Novel molecular markers of disease-association among strains of *Streptococcus suis*: a genomic approach

Thomas Mathew Wileman

This dissertation is submitted for the degree of

Doctor of Philosophy

January 2018

University of Cambridge

Department of Veterinary Medicine

Novel molecular markers of disease-association among strains of *Streptococcus suis*: a genomic approach

Thomas Mathew Wileman

Abstract

This thesis focuses on the use of a genomic approach to identify novel molecular markers to differentiate *Streptococcus suis* (*S. suis*) isolates into two populations, i) disease-associated and ii) non-disease associated. *S. suis* is a Gram-positive coccus that is considered one of the most important zoonotic bacterial pathogens of swine responsible for significant economic losses to swine production worldwide. Importantly, *S. suis* is not only an invasive pathogen but also a very successful coloniser of mucosal surfaces; often endemic in swine populations sampled.

The widescale use of antibiotics to control and prevent the various clinical manifestations caused by *S. suis* has become unsustainable, due to increases in antibiotic resistance and government pressures. Other popular control strategies, such as the development of efficacious vaccines, are hindered by differences in virulence not only between but also within *S. suis* serotypes, as well as, the lack of a detailed understanding of the role in pathogenesis of many proposed virulence-factors. As a result, the detection of *S. suis* in asymptomatic swine herds is of little practical value in predicting the likelihood of future clinical relevance.

This thesis aims to further understanding of the role the *S. suis* genome has in pathogenesis. The value of future surveillance and preventative health management lies in the detection of strains that genetically have increased potential to cause disease in presently healthy animals. The first results chapter of this thesis (chapter 3) describes the use of genome-wide associations studies, a so-far unexploited method for *S. suis*, to identify genetic markers associated with the observed clinical phenotype i) invasive disease or ii) asymptomatic carriage on the palatine tonsils of swine. Chapter 4 then describes the analyses used to select three genetic markers to pathotype *S. suis* - differentiate isolates of the same species based on their ability to cause disease; going on to describe the design and evaluation of a multiplex-PCR tool targeting the three newly defined "pathotyping markers" in comparison to existing methods used to characterise *S. suis*.

These findings were taken further by using the pathotyping markers to screen material scrapped from the palatine tonsils of swine with no obvious signs of streptococcal disease. This produced an interesting result - the production of both invasive disease-associated and non-disease associated multiplex-PCR amplicons from the same experimental sample. Unsurprising in itself, what *was* surprising is the frequency with which this observation was found. Picking single colonies from solid agar plates is a crippling bottleneck of existing *S. suis* diagnostics, and its removal has the potential to improve the sensitivity of surveillance and preventive health management programs. Chapter 5 describes investigation of this surprising observation and indicates that classic culture-based methods of detection are not sensitive enough to confidently report the presence (or absence) of invasive disease-associated *S. suis* strains.

This thesis concludes with the description of efforts to address the lack of a comprehensive understanding of *S. suis* virulence/'virulence-associated' factors. Chapter 6 describes the design of an isogenic mutant knocking out the invasive disease-associated pathotyping marker, SSU1589 (also known as *virA*). That is then evaluated in simple *in vitro* and *in vivo* experimental models in order to understand the role Type I restriction modification proteins have in *S. suis* pathogenesis.

In conclusion, this thesis furthers our understanding that differences in the *S. suis* genome are an important factor in *S. suis* pathogenesis, and describes the identification and evaluation of novel genetic markers for the detection and control of invasive disease-associated *S. suis* strains in intensive pig production systems.

Table of contents

List of figures

List of tables

Abbreviations

A_{595}	Absorbance 595 nm
AFLP	Amplified fragment length polymorphism
APHA	Animal and Plant Health Agency
AHVLA	Animal Health and Veterinary Laboratories Agency
AIC	Akaike information criterion
BAPS	Bayesian Analysis of Population Structure
BBSRC	Biotechnology and Biological Sciences Research Council
BLAST	Basic Local Alignment tool
bp	Base pair
CDCD	Caesarean-derived, colostrum-deprived
CMH	Cochran-Mantel-Haenszel (CMH) Chi-squared test for count data
CP	Capsular polysaccharide
CSF	Cerebrospinal fluid
DAPC	Discriminant Analysis of Principal Components
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EF	Extracellular protein factor
ERIC	Enterobacterial repetitive intergenic consensus
g	Force of gravity
G. mellonella	*Galleria mellonella*
GBS	Group B streptococci
GLM	Generalised linear model
GWAS	Genome-wide association study
h	Hours
HMI	Host Microbe Interactomics
ICE	Integrative conjugative elements
IL	Interleukin

Kb	Kilo base
LoLa	Longer and Larger
Mb	Mega base
MCG	Minimum core genome
mg	Milligrams
min	Minutes
MLEE	Multilocus enzyme electrophoresis
MLST	Multilocus sequence typing
mPCR	Multiplex-polymerase chain reaction
MRP	Muramidase-related protein
ng	Nanograms
nr	Non-redundant
OD_{595}	Optical density 595 nm
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PCV	Porcine circovirus
PFGE	Pulsed-field gel electrophoresis
PRRSV	Porcine reproductive and respiratory syndrome virus
qPCR	Quantitative real-time polymerase chain reaction
REP	Repetitive extragenic palindromic
RFLP	Restriction fragment length polymorphism
RM	Restriction modification
ROC	Receiver operating characteristic
s	Seconds
S. suis	*Streptococcus suis*
SAM	S-adenosylmethionine
SLY	Suilysin

SMRT	Single-molecule real-time
SNP	Single-nucleotide polymorphism
spec	Spectinocmycin
SPF	Specific-pathogen-free
ST	Sequence type
STM	Signature-tagged mutagenesis
STSS	Streptococcal toxic shock syndrome
TBE	Tris base, boric acid and EDTA
THB	Todd-Hewitt broth
TNF-α	Tumour necrosis factor-α
TRD	Target recognition domain
V	Volts
vol	Volume
vol/vol	Volume/Volume
WGS	Whole-genome sequencing
WHO	World Health Organisation
wt/vol	Weight/volume
WIAS	Wageningen Institute of Animal Sciences
WTSI	Wellcome Trust Sanger Institute
α	Alpha
β	Beta
ϕ	Phi

Declaration

This dissertation is the results of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

Dr. Edoardo Zaccaria and Professor Jerry M. Wells gave guidance and practical help with the design of isogenic *virA* operon mutants whilst I was a Visiting Research Fellow in the Department of Animal Sciences, subdivision Host-Microbe Interactomics (HMI) group at Wageningen University, the Netherlands. Dr. Nadya Velikova from the Department of Animal Sciences, subdivision HMI at Wageningen University, and Dr. Xiaoliang (Ibrahim) Ba from the Department of Veterinary Medicine at the University of Cambridge assisted the author with Greater Wax Moth (Lepidoptera: Pyralidae, *Galleria mellonella*) larvae infection model experiments conducting all inoculations. All other protocols were conducted by the author.

Statement of length

This dissertation does not exceed the word limit of 60,000 words (excluding figures, photographs, tables, appendices and bibliography) set by the Degree Committee for the Faculties of Clinical Medicine and Veterinary Medicine.

Acknowledgements

First and foremost I would like to thank my supervisors, Dr Alexander (Dan) W. Tucker and Professor Duncan J. Maskell, for giving me the opportunity to work on a project that I found so engaging, and for giving me the guidance and support necessary to complete my PhD. I would also like to thank Research Associate Dr Lucy A. Weinert for countless helpful conversations, encouragement and for being so generous with her time (and for proof reading this thesis).

I'd like to extend my gratitude to Professor Jerry M. Wells for his invaluable guidance and refreshing support. I am also in his debt for his part in making it possible for me to be a visiting researcher in his Host Microbe Interactomics (HMI) group at Wageningen University (the Netherlands); allowing me to work with his group and cultivate an ongoing and enduring collaboration. Thanks must also go to Dr Kate J. Howell, Dr Sarah E. Peters and Dr Jinhong (Tracy) Wang for their technical guidance, and I must also acknowledge Dr Andrew J. Grant and Dr Trevelyan J. McKinley for helpful conversations and assistance with statistical analyses.

I also wish to extend my gratitude to Dr Mark R. Davies, Professor Gordon Dougan, Dr Christine Hale, Dr Trevor D. Lawley and Dr Derek J. Pickard for showing me how scientists behave, think and draw the best out of one another. Being part of Team 15 at The Wellcome Trust Sanger Institute taught me a huge amount, not least to work harder than anyone else and take nothing for granted; something I saw every single day in Doctor Vanessa 'Nessy Monster' K. Wong.

Thanks to my parents, without whom I would not be where I am today. They have been, and continue to be, an incredible source of inspiration and support, even when they are not sure what I am talking about on the phone; I couldn't have done this without them. Similarly, my heartfelt thanks to my friends for not making me talk about science outside of work – especially those at Pembroke, and a unique group mad enough to drag me around in a rowing boat (in particular, Dr Moritz Matthey, Dr Tim T. Rademacher and Dr Nick Salisbury). A special thank you also needs to go to [Dr] Alessandra Bittante; for your relentless encouragement.

Lastly, I acknowledge the Biotechnology and Biological Sciences Research Council (BBSRC) for their generous funding of this research project. This PhD research project was a BBSRC Knowledge Transfer Network (KTN) CASE studentship co-funded by Zoetis, a company committed to expanding its portfolio of diagnostic support to enable ongoing responsible use of antimicrobials in the UK and global pig industry. Significant contribution was also received from BQP Ltd, a major pig integrator focused on reducing streptococcal disease in their pig breeding system. This research also contributed to the BBSRC Longer and Larger (LoLa) project, grant number: BB/G019274/1, that had the aim to create single platform diagnostics and multivalent vaccines against *Actinobacillus pleuropneumoniae*, *Haemophilus parasuis*, *Mycoplasma hyopneumoniae* and *Streptococcus suis* - all important bacterial pathogens commonly recovered from the upper respiratory tract of swine.

Fellowships

I was awarded a **Wageningen Institute of Animal Sciences (WIAS) research fellowship 2016** (€4,800) to visit Wageningen University (the Netherlands) and work for four months as a visiting researcher in the Department of Animal Sciences, subdivision Host-Microbe Interactomics (HMI) group chaired by Professor Jerry M. Wells.

I was awarded a **Cambridge Global Food Security Early Career Researcher Travel Fund Grant 2018** (£500), co-provided by the Sir Isaac Newton Trust, to travel to Chongqing Yuelai International Convention Centre (China) and give an oral presentation at the 25th International Pig Veterinary Society (IPVS) Congress & 2018 International PRRS Symposium.

WAGENINGEN UR
For quality of life

The Graduate School WIAS
WAGENINGEN INSTITUTE of ANIMAL SCIENCES

P.O. Box 338 | 6700 AH Wageningen | The Netherlands

Wageningen University, Department of Animal Sciences
Host Microbe Interactomics group
Prof.dr. Jerry Wells
Courier 97

Wageningen Institute of Animal Sciences

DATE
September 1, 2016

SUBJECT
WIAS fellowship visiting researcher 2016

OUR REFERENCE
16WIAS0025 / HMI Wells / Mr Thomas Wileman

POSTAL ADDRESS
P.O. Box 338
6700 AH Wageningen
The Netherlands

VISITORS' ADDRESS
Wageningen Campus
Zodiac Building 122
De Elst 1
6708 WD Wageningen
The Netherlands

INTERNET
www.wageningenUR.nl/wias

HANDLED BY
Janneke van Seters

TELEPHONE
+31 (0)317 48 22 87

EMAIL
janneke.vanseters@wur.nl

Dear Jerry,

Hereby, I want to thank you for requesting a WIAS researcher fellowship to invite researchers from abroad to visit Wageningen University for a period up to 6 months for Mr. Thomas Wileman.

Seven applications for this fellowship were received in the second call. The applications were reviewed by me. The applications were evaluated regarding added value of the visiting researcher to WIAS and scientific profile of the candidate. The result of this review is that four requests are accepted. I am pleased to inform you that your request is granted.

Mr. Wileman will receive a total of € 4800 for the stay of 4 months. This fellowship is meant to cover the costs of housing and living during the stay in Wageningen.

When your guest arrives, please take him to Eva van Voorst (assistant controller; room B0013) with a copy of his passport and she will arrange the financial matters. If your guest cancels or shortens his visit, please let Janneke van Seters know. Others may be happy to receive the fellowship!

With kind regards,

Prof. Dr Ir. Johan van Leeuwen
WIAS Scientific Director

Publications

In preparation

Wileman TM, Grant A, Hernandez-Garcia J, Weinert LA and Tucker AW. Surveillance for *Streptococcus suis* must account for relatively low abundance of disease-associated strains relative to non-disease associated strains in carrier pigs. *In preparation.*

Wileman TM, Weinert LA, Wang J, Peters SE, Howell KJ, Williamson SM, Langford PR, Rycroft AN, Wren BW, Maskell DJ and Tucker AW. Pathotyping *Streptococcus suis*: A novel Approach to the Differentiation of Disease-Associated and Non-Disease Associated *Streptococcus suis* Isolates from UK pigs. *Submitted to J Clin Microbiol.*

Kingsley RA, Langridge G, Smith SE, Makendi C, Fookes M, **Wileman TM** *et al*. Functional analysis of *Salmonella* Typhi adaptation to survival in water.
Accepted Envion Microbiol. 2018 Oct 28. doi: 10.1111/1462-2920.14458 [ahead of print].

Published

Atack JM, Weinert LA, Tucker AW, Husna AU, **Wileman TM** *et al*. *Streptococcus suis* contains multiple phase-variable methyltransferases that show a discreet lineage distribution. Nucleic Acids Res. 2018 Oct 10. doi: 10.1093/nar/gky913. [Epub ahead of print].

Zou G, Zhou J, Xiao R, Zhang L, Cheng Y, Jin H, Li L, Zhang L, Wu B, Qian P, Li S, Ren L, Wang J, Oshota O, Hernandez-Garcia J, **Wileman TM** *et al*. Effects of Environmental and Management-Associated Factors on Prevalence and Diversity of *Streptococcus suis* in Clinically Healthy Pig Herds in China and the United Kingdom. Appl Environ Microbiol. 2018 Apr 2;84(8). pii: e02590-17. doi: 10.1128/AEM.02590-17.

Hernandez-Garcia J, Wang J, Restif O, Holmes MA, Mather AE, Weinert LA, **Wileman TM** *et al*. Patterns of antimicrobial resistance in *Streptococcus suis* isolates from pigs with or without streptococcal disease in England between 2009 and 2014. Vet Microbiol. 2017. Aug;207:117-124. doi: 10.1016/j.vetmic.2017.06.002.

Klemm EJ, Gkrania-Klotsas E, Hadfield J, Forbester JL, Harris SR, Hale C, Heath JN, **Wileman TM** *et al*. Emergence of host-adapted *Salmonella* Enteritidis through rapid evolution in an immunocompromised host. Nat Microbiol. 2016 Mar;1(3). pii: 15023. doi: 10.1038/nmicrobiol.2015.23.

Skinner BM, Sargent CA, Churcher C, Hunt T, Herrero J, Loveland JE, Dunn M, Louzada S, Fu B, Chow W, Gilbert J, Austin-Guest S, Beal K, Carvalho-Silva D, Cheng W, Gordon D, Grafham D, Hardy M, Harley J, Hauser H, Howden P, Howe K, Lachani K, Ellis PJ, Kelly D, Kerry G, Kerwin J, Ng BL, Threadgold G, **Wileman TM** *et al*. The pig X and Y Chromosomes: structure, sequence, and evolution. Genome Res. 2016 Jan;26(1):130-9. doi: 10.1101/gr.188839.114.

Wong VK, Pickard DJ, Barquist L, Sivaraman K, Page AJ, Hart PJ, Arends MJ, Holt KE, Kane L, Mottram LF, Ellison L, Bautista R, McGee CJ, Kay SJ, **Wileman TM** *et al*. Characterization of the *yehUT* Two-Component Regulatory System of *Salmonella enterica* Serovar Typhi and Typhimurium. PLoS One. 2013 Dec 30;8(12):e84567. doi: 10.1371/journal.pone.0084567.

Kingsley RA, Kay S, Connor T, Barquist L, Sait L, Holt KE, Sivaraman K, **Wileman TM** *et al*. Genome and Transcriptome Adaptation Accompanying Emergence of the Definitive Type 2 Host-Restricted *Salmonella enterica* Serovar Typhimurium Pathovar. mBio. 2013 Aug 27;4(5):e00565-13. doi: 10.1128/mBio.00565-13.

Perkins TT, Davies MR, Klemm EJ, Rowley G, **Wileman TM** *et al*. ChIP-seq and transcriptome analysis of the OmpR regulon of *Salmonella enterica* serovars Typhi and Typhimurium reveals accessory genes implicated in host colonization. Mol Microbiol. 2013 Feb;87(3):526-38. doi: 10.1111/mmi.12111.

Conferences attended

2018

The **25th International Pig Veterinary Society (IPVS) Congress & 2018 International PRRS Symposium** at the Chongqing Yuelai International Convention Centre, Chongqing, China

The **EBI-Sanger-Cambridge PhD Symposium (eSCAMPS) 2018** at the Gillespie Centre (Clare College), Cambridge, UK

2017

The **20th Lancefield International Symposium on Streptococci and Streptococcal (LISSS) Disease 2017** at the Sofitel Fiji Resort and Spa, Denarau Island, Fiji

The **EBI-Sanger-Cambridge PhD Symposium (eSCAMPS) 2017** at the Gillespie Centre (Clare College), Cambridge, UK

2016

The **24th International Pig Veterinary Society (IPVS) Congress | 8th European Symposium of Porcine Health Management 2016** at the Royal Dublin Society (RDS), Dublin, Ireland

The **EBI-Sanger-Cambridge PhD Symposium (eSCAMPS) 2016** at Magdalene College (Cripps Court), Cambridge, UK

2015

The **EBI-Sanger-Cambridge PhD Symposium (eSCAMPS) 2015** at Magdalene College (Cripps Court), Cambridge, UK

2014

The **EBI-Sanger-Cambridge PhD Symposium (eSCAMPS) 2014** at Magdalene College (Cripps Court), Cambridge, UK

Training courses

2017

Protein Structure Analysis

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

Statistical Analysis using R

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

CRUK: Introduction to Statistical Analysis

CRUK, University of Cambridge, Cambridge, United Kingdom

CRUK: Using the Ensembl Genome Browser

CRUK, University of Cambridge, Cambridge, United Kingdom

2016

Bioinformatics for Protein Biology

EMBL-EBI, Hinxton, United Kingdom

Introduction to genome variation analysis using NGS

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

Introduction to Scientific Figure Design

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

Analysis of DNA Methylation using Sequencing

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

2015

How to write an academic paper and get it published

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

Clear Vision - Presenting data without the fog

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

EMBL-EBI: Introduction to ontologies

EMBL-EBI, Hinxton, United Kingdom

Protein and RNA multiple sequence alignment, protein secondary structure prediction, trees, sub-family and function analysis with Jalview

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

2014

Bioinformatics: An Introduction to Solving Biological Problems with Python

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

Bioinformatics: An Introduction to Solving Biological Problems with R

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

Regression Analysis: A Hands-on Approach

Statistical Services Centre, University of Reading, Reading, United Kingdom

A Review of Basic Statistics

Statistical Services Centre, University of Reading, Reading, United Kingdom

2013

Bioinformatics: Multiple Sequence Alignment and Phylogenetics

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

Bioinformatics: Molecular Phylogenetics

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

Bioinformatics: Introduction to Next Generation Sequencing

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

Data Management

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

Scientific Writing

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

How to Write Your Dissertation (School of Biological Sciences)

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

How to Keep a Lab Notebook

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

Skills To Get You Started

Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom

Chapter 1

Literature review

1 Introduction

Streptococcus suis (*S. suis*) is one of the most important bacterial pathogens of pigs causing significant economic losses to the swine industry worldwide [1]. The infectious agent is responsible for a wide range of clinical manifestations, including septicaemia with sudden death, arthritis, endocarditis, meningitis, and pneumonia amongst other diseases [2-11]. *S. suis* is also a zoonotic pathogen, and although considered a rare and sporadic event, associated with prolonged exposure to pigs or pork-derived products, reports of human infections have increased substantially in recent years, particularly in Southeast Asia [12-17].

Streptococcal infection of the central nervous system has been described in swine since 1912. In 1951, meningo-encephalitis was first experimentally reproduced in pigs, and in 1954, an inability to maintain balance due to meningo-encephalitis resulting from invasion of the brain and its meninges was first described in Great Britain [18]. The incriminated haemolytic streptococci all showed very similar biochemical properties and did not yield an extract that gave a precipitate with any of the Lancefield group sera described [18, 19]. However, it was not until 1987 that *S. suis* was considered to be a genetically homogeneous species and recognised as a new valid species of the genus *Streptococcus* [20].

The prophylactic use of antibiotics in food and drinking water has proven unsuccessful in controlling disease. Antibiotics are becoming less effective because of an increase in resistance among *S. suis* isolates. Blanket antibiotic use is also becoming less accepted because of consumer/government pressure on the swine industry to reduce antimicrobial use [21, 22]. Other popular control strategies, such as the development of effective vaccines are hindered by the lack of a comprehensive understanding of *S. suis* virulence-factors, the number of serotypes and differences in virulence not only among but also within the different serotypes.

Importantly, *S. suis* is not only an invasive pathogen but also a very successful coloniser of mucosal surfaces [23]. In fact, the upper respiratory tract of pigs, in particular the palatine

tonsils, is considered to be both the natural habitat of *S. suis* and a principal route of invasion; although the bacterium can also be recovered from the gastrointestinal and genital tracts [11]. Colonisation of adult pigs is common in almost all pig populations sampled, meaning that transfer of *S. suis* from sow to piglet during parturition and suckling is an important route of transmission [23].

Since S. sus is such a successful coloniser of mucosal surfaces this review starts with an introduction on its biology and the numerous existing approaches for identification, discrimination of phylogenetic diversity and surveillance. The distribution and determinants of health-related states or events (including disease), and the application of this to the control of diseases and other health problems is then discussed (epidemiology). Putative and confirmed virulence factors are then reviewed. The advantages and disadvantages of experimental infection models are reviewed. Before important features in the genome are discussed.

1.1 *Streptococcus suis*

S. suis is a Gram-positive coccus of the order Lactobacillales (lactic acid bacteria), family Streptococcaeceae and genus *Streptococcus*. Other notable pathogenic species of the genus include *Streptococcus orisratti*, the well-studied *Streptococcus pneumoniae* (the pneumococcus) and *Streptococcus pyogenes*. Concerning its biological features, *S. suis* shares similarities with the pneumococcus and group B streptococci (GBS), i.e. the occurrence of several capsular serotypes, the significance of the capsular polysaccharide as a virulence factor, the respiratory (and genital) tract as predominant niche within the host, and clinical consequences of infection (septicaemia, meningitis, pneumonia). *S. suis* is small, less than two micrometres in diameter. *S. suis* is nonmotile, and often found singly, in pairs or (rarely) in short chains. *S. suis* is a chemoorganotroph with facultative anaerobic and fermentative metabolisms. *S. suis* isolates produce typically mucoid grey/white colonies that are α-haemolytic on solid agar plates supplemented with sheep blood. Most isolates are also β-haemolytic on solid agar plates supplemented with horse blood agar [1, 20, 24]. Growth and haemolysis can be enhanced when organisms are cultured anaerobically.

1.1.1 Current approaches for diagnosis and surveillance

Presumptive diagnosis of *S. suis* infection is based on animal age, clinical signs and the presence of macroscopic lesions. Signs of acute infection start with anorexia, depression, fever and lassitude, followed by symptoms depending on the localisation i.e. arthritis, ataxia, blindness, convulsions, dyspnoea, erythrema, lameness, opisthotonus, paddling and tremor [11]. Some animals with acute infection survive, resulting in persistent or chronic infection. In the latter, lameness and residual nervous signs such as otitis interna can be evident, as well as chronic arthritis [9, 25].

Diagnosis is then confirmed by the recognition of microscopic lesions, and the isolation of the infectious agent followed by a combination of simple biochemical tests and serotyping [1]. Gross necropsy findings depend on the clinical presentation. Few lesions are seen in the

septicemic phase, but purulent meningitis may be seen when the blood-brain barrier is crossed, and in arthritis inflammation around the affected joint with mucoid fluid is common. Where an infectious agent is isolated from diseased pigs and serotyping is available, an α-haemolytic *Streptococcus* (on solid agar supplemented with sheep blood) that produces amylase but a negative Voges-Proskauer (acetoin) reaction can be considered to be *S. suis* [26].

Pulmonary lesions are common with *S. suis* infection, but frequently other bacteria contribute to the development of a fibrin-haemorrhagic pneumonia [9]. Indeed, isolation of *S. suis* from the lungs must be interpreted with caution, as the organism is considered endemic in pig herds, often without causing clinical signs of disease[11]. The identification of non-symptomatic carriage of *S. suis* can be achieved through detection of the organism using classic microbiology techniques. However, isolation of bacteria from sites such as the upper respiratory tract is a laborious and time consuming process associated with low sensitivity due to phenotypically similar streptococcal species difficult to distinguish on the basis of colony morphology alone [27]. In fact, detection of *S. suis* from the upper respiratory tract is deemed to be of little practical utility in the diagnosis of disease as colonisation alone is not a good criterion for virulence [11].

1.1.1.1 Biochemical typing

Biochemical typing is an important part of the routine diagnostic procedure for *S. suis*. Table 1.1 summarises the biochemical properties of *S. suis*. *S. suis* is an encapsulated coccus that retains the crystal violet stain when treated by Gram's method (Gram-positive). Different biochemical tests have been used for the identification of *S. suis*, although biochemical variations have been noted among different strains [10, 24]. Since the first study to include a complete scheme for identification [28], the minimum of standardised biochemical tests required to differentiate *S. suis* from the other species of streptococci has been proposed by numerous authors as awareness of the organism increased [10, 24, 26, 29, 30]. The use of at least four tests has been proposed for a presumptive identification of *S. suis*, including i) the production of amylase, ii) the production of acid in salicin and trehalose broth, iii) a negative Voges-Proskauer (acetoin) reaction and iv) the absence of growth on 6.5% NaCl agar [24, 26], where the production of acetoin (or lack of) is critical and the most reliable property for discrimination of *S. suis* from *Streptococcus bovis* [24].

Table 1.1. The biochemical properties of *Streptococcus suis*.

Summary of the biochemical properties of *S. suis*, adapted and updated from the review by Perch *et al.*[28].

Positive reactions	**Negative reactions**	**Variable**
Fermentation of	*Fermentation of*	*Fermentation of*
Glucose	Arabinose	Raffinose
Glycogen	Glycerol	Melibiose
Inulin	Mannitol	
Lactose	Melezitose	
Maltose	Soritol	
Sucrose		
Trehalose		
Hydrolysis of	*Hydrolysis of*	
Arginine	Hippurate	
Esculin		
Salicin		
	Production of	*Production of*
	Acetoin	Hyaluronidase
	Growth	
	At 10 ˚C	
	At 45 ˚C	
	In 6.5 % NaCl	
	On 40 % bile agar	
	On 0.04 % tellurite	
	Sensitivity to optochin	

1.1.1.2 Lancefield grouping

S. suis possesses cell wall antigenic determinates belonging to Lancefield group D [31-33]. Lancefield grouping is the serological differentiation of haemolytic streptococci into distinct and orderly groups [34]. When first described, although cultural and biochemical tests were useful in differentiation of haemolytic streptococci from various sources, a serological test added considerable weight to conclusions regarding the origin of a particular strain. Lancefield's approach, by means of the precipitin reaction [35], is "dependent on the presence in streptococci of substances characteristic of the large groups although not specific for types within groups" [34]. More specifically, Lancefield was referring to 'group-specific' immunological differences in a carbohydrate, the so-called C substance, present in the cell wall of streptococci.

The group antigen of group D streptococci is a glucose-substituted α-glycerophosphate that can be found in two forms: lipid-bound (lipoteichoic acid), or lipid free. Both forms react with group D antiserum and are often found together, but in differing relative proportions, depending on the method of extraction [32, 36]. This disparity in extraction methods is important and has led to considerable confusion resulting in the description of new Lancefield groups R, S, RS and T [19]. However, some years later this was identified as being wrong and the new Lancefield groups R and S shown to be *S. suis* (Lancefield group D) and reclassified as serotype 1 (formerly group S), serotype 2 (formerly group R), and serotype 1/2 (formerly group R/S) [37]. Group T was also reclassified as serotype 15 [38]. The terminology of Lancefield groups R, S, RS, and T is sometimes used in papers describing human infections and should not be [11].

1.1.1.3 Serotyping

Serotyping is an important part of the routine diagnosis procedure for *S. suis* often used in combination with a minimum of simple biochemical tests [26, 39]. Different techniques exist, including i) the capillary precipitation test, ii) the coagglutination test [40] or iii) Neufeld's capsular reaction; all of which require serotype-specific reference antisera. The capsular reaction has previously been reported as the test of choice [24], although, the coagglutination technique is preferred by many laboratories, especially in North America [41].

To date, *S. suis* has been classified into 35 (1-34 and 1/2) serotypes based on differences in the capsular polysaccharide antigens [31, 37, 38, 42-44]. Serotype reference strains have been derived from a variety of sources, including diseased pigs, the nasal cavities of clinically healthy pigs and other animal species including humans. However, since their original descriptions evidence now exists for the reclassification of serotypes 20 and 22 as *Streptococcus orisratti*, and serotypes 26, 32, 33 and 34 as belonging to a bacterial species different from *S. suis* [45-49]. As a result, current opinion now considers the *Streptococcus suis* sp. to comprise of 29 'true' serotypes (1-19, 21, 23-25, 27-31 and 1/2) [50].

Serotype 2 has always been considered the most virulent and the most frequent *S. suis* serotype to be associated with clinical disease in both pigs and humans worldwide [41]. In China, serotype 2 accounts for more than 70% of the systemic *S. suis* disease in piglets [51]. Importantly, not all serotype 2 isolates are virulent and there is variation in the degrees of virulence among strains [52-56]. Indeed, virulence differs both between and within many of the *S. suis* serotypes. There is also a clear geographical effect on the distribution of serotypes, that appears to change over time [17]. For example, i) serotypes 2 and 3 are now most prevalent in Canada and the United States - a similar serotype distribution that might be explained by the fluid movement of animals between the two countries [17], ii) in European countries the Netherlands and Spain serotype 9 is most prevalent in diseased animals [57-59] while iii) other important pig producing (European) countries, such as Belgium, Denmark, France, Germany and

the UK, have not recently reported the distribution of serotypes [17]; although in the UK (in addition to serotype 2) serotypes 1 and 14, in the past, have been documented as being responsible for invasive *S. suis* disease in predominately suckling piglets [60, 61]. Broadly speaking serotypes 1-9 and 1/2 are most frequently recovered from diseased pigs worldwide [17]. It is also possible to isolate multiple *S. suis* serotypes from diseased animals within the same herd, and the same animal [62].

Some *S. suis* isolates cross-react with more than one antisera, indicating the presence of common antigenic determinants [24]. The following cross-reactions are well documented: serotype 1 with type 14 antisera, serotype 1/2 with type 1 and 2 antisera, serotype 6 with type 16 antisera (and vice versa) and serotype 22 with type 2 antisera. In some cases, absorption to obtain monospecific antisera is recommended to reduce the number of cross-reactions [24, 38]. Since most serotypable isolates associated with clinical disease belong to serotypes 1 - 9 and 1/2 it has been suggested that it is impractical for diagnostic laboratories to stock expensive antisera corresponding to all of the 35 originally described *S. suis* serotypes. Instead, to save expense and laborious processing time diagnostic laboratories should use only reference antisera for serotypes 1 - 9 and 1/2, then sending all non-serotypable isolates to a centralised reference laboratory [44].

Only encapsulated isolates can be serotyped [24]. Non-serotypable isolates are commonly recovered from the upper respiratory tract of clinically healthy pigs with similar biochemical and morphological features. It is difficult to be certain if isolates not reacting with the reference antisera are truly non-encapsulated or possess a novel capsular polysaccharide antigen to which reference antisera have not previously been raised [16] - a feasible possibility as traditionally antisera were only raised against capsular polysaccharide antigens associated with clinical disease. Loss of capsule has been reported among some *S. suis* isolates [63, 64]. It has also been proposed that *S. suis* down-regulates expression of the capsular polysaccharide for increased

adhesion to epithelial cells, and in turn up-regulates expression of the capsular polysaccharide for protection against phagocytosis as it enters the bloodstream [41].

It is important to note that some reported cases of human infection have been attributed to *S. suis* serotype 2 based on biochemical analyses obtained using commercial multi-test kits. However, there is no evidence of a correlation between a specific serotype and its biochemical properties [17]. The apiweb database (Biomerieux) identifies isolates as "*Streptococcus suis I*" and "*Streptococcus suis II*" and for clarity these results are biotypes (groups having the same fundamental constitution in terms of genetic factors) of the organism and not serotypes [65].

1.1.1.3.1 Molecular 'serotyping'

Molecular 'serotyping' based on PCR amplification of serotype-specific genes has been explored as an attractive alternative and/or complement to the existing serological tests [50, 66-75]. Serotyping of *S. suis* is one of the most useful methods to understand the epidemiology of a particular outbreak and monitor the prevalence of potentially hazardous strains, however, traditional serotyping techniques are time-consuming, and preparing the antisera is not easy due to the high cost and labor associated with its production. Therefore, the development of more practical and easier serotyping methods is desirable. The *S. suis* capsular polysaccharide synthesis (*cps*) genes are clustered on a single locus of the chromosome, and as in other Gram-positive bacteria, such as *Streptococcus pneumoniae*, the capsular polysaccharides of all *S. suis* serotypes are synthesised by the Wzx/Wzy-dependent pathway [76-78]. The two most recent molecular serotyping schemes for *S. suis*, published within six months of one another, both use multiple multiplex-PCR (mPCR) assays to target serotype-specific differences in the *cps* genes [74, 75]. For example, the Liu *et al.* [74] scheme uses a combination of four mPCRs to assign 33 of the 35 originally described *S. suis* serotypes to isolates (serotypes 32 and 34 were excluded from analyses as these were considered by the authors to have been reassigned to *Streptococcus orisratti*). The first mPCR of which targets the serotypes most commonly associated with clinical disease in pigs (serotypes 1-10, 14 and 1/2), by targeting serotype-specific differences in five

cps genes (I, J, K, L and M) using a total of ten oligonucleotide primer pairs. In comparison, the molecular serotyping scheme described by Okura *et al.* [75] uses just two mPCRs, an initial "Grouping PCR" (targeting *cps* genes G, N, P, Q and R) and then a "Typing PCR" (targeting *cps* genes E, G, H, I, J, K, L, M, N, O and P), to assign the 35 serotypes originally described for *S. suis* to isolates.

Most singleplex and mPCR assays designed to identify *S. suis* serotypes have reported assignment of a 'serotype' to isolates deemed to be non-serotypable using antisera, as well as to isolates deemed to be truly non-encapsulated based on surface hydrophobicity and electron microscopy [79, 80]. However, even with the use of molecular 'serotyping' based on PCR assays, non-serotypeable isolates are still commonly recovered from both clinically healthy and diseased pigs. In recent years, 17 novel *cps* loci have been identified from non-serotypeable *S. suis* isolates and designated novel *cps* loci 1 to 16 and serotype Chz [81-83]. As a result, an 18-plex Luminex assay has been developed to detect these 17 novel *cps* loci and nearly 60% of non-serotypeable strains from healthy pigs identified as carrying one of these novel loci [82].

Accurate *S. suis* serotype identification remains of significant epidemiological importance for the control of swine infections since different serotypes are prevalent in different geographic locales, and based on current data serotypes 2 and 14 are most strongly associated with zoonotic disease [17]. However, neither traditional serotyping techniques nor molecular 'serotyping' can discriminate between serotypes 1 and 14 or between serotypes 2 and 1/2, as each of these pairs cross-react *in vitro* and are also inseparable by mPCR due to the similar gene content of the respective *cps* loci. Recently, development of an automated pipeline has been shown to solve this three-decade longstanding *S. suis* serotyping issue by using differential alignment of short read DNA sequencing data to a custom built *cps* loci database, successfully discriminating between serotypes 1, 2, 14 and 1/2 based on a missense mutation in the *cpsK* gene [84], although the practical usefulness of such an approach is arguable.

1.1.1.4 Molecular typing

In addition to molecular 'serotyping' a number of other methods have been used to identify and discriminate the genetic diversity of *S. suis* as part of epidemiological studies, including 16-23S rDNA intergenic spacer PCR, restriction fragment length polymorphism (RFLP, or a combination of both (ISR_RFLP) [85, 86]), amplified fragment length polymorphism (AFLP) [87], arbitrarily primed PCR [88, 89], multilocus sequence typing (MLST; see section 1.1.1.4.1) [90], pulsed-field gel electrophoresis (PFGE) [57, 91-94], repetitive extragenic palindromic (REP) or enterobacterial repetitive intergenic consensus (ERIC) PCR [95-97], restriction endonuclease analysis [98-101] and ribotyping [102-106]. Of these methods, PFGE has been shown to be highly discriminatory and reproducible [92, 93], but is also laborious and time consuming. MLST has also been shown to have a high discriminatory power and high degree of reproducibility but is difficult to apply to routine diagnostic testing. In fact, each approach has its own limitations either requiring large amounts of sample DNA, which is labour intensive and cumbersome, or requiring high levels of technical competence and so these have been superceded as sequencing technologies have become more accessible.

It has been observed that *S. suis* isolates recovered from clinically healthy animals are more heterogeneous in comparison to the majority of isolates recovered from cases of disease in pigs or humans [91, 101, 106]. For example, most serotype 2 isolates recovered from cases of septicaemia were found to produce a common ribotype that was distinct from heterogeneous ribotypes of less virulent isolates [106]. Similar results, observed using PFGE, have also been reported for *S. suis* isolates recovered from pigs with meningitis and septicaemia which showed a significantly higher degree of genetic homogeneity in comparison to that of isolates recovered from pigs with signs of pneumonia and from healthy pigs. This provides support for the theory of a clonal-type relationship existing among highly virulent strains[91].

An important observation of epidemiological studies using the aforementioned molecular techniques to characterise and sub-type *S. suis* is many used a small sample size of isolates.

Furthermore, epidemiological studies have often been restricted to a single *S. suis* serotype (more often than not serotype 2 [41]) or single geographical locale, and use a variety of traditional approaches that make it difficult to compare results between studies. However, a general conclusion to draw from these studies is that the *S. suis* sp. is genetically very diverse, a feature that should be the focus of future diagnosis, surveillance and strategies to control clinical disease [11].

1.1.1.4.1 Multilocus sequence typing

MLST is a simple and unambiguous procedure for characterising isolates of bacterial species using the sequences of internal fragments of core metabolic (housekeeping) genes [107]. Since first described, using *Neisserria meningitidis* (the meningococcus; also a highly recombinogenic species) as a model organism, MLST has since been applied to numerous pathogenic species, including *Staphylococcus aureus*, *Streptococcus pneumoniae* and *Streptococcus pyogenes* [108-110]. There are two approaches to achieve high-levels of discrimination between bacterial strains within a population. The approach employed by more traditional techniques such as PFGE and ribotyping targets uncharacterised regions of the genome that are highly variable within the bacterial population, and as a consequence the output signal is complex and changes significantly over time as diversity in those regions progresses. The alternative approach, typified by multilocus enzyme electrophoresis (MLEE), uses variation that is accumulating very slowly in the population and that is likely to be selectively neutral. Although a much smaller number of alleles can be identified within the population, high levels of discrimination can be achieved by analyses of multiple loci, ideally spread across the genome. MLST is an adaptation of the useful and proven concepts of MLEE, identifying diversity of alleles from the nucleotide sequences of internal fragments of housekeeping genes rather than by comparing the electrophoresis motilities of the enzymes they encode [107]. The advantage of MLST over other molecular typing methods is that sequence data are truly comparable between laboratories, permitting one expanding global database, per species, to be placed on the Internet.

A MLST scheme for *S. suis* was first described in 2002 by King *et al.* [90]. Although some MLST schemes have a number of loci in common, it has not proved possible to define a set of core metabolic genes that are universally applicable to a wide range of bacterial pathogens [111]. The MLST scheme for *S. suis* uses specific loci of the following seven core metabolic genes: *aroA* (encoding 5-enolpyruvylshikimate 3-phosphate synthase), *cpn*60 (encoding a 60 kDa chaperonin), *dpr* (encoding a putative peroxide resistance protein), *gki* (encoding glucose kinase), *mutS* (encoding a DNA mismatch repair enzyme), *recA* (encoding homologous recombination factor) and *thrA* (encoding aspartokinase/homoserine dehydrogenase) [90]. Originally developed using 294 isolates of *S. suis*, a total of 92 unique sequence types (ST) and three major ST, or clonal, complexes (ST1, ST27 and ST87) were defined. ST1 complex was the largest (representing 165 of the original isolates studied) and most strongly associated with classic *S. suis* invasive diseases, including arthritis, meningitis and septicaemia [90]. In comparison, ST27 and ST87 complexes were found to represent a significantly higher proportion of isolates associated with pneumonia. No evidence was found of *S. suis* clones with a preference to infect humans, in fact the authors speculated that the incidence of human infection may simply follow the dominant clones in the local pig population [90]. Isolates of many of the most common serotypes were distributed widely across the MLST dendrogram. As well as highlighting that serotype is often a poor indicator of genetic relatedness between *S. suis* isolates, these findings suggest that capsular genes may be moving horizontally through the *S. suis* population; a feature that might be responsible for the change in most prevalent serotypes over time in certain geographical locations [90]. The horizontal spread of capsular genes has been demonstrated in *Streptococcus pneumonia* [112], where, as in *S. suis*, capsular genes which are conserved between serotypes flank variable serotype-specific loci encoding antigenic specificity [66] and recombination between the conserved regions results in serotype exchange. In *Streptococcus pneumoniae* the proposed mechanism involves natural genetic transformation (competence) although this has yet to be confirmed as the mechanism in *S. suis*. Alternatively, similar horizontal movement of surface markers such as the M protein is known to occur in other non-naturally transformable streptococci such as *Streptococcus pyogenes* [113].

The reported finding of 161, out of 294, isolates representing one ST is surprising. Such a high occurrence of one ST would suggest the MLST scheme for *S. suis* lacks high power discrimination. However, the mean number of alleles per locus was 40, providing the theoretical potential to distinguish >1.6x 10^{11} different genotypes. Hence, a more likely explanation is that ST1 isolates represent a highly successful clone which arose relatively recently and which has spread around the world. MLST is a popular technique due to the ease of carrying out the technique and ability to share comparable information between laboratories, on the Internet (https://pubmlst.org/ssuis/), before peer review. The repeated finding of ST1 complex isolates causing invasive disease indicates this ST defines strains with an increased capacity to cause disease. This might be reflective of a variety of factors such as increased fitness, the possession of certain virulence factors or allelic variants thereof, or particular antibiotic resistance profiles. Facts that appear to be confirmed by the analysis of whole genome sequencing data of strains belonging to ST1 and ST27 complexes. However, the MLST of 20 *S. suis* isolates recovered from human cases of infection in Thailand assigned 80% to the ST27 complex [114]. This is an interesting finding, as these isolates were recovered from the blood or cerebrospinal fluid suggesting a high degree of invasiveness, and brings into question the proposed lower potential of ST27 complex strains to cause invasive disease in swine. However, as mentioned above despite a high discriminatory power and high degree of reproducibility, MLST is difficult to apply to large scale routine diagnostic testing.

1.1.1.4.2 Minimum core genome sequence typing

Minimum core genome (MCG) sequence typing is a method only recently described for *S. suis* that uses the Bayesian clustering tool *STRUCTURE* to establish population genetics-based sub-divisions for strain identification and typing [115, 116]. MLST has become the preferred method for genotyping many biological species due to its ability to identify major phylogenetic clades, molecular groups, or subpopulations of a species, as well as individual strains or clones. However, MLST is difficult to is difficult to apply to routine diagnostic testing and can sometimes lack the discriminatory power to differentiate bacterial strains into virulent and avirulent sub-populations, which limits its use in epidemiological investigations. MCG sequence typing, rather than characterise isolates based on the slowly accumulating variation in the aforementioned seven core metabolic (housekeeping) genes described for MLST [90], is one approach that exploits advances in next-generation sequencing and uses genomics derived data to identify novel regions of the core-genome that can be used to identify and type *S. suis* into "MCG groups" that can later be associated with clinical phenotypes. A total of ten SNPs distributed across six genes are required for MCG sequence typing, including i) SSGZ1_0088 (SNP positions: 81404, 81419 and 81999), ii) SSGZ1_0114 (SNP position: 107453), iii) SSGZ1_0530 (SNP position: 572576), iv) SSGZ1_ 0776 (SNP position: 822644), v) SSGZ1_0777 (SNP positions: 824818 and 825000) and vi) SSGZ1_1981 (SNP positions: 2028696 and 2028744). Table 1.2 summarises the SNPs per gene used for MCG sequence typing. To date, MCG sequencing typing has successfully been applied to *S. suis* [115-117] and *Legionella pneumophila* [118].

Table 1.2. Single-nucleotide polymorphisms used for Minimum Core Genome Sequence Typing. Summary of the SNPs used for MCG sequencing typing, adapted from Zheng *et al.* [116]. SNP positions correspond to *S. suis* strain GZ1 (GenBank accession: NC_017617). Chen *et al.* [115] defined seven MCG groups, of which MCG group 7 is separated into three subgroups, 7-1, 7-2 and 7-3, as there are no unique SNPs that define MCG group 7. [a]E, epidemic strains, [b]N, ungroupable.

SNP in *S. suis* GZ1	**E[a]**	**1**	**2**	**3**	**4**	**5**	**6**	**7-1**	**7-2**	**7-3**	**N[b]**
2028696	A	G	G	G	G	G	G	G	G	G	G
2028744	A	A	G	G	G	G	G	G	G	G	G
824818	T	T	C	T	T	T	T	T	T	T	T
822644	C	C	C	T	C	C	C	C	C	C	C
107453	A	A	A	A	G	A	A	A	A	A	A
825000	G	G	G	G	G	A	G	G	G	G	G
81999	G	G	G	G	G	G	T	G	G	G	G
81404	A	A	A	A	A	A	A	G	A	A	A
81419	T	T	T	T	T	T	T	T	G	T	T
572576	G	G	G	G	G	G	G	G	G	A	G

1.1.2 Epidemiology

S. suis is a very successful coloniser of mucosal surfaces. Indeed, the upper respiratory tract of pigs, in particular the palatine tonsils and nasal cavities, is considered to be the natural habitat of *S. suis*, although the organism can be recovered from the gastrointestinal and genital tracts [11]. Pigs of any age can be infected with *S. suis*, but susceptibility generally decreases with age post weaning [59]. Broadly speaking serotypes 1-9 and 1/2 are most frequently recovered from diseased pigs worldwide, and serotypes 10-34 most likely to sub clinically colonise the upper respiratory tract and vagina than cause disease [85]. More than one serotype of *S. suis* often colonises individual pigs [11].

1.1.2.1 Infection in pigs

1.1.2.1.1 Clinical signs and pathology

S. suis is considered the incriminated haemolytic streptococci responsible for a wide range of clinical manifestations, including septicaemia with sudden death, arthritis, endocarditis, meningitis, and pneumonia amongst other diseases [2-11]. Other less common outcomes of *S. suis* infection are also possible, including abortion, bronchopneumonia, pericarditis, polyserositis and rhinitis [11]. The most consistent clinical signs of acute infection start with anorexia, depression, fever, lassitude and lateral recumbency, followed by symptoms depending on the localisation i.e. arthritis, ataxia, blindness, convulsions, dyspnoea, erythrema, lameness, opisthotonus, paddling and tremor [11]. Some animals with acute infection survive, resulting in persistent or chronic infection. In the latter, lameness and residual nervous signs such as otitis interna can be evident, as well as chronic arthritis [9, 25]. Sudden unexpected death in pigs has also been reported, and where positive detection of *S. suis* from the brain or meninges was not possible meningo-encephalitic lesions have been found during histopathological analyses [9, 119]. In cases of septicemia, the positive detection of *S. suis* from the brain, heart, liver and/or spleen is common, and inflammatory lesions are often found within organs (but without a typical pattern of lesion) [120]. Suppurative or fibrinopurulent inflammation of the brain, heart,

lungs and serosae are the most common histopathological observations in cases of septicemia [11].

It is also common to recover *S. suis* from the respiratory tract of pigs with signs of respiratory disease, however, it is unclear whether *S. suis* is responsible for pneumonia. *S. suis* is often recovered from the respiratory tract in combination with other confirmed respiratory pathogens, including *Actinobacillus pleuropneumoniae, Bordetella bronchiseptica, Haemophilus parasuis* and/or *Pasteurella multocida* [120, 121]. Indeed, concurrent viral infections might also predispose pigs to infection with *S. suis*, including Aujeszky's disease (*Suid herpesvirus*), porcine circovirus (PCV) and/or porcine reproductive and respiratory syndrome (PRRS) virus [122-124]. It has been suggested that *S. suis* is an opportunistic pulmonary pathogen, a hypothesis supported by the reported difficulty to reproduce observed clinical phenotypes of respiratory disease in experimental models of infection unless pre-infected with *Bordetella bronchiseptica* [53]. In contrast, in pigs with meningitis or meningoencephalitis, *S. suis* is considered a primary pathogen since it is often the only bacterial species isolated from the brain of these pigs [121].

1.1.2.1.2 Transmission

Vertical and horizontal transmission of *S. suis* has been demonstrated in all ages of pigs [98]. Adult pigs and healthy carrier piglets are the major reservoir of this pathogen and key players in the epidemiology of *S. suis* diseases [23]. Young pigs are most susceptible to invasive infections around the time of weaning and mixing (four to eight weeks of age) although disease can also occur later in growing pigs. Vertical transfer from sow to piglet occurs during parturition via contamination from vaginal colonisation, or shortly after during suckling, or due to close contact with the sow, her faeces, and the surrounding environment [98, 125, 126]. It has been proposed that once a pig is infected with *S. suis* serotype 2, it may remain a carrier for life [125]. Horizontal transmission is also important, especially during outbreaks, when diseased animals shed higher numbers of bacteria, increasing transmission by direct contact or aerosol [126]. No seasonal incidence has been noted in pigs, although incidence varies from herd to herd and also

within a herd [4]. Multiple *S. suis* serotypes from diseased animals within the same herd have also been isolated [121].

1.1.2.1.3 Survival in the environment

The resilience of *S. suis* in various environmental conditions has been studied in serotype 2 strains. Viability in water at 4 ˚C is retained for up to two weeks [127]. *S. suis* can also survive in water at 60 ˚C for 10 minutes and as a result it has been suggested that the scalding process in abattoirs could be a possible source of contamination. In experimentally infected faeces, *S. suis* is able to survive at 0 ˚C for up to 104 days, at 9 ˚C for up to ten days and 22-25 ˚C for up to eight days. *S. suis* is also able to survive in dust at 0 ˚C for up to 54 days at 9 ˚C for up to 25 days; however, isolation from dust incubated at room temperature (22-25 ˚C) for 24 h failed [127]. *S. suis* viability is also retained in rotting pig carcasses stored at 4 ˚C for six weeks and in carcasses left at 22-25 ˚C for 12 days, and should be considered as a potential source of contamination into (or onto) birds, dogs, mice and rats [128].

1.1.2.1.4 Pathogenesis of infection

The biological mechanisms that lead to a diseased state in pigs (and in humans) due to *S. suis* infection are not well understood. Hampered by the controversial definition of virulent and avirulent strains, most studies on pathogenesis of infection have been carried out with *S. suis* serotype 2 strains [41]. A well-documented invasive pathogen, *S. suis* is also considered a very successful coloniser of mucosal surfaces and thought to be able to colonise the palatine tonsils of pigs for a long period of time [129]. Not all carrier animals go on to develop clinical disease, however, some colonised pigs do develop septicaemia and/or meningitis and in order for such events to happen *S. suis* must first be able to disseminate from the palatine tonsil (and/or other mucosal surfaces) and reach the cardiovascular or lymphatic system. It is not well understood how *S. suis*, is able to breach the mucosal epithelia in the upper respiratory tract, although both the palatine and pharyngeal tonsils are considered principal points of entry [41, 129-131]. It has

been reported that virulent isolates of *S. suis* can both invade and lyse HEp-2 cells, a continuous laryngeal epithelial cell line of human origin [132]. However, invasion of epithelial cells is controversial since adhesion but failure to invade epithelial cells has been reported in studies of other epithelial cell lines using virulent strains of *S. suis* [133], although importantly both the strains of *S. suis* and the epithelial cell lines were different in these respective studies. More recent studies add to the confusion of how *S. suis* reaches the circulation, reporting the invasion of HEp-2 cells by four non-encapsulated strains (deemed to be non-virulent by the authors), but not by any of ten virulent serotype 2 strains [134, 135].

Once in circulation (cardiovascular or lymphatic), *S. suis* is able to survive and replicate [11]. An early "Trojan horse theory" suggested that, in the absence of specific antibodies, *S. suis* is phagocytosed by monocytes and transits the bloodstream intracellularly crossing the blood-brain barrier inside monocytes [136, 137]. However, subsequent studies in numerous groups indicated that *S. suis* travels in the bloodstream extracellularly, either free in circulation or attached to the surface of monocytes ("modified Trojan horse theory") [41, 131, 138]. Virulence factors associated with *S. suis* infection are discussed in depth in section 1.1.3.

1.1.2.1.5 Immunity

Passive immunisation as well as maternal antibodies can protect against *S. suis* infection. In fact, early studies showed that transfer of serum from convalescent pigs was able to induce complete protection against subsequent *S. suis* serotype 2 infection [2]. Protection has also been shown to be passively transferred to susceptible pigs by the inoculation of sera from pigs repeatedly immunised with live virulent *S. suis* serotype 2 strains, suggesting that humoral immunity plays an important role in *S. suis* infections [139]. Maternal antibodies against *S. suis* serotype 2 can be transferred from vaccinated sows to their piglets. However, it is important to note that sows can respond poorly to vaccination with whole cell vaccines, and good protection of piglets can be obtained only when titers of maternal antibodies reach a high level (25,600 or more, titers were

defined as the reciprocal of the lowest serum dilution having an OD_{540} equal to at least the mean optical density of four wells containing negative controls plus three standard deviations) [140].

1.1.2.1.6 Prevention, treatment and control

In addition to the virulence attributes of individual isolates of *S. suis*, it is important to recognise other additional factors also capable of contributing to the development of a diseased state in pigs. *S. suis* emerged as an important bacterial pathogen early in the 20th Century alongside intensification of the swine industry [11, 141]. *S. suis* has been reported worldwide in both traditional backyard and modern intensive swine operations, and as with other bacterial and viral infections, stress can precipitate *S. suis* infection. Crowding, poor ventilation, sudden weather change, mixing, moving, vaccination and concurrent disease are all stresses that predispose pigs to *S. suis* infections [142]. Understanding and the control of these individual factors, as well as, good farm management practices such as all-in/all-out pig flow can help regulate disease.

1.1.2.1.6.1 Farm management practices

Different management practices and/or the presence of other pathogens have been suggested as predisposing factors of *S. suis* infection [143]. Control of the environment and stress factors such as moving and mixing, (decreasing) population density and (increasing) ventilation, coupled with an all-in/all-out system of housing - with adequate cleaning and drying of rooms between batches should ideally constitute the first line of defense against mortality caused by streptococci [4, 98, 143]. However, in reality these management changes are the most difficult to implement and often-neglected in favour of early medication.

Production technologies such as medicated early weaning and segregated early weaning have been used to improve the health status of pig populations and to eliminate some infectious organisms, but proven ineffective for the elimination of *S. suis* carriage [95, 143, 144]. In fact,

S. suis has been recovered from the palatine tonsils of pigs in the presence of circulating antibodies in the bloodstream, as well as in pigs receiving penicillin via feed. Such a finding might go some way to explain why *S. suis* can persist indefinitely in a herd and why most attempts of control or eradication by blanket medication have failed [127, 142].

1.1.2.1.6.2 Antibiotics

Since current vaccines (see section 1.1.2.1.6.3) provide only partial serotype-specific protection, the use of antimicrobial agents has become increasingly important in treating and controlling the various clinical manifestations caused by *S. suis* infection. Numerous studies have investigated the antimicrobial susceptibility of *S. suis* strains recovered from diseased [21, 22, 145-152] and clinically healthy pigs [145, 147, 152-155] in different countries, reporting differences in the level of resistance and susceptibility to antimicrobials between countries, serotypes and over time [21, 22, 147, 152]. Such observations highlight the importance of monitoring the susceptibility pattern of *S. suis* to antimicrobials such as the β-lactam, penicillin - still the treatment of choice for human and pig infections.

In general, *S. suis* is sensitive to the inhibitory effects on bacterial cell wall biosynthesis of β-lactam antibiotics, a feature that has led to the prophylactic use of this broad-spectrum group of antibiotics at periods of high risk such as weaning. Clinically infected pigs are usually treated with ampicillin or benzylpenicillin (penicillin G), although most *S. suis* isolates are also susceptible to amoxicillin, ceftiofur, cefquinome, cephalexin, enrofloxacin and florfenicol [21, 22, 146, 147, 149, 156, 157]. As in humans the host inflammatory response against bacterial infection may be detrimental in some cases and adjunctive therapy with anti-inflammatory agents is recommended for treatment of *S. suis* meningitis in pigs.

β-lactam antibiotics other than cephalosporins (ceftiofur and cefquinome, antibiotics reserved for use only as a last resort by the UK's Pig Veterinary Society), especially penams (ampicillin, amoxicillin, penicillin) are considered the first option to treat *S. suis* related disease in the UK.

However, in line with the use of other antimicrobials to treat various bacterial and fungal infections cases of *S. suis* infection difficult to treat with penicillin, as well as, several moderately and some resistant isolates are increasingly reported in Asian and European countries; meaning penicillin susceptibility should not be assumed for *S. suis* isolates [152, 154, 155, 158, 159]. The reason for penicillin resistance in *S. suis* isolates has not been confirmed, but has been proposed to be due to modifications in the penicillin-binding proteins of certain field strains rather than a result of modifications in β-lactamase production resistance [11].

In addition to a small number of *S. suis* isolates resistant to penicillin, MIC's have demonstrated varying levels of resistance to clindamycin, erythromycin, gentamycin, kanamycin, spectinomycin, sulfisoxazole, tetracycline, tilmicosin and trimethoprim/sulfamethoxazole [152, 154, 160]. Variations in antimicrobial usage from one country to another, variations in laboratory methodologies or variations in the serotypes tested may all contribute to apparent differences in antibiotic susceptibility. A major point of concern is the relatively high proportion of *S. suis* strains shown to be resistant to antimicrobial agents that are used to treat *S. suis* infections in pigs (tetracycline, trimethoprim/sulphamethoxazole) or, in addition, are used as a feed additive (tylosin). For tetracycline and tylosin (functionally related to tilmicosin) this trend was observed in Denmark and Sweden. In these countries a decrease in susceptibility to older classes of macrolides (including tylosin) and tetracycline was observed among historical (1967–1981) and contemporary (1992–1997) *S. suis* strains isolated from pigs, prompting the authors to suggest that these emerging rates of resistance might be explained by the intensive use of macrolides and tetracycline [21]. Whether this suggestion is also valid for *S. suis* strains isolated in other European countries is not known and can only be confirmed with large-scale epidemiological studies.

In a recent report, Hernandez-Garcia *et al.* found no evidence of a statistically significant change (negative or positive) in *S. suis* resistance to amoxicillin, penicillin or amoxicillin/clavulanate between 2009 and 2014 [152]. An observation thought to be in response to increased public

awareness of antimicrobial residues and governmental pressure on the swine industry to reduce the blanket use of antimicrobials, and ban of the use of antibiotics for non-medicinal purposes, such as in animal feed, in the EU since 2006. This observation concurred with APHA Quarterly Reports based on routine surveillance that describe the isolation of only a small number of penicillin-resistant *S. suis* isolates since 2009 with no evidence of an increasing trend in penicillin MIC's [161]. However, the presence of resistant and intermediate strains to penicillin does suggest the need for a continuous surveillance of the susceptibility pattern of this pathogen.

1.1.2.1.6.3 Vaccination

Passive immunisation (artificially) as well as maternal antibodies (natural) can protect against *S. suis* infections. Indeed, in (1966) early studies Elliot *et al.* [31] showed that the transfer of serum from convalescent piglets provided other pigs with complete protection against homologous challenge with *S. suis* serotype 2 infection, indicating that some *S. suis* antigens are able to provoke a specific immune response that is capable of preventing subsequent infection. However, despite ever increasing approaches and studies no effective vaccine is available that can protect against all *S. suis* infections in pigs or in humans regardless of serotype [11, 162]. Numerous reasons hamper the development of effective vaccines, including but not limited to, the number of existing serotypes, variation in virulence seen both among and within serotypic groups and limited understanding of factors contributing to virulence as well as protection [163]. Adding to the challenge of vaccine development is to find the right adjuvant to provoke an effective immune response, as well as ensure that it is safe [164].

1.1.2.1.6.3.1 Whole cell vaccines

Whole cell vaccines (bacterins) can be split into two groups, i) inactivated and ii) live, attenuated. Autogenous bacterins, prepared from the virulent strain of *S. suis* isolated on a farm with clinical problems and then applied to the same farm, are frequently used in the field [165-167]. However, in comparison to experimental and commercial bacterins, limited (if any) vaccine safety data is produced for autogenous bacterins and severe adverse reactions can occur [168, 169]. Inactivated vaccines are produced by killing the disease-causing agent with chemicals, heat or radiation resulting in a safe and stable microbe that cannot revert back to a disease-causing state. Inactivated (killed) whole-cell vaccines seem to induce significant protection against challenge with a homologous *S. suis* strain, but the protection appears to be serotype specific and inconsistent. For example, a formalin-killed pathogenic *S. suis* serotype 2 bacterin has been shown to stimulate a complete protective response against homologous challenge in piglets - although, did require very high doses ($\sim 10^9$) of formalin-killed bacterins or repeated intravenous injections to induce protection (and could not be repeated using heat-killed organisms) [170]. In contrast, later studies of formalin-killed bacterins found no significant effect on the production of specific antibodies in vaccinated piglets, and the absence of passive protection against homologous *S. suis* serotype 2 challenges in mice [140]. Results obtained from experimental infection challenges were paralleled by field studies of a commercial serotype 2 bacterin that also failed to protect against nursery mortality among vaccinated pigs [171]. However, other experimental pig vaccination trials have reported protection with a bacterin, but that protection was subject to the use of a strong potentiating oil-in-water adjuvant; that also caused serious lesions at the injection sites [163, 172].

In comparison live, attenuated vaccines contain a version of the living microbe that has been weakened so that it cannot cause disease. However, although live, attenuated vaccines are the closest thing to a natural infection that elicit strong cellular and antibody responses, the remote possibility exists that an attenuated microbe could revert to a virulent form and cause disease. Strong humoral immunity in piglets has been induced following challenge with live *S. suis*

serotype 9 bacteria, despite other piglets presenting no detectable specific antibody response after bacterin immunisation [173]. As a result, of this and other similar observations it was suggested that a single dose of live either pathogenic or non-pathogenic *S. suis* could induce a strong protective response against subsequent challenge with homologous strains [123, 139, 170]. Consequently, several attempts to create a safe, attenuated live vaccine for *S. suis* have been reported. For example, temperature-sensitive mutants of *S. suis* serotypes 1, 2, 3 and 1/2 were evaluated in mice as potential vaccine candidates, and found to provide protection against challenge with strains of homologous serotypes but not heterologous serotypes - except for the serotype 1/2 mutant that also provided protection to challenge with serotype 1 and 2 strains [174]. In addition, the potential of a streptomycin-dependent mutant of *S. suis* serotype 1/2 has also been evaluated in mice as a vaccine candidate, where homologous and heterologous trials resulted in complete within serotype protection against challenge with *S. suis* serotypes 1 and 1/2 but only partial protection was observed against a challenge with *S. suis* serotype 2 strains [175]. However, due to *S. suis* zoonotic potential, the risks posed by a live-attenuated vaccine need to be carefully evaluated.

1.1.2.1.6.3.2 Subunit vaccines

As the volume of *S. suis* research increased more recent interest has shifted to the use of subunit vaccines. Subunit vaccines present a specific protein antigen to the immune system, rather than whole cell inactivated or live attenuated bacteria (bacterins). A subunit vaccine based on purified suilysin (see section 1.1.3.1.2) has been shown to protect pigs against challenge with virulent *S. suis* serotype 2 strains [176]. However, the absence of suilysin in a substantial number of isolates recovered from diseased pigs limits the value of this vaccine [177]. Extracellular factor protein and muramidase-released protein (see section 1.1.3.1.3) are both highly immunogenic and recognised by convalescent sera of pigs infected with virulent serotype 2 strains, and have been shown to be as effective as formalin-killed bacterins of *S. suis* serotype 2 [172, 178]. However, like suilysin, extracellular factor protein and muramidase-released protein are not produced by a large percentage of field strains worldwide, again limiting the efficacy of

this subunit vaccine [11]. In addition, a multitude of putative protective candidates for vaccination against *S. suis* infection have been reported by several research groups, using reverse vaccinology and/or different immunoproteomics approaches [179-181]. However, the protective capacities of these candidates remain to be verified experimentally.

1.1.2.1.6.4 Disinfectants

Disinfectants and cleansers commonly used on pig farms can inactivate *S. suis* serotype 2 strains in less than 1 minute, even at concentrations below those recommended by the manufacturers [125, 127]. Indeed, chlorhexadine, formaldehyde, 5% hypochlorite, 3% iodine, phenol and quaternary ammonium have all been shown to be effective disinfectants against *S. suis*, although 70% alcohol was not [182]. Liquid hand soap also inactivates *S. suis* seroype 2 in less than 1 minute at a dilution of 1 in 500 [127].

1.1.2.2 Infection in humans

S. suis is a zoonotic pathogen that causes similar clinical manifestations in humans as in pigs. Purulent or non-purulent meningitis is most common clinical symptom, followed by sepsis and septic shock [16, 17, 162, 183, 184]. In humans, other clinical manifestations include cellulitis, endocarditis, endophtalmitis, pneumonia, rhabdomylysis and uveitis [184]. Arthritis is also common, affecting elbows, hips, the sacroiliac joint, the spine, thumbs and wrists, and in most cases is thought to reflect generalised septicaemia [185]. Some instances of meningitis due to *S. suis* infection lead to sequela such as hearing loss and vestibular dysfunction [186, 187]. For unknown reasons, the recorded incidence of hearing loss due to *S. suis* infection is consistently higher than in patients with meningitis due to other bacterial infections, such as *Haemophilus influenza*, *Neisseria meninginitidis* and *Streptococcus pneumonia*, reaching 50% in Europe and 60% in Asia [188-190]. In addition, septicaemia with streptococcal toxic shock syndrome (STSS) has been documented as a feature of an important outbreak of acute infection in China during 2005 (see section 1.1.2.2.1). However, STSS [191, 192] had not (nor since) been reported due to *S. suis* infection, although has been described in invasive and non-invasive group A (*Streptococcus pyogenes*) [193] and *Staphylococcus aureus* [194, 195] infections.

In general, human *S. suis* infection is considered a rare and sporadic event associated with prolonged exposure to pigs or pork-derived products [190]. Most cases of human infection are reported in adult male dominated professions, such as abattoir workers, butchers, pig farmers, meat inspectors and veterinary practitioners [11]. Since the first human case was reported in Denmark in 1968 [12], in the region of ~1600 human cases of *S. suis* infection have been reported in European (Austria [196], Belgium [197], Croatia [198], France [199-203], Germany [187, 204-207], Greece [208-210], Hungary, Ireland [211], Italy [212-215], the Netherlands [190, 216], Poland [217, 218], Portugal [219, 220], Spain [58, 221-224], Sweden [225], the Untied Kingdom [186, 226-235]) and Asian countries (China [183, 236-238], Hong Kong [14, 158, 239-242], Japan [243], Korea [244], Madagascar [245], Malaysia [246], Singapore [247-249], Taiwan [250, 251], Thailand [13, 114, 252-264], Vietnam [15, 265-267]), as well as, in

Argentina [268], Australia [269, 270], Brazil [271], Canada [272-274], Chile [275], New Zealand [144], Togo [276] and the United States of America [277-279]; with many more probably never diagnosed or misdiagnosed [17]. Surprisingly, no cases have yet been reported in Russia, a country with an increased and developing pig production [17]. However, *S. suis* is increasingly becoming a public health concern in several Asian countries and was recently reported as the most common cause of adult meningitis in Vietnam, the second most common cause in Thailand and the third (behind *Streptococcus pneumoniae* and *Mycobacterium tuberculosis*) most frequent cause of community-acquired bacterial meningitis in Hong Kong [13-15]. Indeed, despite being reviewed as an "old neglected zoonotic pathogen" [16], the scientific community now considers *S. suis* one of the most important emerging/"re-emerging" infectious diseases in Asian countries, where most of the general population have regular contact with raw pork meat (which in certain markets has been found to be contaminated with *S. suis* [241, 280]) and consume "high-risk" lightly cooked pork derived products [162, 281].

In humans, the route of *S. suis* entry into the cardiovascular system is unclear. Under natural conditions, pigs are considered to be infected with *S. suis* serotype 2 via the mucosae of the nasopharynx and the palatine tonsils, whereas humans may be infected via small skin lesions, such as occupational wounds [190]. Popular opinion suggests entry to the cardiovascular or lymphatic system is most probably through lesions in the skin, however, in some cases wounds are not always obvious [162]. *S. suis* may colonise the nasopharynx, as well as, the gastrointestinal tract as suggested by diarrhoea as a prodromal symptom [253]. As in pigs, exposure to *S. suis* may lead to colonisation of the upper respiratory tract without leading to any obvious health consequences [162]. Alcoholism, diabetes mellitus and splenectomy have been proposed as important predisposing factors for the onset of *S. suis* disease in humans [234, 265, 282, 283]. Typically, *S. suis* isolates recovered from humans are genetically more homogeneous than those recovered from pigs, and most isolates recovered from humans share a similar phenotype and genotype to those recovered from swine within the same geographical locale [86, 92, 141]. Curiously, some isolates recovered from human patients are only weakly virulent for

pigs [53, 178]. It is difficult to understand why isolates recovered from humans are weakly virulent for pigs, one explanation could be that different routes of entry require different bacterial properties for the invasion process [284].

1.1.2.2.1 The 2005 outbreak in China's central Sichuan province

An unprecedented outbreak of acute *S. suis* disease in humans was reported across the central Chinese province of Sichuan in the summer of 2005, characterised by the rapid onset of disease affecting a large number of individuals, distinctive clinical manifestations (meningitis, sepsis or STSS) and the major challenges facing public health authorities in both surveillance and control [238]. Within a six-week period, 215 reported cases (66 laboratory confirmed, 149 probable) and 39 deaths were attributed to the outbreak across 12 prefectures and 203 villages; far exceeding the ten cases of meningococcal meningitis expected per month based on routine surveillance during the summer months (June-August) from 2003 to 2005 [238]. Almost all affected individuals were butchers or farmers, of previously good health, who had been involved in the backyard slaughtering practices of pigs that had died of unknown cause or been killed for food because they were ill. In all cases, there was no evidence for human-to-human infections [183, 236]. Decline of the outbreak was connected with measures that prohibited the domestic slaughter and consumption of sick pigs through provincial legislation, that was enforced with prosecution. It is important to mention, a concurrent disease outbreak among the local pig population was responsible for the death of 647 pigs in almost the same area as the outbreak in humans [238].

1.1.2.2.2 The 1998 outbreak in China's eastern Jiangsu province

The Sichuan outbreak in 2005 is the largest reported outbreak of severe acute disease in humans attributed to *S. suis* (attracting worldwide news coverage) surpassing a previous outbreak that took place in the eastern Chinese province of Jiangsu in 1998. Killing 14 of the 25 human patients and responsible for the death or culling of 80,000 pigs [236] little is known outside of China about this outbreak as all scientific reports [285-287] appeared in scientific journals in Chinese.

1.1.2.3 Infection in other animal species

S. suis has been recovered from a wide range of mammalian and avian species, despite the upper respiratory tract of pigs, in particular the palatine tonsil, being considered the natural habitat of the bacterium. As described in pigs the tonsil is a common carriage site, although recovery of viable isolates from the anal flora, genital tract and intestine is possible [11]. Isolation from cats and dogs [288-290], hamsters [291], horses [26, 290, 292], wild boar [293, 294], wild rabbits [295] and ruminants including calves, deer and lambs [44], as well as, birds [296] raise the importance of these animals as reservoirs in infections of pigs and humans. Indeed, the prevalence of *S. suis* in wild boar (*Sus scrofa ferus*) and wild rabbits (*Oryctolagus cuniculus*; an animal species far from pigs) is of particular importance regarding human infection, as numerous reports of hunters and poachers infected with *S. suis* after slaughtering wild animals are described in the literature [199, 206, 297, 298]. As a result *S. suis* infection has been described as an occupational risk affecting these professions/recreational activities [206].

1.1.3 Virulence factors

A wide range of homologs of bacterial virulence factors and virulence-associated factors found in other Gram-positive organisms has been shown, through targeted mutagenesis studies, to influence the virulence of *S. suis* strains [299-301]. Virulence and virulence-associated factors in *S. suis* (summarised in Table 1.3) have been the subject of a number of comprehensive reviews, one in particular by Fittipaldi *et al.* in 2012 [129]. Indeed, the increasing severity of human infections, including a shorter incubation time, more rapid disease progression and a higher mortality, emphasises the critical need to better understand the factors associated with the pathogenesis of *S. suis* infection [162]. However, clear association with specific roles in the onset and development of disease has not been found for many proposed factors [129]. The reason for this is unclear, although could be a product of previous studies being limited to a small number of isolates often restricted to serotype 2 [41].

A major limitation to the identification of *S. suis* virulence factors is the lack of, and urgent need for, a clear definition of virulence [302-304]. The virulence of a pathogen is commonly defined as the relative ability to cause harm to the natural host, but discrepancies in the clinical outcomes of challenge studies have been reported for the same strains of *S. suis* [54, 305, 306]. This feature is presumably due to differences in experimental design, pathogenicity being a quantitative trait with influences of pathogen genetics, host genetics and the environment. Indeed, it could equally be a product of important discrepancies existing in the concept of virulence between research groups. For example, different studies have designated *S. suis* field strains as virulent or avirulent based on i) the observed clinical phenotype of pigs from which the isolate was recovered (loosely grouped as diseased or clinically healthy animals), ii) the positive detection of virulence-associated proteins or iii) the use of different experimental challenge models; using different murine models, or pigs of different ages from either conventional or specific pathogen-free herds or caesarean-derived colostrum-deprived piglets [55, 56, 307-309].

Curiously, the positive detection of some proposed virulence factors in *S. suis* isolates does not necessarily define an isolate as virulent [129]. In fact, some isolates detecting positive for a specific virulence factor are avirulent while other isolates devoid of the same factor are still able to cause disease [129]. The bacterial virulence-associated factors extracellular factor protein and muramidase-released protein as well as the thiol-activated haemolysin known as suilysin are very good examples of this phenomenon (see sections 1.1.3.1.2 and 1.1.3.1.3).

Table 1.3. Putative and confirmed *Streptococcus suis* virulence factors.

Summary of *S. suis* virulence and virulence-associated factors, adapted from Fittipaldi *et al.* [129].

