TBE 10X running buffer was prepared by adding 800mL of distilled water to 108g of tris base, 55g boric acid and 7.5g of EDTA di-sodium salt, the volume was then adjusted to 1 litre by topping it up with more distilled water and the mixture was thoroughly mixed. The agarose gel was prepared using 0.75 g agarose gel, 75 mL 1XTBE, prepared earlier and 7.5 µl SYBR safe dye, this was stirred using a magnetic stirrer and heated at 55°C for 15 mins to reach it's boiling point. The agarose gel was immediately transferred to the gel tanks, the comb was put in place and the gel was allowed to solidify at room temperature for 10 minutes. Once solidified, the comb was removed and the wells were exposed. Nothing was loaded in the side wells (1 and 10). Well 2 was loaded with a 100 – 1000 bp gene ladder and wells 3, 4, 5, 6, 7 & 8 was loaded with the DNA sample.

Gels were run in a FlowGen Gel Electrophoresis system in 1X TBE buffer at Voltage 200V for 3 hours The gel was placed in a Molecular Imager® Gel Doc™ XR System. The image was captured with a Windows computer and the Imager® software.

2.7 Ion Torrent: Next Generation Sequencing

High-quality RNA-free genomic DNA (gDNA) that was isolated from the six *Staphylococcus aureus* isolates (OP100, B51, IP70, B53, S62 & AS3) using the Sigma-Aldrich GenElute™ Bacterial Genomic DNA Kit (see section 2.4), was used with the next generation sequencing system (NGS) (Life Technologies, USA). The manufacturers instructions were followed during stage one of the Ion Xpress™ Plus gDNA Fragment Library Preparation, stage two of the Ion PGM™ Template OT2 400 Kit and the third stage of the Ion PGM™ Sequencing 400 Kit protocols, with minor modifications developed during the trial and error phase.

2.7.1 Ion Xpress™ Plus gDNA Fragment Library Preparation

2.7.1.1 Fragmentation of the DNA

The initial stage of the library preparation included the use of Ion Shear™ Plus Reagents to fragment the gDNA to the desired median fragment size of 350 - 450 base pairs (bp) for use with the 400-library kit. The Ion Shear™ Plus 10X Reaction Buffer and the Ion Shear™ Plus Enzyme Mix II were both vortexed for 5 seconds and centrifuged (using the Eppendorf centrifuge), using the pulse-spin function and placed on ice. The concentration of each *S. aureus* genomic DNA sample was adjusted to 100 ng/μL using Nuclease-free water and 10 μL was subsequently added to a 1.5 mL Eppendorf LoBind® tube. Ion Shear™ Plus 10X Reaction Buffer (5 μL) and Nuclease-free water (25 μL) were also added to the Eppendorf LoBind® tube and were mixed vigorously by vortexing and centrifuging, using the pulse spin feature. A Gilson™ P10 - P20 pipette was then used to add 10 μL of Ion Shear™ Plus Enzyme Mix II to the samples, bringing the total reaction volume to 50 μL. It was then immediately mixed using a Gilson™ P100 - P200 pipette, by pipetting up and down 10 times and the mixture was instantly incubated in a water bath set at 37°C for a total of 10 minutes, as per the protocol, to reach the desired gDNA median fragment size of 350 - 450 bp. After

incubation, the Ion Shear™ Stop Buffer (5 µL) was immediately added to each sample and thoroughly vortexed for 5 seconds to stop the reaction, and placed on ice.

2.7.1.2 Clean up using Agencourt® AMPure® XP Reagents

Agencourt® AMPure® XP Reagents (99 µL) was added to the sheared DNA samples. A pipette was used to mix the bead suspension, by pipetting the mixture up and down 5 times. It was centrifuged with the pulse spin feature and incubated for 5 minutes at room temperature. The tubes were placed in a magnetic DynaMag™-2 rack for 3 minutes and the supernatant was carefully removed without disturbing the bead pellet. With the tube still in the magnetic rack, 500 µL of freshly prepared 70% ethanol (4.9ml of 100% ethanol added to 2.1 mL MilliQ, to prepare 7ml of 70% ethanol) was added to the samples and incubated for 30 seconds, turning the tube around twice in the magnet rack to ensure that the beads move around. Once the solution had cleared, the supernatant was removed and discarded without disturbing the pellet and left to air-dry on the rack at room temperature for 5 minutes. The tubes were then removed from the magnetic rack and 25 µL of Low TE buffer was added directly to the pellet to disperse the beads. It was then mixed by pipetting the suspension up and down 5 times and vortexing the sample for 10 seconds and placed on the magnetic rack for 1 minute, in which time the solution cleared. The supernatant contains the eluted DNA and was transferred to a new 0.2-mL PCR tube without disturbing the pellet, removing 2 µL from each sample for the Bioanalyser® instrument analysis and storing the remaining DNA samples at -10°C to -30°C.

2.7.1.3 Bioanalyser

The fragment size of the eluted DNA (from step 2.7.1b) was checked using the Agilent® 2100 Bioanalyser® instrumentation and Agilent High Sensitivity DNA Kit which will confirm the desired DNA fragment size range. Before beginning the chip preparation protocol, the Chip Priming Station's base plate was adjusted to suit the High Sensitivity DNA Chip and the syringe clip was adjusted to the lowest position (figure 2.1a). The Agilent® 2100 Bioanalyser® was also adjusted to accommodate the DNA Chip by removing the pressure cartridge and inserting the electrode cartridge. The 2100 Expert software that corresponds with the Agilent® 2100 Bioanalyser® instrument was calibrated so that it communicated with the High Sensitivity DNA Chip that was inserted. The Agilent® High sensitivity DNA kit includes a High Sensitivity DNA dye concentrate and also a High Sensitivity DNA gel matrix, both were equilibrated at room temperature (30 mins) ensuring that they are protected from any light sources. Care must also be taken with the High Sensitivity DNA dye-concentrate as it contains dimethyl sulfoxide (DMSO), a potential mutagen that requires the aid of gloves. The vial containing the High Sensitivity DNA dye concentrate is then mixed using a vortex for 10 seconds and using a pipette, 15 µL is added to the High Sensitivity DNA gel matrix vial. The complete gel-dye mix was transferred to the top receptacle of a spin filter and placed in a micro-centrifuge (6000 rpm) for 10 minutes at room temperature. The filter was discarded and the gel-dye mix was stored away from light and equilibrated for 30 minutes at room temperature. A High Sensitivity DNA chip (shown in figure 2.10) was then placed on the chip priming station at position C. Using a pipette, 9.0 µL of the gel-dye mix was dispensed to the bottom of the well marked (G), ensuring that the no air bubbles were formed. The plunger of the syringe on the chip priming station was set to 1 mL. The plunger was subsequently pressed until it was held by the latch, to pressurise the DNA chip and released after exactly 60 seconds. The plunger was visually inspected to ensure that it moved back to the 0.3 mL. Waiting a further 5 seconds, the plunger was slowly pulled back to the 1 mL position on the chop priming station. The gel-dye mix (9.0 µL) was transferred into each of the

wells marked **G**. The High Sensitivity DNA marker (5 μL) was transferred into the wells marked with the ladder symbol and the remaining 11 sample wells, leaving no wells empty. The DNA samples (1 μL) were then loaded in to each of the individual sample wells, noting the location. The High Sensitivity DNA ladder (1 μL) was subsequently loaded into the wells marked with the ladder symbol and the sample wells that have the DNA.

The DNA chip was then placed horizontally in the adapter of the vortex (2000rpm for 60 seconds) and placed carefully in the receptacle of the Agilent® 2100 Bioanalyser® instrument, closing the lid. The computer displayed the DNA chip, and the names of the samples were added to the corresponding well that were noted previously. The Bioanalyser® was run overnight to perform size fractionation and quantification of the DNA samples. Once the assay was complete the chip was removed and the cleaning cycle was run. Using a new electrode cleaner after each run, the wells were filled with deionised analysis-grade water (350 μL) and placed in the instrument for 10 seconds, closing the lid. Once complete, it was immediately removed to allow the water on the electrodes to evaporate by air drying for a further 10 seconds, before closing the lid.

2.7.1.4 Ligate Adapters

The following reagents, provided in the Ion Xpress™ Barcode Adapters Kit were added to a 0.2-mL PCR tube for a total reaction volume of 100 μL : fragmented DNA sample (25 μL), 10x Ligase Buffer (10 μL), Ion P1 Adapter (10 μL), dNTP Mix (2 μL), Nuclease-free water (31 μL), DNA Ligase (4 μL), Nick Repair Polymerase (8 μL) and the Ion Xpress™ Barcodes X (10 μL) (33 to 38). When handling the IonXpress™ Barcoded adapters, gloves were changed frequently to avoid cross contamination as the adapters were specific to each sample. The specific barcode adaptor used were as follows; OP100: 33, B51: 34, IP70: 35, B53: 36, S62: 37 and AS3: 38. The PCR tubes were subsequently placed in the C1000 Thermal Cycler and programed to run the sequence shown in table 2.2.

Stage	Temperature	Time
Hold	25C	15 minutes
Hold	72C	5 minutes
Hold	4C	Hold

Table 2.2: The thermal cycler was programmed to run the sequence shown in the table for this stage.

Once complete, the entire mixture from the PCR tube was added to a 1.5-mL Eppendorf LoBind® tube and was cleaned using the Agencourt® AMPure® XP Reagents described in section 2.7.1.2, using 100 µL as the reaction volume.

2.7.1.5 E-Gel® SizeSelect™ Agarose Electrophoresis

The E-Gel® iBase™ unit and E-Gel® Safe Imager™ were assembled and the “SizeSelect™ 2%” program loaded onto the system. The E-Gel® cassette was taken out from the package, the comb removed from the gel and firmly inserted into the apparatus so that a steady light illuminated. The DNA ladder (cat. No. 10416-014), suitable for 400-base read libraries was diluted in Low TE buffer to achieve a 1:40 dilution. Using a pipette, 10 µL of the diluted DNA ladder was carefully added to the middle well (Figure 2.1), ensuring that the tip does not pierce the agarose at the bottom of the well. In the same fashion, ligated DNA samples (20 µL) was added to the top loading wells numbered #2, #3, #4, #5, #6 & #7 and deionised water was added to the wells marked #1 and #8. Deionised water (25 µL) was also added to the bottom collection well and 10 µL to the well marked M on the bottom row.

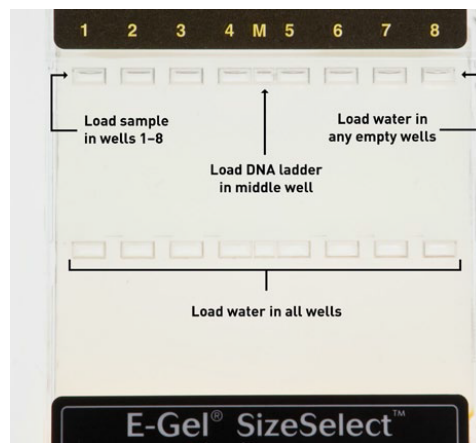


Figure 2.1: An example of the E-Gel® SizeSelect™ gel that was used.

Once loaded, the amber filter was placed over the E-Gel® iBase™ unit and the electrophoresis was initiated, leaving it to run for 16 - 20 minutes, allowing for more time if needed at the end of the run if the band had not reached the reference line. Once the band reached the reference line, the collection well was refilled with $\cong 10$ - μL of nuclease-free water and the E-Gel® was run for a further 2 - 2.5 minutes, stopping once the band of interest had migrated into the collection well. Once complete, the solution was collected using a pipette, without piercing the bottom of the well. The wells were filled with a further 10 μL of nuclease-free water to wash and rinse the collection wells and collected, the total recovered volume from each well was $\cong 60$ μL .

2.7.1.6 Amplify and Purify Libraries

The following reagents: Platinum® PCR SuperMix High Fidelity (200 μL), Library Amplification Primer Mix (10 μL) and the DNA Library (50 μL) were combined in a vial and thoroughly mixed by pipetting up and down. The 260 μL reaction mixture was equally split placed into 0.2-mL PCR tubes and subsequently placed in a thermal cycler following the cycling program indicated in the table 2.3 below.

Stage	Step	Temperature	Time
Holding	Denature	95°C	5 mins
Cycling x 5	Denature	95°C	15 sec
	Anneal	58°C	15 sec
	Extend	70°C	1 min
Holding	-	4°C	Hold

Table 2.3: The thermal cycler was programmed to run the sequence shown in the table, cycling for a maximum of 5 times.

Once complete the entire reaction mixture (260 μ L) was added to a new 1.5-mL Eppendorf LoBind® tube and underwent the clean up step using the Agencourt® AMPure® XP Reagents described in step 2.7.1b. The supernatant which contains the final amplified library was transferred to 3 aliquots and stored at -30°C to -10°C, to reduce the number of freeze-thaw cycles which could potentially damage the DNA.

2.7.1.7 Quantify and Pool barcoded libraries

The barcoded libraries were individually quantified and pooled in equal-molar amounts to ensure equal representation of each library, initially there were several pools at different concentrations. The Bioanalyser® software was used to assess the barcoded library samples (1 μ L) at the highest possible equal-molar concentrations.

2.7.2 Ion PGM™ Template OT2 400 Kit

2.7.2.1 Configuring the machine

The IonOneTouch™ 2 Instrument layout is shown in Figure 2.2 below. Using the materials provided in the Ion PGM™ Template OT2 Supplies 400 Kit and the Ion PGM™ Template OT2 Solutions 400 Kit, the machine was configured.

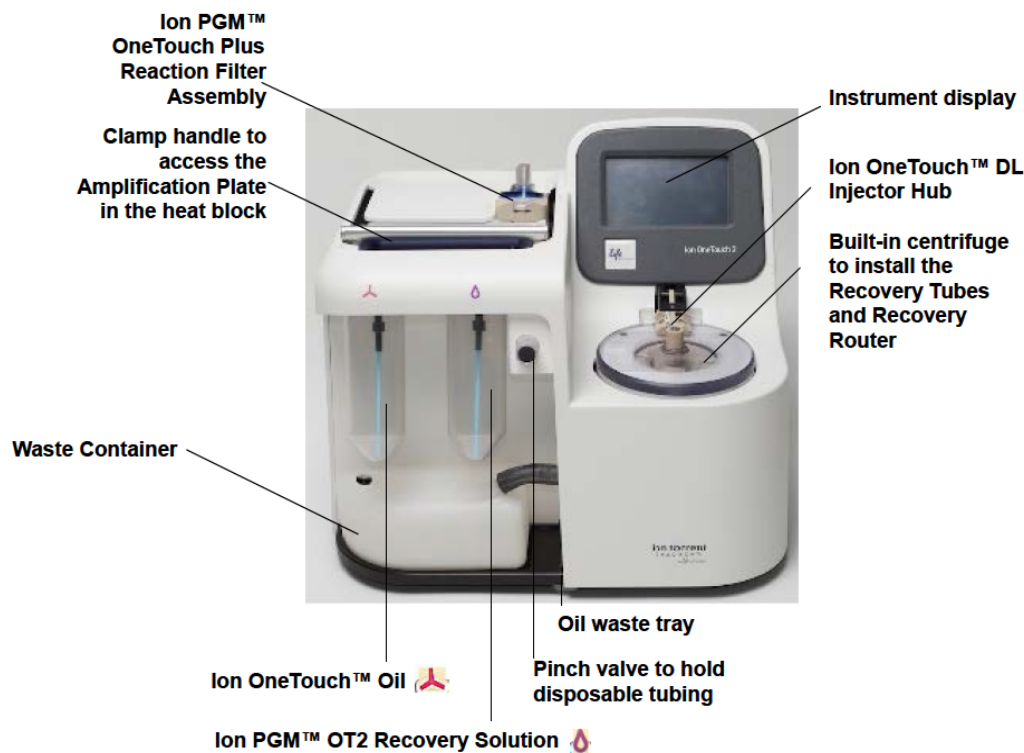


Figure 2.2: Layout of the IonOneTouch™ 2 Instrument. Image taken from the manufacturer manual.

The touch screen display on the instrument was used to open the lid on the centrifuge. The Ion OneTouch™ Recovery Tubes and Ion OneTouch™ Recovery Router were installed into each of the slots in the centrifuge (Figure 2.3), ensuring that it was seated flat and securely in place, the lid of the centrifuge was subsequently closed.

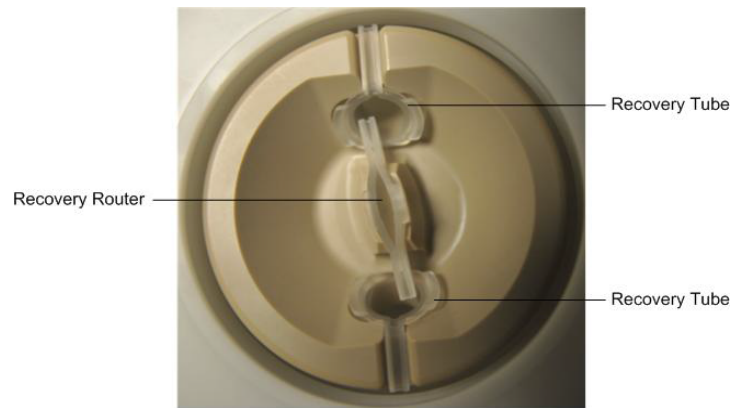


Figure 2.3: An enhanced view of the IonOneTouch™ 2 built in centrifuge with the replaceable recovery tubes (x2) and recovery router (x1) shown *in situ*. Image taken from the manufacturer manual.

The used Ion oneTouch™ 2 Cleaning Adapter was removed from the system and the handles were pushed back to open the heat block. An IonOneTouch™ Amplification Plate, which has a connected disposable tubing and disposable injector, was installed by inserting the amplification plate into the heat block, threading the tubing through the Ion OneTouch™ DL Tubing Catch, closing the handle on the heat block, inserting the disposable tubing into the pinch valve (Figure 2.4) and finally inserting the disposal injector into the port of the Ion OneTouch™ DL injector hub so that it stops at the base of the router (Figure 2.5).



Figure 2.4: The disposable tubing in the pinch valve with the handle on the heat block closed. Image taken from the manufacturer manual.

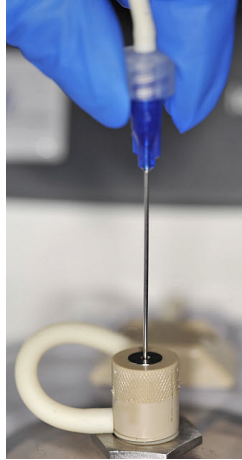


Figure 2.5: The disposable injector being inserted into the injector hub at the base of the router. Image taken from the manufacturer manual.

Using a new reagent tube from the kit and wearing fresh gloves, the Luer-Lok® was attached to the end of a new IonTouch™ Sipper tube to the *left* port of the instrument, ensuring that the Sipper Tube did not touch any surface or become contaminated. The Ion OneTouch™ Oil bottle was inverted 3 times to mix and the Reagent Tube was filled half-way with the oil, minimising any bubbles. The Ion OneTouch™ Reagent Tube with the Ion OneTouch™ Oil was firmly screwed into the front *left* port, one quarter turn on the instrument. The Recovery Solution was heated in a water bath (30°C) water bath until the solution was clear. Using a new Reagent Tube from the kit the Luer-Lok® was attached to the end of a Ion Touch™ Sipper tube to the *right* port of the instrument. The Recovery Solution bottle was inverted 3 times to mix and the Reagent Tube was filled with the solution, quarter-full. The filled Reagent Tube was attached to the *right* port and screwed firmly into place in. The Waste Container was also emptied. The Ion PGM™ Template OT2 400 Reagents Mix and Ion PGM™ Template OT2 400 PCR Reagent B were incubated at room temperature and mixed by vortexing for 30 seconds and 60 seconds, respectively. Both reagents were centrifuged for 2 seconds and visually inspected to ensure so that no residual precipitate was present at the bottom of the tube, they were subsequently stored at room temperate. The Ion PGM™ Template OT2 400 Enzyme Mix

and Ion PGM™ Template OT2 Reagent X were centrifuged for 2 seconds and placed on ice and the PGM™ Template OT2 400 Ion Sphere™ Particles (ISPs) at room temperature.

An aliquot of the barcoded DNA library was incubated at room temperature and 2 µL was added to 23 µL nuclease-free water. The diluted library was mixed by vortexing for 5 seconds, centrifuged for a further 2 seconds. Using a new 1.5 mL Eppendorf LoBind® Tube the following reagents were added: Ion PGM™ Template OT2 400 Reagent Mix (500 µL), Ion PGM™ Template OT2 400 PCR Reagent B (285 µL), Ion PGM™ Template OT2 400 Enzyme Mix (50 µL), Ion PGM™ Template OT2 400 Reagent X (40 µL) and the diluted library (25 µL) to create a total amplification solution mix. The amplification solution was mixed by pipetting up and down, vortexing for 5 seconds and centrifuging for a further 2 seconds. The ISPs were mixed at maximum speed by vortexing for 1 minute and centrifuging for a further 2 seconds to re-suspend the particles, 100 µL was immediately added to the amplification solution and the total amplification solution (1000 µL) was mixed using the vortex for 5 seconds.

2.7.2.2 Loading Ion PGM OneTouch Plus Reaction Filter

A new Ion PGM OneTouch Plus Reaction Filter Assembly was placed into a tube rack so that the 3 ports of filter face up (Figure 2.6).

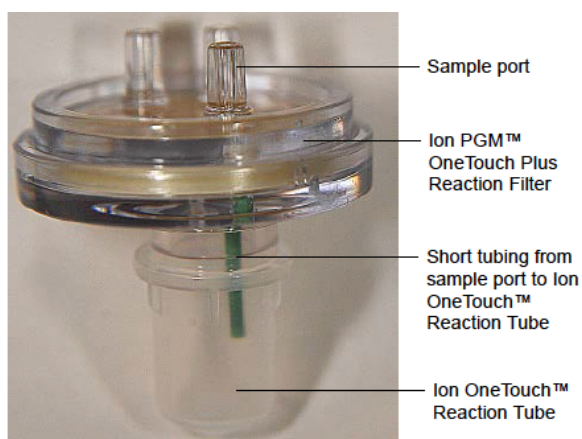


Figure 2.6: An example of the Ion PGM OneTouch Plus Reaction Filter Assembly that was used. Image taken from the manufacturer manual.

Using a Gilson P1000, the amplification solution (1000 μ L) was slowly transferred into the Reaction Filter. This was achieved by ensuring that the tip of the pipette was inserted into the sample port forming a tight seal, once transferred the plunger of the pipette was kept depressed to avoid aspirating the solution from the reaction filter. Using a new tip each time and in the same process described, 1000 μ L and then 500 μ L of the Ion OneTouch™ Reaction Oil was also transferred into the sample port of the Reaction Filter. The Reaction Filter was subsequently oriented in a specific way; first the Reaction Filter was lifted out of the tube rack so that the sample port was on the *left*, it was then rotated to the *right* until the Reaction Tube was inverted and the 3 ports of the Filter face *down* (Figure 2.7), and placed firmly into the three holes on the top stage of the Instrument.

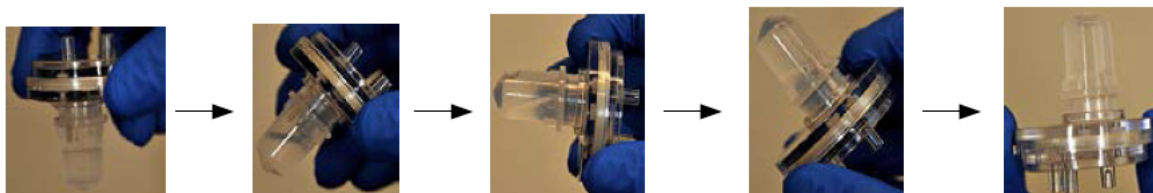


Figure 2.7: A representation of how the Ion PGM OneTouch Plus Reaction Filter was inverted, concluding with the 3 ports of the filter facing down. Image taken from the manufacturer manual.

The centrifuge lid was closed on the OneTouch™ 2 Instrument and the *touch* display screen was used to load the specific 'PGM: Ion PGM™ Template OT2 400 Kit' program from the drop down menu and the on screen instructions were followed to initiate a run, lasting for ≤ 16 hours as per the manufactures instructions.

2.7.2.3 Recover the template-positive Ion PGM™ Template OT2 400 Ion Sphere™ Particles

Before removing the sample, the instrument was used to perform a 10-minute centrifuge run. Once complete, the Recovery Router was removed and discarded, and the Recovery Tubes from the instrument were placed in a tube rack. Using a pipette, all but 100 μL of the Ion OneTouch™ Recovery solution was removed from each Recovery Tube, withdrawing the supernatant from the surface and on the opposite side from the pellet, removing any white flocculent material and ensuring that the ISP pellet was not disturbed. With a new tip and using the same tip for both tubes, the ISPs were resuspended in the remaining Ion OneTouch™ Recovery solution by pipetting the pellet up and down until each pellet dispersed in the solution. Ion OneTouch™ Wash Solution (500 μL) was then added to each Recovery Tube and mixed by pipetting to disperse the ISPs. The suspension was subsequently transferred to a new 1.5-mL Eppendorf LoBind® Tube and heated, using a heat block (50°C) for 2 minutes followed by centrifuging (15,000 x g) for 2.5 minutes. A pipette was then used to remove all but 100 μL of the Wash Solution from the tube, withdrawing the supernatant from the surface and on the opposite side of the pellet.

A 8-well strip (Figure 2.8) was obtained from the Ion OneTouch™ ES Supplies Kit and orientated so that the square-shaped tab on the 8-Well strip was on the left. The suspension from the tube (100 μL) was transferred into Well-1 and a small aliquot of the sample was retained for quality assessment using a Qubit® 2.0 Fluorometer.

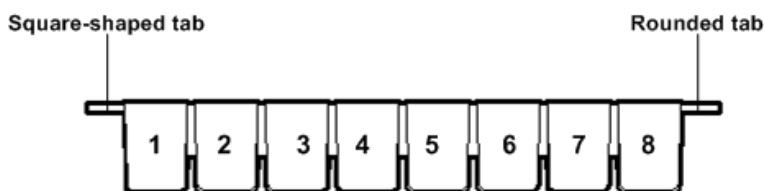


Figure 2.8: A representation of the 8-well strip that was used.

2.7.2.4 Enrich the template-positive Ion PGM™ Template OT2 400 Ion Sphere™ Particles

The tube containing the Dynabeads® MyOne™ Streptavidin C1 Beads was resuspended by vortexing (30 seconds) and centrifuging (2 seconds). The tube was then opened and using a pipette, the beads were further mixed by pipetting the mixture up and down and 13 µL was immediately transferred to a new 1.5-mL Eppendorf LoBind® Tube. The tube was placed on a DynaMag™-2 magnet rack for 2 minutes and the supernatant was carefully removed without disturbing the pellet, and discarded. MyOne™Beads Wash Solution (130 µL) was added to the tube and removed from the magnet rack, vortexed (30 seconds), centrifuged (2 seconds) and subsequently dispensed into the 8-well strip mentioned in Section 2.7.2.3 Well-2. Ion OneTouch™ Wash Solution (300 µL) was transferred to Well-3, Well-4 and Well-5. Fresh melt-off solution was prepared by combining Tween® Solution (280 µL) and 1 M NaOH (40 µL) and 300 µL was transferred to well-7. Well-6 remained empty. The Ion OneTouch™ ES instrument was set up according to the manufacturer's instruction, the 8-Well strip was then inserted into the slot of the Ion OneTouch™ ES, ensuring that the square-shaped tab sits on the left. A new tip was placed in the instrument's Tip Loader arm and a new, opened 0.2 mL PCR tube was loaded into its hole. Before the run was performed, the contents of Well-2 was mixed with the aid of a pipette, being careful not to introduce bubbles into the solution. The run took 35 minutes, and at the end when the alarm sounded, the PCR tube was immediately closed and the enriched IPSs were stored at 2°C to 8°C.

2.7.3 Ion PGM™ Sequencing 400 Kit

Figure 2.9 below shows the Ion PGM™ System with Reagent and Wash Bottles attached.

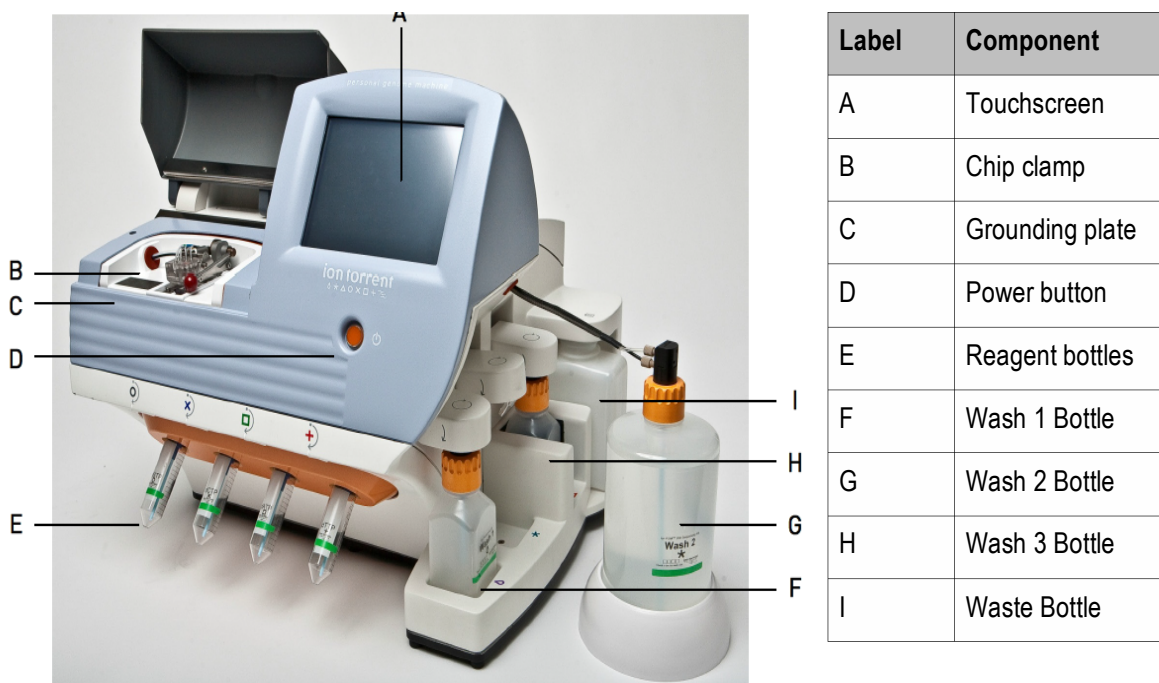


Figure 2.9: The layout of the Ion PGM™ instrument with the reagent and wash bottle. Image taken from the manufacturer manual.

The Torrent Server connected to the Ion PGM™ System is accessible through the Torrent Browser which was accessed to create a planned run, using the planned run wizard. Briefly, in the *Application tab*: whole genome sequencing with a forward run type was selected. *Kits tab*: Ion PGM™ Template OT2 400 Kit was selected and the strain names with the corresponding Ion Xpress™ Barcode Adapter was entered. *Monitor tab*: Monitor the progress of the run and enter a name for the new planned run. Once the plan was created, it was loaded onto the Ion PGM™ Sequencer instrument.

Before starting the run, the cleaning log was checked to ensure that the Ion PGM™ System was cleaned as per the manufactures instructions. The dNTP stock solutions (Ion PGM™ Sequencing 400 dGTP, Ion PGM™ Sequencing 400 dCTP, Ion PGM™ Sequencing 400 dATP and Ion PGM™ Sequencing 400

dTTP) were removed from the freezer and placed on ice to thaw slowly. Wash Bottle 1, Wash bottle 2 and Wash bottle 3 were obtained from the Ion PGM™ Sequencing 400 Kit and rinsed three times with 18 MΩ water. 100 mM NaOH (350 μL) was added to Wash Bottle 1; 2 litres of 18 MΩ water, entire contents of the Ion PGM™ Sequencing 400 W2 solution, and freshly prepared 100 mM NaOH (140 μL) was added to wash bottle 2, the bottle was capped and inverted 5 times to mix; and Ion PGM 1X W3 solution was added to the 50 mL line marked on Wash 3 Bottle. The old sipper tubes from the dNTP ports and the cleaning chip that was used to clean the Ion PGM™ system remained in place during the initialisation steps, and were only changed when instructed to do so. The barcode on the Ion PGM™ Sequencing 400 W2 Bottle was scanned and the corresponding Ion PGM™ Sequencing Kit was loaded on the Ion PGM™ Sequencer. The touch screen display then started the initialisation wizard. The on-screen instructions were followed to verify the gas pressure, once verified and wearing new gloves each time, new sipper tubes were firmly attached to the caps in the three wash bottles, making sure that the sipper tube did not touch any other surface to avoid contamination. The Ion PGM™ System tested the bottles for leaks, fill the Wash 1 Bottle, then adjust the pH of the W2 Solution, this process is automated. After each dNTP stock solution had thawed, it was vortexed and centrifuged to collect the contents, and put back on ice. Four new Reagent Bottles were labelled as dGTP, dCTP, dATP and dTTP, using filtered pipettes and changing gloves after handling every dNTP stock solution, 20 μL of each dNTP stock solution was transferred to its respective labelled bottle. After the wash solutions had initialised, the onscreen prompts instructed to remove the used sipper tubes and collection tray from the dNTP ports, and wearing gloves, new sipper tubes were inserted into each dNTP port. The Reagent Bottle was firmly attached to its corresponding correct dNTP port, ensuring that it was not over tightened. The Ion PGM™ System will check the pressure of the Reagent Bottles and fill each Reagent Bottle with 40 mL of W2 solution, an automated process. At the end of the initialisation, the Ion PGM™ System measured the pH of the Reagents and if every Reagent is in the target pH range, a green Passed screen was displayed.

The contents of the the tube contains the enriched ISPs were mixed by thoroughly pipetting it up and down and centrifuging for 2 minutes (15,000 X g). The supernatant was carefully removed without disturbing the pellet and 15 µL of the enriched ISPs were left in the tube, this was achieved by visually comparing it to an identical tube loaded with 15 µL of liquid. Sequencing Primer (12 µL) was added to the tube and the contents were thoroughly mixed by pipetting up and down gently to disrupt the pellet. The total volume of the tube was 27 µL, this was added to a thermal cycler programmed at 95°C for 2 minutes and then 37°C for a further 2 minutes, using the heated lid option. After cycling, the reaction mixture remained in the cycler at room temperature until the chip check was preformed.

2.7.3.1 Chip Check

Chip Check test ensures that it is functioning properly prior to loading the samples. A new chip was removed from its packaging and labelled to identify the experiment. The chip was placed on the Ion PGM™ Sequencer grounding plate to ensure that the chip will not be damaged from any static electricity. The person operating the chip also grounded their self. The cleaning chip was removed and the new chip was placed in the chip socket, and the chip clamp was closed. Chip check was initiated on the touch screen display, and once complete, the chip was removed from the chip socket, placed on the grounding plate and replaced with the cleaning chip.

2.7.3.2 Loading the Chip

The ISPs were removed from the thermal cycler and 3 μL of Ion PGM™ Sequencing Polymerase was added to it, bringing the total final volume up to 30 μL . This solution was thoroughly mixed by pipetting the mixture up and down and incubated at room temperature for 5 minutes. The chip (Figure 2.10) was tilted 45 degrees so that the loading port became the lower port and a pipette tip was inserted to remove as much liquid as possible. The chip was then placed upside down in the centrifuge adapter bucket in a MiniFuge with the chip tab pointing *in*, it was balanced with another used chip of the same type and was subsequently centrifuged for 5 seconds, ensuring that the chip was completely empty of any liquid.

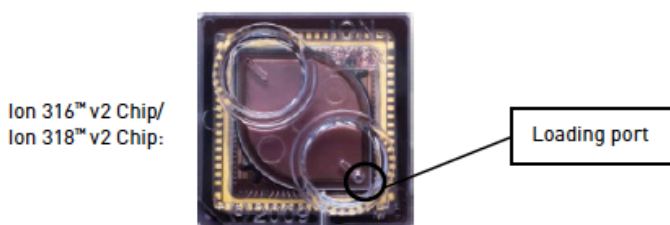


Figure 2.10: A picture representing the - Ion 316™ Chip v2 that was used. Image taken from the manufacturer manual.

The centrifuge adapter bucket was then removed from the MiniFuge and placed on the bench top, a flat stable surface. The entire $\sim 30 \mu\text{L}$ was then collected in a pipette tip, which was inserted into the loading port of the chip. The pipette was then dialled down, extremely slowly, at the rate of $\sim 1 \mu\text{L}$ per second which avoids the introduction of bubbles into the chip, leaving a small amount of sample ($\sim 0.5 \mu\text{L}$) in the tip of the pipette. The chip was then transferred to the MiniFuge, and was centrifuged for 30 seconds. The chip was removed from the bucket and tilted 45 degrees, and using a pipette, with the volume set to 25 μL , the tip is inserted into the loading port and without removing the tip, the sample was slowly pipette in and out of the chip 3 times, to avoid bubbles. This chip mixing method was repeated, this time pipetting the sample in and out of the chip five times. The chip was then tilted to a 45-degree angle and the liquid was slowly removed from the loading port. A quick 5 second spin with the tab pointing *out* was performed to remove any additional

liquid. The chip was also lightly and rapidly tapped against the bench a few times to discard any final liquid that may have been present.

The planned run on the Ion Torrent PGM™ instrument was initiated and when prompted by the instrument, the chip was loaded and clamped into place. Monitoring the run and the output of the sequencing run was made available on the Ion Torrent Server via the Ion Torrent Browser.

2.7.3.3 Ion Torrent

Once the Ion Torrent had completed the sequencing, the output of the sequencing runs was made available on the Ion Torrent server via the Ion Torrent browser. It was possible to download the sequence files as either .BAI, .BAM, .UBAM or FASTQ format.

2.8 CLC Bio

CLC Bio Genomics Workbench (version 8.5.1), a software package that analyses and visualises data from all major Next Generation Sequencing platforms, including Ion Torrent was used to map sequence reads onto reference *S. aureus* genome sequences.

2.8.1 Uploading files to CLC Bio

The downloaded .FATSQ files for the six MRSA genomes were uploaded to the CLC Genomics workbench platform, as shown in Figure 2.11.

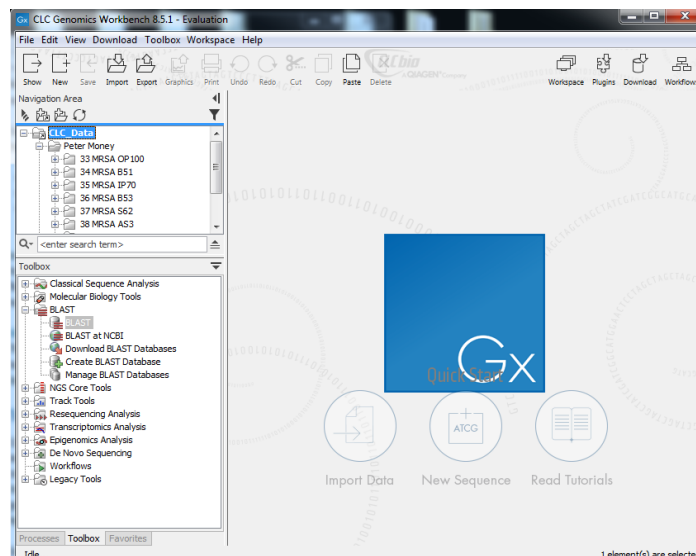


Figure 2.11: The CLC Genomics Workbench platform with the novel MRSA genomes uploaded.

2.8.2 Read Mapping

The read mapping functionality of the CLC Genomic Workbench was initially used. Each un-assembled genome was mapped to both MRSA NCTC 8325 and MRSA MW2 reference strains which were also annotated. Mapping to NCTC 8325 reference strain with a known gene and its corresponding consensus sequence is shown in Figure 2.12.

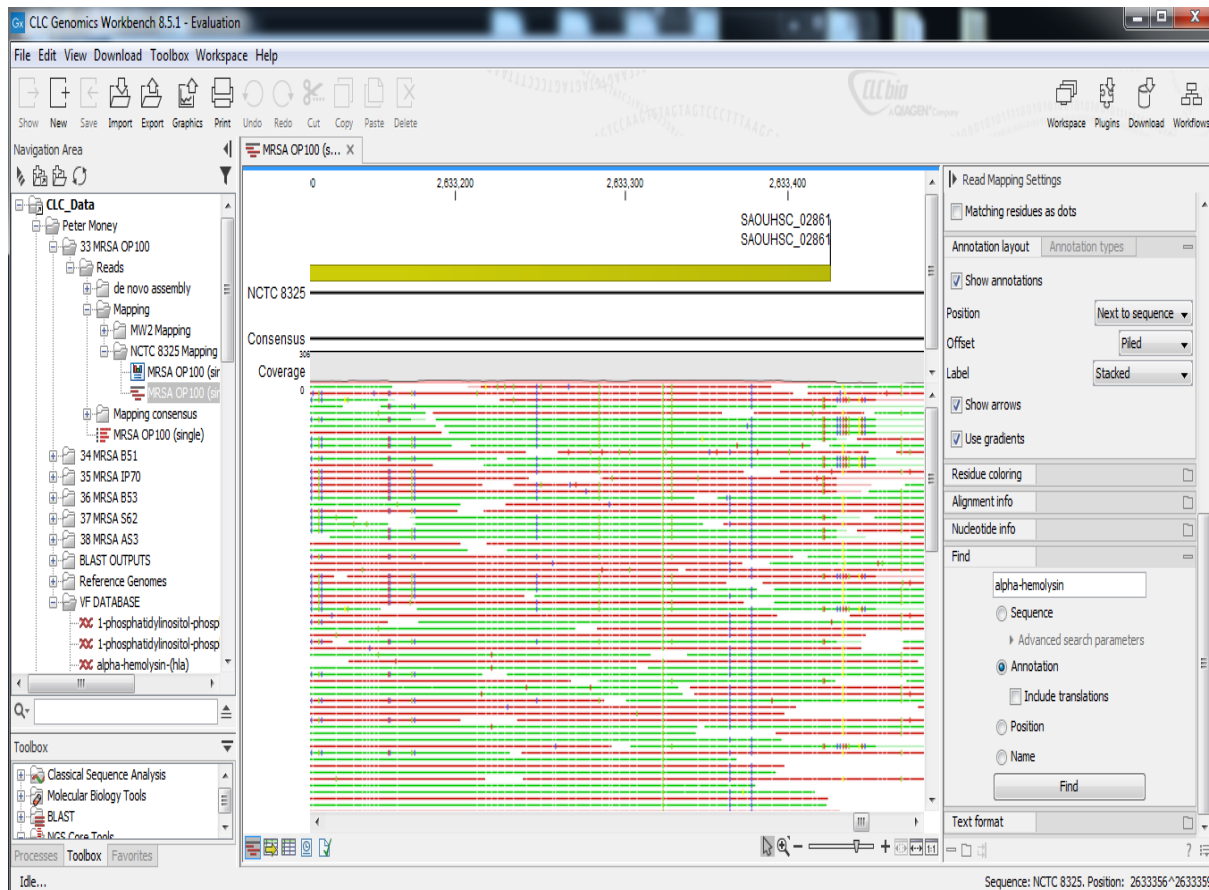


Figure 2.12: An output of the the NCTC 8325 reference strain that has been mapped to a gene (SAOUHSC_02861) and the consensus sequence of the MRSA OP100 isolates displayed.

2.8.3 Virulence factor database

A virulence factor database was created in the CLC Genomics workbench package. The virulence factors were uploaded to the software in .FASTA format, as shown in Figure 2.13

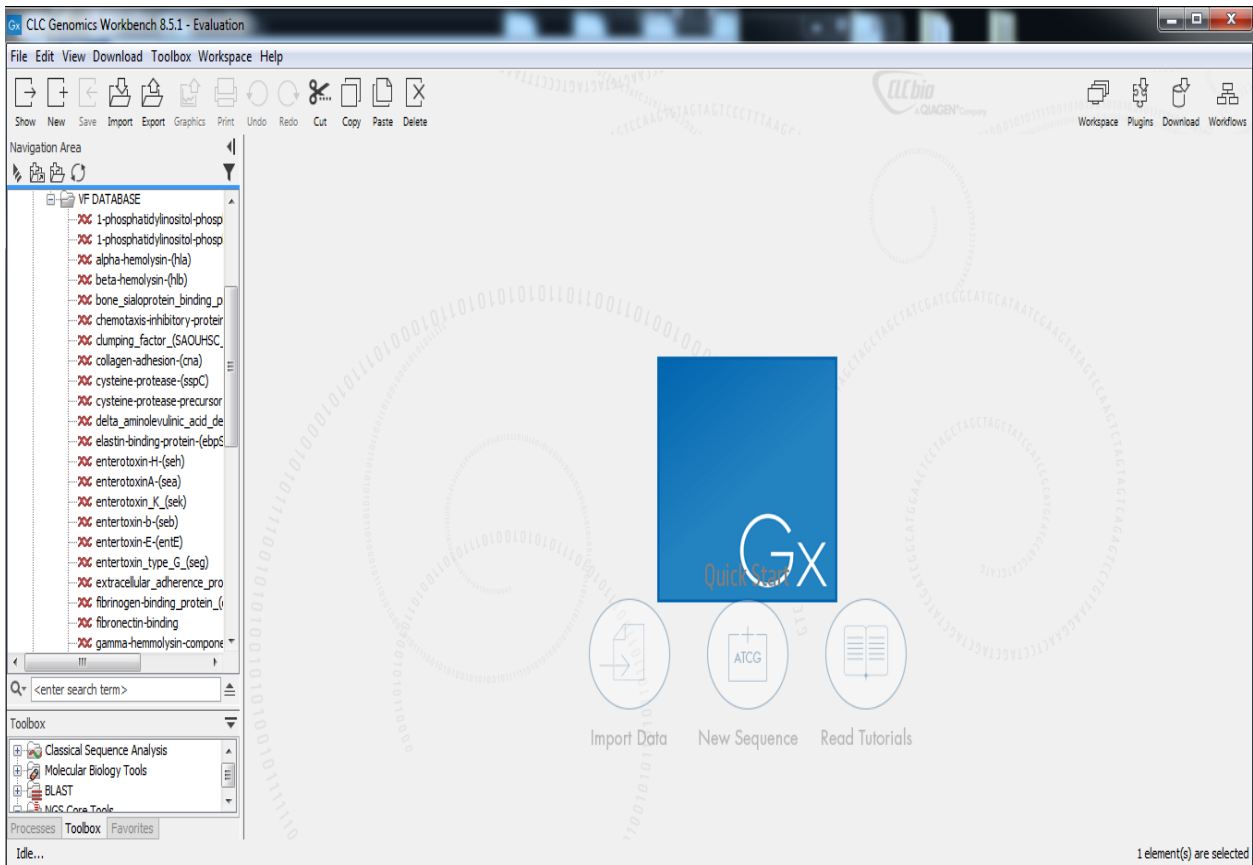


Figure 2.13: CLC Genomics workbench software showing the uploaded virulence factor genes.

2.8.4 Extracting virulence factors.

Because the reference strains are annotated; the virulence factors can be located on the consensus sequence by simply searching for it using its name or symbol (or both) (Figure 2.14). There is also another option to search for the query virulence gene. It is also possible to use BLAST as a method for finding the gene.

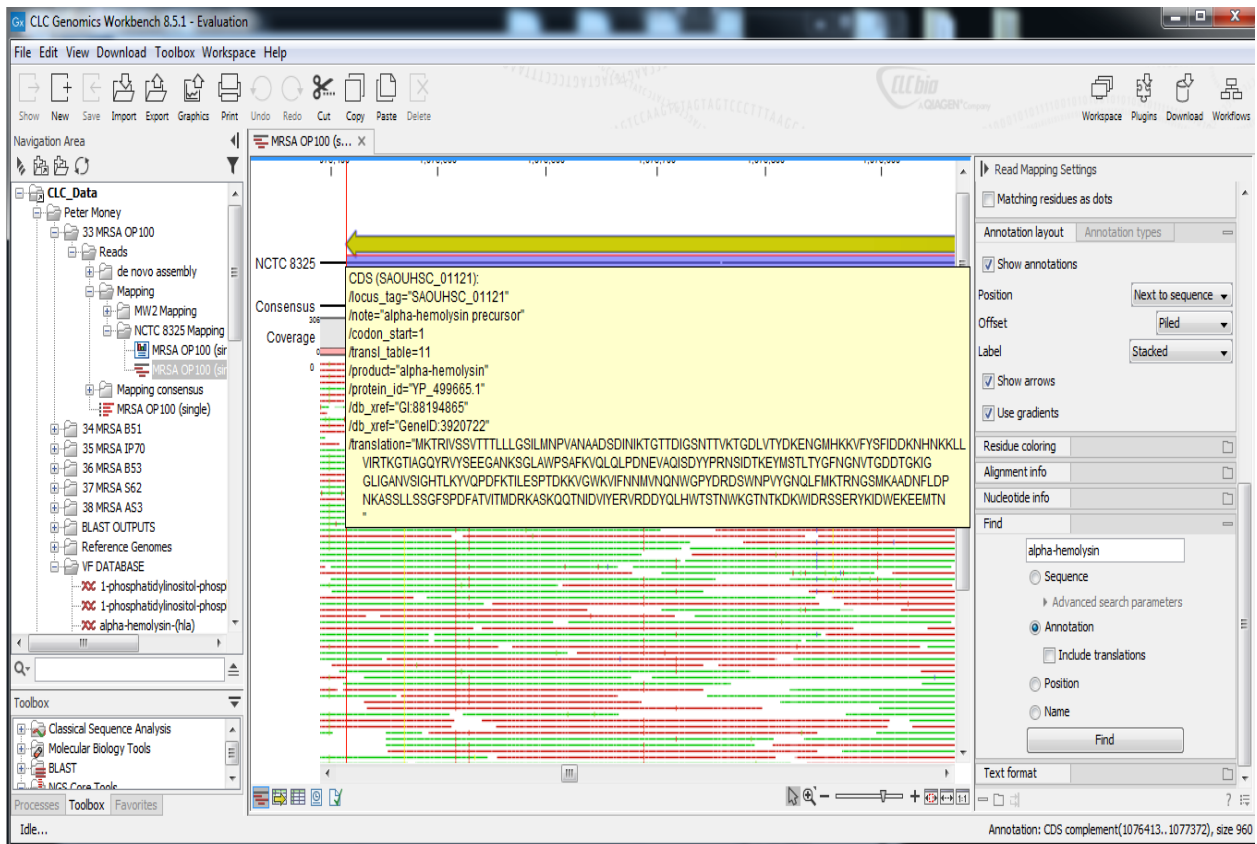


Figure 2.14: The search box of the software package with the 'alpha-hemolysin' gene as the query gene, and an output of the annotated consensus sequence that corresponds to this gene. This consensus sequence can be extracted for further analysis.

2.9 GenBank®

GenBank® (<http://www.ncbi.nlm.nih.gov/genbank>) an online genetic sequence database that contains annotated collections of publicly available DNA sequences and its corresponding bibliographic information was used. GenBank® is built and distributed by the National Centre for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov>), a division of the National Library of Medicine (NLM) (<https://www.nlm.nih.gov>), located at the National Institutes of Health (NIH) in the USA. GenBank® is part of the International Nucleotide Sequence Database Collaboration (INSDC) (<http://www.insdc.org>) that partners with the DNA DataBank of Japan (DDBJ) (<http://www.ddbj.nig.ac.jp>) and the European Molecular Biology Laboratory (EMBL) (<http://www.ebi.ac.uk/ena>) (Benson *et al.*, 2012).

NCBI builds its GenBank® database primarily from the submission of sequence data from numerous research institutions. NCBI, 2015 reports that with the publication of 'Release 3' it brought 606 sequences into the public domain. The records for whole-genome shotgun (WGS) sequences projects processed at GenBank, began with 'Release 129.0' included the publication of 172768 WGS sequences and 16769983 total sequences (NCBI., 2015). To date (October 2015), with the publication of 'Release 210.0', the database stores an overwhelming 188,372,017 non-WGS sequences and 309,198,943 WGS sequences. For downloading purposes, the uncompressed GenBank flat files for this release requires approximately 742 GB of storage space (NCBI., 2015).

2.9.1 Acquiring the Bacterial Genome

A thorough and detailed search of GenBank® was conducted. In the Genome section of GenBank, *Staphylococcus aureus* was used as the keyword in the search box. *Staphylococcus aureus* strains that have

had their genomes mapped and published were found, and the sequences were presented as a dendrogram (Figure 2.15).

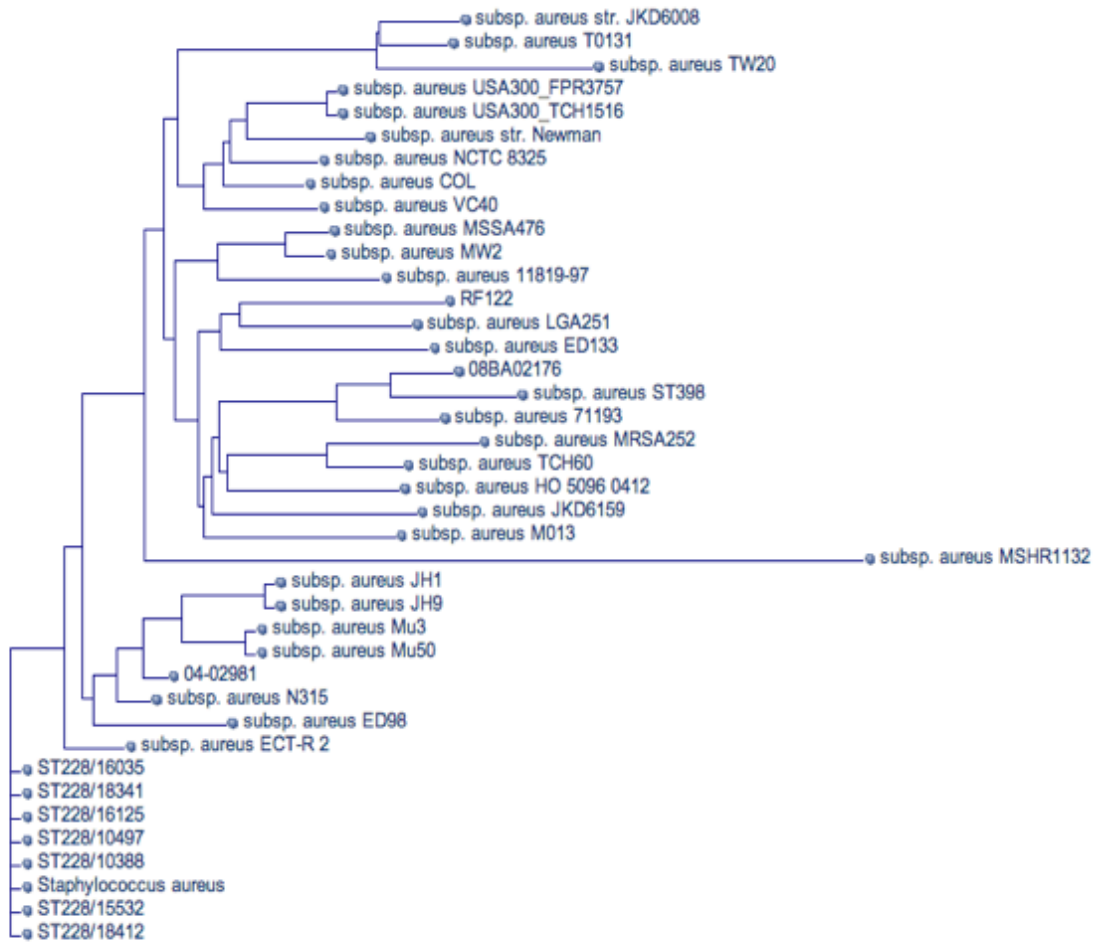


Figure 2.15: A dendrogram showing the *Staphylococcus aureus* genomes that have been mapped.

S. aureus subsp. aureus ED133 was chosen because it was a representative live-stock associated strain that was of interest in the literature. By selecting the corresponding strain name from the dendrogram, the genome assembly and annotation report are available to download in various formats (Figure 2.16).

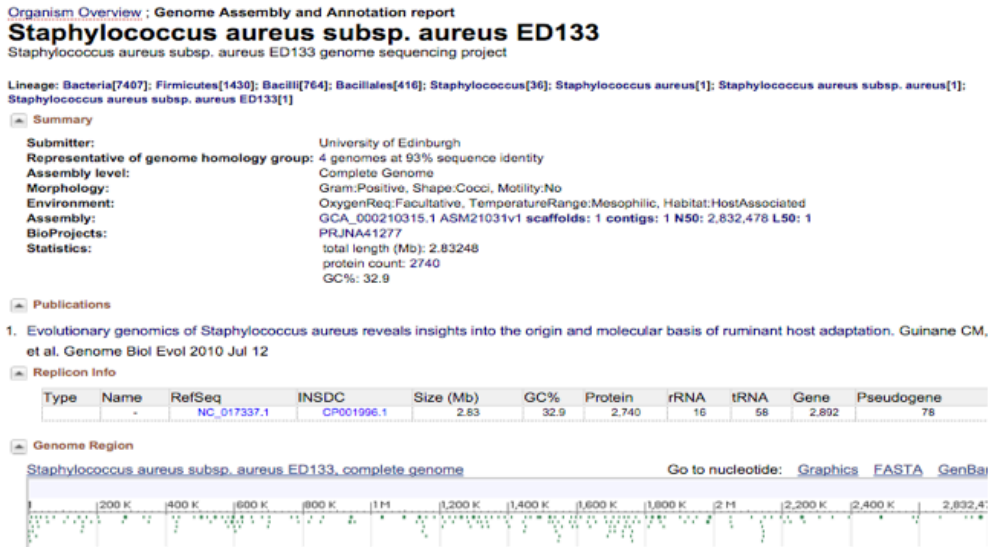


Figure 2.16: An organism overview showing the Genome assembly and annotation report of *Staphylococcus aureus subsp. aureus* ED133.

The *S. aureus* subsp. aureus ED133 sequence was downloaded and saved in FASTA (.fas) format, a text-based format that represents the nucleotide using single letter codes (e.g. TCAG). A sequence in FASTA format begins with a single line description followed by lines of the sequence data, differentiated by a greater-than (>) symbol at the start of the descriptor/ header line.

2.9.2 *S. aureus* literature review

A literature review was carried out on the strains of *S. aureus* that were identified. Two search engines were used to carry out the literature review, which included; PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>) and Science Direct (<http://www.sciencedirect.com/>).

The specific *S. aureus* strain that was identified in GenBank was also used as the key word, for instance “*Staphylococcus aureus* subsp. aureus ED133” was entered into the search field and the relevant articles were viewed and information was collated. This information included; host (animal or human), year it was isolated, country of isolation and any other features. This information was subsequently collated into a table.

2.9.3 Virulence Factor literature review

A literature review was then carried out on *S. aureus* virulence and the virulence factors that are associated with it. Keywords that were used in PubMed and Science Direct included; *S. aureus*, *Staphylococcus aureus*, virulence, virulence determinants, enhanced virulence, pathogenicity. A list of virulence factor determinants and the subsequent genes was collated. Along with information corresponding to the function of the virulence.

2.9.4 Acquiring the Virulence Factor Genes

A list of virulence factors was compiled. The virulence factor genes were searched in the GenBank® database in the same way as the *S. aureus* genomes. Using the GenBank® database search box, the query gene name was used as the key word, for example, alpha hemolysin or its corresponding gene abbreviation, HLA (Figure 2.17). GenBank® was used to search all of the nucleotide sequences that have been uploaded to its database corresponding to the search name, the results page shows the alpha-hemolysin gene from different organisms. Figure 2.17 shows the alpha-hemolysin present in the *S. aureus subsp. aureus* NCTC8325 strain as well as *Cyanidioschyzon merolae*, a unicellular haploid red alga. It is important to select the gene that is located in a *S. aureus* genome, as shown in the description column.

Gene [Create alert](#) [Advanced](#)

Tabular ▾ 20 per page ▾ Sort by Relevance ▾ [Send to: ▾](#)

Search results

Items: 1 to 20 of 260 << First < Prev Page 1 of 13 Next > Last >>

[See also 146 discontinued or replaced items.](#)

Name/Gene ID	Description	Location	Aliases
<input type="checkbox"/> SAOUHSC_01121 ID: 3920722	alpha-hemolysin [<i>Staphylococcus aureus subsp. aureus</i> NCTC 8325]	NC_007795.1 (1076413..1077372, complement)	SAOUHSC_01121
<input type="checkbox"/> CYME_CMM140C ID: 16995167	similar to alpha-hemolysin [<i>Cyanidioschyzon merolae strain 10D</i>]	NC_010139.1 (321343..321897, complement)	CYME_CMM140C
<input type="checkbox"/> FN0003 ID: 991931	alpha-hemolysin [<i>Fusobacterium nucleatum subsp. nucleatum</i> ATCC 25586]	NC_003454.1 (643032..643280)	FN0003
<input type="checkbox"/> HMPREF0351_12042 ID: 13000740	alpha-hemolysin [<i>Enterococcus faecium</i> DO]	NC_017960.1 (1992745..1993041, complement)	HMPREF0351_12042
<input type="checkbox"/> HMPREF0204_RS14150 ID: 23024064	alpha-hemolysin [<i>Chryseobacterium gleum</i> ATCC 35910]		HMPREF0204_RS14150
<input type="checkbox"/> TTHA0444 ID: 3170061	alpha-hemolysin [<i>Thermus thermophilus</i> HB8]	NC_006461.1 (417522..417770, complement)	TTHA0444

Figure 2.17: A results page output of the GenBank® database showing the presence of the alpha-hemolysin gene in both *Staphylococcus aureus* and *Cyanidioschyzon merolae* species.

The alpha-hemolysin gene (SAOUHSC_01121) that is found in the *S. aureus subsp. aureus* NCTC 8325 was selected and an extensive amount of information was presented (Figure 2.18). It was important to check if the gene had been reviewed, this was found under the summary tab and it ensured that the gene was authentic. The gene was then downloaded in FASTA format and GenBank format.

Gene [Advanced](#)

Full Report ▾ [Send to: ▾](#)

SAOUHSC_01121 alpha-hemolysin [*Staphylococcus aureus subsp. aureus* NCTC 8325]

Gene ID: 3920722, updated on 2-Dec-2015

Summary

Gene symbol SAOUHSC_01121
 Gene description alpha-hemolysin
 Locus tag SAOUHSC_01121
 Gene type protein coding
 RefSeq status REVIEWED
 Organism [Staphylococcus aureus subsp. aureus NCTC 8325 \(strain NCTC 8325, sub-species aureus\)](#)
 Lineage Bacteria; Firmicutes; Bacilli; Bacillales; Staphylococcus

Genomic context

Sequence: NC_007795.1 (1076413..1077372, complement)

Genomic regions, transcripts, and products

Genomic Sequence: NC_007795.1 [Go to reference sequence details](#)

[Go to nucleotide:](#) [Graphics](#) [FASTA](#) [GenBank](#)

Figure 2.18: An output of the alpha-hemolysin (SAOUHSC-01121) gene that is present in *Staphylococcus aureus subsp. aureus* NCTC 8325.

2.10 MAUVE

A Bioinformatic software called MAUVE, developed at The Darling Lab at the University of Technology Sydney, was used to compare the *S. aureus* genomes uploaded from GenBank® (Darling *et al.*, 2004); however, it was not used to compare the 6 novel *S. aureus* strains that were sequenced. The software is publicly available to download (<http://darlinglab.org/mauve/download.html>). MAUVE is a multiple genome alignment software for comparative genomic and evolutionary dynamic studies, that works well on closely related species

Genome sequences can undergo small local changes caused by nucleotide substitutions resulting in SNPs, including insertion or deletion of regions of DNA (Indels) or inversions. They can also undergo large-scale changes as they evolve. Recombination can cause major changes which includes gene loss, duplication and rearrangement, the process that creates genetic diversity. MAUVE used this information during the alignment process and identified conserved segments that appeared to be free from genome rearrangements, such regions are referred to as Locally Collinear Blocks (LCBs) and are represented by the coloured blocks in the MAUVE display. LCBs are selected as Mauve requires that each LCB regions appear as genome arrangements and that each region of the alignment meets a 'minimum weight' criteria. The weight of an LCB is a measure of confidence that it is a true genome arrangement, rather than a false match. Mauve also uses the anchored alignment method to align the sequences using inexact, un-gapped matches as alignment anchors (Darling *et al.*, 2004).

Figure 2.19 shows an alignment of six *S. aureus* strains (04-02981, OBA02176, 11819-17, 502A, 55-2053 and 6850) generated with Mauve 2.1 using the progressive Mauve algorithm. The strains were uploaded into the programme in FASTA format. Each genome is laid out horizontally with the position of the nucleotides numbered across the top. The LCBs in the genomes are the coloured blocks and shows differences and similarities between the genome sequence. Blocks with no colour represent regions specific

to that strain, whereas homologous segments are shown as coloured blocks that are connected across the genomes.

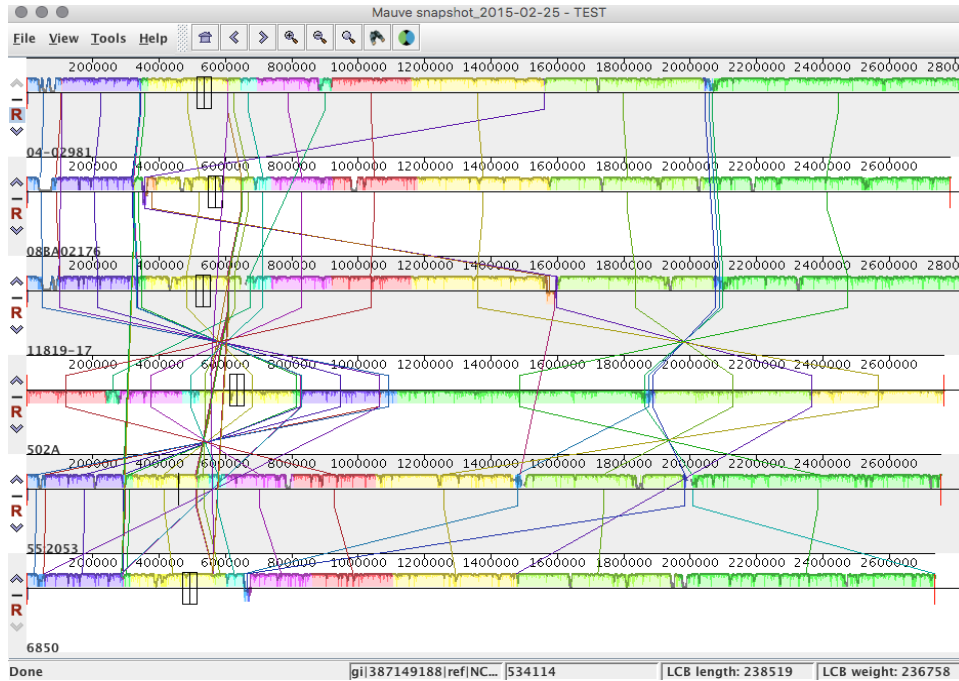


Figure 2.19: An alignment of six *S. aureus* strains (04-02981, OBA02176, 11819-17, 502A, 55-2053 and 6850) generated with Mauve 2.1 using the progressive Mauve algorithm.

The aligned sequences can also be loaded into the programme with its corresponding GenBank® file, which contains the known annotated features of the genome. The annotations only appear once the tool is focused on a sufficiently small region of the genomes e.g. less than 1Mbp of the sequence; so the more that the user zooms in the more likelihood that annotation (if any) will appear. The toolbar at the top of the home screen has a number of tools (Figure 2.20) the helps navigate the alignment.



Figure 2.20: An output showing the Mauve toolbar with the various tools available.

To zoom in on the alignment, the magnifying glass button from the toolbar was used. Once zoomed in, the annotated CDS features showed up as white boxes and the NCBI website can be accessed directly from the Mauve software to find out additional features on that particular gene. The software also has a powerful interface to search genome annotations and its corresponding position in the genome (Figure 2.21).

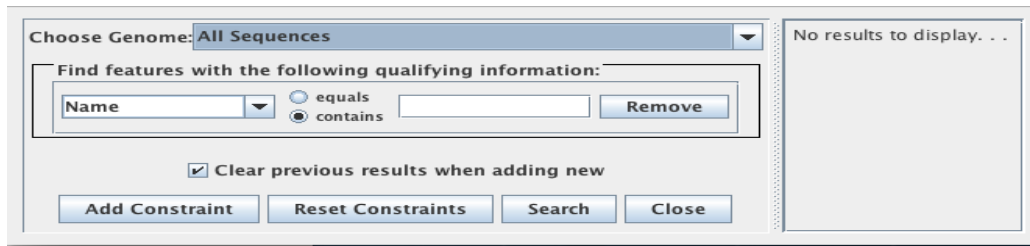


Figure 2.21: An output showing Mauve search tools interface that will also display the corresponding position of the genome.

Annotations can be searched by the identification of the query coding sequence name, amino acid sequence, gene ontology and other fields. It is also possible to search for *S. aureus* proteins associated with pathogenicity, for example, alpha-hemolysin. The software also has the ability to zoom in all the way down to the base level of the DNA so that any changes that occur at the base level of the genomes can be noted (Figure 2.22).

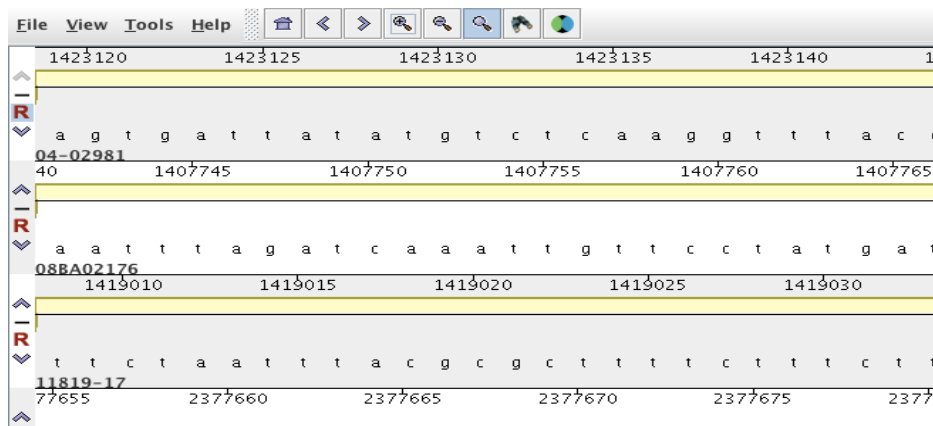


Figure 2.22: An output showing the software's functionality to zoom in all the way down to the base level of the DNA.

2.11 BLAST

An open access tool available at the National Centre for Biotechnology Information (NCBI) website called Basic Local Alignment Search Tool (BLAST) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>.) was used to find query regions, for example virulence genes, of local similarity between a query sequence and the sequences held within the GenBank database or a specified sub-section of the database (Altschul *et al.*, 1990). The tool has the ability to compare query nucleotide or protein sequences to the GenBank® database. BLAST can also read sequences to its complimentary strand or reverse-complement counter part and calculates the statistical significance of the matches.

There are many different BLAST programs to choose from depending on the type of query sequence. The different options include: Nucleotide BLAST (which searches nucleotide databases using a nucleotide query), Protein BLAST (which searches protein databases using a protein query), BLASTx (which searches protein databases using a translated nucleotide query), tBLASTn (which searches translated nucleotide databases using a protein query) and tBLASTx (which searches translated nucleotide databases using a translated nucleotide query).

For this study, Nucleotide BLAST was used to help identify if virulence factor genes were present in some or all of the *S. aureus* strains, shown in Figures 2.23 a – d). The strain containing the virulence factor genes was identified. The specific region of the bacterial genome was subsequently extracted in a FASTA format and recorded electronically by setting up an account with NCBI, as well as downloaded to a local database. The cut of value to determine if the virulence factor sequence was present in a particular strain was set at 90% query coverage, 85% identity and an E value of <0.01; this data is shown in Appendix 2.

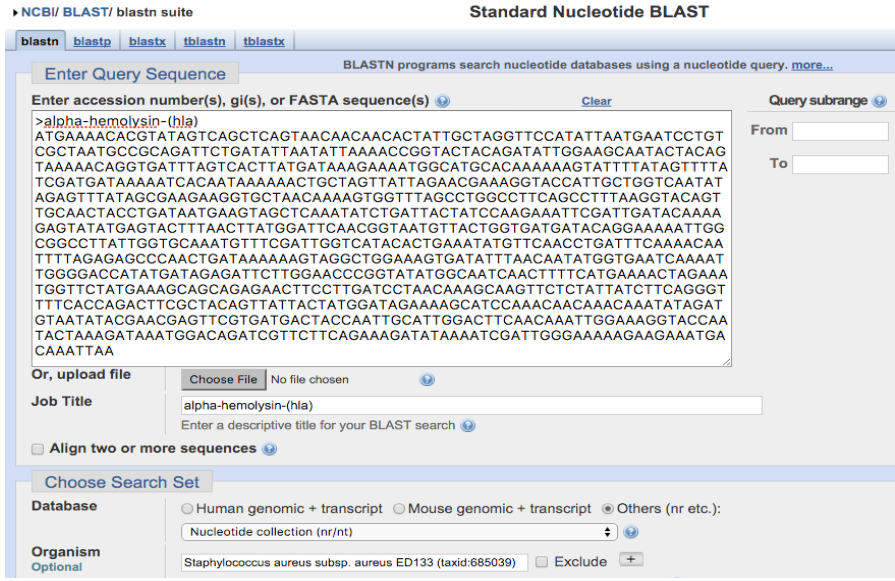


Figure 2.23a. The Complete genome of *S. aureus subsp. aureus* ED133 and the virulence factor gene of alpha-hemolysin, also from *S. aureus* was acquired through GenBank.

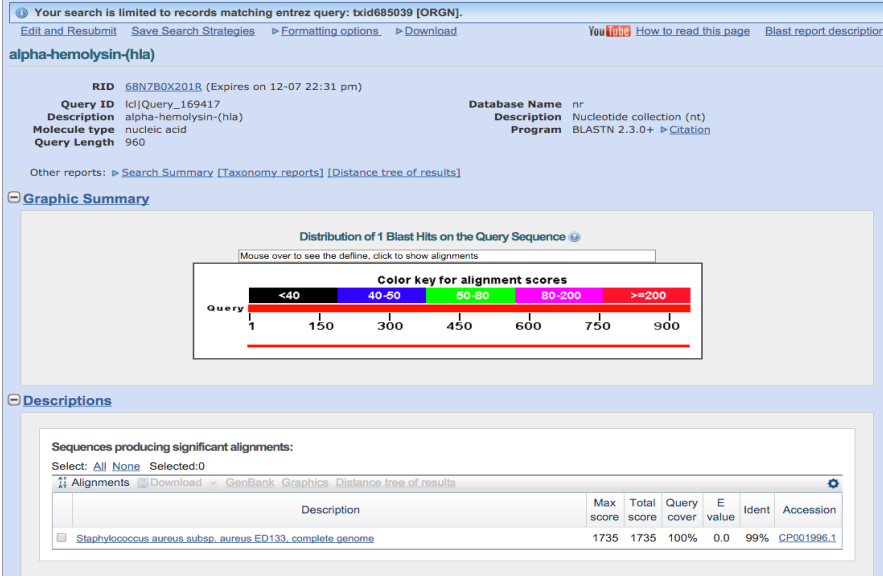


Figure 2.23b. After hitting 'BLAST', the description box is presented which highlights the region of the *S. aureus subsp. aureus* ED133 that corresponds to the alpha-hemolysin gene.

Download ▾ GenBank Graphics

Staphylococcus aureus subsp. aureus ED133, complete genome
Sequence ID: [gb|CP001996.1](#) Length: 2832478 Number of Matches: 1

Range 1: 1183375 to 1184334 [GenBank](#) [Graphics](#) ▾ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
1735 bits(939)	0.0	953/960(99%)	0/960(0%)	Plus/Minus
Query 1	ATGAAAACCGTATAGTCAGCTCAGTAACAACAACACTATTGCTAGGTTCCATATTAATG	60		
Sbjct 1184334	ATGAAAACCGTATAGTCAGCTCAGTAACAACAACACTATTGCTAGGTTCCATATTAATG	1184275		
Query 61	AATCTCTCGCTAATGCGCCAGATTCTGATATTATATAAAACCGGTACTACAGATATT	120		
Sbjct 1184274	AATCTCTCGCTAATGCGCCAGATTCTGATATTATATAAAACCGGTACTACAGATATT	1184215		
Query 121	GGAAGCAATACTACAGTAAAAACAGGTGATTTAGTCACTTATGATAAAGAAAATGGCATG	180		
Sbjct 1184214	GGAAGCAATACTACAGTAAAAACAGGTGATTTAGTCACTTATGATAAAGAAAATGGCATG	1184155		
Query 181	CACAAAAAAGTATTTATAGTTTTATCGATGATAAAAAACACAATAAAAAACGCTAGTT	240		
Sbjct 1184154	CACAAAAAAGTATTTATAGTTTTATCGATGATAAAAAACACAATAAAAAACGCTAGTT	1184095		
Query 241	ATTAGAACGAAAGGTACCATTGCTGGTCAATATAGAGTTTATAGCGAAGAAAGTGTCTAAC	300		
Sbjct 1184094	ATTAGAACGAAAGGTACCATTGCTGGTCAATATAGAGTTTATAGCGAAGAAAGTGTCTAAC	1184035		
Query 301	AAAAGTGGTTAGCCGGCCCTCAGCCCTTAAAGTACAGTTGCAACTACCTGATAATGAA	360		
Sbjct 1184034	AAAAGTGGTTAGCCGGCCCTCAGCCCTTAAAGTACAGTTGCAACTACCTGATAATGAA	1183975		
Query 361	CTAGCTCAAAATFCTGATTAATCTATCCAAGAAATTCGATGATACAAAAGAGTATAGAT	420		
Sbjct 1183974	CTAGCTCAAAATFCTGATTAATCTATCCAAGAAATTCGATGATACAAAAGAGTATAGAT	1183915		
Query 421	ACTTTAACTTATGGATTCAACGGTAATGTTACTGGTGGATGATACAGGAAAAATGGCGGC	480		
Sbjct 1183914	ACTTTAACTTATGGATTCAACGGTAATGTTACTGGTGGATGATACAGGAAAAATGGCGGC	1183855		
Query 481	CTTATGGTGCAAATGTTTCGATTGGTCACTGAAATATGTTCAACCTGATTTCAA	540		
Sbjct 1183854	CTTATGGTGCAAATGTTTCGATTGGTCACTGAAATATGTTCAACCTGATTTCAA	1183795		
Query 541	ACAATTTTAGAGAGCCCACTGATAAAAAAGTAGCCGGAAGTATATTTAACAATATG	600		
Sbjct 1183794	ACAATTTTAGAGAGCCCACTGATAAAAAAGTAGCCGGAAGTATATTTAACAATATG	1183735		
Query 601	GTGAATCAAAATGGGGACCATATGATAGAGATCTTGGAAACCCGGTATATGGCAATCAA	660		
Sbjct 1183734	GTGAATCAAAATGGGGACCATATGATAGAGATCTTGGAAACCCGGTATATGGCAATCAA	1183675		
Query 661	CTTTTCATGAAAACAGAAATGGTCTATGAAAGCAGCAGAACTTCCTTGATCCTAAC	720		
Sbjct 1183674	CTTTTCATGAAAACAGAAATGGTCTATGAAAGCAGCAGAACTTCCTTGATCCTAAC	1183615		
Query 721	AAAGCAAGTCTCTATATCTCAGGGTTTTCCACCAGACTTCGCTACAGTTATTAATATG	780		
Sbjct 1183614	AAAGCAAGTCTCTATATCTCAGGGTTTTCCACCAGACTTCGCTACAGTTATTAATATG	1183555		
Query 781	GATGAAAAGCAATCCAAACAACAACAATAFAGATGTAATATACGAACGAGTTCGCGAT	840		
Sbjct 1183554	GATGAAAAGCAATCCAAACAACAACAATAFAGATGTAATATACGAACGAGTTCGCGAT	1183495		
Query 841	GACTACCAATTCGATTGGACTTCAACAATAATGGAAAGTACCAATACTAAGATAAATGG	900		
Sbjct 1183494	GACTACCAATTCGATTGGACTTCAACAATAATGGAAAGTACCAATACTAAGATAAATGG	1183435		
Query 901	ACAGATCGTCTTCAGAAAGATATAAATCGATTGGGAAAAGGAAGAAATGACAAATTA	960		
Sbjct 1183434	ACAGATCGTCTTCAGAAAGATATAAATCGATTGGGAAAAGGAAGAAATGACAAATTA	1183375		

Figure 2.23c. This box is then presented which shows the query sequence (alpha-hemolysin) aligned with the corresponding section of the subject strain (*S. aureus* ED133) genome sequence.

