Molecular epidemiology and transmission dynamics of *S.uberis* bovine clinical mastitis

Peers L Davies MA VetMB MRCVS

Thesis submitted to the University of Nottingham for degree of Doctor of Philosophy

May 2016

Declaration

Author and Description of Work

Own Work

'The thesis must be the result of the candidate's own work. This requirement does not preclude a candidate obtaining limited assistance with the routine collection and/or processing of data under guidelines and instructions clearly devised by the candidate. When such help is obtained it should be with the prior approval of the supervisor who must be satisfied that the spirit of the 'own work' requirement is not breached'.

Describe any assistance provided in the generation of the work presented here.

1. MLST sanger sequencing was conducted by Oxford University Sequencing Service as a technical subcontractor service provider on S.uberis gDNA samples that were recovered, classified, DNA extracted and purified by Peers Davies. All analysis and interpretation of the MLST results was conducted by Peers Davies.
2. NextSeq DNA sequencing was carried out and DeepSeq, QMC University of Nottingham as technical subcontractor service provider on gDNA isolates selected, extracted and purified by Peers Davies. All analysis and interpretation of NextSeq results was conducted by Peers Davies.
3. QMMS Ltd provided training to Peers Davies on bacterial recovery and MALDI-ToF techniques and extraction protocols. All bacterial isolates were recovered by Peers Davies from archived samples stored at -80 degrees Celsius at QMMS ltd, Wells, Somerset.
4. MALDI-ToF bacterial extraction and ClinProTools analysis training was provided to Peers Davies by Bruker at University of Liverpool and Dr Simon Archer, University of Nottingham.
5. Assistance to Peers Davies with MALDI trace quality control assessment was provided by Dr Selin Cooper at QMMS Ltd. Bacterial proteomic extraction and MALDI-ToF assessment was carried out by Peers Davies and Dr Cooper, All analysis and interpretation of MALDI results a carried out by Peers Davies.
6. Assistance with alternate WGS assembly methods was provided by ADAC, University of Nottingham as cross validation of WGS assembly carried out by Peers Davies.

I confirm that this is an accurate reflection of the assistance provided with the prior approval of my supervisor. I do not believe that the spirit of the 'own work' requirement has been breached.

Prior Publication

'Prior publications of papers arising from the research being undertaken will not adversely prejudice the assessment of the thesis by the Examiners'.

Provide references for any prior publications based in whole or in part on work presented here.

Davies, P. L. et al. Molecular Epidemiology of Streptococcus uberis clinical Mastitis in Dairy Herds: Strain Heterogeneity and Transmission. J. Clin. Microbiol. 54, 1–23 (2016).

I have reviewed the final written submission and the associated 'Turn it in' report with my supervisor. I have not identified any journal copyright issues related to prior publications based in whole or in part on the work presented here, and I do not believe that the spirit of the 'own work' requirement has been breached.

Abstract

Mastitis remains one of the most common, costly and intractable diseases affecting the dairy cattle industry worldwide. In spite of concerted efforts meaningful progress in reducing the incidence of mastitis has been limited over the past thirty years by our partial understanding of the epidemiology of key pathogens, such as *S.uberis*.

The aim of this thesis was to investigate the epidemiology of *S.uberis* mastitis by analysis of the population structure and transmission dynamics of clinical mastitis patterns within and between commercial dairy herds in England and Wales.

In Chapter 3 Multi Locus Sequence Typing (MLST) was used to describe the sub-species (sequence type) genetic heterogeneity of *S.uberis* isolates collected from a longitudinal study of clinical mastitis on 52 farms. The clinical cases were classified according the pattern of occurrence of the sequence types in the herds. The findings suggested that a small subset of MLST sequence types were disproportionally important in the epidemiology of *S.uberis* mastitis, with cow-to-cow transmission of *S.uberis,* potentially occurring in the majority of herds in the United Kingdom; this may be the most important route of *S.uberis* transmission in approximately one third of herds.

In Chapter 4 cow and herd level variables, including monthly recorded milk constituents, yield and parity were evaluated against the clinical case classifications defined in Chapter 3. The temporal relationships between

clinical cases and classifications were also evaluated as potential predictors of transmission dynamics within a herd. The findings indicated that the time interval between clinical cases classified by genotype as potentially contagious transmission was significantly shorter than that between successive mastitis clinical cases attributed to environmental transmission. The distribution of clinical cases throughout lactation also indicated a higher proportion of potentially contagious isolates were cultured from clinical cases originating from lactation rather than the dry period compared to those attributed to environmental transmission.

In Chapter 5 Matrix Assisted Laser Desorption (MALDI) mass spectroscopy was used to generate spectral profiles of *S.uberis* isolates cellular composition. Spectral profiles were used successfully as an alternative method of discriminating between clinical mastitis isolates associated with contagious transmission from those associated with and environmental origin of infection defined in Chapter 3. The findings of this chapter demonstrated marked variation between herds in the spectral profiles of isolates from the same clinical case classification.

In Chapter 6 selected isolates of *S.uberis* associated with contagious transmission and persistent intramammary infections were sequenced using next generation technology and compared by core genome multi locus sequence typing (cgMLST) to evaluate the discriminatory capacity of the MLST and MALDI. The results from this chapter confirmed the importance of the herd unit in the genotypic population structure of *S.uberis* suggested by the results of Chapter 5 and also support the results of Chapter3 which suggest

‘low grade’ contagious transmission of *S.uberis* is superimposed on a ubiquitous, environmental *S.uberis* mastitis pattern in many herds

This thesis demonstrated that the *S.uberis* population is complex with variation at the bacterial, cow and herd level suggesting different patterns of disease and bacterial evolution occurring in different circumstances. Categorisation of sub-species of *S.uberis* within herds as being transmitted via contagious or environmental routes appears to be eminently achievable using modern, high throughput technologies; this could lead to a step change in mastitis control.

Acknowledgements

There are many people I would like to thank for their practical help and moral support, generously given over the past four years and which I have enormously appreciated. First and foremost I would like to thank Professor Martin Green for your, boundless enthusiasm and phenomenal patience and encouragement throughout what must have felt like a very long four years!

I would also like to thank my secondary supervisors Dr Andrew Bradley, Professor James Leigh and Prof Richard Emes for your invaluable insight and support at various points throughout the project.

I would like to thank Dr Adam Blanchard, Dr Selin Cooper, Dr Sharon Egan, Dr Andrew Warry, Dr Simon Archer and Barbara Payne for your technical advice and expertise at various stages of the project.

The project was funded by AHDB Dairy who have provided great support throughout and in particular I would like to thank Dr Jenny Gibbons and Dr Ray Keatinge.

Finally especially important thanks to Heather and Eira for your love and forbearance and for all the sacrifices you have made for my benefit along the way.

Contents

Acknowledgements 6

Contents 7

Figures 16

Tables 20

Abbreviations 23

Chapter 1: Introduction 24

1.1 Importance of Mastitis 24

1.2 Improving mastitis control 25

1.2.1 Historical perspective on the microbiology of mastitis and associated control strategies 25

1.2.2 Structured national mastitis control programs in the United Kingdom and abroad 25

1.2.3 The 'Contagious' vs 'Environmental' mastitis classification system 31

1.2.4 Classification as a 'Contagious' or 'Environmental' mastitis pattern 31

1.2.5 Significance of 'Contagious' and 'Environmental' patterns 32

1.3 The significance of *S.uberis* as a mastitis pathogen..........................36

1.3.1 Response of *S.uberis* to preventive mastitis control measures .38

1.3.2 Effect of antibiotic treatment on the incidence of *S.uberis* clinical mastitis 39

1.4 Methods of sub-species identification of *S.*uberis.............................40

1.4.1 Colony Morphology 40

1.4.2 Biochemical characteristics 40

1.4.3 Ribotyping..41

1.4.4 Serological typing..42

1.4.5 Pulse Gel Field Electrophoresis (PGFE)......................................42

1.4.6 Random Amplified Polymorphic DNA (RAPD) 43

1.4.1 Multi Locus Enzyme Electrophoresis (MLEE)..............................43

1.4.2 Multi Locus Sequence Typing (MLST) 44

1.4.3 Multi Virulence Locus Sequence Typing (MVLST) 47

1.4.4 Core Genome Multi Locus Sequence Typing (cgMLST) 48

1.4.5 MALDI Mass Spectrometry Typing 49

1.5 Epidemiology of *S.uberis* in dairy cattle mastitis. 51

1.5.1 Global *S.uberis* mastitis population diversity and structure assessed by MLST 52

1.5.2 *S.uberis* sub-species diversity and evolutionary dynamics 53

1.5.3 Relative prevalence of *S.uberis* sub-species with sample origin 56

1.5.4 Evidence of predominant sequence types 57

1.5.5 Evidence for persistent *S.uberis* infections 58

1.5.6 Transmission dynamics of *S.uberis* intramammary infections...60

1.5.7 Evidence of adaptation to the cow mammary environment and determinants of bacterial virulence...62

1.6 Cow factors and susceptibility to 'Contagious' or 'Environmental' mastitis...65

1.6.1 Predictive value of SCC and other milk recording data for *S.uberis* mastitis...66

1.7 Summary of objectives for this research...68

Chapter 2: Materials and Methods...70

2.1 Study population and clinical mastitis samples...70

2.2 Milk sample collection and initial pathogen identification...71

2.3 *S.uberis* isolate archive storage...71

2.4 *S.uberis* recovery protocol...72

2.5 gDNA extraction and quantification...72

2.5.1 Method for Purification of chromosomal DNA from *S.uberis*...74

2.5.2 Preparation of Reagents...76

2.5.3 Final gDNA extraction Protocol...77

2.6 Multi Locus Sequence Typing...79

2.6.1 MLST Sequence Analysis...80

2.7 Whole Genome Sequencing...81

2.7.1 Illumina sequencing carried out by DeepSeq, University of Nottingham 81

2.7.2 Library preparation and sequencing...81

2.7.3 Genome Assembly 82

2.7.4 cgMLST Target quality control parameters 83

2.8 Matrix Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI) 86

2.8.1 Isolate selection and recovery..88

2.8.2 Ethanol-Formic Acid Extraction Protocol 89

2.8.3 Isolate quality assessment of spectral profiles..........................90

Chapter 3: Analysis of the population dynamics of *S.uberis* clinical mastitis using multilocus sequence typing. 91

3.1 Methods...92

3.1.1 Mastitis clinical case classifications 93

3.2 Results...94

3.2.1 MLST sequencing performance analysis 95

3.2.2 MLST sequence type population structure 95

3.2.3 Herd and species level mastitis dynamics 96

3.2.4 Herd level MLST descriptive analysis..97

3.2.5 Cow level MLST descriptive analysis..98

3.2.6 Analysis of epidemiology patterns in MLST sequence type results 100

3.2.7 Analysis of herd level variables on MLST patterns 102

3.2.8 Statistically Overrepresented Sequence Types 104

3.2.9 eBURST analysis of MLST groupings 105

3.3 Discussion...109

Chapter 4: Analysis of clinical case classifications and associated cow level information. 117

4.1 Introduction 117

4.1.1 Analysis of National Milk Records (NMR) individual cow data 118

4.1.2 Analysis of cow parity 119

4.1.3 Analysis of impact of stage of lactation on mastitis case classification 120

4.1.4 Temporal Analysis of relationships between clinical cases by classification group (intercase interval) 120

4.2 Methods...121

4.2.1 Statistical Analysis..121

4.2.2 Clinical case classifications, refined from those define in Chapter 3 122

4.2.3 Data structure 123

4.3 Results 125

4.3.1 Aim 1: Evaluation between clinical case classifications of routinely recorded (NMR) milk sample data: somatic cell count; milk yield; protein concentration; fat concentration and lactose concentration.. 125

4.3.2 Aim 2: An evaluation between clinical case classifications of cow level factors such as parity, seasonality and stage of lactation (days in milk, DIM). 134

4.3.3 Aim 3: An evaluation of the temporal relationship between successive clinical cases of the same case classification (the inter-case interval) 137

4.3.4 Aim 4: An evaluation of herd and cow level factors on clinical case classifications....138

4.4 Discussion....139

4.4.1 Aim 1: An evaluation between clinical case classifications of routinely recorded (NMR) milk sample data: somatic cell count; milk yield; protein concentration; fat concentration and lactose concentration. 139

4.4.2 Aim 2: An evaluation between clinical case classifications of cow level factors such as parity, seasonality and stage of lactation (days in milk, DIM). 141

4.4.3 Aim 3: An evaluation of the temporal relationship between successive clinical cases of the same case classification (the inter-case interval) 143

Chapter 5: MALDI phenotype compared to clinical case classification 145

5.1 Introduction 145

5.2 Methods 146

5.2.1 Sample selection 146

5.2.2 Statistical Analysis..151

5.3 Results 156

5.3.1 Aim 1: MALDI Phenotype of MLST sequence types 156

5.3.2 Aim 2: The Discriminatory ability of clinical case classifications by MALDI Phenotype...165

5.3.3 Aim 3 Determine if MALDI spectral profiles are predictive of epidemiological case classifications in a new herd 168

5.3.4 Aim 4: To Determine the between farm variation in MALDI phenotype of Solitary and PTE isolates. 172

5.4 Discussion...177

Chapter 6: Whole genome sequencing of selected *S.uberis* isolates 184

6.1 Introduction 184

6.2 Methods 186

6.2.1 Isolate selection 186

6.2.2 Sequencing Protocols and Genome Assembly 189

6.2.3 Bioinformatic Analysis 189

6.3 Results 191

6.3.1 Aim 1: Validation of classic MLST scheme sequence type classifications by cgMLST 191

6.3.2 Aim 2: Determine the genetic heterogeneity of classical MLST sequence types within a single herd and between multiple herds using cgMLST analysis. 191

6.3.3 Aim 3: Determine if the genetic heterogeneity between potentially contagious sequence types within a herd correlates with the isolate MALDI spectral profiles. 199

6.3.4 Aim 4: Determine if isolates classified as persistent clinical cases in the same mammary gland according to MLST sequence type are true persistent infections or new infections with the same sequence type. .205

6.3.5 Aim 5: Determine if there is evidence of antigenic variation of three virulence genes (SUB 0145/RS00865, SUB 1095/RS5505, SUB 1154/RS5795) as a possible immune escape mechanism in persistent mastitis case. 209

6.4 Discussion...219

Chapter 7: Discussion 230

7.1 Limitations of the project...236

7.2 Further work 238

7.2.1 Updated and extended multi –herd longitudinal study 238

7.2.2 Sampling beyond the udder 239

7.2.3 Further opportunities for MALDI classification of *S.uberis* 239

7.2.4 Further opportunities for whole genome sequencing of *S.uberis* isolates. 240

7.2.5 Investigation of mobile genetic elements and transcriptome analysis. 240

7.3 Conclusions 241

7.3.1 Chapter 3: 241

7.3.2 Chapter 4: 241

7.3.3 Chapter 5: 242

7.3.4 Chapter 6: 243

Chapter 8: References 244

Figures

Figure 1 Incidence rate of clinical mastitis cases per herd caused by *S.uberis* and other pathogens 97

Figure 2 Number of unique multi locus sequence types of clinical *S.uberis* isolates identified in different dairy herds. 98

Figure 3 Scatter plot of the proportion of the total *S.uberis* clinical mastitis classified as potentially contagious (Index/PTE/Persistent) against herd size. 102

Figure 4 Scatter plot of the proportion of *S.uberis* cases per farm classified as potentially contagious (Index/PTE/Persistent) against the incidence rate of *S.uberis* mastitis per herd during the 16 month study period. 103

Figure 5 Scatter plot of the proportion of *S.uberis* cases classified as potentially contagious (Index/PTE/Persistent) against total mastitis incidence rate over the 16 month study period 103

Figure 6 Mastitis clinical cases by classification group 108

Figure 7 eBURST V3 population plot of sequence types identified 108

Figure 8 Somatic cell count (SCC) distributions of four clinical case classifications 126

Figure 9 Somatic cell count (SCC) natural log transformed distributions of four clinical case classifications: 127

Figure 10 Somatic cell count (SCC) natural log transformed distributions of two clinical case classifications: 128

Figure 11 Somatic cell count (SCC) distributions of (OVR PTE) and Solitary (S). ... 129

Figure 12 Distributions of recorded milk yield; 130

Figure 13 Distributions of recorded milk protein concentration; 131

Figure 14 Distributions of recorded milk fat concentration; 132

Figure 15 Distributions of recorded milk lactose concentration; 133

Figure 16 Frequency of cows by parity of all cows from which *S.uberis* MLST profiles were obtained from their clinical mastitis case. 134

Figure 17 Seasonal distribution of clinical case classifications. 135

Figure 18 Individual value plot of *S.uberis* mastitis case diagnosis relative to stage of lactation (lactation week) against the clinical case classifications: 136

Figure 19 Cumulative proportion plot of all intercase intervals less than 8 weeks for clinical case classifications 137

Figure 20 ST6 (red) compared to single locus variants of ST6 156

Figure 21 ST6 (red) compared to double locus variants of ST6 157

Figure 22 ST6 (red) compared to triple locus variants of ST6 158

Figure 23 ST6 (red) compared to four locus variants of ST6 159

Figure 24 ST6 (red) compared to five locus variants of ST6 160

Figure 25 ST6 (red) compared to six locus variants of ST6 161

Figure 26 ST6 (red) compared to seven locus variants of ST6 162

Figure 27 PCA of MLST sequence types grouped according to the number of locus variants (lv) for ST6. 163

Figure 28 First two most discriminatory peaks for All OVR PTE and all Solitary cases isolates 166

Figure 29 First two most discriminatory peaks for OVR PTE(ICI <8wks) and a randomized sample of Solitary cases isolates 167

Figure 30 PCA 3D plot of OVR PTE(ICI <8wks) and a randomized sample of Solitary cases isolates 168
Figure 31 Pairwise PCA plots for PTE and Solitary spectra from Farm 8. 169
Figure 32 Pairwise PCA plots for PTE and Solitary spectra from Farm 20. 170
Figure 33 Pairwise PCA plots for PTE and Solitary spectra from Farm 27. 170
Figure 34 Pairwise PCA plots for PTE and Solitary spectra from Farm 55. 170
Figure 35 Pairwise PCA plots for PTE and Solitary spectra from Farm 65. 171
Figure 36 Pairwise PCA plots of Multi-farm Solitary case model 172
Figure 37 Pairwise PCA plots of Multi-farm PTE case model 173
Figure 38 First two most discriminatory peaks for Solitary isolates 174
Figure 39 PCA plots of Multi-farm PTE case model. 175
Figure 40 First two most discriminatory peaks for PTE isolates (ex Farm27) 176
Figure 41 First two most discriminatory peaks for PTE isolates (inc Farm 27). 177
Figure 42 Schematic diagram of hypothetical phenotypic differences between case classifications 182
Figure 43 Neighbour joining tree of all 62 cgMLST isolates 193
Figure 44 Minimum Spanning Tree of 62 isolates, label & colour = MLST ST 194
Figure 45 Minimum Spanning Tree, 62 isolates, label = Farm of origin, colour = MLST group. 195
Figure 46 Minimum spanning tree of ST6 isolates from Farms 4, 21, 27 & 33. 196
Figure 47 Distributions of four Farm and ST cgMLST groups. 198
Figure 48 Minimum Spanning Tree of isolates selected from Fm27 200
Figure 49 95% confidence intervals about the median of cgMLST distance matrix values 202

Figure 50 95% confidence intervals about the median of MALDI CCI values...... 204

Figure 51 Dendrogram of 27 persistent clinical mastitis case isolates 206

Figure 52 95% confidence interval plot of the median MALDI CCI values for two persistent isolates (red) or two non-persistent case isolates (blue)............... 209

Figure 53 Dendrogram based upon SNP variation at SUB 0145/RS 0865 locus. Isolates of the same MLST sequence type are grouped by colour. 211

Figure 54 Dendrogram based upon SNP variation at SUB 1095/RS 5505 locus Isolates of the same MLST sequence type 212

Figure 55 Dendrogram based upon SNP variation at SUB 1154/RS 5795 locus Isolates of the same MLST sequence type are grouped by colour. 213

Figure 56 pfam diagram of the structure of SUB 0145/RS00865. Regions correspond to mapped areas described in Table 6 and the numeric scale refers to codon number.. 215

Figure 57 pfam diagram of SUB1095/RS5055. Regions correspond to mapped areas described in Table 9. 216

Figure 58 Diagram of a potential *S.uberis* isolate/gene flow in a dairy herd. 234

Tables

Table 1. The frequency of missing loci from the 278 isolates with partial *S.uberis* MLST profiles..95

Table 2. Distribution of the most prevalent MLST sequence types across all farms. 98

Table 3. Distribution of persistent mastitis clinical cases between MLST sequence types of *S.uberis* from all herds...100

Table 4. Number of *S.uberis* clinical mastitis cases grouped by case classification 104

Table 5. All 35 *S.uberis* MLST sequence types from clinical mastitis isolates associated with potential transmission events (PTE)...................................106

Table 6. The number of individual cow milk recordings falling within each of the six lagged time period 124

Table 7. Frequency of isolates per MALDI analysis group............................149

Table 8. Case classification groups used for MALDI analysis........................150

Table 9. Number of MALDI spectra of PTE and Solitary case classification used for SNN and GA model generation per farm. 151

Table 10. SNN and GA models of CC6 compared to Non-CC6 isolates discriminatory capacity of MALDI...163

Table 11. Pairwise comparison of individual MLST ST model classification of MALDI phenotype...164

Table 12. Clinical case classification classes identified by MALDI phenotype 165

Table 13. Individual farm models of PTE v Solitary case classification discrimination ability using MALDI phenotype..169

Table 14. Multi-farm models of OVR PTE vs Solitary clinical case classifications 171

Table 15. Supervised neural network and Genetic algorithm models of the Solitary isolate populations 172

Table 16. Supervised neural network and Genetic algorithm models of the PTE isolate populations 175

Table 17. Isolates selected for whole genome sequencing..........................188

Table 18. Number of isolates selected per farm for whole genome sequencing 188

Table 19. cgMLST distance matrix group comparison.................................197

Table 20. Mood median test of cgMLST distance matrix Fm27 ST6, ST20 & ST233 202

Table 21. Mood Median Test of MALDI CCI values Fm27 ST6,20,233. Quantitative analysis of within and between ST phenotype heterogeneity using MALDI protein mass spectra. 203

Table 22. cgMLST distance matrix and MALDI CCI score for each persistent isolate combination. 208

Table 23. Protein structure diagram of SUB 0145/RS00865 215

Table 24. Table of all non-synonymous SNP's in persistent clinical cases in SUB 0145/RS865. 216

Table 25. Protein structure diagram of SUB 1095/RS5505. 216

Table 26. Table of all non-synonymous, missense SNP's in persistent clinical cases in SUB 1095/RS5505 217

Table 27. Amino acid sequences of SUB 0145/RS 00865 from six isolates form three persistent clinical mastitis cases 218

Abbreviations

bp	Base pairs
cfu	colony forming units
cgMLST	core genome Multi Locus Sequence Typing
CRISPR	Clustered regularly-interspaced short palindromic repeats
GA	Genetic Algorithm
hr	Hour
MALDI	Matrix Assisted Laser Desorption Time of Flight Mass Spectrometry
ml	Millilitres
MLST	Multi Locus Sequence Typing
OVR	Over Represented sequence type
PCA	Principal Component Analysis
postMTD	Post-milking teat disinfection
preMTD	Pre-milking teat disinfection
PTE	Potential Transmission Event
ST	Sequence Type
S.uberis	Streptococcus uberis
SCC	Somatic Cell Count
SNN	Supervised Neural Network

Chapter 1: Introduction

1.1 Importance of Mastitis

Mastitis is arguably the most significant disease affecting the UK dairy industry today. Clinical and sub-clinical intramammary infections cause financial and production losses in addition to impairing animal welfare. Mastitis has been estimated to account for financial losses in excess of £170 million pounds to the UK dairy industry (Bradley 2002) mostly due to reduced production. Mastitis is also the single most costly disease in terms of treatment in the UK dairy industry, representing 38% of total herd disease treatment costs ahead of lameness at 27% (Kossaibati and Esslemont 1997). Mastitis has been shown to increase the risk of mortality and morbidity leading to increased culling and herd replacement costs (Geary et al. 2012). Mastitis is a major animal welfare concern by virtue of its clinical presentation as an inflammatory disease which is widely accepted to cause pain and discomfort to individuals affected by the condition (Medrano-Galarza et al. 2012). The welfare implications of mastitis were highlighted in the UK Farm Animal Welfare Council Report (FAWC, 2009). Mastitis treatment and prevention is a major contributor to antibiotic use in the UK dairy industry and has been the focus of efforts to reduce antibiotic usage in several European countries (Østerås and Sølverød 2009). Finally mastitis poses a real and ever-present risk to the dairy industry as it threatens the perception of milk and milk products as a wholesome, ethically produced foodstuff. Taken

together these issues justify the rationale to better understand the epidemiology of this disease and our attempts to improve the treatment, control and prevention of mastitis.

1.2 Improving mastitis control

1.2.1 Historical perspective on the microbiology of mastitis and associated control strategies

In the 1970's the principal mastitis pathogens in the UK dairy herd were *S.agalactiae*, *S.dysgalactiae* and *S.aureus* accounting for 58.1% of diagnoses (Bradley, 2002). All of these pathogens were believed to reside principally or exclusively in the mammary gland or on the skin and mucus membranes of the cow rather than in the environment. In contrast *S.uberis* and *E.coli* only accounted for 9% of diagnoses combined. These two pathogens had been shown to be present in the gastrointestinal tract of ruminants and in the farm environment (housing and pasture) (Anon 1970; Piessens et al. 2011; Pullinger et al. 2006; Zadoks and Fitzpatrick 2009). This was the context in which the first 'best practice' control strategies were developed, as described below.

1.2.2 Structured national mastitis control programs in the United Kingdom and abroad

1.2.2.1 United Kingdom

The first national mastitis control program was introduced in the UK in the 1960's as a voluntary set of best practice guidelines designed to improve mammary health and improve milk quality with statutory milk quality

standards acting as an check on poor herd performance. The “Five Point Plan” was introduced by the National Institute for Research into Dairying (NIRD). The primary aim was to address the high level of chronic infection and contagious transmission by a combination of generic recommendations as described below from the Agriculture and Horticulture Development Board: Dairy (AHDB Dairy) publication 2012.

1. **Hygienic teat management**: which includes good housing management, effective teat preparation and disinfection for good milk hygiene, teat health and disease control.
2. **Prompt identification and treatment of clinical mastitis cases**: including the use of the most appropriate treatment for the symptoms.
3. **Dry cow management and therapy**: where cows are dried off abruptly and teats are cleaned scrupulously before dry cow antibiotics are administered, including the use of infused teat sealants if appropriate.
4. **Culling chronically affected cows**: cows that become impossible to cure and represent a reservoir of infection for the whole herd.
5. **Regular testing and maintenance of the milking machine**: with regular, recommended teat-cup liner replacement and milking machine servicing and attention paid to items which must be checked on a daily, weekly or monthly basis.

The widespread implementation of the Five Point Plan reduced the overall incidence rate of clinical mastitis in the UK dairy industry from 153 quarter cases per 100 cows per year in the 1960's to 40-50 quarter cases per 100 cows per year by the 1980's (Bradley, 2002). This was principally achieved by culling or preventing chronic cases and reducing contagious transmission of the major mastitis pathogens; *S.aureus*, *S.agalactiae* and *S.dysgalactiae*.
In 2004 the UK dairy industry levy board (DairyCo, renamed latterly AHDB Dairy) funded an intervention study to test multifactorial, individualised herd approach to mastitis control management. This was an attempt to identify an evidence based approach to mastitis control that incorporated the improved understanding of the evolving disease in the UK context since the original Five Point Plan was proposed. Fifty two herds were selected and allocated into either a test or control group. The intervention farms received individualised advice based upon their management practices, as assessed by questionnaire and analysis of milk recording data which led to a duel classification of each herd as fitting either a 'Contagious' or 'Environmental' pattern of either 'Lactation Period Origin (LPO)' or 'Dry Period Origin (DPO)'. A significant reduction in the incidence of clinical mastitis and subclinical mastitis of approximately 20% in the intervention herds was recorded compared with the control herds after one year (Green et al. 2007) from a starting point of ≥35 cases per 100 cows per year. The Dairy Mastitis Control Plan (DMCP) was then implemented nationally using veterinary and non-veterinary advisors to deliver individualised advice to dairy farmers. The uptake and implementation of the recommendations proposed as part of the DMCP were subsequently

assessed, identifying a number of recommendations that were infrequently implemented in spite of being deemed important (Down et al. 2016).

1.2.2.2 USA

In the USA the National Mastitis Council (NMC) collaborated with the National Institute for Research into Dairying (NIRD) in the UK to expand the Five-Point Plan into the NMC 10-point plan entitled “NMC Recommended Mastitis Control Program” (http://nmconline.org/docs/NMCchecklistNA.pdf). The five additions to the original Five Point Plan included: establishment of goals for udder health for herd level SCC; maintenance of a clean, dry, comfortable environment; Correct milking procedures; Record-keeping of clinical mastitis cases and maintenance of herd biosecurity. In addition the NMC recommendations advised periodic review of the herd’s mastitis control program.

1.2.2.3 Norway

Norway introduced a mastitis control program in 1982, the goal of the program at its inception was to improve mammary health by reducing the bulk milk somatic cell count (BMSCC) and also reduce the use of antibiotics to reduce the cost of mastitis at the herd level and also maintain consumer confidence in milk and dairy products. In 1996 a decision was taken at governmental level to reduce the use of antibiotics in all Norwegian food producing animal production systems by 25% by 2001. This added impetus to the mastitis control program and contributed to the early adoption of

selective dry cow therapy program (SDCTP) as an a element of the program in 2006 (Østerås and Sølverød 2009). The results of the Norwegian Mastitis Control Program indicated a 60% reduction in the treatment of clinical mastitis between 1994 and 2007, and a reduction in average bulk milk tank somatic cell count from 250,000 cells per ml to 114,000 cells per ml (Østerås and Sølverød 2009). This reduction has been attributed to a range of management changes such as a change in breeding selection, eradication bovine viral diarrhoea virus (BVDV) as well as improved implementation of the mastitis prevention programme (Østerås and Sølverød 2009).

1.2.2.4 Australia

In 1998 Australia introduced a national mastitis control strategy "Countdown Downunder" primarily focused on reducing Bulk Milk Somatic Cell Counts with a target of achieving 90% of bulk milk recordings with less than 250,000 cells per ml (Brightling et al. 2009). This program was based upon advisor led individual herd level advice using an agreed set of control recommendations (Brightling et al. 1998). By 2006, 95% of bulk milk collections were less than 400,000 cells per ml but only 70.9% were lower than 250,000 cells per ml compared to 90% and 64% respectively at the start of the project (Brightling et al. 2009). No data were available for clinical case incidence or prevalence over the same period.

1.2.2.5 Canada

In contrast to the very applied approaches adopted in the USA and Australia the Canadian approach to mastitis control has been focused on research and dissemination of recommendations in specific areas of management (Reyher et al. 2011) and analysis of variations between regions and variation between farming systems within Canada (Riekerink et al. 2008).

1.2.2.6 Netherlands

In 2004 the Dutch dairy industry and Dutch government started a collaborative national mastitis control strategy program. The program involved a combination of vet/farmer/industry best practice meetings alongside broader motivational strategies using non-technical arguments for changing herd management to improve mastitis management. The strategy was developed after a gradual decrease in the bulk milk somatic cell counts recorded from Dutch herds over the previous decades. The strategy did not show a significant decrease in sub-clinical mastitis rates, 23.0 cases per 100 cows per year in 2004 compared to 22.2 cases per 100 cows per year in 2009 (Lam et al. 2013). However, the clinical mastitis incidence rate experienced a small but significant reduction over the course of the program from 33.5 cases per 100 cows per year to 28.1 cases per 100 cows per year.

1.2.3 The ‘Contagious’ vs ‘Environmental’ mastitis classification system

The ‘Contagious’ vs ‘Environmental’ classification system for mastitis patterns within a herd has been the dominant paradigm in mastitis research and clinical veterinary practice since at least the 1960’s (Anon 1970) although no clear reference could be found for the first definition. There is a clear attraction for both farmers and vets to classify the route source of mastitis infections in a given herd according to the species of the principal pathogen cultured from milk samples. This species level approach is quick, simple and inexpensive as it relies upon culture and species level identification from a small number of milk samples. It has also been undoubtedly a useful broad classification system for reducing mastitis incidence in herds with previously poor mastitis control (Wilesmith 1986). However, it is inherently a radical simplification of the biological diversity found in mastitis causing pathogens. When applied too rigidly it can hinder a more nuanced understanding of mastitis epidemiology.

1.2.4 Classification as a ‘Contagious’ or ‘Environmental’ mastitis pattern

A herd will be classified as experiencing either a ‘Contagious’ mastitis problem caused by a species such as, but not exclusively *S.aureus* or *S.agalactiae* which are well adapted to colonising the cow for long periods. Alternatively the mastitis is classified as ‘Environmental’ in origin, due to pathogens which are more readily recovered from the environment such as *E.coli* and *S.uberis*.

1.2.5 Significance of ‘Contagious’ and ‘Environmental’ patterns

The best evidence to support the validity of the ‘Contagious’ vs ‘Environmental’ classification system is by correlation with the implementation of control measures designed to reduce either contagious, cow-cow transmission, such as the Five Point Plan or measures designed to prevent bacterial invasion from the environment such as pre-milking teat disinfection (preMTD). Control of ‘Contagious’ and ‘Environmental’ mastitis patterns is described below.

1.2.5.1 ‘Contagious’ Mastitis Control

In the 1960’s the very high incidence rate of infection in the UK national herd of 153 clinical mastitis cases per 100 cows per year (Wilson & Kingwill 1975) presented a relatively easy baseline from which rapid improvements could be made. In the 1960’s 58% of clinical cases were attributed to pathogens classified as ‘Contagious’; *S.aureus*, *S.agalactiae* and *S.dysgalactiae* whilst only 9% were attributable to ‘Environmental’ pathogens *S.uberis* and *E.coli* (Wilson & Kingwill 1975). The Five Point Plan was introduced to address this problem and focused on measures that were highly likely to reduce the number of chronic mastitis cases and reduce the risk of transmission via the milking equipment. This resulted in a dramatic reduction in the overall mastitis rate to 40 cases per 100 cows per year in 1982 (Wilesmith 1986) and a reduction in the proportion of cases attributed to ‘Contagious’ mastitis from 58% to 30%. However, the proportion attributed to ‘Environmental’ mastitis increased from 9% in 1967 to 48% in 1982 (Wilesmith 1986).

1.2.5.2 'Environmental' mastitis control

The emerging dominance of *E.coli* and *S.uberis* in mastitis diagnoses during the 1970's – 80's led to an emphasis on the environmental hygiene of bedding during housing periods and more recently to an awareness of pasture contamination levels and the advantages of 2 week rotation of pasture (Compton et al. 2007; Cursons and Leigh 2007; Turner et al. 2013). However, in the UK and USA in particular, much attention was focused on pre-milking teat disinfection (preMTD) as a control method for 'Environmental' mastitis. Unlike the Five Point Plan for 'Contagious' mastitis there has not been such clear evidence of successful control measures for 'Environmental' mastitis. Many studies have investigated the efficacy of pre-milking teat disinfection with a range of different active ingredients. Field trials have identified a significant reduction in mastitis associated with the practice (Pankey et al. 1987) and a smaller UK study demonstrated a non-significant reduction in clinical mastitis with preMTD (Blowey RW 1992). However, there are also several large studies which have failed to show any improvement in mastitis with this practice (Hillerton et al. 1993; Shearn et al 1992). In a more recent randomized control trail of preMTD under grazing conditions in Australia where *S.uberis* was the dominant pathogen (accounting for 47% of all clinical mastitis (range 20%-97%)) there was a correlation between a udder cleanliness and the efficacy of preMTD to reduce 'Environmental' mastitis (Morton et al. 2014). In those herds managing cows in clean, dry conditions there was no additional benefit to preMTD in the clinical mastitis incidence.

These studies indicated a stubborn disease challenge that was unresponsive to additional 'Environmental' control measures once gross contamination and cleanliness had been addressed. This relationship between cleanliness and preMTD efficacy was not correlated with the mastitis incidence at herd level (Morton et al. 2014). The results indicate that high incidence rates of supposedly 'Environmental' *S.uberis* clinical mastitis (equivalent of >40 cases per 100 cows per year) were not controlled by either clean, dry conditions or pre-milking teat disinfection (preMTD) for 'Environmental' mastitis control.

1.2.5.3 Comparison of 'Contagious' and 'Environmental' Control measures for S.uberis mastitis

The rationale for pre-milking treat disinfection (preMTD) was that 'Environmentally' associated mastitis caused by *S.uberis* and *E.coli* were not well controlled by the Five Point Plan (Wilson & Kingwill, 1975) and that better management techniques were needed (Hogan and Smith 2003; Todhunter, Smith, and Hogan 1995 and Smith 1985) to reduce the incidence of these two pathogens. However, the assumption that 'Contagious' control measures were entirely ineffective against all 'Environmental' pathogens was incorrect. Wilson & Kingwill, (1975) found an average mastitis incidence over 3 years due to *S.uberis* of 20.8 cases per 100 cows per year whilst (Wilesmith et al 1986) found that by 1982 the average incidence of *S.uberis* mastitis was 7 cases per 100 cows per year suggesting a substantial reduction in *S.uberis* clinical cases alongside large reductions in the classic 'Contagious' pathogens; *S.aureus*, *S.agalactiae* and *S.dysgalactiae* whilst there was no reduction in the

incidence of *E.coli* mastitis over this period.   This suggests that the Five Point Plan was partly effective in controlling *S.uberis* mastitis, indicating responsiveness to classic 'Contagious' transmission control measures. In contrast there was no such reduction in the incidence of mastitis caused by *E.coli* over the same period suggesting that *S.uberis* and *E.coli* are not equivalent 'Environmental' mastitis pathogens.

1.2.5.4 The importance of identifying the mode of transmission in a herd

The results of the previous fifty years developments in mastitis control have highlighted that initial rapid improvements can be made in the incidence of clinical mastitis of either 'Contagious' or 'Environmental' origin when previous management was poor. However, once the simple, generic actions are undertaken (eg. Five Point Plan or basic cow bedding hygiene) then a more resilient, complex, underlying disease pattern is encountered which is more variable between herds and does not conveniently fit the classic 'Contagious' 'Environmental' classification based on pathogen species but has elements of both. This variability between farms requires a greater understanding of the factors that determine disease transmission in order to develop more efficacious interventions at an individual herd or cow level that will allow continued progress in mastitis control, beyond that which was possible with generic recommendations. The risk of using an oversimplified classification system is that limited resources of labour, time and money are expended on control measures that are unlikely to deliver a benefit while more efficacious ones are overlooked.

1.3 The significance of *S.uberis* as a mastitis pathogen

S.uberis has been repeatedly identified as the most commonly isolated pathogen from clinical and sub-clinical samples in several countries including the United Kingdom, Australia, New Zealand and Belgium (Green, et al. 2007; Shum et al. 2009; Verbeke et al. 2014; Zadoks and Fitzpatrick 2009). In the UK

S.uberis mastitis has become more common both in absolute and relative terms in the two decades following the introduction of the Five Point Plan, accounting for 23.5% of all clinical mastitis cases and approximately one third of sub-clinical, high somatic cell count (HSCC) diagnoses (Bradley et al. 2007). In the UK context *S.uberis* may have come to occupy the 'mammary ecological niche' made vacant by the relatively effective control of the other major pathogens through the implementation of the Five Point Plan. It is also possible that changing management practices within the dairy industry over this time period produced conditions more conducive to *S.uberis* such as increasing herd size, changing housing and bedding management and potentially changes in genetic selection. Whatever the underlying causes for this emergence of *S.uberis* as the most prevalent mastitis pathogen it was identified as the single most significant barrier to achieving any further substantial reduction in the incidence of clinical mastitis in commercial UK dairy herds and therefore justifies disproportionate investigative interest (Bradley 2002; Leigh 1999; White et al. 2001).

In order to investigate the epidemiology of *S.uberis* mastitis it is essential that the population structure of the pathogen can be understood both within a herd and also at a national or international level. With sufficiently discriminatory sub-species typing of bacterial isolates from milk samples and other locations, the transmission dynamics of the disease can be investigated, as discussed later.

1.3.1 Response of *S.uberis* to preventive mastitis control measures

It was reported as early as 1993 that the relative prevalence of the major mastitis pathogen diagnoses before and after the widespread adoption of the Five Point Plan did not support the classification of *S.uberis* as an obligate environmental pathogen (Hillerton et al. 1993) as the prevalence of *S.uberis* was reduced to an extent by the implementation of mastitis control measures designed to reduce Contagious rather than Environmentally acquired infections, as discussed in more detail in the earlier (section 1.2.5.3). Specific contagious control interventions have been identified to be efficacious in the control of *S.uberis* in particular; Wesen and Schultz, (1970) described a substantial and comparable reduction of 53% in the new mastitis infection incidence during lactation due to *S.uberis* and *S.aureus* in the trial quarters (RF & RH) of 125 cows undergoing post-milking teat disinfection with an iodine based preparation compared to a control quarters (LF & LH). Zadoks et al, (2001) described a significant reduction in *S.uberis* intramammary infection incidence during periods when post-milking teat disinfection was practiced in a herd experiencing high incidence of *S.uberis* mastitis and also referred to unpublished data indicating that *S.uberis* was recoverable form teat liners for up to cow cycles after an infected (shedding) cow was milked with that teat cluster. Given that the minimum infectious dose of *S.uberis* has been established experimentally to be <1000 cfu/ml (Leigh 1999) and that infected cows will shed $10^6 – 10^7$ cfu/ml it is plausible that the observed reduction in *S.uberis* incidence was causally linked to the introduction of post-

milking teat disinfection. However, the efficacy of post-milking teat disinfection to prevent very small numbers of bacteria being inoculated into unaffected mammary glands may not have be sufficient to entirely control contagious transmission.

1.3.2 Effect of antibiotic treatment on the incidence of *S.uberis* clinical mastitis

Treatment of clinical cases of mastitis is fundamental to the control of contagious transmission as the antibiotic treatment curtails the infectious period of infective cows. In a true 'Environmental' mastitis pattern where each clinical case is acquired only from the environment each cow is a separate unit in a shared infective environment. This means that antibiotic treatment of any individual should not affect the infectious risk or mastitis incidence in other cows in the herd. A case report in 1996 was one of the first incidences where a contagious type mastitis pattern was linked to *S.uberis* (Cattell 1996) and a change in the antibiotic treatment protocol. A dramatic increase in the incidence of *S.uberis* clinical mastitis from 20% of clinical cases to 73% of clinical cases and a simultaneous exponential rise in the bulk milk somatic cell count occurred in the herd which had suspended antibiotic treatment of all clinical mastitis cases (Cattell 1996). The *S.uberis* mastitis outbreak resolved following the reintroduction of clinical mastitis antibiosis. This case report has been cited as an example of a clinical mastitis pattern that did not fit the expected pattern of an opportunistic environmental pathogen (Zadoks et al, 2001). However, as a clinical case report it

represents only a correlation between factors rather than experimentally robust cause and effect.

1.4 Methods of strain identification of *S.*uberis

There are a range of typing methods for bacterial pathogens that differ in their sub-species discriminatory capability and the characteristics of the bacteria used for identification. The major genomic, proteomic, immunogenic and phenotypic typing methods are discussed below along with their usefulness in molecular epidemiological research.

1.4.1 Colony Morphology

S.uberis has traditionally been identified by several techniques including colony morphology small greyish/white colonies which may be surrounded by a greenish zone. A mucoid variant also exists which is more transparent and irregular. *S.uberis* can be cultured on selective media for identification such as blood agar plates containing esculin and viewed under UV light is surrounded by a darker zone. When grown on blood agar plates containing ferric citrate, a black pigment is produced during the fermentation process.

1.4.2 Biochemical characteristics

Biochemical tests can be used for identification either individually or in multiple API 20 panels including catalase and oxidase and Lancefield groupings, *S.uberis* is a non-Lancefield streptococci but enterococcus is Lancefield group D, which allows its differentiation. *S.uberis* is difficult to

distinguish from the closely related species *S.parauberis* by any of the above methods. They have been distinguished by β-D-glucuronidase assay (Khan et al. 2003) and genetic sequencing technologies discussed below.

1.4.3 Ribotyping

Hill and Leigh, 1989 used ribotyping to discriminate between *S.uberis* strain isolates from the U.K. strains using *Hin*dIII endonuclease digestion. The first international attempt to strain type *S.uberis* used of 45 isolates from the UK, USA and Germany, the study suggested that *S.uberis* rDNA fragment patterns were relatively similar in all three countries (Groschup, Hahn, and Timoney 1991). These methods were insufficiently discriminatory and gave rise to the erroneous early impression that the *S.uberis* population was relatively homogenous.

S.uberis and *S.parauberis* were distinguished using ribotyping of the 16S rRNA genes by PCR amplification and subsequently digestion with the *RsaI* and *AvaII* restriction enzymes to yield species specific restriction patterns (Khan et al. 2003). In the same study both species were additionally identified by amplifying species-specific parts of the genes encoding the 16S rRNA, the 23S rRNA and the 16S-23S rDNA intergenic spacer region, respectively. Whilst useful at this higher level, the highly conserved nature of the ribosomal DNA means that 16S typing is not suitable for more discriminatory molecular epidemiology studies where subtle differences between isolates need to be identified to track transmission patterns through a population over a

relatively short period of time (Knox, et al. 1998; Raemy et al. 2013; Shome et al. 2012).

1.4.4 Serological typing

Serological typing techniques have been established for differentiating many bacterial pathogens and have been demonstrated to answer specific hypotheses relating to virulence genes such as *MtuA* in *S.uberis* (Jones et al. 2004). However, immunological typing is limited by its dependence on a small number of antigenic loci and unpredictable reactivity of label antibodies with different antigenic variants. Serological typing also requires prior knowledge of the target antigens in order to construct the test. These limitations restrict the usefulness of such tests for observational studies which seek to identify novel relationships between isolates at the sub-species level.

1.4.5 Pulse Gel Field Electrophoresis (PGFE)

PFGE uses a restriction endonuclease to cleave the genomic DNA (gDNA) at a small number of sites to produce a few fragments with high-molecular-weight. These fragments can then be separated according to size under electrophoresis on gel channels. PFGE has been used to compare isolates of *S.uberis* from milk samples, identifying substantial heterogeneity between strain types (Baseggio et al. 1997; Douglas et al. 2000; Phuektes et al. 2001). However, PFGE demonstrates poor reproducibility between studies (Cookson et al. 2016; Murchan et al. 2003) and cannot be used to quantify the genetic variation between isolates (Maiden et al. 1998) and is therefore of very

limited use for in depth bioinformatic analysis of the molecular epidemiology of a disease such as *S.uberis* mastitis.

1.4.6 Random Amplified Polymorphic DNA (RAPD)

RAPD is a strain typing method which uses PCR amplified DNA fragments. RAPD uses several short primers which combine with larger pieces of genomic DNA allowing sections of the genome to be amplified (Williams et al. 1990). Unlike PFGE it is not necessary to identify target loci prior to typing as the primers bind to any suitable genomic DNA region. However, the reproducibility of RAPD results is highly dependent on individual laboratory protocols and also DNA quality due to the inclusion of a PCR stage. Similar to PFGE methods, RAPD has also been used to assess strain similarity of *S.uberis* isolated from milk samples with similar results, identifying a large variety of strains within a herd (Zadoks, 2003) from both milk samples and milking equipment with some strains found in both locations.

1.4.1 Multi Locus Enzyme Electrophoresis (MLEE)

Multi Locus Enzyme Electrophoresis (MLEE), which differentiates strains based on their different electrophoretic mobility's (EM) of a small number of highly conserved cellular proteins which produces less ambiguous but also less discriminatory results than PGFE (Maiden et al. 1998). MLEE has some similarities with MALDI (discussed in section 1.4.5) in that it utilises transcription products rather than the genetic material itself, however the

limited number of loci used in MLEE limits the amount of information about strain expression diversity that can be assessed.

1.4.2 Multi Locus Sequence Typing (MLST)

Typing techniques such as pulsed-field gel electrophoresis (PFGE), ribotyping, and PCR-based fingerprinting can all be used to assess strain identity between isolates in a binary fashion (Same Strain vs Different Strain). However, these DNA banding-based subtyping techniques do not allow bioinformatic evolutionary analysis (Coffey, et al. 2006) and do not allow the integration of results from one study with a wider database of isolates from other studies. Multi Locus Sequence Typing (MLST) was developed to overcome this problem (Maiden et al. 1998).

Multi Locus Sequence Typing (MLST) uses base pair differences at specific loci in highly conserved housekeeping genes to provide a portable, highly discriminatory, objective and robust method of typing bacteria. Whilst whole genome sequencing yields the most definitive identification of an isolate this is not necessarily the most practical approach for achieving all the aims of a project due to the expense of sequencing several hundred isolates. Genetic techniques such as PGFE and MLST (Douglas et al. 2000; Pullinger et al. 2007; Zadoks, Schukken, and Wiedmann 2005) have allowed us to examine the epidemiologic and evolutionary behaviour of *S.uberis* in unique ways at the sub-species level.

MLST was used for this research for several reasons; the allelic profiles produced can be compared directly against a large, global library of MLST

profiles (www.*pubMLST*.org) from other *S.uberis* studies greatly improving our ability to understand the behaviour of this pathogen on the largest scale. The technique is well established having been used in many similar, smaller studies (Gilchrist et al. 2012; Leigh 1999; Pullinger et al. 2007; Shome et al. 2012; Tomita, et al. 2008). MLST sequencing can also be used as an initial screening tool to identify isolates contributing to disease patterns in clinical case data which warrant deeper investigation by more discriminatory typing methods, such as whole genome sequencing, to provide more detailed bioinformatic information.

For *S.uberis* the MLST scheme uses the following genes: Glucose kinase (gki); Transketolase (recP); D-ala-D-ala ligase (ddl); Thymidine kinase (tdk); Carbamate kinase (arcC); Triosephosphate isomerase (tpi); Acetyl CoA acetyl-transferase (yqiL). Each variant of the gene is assigned an allele number and the isolate is then assigned an allelic profile made up of all the allele numbers for example the reference strain O140J has been assigned the profile 1111111. Allele numbers are assigned as variants are identified; the allele number does not indicate similarity to other alleles of that gene. Clonal complexes are defined somewhat arbitrarily as allelic profiles with at least four of the seven alleles in common with a reference strain eg CC5 uses ST-5 as the reference, or 'founder' strain "Members of the complex were defined as STs that shared at least four alleles with ST-5" (Coffey, et al. 2006). This is a considerably broader clonal complex definition than used by other bacterial nomenclatures such as *S.aureus* which uses a more restrictive definition of

every ST shares at least five of seven identical loci with at least one other ST in the group (Day et al. 2001).

Unlike its predecessor Multi Locus Enzyme Electrophoresis (MLEE), which differentiates strains based on their different electrophoretic mobility's (EM) of multiple core metabolic enzymes, MLST does not use protein products as proxies for the allele profiles. MLST can offer, if appropriate loci are selected, portable, high resolution genomic discriminatory power that is not influenced by post transcription or epigenetic factors unlike PGFE or MLEE for several reasons, all arising from the fact that phenotype diversity is only a proxy for DNA sequence diversity. Firstly, cellular proteins from different strains may have different amino acid sequences without having sufficiently different Electrophoretic modalities (EM) to give distinct bands. Secondly, synonymous mutations can alter the gDNA sequence of an allele whilst still coding for the same amino acid, this would not be detected as a difference by gel techniques but would be detected by MLST. Thirdly, the protein expression phenotype of the cell can be altered in response to environmental conditions which affects the reproducibility of MLEE and PGFE results (Jolley, Feil, and Maiden 2001; Pullinger et al. 2006). In contrast MLST can be easily standardised and analysed as sequence data making it significantly more robust. However, due to the sequence conservation in housekeeping genes, MLST sometimes lacks the discriminatory power to differentiate bacterial strains. The small percentage of genes sequenced for MLST (7 of 1858 in *S.uberis*) can present problems when one or more of the loci are duplicated or deleted. These issues limit the usefulness of MLST in epidemiological investigations.

1.4.2.1 Analysis of MLST data

In bacterial lineages with little or no recombination where evolution has been restricted to clonal expansion by point mutations, *BURST* analysis generates accurate dendrograms depicting the evolutionary descent of the strain types. However, in the case of bacterial species such as *S.uberis* where recombination is common, the software attempts to make vertical linkages between strains where in reality the genetic material was transferred horizontally by recombination (Turner et al. 2007). Cluster analysis of *S.uberis* MLST results using *eBURSTv3* software (Feil et al. 2004; Jolley et al. 2001) is more appropriate given the recombinatorial nature of the population (discussed in section 1.5.2) and it allows useful visualisation of relationships.

1.4.3 Multi Virulence Locus Sequence Typing (MVLST)

To improve the discriminatory power of MLST, a multi-virulence-locus sequence typing (MVLST) approach has been developed using *Listeria monocytogenes* (Zhang et al. 2004). MVLST retains the low cost benefits of a classical MLST scheme but additionally targets one or more virulence genes, which may be more polymorphic and therefore more discriminatory than housekeeping genes. However in order for MVLST to be used for this study we need to know which virulence genes to target. Currently we have several virulence gene targets such as *hasA,* a capsule protein gene, (Coffey, et al. 2006; Pullinger et al. 2006). However, there is no established association between strain type or complex type and disease manifestation or transmission route. All known virulence determinants appear to reside in the

core genome (Hossain et al. 2015) but diversity between virulence loci of isolates within a herd may help elucidate the infection/transmission dynamics even between isolates of the same MLST sequence type.

1.4.4 Core Genome Multi Locus Sequence Typing (cgMLST)

Whole genome sequencing (WGS) using ‘next-generation’ sequencing technology allows investigation of variation at several hundreds or thousands of gene loci rather than the 7 predefined loci utilised by the conventional MLST scheme. Analysis of sequence variation between isolates across a larger number of loci throughout the genome allows greater discrimination between isolates and therefore more appropriate clustering of isolates for molecular epidemiological studies. Several recent studies have used core genome MLST (cgMLST) analysis to elucidate epidemiological relationships in human disease outbreaks involving several bacterial species including; *Mycobacterium tuberculosis* (Kohl et al. 2014)*, Klebsiella* (Weterings et al. 2015) and *Listeria monocytogenes* (Ruppitsch et al. 2015). In this project cgMLST provided the ability to investigate the relatedness of specific isolates which were representative of the most significant disease patterns both within and between herds. Core genome MLST of 1622 target loci was used to validate results of the classical MLST scheme and to understand if the MLST sequence type is the most discriminatory or informative categorisation system of molecular epidemiology research into *S.uberis.*

1.4.5 MALDI Mass Spectrometry Typing

Whilst MLST is a highly discriminatory and reproducible tool for investigating the molecular epidemiology of *S.uberis* isolate strains in this study, MLST is not a suitable tool for clinical microbiology given the time and expense required to produce results. In order to allow strain type data to be integrated into a predictive model for practitioners in new outbreaks a means of cheap, accurate and rapid strain type identification is required. MALDI-ToF (Matrix Assisted Laser Desorption Time of Flight mass spectrometry (MALDI)) is one possible tool which may full fill these requirements.