Virulence factor	Putative or confirmed function/role in virulence	Reference
38 kDa protein	Metabolism: phosphoglycerate mutase	[179]
44 kDa membrane protein	Unknown	[310]
6-phosphogluconate-dehydrogenase	Adherence: human epithelial cells HEp-2, and HeLa cells	[311, 312]
AdcR	Zinc-uptake regulator responsible for resistance to peroxide-induced oxidative damage	[313]
Adhesin P (Streptococcal adhesin Protein, SadP)	Hemagglutinin recognises the disaccharide galactosyl-α1-4-galactose (Galα1- 4Gal) in cell surface glycolipids	[314]
Protein homologous to *S. pyogenes* SAGP (adiS)	Stress-regulated. Part of the arginine deiminase system	[315, 316]
Amylopullulanase	Adherence: porcine epithelium NPTr (newborn pig tracheal cells), and porcine mucus	[317]
Catabolite control protein A (CcpA)	Gene expression repressor in the presence of excess sugar (sugar catabolism regulator)	[318]
Gene homologous to *S. pneumoniae* SP0844 (*cdd*)	Metabolism: cytidine deaminase	[319]
CiaRH	Two-component regulatory system (signal transduction)	[320]
Collagenase	Protease-U32 peptidase family (collagen degradation)	[319]
CovR	Orphan response regulator	[321]
Cps2C	CPS biosynthesis (chain length determination /export)	[76]
CpsE/F	CPS biosynthesis (glycosyltransferase)	[76, 322]
Di-peptidyl peptidase IV (DPP IV)	Adherence: human fibronectin-binding	[323, 324]
Dlt	Incorporation of $_{D}$-alanine residues into lipoteichoic acids	[325]
Dpr	Resistance to iron-mediated toxicity (tolerance of H_2O_2)	[326, 327]
Endo-β-*N*-	Processing of free and host surface oligosaccharides in the	[319]

acetylglucosaminidase D	cytosol	
Enolase (SsEno)	Adherence: human fibronectin-/plasminogen-binding	[328]
Extracellular protein factor	Unknown	[329, 330]
F8D	ATPase component of ABC-type multidrug transport system	[331]
Fibronectin (FN)- and fibrinogen (FGN)-binding protein (FBPS)	Adherence: human fibronectin- /fibrinogen-binding	[332]
FeoB	Iron transporter	[333]
Ferric uptake regulator (Fur)	Iron transport regulator	[313]
Glutamate dehydrogenase (GDH)	Metabolism: catalyses the (reversible) oxidative deamination of L-glutamate to α-ketoglutarate and NH_3	[69]
Glutamine synthetase (GlnA)	Adherence: human epithelial cells HEp-2	[299]
Gene homologous to *S. pyogenes glnH*	Unknown	[319]
Glyceraldehyde-3-phosphate dehydrogenase	Adherence: human epithelial cells HEp-2	[334, 335]
Gene homologous to *L. monoctogenes gtfA*	Metabolism: Sucrose phosphorylase	[319]
Gene homologous to *S. agalactiae guaA*	Metabolism: guanosine monophosphate (GMP) synthase	[319]
Gene homologous to *S. thermophilus guaB*	Metabolism: Inosine monophosphate dehydrogenase	[319]
H4F	Bacteriocin operon protein	[331]
Hyaluronate lyase	Catalyses the degradation of hyaluronic acid	[336]
IgA1 protease	Cleavage of human IgA1	[337, 338]
Lgt	Prolipoprotein diacylglyceryl transferase	[339]
Lipoprotein 103 (SSU0308)	ATP-binding cassette transporter implicated in zinc uptake	[340]
Lipoprotein signal peptidase (Lsp)	Signal peptide cleavage	[341]
Gene homologous to *S. mutans lpp*	Lipoprotein	[319]
LuxS	Quorum sensing	[342-344]
Mannose-specific PTS IID (manN)	Mannose-specific phosphotransferase system (PTS) transport	[319]
Muramidase-released protein (MRP)	Unknown	[330, 345]
Gene homologous to	Transcriptional regulator	[319]

S. thermophilus nadR		
NeuB	Sialic acid synthase	[77, 346]
Opacity factor of *S. suis* (OFS)	Surface-associated serum opacification	[300, 347]
Product of gene *pgdA*	Peptidoglycan N-deacetylation	[348, 349]
Phospholipase C	Induction of host arachidonic acid production	[350]
Protein homologous to *S. mutans* UA159 (purA)	Metabolism: adenylosuccinate synthetase	[319]
purD	Metabolism: phosphoribosylamine-glycine ligase	[319]
Response regulator of *S. suis* (RevS)	Orphan response regulator (part of a two-component signal transduction system without a flanking histidine protein kinase)	[351]
Protein homologous to *S. aureus* AgrA (RevSC21)	Orphan response regulator, influencing the attachment to the human epithelial cell line Hep-2	[352]
Rgg-like protein	Transcriptional regulator	[353]
Gene homologous to *S. agalactiae* SAG0907	Lipoprotein	[319]
Proteins homologous to *S. salivarius* SalK/SalR	Two-component signal transduction system	[354]
Gene homologous to *S. pneumoniae* SP1724 (*scrB*)	Metabolism: Sucrose-6-phosphate hydrolase	[319]
Gene homologous to *S. pneumoniae* SP1725 (*scrR*)	Metabolism: Sucrose operon repressor	[319]
Gene homologous to *S. mutans* SMU_61	Transcriptional regulator	[319]
Sortase A (SrtA)	Membrane-bound thiol transpeptidase enzyme (adhesion)	[355-358]
Gene homologous to *S. pneumoniae* spr1018	Unknown	[319]
Gene homologous to *S. pyogenes* spyM3_0908	Permease (amino acid ABC transporter)	[319]
Product of gene *srtF*	Pili adhesin	[348, 359, 360]
Product of gene *srtG*	Pili adhesin	[359, 361]
Secreted nuclease (SsnA)	Secreted, cell wall anchored DNase (degradation of DNA)	[362]
Surface-associated subtilisin-like protease (SspA)	Modulate (induce) cytokine secretion by macrophages	[363-365]
Streptococcal histidine triad protein	Unknown	[366]

Suilysin (SLY)	Thiol-activated toxin hemolysin	[309, 367-370]
Superoxide dismutase (SOD)	Intracellular survival in macrophages	[371]
Surface antigen one (Sao)	Unknown	[372-375]
Trag	Unknown	[301]
Gene orthologous to *C. perfingens treR*	Transcriptional regulator—trehalose utilization	[319]
TroA	High-affinity metal binding lipoprotein (manganese acquisition)	[376, 377]
VirA	Unknown	[319, 378]
Zinc uptake regulator (Zur)	Control of zinc homeostasis (resistance to zinc-mediated toxicity)	[379]

1.1.3.1 'Classic' virulence factors

Since the early 1990s and for several years thereafter, the capsular polysaccharide, the thiol-activated haemolysin known as suilysin, and the two proteins extracellular factor and muramidase-released protein were considered the most important virulence-associated factors of *S. suis*, resulting in the incorporation of serotyping and virulence-associated gene profiling into the routine diagnosis procedure for *S. suis* [11, 131]. Although the exact contribution of some of these factors in *S. suis* virulence remains unclear, for historical reasons, comparison is drawn against these factors in chapters 3 and 4 of this thesis and as a result they are described in depth here first under the title "classic virulence factors".

1.1.3.1.1 Capsule

It has been demonstrated by two independent laboratories that the capsule of *S. suis* serotype 2 is an important virulence factor [76, 305]. The capsular polysaccharide (CP) is recognised to play an important function in virulence of many Gram-positive and Gram-negative bacteria [380]. *S. suis* possesses an integral, cell associated polysaccharide capsule [381]. Early studies reported that virulent isolates of *S. suis* were able to resist phagocytosis in porcine blood absent of anti-*S. suis* specific antibodies, observations supported by the production of non-encapsulated isogenic mutants from virulent serotype 2 parent strains [11]. The absence of the CP correlated with increased hydrophobicity of *S. suis*, and increased phagocytosis by murine macrophages and both porcine macrophages [76, 305] and neutrophils [322]. As a result, non-encapsulated isogenic mutants were deemed to have become avirulent in murine and pig experimental models of infection [76, 305].

However, non-encapsulated *S. suis* strains might also invade host tissue, though potentially to a lower degree. This was shown by the isolation of isogenic non-encapsulated mutants from the central nervous system, joints and serosa using an intranasal experimental model of infection [76]. Indeed, significant differences in virulence have been described for strains belonging to

the same serotype [53], indicating that the capsule of *S. suis* is not sufficient for full virulence and that other factors have important functions in the pathogenesis of *S. suis*.

Cell wall components of *S. suis* induce a significant release of host proinflammatory cytokines (TNF-α, IL-1β and IL-6) and chemokines (IL-8). It is thought that the CP reduces the production of such cytokines by blocking the interaction of cell wall components with pattern-recognition receptors such as Toll-like receptor 2 [138, 382, 383]. Isolates of *S. suis* serotype 2 recovered from diseased pigs have been reported to possess thicker capsule in comparison to isolates recovered from clinically healthy animals [43, 384-386]. An increase of capsular thickness following growth *in vivo* has also been described for virulent but not avirulent isolates [384]. As a result, it has been proposed that *S. suis* down-regulates CP expression for increased adhesion to epithelial cells and then up-regulates CP expression for protection against phagocytosis as it enters the bloodstream [41]. However, other studies have been unable to reproduce any correlation between thickness of the capsular material and virulence [11]. *S. suis* serotype 2 is most frequently associated with clinical disease (see section 1.1.1.3). However, cells of the serotype 2 reference strain have not been shown to be protected by a significantly thicker layer of CP in comparison to the reference strains of other *S. suis* serotypes [11].

It has been proposed that the ability of *S. suis* serotype 2 isolates to cause invasive disease may be due to the composition of the capsular material itself, which contains N-acetyl neuraminic acid (sialic acid). Sialic acid has been proposed as an important virulence factor for several species of encapsulated pathogens responsible for meningitis, including *Escherichia coli* and group B Streptococcus [387]. The CP of serotypes 1 and 2 are composed of sialic acid and four additional sugars (namely glucose, galactose, N-acetyl glucosamine and N-acetyl galactosamine (serotype 1) or rhamnose (serotype 2) [33]). Serotypes 1, 2, 14, 27 and 1/2 have all been shown to harbour the genes involved in sialic acid synthesis [77]. Surface-associated sialic acid contributes to adhesion of *S. suis* serotype 2 isolates to murine macrophages, an adhesion not accompanied by phagocytosis [138]. As a result, the "modified Trojan horse theory" was

proposed that suggested *S. suis* isolates are able to cross the blood-brain-barrier through adherence to immune cells [41]. However, it has been shown that the capsule of both virulent and avirulent isolates has similar concentrations of sialic acid and proposed its association with virulence might be coincidental [386]. After all, despite the CP seeming to be a major virulence factor, most avirulent strains are also encapsulated [11]. Nevertheless, it is clearly an important feature for blood survival even if not essential for the initial penetration from epithelia.

1.1.3.1.2 Suilysin

Suilysin (SLY) is a pore-forming thiol-activated cholesterol-dependant cytolysin, similar in structure and function to other haemolysins, such as listeriolysin O (*Listeria monocytogenes*), pneumolysin (*Streptococcus pneumoniae*) and streptolysin O (*Streptococcus pyogenes*) [367, 368, 388-391]. *S. suis* is α-haemolytic on solid agar supplemented with sheep blood and can produce β-haemolysis on solid agar supplemented with horse blood, an ability attributed to SLY [11]. *In vitro*, SLY is an extracellular protein produced in the culture supernatant at the end of the exponential growth phase, peaking around 15 - 24 h. Haemolysis, in haemolytic units per mg, is comparable to listeriolysin O and pneumolysin, and is completely inhibited by free cholesterol [367, 368]. As with other thiol-activated toxins, haemolysis is significantly increased when incubated with a reducing agent, such as Dithiothreitol (DTT). Indeed, treatment of cell-free supernatant with oxygen significantly reduces haemolytic activity, but is fully reversible when then treated with DTT [367, 368]. Human group O, horse, pig and sheep erythrocytes are all similarly susceptible to haemolysis by SLY, via a multi-hit mechanism of action (multiple SLY to one erythrocyte) [368].

Thiol-activated toxins are produced by several genera of pathogenic, primarily Gram-positive, bacteria that cause very different diseases, and as more information becomes available about their role as important virulence factors it is apparent that their effects might be much more subtle than simple destruction through pore formation [388]. First purified and characterised from European virulent reference strains P1/7 [367, 391] and Henrichsen S735 [368], SLY

haemolysis was associated with the pathogenic attributes of these isolates. However, although described in pathogenic strains across an extensive number of serotypes in Europe [392], an equivalent number of other virulent strains, including many North American serotype 2 isolates [80, 393], were shown not to produce SLY; indicating SLY may not play a critical role in virulence [302, 394]. Numerous studies using SLY isogenic mutants and ST104 strains, with reduced SLY production due to two nucleotide insertions in the *sly* promoter [395], have been used to understand the actual role of SLY in virulence but the results are inconclusive. Indeed, murine intraperitoneal infection model experiments show SLY has an unequivocal role in virulence [309, 395, 396], while swine intranasal or intravenous infection models showed no difference between an allelic-replacement mutant and wild-type strains [309, 370]. Despite the discrepancies between the animal infection models (a common theme in the literature on *S. suis* [307, 308], see section 1.1.4) other *in vitro* studies have shown SLY plays a role in cytotoxicity to epithelial cells [133], endothelial cells [397, 398], neutrophils and macrophages [138, 322]. As well as, being shown to have antiphagocytic and antibactericidal properties directed against neutrophils and macrophages/monocytes [322, 399]. In summary, SLY cannot be considered a critical *S. suis* virulence factor [394], although if present SLY does probably contribute to the pathogenicity traits of SLY^{+} strains [302, 309, 370].

1.1.3.1.3 Extracellular factor and muramidase-released proteins

The two proteins, known as extracellular factor (EF) protein and muramidase-released protein (MRP) were first reported as potential markers capable of differentiating the different virulence properties of *S. suis* serotype 2 isolates in newborn germfree pigs [52, 178]. To date, the function(s) of both EF protein and MRP in the pathogenesis of the *S. suis* infection remain unclear. However, the association of EF and MRP with virulence is observed with strains from certain (mainly European) countries, and has led to the incorporation of the gene profiling of *epf* and *mrp* into the routine diagnosis procedure of *S. suis* in several laboratories [131]. Despite the suggestion that the production of EF and MRP may occur only coincidently in association with other 'true' virulence factors [330].

1.1.3.1.3.1 Extracellular factor protein

The 110 kDa EF protein (encoded by the gene *epf*) was identified due to its association with virulent serotype 1 and serotype 2 strains in newborn germfree pigs [178]. To date, the function of EF is still unknown. Isogenic *epf* mutants of serotype 1 and 2 strains have been shown to be as virulent as the parental wild-type strain in experimental infections indicating that EF is associated but not essential for virulence [330]. Large molecular weight (EF*; >110 kDa) variants of EF are expressed by some serotype 2 strains, and are characterised by long C-terminal tandem repeats (each 76 amino acids long). Based on the number of repeated sequence five different classes of EF* proteins have been differentiated [329]. Monoplex- and multiplex-PCR assays are available for differentiation of the different size variants of *epf* [68, 400]. In contrast, the N-terminus, namely the first 811 amino acids, of EF and EF* are nearly identical.

The 110 kDa EF and the high molecular variants of EF are immunogenic proteins. Convalescent phase sera from animals infected either with MRP+ EF+ or MRP+ EF* serotype 2 strains generally contain high antibody titers against EF [401]. Immunisation with EF alone did not elicit protection against MRP+ EF+ serotype 2 experimental infection challenges, but a vaccine containing purified MRP and EF together with a water-in-oil adjuvant was as protective as a

bacterin (see section 1.1.2.1.6.3.1) [172]. However, prevalence of EF and the high molecular variants of EF positive strains show substantial geographical variation. Indeed, a major problem of an EF-based vaccine is that many virulent *S. suis* strains, including all serotype 9 strains that are of considerable importance in North America, do not express this factor [401].

1.1.3.1.3.2 Muramidase-released protein

MRP is a 136 kDa protein, discovered as a factor released from virulent (in newborn germfree pigs) serotype 2 strains after treatment with muramidase (lysozyme) [178]. MRP is predominantly present in protoplast supernatant. Sequence analysis of the 3771 bp gene encoding MRP (*mrp*; SSU0706) and corresponding 1256 amino acid sequence revealed an LPXTG-motif at the C-terminus of the protein, a finding that indicates MRP is a cell envelope-associated protein [345]. In addition, at the N-terminus the first 47 amino acids were found to have the characteristics of a typical signal peptide, and taken together these findings resemble similar regions in several other surface proteins from Gram-positive bacteria. Indeed, MRP has since been shown to be a substrate of sortase A, a transpeptidase that anchors secreted proteins covalently to the cell wall [355]. Other peptidoglycan-associated proteins of streptococci have been described, such as the M protein in group A streptococci that is an important virulence factor as only M protein-positive strains resist phagocytosis and adhere to humans cells [402, 403]. The N-terminal of MRP shares some similarity (17.2% across 377-amino acids) with the sequence of the fibrinogen-binding protein of *Staphylococcus aureus* [404], although the binding of MRP to human fibrinogen could not be detected [345]. However, in contrast to the original finding more recent studies indicate MRP can in fact bind to the blood protein fibrinogen and this could be an essential step in the development of *S. suis* meningitis [405, 406].

Large (MRP*; >136 kDa) and small (MRP^S; <136 kDa) molecular weight variants have been described [400]. Duplication of 162 bp units of repetitive sequence present in *mrp* encoding for the 136 kDa protein is responsible for the larger size of MRP* [345]. The variability of the number of repeats found among the different *mrp* variants is reminiscent of the mechanism of

size variation in other streptococcal surface-associated proteins, such as SfbI [407]. In contrast, MRP^S is found in culture supernatant and thought to be due to MRP breaking from the peptidoglycan layer during growth of bacteria, resulting in unbound MRP in the culture supernatant that has a lower molecular weight than bound MRP in the protoplast supernatant [178]. Due to the correlation of MRP with virulence, despite the role of MRP remaining unclear, the determination of *mrp* genotype (or the respective phenotype) has been performed in numerous epidemiological studies [400, 408, 409], and is now included in the routine diagnosis of *S. suis* in several laboratories [131]. No difference in virulence was seen between isogenic mutants, created by disruption of *mrp* by insertion of an antibiotic (*spc*) resistance gene, in comparison to wild-type parent strains [330]. Neither genotype or phenotyping of *mrp* allows classification of a strain as virulent or not, and it is entirely possible that MRP merely occurs in association with true virulence factors.

1.1.3.2 Other proposed virulence-associated factors of *Streptococcus suis*

This review of virulence factors so far focuses on four "classical virulence factors" for *S. suis*, including the CP, the thiol-activated haemolysin known as SLY, and the two virulence-associated proteins EF and MRP. In fact, a wide range of homologs of virulence factors found in other Gram-positive organisms has been shown to affect the virulence of *S. suis* strains through targeted mutagenesis studies [299-301]. For example, other factors associated with *S. suis* disease include but are by no means limited to enolase [181, 328, 410, 411], hyaluronidase [336, 412], opacity-factor [300, 347, 413], peptidoglycan [349, 414, 415], phospholipase C [350], pili [348, 381], Sao (surface antigen one) [181, 372, 375, 401, 416, 417], sortase [355-358, 418], and fibrinogen/fibrinogen-binding proteins [180, 332, 419-421]. All have been the subject of a number of comprehensive reviews [129, 131, 302], the most recent of which “Critical *Streptococcus suis* virulence factors: are they all really critical?” reemphasising clear association with specific roles in the development of disease has not been found for many proposed factors [302].

1.1.4 Experimental infection models

There is an urgent need for a convenient, reliable and standardised animal model to assess the virulence potential of *S. suis* isolates [302-304]. At present, molecular markers are unable to distinguish virulent isolates of *S. suis* from avirulent commensal-like isolates, emphasising the need for better disease-associated genomic markers and/or animal models to quickly assess the virulence potential of different isolates. One of the most commonly used models for studying microbial infections is the rodent (mice and rats) model. However, as with pigs, there are budgetary, ethical and logistical hurdles associated with rodents as infection models. Firstly, maintaining a sufficient number of animals required to generate statistically relevant data is expensive and often regarded as ethically objectionable. Secondly, mammals have lengthy reproduction times, which slow the progress of experiments. The 3Rs (Replacement, Reduction and Refinement) are increasingly seen as a framework for conducting high quality science in the academic and industrial sectors with more focus on developing alternative approaches that avoid the use of animals [422]. It is therefore useful to contemplate alternative invertebrate models, if these can be shown to yield useful data. As a result, invertebrate hosts, such as *Caenorhabditis elegans*, *Danio rerio*, *Drosophila melanogaster* and *Galleria mellonella* larvae, have been introduced as alternative *in vivo* models to study microbial infections. Such models have contributed substantially to biomedical research over the last decade, and recently have been adapted to explore and better understand *S. suis* pathogenesis [423, 424].

1.1.4.1 Swine experimental infection models

Certainly, the pig remains the most valuable experimental model of infection [302]. Popular opinion considers field isolates of *S. suis* to breach the mucosal epithelium in the upper respiratory tract of pigs (see section 1.1.2.1.4). However, it has proved challenging to replicate clinical disease using natural routes of exposure producing consistent and reproducible results. As a result, a number of models have been developed where bacteria are inoculated i) intraperitoneally [425], ii) intravenously [426], iii) intranasally [52, 53] or iv) subcutaneously [427]. However, as with other well-established experimental models of infection variations in

genetic background complicate pathogenesis studies in the natural host. Age is also an important factor that should be considered, which might have an impact on susceptibility based on the presence of maternal immunity, as should the health status of the herd.

A number of studies have reported the use of pigs from specific-pathogen-free (SPF) herds [358, 428, 429]. The advantage of the SPF method is its convenience, cost efficiency and low technical requirement [430]. However, when the research requires freedom of infection with pathogens that are highly prevalent in pig populations, such as *S. suis* this method may be inadequate, as most pigs have antibodies against these pathogens, either maternal or acquired, or are actively infected with the pathogen of interest. Indeed, *S. suis* is considered a normal member of the swine microflora, and the carrier rates of this "early colonizer" can be up to 100%; with prevalence of disease generally being not more than 5% [431, 432]. As a result, researchers may have to screen a large number of farms and pigs to obtain a reliable pig source and then select pigs after the level of maternally derived antibodies has waned. Unfortunately, despite being developed over the last 40 years, the definition of SPF pigs is still controversial and does not include being free from *S. suis*; which would be almost impossible to achieve as *S. suis* is vertically transmitted from sow to piglet during parturition and suckling [23].

An alternative approach has been to use caesarian-derived colostrum-deprived (CDCD) pigs to study disease caused by *S. suis* [319], as well as, other pathogens such as *Haemophilus parasuis* [433-437]. The CDCD and gnotobiotic methods use cesarean section to obtain term piglets from pregnant sows. The CDCD pigs are raised in sterile compartments for several days and then in a clean room [438]. Gnotobiotic pigs are raised entirely in sterile compartments. CDCD pigs are an experiential infection model devoid of maternal immunity, and where the upper respiratory tract is confirmed to be free of *S. suis*. Although the CDCD and gnotobiotic methods are reliable for obtaining pathogen-free pigs, they have several disadvantages, including the need for surgery, specialised facilities, and sterile compartments, and a greater cost. CDCD pigs are considered to produce fairly consistent results, however, results from CDCD pigs should be

treated with caution as animals in general are highly susceptible to infection in comparison to 'normal' piglets [302].

1.1.4.2 Murine experimental infection models

One of the most commonly used models for studying microbial infections is the murine model. Intraperitoneal and intravenous mouse models have been used to better understand *S. suis* pathogenesis [439, 440]. However, a number of studies have reported isolates associated with disease in pigs to appear much less virulent or avirulent in murine models, and vice-versa [307, 308]. Animal age and the volume of inoculum have been documented as influencing murine models. In fact, mucosal irritation by acetic acid is often required in order to observe disease in mice. Such observations indicate that murine models do not mimic events in pigs, and that *S. suis* pathogenesis could be species dependant. As a result, wherever ethically possible experimental model infections should be performed in the natural host (the pig) [441].

In summary, there remains an urgent need to reach a consensus in i) the experimental conditions, ii) which inbred/outbred mouse strain to use, iii) the most appropriate and reproducible swine model, iv) the route of infection in both mice and pigs, and v) the clinical scores to measure disease outcome.

1.1.4.3 Zebrafish

The zebrafish (*Danio rerio)* is a derived member of the genus *Danio* in the family Cyprinidae, and was first identified as a genetically tractable organism in the 1980s. Although its contribution has already been substantial (notably to understanding basic vertebrate biology and vertebrate development, as well as, factors controlling the specification of cell types, organ systems and body axes of vertebrates [442-444]), zebrafish research holds further promise to enhance understanding of the detailed roles of specific genes in disease, both rare and common [445]. Zebrafish are members of the teleostei infraclass, a monophyletic group that is thought to have arisen approximately 340 million years ago [446]. Although mammals may appear to be extremely different than zebrafish, a direct comparison of the zebrafish and human protein-coding genes revealed 71.4% of human genes have at least one zebrafish orthologue, and reciprocally 69% of zebrafish genes have at least one human orthologue. Moreover, zebrafish have bile ducts, blood, bone, brain, cartilage, ear, esophagus, eyes, heart, intestine, kidney, liver, mouth, muscle, nose, pancreas, spinal cord and teeth. Many of the genes and critical pathways that are required to grow these features are highly conserved between humans and zebrafish. Thus, any type of disease that causes changes in these body parts in mammals could theoretically be modeled in zebrafish. In fact, zebrafish have been used extensively as an animal infection model for a wide range of Gram-negative and -positive bacterial pathogens, including but not limited to *Salmonella typhimurium* [447, 448], *Staphylococcus aureus* [449] and numerous *Streptococcus* species. [450].

The use of adult [451-453] and pre-feeding stage larvae [424] zebrafish have been investigated as a convenient and reliable infection model to assess the virulence potential of *S. suis* isolates. Zebrafish breed readily (~ every 10 d) and can produce as many as 50-300 eggs at a time, this is different to mice that generally produce litters of 1-10 pups and can bear ~ three litters in their lifetime. Pre-feeding zebrafish larvae (up to six days post fertilization) are cheap to rear in large numbers, and exempt from ethical legislation. The zebrafish embryonal and larval innate immune system develops rapidly, resulting in the presence of functional phagocytes,

complement factors, and antimicrobial enzymes in the embryo before or soon after hatching [454-456]. This has allowed high-throughput screens to be performed that would be much less feasible in mammalian experimental models of infection. Pre-feeding zebrafish larvae at 72 h post fertilization have been shown to be susceptible to infection by *S. suis* isolates virulent in pigs [424]. In fact, microinjection of different *S. suis* isolates into pre-feeding zebrafish larvae resulted in highly reproducible dose-dependent and strain-dependent larval death, correlating with the original virulence of the isolates in pigs. As a result, Zaccaria *et al*. consider pre-feeding stage zebrafish larvae to be a rapid and reliable model to assess the virulence of *S. suis* isolates causing clinical disease in pigs [424].

Adult zebrafish have been used to evaluate the virulence potential of *S. suis* serotype 2 isolates [451]. Adult zebrafish have also been used in transcriptome profiling experiments, resulting in the identification of 189 genes (125 genes up- and 64 genes down-regulated) differentially expressed during *S. suis* infection [452, 453]. However, as with the natural mammalian host, adult zebrafish are not exempt from ethical legislation limiting the usefulness of the model. In addition, an important caveat of the zebrafish experimental infection model is that experiments must be performed at 28 ˚C [424].

1.1.4.4 Greater Wax Moth

The Greater Wax Moth, *Galleria mellonella*, is a member of the Galleriinae subfamily within the family Pyralidae of Lepidopteran order. Greater and close relative Lesser (Lepidoptera: Pyralidae, *Achroia grisella*) wax moths are ubiquitous pests of honeybee colonies. The economic importance of wax moths has led to a number of investigations into wax moth behaviour, biology, ecology, molecular biology, physiology and control. However, despite the importance of wax moths to the apicultural industry, they are investigated considerably more as a model organism for studies in genomics, insect physiology and proteomics; this is especially true for the Greater Wax Moth [457].

Greater Wax Moth larvae have recently been explored as a simple and cost-effective *in vivo* experimental infection model to screen the virulence potential of *S. suis* isolates [423]. Experimental infection with *S. suis* isolates belonging to different serotypes killed Greater Wax Moth larvae in a dose-dependent manner. Crucially, the observed virulence was in agreement with virulence observed in piglets and humans, as well as, other experimental infection models such as mice and zebrafish larvae [424]. Infection with heat-inactivated bacteria or bacteria-free culture supernatants showed that in most cases live bacteria are needed to cause mortality in Greater Wax Moth larvae; which should be considered when designing novel *S. suis* diagnostics. In addition, Greater Wax Moth larvae infected with the virulent serotype 2 strain, *S. suis* S3881/S10, were able to be rescued using antibiotic therapy. As a result, Velikova *et al*. consider Greater Wax Moth larvae to be a simple and cost effective model to assess the virulence potential of *S. suis* isolates [423].

It is the caterpillar larvae (or 'wax worm') and not the adult moth that is used as an animal model. Larvae of the Greater Wax Moth have been shown to be an excellent model organism for *in vivo* models of microbial infection and for antimicrobial drug screening, replacing the use of small mammals [458-464]. The Greater Wax Moth larvae do not require ethical approval and their short life span makes them ideal for high-throughput studies. In fact, when compared with

traditional mammalian model hosts, the Greater Wax Moth larvae are cheaper to establish and easier to maintain as they do not require specialist equipment.

The Greater Wax Moth is a typical holometabolous insect and develops through four distinct life stages, i) egg, ii) larva, iii) pupa and iv) adult (imago) (Figure 1.1). Greater Wax Moth larvae are polipod (eruciform), with six legs on the thorax and a number of prolegs on the third to sixth abdominal segments. The larvae are cream white in colour, with sclerotized body parts, although darken with each successful molt [465]. At the larval stage, sexing into male and female is not possible due to the absence of sex specific external morphological characters. Upon hatching, Greater Wax Moth larvae are ~1–3 mm in length and 0.12–0.15 mm in diameter. Larvae undergo 8–10 moulting stages and spin silk threads across all stages, but only the last instar spins a cocoon. Prior to pupation, late instar stage larvae are ~25–30 mm in length and ~5–7 mm in diameter.

Figure 1.1. Lifecycle of the Greater Wax Moth (Lepidoptera: Pyralidae, *Galleria mellonella*).

Overview of the Greater Wax Moth lifecycle, adapted and updated from Kang *et al.* [466]. The Greater Wax Moth is a typical holometabolous insect and develops through four distinct life stages, i) egg, ii) larva (8-10 instars), iii) pupa and iv) adult (imago).

The duration taken to complete the life cycle varies from weeks to months and is affected by both biotic and abiotic factors [467]. Diet has been demonstrated to affect larva development [468]. Diet is also deemed to boost immunity, as nutrient deprived larvae became susceptible to *Candida albicans* [469]. Abiotic factors such as temperature and relative humidity are crucial to the entire life cycle. However, unlike other invertebrate models such as *Caenorhabditis elegans*

and *Drosophila melanogaster*, the Greater Wax Moth larvae can survive at 37 ˚C (body temperature of the *S. suis* natural host, the pig) and therefore allow the investigation of temperature-dependent microbial virulence factors [470, 471].

The last instar larvae, which develop from the egg in about five weeks, are used for experimental studies. It is at this stage that Greater Wax Moth larvae are large enough for the most common infection route, intrahemocoelic injection of test compounds through the last posterior left proleg [472]. Greater Wax Moth larvae can be stored at 15 ˚C before use, and it is recommended to starve the larvae for 24 h before infection [473]. Oral infection has also been described, but has the draw back that exact infection doses are difficult to obtain [474]; a problem that can be overcome by using a technically more challenging force-feeding method [473]. Microbial inoculums should be washed prior to infection, to minimise the introduction of virulence factors secreted during *in-vitro* growth of microorganisms. It is recommended to apply a placebo inoculum as a control for potential physical trauma due to the injection [459]. At least 10–20 larvae for each experimental condition should be used, in order to generate statistically relevant data [475].

Despite its advantages, uptake of the Greater Wax Moth model has been limited due to significant variation in the behavior of larvae both within and between batches. Greater Wax Moth larvae for experimental infection models are often purchased from unregulated pet shops, such as Livefood UK Ltd. In response to this, BioSystems Technology Ltd. has developed TruLarv; standardised research grade Greater Wax Moth larvae that are age and weight defined, surface decontaminated and inbred from a genome sequenced dedicated breeding colony without the use of antimicrobials, hormones or other drugs. As a result, TruLarv behave more consistently and reproducibly than Greater Wax Moth larvae from other sources, such as pet shops.

1.1.5 The genome

At present (March 2018), a total of 42 *S. suis* 'finished' complete genome sequences and 1,219 *S. suis* 'unfinished' draft genome sequences can be found in public databases. However, in October 2013, at the beginning of this PhD project, just 14 *S. suis* complete genome sequences had been uploaded to GenBank; sequenced at either the Beijing Institute of Genomics (China) or Wellcome Trust Sanger Institute (Cambridge, UK) [476-482]. The reason for this explosion in the amount of detailed genetic information for *S. suis* is i) rapid advances and reductions in cost allowing the whole-genome sequencing of multiple strains of the same species, ii) a surge in popularity of this approach to reassess the diagnostic, prevention, and surveillance programs of both commensal and invasive bacteria and iii) a renewed interest in *S. suis* as an emerging/re-emerging zoonotic pathogen with the introduction of limitations on the usage of antimicrobials.

Table 1.4 summarises the characteristics of the 14 *S. suis* 'finished' complete genome sequences available in October 2013, of which half are serotype 2. The *S. suis* genome typically consists of a single circular chromosome with an approximate size of 2.1 (median 2.09612) Mb in length [477]. Median GC content is 41.2%. Approximately 2,000 (median 1981) predicted protein-encoding sequences are contained within the *S. suis* genome, resulting in a genome coding density of ~88%. Comparative genomic analysis of 391 *S. suis* isolates by Weinert *et al.* identified a *S. suis* core-genome (i.e. prevalent in all isolates, as opposed the accessory-genome i.e. genes absent from one or more isolates or unique to a given isolate) consisting of 793 protein-encoding sequences [141].

Table 1.4. Characteristics of 14 *Streptococcus suis* 'finished' complete genome sequences available in GenBank, October 2013.

Isolate	Origin	Host	Serotype	Size (bp)	GC content (%)	CDSs	Coding density (%)	Genome Islands	ICE	89K	Plasmids	GenBank	Reference
05ZYH33	China	Human	2	2,096,331	41.1	2194	-	-	-	Yes	0	CP000407	Chen *et al.* [476]
98HAH33	China	Human	2	2,095,720	41.1	2191	-	-	-	Yes	0	CP000408	Chen *et al.* [476]
BM407	Vietnam	Human	2	2,146,229	41.1	2,040	83.9	4	32	-	1	FM252032	Holden *et al.* [477]
P1/7	UK	Pig	2	2,007,491	41.3	1908	85.1	4	27	No	0	AM946016	Holden *et al.* [477]
SC84	China	Human	2	2,095,898	41.1	1985	84.8	3	28	No	0	FM252031	Holden *et al.* [477]
GZ1	China	Human	2	2,038,034	41.4	1987	-	-	-	-	0	CP000837	Ye *et al.* [483]
A7	China	Pig	2	2,038,409	41.2	1974	88.7	-	25	No	0	CP002570	Hu *et al.* [480]
JS14	China	Pig	14	2,137,435	41.2	2106	88.4	17	26	-	0	CP002465	Hu *et al.* [479]
ST3	China	Pig	3	2,028,815	41.3	2031	87.8	-	28	-	0	CP002633	Hu *et al.* [478]
D9	China	-	7	2,177,656	41.0	2136	88.3	16	27	-	0	CP002641	Zhang *et al.* [481]
D12	China	-	9	2,183,059	41.3	2124	88.1	23	27	-	0	CP002644	Zhang *et al.* [481]
SS12	China	-	1/2	2,096,866	41.2	2091	88.4	16	27	-	0	CP002640	Zhang *et al.* [481]
ST1	China	-	1	2,034,321	41.4	2030	87.7	6	23	-	0	CP002651	Zhang *et al.* [481]
TL13	China	Pig	16	2,038,146	41.3	1,950	97.2	-	-	No	0	CP003993	Wang *et al.* [482]

Comparative genomic analysis, based on 16S rRNA and *rnpB* gene sequences, has shown *S. suis* is phylogenetically distinct from the other *Streptococcus* species for which genome sequences are available (Figure 1.2) [477]. Accordingly, ~40% of the *S. suis* genome is unique in comparison to other *Streptococcus* species [477]. Importantly, it should be noted that *S. suis* phylogenies should be treated with caution as the generation of a reliable phylogenetic tree of the *S. suis* core-genome is impossible given the evidence of widespread recombination [141].

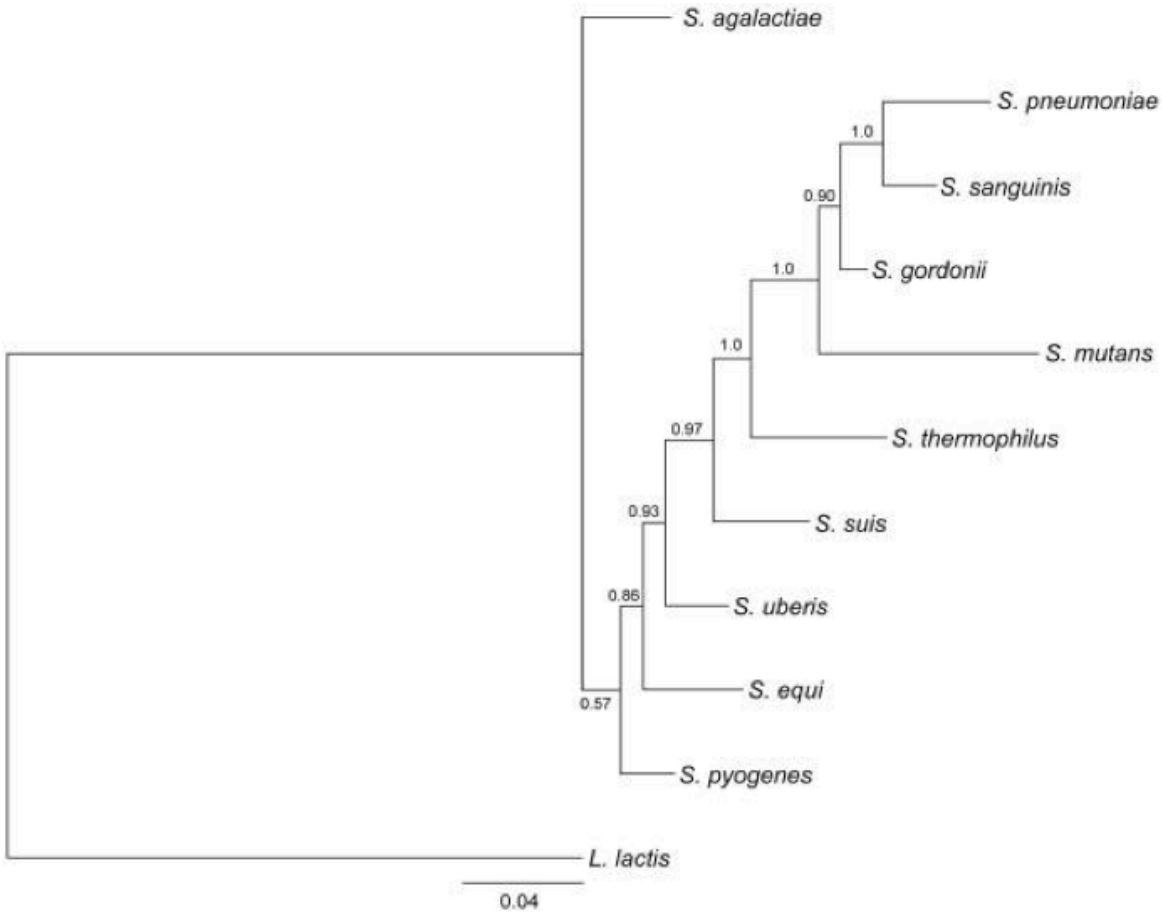

Figure 1.2. Phylogenic tree of the relationships of *Streptococcus suis* and the other streptococci. Unrooted majority-rule tree built from 16S rRNA and *rnpB* gene sequences taken from Holden *et al.* [477]. The other genome sequence *Streptococcus* species include: (*S. suis* P1/7) *S. agalactiae* NEM316, *S. equi* 4047, *S. gordonii* Challis CH1, *S. mutans* UA159, *S. pneumoniae* TIGR4, *S. pyogenes* Manfredo, *S. sanguinis* SK36, *S. thermophilus* CNRZ1066 and *S. uberis* 0140J. *Lactococcus lactis* subsp. *lactis* IL1403 was included as an outgroup. The numbers at the branches are posterior probabilities indicating the support for the branch. The bar indicates the number of substitutions per site (4 per 100 sites).

In generating complete genome sequences for *S. suis* strains BM407, P1/7 and SC84 Holden *et al.* identified four regions (designated genomic islands) in the P1/7 genome that had properties indicative of recent acquisition, which included one prophage-like region and three miscellaneous islands [477]. None of these genomic islands carried characterised or putative virulence factors. Equivalent genomic islands were also observed in the SC84 genome, however, BM407 contained only three of the genomic islands present in P1/7 [477]. In contrast to other streptococcal species, such as S. pyogenes [484], Holden *et al.* observed that the proportion of mobile genetic elements in the *S. suis* genome is relatively small (1.8% of P1/7) suggesting that the pan-genome (the entire gene set of all strains of a species) is smaller than other streptococci [485, 486].

Integrative conjugative elements (ICE) are commonly found in streptococcal genomes. Tn916-type and Tn5252-type elements are most common. Analogous with the bacterial chromosome ICE have core and accessory components to their gene complement. The conjugative machinery appears to be relatively well conserved between the ICE, with all of the streptococcal Tn5252-like elements sharing a large recombination protein, as well as VirB4-type and VirD4-type proteins. Furthermore, a genus-wide comparison of streptococcal ICE reveals a surprising level of similarity in their complement of cargo genes. The pezAT addiction toxin system, first identified in the chromosome of *Streptococcus pneumoniae*, is found on elements present in *Streptococcus agalactiae*, *Streptococcus pneumoniae* and *S. suis*, presumably aiding fixation of the transposon following integration. Bacteriocins, either alone or as part of a gene cluster encoding the accompanying processing machinery, are present on all the sequenced *Streptococcus pneumoniae* [487] and *S. suis* elements. Of greater clinical importance, sequenced streptococcal ICE can carry genes for resistance to tetracycline, erythromycin, chloramphenicol, trimethoprim, aminoglycosides and streptomycin.

One ICE, an additional 89 Kb DNA fragment designated "89K", has been reported in highly virulent *S. suis* isolates [476]. 89K was originally identified in *S. suis* isolates 98HAH12 and

05ZYH33, using the combination of a genome-wide display of GC content and co-linearity comparison. Further investigation showed it was possible to detect 89K in 23 highly virulent Chinese isolates recovered during the 1998 Jiangsu and 2005 Sichuan outbreaks, characterised by the rapid onset of disease affecting a large number of individuals and distinctive clinical manifestations (meningitis, sepsis or STSS; see sections 1.1.2.2.1 and 1.1.2.2.2). The fact that 89K could not be positively detected in avirulent Chinese isolates, such as 05HAS68, led to the hypothesis that ICE 89K was a specific pathogenicity island candidate in Chinese highly virulent *S. suis* serotype 2 strains. 89K encodes ~70 potential protein-encoding sequences, including three ABC transporter cassettes linked to important roles in controlling efflux and influx across cell membranes, two Two-component systems linked to quorum sensing to initiate the infection cycle through regulating efficient bacterial density and adapting to different environments and three different components of Type IV secretion systems. In fact, the major features of typical pathogenicity islands are all present in 89K. However, despite 89K being present in Chinese highly virulent isolates and absent in Chinese avirulent isolates, indicating its potential relationship to high pathogenicity, 89K could not be positively detected in any Canadian or European isolates tested (including, but not limited to P1/7, S10, S735, SS2-N, T15) regardless of virulence potential [476].

Plasmids have been reported to be present in three of the 1,261 *S. suis* isolates available in GenBank. Table 1.5 summarises the characteristics of the four *S. suis* plasmids pBM407, pNSU1060a, pNSU1060b and pSSU1; of which pNSU1060a and pNSU1060b were recovered from the same isolate, NSU1060 (serotype 2, ST25). Plasmids reported in *S. suis* range in size from 4,975 bp to 24,579 bp, and have been reported to harbour up to 18 predicted protein-encoding sequences (pBM407). While chromosomes are large and encode genetic information essential for living under 'normal' conditions, in nature plasmids are typically small and harbour additional genes that may benefit the survival of an organism in certain conditions; such as, but not limited to antibiotic resistance. Conjugation is the process streptococci most commonly use to mediate transfer of plasmids, and was first described by Jacob and Hobbs; who noted conjugal

transfer of multiple antibiotic resistance by a plasmid [488]. However, attempts to use plasmid DNA from *S. suis* resistant to erythromycin, clindamycin and tetracycline to transform antibiotic sensitive recipients were unsuccessful [489]. No known antibiotic resistance genes were reported to be encoded on the plasmids listed in Table 1.5, and other mechanisms exist for the exchange of genetic material in *S. suis*, for example pheromone induced natural competence [490-492].

Table 1.5. Plasmids recovered from *Streptococcus suis* isolates available in GenBank, March 2018.

Plasmid	Size (bp)	GC content (%)	CDSs	Coding density (%)	GenBank	Reference
pBM407	24,579	36.4	18	-	FM252033	Holden *et al.* [477]
pNSUI060a	11,393	40.7	16	-	CP012912	Athey *et al.* [493]
pNSUI060b	5,581	37.7	6	-	CP012913	Athey *et al.* [493]
pSSU1	4,975	36.9	6	75.6	AB019522	Takamatsu *et al.* [494]

A number of prophages have been identified in the whole-genome sequences of *S. suis* isolates [479, 495-497]. A prophage is a bacteriophage genome inserted and integrated into the bacterial chromosome or existing as a plasmid. This is a latent form of a phage, in which the viral genes are present in the bacterium without causing disruption of the bacterial cell. However, due to their highly specific host recognition, bacteriophages have potential as therapeutic agents in the treatment of infections [498, 499]. Phage can be classified by their lifestyle, as either lytic phage or temperate phage [500]. Temperate phage, such as *S. suis* phage Ss1 [495] or ϕNJ2 (GenBank accession number: JX879087) [497], have been the subject of attention because they play an important role in bacterial pathogenesis and the dissemination of genes encoding virulence factors by horizontal transfer of DNA [501-503], whereas lytic phage provide potential tools for classifying bacterial strains and combating the antibiotic resistance crisis [504]. To date, phage SMP is the only reported lytic phage of *S. suis*, isolated from *S. suis* serotype 2 strains recovered from nasal swabs of healthy Bama minipigs [505]. Phage SMP was reported to have an isometric head of 50 nm, a noncontractile tail of ~135 nm and a linear double-stranded DNA genome. The host range of phage SMP was limited to two of 24 *S. suis* serotype 2 strains tested, and determining the genome sequence was an approach suggested by the authors that could offer clues to understanding the evolution and heredity of the host [505].

1.2 Thesis objectives

Even before the explosion in the amount of detailed genetic information for *S. suis* molecular typing approaches had shown the *S. suis* species to be genetically very heterogeneous. A feature that should be the focus of and exploited by future programs to control *S. suis* disease. Population genetic studies have described at least 121 sequence types for *S. suis*, but neither the authors or follow on studies have been able to observe a strict relationship between sequence type and disease (or serotype) in pigs or humans [90]. One of the drawbacks of population genetic studies such as MLST is that they can lack discriminatory power due to high levels of sequence conservation (and as a result slow accumulation of genetic variation) in the target housekeeping genes, resulting in an inability to differentiate bacterial strains. In addition to this, MLST struggles to monitor virulence factors, genes typically associated with high levels of recombination and mobility in comparison to the population genetic framework. In terms of diagnostics, it is more important to identify virulence factors than have a population genetics-based evaluation of prevalent strains.

Recent genome analyses have identified the presence of "89K", a potential pathogenicity island thought to be responsible for antibiotic resistance in *S. suis* isolates causing human disease in China [477]. However, despite 89K being present in Chinese highly virulent isolates and absent in Chinese avirulent isolates, indicating its potential relationship to high pathogenicity, 89K could not be positively detected in Canadian or European isolates tested regardless of virulence potential [476]. Other analyses have identified SNPs resulting from point mutations and recombination, and have been proposed to be responsible for subtle effects on the expression and function of virulence genes [115, 477]. Importantly whole-genome sequencing and the detailed annotation of *S. suis* genomes is providing the valuable data for studying evolutionary events that are shaping the virulence and drug resistance of this pathogen.

Indeed, bioinformatics analyses in our group have revealed significant genomic differences between non-clinical isolates of *S. suis* recovered from the upper respiratory tract of pigs and

isolates causing systemic disease [141]. In doing so post-doctoral research associate Dr Lucy Weinert and colleagues showed *S. suis* isolates causing systemic disease had substantially fewer genes (on average) than non-clinical isolates from the upper respiratory tract, inferring that the loss of protein-encoding sequences had led to a smaller genome with increased virulence potential; a phenomenon that has also been reported in other bacteria, such as *Salmonella enterica*, *Shigella flexneri* and *Yersinia pestis* [506-509]. Curiously, despite this genomic reduction, an overrepresentation of genes encoding previously reported "virulence-factors" associated with *S. suis* was also found in isolates causing systemic disease in pigs [141]. Crucially, Dr Lucy Weinert and colleagues realised the potential of using whole-genome sequencing data for *S. suis* to identify novel virulence markers, leading to the inception of this PhD project.

The research described in this thesis focuses on the design and evaluation of innovative methods for the detection and control of invasive disease-associated *S. suis* strains in intensive pig production systems, a need echoed in the literature "given the cost of eliminating *S. suis*, the risk of failure, the difficulty in maintaining a free herd, and the lack of tools to monitor herd status, it would appear reasonable to direct resources toward control measures rather than eradication" [11]. The main objective of this research was to identify novel associations between genotype and phenotype by combing whole-genome sequencing data with carefully defined clinical metadata. These genetic markers would then be used to predict the invasive disease causing potential of isolates being carried asymptomatically in pig populations on farms in England and Wales. With access to the whole-genome sequencing data of an initial (training) collection of 115 *S. suis* isolates and corresponding detailed clinical metadata I set out to achieve the following objectives:

1. To identify significant associations between genotype and phenotype by combining whole-genome sequencing data with carefully defined clinical metadata – in doing so identifying regions of the genome associated with invasive disease caused by *S. suis*
2. To design and evaluate a pathotyping tool to predict the potential to cause invasive disease of *S. suis* isolates recovered from intensive pig production systems in England and Wales
3. To use the newly-defined "pathotyping markers" to estimate the bacterial load of disease-associated and non-disease associated strains of *S. suis* on the palatine tonsils of pigs - in doing so drawing comparisons against current culture-based approaches for surveillance
4. To design and evaluate an isogeneic knockout mutant to investigate the function and biological role in pathogenesis of one the invasive disease-associated "pathotyping markers", SSU1589 (*virA*)

Chapter 2

Materials and methods

2.1 Materials

2.1.1 Apparatus

AC88 Balance	Mettler, Zurich, Switzerland
Airstream Class II Biological Safety Cabinet	Esco Micro Pte. Ltd., Singapore
Bioscreen C	OY Growth Curves, Finland
Centrifuge 5415C, 5804	Eppendorf, Stevenage, UK
Dri-Block DB-2D	Techne Ltd., Cambridge, UK
Falcon 30 autoclave	LTE Scientific Ltd., Oldham, UK
Gel Doc XR+ Gel Documentation System	Bio-Rad Laboratories, Inc., Hertfordshire, UK
Gel electrophoresis equipment	Bio-Rad Laboratories, Inc., Hertfordshire, UK
Heraeus Megafuge 16R Centrifuge	Thermo Fisher Scientific, Paisley, UK
Heraeus Pico17 Centrifuge	Thermo Fisher Scientific, Paisley, UK
Innova43 Incubator shaker	New Brunswick Scientific Ltd, Maldon, UK
Gilson Pipette Pipetman, adjustable	Gilson Scientific Ltd., Dunstable, UK
Power Pac 300	Bio-Rad Laboratories, Inc., Hertfordshire, UK
Qubit 2.0 Fluorometer	Invitrogen, Paisley, UK
Rotor-Gene Q PCR cycler	Qiagen, Manchester, UK
SpectraMaxM5 Microplate Reader	Molecular Devices LLC, California, USA
SUB Aqua 12 Plus Water bath	Grant Instruments, Cambridge, UK
T100 Thermal Cycler	Bio-Rad Laboratories Inc., Hertfordshire, UK
Tridak STEPPER	Dymax Corporation, Connecticut, USA
Ultra Low Temp -85 ˚C Freezer U57085	New Brunswick Scientific Ltd, Maldon, UK
UV1101 Biotech Photometer	WPA, Cambridge, UK
UV Transilluminator	UVP, Upland, USA
Vortex-Genie 2	Scientific Industries Inc., New York, USA

2.1.2 Buffers and growth media

Chloroform-isoamyl alcohol mixture	Sigma-Aldrich, Dorset, UK
Columbia Agar Base with Horse Blood plates	Oxoid Ltd, Basingstoke, UK
Columbia Agar Base with Sheep Blood plates	Oxoid Ltd, Basingstoke, UK
Crystal violet, 1%, aqueous solution	Sigma-Aldrich, Dorset, UK
DNA Loading Buffer, Blue	Bioline, London, UK
Glycerol	Fisher Scientific, Loughborough, UK
Nuclease-Free water	Promega, Southampton, UK
Phenol: Chloroform:Isoamyl Alcohol 25:24:1	Sigma-Aldrich, Dorset, UK
Phosphate buffered saline (PBS; pH 7.2)	Sigma-Aldrich, Dorset, UK
Sodium acetate buffer solution, pH 5.2	Sigma-Aldrich, Dorset, UK
TBE buffer (10x)	Promega Corporation, Madison, USA
TE buffer	Invitrogen, Paisley, UK
Tris-Cl (pH 8.5)	Qiagen, Manchester, UK
Todd-Hewitt broth	Oxoid Ltd, Basingstoke, UK
Yeast Extract	Fisher Scientific, Loughborough, UK

2.1.3 Consumables

Centrifuge Tubes; 15 mL, 50 mL	Corning Inc., Wiesbaden, Germany
Cryogenic vial, ext. thread (2.0 mL)	Corning Inc., Wiesbaden, Germany
Disposable Scalpel blades	Swann Morton, Sheffield, UK
Disposable Serological Pipette	Thermo Fisher Scientific, Paisley, UK
Dry swab, Fine tip plastic	Medical Wire & Equipment, Corsham, UK
Dry swab, Standard plastic	Medical Wire & Equipment, Corsham, UK
Filter tips, beveled; 20, 200, 300, 1000 µL	Star Lab, Milton Keynes, UK
Flat-bottom 96-well tissue culture plates	Greiner Bio-One Ltd. Gloucestershire, UK
Inoculation Loops (1 µL)	Copan Diagnostics Inc., Murrieta, USA
Microlance hypodermic needles	Becton, Dickinson and Company Ltd., Ireland
Multiplate PCR Plates 96-well, clear	Bio-Rad Laboratories, Inc., Hertfordshire, UK
PCR Tube Strips Domed 8-cap strips	Bio-Rad Laboratories, Inc., Hertfordshire, UK
Petri dishes (90 mm)	Thermo Fisher Scientific, Paisley, UK
Skirted 96-well PCR Plate	Thermo Fisher Scientific, Paisley, UK
Strip Tubes and Caps (0.1 mL)	Qiagen, Manchester, UK
Syringe driven filter unit (0.22 µm)	Millex (Merck), Darmstadt, Germany
Transwab, Standard plastic	Medical Wire & Equipment, Corsham, UK

2.1.4 Kits

API® 20 STREP	Biomerieux, Inc., Basingstoke UK
API® RAPID ID 32 STREP	Biomerieux, Inc., Basingstoke UK
GenElute Bacterial Genomic DNA Kit	Sigma-Aldrich, Haverhill, UK
Genomic-tip 100/G Columns	Qiagen, Manchester, UK
Genomic DNA Buffer Set	Qiagen, Manchester, UK
GenomiPhi V2 DNA Amplification Kit	GE Healthcare, Little Chalfont, UK
KOD Hot Start DNA Polymerase Kit	Novagen (Merck), Darmstadt, Germany
MasterPure Gram Positive DNA Purification	Epicentre, Madison, USA
Multiplex PCR *Plus* Kit	Qiagen, Manchester, UK
Qubit dsDNA BR Assay Kit	Invitrogen, Paisley, UK
SensiFAST SYBR No-ROX Kit	Bioline, London, UK
Wizard SV Gel and PCR Clean-Up System	Promega Corporation, Madison, USA

2.1.5 Bacterial isolates and strains

Two groups of *S. suis* isolates were used in this study i) a training collection of 115 isolates and ii) an out-of-sample test collection of 138 previously uncharacterised isolates. **The original training collection** was used to identify genetic markers associated with i) invasive disease or ii) asymptomatic carriage on the palatine tonsils of pigs, with an aim to in the future differentiate *S. suis* isolates into i) disease-associated and ii) non-disease associated phenotypic groups. The 'training' collection consisted of laboratory reference strain P1/7 (NC_012925) originally recovered from an ante-*mortem* blood culture from a pig that died with meningitis in the United Kingdom [477, 510]. The other 114 isolates of the training collection were a subset recovered from pigs on farms in England and Wales during routine diagnostic investigations of post-*mortem* material at the Animal Health and Veterinary Laboratories Agency (AHVLA; now the Animal and Plant Health Agency, APHA) in 2010, and contribute to a larger collection (denoted: the "LoLa collection") previously described in 2015 by Weinert *et al.* [141]. Well-defined phenotypic metadata were available allowing each isolate to be categorised as being associated with invasive *S. suis* disease (n=53; recovered from systemic sites in the presence of clinical signs (arthritis, meningitis, septicaemia) and/or gross pathology consistent with *S. suis* infection) or as being non-disease associated (n=62; recovered from the tonsil or trachea-bronchus of pigs without any typical signs of infection, but diagnosed with disease unrelated to *S. suis*, such as enteric disease). **The out-of-sample test collection** was used to evaluate the molecular pathotyping tool. Out-of-sample forecasting is a common approach used to evaluate the performance of binary diagnostic tests. To avoid reducing statistical power, rather than split the original training collection, an additional out-of-sample 'test' collection was put together consisting of 69 invasive disease-associated (recovered from systemic, non-respiratory locations of pigs diagnosed with *S. suis* disease at the APHA during 2013) and 69 non-disease associated isolates (recovered from tonsil swabs of pigs on farms in England and Wales exhibiting no signs of *S. suis* disease between June 2013 and May 2014). Table 2.1 summarises the site of recovery, ante- and post-*mortem* findings of all isolates used to produce the data described in this thesis.

Table 2.1 *Streptococcus suis* isolates used to produce the data described in this thesis.

Summary of the i) site of recovery, ii) ante- and post-*mortem* findings and iii) categorisation (disease-/non-disease associated) of all 253 *S. suis* isolates used to produce the data described in this thesis.

Isolate	Collection	Ante	Post	Origin	Phenotype
P1/7	Training	-	Meningitis	Blood	Disease-associated
SS002	Training	Nervous	Meningitis	Brain	Disease-associated
SS004	Training	Found dead	Meningitis	Meninges	Disease-associated
SS005	Training	Swollen joints	Arthritis	Joint	Disease-associated
SS006	Training	Found dead	Arthritis	Brain	Disease-associated
SS007	Training	Nervous	Septicaemia	Brain	Disease-associated
SS008	Training	Found dead	Pericarditis	Pericardium	Disease-associated
SS010	Training	Nervous	Meningitis	Brain	Disease-associated
SS012	Training	Lameness	Arthritis	Joint	Disease-associated
SS013	Training	Lameness	Arthritis	Left carpus	Disease-associated
SS015	Training	Found dead	Aortic stenosis	Brain	Disease-associated
SS016	Training	Lameness	Arthritis	Joint	Disease-associated
SS017	Training	-	-	Joint	Disease-associated
SS018	Training	Ill thrift	Pneumonia	Pericardium	Disease-associated
SS021	Training	Found dead	Septicaemia	Joint	Disease-associated
SS022	Training	Lameness	Arthritis	Brain	Disease-associated
SS024	Training	Ill thrift	Arthritis	Brain	Disease-associated
SS025	Training	-	-	Brain	Disease-associated
SS028	Training	Found dead	Meningitis	Brain	Disease-associated
SS029	Training	Nervous	Meningitis	Joint	Disease-associated
SS035	Training	Found dead	Septicaemia	Joint	Disease-associated
SS036	Training	Nervous	Septicaemia	Brain	Disease-associated
SS038	Training	-	-	Liver	Disease-associated
SS041	Training	Found dead	Polyserositis	Brain	Disease-associated
SS042	Training	Nervous	Meningitis	Brain	Disease-associated
SS044	Training	Found dead	Arthritis	Joint	Disease-associated
SS045	Training	Found dead	Meningitis	Meninges	Disease-associated
SS046	Training	Found dead	Meningitis	Brain	Disease-associated
SS053	Training	Found dead	Haemorrhagic enteropathy	Brain	Disease-associated
SS057	Training	Nervous	Meningitis	Liver	Disease-associated
SS058	Training	Nervous	Meningitis	Brain	Disease-associated
SS060	Training	Nervous	Meningitis	Brain	Disease-associated
SS062	Training	Found dead	Polyserositis	Brain	Disease-associated
SS063	Training	Found dead	Arthritis	Joint	Disease-associated
SS065	Training	Found dead	Septicaemia	Lymph node	Disease-associated
SS066	Training	Ill thrift	Polyserositis	Brain	Disease-associated
SS068	Training	Ill thrift	Pneumonia	Brain	Disease-associated
SS069	Training	Found dead	Pneumonia	Liver	Disease-associated
SS071	Training	Nervous	Pneumonia	Brain	Disease-associated
SS072	Training	Nervous	Arthritis	Joint	Disease-associated

SS077	Training	Lameness	Meningitis	Brain	Disease-associated
SS083	Training	Ill thrift	Pneumonia	Meninges	Disease-associated
SS085	Training	Ill thrift	Pneumonia	Brain	Disease-associated
SS086	Training	-	-	Peritoneum	Disease-associated
SS087	Training	Nervous	Meningitis	Brain	Disease-associated
SS088	Training	Nervous	Pneumonia	Joint	Disease-associated
SS093	Training	Found dead	Meningitis	Brain	Disease-associated
SS095	Training	Found dead	Meningitis	Brain	Disease-associated
SS096	Training	Found dead	Meningitis	Brain	Disease-associated
SS097	Training	Found dead	Anaemia	Spleen	Disease-associated
SS098	Training	Found dead	Meningitis	Brain	Disease-associated
SS099	Training	Found dead	Meningitis	Brain	Disease-associated
SS100	Training	-	-	Brain	Disease-associated
LSS001	Training	Ill thrift	Pneumonia	Tracheo-bronchus	Non-disease associated
LSS003	Training	Found dead	Pneumonia	Tracheo-bronchus	Non-disease associated
LSS009	Training	Ill thrift	-	Palatine tonsils	Non-disease associated
LSS011	Training	Nervous	Water deprivation	Palatine tonsils	Non-disease associated
LSS018	Training	Ill thrift	Enterocolitis	Tracheo-bronchus	Non-disease associated
LSS020	Training	Ill thrift	Enterocolitis	Tracheo-bronchus	Non-disease associated
LSS023	Training	Found dead	Septicaemia	Tracheo-bronchus	Non-disease associated
LSS025	Training	Found dead	PCVAD	Tracheo-bronchus	Non-disease associated
LSS027	Training	Lameness	Anaemia	Palatine tonsils	Non-disease associated
LSS030	Training	Found dead	Septicaemia	Palatine tonsils	Non-disease associated
LSS031	Training	Found dead	Gut torsion	Palatine tonsils	Non-disease associated
LSS032	Training	Found dead	Gut torsion	Palatine tonsils	Non-disease associated
LSS034	Training	-	-	Nasal	Non-disease associated
LSS037	Training	Found dead	Peritonitis	Tracheo-bronchus	Non-disease associated
LSS038	Training	Found dead	Peritonitis	Palatine tonsils	Non-disease associated
LSS039	Training	Found dead	Peritonitis	Tracheo-bronchus	Non-disease associated
LSS040	Training	Ill thrift	Arthritis	Tracheo-bronchus	Non-disease associated
LSS041	Training	Ill thrift	Arthritis	Tracheo-bronchus	Non-disease associated
LSS042	Training	Ill thrift	Arthritis	Palatine tonsils	Non-disease associated
LSS044	Training	Ill thrift	Arthritis	Palatine tonsils	Non-disease associated
LSS047	Training	Found dead	Septicaemia	Palatine tonsils	Non-disease associated
LSS048	Training	Found dead	Septicaemia	Tracheo-bronchus	Non-disease associated
LSS049	Training	Found dead	Enterocolitis	Tracheo-bronchus	Non-disease associated
LSS053	Training	Nervous	Neuropathy	Tracheo-bronchus	Non-disease associated
LSS054	Training	Ill thrift	Navel illl	Palatine tonsils	Non-disease associated
LSS055	Training	Ill thrift	Navel illl	Palatine tonsils	Non-disease associated
LSS057	Training	Found dead	PCVAD	Tracheo-bronchus	Non-disease associated
LSS058	Training	Nervous	Enterocolitis	Tracheo-bronchus	Non-disease associated
LSS059	Training	Nervous	Enterocolitis	Tracheo-bronchus	Non-disease associated
LSS060	Training	Nervous	Enterocolitis	Palatine tonsils	Non-disease associated
LSS061	Training	Nervous	Enterocolitis	Palatine tonsils	Non-disease associated
LSS062	Training	Nervous	Enterocolitis	Tracheo-bronchus	Non-disease associated
LSS063	Training	Nervous	Enterocolitis	Tracheo-bronchus	Non-disease associated
LSS064	Training	Ill thrift	Pneumonia	Palatine tonsils	Non-disease associated
LSS065	Training	Found dead	Septicaemia	Palatine tonsils	Non-disease associated

LSS067	Training	Found dead	Pneumonia	Palatine tonsils	Non-disease associated
LSS068	Training	Found dead	Pneumonia	Tracheo-bronchus	Non-disease associated
LSS069	Training	Found dead	Pneumonia	Tracheo-bronchus	Non-disease associated
LSS070	Training	Lameness	Fracture	Tracheo-bronchus	Non-disease associated
LSS071	Training	Ill thrift	PCVAD	Tracheo-bronchus	Non-disease associated
LSS072	Training	Ill thrift	Pneumonia	Tracheo-bronchus	Non-disease associated
LSS075	Training	Ill thrift	Pneumonia	Palatine tonsils	Non-disease associated
LSS076	Training	-	-	Palatine tonsils	Non-disease associated
LSS077	Training	Nervous	Neuropathy	Tracheo-bronchus	Non-disease associated
LSS078	Training	Found dead	Gut torsion	Tracheo-bronchus	Non-disease associated
LSS079	Training	Found dead	Gut torsion	Palatine tonsils	Non-disease associated
LSS080	Training	Found dead	Gut torsion	Tracheo-bronchus	Non-disease associated
LSS081	Training	Found dead	Polyserositis	Palatine tonsils	Non-disease associated
LSS082	Training	Ill thrift	Arthritis	Tracheo-bronchus	Non-disease associated
LSS083	Training	Respiratory	Pneumonia	Palatine tonsils	Non-disease associated
LSS084	Training	Respiratory	Pneumonia	Tracheo-bronchus	Non-disease associated
LSS085	Training	Respiratory	Pneumonia	Palatine tonsils	Non-disease associated
LSS086	Training	Respiratory	Pneumonia	Tracheo-bronchus	Non-disease associated
LSS088	Training	Ill thrift	Lymphadenitis	Palatine tonsils	Non-disease associated
LSS089	Training	Respiratory	Pneumonia	Palatine tonsils	Non-disease associated
LSS090	Training	Respiratory	Pneumonia	Palatine tonsils	Non-disease associated
LSS091	Training	Diarrhoea	Colibacillosis	Palatine tonsils	Non-disease associated
LSS092	Training	Diarrhoea	Colibacillosis	Palatine tonsils	Non-disease associated
LSS094	Training	Diarrhoea	Colitis	Tracheo-bronchus	Non-disease associated
LSS095	Training	Diarrhoea	Enterocolitis	Palatine tonsils	Non-disease associated
LSS096	Training	Diarrhoea	Enterocolitis	Palatine tonsils	Non-disease associated
SS027	Training	Found dead	Enteropathy	Trachea	Non-disease associated
tmw001	Out-of-sample	-	-	-	Disease-associated
tmw002	Out-of-sample	-	-	-	Disease-associated
tmw003	Out-of-sample	-	-	-	Disease-associated
tmw004	Out-of-sample	-	-	-	Disease-associated
tmw005	Out-of-sample	Found dead	-	-	Disease-associated
tmw006	Out-of-sample	Lameness	Endocarditis	Joint	Disease-associated
tmw007	Out-of-sample	Nervous	Meningitis	Brain	Disease-associated
tmw008	Out-of-sample	-	-	-	Disease-associated
tmw010	Out-of-sample	Respiratory	-	-	Disease-associated
tmw011	Out-of-sample	Found dead	Endocarditis	-	Disease-associated
tmw012	Out-of-sample	-	-	-	Disease-associated
tmw013	Out-of-sample	Ill thrift	Pericarditis	Pericardium	Disease-associated
tmw014	Out-of-sample	-	-	-	Disease-associated
tmw015	Out-of-sample	Nervous	-	-	Disease-associated
tmw016	Out-of-sample	Nervous	-	-	Disease-associated
tmw017	Out-of-sample	Lameness	Meningitis	-	Disease-associated
tmw020	Out-of-sample	-	-	-	Disease-associated
tmw021	Out-of-sample	Malaise	Pericarditis	Pericardium	Disease-associated
tmw022	Out-of-sample	Found dead	-	Brain	Disease-associated
tmw023	Out-of-sample	Found dead	Septicaemia	Liver	Disease-associated
tmw025	Out-of-sample	-	-	-	Disease-associated

tmw029	Out-of-sample	Abortion	-	Foetal stomach content	Disease-associated
tmw030	Out-of-sample	Lameness	Arthritis	Joint	Disease-associated
tmw031	Out-of-sample	-	-	-	Disease-associated
tmw034	Out-of-sample	Nervous	Meningitis	-	Disease-associated
tmw035	Out-of-sample	-	-	-	Disease-associated
tmw036	Out-of-sample	-	-	-	Disease-associated
tmw037	Out-of-sample	-	-	-	Disease-associated
tmw038	Out-of-sample	Found dead	Meningitis	Brain	Disease-associated
tmw040	Out-of-sample	Found dead	Meningitis	Brain	Disease-associated
tmw047	Out-of-sample	-	-	-	Disease-associated
tmw049	Out-of-sample	-	-	-	Disease-associated
tmw051	Out-of-sample	Found dead	Pneumonia	Brain	Disease-associated
tmw052	Out-of-sample	-	-	-	Disease-associated
tmw053	Out-of-sample	-	-	-	Disease-associated
tmw055	Out-of-sample	Found dead	Pericarditis	-	Disease-associated
tmw056	Out-of-sample	Found dead	-	-	Disease-associated
tmw057	Out-of-sample	-	-	-	Disease-associated
tmw058	Out-of-sample	-	-	-	Disease-associated
tmw059	Out-of-sample	Nervous	Meningitis	Brain	Disease-associated
tmw060	Out-of-sample	Nervous	Meningitis	Brain	Disease-associated
tmw061	Out-of-sample	Found dead	Arthritis	Joint	Disease-associated
tmw062	Out-of-sample	-	-	-	Disease-associated
tmw065	Out-of-sample	-	-		Disease-associated
tmw066	Out-of-sample	Found dead	-	-	Disease-associated
tmw069	Out-of-sample	Nervous	Meningitis	-	Disease-associated
tmw071	Out-of-sample	Malaise	Peritonitis	Brain	Disease-associated
tmw072	Out-of-sample	-	-	-	Disease-associated
tmw075	Out-of-sample	Diarrhoea	-	Navel	Disease-associated
tmw077	Out-of-sample	Lameness	Arthritis	-	Disease-associated
tmw078	Out-of-sample	Malaise	Meningitis	Kidney	Disease-associated
tmw080	Out-of-sample	Respiratory	Pneumonia	Spleen	Disease-associated
tmw081	Out-of-sample	Found dead	Meningitis	Brain	Disease-associated
tmw083	Out-of-sample	Nervous	Meningitis	Brain	Disease-associated
tmw084	Out-of-sample	-	-	-	Disease-associated
tmw086	Out-of-sample	Nervous	Meningitis	Brain	Disease-associated
tmw089	Out-of-sample	Lameness	Endocarditis	-	Disease-associated
tmw090	Out-of-sample	Nervous	Arthritis	Brain	Disease-associated
tmw092	Out-of-sample	Nervous	Meningitis	-	Disease-associated
tmw093	Out-of-sample	Wasted recumbent	Septicaemia	Liver	Disease-associated
tmw094	Out-of-sample	-	-	-	Disease-associated
tmw095	Out-of-sample	-	-	-	Disease-associated
tmw096	Out-of-sample	Found dead	Septicaemia	Brain	Disease-associated
tmw097	Out-of-sample	-	-	-	Disease-associated
tmw098	Out-of-sample	-	-	-	Disease-associated
tmw099	Out-of-sample	Lameness	Arthritis	Pericardium	Disease-associated
tmw100	Out-of-sample	Respiratory	Pneumonia	-	Disease-associated

tmw101	Out-of-sample	-	Pneumonia	-	Disease-associated
tmw102	Out-of-sample	Found dead	-	-	Disease-associated
083/01C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
083/02A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
083/02B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
083/03A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
083/13A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
083/17A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
083/17B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
083/17C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
083/22A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
083/24C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
083/25C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
125/01A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
270/06A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
270/16A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
207/21B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
270/21C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
270/23C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
359/06B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
427/05B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
427/07C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
427/17B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
427/17C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
632/04A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
632/07B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
632/07C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
632/09A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
632/09C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
635/03A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
635/06B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
635/06C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
635/08C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
635/09A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
635/09C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
635/10C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/01B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/02A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/03A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/03B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/03C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/04A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/05A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/05B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/06A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/06B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/06C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/07C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated

684/08A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/09B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/10A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/11B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/11C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/12A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/12B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/12C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/13A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/13C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/14B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/14C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/15B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/16A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/17C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/18B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/18C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/19B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/20A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/20C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/21A	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/21B	Out-of-sample	-	-	Palatine tonsils	Non-disease associated
684/21C	Out-of-sample	-	-	Palatine tonsils	Non-disease associated

2.1.6 Whole-genome assemblies from next generation sequencing data

The research described in this thesis follows on from previous work carried out by post-doctoral Research Associate Dr Lucy A. Weinert at the University of Cambridge Department of Veterinary Medicine, as part of the BBSRC Longer and Larger (LoLa) research grant BB/G019274/1. The original LoLa collection was put together investigate the evolution of *S. suis* and address two specific questions i) the evolution of *S. suis* disease in swine and ii) the evolution of *S. suis* disease in humans, and resulted in the Nature Communications publication "Genomic signatures of human and animal disease in the zoonotic pathogen *Streptococcus suis*" [141]. For completeness this section briefly describe the WGS and initial processing of the raw sequence data.

Paired-end **whole-genome sequencing** was performed on Illumina HiSeq 2000 instruments according to the manufacturer's instructions at The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK. Single colonies were picked from 90 mm Columbia agar base with sheep blood plates (Oxoid Ltd.) incubated at 37 ˚C overnight, and used to inoculate Todd-Hewitt broth (Oxoid) supplemented with 0.2% Yeast Extract (Fisher Scientific) then incubated at 37 ˚C overnight without agitation. All genomic and plasmid DNA was extracted from overnight cultures using the MasterPure Gram Positive DNA Purification Kit (epicentre) according to the manufacturer's instructions. ***De novo* assemblies** were generated using Velvet [511] version 1.1 with parameter selection taken from VelvetOptimiser and the quality of draft genome assemblies checked by mapping the raw reads back to the assemblies. Annotation was performed using Prokka. The ***S. suis* core-genome** (ie. genes present in all isolates) was identified by extracting all annotated protein-encoding sequences from the published genome of *S. suis* strain P1/7 (GenBank accession: AM946016) [477, 510], and using them as BLASTp queries against a bespoke BLAST database of all *de novo* assemblies and known complete *S. suis* genome sequences. Protein-encoding sequences that showed at least 80% identity over at least 80% of the length of the equivalent protein-encoding sequence extracted from P1/7 were retained and the core-genome defined as the protein-encoding sequences that were identified in

all assemblies using these criteria. The ***S. suis* accessory genome** (i.e. genes absent from one or more isolates or unique to a given isolate) was reconstructed using the program OrthoMCL [512-514]. Protein-encoding sequences were predicted using default parameters in Prodigal and combined with those extracted from published genome sequences. The proteins encoded by each predicted protein-encoding sequence were used as BLASTp queries against a bespoke BLAST database of all predicted protein sequences. The all-against-all BLAST output was parsed using the OrthoMCLBlastParser Perl script to compute the percent match of each hit. OrthoMCL was then used to identify reciprocal best similarity protein pairs, and homology groups identified by clustering with MCL using a low inflation value parameter. Assignment of protein-encoding sequences to homology groups can be inaccurate for a number of reasons, including: sequencing errors, assembly errors, failure to distinguish between orthologues and paralogues, failure to distinguish between recent duplicates, and recent pseudogenes; to address these issues, homology groups were subject to extensive quality control. This work was perfomed by post-doctoral Research Associate Dr Lucy A. Weinert as part of the BBSRC LoLa research grant BB/G019274/1.