Similarity between the ED133 and *hla* virulence factor sequences is 99% suggesting that the sequences are highly related. The comparison shows a total of 7 SNPs of which 6 are synonymous mutations and only one G>T transversion at 702 is nonsynonymous giving a E to D amino acid change (Glutamic acid to aspartic acid).

FASTA ▾

Staphylococcus aureus subsp. aureus ED133, complete genome

NCBI Reference Sequence: NC_017337.1

[GenBank](#) [Graphics](#)

>gi|384546269:c1184334-1183375 Staphylococcus aureus subsp. aureus ED133, complete genome

```
ATGAAAACCGTATAGTCAGCTCAGTAACAACAACACTATTGCTAGGTTCCATATTAATGAATCCTGCG
CTAATGCCCGAGATTCGATATTAATATAAAACCGGTACTACAGATATGGAAAGCAATACTACAGTAAA
AACAGGTGATTTAGTCACTTATGATAAAGAAAATGGCATGCAAAAAGATATTTATAGTTTTATCGAT
GATAAAAATCACATAAAAAACGCTAGTTTATAGAACGAAAGTACCATTGCTGGTCAATATAGAGTTT
ATAGCGAAGAGCTGCTAACAATAAGTGGTTAGCCCTGGCCCTCAGCCCTTAAAGTACAGTTGCAACTACC
TGATAATGAAGTAGCTCAAAATATCTGATTAATCTATCCAAGAAATTCGATGATACAAAAGAGTATAGAT
ACTTTAACTTATGGATTCAACGGTAATGTTACTGGTGGATGATACAGGAAAAATGGTGGCCTTATGGTG
CAAAATGTTTCGATTGGTCACTCAAAATATGTTCAACCCGGATTTCAAAACAATTTAGAGAGCCCAAC
TGATAAAAAGTAGGCTGAAAGTATATTAACAATAATGTTGAAATGAAATGGGACCAATATGATAGA
GATTCCTGGAACCCGGTATATGGCAATCAACTTTTCATGAAACGAAATGGCTCTATGAAAGCAGCAG
ATAACTTCCTGATCCTAACAAAGCAAGTCTCTATATCTCAGGTTTTTCACAGACTTCGCTACAGT
TATTAATGATGATAGAAAAGCAATCCAAACAACAACAATAFAGATGTAATATACGAACGAGTTCGCGAT
GACTACCAATTCGATTGGACTTCAACAATAATGGAAAGTACCAATACTAAGATAAATGGACAGATCCT
CTTCAGAAAGATATAAATCGATTGGGAAAAGGAAGAAATGACAAATTA
```

Send: ▾

Change region shown

Whole sequence

Selected region

from: 1183375 to: 1184334

Update View

Customize view

Display options

Show reverse complement

Update View

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Figure 2.23d. The region of the query genome was then entered ('1184334 - 1183375') into the 'selected region' view. The FASTA output shown is the region of the ED133 genome that coincides with the coding region of the alpha-hemolysin gene.

2.12 MEGA

A publicly available bioinformatics software called MEGA (Molecular Evolutionary Genetics Analysis) is used for comparative analysis of DNA and protein sequences, as well as estimating evolutionary distances and constructing phylogenetic trees (Kumar *et al.*, 2004; Tamura *et al.*, 2013). The software was downloaded from <http://www.megasoftware.net/>.

2.12.1 Load Sequence

Load MEGA (version 6.0.6) and the home screen is presented with numerous options (Figure 2.24).

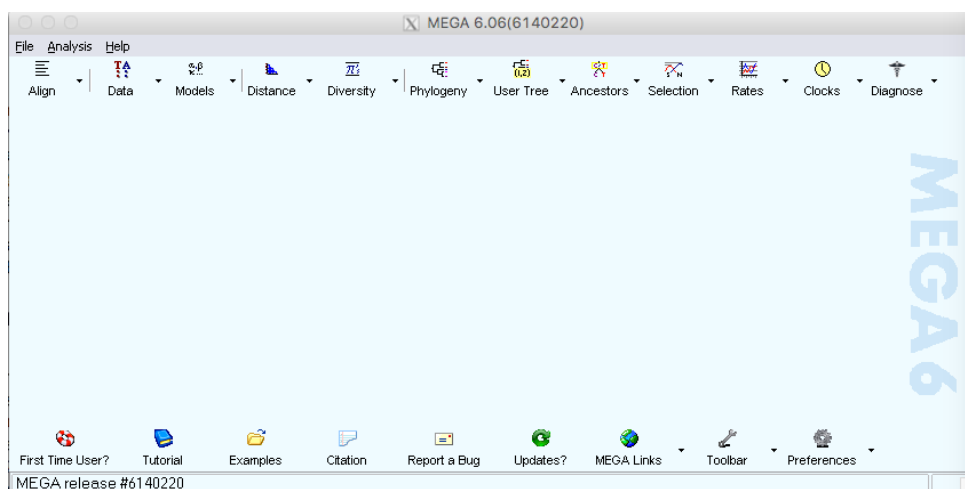


Figure 2.24: An output showing the home screen of MEGA (version 6.0.6).

The Align Tab was selected and the Edit/ Build Alignment option was chosen; a new screen was then presented. Select create a new alignment and then click OK as show below. Another screen will then be generated asking “Are you building a DNA or Protein sequence alignment?”. On this occasion, click DNA (Figure 2.25).

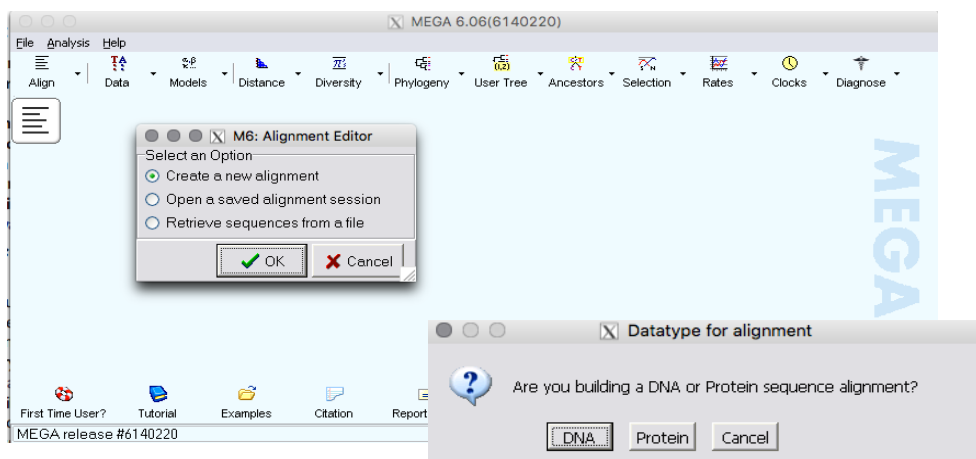


Figure 2.25: An output of MEGA showing the options needed to be selected for a DNA alignment.

Subsequently, select 'Data' from the home screen and choose 'Retrieve sequences from file?'. Open up the FASTA file for the virulence factor and import the aligned sequences. The Alignment Explorer screen will be presented (Figure 2.26 in the MEGA program. Click on the data tab, select the 'Select Genetic Code' tab and ensure that the standard genetic code has been selected.

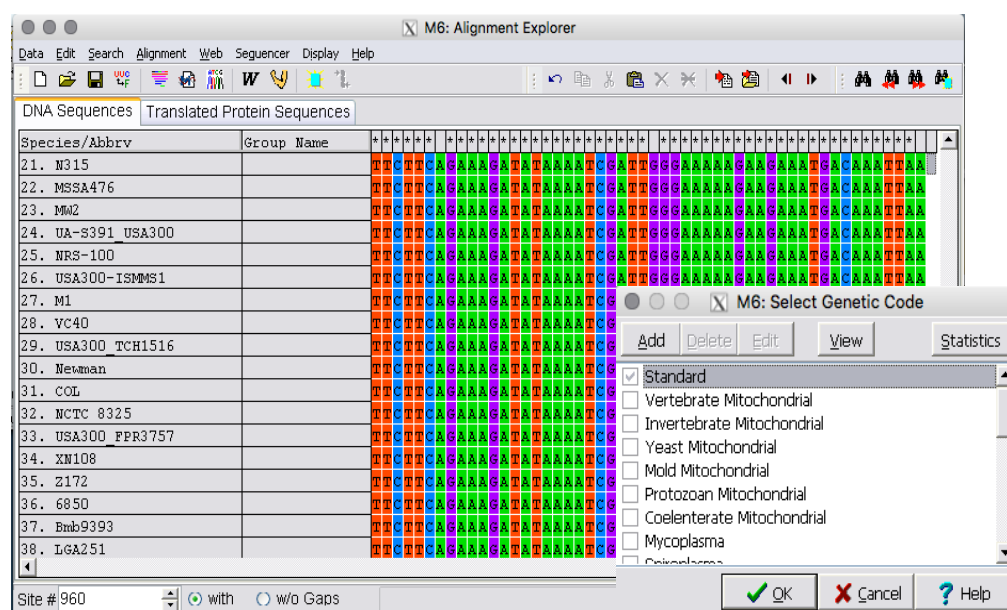


Figure 2.26: An output showing the Alignment Explorer screen with the aligned *S. aureus* sequences.

Reduce the size of the alignment window so that it looks like figure 2.27. The sequences are now ready to be analysed phylogenetically.

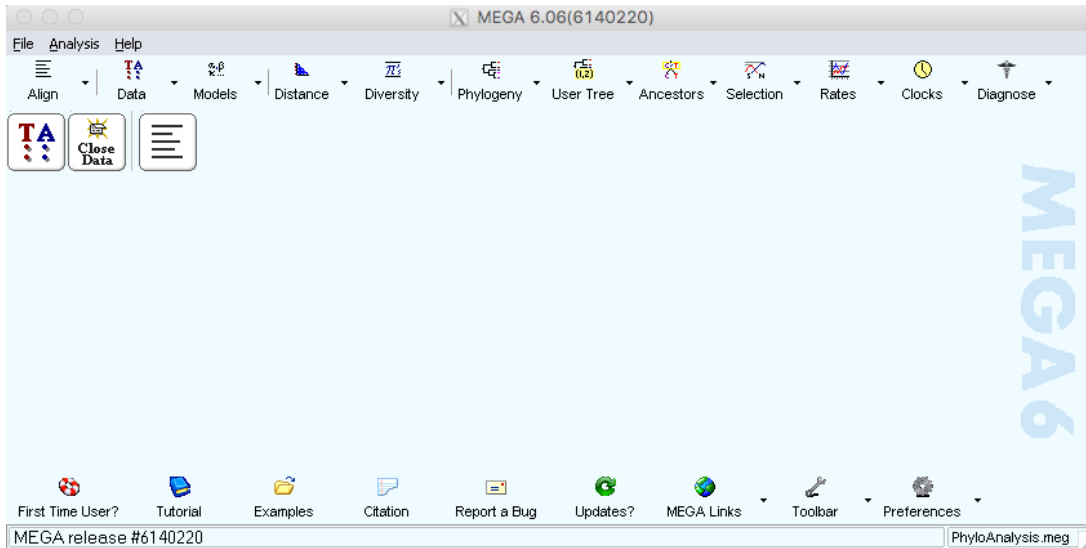


Figure 2.27: An output of the MEGA home screen with the aligned sequences now ready to be analysed phylogenetically.

2.12.2 Constructing a Neighbour-joining Tree

The neighbour joining trees are constructed based on genetic distance measured by comparing the number of nucleotide differences between each sequence. To construct a tree, select the 'Phylogeny' icon and click the 'Construct/ Test Neighbour-joining tree' for the aligned sequence data. A new window will be presented; change the parameters so that they match Figure 2.28.

Briefly, for Test of Phylogeny, select 'Bootstrap-method'; '500' for the number of bootstrap replications; substitution type should be 'Nucleotide'; select the 'kimura 2-parameter model'; substitution to include 'Transitions and Transversions' and finally 'Complete Deletion' should be selected for gaps/ missing data treatment (Figure 2.28).

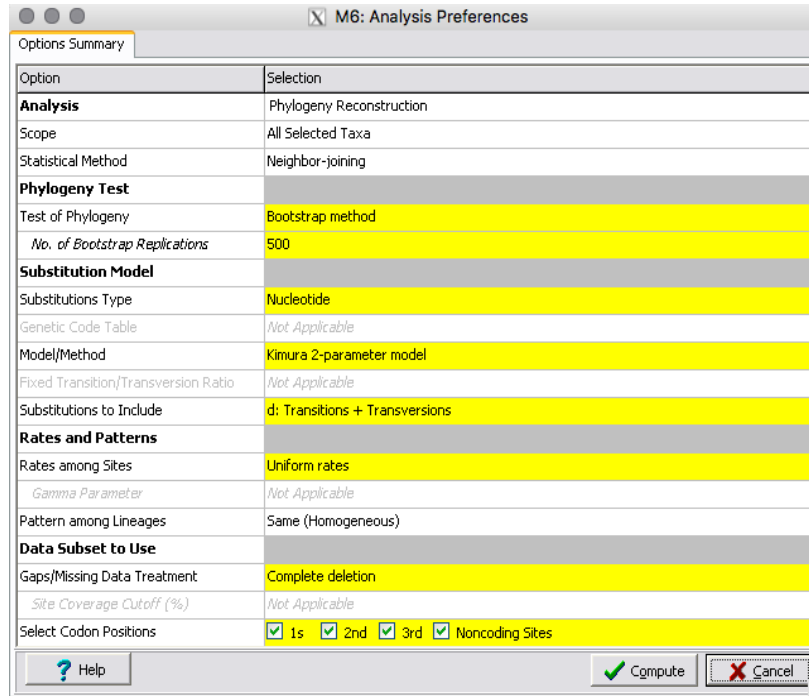


Figure 2.28: An output showing the Analysis Preferences in order to construct a Neighbour Joining Tree.

Once the parameters have been selected, press compute and a progress screen will be presented (Figure 2.29).

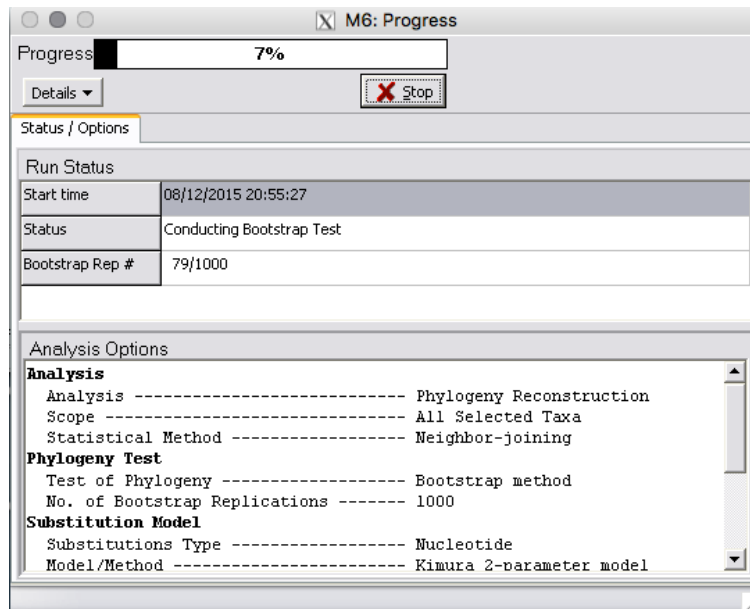


Figure 2.29: An output of the progress screen.

Once the analysis has been completed, the neighbour-joining tree will be presented in a new window. The neighbour-joining tree should be saved in both standard format and Newick format, so that the tree can be edited at a later stage.

2.12.3 Constructing a Maximum-parsimony Tree

This tree takes into account the evolutionary process affecting sequences. Maximum-parsimony chooses the tree(s) that require the fewest evolutionary changes and involves no substitution model to guide the evolutionary process. To construct the tree, select the Phylogeny icon and click the 'Construct/ Test Maximum-parsimony tree and confirm that you want to use the currently active data and a window will open (Figure 2.30). In the window, change the parameters so that they match the figure below. Briefly, for Test of Phylogeny, select 'Bootstrap-method'; setting the number of Bootstrap Replications to '500'; substitution type should be 'Nucleotide'; and finally 'Complete Deletion' should be selected for gaps/ missing data treatment.

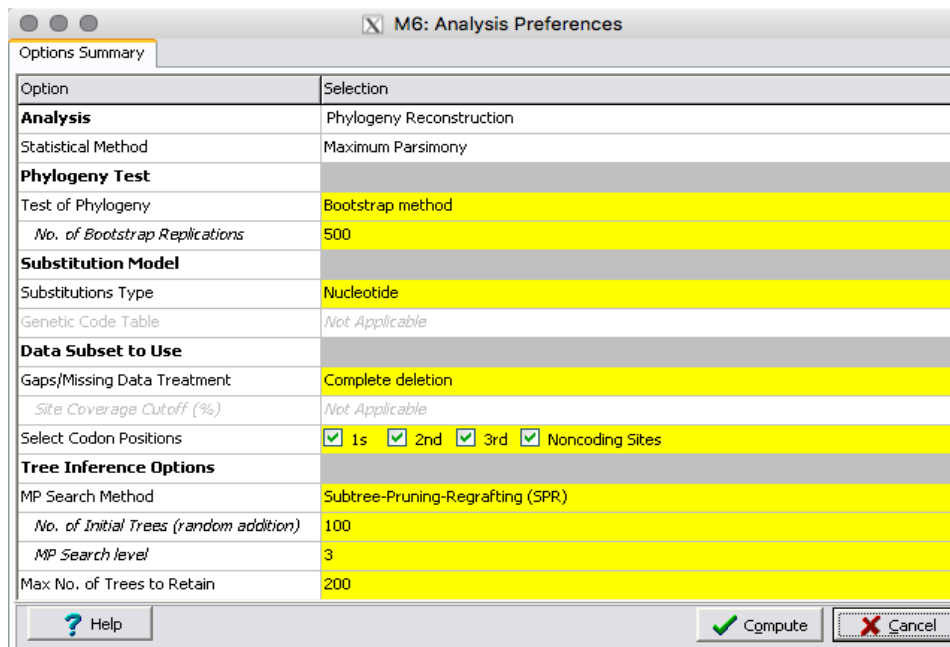


Figure 2.30: An output showing the Analysis Preferences in order to construct a Maximum Parsimony Tree.

Click compute and the progress screen will be shown. Once the analysis completes, a new screen will open presenting the phylogenetic tree; this is one of many trees that the maximum-parsimony analysis will produce (Figure 2.31). This alpha-hemolysin gene has produced 97 trees.

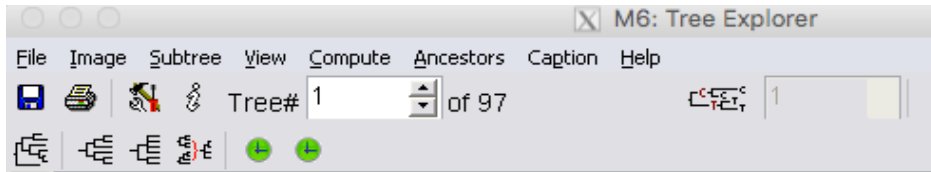


Figure 2.31: An output of the Tree Explorer window which shows that the alpha-hemolysin gene has produced 97 trees.

The next step is to produce a consensus tree based upon all the trees generated by the. To do this, press the compute consensus command in the toolbar (Figure 2.32). A new window will be presented 'Tree Options' and will state 'Cut-off Value for Consensus Tree 50%'. At this stage press OK, this will produce the maximum-parsimony consensus tree. Once complete, save the tree in both standard format and Newick format. To ensure that it can be edited for further studies.

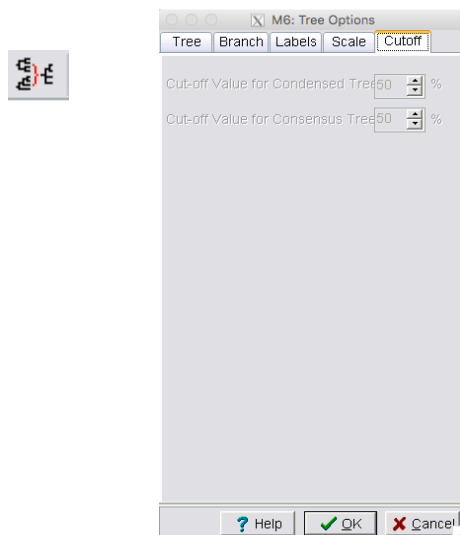


Figure 2.32: An output showing the Tree Options windows to compute a consensus tree.

2.12.4 Constructing a Maximum-Likelihood Tree

The maximum likelihood tree will also be constructed. This analysis takes into account the patterns of evolution in the sequence alignment and uses a specified molecular model to construct and infer the evolutionary histories and relationships.

Begin by identifying the correct model for the virulence factor that is being queried. In this example begin with the alpha-hemolysin alignment. Start by selecting the 'Models' tab and choosing the 'Find Best DNA/ Protein Models (ML)', see figure 2.27 (MEGA homescreen). A new window will be presented (Figure 2.33), keep the default settings and click compute.

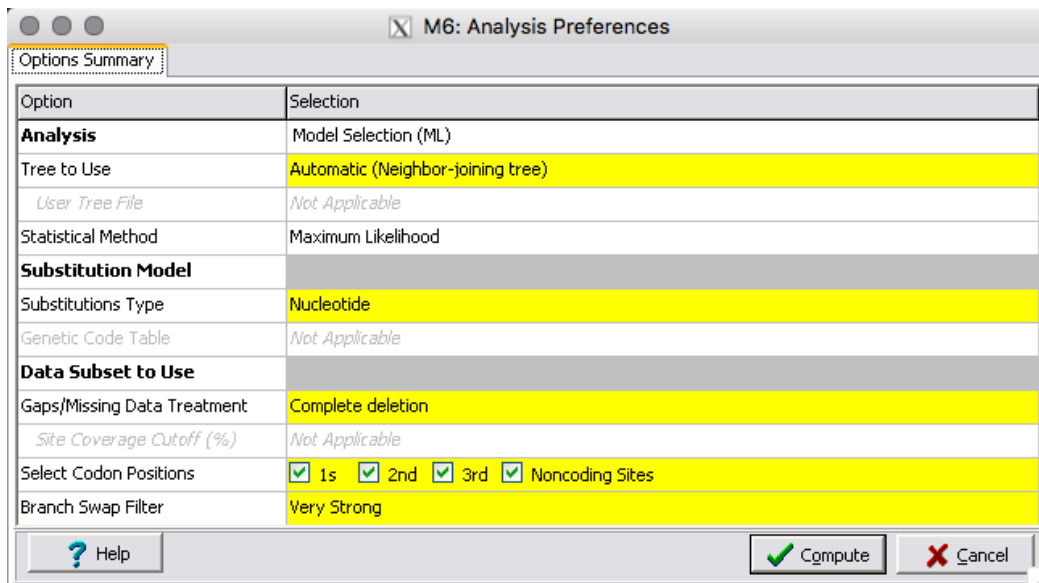


Figure 2.33: An output of the Analysis preference for finding the best DNA Model for a particular virulence factor.

A new table will be presented showing the models that fits best with the data (Figure 2.34). At the bottom of the table (not shown), there is a description of all the codes for the different parameters and components of the models. Ensure that the results of the top 3 models are tabulated in a local database that best fits the data. The abbreviations for the models are as follows: GTR; General Time Reversible, HKY; Hasegawa-Kisaino-Yano, T93; Tamaura-Nei, T92; Tamura 3-parameter, K2; Kimura 2-parameter and JC; Jukes-Cantor.

MEGA Caption Expert

File Edit View Help

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R	f(A)	f(T)
T92+G	112	4943.334	3947.870	-1861.699	n/a	0.09	2.26	0.330	0.330
TN93+G	115	4943.982	3921.867	-1845.685	n/a	0.16	2.26	0.383	0.277
TN93	114	4952.869	3939.637	-1855.574	n/a	n/a	2.03	0.383	0.277
T92+G+I	113	4953.857	3949.509	-1861.514	0.52	0.47	2.29	0.330	0.330
TN93+G+I	116	4954.406	3923.407	-1845.450	0.50	0.63	2.29	0.383	0.277
HKY+G	114	4954.705	3941.474	-1856.492	n/a	0.07	2.37	0.383	0.277
T92	111	4956.012	3969.432	-1873.484	n/a	n/a	2.02	0.330	0.330

Figure 2.34: An output of a table showing the Maximum likelihood fits of 24 different nucleotide substitutions models.

For this dataset, T92+G is the best model to be used. It is now possible to construct the maximum-likelihood phylogeny. Select the phylogeny tab and select Construct/ Test Maximum-likelihood Tree. You will be presented with a windows relating to the parameters of the analysis (Figure 2.35).

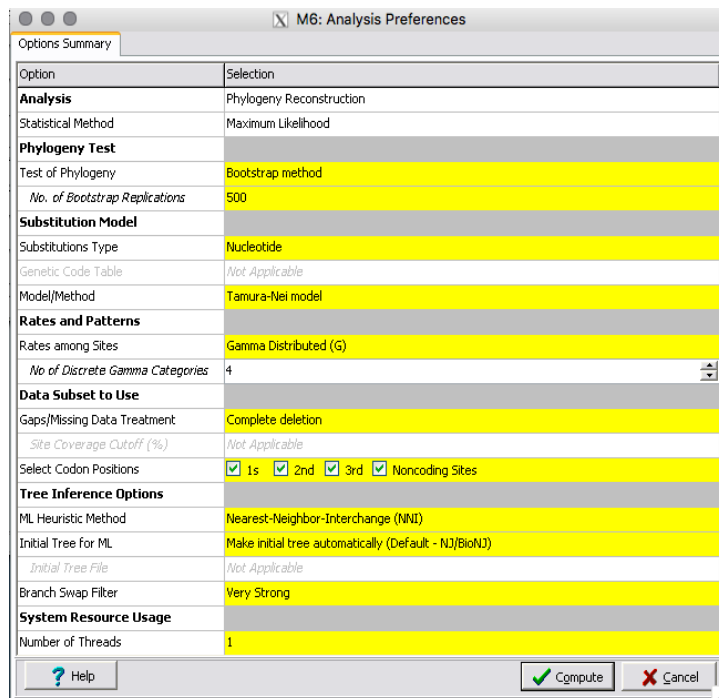


Figure 2.35: An output showing the Analysis Preferences in order to construct a Maximum Likelihood Tree.

Briefly, ensure that the Bootstrap method and 500 replicates is selected. Specify that the model that has been predicted for the data model, in this case T92+G, so select 'Tamura-Nei' model from the drop down menu.

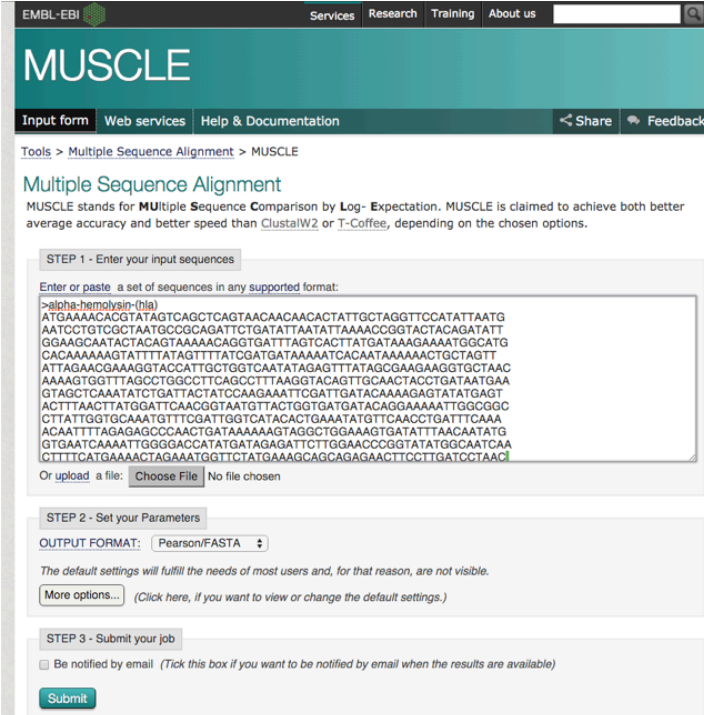
Select 'Gamma Distributed (G)' from the Rates among Sites tab, this corresponds with the G from the T92+G model. Once parameters have been set, click compute. Once analysis have been completed, the maximum-likelihood tree will be produced.

Once complete, save the tree in both standard format and Newick format. To ensure that it can be edited for further studies.

2.13 Muscle

A publicly available software called MUSCLE (Multiple Sequence Comparison by Log-Expectation) (<http://www.ebi.ac.uk/Tools/msa/muscle>) was used to align the *S. aureus* genome sequences to arrange the sequences so that regions of similarity can be readily identified. MUSCLE is a fast method for aligning large numbers of sequences and when compared it achieved greater accuracy than both ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2>) and T-Coffee (<http://www.ebi.ac.uk/Tools/msa/tcoffee>), two other widely used multiple sequence alignment programmes (Edgar, 2004). Multiple alignment is an essential prerequisite to any further phylogenetic analysis, it allows DNA sequences to be compared to each other and also allows nucleotide difference to be highlighted.

On the MUSCLE tool homepage, shown in figure 2.36, a set of virulence factor gene sequences (alpha-hemolysin in the example shown) was pasted, in FASTA format into the input window ensuring that the output format is set to Pearson/ FASTA, and the job was submitted.



The screenshot shows the MUSCLE tool homepage. The header includes the EMBL-EBI logo and navigation links for Services, Research, Training, and About us. Below the header, there are links for Input form, Web services, and Help & Documentation, along with Share and Feedback buttons. The main content area is titled "Multiple Sequence Alignment" and includes a brief description of the tool. The "STEP 1 - Enter your input sequences" section contains a text input field with a "Choose File" button and a "No file chosen" message. The input field contains a FASTA sequence for alpha-hemolysin. The "STEP 2 - Set your Parameters" section shows the "OUTPUT FORMAT" set to "Pearson/FASTA" and a "More options..." link. The "STEP 3 - Submit your job" section has a checkbox for "Be notified by email" and a "Submit" button.

```
>alpha-hemolysin-1(hia)
ATGAAAACAGTATAGTCAGCTCAGTAAACAACAACACTATTGCTAGGTTCCATTAATG
AATCCTGTGCGCTAATGCCGAGATTCTGATATTAATATTAACCCGGTACTACAGATATT
GGAAGCAACTACTACAGTAAAAACAGGTGATTTAGTCACTTATGATAAAGAAATGGCATG
CACAAAAAGTATTTATAGTTTATCGATGATAAAAATCACAATAAAAACTGCTAGTT
ATTAGAAGCAAGGTACCATTGCTGGTCAATATAGAGTTTATAGCCAAAGAGGTGCTAAC
AAAAGTGGTTTAGCCTGGCCTTCAGCCTTAAAGGTACAGTTGCAACTACCTGATAATGAA
GTAGCTCAAATCTGATTACTATCCAAGAAATTCGATTGATACAAAAGAGATATAGAGT
ACTTTAACTTATGGATTCAACGGTAAATGTTACTGGTATGATACAAAAGAAATGGCGGC
CTTATTGGTGCAAAATGTTTCGATTGGTCATACACTGAAATATGTTCAACCTGATTTCAA
ACAATTTTAGAGAGCCCACTGATAAAAAGTAGGCTGGAAAGCTGATTTTAAACAATAG
GTGAATCAAAATGGGGACCATATGATAGAGATCTTGGAAACCGGTATATGGCAATCAA
CTTTTCATGAAAACAGAAATGGTTCTATGAAAAGCAGCAGAGAAGCTTCTTGATCCTAAC
```

Figure 2.36. An output showing the MUSCLE tool homepage with the alpha-hemolysin virulence factor pasted in FASTA format.

The job will only take a few minutes to appear in a new browser and the results will also be saved for 7 days on the server. Once completed, the aligned sequence results will be shown as FASTA format in the browser window (Figure 2.37). Simply 'copy and paste' the aligned FASTA format sequence into a text-file, this conserves the FASTA format. This process was preformed for all the sets of virulence factor gene sequences.

EMBL-EBI Services Research Tr

MUSCLE

Input form Web services Help & Documentation

Tools > Multiple Sequence Alignment > MUSCLE

Results for job muscle-l20151208-111056-0421-395029

Alignments Result Summary Phylogenetic Tree Submission Details

Download Alignment File Send to ClustalW2_Phylogeny

```
>JKD6159
ATGAAAACACGTATAGTCAGCTCAGTAACAACAACACTATTGCTAGGTTCCATATTAATG
AATCCTGTCGCTAATGCCGCAGATTCGATATTAATATTAACCCTGACTACAGATATT
GGAAGCAATACTACAGTAAAAACAGGTGATTAGTCACTTATGATAAAGAAAATGGCATG
CACAAAAAGTATTTATAGTTTATCGATGATAAAAATCACAATAAAAACTGCTAGTT
ATTAGAACGAAAGGTACCATTGCTGGTCAATATAGAGTTTATAGCGAAGAAGGTGCTAAC
AAAAGTGGTTAGCCTGGCCTTCAGCCTTTAAGGTACAGTTGCAACTACCTGATAATGAA
GTAGCTCAAATATCTGATTACTATCCAAGAAAATTCGATTGATACAAAAGATATATGAGT
ACTTTAACTTATGGATTCAACGGTAATGTTACTGGTGTATGATACAGGAAAAATGGCGGC
CTTATTGGTGCAAATGTTTCGATTGGTCATACACTGAAATATGTTCAACCTGATTTCAAA
ACAATTTTAGAGAGCCCAACTGATAAAAAAGTAGGCTGGAAAGTGATATTTAACAATATG
GTAAATCAAATTTGGGACCATATGATAGAGATTC TTGGAACCCAGTATATGGCAATCAA
CTTTTCATGAAAAGTAGAAAATGGTTCTATGAAAGCAGCAGATAACTTCCTTGATCCTAAC
AAAGCAAGTTCCTATTATCTTCAGGGTTTTCACCAGACTTCGCTACAGTTATTAAGTATG
GACAGAAAAGCATCCAACAACAACAATAATAGATGTAATATATGAACGAGTTCGTGAT
GACTATCAATTACATTGGACTTCAACAACTGGAAGGTACTAATACTAAGATAAATGG
ACAGATCGTTCTTCAGAAAAGATATAAAAATCGATTGGGAAAAAGAAGAAATGACAAATTA
```

Figure 2.37: An output of the aligned sequence results shown on the MUSCLE homepage.

2.14 Phylogeny.fr

Phylogeny fr. (<http://www.phylogeny.fr/index.cgi>) is a publicly available online tool used for the identification of homologous sequences, their multiple alignment, phylogenetic tree reconstruction and the graphical representation of the inferred tree (Dereeper *et al.*, 2008; Dereeper *et al.*, 2010). Phylogeny fr. has three modes of use. The 'One Click' Mode is designed for users who have little experience in the field of phylogenetic's. Insert a set of FASTA sequences into the submission box and this mode will use a specific workflow that includes MUSCLE for multiple alignment, PhyML tool for tree building and TreeDyn for tree rendering (Dereeper *et al.*, 2008; Guindon & Gascuel., 2003). The 'Advance Mode' allows the parameters of each of the programs, in the above workflow to be customised. The 'A La Carte' mode allows a lot more flexibility. The user can create a customised phylogeny workflow using different programs from its server. (Dereeper *et al.*, 2008; 2010).

Select the 'One Click' mode and paste the set of alpha-hemolysin sequences into the field in FASTA format (Figure 2.38). Name the analysis, choosing a name that reflects the data set. Enter an email address to receive the results and simply click submit. Once complete, the window will show a phylogenetic tree. Save as PDF and Newick file.

Home Phylogeny Analysis Blast Explorer Online Programs Your Workspace Documentation Downloads

"One Click" Mode Alignment MUSCLE Curation Gblocks Phylogeny PhyML Tree Rendering TreeDyn

1. Overview 2. Data & Settings

Name of the analysis (optional):

Upload your set of sequences in FASTA, EMBL or NEXUS format from a file:

Or paste it here (load example of sequences)

```
>alpha-hemolysin (hla)
ATGAAAACACGTTAGTCAGCTCAGTAAACAACACTATTGCTAGGTTCCATATTAATG
AATCCTGTCGCTAATGCCGAGATTCTGATTAATATAAACCCGCTACTACAGATATT
GGAAGCAACTACTACAGTAAAAACAGTGATTAGTCACTTATGATAAAGAAAATGGCATG
CACAAAAAAGTATTTTATAGTTTTATCGATGATAAAAAACACAATAAAAACTGCTAGTT
ATTAGAACGAAAGGTACCATTGCTGGTCAATATAGAGTTTATAGCGAAGAAGGTGCTAAC
AAAAGTGGTTTAGCCTGGCCCTTCAGCCCTTAAAGGTACAGTTGCAACTACCTGATAATGAA
GTAGCTCAAATATCTGATTACTATCCAAAGAAATTCGATTGATCAAAAGAGTATATGAGT
ACTTTAAGCTTATGGATTCAACGGTAATGTTACTGGTGATGATACAGGAAAAATGGCGGC
CTATTGGTGCAAAATGTTTCGATTGGTCATACACTGAAATATGTTCAACCTGATTTCAA
ACAATTTTAGAGAGCCCAACTGATAAAAAAGTAGGCTGGAAGTGATATTTAACAATAG
GTGAATCAAAATGGGGACCATATGATAGAGATCTTGGAAACCCGGTATATGGCAATCAA
CTTTTCATGAAAACAGAAATGGTTCTATGAAAGCAGCAGAGAACTTCTTGATCCTAAC
AAAGCAAGCTTCGTATATCTTCAGGGTTTTCAGCAGACTTCGCTACAGTTACTACTG
GATAGAAAAGCATCCAAACAACAACAATATAGATGTAATATACGAACGAGTTCTGTGAT
```

Maximum number of sequences is 200 for proteins and 200 for nucleic acids.
Maximum length of sequences is 2000 for proteins and 6000 for nucleic acids.

Use the Gblocks program to eliminate poorly aligned positions and divergent regions

To receive the results by e-mail, enter your address(es):

Figure 2.38. An output showing a set of alpha-hemolysin FASTA sequences pasted in to the phylogeny.fr query box.

2.15 DNA Sequence Polymorphism (DnaSP)

DnaSP, is a publicly available software package, developed for molecular population geneticists for the analysis of nucleotide polymorphism from aligned DNA sequence data and can handle a large number of sequences (Rozas & Rozas., 1995). DnaSP is limited to run on Microsoft Windows® but can run on Apple and Linux based platforms with the use of an emulator software (Rozas *et al.*, 2003).

In this analysis DnaSP version 5.10.01 was used. It is available to academic users from: <http://www.ub.edu/dnasp>. DnaSP can read and export multiple-aligned nucleotide sequence file formats in FASTA, MEGA, NBRF/ PIR, NEXUS and PHYLIP (Rozas *et al.*, 2003). DnaSP can estimate several measures of DNA sequence variation between populations, in noncoding, synonymous or non synonymous sites and analyse the evolutionary pattern of preferred and unpreferred codons (Rozas *et al.*, 2003). DnaSP v5 can also allow for analysis on multiple files, haplotype phasing and analysis on insertion/ deletion polymorphism data and implements new algorithms and methods to identify conserved DNA regions, a useful feature for phylogenetic foot printing based analysis (Librado *et al.*, 2009). DnaSP can also carry out several tests of neutrality, Tajimas test (Tajima., 1989) and Fu & Li's test (Fu & li., 1993) are statistical tests for testing neutral mutation hypothesis.

2.15.1 Load Sequence Alignment

To begin, load the DnaSP program and you will be presented with the home screen (Figure 2.39a)

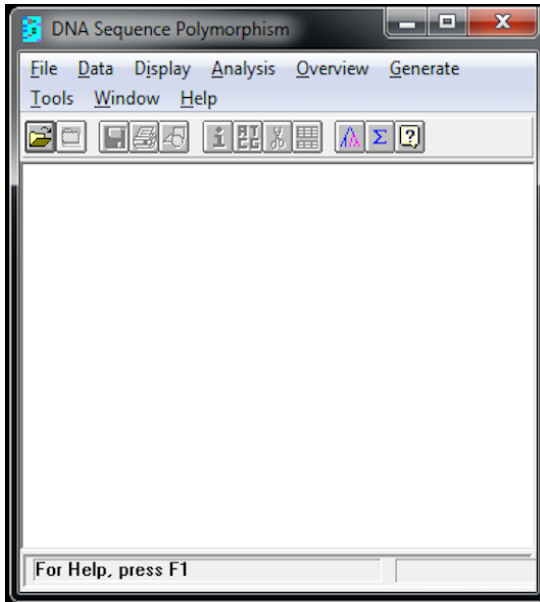
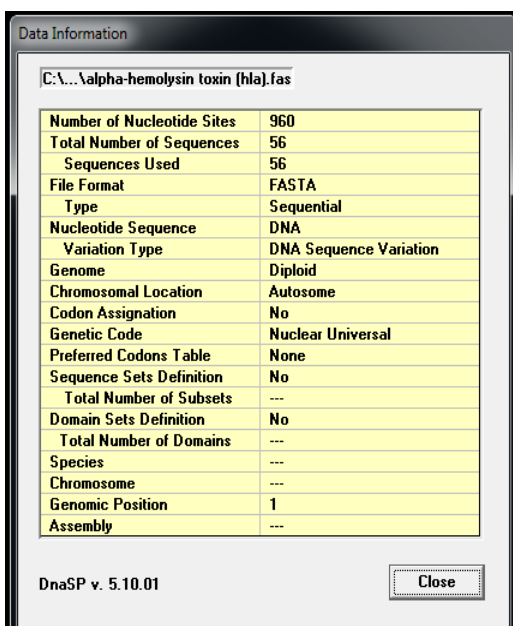


Figure 2.39a. An output of the DnaSP program showing the home screen.

Select 'File' in the top left hand of the window and choose the option to open a data file. Open a sequence alignment of your choice, in this example alpha-hemolysin is used and a new window will be presented (Figure 2.39b).

The image shows a dialog box titled "Data Information" with a yellow background. It displays metadata for a file named "C:\...\alpha-hemolysin toxin [hla].fas". The information is organized into a table with two columns: a label and a value. At the bottom left, it says "DnaSP v. 5.10.01" and at the bottom right, there is a "Close" button.

Data Information	
C:\...\alpha-hemolysin toxin [hla].fas	
Number of Nucleotide Sites	960
Total Number of Sequences	56
Sequences Used	56
File Format	FASTA
Type	Sequential
Nucleotide Sequence	DNA
Variation Type	DNA Sequence Variation
Genome	Diploid
Chromosomal Location	Autosome
Codon Assignment	No
Genetic Code	Nuclear Universal
Preferred Codons Table	None
Sequence Sets Definition	No
Total Number of Subsets	---
Domain Sets Definition	No
Total Number of Domains	---
Species	---
Chromosome	---
Genomic Position	1
Assembly	---

DnaSP v. 5.10.01 Close

Figure 2.39b. An output of the Data Information for this virulence factor.

The information presented in this screen needs to be changed, for example it is not a diploid number cell but instead haploid. Continue by closing this screen and you will return back to the home-screen and start formatting the data so that it matches your dataset. Select the 'Data' tab and choose 'Format', a new screen will be presented to you (Figure 2.39c). Ensure that 'DNA' Nucleotide Sequence, 'Haploid' Genomic State and 'Prokaryotic' Chromosomal Location is selected. Click OK and return back to the home screen.

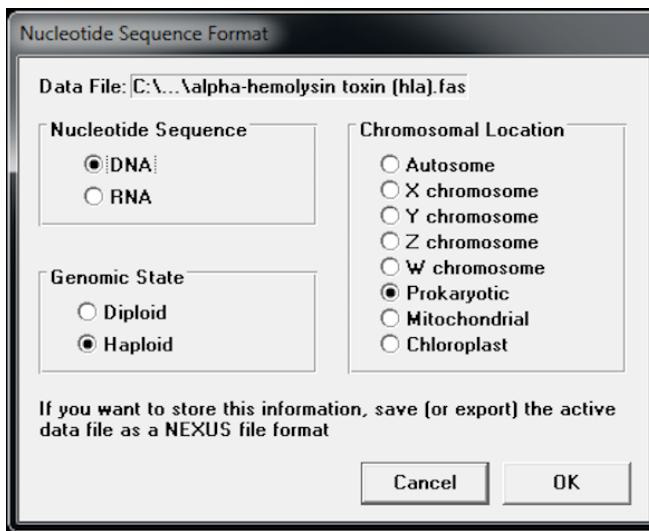


Figure 2.39c. An output showing the Nucleotide Sequence Format for alpha-haemolysin.

Select the 'Data' tab again and choose the 'Protein Coding Regions' tab, and a new window will be presented (Figure 2.39d). Ensure that the 'Assigned the selected region as a protein coding region' position has been set to '1'. Select 'OK' and another window stating that assignment has been done successful will open. It will then prompt for 'Do you want to assign more regions?' Select 'No' and return back to the home screen.

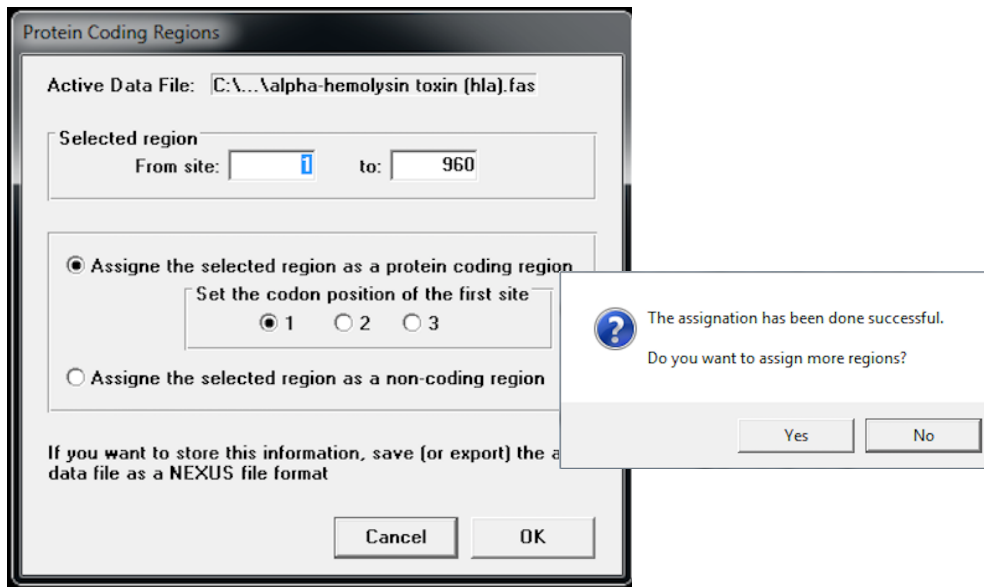


Figure 2.39d: An output showing the 'Protein Coding Regions' that need to be selected for this data set.

Select the 'Data' tab for the third time and choose 'Assign Genetic Code'. A new screen will be presented, ensure that 'Nuclear Universal' is selected from the drop down menu and select 'OK' (Figure 2.39e).

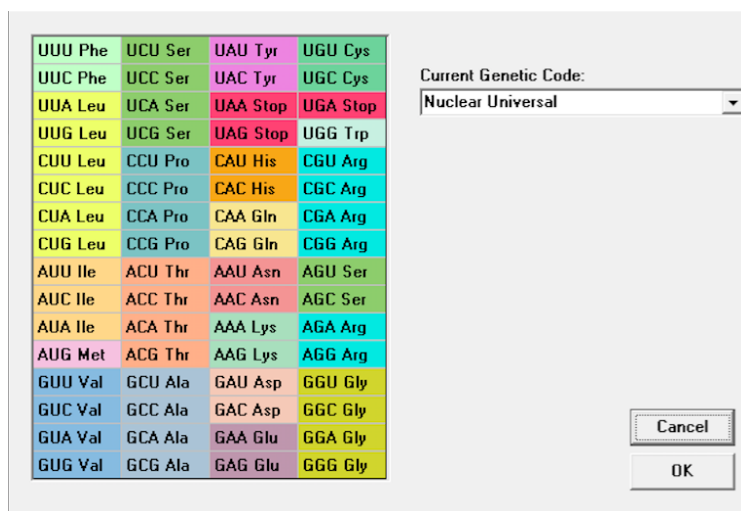


Figure 2.39e: An output showing that the 'Current Genetic Code' option need to be set to 'Nuclear Universal'.

Select the 'Display' tab and choose 'Data Info', a new screen will be presented showing the updated information, confirm that everything is OK. The sequence alignment data is now ready for DnaSP analysis, Figure 2.39f.

Number of Nucleotide Sites	960
Total Number of Sequences	56
Sequences Used	56
File Format	FASTA
Type	Sequential
Nucleotide Sequence	DNA
Variation Type	DNA Sequence Variation
Genome	Haploid
Chromosomal Location	Prokaryotic
Codon Assignment	Yes
Genetic Code	Nuclear Universal
Preferred Codons Table	None
Sequence Sets Definition	No
Total Number of Subsets	---
Domain Sets Definition	No
Total Number of Domains	---
Species	---
Chromosome	---
Genomic Position	1
Assembly	---

Figure 2.39f: An output showing the updated information of the current dataset (alpha-hemolysin).

2.15.2 Generate Polymorphism Data

On the home screen select the 'Overview' tab and choose 'Polymorphism Data'. A new screen will be presented (Figure 2.40). Tabulate the data in an excel sheet ensure that it includes the 'Number of sequences', 'Number of variable sites', 'Number of haplotypes', 'Haplotype diversity', 'Standard deviation of haplotype diversity', 'Nucleotide diversity', 'Standard deviation of Pi', 'Average number of nucleotide differences', 'Tajimas's test of neutrality' and 'Fu & Li's test of neutrality'.

```
Overview: Polymorphism Data

Input Data File: C:\... \alpha-hemolysin toxin (hla).fas
Number of sequences: 56  Number of sequences used: 56
Selected region: 1-960  Number of sites: 960
Total number of sites (excluding sites with gaps / missing data): 960

Number of variable sites, S: 79
Total number of mutations, Eta: 81

===== G+C content =====
G+C content at noncoding positions, G+Cn: 0.000  (3.00 sites)
G+C content at coding positions, G+Cc: 0.341  (957.00 sites)
G+C content -Total-, G+C: 0.339  (960.00 sites)

===== Haplotype/Nucleotide Diversity =====
Number of Haplotypes, h: 15
Haplotype [gene] diversity, Hd: 0.868
  Variance of Haplotype diversity: 0.00079
  Standard Deviation of Haplotype diversity: 0.028

Nucleotide diversity (per site), Pi: 0.02137
  Sampling variance of Pi: 0.0000088
  Standard deviation of Pi: 0.00296
Average number of nucleotide differences, k: 20.51883
Theta [per sequence] from Eta: 17.63318
Theta [per site] from Eta: 0.01837

===== Neutrality Tests =====
Tajima's D: 0.56949
  Statistical significance: Not significant, P > 0.10
Fu and Li's D* test statistic: 0.94103
  Statistical significance: Not significant, P > 0.10
Fu and Li's F* test statistic: 0.95718
  Statistical significance: Not significant, P > 0.10
Fu's Fs statistic: 9.782
Strobeck's S statistic: 0.000
  (Probability that NHap <= 15)
Probability that [NHap = 15]: 0.000

Calculated using the total number of mutations
```

Figure 2.40: An overview screen showing the Polymorphism Data for the dataset (alpha-hemolysin).

2.15.3 Generate Polymorphic Sites

On the home screen select the 'Analysis' tab and choose 'Polymorphic Sites'. A new screen will be presented (Figure 2.41). Tabulate the 'Total number of synonymous mutations' and 'Total number of replacement changes' (non-synonymous mutations)

```
Polymorphic Sites
Parsimony informative sites: 69
Singleton variable sites (two variants): 10
Site positions: 339 340 603 696 777 783 831 870 915 935
Parsimony informative sites (two variants): 67
Site positions: 11 37 51 120 134 222 232 234 235 240 243 255 270
294 300 306 327 337 349 366 372 384 387 396 405 411 423
447 450 463 465 477 480 481 483 501 513 516 552 555 576
585 624 627 645 651 675 681 684 687 702 714 733 750 762
786 793 805 825 843 846 850 852 855 882 902 903
Singleton variable sites (three variants): 0
Parsimony informative sites (three variants): 2
Site positions: 531 747
Variable sites (four variants): 0

===== Protein Coding Region =====
Genetic Code: Nuclear Universal
Protein Coding, and Non-Coding Regions:
Number of protein coding regions (exons): 1
Number of noncoding regions (intronic and flanking regions): 1
Protein coding region, from site: 1 to 957
Non coding region, from site: 958 to 960

Total number of sites in coding regions: 957
Total number of sites excluding complex codons or codons with gaps: 957

===== Synonymous/Replacement Changes =====
Segregating sites: 79 Total number of mutations: 81
Total number of Synonymous changes: 70
37 51 120 222 234 235 240 243 255 270 294 300 306
327 339 340 366 372 384 387 396 405 411 423 447 450 465
477 480 481 483 501 513 516 531 531 552 555 576 585 603
624 627 645 651 675 681 684 687 696 714 733 747 747 750
762 777 783 786 825 831 843 846 850 852 855 870 882 903
915
Total number of Replacement changes: 11
11 134 232 337 349 463 702 793 805 902 935

!, Some codons with multiple evolutionary paths were detected.
The number of synonymous and nonsynonymous mutations were estimated
using the DnaSP conservative criteria (see the help file).
Sites affected: 234 481 483 903
```

Figure 2.41: An overview screen showing the Polymorphic Sites for the dataset (alpha-hemolysin).

2.15.4 Generate Synonymous and non-Synonymous Substitutions

On the home screen, select the 'Analysis' tab and choose the 'Synonymous and non-Synonymous' tab. A new screen will be presented (Figure 2.42), confirm that the 'In Coding and Non-coding regions' is selected in the 'Silent Substitutions considered' tab, and select 'OK'.

The screenshot shows a dialog box with the following fields and options:

- Data Set:** All Included Sequences (n = 56)
- Region to Analyze:** From site: 1 to: 960
- Silent Substitutions Considered:**
 - Substitutions in Coding Regions
 - In Coding and Noncoding Regions
- Buttons:** Cancel, OK

Figure 2.42: An overview screen showing the options that need to be selected to generate synonymous and non-synonymous substitutions.

A new file will be generated (Figure 2.43), export the file and open in Microsoft excel. Subsequently, calculate the average of the Ks and Ka column and the standard deviation of the averages. Tabulate the results in a spreadsheet.

Seq 1	Seq 2	SilentDif	SilentPos	Ks	NSynDif	NSynPos	Ka
alpha-hemo	MRSA_OP10	1.00	214.67	0.0047	1.00	745.33	0.0013
alpha-hemo	MRSA_B51	0.00	214.67	0.0000	0.00	745.33	0.0000
alpha-hemo	MRSA_IP70	44.50	213.42	0.2443	5.50	746.58	0.0074
alpha-hemo	MRSA_B53	37.50	213.67	0.1999	5.50	746.33	0.0074
alpha-hemo	MRSA_S62	37.50	213.67	0.1999	5.50	746.33	0.0074
alpha-hemo	MRSA_AS3	37.50	213.67	0.1999	5.50	746.33	0.0074
alpha-hemo	502A	0.00	214.67	0.0000	0.00	745.33	0.0000
alpha-hemo	CN1	0.00	214.67	0.0000	0.00	745.33	0.0000
alpha-hemo	ST228_1858	0.00	214.67	0.0000	0.00	745.33	0.0000
alpha-hemo	ST228_1612	0.00	214.67	0.0000	0.00	745.33	0.0000
alpha-hemo	ST228_1603	0.00	214.67	0.0000	0.00	745.33	0.0000
alpha-hemo	1819-97	0.00	214.67	0.0000	0.00	745.33	0.0000
alpha-hemo	ECT-R2	0.00	214.67	0.0000	0.00	745.33	0.0000
alpha-hemo	04-02981	0.00	214.67	0.0000	0.00	745.33	0.0000
alpha-hemo	ED98	0.00	214.67	0.0000	0.00	745.33	0.0000
alpha-hemo	Mu3	0.00	214.67	0.0000	0.00	745.33	0.0000
alpha-hemo	JH1	0.00	214.67	0.0000	0.00	745.33	0.0000
alpha-hemo	JH9	0.00	214.67	0.0000	0.00	745.33	0.0000
alpha-hemo	Mu50	0.00	214.67	0.0000	0.00	745.33	0.0000
alpha-hemo	N315	0.00	214.67	0.0000	0.00	745.33	0.0000

Figure 2.43: The output file generated in DnaSP showing the synonymous and non-synonymous mutations for this dataset (alpha-hemolysin).

2.15.5 Define Datasets

On the home screen, select the 'Data' tab and choose 'Define Sequence Sets' and a new screen will be presented (Figure 2.44). Create four new sequence sets. Move over all of the sequences that are being used in the study to the 'Included List', removing any strains that are no longer being used, for example the alpha-hemolysin reference sequence. Subsequently create a Hospital-associated, Community-associated and Livestock-associated sequence set, using information from the literature research. Save the file in an output file

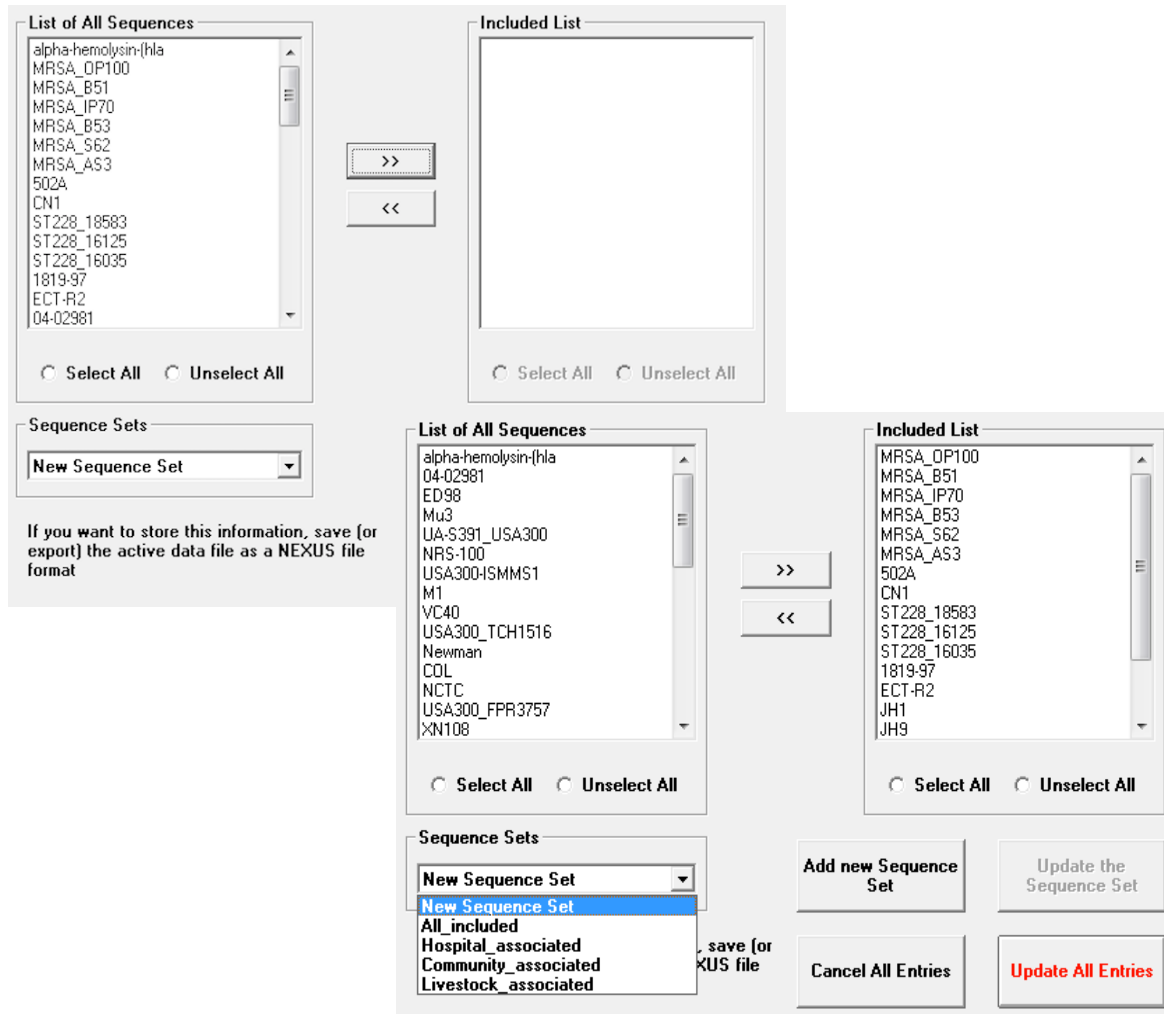


Figure 2.44. An overview showing the 'define data set' screen with four data sets; all sequences, hospital associated, community associated and livestock associated.

2.15.6 DNA Divergence

Select the 'Analysis' tab and choose 'DNA Divergence Between Populations' and a new screen will be presented (Figure 2.45). Generate DNA divergence between different populations, i.e. 'Livestock-associated vs Community-associated', 'Community-associated vs Hospital-associated' and 'Hospital-associated vs Livestock-associated'. To do this, change the datasets by changing the 'Population #1' and 'Population #2' drop down menu, select OK and a new screen will be presented with the output (Figure 2.46).

The dialog box contains the following fields and options:

- Region to Analyze:** From site: 1 to: 960
- Data Sets:**
 - Population #1: Livestock_associated (n = 6)
 - Population #2: Community_associated (n = 6)
- Sliding Window:**
 - Compute
 - Window length: 100 sites
 - Step size: 25 sites
 - Excluding sites with gaps
- Options:**
 - Compute Variances of P_i , D_{xy} , D_a (Jukes and Cantor)

Buttons: Cancel, OK

Figure 2.45: An overview screen showing the options that need to be selected to generate the DNA divergence data.

Tabulate the average number of nucleotide differences between populations, the average number of nucleotide substitutions per site between populations (D_{xy}) and the number of net nucleotide substitutions per site between populations (D_a)

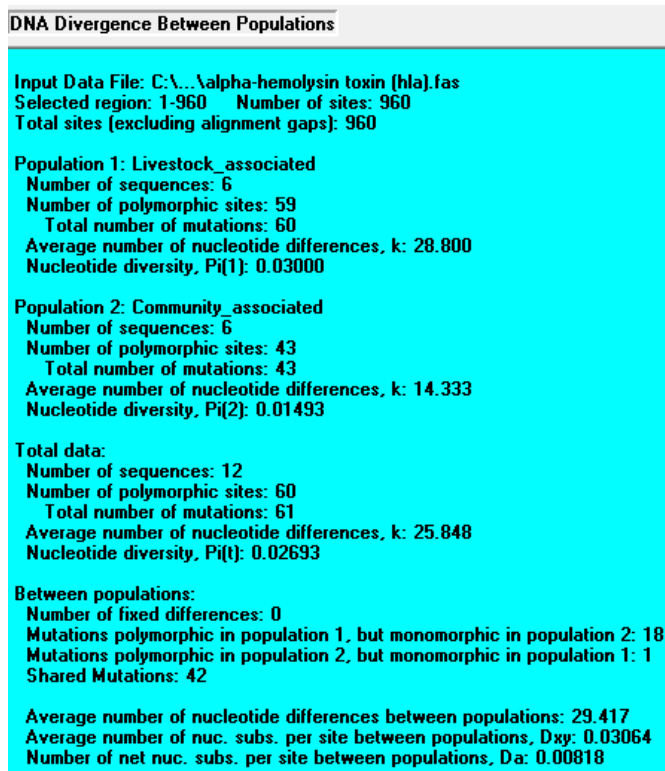


Figure 2.46: An overview screen showing the results for the DNA divergence between populations data.

2.15.7 Generate Haplotype Data

On the home screen, select the 'Generate' tab and select 'Haplotype Data File' and a new window will be presented to you (Figure 2.47).

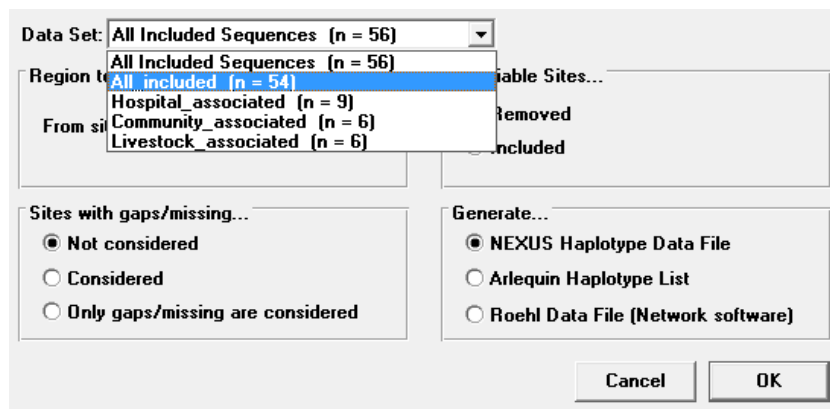


Figure 2.47: An overview screen showing the options that need to be selected to generate the Haplotype data.

Select 'All included (n=54) sequences' from the data set drop down menu. ensure that Generate NEXUS Haplotype Data File is selected and press OK. A new window will be opened (Figure 2.48), Save the file in NEXUS format (*.nex) and in the Current Output format (*.out). Tabulate the Haplotype's with its corresponding sequence name.

```

Total number of sites (excluding sites with gaps / missing data): 960

Sites with alignment gaps: not considered
Number of variable sites: 79

===== Haplotype Distribution =====
Number of haplotypes, h: 15
Haplotype diversity, Hd: 0.8675

Hap_1: 16 [1 3 8-21]
Hap_2: 3 [2 22-23]
Hap_3: 2 [4 56]
Hap_4: 6 [5-7 48-50]
Hap_5: 10 [24-33]
Hap_6: 6 [34-35 37 39-41]
Hap_7: 1 [36]
Hap_8: 1 [38]
Hap_9: 1 [42]
Hap_10: 3 [43-45]
Hap_11: 1 [46]
Hap_12: 1 [47]
Hap_13: 2 [51-52]
Hap_14: 2 [53 55]
Hap_15: 1 [54]

Hap_1: 16 [alpha-hemolysin-(hla MRSA_B51 502A CN1 ST228_18583 ST228_16125
ST228_16035 1819-97 ECT-R2 04-02981 ED98 Mu3 JH1 JH9 Mu50 N315]
Hap_2: 3 [MRSA_OP100 MSSA476 MW2]
Hap_3: 2 [MRSA_IP70 MRSA252]
Hap_4: 6 [MRSA_B53 MRSA_S62 MRSA_AS3 08BA02176 71193 ST398]
Hap_5: 10 [UA-S391_USA300 NRS-100 USA300-ISMM51 M1 VC40 USA300_TCH1516
Newman COL NCTC USA300_FPR3757]
Hap_6: 6 [XN108 Z172 Bmb9393 T0131 JKD6008 TW20]
Hap_7: 1 [6850]
Hap_8: 1 [LGA251]
Hap_9: 1 [RF122]
Hap_10: 3 [SA40 SA957 M013]
Hap_11: 1 [ED133]
Hap_12: 1 [JKD6159]
Hap_13: 2 [H-EMRSA-15 H050960412]
Hap_14: 2 [55/2053 TCH60]
Hap_15: 1 [CA-347]

```

Figure 2.48: An overview screen showing the results for the Haplotype

2.15.8 BioEdit

Open the sequence alignment file (alpha-hemolysin) in BioEDIT, a biological sequence alignment editor that is publicly available to download at: <http://www.mbio.ncsu.edu/bioedit/bioedit.html>. Once open, the screen will show the alignment (Figure 2.49) (Hall., 1999).



Figure 2.49: An overview screen of BioEDIT showing the sequence alignment file for alpha-hemolysin.

The names of the *Staphylococcus aureus* sequences are on the left hand side and its corresponding nucleotide sequence is on the right, altogether there are 54 sequences. Select the drop down menu under the 'Mode' tab and change to Edit. Using the haplotype data from section H, delete all *S. aureus* sequences that are duplicate haplotypes, for example, Haplotype_2 consists of 3 haplotype sequences, namely MRSA OP100, MSSA476 & MW2. Select the 2 latter sequences on the left hand column and simply delete. The end process should include only 15 haplotype sequences. Using BioEdit, export the file as a FASTA file.

2.15.9 Generating a NEXUS file

Open a new window on DnaSP and load the haplotype FASTA file that was formatted in BioEdit. On the home screen, select the 'Generate' tab and choose 'Haplotype Data File'. Save the output as a NEXUS file. The file generated will be used for further studies.

2.16 Population Analysis with Reticulate Trees (PopART)

PopART is a software package publicly available at: <http://popart.otago.ac.nz>. PopART version 1.7 was downloaded and used in this project (Popart., 2016). PopART provides a powerful and user friendly interface capable of producing minimum spanning, median-joining and TCS haplotype network (Leigh & Bryant., 2015). If the haplotype data also has corresponding geographical data, PopART can also use the map feature to place the the nucleotide sequences on a map viewer (Leigh & Bryant, 2015). PopART reads NEXUS-format alignments. Start by opening the file created in DnaSP *section 2.15.9*, in a text editor program, such as notepad and edit the data file. Arrange the data so that it is presented as shown in the Figure 2.50 below. In this example the format labels correspond with the sequence data sets for alpha-hemolysin, created in *section section 2.15.5*. The define sequence data sets include 3 traits: Livestock_acquired, Hospital_acquired and Community_acquired. The haplotype that corresponds with that specific trait is grouped together to indicate how many are present in each trait.

```
BEGIN TRAIT;
[The traits block is specific to PopART. The numbers in the matrix are number of
samples associated with each trait. The order of the columns must match the
order of TraitLabels. Separator can be comma, space, or tab.]
Dimensions NTRAITS=3;
Format labels=yes missing=? separator=Comma;
TraitLabels Livestock_acquired Hospital_acquired Community_acquired;
Matrix
alpha_hemolysin_Hap_1 0,1,2
alpha_hemolysin_Hap_2 0,1,0
alpha_hemolysin_Hap_3 3,0,0
alpha_hemolysin_Hap_4 0,0,5
alpha_hemolysin_Hap_5 0,0,2
alpha_hemolysin_Hap_6 1,0,1
alpha_hemolysin_Hap_7 0,2,0
alpha_hemolysin_Hap_8 0,4,0
alpha_hemolysin_Hap_9 0,0,1
```

Figure 2.50: An example of a NEXUS file showing the haplotype data for alpha-hemolysin.

Open PopART (Figure 2.52). Using the toolbar at the top of the viewer, the nex tab, load the edited aligned NEXUS file. The file will appear in the data viewer and will display the 3 traits and the number of sequences and samples that correspond to it.

Using the Network tab, select the type of network you wish to use, for example TCS network. It will be displayed in the viewer where the TCS network can be viewed and evaluated.

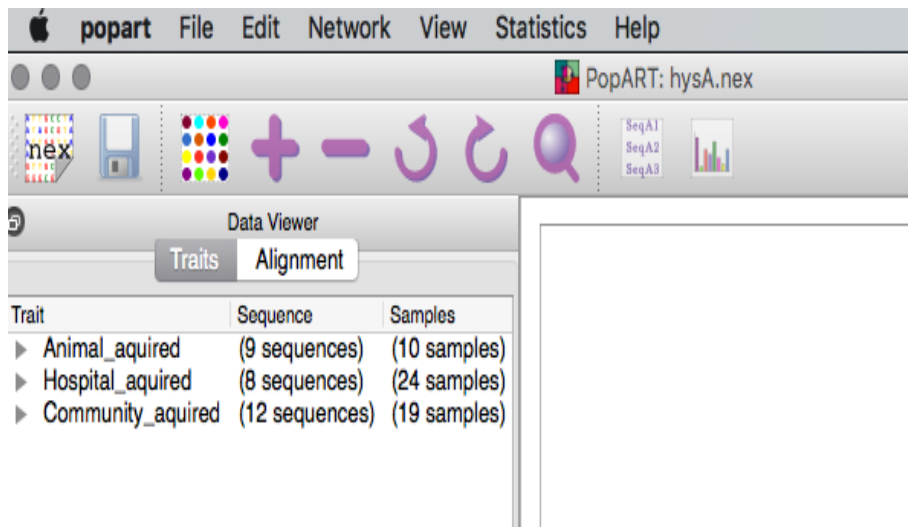


Figure 2.52: An overview screen showing the PopART home screen with the toolbar and loaded NEXUS files.

Chapter 3:

Strain Selection, Strain Classification and Virulence Factor Extraction

3.1 Introduction

Methicillin-resistant *S. aureus* (MRSA) is a major pathogen that can cause a diverse array of clinical conditions and diseases in humans ranging from superficial infections of the skin to more severe soft tissue infections (Casey *et al.*, 2013). MRSA infections are not restricted to humans; this bacterium has the ability to invade and cause harm to a diverse range of animal species including commercially important farm animals such as cattle, sheep, swine and poultry. Other animals such as horses, rabbits and companion animals, e.g. cats and dogs, are also susceptible. They can cause a variety of clinical symptoms and diseases which can cause illness and death.

Historically, MRSA only occurred in health-care settings and was therefore defined in the literature as a hospital-associated MRSA. (Sydnor & Perl, 2011). Eventually, MRSA was noted in the community, affecting patients who had no healthcare-related predisposing factors; these MRSA were subsequently termed community-associated MRSA (Deurenberg & Stobberingh, 2008). The community-associated MRSA became increasingly endemic and have now become the leading cause of skin and soft tissue infections (Chen *et al.*, 2011). MRSA have also been identified in food-producing animals since the early 1970s, however humans were not implicated with the infection, instead the infection was limited to the animal (Weese, 2010). More recently, a third group of MRSA strains have been characterised with the identification of a novel MRSA reservoir associated with farm animals, and in particularly swine. These strains have been termed livestock-associated MRSA; and these MRSA have been identified globally as an emerging zoonotic agent with livestock constituting a reservoir of infection (Arman-Lefevre *et al.*, 2005; Lindsay, 2010; Kimman *et al.*, 2013). The present study will investigate the presence of genes encoding virulence factors that are associated with the *S. aureus* strains and determine what differentiates the classifications (HA, CA, LA), as

well as determining if the presence of certain virulence factor genes were responsible for the difference in pathogenicity between the different classifications.

3.2 Aims

The aims of this present study were to:

- Identify previously published sequenced genomes of known human and animal associated *Staphylococcus aureus* strains and classify them into either hospital-associated, community-associated or livestock-associated strains as per the literature. The classification that was designated by the author was used in each case to designate if it was a HA, CA or LA strain.
- Sequence and assemble into contigs the genomes of novel characterised MRSA strains from various host species that were acquired from The Veterinary & Agrochemical Research Centre in Brussels, Belgium as well as clinical human isolates collected from The Manchester Metropolitan University, UK.
- From the literature, identify known *S. aureus* virulence factor genes and acquire their respective nucleotide sequences to establish if the virulence factors are present in both the previously published and novel strains; and extract its corresponding sequence.
- Use MAUVE to compare the published *S. aureus* genome sequences.

3.3 Materials and Methods

- The bacterial strains were cultured and maintained as carried out in Section 2.1, 2.2 and 2.3 and the isolation of Bacterial Genomic DNA using The GenElute™, as carried out in Section 2.4
- The method for bacterial genomic DNA concentration and purity checks using spectrophotometric analysis and gel electrophoresis, as carried out in Section 2.5 & Section 2.6
- The sequencing of the novel strains of *S. aureus* using Ion Torrent next generation sequencing, as carried out in Section 2.7.
- The method for analysing and visualizing data from the novel *S. aureus* Next Generation Sequencing (NGS) as described in section 2.8.
- The method for acquiring *S. aureus* genomes and virulence factors that have already been published as carried out in Section 2.9.
- The method for identifying if virulence factor genes are present in some or all of the *S. aureus* strains as carried out in Section 2.11.
- Using MAUVE for visually viewing and comparing the published *S. aureus* genomes as carried out in Section 2.10.

3.4 Results

3.4.1 Strain Selection

A total of 50 *Staphylococcus aureus* strains from human and animal origin were identified from GenBank®, an online genetic sequence database that is maintained by the National Centre for Biotechnology Information (NCBI). They were downloaded from the online database as described in section 2.9. Only complete, fully assembled and annotated genomes were collected into a local database. Important features, such as recording the NCBI accession number and the GenInfo Identifier (GI) number, the latter which the NCBI intend to discontinue from September 2016 were also recorded into the database. This is important as the sequence in the online database can be updated or changed and may differ if this study is reproduced. The *S. aureus* strains that were selected are listed in Table 3.1

3.4.2 Strain classification

Once the *S. aureus* genomes were acquired and placed in the local database, they were subsequently classified as per the literature as either; hospital-associated, community-associated or live-stock associated *S. aureus* strains.

At the initiation of this study there was no general consensus on a definitive definition of hospital-associated, community-associated or livestock associated *S. aureus* strain. So for the purpose of this study it was decided that a working definition that matched the descriptive definitions used in the range of literature associated with *S. aureus* was used. Subsequent to the investigation of this study an almost identical set of definitions (that were described in Sections 1.8.1, 1.8.2 & 1.8.3) have been published by Bal *et al.* (2016). It is noteworthy to point out that information on the strains downloaded from GenBank (Table 3.1) was very limited with many journals not mentioning where the strain was isolated or under what circumstances (i.e hospital setting or community setting), so it would have been very difficult to determine the classification

of the strain. However, during the extensive literature review, the authors of the papers of the where the respective strain was studied almost always concluded what the classification of the strain was.

A thorough search of PubMed and GenBank was conducted and the table below was subsequently produced (Table 3.1). This table identifies the *S. aureus* strain, the NCBI/ GI accession numbers, the length of the genome, the classification and its corresponding literature reference. Not all strains are classified, as there was limited information present in the literature to classify them. In total, 3 strains could not be classified in Table 3.1; this included strains TCH60, NRS100 & 55/2053. These strains were excluded from all further analysis.

Table 3.1: Table identifying the the *S. aureus* strain acquired from GenBank, its corresponding NCBI/ GI accession numbers, the length of the genome in base pairs, classification and its corresponding reference in the literature.