MALDI was developed initially for rapid, species level, identification of bacteria for medical clinical diagnostics. MALDI has since been used in milk quality laboratories for diagnostics at the species level and also in research, to facilitate bacterial species identification (Bradley et al. 2015; Tomazi et al. 2014). The technique has also been applied for rapid strain typing of other pathogenic organisms including *Listeria* (Barbuddhe et al. 2008), *Yersinia* (Rizzardi, Wahab, and Jernberg 2013), *Staphylococci* (Dubois et al. 2010) and *Corynebacterium* (Goncalves et al. 2014).

The technique involves mixing a whole cell culture or protein lysate extract with a matrix which is then ionised and desorbed with a lazar. The ionised protein particles are then subjected to an electromagnetic field which accelerates the particles across a vacuum to impact on a detector. The time delay correlates to the mass of the particles and the amplitude of the molecular weight peaks reflects the abundance of the molecule in the extract.

This procedure is repeated 40 times per samples to achieve a characteristic 'fingerprint' for the isolate. The position and intensity of the individual peaks are then compared to a reference library file of isolates to identify the best match. The technique produces a characteristic fingerprint of the all charged cellular products (mostly proteins) produced by the bacteria at the time that they are harvested from the culture.

"Direct typing" involves the colonies being spread whole on the target plate and a binding matrix applied prior to laser desorption. This method provides acceptable data quality for species level identification. However, the method generates protein mass spectra with significantly more signal noise especially in the lower molecular weight region. In this research a formic acid extraction protocol was used to improve the quality of the signal generated across the spectrum.

The pattern of protein (and polysaccharide) secretion by *S.uberis*, as with all other genera is determined to a greater or lesser extent by the environment in which it is growing. The physical and metabolic stressors acting on the bacteria is expected to drive expression of particular proteins in response to the environment (Kallow et al. 2006; Valentine et al. 2005). In every spectral profile there will be peaks relating to essential, conserved or 'housekeeping' proteins which are ever-present and peaks relating to 'response proteins' which are variable and specific to the individual isolate genotype and epigenetic regulation.

Analysis of the MALDI spectral profiles can take place at multiple levels from pairwise comparison of isolates at the lowest level to principal component analysis of groups and discriminatory modelling of a large number of isolates in multiple groups based upon characteristic spectral peaks (Ketterlinus et al. 2005). These techniques have been applied to molecular epidemiological studies of *S.pneumoniae* (S Nakano et al. 2015), *B.pseudomallei* (Niyompanich et al. 2014), *Leptospira spp* (Rettinger et al. 2012) and *M.pneumoniae* (Xiao et al. 2014b). The technical capabilities of the models are discussed in Chapter 2.

The purpose of using MALDI in this project was to determine if the MALDI spectral profiles could be used to discriminate between groups of isolates defined by genotype or by case classification as a predictive diagnostic tool for contagious or environmental transmission behaviour in *S.uberis* mastitis cases.

1.5 Epidemiology of *S.uberis* in dairy cattle mastitis.

S.uberis is a coagulase and oxidase negative, gram positive coccus. It is a facultative anaerobe related to several other significant major mastitis pathogens; *S.dysgalactiae*, *S.agalactiae* and *E.faecalis* as well as several other bacteria such as *S.pneumoniae* that do not cause mastitis. It is found commonly in the environment, particularly straw in bedding and in soil on grazing cattle pastures (Lopez-Benavides et al. 2007). *S.uberis* is also a commensal bacteria found at many sites including tonsils, genital tract and skin (Leigh 1999) as well as the gastrointestinal tract of cattle (Döpfer et al.

2008). Traditionally *S.uberis* infections were assumed to have been acquired from the environment rather than from other infected cows because of the ubiquitous presence of the bacteria in the environment. This view has been challenged by a number of studies which relate to different assumptions relating to the expected characteristics of an environmental opportunist pathogen. In the following sections the evidence on the genetic diversity of *S.uberis* and the current evidence on the epidemiology of *S.uberis* will be reviewed with particular reference to those pathogen characteristics that are associated with 'Contagious' vs 'Environmental' transmission.

1.5.1 Global *S.uberis* mastitis population diversity and structure assessed by MLST

Previous MLST studies have demonstrated there are a small number of clonal complexes (CC) which represent a large proportion of the isolates collected from clinical intramammary infections in multiple countries, these are termed global clonal complexes (GCC). Coffey et al, (2006) identified 112 isolates grouped as GCC5 of 160 *S.uberis* isolates. GCC5 is the dominant GCC isolated from European mastitis cases. GCC143 predominates In India (Shome et al. 2012) and Australasia (Tomita, Meehan, Wongkattiya, Malmo, et al. 2008). A small number of identical sequence types have been identified in both continents.

A large proportion of *S.uberis* isolates from clinical mastitis cases previously subjected to MLST in earlier studies were classified as 'singleton' sequence types, aligned to no clonal complex (Coffey, et al. 2006; Tomita, et al. 2008;

Zadoks, et al. 2005). It was not clear from the results of these previous studies whether the high prevalence of singleton sequence types was an accurate description of the *S.uberis* population heterogeneity or an artefact of an insufficient sample size of isolates. The relationship between clonal complexes or singleton sequence types and clinical characteristics, such as transmission route has not previously been carried out; this was a main aim of this research.

1.5.2 *S.uberis* sub-species diversity and evolutionary dynamics

S.uberis is a very genetically diverse pathogen which is partly attributed to a high genomic recombination rate. Recombinatorial parameters were first established using MLST data in *S.pneumoniae* by (Feil et al. 2000) and has since been applied to *S.uberis* by (Zadoks, Schukken, et al. 2005) who were the first to develop an MLST/MVLST scheme for *S.uberis*. In this study six loci were used *cpn60, gapC, soda* and *tuf* were housekeeping genes while *pauA* and *oppF* were regarded as virulence genes. Fifty isolates were compared from mastitis samples from USA (n = 31 isolates from 31 herds) and the Netherlands (n = 20 isolates from 2 herds) which allowed differentiation of 40 sequence types. There were no common sequence types USA and Netherlands. Between herd variation amongst the American samples could not be established as there was only one sample per herd. Between the two Dutch herds there was one common sequence type. The MLST scheme described by (Zadoks, Schukken, et al. 2005) was more discriminatory than ribotyping and provided sub-species data of more epidemiological relevance

by discriminating isolates with identical ribotypes obtained from different herds. Phylogenetic analyses of MLST data in this study indicated that recombination within genes was limited but recombination between genes was high. Indicating that, like several other streptococcal species, *S.uberis* had a recombinatorial population structure.

Coffey et al, (2006) published an MLST scheme using seven housekeeping genes as per the description in section 1.4.1. The study used a sample of 160 mastitis isolates from 6 herds (2 research herds and 4 'Organic' herds). Fifty seven sequence types were identified and one major clonal complex CC5 which represented 24 sequence types and 112 of the 160 isolates. The sample distribution in this study was heavily skewed toward the two research herds which contributed 131 of the 160 isolates. Within those herds two sequence types (ST5 and ST6) were found to predominate and were also identified in other herds. This was the first study to show potential for common, prevalent sequence types of *S.uberis* to dominate the mastitis pattern in a herd in a way that would be more akin to a 'Contagious' pathogen than a purely opportunistic 'Environmental' pathogen. The scheme developed by (Coffey et al, 2006) has been adopted by most subsequent studies as the international format for *S.uberis* MLST with an open access database at (www.pubMLST.org) hosted by Oxford University and curated by Prof Richard Emes , University of Nottingham. The details of the scheme are described in section 1.4.1.

Shome et al, (2012) used this MLST scheme to investigate *S.uberis* diversity in four Indian herds from 209 composite milk samples which yielded 13 *S.uberis* isolates from three sequence types. The dominant clonal complex in this study was CC143. The sequence types were all restricted to just one herd each. This was a very small study and only really serves to highlight the potential for both; substantial geographical variation in dominant clonal complexes and also farm specific *S.uberis* populations.

Pullinger et al, (2006) compared the distribution of MLST diversity in New Zealand and UK isolates using 253 NZ isolates (84 milk, 68 cow & 101 environment origin samples) with the UK study of 160 isolates described above (Coffey et al, 2006). There were no common sequence types between the two countries and NZ samples were dominated by CC143 and CC86 sequence types rather than the CC5 sequence types which were dominant in the UK sample population.

Using the scheme developed by (Coffey, et al. 2006) 1482 *S.uberis* isolates have been previously sequenced by MLST, generating 637 unique sequence types with a mean and median number of isolates per sequence type of 2.33 and 1 respectively. Sequence type ST6 was the most numerous with 91 isolates recorded from all studies combined (www.pubMLST.org & (Davies et al. 2016)). These figures highlight the apparent genetic diversity of the pathogen but also the asymmetry in sequence type prevalence discussed in greater detail in Chapter 3.

1.5.3 Relative prevalence of *S.uberis* sub-species with sample origin

The majority of molecular epidemiology studies of *S.uberis* have concentrated on isolates collected from milk cultures. It is important however, to understand the population structure in all bacterial reservoirs when considering the evidence of infectivity from one reservoir over another. Lopez-Benavides et al, (2007), found that in a study of *S.uberis* isolates from New Zealand the frequency with which *S.uberis* strains were identified from mammary gland milk samples and the environment (bedding, pasture, cow tracks) was not proportional to the overall distribution of *S.uberis* strains as defined by (MLST sequence type) isolated from those sites. Of the 76 isolates cultured by (Lopez-Benavides et al. 2007) 58 were cultured from clinical or sub-clinical mastitis infections while only 18 were cultured from the environment. Of the 38 sequence types isolated in this study only 7 were present in samples collected from the environment. Only one of the sequence types cultured from the environment caused more than a single clinical or sub-clinical mastitis case. (Pullinger et al. 2006) also identified a very diverse population of isolates in both milk and environmental samples both dominated by the CC143 complex (unlike the UK) but only a limited positive correlation at the individual sequence type level.

The populations of *S.uberis* isolates reported by (Lopez-Benavides et al. 2007; Pullinger et al. 2006) were very diverse in both the environment, subclinical and clinical cases; however, the distribution of sequence types found in the cow did not closely reflect that found in the environment. The (Lopez-

Benavides et al. 2007) study also highlighted another important aspect of *S.uberis* ecology by demonstrating a positive correlation between the stocking density, duration on pasture and weather conditions with the penetration of soil and the persistence of viable bacteria with time. The results indicate that the cow is essential for maintenance of the environmental population possibly by replenishment from the gastrointestinal tract.

1.5.4 Evidence of predominant sequence types

Predominant *S.uberis* strains within a *S.uberis* mastitis isolate population in a herd is of significance as strain predominance is more plausibly explained by contagious transmission of a clonal isolate than by coincidental acquisition from the environment. Two molecular epidemiological studies using PFGE and MLST have been described the presence of predominant sequence types in individual herds (Phuektes et al. 2001; Pullinger et al. 2007). (Phuektes et al. 2001) used pulsed-field gel electrophoresis (PFGE) to investigate the epidemiology of *S.uberis* mastitis in four herds. A total of 62 different strains were identified among 138 *S.uberis* isolates. A range of 10 to 26 different strains were found per herd. There were no common strains between the four herds but predominant strains were present in two herds as were persistent infections. Within a herd identical strains of *S.uberis* were isolated from different mammary gland quarters of individual cows and also from different cows within the same herd. The authors suggested that "transmission from quarter to quarter and cow to cow had

occurred"(Phuektes et al. 2001). The limitations of this study were that only four herds were studied which limits the ability to extrapolate the results to a wider population and the PFGE strain typing is difficult to faithfully reproduce between laboratories, limiting the transferability of the strain typing information.

MLST has been used to identify predominant sequence types in three studies (Coffey et al. 2006; Davies et al. 2016; Pullinger et al. 2007) have all identified ST5, ST6 and ST20 as very common and predominant strains in studies of individual and multiple herds. The methodology and results of the largest, multi herd study (Davies et al. 2016) are discussed fully in Chapter 3.

1.5.5 Evidence for persistent *S.uberis* infections

Persistence of an intramammary infection is an essential prerequisite for successful contagious transmission. The number of transmission opportunities via the milking equipment increases for each additional day that an infectious titre of bacteria is present in the milk of a persistently infected, sub-clinical cow. Persistent *S.uberis* intramammary infections have been demonstrated in a number of studies using a range of molecular epidemiological techniques. (Phuektes et al. 2001) found that of 47 pairs of *S.uberis* isolates collected from successive periodic samplings (4-6 weeks apart) during a lactation from a single Australian herd all but 3 pairs were confirmed as being of the same strain indicating a high level of persistence compared to re-infection over this period. In contrast, in the same study only 13 of 64 *S.uberis* infected quarters prior to dry-off were infected with *S.uberis*

in the subsequent lactation and only 1 of these quarters was infected with the same strain of *S.uberis*, indicating a high cure rate of infection during the dry period with some limited re-infection with a different strain of *S.uberis*. A second Australian study (Abureema et al. 2014) used PFGE to establish that post antibiotic treatment 5 of 27 recurrent clinical *S.uberis* mastitis cases were identified as the same PFGE strain as the pre-treatment isolate whereas the remaining isolates were of different PFGE strain indicating a new infection.

MLST has been used to examine the phenomenon of persistent *S.uberis* infections, firstly by (Pullinger et al. 2007) in milk samples collected from one UK research herd. In this study a range of sequence types were identified as causing both short and persistent infections. The study demonstrated persistent infections with the same sequence type in 24 of 33 the mammary gland quarters sampled with duration of infection lasting up to several months. A range of sequence types were identified as able to establish persistent infections with no one sequence type significantly more likely to be persistent. However, in light of the recent finding of this project (Davies et al. 2016) it has been shown that a small subset of sequence types representing only 5% of the sequence type diversity were implicated in 85% of the persistent intramammary infections. Significantly ST6 and ST20 were frequently identified in both studies (Davies et al. 2016; Pullinger et al. 2007) as persistent sequence types and frequently associated with potential contagious transmission.

1.5.6 Transmission dynamics of *S.uberis* intramammary infections

As discussed previously the principal dynamic characteristic of any contagious infectious disease, including bovine mastitis is the increased risk of disease experienced by animals within a population according to the number of infective individuals within the same population. This is not the case with environmentally acquired infections where the size or infectivity of the reservoir is independent of number of infected individuals within the population. In mastitis these dynamic relationships were explored by (R N Zadoks, et al. 2001) in a longitudinal study of one herd. The study calculated the transmission parameters, identifying that the number of infected mammary gland quarter-days in the preceding three week period was a significant predictor of the likelihood of new intramammary infections. Implying that exposure to a high number of shedders increases the risk of new clinical mastitis cases. The weakness of this study was that sub-species identification of the *S.uberis* isolates was not carried out to determine if successive clinical cases were related or unrelated. The significance of this question is that if the environmental conditions which predispose to environmentally acquired conditions are not constant then you might also expect variation in infectious risk. However, isolates from an environmental reservoir would be expected to be more diverse than those acquired form contagious transmission given that the majority (70%) of *S.uberis* intramammary infections are mono-clonal (Pryor, Cursons, and Williamson 2009) whereas the environment is inherently multi-clonal. A more recent

study compared the transmission dynamics of *S.uberis* and *S.agalactiae* (Leelahapongsathon et al. 2015) by the basic reproduction number (Rø) metric commonly used in epidemiology of contagious disease to indicate the ease of transmission of a disease in a susceptible population (Diekmann, Heesterbeek 1990), an Rø value equal to one means each infected animal, on average, transmits infection to one other animal (a stable, endemic state). Whereas a value greater than one indicates an expanding epidemic and a value less than one indicates a pattern of transmission that will lead to extinction of disease. The metric depends on the infectivity of the pathogen and the duration of infection as well as the size of susceptible population and the exposure to that susceptible population. Leelahapongsathon et al. (2015) conducted a 10 month longitudinal study of two, small herds with poor mastitis management in which the basic reproduction ratio and duration of infection were calculated. Rø values (95% confidence intervals) of 2.91 (0.63–13.47) and 1.86 (0.21–16.61) were calculated for *S.uberis* and *S.agalactiae* intramammary infections respectively indicating that in these two herds infection dynamics between these two pathogens traditionally attributed to different transmission routes were similar. However, the *S.uberis* intramammary infections were of significantly shorter duration in this study, 188 days compared to 271 days for *S.agalactiae.* The shorter duration of infection may indicate a different transmission mechanism or attribute was potentially compensating for the shorter infection duration of *S.uberis, i.e*. environmentally acquired infections. The potential of combination of different infection sources within a herd (Infection from another cow or from

the environment) represents a particular challenge to Rø metric calculations because different model structures need to be considered when free-living pathogens are involved (Bani-Yaghoub et al. 2012). The comparison of *S.uberis* and *S.agalactiae* is relevant as the role of *S.agalactiae* as an obligate 'Contagious' pathogen has been questioned by a recent Norwegian study (Jørgensen et al. 2015) which identified *S.agalactiae* in the environment and gastrointestinal tracts of cows in 10/15 (66%) of herds with diagnosed *S.agalactiae* intramammary infections. (Jørgensen et al. 2015) is the first study to clearly identify a potential 'Environmental' mastitis pattern for *S.agalactiae*, in herds which good contagious control. Whilst this may or may not be disease manifestation that is particular to Norwegian conditions it demonstrates the diverse range of transmission patterns which can be demonstrated in *Streptococci spp* and highlights the oversimplification of a 'Contagious' vs 'Environmental' mastitis pattern classification system based on bacterial species.

1.5.7 Evidence of adaptation to the cow mammary environment and determinants of bacterial virulence

Prolonged high somatic cell count (HSCC) and persistent infections were not traditionally regarded as a feature of *S.uberis* mastitis (Bradley 2002; Green 2012). However, there are now many reports of mastitis disease patterns on farms which display these characteristics as previously discussed (Section 1.5.5) and *S.uberis* isolates have also been shown to persist within the mammary gland for up to 260 days in commercial dairy herds (Zadoks , et al,

2009; Oviedo-Boyso et al., 2007; Zadoks et al., 2001). The ability to establish infections of that resist immunological clearance are exhibiting adaptations to the host immune defences and are hence 'cow-adapted'. (Ward et al. 2009) demonstrated through genomic analysis of one *S.uberis* isolate (*O140J*) the ability of the pathogen to use diverse metabolic pathway options to successfully occupy a 'discrete ecological niche' in the environment and also in the cow. This flexibility may also underpin the evidence of antibiotic resistance amongst some *S.uberis* isolates in a study by (Rato et al. 2013) which demonstrated higher levels of antibiotic resistance to erythromycin, pirlimycin, cefoperazone, tetracycline, gentamicin and streptomycin amongst the *S.uberis* isolates than any of the other streptococcal spp. In the same study farm specific patterns of resistance were observed within the *S.uberis* isolates across the 11 herds, indicating adaption to the particular selection pressures in each herd dictated, in part, by their antibiotic treatment management.

A number of virulence mechanism have been identified for *S.uberis*, principally resisting phagocytosis, establishing protective biofilms and invading epithelial cells to establish more persistent infections (Bradley 2002; Crowley et al. 2011; Günther et al. 2016; Leigh 1999). To operate these virulence mechanisms specific virulence genes, coding for functional proteins essential for the pathogenesis are required. Several virulence genes have been suggested previously including *pauA* and *pauB* (Gilchrist et al. 2012; Khan et al. 2003; Raemy et al. 2013; Zadoks et al. 2005) which have shown that they are present with different prevalence in different species (Gilchrist

et al. 2013) and in different sequence types. The capsule forming gene *hasA* has also been implicated as a virulence gene (Coffey et al. 2006; Pullinger et al. 2006; Tomita, Meehan, Wongkattiya, Pullinger, et al. 2008; Zadoks and Fitzpatrick 2009) as a means of blocking neutrophil action however, these studies showed that this gene was not present in every mastitis isolate and neither *pauA/B* or *hasA* have been shown to correlate with isolates demonstrating contagious rather than environmental behaviour. Significantly different inflammatory reactions to *S.uberis* sequence types ST439 and ST475 which carry different combinations of *cfu, S*kc and *sua* virulence genes has been recently demonstrated in mouse mammary models using bovine mastitis isolates (Mitra et al. 2016). Indicating that the host immune response and the likelihood of persistence is influenced by the virulence attributes of the pathogen at the sub-species level. Another group of sortase anchored cell wall proteins were identified that in mutant knockout experiments reduced or prevented virulence (Crowley et al. 2011; Leigh et al. 2010). Two of these proteins (SUB 0145/RS00865 and SUB 1095/RS5505) possessed a hypervariable region. Hypervariable regions may allow immune escape of bacterium by presenting an alternative antigenic profile to the adaptive immune response and this could be an important mechanism for establishment of persistent infections (discussed further in Chapter 6).

It is unclear at present if different virulence mechanisms are employed by *S.uberis* sub-species that are able to transmit contagiously from quarter to quarter compared to from the environment to the quarter? Or is contagious transmission a consequence of particular cow factors that influence

susceptibility to the low infection from any available *S.uberis* that may be present on vectors in the parlour (liners, gloves, hands etc) following the milking of an infected cow?

1.6 Cow factors and susceptibility to 'Contagious' or 'Environmental' mastitis

There are many cow factors such as age, parity, breed and milk constituent parameters which have been suggested as predictive indicators of mammary gland infection risk. Cow level factors such as parity and stage of lactation have previously been identified as important covariates with risk of clinical mastitis in; UK herds (Breen, et al 2009); worldwide (Oliveira et al. 2015; R N Zadoks, Allore, Barkema, Sampimon, Wellenberg, et al. 2001); and specifically in respect of *S.uberis* infections, (R.N. Zadoks et al. 2001) identified in a study of 93 new *S.uberis* clinical cases that parity and days in milk (DIM) were associated with risk of mastitis as was a history of a pervious mastitis clinical case however, SCC did not correlate with increased risk in the multivariate analysis for cow or quarter factors. This study did not consider sub-species difference between the 93 isolates, considering the risk of any *S.uberis* case as the outcome. In contrast (Pullinger et al. 2007) examined the MLST sequence types of 33 persistent infections of 50-260 days duration and 17 short duration infections that were present for <30days. This study found no significant difference in the MLST sequence types that caused these two groups of cases and concluded on that basis that cow factors were more important than bacterial specialisation in the establishment of persistent

infections. In light of the recent finding of a far larger multi-farm study (Davies et al. 2016)(Chapter 3) identified a strong bias toward a small subset of MLST sequence types being responsible for the majority (85%) of persistent infections.

1.6.1 Predictive value of SCC and other milk recording data for *S.uberis* mastitis

There has been a great deal of interest over many years in using milk constituents information in particular as a predictive or diagnostic aid in the control of mastitis (Jones TO 1986) as well as for other management purposes such as nutrition and fertility management (Dronawat et al. 1966).

Milk constituent records are a potentially attractive resource because the regular, routine, individual measurement of milk sample parameters is a common commercial practice in the UK dairy industry. Approximately 60% of cows in the UK are sampled and recorded on a monthly basis (NMR website). The recorded parameters are: volume of milk (yield); somatic cell concentration (somatic cell count SCC); milk fat percentage; milk protein percentage and milk lactose percentage. The sampling protocol typically involves combining a pooled milk sample from all lactating mammary glands at that milking event. Additional cow level parameters are also recorded to accompany the monthly milk sampling data: parity, days since calving (days in milk DIM). This data provides a rich source of easily available information. Changes in one or more of the recorded parameters may correlate with intramammary infections in several ways. These relationships may then be

used to predict the nature of a new infection based on variables relating to that individual cow's previous recordings or the pattern of variables within the herd as a whole that may indicate a change in the risk of infection.

1.6.1.1 Milk somatic cell count (SCC)

Somatic cell count is used as an indicator of mammary gland health status with a measurement 200,000 cells/ml widely accepted as a threshold between 'uninfected' and 'infected' glands. Other factors such stage of lactation, heat stress and stage of reproductive cycle have been shown to have only a small effect on SCC (Harmon 1994). Infection and colonisation of the mammary gland by microorganisms, principally bacteria, are the most common explanation for elevated SCC measurements. The pattern of the SCC response to intramammary infection has been used to assist in the diagnosis of clinical mastitis (Green et al. 2004) and predict the causal pathogens (Green et al. 2004; de Haas et al. 2004; Haas et al. 2005).

1.6.1.2 Milk fat, protein and lactose as mastitis predictors

Milk fat and protein concentration and fat : protein ratio have been investigated as predictors for clinical mastitis from herd records (Zoche-Golob, Heuwieser, and Krömker 2015) and from a prospective inoculation study using *S.uberis* which included analysis of milk lactose concentration (Kester, Sorter, and Hogan 2015). The studies revealed reduced lactose concentration post *S.uberis* infection and a correlation with Fat : Protein ratios. Results between these studies were not entirely reproducible and indicate that under certain circumstances useful information may be derived

from these measures and therefore justified further analysis to identify if differences occurred according to the sub-species of *S.uberis* or the potential transmission route.

1.7 Summary of objectives for this research

Currently we know that *S.uberis* is the most significant single mastitis pathogen and that it is genetically very diverse. There is evidence from genotyping studies and other sources suggesting a diverse clinical disease pattern as well with both 'Contagious' and 'Environmental' transmission. We do not know how common 'Contagious' mastitis transmission is or how important it is as a proportion of the mastitis cases within a herd and we do not know what virulence mechanisms are used by *S.uberis* to achieve it.

The primary aim of this research project was to estimate the prevalence of potential contagious transmission at herd level and the significance of contagious transmission within a sample of 52 commercial UK dairy herds using multilocus sequence typing (Chapter3).

The next aim of the project was to use readily available herd and cow level data alongside the clinical case and MLST data to identity parameters that might be used to predict the transmission dynamics of *S.uberis* infection (Chapter 4).

Using the MLST results to categorise the *S.uberis* isolates the next aim was to investigate the discriminatory capacity of MALDI spectral profiles firstly as a typing tool for sub-species identification and then as a potential tool for rapid

identification of mastitis isolates likely to be the result of contagious transmission. The predictive capacity of the MALDI spectral profiles for 'Contagious' vs 'Environmental' isolates was then tested for isolates from different farms to build and validate statistical models (Chapter 5).

The final aim was to investigate in greater depth the genetic heterogeneity of potentially contagious *S.uberis* isolates at both the herd level and the sequence type level using whole genome sequencing (Chapter 6). As well as investigating genetic diversity in the whole genome of these potentially contagious isolates, pairwise comparisons were made of selected virulence genes from isolates originating from the same persistent intramammary infection to explore evidence of immune evasion mechanisms (Chapter 6).

Chapter 2: Materials and Methods

2.1 Study population and clinical mastitis samples

The study utilised isolates of *S.uberis* from milk samples collected during a previous, large scale study (Green et al. 2007) of UK dairy herds as part of the dairy industry mastitis initiative (DairyCo Mastitis Control Plan). Fifty two herds were recruited using a database administered by National Milk Records (NMR), Chippenham UK. Two hundred and fifty herds with a record of more than 35 cases of clinical mastitis per 100 cows during the previous 12 months (2003-4) were selected. A recruitment letter was sent to these farmers inviting them to participate in the study and 68 responded positively. Monthly recordings of somatic cell counts (SCCs) from these herds were assessed, 26 herds with an annual arithmetic mean SCC of more than 200,000 cells/ml ('High' group) and 26 with an annual arithmetic mean SCC of less than 200,000 cells/ml ('Low' group) were selected. A final selection was made from farms located in one of three regions of England and Wales; a region to the north of a line from the Severn estuary to the Wash, and two regions to the south of this line divided to the east and west by a line joining Oxford and Portsmouth. Within each region herds were paired according to their annual mean SCC (High and Low). Average herd size over the study period ranged from 55 to 320 cows with mean of 164 and median 163 cows per herd. The median incidence of reported clinical mastitis during the original study period (March 2004 – June 2005) in the 52 recruited herds was 66 (mean 75) cases per one hundred cows per 365 days; with a range from 16 to 146. The mean

percentage of clinical cases diagnosed as *S.uberis* was 28% during the study period with a range in individual herds from 7% to 64%. All herds were housed during the winter period. Year round and seasonal block calving herds were represented within the sample group. For the purposes of the current study this farm selection and categorisation provided a reasonable cross section of commercial dairy herds in England and Wales.

2.2 Milk sample collection and initial pathogen identification

Milk samples were requested from all clinical mastitis cases that occurred in each herd during the study period. Clinical mastitis diagnosis and sample collection was performed in each herd by farm employees according to a standardised operating procedure (Green et al. 2004). Pathogens were identified and classified according to the standard bacterial classification technique (Bradley et al. 2007) by Quality Milk Management Services Ltd (QMMS), a commercial milk laboratory. More specifically *S.uberis* were identified based on colony morphology on blood and Edwards agar after 72 hours incubation at 37°C as well as gram staining, catalase and Lancefield grouping; where necessary further biochemical identification was performed using the API-20 STREP system (Biomerieux, Basingtoke, UK).

2.3 *S.uberis* isolate archive storage

All isolates were stored using a bead based micro-preservation system (Protect, Technical Service Consultants; Heywood, UK) and stored at -80C. Each isolate was 'subbed' during the identification process onto an

individual blood agar plate and subsequently two to three colonies (from a pure plate) were selected for preservation.

2.4 *S.uberis* recovery protocol

Samples of each isolate were cultured on brain heart infusion (BHI) agar medium and incubated for 20-36hrs at 38°C according to the abundance of colonies identified. Initial confirmation of classification was carried out by assessment of colony morphology. Any recovered colonies subjectively judged to display atypical morphology, such as mucoid colonies, small colony size or irregular colony margins were subjected to further confirmatory tests using both API biochemical strips (32 STREP; BioMérieux Canada, Inc., Saint-Laurent, Quebec, Canada) and MALDI-ToF mass spectrometry (Bruker Biotyper) using an ethanol/formic acid extraction protocol described in section 2.8.2 (Raemy et al. 2013).

Isolates which generated only light or nil growths under the primary recovery protocol underwent a second recovery protocol where 10ml aliquots of BHI broth were inoculated with a sample of each isolate and incubated for 24 hours at 38°C. Samples of enriched incubated broth culture were then used to inoculate BHI agar plates which were then incubated for 24hrs at 38°C and underwent the same confirmatory tests as other isolates.

2.5 gDNA extraction and quantification

In an attempt to find an extraction protocol which was both quick and efficient as possible for the large number of isolates to be processed a

number of possibilities were considered and tested as follows. Proprietary kits were ruled out on the basis of prior unsatisfactory and inconsistent results with *S.uberis* (Leigh, personal communication, Cogan personal communication) and their relatively high cost. Boilate methods were attempted as below based on the protocol described by Tomita et al (2008):

1) Sweep of colonies transferred to 0.5ml PBS and incubated with 0.5ml Lysozyme (20mg/ml) and incubated at 37°C for 15 min them boiled at 100°C for 20 min. Cell wall debris was then separated by centrifuging at 4500 rpm for 10 min. gDNA concentration was quantified using 'Qubit-HR' fluorescence dye binding system specific to double stranded DNA.

- Very low gDNA yields were obtained by this method ranging from 0.008-0.027ng/ul. (Target >20ng/ul for MLST)

- Examination of the lysate under high power microscopy revealed inadequate lysis of a high proportion of the bacteria.

2) To improve the rate of lysis, incubation times were extended to 60min at 37°C with lysozyme followed by 60min at 100°C with 50ul SDS (20% w/v).

- Adequate lysis was achieved with this protocol as accessed microscopically

- Yields of gDNA with this, modified protocol ranged from 0.01-0.039 ng/ul

3) To prevent the possibility degradation of gDNA by DNase enzymes extractions were performed according to method (2) with the addition of EDTA and Depsipeptide with SDS to denature DNase's.

- This additional step did not increase yields of gDNA.

Following these unsuccessful attempts a series of trial extractions by the alternative technique as described by (Coffey et al 2006) using ProteinkinaseK and Phenol/Chloroform to purify the gDNA product was carried out. gDNA yield was quantifying using a 'Nanodrop 8000' mass spectrometer to assess absorption ratios correlated to DNA concentration and protein contamination.

2.5.1 Method for Purification of chromosomal DNA from *S.uberis*

(Coffey et al 2006, pubMLST website reproduced verbatim)

1. Grow *S.uberis* strains at 37°C for 16-20h in "Universals" containing 10 ml of liquid media (Todd Hewitt Broth or Brain Heart Infusion).

2. Prepare chromosomal DNA by harvesting bacterial cells from multiples of 1.5 ml of culture by centrifugation (13000 x g for 5 min at room temperature). Typically 3ml of bacterial culture yields 50-100µg DNA.

3. Re-suspend cell pellet by vortexing with 0.5 ml of TE buffer[1] [10mM Tris, 5 mM EDTA (pH 7.8)]. Re-pellet cells (13000 x g for 5 min

at room temperature), remove supernatant and re-suspend in 375µl of fresh cell wall disruption buffer[2].

4. Incubate at 37°C for 30 min.

5. Lyse cells by addition of 20µl Lysis buffer[3] and 3µl of Proteinase K[4] - shake briefly to mix (the sample should clear). Incubate at 37°C for 1h.

6. Add 200µl saturated NaCl (approx. 6.0M) to precipitate protein cell wall material and agitate for 15s. Centrifuge (13,000 x g) for 10 min to obtain a firm pellet.

7. Remove 400-450µl of supernatant carefully to a fresh, labelled tube. Re-pellet if strings of precipitated material are lifted, and discard tube containing pellet. Add an equal volume of Tris-equilibrated phenol:chloroform:isoamyl alcohol (25:24:1 molecular biology grade, Sigma).

8. Agitate to mix and centrifuge (13,000 x g at RT) for 3 min to separate phases.

9. Retain 300-450µl of the upper aqueous phase avoiding the interface (leave some if in doubt). Precipitate DNA by adding 2 volumes of cold ethanol and hold at 4°C for up to 2h.

10. Precipitate DNA by centrifugation at 13,000 x g for 5 min. Wash with cold 70% ethanol (centrifuge again at 13,000 x g for 5 min, and discard ethanol). Thoroughly air-dry the pellet.

11. Allow to re-suspend slowly (~1/2h) at 4°C in 30-50µl of TE buffer (10mM Tris; 1mM EDTA pH 8.0) containing RNAase A (made DNAase-free by boiling for 5 min) at 20µg/ml. Do not shear by pipetting or vortexing. Incubate for 15-30 min at 37°C.

12. Determine DNA concentration by assaying the absorbance of a dilution (typically 1/50 in water) of the preparation at 260nm. 1.0 OD_{260} corresponds to 50µg/ml of chromosomal DNA.

13. Store DNA at -20°C (but try to avoid repeated freeze thaw).

2.5.2 Preparation of Reagents

TE buffer: for 10 ml, mix 0.1ml of 1M Tris-HCl pH 7.5 with 0.1ml of 0.5M EDTA pH 8.0 and 9.8ml of sterile molecular biology grade water.

Cell wall disruption buffer:[30u/ml mutanolysin; 10mg/ml lysozyme in TE buffer]. For 10 ml, add 100mg desiccated lysozyme and 60µl of stock mutanolysin reagent (stock prepared in sterile water at 5000 units per ml) to 10ml TE buffer (prepared as above). Mix, store on ice until used.

Lysis buffer:SDS (20% w/v) in 50 mM Tris, 20 mM EDTA pH 7.8. For 100ml, dissolve 20g SDS in 80ml molecular biology grade water, add

5ml 1M Tris-HCl (pH 7.5), 4ml 0.5M EDTA (pH 8.0), make up to 100ml with water, mix & store buffer at ambient temperature.

Proteinase K:[stock 20mg/ml]. For 5ml, dissolve 0.1g in 5ml of sterile molecular biology grade water, aliquot and store frozen at < -20°C

The gDNA yield obtained by this method was much improved on the previous techniques, ranging from 6ng/ul- 13500ng/ul. However the length and complexity of the protocol reduces the number of isolates that could be extracted to 30-40 per day. In addition approximately 10% of isolates extracted by this method failed to reach the 20ng/ul threshold for successful MLST sequencing.

The protocol was subsequently adapted to further increase the yield of gDNA by increasing the number of bacteria subjected to the extraction method and combined some steps and adjusted incubation times to maximise the lysis of the bacteria. A further modification to the technique replacing the phenol/chloroform extraction with a spin column as described be (Gilchrist et al. 2013) was planned in order to reduce the number of individual steps and increase the efficiency of the protocol but was not finally conducted due to lack of availability of materials. The final protocol is described below:

2.5.3 Final gDNA extraction Protocol

1. Incubation at 37°C for 16-20h in 10 ml of liquid Brain Heart Infusion media.

2. Prepare chromosomal DNA by harvesting bacterial cells by centrifugation whole cylinder at 4500 x g for 5 min at room temperature.

3. Re-suspend cell pellet by vortexing with 0.5 ml of TE buffer[1] [10mM Tris, 5 mM EDTA (pH 7.8)]. Re-pellet cells (13000 x g for 5 min at room temperature), remove supernatant and re-suspend in 375µl of fresh cell wall disruption buffer[2]. Lyse cells by addition of 20µl Lysis buffer[3] and 3µl of Proteinase K[4] - shake briefly to mix (the sample should clear). Incubate at 37°C for 2h.

4. Add 200µl saturated NaCl (approx. 6.0M) to precipitate protein cell wall material and agitated for 15s. Centrifuge (13,000 x g) for 10 min to obtain a firm pellet.

5. Remove 400-450µl of supernatant carefully to a fresh, labelled tube. Re-pellet if strings of precipitated material are lifted, and discard tube containing pellet. Add an equal volume of Tris-equilibrated phenol:chloroform:isoamyl alcohol (25:24:1 molecular biology grade, Sigma).

6. Agitate to mix and centrifuge (13,000 x g at RT) for 3 min to separate phases.

7. Retain 300-450µl of the <u>upper aqueous phase</u> avoiding the interface (leave some if in doubt). Precipitate DNA by adding 2 volumes of cold ethanol and hold at 4°C for up to 2h.

8. Precipitate DNA by centrifugation at 13,000 x g for 5 min. Wash with cold 70% ethanol (centrifuge again at 13,000 x g for 5 min, and discard ethanol). Thoroughly air-dry the pellet.

9. Allow to re-suspend slowly (~1/2h) at 4°C in 30-50µl of TE buffer (10mM Tris; 1mM EDTA pH 8.0) containing RNAase A (made DNAase-free by boiling for 5 min) at 20µg/ml. Incubate for 15-30 min at 37°C.

Each sample was cultured and extracted using the above protocol. The gDNA concentration was quantified using Nanodrop 8000 to provide a minimum of 5ug of high molecular weight gDNA at a concentration greater than 50ng/ul.

2.6 Multi Locus Sequence Typing

Multi Locus Sequence Typing was performed by University of Oxford, Department of Zoology using the internationally recognised MLST scheme (Coffey, et al. 2006) using the following highly conserved genes; arcC, ddl, gki, recP, tdk, tpi and yqiL (Ward et al. 2009). The methods described here are a more detailed version of those detailed in the resulting paper (Davies et al. 2016). Initial PCR amplifications were carried out using standard 96 well format with a positive and negative control using the following reagents per 96 well plate: H_2O-2975.00µl; Qsolution-1000.00µl; 10xBuffer-500.00µl; Forward primer - 100.00µl (10µm stock); Reverse primer - 100.00µl (10µm stock); dNTP-100.00µl; TAQ-25.00µl. This protocol was successful for loci arcC, ddl, yqil, recP and gki, generating sufficient quantities of DNA but insufficient

for the tdk and tpi loci. Since these two genes have longer PCR products, PCR quantities were adjusted as follows: H2O-2815.00µl; Q soln-1100.00µl; 10xBuffer-500.00µl; Forward primer - 120.00µl; reverse primer -120.00µl; dNTP-120.00µl; TAQ- 25.00µl., The adjusted quantities improved the results for the tpi locus but not the tdk. Therefore tdk PCR quantities were used at a reduced annealing temperature from 72°C to 54°C. This improved the yield sufficiently for Sanger sequencing.

2.6.1 MLST Sequence Analysis

The MLST scheme was selected because of the large number of *S.uberis* isolates already sequenced and the readily accessible online format which facilitates analysis of these data in the broadest, international context. Analysis of sequence traces was carried out using *Ridom SeqSphere* for automatic and manual assignment of sequence traces to alleles. Sequence profiles were catalogued for inclusion in the pubMLST database (pubMLST.org). Phylogenetic and cluster analysis was performed using eBURST version3 (Feil et al. 2004) statistical analysis of MLST data was conducted in *MiniTab16 & 17* (Minitab 17 Inc 2015) using Fisher's exact Chi-squared tests.

2.7 Whole Genome Sequencing

2.7.1 Illumina sequencing carried out by DeepSeq, University of Nottingham

Whole genome sequencing was conducted on 69 selected isolates following analysis of MLST sequencing data discussed in Chapter 3. The criteria for sample selection and the results of whole genome sequencing analysis are discussed in Chapter 6. Samples were recovered and gDNA extracted according to the protocols described previously (2.5). Sixty eight samples were processed via TruSeq PCR free libraries and were sequenced using one fifth of an Illumine NextSeq500 high output run with the aim of providing an average of 60x coverage for each sample from 2 x 150bp reads. This sequencing format was selected as a compromise between assembly integrity vs number of isolates that could be sequenced for a limited budget to provide the most useful data for analysis.

2.7.2 Library preparation and sequencing

Library preparation and sequencing of each strain was conducted by DeepSeq, Queens Medical Centre, University of Nottingham as follows. gDNA samples were incubated with RNAse A at a final concentration of 100ng/µl at 37°C for 30mins to remove contaminating RNA. Samples were purified with equal volume of AMPure XP beads (Beckman Coulter, A63881). Purified gDNA was quantified using Quant-it dsDNA assay kit, high sensitivity (ThermoFisher Scientific, Q-33120). Samples were diluted to 0.2ng/µl with nuclease free water. Sequencing libraries were prepared using Nextera XT library

preparation kit (Illumina, FC-131-1096, FC-131-1002) according to the Nextera XT library preparation manual. Library QC was performed using Agilent Bioanalyser HS kit (Agilent biotechnologies, 5067-4626). Sequencing library concentrations were measured using Quant-it dsDNA HS assay kit. Equimolar library pool was created and the library pool concentration was measured with qPCR (Kapa Biosystems, KK4824). Sequencing was done on Illumina NextSeq500 sequencing platform using a MID output 300bp V2 chemistry according to manufactures' guidelines (Illumina, FC-404-2003).

2.7.3 Genome Assembly

Bacterial genome assembly and annotation was conducted in parallel by two de novo methods for comparison and validation. The first assembly method used the *Ridom Seqsphere Pipeline* function to assemble sequence reads from sample Fastq files using internal *Velvet* assembler with trimming of reads at 5' and 3' ends to remove adaptors and until average quality was ≥30 in a window of 20 bases resulting in average good quality read length of 98 bases. In parallel, bioinformatic services were provided by Andrew Warry (University of Nottingham - Advanced Data Analysis Centre (ADAC)) to assemble the sample genomes by an alternative method that did not use *Ridom Seqsphere* for cross validation. The assembly process used was as follows: 1) Fastq sequence files were adapter/ end trimmed to remove low quality data with *Cutadapt* and *Sickle*. 2) Overlapping read pairs were merged using *Flash*. 3) De novo assembly with merged and unmerged reads was carried out using

SPAdes. 4) Scaffolds were filtered to produce Fasta sequence files from scaffolds with >500 bp length and >5 reads coverage.

Core genome MLST (cgMLST) analysis was performed in *Ridom SeqShpere+* using the local nomenclature ad hoc multiple outbreak approach. The reference used was strain NC_012004.1 0140J from the NCBI database. This strain was selected on the basis that it is the only complete genome available and the common reference strain used in previous studies (Davies et al. 2016; Eigh et al. 2010; Hossain et al. 2015; Pullinger et al. 2007).

2.7.4 cgMLST Target quality control parameters

Gene targets from all inputted sequence isolate FASTA files were categorised as either 'cgMLST targets' used for analysis or 'Accessory Targets' not used of analysis. Genes were designated as accessory targets if they met any of the following criteria:

- They are not found in each of the query genomes
- They are found more than once in at least one query genome
- They overlap in the reference genome
- They do not contain correct number of stop codons in more than 80% of The query genomes

(Ridom SeqSphere+ manual)

A second set of filters was then applied to both the reference genome and also to the sample of query genomes. Any gene which failed was either

discarded or moved to Accessory targets. The following filter definitions were used (reproduced from the Ridom SeqSphere manual):

2.7.4.1 Filters for reference genome: these filters are only applied to the reference genome.

- Minimum Length Filter: Discarded genes that are shorter than 50 bases.

- Start Codon Filter: Discarded all genes that contain no start codon at the beginning of the gene.

- Stop Codon Filter: Discarded all genes that contain no stop codon, more than 1 stop codon or if the stop codon is not at the end of the gene.

- Homologous Gene Filter: Discarded all genes that have fragments that occur in multiple copies in a genome (with identity >= 90% and more than 100 bases overlap).

- Gene Overlap Filter: If two genes from the reference genome overlap more than 4 bases, move the shorter gene to the Accessory targets.

2.7.4.2 Filters for query genomes: these filters are applied to the genes found by BLAST in each query genome.

- Start Codon Filter: Moves all genes to Accessory targets that contain no start codon at the beginning of the gene in at least one query genome.

- Stop Codon Filter: Moves all genes to Accessory targets that contain no stop codon, more than 1 stop codon or if the stop codon is not at the end of the gene in at least one query genome.

- Stop Codon Percentage Filter: Move all genes to Accessory targets that fulfil the following condition in more than 80% of the query genomes:

- The gene contains no stop codon, more than 1 stop codon or the stop codon is not at the end of the gene.

Finally gene targets from individual study isolates were discarded from their category as either cgMLST Targets or Accessory Targets if any of the following conditions were met:

- The length of the consensus sequence was not equal to that of the reference by more or less than nine base pairs

- If a frame shift exists in the translatable consensus area

- If there were any ambiguous codons in consensus sequence

- If a substitution variant to the reference sequence has a frequency of less than 75% of reads

- If minimum coverage is less than 5.

The matching thresholds for scanning targets required 90% similarity to the reference sequence and 99% aligned to the reference sequence. Accessory targets were not used in the following analysis.

Following quality control of the data a cgMLST scheme was created using 1622 cgMLST targets from the 1858 reference genome targets. Of the cgMLST targets the number of missing targets per isolate was normally distributed ranging from 2 – 11 targets. This equated to a percentage of useable targets ranging from 99.3% to 99.9%.

Cluster analysis was carried out in Ridom Seqsphere using neighbour joining trees and minimum spanning trees. During distance matrix calculation the missing target values for individual isolates was assessed by two different methods; pairwise omission and as a separate column of values at each stage of analysis. However, no significant difference to the clustering outcome was identified between these two methods in any of the analyses (see Chapter 6).

2.8 Matrix Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI)

MALDI is a widely used in microbiology as a typing method, most frequently for species level diagnostic and research purposes (Barbuddhe et al. 2008; Christner et al. 2014; Goncalves et al. 2014; Mehta and Silva 2015; Nakano et al. 2015; Rettinger et al. 2012; Tagg et al. 2015). The process involves the vaporisation of an ionized mixture of biological material and matrix solution on a metal target plate by a high power laser followed by the acceleration of those charged particles across a vacuum to impact upon a detector. The delay between vaporisation and reception at the detector correlates with the mass of the charged particle in question while the intensity of the signal correlates with the abundance of that charged particle.

Over multiple repeat laser desorption's (40 desorptions x 6 technical replicates), at different random locations on the target pate a spectral profile is built up of multiple intensity regions or 'peaks' distributed across a wide range of molecular weights. Individual spectra, generated from a single specific laser desorption are excluded automatically if they deviate significantly from the average of all spectra for that isolate. This quality control function removes un-representative spectra that may be the result of a desorption at an area of low sample density or poor matrix coverage on the target plate for example. The averaged individual spectra from each desorption are then averaged to produce the isolate spectral profile.

Spectral profiles can be compared mathematically in a number of ways (Trede et al. 2012). In this study pairwise analysis and group analysis were conducted. Composite correlation index (CCI) matrices were generated for pairwise, quantitative comparison of whole spectral profiles of two isolates using the average of each pairwise comparison of their technical replicates. Group modelling of isolate data incorporated a higher level of comparison between isolates. This was conducted using two alternative modelling strategies that use different methods for isolate discrimination as described below, and in greater detail in Chapter 5, section 5.2.2.

The Supervised Neural Network (SNN) model algorithm use a prototypical typing system to assign isolates according to their relative proximity in principal component analysis terms, to a predefined 'prototypical spectra'. The model uses prototypical spectra based upon the pre-existing group

structure that is defined in the training set by the operator hence a 'supervised' neural network. An alternative approach was adopted using the Genetic Algorithm (GA) model function which selects a small number of individual peaks into combinations which are then tested and reasserted to generate a group with the greatest discriminatory ability for the predefined groups. The GA model therefore explicitly selects only a very limited proportion of the spectral profile and ignores the remainder.

Model fit for both approaches was then tested by a process of internal cross validation by random exclusion of 20% of the input data, recalculation of the model using the remaining 80% and refitting of the excluded 20% as a test dataset. The process is repeated ten times. More details of these model are provided in Chapter 5, (5.2.2).

2.8.1 Isolate selection and recovery

All 494 isolates which had previously been successfully recovered and successfully generated complete MLST profiles underwent recovery for MALDI-ToF analysis, along with 278 isolates which had generated partial MLST allele profiles. Isolate recovery, extraction and quality control were carried out at QMMS Ltd by the author and Dr Selin Cooper. Further optimisation and quality control was carried out in collaboration with Dr Cooper. Data organisation of CCI matrices was carried out with the help of Dr Simon Archer.

Recovered putative *S.uberis* colonies were incubated on blood agar rather than BHI agar to provide optimum internal reliability between technical replicates based upon previous observations of increased spectral heterogeneity amongst technical replicates from isolates cultured on BHI compared to blood agar (Cooper personal communication 2014). Isolates were incubated at 37 °C for 18 to 24 hours. Bacterial species level identification was carried out by the direct whole culture MALDI prior to ethanol-formic acid protein extraction to obtain a high quality lysate solution for analysis according the protocol below.

2.8.2 Ethanol-Formic Acid Extraction Protocol

One colony (or a few colonies in the case of a small colony size or confluent growth) was suspended in 300 µl of molecular-grade water (Sigma-Aldrich, St. Louis, MO) and vortexed. 900 µl of 100% ethanol (Sigma-Aldrich, ≥99.8%) was added and the suspension vortexed for 5 seconds. Suspension was then centrifuged (20,800 × *g*) for 3 min. The supernatant was decanted and discarded. The pellet was dried at room temperature. Fifty microliters of 70% formic acid (Fluka [Sigma-Aldrich], St. Louis, MO) and 50 µl of acetonitrile (Fluka) were added and thoroughly mixed by pipetting followed by centrifugation (20,800 × *g*) for 2 min. Two microliters of supernatant was spotted onto the 24-spot plate and allowed to dry at room temperature before the addition of 2 µl HCCA matrix solution. The sample spot was then air dried before analysis. This is a modification of the protocol described previously (Barreiro et al. 2010).

2.8.3 Isolate quality assessment of spectral profiles.

MALDI analysis confirmed 521 isolates as *S.uberis.* Each isolate analysis was repeated to generate 6 technical replicates of the spectra in the range 4,000 to 10,000 Da. Spectra were compared visually using Biotyper3.0 (S. Nakano et al. 2015). Those technical replicate spectra with insufficient resolution, low intensity, or substantial background noise were excluded from the analysis. Technical replicates of each isolate were then compared using Composite Correlation Indices (CCI). CCI's are percentage similarity scores applied to a pair of isolates based upon the similarity of their spectra, specifically the number of shared peaks compared to the number of non-shared peaks and the relative peak intensity of those shared peaks. Technical replicate spectra from the same isolate, extracted from the same colonies at the same time should very similar so CCI scores of each and every pair of technical replicate of an isolate was used to remove unreliable, dissimilar spectra with a CCI < 0.99 (Arnold and Reilly 1998). CCI matrices of all isolate combinations were created for each herd from the confirmed 521 isolates.

Chapter 3: Analysis of the population dynamics of *S.uberis* clinical mastitis using multilocus sequence typing.

An accurate understanding of the likely transmission route is fundamental when designing risk based mastitis control strategies. *S.uberis* has been repeatedly identified as the most commonly isolated pathogen from clinical and sub-clinical samples in several countries including Australia, UK, New Zealand and Belgium (Green et al. 2007; Shum et al. 2009; Verbeke et al. 2014; Zadoks and Fitzpatrick 2009) and has been implicated in contagious transmission (Cattell 1996; R N Zadoks, Allore, Barkema, Sampimon, Gröhn, et al. 2001). To date there have been no studies conducted to estimate the relative prevalence of possible contagious transmission of *S.uberis* in the UK dairy population or describe the variation between herds in contagious versus environmental transmission of *S.uberis* clinical mastitis. To address this knowledge gap a uniquely large and comprehensive existing archive of UK clinical mastitis samples was analysed using Multi Locus Sequence Typing (MLST). MLST has been shown to provide a highly discriminatory and reproducible means of investigating the molecular epidemiology of *S.uberis* (Coffey et al. 2006; Pullinger et al. 2006, 2007; Shome et al. 2012; Wang et al. 2009; Zadoks, Schukken, et al. 2005). Previously this has been undertaken primarily on isolates from clinical and subclinical mastitis, which have provided insights into the population diversity (Coffey et al. 2006; Pullinger et

al. 2007) as well as identifying distinct geographical variation in strain prevalence (Tomita, Meehan, Wongkattiya, Malmo, et al. 2008).

The aim of this chapter was to evaluate the heterogeneity of *S.uberis* strain types between and within herds, and assess the extent of possible within herd contagious transmission of *S.uberis* using MLST as a method of sub-species typing.

3.1 Methods

The results and analysis presented in this chapter refer the Herd selection criteria, sample collection, isolate recovery and MLST protocols discussed in detail in Chapter 2 (section 2.6) and in (Davies et al. 2016). MLST sequence types were used to assign clinical cases into one of five, mutually exclusive categories according to a combination of genotype observed epidemiological patterns. The following case classifications were defined as shown below.

Data analysis was conducted in *Microsoft Excel, Microsoft Access, MiniTab17*. Locus variant cluster assignment was conducted in *eBURST Version3.0* (http://eburst.mlst.net/v3) with nodes of single, double and triple locus variants connected. Node area was scaled linearly according to sequence type frequency.

3.1.1 Mastitis clinical case classifications

Index case (I) ; An index case is the first chronological recorded mastitis case attributed to an MLST sequence type in a herd where that same sequence type is identified in more than one cow during the study period. Subsequent mastitis cases with the same ST in a different cow in the same herd are classified as potential transmission events (Davies et al. 2016).

Potential Transmission Event (PTE); Following an index case in a herd subsequent clinical cases caused by **that specific sequence type** in another cow in that herd were classified as a 'Potential Transmission Event'(PTE) on the first occasion in that new cow. Subsequent identification of that same sequence type in that cow were classified as Persistent infection of the first PTE clinical case identified in that cow (Davies et al. 2016).

Persistent (P); Using the established definition described previously (Baseggio et al. 1997). When **the same sequence type** was identified on two or more occasions from the same mammary gland quarter all but the first identified case caused by that sequence type were classified as persistent infections (Davies et al. 2016).