Bayesian inference of the **genetic population structure in *S. suis*** [141] was performed using Bayesian Analysis of Population Structure (BAPS) software version 6.0, in particular the hierBAPS module that fits lineages to genome data using nested clustering. Three nested levels of molecular variation were fitted to core-genome alignments and five clusters at level one of the hierarchy identified. This work was perfomed by post-doctoral Research Associate Dr Jukka Corander (Department of Mathematics and Statistics, University of Helsinki, Finland) as part of the BBSRC LoLa research grant BB/G019274/1.

2.1.7 Swabs of oral fluid and material scraped from the palatine tonsils of pigs

Swabs of oral fluid or material scraped from the palatine tonsils of pigs on UK farms were collected between June 2013 and June 2014, and used to investigate the prevalence of invasive disease-associated and non-disease associated strains of *S. suis* in herds on farms in England and Wales not showing obvious signs of streptococcal disease. A total of three farms were sampled at two time points, five weeks and 20 weeks of age. Each farm operated an all-in-all-out management system, holding pigs from weaning (four weeks of age) until slaughter (~20-25 weeks of age). At each time point, per farm, 25 randomly selected pigs from the same group within the herd were selected for sampling. Pigs were caught and held using a snare around the upper jaw, and a custom designed iron wedge pushed between the teeth to enable scrapings to be collected. Material scraped from both palatine tonsils was immediately transferred to sterile Dryswabs (Medical Wire & Equipment) and transported at ambient temperature. In the laboratory, swabs were used to inoculate 90 mm Columbia Agar Base with Sheep Blood plates (Oxoid) and swabs then stored at -80 ˚C.

2.2 Methods

2.2.1 Statistical methods

Unless otherwise stated, all statistical analyses were performed using the standard R environment for statistical computing and graphs version 3.1.1 "Bug in Your Hair" [515] on a MacBook Pro (late 2013) 2.8 GHz Intel Core i7 16 GB 1600 MHz DDR3 OS X version 10.9.5.

2.2.1.1 Chi-squared test for independence

The Chi-squared test for independence, implemented in the R package: *stats* [516], was used to identify statistically significant associations between the positive detection of protein-encoding sequences (genotype) and phenotype (invasive disease-associated and non-disease associated groups). The Chi-squared test for independence compared observed frequencies with expected frequencies, and in doing so calculated a test statistic that if greater than the critical value was reason enough to reject the null hypothesis of independence (p-value <0.05). Bonferroni adjustment (α/n) was used to control for family-wise error associated with multiple sampling.

2.2.1.2 Cochran-Mantel-Haenszel Chi-squared test for count data

The Cochran-Mantel-Haenszel (CMH) Chi-squared test for count data, implemented in the R package: *stats* [516], was used to minimise the number of spurious associations due to an underlying bacterial population structure. The Chi-squared test for independence model compares observed frequencies with expected but does not take into consideration bacterial population structure. Population groups identified by Bayesian Analysis of Population Structure (BAPS) have previously been described for *S. suis* by Weinert *et al.* [141] (see Appendix 8.x). The CMH Chi-squared test for count data generated a test statistic that across all five BAPS groups was the odds of a protein-encoding sequence being present in invasive disease-associated isolates compared with non-disease associated isolates, a p-value was then calculated to assess the statistical significance of the common odds ratio. Bonferroni adjustment (α/n) was used to control for family-wise error associated with multiple sampling (α = 0.05).

2.2.1.3 Discriminant Analysis of Principal Components

The Discriminant Analysis of Principal Components (DAPC) [517, 518], implemented in the R package: *adegenet* [519, 520], was used to identify genetic differences between pre-defined (invasive disease-associated and non-disease associated) phenotypic groups. DAPC is a multivariate genome-wide association methodology that identified genetic structures without making assumptions about underlying population genetic models. Identified by principal component analysis (PCA) overall genetic variation was partitioned and then the best discrimination of between-group variation summarised while within-group variation overlooked. As a result, the contribution of alleles (detection of protein-encoding sequences) to the genetic structures identified allowed identification of regions of the genome driving genetic divergence among groups [518].

The total amount of original variation retained in the DAPC model affected which protein-encoding sequences contributed most to the separation of genetic structures. To assess what impact the total amount of original variation retained had on the DAPC model four independent DAPC analyses were performed retaining 60, 70, 80 or 90% of the original genetic variation. The 1% of ranked protein-encoding sequences contributing most to the discrimination of pre-defined phenotypic groups was then analysed, and those protein-encoding sequences consistently output by two or more DAPC analyses taken forward as candidates for pathotyping.

2.2.1.4 Logistic regression analysis

Logistic regression in the form of a generalised linear model (GLM), implemented in the R package: *logistf* [521], was used for the final selection of genetic markers to pathotype *S. suis*. Logistic regression is a statistical method for the analysis of a dataset in which there are one or more independent explanatory variables (shortlisted genetic markers) that are expected to influence the binary dependent variable (observed clinical phenotype; invasive disease-

associated: 1, non-disease associated: 0). Backwards step-wise selection, using penalised likelihood ratio tests, was used to select the fewest statistically significant genetic markers to differentiate *S. suis* isolates into disease and non-disease associated phenotypic groups. The number of explanatory variables for the logistic regression model was limited to no more than 10% of the sample size, a commonly accepted rule-of-thumb approach to avoid overfitting [522].

2.2.1.5 Interpretation of the real-valued output of a logistic regression model

2.2.1.5.1 Receiver operating characteristic curves

A receiver operating characteristic (ROC) curve, implemented in the R package: *ROCR* [523], was used to visualise at different cutoff thresholds the GLM performance metrics true positive rate (sensitivity) and false positive rate (1-specificity) in comparison to the observed clinical phenotype (considered to be the 'gold-standard'). DeLong's test for two correlated ROC curves, implemented in the R package: *pROC* [524], was used to identify statistically significant (p-value <0.05) differences between two ROC curves generated using different explanatory variables from the same data set.

2.2.1.5.2 Selection of a cutoff to convert the real-valued output of a generalised linear model into a binary classification decision

The real-valued output of a logistic regression model, i.e. the probability of causing disease, was converted into a binary class decision (disease-/non-disease associated) by selecting a cutoff threshold. Typical measures of binary diagnostic tests are the performance metrics: true positive rate (sensitivity; i.e. the proportion of isolates recovered from systemic sites and predicted to be invasive disease causing isolates), the true negative rate (specificity; i.e. the proportion of isolates recovered from the upper respiratory tract of a healthy pig without any typical signs of *S. suis* infection and predicted to be non-disease causing isolates), positive predictive value (PPV, precision; i.e. the proportion of isolates predicted to be disease causing isolates that were recovered from a systemic site), negative predictive value (NPV; i.e. the

proportion of isolates predicted to be non-disease causing isolates that were recovered from the upper respiratory tract of a healthy pig without any typical signs of *S. suis* infection) and accuracy. Performance metrics are heavily dependent on the cutoff chosen. As no cutoff was optimal according to all possible performance criteria, cutoff choice involved a trade-off of different performance metrics. The false negative rate (analogous to Type II error) was chosen as the most valuable performance metric for pathotyping *S. suis*, with a view to establish and then maintain a pig population free of invasive disease-associated strains. False negative rate is equivalent to 1 – sensitivity, therefore, selection of a cutoff producing a low false negative rate is the equivalent of selecting a cutoff producing a high sensitivity. False negative rate was calculated using the equation:

$$\text{False negative rate} = \frac{\text{False negative (Type II error)}}{\text{Phenotype positive}}$$

In order to establish and maintain a population free of disease-associated strains the F_1 score was also deemed to be a valuable performance metric. In statistical analysis of binary classification the F_1 score is a measure of accuracy that is the harmonic average of the sensitivity and precision of a test, and was considered more valuable than accuracy deduced from the number of true positives and the number of true negatives alone. The F_1 score was calculated using the equation:

$$F_1 \text{ score} = \frac{2}{\frac{1}{\text{True positive rate}} + \frac{1}{\text{Positive predictive value}}}$$

In summary, cutoff selection simply traded-off the closest F_1 score to 1 against a false negative rate closest to 0. Table 2.2 shows an extended 2x2 contingency table of commonly used performance metrics to compare binary diagnostic tests, and the equations used to calculate each performance metric.

Table 2.2. Performance metrics used to summarise and compare binary diagnostic tests in a paired study.

Performance metrics were calculated using the equations indicated in blue in the standard R environment for statistical computing and graphs version 3.1.1.

(a) F_1 score is the harmonic mean of the true positive rate (sensitivity) and positive predictive value (precision) of a test and calculated using the equation:

F_1 score = 2/((1/true positive rate)+(1/positive predictive value)).

		Phenotype			
	Total population	**Phenotype positive**	**Phenotype negative**		
mPCR result	**mPCR result positive**	**True positive**	Type I error **False positive**	True positive mPCR result positive **Positive predictive value**	False positive mPCR result positive **False discovery rate**
	mPCR result negative	Type II error **False negative**	**True negative**	False negative mPCR result negative **False omission rate**	True negative mPCR result negative **Negative predictive value**
		True positive Phenotype positive **True positive rate**	False positive Phenotype negative **False positive rate**	a **F_1 score**	
		False negative Phenotype positive **False negative rate**	True negative Phenotype negative **True negative rate**		

2.2.1.6 Comparison of generalised linear model performance metrics

McNemar's Chi-squared Test for Count Data, implemented in the R package: *stats* [516], was used to test for a statistically significant difference in the sensitivity and specificity of two binary diagnostic tests in a paired study. The Weighted Generalised Score Statistic for Comparison of Predictive Values as proposed by Kosinski [525], implemented in the R package: *DTComPair* [526], was used to test for statistically significant differences in (positive and negative) predictive values of two binary diagnostic tests.

2.2.1.7 McNemar's Chi-squared Test for Count Data

McNemar's Chi-squared Test for Count Data, implemented in the R package: *stats* [516], was used to test for significant differences in the sensitivities and specificities of two binary diagnostic tests in a paired study. McNemar's Chi-squared Test for Count Data is applied to a 2x2 contingency table, and tests for consistency in responses across two variables.

2.2.1.8 The Weighted Generalised Score Statistic for Comparison of Predictive Values

The Weighted Generalised Score Statistic for Comparison of Predictive Values as proposed by Kosinski[525], implemented in the R package: *DTComPair* [526], was used to test for significant differences in (positive and negative) predictive values of two binary diagnostic tests.

2.2.1.9 Comparison of the proportion of tonsil swabs testing positive for *Streptococcus suis* using two different detection strategies

Proportions were calculated by dividing the number of swabs confirmed to be positive for *S. suis* by the number of swabs collected per group sampled (n=25). To test for difference between two proportions, swabs testing positive for *S. suis* using qPCR in comparison to using culture-based methods, Z-values were calculated using the equation:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

where $\bar{p}$ is the proportion of swabs testing positive for *S. suis* obtained by pooling all *n* data: $((X_1+X_2)/(n_1+n_2))$, and $\bar{q} = 1 - \bar{p}$. Testing the null hypothesis (H_0): $p_1=p_2$

The Z-table was used to find probabilities for a statistical sample with a standard (Z-) normal distribution; and *p*-values < 0.05 considered significant. Barplot, implemented in the R package: *ggplot2*, was used to visualise the proportion of swabs confirmed to be positive for *S. suis*.

2.2.1.10 Kaplan–Meier curves of survival

Kaplan–Meier curves of survival, implemented in the R package: *survival* [527, 528], were used to visualise Greater Wax Moth (*G. mellonella*) larvae survival over 144 h post infection. The log-rank test, implemented in the R package: *survival*, was used compare Kaplan–Meier curves of survival and in all instances *p*-values <0.05 were considered statistically significant.

2.2.2 Comparative genomics methods

2.2.2.1 Investigation of molecular serotype prevalence in the training collection

Traditional serotyping (by capillary precipitation) data was unavailable for all *S. suis* isolates described in this thesis, therefore, molecular 'serotyping' was performed using an adaptation (for *in silico* use) of the mPCR assays described by Liu *et al.* [74]. Primer sequences were used as BLASTn queries (word_size 10) against a bespoke BLAST database consisting of the draft genome assemblies of all isolates described in this thesis. Nucleotide level matches to >95% of the total length of each primer sequence were considered hits, and the distance between hits compared to reported PCR amplicon sizes [74]. Isolates that could not be assigned to one of the 35 (1-34 and 1/2) originally described *S. suis* serotypes were deemed non-serotypable (NT).

2.2.2.2 Investigation of published putative "virulence-factor" prevalence in the training collection

Investigation into the prevalence of published putative "virulence-factors", extracted from a previous comprehensive review [129], in the training collection was performed *in silico*. The protein-encoding sequences of the list of 71 putative "virulence-factors" (Table 1.3) was extracted from GenBank. Protein-encoding sequences were used as tBLASTn queries against a bespoke BLAST database consisting of the draft genome assemblies of all isolates described in this thesis (Table 2.1). Amino acid level matches to >80% of >80% of the total length of each translated protein-encoding sequence were considered hits. The function geom_bar, implemented in the R package: *ggplot2* [529], was used to plot bars representative of the proportion of isolates testing positive for bacterial virulence and virulence-associated factors.

2.2.2.3 Estimation of the minimum number of genetic markers associated with observed clinical phenotype required to pathotype *Streptococcus suis*

To estimate the minimum number of genetic markers required to pathotype *S. suis*, the proportion of the disease-associated phenotype explained by increasing the cumulative number of genetic markers was calculated using a bespoke R script written by Post Doctoral Research Associate Dr Lucy Weinert (Department of Veterinary Medicine, University of Cambridge). Estimates were generated by simple random sampling of 1 of the 14 shortlisted genetic markers and calculating the proportion of the invasive disease-associated isolates in the training collection (n=53) positive for that marker, and this process then repeated for 10,000 random sampling events. To investigate the effect of increasing the cumulative number of genetic markers, the number of genetic markers randomly sampled from was increased to *x*, where *x* takes a value between 1 and 14, and the proportion of invasive disease-associated isolates in the training collection testing positive for *x* markers calculated; again this process was then repeated for 10,000 random sampling events. The process of increasing the number of candidate markers randomly sampled from by one was repeated until *x*=14. Boxplot, implemented in the R package: *graphics*, was used to plot boxes representative of the 1st quartile, a typical value of the proportion of disease phenotype (median) and 3rd quartile (respectively), as well as, whiskers representative of 1.5x the interquartile range; outliers were plotted as unfilled circles.

2.2.2.4 'Virulence-associated' gene profiling (*epf*, *mrp* and *sly*)

Virulence-associated gene (*epf*, *mrp* and *sly*) profiling was performed using an adaptation, for *in silico* use, of the method described by Silva *et al.* [400]. mPCR and singleplex-PCR primer sequences were used as BLASTn queries (word_size 10) against a bespoke BLAST database consisting of the (unfinished) draft genome assemblies of all 253 isolates described in this thesis (Table 2.1). Nucleotide level matches of >95% to the total length of each query sequence were considered hits, and the distance between hits compared to reported mPCR amplicon sizes.

2.2.2.4.1 Profiling the large and small molecular weight variants of extracellular factor and muramidase-released protein

Large and small molecular weight variants [53, 409, 530] of extracellular factor and muramidase-released protein were determined using an adaptation, for *in silico* use, of the gene profiling method described by Silva *et al.* [400]. PCR primer sequences were used as BLASTn queries (word_size 10) against a bespoke BLAST database consisting of the (unfinished) draft genome assemblies of all 253 isolates described in this thesis (Table 2.1). Nucleotide matches to >95% of the total length of each query sequence were considered hits, and the distance between hits compared to anticipated PCR amplicon sizes (747 bp for mrp^{S}, 1148 bp for mrp^{+}, 1556 bp for mrp^{*}, ~1600 bp for mrp^{**}, ~2000 bp for mrp^{***} and ~2400 bp for mrp^{****}). No epf^{*} variants were found.

2.2.2.5 Minimum core genome sequence typing

Minimum core genome (MCG) sequence typing is one of the most recent methods used to characterise and sub-type *S. suis*, and was performed using an adaptation, for *in silico* use, of the method described by Zheng *et al.* [116]. Multiplex-PCR and singleplex-PCR primer sequences were used as BLASTn queries (word_size 10) against a bespoke BLAST database consisting of the (unfinished) draft genome assemblies of all 253 isolates described in this thesis. Nucleotide level matches to >95% of the total length of each query sequence were considered hits, and the distance between hits compared to reported multiplex-PCR amplicon sizes. Nucleotide sequences between primer sequence matches were then extracted, aligned (Figure 2.1) against the MCG typing reference strain GZ1 (GenBank: CP000837), and the 10 SNPs of interest called allowing isolates to be assigned to one of the seven MCG groups.

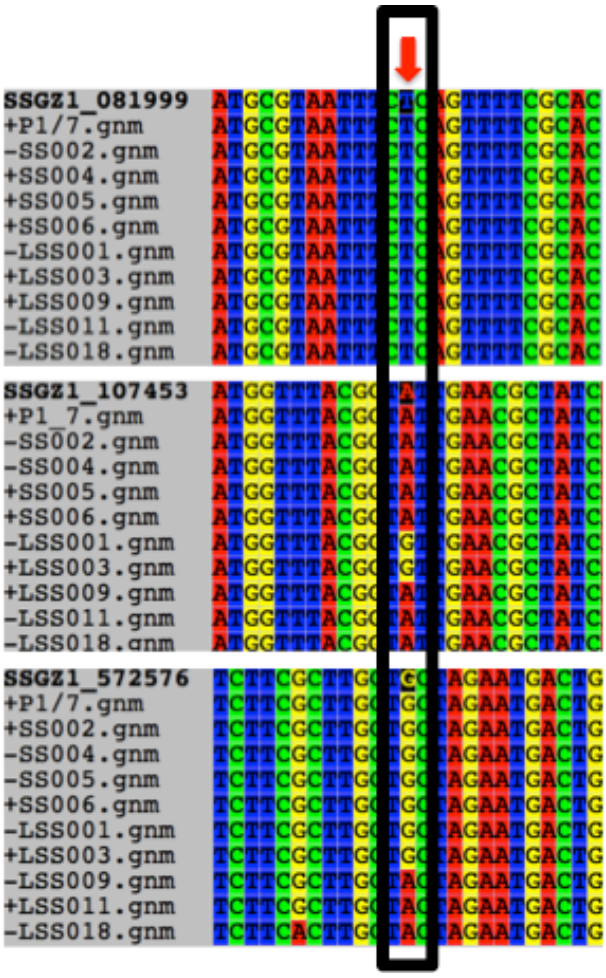

Figure 2.1. Example multiple sequence alignment used to assign MCG sequence type.

Example multiple sequence alignment of three MCG sequence typing SNPs, 81999, 107453 and 572576 (out of 10), in reference strain GZ1 (GenBank: CP000837).

2.2.3 Bacterial culture

Unless otherwise stated, *S. suis* isolates were recovered from storage (-80 ˚C in 20% (v/v) glycerol) on 90 mm Columbia agar base with 5% sheep blood plates (Oxoid Ltd.) at 37 ˚C for 18 – 24 h in the presence of 5% CO_2. Liquid cultures, inoculated with a single colony pick first grown on solid agar, were grown in Todd-Hewitt broth (THB; Oxoid Ltd.) [531] supplemented with 0.2% Yeast Extract (Fisher Scientific) at 37 ˚C for 18 - 24 h without agitation. Prior to use THB was sterilised by autoclaving at 115 ˚C for 10 minutes in a Falcon 30 autoclave (LTE Scientific Ltd.). When appropriate growth medium was supplemented, after being autoclaved, with 100 µg/mL Spectinomycin (Sigma-Aldrich) sterilised using a 0.22 µM syringe driven filter unit (Millex).

2.2.3.1 Serotyping

Serotyping is the most widely used of the methods available to characterise and subtype *S. suis*, and continues to be an important part of the routine diagnostic procedure [26, 39]. Serotyping of the *S. suis* isolates in the original training collection was performed using reference antisera and the capillary precipitation technique at the APHA. Isolates that could not be assigned to one of the 35 (1-34 and 1/2) originally described *S. suis* serotypes were deemed to be non-serotypable (NT).

2.2.3.2 Growth curves

Independent overnight THB cultures were diluted to an OD_{595} = 0.05 with fresh pre-warmed THB, and 300 µL added to wells of a sterile 10x10-well flat-bottom Honeycomb plate (OY Growth Curves). Bacterial cultures were incubated at 37 ˚C for 24 h without agitation on a Bioscreen C instrument (OY Growth Curves), and A_{595}, used as a proxy for bacterial growth, measured every 1 h. Bacterial growth data was analysed using Prism 6 (GraphPad Software, CA, USA). Multiple t-tests were performed to identify statistically significant (*p*-values <0.05) differences between growth curves. The Holm-Sidak method was used to control for family-wise error associated with multiple sampling.

2.2.3.3 Biofilm production

Bacterial isolates were tested for production of biofilm using the Microtiter plate biofilm formation assay protocol originally described by Grenier and colleagues [532, 533] with modifications taken from Wang *et al.* [534]. Independent overnight THB cultures were diluted to an OD_{595} of 0.2 with THB, and 100 µL added to wells of a sterile 96-well flat-bottom tissue culture plate (Greiner Bio-One Ltd.) containing 100 µL fresh THB. Plates were incubated at 37 ˚C for 24 h without agitation. Medium and planktonic bacteria were then removed and wells washed three times with 250 µL PBS (pH 7.2; Sigma-Aldrich). Remaining biofilm was fixed with 200 µL methanol for 10 min and stained with 200 µL 1 % (w/v) crystal violet (Sigma-Aldrich) for 10 min. Wells were washed five times with 250 µL PBS to removed unbound crystal violet and dried at 37 ˚C for 2 h. To release the crystal violet stain from fixed biofilms 200 µL of 95 % (v/v) ethanol was added to each well and plates agitated at 500 rpm for 10 min. A_{595nm} was measured with a SpectraMaxM5 Microplate Reader (Molecular Devices LLC). Wells with sterile medium were used as controls. Unless otherwise stated, all assays were performed in triplicate and repeated starting from new overnight cultures.

2.2.4 Molecular methods

2.2.4.1 DNA extraction

Unless otherwise stated, extraction of genomic DNA from bacteria was performed using the GenElute Bacterial DNA Kit (Sigma-Aldrich) according the manufacturer's recommendations.

2.2.4.1.1 DNA extraction without culture and first isolating single colonies

Extraction of genomic DNA from swabs without culture and first isolating single colonies/clones was performed using the GenElute Bacterial DNA Kit (Sigma-Aldrich) with adaptation of the manufacturer's recommendations for sample preparation. Material on a swab was resuspended in 1 mL 1x PBS and vortexed for 10 s. A total of 500 µL of resuspended material was then centrifuged at 16,000 rpm for 1 min, and the supernatant discarded. The remaining pellet was used as the 'starting material' and genomic DNA extraction thereafter performed according to the manufacturer's recommendations; with DNA recovered from the spin column using 100 µL of nuclease-free water.

2.2.4.2 Polymerase chain reaction

The polymerase chain reaction (PCR) was performed using the KOD Hot Start DNA Polymerase Kit (Novagen). Unless otherwise stated, all PCR experiments contained the same reagents except for primers and template DNA. The reaction mixture (50 µL) for each PCR consisted of 1x KOD Hot Start DNA Polymerase buffer, 1.5 mM $MgSO_4$, 0.2 mM dNTPs (each), 0.3 mM of each primer, 0.02 U/µL KOD Hot Start DNA Polymerase and 10 ng DNA, made up to 50 µL with RNase-free water. The three-step thermal cycling program for all PCR experiments was as follows, 95 ˚C for 2 minutes, followed by 35 cycles of (denaturation) 95 ˚C for 20 s, (annealing) 68 ˚C for 10 s, and (extension) 70 ˚C for 20 s/kb using a T100 Thermal Cycler (Bio-Rad).

2.2.4.3 Agarose gel electrophoresis

Unless otherwise stated, PCR products were analysed by gel electrophoresis using 2% (wt/vol) UltraPure Agarose (Invitrogen) gels made with 1x Tris base, boric acid and EDTA (TBE) buffer; and contained 1x SYBR Safe DNA gel stain (Invitrogen). Running time was 60 min at a constant 100 V. Results were visualised using a Gel Doc XR+ Gel Documentation System (BioRad).

2.2.4.4 Purification of DNA from agarose gels

PCR products were purified using the Wizard SV Gel and PCR Clean-Up System (Promega) according to the manufacturer's instructions. When gel extraction was required, PCR amplicons were excised using disposable scalpel blades (Swann Morton) in the thinnest slice of agarose gel possible. A UV transilluminator (UVP) was used to visualise amplicons with exposure kept to a minimum.

Where appropriate, PCR products were Sanger sequenced using the Source Bioscience Lifesciences Sanger Sequencing Service. Sequencing data was returned using the SpeedREAD Automated Data Delivery System and aligned with reference sequences of the target protein-encoding sequence using CodonCode Aligner software version 7.0.1 (CodonCode Corporation).

2.2.4.5 Multiplex polymerase chain reaction

Multiplex-PCR (mPCR) was performed using the Multiplex PCR *Plus* Kit (Qiagen). Unless otherwise stated, all mPCRs contained the same reagents except for primers and template DNA. The reaction mixture (50 µL) for each mPCR consisted of 25 µL 2x Multiplex PCR Master Mix, 5 µL 10x CoralLoad Dye, 10 µL RNase-free water, 0.2 µM (final concentration) of each primer, and 10 ng template DNA. The three-step thermal cycling program for all reactions was as follows: 95 ˚C for 5 min followed by 35 cycles of (denaturation) 95 ˚C for 30 s, (annealing) 66 ˚C for 90 s, and (extension) 72 ˚C for 90 s; with a final extension of 68 ˚C for 10 min performed on a T100 Thermal Cycler (Bio-Rad).

2.2.4.5.1 Multiplex-PCR oligonucleotide primer design

The online software primer3 version 4.0.0 (http://primer3.ut.ee) was used to design mPCR oligonucleotide primers. Multiplex-PCR primers were designed to target conserved regions within the protein-encoding sequence of markers (as opposed to flanking regions). Primers were designed to have similar physical characteristics, enabling simultaneous amplification under the same thermal cycling conditions and in multiplex reactions. Primer length (21-30 bp), GC content (40-60%), melting temperature (>68 ˚C if possible; but at least 60 ˚C), and expected amplicon size (100-1000 bp) were based on the manufacturer's recommendations for primer design from the Multiplex PCR *Plus* Kit (Qiagen). Prior to ordering, all primers were queried against the NCBI non-redundant nucleotide database to check for matches to other non-*S. suis* bacterial species. Primers were synthesised by Sigma-Aldrich (Haverhill, UK) and delivered in solution (TE buffer; 10 mM Tris-Cl, 1 mM EDTA [pH 8.0]) at a stock concentration of 100 µM; primers were used at a working stock concentration of 20 µM.

2.2.4.6 Quantitative real-time polymerase chain reaction

Quantitative real-time PCR (qPCR) was performed using the SensiFAST™ SYBR® No-ROX Kit (Bioline) in 0.1 mL Strip Tubes and Caps (Qiagen), according to the manufacturer's instructions. Unless otherwise stated, all qPCRs contained the same reagents except for primers and template DNA. The reaction mixture (20 µL) for each qPCR consisted of 10 µL 2x SensiFAST™ SYBR® No-ROX Mix, 3.4 µL nuclease-free water, 400 nM (final concentration) of each primer and 5 µL template DNA. The three-step thermal cycling program for all reactions was as follows: 95 ˚C for 3 minutes, followed by 40 cycles of 95 ˚C for 5 s (denaturation), 65 ˚C for 10 s (annealing) and 72 ˚C for 5 s (extension); with a final melt curve analysis from 65 ˚C to 95 ˚C (increasing by increments of 1 ˚C waiting for 5 s before each acquisition) performed on a Rotor-Gene Q PCR cycler (Qiagen).

Where possible, consumables were purchased in bulk and from the same batches. To limit the effect of differences in pH or salt concentration, a pre-mix of working concentration qPCR reagents was created. The pre-mix, large enough for the whole experiment, was then aliquoted into single-use volumes and stored at -20 ˚C. Those qPCR reagents unable to be premixed (primers and template DNA) were also aliquoted into single-use volumes and stored at -20 ˚C. All qPCRs were run in triplicate and negative controls, without primers or template DNA, included with each run. All analyses were performed using Rotor-Gene Q software, version 2.3.1 (Qiagen).

2.2.4.6.1 Oligonucleotide primer design

The online software primer3 version 4.0.0 (http://primer3.ut.ee) was used to design qPCR oligonucleotide primers. Quantitative real-time PCR primers were designed to target conserved regions within the protein-encoding sequence of markers (as opposed to flanking regions). Primers were designed to have similar physical characteristics, enabling simultaneous amplification under the same thermal cycling conditions. Primer length (18-24 bp), GC content (~50%), melting temperature (~60 ˚C; and within 1 ˚C of each other), and expected amplicon size (80-200 bp; not exceeding 400 bp) were based on the manufacturer's recommendations for primer design from the SensiFAST™ SYBR® No-ROX Kit (Bioline). Prior to ordering, all primers were queried against the NCBI non-redundant nucleotide database to check for matches to other non-*S. suis* bacterial species. Primers were synthesised, purification scale: RP1 (Reverse-Phase Cartridge), by Sigma-Aldrich (Haverhill, United Kingdom) and delivered in solution (TE buffer; 10 mM Tris-Cl, 1 mM EDTA [pH 8.0]) at a stock concentration of 100 µM; primers were used at a working stock concentration of 10 µM.

2.2.4.6.2 Oligonucleotide primer efficiencies

Robust and precise qPCR experiments are associated with high PCR amplification efficiencies [535]. PCR efficiencies are a measurement of the rate at which a PCR amplicon is generated, and are dependent on the primers used. Amplification efficiencies (or 'primer efficiencies'), per primer pair, were calculated over the range ~1x 10^0 - 1x 10^7 genome copies per reaction using a calibration curve and the equation:

$$\mathrm{E} = 10^{(-1/m)} - 1$$

where E denotes PCR amplification efficiency; and *m* the slope of the linear regression line, when the logarithm of the initial template concentration (the independent variable) is plotted on the X-axis and the C_t (the dependent variable) plotted on the Y-axis

2.2.4.6.3 Number of genome copies per reaction

To generate calibration/standard curves in units of genome copies per reaction, the number of genome copies per nanogram of *S. suis* genomic DNA was estimated using the equation:

$$C = \frac{(\text{Mass of DNA (ng)})(\text{Avogadro's constant})}{(\text{Molecular weight of a deoxynucleotide base pair (Daltons)})(\text{Genome size (bp)})}$$

where C denotes the number of genome copies. The average molecular weight of a deoxynucleotide base pair was assumed to be 650 Daltons; and genome size for *S. suis* assumed to be 2.1 Mb (2.1×10^6 bp)

Rearrangement of the number of genome copies equation to make mass of DNA (ng) the subject, estimated 23 ng of *S. suis* genomic DNA to be equivalent to $\sim 1 \times 10^7$ genome copies. To cover the range $\sim 1 \times 10^0$ - 1×10^7 genome copies per reaction, 10-fold serial dilutions were performed in 10 mL volumes (total) using nuclease-free water (Promega). Dilutions of template genomic DNA for qPCR standards were then aliquoted into single use volumes and stored at -20 ˚C.

2.2.4.6.4 Standard curve equations

Standard curve equations, per primer pair, were used to estimate the unknown number of genome copies in experimental samples. A standard curve was generated for each primer pair, over the range ~1x 10^1 - 1x 10^5 genome copies per reaction. Each standard curve was generated from four biological repeats using independent single-use qPCR premixes, and each biological repeat performed in triplicate per 10-fold serial dilution of template DNA. The concentration of DNA was linked to the number of thermal cycles (C_t) required to reach the fluorescence threshold at which point individual amplification reactions contained identical amounts of DNA using the standard curve equation:

$$\text{Concentration} = 10^{a(\text{Ct})+b}$$

where a denotes the slope; and b the Y-axis intercept of the linear regression line, when the logarithm of the initial template concentration is plotted on the X-axis and the C_t plotted on the Y-axis

Limited to a 72-well qPCR rotor it was not possible to generate a full standard curve per primer pair for each rotor run. Instead, adaptation of a method described by Grant *et al.* [536] was used to calculate robust estimates of the standard curve equation for each primer pair, that were used to assess both i) the performance and variability between rotor runs and ii) estimate the number of genome copies in experimental samples. The "Auto-Find Threshold" function of Rotor-Gene Q analysis software version 2.3.1 (Qiagen) was used to find the fluorescence threshold that delivered the best-fit of each standard curve to the samples defined as standards. However, to draw direct comparison between C_t values from separate rotor runs, the same fluorescence threshold must be used to analyse each rotor run [537]. The mean fluorescence threshold, for each primer pair, was used to re-analyse the four biological repeats of each standard curve. A regression line, per biological repeat, was fitted and used to generate estimates of the variables a and b, which were then averaged across the repeats to give a final standard curve equation per primer pair. These robust estimates of the standard curve equation, per primer pair, were then used to predict the unknown number of genome copies in an experimental sample.

2.2.4.6.5 Quality control

The performance and variability between qPCR rotor runs was assessed using a method described by Grant *et al.* [536]. In brief, a single linear regression line was fitted to the C_t values of the four biological repeats of each standard curve. Treating C_t values as random samples from a normal distribution, variability reflected in the confidence intervals was deemed to be an artefact of qPCR rotor runs. Standards of known concentration (10^3 genome copies per reaction; arbitrarily chosen) were used to pass/fail rotor runs based on the following criteria: to pass i) the standard deviation of C_t values must be less than 1.0 and ii) the mean C_t value must fall within +/- two standard deviations of the standard curve. If any one of the criteria listed above were not met, qPCR rotor runs were deemed to have failed and all experiments repeated. Individual experimental samples were also subject to quality control based on the following criteria: to pass, the standard deviation of C_t values must be less than 1.0; if not, experiments were deemed to have failed and repeated.

2.2.4.6.6 Confidence intervals around the estimated number of genome copies

Confidence intervals around the estimated number of genome copies of experimental samples were generated using a Bayesian approach, as previously described by Grant *et al.* [536]. In brief, using simple linear regression to estimate the standard curve equation variables *a* and *b* also allowed confidence intervals around the mean number of genome copies to be derived on the log10 scale. To make interference about individual experimental samples *prediction intervals* were generated to provide a measure of the degree of uncertainty surrounding the estimated number of genome copies. However, converting *prediction intervals* from the log10 back to the original scale is non-trivial, and so a Bayesian approach to model fitting was used. This approach used a numerical Markov chain Monte Carlo algorithm, implemented in the open source package WinBUGS [538, 539], to fit the model. For each experimental sample, 95% credible intervals for the predicted number of genome copies were calculated by taking the 2.5th and 97.5th percentiles in each case. Instead of measuring uncertainty in the mean estimates, this interval is generated centred about the median of the estimated number of genome copies, reflecting the fact that there is the same probability of lying above the interval as below it. This work was performed using adaptation of a bespoke R script written by post-doctoral research associate Dr Trevelyan J. McKinley (Department of Veterinary Medicine, University of Cambridge; current address College of Engineering, Mathematics and Physical Sciences, University of Exeter) for a previous research project.

2.2.4.6.7 Limit of detection

Analytical sensitivity refers to the ability of an assay to accurately detect very low concentrations of a given substance. Typically, for qPCR analytical sensitivity is expressed as the limit of detection, which is the lowest concentration that can be detected with reasonable certainty; usually 95% probability (theoretically, three copies per PCR is the most sensitive limit of detection possible) [535]. Limit of detection was determined by performing ten replicates per dilution over the range $\sim 10^0$ - 10^2 genome copies per reaction. The dilution with the lowest concentration meeting the following requirements was deemed to be the limit of detection: i) the standard deviation of C_t values must be less than 1.0 and ii) the number of replicates with positive detection must be greater than 95%. The mean C_t value of the lowest concentration was then used to calculate the limit of detection and expressed as a number of genome copies per reaction for each primer pair.

2.2.4.6.8 Limit of quantification

The limit of quantification for quantitative real-time PCR is the lowest concentration that can be accurately quantified with 95% confidence [535]. Limit of quantification was calculated, per primer pair, using the equation:

$$C_{t\ \text{(Limif of quantification)}} = C_{t\ \text{(Limit of detection)}} - 2\left(\sigma_{C_{t\ \text{(Limit of detection)}}}\right)$$

where $\sigma_{C_{t\ \text{(Limit of detection)}}}$ is the standard deviation of the $C_{t\ \text{(Limit of detection)}}$

2.2.5 Construction of isogenic mutants

2.2.5.1 Identification of *S.suis* isolates capable of natural genetic transformation

The addition of exogenous ComS has been shown to induce the transient physiological bacterial state of natural genetic transformation (competence) in *S. suis* under laboratory conditions [490]. The ComS protein-encoding sequence (SSU0050) was used as a BLASTn query to identify all isolates of the original training collection with 100% matches to the 66 bp nucleotide sequence. Isolates meeting these criteria were deemed able to enter the transient physiological bacterial state of natural competence in the presence of exogenous ComS (see Appendix Table 8.1).

2.2.5.2 Design and construction of isogeneic *virA* operon mutant strains

Guidance and practical help with the design of isogenic *virA* operon mutant strains was provided by Postdoctoral Research Associate Dr. Edoardo Zaccaria (Department of Animal Sciences, subdivision HMI, Wageningen University, the Netherlands), while I was a Visiting Research Fellow and guest of Prof. Jerry M. Wells' research group. Standard spliced overlap extension PCR techniques [540] were used to create, for transformation, the double-stranded DNA construct consisting of three parts, including i) a downstream of the operon flanking region, ii) a spectinomycin resistance cassette and iii) a upstream of the operon flanking region. Individual fragments were synthesised using the oligonucleotide primers in Table 2.3, and then annealed to generate full length constructs by spliced overlap extension PCR [540]. Individual fragments were amplified by PCR, resolved on a 0.7% agarose gel, gel extracted using the Wizard SV Gel and PCR Clean-up System (Promega) and then used as the template for a second-round PCR to anneal two fragments using flanking primers. All constructs required three rounds of PCR. Final naked double-stranded DNA constructs were integrated into the wild-type genome by a double crossover event.

Table 2.3. Oligonucleotide primers designed to generate *virA* operon mutants.

The online software primer3 version 4.0.0 (http://primer3.ut.ee) was used to design oligonucleotide primers. Primers were designed to have similar physical characteristics, enabling simultaneous amplification under the same thermal cycling conditions. Primers were synthesised, purification scale: RP1 (Reverse-Phase Cartridge), by Sigma-Aldrich (Haverhill, United Kingdom) and delivered in solution (TE buffer; 10 mM Tris-Cl, 1 mM EDTA [pH 8.0]) at a stock concentration of 100 μM; primers were used at a working stock concentration of 10 μM.

Primer name	Sequence (5'-3')
1a/4a.S10Ex_UpFor	ATCCACGACAGTAGTGTTCCC
1b.S10Ex_UpRev	GACCTGCAGGCATCAAGAGACGGTTCTGTCACCATTCCAGAAG
2a.S10Ex_SPECFor	GGAATGGTGACAGAACCGTCTCTTGATGCCTGCAGGTCGACTC
2b/5b.LSS034Ex_SPECRev	GTTGCTGCGTTACTTGCCGGTACCCTATGCAAGGGTTTATTG
3a/6a.S10Ex_DownFor	AACCCTTGCATAGGGTACCGGCAAGTAACGCAGCAACTAGAAC
3b.S10Ex_DownRev	CGCAGAAGGCTACCAAGAAATC
4b.LSS034Ex_UpRev	TCGACCTGCAGGCATCAAGAACGGTTCTGTCACCATTCCAGAAG
5a.LSS034Ex_SPECFor	TGGAATGGTGACAGAACCGTTCTTGATGCCTGCAGGTCGACTC
6b.LSS034Ex_DownRev	GACGGAGTAGCTGCCAAATTAG
7a.S10KO_UpFor	CCGCCGTCTCCTTTATAATTCC
7b.S10KO_UpRev	CGACCTGCAGGCATCAAGCCAAGGCACCTAGCTGTAAACT
8a.S10KO_SPECFor	TACAGCTAGGTGCCTTGGCTTGATGCCTGCAGGTCGACTC
8b.S10KO_SPECRev	TGCTGCGTTACTTGCCCGGTACCCTATGCAAGGGTTTATT
9a.S10KO_DownFor	ACCCTTGCATAGGGTACCGGGCAAGTAACGCAGCAACTAGA
9b.S10KO_DownRev	AGAAACTGTCTGCGGATGTG
10a.LSS034KO_UpFor	GGGCTTCGGAGAATCGAATACC
10b.LSS034KO_UpRev	GACCTGCAGGCATCAAGCCGCTTCTGCCGTAAGGCTAATCC
11a.LSS034KO_SPECFor	TTAGCCTTACGGCAGAAGCGGCTTGATGCCTGCAGGTCGAC
11b.LSS034KO_SPECRev	CTTGCCCGAGATCAACGGTACCCTATGCAAGGGTTTATTG
12a.LSS034KO_DownFor	CCTTGCATAGGGTACCGTTGATCTCGGGCAAGTAACGCAGC
12b.LSS034KO_DownRev	TCCAAATACGCCGCACGATTC
x1.Ctrl_S10	TTACGGACGCTTGGAAACAG
x2.Ctrl_S10	GATTCAGCCACTGCATTTCC
x3.Ctrl_S10	TATTGCGGGAAATGCAGTGG
x4.Ctrl_S10	ATGTGAAGCTGGATGGAACC
x5.Ctrl_LSS034	GGCCGAAAGTTTCTTATGCC
x6.Ctrl_LSS034	GAATCATCCTCCCAAACAAG
x7.Ctrl_LSS034	TAGAACGTGCCCTCTACAAC

To transform *S. suis* isolates, overnight cultures were diluted 1:40 with fresh pre-warmed THB and incubated at 37 ˚C without agitation. After 1 h, 100 µL aliquots were removed from the reset culture and added to 5 µL of ComS (final concentration 250 mM) and 1 µL of naked double-stranded donor DNA (~100 ng/µL). The ComS peptide, to induce the transient physiological bacterial state of natural competence, was purchased from JPT Peptide Technologies (Berlin, Germany). Samples were incubated for 2h at 37 ˚C, and then diluted and plated onto Todd-Hewitt agar plates, supplemented with spectinomycin (100 µg/mL). Plates were incubated overnight at 37 ˚C with 5% CO_2.

2.2.6 Experimental infection of Greater Wax Moth larvae

Experimental infection of Greater Wax Moth (*G. mellonella*; see section 1.1.4.4) larvae with isolates of *S. suis* was performed as previously described by Velikova *et al.* [423]. Greater Wax Moth caterpillars in the final-instar larval stage were purchased from Livefood UK Ltd. (Somerset, UK), and stored in the dark at 15 ˚C. All larvae were used within five days of shipment. Before use, larvae were separated by weight and only larvae between 0.2 g - 0.3 g used for infection model experiments.

Unless otherwise stated, a 20 µL inoculum of bacteria adjusted in PBS to $1x\ 10^9$, $1x\ 10^8$, $1x\ 10^7$, $1x\ 10^6$ or $1x\ 10^5$ CFU/mL was used to infect the hemocel of *G. mellonella* larvae via the last posterior left proleg. Before experimental infection the area of injection was cleaned using an alcohol swab. Injections were performed using a Tridak STEPPER with 27G x 0.5 microlance hypodermic needles (Becton, Dickinson and Company Ltd.). Post experimental infection larvae were stored in 90 mm Petri dishes in the dark and incubated at 37 ˚C with 5% CO_2. Larvae survival was inspected every 24 h up to 144 h post-infection. Larvae were considered dead if unresponsive to touch. All experiments were performed using groups of 15 larvae, and experiments repeated using larvae from a different batch; data from replicate experiments was pooled to achieve n=30 per group. Two negative control groups were included in every experiment i) a no treatment control was included to assess background larval mortality and ii) a PBS only control was included to assess background larval mortality due to physical trauma.

Chapter 3

Identification of genetic markers to pathotype *Streptococcus suis*

3.1 Introduction

The detection of the *S. suis* species in asymptomatic carrier herds is of little practical value in predicting likelihood of future clinical relevance [11]. Instead, the value of future molecular tools for surveillance and preventive health management lies in the detection of *S. suis* strains that have genetically increased potential to cause disease and are colonising presently healthy animals [11]. At present, numerous methods exist to characterise and sub-type *S. suis* isolates as part of epidemiological studies, the most recent of these have been reviewed in chapter 1 (section 1.1.1). However, despite some of these methods being highly discriminatory and reproducible, these methods remain laborious, costly and difficult to apply to routine diagnostic testing outside of high-income countries [86].

A so-far unexploited method for the identification of virulence-associated factors in *S. suis* are the genome-wide association studies (GWAS). Although the use of GWAS is common for SNP and small indel analyses of eukaryotic genomes, only relatively recently have these methods been adapted for bacterial genomics. At present, as a result of advances and reductions in cost, the whole-genome sequencing of multiple isolates of the same species has opened up exciting possibilities to reassess diagnostic and surveillance programs for bacteria that behave as asymptomatic commensal-like strains contributing to the normal oral microflora, as well as, an infectious agent responsible for invasive disease; such as *S. suis* [141, 477, 493, 541]. In fact, this explosion in the amount of whole-genome sequencing data has provided the necessary statistical power to allow the identification of robust associations between genotype and phenotypes including virulence, and is continually growing in popularity being applied to many different bacterial species [542, 543].

Descriptive comparisons between commensal and pathogenic isolates of the same species began with *Escherichia coli* [544-546], and have since progressed to the investigation of recombination and the acquisition of antimicrobial resistance genes as studied in *Streptococcus pneumonia* [547]. The availability of more complex methods has made it possible to detect statistically

significant associations of SNPs and indels with phenotypes. For example, tree-based methods have been used to detect *Campylobacter jejuni* genotypes associated with host-restriction [548], the Discriminant Analysis of Principal Components method (DAPC) has been used to detect genotypes associated with temporal variation in seasonal influenza (H3N2) [517, 518], and PLINK [549] has been used to detect genotypes associated with the phenotypes toxicity (the ability to destroy host cell membranes) and adhesion (the ability to adhere to human tissues) in *Staphylococcus aureus* [550]. Such studies sought to identify associations between phenotype and genotype that would improve surveillance and diagnostic tools, as well as contribute to an increased understanding of the respective pathogens – potentially leading to new treatment strategies.

As mentioned in chapter 1, previous bioinformatics analyses in our group revealed significant genomic differences between invasive disease-associated isolates and non-disease associated *S. suis* isolates recovered from the upper respiratory tract of pigs on farms in England and Wales [141]. In fact, in the corresponding Nature Communications publication disease-associated isolates were described as having smaller genomes, and as a result, a reduced gene content, in comparison to non-disease associated isolates. However, despite this it was also found that *S. suis* isolates causing invasive disease in pigs had an overrepresentation of genes previously reported to be associated with virulence [141, 360]. This chapter builds on these findings and describes first the identification and then compilation of a shortlist of genetic markers in *S. suis* isolates that were found to have statistically significant associations with the phenotypes i) invasive disease or ii) asymptomatic carriage on the palatine tonsils of pigs on UK farms. The work described in this chapter used the whole-genome sequencing data of a collection of 115 isolates of *S. suis* recovered from pigs on farms in England and Wales, that had previously been categorised as being either associated with invasive disease or asymptomatic carriage in the upper respiratory tract of pigs based on detailed clinical metadata (see Materials and methods, Table 2.1).

In order to define a shortlist of genetic markers associated with our phenotypic groups that could potentially be exploited to pathotype *S. suis*, three complementary approaches are described i) a comparison of observed and expected frequencies of protein-encoding sequences using the Chi-squared test for independence, ii) a consideration of the impact that bacterial population structure had on these findings, using the Cochran-Mantel-Haensael Chi-squared test for count data (an approach also used in the software PLINK [549] and iii) the multivariate Discriminant Analysis of Principal Components [518]. All of these approaches can be used to investigate both SNPs within the core-genome and/or variation in the pan-genome (in the form of presence and absence of protein-encoding sequences). However, with future intentions to create a simple molecular tool to pathotype *S. suis*, using techniques such as a mPCR, the decision was made to investigate the variation in the presence and absence of protein-encoding sequences in the accessory genome of the 115 isolates of the original training collection. In doing so, any suitable genetic markers could be built into a mPCR based pathotyping tool – a simple, rapid and comparable technique available to most laboratories worldwide.

3.1.1 Chapter objectives

1. To investigate serotype and published virulence factor genotype in the original training collection of 115 *S. suis* isolates with defined clinical phenotypes
2. To identify genetic markers (protein-encoding sequences) in the *S. suis* accessory genome associated with observed clinical phenotype
3. To compile a shortlist protein-encoding sequences that would be suitable targets for a multiplex-PCR tool to pathotype *S. suis*
4. To identify and compile a shortlist of *S. suis* species-specific markers to complement any molecular pathotyping markers

3.2 Results

3.2.1 Investigation of serotype prevalence in the training collection

Serotyping is the most widely used of the methods available to characterise and subtype *S. suis*, and to date, continues to be an important part of the routine diagnostic procedure [26, 39]. In fact, historically serotyping has commonly been used as a proxy for virulence potential. Therefore, investigation of the serotype prevalence in the *S. suis* isolates of the training collection was an obvious starting point of this PhD project. Table 3.1 shows the serotype assigned to each of the 115 training collection isolates using two approaches, i) traditional capillary precipitation and ii) adaptation, for *in silico* use, of the mPCR assays described by Liu *et al.* [74]. Figure 3.1 summarises the information in Table 3.1, and shows the cumulative frequency of isolates assigned to one of the 35 (1-34 and 1/2) serotypes originally described for *S. suis*. Serotype 2 was the most frequently identified serotype in the training collection, regardless of observed clinical phenotype. Serotype 2, based on the 'gold-standard' capillary precipitation data, was also the serotype most frequently assigned (n=28, of a total 53; 54%) to the invasive disease-causing isolates, in agreement with previously published reports in the literature. Despite good agreement, discrepancies do exist between the traditional and *in silico* approaches used to generate the serotyping data in Table 3.1. The most obvious being that adaptation, for *in silico* use, of the mPCR assays described by Liu *et al.* was able to assign a serotype to 34 isolates deemed to be non-serotypable (NT) using the traditional capillary precipitation technique. Six other discrepancies also exist between the *in silico* approach and the 'gold-standard' capillary precipitation approach undertaken at the APHA. For example, the three non-disease associated isolates (LSS078, LSS095 and LSS096) assigned to serotype 1 by the capillary precipitation technique were deemed to be serotypes 12, 11 and 11 respectively using the *in silico* approach. Other important observations include not all serotype 2 isolates are associated with invasive disease, in fact, differences in disease association are present in numerous serotypic groups. Molecular 'serotyping' assigned just one invasive disease-associated isolate in the training collection to a serotype other than 1-10 and 14 (the most common serotypes recovered from clinical samples [74]).

Table 3.1. The serotype assigned to *Streptococcus suis* isolates in the training collection using traditional capillary precipitation and *in silico* approaches.

Serotyping was performed using two techniques (i) 'gold-standard' capillary precipitation at the Animal and Plant Health Agency (APHA) and (ii) using adaptation, for *in silico* use, of the multiplex-PCR assays described by Liu *et al.* [74]. The Liu *et al.* molecular 'serotyping' method, the most recent at the time of undertaking this work, is unable to differentiate all reported *S. suis* serotypes due to closely related *cps* genes. Isolates that could not be assigned to one of the 35 originally described *S. suis* serotypes were designated non-serotypable (NT). Grey-scale has been used to highlight discrepancies between the two techniques used to serotype isolates.

			Serotyping	
Isolate	**Collection**	**Phenotype**	**Capillary precipitation**	***In silico* (Liu *et al.* [74])**
P1/7	Original training	Disease-associated	2	2&1/2
SS002	Original training	Disease-associated	2	2&1/2
SS004	Original training	Disease-associated	2	2&1/2
SS005	Original training	Disease-associated	1	1&14
SS006	Original training	Disease-associated	2	2&1/2
SS007	Original training	Disease-associated	2	2&1/2
SS008	Original training	Disease-associated	2	2&1/2
SS010	Original training	Disease-associated	2	2&1/2
SS012	Original training	Disease-associated	14	1&14
SS003	Original training	Disease-associated	14	1&14
SS015	Original training	Disease-associated	9	9
SS016	Original training	Disease-associated	14	1&14
SS017	Original training	Disease-associated	1	1&14
SS018	Original training	Disease-associated	7	7
SS021	Original training	Disease-associated	1	1&14
SS022	Original training	Disease-associated	1	1&14
SS024	Original training	Disease-associated	7	7
SS025	Original training	Disease-associated	2	2&1/2
SS028	Original training	Disease-associated	2	2&1/2
SS029	Original training	Disease-associated	14	1&14
SS035	Original training	Disease-associated	2	2&1/2
SS036	Original training	Disease-associated	2	2&1/2
SS038	Original training	Disease-associated	2	2&1/2
SS041	Original training	Disease-associated	7	7
SS042	Original training	Disease-associated	2	2&1/2
SS044	Original training	Disease-associated	2	2&1/2
SS045	Original training	Disease-associated	1	1&14
SS046	Original training	Disease-associated	2	2&1/2
SS053	Original training	Disease-associated	3	3
SS057	Original training	Disease-associated	2	2&1/2

SS058	Original training	Disease-associated	2	2&1/2
SS060	Original training	Disease-associated	2	2&1/2
SS062	Original training	Disease-associated	4	4
SS063	Original training	Disease-associated	14	1&14
SS065	Original training	Disease-associated	2	2&1/2
SS066	Original training	Disease-associated	1	1&14
SS068	Original training	Disease-associated	8	8
SS069	Original training	Disease-associated	2	2&1/2
SS071	Original training	Disease-associated	1	1&14
SS072	Original training	Disease-associated	1	1&14
SS077	Original training	Disease-associated	14	1&14
SS083	Original training	Disease-associated	7	7
SS085	Original training	Disease-associated	2	2&1/2
SS086	Original training	Disease-associated	2	2&1/2
SS087	Original training	Disease-associated	2	2&1/2
SS088	Original training	Disease-associated	9	8
SS093	Original training	Disease-associated	2	2&1/2
SS095	Original training	Disease-associated	2	2&1/2
SS096	Original training	Disease-associated	2	2&1/2
SS097	Original training	Disease-associated	16	16
SS098	Original training	Disease-associated	2	2&1/2
SS099	Original training	Disease-associated	2	2&1/2
SS100	Original training	Disease-associated	1/2	2&1/2
LSS001	Original training	Non-disease associated	2	2&1/2
LSS003	Original training	Non-disease associated	1/2	2&1/2
LSS009	Original training	Non-disease associated	NT	10
LSS011	Original training	Non-disease associated	NT	31
LSS018	Original training	Non-disease associated	NT	16
LSS020	Original training	Non-disease associated	4	4
LSS023	Original training	Non-disease associated	NT	24
LSS025	Original training	Non-disease associated	1/2	2&1/2
LSS027	Original training	Non-disease associated	3	10
LSS030	Original training	Non-disease associated	NT	31
LSS031	Original training	Non-disease associated	NT	24
LSS032	Original training	Non-disease associated	NT	24
LSS034	Original training	Non-disease associated	NT	8
LSS037	Original training	Non-disease associated	NT	24
LSS038	Original training	Non-disease associated	NT	NT
LSS039	Original training	Non-disease associated	NT	31
LSS040	Original training	Non-disease associated	NT	6
LSS041	Original training	Non-disease associated	NT	21
LSS042	Original training	Non-disease associated	NT	16
LSS044	Original training	Non-disease associated	NT	15
LSS047	Original training	Non-disease associated	NT	16
LSS048	Original training	Non-disease associated	NT	31
LSS049	Original training	Non-disease associated	NT	NT
LSS053	Original training	Non-disease associated	NT	21
LSS054	Original training	Non-disease associated	NT	NT

LSS055	Original training	Non-disease associated	NT	9
LSS057	Original training	Non-disease associated	4	NT
LSS058	Original training	Non-disease associated	NT	31
LSS059	Original training	Non-disease associated	NT	10
LSS060	Original training	Non-disease associated	2	31
LSS061	Original training	Non-disease associated	2	2&1/2
LSS062	Original training	Non-disease associated	2	2&1/2
LSS063	Original training	Non-disease associated	2	2&1/2
LSS064	Original training	Non-disease associated	NT	21
LSS065	Original training	Non-disease associated	NT	10
LSS067	Original training	Non-disease associated	NT	8
LSS068	Original training	Non-disease associated	NT	8
LSS069	Original training	Non-disease associated	NT	31
LSS070	Original training	Non-disease associated	NT	9
LSS071	Original training	Non-disease associated	2	2&1/2
LSS072	Original training	Non-disease associated	2	2&1/2
LSS075	Original training	Non-disease associated	NT	31
LSS076	Original training	Non-disease associated	NT	8
LSS077	Original training	Non-disease associated	1/2	2&1/2
LSS078	Original training	Non-disease associated	1	12
LSS079	Original training	Non-disease associated	NT	15
LSS080	Original training	Non-disease associated	NT	31
LSS081	Original training	Non-disease associated	NT	15
LSS082	Original training	Non-disease associated	NT	16
LSS083	Original training	Non-disease associated	7	7
LSS084	Original training	Non-disease associated	7	7
LSS085	Original training	Non-disease associated	7	7
LSS086	Original training	Non-disease associated	1/2	2&1/2
LSS088	Original training	Non-disease associated	NT	NT
LSS089	Original training	Non-disease associated	NT	8
LSS090	Original training	Non-disease associated	NT	31
LSS091	Original training	Non-disease associated	6	6
LSS092	Original training	Non-disease associated	NT	9
LSS094	Original training	Non-disease associated	4	4
LSS095	Original training	Non-disease associated	1	11
LSS096	Original training	Non-disease associated	1	11
SS027	Original training	Non-disease associated	NT	NT

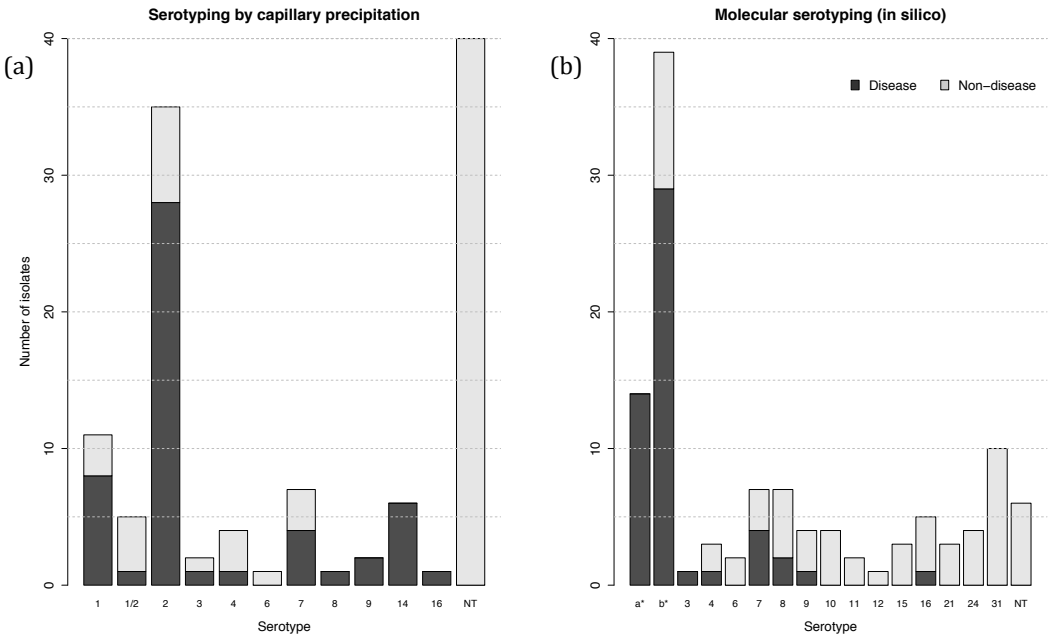

Figure 3.1. Cumulative frequency of serotypes assigned to *Streptococcus suis* isolates in the training collection using traditional capillary precipitation and *in silico* approaches.

Cumulative frequency of the originally described *S. suis* serotypes (1-34 and 1/2) assigned to isolates in the training collection (n=115). Serotyping was performed using two techniques: (**a**) 'gold-standard' capillary precipitation at the Animal and Plant Health Agency (APHA) and (**b**) using adaptation, for *in silico* use, of the multiplex-PCR assays described by Liu *et al.* [74]. The Liu *et al.* molecular 'serotyping' method, the most recent at the time of undertaking this work, is unable to differentiate all reported *S. suis* serotypes due to closely related *cps* genes. Groups "a*" represent serotypes 1 and 14 and "b*" represent serotypes 2 and 1/2. Isolates that could not be assigned to one of the 35 originally described *S. suis* serotypes were designated non-serotypable (NT). Colour has been used to represent the observed clinical phenotypes i) invasive disease-associated (black) and ii) asymptomatic carriage on the palatine tonsils of pigs on UK farms (grey-scale).

3.2.2 Investigation of published putative "virulence-factor" prevalence in the training collection

A wide range of homologs of bacterial virulence factors and virulence-associated factors found in other Gram-positive organisms has been shown to affect the virulence of *S. suis* strains through targeted mutagenesis studies [299-301]. Therefore, investigation into the prevalence of published putative "virulence-factors", extracted from a previous comprehensive review by Fittipaldi *et al*. [129], was another obvious requirement for this PhD project. To do this, the protein-encoding sequence of 71 putative "virulence-factors" (Table 1.3) was extracted from GenBank, and then used as tBLASTn queries against a bespoke BLAST database consisting of the draft genome assemblies of all isolates described in this thesis. Figure 3.2 shows the proportion of invasive disease-associated isolates (n=53) and the proportion of non-disease associated (n=62) isolates in the training collection with matches to each of 31 putative "virulence-factors" investigated. Originally a total of 71 protein-encoding sequences were extracted from GenBank and their prevalence in the training collection investigated. However, over half (n=40, of a total 71; 56%) were either i) detected in the *S. suis* core-genome (i.e. prevalent in all isolates; n=38) or ii) not detected by our methods in any of the 115 isolates of the training collection (n=2; data not shown). The Chi-squared test for independence was used to test for statistically significant (p-value <0.05) associations between the detection of protein-encoding sequences (genotype) and observed clinical phenotype, identifying 14 statistically significant associations (indicated with an asterisk) between published putative "virulence-factors" and the invasive disease-associated phenotype. Of note, the 'virulence-associated markers' (rather than virulence factors *per se*) extracellular protein factor, muramidase-related protein, as well as, the thiol-activated toxin hemolysin, known as Suilysin were amongst the 14 published factors with statistically significant association with the invasive disease phenotype; each of which has been extensively used to characterise and predict the virulence potential of *S. suis* strains in certain, mainly European (UK, Austria, Germany, and Spain), countries[129, 408, 409]. In fact, further analyses revealed, independent of one another, *epf*, *mrp* and *sly* were overrepresented in invasive disease-associated and underrepresented in non-disease associated isolates.

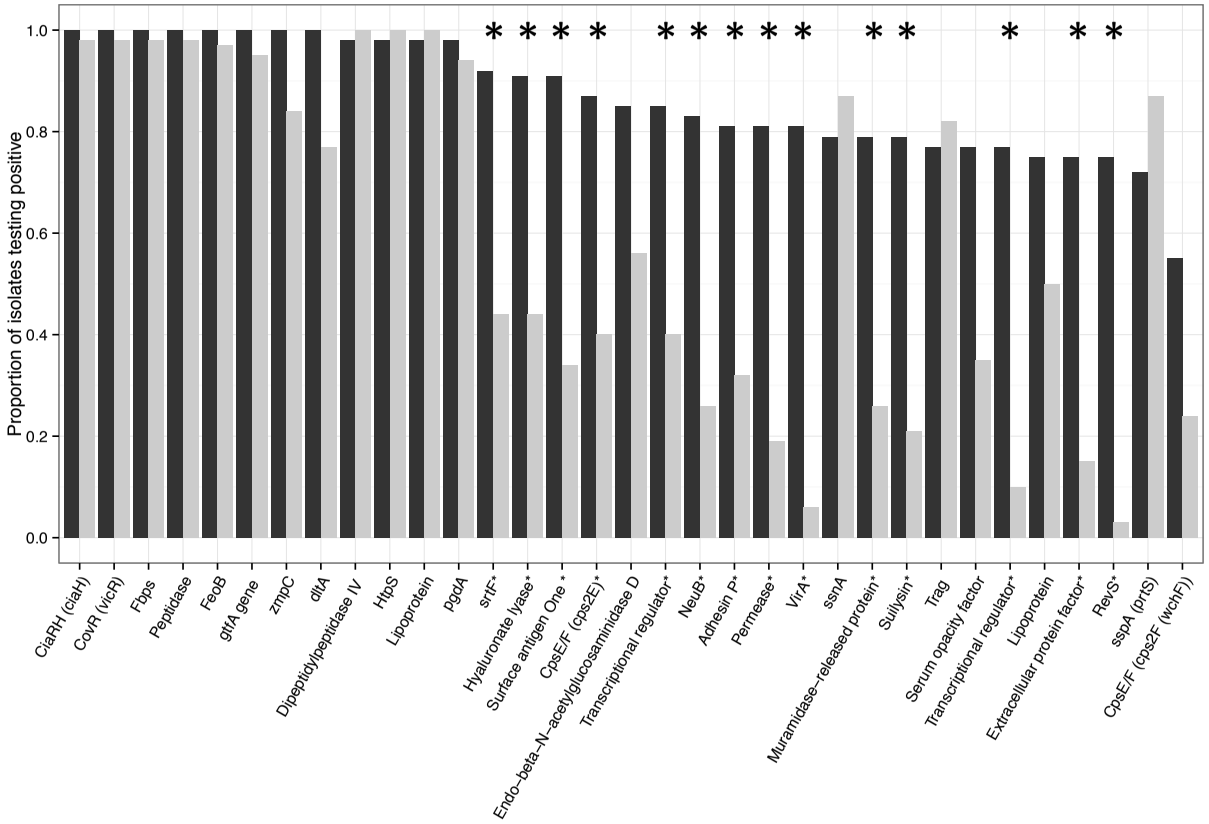

Figure 3.2. Proportion of isolates in the training collection testing positive for published putative "virulence-factors" associated with *Streptococcus suis*.

Investigation into the prevalence of published putative "virulence-factors", extracted from the previous comprehensive review by Fittipaldi *et al.* [129]. The training collection consists of 53 invasive disease-associated isolates and 62 non-disease associated isolates. Detection of protein-encoding sequences was performed *in silico*, where virulence-factors were used as BLASTn queries against a bespoke BLAST database consisting of the draft genome assemblies of all isolates in the training collection. The Chi-squared test for independence was used to identify statistically significant (p-value <0.05) association between observed clinical phenotype. Colour has been used to represent the observed clinical phenotypes i) invasive disease-associated (black) and ii) asymptomatic carriage on the palatine tonsils of pigs on UK farms (grey-scale).

3.2.3 Identification of genetic markers in the *Streptococcus suis* accessory-genome associated with observed clinical phenotype

The identification of genetic markers associated with observed clinical phenotype was achieved using prevalence data for 7261 protein-encoding sequences deemed to be in the *S. suis* accessory-genome (i.e. genes absent from one or more isolates or unique to a given isolate). A combination of three complementary approaches was used to define a preliminary shortlist of markers, that included i) 7261 independent Chi-squared tests for independence were used to identify statistically significant associations between genotype (protein-encoding sequences) and observed clinical phenotype (n=383 identified), ii) a consideration of any impact bacterial population structure had on findings using the Cochran-Mantel-Haenszel Chi-squared test for count data and iii) the multivariate DAPC (n=221). That resulted in the identification of 497 unique candidate genetic markers to pathotype *S. suis*.

3.2.3.1 The Chi-squared test for independence identified 383 statistically significant associations between genotype and phenotype

The Chi-squared test for independence was used to identify statistically significant associations between protein-encoding sequence prevalence (genotype) and observed clinical phenotype. Assuming genotype and phenotype to be independent, the Chi-squared test compared observed frequencies with expected, and Bonferroni adjustment (0.05/7261) was used to control for family-wise error associated with multiple sampling. Figure 3.3 shows a visual representation of the -log10-transformed *p*-value for 7261 independent Chi-squared tests for independence, of which 383 were deemed statistically significant (above the horizontal dotted line). Further analyses of the protein-encoding sequences deemed to be associated with observed clinical phenotype identified 14 (filled red) to correspond to the published putative "virulence-factors" extracted from a previous comprehensive review by Fittipaldi *et al*. [129] (information that was layered onto Figure 3.2). Of the 383 statistically significant associations identified, a total of 257 protein-encoding sequences were found to be over represented in disease-associated isolates

(>50% of disease-associated and <50% of non-disease associated isolates), and vice versa 47 protein-encoding sequences were found to be over represented in non-disease associated isolates (<50% of disease associated and >50% of non-disease-associated isolates).

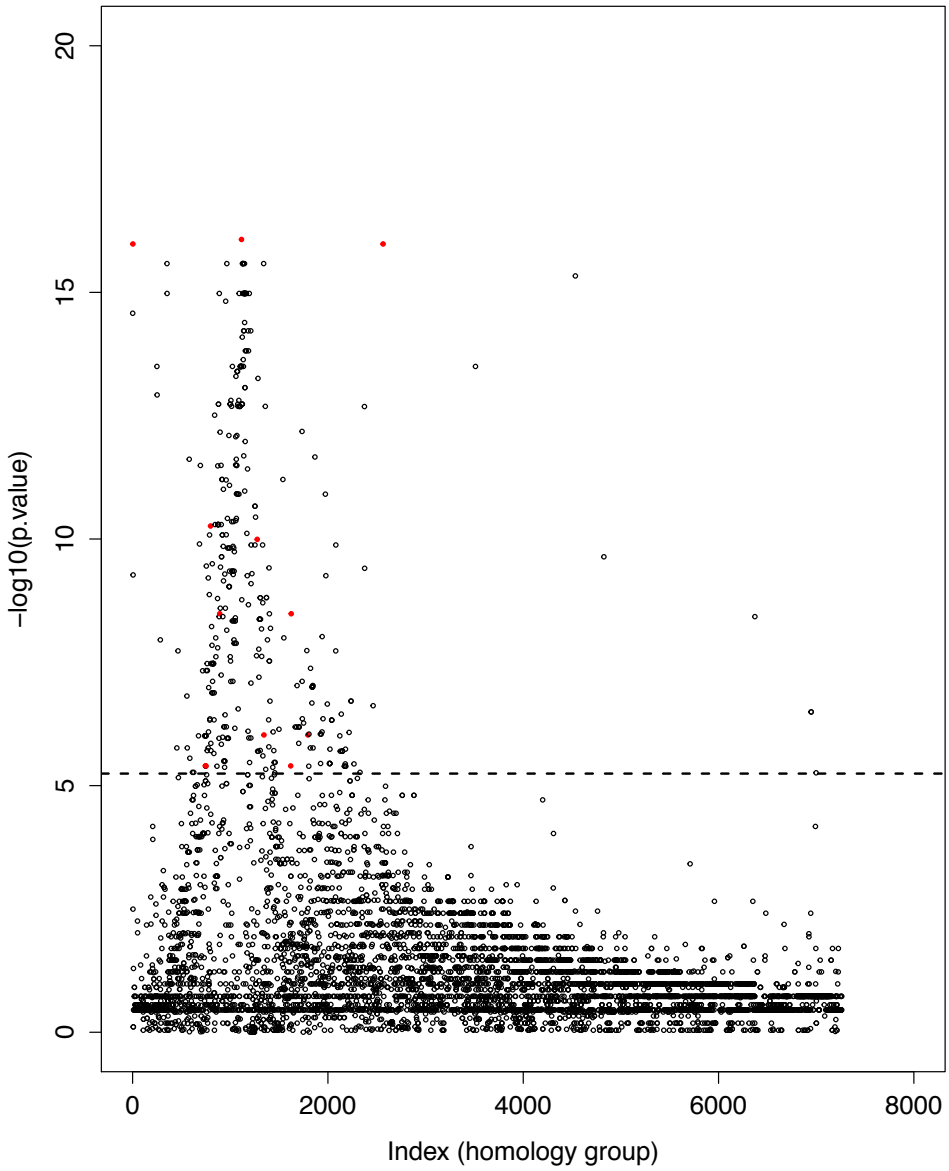

Figure 3.3. Visual representation of the *p*-value for 7261 Chi-squared tests for independence.

Independent Chi-squared tests for independence were used to identify statistically significant associations between genotype (detection of protein-encoding sequences) and observed clinical phenotype : i) invasive disease and ii) asymptomatic carriage on the palatine tonsils of pigs on UK farms. The dotted horizontal line indicates the *p*-value significance threshold after Bonferroni adjustment (α/n, used to control for family-wise error associated with multiple sampling). Data points filled red (n=14) represent statistically significant *p*-values corresponding to published putative "virulence-factors", extracted from the previous comprehensive review by Fittipaldi *et al.* [129].

3.2.3.2 The Cochran-Mantel-Haenszel Chi-squared test for count data deemed no statistically significant associations exist between genotype and observed clinical phenotype

The CMH Chi-squared test for count data was used to minimise the number of spurious associations due to an underlying bacterial population structure. The Chi-squared model compares observed frequencies with expected, but does not take into consideration bacterial population structure. Population groups defined by Bayesian Analysis of Population Structure (BAPS) have previously been described for *S. suis* by Weinert *et al.* [141] (summarised in section 2.1.6). Here the CMH Chi-squared test for count data generated a test statistic that across all five BAPS groups (previously described for the study population used in this thesis) was the odds of a protein-encoding sequence being present in invasive disease-associated isolates compared with non-disease associated isolates. Independent CMH Chi-squared test for count data were performed on the 383 protein-encoding sequences for which the detection of was deemed to be associated with observed clinical phenotype using the Chi-squared test for independence. However, after using Bonferroni adjustment to control for family-wise error associated with preforming multiple comparisons no p-values were below the significance threshold (0.05/383) to reject the null hypothesis of independence.

Statistical inference logic is based on rejection of a null hypothesis if the likelihood of observed data is low. The problem of multiplicity arises from the fact that as the number of independent tests increases so does the likelihood of a rare event, and with it the likelihood of incorrectly rejecting the null hypothesis (Type I error). The classic approach to the multiple comparison problem is to control the familywise error rate, and the most common method is with a Bonferroni adjustment (α/n). However, a common criticism of the Bonferroni correction is that it can be somewhat conservative for a large number of tests, at the cost of increasing the probability of producing false negatives (Type II error). Without Bonferroni adjustment, the CMH Chi-squared test for count data identified nine protein-encoding sequences deemed to be associated with observed clinical phenotype (p-values <0.05; summarised in Table 3.2).

However, close investigation of the length and predicted biological function of these nine protein-encoding sequences revealed none to be suitable targets for a mPCR pathotyping assay/tool. Based on the mean sequence length seven protein-encoding sequences were considered inappropriate as potential pathotyping markers because they were less than 500 bp in length (based on the manufacturer's recommendations for primer design using the Qiagen Multiplex PCR *Plus* kit). The two other protein-encoding sequences were also ruled out as potential pathotyping markers based on the predicted biological function of the protein-encoding sequences as transposases, based on Prokka annotations

Table 3.2. Protein-encoding sequences deemed to be associated with observed clinical phenotype based on the output of the Cochran-Mantel-Haenszel Chi-squared test for count data.