S. aureus strain	NCBI accession & GI number	Length (bp)	Information about the Strain	Classification	Reference:
08BA0216	NC_018608.1 GI: 404477334	2782313	A livestock associated MRSA strain that was isolated in 2008 in Canada from a human postoperative surgical site. This strain contains a novel staphylococcal cassette chromosome <i>mec</i> element type V (SCC <i>mecV</i>).	Livestock Associated	Golding <i>et al.</i> , 2012
04-02981	NC_017340 GI: 387149188	2821452	A highly prevalent strain found in hospitals in Central Europe that was initially isolated in 2004 from Köln, Germany.	Hospital Associated	Xue, <i>et al.</i> , 2011 Nübel <i>et al.</i> , 2010
11819-97	NC_017351.1 GI: 385780298	2846546	This 11819-97 is a ST80-IV representative that was isolated from a skin abscess from Denmark, in 1997. ST80-IV was initially identified in the early 1990's and has historically dominated community associated infections in major parts of Europe that caused mostly skin and soft tissue infection.	Community Associated	Stegger <i>et al.</i> , 2012
55/2053	NC_022113.1 GI: 532358222	2756919	Strain was sequenced on the 15 th August 2015. No information found in literature.		Genome acquired from GenBank

502A	CP007454.1 GI: 595636499	2764699	<p>This strain was initially isolated from a nurse who become colonised when caring for 40 neonates, in 1963.</p> <p>In the 1960's and 1970's the 502A strain was used in bacterial interference programs. Infants were intentionally colonised with the 502A strain to prevent colonisation with more invasive strains. This was eventually stopped as it caused several infections and deaths.</p>	Hospital Associated	Parker <i>et al.</i> , 2014
6850	NC_022222.1 GI:537441500	2736560	<p>6850 is a highly cytotoxic and clinically virulent Methicillin Sensitive <i>S. aureus</i> strain (MSSA).</p> <p>It was isolated from a patient with complicated <i>S. aureus</i> bacteraemia associated with osteomyelitis and septic arthritis.</p>	Hospital Associated	<p>Fraunholz <i>et al.</i>, 2013</p> <p>Vann & Proctor, 1987</p>
71193	NC_017673.1 GI:386727822	2715000	<p>The initial ST398 MSSA strain was designated 71193 or ST398NMO1. It was isolated in 2004, with swine being identified as the major reservoir for infection.</p>	Livestock Associated	<p>Bhat <i>et al.</i>, 2009</p> <p>Uhlemann <i>et al.</i>, 2012</p>
Bmb9393	NC_021670.1 GI:521210823	2980548	<p>This strain was isolated in 2003 from a case of nosocomial bloodstream infections in Rio de Janeiro. It is now widely disseminated in Brazilian hospitals.</p>	Hospital Associated	<p>Amaral <i>et al.</i>, 2005</p> <p>Costa <i>et al.</i>, 2013</p>

COL	NC_002951.2 GI:57650036	2809422	A strain that was isolated from an operating theatre in Colindale, England in the early 1960's. It was initially identified as a penicillinase-negative strain.	Hospital Associated	Gill <i>et al.</i> , 2005
CN1	NC_022226.1 GI:537459744	2751266	This isolate CN1 is also termed HM1. It was initially obtained from the pus of a necrotising fasciitis infection in an 80-year old patient in the Seoul area of South Korea. Subsequently, it was also isolated from elderly patients between 2005 to 2007; the sputum of a 68-year old male, the blood of a 71-year old female, the pus of an 82-year old male and 66-year old female. These are the most common strains of MRSA circulating the community in South Korea.	Community Associated	Li <i>et al.</i> , 2010 Lee <i>et al.</i> , 2010 Chen <i>et al.</i> , 2013
CA-347	NC_021554.1 GI:514064966	2850503	A community USA600 strain that was isolated from a bacteremia patient in California, in 2005.	Community Associated	Stegger <i>et al.</i> , 2013
ECT-R2	NC_017343.1 GI:384863396	2729540	This strain caused a clonal outbreak in Ostergotland County, Sweden. It was originally isolated from a wound and was identified as a multi resistant methicillin-susceptible <i>S. aureus</i> (MR-MSSA).	Community Associated	Lindqvist <i>et al.</i> , 2012

ED133	NC_017337.1 GI:384546269	2832478	Strain 1774, designated ED133 (CC133 lineage/ ST133) was isolated from a case of bovine mastitis in France, in 1997.	Livestock Associated	Ben Zakour <i>et al.</i> , 2008 Guinane <i>et al.</i> , 2010
ED98	NC_013450.1 GI:269201690	2824404	This strain was isolated from a chicken with bacterial chondronecrosis with osteomyelitis (BCO) in Ireland. Isolation date 1996 - 1997.	Livestock Associated	Rodgers <i>et al.</i> , 1999 Lowder <i>et al.</i> , 2009
H-EMRSA-15	NZ_CP007659.1 GI:749295050	2846320	A predominant health-care associated MRSA that was characterised as a prolific biofilm produce, emerged in the UK, in 1991 and rapidly displaced most other epidemic methicillin-resistant <i>Staphylococcus aureus</i> (EMRSA) strains.	Hospital Associated	O'Neil <i>et al.</i> , 2001 Sabirova <i>et al.</i> , 2014
HO 5096 0412	NC_017763.1 GI:386829725	2832299	HO 5096 0412 is representative of the EMRSA-15 and is 99.4% similar to it. It was isolated in Suffolk, from a fatal neonatal infection, in 2005.	Hospital Associated	Sabirova <i>et al.</i> , 2014 Holden <i>et al.</i> , 2013
JKD6008	NC_017341.1 GI:384860682	2924344	A strain that shows intermediate-level vancomycin resistance isolated in 2003 from the blood stream of a patient in New Zealand.	Hospital Associated	Howden <i>et al.</i> , 2010 Howden <i>et al.</i> , 2006

JKD6159	NC_017338.1 GI:384548923	2811435	A dominant Australian community-associated MRSA clone. It was identified in an intravenous drug user causing septicaemia, musculo-skeletal abscess and outbreak of skin infection.	Community Associated	Chua <i>et al.</i> , 2010
JH1	NC_009487.1 GI:148266447	2906700	JH1, JH2, JH3, JH4, JH5, JH6 & JH9 are a series of MRSA strains that were isolated from a patient with congenital heart disease and endocarditis.	Hospital Associated	Mwangi <i>et al.</i> , 2007 Sieradzki <i>et al.</i> , 2003
JH9	NC_009632.1 GI:150392480	2906507	This patient underwent antibiotic therapy which included vancomycin, rifampin and the β -lactam antibiotic imipenem and these MRSA evolved as a result of the impact of the therapy, the patient eventually died. JH1 was isolated at the beginning of this chemotherapy on the 20/7/2000 and is susceptible to vancomycin and JH9 was isolated at the end of the chemotherapy and showed decreased susceptibility to vancomycin on the 13/10/2000.		
LGA251	NC_017349.1 GI:387779217	2750834	This strain was isolated from a bulk milk sample from a farm in England, in May 2007.	Livestock Associated	Garcia-Alvarez <i>et al.</i> , 2011

M1	NC_021059.1 GI:479328021	2864125	This strain was initially isolated in Denmark, in 2003. It subsequently spread through the local hospital and eventually the nursing homes in the surrounding areas, affecting 501 patients and staff. This strain was not fully under control until 2009.	Community Associated	Larner-Svensson <i>et al.</i> , 2013
Mu50	NC_002758.2 GI:57634611	2878529	A clinical strain which showed resistance to vancomycin therapy from a 4-month old male Japanese infant. 4-month old infant Japanese male patient who underwent heart surgery, in 1996. The site of the incision developed a purulent discharge which showed MRSA.	Hospital Associated	Gill <i>et al.</i> , 2005 Hiramatsu <i>et al.</i> , 1997
Mu3	NC_009782.1 GI:156978331	2880168	Isolated from the purulent sputum of a 64-year old Japanese male after a surgical operation for lung cancer, in 1996.	Hospital Associated	Hiramatsu <i>et al.</i> , 1997
MW2	NC_003923.1 GI:21281729	2820462	A highly virulent strain that was linked to four paediatric deaths in Minnesota and North Dakota. It was initially isolated from a 16-month old female from North Dakota, in 1998.	Community Associated	CDC, 1999 Baba <i>et al.</i> , 2002
MSSA476	NC_002953.3 GI:49484912	2799802	Isolated from a 9-year old male, in 1998 with primary tibial osteomyelitis and bacteraemia who had no predisposing risk factors.	Community Associated	Holden <i>et al.</i> , 2004

MSHR1132	NC_016941.1 GI:379794527	2762785	Isolated from a blood culture of an indigenous female with necrotising fasciitis in September 2006. It is a strain that is dominant among indigenous communities in the Northern Territory of Australia.	Community Associated	Holt <i>et al.</i> , 2006 McDonald <i>et al.</i> , 2006
MRSA252	NC_002952.2 GI:49482253	2902619	Isolated in 1997 from a 64-year-old female who acquired MRSA postoperatively. The organism was part of a miniature outbreak at an intensive treatment unit where all three of the infected patients who became infected died as a direct result of MRSA septicaemia.	Hospital Associated	Holden <i>et al.</i> , 2004
M013	NC_016928.1 GI:379019812	2788636	A PVL positive community-associated Taiwanese strain that was isolated from a wound specimen of a paediatric outpatient in 2002.	Community Associated	Huang <i>et al.</i> , 2012
Newman	NC_009641.1 GI:151220212	2878897	Isolated from a male, in 1952 from a case of secondarily infected tubercular osteomyelitis.	Hospital Associated	Baba <i>et al.</i> , 2008 Duthie & Lorenz, 1952
NRS100	CP007539.1 GI:610395799	2823087	NRS100 is an identifier from the Network on Antimicrobial Resistance in <i>Staphylococcus aureus</i> (NARSA), and may be the same as strain COL.		Genome acquired from GenBank Gill, 2016

NCTC 8325	NC_007795.1 GI:88193823	2821361	This is a registered strain of the National Collection of Type Cultures and has been widely used as laboratory <i>S. aureus</i> strains.	Hospital Associated	Baba <i>et al.</i> , 2002
N315	NC_002745.2 GI:29165615	2814816	A Japanese strain identified in 1982 from a pharyngeal smear. It was named after Nagasaki University Hospital, Japan, Isolate number 315.	Hospital Associated	Gill <i>et al.</i> , 2005 Kuroda <i>et al.</i> , 2001 Baba <i>et al.</i> , 2002
RF122	NC_007622.1 GI:82749777	2742531	A bovine mastitis strain isolated from Ireland, in 1993.	Livestock Associated	Fitzgerald <i>et al.</i> , 2001 Herron-Olsen <i>et al.</i> , 2007 Sung <i>et al.</i> , 2008
SA40	NC_022443.1 GI:545636471	2728308	A commensal Asian Pacific clone, that frequently colonises healthy children, isolated from a nasopharyngeal swab in 2005.	Community Associated	Chen <i>et al.</i> , 2013
SA957	NC_022442.1 GI:545633528	2789538	A virulent Taiwanese clone that was isolated from a 16-year male, in 2000. It shows enhanced virulence in both humans and animal models.	Community Associated	Chen <i>et al.</i> , 2013
ST228 isolate 18385	NC_020568.1 GI:470223734	2759328	Eight different isolates of <i>S. aureus</i> belonging to ST228 lineage (spa type t041) collected between 2001 - 2008. at the tertiary care hospital in of Lausanne.		

ST228 isolate 16125	HE579067.1 GI:408430526	2759457	The study concludes that eight isolates belonging to a lineage that spread over ten years within a tertiary care hospital with genomes that were extremely closely related. Only three of the eight isolates were included in this study because the dataset would have been too large.	Hospital Associated	Vogel <i>et al.</i> , 2010
ST228 isolate 16035	HE579065.1 GI:408428537	2759835			
ST398 (S0385)	NC_017333.1 GI:387601291	2872582	The initial ST398 strain was designated 71193 or ST398NMO1. It was isolated in 2004 from swine. A strain that poses a risk factor for humans working with intensively farmed animals such as pigs, cattle and poultry.	Livestock Associated	Schijffelen <i>et al.</i> , 2010 Hernandez <i>et al.</i> , 2013
T0131	NC_017347.1 GI:384868588	2913900	An epidemic hospital-associated strain that was recovered from an 87-year old patient in China, in 2006.	Hospital Associated	Li <i>et al.</i> , 2011
TW20	NC_017331.1 GI:387141638	3043210	Isolated in 2003 during a 2-year outbreak of a highly transmissible MRSA, Sequence Type 239 in an intensive care unit in London.	Hospital Associated	Edgeworth <i>et al.</i> , 2007 Holden <i>et al.</i> , 2010
TCH60	NC_017342.1 GI:384865886	2802675	A PVL (Panton-Valentine Leukocidin) positive strain that was isolated from the skin of a patient in Texas, USA.		McGavin <i>et al.</i> , 2012.

USA300_TCH151 6	NC_010079.1 GI:161508266	2872915	USA300 strains are implicated in epidemiologically un-associated sporadic outbreaks of skin and soft tissue infections in healthy individuals. Isolate was obtained from an adolescent patient with severe sepsis syndrome.	Community Associated	Highlander <i>et al.</i> , 2007
USA300-ISMMS1	CP007176.1 GI:584463070	2921008	Transmission of MRSA via a deceased donor liver transplantation from a 40-year old woman who presented with drug overdose and multifocal embolic cerebrovascular accident and MRSA endocarditis to a 64-year old man with cirrhosis from hepatitis C virus and hepatocellular carcinoma.	Community Associated	Altman <i>et al.</i> , 2014.
USA300_FPR373 7	NC_007793.1 GI:87159884	2872769	This strain was isolated from the wrist of a 36-year old HIV-positive male who has a history of intravenous drug use in September 2000. It was isolated by the clinical laboratory of a San Francisco hospital.	Community Associated	Diep <i>et al.</i> , 2006
UA-S391_USA300	CP007690.1 GI:640834542	2872916	A USA300 clonal lineage what was found to have a genome map identical to that of USA300_TCH1516. This strain is a prolific biofilm producer.	Community Associated	Sabirova <i>et al.</i> , 2014

VC40	NC_016912.1 GI:379013365	2692570	<p>The overuse of Vancomycin has led to the emergence of vancomycin-intermediate (VISA) and vancomycin-resistant (VRSA) <i>S. aureus</i> strains. It is termed VC40 because it is resistant to a vancomycin concentration of $\geq 40\mu\text{g/ml}$.</p> <p>VC40 has been generated by serial passage of RN4220 strain with a <i>mutS</i> gene mutation.</p>	Hospital Associated	<p>Schaaff <i>et al.</i>, 2002</p> <p>Sass <i>et al.</i>, 2012</p> <p>Kreiswirth <i>et al.</i>, 1983</p>
XN108	CP007447.1 GI:657154530	3052055	<p>Isolated from a 34-year old patient who had a wound infection and was hospitalised in February 2004 and underwent two skin operations, in mainland China.</p>	Hospital Associated	<p>Zhang <i>et al.</i>, 2013</p> <p>Zhang <i>et al.</i>, 2014</p>
Z172	NC_022604.1 GI:554642795	2987966	<p>A strain that was isolated from a blood specimen of an elderly intensive care unit patient. It is among the most prevalent of HA-MRSA strains isolated in Taiwan.</p>	Hospital Associated	<p>Chen <i>et al.</i>, 2013</p>

3.4.3 Mauve analysis

MAUVE was used to compare the known and published *S. aureus* genomes that were extracted from GenBank®, as described in Section 2.10.

They were initially compared as total *S. aureus* sequences (Figure 3.1) and then as groups: Hospital-associated (Figure 3.2), Community-associated (Figure 3.3), Livestock-associated (Figure 3.4) & Comparing the 3 groups (Figure 3.5).

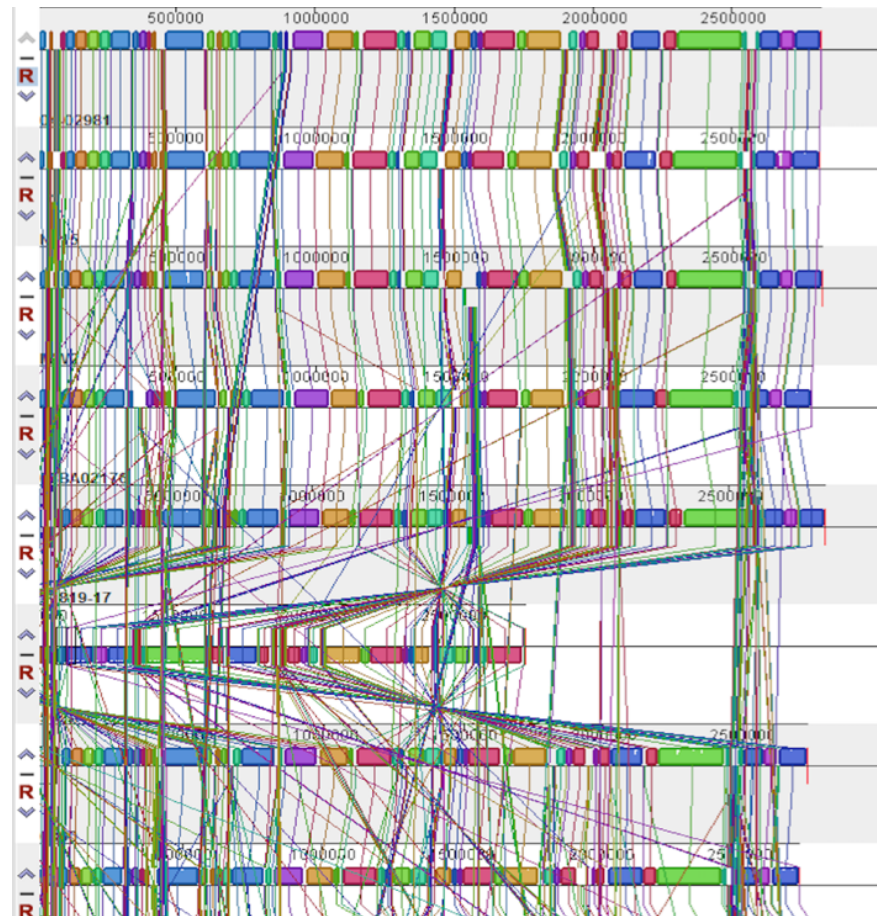
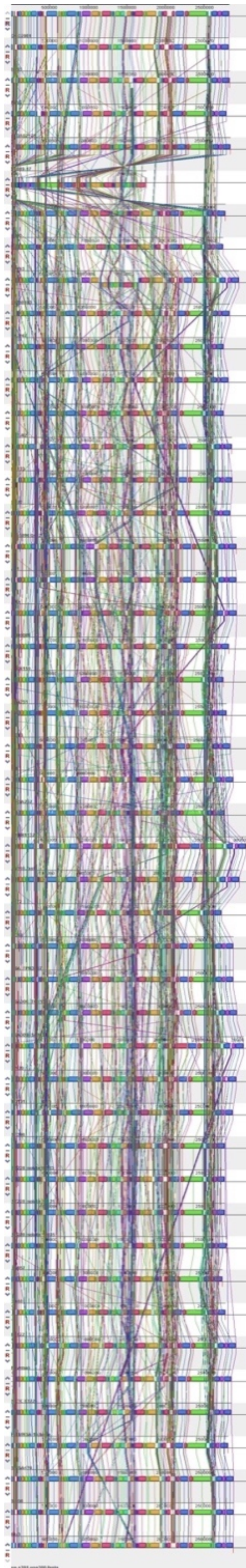


Figure 3.1: An output of MAUVE representing 8 of the 47-published Methicillin-Resistant *S. aureus* genomes from Table 3.1. 3 strains which could not be classified were removed from this analysis (NRS100, TCH60 & 55/2053).

The in-set image is a magnified view of the MAUVE output and shows 8-MRSA strains. This is shown for representative purposes.

The 6th sequence seems to be shorter in the inset figure, however, this is caused by a difference in the 'origin' used in assembling the genome.

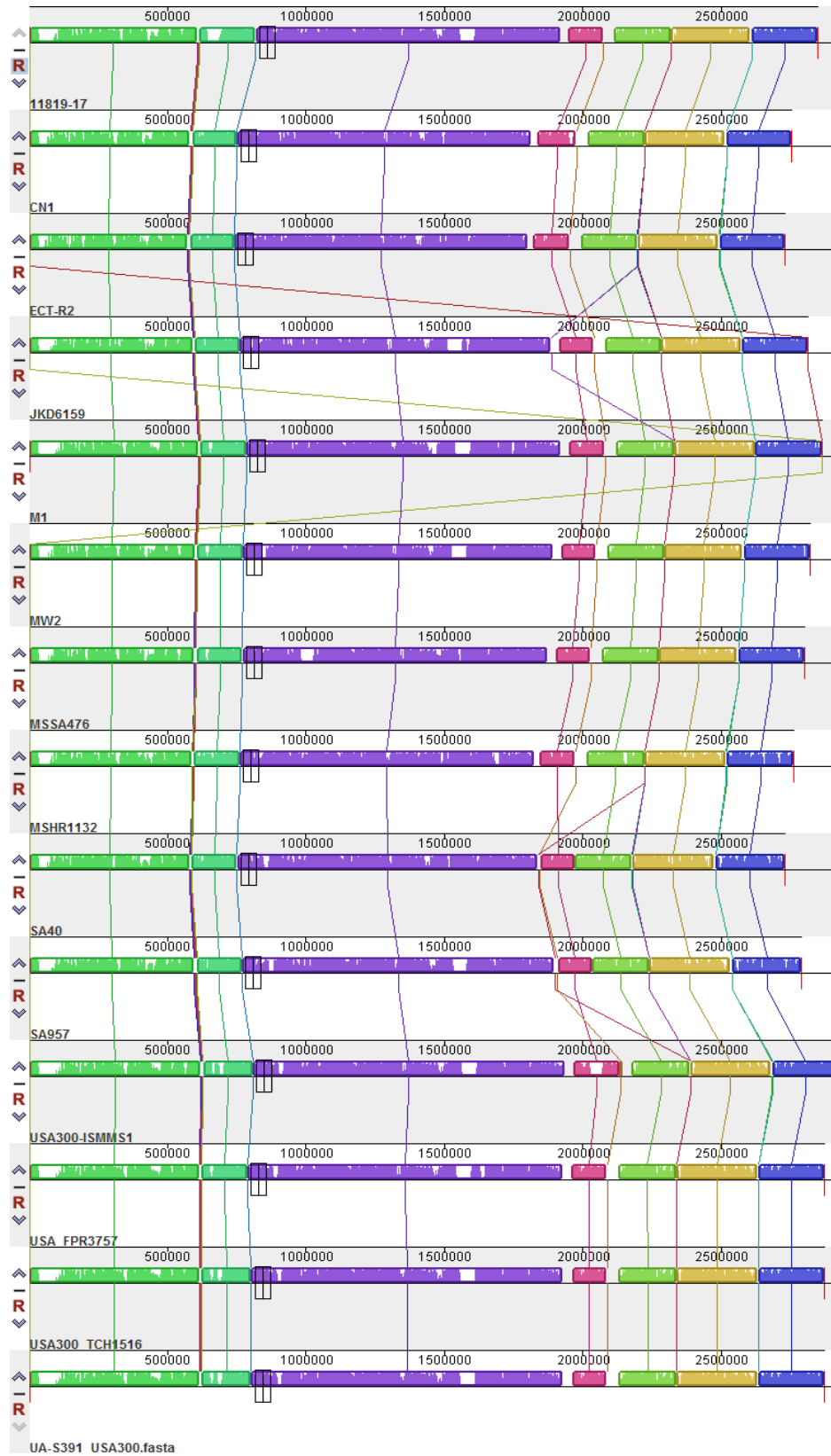


Figure 3.3: An output of MAUVE representing the 16-published Community-associated *S. aureus* genomes from Table 3.1.

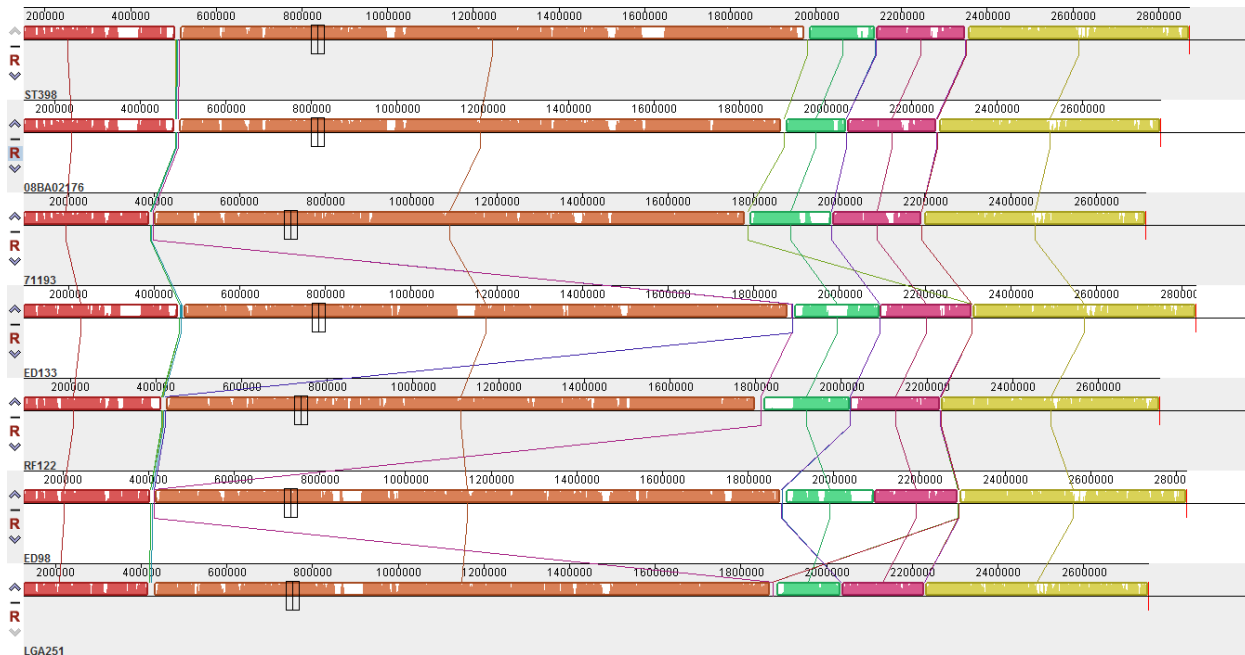


Figure 3.4: An output of MAUVE representing the 7-published Livestock-associated *S. aureus* genomes from Table 3.1.

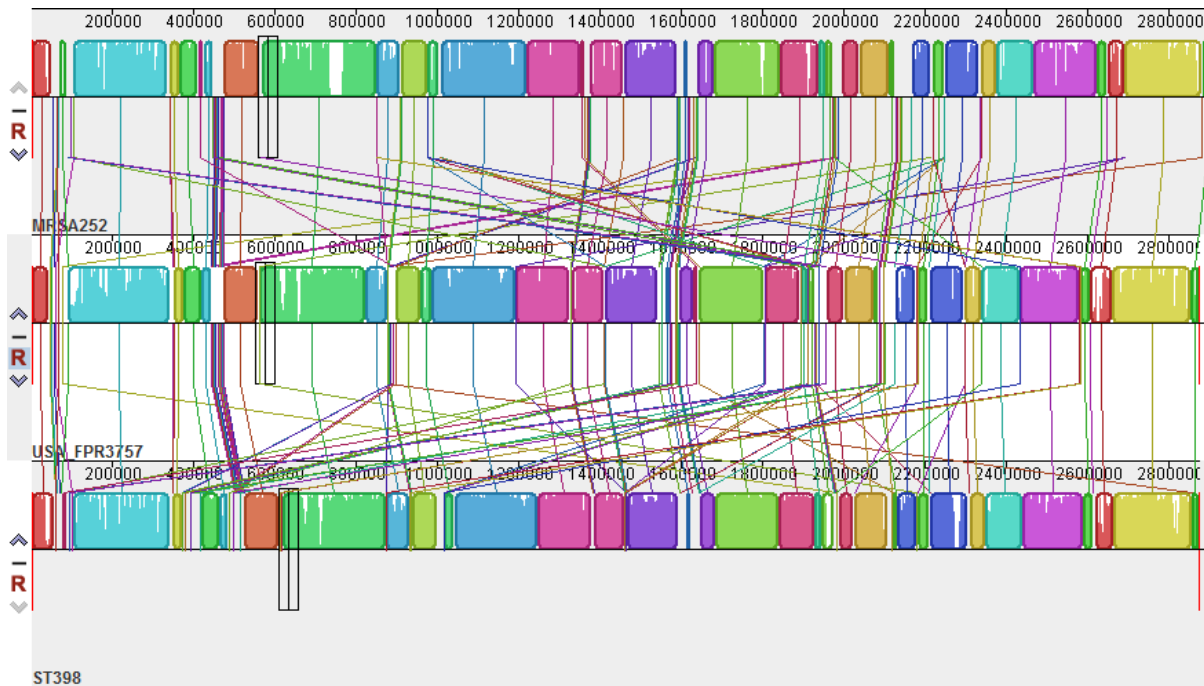


Figure 3.5: An output of MAUVE comparing a Hospital-associated (MRSA252), Community-associated (USA300_FPR3757) and Livestock-associated (ST398) *S. aureus* strains.

An output of MAUVE representing the 47-published *S. aureus* genomes is shown in Figure 3.1. This was an important step to visually see which genes these genomes share and which regions were unique to a particular genome(s).

It is evident from Figure 3.1 that the 47-strains of *S. aureus* share regions that are homologous, but it is equally evident that the strains have regions that are unique to that strain. It was very difficult to view all 47 strains at one time and extract any information that may be useful. Figure 3.1 also shows an inset image, it is a magnified view of the MAUVE output and shows 8 -MRSA strains.

It was decided that a smaller number of strains as well as comparing the three classes to examine the similarities and differences within each class would be applied to any future MAUVE analysis.

Outputs of MAUVE representing the Hospital-associated, Community-associated and Livestock-associated groups were then created. Figure 3.2 is the output of Hospital-associated *S. aureus* (n=24). It is noticeable on the first instance that strain 502A differs greatly from the others (3rd strain down). All strains contain similar homologous segments; however, strain 502A's segments position differs greatly. The reason for this is that it is not a genuine rearrangement of the sequence, it is most likely caused by a different 'origin' when assembling the genome. It is also possible that the sequence in the database is for the complementary strand, when comparing it to the other sequences in this output. Although it is possible to undertake this analysis with the complementary sequence for 502A it would not be an easy task to rearrange the sequence. Strain 502A is unique because it was used to intentionally colonise infants with *S. aureus* to prevent colonisation with more invasive strains, this practice was stopped as it caused severe infection and deaths (Parker *et al.*, 2014). The next noticeable strain is Bmb9393 (the 18th strain down from the top). This strain has undergone inversion of 3 regions or local collinear blocks, which is clearly shown in the output. It is interesting to note that this specific hospital-associated strain isolated from a case of nosocomial bloodstream

infection in Rio de Janeiro in 2003 is still widely disseminated in Brazilian hospitals (Costa *et al.*, 2013). The average lengths of the individual hospital-associated genomes vary; from 2.6 Mbp for VC40 to 3.0 Mbp for XN108. It is worth noting that the whole-genome sequencing of *S. aureus* strain T0131 generated was 2.9 Mbp (Li *et al.*, 2011). It is therefore reasonable to suggest that Insertion/Deletion (Indel) events have taken place amongst the various genomes. The source of the strain may also be an indicator for the length of the strains. Strain 6850 was isolated from a patient with complicated *S. aureus* bacteraemia associated with osteomyelitis and septic arthritis whereas VC40 was generated by serial passage of RN422 strain that is resistant to Vancomycin.

Figure 3.3 is an output of MAUVE representing the Community-associated *S. aureus* strains (n=14). The Community-associated strains look very similar, this is demonstrated by the homologues segments and their respective nucleotide positions correlating between the various strains. It is noticeable that there is not much difference between the four USA300 strains (UA-S391_USA300, USA300_TCH1516, USA300_FPR3757 & USA300-ISMMS1). The alignment of these strains look almost identical with the exception of strain USA300-ISMMS1 that differs slightly at nucleotide position 2000000. The USA300 strains are all 2.8 Mbp, again with the exception of USA300-ISMMS1 which is 2.9 Mbp. This study coincides with the findings of Highlander *et al* (2007) and suggests that because the strains are highly similar based on the whole genome alignments undertaken with MAUVE, the differences in pathogenesis are likely due to very subtle changes of the virulence factor genes that have ultimately enhanced the organism's fitness and pathogenicity. These subtle changes can be down to Single Nucleotide Polymorphism (SNP), a DNA sequence variation occurring when a single nucleotide changes and has the potential to alter the DNA and subsequent protein sequence.

Figure 3.4 is an output of MAUVE representing the Livestock-associated *S. aureus* strains (n=7). Strain ST398 and 08BA0126 are strikingly similar sequences of which both were isolated from humans. 71193 a methicillin-sensitive *S. aureus* (MSSA) strain isolated from humans also differs considerably at the segment

(coloured green) at points 1794397 to 1976645. Although there were no annotations at these points to determine what genes could be present, a blast search of the sequence at these point also didn't yield any results. ED133 and RF122 were both isolated from cases of bovine mastitis and it was interesting to note the two strains are significantly different from each other. Strains LGA251 and ED98 were isolated from a bulk milk sample and a chicken with chondronecrosis, respectively. Both strains share characteristics that are extremely similar to each other but different from the other LA-MRSA strains. For example, the red segment (located at nucleotide points 125436 to 413995) and orange segment (located at nucleotide points 429183 to 1861135) are similar but differ greatly from the green segment (located at nucleotide points 1887170 to 2030365), strain ED98 has regions that are strain specific and need further investigation.

Finally, Figure 3.5 is an output of MAUVE comparing a strain of hospital-associated (MRSA252), community-associated (USA300_FPR3757) and livestock-associated (ST3980) *S. aureus* strains. This is an interesting figure as it shows that not much difference exists between the various strains. The homologous regions that are shown in coloured blocks are shared throughout the 3 sequences. There are of course regions within the homologous blocks regions that are unique to the strain.

3.4.4 Novel bacterial strain selection and classification:

Along with the 50 published *S. aureus* isolates that were identified in this study, a selection of novel MRSA isolates from human and animal hosts were also investigated as described in section 2.1.

The information for these strains was provided by the collaborators. They were isolated from either human or animal sources, and were designated as the following: AS3, swine; B53, cattle; S62, chicken; B51, burns patient; IP70, inpatient and OP100, outpatient. It was confirmed by Dr Patrick Butaye (Veterinary & Agrochemical Research Centre, Belgium) and Prof. Valerie Edward-Jones (Manchester Metropolitan University) that the novel strains were all methicillin-resistant *S. aureus* isolates. The six strains were classified as either hospital-associated or live-stock associated MRSA strains, as shown in Table 3.2. They

were classified according to the information that was provided about them and their classification provided to us by our collaborators.

<i>Staphylococcus aureus</i> Strain	Isolation Source	Classification	Reference:
AS3	Swine Strain	Livestock Associated	Strain acquired from Dr Patrick Butaye. The Veterinary & Agrochemical Research Centre, Belgium
B53	Cattle Strain	Livestock Associated	
S62	Chicken Strain	Livestock Associated	
B51	Burns Patient	Hospital Associated	Strain acquired from Prof Valerie Edward-Jones. Manchester Metropolitan University, UK
IP70	In-patient strain	Hospital Associated	
OP100	Out-patient	Hospital Associated	

Table 3.2: Table identifying the the *S. aureus* strain acquired from the collaborators indicating its host, and the classification it has been designated.

3.4.5 Bacterial genomic DNA isolation

The six bacterial strains were cultured and maintained as described in section 2.1, 2.2 and 2.3. The isolates were confirmed as being methicillin-resistant *S. aureus* (MRSA) samples so isolate identity was not carried out. Pure RNA-free genomic DNA was isolated from the six strains, as described in section 2.3.

3.4.6 Spectrophotometric analysis

The purity and concentration of the DNA was determined by spectrophotometric analysis using the Nanodrop equipment as described in Section 2.5 and the results are given in Table 3.3. The results from all samples indicate that the DNA isolation had been successful. The A260/ A280 reading for DNA should be in the region 1.8 and RNA should be in the 2.0 range. It is evident from the results that pure DNA has been successfully isolated because all six of the reading were in the 1.8 range. The concentration is also in the expected region of 84.5 ng/μl – 190.0 ng/μl as indicated in the equipment's user manual.

	OP100	B51	IP70	B53	S62	AS3
	Out patient	Burns Patient	In Patient	Cattle Strain	Chicken Strain	Swine Strain
Concentration (ng/ul)	84.5	192.0	152.7	95	85.5	108.0
A260/ A280	1.837	1.82	1.848	1.845	1.839	1.846
A260/ A230	3.25	2.233	3.211	1.9	2.375	2.077

Table 3.3a: The results of the spectrophotometric analysis for the six *S. aureus* samples; which include the average concentration and the A260/ A280 readings.

3.4.7 Gel Electrophoresis

The isolated genomic bacterial DNA samples were analysed by gel electrophoresis as described in Section 2.6 and the results are given in Figure 3.6.

The results of the Gel Electrophoresis indicate that the DNA is intact and RNA-free as expected. Although there is little streaking and migration the majority of the DNA appears as large bands at the top of the gel; with the primary band for the isolated DNA being far larger than the largest fragment in the ladder (1000 bp). This is relevant because it shows that DNA fragments are much greater than the 100 – 1000 bp ladder that was used. This clearly indicates the successful isolation of RNA-free DNA isolation of these bacteria.

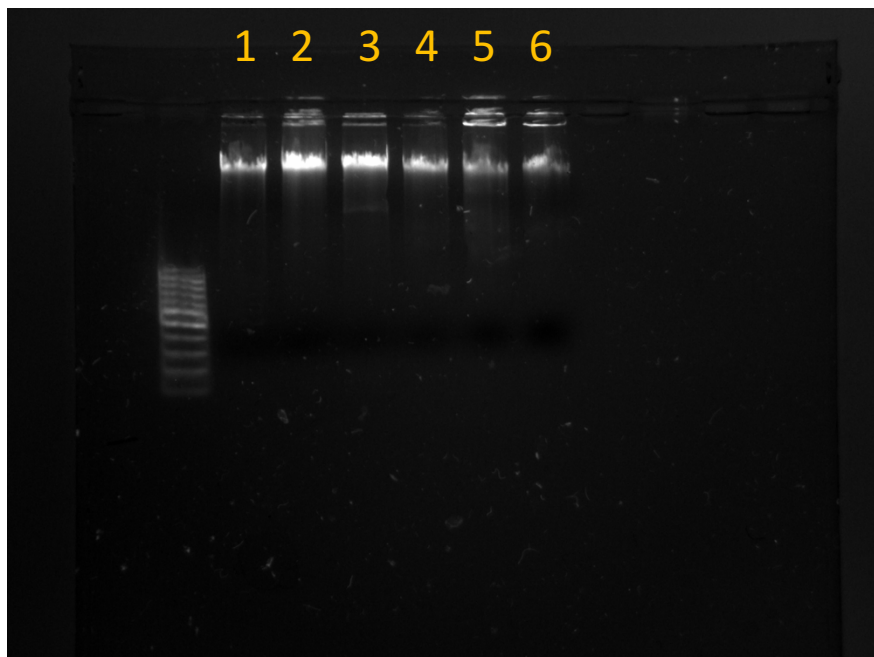


Figure 3.6: An image of the gel electrophoresis results, the bands from left to right show; the ladder, MRSA OP100, B51, IP70, B53, S62 and AS3. The gel was run in a FlowGen Gel Electrophoresis system at a voltage of 200V for 3 hours and viewed using a Molecular Imager® Gel Doc™ XR System.

3.4.8 Next Generation Sequencing

The high quality RNA-free genomic DNA (gDNA) that was isolated from the six MRSA isolates (OP100, B51, IP70, B53, S62 & AS3), was used with the Ion Torrent™ next generation sequencing kit; and carried out as described in Section 2.7. The quality assessment results are from each stage of sequencing is mentioned below.

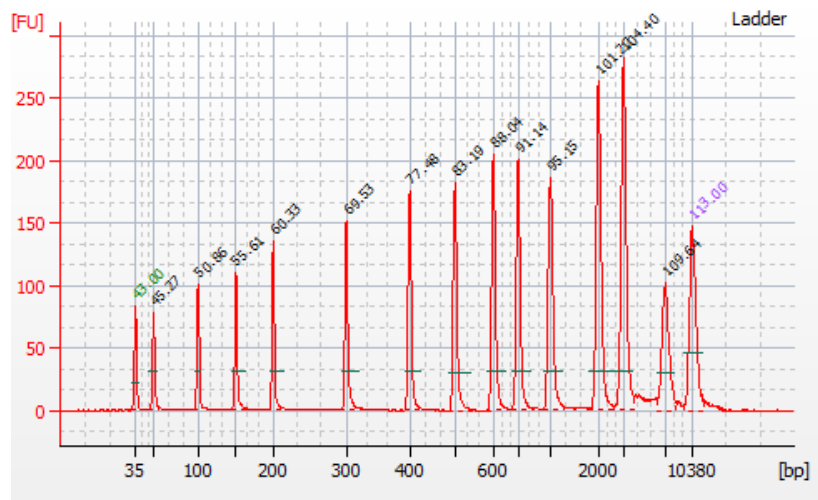
3.4.8.1 Fragmentation of the DNA

The initial stage of the sequencing, involved the fragmentation of the DNA and then the subsequent clean up of the DNA as described in Section 2.7.1a and Section 2.7.1b, respectively.

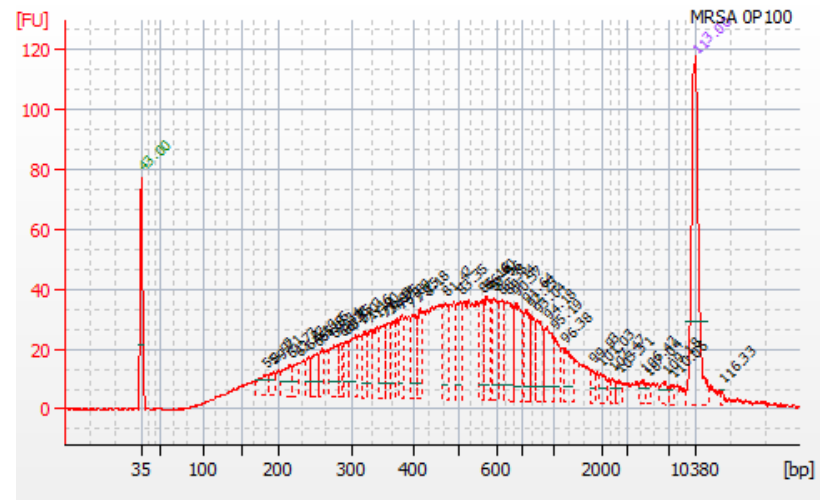
To ensure that the fragmented DNA was within the required size after the initial step, the Bioanalyser® instrumentation was used to assess its average size as described in Section 2.7.1c. The results of this is given in Figure 3.7a-g.

Figure 3.7a is an electropherogram of the high sensitivity ladder showing the upper and lower markers. These markers are indicators of a successful run. The ladder also shows the size of base pairs that can be expected; which range from 35 bp to 10380 bp.

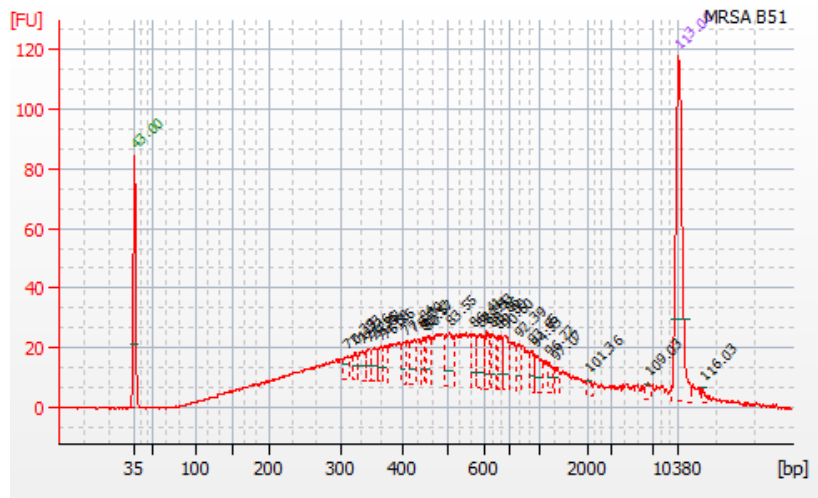
A 400 bp ladder was used as the expected size of the DNA once fragmented should be between 350 – 450 bp's. The results shown in Figures 3.7b-g that the six isolates have been fragmented near to the expected fragment size, indicated by the peak in the graph. The results show the following: OP100 peaks at 481 bp; B51 peaks at 479 bp; IP70 peaks at 483 bp; B53 peaks at 478 bp; S62 peaks at 481 bp and AS3 peaks at 488 bp.



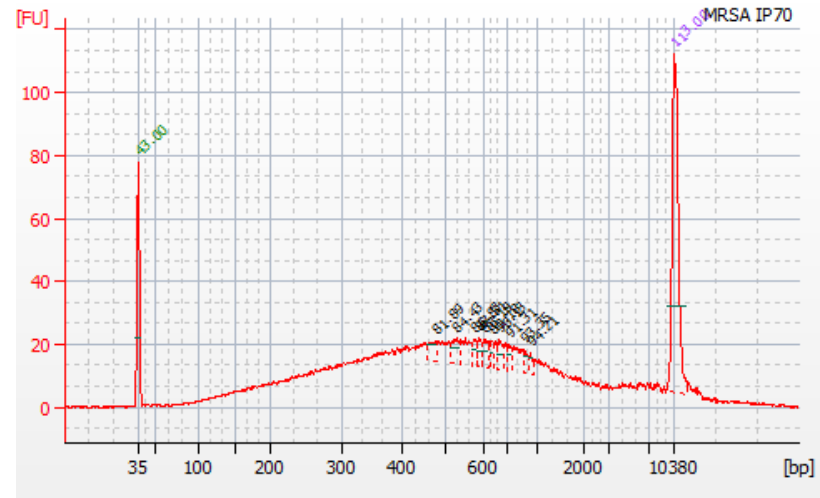
(A)



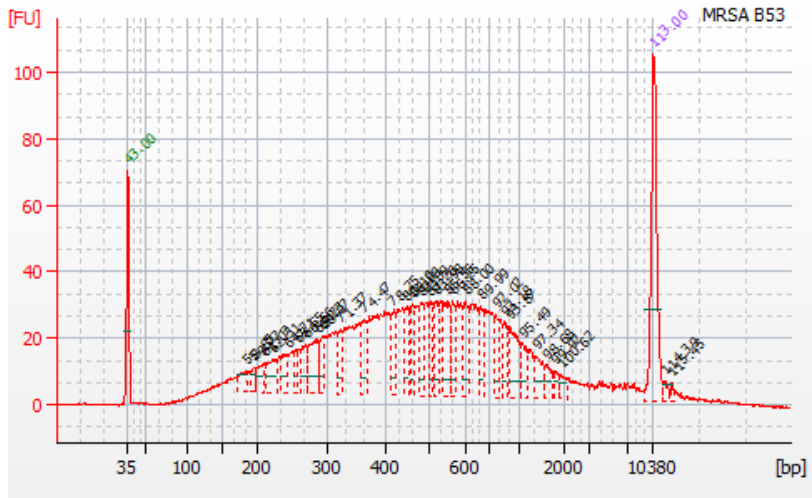
(B)



(C)



(D)



3.4.8.2 Ligation to Adaptors

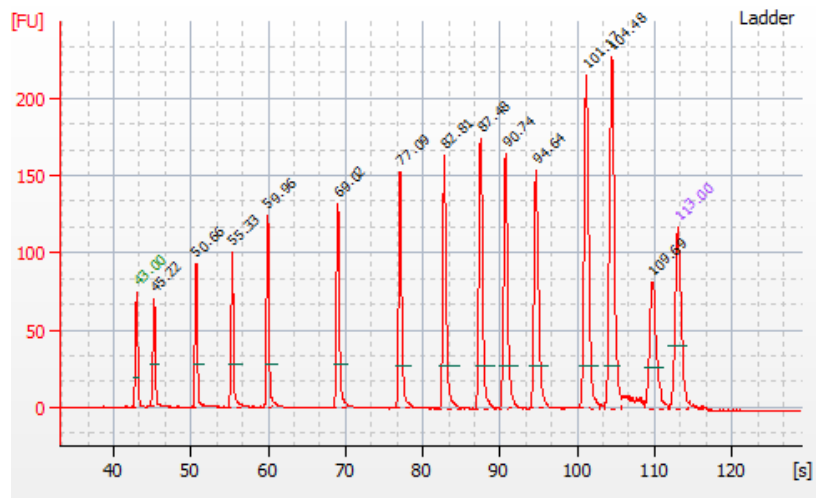
Once the fragmented DNA size was checked; the samples were ligated to their very own specific barcode adaptors as described in section 2.7.1d and as shown in Table 3.3.

The ligated-DNA samples were then added to the E-Gel® SizeSelect™ agarose cassette, and once the samples had migrated to the 400-base pair collection well, it was extracted; as described in section 2.7.1e. The six samples were then amplified and purified as described in 2.7.1f. The six samples were subsequently quantified and pooled in equal-molar concentration in to two pools, as described in section 2.7.1g; the pools are shown in Table 3.3.

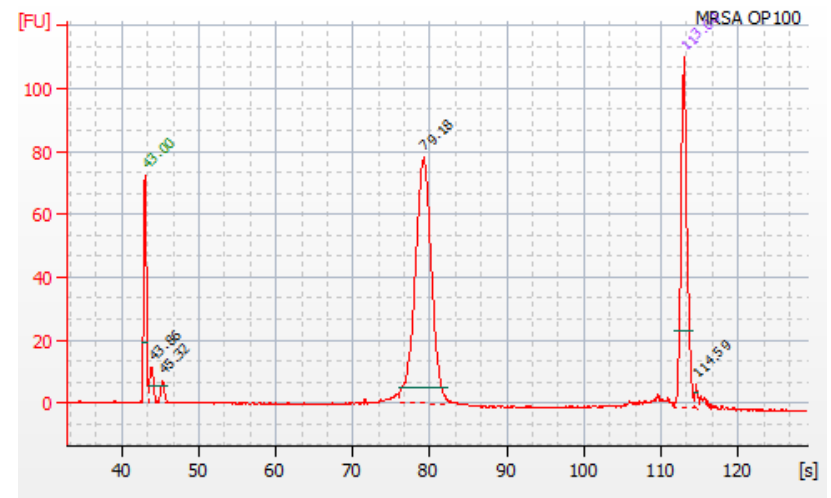
The barcoded library samples were then analysed by the Bioanalyser® instrumentation, as described in section 2.7.1c. The results of this is given in Table 3.3 and Figures 3.8a-g

	OP100	B51	IP70	B53	S62	AS3
Specific Barcode adaptor	33	34	35	36	37	38
Pool Number	1	1	1	2	2	2
Size in Base pairs	437	451	464	489	492	495
Molarity (pmol/l)	988.6	289	438.4	182.5	284.5	212
Aligned migration time (seconds)	79.18	80.01	80.74	82.16	82.37	82.54

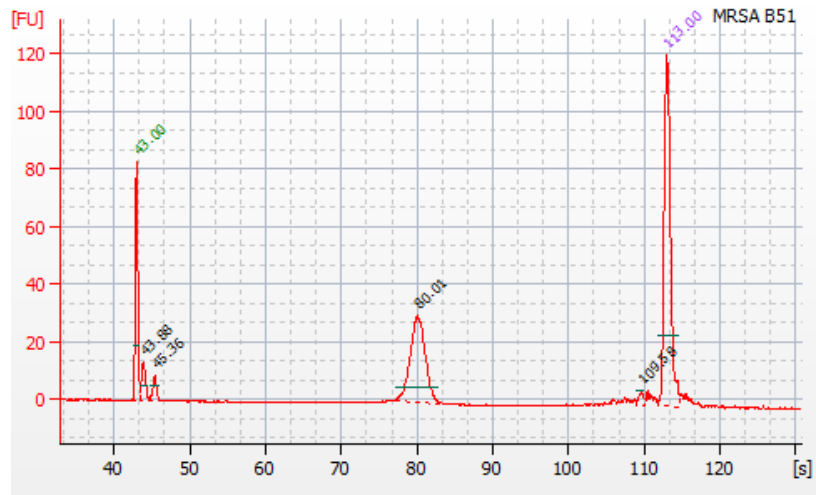
Table 3.3: Showing the assigned barcode adaptors, the pool number and output of the Bioanalyser®; which includes the size of the isolate in base pairs, its molarity and the aligned migration time.



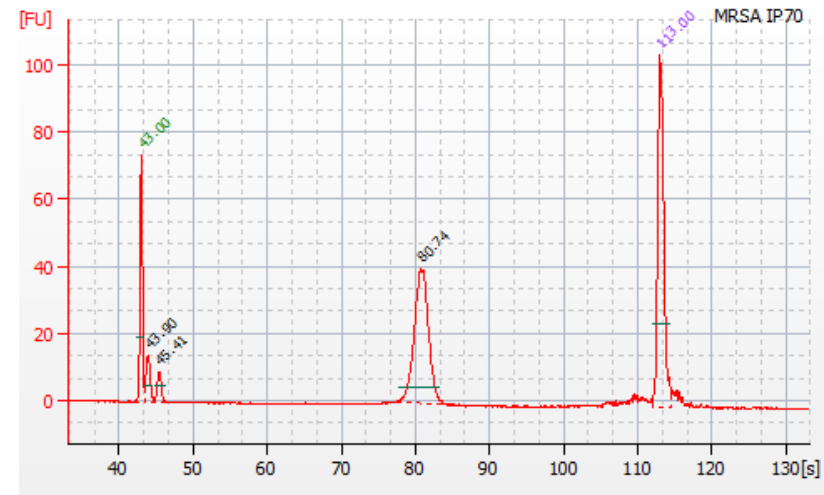
(A)



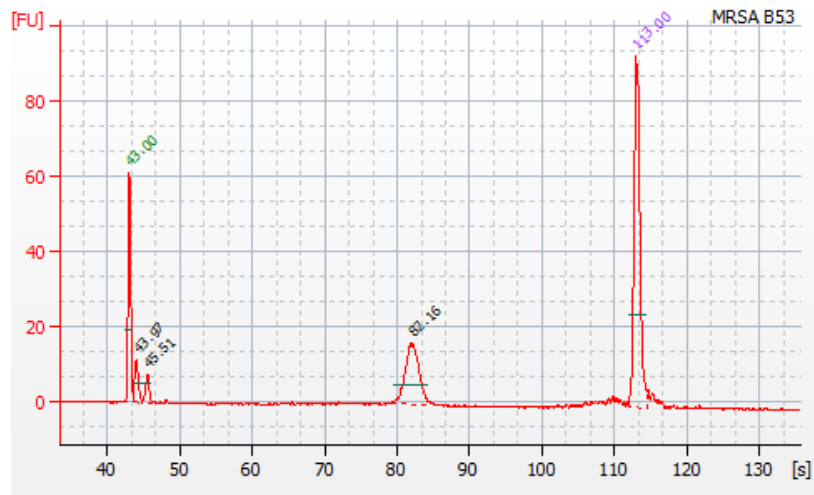
(B)



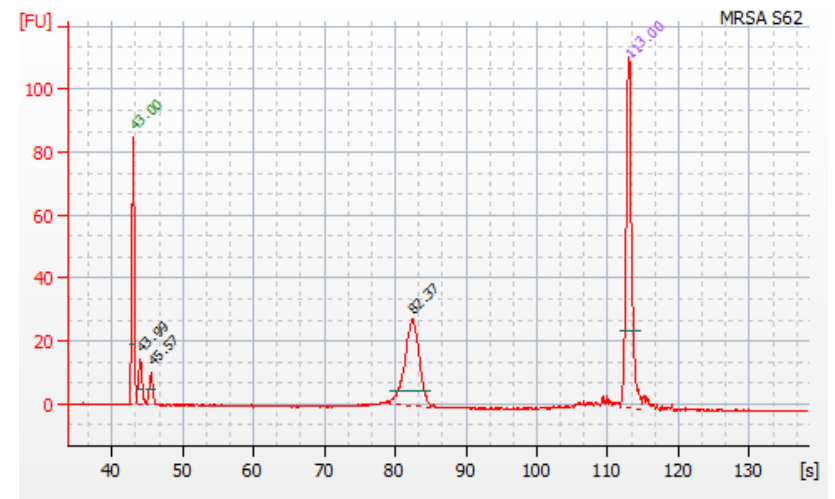
(C)



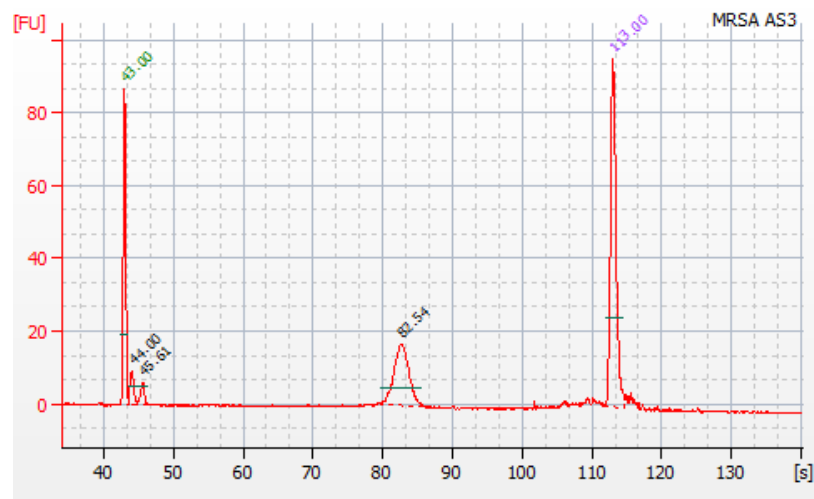
(D)



(E)



(F)



(G)

Figure 3.8a: Electropherogram showing the fluorescence intensity (FU) versus the size/ migration time for the ladder markers.

Figures 3.8b-g: Shows the fluorescence intensity (FU) versus the size/ migration time for the six samples.

3.4.8.3 Chip loading

The 2 pools of MRSA samples were then loaded into the ports of two Ion 316™ v2 Chip as described in section 2.7.3d and a representation of how successful the loading was shown in Figure 3.3 and Table 3.4

It is interesting to note from Figures 3.9a, 3.9b and Table 3.4 that 71% and 72% of wells were loaded in both chips. The samples that were loaded in the wells were 100% enriched with Ion Sphere Particles (ISP). Both pools had a greater percentage of clonal ISPs which indicate that DNA is cloned from a single original template, as opposed to polyclonal, which indicate that the ISPs is carrying clones from two or more templates. For both pools, the number of usable reads was $\geq 50\%$, which is again an indication of a good analysis.

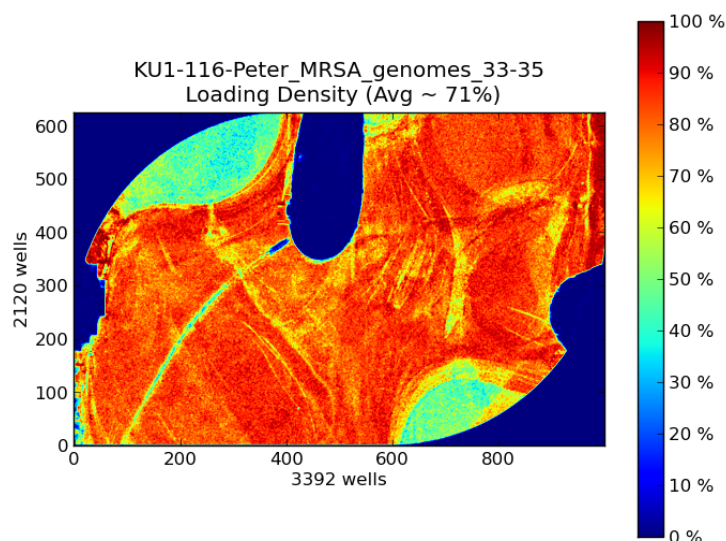


Figure 3.9a: A representation of the loading density for pool two which includes MRSA OP100, B51 & IP70, on to the Ion 316™ v2 Chip. This heat map shows that the analysis has achieved an average loading density of 71%. As shown on the indicator bar; red shows a good loading density and blue indicates a poor loading density on the specific area of the chip.

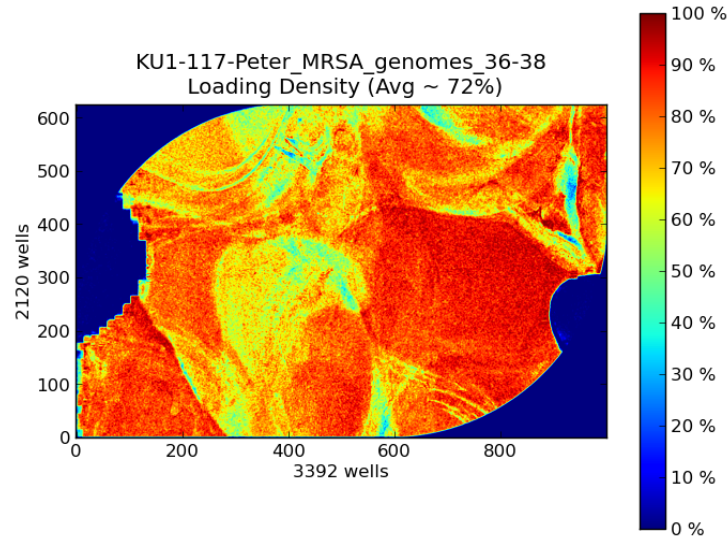


Figure 3.9b: A representation of the loading density for pool two which includes MRSA B53, S62 & AS3, on to the Ion 316™ v2 Chip. This heat map shows that the analysis has achieved an average loading density of 72%. As shown on the indicator bar; red shows a good loading density and blue indicates a poor loading density on the specific area of the chip.

	Pool One	Pool two
Usable Reads	53%	50%
% Loaded wells	71%	72%
% Empty wells	29%	28%
% Enrichment	100%	100%
% Clonal	59%	56%
% Polyclonal	41%	44%

Table 3.4a: Showing the success rate of Pool One and Pool Two chip loading; which is represented by % usable reads, % loaded / empty wells, % enrichment and % clonal/ polyclonal.

3.4.8.4 Ion Torrent summary

Once the Ion Torrent had completed sequencing the six strains, they were made available on the Ion Torrent Sever via the Ion Torrent browser and downloaded, as described in section 2.7.3e. The results from the Ion Torrent Browser analysis report indicated that a good run was performed. The results of the run are shown below in Table 3.4b. The results in the table include the total number of sequence reads, average length of the read, the total number of contigs and the average length of the contigs.

3.4.8.5 Genome Assembly

The CLC Bio Genomics workbench software package was used to assemble the six MRSA isolates using the read mapping function as described in section 2.8. The sequences were mapped to two reference *S. aureus* genomes; NCTC 8325 and MW2. NCTC 8325 was chosen because it is a laboratory reference genome and MRSA MW2 was chosen because it is a highly virulent strain that includes a majority of the virulence factors (n=40) identified in the literature. The results of the MW2 read mapping is shown below in table 3.4b. It was not possible to fully assemble the strains as there was not a sufficient density of cover to allow one to get a fully assembly. The coverage varied between 75.94% to 94.05% of the mapped read. If one Ion Torrent chip was used per sample, then it may have been possible to achieve a greater coverage and potentially assemble each strain, however this was not the case. The table shows the total number of MW2 mapped reads, the % of MW2 mapped reads as well as the average length of the mapped reads.

	OP100	B51	IP70	B53	S62	AS3
Total sequence Reads	841,702	755,056	715,221	739,943	791,261	661,387
Average length of Read	297.03	309.63	321.34	307.3	292.44	305.85
Total No. of Contigs	1,662	80	93	162	164	137
Average length of Contig	3,838	34,580	30,051	17,534	17,415	20,453
No. of MW2 mapped reads	756,556	710,150	641,185	624,788	600,896	584,528
% of MW2 mapped reads	89.88%	94.05%	89.65%	84.44%	75.94%	88.38%
Avg length of mapped reads	298.43	311.02	323.08	312.17	298.69	309.34

Table 3.4b. This table shows the results of the ion torrent run which includes the total number of contigs as well as the average length of the contigs. It also shows the % of mapped reads including the average length of the mapped reads.

3.4.9 Selecting and acquiring the virulence factor gene sequences

A literature review was then undertaken to identify known *S. aureus* virulence factors and acquire their genomic sequences as described in section 2.9.3. Table 3.5 presents a list of virulence factors and their respective gene symbols that were compiled. Virulence factors refer to the properties that will enable *S. aureus* to establish itself in the host and enhance its potential to cause disease. This list includes the major virulence factor genes, i.e. adherence genes, cytotoxin genes, immune evasion genes, enzyme genes, super antigen protein genes and enterotoxin genes, as well as the genes that encode for methicillin resistance. This list includes a wide variety of virulence factors present for many different roles as described in Section 1.5.

Gene	Description
<u>Methicillin resistance gene</u>	
<i>mecA</i>	Encodes for methicillin-resistance
<i>PBP2</i>	Penicillin binding protein-2
<u>Cytotoxin</u>	
<i>hla</i>	alpha-hemolysin gene
<i>hlb</i>	beta-hemolysin phospholipase gene
<i>hlgA</i>	Gamma-hemolysin component A
<i>lukS-PV</i>	Panton-Valentine leukocidin
<i>lukS</i>	leucocidin S component
<i>lukF-PV</i>	Panton-Valentine leukocidin subunit F
<u>Immune Evasion</u>	
<i>chp</i>	chemotaxis inhibitory protein CHIPS
<i>sbi</i>	IgG binding protein
<i>sak</i>	Staphylokinase
<u>Enzymes and other secreted proteins</u>	
<i>sspA</i>	serine protease; v8 protease; glutamyl endopeptidase
<i>sspB</i>	cysteine protease precursor
<i>sspC</i>	cysteine protease
<i>vwb</i>	von Willebrand factor-binding protein
<i>lip</i>	lipase
<i>lip1</i>	tricylglycerol lipase
<i>coa</i>	staphylocoagulase
<i>lipA</i>	lipoic acid synthetase
<i>plc</i>	1-phosphatidylinositol phosphodiesterase/ precursor
<i>nuc</i>	nuclease
<i>nucA</i>	thermonuclease
<i>hysA, hysA1, hysA2</i>	hyaluronate lyase

<i>aur</i>	zinc metallo proteinase aureolysin
<u>Involved in Persistence</u>	
<i>hemB</i>	Delta-aminolevulinic acid dehydratase
<i>sepA</i>	multidrug resistance efflux pump
<u>Adherence</u>	
<i>cna</i>	Collagen adhesion protein
<i>efb</i>	fibrinogen-binding protein
<i>fnb</i>	fibronectin-binding protein
<i>ebpS</i>	elastin-binding protein
<i>bbp</i>	bone sialoprotein-binding protein
<i>eap</i>	extracellular adherence protein
<i>clf</i>	clumping factor
<u>Enterotoxins</u>	
<i>entE</i>	enterotoxin E
<i>sea</i>	enterotoxin A
<i>seb</i>	enterotoxin B
<i>seh</i>	enterotoxin H
<i>sek</i>	enterotoxin K
<i>seg</i>	enterotoxin G
<i>tst</i>	toxic shock syndrome

Table 3.5: The *S. aureus* virulence factors and their respective genes.

3.4.10 Extracting virulence factors from published *S. aureus* strains

The Nucleotide BLAST tool (Altschul *et al.*, 1990) was used to identify if the virulence factors, acquired through literature research and GenBank, were present in the published *S. aureus*. If the corresponding virulence factor gene was present, the gene sequence was extracted in FASTA format and GenBank format, as described in section 2.11. A representation of the presence of virulence factors in Hospital-associated (Table 3.6), Community-associated (Table 3.7) and Livestock-associated (Table 3.8) *S. aureus* strains are given in Tables 3.6 to 3.8. The cut of value to determine if the virulence factor sequence was present in a particular strain was set at 90% query coverage, 95% identity and an E value of <0.01, this data is shown in Appendix 2. BLASTn results were checked for evidence of gene duplication (more than one homologue gene in a strain); however, no evidence of this was observed.

3.4.11 Extracting virulence factors from novel MRSA strains

CLC Genomics Workbench was initially used to create a reference library of the novel strains using the read mapping functionality, as described in Section 2.8.2. This ensured that the novel sequences were annotated. A virulence factor database was then created, as described in Section 2.8.3. The virulence factors could then be located on the newly sequenced consensus sequence by simply searching for it using the name or symbol of the query virulence factor, it was also possible to use the BLASTn tool to locate them, as described in Section 2.8.5. A representation of the presence of virulence factors of the newly sequenced strains is given in Table 3.9 below.

Once all of the virulence factor genes had been extracted from both the published and novel strains, a graph (Figure 3.10) showing the overall presence of virulence factor genes in the *S. aureus* strains, based on the percentage of sequence similarity was created. This graph shows that although the the frequency of virulence factor genes varies amongst the *S. aureus* strains, a majority of genes are shared as well. A clustered column

bar graph (Figure 3.11) representing the frequency of virulence factor genes in hospital-associated (HA), community-associated (CA) and livestock-associated (LA) was also produced. It is again evident from this graph that a majority of the virulence factor genes are shared amongst the HA, CA and LA strains, with a few exceptions. It is noticeable that the enterotoxins are not present in many strains along with the *cna* gene and *lukSpV* gene.

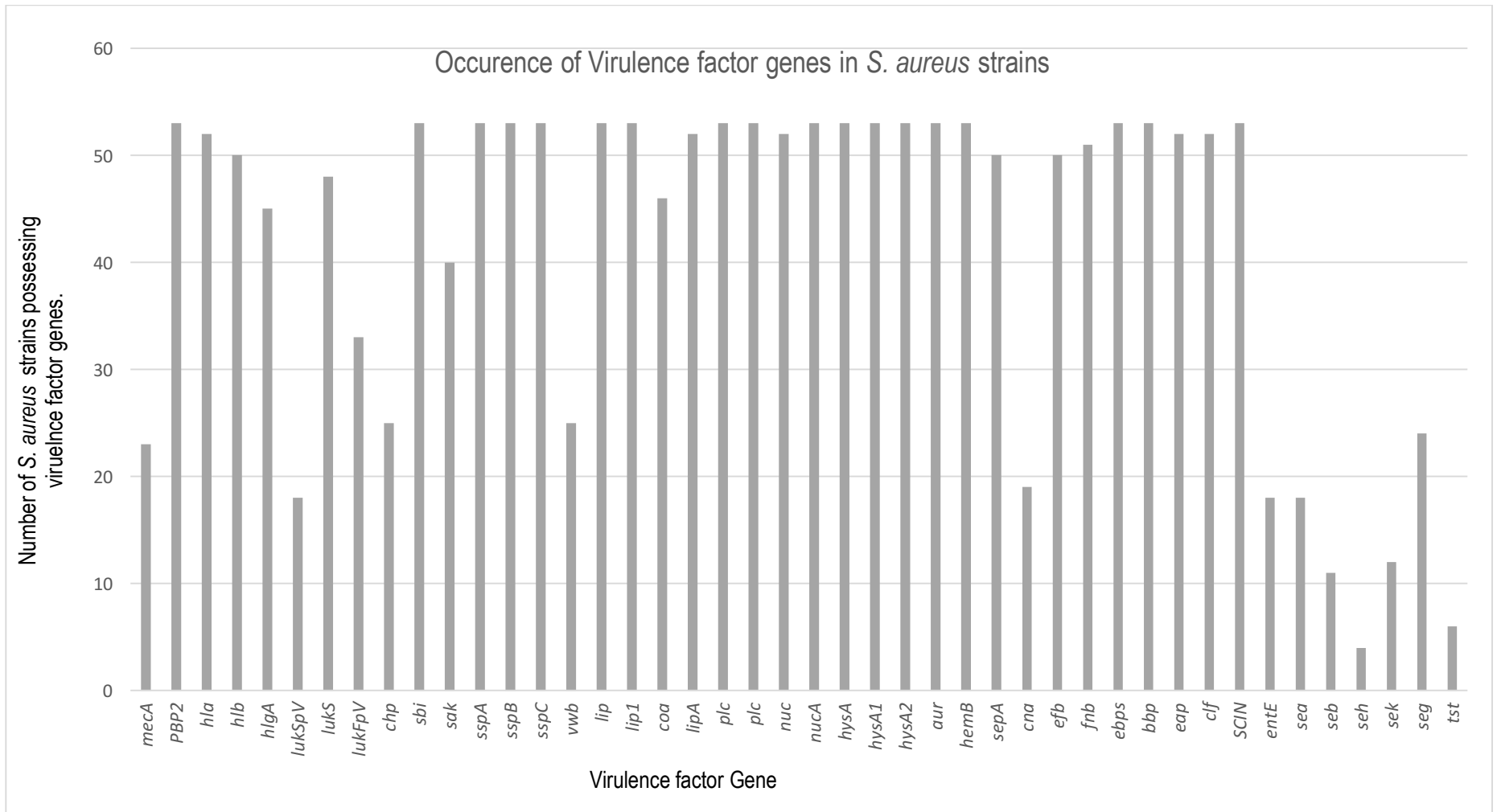


Figure 3.10: A histogram representing the the prevalence of virulence factor genes in the 53 *S. aureus* strains.

Gene	04-02981	502A	6850	Bmb9393	COL	H-EMRSA15	H050960412	JKD6008	JH1	JH9	Mu50	Mu3	MRSA252	Newman	NCTC8325	N315	ST228 18385	ST228 16125	ST228 16035	TO131	TW20	VC40	XN108	Z172
<i>mecA</i>																								
<i>PBP2</i>																								
<i>hla</i>																								
<i>hly</i>																								
<i>hlyA</i>																								
<i>lukSpV</i>																								
<i>lukS</i>																								
<i>lukFpV</i>																								
<i>chp</i>																								
<i>sbi</i>																								
<i>sak</i>																								
<i>sspA</i>																								
<i>sspB</i>																								
<i>sspC</i>																								
<i>vwb</i>																								
<i>lip</i>																								
<i>lip1</i>																								
<i>coa</i>																								
<i>lipA</i>																								
<i>plc</i>																								
<i>plc</i>																								
<i>nuc</i>																								
<i>nucA</i>																								
<i>hysA</i>																								
<i>hysA1</i>																								
<i>hysA2</i>																								
<i>aur</i>																								
<i>hemB</i>																								
<i>sepA</i>																								
<i>cna</i>																								
<i>efb</i>																								
<i>fnb</i>																								
<i>ebps</i>																								
<i>bbp</i>																								
<i>eap</i>																								
<i>clf</i>																								
<i>SCIN</i>																								
<i>entE</i>																								
<i>sea</i>																								
<i>seb</i>																								
<i>seh</i>																								
<i>sek</i>																								
<i>seg</i>																								
<i>tst</i>																								

Table 3.6: A representation of the presence of virulence factor genes present in 24 published hospital-associated strains. The *S. aureus* strains are given across the top of the figure and the genes along the side. The black colour indicates that a gene is present in the *S. aureus* and the white indicates that it is absent.

	11819-07	CN1	CA-347	ECT-R2	JKD6159	M1	MW2	MSSA476	MSHR1132	M013	SA40	SA957	UAS391- USA300	USA300- ISMMS1	USA300- FPR3737	USA300_T CH1516
<i>mecA</i>																
<i>PBP2</i>																
<i>hla</i>																
<i>hlb</i>																
<i>hlgA</i>																
<i>lukSpV</i>																
<i>lukS</i>																
<i>lukFpV</i>																
<i>chp</i>																
<i>sbi</i>																
<i>sak</i>																
<i>sspA</i>																
<i>sspB</i>																
<i>sspC</i>																
<i>vwb</i>																
<i>lip</i>																
<i>lip1</i>																
<i>coa</i>																
<i>lipA</i>																
<i>plc</i>																
<i>plc</i>																
<i>nuc</i>																
<i>nucA</i>																
<i>hysA</i>																
<i>hysA1</i>																
<i>hysA2</i>																
<i>aur</i>																
<i>hemB</i>																
<i>sepA</i>																
<i>cna</i>																
<i>efb</i>																
<i>fnb</i>																
<i>ebps</i>																
<i>bbp</i>																
<i>eap</i>																
<i>clf</i>																
<i>SCIN</i>																
<i>entE</i>																
<i>sea</i>																
<i>seb</i>																
<i>seh</i>																
<i>sek</i>																
<i>seg</i>																
<i>tst</i>																

Table 3.7: A representation of the presence of virulence factor genes present in 16 published community-associated strains. The *S. aureus* strains are given across the top of the figure and the genes along the side. The coloured block indicates that a gene is present in the *S. aureus* and the white indicates that it is not present.

	HA-MRSA	HA-MRSA	HA-MRSA	LA-MRSA	LA-MRSA	LA-MRSA	LA-MRSA
Gene	08BA0216	71193	ED133	ED98	LG1251	RF122	ST398
<i>mecA</i>							
<i>PBP2</i>							
<i>hla</i>							
<i>hlb</i>							
<i>hlgA</i>							
<i>lukSpV</i>							
<i>lukS</i>							
<i>lukFpV</i>							
<i>chp</i>							
<i>sbi</i>							
<i>sak</i>							
<i>sspA</i>							
<i>sspB</i>							
<i>sspC</i>							
<i>vwb</i>							
<i>lip</i>							
<i>lip1</i>							
<i>coa</i>							
<i>lipA</i>							
<i>plc</i>							
<i>plc</i>							
<i>nuc</i>							
<i>nucA</i>							
<i>hysA</i>							
<i>hysA1</i>							
<i>hysA2</i>							
<i>aur</i>							
<i>hemB</i>							
<i>sepA</i>							
<i>cna</i>							
<i>efb</i>							
<i>fnb</i>							
<i>ebps</i>							
<i>bbp</i>							
<i>eap</i>							
<i>clf</i>							
<i>SCIN</i>							
<i>entE</i>							
<i>sea</i>							
<i>seb</i>							
<i>seh</i>							
<i>sek</i>							
<i>seg</i>							
<i>tst</i>							

Table 3.8: A representation of the presence of virulence factor genes present in published livestock-associated *S. aureus* strains. The *S. aureus* strains are given across the top of the figure and the genes along the side. The coloured block indicates that the gene is present in the *S. aureus* strain and the white indicates that it is not.

	MRSA OP100	MRSA B51	MRSA IP70	MRSA B53	MRSA S62	MRSA AS3
<i>mecA</i>						
<i>PBP2</i>						
<i>hla</i>						
<i>hlb</i>						
<i>hlgA</i>						
<i>lukSpV</i>						
<i>lukS</i>						
<i>lukFpV</i>						
<i>chp</i>						
<i>sbi</i>						
<i>sak</i>						
<i>sspA</i>						
<i>sspB</i>						
<i>sspC</i>						
<i>vwb</i>						
<i>lip</i>						
<i>lip1</i>						
<i>coa</i>						
<i>lipA</i>						
<i>plc</i>						
<i>plc</i>						
<i>nuc</i>						
<i>nucA</i>						
<i>hysA</i>						
<i>hysA1</i>						
<i>hysA2</i>						
<i>aur</i>						
<i>hemB</i>						
<i>sepA</i>						
<i>cna</i>						
<i>efb</i>						
<i>fnb</i>						
<i>ebps</i>						
<i>bbp</i>						
<i>eap</i>						
<i>clf</i>						
<i>SCIN</i>						
<i>entE</i>						
<i>sea</i>						
<i>seb</i>						
<i>seh</i>						
<i>sek</i>						
<i>seg</i>						
<i>tst</i>						

Table 3.9: A representation of the presence of virulence factor genes present in the Novel MRSA strains. The MRSA strains are given across the top of the figure and the genes along the side. The coloured block indicates that the gene is present in the *S. aureus* strain and the white indicates that it is not.

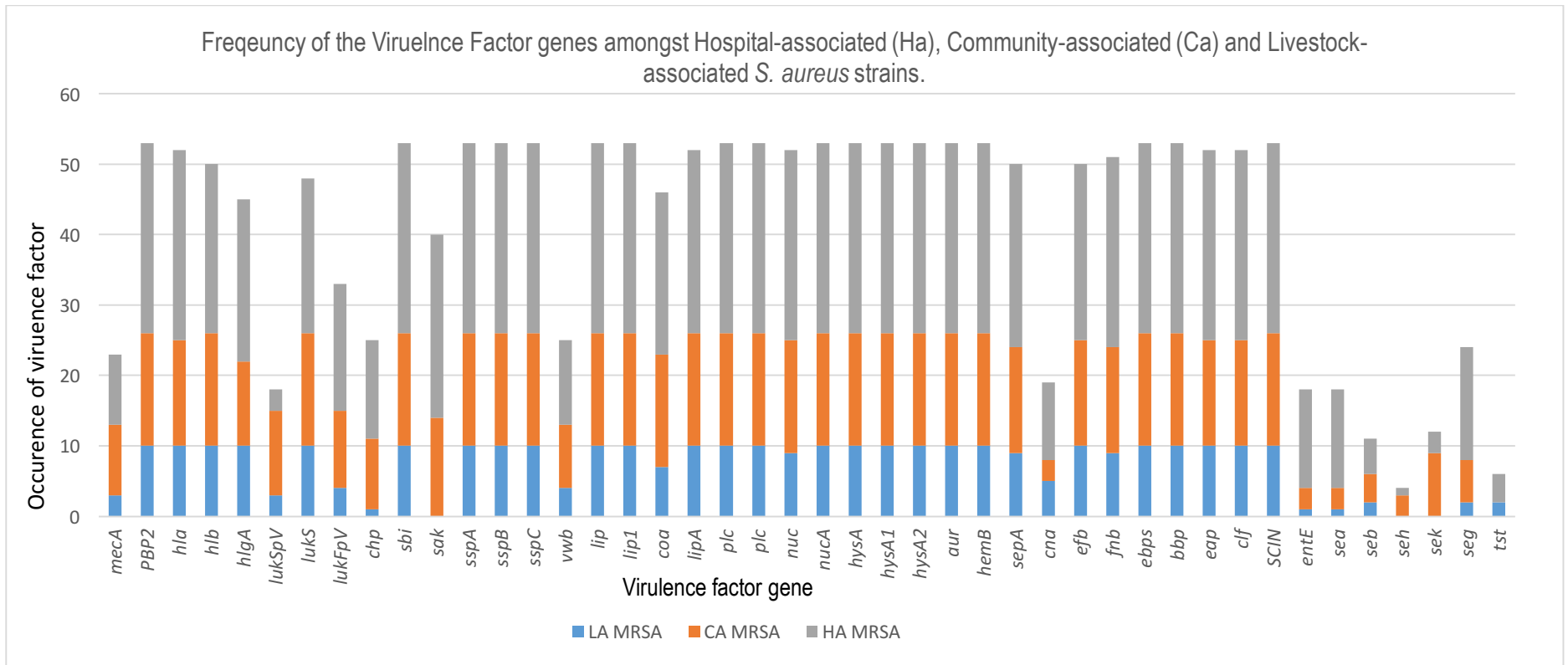


Figure 3.11: A clustered column bar graph representing the frequency of virulence factors amongst the hospital-associated, community-associated and livestock-associated published and novel *S. aureus* strains.