Unclassified (U); Clinical cases where the identified sequence type was isolated only once in a particular herd but was also identified (as a Persistent, Index or PTE) in another herd (Davies et al. 2016).

Solitary (S); Sequence types identified in only one clinical case throughout the study period (Davies et al. 2016).

3.2 Results

The results described here are a more detailed version of those provided in the paper (Davies et al. 2016). Of the 3918 clinical mastitis case milk samples submitted for bacteriology 1099 samples cultured were recorded as *S.uberis* infections during the original DCMCP study (Green et al. 2007). Of the 1099 *S.uberis* samples 944 were categorised as pure *S.uberis* cultures and 155 were categorised as mixed infections with *S.uberis* and one or more other pathogens including most commonly E.coli (52 samples), Bacillus (49 samples), Aspergillus (11 samples), Staph aureus (7 samples) also Klebsiella, Mucor, Yeast, Pasteurella, other Streptococci spp, coagulase positive Staphylococci spp, Proteus and Serratia.

Following sequencing 54 of the 494 isolates analysed had originated from mixed infection samples. The origin of the isolates from a pure or mixed infection did not have any significant association or trend with MLST sequence type or clinical case classification.

994 of the 1099 submitted samples could be paired with an individual cow identifier with production records. Of the 994 samples, 624 cows were diagnosed with a single case of *S.uberis* mastitis during the study period. 113 cows were diagnosed with two cases, 29 cows with three cases, 8 cows with four cases and 5 cows with five cases of *S.uberis* mastitis during the study period.

3.2.1 MLST sequencing performance analysis

In total 854 *S.uberis* isolates (86%) were successfully recovered from the 994 stored *S.uberis* samples linked to an individual cow identifier. Of these, 494 isolates generated complete MLST profiles at all seven loci. A further 278 isolates generated partial profiles with 1-6 sequenced loci. No loci could be sequenced in the remaining 82 isolates. The frequency of locus absence is presented in Table 1. Of the remaining 222 isolates, 81 isolates produced no sequences using the MLST primers, 25 were excluded due to the recovered isolate being classified as a non-*S.uberis* species and 116 could not be recovered from the stored sample.

Table 1. The frequency of missing loci from the 278 isolates with partial *S.uberis* MLST profiles where between 1 – 6 loci were successfully sequenced.

MLST Locus	Number and % of un-sequenced loci from confirmed *S.uberis* isolates by MALDI-ToF MS.	
arcC	53	7%
ddl	153	20%
gki	63	8%
recP	47	6%
tpi	218	28%
tdk	157	22%
yqiL	68	9%

3.2.2 MLST sequence type population structure

Cluster analysis was performed using pubMLST webtools (www.pubMLST.org) to establish clonal complex structure and allow comparison with the wider pubMLST *S.uberis* database. Using the accepted clonal complex definition of 4 shared alleles of the 7 loci between the group of sequence types in the

complex. Sequence types from the three principal clonal complexes previously described (Coffey et al. 2006; Tomita, Meehan, Wongkattiya, Malmo, et al. 2008) were identified in this study along with singleton sequence types. Clonal complex five (CC5) formed the largest group accounting for 42.5% of all sequence types and 63% of clinical cases. Clonal complexes 86 and 143 accounted for 2.1% and 3.1% of sequence types and 1.4% and 1.2% of clinical cases, respectively whilst the remaining 52.3% of sequence types were singletons.

3.2.3 Herd and species level mastitis dynamics

Average herd size over the study period ranged from 55 to 320 cows, with a mean of 164 and a median of 163 cows per herd. The median incidence of reported clinical mastitis during the original study period in the 52 recruited herds was 66 (mean, 75) cases per 100 cows per 365 days, with a range of 16 to 146 cases. The mean percentage of clinical cases diagnosed as *S.uberis* was 28%, with a range in individual herds of 7% to 64%. No variation in mastitis incidence or *S.uberis* prevalence was observed by geographical region.

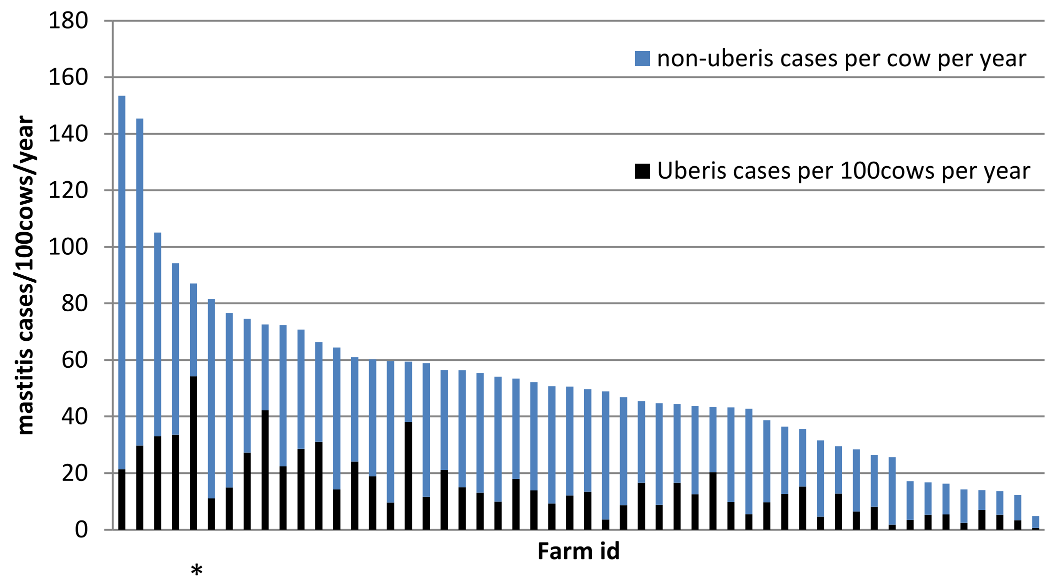

Figure 1. Incidence rate of clinical mastitis cases per herd caused by *S.uberis* (Black) and other pathogens (Blue). * = Farm27 (highest % of mastitis cases attributed to *S.uberis*).

3.2.4 Herd level MLST descriptive analysis

S.uberis MLST sequence types were identified from clinical mastitis samples collected from 51 of the 52 herds. One hundred and ninety five sequence types were identified across all herds. The majority of the sequence types (148 of 195, 76%) were identified in only a single herd. Of these 127 sequence types, were cultured from only one mastitis sample in that single herd. The remaining 47 sequence types were cultured from more than one herd: 26 STs were identified in 2 herds, 16 STs were identified in 3-9 herds while 5 STs were identified in isolates from 11 to 20 of the 52 herds. The mean number of sequence types (STs) identified per herd was 3.75 and median 6 STs the range extending from zero to twenty sequence types per herd (Figure 2). The most widespread sequence types were also the most numerous sequence types (Table 2).

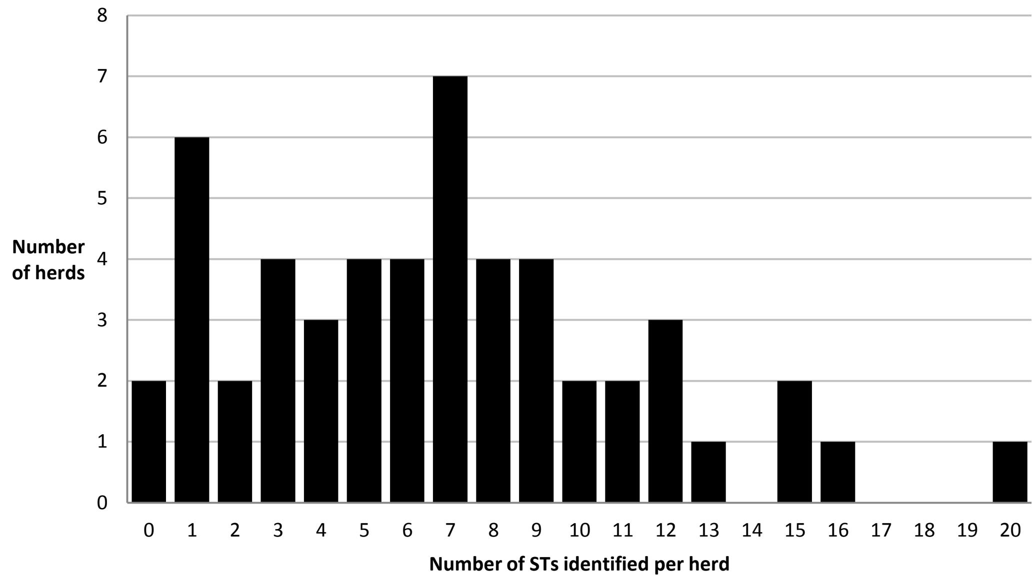

Figure 2. Number of unique multi locus sequence types of clinical *S.uberis* isolates identified in different dairy herds.

Table 2. Distribution of the most prevalent five MLST sequence types across all farms and the five most prevalent Sequence Types (ST's) as a percentage of all MLST sequenced clinical cases from all herds.

Most widespread MLST sequence type across herds. (% of farms with identified ST's)		Most common MLST sequence type (% of all clinical cases attributed to that sequence type from all herds)	
Sequence Type	% Farms with that ST	Sequence Type	% of all clinical cases that are attributed to that ST
6	38%	6	10%
22	38%	22	6%
5	29%	5	5%
35	21%	35	4%
24	21%	20	4%

3.2.5 Cow level MLST descriptive analysis

In total 494 complete MLST sequence profiles were obtained from bacterial isolates cultured from clinical mastitis samples collected from 419 cows for

which accompanying cow identifier information was available. Of these, 365 (87%) cows experienced a single quarter case, 46 (11%) cows experienced two quarter cases and 8 (2%) cows experienced 3, 4 or 5 clinical mastitis quarter cases attributed to *S.uberis* isolates with a complete MLST sequence profile.

Forty four quarters in separate cows experienced more than one clinical case of *S.uberis*. Twenty two of these quarters experienced persistent infections with the same sequence type identified on two or three separate occasions (each separate culture isolate was classified as a clinical case even if subsequently classed as a continuation of an earlier infection). The remaining twelve were reinfections with a different sequence type. Sixteen sequence types caused persistent infections, which were defined as cases where the same sequence type was identified in same quarter on successive occasions. The average interval between sampling events was 32 days (Table 3).

Table 3. Distribution of persistent mastitis clinical cases between MLST sequence types of *S.uberis* from all herds.

MLST Sequence Type	Number of Persistent Infection isolates	Mean Interval between consecutive cases (days)
6	6	13
67	4	35
233	4	74
372	4	79
5	4	29
20	3	17
343	3	12
537	3	53
512	2	105
597	2	8
510	2	14
528	2	41
361	2	23
35	2	7
23	2	25
24	2	8

3.2.6 Analysis of epidemiology patterns in MLST sequence type results

Isolates from all clinical cases were assigned to one of the five clinically defined classifications according to the observed epidemiological patterns

described in section (3.1): Index (I); Potential Transmission Event (PTE); Persistent (P); Solitary (S) and Unclassified (U) (Table 4).The 'Solitary' sequence types accounted for 65% of the of all the MLST sequence types identified within the dataset and were identified in 90% of the study herds. However, 'Solitary' sequence types only accounted for 25.7% of the clinical mastitis cases. The majority, 279 of 494 isolates sequenced in this dataset (56.48%) were attributed to one of 35 sequence types, all of which caused PTE cases in at least one herd (Table 5). Of these 279 isolates, 198 (71%) isolates were 'I', 'PTE' or associated 'P' cases. The remaining 81 isolates attributed to one of the PTE causing sequence types were assigned as 'U' cases, identified only once in their herd. Potential Transmission Events 'PTE' clinical cases were identified in 33 of 52 herds (63.4%) (

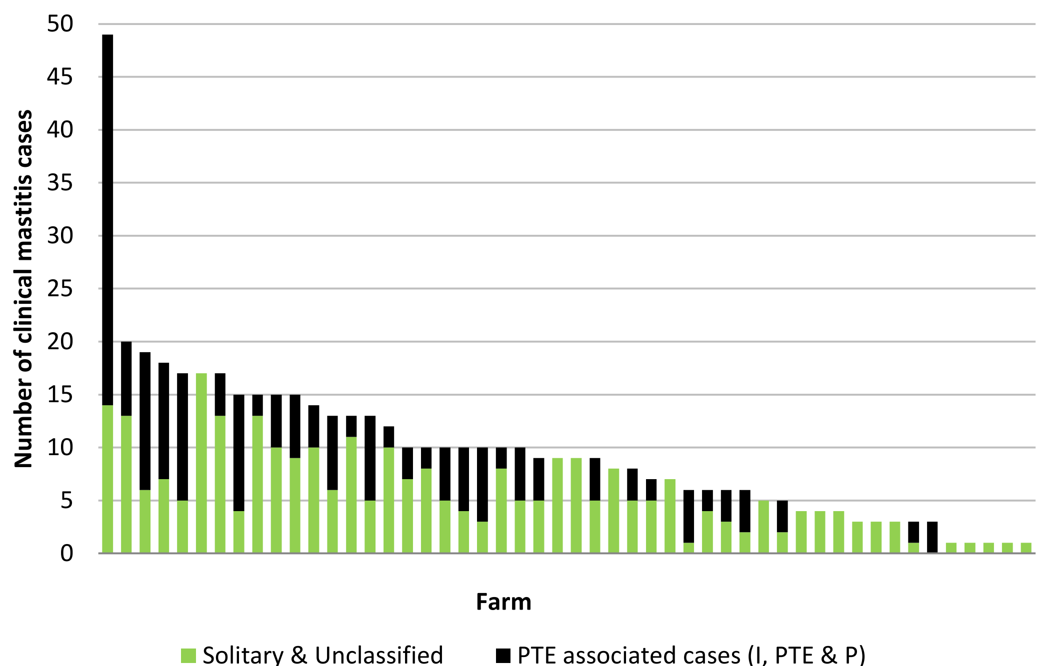

Figure 6). In 17 herds (33%) Index/PTE/Persistent clinical cases accounted for more than 50% of the clinical cases illustrated in (

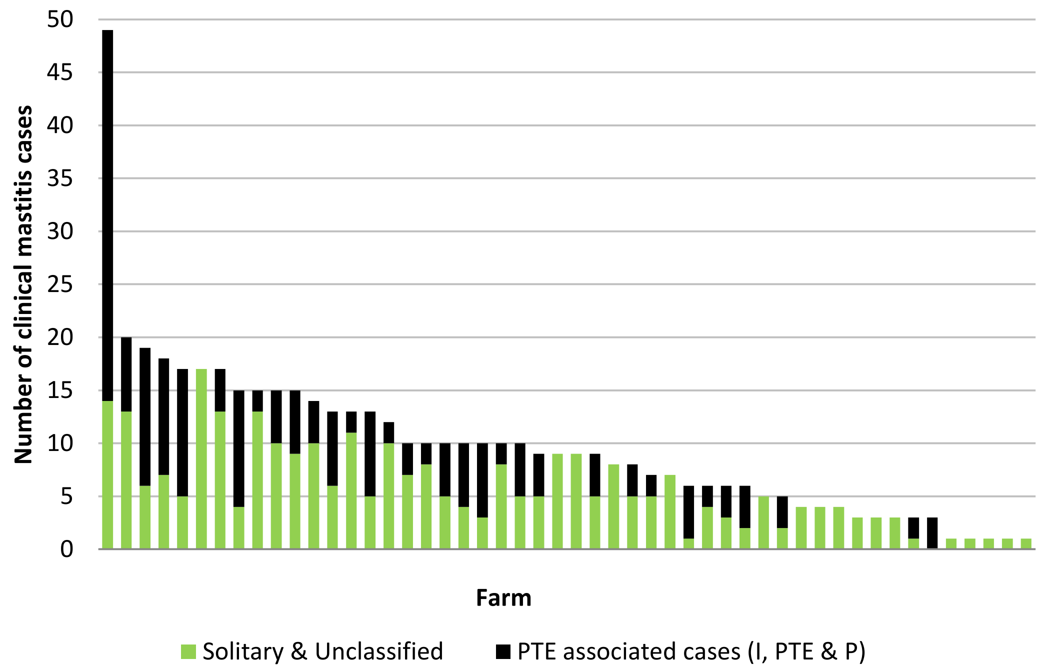

Figure 6). The mean number of PTE mastitis cases following an Index case was 1.6 with a range from 1 to 15 clinical cases. There was no significant correlation between the proportion of the isolates classified as Index, PTE or Persistent and herd size, total herd mastitis prevalence or prevalence of *S.uberis* within a herd.

3.2.7 Analysis of herd level variables on MLST patterns

Herd size, total mastitis incidence rate and *S.uberis* mastitis incidence were compared graphically to the percentage of *S.uberis* cases attributed to the potentially contagious case classifications (Index/PTE/Persistent) using scatter plots (Figure 3, Figure 4, Figure 5).

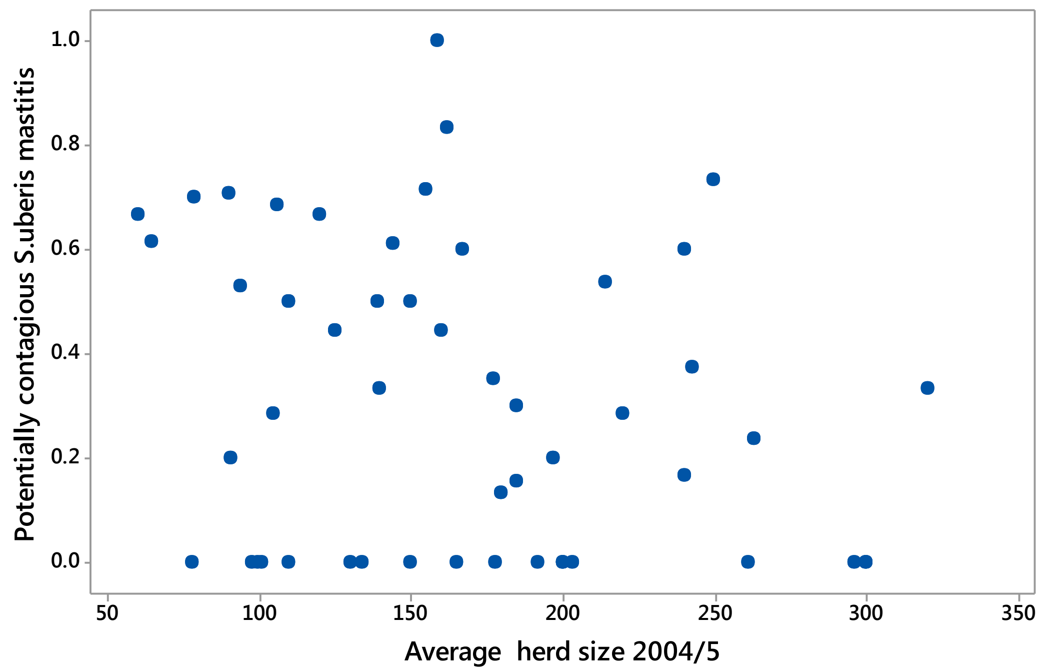

Figure 3. Scatter plot of the proportion of the total *S.uberis* clinical mastitis classified as potentially contagious (Index/PTE/Persistent) against herd size.

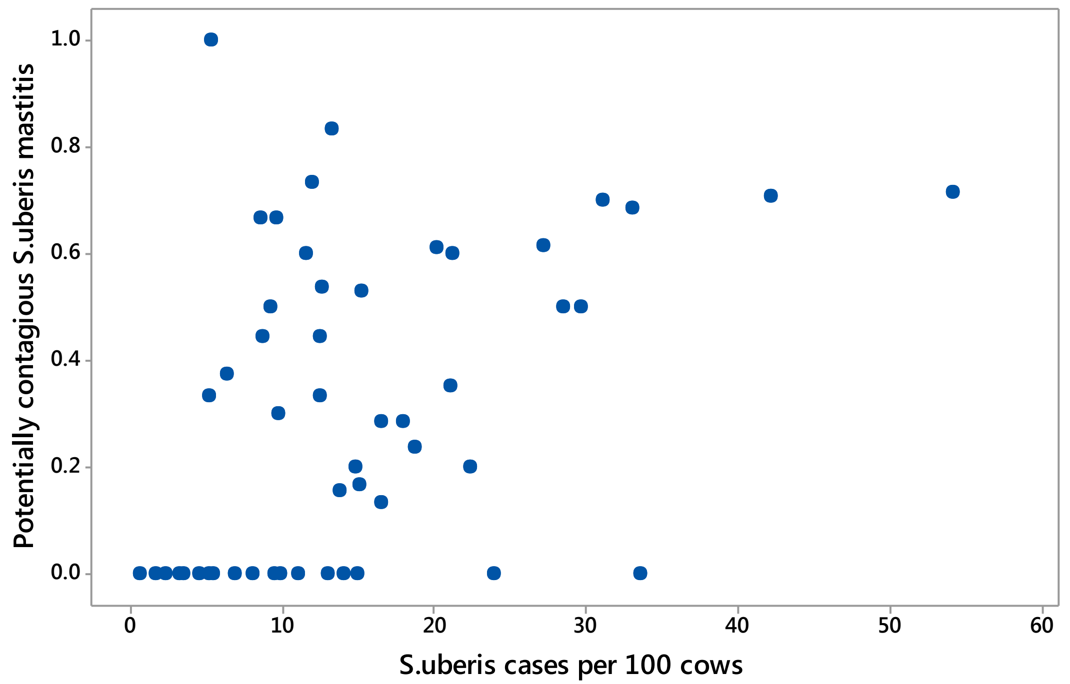

Figure 4. Scatter plot of the proportion of *S.uberis* clinical cases per farm classified as potentially contagious (Index/PTE/Persistent) against the incidence rate of *S.uberis* clinical mastitis per herd during the 16 month study period.

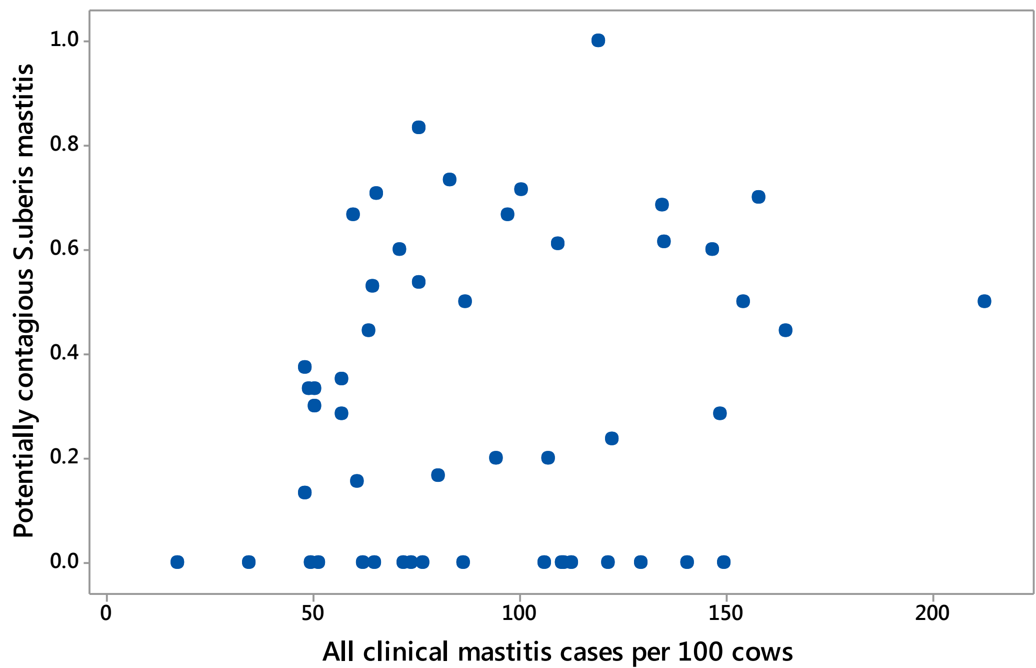

Figure 5. Scatter plot of the proportion of *S.uberis* clinical cases classified as potentially contagious (Index/PTE/Persistent) against total clinical mastitis incidence rate over the 16 month study period including all S.uberis clinical mastitis per herd.

3.2.8 Statistically Overrepresented Sequence Types

Out of all 195 sequence types nine were significantly overrepresented, causing ten or more clinical cases across all herds (p=0.045, Fisher's exact Chi-squared test). These nine overrepresented (OVR) sequence types caused 38.25% of all the clinical mastitis cases. Eight of the nine OVR sequence types were found in multiple herds. Eight of the nine sequence types were also identified in persistent infections (Table 5).

The PTE STs were significantly more likely to be classified as persistent cases, accounting for 85% (23 of 27) of the persistent cases Table 3 & Table 5. Within the 35 PTE sequence types, the OVR subgroup mastitis cases were significantly more likely to be classified as persistent, accounting for 63% (17 of 27) of the identified persistent infections in the study.

Table 4. Number of *S.uberis* clinical mastitis cases grouped by case classification for all MLST sequence types across all herds. (I) Index case; 1st case of an ST causing >1 mastitis case in different cows in the same herd. (PTE) Potential Transmission Event; Following an index case in a herd each subsequent clinical case caused by that ST another cow (P) Persistent; 2nd or subsequent isolation of same ST from same quarter. (U) Unclassified; ST identified in >1 herd but the only case in that herd. (S) Solitary; ST occurred in only one clinical case throughout the study period.

Clinical Case Classification	Number and % of clinical cases per classification
Index case (I)	68 (14%)
Potential Transmission Event (PTE)	108 (22%)
Persistent case (P)	27 (5%)
Solitary(S)	127 (26%)
Unclassified (U)	164 (33%)

3.2.9 eBURST analysis of MLST groupings

A small number of sequence types of common lineage were identified that caused approximately 40% of clinical mastitis cases attributable to *S.uberis* (Table 5). Clustering analysis (Figure 7) illustrates the relative similarity of eight of the nine OVR sequence types ST5, ST6, ST22, ST24, ST35, ST233, ST361, ST512 and ST20 (separate lineage). With the exception of ST20 the other over represented sequence types form part of the Clonal Complex previously defined as CC5 (Coffey et al. 2006). The singleton sequence types varied by more than 3 alleles both from the CC5 cluster and from each other, leading to the majority remaining un-grouped. In this study eBURST V3 defined ST6 as the 'founder' and four of the OVR sequence types are single locus variants and two double locus variants (Table 5). ST20 is the fifth most commonly isolated ST and the only over represented ST found to be in a separate lineage.

Table 5. All 35 *S.uberis* MLST sequence types from clinical mastitis isolates associated with potential transmission events (PTE). Cases listed in descending order of their overall prevalence as distribution between classification groups. Index Cases (I), Potential Transmitted Events (PTE) or Persistent (P) cases and Unclassified (U) cases. Statistically over represented (OVR) sequence types are indicated by (*).

MLST Sequence type	Number (%) of all mastitis cases in all herds	Number of 'Index' case = number of herds	Number of 'PTE' cases	Number Persistent cases	Number of 'Unclassified' cases	Number of non-identical loci from founder (ST6)
6*	49 (10.1)	9	23	5	12	0
22*	30 (6.2)	8	11	-	11	2
5*	23 (4.7)	6	7	2	8	1
35*	20 (4.1)	5	6	2	7	2
20*	17 (3.5)	3	8	2	4	5
233*	15 (3.1)	3	5	3	4	1
24*	14 (2.9)	1	2	1	10	3
361*	11 (2.3)	2	3	1	5	1
512*	10 (2.1)	3	3	1	3	1
67	9 (1.9)	3	3	2	1	3
343	9 (1.9)	1	2	2	4	6
10	6 (1.2)	1	2	-	3	2
595	6 (1.2)	1	1	-	4	6
523	4 (0.8)	1	2	-	1	4

Table 5 continued from previous page						
MLST Sequence type	Number (%) of all mastitis cases in all herds	Number of 'Index' case = number of herds	Number of 'PTE' cases	Number Persistent cases	Number of 'Unclassified' cases	Number of non-identical loci from founder (ST6)
528	4 (0.8)	1	1	1	1	6
553	4 (0.8)	1	3	-	-	3
597	4 (0.8)	1	2	1	-	1
496	3 (0.6)	1	1	-	1	7
544	3 (0.6)	1	1	-	1	2
545	3 (0.6)	1	1	-	1	3
497	3 (0.6)	1	2	-	-	3
501	3 (0.6)	1	2	-	-	3
507	3 (0.6)	1	2	-	-	5
511	3 (0.6)	1	2	-	-	3
577	3 (0.6)	1	2	-	-	4
9	2 (0.4)	1	1	-	-	2
476	2 (0.4)	1	1	-	-	6
480	2 (0.4)	1	1	-	-	5
509	2 (0.4)	1	1	-	-	3
526	2 (0.4)	1	1	-	-	2
546	2 (0.4)	1	1	-	-	3
552	2 (0.4)	1	1	-	-	2
560	2 (0.4)	1	1	-	-	3
561	2 (0.4)	1	1	-	-	5
614	2 (0.4)	1	1	-	-	3

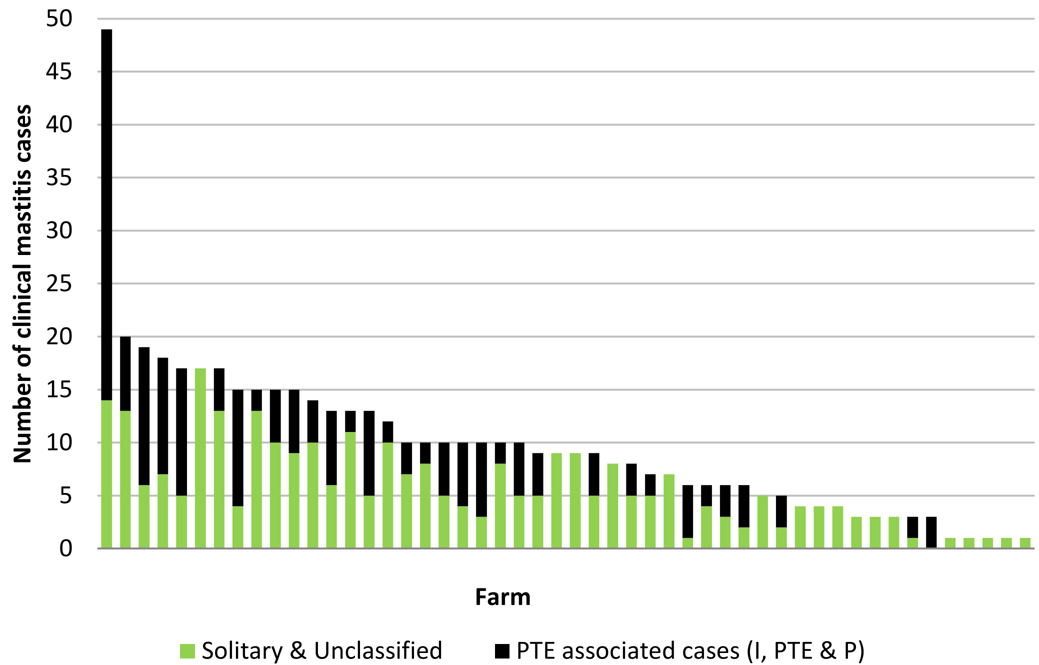

Figure 6. Mastitis clinical cases by classification group for each farm where *S.uberis* mastitis cases were identified with complete MLST profiles (in descending number of total number of identified isolates - Farm 27 at extreme left).

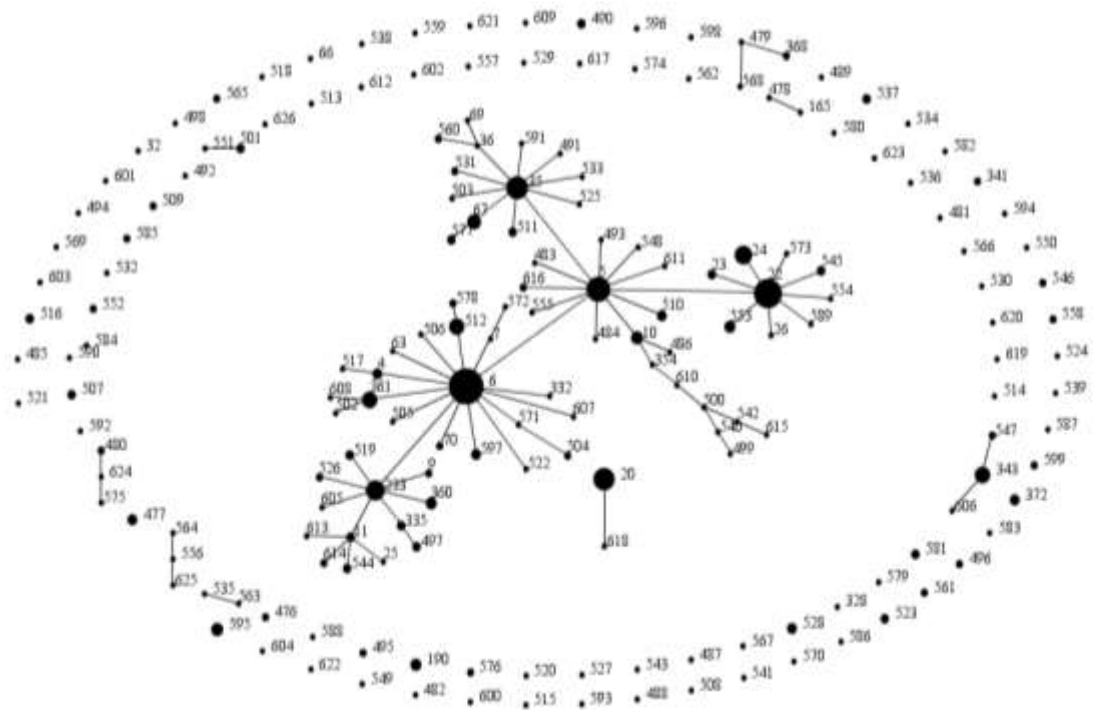

Figure 7. eBURST V3 population plot of ST's identified in this study the node size proportionate to the number of clinical mastitis isolates attributed to that sequence type. ***Joining lines indicate a single locus variation (SLV) between isolates. Double locus variants (DLV) are only linked if a common SLV is also present in the sample set. The presumed founder (ST6) is defined as the sequence type with the highest frequency of single locus variants. Eight of the nine statistically over represented sequence types are found within this group (ST 5,6,22,24,35,233,361,and 512).***

3.3 Discussion

S.uberis has been generally classified as an opportunistic 'environmental' pathogen since the early 1970s following its identification in straw bedding and as a commensal organism on skin and in faeces (Anon 1970) as well as in infected bovine mammary glands. This is in contrast to the more prevalent, obligate pathogens of that era, principally *S. agalactiae*, which appeared to reside principally or exclusively in the mammary gland, which were in turn classified as 'contagious' pathogens. This broad classification is likely to be an over simplification of the epidemiology (Hillerton et al. 1993; Milne et al. 2005; Phuektes et al. 2001). Several studies investigating *S.uberis* mastitis have identified temporal patterns of clinical mastitis which are suggestive of contagious transmission (Cattell 1996; R N Zadoks, Allore, Barkema, Sampimon, Gröhn, et al. 2001) occurring as outbreaks in individual herds. There are also several published studies using MLST on a relatively small numbers of samples collected from a small number of herds which have demonstrated a wide variety of sequence types, indicating that infections were likely to have been acquired from a diverse environmental population (Coffey et al. 2006; Pullinger et al. 2007; Tomita 2008). Studies have also shown that one pathogenic sequence type can predominate in a herd (Coffey et al. 2006; Phuektes et al. 2001; Pullinger et al. 2007). The same STs (ST 5, 6 and 20), previously identified as predominant in different herds have also been identified as over represented sequence types in this study. The prevalence of ST's in mastitis samples compared to environmental samples

has been shown to be significantly different (Pullinger et al. 2006) suggesting either a variable fitness to colonise exists between ST's in the environment exists or some degree of cow to cow transmission occurs. Given the sequence type diversity catalogued in the *S.uberis* MLST database (635 separate sequence types), the heterogeneity of environmental populations (Zadoks, Tikofsky, and Boor 2005) and the limited number of clinical cases per herd during the study period, it would seem implausible that multiple cows in the same herd would become infected with the same sequence type if there were no additional fitness or contagious phenotypic attributes, specific to these highly prevalent strains. The potential for contagious transmission has also been described by several studies which demonstrated temporal patterns (R N Zadoks, Allore, Barkema, Sampimon, Gröhn, et al. 2001) and herd level mastitis treatment and control protocols (Cattell 1996; Hillerton et al. 1993) which support the contagious transmission hypothesis.

Analysis of the prevalence of sequence types across all herds identified nine strains which were significantly more prevalent (OVR sequence types) than would be expected if infection was a 'chance' event process as would be expected if opportunistic infection from the environment was the main transmission route. The nine over represented STs accounted for only 4.7% of the ST identified in the study. However, these sequence types were diagnosed in 38% of sequenced *S.uberis* clinical mastitis cases and 63% of persistent mastitis infections. Whilst it should be recognised that the distribution of mastitis prevalence could reflect an underlying distribution within the *S.uberis*

population in the farm environment, this would seem unlikely given the mastitis and environment populations have previously been shown to be significantly different (Pullinger et al. 2006).

It is undeniable that the results of this study would have been substantially enhanced by the inclusion of *S.uberis* isolates collected periodically throughout the study period from the environment, such as bedding and pasture as well as other potential bacterial resevoirs and sources of contamination such as the bovine gastrointestinal tract and sub-clinically infected mammary glands (High SCC cows). Unfortunatly no such samples were collected during the original data collection period in 2004-5. The information of relative sequence type prevalence would have provided a useful baseline to interpret the plausibility of cow to cow transmission as opposed to direct environmental infection.

Of the 52 herds studied 63% had at least one pair of mastitis cases in different cows caused by the same sequence type. Whilst not conclusive, this result suggests that either contagious transmission or enhanced infectivity amongst the environmental population may be more common and more important than previously thought. Pathogens which acquire virulence genes for intramammary infection may reside in the environment and cause new infections as an environmental opportunist. However, it is logical to assume that environmental pathogens with enhanced infectivity will therefore present an increased contagious risk, and poses the ability to maintain bacterial numbers at a high enough level for long enough to contaminate

milking equipment with the requisite infectious dose. It would therefore be reasonable to hypothesise that the interval between clinical mastitis cases (the inter-case interval) caused by contagious transmission would tend to be shorter than the interval between opportunistic infections from the environment (Chapter 4). Especially if their virulence mechanisms involve host immune evasion and the establishment of persistent infections as has been suggested by previous studies (Almeida and Oliver 2006). Persistent infections, by definition, are present during more milking periods (Anon 1999) than transient infections and therefore present an increased number of transmission opportunities for an infectious pathogen. In this study the overwhelming dominance of PTE sequence types and particularly the OVR sequence types in the Persistent mastitis classification suggests that these sequence types would have more opportunities to be transmitted from cow to cow during milking.

However, defining what is indeed a persistent mastitis case rather than antibiotic treatment failure or indeed antibiotic resistance. Previous studies have defined clinical mastitis cases in the same quarter gland as a continuation of the same clinical case rather than a new case unless the interval between samples exceeded 14 days (Barkema et al. 1998). In this study several of the intervals were less than 14 days and given the relatively small number of mastitis cases coded as Persistent it is possible that some of these cases were the result of treatment failure rather than a fundamental

difference in their biological ability to establish a persistent infection related to their sequence type.

Previous studies have identified genetic sequences associated with important infectious processes such as colonisation (Reinoso et al. 2011; Ward et al. 2009). Genomic analysis using isolates from clinical and sub-clinical mastitis cases show that complex multi-gene arrangement, rather than the simple presence or absence of virulence determinants, influences bacterial pathogenicity and the outcome of the infection (Hossain et al. 2015).

The results of this study confirm the predominant clonal complex in the UK to be CC5 in contrast to Australia and New Zealand where CC86 and CC143 are more common (Pullinger et al. 2006, 2007; Tomita, Meehan, Wongkattiya, Malmo, et al. 2008). Clonal complex 143 has also been identified in China (Wang et al. 2013) whereas in India only singleton ST's were identified (Shome et al. 2012). Whilst some of these studies involved only a limited number of isolates their results do indicate geographical variation in the *S.uberis* population.

The more extensive management systems practiced in some countries and different pre and post milking hygiene protocols may predispose to different patterns of *S.uberis* mastitis in different countries and it is plausible that contagious transmission may not be as likely under those conditions. This study provides the first estimate of the prevalence of these 'potentially contagious patterns' between UK herds and the distribution of incidence rates of potentially contagious transmission events within herds. The finding

that potential contagious transmission contributed > 50% of the total clinical mastitis in 33% of herds studied suggests that contagious transmission may represent a major barrier to conventional control protocols designed to limit new *S.uberis* infections which focus on environmental factors alone. It is unclear how effective individual contagious mastitis control measures are against *S.uberis* in comparison to *S.aureus* for example as appropriately powered randomised control trials have not been conducted. However, (Wesen and Schultz 1970) did show a comparable trend in the reduction of new mastitis cases due to *S.uberis* and *S.aureus* in response to post-milking teat disinfectant compared to a control group. This study demonstrated a trend that indicates some efficacy of contagious control measures for *S.uberis* but the findings of the current project would imply that this effect may depend upon the *S.uberis* sequence types present in the herd. The low infectious dose required to establish *S.uberis* intramammary infections (Leigh 1999) may also present a significant challenge for preventative control measures.

Previous studies have discussed contagious transmission as a means of explaining dramatic clinical mastitis outbreaks (Cattell 1996; R N Zadoks, Allore, Barkema, Sampimon, Gröhn, et al. 2001) in individual herds. In this study 1 of the 52 herds (Farm27) displayed a temporal pattern of clinical cases that would have fallen within the definition of an outbreak that has been previously proposed (R N Zadoks, Allore, Barkema, Sampimon, Gröhn, et al.

2001). In this herd 71% of the clinical cases were classified as PTEs (

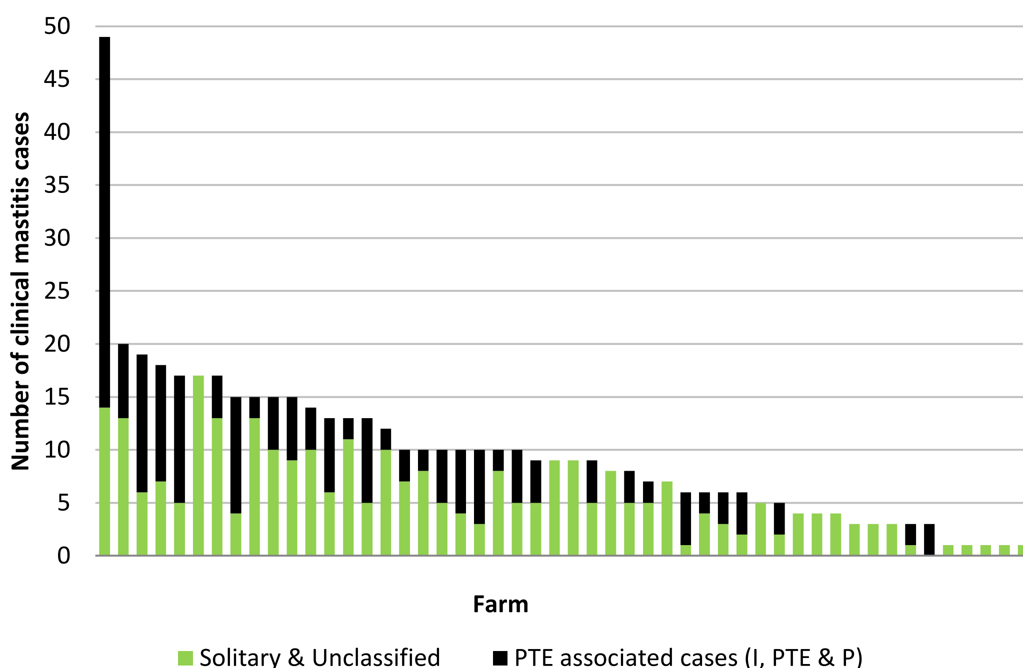

Figure 6, Farm27) with high ratios of Index:PTE cases for several OVR STs (eg. ST6 1:15 and ST20 1:6). However, the current findings highlights the potential role of insidious, low-level cow to cow transmission at Index:PTE ratios of 1:1 or 1:2 as a major component of mastitis cases in the majority (63%) of herds. This insidious manifestation of potentially contagious mastitis may arguably be more important and more challenging than the rare, dramatic mastitis outbreaks, it is only with the aid of molecular techniques that estimates of the relative importance of these two modes of transmission can be attempted.

The selection criteria used in the recruitment of herds provided a reasonable cross section of herds and management practices at the time of data collection. However, herds with substantially higher incidence of *S.uberis* clinical mastitis or significantly different management practices may exhibit higher rates of PTE than those identified in this study. This study has

identified several, specific sequence types which warrant further investigation due to their higher prevalence and wider distribution. Further work is required to elucidate the underlying molecular mechanisms, host-pathogen interactions and the influence of management interventions in the colonisation of mammary gland by *S.uberis*. This may allow us to understand how and why different patterns of ‘potentially contagious’ or ‘environmental’ mastitis appear to predominate in different herds. The variability observed between herds in this study may provide some explanation of the apparently intractable difficulties faced by the dairy industry in reducing *S.uberis* mastitis prevalence in spite of the adoption of control measures for 'environmental' pathogens as has been suggested previously (Hillerton et al. 1993) . A more tailored, risk based approach using a combination of pathogen classification, analysis of mastitis patterns and individual herd management practices may be more successful.

The areas discussed in this section (3.3) are a more complete analysis of those described previously in (Davies et al. 2016).

Chapter 4: Analysis of clinical case classifications and associated cow level information.

4.1 Introduction

In Chapter 3, clinical case classification of *S.uberis* mastitis was carried out using MLST sequence data. The case classifications grouped S.uberis isolates as Index, Potential Transmission Events, Persistent, Solitary and Unclassified. The next objective was to develop an understanding of the epidemiology of those case classifications by making comparisons between groups in terms of readily available cow level data from monthly milk constituent records, parity and stage of lactation and the relationship between clinical cases over time in a herd. The following specific aims were identified to address this objective.

The aims were to conduct:

1. An evaluation between clinical case classifications of routinely recorded (NMR) milk sample data: somatic cell count; milk yield; protein concentration; fat concentration and lactose concentration.
2. An evaluation between clinical case classifications of cow level factors such as parity, seasonality and stage of lactation (days in milk, DIM).
3. An evaluation between clinical case classifications of the temporal relationship between successive clinical cases of the same case classification (the inter-case interval)

4. An evaluation of herd and cow level factors on clinical case classifications.

4.1.1 Analysis of National Milk Records (NMR) individual cow data

Analysis of cow level information, in particular monthly somatic cell count (SCC) records would potentially be useful because of the well-established correlation between SCC and the infectious status of the mammary gland (Harmon 1994). It is biologically plausible that the attributes of a more contagious *S.uberis* sub-group which may have the ability to establish persistent infections and the ability to evade or avoid an aggressive immune response, could be indicated by differences in the SCC pattern. Changes in the SCC pattern could also be the simplest and cheapest means of identifying quantifiable variation in the host-pathogen immune response for commercial dairy producers or other stakeholders within the industry such as diagnostic laboratories, veterinary advisors, milk processors or regulatory authorities.

SCC patterns before and after a clinical case have been shown to differ between pathogens which are associated with alternate transmission routes (Green et al. 2004; de Haas et al. 2004, 2005). The characteristic SCC pattern of “environmental” pathogens such as *E. coli* has been reported to be a rapid increase in SCC prior to the diagnosis followed by a rapid decline in SCC associated with clinical cure and removal of the pathogen (de Haas et al. 2004). In contrast, contagious pathogens such as *S.aureus* are understood to be better adapted to the mammary gland environment and are able colonise for a longer period. This was found to be reflected by a slower rise in SCC and

a lower peak SCC as well as a more prolonged elevation in SCC after initial clinical diagnosis (de Haas et al. 2004). This long pre-clinical or sub-clinical period of infection provides many more opportunities for pathogen transfer, via multiple daily milkings to uninfected mammary glands which share the same milking equipment or other common vectors such as the milking technicians' hands. However, *S.uberis* and *S.dysgalactiae* mastitis cases where more variable in their SCC responses some cases more akin to that of *E.coli*, other closer to the *S.aureus* pattern (de Haas et al. 2004). Previous studies have not differentiated between SCC patterns at the sub-species level. In this study we can make these comparisons based upon genotype or clinical case classification and a combination of both to investigate the heterogeneity of somatic cell count patterns to infection with *S.uberis*. It was unclear how or whether concentrations of other milk constituents (fat, protein and lactose) would vary with clinical case classifications; but since their data were also available an assessment was considered worthwhile.

4.1.2 Analysis of cow parity

Previous studies have indicated cows of parity three or above to be at increased risk of *S.uberis* clinical mastitis (Zadoks, et al. 2001) but it is not clear how this relative risk might be influenced by the transmission route of the pathogen. Several potential scenarios were envisaged that justified the examination of parity: potentially parity could change the susceptibility to infection due to age related immune modulation as has been demonstrated previously (Wojdak-Maksymiec, Szyda, and Strabel 2013) or changing

susceptibility due to underlying differences in the management of the parity groups, especially heifers; variation in management by parity could also expose groups to different infection pressures, resulting in different disease prevalence; alternatively variation by parity, if it occurs could be a covariate of other variables such as milk yield which vary with age.

4.1.3 Analysis of impact of stage of lactation on mastitis case classification

It is well established that mastitis cases identified in the first 30 days post calving are predominantly acquired during the dry period (Dry Period Origin) (Green et al. 2002) whereas those infections diagnosed after 30 days post calving are predominantly acquired during lactation (Lactation Period Origin). Is was reasonable to hypothesise that the subset of *S.uberis* sequence types that were implicated in contagious transmission (PTE & I) would be over represented during the period of lactation acquired infections, reflecting the additional contagious transmission behaviour of these sub-species which should only take place during lactation and not during the dry period.

4.1.4 Temporal Analysis of relationships between clinical cases by classification group (intercase interval)

Similarly it could be expected that the interval between successive clinical cases (intercase interval) that were the result of contagious transmission would be, on average, shorter than the interval between successive environment derived infections, when the whole study period is considered in spite of seasonal variation. For a contagious pathogen each additional

infective individual in the herd increases the risk of infection for the susceptible cows at each milking whereas the mastitis infection risk posed to a susceptible cow by an obligate environmental pathogen is unrelated to the number of infected herdmates (Down, et al, 2013) and as shown by (Zadoks, et al. 2001) in an outbreak of *S.uberis* mastitis. As the infectivity of a clinical mastitis case will depend upon the duration of infection the contagious transmission is only able to occur for a limited period, it has been shown that this period can range from <30days to 260days, mean 134days (Pullinger et al. 2007) in the absence of antibiotic treatment. In contrast no such limits exist for an environmental reservoir of infection. The intercase interval of environmental infection would be affected by environmental conditions such as seasonality, or management affecting the abundance of a diverse environmental population of opportunistic *S.uberis* (Lopez-Benavides et al. 2007; Steeneveld et al. 2009; Zadoks, Tikofsky, et al. 2005). In addition to variation in the intercase interval between clinical case classifications it was hypothesised that there would be less seasonality in the *S.uberis* mastitis incidence rate amongst the PTE mastitis cases compared to Solitary mastitis cases because seasonality should not affect contagious transmission that is occurring during the milking process.

4.2 Methods

4.2.1 Statistical Analysis

Statistical analysis was carried out using the software *Minitab17*. When the data distributions for the outcome variables did not conform to a normal distribution, non-parametric, Mann-Whitney tests were conducted. The

highly skewed nature of many of the parameters, in particular SCC, necessitated log transformation of the distributions to allow visual comparison between the classification groups and also analysis by ANOVA (where data were approximately normally distributed) to identify variation between categories.

Multivariable analysis was conducted in *MLwiN* (Rasbash et al, 2012) using a two level, linear regression model with clinical case as the lower level and farm as the second level. Days in milk (DIM) and Intercase interval (ICI) were tested as the outcome variables with the following explanatory variable groups: Case Classification; Parity and lagged milk parameters (SCC, Yield, Protein%, Fat%, Lactose%).

Multi-level mixed effects linear regression modelling was planned for analyis of lagged milk recording data to account of repeated measures with variable time intervals within lagged periods and substantial missing data. However, it was judged unnecessary following preliminary analyis of the data.

4.2.2 Clinical case classifications, refined from those define in Chapter 3

(I) Index case; The first clinical case isolate of a multi-isolate (PTE causing) sequence type in herd

(PTE) Potential Transmission Event; Following an index case in a herd, the subsequent clinical cases caused by that specific sequence type in another cow in that herd was classified as a 'Potential Transmission Event'(PTE).

(R) Persistent; Using the established definition described previously (Baseggio et al. 1997). When the same sequence type was identified on two or more occasions from the same mammary gland quarter all but the first case caused by that sequence type were classified as persistent infections.

(OVR PTE) Overrepresented Sequence Type Potential Transmission Event; A subset of the PTE category of mastitis cases including only those isolates of an OVR sequence type (ST5, ST6, ST20, ST22, ST24, ST35, ST233, ST361, ST512) see Chapter3.

(S) Solitary; Sequence types occurred in only one clinical case throughout the study period.

4.2.3 Data structure

Individual monthly milk recording data for all cows in each of the 52 herds were indexed according to recording date and Herd Animal ID (HAID) which uniquely identified each herd and each cow within each herd. Ten cows and associated clinical mastitis cases were excluded from the analysis because their records did not include the HAID number required to link their relevant monthly milk records. The remaining cows and all associated milk records were organised in an Excel spreadsheet format in numerical order by HAID and then by recording date in chronological order. Records were lagged into six time periods around the mastitis date. The time periods were calculated in 30 day blocks, 90 days (3 blocks) prior to the clinical case and 90 days (three blocks) after the case. The records were then lagged to allow analysis of

successive milk recordings (which occurred approximately monthly) against each mastitis case. Lagged milk records which related to a previous or subsequent lactation (ie. records with an intervening calving date) were excluded.

Table 6. The number of individual cow milk recordings falling within each of the six lagged time periods is displayed for each of the following clinical case classification: (Index), (PTE), (OVR PTE), Persistent (R), and Solitary (S). The day of mastitis sample collection is assigned Day Zero. Lagged period prior to the clinical sample collection date are shown in red. Lagged periods following the sample collection date are shown in black.

Lagged Data period (days)	Clinical Case Classification				
	Index (I)	Potential Transmission Event (PTE)	Overrepresented Sequence type Potential Transmission Event (OVR PTE)	Persistent (R)	Solitary (S)
90-60	36	51	25	9	39
59-30	32	52	40	11	40
29-0	36	51	40	9	40
1-30	41	51	42	15	43
31-60	49	51	47	9	50
61-90	39	43	34	7	44

Two further time periods were created for comparison of OVR PTE and Solitary classifications in the period 7 days prior to the case up to an including day of the mastitis sample collection (-7) and 7 days following the mastitis diagnosis day (+7). These two additional blocks allowed examination of milk constituent parameters that closely coincided with the recorded mastitis case. The limited time period reduced the number of available records in each

group and time period as follows: OVR PTE with a total of 71 isolates (-7) 12 records from 71 isolates, (+7) 16 records , Solitary (S) with a total of 124 isolates (-7) 24 records, (+7) 9 records.