Independent Cochran-Mantel-Haenszel Chi-squared test for count data were performed on the 383 protein-encoding sequences for which the detection of was deemed to be associated with observed clinical phenotype using the Chi-squared test for independence. Close inspection of the length and predicted biological function of these nine protein-encoding sequences revealed none to be suitable targets for a multiplex-PCR pathotyping assay/tool. Grey-scale has been used to indicate the reason why protein-encoding sequences were rejected as potential genetic markers for pathotyping *S. suis*.

OrthoMCL identifier	*p*-value	Proportion of disease-associated isolates testing positive	Proportion of non-disease associated isolates testing positive	Sequence length (bp)	Predicted biological function (Prokka)
0001	0.03	0.83	0.06	939	Transposase
1668	0.01	1.00	0.47	261	Copper sensing transcriptional repressor
1846	0.03	0.81	0.05	204	Hypothetical protein
1950	0.02	0.55	0.02	360	Hypothetical protein
2740	0.03	0.83	0.06	939	Transposase
0002	0.01	0.06	0.47	480	Transposase
2141	0.03	0.06	0.44	150	Hypothetical protein
2145	0.02	0.06	0.58	183	YcfA-like protein
2229	0.01	0.06	0.47	480	Transposase

3.2.3.3 The Discriminant Analysis of Principal Components identified 221 statistically significant genetic differences between phenotypic groups

In addition to univariate methods adjusted for multiple sampling the DAPC [517, 518] was also used to identify genetic differences between pre-defined (disease-associated and non-disease associated groups) phenotypic groups. DAPC is a multivariate genome-wide association approach that identified genetic structures without making assumptions about underlying population genetic models. DAPC is a two-step approach that i) identified genetic variation using a principal components analysis (Figure 3.4 (a)) and ii) then summarised the best discrimination of between group variation with a discriminant analysis (Figure 3.4 (b)). Figure 3.4 (a) shows the first two principal components of the data. Principal components are ordered so that the first few represent as much of the variation in the original dataset as possible. PCA showed 81% (n=43) of disease-associated isolates (red) form a distinct cluster based on information held in the first principal component (X-axis), indicating differences in the positive detection of protein-encoding sequences are present between phenotypic groups. Figure 3.4 (b) shows the Discriminant Analysis (of the Principal Component Analysis) that uses data transformation of the PCA to achieve the best discrimination of isolates into pre-defined (phenotypic) groups. Clear separation between invasive disease-associated (red) and non-disease associated (grey) phenotypic groups, despite a small number of exceptions, indicated differences in the prevalence of protein-encoding sequences in the accessory-genome could be responsible for invasive *S. suis* disease. Importantly, the total amount of original variation retained in the DAPC model affected which genetic markers contributed most to the separation of genetic structures in the training collection. As a result, to assess the impact the total amount of original variation retained had on the DAPC model four independent DAPC analyses were performed retaining 60, 70, 80 or 90% of the original genetic variation. The 1% of genetic markers ranked (highest-to-lowest) based on their contribution to the first principal component of the DAPC was then extracted (n=221; Table 3.3) and genetic markers consistently contributing most to the discrimination of the pre-defined phenotypic groups output by two or more DAPC analyses were taken forward as candidates for pathotyping *S. suis* (n=107).

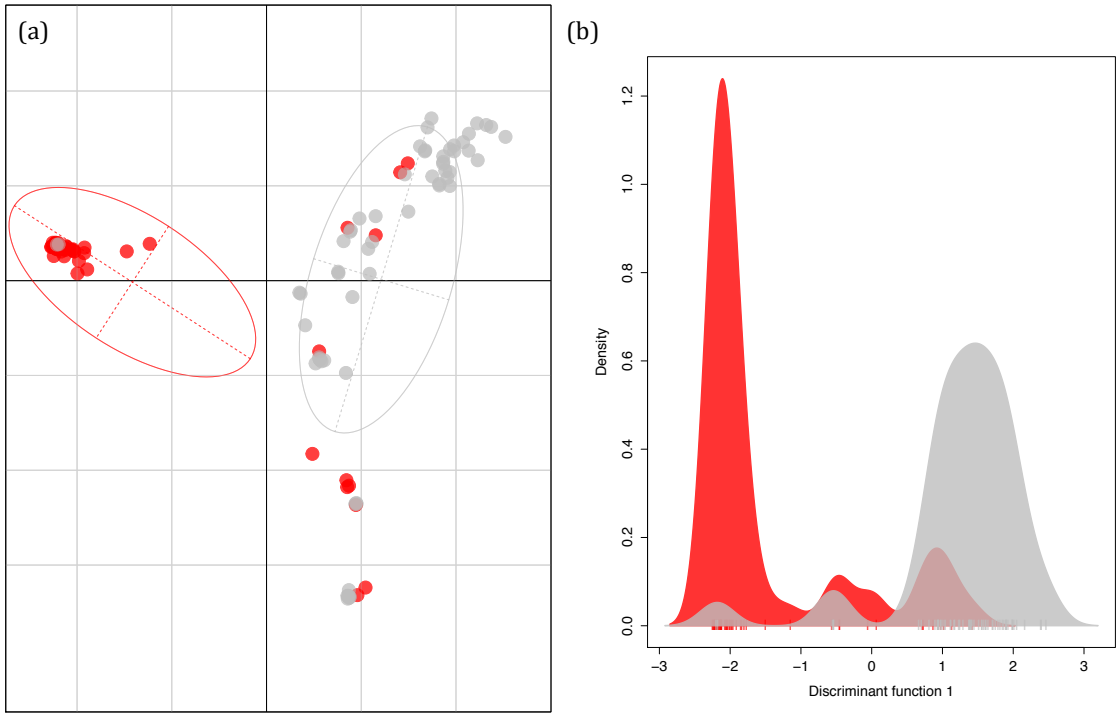

Figure 3.4. Visualisation of genomic differences identified between invasive disease-associated and non-disease associated isolates of *Streptococcus suis* in the training collection.

Discriminant Analysis of Principal Components (DAPC) applied to prevalence data for protein-encoding sequences (n=7621) in the accessory-genome of *S. suis* isolates in the training collection. DAPC is a two-step process that a) identified genetic variation using a principal components analysis (PCA) and then b) summarised the best discrimination of between group variation with a discriminant analysis. PCA showed 81% (n=43) of invasive disease-associated isolates (red) form a distinct cluster based on information held in the first principal component (X-axis). Individual isolates are represented by filled circles coloured by phenotype (invasive disease-associated (red) and non-disease associated (grey-scale)), and positioned on the plane using their values for two variables, principal component 1 (X-axis) and principal component 2 (Y-axis). Discriminant Analysis (of Principal Component Analysis) that uses the data transformation of the PCA to achieve the best discrimination of isolates into pre-defined (phenotypic) groups. Clear separation between invasive disease-associated (red) and non-disease associated (grey-scale) phenotypic groups indicated differences in the prevalence of protein-encoding sequences in the accessory-genome could be responsible for invasive *S. suis* disease.

Table 3.3. The 1% of ranked protein-encoding sequences contributing most to the discrimination of pre-defined phenotypic groups using a Discriminant Analysis of Principal Components.

The total amount of original variation retained in the DAPC model affected which genetic markers contributed most to the separation of genetic structures in the training collection. To assess the impact the total amount of original variation retained had on the DAPC model, four independent DAPC analyses were performed retaining 60, 70, 80 or 90% of the original genetic variation. The 1% of genetic markers ranked (highest-to-lowest) based on their contribution to the first principal component of the DAPC was then extracted and genetic markers consistently contributing most to the discrimination of the pre-defined phenotypic groups output by two or more DAPC analyses were taken forward as candidates for pathotyping *S. suis*.

n	OrthoMCL identifier	60% variation retained	70% variation retained	80% variation retained	90% variation retained
1	1_a	1	1	1	0
2	1_b	1	1	1	1
3	10_a	0	0	1	0
4	1263	0	0	0	1
5	1282_a	0	0	0	1
6	1342_a	1	0	0	0
7	1342_b	1	0	0	0
8	1375	0	0	0	1
9	1406_a	1	1	1	0
10	1406_b	1	1	1	0
11	1416	0	0	0	1
12	1529	0	0	0	1
13	1532_b	1	1	0	1
14	1541	0	0	0	1
15	1542	0	0	0	1
16	1572	0	0	0	1
17	1586	0	0	1	0
18	1630	0	0	1	1
19	1631	1	1	0	1
20	1636	1	1	1	1
21	1642	0	1	0	0
22	1643	0	1	1	1
23	1647	0	0	1	0
24	1648	0	0	1	0
25	1655	1	1	1	1
26	1668	0	0	1	1
27	1687_a	1	1	1	0
28	1688_b	0	1	1	0
29	1692	0	0	1	0

30	1702	0	1	1	0
31	1705	1	1	1	0
32	1706	1	1	1	0
33	1707	1	1	1	0
34	1708	1	1	1	0
35	1709	1	1	1	0
36	1710	1	1	1	0
37	1711_a	0	0	1	0
38	1712	0	1	0	0
39	1713_a	1	1	1	0
40	1715_b	1	1	1	0
41	1716	0	1	0	0
42	1717	1	1	1	0
43	171	1	1	1	0
44	1720	1	1	1	0
45	1721	1	1	1	1
46	1722	1	1	1	0
47	1724	1	1	0	0
48	1725	1	1	0	0
49	1726	1	1	1	0
50	1729	1	1	1	0
51	1730	0	0	1	0
52	1731	0	0	1	0
53	1733	1	1	1	0
54	1740	0	0	0	1
55	1746	1	1	0	0
56	1749	0	0	1	0
57	1750	1	0	0	0
58	1751_b	1	1	0	0
59	1752_b	1	1	1	0
60	1753	0	0	1	0
61	1758	1	1	1	1
62	1764	0	0	1	0
63	1773	0	0	1	0
64	1776	1	1	1	0
65	1779	0	0	1	0
66	1782_b	1	1	0	0
67	1784_b	1	1	1	0
68	1789_a	1	1	0	0
69	1796	1	0	1	0
70	18_a	0	1	0	0
71	1806	0	1	0	0
72	1807	0	0	1	0
73	1808	0	0	1	0
74	1809	1	1	0	0
75	1811	1	1	0	1
76	1818	1	1	1	0
77	1819	0	1	0	0

78	1820	1	1	0	0
79	1821	1	1	1	0
80	1823	1	0	0	0
81	1827	1	0	1	0
82	1828	0	1	0	0
83	1829_b	1	1	1	0
84	1846	1	1	1	1
85	1850	1	1	1	0
86	1852	1	1	1	0
87	1853	1	1	1	0
88	1854	1	1	0	0
89	1855	1	1	1	0
90	1858	1	1	1	0
91	1859	1	1	1	1
92	1860	1	1	0	0
93	1861	1	1	0	0
94	1862	1	1	0	0
95	1863	1	0	0	0
96	1864	1	1	1	0
97	1865	1	1	0	0
98	1866	1	1	1	0
99	1867	1	1	1	0
100	1868	1	1	1	0
101	1869	1	1	1	0
102	1870	1	1	1	0
103	1871	1	1	1	0
104	1872	1	1	0	0
105	1874	1	1	1	0
106	1875	1	1	0	0
107	1876	1	1	0	0
108	1877	1	1	1	0
109	1878	1	0	0	0
110	1879	1	1	1	0
111	1880	1	1	0	0
112	1881	1	1	1	0
113	1883	1	1	1	0
114	1885	1	1	0	0
115	1886	1	1	1	0
116	1893	0	0	1	0
117	1894	0	0	1	0
118	1895_b	0	0	1	0
119	1896_a	1	1	0	0
120	1898	1	1	0	0
121	1900_a	0	0	1	1
122	1902	1	1	1	0
123	1905	0	0	1	0
124	1907	0	0	1	0
125	1911	1	1	1	0

126	1912	0	0	1	1
127	1914	0	0	1	1
128	1915	0	0	1	1
129	1950	0	0	0	1
130	1955	1	0	0	0
131	1972	0	1	0	1
132	1982	1	1	0	0
133	1996	1	1	1	1
134	1999_a	0	0	0	1
135	2_a	1	1	1	0
136	2_c	0	0	0	1
137	2000	1	0	0	0
138	2003	0	1	0	1
139	2004_b	0	1	0	0
140	2007	0	1	0	0
141	2012	0	0	0	1
142	2015	0	0	1	0
143	2028	1	1	1	0
144	2031	0	0	1	0
145	2039	0	0	0	1
146	2040	0	1	1	0
147	2042_a	0	0	0	1
148	2056_a	1	0	0	0
149	2073	0	0	0	1
150	2074_b	0	0	0	1
151	2096	0	1	0	0
152	2111_b	0	0	0	1
153	2120	1	0	0	0
154	2121	1	0	0	0
155	2123	0	0	0	1
156	2125	0	0	0	1
157	2126	0	0	0	1
158	2127	0	0	0	1
159	2128	0	0	0	1
160	2131	0	0	0	1
161	2137	0	0	0	1
162	2141	0	0	0	1
163	2145	0	0	1	1
164	2148	1	0	0	0
165	2174	0	0	0	1
166	2195	1	0	0	0
167	2229	0	0	0	1
168	2258	0	0	1	1
169	2259	0	0	1	1
170	2260	0	0	1	1
171	2261	0	0	1	1
172	2262	0	0	1	1
173	2263	0	0	1	1

174	2264	0	0	1	1
175	2289	0	0	0	1
176	2295	0	0	1	0
177	2297_b	0	0	0	1
178	2305	0	0	0	1
179	2312	0	0	0	1
180	2318	0	0	0	1
181	2329	0	0	0	1
182	2356	0	0	0	1
183	2417	0	0	1	1
184	2445	0	0	1	0
185	2456	0	0	0	1
186	2482	0	0	0	1
187	2555_b	0	0	0	1
188	26_a	1	1	1	1
189	2602	0	0	1	0
190	2621	0	0	0	1
191	2676	0	0	1	0
192	2740	1	1	1	1
193	2749	0	0	0	1
194	2776	0	0	0	1
195	2851	0	0	0	1
196	2932	0	0	0	1
197	2996	0	0	0	1
198	3222	0	0	0	1
199	3423	0	0	0	1
200	3432	1	0	0	0
201	3614_b	0	0	0	1
202	3615_b	0	0	0	1
203	37_b	0	0	0	1
204	3744	0	0	0	1
205	3749	0	0	0	1
206	4072	0	0	0	1
207	4258	0	0	0	1
208	43_b	1	1	1	1
209	4547	1	1	0	0
210	6_b	0	0	0	1
211	6053	0	0	0	1
212	7507	0	0	0	1
213	7508	0	0	0	1
214	7510	0	0	0	1
215	7511	0	0	0	1
216	7512	0	0	0	1
217	7513	0	0	0	1
218	7514	0	0	0	1
219	7515	0	0	0	1
220	7516	0	0	0	1
221	7517	0	0	0	1

3.2.4 Reduction of the preliminary list of genetic markers associated with observed clinical phenotype to a suitable number for logistic regression

A preliminary list of 497 genetic markers with the potential to pathotype S. suis was created by combining the output of i) the Chi-squared test for independence (n=383) and ii) the DAPC (n=221; total n=497 unique protein-encoding sequences). To reduce the preliminary list of genetic markers associated with observed clinical phenotype to a suitable number for logistic regression analysis a multi-step process was followed. Here genetic markers were retained only if i) detected in >50% of invasive disease-associated and <50% of non-disease associated isolates (and vice versa <50% of invasive disease-associated and <50% of non-disease associated isolates; n=88 remaining), ii) protein-encoding sequence length was >500 bp (based on the manufacturer's recommendations for primer design using the Qiagen Multiplex PCR *Plus* kit; n=44 remaining), and iii) not predicted to be a mobile genetic element, such as a phage gene, integrase or transposon (based on Prokka annotations; n=14 remaining). Table 3.4 summarises the fourteen potential molecular markers, shortlisted using the above criteria, for logistic regression analysis.

Table 3.4. Final shortlist of potential genetic markers to pathotype *Streptococcus suis*.

OrthoMCL identifier	Proportion of disease-associated isolates testing positive (n=53)	Proportion of non-disease associated isolates testing positive (n=62)	Sequence length (bp)	Predicted biological function (Prokka)
0043	0.81	0.06	1210	Type I restriction-modification system S protein (virA)
1655	1.00	0.44	1655	Pyridine nucleotide-disulfide oxidoreductase
1708	0.92	0.32	759	Binding protein-dependant transport system protein
1720	0.91	0.26	837	AraC family transcriptional regulator
1721	0.91	0.42	699	GNAT family acetyltransferase
1726	0.92	0.29	1017	Replication initiation factor
1733	0.85	0.40	864	Transcriptional regulator
1796	0.74	0.15	1350	Deoxyguanosinetriphosphate triphosphohydrolase
1821	0.77	0.10	1464	Type I restriction-modification system M protein
1844	0.75	0.06	786	Hydroxyethylthiazole kinase
4547	0.91	0.32	959	Cation transport ATPase
2003	0.19	0.52	693	Hypothetical protein
2028	0.11	0.65	1221	Pectinacetylesterase
2261	0.02	0.55	1701	Putative sugar ABC transporter substrate-binding protein

3.2.5 Identification of the most conserved *Streptococcus suis* species-specific genetic markers to complement the pathotyping markers

The most conserved protein-encoding sequences of the *S. suis* core-genome (i.e. prevalent in all isolates) were used to select a species-specific marker to complement the pathotyping markers. To do this, all annotated protein-encoding sequences of isolate P1/7 (GenBank accession: AM946016) [477, 510], were used as BLASTn queries against a bespoke BLAST database of all draft genome assemblies and known published complete genome sequences for *S. suis*. Protein-encoding sequences with identities >95% across >80% of the total length of each query were then used to query the NCBI non-redundant (nr) database to identify matches only to *S. suis*. Table 3.5 summarises the three most conserved protein-encoding sequences of the *S. suis* core-genome that have a minimum sequence identity in excess of 98% across >80% of the open reading frame. The genetic markers listed in Table 3.5 are not ranked, but instead ordered numerically based on their GenBank identifier.

Table 3.5. Final shortlist of protein-encoding sequences deemed potential genetic markers of the *Streptococcus suis* species.

GenBank identifier	Sequence length (bp)	Minimum sequence identity	Predicted biological function (GenBank)	Predicted biological function (InterPro)
SSU0220	312	98.08	Conserved hypothetical protein	Colicin V production CvpA
SSU0323	675	98.07	Putative membrane protein	Aldolase-type TIM barrel
SSU0577	918	98.15	Conserved hypothetical protein	Sporulation regulator WhiA

3.3 Discussion

This chapter describes the identification of genetic markers in the *S. suis* accessory-genome (i.e. genes absent from one or more isolates or unique to a given isolate) associated with the observed clinical phenotypes i) invasive disease or ii) asymptomatic carriage on the palatine tonsils of pigs on UK farms. Initial analyses of the original training collection were unable to identify any single genetic marker of invasive disease prevalent in >95% of invasive disease-associated isolates and not positively identifiable in <5% of non-disease associated isolates. To avoid restricting our analyses to previously published reports and not taking full advantage of the statistical power of our WGS data set, we used three complementary genome-wide association studies and then a multi-step process to reduce the number of preliminarily listed genetic markers associated with observed clinical phenotype to a number suitable for logistic regression analysis. In the future, using logistic regression analysis for the final selection of genetic markers to pathotype *S. suis* (described later in chapter 4) will allow for the possibility that multiple genetic markers might best describe the *S. suis* pathotype. Also described in this chapter is the identification of the three most conserved protein-encoding sequences of the *S. suis* core-genome, in order to complement the pathotyping markers in any future mPCR assay. In doing so, an evaluation of serotype and bacterial "virulence factors" previously reported [129] to be associated with *S. suis* has been described in this chapter.

Objective 1. Investigation of serotype prevalence revealed, in agreement with previously published observations [41], serotype 2 was the most frequently identified serotype in the training collection regardless of observed clinical phenotype. In addition, serotype 2 was also the serotype most frequently assigned (n=28, of a total 53; 54%) to the invasive disease-causing isolates in the training collection. Serotyping is the most widely used of the methods available to characterise and subtype *S. suis*, and continues to be an important part of the routine diagnostic procedure [26, 39]. However, traditional serotyping techniques are time-consuming, and preparing the reference antisera is not easy due to the high cost and labor associated with its production. Molecular 'serotyping' based on PCR amplification of serotype-specific genes has

been explored as an attractive alternative and/or complement to the existing serological tests [50, 66-75]. Using adaptation (for *in silico* use) of the mPCR molecular 'serotyping' assays described by Liu *et al.* [74] it was possible to assign a molecular serotype to 34 isolates deemed to be non-serotypable using the traditional capillary precipitation technique. These differences between the two approaches to identify the capsular type of isolates could be due to a number of reasons, for example i) user interpretation of the capillary precipitation technique, ii) the diagnostic laboratory (APHA) not keeping the expensive stocks of reference antisera for all 35 serotypes or iii) potential down-regulation of capsule gene expression [41]. In addition, *in silico* analyses assigned a 'serotype' to six isolates that was different to that identified by the 'gold-standard' capillary precipitation technique at the APHA. For example, three non-disease associated isolates were deemed to be serotype 1 by capillary precipitation, but based on *in silico* analyses were assigned to serotypes 11 (n=2) and 12 (Table 3.1). The reason for this disparity between typing techniques is unclear, and is not one of the common cross-reactions (serotype 1 with type 14 antisera, serotype 1/2 with type 1 and 2 antisera, serotype 6 with type 16 antisera (and vice versa) and serotype 22 with type 2 antisera [24]) due to closely related *cps* genes described in the literature [17]. Differences in virulence within the serotypic groups was also evident in this collection of isolates recovered from pigs on farms in England and Wales, in agreement with previous observations in the literature [52-56]. However, most disease-associated isolates belong to relatively few serotypes (1-9, 14 or 1/2) indicating that serotype could be a good indicator and useful proxy for virulence, although there is room for improvement.

Investigation into the prevalence of published putative "virulence-factors", extracted from a previous comprehensive review by Fittipaldi *et al*. [129], did not show a strong relationship with observed clinical phenotype as over half (n=40, of a total 71; 56%) were either i) detected in the *S. suis* core-genome (i.e. prevalent in all isolates; n=38) or ii) not detected by our methods in any of the 115 isolates of the training collection (n=2). The reason for this is unclear, although could be an effect of previous studies being limited to small numbers of isolates often restricted to

serotype 2 [41]. Further analyses revealed 14 of the 31 putative "virulence-factors" whose detection varied between phenotypic groups to have statistically significant association with the invasive-disease phenotype. Of note, the protein-encoding sequences of the 'virulence-associated factors' EF (encoded by the *epf* gene [329]), MRP (encoded by the *mrp* gene) [178]) and SLY (encoded by the *sly* gene) [367, 369]) were amongst the 14 published factors with statistically significant association with the invasive disease phenotype. Closer inspection revealed these protein-encoding sequences (independent of one another) were overrepresented in disease-associated isolates and underrepresented in non-disease associated isolates. Originally associated with *S. suis* disease in the 1990s, to date the exact roles of EF and MRP in virulence remain unclear. In fact, it has been suggested that the synthesis of these proteins might be only coincidently associated with virulence, rather than them being virulence-factors *per se*. However, from an evolutionary perspective this is confusing. The synthesis of these proteins, and secretion (EF) or translocation to the surface of the bacterial cell (MRP), would expend energy and it is difficult to understand why this happens in a high proportion of European isolates associated with disease if not contributing in some capacity to pathogenesis. Regardless, the fact that the practical application of these proteins as genetic markers of invasive disease is limited to certain countries (mainly European, such as the UK, Austria, Germany, and Spain) and not others such as North America [129, 408, 409] indicates that other, potentially more important, factors are responsible for virulence.

Other published putative "virulence-factors", amongst the 31 whose detection varied between phenotypic groups (Figure 3.2), include the protein-encoding sequences *cps2E* and *cps2F* (encoding different glycosyltransferases involved in the synthesis of the serotype 2 CPS [78]). As both feature in Figure 3.2, this could potentially indicate that serotype 2 is an important virulence factor. However, the fact that two genes involved in the synthesis of the serotype 2 CPS appear amongst the 31 protein-encoding sequences should not simply be interpreted as indicating serotype 2 is a more important virulence factor for this UK-biased strain collection. Indeed, the genes required for the biosynthesis of the CPS are clustered on a single locus of the

chromosome [76, 78, 551] and if both genes were detected in the same isolates would be a good example of the effect of bacterial population structure could have on our approach to identify novel genetic markers associated with disease. In this instance, this is not the case and curiously only the *cps2E* gene is deemed to be significantly associated with the invasive disease phenotype, and without further targeted investigation the reason for this is unclear.

The protein-encoding sequences *revS* and *virA*, both significantly associated with the invasive disease phenotype, attract attention in Figure 3.2 because they were detected in a large proportion (>0.75) of invasive disease-associated isolates and a comparatively small proportion (<0.10) of non-disease associated isolates. Identification of such striking differences in prevalence is exciting and might indicate that a molecular approach to diagnosis and surveillance, targeting these genes, could be a very powerful tool if such a distribution is found in other isolate collections. Interestingly, the protein encoded by *revS* is considered to be a response regulator [351], while the protein encoded by *virA* considered to be a Type I restriction modification system specificity domain [378]. However, neither has been the subject of *S. suis* research publications for a number of years [378] until very recently when Willemse *et al*. described the distribution of Type I restriction modification systems in *S. suis* isolates recovered from pigs in the Netherlands [552]. In doing so, Willemse *et al*. proposed that the role of Type I restriction modification systems in *S. suis* is not only in host defence against foreign DNA borne by bacteriophage, but also in stabilising mobile genetic elements or gene regulation [552]. Indeed, switching of specificity subunits in the Type I restriction modification system SpnD39III in *Streptococcus pneumonia* has been shown to be linked to a "switch" from a carriage state to a virulent state [553].

One other published putative "virulence-factor", amongst the 14 with statistically significant association with the invasive disease phenotype, was Adhesin P (Streptococcal adhesin Protein, SadP). Adhesin P is a cell-wall adhesin that can recognise galactosyl-α1-4galactose present on erythrocytes [554]. Adhesin P has also been shown to be involved in *S. suis* adhesion to human

intestinal epithelium [555], as well as, adhesion to monocyte/macrophage immune cells [41]. Indeed SadP has also been proposed to be an important virulence factor involved in the development of meningitis in pigs on farms in the UK as per the "modified Trojan horse theory" [41, 556]. The fact that SadP is a bacterial surface-associated factor makes it an attractive candidate for biotech companies interested in developing tools such as lateral flow devises, akin a pregnancy test, for the surveillance of *S. suis* strains with increased potential to cause invasive disease.

Objective 2. The identification of genetic markers associated with observed clinical phenotype was achieved using a combination of three complementary approaches, resulting in the compilation of a preliminary list of 497 candidates to pathotype *S. suis*. An important caveat of this approach is that the Chi-squared test for independence nor DAPC control for any potential effect of bacterial population structure. Attempts to minimise the number of potentially spurious associations due to any underlying *S. suis* population structure (defined by Bayesian Analysis of Population Structure previously described for *S. suis* by Weinert *et al.* [141]) reduced the number of candidate pathotyping markers to nine, all of which were deemed to be unsuitable to pathotype *S. suis* using a mPCR assay/tool. Therefore, it is important to acknowledge that any association between the 497 potential molecular markers in the preliminary list and phenotype is confounded with bacterial population structure. This may mean that there are different evolutionary routes to pathogenicity in the different BAPS groups (previously described by Weinert *et al.* [141], see section 2.1.6). It may also be the case that the limited number of isolates in this study mean that we lack the power for a GWAS. Unfortunately, given that statistical power will be heavily influenced by amount of recombination, it is not straightforward to perform a power calculation.

Another important caveat of pathotyping tool design is consideration of the observed clinical phenotype associated with each isolate as the 'gold standard' to characterise *S. suis* isolates as disease-associated or non-disease associated. In the absence of an agreed superior approach,

clinical metadata was used to assign *S. suis* isolates to one of these two phenotypic groups and it is acknowledged that such an approach is not perfect and should be treated with caution as not all additional factors can be accounted for, such as i) host-immune status, ii) concurrent infections, or iii) environmental conditions that could influence the susceptibility of a host to *S. suis*-associated disease. Indeed, reports of *in vivo* challenge studies can be readily found in the *S. suis* literature, although most describe data limited to a small number of isolates that is often restricted to serotype 2 [41], and under very different conditions making the extrapolation of findings difficult to interpret. An ideal standard would require an agreed panel of isolates for which a series of consistently controlled experimental infection challenge studies had been undertaken using pigs of identical immune status and genetics. However, in order for this to happen experts in the field must first agree on a suitable model and set of well-defined criteria to score virulence [302-304].

Objective 3. To allow for the possibility that multiple markers might best describe the *S. suis* pathotype, logistic regression analysis in the form of a generalised linear model was chosen for the final selection of genetic markers for pathotyping (described later in chapter 4). As no single protein-encoding sequence had an overbearing contribution to the DAPC it was inferred that a combination of multiple factors with potentially subtle cumulative contributions (similar in nature to diabetes susceptibility genes [557, 558]) are required for invasive *S. suis* disease. This fits with clinical observations that *S. suis* associated disease shows complex dependence on environmental factors and pig immune status. To reduce the preliminary list of genetic markers associated with the invasive disease and non-disease associated phenotypes (n=497) to a number suitable for logistic regression a multi-step process was followed that resulted in a final shortlist of 14 genetic markers for generalised linear modelling (11 associated with the phenotype of invasive disease and three associated with asymptomatic carriage on the palatine tonsils of pigs on UK farms). Curiously, two Type I restriction modification system proteins appear in the shortlist of 11 molecular markers associated with the invasive disease phenotype (one of which, *virA* has been discussed earlier due to its previous association with virulence

[319, 378]). Additional analyses revealed the protein-encoding sequences correspond to two separate Type I restriction modification systems, of which there is a total of three such systems present in *S. suis* [552], each located a considerable distance apart in the genome of isolate P1/7 (GenBank accession: AM946016) [477, 510].

Out of the genetic markers associated with the invasive disease phenotype in Table 3.4, the "Type I restriction modification system S protein (virA)", could arguably be considered the 'best' single molecular marker to pathotype *S. suis*; due to the fact it was detected in a high proportion (0.81) of invasive disease-associated isolates and a comparatively small proportion (0.06) of non-disease associated isolates. Traditionally Type I restriction modification systems are protein complexes associated with defence of the host bacterium against foreign DNA such as that borne by bacteriophages [559]. The reason why a protein-encoding sequence linked with host bacterium defence is associated with the invasive disease-associated phenotype is unclear and could be spurious. Although, as discussed earlier published information on the range of functions that Type I restriction modification systems may have is expanding and proposed to include stabilising mobile genetic elements and gene regulation, potentially providing evolutionary fitness advantages and virulence under certain conditions [560]. For example, fixing plasmids or a prophage in a bacterial chromosome that carries tetracycline resistance genes would be beneficial to that clone in a bacterial population. As a result, it has been proposed that the role of Type I restriction modification systems in protection against foreign DNA may merely be a coincidental benefit of other functions [561].

Objective 4. The three most conserved (minimum sequence identity >98%) protein-encoding sequences in the *S. suis* core-genome were identified and shortlisted in order to select a species-specific marker to complement the pathotyping markers (Table 3.5). The longest in length and most highly conserved of the three shortlisted protein-encoding sequences is SSU0577, features that potentially provide the most flexibility with respect to future mPCR primer design. Annotated as a hypothetical protein by Prokka and in GenBank, SSU0577 was identified by the

Interpro database as sharing sequence similarity with the sporulation regulator *whiA*. Unlike members of the *Bacillus* and *Clostridium*, *S. suis* is considered a non-spore forming coccus, and it is unclear why a protein-encoding sequence associated with sporulation would be such a highly conserved component of the *S. suis* core-genome. It is interesting to speculate that maybe *S. suis* is able to form an endospore and exist in a stripped back dormant state that is revived in the nutrient rich bloodstream or on solid agar plates in the laboratory. However, a common feature of endospores is a tolerance of high temperatures and chemical disinfectants, something that *S. suis* does not demonstrate [127] (see sections 1.1.2.1.3 and 1.1.2.1.6.4). In contrast, it is possible that *S. suis* has lost the ability to form endospores and that the presence of SSU0577 is a remnant of a previously functional operon, although this is probably not true. Indeed, a more likely explanation is that *whiA* is a regulator that has been associated solely with sporulation in Gram-positive bacteria and actually has other roles too.

In conclusion, this chapter has described the work undertaken to identify genetic markers in the *S. suis* accessory genome associated with defined clinical phenotype, and then the compilation of a list of 14 genetic markers with the potential to pathotype *S. suis* isolates (Table 3.4). To allow for the possibility that multiple markers might best describe the *S. suis* pathotype, logistic regression analysis in the form of a generalised linear model was chosen for the final selection of genetic markers for pathotyping – a process that is described in detail in chapter 4.

Chapter 4

Design and evaluation of a multiplex-PCR pathotyping tool for *Streptococcus suis*

4.1 Introduction

Numerous methods exist to characterise and subtype *S. suis* isolates, that have been reviewed in chapter 1 (section 1.1.1). Indeed, simple biochemical tests are capable of species level identification from diseased pigs. However, isolation and identification of *S. suis* isolates from carrier pigs is much more complicated. The fact that most pigs harbour *S. suis* on their tonsils means that detection in asymptomatic carrier herds is often of little practical value in predicting the likelihood of future clinical relevance. Due to this, chapter 3 of this thesis describes the identification of 14 genetic markers, in the *S. suis* accessory-genome, associated with observed clinical phenotype that have the potential to be used as pathotyping markers (Table 3.4).

Serotyping is the most widely used of the methods available to characterise and subtype *S. suis* isolates, and remains an important part of the routine diagnostic procedure [11, 17]. The observation of multiple serovars, as well as, non-serotypable strains present in the same animal is common. In general, *S. suis* serotype 2 predominates among clinical cases of disease in most countries, although serovars 1-9, 14 and 1/2 have been reported to be important serotypes associated with *S. suis* disease in certain geographical locales [42, 60, 393, 409, 562]. Chapter 3 describes a similar serotype distribution for the training collection of 115 *S. suis* isolates recovered from pigs on farms in England and Wales (Table 3.1 and Figure 3.1), where all but one invasive disease-associated isolate were assigned a serotype (or molecular 'serotype') of either 1-9, 14 or 1/2 (the most common serotypes recovered from clinical samples [17]). Historicaly, serotyping has been used as a proxy for virulence, resulting in numerous singleplex and mPCR assays/tools to amplify serotype-specific *cps* genes and assign molecular 'serotypes' to *S. suis* isolates. However, while serotype is still important for the formulation of vaccine strategy against this bacterium, its relationship with virulence has long been discussed but never proven, and differences in virulence not only between but also within the *S. suis* serotypic groups limits the interpretation of epidemiological studies that include only serotyping.[53, 178, 400].

Discussed in chapter 3, a wide range of homologs of virulence factors found in other Gram-positive organisms has been shown to affect the virulence of *S. suis* strains through targeted mutagenesis studies [299-301]. However, clear associations with specific roles in the development of disease have not been found for many proposed factors [129]. Despite this, the 'virulence-associated markers' (rather than virulence factors *per se*) EF and MRP, as well as, the thiol-activated toxin hemolysin SLY have been extensively used to predict the virulence potential of *S. suis* strains in certain, mainly European (UK, Austria, Germany, and Spain), countries [129, 408, 409]. Large and small size variants of *epf* and *mrp* (reviewed in section 1.1.3.1.3) have been described leading to determination of the genotype of these protein-encoding sequences being included in the routine diagnostic typing of *S. suis* in several laboratories [131, 329]. Unfortunately, genotyping *epf, mrp*, and/or *sly* fails to provide clear classification of a *S. suis* isolate as virulent, because isogenic mutants devoid of such factors were found to be as virulent as their respective parental strains [302, 330, 395]. Further to this, the fact that the usefulness of these proteins is limited to certain countries and not all important pig producing countries, such as North America [129, 408, 409], indicates that other potentially more important factors are responsible for virulence.

One of the most recent methods used to characterise and sub-type *S. suis* is minimum core genome (MCG) sequence typing, a method that uses the Bayesian clustering to establish population genetics-based sub-divisions for strain identification and typing [115, 116]. Multilocus sequence typing has become the preferred method for genotyping many biological species, due to its ability to identify major phylogenetic clades, molecular groups, or subpopulations of a species, as well as individual strains or clones. However, MLST can sometimes lack the discriminatory power to differentiate bacterial strains into virulent and avirulent sub-populations, which limits its use in epidemiological investigations. Rather than characterise isolates based on the slowly accumulating variation in the seven core metabolic (housekeeping) genes described for MLST[90], MCG sequence typing is one approach that exploits advances in next-generation sequencing and uses genomics derived data to identify

novel regions of the core-genome that can be used to identify and type *S. suis* into "MCG groups" that can later be associated with clinical phenotypes. MCG sequence typing targets SNPs in the *S. suis* core-genome (Table 1.2), and during its design all highly virulent isolates and epidemic isolates tested were reported to be assigned to MCG group 1 [115].

Ease of use and reliability are important factors that should be considered when designing a novel diagnostic tool. The term "pathotyping" tool is used to describe a diagnostic assay/tool that can predict the potential of an isolate to cause invasive clinical disease, based on the detection (or not) of biomarkers targeted by the assay. This is particularly relevant for *S. suis*, a bacterium that can be found as an asymptomatic commensal-like strains or (primary or secondary) pathogen. Thus, the ability to predict whether an isolate has the potential to cause invasive clinical disease would be useful and highly desirable. Furthermore, the ideal diagnostic tool from the end users' perspective should include i) minimal culture/purification steps, ii) having a high sensitivity and specificity, iii) allowing for detection in a mixed sample and iv) having a short turn-around time for the result.

This chapter describes the design and evaluation of a pathotyping tool (mPCR and generalised linear model) for *S. suis*. Starting with the shortlist of 14 genetic markers associated with observed clinical phenotype compiled in chapter 3 (Table 3.4), this chapter describes the use of logistic regression to select three statistically significant markers for pathotyping. Subsequent work focuses on evaluating the ability of these three genetic markers to predict the potential of *S. suis* isolates in the training collection to cause invasive disease. Evaluation is then extended to include an out-of-sample 'test' collection of 138 previously uncharacterised isolates of *S. suis*, and in doing so comparisons are drawn between the newly-defined pathotyping markers and other popular published methods used to characterise and subtype *S. suis* isolates. Finally, this chapter describes the investigation of the newly-defined pathotyping markers to screen oral fluid as well as swabs of material scraped from the palatine tonsils of pigs without first culturing and isolating single colonies/clones.

4.1.1 Chapter objectives

1. To select the fewest possible genetic markers associated with observed clinical phenotype for inclusion in a mPCR assay/tool that is able to reliably identify invasive disease-associated isolates of *S. suis*
2. To evaluate the ability of any genetic markers selected to predict the potential of isolates to cause invasive disease in comparison to observed clinical phenotype
3. To evaluate the ability of any genetic markers selected to predict the potential of isolates to cause invasive disease in comparison to observed clinical phenotype using an out-of-sample 'test' collection of *S. suis* isolates
4. To evaluate the ability of any genetic markers selected to predict the potential of isolates to cause invasive disease in comparison to popular published methods used to characterise and subtype *S. suis* isolates
5. To determine whether the mPCR component of the pathotyping tool can be used to screen oral fluid and swabs of material scraped from the palatine tonsils of pigs without first culturing and isolating single colonies/clones

4.2 Results

4.2.1 Design of a multiplex-PCR pathotyping tool for *Streptococcus suis*

4.2.1.1 Three genetic markers associated with defined clinical phenotype are required to explain ~95% of the invasive disease-associated phenotype

The complexity of multiplex-PCR design increases as the number of targets increases, and compromises in optimal size differences between amplicons and in primer placement become inevitable. Working with a UK-biased collection a conscious decision of design was to create the simplest mPCR with a minimum number of markers possible so that as genome analyses are expanded to include isolates from other countries additional markers important to other geographical locales can easily be incorporated into our mPCR pathotyping tool. To trade-off simplicity against the ability to reliably identify the disease-associated isolates the minimum number of genetic markers required to explain the invasive disease-associated phenotype was estimated. To do this, the minimum number of genetic markers (shortlisted in chapter 3, see Table 3.4) required to pathotype *S. suis*, the proportion of the invasive disease-associated phenotype explained by increasing the cumulative number of markers was calculated using a bespoke R script written by Post Doctoral Research Associate Dr Lucy Weinert (Department of Veterinary Medicine, University of Cambridge). Figure 4.1 shows 11 of the 14 shortlisted genetic markers are required for the median (a typical value) proportion of the invasive disease-associated phenotype and interquartile range to both equal 1.0. In comparison, seven markers are required for the median to equal 1.0 but not the interquartile range, and as few as three of the shortlisted genetic markers associated with the observed clinical phenotype to be required for the median to explain ~95% of the invasive disease-associated phenotype.

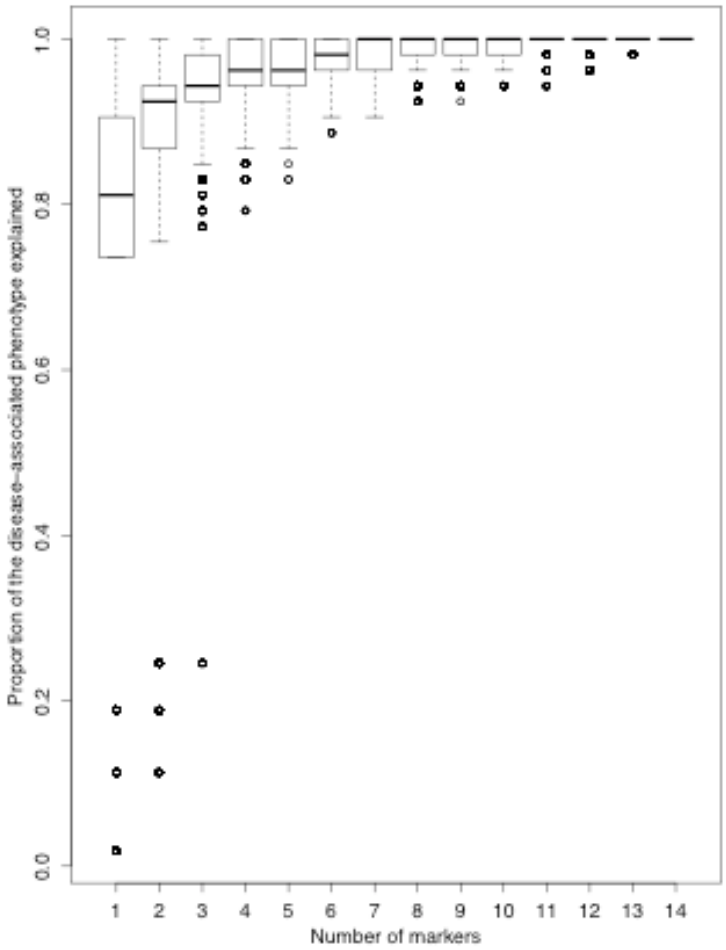

Figure 4.1. Visual representation of estimating the proportion of disease-associated isolates in the training collection captured by increasing the cumulative number of genetic markers.

Estimates were generated by simple random sampling of 1of the 14 shortlisted genetic markers associated with observed clinical phenotype listed in Table 3.4 (chapter 3) and calculating the proportion of the invasive disease-associated isolates in the training collection (n=53) positive for that marker, this process was then repeated for 10,000 random sampling events. To investigate the effect of increasing the cumulative number of genetic markers the number of genetic markers in Table 3.4 randomly sampled from was increased to *x*, where *x* takes a value between 1 and 14, and the proportion of invasive disease-associated isolates in the training collection testing positive for *x* markers calculated; again this process was then repeated for 10,000 random sampling events. The process of increasing the number of candidate markers randomly sampled from by one was repeated until *x*=14. Boxplot, implemented in the R package: *graphics*, was used to plot boxes representative of the 1st quartile, a typical value (median) and 3rd quartile (respectively) for 10,000 random sampling events with whiskers representing 1.5x the interquartile range. Outliers were plotted as unfilled circles.

4.2.1.2 Selection of three genetic markers associated with observed clinical phenotype to pathotype *Streptococcus suis*

Logistic regression analysis in the form of a generalised linear model (GLM) was used for the final selection of genetic markers to pathotype *S. suis*. Logistic regression is a statistical method for the analysis of a dataset in which there are one or more independent explanatory variables (shortlisted pathotyping markers) that are expected to influence the binary dependent variable (the observed clinical phenotype, invasive disease = 1 and non-disease = 0). The first GLM fitted to protein-encoding sequence prevalence data for the training collection included all 14 shortlisted genetic markers described in chapter 3 (Table 3.4) and produced an Akaike information criterion (AIC) = -63.34. The addition of explanatory variables to logistic regression models will always increase statistical validity because addition of another explanatory variable will always explain a little more variance of the log odds. However, adding more and more explanatory variables to a model is inefficient and leads to overfitting (and, as a result, the inclusion of unnecessary primers and amplicons to resolve in our mPCR assay). Therefore, backwards step-wise selection, using penalised likelihood ratio tests, was used to select the fewest statistically significant genetic markers to differentiate invasive disease-associated and non-disease associated phenotypic groups (AIC = -76.84). The AIC is an estimator of the relative quality of a statistical model for a given data set and useful for comparison and selection of models, although, is also a measure that should be treated with caution as it gives no indication of absolute quality of a model. The AIC was used as a statistical sanity check, and the AIC of -76.84 in comparison to -63.34 indicated the trade-off of complexity and goodness of fit of the model favoured the use of the three markers in Table 4.1 (as opposed to all 14 shortlisted candidate genetic markers). Table 4.1 shows the three explanatory variables whose elimination from the GLM could not be done without a statistically significant loss of fit (p-value <0.05) – the newly defined pathotyping markers.

Table 4.1. Genetic markers selected to pathotype *Streptococcus suis*.

The proportion of invasive disease-associated (n=53) and non-disease associated (n=62) isolates testing positive for each genetic marker was calculated using the isolates in the training collection (n=115). GenBank identifier prefixes "SSU" and "SSUST3" correspond to *S. suis* P1/7 [477] and *S. suis* ST3 [478].

OrthoMCL identifier	GenBank identifier	Marker of	Proportion of disease-associated isolates testing positive	Proportion of non-disease associated isolates testing positive	Contribution to the GLM (*p*-value)
0043	SSU1589	Disease-association	0.81	0.06	2.57e-06
2261	SSUST30534	Non-disease association	0.02	0.55	4.37e-02
4547	SSU0207	Disease-association	0.91	0.32	1.23e-02
-	SSU0577	*Streptococcus suis*	1.00	1.00	-

4.2.2 Evaluation of a multiplex-PCR pathotyping tool for *Streptococcus suis*

4.2.2.1 In-sample testing

The mPCR for pathotyping *S. suis* was first evaluated using genomic DNA extracted from each of the 115 isolates of the training collection. For all isolates, the mPCR assay consistently produced amplicons as anticipated based on *in silico* analyses. Figures 4.2 and 4.3 show the mPCR amplicon patterns after gel electrophoresis on a 2% (wt/vol) agarose gel photographed under UV transillumination. Amplicons of size 722 bp correspond to the *S. suis* species-specific marker (SSU0577), and were produced by all isolates of the training collection irrespective of invasive disease-associated/non-disease associated phenotype or genotype. Other amplicons, of size 211 bp (or 190 bp) and 347 bp correspond to the disease-associated markers SSU0207 and SSU1589 respectively (primers for marker SSU0207 produce a 211 bp or 190 bp amplicon due to a 21 bp indel in the protein-encoding sequence, unavoidable during primer design). In addition, amplicons of size 892 bp correspond to the non-disease associated marker SSUST30534.

Figure 4.3 shows the mPCR amplicons for all 63 non-disease associated isolates in the training collection. Based on the observed clinical phenotype, isolates LSS062 and LSS063 were deemed to be non-disease associated as they were recovered from the trachea-bronchus of two six-week old pigs both with ante and post-*mortem* findings of diarrhoea and enterocolitis. However, the mPCR amplicon pattern (as anticipated based on *in silico* analyses) of these two isolates resembles that of the invasive disease-associated isolates of the training collection. Sanger sequencing was used to confirmed mPCR amplicons to be the targeted regions of invasive disease-associated pathotyping markers SSU0207 (211 bp) and SSU1589 (347 bp; data not shown).

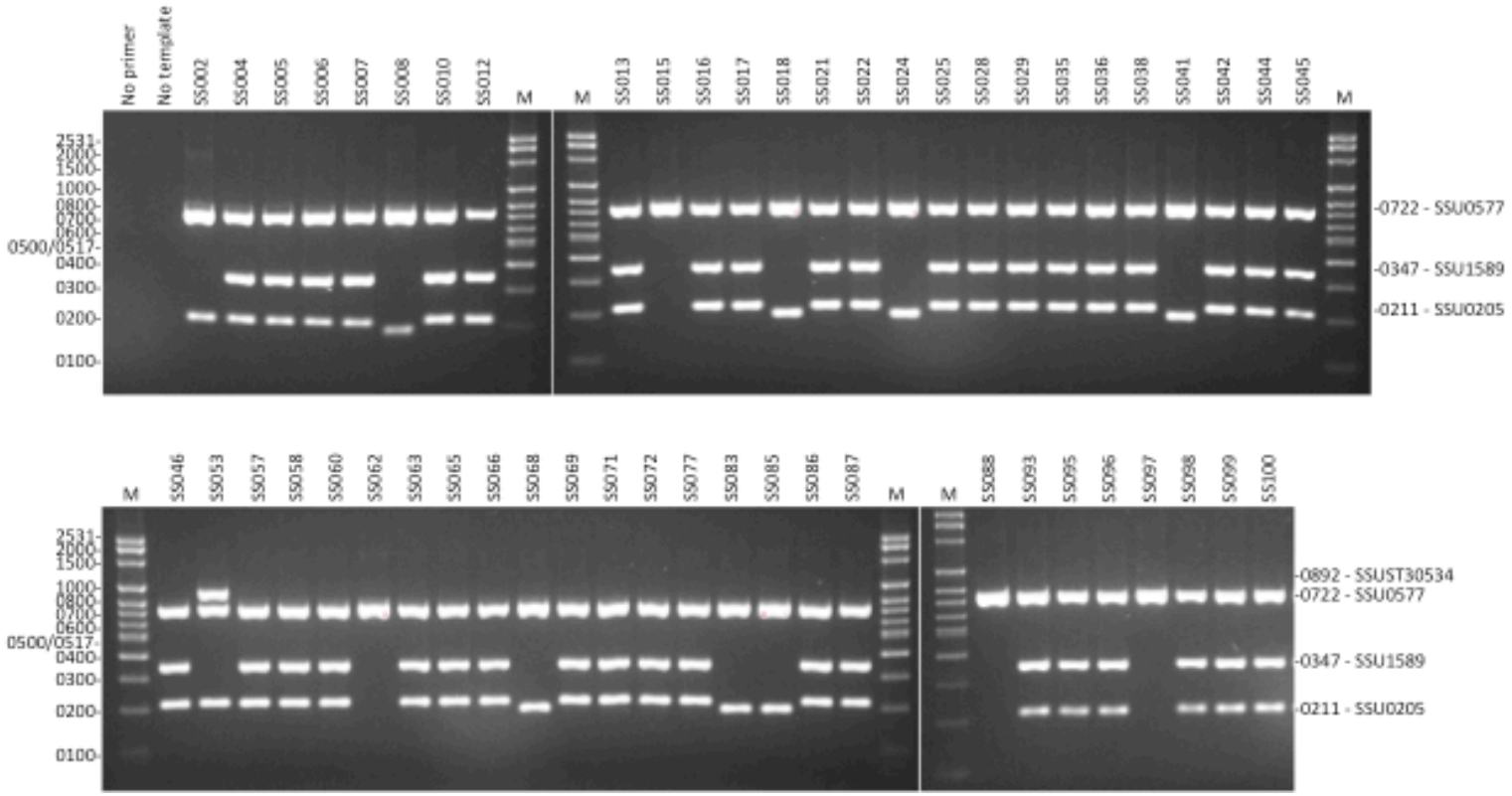

Figure 4.2. Multiplex-PCR amplicons produced from template genomic DNA extracted from all 53 invasive disease-associated isolates in the training collection.

Agarose gel containing mPCR amplicons produced from *S. suis* genomic DNA of all 53 disease-associated isolates in the training collection. PCR amplicons were electrophoresed on a 2% (wt/vol) agarose gel containing 1x SYBR Safe DNA gel stain for 60 min at a constant 100 V, and photographed under UV transillumination. In each instance multiplex-PCR amplicon patterns matched anticipated amplicon patterns based on *in silico* analyses. Isolate names are indicated above lanes. Lanes labelled M contain 1x Bioline HyperLadder 100 bp Plus DNA ladder with sizes indicated on the left. Multiplex-PCR amplicon sizes are indicated on the right. Primers for invasive disease-associated marker SSU0205 produce a 211 bp or 190 bp amplicon, due to a 21 bp indel in the protein-encoding sequence unavoidable during primer design.

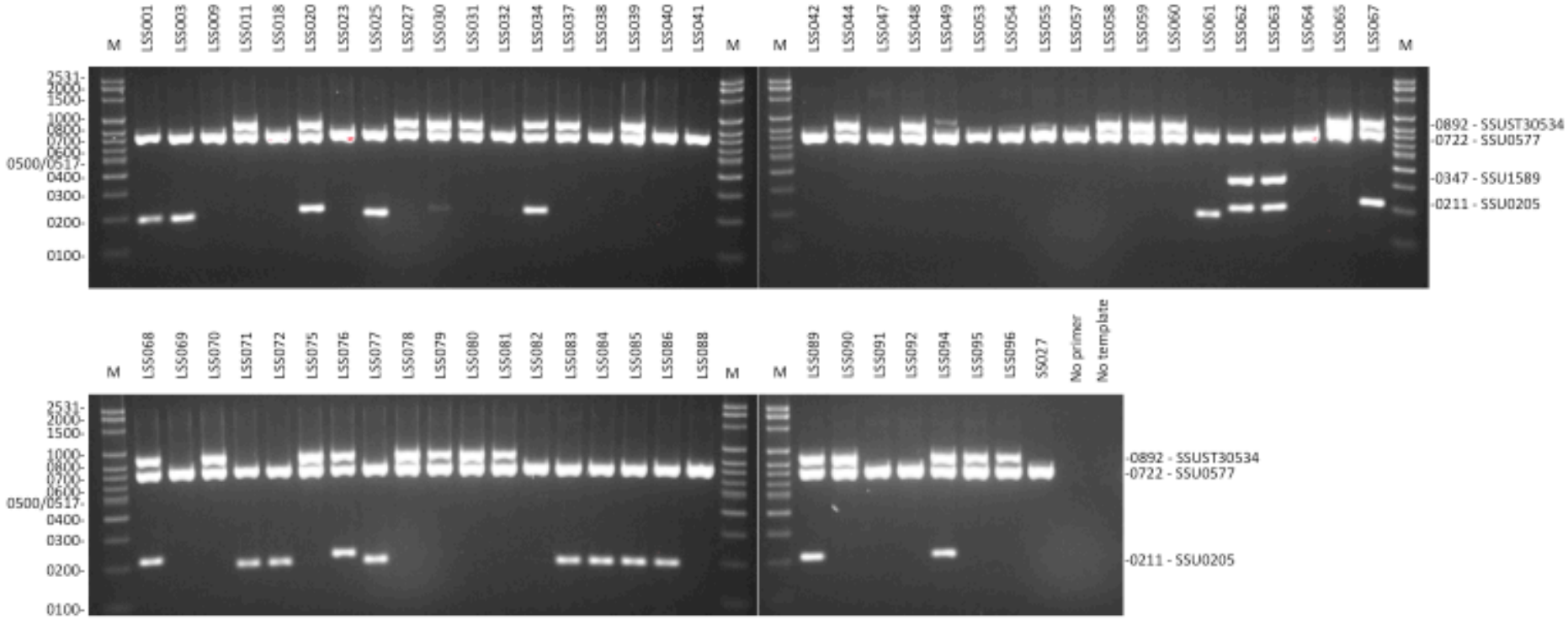

Figure 4.3. Multiplex-PCR amplicons produced from template genomic DNA extracted from all 63 non-disease associated isolates in the training collection.

Agarose gel containing mPCR amplicons produced from *S. suis* genomic DNA of all 62 non-disease associated isolates in the training collection. PCR amplicons were electrophoresed on a 2% (wt/vol) agarose gel containing 1x SYBR Safe DNA gel stain for 60 min at a constant 100 V, and photographed under UV transillumination. In each instance multiplex-PCR amplicon patterns matched anticipated amplicon patterns based on *in silico* analyses. Isolate names are indicated above lanes. Lanes labelled M contain 1x Bioline HyperLadder 100 bp Plus DNA ladder with sizes indicated on the left. Multiplex-PCR amplicon sizes are indicated on the right. Primers for invasive disease-associated marker SSU0205 produce a 211 bp or 190 bp amplicon, due to a 21 bp indel in the protein-encoding sequence unavoidable during primer design.

4.2.2.1.1 Selection of the 0.43 cutoff to convert the fitted values of the generalised linear model into a binary classification decision

The continuous real-valued output (fitted values), of the three explanatory variable GLM fitted to the training collection, was converted into a binary class decision (invasive disease = 1 and non-disease = 0) by selecting a 0.43 cutoff threshold. As no cutoff was optimal according to all possible performance criteria, cutoff choice involved a trade-off between different performance metrics (Table 4.2). Low false negative rate (analogous to Type II error) was chosen as the most valuable performance metric for pathotyping *S. suis*, with a view to establish and maintain a pig population free of invasive disease-associated strains. Figure 4.3.3 is a visual representation of the data presented in Table 4.3, and demonstrates how the choice of cutoff threshold affected the performance metrics false negative rate and F_1 score (a weighted measure of accuracy) for the training collection. To choose the cutoff of 0.43 the closest F_1 score to 1 was traded-off against a false negative rate closest to 0.

Using the 0.43 cutoff threshold the pathotyping tool predicted correct classification, in comparison to the observed clinical metadata, of 84% (accuracy) of the 115 isolates in the original training collection. Table 4.2(iii) is an extended 2x2 contingency table that reports a number of other performance metrics important for pathotyping *S. suis*. Commonly used measures of binary diagnostic tests are the performance metrics i) true positive rate (sensitivity, 0.91), ii) true negative rate (specificity, 0.79), iii) positive predictive value (PPV; precision, 0.79) and iii) negative predictive value (NPV, 0.91). Calculated using the cutoff threshold of 0.43, false negative rate = 0.09 and F_1 score (a measure of accuracy, that is the average of the sensitivity and positive predictive value (precision) of a test) = 0.84.

Table 4.2. Performance metrics calculated for the pathotyping tool when applied to the training collection and compared to the 'gold-standard' observed clinical phenotype.

An extended 2x2 contingency table for calculating the performance metrics of two binary diagnostic tests.

(i) Cutoff: 0.10

		Phenotype			
	115 Total population	53 Phenotype positive	62 Phenotype negative		
mPCR result	93 mPCR result positive	53 True positive	40 False positive	0.57 Positive predictive value	0.43 False discovery rate
	22 mPCR result negative	0 False negative	22 True negative	0.00 False omission rate	1.00 Negative predictive value
		1.00 True positive rate	0.65 False positive rate	0.73 F1 score	
		0.00 False negative rate	0.35 True negative rate		

(ii) Cutoff: 0.15

		Phenotype			
	115 Total population	53 Phenotype positive	62 Phenotype negative		
mPCR result	85 mPCR result positive	52 True positive	33 False positive	0.61 Positive predictive value	0.39 False discovery rate
	30 mPCR result negative	1 False negative	29 True negative	0.03 False omission rate	0.97 Negative predictive value
		0.98 True positive rate	0.53 False positive rate	0.75 F1 score	
		0.02 False negative rate	0.47 True negative rate		

(iii) Cutoff: 0.43

		Phenotype			
	115 Total population	53 Phenotype positive	62 Phenotype negative		
mPCR result	61 mPCR result positive	48 True positive	13 False positive	0.79 Positive predictive value	0.21 False discovery rate
	54 mPCR result negative	5 False negative	49 True negative	0.09 False omission rate	0.91 Negative predictive value
		0.91 True positive rate	0.21 False positive rate	0.84 F1 score	
		0.09 False negative rate	0.79 True negative rate		

(iv) Cutoff: 0.93

		Phenotype			
	115 Total population	53 Phenotype positive	62 Phenotype negative		
mPCR result	42 mPCR result positive	40 True positive	2 False positive	0.95 Positive predictive value	0.05 False discovery rate
	73 mPCR result negative	13 False negative	60 True negative	0.18 False omission rate	0.82 Negative predictive value
		0.75 True positive rate	0.03 False positive rate	0.84 F1 score	
		0.25 False negative rate	0.97 True negative rate		

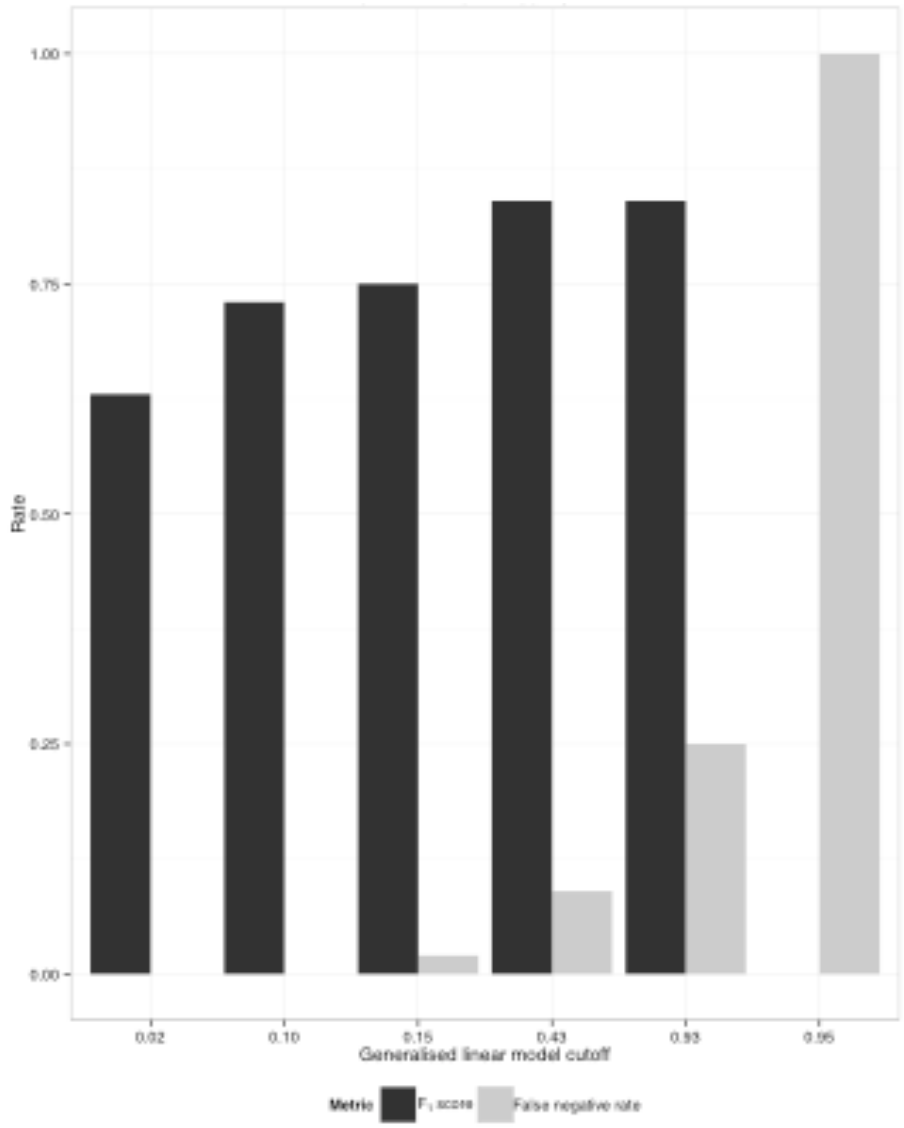

Figure 4.4. Visual representation to view and trade-off the performance metrics false negative rate and F_1 score, at different cutoff thresholds.

The continuous real-valued output (fitted values) of the GLM, fitted to the training collection protein-encoding sequence prevalence data, was converted into a binary class decision (disease-associated (1) or non-disease associated (0)) by selecting a cutoff threshold. In order to select the cutoff threshold of 0.43, the closest F_1 score to 1 was traded-off against a false negative rate closest to 0.

4.2.2.1.2 Estimation of a ~0.0001 ng limit of detection for the multiplex-PCR component of the pathotyping tool for *Streptococcus suis*

To determine the analytical sensitivity of the mPCR assay the approximate limit of detection was estimated from 10-fold serial dilutions of *S. suis* genomic DNA of known concentration. DNA extracted from four isolates (two disease-associated and two non-disease associated) of the training collection was mixed in equal quantities so that templates for each mPCR amplicon would be present in all reactions, and a series of 10-fold dilutions then performed to create mPCR templates decreasing in DNA concentration. Figure 4.5 shows mPCR amplicons following gel electrophoresis, photographed under UV transillumination. The limit of detection was estimated to be ~0.0001 ng of *S. suis* genomic DNA (equivalent to ~45 genome copies), the lowest concentration of template DNA from which all predicted mPCR amplicons, after 35 thermal cycles, were easily visible under UV transillumination.

4.2.2.1.3 Investigation into the specificity for *Streptococcus suis* of the multiplex-PCR component of the pathotyping tool

To evaluate specificity of the mPCR assay for the *S. suis*, field isolates of Streptococcaceae commonly recovered from the upper respiratory tract of pigs on farms in England and Wales were used as a panel of negative controls. Sourced from research project: BB/L003902/1, the collection included isolates of *Streptococcus gallolyticus*, *Streptococcus orisratti, Streptococcus pneumoniae*, and *Streptococcus uberis*. In addition, commensal Pasteurellaceae including *Actinobacillus indolicus, Actinobacillus minor, Actinobacillus porcinus,* and *Haemophilus parasuis* (Nagasaki and SW140) were also included, as well as, an Alcaligenaceae isolate of *Bordetella bronchiseptica* RB50 (NC_002927) [563]. No mPCR amplicons, after 35 thermal cycles and gel electrophoresis, were visible under UV transillumination for any of the panel of ten negative controls (data not shown).

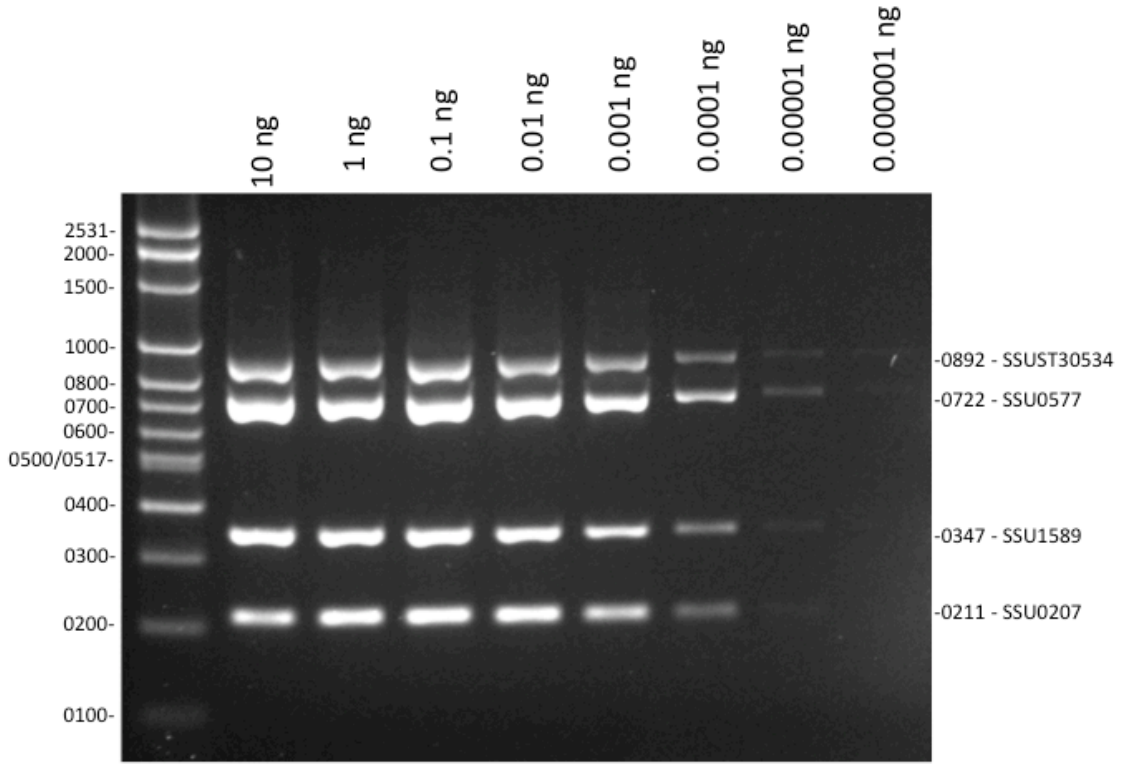

Figure 4.5. The detection limit of the multiplex-PCR component of the pathotyping tool was estimated to be ~0.0001 ng of *Streptococcus suis* genomic DNA.

Agarose gel containing multiplex-PCR amplicons produced from 10-fold serial dilutions of *S. suis* genomic DNA over the range 1 $x10^{1}$ - $1x10^{-6}$ ng per reaction. Purified genomic DNA from four *S. suis* isolates: SS002, SS004, LSS011 and LSS027 was mixed in equal quantities so that templates for each mPCR amplicon would be present in all reactions, and a series of 10-fold dilutions then performed to create mPCR templates decreasing in DNA concentration. mPCR amplicons were electrophoresed on a 2% (wt/vol) agarose gel containing 1x SYBR Safe DNA gel stain for 60 min at a constant 100 V before being photographed under UV transillumination. The detection limit of ~0.0001 ng of *S. suis* genomic DNA was estimated as the lowest concentration of template DNA from which all anticipated mPCR amplicons, after 35 thermal cycles, were easily visible under UV transillumination on a Gel Doc XR+ Gel Documentation System (Bio-Rad). Approximate template starting concentrations per reaction are indicated above lanes; M contains the Bioline HyperLadder 100 bp Plus DNA ladder with sizes indicated on the left. Multiplex-PCR amplicon sizes are indicated on the right.

4.2.2.2 Out-of-sample testing

To evaluate further the mPCR pathotyping tool (GLM and mPCR) out-of-sample testing of 138 (69 disease-associated and 69 non-disease associated) previously uncharacterised *S. suis* isolates was performed *in vitro*. Out-of-sample forecasting is a common approach used to evaluate the performance of binary diagnostic tests. Figures 4.6 and 4.7 show the mPCR amplicons produced from genomic DNA extracted from each of the 138 'test' collection isolates. A number of the gel lanes in Figure 4.6 have not been labelled as the mPCR was performed before detailed clinical metadata corresponding to the invasive disease-associated isolates (originally a collection of 105 isolates) was made available by the APHA. Unlabelled isolates were excluded from the research described in this thesis due to the clinical metadata being either incomplete or not allowing assignment to the well defined invasive disease-associated or non-disease associated phenotypic groups previously used to characterise the original 'training' collection. For example, a number of isolates were recovered from the lung in the presence of pneumonia.

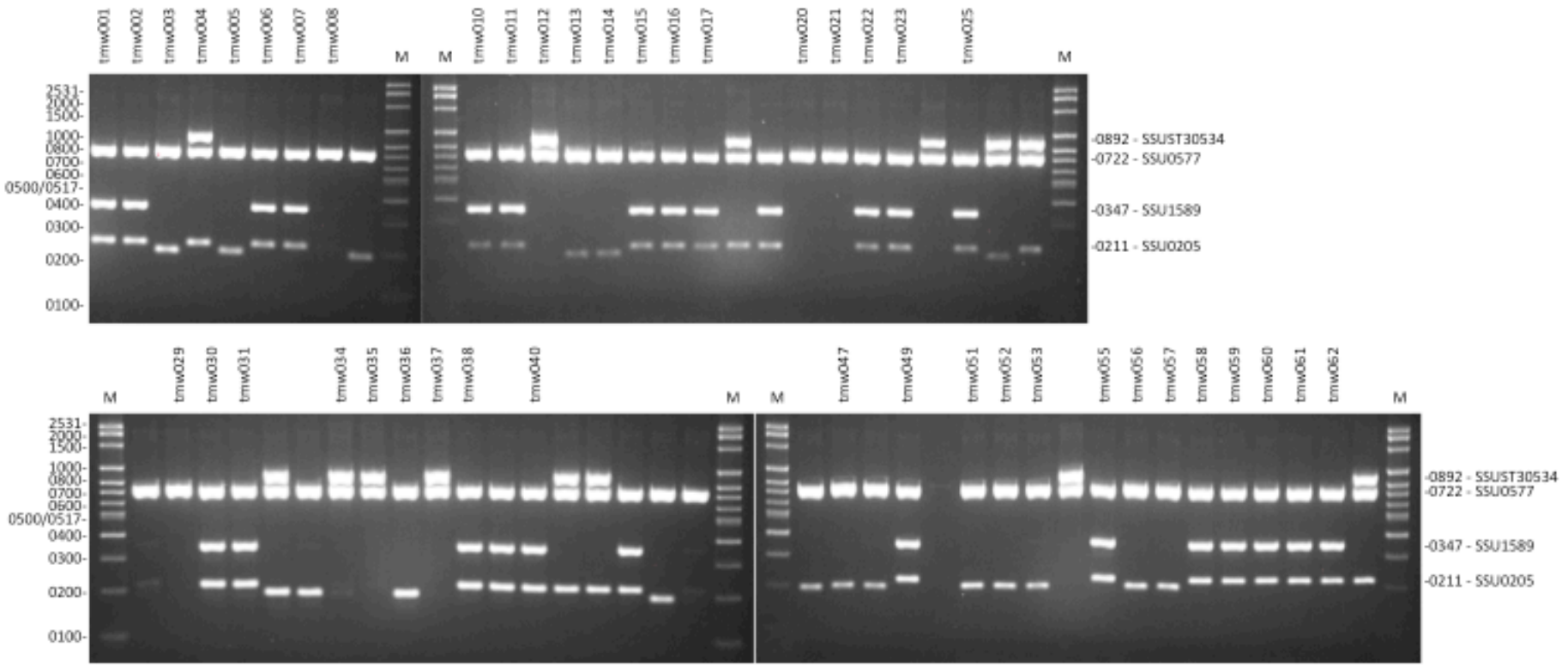

Figure 4.6 (a). Multiplex-PCR amplicons produced from template genomic DNA extracted from 43/69 invasive disease-associated isolates in the test collection. Agarose gel containing mPCR amplicons produced from *S. suis* genomic DNA of 43 out of 69 invasive disease-associated isolates in the out-of-sample test collection. PCR amplicons were electrophoresed on a 2% (wt/vol) agarose gel containing 1x SYBR Safe DNA gel stain for 60 min at a constant 100 V, and photographed under UV transillumination. In each instance multiplex-PCR amplicon patterns matched anticipated amplicon patterns based on *in silico* analyses. Isolate names are indicated above lanes. Lanes labelled M contain 1x Bioline HyperLadder 100 bp Plus DNA ladder with sizes indicated on the left. Multiplex-PCR amplicon sizes are indicated on the right. Primers for marker SSU0205 produce a 211 bp or 190 bp amplicon, due to a 21 bp indel in the protein-encoding sequence unavoidable during primer design.