3. 5 Discussion

A number of published complete *S. aureus* genome sequences (n=50) were acquired for this study. *S. aureus* strains were classified as either hospital-associated (HA), community-associated (CA) or live-stock associated (LA), following an extensive literature review as shown in Table 3.1. For instance, *S. aureus* H-EMRSA-15 has been described as a strain that appeared in 1991 and rapidly disseminated across UK hospitals affecting individuals who had been in the healthcare setting (O'Neil *et al.*, 2001). This strain was therefore characterised as hospital-associated (HA). In contrast, *S. aureus* strain USA300_FPR3757 was described as causing sporadic outbreaks of skin and soft tissue infection in otherwise healthy individuals (Diep *et al.*, 2007). This strain has therefore been designated as an epidemic community-associated (CA) *S. aureus* strain. Another isolate; *S. aureus* ST398 has been described as a livestock-associated (LA) because it was originally isolated from swine in 2004. Currently, this strain has been reported as the causative agent causing infection in humans working with intensively farmed animals (Schijffelen *et al.*, 2010). Forty-seven *S. aureus* strains were successfully classified as either HA (n=24), CA (n=16) or LA (n=7). Three strains could not be classified due to the lack of information present in the literature. *Staphylococcus aureus* strains 55/2053, NRS100 & TCH60 were therefore omitted from further investigation. The methicillin-sensitive *S. aureus* (MSSA) strains were also included (MSSA476 & 71193) as a comparison to the MRSA strains, as they are effectively progenitor strains, for example 71193 was the precursor to ST398 (MRSA) (Bhat *et al.*, 2009; Uhlemann *et al.*, 2012). The MAUVE analysis as well as the BLASTn analysis revealed that there were very minor differences between 71193 and ST398, with 71193 harbouring only two additional virulence factor genes (*chp* & *vwb*) than ST398.

Initial genome comparative analysis using MAUVE was carried out. Figure 3.5, an output comparing the three classifications was produced. Initially three classifications were expected to be different with more unique regions and segments, however, it was interesting as the strains were similar. The homologous regions represented by the coloured blocks were shared throughout the 3 sequences, with minor differences occurring within the blocks

indicating regions that are unique to the strain. More in-depth analysis using MAUVE observed that nucleotide differences did exist amongst the virulence factor genes but further analysis is required to test what the changes may cause phenotypically.

A further six novel MRSA strains were also investigated. These strains were also classified as either HA (n=3) or LA (n=3) according to the information that was provided by the collaborators. The presence of virulence factors from the published and novel strains were then examined further. It is observed that there is a minor difference in the presence of virulence factor genes amongst the classified strains, evidenced by Figure 3.11.

The presence of these genes amongst the strains was further analysed to try and suggest why the genes are present in some strains but not others. The first step in *S. aureus* colonisation is mediated by the products of a number of adhesion genes. One such gene, the *cna* gene, encodes the collagen-binding adhesion protein which targets collagen (Foster & Hook, 1998). Collagen-binding adhesion protein binds to collagen with a high degree of specificity (Rhem *et al.*, 2000). It is therefore reasonable to suggest that the presence of the *cna* gene in livestock-associated (LA) *S. aureus* may contribute to the fact that it acts as a virulence factor, enhancing the bacteria's chance of adhesion to the hosts cells and tissue. Although the sample size is small, this study has demonstrated that the *cna* gene is present in two LA strains that cause outbreaks in humans, however it is absent from the strains that are restricted to animals (ED133, ED98 & RF122), this characteristic is shown in Table 3.8. LA strain ST398, originally designated 71193 was isolated from healthy pig farmers and their pigs (Armand-Lefevre *et al.*, 2005). At the time, Armand-Lefevre *et al.*, (2005) strongly suggested that *S. aureus* colonisation occurred in the nostrils of the pig farmers and that their daily interaction with the pigs most likely increased their chances of being colonised greatly with this bacterium. Whereas LA strain 08BA0216 was isolated from a human postoperative surgical site (Golding *et al.*, 2012). It is suggested that that the adhesion protein was not absolutely necessary to cause an infection in this instance due to the presence of an open wound, nevertheless this strain still possessed the relevant gene to adhere to the collagen, abundantly present in the connective tissue of the skin. It is further noted that LA *S. aureus* which do not contain the *cna* gene have not effected humans. Strains

ED133 and ED98 have been implicated in an outbreak of bovine mastitis and a chicken with bacterial chondronecrosis, respectively (Lowder *et al.*, 2009; Guinane *et al.*, 2010). It is therefore suggested that these strains do not possess the ability to bind to collagen in humans or animals. The exception to this trend was strain LA LGA251 which was isolated from a milk sample (Garcia-Alvarez *et al.*, 2011; Paterson *et al.*, 2014). It is suggested however, that since this strain was isolated from a bovine bulk tank milk sample and not directly from the cattle, it only serves as a contaminant of the milk. This suggests that this strain may still have the potential to interact with human collagen, this remains to be seen. When the novel LA strains sequenced in the study were examined, only LA-MRSA S62 carried the *cna* gene, it was not carried by either LA-MRSA B53 or LA-MRSA AS3; however, it is not known if these particular strains were isolated exclusively from animals, produce contamination or animal-to-human interaction. A conclusion cannot therefore be made due to lack of information. Typically, the LA-MRSA strains examined did carry a high proportion of the adherence genes that were being investigated: *efb*, *ebpS*, *bbp*, *eap* & *clf*. The exception was the *fnb* gene, which encodes the fibronectin-binding protein, was absent from one of the sequenced strains (MRSA AS3); however, the *eap* gene and its subsequent extracellular adherence protein has the ability to bind to numerous plasma protein which includes fibronectin. A majority of HA-MRSA (58%) did not carry the *cna* gene. It was present in strain MRSA252, an important epidemic methicillin-resistant *S. aureus* EMRSA-16 clone. However, all the other adherence genes (*efb*, *ebpS*, *bbp*, *eap* & *clf*) are present in the HA-MRSA strains; except EMRSA and HO 50960412 that lack *efb*. It is not necessary for all of the adherence genes to be present because health-care institutions have other risk factors that can contribute to MRSA colonisation in humans and not just adherence. For instance, Strain Mu50 was isolated from a patient who underwent heart surgery. Hiramatsu *et al.* (1997) suggested that the full complement of adhesion related genes to colonise the host are not required. The majority of CA-MRSA (81%) also do not possess the *cna* gene; again this is not the expected trend, however with the exception of the MSHR1132 strain; they do exhibit all of the other adherence related genes.

Staphylokinase, encoded by the *sak* gene, promotes the establishment of *S. aureus* skin infections and increases the bacterial penetration through skin barriers (Kwienski *et al.*, 2013). Once penetration has occurred through the skin barriers, staphylokinase activates plasminogen molecules to plasmin; an enzyme that degrades fibrin, a protein of the extra-cellular matrix (ECM) (Kwienski *et al.*, 2013). This invasion of the host tissue results in the spread of the bacterium and increased infection. It is interesting to note that the majority of the highly virulent community-associated (CA) (88%) and the established hospital-associated (HA) (96%) including the newly sequenced HA *S. aureus* (100%) possess the *sak* gene. In contrast to this, all of the livestock-associated strains, including the newly sequenced LA strains lack the *sak* gene and therefore lack staphylokinase production. This study confirms that *S. aureus* from veterinary sources lack staphylokinase production and suggests that it may have an impact on the virulence of LA MRSA by promoting the establishment of *S. aureus* and causing increased bacterial penetration through the skin barriers (Kwienski *et al.*, 2013).

The chemotaxis inhibitory protein virulence factor, which is encoded by the *chp* gene, specifically impairs the response of neutrophils and monocytes by reducing neutrophil recruitment towards complement activation (C5a) (de Haas *et al.*, 2004). This gene is not present in the LA strains, with the exception of 71193. This is not the expected trend because MRSA ST398 was initially designated 71193 so the same genotypic profile was expected. The *chp* gene is also present in 62.5% of the HA strains and 50% of the CA strains including the highly virulent and endemic USA300 strains. This profile is an indication that there is a minor difference in the LA strains which differ from both HA and CA strains and suggests that the possession of different virulence genes may possibly alter pathogenicity.

The presence of the Pantone-Valentine Leucocidin (PVL) in MRSA is associated with an increased virulence in humans and is typically found in CA (Sola *et al.*, 2012) (1.5.6.3, Pg 1.8.2). Functional PVL consists of two components encoded by the *lukS-PV* and *lukF-PV* genes, thus both genes are needed to establish PVL

as a virulence factor (Rahman *et al.*, 1991). PVL gene *lukS-PV* is absent from 100% of the HA strains and the *lukF-PV* is present in 62.5% of the HA strains. Since both genes are needed to encode for the PVL toxin, this study confirms that none of the HA-MRSA strains have ability to produce functional PVL. The *lukS-PV* is present in 57% of the CA strains and *lukF-PV* is present in 69% of the CA strains. This study also confirms that the endemic hyper virulent USA300 strains (UA-S391_USA300, USA300-ISMMS1, USA300_FPR3737 and USA300_TCH1516) contain both genes and therefore have the potential to produce the PVL virulence factor, more examination is required to make a definitive conclusion on the production of this virulence factor. Studies by Blanco *et al.* (2011) concluded that the USA300 clones are predominant among the PVL-positive clones in Spain and throughout Europe and are normally implicated in severe skin and soft tissue infections (SSTIs). In the literature, for an isolate to be typically more pathogenic and classified as a community-associated strain, the presence of the PVL toxin is necessary with PVL positive genes being used as a marker (Bhatta *et al.*, 2016). This study suggests that this association should no longer be applied to identify community-associated strains. The data also indicates that 100% of the published LA-MRSA strains encode the *lukS-PV* gene and only 1 strain (ED98) encodes the *lukF-PV* gene; again this would suggest that PVL is not functional in 6 out of the 7 strain. Surprisingly, the newly sequenced HA-MRSA and LA-MRSA strains harbour both of these genes and therefore the presence of PVL gene. This study concludes that livestock-associated strains possess the genes necessary to produce PVL but more work is required to demonstrate if it a functional gene. The conclusions drawn here support the findings by Peeters *et al.* (2015). Studies by Peeters *et al.* (2015) concluded that ST80 strains carried the Panton-Valentine leukocidin gene.

During infection, *S. aureus* can secrete two coagulases; Staphylococcal coagulase (*coa* gene) and von Willebrand factor binding protein (vWbp) (*vwb* gene), both promote coagulation by activating prothrombin, which in turn converts fibrinogen into fibrin and promotes clotting of the plasma or blood and enables the establishment of *S. aureus* disease (McAdow *et al.*, 2012). The *coa* gene is present in all HA-MRSA strains, excluding MRSA

6850 which does however encode the *vwb* gene. The *vwb* gene also is present in a further 42% of all HA-MRSA strains. The *coa* gene is encoded by 100% of the CA-MRSA strains. A further 56% of the CA-MRSA also encode the *vwb* gene, including the highly virulent USA300 strains. The *coa* gene is encoded by 100% of the LA-MRSA strains, however, only the 71193 strain encodes the *vwb* gene. The novel sequenced strains do not encode the *coa* gene, however the trend is reverse, because excluding MRSA IP70, they all encode the *vwb* gene. These trends suggest that the presence of both the *coa* and *vwb* genes in the bacteria will determine a more virulent strain. This statement is supported by the fact that the highly virulent USA300 strains do encode both genes.

The *mecA* gene allows *S. aureus* to be resistant to β -lactam antibiotics by encoding the protein penicillin binding protein 2a (PBP2a). PBP2a has a low affinity for β -lactams such as methicillin, therefore allows the transpeptidation reactions to proceed, allowing cell wall synthesis at concentrations which would otherwise have been inhibited (Lim & Strynadka, 2002; Rice, 2012). It is interesting to note that all strains that were investigated in this study have the presence of PBP2, this ensures that these strains are methicillin-resistant. Surprisingly, not all strains show the presence of the *mecA* gene; only 42% of HA-MRSA, 63% of CA-MRSA, 0% of LA-MRSA, 0% newly sequenced HA-MRSA and 100% of newly sequenced LA-MRSA encode for the *mecA* gene. This was not the expected trend but can be explained by the fact that in this study only one *mecA* gene was investigated. Studies have suggested a divergent *mecA* gene in livestock-associated LGA251 has been observed which is unique because it is phenotypically resistant to methicillin, however it has tested negative for the *mecA* gene (Garcia-Alvarez *et al.*, 2011); this divergent *mecA*_{LGA251} gene was not investigated in this current study, due to lack of time. In this study BLAST was used to help identify if the virulence factor genes were present in some or all of the *S. aureus* strains. The strains containing the virulence factor genes were only identified if they were highly similar. Various divergent *mecA* genes such as *mecC* were not recognised by the BLAST search, because it will only determine sequences that are similar. The cut-off value to determine if the virulence factor sequence was present in a particular strain was set at 90% query coverage, 95% identity and an E value of <0.05. *mecA*_{LGA251} has been recognised to be 70% identical to the *S. aureus* *mecA* gene (Garcia-Alvarez *et al.*, 2011), this did not

therefore show up in the BLAST search. Further gene analysis is needed to confirm the presence of the *mecA* gene in all of these strains.

It has been shown that the presence/ absence of virulence factor genes can help predict if an organism will be pathogenic. However, it has also been demonstrated that the majority of these genes are present in the three classifications that have been given in literature (HA, CA & LA), with a minor difference within livestock-associated strains. It is therefore reasonable to suggest that classifications that have been given in literature may possibly be artificial and not universally applicable and that at a molecular level the strains are in fact very similar. This conclusion coincides with the findings by Bal *et al.* (2016) who published their paper a month before this thesis was submitted. Bal *et al.* (2016) concluded that the traditional classification of MRSA into hospital-associated and community-associated was no longer relevant and that genomic studies had led to a better epidemiological understanding of this bacterium with the authors describing the 'blurring' of these classifications (Bal *et al.*, 2016).

Chapter 4:

Phylogenetic analysis of *S. aureus* strains

4.1 Introduction

Phylogenetics is the science of classification of living organisms. A phylogenetic analysis attempts to group organisms, or gene sequences based on similarity and is dependent on the accurate estimates of evolutionary relationships (Gaucher *et al.*, 2010). The phylogenetic trees that are produced are the inferred graphical representation of the evolutionary relatedness among the species.

A variety of models are used to reconstruct phylogenetic trees. The neighbour-joining tree (NJ) is constructed by calculating the genetic distance measured by comparing the nucleotide sequence; it is therefore a distance based method. NJ is the most widely used method for building phylogenetic trees, with the original research paper by Saitou & Nei being cited about 13,000 times since 1987 (Gascuel & Steel, 2006). The maximum-parsimony tree (MP) is created by using the model that minimises the number of changes required to generate the observed variation in the sequences and chooses the tree that require the fewest mutations to explain the data. Finally, the maximum-likelihood tree (ML) uses the method that first incorporates the Bayesian statistics to determine a model, a tree is then created which gives the highest likelihood of the observed data based on the selected model of evolution (Saitou & Nei, 1987; Yang & Rannala, 2012).

In this study all three methods of phylogenetic tree reconstruction have been used to examine and analyse the phylogenetic relationship between the nucleotide sequences of the *S. aureus* virulence factor genes. This will help determine the relationship between the strains to see how they cluster together and infer how related they are.

4.2 Aims

The aims of this present study were to:

- Collate the virulence factor sequences that were extracted from the known (n=47) and newly assembled (n=6) *S. aureus* genomes into one FASTA file.
- Use MUSCLE, a multiple sequence alignment software to align the collated virulence factor sequences. Repeating for all virulence factors (n=43) that were extracted (Edgar., 2004).
- Use Phylogeny.fr for initial phylogenetic analysis and produce evolutionary trees that show similarities and differences in the genetic characteristics of the nucleotide sequences. The ‘One Click mode’ pipeline was used to construct the trees within the phylogeny.fr program. No parameters were changed and the default settings were utilised within the software. This included the use of MUSCLE 3.7 to align the sequences, Gblocks 0.91b for alignment refinement by removing poorly aligned positions and divergent regions, PhyML 3.0 to undertake the phylogeny with the HKY85 substitution model which is suitable for calculating DNA substitutions such as transition and transversions and finally the tree rendering was undertaken using TreeDyn 198.3 (Dereeper *et al.*, 2008).
- Use MEGA, a molecular evolutionary genetics analysis software to construct Neighbour-Joining Trees (NJ), Maximum-Parsimony Trees (MP) and Maximum-Likelihood Trees (ML) to further compare the virulence factors of interest to determine if the three classifications of *S. aureus* (HA, CA & LA) are evident (Kumar *et al.*, 1994).

4.3 Materials and Methods

- The virulence factor sequences that were extracted from the known and newly assembled *S. aureus* genomes, as carried out in Sections 2.9 and 2.11, respectively, were collated in to a single FASTA file.
- The method for aligning the virulence factor sequences using MUSCLE, as carried out in Section 2.13.
- The method for initial phylogenetic analysis using Phylogeny.fr as carried out in Section 2.14.
- Use MEGA to construct construct Neighbour-Joining Trees, Maximum-Parsimony Trees and Maximum-Likelihood Trees as carried out in Sections 2.12.1, 2.12.2, 2.12.3, 2.12.4.

4.4 Results

The *S. aureus* virulence factors (n=43) that were compiled during the literature research and their respective genes are shown in Table 3.5. The nucleotide sequences of these virulence factors were extracted, if present, from the *S. aureus* genomes. Nucleotide BLAST was used to identify and extract the query virulence factor from the known *S. aureus* strains, as described in Section 2.11. CLC Bio was then used to identify and extract the query virulence factor from the novel *S. aureus* strains, as described in Section 2.8. A representation of the presence of the query virulence factors for Hospital-associated, Community-associated, Livestock-associated and the newly sequenced strains are given in Table 3.6, Table 3.7, Table 3.8 and Table 3.9, respectively.

The extracted virulence factor gene sequences were then collated into a single FASTA file for each virulence factor. This file consists of a single-line description (the sequence name), with the first character of the description line preceded by a greater-than (>) symbol, followed by a text-based format for representing the nucleotide sequence base pairs. As an example, the *mecA* gene was present in n=10 Hospital-associated strains, n=10 Community-associated strains, n=7 Livestock-associated strains and n=3 newly sequenced strains. These were all collated into a single FASTA file so that comparative analysis could be done on this virulence factor gene. This was then carried out for all virulence factors (n=43).

4.3.1 Aligning the *S. aureus* Sequences

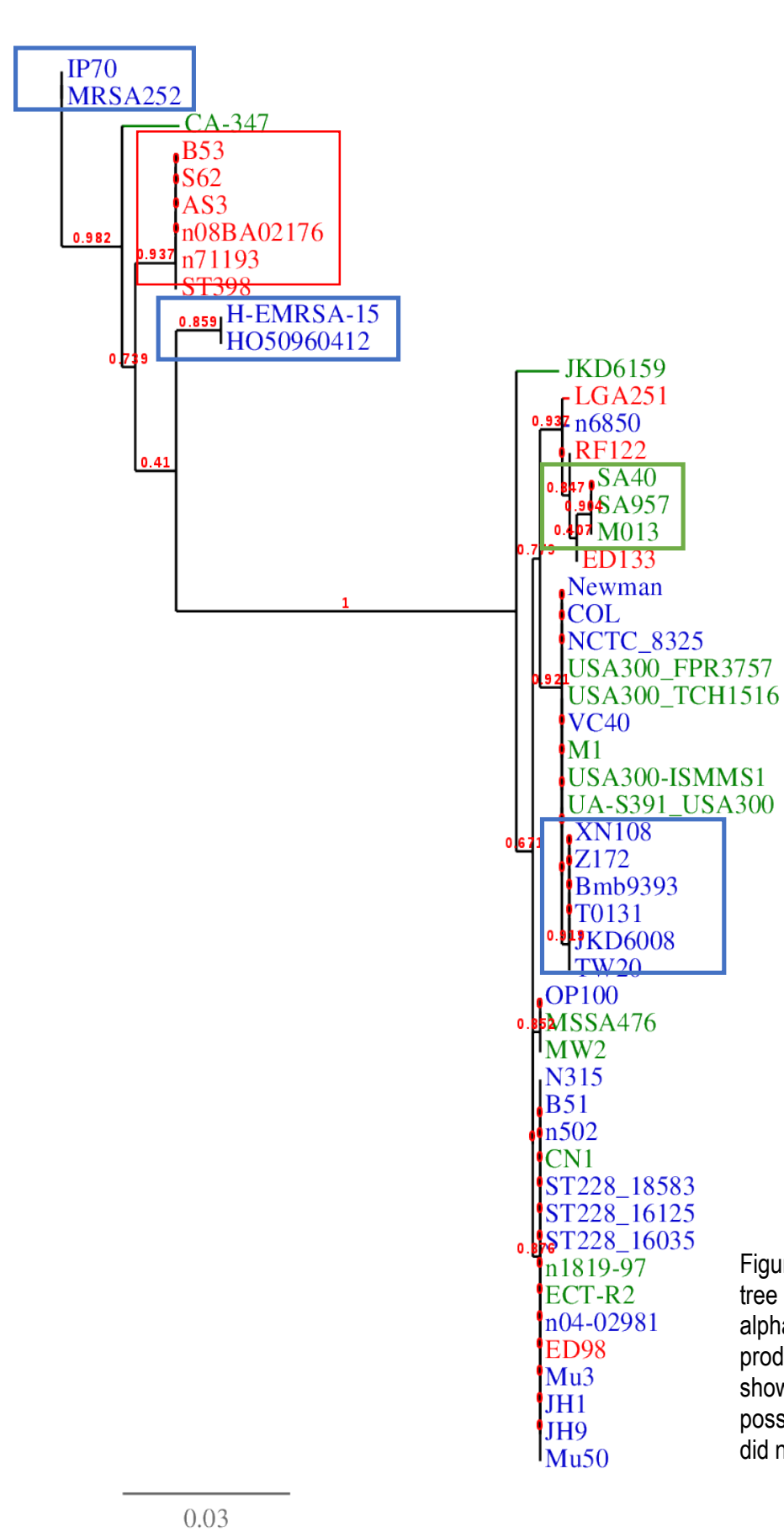
The *S. aureus* sequences were subsequently aligned using MUSCLE, as described in Section 2.13. The output file was in FASTA format. Multiple alignment of sequences is an important step in many applications that will be carried out in this study; including phylogenetic tree estimations and the analysis of DNA polymorphisms (using DnaSP). Both are important bioinformatics applications that are used to understand the evolutionary process shaping genomic variation.

4.3.2 Phylogeny.fr

Phylogeny.fr was then initially used to create phylogenetic trees, as described in Section 2.14. Phylogenetic trees were created for all of the virulence factor genes, they are given in Appendix 2. An example of one phylogenetic tree produced using phylogeny.fr is shown in Figure 4.1a.

The phylogenetic tree is for the *hla* gene which encodes alpha-haemolysin. The virulence factors were chosen as it contains *S. aureus* strains from hospital-associated (HA), community-associated (CA), livestock-associated (LA) as well as the newly sequenced strains.

A phylogenetic analysis shown in Figure 4.1a was created using Phylogeny.fr and shows the relationship between alpha-hemolysin virulence factor gene that was extracted from the various *S. aureus* strains. The 'red' box is showing a number of livestock-associated strains that have formed a single clade. The 3 'blue' boxes are highlighting hospital-associated strains that have clustered together to form 3 clades and the 'green' is highlighting the community-associated strains that have formed 1 clade. It is worth noting that not all LA strains fall in the same clade, the same is true of the HA and LA strains.



Key:
 Blue: HA-MRSA
 Red: LA-MRSA
 Green: CA-MRSA

Figure 4.1a: A phylogenetic tree for the virulence factor alpha-hemolysin (*hla*) produced using phylogeny.fr showing 52 out of a total of 53 possible strains (MSHR1132 did not harbour the *hla* gene..

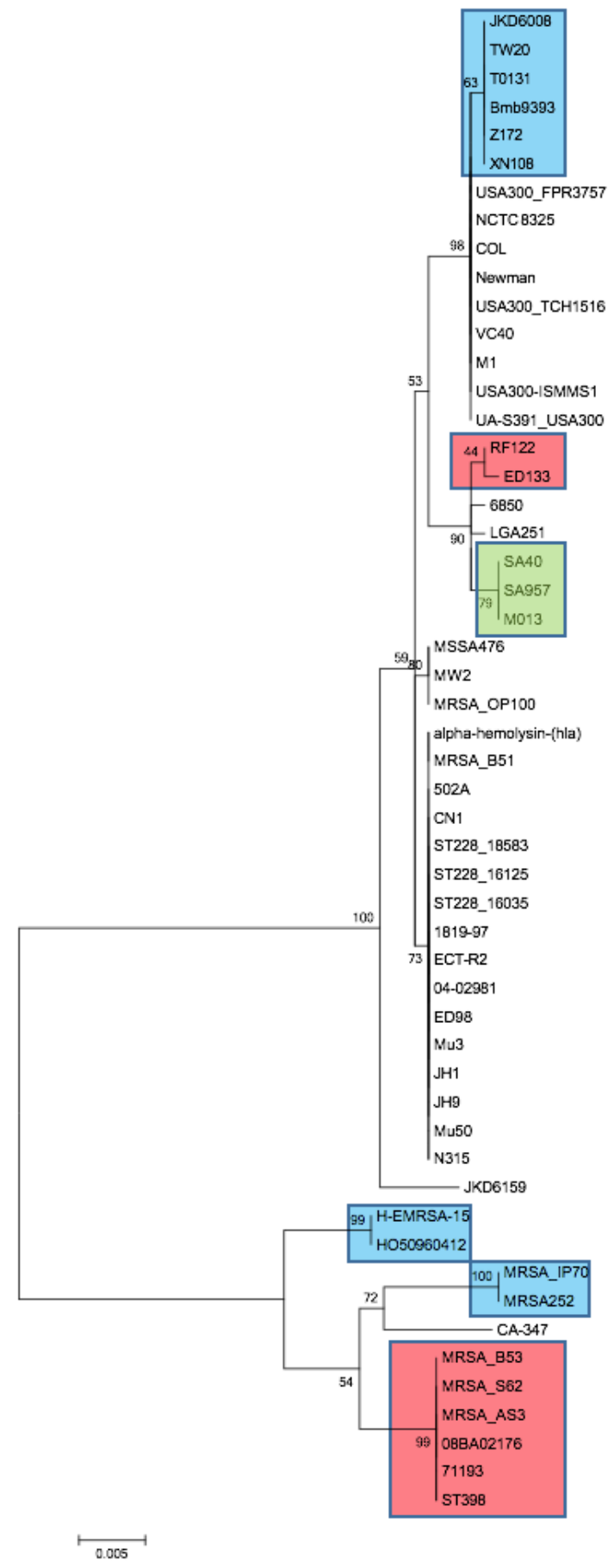


Figure 4.1b: A maximum-likelihood tree for the virulence factor alpha-hemolysin (*hla*) produced using MEGA.

4.3.3 MEGA analysis

A more detailed phylogenetic analysis was then undertaken using MEGA, as described in section 2.12. Phylogenetic trees were created using the neighbour-joining method, maximum-parsimony method and maximum-likelihood method. In total 132 phylogenetic trees were created, however only six will be presented. The phylogenetic trees that will be presented are maximum-likelihood trees that are representative of the gene family classification.

Figure 4.1b is a maximum-likelihood tree created using MEGA and shows the alpha-hemolysin virulence factor. A comparison between the two trees (Figure 4.1a and 4.1b) indicate that many of the clusters are similar, however not identical. For instance, the hospital-associated strains (HA) H-EMRSA-15 & H050960412 cluster together as do IP70 & MRSA252 and finally a larger clade consisting of XN108, Z172, Bmb9393, T0131, JKD6008 & TW20. The same is true of the community-associated (CA) clade (SA40, SA957 & M013) indicated by the green box. However, the notable difference can be observed in the livestock-associated (LA) strains that have been clustered together (red box). In both figures one clade that consists of LA strains (B53, S62, AS3, O8B02176, ST398 & 71193) have clustered together. The other LA strains (LGA251, ED133, RF122 & ED98) that have clustered in Figure 4.1b but have not in Figure 4.1a with the exception of strain ED98. These differences concluded that MEGA maximum-likelihood trees were chosen over the other methods and program due to the nature of modelling. DNA sequences require a model to understand the largely stochastic process. Phylogeny.fr uses the HGY85 model, which can not be changed, whereas MEGA utilises a model that best fits the data; for alpha-hemolysin the Tamura 3-parameter model was used and each virulence factor gene had it own explicit model of evolution to fit to the specific dataset.

Figure 4.2 is a bar graph representing the nodal support values that were achieved from a maximum-likelihood tree for all of the virulence factor genes. These nodal support values are the bootstrap support

values (BP) for the splits or branches in the phylogenetic tree. The error bars represent the standard deviation of the nodal values and gives a general idea of how precise a measurement is from the true value. Phylogenetic bootstrapping is a technique for inferring confidence values on phylogenetic trees that is based on reconstructing many tree from minor variations of the input data, called replicates (Pattengale *et al*, 2010). High bootstrap support values for a branch is highly suggestive that the split actually occurs in the tree, conversely a low bootstrap support values suggests that this might not be true split. It can also suggest that recent expansion has occurred in the organism's gene. A study by Pattengale *et al.* (2010) concluded that bootstrap replicates of 100 – 500 replicates is sufficient as it produces support values that correlate at better than 95%. For the ML analysis in this study the bootstrap was set to 500 replications, as shows in Section 2.12.3.

Figure 4.2 shows the mean nodal support values calculated from the maximum-likelihood trees of the *S. aureus* virulence factor genes. The majority of the ML trees nodal values are exceeding >50. This high nodal values therefore suggests that the split occurring in the trees have actually occurred. A number of genes are also showing low nodal values.

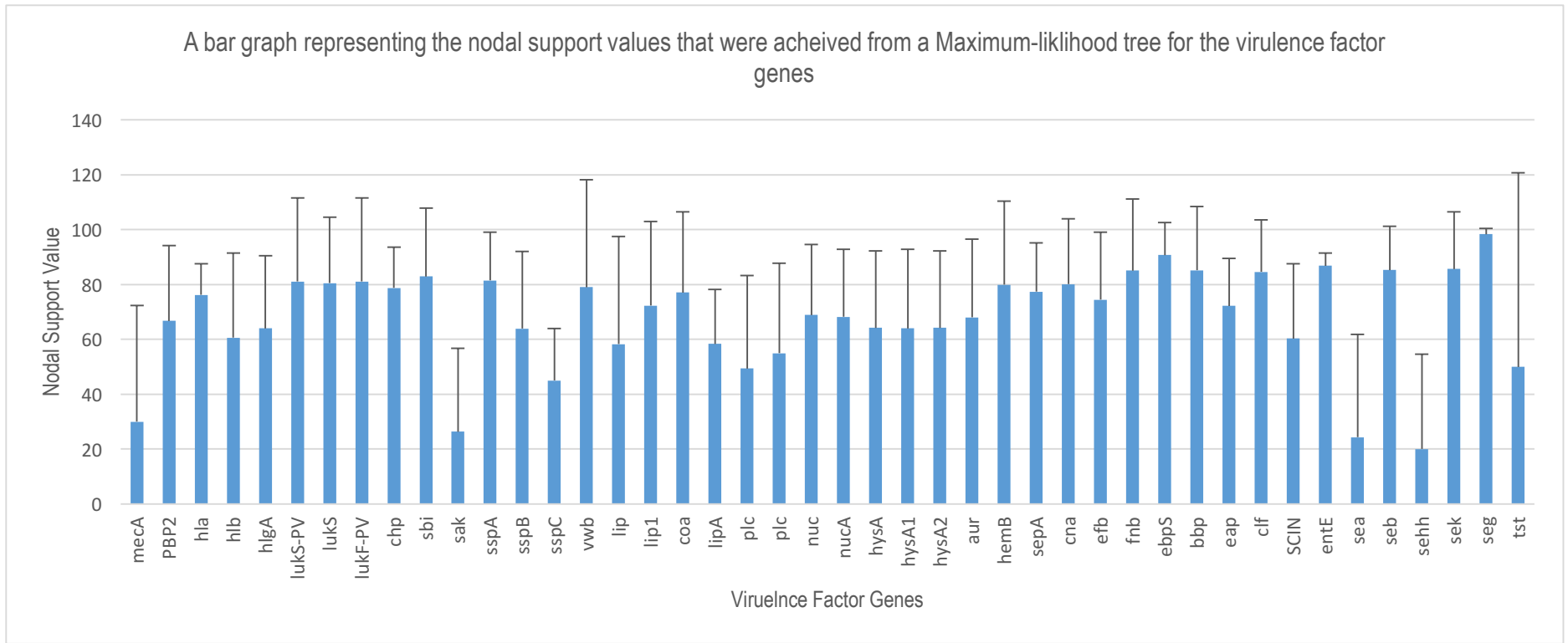


Figure 4.2. A bar graph representing the mean nodal support values (achieved from 500 bootstrap replications) from maximum-likelihood trees that were produced for the *S. aureus* virulence factor genes. The error bars represent the standard deviation of the nodal support values.

Figure 4.3 is a maximum-likelihood phylogenetic analysis revealing the relationship of the *S. aureus* strains for the *pbp2* gene which encodes penicillin-binding protein 2. The coloured boxes indicate clades that have formed as part of the classification that is observed in the literature. The red box indicates two livestock-associated clades. The first clade are the LA *S. aureus* strains that have caused disease in humans and the second clade are the rest of the LA strains that have been isolated from the animals. The blue boxes indicate that 3 hospital-associated clades that are being observed. It is interesting to note that no community-associated *S. aureus* strains were grouped in a clade, instead they are part of a larger clade that also includes HA and LA strains.

Figure 4.4 is a maximum-likelihood phylogenetic analysis revealing the relationship of the *S. aureus* strains for the *lukf-pv* gene which encodes the Pantone-Valentine leucocidin subunit S and PVL toxin. This gene was chosen as it is a representative of a cytotoxin gene. The coloured boxes again denote the clades that have formed as part of the classification in the literature. The clades are well defined for this gene.

Figure 4.5 is a maximum-likelihood tree that reveals the relationship of *S. aureus* strains for the *vwb* genes which encodes the von Willebrand factor-binding protein. This gene was chosen because it is a representative secreted protein that enhances virulence. The red box shows the novel LA strains and the blue box show the novel HA strains. Again CA does not form its own clade but is instead clustered among the LA and HA strains.

Figure 4.6 is a maximum-likelihood tree that reveals the relationship of *S. aureus* strains for the *seb* gene that encodes enterotoxin B. This gene was chosen because it is a representative enterotoxin. The red box shows the LA strains that do not cause disease in humans. A single CA strain sits apart from the others and 2 HA clades are also present.

Figure 4.7 is a maximum-likelihood tree representing the relationship for the *clf* gene which encodes clumping factor. The LA strains (red box) have formed the clades that do and do not cause disease in human. Again the CA strains are clustered amongst the HA and LA strains.

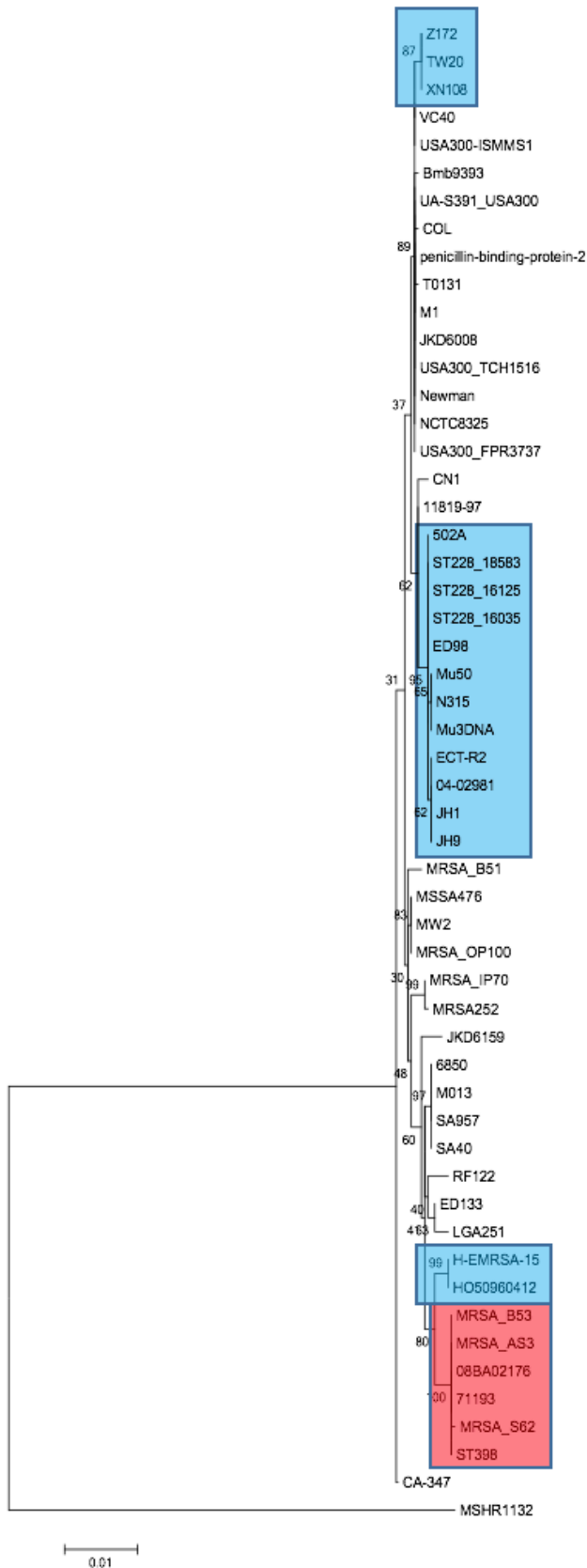


Figure 4.3: A Maximum Likelihood phylogenetic analysis, based on the General Time Reversible (GTR-G) evolution model. The relationship of HA, CA & LA *S. aureus* strains for the *PBP2* gene which encodes Penicillin-binding protein 2. The numbers at the nodes are bootstrap percent probabilities values based on 500 replications. Coloured boxes are consistent with strain classifications.

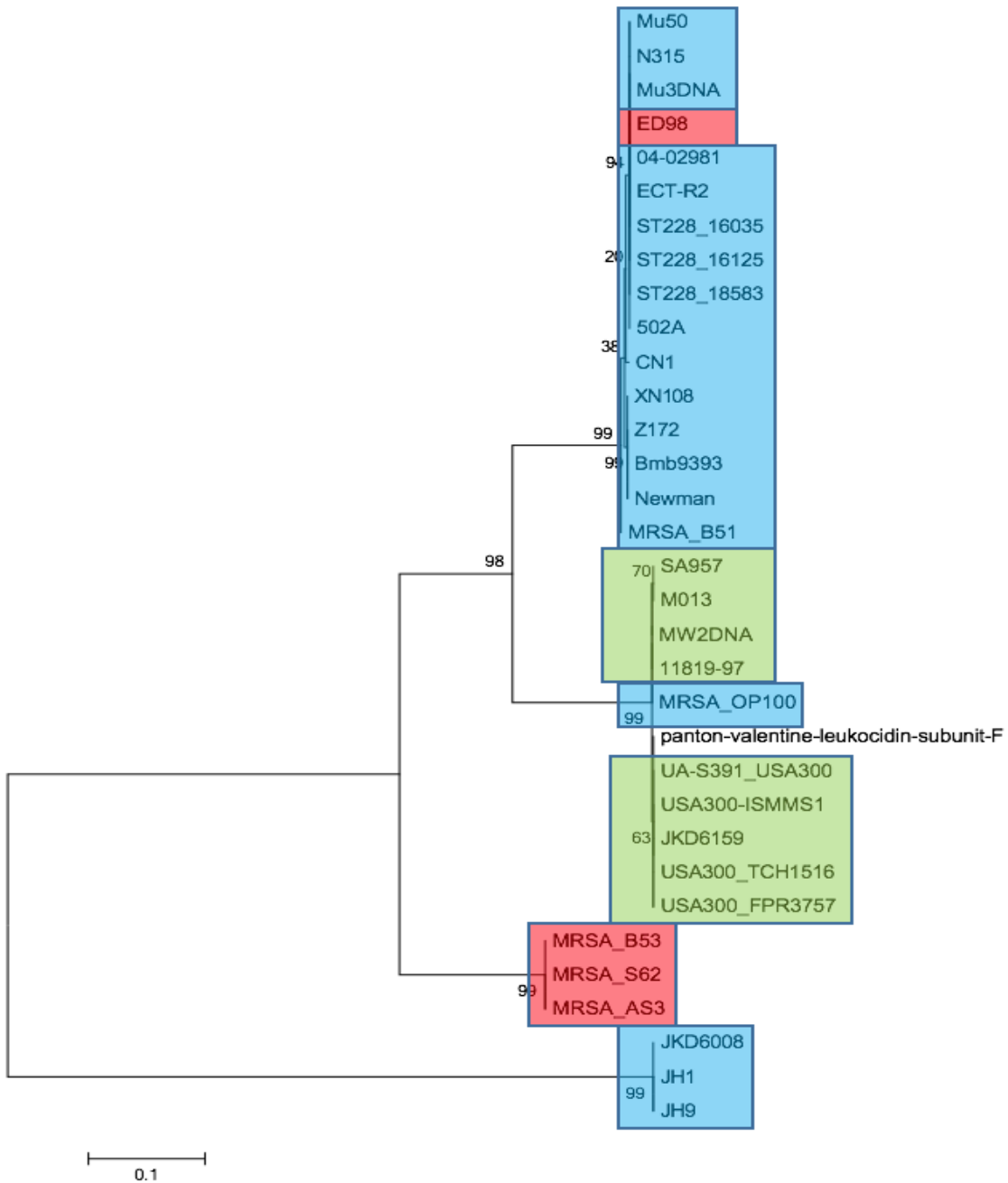


Figure 4.4: A Maximum Likelihood phylogenetic analysis, based on the Tamura Nei (TN93+G) model of evolution. The relationship of HA, CA & LA *S. aureus* strains for the *lukF-PV* gene which encodes Panton-Valentine leucocidin subunit F. The numbers at the nodes are bootstrap percent probabilities values based on 500 replications. Coloured boxes are consistent with strain classifications.

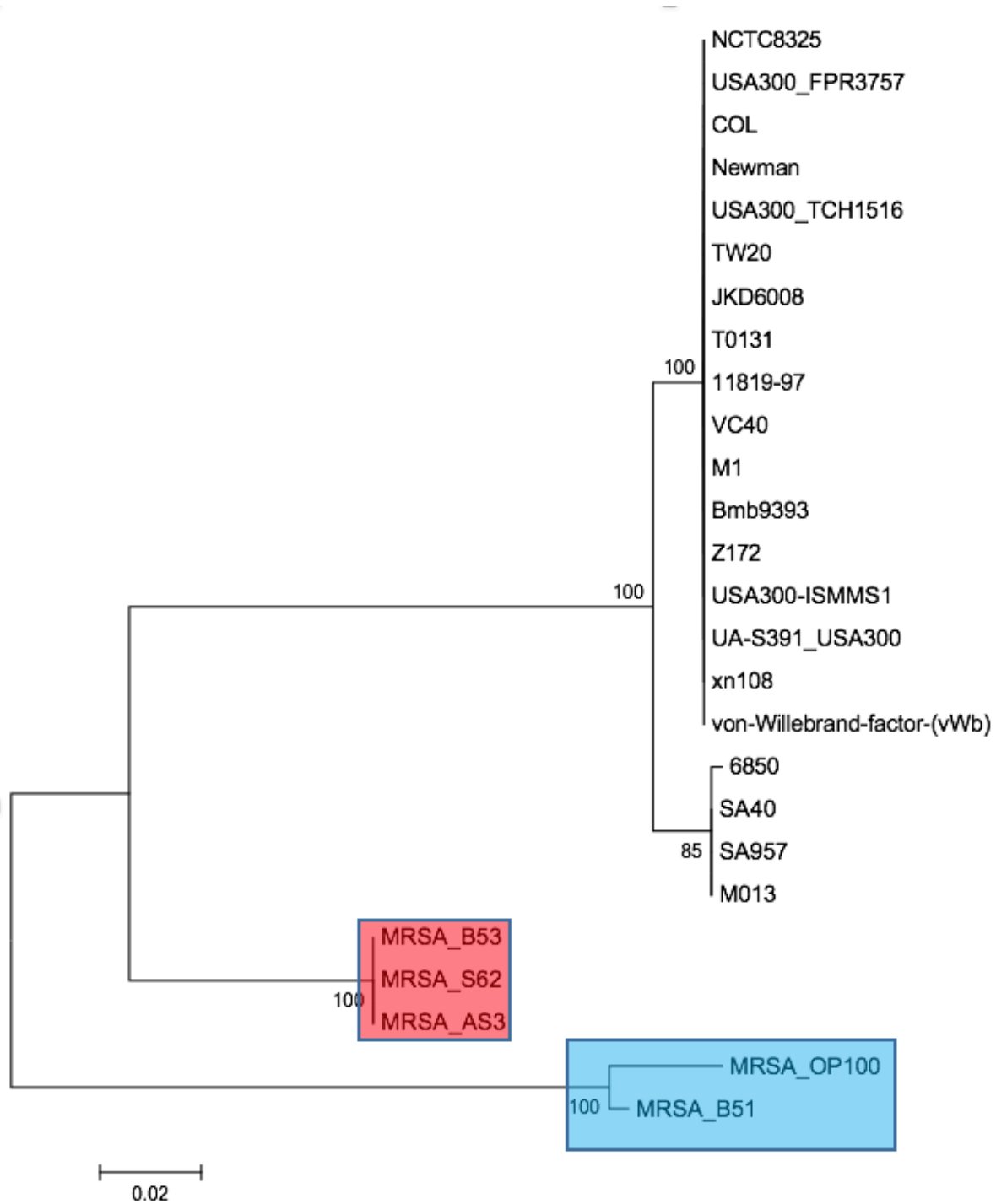


Figure 4.5: A Maximum Likelihood phylogenetic analysis, based on the Tamura 3-parameter (T92). The relationship of HA, CA & LA *S. aureus* strains for the *vwb* gene which encodes von Willebrand factor-binding protein. The numbers at the nodes are bootstrap percent probabilities values based on 500 replications. Coloured boxes are consistent with strain classifications.

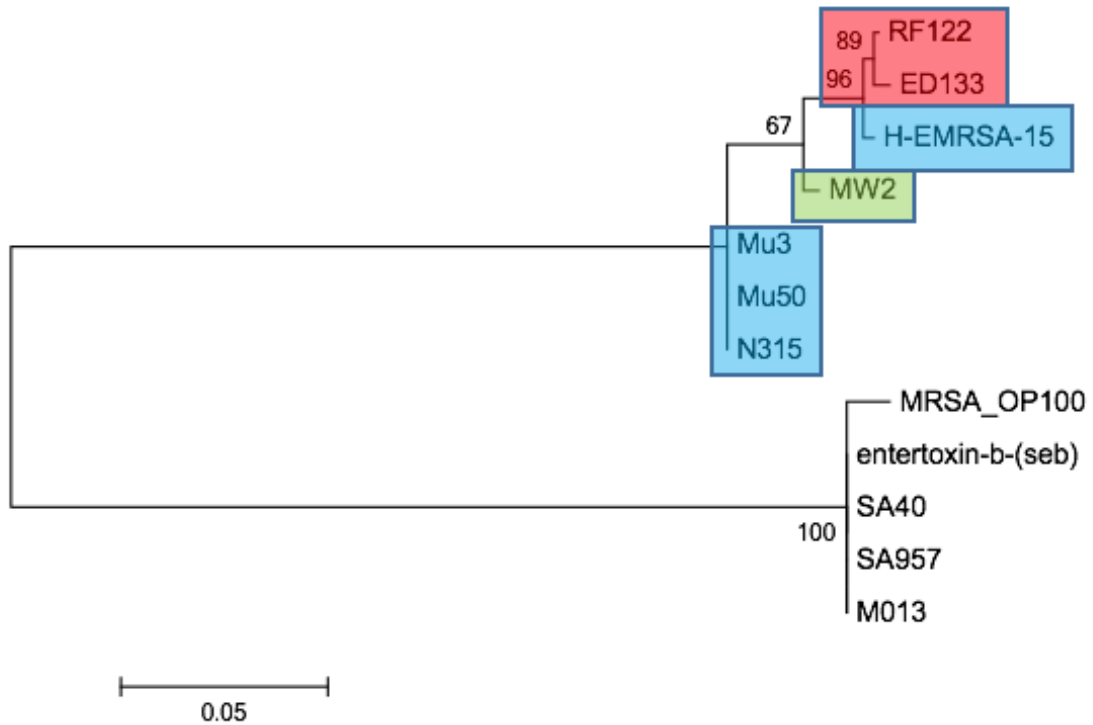


Figure 4.6: A Maximum Likelihood phylogenetic, based on the Hasegawa-Kisaino-Yano parameter (HKY). The relationship of HA, CA and LA *S. aureus* strains for the *seb* gene which encodes enterotoxin B. The numbers at the nodes are bootstrap percent probabilities values based on 500 replications Coloured boxes are consistent with strain classifications.

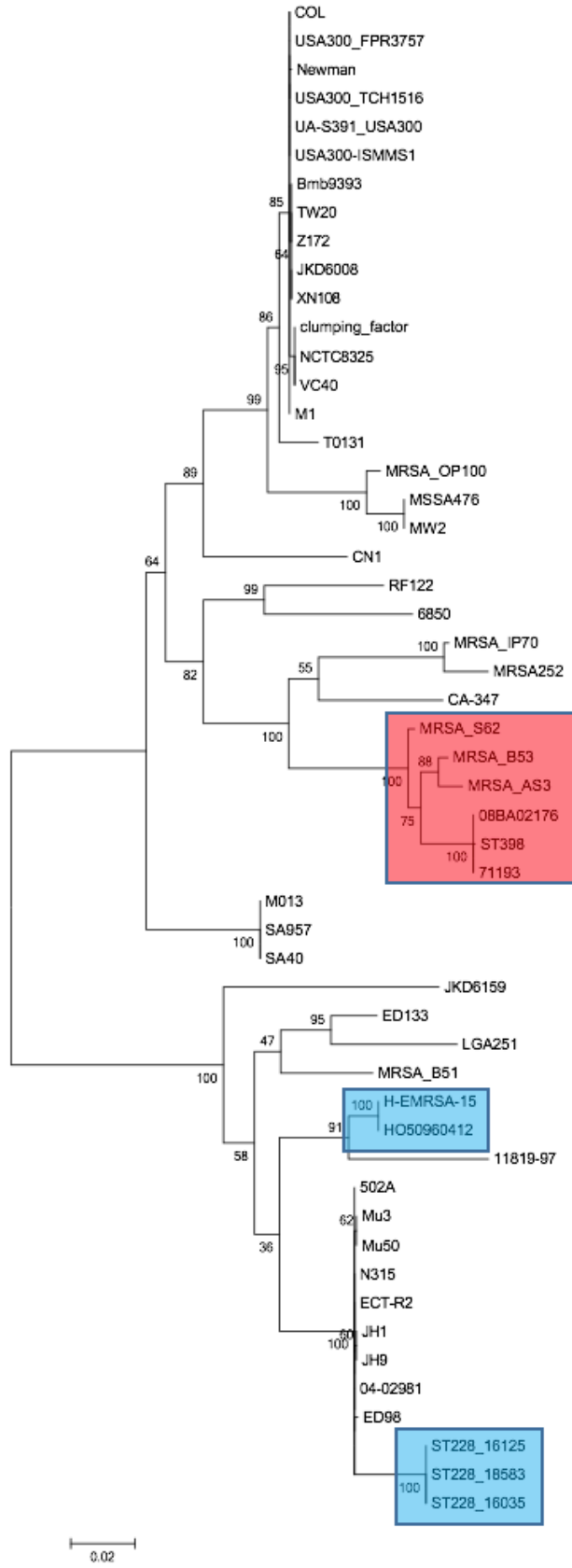


Figure 4.7: A Maximum Likelihood phylogenetic analysis, based on the General Time Reversible (GTR-G) parameter. The relationship of HA, CA & LA *S. aureus* strains for the *clf* gene which encodes clumping factor. The numbers at the nodes are bootstrap percent probabilities values based on 500 replications. Coloured boxes are consistent with strain classifications.

4.5 Discussion:

It was concluded by a comparative analysis of two phylogenetic trees (Figure 4.1a & 4.1b) created using two different techniques (Phylogeny.fr & MEGA) that the optimum methods of phylogeny reconstruction to be used in this study will be the maximum-likelihood method, using MEGA. This is further supported by the use of Bayesian statistical models that is incorporated within this method. It was observed in Figure 4.2 that a large majority of ML trees produced show high nodal support values. This suggests that the splits and subsequent branching occurring in the trees have actually occurred.

Figure 4.1b is a phylogenetic tree for the virulence factor alpha-hemolysin (*hla*) that is found in the majority of the samples being observed ($n = 52$). The LA strains that have been clustered together are highlighted with a 'red' box and two clades are presented. The first clade (B53, S62, AS3, 08BA02176, ST398 & 71193) consists of strains implicated with infection and diseases in humans. Subsequently, the second clade consists of LA strains that have not been implicated in human disease and instead have been isolates from either milk samples (LGA251) or bovine mastitis (ED133 & RF122). It is notable that ED98 has not clustered with this clade. It is suggested that ED98 has not clustered with the other LA strains because it is not a bovine strain or has not been implicated with humans, instead it has been isolated from a chicken with bacterial chondronecrosis and osteomyelitis. It is also notable that this second LA clade is part of a large clade that also includes hospital-associated (HA) and community-associated (CA). This general trend seems to exist within all of the ML phylogenetic trees that were created for the various virulence factors.

Another notable trend in a majority of the ML trees, with the exception of the gene *seb*, is the clustering of HA strains H-EMRSA-15 and H050960412. This is not that surprising as it has been previously reported that both of these strains are predominant in the UK, and both have previously been profiled to being 99.45% similar to each other (Sabirova *et al.*, 2014). This observation along with the consistently high nodal values, which are typically in the 90 – 99 percentile, suggests that the ML phylogenetic trees that have been produced are reflective of a true split.

Another pattern that seems to be notable within the majority of the ML trees is the clade that consists of HA strains XN108, Z172, Bmb9393, T0131, JKD6008, TW20. Interestingly these HA strains are disseminated globally (China, Taiwan, Brazil, New Zealand & UK) and do not possess any characteristic apart from the fact that they are hospital-associated. Again they are clustered in the majority of the ML trees that were produced. It is suggested that the possible cause for these HA strains clustering together could be the antibiotics that were used to treat them. If the same antibiotic has been used in different countries, it represents a common factor that is acting as a selective pressure being exerted on the *S. aureus* strain. This suggestion needs further investigation. Interestingly, these strains are part of a larger clade that also include the USA300 strains which are implicated in epidemiologically non-associated sporadic outbreaks in healthy individuals.

The final HA clade that appears in the majority of ML trees consists of strains MRSA252 & MRSA IP70. MRSA252 is a strain that was implicated in an outbreak at an ICU where all of the infected individuals died as a direct result of MRSA septicaemia. Whereas MRSA IP70, a newly sequenced strain was isolated from an inpatient, however no other information is available. Strains OP100 and B51 were isolated from outpatients, who by definition were hospitalised for less than 24 hours. These two novel strains are clustered with the other CA strains (MMSA476 and MW2) that have been implicated with paediatric deaths; it is reasonable to suggest that these strains may have been community-associated and not hospital-associated and further confirms this observation.

The green box (Figure 4.1) indicates a community-associated clade. Interestingly these three CA strains (SA40, SA957 and M013) are also Taiwanese clones so it is not surprising that they group together. It is therefore suggested that this CA clade is more likely due to a geographic relationship.

The maximum-likelihood tree that is of notable difference is Figure 4.4 and represents the *lukF-PV* gene. This ML tree represents an almost true split of the classifications (HA, CA & LA) of *S. aureus* that were initially to be expected before this study was undertaken. The exception was strain OP100 that is clustered

within a clade of CA strains. However, this coincides with the previous suggestion that because OP100 is from an outpatient, this strain may have been classified incorrectly. However, this observation is not consistent with all of the virulence factor genes. It was also previously concluded that PVL genes *lukS-PV* and *lukF-PV* that encodes the bi-component Panton-Valentine leucocidin (PVL) play a crucial role in the lysis of erythrocytes and leukocytes and are present in the novel livestock-associated strains. Two novel HA strains also possess both components of the PVL (B51 & OP100), again this has not been identified before.

The final ML tree presented in Figure 4.7 shows *clf* gene which encodes clumping factor protein. This is an important adherence gene that has a dual role. Clumping factor A bind to sites on fibrinogen which plays a major role in blood clotting and clumping factor B is a major determinant of *S. aureus* nasal persistence and colonisation in hosts (Wertheim *et al*, 2008; Xiang *et al*, 2010). It has been observed that in a healthy population, approximately 20% are colonised with *S. aureus* and a further 30% of the population are intermittently colonised (Kluytmans *et al*, 1997). It is interesting to note that the CA strains do not cluster in a single clade. This observation suggests that nasal persistence of *S. aureus* is not strain dependent as the majority of both HA and CA strains have the same gene for clumping factor. That being said, it is also observed that LA strains do show a difference but nevertheless still possess the clumping factor protein. This is further supported by the origin of the LA ST398 strain. Armand-Armand-Lefevre *et al* (2005) observed that pig farming is a risk factor for increased nasal *S. aureus* colonisation.

In conclusion these observations are suggesting that although only a small sample set of virulence factors phylogenies were presented, the overall trend observed in the ML analysis showed consistently similar results. This data does show some clusters, notably in the livestock-associated strains and less so in the hospital-associated strains. Almost no clusters exist within the community-associated strains with the exception of one but it was proposed that this was likely due to a geographical relationship. It is therefore reasonable to suggest that the lack of clustering into HA, CA and LA clades again indicates that the

classifications given to these *S. aureus* strains by the authors of their respective papers may be artificial and leads to the finding presented by Bal *et al.* (2016) that there is a blurring of these epidemiological classifications.

It is important to mention that only one virulence factor was phylogenetically analysed at one time. The differentiation between the virulence factor phylogenies are also suggestive of the fact that different selective pressures may be acting on the genes which are therefore evolving at different rates. This needs further investigation.

Chapter 5

Investigating positive selection in *S. aureus* and divergence between populations

5.1 Introduction:

The literature review discussed in Section 1.1 revealed that *S. aureus* is an established commensal organism that is relatively common as part of the human micro-flora, however it is also an opportunistic pathogen that can cause disease, especially if the immune system is compromised (Sousa & Lenastre, 2004). For this organism to become pathogenic, they must have the ability to establish infection and cause disease. *S. aureus* accomplishes this task by expressing virulence factors that will enable it to successfully attach and establish itself in/on the host and evade its defences. However, in order for *S. aureus* to survive in the host, it must adapt to any selective pressures such as changes to the hosts biology, the hosts immune system and any antibiotic therapy that may administered.

Genomes can undergo small changes as well as large-scale evolutionary changes. Frequent genome rearrangements are caused by genomic recombination events. Horizontal transfer has the ability to introduce new sequences into the bacterial chromosomes whereas deletions remove portions of the genomes (Darling *et al.*, 2004). Substitutions of bases also readily occur. Substitutions are mutations that can exchange one base of the genome for another, this could result in a change of the codon that can then encode a different amino acid and consequently change the protein that will be produced. A synonymous substitution is a change in the DNA sequence that codes for amino acids in a protein sequence, but does not change the encoded amino acid sequence. In comparison, a nonsynonymous substitution is a nucleotide mutation that alters the amino acid of a protein. Sequence alignments are therefore absolutely necessary for bioinformatics studies. Sequence alignments arrange the DNA nucleotide sequences to identify regions of homology among the strains. These regions of similarity/ dissimilarity are presented as rows and identifies nucleotide residues that may be conserved, not conserved (or mismatch) or not aligned. The mismatch of the DNA sequences can be caused by substitutions or point mutations and the non-alignment can be caused by the insertion and/ or deletion of base(s) (Indels) (Edgar, 2004).

The genes encoding the various *S. aureus* virulence factors will be investigated and analysed to

determine if they are under positive selection by comparing the rates of synonymous versus nonsynonymous mutation (dN/dS or ω). DnaSP (Librado & Rozas., 2009) generated average number of nucleotide differences between the sequences per nonsynonymous site (dN) and also the average number of nucleotide differences between sequences per synonymous sites (dS). The ratio $\omega = dN/dS$ can therefore be calculated and is a measure of the difference between the two sites (Yang & Bielawski, 2000). The virulence genes have different ω (dN/dS) values and inferences about the evolutionary mechanisms acting on the specific gene can therefore be made. If $\omega = 1$ it implies that the amino acid change is neutral and that the DNA sequence is evolving neutrally, if $\omega < 1$ it implies that purifying selection is acting on the amino acid changes to filter out deleterious mutations and remove any non-synonymous substitutions from the population in order to preserve the function of the protein, only when $\omega > 1$ does it imply that the amino acid change for the gene offers a selective advantage, providing evidence of diversifying or positive selection resulting in some kind of phenotypical outcome that is of benefit (Yang & Bielawski, 2000).

5.2 Aims:

The aims of this present study were to:

- To perform a genome-wide analysis on *S. aureus* virulence factor genes, in order to identify genes that are undergoing adaptive evolution.
- Investigate divergence between populations of *S. aureus*

5.3 Materials and Methods:

- DnaSP analysis was carried out to determine nucleotide polymorphism from aligned DNA sequence data. The query virulence factor genes were loaded (Section 2.15.1). The polymorphism data was generated (2.15.2). The polymorphic sites data was generated (2.15.3). The synonymous and nonsynonymous substitutions were generated (Section 2.15.4). And the DNA divergence was also calculated

5.4 Results

5.4.1 Haplotype and nucleotide diversity

The data that was produced from the DnaSP analysis is available in full in Appendix 1. To determine the nucleotide DNA sequence variation in the *S. aureus* population, the proportion of nucleotide positions that are polymorphic can be determined. Polymorphic or variable sites (s) were determined from the data. Polymorphic sites within the *S. aureus* DNA can be calculated by measuring the haplotype diversity (h) and nucleotide diversity per site (π) of the virulence factors. Haplotype diversity (h) is a measure of the probability that two randomly sampled alleles are different (de Jong *et al.*, 2011). Whereas nucleotide diversity per site (π) is a measure of the average number of nucleotide substitutions between two populations or species and is used to measure the degree of polymorphism within a population (Nei & Li, 1979). Studies carried out by Goodall-Copestake *et al* (2012) estimated the genetic diversity of animal *cox1* genes by calculating the h and π . The effect of sample size on haplotype diversity and nucleotide diversity were investigated; it was suggested that samples analysis of $n \leq 5$ individuals has shown to hinder accurate h and π comparisons where as sample sizes ≥ 25 are recommended for greater accuracy (Goodall-Copestake *et al.*, 2012).

In this study the sample size of the *S. aureus* organisms was 52 strains with the frequency of the genes that are present ranging from 7 strains to 52, with the exception of the *seh* gene that was present in 5 strains, suggesting that an overall good sample size is being tested.

A graph (Figure 5.1) was subsequently produced to show the the nucleotide diversity (π) within the virulence factor genes. The graph shows that the overall nucleotide diversity of the virulence factor genes varies greatly. This suggests that variation does exist between populations of *S. aureus*, as demonstrated by the π of the virulence factor genes.

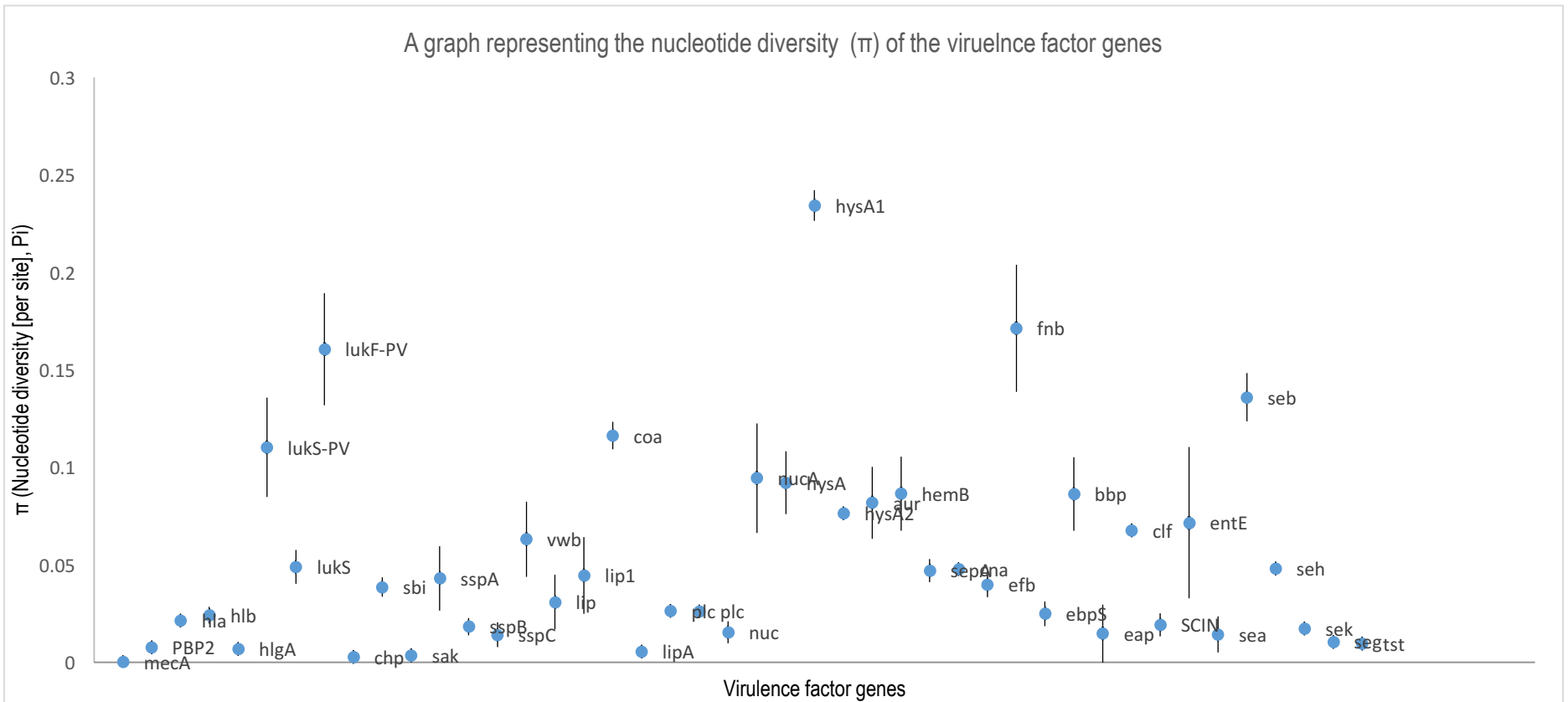


Figure 5.1: A graph representing the nucleotide diversity (π) of the *S. aureus* virulence factor genes. The error bars represent the standard deviation of the nucleotide diversity. Data was generated using DnaSP. The complete data generated is shown in Appendix 1.

5.4.2 Positive selection

To examine and subsequently explain the genetic variation that was observed in the *S. aureus* virulence factors, the dN/dS (ω) ratios were calculated. The ω was calculated by using orthologous gene sequences that were found within in multiple *S. aureus* genome strains. This ratio determines if selective pressure is being exerted on the gene. The summary data in Table 5.1 were extracted from Appendix 1 and it shows the virulence factor genes that are under positive selection including the statistical significance of this observation.

Virulence factor gene	N (No. of sequences)	S (No. of variable sites)	Hd (Haplotype diversity)	dN	dS	ω (dN/dS)	Tajimas D	Statistical significance	Fu & Li's D	Statistical significance
<i>PBP2</i>	57	2	0.94 (+/- 0.015)	26	202	0.12	-2.41	p<0.01	-6.11	p<0.02
<i>hlyB</i>	52	134	0.95 (+/- 0.019)	128	16	8	-1.39		-2.46	p<0.05
<i>hlyA</i>	56	58	0.87 (+/-0.032)	13	47	0.28	-1.79		-4.49	p<0.02
<i>lukF-PV</i>	34	570	0.86 (+/- 0.034)	161	123	1.31	-0.38		2.04	p<0.02
<i>sspA</i>	57	411	0.86 (+/- 0.029)	219	88	2.49	-2.47	p< 0.01	-6.01	p<0.02
<i>sspB</i>	57	201	0.89 (+/- 0.021)	65	142	0.46	-1.88	p<0.05	-4.87	p<0.02
<i>sspC</i>	57	69	0.82 (+/- 0.035)	24	75	0.32	-2.47	P<0.01	-6.06	p<0.02
<i>vwb</i>	27	134	0.55 (+/- 0.107)	78	42	1.86	-0.71		1.39	p<0.05
<i>lip</i>	57	970	0.86 (+/- 0.024)	457	361	1.27	-2.63	p<0.001	-6.69	p<0.02
<i>lip1</i>	56	845	0.93 (+/- 0.018)	365	143	2.55	-2.64	p<0.001	-6.49	p<0.02
<i>lipA</i>	56	56	0.83 (+/- 0.033)	1	59	0.02	-2.11	p<0.05	-5.22	p<0.02
<i>plc</i>	57	132	0.89 (+/- 0.022)	43	95	0.45	-0.61		-2.64	p<0.05
<i>plc</i>	57	132	0.89 (+/-0.022)	43	95	0.45	-0.64		-2.64	p<0.05
<i>nuc</i>	56	136	0.85 (+/- 0.032)	60	71	0.85	-2.38	p<0.01	-5.99	p<0.02
<i>hysA</i>	57	1500	0.89 (+/- 0.028)	626	235	2.66	-1.66		-4.76	p<0.02
<i>aur</i>	57	565	0.90 (+/- 0.021)	253	100	2.53	-1.78		-5.77	p<0.02
<i>sepA</i>	55	65	0.85 (+/- 0.031)	8	59	0.14	1.75		1.66	p<0.02
<i>cna</i>	22	225	0.89 (+/- 0.055)	127	206	0.61	2.01	p<0.05	1.15	
<i>ebpS</i>	56	335	0.89 (+/- 0.028)	253	71	3.56	-1.96	p<0.05	-5.60	p<0.02
<i>SCIN</i>	57	79	0.82 (+/- 0.042)	37	37	1	-2.17	p<0.05	-5.29	p<0.02
<i>entE</i>	19	374	0.54 (+/-0.136)	241	62	3.89	-2.39	p<0.01	-3.16	p<0.02
<i>sea</i>	19	84	0.45 (+/-0.128)	39	84	0.46	-2.24	p<0.01	-3.37	p<0.02
<i>seb</i>	14	213	0.80 (+/- 0.094)	138	68	2.03	2.59	p<0.01	1.36	
<i>seh</i>	5	81	0.4 (+/- 0.237)	26	55	0.47	-1.26		-1.26	p<0.02
<i>sek</i>	18	34	0.82 (+/- 0.237)	30	4	7.5	1.13		1.66	p<0.02

Table 5.1: Table showing the virulence factor genes that are under positive selection and the statistical significance tests. Data was generated using DnaSP. The complete data generate dis shown in Appendix 1.

Similar studies have previously demonstrated the role of the dN/dS. Figure 5.2a was produced from the dN/dS data generated for all *S. aureus* strains and all virulence genes. This graph shows the various virulence factors along the *x* axis and the nonsynonymous to synonymous ratio ($\omega = \text{dN/dS}$) along the *y* axis. This graph shows an array of results which suggests that the various virulence factor genes are either evolving neutrally ($\omega = 1$), under purifying selection ($\omega < 1$) or under positive selection ($\omega > 1$). The data suggests that the vast majority of the virulence factor genes ($n = 21$) are undergoing positive selection, $\omega > 1$. Likewise, many virulence factor genes ($n = 16$) are also demonstrating purifying selection, $\omega < 1$. Only a small minority of the genes ($n = 3$) are evolving neutrally ($\omega = 1$), this is demonstrated on the graph by the placement of the gene being on the *x* axis.

In this study it is seen that genes under positive selection have been found in all virulence factor categories, and that the genes include; *LukS-PV*, *lukF-PV*, *chp*, *sspA*, *vwb*, *lip*, *lip1 nucA*, *hysA*, *hysA1*, *aur*, *fnb*, *ebpS*, *bbp*, *entE*, *seb* and *tst*. It is interesting to note that the dN/dS value for the *mecA* gene is $\omega = 1$ which suggests that the amino acid substitution may be largely neutral. The *mecA* gene, which is carried on a Staphylococcal Cassette Chromosome (SCC*mec*) is present in many strains of *S. aureus* that are resistant to methicillin; it is observed that the various strains have acquired the SCC*mec* by horizontal transfer that appears to be neutral, and not evolving (Ito & Hiramatsu, 1998; Katayama *et al.*, 2000). Where the *mecA* was not present it is likely that divergent *mecA* genes such as the *mecIga251* or *mecC* are present but were not identified in this study.

The next stage of the analysis focused on the livestock-associated strains only. This was because the livestock-associated *S. aureus* strains were only recently noted and it would be interesting to investigate the genes that are under positive selection. Figure 5.2b is a graph representing the dN/dS ratio of all of the virulence factor genes present in only livestock-associated *S. aureus* strains. The sample size of the

livestock-associated strains was comparably smaller ($n = 10$) than all of the combined strains ($n = 52$). Similar genes as before are showing the same characteristics, for instance the *hly* gene is amongst the virulence genes that is showing considerable positive selection ($\omega > 10.5$). It is also interesting to note that the *mecA* gene is again evolving neutrally ($\omega = 1$). Other genes have shown the same rate of evolution and are evolving under purifying selection ($\omega < 1$) namely *PBP2*, *hly*, *hlyA*, *spA*, *sspC*, *lipA*, *plc*, *sepA*, *cna*, *ebb* and *eap*. The genes that are under purifying selection would not be expected to change. The notable difference in the *clf* gene which is under a considerable amount of positive selection in the livestock-associated *S. aureus* strains ($\omega > 6.03$) whilst it is under purifying selection when collated with all of the other strains ($\omega > 0.85$). The other notable difference is the absence of the enterotoxins from the livestock-associated strains.

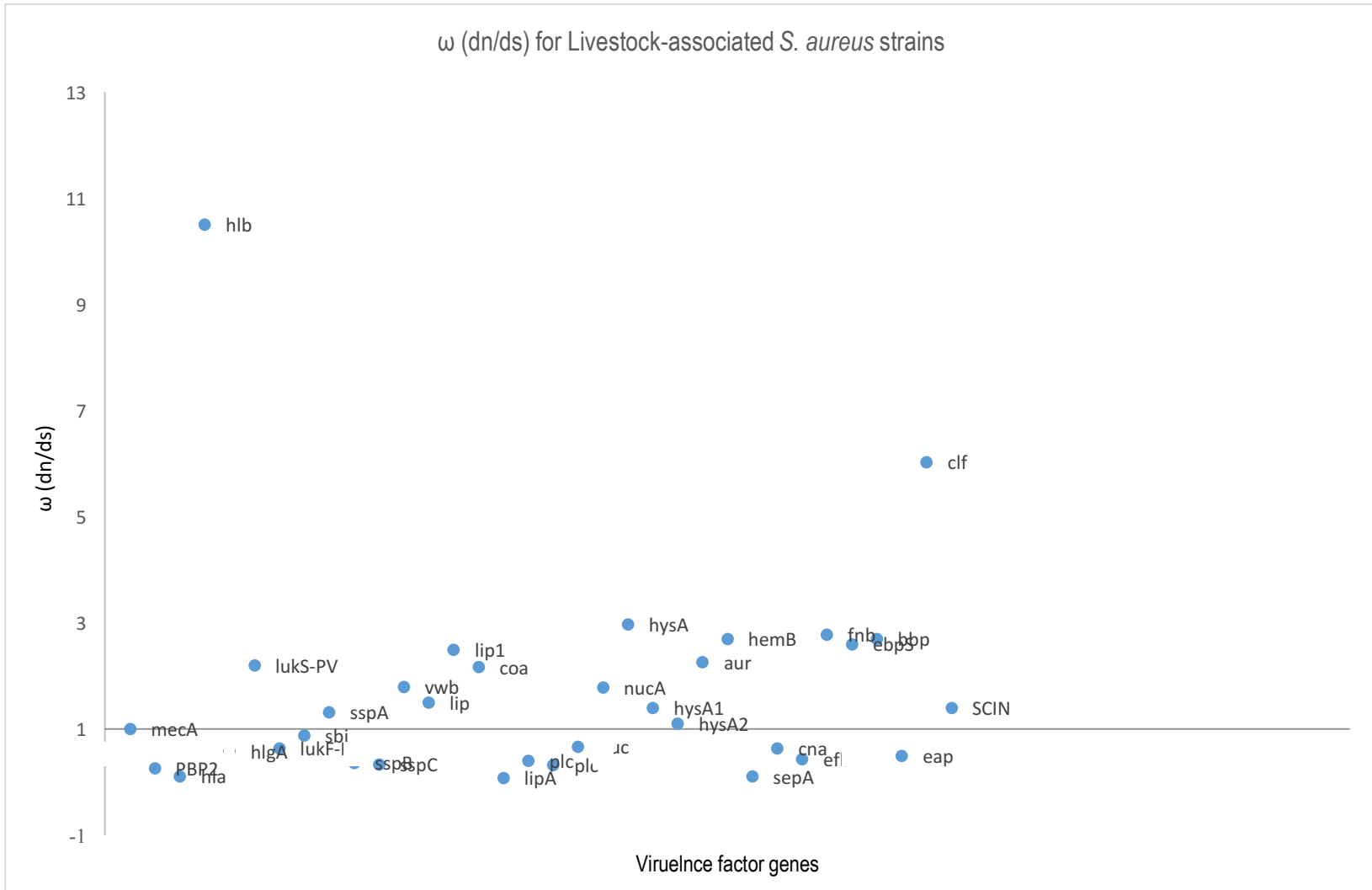


Figure 5.3a and Figure 5.3b were subsequently produced from the data that is presented in Appendix 1. Figure 5.3 is a regression analysis that shows the length of the virulence factor gene sequences along the x axis and the nucleotide diversity of the virulence factor genes along the y axis. The length of the virulence factor genes nucleotide sequence ranges from 330 bp (base pairs) for the *sspC* gene to 4872 bp for the *hysA1* gene. A single nucleotide polymorphism (SNP) is a variation in a single nucleotide that occurs at a specific position along the nucleotide sequence. Since nucleotide diversity is used to measure the degree of polymorphism within a population the expected trend from the regression analysis would be that the longer the virulence factor nucleotide sequence the greater the chance of a SNP occurring, this graph shows that this is statistically highly significant ($P < 0.001$). As the graph shows, the nucleotide diversity increases as the length of the nucleotide sequence increases.