The Clinical case classifications previously defined in Chapter 3 were used with the addition of a further group 'Overrepresented Sequence Type Potential Transmission Event' (OVR PTE) which was a subset of the PTE category of mastitis cases including only those isolates of an OVR sequence type (ST5, ST6, ST20, ST22, ST24, ST35, ST233, ST361, ST512).

4.3 Results

4.3.1 Aim 1: Evaluation between clinical case classifications of routinely recorded (NMR) milk sample data: somatic cell count; milk yield; protein concentration; fat concentration and lactose concentration.

4.3.1.1 Somatic Cell Count (SCC)

Arithmetic mean and median SCC values, calculated for each herd over the entire study period, using all individual cow SCC recordings (independent of individual milk yield) ranged from 106 to 337 thousand cells per ml and 31 to 160 thousand cells per ml respectively. Neither mean nor median SCC correlated with the percentage of mastitis clinical cases classified as either PTE or S when assessed graphically.

Graphical analysis of the lagged somatic cell count (SCC) patterns for each case classification over a 6 month period spanning the sample collection date did not identify any notable differences between case classifications or between lagged time points (Figure 8). A natural log transformation, normalising the data distributions confirmed the initial result (Figure 9).

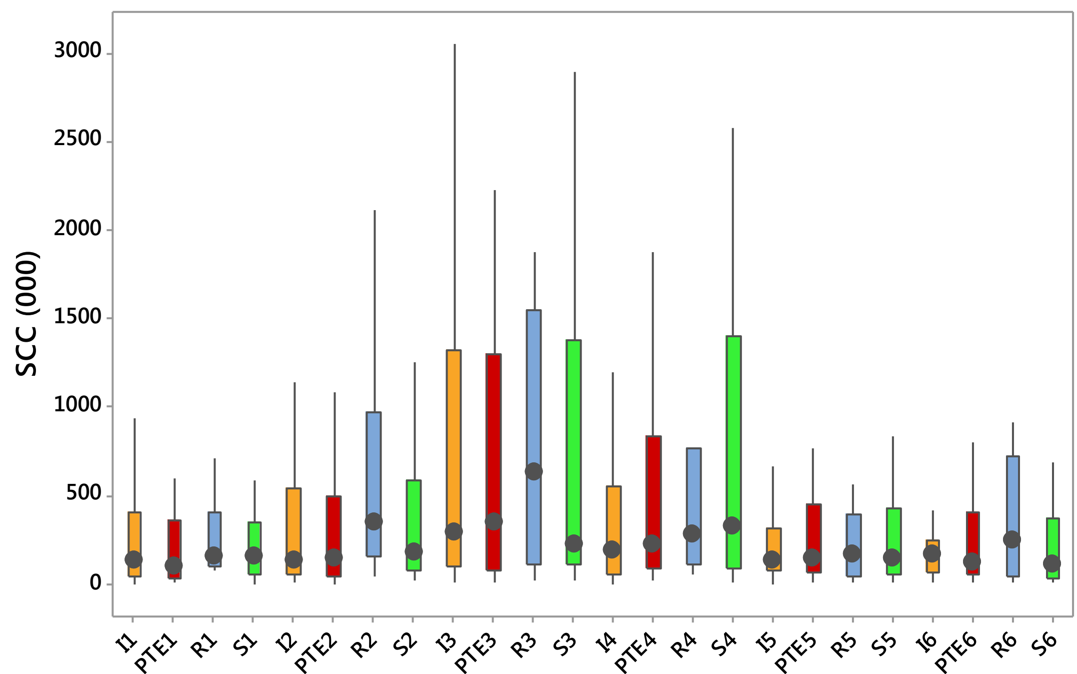

Figure 8. Somatic cell count (SCC) distributions of four clinical case classifications: Index (I, orange), Potential transmission event(PTE, red), Persistent (R, blue), and Solitary (S, green) over six, lagged time periods each of 30 days, three periods (months) before the recorded mastitis sample collection date and three following the recorded collection date. Notation of time periods as follows: 1 = 90-60 days prior to diagnosis of the mastitis case, 2 = 59-30 days prior to diagnosis of the mastitis case, 3 = 29 days prior and including the day of mastitis case, 4 = 1-30 days following the diagnosis of the mastitis case, 5 = 31-60 days following the diagnosis of the mastitis case, 6 = 61-90 days following the diagnosis of the mastitis case. Vertical black line = 95% confidence interval, coloured box = interquartile range, horizontal line = mean, black circle = median.

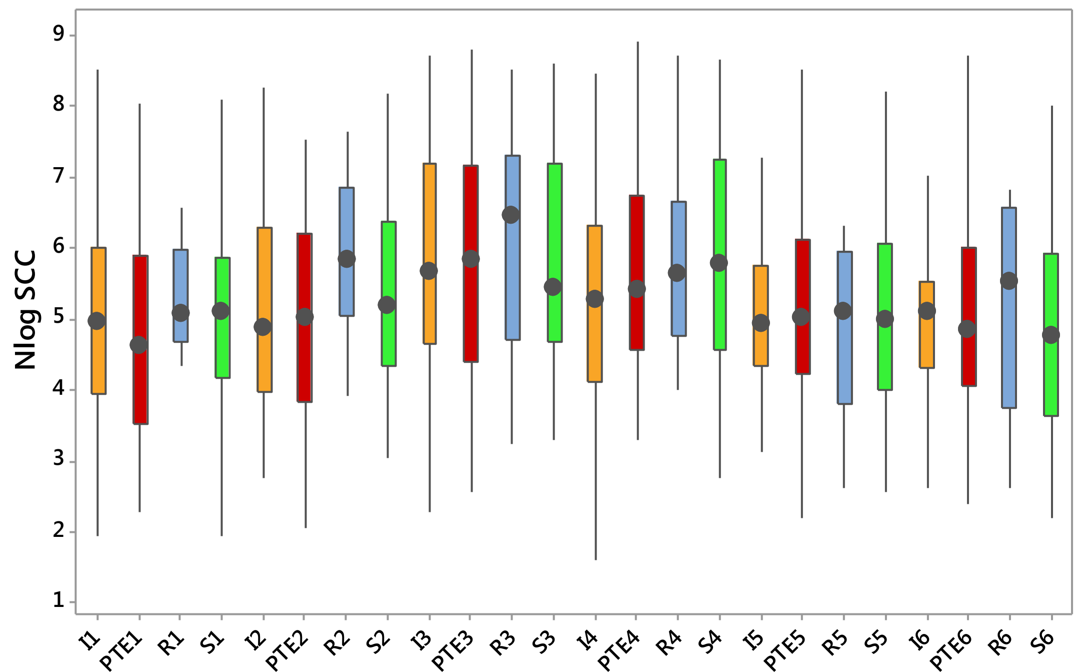

Figure 9. Somatic cell count (SCC) natural log transformed distributions of four clinical case classifications: Index (I, orange), Potential transmission event(PTE, red), Persistent (R, blue), and Solitary (S, green) over six, lagged time periods each of 30 days, three periods (months) before the recorded mastitis sample collection date and three following the recorded collection date. Notation of Time periods as follows: 1 = 90-60 days prior to diagnosis of the mastitis case, 2 = 59-30 days prior to diagnosis of the mastitis case, 3 = 29 days prior and including the day of mastitis case, 4 = 1-30 days following the diagnosis of the mastitis case, 5 = 31-60 days following the diagnosis of the mastitis case, 6 = 61-90 days following the diagnosis of the mastitis case. Vertical black line = 95% confidence interval, coloured box = interquartile range, horizontal line = mean, black circle = median.

A trend was observed in all four case classification groups for the SCC to increase in the two months (60 days) prior to the diagnosis of the clinical mastitis case. There was substantial variation in SCC between individual cows within each case classification at each time point resulting in a skewed distribution with a very long right tail of a small number of very high SCC values.

An equivalent (OVR PTE) and the Solitary (S) isolates, identified the same trend in SCC before and after the sample collection date but did not identify any notable differences between the two case classifications at any time point (Figure 10).

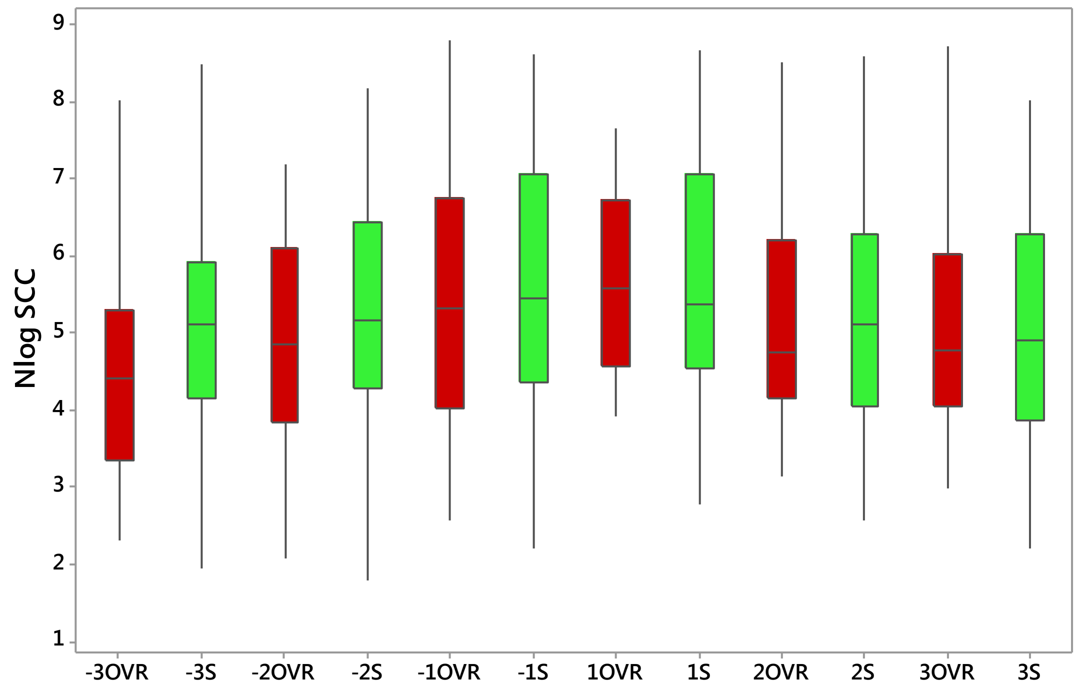

Figure 10. Somatic cell count (SCC) natural log transformed distributions of two clinical case classifications: Potential transmission event isolates of over represented MLST sequence types only (OVR PTE) = red and Solitary (S) = green) over six, lagged time periods each of 30 days, three periods (months) before the recorded mastitis sample collection date and three following the recorded collection date. Notation of time periods as follows: 1 = 90-60 days prior to diagnosis of the mastitis case, 2 = 59-30 days prior to diagnosis of the mastitis case, 3 = 29 days prior and including the day of mastitis case, 4 = 1-30 days following the diagnosis of the mastitis case, 5 = 31-60 days following the diagnosis of the mastitis case, 6 = 61-90 days following the diagnosis of the mastitis case. Vertical black line = 95% confidence interval, coloured box = interquartile range, horizontal line = mean.

The 7 day period immediately prior to a case of *S.uberis* clinical mastitis and following the case diagnosis were examined in more detail as a distribution of SCC records and as the natural log of the distribution. Again, no differences were identified between the two case classifications in the week prior to the mastitis case diagnosis or after it (Figure 11).

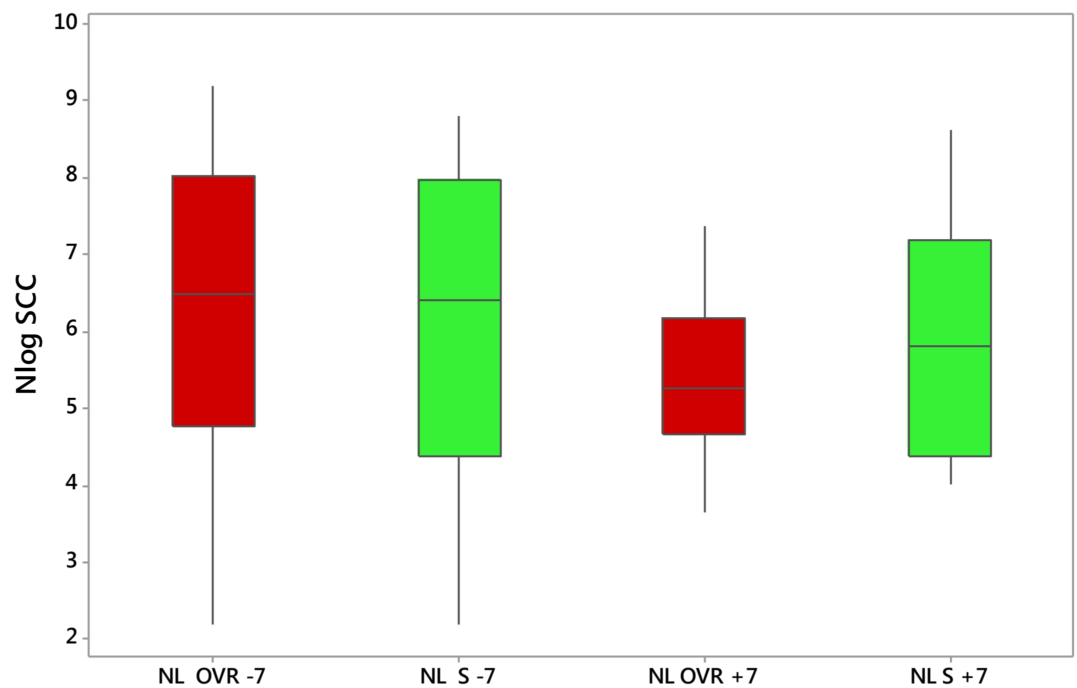

Figure 11. Somatic cell count (SCC) distributions; (OVR PTE) = red and Solitary (S) = green over two, lagged time periods each of 7 days. The first period covers the 7 days prior to and including the sample collection/mastitis diagnosis date (-7). The second period covers the 7 days following the sample collection/mastitis diagnosis date (+7). Vertical black line = 95% confidence interval, coloured box = interquartile range, horizontal line = mean.

4.3.1.2 Milk Yield

Milk yield on the day of recording was compared using the same group structure of OVR PTE and Solitary (S) case classifications as described above

(4.2.1). There were no significant differences observed between these two case classifications in any time lagged period (Figure 12).

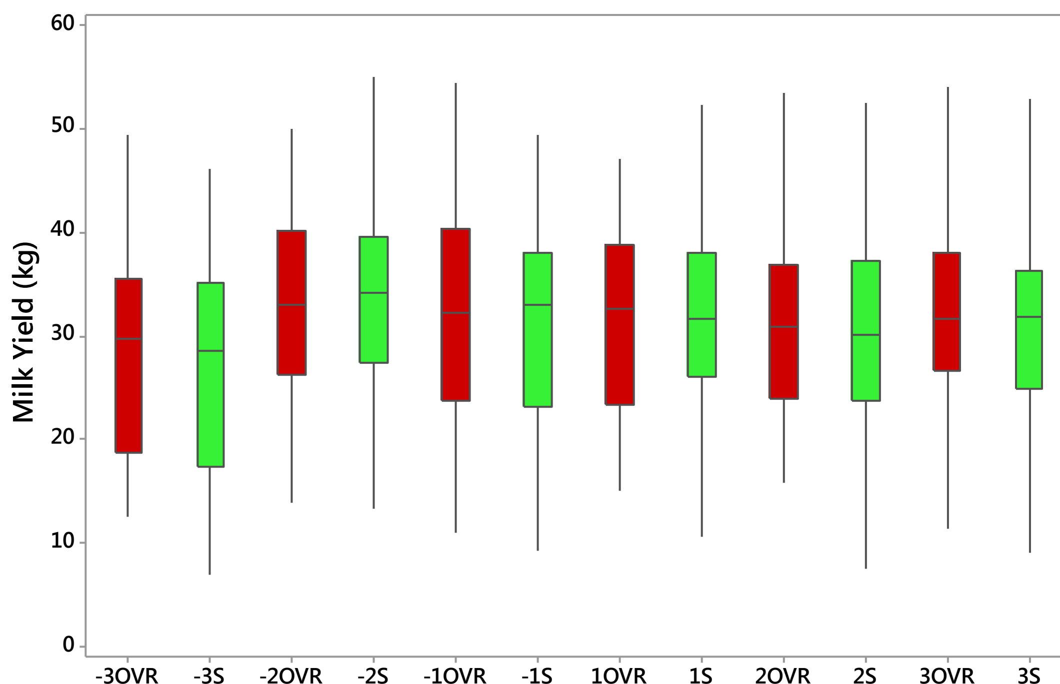

Figure 12. Distributions of recorded milk yield; (OVR PTE) = red and Solitary (S) = green, over six, lagged time periods each of 30 days, three periods (months) before the recorded mastitis sample collection date and three following the recorded collection date. Notation of Time periods as follows: 1 = 90-60 days prior to diagnosis of the mastitis case, 2 = 59-30 days prior to diagnosis of the mastitis case, 3 = 29 days prior and including the day of mastitis case, 4 = 1-30 days following the diagnosis of the mastitis case, 5 = 31-60 days following the diagnosis of the mastitis case, 6 = 61-90 days following the diagnosis of the mastitis case. Vertical black line = 95% confidence interval, coloured box = interquartile range, horizontal line = mean.

4.3.1.3 Milk Protein

Milk protein concentration on the day of recording was compared using the same group structure of OVR PTE and Solitary (S) case classifications as

described above (1.2.2). There were no notable differences observed between these two case classifications in any time lagged period (Figure 13).

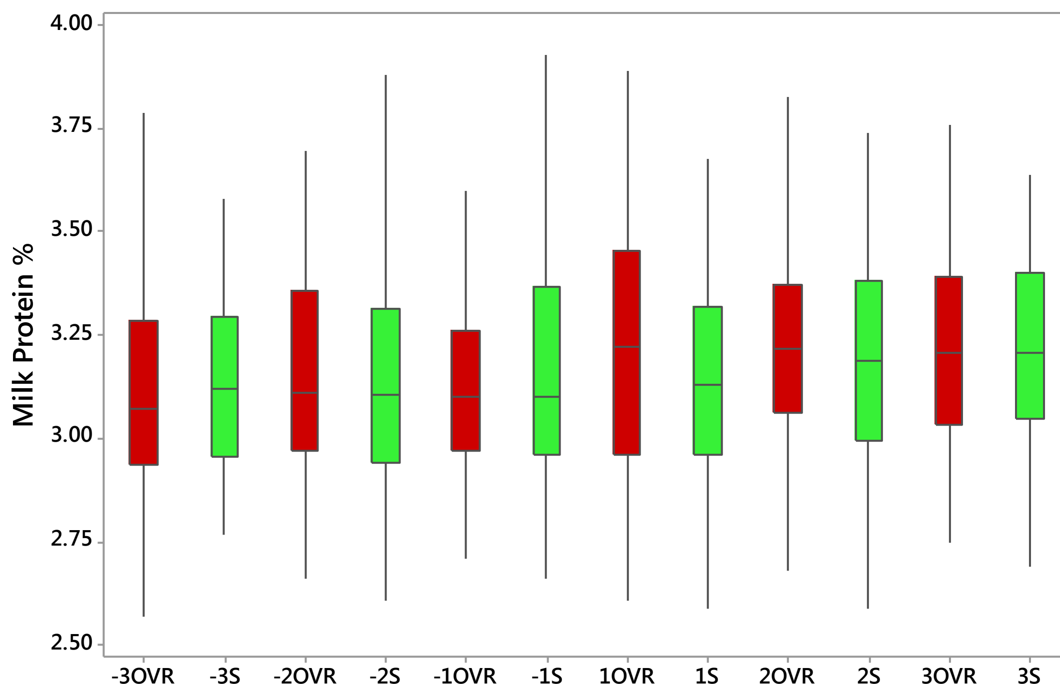

Figure 13. Distributions of recorded milk protein concentration; (OVR PTE) = red and Solitary (S) = green, over six, lagged time periods each of 30 days, three periods (months) before the recorded mastitis sample collection date and three following the recorded collection date. Notation of Time periods as follows: 1 = 90-60 days prior to diagnosis of the mastitis case, 2 = 59-30 days prior to diagnosis of the mastitis case, 3 = 29 days prior and including the day of mastitis case, 4 = 1-30 days following the diagnosis of the mastitis case, 5 = 31-60 days following the diagnosis of the mastitis case, 6 = 61-90 days following the diagnosis of the mastitis case. Vertical black line = 95% confidence interval, coloured box = interquartile range, horizontal line = mean.

4.3.1.4 Milk Fat

Milk fat concentration on the day of recording was compared using the same group structure of OVR PTE and Solitary (S) case classifications as described

above (1.2.2). There were no notable differences observed between these two case classifications in any time lagged period (Figure 14).

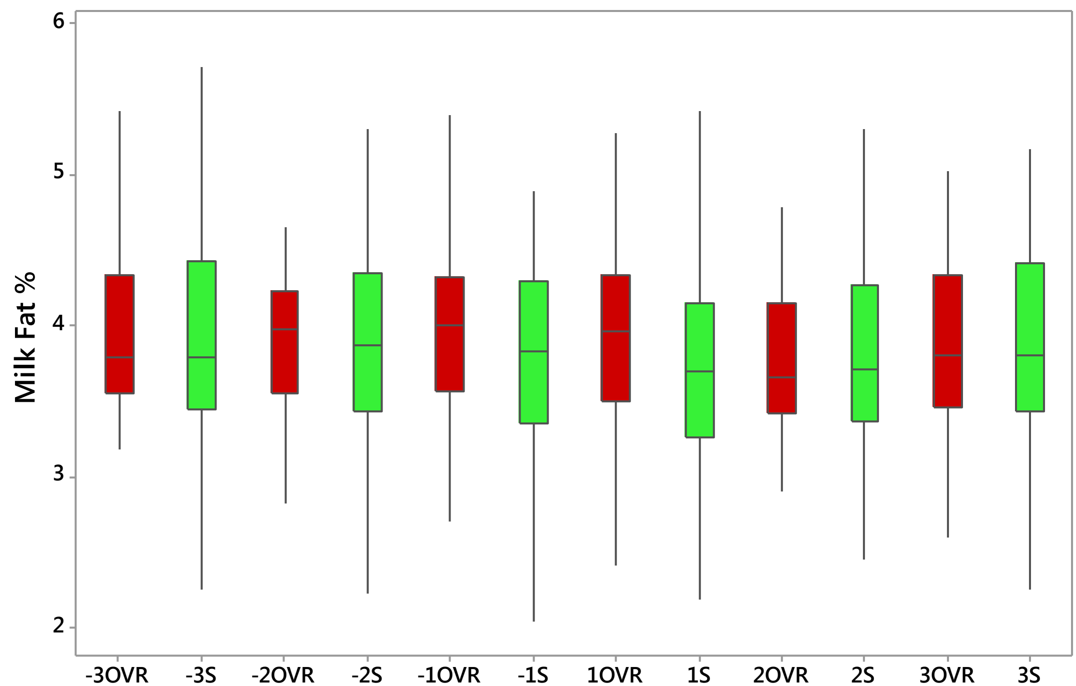

Figure 14. Distributions of recorded milk fat concentration; (OVR PTE) = red and Solitary (S) = green, over six, lagged time periods each of 30 days, three periods (months) before the recorded mastitis sample collection date and three following the recorded collection date. Notation of Time periods as follows: 1 = 90-60 days prior to diagnosis of the mastitis case, 2 = 59-30 days prior to diagnosis of the mastitis case, 3 = 29 days prior and including the day of mastitis case, 4 = 1-30 days following the diagnosis of the mastitis case, 5 = 31-60 days following the diagnosis of the mastitis case, 6 = 61-90 days following the diagnosis of the mastitis case. Vertical black line = 95% confidence interval, coloured box = interquartile range, horizontal line = mean.

4.3.1.5 Milk Lactose

Milk lactose concentration on the day of recording was compared using the same group structure of OVR PTE and Solitary (S) case classifications as

described above (1.2.2). There were no notable differences observed between these two case classifications in any time lagged period (Figure 15).

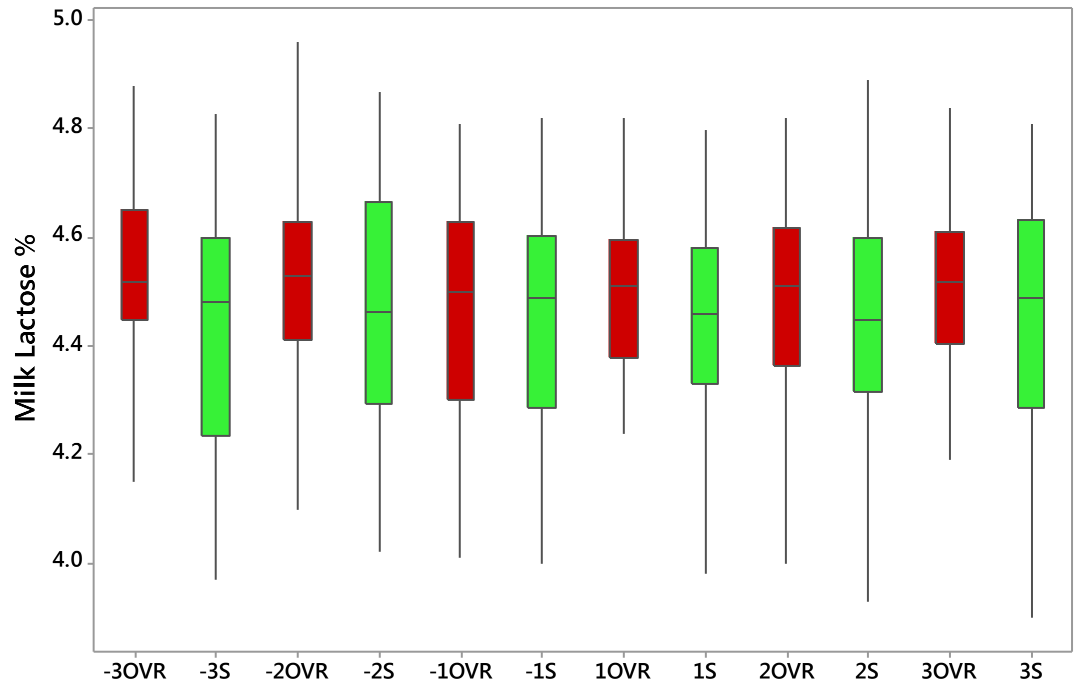

Figure 15. Distributions of recorded milk lactose concentration; (OVR PTE) = red and Solitary (S) = green, over six, lagged time periods each of 30 days, three periods (months) before the recorded mastitis sample collection date and three following the recorded collection date. Notation of Time periods as follows: 1 = 90-60 days prior to diagnosis of the mastitis case, 2 = 59-30 days prior to diagnosis of the mastitis case, 3 = 29 days prior and including the day of mastitis case, 4 = 1-30 days following the diagnosis of the mastitis case, 5 = 31-60 days following the diagnosis of the mastitis case, 6 = 61-90 days following the diagnosis of the mastitis case. Vertical black line = 95% confidence interval, coloured box = interquartile range, horizontal line = mean.

4.3.2 Aim 2: An evaluation between clinical case classifications of cow level factors such as parity, seasonality and stage of lactation (days in milk, DIM).

4.3.2.1 Parity

Cow parity at the time of the mastitis case diagnosis for clinical case classifications; Solitary (S), Persistent (R), Potential Transmission event (PTE) and its subgroup, (OVR PTE) did not identify any significant differences between classifications with mean parity of 4 for animals experiencing a clinical case of any classification. Parity four was also the mode and median parity for the population of cows with diagnosed *S.uberis* mastitis.

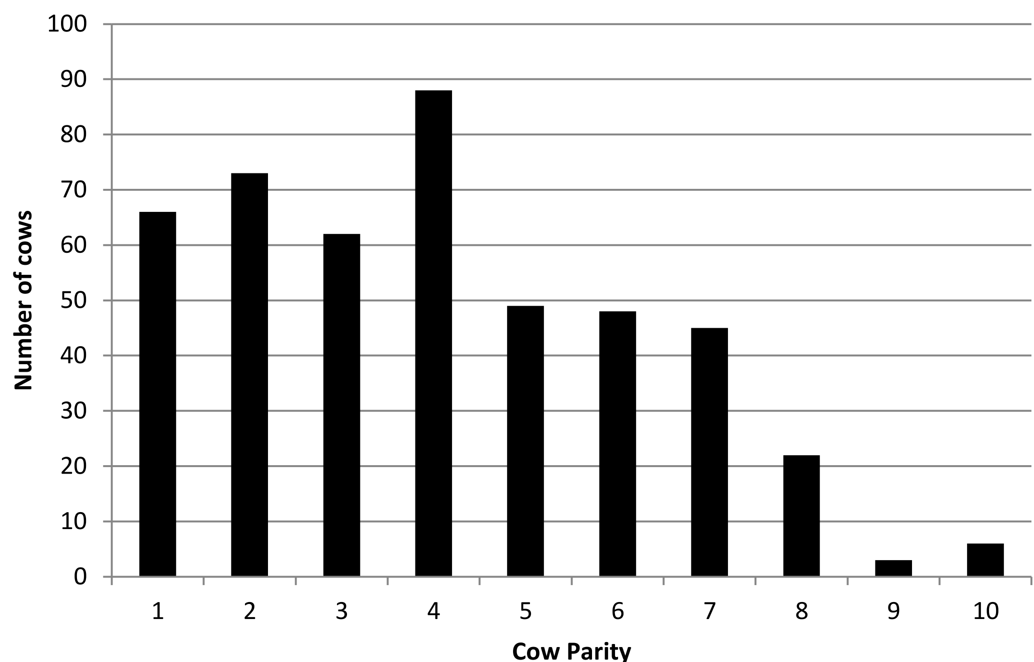

Figure 16. Frequency of cows by parity of all cows from which *S.uberis* MLST profiles were obtained from their clinical mastitis case. Mean = 4.1 median = 4 and mode parity = 4.

4.3.2.2 Seasonal Prevalence of S.uberis mastitis by Case Classification

A clear seasonal trend was identified, mastitis incidence increased from the nadir in June to a winter peak, due to both; potentially contagious (PTE & I) and suspected environmental (S) case classifications from June to December (Figure 17). From January to April the number of PTE associated clinical cases (PTE & I) diagnosed was consistently higher than the number attributed to the (S) classification. In May there is a dramatic peak in the Solitary mastitis cases.

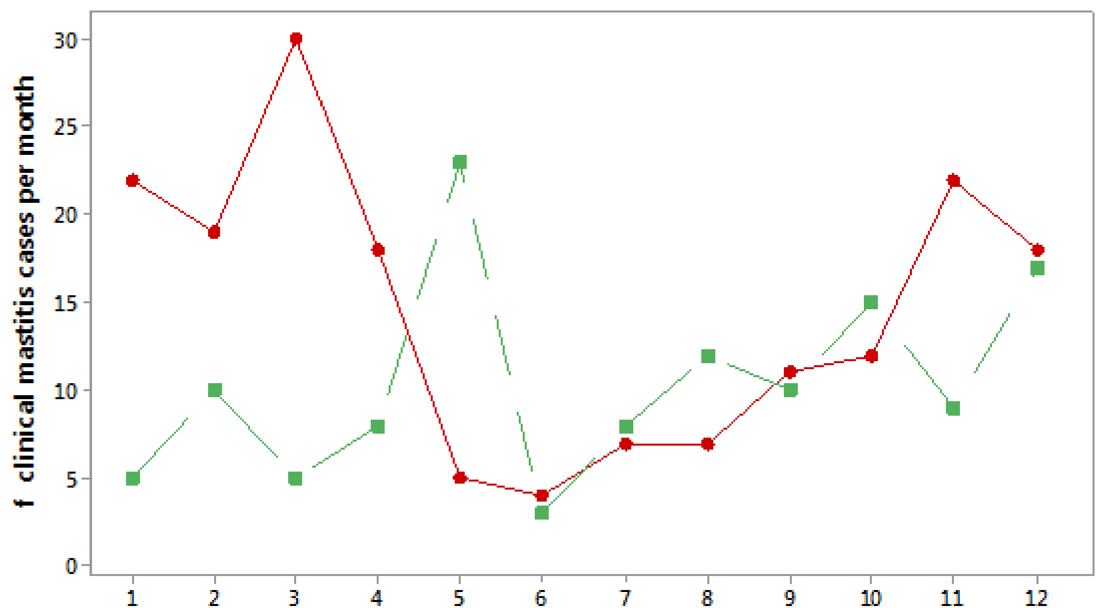

Figure 17. Seasonal frequency distribution of clinical mastitis cases by classification group. January = 1, December = 12. Red line represents the potentially contagious case classifications (PTE and Index) while the Green line represents the Solitary case classifications.

4.3.2.3 Stage of lactation

Clinical cases of *S.uberis* mastitis occurred throughout lactation. However, the distribution of cases throughout lactation differed between clinical case

classifications. The potentially contagious clinical cases, classified as potential transmission events (PTE), occurred significantly later in lactation than the Solitary (S) isolates (p=0.0095 Mann-Whitney Test) with median lactation weeks of 17.0 for PTE cases compared to 12.5 for Solitary cases. The difference was even greater when only the OVR PTE cases were compared to the Solitary isolates medians 17.5 vs 12.5 lactation weeks respectively (p=0.0058 Mann-Whitney Test) Figure 18.

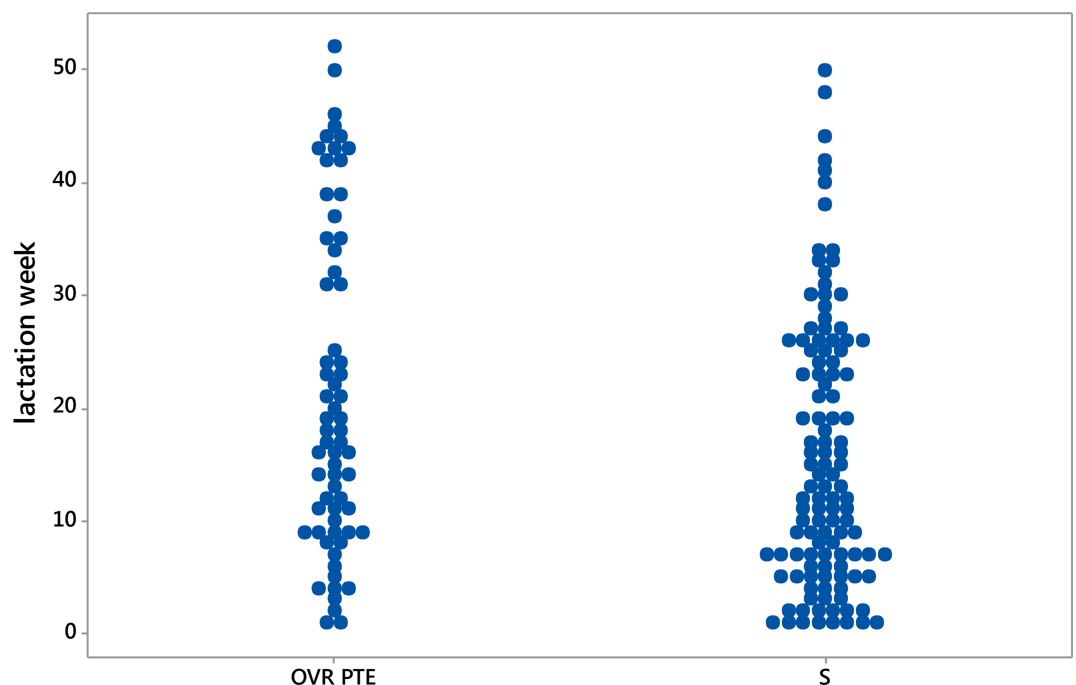

Figure 18. Individual value plot of *S.uberis* mastitis case diagnosis relative to stage of lactation (lactation week) against the clinical case classifications: potential transmission events due to over represented sequence types (OVR PTE) and Solitary (S) isolates.

When lactation weeks were aggregated into two blocks; Dry Period Origin (DPO = mastitis date less than 4 weeks after calving) and Lactation Period Origin (LPO = mastitis date more than 4 weeks after calving), a significantly

smaller proportion of PTE isolate mastitis cases were classified as of dry period origin (10%) compared to the Solitary isolates (18%) (p=0.042, Chi-squared Test).

4.3.3 Aim 3: An evaluation of the temporal relationship between successive clinical cases of the same case classification (the inter-case interval)

Intercase intervals calculated between each successive clinical case within each herd ranged from zero weeks to fifty weeks. A higher proportion of intercase intervals between succesive Index (I) and potential transmission event (PTE) cases (of the same MLST sequence type) were ≤8weeks compared to the Solitary (S) clinical cases (Figure 19).

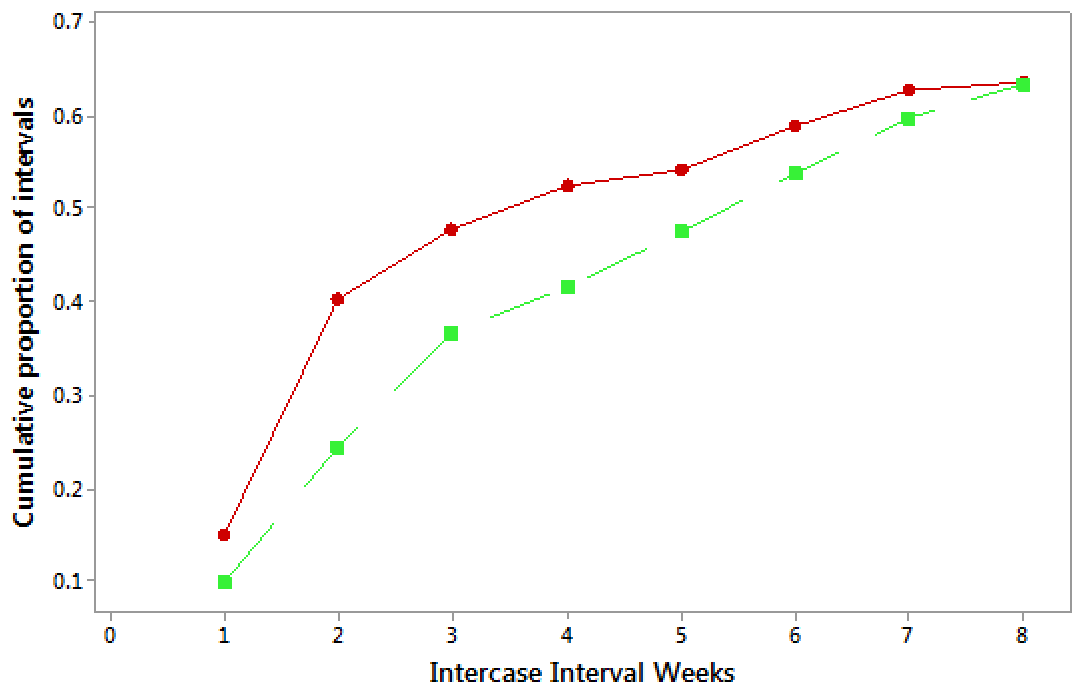

Figure 19. Cumulative proportion plot of all intercase intervals less than 8 weeks for clinical case classifications associated with: The Red line and circles = Index and PTE classifications and the Green line and squares = Solitary classification.

In the 8 weeks following a potentially contagious mastitis case (Index (I), Potential transmission event (PTE)) or a Solitary (S) case, the intercase intervals between of the potentially contagious mastitis cases were significantly shorter (p=0.0097, Mann-Whitney Test) than the intercase intervals between mastitis cases which were classified as Solitary (S) isolates. The mean intercase interval of the potentially contagious group was 1.8 weeks (12.6 days) for 68 intercase intervals while the mean intercase interval between solitary isolates during the same period was 2.7 weeks (18.8 days) for 52 intervals.

4.3.4 Aim 4: An evaluation of herd and cow factors on clinical case classifications.

The case classifications were used as independent variables in multivariable analysis in a series of multi-level models of milk recording records and cow variables which were considered individually and in combination. The base level was clinical case and the upper level was farm. No significant differences were identified between case classifications for any of the following outcome variables; somatic cell count, milk yield, milk protein, milk fat, milk lactose concentrations. Similarly no significant effect was identified for case classification on parity, quarter or stage of lactation. However, a significantly shorter intercase intervals between potentially contagious clinical cases was identified using the model with a coefficient of -0.91 (SE=0.37) indicating that on average, the combined Index and PTE group had an intercase interval 0.9 weeks (6.3 days) less than the Solitary group.

4.4 Discussion

4.4.1 Aim 1: An evaluation between clinical case classifications of routinely recorded (NMR) milk sample data: somatic cell count; milk yield; protein concentration; fat concentration and lactose concentration.

No such significant associations were identified between SCC, yield, fat, protein or lactose and the case classification of *S.uberis* clinical mastitis cases. Of particular interest was the somatic cell count of cows that experienced potentially contagious *S.uberis* infections, because of the role of the innate immune response, reflected by the SCC, in combatting infectious pathogens and the evidence of differential patterns in SCC between pathogens associated with either contagious transmission or environmental transmission (de Haas et al. 2004). It was hypothesised that these species level differences in SCC may be replicated at the sub-species level in *S.uberis* between 'contagious' and 'environmental' sub-groups. However, at a herd SCC level there was apparent correlation between the proportion of cases assigned to the alternative classifications and at the individual cow level the SCC distributions before and after a potentially contagious (PTE/I) clinical case were indistinguishable from the distribution of the most likely 'environmentally' acquired infections which were classified as Solitary (S) isolates. This similarity could be explained by there being no distinguishable or detectable difference in the cell mediated immune response to these two subgroups of *S.uberis* when a case of clinical mastitis occurs. Alternatively there may be real immunological differences which are not detectable in this data for a number of reasons The first of these possible reasons is that the

milk records were based on samples collected on a monthly basis which may be too infrequent to provide the necessary data granularity to identify subtle changes in the SCC around the time of infection or development of clinical signs. The second possible reason could be that the inherent high variability in SCC between individual cows reduced the statistical power such that a larger sample size would be required to allow detection of relatively small significant differences between the case classifications. Thirdly, the method of milk recording involves pooling of all four quarters into a single sample whereas the clinical mastitis case and the immune response to it are largely restricted to one quarter. Therefore it is possible that this pooling masks the real differences that may be detectable in individualised quarter samples. Finally, there is the possibility that real immunological differences exist but these differences do not manifest themselves as changes in the SCC but instead are manifested as changes in the activation of proinflammatory cytokines such as TNF-α or IL-1β (Almeida and Oliver 2006; Oviedo-Boyso et al. 2007) by the innate immune response. However, the purpose of this project was to attempt to characterise the modes of transmission and identify means of diagnosing transmission behaviour in a commercial farm setting. At this point in time it does not appear that the established monthly milk recording parameters are in themselves useful in the discrimination of isolates according to transmission route.

A robust comparison of the SCC values between *S.uberis* mastitis cases and mastitis cases of other, 'contagious' or 'environmental' pathogens, such as

S.aureus and *E.coli* was not possible due to the very small number of *S.aureus* cases identified during the study (76).

4.4.2 Aim 2: An evaluation between clinical case classifications of cow level factors such as parity, seasonality and stage of lactation (days in milk, DIM).

Whilst there did not appear to be an effect of parity based on the univariate or multivariate analysis there was an intriguing seasonality pattern in the data. The seasonal increase in mastitis incidence from mid-summer to mid-winter might be expected due to increasing infectious pressure associated with housing on organic bedding (Lopez-Benavides et al. 2007; Steeneveld et al. 2009). The prolonged high incidence rate of the contagious associated case classifications (PTE & I) during the late winter and early spring could be the result of contagious transmission of infections originally acquired from the environment in the preceding months which are now perpetuated through cow-cow transmission rather than being reliant on reinfection from the environment for every new case. Theoretically an obligate mammary pathogen which is transmitted exclusively in a cow to cow contagious manner during milking should not exhibit much seasonality in disease incidence as it is not dependent on environmental conditions or environmental/management factors for its infectivity. The potentially contagious PTE & I case classifications did exhibit some seasonality but arguably less seasonality than the Solitary (S) classification group which is presumed to be principally an environmental opportunist. The reduced seasonality of the potentially contagious *S.uberis* isolates supports the hypothesis described in the previous

Chapter of 'Low grade' transmission where an infection may be acquired originally from the environment, transmitted via the milking equipment in a contagious manner to a limited number of cows before dying out. In this scenario the contagious *S.uberis* sub-groups which possess the limited ability to establish, persist and transmit to another cow, are maintained by non-mammary reservoirs in the environment or gastrointestinal tract and emerge initially as opportunistic environmental and then transmit in a contagious manor when and for as long as favourable conditions for that contagious transmission allow.

The Solitary (S) clinical cases which were assumed to be opportunistic, environment infections were significantly more likely to occur in the first month after calving than the potentially contagious (PTE/I) clinical cases. Given that those intramammary infections manifesting as clinical mastitis, occurring in the first 30 days post calving are more likely to have been acquired during the dry period (Green et al. 2002) the fact that some of the PTE and Index cases occur during this period suggests that these potentially contagious MLST sequence types are present in the environment and are capable of transmitting in the classic opportunistic fashion associated with *S.uberis*. However, the fact they were significantly more likely to occur in during lactation implies that the potentially contagious MLST sequence types are better adapted for colonisation during this period. Either the potentially contagious sequence types were transmitted from cow to cow via milking, thereby taking advantage of an additional infection opportunity not present

during the dry period, or these sequence types are better adapted to colonise the lactating udder directly from the environment than their Solitary (S) counterparts.

4.4.3 Aim 3: An evaluation of the temporal relationship between successive clinical cases of the same case classification (the inter-case interval)

The temporal pattern of clinical cases showed that PTE cases occurred significantly closer together in time than successive Solitary cases during the first eight weeks after a clinical case of that category. The intercase interval is a potentially important indicator of transmission behaviour because the majority of intramammary infections have a limited window of opportunity during which they can transmit an infectious dose of bacteria to another host before they are cleared by the immune system. In contrast, environmental opportunist mastitis pathogens are maintained and replicate outside the udder and are not subject to the immune pressure that contagious pathogens are under. It would therefore be anticipated that contagious transmission events involving the same MLST sequence types between two different cows would occur, on average, in closer succession than two independent, opportunistic infections from the environment. Previous studies have identified herds with outbreaks of *S.agalactiae* (an obligate contagious pathogen) and outbreaks of *S.uberis* with similar Rø values 'the average number of secondary infections resulting from the introduction of one infectious individual into a fully susceptible population' (Leelahapongsathon et al. 2015) indicating similar transmissibility in spite of a shorter duration of

infection with *S.uberis* (Leelahapongsathon et al. 2015). This could be explained by a mixed pattern of *S.uberis* transmission from cow to cow but also from the environment.

S.uberis is a very diverse and adaptable bacteria, the classification system used in this project will inevitably misclassify some clinical cases in each category. This means many of the patterns we would expect to clearly differentiate transmission route in pathogens that are either obligate, host adapted pathogens or opportunistic environmental pathogens do not have the same clarity in the case of *S.uberis*. However, this analysis has identified several observations that support the hypothesis of cow to cow transmission of a small minority of Sequence types but it is probable that even these sequence types are also acquired sporadically from the environment thereby blurring the observed temporal patterns and complicating the transmission dynamics of the pathogen.

Chapter 5: MALDI phenotype compared to clinical case classification

5.1 Introduction

In Chapter 3, epidemiological patterns consistent with contagious transmission of *S.uberis* between cows was identified in 33 out of the 52 study herds. The technique used, multilocus sequence typing, is a time consuming and relatively expensive method for diagnostic purposes. In clinical practice accurate identification of contagious transmission would be very useful but would need to be conducted in a timely manner to allow corrective management changes to be implemented. Therefore, a fast and relatively cheap procedure is needed to differentiate strains of *S.uberis* according to likelihood of contagious behaviour.

The objective of this Chapter was to investigate an alternative method to classify cases of *S.uberis* clinical mastitis which would be more practical for rapid throughput diagnostics. The method chosen, matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI) is a fast throughput technology which has been demonstrated to be capable of sub-species typing in *Leptospira, Yersinia* and *Mycoplasma* (Rettinger et al. 2012; Rizzardi et al. 2013; Xiao et al. 2014a). Since it is biologically plausible that the bacterial mechanisms which facilitate contagious transmission would be associated with the expression of specific proteins, MALDI charged molecule

(mostly protein) mass spectrometry could provide a viable alternative method to classify sub-species of *S.uberis*.

Therefore, to evaluate the extent to which MALDI spectral profiles aligned with MLST profiles and epidemiological classifications of *S.uberis*, the following aims were addressed:

1. To determine if MLST genotype correlated with distinguishable MALDI phenotypes
2. To determine if the epidemiological case classifications described in Chapter 3 could be differentiated using MALDI
3. To determine if MALDI spectral profiles could be used to predict epidemiological case classifications of isolates of *S.uberis* in a new herd
4. To determine the between farm variation in MALDI phenotype of Solitary and PTE isolates.

5.2 Methods

5.2.1 Sample selection

In total 426 of the 494 isolates with MLST sequence types were phenotyped using a Bruker Biotyper3.0 MALDI system. Samples were categorised according to their farm of origin and their case classification (I, PTE, P, S and U) definitions described in Chapter 3.

5.2.1.1 *Aim 1: To determine if MLST genotype correlated with distinguishable MALDI phenotypes*

To address the first aim sample isolates of each ST were grouped into seven distinct categories according to the number of locus variants of the most prevalent sequence type, ST6 (6, 5, 4, 3, 2, 1, 0 common alleles with ST6) see Table 7. Initial analysis was conducted by pairwise comparison of the two most discriminatory peaks between the ST6 group and each of the seven common allele groups. A simplified class structure of 4 groups comprised of amalgamated locus variant (lv) classes (ST6&1lv, 2lv&3lv, 4lv&5lv, 6lv&7lv (see Figure 27) was used to facilitate principal component analysis by reducing disparity in isolate number between the largest and smallest groups from a multiple of 5.9 to 2.3. The simplification of the group structure reduced potential statistical bias within the PCA model towards the groups with more isolates.

The groups were then amalgamated into two larger groups for analysis to reduce statistical bias in model generation using *ClinProTools3.0* due to divergent relative group sizes. The first class (CC6) represented a clonal complex defined by each isolate within the complex sharing at least four common alleles with the founder ST6. The second group (Non-CC6) were all remaining isolates which shared fewer than four alleles with ST6 or with any other isolate. This grouping resembled the overall MLST population structure described previously in Chapter 3. The CC6 group comprised 84 sequence types including 24 of the 35 PTE causing sequence types and 8 of the 9 over

represented sequence types (OVR). Within the CC6 group there were isolates of different case classifications (I, PTE, P, U and S). The CC6 and Non-CC6 groups were compared using Supervised Neural Networks (SNN) and Genetic Algorithm (GA) model algorithms.

Finally, pairwise analysis of all isolates of three individual sequence types was conducted using (ST6 , ST5 and ST20). ST5 was selected as it was the most numerous single MLST locus variant of ST6 which allowed comparison of the MALDI discriminatory ability between very similar isolates. ST6 and ST20 have five MLST locus variants and were chosen to compare very the MALDI discriminatory ability of very dissimilar sequence types. Fourteen isolates were selected for each sequence type. Equal numbers of isolates were selected for each ST to remove statistical model bias. The maximum number of fourteen was dictated by the sequence type group with the fewest MALDI isolates (ST20). The *Minitab* random number generator function was used to select the 14 isolates from the ST6 and ST5 populations. Ten isolates were selected randomly using the same method from each sequence type for model generation and internal validation and the remaining four were used for external validation. All six ST groups; for model training and external validation contained isolates from more than one herd.

Table 7. Frequency of isolates per MALDI analysis group used of model generation in ClinProTools.

Group	Number of isolates
Sequence Type 6	40
1 locus variant ST6	73
2 locus variants ST6	95
3 locus variants ST6	74
4 locus variants ST6	27
5 locus variants ST6	45
6 locus variants ST6	56
7 locus variants ST6	16
Clonal complex 6 (CC6)	282
Non-CC6	144

5.2.1.2 Aim 2: To determine if the epidemiological case classifications described in Chapter 3 could be differentiated using MALDI

To address the second aim isolates were selected based upon their case classification as defined previously in Chapter 3. The first group was comprised of all isolates of OVR sequence types (ST5,6,20,22,24,35,233,361&512) which had been classified as either an index case, potential transmission event or persistent case (All OVR PTE (I,PTE,P)) . This group was then compared to a group of all isolates classified as Solitary clinical cases (All Solitary). The (All OVR PTE (I,PTE,P)) group was then refined in line with the results of modelling results discussed in Chapter 4 to include only those isolates with an inter case interval of less than 8 weeks (OVR PTE <8ICI wks). This refined grouping was intended to increase the likelihood that isolates within the new group were the result of contagious transmission rather than environmentally acquired infections coincidentally of an OVR

sequence type. This group was then compared to a randomised sample of Solitary isolates, ensuring each group contained an equal number of isolates, see Table 8.

Table 8. Number of isolates in each case classification group used for MALDI analysis and model generation in ClinProTools.

Group	Number of isolates
All OVR PTE (I,PTE,P)	102
OVR PTE <8week Intercase interval	36
All Solitary	109
Solitary Randomised	36

The OVR PTE group was largely a refinement of the CC6 group with the inclusion of only one sequence type, ST20, from the Non-CC6 group. The Solitary group was comprised isolates from both the CC6 group (38.5%) and the Non-CC6 group (61.5%).

5.2.1.3 Aim 3: To determine if MALDI spectral profiles could be used to predict epidemiological case classifications of isolates of S.uberis in a new herd

To assess the predictive ability of models using the MALDI phenotype five herds were selected based upon the criteria that a minimum of 12 isolate spectra classified as PTE and Solitary were available, see Table 9. The discriminatory ability of the MALDI for each farm was assessed individually using SNN and GA model algorithms. The farms were then amalgamated to create multi-farm models using 4 herds and the classification test accuracy validated with the PTE and Solitary isolates from the fifth herd.

Table 9. Number of MALDI spectra of PTE and Solitary case classification used for SNN and GA models generated for each of the five farms.

Farm of origin	Number of PTE spectra	Number of Solitary spectra
8	29	13
20	25	19
27	67	34
55	18	16
65	13	23

5.2.1.4 Aim 4: To determine the between farm variation in MALDI phenotype of Solitary and PTE isolates.

The Solitary isolate populations from Farm 8, Farm20, Farm55 and Farm65 were compared as separate groups within the same GA and SNN model. Farm 27 was excluded because the larger number of isolates in that herd compared to the others increased the risk of statistical bias within the models. The same process was carried out on the PTE isolate populations in a separate model. PCA and two dimensional best peak analysis was conducted on the isolates from each model. Finally two dimensional best peak analysis was conducted using all five herds, including Farm27 to further illustrate the between farm variability in the PTE isolate populations.

5.2.2 Statistical Analysis

Analysis was conducted in *ClinProTools3.0* using principal component analysis (PCA), two dimensional most discriminatory peak analysis and two alternative statistical model algorithms; Supervised Neural Network modelling (SNN) and Genetic Algorithm (GA). The fit of both model algorithms were assessed using

a combination of two parameters; recognition capacity and internal cross validation which are described in section 0 and 5.2.2.7. The model with the highest combined percentage success over these two parameters was then adopted and the model further tested by external validation where suitable data was available.