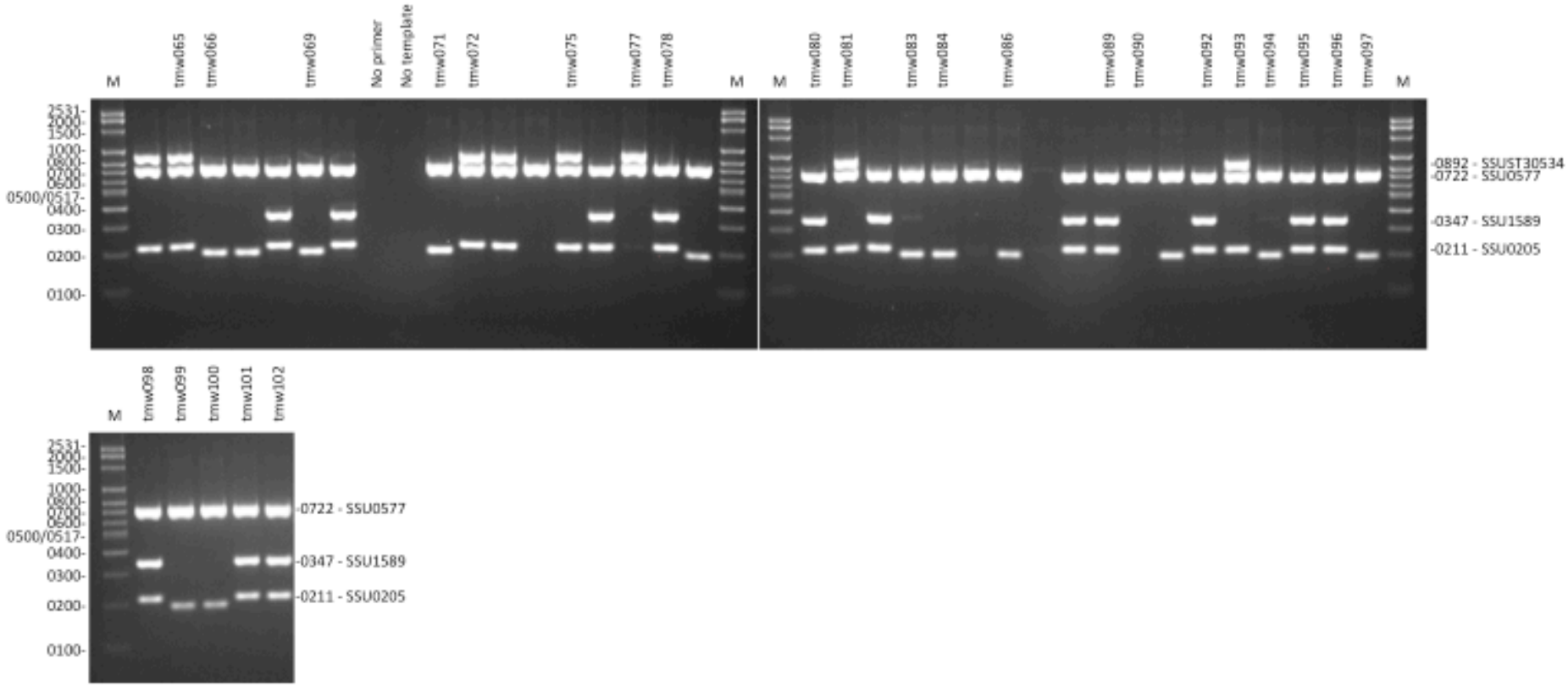

Figure 4.6 (b). Multiplex-PCR amplicons produced from template genomic DNA extracted from 26/69 invasive disease-associated isolates in the test collection.

Agarose gel containing mPCR amplicons produced from *S. suis* genomic DNA of 26 out of 69 invasive disease-associated isolates in the out-of-sample test collection. PCR amplicons were electrophoresed on a 2% (wt/vol) agarose gel containing 1x SYBR Safe DNA gel stain for 60 min at a constant 100 V, and photographed under UV transillumination. In each instance multiplex-PCR amplicon patterns matched anticipated amplicon patterns based on *in silico* analyses. Isolate names are indicated above lanes. Lanes labelled M contain 1x Bioline HyperLadder 100 bp Plus DNA ladder with sizes indicated on the left. Multiplex-PCR amplicon sizes are indicated on the right. Primers for marker SSU0205 produce a 211 bp or 190 bp amplicon, due to a 21 bp indel in the protein-encoding sequence unavoidable during primer design.

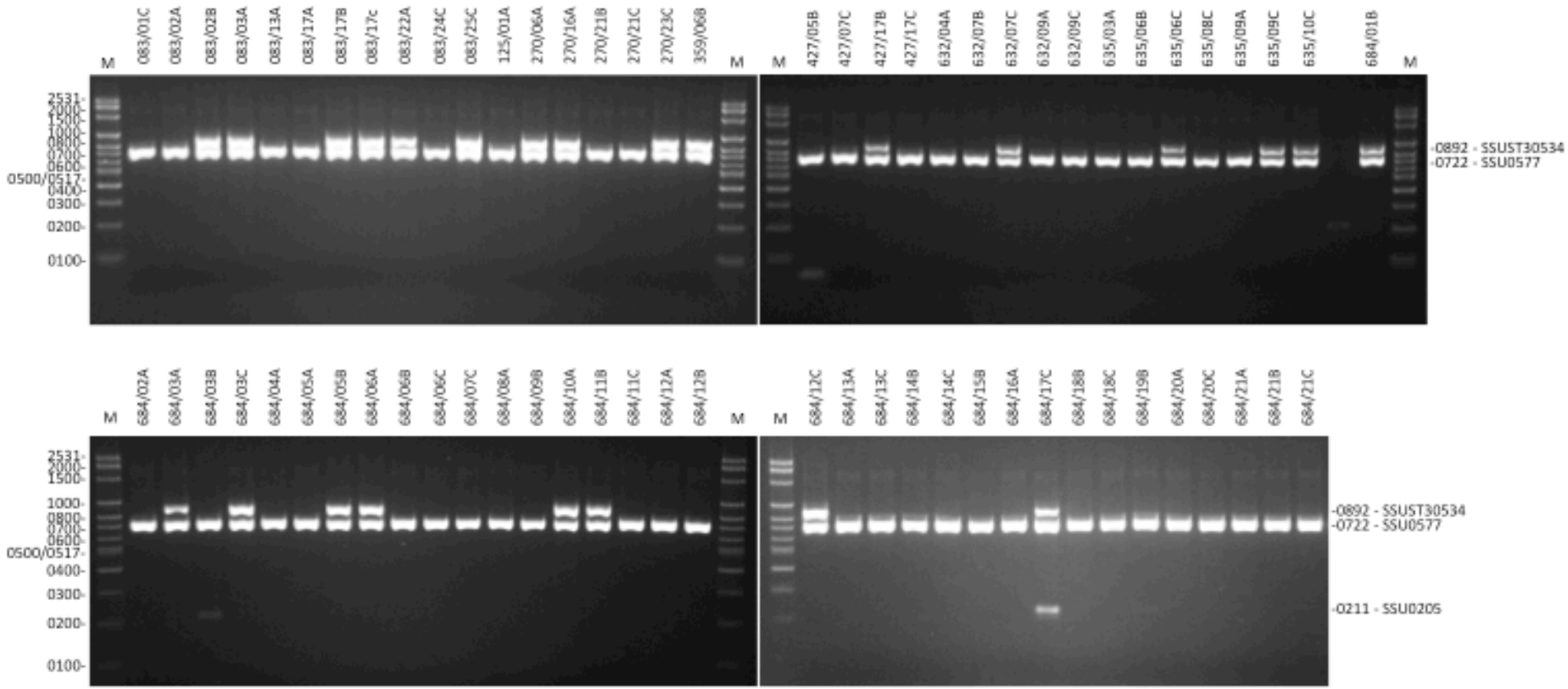

Figure 4.7. Multiplex-PCR amplicons produced from template genomic DNA extracted from 69 non-disease associated isolates in the test collection.

Agarose gel containing mPCR amplicons produced from *S. suis* genomic DNA extracted from 69 non-disease associated isolates in the out-of-sample test collection. PCR amplicons were electrophoresed on a 2% (wt/vol) agarose gel containing 1x SYBR Safe DNA gel stain for 60 min at a constant 100 V, and photographed under UV transillumination. In each instance multiplex-PCR amplicon patterns matched anticipated amplicon patterns based on *in silico* analyses. Isolate names are indicated above lanes. Lanes labelled M contain 1x Bioline HyperLadder 100 bp Plus DNA ladder with sizes indicated on the left. Multiplex-PCR amplicon sizes are indicated on the right. Primers for marker SSU0205 produce a 211 bp or 190 bp amplicon, due to a 21 bp indel in the protein-encoding sequence unavoidable during primer design.

All 138 isolates of the test collection produced the mPCR amplicon of size 722 bp corresponding to the *S. suis* species-specific marker. Protein-encoding sequence detection data, based on the production of mPCR amplicons (corresponding to the pathotyping markers), was then input into the pathotyping GLM and the predict function, implemented in the R package: *logistf* [521], used to generate fitted values for each isolate. Table 4.3 shows the performance metrics of the pathotyping tool once the 0.43 cutoff threshold was applied to the fitted values in order to generate a binary class decision (disease = 1/non-disease =0) for each isolate. The pathotyping tool predicted correct classification, in comparison to the observed clinical metadata, of 80% (accuracy) of the 138 isolates. False negative rate deemed the most valuable performance metric for pathotyping was 0.20, assuming clinical metadata as the 'gold standard'. Other popular performance metrics commonly used to compare binary diagnostic tests were; sensitivity = 0.80, specificity = 1.00, positive predictive value = 1.00 and negative predictive value = 0.83.

Table 4.3. Performance metrics calculated for the pathotyping tool when applied to the out-of-sample 'test' collection and compared to the 'gold-standard' observed clinical phenotype.

An extended 2x2 contingency table for calculating the performance metrics of two binary diagnostic tests.

		Phenotype			
	138 Total population	69 Phenotype positive	69 Phenotype negative		
mPCR result	55 mPCR result positive	55 True positive	0 False positive	1.00 Positive predictive value	0.00 False discovery rate
	83 mPCR result negative	14 False negative	69 True negative	0.17 False omission rate	0.83 Negative predictive value
		0.80 True positive rate	0.00 False positive rate	0.89 F1 score	
		0.20 False negative rate	1.00 True negative rate		

4.2.2.2.1 Comparison of the pathotyping tool to published methods used to characterise disease associated isolates of *Streptococcus suis*

To compare the pathotyping tool to published methods used to characterise and sub-type disease-associated isolates of *S. suis*, the molecular serotype, 'virulence-associated' gene (*epf*, *mrp*, and *sly*) profile and minimum core-genome sequence type were first determined *in silico*. All comparisons were then drawn between typing methods using the test collection of 138 isolates.

4.2.2.2.1.1 Serotypes 1-10, 14 and 1/2 classified isolates of *Streptococccus suis* with a superior sensitivity in comparison to the pathotyping tool but the trade-off is a significantly worse specificity and positive predictive value

Traditional (coagglutination) serotyping data was unavailable for all isolates of the out-of-sample 'test' collection, therefore, molecular 'serotyping' was performed and used for comparison against the pathotyping tool. Figure 4.8 shows the cumulative frequency of isolates assigned to the 35 (1-34 and 1/2) originally reported *S. suis* serotypes. Serotype 2 (and 1/2, indistinguishable by the Liu *et al.* [74] method due to closely related *cps* genes) was assigned most frequently to isolates in test collection. Four (a total of 6% of) invasive disease-associated isolates in the test collection were assigned 'serotypes' other than 1-10, 14 and 1/2 (the most common serotypes recovered from clinical samples). A total of 16 non-disease associated isolates were assigned 'serotypes' other than 1-10, 14 and 1/2. Isolates that could not be assigned to one of the 35 originally described *S. suis* serotypes were present in test collection, although this non-serotypable 'group' was dominated by non-disease associated isolates (two invasive disease-associated and 13 non-disease associated).

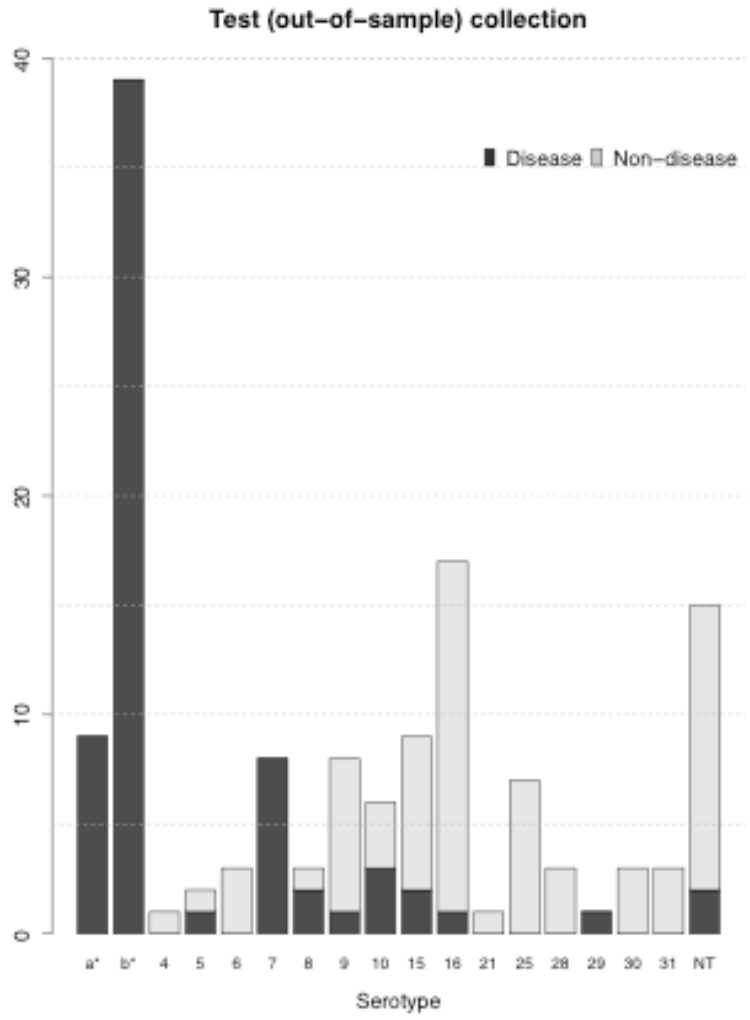

Figure 4.8. Cumulative frequency of molecular 'serotypes' assigned to *Streptococcus suis* isolates in the out-of-sample 'test' collection using an *in silico* based approach.

Cumulative frequency of the originally described *S. suis* serotypes (1-34 and 1/2) assigned to isolates in the test collection (n=138). Serotyping was performed using an adaptation, for *in silico* use, of the multiplex-PCR assays described by Liu *et al.* [74]. The Liu *et al.* molecular 'serotyping' method, the most recent at the time of undertaking this work, is unable to differentiate all reported *S. suis* serotypes due to closely related *cps* genes. Groups "a*" represent serotypes 1 and 14 and "b*" represent serotypes 2 and 1/2. Isolates that could not be assigned to one of the 35 originally described *S. suis* serotypes were designated non-serotypable (NT). Colour has been used to represent the observed clinical phenotype i) invasive disease-associated (black) and ii) asymptomatic carriage on the palatine tonsils of pigs on UK farms (grey-scale).

Li *et al.* [74] describe a total of four mPCRs to generate molecualr serotypes for *S. suis* isolates, the first of which targets the *cps* genes of serotypes 1-10, 14 and 1/2. To draw direct and meaningful comparison to a published tool to characterise and subtype *S. suis* isolates serotypes targeted by the first Liu *et al.* mPCR were used as a binary classifier to predict disease association. To compare the pathotyping tool to the use of serotype as a proxy for virulence potential, the serotypes 1-10, 14 and 1/2 were considered markers of disease association and all other serotypes considered markers of non-disease association (including isolates deemed to be non-serotypable). Table 4.4 shows use of the 12 serotypes (1-10, 14 and 1/2) to predict disease-association performed with a sensitivity of 0.91 (n=6 type II errors), superior in comparison to the mPCR pathotyping tool (sensitivity = 0.80, n=14 type II errors; McNemar's Chi-squared Test for Count Data *p*-value = 0.00468). However, the trade-off for a high sensitivity (equivalent to a low false-negative rate) is a statistically worse positive predictive value (0.80 in comparison to 1.00; weighted generalised score statistic for comparison of predictive values *p*-value = 0.00016) and a worse specificity (0.77 in comparison to 1.00; McNemar's chi-squared test for count data *p*-value = 0.00006); no significant difference in negative predictive value was observed (*p*-value 0.06025).

Table 4.4. Performance metrics calculated for the use of serotyping as a proxy for virulence when applied to the 'test' collection and compared to the 'gold-standard' observed clinical phenotype.

An extended 2x2 contingency table for calculating the performance metrics of two binary diagnostic tests. (**a**) To draw direct comparison to a published tool to characterise *S. suis* isolates, serotypes 1-10, 1/2 and 14 were considered a marker of disease association and all other serotypes considered markers of non-disease association (including isolates deemed to be non-serotypable). Traditional serotyping data was unavailable for the out-of-sample test collection, therefore, molecular 'serotyping' was performed using an adaptation, for *in silico* use, of the mPCR assay described by Liu *et al.* [74]. (**b**) For comparison. Performance metrics calculated for the pathotyping tool when applied to the out-of-sample 'test' collection and compared to the 'gold-standard' observed clinical phenotype.

(a) Serotypes 1-10, 14 and 1/2

		Phenotype			
	138 Total population	69 Phenotype positive	69 Phenotype negative		
mPCR result	79 mPCR result positive	63 True positive	16 False positive	0.80 Positive predictive value	0.20 False discovery rate
	59 mPCR result negative	6 False negative	53 True negative	0.10 False omission rate	0.90 Negative predictive value
		0.91 True positive rate	0.23 False positive rate	0.85 F1 score	
		0.09 False negative rate	0.77 True negative rate		

(b) Multiplex-PCR cutoff: 0.43

		Phenotype			
	138 Total population	69 Phenotype positive	69 Phenotype negative		
mPCR result	55 mPCR result positive	55 True positive	0 False positive	1.00 Positive predictive value	0.00 False discovery rate
	83 mPCR result negative	14 False negative	69 True negative	0.17 False omission rate	0.83 Negative predictive value
		0.80 True positive rate	0.00 False positive rate	0.89 F1 score	
		0.20 False negative rate	1.00 True negative rate		

4.2.2.2.1.2 'Virulence-associated' factors classified isolates of *Streptococcus suis* with a superior sensitivity in comparison to the pathotyping tool but the trade-off is positive and negative predictive value no better than chance

Numerous other markers have been used to characterise and subtype *S. suis* isolates, one such approach for the identification of virulent strains is the detection of 'virulence-associated' factors EF (encoded by *epf*), MRP (encoded by *mrp*) and the thiol-activated haemolysin known as SLY (encoded by *sly*). To draw direct and meaningful comparison to a published tool to characterise *S. suis* isolates adaptation, for *in silico* use, of the method described by Silva *et al.* [400] was used for virulence-associated gene profiling of all *S. suis* isolates described in this thesis (n=253). Figure 4.9 shows the proportion of isolates, per strain collection, testing positive for each of the three genes: *epf*, *mrp* and *sly* (independent of one another). In both the training collection and out-of-sample 'test' collection each gene was independently detected in over half of invasive disease-associated isolates (black). Interestingly, all three genes were positively detected in the same isolate in 75% and 52% of invasive disease-associated isolates in the training collection and test collection respectively (data not shown).

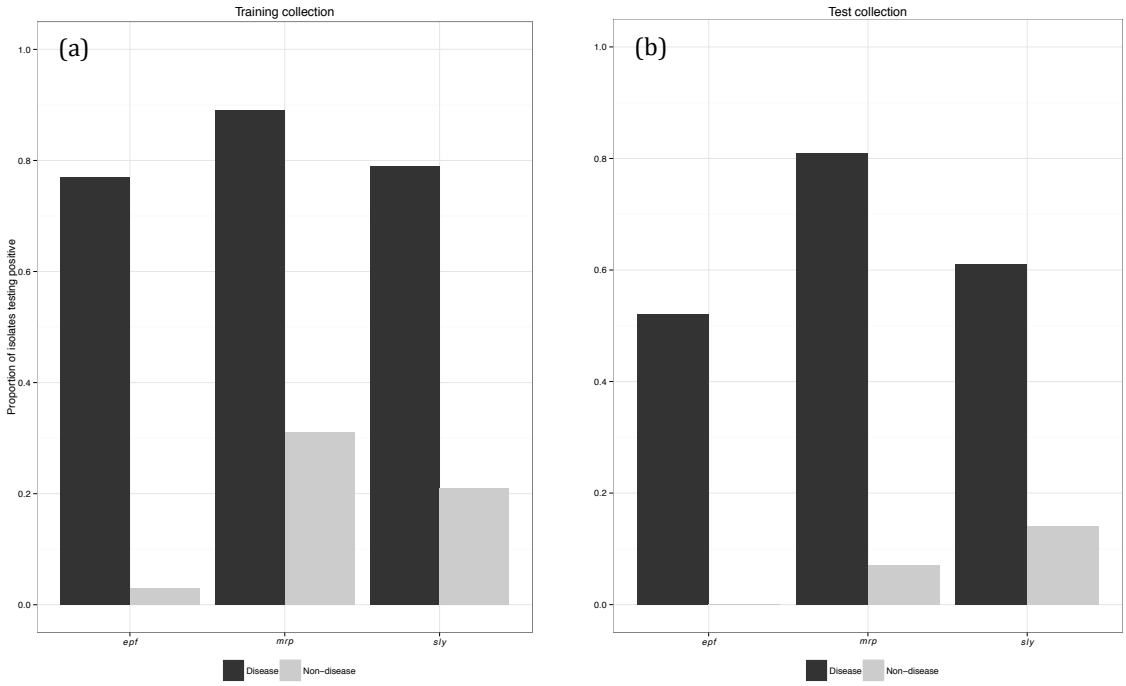

Figure 4.9. Proportion of training collection and out-of-sample 'test' collection isolates testing positive for the protein-encoding sequences of virulence-associated factors *epf*, *mrp* and *sly*.

The proportion of *S. suis* isolates in the (a) training collection (n=115) and (b) out-of-sample 'test' collection (n=138) testing positive for the virulence-associated factors extracellular factor (encoded by *epf*), muramidase-released protein (encoded by *mrp*) and the thiol-activated haemolysin suilysin (encoded by *sly*). Positive detections (independent of one another) were determined using an adaptation, for *in silico* use, of the virulence-associated gene profiling method described by Silva *et al.* [400].

Large and small molecular weight variants of EF and MRP have been described [53, 409, 530]. *In silico* analyses of the distance between primer sequence matches targeting the *epf* gene revealed all positive matches to be 744 bp apart, corresponding to the *epf*$^{+}$ protein-encoding sequence, and not one of the five classes of *epf** variants. The mPCR assay described by Silva *et al.* [400] was intentionally designed to generate a single 188 bp amplicon to infer the presence of *mrp* irrespective of the *mrp* variant. In the same publication, Silva *et al.* [400] also describe an additional (singleplex) PCR assay to differentiate the genes that encode the known variants of MRP. Investigation of the distance between *mrp* singleplex primer sequence matches identified five out of the six known *mrp* size variants in the 253 isolates described in this thesis. Figure 4.10 shows the proportion of isolates testing positive for each known *mrp* size variant, of which the 1148 bp 'amplicon' corresponding to the *mrp*$^{+}$ protein-encoding sequence was the most common.

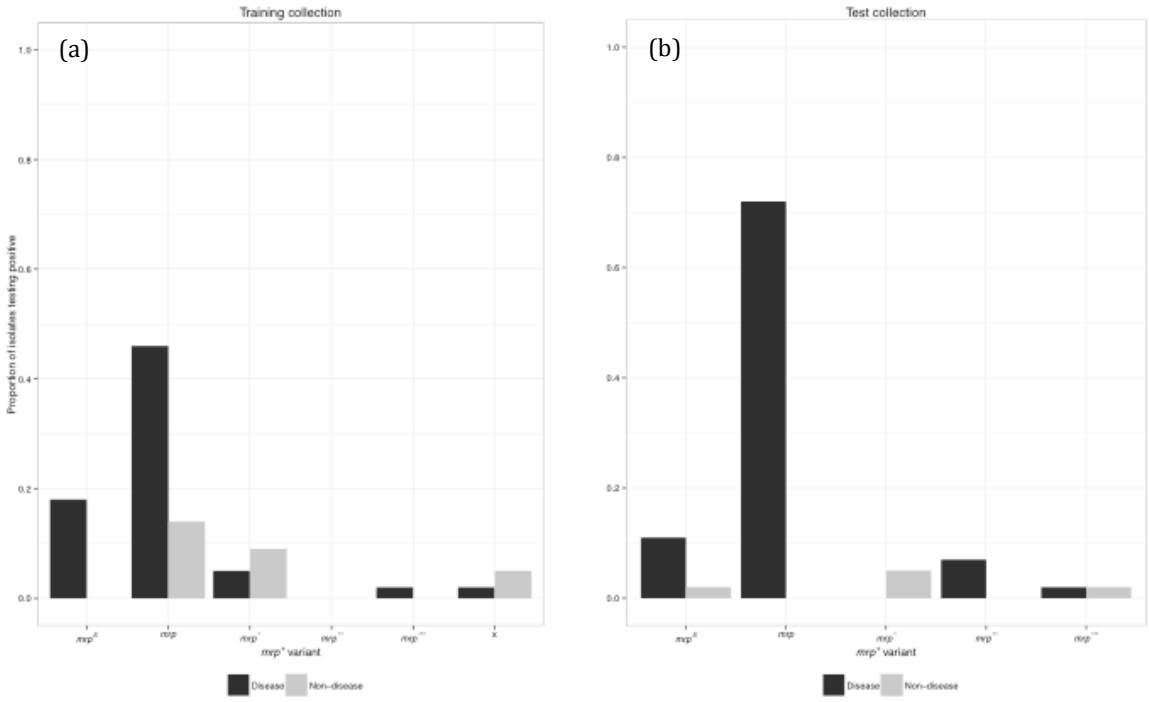

Figure 4.10. Proportion of training collection and out-of-sample 'test' collection isolates testing positive for the known variants of the 'virulence-associated' factor *mrp*.

The proportion of *S. suis* isolates in the (a) training collection (n=115) and (b) out-of-sample 'test' collection (n=138) testing positive for the protein-encoding sequences known to encode variants of the virulence-associated factor muramidase-released protein (encoded by *mrp*). "*x*" represents *mrp* size variants that could not be identified *in silico* as oligonucleotide matches (both >95%) were to separate contigs of the unfinished draft genome assemblies.

In order to draw robust and meaningful comparison between the pathotyping tool and the use of the virulence-associated genes *epf*, *mrp* and *sly*, first a GLM was fitted to the prevalence data of *epf*, *mrp* and *sly* in the original 'training' collection. Following the same logistic regression analysis methodology as before a GLM was fitted to the 'training' collection and then a cutoff threshold selected (0.12) to convert the fitted values into a binary class decision. Table 4.5 shows the influence cutoff threshold had on performance metrics. Again, the closest F_1 score to 1 was traded-off against a false negative rate closest to 0 to select the cutoff threshold. The predict function, implemented in the R package: *logistf* [521], was then used to generate fitted values for the detection data of *epf*, *mrp* and *sly* for each of the 138 isolates in the test collection. Table 4.6 summarises the classification of isolates as disease/non-disease associated based on the detection of *epf*, *mrp* and *sly* in comparison to the 'gold-standard' observed clinical phenotype. The effect of *mrp* length variations on virulence to date remains unclear, therefore, in order to make fair comparisons to an existing sub-typing tool the decision was made to draw comparison against the Silva *et al.* [400] mPCR assay designed to capture *epf*, *mrp* and *sly* irrespective of the size variations in *mrp*. Table 4.6 summaries the classification of the test collection and performance of *epf*, *mrp* and *sly* in comparison to the clinical metadata (assumed to be the 'gold standard'). The virulence-associated markers perform with a sensitivity of 0.93 (n=5 type II errors), that is superior in comparison to th pathotyping tool (sensitivity = 0.80, n=14 type II errors; McNemar's Chi-squared test for Count Data *p*-value = 0.00270). However, again the trade-off for a high sensitivity (equivalent to a low false-negative rate; 0.07) is a statistically worse positive predictive value (0.50 in comparison to 1.00; weighted generalised score statistic for comparison of predictive values *p*-value = 4.77396e-15) – incidentally performing no better than chance (Exact binomial test *p*-value = 1). Specificity is also substantially worse (0.09 in comparison to 1.00; McNemar's Chi-squared Test for Count Data *p*-value = 0.00006), as is the negative predictive value (0.55 in in comparison to 1.00; weighted generalised score statistic for comparison of predictive values *p*-value = 0.00272) – again performing no better than chance (Exact binomial test *p*-value = 1).

Table 4.5. Performance metrics calculated for the pathotyping tool when applied to the training collection and compared to the 'gold-standard' observed clinical phenotype.

An extended 2x2 contingency table for calculating the performance metrics of two binary diagnostic tests.

(i) Cutoff: 0.12

		Phenotype			
	115 Total population	53 Phenotype positive	62 Phenotype negative		
mPCR result	60 mPCR result positive	52 True positive	8 False positive	0.87 Positive predictive value	0.13 False discovery rate
	55 mPCR result negative	1 False negative	54 True negative	0.02 False omission rate	0.98 Negative predictive value
		0.98 True positive rate	0.13 False positive rate	0.92 F_1 score	
		0.02 False negative rate	0.87 True negative rate		

(ii) Cutoff: 0.21

		Phenotype			
	115 Total population	53 Phenotype positive	62 Phenotype negative		
mPCR result	67 mPCR result positive	48 True positive	19 False positive	0.72 Positive predictive value	0.28 False discovery rate
	48 mPCR result negative	5 False negative	43 True negative	0.10 False omission rate	0.90 Negative predictive value
		0.91 True positive rate	0.31 False positive rate	0.80 F_1 score	
		0.09 False negative rate	0.69 True negative rate		

(iii) Cutoff: 0.30

		Phenotype			
	115 Total population	53 Phenotype positive	62 Phenotype negative		
mPCR result	64 mPCR result positive	48 True positive	16 False positive	0.75 Positive predictive value	0.25 False discovery rate
	51 mPCR result negative	5 False negative	46 True negative	0.10 False omission rate	0.90 Negative predictive value
		0.91 True positive rate	0.26 False positive rate	0.82 F_1 score	
		0.09 False negative rate	0.74 True negative rate		

(iv) Cutoff 0.84

		Phenotype			
	115 Total population	53 Phenotype positive	62 Phenotype negative		
mPCR result	43 mPCR result positive	41 True positive	2 False positive	0.95 Positive predictive value	0.05 False discovery rate
	72 mPCR result negative	12 False negative	60 True negative	0.17 False omission rate	0.83 Negative predictive value
		0.77 True positive rate	0.03 False positive rate	0.85 F_1 score	
		0.23 False negative rate	0.97 True negative rate		

(v) Cutoff 0.94

		Phenotype			
	115 Total population	53 Phenotype positive	62 Phenotype negative		
mPCR result	42 mPCR result positive	40 True positive	2 False positive	0.95 Positive predictive value	0.05 False discovery rate
	73 mPCR result negative	13 False negative	60 True negative	0.18 False omission rate	0.82 Negative predictive value
		0.75 True positive rate	0.03 False positive rate	0.84 F_1 score	
		0.25 False negative rate	0.97 True negative rate		

Table 4.6. Performance metrics calculated for the use of the *epf*, *mrp* and *sly* markers as a proxy for virulence when applied to the out-of-sample 'test' collection and compared to the 'gold-standard' observed clinical phenotype.

An extended 2x2 contingency table for calculating the performance metrics of two binary diagnostic tests. (**a**) To draw direct comparison to a published tool to characterise *S. suis* isolates, the virulence-associated factors extracellular factor (encoded by *epf*), muramidase-released protein (encoded by *mrp*) and the thiol-activated haemolysin suilysin (encoded by *sly*) were used as GLM explanatory variables. Positive detections (independent of one another) were determined using an adaptation, for *in silico* use, of the virulence-associated gene profiling method described by Silva *et al.* [400]. (**b**) For comparison. Performance metrics calculated for the pathotyping tool when applied to the out-of-sample 'test' collection and compared to the 'gold-standard' observed clinical phenotype.

(a) *epf*/*mrp*/*sly* cutoff: 0.12

		Phenotype			
	138 Total population	69 Phenotype positive	69 Phenotype negative		
mPCR result	127 mPCR result positive	64 True positive	63 False positive	0.50 Positive predictive value	0.50 False discovery rate
	11 mPCR result negative	5 False negative	6 True negative	0.45 False omission rate	0.55 Negative predictive value
		0.93 True positive rate	0.91 False positive rate	0.65 F1 score	
		0.07 False negative rate	0.09 True negative rate		

(b) Multiplex-PCR cutoff: 0.43

		Phenotype			
	138 Total population	69 Phenotype positive	69 Phenotype negative		
mPCR result	55 mPCR result positive	55 True positive	0 False positive	1.00 Positive predictive value	0.00 False discovery rate
	83 mPCR result negative	14 False negative	69 True negative	0.17 False omission rate	0.83 Negative predictive value
		0.80 True positive rate	0.00 False positive rate	0.89 F1 score	
		0.20 False negative rate	1.00 True negative rate		

4.3.2.2.1.3 Assignment to minimum core genome sequence typing group 1 as a proxy for virulence classified isolates with a significantly worse sensitivity and negative predictive value

Lastly, the pathotyping tool was compared to the use of one of the most recent methods used to characterise and sub-type *S. suis*, minimum core genome sequence typing [115, 116]. Figure 4.11 shows the proportion of isolates, per strain collection, assigned to one of the nine originally described MCG sequence typing groups. MCG group 1 was reported as being assigned to all highly virulent isolates and epidemic isolates tested during its design [115], an observation also seen in the invasive disease-associated isolates of the training collection (Figure 4.11 (a)). In fact, MCG group 1 was assigned to 77% of the disease-associated isolates (black; n=41) in the training collection. Therefore, assignment to MCG group 1 was used as a binary classifier to indicate disease association. Table 4.7 summaries the classification of the out-of-sample 'test' collection and performance of assignment to MCG group 1 as an indicator of disease association in comparison to the clinical metadata (assumed to be the 'gold standard'). Here. assignment to MCG group 1 performs with a statistically worse sensitivity of 0.52 (n=33 type II errors) in comparison to the pathotyping tool (sensitivity = 0.80, n=14 type II errors; McNemar's Chi-squared Test for Count Data p-value = 0.00001). Use of MCG group 1 also produces a statistically worse negative predictive value (0.68 in comparison to 0.83; weighted generalised score statistic for comparison of predictive values p-value = 0.00001); while other performance metrics (specificity and positive predictive value) were not found to be statistically different.

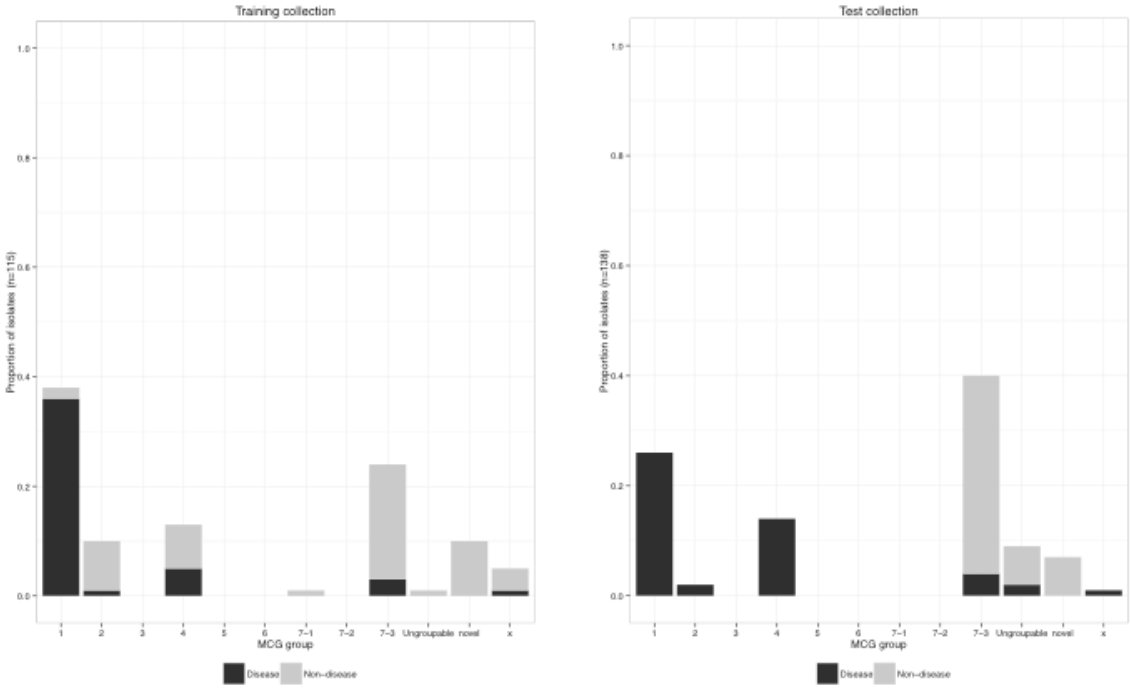

Figure 4.11. Proportion of training collection and out-of-sample 'test' collection isolates assigned to the previously described minimum core-genome sequence typing groups.

The proportion of *S. suis* isolates in the (a) training collection (n=115) and (b) out-of-sample 'test' collection (n=138) assigned to the previously described minimum core-genome sequence typing groups [115, 116].). "*x*" represents MCG groups that could not be identified *in silico* as oligonucleotide matches (both >95%) were to separate contigs of the unfinished draft genome assemblies.

Table 4.7. Performance metrics calculated for the use of assignment to minimum core genome sequence type group 1 as a proxy for virulence when applied to the out-of-sample 'test' collection and compared to the 'gold-standard' observed clinical phenotype.

An extended 2x2 contingency table for calculating the performance metrics of two binary diagnostic tests. (**a**) To draw direct comparison to a published tool to characterise *S. suis* isolates, minimum core genome sequence typing was performed using an adaptation, for *in silico* use, of the method described by Zheng *et al*. [116]. (**b**) For comparison. Performance metrics calculated for the pathotyping tool when applied to the out-of-sample 'test' collection and compared to the 'gold-standard' observed clinical phenotype.

(a) MCG group 1

		Phenotype			
	138 Total population	69 Phenotype positive	69 Phenotype negative		
mPCR result	36 mPCR result positive	36 True positive	0 False positive	1.00 Positive predictive value	0.00 False discovery rate
	102 mPCR result negative	33 False negative	69 True negative	0.32 False omission rate	0.68 Negative predictive value
		0.52 True positive rate	0.00 False positive rate	0.69 F1 score	
		0.48 False negative rate	1.00 True negative rate		

(b) Multiplex-PCR cutoff: 0.43

		Phenotype			
	138 Total population	69 Phenotype positive	69 Phenotype negative		
mPCR result	55 mPCR result positive	55 True positive	0 False positive	1.00 Positive predictive value	0.00 False discovery rate
	83 mPCR result negative	14 False negative	69 True negative	0.17 False omission rate	0.83 Negative predictive value
		0.80 True positive rate	0.00 False positive rate	0.89 F1 score	
		0.20 False negative rate	1.00 True negative rate		

4.2.3 Pilot study evaluation of the pathotyping tools ability to screen pooled oral fluid collected from cotton chew ropes

Before screening valuable field swabs a small pilot study was performed to investigate if the pathotyping tool could be used to screen material from the upper respiratory tract of pigs without culture and first isolating single clones/colonies. To do this, total genomic DNA, extracted directly from transport swabs and used as mPCR templates. PCR inhibitors in salivary fluid have been reported to be problematic in various studies working with oral fluid, a problem that can be offset by diluting PCR template material. Figure 4.12 (a) shows PCR products amplified using the universal primers (U16SRT-F and U16SRT-R; PCR product size ~200 bp) designed by Clifford *et al*. [564] to target the bacterial *16S* gene, and were used here to confirm the presence of bacterial DNA in DNA extraction eluate. Figure 4.12 (b) then shows mPCR amplicons produced from DNA extracted directly from oral fluid and subjected to a series of 10-fold dilutions. From undiluted DNA five (including both size variants of SSU0207) mPCR amplicons can easily be seen under UV transillumination that become increasingly more difficult to see as PCR template DNA template was diluted 10-fold. All amplicons were Sanger sequenced using the Source Bioscience Lifesciences sequencing service and confirmed to be the correct target regions of the mPCR primers (data not shown). In addition to the 722 bp mPCR amplicon corresponding to the *S. suis* species-specific marker (SSU0577), both disease-associated and non-disease associated markers were also amplified from the same experimental sample. This was not observed in either the training collection (n=115) or out-of-sample 'test' collection (n=138). Indeed, in both the training collection and the out-of-sample test collection the disease-associated marker SSU1589 (347 bp) and non-disease associated marker (892 bp) are mutually exclusive and have not yet been detected in the same isolate. Therefore, it was inferred that at least two (heterogeneous) strains of *S. suis* were present in the pooled oral fluid collected from cotton chew ropes.

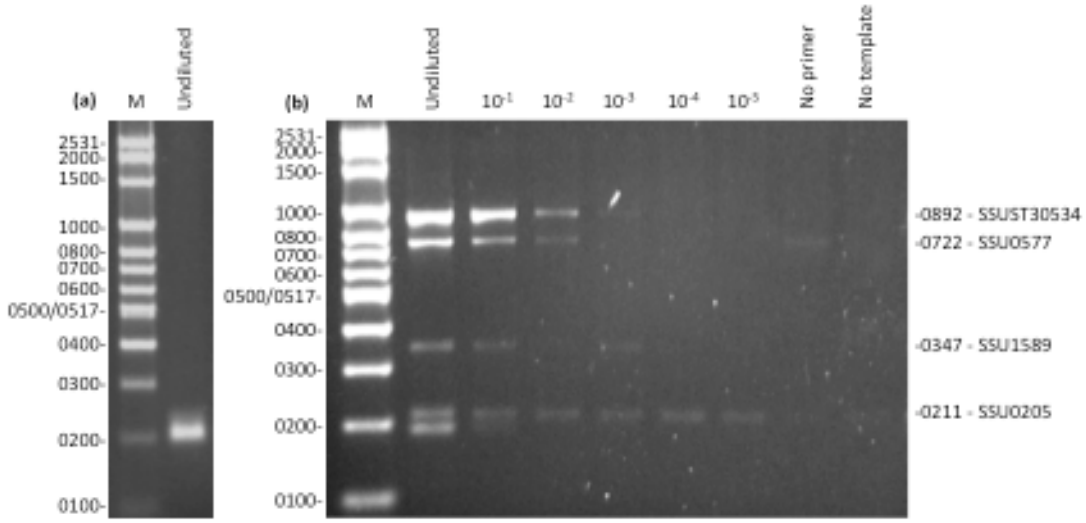

Figure 4.12. Pathotyping multiplex-PCR amplicons produced from template genomic DNA extracted from oral fluid collected from cotton chew ropes without culture and first isolating single colonies.

(**a**) Agarose gel containing PCR amplicons produced using universal primers (U16SRT-F and U16SRT-R; PCR product size ~200 bp) designed by Clifford *et al*. [564] to target the bacterial *16S* gene – used to confirm the presence of bacterial genomic DNA in DNA extraction eluate. (**b**) Agarose gel containing pathotyping tool mPCR amplicons produced from genomic DNA extracted directly from pooled oral fluid collected from cotton chew ropes without culture and first isolating single colonies/clones. PCR amplicons were electrophoresed on a 2% (wt/vol) agarose gel containing SYBR Safe DNA gel stain for 60 min at a constant 100 V and photographed under UV transillumination. Template DNA dilutions are indicated above lanes. Lane 1 contains 1x Bioline HyperLadder 100 bp Plus DNA ladder with sizes indicated on the left (bp).

4.2.3.1 Screening swabs of oral fluid and swabs of material scraped from the palatine tonsils of pigs for invasive disease-associated strains of *Streptococcus suis* revealed invasive disease-associated and non-disease associated strains frequently coexist on the tonsils of colonised pigs

The pathotyping tool was used to screen (paired samples) swabs of oral fluid and swabs of material scraped from the palatine tonsils of pigs on three UK farms (without obvious signs of streptococcal disease) for invasive disease-associated strains of *S. suis.* Figure 4.13 shows mPCR amplicons produced from DNA recovered from five randomly selected swabs from each of six groups sampled (three farms at two time points, five weeks and 20 weeks of age). In agreement with reports of *S. suis* being endemic in all sampled pig populations, the amplicon of size 722 bp corresponding to the *S. suis* species specific marker was produced in all groups sampled. Surprisingly, more often than not the disease-associated (SSU1589; 347 bp) and non-disease associated (SSUST30534; 892 bp) mPCR amplicons were produced from the same experimental sample, in both the training collection and test collection these pathotyping markers have not yet been detected in the same isolate. No difference in the prevalence of mPCR amplicons was obvious between the three farms sampled, or on a farm between the two time points sampled.

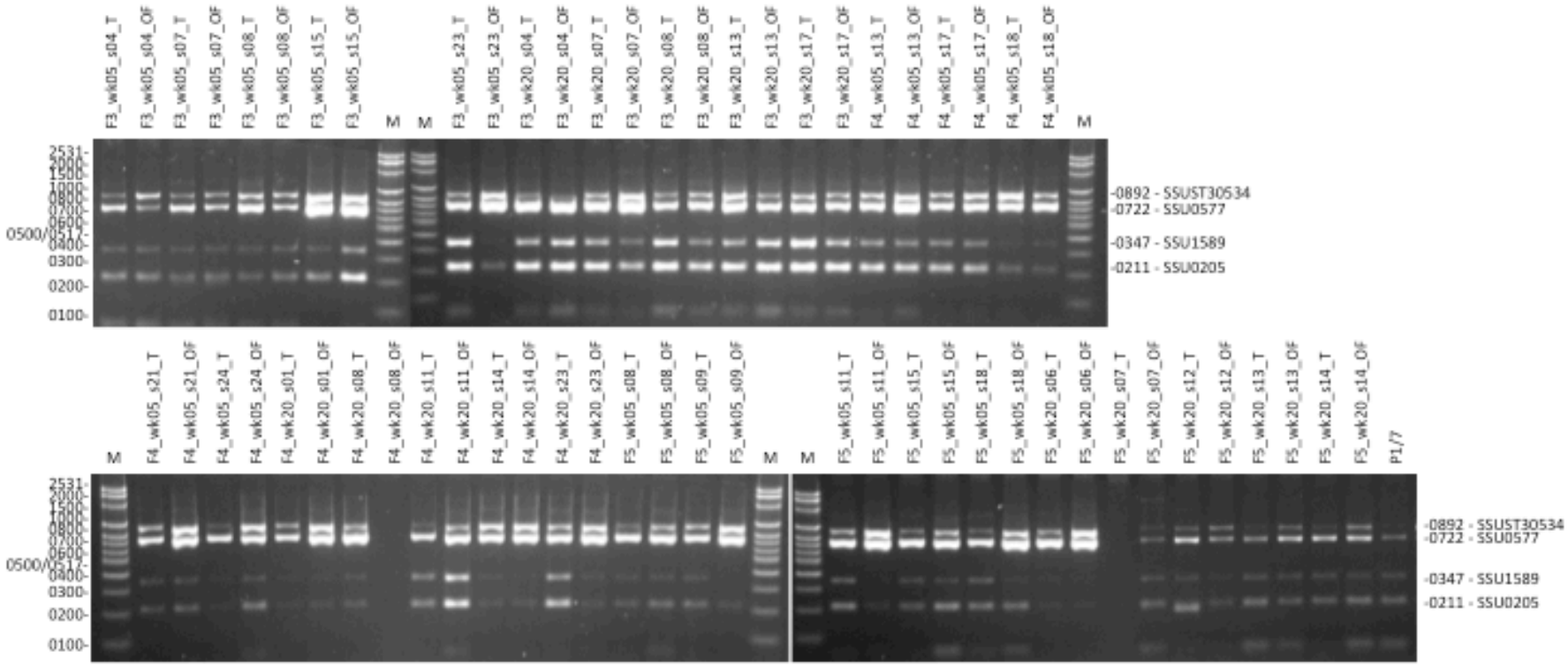

Figure 4.13. Multiplex-PCR amplicons produced from template genomic DNA extracted from paired samples of oral fluid and material scrapped from the palatine tonsils of pigs without obvious signs of streptococcal disease on farms in England and Wales.

Agarose gel containing mPCR amplicons produced from *S. suis* genomic DNA extracted from extracted from swabs of oral fluid or swabs of material scraped from the palatine tonsils of pigs without obvious signs of streptococcal disease. PCR amplicons were electrophoresed on a 2% (wt/vol) agarose gel containing 1x SYBR Safe DNA gel stain for 60 min at a constant 100 V, and photographed under UV transillumination. In each instance multiplex-PCR amplicon patterns matched anticipated amplicon patterns based on *in silico* analyses. Isolate names are indicated above lanes. Lanes labelled M contain 1x Bioline HyperLadder 100 bp Plus DNA ladder with sizes indicated on the left. Multiplex-PCR amplicon sizes are indicated on the right

4.3 Discussion

Objective 1. Starting with the shortlist of 14 genetic markers associated with observed clinical phenotype compiled in chapter 3 (Table 3.4), this chapter describes the use of logistic regression in the form of a GLM to select three statistically significant genetic markers to pathotype *S. suis*. Two genetic markers were positively associated with the invasive disease phenotype, while a third (SSUST30534) was positively associated with the non-disease associated phenotype (asymptomatic commensal-like carriage on the palatine tonsils of pigs on UK farms). The practical usefulness of the genetic marker positively associated with asymptomatic carriage might not be immediately obvious, but its statistical significance in the GLM did prompt discussions about the concept of antivirulence genes and gene loss in the evolution from non-pathogenic commensals to bacterial pathogens as a mechanism of fine tuning pathogen genomes for maximal fitness in new host environments; in short when regulation of the invasion, replication and transmission processes is altered, virulence can emerge. Genome reduction via gene loss and pseudogenisation resulting in enhanced pathogenicity has previously been described in other bacteria, such as *Rickettsia* spp., *Shigella* spp. and *Yersinia* spp. [509]. In fact, genome reduction through the loss of genes, potentially interfering with host infection, has also been proposed in *S. suis* [141]. Therefore, as the elimination of the genetic marker associated with asymptomatic commensal-like carriage from the GLM could not be done without a statistically significant loss of fit (p-value <0.05) it was retained and its practical usefulness evaluated further.

Objective 2. *In vitro* evaluation of the three genetic markers selected to pathotype *S. suis* showed 100% agreement with *in silico* analyses. In order to evaluate the ability of the genetic markers to predict the potential of isolates to cause invasive clinical disease a cutoff threshold (0.43) was chosen to convert the fitted values (or probability of causing disease) from the GLM into a binary class decision, disease (1) or non-disease (0). To do this, the false negative rate (0.09; equivalent to 1 – true positive rate (sensitivity)) was chosen as the most valuable performance metric for a *S. suis* pathotyping tool with a view to establish and then maintain a

pig population free of invasive disease-associated strains. Using the cutoff of 0.43, the pathotyping tool was found to predict the correct classification of 84% (accuracy) of the isolates of the original training collection (n=115) in comparison to the observed clinical metadata (Table 4.2(iii)), assuming the observed clinical metadata as the 'gold-standard').

The predicted classifications of five isolates of the training collection were considered to be false negatives (analogous to type II error, i.e. predicted to be non-disease associated but originally classified as being responsible for an invasive disease phenotype). Often *S. suis* strains are described as opportunistic or secondary pathogens that without a weakened host immune status (due to stress or concurrent infection) would normally be carried asymptomatically contributing to the normal oral microflora of pigs [121]. This is the most likely explanation for the differences observed between the pathotyping tool prediction and phenotype based on observed clinical metadata. Indeed, it is important to emphasise the fallibility of the phenotype assigned when it is based on field sampling without carefully controlled infection challenge data. Again, as discussed in chapter 3, an ideal standard would require an agreed panel of isolates for which a series of consistently controlled challenge experiments had been undertaken using pigs of identical immune status and genetics, however, this is remains unavailable for *S. suis* [302-304].

Despite being ranked as a lower priority in order to establish and then maintain a pig population free of invasive disease-associated strains of *S. suis*, it is equally important to discuss predictions considered to be false positives (type I errors). Table 4.2(iii) shows 13 isolates of the original training collection, that were recovered from the tonsil or trachea of pigs not showing obvious signs of streptococcal disease, were predicted by the pathotyping tool to have the ability to cause invasive disease. Deemed to be 'false' positives (in comparison to the 'gold-standard' observed clinical metadata) these predictions could in fact be correct predictions and examples of what the research described in this thesis is attempting to achieve – the identification of *S. suis* strains with the potential to cause invasive disease in the upper respiratory tract of pigs. An

important unmentioned caveat of pathotyping tool design is the source of *S. suis* isolates considered to be non-disease associated. While all efforts were made to accurately define disease-associated and non-disease associated phenotypic groups it is important to acknowledge that the non-disease associated isolates of the original training collection were recovered from routine submissions to the APHA, and that these pigs were not healthy even though they did not show signs of typical streptococcal disease; instead clinical features were consistent with disease caused by other non-*S. suis* infectious agents. The palatine tonsils of pigs are considered the natural habitat of *S. suis*. Therefore, it is feasible that the mortality of these 13 pigs was due to clones of isolates recovered from the tonsil or trachea-bronchus yet was not identified due to a concurrent infection presenting a more obvious phenotype such as diarrhoea. Despite this caveat there is evidence in the literature that virulent strains of *S. suis* can be isolated from tonsils of pigs without obvious streptococcal disease [70, 85]. This is likely to represent carriage of hot strains by pigs that have mounted an effective immune response that enable the virulent strain(s) to be kept in check in the tonsil (and not go systemic).

It is important to mention that although primer (BLASTn) and mPCR specificity for *S. suis* has been assessed to the best of our ability at the time, additional experiments should be considered. The most obvious is a control of pig (and human) genomic DNA, in order to be certain no false positive mPCR amplicons are produced due to host DNA. Indeed, although a panel of ten field isolates of Streptococcaceae, Pasteurellaceae, and Alcaligenaceae commonly recovered from the upper respiratory tract of pigs on farms in England and Wales were used as a panel of negative controls this should be expanded to include other important swine pathogens known to colonise the upper respiratory tract, such as (but not limited to) *Actinobacillus pleuropneumoniae*, *Haemophilus parasuis* and *Mycoplasma hyopneumoniae.*

Objective 3. Further evaluation of the pathotyping tool was performed *in vitro* using an out-of-sample 'test' collection of 138 previously uncharacterised *S. suis* isolates. Out-of-sample forecasting is a common approach used to evaluate the performance of binary diagnostic tests.

The same well-defined criteria were used to characterise the isolates of the test collection into phenotypic groups (disease/non-disease associated) based on the observed clinical metadata. Comparison of the predicted classification of isolates to the observed clinical phenotype (again assumed to be the 'gold-standard') found the pathotyping tool performed with a false negative rate = 0.20 (sensitivity = 0.80), accuracy = 0.80, true negative rate (specificity) = 1, positive predictive value = 1 and a negative predictive value = 0.83 (Table 4.3). Out-of-sample forecasting often produces worse performance metrics in comparison to in-sample testing where the model trained on the original data set is evaluated back on itself. Chosen as the most valuable performance metric for a *S. suis* pathotyping tool the 0.20 false negative rate corresponds to 14 false negatives (or type II errors); where non-disease associated pathotyping tool predictions were made for isolates linked with invasive disease clinical metadata. Speculation as to the reason for such observations has been discussed earlier. As with any new diagnostic tool it is important to draw comparison against an agreed 'gold-standard' allowing meaningful comparisons to be drawn regarding the commonly used performance metrics, such as sensitivity, specificity, and the overall accuracy of the new tool - and an ideal standard has been discussed earlier. The pathotyping tool described so far highlights the value of future controlled challenged studies to better understand the behaviour of these anomalous isolates under controlled, standardised conditions.

Objective 4. Historically serotyping has commonly been used as a proxy for virulence potential [17]. Considering the assignment to molecular 'serotypes' 1-10, 14 and 1/2 as markers of disease-association predicted the correct classification of isolates to the observed phenotypic groups with a superior sensitivity (0.91; i.e. the proportion of isolates actually recovered from systemic sites and predicted to be disease isolates) in comparison to the pathotyping tool (0.80; *p*-value = 0.00468). However, while virulence does appear to differ between *S. suis* serotypes it also differs within the serotypic groups [17, 53, 178, 400] – a feature also seen in the isolates of the original training collection recovered from pigs on farms in England and Wales (Table 3.1 and Figure 3.1). Indeed, even non-encapsulated strains have been recovered from cases of

invasive disease and described as virulent in experimental infection models [134, 135], indicating that the capsular polysaccharide is not a good indicator of virulence [11]. Therefore, it is unsurprising to see the trade-off for a high sensitivity is a statistically worse positive predictive value (i.e. the proportion of isolates predicted to be disease isolates that were actually recovered from a systemic site; 0.80 vs. 1.0, *p*-value = 0.00016) and specificity (i.e. the proportion of isolates actually recovered from the upper respiratory tract of a healthy pig and predicted to be non-disease isolates; 0.77 vs. 1.0, *p*-value = 0.00006). Indicating, that while the use of serotype as a proxy for virulence is very sensitive it could also be contributing to unnecessary prophylactic use of antibiotics in food and drinking water.

Molecular 'serotyping' methods [50, 68, 70, 74, 75] targeting serotype-specific differences in the *cps* genes have reduced the scope for user error/interpretation associated with traditional serotyping techniques, such as capillary precipitation. However, the first round of the Liu *et al.* [74] mPCR serotyping scheme requires a combination of 20 primers to target the most common *S. suis* serotypes (1-10, 14 and 1/2) associated with clinical disease. In comparison, the mPCR assay of the newly-described pathotyping tool (mPCR and GLM) is considerably simpler using just eight primers (of which two are for a *S. suis* species-specific marker), to which additional markers/primers could easily be added to as genome analyses are expanded to include isolates from important pig producing countries other than the UK. In fact, only using molecular serotypes 1, 2, 14 and 1/2 (the most common serotypes associated with *S. suis* disease in the UK) predicts disease-association with a sensitivity significantly worse than the mPCR pathotyping tool (*p*-value <0.05). However, in order to draw fair and robust comparison to published tools serotypes 1-10, 14 and 1/2 (targeted by the first of a total of four mPCR assay described by Liu *et al.* [74]) were considered markers of disease-association.

As discussed earlier differences in virulence not only between *S. suis* serotypes but also within the serotypic groups limits the interpretation of epidemiological have been well documented [53, 178, 400], and limits the interpretation of epidemiological studies that include only

serotyping. Indeed, other markers have been suggested to characterise and subtype *S. suis* isolates, one such approach for the identification of virulent strains is the detection of the 'virulence-associated' factors EF [329], MRP [345] and the thiol-activated haemolysin SLY. Using a GLM fit to prevalence data of the 'virulence-associated' genes *epf*, *mrp* and/or *sly* as a predictor of disease association also predicted the correct classification of isolates with a superior sensitivity (0.93) in comparison to the mPCR pathotyping tool (0.80; *p*-value =0.00270). However, again the trade-off for a high sensitivity was a statistically worse positive predictive value (0.50; *p*-value = 4.77396e-15 - no better than chance) and a specificity of just 0.09 (*p*-value = 0.00006). Indicating, like serotyping, that using the detection of these genes as a proxy for disease is capturing all invasive disease-associated isolates but it comes at the cost of a large number of false-positives. Despite first being associated with *S. suis* disease in the early 1990s [178, 329, 367, 369] and the genotyping of *epf* and *mrp* being incorporated into the routine diagnosis procedure for *S. suis* in several laboratories, the exact contribution of proteins EF and MRP to virulence still remains unclear [131]. Further confounding evidence indicates the use of *epf*, *mrp*, and *sly*, considered to be 'virulence-associated' markers (rather than virulence factors *per se*), is limited to mainly European countries and not worldwide [129, 408, 409].

Lastly, the pathotyping tool was also compared to MCG sequence typing [115, 116], one of the most recent methods used to characterise *S. suis* isolates. Here MCG sequencing typing was found to perform with a significantly worse sensitivity (0.80 vs. 0.50; *p*-value = 0.00001) and worse negative predictive value (0.68 vs. 0.83; *p*-value = 0.00001). During its design all highly virulent isolates and epidemic isolates tested were reported to be assigned to MCG group 1 [115], an observation also seen for most of the invasive disease-associated isolates in the original training collection (Figure 4.11). Assignment to MCG group 1 was used as a binary classifier to indicate disease association and showed our approach targeting the accessory-genome and the large amounts of variation in gene prevalence that can exist between isolates of the same species to be superior for the purpose of pathotyping *S. suis*.

Objective 5. To determine if the pathotyping tool could be used to screen oral fluid and swabs of material scraped from the palatine tonsils of pigs for invasive disease-associated *S. suis* strains, total genomic DNA was extracted from experimental samples without culture and first isolating single colonies/clones and used as mPCR assay/tool templates. *S. suis* is considered endemic in pig populations, but despite widespread colonisation few invasive disease-associated isolates are recovered from the upper respiratory tract [141]. In pig herds infected with multiple strains of *S. suis* sampling error, taking one or two samples or sampling from only one or two pigs within a herd, may result in failure to identify a strain associated with a recent outbreak of disease herds [99, 121]. The reliance of pathogen detection on the culture of microorganisms and isolation of single colonies/clones is a major drawback that creates a crippling bottleneck, and if removed has the potential to improve the true positive rate (sensitivity) of *S. suis* surveillance programs – a feature that would be of considerable benefit to the swine industry.

S. suis is a very successful coloniser of mucosal surfaces and the upper respiratory tract, in particular the deep crypts of the palatine tonsils, considered the natural habitat of the organism [11]. The palatine tonsils are also considered a potential site *S. suis* uses to gain entry to the cardiovascular or lymphatic system and is arguably the best site to sample pigs [70, 85]. However, in comparison, oral fluid is much easier to recover from pigs! In agreement with reports of *S. suis* being endemic in all sampled pig populations the mPCR amplicon of size 722 bp that corresponded to the *S. suis* species specific marker was produced in all groups of swabs sampled. Surprisingly, mutually exclusive disease-associated (SSU1589; 347 bp) and non-disease associated (SSUST30534; 892 bp) mPCR amplicons were frequently produced from the total genomic DNA extracted from a single swab, all of which were recovered form herds not showing obvious signs of disease, on farms in England and Wales without a history of streptococcal disease. Such an observation indicates a higher prevalence of invasive disease associated strains than previously described in healthy pigs, and the carriage of both disease-associated and non-disease associated *S. suis* strains in the same pig. Alternatively, this

observation indicates the detection of a strain possessing all four pathotyping markers, something that was obsevered in the original training collection or out-of-sample test collection.

In conclusion, this chapter describes the design and evaluation of a pathotyping tool (mPCR and GLM) for *S. suis* that predicted correct classification of 84% (accuracy) of the isolates in the original training collection in comparison to the observed clinical metadata, managing to predict the correct classification of 91% (sensitivity) of invasive disease-associated isolates (Table 4.2(iii)). Out-of-sample forecasting using a collection of previously uncharacterised *S. suis* isolates unsurprisingly produced worse performance metrics (sensitivity = 80%, accuracy = 80%, specificity = 100%, positive predictive value = 100% and negative predictive value = 83%), for reasons discussed earlier. Regardless, these performance metrics are indicative of a well-rounded binary diagnostic test. In comparison to other commonly used methods to characterise and sub-type *S. suis* the pathotyping tool performed with a worse sensitivity. However, the trade-off for a significantly superior sensitivity was, in the case of the virulence-associated genes *epf*, *mrp* and *sly*, a positive predictive value no better than chance (Exact binomial test *p*-value = 1). While false negative rate (1-sensitivty) was chosen as the most valuable performance metric for pathotyping *S. suis* (with a view to establish and maintain a pig population free of invasive disease associated isolates) sacrificing other performance metrics such as PPV and specificity would result in the unnecessary removal of animals from a herd. Therefore, the mPCR pathotyping tool is a good proof of principal that with the inclusion of additional biomarkers could become a very powerful diagnostic tool for *S. suis*. Interestingly, the frequency with which the disease-associated marker (SSU1589) was detected in material from the upper respiratory tract of pigs was unexpected and is explored further in chapter 5, as removal of the culture and isolation of single colonies/clones from the current diagnosis procedure would be of considerable benefit to the swine industry.

Chapter 5

Screening for disease-associated strains of *Streptococcus suis* on the tonsils of pigs on UK farms using quantitative real-time PCR

5.1 Introduction

S. suis is considered endemic in pig herds, often without causing signs of clinical disease [142, 565, 566]. Some strains act as primary pathogens, causing encephalitis and meningitis, although others are thought to act as opportunistic or secondary pulmonary pathogens [121]. The economic impact of *S. suis* infection in countries with an established swine industry is substantial, estimated to be more than $300 million dollars per annum to the swine industry in the US alone [11, 85, 567]. It is difficult to estimate the true cost to swine production worldwide as losses include i) mortality and medical treatment, ii) labour costs associated with nursing of sick pigs, iii) costs of preventative measures, as well as, iv) reduced product yields due to impaired growth rates of pigs [1]. The prophylactic use of antibiotics in food and drinking water has proven unsuccessful in controlling disease. Antibiotics are becoming less effective because of an increase in resistance among *S. suis* isolates, and blanket antibiotic use is now less accepted because of government pressure on the swine industry to reduce antimicrobial use [21, 22]. Other popular control strategies, such as the development of effective vaccines are hindered by the number of serotypes, the lack of a comprehensive understanding of *S. suis* virulence-factors, differences in virulence not only among serotypes but also within the different serotypes and a lack of serotype cross-protection.

Increased awareness of *S. suis* as a zoonotic pathogen has directed attention to the role of clinically healthy pigs as a reservoir for disease-causing strains, where there has been little focus on potential predisposing differences in prevalence and on diversity of carriage. High levels of genomic diversity between isolates of *S. suis*, combined with evidence for high levels of recombination indicate potential for emergence of novel disease-associated strains [141]. In fact, because *S. suis* undergoes such extensive recombination widespread across the genome, an increase in virulence anywhere in the world could have a global impact over a short timescale [141]. As a result, there is importance to developing a more comprehensive understanding of *S. suis* population diversity in the upper respiratory tract of clinically healthy pigs.

Despite widespread colonisation, relatively few disease-associated isolates of *S. suis* are recovered from the upper respiratory tract of pigs [141]. In pig herds infected with multiple strains of *S. suis*, sampling bias (taking one or two samples, or sampling from only one or two animals in a herd) may result in failure to identify a strain associated with a recent outbreak; a feature especially problematic in endemic herds [99, 121]. Reliance of pathogen detection on the culture of microorganisms and the isolation of single colonies/clones is a major drawback of pathogen detection. Detection of *S. suis* using classic culture-based techniques is simple enough [11]. However, the laborious and time-consuming process of picking single colonies from solid agar plates creates a crippling bottleneck that if removed has the potential to improve the sensitivity of *S. suis* surveillance programs, and would be of considerable benefit to the swine production worldwide.

In chapter 4, evaluation of the newly defined pathotyping tool on a collection of 138 previously uncharacterised *S. suis* isolates performed with a sensitivity of 0.80, specificity of 1.0, positive predictive value of 1.0 and negative predictive value of 0.83 (assuming the observed clinical phenotype to be the 'gold-standard'). Attempts to then use the *S. suis* pathotyping tool to screen swabs of material scraped from the palatine tonsils of pigs without culture and first isolating single colonies produced an interesting result - the production of both invasive disease-associated and non-disease associated mPCR amplicons from the same experimental sample. Detection of more than one strain of *S. suis* on the palatine tonsils of a pig is unsurprising, and supports published observations of more than one serotype, sequence type and/or 'virulence-associated' factor profile being found when multiple samples are taken from a single animal [568, 569]. However, what *is* surprising was to detect so frequently the invasive disease-associated genetic markers (and infer invasive disease-associated strains of *S. suis*) in pig herds not showing obvious signs of disease on farms in England and Wales without a history of streptococcal disease. This unexpected finding prompted investigation of the hypothesis that the prevalence of invasive disease-associated isolates of *S. suis* on farms in England and Wales free of obvious signs of infection is more common than currently thought.

5.1.1 Quantitative real-time PCR

Quantitative real-time PCR (qPCR) provides fast, precise and accurate results, and like 'traditional' PCR has become an established method for molecular diagnostics, in particular pathogen detection [535, 570, 571]. However, unlike 'traditional'/endpoint PCR that requires laborious post reaction processing (gel electrophoresis and image analysis), qPCR uses fluorescence-based methods to detect the accumulation of amplicons during the exponential amplification phase of the reaction after each thermal cycle in "real-time". A number of fluorescence-based qPCR chemistries exist, the most common use non-specific intercalating dyes such as SYBR Green I, or sequence-specific hydrolysis probes such as a dual labelled (reporter/quencher) TaqMan probe. In brief, both approaches are designed to generate fluorescence during the PCR, taking advantage of the linear relationship between detectable fluorescence signal and the amount of DNA. Such an approach provides researchers with a more accurate capacity to detect and quantify minute amounts of target nucleic acids, which can be used to determine the initial quantities of targets with great precision.

5.1.1.1 Quantitative real-time PCR using intercalating dyes

The simplest form of fluorescence-based qPCR uses DNA-binding dyes. Intercalating dyes are inexpensive (in comparison to TaqMan probes), accounting for their popular use. SYBR Green I is the most commonly used, and has an excitation and emission maxima of 494 nm and 521 nm respectively. SYBR Green I, preferentially binds double-stranded DNA over single-stranded DNA. Fluorescent in its own right, SYBR Green I, in the presence of double-stranded DNA intercalates with the minor groove of the DNA double helix. This alters the structure of the dye and excitation of DNA-bound SYBR Green I produces a much stronger fluorescent signal than unbound dye. The advantages of this approach are low cost and simplicity, ideal for research based pilot studies. However, SYBR Green I non-selectively binds to all double-stranded DNA including both target and non-target amplicons requiring post-reaction melt curve analysis.

Quantitative real-time PCR using an intercalating dye was chosen for a pilot investigation into the prevalence of invasive disease-associated and non-disease associated isolates of *S. suis* in material scraped from the palatine tonsils of clinically healthy pigs on farms in England and Wales. Probe-based chemistries are associated with superior specificity. However, a recurring feature of *S. suis* genome analyses is a high level of recombination within the species, widespread throughout the genome [141]. It is acknowledged that using a probe-based approach would increase specificity and increase confidence that qPCR amplicons and corresponding C_t values are derived from *S. suis* strains, but due to high levels of recombination such an approach could lead to a number of false negatives and an inaccurate representation of the true prevalence of invasive disease-associated strains in herds not showing obvious signs of *S. suis* disease.

This chapter describes the use of qPCR, targeting the genetic markers used to pathotype *S. suis*, in order to screen swabs of material scraped from the palatine tonsils of pigs on three farms (sampled at two time points, five weeks and 20 weeks of age) in England and Wales for invasive disease-associated strains of *S. suis*. Once the qPCR technique has been established this chapter describes three comparisons, drawn against i) the ambient temperature at the time of sampling, ii) pig age and iii) culture-based methods of detection in use today. Of which, the most interesting finding suggests that culture-based approaches to surveillance may underestimate the true positive rate of *S. suis* on the palatine tonsils of pigs on farms in England and Wales.

5.1.2 Chapter objectives

1. To determine whether the newly defined pathotyping markers described previously in this thesis (chapters 3 and 4) can be used to screen material scraped from the palatine tonsils of pigs for *S. suis* without culture and first isolating single colonies/clones
2. To determine the prevalence of *S. suis* strains with the potential to cause invasive disease on the tonsils of pigs not showing obvious signs of streptococcal disease
3. To determine whether picking three single colonies based on morphology and haemolysis on blood agar plates is enough to produce an accurate evaluation of *S. suis* on the tonsils of pigs not showing obvious signs of streptococcal disease

5.2 Results

5.2.1 Establishment of quantitative real-time PCR to screen tonsil swabs for disease- and non-disease associated strains of *Streptococcus suis*

Quantitative real-time PCR was used to assess the prevalence of invasive disease-associated and non-disease associated strains of *S. suis* on the palatine tonsils of pigs. The online software, primer3 version 4.0.0 was used to design qPCR primers (Table 5.1). Primers were designed to have similar physical characteristics, enabling simultaneous amplification under the same thermal cycling conditions. Primer length (18-24 bp), GC content (~50%), melting temperature (~60 ˚C; and within 1 ˚C of each other), and expected amplicon size (80-200 bp; not exceeding 400 bp) were based on the manufacturer's recommendations for primer design using the SensiFAST™ SYBR® No-ROX Kit (Bioline). Two primer pairs were designed per target. All primers were queried against the NCBI non-redundant nucleotide database to check for matches to other non-*S. suis* bacterial species, and redesigned if similarities to other organisms found.

Table 5.1. Primer sequences for quantitative real-time PCR to screen tonsil swabs for disease-associated strains of *Streptococcus suis*.

Two primer pairs were designed per target. The online software, primer3 version 4.0.0 was used to design qPCR primers. Primers were designed to have similar physical characteristics, enabling simultaneous amplification under the same thermal cycling conditions. Primer length (18-24 bp), GC content (~50%), melting temperature (~60 ˚C; and within 1 ˚C of each other), and expected amplicon size (80-200 bp; not exceeding 400 bp) were based on the manufacturer's recommendations for primer design using the SensiFAST™ SYBR® No-ROX Kit (Bioline). All primers were queried against the NCBI non-redundant nucleotide database to check for matches to other non-*S. suis* bacterial species, and redesigned if similarities to other organisms found.

Primer pair	Phenotype	Sequence (Forward/Reverse)
SSU1589:0874F/0975R	Disease-associated	TCAGGCCAGTTTATATATGGGA/AGCAGGCGAATCAAGAGTTG
SSU1589:0958F/1101R	Disease-associated	ACTCTTGATTCGCCTGCTTTTG/AGTTTCTGTGTGGATTCTTTTAGC
SSUST30534:0526F/0633R	Non-disease associated	GATTGGGAAAGTAAGTGGGATG/GAAGAATCGGCAGCATGAC
SSUST30534:1236F/1319R	Non-disease associated	AGCAGTGTATGGAGAGCTCG/TGAGTAGGGGTCCAAATTTGTG
SSU0577:0462F/0562R	*Streptococcus suis*	CAACGTGCAGAGGACATTATGG/TCATTGCGAGCCTCACGAAG
SSU0577:0544F/0683R	*Streptococcus suis*	TTCGTGAGGCTCGCAATGAC/ACCTGACAACTGATCTAGACCG
SSU1462:0032F/0178R	*Streptococcus suis* (*recN*)	GCTGACCTGCTCGTTCAACC/ATGGTGGAACTGTCGATGATG
SSU1462:0156F/0238R	*Streptococcus suis* (*recN*)	AACATCATCGACAGTTCCACC/GCAATCGTCTCATGCAATTGG

5.2.1.1 Primer placement

Limited to a 72-well qPCR rotor, the work described in this chapter was restricted to targeting the newly defined *S. suis* species-specific marker (SSU0577) and one each of the disease-associated (SSU1589) and non-disease associated (SSUST30534) markers targeted in the pathotyping tool described in chapter 4. Of the two disease-associated pathotyping markers described in chapter 4, SSU1589 was chosen for qPCR in favour of SSU0207 due to superior sequence homology in invasive disease-associated isolates recovered from pigs on farms in England and Wales. In addition, primers were also designed to target the core-metabolic gene *recN*, encoding a recombination/repair protein, to allow further evaluation of SSU0577 as a robust species-specific marker for *S. suis*. Previously, the *recN* gene has been used to differentiate members of the genus *Streptococcus* based on partial gene sequencing [572], and most recently as the *S. suis* species-specific marker in a mPCR for the identification of six clinically relevant streptococci [573].

Where possible, qPCR primers were designed to target the same, albeit shorter (80-200 bp), regions as the mPCR pathotyping tool described in chapter 4. However, for the invasive disease-associated marker SSU1589 it was not possible to design qPCR primers that targeted the same region used for mPCR, due to the presence of two independent direct repeats and the conserved 5' end (1-613 bp) of the protein encoding sequence across *S. suis* isolates regardless of observed clinical phenotype. Figure 5.1 shows the placement of qPCR primers SSU1589:0958F/1101R, designed to target the 3' end of the protein-encoding sequence present in invasive disease-associated isolates and 'missing' from non-disease associated isolates recovered from pigs on farms in England and Wales.