A second regressions analysis was also undertaken (Figure 5.b), to determine if there was a relationship between the observed dN/dS ratio of the virulence factor gene and the length of its nucleotide sequence. The graph produced shows the length of the virulence factor nucleotide sequences along the x axis and the dN/dS ratio (ω) along the y axis. Since dN/dS ratio is the numbers of synonymous mutations versus the nonsynonymous mutation that are being observed on the nucleotide sequence of the virulence factor gene; the expected trend would be to observe a greater number of mutations occurring as the length of the virulence factor genes nucleotide sequence increases, however this is not the case. This graph suggests that this relationship does not exist, instead there is a large dN/dS clusters near the 150 bp to 1500 bp mark ($P = 0.704$).

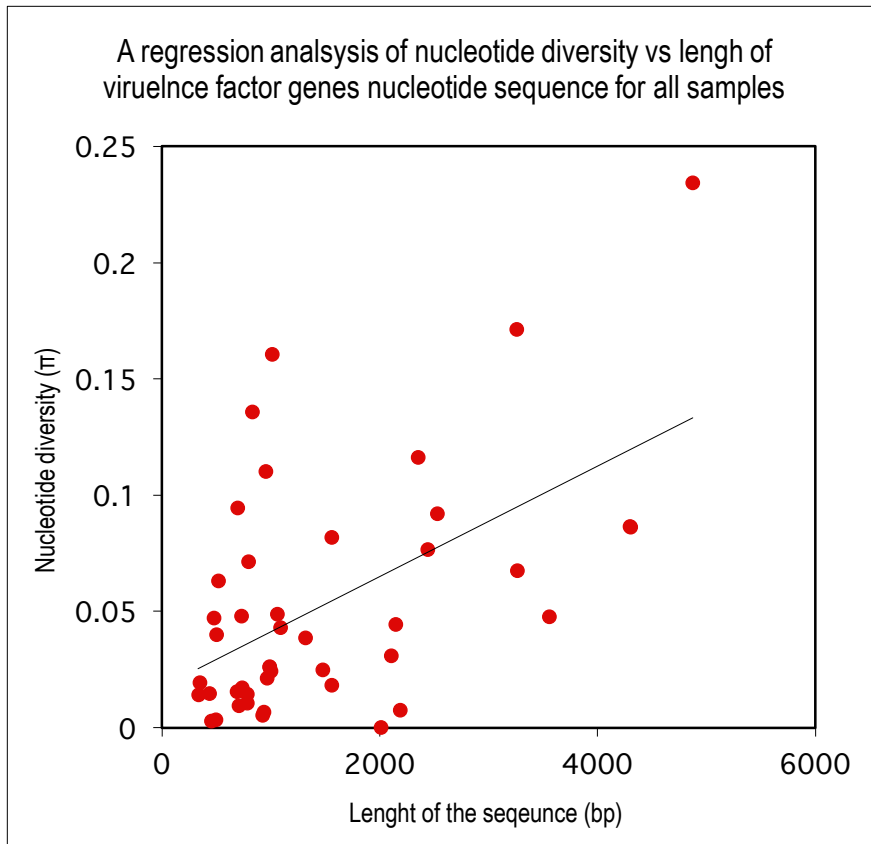


Figure 5.3a: A regressions analysis showing the observed nucleotide diversity (π) vs the length of the virulence factor gene. The trend line is shown. (Slope: 0.000024; Intercept 0.017501; r^2 0.292441; DoF 42; $P=0.000151$ therefore $P<0.001$).

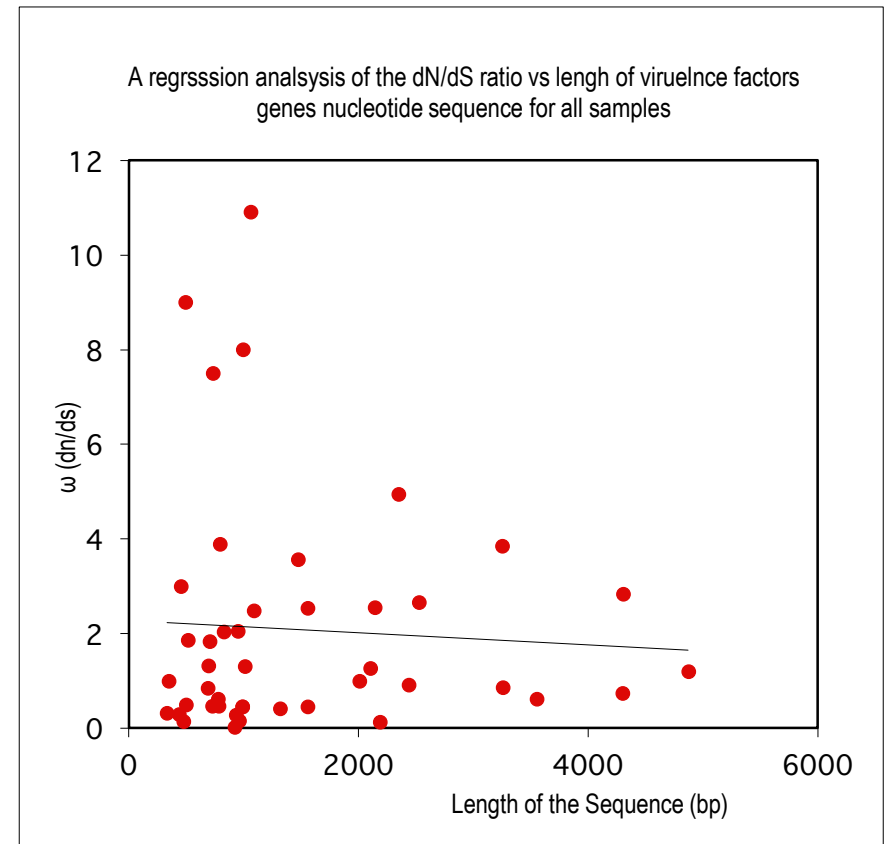


Figure 5.3b: A regression analysis showing the observed dN/dS ratio (ω) vs the length of the virulence factor gene. The trend line is shown. (Slope: -0.000127; Intercept 2.271417; r^2 0.003483; DoF 42; $P=0.704$).

A second regression analyses was then undertaken of the livestock-associated strains to determine the significance of the data. Figure 5.3a was then produced by the data that is presented in Appendix 2. This regression graph shows the length of the virulence factor nucleotide sequences along the x axis and the nucleotide diversity of the virulence factor genes along the y axis. This graph suggests that there is correlation between the two variables. Although there a major cluster between 250 bp to 1000 bp, the graph does suggest that as the length of the nucleotide sequence increases there so does the nucleotide diversity, this is supported by the $P < 0.001$. The number of *S. aureus* sequences varied from $n = 3$ for the *chp* gene to a maximum of $n = 10$ for the *hla* gene. Goodall-Copestake *et al.* (2012) concluded that a sample size of ≥ 25 is recommended for greater accuracy, this has not been the case for this analysis. Since the number of *S. aureus* sequences sample size did not exceed ≥ 25 , it is suggested that a greater sample set is required to improve the accuracy of this observation.

A second regressions analysis was again undertaken (Figure 5.3.b), to determine if there was a relationship between the observed dN/dS ratio of the virulence factor gene and the length of its nucleotide sequence. The graph produced shows the length of the virulence factor nucleotide sequences along the x axis and the dN/dS ratio (ω) along the y axis. This graph suggests that this relationship does not exist, instead there is a large dN/dS clusters near the 150 bp to 1500 bp mark, with a P value this is $P = 0.145$. This again suggests no significance.

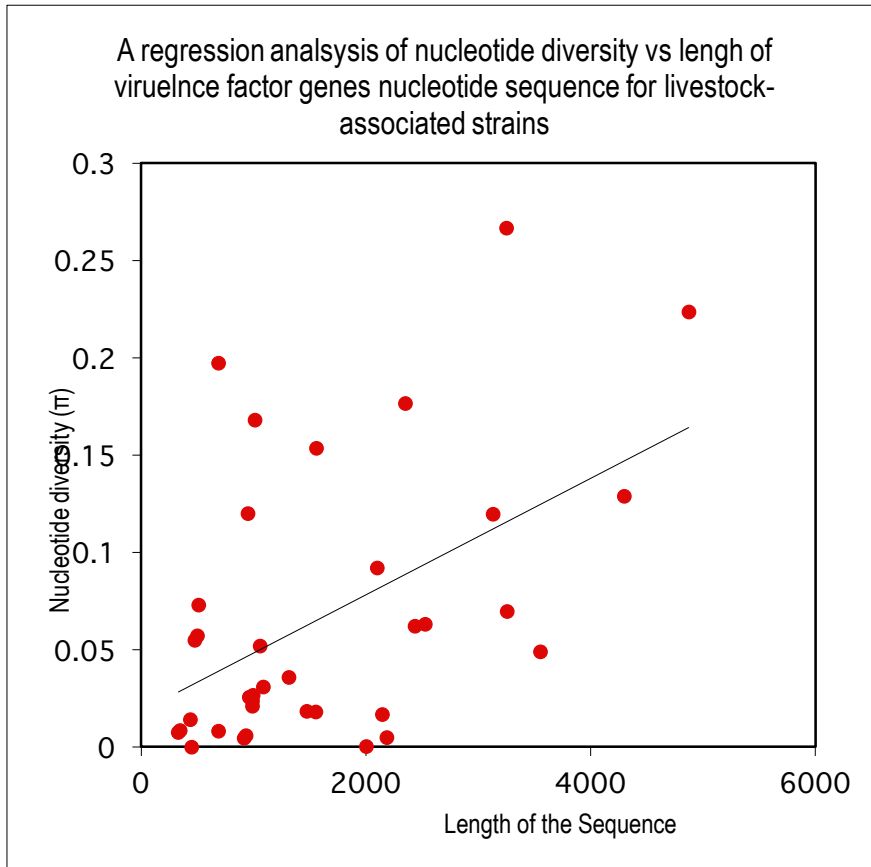


Figure 5.4a: A regressions analysis showing the observed nucleotide diversity (π) vs the length of the virulence factor gene in livestock-associated strains. The trend line is shown. (Slope: 0.00003; Intercept 0.018278; r^2 0.248114; DoF 34; $P=0.001991$)

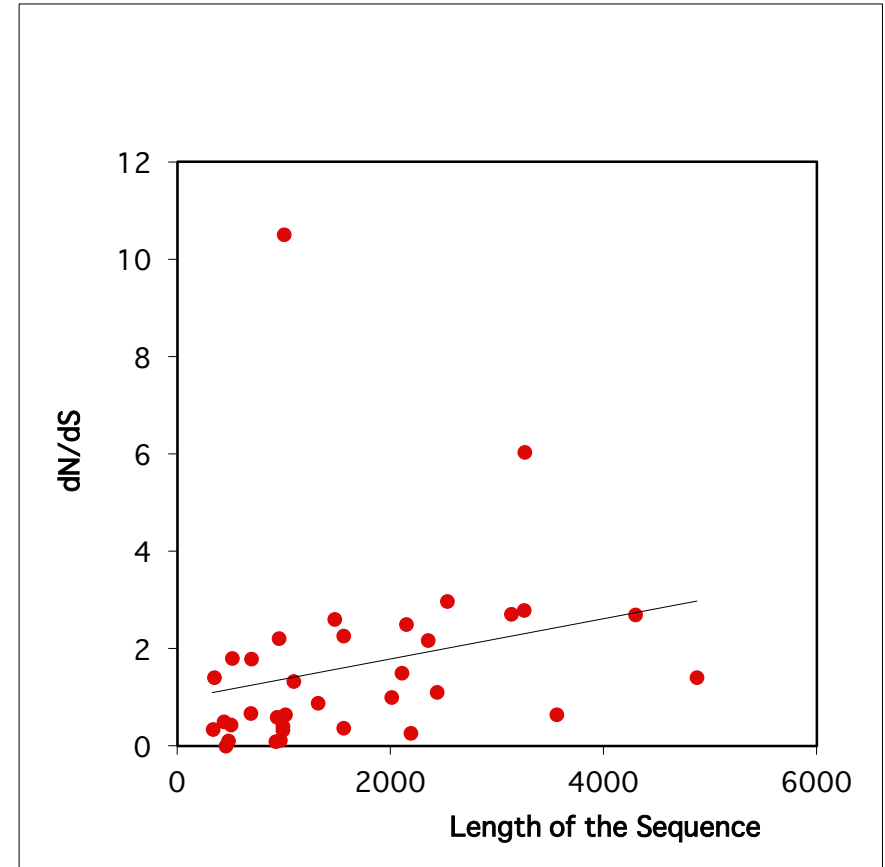


Figure 5.4b: A regressions analysis showing the observed nucleotide diversity (π) vs the length of the virulence factor gene in livestock-associated strains. The trend line is shown. (Slope: 0.000414; Intercept 0.951858; r^2 0.061562; DoF 34; $P=0.145$)

5.4.3 Divergence

The next step was to investigate the divergence (Dxy) between the populations of *S. aureus* that have been classified, namely hospital-associated, community-associated and livestock-associated. The Dxy measure corresponds to the average number of sequence differences per nucleotide position between sequences sampled from two nucleotide populations, this was implemented by using DnaSP (Westram *et al.*, 2016).

Figure 5.5 is a graph representing the divergence that was observed between populations of *S. aureus*. The observations suggest that the populations are not very divergent. A similar study by Jiufeng *et al.* (2013) investigated the divergence of *Clonorchis sinensis*, a liver fluke species in China. The authors obtained 255 isolates that spanned over 17 provinces of China. Differentiation analysis revealed three clusters (A, B & C); however, it was concluded that there was low divergence within the populations with Dxy ranging from 0.4% to 0.8%, nevertheless a geographical link was observed (Jiufeng *et al.*, 2013). Geographical linkage in *S. aureus* was investigated further.

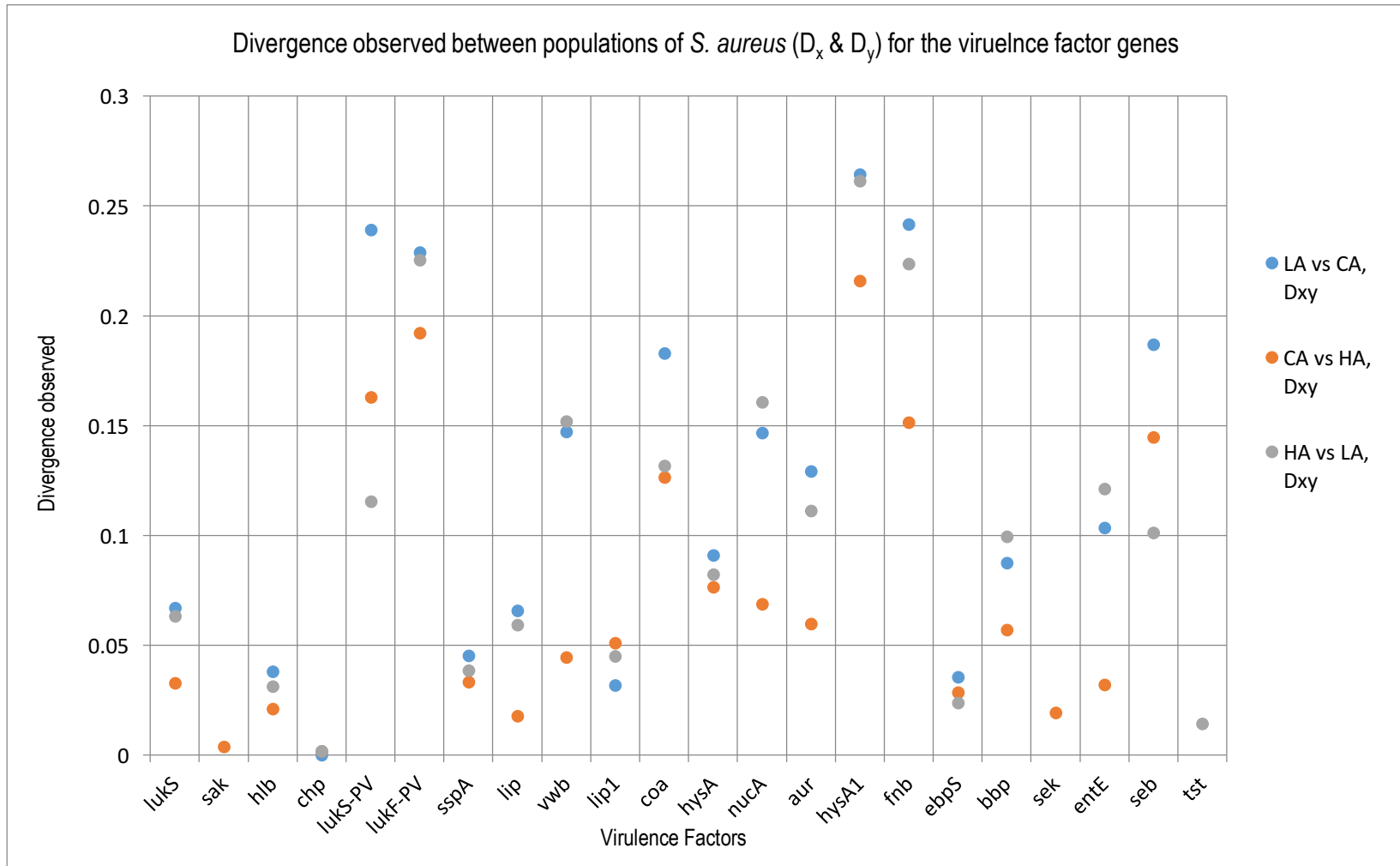


Figure 5.5: A graph representing the pairwise nucleotide differences observed between two populations (D_x & D_y). This is used to quantify divergence within the populations.

Discussion

Bacteria need to adapt to its ever changing environment in order for it to survive and grow in the host. It is hypothesised that selective pressures, that are put on the bacteria, are the driving forces behind bacterial evolution (Lambert & Kussell, 2015). Investigating and analysing evolutionary events that occur in wild populations that can be possibly related to nucleotide changes is not an easy task and relies on assumptions. It is however possible to investigate 'selective pressures' in a controlled setting. For instance, Mwangi *et al.* (2007) conducted an experiment where a patient was treated with Vancomycin for three months due to reoccurring *S. aureus* infections. Samples of the bacterium were collected before and during intervals of the therapy, for instance the first isolate (JH1) recovered at the beginning of the treatment was fully susceptible to Vancomycin (MIC = 1 µg/ml). The last isolate recovered at the end of the therapy (JH9) showed a decrease susceptibility to the chemotherapy (MIC = 8 µg/ml), the bacterium ultimately developed resistance to this antibiotic and the treatment ultimately failed. There were only 35 point mutations identified in 31 loci between JH1 and JH9. The resistance patterns were correlated by observing the nucleotide substitutions. The first and second isolates (JH1 & JH2) only differed by six nucleotide substitutions (increasing the MIC from 1.0 to 4.0 µg/ml); and between the second and fifth mutation (JH2 and JH5) there was a further increase in tolerance to Vancomycin (from 4 - 6 µg/ml) caused by a single additional mutation in gene SA1249, of unknown function (Mwangi *et al.*, 2007). It is interesting to note that only a few mutations can cause a bacterium's tolerance to chemotherapy increase. The impact of nucleotide polymorphisms on virulence has also recently been demonstrated by identification of InDel's and SNPs in the comparative analysis of two methicillin resistant *S. aureus* strains, USA300_FPR3757 and USA300-HOU-MR (Diep *et al.*, 2006; Highlander *et al.*, 2007). Highlander *et al.* (2007) identified an Indel in an operon (snoABCDEFGG) in USA300-HOU-MR which results in reduction in susceptibility to thrombin-induced platelet microbicidal protein 1 (tPMP-1). It has been suggested that this allows USA300-HOU-MR to evade platelet-mediated killing in the blood stream and can therefore lead to a more severe disease (Highlander *et al.*, 2007).

Evolutionary analysis has been undertaken on many important human pathogens. A study by Lefebure & Stanhope (2007) investigated the genus *Streptococcus*. Analysis was carried out on 26 *Streptococcus* strains to determine levels of recombination and positive selection. This study reported that positive selection was observed in 11% to 34% of the core genome with the authors further concluding that positive selection was a much slower process on genome structure than recombination but nevertheless played an important role in species differentiation and host adaptation of different species to different hosts (Lefebure & Stanhope, 2007). Another similar study by Lefebure & Stanhope (2009) investigated the genus *Campylobacter* it was concluded that 92.5% of the non-recombinant genome of one lineage was under positive selection. Again, it was suggested that this may be due to the various species of bacteria in competition for resources in the gastrointestinal tract of animals and humans, resulting in a 'Red Queen macroevolutionary dynamic Hypothesis' whereby each competing bacteria must constantly adapt in order to maintain its position within the ecosystem (Lefebure & Stanhope, 2007). Studies by Soyer *et al.* (2009) have shown evidence of positive selection that has also been noted in selected *Salmonella* serotypes. *Salmonella* causes disease in humans and different animal species and is another pathogen of great interest. Five *Salmonella* genome sequences were investigated that represented a broad host range and were used to identify core chromosomal genes that show evidence for recombination and positive selection. In total, 41 genes with confirmed or likely roles in virulence were found to be under positive selection, interestingly these were all located in *Salmonella* pathogenicity islands (SPIs). It was also concluded that positive selection contributes to the evolution of host restricted *Salmonella* serotypes (Soyer *et al.*, 2009).

Positive selection was investigated in *S. aureus* species that are not only present in humans but also in a variety of different host species. These hosts will have distinctly different immune systems as well as different biological pressures exerted on the bacterium. In addition to this, the different hosts will also have

different therapeutic pressures exerted on to it which will cause *S. aureus* to evolve differently in the hosts if the organism is to thrive. It was concluded in this study that although there is a relationship between the nucleotide diversity (π) and length of the virulence factor gene, there is no significance between the dN/dS ratio of the virulence factor and its respective nucleotide significance indicating that the genes are instead exerted under positive selection.

The genes encoding for virulence factors were investigated and analysed to determine if they are under positive selection. It was observed that a majority of the genes were indeed under positive selection including *LukS-PV*, *lukF-PV*, *chp*, *sspA*, *vwb*, *lip*, *lip1*, *nucA*, *hysA*, *hysA1*, *aur*, *fnb*, *ebpS*, *bbp*, *entE*, *seb* and. It is also observed that genes *PBP2*, *hlgA*, *hla*, *sspB*, *sspC*, *lipA*, *plc*, *nuc*, *cna*, *sea* and *seh* are undergoing purifying selection.

The *hlyB* gene was under a significant amount of positive selection as observed by the significantly high dN/dS ratio of 8. Beta-hemolysin which is encoded by the *hlyB* gene functions as an exotoxin that is involved in the haemolytic activity on erythrocytes (Kataama *et al.*, 2012). Recent studies have also suggested that the *hla* gene may have an important secondary role in skin colonisation with *in vitro* studies suggesting that colonisation efficiency was 50-fold greater with the presence of the gene than without it (Katayama *et al.*, 2012). In comparison alpha-hemolysin, a pore-forming cytotoxin encoded by the *hla* gene, does not have any secondary role and is instead involved in only erythrocyte lysis. The *hla* gene had a dN/dS ratio of 0.16, concluding that it is not under any kind of positive selection but is instead under purifying selection. This observation concludes that the selective pressure acting on this gene may be due to its secondary function. This phenomenon has altered the fitness of the *S. aureus* and it suggested that it is the driving force of evolution because it has to compete with the micro-flora of the skin if it is to colonise the host and thrive. The *hlyB* gene will therefore be under continuous positive selection. This is further supported by Figure 5.2b. Again

the *hly* was under positive selection and has a dN/dS ratio of 10.5. It is suggested that this gene may be needed for skin colonisation of the *S. aureus* to the host and it is under positive selection because it is competing with other micro-organisms. Armand-Lefere *et al.* (2005) isolated LA-MRSA ST398 from healthy pig farmers and their pigs suggesting that colonisation occurred primarily in the nostrils however the colonisation of skin was not investigated during that study.

Since competing with other micro-organisms can drive selective pressure and cause evolutionary changes, it is not surprising that a majority of the adherence genes are under positive selection. Figure 5.2a shows a cluster of adherence genes that are evolving under positive selection. For instance, adherence genes that encode fibronectin-binding protein (*fnb*), Elastin-binding protein (*ebpS*) and bone- sialoprotein binding protein (*bbp*) have dN/dS ratios of 3.85, 3.56 and 2.83, respectively. This data strongly suggests that competition of other micro-organisms adhering to the host results in selective pressure acting on the gene.

Clumping factor (*clf*), found in a majority of the strains ($n = 54$), binds to different sites on the fibrinogen (*clfA* and *clfB*). It is observed from Figure 5.2a that the *clf* gene has a dN/dS ratio of 0.85, suggesting that the gene may be evolving away from neutrality. *clf* is neutral in the full data set but very positive in the LA data set (Figure 5.2), it is suggested host-pathogen interaction may play a role in this difference because interaction with host tissue plays a critical role in establishing, for example, mastitis infection (Kerro *et al.*, 2002).

Three enterotoxins, *entE*, *seb* and *sek*, dN/dS values of 3.88, 2.03 and 7.7, respectively, which cause sporadic food poisoning outbreaks are released by *S. aureus* strains also show considerable positive selection. The gastrointestinal tract is a very volatile environment to a microorganism due to competition with other bacteria, viruses, archaea and fungi with the human intestine being home to 10^{14} bacterial cells and thousands of strains (Kaiko & Stappenbeck, 2014). It is suggested that *S. aureus* has to compete in the hosts environment and could possibly acquire genetic information by natural genetic transformation which is the uptake of free DNA by bacterial cells, a type of horizontal gene transfer (Lorenz & Wackernagel, 1994).

The human complement system is the first defence against bacteria, once activated it attracts neutrophils to the site of infections and opsonises bacteria to facilitate phagocytosis (Laarman *et al.*, 2011). *Staphylococcus aureus* have successfully developed ways to evade the human complement system. Metalloprotease aureolysin, encoded by the *aur* gene, is a complement inhibitor that prevents phagocytosis of the bacteria and thus ensures survival in the host. It is therefore not surprising that the *aur* gene has a dN/dS ratio of 2.53. Other genes that may be evolving due to selective pressure from the hosts defenses are both components of the PVL toxin, *lukf-pv* and *luks-pv* that have a dN/dS ratio of 1.3 and 2.04 respectively; and von Willebrand factor binding protein which is encoded by the *vwb* gene (dN/dS = 1.86). These secreted toxins and proteins will be considered foreign by the host and will be targeted by the immune response. The *S. aureus* will therefore have to evolve and adapt if they are to survive in the host.

It is concluded that selective pressures act on various genes independently and that the size of the gene does not play a role in the rate of mutation. It is suggested that genes are therefore evolving at different rates as different genes are under different selective pressures. In addition to this, observations from the divergence studies suggest that there is very little variation amongst the hospital-associated, community-associated and livestock associated populations, as seen in Figure 5.5. It is concluded that the three *S. aureus* populations are not significantly differentiated.

Chapter 6:

Population genetics of *S. aureus*

6.1 Introduction

Population genetics is the study of the distribution and change in frequency of alleles within a population and as such sits in the field of evolutionary biology; it examines the genetic variability within a population and its primary purpose is to understand the evolutionary forces that act upon it (Brookfield, 1996). It has been stated that pathogenic bacteria exist as population, with studies suggesting that members of these populations exhibit varying degrees of virulence (Spratt & Maiden, 1999). Many of the genes, as shown by this study, will be polymorphic and as such exist in a number of forms. It is important to understand that the focus of population genetics is to understand the population or species as a whole and not an individual strain.

A software called DnaSP was implemented in this study to determine the level and distribution of genetic polymorphism within natural populations by comparing the respective alleles. In this study virulence factor genes that occur within a population of *S. aureus* were compared. The evolutionary forces which can include mutation, migration, selection and genetic drift are examined by DnaSP to determine the pattern of genetic variation that is being observed and is supported by a number of statistical methods and tests (Rozas *et al.*, 2009). The 'Neutral Theory' of molecular evolution plays a central role in the analysis of genetic data. The Neutral Theory states that the majority of evolutionary change and most of the variation within a species at the molecular level are not caused by natural selection but instead driven by random drift of selectively mutant alleles that are neutral (Crow, 2000). DnaSP was also used to perform statistical neutrality tests on the data to determine if a DNA sequence is evolving randomly (neutrally) or evolving under a non-random process, i.e selection. Neutrality tests applied to the DNA sequences was Tajima's *D* which is computed as the difference between two measures of genetic diversity i.e. the mean number of pairwise differences and the number of segregating sites and Fu & Li's *D* (Tajima, 1989). Another statistical test that examines the polymorphism distributions have been designed, namely the Fu and Li's *D* and extends Tajima's *D* by measuring the differences between the number of mutations appearing only once among the sequences and the total number of mutations (Fu & Li, 1993; Rozas *et al.*, 2009). Other statistical tests, such as the

Ramos-Onsins & Rozas index, R_2 and the raggedness index, r were used to determine the evolutionary history of the species by detecting population growth (Ramos-Onsins & Rozas, 2002).

The use of haplotype networks to see a visual relationship between individual genotypes at the population level was also utilised in this study, by using a software called PopArt (Leigh & Bryant, 2015).

6.2 Aims

The aims of the present study include:

- To quantify the level of DNA polymorphism and subsequent genetic diversity in the *S. aureus* population.
- To investigate the geographical relationship of the *S. aureus* population using haploptype networks
- To investigate the classifications of *S. aureus* (hospital-associated, community-associated and livestock-associated) at a population level and infer a relationship.
- To investigate the diversity of synonymous and nonsynonymous mutations along the length of the gene using sliding window analysis.
- Carry out demographic pairwise analysis of the *S. aureus* virulence factor genes to infer population growth events.

6.3 Materials and Methods

- DnaSP analysis was carried out, in Section 2.15. Initially the sequence alignment of the virulence factor was loaded, as carried out in Section 2.15.1. The polymorphism data, which included the number of variable sites, the haplotype diversity, nucleotide diversity and the tests of neutrality, as carried out in Section 2.15.5 and Section 2.15.3
- Initially the virulence factor NEXUS file was generated in DnaSP, as carried out in Section 2.15.7. The NEXUS file was then formatted in BioEdit, as carried out in Section 2.15.8. And the method for Haplotype network analysis and the creation of haplotype networks using PopArt, as carried out in Section 2.16

6.4 Results

6.4.1 Quantifying DNA polymorphism in the *S. aureus* population

Quantifying the level of DNA polymorphism in the *S. aureus* population was accomplished by measuring the nucleotide diversity per site (π), the number of variable sites (s), the number of haplotypes (h) and also the haplotype diversity (Hd) (Tajima, 1989; Rozas, 2009). This data was produced using DnaSP analysis which is available in full in Appendix 1.

Figure 6.1a is a graph showing the number of variable sites (s) along the \mathcal{X} axis and the nucleotide diversity (π) along the \mathcal{Y} axis. This graph suggest that as the number of variable sites (s) (or polymorphic sites) between the *S. aureus* samples increases so does the nucleotide diversity (π). This demonstrates that there is overall diversity amongst the *S. aureus* population. This is further supported by Figure 6.1b which shows the number of haplotypes (h) along the \mathcal{X} axis and the nucleotide diversity per site (π) along the \mathcal{Y} axis. The extent of nucleotide diversity in the *S. aureus* population is represented in this graph. This graph suggests that the nucleotide diversity varies greatly within the overall population.

However, Figure 6.1c is a graph showing the number of variable sites (s) along the \mathcal{X} axis and the haplotype diversity (Hd) along the \mathcal{Y} axis. This graph shows the trend that as the number of variable sites increases so does the haplotype diversity, although there is a plateau. This plateau is illustrative of gene maintenance in the *S. aureus* population. Since haplotype diversity (Hd) is a measure of the probability that two randomly sampled alleles are different, the plateau suggests that the gene function is being maintained despite the differing number of variable sites (s). These differences in the number of variable sites (s) can be due to the ratio of synonymous and nonsynonymous nucleotide mutations being expressed; synonymous nucleotide substitutions do not alter amino acids sequence, in contrast nonsynonymous nucleotide mutations alter the amino acid sequence of the protein.

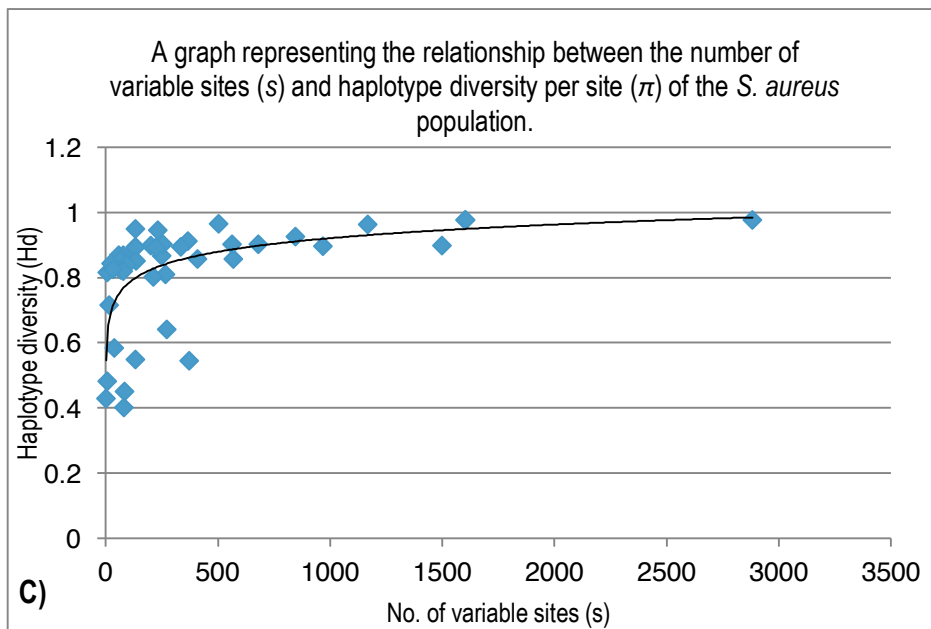
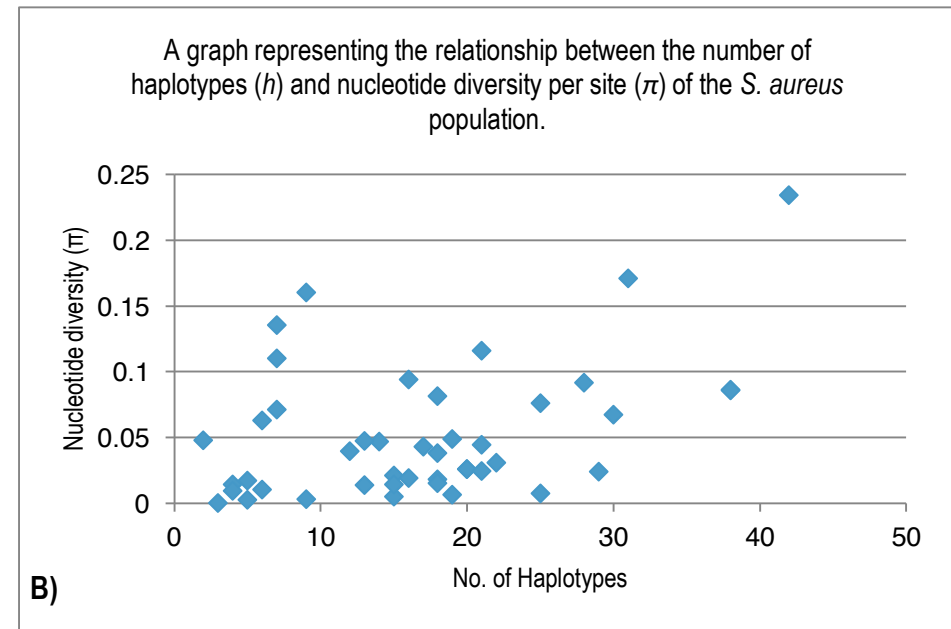
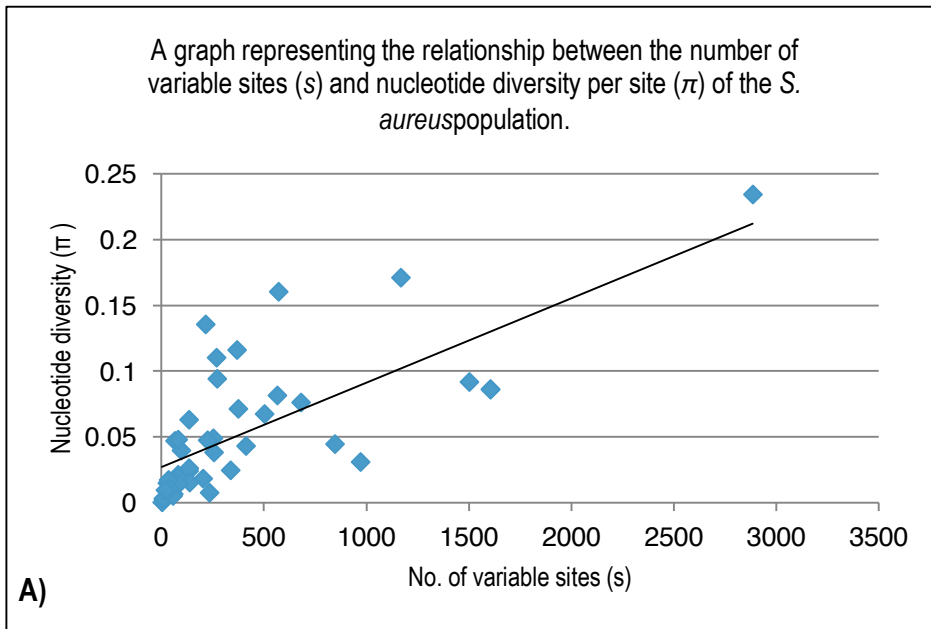


Figure 6.1a is a graph representing the relationship between the number of variable sites and nucleotide diversity per site of the *S. aureus* population. Figure 6.1b is a graph representing the relationship between the number of haplotypes and the nucleotide diversity. Figure 6.1c is a graph representing the relationship between the number of variable sites and haplotype diversity per site of the *S. aureus* populations.

These graphs suggest that although there is observed diversity amongst the *S. aureus* population and its respective genes, the function of the gene is being maintained.

6.4.2 Investigating the geographical relationship of the *S. aureus* population

This study included a large number of *S. aureus* isolates (n=53). Table 6.1 shows the *S. aureus* strains that were used in this study along with countries that they were isolated from. The information for this table was extracted from Table 3.1, found in Section 3.4. It should be noted that it was not possible to determine the country of origin for strain NCTC8325 so it was excluded from this study. It is obvious from Table 6.1 that the *S. aureus* strains from this study have been isolated from all over the world and are therefore a good representation of the strains that have been isolated (n=52). Haplotypes of virulence factor genes shared amongst the *S. aureus* strains were identified using DnaSP, as described in Section 2.15. Haplotypes can be used to compare individuals between populations in order to infer which populations are most closely related with each other.

Haplotype networks to infer geographic relationships were created using data for the *pbp2* and *scn* virulence factor genes. These two genes were used as they were present in the majority of the strains and were therefore a good representation of the wider *S. aureus* population. Figure 6.2a and Figure 6.2b are haplotype networks created using PopArt, as described in Section 2.16. The areas of the circles are proportional to the number of individual strains harbouring the specific haplotype and the different colours represent the geographic location of the isolate, this is identified by the key. The numbers in the brackets adjacent to the lines linking the circles represent the mutational steps between the haplotypes. It is interesting to note that there is no set pattern of the distribution in haplotypes across the entire network for either the *PBP2* and *SCIN* genes. Haplotype network analysis based on the *PBP2* and *SCIN* gene show little association between the multiple countries, with the majority of the haplotypes being unique to the country of isolation, however *SCIN* Hap_9 is shared by isolates from nine different countries. It is important to point out that the mutational steps between the haplotypes, which are represented by the numbers in the brackets do not differ by a great amount, with the majority of the haplotypes only differing by 1 - 5 mutational steps. The exception to this is *SCIN* Hap_16 which differs by 60 mutational steps and *PBP2* Hap_24 which differs by 185 mutational steps. Despite these difference in the mutational steps between the haplotypes, it is important to note that the virulence factor gene is still recognised as either a *SCIN* or *PBP2* gene.

Strain	Isolated
04-02981	Koln, Germany
08BA0216	Canada
11819-97	Denmark
502A	USA
6850	USA
71193	Dominican Republic
AS3	Belgium
B51	UK
B53	Belgium
Bmb9393	Rio de Janeiro, Brazil
COL	Colindale, England
CN1	Seoul, South Korea
CA-347	California, USA
ECT-R2	Ostergotland, Sweden
ED133	France
ED98	Ireland
H-EMRSA-15	England
H0 5096 0412	Suffolk
IP70	UK
JKD6008	New Zealand
JKD6159	Australian
JH9	Maryland, USA
JH1	Maryland, USA
LGA251	England
M1	Denmark
Mu50	Japan
Mu3	Japan
MW2	North Dakota, USA
MSSA476	Oxford, UK
MSHR1132	Northern Territory, Australia
MRSA252	UK
M013	Taiwan
Newman	UK
N315	Japan
NCTC8325	Not Known.
OP100	UK
RF122	Ireland
S62	Belgium
SA40	Taiwan
SA957	Taiwan
ST398 (S0385)	Germany
ST228 isolate 18385	Lausanne, Switzerland
ST228 isolate 16125	Lausanne, Switzerland
ST228 isolate 16035	Lausanne, Switzerland
T0131	China
TW20	London, UK
UA-S391_USA300	USA
USA300-ISMMS1	USA
USA300_FPR3737	USA
USA300_TCH1516	San Fransisco, USA
VC40	USA
XN108	China
Z172	Taiwan

Table 6.1: A table showing the *Staphylococcus aureus* strains that were used in this study along with the countries that they were isolated from. The information was extracted from Table 3.1.

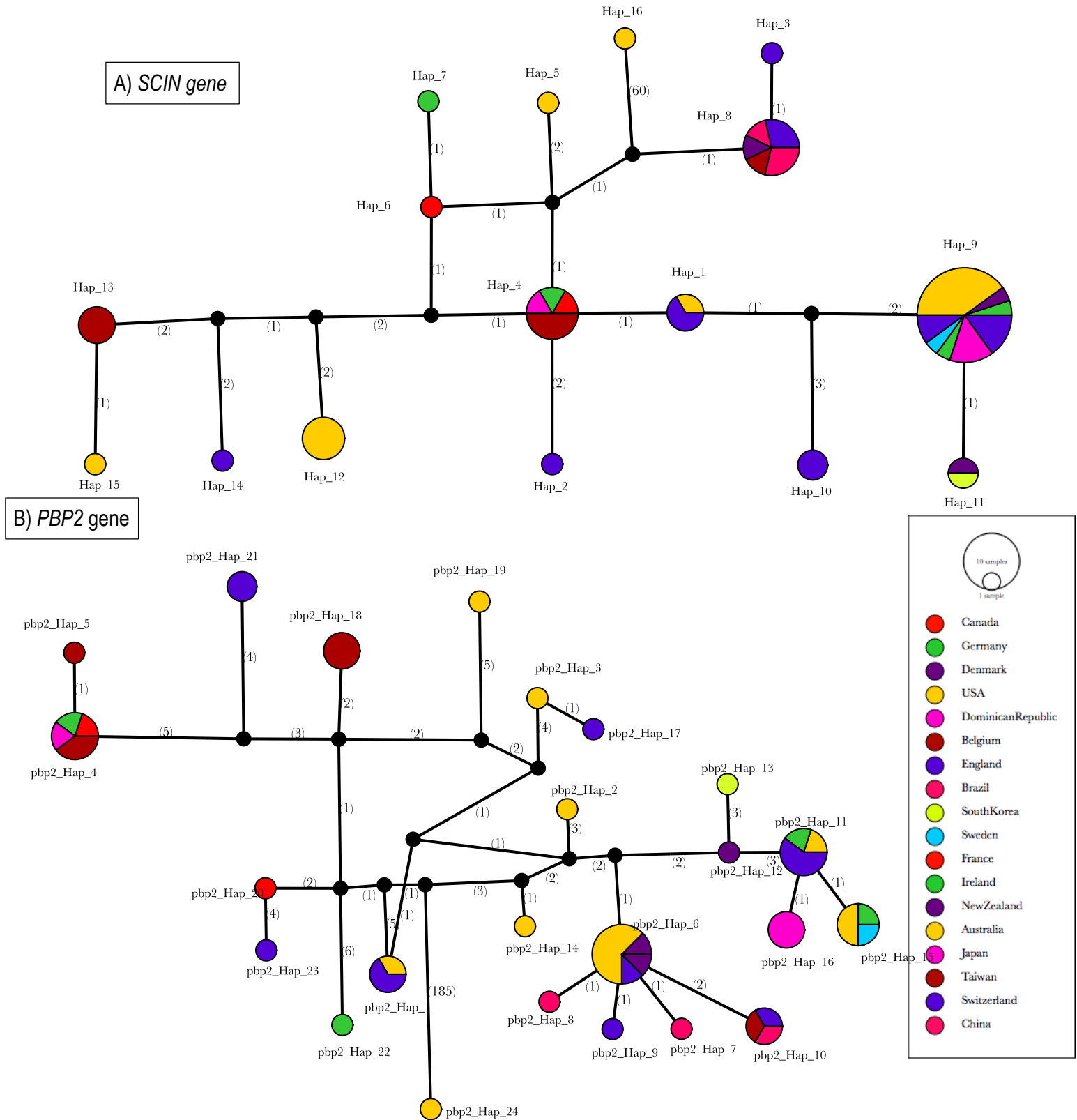


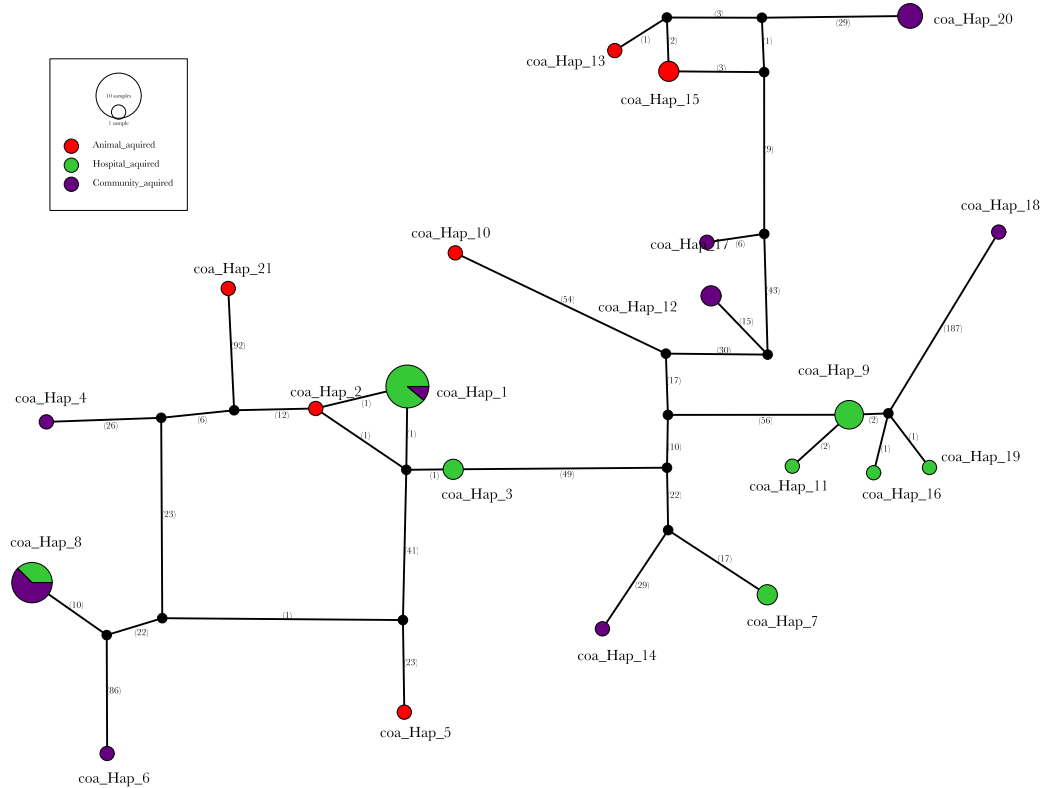
Figure 6.2a & Figure 6.2b: Are haplotype network representing the *SCIN* & *PBP2* virulence factor genes, created using the TCS methodology. The areas of the circle are proportionate to the number of individual strains harbouring the specific haplotype and the different colours represent the country of origin of the isolate. The numbers in brackets represent the mutational steps between the haplotypes and the small black circles indicate missing haplotypes.

6.4.3 Investigating the relationship of the *S. aureus* population

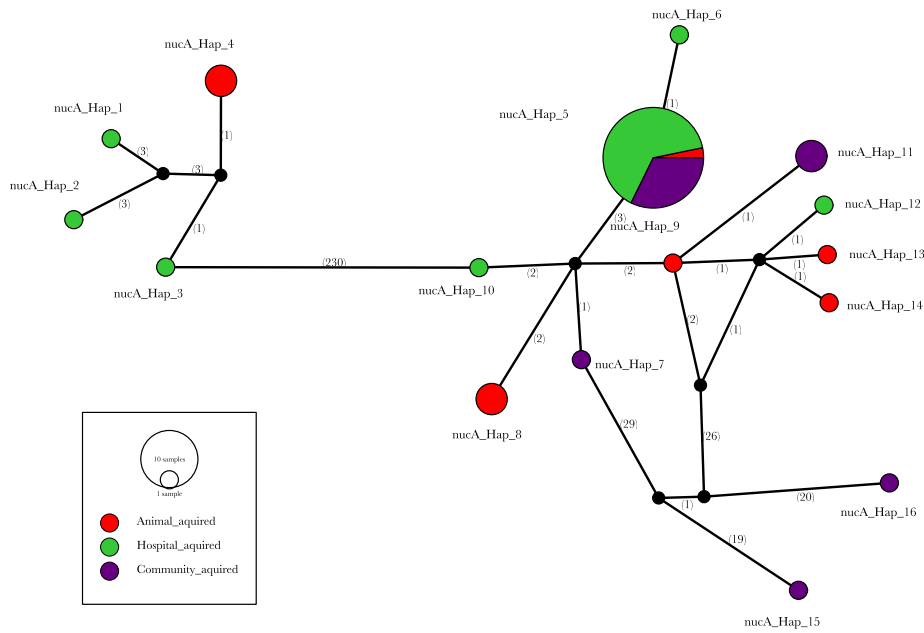
Haplotype networks were also created for *S. aureus* strains that represented hospital-associated, community-associated and livestock-associated to determine if any relationship existed between this classification in the wider population. Figures 6.3a-e are haplotype networks that were created for a number of *S. aureus* virulence factor genes: a) Staphylocoagulase (*coa*), b) PVL toxin (*lukF-PV*), c) von Willebrand factor-binding protein (*vwb*), d) enterotoxin B (*seb*) and e) thermonuclease (*nucA*). These haplotype networks represent *S. aureus* virulence factor genes that have a variety of different functions. These particular virulence factor genes were also chosen as they represented virulence factor genes that were exhibiting the greatest amount of divergence between the given populations i.e hospital-associated, community-associated and livestock-associated; as demonstrated in Section 5, Figure 5.6a-e. The 'populations' for these virulence factor were designated as either hospital-associated, community-associated or livestock-associated and are shown on the haplotypes networks as green, purple and red circles, respectively. The areas of the circle are proportionate to the number of individual strains harbouring the specific haplotype. Where more than one haplotype was observed to be in the same 'population'; the circle is divided into a pie diagram where the slices correspond to the proportion of the individual strains from a given population. Again, the numbers in the brackets represent the mutational steps between the haplotypes.

It is interesting to note from all of these haplotypes networks that there is a limited relationship between the 'populations' defined in literature. This has been suggested because the 3 classifications do not cluster together into separate clades, instead the hospital-associated, community-associated and livestock-associated haplotypes are present across the whole population. The *coa* Hap_1 and Hap_8 share haplotypes that are present in both hospital-associated (HA) and community-associated (CA) populations. This relationship is also present in *vwb* Hap_4, *seb* Hap_2 and *lukF-PV* Hap_1. The *nucA* Hap_3 and *lukF-PV* Hap_6 haplotypes are present in all three *S. aureus* populations, including livestock-associated (LA). It is also observed that the mutational steps between the haplotypes vary greatly. For instance, *coa* Hap_1 which a haplotype for both HA and CA strains differs by only 1 mutational step from Hap_2 which is apparently specific to LA strains. Another example is the LA specific *seb* Hap_4 and Hap_6 haplotypes which appear to be significantly different, however, they only differ by 6 mutational steps.

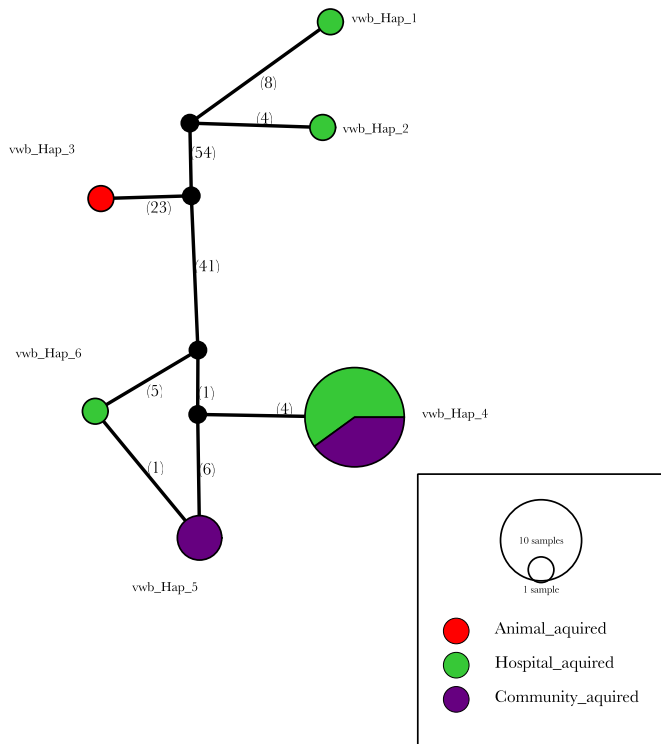
The final example is the *nucA* Hap_12, Hap_13, Hap_14, Hap_11 and Hap_9 haplotypes which represent individual HA, CA and LA strains that only differ by a few mutational steps. It is therefore reasonable to conclude that the classification into the populations' implied by the literature is artificial.



Figures 6.3a: A haplotype network that was created for the Staphylocoagulase (*coa*) virulence factor gene, created using the TCS methodology. The areas of the circle are proportionate to the number of individual strains harbouring the specific haplotype and the different colours represent the country of origin of the isolate. The numbers in brackets represent the mutational steps between the haplotypes and the small black circles indicate missing haplotypes.



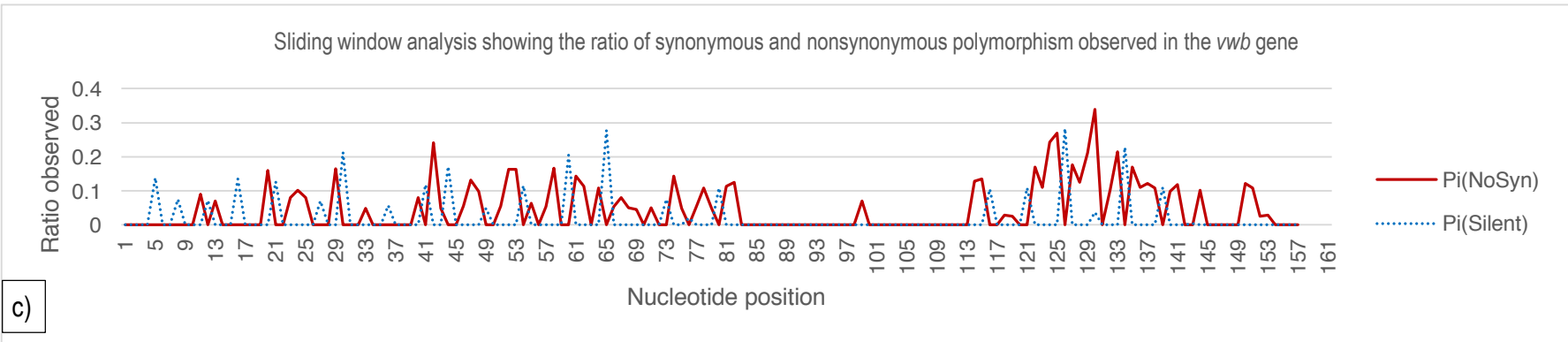
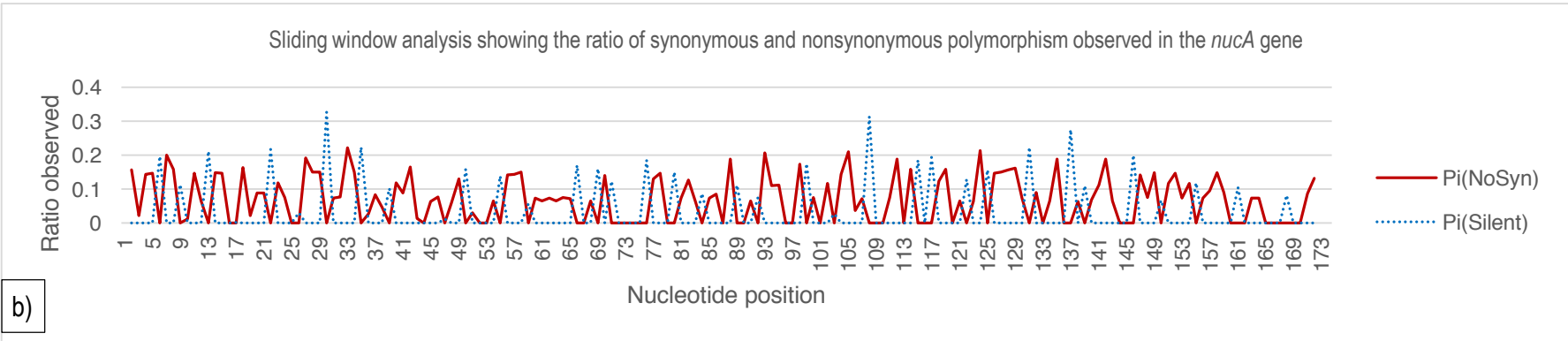
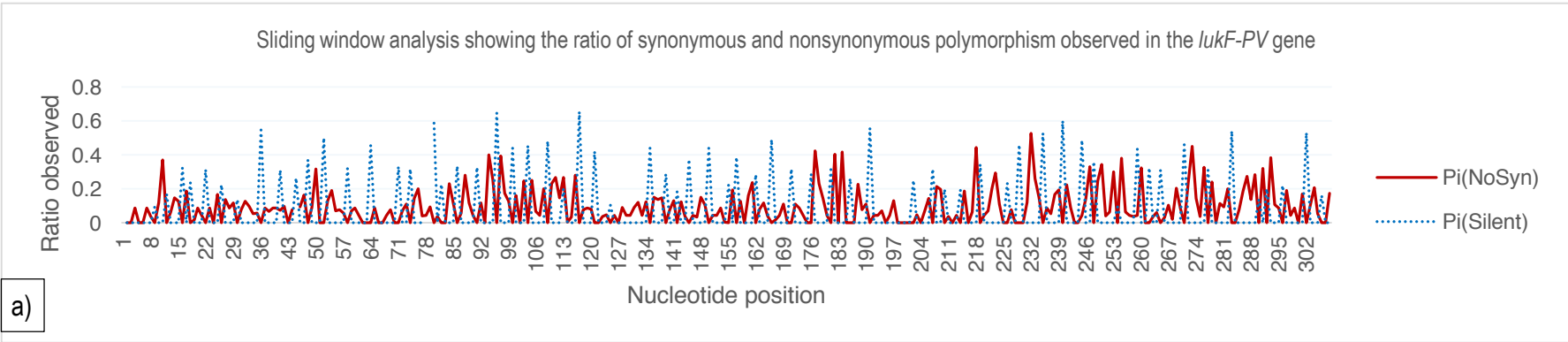
Figures 6.3b: A haplotype network that was created for the Thermonuclease (*nucA*) virulence factor gene, created using the TCS methodology. The areas of the circle are proportionate to the number of individual strains harbouring the specific haplotype and the different colours represent the country of origin of the isolate. The numbers in brackets represent the mutational steps between the haplotypes and the small black circles indicate missing haplotypes.



Figures 6.3c: A haplotype network that was created for the Thermonuclease (*nucA*) virulence factor gene, created using the TCS methodology. The areas of the circle are proportionate to the number of individual strains harbouring the specific haplotype and the different colours represent the country of origin of the isolate. The numbers in brackets represent the mutational steps between the haplotypes and the small black circles indicate missing haplotypes.

6.4.4 Sliding window analysis of synonymous and nonsynonymous mutations

The sliding window analysis is a useful tool to uncover synonymous and nonsynonymous rate variation along a sequence that can detect regions under differing selection (Schmid & Yang, 2008). Llopart & Aguade (1999) have previously used this analysis to identify the variations in synonymous and nonsynonymous within genes in the *Drosophila* species. The sliding window analysis method was used to compare the ratio of synonymous and non synonymous polymorphisms, for a number of virulence factor genes: a) *lukF-PV*, b) *nucA*, c) *vwb*, d) *coa* and e) *seb*, Figures 6.4a-e. The nonsynonymous mutations are highlighted in red and the synonymous or silent mutations are highlighted in blue. In Table 5.1 (Section 5.7) it was observed that a number of virulence factor genes were under positive selection, this was identified by a ratio $\omega > 1$. It is interesting to note that although *lukF-PV* ($\omega = 1.308$), *nucA* ($\omega = 1.315$), *vwb* ($\omega = 1.857$), *col* ($\omega = 4.9$) and *seb* ($\omega = 2.029$) are all evolving under positive selection there is still a high level of diversity of synonymous and nonsynonymous rates being observed along the gene, as shown in Figures 6.4a-e. Examining these graphs illustrates the trend of mutations along the length of the gene and suggest that there is diversity, however it is also apparent that the majority of polymorphisms being observed in these genes are mostly non synonymous. Some parts of the genes are also free of any mutations (synonymous or nonsynonymous), this is particularly apparent in region 85 – 113, in the *vwb* gene and 225 – 253, in the *coa* genes.



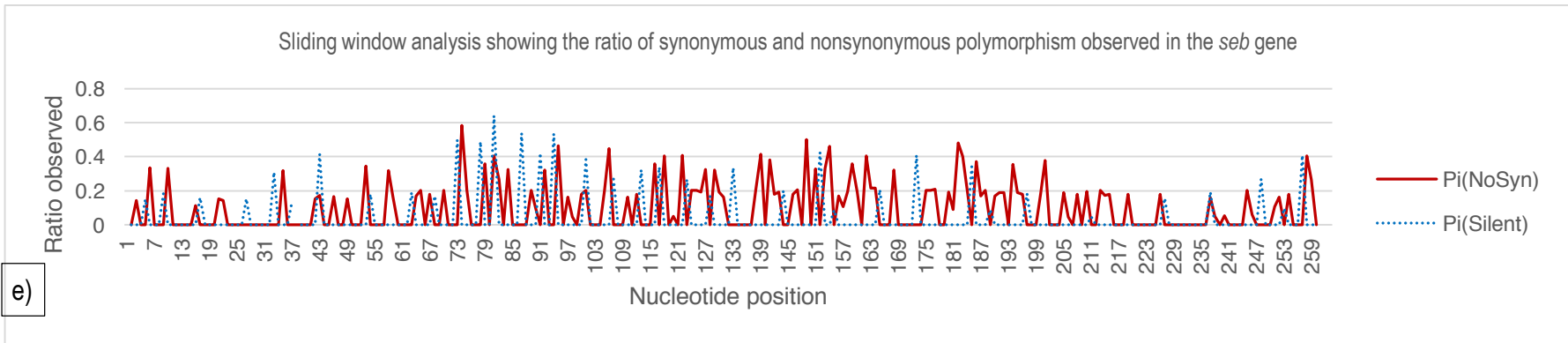
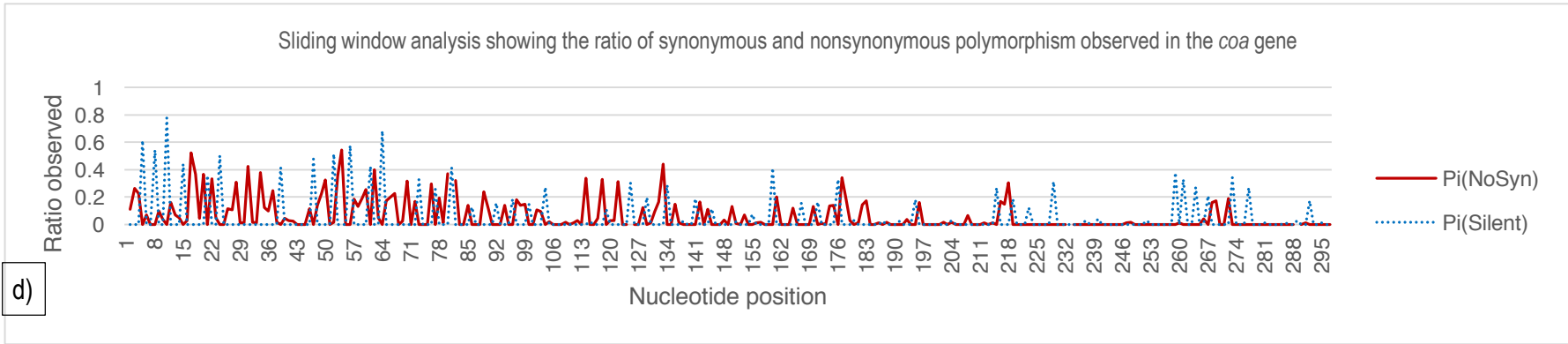


Figure 6.4a-e: Sliding window analysis showing the ratio of synonymous and nonsynonymous polymorphism observed in a number of *S. aureus* virulence factor genes: a) *lukF-PV*, b) *nucA*, c) *vwb*, d) *coa* and d) *seb*. The nonsynonymous mutations are highlighted in red and the synonymous or silent mutations are highlighted in blue.

6.3.5 Pairwise comparisons of *S. aureus* virulence factor genes

The distribution of pairwise differences in a set of sequences is an important tool for molecular evolutionary studies (Mousset *et al.*, 2004). DnaSP was used to generate the theoretical population growth distribution by measuring the number of pairwise differences between pairs of sequences (Roza, 2009). The mismatch distribution is a graphic way of visualising the signature of an expansion event. A frequency graph of pairwise differences between alleles that exhibit unimodal properties are indicative of populations that have recently passed through a recent demographic expansion (Excoffier, 2003). In comparison a bimodal or multimodal distribution frequency of sharp peaks are found in populations that have been constant over time and are therefore considered to be nonsignificant (Sousa *et al.*, 2012). Studies have also shown that a unimodal distribution for the pairwise analysis is suggestive of a process of expansion which is supported by neutrality tests that indicate a deviation from neutrality (Sodre *et al.*, 2012). To test the goodness-of-fit for the observed distribution of frequency curves the raggedness index, r and Ramos-Onsins & Rozas index, R_2 were also calculated, and if significant were indicative of population growth (Ramos-Onsins & Rozas, 2002). To determine whether the sequences also conformed to the expectations of neutrality of evolution, Tajima's D statistical test was used (Tajima, 1989). Previously, The Tajima's D test showing negative values has indicated deviations from neutrality and has suggested recent population expansion (Sharma *et al.*, 2013).

Figure 6.5a-e shows the pairwise distribution for a number of *S. aureus* virulence factor genes a) *lukF-PV*, b) *nucA*, c) *vwb*, d) *coa* and d) *seb*. All the pairwise distribution analysis are showing a multimodal distribution and as such it is indicative that the population have not passed through a recent expansion and are instead showing that the populations have been constant over some time. This is supported by the the r and R_2 index, shown in Table 6.2, which support this observation. However, conflicting evidence from the both the majority of the Tajima's D and Fu & Li's D negative values deviating away from neutrality are supporting the hypothesis that the whole sample has recently passed through a population expansion event. This is observed for the *lukF-PV*, *nucA*, *vwb*, and *coa* genes. The *seb* gene has a Tajima's D of 2.59 ($p < 0.01$)

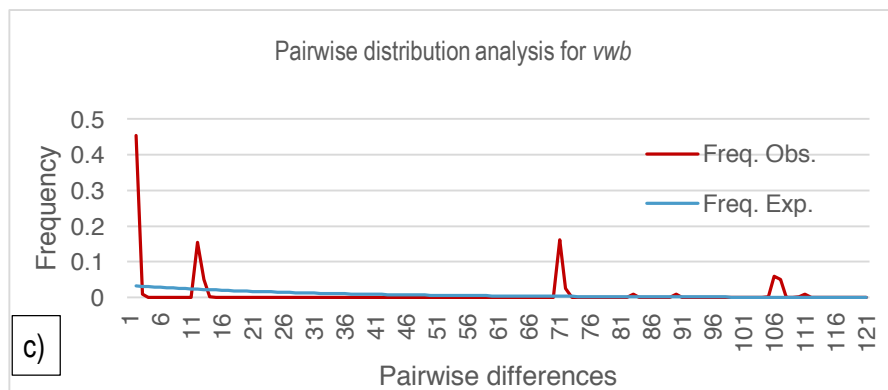
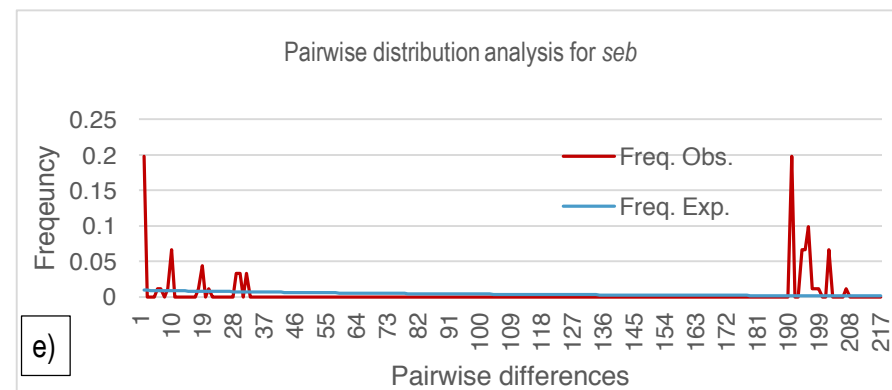
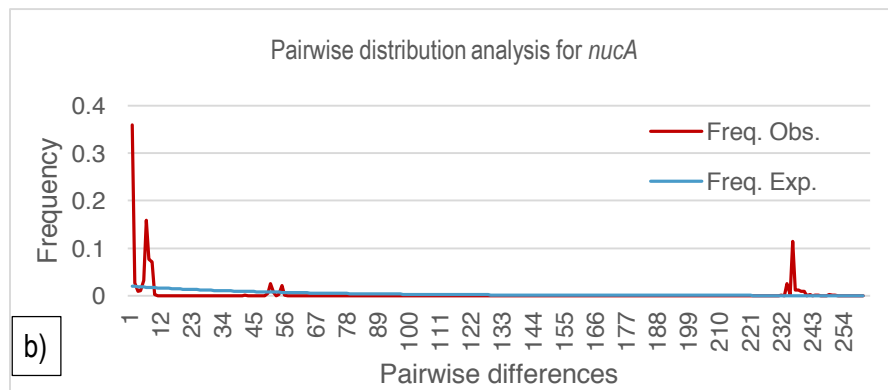
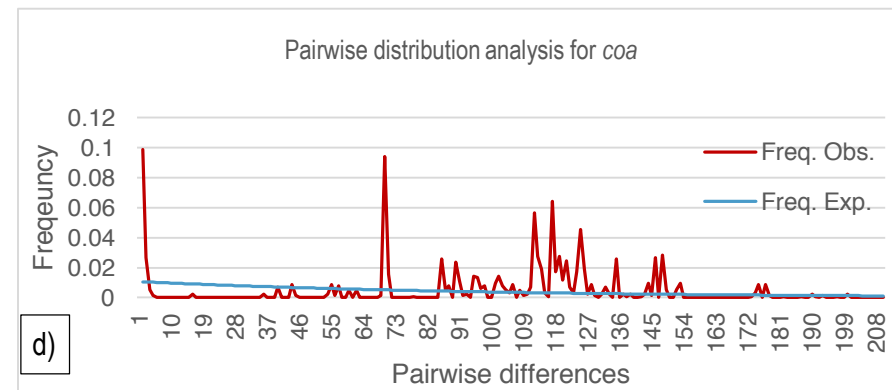
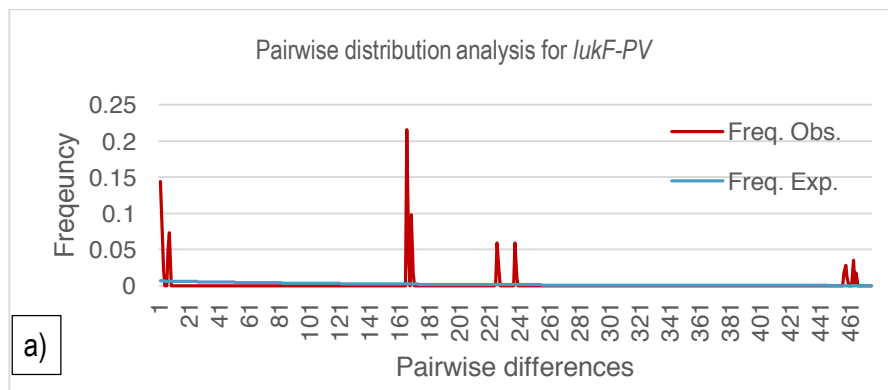


Figure 6.5a-e: Pairwise distributio analysis for a number of *S. aureus* virulence factor genes: a) *lukF-PV*, b) *nucA*, c) *vwb*, d) *coa* and d) *seb*. All analysis are showing a multimodal distributon and as such are not significant.

Virulence factor	<i>lukF-PV</i>	<i>nucA</i>	<i>vwb</i>	<i>coa</i>	<i>seb</i>
Tajima's <i>D</i>	-0.38234	-0.8496	-0.71047	-0.11428	2.5904 (p<0.01)
Fu & Li's <i>D</i>	2.04025 (p<0.02)	0.88805	1.39427 (p<0.05)	-1.42261	1.36471 (p<0.5)
<i>r</i>	0 (p<0.000)	0.927 (p<0.036)	0.253 (p<0.088)	0.205 (p<0.007)	0.01 (p<0.040)
<i>R2</i>	0 (p<0.000)	0.937 (p<0.184)	0.765 (p<0.225)	0.864 (p<0.15)	0.85 (p<0.2440)

Table 6.2: A table showing the statistical tests of neutrality and the goodness-of-fit indexes for population expansion for *S. aureus* virulence factors.

6.5 Discussion:

This study has suggested that the observed diversity amongst the *S. aureus* populations varies greatly, this trend is shown in Figure 6.1a-b. Furthermore, although there is genetic diversity it is shown that the virulence factor genes are still recognized as their respective genes and the population is being maintained, as shown in Figure 6.1c. This diversity is also confirmed by the use of geographic haplotypes that were created for the *scin* and *pbp2* virulence factors that were present in the majority of *S. aureus* strain in this study. The haplotypes for both genes are promiscuously distributed and show no geographic clustering. Haplotype networks were also created for the *S. aureus* strain 'populations' that were defined in literature, namely hospital-associated, community-associated and livestock-associated. It was noted that only a limited relationship was observed with the majority of the haplotypes not clustering into clades or any defined pattern. There were even occurrences of the 3 classifications sharing a haplotype, denoted as a pie chart with the three colours in Figure 6.3b and 6.3e. This was observed in LukF-PF Hap_6 and nucA Hap_5 demonstrates sharing of the haplotypes between hospital-associated, community-associated and livestock-associated strains. This was further evidence that these classifications have become blurred over time and do not represent genuine divisions in the wider *S. aureus* population; this was also evidenced in the phylogenetic trees that also showed a mixed distribution of clades for the HA, CA and LA strains.

The sliding window analysis of the virulence factor genes also showed a high level of diversity along the nucleotide sequence. It was concluded that the five virulence factors that were analysed (*LukF-PV*, *nucA*, *vwb*, *coa* & *seb*) were under positive selection and this was further confirmed by the greater frequency of nonsynonymous mutations than synonymous mutations present in the sliding window analysis (Figure 6.4a-e). Recent studies by Agashe *et al.* (2016) have however suggested that synonymous mutations can also evolve under extremely strong positive selection, this needs further analysis as the mechanisms for this were unknown. The pairwise comparisons for *S. aureus* virulence genes showed some conflicting evidence. On the one hand, pairwise analyses were showing a multimodal distribution which is indicative of a population that have not passed through a recent expansion event, however on the other hand, the tests for neutrality, Tajima's *D* and Fu & Li's *D* negative values has indicated deviations from neutrality, which suggests a recent population expansion. If *S. aureus* were evolving neutrally, it

would mean that they are evolving randomly, however since they are evolving away from neutrality and under a non-random process, it is suggested that selective forces are acting on the population. This is further evidence of positive selection and local adaptation, and this has been demonstrated throughout this thesis and is further supported by the array of evidence that is presented in this chapter, including the lack of relationship in the haplotype networks that show notable diversity and no relationship between the hospital-associated, community-associated and livestock-associated.

In conclusion, because the selected *S. aureus* population lack a geographic structure, shown by the lack of clustering in the haplotypes, but a high diversity It is therefore highly suggestive that these *S. aureus* strains are under evolutionary processes acting at a local level, and the bacteria are adapting and evolving at a local level through positive selection.

7.0 General discussion and future work

7.1 General discussion

The aim of this study was to initially investigate the three classifications of *S. aureus* that been defined in literature, namely hospital-associated (HA), community-associated (CA) and livestock-associated (LA) and to determine if this classification was also recognised at a genomic level. As stated in Section 1.8, MRSA initially occurred only in healthcare-settings so they were subsequently defined as hospital-associated (HA). MRSA were then being isolated from healthy persons in the community who had no links with healthcare facilities, these were classified as community-associated (CA). More recently, a third group of MRSA have been identified. This group has emerged from a novel zoonotic reservoir; since farm animals such as pigs were implicated, this group were classified as livestock-associated (LA) MRSA. Fifty-three *S. aureus* strains (Table 3.1 and 3.2) have been used in this study, which includes previously sequenced and published strains obtained through the NCBI database. A number of novel strains were also gifted from two collaborators; Prof. Valerie Edward-Jones (Manchester Metropolitan University, UK) provided isolates from hospital patients and Dr Patrick Butaye (The Veterinary & Agrochemical Research Centre, Belgium) provided strains from a number of farm animals (cattle, chicken & swine) and they were subsequently sequenced using the Ion Torrent Next Generation sequencing protocol. A comprehensive literature review was then undertaken on these strains and an array of virulence factor genes which are produced by *S. aureus* to facilitate pathogenesis were identified (Table 3.5). A major emphasis in the literature has been placed on *S. aureus* classification into HA, CA and LA strains. The authors of the papers where the respective strain was studied almost always concluded what the classification of the strain was (i.e HA, CA OR LA). It was hypothesised that the presence/ absence of virulence factor genes that are expressed on the genomes of these *S. aureus* may explain the virulence that is associated with the different classes.

In Chapter three, the virulence factor genes from each isolate of *S. aureus* was extracted, if present, to determine if there was a difference in the virulence genes associated with the three classes of *S. aureus*. This

study has suggested that livestock-associated strains that cause infection to humans seem to be subtly different from the other strains, but overall the majority of the virulence genes are found within all strains. This study showed that the *cna* adhesin gene was present in 3 LA strains that cause outbreaks in humans, but was absent from the strains that are restricted to animals. It was therefore suggested that the presence of this gene may possibly contribute to infections in humans. The presence of the Panton-Valentine Leucocidin (PVL) in MRSA is associated with an increased virulence in humans and is typically found in community-associated strains. Functional PVL consists of two components encoded by the *lukS-PV* and *lukF-PV* genes, thus both genes are needed to establish PVL as a virulence factor (Rahman *et al.*, 1991). This study has confirmed that the USA300 strains contain both functional genes, this has been reported before. (Highlander *et al.*, 2007). However, this study has also confirmed that the novel LA strains also contain both functional genes. This study concludes that livestock-associated strains possess the genes necessary to produce PVL. The finding drawn here support the conclusions by Peeters *et al.* (2015) who found that livestock-associated ST80 strains carried the Panton-Valentine leukocidin gene. It was also noted in this study that not all isolates of *S. aureus* encoded the *mecA* gene but all of the isolates had the presence of the *PBP2A* gene. This was an interesting trend because it suggests that all isolates of *S. aureus* had the presence of the penicillin binding protein 2 gene, which is needed for β -lactam resistance; however, it also suggested the presence of divergent *mecA* genes that were not investigated in this study (Garcia-Alvarez *et al.* (2011).