5.2.2.1 2D Peak Distribution

This is the most basic form of group comparison which is useful in this analysis to demonstrate substantial variation between groups. The two dimensional most discriminatory peak analysis shows the first and second peaks as x and y axis and plots all the isolates with their respective group colour against them. The ellipses represent the 95% confidence interval of the class average of the peak areas/intensities. The scales are arbitrary units assigned for appropriate visual scaling of the plot.

5.2.2.2 Principal component analysis (PCA)

PCA analysis was used to visualise the variation between the groups. The PCA information also forms an integral part of the hierarchical unsupervised clustering analysis. The analysis combines multiple dimensions of data into a small number of components that encompass the majority of the variance within the dataset. Mapping of these components against each other allows labelled groups of individual isolates to be compared and ranked against each other. This is more useful for examining more completely the relationship and separation between groups using the maximum number of spectral features.

5.2.2.3 Model generation

Clinprotools3.0 Biotyper MALDI analysis software provided several model classification algorithms for groups of spectral profiles. In this analysis two alternative machine learning algorithms were compared; Genetic Algorithm (GA) and Supervised Neural Network (SNN) as described below:

5.2.2.4 Genetic Algorithm (GA)

In the GA model combinations of spectra peaks which perform best in separating the classes are selected by pattern determination to find an optimal set of spectral peaks, rather than using a systematic trial of all combinations. Pairs of peaks were ranked according to their discriminatory characteristics between the classes and are then a proportion are then recombined randomly with other pairs in a crossover analogous to genetic recombination to produce the next generation of peak combinations. Strongly discriminatory pairs are maintained while weak ones are discarded. The model recombined peak pairs over 50 generations or until no better combinations are identified. A 'mutation' factor is also added whereby a proportion of the combinations are modified to maintain diversity and prevent premature convergence upon a local optimum within the spectrum. The overall result should be an increase in the rank scores of the combinations that are most fit.

The model parameters were set with a maximal number of peaks set at 5. The mutation rate was set at 0.2 and the crossover rate at 0.5 and the K-number

near neighbours was set at 3. The K-nearest Neighbour classification is used within the GA model to determine the final classification by comparing the distance between a data point of a given class and the class of their 3 nearest neighbour data points.

5.2.2.5 Supervised Neural Network (SNN)

SNN models use a prototype-based classification algorithm system where a subset of spectra was identified as prototypical of their class. The whole data region is then divided up into areas associated with each of these prototypes. Each non-prototype data point is then located in a region assigned to a prototype and that non-prototype data point will then be assigned to the class of that prototype. The selection of the prototypes is therefore fundamental to the ability of the model to correctly assign new data points to the correct class. The data regions are created from pairs of discriminatory peaks plotted against each other. The prototypes are initially distributed across the whole data region according to the data point density. The positions of the prototypes are then optimised with respect to the classes in the model. Finally the model runs multiple iterations to select the most discriminatory prototypes. The number and position of peaks were selected automatically in the range (1-25), prototypes were detected automatically and optimisation used 5000 iteration cycles as an upper limit.

Model fit was assessed by the following methods:

5.2.2.6 Recognition capacity

The recognition capacity describes the performance of the classification algorithm. It is calculated for a determined model as “the relative number of correctly classified data points by the classifier for the given model under the constraint that all the tested data is previously used for the determination of the model or training of the classifier” *Clinprotools3.0 manual*.

5.2.2.7 Internal Cross Validation

A random subset of 20% of data points were taken from all groups and omitted from the model generation procedure. The model was calculated with the remaining data points and the random set of data points classified against the model. The process was repeated 10 times and the averaged classification results provided the prediction capacity of the model. In a model comprising two classes a cross validation of 50% is equivalent to that achieved by random assignment.

5.2.2.8 External Cross Validation

Additional isolate spectra not used in the generation or training of the model but being of the same group definition as used in the model are classified against the generated model to test the predictive accuracy of the model. In this study external cross validation was recorded as Test Accuracy (Number of correctly classified spectra divided by total number of spectra).

5.3 Results

5.3.1 Aim 1: MALDI Phenotype of MLST sequence types

To address the first aim pairwise comparison was made between the ST6 group and each group of isolates classified by number of identical alleles with ST6. Due to the wide disparity in group sizes and the small number of isolates in several of the seven locus variant groups a PCA, SNN or GA model would have been unreliable so analysis of those classes was restricted to pairwise analysis of each class with ST6 of the two most discriminatory peaks see Figure 20-26.

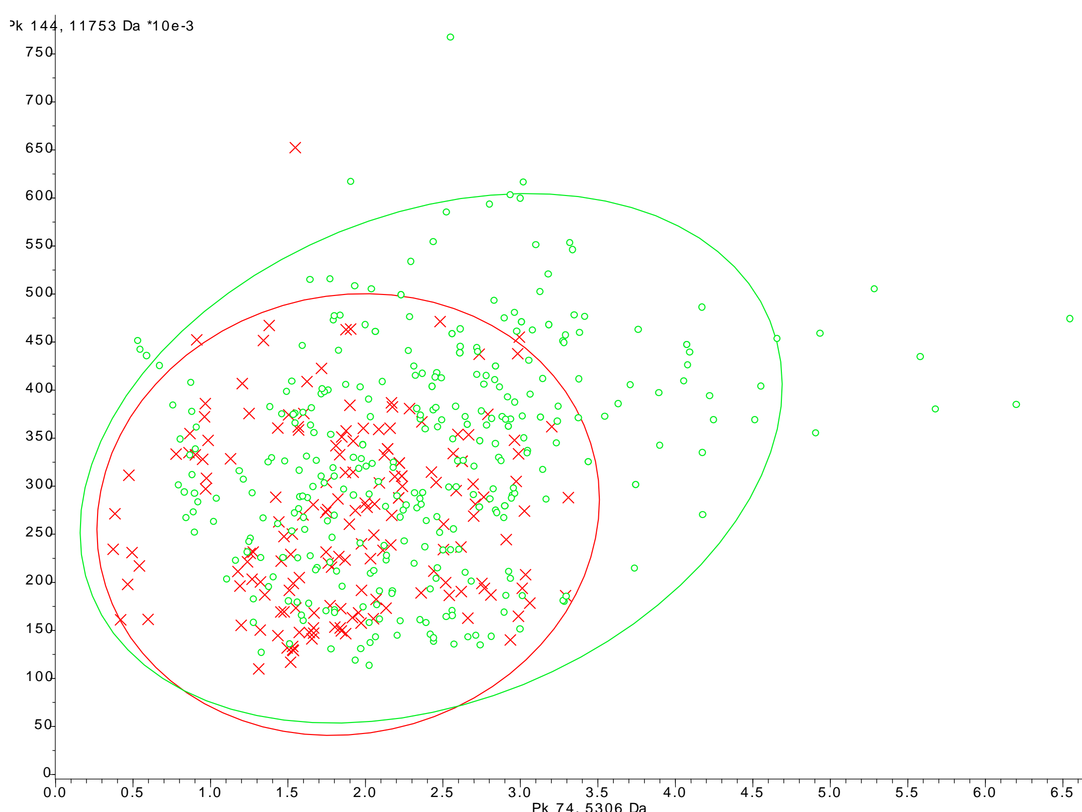

Figure 20. ST6 (red) compared to single locus variants of ST6 by two dimensional plot of the first two most discriminatory spectral peaks (first peak on X axis, second peak on y axis). Peak weight in Da is indicated on each axis and signal intensity is indicated in arbitrary units defined automatically by ClinProTools on each axis.

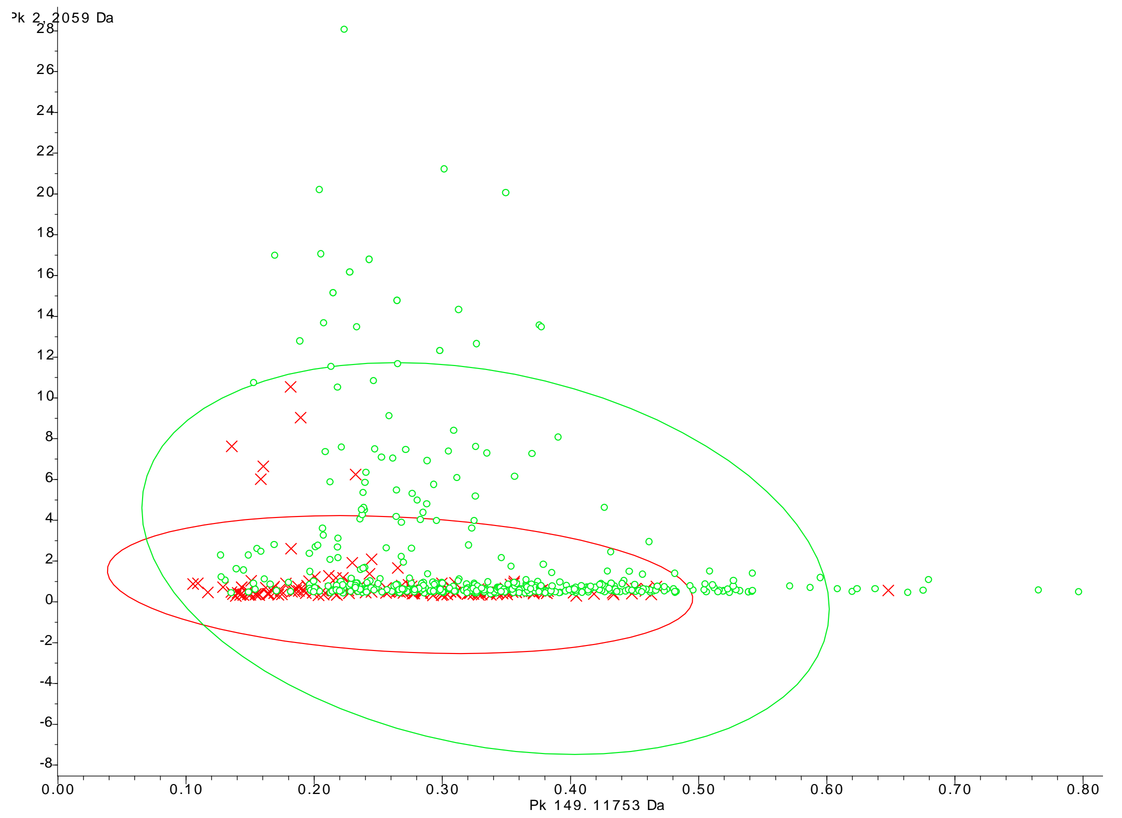

Figure 21. ST6 (red) compared to double locus variants of ST6 by two dimensional plot of the first two most discriminatory spectral peaks (first peak on X axis, second peak on y axis). Peak weight in Da is indicated on each axis and signal intensity is indicated in arbitrary units defined automatically by ClinProTools on each axis.

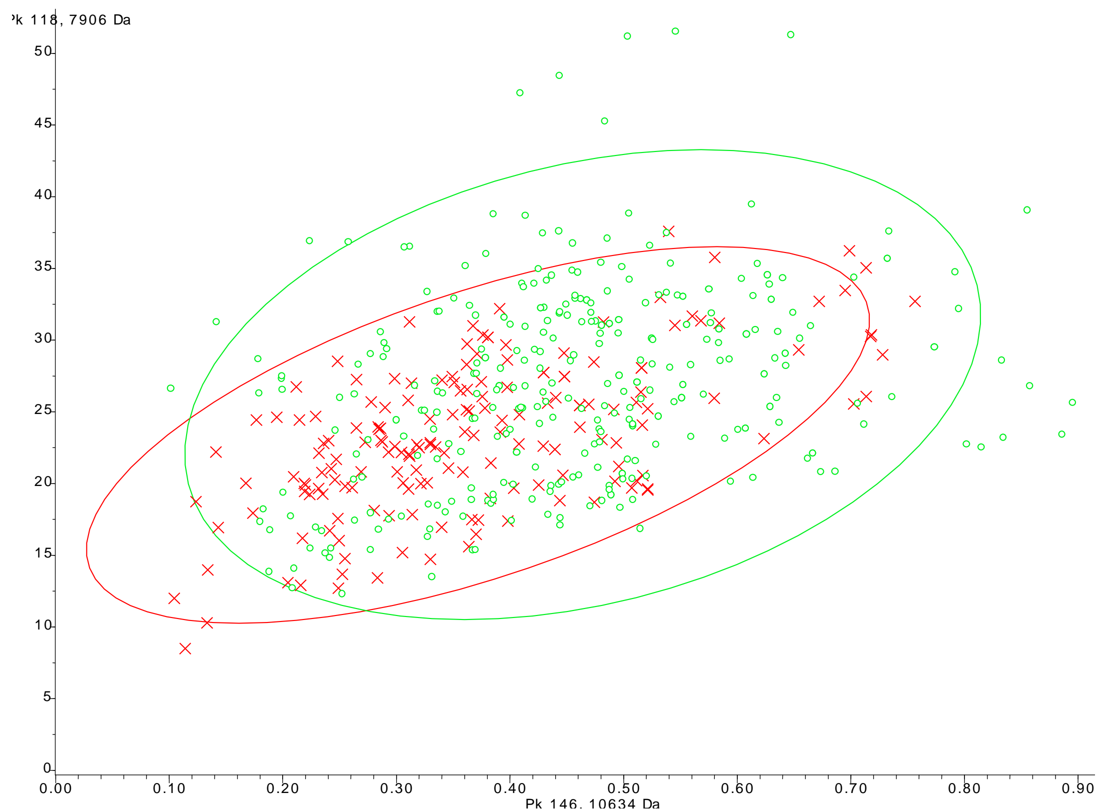

Figure 22. ST6 (red) compared to triple locus variants of ST6 by two dimensional plot of the first two most discriminatory spectral peaks (first peak on X axis, second peak on y axis). Peak weight in Da is indicated on each axis and signal intensity is indicated in arbitrary units defined automatically by ClinProTools on each axis.

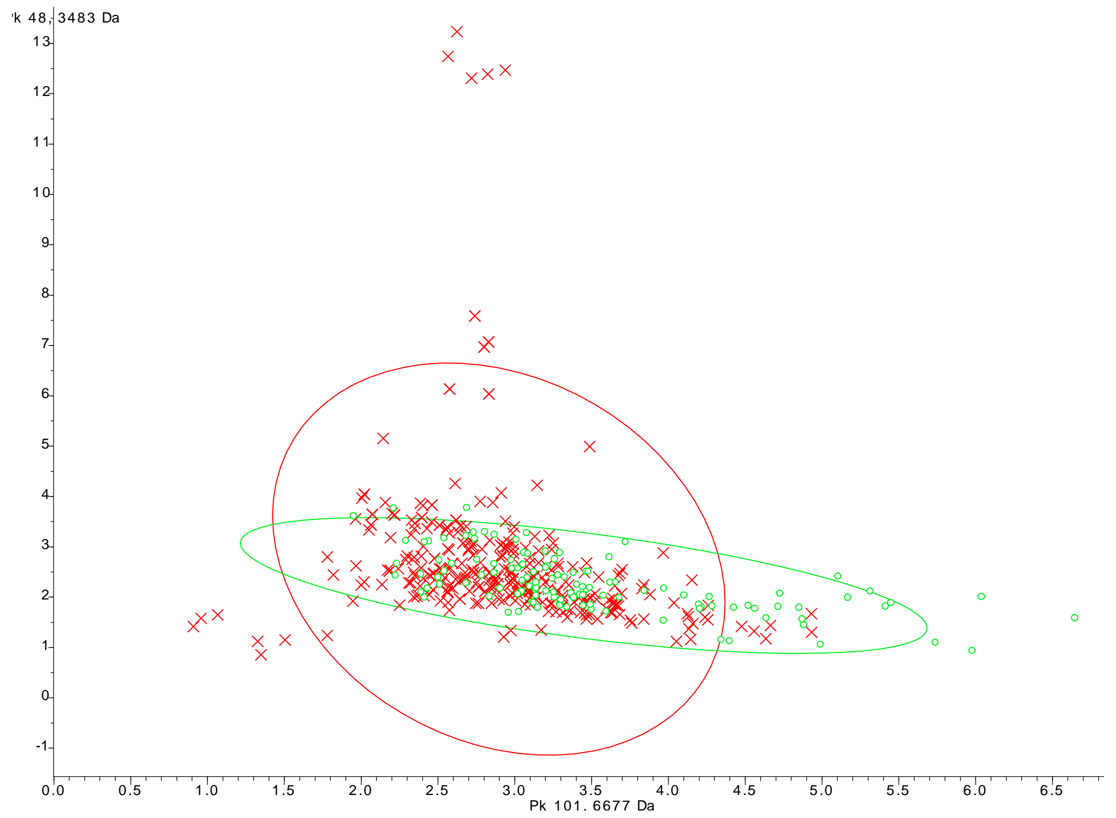

Figure 23. ST6 (red) compared to four locus variants of ST6 by two dimensional plot of the first two most discriminatory spectral peaks (first peak on X axis, second peak on y axis). Peak weight in Da is indicated on each axis and signal intensity is indicated in arbitrary units defined automatically by ClinProTools on each axis.

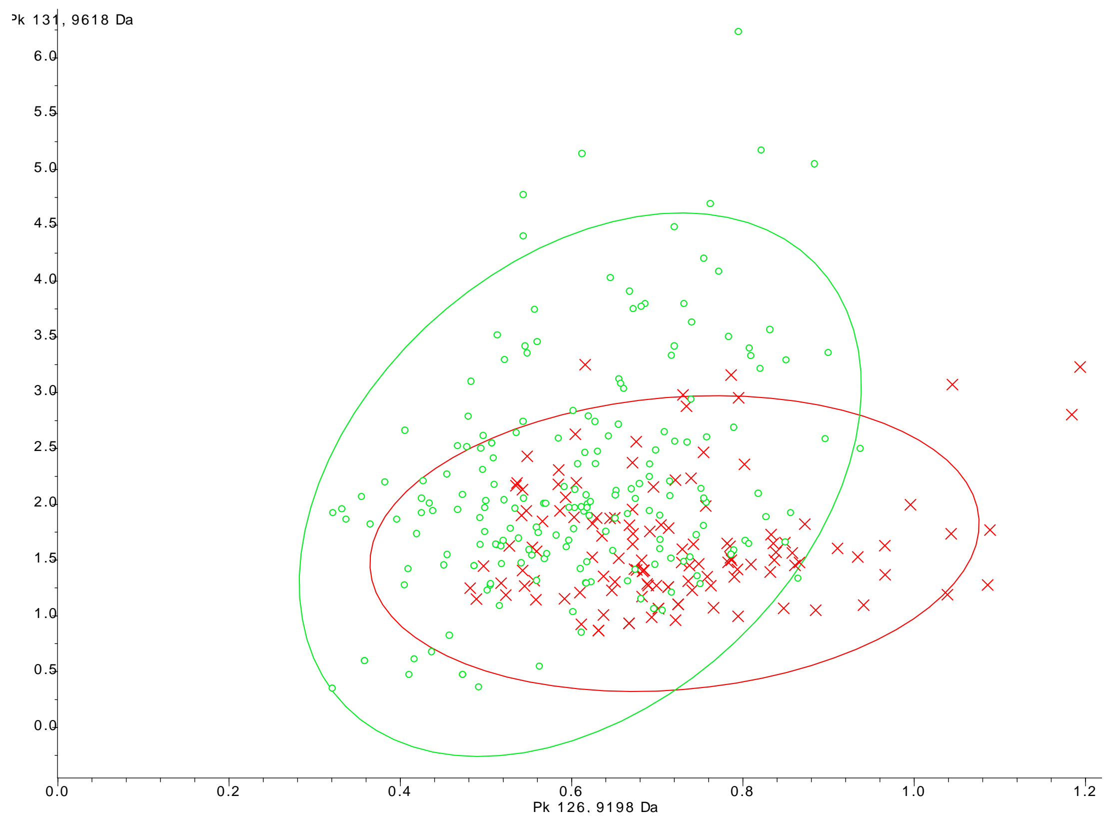

Figure 24. ST6 (red) compared to five locus variants of ST6 by two dimensional plot of the first two most discriminatory spectral peaks (first peak on X axis, second peak on y axis). Peak weight in Da is indicated on each axis and signal intensity is indicated in arbitrary units defined automatically by ClinProTools on each axis.

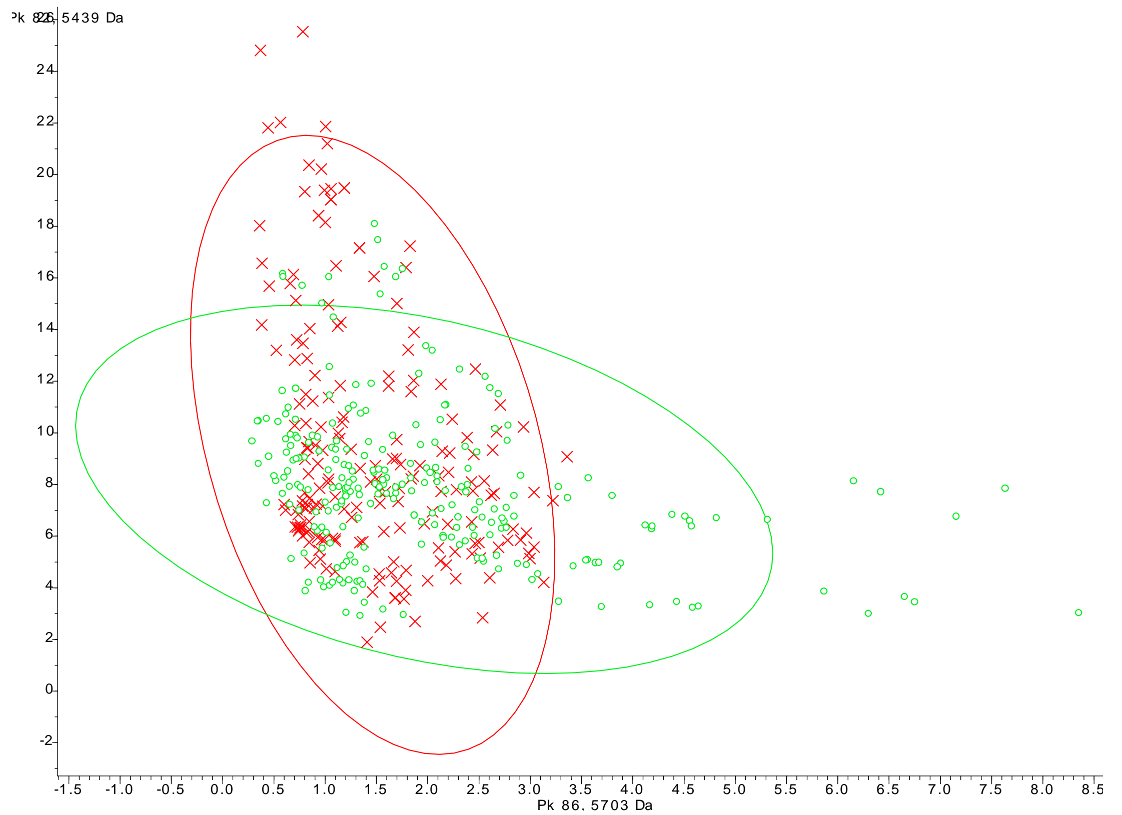

Figure 25. ST6 (red) compared to six locus variants of ST6 by two dimensional plot of the first two most discriminatory spectral peaks (first peak on X axis, second peak on y axis). Peak weight in Da is indicated on each axis and signal intensity is indicated in arbitrary units defined automatically by ClinProTools on each axis.

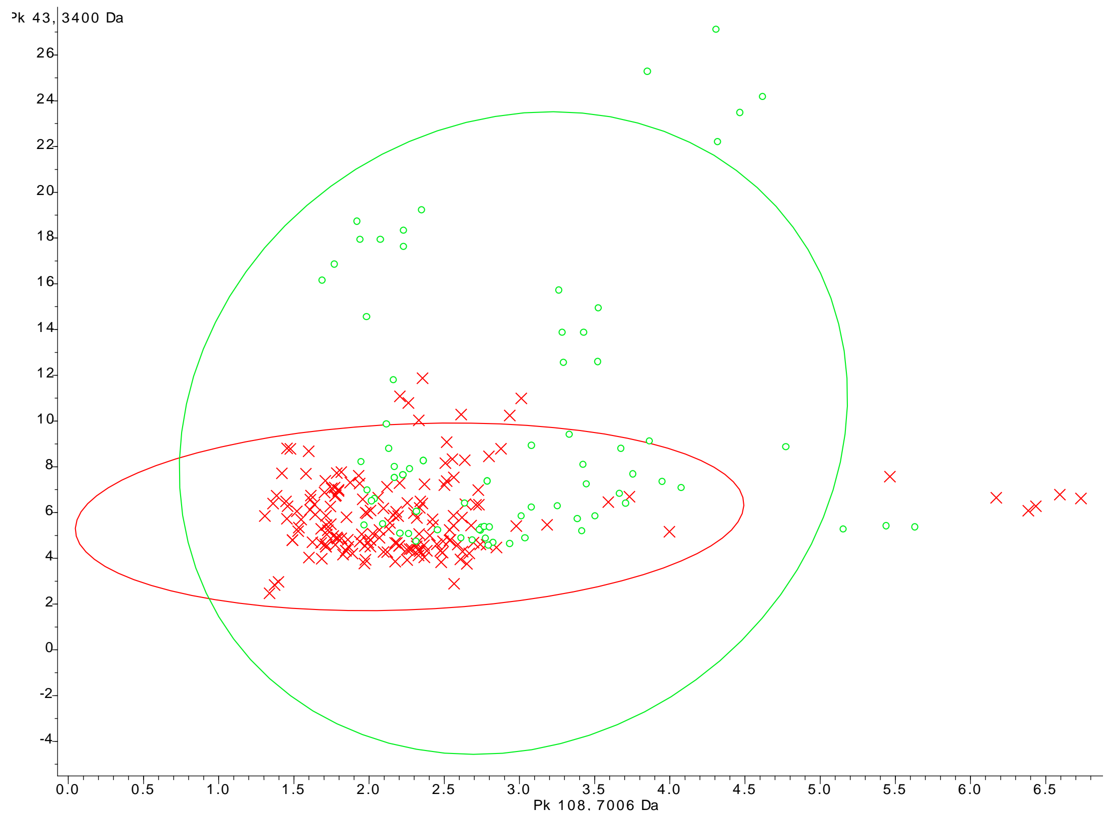

Figure 26. ST6 (red) compared to seven locus variants of ST6 by two dimensional plot of the first two most discriminatory spectral peaks (first peak on X axis, second peak on y axis). Peak weight in Da is indicated on each axis and signal intensity is indicated in arbitrary units defined automatically by ClinProTools on each axis.

Pairwise analysis of each group did not reveal a discernible trend between MLST sequence type similarity and best discriminatory peak values.

Principal component analysis was carried out using a simplified class structure of 4 classes comprised of amalgamated locus variant (lv) classes (ST6&1lv, 2lv&3lv, 4lv&5lv, 6lv&7lv) see Figure 27. All groups were heavily represented in the same region with outlier members of each class located along the extremities of each principal component plane.

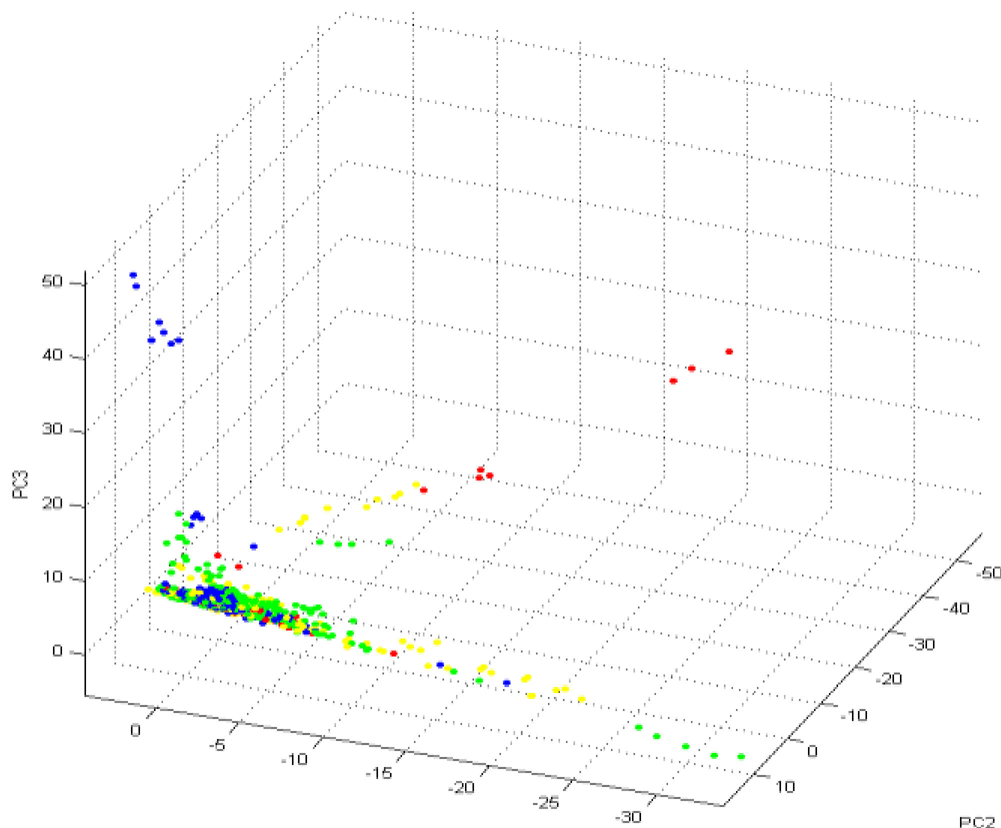

Figure 27. PCA analysis of MALDI spectral profiles for all *S.uberis* isolates grouped by their MLST sequence types grouped according to the number of locus variants (lv) from ST6. ST6&1lv = Red, 2lv&3lv = Green, 4lv&5lv = Blue, 6lv&7lv =Yellow.

To investigate the MALDI phenotype variation between the two largest groupings all isolates were assigned to either CC6 or Non-CC6. These were compared using a SNN and a GA model (Table 10). Model fit was quantified by recognition capability and internal cross validation as described in section 5.2.2.7. Model fit was moderate for classification of isolates based upon MLST sequence type information alone in both models.

Table 10. SNN and GA models comparing isolates of CC6 to Non-CC6 isolates testing the discriminatory capacity of MALDI in ClinProTools.

Model Name	Supervised Neural Network		Genetic Algorithm	
	Cross	Recognition	Cross	Recognition

	Validation	Capability	Validation	Capability
CC6 v Non-CC6	64.1%	64.0%	68.6%	74.1%

To test the discriminatory ability of MALDI phenotype for individual MLST sequence types of differing similarity, rather than clonal complexes, ten isolates were randomly selected from three of the most prevalent sequence types (ST5, ST6 and ST20). ST 5 is a single locus variant of ST6 while ST20 shares only two common loci with ST6. An additional four isolates of each ST group were used for external validation of the model (Table 11). The individual sequence types investigated were successfully classified by the both model algorithms with superior model fit compared to the clonal complex group models. The external validation test accuracy was substantially higher in both the SNN and GA models for the ST6 – ST20 (genetically diverse) classification compared to the ST6 – ST5 (genetically similar) classification models.

Table 11. Pairwise comparison of discriminatory ability based on MLST group using multiple isolates of two MLST Sequence Types each as its own group by their MALDI spectral profiles using SNN and GA model classification algorithms in ClinProTools. Pairwise comparison was made of ST5 to ST6 as representatives of very similar sequence types and ST6 vs ST20 as representatives of very dissimilar sequence types.

Pairwise comparison of individual MLST sequence types	Number of non-identical MLST Loci	SNN			GA		
		Cross Validation	Recognition Capability	Test Accuracy	Cross Validation	Recognition Capability	Test Accuracy
ST6 vs ST5	1	82.3%	93.6%	46.0%	86.9%	97.3%	49.0%
ST6 vs ST20	5	88.8%	97.6%	82.0%	91.7%	99.2%	84.0%

5.3.2 Aim 2: The Discriminatory ability of clinical case classifications by MALDI Phenotype

To address the second aim, the two most important case classifications, OVR PTE associated cases and Solitary isolates were compared. Analysis was conducted simultaneously in a SNN and GA model framework to compare model fit, (Table 12). The two most discriminatory peak analysis for these groups indicate a limited degree of separation between groups at this basic level (Figure 28). The GA model results indicate a higher discriminatory ability than in the model based exclusively on MLST sequence type (Table 10).

To increase the confidence that only the most likely contagious case isolates were used the comparison the groups were further refined using only OVR PTE cases with an inter-case interval of less than 8 weeks. The 95% confidence interval regions of the first two most discriminatory peaks overlapped less in the OVR PTE ICI<8wks Vs Solitary analysis (Figure 29), compared to the All OVR PTE Vs All Solitary analysis Figure 28. The PCA analysis (Figure 30) illustrated clear group phenotype separation.

The fit of both SNN and GA models was enhanced by the refining the group definition to isolates with short inter-case intervals, results are summarised in Table 12.

Table 12. Discriminatory capacity of *S.uberis* isolates of different case classifications using their MALDI spectral profile in SNN and GA models generated in ClinProTools. First model used case classifications used were all isolates of the nine over represented (OVR) sequence types which had

also been classified individually as a PTE (OVR PTE) case against all isolates classified as S, Solitary clinical cases. The second model used a refined grouping based upon only those OVR PTE cases with and intercase interval less than eight weeks and compared to a random sample of solitary cases as an equally sized group for analysis.

Model Name	Supervised Neural network		Genetic Algorithm	
	Cross Validation	Recognition Capability	Cross Validation	Recognition Capability
All OVR PTE v All Solitary	53.6%	53.3%	74.9%	86.9%
OVR PTE(<8wks ICI) vs Solitary (matched randomised sample)	87.8%	92.0%	85.6%	94.5%

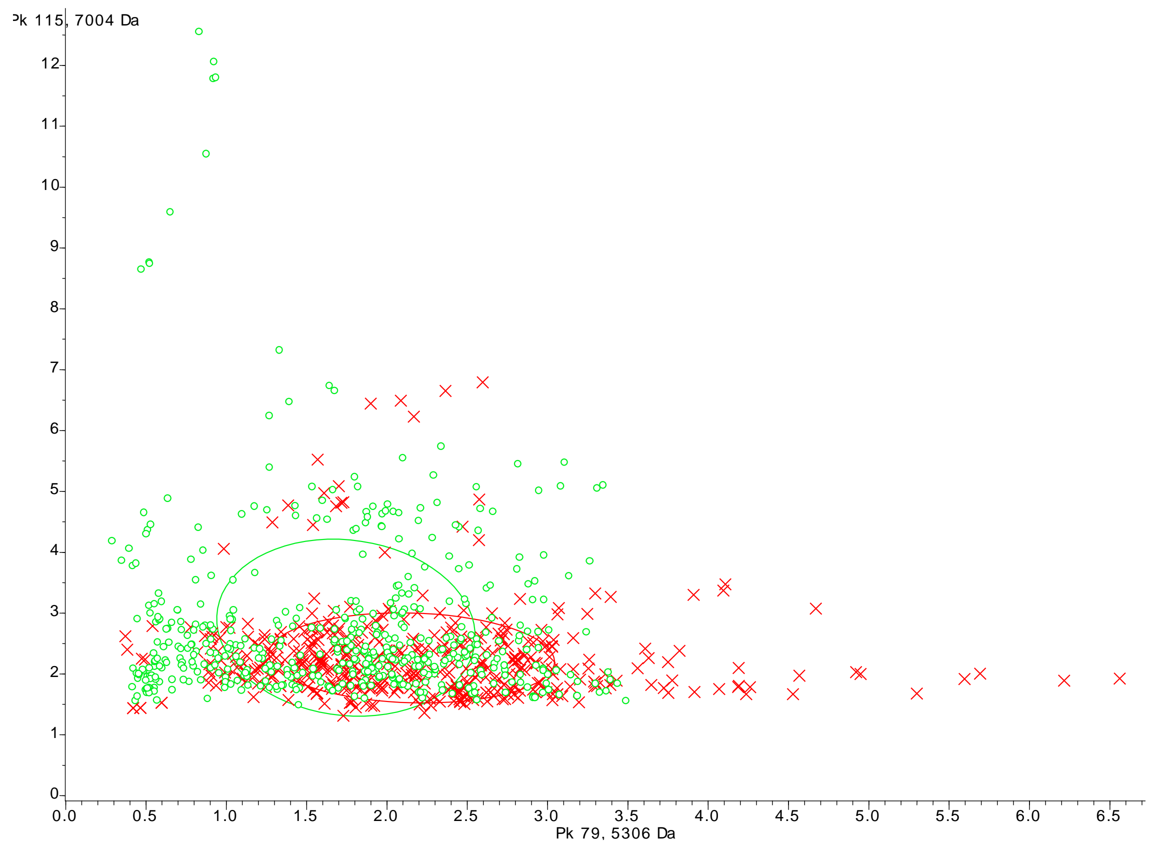

Figure 28. *S.uberis* isolates grouped by their case classification and MLST sequence type as (All OVR PTE) =red and all (Solitary) cases isolates = green plotted as a two dimensional analysis of the first two most discriminatory spectral peaks for the identification of these two groups (first peak on X axis,

second peak on y axis). Peak weight in Da is indicated on each axis and signal intensity is indicated in arbitrary units defined automatically by ClinProTools on each axis.

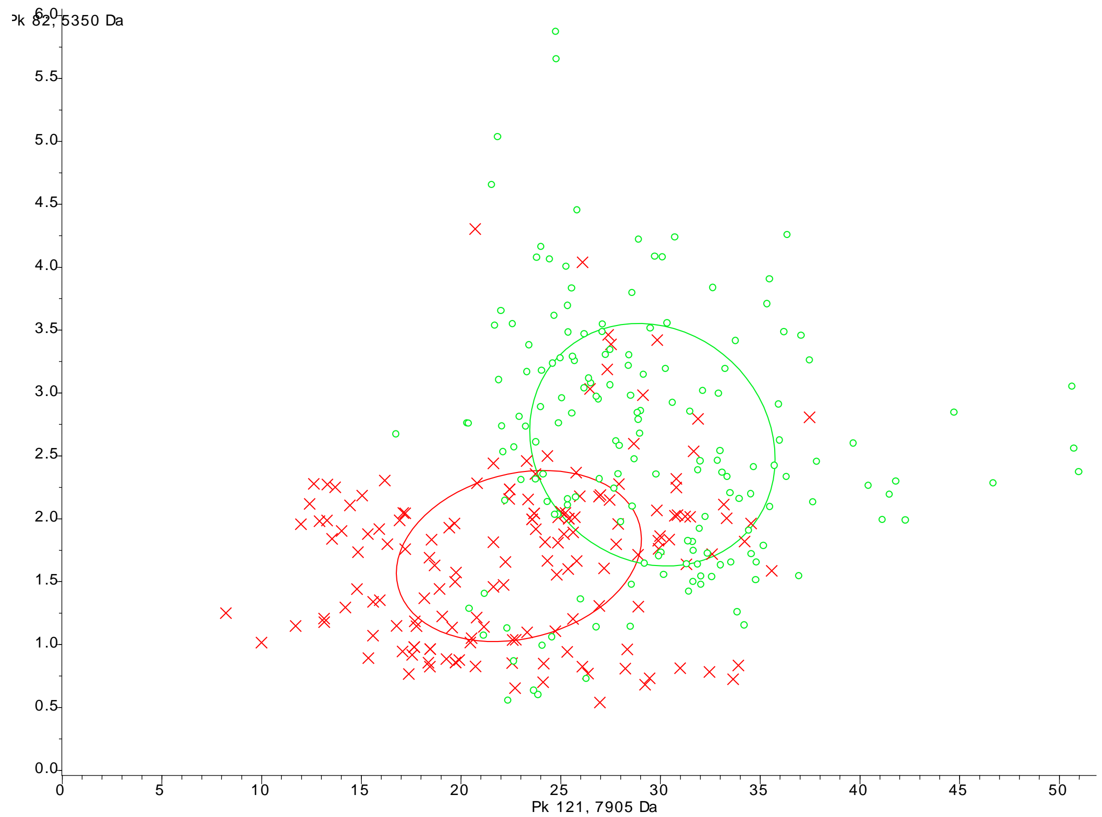

Figure 29. ***S.uberis*** **isolates grouped by their case classification and MLST sequence type as All OVR PTE(ICI <8wks) =red and a randomized sample of Solitary cases isolates = green plotted as a two dimensional analysis of the first two most discriminatory spectral peaks for the identification of these two groups (first peak on X axis, second peak on y axis). Peak weight in Da is indicated on each axis and signal intensity is indicated in arbitrary units defined automatically by ClinProTools on each axis.**

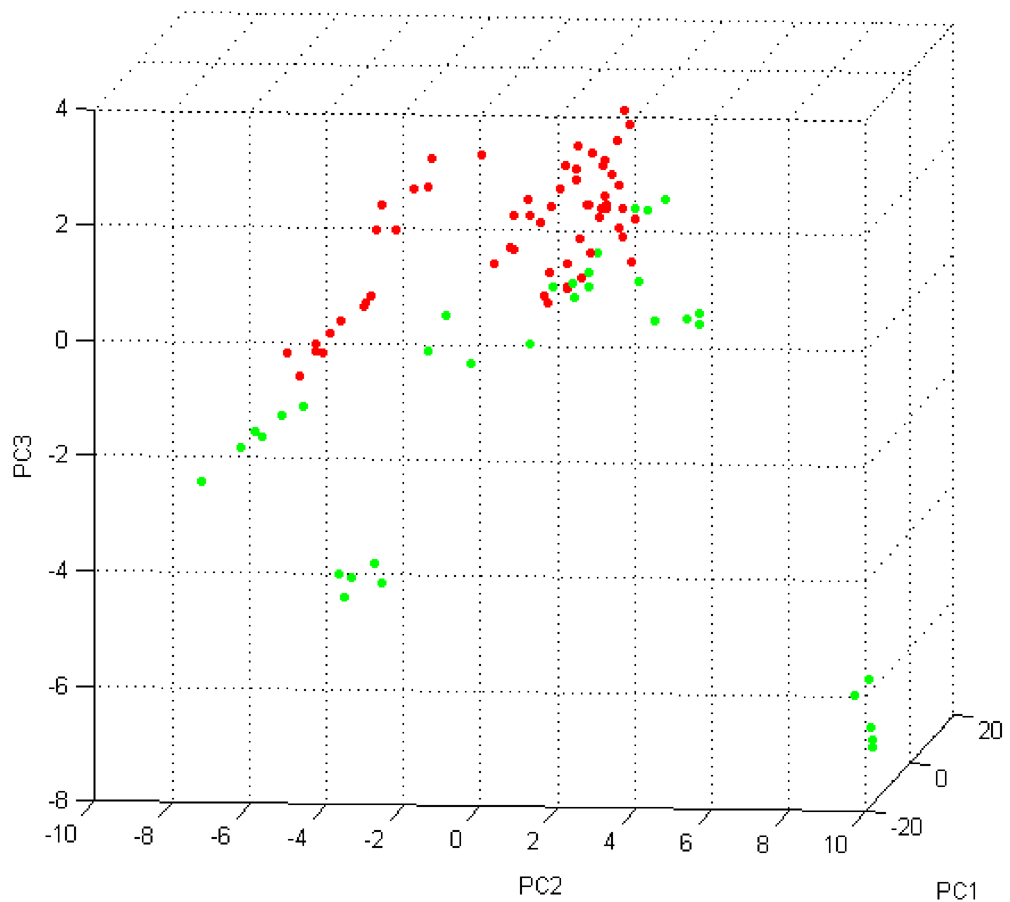

Figure 30. PCA 3D plot of MALDI spectral profiles grouped according to their case classification and MLST sequence type and the intercase interval between successive clinical cases within their herd of origin. All OVR PTE(ICI <8wks) in red and a randomized sample of Solitary cases isolates in green.

5.3.3 Aim 3 Determine if MALDI spectral profiles are predictive of epidemiological case classifications in a new herd

Herds were modelled initially as isolated epidemiological systems using PCA analysis (Figure 31-16) as well as by GA and SNN modelling Table 13. Within each herd model fit was good for all GA models and three of five SNN models. Individual, within farm models were more discriminatory than previous

models which compared isolates drawn from all farms. However, multi-farm models, (Pairwise PCA plots of *S.uberis* isolate MALDI profiles from farm 65 grouped according to case classification PTE (Red) and Solitary (Green).

Table 14), demonstrated high recognition capability and high cross validation whilst the test accuracy of the external validation, using isolate data from a farm not used for model generation was relatively poor in comparison in each of the five combinations analysed.

Table 13. Individual farm models of PTE verses Solitary case classification in both SNN and GA models testing discrimination ability between these case classifications within each farm based upon the MALDI spectral profile phenotypes of the isolates.

OVR PTE v Solitary Model for	Supervised Neural Network		Genetic Algorithm	
	Cross Validation	Recognition Capability	Cross Validation	Recognition Capability
Farm 8	68.9%	51.7%	96.6%	100.0%
Farm 20	88.0%	100.0%	100.0%	100.0%
Farm 27	86.0%	97.1%	97.1%	100.0%
Farm 55	92.0%	100.0%	97.4%	100.0%
Farm 65	74.9%	52.2%	100.0%	100.0%

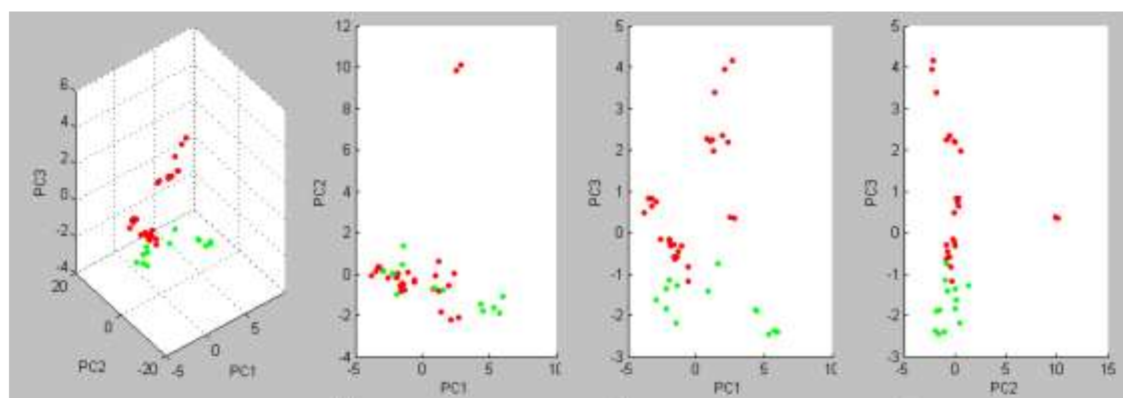

Figure 31. Pairwise PCA plots of *S.uberis* isolate MALDI profiles from farm 8 grouped according to case classification PTE (Red) and Solitary (Green).

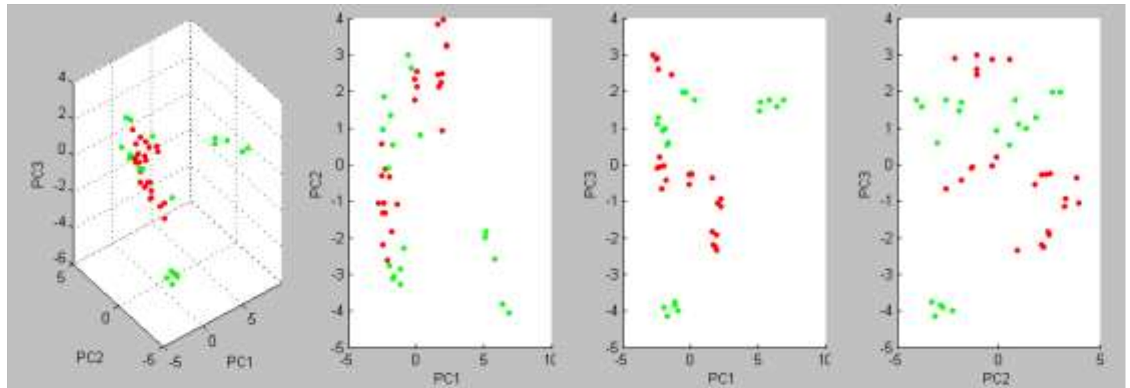

Figure 32. Pairwise PCA plots of *S.uberis* isolate MALDI profiles from farm 20 grouped according to case classification PTE (Red) and Solitary (Green).

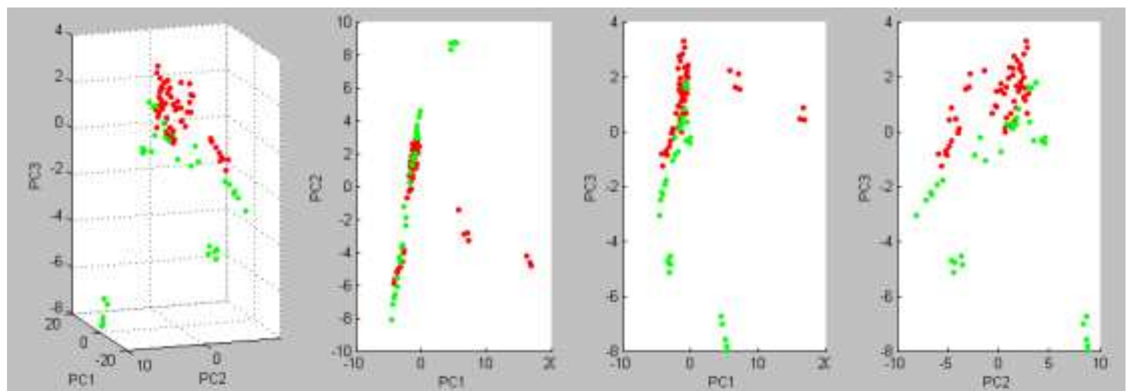

Figure 33. Pairwise PCA plots of *S.uberis* isolate MALDI profiles from farm 27 grouped according to case classification PTE (Red) and Solitary (Green).

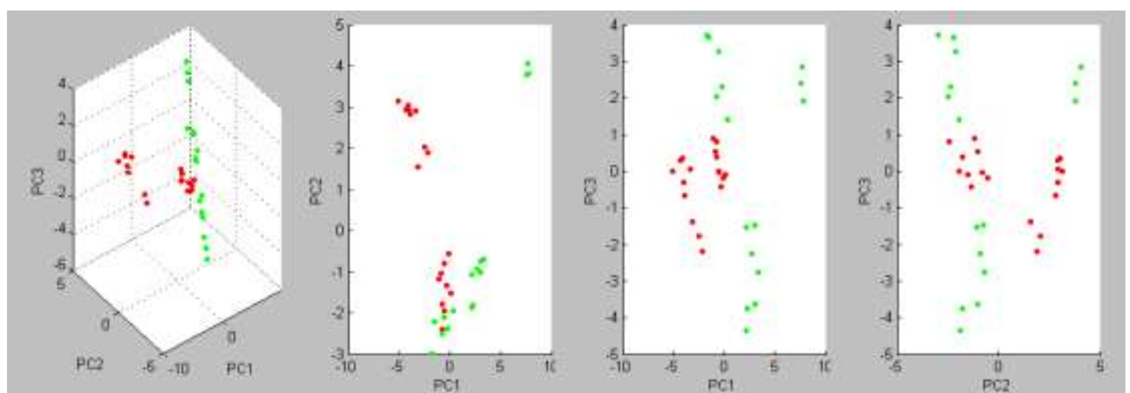

Figure 34. Pairwise PCA plots of *S.uberis* isolate MALDI profiles from farm 55 grouped according to case classification PTE (Red) and Solitary (Green).

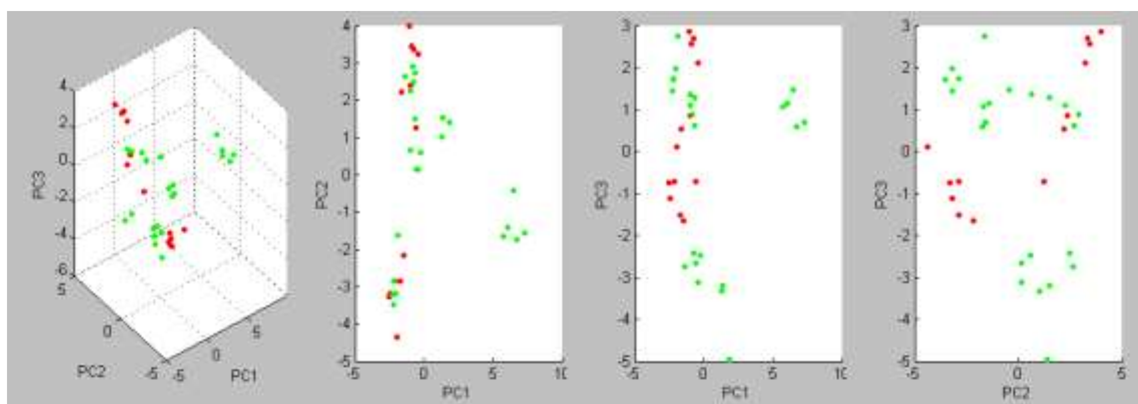

Figure 35. Pairwise PCA plots of *S.uberis* isolate MALDI profiles from farm 65 grouped according to case classification PTE (Red) and Solitary (Green).

Table 14. Multi-farm SNN and GA models of isolates grouped as OVR PTE or Solitary by clinical case classifications using four farms for model training and a fifth farm for external validation. Model performance is quantified by internal measures of 'Recognition capacity' and 'internal cross validation' and by external validation quantified as test accuracy.

OVR PTE vs Solitary Model Farms: 8,20,27,55,65	SNN			GA		
	Internal Cross Validation	Recognition Capability	Test Accuracy	Internal Cross Validation	Recognition Capability	Test Accuracy
(8,20,55,65) Fm27 used for external validation	71.1%	51.8%	34.0%	90.0%	97.3%	47.0%
(20,27,55,65) Fm8 used for external validation	63.5%	80.6%	64.0%	83.7%	99.1%	59.0%
(8,27,55,65) Fm20 used for external validation	61.1%	50.4%	43.2%	90.3%	97.1%	43.2%
(8,20,27,65) Fm55 used for external validation	75.5%	89.5%	38.2%	83.4%	91.2%	52.9%
(8,20,55,27) Fm65 used for external validation	53.1%	74.9%	25.0%	85.0%	96.6%	52.8%

5.3.4 Aim 4: To Determine the between farm variation in MALDI phenotype of Solitary and PTE isolates.

5.3.4.1 Between farm variation amongst Solitary isolates

The Genetic Algorithm method generated a classification model with higher recognition capability and internal cross validation than the Supervised Neural Network (Table 15) The PCA (Figure 36 & Figure 37) and two dimensional most discriminatory peak analysis (Figure 38) illustrate the farm specific variability in MALDI spectral characteristics in the Solitary isolate populations.

Table 15. Supervised neural network (SNN) and Genetic algorithm (GA) models of the Solitary isolate populations on four separate farms with approximately equal number of isolates where each farm was a modelled as a separate group.