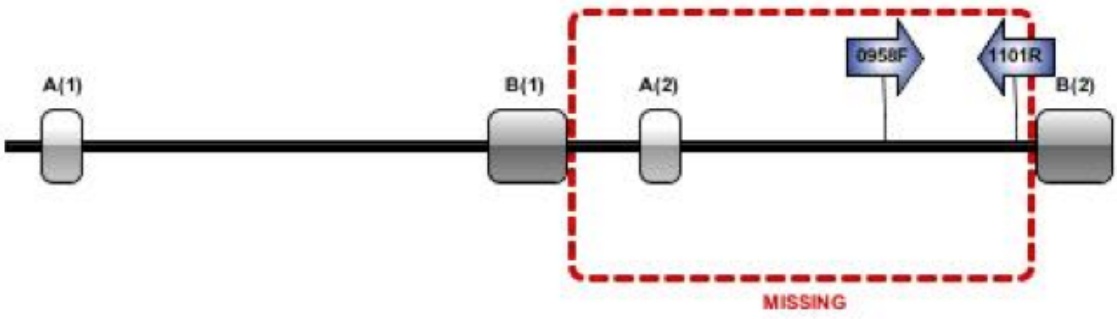

Figure 5.1. Placement of quantitative real-time PCR primers targeting the *Streptococcus suis* invasive disease-associated pathotyping marker SSU1589.

Schematic representation of SSU1589 (1206 bp) showing the placement of qPCR primers SSU1589:0958F and SSU1589:1101R. Primer design targeted the 3' end (614-1206 bp) of SSU1589 that is conserved in invasive disease-associated isolates but missing from non-disease associated isolates of *S. suis* recovered from pigs on farms in England and Wales. Primer design intentionally avoided the repeat regions ($A_{(1)}$(44 bp), $A_{(2)}$, $B_{(1)}$(82 bp) and $B_{(2)}$), as well as, the 5' end of SSU1589 (1-613 bp) that is conserved, regardless of observed clinical phenotype, across *S. suis* isolates recovered from pigs on farms in England and Wales.

5.2.1.2 Primer efficiencies

Robust and precise qPCR experiments are associated with high PCR amplification efficiencies [535]. PCR efficiencies are a measurement of the rate at which a PCR amplicon is generated, and are dependent on the primers used. Amplification efficiencies (or 'primer efficiencies'), per primer pair, were calculated over the range ~1x 10^0 - 1x 10^7 genome copies per reaction using a calibration curve (Figure 5.2 and Figure 5.3) and the equation:

$$E = 10^{(-1/m)} - 1$$

where E denotes PCR amplification efficiency; and *m* the slope of the linear regression line, when the logarithm of the initial template concentration (the independent variable) is plotted on the X-axis and the C_t (the dependent variable) plotted on the Y-axis

Two primer pairs were designed for each qPCR marker. Table 5.2 shows the primer efficiencies calculated for all qPCR primer pairs designed to investigate the prevalence of invasive disease-associated and non-disease associated strains of *S. suis* on swabs of material scraped from the palatine tonsils of pigs on farms in England and Wales. One primer pair, per marker, was selected (black) for downstream qPCR experiments based on primer efficiencies. Primer pairs with amplification efficiencies most similar to one another, and between the range considered acceptable for robust and precise qPCR (rule-of-thumb: 95%-105%) were selected.

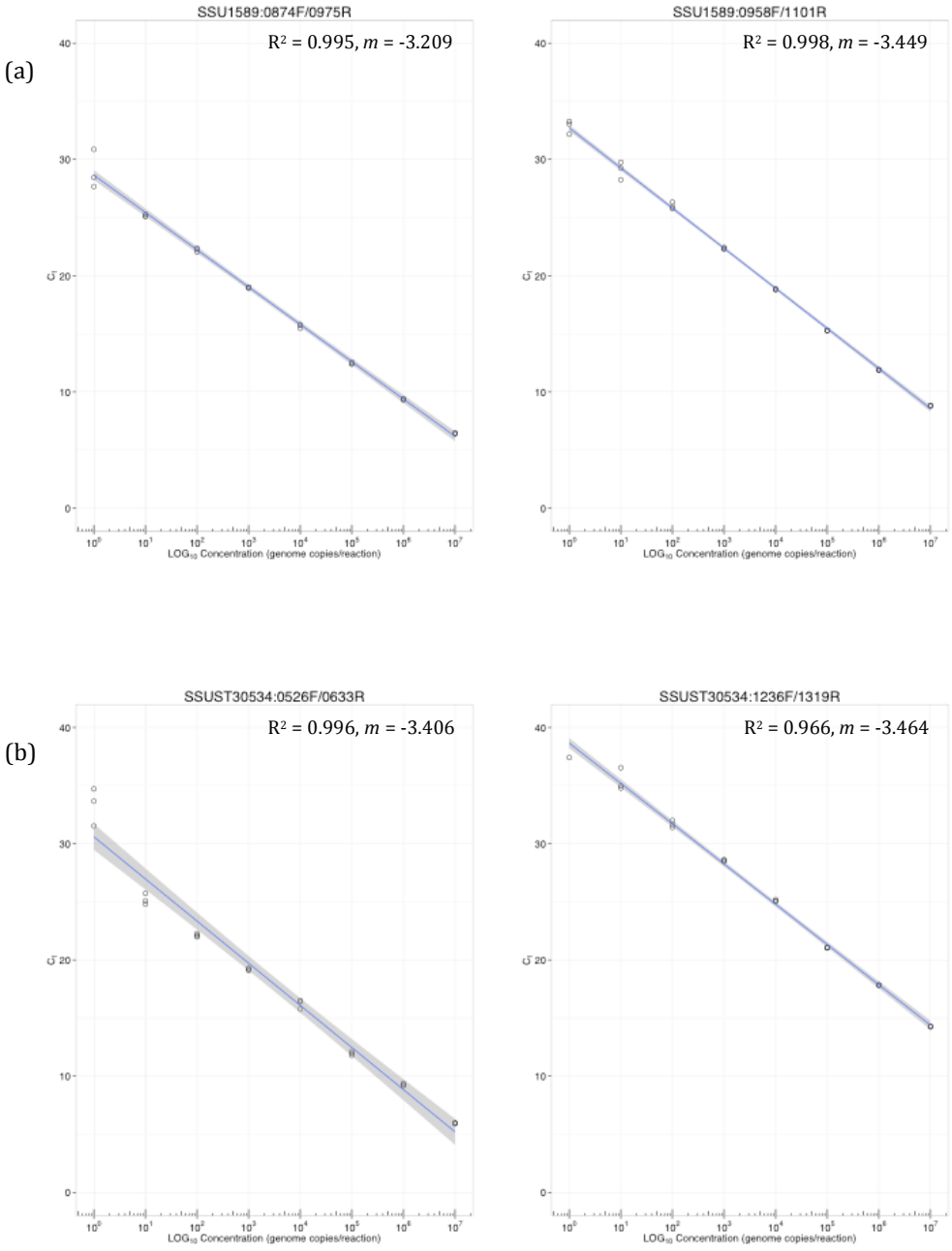

Figure 5.2. Standard curves used to calculate amplification efficiency for invasive disease-associated and non-disease associated quantitative real-time PCR primers.

Visual representation used to calculate primer efficiencies for qPCR primers designed to screen tonsil swabs for invasive disease-associated or non-disease associated strains of *S. suis*. Two primer pairs were designed per target, (**a**) SSU1589 and (**b**) SSUST30534. Amplification efficiency, per primer pair, was then calculated over the range ~1x 10^0 - 1x 10^7 genome copies per reaction by means of a calibration curve and the equation: $E = 10^{(-1/m)} - 1$; where E denotes PCR amplification efficiency, and *m* the slope of the linear regression line when the logarithm of the initial template concentration (the independent variable) is plotted on the X-axis and the C_t (the dependent variable) plotted on the Y-axis.

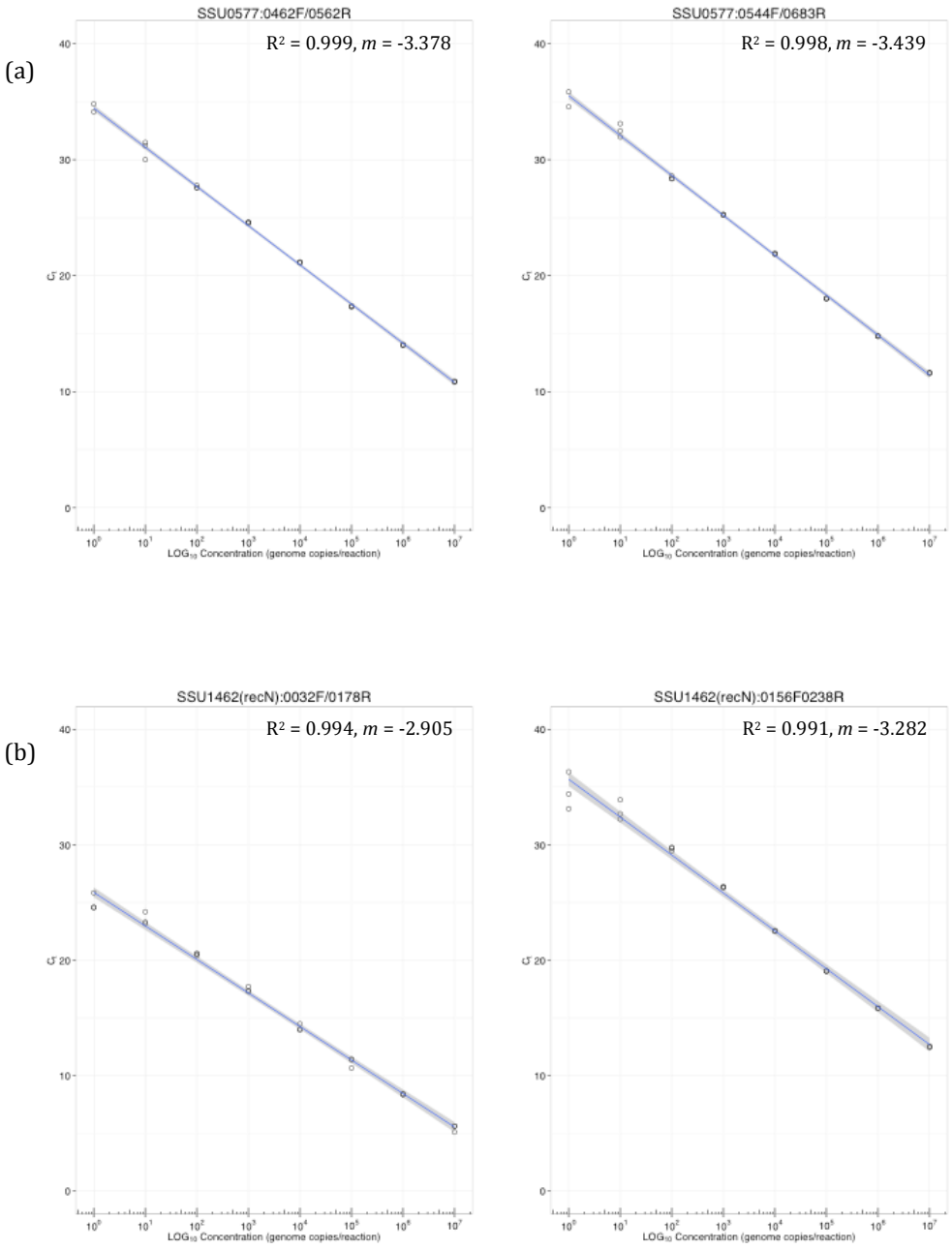

Figure 5.3. Standard curves used to calculate amplification efficiency for *Streptococcus suis* species-specific quantitative real-time PCR primers.

Visual representation used to calculate primer efficiencies for qPCR primers designed to screen tonsil swabs for the *S. suis* species. Two primer pairs were designed per target, (**a**) SSU0577 and (**b**) *recN*. Amplification efficiency, per primer pair, was then calculated over the range ~1x 10^0 - 1x 10^7 genome copies per reaction by means of a calibration curve and the equation: $E = 10^{(-1/m)} - 1$; where E denotes PCR amplification efficiency, and *m* the slope of the linear regression line when the logarithm of the initial template concentration (the independent variable) is plotted on the X-axis and the C_t (the dependent variable) plotted on the Y-axis.

Table 5.2. Amplification efficiencies for quantitative real-time PCR primer pairs.

Amplification efficiencies calculated for qPCR primer pairs designed to investigate the prevalence of *S. suis* on swabs of material scraped from the palatine tonsils of pigs on farms in England and Wales. Originally two primer pairs were designed per target. The four primer pairs (one per target, black) with most similar amplification efficiencies were selected for downstream qPCR experiments. Amplification efficiency, per primer pair, was calculated over the range ~1x 10^0 - 1x 10^7 genome copies per reaction by means of a calibration curve and the equation: $E = 10^{(-1/m)} - 1$; where E denotes PCR amplification efficiency, and *m* the slope of the linear regression line when the logarithm of the initial template concentration (the independent variable) is plotted on the X-axis and the C_t (the dependent variable) plotted on the Y-axis.

Primer pair	Phenotype	Amplicon size (bp)	R^2	*m*	Efficiency	qPCR
SSU1589:0874F/0975R	Disease-associated	101	0.995	-3.209	1.05	-
SSU1589:0958F/1101R	Disease-associated	143	0.998	-3.449	0.95	Yes
SSUST30534:0526F/0633R	Non-disease associated	107	0.996	-3.406	0.97	Yes
SSUST30534:1236F/1319R	Non-disease associated	83	0.966	-3.464	0.94	-
SSU0577:0462F/0562R	*Streptococcus suis*	100	0.999	-3.378	0.98	Yes
SSU0577:0544F/0683R	*Streptococcus suis*	139	0.998	-3.439	0.95	-
SSU1462:0032F/0178R	*Streptococcus suis* (*recN*)	146	0.994	-2.905	1.21	-
SSU1462:0156F/0238R	*Streptococcus suis* (*recN*)	82	0.991	-3.282	1.02	Yes

5.2.1.3 Standard curve equations

Standard curve equations, per primer pair, were used to estimate the unknown number of genome copies in experimental samples (section 5.2.2 onwards). A standard curve was generated, for each primer pair, over the range ~1x 10^1 - 1x 10^5 genome copies per reaction. Each standard curve was generated from four biological repeats using independent single-use qPCR premixes, and each biological repeat performed in triplicate per 10-fold serial dilution of template DNA. The concentration of DNA was linked to the number of thermal cycles (C_t) required to reach the fluorescence threshold at which point individual amplification reactions contained identical amounts of DNA using the standard curve equation:

$$\text{Concentration} = 10^{a(\text{Ct})+b}$$

where a denotes the slope; and b the Y-axis intercept of the linear regression line, when the logarithm of the initial template concentration is plotted on the X-axis and the C_t plotted on the Y-axis

Limited to a 72-well qPCR rotor, it was not possible to generate a full standard curve, per primer pair, for each rotor run. Instead, adaptation of a method described by Grant *et al.* [536] was used to calculate robust estimates of the standard curve equation for each primer pair, that were then used to both assess the performance/variability between rotor runs and estimate the number of genome copies in experimental samples. Figure 5.4(a) shows the four biological repeats used to calculate estimates of the standard curve equation for the invasive disease-associated marker primer pair SSU1589:0958F/1101R. The "Auto-Find Threshold" function of Rotor-Gene Q analysis software version 2.3.1 (Qiagen) was used to find the fluorescence threshold that delivered the best-fit of each standard curve to the samples defined as standards. However, to draw direct comparison between C_t values from separate rotor runs, the same fluorescence threshold must be used to analyse each rotor run [537]. The mean fluorescence threshold, for each primer pair, was used to re-analyse the four biological repeats of each standard curve. A regression line, per biological repeat, was fitted and used to generate estimates of the variables a and b, which were then averaged across the repeats to give a final standard curve equation per primer pair. These robust estimates of the standard curve

equation, per primer pair, were then used to predict the unknown number of genome copies in an experimental sample. Figure 5.4(b) shows the same four biological repeats (for primer pair SSU1589:0958F/1101R) re-analysed using the mean fluorescence threshold, from which robust estimates of the standard curve equation variables *a* and *b* were calculated. Table 5.3 shows the standard curve equation, per primer pair, used to estimate the number of genome copies in experimental samples.

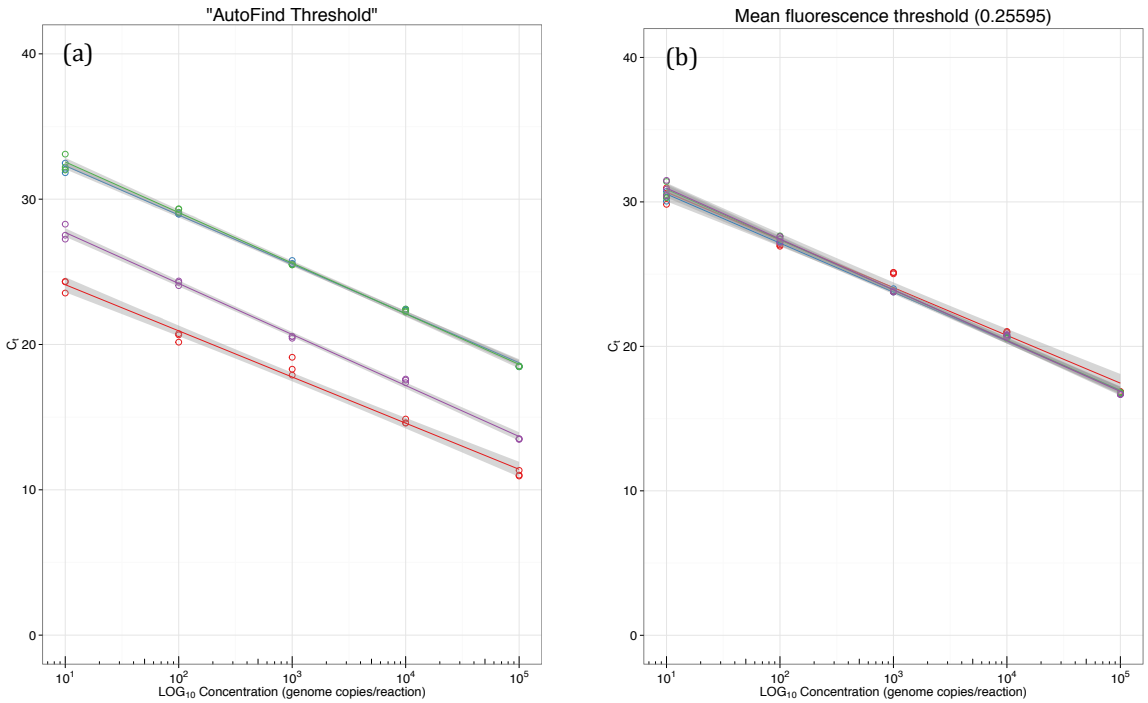

Figure 5.4. Visual representation of the use of the mean florescence threshold to generate robust estimates of the standard curve equation variables *a* and *b*.

(**a**) Four biological repeats of the standard curve for primer pair SSU1589:0958F/1101R generated over the range 10^5 - 10^1 genome copies per reaction. Each dilution, per biological repeat, was performed in triplicate. The "AutoFind Threshold" function of Qiagen Rotor Gene Q software version 2.3.1 was used to find the fluorescence threshold that delivered the best-fit of each standard curve. To draw direct comparison between C_t values from separate rotor runs the same fluorescence threshold must be used to analyse each rotor run. (**b**) Reanalysis of the four biological repeats for primer pair SSU1589:0958F/1101R using the mean fluorescence threshold (0.25595). A single regression line was fitted, per biological repeat, and used to generate estimates of the standard curve equation variables *a* and *b*; which were averaged and used to generate a final standard curve equation from which the unknown number of *S. suis* genome copies in experimental samples was estimated.

Table 5.3. Final standard curve equations for quantitative real-time PCR primer pairs.

Estimates of the standard curve equation variables *a* and *b*, per primer pair, generated using the mean fluorescence threshold. Standard curves were generated from four biological repeats over four rotor runs. A regression line, per biological repeat, was used to generate estimates of the variables *a* and *b*, which were then averaged across the standards to give a final standard curve equation per primer pair. The standard curve equation, per primer pair, was then used to predict the unknown number of genome copies in an experimental sample.

Primers	Phenotype	Mean fluorescence threshold	*a*	*b*	Final standard curve equation
SSU1589:0958F/1101R	Disease-associated	0.25595	-0.292	9.987	Conc. = $10^{(-0.292(Ct)+\ 9.987)}$
SSUST30534:0526F/0633R	Non-disease associated	0.21078	-0.291	10.132	Conc. = $10^{(-0.291(Ct)+10.132)}$
SSU0577:0462F/0562R	*Streptococcus suis*	0.20513	-0.296	10.013	Conc. = $10^{(-0.296(Ct)+10.013)}$
SSU1462:0156F/0238R	*Streptococcus suis* (*recN*)	0.26480	-0.294	10.249	Conc. = $10^{(-0.294(Ct)+10.249)}$

5.2.1.4 Quality control

Performance and variability between qPCR rotor runs was assessed using the method described by Grant *et al.* [536]. In brief, treating C_t values as random samples from a normal distribution, variability reflected in the confidence intervals was deemed to be an artefact of qPCR rotor runs. Standards of known concentration (10^3 genome copies per reaction; arbitrarily chosen) were used to pass/fail rotor runs based on the following criteria: to pass i) the standard deviation of C_t values must be less than 1.0, and ii) the mean C_t value must fall within +/- two standard deviations of the standard curve. Table 5.4 shows the C_t range, per primer pair, considered acceptable in order to pass qPCR quality control. If the criteria listed above were not met, qPCR rotor runs were failed and all experiments repeated. Individual experimental samples were also subject to quality control based on the following criteria: i) to pass the standard deviation of C_t values must be less than 1.0, if not experiments were repeated.

Table 5.4. C_t value limits considered acceptable to pass quantitative real-time PCR quality control. Limited to a 72-well quantitative real-time PCR rotor, it was not possible to generate a full standard curve, per primer pair, for each rotor run. Instead, standards of 10^3 genome copies per reaction were used to pass/fail rotor runs based on the following criteria: to pass i) the standard deviation of C_t values must be less than 1.0, and ii) the mean C_t value must have been within +/- two standard deviations of the standard curve. Mean C_t and standard deviation (sd) were calculated, per primer pair, from the four biological repeats of each standard curve.

Primers	Phenotype	Marker	Mean C_t	sd	-2sd	+2sd
SSU1589:0958F/1101R	Disease-associated	SSU1589	24.13	0.63	23.51	24.76
SSUST30534:0526F/0633R	Non-disease associated	SSUST30534	24.63	0.74	23.89	25.37
SSU0577:0462F/0562R	*Streptococcus suis*	SSU0577	23.86	0.48	23.38	24.34
SSU1462:0156F/0238R	*Streptococcus suis* (*recN*)	SSU1462	24.80	0.68	24.13	25.48

5.2.1.5 Limits of detection and quantification

Analytical sensitivity refers to the ability of an assay to accurately detect very low concentrations of a given substance. For qPCR, analytical sensitivity is typically expressed as the limit of detection, which is the lowest concentration that can be detected with reasonable certainty, usually 95% probability (theoretically, three copies per PCR is the most sensitive limit of detection possible) [535]. Limit of detection was determined by performing ten replicates per dilution over the range $\sim 10^0$ - 10^2 genome copies per reaction. The dilution with the lowest concentration meeting the following requirements was deemed to be the limit of detection: i) the standard deviation of C_t values must be less than 1.0 and ii) the number of replicates with positive detection must be greater than 95%. The mean C_t value of the lowest concentration was then used to calculate the limit of detection and expressed as a number of genome copies per reaction for each primer pair. Table 5.5 lists the limit of detection calculated for each qPCR primer pair. In comparison, the limit of quantification for qPCR is the lowest concentration that can be accurately quantified with 95% confidence [535]. The limit of quantification, per primer pair, is also included in Table 5.5.

Table 5.5. Limit of detection and quantification for quantitative real-time PCR primer pairs.

Estimates of the limit of detection, the lowest concentration of a given substance that can be detected with 95% probability, and the limit of quantification, the lowest concentration of a given substance that can be accuracy quantified with 95% confidence.

Primers	Phenotype	Limit of detection (genome copies per reaction)	Limit of quantification (genome copies per reaction)
SSU1589:0958F/1101R	Disease-associated	14	24
SSUST30534:0526F/0633R	Non-disease-associated	6	10
SSU0577:0462F/0562R	*Streptococcus suis*	13	28
SSU1462:0156F/0238R	*Streptococcus suis* (*recN*)	50	67

5.2.1.6 Pilot evaluation of the effect mixed template samples of closely related strains has on quantitative real-time PCR performance is minimal

Any molecular test for *S. suis* must be specific. The upper respiratory tract of pigs carries a diverse bacterial population, and being able to remove the culture and isolation of single colonies from the *S. suis* detection/diagnosis procedure would require any alternative assay to be capable of specifically detecting *S. suis* from a mixed sample. Accurately replicating the diversity of the upper respiratory tract *in vitro* is extremely difficult, if not impossible. Therefore, to test for non-*S. suis* amplification of qPCR markers, genomic DNA from field isolates of closely related Streptococcaceae, Pasteurellaceae and Alcaligenaceae commonly recovered from the upper respiratory tract of pigs were used as a panel of negative controls. No qPCR amplicons were produced by qPCR primer pairs targeting the pathotyping markers described in chapter 4. However, primer pairs targeting SSU1462(*recN*) did produce qPCR amplicons using template DNA from strains of *Streptococcus gallolyticus* and *Streptococcus pneumoniae*. Melt curve analyses indicated that qPCR amplicons could be differentiated from those corresponding to *S. suis* based on the higher temperature required for dissociation of dsDNA (data not shown; confirmed using uMELT).

The performance of qPCR primers with a mixed sample template was then assessed using two *S. suis* strains, P1/7 (invasive disease-associated) and LSS034 (non-disease associated). Mixed templates were created by combining genomic DNA of the two *S. suis* strains in different ratios: 1:1, 1:10, 1:100, 1:1,000 and 1:10,000, to reflect the hypothesis that invasive disease-associated strains are prevalent at low levels in comparison to non-disease associated strains in the upper respiratory tract of pigs. Figure 5.5 shows (a) qPCR amplification traces and (b) endpoint melt curve analysis for the mixed sample templates. All amplification traces and the endpoint melt curve analysis produced profiles consistent with those produced by qPCR when using template DNA from a single source (data not shown), indicating qPCR primers did not cross react with very closely related isolates (of the same species).

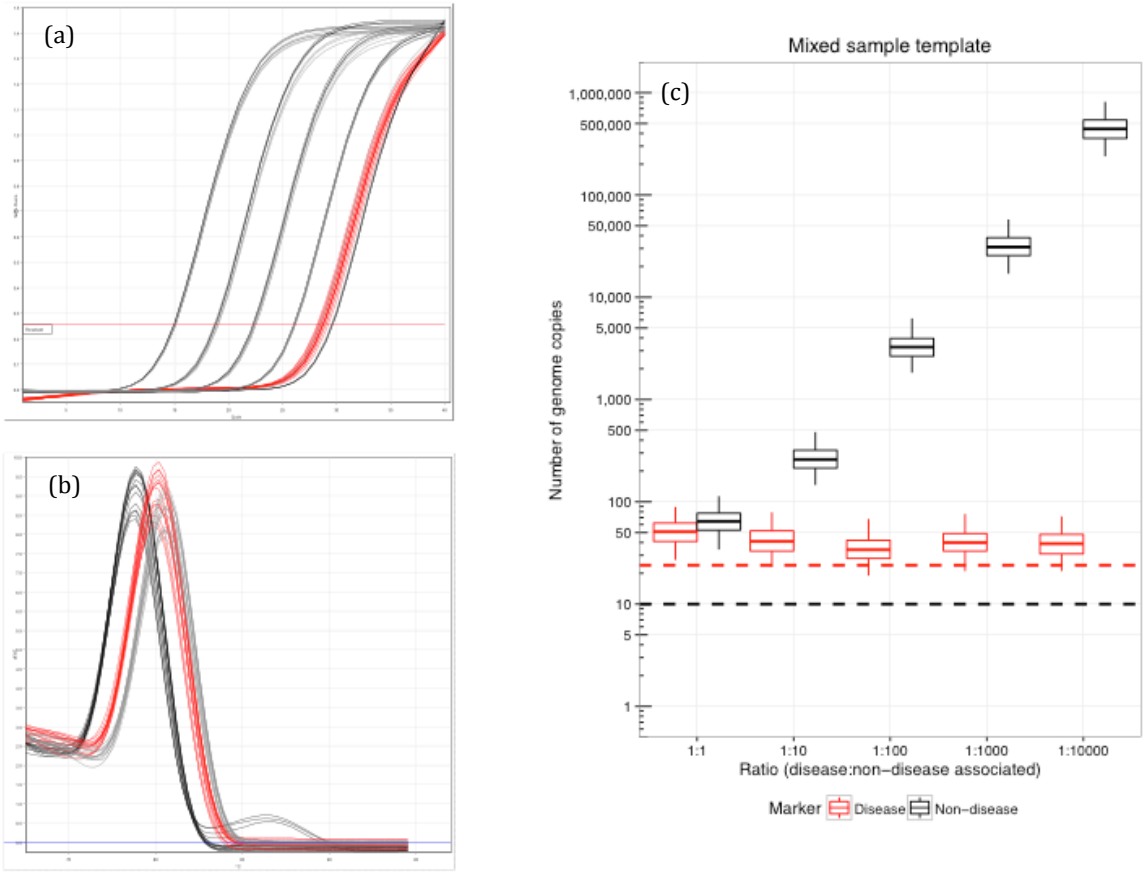

Figure 5.5. Pilot evaluation showing the effect mixed template samples of closely related *Streptococcus suis* strains has on quantitative real-time PCR performance is minimal.

(**a**) Amplification traces and (**b**) endpoint melt curve analysis of invasive disease-associated (red), non-disease associated (black) and *S. suis* species-specific (grey) qPCR amplicons, performed using Qiagen Rotor-Gene Q software version 2.3.1. All amplification traces and melt curves resembled profiles consistent with those produced by qPCR when using template DNA from a single source (data not shown). (**c**) Boxplot, implemented in the R package: *ggplot2*, of the number of invasive disease-associated and non-disease associated genome copies estimated from qPCR using mixed sample DNA templates of known concentration. Boxes represent the median, 25th and 75th percentiles, and whiskers 95% confidence intervals (2.5th and 97.5th percentiles respectively). Mixed templates were created *in vitro* by mixing genomic DNA from two *S. suis* strains, P1/7 (disease-associated) and LSS034 (non-disease associated) in different ratios. As few as 38 invasive disease-associated genome copies could be detected by in a mixed template sample of ~500,000 genome copies. The horizontal lines at 24 genome copies (red; disease associated) and 10 genome copies (black; non-disease associated) represent the qPCR limits of quantification per marker.

Figure 5.5 also shows (c) a boxplot of the estimated number of invasive disease-associated (red) and non-disease associated (black) genome copies per reaction, with 95% confidence intervals. Estimates of the number of invasive disease-associated genome copies were consistent across the five mixed sample templates (range: 33-49, mean: 41; aim: ~50 genome copies per reaction), and all within the 95% confidence interval (28-81 genome copies) calculated for 50 disease-associated genome copies detected using qPCR primers SSU1589:0958F/1101R. In comparison, estimates of the number of non-disease associated genome copies increased ~10-fold, consistent with estimates when using template DNA from a single source.

Detection of the reverse ratio (non-disease associated strains at low levels in comparison to invasive disease-associated strains) was also possible and produced very similar results. Table 5.6 shows the number of genome copies, per marker, estimated using qPCR and the corresponding ratio of non-disease associated: disease-associated genome copies per reaction, indicating very low copy numbers in a mixed sample of closely related strains can be detected by this method.

Table 5.6. Ratio of disease-associated: non-disease associated genome copies estimated using quantitative real-time PCR from a mixed template sample of known concentration.

The ratio of disease-associated: non-disease associated genome copies estimated from a mixed template sample. Mixed template samples were created *in vitro* by combining genomic DNA in different ratios from two *S. suis* strains, P1/7 (disease-associated) and LSS034 (non-disease associated). Mixed templates were created over the range ~50-500,00 genome copies per isolate. The number of genome copies, per marker, was estimated from C_t values using robust estimates of the standard curve equation for each primer pair.

	----- Estimates generated from quantitative real-time PCR data -----			
Ratio in (genome copies; disease: non-disease)	**Disease (genome copies)**	**Non-disease (genome copies)**	***Streptococcus suis* (genome copies)**	**Ratio out (genome copies; non-disease: disease)**
1:10,000	38	453,649	481,415	11,923:1
1:1,000	40	32,368	33,055	801:1
1:100	33	3,218	3,063	97:1
1:10	41	256	262	7:1
1:1	49	62	67	1:1
10,000:1	478,630	45	458,987	1:10,637
1,000:1	64,977	58	58,998	1:1,121
100:1	3,756	47	3,583	1:80
10:1	267	36	302	1:8
1:1	37	43	66	1:1

5.2.1.7 Two different species-specific markers detect *Streptococcus suis* on the tonsil swabs but estimate different bacterial load

Chapter 3 of this thesis describes the identification of SSU0577, one of the most conserved protein-encoding sequences in the *S. suis* core-genome. SSU0577 is then evaluated as a novel species-specific genetic marker for *S. suis* in chapter 4. To further evaluate SSU0577 as a robust species-specific marker its ability to detect and estimate the number of genome copies was compared against the core-metabolic gene *recN*. Amongst other important core-metabolic genes *recN* has been used to investigate the genetic relatedness among closely associated strains of the genus *Streptococcus* [572]. To draw comparison between the two markers all tonsil swabs were used as qPCR templates. Table 5.7 shows a 2x2 contingency table summarising the outcome of *S. suis* detection using i) SSU0577 and ii) SSU1462(*recN*), in the interest of fairness a total of 23 swabs have been removed from statistical analyses at this point because experimental samples did not pass qPCR quality control. Table 5.7 shows that qPCR targeting i) SSU0577 or ii) SSU1462(*recN*), agreed for 71% of the swabs ((62+29)/127). Of the other outcomes, *S. suis* was only detected on 13% of swabs using qPCR targeting SSU0577 (n=16), and in comparison 6% of swabs tested positive using qPCR targeting *recN* only (n=7).

Table 5.7. The number of tonsil swabs testing positive for *Streptococcus suis* using quantitative real-time PCR targeting species-specific markers SSU0577 and the core metabolic gene *recN*.

A 2x2 contingency table summarising the number of swabs (n=127) of material scraped from the palatine tonsils of pigs on farms in England and Wales from which *S. suis* was detected with 95% confidence.

		SSU1462(*recN*)	
		+	-
SSU0577	+	62	16
	-	7	29

In addition to using a binary approach (detection/no detection) the estimated numbers of genome copies per swab were also compared. Table 5.8 shows the test statistics and corresponding *p*-values used to identify statistically significant differences. qPCR targeting SSU1462(*recN*) consistently estimated significantly higher numbers of genome copies (used as a proxy for bacterial load), per tonsil swab, in comparison to SSU0577 (*p*-value <0.05). In section 5.2.1.6, it is reported that primer pairs targeting SSU1462(*recN*) produced qPCR amplicons using template DNA from strains of *Streptococcus gallolyticus* and *Streptococcus pneumoniae*. This was inferred to be the reason for the higher estimates of the number of genome copies on tonsil swabs, therefore, where necessary to draw comparison to a species-specific marker in this chapter all comparisons are hereafter drawn against the newly defined *S. suis* species-specific marker SSU0577.

Table 5.8. Test statistics and corresponding *p*-values used to idenitfy significant differences between bacterial load estimated by *Streptococcus suis* species-specific marker SSU0577 and *recN*.

Statistical tests were used to determine if estimates of the number of genome copies per experimental sample were the same when using different species-specific markers for *S. suis*, SSU0577 and the core metabolic gene *recN*. The Shapiro-Wilk Normality test assumes normality unless *p*-values are <0.05. As data was not well modelled by a normal distribution, the Student's t-test was inappropriate and, the non-parametric Wilcoxon Rank Sum test was used that assumes the true location shift is equal to 0, unless *p*-values are <0.05. All tests were performed using log10-transformed data.

Marker	Shapiro-Wilk Normality test (W; ***p*-value**)	Mean	Variance	Student's t-test (t; ***p*-value**)	Wilcoxon Rank Sum test (W; ***p*-value**)
SSU0577	0.87992; **0.00002**	-	-	-	1718.5; **0.00465**
SSU1462(*recN*)	0.94177; **0.00154**	-	-		

5.2.2 Quantitative real-time PCR can be used to detect *Streptococcus suis* on tonsil swabs without culture and first isolating single colonies/clones

To investigate if qPCR could be used to detect *S. suis* on tonsil swabs without culture and first isolating single colonies, material scraped from the palatine tonsils of pigs was resuspend in PBS and the DNA extracted (see section 2.2.4.1.1). Figure 5.6 (five-weeks) and Figure 5.7 (20-weeks) show the number of genome copies, per experimental sample, estimated using qPCR targeting the newly defined *S. suis* species-specific marker SSU0577 (identified as one of the most conserved protein-encoding sequences in the *S. suis* core-genome, see section 3.3.5). The data presented in Figures 5.6 and 5.7 show it is possible to detect very low (~14) to comparatively high (~20,000) numbers of genome copies in experimental samples using the qPCR-based approach described in this chapter. In fact, in agreement with other studies using culture-based techniques, it was possible to detect *S. suis* in experimental samples from all six groups sampled; a finding that supports opinion that *S. suis* is endemic in all pig populations [142, 565, 566].

Despite being detected in all six groups sampled, it was not possible to detect *S. suis* in all experimental samples. In Figure 5.6 and Figure 5.7, it is important to draw attention to the data presented as an '*x*', which has been done to indicate that estimates of the number of genome copies are not reported with 95% confidence. It is usual not to report estimates of detection below the limit of detection or to report estimates of the number of genome copies where qPCR rotor runs are deemed to have failed quality control. However, due to time restrictions, repeating all qPCR rotor runs deemed to have failed quality control was not possible. Inclusion of estimates of the number of genome copies unable to be reported with 95% confidence was done so as to help highlight potential trends and avoid presenting misleading reports of detection in this work. It has also served the purpose of highlighting additional work required in order for publication in a peer-reviewed journal. Importantly, although included in Figure 5.6 and Figure 5.7 all estimates of the number of genome copies in experimental samples deemed unable to be reported with 95% confidence have been excluded from all statistical analyses in this chapter.

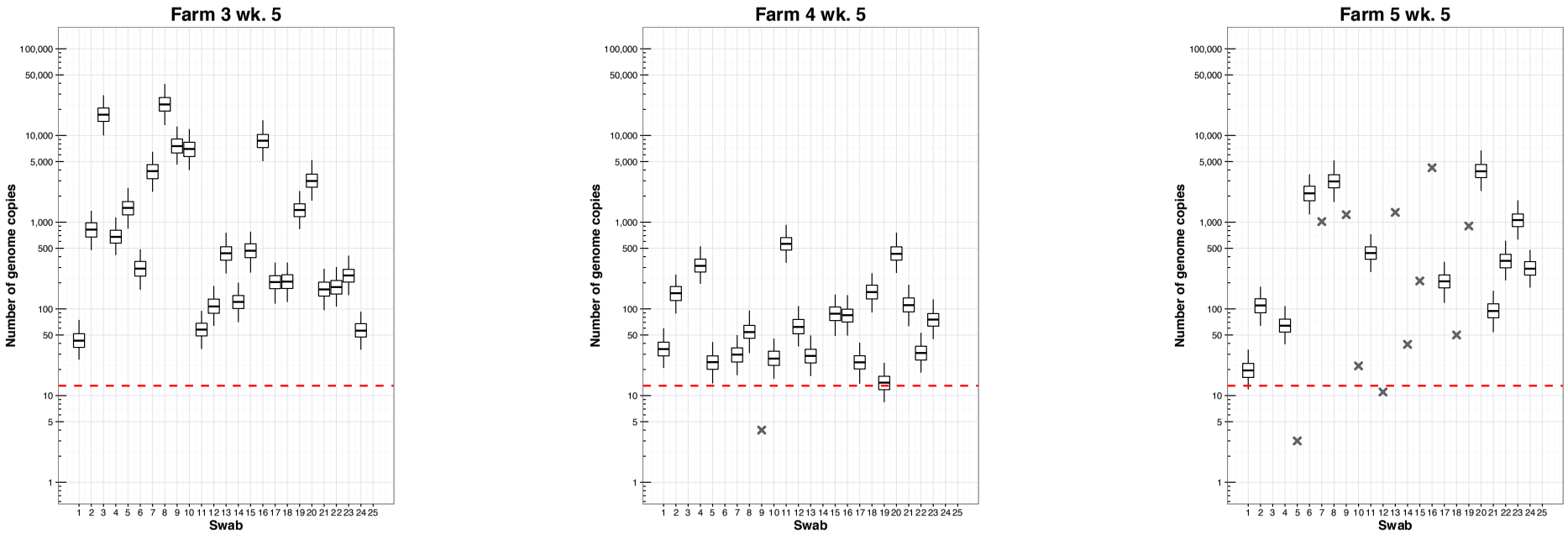

Figure 5.6. The number of *Streptococcus suis* genome copies on swabs of material scrapped from the palatine tonsils of five-week old pigs estimated from quantitative real-time PCR targeting the novel species-specific marker SSU0577.

Quantitative real-time PCR was performed using DNA extracted from material scraped from the palatine tonsils of pigs without culture and first isolating single colonies. The number of genome copies, per experimental sample, was estimated from C_t values using the standard curve equation for SSU0577. Boxplot, implemented in the R package: *ggplot2*, was used to visualise the number of genome copies and 95% confidence intervals (generated using a Bayesian approach). Boxes represent the median, 25th and 75th percentiles. Whiskers 95% confidence intervals (2.5th and 97.5th percentiles respectively). The horizontal line at 13 genome copies represents the qPCR limit of detection for SSU0577. Data presented as an '*x*' was done so to indicate that estimates of the number of genome copies are not reported with 95% confidence.

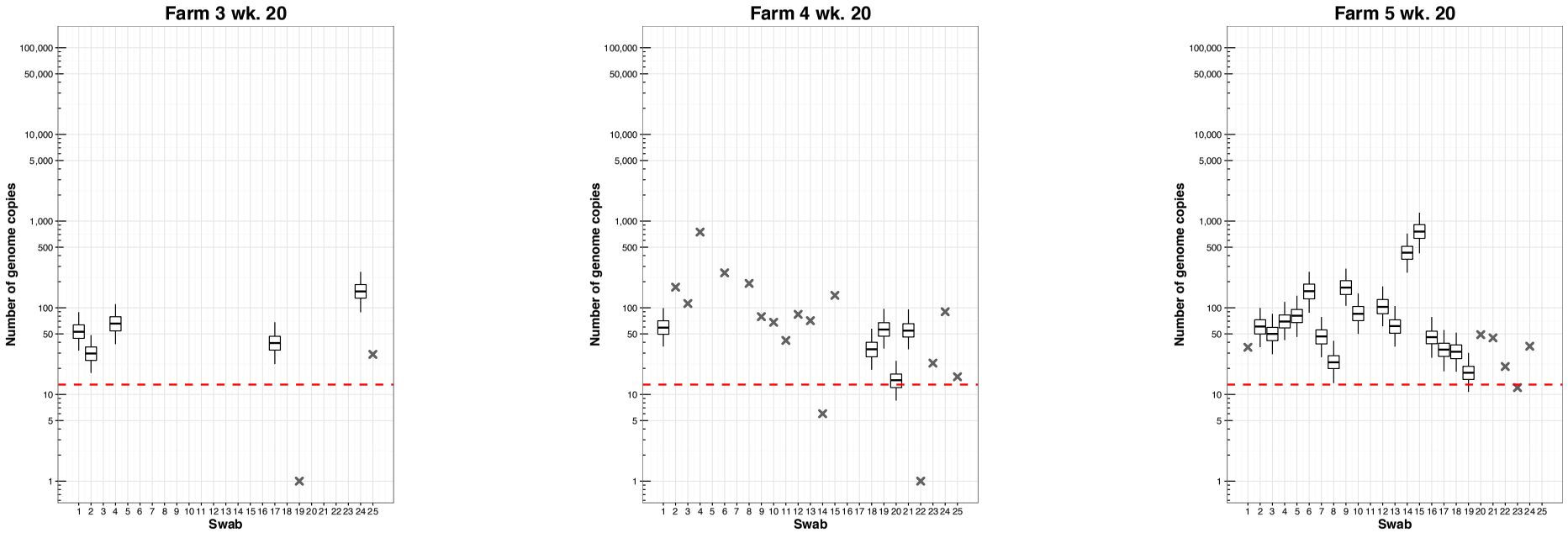

Figure 5.7. The number of *Streptococcus suis* genome copies on swabs of material scrapped from the palatine tonsils of 20-week old pigs estimated from quantitative real-time PCR targeting the novel species-specific marker SSU0577.

Quantitative real-time PCR was performed using DNA extracted from material scraped from the palatine tonsils of pigs without culture and first isolating single colonies. The number of genome copies, per experimental sample, was estimated from C_t values using the standard curve equation for SSU0577. Boxplot, implemented in the R package: *ggplot2*, was used to visualise the number of genome copies and 95% confidence intervals (generated using a Bayesian approach). Boxes represent the median, 25th and 75th percentiles. Whiskers 95% confidence intervals (2.5th and 97.5th percentiles respectively). The horizontal line at 13 genome copies represents the qPCR limit of detection for SSU0577. Data presented as an '*x*' was done so to indicate that estimates of the number of genome copies are not reported with 95% confidence.

5.2.2.1 Genome copy number as an indication of bacterial load on tonsil swabs is higher in pigs at five weeks of age in comparison to 20 weeks of age

Estimates of the number of *S. suis* genome copies on swabs of material scraped from the palatine tonsils of pigs was compared at two time points, five weeks and 20 weeks of age. The Shapiro-Wilk test for normality alongside Quantile-Quantile (qq)-plots was used to determine if the distribution of genome copies per experimental sample for a given farm group was well modelled by a normal distribution, and if so Student's t-test was used to determine if the mean number of genome copies at five weeks was significantly different in comparison to 20 weeks of age (p-value <0.05). Where sample size was too small to be modelled by a normal distribution, non-parametric hypothesis testing was performed. Table 5.9 summarises the test statistics and corresponding p-values used to infer that bacterial load at five weeks was consistently higher than in comparison to 20 weeks of age for all three farms sampled.

Table 5.9. Test statistics and corresponding *p*-values for statistical tests used to determine if the number of *Streptococcus suis* genome copies on swabs of material scraped from the palatine tonsils of pigs was significantly different at five weeks in comparison to 20 weeks of age.

The Shapiro-Wilk Normality test, Student's t-test and Wilcoxon Rank Sum test were all performed using log10-transformed data. Statistical tests were used to determine if the number of *S. suis* genome copies on swabs of material scraped from the palatine tonsils of pigs on farms in England and Wales was significantly different at two time points, five weeks and 20 weeks of age. The Shapiro-Wilk Normality test assumes normality unless p-values are <0.05. The Student's t-test assumes the true difference in means is equal to 0, unless p-values are <0.05. The Wilcoxon-Rank Sum test assumes the true location shift is equal to 0, unless p-values are <0.05.

Farm	Age	Shapiro-Wilk Normality test (W; ***p*-value**)	Mean	Variance	Student's t-test (t; ***p*-value**)	Wilcoxon Rank Sum test (W; ***p*-value**)
3	5	0.94145; **0.176**	2.829	0.680	5.1635; **0.00004**	-
3	20	0.94562; **0.706**	1.754	0.075		
4	5	0.93010; **0.274**	1.965	0.177	-	111; **0.00355**
4	20	0.72131; **0.020**	-	-		
5	5	0.93530; **0.467**	2.655	0.388	3.3252; **0.00430**	-
5	20	0.88414; **0.055**	1.944	0.157		

5.2.2.2 Ambient temperature has no detectable effect on the copy number as an indication of bacterial load of *Streptococcus suis* on tonsil swabs

To investigate if ambient air temperature had an effect on the number of *S. suis* genome copies on swabs of material scraped from the palatine tonsils of pigs, an analysis of variance (ANOVA) was used. High ambient temperature has been proposed as a risk factor responsible for high *S. suis* bacterial load in pork derived food items [265, 280]. Tonsil swabs used in this pilot study were originally collected over a 12-month period as part of a separate longitudinal study designed to enable any effect of air temperature on the positive isolate rate of *S. suis* to be evaluated. No significant effect (p-value = 0.211) of ambient air temperature on the number of genome copies was found when using an ANOVA to partition the observed variance into different sources of variation.

5.2.3 Invasive disease-associated strains of *Streptococcus suis* detected on tonsil swabs at five weeks were unable to be detected at 20 weeks of age

To investigate if qPCR could be used to screen swabs of material scraped from the palatine tonsils of pigs for invasive disease-associated strains of *S. suis*, primers were designed to target the newly defined pathotyping marker SSU1589. Figure 5.8 (five-weeks) and Figure 5.9 (20-weeks) show the number of invasive disease-associated genome copies, per experimental sample, estimated using qPCR targeting the newly defined *S. suis* invasive disease-associated pathotyping marker SSU1589. On all three farms sampled invasive disease-associated strains of *S. suis* could be detected by our methods. However, comparison of Figure 5.8 with Figure 5.9 revealed invasive disease-associated genome copies were detected in fewer experimental samples at 20-weeks of age despite numerous swabs testing positive (with 95% confidence) at five weeks of age.

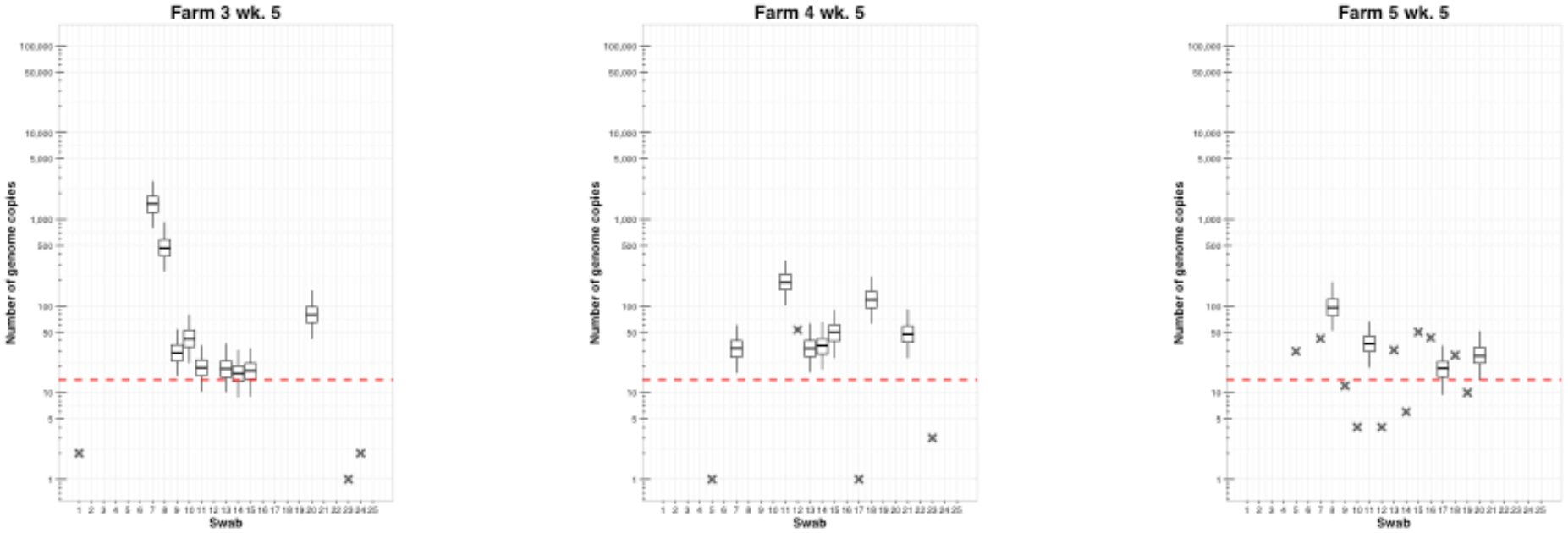

Figure 5.8. The number of invasive disease-associated *Streptococcus suis* genome copies on swabs of material scrapped from the palatine tonsils of five-week old pigs estimated from quantitative real-time PCR targeting the newly defined invasive disease-associated pathotyping marker SSU1589.

Quantitative real-time PCR was performed using DNA extracted from material scraped from the palatine tonsils of pigs without culture and first isolating single colonies. The number of genome copies, per experimental sample, was estimated from C_t values using the standard curve equation for SSU1589. Boxplot, implemented in the R package: *ggplot2*, was used to visualise the number of genome copies and 95% confidence intervals (generated using a Bayesian approach). Boxes represent the median, 25th and 75th percentiles. Whiskers 95% confidence intervals (2.5th and 97.5th percentiles respectively). The horizontal line at 14 genome copies represents the qPCR limit of detection for SSU1589. Data presented as an '*x*' was done so to indicate that estimates of the number of genome copies are not reported with 95% confidence.

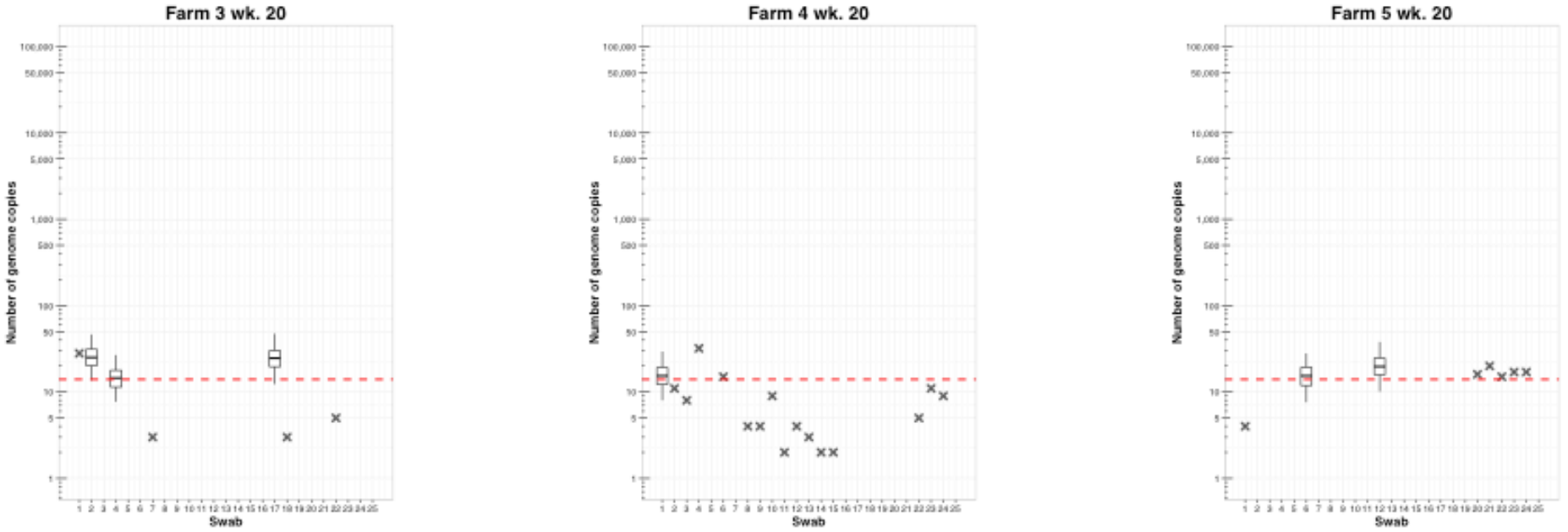

Figure 5.9. The number of invasive disease-associated *Streptococcus suis* genome copies on swabs of material scrapped from the palatine tonsils of 20-week old pigs estimated from quantitative real-time PCR targeting the newly defined invasive disease-associated pathotyping marker SSU1589.

Quantitative real-time PCR was performed using DNA extracted from material scraped from the palatine tonsils of pigs without culture and first isolating single colonies. The number of genome copies, per experimental sample, was estimated from C_t values using the standard curve equation for SSU1589. Boxplot, implemented in the R package: *ggplot2*, was used to visualise the number of genome copies and 95% confidence intervals (generated using a Bayesian approach). Boxes represent the median, 25th and 75th percentiles. Whiskers 95% confidence intervals (2.5th and 97.5th percentiles respectively). The horizontal line at 14 genome copies represents the qPCR limit of detection for SSU1589. Data presented as an '*x*' was done so to indicate that estimates of the number of genome copies are not reported with 95% confidence.

Having estimated the number of invasive disease-associated genome copies (Figures 5.8 and 5.9) it seemed logical to also estimate the number of non-disease associated genome copies, then compare the two and investigate the ratio of non-disease associated: disease-associated strains within the same experimental sample. Despite widespread colonisation few disease-associated isolates of *S. suis* are recovered from the upper respiratory tract [141], and in pig herds carrying multiple strains of *S. suis* sampling error (taking one or two samples, or sampling from only one or two pigs) may result in failure to identify a strain associated with a recent outbreak; a feature especially problematic in endemic herds [99, 121]. Figure 5.10 and Figure 5.11 show the number of non-disease associated genome copies, per experimental sample, estimated using qPCR targeting the newly defined *S. suis* non-disease associated pathotyping marker SSUST30534. On all three farms at both time points sampled non-disease associated strains of *S. suis* could be detected by our methods, in good agreement with Figures 5.6 and 5.7. The number of non-disease associated genome copies is estimated to be more consistent and higher in comparison to the number of invasive disease-associated genome copies.

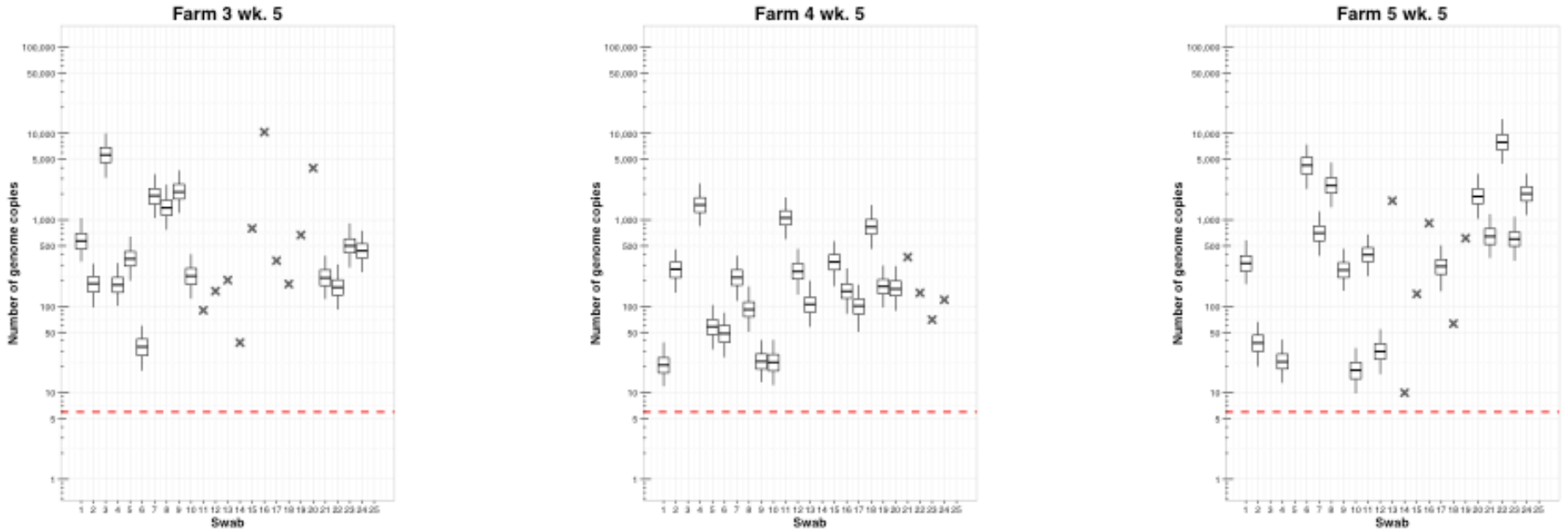

Figure 5.10. The number of non-disease associated *Streptococcus suis* genome copies on swabs of material scrapped from the palatine tonsils of five-week old pigs estimated from quantitative real-time PCR targeting the newly defined invasive disease-associated pathotyping marker SSUST30534.

Quantitative real-time PCR was performed using DNA extracted from material scraped from the palatine tonsils of pigs without culture and first isolating single colonies. The number of genome copies, per experimental sample, was estimated from C_t values using the standard curve equation for SSUST30534. Boxplot, implemented in the R package: *ggplot2*, was used to visualise the number of genome copies and 95% confidence intervals (generated using a Bayesian approach). Boxes represent the median, 25th and 75th percentiles. Whiskers 95% confidence intervals (2.5th and 97.5th percentiles respectively). The horizontal line at 6 genome copies represents the qPCR limit of detection for SSUST30534. Data presented as an '*x*' was done so to indicate that estimates of the number of genome copies are not reported with 95% confidence.

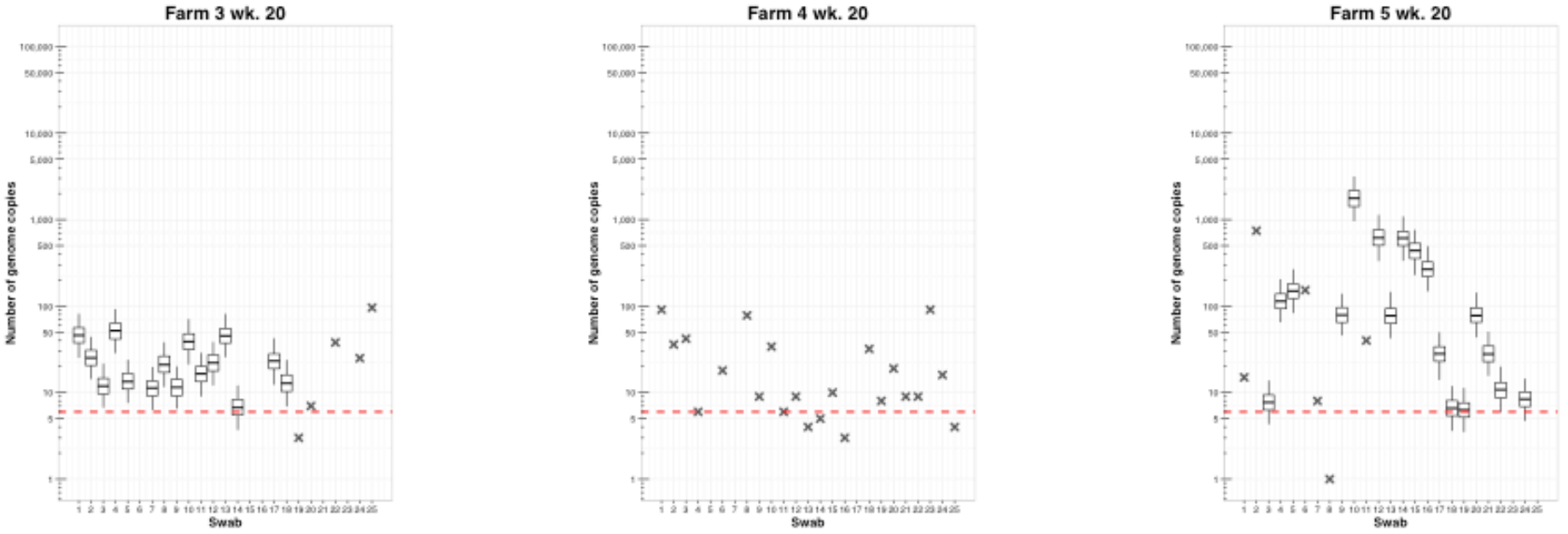

Figure 5.11. The number of non-disease associated *Streptococcus suis* genome copies on swabs of material scrapped from the palatine tonsils of 20-week old pigs estimated from quantitative real-time PCR targeting the newly defined invasive disease-associated pathotyping marker SSUST30534.

Quantitative real-time PCR was performed using DNA extracted from material scraped from the palatine tonsils of pigs without culture and first isolating single colonies. The number of genome copies, per experimental sample, was estimated from C_t values using the standard curve equation for SSUST30534. Boxplot, implemented in the R package: *ggplot2*, was used to visualise the number of genome copies and 95% confidence intervals (generated using a Bayesian approach). Boxes represent the median, 25th and 75th percentiles. Whiskers 95% confidence intervals (2.5th and 97.5th percentiles respectively). The horizontal line at 6 genome copies represents the qPCR limit of detection for SSUST30534. Data presented as an '*x*' was done so to indicate that estimates of the number of genome copies are not reported with 95% confidence.

To test the hypothesis that, despite widespread colonisation, the reason few disease-associated isolates are recovered from the upper respiratory tract of pigs is because the ratio of non-disease-associated: disease associated strains is greater than 3:1 (respectively), the number of genome copies estimated using primers targeting the newly defined pathotyping markers SSU1589 (invasive disease-associated) and ST30534 (non-disease) were compared. Where, although chosen for practical reasons, it was hypothesised that three single colony picks from solid agar plates is not enough to generate accurate reports of the true prevalence of *S. suis* strains on the tonsils of pigs. Table 5.10 shows the estimated number of genome copies and ratio (non-disease-associated: disease associated; rounded to one significant figure) where both invasive disease-associated and non-disease associated markers were detected in the sample experimental sample. In general, the number of non-disease associated: disease-associated genome copies is greater than 3:1, indicating that three single colony picks from solid agar plates is not be enough to detect low-level prevalence of invasive disease-associated strains of *S. suis*.

Table 5.10. Ratio of non-disease associated: disease associated strains of *Streptococcus suis* detected on swabs of material scrapped from the palatine tonsils of pigs.

Quantitative real-time PCR was performed using DNA extracted from material scraped from the palatine tonsils of pigs without culture and first isolating single colonies. The number of genome copies, per experimental sample, was estimated from C_t values using the standard curve equation calculated for each primer pair. This table includes only experimental samples for which both disease-associated and non-disease associated estimates of the number of genome copies could be reported with 95% confidence.

		----- Estimates generated from quantitative real-time PCR data -----			
Farm	**Swab**	***Streptococcus suis* (genome copies)**	**Disease (genome copies)**	**Non-disease (genome copies)**	**Ratio (genome copies; non-disease: disease)**
F3_w05	7	3862	1505	1901	1:1
F3_w05	8	2350	477	1416	3:1
F3_w05	9	7636	28	2116	80:1
F3_w05	10	7036	42	223	5:1
F3_w05	20	3001	78	-	-
F4_w05	7	29	7	32	5:1
F4_w05	11	557	186	1054	6:1
F4_w05	13	28	32	104	3:1
F4_w05	14	-	34	-	-
F4_w05	15	87	49	326	7:1
F4_w05	18	157	119	845	7:1
F4_w05	21	111	47	-	-
F5_w05	8	3022	97	2570	30:1
F5_w05	11	436	36	394	10:1

5.2.4 Quantitative real-time PCR is more sensitive in comparison to classic culture-based detection methods for *Streptococcus suis*

To compare the positive detection of *S. suis* using quantitative real-time PCR targteing the novel species specfic marker (SSU0577) to detection using the standard culture-based method, all swabs were subjected to both detection methods. Classic culture-based isolation of *S. suis* from the upper respiratory tract is a laborious and time-consuming process associated with low sensitivity, particularly due to phenotypically similar streptococcal species difficult to distinguish on the basis of colony morphology alone [27]. In contrast, qPCR provides fast, precise and accurate results, and has become an established method for pathogen detection [571]. Table 5.11 shows a 2x2 contingency table summarising the outcome of both detection methods for 114 swabs, in the interest of fairness a total of 36 swabs have been removed from statistical analyses at this point because experimental samples did not pass qPCR quality control. Table 5.11 shows that detection techniques agreed for 44% of the swabs ((25+25)/114). Of the other outcomes it was not possible using qPCR to detect *S. suis* on 11 (or 10%) swabs that were confirmed to be positive for *S. suis* using culture-based methods (then confirmed with whole-genome sequencing). In comparison, a total of 53 swabs (or 46%) were found to be positive for *S. suis* using the qPCR-based approach but none of the three colony picks from solid agar plates were confirmed to be *S. suis*. Indicating qPCR might be more effective than standard culture based detection techniques for *S. suis*.

Table 5.11. Contingency table of the number of tonsil swabs positive for *Streptococcus suis* based on quantitative real-time PCR in comparison to classic culture-based detection techniques.

A 2x2 contingency table summarising the number of swabs (n=114) of material scraped from the palatine tonsils of pigs from which *S. suis* was detected with 95% confidence.

		API20 Strep, confirmed by WGS	
		+	-
Quantitative real-time PCR	+	25	53
	-	11	25

To identify if the difference between qPCR and culture-based methods detecting *S. suis* on tonsil swabs was statistically significant, the proportion of swabs testing positive using each technique was calculated. Proportions were calculated by dividing the number of swabs confirmed to be positive for *S. suis* by the number of swabs collected (n=114, 25 swabs per group sampled minus the number of swabs failing qPCR quality control). Comparison of these proportions produced a *Z*-value of 2.137 (*p*-value = 0.033; see section 2.2.1.9), indicating qPCR is more effective to detect *S. suis* in comparison to detection using microbiology techniques to isolate single colonies and then biochemical profiles to group isolates to the bacterial species.

Table 5.12 summarises the number of single colonies picked from solid agar plates that had biochemical profiles (API20 Strep) associated with *S. suis*, and then of these how many were confirmed to be *S. suis* using whole-genome sequencing data. The information in Table 5.12 highlights the difficulty, even for experienced technicians at the Scottish Agricultural College Consulting Service, of identifying *S. suis* based on colony morphology and α/β haemolysis on solid agar plates alone. For example, for five of the six groups sampled from 75 single colony picks per group (three picks, from 25 swabs) suspected of being *S. suis* single figure numbers were confirmed to be *S. suis* using whole-genome sequencing. Figure 5.12 is a visual representation of the data in Table 5.12, showing the proportion of swabs, per group sampled, testing positive using qPCR (black) in comparison to swabs testing positive for *S. suis* using culture-based methods (grey-scale). The five groups sampled where statistically significant (*p*-value <0.05) differences exist between the proportions of swabs testing positive for *S. suis* using qPCR or culutr-based methods are indicated with an asterisk.

Table 5.12. Proportions of swabs testing positive for *Streptococcus suis* using culture-based techniques in comparison to testing positive using quantitative real-time PCR.

(**API20 Strep**) The number of single colony picks deemed to be *S. suis* based on API20 Strep profiles. Three single colony picks from solid agar plates were taken per swab (n=75) and API 20 Strep profiles deduced. (**API20 Strep, confirmed by WGS**) The number of single colony picks confirmed to be *S. suis* based on Illumina whole-genome sequencing data. Proportions were then calculated by dividing the number of swabs confirmed to be positive for *S. suis* by the number of swabs collected per group sampled. ***Z*-values** were calculated to test for difference between two proportions i) swabs testing positive for *S. suis* using qPCR in comparison to ii) testing positive using culture-based methods then confirmed to be *S. suis* using whole-genome sequencing data. The *Z*-table was used to find probabilities for a statistical sample with a standard (*Z*-) normal distribution, and *p*-values <0.05 considered significant.

Farm	API20 Strep	API20 Strep, confirmed by WGS	Proportion positive for *S. suis*	qPCR proportion positive for *S. suis*	*Z*-value	*p*-value
F3_w05	52	16	0.64	0.96	-2.82843	0.002342
F3_w20	13	4	0.16	0.22	-0.36811	0.356437
F4_w05	11	7	0.28	0.75	-3.39683	0.000342
F4_w20	8	4	0.16	0.50	-3.12889	0.000880
F5_w05	9	1	0.04	0.79	-3.54654	0.000196
F5_w20	18	5	0.20	0.84	-3.41882	0.000315

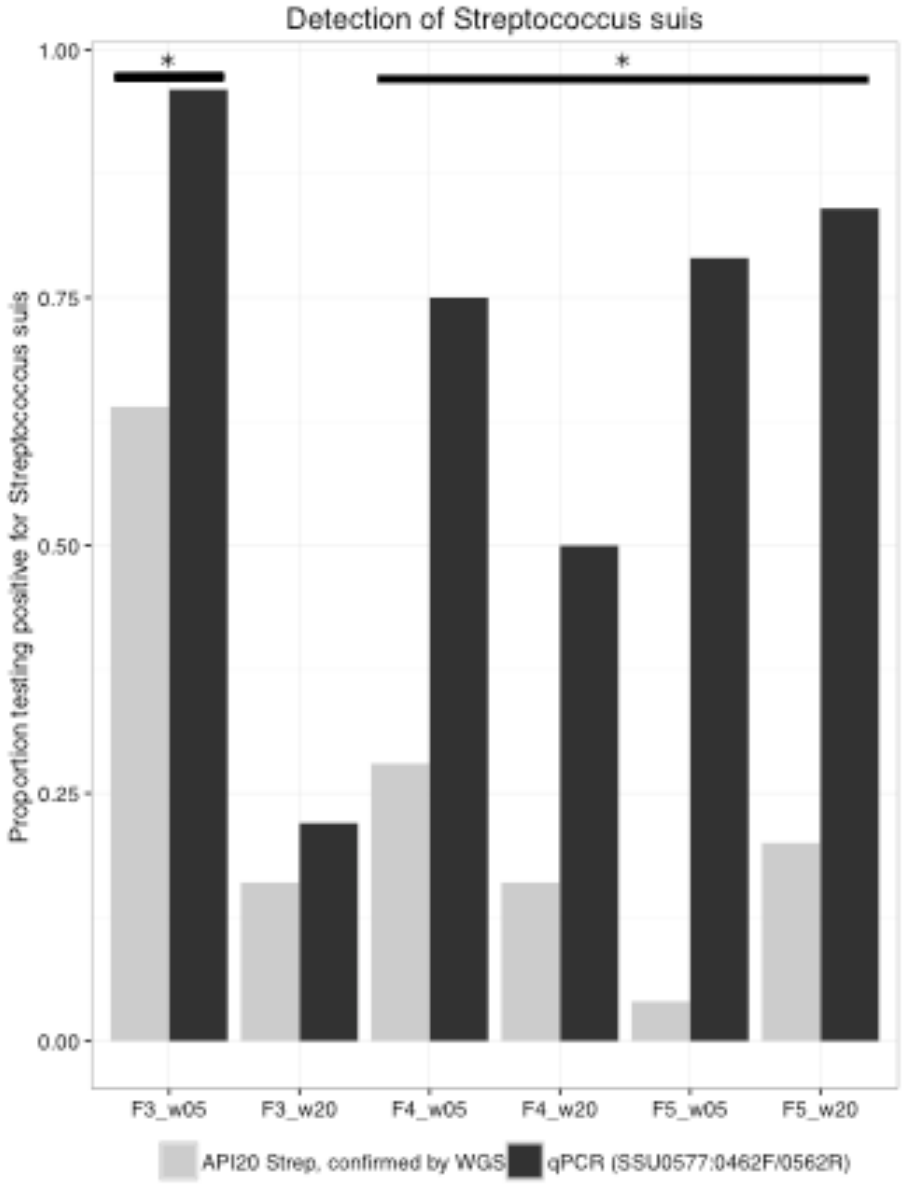

Figure 5.12. Proportions of tonsil swabs testing positive for *Streptococcus suis* using quantitative real-time PCR in comparison to classic culture-based techniques.