In chapter 4, a maximum-likelihood phylogenetic analysis was carried out to determine if a molecular relationship was observed between the classifications of *S. aureus* denoted in the literature (HA, CA & LA). It was interesting to note that there were some similarities but not enough to warrant the usage of the exclusive terms hospital-associated, community-associated and livestock-associated. Many of the strains that formed clades did so because of other reasons, for instance, three strains that have been designated

community-associated (SA40, SA957 & M013) were also all Taiwanese strains, and a geographical relationship existed between them.

Another interesting relationship was between the cluster of livestock-associated strains that cause disease in humans (B53, S62, AS3, O8B02176, ST398 & 71193); although, not much information is known about the novel strains from this cluster (B53, S62 & AS3). The novel strains were isolated from the Belgium, however, it is not known if they are from the same farm or region of Belgium. Hospital strains (XN108, Z172, Bmb9393, T0131, JKD6008, TW20) that have been disseminated across the world did cluster into a clade that seem to show no geographical relationship, the only obvious relationship was the fact that they had all been isolated from hospitals. It is important to remember that many of the hospital-associated *S. aureus* were among the first group of strains that were isolated. It is therefore suggested that the use of the same antibiotics across the different hospitals around the world, represents a common factor that is possibly acting as a selective pressure on the bacteria. This was further supported by the differentiation observed between the phylogenetic trees that were produced. The virulence factor genes were subjected to phylogenetic analysis one at a time and the subsequent trees that were produced, varied greatly. This was suggestive of the fact that different selective pressures may be acting on the genes which are therefore evolving at different times. Although it is also possible that some virulence factor genes are also gained by horizontal transfer (Brody *et al.*, 2008), therefore whilst some virulence genes will differentiate between two strains, the ones gained by horizontal transfer will not. It was concluded that a relationship did not exist between the classifications of *S. aureus* that have been given in literature is no longer relevant.

The suggestion of different selective pressures acting on different genes which are evolving at different times was investigated in Chapter 5. It has been shown that *S. aureus* is relatively common commensal of the human flora; however, it can also be an opportunistic pathogen that can cause disease

(Sousa & Lenastre, 2004). In order for *S. aureus* to survive in the host, it must adapt to any selective pressures such as changes to the hosts biology, the hosts immune system and any antibiotic therapy that may be administered. A study by Kennedy *et al.* (2008) demonstrated that subtle nucleotide polymorphisms have a profound affect on the virulence of staphylococcal strains. It was observed that the highly virulent USA300 isolates may have resulted in a recent clonal expansion which resulted in the emergence of virulent isolates that were the consequence of DNA polymorphisms and not due to the acquisition of new or novel virulence factor genes (Kennedy *et al.*, 2008). This study has concluded that genes under positive selection have been found in the majority of the virulence factor categories, for instance; virulence genes, adhesin gene, enterotoxin genes, enzymes and other secreted protein genes. This include; *lukS-PV*, *lekF-PV*, *chp*, *sspA*, *vwb*, *lip*, *lip1*, *nucA*, *hysA*, *hysA1*, *aur*, *fnb*, *ebpS*, *bbp*, *entE*, *seb* and *tst*. It was further concluded that selective pressures act on various genes independently, with the rate of mutation not correlating with the length of the gene sequence. It was suggested that genes are therefore evolving at different rates because different genes are under different pressures. It was also observed that there was very little divergence or variation amongst the hospital-associated, community-associated and livestock-associated strains. This supports the observed lack of HA, CA and LA specific clades in the phylogenetic analyses and again reiterated the conclusion that the classification that has been denoted in the literature may be artificial.

In Chapter 6, the population genetics of *S. aureus* was investigated to determine the genetic diversity of the population, to investigate a geographical relationship between the population and also to investigate the classification that has been denoted in the literature. It was concluded that the observed diversity amongst the *S. aureus* population varied greatly, however, despite this diversity the gene function was being maintained. A geographical linkage was also investigated, and it was concluded that haplotypes for *S. aureus* genes (*pbp2* & *scin*) were promiscuously distributed and showed no geographic clustering. It was also further concluded; with the aid of haplotype networks, that the 3 classifications denoted in the literature are artificial

and do not represent genuine divisions, this was initially concluded in Chapter 4, based upon the phylogenetic analyses and Chapter 5, based upon polymorphism and divergence studies.

In conclusion, an extensive understanding of genomic variation at the single nucleotide level is extremely important to determine the events that define bacterial virulence. *S. aureus* populations are highly diverse and have a broad host range. It has previously been suggested that pathogens with a broad host range have the greatest probability of being transmitted to humans (Taylor et al., 2001). It is obvious from the data that *S. aureus* are present in a broad range of farm animals as well as other animals that are in close contact to humans. It is also evident from the data that the *S. aureus* population is highly diverse. This study employs the computational aspects of biology to determine the genes that are subject to positive and selective pressures. It is suggested that *S. aureus* are under evolutionary processes acting at a local level and that they are adapting and evolving at a local level through positive selection, which may be dependent on the hosts environment.

7.2 Future work

A future aim of this study would be to evaluate a greater number of *S. aureus* strains that have been sequenced. It should be noted, that when this thesis was undertaken, all of the *S. aureus* strains that had complete assembled genome sequences published in GenBank were utilised in this study and at the time there were approximately 50 *S. aureus* strains. Undertaking a current search of *S. aureus* in GenBank would yield in excess of 116 fully sequenced strains. Also, the collaborators; Prof. Valerie Edward-Jones (Manchester Metropolitan University, UK) provided isolates from hospital patients and Dr Patrick Butaye (The Veterinary & Agrochemical Research Centre, Belgium) provided strains from a number of farm animals, did in fact provide many more strains that were not utilised in this study. In total, 20 strains were provided of which only 6 were sequenced using Ion Torrent, due to the time and financial constraints of the study. It is suggested that a greater number of strains are investigated in the future to confirm and further the findings observed.

It should also be noted that currently only contigs were used in this study from the sequenced strains and that the novel strains of *S. aureus* were not full assembled. This was due to the fact that full coverage of the contigs was not observed, instead there was some gaps in the genome when compared to the existing published *S. aureus* strains. Any future works that may arise from this study should use one high sensitivity Ion Torrent DNA chip per isolate, and not three as has been currently done. Although it has been enough for the purpose of contigs, it should be noted that one high sensitivity DNA chip per sample will yield a greater number of contigs and that the assembly of the entire genome would be possible.

Although a relatively high number of *S. aureus* virulence factor genes were identified (43) from a broad range of categories, it should be noted that a greater number of virulence factor genes are available for investigation. Figure 1.2 identified approximately 150 genes that may be involved in the virulence of *S. aureus*. Investigating a greater number of virulence factor genes across the whole genome would be a good step to determine if similar findings are observed.

This study utilised the power of computational biology to determine if the virulence factor genes are present or absent at a genotypic level. The genotypic properties were investigated, however, it would also be beneficial to investigate if the gene are expressed phenotypically. This can be determined by initially translating the amino acid sequences to see if the protein function was maintained. Following on from the DnaSP work the range of nonsynonymous mutations in the virulence genes could be investigated to determine any potential in the disruption of the protein structures that may lead to altered function. This would require analysis of 3-dimensional structures. The auto-translation of the nucleotide sequences using EMBOSS translate tool (of both the reference and identified query sequence) would be initially undertaken; the altered amino acid sites would be identified and these would be mapped on a 3D structure of the reference protein to determine if they affect the functional sites (active sites/ binding sites of the protein).

A large number of polymorphisms that were identified within the virulence factor haplotypes were single nucleotide polymorphisms. It would be interesting to investigate how these changes to the nucleotide sequence ultimately change the protein it is encoding. It was stated that the gene function was being maintained, this can be further investigated by utilising protein modelling to determine if the functions of the protein remains the same or differs. Undertaking culture based experiments to confirm if the bacterium is in fact producing the products of the gene, for example, is the bacterium with the collagen binding gene actually producing collagen.

Bibliography

Agashe D, Sane M, Phalnikar K, Diwan GD, Habibullah A, Martinez-Gomez NC, Sahasrabudhe V, Polachek W, Wang J, Chubiz LM & Marx. Large-effect beneficial synonymous mutations mediate rapid and parallel adaptation in a bacterium. *The Society of Molecular Biology and Evolution*. 2016; 33(6): 1542 - 1553.

Al Laham N, Mediavilla JR, Chen L, Abdelateef N, Elamreen RA, Ginocchio CG, Pierard D, Becker K, Kreiswirth BN. MRSA Clonal Complex 22 Strains Harboring Toxic Shock Syndrome (TSST-1) Are Endemic in the Primary Hospital in Gaza, Palestine. *PLoS One*. 2015; 10(3):

Altman DR, Sebra R, Hand J, Attie O, Deikis G, Carpini KWD, Patel G, Rana M, Arvelakis A, Grewal P, Dutta J, Rose H, Shopsin B, Daefler S, Schadt E, Kasarskis A, van Bakel H, Bashir A & Huprikar S. Transmission of Methicillin of Methicillin-Resistant *Staphylococcus aureus* via Deceased Donor Liver Transplantation Confirmed by Whole Genome Sequencing. *American Journal of Transplantation*. 2014; xx: 1 - 5.

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ. Basic local alignment tool. *The Journal of Molecular Biology*. 1990; 215(3): 403 - 410.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. 1997; 25: 3389 - 3402.

Amaral MM, Coelho LR, Flores RP, Souza RR, Silvia-Carvalho MC, Teixeira LA, Ferreira-Carvalho BT & Figueiredo MS. The predominant variant of the Brazilian epidemic clonal complex of Methicillin-Resistant *Staphylococcus aureus* has an enhanced ability to produce biofilm and adhere to and invade airway epithelial cells. *The Journal of Infectious Diseases*. 2005; 192: 801 - 810.

Ammerlaan HSM, Kluytmans JAJW, Wertheim HFL, Nouwen JL and Bonten MJN. Eradication of Methicillin-Resistant *Staphylococcus aureus* Carriage: A Systemic Review. *Healthcare Epidemiology*. 2009; 48: 922 - 930.

Andrey DO, Jousselein A, Villanueva M, Renzoni A, Monod A, Barras C, Rodriguez N & Kelley WL. Impact of the Regulators *SigB*, *Rot*, *SarA* and *sarS* on the Toxic Shock *Tst* Promoter and TSST-1 Expression in *Staphylococcus aureus*. *PLoS One*. 2015; DOI: 10.1371/journal.pone.0135579.

Applebaum P C. The emergence of vancomycin-intermediate and vancomycin-resistant *Staphylococcus aureus*. *Clinical Microbiology and Infection*. 2006; 12(1): 16 - 23.

Armand-Lefevre L, Ruimy R & Andremont A. Clonal comparison of *Staphylococcus aureus* isolates from healthy pig farmers, human controls, and pigs. *Emerging Infectious Diseases*. 2005; 11(5): 711 - 714.

Atkins KL, Burman JD, Chamberlain ES, Cooper JE, Poutrel B, Bagby S, Jenkins ATA, Feil EJ, Van den Elsen JMH. *S. aureus* IgG-binding proteins SpA and Sbi: Host specificity and mechanisms of immune complex formation. *Molecular Immunology*. 2008; 45: 1600 - 1611.

Baba T, Takeuchi F, Kuroda M, Yuzawa H, Aoki K, Oguchi A, Nagai Y, Iwama N, Asano K, Naimi T, Kuroda H, Cui L, Yamamoto K & Hiramatsu K. Genome and virulence determinants of high virulence community-acquired MRSA. *The Lancet*. 2002; 359: 1819 - 1827.

Baba T, Bae T, Schneewind O, Takeuchi F & Hiramatsu K. Genome sequence of *Staphylococcus aureus* strain Newman and comparative analysis of Staphylococci genomes: Polymorphisms and evolution of two major pathogenicity islands. *Journal of Bacteriology*. 2008; 190(1): 300 - 310. *Journal of Global Antimicrobial Resistance*. 2016; 6: 95 – 101.

Bal AM, Coomb GW, Holden MTG, Lindsay JA, Nimmo GR, Tattavin P & Skov RL. Genomic insights into the emergence and spread of international clones of healthcare-, community- and livestock-associated methicillin-resistant *Staphylococcus aureus*: Blurring of the traditional definitions.

Balaban N & Rasooly a. Staphylococcal Enterotoxins. *International Journal of Food Microbiology*. 2000; 61(1): 1 - 10.

Bansal AK. Bioinformatics in microbial biotechnology - a mini review. *Microbial Cell Factories*. 2005; 4:19; doi:10.1186/1475-2859-4-19.

Baptiste KE, Williams K, Williams NJ, Wattret A, Clegg PD, Dawson S, Corkill JE, O'Neil T & Hart CA. Methicillin-resistant staphylococci in companion animals. *Emerging Infectious Diseases*. 2005; 11(12): 1942 - 1944.

Becker K, Friedrich AW, Lubritz G, Weilert M, Peters G & von Eiff C. Prevalence of Genes Encoding Pyrogenic Toxic Superantigens and Exfoliative among Strains of *Staphylococcus aureus* Isolated from Blood and Nasal Specimens. *The Journal of Clinical Microbiology*. 2003; 41(4): 1434 - 1439.

Ben Zakour NL, Sturdevant DE, Even S, Guinane CM, Barbey C, Alves PD, Cochet MF, Gautier M, Otto M, Fitzgerald JR & Loir YL. Genome-Wide Analysis of Ruminant *Staphylococcus aureus* Reveals Diversification of the Core Genome. *Journal of Bacteriology*. 2008; 190(19): 6302 - 6317.

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J and Sayers EW. GenBank. *Nucleic Acids Research*. 2012; 41.

Bhat M, Dumortier C, Taylor BS, Miller M, Vasquez G, Yunen J, Brudney K, Sanchez JE, Rodriguez-Taveras C, Rojas R, Leon P & Lowy FD. *Staphylococcus aureus* ST398, New York City and Dominican Republic. *Emerging Infectious Diseases*. 2009; 15(2): 285 - 297.

Bhatta DR, Cavaco LM, Nath G, Kumar K, Gaur A, Gokhale S & Bhatta DR. Association of Pantone Valentin Leukocidin (PVL) genes with methicillin resistant *Staphylococcus aureus* (MRSA) in Western Nepal: a matter of concern for community infection (a hospital based prospective study). *BMC Infectious Diseases*. 2016; 16: 199.

Bhakti S & Trantum-Jensen J. Alpha-Toxin of *Staphylococcus aureus*. *Microbiology Reviews*. 1991; 55(4): 733 - 751.

Bjerketorp J, Nilsson M, Ljungh A, Flock JI, Jacobsson K & Frykberg L. A novel von Willebrand factor binding protein expressed by *Staphylococcus aureus*. *Microbiology*. 2002; 148: 2037 - 2044.

Bjerketorp J, Jacobsson K & Frykberg L. The von Willebrand factor-binding protein (vWbp) of *Staphylococcus aureus* is a coagulase. *FEMS Microbiology Letters*. 2004; 234: 309 - 314.

Blanco R, Tristan A, Ezpeleta G, Larsen AR, Bes M, Etienne J, Cisterna R & Laurent F. Molecular epidemiology of Pantón-Valentine Leukocidin-positive *Staphylococcus aureus* in an autochthonous population. *Journal of Clinical Microbiology*. 2011; 49(1): 433 - 436.

Boyle-Vavra S & Daum RS. Community-acquired methicillin-resistant *Staphylococcus aureus*: the role of Pantón-Valentine leukocidin. *Laboratory Investigation*. 2007; 87: 3- 9.

Brody T, Yavatkar AS, Lin Y, Ross J, Kuzin A, Kundu M, Fann Y & Odenwald WF. Horizontal gene transfer link a human MRSA pathogen to contagious bovine mastitis bacteria. *PLoS ONE*. 2008; 3(8): e3074.

Burlak C, Hammer CH, Robinson MA, Whitney AR, McGavin MJ, Kreiswirth BN & DeLeo FR. Global analysis of community-associated methicillin-resistant *Staphylococcus aureus* exoproteins reveals molecules produced *in vitro* and during infection. *Cellular Microbiology*. 2007; 9(5): 1172 - 1190.

Burnside K, Lembo A, de los Reyes M, Iliuk A, BinhTran, Connelly JE, Lin WJ, Schmidt BZ, Richardson AR, Fang FC, Tao WA & Rajagopal L. Regulation of Hemolysin Expression and Virulence of *Staphylococcus aureus* by a Serine/Threonine Kinase and Phosphate. *PLoS*. 2010; 5(6): e11071.

Burman JD, Leung E, Atkins KL, O'Seaghdha MN, Lango I, Bernado P, Bagby S, Svergun DI, Foster TJ, Isenman DE & Van den Elsen JMH. Integration of Human Complement with Sbi, a Staphylococcal Immunoglobulin-binding Protein. *The Journal of Biological Chemistry*. 2008; 283(25): 17579 - 17593.

Busier S, Smaczny C, Von Mallinckrodt C, Kraal A, Ackerman H, Brade V & Wichelhaus TA. Prevalence and Clinical Significance of *Staphylococcus aureus* Small-Colony Variants in Cystic Fibrosis Lung Disease. *The Journal of Clinical Microbiology*. 2007; 45(1): 168 - 172.

Casey JA, Cosgrove SE, Stewart WF, Pollak J & Schwartz BS. A population-based study of the epidemiology and clinical features of methicillin-resistant *Staphylococcus aureus* infection in Pennsylvania, 2001 - 2010. *Epidemiology and Infection*. 2013; 141(6): 116 - 1179.

CDC. Four paediatric deaths from community-acquired methicillin-resistant *Staphylococcus aureus*: Minnesota and North Dakota, 1997-1999. *Morbidity and Mortality Weekly Report*. 1999; 48: 707 - 10. <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm4832a2.htm#tab1>

Chang S, Sievert DM, Hageman JC, Boulton ML, Tenover FC, Downes FP, Shah S, Rudrik JT, Pupp GR, Brown WJ, Cardo D & Fridkin SK. Infection with Vancomycin-Resistant *Staphylococcus aureus* containing the vanA resistance gene. *The New England Journal of Medicine*. 2003; 348(14): 1342 - 1347.

Chen Y, Chatterjee SS, Porcella SF, Yu YS & Otto M. Complete Genome Sequence of a Pantón-Valentine Leukocidin-Negative Community-Associated Methicillin-Resistant *Staphylococcus aureus* Strain of Sequence type 72 from Korea. *PLoS One*. 2013; 8(8): e72803

Chen FJ, Lauderdale TL, Wang LS, & Huang IW. Complete Genome Sequence of *Staphylococcus aureus* Z172, a Vancomycin-Intermediate and Daptomycin-Nonsusceptible Methicillin-Resistant Strain Isolated in Taiwan. *Genome Announcements*. 2013; 1(6): e01011-13.

Chen CJ, Unger C, Hoffmann W., Lindsay JA, Huang YC & Götz F. Characterisation and Comparison of 2

Distinct Epidemic Community-Associated Methicillin-Resistant *Staphylococcus aureus* Clones of ST59 Lineage. *PLoS One*. 2013; 8(9): e63210.

Chu C, Yu C, Lee Y & Su Y. Genetically divergent methicillin-resistant *Staphylococcus aureus* and sec-dependant mastitis of dairy goats in Taiwan. *BMC Veterinary Research*. 2012; 8(39): e1746-6148/8/39.
Chua K, Seeman T, Harrison PF, Davies JK, Coutts SJ, Chen H, Haring V, Moore R, Howden BP & Stinear TP. Complete genome sequence of *Staphylococcus aureus* Strain JKD6159 a unique Australian Clone of ST398-IV Community Methicillin-Resistant *Staphylococcus aureus*. *Journal of Bacteriology*. 2010; 192(20): 5556 - 5557.

Cowan S T, Shaw C & Williams R E O. Type strain for *Staphylococcus aureus* Rosenbach. *Journal of General Microbiology*. 1954; 10: 174-176.

Cosgrove SE, Sakoulas G, Perencevich EN, Schwaber MJ, Karchmer A & Carmeli Y. Comparison of mortality associated with methicillin-resistant and methicillin-sensitive *Staphylococcus aureus* bacteraemia: a meta analysis. *Clinical Infectious Disease*. 2003; 36: 53 - 59.

Costa MOC, Beltrame CO, Ferreira FA, Botelho AMN, Lima NCB, Souza RC, de Almeida LGP, Vasconelos ATR, Nicolas MF & Figueiredo AMS. 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*. 2013; 1(4): e00576-13.

Crombe F, Argudin MA, Vanderhaeghen W, Hermans K, Haesebrouck F & Butaye P. Transmission dynamics of methicillin-resistant *Staphylococcus aureus* in pigs. *Frontiers in Microbiology*. 2013; 4(57): doi: 10.3389/fmicb.2013.00057.

Cuny C, Kuemmerle J, Stanek C, Willey B, Strommenger B & Witte W. Emergence of MRSA infections in horses in a veterinary hospital: Strain characterisation and comparison with MRSA from humans. *Eurosurveillance*. 2006; 11(1-3): 44 - 47.

Cuny C, Kock R & Witte W. Livestock associated MRSA (LA-MRSA) and its relevance for humans in Germany. *International Journal of Medical Microbiology*. 2013; 303: 331 - 337.

Darling ACE, Mau B, Plattner FR & Perna NT. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Research*. 2004; 14: 1394 - 1403.

Daum L, Rodriguez JD, Worthy SA, Ismail NA, Omar S, Dreyer AW, Fourie PB, Hoosen AA, Chambers JP & Fischer GW. Next-generation Ion Torrent sequencing of drug resistance mutations in *Mycobacterium tuberculosis* strains. *Journal of Clinical Microbiology*. 2012; 50(12): 3831 - 3837.

Deogo KO, van Dijk JE & Nederbuigt. Factors involved in the early pathogenesis of bovine *Staphylococcus aureus* mastitis with emphasis on bacterial adhesion and invasion. A review. *Veterinary Quarterly*. 2002; 24(4): 181 - 198.

de Haas CJC, Veldkamp KE, Peschel A, Weerkamp F, van Wamel JB, Heezius ECJM, Poppelier MJJGP, Van Kessel & van Strijp JAG. Chemotaxis Inhibitory Protein of *Staphylococcus aureus*, a Bacteria Anti-inflammatory Agent. *The Journal of Experimental Medicine*. 2004; 199(5): 687 - 695.

de Jong MA, Wahlberg N, van Eijk M, Brakheid PM & Zwaan BJ. Mitochondrial DNA signature for range-wide populations of *Bicyclus anynana* suggests a rapid expansion from recent refugia. *PLoS one*. 2011; 6(6); e21385.

de Vries AS, Leshner L, Schlievert PM, Rogers T, Villaume LG, Danila R & Lynfield R. Staphylococcal toxic shock syndrome 2000 - 2006: Epidemiology, clinical features, and molecular characteristics. *PLoS One*. 2011; 6(8): e22997.

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM & Gascuel O. [phylogeny.fr](http://www.phylogeny.fr): robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*. 2008; 36: 465 - 469.

Dereeper A, Audic S, Claverie JC & Blanc G. BLAST EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evolutionary Biology*. 2010; 10:8. doi:10.1186/1471-2148-10-8

Deurenberg RH & Stobberingh EE. The evolution of *Staphylococcus aureus*. *Infection, Genetics and Evolution*. 2008; 8: 747 - 763.

Deurenberg RH, Vink C, Kalenic S, Freidrich AW, Bruggeman CA & Stobberingh EE. The molecular of methicillin-resistant *Staphylococcus aureus*. *Clinical Microbiology and Infections*. 2006; 13: 222 -235.

Diep A B, Gill S R, Chang R F, Phan T H, Chen J H, Davidson M G, Lin F, Lin J, Carleton H A, Mongodin E F, Sensabaugh G F & Perdreau-Remington F. Complete genome sequence of USA300, an epidemic clone of community-acquired methicillin-resistant *Staphylococcus aureus*. *The Lancet*. 2006; 367: 731 - 39.

Dinges MM, Erwin PM & Schlievert PM. Exotoxins of *Staphylococcus aureus*. *Clinical Microbiology Reviews*. 2000; 13(1): 16 - 34.

Downer R, Roch F, Park WP, Mecham RP & Foster TJ. The Elastin-binding Protein of *Staphylococcus aureus* (EbpS) Is Expressed at the Cell Surface as an Integral Membrane Protein and Not as a Cell Wall-associated Protein. *The Journal of Biological Chemistry*. 2002; 277(1): 243 - 250.

Drapeau GR, Boily Y & Houmard J. Purification and Properties of an Extracellular Protease of *Staphylococcus aureus*. *The Journal of Biological Biochemistry*. 1972; 247(20): 6720 - 6726.

Duthie E S & Lorenz L L. Staphylococcal Coagulase: Mode of action and antigenicity. *Journal of General Microbiology*. 1952; 6: 95 - 107.

ECDC-1. European Centre for Disease Prevention and Control. Antimicrobial resistance surveillance in Europe 2009. Annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net). Stockholm, Sweden: ECDC; 2010.

ECDC-2. European Centre for Disease Prevention and Control. Antimicrobial resistance surveillance in Europe 2014 Annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net). Stockholm, Sweden: ECDC; 2015.

Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004; 5:113. doi:10.1186/1471-2105-5-113.

Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 2004; 32(5): 1792 - 1797.

Edgeworth J D, Yadegarfar G, Pathak S, Batra R, Cockfield J D, Wyncoll D, Beale R & Lindsay J A. 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 bacteraemia. *Clinical Infectious Diseases*. 2007; 44: 493 - 501.

Excoffier L. Patterns of DNA sequence diversity and genetic structure after range expansion: lessons from the infinite-island model. *Molecular Ecology*. 2004; 13: 853 – 864.

Fisher JF & Mobashery S. 2007 Astellas USA Foundation Award: Host-Guest Chemistry of the Peptidoglycan. *Journal of Medical Chemistry*. 2010; 8(53): 4813 - 4829.

Fitzgerald JR, Monday SR, Foster TJ, Bohach GA, Hartigan PJ, Meaney WJ, & Smyth CJ. Characterisation of a putative pathogenicity island from bovine *Staphylococcus aureus* encoding multiple superantigens. *Journal of Clinical Microbiology*. 2001; 183: 63 - 70.

Fitzgerald JR, Sturdevant DE, Mackie SM, Gill SR & Musser JM. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *PNAS*. 2001; 98(15): 8821 - 8826.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougerty BA & Merrick JM. Whole-genome random sequencing and assembly of *Haemophilus influenzae*. *Science*. 1995; 269(5223): 496 - 512.

Fleming A. On the antibacterial action of cultures of a penicillim, with special reference to their use in the isolation of *B. influenzae*. *The British Journal of Experimental Pathology*. 1929; 10: 780 - 790.

Foster T and Hook M. Surface protein adhesions of *Staphylococcus aureus*. *Trends in Microbiology*. 1998; 6(12): 484 - 488.

Frank DN, Feazel LM, Bessesen MT, Price CS, Janoff EN & Pace NR. The Human Nasal Microbiota and *Staphylococcus aureus* carriage. *PLoS One*. 2010. 5(5)

Fraunholz M, Bernhardt J, Schuldes J, Daniel R, Hecker M & Sinah B. Complete Genome Sequence of *Staphylococcus aureus* 6850, a Highly Cytotoxic and Clinically Virulent Methicillin-Sensitive Strain with Distant Relatedness to Prototype Strains. *Genome Announcements*. 2013; 1(5): e00775-13.

Frazer BW, Lynn J, Charlebois ED, Lambert L, Lowery D & Perdreau-Remington F. High prevalence of methicillin-resistant *Staphylococcus aureus* in emergency department skin and soft tissue infections. *Annals of Emergency Medicine*. 2003; 47(1): 196 - 203.

Fu YX & Li WE. Statistical Tests of Neutrality of Mutations. *Genetics*. 1993; 133: 693 - 709.

García-Álvarez L, Holden MT, Lindsay H, Webb CR, Brown DF, Curran MD, Walpole E, Brooks K, Pickard DJ, Teale C, Parkhill J, Bentley SD, Edwards GF, Girvan EK, Kearns AM, Pichon B, Hill RLR, Larsen AR, Skov RL, Peacock SJ, Maskell DJ & Holmes. Methicillin-resistant *Staphylococcus aureus* with a novel *mecA* homologue in human and bovine populations in the UK and Denmark: a descriptive study. *The Lancet Infectious Diseases*. 2011; 11(8): 595 – 603.

Gaucher EA, Kratzer JT & Randall RN. Deep phylogeny - How a tree can help characterise early life on earth. *Cold Spring Harbour Perspectives in Biology*. 2010; 2: a002238.

Gill SR. *Staphylococcus aureus*, Strain COL: Product Information Sheet for NR-45906. The following reagent was provided by the Network on Antimicrobial Resistance in *Staphylococcus aureus* (NARSA) for distribution by BEI Resources, nIAD, NIH: *Staphylococcus aureus*, Strain COL, NR-45906. <https://www.beiresources.org/ProductInformationSheet/tabid/784/Default.aspx?doc=47458.pdf>. (Accessed May 2016).

Gill SR, Fouts DE, Acrher GL, Mongodin EF, DeBoy RT, Ravel J, Paulson IT, Kolonay JF, Brinkac L, Beaman M, Dodson RJ, Daugherty SC, Madupu R, Angiuoli SV, Durkin AS, Haft DH, Vamathevan J, Khouri H, Utterback T, Lee C, Dimitrov G, Jiang L, Qin H, Weidman J, Tran, Kang K, Hance IR, Nelson KE & Fraser CM. Insights on evolution of virulence and resistance from the complete genome analysis of an early Methicillin-Resistant *Staphylococcus aureus* strain and a biofilm-producing Methicillin-Resistant *Staphylococcus epidermidis* strain. *Journal of Bacteriology*. 2005; 187(7): 2426-22438.

Golding GR, Bryden L, Nevett PN, McDonald RR, Wong R, Graham MR, Tyler S, Van Domselaar G, Mabon P, Kent H, Butaye P, Smith TC, Kadlec K, Schwarz S, Weese SJ & Mulvey MR. Whole-Genome Sequence of Livestock-Associated ST398 Methicillin-Resistant *Staphylococcus aureus* Isolated from Humans in Canada. *Journal of Bacteriology*. 2012; 194(23): 6627 - 6628.

Goodall-Copestake WP, Tarling GA & Murphy EJ. On the comparison of population-level estimates of haplotype and nucleotide diversity: a case study using the gene *cox1* in animals. *Heredity*. 2012; 109: 50 - 56.

Goss CH & Muhlebach MS. Review: *Staphylococcus aureus* and MRSA in cystic fibrosis. *Journal of Cystic Fibrosis*. 2011; 10, 298 - 306.

Gouaux E, Hobaugh M & Song L. α -Hemolysin, γ -hemolysin, and leukocidin from *Staphylococcus aureus*: Distant in sequence but similar in structure. *Protein Science*. 1997; 6(12): 2631 - 2635.

Gregory PD, Lewis RA, Curnock P and Dyke KGH. Studies of the repressor (Bl_l) of beta-lactase synthesis in *Staphylococcus aureus*. *Molecular Microbiology*. 1997; 24(5): 1025 - 1037.

Grinberg A, Hittman A, Leyland M, Rogers L & Quesne BLE. Epidemiological and molecular evidence of a monophyletic infection with *Staphylococcus aureus* causing a purulent dermatitis in a dairy farmer and multiple cases of mastitis in his cows. *Epidemiology Infection*. 2004; 132: 507 - 513.

Guinane CM, Ben Zakour NL, Tormo-Mas MA, Weinert LA, Lowder BV, Cartwright RA, Smyth DS, Smyth

CJ, Lindsay JA, Gould KA, Witney A, Hinds J, Bollback JP, Rambaut A, Pendadés JR, Fitzgerald JR: Evolutionary genomics of *Staphylococcus aureus* reveals insight into the origin and molecular basis of ruminant host adaptation. *Genome Biology and Evolution*. 2010; 2: 454–466.

Guindon S & Gascuel O. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*. 2003; 52(5): 696 - 704.

Haggar A, Hussain M, Lonnie H, Herrmann M, Norrby-Teglund A & Flock JI. Extracellular Adherence Protein from *Staphylococcus aureus* Enhances Internalisation into Eukaryotic Cells. *Infection and Immunity*. 2003; 71(5); 2310 - 2317.

Hall TA. BioEDIT: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 1999; 41: 95 - 98.

Hamed DM & Youssef AI. Clinical features of methicillin-resistant *Staphylococcus aureus* (MRSA) infection in rabbits and its zoonotic potentials. *Pakistan Journal of Nutrition*. 2013; 12(3): 244 - 249.

Haran KP, Godden SM, Boxrud D, Jawahir S, Bender JB & Sreevatsan S. Prevalence and characterisation of *Staphylococcus aureus*, including methicillin-resistant *Staphylococcus aureus*, Isolated from bulk tank milk from Minnesota dairy farms. *Journal of Clinical Microbiology*. 2012; 50(3): 688 - 695.

Harraghy N, Hussain M, Haggar A, Chavakis T, Sinah B, Hermann M & Flock JI. The adhesive and immunomodulating properties of the multifunctional *Staphylococcus aureus* protein Eap. *Microbiology*. 2003; 149: 2701 - 2707.

Hernandez D, van der Mee-Marquet N, Kluytmans J, Donnio PY, Quentin R, Corvaglia AR & Francois P. Whole-Genome Sequence of *Staphylococcus aureus* ST398 Strains of Animal Origin. *Genome Announcements*. 2013; 1(5): e00689-13

Herold BC, Immergluck LC, Marana MC, Lauderdale DS, Gaskin RE, Boyle-Vavra S, Leitch CD & Daum RS. Community-acquired methicillin-resistant *Staphylococcus aureus* in children with no identified predisposing risk. *The Journal of the American Medical Association*. 1998; 279: 593 - 598.

Highlander S K, Hulten K G, Qin X, Jiang H, Yerrapragada S, Mason E O, Shang Y, Williams T M, Fortunov R M, Liu Y, Igbowli O, Petrosino J, Tirumalai M, Uzam A, Fox G E, Cardenas A M, Muzny D M, Hemphill L, Ding Y, Dugan S, Blyth P R, Buhay C J, Dinh H H, Hawes A C, Holder M, Kovar C L, Lee S L, Liu W, Nazareth L V, Wang Q, Zhou J, Kaplan S L & Weinstock G M. Subtle genetic changes enhance virulence of methicillin resistant and sensitive *Staphylococcus aureus*. *BMC Microbiology*. 2007; 7: 99.

Hiramatsu K, Hanaki H, Ino T, Yabuta K, Oguri T & Tenover FC. Methicillin-resistant *Staphylococcus aureus* clinical strain with reduced vancomycin susceptibility. *Journal of Antimicrobial Chemotherapy*. 1997; 40: 135 - 146.

Hiramatsu K, Aritaka N, Hanaki H, Kawasaki S, Hosoda Y, Hori S, Fukuchi Y & Kobayashi I. Dissemination in Japanese hospitals of strains of *Staphylococcus aureus* heterogeneously resistant to vancomycin. *The Lancet*. 1997; 350: 1670 - 1673.

Holden MTG, Feil EJ, Lindsay JA, Peacock SJ, Day NPJ, Enright MC, Foster TJ, Moore CE, Hurst L, Atkin R, Barron A, Bason N, Bentley SD, Chillingworth T, Churcher C, Clark L, Corton C, Cronin A, Doggett J, Dowd L, Feltwell T, Hance Z, Harris B, Hauser H, Holroyd S, Jagels K, James KD, Lennard N, Line A, Mayes B, Moule S, Mungall K, Ormond D, Quail MA, Rabinowitsch E, Rutherford K, Sanders M, Sharp S, Simmonds M, Stevens K, Whitehead S, Barrell BG, Spratt BG & Parkhill J. Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. *Proceeding National Academy of Sciences USA*. 2004; 101: 9786 – 9791.

Holden M T G, Lindsay J A, Corton C, Quail M A, Cockfield J D, Pathak S, Batra R, Parkhil J, Bently & Edgeworth J D. 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*. 2010; 192(3): 888 - 892.

Holden MTG, Hsu LY, Kurt K, Weinert LA, Mather AE, Harris SR, Strommenger B, Layer F, Witte W, de Lencaster H, Skov R, Westh H, Zemlickova H, Coombs G, Kearns AM, Hill RLR, Edgeworth J, Gould I, Gant V, Cooke J, Edwards GF, McAdam PR, Templeton KE, McCann A, Zhou Z, Castillo-Ramirez S, Feil EJ, Hudson LO, Enright MC, Balloux F, Aanensan DM, Spratt BG, Fitzgerald JR, Parkhill J, Achtman M, Bentley & Nubel U. A genomic portrait of the emergence, evolution and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Research*. 2013; 23: 653 - 664.

Holt DC, Holden MTG, Tong SYC, Castillo-Ramirez S, Clarke L, Quail MA, Currie BJ, Parkhill J, Bentley SD, Feil EJ & Giffard PM. A very early branching *Staphylococcus aureus* lineage lacking the carotenoid pigment Staphyloxanthin. *Genome, Biology & Evolution*. 2011; 3: 881 - 895.

Howden BP, Johnson PDR, Ward PB, Stinear TP, & Davies JK. Isolates with Low-Level Vancomycin Resistance Associated with Persistent Methicillin-Resistant *Staphylococcus aureus* Bacteraemia. *Antimicrobial Agents and Chemotherapy*. 2006; 50(9): 3039 - 3047.

Howden BP, Seeman T, Harrison PF, McEvoy CR, Stanton JL, Rand CJ, Mason CW, Jenson SO, Firth N, Davies JK, Johnson PDR & Stinear TP. Complete Genome Sequence of *Staphylococcus aureus* Strain JKD6008, an ST239 Clone of Methicillin-Resistant *Staphylococcus aureus* with Intermediate-Level Vancomycin Resistance. *Journal of Bacteriology*. 2010; 92(21): 5848 - 5849.

HPA-1. Health Protection Agency. Annual counts and rates of MRSA bacteraemia (April 2007 - March 2011). Available from: <https://www.gov.uk/government/statistics/mrsa-bacteraemia-annual-data>. 2011. (Accessed 06/06/17).

Hu C, Xiong N, Zhang Y, Rayner & Chen S. Functional characterisation of lipase in the pathogenesis of *Staphylococcus aureus*. *Biochemical and Biophysical Research Communications*. 2012; 419: 617 - 620.

Hu Y, Meng J, Sei C, Hervin K, Fratamico PM & Shi X. Characterisation and comparative analysis of a second thermonuclease from *Staphylococcal aureus*. *Microbiological Research*. 2013; 168: 174 - 182.

Huang TW, Chen FJ, Miu WC, Liao TL, Lin AC, Huang IW, Wu KM, Tsai SF, Chen YT & Lauderdale TL. Complete genome sequence of *Staphylococcus aureus* M013, a *pvl*-positive, ST59-SCC*mec* type V strain isolated in Taiwan. *Journal of Bacteriology*. 2012; 194(5): 1256 - 1257.

- Ito T & Hiramastu. Acquisition of Methicillin Resistance and Progression of Multi-antibiotic Resistance in Methicillin-Resistant *Staphylococcus aureus*. *Yonsei Medical Journal*. 1998; 39(6): 526 - 533.
- Ito Y, Yoh MF, Toda K, Shimazaki M, Nakamura T & Morita E. Staphylococcal scalded-skin syndrome in an adult due to methicillin-resistant *Staphylococcus aureus*. *Journal of Infectious Chemotherapy*. 2002; 8: 256 - 261.
- James CW, Gurk-Turner C. Cross-reactivity of beta-lactam antibiotics. *BUMC Proceedings*. 2001; 14: 106 - 107.
- Jamrozny DM, Fielder MD, Butaye P & Coldham NG. Comparative Genotypic and Phenotypic Characterisation of Methicillin-Resistant *Staphylococcus aureus* ST398 Isolated from Animals and Humans. *PLoS One*. 2010; 7(70).
- Jevons MP. Celbenin-resistant Staphylococci. *British Journals of Medicine*. 1961; 1:124 - 125.
- Jiufeng S, Huang Y, Huang H, Liang P, Wang X, Mao Q, Men J, Chen W, Deng C, Zhou X, Lv X, Zhou J, Zhang F, Li R, Tian Y, Lei H, Liang C, Hu C, Li X & Yu X. Low divergence of *Clonorchis Siensis* in China based on multi-locus analysis. *PLoS One*. 2013; 8(6): e67006.
- Jongerius I, Puister M, Wu J, Ruyken M, Van Strijp JAG & Rooijackers SHM. Staphylococcal Complement Inhibitor Modulates Phagocyte Responses by Dimerisation of Convertases. *The Journal of Immunology*. 2009; 184: 420 - 425.
- Jongerius I, Kohl J, Pandey MK, Ruyken M, Van Kessel KPM, Van Strijp JAG & Rooijackers SHM. Staphylococcal complement evasion by various convertase-blocking molecules. *The Journal of Experimental Medicine*. 2007; 204(10): 2461 - 2471.
- Kafala B & Sasarman A. Cloning and sequence analysis of the *hemB* gene of *Staphylococcus aureus*. *Canadian Journal of Microbiology*. 1994; 40(8): 651 - 657.
- Kaiko GE & Stappenbeck TS. Host-microbe interactions shaping gastrointestinal environment. *Trends in Immunology*. 2014; 35(11): 538 - 548.
- Kanehisa M. The KEGG database. *Novartis Foundation Symposia*. 2002; 247: 91 – 101.
- Kaplan NM. Use of thermonuclear testing to identify *Staphylococcus aureus* by direct examination of blood cultures. *East Mediterranean Health Journal*. 2003; 9(1/2): 185 - 190.
- Katyama Y & Hiramatsu K. A new class of genetic element, *Staphylococcus* cassette chromosome *mec*, encodes methicillin resistance in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*. 2000; 44(6): 1549 - 1555.
- Katayama Y, Baba T, Fukuda M & Hiramatsu K. Beta-Hemolysin Promotes Skin Colonisation by *Staphylococcus aureus*. *The Journal of Bacteriology*. 2013; 195(6): 1194 - 1203.
- Kazmierczak Z, Gorski A & Dabrowska. Facing Antibiotic Resistance: *Staphylococcus aureus* Phages as a

medical tool. *Viruses*. 2014. 6, 2551 - 2570.

Kennedy AD, Otto M, Braughton KR, Whitney AR, Chen L, Mathema B, Mediavilla JR, Byrne KA, Parkins LD, Tenover FC, Kreiswirth BN, Musser JM & DeLeo RF. Epidemic community-associated methicillin-resistant *Staphylococcus aureus*: Recent clonal expansion and diversification. *Proceedings of the National Academy of Science*. 2008; 105(4): 1327 – 1332.

Khawcharoenporn T, Tice AD, Grandinetti A & Chow D. Risk factors for community-associated methicillin-resistant *Staphylococcus aureus* cellulitis - and the value of recognition. *Hawaii Medical Journal*. 2010; 69: 232 - 236.

Kim SH, Chun HS, Han MH, Park NY & Suk K. A Novel Variant of Staphylokinase Gene from *Staphylococcus aureus* ATTC 29213. *Thrombosis Research*. 1997; 87(4): 387 - 395.

Kimman T, Hoek M & de Jong MCM. Assessing and controlling health risks from animal husbandry. *NJAS Wageningen of Life Sciences*. 2013; 66: 7 - 14.

Klein E, Smith DL & Laxminarayan R. Hospitalisations and deaths caused by methicillin-resistant *Staphylococcus aureus*, United States, 1999 - 2005. *Emerging Infectious Diseases*. 2007; 13(12): 1840 - 1846.

Kluytmans J, Van Belkum A and Verbrugh H. Nasal Carriage of *Staphylococcus aureus*: Epidemiology, Underlying Mechanisms, and Associated Risks. *The Journal of Clinical Microbiology Reviews*. 1997; 10(3): 505 - 520.

Kock R, Harlizius J, Bressan N, Laerberg R, Wieler LH, Witte W, Deurenberg RH, Voss A, Becker K & Friedrich AW. Prevalence and Molecular characteristics of methicillin-resistant *Staphylococcus aureus* (MRSA) among pigs on German farms and import of livestock-related MRSA into hospitals. *European Journal of Clinical Microbiology & Infectious Disease*. 2009; 28: 1375-1382.

Krawczyk-Balska A & Bielecki J. *Listeria monocytogenes* listeriolysin O and phosphatidylinositol-specific phospholipase C affect adherence to epithelial cells. *Canadian Journal of Microbiology*. 2005; 51(9): 745 - 751.

Kreisel KM, Stine OC, Johnson JK, Perencevich EN, Sardell MD, Lesse AJ, Gordin FM, Climo MW & Roughmann MC. USA300 Methicillin-resistant *Staphylococcus aureus* bacteremia and the risk of severe sepsis: Is USA300 MRSA associated with more severe infections? *Diagnostic Microbiology and Infectious Disease*. 2011; 70(3): 285 - 290.

Kreiswirth BN, Lofdahl S, Betley MJ, O'Reilly M, Schlieret PM, Bergdoll MS & Novick RP. The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature*. 1983; 20: 709 - 712.

Kreiswirth B, Kornblum J, Arbeit RD, Eisner W, Maslow JN, McGeer A, Low DE & Novick RP. Evidence for a clonal origin of methicillin-resistance in *Staphylococcus aureus*. *SCIENCE*. 1993; 259: 227 - 230.

Kumar S, Tamura K & Nei M. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Computer Application in Bioscience*. 1994; 10(2): 189 - 191.

- Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki KI, Nagai Y, Lian T, Ito T, Kanamori M, Matsumaru H, Maruyama A, Murakami H, Hosoyama A, Mizutani-Ui Y, Takahashi NK, Sawano T, Inoue RI, 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. Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *The Lancet*. 2001; 357(9264): 1225 - 1240.
- Kwiecinski J, Jacobson G, Karlsson M, Zhu X, Wang W, Bremell T, Josefsson E & Jin T. Staphylokinase Promotes the Establishment of *Staphylococcus aureus* Skin Infections While Decreasing Disease Severity. *The Journals of Infectious Diseases*. 2013; 208: 990 - 998.
- Kwok AYC, Su S, Reynolds RP, Bay SJ, Av-Gay Y, Dovichi NJ & Chow AW. Species identification and phylogenetic relationships based on partial HSP60 gene sequences within the genus *Staphylococcus*. *International Journal of Systemic Bacteriology*. 1999; 49, 1181-1192.
- Laarman AJ, Ruyken M, Malone CL, van Strijp JAG, Horswill AR & Rooijackers SHM. *Staphylococcus aureus* Metalloprotease Aureolysin Cleaves Complement C3 to Mediate Immune Evasion. *The Journal of Immunology*. 2011;186: 6445 - 6453.
- Lambert G & Kussell E. Quantifying selective pressure driving bacterial evolution using lineage analysis. *Physical Reviews*. 2015; 5(1).
- Larner-Sevensson H, Worning P, Bartels MD, Hansen LH, Boye K & Westh H. Complete Genome of *Staphylococcus aureus* Strain M1, a Unique t024-ST8-IVa Danish Methicillin-Resistant *S. aureus* Clone. *Genome Announcements*. 2013; 1(3): e00336-13.
- Lee SS, Kim YJ, Chung DR, Jung KS & Kim JS. Invasive Infection Caused by a Community-Associated Methicillin-Resistant *Staphylococcus aureus* Strain Not Carrying Pantone-Valentine Leukocidin in South Korea. *The Journal of Clinical Microbiology*. 2010; 48(1): 311 - 313.
- Lee BY, Singh A, David MZ, Bartsch SM, Slayton RB, Huang SS, Zimmer SM, Potter MA, Macal CM, Lauderdale DS, Miller LG & Daum RS. The economic burden of community-associated methicillin-resistant *Staphylococcus aureus* (CA-MRSA). *Clinical Microbiology and Infection*. 2013; 19(6): 528 - 536.
- Lefebure T & Stanhope MJ. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biology*. 2007; 8:R71.
- Lefebure T & Stanhope MJ. Pervasive, genome-wide positive selection leading to functional divergence in the bacterial genus *Campylobacter*. *Genome Research*. 2009; 19: 1224 - 1232.
- Leigh JW & Bryant D. POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*. 2015; 6: 1110 - 1116.
- Lim D & Strynadka NCJ. Structural basis for the β -lactam resistance of the PBP2a from methicillin-resistant *Staphylococcus aureus*. *Natural Structural Biology*. 2002; 9(11): 870 - 876.

- Lin D & Strynadka NC. Structural basis for the beta lactic resistance of PBP2a from methicillin-resistant *Staphylococcus aureus*. *Natural Structure and Molecular Biology*. 2002; 9(11): 870 - 876.
- Lindqvist M, Isakasson B, Grub C, Jonassen TO & Hallgren A. Detection and characterisation of SCCmec remnants in multiresistant methicillin-susceptible *Staphylococcus aureus* causing a clonal outbreak in a Swedish county. *European Journal of Clinical Microbial Infectious Diseases*. 2012; 31: 141 - 147.
- Lister J. On the Antiseptic Principle in the Practice of Surgery. *British Medical Journal*. 1867; 21; 2(351): 246-248
- Li M, Cheung GYC, Hu J, Wang D, Joo, HS, DeLeo FR & Otto M. Comparative Analysis of Virulence and Toxin Expression of Global Community-Associated Methicillin-Resistant *Staphylococcus aureus* Strains. *The Journal of Infectious Diseases*. 2010; 202: 1866 – 1876.
- Li Y, Cao B, Zhang Y, Zhou J, Yang B & Wang L. Complete genome sequence of *Staphylococcus aureus* T0131, an ST239-MRSA-SCCmec Type III clone isolated in China. *Journal of Bacteriology*. 2011; 193(13): 3411 - 3412.
- Librado P & Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009; 25(11): 1451 - 1452.
- Lipovy B, Brychta P, Chaloupkova Z & Suchanek I. Staphylococcal scalded skin syndrome in the Czech Republic an epidemiological study. *Burns*. 2012; 38: 296 - 300.
- Llopart A & Aguade M. Synonymous rates *RplI215* gene of *Drosophila*: Variation among species and across the coding region. *Genetics*. 1999; 152: 269 – 280.
- Lorenz MG & Wackernagel W. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiological Reviews*. 1994; 58(3): 563 - 602.
- Lowder BV, Guinane CM, Zakour NB, Weinert LA, Conway-Morris A, Cartwright RA, Simpson AJ, Rambout A, Nubel U & Fitzgerald JR. Recent human-to-poultry host jump, adaptation and pandemic spread of *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences*. 2009; 106(46): 19545 - 19550.
- Lowy FD. Antimicrobial resistance: the example of *Staphylococcus aureus*. *The Journal of Clinical Investigation*. 2003; 111(9): 1265 - 1273.
- Markis G, Wright JD, Ingham E & Holland KT. The hyaluronate lyase of *Staphylococcus aureus* - a virulence factor? *Microbiology*. 2004; 150: 2005 - 2013.
- Martin JF, Ullan RV & Garcia-Estrada C. Regulation and compartmentalisation of beta-lactam biosynthesis. *Microbial Biotechnology*. 2010; 3(3): 285 - 299.
- McAdow M, Missiakas DN & Schneewind O. *Staphylococcus aureus* Secretes Coagulase and von Willebrand Factor Binding Protein to Modify the Coagulation Cascade and Establish Host Infections. *Journal of Innate Immunity*. 2012; 4: 141 - 148.

- McCarthy AJ & Lindsay JA. The distribution of plasmids that carry virulence and resistance genes in *Staphylococcus aureus* is lineage associated. *BMC Microbiology*. 2012; 12: 104
- McDonald M, Dougall A, Holt D, Huygens F, Oppedisano F, Gifford PM, Inman-Barber J, Stephens AJ, Towers R, Carapetis JR & Currie BJ. Use of a Single-Nucleotide Polymorphism Genotyping System To Demonstrate the Unique Epidemiology of Methicillin-Resistant *Staphylococcus aureus* in Remote Aboriginal Communities. *Journal of Clinical Microbiology*. 2006; 44(10): 3720 - 3727.
- McGavin MJ, Arsic B & Nickerson NN. Evolutionary blueprint for host- and niche-adaptation in *Staphylococcus aureus* clonal complex CC30. *Frontiers in Cellular and Infection Microbiology*. 2012; 2(48).
- McGavin MJ, Zahradka C, Rice K & Scott EJ. Modification of the *Staphylococcus aureus* Fibronectin Binding Phenotype by V8 Protease. *Infection and Immunity*. 1997; 65(7): 2621 - 2628.
- McNamee PT, Symth JA & Smyth JA. Bacterial chondronecrosis with osteomyelitis ('femoral head necrosis') of broiler chickens: A review. *Avian Pathology*. 2000; 29: 253 - 270.
- Meroueh SO, Bencze KZ, Heseck D, Lee M, Fisher JF, Stemmler TL & Mobashery S. Three-dimensional structure of the bacterial cell wall peptidoglycan. *PNAS*. 2006; 103(12): 4404 - 4409.
- Molne L, Verdrengh MA & Tarkowski A. Role of Neutrophil Leukocytes in Cutaneous Infection Caused by *Staphylococcus aureus*. *Infection and Immunity*. 2000; 68(11): 6162 - 6167.
- Monaco M, Pedroni P, Sanchini A, Bonomini A, Indelicato A & Pantosti A. Livestock-associated methicillin-resistant *Staphylococcus aureus* responsible for human colonisation and infection in an area of Italy with high density of pig farming. *BMC Infectious Diseases*. 2013; 13: 258.
- Mongodin E, Bajolet O, Cutrona J, Bonnet N, Dupuit F, Puchelle E & de Bentzmann S. Fibronectin-Binding Proteins of *Staphylococcus aureus* Are Involved in Adherence to Human Airway Epithelium. *Infection and Immunity*. 2002; 70(2): 620 - 630.
- Mosesson MW. Fibrinogen and fibrin structure and functions. *Journal of Thrombosis and Haemostasis*. 2005; 3: 1984 - 1904.
- Mousset S, Derome N & Veuille M. A test of neutrality and constant population size based on the mismatch distribution. *Molecular Biology and Evolution*. 2004; 21(4): 724 - 731.
- Mwangi MM, Wu SW, Zhou Y, Sieradzki K, de Lencastre H, Richardson P, Bruce D, Rubin E, Myers E, Siggia ED, & Tomasz A. Tracking the in vivo evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proceedings of the National Academy of Sciences*. 2007; 104(22): 9451 - 9456.
- NCBI. (2015). Available from: <http://www.ncbi.nlm.nih.gov/genbank/statistics>. [Accessed: 7th December 2015].
- Nei M & Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *PNAS*. 1979; 76(10): 5629 - 5273.

Newsom S W B. Ogston's coccus. *Journal of hospital Infection*. 2008; 70: 369-372.

Nichols RL & Florman S. Clinical presentations of soft-tissue infections and surgical site infections. *Clinical Infectious Diseases*. 2001; 33(2): 84 - 93.

Nubel U, Dordel J, Kurt K, Strommenger B, Westh H, Shukla SK, Zemlickova H, Lebois R, Wirth T, Jombart T, Balloux F & Witte W. A TimeScale for Evolution, Population Expansion, and Spatial Spread of an Emerging Clone of Methicillin-Resistant *Staphylococcus aureus*. *PLoS One*. 2010; 6(4): e1000855.

Ogsten A. Report upon Micro-organisms in surgical Diseases. *British Medical Journals*. 1881; 1(1054): 369.b2-375.

O'Neill GL, Murchan S, Gil-Setas A & Aucken HM. Identification and Characterisation of the Phage Variants of a Strain of Epidemic Methicillin-Resistant *Staphylococcus aureus* (EMRSA-15). *Journal of Clinical Microbiology*. 2001; 39(4): 1540 - 1548.

ONS-1. Office for national statistics. Deaths involving *Staphylococcus aureus* and MRSA, England & Wales, 1993 - 2006. Available from: http://webarchive.nationalarchives.gov.uk/20160107063144/http://www.ons.gov.uk/ons/dcp171778_324558.pdf. (Accessed 06/06/17).

Ortega E, Abriouel H, Lucas R & Galvez A. Multiple Roles of *Staphylococcus aureus* Enterotoxins: Pathogenicity, Super-antigenic Activity, and Correlation to Antibiotic Resistance. *Toxins*. 2010; 2: 2117 - 2131.

Palma M, Shannon O, Quezada HO, Berg A & Flock JI. Extracellular Fibrinogen-binding Protein, EFB, from *Staphylococcus aureus* Blocks Platelet Aggregation Due to Its Binding to the α -Chain. *The Journal of Biological Chemistry*. 2001; 276(34): 31691 - 31697.

Pantosti A. Methicillin-resistant *Staphylococcus aureus* associated with animals and its relevance to human health. *Frontiers in Microbiology*. 2012; 3(127): doi.10.3389/fmicb.2012.00127.

Parker D, Narechania A, Sebra R, Deikus G, LaRussa S, Ryan C, Smith H, Prince A, Mathema B, Ratner AJ, Kreiswirth B & Planet PJ. Genome Sequence of Bacterial Interference Strain of *Staphylococcus aureus* 502A. *Genome Announcements*. 2014; 2(2): e00284-14.

Parkhill J & Wren BW. Bacterial epidemiology and biology - lessons from genome sequencing. *Genome Biology*. 2011; 12: 230.

Paterson GK, Harrison EM & Holmes MA. The emergence of *mecC* methicillin-resistant *Staphylococcus aureus*. *Trends in Microbiology*. 2014; 22(1): 42 - 47.

Pattengale ND, Alipour M, Bibinda-Emonds OR, Moret BM & Stamatakis A. How many bootstraps replicates are necessary? *Journal of Computational Biology*. 2010; 17(3): 337 - 354.

Patti JM, Jonsson H, Guss B, Switalski LM, Wiberg K, Lindberg M & Hook M. Molecular Characterisation and

Expression of a Gene Encoding a *Staphylococcus aureus* Collagen Adhesion. *The Journal of Biological Chemistry*. 1992; 267(7); 4766 - 4722

Peeters LEJ, Argudin MA, Azadikhah S, Butaye P. Antimicrobial resistance and population structure of *Staphylococcus aureus* recovered from pig farms. *Veterinary Microbiology*. 2015; 180: 151 – 156.

Peetermans M, Vanassche T, Liesenbroghs L, Claes J, Velde GV, Kwiecinski J, Jin T, De Geest B, Hoylaert MF & Verhamme P. Plasminogen activation by staphylokinase enhances local spreading of *S. aureus* in skin infections. *BMC Microbiology*. 2014; 14: 310.

Petkau A, Stuart-Edwards M, Stothard P, and Van Domselaar G. Interactive microbial genome visualisation with GView. *Bioinformatics* 2010; 26; 3125-3126.

Pereira LB. Impetigo - review. *Anais Brasileiros de Dermatologia*. 2014; 89(2): 293 - 299.

Persson L, Johnansson C and Ryden. Antibodies to *Staphylococcus aureus* Bone Sialoprotein-Binding Protein Indicate Infectious Osteomyelitis. *Clinical and Vaccine Immunology*. 2009; 16(6): 949 - 952.

Pinchuk IV, Beswick EJ & Reyes VE. Staphylococcal Enterotoxins. *Toxins*. 2012; 2, 2177-2197.

Plata K, Rosato AE, Wegrzyn G. *Staphylococcus aureus* as an infectious agent: overview of biochemistry and molecular genetics of its pathogenicity. *Acta Biochemical Polonica*. 2009; 56(4), 597-612.

Popart-1. PopART. 2016. Available at: <http://popart.otago.ac.nz>. (Accessed 9th January 2016).

Prasad GS, Radhakrishnan R, Mitchell DT, Earhart CA, Dinges MM, Cook WJ, Schlieret PM & Ohlendorf HH. Refined Structures of the crystal form of toxic shock syndrome toxin-1 and of a tetramutant with reduced activity. *Protein Science*. 1997; 6: 1220 - 1227.

Ragle BE & Wardenburg JB. Anti-Alpha-Hemolysin Monoclonal Antibodies Mediate Protection against *Staphylococcus aureus* Pneumonia. *Infection and Immunity*. 2009; 77(7): 2712 - 2718.

Rahman A, Izaki K, Kato I & Kamio Y. Nucleotide sequence of leukocidin S-component gene (*lukS*) from methicillin resistant *Staphylococcus aureus*. *Biochemical and Biophysical Research Communication*. 1991; 181(1): 138 - 144.

Ramos-Onsins SE & Rozas J. Statistical properties of new neutrality tests against population growth. *Molecular Biology & Evolution*. 2002; 19(12): 2092 – 2100.

Rangarajan ES & Shankar V. Sugar non-specific endonucleases. *FEMS Microbiology Reviews*. 2001; 25(5): 583 - 613.

Rasooly R & Hernlem BJ. Quantitative Analysis of *Staphylococcus* Enterotoxin A by Differential Expression of IFN- γ in Splenocyte and CD4+ T-Cells. *Sensors*. 2014; 14: 8869 - 8876.

Rhem MN, Lech EM, Patti JM, McDevitt D, Hook M, Jones DB & Wilhelmus KR. The Collagen-Binding

Adhesion is a Virulence Factor in *Staphylococcus aureus* Keratitis. *Infection and Immunity*. 2000; 68(6): 3776 - 3779.

Rice K, Peralta R, Bast D, de Azavedo J & McGavin MJ. Description of *Staphylococcus* Serine Protease (*ssp*) Operon in *Staphylococcus aureus* and Non-polar Inactivation of *sspA*-Encoded Serine Protease. *Infection and Immunity*. 2001; 69(1): 159 - 169.

Rice LB. Mechanisms of resistance and clinical relevance of resistance to beta-lactams, glycopeptides and fluoroquinolones. *Symposium on Antimicrobial Therapy*. 2012; 87(2): 198 - 208.

Rollof J, Vinge E, Nilsson-Ehle P & Braconier JH. Aggregation of human granulocytes by *Staphylococcus aureus* lipase. *Journal of Medical Microbiology*. 1992; 36: 52 - 55.

Rooijackers SHM, Milder FJ, Bardoel BW, Ruyken M, Van Strijp JAG & Gros P. Staphylococcal Complement Inhibitor: Structure and Active Sites. *The Journal of Immunology*. 2007; 179(5): 2989 - 2998.

Rodgers JD, McCullagh JJ, McNamee PT, Smyth JA & Ball HJ. Comparison of *Staphylococcus aureus* recovered from personnel in a poultry hatchery and in broiler parent farms with those isolated from skeletal disease in broilers. *Veterinary Microbiology*. 1999; 69: 189 - 198.

Rolinson GN. Forty years of beta-lactam research. *Journal of Antimicrobial Chemotherapy*. 1998; 41: 589 - 603.

Rollof J & Normak S. *in vivo* Processing of *Staphylococcus aureus* Lipase. *Journal of Bacteriology*. 1992; 174(6): 1844 - 1847.

Rosenstein R & Götz F. Staphylococcal lipases: Biochemical and Molecular characterisation. *Biochimie*. 2000; 82: 1005 - 1014.

Rothberg JM & Laemon JH. The development and impact of 454 sequencing. *Nature Biotechnology*. 2008; 26(10): 1117 - 1124.

Rozas J & Rozas R. DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Computational Applications Bioscience*. 1995; 11(6): 621 - 625.

Rozas J, Sanchez-DelBarrio JC, Messeguer X & Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*. 2003; 19(18): 2496 - 2497.

Rozas, J. DNA Sequence Polymorphism Analysis using DnaSP. Pp. 337-350. In Posada, D. (ed.) *Bioinformatics for DNA Sequence Analysis; Methods in Molecular Biology Series*. 2009; Vol. 537. Humana Press, NJ, USA.

Ryden C, Yacoub AI, Maxe I, Heingard D, Oldsberg A, Franzen A, Ljunch A and Rubin K. Specific binding of bone sialoprotein to *Staphylococcus aureus* isolated from patients with osteomyelitis. *The European Journal of Biochemistry*. 1989; 184: 331 - 336.

Sabat A, Kosowska K, Poulsen K, Kasproicz A, Sekowska A, van den Burg B, Travis J & Potempa J. Two

Allelic Forms of the Aurolysin Gene (*aur*) within *Staphylococcus aureus*. *Infection and Immunity*. 2000; 68(2): 973 - 976.

Sabat A, Melles DC, Martirosian G, Grundmann H, van Beckum A and Hryiewicz. Distribution of the Serine-Aspartate Repeat Protein-Encoding *sdr* Genes among Nasal-Carriage and Invasive *Staphylococcus aureus* Strains. *The Journal of Clinical Microbiology*. 2006; 44(3): 1135 - 1138.

Sabirova, J. S., Xavier, B. B., Hernalsteens, J.-P., De Greve, H., Ieven, M., Goossens, H., & Malhotra-Kumar, S. Complete Genome Sequences of Two Prolific Biofilm-Forming *Staphylococcus aureus* Isolates Belonging to USA300 and EMRSA-15 Clonal Lineages. *Genome Announcements*. 2014; 2(3): e00610–14.

Schijffelen MJ, Boel CHE, van Strijp JAG & Fluit AD. Whole genome analysis of a livestock-associated methicillin-resistant *Staphylococcus aureus* ST398 isolate from a case of human endocarditis. *BMC Genomics*. 2010; 11: 376.

Saitou N & Nei M. The Neighbour-Joining method: A new method of reconstructing phylogenetic trees. *Molecular Biology & Evolution*. 1987; 4(4): 406 - 425.

Sanger F, Nicklen S & Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Science USA*. 1977; 74(12): 5463 - 5467.

Sass P, Berscheid A, Jansen A, Oedenkoven M, Szekat C, Strittmatter A, Gottschalk G & Bierbaum G. Genome sequence of *Staphylococcus aureus* VC40, a Vancomycin- and Daptomycin-resistant strain, to study the genetics of development of resistance to currently applied last-resort antibiotics. *Journal of Bacteriology*. 2012; 194(8): 2107 - 2108.

Schaaff F, Reipert A & Bierbaum. An Elevated Mutation Frequency Favours Development of Vancomycin Resistance in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*. 2002; 46(11): 3540 - 3548.

Schmid K & Yang Z. The trouble with sliding windows and the selective pressure in BRCA1. *PLoS ONE*. 2008; 3(11): e3746.

Shannon O, Uekotter A & Flock JI. The Neutralising Effects of Hyperimmune Antibodies Against Fibrinogen-Binding Protein, Efb, from *Staphylococcus aureus*. *Scandinavian Journal of Immunology*. 2005; 63: 184 - 190.

Sheng-Yun L, Fang-Yu C, Ching-Chung Cheng, Keong-Diong L & Yhu-Chering H. Methicillin-resistant *Staphylococcus aureus* nasal colonisation among adult patients visiting emergency department in a medical centre in Taiwan. *PLoS One*. 2011; 6(6): e18620.

Shi D, Ishii S, Sato T, Yamasaki H, Matsunaga M, Higuchi W, Takano T, Yabe S, Tanaka K & Yamamoto T. Staphylococcal scalded skin syndrome in an extremely low-birth-weight neonate: Molecular characterisation and rapid detection by multiplex and real-time PCR of methicillin-resistant *Staphylococcus aureus*. *Paediatrics International*. 2011; 53: 211 - 217.

Shinji H, Yosizawa Y, Tajima A, Iwase T, Sugimoto S, Seki K & Mizunoe. Role of Fibronectin-Binding proteins A and B in *In Vitro* Cellular Infections by *Staphylococcus aureus*. *Infection and Immunity*. 2011; 79(6): 2215 - 2223.

Shopsin B, Gomez M, Waddington M, Riehman M & Kreiswirth BN. Use of Coagulase Gene (*coa*) Repeat Region Nucleotide Sequences for Typing of Methicillin-Resistant *Staphylococcus aureus* Strain. *Journal of Clinical Microbiology*. 2000; 38(9): 3453 - 3456.

Sieradzki K, Leski T, Dick J, Borio L & Tomasz. Evolution of a Vancomycin-intermediate *Staphylococcus aureus* Strain In Vivo: Multiple Changes in the Antibiotic Resistance Phenotypes of a Single Lineage of Methicillin-Resistant *S. aureus* under the Impact of Antibiotics Administered for Chemotherapy. *The Journal of Clinical Chemotherapy*. 2003; 41(4): 1687 - 1693.

Sifri CD, Baresch-Bernal A, Calderwood SB & Von Eiff C. Virulence of *Staphylococcus aureus* Small Colony Variants in the *Caenorhabditis elegans* Infection Model. *Infection and Immunity*. 2006; 74(2): 1091 - 1096.

Smith TL, Pearson ML, Wilcox KR, Cruz C, Lancaster MV, Robinson-Dunn B, Tenover FC, Zervos MJ, Band JD, White E & Jarvis WR. Emergence of Vancomycin resistance in *Staphylococcus aureus*. *The New England Journal of Medicine*. 1999; 340(7): 493 - 501.

Smith EJ, Visai L, Kerrigan SW, Speziale P & Foster TJ. The *sbi* Protein is a Multifunctional Immune Evasion Factor of *Staphylococcus aureus*. *Infection and Immunity*. 2011; 79(9): 3801 - 3809.

Smith K A. Louis Pasteur, the father of immunology? *Frontiers in Immunology*. 2012; 68(3): 1-10.

Sodre D, Rodriguis-Filho LFS, Souza RFC, Rego PS, Schneider H, Sampaio I & Vallinoto M. Inclusion of South American samples reveals new population structuring of the blacktip shark (*Carcharhinus limbatus*) in the western Atlantic. *Genetics and Molecular Biology*. 2012; 35(4): 752 – 760.

Sola C, Paganini H, Egea AL, Moyano AJ, Garnera A, Kevric I, Culasso C, Vindel A, Study Group of CA-MRSA in Children, Argentina-2007, Lopardo H & Bocco JL. Spread of epidemic MRSA-ST5-IV clone encoding PVL as a major cause of community onset Staphylococcal infectious in Argentinean children. *PLoS One*. 2012; 7(1): e30487.

Solar GD, Giraldo R, Ruiz-Echevarria MJ, Espinosa M & Diaz-Orejas. Replication and control of circular bacterial plasmids. *Microbiology and Molecular Biology Reviews*. 1998; 62(2): 434 – 464.

Sousa MA & de Lencastre H. Bridges from hospitals to the laboratory: genetic portraits of methicillin-resistant *Staphylococcus aureus* clones. *FEMS Immunology and Medical Microbiology*. 2004; 40: 101 - 111.

Soyer Y, Orsi RH, Rodriguez-Rivera LD, Sun Q & Wiedmann M. Genome wide evolutionary analyses reveal serotype specific patterns of positive selection in selected *Salmonella* serotypes. *BMC Evolutionary Biology*. 2009; 9:264: doi:10.1186/1471-2148-9-264.

Spaan AN, Vrieling M, Wallet P, Badiou C, Reyes-Robles T, Ohneck EA, Benito Y, de Haas, CJC, Day CJ, Jennings MP, Lina G, Vandenesch F, van Kessel KPM, Torres VJ, van Strijp JAG & Henry. The staphylococcal toxins γ -hemolysin AB and CB differentially target phagocytes by implying specific chemokine receptors. *Nature Communication*. 2014; 5: 5438.

Stegger M, Price LB, Larsen AR, Gillece JD, Waters AE, Skov R & Anderson PS. Genome sequence of

Staphylococcus aureus 11819-97, an ST80-IV European community-acquired methicillin-resistant isolate. *Journal of Bacteriology*. 2012; 194(6): 1625-1626.

Stegger M, Driebe EM, Roe C, Lemmar D, Bowers JR, Engethaler DM, Keim P & Anderson PS. Genome Sequence of *Staphylococcus aureus* Strain CA-347, a USA600 Methicillin-Resistant Isolate. *Genome Announcements*. 2013; 1(4): e00517-13.

Sun J, Deng & Yan. Bacterial multidrug efflux pumps: Mechanisms, physiology and pharmacological exploitations. *Biochemical and Biophysical Research Communications*. 2014; 453: 254 - 267.

Sung JML, Lloyd DH & Lindsay. *Staphylococcus aureus* host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiology*. 2008; 154: 1949 - 1959.

Suzuki H, Lefebure T, Bitar PP & Stanhope MJ. Comparative genomic analysis of the genus *Staphylococcus* including *Staphylococcus aureus* and its newly described sister species *Staphylococcus simiae*. *BMC Central*. 2012; 13(38): doi: 10.1186/1471-2164-13-38.

Sydnor ERM & Perl TM. Hospital epidemiology and infection control in acute-care settings. *Clinical Microbiology Reviews*. 2011; 24(1): 141 - 173.

Tang J, Zhou R, Shi X, Kang M, Wang H & Chen H. Two thermostable nucleases coexisted in *Staphylococcus aureus*: evidence from mutagenesis and *in vitro* expression. *FEMS Microbiology Letter*. 2008; 284: 176 - 183.

Tamer K, Stecher G, Peterson D, Filipski A & Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*. 2013; 30(12): 2725 - 2729.

Tajima F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics*. 1989; 123: 585 - 595.

Taubes G. The bacteria fight back. *SCIENCE*. 2008; 321: 356 - 361.

Taylor LH, Lathan SP & Woolhouse MEJ. Risk factors for human disease emergence. *The Royal Society*. 2001; 256: 983 – 989.

Tung HS, Guss B, Hellman U, Persson L, Rubin K & Ryden C. A bone sialoprotein-binding protein from *Staphylococcus aureus*: a member of the staphylococcal Sdr family. *Biochemical Journal*. 2000; 345: 611 - 619.

Uhlemann AC, Porcella SF, Trivedi S, Sullivan SB, Hafer C, Kennedy AD, Barbian KD, McCarthy AJ, Street C, Hirschberg DL, Lipkin WI, Lindsay JA, DeLeo FR & Lowy FD. Identification of a Highly Transmissible Animal-Independent *Staphylococcus aureus* ST398 Clone with Distinct Genomic and Cell Adhesion Properties. *The American Society for Microbiology*. 2012; 3(2): e00027-12.

Vann JM & Proctor RA. Ingestion of *Staphylococcus aureus* by Bovine Endothelial Cells Results in Time- and Inoculum-Dependent Damage to Endothelial Cell Monolayers. *Infection and Immunity*. 1987; 55(9): 2155 - 2163.

van Cleef BA, Verkade EJM, Wulf MW, Buiting AG, Voss A, Huijasedens XW, van Pelt W, Mulders MN & Kluytmans JA. Prevalence of Livestock-Associated MRSA in Communities with High Pig-Densities in The Netherlands. *PLoS One*. 2010; 5(2).

van Duijkeren E, Jansen MD, Flemming SC, de Heeling H, Wagenaar JA, Schoormans AHW, van Nes A & Fluit AdC. Methicillin-resistant *Staphylococcus aureus* in pigs with exudative epidermitis. *Emerging Infectious Diseases*. 2007; 13(9): 1408 - 1410.

van Rijen MML, Bosch T, Verkade EJM, Schouls L & Kluytmans JA. Livestock-Associated MRSA Carriage in Patients with Ought Direct Contact with Livestock. *PLoS One*. 2014; 9(6).

Van Wamel WJB, Rooijackers SHM, Ruyken M, Van Kessel KPM, & Van Strijp JAG. The Innate Modulators Staphylococcal Complement Inhibitor Protein of *Staphylococcus aureus* Are Located on β -Hemolysin-Converting Bacteriophages. *The Journal of Bacteriology*. 2006; 188(4): 1310 - 1315.

Vaudaux P, Francois P, Bisognano C, Kelley WL, Lew DP, Schrenzel J, Proctor RA, McNamara PJ, Peters G & Von Eiff C. Increased Expression of Clumping Factor and Fibronectin-binding proteins by *hemB* Mutants of *Staphylococcus aureus* Expressing Small Colony Variants. *Infection and Immunity*. 2002; 70(10): 5428 - 5437.

Vautor E, Cockfield J, Le Marcechal C, Le Loir Y, Chevalier M, Robinson DA, Thiery R & Lindsay J. Differences in virulence between *Staphylococcus aureus* isolates causing gangrenous mastitis versus subclinical mastitis in a dairy sheep flock. *Veterinary Research*. 2009; 40(56): DOI. 10.1051/vetres/2009039.

Vazquez V, Liang X, Horndhal JK, Ganesh VK, Smeds E, Foster TJ & Hook M. Fibrinogen Is a Ligand for the *Staphylococcus aureus* Microbial Surface Components Recognising Adhesive Matrix Molecules (MSCRAMM) Bone Sialoprotein-binding Protein (Bbp). *The Journal of Biological Chemistry*. 2011; 286(34): 29797 - 29805.

Ventola C.L. The antibiotic resistance crisis Part 2: Management Strategies and New Agents. *Pharmacy and Therapeutics*. 2015; 20(5):344 – 352.

Vogel V, Falquet L, Calderon-Copete SP, Basset P & Blanc DS. Short term evolution of a highly transmissible methicillin-resistant *Staphylococcus aureus* clone (ST228) in a tertiary care hospital. *PLoS One*. 2012; 7(12): e38969.

Von Eiff C, Heilmann C, Proctor RA, Waltz C, Peters G & Gotz F. A Site-Directed *Staphylococcus aureus hemB* Mutant Is a Small-Colony Variant Which Persists Intracellularly. *The Journal of Bacteriology*. 1997; 179(15): 4706 - 4712.

Wann ER, Gurusiddappa S & Hook M. The Fibronectin-binding MSCRAMM FnbpA of *Staphylococcus aureus* Is a Bifunctional Protein That Also Binds to Fibrinogen. *The Journal of Biological Chemistry*. 2000; 275(18):13863 - 13871.

Wakeley J. The variance of pairwise Nucleotide Difference in two populations with migration. *Theoretical Population Biology*. 1994; 49(2): 39 - 57.

Weese JS, Rousseau J, Willey BM, Archambault M, McGeer A & Low DE. Methicillin-resistant *Staphylococcus aureus* in horses at a veterinary teaching hospital: Frequency, characterisation, and association with clinical disease. *Journal of Veterinary Internal Medicine*. 2006; 20: 182 - 186.

Weese JS. Methicillin-resistant *Staphylococcus aureus* in animals. *Oxford Journals*. 2010; 51(3): 233 - 244.
Westram AM, Panova M, Galindo J & Butlin RK. Targeted resequencing reveals geographical patterns of differentiation for loci implicated in parallel evolution. *Molecular Ecology*. 2016. In Press: <http://dx.doi.org/10.1111/mec.13640>.

WHO-1. World Health Organization Europe. Hospital acquired infections; The European context. Available from: http://www.who.int/patientsafety/events/05/apspd/EURO_Initiatives_on_blood_safety_Hafner.pdf. 2012. (Accessed 01/06/12).

WHO-2. World Health Organization. A global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics: <http://www.who.int/medicines/publications/global-priority-list-antibiotic-resistant-bacteria/en/>. 2017. (Accessed 06/06/17).

Wielders CLC, Fluit AC, Brisse S, Verhoef J & Schmitz FJ. *mecA* Gene is Widely Disseminated in *Staphylococcus aureus* population. *The Journal of Clinical Microbiology*. 2002; 40(11): 3970 - 3975.

Wilke MS, Lovering AL & Strynadka NCJ. β -Lactam antibiotic resistance: a current structural perspective. *Current Opinion in Microbiology*. 2005; 8: 525 - 533.

Wei Z, Schnupf P, Poussin MA, Zenewicz LA, Shen H & Goldfine H. Characterisation of *Listeria monocytogenes* Expressing Anthrolysin O and Phosphatidylinositol-Specific Phospholipase C from *Bacillus anthracis*. *Infection and Immunity*. 2005; 73(10): 6639 - 6646.

Wertheim HFL, Walsh E, Choudhury R, Melles DC, Boelens HAM, Miajlovic H, Verburgh HA, Foster T & van Belkum A. Key Role for Clumping Factor B in *Staphylococcus aureus* Nasal Colonisation of Human. *PLoS One*. 2008; 5(1).

Williams REO. Healthy Carriage of *Staphylococcus aureus*: Its Prevalence and Importance. *Microbiology and Molecular Biology Reviews*. 1963; 27(1): 56 - 71.