Multi-farm Model (8,20,55,65)	SNN		GA	
	Cross Validation	Recognition Capability	Cross Validation	Recognition Capability
Solitary	46.2%	44.2%	93.4%	100.0%

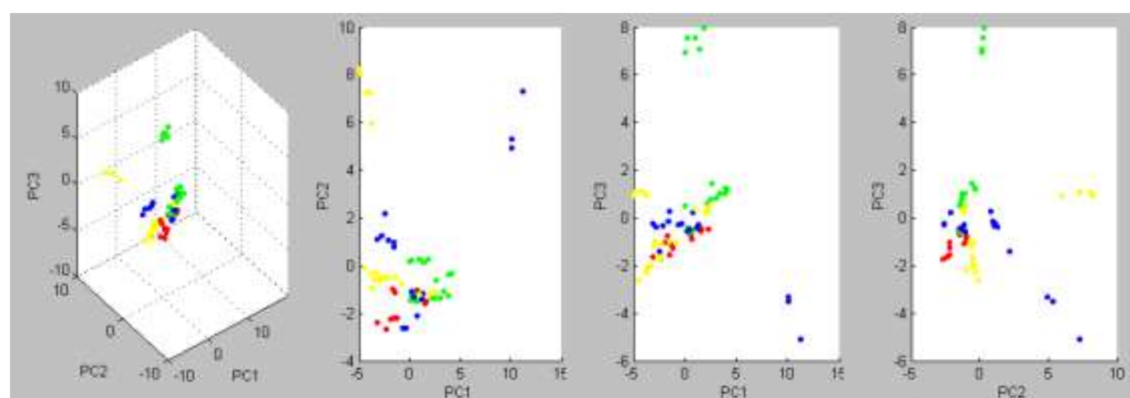

Figure 36. Pairwise PCA plots of MALDI spectral profiles from isolates classified as Solitary clinical cases and grouped according to their farm of origin by colour as follows: Farm 8 (red), Farm20 (green), Farm 55 (blue), Farm 65 (yellow).

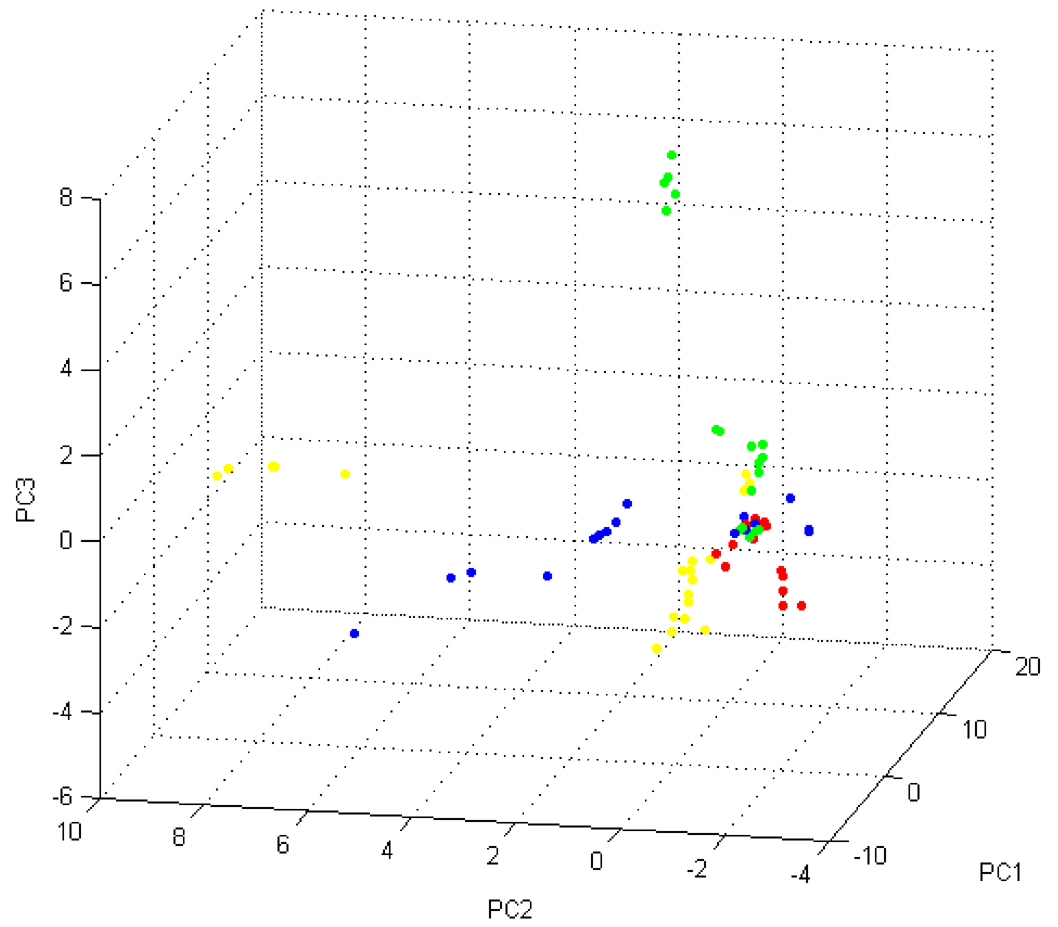

Figure 37. PCA plot of MALDI spectral profiles from *S.uberis* clinical cases classified as PTE isolates from Farm 8 (red), Farm 20 (Green), Farm 55 (Blue) and Farm 65 (Yellow) NB: Farm 27 not included.

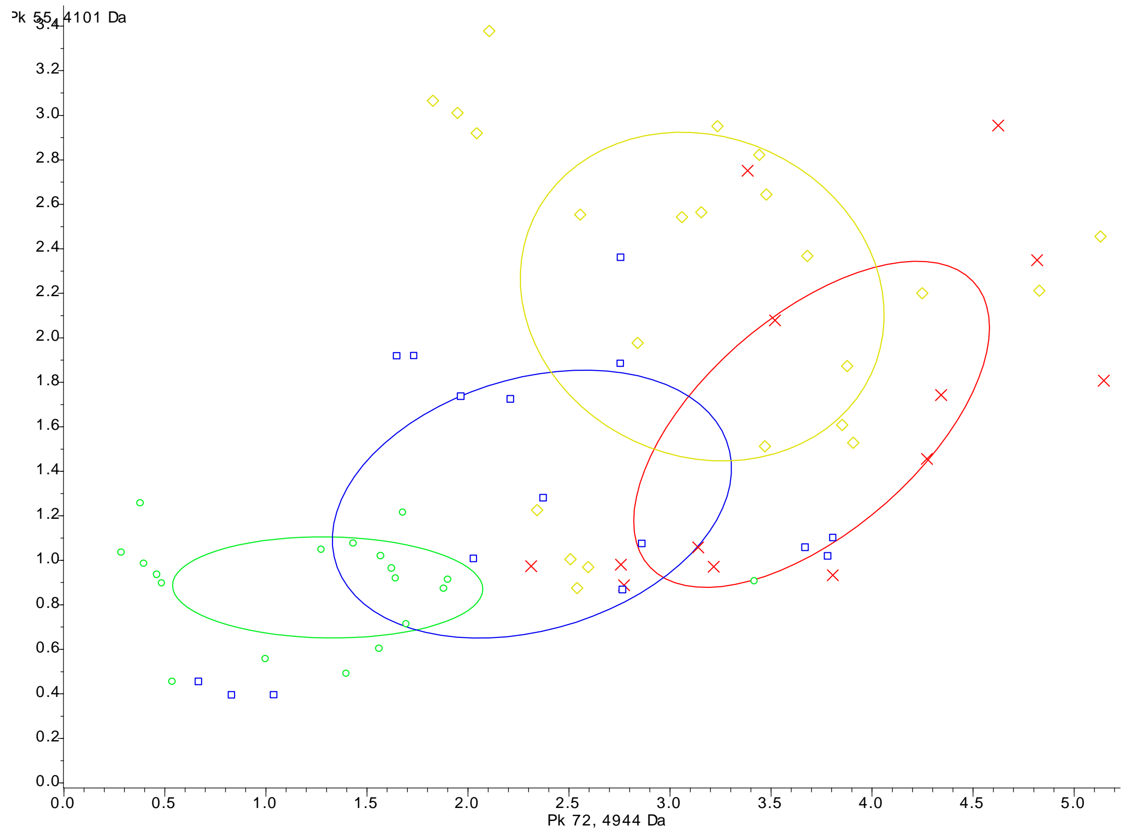

Figure 38. Two dimensional plot representation of the first two most discriminatory MALDI spectral peaks for Solitary isolates from Farm 8 (red), Farm 20 (Green), Farm 55 (Blue) and Farm 65 (Yellow) NB: Farm 27 not included.

5.3.4.2 *Between farm variation amongst PTE isolates.*

The Genetic Algorithm again generated a model for classifying the four farm groups with very high recognition capacity and high cross validation whereas the supervised neural network method, performed particularly poorly in this analysis (Table 16).

Table 16. Supervised neural network and Genetic algorithm models of the PTE isolate populations on four separate farms where each farm was a modelled as a separate group.

Multi-farm Model (8,20,55,65)	SNN		GA	
	Cross Validation	Recognition Capability	Cross Validation	Recognition Capability
PTE	73.3%	26.9%	91.0%	100.0%

The PCA and two dimensional, best discriminatory peak analysis (Figure 39 & Figure 40) both illustrate distinct segregation of isolate MALDI spectral profiles according to the farm of origin.

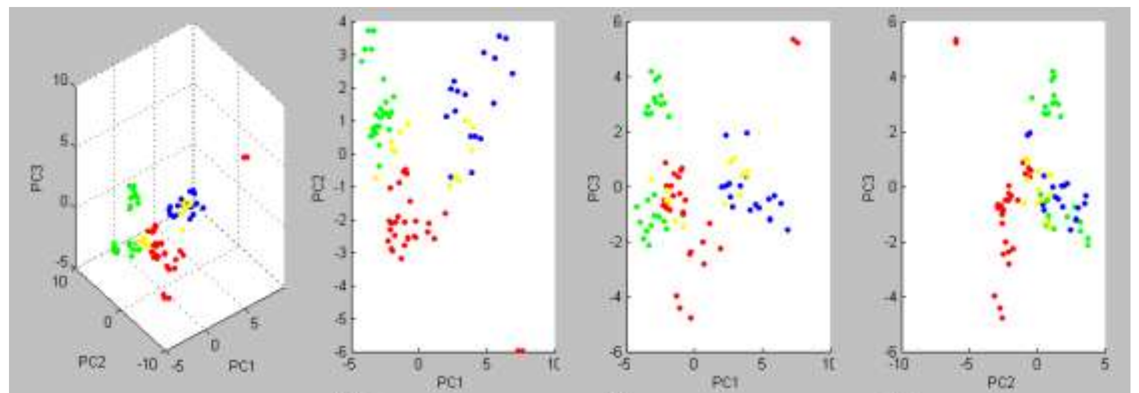

Figure 39. Pairwise PCA plots of MALDI spectral profiles from isolates classified as PTE clinical cases and grouped according to their farm of origin by colour as follows. Farm 8 (red), Farm20 (green), Farm 55 (blue), Farm 65 (yellow).

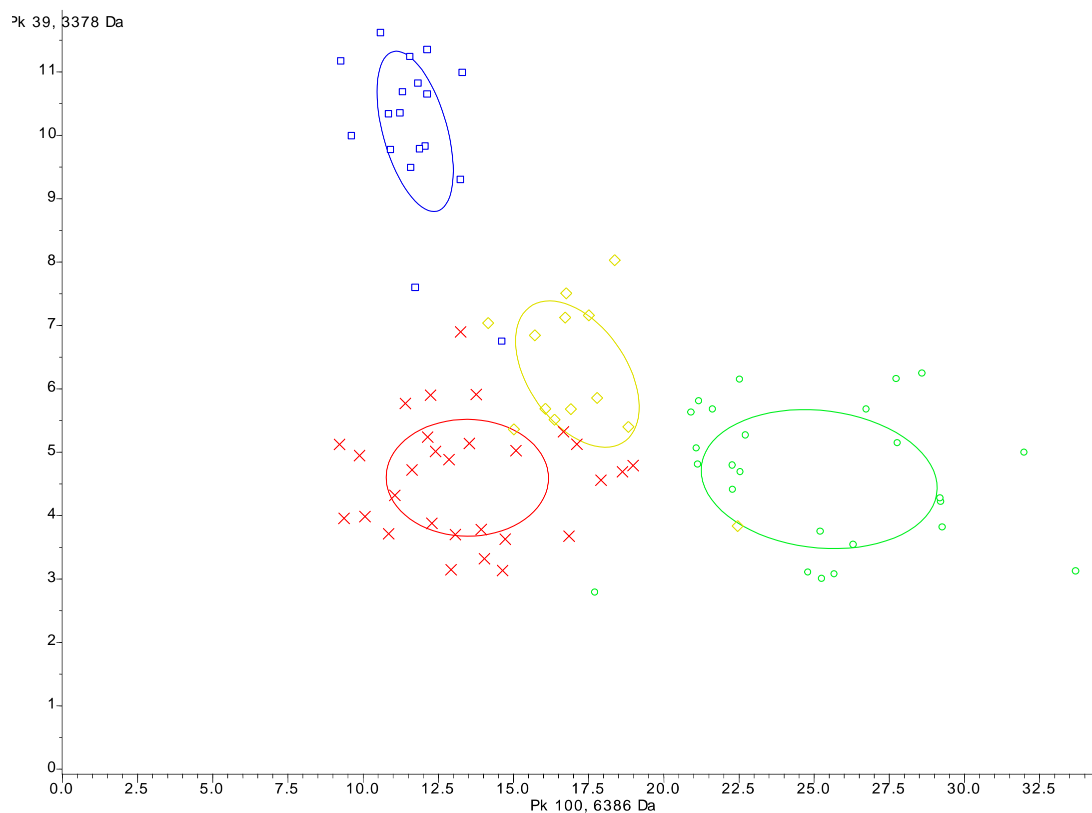

Figure 40. Two dimensional representation of the first two most discriminatory MALDI spectral peaks for *S.uberis* clinical mastitis cases classified as PTE isolates grouped according to farm of origin as follows: Farm 8 (red), Farm 20 (Green), Farm 55 (Blue) and Farm 65 (Yellow) NB: Farm 27 not included.

In contrast, when PTE isolates from Farm27 were included in the analysis, peak 41 (3379Da) and peak 105 (6387Da) were identified as the most discriminatory. However, the PTE isolates from Farm 27 had more variable peak characteristics, forming a larger and less distinct group than was the case with the other four farms (Figure 41). The group of PTE isolates from Farm 27 shared peak characteristics with all four other herds. In the 2 dimensional, most discriminatory peak analysis Farm 55 appeared to be the most distinct population.

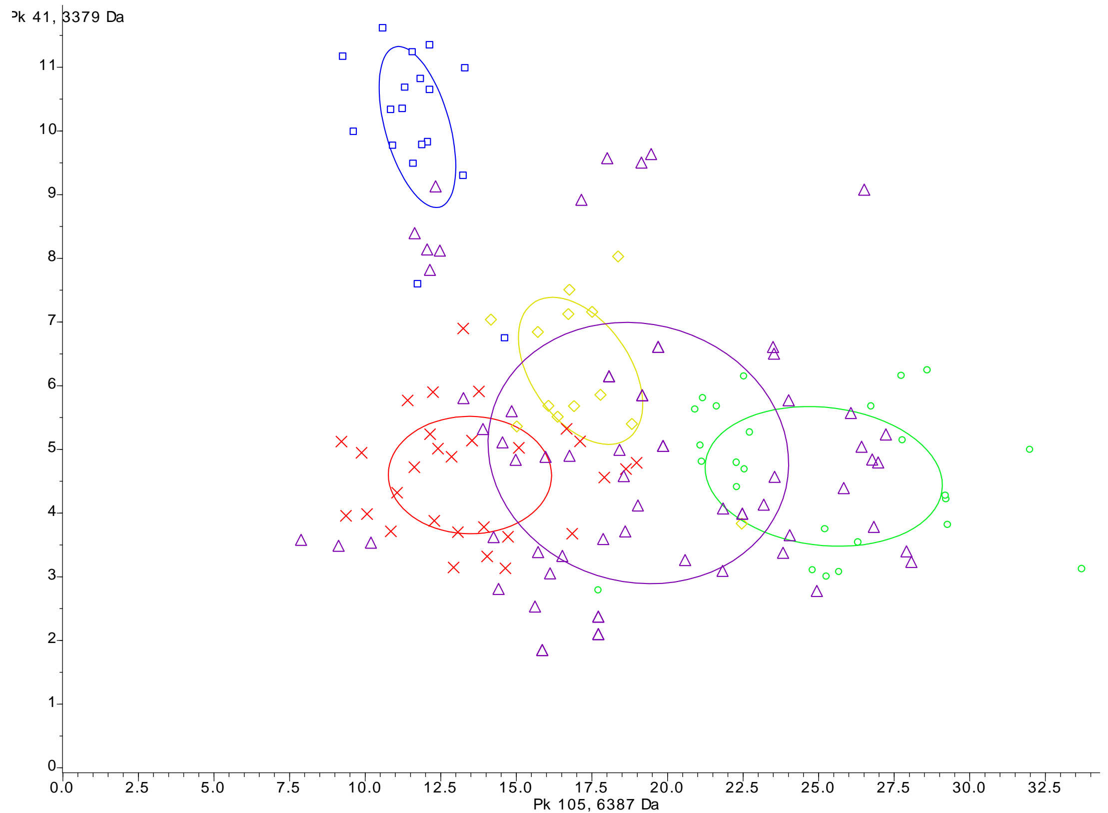

Figure 41. Two dimensional representation of the first two most discriminatory MALDI spectral peaks for S.uberis clinical mastitis cases classified as PTE isolates grouped according to farm of origin as follows: Farm 8 (red), Farm 20 (Green), Farm 55 (Blue) and Farm 65 (Yellow) NB: Farm 27 not included. Farm 8 (red), Farm 20 (Green), Farm 55 (Blue) and Farm 65 (Yellow) and Farm 27 (Purple).

5.4 Discussion

The first aim of this analysis was to establish, by using a large data number of isolates if genotype groupings could be differentiated using MALDI spectral profiles. The results indicated that the clonal complex of isolates which were genetically similar to ST6 (CC6) could be differentiated moderately well from the more genetically diverse and distantly related MLST sequence types by the MALDI phenotype (Table 10). This was supported by the findings of the pairwise comparison of ST6 to ST5 and ST6 to ST20. This may be expected because the MLST genes were originally chosen to be discriminatory and well

dispersed throughout the genome to limit the risk of major genetic variation due to recombination causing substantial, undetected genetic differences. Agreement between MLST and protein MALDI has been assessed in *Staphylococcus haemolyticus* (Kornienko et al. 2016) and *Leptospira species* (Rettinger et al. 2012) using similar techniques with some success. However, these studies did not test the sub-species discriminatory resolution of the protein MALDI discrimination to the same extent as has been attempted here or using such large numbers of isolates and MLST sequence types. Alternative MALDI protocols such as nucleic acid MALDI has been shown to successfully replicate MLST typing in *E.coli* (Tagg et al. 2015) with very fidelity, this technique presents very interesting possibilities as a rapid genotyping tool for further research and rapid strain diagnostics but no such comparison has been attempted with *S.uberis* or with such large numbers of isolates to date. Nucleic acid MALDI does not provide the protein expression phenotype information that is available by the cellular protein MALDI technique used in this study. The moderate discriminatory power of the MALDI protein spectra profile for MLST sequence type indicated that the MLST sequence type of an isolate is associated with a reasonable proportion of the phenotypic variation between isolates. However, other genetic or epigenetic factors must also be responsible for a substantial, detectable proportion of the variation that could not be accounted for in the MLST based classification based models.

The second aim was to determine if the case classifications assigned to each isolate on the basis of their MLST sequence type and the pattern of clinical

cases of that sequence type within their herd, correlated with a detectable difference in peak detection within the MALDI spectral profile. Initially all the potential transmission event case isolates from the nine over represented MLST sequence types (OVR PTE) were compared to all of the Solitary isolates. This model was more discriminatory between these two groups than the model based upon MLST clonal complex (CC6 vs Non-CC6). This indicates that the case classifications correlate better with the distribution of MALDI spectra profiles than does the MLST clonal complex classification scheme. The model was refined to further increase the likelihood of all isolates being the product of contagious transmission. Only those over represented sequence type isolates that were classified as PTE and occurred within 8 weeks of the previous case within that herd of that sequence type, in light of the multi-level modelling results described in Chapter 4. The model fit was improved by this refinement of the group definition, further supporting the hypothesis that genotype and clinical case dynamics within a herd can be used to predict the MALDI spectral profile. This is the first instance where MALDI has been used to type bacterial groups according to epidemiologic behaviour classification in the absence of a defined molecular definition of the groups, although the technology has been widely used as a typing tool based upon previously defined genotypes or serogroups (S Nakano et al. 2015; Rettinger et al. 2012; Sousa et al. 2014).

The third aim was to determine if an accurate prediction case classification could be made on the basis of MALDI spectral profile. The five herds selected

could all be readily classified by the Genetic Algorithm model and three herds were successfully classified with high recognition capability and internal cross validation by both GA and SNN model algorithms. However, external validation using a novel herd, not used in the model generation resulted in poor classification performance and low test. This could potentially be explained by substantial between farm heterogeneity in MALDI phenotype in addition to variation within a herd between isolates of different case classifications. This is plausible if we consider the limited contact between farms and therefore the likelihood of divergent evolution under such circumstances. A similar pattern of diversity within a bacterial species has been demonstrated in Campylobacter populations on broiler farms in Switzerland (Wittwer et al. 2005) and at a much larger scale divergent evolution of Helicobacter pylori has been described in human populations, in different geographical regions, which have had migrated from a common origin (Falush et al. n.d.).

The final aim in this Chapter was to determine heterogeneity of isolate populations of the same case classification found in different herds. Within each of the case classifications (PTE and Solitary) the isolates from each individual farms could be distinguished effectively by the GA model. This indicates that both the 'Solitary /presumed environmental' *S.uberis* populations and the 'potentially contagious' *S.uberis* populations have MALDI profile characteristics which correlate both to their pathogenic mechanisms and also spectral characteristics which are more similar to with their

particular farm of origin as illustrated in Figure 42. The substantial genetic diversity of *S.uberis* identified in all epidemiological studies using MLST (Barkema et al. 2001; Cattell 1996; Coffey et al. 2006; Cursons and Leigh 2007; Pullinger et al. 2006, 2007; Tomita, Meehan, Wongkattiya, Pullinger, et al. 2008; Zadoks, Schukken, et al. 2005) were unable to identify a specific population structure because of the discriminatory limitations of the classical MLST scheme. The MALDI ToF mass spectrometry technique provides one alternative typing method that can cluster isolates according to a combination of their genotype and the epigenetic regulation of that genotype. This facilitates a different level of isolate discrimination which may be more appropriate for certain epidemiologic analysis of very genetically heterogeneous bacterial species such as *S.uberis*.

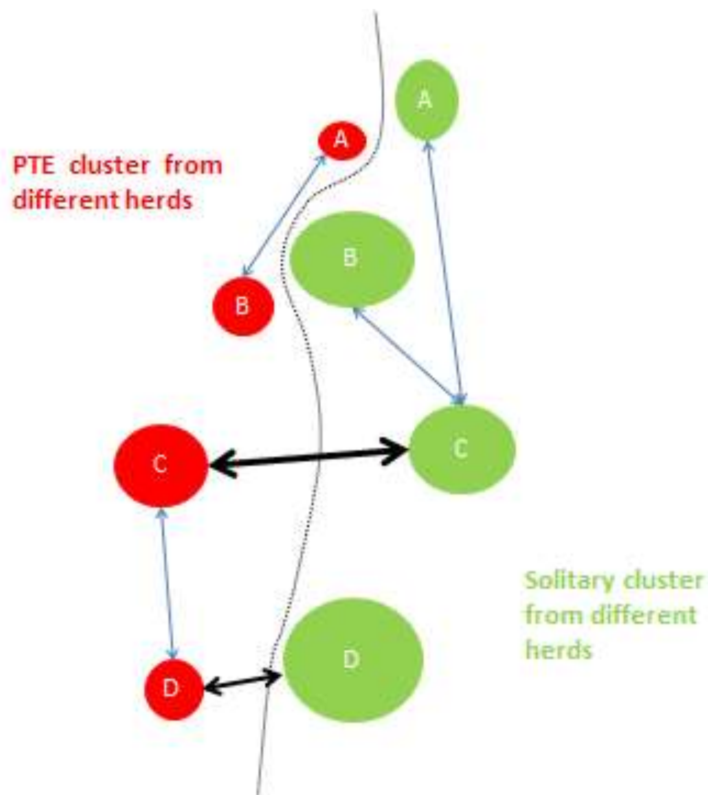

Figure 42. Schematic diagram of hypothetical phenotypic differences between case classifications in Red and Green and farms A,B,C,D. Dotted line illustrating model boundary. The black arrows represent the MALDI spectral heterogeneity between isolates from the same herd but from different case classifications whereas the blue arrows refer to the MALDI spectral heterogeneity between isolates of the same case classification in different herds.

It was observed that the performance of the Genetic Algorithm (GA) was substantially better than that of the Supervised Neural Network (SNN) in this study. In a larger related data set of *S.uberis* mastitis isolates the performance of the two models was comparable (Archer S, personal communication 2016). It was suspected that the SNN methodology requires a larger range of potential prototype spectra in order to optimise the model whereas the GA model is constructed on a limited number of peaks within each spectra which is not as reliant on the number of sample spectra for good model fit. Both model algorithms have been used successfully to discriminate between

bacterial species sub-populations in *Leptospira* and *Mycoplasma* (2,16) respectively.

The analysis of farm levels MALDI patterns has highlighted the importance of understanding the limitations of the classic *S.uberis* MLST scheme in understanding the epidemiology of transmission. The analysis indicated that there was indeed a protein expression difference between potentially contagious and presumed opportunistic, environmental isolates of *S.uberis*.

There are clear potential benefits in a more comprehensive analysis of the genome to investigate the possible genetic elements underpinning the case classifications and also a more nuanced analysis of the genome and potentially the regulatory modulation of the genome that could also contribute to the different patterns of disease that have been observed.

Chapter 6: Whole genome sequencing of selected *S.uberis* isolates

6.1 Introduction

Having demonstrated in the previous Chapter that the clinical case classifications consistently correlate with different MALDI spectral profiles at the farm level, two objectives were identified to explore the underpinning biology of the most likely contagious isolates in greater depth.

Objective 1: Evaluate genetic and phenotypic heterogeneity between isolates of MLST sequence types associated with potential contagious transmission.

Objective 2: Evaluate evidence of specific pathogenic mechanisms in potentially contagious isolates.

Both of these objectives required more information about the genome of the bacterial isolates to increase the discriminatory capability and allow direct, detailed comparison of specific gene targets in multiple isolates. Whole genome sequencing (WGS) using ‘next-generation’ sequencing technology allows investigation of variation at several hundreds or thousands of gene loci rather than the 7 predefined loci utilised by the conventional MLST scheme. Analysis of sequence variation between isolates across a larger number of loci throughout the genome allows greater discrimination between isolates and therefore more defined clustering of isolates for molecular epidemiological studies. Several recent studies have used core genome MLST (cgMLST)

analysis to elucidate epidemiological relationships in human disease outbreaks involving several bacterial species including; *Mycobacterium tuberculosis* (Kohl et al. 2014)*, Klebsiella* (Weterings et al. 2015) and *Listeria monocytogenes* (Ruppitsch et al. 2015). In this project WGS provided the ability to investigate the similarity of isolates which have been categorised as potential transmission events (PTE) and persistent (P) clinical cases both within and between herds. Pairwise analysis of the WGS data from multiple isolates from the same persistent mammary gland infection allowed investigation of specific virulence genes which have been suggested as contributors to an immune escape/evasion mechanism (Leigh et al. 2010). Whereby the bacteria alters the antigenic proteins which are expressed, normally on the cell surface, to which the immune system has developed a response, delaying effective targeting of the bacteria by the host's adaptive immune response.

In order to understand what determines the behaviour of an isolate at the cow or farm level it is important to understand not only the genotype diversity but also how that relates to the proteins actually expressed by the isolates in question. In the investigation of both these objectives the protein expression phenotype of the bacteria was compared to cgMLST typing using MALDI composite correlation indexes (CCI) for accurate, quantitative, analysis of similarity between specific isolates.

To meet the two objectives the following 5 aims were identified:

Aims to meet Objective 1

1. Validate classical MLST scheme sequence type classifications by cgMLST
2. Determine the genetic heterogeneity of classical MLST sequence types within a single herd and between multiple herds using cgMLST analysis.
3. Determine if the genetic heterogeneity between potentially contagious sequence types within a herd correlates with heterogeneity in isolate MALDI spectral profiles.

Aims to meet Objective 2

4. Determine if isolates classified as persistent clinical cases in the same mammary gland according to MLST sequence type are true persistent infections or new infections with the same sequence type.
5. Determine if there is evidence of antigenic variation of three virulence genes (SUB 0145/RS00865, SUB 1095/RS 5505,SUB 1154/RS5795) as a possible immune escape mechanism in persistent mastitis case.

6.2 Methods

6.2.1 Isolate selection

Only a limited subset (68 of 494) of isolates could be analysed at the whole genome sequence level. To meet all the aims isolates were selected for WGS based upon criteria which selected only persistent infection isolates from MLST sequence types which were associated with PTE cases. Isolates of PTE associated sequence types were then selected from those herds from which

persistent infection isolates had already been identified and additionally from other herds where isolates of that sequence type had been classified as PTE cases. In this way many of the persistent infection isolates that were required for the investigation of Aim 4 & 5 were also suitable for wider analysis to fulfil aims 1-3.

Selection criteria:

1. All isolates of the Persistent (P) case classification along with their accompanying initial case of that ST in that quarter (that first case could be either an I, T or U classification (as defined in Chapter 3)) if that ST had previously been classified as a PTE.
2. Isolates classified as PTE in the same herd and of the same ST as the selected Persistent isolates.

Final selection of the isolates balanced the number of isolates per herd to optimise the inter-herd sampling breadth and intra-herd depth of sampling for Farm 27 where the majority of candidate isolates from the initial selection originated (Table 17 & 18).

Table 17. *S.uberis* Isolates selected for whole genome sequencing by their MLST sequence type and case classifications.

MLST ST	Persistent isolates	PTE isolates	Number of herds	Number of isolates
5	4	2	2	6
6	6	13	3	19
20	3	6	1	9
23	2	0	1	2
24	2	2	1	4
35	2	1	1	3
67	4	2	2	6
233	4	2	2	6
343	3	0	1	3
361	2	1	1	3
372	4	0	2	4
512	2	1	1	3
Total	38	30	15	68

Table 18. Number of isolates selected per farm for whole genome sequencing

Farm id number	Isolates per farm
3	3
4	3
15	2
16	2
21	2
26	3
27	31
33	2
39	6
41	3
54	2
56	3
58	2
62	4
14	68

Farm 27 contributed 30 of the 62 clinical case isolates which were successfully whole genome sequenced. This represented 61% of all the MLST sequenced isolates from Farm27 and 88% of all the PTE, I and R classified mastitis cases. The high percentage of isolates sequenced from Farm27 allowed analysis to be conducted on the temporal relationship of clinical mastitis isolates within a herd comparing the results of the original, classical MLST to core genome MLST (cgMLST) and (MALDI) phenotypes.

To assess within herd concordance between cgMLST and MALDI- 27 isolates of three MLST sequence types, from Farm27 were selected. These included only isolates with both MALDI and cgMLST data; 15 x ST6 isolates, 8 x ST20 isolates and 4 x ST233 isolates.

6.2.2 Sequencing Protocols and Genome Assembly

All protocols for isolate culture, preparation, gDNA extraction, whole genome sequencing, sequence data quality control and genome assembly are detailed in Chapter 2.

6.2.3 Bioinformatic Analysis

Assembled genomes generated in *Ridom Seqsphere* and by ADAC were compared. Dendrograms were generated for total SNP in *MEGA7.0* for comparison with the cgMLST dendrograms generated in *Ridom SeqSphere*. The node structure of trees generated in *MEGA7.0* resolved with high bootstrap values and clustering of isolates and clade structure was

comparable between both analysis typing methods and both assembly techniques. *Ridom SeqSphere* was adopted as principal analysis tool for this analysis due to its greater flexibility for clustering analysis of cgMLST isolate data. Clustering was assessed by minimum spanning trees and neighbour joining trees generated using *Ridom SeqSphere*. Quantitative analysis of genetic relatedness was conducted using distance matrices generated from accumulated single nucleotide polymorphisms within all the cgMLST target loci by *Ridom Seqsphere*. MALDI spectral profiles were assessed using composite correlation indices (CCI) matrices generated from pairwise isolate comparisons of every spectral peak by molecular weight and signal intensity in Bruker *MALDI Biotyper3.0*.

Data output in the form of cgMLST distance matrices and MALDI CCI matrices were exported into Windows Excel 2010 for data processing and MiniTab17 for statistical analysis using nonparametric Moods Median tests using a 95% confidence interval about the median.

Virulence gene analysis was conducted using the online tools BLAST (http://blast.ncbi.nlm.nih.gov) and SMART (http://smart.embl-heidelberg.de) and pairwise alignment for SNP and amino acid substitution identification was conducted in Ridom SeqSphere+. Pfam (Finn et al. 2015)(http://pfam.xfam.org) was used to confirm identity of protein products and region mapping.

6.3 Results

6.3.1 Aim 1: Validation of classic MLST scheme sequence type classifications by cgMLST

Whole genome sequencing confirmed the sequence type identity assigned to each isolate by MLST (Sanger sequencing) in 61 of the 62 samples tested. One sample which had originally been classified as ST361 was subsequently assigned to a novel ST group with a previously unknown allelic profile. The novel MLST profile diverged from that of ST361 at the ddl locus a SNP comparison between the isolate and the ST361 reference revealed eight silent SNPs distributed along the sequence.

Of the 68 genomic DNA samples, Two were shown to be contaminated and shown to contain DNA from either *Lactococcus garvieae* or *Enterococcus faecalis* and four further samples could not be sequenced due to issues of DNA quality and or quantity. The following analysis was therefore carried out on 62 confirmed *S.uberis* isolates for which whole genome sequences were generated, unambiguously.

6.3.2 Aim 2: Determine the genetic heterogeneity of classical MLST sequence types within a single herd and between multiple herds using cgMLST analysis.

The cgMLST data was displayed as a neighbour joining tree dendrogram and as a minimum spanning tree for visual assessment of clustering Figure 43 & Figure 44. Two large, mutually exclusive clades of isolates all sharing a common node could be readily identified from the cgMLST dendrogram,

Figure 43, which were labelled (C20) and (C6) after their principal members. Within the C6 clade there was evidence of substantial heterogeneity with 'cross-clustering' between sequence types and farms. This could be seen with several sequence types; ST5, ST6, ST67, ST233 which clustered more closely with isolates of different sequence types in their own herd or other herds than they did with isolates of their own sequence type. C20 had fewer isolates and no sequence types which appeared in more than one herd so a direct comparison with the heterogeneity evident in C6 was not possible.

Within farm 27 there was evidence of separate clades within the ST6 sequence type which clustered more closely with isolates of other sequence types in separate herds than with the alternate clades of ST6 from their own herd, farm 27, as depicted in the minimum spanning trees Figure 44 – 46.

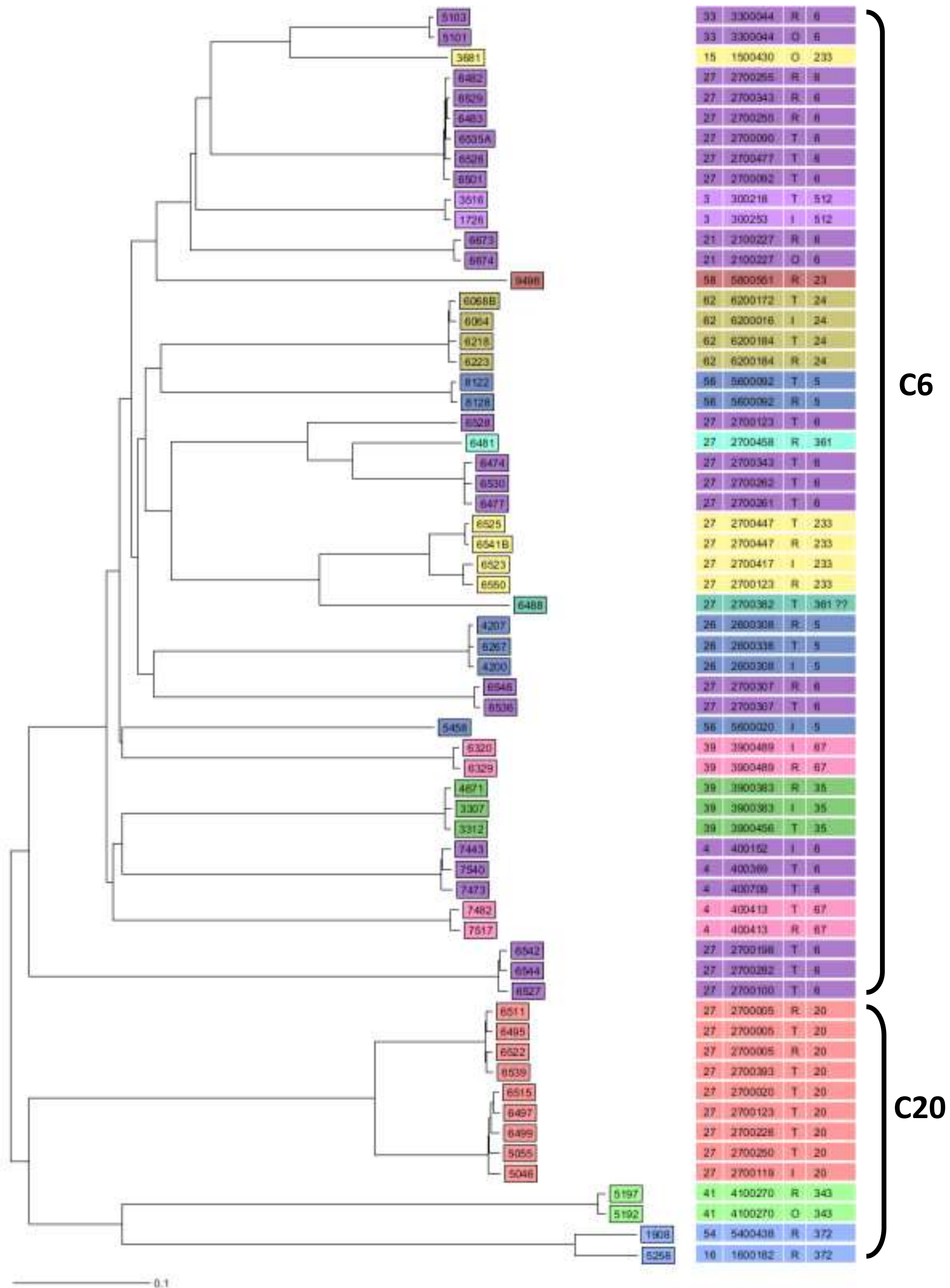

Figure 43. Neighbour joining tree of all 62 cgMLST isolates generated in Ridom SeqSphere+. Sample ID located at each terminal branch, Farm of origin is displayed in the left column and MLST sequence type (ST) is grouped by colour and displayed in the right column. The horizontal scale represents the genetic distances in substitutions per nucleotide.

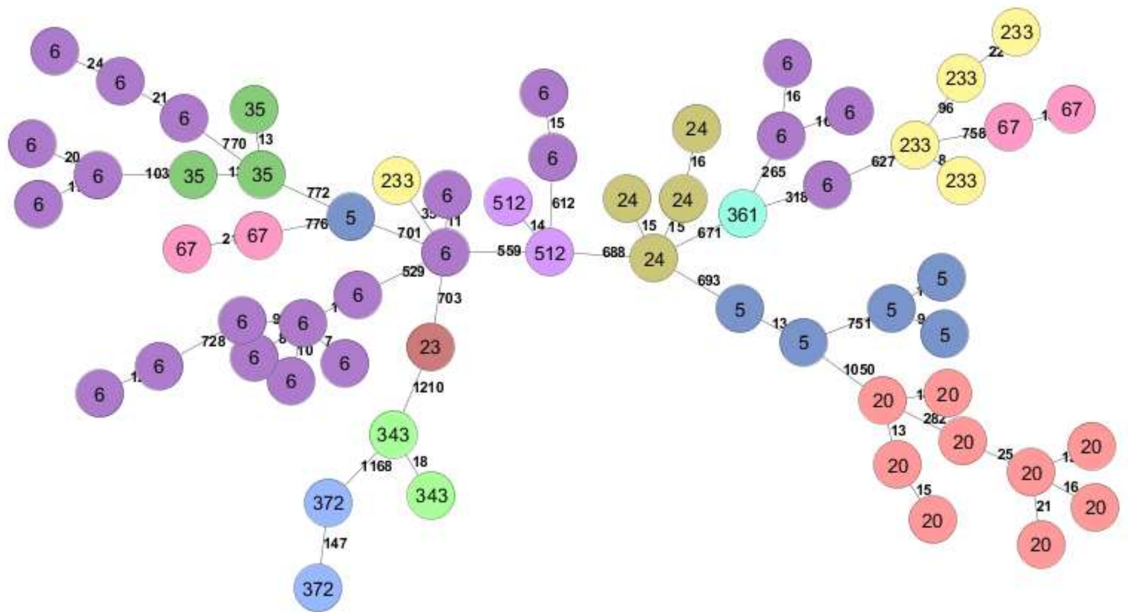

Figure 44. Minimum Spanning Tree of all 62 *S.uberis* isolates by cgMLST. Branch lengths scaled log10 represent the genetic distances in substitutions per nucleotide. Node label & colour represent the MLST sequence type. Seqsphere+.

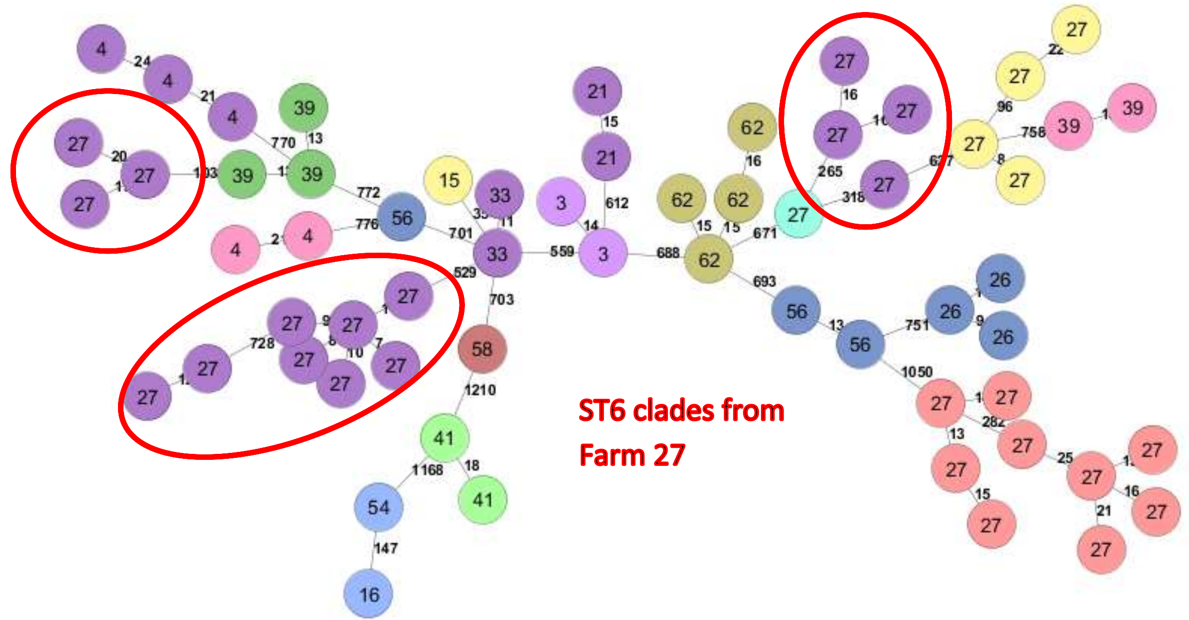

Figure 45. Minimum Spanning Tree of all 62 *S.uberis* isolates by cgMLST. Branch lengths scaled log10 represent the genetic distances in substitutions per nucleotide. Node label & colour represent the MLST sequence type. Seqsphere+.

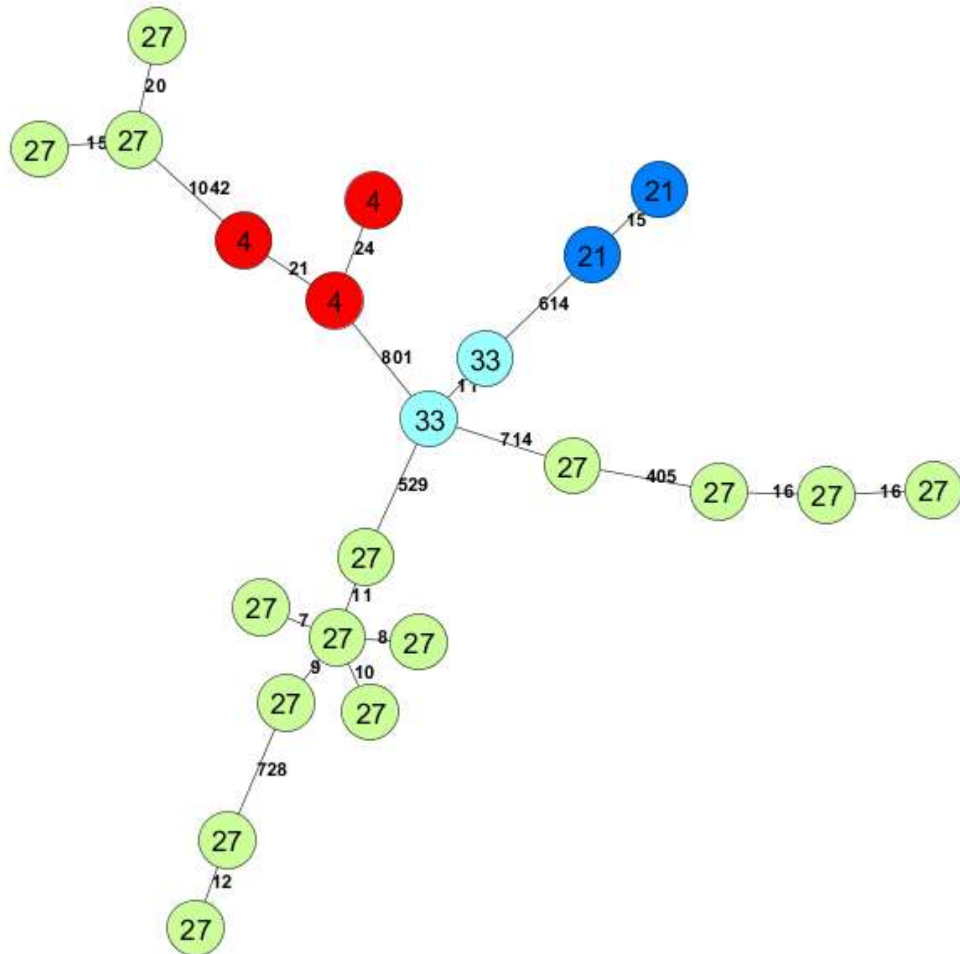

Figure 46. Minimum spanning tree of ST6 isolates from Farms 4, 21, 27 & 33. Line distances are scaled log base 10 representing substitutions per nucleotide. Nodes are colour grouped according to farm of origin and label nomenclature is Farm-ST. Significant diversity is evident with the Farm 27 population with three clearly separated clades.

All isolate distance matrix values were assigned to one of the four following, mutually exclusive groups based upon the characteristics of the isolates for statistical analysis of relative genetic heterogeneity by farm and by MLST sequence type.

- Same Farm (Fm) & Same Sequence Type (ST)
- Same Farm (Fm) & Different Sequence Type (ST)
- Different Farm (Fm) & Same Sequence Type (ST)

- Different Farm (Fm) & Different Sequence Type (ST)

The matrix distances for isolate similarity were not normally distributed within each category. A 95% confidence interval about the median distance was used to compare inter and intra farm and inter and intra sequence type (ST) heterogeneity and a Mood median test was performed in Minitab17 for nonparametric statistical comparison of the categories. This test was selected due to the dissimilar shapes of the underlying distributions of the categories Table 19 & Figure .

Table 19. Median and 95% confidence intervals of distance matrix values between cgMLST isolates grouped by a 2 x 2 factorial design according to their farm or origin (Same Farm or Different Farm) and their original MLST sequence type (Same ST or Different ST). In addition ST6 isolates from Farm27 were considered as a separate group as well as integrated into the main analysis.

Category (in ascending order of similarity, least to most)	Median Distance matrix value	95% confidence interval of the median		
		Lower	Upper	Range
Same Farm & Different ST	0.68	0.67	0.68	0.01
Different Farm & Different ST	0.53	0.53	0.54	0.01
Different Farm & Same ST	0.50	0.50	0.50	0.01
ST6 from Farm 27	0.47	0.45	0.48	0.03
Same Farm & Same ST	0.22	0.18	0.45	0.28
Same Farm & Same ST (Excluding Fm27, ST6 isolates)	0.01	0.01	0.06	0.05

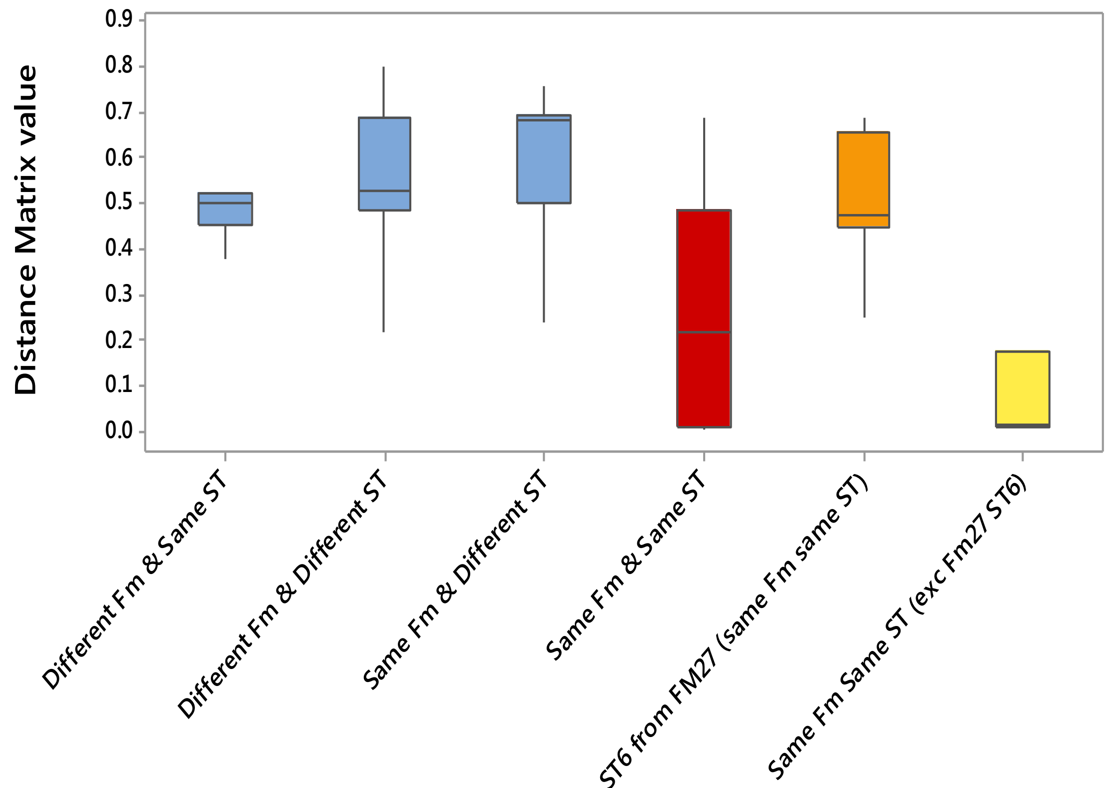

Figure 47. Distributions of cgMLST distance matrix values between cgMLST isolates grouped by a 2 x 2 factorial design according to their farm or origin (Same Farm or Different Farm) and their original MLST sequence type (Same ST or Different ST). In addition ST6 isolates from Farm27 were considered as a separate group as well as integrated into the main analysis. ST6 from Farm27(Orange) is displayed as a subset of SameFm&SameST (Red) and the remainder of that group with the Farm27 ST6 isolates removed is displayed on the right as Same FM Same ST6 (exc FM27 ST6) (Yellow). Vertical black line = 95% confidence interval, coloured box = interquartile range, horizontal line = mean.

The most genetically heterogeneous category was “Same Fm & Different ST” which had a significantly higher median distance than any other group. The “Different Fm & Different ST” group had the next highest median distance and was significantly higher than the “Different Fm & Same ST” group.

The most genetically similar isolate category with the lowest median distance value between isolates within the category, was the "Same Fm & Same ST" category. However, the "Same Farm & Same ST" category had the widest confidence interval range of the four groups due primarily to the variation between ST6 isolate clades from Farm 27. When these Fm27 ST6 isolates were excluded from the "Same Farm & Same ST" category the confidence interval range around the median was comparable to the ranges for the other three categories and the median value of the Same Farm & Same ST group (excluding Farm27 ST6 isolates) decreased by 93.1% indicating a significantly greater genetic similarly within this group compared to any of the other four groups ($p < 0.0001$) see Figure 47.

6.3.3 Aim 3: Determine if the genetic heterogeneity between potentially contagious sequence types within a herd correlates with the isolate MALDI spectral profiles.

Statistical comparisons were conducted on cgMLST distance matrices and MALDI CCI matrices of isolates from Farm 27 of ST6, ST20 and ST233 for which data in both modalities was available. The genetic relationships were visualised graphically using a minimum spanning tree Figure 48. The underlying data distributions for each category in each modality were assessed graphically. The distributions were not normally distributed and the distribution shapes varied, therefore a Mood's median test was used for statistical comparison.