Histogram of proportions of swabs of material scrapped from the palatine tonsils of pigs testing positive for *S. suis* using qPCR (black) in comparison to testing positive using culture-based techniques (grey-scale) for each group sampled. Proportions were calculated by dividing the number of swabs confirmed to be positive for *S. suis* by the number of swabs collected per group sampled. *Z*-values were calculated to test for difference between two proportions i) swabs testing positive for *S. suis* using qPCR in comparison to ii) testing positive using culture-based methods then confirmed to be *S. suis* using whole-genome sequencing data. The *Z*-table was then used to find probabilities for a statistical sample with a standard (*Z*-) normal distribution, and *p*-values <0.05 considered significant (indicated with an asterisk),

5.3 Discussion

This chapter describes the use of qPCR to screen swabs of material scraped from the palatine tonsils of pigs for invasive disease-associated strains of *S. suis*. qPCR is a technique capable of fast, precise and accurate results that has become an established method for pathogen detection [571]. In chapter 4, attempts to use our *S. suis* pathotyping tool to screen swabs of material scraped from the palatine tonsils of pigs without culture and first isolating single colonies produced an interesting result - the production of both invasive disease-associated and non-disease associated mPCR amplicons from the same experimental sample, that in both the training collection (n=115) and the out-of-sample 'test' collection (n=138) have not yet been detected in the same isolate clone. Therefore, we inferred that at least two (heterogeneous) strains of *S. suis* were present in the pooled oral fluid collected from cotton chew ropes. The detection of the *S. suis* spp. in asymptomatic carrier herds is of little practical value in predicting likelihood of future clinical relevance [11]. Instead, the value of future molecular tools for surveillance and preventive health management lies in the detection of strains of *S. suis* that have genetically increased potential to cause disease and are colonising presently healthy animals [11]. Detection of genetic markers without the need to culture and perform the laborious and costly process of picking single colonies from solid agar plates also has the potential to improve the sensitivity of *S. suis* surveillance programs, a feature that would be of considerable benefit to better the understanding of subclinical infection dynamics to support future intervention programs.

Objective 1. Limited to a 72-well qPCR rotor, this study was restricted to targeting four molecular markers for practical reasons. Evaluation of analytical sensitivity, typically expressed as the limit of the detection, showed the primer pairs designed for this assay were able to detect very low copy numbers of *S. suis* genomes, numbers impossible to see on solid agar plates without culture at 37 ˚C for 18-24 hours. Screening herds for tonsil carriers of *S. suis* has been explored in the past, and the tonsil remains a popular sampling site due to it being considered the natural habitat of *S. suis* [142, 426, 569]. The upper respiratory tract is a diverse environment and accurately replicating the diversity of such a highly contaminated site *in vitro* is extremely

difficult, if not impossible. Evaluation of species-specificity was done using field isolates of closely related Streptococcaceae, Pasteurellaceae and Alcaligenaceae commonly recovered from the upper respiratory tract of pigs. No qPCR amplicons were produced indicating the qPCR assay described in this chapter reliably targets *S. suis*, and false-positives due to phenotypically similar and closely-related field isolates are low. However, it is important to acknowledge that none of the species-specificity experiments contained a pig (or human) genomic DNA control, which is important to be absolutely sure that no false positive bands have been produced.

It is important to mention the primer pairs targeting SSU1462 (*recN*) did produce qPCR amplicons using template DNA from strains of *Streptococcus gallolyticus* and *Streptococcus pneumoniae*. The core-metabolic gene *recN*, encoding a recombination/repair protein, was included to continue to evaluate SSU0577 as a robust species-specific marker for *S. suis* due to its previous use as a marker to differentiate members of the genus *Streptococcus* [572, 573]. Melt curve analyses indicated that these qPCR amplicons could be differentiated from those corresponding to *S. suis* based on the higher temperature required for dissociation of dsDNA, indicating a different length or GC content of these non-*S. suis* amplicons. Indeed, it was curious to find primer pairs targeting SSU1462 (*recN*) consistently produced higher estimates of the number of genome copies per experiential sample in comparison to SSU0577; a feature we infer was due to non-specific primer binding.

The performance of qPCR primers was further evaluated with a mixed sample template using genomic DNA from two closely related lab strains of *S. suis*, P1/7 (invasive disease-associated) and LSS034 (non-disease associated). Mixed templates were created to reflect the hypothesis that disease-associated strains of *S. suis* are prevalent at low levels in comparison to non-disease associated strains in the upper respiratory tract of pigs. Although by no means representative of the true diversity in the upper respiratory tract, a mixed sample template of closely related strains did not affect qPCR performance and served to highlight that very low copy numbers of

invasive disease associated markers in a mixed sample of closely related strains can be detected by our method.

In order to investigate if qPCR could actually be used to detect *S. suis* (the species, regardless of observed clinical phenotype) on tonsil swabs without the need to first culture and isolate single colonies/clones from solid agar plates, material scraped from the palatine tonsil was resuspended in PBS and a DNA extraction performed using the Sigma-Aldrich GenElute Bacterial DNA Kit. Despite the acknowledged reservations, it was possible to use qPCR targeting SSU0577 to detect *S. suis* on swabs of material scraped from the palatine tonsils of pigs without the need to first culture and isolate single colonies. The *S. suis* gene SSU0577, was evaluated in chapter 4 as a novel species-specific marker, after being identified as one of the most conserved protein-encoding sequences in the *S. suis* core-genome (chapter 3). Inclusion of a *S. suis* species-specific marker in the mPCR pathotyping tool described in chapter 4 was for two purposes: i) a positive control, to confirm PCRs were successful and ii) to infer disease-/non-disease associated PCR amplicons were due to the presence of *S. suis*. In qPCR, SSU0577 was used for the same purposes. However, qPCR-targeting SSU0577 also produced estimates of the number of *S. suis* genome copies per experimental sample regardless of association with a particular clinical phenotype. Of the three farms sampled, none had a recent history of *S. suis* disease or house pigs showing obvious signs of streptococcal disease. Detection of *S. suis* was possible in all six groups sampled (three farms, sampled at two time points five weeks and 20 weeks of age), supporting previous observations that consider *S. suis* endemic in pig herds often without causing signs of clinical disease [142, 565, 566]. However, although detected in all six groups sampled, *S. suis* was not detected in all experimental samples (i.e. tonsil swabs). Possible reasons for this could include failure of the sampling techniques to draw material from deep within the crypts of the palatine tonsils, where *S. suis* tends to form confined groups [574] or the effect of PCR inhibitors in salivary fluids such as DNA binding agents or compounds added as a result of sample processing which could potentially lead to false negatives [575]. Equally, it is feasible that 'true negatives' existed within the herds sampled. Regardless, *S. suis was* detected in all three herds sampled and

this finding highlights the importance of sampling a statistically robust sample population based on expected prevalence of carriers.

Objective 2. Most *S. suis* strains are thought to act as opportunistic or secondary pathogens, although some can act as primary pathogens causing invasive disease such as encephalitis and meningitis [121]. With this in mind this chapter reported the detection of not only the *S. suis* species but also detection of markers associated with invasive disease-causing isolates on swabs of material scraped from the palatine tonsils of pigs. The detection of the invasive-disease associated genetic markers is an interesting finding in itself, as all three farms sampled did not have a history of *S. suis* disease or at the point of sampling house pigs showing obvious signs of streptococcal disease. This finding supports opinion that *S. suis* strains with a heightened genetic capacity to cause invasive disease are present on farms without signs of streptococcal disease, and potentially that disease onset is dependent on animal stress arising from environment and farm management practices [85, 142].

A notable finding described in this chapter was that despite 20 (27%, of n=75) swabs testing positive (with 95% confidence) for the disease-associated marker of *S. suis* at five weeks of age, the same genetic marker could not be detected with 95% confidence in experimental samples collected at 20 weeks of age. This finding supports the previous observation that *S. suis* associated disease is less prevalent in older pigs approaching slaughter age, instead being most prevalent during the post-weaning period when pigs are commonly mixed and maternal derived passive antibody titres are in decline [126, 426]. It is interesting to speculate that this observation is linked to previous descriptions of disease-associated strains of *S. suis* having a smaller genome size in comparison to non-disease associated strains [141]. Clear genetic differences between disease-associated and non-disease associated strains have been described and although disease-associated strains have an overrepresentation of putative virulence factors this finding could indicate that having fewer metabolic genes could lead to disease-associated strains being less able to survive in the diverse and highly competitive environment of the upper

respiratory tract for long periods of time. Equally, it is also important to acknowledge that this observation could be due to the effect of immunity against recently encountered strains of *S. suis* at five weeks versus the background of a maturing and more efficient immune system as the pigs reach older ages.

Ambient temperature at the time of sampling was also investigated as a significant factor influencing the detection rate of *S. suis* on farms in England and Wales. Previous reports have proposed high ambient temperature as a potential risk factor responsible for high *S. suis* bacterial load in pork derived food items and/or human disease in Asian countries [242, 265, 280]. Such an effect could potentially lead to enhanced carriage and shedding from clinically healthy pigs during warmer months. However, such a relationship is likely to be complicated. Interestingly, the opposite relationship between temperature and transmission of *Streptococcus pneumonia* has recently been described – where temperature and prevalence of carriage are described as being repeatedly enhanced during cooler months [576]. In this study, no statistically significant (*p*-value <0.05) effect of ambient temperature on the number of genome copies estimated using qPCR was found. This could be due to a number of cultural or experimental reasons, although it is difficult to speculate further as there is no immune response in dead meat and such a comparison with the bacterial load found on a carcass local supermarkets or a retail wet market in Asia are inappropriate.

Objective 3. *S. suis* is considered endemic in pig herds [142, 565, 566], an observation also described in this chapter using qPCR. However, despite widespread colonisation, few disease-associated isolates of *S. suis* are recovered from the upper respiratory tract [141]. In pig herds infected with multiple strains of *S. suis*, sampling bias might result in failure to identify a strain associated with a recent outbreak; a feature especially problematic in endemic herds [99, 121]. Currently, the *S. suis* diagnosis procedure relies on culture and isolation of single colonies. This is a major limiting factor of the detection/diagnosis procedure, and removal of the bottleneck created by picking single colonies from solid agar plates has the potential to improve the sensitivity of surveillance programs and enable more detailed investigation of infection dynamics. Comparison of any new diagnostic tool should be to an agreed 'gold standard', and in this study was drawn against classic culture-based techniques to date still routinely used in diagnostic laboratories. Comparing qPCR to classic culture-based techniques showed 46% of swabs tested to be positive for *S. suis* using qPCR were deemed negative for *S. suis* based on culture. While this is an interesting observation, it is important to acknowledge that this comparison was comparing two different things, live recoverable bacteria verses a DNA fragment.

An important observation described in this chapter is the number of non-disease associated: disease-associated genome copies detected is greater than 3:1, indicating that three single colony picks from solid agar plates is insufficient to be confident of detecting strains of *S. suis* with the potential to cause invasive disease. For practical and budgetary reasons three single colonies suspected of being *S. suis* were picked from solid agar plates. It should be acknowledged that some diagnostic laboratories now use five single colony picks, this is important and when this data is reviewed for publication in a peer reviewed scientific journal any additional analyses (where possible) should be adjusted to reflect this. Regardless, increasing this number of single colony picks to exceed five would be impractical for routine surveillance. Equally using whole-genome sequencing data to confirm isolates recovered using culture-based techniques are *S. suis* is excessive (but in this case available due to data being generated as part of a separate research project) and although potentially imminent in human diagnostic laboratories at present is

impractical. Until whole-genome sequencing becomes even more accessible the potential of qPCR to accurately detect minute amounts of target nucleic acids with great precision and without the need to culture and first isolate single colonies would be a useful surveillance tool for the swine industry.

Finally, this study describes the use of qPCR using an intercalating dye to detect dsDNA PCR amplicons due to its low cost and simplicity. To increase specificity development of a probe-based approach should be considered, although such an approach might not be possible for such a highly recombinagenic species. Probe-based qPCR is designed to be more specific than using an intercalating dye. A number of approaches are available, including: hybridisation probes, hydrolysis probes or molecular beacon probes. Hydrolysis probes, such as TaqMan or PerfectProbe, are among the most popular. These are fluorescently labelled oligonucleotides designed to bind downstream of one of the qPCR primers. The TaqMan assay, is based on the 5'-3' exonuclease activity associated with *Taq* DNA polymerase and the phenomenon of fluorescence resonance energy transfer (FRET). Each TaqMan probe contains a covalently attached 5' fluorescent reporter (FAM; 6-carboxy-fluorescein is a green reporter; others include VIC, JOE and CY5 which emit light at different wavelengths); and a 3' quencher molecule (TAMRA, 6-carboxy-tetramethyl-rhodamine). Before the qPCR begins, the reporter and quencher are in close proximity to each other, allowing FRET to occur; meaning the fluorescent signal of the reporter is quenched. During the qPCR, the primers and probe anneal to the target; and DNA polymerase extends the primer upstream, of the probe. If the probe is bound to the correct target sequence 5'-3; exonuclease activity of the *Taq* DNA polymerase cleaves the probe, releasing a fragment containing the reporter. Cleavage of the probe separates the report and quencher dyes; resulting in fluorescence. The use of a probe-based approach is attractive and should be carefully considered if the use of qPCR is a developed into a widely used diagnostic tool for *S. suis*, alone or in combination with other currently available techniques.

In conclusion, this chapter presents evidence for the development of a molecular-based diagnostic tool to screen material scraped from the palatine tonsils of pigs for invasive disease-associated strains of *S. suis* without the need for culture and isolation of single colonies from solid agar plates. Due to time restrictions is was not possible to repeat all experimental samples that failed qPCR quality control, and to avoid reducing sample size and the statistical power of this study should be the first task of future work for a complete proof of concept. An interesting experiment to add to evaluation of proof of concept would be plating aliquots of material from the palatine tonsils of pigs resuspend in PBS from which DNA extractions were performed to see if ratios of invasive disease-associated:non-disease associated strains can be replicated at the bench quickly using colony PCR. Finally, a feature of *S. suis* research is the lack of a comprehensive understanding of the role virulence-factors have in pathogenesis. At present the role(s) of the pathotyping markers identified and evaluated in this thesis are also unclear and efforts should be put into identifying their roles, such as using site-directed mutagenesis studies (as described in chapter 6).

Chapter 6

Design, construction and evaluation of isogenic *virA* operon knockout mutants in *Streptococcus suis*

6.1 Introduction

Recent progress in *S. suis* 'omics' research has generated a large amount of data on putative virulence-associated factors (summarised in Table 1.3), however experimental evidence for possible function and the biological role in infection is available for only a few of these factors [129, 131, 302]. This chapter attempts to address this, describing the design and evaluation of two isogenic mutants designed to begin to better understand the role in pathogenesis of the newly defined pathotyping marker SSU1589. Described in chapter 3, SSU1589 was identified as a genetic marker associated with the observed clinical phenotype of invasive *S. suis* disease in pigs on farms in the UK. Described in chapters 4 and 5, SSU1589 was then evaluated as a novel marker of invasive disease-association, indicating that *S. suis* strains capable of causing invasive disease co-exist with non-disease associated asymptomatic carriage strains on the palatine tonsils of colonised pigs. Leading to the hypothesis that culture-based approaches to surveillance may be underestimating the true prevalence of strains with increased capacity to causing invasive disease.

The protein-encoding sequence SSU1589, also known as *virA*, is included as part of the Fittipaldi *et al.* review of putative virulence factors (summarised in Table 1.3) involved in the pathogenesis of the infection caused by *S. suis* [129]. To our knowledge *virA* was first associated with virulence by Wilson *et al.* who used a novel signature-tagged mutagenesis (STM) system for *S. suis* on serotype 2 strain Henrichsen S735 [319]. One of 22 unique STM mutants, the "*lin*0523" protein was predicted to have the putative function "similar to specificity determinant HsdS (*Listeria innocua* Clip11262)", and (the STM mutant was) shown to be attenuated in experimentally infected CF-1 mice and in both intravenously and intraperitoneally infected CDCD pigs [319]. Since 2007, Li *et al.* has been the only group to publish research aimed at understanding the role of SSU1589 (hereafter referred to as *virA*, denoting virulence gene A) in pathogenesis and virulence of *S. suis*; reporting that the intact *virA* protein-encoding sequence (1206 bp) existed only in virulent isolates tested, and that it played an important role in pathogenesis of *S. suis* serotype 2 strain ZY458 in a rabbit experimental infection model [378].

However, Li *et al*. were unable to propose a function or role in pathogenesis for *virA*, although did question the proposed role of *virA* as a Type I restriction modification system hsdS protein as its annotation as a "type I restriction-modification system S protein" was based solely on the presence of a conserved sequence in this gene with homology to the sequence of the restriction endonuclease S subunit [378].

6.1.1 Type I restriction modification systems in *Streptococcus suis*

In general, restriction endonucleases do not correlate with virulence in bacteria [378]. Since it was learnt that modification of a cell's DNA by methylation protected DNA whereas the absence of modification on bacteriophage DNA rendered it susceptible to restriction by endonucleases, Type I restriction modification (RM) enzymes have been associated with defence of the host bacterium against foreign DNA [559]. However, advances in bioinformatics, whole-genome sequencing and very recently methylome analysis by next generation single-molecule real-time (SMRT) sequencing has allowed our understanding of the range of functions that RM systems have to expand to include stabilising mobile genetic elements and gene regulation potentially providing evolutionary fitness advantages and virulence under certain conditions [560]. For example, this might include fixing an antibiotic resistance gene within the bacterial chromosome, where the gene was originally introduced to the bacterial cell by a bacteriophage or on plasmid. As a result, the proposed role in protection against foreign DNA such as that borne by bacteriophages and plasmids may merely be a coincidental benefit of these functions [561].

Type I RM systems are large pentameric (2M+2R+S) protein complexes that require ATP, Mg^{2+} and S-adenosylmethionine (SAM) for activity, and display both restriction endonuclease and modification methyltransferase activities. Trimers of 2M+S can act solely as a methyltransferase [577]. Type I RM systems are encoded by three host specificity determinant genes (*hsdM*, modification; *hsdR*, restriction; and *hsdS*, specificity). The S subunit of Type I RM enzymes is responsible for DNA sequence recognition. Characteristically, the specificity protein of Type I RM enzymes has a duplicated organisation of two target recognition domains (TRDs) in tandem, with

each TRD specifying one half of a bi-partite recognition sequence separated by a gap (AAC-N6-GTGC; where N = any base). TRDs occur as direct repeats in the linear amino acid sequence, but assume an inverted orientation in the folded protein as α-helices associate in antiparallel to make a coiled-coil 'leucine-zipper' (Figure 6.1). Recombination between TRDs generates new sequence specificities and is a powerful driver of Type I RM system diversification [553, 578]. Recently, this switching of S subunits in the Type I RM system SpnD39III has been demonstrated to be linked to a "switch from a carriage state to a virulent state" in *Streptococcus pneumonia* [553].

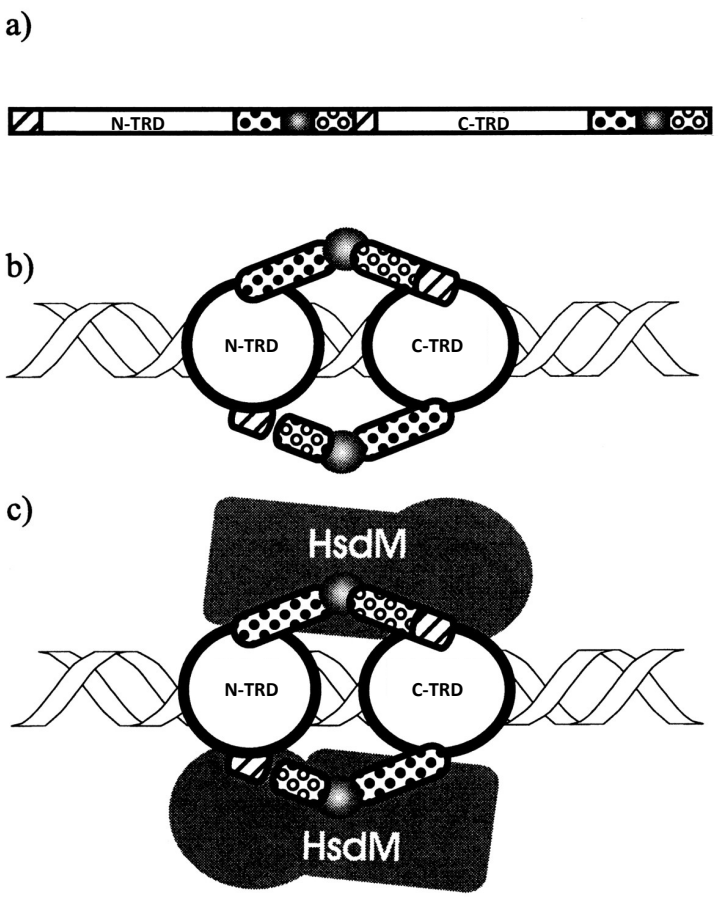

Figure 6.1. Domain organisation of the Type I restriction modification S subunit.

Schematic representation of the Type I RM HsdS subunit, adapted and updated from the review by Murray [579]. (**a**) Organisation indicating the two variable target recognition domains (TRDs), and regions conserved in sequence for all members of the Type I RM family. (**b**) Adaptation of the model by Kneale [580] indicating the repeated sequences form bridges between the variable TRDs. (**c**) Representation of a methyltransferase trimer (2M+S), in which the two HsdM subunits bind to the bridge region between TRDs forming an enzyme with pseudodyad symmetry.

The distribution of Type I RM systems in *S. suis* has recently been described, of which there are three in isolate P1/7 [552]. Since the efforts of Sekizaki *et al.*, focusing on the Type II RM system SsuDAT1I, in the early 2000s [581-583] the distribution and RM systems in *S. suis* (until recently) and the epigenetic landscape of *S. suis* has not been investigated. Interestingly, in the recent Willemse *et al*. description of the distribution of Type I RM systems in *S. suis* the authors propose that "*S. suis* uses a 'core' type I R–M system (SsuPORF1588P) to protect itself from foreign DNA and that additional type I R–M systems might allow for further phage defense and possibly for regulatory control" [552]. However, we found that "SsuPORF1588P" is unable to be detected in all of the *S. suis* isolates that we have sequenced (recovered from pigs on farms in the UK, as opposed to the Netherlands) and therefore should not be simply defined as a core-metabolic (housekeeping) system; although this does highlight that effort to understand the role of epigenetic modification in *S. suis* is currently being undertaken. Indeed, as this thesis is written a *hsdS* gene (locus_tag ZY05719_RS06855) belonging to Type I RM system "SsuZY05719II" in the Chinese virulent serotype 2 strain ZY05719 was reported to facilitate anti-phagocytosis and survival in microglia and whole blood through positively impacting the expression of virulence-associated factors [584]. It is worth noting that SsuZY05719II corresponds to operon SSU1272-1274 in P1/7 and is neither the Type I RM system the work described in this thesis focuses on (SSU1588-1590), nor the original Type I RM system unconventionally named "SsuCC20P" (SSU0651-0653) of particular interest to Willemse *et al.* [552].

6.1.2 Natural competence in *Streptococcus suis*

A pheromone-induced mechanism of natural genetic transformation (or competence) has been proposed as an approach for rapid targeted gene manipulation of *S. suis* [490]. Natural competence is a form of horizontal (lateral) gene transfer used by bacteria to uptake, process and recombine with short fragments of 'naked' exogenous DNA from the environment. Manipulation, *in vitro*, of the competence system in *S. suis* has been shown to have the potential to replace the use of low-efficiency genetic approaches that use *Escherichia coli* shuttle vectors and suicide vectors [490]. This opens up the possibility of routine genetic manipulation and gene deletion using linear DNA fragments assembled with common PCR-based techniques.

Homologs of the genes associated with natural competence are widespread among both Gram-negative and Gram-positive bacteria, although experimental evidence for natural competence is limited to ~80 bacterial species; that includes 16 species of the genus *Streptococcus* [585]. All known naturally competent bacteria, with the exception of *Helicobacter pylori*, are thought to share a common mechanism of DNA uptake, with processing and recombination controlled by a single master regulator. In the streptococci the single master regulator is thought to be the alternative sigma factor SigX (also known as ComX). The pheromone regulatory system for activation of competence used to generate the data described in this thesis has, in addition to *S. suis*, also previously been described in *Streptococcus mutans*, *Streptococcus pyogenes* and *Streptococcus thermophiles* [490]. Where a Rgg-family transcriptional activator (ComR) positively regulates expression of ComX through allosteric interaction with a extracellular pheromone encoded by ComS (SSU0050) [490]. Therefore, the addition of exogenous ComS can be used to induce the transient physiological bacterial state of natural competence in *S. suis* under laboratory conditions (Figure 6.2) [490].

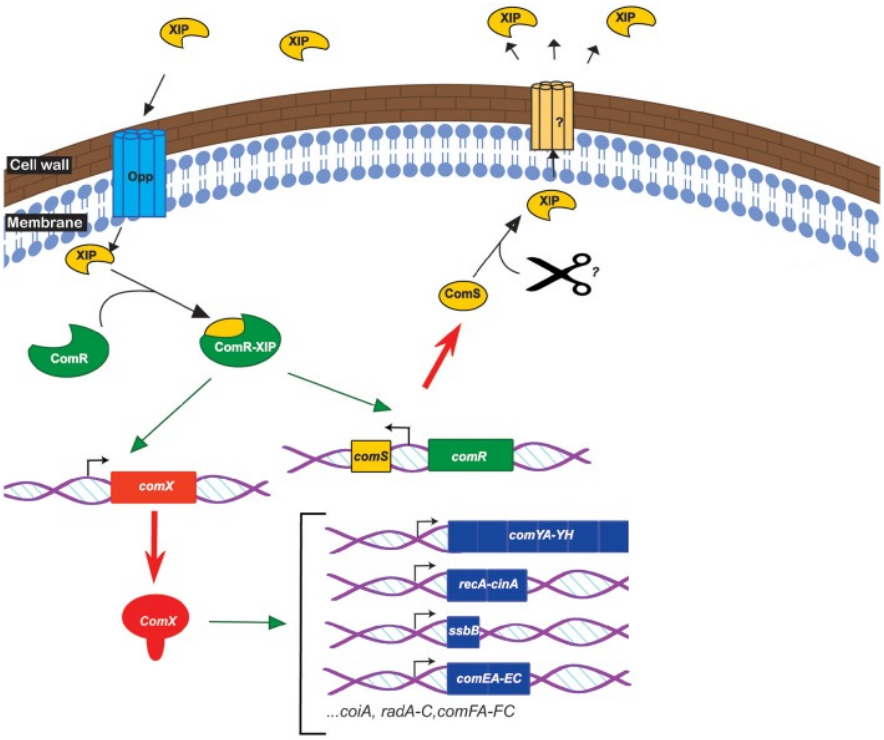

Figure 6.2. Model for the induction of natural transformation in *Streptococcus suis*.

Schematic representation of the current understanding of the process of natural transformation (or competence) in *S. suis*, adapted and updated from Zaccaria *et al.* [491]. Extracellular SigX-inducting peptide (XIP) enters the bacterial cell through the Opp transporter system (blue). Once intracellular, XIP binds to the transcriptional regulator (ComR), and the ComR:XIP complex promotes (green arrows) the expression of *comS* (encoding the full-length unprocessed form of the XIP pheromone) and *comX*. The ComX protein promotes expression (green arrow) of the late-competence genes. The ComS protein is processed by an unknown mechanism and secreted (also an unknown mechanism) inducing a positive feedback loop.

This chapter describes the design, construction and evaluation of two isogenic *virA* operon knockout mutants, one each in *S. suis* isolates LSS089 and P1/7. Preliminary studies looking at the positive detection of *virA* protein-encoding sequence in the original training collection revealed a number of non-disease associated isolates to possess a truncated *virA*, and are missing the second variable TRD (Figure 6.3). As the immediate flanking regions of disease-associated and non-disease associated isolates examined are almost identical, the removal of both versions of the *virA* (and *virA**) operon is described. Subsequent analyses evaluate the successfully generated LSS089Δ*virA* and P1/7Δ*virA* operon mutants and investigate the effect of knocking out the *virA* operon on growth characteristics and biofilm formation. Other assays to deduce function were planned, such as whole blood killing or phagocytosis assays, but could not be performed due to difficulties acquiring fresh blood at the time. Instead, this chapter describes investigation into the effect knocking out the *virA* operon in LSS089 and P1/7 had on the survival of the increasingly popular Greater Wax Moth experimental infection model (reviewed in section 1.4.4).

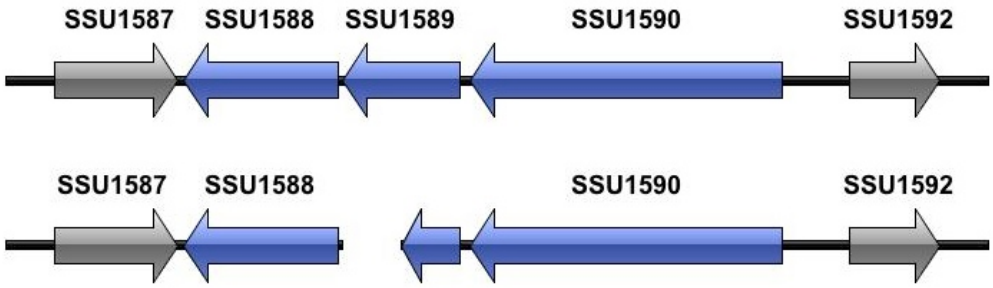

Figure 6.3. Schematic representation of the P1/7 *virA* operon (SSU1588-1590) in comparison to a truncated *virA operon in the non-disease associated field isolate LSS089.**

The *virA* operon (SSU1588-1590) of *S. suis* isolate (a) P1/7 (invasive disease-associated) in comparison to (b) the truncated *virA** operon of LSS089 (non-disease associated). Protein-encoding sequences of the *virA* operon are shown as blue arrows, with flanking region protein-encoding sequences in grey.

6.1.3 Chapter objectives

1. Design and construct an isogenic *virA* operon knockout mutant in *S. suis* isolates LSS089 and P1/7 using natural genetic transformation (competence)
2. Begin to understand the role of virA in pathogenesis by evaluating LSS089Δ*virA* and P1/7Δ*virA* operon knockout mutant strains in simple functional assays

6.2 Results

6.2.1 Differences in the *virA* protein-encoding sequence between invasive disease-associated and non-disease associated *Streptococcus suis* isolates

An investigation into the positive detection of the *virA* protein-encoding sequence revealed striking differences between invasive disease-associated and non-disease associated isolates in the original training collection. The *virA* protein-encoding sequence is 1206 bp in length, and was positively detected in 81% of the 53 invasive disease-associated isolates. In comparison, the 1206 bp *virA* protein-encoding sequence was positively detected in just 6% of the 62 non-disease associated isolates in the original training collection. Closer inspection of the BLASTn output revealed the *virA* operon (SSU1588-1590) unable to be detected in 71% of the 62 non-disease associated isolates. However, in the other 14 non-disease associated isolates a 618 bp truncated *virA** protein-encoding sequence was detected, that was missing the C-terminus TRD (Figure 6.4).

To begin to understand the role of *virA* in *S. suis* pathogenesis work began to design four isogenic mutants strains in LSS034 (arbitrarily chosen from the non-disease associated isolates with a truncated *virA** protein-encoding sequence) and P1/7 background, including i) a P1/7Δ*virA* operon knockout ii) a LSS034Δ*virA* operon knockout, as well as, two *virA* operon 'exchange' mutants where the *virA* operon of P1/7 was exchanged for the LSS034 *virA** operon and vice versa. Figure 6.5 shows a schematic representation of the strategy employed to knockout the *virA* operon in *S. suis* isolate P1/7. The same approach was employed to knockout the truncated *virA** operon in the non-disease associated field isolates LSS034. To avoid confusion, a schematic representation of the strategy employed to generate the two *virA* operon 'exchange' mutants has not been included in this thesis, as it was not possible to generate these desired mutants.

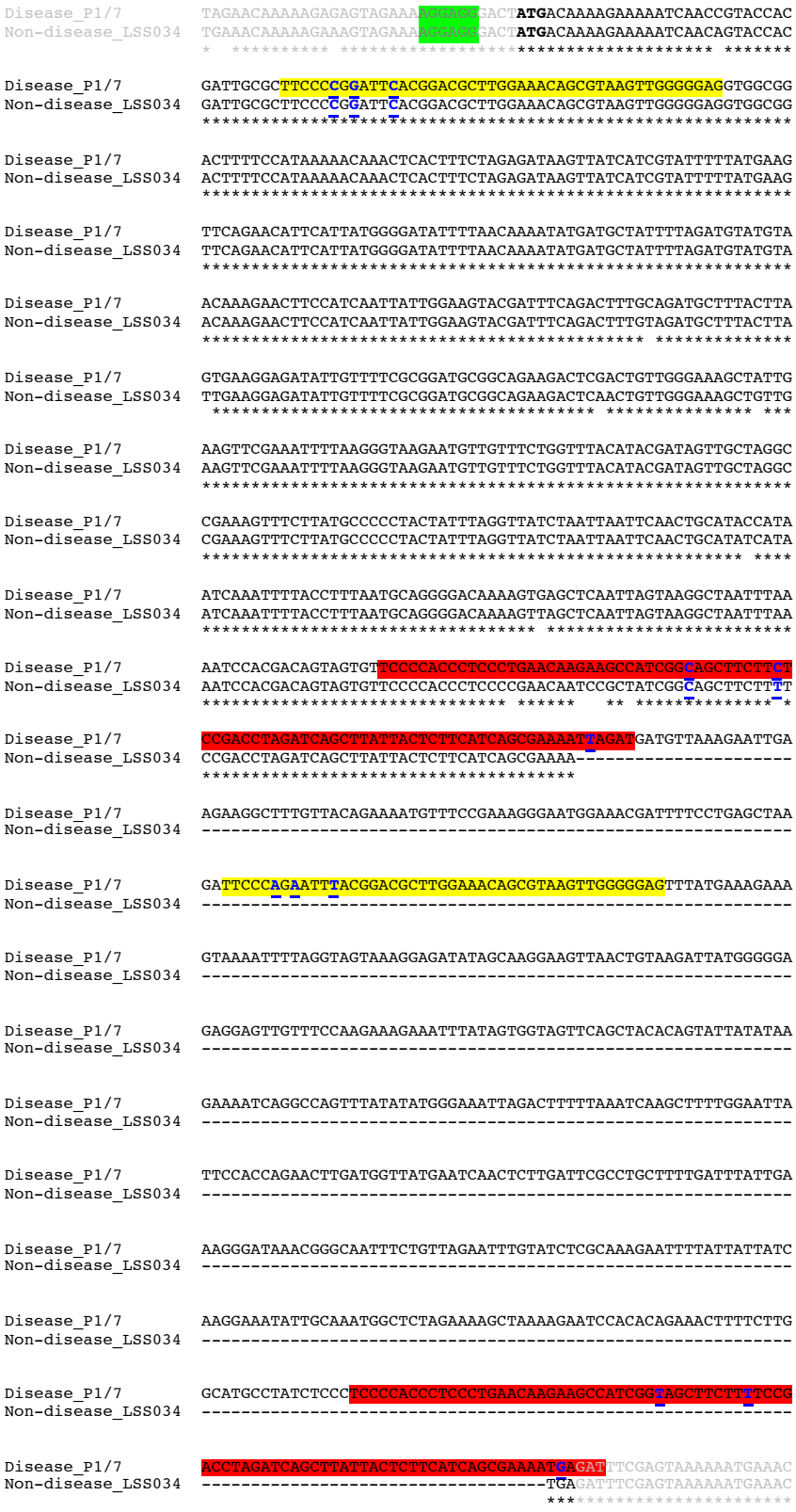

Figure 6.4. Clustal Omega alignment of the truncated *virA protein encoding sequence in LSS034 against the 'functional' *virA* in P1/7.**

Conserved repeat regions have been highlighted in red and yellow. Single nucleotide polymorphisms in conserved regions have been underlined in blue. The Shine-Dalgarno sequence is highlighted in green.

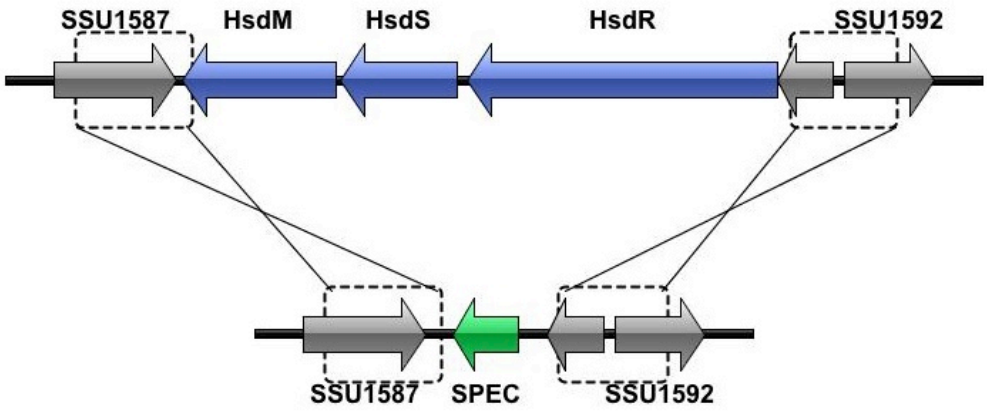

Figure 6.5. Design strategy to knockout the *virA* operon in *Streptococcus suis* isolate P1/7.

Schematic representation of the strategy employed to knockout the *virA* operon (SSU1588-1590) in *S. suis* isolate P1/7. Arrows represent protein-encoding sequences. The *virA* operon is shown in blue, and flanking regions in grey. Flanking regions amplified using standard spliced overlap extension PCR techniques are shown as dotted lines. The spectinomycin cassette is shown in green. The predicted double-crossover recombination event to integrate DNA constructs into the wild-type genome is shown.

6.2.2 Evidence for the construction of isogenic *virA* operon knockout mutants in LSS034 and P1/7 backgrounds

Standard spliced overlap extension PCR techniques [540] were used to create the double-stranded DNA constructs consisting of three parts i) a flanking region downstream of the operon, ii) a spectinomycin resistance cassette and iii) a flanking region upstream of the operon (Figure 6.6 (a)). Figure 6.6 (b) shows successful first round PCR amplification of the individual parts of each construct. Figure 6.6 (c) shows successful third round (final) spliced overlap extension PCR amplification of the DNA constructs to be transformed into LSS034 and P1/7. DNA constructs were purified before being checked for SNPs by Sanger sequenced (data not shown). Transformants cultured on solid agar plates supplemented with spectinomycin were checked using bespoke oligonucleotide primers. No transformants were generated using the LSS034 background strain. However, LSS034 and LSS089 have identical arrangements of the *virA** operon, and using the same experimental conditions LSS089Δ*virA* was successfully created.

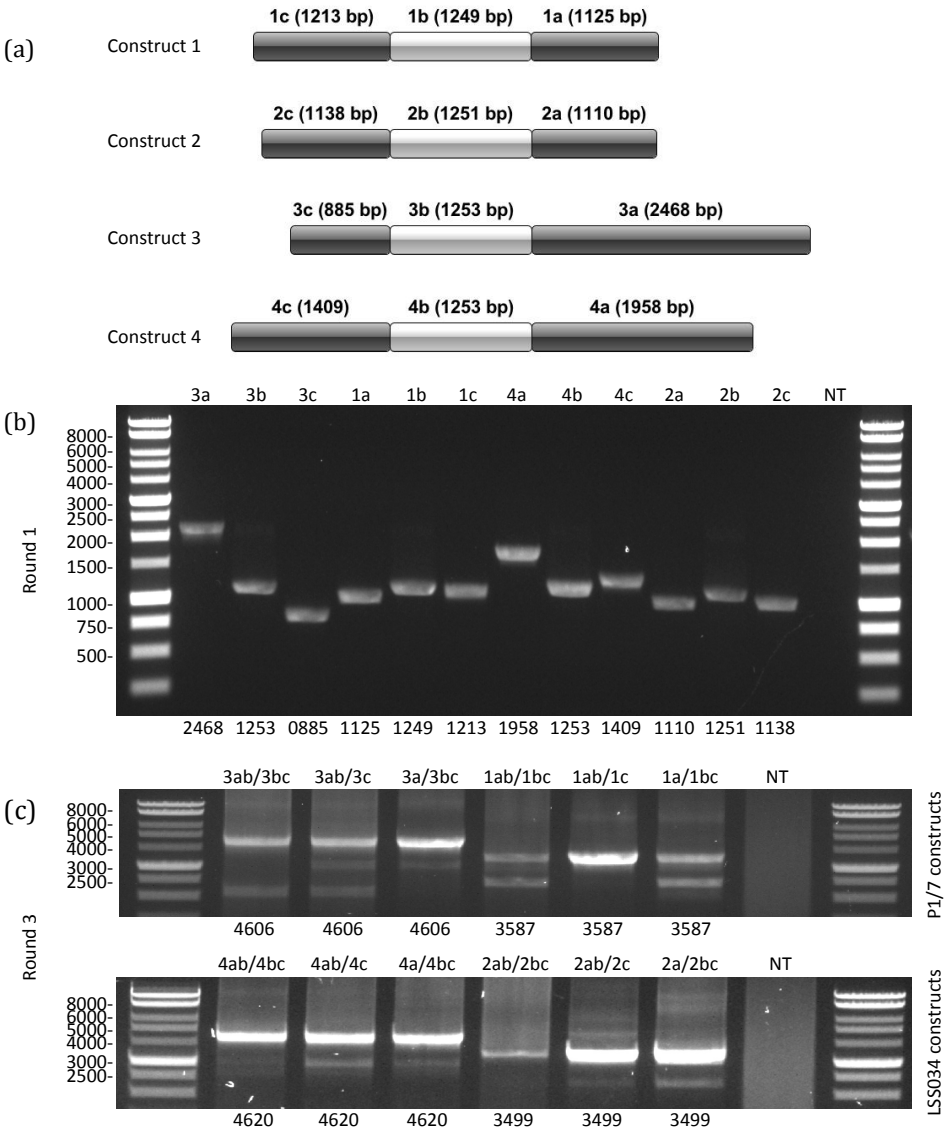

Figure 6.6. Construction of linear fragments of DNA for transformation into LSS034 and P1/7.

(a) Schematic overview of fragments required for synthesis of isogenic *virA* operon mutants. Construct 1 (P1/7) and construct 2 (LSS034) were designed to knockout the *virA* operon (SSU1588-1590). Construct 3 and construct 4 were designed for exchange of the *virA* operon between disease-associated (P1/7) and non-disease associated isolates (LSS034). (b) PCR products generated during the first round of spliced overlap extension PCR. (c) PCR fragments generated during the third (and final) round of spliced overlap extension PCR. All agarose gels contain PCR products generated from genomic DNA extracted from *S. suis* isolates LSS034 and P1/7 and LSS034. PCR amplicons were electrophoresed on a 0.7% (wt/vol) agarose gel containing SYBR Green DNA stain for 45 minutes at a constant 100 V and photographed under UV transillumination. The name of construct subunits is indicated above lanes. The anticipated size of PCR amplicons is indicated below lanes (bp) and corresponds to those in section (a). The first and last lane of each agarose gel contains 1x Promega 1 Kb DNA ladder with sizes indicated on the left (bp).

6.2.3 Knock out of the *virA* operon had inconsistent effects on *in vitro* growth

To investigate if knocking out the *virA* operon (SSU1588-SSU1590) had an effect on bacterial growth *in vitro* the strains LSS089Δ*virA* and P1/7Δ*virA*, as well as their respective parent strains, were grown under standard conditions in THB supplemented with 0.2% yeast extract and a growth curve performed. OD_{595} measurements were used as a proxy for bacterial growth, and were recorded every 1-hour over a 24-hour period. Figure 6.7 (a) shows two growth curves, one for the P1/7Δ*virA* operon mutant (grey-scale) and the second wild-type P1/7 (black). Both mutant and wild-type strains show similar growth characteristics, undergoing i) an initial one hour lag phase, ii) an exponential growth phase between one and five hours post inoculation and iii) a stationary phase between five and 12-hours post inoculation. At 12-hours post inoculation, rather than an expected decline in OD_{595} measurements typically associated with death phase, OD_{595} measurements for both mutant and wild-type strains increased in a manner characteristic of i) a second lag phase (12-15 hours post inoculation), ii) a second exponential growth phase (15-21 hours post inoculation) and iii) a second stationary phase (21-22 hours post inoculation) before a decline typical of death phase was observed in OD_{595} measurements (22-hours post inoculation onwards).

The non-disease associated *S. suis* field isolate LSS089 was observed to have significantly (*p*-value <0.05) different growth characteristics in comparison to the disease-associated wild-type P1/7 (Figure 6.7 (b)). Despite producing a similar general pattern of growth wild-type LSS089 was found to have a much shorter stationary phase (if at all) at approximately five to six hours post inoculation. Instead, in contrast to P1/7, wild-type LSS089 almost immediately entered a second lag phase (at five to six hours post inoculation) followed by a second exponential phase (8-17 hours post inoculation) before a decline in OD_{595} measurements typical of death phase began approximately 18-hours post inoculation.

Figure 6.7 (b) also shows the growth characteristics of the LSS089Δ*virA* operon mutant (grey-scale). In comparison to the wild-type parent strain, the LSS089Δ*virA* operon mutant had a

similar one hour lag phase. However, between two and five hours post inoculation the OD_{595} measurements of the wild-type increased significantly quicker than the mutant strain (*p*-value <0.05), indicating slower growth. Then at six hours post inoculation, as the wild-type strain briefly entered an initial stationary phase the LSS089Δ*virA* operon mutant did not, and instead appeared to continue to grow exponentially until ten hours post inoculation when it reached a peak OD_{595} of ~0.3 and entered stationary phase. OD_{595} remains constant until 15 h at which point LSS089Δ*virA* operon mutant entered death phase approximately three hours before the wild type strain.

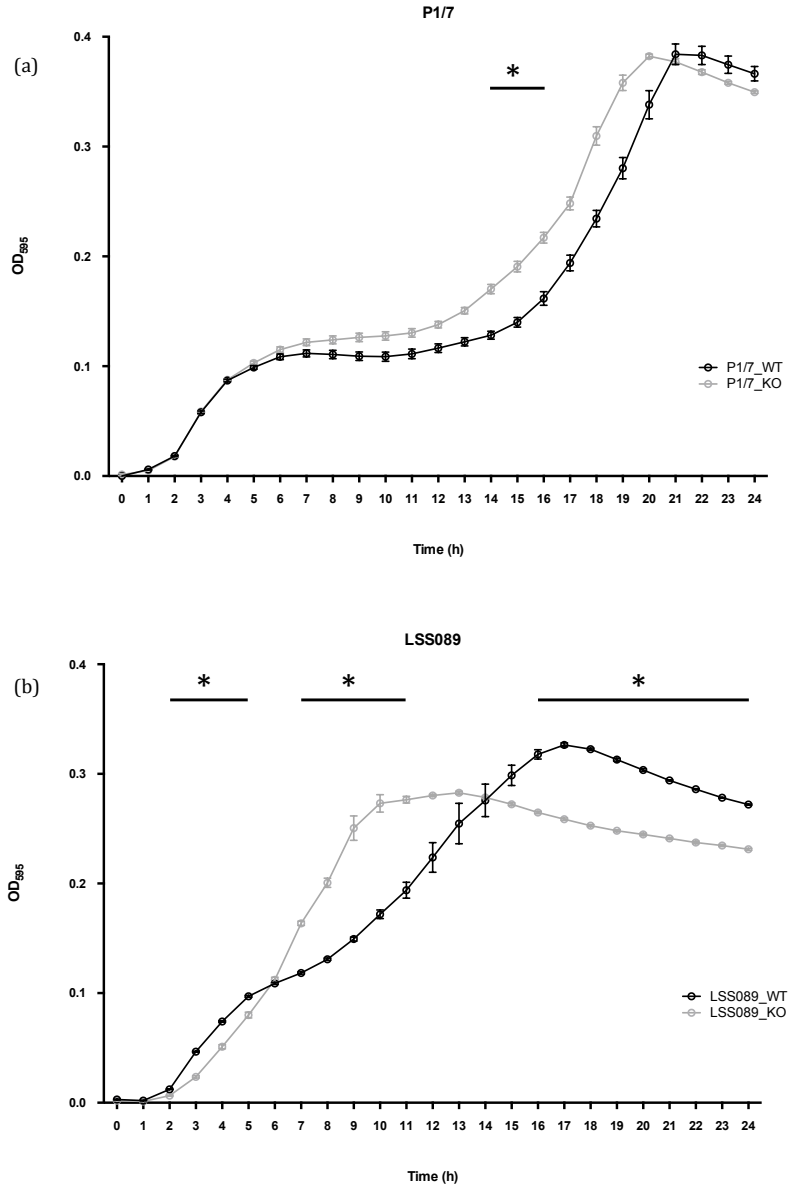

Figure 6.7. Growth curves for isogenic *virA* operon mutant and wild type parent strains.

Representation of *in vitro* growth characteristics for (**a**) P1/7Δ*virA* and (**b**) LSS089Δ*virA* mutants and their respective wild-type parent strains. An overnight culture of each strain was used to inoculate (1 in 1,000) fresh pre-warmed Todd-Hewitt broth supplemented with 0.2% yeast extract and then grown at 37 ˚C for 24 h without shaking. OD_{595} measurements were used as a proxy for bacterial growth, and were recorded every 1-hour over a 24-hour period. Each data point is representative of the mean and standard error of the mean (SEM) of three independent experiments (biological repeats). Multiple t-tests were performed to identify statistically significant (*p*-value <0.05, indicated by "*") differences between the growth curve of each isogenic mutant and its respective parent strain. The Holm-Sidak method was used to control for family-wise error associated with multiple sampling.

6.2.4 Knock out of the *virA* operon reduced biofilm formation

The ability to form a biofilm has been shown to be a factor associated with virulence in a number of bacteria, including *S. suis* [343, 534, 586]. To investigate if knocking out the *virA* operon had an effect on biofilm formation *in vitro* the strains LSS089Δ*virA* and P1/7Δ*virA*, as well as their respective parent strains, were grown under standard conditions in THB supplemented with 0.2% yeast extract and the formation of a biofilm assessed using a crystal violet microtiter plate assay (section 2.2.2.1). Figure 6.8 shows OD_{595} measurements after planktonic bacteria were removed at 24 hours post inoculation, and were used as a proxy for biofilm formation. Both mutant strains in comparison to their respective wild-type parent strains had a statistically significant (*p*-value <0.05) lower A_{595}, indicating biofilm formation was disrupted in these strains. No statistically significant difference in A_{595} was observed between wild-type parent strains (data not shown). Table 6.1 shows the test statistic and *p*-values calculated to identify significant differences in biofilm formation.

To better understand the effect of knocking out the *virA* operon on biofilm formation, all isolates of the original training collection (n=115) were subject to the crystal violet microtiter plate assay. Figure 6.9 shows OD_{595} measurements after planktonic bacteria were removed at 24-hours post inoculation, arranged in ascending order and coloured based on phenotype.

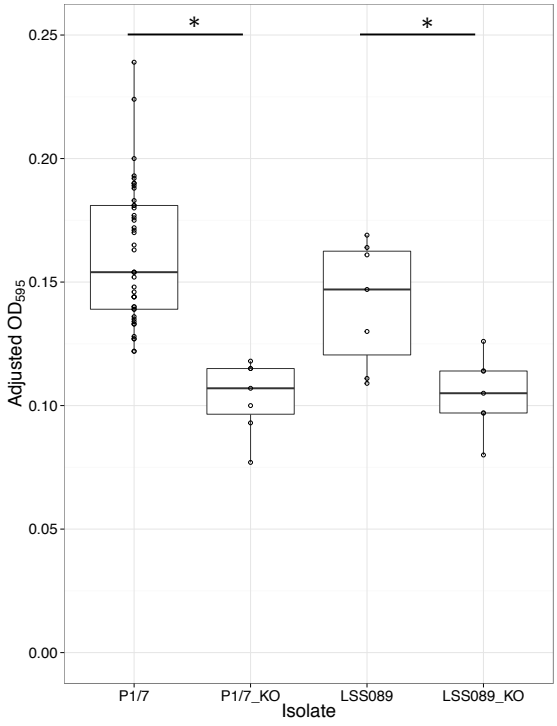

Figure 6.8. Biofilm formation *in vitro* by *virA* operon mutants was significantly lower in comparison to wild-type *Streptococcus suis* parent strains.

Biofilm formation was assessed using the crystal violet microtiter plate assay, where OD_{595} measurements after removal of planktonic bacteria were used as a proxy for biofilm formation. Student's t-test was used to test for statistically significant (indicated with an asterisk) differences in biofilm formation, assuming the true difference in means is equal to 0 unless *p*-value < 0.05.

Table 6.1. Test statistics and corresponding *p*-values to determine if knocking out the *virA* operon had an effect on *Streptococcus suis* biofilm formation *in vitro.*

Biofilm formation was assessed using the crystal violet microtiter plate assay, where OD_{595} measurements after removal of planktonic bacteria were used as a proxy for biofilm formation. Student's t-test was used to test for statistically significant differences in biofilm formation, assuming the true difference in means is equal to 0 unless *p*-value < 0.05.

Isolate	**Mean**	**Variance**	**Student's t-test** (t; ***p*-value**)
P1/7	0.16123	0.00076	8.2742; **$9.679e^{-7}$**
P1/7Δ*virA*	0.10357	0.00022	
LSS089	0.14157	0.00063	3.3294; **0.00783**
LSS089Δ*virA*	0.10471	0.00023	

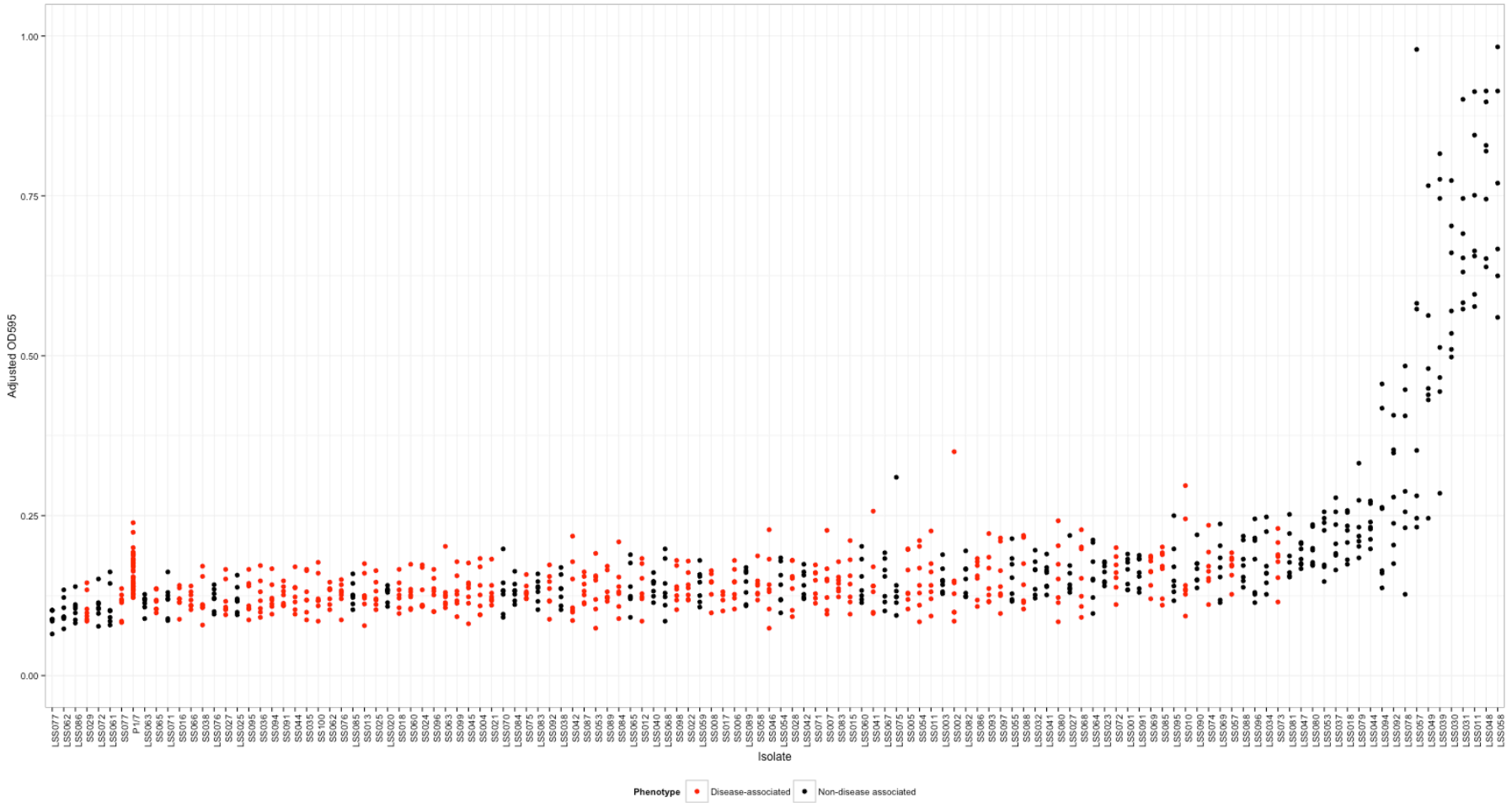

Figure 6.9. Biofilm formation *in vitro* by the 115 *Streptococcus suis* isolates of the original training collection.

Biofilm formation was assessed using the crystal violet microtiter plate assay assay, where OD_{595} measurements after removal of planktonic bacteria were used as a proxy for biofilm formation. Colours represent disease- (red)/non-disease (black) associated phenotypes. Isolates are ordered lowest to highest based on mean A_{595}.

6.2.5 Establishment of a Greater Wax Moth larvae experimental infection model

Greater Wax Moth (*G. mellonella*) larvae have recently been explored as a simple and cost-effective *in vivo* experimental infection model to screen the virulence potential of *S. suis* isolates [423]. To establish and evaluate the Greater Wax Moth larvae experimental infection model in our laboratory, *S. suis* isolates P1/7 (disease-associated) and Henrichsen S735 (deemed to be "weakly virulent" in newborn gnotobiotic pigs [284]) were used to infect groups of final instar stage caterpillars and larvae survival recorded every 24-hours over a 144-hour period post infection. Figure 6.10 shows similar dose-dependent larvae survival and an LD_{50} ($1x\ 10^7$ - $1x\ 10^8$ CFU/mL) to that previously reported for P1/7 and Henrichsen S735 [423]. The log-rank test was used to compare Kaplan-Meier curves of larvae survival and identified $1x\ 10^7$ CFU/mL to be the inocula dose capable of discriminating invasive disease-associated and "weakly virulent" isolates of *S. suis* (Table 6.2, *p*-value = 0.035). As a result, this technique was subsequently used to investigate if to investigate if knocking out the *virA* operon had an effect on virulence.

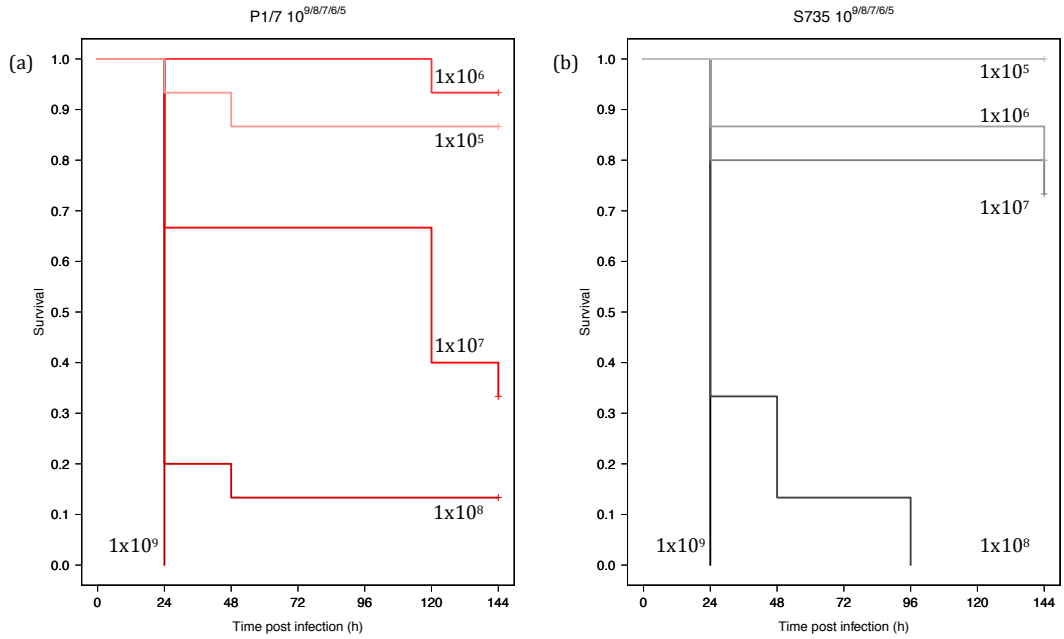

Figure 6.10. Kaplan-Meier curves of Greater Wax Moth larvae survival showing similar dose-dependant larvae survival to that previously reported following infection with *Streptococcus suis* isolates P1/7 and Henrichsen S735.

Kaplan-Meier curves of Greater Wax Moth larvae survival following experimental infection with *S. suis* isolate (**a**) P1/7 (invasive disease-associated) and (**b**) Henrichsen S735 (deemed to be "weakly virulent" in newborn gnotobiotic pigs [284]). A 20 µL inoculum of bacteria adjusted in PBS to 1x 10^9, 1x 10^8, 1x 10^7, 1x 10^6 or 1x 10^5 CFU/mL was used to infect groups of 15 final instar stage caterpillars via the last posterior left proleg. All experiments were repeated using larvae from a different batch, and data from replicate experiments pooled to achieve n=30. Colour/grey-scale has been used to indicate inocula dose.

Table 6.2. Comparison of Kaplan-Meier curves of Greater Wax Moth larvae survival following infection with *Streptococcus suis* isolates P1/7 and Henrichsen S735.

The log-rank test was used to compare Kaplan-Meier curves of larvae survival infected with inocula of different isolates (but the same dose). The inoculum of 1x 10^7 CFU/mL was identified as the dose capable of discriminating (*p*-value <0.05) P1/7 (disease-associated) and Henrichsen S735 (deemed to be "weakly virulent" in newborn gnotobiotic pigs [284]).

Inoculum (CFU/mL)	**1x 10^5**	**1x 10^6**	**1x 10^7**	**1x 10^8**	**1x 10^9**
Log-rank test (*p*-value)	0.150	0.287	0.035	0.746	-

6.2.5.1 The Greater Wax Moth larvae experimental infection model can differentiate disease-associated from non-disease associated strains

To further evaluate the use of Greater Wax Moth larvae survival as a robust experimental infection model for *S. suis*, four invasive disease-associated (Figure 6.11 (a-d)) and four non-disease associated (Figure 6.12 (a-d)) field strains recovered from pigs on UK farms were used to infect groups of final instar stage caterpillars and larvae survival recorded every 24-hours over a 144-hour period post infection. Figure 6.11 shows dose-dependent larvae survival for the four (a-d) invasive disease-associated isolates tested. Infection with an inoculum of 1x 10^9 CFU/mL resulted in no Greater Wax Moth larvae survival at 24-hours post infection. Infection with a 1x 10^8 CFU/mL inoculum of SS1008 resulted in ~40% larvae survival after 24-hours, reducing to ~20% at the end of the 144-hour post-infection time course. Other invasive disease-associated isolates resulted in less larvae survival at 24-hours post infection, SS1024: ~20%, SS1066: ~10% and SS1072: 0%. However, infection with an inoculum of 1x 10^7 CFU/mL invasive disease-associated strains resulted in significantly less larvae mortality (log-rank test; p-value <0.05),, with a minimum 60% larvae survival observed 144-hours post infection.

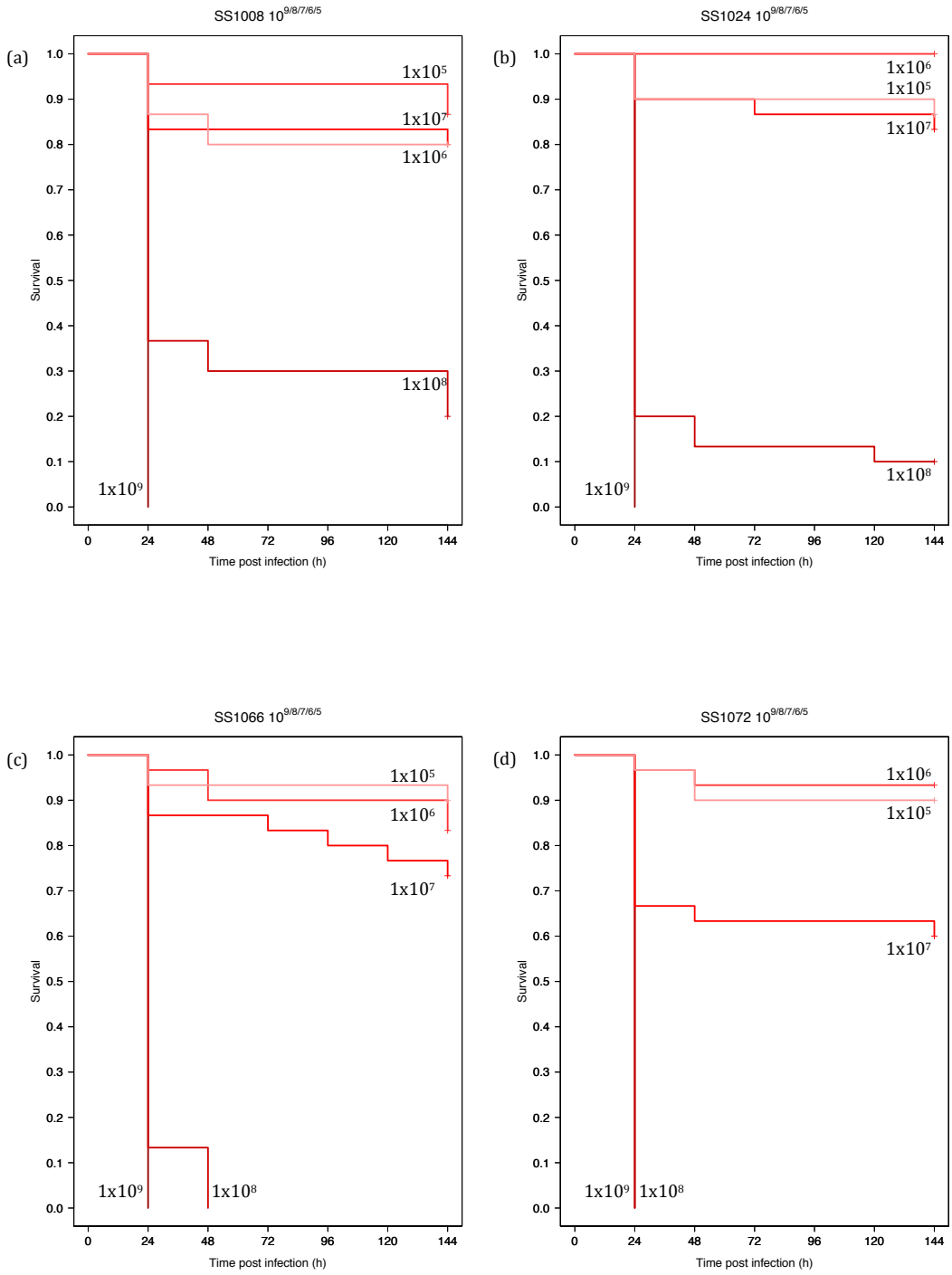

Figure 6.11. Kaplan-Meier curves of Greater Wax Moth larvae survival following experimental infection with invasive disease-associated isolates of *Streptococcus suis* recovered from UK pigs.

Kaplan-Meier curves of Greater Wax Moth final instar stage larvae survival following experimental infection with invasive disease-associated *S. suis* isolates (**a**) SS1008, (**b**) SS1024, (**c**) SS1066 and (**d**) SS1072. A 20 µL inoculum of bacteria adjusted in PBS to 1x 10^9, 1x 10^8, 1x 10^7, 1x 10^6 or 1x 10^5 CFU/mL was used to infect groups of 15 final instar stage caterpillars. All experiments were repeated using larvae from a different batch, and data from replicate experiments pooled to achieve n=30 per inoculum.

Infection of Greater Wax Moth larvae with non-disease associated *S. suis* isolates also resulted in dose-dependent larvae survival. Figure 6.12 shows dose-dependent larvae survival for the four (a-d) non-disease associated isolates tested. As reported for disease-associated isolates, infection with an inoculum of 1x 10^9 CFU/mL resulted in no Greater Wax Moth larvae survival at 24-hours post infection. Infection with an inoculum of 1x 10^8 CFU/mL resulted in larvae survival between 17% (632/06A) and 56% (427/05B) after the 144-hour post infection time course, with no strains causing 100% larvae mortality. Infection with an inoculum of 1x 10^7 CFU/mL resulted in at least 75% larvae survival being observed 144-hours post infection.

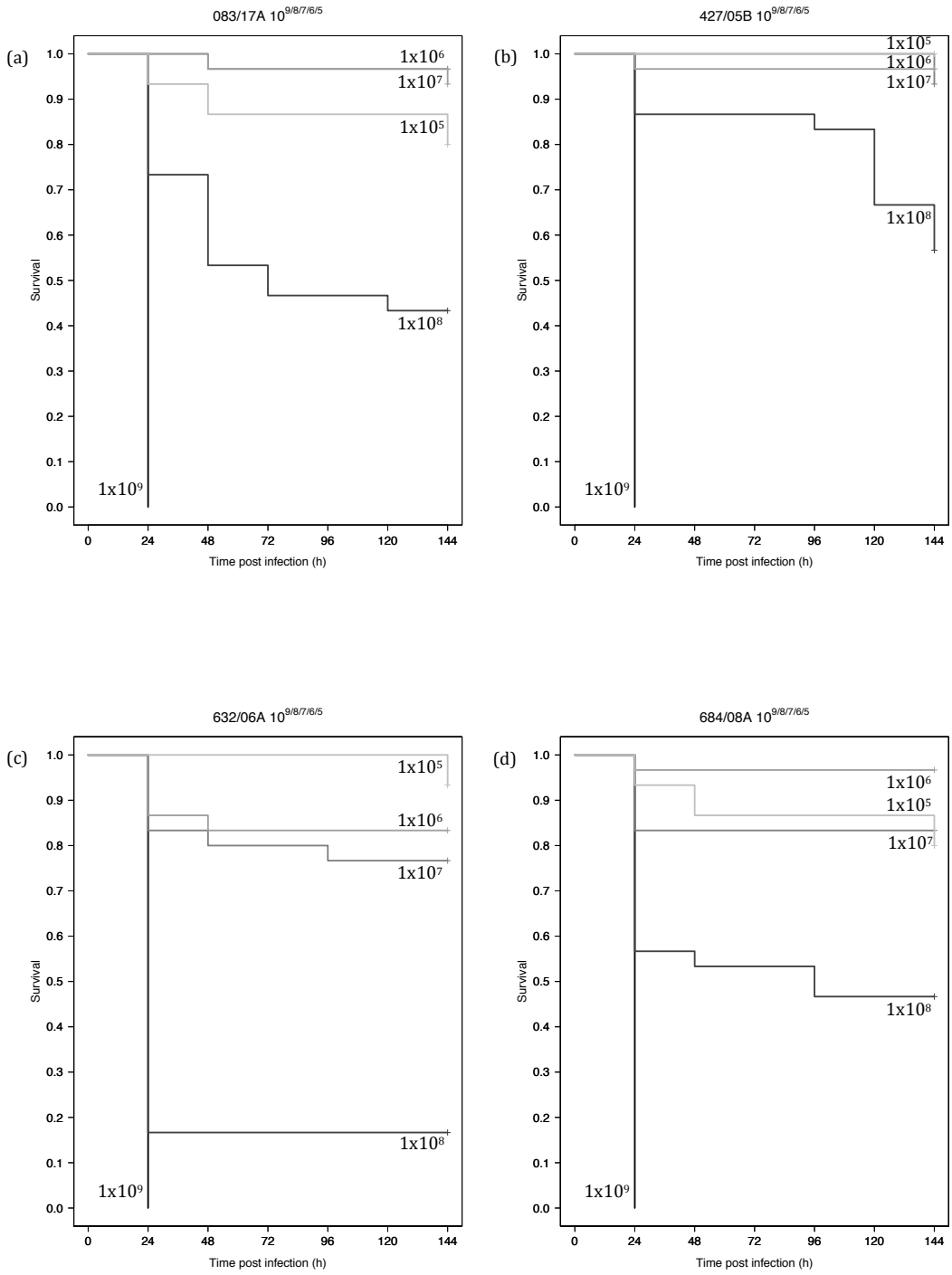

Figure 6.12. Kaplan-Meier curves of Greater Wax Moth larvae survival following experimental infection with non-disease associated isolates of *Streptococcus suis* recovered from UK pigs. Kaplan-Meier curves of Greater Wax Moth final instar stage larvae survival following experimental infection with non-disease associated *S. suis* isolates (**a**) 083/17A, (**b**) 427/05B, (**c**) 632/06A and (**d**) 684/08A. A 20 µL inoculum of bacteria adjusted in PBS to 1x 10^9, 1x 10^8, 1x 10^7, 1x 10^6 or 1x 10^5 CFU/mL was used to infect groups of 15 final instar stage caterpillars. All experiments were repeated using larvae from a different batch, and data from replicate experiments pooled to achieve n=30 per inoculum.

To evaluate differences in Greater Wax Moth larvae survival following infection with isolates grouped into invasive disease-associated and non-disease associated phenotypic groups, larvae survival for all strains of a phenotypic group (disease/non-disease associated) were drawn on one plot. Figure 6.13 (a) shows the larvae survival following infection with 1x 10^8 CFU/mL of the four invasive disease-associated (red) and four non-disease associated (black) strains. The mean larvae survival of phenotypic groups (disease-associated/non-disease associated) was then calculated and plotted with 95% confidence bounds (Figure 6.13 (b)). The log-rank test was then used to demonstrate a statistically significant difference in the mean larvae survival between phenotypic groups (*p*-value = 0.00000598). No significant difference in larvae survival was observed between phenotypic groups when using an inoculum of 1x 10^7 CFU/mL.

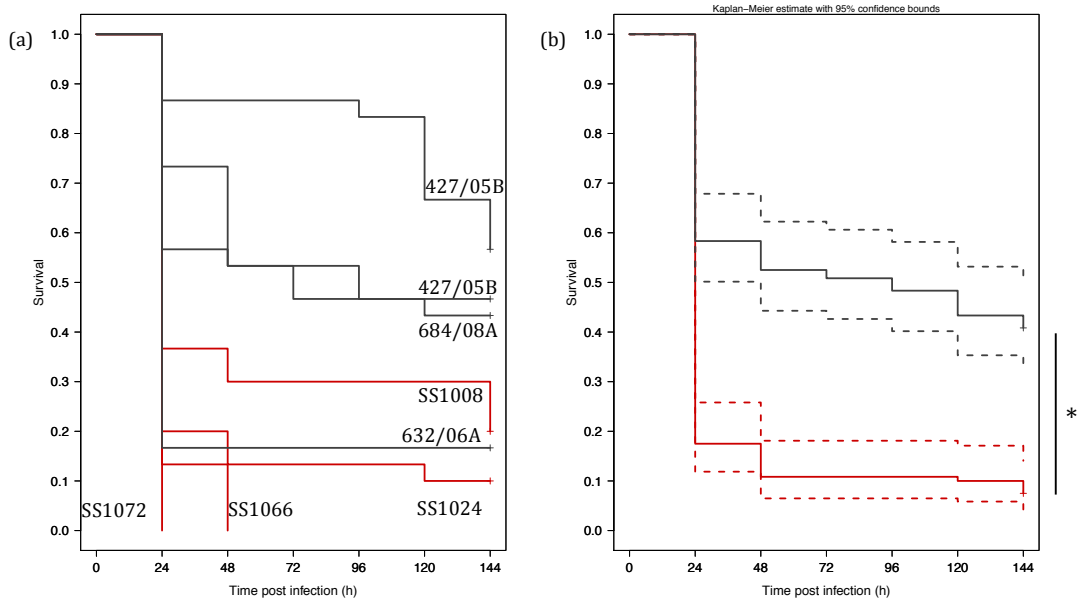

Figure 6.13. Kaplan-Meier curves of Greater Wax Moth larvae survival following experimental infection with disease- or non-disease associated *Streptococcus suis* inocula at 1x 10^8 CFU/mL.

Kaplan-Meier curves of Greater Wax Moth final instar stage larvae survival following experimental infection with invasive disease-associated (red) and non-disease associated (black) *S. suis* isolates. (**a**) To evaluate differences in larvae survival following infection with isolates grouped into invasive disease-/non-disease associated phenotypic groups, larvae survival for all strains of a phenotypic group were drawn on one plot. (**b**) Mean larvae survival of phenotypic groups was calculated and plotted with 95% confidence bounds (dotted lines). The log-rank test was used to identify statistically significant differences between mean larvae survival, and is indicated with an asterisk (*p*-value <0.05).