Wilson L G. The early recognition of streptococci as causes of disease. *Medical History*. 1987; 31: 403-414.

Xiang H, Feng Y, Wang J, Liu B, Chen Y, Liu L, Deng X & Yang M. Crystal Structures Reveal the Multi-ligand Binding Mechanism of *Staphylococcus aureus* clfB. *PLoS One*. 2010; 8(6).

Xiong N, Hu C, Zhang Y & Chen S. Interaction of sortase A and lipase 2 in the inhibition of *Staphylococcus aureus* biofilm formation. *Archives of Microbiology*. 2009; 191: 879 - 884.

Xue H, Lu H & Zhao X. Sequence Diversities of Serine-Aspartate Repeat Genes among *Staphylococcus aureus* Isolates from Different Hosts Presumably by Horizontal Gene Transfer. *PLoS One*. 2011; 6(5):e20332.

Yadav PK, Singh G, Singh, S, Gautam B & Saad EI. Potential therapeutic drug target identification in

Community Acquired-Methicillin Resistant *Staphylococcus aureus* (CA-MRSA) using computational analysis. *Bioinformatics*. 2012; 8(14): 664-672.

Yang Z & Bielawski JP. Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*. 2000; 15(12): 496 - 503.

Yang Z & Rannala B. Molecular phylogenetics: Principles and practise. *Nature reviews*. 2012; 13: 303 - 314.

Zhang X, Hu Q, Yuan W, Shang W, Cheng H, Yuan J, Zhu J, Hu Z, Li S, Chen W, Hu X & Rao X. First report of a sequence type 239 vancomycin-intermediate *Staphylococcus aureus* isolate in mainland China. *Diagnostic Microbiology and Infectious Disease*. 2013; 77:64 - 68.

Zhang HZ, Hackbarth CJ, Chansky KM & Chambers HF. A proteolytic transmembrane signalling pathway and resistance to beta-lactams in Staphylococci. *Science*. 2001; 291: 1962 - 1965.

Zhang X, Xu X, Yuan W, Hu Q, Shang W, Hu, X, Tong Y & Rao X. Complete Genome Sequence of *Staphylococcus aureus* XN108, an ST239-MRSA-SCC*mec* III Strain with Intermediate Vancomycin Resistance Isolated in Mainland China. *Genome Announcements*. 2014; 2(4): e00449–14.

	(L) Length of Sequence	N (No. of Sequences)	S (No. of Variable sites)	h (No. of Haplotypes)	Hd (Haplotype diversity)	SD of Haplotype diversity	π (Nucleotide diversity)	SD of Nucleotide diversity	k (Avg no. of nucleotide difference)	Tajimas D	Statistical Significance	Fu & Li's D	Statistical Significance	dn (Replacement changes)	ds (Synonymous changes)	ω (dn/ds)	Ka	SD	Ks	SD
<i>mecA</i>	2007	26	2	3	0.428	0.095	0.00024	0.0006	0.48615	-0.15513		-0.68907		1	1	1	0.000245538	0.00029547	0.000184615	0.000640513
<i>PBP2</i>	2187	57	232	25	0.944	0.015	0.00769	0.00286	16.73496	-2.41322	p<0.01	-6.1118	p<0.02	26	202	0.128712871	0.002178446	0.002943435	0.031456078	0.08315089
<i>hla</i>	960	56	79	15	0.868	0.028	0.02137	0.00296	20.5188	0.056949		0.94103		11	70	0.157142857	0.004062857	0.003358898	0.093564416	0.10243034
<i>h1b</i>	995	52	134	29	0.948	0.019	0.0243	0.00424	19.97285	-1.39415		-2.45919	p<0.05	128	16	8	0.028095249	0.030307735	0.011451056	0.011060187
<i>h1gA</i>	930	56	58	19	0.868	0.032	0.00684	0.00114	6.35	-1.79359		-4.49654	p<0.02	13	47	0.276595745	0.001355	0.001558075	0.026948117	0.029635237
<i>LukSPV</i>	949	17	268	7	0.809	0.068	0.1102	0.02538	103.47794	0.81837		0.46684		188	92	2.043478261	0.112825735	0.121985566	0.236900735	0.253508508
<i>lukS</i>	1057	53	251	19	0.866	0.035	0.04894	0.00876	51.38679	-63385		-0.93617		229	21	10.9047619	0.059494122	0.068582124	0.02568418	0.026634222
<i>lukF-PV</i>	1011	34	570	9	0.856	0.034	0.16055	0.02883	149.95009	-0.38234		2.04025	p<0.02	161	123	1.308943089	0.150054902	0.229543728	0.593016756	0.46379684
<i>chp</i>	451	28	8	5	0.481	0.108	0.0028	0.00076	1.2619	-1.19239		-0.54024		6	2	3	0.002622222	0.003462406	0.003793915	0.006892478
<i>sbi</i>	1314	57	255	18	0.902	0.02	0.03859	0.00488	48.20175	-0.79584		-1.4108		64	157	0.407643312	0.023576629	0.019564748	0.108159211	0.111625841
<i>sak</i>	491	41	9	9	0.815	0.038	0.0036	0.00032	1.77073	-0.46005		-1.16204		9	0	9	0.004669024	0.00334164	0	0
<i>sspA</i>	1087	57	411	17	0.857	0.029	0.04314	0.01652	31.44862	-2.46653	p<0.01	-6.01246	p<0.02	219	88	2.488636364	0.051335088	0.144144723	0.085016729	0.211448419
<i>sspB</i>	1555	57	201	18	0.898	0.021	0.01832	0.00444	21.63409	-1.86734	p<0.05	-4.87137	p<0.02	65	142	0.457746479	0.007220739	0.011927683	0.071325	0.116519437
<i>sspC</i>	330	57	69	13	0.823	0.035	0.01408	0.00629	4.64599	-2.47292	P<0.01	-6.0626	p<0.02	24	75	0.32	0.006732644	0.021610558	0.053919799	0.148814427
<i>vwb</i>	512	27	134	6	0.547	0.107	0.06312	0.01937	29.72934	-0.71047		1.39427	p<0.05	78	42	1.857142857	0.061724501	0.085825083	0.116790883	0.150017703
<i>lip</i>	2100	57	970	22	0.895	0.024	0.03093	0.01393	62.32331	-2.62849	p<0.001	-6.68815	p<0.02	457	361	1.265927978	0.025712782	0.091913042	0.104507331	0.271810229
<i>lip1</i>	2143	56	845	21	0.926	0.018	0.04453	0.01958	56.24156	-2.64339	p<0.001	-6.48876	p<0.02	365	143	2.552447552	0.059937727	0.185391967	0.088118182	0.265171001
<i>coa</i>	2349	51	367	21	0.911	0.023	0.11623	0.00698	104.83843	-0.11428		-1.42261		198	40	4.95	0.130320627	0.067478835	0.115796235	0.071819499
<i>lipA</i>	918	56	56	15	0.832	0.033	0.00549	0.00163	5.03701	-2.10997	p<0.05	-5.21607	p<0.02	1	59	0.016949153	0.00005	0.000259892	0.026146818	0.048110514
<i>plc</i>	987	57	132	20	0.893	0.022	0.02638	0.00179	26.00689	-0.61081		-2.64197	p<0.05	43	95	0.452631579	0.009644925	0.005997359	0.095161905	0.057472722
<i>plc</i>	987	57	132	20	0.892	0.022	0.02615	0.00184	25.78321	-0.63594		-2.64197	p<0.05	43	95	0.452631579	0.009626065	0.005966283	0.096446805	0.057435175
<i>nuc</i>	687	56	136	18	0.851	0.032	0.01551	0.00567	10.37403	-2.37667	p<0.01	-5.99137	p<0.02	60	71	0.845070423	0.008314221	0.020460035	0.049946688	0.110560148
<i>nucA</i>	690	57	272	16	0.641	0.07303	0.0945	0.02796	50.08709	-0.8496		0.88805		121	92	1.315217391	0.122866165	0.246672777	0.235605451	0.385913198
<i>hysA</i>	2528	57	1500	28	0.898	0.028	0.09215	0.01609	211.86028	-1.66335		-4.75547	p<0.02	626	235	2.663829787	0.109446366	0.167450506	0.135715539	0.231318698
<i>hysA1</i>	4872	57	2884	42	0.977	0.012	0.23432	0.00777	1111.61529	0.26073		-0.45342		419	349	1.200573066	0.272862531	0.139884799	0.390797431	0.164870006
<i>hysA2</i>	2436	57	680	25	0.902	0.026	0.07658	0.00315	183.56078	0.29673		-0.47924		259	283	0.915194346	0.055963534	0.024159787	0.185718922	0.086187476
<i>aur</i>	1558	57	565	18	0.901	0.021	0.08186	0.01848	70.07018	-1.77995		-5.77113	p<0.02	253	100	2.53	0.100618296	0.194420703	0.155057393	0.330170276
<i>hemB</i>	4296	57	1603	38	0.977	0.009	0.0865	0.01886	262.87845	-1.29973		0.64854		722	976	0.739754098	0.099063409	0.157451314	0.13095119	0.161847954
<i>sepA</i>	474	55	65	14	0.853	0.031	0.04711	0.00588	22.04579	1.74923		1.65977	p<0.02	8	59	0.13559322	0.005990572	0.006430586	0.265850976	0.293161811
<i>cna</i>	3554	22	225	13	0.892	0.055	0.04778	0.00303	94.97835	2.01082	p<0.05	1.15284		127	206	0.616504854	0.041912121	0.035334277	0.075595671	0.046459595
<i>efb</i>	498	54	96	12	0.838	0.03	0.03996	0.00659	19.90077	-0.35967		0.05541		32	65	0.492307692	0.016903215	0.018462957	0.151025437	0.171726808
<i>fnb</i>	3252	55	1168	31	0.963	0.013	0.17129	0.03249	245.28283	-1.20642		-1.46986		177	46	3.847826087	0.251516229	0.349209622	0.345513939	0.4613624
<i>ebpS</i>	1471	56	335	21	0.894	0.028	0.02499	0.00628	35.43247	-1.95648	p<0.05	-5.60148	p<0.02	253	71	3.563380282	0.025553831	0.040772489	0.032687597	0.047225734
<i>bbp</i>	4302	57	1604	38	0.977	0.009	0.08631	0.01884	262.97494	-1.29888		0.64854		721	255	2.82745098	0.098827381	0.15730697	0.130908584	0.161439527
<i>eap</i>	435	55	26	15	0.842	0.033	0.01481	0.01481	5.99933	0.0541		0.00312		7	24	0.291666667	0.005684444	0.004132987	0.048569966	0.037354995
<i>clf</i>	3258	56	503	30	0.965	0.011	0.06763	0.00316	143.17078	0.32648		0.05801		389	455	0.854945055	0.075674545	0.040644596	0.050043312	0.028239989
<i>SCIN</i>	345	57	79	16	0.818	0.042	0.01931	0.00594	6.66103	-2.1727	p<0.05	-5.29299	p<0.02	37	37	1	0.014079073	0.02470169	0.045222431	0.100294579
<i>entE</i>	791	19	374	7	0.544	0.136	0.0716	0.0388	53.33918	-2.39421	p<0.01	-3.16397	p<0.02	241	62	3.887096774	0.092631579	0.190466238	0.100051462	0.17933361
<i>sea</i>	780	19	84	4	0.45	0.128	0.01436	0.00921	11.11696	-2.24588	p<0.01	-3.37062	p<0.02	39	84	0.464285714	0.008961404	0.018856668	0.040578363	0.093724416
<i>seb</i>	826	14	213	7	0.802	0.094	0.13588	0.01238	107.48352	2.5904	p<0.01	1.36471	p<0.5	138	68	2.029411765	0.138303297	0.119123102	0.290992308	0.249186388
<i>seh</i>	726	5	81	2	0.4	0.237	0.04807	0.0008134	32.4	-1.26452		-1.26452	p<0.02	26	55	0.472727273	0.02092	0.027007604	0.22348	0.288511439
<i>sek</i>	730	18	34	5	0.824	0.045	0.01735	0.00219	12.64706	1.13049		1.65638	p<0.02	30	4	7.5	0.019051634	0.01246854	0.011477124	0.011477124
<i>seg</i>	777	24	40	6	0.583	0.101	0.01056	0.00262	8.19203	-0.9755		-2.20829		25	41	0.609756098	0.007194565	0.008263361	0.024244565	0.026700738
<i>tst</i>	705	7	17	4	0.714	0.181	0.00959	0.00342	6.7619	-0.14193		0.21786		11	6	1.833333333	0.00817619	0.007906826	0.015371429	0.015368609

Appendix 2:

The following outputs show the results of the BLASTn Search. The BLASTn tool was used to identify if virulence factor genes were present in some or all of *the S. aureus strains*. The cut off value was set at 90% coverage, 85% identity and an E value of <0.01.

Description	Max score	Total score	Query cover	E value	Ident	Accession
MRSA252	6305	6305	100%	0.0	100%	Query_216882
TCH60	6122	6122	100%	0.0	99%	Query_216883
MRSA_IP70	6067	6067	100%	0.0	99%	Query_216878
55/2053	5801	6760	100%	0.0	99%	Query_216884
MRSA_OP100	5422	5422	100%	0.0	95%	Query_216876
RF122	4870	5597	100%	0.0	94%	Query_216885
MSHR1132	4638	4638	94%	0.0	93%	Query_216886
LGA251	4211	4211	100%	0.0	89%	Query_216887
M013	4200	4200	100%	0.0	89%	Query_216890
SA957	4200	4200	100%	0.0	89%	Query_216889
SA40	4200	4200	100%	0.0	89%	Query_216888
N315	4167	4167	100%	0.0	89%	Query_216891
Mu50	4161	4161	100%	0.0	89%	Query_216894
Mu3	4161	4161	100%	0.0	89%	Query_216893
ED98	4161	4161	100%	0.0	89%	Query_216892
MW2	4145	4145	100%	0.0	89%	Query_216896
MSSA476	4145	4145	100%	0.0	89%	Query_216895
ST228_16035	4128	4128	100%	0.0	89%	Query_216898
ST228_16125	4128	4128	100%	0.0	89%	Query_216897
MRSA_B51	4128	4128	100%	0.0	89%	Query_216877
ST228_18583	4120	4120	100%	0.0	89%	Query_216899
Z172	4100	4100	100%	0.0	88%	Query_216900
JKD6008	4084	4084	100%	0.0	88%	Query_216904
71193	4084	4084	100%	0.0	88%	Query_216903
08BA02176	4084	4084	100%	0.0	88%	Query_216902
CN1	4084	4084	100%	0.0	88%	Query_216901
TW20	4078	4078	100%	0.0	88%	Query_216905
MRSA_S62	4050	4050	100%	0.0	88%	Query_216880
T0131	4045	4045	100%	0.0	88%	Query_216906
MRSA_B53	4034	4034	100%	0.0	88%	Query_216879
6850	4032	4842	100%	0.0	89%	Query_216907
M1	4008	4008	100%	0.0	88%	Query_216908
MRSA_AS3	3978	3978	100%	0.0	88%	Query_216881
JKD6159	3973	3973	100%	0.0	88%	Query_216909
11819-97	3967	3967	100%	0.0	88%	Query_216910
USA300_FPR3757	3947	3947	100%	0.0	88%	Query_216914
USA300_TCH1516	3947	3947	100%	0.0	88%	Query_216913
USA300-ISMMS1	3947	3947	100%	0.0	88%	Query_216912
UA-S391_USA300	3947	3947	100%	0.0	88%	Query_216911
HO50960412	3940	4727	100%	0.0	89%	Query_216915
H-EMRSA-15	3934	4722	100%	0.0	89%	Query_216916
04-02981	3877	4692	100%	0.0	88%	Query_216919
ECT-R2	3877	4692	100%	0.0	88%	Query_216918
502A	3877	4692	100%	0.0	88%	Query_216917
JH9	3871	4681	100%	0.0	88%	Query_216921
JH1	3871	4681	100%	0.0	88%	Query_216920
ED133	3794	4432	100%	0.0	89%	Query_216922
Bmb9393	3788	4548	100%	0.0	88%	Query_216923
CA-347	3749	4509	100%	0.0	88%	Query_216924
Newman	3723	4489	100%	0.0	88%	Query_216926
XN108	3723	4445	100%	0.0	88%	Query_216925
COL	3718	4484	100%	0.0	88%	Query_216928
NRS100	3718	4484	100%	0.0	88%	Query_216927

bone sialoprotein-binding protein (*bbp*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
MW2	1341	1341	100%	0.0	100%	Query_69088
MSSA476	1341	1341	100%	0.0	100%	Query_69087
MRSA_OP100	1341	1341	100%	0.0	100%	Query_69086

enterotoxin H gene (*seh*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
MW2DNA	610	610	100%	9e-178	100%	Query_155833
MSSA476	610	610	100%	9e-178	100%	Query_155832
MRSA_B51	610	610	100%	9e-178	100%	Query_155827
MRSA_OP00	610	610	100%	9e-178	100%	Query_155826
N315	604	604	100%	4e-176	99%	Query_155865
USA300_FPR3757	604	604	100%	4e-176	99%	Query_155864
NCTC8325	604	604	100%	4e-176	99%	Query_155863
COL	604	604	100%	4e-176	99%	Query_155862
Mu50	604	604	100%	4e-176	99%	Query_155861
JH9	604	604	100%	4e-176	99%	Query_155860
JH1	604	604	100%	4e-176	99%	Query_155859
Newman	604	604	100%	4e-176	99%	Query_155858
Mu3	604	604	100%	4e-176	99%	Query_155857
USA300_TCH1516	604	604	100%	4e-176	99%	Query_155856
ED98	604	604	100%	4e-176	99%	Query_155855
TW20	604	604	100%	4e-176	99%	Query_155854
JKD6008	604	604	100%	4e-176	99%	Query_155853
04-02981	604	604	100%	4e-176	99%	Query_155852
ECT-R2	604	604	100%	4e-176	99%	Query_155851
T0131	604	604	100%	4e-176	99%	Query_155850
LGA251	604	604	100%	4e-176	99%	Query_155849
11819-97	604	604	100%	4e-176	99%	Query_155848
VC40	604	604	100%	4e-176	99%	Query_155847
HO50960412	604	604	100%	4e-176	99%	Query_155846
ST228_18035	604	604	100%	4e-176	99%	Query_155845
ST228_16125	604	604	100%	4e-176	99%	Query_155844
ST228_18583	604	604	100%	4e-176	99%	Query_155843
M1	604	604	100%	4e-176	99%	Query_155842
Bmb9393	604	604	100%	4e-176	99%	Query_155841
CN1	604	604	100%	4e-176	99%	Query_155840
Z172	604	604	100%	4e-176	99%	Query_155839
USA300-ISMM51	604	604	100%	4e-176	99%	Query_155838
NRS100	604	604	100%	4e-176	99%	Query_155837
H-EMRSA-15	604	604	100%	4e-176	99%	Query_155836
UA-S391_USA300	604	604	100%	4e-176	99%	Query_155835
XN108	604	604	100%	4e-176	99%	Query_155834
ST398	599	599	100%	2e-174	99%	Query_155870
ED133	599	599	100%	2e-174	99%	Query_155869
71193	599	599	100%	2e-174	99%	Query_155868
08BA02176	599	599	100%	2e-174	99%	Query_155867
502A	599	599	100%	2e-174	99%	Query_155866
MRSA_AS3	599	599	100%	2e-174	99%	Query_155831
MRSA_S62	599	599	100%	2e-174	99%	Query_155830
MRSA_B53	599	599	100%	2e-174	99%	Query_155829
MRSA252	593	593	100%	9e-173	99%	Query_155873
TCH60	593	593	100%	9e-173	99%	Query_155872
55/2053	593	593	100%	9e-173	99%	Query_155871
MRSA_IP70	593	593	100%	9e-173	99%	Query_155828
CA-347	588	588	100%	4e-171	99%	Query_155875
6850	588	588	100%	4e-171	99%	Query_155874
RF122	582	582	100%	2e-169	98%	Query_155880
JKD6159	582	582	100%	2e-169	98%	Query_155879
M013	582	582	100%	2e-169	98%	Query_155878
SA957	582	582	100%	2e-169	98%	Query_155877
SA40	582	582	100%	2e-169	98%	Query_155876
MSHR1132	259	259	90%	4e-72	82%	Query_155881

cysteine protease (*sspC*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> COL	1834	1834	100%	0.0	100%	Query_57190
<input type="checkbox"/> NRS100	1834	1834	100%	0.0	100%	Query_57189
<input type="checkbox"/> VC40	1829	1829	100%	0.0	99%	Query_57191
<input type="checkbox"/> MRSA_OP100	1818	1818	100%	0.0	99%	Query_57183
<input type="checkbox"/> MRSA_B51	1801	1801	100%	0.0	99%	Query_57184
<input type="checkbox"/> LGA251	1790	1790	100%	0.0	99%	Query_57192
<input type="checkbox"/> M013	1779	1779	100%	0.0	99%	Query_57195
<input type="checkbox"/> SA957	1779	1779	100%	0.0	99%	Query_57194
<input type="checkbox"/> SA40	1779	1779	100%	0.0	99%	Query_57193
<input type="checkbox"/> RF122	1773	1773	100%	0.0	99%	Query_57198
<input type="checkbox"/> ED133	1773	1773	100%	0.0	99%	Query_57197
<input type="checkbox"/> 6860	1773	1773	100%	0.0	99%	Query_57196
<input type="checkbox"/> MRSA_IP70	1622	1622	100%	0.0	96%	Query_57185
<input type="checkbox"/> ST398	1580	1580	100%	0.0	95%	Query_57200
<input type="checkbox"/> O88A02176	1580	1580	100%	0.0	95%	Query_57199
<input type="checkbox"/> MRSA_B53	1580	1580	100%	0.0	95%	Query_57186
<input type="checkbox"/> MRSA_S62	1574	1574	100%	0.0	95%	Query_57187
<input type="checkbox"/> MRSA_AS3	1572	1572	100%	0.0	95%	Query_57188
<input type="checkbox"/> USA300_TCH1516	1522	1522	82%	0.0	100%	Query_57232
<input type="checkbox"/> USA300_FPR3757	1522	1522	82%	0.0	100%	Query_57209
<input type="checkbox"/> Newman	1522	1522	82%	0.0	100%	Query_57208
<input type="checkbox"/> JKD6008	1522	1522	82%	0.0	100%	Query_57207
<input type="checkbox"/> T0131	1522	1522	82%	0.0	100%	Query_57206
<input type="checkbox"/> Bmb9393	1522	1522	82%	0.0	100%	Query_57205
<input type="checkbox"/> Z172	1522	1522	82%	0.0	100%	Query_57204
<input type="checkbox"/> USA300-ISMMS1	1522	1522	82%	0.0	100%	Query_57203
<input type="checkbox"/> UA-S391_USA300	1522	1522	82%	0.0	100%	Query_57202
<input type="checkbox"/> XN108	1522	1522	82%	0.0	100%	Query_57201
<input type="checkbox"/> NCTC8325	1517	1517	82%	0.0	99%	Query_57210
<input type="checkbox"/> M1	1509	1509	82%	0.0	99%	Query_57211
<input type="checkbox"/> MW2	1506	1506	82%	0.0	99%	Query_57212
<input type="checkbox"/> CN1	1495	1495	82%	0.0	99%	Query_57213
<input type="checkbox"/> N315	1489	1489	82%	0.0	99%	Query_57216
<input type="checkbox"/> ED98	1489	1489	82%	0.0	99%	Query_57215
<input type="checkbox"/> 502A	1489	1489	82%	0.0	99%	Query_57214
<input type="checkbox"/> 04-02961	1483	1483	82%	0.0	99%	Query_57233
<input type="checkbox"/> Mu50	1483	1483	82%	0.0	99%	Query_57222
<input type="checkbox"/> JH9	1483	1483	82%	0.0	99%	Query_57221
<input type="checkbox"/> JH1	1483	1483	82%	0.0	99%	Query_57220
<input type="checkbox"/> Mu3	1483	1483	82%	0.0	99%	Query_57219
<input type="checkbox"/> ECT-R2	1483	1483	82%	0.0	99%	Query_57218
<input type="checkbox"/> 11819-97	1483	1483	82%	0.0	99%	Query_57217
<input type="checkbox"/> JKD6199	1472	1472	82%	0.0	99%	Query_57225
<input type="checkbox"/> H050960412	1472	1472	82%	0.0	99%	Query_57224
<input type="checkbox"/> H-EMRSA-15	1472	1472	82%	0.0	99%	Query_57223
<input type="checkbox"/> 55/2053	1362	1362	82%	0.0	96%	Query_57226
<input type="checkbox"/> MRSA252	1351	1351	82%	0.0	96%	Query_57227
<input type="checkbox"/> TCH60	1345	1345	82%	0.0	96%	Query_57228
<input type="checkbox"/> 71193	1328	1328	82%	0.0	96%	Query_57229
<input type="checkbox"/> CA-347	1223	1223	82%	0.0	93%	Query_57230
<input type="checkbox"/> MSHR1132	1035	1035	82%	0.0	89%	Query_57231

beta-hemolysin gene (*hly*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> COL	1480	1480	100%	0.0	100%	Query_198004
<input type="checkbox"/> M013	1480	1480	100%	0.0	100%	Query_198003
<input type="checkbox"/> SA957	1480	1480	100%	0.0	100%	Query_198002
<input type="checkbox"/> SA40	1480	1480	100%	0.0	100%	Query_198001
<input type="checkbox"/> NRS100	1480	1480	100%	0.0	100%	Query_198000
<input type="checkbox"/> MRSA_OP100	1430	1430	100%	0.0	99%	Query_197999
<input type="checkbox"/> N315	414	414	99%	1e-118	77%	Query_198007
<input type="checkbox"/> Mu50	414	414	99%	1e-118	77%	Query_198006
<input type="checkbox"/> Mu3	414	414	99%	1e-118	77%	Query_198005
<input type="checkbox"/> RF122	403	403	99%	2e-115	76%	Query_198008
<input type="checkbox"/> H-EMRSA-15	392	392	99%	5e-112	76%	Query_198009
<input type="checkbox"/> ED133	387	387	99%	2e-110	76%	Query_198010
<input type="checkbox"/> MW2	359	359	99%	5e-102	75%	Query_198011

enterotoxin B (*seb*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> MRSA252	6305	6305	100%	0.0	100%	Query_241192
<input type="checkbox"/> TCH80	6122	6122	100%	0.0	99%	Query_241193
<input type="checkbox"/> MRSA_IP70	6067	6067	100%	0.0	99%	Query_241189
<input type="checkbox"/> 55/2053	5801	6760	100%	0.0	99%	Query_241194
<input type="checkbox"/> MRSA_OP100	5435	5435	100%	0.0	95%	Query_241190
<input type="checkbox"/> RF122	4870	5597	100%	0.0	94%	Query_241195
<input type="checkbox"/> MSHR1132	4638	4638	94%	0.0	93%	Query_241196
<input type="checkbox"/> LGA251	4211	4211	100%	0.0	89%	Query_241197
<input type="checkbox"/> M013	4200	4200	100%	0.0	89%	Query_241200
<input type="checkbox"/> SA957	4200	4200	100%	0.0	89%	Query_241199
<input type="checkbox"/> SA40	4200	4200	100%	0.0	89%	Query_241198
<input type="checkbox"/> N315	4167	4167	100%	0.0	89%	Query_241201
<input type="checkbox"/> Mu50	4161	4161	100%	0.0	89%	Query_241204
<input type="checkbox"/> Mu3	4161	4161	100%	0.0	89%	Query_241203
<input type="checkbox"/> ED98	4161	4161	100%	0.0	89%	Query_241202
<input type="checkbox"/> MW2	4145	4145	100%	0.0	89%	Query_241206
<input type="checkbox"/> MSSA476	4145	4145	100%	0.0	89%	Query_241205
<input type="checkbox"/> ST228 isolate 16035	4128	4128	100%	0.0	89%	Query_241208
<input type="checkbox"/> ST228 16125	4128	4128	100%	0.0	89%	Query_241207
<input type="checkbox"/> MRSA_B51	4128	4128	100%	0.0	89%	Query_241187
<input type="checkbox"/> ST228 18583	4120	4120	100%	0.0	89%	Query_241209
<input type="checkbox"/> Z172	4100	4100	100%	0.0	88%	Query_241210
<input type="checkbox"/> JKD8008	4084	4084	100%	0.0	88%	Query_241214
<input type="checkbox"/> 71193	4084	4084	100%	0.0	88%	Query_241213
<input type="checkbox"/> 08BA02176	4084	4084	100%	0.0	88%	Query_241212
<input type="checkbox"/> CN1	4084	4084	100%	0.0	88%	Query_241211
<input type="checkbox"/> TW20	4078	4078	100%	0.0	88%	Query_241215
<input type="checkbox"/> MRSA_S62	4050	4050	100%	0.0	88%	Query_241191
<input type="checkbox"/> T0131	4045	4045	100%	0.0	88%	Query_241216
<input type="checkbox"/> MRSA_B53	4034	4034	100%	0.0	88%	Query_241188
<input type="checkbox"/> 6850	4032	4842	100%	0.0	89%	Query_241217
<input type="checkbox"/> M1	4008	4008	100%	0.0	88%	Query_241218
<input type="checkbox"/> AS3	3978	3978	100%	0.0	88%	Query_241186
<input type="checkbox"/> JKD6159	3973	3973	100%	0.0	88%	Query_241219
<input type="checkbox"/> 11819-97	3967	3967	100%	0.0	88%	Query_241220
<input type="checkbox"/> USA300_FPR3757	3947	3947	100%	0.0	88%	Query_241224
<input type="checkbox"/> USA300_TCH1516	3947	3947	100%	0.0	88%	Query_241223
<input type="checkbox"/> USA300-ISMM51	3947	3947	100%	0.0	88%	Query_241222
<input type="checkbox"/> UA-S391_USA300	3947	3947	100%	0.0	88%	Query_241221
<input type="checkbox"/> HO50960412	3940	4727	100%	0.0	89%	Query_241225
<input type="checkbox"/> H-EMRSA-15	3934	4722	100%	0.0	89%	Query_241226
<input type="checkbox"/> 04-02981	3877	4692	100%	0.0	88%	Query_241229
<input type="checkbox"/> ECT-R2	3877	4692	100%	0.0	88%	Query_241228
<input type="checkbox"/> 502A	3877	4692	100%	0.0	88%	Query_241227
<input type="checkbox"/> JH9	3871	4681	100%	0.0	88%	Query_241231
<input type="checkbox"/> JH1	3871	4681	100%	0.0	88%	Query_241230
<input type="checkbox"/> ED133	3794	4432	100%	0.0	89%	Query_241232
<input type="checkbox"/> Bmb9393	3788	4548	100%	0.0	88%	Query_241233
<input type="checkbox"/> CA-347	3749	4509	100%	0.0	88%	Query_241234
<input type="checkbox"/> Newman	3723	4489	100%	0.0	88%	Query_241236
<input type="checkbox"/> XN108	3723	4445	100%	0.0	88%	Query_241235
<input type="checkbox"/> COL	3718	4484	100%	0.0	88%	Query_241238

delta-aminolevulinic acid dehydratase (*hemB*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> CA-347	865	865	100%	0.0	100%	Query_82778
<input type="checkbox"/> 71193	854	854	100%	0.0	99%	Query_82780
<input type="checkbox"/> 08BA02176	854	854	100%	0.0	99%	Query_82779
<input type="checkbox"/> MRSA_S62	854	854	100%	0.0	99%	Query_82776
<input type="checkbox"/> MRSA262	848	848	100%	0.0	99%	Query_82784
<input type="checkbox"/> ST398	848	848	100%	0.0	99%	Query_82783
<input type="checkbox"/> TCH60	848	848	100%	0.0	99%	Query_82782
<input type="checkbox"/> 55/2053	848	848	100%	0.0	99%	Query_82781
<input type="checkbox"/> MRSA_AS3	848	848	100%	0.0	99%	Query_82777
<input type="checkbox"/> MRSA_B53	848	848	100%	0.0	99%	Query_82775
<input type="checkbox"/> MRSA_IP70	848	848	100%	0.0	99%	Query_82774
<input type="checkbox"/> HO50960412	760	760	100%	0.0	96%	Query_82786
<input type="checkbox"/> H-EMRSA-15	760	760	100%	0.0	96%	Query_82785
<input type="checkbox"/> USA300_FPR3757	597	597	99%	1e-173	90%	Query_82802
<input type="checkbox"/> NCTC8325	597	597	99%	1e-173	90%	Query_82801
<input type="checkbox"/> COL	597	597	99%	1e-173	90%	Query_82800
<input type="checkbox"/> Newman	597	597	99%	1e-173	90%	Query_82799
<input type="checkbox"/> USA300_TCH1516	597	597	99%	1e-173	90%	Query_82798
<input type="checkbox"/> TW20	597	597	99%	1e-173	90%	Query_82797
<input type="checkbox"/> JKD6008	597	597	99%	1e-173	90%	Query_82796
<input type="checkbox"/> T0131	597	597	99%	1e-173	90%	Query_82795
<input type="checkbox"/> VC40	597	597	99%	1e-173	90%	Query_82794
<input type="checkbox"/> M1	597	597	99%	1e-173	90%	Query_82793
<input type="checkbox"/> Bmb9393	597	597	99%	1e-173	90%	Query_82792
<input type="checkbox"/> Z172	597	597	99%	1e-173	90%	Query_82791
<input type="checkbox"/> USA300-ISMM51	597	597	99%	1e-173	90%	Query_82790
<input type="checkbox"/> NRS100	597	597	99%	1e-173	90%	Query_82789
<input type="checkbox"/> UA-S391 USA300	597	597	99%	1e-173	90%	Query_82788
<input type="checkbox"/> XN108	597	597	99%	1e-173	90%	Query_82787
<input type="checkbox"/> MRSA_OP100	593	593	100%	2e-172	90%	Query_82772
<input type="checkbox"/> MW2	592	592	99%	7e-172	90%	Query_82805
<input type="checkbox"/> MSSA476	592	592	99%	7e-172	90%	Query_82804
<input type="checkbox"/> 11819-97	592	592	99%	7e-172	90%	Query_82803
<input type="checkbox"/> MRSA_B51	566	566	100%	4e-164	88%	Query_82773
<input type="checkbox"/> N315	564	564	99%	1e-163	88%	Query_82820
<input type="checkbox"/> Mu50	564	564	99%	1e-163	88%	Query_82819
<input type="checkbox"/> JH9	564	564	99%	1e-163	88%	Query_82818
<input type="checkbox"/> JH1	564	564	99%	1e-163	88%	Query_82817
<input type="checkbox"/> Mu3	564	564	99%	1e-163	88%	Query_82816
<input type="checkbox"/> ED98	564	564	99%	1e-163	88%	Query_82815
<input type="checkbox"/> ED133	564	564	99%	1e-163	88%	Query_82814
<input type="checkbox"/> 04-02981	564	564	99%	1e-163	88%	Query_82813
<input type="checkbox"/> ECT-R2	564	564	99%	1e-163	88%	Query_82812
<input type="checkbox"/> LGA251	564	564	99%	1e-163	88%	Query_82811
<input type="checkbox"/> ST228_16035	564	564	99%	1e-163	88%	Query_82810
<input type="checkbox"/> ST228_16125	564	564	99%	1e-163	88%	Query_82809
<input type="checkbox"/> ST228_18583	564	564	99%	1e-163	88%	Query_82808
<input type="checkbox"/> CN1	564	564	99%	1e-163	88%	Query_82807
<input type="checkbox"/> 502A	564	564	99%	1e-163	88%	Query_82806
<input type="checkbox"/> M013	547	547	99%	1e-158	88%	Query_82823
<input type="checkbox"/> SA957	547	547	99%	1e-158	88%	Query_82822
<input type="checkbox"/> SA40	547	547	99%	1e-158	88%	Query_82821
<input type="checkbox"/> JKD6159	542	542	99%	7e-157	88%	Query_82825
<input type="checkbox"/> 6850	542	542	99%	7e-157	88%	Query_82824

multidrug resistance efflux pump (*sepA*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> NCTC8325	5142	9096	100%	0.0	100%	Query_63853
<input type="checkbox"/> VC40	5136	9063	100%	0.0	99%	Query_63854
<input type="checkbox"/> TW20	5120	8606	100%	0.0	99%	Query_63855
<input type="checkbox"/> Z172	5108	8578	100%	0.0	99%	Query_63856
<input type="checkbox"/> USA300_FPR3757	5042	8619	100%	0.0	99%	Query_63862
<input type="checkbox"/> COL	5042	8619	100%	0.0	99%	Query_63861
<input type="checkbox"/> USA300_TCH1516	5042	8619	100%	0.0	99%	Query_63860
<input type="checkbox"/> USA300-ISMMS1	5042	8619	100%	0.0	99%	Query_63859
<input type="checkbox"/> NRS100	5042	8619	100%	0.0	99%	Query_63858
<input type="checkbox"/> UA-S391_USA300	5042	8619	100%	0.0	99%	Query_63857
<input type="checkbox"/> Newman	5036	8613	100%	0.0	99%	Query_63865
<input type="checkbox"/> JKD6008	5036	8613	100%	0.0	99%	Query_63864
<input type="checkbox"/> Bmb9393	5036	8613	100%	0.0	99%	Query_63863
<input type="checkbox"/> XN108	5003	9126	100%	0.0	99%	Query_63866
<input type="checkbox"/> MRSA_OP100	4721	9286	100%	0.0	97%	Query_63847
<input type="checkbox"/> MSSA476	4567	9594	100%	0.0	96%	Query_63867
<input type="checkbox"/> M1	4383	8015	100%	0.0	95%	Query_63868
<input type="checkbox"/> MW2	4324	8314	100%	0.0	94%	Query_63869
<input type="checkbox"/> MRSA_IP70	4037	7189	100%	0.0	93%	Query_63849
<input type="checkbox"/> MRSA_S62	4002	6142	100%	0.0	93%	Query_63851
<input type="checkbox"/> T0131	3951	7255	100%	0.0	96%	Query_63870
<input type="checkbox"/> MRSA_AS3	3945	5968	100%	0.0	92%	Query_63852
<input type="checkbox"/> MRSA_B53	3923	6449	100%	0.0	92%	Query_63850
<input type="checkbox"/> RF122	3797	7792	100%	0.0	92%	Query_63871
<input type="checkbox"/> MRSA_B51	3762	7076	100%	0.0	91%	Query_63848
<input type="checkbox"/> SA40	3747	7529	100%	0.0	93%	Query_63874
<input type="checkbox"/> SA957	3747	7529	100%	0.0	93%	Query_63872
<input type="checkbox"/> M013	3637	7418	97%	0.0	93%	Query_63873
<input type="checkbox"/> Mu50	3624	8181	100%	0.0	90%	Query_63876
<input type="checkbox"/> Mu3	3624	8181	100%	0.0	90%	Query_63875
<input type="checkbox"/> 08BA02176	3602	6795	100%	0.0	90%	Query_63877
<input type="checkbox"/> CN1	3598	5898	100%	0.0	92%	Query_63878
<input type="checkbox"/> ST398	3578	5779	100%	0.0	90%	Query_63879
<input type="checkbox"/> 6850	3555	5916	100%	0.0	92%	Query_63880
<input type="checkbox"/> HO50960412	3513	10761	100%	0.0	89%	Query_63883
<input type="checkbox"/> CA-347	3513	7810	100%	0.0	91%	Query_63882
<input type="checkbox"/> H-EMRSA-15	3513	10761	100%	0.0	89%	Query_63881
<input type="checkbox"/> ED133	3452	7074	100%	0.0	89%	Query_63884
<input type="checkbox"/> TCH60	3432	11119	100%	0.0	91%	Query_63885
<input type="checkbox"/> MRSA252	3400	8422	100%	0.0	91%	Query_63886
<input type="checkbox"/> 55/2053	3393	8413	100%	0.0	90%	Query_63887
<input type="checkbox"/> 71193	3310	7084	100%	0.0	90%	Query_63888
<input type="checkbox"/> N315	3181	7544	100%	0.0	89%	Query_63890
<input type="checkbox"/> 502A	3181	7544	100%	0.0	89%	Query_63889
<input type="checkbox"/> 04-02981	3169	7784	100%	0.0	90%	Query_63892
<input type="checkbox"/> ECT-R2	3169	7784	100%	0.0	90%	Query_63891
<input type="checkbox"/> JH9	3164	7778	100%	0.0	90%	Query_63895
<input type="checkbox"/> JH1	3164	7778	100%	0.0	90%	Query_63894
<input type="checkbox"/> ED98	3153	8516	99%	0.0	89%	Query_63893
<input type="checkbox"/> LGA251	3131	7658	100%	0.0	89%	Query_63896
<input type="checkbox"/> 11819-97	3125	11366	100%	0.0	89%	Query_63897
<input type="checkbox"/> JKD6159	2950	7828	100%	0.0	88%	Query_63898
<input type="checkbox"/> ST228_16035	2776	7903	91%	0.0	88%	Query_63901
<input type="checkbox"/> ST228_16125	2776	7903	91%	0.0	88%	Query_63900
<input type="checkbox"/> ST228_18583	2776	7903	91%	0.0	88%	Query_63899

clumping factor (*clf*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
USA300_FPR3757	804	804	100%	0.0	100%	Query_179282
NCTC8325	804	804	100%	0.0	100%	Query_179281
COL	804	804	100%	0.0	100%	Query_179280
Newman	804	804	100%	0.0	100%	Query_179279
USA300_TCH1516	804	804	100%	0.0	100%	Query_179278
TW20	804	804	100%	0.0	100%	Query_179277
JKD6008	804	804	100%	0.0	100%	Query_179276
T0131	804	804	100%	0.0	100%	Query_179275
VC40	804	804	100%	0.0	100%	Query_179274
M1	804	804	100%	0.0	100%	Query_179273
Bmb9393	804	804	100%	0.0	100%	Query_179272
Z172	804	804	100%	0.0	100%	Query_179271
USA300-ISMS1	804	804	100%	0.0	100%	Query_179270
NRS100	804	804	100%	0.0	100%	Query_179269
UA-S391_USA300	804	804	100%	0.0	100%	Query_179268
XN108	804	804	100%	0.0	100%	Query_179267
MW2	798	798	100%	0.0	99%	Query_179285
MSSA476	798	798	100%	0.0	99%	Query_179284
11819-97	798	798	100%	0.0	99%	Query_179283
MRSA_OP100	798	798	100%	0.0	99%	Query_179262
N315	782	782	100%	0.0	99%	Query_179298
Mu50	782	782	100%	0.0	99%	Query_179297
JH9	782	782	100%	0.0	99%	Query_179296
JH1	782	782	100%	0.0	99%	Query_179295
Mu3	782	782	100%	0.0	99%	Query_179294
ED98	782	782	100%	0.0	99%	Query_179293
04-02981	782	782	100%	0.0	99%	Query_179292
ECT-R2	782	782	100%	0.0	99%	Query_179291
ST228_16035	782	782	100%	0.0	99%	Query_179290
ST228_16125	782	782	100%	0.0	99%	Query_179289
ST228_18583	782	782	100%	0.0	99%	Query_179288
CN1	782	782	100%	0.0	99%	Query_179287
502A	782	782	100%	0.0	99%	Query_179286
H050960412	776	776	100%	0.0	99%	Query_179300
H-EMRSA-15	776	776	100%	0.0	99%	Query_179299
MRSA_B51	771	771	100%	0.0	99%	Query_179263
ED133	760	760	100%	0.0	98%	Query_179301
JKD6159	754	754	100%	0.0	98%	Query_179304
LGA251	754	754	100%	0.0	98%	Query_179303
8850	754	754	100%	0.0	98%	Query_179302
MRSA252	749	749	100%	0.0	98%	Query_179309
RF122	749	749	100%	0.0	98%	Query_179308
TCH60	749	749	100%	0.0	98%	Query_179307
CA-347	749	749	100%	0.0	98%	Query_179306
55/2053	749	749	100%	0.0	98%	Query_179305
MRSA_IP70	749	749	100%	0.0	98%	Query_179264
SA40	743	743	100%	0.0	97%	Query_179314
M013	743	743	100%	0.0	97%	Query_179311
SA957	743	743	100%	0.0	97%	Query_179310
71193	737	737	100%	0.0	97%	Query_179312
MRSA_AS3	737	737	100%	0.0	97%	Query_179266
MRSA_B53	737	737	100%	0.0	97%	Query_179265
08BA02176	732	732	100%	0.0	97%	Query_179313
ST398	673	673	91%	0.0	97%	Query_179315

extracellular adherence protein (*eap*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> MRSA252	1435	1435	100%	0.0	100%	Query_245771
<input type="checkbox"/> TCH60	1435	1435	100%	0.0	100%	Query_245770
<input type="checkbox"/> 55/2053	1435	1435	100%	0.0	100%	Query_245769
<input type="checkbox"/> MRSA IP70	1435	1435	100%	0.0	100%	Query_245768
<input type="checkbox"/> RF122	1384	1384	100%	0.0	99%	Query_245772
<input type="checkbox"/> N315	1352	1352	100%	0.0	98%	Query_245786
<input type="checkbox"/> Mu50	1352	1352	100%	0.0	98%	Query_245785
<input type="checkbox"/> JH9	1352	1352	100%	0.0	98%	Query_245784
<input type="checkbox"/> JH1	1352	1352	100%	0.0	98%	Query_245783
<input type="checkbox"/> Mu3	1352	1352	100%	0.0	98%	Query_245782
<input type="checkbox"/> ED98	1352	1352	100%	0.0	98%	Query_245781
<input type="checkbox"/> ECT-R2	1352	1352	100%	0.0	98%	Query_245780
<input type="checkbox"/> HO50960412	1352	1352	100%	0.0	98%	Query_245779
<input type="checkbox"/> ST228_16035	1352	1352	100%	0.0	98%	Query_245778
<input type="checkbox"/> ST228_16125	1352	1352	100%	0.0	98%	Query_245777
<input type="checkbox"/> ST228_18583	1352	1352	100%	0.0	98%	Query_245776
<input type="checkbox"/> CA_347	1352	1352	100%	0.0	98%	Query_245775
<input type="checkbox"/> 502A	1352	1352	100%	0.0	98%	Query_245774
<input type="checkbox"/> H-EMRSA-15	1352	1352	100%	0.0	98%	Query_245773
<input type="checkbox"/> MRSA_B51	1352	1352	100%	0.0	98%	Query_245767
<input type="checkbox"/> 04-02981	1347	1347	100%	0.0	98%	Query_245788
<input type="checkbox"/> CN1	1347	1347	100%	0.0	98%	Query_245787
<input type="checkbox"/> MSHR1132	1297	1297	100%	0.0	97%	Query_245789

entertotoxin type G (*seg*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> MRSA_IP70	760	760	100%	0.0	84%	Query_212553
<input type="checkbox"/> MW2	756	756	99%	0.0	84%	Query_212567
<input type="checkbox"/> MSSA476	756	756	99%	0.0	84%	Query_212566
<input type="checkbox"/> Mu50	756	756	99%	0.0	84%	Query_212565
<input type="checkbox"/> Newman	756	756	99%	0.0	84%	Query_212564
<input type="checkbox"/> Mu3	756	756	99%	0.0	84%	Query_212563
<input type="checkbox"/> TW20	756	756	99%	0.0	84%	Query_212562
<input type="checkbox"/> JKD6008	756	756	99%	0.0	84%	Query_212561
<input type="checkbox"/> T0131	756	756	99%	0.0	84%	Query_212560
<input type="checkbox"/> ST228_16035	756	756	99%	0.0	84%	Query_212559
<input type="checkbox"/> ST228_16125	756	756	99%	0.0	84%	Query_212558
<input type="checkbox"/> ST228_18583	756	756	99%	0.0	84%	Query_212557
<input type="checkbox"/> M1	756	756	99%	0.0	84%	Query_212556
<input type="checkbox"/> Z172	756	756	99%	0.0	84%	Query_212555
<input type="checkbox"/> XN108	756	756	99%	0.0	84%	Query_212554
<input type="checkbox"/> MRSA_OP100	754	754	100%	0.0	84%	Query_212552
<input type="checkbox"/> ED133	704	704	99%	0.0	83%	Query_212569

entertoxin E (*entE*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> MW2	1735	1735	100%	0.0	100%	Query_158704
<input type="checkbox"/> USA300_FPR3757	1735	1735	100%	0.0	100%	Query_158703
<input type="checkbox"/> USA300_TCH1516	1735	1735	100%	0.0	100%	Query_158702
<input type="checkbox"/> JKD6159	1735	1735	100%	0.0	100%	Query_158701
<input type="checkbox"/> USA300-ISMM51	1735	1735	100%	0.0	100%	Query_158700
<input type="checkbox"/> TCH60	1724	1724	100%	0.0	99%	Query_158708
<input type="checkbox"/> M013	1724	1724	100%	0.0	99%	Query_158707
<input type="checkbox"/> 55/2053	1724	1724	100%	0.0	99%	Query_158706
<input type="checkbox"/> SA957	1724	1724	100%	0.0	99%	Query_158705
<input type="checkbox"/> MRSA_OP100	1724	1724	100%	0.0	99%	Query_158694
<input type="checkbox"/> 11819-97	1712	1712	100%	0.0	99%	Query_158709

Panton-Valentine leukocidin - *lukS-PV*

Alignments Download Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> USA300_FPR3757	1807	1807	100%	0.0	100%	Query_115258
<input type="checkbox"/> USA300_TCH1516	1807	1807	100%	0.0	100%	Query_115257
<input type="checkbox"/> JKD6159	1807	1807	100%	0.0	100%	Query_115256
<input type="checkbox"/> USA300-ISMM51	1807	1807	100%	0.0	100%	Query_115255
<input type="checkbox"/> UA-S391_USA300	1807	1807	100%	0.0	100%	Query_115254
<input type="checkbox"/> MW2DNA	1801	1801	100%	0.0	99%	Query_115261
<input type="checkbox"/> TCH60	1801	1801	100%	0.0	99%	Query_115260
<input type="checkbox"/> 11819-97	1801	1801	100%	0.0	99%	Query_115259
<input type="checkbox"/> MRSA_OP100	1801	1801	100%	0.0	99%	Query_115249
<input type="checkbox"/> M013	1796	1796	100%	0.0	99%	Query_115264
<input type="checkbox"/> 55/2053	1796	1796	100%	0.0	99%	Query_115263
<input type="checkbox"/> SA957	1796	1796	100%	0.0	99%	Query_115262
<input type="checkbox"/> MRSA_B51	870	870	100%	0.0	83%	Query_115250
<input type="checkbox"/> N315	854	854	99%	0.0	83%	Query_115276
<input type="checkbox"/> Mu50	854	854	99%	0.0	83%	Query_115275
<input type="checkbox"/> Mu3DNA	854	854	99%	0.0	83%	Query_115272
<input type="checkbox"/> ED98	854	854	99%	0.0	83%	Query_115271
<input type="checkbox"/> 04-02981	854	854	99%	0.0	83%	Query_115270
<input type="checkbox"/> ECT-R2	854	854	99%	0.0	83%	Query_115269
<input type="checkbox"/> ST228_16035	854	854	99%	0.0	83%	Query_115268
<input type="checkbox"/> ST228_16125	854	854	99%	0.0	83%	Query_115267
<input type="checkbox"/> ST228_18583	854	854	99%	0.0	83%	Query_115266
<input type="checkbox"/> 502A	854	854	99%	0.0	83%	Query_115265
<input type="checkbox"/> Newman	837	837	99%	0.0	82%	Query_115281
<input type="checkbox"/> Bmb6393	837	837	99%	0.0	82%	Query_115280
<input type="checkbox"/> CN1	837	837	99%	0.0	82%	Query_115279
<input type="checkbox"/> Z172	837	837	99%	0.0	82%	Query_115278
<input type="checkbox"/> XN108	837	837	99%	0.0	82%	Query_115277

Panton-Valentine leukocidin subunit F (*lukF-PV*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> JKD8159	638	638	100%	0.0	100%	Query_228678
<input type="checkbox"/> ED133	625	625	99%	0.0	99%	Query_228679
<input type="checkbox"/> ST398	621	621	100%	0.0	99%	Query_228682
<input type="checkbox"/> 71193	621	621	100%	0.0	99%	Query_228681
<input type="checkbox"/> 08BA02176	621	621	100%	0.0	99%	Query_228680
<input type="checkbox"/> MRSA_AS3	621	621	100%	0.0	99%	Query_228677
<input type="checkbox"/> MRSA_S62	621	621	100%	0.0	99%	Query_228676
<input type="checkbox"/> MRSA_B53	621	621	100%	0.0	99%	Query_228675
<input type="checkbox"/> RF122	619	619	99%	2e-180	99%	Query_228683
<input type="checkbox"/> MW2	616	616	100%	2e-179	99%	Query_228685
<input type="checkbox"/> MSSA476	616	616	100%	2e-179	99%	Query_228684
<input type="checkbox"/> MRSA_OP100	616	616	100%	2e-179	99%	Query_228672
<input type="checkbox"/> MRSA_IP70	610	610	100%	1e-177	99%	Query_228674
<input type="checkbox"/> MRSA_B51	610	610	100%	1e-177	99%	Query_228673
<input type="checkbox"/> MRSA252	608	608	99%	4e-177	99%	Query_228694
<input type="checkbox"/> TW20	608	608	99%	4e-177	99%	Query_228693
<input type="checkbox"/> JKD8008	608	608	99%	4e-177	99%	Query_228692
<input type="checkbox"/> TCH80	608	608	99%	4e-177	99%	Query_228691
<input type="checkbox"/> T0131	608	608	99%	4e-177	99%	Query_228690
<input type="checkbox"/> Bmb9393	608	608	99%	4e-177	99%	Query_228689
<input type="checkbox"/> 55/2053	608	608	99%	4e-177	99%	Query_228688
<input type="checkbox"/> Z172	608	608	99%	4e-177	99%	Query_228687
<input type="checkbox"/> XN108	608	608	99%	4e-177	99%	Query_228686
<input type="checkbox"/> N315	603	603	99%	2e-175	98%	Query_228716
<input type="checkbox"/> USA300_FPR3757	603	603	99%	2e-175	98%	Query_228715
<input type="checkbox"/> NCTC8325	603	603	99%	2e-175	98%	Query_228714
<input type="checkbox"/> COL	603	603	99%	2e-175	98%	Query_228713
<input type="checkbox"/> Mu50	603	603	99%	2e-175	98%	Query_228712
<input type="checkbox"/> JH9	603	603	99%	2e-175	98%	Query_228711
<input type="checkbox"/> JH1	603	603	99%	2e-175	98%	Query_228710
<input type="checkbox"/> Newman	603	603	99%	2e-175	98%	Query_228709
<input type="checkbox"/> Mu3	603	603	99%	2e-175	98%	Query_228708
<input type="checkbox"/> USA300_TCH1516	603	603	99%	2e-175	98%	Query_228707
<input type="checkbox"/> ED98	603	603	99%	2e-175	98%	Query_228706
<input type="checkbox"/> 04-02981	603	603	99%	2e-175	98%	Query_228705
<input type="checkbox"/> ECT-R2	603	603	99%	2e-175	98%	Query_228704
<input type="checkbox"/> VC40	603	603	99%	2e-175	98%	Query_228703
<input type="checkbox"/> ST228_16035	603	603	99%	2e-175	98%	Query_228702
<input type="checkbox"/> ST228_18583	603	603	99%	2e-175	98%	Query_228701
<input type="checkbox"/> ST228_16125	603	603	99%	2e-175	98%	Query_228700
<input type="checkbox"/> M1	603	603	99%	2e-175	98%	Query_228699
<input type="checkbox"/> USA300-ISMM51	603	603	99%	2e-175	98%	Query_228698
<input type="checkbox"/> 502A	603	603	99%	2e-175	98%	Query_228697
<input type="checkbox"/> NRS100	603	603	99%	2e-175	98%	Query_228696
<input type="checkbox"/> UA-S391_USA300	603	603	99%	2e-175	98%	Query_228695
<input type="checkbox"/> 11819-97	597	597	99%	8e-174	98%	Query_228721
<input type="checkbox"/> HO50980412	597	597	99%	8e-174	98%	Query_228720
<input type="checkbox"/> CA-347	597	597	99%	8e-174	98%	Query_228719
<input type="checkbox"/> CN1	597	597	99%	8e-174	98%	Query_228718
<input type="checkbox"/> H-EMRSA-15	597	597	99%	8e-174	98%	Query_228717
<input type="checkbox"/> LGA251	592	592	99%	4e-172	98%	Query_228725
<input type="checkbox"/> M013	592	592	99%	4e-172	98%	Query_228724
<input type="checkbox"/> SA957	592	592	99%	4e-172	98%	Query_228723
<input type="checkbox"/> SA40	592	592	99%	4e-172	98%	Query_228722
<input type="checkbox"/> 6850	586	586	99%	2e-170	97%	Query_228726

Staphylococcal complement inhibitor SCIN (*scn*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> N315	1823	1823	100%	0.0	100%	Query_229902
<input type="checkbox"/> Mu50	1823	1823	100%	0.0	100%	Query_229901
<input type="checkbox"/> JH9	1823	1823	100%	0.0	100%	Query_229900
<input type="checkbox"/> JH1	1823	1823	100%	0.0	100%	Query_229899
<input type="checkbox"/> Mu3	1823	1823	100%	0.0	100%	Query_229898
<input type="checkbox"/> ED98	1823	1823	100%	0.0	100%	Query_229897
<input type="checkbox"/> 04-02981	1823	1823	100%	0.0	100%	Query_229896
<input type="checkbox"/> ST228_16035	1823	1823	100%	0.0	100%	Query_229895
<input type="checkbox"/> ST228_16125	1823	1823	100%	0.0	100%	Query_229894
<input type="checkbox"/> ST228_18583	1823	1823	100%	0.0	100%	Query_229893
<input type="checkbox"/> 502A	1823	1823	100%	0.0	100%	Query_229892
<input type="checkbox"/> ECT-R2	1818	1818	100%	0.0	99%	Query_229903
<input type="checkbox"/> 11819-97	1796	1796	100%	0.0	99%	Query_229904
<input type="checkbox"/> ST398	1718	1718	100%	0.0	98%	Query_229907
<input type="checkbox"/> Z1193	1718	1718	100%	0.0	98%	Query_229906
<input type="checkbox"/> CA-347	1718	1718	100%	0.0	98%	Query_229905
<input type="checkbox"/> MRSA_AS3	1718	1718	100%	0.0	98%	Query_229891
<input type="checkbox"/> MRSA_S62	1718	1718	100%	0.0	98%	Query_229890
<input type="checkbox"/> MRSA_B53	1718	1718	100%	0.0	98%	Query_229889
<input type="checkbox"/> MRSA_B51	1718	1718	100%	0.0	98%	Query_229887
<input type="checkbox"/> 08BA02176	1712	1712	100%	0.0	98%	Query_229908
<input type="checkbox"/> ED133	1690	1690	100%	0.0	98%	Query_229909
<input type="checkbox"/> USA300_FPR3757	1685	1685	100%	0.0	97%	Query_229919
<input type="checkbox"/> NCTC-8325	1685	1685	100%	0.0	97%	Query_229918
<input type="checkbox"/> COL	1685	1685	100%	0.0	97%	Query_229917
<input type="checkbox"/> Newman	1685	1685	100%	0.0	97%	Query_229916
<input type="checkbox"/> USA300_TCH1516	1685	1685	100%	0.0	97%	Query_229915
<input type="checkbox"/> VC40	1685	1685	100%	0.0	97%	Query_229914
<input type="checkbox"/> M1	1685	1685	100%	0.0	97%	Query_229913
<input type="checkbox"/> USA300-ISMM51	1685	1685	100%	0.0	97%	Query_229912
<input type="checkbox"/> NRS-100	1685	1685	100%	0.0	97%	Query_229911
<input type="checkbox"/> UA-S391_USA300	1685	1685	100%	0.0	97%	Query_229910
<input type="checkbox"/> MW2	1679	1679	100%	0.0	97%	Query_229921
<input type="checkbox"/> JKD6159	1679	1679	100%	0.0	97%	Query_229920
<input type="checkbox"/> MRSA_OP100	1679	1679	100%	0.0	97%	Query_229886
<input type="checkbox"/> CN1	1674	1674	100%	0.0	97%	Query_229922
<input type="checkbox"/> MSSA476	1672	1672	100%	0.0	97%	Query_229923
<input type="checkbox"/> LGA251	1663	1663	100%	0.0	97%	Query_229924
<input type="checkbox"/> MRSA252	1652	1652	100%	0.0	97%	Query_229933
<input type="checkbox"/> TW20	1652	1652	100%	0.0	97%	Query_229932
<input type="checkbox"/> JKD6008	1652	1652	100%	0.0	97%	Query_229931
<input type="checkbox"/> TCH60	1652	1652	100%	0.0	97%	Query_229930
<input type="checkbox"/> T0131	1652	1652	100%	0.0	97%	Query_229929
<input type="checkbox"/> Bmb9393	1652	1652	100%	0.0	97%	Query_229928
<input type="checkbox"/> 55-2053	1652	1652	100%	0.0	97%	Query_229927
<input type="checkbox"/> Z172	1652	1652	100%	0.0	97%	Query_229926
<input type="checkbox"/> XN108	1652	1652	100%	0.0	97%	Query_229925
<input type="checkbox"/> MRSA_IP70	1652	1652	100%	0.0	97%	Query_229888
<input type="checkbox"/> M013	1646	1646	100%	0.0	97%	Query_229936
<input type="checkbox"/> SA957	1646	1646	100%	0.0	97%	Query_229935
<input type="checkbox"/> SA40	1646	1646	100%	0.0	97%	Query_229934
<input type="checkbox"/> HO-50960412	1640	1640	100%	0.0	97%	Query_229938
<input type="checkbox"/> H-EMRSA-15	1640	1640	100%	0.0	97%	Query_229937
<input type="checkbox"/> 6850	1624	1624	100%	0.0	96%	Query_229939
<input type="checkbox"/> RF122	1618	1618	100%	0.0	96%	Query_229940

1-phosphatidylinositol-phosphodiesterase-precursor (*plc*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> MRSA252	1823	1823	100%	0.0	100%	Query_43134
<input type="checkbox"/> TW20	1823	1823	100%	0.0	100%	Query_43133
<input type="checkbox"/> JKD6008	1823	1823	100%	0.0	100%	Query_43132
<input type="checkbox"/> TCH60	1823	1823	100%	0.0	100%	Query_43131
<input type="checkbox"/> T0131	1823	1823	100%	0.0	100%	Query_43130
<input type="checkbox"/> Bmb9393	1823	1823	100%	0.0	100%	Query_43129
<input type="checkbox"/> 55/2053	1823	1823	100%	0.0	100%	Query_43128
<input type="checkbox"/> Z172	1823	1823	100%	0.0	100%	Query_43127
<input type="checkbox"/> XN108	1823	1823	100%	0.0	100%	Query_43126
<input type="checkbox"/> MRSA_IP70	1823	1823	100%	0.0	100%	Query_43122
<input type="checkbox"/> 6850	1674	1674	100%	0.0	97%	Query_43135
<input type="checkbox"/> 11819-97	1668	1668	100%	0.0	97%	Query_43136
<input type="checkbox"/> JKD6159	1657	1657	100%	0.0	97%	Query_43137
<input type="checkbox"/> N315	1652	1652	100%	0.0	97%	Query_43149
<input type="checkbox"/> Mu50	1652	1652	100%	0.0	97%	Query_43148
<input type="checkbox"/> JH9	1652	1652	100%	0.0	97%	Query_43147
<input type="checkbox"/> JH1	1652	1652	100%	0.0	97%	Query_43146
<input type="checkbox"/> Mu3	1652	1652	100%	0.0	97%	Query_43145
<input type="checkbox"/> ED98	1652	1652	100%	0.0	97%	Query_43144
<input type="checkbox"/> 04-02981	1652	1652	100%	0.0	97%	Query_43143
<input type="checkbox"/> ST228_16035	1652	1652	100%	0.0	97%	Query_43142
<input type="checkbox"/> ST228_16125	1652	1652	100%	0.0	97%	Query_43141
<input type="checkbox"/> ST228_18583	1652	1652	100%	0.0	97%	Query_43140
<input type="checkbox"/> CA-347	1652	1652	100%	0.0	97%	Query_43139
<input type="checkbox"/> 502A	1652	1652	100%	0.0	97%	Query_43138
<input type="checkbox"/> ED133	1646	1646	100%	0.0	97%	Query_43152
<input type="checkbox"/> ECT-R2	1646	1646	100%	0.0	97%	Query_43151
<input type="checkbox"/> HO50980412	1646	1646	100%	0.0	97%	Query_43150
<input type="checkbox"/> H-EMRSA-15	1640	1640	100%	0.0	97%	Query_43153
<input type="checkbox"/> USA300_FPR3757	1635	1635	100%	0.0	97%	Query_43164
<input type="checkbox"/> NCTC8325	1635	1635	100%	0.0	97%	Query_43163
<input type="checkbox"/> COL	1635	1635	100%	0.0	97%	Query_43162
<input type="checkbox"/> Newman	1635	1635	100%	0.0	97%	Query_43161
<input type="checkbox"/> USA300_TCH1516	1635	1635	100%	0.0	97%	Query_43160
<input type="checkbox"/> LGA251	1635	1635	100%	0.0	97%	Query_43159
<input type="checkbox"/> VC40	1635	1635	100%	0.0	97%	Query_43158
<input type="checkbox"/> M1	1635	1635	100%	0.0	97%	Query_43157
<input type="checkbox"/> USA300-ISMM51	1635	1635	100%	0.0	97%	Query_43156
<input type="checkbox"/> NRS100	1635	1635	100%	0.0	97%	Query_43155
<input type="checkbox"/> UA-S391_USA300	1635	1635	100%	0.0	97%	Query_43154
<input type="checkbox"/> MW2	1629	1629	100%	0.0	96%	Query_43168
<input type="checkbox"/> M013	1629	1629	100%	0.0	96%	Query_43167
<input type="checkbox"/> SA957	1629	1629	100%	0.0	96%	Query_43166
<input type="checkbox"/> SA40	1629	1629	100%	0.0	96%	Query_43165
<input type="checkbox"/> MRSA_OP100	1629	1629	100%	0.0	96%	Query_43120
<input type="checkbox"/> ST398	1624	1624	100%	0.0	96%	Query_43170
<input type="checkbox"/> 71193	1624	1624	100%	0.0	96%	Query_43169
<input type="checkbox"/> MRSA_AS3	1624	1624	100%	0.0	96%	Query_43125
<input type="checkbox"/> MRSA_S62	1624	1624	100%	0.0	96%	Query_43124
<input type="checkbox"/> MRSA_B53	1624	1624	100%	0.0	96%	Query_43123
<input type="checkbox"/> MRSA_B51	1624	1624	100%	0.0	96%	Query_43121
<input type="checkbox"/> MSSA476	1622	1622	100%	0.0	96%	Query_43171
<input type="checkbox"/> 08BA02176	1618	1618	100%	0.0	96%	Query_43172
<input type="checkbox"/> CN1	1613	1613	100%	0.0	96%	Query_43173
<input type="checkbox"/> RF122	1602	1602	100%	0.0	96%	Query_43174

1-phosphatidylinositol phosphodiesterase (*p/c*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> N315	920	920	100%	0.0	100%	Query_16844
<input type="checkbox"/> Mu50	920	920	100%	0.0	100%	Query_16843
<input type="checkbox"/> JH9	920	920	100%	0.0	100%	Query_16842
<input type="checkbox"/> JH1	920	920	100%	0.0	100%	Query_16841
<input type="checkbox"/> Mu3	920	920	100%	0.0	100%	Query_16840
<input type="checkbox"/> ED98	920	920	100%	0.0	100%	Query_16839
<input type="checkbox"/> 04-02981	920	920	100%	0.0	100%	Query_16838
<input type="checkbox"/> ECT-R2	920	920	100%	0.0	100%	Query_16837
<input type="checkbox"/> ST228_16035	920	920	100%	0.0	100%	Query_16836
<input type="checkbox"/> ST228_16125	920	920	100%	0.0	100%	Query_16835
<input type="checkbox"/> ST228_18583	920	920	100%	0.0	100%	Query_16834
<input type="checkbox"/> 502A	920	920	100%	0.0	100%	Query_16833
<input type="checkbox"/> MRSA_B51	904	904	100%	0.0	99%	Query_16828
<input type="checkbox"/> USA300_FPR3757	898	898	100%	0.0	99%	Query_16860
<input type="checkbox"/> NCTC8325	898	898	100%	0.0	99%	Query_16859
<input type="checkbox"/> COL	898	898	100%	0.0	99%	Query_16858
<input type="checkbox"/> Newman	898	898	100%	0.0	99%	Query_16857
<input type="checkbox"/> USA300_TCH1516	898	898	100%	0.0	99%	Query_16856
<input type="checkbox"/> TW20	898	898	100%	0.0	99%	Query_16855
<input type="checkbox"/> JKD6008	898	898	100%	0.0	99%	Query_16854
<input type="checkbox"/> T0131	898	898	100%	0.0	99%	Query_16853
<input type="checkbox"/> VC40	898	898	100%	0.0	99%	Query_16852
<input type="checkbox"/> M1	898	898	100%	0.0	99%	Query_16851
<input type="checkbox"/> Bmb9393	898	898	100%	0.0	99%	Query_16850
<input type="checkbox"/> Z172	898	898	100%	0.0	99%	Query_16849
<input type="checkbox"/> USA300-ISMMS1	898	898	100%	0.0	99%	Query_16848
<input type="checkbox"/> NRS100	898	898	100%	0.0	99%	Query_16847
<input type="checkbox"/> UA-S391_USA300	898	898	100%	0.0	99%	Query_16846
<input type="checkbox"/> XN108	898	898	100%	0.0	99%	Query_16845
<input type="checkbox"/> MW2	893	893	100%	0.0	99%	Query_16863
<input type="checkbox"/> MSSA476	893	893	100%	0.0	99%	Query_16862
<input type="checkbox"/> CN1	893	893	100%	0.0	99%	Query_16861
<input type="checkbox"/> MRSA_OP100	893	893	100%	0.0	99%	Query_16827
<input type="checkbox"/> RF122	887	887	100%	0.0	99%	Query_16870
<input type="checkbox"/> ED133	887	887	100%	0.0	99%	Query_16869
<input type="checkbox"/> LGA251	887	887	100%	0.0	99%	Query_16868
<input type="checkbox"/> M013	887	887	100%	0.0	99%	Query_16867
<input type="checkbox"/> 6850	887	887	100%	0.0	99%	Query_16866
<input type="checkbox"/> SA957	887	887	100%	0.0	99%	Query_16865
<input type="checkbox"/> SA40	887	887	100%	0.0	99%	Query_16864
<input type="checkbox"/> 11819-97	876	876	100%	0.0	98%	Query_16871
<input type="checkbox"/> MRSA252	732	732	100%	0.0	93%	Query_16874
<input type="checkbox"/> TCH60	732	732	100%	0.0	93%	Query_16873
<input type="checkbox"/> 55/2053	732	732	100%	0.0	93%	Query_16872
<input type="checkbox"/> MRSA_IP70	732	732	100%	0.0	93%	Query_16829
<input type="checkbox"/> JKD6159	665	665	100%	0.0	91%	Query_16875
<input type="checkbox"/> ST398	649	649	100%	0.0	90%	Query_16878
<input type="checkbox"/> 71193	649	649	100%	0.0	90%	Query_16877
<input type="checkbox"/> 08BA02176	649	649	100%	0.0	90%	Query_16876
<input type="checkbox"/> MRSA_S62	649	649	100%	0.0	90%	Query_16832
<input type="checkbox"/> MRSA_AS3	649	649	100%	0.0	90%	Query_16831
<input type="checkbox"/> MRSA_B53	649	649	100%	0.0	90%	Query_16830
<input type="checkbox"/> CA-347	593	593	100%	2e-172	88%	Query_16879

fibrinogen-binding protein (*efb*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> N315	1303	1303	100%	0.0	100%	Query_112891
<input type="checkbox"/> Mu50	1303	1303	100%	0.0	100%	Query_112890
<input type="checkbox"/> Mu3DNA	1303	1303	100%	0.0	100%	Query_112889
<input type="checkbox"/> MRSA_IP70	1297	1297	100%	0.0	99%	Query_112888
<input type="checkbox"/> RF122	1236	1236	100%	0.0	98%	Query_112892
<input type="checkbox"/> ED133	1225	1225	100%	0.0	98%	Query_112893

Toxic shock syndrome (*tst*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> N315	4488	4488	100%	0.0	100%	Query_106439
<input type="checkbox"/> Mu50	4488	4488	100%	0.0	100%	Query_106438
<input type="checkbox"/> JH9	4488	4488	100%	0.0	100%	Query_106437
<input type="checkbox"/> JH1	4488	4488	100%	0.0	100%	Query_106436
<input type="checkbox"/> Mu3	4488	4488	100%	0.0	100%	Query_106435
<input type="checkbox"/> ED98	4488	4488	100%	0.0	100%	Query_106434
<input type="checkbox"/> 04-02981	4488	4488	100%	0.0	100%	Query_106433
<input type="checkbox"/> ECT-R2	4488	4488	100%	0.0	100%	Query_106432
<input type="checkbox"/> ST228_16035	4488	4488	100%	0.0	100%	Query_106431
<input type="checkbox"/> ST228_16125	4488	4488	100%	0.0	100%	Query_106430
<input type="checkbox"/> ST228_18583	4488	4488	100%	0.0	100%	Query_106429
<input type="checkbox"/> 502A	4488	4488	100%	0.0	100%	Query_106428
<input type="checkbox"/> MRSA_OP100	3764	3764	100%	0.0	95%	Query_106422
<input type="checkbox"/> MW2	3759	3759	99%	0.0	95%	Query_106442
<input type="checkbox"/> MSSA476	3759	3759	99%	0.0	95%	Query_106441
<input type="checkbox"/> CN1	3759	3759	99%	0.0	95%	Query_106440
<input type="checkbox"/> RF122	3546	3546	100%	0.0	93%	Query_106443
<input type="checkbox"/> MRSA_B51	3443	3443	100%	0.0	92%	Query_106423
<input type="checkbox"/> 71193	3419	3419	100%	0.0	92%	Query_106445
<input type="checkbox"/> 08BA02176	3419	3419	100%	0.0	92%	Query_106444
<input type="checkbox"/> ST398	3413	3413	100%	0.0	92%	Query_106446
<input type="checkbox"/> USA300_FPR3757	3402	3402	100%	0.0	92%	Query_106459
<input type="checkbox"/> NCTC8325	3402	3402	100%	0.0	92%	Query_106458
<input type="checkbox"/> COL	3402	3402	100%	0.0	92%	Query_106457
<input type="checkbox"/> Newman	3402	3402	100%	0.0	92%	Query_106456
<input type="checkbox"/> USA300_TCH1516	3402	3402	100%	0.0	92%	Query_106455
<input type="checkbox"/> TW20	3402	3402	100%	0.0	92%	Query_106454
<input type="checkbox"/> T0131	3402	3402	100%	0.0	92%	Query_106453
<input type="checkbox"/> VC40	3402	3402	100%	0.0	92%	Query_106452
<input type="checkbox"/> M1	3402	3402	100%	0.0	92%	Query_106451
<input type="checkbox"/> Bmb9393	3402	3402	100%	0.0	92%	Query_106450
<input type="checkbox"/> Z172	3402	3402	100%	0.0	92%	Query_106449
<input type="checkbox"/> NRS100	3402	3402	100%	0.0	92%	Query_106448
<input type="checkbox"/> XN108	3402	3402	100%	0.0	92%	Query_106447
<input type="checkbox"/> USA300-ISMMS1	3397	3397	100%	0.0	92%	Query_106461
<input type="checkbox"/> UA-S391 USA300	3397	3397	100%	0.0	92%	Query_106460
<input type="checkbox"/> MRSA_S62	3395	3395	100%	0.0	92%	Query_106426
<input type="checkbox"/> JKD6008	3391	3391	100%	0.0	92%	Query_106465
<input type="checkbox"/> M013	3391	3391	100%	0.0	92%	Query_106464
<input type="checkbox"/> 11819-97	3391	3391	100%	0.0	92%	Query_106463
<input type="checkbox"/> SA957	3391	3391	100%	0.0	92%	Query_106462
<input type="checkbox"/> SA40	3386	3386	100%	0.0	92%	Query_106466
<input type="checkbox"/> HO50960412	3382	3382	100%	0.0	92%	Query_106468
<input type="checkbox"/> H-EMRSA-15	3382	3382	100%	0.0	92%	Query_106467
<input type="checkbox"/> 8850	3380	3380	100%	0.0	92%	Query_106469
<input type="checkbox"/> LGA251	3376	3376	100%	0.0	92%	Query_106470
<input type="checkbox"/> ED133	3374	3374	100%	0.0	92%	Query_106471
<input type="checkbox"/> MRSA252	3360	3360	100%	0.0	92%	Query_106472
<input type="checkbox"/> MRSA_B53	3360	3360	100%	0.0	92%	Query_106425
<input type="checkbox"/> 55/2053	3354	3354	100%	0.0	92%	Query_106473
<input type="checkbox"/> MRSA_IP70	3352	3352	100%	0.0	92%	Query_106424
<input type="checkbox"/> TCH60	3349	3349	100%	0.0	92%	Query_106474
<input type="checkbox"/> MRSA_AS3	3293	3293	100%	0.0	91%	Query_106427
<input type="checkbox"/> CA-347	3221	3221	100%	0.0	91%	Query_106475
<input type="checkbox"/> MSHR1132	2959	2959	100%	0.0	89%	Query_106476
<input type="checkbox"/> JKD6159	2754	2754	100%	0.0	87%	Query_106477

hyaluronate lyase precursor (*hysA*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
USA300_FPR3737	1718	1718	100%	0.0	100%	Query_159438
NCTC8325	1718	1718	100%	0.0	100%	Query_159437
Newman	1718	1718	100%	0.0	100%	Query_159435
USA300_TCH1516	1718	1718	100%	0.0	100%	Query_159434
TW20	1718	1718	100%	0.0	100%	Query_159433
JKD6008	1718	1718	100%	0.0	100%	Query_159432
T0131	1718	1718	100%	0.0	100%	Query_159431
VC40	1718	1718	100%	0.0	100%	Query_159430
M1	1718	1718	100%	0.0	100%	Query_159429
9mb9393	1718	1718	100%	0.0	100%	Query_159428
Z172	1718	1718	100%	0.0	100%	Query_159427
USA300-ISMM51	1718	1718	100%	0.0	100%	Query_159426
NRS100	1718	1718	100%	0.0	100%	Query_159425
UA-S391_USA300	1718	1718	100%	0.0	100%	Query_159424
XN108	1718	1718	100%	0.0	100%	Query_159423
COL	1711	1711	100%	0.0	99%	Query_159436
M013	1696	1696	100%	0.0	99%	Query_159442
8850	1696	1696	100%	0.0	99%	Query_159441
SA957	1696	1696	100%	0.0	99%	Query_159440
SA40	1696	1696	100%	0.0	99%	Query_159439
MW2	1690	1690	100%	0.0	99%	Query_159447
MSSA476	1690	1690	100%	0.0	99%	Query_159446
ED98	1690	1690	100%	0.0	99%	Query_159445
11819-97	1690	1690	100%	0.0	99%	Query_159444
CN1	1690	1690	100%	0.0	99%	Query_159443
MRSA_B51	1690	1690	100%	0.0	99%	Query_159418
MRSA_OP100	1690	1690	100%	0.0	99%	Query_159417
RF122	1685	1685	100%	0.0	99%	Query_159461
N315	1685	1685	100%	0.0	99%	Query_159460
Mu50	1685	1685	100%	0.0	99%	Query_159459
JH9	1685	1685	100%	0.0	99%	Query_159458
JH1	1685	1685	100%	0.0	99%	Query_159457
Mu3	1685	1685	100%	0.0	99%	Query_159456
04-02981	1685	1685	100%	0.0	99%	Query_159455
TCH60	1685	1685	100%	0.0	99%	Query_159454
LGA251	1685	1685	100%	0.0	99%	Query_159453
ST228_18035	1685	1685	100%	0.0	99%	Query_159452
ST228_18125	1685	1685	100%	0.0	99%	Query_159451
ST228_18583	1685	1685	100%	0.0	99%	Query_159450
55/2053	1685	1685	100%	0.0	99%	Query_159449
502A	1685	1685	100%	0.0	99%	Query_159448
MRSA_IP70	1685	1685	100%	0.0	99%	Query_159419
MRSA252	1679	1679	100%	0.0	99%	Query_159465
ED133	1679	1679	100%	0.0	99%	Query_159464
ECT-R2	1679	1679	100%	0.0	99%	Query_159463
CA-347	1679	1679	100%	0.0	99%	Query_159462
H050960412	1668	1668	100%	0.0	99%	Query_159467
H-EMRSA-15	1668	1668	100%	0.0	99%	Query_159466
ST398	1663	1663	100%	0.0	99%	Query_159469
71193	1663	1663	100%	0.0	99%	Query_159468
MRSA_AS3	1663	1663	100%	0.0	99%	Query_159422
MRSA_S62	1663	1663	100%	0.0	99%	Query_159421
MRSA_B53	1663	1663	100%	0.0	99%	Query_159420
08BA02176	1657	1657	100%	0.0	99%	Query_159470
JKD6159	1535	1535	100%	0.0	96%	Query_159471

gamma-hemolysin, component A (*hlgA*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
MRSA252	8946	12793	100%	0.0	99%	Query_199211
55/2053	8929	12782	100%	0.0	99%	Query_199212
TCH60	8924	12771	100%	0.0	99%	Query_199213
MRSA_IP70	8874	12741	100%	0.0	99%	Query_199207
RF122	7173	11029	100%	0.0	93%	Query_199214
71193	7112	11051	100%	0.0	93%	Query_199216
08BA02176	7112	11051	100%	0.0	93%	Query_199215
ST398	7040	10885	100%	0.0	93%	Query_199217
MRSA_AS3	6868	10907	100%	0.0	92%	Query_199210
MRSA_S62	6868	10896	100%	0.0	92%	Query_199209
LGA251	6817	10854	100%	0.0	92%	Query_199218
JKD6159	6663	10497	100%	0.0	91%	Query_199219
8550	6396	10809	100%	0.0	91%	Query_199220
MRSA_B53	6010	10907	100%	0.0	89%	Query_199208
MSHR1132	5821	10619	100%	0.0	88%	Query_199221
MRSA_OP100	5504	10715	100%	0.0	87%	Query_199205
TW20	5417	10525	100%	0.0	87%	Query_199229
T0131	5417	10525	100%	0.0	87%	Query_199228
VC40	5417	10525	100%	0.0	87%	Query_199227
M1	5417	10525	100%	0.0	87%	Query_199226
Bmb9393	5417	10525	100%	0.0	87%	Query_199225
Z172	5417	10525	100%	0.0	87%	Query_199224
NRS100	5417	10525	100%	0.0	87%	Query_199223
XN108	5417	10525	100%	0.0	87%	Query_199222
Newman	5393	10472	100%	0.0	87%	Query_199231
USA300_TCH1516	5393	10472	100%	0.0	87%	Query_199230
COL	5384	10475	100%	0.0	87%	Query_199232
NCTC8325	5373	10597	100%	0.0	87%	Query_199233
USA300_FPR3737	5367	10619	100%	0.0	87%	Query_199234
USA300-ISMMS1	5317	10641	100%	0.0	87%	Query_199236
UA-S391_USA300	5317	10641	100%	0.0	87%	Query_199235
JKD6008	5312	10630	100%	0.0	87%	Query_199237
MRSA_B51	5312	10632	100%	0.0	87%	Query_199206
ED133	5245	10619	100%	0.0	86%	Query_199238
CA-347	5149	10377	100%	0.0	86%	Query_199239
M013	5129	10448	100%	0.0	86%	Query_199241
SA957	5129	10448	100%	0.0	86%	Query_199240
SA40	5123	10436	100%	0.0	86%	Query_199242
ST228_18583	5101	10525	100%	0.0	86%	Query_199244
502A	5101	10525	100%	0.0	86%	Query_199243
ST228_16125	5099	10339	100%	0.0	86%	Query_199245
ECT-R2	5096	10514	100%	0.0	86%	Query_199247
ST228_16035	5096	10514	100%	0.0	86%	Query_199246
JH9	5090	10503	100%	0.0	86%	Query_199252
JH1	5090	10320	100%	0.0	86%	Query_199251
Mu3	5090	10459	100%	0.0	86%	Query_199250
ED98	5090	10459	100%	0.0	86%	Query_199249
04-02981	5090	10320	100%	0.0	86%	Query_199248
N315	5084	10459	100%	0.0	86%	Query_199254
Mu50	5084	10309	100%	0.0	86%	Query_199253
11819-97	4983	10200	100%	0.0	85%	Query_199255
H-EMRSA-15	4761	10566	100%	0.0	85%	Query_199256
H050960412	4667	10239	100%	0.0	84%	Query_199257

hyaluronate lysase precursor 1 (*hysA1*)