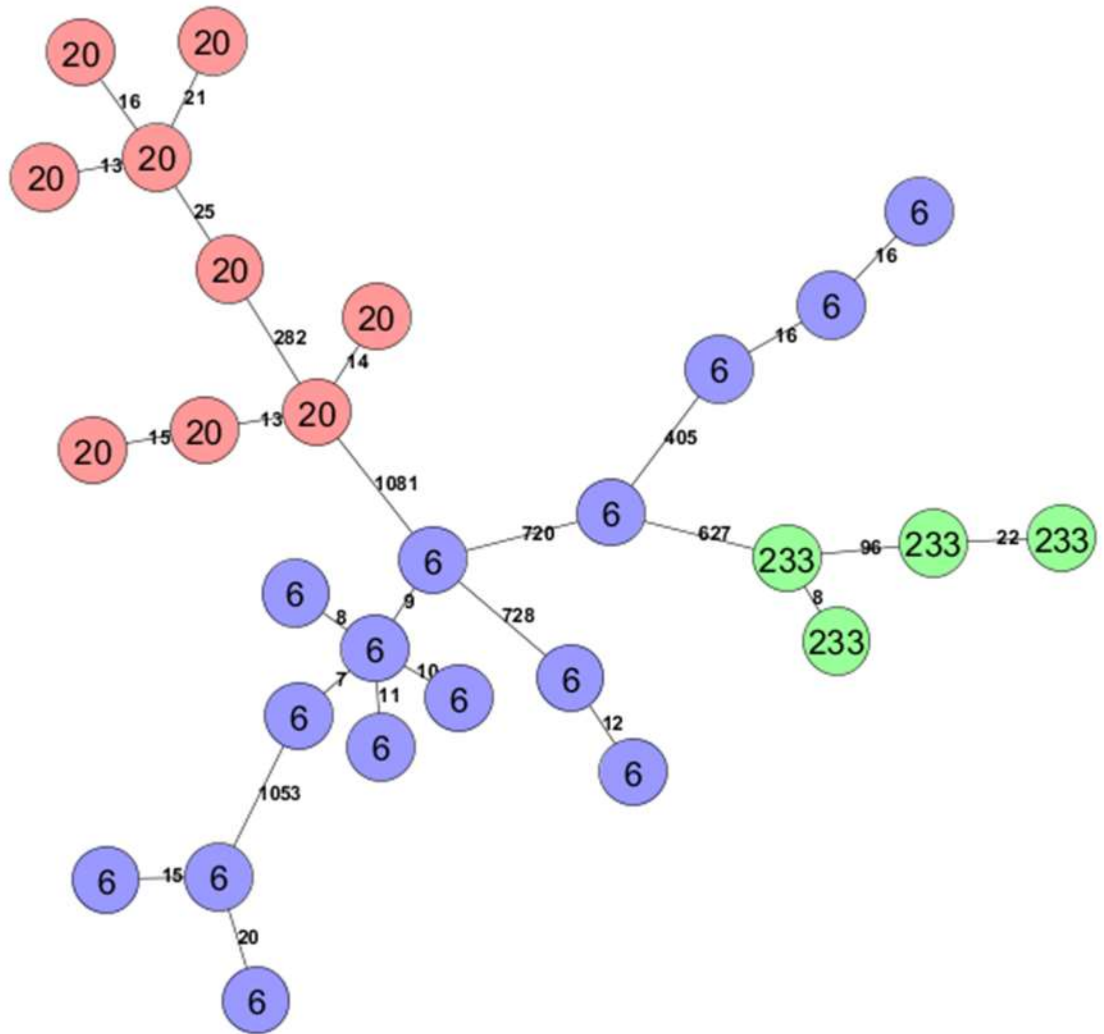

Figure 48. Minimum Spanning Tree of isolates selected from Fm27, node labels refer to the MLST sequence types ST6, ST20 and ST233. Line distances are scaled to log10 representing substitutions per nucleotide. ST233 is a single locus variant of ST6, ST20 is a five locus variant of ST6 and ST233 is a four locus variant of ST20 according to the classical MLST scheme.

Within Farm 27 analysis of cgMLST data identified substantial variation within the MLST ST6 group that is not apparent in any of the other sequence types Figure 48. The ST6 population from Farm 27 can be divided into three distinct subgroups based upon the cgMLST clustering. In contrast all isolates of ST20 form a very distinct separate clade both within Farm27 and also within the whole dataset of 62 isolates. There is only evidence of one clade of ST233

within farm 27 however; another ST233 isolate from Farm15 was identified as genetically quite dissimilar Figure 44 & 45. The similarity between and within the MLST ST groups was described statistically using 95% confidence intervals about the median distance matrix value for each group.

To assess the similarity between isolates of the same ST group median distance matrix distributions were calculated for each of the ST groups ST6, ST20 and ST233. To assess the similarity between to ST groups, median distance matrix distributions were also calculated for each of the pairwise combinations of ST group isolates ST6-ST20, ST6-ST233, ST20-ST233 (Table 20 & Figure 49). Of the three single ST groups the ST6 group was the most divergent with a significantly higher median distance value than either ST20 or ST233. All three inter-ST groups: ST6-ST20, ST6-ST233 and ST20-ST233 have significantly higher median distance values than either the ST20 or ST233 group. The ST6-ST233 group had a significantly lower median distance value than either of the other two inter-ST groups.

Table 20. Mood median test: cgMLST distance matrix Fm27 ST6, ST20 & ST233 (Minitab17). Isolate combinations of the same ST group shown in black and isolates from different ST groups of each possible combination shown in red. Low confidence interval median distance matrix distributions indicate greater similarity of the gDNA sequence.

Factor	N≤ Overall median	N> Overall median	Median	Interquartile range
ST6	86	19	0.47	0.21
ST20	36	0	0.18	0.17
ST233	6	0	0.06	0.05
ST6-ST20	0	91	0.69	0.02
ST6-ST233	20	3	0.50	0.17
ST20-ST233	0	34	0.70	0.01

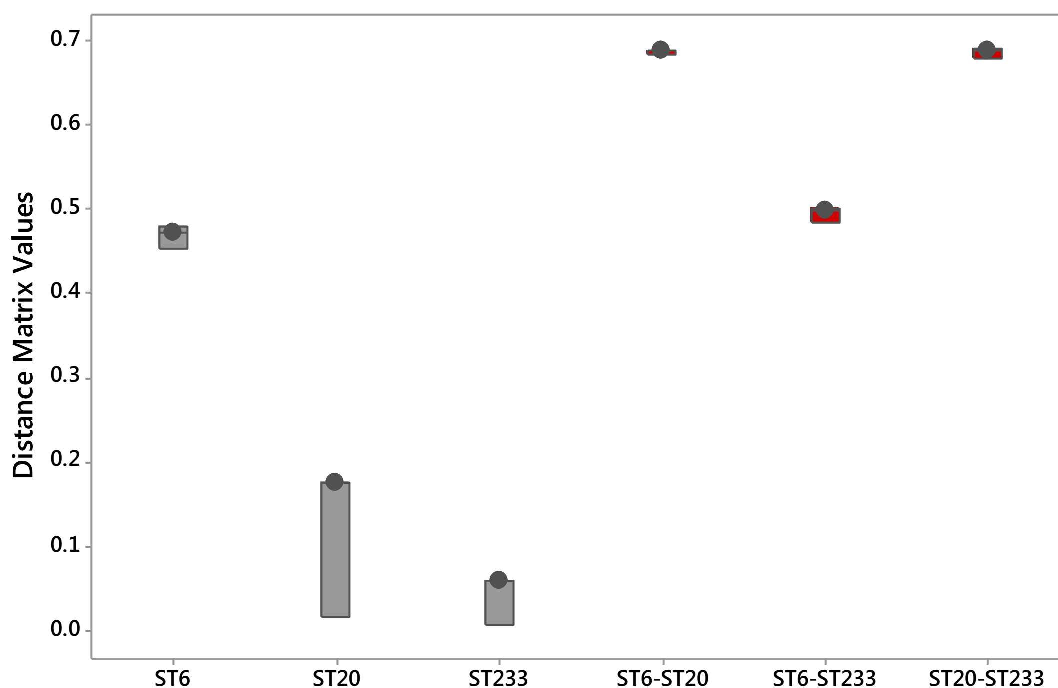

Figure 49. 95% confidence intervals about the median of cgMLST distance matrix values for three populations of MLST sequence types (ST6, ST20 and ST233) (grey) and of three combinations of all isolates from different sequence types (red) as a means of comparing heterogeneity within sequence types and between sequence types. Low values indicate greater similarity of the genotypes. Coloured box = 95% confidence interval, horizontal line = mean, black dot = median.

To compare the heterogeneity in MALDI protein expression between isolates of different sequence types within a herd quantitative analysis was carried out using pairwise CCI scores using the same categories defined for the analogous cgMLST distance matrix analysis. Mood median test was used as the underlying CCI values were not normally distributed and the shape of the distributions varied. See Table 21 & Figure 50.

Table 21. Mood Median Test: MALDI CCI values Fm27 ST6,20,233. Quantitative analysis of within and between ST phenotype heterogeneity using MALDI protein mass spectra. Higher CCI values indicate greater similarity between the isolates in the comparison. The single ST group CCI results are displayed in black and the inter-ST CCI comparisons are displayed in red for each combination.

Factor	N≤ Overall Median	N> Overall Median	Median CCI	Interquartile range
ST6	35	30	0.93	0.07
ST20	2	13	0.97	0.02
ST233	4	4	0.94	0.05
ST6-ST20	76	54	0.93	0.07
ST6-ST233	24	28	0.94	0.07
ST20-ST233	14	26	0.96	0.06

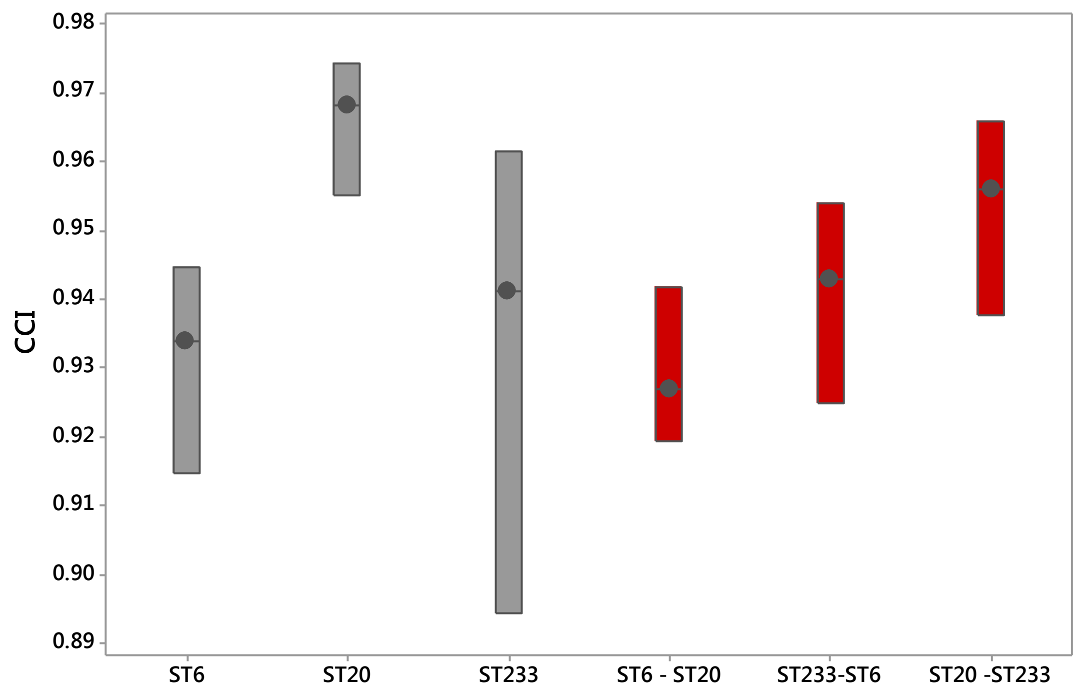

Figure 50. 95% confidence intervals about the median of MALDI CCI values for three populations of MLST sequence types (ST6, ST20 and ST233) (grey) and of three combinations of all isolates from different sequence types (red) as a means of comparing heterogeneity within sequence types and between sequence types. High values indicate greater similarity of the MALDI spectral profile phenotypes. Coloured box = 95% confidence interval, horizontal line = mean, black dot = median.

The highest median CCI score and the narrowest 95% confidence interval was observed in the ST20 group. The median CCI of the ST20 group was significantly higher that of the ST6 group, which the lowest median CCI of any of the single ST groups and a wider 95% confidence interval than ST20. The ST233 group contributed fewer isolates than ST6 or ST20, therefore fewer possible isolate combinations for CCI were available for analysis. The median CCI for ST233 was intermediate between the other two single ST groups. However, the confidence interval for ST233 was the broadest, encompassing the 95% of ST6 and overlapping that of ST20.

The inter–ST factors use CCI scores to estimate relative similarity between three paired combinations of groups of isolates based upon their different sequence types. The higher the median CCI score the more similar the two isolates in the comparison according to the MALDI spectra. The analysis demonstrates that ST20 vs ST233 CCI scores were significantly higher than ST20 vs ST6 CCI scores. The ST233 vs ST6 CCI distribution was not significantly different from either of the other two.

6.3.4 Aim 4: Determine if isolates classified as persistent clinical cases in the same mammary gland according to MLST sequence type are true persistent infections or new infections with the same sequence type.

Twenty seven isolates collected from thirteen cows on eight farms, involving eight sequence types were classified as emanating from a persistent infection (Figure 51). Persistent mastitis infections were defined as the isolation of the same ST from the same mammary quarter on more than one occasion (12 x two occasions and 1 x 3 occasions). These specific persistent cow cases were selected for whole genome sequencing to simultaneously provide a broad sample across multiple farms and multiple sequence types.

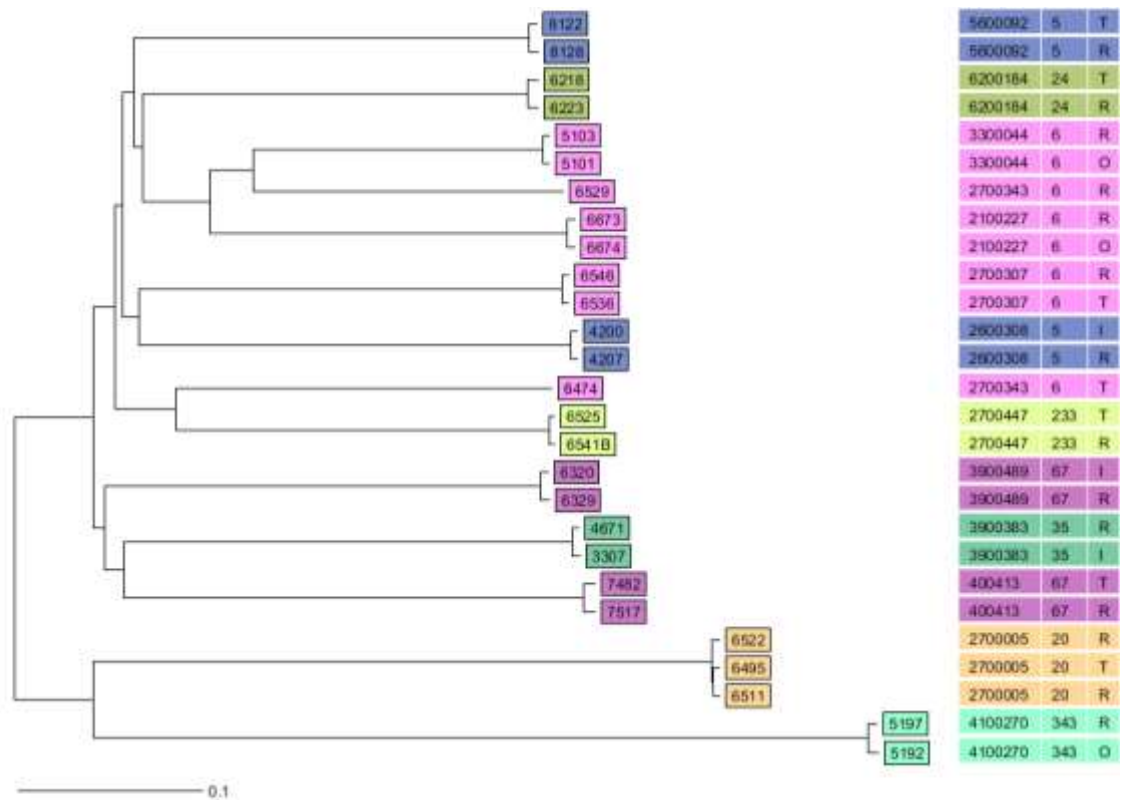

Figure 51. Dendrogram of 27 persistent clinical mastitis case isolates using 1622 cgMLST target loci. Isolates are grouped by colour according to their previously defined MLST sequence type ST number. Dendrogram labels indicate unique sample id number. Columns indicate (left to right) Individual cow id, MLST sequence type and case classification (I = Index case, T = Potential Transmission Event(PTE), R = repeat isolation of the same ST from the same mammary gland, O = Unclassified case). The horizontal scale representing substitutions per nucleotide.

Twelve of the thirteen cow cases clustered both (and in one case all 3) isolates from that case at the same terminal and shared that terminal node exclusively with isolates from that persistent cow case as anticipated for a true persistent infection. This clustering pattern was repeated in the cgMLST analysis of all 62 isolates. One pair of ST6 isolates (id 6474 & 6529) were significantly more genetically diverse, separated by 6 nodes and significantly higher distance matrix value than all other combinations of potentially persistent isolates see Table 22.

Eighteen of the persistent cow case isolates selected for cgMLST analysis also underwent MALDI analysis. These isolates were drawn from ten persistent cow cases, originating from five herds. MALDI CCI scores were calculated within their herds of origin between those persistent case isolates and other isolates from those herds which had not been classified as Persistent.

Within each farm cases were assigned a CCI score for each unique isolate combination. The CCI score for combinations of isolates identified by cgMLST as likely persistent cases were compared to the CCI scores for each of the putative persistent isolates with all other (non-Persistent) isolates within their own farm, see Table 22.

Table 22. cgMLST distance matrix and MALDI CCI score for each persistent isolate combination. The Isolate pair in red was originally classified as Persistent by MLST but appears to be two separate cgMLST genotypes and dissimilar MALDI phenotypes with significantly higher genetic distance value and significantly lower CCI value.

Persistent Isolates		Distance Matrix value Based upon SNP counts	MALDI CCI	FARM of origin	MLST Sequence Type
4200	**4207**	0.0062	0.999	26	5
6529	**6474**	0.4712	0.860	27	6
6536	**6546**	0.0074	0.938	27	6
6525	**6541**	0.0049	0.935	27	233
6495	**6511**	0.0080	0.969	27	20
6511	**6522**	0.0087	0.990	27	20
6495	**6522**	0.0099	0.951	27	20
5101	**5103**	0.0068	0.996	33	5
6320	**6329**	0.0080	0.958	39	67
8122	**8128**	0.0081	0.916	56	5
6218	**6223**	0.0105	**No paired MALDI data**	62	24
6673	**6674**	0.0093		21	6
7482	**7517**	0.0130		4	67
5197	**5192**	0.0111		41	343

The CCI score for 6529-6474 was significantly lower than that of the other combinations. The isolate pair 6529-6474 was excluded from further analysis due to the decision that they did not represent a persistent infection. All CCI were assigned to one of two CCI groups were designated for each farm as (P) Persistent and (Non-P) non-Persistent. The distribution of CCI values deviated from the normal in both groups.

Within herd median CCI scores were significantly higher between isolates that formed a persistent cow case (P group) than between two non-persistent isolates (p=0.04) see Figure 52.

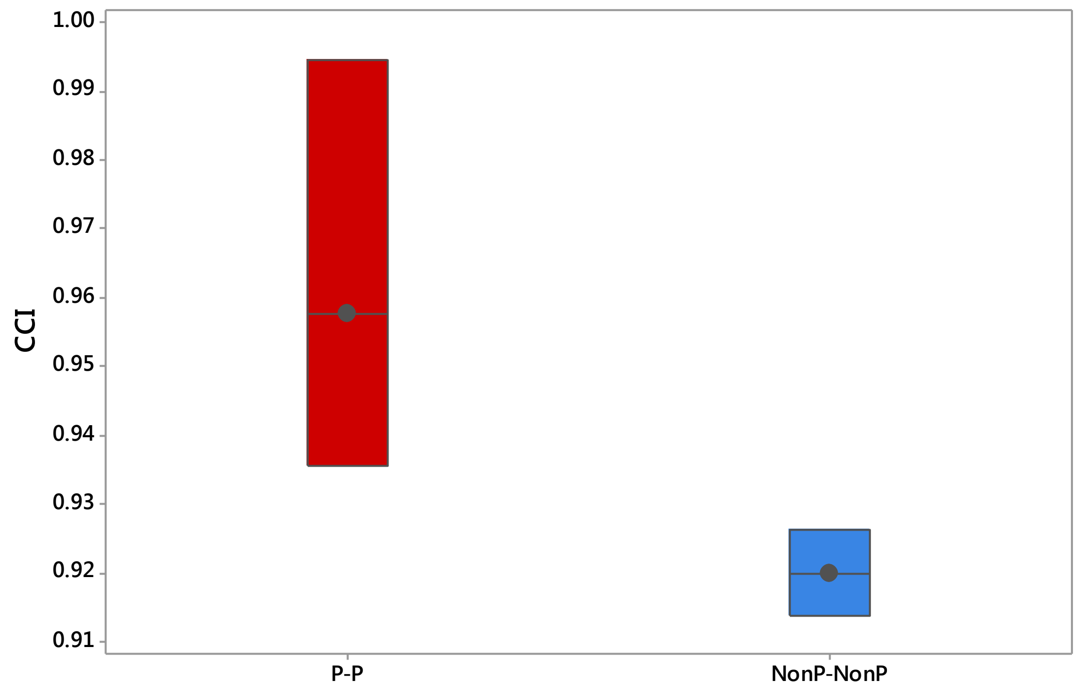

Figure 52. 95% confidence interval plot of the median MALDI CCI values between all pairs of isolates classified as persistent clinical cases (red) and MALDI CCI values for all pairs of isolates that were not classified as persistent cases (blue). Coloured box = 95% confidence interval, horizontal line = mean, black dot = median.

6.3.5 Aim 5: Determine if there is evidence of antigenic variation of three virulence genes (SUB 0145/RS00865, SUB 1095/RS5505, SUB 1154/RS5795) as a possible immune escape mechanism in persistent mastitis case.

Variation at three gene targets of known importance in mastitis virulence (SUB 0145/RS00865, SUB 1095/RS5505 and SUB 1154/RS5795) were used to investigate possible evidence of antigenic variation as mechanism for immune

evasion in persistent infections. Dendrograms were generated in Ridom SeqSphere individually for each locus for all isolates as a means of initially assessing variability within the population and specifically the relative heterogeneity between the three loci Figure 53 – 55.

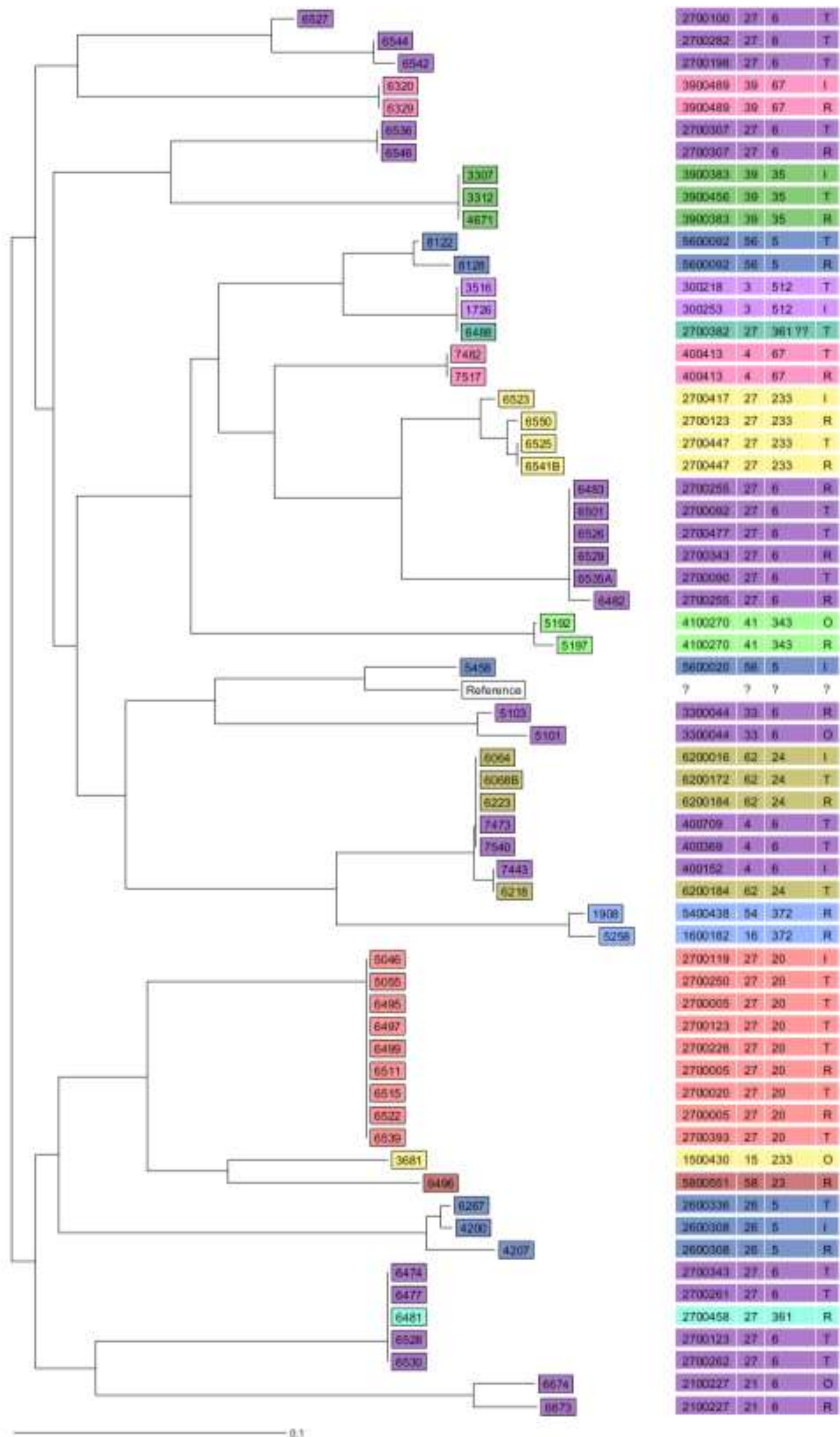

Figure 53. Neighbour joining tree dendrogram of all cgMLST *S.uberis* isolates based upon 132 SNP positions at the locus (SUB 0145/RS00865). Scale represents the genetic distances in substitutions per nucleotide Isolates of the same MLST sequence type are grouped by colour. Sample id at the end of each branch. Columns (left to right): Cow Id, Farm id, MLST ST, Clinical Case Classification.

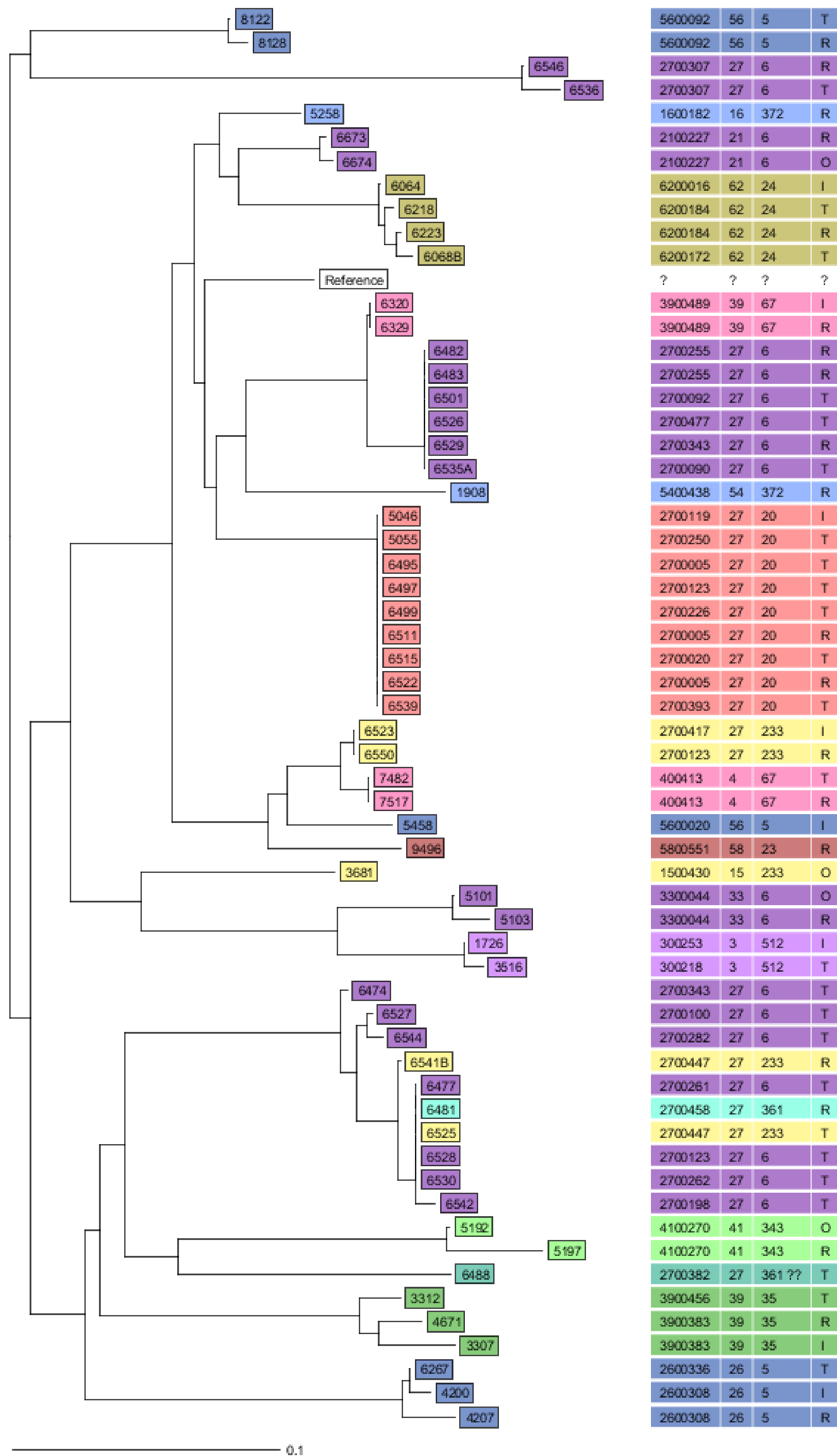

Figure 54. Neighbour joining tree dendrogram of all cgMLST *S.uberis* isolates based upon based upon 67 SNP positions at the locus (SUB 1095/RS 5505). Scale represents the genetic distances in substitutions per nucleotide Isolates of the same MLST sequence type are grouped by colour. Sample id at the end of each branch. Columns (left to right): Cow Id, Farm id, MLST ST, Clinical Case Classification.

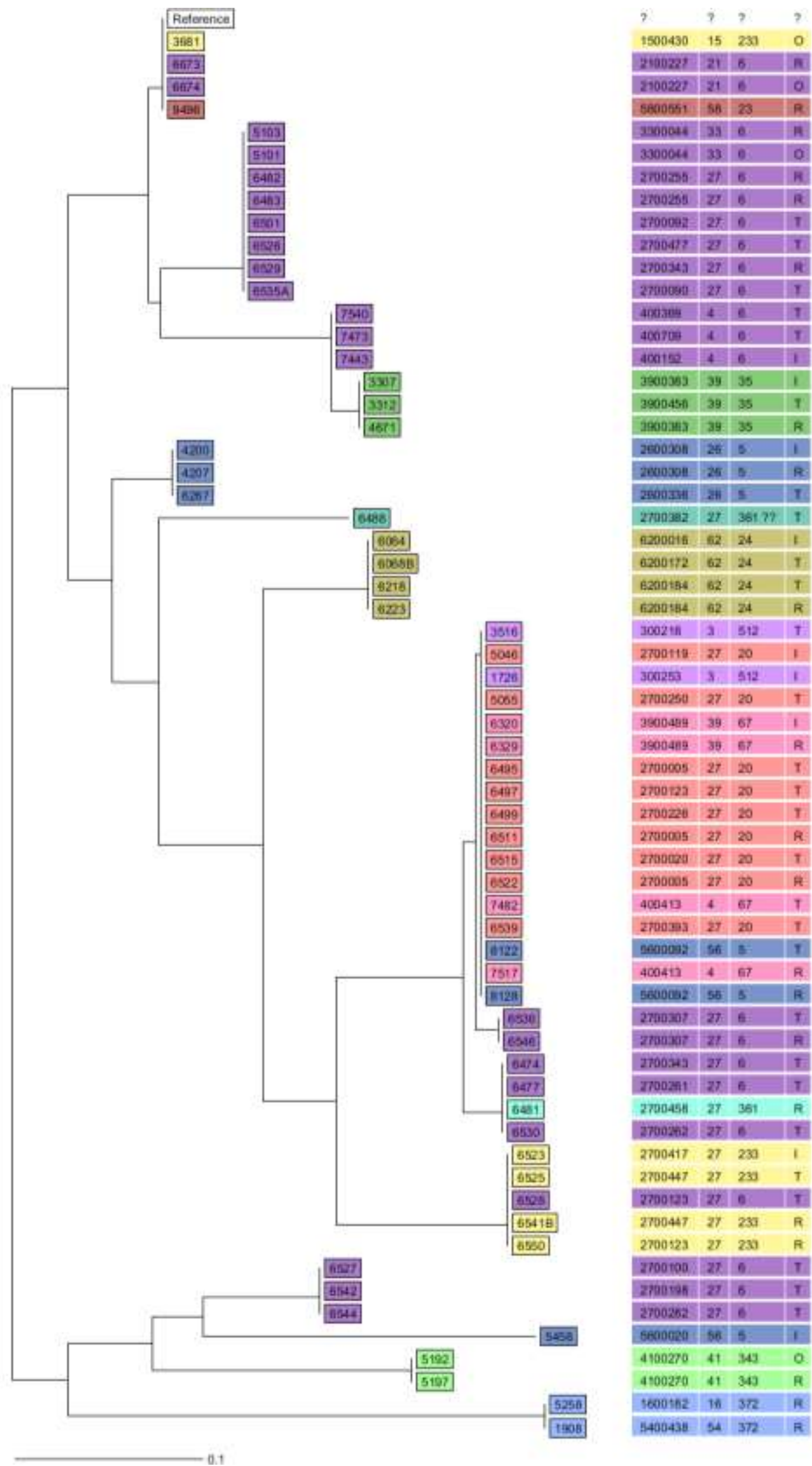

Figure 55. Neighbour joining tree dendrogram of all cgMLST *S.uberis* isolates based upon 134 SNP positions at locus (SUB 1154/RS 5795). Scale represents the genetic distances in substitutions per nucleotide Isolates of the same MLST sequence type are grouped by colour. Sample id at the end of each branch. Columns (left to right): Cow Id, Farm id, MLST ST, Clinical Case Classification.

Dendrogram tree complexity varied substantially between the three loci with 31, 36 and 14 dividing nodes respectively. The degree of variation within MLST groups is less than is seen in the cgMLST analysis however, ST6 is again the most diverse ST with isolates in several clades for each locus. ST20 in contrast showed no variation between isolates at any of the loci. There was no discernible clustering by farm of origin.

Of the twelve plausible persistent infections SNP frequency and position at three loci; SUB 1154/RS5795, SUB 1095/RS5505 and SUB 0145/RS00865 were analysed.

Three infections were true clones, with no SNPs in any of the three loci between their two isolates. Two infections only had synonymous SNPs at either SUB 1095/RS05505 or SUB 0145/RS00865.

The SUB 1154/RS05795 locus was conserved in all persistent infections with no SNP between isolates from the same infection.

Non-synonymous, missense SNPs were identified between isolates from the same infection at SUB 1095/RS05505 and SUB 0145/RS00865 in the remaining seven infections.

Three infections had non-synonymous SNPs in SUB 1095/RS05505 and SUB 0145/RS00865.

Six infections had non-synonymous missense SNPs in SUB 0145/RS00865 Table 23, Table 24 & Figure 56. Three of these SNP's were in all at the same base pair position (126) at the distal portion and the transmembrane region

furthest from the cell wall. Three infections had non-synonymous missense SNPs in SUB 0145/RS00865 and no SNP at SUB 1095/RS05505. Four infections had non-synonymous missense SNPs in SUB 1095/RS05505. One infection had a non-synonymous missense SNP in SUB1095/RS05505 and no SNPs at SUB 0145/RS00865. All but one of the non-synonymous SNPs at SUB1095/RS05505 were in the non-coding region of the sequence, see Table 25 & 26 & Figure 57.

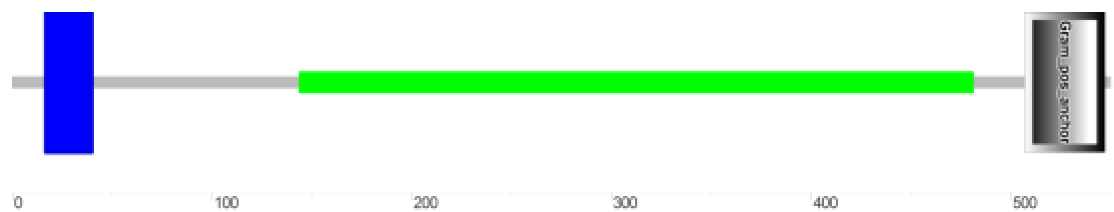

Figure 56. pfam diagram of the structure of SUB 0145/RS00865. Regions correspond to mapped areas described in Table 23 and the numeric scale refers to codon number.

Table 23. Protein structure diagram of SUB 0145/RS00865 reproduced from pfam.

Name of region	Start codon	End codon
Transmembrane region	17	39
Unknown protein region	40	143
Coiled region	144	480
Unknown protein region	481	506
Gram positive anchor	507	547

Table 24. Table of all non-synonymous SNP's in SUB 0145/RS865 from isolates of persistent clinical cases including the base change; amino acid substitution; codon position and the Pfam region of the SNP.

Persistent *S.uberis* clinical cases isolate ID's		Number of non-synonymous SNPs in SUB 0145/RS00865	Amino Acid substitution	Base change	Codon Position from N to C Terminus	Pfam Region
8812	8128	2 SNP in 1 codon	Ala:Val	C:T G:T	286	coiled
6218	6223	1	Asp:Glu	A:T	42	Unknown
5101	5103	2 SNP in 1 codon	Ala:Gly	C:G, A:G	317	coiled
6673	6674	3 SNP in 1 codon	Lys:Ala	C:A, C:A, G:A	356	coiled
4200	4207	1	Asp:Glu	A:T	42	Unknown
5192	5197	1	Asp:Glu	A:T	42	Unknown

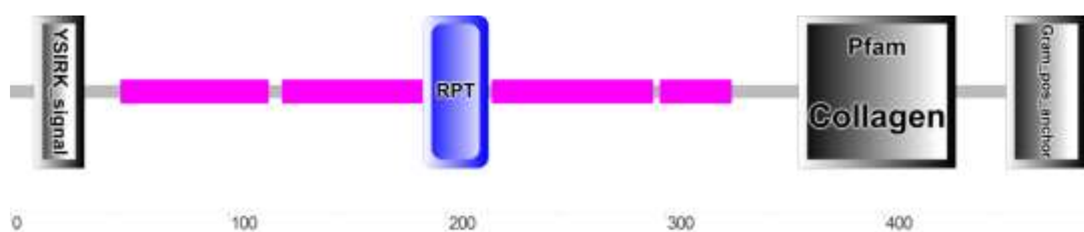

Figure 57. pfam diagram of SUB1095/RS5055. Regions correspond to mapped areas described in Table 25.

Table 25. Protein structure diagram of SUB 1095/RS5505 reproduced from pfam.

Name of region	Start codon	End codon
Pfam:YSIRK_signal	11	34

Pfam:Collagen	89	152
Pfam:Collagen	113	201
Pfam:Collagen	278	338
low complexity	369	397
low complexity	402	415
low complexity	418	429
Pfam: Gram positive anchor	438	475

Table 26. Table of all non-synonymous, missense SNP's in persistent clinical cases in SUB 1095/RS5505 including the base change; amino acid substitution; codon position and the Pfam region of the SNP.

Persistent *S.uberis* clinical cases isolate ID's		Number of non-synonymous SNPs in SUB 1095/RS05505	Amino Acid substitution	Base change	Codon Position	Pfam Region
6218	6223	1	His:Gln	C:A	443	Gram positive anchor
4200	4207	2 SNP in 1 codon	Lys:Gln	G:T, A:G	185	Collagen
5192	5197	2 SNP in 2 codons	His:Arg, Asn:Lys	T:C, A:C	117, 249	Collagen
6525	6541b	1	Asp:Asn	C:T	213	Collagen

Table 27. Amino acid sequences of SUB 0145/RS 00865 from six isolates form three persistent clinical mastitis cases (4200 & 4207, 5192 & 5197, 6218 & 6223) with the same amino acid substitution (Asp, Aspartic acid D/Glu, Glutamate E) at codon 42 highlighted in green.

4200	4207
MEIKQKHRKHALRKAVTAAVLAGTAFSS LGGFAGAVTTVKAEDLFTINNSEVQDKLE SKVKQLLEAQRKGEDISEKLRELLSELPTDI LKDIMLSNIEADYLLGFLKPAVEEMVRRS EQNDERWKDITEKTLALEALKDSEREIRK EKEKLEDEVQLAKVKIETKESELNDLKKDY IDTREELADTIEELDEVKNSIEEKEAKVKGL EEKLRDLEKELGDYDKKLSEAAKQNSDLS NENKELKENLDTAENITVELQKKSHELEK TKKEVELELKAEKEALEAEKVKLAEANEA NDKLSEERDAAKKEAEKVPELEEQVEKLV EEITNAKKEAEELQAKAEGLEKDFEAVKA EKEALEAEIAKLKEDHQKEVDALNALLAD KEKMLKNLQDQLDKAKEEAMKNEQMS QEEKAKLQAELDKAKQELAEKIKDMPNK VAPQAEGKANAGQAAPNQNQNNQAQ ANQTKNGNNLPSTGDKPVNPLLVASGLS LMIGAGAFVYAGKRKKG*	MEIKQKHRKHALRKAVTAAVLAGTAFSS LGGFAGAVTTVKADDLFTINNSEVQDKL ESKVKQLLEAQRKGEDISEKLRELLSELPT DILKDIMLSNIEADYLLGFLKPAVEEMVR RSEQNDERWKDITEKTLALEALKDSEREI RKEKEKLEDEVQLAKVKIETKESELNDLKK DYIDTREELADTIEELDEVKNSIEEKEAKVK GLEEKLRDLEKELGDYDKKLSEAAKQNSD LSNENKELKENLDTAENITVELQKKSHELE KTKKEVELELKAEKEALEAEKVKLAEANE ANDKLSEERDAAKKEAEKVPELEEQVEKL VEEITNAKKEAEELQAKAEGLEKDFEAVK AEKEALEAEIAKLKEDHQKEVDALNALLA DKEKMLKNLQDQLDKAKEEAMKNEQM SQEEKAKLQAELDKAKQELAEKIKDMPN KVAPQAEGKANAGQAAPNQNQNNQA QANQTKNGNNLPSTGDKPVNPLLVASG LSLMIGAGAFVYAGKRKKG*
5192	**5197**
MEIKQKHGKHALRKAVTAAVLAGTAFSS LGGFAGAVTTIKAEDLFTINNSEVQDKLE SKVKQLLEAQRKGEDISEKLRELLSELPTDI LKDIMLSNIEADYLLGFLKPAVEEMVRRS EQNDERWKDITEKTLALEALKDSEREIRK EKEKLEDEVQLAKVKIETKESELNDLKKDY IDTREELADTIEELDEVKNSIVEKEAKVKG LEEKLRDLEKELGDYDKKLSEAAKQNSDL SNENKELKENLDTAENITVELQKKSHELE KTKKEVELELKAEKEALEAEKAKLAEAKEA NDKLTEERDAAKKEAEKVPELEEQVEKLV EEITAAKKEAEDLQAKAEGLEKDFEAVKA EKEALEAEIAKLKEDHQKEVDALNALLAD KEKILKNLQDQLDKAKEEAMKNEQMSQ	MEIKQKHGKHALRKAVTAAVLAGTAFSS LGGFAGAVTTIKADDLFTINNSEVQDKLE SKVKQLLEAQRKGEDISEKLRELLSELPTDI LKDIMLSNIEADYLLGFLKPAVEEMVRRS EQNDERWKDITEKTLALEALKDSEREIRK EKEKLEDEVQLAKVKIETKESELNDLKKDY IDTREELADTIEELDEVKNSIVEKEAKVKG LEEKLRDLEKELGDYDKKLSEAAKQNSDL SNENKELKENLDTAENITVELQKKSHELE KTKKEVELELKAEKEALEAEKAKLAEAKEA NDKLTEERDAAKKEAEKVPELEEQVEKLV EEITAAKKEAEDLQAKAEGLEKDFEAVKA EKEALEAEIAKLKEDHQKEVDALNALLAD KEKILKNLQDQLDKAKEEAMKNEQMSQ

EEKAKLQAELDKAKQELADKIKDMPNKV APQAEGKANAGQAAPNQNQNNQAQA NQAKNANNLPSTGDKPVNPLLVASGLSL MIGAGAFVYAGKRKKG*	EEKAKLQAELDKAKQELADKIKDMPNKV APQAEGKANAGQAAPNQNQNNQAQA NQAKNANNLPSTGDKPVNPLLVASGLSL MIGAGAFVYAGKRKKG*
6218 MEIKQKHGKHALRKAVTAAVLAGTAFSS LGGFAGAVTTVKADDLFTINNSEVQDKL ESKVKQLLEAQRKGEDISEKLRELLSELPT DILKDIMLSNIEADYLLGFLKPAVEEMVR RSEQNDERWKDITEKTLALEALKDSEREI RKEKEKLEDEVQLAKVKIETKESELNDLKK DYIDTREELADTIEELDEVKNSIEEKEAKVK GLEEKLRDLEKELGDYDKKLSEAAKQNSD LSNENKELKENLDTAENITVELQKKSHELE KTKKEVELELKAEKEALEAEKVKLAEANE ANDKLSEERDAAKKEAEKVPELEAQVEKL VEEITTAKKEAEELQAKAEALEKDFEAVK AEKEKLEAEIAKMKEDHQKEVDALNALL ADKEKMLKSLHEQLDKAKEEAMKNEQM SQEEKAKLQAELDKAKQELGDKIKDMPN KVAPQAEGKANAGQAAPNQNQNNQA QANQAKNANNLPSTGDKPVNPLLVASG LSLMIGAGAFVYAGKRKKG*	**6223** MEIKQKHGKHALRKAVTAAVLAGTAFSS LGGFAGAVTTVKAEDLFTINNSEVQDKLE SKVKQLLEAQRKGEDISEKLRELLSELPTDI LKDIMLSNIEADYLLGFLKPAVEEMVRRS EQNDERWKDITEKTLALEALKDSEREIRK EKEKLEDEVQLAKVKIETKESELNDLKKDY IDTREELADTIEELDEVKNSIEEKEAKVKGL EEKLRDLEKELGDYDKKLSEAAKQNSDLS NENKELKENLDTAENITVELQKKSHELEK TKKEVELELKAEKEALEAEKVKLAEANEA NDKLSEERDAAKKEAEKVPELEAQVEKLV EEITTAKKEAEELQAKAEALEKDFEAVKAE KEKLEAEIAKMKEDHQKEVDALNALLAD KEKMLKSLHEQLDKAKEEAMKNEQMSQ EEKAKLQAELDKAKQELGDKIKDMPNKV APQAEGKANAGQAAPNQNQNNQAQA NQAKNANNLPSTGDKPVNPLLVASGLSL MIGAGAFVYAGKRKKG*

6.4 Discussion

The first objective of this analysis was to evaluate genetic and phenotypic heterogeneity in MLST sequence type isolates associated with potential contagious transmission. The analysis was divided into three specific aims:

1. Validate classical MLST scheme sequence type classifications by cgMLST

2. Determine the genetic heterogeneity of MLST sequence types within a single herd and between multiple herds using cgMLST analysis.
3. Determine if the genetic heterogeneity between potentially contagious sequence types within a herd correlates with heterogeneity in isolate MALDI spectral profiles.

The results obtained addressed the first, basic aim of the whole genome sequencing, reproducing very accurately the MLST sequence type results of the large scale sanger sequencing analysis. Whole genome sequencing and cgMLST analysis was used to address the second aim to investigation of the diversity within and between sequence types and herds than was previously possible and allowing evaluation of the heterogeneity of MLST sequence types within a single herd and between multiple herds.

The four way contingency comparison (Farm, Sequence type and Same, Different) in Table 19 & Figure 7. Shows that the 'Farm of Origin' discriminates isolates almost as well as the MLST sequence type. This supports the finding of the farm level MALDI modelling which identified strong farm level discrimination ability within case classification groups. The results of this cgMLST analysis demonstrate that the farm specific variability transcends the MLST sequence type definitions between PTE associated isolates.

This divergent evolution of sequence types was not unexpected given the high recombination rate of *S.uberis* described by (Coffey et al 2006) and a

diverse population of *S.uberis* present in each farm environment, in multiple potential reservoirs including: pasture; bedding; gastrointestinal tracts and mammary glands. In contrast the isolates of the same sequence type from a given herd were, genetically very homogeneous (Figure 47). This in turn could be the result of a very limited population of bacteria present in the mammary gland restricting the opportunity for recombination between bacteria. In the period between the isolation of one clinical case and contagious transmission of that bacterial isolate, via milking equipment for example to a new host from which the next clinical case isolate was recovered, there would be very little opportunity for recombination between sequence types unless multiple sequence types were present as a mixed infection. Therefore the pattern of clinical cases which we see, especially in Farm 27 with ST6 isolates is what we would expect from the 'low grade' contagious transmission which we hypothesised we occurring in (Davies et al. 2016).

This data indicated that the classical MLST scheme is largely appropriate as a molecular epidemiological tool for *S.uberis* in multi herd studies however, there are instances where the discriminatory power of MLST varies significantly between ST's within a herd and should be considered when interpreting results or when considering the appropriateness of using a more discriminatory technology such as cgMLST.

The diverse clade structure of ST6 in farm 27 was particularly interesting and could be interpreted as supporting the hypothesis of low grade contagious transmission suggested previously (Davies et al. 2016) as a manifestation of a

range of environmental isolates acquiring an enhanced ability to transmit between a limited number of cows for a short period rather than a single, dominant, highly competitive isolate being propagated throughout the herd to the exclusion of all others.

In order to fulfil the aim of addressing a number of separate hypothesis relating to both the population level heterogeneity and also the individual heterogeneity of specific virulence genes the selection of a limited number of isolates for whole genome sequencing involved non-random selection of isolates. The selection of isolates for cgMLST analysis included a relatively high proportion of isolates assumed to be persistent infections (isolated from the same mammary gland and the same MLST ST on two occasions) compared to the prevalence of such cases in the original study. The increased likelihood of these persistent cases being genetically very similar was considered a potential source of bias in the assessment of heterogeneity within and between herds and within and between sequence types. To account for this the analysis was conducted twice, first with all isolates contributing to the categories and second with all isolates except the second member of each persistent case pair. The results of both analyses were comparable and results were significant in both case. It therefore appears reasonable to conclude that the high proportion of persistent cases included in the selection for cgMLST analysis does not bias the results in this respect.

The selection of exclusively PTE associated clinical cases in this sample may present a bias to the assessment of heterogeneity between herds by cgMLST

as low prevalence sequence types were not sequenced for the cgMLST analysis. The interpretation provided here of the relative heterogeneity within and between herds and the relative heterogeneity within and between sequence types which had been previously identified as disproportionately prevalent and potentially being transmitted in a contagious manner.

The third aim was to determine if the genetic heterogeneity between potentially contagious sequence types within a herd correlates with heterogeneity in isolate MALDI spectral profiles. For this aim

A comparison between MALDI spectral profiles and cgMLST genotype was conducted examine the within farm variability between three MLST sequence type (ST6, ST20 and ST233) from one herd. Analysis of the within herd comparisons between genotype and phenotype revealed that the CCI results reflect the cgMLST clustering which identified ST20 as a distant clade relative to ST6 and ST233. The CCI data also suggests that the ST20 cluster is relatively homogenous which supports the cgMLST results. The CCI results indicate that ST20 is more similar to ST233 than ST6 in this herd and that ST6 is more similar to ST233 than ST20. However, cgMLST distance matrix analysis indicated that the ST6-ST233 group had a significantly lower median distance value than either of the other two inter-ST groups indicating that there was greater similarity between the ST6 isolate population and the ST233 population than between either of these sequence types and the ST20 group.

The data indicate that the cgMLST genotype was not an entirely consistent predictor of the MALDI phenotype. These results indicate that additional

regulatory mechanisms of gene expression or mobile genetic elements, undetected by the current analysis may be responsible for a sufficiently large proportion of the variation in protein expression to affect the MALDI spectral phenotype. This result correlates with the results of the previous Chapter which highlighted the limited discriminatory ability of the MALDI profile for MLST sequence type complex. The more detailed analysis of specific isolates of one case classification implicated as potential contagious pathogens highlights that there is substantial variation in protein expression of those isolates that cannot be attributed to the chromosomal gDNA sequenced in this study. This interpretation would seem to be supported by previous studies which have identified "the gain/loss of mobile genetic elements such as CRISPRs and prophage are a potential driving force for evolutionary change"(Hossain et al. 2015) that can include enhanced pathogenicity and immune evasion mechanisms.

It is acknowledged that this interpretation has been based upon data from a limited number of sequence types and the sampling was, by necessity, biased toward one herd with a substantial number of isolates that was required for robust statistical analysis. However, the sequence types selected were previously found to be widely dispersed between herds (Davies et al. 2016) and are therefore likely to be important and relevant to the molecular epidemiology of *S.uberis* mastitis.

The second objective was to evaluate evidence of specific a pathogenic mechanism in potentially contagious isolates by addressing two specific aims.

Firstly determine if isolates classified as persistent clinical cases in the same mammary gland according to MLST sequence type are true persistent infections or new infections with the same sequence type.

Persistent infections pose an enormous potential risk for contagious transmission because they expose susceptible cows in the herd to repeated infectious pressure at each milking. Furthermore the persistent clinical case infections expose those susceptible cows to a population of *S.uberis* isolates which have been, by the very definition of a persistent infection, selected for successful mammary colonisation.

The analysis of the cgMLST dendrogram structure of the persistent isolates implied that 12 of the 13 cases are likely to be true persistent mastitis cases with minimal variation between isolates from the same mammary gland. One pair of ST6 isolates (id 6474 & 6539) clearly did not cluster at the same terminal node in the same manner as all the other paired isolates. There are two possible explanations for this, either; a substantial recombination event occurred between isolates in the mammary gland but did not affect any of the MLST genes or the divergent result is due to two independent infections from different sources that coincidentally occurred with the same ST. The first explanation seems implausible given the distribution of the selected MLST genes throughout the genome should be detected by a change in at least one of the seven MLST genes. The second explanation may be more plausible given the high relative and absolute prevalence of ST6 clinical mastitis cases identified in the herd (Farm27). Overall the cgMLST data suggest that the

classic MLST scheme and the persistent case definition used in case classification is accurate in a very high proportion of cases (92%).

Phenotype analysis using MALDI CCI scores were found to be significantly higher in persistent cases compared to other case combinations in their respective herds. This implied that isolates collected from the same mammary gland are more likely to express the same functional protein products as well as being genetically very similar, as has been demonstrated earlier. It has been suggested that 'immune escape' could be a potential mechanism for *S.uberis* to establish and maintain longer term infections (Günther et al. 2016; Leigh et al. 2010). If this were the case then there would be the possibility that variation in protein expression of antigenic cell surface molecules responsible for virulence could be detectable by MALDI if the degree of variation was sufficiently profound and the expression patterns employed by the bacteria in the mammary gland are replicated in the culture process.

Identification of specific divergent peaks within the spectra would provide a means of gathering information on possible protein products that confer enhanced pathogenicity based upon their mass and abundance. This approach could provide an alternative, at the early stages to a sequencing approach to identify potential antigenic targets. As a there is the possibility that successful persistent mastitis strains of *S.uberis* undertake immune evasion by epigenetic gene regulation or by the acquisition of mobile genetic elements, to alter the gene product expression without the need to modify the genomic sequence code itself and has been demonstrated in

S.pneumoniae (Manso et al. 2014) other bacterial species (Seib et al. 2015). This would mean that sequence data alone would not identify this behaviour.