USA300_FPR3737	100%	0.0	100%
NCTC8325	100%	0.0	100%
COL	100%	0.0	100%
Newman	100%	0.0	100%
USA300_TCH1516	100%	0.0	100%
T0131	100%	0.0	100%
VC40	100%	0.0	100%
M1	100%	0.0	100%
Bmb9393	100%	0.0	100%
Z172	100%	0.0	100%
USA300-ISMMS1	100%	0.0	100%
NRS100	100%	0.0	100%
UA-S391_USA300	100%	0.0	100%
XN108	100%	0.0	100%
TW20	100%	0.0	99%
JKD6008	100%	0.0	99%
MRSA_B51	100%	0.0	99%
H050960412	100%	0.0	99%
H-EMRSA-15	100%	0.0	99%
JH9	100%	0.0	98%
JH1	100%	0.0	98%
ED98	100%	0.0	98%
04-02981	100%	0.0	98%
ECT-R2	100%	0.0	98%
ST228_16035	100%	0.0	98%
ST228_16125	100%	0.0	98%
ST228_18583	100%	0.0	98%
CN1	100%	0.0	98%
502A	100%	0.0	98%
MW2	100%	0.0	98%
N315	100%	0.0	98%
MSSA476	100%	0.0	98%
Mu50	100%	0.0	98%
Mu3	100%	0.0	98%
11819-97	100%	0.0	98%
MRSA_OP100	100%	0.0	98%
JKD6159	100%	0.0	98%
M013	100%	0.0	97%
SA957	100%	0.0	97%
SA40	100%	0.0	97%
MRSA252	100%	0.0	97%
55/2053	100%	0.0	97%
MRSA_IP70	100%	0.0	97%
TCH60	100%	0.0	97%
ST398	100%	0.0	97%
71193	100%	0.0	97%
MRSA_AS3	100%	0.0	97%
MRSA_S62(2)	100%	0.0	97%
MRSA_S62	100%	0.0	97%
08BA02176	100%	0.0	97%
RF122	100%	0.0	97%
ED133	100%	0.0	96%
LGA251	100%	0.0	96%
CA-347	100%	0.0	96%

elastin binding protein (*ebp*)

USA300_FPR3757	100%	0.0	100%
NCTC8325	100%	0.0	100%
COL	100%	0.0	100%
Newman	100%	0.0	100%
USA300_TCH1516	100%	0.0	100%
VC40	100%	0.0	100%
M1	100%	0.0	100%
USA300-ISMMS1	100%	0.0	100%
NRS100	100%	0.0	100%
UA-S391_USA300	100%	0.0	100%
MW2	100%	0.0	99%
MSSA476	100%	0.0	99%
11819-97	100%	0.0	99%
MRSA_OP100	100%	0.0	99%
RF122	100%	0.0	99%
N315	100%	0.0	99%
Mu50	100%	0.0	99%
JH9	100%	0.0	99%
JH1	100%	0.0	99%
Mu3	100%	0.0	99%
ED98	100%	0.0	99%
04-02981	100%	0.0	99%
ECT-R2	100%	0.0	99%
ST228_16035	100%	0.0	99%
ST228_18125	100%	0.0	99%
ST228_18583	100%	0.0	99%
502A	100%	0.0	99%
MRSA_B51	100%	0.0	99%
ED133	100%	0.0	99%
LGA251	100%	0.0	99%
6850	100%	0.0	99%
M013	100%	0.0	99%
CN1	100%	0.0	99%
SA957	100%	0.0	99%
SA40	100%	0.0	99%
JKD6159	100%	0.0	99%
HO50960412	100%	0.0	90%
H-EMRSA-15	100%	0.0	90%
ST398	100%	0.0	89%
71193	100%	0.0	89%
08BA02176	100%	0.0	89%
MRSA_AS3	99%	0.0	89%
MRSA_S62	100%	0.0	89%
MRSA_B53	100%	0.0	89%
MRSA252	100%	0.0	89%
TCH60	100%	0.0	89%
55/2053	100%	0.0	89%
MRSA_IP70	100%	0.0	89%
TW20	100%	0.0	89%
JKD6008	100%	0.0	89%
T0131	100%	0.0	89%
Bmb0393	100%	0.0	89%
Z172	100%	0.0	89%
XN108	100%	0.0	89%

zinc metalloproteinase aureolysin (*aur*)

Alignments Download Graphics Distance tree of results			
Description	Query cover	E value	Ident
<input type="checkbox"/> TW20	100%	0.0	100%
<input type="checkbox"/> JKD6008	100%	0.0	100%
<input type="checkbox"/> T0131	100%	0.0	100%
<input type="checkbox"/> Bmb9393	100%	0.0	100%
<input type="checkbox"/> Z172	100%	0.0	100%
<input type="checkbox"/> MRSA_IP70	100%	0.0	99%
<input type="checkbox"/> XN108	100%	0.0	99%
<input type="checkbox"/> MRSA_S62	100%	0.0	99%
<input type="checkbox"/> TCH60	100%	0.0	99%
<input type="checkbox"/> MRSA252	100%	0.0	99%
<input type="checkbox"/> S5/2053	89%	0.0	99%
<input type="checkbox"/> O8BA02176	100%	0.0	99%
<input type="checkbox"/> ST398	100%	0.0	98%
<input type="checkbox"/> CA-347	100%	0.0	98%
<input checked="" type="checkbox"/> MSSA476	100%	0.0	98%
<input type="checkbox"/> HO50960412	100%	0.0	98%
<input type="checkbox"/> H-EMRSA-15	100%	0.0	98%
<input type="checkbox"/> MW2	100%	0.0	98%
<input type="checkbox"/> 6850	100%	0.0	98%
<input type="checkbox"/> LGA251	100%	0.0	98%
<input type="checkbox"/> 71193	100%	0.0	99%

Collagen adhesion gene (*cna*)

Alignments Download Graphics Distance tree of results			
Description	Query cover	E value	Ident
<input type="checkbox"/> USA300_FPR3757	100%	0.0	100%
<input type="checkbox"/> NCTC8325	100%	0.0	100%
<input type="checkbox"/> COL	100%	0.0	100%
<input type="checkbox"/> Newman	100%	0.0	100%
<input type="checkbox"/> USA300_TCH1516	100%	0.0	100%
<input type="checkbox"/> TW20	100%	0.0	100%
<input type="checkbox"/> JKD6008	100%	0.0	100%
<input type="checkbox"/> T0131	100%	0.0	100%
<input type="checkbox"/> 11819-97	100%	0.0	100%
<input type="checkbox"/> VC40	100%	0.0	100%
<input type="checkbox"/> M1	100%	0.0	100%
<input type="checkbox"/> Bmb9393	100%	0.0	100%
<input type="checkbox"/> Z172	100%	0.0	100%
<input type="checkbox"/> USA300-ISMM51	100%	0.0	100%
<input type="checkbox"/> NRS100	100%	0.0	100%
<input type="checkbox"/> UA-S391_USA300	100%	0.0	100%
<input type="checkbox"/> xn108	100%	0.0	100%
<input type="checkbox"/> M013	100%	0.0	98%
<input type="checkbox"/> SA957	100%	0.0	98%
<input type="checkbox"/> SA40	100%	0.0	98%
<input type="checkbox"/> 6850	100%	0.0	97%
<input type="checkbox"/> MRSA_AS3	100%	2e-144	86%
<input type="checkbox"/> MRSA_S62	100%	2e-144	86%
<input type="checkbox"/> MRSA_B53	100%	2e-144	86%

von Willebrand factor (*vWb*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
MRSA252	4482	4482	100%	0.0	100%	Query_144370
55/2053	4477	4477	100%	0.0	99%	Query_144371
TCH60	4471	4471	100%	0.0	99%	Query_144372
MRSA_IP70	4414	4414	100%	0.0	99%	Query_144366
MRSA_OP100	3546	3546	100%	0.0	93%	Query_144364
MW2	3530	3530	99%	0.0	93%	Query_144375
MSSA476	3530	3530	99%	0.0	93%	Query_144374
CN1	3530	3530	99%	0.0	93%	Query_144373
RF122	3426	3426	100%	0.0	92%	Query_144376
MRSA_B51	3413	3413	100%	0.0	92%	Query_144365
N315	3360	3360	100%	0.0	92%	Query_144388
Mu50	3360	3360	100%	0.0	92%	Query_144387
JH9	3360	3360	100%	0.0	92%	Query_144386
JH1	3360	3360	100%	0.0	92%	Query_144385
Mu3	3360	3360	100%	0.0	92%	Query_144384
ED98	3360	3360	100%	0.0	92%	Query_144383
04-02981	3360	3360	100%	0.0	92%	Query_144382
ECT-R2	3360	3360	100%	0.0	92%	Query_144381
ST228_16035	3360	3360	100%	0.0	92%	Query_144380
ST228_16125	3360	3360	100%	0.0	92%	Query_144379
ST228_18583	3360	3360	100%	0.0	92%	Query_144378
502A	3360	3360	100%	0.0	92%	Query_144377
HO50960412	3336	3336	100%	0.0	92%	Query_144390
H-EMRSA-15	3336	3336	100%	0.0	92%	Query_144389
11819-97	3326	3326	100%	0.0	91%	Query_144391
8850	3315	3315	100%	0.0	91%	Query_144392
ED133	3310	3310	100%	0.0	91%	Query_144393
USA300_FPR3737	3265	3265	100%	0.0	91%	Query_144406
NCTC8325	3265	3265	100%	0.0	91%	Query_144405
COL	3265	3265	100%	0.0	91%	Query_144404
Newman	3265	3265	100%	0.0	91%	Query_144403
USA300_TCH1516	3265	3265	100%	0.0	91%	Query_144402
TW20	3265	3265	100%	0.0	91%	Query_144401
T0131	3265	3265	100%	0.0	91%	Query_144400
VC40	3265	3265	100%	0.0	91%	Query_144399
M1	3265	3265	100%	0.0	91%	Query_144398
8mb9393	3265	3265	100%	0.0	91%	Query_144397
Z172	3265	3265	100%	0.0	91%	Query_144396
NRS100	3265	3265	100%	0.0	91%	Query_144395
XN108	3265	3265	100%	0.0	91%	Query_144394
LGA251	3264	3264	100%	0.0	91%	Query_144407
USA300-ISMMS1	3260	3260	100%	0.0	91%	Query_144409
UA-S391 USA300	3260	3260	100%	0.0	91%	Query_144408
JKD6008	3254	3254	100%	0.0	91%	Query_144414
M013	3254	3254	100%	0.0	91%	Query_144413
71193	3254	3254	100%	0.0	91%	Query_144412
08BA02176	3254	3254	100%	0.0	91%	Query_144411
SA957	3254	3254	100%	0.0	91%	Query_144410
MRSA_AS3	3254	3254	100%	0.0	91%	Query_144369
MRSA_S62	3254	3254	100%	0.0	91%	Query_144368
MRSA_B53	3254	3254	100%	0.0	91%	Query_144367
ST398	3249	3249	100%	0.0	91%	Query_144416
SA40	3249	3249	100%	0.0	91%	Query_144415
CA437	3164	3164	100%	0.0	90%	Query_144417
MSHR1132	2955	2955	100%	0.0	89%	Query_144418
JKD6159	2861	2861	100%	0.0	88%	Query_144419

triacylglycerol lipase (*lip1*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> USA300_FPR3757	1951	1951	100%	0.0	100%	Query_26026
<input type="checkbox"/> NCTC8326	1951	1951	100%	0.0	100%	Query_26025
<input type="checkbox"/> COL	1951	1951	100%	0.0	100%	Query_26024
<input type="checkbox"/> Newman	1951	1951	100%	0.0	100%	Query_26023
<input type="checkbox"/> USA300_TCH1516	1951	1951	100%	0.0	100%	Query_26022
<input type="checkbox"/> TW20	1951	1951	100%	0.0	100%	Query_26021
<input type="checkbox"/> JKD6008	1951	1951	100%	0.0	100%	Query_26020
<input type="checkbox"/> T0131	1951	1951	100%	0.0	100%	Query_26019
<input type="checkbox"/> VC40	1951	1951	100%	0.0	100%	Query_26018
<input type="checkbox"/> M1	1951	1951	100%	0.0	100%	Query_26017
<input type="checkbox"/> Bmb89393	1951	1951	100%	0.0	100%	Query_26016
<input type="checkbox"/> Z172	1951	1951	100%	0.0	100%	Query_26015
<input type="checkbox"/> USA300-ISMMS1	1951	1951	100%	0.0	100%	Query_26014
<input type="checkbox"/> NRS100	1951	1951	100%	0.0	100%	Query_26013
<input type="checkbox"/> UA-S391 USA300	1951	1951	100%	0.0	100%	Query_26012
<input type="checkbox"/> XN108	1951	1951	100%	0.0	100%	Query_26011
<input type="checkbox"/> N315	1912	1912	100%	0.0	99%	Query_26034
<input type="checkbox"/> JH9	1912	1912	100%	0.0	99%	Query_26033
<input type="checkbox"/> JH1	1912	1912	100%	0.0	99%	Query_26032
<input type="checkbox"/> Mu3	1912	1912	100%	0.0	99%	Query_26031
<input type="checkbox"/> ED98	1912	1912	100%	0.0	99%	Query_26030
<input type="checkbox"/> 04-02981	1912	1912	100%	0.0	99%	Query_26029
<input type="checkbox"/> ECT-R2	1912	1912	100%	0.0	99%	Query_26028
<input type="checkbox"/> 502A	1912	1912	100%	0.0	99%	Query_26027
<input type="checkbox"/> Mu50	1906	1906	100%	0.0	99%	Query_26035
<input type="checkbox"/> ED133	1901	1901	100%	0.0	99%	Query_26036
<input type="checkbox"/> JKD6159	1895	1895	100%	0.0	99%	Query_26040
<input type="checkbox"/> 6850	1895	1895	100%	0.0	99%	Query_26039
<input type="checkbox"/> SA957	1895	1895	100%	0.0	99%	Query_26038
<input type="checkbox"/> SA40	1895	1895	100%	0.0	99%	Query_26037
<input type="checkbox"/> RF122	1890	1890	100%	0.0	99%	Query_26041
<input type="checkbox"/> M013	1888	1888	100%	0.0	99%	Query_26042
<input type="checkbox"/> LGA251	1882	1882	100%	0.0	99%	Query_26043
<input type="checkbox"/> B51	1869	1869	100%	0.0	99%	Query_26007
<input type="checkbox"/> 11819-97	1864	1864	100%	0.0	99%	Query_26046
<input type="checkbox"/> H050960412	1864	1864	100%	0.0	99%	Query_26045
<input type="checkbox"/> H-EMRSA-15	1864	1864	100%	0.0	99%	Query_26044
<input type="checkbox"/> CN1	1853	1853	100%	0.0	98%	Query_26047
<input type="checkbox"/> MW2	1847	1847	100%	0.0	98%	Query_26049
<input type="checkbox"/> MSSA476	1847	1847	100%	0.0	98%	Query_26048
<input type="checkbox"/> ST398	1419	1419	100%	0.0	91%	Query_26051
<input type="checkbox"/> 08BA02-176	1419	1419	100%	0.0	91%	Query_26050
<input type="checkbox"/> MRSA_S62	1419	1419	100%	0.0	91%	Query_26010
<input type="checkbox"/> MRSA_B53	1419	1419	100%	0.0	91%	Query_26008
<input type="checkbox"/> AS3	1419	1419	100%	0.0	91%	Query_26006
<input type="checkbox"/> 71193	1413	1413	100%	0.0	91%	Query_26052
<input type="checkbox"/> MSHR1132	1247	1247	97%	0.0	89%	Query_26053
<input type="checkbox"/> TCH60	1125	1125	100%	0.0	86%	Query_26055
<input type="checkbox"/> 55/2053	1125	1125	100%	0.0	86%	Query_26054
<input type="checkbox"/> MRSA_IP70	1125	1125	100%	0.0	86%	Query_26009
<input type="checkbox"/> MRSA252	1120	1120	100%	0.0	86%	Query_26056
<input type="checkbox"/> CA-347	1109	1109	100%	0.0	86%	Query_26057

leucocidin S component (*lukS*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> N315	3779	3779	100%	0.0	100%	Query_220021
<input type="checkbox"/> Mu90	3779	3779	100%	0.0	100%	Query_220020
<input type="checkbox"/> JH9	3779	3779	100%	0.0	100%	Query_220019
<input type="checkbox"/> JH1	3779	3779	100%	0.0	100%	Query_220018
<input type="checkbox"/> Mu3	3779	3779	100%	0.0	100%	Query_220017
<input type="checkbox"/> ED98	3779	3779	100%	0.0	100%	Query_220016
<input type="checkbox"/> 04-02981	3779	3779	100%	0.0	100%	Query_220015
<input type="checkbox"/> ECT-R2	3779	3779	100%	0.0	100%	Query_220014
<input type="checkbox"/> ST228_16035	3779	3779	100%	0.0	100%	Query_220013
<input type="checkbox"/> ST228_16125	3779	3779	100%	0.0	100%	Query_220012
<input type="checkbox"/> ST228_18583	3779	3779	100%	0.0	100%	Query_220011
<input type="checkbox"/> 502A	3779	3779	100%	0.0	100%	Query_220010
<input type="checkbox"/> 11819-97	3696	3696	100%	0.0	99%	Query_220022
<input type="checkbox"/> CN1	3668	3668	100%	0.0	99%	Query_220023
<input type="checkbox"/> MW2	3663	3663	100%	0.0	99%	Query_220033
<input type="checkbox"/> MSSA476	3663	3663	100%	0.0	99%	Query_220032
<input type="checkbox"/> USA300_FPR3757	3663	3663	100%	0.0	99%	Query_220031
<input type="checkbox"/> NCTC8325	3663	3663	100%	0.0	99%	Query_220030
<input type="checkbox"/> COL	3663	3663	100%	0.0	99%	Query_220029
<input type="checkbox"/> USA300_TCH1516	3663	3663	100%	0.0	99%	Query_220028
<input type="checkbox"/> VC40	3663	3663	100%	0.0	99%	Query_220027
<input type="checkbox"/> USA300-ISMMS1	3663	3663	100%	0.0	99%	Query_220026
<input type="checkbox"/> NRS100	3663	3663	100%	0.0	99%	Query_220025
<input type="checkbox"/> UA-S391_USA300	3663	3663	100%	0.0	99%	Query_220024
<input type="checkbox"/> Newman	3657	3657	100%	0.0	99%	Query_220035
<input type="checkbox"/> M1	3657	3657	100%	0.0	99%	Query_220034
<input type="checkbox"/> MRSA_OP100	3657	3657	99%	0.0	99%	Query_220004
<input type="checkbox"/> LGA251	3651	3651	100%	0.0	99%	Query_220036
<input type="checkbox"/> M013	3640	3640	100%	0.0	99%	Query_220039
<input type="checkbox"/> SA957	3640	3640	100%	0.0	99%	Query_220038
<input type="checkbox"/> SA40	3640	3640	100%	0.0	99%	Query_220037
<input type="checkbox"/> 8850	3635	3635	100%	0.0	99%	Query_220040
<input type="checkbox"/> MRSA_BS1	3635	3635	99%	0.0	99%	Query_220005
<input type="checkbox"/> RF122	3613	3613	100%	0.0	99%	Query_220043
<input type="checkbox"/> ED133	3613	3613	100%	0.0	99%	Query_220042
<input type="checkbox"/> JKD6159	3613	3613	100%	0.0	99%	Query_220041
<input type="checkbox"/> TW20	3585	3585	100%	0.0	98%	Query_220051
<input type="checkbox"/> JKD6008	3585	3585	100%	0.0	98%	Query_220050
<input type="checkbox"/> T0131	3585	3585	100%	0.0	98%	Query_220049
<input type="checkbox"/> HO50960412	3585	3585	100%	0.0	98%	Query_220048
<input type="checkbox"/> Bmb9383	3585	3585	100%	0.0	98%	Query_220047
<input type="checkbox"/> 55/2053	3585	3585	100%	0.0	98%	Query_220046
<input type="checkbox"/> Z172	3585	3585	100%	0.0	98%	Query_220045
<input type="checkbox"/> H-EMRSA-15	3585	3585	100%	0.0	98%	Query_220044
<input type="checkbox"/> XN108	3583	3583	100%	0.0	98%	Query_220052
<input type="checkbox"/> TCH60	3579	3579	100%	0.0	98%	Query_220053
<input type="checkbox"/> MRSA252	3574	3574	100%	0.0	98%	Query_220054
<input type="checkbox"/> MRSA_IP70	3574	3574	99%	0.0	98%	Query_220006
<input type="checkbox"/> CA-347	3452	3452	100%	0.0	97%	Query_220055
<input type="checkbox"/> ST398	3446	3446	100%	0.0	97%	Query_220058
<input type="checkbox"/> 71193	3446	3446	100%	0.0	97%	Query_220057
<input type="checkbox"/> 08BA02176	3446	3446	100%	0.0	97%	Query_220056
<input type="checkbox"/> MRSA_BS3	3441	3441	99%	0.0	97%	Query_220007
<input type="checkbox"/> MRSA_S62	3435	3435	99%	0.0	97%	Query_220008

lipase gene (*lip*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
MRSA252	4482	4482	100%	0.0	100%	Query_144370
55/2053	4477	4477	100%	0.0	99%	Query_144371
TCH160	4471	4471	100%	0.0	99%	Query_144372
MRSA_IP70	4414	4414	100%	0.0	99%	Query_144366
MRSA_OP100	3546	3546	100%	0.0	93%	Query_144364
MW2	3530	3530	99%	0.0	93%	Query_144375
MSSA476	3530	3530	99%	0.0	93%	Query_144374
CN1	3530	3530	99%	0.0	93%	Query_144373
RF122	3426	3426	100%	0.0	92%	Query_144376
MRSA_B51	3413	3413	100%	0.0	92%	Query_144365
N315	3360	3360	100%	0.0	92%	Query_144388
Mu50	3360	3360	100%	0.0	92%	Query_144387
JH9	3360	3360	100%	0.0	92%	Query_144386
JH1	3360	3360	100%	0.0	92%	Query_144385
Mu3	3360	3360	100%	0.0	92%	Query_144384
ED98	3360	3360	100%	0.0	92%	Query_144383
04-Q2981	3360	3360	100%	0.0	92%	Query_144382
ECT-R2	3360	3360	100%	0.0	92%	Query_144381
ST228_16035	3360	3360	100%	0.0	92%	Query_144380
ST228_16125	3360	3360	100%	0.0	92%	Query_144379
ST228_18583	3360	3360	100%	0.0	92%	Query_144378
502A	3360	3360	100%	0.0	92%	Query_144377
H050960412	3336	3336	100%	0.0	92%	Query_144390
H-EMRSA-15	3336	3336	100%	0.0	92%	Query_144389
11819-97	3326	3326	100%	0.0	91%	Query_144391
6850	3315	3315	100%	0.0	91%	Query_144392
ED133	3310	3310	100%	0.0	91%	Query_144393
USA300_FPR3737	3265	3265	100%	0.0	91%	Query_144406
NCTC8325	3265	3265	100%	0.0	91%	Query_144405
COL	3265	3265	100%	0.0	91%	Query_144404
Newman	3265	3265	100%	0.0	91%	Query_144403
USA300_TCH1516	3265	3265	100%	0.0	91%	Query_144402
TW20	3265	3265	100%	0.0	91%	Query_144401
T0131	3265	3265	100%	0.0	91%	Query_144400
VC40	3265	3265	100%	0.0	91%	Query_144399
M1	3265	3265	100%	0.0	91%	Query_144398
Bmb9393	3265	3265	100%	0.0	91%	Query_144397
Z172	3265	3265	100%	0.0	91%	Query_144396
NRS100	3265	3265	100%	0.0	91%	Query_144395
XN108	3265	3265	100%	0.0	91%	Query_144394
LGA251	3264	3264	100%	0.0	91%	Query_144407
USA300-ISMM51	3260	3260	100%	0.0	91%	Query_144409
UA-S391 USA300	3260	3260	100%	0.0	91%	Query_144408
JKD6008	3254	3254	100%	0.0	91%	Query_144414
M013	3254	3254	100%	0.0	91%	Query_144413
71193	3254	3254	100%	0.0	91%	Query_144412
08BA02176	3254	3254	100%	0.0	91%	Query_144411
SA957	3254	3254	100%	0.0	91%	Query_144410
MRSA_AS3	3254	3254	100%	0.0	91%	Query_144369
MRSA_S62	3254	3254	100%	0.0	91%	Query_144368
MRSA_B53	3254	3254	100%	0.0	91%	Query_144367
ST398	3249	3249	100%	0.0	91%	Query_144416
SA40	3249	3249	100%	0.0	91%	Query_144415
CA437	3164	3164	100%	0.0	90%	Query_144417
MSHR1132	2955	2955	100%	0.0	89%	Query_144418
JKD6159	2861	2861	100%	0.0	88%	Query_144419

hyaluronate lyase precursor precursor 2 (*hysA*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
MW2	2427	2427	100%	0.0	100%	Query_152097
MSSA476	2427	2427	100%	0.0	100%	Query_152096
MRSA_B51	2410	2410	100%	0.0	99%	Query_152064
11819-97	2313	2313	100%	0.0	98%	Query_152094
N315	2302	2302	100%	0.0	98%	Query_152091
JH9	2302	2302	100%	0.0	98%	Query_152090
JH1	2302	2302	100%	0.0	98%	Query_152089
ED98	2302	2302	100%	0.0	98%	Query_152088
ECT-R2	2302	2302	100%	0.0	98%	Query_152087
ST228_16035	2302	2302	100%	0.0	98%	Query_152086
ST228_16125	2302	2302	100%	0.0	98%	Query_152085
ST228_18583	2302	2302	100%	0.0	98%	Query_152084
04-02981	2298	2298	99%	0.0	98%	Query_152092
502A	2296	2296	100%	0.0	98%	Query_152093
USA300_FPR3757	2279	2279	100%	0.0	98%	Query_152077
NCTC_8325	2279	2279	100%	0.0	98%	Query_152076
COL	2279	2279	100%	0.0	98%	Query_152075
Newman	2279	2279	100%	0.0	98%	Query_152074
USA300_TCH1516	2279	2279	100%	0.0	98%	Query_152073
VC40	2279	2279	100%	0.0	98%	Query_152072
M1	2279	2279	100%	0.0	98%	Query_152071
USA300-ISMMS1	2279	2279	100%	0.0	98%	Query_152070
UA-S391 USA300	2279	2279	100%	0.0	98%	Query_152069
TW20	2274	2274	100%	0.0	98%	Query_152083
JKD6008	2274	2274	100%	0.0	98%	Query_152082
T0131	2274	2274	100%	0.0	98%	Query_152081
Bmb9393	2274	2274	100%	0.0	98%	Query_152080
Z172	2274	2274	100%	0.0	98%	Query_152079
XN108	2274	2274	100%	0.0	98%	Query_152078
LGA251	2235	2235	100%	0.0	97%	Query_152095
Mu50	2108	2108	100%	0.0	96%	Query_152099
Mu3	2108	2108	100%	0.0	96%	Query_152098
ED133	2041	2041	100%	0.0	95%	Query_152101
MRSA252	2039	2039	100%	0.0	95%	Query_152100
MRSA_IP70	2039	2039	100%	0.0	95%	Query_152065
RF122	1986	1986	97%	0.0	95%	Query_152111
M013	1986	1986	97%	0.0	95%	Query_152110
6850	1986	1986	97%	0.0	95%	Query_152108
SA957	1986	1986	97%	0.0	95%	Query_152107
SA40	1986	1986	97%	0.0	95%	Query_152106
CN1	1975	1975	100%	0.0	94%	Query_152102
CA-347	1973	1973	100%	0.0	94%	Query_152109
ST398	1953	1953	100%	0.0	94%	Query_152105
Z1193	1953	1953	100%	0.0	94%	Query_152104
08BA02176	1953	1953	100%	0.0	94%	Query_152103
MRSA_AS3	1953	1953	100%	0.0	94%	Query_152068
MRSA_S62	1953	1953	100%	0.0	94%	Query_152067
MRSA_B53	1953	1953	100%	0.0	94%	Query_152066
H050960412	1925	1925	97%	0.0	94%	Query_152113
H-EMRSA-15	1925	1925	97%	0.0	94%	Query_152112
JKD6159	1801	1801	100%	0.0	91%	Query_152114

igG-binding protein gene (*sbi*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
N315	3651	6345	100%	0.0	100%	Query_150402
Mu50	3651	6345	100%	0.0	100%	Query_150401
Mu3	3651	6345	100%	0.0	100%	Query_150400
04-02981	3651	6345	100%	0.0	100%	Query_150399
502A	3651	6345	100%	0.0	100%	Query_150398
JH9	3646	6318	100%	0.0	99%	Query_150406
JH1	3646	6318	100%	0.0	99%	Query_150405
ED98	3646	6340	100%	0.0	99%	Query_150404
ECT-R2	3646	6312	100%	0.0	99%	Query_150403
ST228_16035	3607	6325	100%	0.0	99%	Query_150408
ST228_18583	3607	6325	100%	0.0	99%	Query_150407
ST228_16125	3323	5659	100%	0.0	99%	Query_150409
JKD6159	3142	4283	100%	0.0	95%	Query_150410
LGA251	3031	5647	100%	0.0	95%	Query_150411
CN1	2058	4047	100%	0.0	86%	Query_150412
HO-50960412	1962	3664	100%	0.0	85%	Query_150439
HOS0960412	1962	3664	100%	0.0	85%	Query_150414
H-EMRSA-15	1962	3664	100%	0.0	85%	Query_150413
NCTC-8325	1936	3631	100%	0.0	86%	Query_150419
COL	1936	3631	100%	0.0	86%	Query_150418
Newman	1936	3631	100%	0.0	86%	Query_150417
M1	1936	3631	100%	0.0	86%	Query_150416
NRS100	1936	3631	100%	0.0	86%	Query_150415
VC40	1930	3600	100%	0.0	85%	Query_150420
USA300_FPR3757	1836	3135	100%	0.0	85%	Query_150424
USA300_TCH1516	1836	3135	100%	0.0	85%	Query_150423
USA300-ISMMS1	1836	3135	100%	0.0	85%	Query_150422
UA-S391_USA300	1836	3135	100%	0.0	85%	Query_150421
RF122	1624	2543	100%	0.0	83%	Query_150443

staphylocoagulase (*coa*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> MW2	987	987	100%	0.0	100%	Query_110154
<input type="checkbox"/> N315	987	987	100%	0.0	100%	Query_110153
<input type="checkbox"/> MSSA476	987	987	100%	0.0	100%	Query_110152
<input type="checkbox"/> USA300_FPR3757	987	987	100%	0.0	100%	Query_110151
<input type="checkbox"/> NCTC8325	987	987	100%	0.0	100%	Query_110150
<input type="checkbox"/> COL	987	987	100%	0.0	100%	Query_110149
<input type="checkbox"/> Mu50	987	987	100%	0.0	100%	Query_110148
<input type="checkbox"/> JH9	987	987	100%	0.0	100%	Query_110147
<input type="checkbox"/> JH1	987	987	100%	0.0	100%	Query_110146
<input type="checkbox"/> Newman	987	987	100%	0.0	100%	Query_110145
<input type="checkbox"/> Mu3	987	987	100%	0.0	100%	Query_110144
<input type="checkbox"/> USA300_TCH1516	987	987	100%	0.0	100%	Query_110143
<input type="checkbox"/> ED98	987	987	100%	0.0	100%	Query_110142
<input type="checkbox"/> TW20	987	987	100%	0.0	100%	Query_110141
<input type="checkbox"/> JKD6008	987	987	100%	0.0	100%	Query_110140
<input type="checkbox"/> ECT-R2	987	987	100%	0.0	100%	Query_110139
<input type="checkbox"/> T0131	987	987	100%	0.0	100%	Query_110138
<input type="checkbox"/> 11819-97	987	987	100%	0.0	100%	Query_110137
<input type="checkbox"/> VC40	987	987	100%	0.0	100%	Query_110136
<input type="checkbox"/> HO50960412	987	987	100%	0.0	100%	Query_110135
<input type="checkbox"/> ST228_isolate_16035	987	987	100%	0.0	100%	Query_110134
<input type="checkbox"/> ST228_16125	987	987	100%	0.0	100%	Query_110133
<input type="checkbox"/> ST228_18583	987	987	100%	0.0	100%	Query_110132
<input type="checkbox"/> M1	987	987	100%	0.0	100%	Query_110131
<input type="checkbox"/> Bmb9393	987	987	100%	0.0	100%	Query_110130
<input type="checkbox"/> CN1	987	987	100%	0.0	100%	Query_110129
<input type="checkbox"/> Z172	987	987	100%	0.0	100%	Query_110128
<input type="checkbox"/> USA300-ISMMS1	987	987	100%	0.0	100%	Query_110127
<input type="checkbox"/> 502A	987	987	100%	0.0	100%	Query_110126
<input type="checkbox"/> NRS100	987	987	100%	0.0	100%	Query_110125
<input type="checkbox"/> H-EMRSA-15	987	987	100%	0.0	100%	Query_110124
<input type="checkbox"/> UA-S391_USA300	987	987	100%	0.0	100%	Query_110123
<input type="checkbox"/> XN108	987	987	100%	0.0	100%	Query_110122
<input type="checkbox"/> 04-02981	981	981	100%	0.0	99%	Query_110155
<input type="checkbox"/> CA-347	965	965	100%	0.0	99%	Query_110156
<input type="checkbox"/> MRSA252	959	959	100%	0.0	99%	Query_110163
<input type="checkbox"/> RF122	959	959	100%	0.0	99%	Query_110162
<input type="checkbox"/> ST398	959	959	100%	0.0	99%	Query_110161
<input type="checkbox"/> TCH60	959	959	100%	0.0	99%	Query_110160
<input type="checkbox"/> 71193	959	959	100%	0.0	99%	Query_110159
<input type="checkbox"/> 08BA02176	959	959	100%	0.0	99%	Query_110158
<input type="checkbox"/> 55/2053	959	959	100%	0.0	99%	Query_110157
<input type="checkbox"/> M013	953	953	100%	0.0	99%	Query_110166
<input type="checkbox"/> SA957	953	953	100%	0.0	99%	Query_110165
<input type="checkbox"/> SA40	953	953	100%	0.0	99%	Query_110164
<input type="checkbox"/> ED133	948	948	100%	0.0	99%	Query_110169
<input type="checkbox"/> LGA251	948	948	100%	0.0	99%	Query_110168
<input type="checkbox"/> 6850	948	948	100%	0.0	99%	Query_110167
<input type="checkbox"/> JKD6159	715	715	100%	0.0	91%	Query_110170

thermonuclease (*nucA*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
N315	1773	1773	100%	0.0	100%	Query_26228
Mu50	1773	1773	100%	0.0	100%	Query_26227
JH9	1773	1773	100%	0.0	100%	Query_26226
JH1	1773	1773	100%	0.0	100%	Query_26225
Mu3	1773	1773	100%	0.0	100%	Query_26224
ED98	1773	1773	100%	0.0	100%	Query_26223
04-02981	1773	1773	100%	0.0	100%	Query_26222
ECT-R2	1773	1773	100%	0.0	100%	Query_26221
1819-97	1773	1773	100%	0.0	100%	Query_26220
ST228_16035	1773	1773	100%	0.0	100%	Query_26219
ST228_16125	1773	1773	100%	0.0	100%	Query_26218
ST228_18583	1773	1773	100%	0.0	100%	Query_26217
CN1	1773	1773	100%	0.0	100%	Query_26216
502A	1773	1773	100%	0.0	100%	Query_26215
MRSA_B51	1773	1773	100%	0.0	100%	Query_26210
MW2	1762	1762	100%	0.0	99%	Query_26230
MSSA476	1762	1762	100%	0.0	99%	Query_26229
MRSA_OP100	1762	1762	100%	0.0	99%	Query_26209
USA300_FPR3757	1746	1746	100%	0.0	99%	Query_26240
NCTC_8325	1746	1746	100%	0.0	99%	Query_26239
COL	1746	1746	100%	0.0	99%	Query_26238
Newman	1746	1746	100%	0.0	99%	Query_26237
USA300_TCH1516	1746	1746	100%	0.0	99%	Query_26236
VC40	1746	1746	100%	0.0	99%	Query_26235
M1	1746	1746	100%	0.0	99%	Query_26234
USA300-ISMMS1	1746	1746	100%	0.0	99%	Query_26233
NRS-100	1746	1746	100%	0.0	99%	Query_26232
UA-S391 USA300	1746	1746	100%	0.0	99%	Query_26231
RF122	1740	1740	100%	0.0	99%	Query_26249
TW20	1740	1740	100%	0.0	99%	Query_26248
JKD6006	1740	1740	100%	0.0	99%	Query_26247
T0131	1740	1740	100%	0.0	99%	Query_26246
LGA251	1740	1740	100%	0.0	99%	Query_26245
Bmb9393	1740	1740	100%	0.0	99%	Query_26244
6850	1740	1740	100%	0.0	99%	Query_26243
Z172	1740	1740	100%	0.0	99%	Query_26242
XN108	1740	1740	100%	0.0	99%	Query_26241
ED133	1735	1735	100%	0.0	99%	Query_26253
M013	1735	1735	100%	0.0	99%	Query_26252
SA957	1735	1735	100%	0.0	99%	Query_26251
SA40	1735	1735	100%	0.0	99%	Query_26250
JKD6159	1724	1724	100%	0.0	99%	Query_26254
ST398	1535	1535	100%	0.0	96%	Query_26257
71193	1535	1535	100%	0.0	96%	Query_26256
08BA02176	1535	1535	100%	0.0	96%	Query_26255
MRSA_AS3	1535	1535	100%	0.0	96%	Query_26214
MRSA_S62	1535	1535	100%	0.0	96%	Query_26213
MRSA_B53	1535	1535	100%	0.0	96%	Query_26212
HO50960412	1524	1524	100%	0.0	95%	Query_26259
H-EMRSA-15	1524	1524	100%	0.0	95%	Query_26258
TCH60	1502	1502	100%	0.0	95%	Query_26262
CA-347	1502	1502	100%	0.0	95%	Query_26261
55/2053	1502	1502	100%	0.0	95%	Query_26260
MRSA252	1496	1496	100%	0.0	95%	Query_26263
MRSA_IP70	1496	1496	100%	0.0	95%	Query_26211

alpha-hemolysin toxin (*hla*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
MW2	4034	4034	100%	0.0	100%	Query_61790
MSSA476	4034	4034	100%	0.0	100%	Query_61789
CA-347	4006	4006	100%	0.0	99%	Query_61793
USA300_FPR3737	4006	4006	100%	0.0	99%	Query_61780
NCTC8325	4006	4006	100%	0.0	99%	Query_61779
Newman	4006	4006	100%	0.0	99%	Query_61778
USA300_TCH1516	4006	4006	100%	0.0	99%	Query_61777
JKD8008	4006	4006	100%	0.0	99%	Query_61776
VC40	4006	4006	100%	0.0	99%	Query_61775
M1	4006	4006	100%	0.0	99%	Query_61774
USA300-ISMM51	4006	4006	100%	0.0	99%	Query_61773
UA-S391 USA300	4006	4006	100%	0.0	99%	Query_61772
MRSA_B51	4006	4006	100%	0.0	99%	Query_61767
TCH80	4000	4000	100%	0.0	99%	Query_61805
11819-97	4000	4000	100%	0.0	99%	Query_61788
COL	4000	4000	100%	0.0	99%	Query_61784
T0131	4000	4000	100%	0.0	99%	Query_61783
Bmb9393	4000	4000	100%	0.0	99%	Query_61782
NRS100	4000	4000	100%	0.0	99%	Query_61781
MRSA_IP70	4000	4000	100%	0.0	99%	Query_61768
MRSA252	3995	3995	100%	0.0	99%	Query_61807
5/2053	3995	3995	100%	0.0	99%	Query_61806
CN1	3995	3995	100%	0.0	99%	Query_61792
TW20	3995	3995	100%	0.0	99%	Query_61787
Z172	3995	3995	100%	0.0	99%	Query_61786
XN108	3995	3995	100%	0.0	99%	Query_61785
ED133	3989	3989	100%	0.0	99%	Query_61813
M013	3989	3989	100%	0.0	99%	Query_61811
6850	3989	3989	100%	0.0	99%	Query_61810
SA957	3989	3989	100%	0.0	99%	Query_61809
SA40	3989	3989	100%	0.0	99%	Query_61808
ED98	3984	3984	100%	0.0	99%	Query_61797
ST228_16035	3984	3984	100%	0.0	99%	Query_61796
ST228_16125	3984	3984	100%	0.0	99%	Query_61795
ST228_18583	3984	3984	100%	0.0	99%	Query_61794
502A	3984	3984	100%	0.0	99%	Query_61791
N315	3978	3978	100%	0.0	99%	Query_61804
Mu50	3978	3978	100%	0.0	99%	Query_61803
JH1	3978	3978	100%	0.0	99%	Query_61801
Mu3DNA	3978	3978	100%	0.0	99%	Query_61800
04-02981	3978	3978	100%	0.0	99%	Query_61799
ECT-R2	3978	3978	100%	0.0	99%	Query_61798
RF122	3973	3973	100%	0.0	99%	Query_61816
JKD6159	3973	3973	100%	0.0	99%	Query_61812
JH9	3971	3971	100%	0.0	99%	Query_61802
LGA251	3967	3967	100%	0.0	99%	Query_61817
H050960412	3967	3967	100%	0.0	99%	Query_61815
H-EMRSA-15	3967	3967	100%	0.0	99%	Query_61814
ST398	3956	3956	100%	0.0	99%	Query_61820
71193	3956	3956	100%	0.0	99%	Query_61819
08BA02176	3956	3956	100%	0.0	99%	Query_61818
MRSA_AS3	3956	3956	100%	0.0	99%	Query_61771
MRSA_B53	3956	3956	100%	0.0	99%	Query_61769
MRSA_S62	3951	3951	100%	0.0	99%	Query_61770

Penicillin binding protein (*pbp*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> N315	1901	1901	100%	0.0	100%	Query_128713
<input type="checkbox"/> Mu50	1901	1901	100%	0.0	100%	Query_128712
<input type="checkbox"/> JH9	1901	1901	100%	0.0	100%	Query_128711
<input type="checkbox"/> JH1	1901	1901	100%	0.0	100%	Query_128710
<input type="checkbox"/> Mu3	1901	1901	100%	0.0	100%	Query_128709
<input type="checkbox"/> ECT-R2	1901	1901	100%	0.0	100%	Query_128708
<input type="checkbox"/> ST228_16035	1901	1901	100%	0.0	100%	Query_128707
<input type="checkbox"/> ST228_16125	1901	1901	100%	0.0	100%	Query_128706
<input type="checkbox"/> ST228_18583	1901	1901	100%	0.0	100%	Query_128705
<input type="checkbox"/> 502A	1901	1901	100%	0.0	100%	Query_128704
<input type="checkbox"/> ED98	1893	1893	100%	0.0	99%	Query_128714
<input type="checkbox"/> 04-02981	1784	1784	100%	0.0	98%	Query_128715
<input type="checkbox"/> 11819-97	1746	1746	100%	0.0	98%	Query_128717
<input type="checkbox"/> CN1	1746	1746	100%	0.0	98%	Query_128716
<input type="checkbox"/> MRSA_OP100	1746	1746	100%	0.0	98%	Query_128698
<input type="checkbox"/> MRSA_B51	1735	1823	100%	0.0	97%	Query_128699
<input type="checkbox"/> USA300_FPR3737	1729	1729	100%	0.0	97%	Query_128732
<input type="checkbox"/> NCTC8325	1729	1729	100%	0.0	97%	Query_128731
<input type="checkbox"/> Newman	1729	1729	100%	0.0	97%	Query_128730
<input type="checkbox"/> USA300_TCH1516	1729	1729	100%	0.0	97%	Query_128729
<input type="checkbox"/> TW20	1729	1729	100%	0.0	97%	Query_128728
<input type="checkbox"/> JKD6008	1729	1729	100%	0.0	97%	Query_128727
<input type="checkbox"/> T0131	1729	1729	100%	0.0	97%	Query_128726
<input type="checkbox"/> VC40	1729	1729	100%	0.0	97%	Query_128725
<input type="checkbox"/> M1	1729	1729	100%	0.0	97%	Query_128724
<input type="checkbox"/> Bmb9393	1729	1729	100%	0.0	97%	Query_128723
<input type="checkbox"/> Z172	1729	1729	100%	0.0	97%	Query_128722
<input type="checkbox"/> USA300-ISMM51	1729	1729	100%	0.0	97%	Query_128721
<input type="checkbox"/> NRS100	1729	1729	100%	0.0	97%	Query_128720
<input type="checkbox"/> UA-S391 USA300	1729	1729	100%	0.0	97%	Query_128719
<input type="checkbox"/> XN108	1729	1729	100%	0.0	97%	Query_128718
<input type="checkbox"/> COL	1724	1724	100%	0.0	97%	Query_128733
<input type="checkbox"/> MSSA476	1694	1826	98%	0.0	99%	Query_128734
<input type="checkbox"/> MW2	1666	1815	100%	0.0	99%	Query_128735
<input type="checkbox"/> ED133	1592	1592	100%	0.0	95%	Query_128736
<input type="checkbox"/> H-EMRSA-15	1589	1833	100%	0.0	96%	Query_128737
<input type="checkbox"/> ST398	1580	1651	100%	0.0	94%	Query_128741
<input type="checkbox"/> 71193	1580	1651	100%	0.0	94%	Query_128740
<input type="checkbox"/> 08BA02176	1580	1651	100%	0.0	94%	Query_128739
<input type="checkbox"/> JKD6159	1568	1568	100%	0.0	94%	Query_128742
<input type="checkbox"/> 6850	1546	1546	100%	0.0	94%	Query_128743
<input type="checkbox"/> LGA251	1544	1660	95%	0.0	95%	Query_128744
<input type="checkbox"/> M013	1530	1530	100%	0.0	93%	Query_128747
<input type="checkbox"/> SA957	1530	1530	100%	0.0	93%	Query_128746
<input type="checkbox"/> SA40	1530	1530	100%	0.0	93%	Query_128745
<input type="checkbox"/> MRSA_S62	1522	1522	92%	0.0	95%	Query_128702
<input type="checkbox"/> MRSA_B53	1513	1513	100%	0.0	93%	Query_128701
<input type="checkbox"/> MRSA_AS3	1511	1582	96%	0.0	95%	Query_128703
<input type="checkbox"/> MRSA_IP70	1507	1566	100%	0.0	93%	Query_128700
<input type="checkbox"/> RF122	1495	1630	100%	0.0	96%	Query_128748
<input type="checkbox"/> CA-347	1493	1493	100%	0.0	93%	Query_128749
<input type="checkbox"/> 55/2053	1474	1722	99%	0.0	94%	Query_128750
<input type="checkbox"/> MRSA252	1469	1717	100%	0.0	93%	Query_128752

serine protease; V8 protease; glutamyl endopeptidase (*sspA*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
Mu50	5186	8861	100%	0.0	100%	Query_201955
Mu3	5186	8861	100%	0.0	100%	Query_201954
H050960412	4366	11204	100%	0.0	95%	Query_201957
H-EMRSA-15	4366	11204	100%	0.0	95%	Query_201956
N315	4361	11311	100%	0.0	98%	Query_201960
502A	4361	11311	100%	0.0	98%	Query_201959
ED98	4353	9993	99%	0.0	98%	Query_201958
04-02981	4211	8565	100%	0.0	98%	Query_201962
JH9	4205	8560	100%	0.0	98%	Query_201964
JH1	4205	8560	100%	0.0	98%	Query_201963
ED133	4128	5085	100%	0.0	93%	Query_201965
ST228_16036	3952	10936	92%	0.0	97%	Query_201968
ST228_16126	3952	10936	92%	0.0	97%	Query_201967
ST228_18583	3952	10936	92%	0.0	97%	Query_201966
11819-97	3766	12007	100%	0.0	93%	Query_201969
MSSA476	3692	7772	100%	0.0	91%	Query_201970
TW20	3629	8238	100%	0.0	90%	Query_201971
MW2	3626	8724	100%	0.0	90%	Query_201972
USA300_FPR3757	3624	8235	100%	0.0	90%	Query_201979
NCTC8325	3624	7607	100%	0.0	90%	Query_201978
COL	3624	8235	100%	0.0	90%	Query_201977
USA300_TCH1516	3624	8235	100%	0.0	90%	Query_201976
USA300-ISMM51	3624	8235	100%	0.0	90%	Query_201975
NRS100	3624	8235	100%	0.0	90%	Query_201974
UA-S391_USA300	3624	8235	100%	0.0	90%	Query_201973
Newman	3618	8229	100%	0.0	90%	Query_201984
JKD6008	3618	8229	100%	0.0	90%	Query_201983
VC40	3618	7582	100%	0.0	90%	Query_201982
Bmb9393	3618	8229	100%	0.0	90%	Query_201981
Z172	3618	8190	100%	0.0	90%	Query_201980
LGA251	3602	8670	100%	0.0	92%	Query_201985
JKD6159	3531	7603	100%	0.0	92%	Query_201986
08BA02176	3369	5040	100%	0.0	88%	Query_201987
ST398	3339	5390	100%	0.0	88%	Query_201988
XN108	3251	8956	100%	0.0	90%	Query_201989
M1	3184	12592	100%	0.0	90%	Query_201990
SA957	3158	7248	100%	0.0	89%	Query_201992
SA40	3158	7248	100%	0.0	89%	Query_201991
M013	3153	7243	100%	0.0	89%	Query_201993
CN1	3133	8905	100%	0.0	89%	Query_201994
T0131	3097	5743	100%	0.0	90%	Query_201995
6850	3007	5846	100%	0.0	88%	Query_201996
RF122	2966	6700	100%	0.0	88%	Query_201997
71193	2963	6231	100%	0.0	88%	Query_201998
MRSA262	2929	8376	100%	0.0	88%	Query_201999
CA-347	2918	5711	100%	0.0	87%	Query_202000
55/2053	2880	6887	100%	0.0	87%	Query_202001

fibronectin-binding protein (*fnb*)

Description	Query cover	E value	Ident
<input type="checkbox"/> M1	100%	0.0	100%
<input type="checkbox"/> USA300_FPR3757	100%	0.0	100%
<input type="checkbox"/> NCTC8325	100%	0.0	100%
<input type="checkbox"/> COL	100%	0.0	100%
<input type="checkbox"/> Newman	100%	0.0	100%
<input type="checkbox"/> USA300_TCH1516	100%	0.0	100%
<input type="checkbox"/> VC40	100%	0.0	100%
<input type="checkbox"/> USA300-ISMMS1	100%	0.0	100%
<input type="checkbox"/> NRS100	100%	0.0	100%
<input type="checkbox"/> UA-S391_USA300	100%	0.0	100%
<input type="checkbox"/> TW20	100%	0.0	99%
<input type="checkbox"/> JKD6008	100%	0.0	99%
<input type="checkbox"/> T0131	100%	0.0	99%
<input type="checkbox"/> Bmb9393	100%	0.0	99%
<input type="checkbox"/> Z172	100%	0.0	99%
<input type="checkbox"/> XN108	100%	0.0	99%
<input type="checkbox"/> 11819-97	100%	0.0	99%
<input type="checkbox"/> CN1	100%	0.0	99%
<input type="checkbox"/> HO50960412	100%	0.0	99%
<input type="checkbox"/> H-EMRSA-15	100%	0.0	99%
<input type="checkbox"/> N315	100%	0.0	99%
<input type="checkbox"/> Mu50	100%	0.0	99%
<input type="checkbox"/> JH9	100%	0.0	99%
<input type="checkbox"/> JH1	100%	0.0	99%
<input type="checkbox"/> Mu3	100%	0.0	99%
<input type="checkbox"/> ED98	100%	0.0	99%
<input type="checkbox"/> 04-02981	100%	0.0	99%
<input type="checkbox"/> ECT-R2	100%	0.0	99%
<input type="checkbox"/> ST228_16035	100%	0.0	99%
<input type="checkbox"/> ST228_16125	100%	0.0	99%
<input type="checkbox"/> ST228_18583	100%	0.0	99%
<input type="checkbox"/> 502A	100%	0.0	99%
<input type="checkbox"/> M013	100%	0.0	99%
<input type="checkbox"/> SA957	100%	0.0	99%
<input type="checkbox"/> SA40	100%	0.0	99%
<input type="checkbox"/> MW2	100%	0.0	99%
<input type="checkbox"/> MSSA476	100%	0.0	99%
<input type="checkbox"/> MRSA_B51	100%	0.0	99%
<input type="checkbox"/> MRSA_OP100	100%	0.0	99%
<input type="checkbox"/> LGA251	100%	0.0	99%
<input type="checkbox"/> 6850	100%	0.0	99%
<input type="checkbox"/> RF122	100%	0.0	99%
<input type="checkbox"/> ED133	100%	0.0	99%
<input type="checkbox"/> JKD6159	100%	0.0	99%
<input type="checkbox"/> CA-347	100%	0.0	97%
<input type="checkbox"/> TCH60	100%	0.0	97%
<input type="checkbox"/> 55/2053	100%	0.0	97%
<input type="checkbox"/> MRSA252	100%	0.0	97%
<input type="checkbox"/> MRSA_IP70	100%	0.0	97%
<input type="checkbox"/> ST398	100%	0.0	97%
<input type="checkbox"/> 71193	100%	0.0	97%
<input type="checkbox"/> 08BA02176	100%	0.0	97%
<input type="checkbox"/> MRSA_AS3	100%	0.0	97%
<input type="checkbox"/> MRSA_S62	100%	0.0	97%
<input type="checkbox"/> MRSA_B53	100%	0.0	97%

cysteine protease (*sspC*)

Description	Max score	Total	Query cover	E value	Ident	Accession
<input type="checkbox"/> N315	1269	1269	100%	0.0	100%	Query_43757
<input type="checkbox"/> Mu50	1269	1269	100%	0.0	100%	Query_43756
<input type="checkbox"/> JH9	1269	1269	100%	0.0	100%	Query_43755
<input type="checkbox"/> JH1	1269	1269	100%	0.0	100%	Query_43754
<input type="checkbox"/> Mu3	1269	1269	100%	0.0	100%	Query_43753
<input type="checkbox"/> ED98	1269	1269	100%	0.0	100%	Query_43752
<input type="checkbox"/> 0402981	1269	1269	100%	0.0	100%	Query_43751
<input type="checkbox"/> ECT-R2	1269	1269	100%	0.0	100%	Query_43750
<input type="checkbox"/> ST228_16035	1269	1269	100%	0.0	100%	Query_43749
<input type="checkbox"/> ST228_16125	1269	1269	100%	0.0	100%	Query_43748
<input type="checkbox"/> ST228_18583	1269	1269	100%	0.0	100%	Query_43747
<input type="checkbox"/> 502A	1269	1269	100%	0.0	100%	Query_43746
<input type="checkbox"/> MRSA_B51	1269	1269	100%	0.0	100%	Query_43742
<input type="checkbox"/> H050960412	1253	1253	100%	0.0	99%	Query_43759
<input type="checkbox"/> H-EMRSA-15	1253	1253	100%	0.0	99%	Query_43758
<input type="checkbox"/> ED133	1242	1242	100%	0.0	99%	Query_43762
<input type="checkbox"/> LGA251	1242	1242	100%	0.0	99%	Query_43761
<input type="checkbox"/> CN1	1242	1242	100%	0.0	99%	Query_43760
<input type="checkbox"/> RF122	1236	1236	100%	0.0	99%	Query_43768
<input type="checkbox"/> JKD6159	1236	1236	100%	0.0	99%	Query_43767
<input type="checkbox"/> M013	1236	1236	100%	0.0	99%	Query_43766
<input type="checkbox"/> 8850	1236	1236	100%	0.0	99%	Query_43765
<input type="checkbox"/> SA957	1236	1236	100%	0.0	99%	Query_43764
<input type="checkbox"/> SA40	1236	1236	100%	0.0	99%	Query_43763
<input type="checkbox"/> MW2	1230	1230	100%	0.0	99%	Query_43770
<input type="checkbox"/> MSSA476	1230	1230	100%	0.0	99%	Query_43769
<input type="checkbox"/> MRSA_OP100	1230	1230	100%	0.0	99%	Query_43741
<input type="checkbox"/> USA300_FPR3757	1225	1225	100%	0.0	99%	Query_43786
<input type="checkbox"/> NCTC8325	1225	1225	100%	0.0	99%	Query_43785
<input type="checkbox"/> COL	1225	1225	100%	0.0	99%	Query_43784
<input type="checkbox"/> Newman	1225	1225	100%	0.0	99%	Query_43783
<input type="checkbox"/> USA300_TCH1516	1225	1225	100%	0.0	99%	Query_43782
<input type="checkbox"/> TW20	1225	1225	100%	0.0	99%	Query_43781
<input type="checkbox"/> JKD6008	1225	1225	100%	0.0	99%	Query_43780
<input type="checkbox"/> T0131	1225	1225	100%	0.0	99%	Query_43779
<input type="checkbox"/> VC40	1225	1225	100%	0.0	99%	Query_43778
<input type="checkbox"/> M1	1225	1225	100%	0.0	99%	Query_43777
<input type="checkbox"/> Bmb9393	1225	1225	100%	0.0	99%	Query_43776
<input type="checkbox"/> Z172	1225	1225	100%	0.0	99%	Query_43775
<input type="checkbox"/> USA300-ISMMS1	1225	1225	100%	0.0	99%	Query_43774
<input type="checkbox"/> NRS100	1225	1225	100%	0.0	99%	Query_43773
<input type="checkbox"/> UAS391_USA300	1225	1225	100%	0.0	99%	Query_43772
<input type="checkbox"/> XN108	1225	1225	100%	0.0	99%	Query_43771
<input type="checkbox"/> ST398	1219	1219	100%	0.0	99%	Query_43792
<input type="checkbox"/> TCH60	1219	1219	100%	0.0	99%	Query_43791
<input type="checkbox"/> 11819-97	1219	1219	100%	0.0	99%	Query_43790
<input type="checkbox"/> 71193	1219	1219	100%	0.0	99%	Query_43789
<input type="checkbox"/> 08BA02176	1219	1219	100%	0.0	99%	Query_43788
<input type="checkbox"/> 55/2053	1219	1219	100%	0.0	99%	Query_43787
<input type="checkbox"/> MRSA_S62	1219	1219	100%	0.0	99%	Query_43745
<input type="checkbox"/> MRSA_B53	1219	1219	100%	0.0	99%	Query_43744
<input type="checkbox"/> MRSA_IP70	1219	1219	100%	0.0	99%	Query_43743
<input type="checkbox"/> MRSA252	1214	1214	100%	0.0	99%	Query_43793
<input type="checkbox"/> CA-347	1203	1203	100%	0.0	98%	Query_43794

nuclease (*nuc*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
N315	909	909	100%	0.0	100%	Query_80935
04-02981	909	909	100%	0.0	100%	Query_80934
ECT-R2	909	909	100%	0.0	100%	Query_80933
MSSA476	904	904	100%	0.0	99%	Query_80943
Mu50	904	904	100%	0.0	99%	Query_80942
Mu3	904	904	100%	0.0	99%	Query_80941
11819-97	904	904	100%	0.0	99%	Query_80940
ST228_16036	904	904	100%	0.0	99%	Query_80939
ST228_16125	904	904	100%	0.0	99%	Query_80938
ST228_18583	904	904	100%	0.0	99%	Query_80937
M1	904	904	100%	0.0	99%	Query_80936
MRSA_OP100	904	904	100%	0.0	99%	Query_80930
MW2	898	898	100%	0.0	99%	Query_80952
NCTC8325	898	898	100%	0.0	99%	Query_80951
TW20	898	898	100%	0.0	99%	Query_80950
JKD6008	898	898	100%	0.0	99%	Query_80949
T0131	898	898	100%	0.0	99%	Query_80948
6850	898	898	100%	0.0	99%	Query_80947
Z172	898	898	100%	0.0	99%	Query_80946
502A	898	898	100%	0.0	99%	Query_80945
XN108	898	898	100%	0.0	99%	Query_80944
MRSA_B51	898	898	100%	0.0	99%	Query_80931
MRSA252	893	893	100%	0.0	99%	Query_80965
USA300_FPR3757	893	893	100%	0.0	99%	Query_80964
JH9	893	893	100%	0.0	99%	Query_80963
JH1	893	893	100%	0.0	99%	Query_80962
Newman	893	893	100%	0.0	99%	Query_80961
USA300_TCH1516	893	893	100%	0.0	99%	Query_80960
JKD6159	893	893	100%	0.0	99%	Query_80959
TCH60	893	893	100%	0.0	99%	Query_80958
CA-347	893	893	100%	0.0	99%	Query_80957
Bmb9393	893	893	100%	0.0	99%	Query_80956
55-2053	893	893	100%	0.0	99%	Query_80955
USA300-ISMMS1	893	893	100%	0.0	99%	Query_80954
UA-S391_USA300	893	893	100%	0.0	99%	Query_80953
MRSA_IP70	893	893	100%	0.0	99%	Query_80932
MSHR1132	887	887	100%	0.0	99%	Query_80969
HOS0980412	887	887	100%	0.0	99%	Query_80968
CN1	887	887	100%	0.0	99%	Query_80967
H-EMRSA-15	887	887	100%	0.0	99%	Query_80966

staphylokinase gene (*Sak*)

Alignments Download Graphics Distance tree of results

Description	Query cover	E value	Ident
<input type="checkbox"/> MRSA252	100%	0.0	100%
<input type="checkbox"/> USA300_FPR3757	100%	0.0	100%
<input type="checkbox"/> JH9	100%	0.0	100%
<input type="checkbox"/> JH1	100%	0.0	100%
<input type="checkbox"/> Newman	100%	0.0	100%
<input type="checkbox"/> USA300_TCH1516	100%	0.0	100%
<input type="checkbox"/> JKD6159	100%	0.0	100%
<input type="checkbox"/> TCH80	100%	0.0	100%
<input type="checkbox"/> M013	100%	0.0	100%
<input type="checkbox"/> 71193	100%	0.0	100%
<input type="checkbox"/> HD50960412	100%	0.0	100%
<input type="checkbox"/> CA-347	100%	0.0	100%
<input type="checkbox"/> CN1	100%	0.0	100%
<input type="checkbox"/> SA957	100%	0.0	100%
<input type="checkbox"/> USA300-ISMMS1	100%	0.0	100%
<input type="checkbox"/> H-EMRSA-15	100%	0.0	100%
<input type="checkbox"/> UA-S391_USA300	100%	0.0	100%
<input type="checkbox"/> MRSA_IP70	100%	0.0	100%
<input type="checkbox"/> Bmb9393	100%	0.0	99%
<input type="checkbox"/> 6850	100%	0.0	99%
<input type="checkbox"/> NCTC8325	100%	0.0	99%
<input type="checkbox"/> 502A	100%	0.0	99%
<input type="checkbox"/> MRSA_851	100%	0.0	99%
<input type="checkbox"/> N315DNA	100%	0.0	99%
<input type="checkbox"/> 04-02981	100%	0.0	99%
<input type="checkbox"/> ECT-R2	100%	0.0	99%
<input type="checkbox"/> 55/2053	100%	0.0	99%

chemotaxis inhibitory protein (*chps*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
ECT-R2	1696	1696	100%	0.0	100%	Query_167497
N315	1696	1696	100%	0.0	100%	Query_167458
Mu50	1696	1696	100%	0.0	100%	Query_167457
JH9	1696	1696	100%	0.0	100%	Query_167456
JH1	1696	1696	100%	0.0	100%	Query_167455
Mu3	1696	1696	100%	0.0	100%	Query_167454
ED98	1696	1696	100%	0.0	100%	Query_167453
04-02981	1696	1696	100%	0.0	100%	Query_167452
ST228_18035	1696	1696	100%	0.0	100%	Query_167451
ST228_16125	1696	1696	100%	0.0	100%	Query_167450
ST228_18583	1696	1696	100%	0.0	100%	Query_167449
502A	1696	1696	100%	0.0	100%	Query_167448
MW2	1690	1690	100%	0.0	99%	Query_167476
MSSA476	1690	1690	100%	0.0	99%	Query_167475
USA300_FPR3757	1690	1690	100%	0.0	99%	Query_167474
NCTC8325	1690	1690	100%	0.0	99%	Query_167473
COL	1690	1690	100%	0.0	99%	Query_167472
Newman	1690	1690	100%	0.0	99%	Query_167471
USA300_TCH1516	1690	1690	100%	0.0	99%	Query_167470
TW20	1690	1690	100%	0.0	99%	Query_167469
JKD6008	1690	1690	100%	0.0	99%	Query_167468
T0131	1690	1690	100%	0.0	99%	Query_167467
VC40	1690	1690	100%	0.0	99%	Query_167466
M1	1690	1690	100%	0.0	99%	Query_167465
Bmb9393	1690	1690	100%	0.0	99%	Query_167464
Z172	1690	1690	100%	0.0	99%	Query_167463
USA300-ISMMS1	1690	1690	100%	0.0	99%	Query_167462
NRS100	1690	1690	100%	0.0	99%	Query_167461
UA_S391_USA300	1690	1690	100%	0.0	99%	Query_167460
XN108	1690	1690	100%	0.0	99%	Query_167459
HO50980412	1685	1685	100%	0.0	99%	Query_167478
H-EMRSA-15	1685	1685	100%	0.0	99%	Query_167477
11819-97	1679	1679	100%	0.0	99%	Query_167479
CN1	1674	1674	100%	0.0	99%	Query_167480
MRSA_B51	1674	1674	100%	0.0	99%	Query_167443
RF122	1668	1668	100%	0.0	99%	Query_167482
CA-347	1668	1668	100%	0.0	99%	Query_167481
ST398	1663	1663	100%	0.0	99%	Query_167490
ED133	1663	1663	100%	0.0	99%	Query_167489
M013	1663	1663	100%	0.0	99%	Query_167488
71193	1663	1663	100%	0.0	99%	Query_167487
088A02176	1663	1663	100%	0.0	99%	Query_167486
6850	1663	1663	100%	0.0	99%	Query_167485
SA957	1663	1663	100%	0.0	99%	Query_167484
SA40	1663	1663	100%	0.0	99%	Query_167483
MRSA_AS3	1663	1663	100%	0.0	99%	Query_167447
MRSA_S62	1663	1663	100%	0.0	99%	Query_167446
MRSA_B53	1663	1663	100%	0.0	99%	Query_167445
MRSA262	1657	1657	100%	0.0	99%	Query_167493
TCH60	1657	1657	100%	0.0	99%	Query_167492
55/2053	1657	1657	100%	0.0	99%	Query_167491
MRSA_IP70	1657	1657	100%	0.0	99%	Query_167444
LGA251	1652	1652	100%	0.0	99%	Query_167494
JKD6159	1646	1646	100%	0.0	99%	Query_167495

lipic acid synthetase (*lipA*)

Alignments Download Graphics Distance tree of results						
Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> USA300_FPR3757	3829	3829	100%	0.0	100%	Query_53554
<input type="checkbox"/> USA300_TCH1516	3829	3829	100%	0.0	100%	Query_53553
<input type="checkbox"/> USA300-ISMMS1	3829	3829	100%	0.0	100%	Query_53552
<input type="checkbox"/> UA-S391_USA300	3829	3829	100%	0.0	100%	Query_53551
<input type="checkbox"/> NCTC8325	3823	3823	100%	0.0	99%	Query_53556
<input type="checkbox"/> M1	3823	3823	100%	0.0	99%	Query_53555
<input type="checkbox"/> VC40	3818	3818	100%	0.0	99%	Query_53557
<input type="checkbox"/> MW2	3784	3784	100%	0.0	99%	Query_53559
<input type="checkbox"/> MSSA476	3784	3784	100%	0.0	99%	Query_53558
<input type="checkbox"/> MRSA_OP100	3784	3784	100%	0.0	99%	Query_53545
<input type="checkbox"/> 11819-97	3731	3731	100%	0.0	99%	Query_53561
<input type="checkbox"/> CN1	3731	3731	100%	0.0	99%	Query_53560
<input type="checkbox"/> N315	3725	3725	100%	0.0	99%	Query_53571
<input type="checkbox"/> Mu50	3725	3725	100%	0.0	99%	Query_53570
<input type="checkbox"/> Mu3	3725	3725	100%	0.0	99%	Query_53569
<input type="checkbox"/> ED98	3725	3725	100%	0.0	99%	Query_53568
<input type="checkbox"/> 04-02981	3725	3725	100%	0.0	99%	Query_53567
<input type="checkbox"/> ECT-R2	3725	3725	100%	0.0	99%	Query_53566
<input type="checkbox"/> ST228_16035	3725	3725	100%	0.0	99%	Query_53565
<input type="checkbox"/> ST228_16125	3725	3725	100%	0.0	99%	Query_53564
<input type="checkbox"/> ST228_16583	3725	3725	100%	0.0	99%	Query_53563
<input type="checkbox"/> 502A	3725	3725	100%	0.0	99%	Query_53562
<input type="checkbox"/> MRSA_B51	3692	3692	100%	0.0	99%	Query_53546
<input type="checkbox"/> M013	3598	3598	100%	0.0	98%	Query_53574
<input type="checkbox"/> SA957	3598	3598	100%	0.0	98%	Query_53573
<input type="checkbox"/> SA40	3598	3598	100%	0.0	98%	Query_53572
<input type="checkbox"/> LGA251	3592	3592	100%	0.0	98%	Query_53575
<input type="checkbox"/> 6850	3587	3587	100%	0.0	98%	Query_53576
<input type="checkbox"/> HO50960412	3581	3581	100%	0.0	98%	Query_53578
<input type="checkbox"/> H-EMRSA-15	3581	3581	100%	0.0	98%	Query_53577
<input type="checkbox"/> MRSA252	3565	3565	100%	0.0	98%	Query_53583
<input type="checkbox"/> JKD6008	3565	3565	100%	0.0	98%	Query_53582
<input type="checkbox"/> T0131	3565	3565	100%	0.0	98%	Query_53581
<input type="checkbox"/> Bmb9393	3565	3565	100%	0.0	98%	Query_53580
<input type="checkbox"/> 55/2053	3565	3565	100%	0.0	98%	Query_53579
<input type="checkbox"/> MRSA_IP70	3565	3565	100%	0.0	98%	Query_53547
<input type="checkbox"/> COL	3563	3563	93%	0.0	99%	Query_53586
<input type="checkbox"/> Newman	3563	3563	93%	0.0	99%	Query_53585
<input type="checkbox"/> NRS100	3563	3563	93%	0.0	99%	Query_53584
<input type="checkbox"/> TCH80	3559	3559	100%	0.0	98%	Query_53587
<input type="checkbox"/> TW20	3546	3546	100%	0.0	98%	Query_53595
<input type="checkbox"/> JKD6159	3537	3537	100%	0.0	97%	Query_53588
<input type="checkbox"/> MRSA_B53	3515	3515	100%	0.0	97%	Query_53548
<input type="checkbox"/> 71193	3509	3509	100%	0.0	97%	Query_53589
<input type="checkbox"/> MRSA_AS3	3509	3509	100%	0.0	97%	Query_53550
<input type="checkbox"/> MRSA_S62	3504	3504	100%	0.0	97%	Query_53549
<input type="checkbox"/> JH9	3470	3470	93%	0.0	99%	Query_53591
<input type="checkbox"/> JH1	3470	3470	93%	0.0	99%	Query_53590
<input type="checkbox"/> ED133	3376	3376	93%	0.0	98%	Query_53592
<input type="checkbox"/> RF122	3343	3343	93%	0.0	98%	Query_53593
<input type="checkbox"/> ST398	3293	3293	93%	0.0	97%	Query_53596
<input type="checkbox"/> 08BA02176	3288	3288	93%	0.0	97%	Query_53597

triacylglycerol lipase (*lip1*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> COL	3707	3707	100%	0.0	100%	Query_8678
<input type="checkbox"/> 11819-97	3707	3707	100%	0.0	100%	Query_8677
<input type="checkbox"/> HOS0960412	3707	3707	100%	0.0	100%	Query_8676
<input type="checkbox"/> SA40	3707	3707	100%	0.0	100%	Query_8675
<input type="checkbox"/> H-EMRSA-15	3707	3707	100%	0.0	100%	Query_8674
<input type="checkbox"/> MRSA252	3701	3701	100%	0.0	99%	Query_8695
<input type="checkbox"/> USA300_FPR3757	3701	3701	100%	0.0	99%	Query_8694
<input type="checkbox"/> JH9	3701	3701	100%	0.0	99%	Query_8693
<input type="checkbox"/> JH1	3701	3701	100%	0.0	99%	Query_8692
<input type="checkbox"/> Mu3	3701	3701	100%	0.0	99%	Query_8691
<input type="checkbox"/> USA300_TCH1516	3701	3701	100%	0.0	99%	Query_8690
<input type="checkbox"/> JKD6008	3701	3701	100%	0.0	99%	Query_8689
<input type="checkbox"/> JKD6159	3701	3701	100%	0.0	99%	Query_8688
<input type="checkbox"/> 04-02981	3701	3701	100%	0.0	99%	Query_8687
<input type="checkbox"/> TCH60	3701	3701	100%	0.0	99%	Query_8686
<input type="checkbox"/> T0131	3701	3701	100%	0.0	99%	Query_8685
<input type="checkbox"/> MSHR1132	3701	3701	100%	0.0	99%	Query_8684
<input type="checkbox"/> M1	3701	3701	100%	0.0	99%	Query_8683
<input type="checkbox"/> CA-347	3701	3701	100%	0.0	99%	Query_8682
<input type="checkbox"/> CN1	3701	3701	100%	0.0	99%	Query_8681
<input type="checkbox"/> USA300-ISMMS1	3701	3701	100%	0.0	99%	Query_8680
<input type="checkbox"/> UA-S391_USA300	3701	3701	100%	0.0	99%	Query_8679
<input type="checkbox"/> MRSA_AS3	3701	3701	100%	0.0	99%	Query_8673
<input type="checkbox"/> MRSA_S62	3701	3701	100%	0.0	99%	Query_8672
<input type="checkbox"/> MRSA_B53	3701	3701	100%	0.0	99%	Query_8671

Mec A gene (*meca*)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> USA300_FPR3757	1347	1347	100%	0.0	100%	Query_30799
<input type="checkbox"/> USA300_TCH1516	1347	1347	100%	0.0	100%	Query_30798
<input type="checkbox"/> USA300-ISMMS1	1347	1347	100%	0.0	100%	Query_30797
<input type="checkbox"/> UA-S391_USA300	1347	1347	100%	0.0	100%	Query_30796
<input type="checkbox"/> COL	1269	1269	100%	0.0	98%	Query_30801
<input type="checkbox"/> NRS100	1269	1269	100%	0.0	98%	Query_30800
<input type="checkbox"/> TW20	1264	1264	100%	0.0	98%	Query_30808
<input type="checkbox"/> JKD6008	1264	1264	100%	0.0	98%	Query_30807
<input type="checkbox"/> T0131	1264	1264	100%	0.0	98%	Query_30806
<input type="checkbox"/> M013	1264	1264	100%	0.0	98%	Query_30805
<input type="checkbox"/> SA957	1264	1264	100%	0.0	98%	Query_30804
<input type="checkbox"/> SA40	1264	1264	100%	0.0	98%	Query_30803
<input type="checkbox"/> Z172	1264	1264	100%	0.0	98%	Query_30802
<input type="checkbox"/> XN108	1258	1258	100%	0.0	98%	Query_30809
<input type="checkbox"/> MW2	1208	1208	100%	0.0	97%	Query_30811
<input type="checkbox"/> MSSA476	1208	1208	100%	0.0	97%	Query_30810
<input type="checkbox"/> MRSA_OP100	1208	1208	100%	0.0	97%	Query_30795

enterotoxin K (*sek*)