The second aim was to determine if there is evidence of antigenic variation of three virulence genes (SUB 0145/RS00865, SUB 1095/RS5505 and SUB 1154/RS5795) as a possible immune escape mechanism in persistent mastitis case.

If the role of variation within the three target virulence genes is to allow immune evasion by presentation of alternate antigenic surface proteins to the immune system as has been previously suggested (Foley 2015) then it is logical to assume that variation in the amino acid sequence would occur in the extracellular portion of the gene and potentially in the distal portion of the gene where contact with host cells would be most likely. However, the results of the pairwise comparison of the three target virulence genes from persistent mastitis clinical cases revealed that SUB1154/RS5795 was faithfully conserved within each pair and variation at the SUB1095/RS5055 target was only detected between five of the twelve isolate pairs. Of these five pairs only one of these possessed a base substitution which caused a non-synonymous missense event in a coding region between the transmembrane signal protein and the gram positive membrane anchor domain. It would therefore appear that neither SUB1095/RS5505 nor SUB1154/RS5795 exhibited the pattern of variation in these persistent cases that would have been expected in antigenic immune escape. However, at the SUB 0145/RS00865 locus 25% of the persistent clinical case isolate pairs (3/12) had the same missense substitution

between the first and second isolate. The A to T substitution at position 126 converted the codon between Glutamic acid to Aspartic acid in a protein coding region at the distal portion of the gene. It was unexpected to find exactly the same functional substitution in multiple pairs of isolates and this would be highly unlikely to be the result of multiple, identical, coincidental mutations. It may be significant that modification at this site could potentially change the antigenicity of this protein. Whilst there is no uniform pattern across all the persistent isolate pairs which would be consistent with immune evasion playing universal role in persistence of infection there are some interesting, unexpected findings which would justify further investigation as it is possible to envisage that immune evasion may be only one of many strategies employed by the *S.uberis* to adapt and persist within the mammary gland so 25% of isolates demonstrating one specific pattern may still be significant as an indication of one specific adaptive mechanism.

In the wider context of the epidemiology of *S.uberis* transmission within a herd the role of potential role persistent mastitis infections should not be under estimated. Each successive machine milking event to which a persistently infected cow is exposed presents an additional opportunity for transmission to subsequent cows using the same equipment. From the cgMLST data it can be observed that there are several PTE cases of very similar genotype to some of the persistent mastitis cases. However, a larger sample size would be required to quantify the actual association between the two case types. None the less the current evidence would suggest that a

combination of MALDI CCI and case information such as inter-case interval could allow improved detection of persistent infections allowing prompt management changes to be implemented. These could include well established management protocols such as segregation and milking persistently infected cows last as well as alternate treatment options or premature dry-off.

Chapter 7: Discussion

This research used MLST, MALDI and cgMLST to explore the sub-species diversity and transmission dynamics of *S.uberis* clinical mastitis in 52 dairy herds in England and Wales over a one year period. The purpose of the research was to advance the understanding of the importance of transmission route in the epidemiology of *S.uberis* clinical mastitis. Evidence for contagious transmission of *S.uberis* in addition to an environmental reservoir of infection has been growing over the past twenty five years with multiple reports as described at length in Chapter 1. Until now there was no evidence to suggest what proportion of herds were affected by contagious transmission events or how significant contagious transmission may be in the overall context of *S.uberis* mastitis in those affected herds. The results of this study, described in Chapter 3 and (Davies et al. 2016) highlight for the first time the likelihood that contagious transmission of *S.uberis* is not only common at a herd level, affecting two thirds of herds but also responsible for a substantial proportion of the *S.uberis* mastitis clinical cases. It is none the less clear that the environmental reservoir remains fundamental to the understanding of *S.uberis* mastitis. The contagious element of the epidemiology of the pathogen appears to be additional to a ubiquitous environmental disease pattern. The dynamics of the pathogen or virulence transmission are potentially more complex than has been demonstrated here, involving multiple reservoirs of infection as discussed below and in Figure 58.

The findings of this research suggest that *S.uberis* sequence types isolated from clinical mastitis and attributed to contagious transmission are not disseminated widely in a clonal, epidemic pattern throughout the herd as might be expected in the case of a highly contagious pathogen. The evidence suggests that rather than one, very successful *S.uberis* sub-group, we see several, similar sub-groups represented by sequence types that are common to many herds and able to be moderately successful in transmitting between cows while the vast majority of *S.uberis* sequence types in those herds show no evidence of transmission between cows and are therefore assumed to be, exclusively environmental opportunists.

From this research the is evidence that even within *S.uberis* sequence types there are multiple, apparently distinct, simultaneous contagious transmission events occurring within the same herd, as described in Chapter 6 with reference to Farm27 and three separate clades of Sequence Type 6 (ST6). This research has illustrated that *S.uberis* cases can be classified as likely to be contagious (PTE) or environmental (S) in origin based on their genomic DNA (Chapters 3 and 6) and that these classifications are recognisable as distinct groups from their cellular composition (principally protein) as depicted by their MALDI spectral profiles (Chapter 5). The MALDI spectral profiles also highlighted the powerful herd (farm) effect on the *S.uberis* isolate populations of both transmission case classifications (PTE and S) in Chapter 5. The influence of the individual farm was confirmed by more discriminatory cgMLST genomic analysis in Chapter 6. These results suggest that the classical

MLST scheme based upon only seven loci of the >1858 known loci (<0.3%) provides only a very simplified picture of the *S.uberis* population structure. This is not unexpected as the high recombination rate of *S.uberis*, previously discussed in Chapter1, provides a mechanism for individual bacteria to exchange very significant portions of their genome within an ecosystem. This recombination and exchange of genetic material could be in the form of either gDNA, plasmids, prophage or CRIPR's but by which ever mechanism, the effect on the epidemiology of S.uberis mastitis could be profound.

Dairy herd management provides a number of opportunities for *S.uberis* to adapt and evolve into a new niche as illustrated in Figure 58 below. The dairy cow "provides" *S.uberis* with multiple different environments, each with their own characteristics. Of these the mammary gland and the gastrointestinal tract are the two most potentially significant. The gastrointestinal tract provides an ideal environment for the maintenance of a large and diverse population of *streptococci spp* which is an ideal environment for widespread exchange of genetic material. In contrast the mammary tissue provides a more challenging environment which selects heavily for isolates that can survive the specific immunological responses as shown by Pryor et al (2009). The significance of our management of the dairy herd is that we inadvertently provide the heavily selected, 'cow adapted' isolates with opportunities to infect other cows via the vector of the milking machine or milking technicians hands. Many herds may also be exposing the developing gastrointestinal tract of herd replacement heifer calves to these 'cow adapted' 'potentially

contagious’ *S.uberis* isolates when they feed high somatic cell count milk or antibiotic treated quarter milk to those calves. It is unknown if the phenomenon is significant or if other routes of exposure of potentially high levels of heavily selected ‘contagious’ *S.uberis* isolates to a wider environmental population are important such as through discarding mastitic milk into the slurry system for spreading on pasture. It has been shown that pasture contamination levels of *S.uberis* are likely to be related to faecal loading (Lopez-Benavides et al. 2007) and it is possible that exposure of the wider environmental *S.uberis* population to ‘contagious’, ‘cow adapted’ isolates may allow dissemination of the genetic elements which confer that trait and thereby increase the likelihood that new environmental infections will then translate into additional contagiously acquired infections.

In this context a comparison with the proliferation of antibiotic resistance genes is appropriate because some of the same mechanisms apply and in the same commercial dairy setting. Shedding of antibiotic resistant *E.coli* has been demonstrated to be higher in a calves fed waste milk from clinical mastitis cases compared to those fed milk replacer and this shedding was shown to persist postweaning (Brunton et al. 2014). Antibiotic resistance has been shown amongst E.coli cultured from slurry systems and associated with previous mastitis antibiotic treatment practices (Ibrahim et al. 2016) and mathematical modelling of gene transfer in this slurry system suggested that for a pathogen such as *S.uberis*, with a high gene transfer rate, the propagation of antimicrobial resistance would be rapid and difficult to control

(Baker et al. 2016). Of the 52 herds participating in this study 48 herds fed heifer replacement calves waste milk from high cell count cows, clinical mastitis case or both.

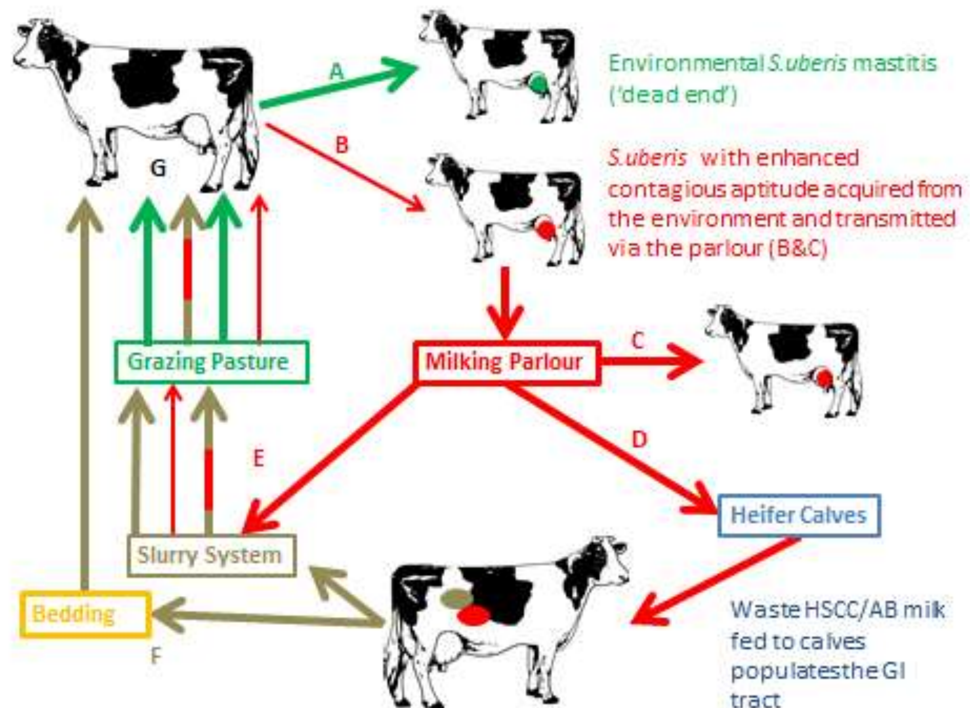

Figure 58. Diagram of a potential *S.uberis* isolate/gene flow in a dairy herd. A - 'Environmental' origin mastitis case with no onward transmission. B – Initial infection from environment with an isolate of contagious potential. C – Contagious transmission via the milking parlour. D – Waste milk containing 'Contagious' *S.uberis* fed to herd replacement heifer calves, establish gut flora and recombination opportunities. E – Waste milk containing 'Contagious' *S.uberis* and faeces with diverse Streptococcal spp mix in slurry system. F – Cows with gut flora containing 'Contagious' *S.uberis* isolates contaminate bedding. G – range of environmental infection sources (Bedding/pasture) with varying degrees of exposure to 'Contagious' *S.uberis*.

This theory of management practices influencing the movement and propagation of bacterial virulence mechanisms through a population could explain the wide variation in herd mastitis patterns observed in this study

differences from other counties, in particular New Zealand where very different pasture and slurry management is practiced.

The predominance of a small number of similar sequence types across a large number of herds (Chapter 3) suggests that the virulence determinants for contagious transmission are dependent in some way upon specific features of this group. These may be epigenetic regulatory mechanisms, rather than specific virulence genes, acting in a way that enhances infectivity or immune evasion. Such mechanisms could act upon mobile genetic elements such as 'Clustered regularly-interspaced short palindromic repeats' (CRISPR) (Hossain et al. 2015) that are otherwise unremarkable but when matched with the appropriate epigenetic manipulation, triggered possibly by external stimuli such as an immune response, they enable the bacterium to survive and replicate sufficiently to increase the probability of contagious transmission.

This research has demonstrated the importance of an understanding of the transmission dynamics at the herd level and the inadequacy of generic recommendations based upon assumptions of pathogen ecology at the species level. A better understanding of the factors that affect the ecology, infectivity and genetic transfer of *S.uberis* at a whole farm level is required alongside a deeper understanding of the mechanisms used by the bacterium to infect and establish intramammary infections. These knowledge gaps are discussed below (7.2).

7.1 Limitations of the project

The limited time period of the study may have resulted in the misclassification of some clinical case isolates assigned to the 'Unclassified' (U) case classification due to potential PTE cases of that same sequence type occurring outside the study period. Only clinical mastitis cases were collected so sub-clinical *S.uberis* infections could be assessed. The clinical mastitis rate is inevitably an underestimate of the true infection prevalence and incidence, if we assume that not all infections develop into overt clinical disease. The clinical case diagnosis and sample submission relied upon the detection sensitivity and specificity of the herdsman in each herd and their compliance with the sample collection protocol. These limitations were likely to have resulted in an underestimate of the potential contagious transmission of *S.uberis* according to the classification criteria used in this study.

The interval between successive milk recording events was too long to capture potentially significant relationships in the pattern of SCC before and after the clinical case. There is also the possibility that infected quarters were prematurely dried off individually or infected cows were dried off entirely thereby removing them from the sampled population at milk recording. This could have biased the results to an extent. Therefore the exact relationship between SCC and sub-species classification of *S.uberis* could not be evaluated in a robust way.

Accurate model generation in *ClinProTools3.0* required predefined groups to be of approximately equal sizes which limited the number of herds which could be used in analysis of between herd classification of PTE or Solitary cases. The high degree of between herd variability and the limited number of suitable herds with equal proportions of PTE and Solitary clinical cases limited the training dataset for an across herd PTE verses Solitary classification model. A larger number of herds with PTE/Solitary isolate group information would allow more of the between farm variability to be captured and would improve the predictive ability of the model.

Economic necessity prevented whole genome sequencing of all the available *S.uberis* isolates. A subset of isolates were selected to address specific questions, principally related to heterogeneity of the most prevalent potentially contagious (PTE) sequence types within and between herds. The selection of isolates excluded Solitary clinical case sequence types which prevented an assessment of the diversity between herds of the *S.uberis* isolates which most closely resemble an environmentally derived population by the more discriminatory cgMLST technique.

Finally this study did not seek to address epigenetic variation between isolates or the potential contribution of mobile genetic elements such as prophage and CRIPR's which have previously been demonstrated to be important (Hossain et al. 2015). Some of these factors could be addressed from the collected data however, alternative sequencing methodologies

would be likely to generate better results in these areas as discussed in section 7.2.5 and 0.

7.2 Further work

7.2.1 Updated and extended multi –herd longitudinal study

The results presented in this research reflect the *S.uberis* clinical mastitis epidemiology of a large sample of UK dairy herds ten to twelve years ago and an updated study on a similar scale that included high somatic cell count cows would address two important weaknesses in our current understanding of the epidemiology. Firstly we do not know how the Overrepresented (OVR) *S.uberis* strains or sub-species behave in the sub-clinical, pre-clinical or post-clinical phases of an intramammary infection. This is important in developing control measures, particularly around segregation and duration of increased shedding risk. Secondly an updated study would allow a better understanding of the trend over time in the prevalence if potential contagious transmission. Alongside this, a large scale, updated study would provide the opportunity to investigate the differences between the divergent dairy production systems which have become more common throughout the industry over the intervening decade. In particular the differences between 'New Zealand Style' grazing systems at one extreme compared to '365 housed' systems with no grazing exposure as well as milking systems such as robotic, automatic milking.

7.2.2 Sampling beyond the udder

Very little is known about the genetic diversity, population structure and ecology of *S.uberis* beyond the udder. The role of the bovine gastrointestinal tract in the epidemiology of *S.uberis* in particular is not well understood at present and whilst there are is some information from other countries on isolates cultured from the environment and GI tract (Cursons and Leigh 2007; Döpfer et al. 2008) it is superficial. A thorough, simultaneous longitudinal study, collecting samples from all likely *S.uberis* reservoirs (mammary glands, GI tract, slurry, pasture, bedding and calves) would be enormously helpful in tracking the flow of bacteria or bacterial genes through the ecosystem and through a herd.

7.2.3 Further opportunities for MALDI classification of *S.uberis*

An updated survey as discussed above would also provide *S.uberis* case classification data from a larger number of herds would assist in the optimisation of the MALDI 'PTE vs Solitary' model by capturing more of the farm or herd specific variability between the *S.uberis* isolates.

From the data already collected further analysis could be undertaken to identify the specific biomarkers with the highest discriminatory ability between the two classifications (PTE vs Solitary) using the specific selected peaks as described by (Ketterlinus et al. 2005), that were identified in the genetic algorithm model generated by *ClinProTools3.0*. These biomarkers could then assist in the investigation of the molecules which are associated

with the alternate transmission patterns and add to our understanding of the virulence mechanisms by a bottom-up approach to complement the genomic, top-down methodology.

7.2.4 Further opportunities for whole genome sequencing of *S.uberis* isolates.

The current study was only able to fully sequence a small subset of all the available isolates. This allowed certain specific, circumscribed aims to be addressed but did not allow for a large scale genetic comparison of potentially contagious isolates (PTE) and the presumed obligate environmental opportunist (Solitary). Principal component analysis of these two groups and analysis of these groups against isolates recovered from pasture, slurry, faeces and sub-clinical intramammary infections could provide a more complete understanding of the genetic component of *S.uberis* pathogenesis.

7.2.5 Investigation of mobile genetic elements and transcriptome analysis.

The NextSeq and Miseq technologies commonly used for whole genome sequence the isolates in as in this study generate reads of an average (post trim) length of 90 - 230 bases. These reads lengths are unable to span the entire genome reliably, especially when repetitive elements are present resulting in more gaps and more lost data including mobile genetic elements. One solution is to use ultra-long read sequencing systems to capture all of this data allowing analysis of all the genetic material (Huddleston et al. 2014).

The possibility of epigenetic variation between genetically very similar *S.uberis* isolates was discussed in Chapter 6.Transcriptome analysis by whole genome RNA-sequencing would allow detailed, quantitative assessment of the epigenetic component of bacterial diversity (Creecy and Conway 2015; Lawless et al. 2013). In conjunction with gDNA whole genome sequencing, transcriptomics would assist in our understanding of how apparently different transmission dynamics are controlled in *S.uberis*.

7.3 Conclusions

7.3.1 Chapter 3:

The results of this research demonstrated that contagious transmission of *S.uberis* may be occurring in at least two thirds of UK dairy herds and may be more common than environmentally derived transmission for one in three herds. The results suggested that persistent infections are significantly more likely to be caused by the same *S.uberis* sequence types associated with potential contagious transmission. The results indicated that more emphasis should be placed on identifying the most the dominant *S.uberis* transmission route operating within a herd and adopting appropriate control measures accordingly.

7.3.2 Chapter 4:

There was no evidence from this research that the readily available, monthly milk recording data can be used to predict the likely transmission route of *S.uberis* intramammary infections. However, individual clinical case

information such as the average stage of lactation in which the clinical mastitis case occurs was significantly later for the potentially contagious *S.uberis* transmission route and the proportion of clinical cases that occur in quick succession was significantly higher for potentially contagious cases. These findings are interesting but would still require a reliable and practical form of sub-species typing to be truly useful in constructing a complete picture of pathogen dynamics in a herd.

7.3.3 Chapter 5:

The results of this research suggested that MLST profiles cannot be reliably distinguished by MALDI however, potentially contagious (PTE) clinical case isolates could be successfully distinguished from suspected environmentally derived mastitis isolates (Solitary) within a herd by MALDI. This implied that there may be common transcribed product characteristics of these two groups which transcend the sequence type definitions.

The multi-herd models did not demonstrate the same discriminatory capability as the within herd models for 'Contagious' verses 'Environmental' type mastitis cases when data from a novel herd was used for validation, indicating substantial between herd variation in the populations of PTE and Solitary isolates.

7.3.4 Chapter 6:

The cgMLST results confirmed the substantial genetic similarity of *S.uberis* isolates within a herd observed in the MALDI spectral profiles (Chapter 5) which cuts across the MLST sequence type definitions. These patterns of genetic similarity between isolates of different sequence types within a herd and genetic heterogeneity between isolates of the same sequence type within a herd demonstrate the limitations of the classical MLST scheme in a highly recombinatorial bacterial species such as *S.uberis*.

Using alignments generated in *Ridom Seqsphere,* pairwise comparison of two, sortase anchored virulence genes found in successive bacterial isolates from the same persistent intramammary infection revealed only limited variation distributed throughout the genes. One notable exception was a common amino acid substitution in the SUB 0145/RS00865 gene which occurred in three of the twelve persistent infections. This result may indicate a functional advantage in alteration of the antigenicity of this protein.

Chapter 8: References

Abureema, Salem, Peter Smooker, Jakob Malmo, and Margaret Deighton. 2014. "Molecular Epidemiology of Recurrent Clinical Mastitis due to Streptococcus Uberis : Evidence of Both an Environmental Source and Recurring Infection with the Same Strain." *Journal of Dairy Science* 97(1):285–90. Retrieved (http://dx.doi.org/10.3168/jds.2013-7074).

Almeida, Raul a., and Stephen P. Oliver. 2006. "Trafficking of Streptococcus Uberis in Bovine Mammary Epithelial Cells." *Microbial Pathogenesis* 41(2-3):80–89.

Anon. 1970. "Occurrence and Ecologic Features of Streptococcus Uberis in the Dairy Cow ." 31(7):1970.

Anon. 1999. "Epidemiological Typing of Bovine Streptococci by Pulsed-Field Gel Electrophoresis." 317–24.

Arnold, Randy J., and James P. Reilly. 1998. "Fingerprint Matching of E . Coli Strains with Matrix-Assisted Laser Desorption / Ionization Time-of-Flight Mass Spectrometry of Whole Cells Using a Modified Correlation Approach." 636:630–36.

Baker, Michelle, Jon L. Hobman, Christine E. R. Dodd, Stephen J. Ramsden, and Dov J. Stekel. 2016. "Mathematical Modelling of Antimicrobial Resistance in Agricultural Waste Highlights Importance of Gene Transfer Rate." *FEMS Microbiology Ecology* (February):fiw040.

Bani-Yaghoub, Majid, Raju Gautam, Zhisheng Shuai, P. van den Driessche, and Renata Ivanek. 2012. "Reproduction Numbers for Infections with Free-Living Pathogens Growing in the Environment." *Journal of Biological Dynamics* 6(2):923–40.

Barbuddhe, Sukhadeo B. et al. 2008. "Rapid Identification and Typing of Listeria Species by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry." *Applied and Environmental Microbiology* 74(17):5402–7.

Barkema, H. W. et al. 1998. "Incidence of Clinical Mastitis in Dairy Herds Grouped in Three Categories by Bulk Milk Somatic Cell Counts." *Journal of Dairy Science* 81(2):411–19. Retrieved (http://dx.doi.org/10.3168/jds.S0022-0302(98)75591-2).

Barkema, H. W., O. C. Sampimon, Y. T. Gro, and Y. H. Schukken. 2001. "Analysis of an Outbreak of Streptococcus Uberis Mastitis." 590–99.

Barreiro, J. R. et al. 2010. "Short Communication : Identification of Subclinical Cow Mastitis Pathogens in Milk by Matrix-Assisted Laser Desorption / Ionization Time-of-Flight Mass Spectrometry." *Journal of Dairy Science* 93(12):5661–67. Retrieved (http://dx.doi.org/10.3168/jds.2010-3614).

Baseggio, N., P. D. Mansell, J. W. Browning, and G. F. Browning. 1997. "Strain Differentiation of Isolates of Streptococci from Bovine Mastitis by Pulsed-Field Gel Electrophoresis." *Molecular and cellular probes* 11(5):349–54.

Blowey RW, Collis K. 1992. "Effect of Premilking Teat Disinfection on Mastitis

Incidence, Total Bacterial Count, Cell Count and Milk Yield in Three Dairy Herds." *Vet Record.* 29(130):175–78.

Bradley, A. J. 2002. "Bovine Mastitis: An Evolving Disease." *The Veterinary Journal* 164(2):116–28. Retrieved (http://linkinghub.elsevier.com/retrieve/pii/S1090023302907240).

Bradley, A. J., J. E. Breen, B. Payne, V. White, and M. J. Green. 2015. "An Investigation of the Efficacy of a Polyvalent Mastitis Vaccine Using Different Vaccination Regimens under Field Conditions in the United Kingdom." *Journal of dairy science* 98(3):1706–20.

Bradley, A. J., K. A. Leach, J. E. Breen, L. E. Green, and M. J. Green. 2007. "Papers & Articles Survey of the Incidence and Aetiology of Mastitis on Dairy Farms in England and Wales." (January 2005).

Breen, J. E., M. J. Green, and A. J. Bradley. 2009. "Europe PMC Funders Group Quarter and Cow Risk Factors Associated with the Occurrence of Clinical Mastitis in Dairy Cows in the United Kingdom." 92(6):2551–61.

Brightling, Pb, Rd Dyson, Af Hope, and J. Penry. 2009. "A National Programme for Mastitis Control in Australia: Countdown Downunder." *The Veterinary Journal* 62 Suppl 4:S52–58.

Brunton, L. A., H. E. Reeves, L. C. Snow, and J. R. Jones. 2014. "A Longitudinal Field Trial Assesing the Impact of Feeding Waste Milk Containing Antibiotic Residues on the Prevalence of ESBL-Producing Escherichia Coli in Calves." *Preventive Veterinary Medicine* 117(2):403–12.

Cattell, Marguerita B. 1996. "An Outbreak of Streptococcus Uberis As a Consequence of Adopting a Protocol of No Antibiotic Therapy for Clinical." 123–26.

Christner, Martin et al. 2014. "Rapid MALDI-TOF Mass Spectrometry Strain Typing during a Large Outbreak of Shiga-Toxigenic Escherichia Coli." 9(7):1–11.

Coffey, Tracey J. et al. 2006. "First Insights into the Evolution of Streptococcus Uberis : A Multilocus Sequence Typing Scheme That Enables Investigation of Its Population Biology First Insights into the Evolution of Streptococcus Uberis : A Multilocus Sequence Typing Scheme That Enabl."

Compton, C. W. R., C. Heuer, K. Parker, and S. Mcdougall. 2007. "Epidemiology of Mastitis in Pasture-Grazed Peripartum Dairy Heifers and Its Effects on Productivity." *Journal of Dairy Science* 90(9):4157–70. Retrieved (http://dx.doi.org/10.3168/jds.2006-880).

Cookson, D. et al. 2016. "Inter-Centre Comparison of Pulsed-F Ield Gel Electrophoresis for the Typing of Methicillin- Res Is T a N T Staphylococcus Aureus." 44(1 996):179–84.

Creecy, James P., and Tyrrell Conway. 2015. "Quantitative Bacterial Transcriptomics with RNA-Seq." *Current Opinion in Microbiology* 23:133–40.

Crowley, R. C., J. A. Leigh, P. N. Ward, and L. D. Bowler. 2011. "Differential Protein Expression in Streptococcus Uberis under Planktonic and Biofilm Growth Conditions ▯ †." 77(1):382–84.

Cursons, R. T., and J. A. Leigh. 2007. “Field Observations on the Variation of Streptococcus Uberis Populations in a Pasture-Based Dairy Farm.” *Journal of Dairy Science* 90(12):5558–66. Retrieved (http://dx.doi.org/10.3168/jds.2007-0194).

Davies, P. L. et al. 2016. “Molecular Epidemiology of Streptococcus Uberis Clinical Mastitis in Dairy Herds : Strain Heterogeneity and Transmission.” 54(1):68–74.

Day, Nicholas P. J. et al. 2001. “A Link between Virulence and Ecological Abundance in Natural Populations of Staphylococcus Aureus.” *Science (New York, N.Y.)* 292(5514):114–16.

Diekmann, Heesterbeek, Metz. 1990. “On the Definition and the Computation of the Basic Reproduction Ratio R 0 in Models for Infectious Diseases in Heterogeneous Populations.” *Journal of mathematical biology* 28(4):365–82.

Döpfer, D. et al. 2008. “Assessing Genetic Heterogeneity within Bacterial Species Isolated from Gastrointestinal and Environmental Samples: How Many Isolates Does It Take?” *Applied and environmental microbiology* 74(11):3490–96. Retrieved (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2423023&tool=pmcentrez&rendertype=abstract).

Douglas, V. L., S. G. Fenwick, D. U. Pfeiffer, N. B. Williamson, and C. W. Holmes. 2000. “Genomic Typing of Streptococcus Uberis Isolates from Cases of Mastitis, in New Zealand Dairy Cows, Using Pulsed-Field Gel Electrophoresis.” *Veterinary microbiology* 75(1):27–41.

Down, P. M., A. J. Bradley, J. E. Breen, C. D. Hudson, and M. J. Green. 2016. “Current Management Practices and Interventions Prioritised as Part of a Nationwide Mastitis Control Plan.” *Veterinary Record* vetrec – 2015–103203.

Down, P. M., M. J. Green, and C. D. Hudson. 2013. “Rate of Transmission: A Major Determinant of the Cost of Clinical Mastitis.” *Journal of dairy science* 96(10):6301–14.

Dronawat, N. S., R. W. Stanley, E. Cobb, and K. Morita. 1966. “Effect of Feeding Limited Roughage and a Comparison between Loose and Pelleted Pineapple Hay on Milk Production, Milk Constituents, and Fatty Acid Composition of Milk Fat.” *Journal of dairy science* 49(1):28–31.

Dubois, Damien et al. 2010. “Identification of a Variety of Staphylococcus Species by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry.” *Journal of Clinical Microbiology* 48(3):941–45.

Eigh, James A. L., Sharon A. E. Gan, Philip N. W. Ard, Terence R. F. Ield, and Tracey J. C. Offey. 2010. “Original Article Sortase Anchored Proteins of Streptococcus Uberis Play Major Roles in the Pathogenesis of Bovine Mastitis in Dairy Cattle.”

Falush, Daniel et al. n.d. “Traces of Human Migrations in Helicobacter Pylori Populations.”

Feil, E. J., J. M. Smith, M. C. Enright, and B. G. Spratt. 2000. “Estimating Recombinational Parameters in Streptococcus Pneumoniae from Multilocus Sequence Typing Data.” *Genetics* 154(4):1439–50.

Feil, Edward J. et al. 2004. “eBURST : Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data eBURST : Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequen.”

Finn, Robert D. et al. 2015. “The Pfam Protein Families Database: Towards a More Sustainable Future.” *Nucleic acids research* 44(December 2015):gkv1344.

Foley, Janet. 2015. “Mini-Review : Strategies for Variation and Evolution of Bacterial Antigens.” *CSBJ* 13:407–16. Retrieved (http://dx.doi.org/10.1016/j.csbj.2015.07.002).

Geary, U. et al. 2012. “Estimating the Effect of Mastitis on the Profitability of Irish Dairy Farms.” *Journal of dairy science* 95(7):3662–73. Retrieved (http://dx.doi.org/10.3168/jds.2011-4863\nhttp://www.ncbi.nlm.nih.gov/pubmed/22720924).

Gilchrist, T. L., D. G. E. Smith, J. L. Fitzpatrick, R. N. Zadoks, and M. C. Fontaine. 2013. “Comparative Molecular Analysis of Ovine and Bovine Streptococcus Uberis Isolates.” 1–9.

Gon??alves, Juliano Leonel et al. 2014. “Identification of Corynebacterium Spp. Isolated from Bovine Intramammary Infections by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry.” *Veterinary Microbiology* 173(1-2):147–51.

Green, M. .. 2012. *Dairy Herd Health*.

Green, M. J. et al. 2004. “Somatic Cell Count Distributions During Lactation Predict Clinical Mastitis.” *J. Dairy Sci.* 87(5):1256–64. Retrieved (http://jds.fass.org/cgi/content/abstract/87/5/1256).

Green, M. J., L. E. Green, G. F. Medley, Y. H. Schukken, and a J. Bradley. 2002. “Influence of Dry Period Bacterial Intramammary Infection on Clinical Mastitis in Dairy Cows.” *Journal of dairy science* 85(10):2589–99. Retrieved (http://dx.doi.org/10.3168/jds.S0022-0302(02)74343-9).

Green, M. J., K. A. Leach, J. E. Breen, L. E. Green, and A. J. Bradley. 2007. “Papers & Articles National Intervention Study of Mastitis Control in Dairy Herds in England and Wales.”

Groschup, M. H., G. Hahn, and J. F. Timoney. 1991. “Antigenic and Genetic Homogeneity Fo Streptococcus Uberis Strains from the Bovine Udder.” *Epidemiology Infections* 107:297–310.

Günther, Juliane et al. 2016. “Streptococcus Uberis Strains Isolated from the Bovine Mammary Gland Evade Immune Recognition by Mammary Epithelial Cells , but Not of Macrophages.” *Veterinary Research* 1–14. Retrieved ("http://dx.doi.org/10.1186/s13567-015-0287-8).

Haas, Y. De, H. W. Barkema, Y. H. Schukken, and R. F. Veerkamp. 2005. “Associations between Somatic Cell Count Patterns and the Incidence of Clinical Mastitis.” 67:55–68.

de Haas, Y., R. F. Veerkamp, H. W. Barkema, Y. T. Gröhn, and Y. H. Schukken. 2004.

"Associations Between Pathogen-Specific Cases of Clinical Mastitis and Somatic Cell Count Patterns." *Journal of Dairy Science* 87(1):95–105. Retrieved (http://linkinghub.elsevier.com/retrieve/pii/S002203020473146X).

Harmon, R. J. 1994. "Physiology of Mastitis and Factors Affecting Somatic Cell Counts." *Journal of Dairy Science* 77(7):2103–12. Retrieved (http://linkinghub.elsevier.com/retrieve/pii/S0022030294771538).

Hill, AW, and JA Leigh. 1989. "DNA Fingerprinting of Streptococcus Uberis: A Useful Tool for Epidemiology of Bovine Mastitis." *Epidemiology and infection* 103:165–71.

Hillerton, J. E., M. F. H. Shearn, R. M. Teverson, S. Langridge, and J. M. Booth. 1993. "Effect of Pre-Milking Teat Dipping on Clinical Mastitis on Dairy Farms in England." *Journal of Dairy Research* 60(1):31–41.

Hogan, J. S., and K. L. Smith. 2003. "Environmental Streptococcal Mastitis, Facts, Fables And Fallacies." *National Mastitis Council Annual Meeting* 162–71.

Hossain, Maqsud et al. 2015. "Virulence Related Sequences ; Insights Provided by Comparative Genomics of Streptococcus Uberis of Differing Virulence." 1–13.

Huddleston, John et al. 2014. "Reconstructing Complex Regions of Genomes Using Long-Read Sequencing Technology Reconstructing Complex Regions of Genomes Using Long-Read Sequencing Technology." *Genome research* 688–96.

Ibrahim, Delveen R., Christine E. R. Dodd, Dov J. Stekel, Stephen J. Ramsden, and Jon L. Hobman. 2016. "Multi Drug and Extended Spectrum Beta-Lactamase Resistant Escherichia Coli Isolated from a Dairy Farm ." (February):1–13.

Jolley, K. A., E. J. Feil, and M. C. J. Maiden. 2001. "Tests (START)." 17(12):1230–31.

Jones TO, Jones PC. 1986. "Cow Milk Yield and Composition before Development of Escherichia Coli Mastitis." *Vet Record.* 27(119(13)):319–21.

Jones, C. L. et al. 2004. "Localization of MtuA, an LraI Homologue in Streptococcus Uberis." *Journal of Applied Microbiology* 97(1):149–57.

Jørgensen, H. J. et al. 2015. "Streptococcus Agalactiae in the Environment of Bovine Dairy Herds – Rewriting the Textbooks?" *Veterinary Microbiology* 184:64–72.

Kallow, Wibke et al. 2006. "Microbial Strain Characterisation by MALDI-TOF MS-Possibilities and Limits." *Annual General Meeting of the European Culture Collections Organisation - ECO XXV: Proceedings*.

Kester, H. J., D. E. Sorter, and J. S. Hogan. 2015. "Activity and Milk Compositional Changes Following Experimentally Induced Streptococcus Uberis Bovine Mastitis." *Journal of Dairy Science* 98(2):999–1004. Retrieved (http://dx.doi.org/10.3168/jds.2014-8576).

Ketterlinus, Ralf, Sen-Yung Hsieh, Shih-Hua Teng, Helen Lee, and Wolfgang Pusch. 2005. "FISHING FOR BIOMARKERS: ANALYZING MASS SPECTROMETRY DATA WITH THE NEW ClinProTools Software." *BioTechniques* 38(June):S37–40.

Khan, I. U., A. A. Hassan, A. Abdulmawjood C. Lämmler, W. Wolter, and M. Zschöck.

2003. “Identification and Epidemiological Characterization of Streptococcus Uberis Isolated from Bovine Mastitis.” *Journal of Veterinary Science* 4(3):213–23.

Knox, C. Michele, Vickey Cevellos, and Deborah Dean. 1998. “16S Ribosomal DNA Typing for Identification of Pathogens in Patients with Bacterial Keratitis.” *Journal of Clinical Microbiology* 36(12):3492–96.

Kohl, Thomas A. et al. 2014. “Whole-Genome-Based Mycobacterium Tuberculosis Surveillance : A Standardized , Portable , and Expandable Approach.” 52(7):2479–86.

Kornienko, Maria et al. 2016. “Infection , Genetics and Evolution Analysis of Nosocomial Staphylococcus Haemolyticus by MLST and MALDI-TOF Mass Spectrometry.” *MEEGID* 39:99–105. Retrieved (http://dx.doi.org/10.1016/j.meegid.2015.12.015).

Kossaibati, M. a, and R. J. Esslemont. 1997. “The Costs of Production Diseases in Dairy Herds in England.” *Veterinary journal (London, England : 1997)* 154(1):41–51.

Lam, T. J. G. M. et al. 2013. “Improving Bovine Udder Health: A National Mastitis Control Program in the Netherlands.” *Journal of dairy science* 96(2):1301–11.

Lawless, Nathan, Amir B. K. Foroushani, Matthew S. McCabe, Cliona O’Farrelly, and David J. Lynn. 2013. “Next Generation Sequencing Reveals the Expression of a Unique miRNA Profile in Response to a Gram-Positive Bacterial Infection.” *PLoS ONE* 8(3).

Leelahapongsathon, Kansuda, Ynte Hein Schukken, Tanu Pinyopummintr, and Witaya Suriyasathaporn. 2015. “Comparison of Transmission Dynamics between Streptococcus Uberis and Streptococcus Agalactiae Intramammary Infections.” *Journal of Dairy Science* 99(2001):1–9.

Leigh, J. a. 1999. “Streptococcus Uberis: A Permanent Barrier to the Control of Bovine Mastitis?” *Veterinary journal (London, England : 1997)* 157(3):225–38.

Leigh, James A., Sharon A. Egan, Philip N. Ward, Terence R. Field, and Tracey J. Coffey. 2010. “Sortase Anchored Proteins of Streptococcus Uberis Play Major Roles in the Pathogenesis of Bovine Mastitis in Dairy Cattle.” *Veterinary Research* 41(5).

Lopez-Benavides, M. G. et al. 2007. “Field Observations on the Variation of Streptococcus Uberis Populations in a Pasture-Based Dairy Farm.” *Journal of dairy science* 90(12):5558–66.

Maiden, M. C. et al. 1998. “Multilocus Sequence Typing: A Portable Approach to the Identification of Clones within Populations of Pathogenic Microorganisms.” *Proceedings of the National Academy of Sciences of the United States of America* 95(6):3140–45.

Manso, Ana Sousa et al. 2014. “Epigenetic Changes.” *Nature Communications* 5:1–9. Retrieved (http://dx.doi.org/10.1038/ncomms6055).

Medrano-Galarza, C., J. Gibbons, S. Wagner, A. M. de Passillé, and J. Rushen. 2012.

“Behavioral Changes in Dairy Cows with Mastitis.” *Journal of dairy science* 95(12):6994–7002.

Mehta, Angela, and Luciano P. Silva. 2015. “MALDI-TOF MS Profiling Approach : How Much Can We Get from It ?” 6(March):2014–16.

Milne, M. H. et al. 2005. “Treatment of Persistent Intramammary Infections with Streptococcus Uberis in Dairy Cows .” *Veterinary Record* 157:245–50.

Minitab 17 Inc. 2015. “Minitab 17.” *Minitab Statistical Software*.

Mitra, Susweta Das et al. 2016. “Streptococcus Uberis ST439 and ST475 Induce Differential Inflammatory Responses in a Mouse Intramammary Infection Model.” *Gene*. Retrieved April 14, 2016 (http://www.sciencedirect.com/science/article/pii/S037811191630244X).

Morton, John M., John F. Penry, Jakob Malmo, and Graeme A. Mein. 2014. “Premilking Teat Disinfection: Is It Worthwhile in Pasture-Grazed Dairy Herds?” *Journal of dairy science* 97(12):7525–37.

Murchan, Stephen et al. 2003. “Harmonization of Pulsed-Field Gel Electrophoresis Protocols for Epidemiological Typing of Strains of Methicillin-Resistant.” *Society* 41(4):1574–85.

Nakano, S. et al. 2015. “Development and Evaluation of MALDI-TOF MS-Based Serotyping for Streptococcus Pneumoniae.” 2191–98.

Nakano, S. et al. 2015. “Development and Evaluation of MALDI-TOF MS-Based Serotyping for Streptococcus Pneumoniae.” *European Journal of Clinical Microbiology and Infectious Diseases* 34(11):2191–98.

Niyompanich, Suthamat, Janthima Jaresitthikunchai, Kitima Srisanga, and Sittiruk Roytrakul. 2014. “Source-Identifying Biomarker Ions between Environmental and Clinical Burkholderia Pseudomallei Using Whole-Cell Matrix-Assisted Laser Desorption / Ionization Time-of-Flight Mass Spectrometry.” 9(6).

Oliveira, C. S. F. et al. 2015. “Cow-Specific Risk Factors for Clinical Mastitis in Brazilian Dairy Cattle.” *Preventive Veterinary Medicine* 121(3-4):297–305. Retrieved (http://linkinghub.elsevier.com/retrieve/pii/S0167587715002652).

Østerås, O., and L. Sølverød. 2009. “Norwegian Mastitis Control Programme.” *Irish Veterinary Journal* 62(Suppl 4):S26.

Oviedo-Boyso, Javier et al. 2007. “Innate Immune Response of Bovine Mammary Gland to Pathogenic Bacteria Responsible for Mastitis.” *The Journal of infection* 54(4):399–409.

Pankey, J. W., E. E. Wildman, P. A. Drechsler, and J. S. Hogan. 1987. “Field Trial Evaluation of Premilking Teat Disinfection.” *Journal of dairy science* 70(4):867–72. Retrieved April 10, 2016 (http://www.sciencedirect.com/science/article/pii/S0022030287800851).

Phuektes, P. et al. 2001. “Molecular Epidemiology of Streptococcus Uberis Isolates from Dairy Cows with Mastitis.” *Journal of clinical microbiology* 39(4):1460–66.

Piessens, V. et al. 2011. “Distribution of Coagulase-Negative Staphylococcus Species from Milk and Environment of Dairy Cows Differs between Herds.” *Journal of Dairy Science* 94(6):2933–44. Retrieved (http://dx.doi.org/10.3168/jds.2010-3956).

Pryor, S. M., R. T. Cursons, and J. H. Williamson. 2009. “Experimentally Induced Intramammary Infection with Multiple Strains of Streptococcus Uberis.” *Journal of Dairy Science* 92(11):5467–75. Retrieved (http://dx.doi.org/10.3168/jds.2009-2223).

Pullinger, Gillian D. et al. 2006. “Application of Streptococcus Uberis Multilocus Sequence Typing : Analysis of the Population Structure Detected among Environmental and Bovine Isolates from New Zealand and the United Kingdom.” 72(2):1429–36.

Pullinger, Gillian D., Tracey J. Coffey, Martin C. Maiden, and James A. Leigh. 2007. “Multilocus-Sequence Typing Analysis Reveals Similar Populations of Streptococcus Uberis Are Responsible for Bovine Intramammary Infections of Short and Long Duration.” 119:194–204.

Raemy, Andreas et al. 2013. “Phenotypic and Genotypic Identification of Streptococci and Related Bacteria Isolated from Bovine Intramammary Infections.” *Acta Veterinaria Scandinavica* 55(1):1. Retrieved (Acta Veterinaria Scandinavica).

Rato, G., Ricardo Bexiga, Carlos Florindo, Lina M. Cavaco, and Cristina L. Vilela. 2013. “Antimicrobial Resistance and Molecular Epidemiology of Streptococci from Bovine Mastitis.” 161:286–94.

Reinoso, Elina B., Mirta C. Lasagno, Silvana a Dieser, and Liliana M. Odierno. 2011. “Distribution of Virulence-Associated Genes in Streptococcus Uberis Isolated from Bovine Mastitis.” *FEMS microbiology letters* 318(2):183–88.

Rettinger, Anna et al. 2012. “Leptospira Spp . Strain Identification by MALDI TOF MS Is an Equivalent Tool to 16S rRNA Gene Sequencing and Multi Locus Sequence Typing (MLST).”

Reyher, K. K. et al. 2011. “The National Cohort of Dairy Farms-A Data Collection Platform for Mastitis Research in Canada.” *J Dairy Sci* 94(3):1616–26.

Riekerink, R. G. M. Olde, H. W. Barkema, D. F. Kelton, and D. T. Scholl. 2008. “Incidence Rate of Clinical Mastitis on Canadian Dairy Farms.” *Journal of Dairy Science* 91(4):1366–77. Retrieved (http://dx.doi.org/10.3168/jds.2007-0757).

Rizzardi, Kristina, Tara Wahab, and Cecilia Jernberg. 2013. “Rapid Subtyping of Yersinia Enterocolitica by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDITOF MS) for Diagnostics and Surveillance.” *Journal of Clinical Microbiology* 51(12):4200–4203.

Ruppitsch, Werner et al. 2015. “Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Whole-Genome Sequence-Based Typing of Listeria Monocytogenes.” 53(9):2869–76.

Seib, Kate L. et al. 2015. “Epigenetic Regulator N 6 -Adenine DNA Methyltransferases of Neisseria Meningitidis.” 1–13.

Shearn MF, Langridge S, Teverson RM, Booth JM, Hillerton JE. 1992. “Effect of Pre-Milking Teat Dipping on Clinical Mastitis.” *Vet Record.* 21(131):488–89.

Shome, Bibek Ranjan et al. 2012. “Molecular Characterization of Streptococcus Agalactiae and Streptococcus Uberis Isolates from Bovine Milk.” *Tropical animal health and production* 44(8):1981–92.

Shum, L. W. C., C. S. McConnel, a. a. Gunn, and J. K. House. 2009. “Environmental Mastitis in Intensive High-Producing Dairy Herds in New South Wales.” *Australian Veterinary Journal* 87(12):469–75.

Sousa, C. et al. 2014. “MALDI-TOF Mass Spectrometry as a Tool for the Discrimination of High-Risk Escherichia Coli Clones from Phylogenetic Groups B2.” 1391–99.

Steeneveld, W., L. C. Van Der Gaag, H. W. Barkema, and H. Hogeveen. 2009. “Providing Probability Distributions for the Causal Pathogen of Clinical Mastitis Using Naive Bayesian Networks.” *Journal of Dairy Science* 92(6):2598–2609. Retrieved (http://dx.doi.org/10.3168/jds.2008-1694).

Tagg, Kaitlin A., Andrew N. Ginn, Sally R. Partridge, and Jonathan R. Iredell. 2015. “MALDI-TOF Mass Spectrometry for Multilocus Sequence Typing of Escherichia Coli Reveals Diversity among Isolates Carrying Bla CMY-2 -Like Genes.” 1–10.

Todhunter, D. A., K. L. Smith, and J. S. Hogan. 1995. “Environmental Streptococcal Intramammary Infections of the Bovine Mammary Gland.” *J. Dairy Sci.* 78(11):2366–74.

Tomazi, T. et al. 2014. “Identification of Coagulase-Negative Staphylococci from Bovine Intramammary Infection by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry.” *Journal of Clinical Microbiology* 52(5):1658–63.

Tomita, Takehiro, Brian Meehan, Nalin Wongkattiya, Gillian Pullinger, et al. 2008. “Identification of Streptococcus Uberis Multilocus Sequence Types Highly Associated with Mastitis Identification of Streptococcus Uberis Multilocus Sequence Types Highly Associated with Mastitis □.”

Tomita, Takehiro, Brian Meehan, Nalin Wongkattiya, Jakob Malmo, et al. 2008. “Identification of Streptococcus Uberis Multilocus Sequence Types Highly Associated with Mastitis.” *Applied and environmental microbiology* 74(1):114–24.

Tomita, Takehiro. 2008. “Subspecies Distribution and Virulence Factors of Streptococcus Uberis.” (November).

Trede, Dennis et al. 2012. “On the Importance of Mathematical Methods for Analysis of MALDI-Imaging Mass Spectrometry Data.” *Journal of integrative bioinformatics* 9(1):189.

Turner, Katherine M. E., William P. Hanage, Christophe Fraser, Thomas R. Connor, and Brian G. Spratt. 2007. “Assessing the Reliability of eBURST Using Simulated Populations with Known Ancestry.” *BMC microbiology* 7:30.

Turner, Sally-anne, John H. Williamson, S. Jane Lacy-hulbert, and J. Eric Hillerton.

2013. “Relationship between Previous History of Streptococcus Uberis Infection and Response to a Challenge Model.” 360–66.

Unit, Ruminant Health, Animal Health Service, and Diagnostic Sciences. 2003. “Clinical , Epidemiological and Molecular Characteristics of Streptococcus Uberis Infections in Dairy Herds.” 335–49.

Valentine, Nancy, Sharon Wunschel, David Wunschel, Catherine Petersen, and Karen Wahl. 2005. “Effect of Culture Conditions on Microorganism Identification by Matrix-Assisted Laser Desorption Ionization Mass Spectrometry Effect of Culture Conditions on Microorganism Identification by Matrix-Assisted Laser Desorption Ionization Mass Spectrometry.” *Applied and environmental microbiology* 71(1):58–64.

Verbeke, Joren, Sofie Piepers, Karlien Supré, and Sarne De Vliegher. 2014. “Pathogen-Specific Incidence Rate of Clinical Mastitis in Flemish Dairy Herds , Severity , and Association with Herd Hygiene.” *Journal of Dairy Science* 97(11):6926–34. Retrieved (http://dx.doi.org/10.3168/jds.2014-8173).

Wang, Lijun, Wei Chen, Lili Zhang, and Yaxin Zhu. 2013. “Full Paper Genetic Diversity of Streptococcus Uberis Isolates from Dairy Cows with Subclinical Mastitis in Southern Xinjiang Province , China.” 293.

Wang, Wei et al. 2009. “Infection , Genetics and Evolution Gene Content Differences across Strains of Streptococcus Uberis Identified Using Oligonucleotide Microarray Comparative Genomic Hybridization.” 9:179–88.

Ward, Philip N. et al. 2009. “Evidence for Niche Adaptation in the Genome of the Bovine Pathogen Streptococcus Uberis.” *BMC genomics* 10:54.

Wesen, D. P., and L. H. Schultz. 1970. “Effectiveness of a Post-Milking Teat Dip in Preventing New Udder Infections.” *Journal of dairy science* 53(10):1391–1403.

Weterings, V., K. Zhou, J. W. Rossen, D. Van Stenis, and E. Thewessen. 2015. “An Outbreak of Colistin-Resistant Klebsiella Pneumoniae Carbapenemase-Producing Klebsiella Pneumoniae in the Netherlands (July to December 2013), with Inter-Institutional Spread.” 1647–55.

White, L. J., Y. H. Schukken, T. J. G. M. Lam, G. F. Medley, and M. J. Chappell. 2001. “A Multispecies Model for the Transmission and Control of Mastitis in Dairy Cows.”

Wilesmith. 1986. “No Title.” *Veterinary Record*.

Williams, J. G., a R. Kubelik, K. J. Livak, J. a Rafalski, and S. V Tingey. 1990. “DNA Polymorphisms Amplified by Arbitrary Primers Are Useful as Genetic Markers.” *Nucleic acids research* 18(22):6531–35.

Wittwer, M. et al. 2005. “Genetic Diversity and Antibiotic Resistance Patterns in a Campylobacter Population Isolated from Poultry Farms in Switzerland.” 71(6):2840–47.

Wojdak-Maksymiec, Katarzyna, Joanna Szyda, and Tomasz Strabel. 2013. “Parity-Dependent Association between TNF-α and LTF Gene Polymorphisms and Clinical Mastitis in Dairy Cattle.” *BMC veterinary research* 9(1):114.

Xiao, Di, Fei Zhao, Huifang Zhang, Fanliang Meng, and Jianzhong Zhang. 2014a. “A Novel Strategy for Typing Mycoplasma Pneumoniae Using MALDI-TOF.” (June):1–20.

Xiao, Di, Fei Zhao, Huifang Zhang, Fanliang Meng, and Jianzhong Zhang. 2014b. “Novel Strategy for Typing Mycoplasma Pneumoniae Isolates by Use of Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry Coupled with ClinProTools.” *Journal of clinical microbiology* 52(8):3038–43.

Zadoks, R. N., H. G. Allore, H. W. Barkema, O. C. Sampimon, Y. T. Gröhn, et al. 2001. “Analysis of an Outbreak of Streptococcus Uberis Mastitis.” *Journal of dairy science* 84(3):590–99.

Zadoks, R. N. et al. 2001. “Cow- and Quarter-Level Risk Factors for Streptococcus Uberis and Staphylococcus Aureus Mastitis.” *Journal of Dairy Science* 84(12):2649–63.

Zadoks, R. N., H. G. Allore, H. W. Barkema, O. C. Sampimon, G. J. Wellenberg, et al. 2001. “Cow- and Quarter-Level Risk Factors for Streptococcus Uberis and Staphylococcus Aureus Mastitis.” *Journal of dairy science* 84(12):2649–63.

Zadoks, Rn, and Jl Fitzpatrick. 2009. “Changing Trends in Mastitis.” *Irish Veterinary Journal* 62(Suppl 4):S59.

Zadoks, Ruth N., Ynte H. Schukken, and Martin Wiedmann. 2005. “Multilocus Sequence Typing of Streptococcus Uberis Provides Sensitive and Epidemiologically Relevant Subtype Information and Reveals Positive Selection in the Virulence Gene pauA.” 43(5):2407–17.

Zadoks, Ruth N., Linda L. Tikofsky, and Kathryn J. Boor. 2005. “Ribotyping of Streptococcus Uberis from a Dairy ’ S Environment , Bovine Feces and Milk.” 109:257–65.

Zhang, Wei, Bhushan M. Jayarao, Stephen J. Knabel, and A. P. P. L. E. Nviron M. Icrobiol. 2004. “Multi-Virulence-Locus Sequence Typing of Listeria Monocytogenes.” 70(2):913–20.

Zoche-Golob, V., W. Heuwieser, and V. Krömker. 2015. “Investigation of the Association between the Test Day Milk Fat–protein Ratio and Clinical Mastitis Using a Poisson Regression Approach for Analysis of Time-to-Event Data.” *Preventive Veterinary Medicine* 121(1-2):64–73. Retrieved (http://linkinghub.elsevier.com/retrieve/pii/S0167587715002305).