6.2.5.2 Removal of the *virA* operon had no effect on Greater Wax Moth larvae survival in comparison to wild-type parent strains

To investigate if knocking out the *virA* operon had an effect on Greater Wax Moth larvae survival the strains LSS089Δ*virA* and P1/7Δ*virA*, as well as their respective parent strains, were grown under standard conditions and then used to inoculate final instar stage caterpillars and survival recorded every 24-hours over a 144-hour post infection time course. Observing 3R guidelines only the highest three inocula doses (1x 10^9, 1x 10^8 and 1x 10^7 CFU/mL) were used to *reduce* the number of larvae per experiment as no significant difference was observed in larvae survival infected with *S. suis* at 1x 10^6 and 1x 10^5 CFU/mL (Figures 6.10/11/12). Figure 6.14 shows larvae survival following experimental infection with (a) wild-type P1/7, (b) P1/7Δ*virA* operon mutant, (c) wild-type LSS089, and (d) LSS089Δ*virA* operon mutant. No statistically significant (*p*-value <0.05) difference at any of the three inocula doses were observed between wild-type P1/7 and the P1/7Δ*virA* operon mutant (log-rank test). A similar result, where no statistically significant (*p*-value <0.05) difference at any of the three inocula doses, was also observed between wild-type LSS089 and the LSS089Δ*virA* operon mutant. In addition to the aforementioned comparisons, a comparison of Greater Wax Moth larvae survival experimentally infected with each wild-type strain was performed and also showed no significant difference at any of the three inocula doses.

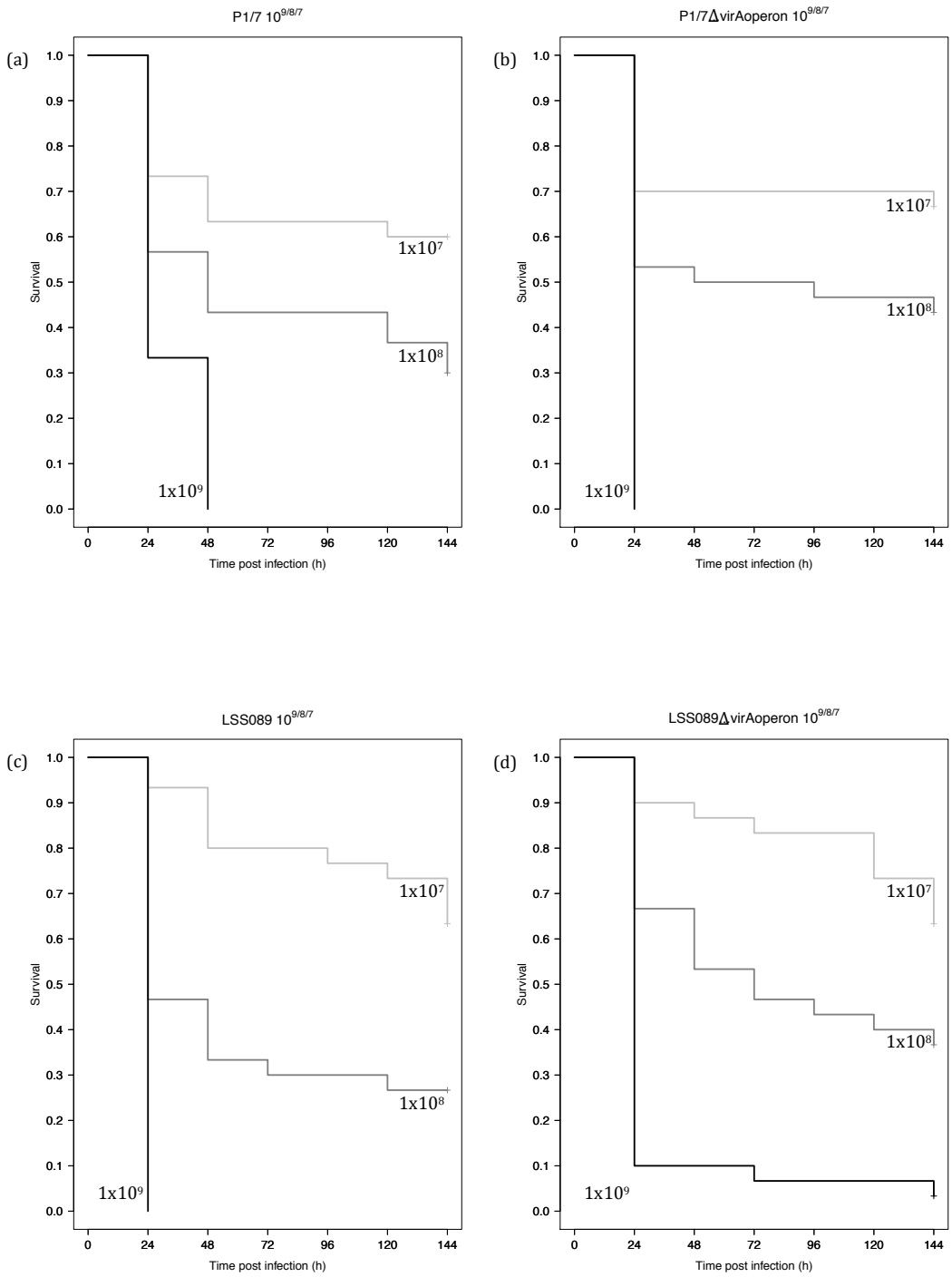

Figure 6.14. Kaplan-Meier curves of Greater Wax Moth larvae survival following experimental infection with *virA* operon mutants.

Kaplan-Meier curves of Greater Wax Moth final instar stage larvae survival following experimental infection with (**a**) P1/7, (**b**) P1/7Δ*virA* operon mutant, (**c**) LSS089, and (**d**) LSS089Δ*virA* operon mutant. A 20 µL inoculum of bacteria adjusted in PBS to 1x 10^9, 1x 10^8 and 1x 10^7 CFU/mL was used to infect groups of 15 final instar stage caterpillars. All experiments were repeated using larvae from a different batch, and data from replicate experiments pooled to achieve n=30 per inoculum.

6.3 Discussion

Objective 1. Having described, in chapter 3, the identification of novel genetic markers associated with *S. suis* isolates causing invasive disease in pigs on UK farms, the research described in this chapter attempted to begin to address the lack of experimental evidence for the newly defined pathotyping marker SSU1589 (also known as *virA* [378]) by ambitiously setting out to create four independent isogenic *virA* operon mutants using the only recently described in *S. suis* process of natural genetic transformation (competence) [490-492]. Two *virA* operon (SSU1588-1590) knockout mutants were successfully generated one each in the invasive disease-associated isolate P1/7 and the non-disease associated field isolate LSS089. Frustratingly, it was not possible to generate faithful *virA* operon 'exchange' mutants, where the operon of P1/7 was replaced with the truncated *virA** operon of LSS089 (and vice-versa). Instead, short-read Sanger sequencing of the operon 'exchange' transformants revealed that two unwanted hybrids had been created, whereby the recombination events occurred at undesirable locations. Clearly, the cloning strategy for operon exchange was suboptimal. With more time, two possible strategies could be implemented to create the desired operon 'exchange' mutants. Firstly, rethinking the cloning strategy using shorter flanking regions may prevent the incorrect crossover event. Alternatively, the two *virA* operon knockout mutants described in this chapter could be utilised for subsequent transformation with *virA* operons from P1/7 and LSS089 respectively - akin to a complement mutant design strategy.

Natural competence has been described in ~80 bacterial species, however a complete understanding of the mechanisms underlying this process is not available [490, 585]. What is known is that competence in *S. suis* can be induced by the extracellular pheromone encoded by *ComS* (SSU0050) [490]. Due to this, the isolates of the training collection were screened for the presence of *ComS*, and only isolates with 100% matches to *ComS* (Appendix Table 8.1) selected as candidates for mutation. However, despite non-disease associated field isolate LSS034 encoding a 100% match to *ComS* in P1/7 it was not possible to induce the process of competence in this strain. The reason for this is unclear and could be due to a number of reasons, including but not

limited to i) other genes of the natural competence cascade being absent from LSS034, ii) premature switching off of natural competence, due to a lack of our current understanding of the mechanisms mediating shut down, iii) composition of the growth medium, iv) bacterial density, or v) a lack of other necessary environmental cues [491]. Because there are many factors that have the potential to affect this tightly regulated process, it is difficult to determine which factor played the/an important role in preventing the process of natural competence in LSS034. Curiously, under the same experimental conditions, it was possible to induce the process of natural competence in non-disease associated field isolate LSS089, and go on to generate the desired *virA** operon knockout mutant. Therefore, we hypothesise that the experimental conditions and the competence peptide itself were suitable for the induction of natural competence, and that the inability to induce competency was with the machinery of LSS034.

Objective 2. Evaluation of the growth characteristics of the *virA* operon mutants in comparison to their wild-type parent strain produced interesting results. The P1/7Δ*virA* operon mutant and wild-type P1/7 strain showed very similar growth characteristics *in vitro* across the duration of the experiment (24-hours), indicating the 'functional' *virA* operon does not play a role in regulating growth in artificial growth media (Figure 6.7 (a)). Although, it is important to acknowledge that this might not be the case in the natural host, where such a high abundance of growth favouring nutrients is unlikely to be found. In comparison, knocking out the truncated *virA** operon in LSS089 appeared to result in significantly faster bacterial growth when compared to the wild-type parent strain (reaching peak OD_{595} sooner because of the skipped second lag phase). The reason for this is unclear, although it is interesting to speculate that the synthesis of the proteins encoded by the *virA* operon require a significant energy expenditure and that the subsequent removal of the operon allows the energy to be redirected to bacterial growth. However, this observation was not replicated when knocking out the full-length 'functional' *virA* operon in P1/7 and is therefore unlikely to be true. Alternatively, the methyltransferase activity of the protein encoded by SSU1588 could be utilised by one and/or both of the two other Type I RM systems known to exist in *S. suis* [552]. Therefore, it might be possible that knockout of the

whole *virA* operon (that includes SSU1588) could be resulting in reduced methylation of promoter sequences of genes restricting/limiting exponential bacterial growth.

The ability of *S. suis* isolates to produce a biofilm has previously been reported [532-534, 587, 588] and as in other bacteria proposed to be a virulence factor [589], where establishing a community structure provides i) enhanced protection from environmental stress, ii) increased exchange of genetic material and iii) increased resilience to antibiotics, all of which result in persistent colonisation [590]. Knocking out the *virA* operon in both LSS089 and P1/7 resulted in a statistically signification (p-value <0.05) reduction in biofilm, indicating that *virA* is in some capacity involved in biofilm production; most likely in regulation of the genes responsible for biofilm production rather than the actual process itself [378]. To better understand the effect and context of knocking out the *virA* operon on biofilm formation, all isolates of the original training collection (n=115) were subject to the crystal violet microtiter plate assay (Figure 6.9). Comparison to the mutants to the data presented in Figure 6.5 indicated that while statistically significant in comparison to their respective wild-type parent strains (t-test; p.value<0.05), the effect of knocking out the *virA* operon on biofilm production is probably not biologically significant. Indeed, the observation that removing the virA operon had a significant effect on biofilm formation may be an effect of the limited sensitivity of the microtiter plate assay. It might also be an effect of using 96-well tissue culture plates that had been treated to enhance cell adhesion, which with hindsight should have been avoided. As an alternative, the presence of biofilms could be visualised using scanning electron microscopy. Although for improved visualisation of the structural architecture, transmission electron microscopy should be used [533]. However, both are methods generate qualitative rather than quantitative measures and precise differences between strains would be difficult to determine.

Greater Wax Moth larvae have recently been explored as a simple and cost-effective *in vivo* experimental infection model to screen the virulence potential of *S. suis* isolates [423]. Described within this chapter is evidence that the increasingly popular Greater Wax Moth larvae

experimental infection model has been successfully established for the first time at the Department of Veterinary Medicine (University of Cambridge). In fact, the characterisation of P1/7 and Henrichsen S735 was reproducible from previously published work (Figure 6.10) [423]. Positively, a statistically significant (*p*-value <0.05) difference in larvae survival experimentally infected with 10^8 CFU/mL was observed between the invasive disease-associated P1/7 and "weakly virulent" [284] Henrichsen S735. Such an observation was also observed between disease-associated and non-disease associated field isolates recovered from pigs on farms in the UK (Figure 6.13), indicating the Greater Wax Moth larvae is potentially a reproducible and robust alternative experimental infection model for *S. suis.*

To investigate if knocking out the *virA* operon had an effect on Greater Wax Moth larvae survival the strains LSS089Δ*virA* and P1/7Δ*virA*, as well as their respective parent strains, were grown under standard conditions and then used to inoculate final instar stage caterpillars and survival recorded every 24-hours over a 144-hour post infection time course. Based on observed phenotypic data and the site of isolation LSS089 was deemed to be a non-disease associated *S. suis* isolate (see Table 2.1). However, Greater Wax Moth larvae survival following infection with LSS089 was similar to that of larvae infected with the invasive disease-associated isolate P1/7, and indeed any differences between larvae survival was not deemed to be statistically significant (*p*-value <0.05). The reason for this observation is unclear, although could again be due to a number of reasons. Firstly, it may be that the effect of knocking out the *virA* operon is masked by one or both of the two other Type I RM systems also known to be present in P1/7 [552]. To address this, it may be useful to create a P1/7 mutant strain that has had all three Type I RM systems knocked out, although this would not allow identification of the exact biological function specifically of *virA* but the effect of type I RM systems in general. Equally, it might be that the Greater Wax Moth larvae are unable to show subtle differences of knocking out the *virA* operon. Other experimental infection models exist for *S. suis*, and have been reviewed in chapter 1 (see section 1.1.4). Another inexpensive model, also exempt from ethical legislation protecting terrestrial mammals, is the zebrafish (*Danio rerio*) experimental infection model [424], that might

show further phenotypic characteristics and is an option for further rapid screening of the *virA* operon mutants described in this chapter. Another alternative, and to my knowledge unexplored, non-mammalian infection model is the *Caenorhabditis elegans* model. However, establishment of this model is not straightforward and would require working at a highest temperature of 21 ˚C rather than 37 ˚C (personal communication with Professor Jerry Wells and Professor Jan Kammenga, University of Wageningen, the Netherlands). Other previously used experimental infection models for *S. suis* include CF-1 mice [319], however, the extrapolation of findings in mice to pigs has been inconsistent with strains found to be virulent in mice being deemed non-virulent in pigs (and vice-versa) [336, 440]. The optimum experimental infection model is and will always be the natural host, and a number of options exist. Since *S. suis* is endemic in almost all pig populations sampled [11], it has been suggested that CDCD pigs are the optimum model. But CDCD pigs are expensive and are often very sensitive, meaning that the extrapolation of findings in CDCD pigs to pigs born naturally and exposed to maternal antibodies has been questioned [302]. Most recently, whole blood killing and phagocytosis assays have been used to show significant differences between Type I RM system isogenic mutants and their parent *S. suis* strains [584]. Similar experiments were planned in our laboratory but could not be performed due to difficulties acquiring fresh blood at the time, and are an important model that should be explored in the future.

In conclusion, knocking out the *virA* operon in P1/7 resulted in no significance difference in growth characteristics, biofilm formation (no biological significance) or in Greater Wax Moth larvae survival. Future evaluation of the role and biological function of *virA* will likely use SMRT DNA sequencing (Pacific Biosciences) with methylome analysis to identify the exact nucleotide target sequence recognised by the protein encoded by *virA* (SSU1589). SMRT DNA sequencing will also be able to provide information into whether the effect of knocking out the *virA* operon is masked by the two other Type I RM systems in P1/7. Other work, not related to *virA*, could take advantage of the biofilm formation data presented in this chapter, and using a GWAS it may be

possible to identify all of the genes involved in *S. suis* biofilm formation, as well as, assess their potential as genetic markers of invasive disease.

Chapter 7

General discussion, conclusions and future work

The work described in this thesis aimed to identify novel genetic markers of disease-association among strains of *S. suis*. With access to an initial (training) collection of 115 isolates recovered from pigs on UK farms and detailed clinical metadata, I aimed to identify regions of the genome associated with invasive disease caused by *S. suis*. I then investigated the potential practical utility of these genetic markers, and compared their use in predicting the virulence potential of an out-of-sample collection of *S. suis* isolates to the use of commonly used virulence-associated factors also used to predict virulence potential. Further investigations, using these novel genetic markers, estimated the bacterial load of invasive disease-associated and non-disease associated strains of *S. suis* on the palatine tonsils of pigs, in doing so allowing comparisons to be drawn against culture-based approaches for surveillance. Finally, attempts to investigate the function and biological role of one of the recently defined pathotyping markers, SSU1589 (also know as *virA*, denoting virulence gene A [378]), have been described that has resulted in the generation of two isogenic *virA* operon knockout mutant strains; LSS089Δ*virA* and P1/7Δ*virA*.

As a result of advances and reductions in cost, the whole-genome sequencing of multiple isolates of the same species has opened up exciting possibilities to reassess diagnostic and surveillance programs for bacteria, including *S. suis*. Taking advantage of these advances I have presented evidence for the use of three genetic markers (two associated with invasive disease and one associated with asymptomatic carriage on the palatine tonsils of pigs on UK farms) and a novel species-specific marker to pathotype *S. suis*. There is an urgent need for an agreed set of criteria to define an isolate of *S. suis* as virulent and pathotyping, differentiating isolates based on their ability to cause invasive disease, could be the approach to do this. The pipelines put together to generate the data in this thesis can quickly be adapted to incorporate additional strain collections, for example from Canada, China and/or Vietnam, and add to those markers already described.

Having used a UK biased collection of isolates in this study, one of the original aims of this PhD research was to return to and rerun the pipeline of statistical analyses I have built to identify genetic markers of disease association, but also include the whole-genome sequencing data of

S. suis isolates from different geographic locales. Recently the whole-genome sequencing data of a collection in excess of 250 *S. suis* isolates recovered from pigs on farms in Canada has become available (for which I undertook the sample preparation for whole-genome sequencing). Further to the research presented in this thesis, it is an exciting prospect to be able to identify and inspect the genetic markers associated with invasive disease in these Canadian isolates. It is also an exciting prospect to see how adapting our methods to include a random forest machine learning statistical approach [591-595] to identify and rank genetic markers would influence the markers shortlisted as potential pathotyping markers for *S. suis* – an approach not used in this thesis. Other unpublished collections of *S. suis* isolates from China and Vietnam are now also available to our group and incorporation of all of these isolate collections would only add statistical power to any subsequent analyses; increasing the size of the dataset should not be a problem for the computationally efficient statistical tools employed, such as the Discriminant analysis of principal components [518]. Importantly, each strain collection has its own caveats, one example is that the Vietnamese collection has a strong serotype 2 bias due to the sampling strategy used by our collaborators, a bias I attempted to avoid during the research presented here by not restricting analyses to a single serotype. One common limitation that affects all isolate collections available is, in the absence of an agreed superior alternative, the use of observed clinical metadata as the 'gold-standard' to characterise isolates as virulent. This is an approach not without fault for reasons discussed throughout this thesis, including i) host immune status, ii) concurrent infections and iii) welfare conditions that could influence the susceptibility of a host to this pathogen. In fact, due to these uncertainties over the 'gold standard' any comparison of sensitivity and specificity against this should be treated with caution.

With more time, one of the first and most interesting follow on analyses to this thesis would be to screen the Canadian isolates for the newly-defined pathotyping marker *virA*. Clear geographic effects have been described in the *S. suis* literature regarding serotype [17], sequence type [90] and the positive detection of the virulence-associated proteins EF, MRP and SLY [11]. While the positive detection of serotype 2, *epf*, *mrp* and/or *sly* is considered a good indicator of disease in

certain European countries the genes encoding EF, MRP and SLY are often absent from both disease-associated and non-disease associated isolates recovered from pigs on farms in North America [129, 393, 408, 409, 596]. By expanding the number of genome sequences included in analyses it would be desirable to be able to create a diagnostic/surveillance tool that could be used globally. However, as disease caused by *S. suis* appears to be influenced by the environment and immune status of the host (and not solely by *S. suis* genetics) this might not be possible, resulting in surveillance tools designed to target specific continents, countries or a combination of geographic locales. Indeed, it would also be interesting to screen the strain collection of 116 isolates recently recovered from pigs on farms in the Netherlands and used to assess *S. suis* population structure and the distribution of Type I RM systems in *S. suis* [552, 597, 598]. As numerous non-disease associated isolates in our UK biased strain collection do not possess the Type I RM enzyme encoded by the operon SSU1588-1590, I suggest this is not, as proposed by Willemse *et al.* [552], a housekeeping operon. However, if all 116 Dutch isolates have been recovered from systemic locations of diseased pigs this is further evidence that *virA* could be a robust genetic marker of virulence in European countries. Indeed, arguably *virA* could be described as the 'best' single molecular marker to pathotype *S. suis* that I have identified. Having a single genetic marker, as opposed to multiple markers, is preferable for future infection dynamics studies, because multiple strains of *S. suis* can colonise the same tonsil and this would complicate the interpretation of multiplex-PCR results.

Alternatively, an option not yet explored is the incorporation of *epf*, *mrp* and *sly* into the newly described pathotyping tool for *S. suis*. Due to the usefulness of *epf*, *mrp* and *sly* being restricted to European countries and the original aim of this project to develop a pathotyping tool for global use, the approach I used to identify genetic markers tried not to place substantial emphasis on predicted biological function, as current annotations can be inaccurate. This strategy should also prevent the selection of genetic markers for *S. suis* that may have simply been associated with virulence in other bacterial species. Indeed, *virA* exemplifies this; before more recent papers proposed epigenetic modification by Type I RM systems is involved, not only in host bacterium

defence against foreign DNA, but also in the stabilisation of mobile genetic elements and in gene regulation numerous questions have been asked as to why or how a Type I RM system is involved in pathogenesis and virulence (personal communication with Professor Duncan Maskell, University of Cambridge).

In some isolates of the training collection, *epf, mrp, sly* and *virA* were all positively identified to be present in the same isolate (chapter 4). Therefore, incorporation of these genes into our pathotyping tool for *S. suis* might not significantly improve the sensitivity and/or specificity of binary classification. Based on this observation, it is interesting to speculate that the methyltransferase activity of the Type I RM system, for which *virA* encodes the specificity domain, has a regulatory effect on the expression of EF, MRP and SLY by modifying the bases of the upstream promoter regions and masking (or revealing) the DNA from transcription enzymes. Exciting follow up studies are planned to use RNA sequencing microarrays to investigate the effect of knocking out the *virA* operon (SSU1588-1590) in *S. suis* isolate P1/7 has on transcription, and could reveal which genes are susceptible to regulation by *virA*. In addition, the isogenic P1/7Δ*virA* operon knockout mutant strain is planned to be sequenced using SMRT DNA sequencing with methylome analysis and, with comparison to the wild-type P1/7 parent strain, this might reveal the exact bipartite DNA sequence recognised by the protein encoded by *virA*.

An important finding of the work described in chapter 5 is that isolates of *S. suis* associated with invasive disease were found to coexist with non-disease associated isolates on the palatine tonsil of pigs on farms without obvious signs or a recent history of streptococcal disease. When compared to culture-based surveillance technologies, the frequency of this observation was higher and could indicate that culture-based approaches are underestimating the true positive rate of subclinical carriage of disease-associated strains in UK pig populations. With more time this observation would have been followed up by using quantitative bacteriology to validate the ratios predicted by qPCR of non-disease associated:disease-associated for the same swab. It is important to mention that this aim of validating the ratio of non-disease associated:disease-

associated isolates through quantitative bacteriological methods might not be possible as molecular techniques, although typically more sensitive, detect DNA and not necessarily only the DNA of viable bacteria, something that should at least be acknowledged in future diagnosis or surveillance developments. From a practical perspective, increasing the number of single colonies isolated for characterisation from solid agar plates is not a realistic option to improve wide scale surveillance. Indeed, the procedure for collection of palatine tonsils swabs from pigs is not straightforward and it may be interesting to investigate the potential use of alternative sampling sites, for example the nasal passage, anus or even housing environment. Another potential option is to sample oral fluid collected by swabs or a rope chew. Mentioned in chapter 4, it was shown to be possible to positively detect *S. suis* DNA from oral fluid using qPCR.

As *S. suis* is considered endemic in pig populations [17], it is interesting to speculate about the role non-disease associated isolates have on the development of invasive disease. From an evolutionary perspective crossing the blood brain barrier and causing meningitis is self-limiting as the infectious organism will kill the host and itself. It could be the case that lethal invasive disease caused by *S. suis* is accidental and a product of being in possession of a significantly smaller genome (albeit enriched in proposed virulence-associated factors) [141]. Previous analysis of the original training collection identified non-disease associated isolates to have a larger genome in comparison to isolates associated with invasive disease. Presented in chapter 6, the observation that the growth characteristics of non-disease associated isolate LSS089 revealed a much shorter stationary phase between five and 12-hours could indicate that isolates with a larger genome are able to move to a second carbon source more efficiently and as a result out compete invasive disease causing strains of *S. suis* in the upper respiratory tract. As a result, disease associated strains might be causing invasive disease inadvertently as a consequence of adapting to a less competitive environment than the upper respiratory tract.

A common theme in the literature for *S. suis* is the urgent need for an agreed set of criteria to define strains as virulent [302-304]. Worldwide, *S. suis* is one of the most important bacterial

swine pathogens to affect both traditional backyard and modern intensive farming practices; the incidence and awareness of which has been proposed to have increased with the intensification of pig production in the 1920s [11, 141]. As research groups continue to define virulence differently, the extrapolation of findings between groups is difficult [302, 425]. Indeed, in this thesis we used observed clinical phenotypes to define isolates as disease-associated or non-disease associated, a method that has been acknowledged not to be without caveats. Due to the advances in whole-genome sequencing and a marked increase in the amount of *S. suis* isolates being sequenced, the need for an agreed set of criteria or experimental infection model to screen potential virulence markers has never been more urgent. Work presented in chapter 6 confirmed the simple non-mammalian Greater Wax Moth larvae experimental infection model [423] can be used to detect differences in the virulence potential of invasive disease-associated and non-disease associated field isolates recovered from pigs on UK farms. However, before such an experimental infection model could be agreed upon as the standard model defining strains of *S. suis* as virulent, more thorough evaluation must be performed. Indeed, other simple experimental infection models also exist that are being explored for *S. suis*, such as a *Caenorhabditis elegans* model. Pilot studies I conducted for this model, not presented in this thesis, suggest that when grown on bacterial lawns of disease-associated or non-disease associated *S. suis* the differences in nematode development is very subtle. These early observations indicated that this model would therefore be unsuitable for worldwide use in laboratories without extensive experience of the stages of *C. elegans* development. At present, one rapidly advancing field of science is the development of organoids, (an in vitro 3D cellular cluster derived exclusively from primary tissue [599]) and it will be exciting to see if and how porcine organoid models contribute to improving the understanding of *S. suis* pathogenesis in the future.

In conclusion, the complexity of *S. suis* has allowed extensive discussion regarding the factors associated with pathogenesis and virulence. This thesis only describes a percentage of the current exploratory studies being undertaken worldwide to improve understanding of how this

both commensal and invasive organism can be better controlled by the swine industry. It will be exciting to see how future work on *S. suis* develops and if any of the outcomes of the work presented in this thesis can contribute to improving *S. suis* surveillance and/or diagnostics.

Appendix

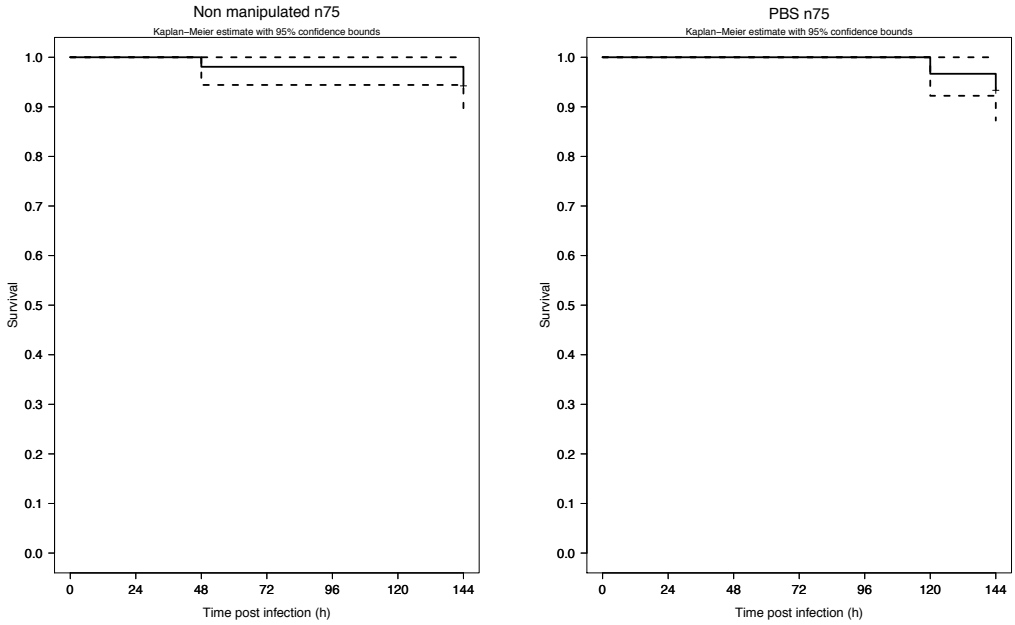

Figure 8.1. Kaplan-Meier curves of Greater Wax Moth larvae survival for negative control groups. Kaplan-Meier curves of Greater Wax Moth final instar stage larvae survival for negative control groups (a) no treatment control, included to assess background larval mortality and (b) PBS (20 µL) only control, included to assess background larval mortality due to the physical trauma of injection. Groups of 15 final instar stage caterpillars were used in each case.

Table 8.1. *Streptococcus suis* isolates of the original training collection with 100% matches to the protein-encoding sequence of the natural competence inducing peptide ComS.

The 66 bp protein-encoding sequence of the natural competence inducing peptide ComS (SSU0050) was used as a BLASTn query against a bespoke BLAST database consisting of the whole-genome sequences of the 115 isolates of the original training collection. Matches to 100% of the total length of the query sequence were considered hits deemed able to enter the transient physiological bacterial state of natural competence in the presence of exogenous ComS.

BLASTn query	Isolate	Phenotype	% identity	Alignment length	Evalue
SSU0050_ComS	LOLA-SS001.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	LOLA-SS002.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	LOLA-SS003.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	LOLA-SS005.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	LOLA-SS010.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	LOLA-SS011.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	P1/7.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1001.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1002.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1004.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1008.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1010.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1011.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1012.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1023.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1024.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1026.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1028.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1029.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1032.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1033.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1036.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1038.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1039.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1054.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1055.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1056.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1063.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1065.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1066.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1070.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1072.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS1073.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS970.gnm	Disease-associated	100	66	9.67E-28

SSU0050_ComS	SS972.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS973.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS976.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS979.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS982.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS983.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS987.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS991.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS994.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	SS995.gnm	Disease-associated	100	66	9.67E-28
SSU0050_ComS	LSS20.gnm	Non-disease associated	100	66	9.67E-28
SSU0050_ComS	LSS34.gnm	Non-disease associated	100	66	9.67E-28
SSU0050_ComS	LSS44.gnm	Non-disease associated	100	66	9.67E-28
SSU0050_ComS	LSS60.gnm	Non-disease associated	100	66	9.67E-28
SSU0050_ComS	LSS62.gnm	Non-disease associated	100	66	9.67E-28
SSU0050_ComS	LSS63.gnm	Non-disease associated	100	66	9.67E-28
SSU0050_ComS	LSS67.gnm	Non-disease associated	100	66	9.67E-28
SSU0050_ComS	LSS68.gnm	Non-disease associated	100	66	9.67E-28
SSU0050_ComS	LSS70.gnm	Non-disease associated	100	66	9.67E-28
SSU0050_ComS	LSS76.gnm	Non-disease associated	100	66	9.67E-28
SSU0050_ComS	LSS79.gnm	Non-disease associated	100	66	9.67E-28
SSU0050_ComS	LSS81.gnm	Non-disease associated	100	66	9.67E-28
SSU0050_ComS	LSS89.gnm	Non-disease associated	100	66	9.67E-28
SSU0050_ComS	LSS90.gnm	Non-disease associated	100	66	9.67E-28
SSU0050_ComS	LSS94.gnm	Non-disease associated	100	66	9.67E-28

References

[1] Staats JJ, Feder I, Okwumabua O, Chengappa MM. Streptococcus suis: past and present. Vet Res Commun. 1997;21:381-407.

[2] Elliott SD, Alexander TJ, Thomas JH. Streptococcal infection in young pigs. II. Epidemiology and experimental production of the disease. J Hyg (Lond). 1966;64:213-20.

[3] Windsor RS. Meningitis in pigs caused by Streptococcus suis type II. Vet Rec. 1977;101:378-9.

[4] Clifton-Hadley FA. Streptococcus suis type 2 infections. Br Vet J. 1983;139:1-5.

[5] Cook RW, Jackson AR, Ross AD. Streptococcus suis type 1 infection of sucking pigs. Aust Vet J. 1988;65:64-5.

[6] Higgins R, Gottschalk M, Mittal KR, Beaudoin M. Streptococcus suis infection in swine. A sixteen month study. Can J Vet Res. 1990;54:170-3.

[7] Hariharan H, MacDonald J, Carnat B, Bryenton J, Heaney S. An investigation of bacterial causes of arthritis in slaughter hogs. J Vet Diagn Invest. 1992;4:28-30.

[8] Chanter N, Jones PW, Alexander TJ. Meningitis in pigs caused by Streptococcus suis--a speculative review. Vet Microbiol. 1993;36:39-55.

[9] Reams RY, Glickman LT, Harrington DD, Thacker HL, Bowersock TL. Streptococcus suis infection in swine: a retrospective study of 256 cases. Part II. Clinical signs, gross and microscopic lesions, and coexisting microorganisms. J Vet Diagn Invest. 1994;6:326-34.

[10] Tarradas C, Arenas A, Maldonado A, Luque I, Miranda A, Perea A. Identification of Streptococcus suis isolated from swine: proposal for biochemical parameters. J Clin Microbiol. 1994;32:578-80.

[11] Gottschalk M. Streptococcosis. In: Zimmerman J, Karriker L, Ramirez A, Schwartz K, Stevenson G, editors. Diseases of Swine. Tenth Edition ed: John Wiley & Sons, Inc.; 2012. p. 841-55.

[12] Perch B, Kristjansen P, Skadhauge K. Group R streptococci pathogenic for man. Two cases of meningitis and one fatal case of sepsis. Acta Pathol Microbiol Scand. 1968;74:69-76.

[13] Suankratay C, Intalapaporn P, Nunthapisud P, Arunyingmongkol K, Wilde H. Streptococcus suis meningitis in Thailand. Southeast Asian J Trop Med Public Health. 2004;35:868-76.

[14] Hui AC, Ng KC, Tong PY, Mok V, Chow KM, Wu A, et al. Bacterial meningitis in Hong Kong: 10-years' experience. Clin Neurol Neurosurg. 2005;107:366-70.

[15] Mai NT, Hoa NT, Nga TV, Linh le D, Chau TT, Sinh DX, et al. Streptococcus suis meningitis in adults in Vietnam. Clin Infect Dis. 2008;46:659-67.

[16] Gottschalk M, Xu J, Calzas C, Segura M. Streptococcus suis: a new emerging or an old neglected zoonotic pathogen? Future Microbiol. 2010;5:371-91.

[17] Goyette-Desjardins G, Auger J-P, Xu J, Segura M, Gottschalk M. Streptococcus suis, an important pig pathogen and emerging zoonotic agent—an update on the worldwide distribution based on serotyping and sequence typing. Emerging Microbes & Infections. 2014;3:e45.

[18] Field HI, Buntain D, Done JT. Studies on piglet mortality. I. Streptococcal meningitis and arthritis. Vet Rec. 1954;66:453-5.

[19] de Moor CE. Septicaemic Infections in Pigs, Caused by Haemolytic Streptococci of New Lancefield Groups Designated R, S, and T. Antonie Van Leeuwenhoek. 1963;29:272-80.

[20] Kilpper-Balz R, Schleifer KH. Streptococcus suis sp. nov., nom. rev. International Journal of Systematic Bacteriology. 1987;37:160-2.

[21] Aarestrup FM, Rasmussen SR, Artursson K, Jensen NE. Trends in the resistance to antimicrobial agents of Streptococcus suis isolates from Denmark and Sweden. Vet Microbiol. 1998;63:71-80.

[22] Aarestrup FM, Jorsal SE, Jensen NE. Serological characterization and antimicrobial susceptibility of Streptococcus suis isolates from diagnostic samples in Denmark during 1995 and 1996. Vet Microbiol. 1998;60:59-66.

[23] Clifton-Hadley FA, Alexander TJ. The carrier site and carrier rate of Streptococcus suis type II in pigs. Vet Rec. 1980;107:40-1.

[24] Higgins R, Gottschalk M. An Update on Streptococcus Suis Identification. Journal of Veterinary Diagnostic Investigation. 1990;2:249-52.

[25] Madsen LW, Svensmark B, Elvestad K, Jensen HE. Otitis interna is a frequent sequela to Streptococcus suis meningitis in pigs. Vet Pathol. 2001;38:190-5.

[26] Devriese LA, Ceyssens K, Hommez J, Kilpper-Balz R, Schleifer KH. Characteristics of different Streptococcus suis ecovars and description of a simplified identification method. Vet Microbiol. 1991;26:141-50.

[27] Gottschalk M, Lacouture S, Odierno L. Immunomagnetic isolation of Streptococcus suis serotypes 2 and 1/2 from swine tonsils. J Clin Microbiol. 1999;37:2877-81.

[28] Perch B, Kjems E, Slot P, Pedersen KB. Biochemical and serological properties of R, S, and RS streptococci. Acta Pathol Microbiol Scand B. 1981;89:167-71.

[29] Vecht U, van Leengoed LA, Verheijen ER. Streptococcus suis infections in pigs in the Netherlands (Part I). Vet Q. 1985;7:315-21.

[30] Hommez J, Devriese LA, Henrichsen J, Castryck F. Identification and characterization of Streptococcus suis. Vet Microbiol. 1986;11:349-55.

[31] Elliott SD. Streptococcal infection in young pigs. I. An immunochemical study of the causative agent (PM streptococcus). J Hyg (Lond). 1966;64:205-12.

[32] Elliott SD, McCarty M, Lancefield RC. Teichoic acids of group D streptococci with special reference to strains from pig meningitis (Streptococcus suis). J Exp Med. 1977;145:490-9.

[33] Elliott SD, Tai JY. The type-specific polysaccharides of Streptococcus suis. J Exp Med. 1978;148:1699-704.

[34] Lancefield RC. A Serological Differentiation of Human and Other Groups of Hemolytic Streptococci. J Exp Med. 1933;57:571-95.

[35] Swift HF, Wilson AT, Lancefield RC. Typing Group a Hemolytic Streptococci by M Precipitin Reactions in Capillary Pipettes. J Exp Med. 1943;78:127-33.

[36] Joseph R, Shockman GD. Synthesis and excretion of glycerol teichoic acid during growth of two streptococcal species. Infect Immun. 1975;12:333-8.

[37] Windsor RS, Elliott SD. Streptococcal infection in young pigs. IV. An outbreak of streptococcal meningitis in weaned pigs. J Hyg (Lond). 1975;75:69-78.

[38] Gottschalk M, Higgins R, Jacques M, Mittal KR, Henrichsen J. Description of 14 new capsular types of Streptococcus suis. J Clin Microbiol. 1989;27:2633-6.

[39] Van Calsteren MR, Goyette-Desjardins G, Gagnon F, Okura M, Takamatsu D, Roy R, et al. Explaining the Serological Characteristics of Streptococcus suis Serotypes 1 and 1/2 from Their Capsular Polysaccharide Structure and Biosynthesis. J Biol Chem. 2016;291:8387-98.

[40] Gottschalk M, Higgins R, Boudreau M. Use of polyvalent coagglutination reagents for serotyping of Streptococcus suis. J Clin Microbiol. 1993;31:2192-4.

[41] Gottschalk M, Segura M. The pathogenesis of the meningitis caused by Streptococcus suis: the unresolved questions. Vet Microbiol. 2000;76:259-72.

[42] Perch B, Pedersen KB, Henrichsen J. Serology of capsulated streptococci pathogenic for pigs: six new serotypes of Streptococcus suis. J Clin Microbiol. 1983;17:993-6.

[43] Gottschalk M, Higgins R, Jacques M, Beaudoin M, Henrichsen J. Characterization of six new capsular types (23 through 28) of Streptococcus suis. J Clin Microbiol. 1991;29:2590-4.

[44] Higgins R, Gottschalk M, Boudreau M, Lebrun A, Henrichsen J. Description of Six New Capsular Types (29-34) of Streptococcus Suis. Journal of Veterinary Diagnostic Investigation. 1995;7:405-6.

[45] Chatellier S, Harel J, Zhang Y, Gottschalk M, Higgins R, Devriese LA, et al. Phylogenetic diversity of Streptococcus suis strains of various serotypes as revealed by 16S rRNA gene sequence comparison. Int J Syst Bacteriol. 1998;48 Pt 2:581-9.

[46] Brousseau R, Hill JE, Prefontaine G, Goh SH, Harel J, Hemmingsen SM. Streptococcus suis Serotypes Characterized by Analysis of Chaperonin 60 Gene Sequences. Applied and Environmental Microbiology. 2001;67:4828-33.

[47] Hill JE, Gottschalk M, Brousseau R, Harel J, Hemmingsen SM, Goh SH. Biochemical analysis, cpn60 and 16S rDNA sequence data indicate that Streptococcus suis serotypes 32 and 34, isolated from pigs, are Streptococcus orisratti. Vet Microbiol. 2005;107:63-9.

[48] Tien le HT, Nishibori T, Nishitani Y, Nomoto R, Osawa R. Reappraisal of the taxonomy of Streptococcus suis serotypes 20, 22, 26, and 33 based on DNA-DNA homology and sodA and recN phylogenies. Vet Microbiol. 2013;162:842-9.

[49] Okura M, Osaki M, Nomoto R, Arai S, Osawa R, Sekizaki T, et al. Current Taxonomical Situation of Streptococcus suis. Pathogens. 2016;5.

[50] Kerdsin A, Akeda Y, Hatrongjit R, Detchawna U, Sekizaki T, Hamada S, et al. Streptococcus suis serotyping by a new multiplex PCR. J Med Microbiol. 2014;63:824-30.

[51] Wei Z, Li R, Zhang A, He H, Hua Y, Xia J, et al. Characterization of Streptococcus suis isolates from the diseased pigs in China between 2003 and 2007. Vet Microbiol. 2009;137:196-201.

[52] Vecht U, Arends JP, Vandermolen EJ, Vanleengoed LAMG. Differences in Virulence between 2 Strains of Streptococcus-Suis Type-II after Experimentally Induced Infection of Newborn Germ-Free Pigs. American Journal of Veterinary Research. 1989;50:1037-43.

[53] Vecht U, Wisselink HJ, van Dijk JE, Smith HE. Virulence of Streptococcus suis type 2 strains in newborn germfree pigs depends on phenotype. Infect Immun. 1992;60:550-6.

[54] Kataoka Y, Haritani M, Mori M, Kishima M, Sugimoto C, Nakazawa M, et al. Experimental infections of mice and pigs with Streptococcus suis type 2. J Vet Med Sci. 1991;53:1043-9.

[55] Beaudoin M, Higgins R, Harel J, Gottschalk M. Studies on a murine model for evaluation of virulence of Streptococcus suis capsular type 2 isolates. FEMS Microbiol Lett. 1992;78:111-6.

[56] Galina L, Pijoan C, Sitjar M, Christianson WT, Rossow K, Collins JE. Interaction between Streptococcus suis serotype 2 and porcine reproductive and respiratory syndrome virus in specific pathogen-free piglets. Vet Rec. 1994;134:60-4.

[57] Vela AI, Goyache J, Tarradas C, Luque I, Mateos A, Moreno MA, et al. Analysis of genetic diversity of Streptococcus suis clinical isolates from pigs in Spain by pulsed-field gel electrophoresis. J Clin Microbiol. 2003;41:2498-502.

[58] Tarradas C, Perea A, Vela AI, Goyache J, Dominguez L, Fernandez-Garaizabal JF, et al. Distribution of serotypes of Streptococcus suis isolated from diseased pigs in Spain. Vet Rec. 2004;154:665-6.

[59] Luque I, Blume V, Borge C, Vela AI, Perea JA, Marquez JM, et al. Genetic analysis of Streptococcus suis isolates recovered from diseased and healthy carrier pigs at different stages of production on a pig farm. Vet J. 2010;186:396-8.

[60] Heath PJ, Hunt BW. Streptococcus suis serotypes 3 to 28 associated with disease in pigs. Veterinary Record. 2001;148:207-8.

[61] Heath PJ, Hunt BW, Duff JP, Wilkinson JD. Streptococcus suis serotype 14 as a cause of pig disease in the UK. Vet Rec. 1996;139:450-1.

[62] Higgins R, Gottschalk M. Distribution of Streptococcus suis capsular types in 1998. Can Vet J. 1999;40:277.

[63] Pedersen KB, Kjems E, Perch B, Slot P. Infection with RS streptococci in pigs. Acta Pathol Microbiol Scand B. 1981;89:161-5.

[64] Lakkitjaroen N, Takamatsu D, Okura M, Sato M, Osaki M, Sekizaki T. Loss of capsule among Streptococcus suis isolates from porcine endocarditis and its biological significance. J Med Microbiol. 2011;60:1669-76.

[65] MacLennan M, Foster G, Dick K, Smith WJ, Nielsen B. Streptococcus suis serotypes 7, 8 and 14 from diseased pigs in Scotland. Vet Rec. 1996;139:423-4.

[66] Smith HE, van Bruijnsvoort L, Buijs H, Wisselink HJ, Smits MA. Rapid PCR test for Streptococcus suis serotype 7. FEMS Microbiol Lett. 1999;178:265-70.

[67] Smith HE, Veenbergen V, van der Velde J, Damman M, Wisselink HJ, Smits MA. The cps genes of Streptococcus suis serotypes 1, 2, and 9: development of rapid serotype-specific PCR assays. J Clin Microbiol. 1999;37:3146-52.

[68] Wisselink HJ, Joosten JJ, Smith HE. Multiplex PCR Assays for Simultaneous Detection of Six Major Serotypes and Two Virulence-Associated Phenotypes of Streptococcus suis in Tonsillar Specimens from Pigs. Journal of Clinical Microbiology. 2002;40:2922-9.

[69] Okwumabua O, O'Connor M, Shull E. A polymerase chain reaction (PCR) assay specific for Streptococcus suis based on the gene encoding the glutamate dehydrogenase. FEMS Microbiol Lett. 2003;218:79-84.

[70] Marois C, Bougeard S, Gottschalk M, Kobisch M. Multiplex PCR assay for detection of Streptococcus suis species and serotypes 2 and 1/2 in tonsils of live and dead pigs. J Clin Microbiol. 2004;42:3169-75.

[71] Wang K, Fan W, Wisselink H, Lu C. The cps locus of Streptococcus suis serotype 16: development of a serotype-specific PCR assay. Vet Microbiol. 2011;153:403-6.

[72] Wang K, Sun X, Lu C. Development of rapid serotype-specific PCR assays for eight serotypes of Streptococcus suis. J Clin Microbiol. 2012;50:3329-34.

[73] Kerdsin A, Dejsirilert S, Akeda Y, Sekizaki T, Hamada S, Gottschalk M, et al. Fifteen Streptococcus suis serotypes identified by multiplex PCR. J Med Microbiol. 2012;61:1669-72.

[74] Liu Z, Zheng H, Gottschalk M, Bai X, Lan R, Ji S, et al. Development of multiplex PCR assays for the identification of the 33 serotypes of Streptococcus suis. PLoS One. 2013;8:e72070.

[75] Okura M, Lachance C, Osaki M, Sekizaki T, Maruyama F, Nozawa T, et al. Development of a two-step multiplex PCR assay for typing of capsular polysaccharide synthesis gene clusters of Streptococcus suis. J Clin Microbiol. 2014;52:1714-9.

[76] Smith HE, Damman M, van der Velde J, Wagenaar F, Wisselink HJ, Stockhofe-Zurwieden N, et al. Identification and characterization of the cps locus of Streptococcus suis serotype 2: the capsule protects against phagocytosis and is an important virulence factor. Infection and Immunity. 1999;67:1750-6.

[77] Smith HE, de Vries R, van't Slot R, Smits MA. The cps locus of Streptococcus suis serotype 2: genetic determinant for the synthesis of sialic acid. Microb Pathog. 2000;29:127-34.

[78] Okura M, Takamatsu D, Maruyama F, Nozawa T, Nakagawa I, Osaki M, et al. Genetic analysis of capsular polysaccharide synthesis gene clusters from all serotypes of Streptococcus suis: potential mechanisms for generation of capsular variation. Appl Environ Microbiol. 2013;79:2796-806.

[79] Bonifait L, Gottschalk M, Grenier D. Cell surface characteristics of nontypeable isolates of Streptococcus suis. FEMS Microbiol Lett. 2010;311:160-6.

[80] Gottschalk M, Lacouture S, Bonifait L, Roy D, Fittipaldi N, Grenier D. Characterization of Streptococcus suis isolates recovered between 2008 and 2011 from diseased pigs in Quebec, Canada. Vet Microbiol. 2013;162:819-25.

[81] Pan Z, Ma J, Dong W, Song W, Wang K, Lu C, et al. Novel variant serotype of streptococcus suis isolated from piglets with meningitis. Appl Environ Microbiol. 2015;81:976-85.

[82] Qiu X, Bai X, Lan R, Zheng H, Xu J. Novel Capsular Polysaccharide Loci and New Diagnostic Tools for High-Throughput Capsular Gene Typing in Streptococcus suis. Appl Environ Microbiol. 2016;82:7102-12.

[83] Zheng H, Ji S, Liu Z, Lan R, Huang Y, Bai X, et al. Eight Novel Capsular Polysaccharide Synthesis Gene Loci Identified in Nontypeable Streptococcus suis Isolates. Appl Environ Microbiol. 2015;81:4111-9.

[84] Athey TB, Teatero S, Lacouture S, Takamatsu D, Gottschalk M, Fittipaldi N. Determining Streptococcus suis serotype from short-read whole-genome sequencing data. BMC Microbiol. 2016;16:162.

[85] Marois C, Le Devendec L, Gottschalk M, Kobisch M. Detection and molecular typing of Streptococcus suis in tonsils from live pigs in France. Can J Vet Res. 2007;71:14-22.

[86] Marois C, Le Devendec L, Gottschalk M, Kobisch M. Molecular characterization of Streptococcus suis strains by 16S-23S intergenic spacer polymerase chain reaction and restriction fragment length polymorphism analysis. Can J Vet Res. 2006;70:94-104.

[87] Rehm T, Baums CG, Strommenger B, Beyerbach M, Valentin-Weigand P, Goethe R. Amplified fragment length polymorphism of Streptococcus suis strains correlates with their profile of virulence-associated genes and clinical background. J Med Microbiol. 2007;56:102-9.

[88] Chatellier SG, M. Higgins, R. Brousseau, R. and Harel, J. Relatedness of Streptococcus suis Serotype 2 Isolates from Different Geographic Origins as Evaluated by Molecular Fingerprinting and Phenotyping. JOURNAL OF CLINICAL MICROBIOLOGY. 1999;37:362-6.

[89] Martinez G, Harel J, Lacouture S, Gottschalk M. Genetic diversity of Streptococcus suis serotypes 2 and 1/2 isolates recovered from carrier pigs in closed herds. Can J Vet Res. 2002;66:240-8.

[90] King SJ, Leigh JA, Heath PJ, Luque I, Tarradas C, Dowson CG, et al. Development of a Multilocus Sequence Typing Scheme for the Pig Pathogen Streptococcus suis: Identification of Virulent Clones and Potential Capsular Serotype Exchange. Journal of Clinical Microbiology. 2002;40:3671-80.

[91] Allgaier A, Goethe R, Wisselink HJ, Smith HE, Valentin-Weigand P. Relatedness of Streptococcus suis isolates of various serotypes and clinical backgrounds as evaluated by macrorestriction analysis and expression of potential virulence traits. Journal of Clinical Microbiology. 2001;39:445-53.

[92] Berthelot-Herault F, Marois C, Gottschalk M, Kobisch M. Genetic diversity of Streptococcus suis strains isolated from pigs and humans as revealed by pulsed-field gel electrophoresis. J Clin Microbiol. 2002;40:615-9.

[93] Tian Y, Aarestrup FM, Lu CP. Characterization of Streptococcus suis serotype 7 isolates from diseased pigs in Denmark. Vet Microbiol. 2004;103:55-62.

[94] Luey CK, Chu YW, Cheung TK, Law CC, Chu MY, Cheung DT, et al. Rapid pulsed-field gel electrophoresis protocol for subtyping of Streptococcus suis serotype 2. J Microbiol Methods. 2007;68:648-50.

[95] Torremorell M, Pijoan C. Prolonged persistence of an epidemic Streptococcus suis strain in a closed pig population. Vet Rec. 1998;143:394-5.

[96] Torremorell M, Calsamiglia M, Pijoan C. Colonization of suckling pigs by Streptococcus suis with particular reference to pathogenic serotype 2 strains. Can J Vet Res. 1998;62:21-6.

[97] Oliveira S, Batista L, Torremorell M, Pijoan C. Experimental colonization of piglets and gilts with systemic strains of Haemophilus parasuis and Streptococcus suis to prevent disease. Can J Vet Res. 2001;65:161-7.

[98] Amass SF, SanMiguel P, Clark LK. Demonstration of vertical transmission of Streptococcus suis in swine by genomic fingerprinting. J Clin Microbiol. 1997;35:1595-6.

[99] Mogollon JD, Pijoan C, Murtaugh MP, Collins JE, Cleary PP. Identification of epidemic strains of Streptococcus suis by genomic fingerprinting. J Clin Microbiol. 1991;29:782-7.

[100] Mogollon JD, Pijoan C, Murtaugh MP, Kaplan EL, Collins JE, Cleary PP. Characterization of prototype and clinically defined strains of Streptococcus suis by genomic fingerprinting. J Clin Microbiol. 1990;28:2462-6.

[101] Beaudoin M, Harel J, Higgins R, Gottschalk M, Frenette M, MacInnes JI. Molecular analysis of isolates of Streptococcus suis capsular type 2 by restriction-endonuclease-digested DNA

separated on SDS-PAGE and by hybridization with an rDNA probe. J Gen Microbiol. 1992;138:2639-45.

[102] Okwumabua O, Staats J, Chengappa MM. Detection of Genomic Heterogeneity in Streptococcus-Suis Isolates by DNA Restriction-Fragment-Length-Polymorphisms of Ribosomal-Rna Genes (Ribotyping). Journal of Clinical Microbiology. 1995;33:968-72.

[103] Smith HE, Rijnsburger M, StockhofeZurwieden N, Wisselink HJ, Vecht U, Smits MA. Virulent strains of Streptococcus suis serotype 2 and highly virulent strains of Streptococcus suis serotype 1 can be recognized by a unique ribotype profile. Journal of Clinical Microbiology. 1997;35:1049-53.

[104] Rasmussen SR, Andresen LO. 16S rDNA sequence variations of some Streptococcus suis serotypes. Int J Syst Bacteriol. 1998;48 Pt 3:1063-5.

[105] Rasmussen SR, Aarestrup FM, Jensen NE, Jorsal SE. Associations of Streptococcus suis serotype 2 ribotype profiles with clinical disease and antimicrobial resistance. Journal of Clinical Microbiology. 1999;37:404-8.

[106] Staats JJ, Plattner BL, Nietfeld J, Dritz S, Chengappa MM. Use of ribotyping and hemolysin activity to identify highly virulent Streptococcus suis type 2 isolates. J Clin Microbiol. 1998;36:15-9.

[107] Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A. 1998;95:3140-5.

[108] Enright MC, Spratt BG. A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. Microbiology. 1998;144 (Pt 11):3049-60.

[109] Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of Staphylococcus aureus. J Clin Microbiol. 2000;38:1008-15.

[110] Enright MC, Spratt BG, Kalia A, Cross JH, Bessen DE. Multilocus sequence typing of Streptococcus pyogenes and the relationships between emm type and clone. Infect Immun. 2001;69:2416-27.

[111] Maiden MC. Multilocus sequence typing of bacteria. Annu Rev Microbiol. 2006;60:561-88.

[112] Coffey TJ, Enright MC, Daniels M, Morona JK, Morona R, Hryniewicz W, et al. Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of Streptococcus pneumoniae. Molecular Microbiology. 1998;27:73-83.

[113] Whatmore AM, Kapur V, Sullivan DJ, Musser JM, Kehoe MA. Non-Congruent Relationships between Variation in Emm Gene-Sequences and the Population Genetic-Structure of Group-a Streptococci. Molecular Microbiology. 1994;14:619-31.

[114] Takamatsu D, Wongsawan K, Osaki M, Nishino H, Ishiji T, Tharavichitkul P, et al. Streptococcus suis in humans, Thailand. Emerg Infect Dis. 2008;14:181-3.

[115] Chen C, Zhang W, Zheng H, Lan R, Wang H, Du P, et al. Minimum core genome sequence typing of bacterial pathogens: a unified approach for clinical and public health microbiology. J Clin Microbiol. 2013;51:2582-91.

[116] Zheng H, Ji S, Lan R, Liu Z, Bai X, Zhang W, et al. Population analysis of Streptococcus suis isolates from slaughtered swine by use of minimum core genome sequence typing. J Clin Microbiol. 2014;52:3568-72.

[117] Du P, Zheng H, Zhou J, Lan R, Ye C, Jing H, et al. Detection of Multiple Parallel Transmission Outbreak of Streptococcus suis Human Infection by Use of Genome Epidemiology, China, 2005. Emerg Infect Dis. 2017;23:204-11.

[118] Qin T, Zhang W, Liu W, Zhou H, Ren H, Shao Z, et al. Population structure and minimum core genome typing of Legionella pneumophila. Sci Rep. 2016;6:21356.

[119] Reams RY, Glickman LT, Harrington DD, Bowersock TL, Thacker HL. Streptococcus suis infection in swine: a retrospective study of 256 cases. Part I. Epidemiologic factors and antibiotic susceptibility patterns. J Vet Diagn Invest. 1993;5:363-7.

[120] Galina L, Collins JE, Pijoan C. Porcine Streptococcus suis in Minnesota. J Vet Diagn Invest. 1992;4:195-6.

[121] Reams RY, Harrington DD, Glickman LT, Thacker HL, Bowersock TL. Multiple serotypes and strains of Streptococcus suis in naturally infected swine herds. J Vet Diagn Invest. 1996;8:119-21.

[122] Feng W, Laster SM, Tompkins M, Brown T, Xu JS, Altier C, et al. In utero infection by porcine reproductive and respiratory syndrome virus is sufficient to increase susceptibility of piglets to challenge by Streptococcus suis type II. J Virol. 2001;75:4889-95.

[123] Schmitt CS, Halbur PG, Roth JA, Kinyon JM, Kasorndorkbua C, Thacker B. Influence of ampicillin, ceftiofur, attenuated live PRRSV vaccine, and reduced dose Streptococcus suis exposure on disease associated with PRRSV and S. suis coinfection. Vet Microbiol. 2001;78:29-37.

[124] Pallares FJ, Halbur PG, Opriessnig T, Sorden SD, Villar D, Janke BH, et al. Porcine circovirus type 2 (PCV-2) coinfections in US field cases of postweaning multisystemic wasting syndrome (PMWS). J Vet Diagn Invest. 2002;14:515-9.

[125] Robertson ID, Blackmore DK, Hampson DJ, Fu ZF. A longitudinal study of natural infection of piglets with Streptococcus suis types 1 and 2. Epidemiol Infect. 1991;107:119-26.

[126] Cloutier G, D'Allaire S, Martinez G, Surprenant C, Lacouture S, Gottschalk M. Epidemiology of Streptococcus suis serotype 5 infection in a pig herd with and without clinical disease. Vet Microbiol. 2003;97:135-51.

[127] Clifton-Hadley FA, Enright MR. Factors affecting the survival of Streptococcus suis type 2. Vet Rec. 1984;114:584-6.

[128] Clifton-Hadley FA, Alexander TJ, Enright MR. Monitoring herds for Streptococcus suis type 2: chance contamination of slaughter pigs. Vet Rec. 1986;118:274.

[129] Fittipaldi N, Segura M, Grenier D, Gottschalk M. Virulence factors involved in the pathogenesis of the infection caused by the swine pathogen and zoonotic agent Streptococcus suis. Future Microbiol. 2012;7:259-79.

[130] Madsen LW, Bak H, Nielsen B, Jensen HE, Aalbaek B, Riising HJ. Bacterial colonization and invasion in pigs experimentally exposed to Streptococcus suis serotype 2 in aerosol. J Vet Med B Infect Dis Vet Public Health. 2002;49:211-5.

[131] Baums CG, Valentin-Weigand P. Surface-associated and secreted factors of Streptococcus suis in epidemiology, pathogenesis and vaccine development. Anim Health Res Rev. 2009;10:65-83.

[132] Norton PM, Rolph C, Ward PN, Bentley RW, Leigh JA. Epithelial invasion and cell lysis by virulent strains of Streptococcus suis is enhanced by the presence of suilysin. FEMS Immunol Med Microbiol. 1999;26:25-35.

[133] Lalonde M, Segura M, Lacouture S, Gottschalk M. Interactions between Streptococcus suis serotype 2 and different epithelial cell lines. Microbiology. 2000;146 (Pt 8):1913-21.

[134] Benga L, Goethe R, Rohde M, Valentin-Weigand P. Non-encapsulated strains reveal novel insights in invasion and survival of Streptococcus suis in epithelial cells. Cell Microbiol. 2004;6:867-81.

[135] Valentin-Weigand P. Intracellular invasion and persistence: survival strategies of Streptococcus suis and Mycobacterium avium ssp. paratuberculosis. Berl Munch Tierarztl Wochenschr. 2004;117:459-63.

[136] Williams AE, Blakemore WF. Pathogenesis of meningitis caused by Streptococcus suis type 2. J Infect Dis. 1990;162:474-81.

[137] Williams AE, Blakemore WF. Monocyte-mediated entry of pathogens into the central nervous system. Neuropathol Appl Neurobiol. 1990;16:377-92.

[138] Segura M, Gottschalk M. Streptococcus suis interactions with the murine macrophage cell line J774: Adhesion and cytotoxicity. Infection and Immunity. 2002;70:4312-22.

[139] Holt ME, Enright MR, Alexander TJ. Immunisation of pigs with live cultures of Streptococcus suis type 2. Res Vet Sci. 1988;45:349-52.

[140] Blouin C, Higgins R, Gottschalk M, Simard J. Evaluation of the antibody response in pigs vaccinated against Streptococcus suis capsular type 2 using a double-antibody sandwich enzyme-linked immunosorbent assay. Can J Vet Res. 1994;58:49-54.

[141] Weinert LA, Chaudhuri RR, Wang J, Peters SE, Corander J, Jombart T, et al. Genomic signatures of human and animal disease in the zoonotic pathogen Streptococcus suis. Nat Commun. 2015;6:6740.

[142] Clifton-Hadley FA, Alexander TJ, Enright MR, Guise J. Monitoring herds for Streptococcus suis type 2 by sampling tonsils of slaughter pigs. Vet Rec. 1984;115:562-4.

[143] Dee SA, Carlson AR, Winkelman NL, Corey MM. Effect of Management-Practices on the Streptococcus-Suis Carrier Rate in Nursery Swine. Journal of the American Veterinary Medical Association. 1993;203:295-9.

[144] Robertson ID, Blackmore DK. Occupational exposure to Streptococcus suis type 2. Epidemiol Infect. 1989;103:157-64.

[145] Tarradas MC, Arenas A, Maldonado A, Vicente S, Miranda A, Perea A. Susceptibility of Streptococcus-Suis to Various Antimicrobial Agents. Journal of Veterinary Medicine Series B-Zentralblatt Fur Veterinarmedizin Reihe B-Infectious Diseases and Veterinary Public Health. 1994;41:685-8.

[146] Martel A, Baele M, Devriese LA, Goossens H, Wisselink HJ, Decostere A, et al. Prevalence and mechanism of resistance against macrolides and lincosamides in Streptococcus suis isolates. Vet Microbiol. 2001;83:287-97.

[147] Marie J, Morvan H, Berthelot-Herault F, Sanders P, Kempf I, Gautier-Bouchardon AV, et al. Antimicrobial susceptibility of Streptococcus suis isolated from swine in France and from humans in different countries between 1996 and 2000. J Antimicrob Chemother. 2002;50:201-9.

[148] Vela AI, Moreno MA, Cebolla JA, Gonzalez S, Latre MV, Dominguez L, et al. Antimicrobial susceptibility of clinical strains of Streptococcus suis isolated from pigs in Spain. Vet Microbiol. 2005;105:143-7.

[149] Wisselink HJ, Veldman KT, Van den Eede C, Salmon SA, Mevius DJ. Quantitative susceptibility of Streptococcus suis strains isolated from diseased pigs in seven European countries to antimicrobial agents licensed in veterinary medicine. Vet Microbiol. 2006;113:73-82.

[150] Chen L, Song Y, Wei Z, He H, Zhang A, Jin M. Antimicrobial Susceptibility, Tetracycline and Erythromycin Resistance Genes, and Multilocus Sequence Typing of Streptococcus suis Isolates from Diseased Pigs in China. Journal of Veterinary Medical Science. 2013;75:583-7.

[151] Glass-Kaastra SK, Pearl DL, Reid-Smith R, McEwen B, Slavic D, Fairles J, et al. Multiple-class antimicrobial resistance surveillance in swine Escherichia coli F4, Pasteurella multocida and

Streptococcus suis isolates from Ontario and the impact of the 2004-2006 Porcine Circovirus type-2 Associated Disease outbreak. Prev Vet Med. 2014;113:159-64.

[152] Hernandez-Garcia J, Wang J, Restif O, Holmes MA, Mather AE, Weinert LA, et al. Patterns of antimicrobial resistance in Streptococcus suis isolates from pigs with or without streptococcal disease in England between 2009 and 2014. Vet Microbiol. 2017;207:117-24.

[153] Prieto C, Garcia FJ, Suarez P, Imaz M, Castro JM. Biochemical traits and antimicrobial susceptibility of Streptococcus suis isolated from slaughtered pigs. Zentralbl Veterinarmed B. 1994;41:608-17.

[154] Zhang C, Ning Y, Zhang Z, Song L, Qiu H, Gao H. In vitro antimicrobial susceptibility of Streptococcus suis strains isolated from clinically healthy sows in China. Vet Microbiol. 2008;131:386-92.

[155] Callens BF, Haesebrouck F, Maes D, Butaye P, Dewulf J, Boyen F. Clinical resistance and decreased susceptibility in Streptococcus suis isolates from clinically healthy fattening pigs. Microb Drug Resist. 2013;19:146-51.

[156] Salmon SA, Watts JL, Case CA, Hoffman LJ, Wegener HC, Yancey RJ, Jr. Comparison of MICs of ceftiofur and other antimicrobial agents against bacterial pathogens of swine from the United States, Canada, and Denmark. J Clin Microbiol. 1995;33:2435-44.

[157] Priebe S, Schwarz S. In vitro activities of florfenicol against bovine and porcine respiratory tract pathogens. Antimicrob Agents Chemother. 2003;47:2703-5.

[158] Woo J, Li EK. Streptococcus suis meningitis requires prolonged treatment with penicillin. Infection. 1987;15:129-30.

[159] Gottschalk M, Turgeon P, Higgins R, Beaudoin M, Bourgault AM. Susceptibility of Streptococcus Suis to Penicillin. Journal of Veterinary Diagnostic Investigation. 1991;3:170-2.

[160] Escudero JA, San Millan A, Catalan A, de la Campa AG, Rivero E, Lopez G, et al. First characterization of fluoroquinolone resistance in Streptococcus suis. Antimicrob Agents Chemother. 2007;51:777-82.

[161] APHA. GB Emerging Threats Quarterly Report Pig Diseases. Quarterly Report January to March 2015. 2015.

[162] Gottschalk M, Segura M, Xu J. Streptococcus suis infections in humans: the Chinese experience and the situation in North America. Anim Health Res Rev. 2007;8:29-45.

[163] Wisselink HJ, Stockhofe-Zurwieden N, Hilgers LA, Smith HE. Assessment of protective efficacy of live and killed vaccines based on a non-encapsulated mutant of Streptococcus suis serotype 2. Vet Microbiol. 2002;84:155-68.

[164] Segura M. Streptococcus suis vaccines: candidate antigens and progress. Expert Rev Vaccines. 2015;14:1587-608.

[165] Amass SF, Stevenson GW, Knox K, Reed A. Efficacy of an autogenous vaccine for preventing streptococcosis in piglets. Vet Med. 1999;94:480-4.

[166] Lapointe L, D'Allaire S, Lebrun A, Lacouture S, Gottschalk M. Antibody response to an autogenous vaccine and serologic profile for Streptococcus suis capsular type 1/2. Can J Vet Res. 2002;66:8-14.

[167] Baums CG, Bruggemann C, Kock C, Beineke A, Waldmann KH, Valentin-Weigand P. Immunogenicity of an autogenous Streptococcus suis bacterin in preparturient sows and their piglets in relation to protection after weaning. Clin Vaccine Immunol. 2010;17:1589-97.

[168] Chase CC. Autogenous vaccines: current use in the field in the U.S. cattle and hog industry. Dev Biol (Basel). 2004;117:69-71.

[169] Haesebrouck F, Pasmans F, Chiers K, Maes D, Ducatelle R, Decostere A. Efficacy of vaccines against bacterial diseases in swine: what can we expect? Vet Microbiol. 2004;100:255-68.

[170] Holt ME, Enright MR, Alexander TJ. Immunisation of pigs with killed cultures of Streptococcus suis type 2. Res Vet Sci. 1990;48:23-7.

[171] Torremorell M, Pijoan C, Trigo E. Vaccination against Streptococcus suis: Effect on nursery mortality. Swine Health and Production. 1997;5:139-43.

[172] Wisselink HJ, Vecht U, Stockhofe-Zurwieden N, Smith HE. Protection of pigs against challenge with virulent Streptococcus suis serotype 2 strains by a muramidase-released protein and extracellular factor vaccine. Vet Rec. 2001;148:473-7.

[173] Buttner N, Beineke A, de Buhr N, Lilienthal S, Merkel J, Waldmann KH, et al. Streptococcus suis serotype 9 bacterin immunogenicity and protective efficacy. Vet Immunol Immunopathol. 2012;146:191-200.

[174] Kebede M, Chengappa MM, Stuart JG. Isolation and characterization of temperature-sensitive mutants of Streptococcus suis: efficacy trial of the mutant vaccine in mice. Vet Microbiol. 1990;22:249-57.

[175] Foster N, Staats JJ, Chengappa MM. Isolation, characterization and protection studies in mice of a streptomycin-dependent mutant of Streptococcus suis type 1/2. Vet Res Commun. 1994;18:155-63.

[176] Jacobs AA, van den Berg AJ, Loeffen PL. Protection of experimentally infected pigs by suilysin, the thiol-activated haemolysin of Streptococcus suis. Vet Rec. 1996;139:225-8.

[177] Segers RP, Kenter T, de Haan LA, Jacobs AA. Characterisation of the gene encoding suilysin from Streptococcus suis and expression in field strains. FEMS Microbiol Lett. 1998;167:255-61.

[178] Vecht U, Wisselink HJ, Jellema ML, Smith HE. Identification of Two Proteins Associated with Virulence of Streptococcus suis Type 2. Infection and Immunity. 1991;59:3156-62.

[179] Okwumabua O, Chinnapapakkagari S. Identification of the gene encoding a 38-kilodalton immunogenic and protective antigen of Streptococcus suis. Clin Diagn Lab Immunol. 2005;12:484-90.

[180] Liu L, Cheng G, Wang C, Pan X, Cong Y, Pan Q, et al. Identification and experimental verification of protective antigens against Streptococcus suis serotype 2 based on genome sequence analysis. Curr Microbiol. 2009;58:11-7.

[181] Zhang A, Chen B, Mu X, Li R, Zheng P, Zhao Y, et al. Identification and characterization of a novel protective antigen, Enolase of Streptococcus suis serotype 2. Vaccine. 2009;27:1348-53.

[182] Dee SA, Corey MM. The survival of Streptococcus suis on farm and veterinary equipment. J Swine Health Prod. 1993;1:17-20.

[183] Tang JQ, Wang CJ, Feng YJ, Yang WZ, Song HD, Chen ZH, et al. Streptococcal toxic shock syndrome caused by Streptococcus suis serotype 2. Plos Medicine. 2006;3:668-76.

[184] Huong VT, Ha N, Huy NT, Horby P, Nghia HD, Thiem VD, et al. Epidemiology, clinical manifestations, and outcomes of Streptococcus suis infection in humans. Emerg Infect Dis. 2014;20:1105-14.

[185] Gottschalk M. Porcine *Streptococcus suis* strains as potential sources of infections in humans: an underdiagnosed problem in North America? Journal of Swine Health and Production. 2004;12:197-9.

[186] Shneerson JM, Chattopadhyay B, Murphy MF, Fawcett IW. Permanent perceptive deafness due to Streptococcus suis type II infection. J Laryngol Otol. 1980;94:425-7.

[187] Lutticken R, Temme N, Hahn G, Bartelheimer EW. Meningitis caused by Streptococcus suis: case report and review of the literature. Infection. 1986;14:181-5.

[188] Bohr V, Hansen B, Kjersem H, Rasmussen N, Johnsen N, Kristensen HS, et al. Sequelae from bacterial meningitis and their relation to the clinical condition during acute illness, based on 667 questionnaire returns. Part II of a three part series. J Infect. 1983;7:102-10.

[189] Dodge PR, Davis H, Feigin RD, Holmes SJ, Kaplan SL, Jubelirer DP, et al. Prospective evaluation of hearing impairment as a sequela of acute bacterial meningitis. N Engl J Med. 1984;311:869-74.

[190] Arends JP, Zanen HC. Meningitis caused by Streptococcus suis in humans. Rev Infect Dis. 1988;10:131-7.

[191] Brown EJ. The molecular basis of streptococcal toxic shock syndrome. N Engl J Med. 2004;350:2093-4.

[192] Lappin E, Ferguson AJ. Gram-positive toxic shock syndromes. Lancet Infect Dis. 2009;9:281-90.

[193] Cone LA, Woodard DR, Schlievert PM, Tomory GS. Clinical and bacteriologic observations of a toxic shock-like syndrome due to Streptococcus pyogenes. N Engl J Med. 1987;317:146-9.

[194] Shands KN, Schmid GP, Dan BB, Blum D, Guidotti RJ, Hargrett NT, et al. Toxic-shock syndrome in menstruating women: association with tampon use and Staphylococcus aureus and clinical features in 52 cases. N Engl J Med. 1980;303:1436-42.

[195] Cohen ML, Falkow S. Protein antigens from Staphylococcus aureus strains associated with toxic-shock syndrome. Science. 1981;211:842-4.

[196] Spiss HK, Kofler M, Hausdorfer H, Pfausler B, Schmutzhard E. [Streptococcus suis meningitis and neurophysiology of the acoustic system. First case report from Austria]. Nervenarzt. 1999;70:738-41.

[197] Colaert J, Allewaert M, Magerman H, Vandeven J, Vandepitte J. Streptococcus suis meningitis in man. First reported observation in Belgium. Acta Clin Belg. 1985;40:314-7.

[198] Kopic J, Paradzik MT, Pandak N. Streptococcus suis infection as a cause of severe illness: 2 cases from Croatia. Scand J Infect Dis. 2002;34:683-4.

[199] Bonmarchand G, Massari P, Humbert G, Leroy J, Morel A, Lemeland JF, et al. Group R streptococci: wild boars as a second reservoir. Scand J Infect Dis. 1985;17:121-2.

[200] Dupas D, Vignon M, Geraut C. Streptococcus suis meningitis. A severe noncompensated occupational disease. J Occup Med. 1992;34:1102-5.

[201] Tayoro J, Besnier JM, Laudat P, Cattier B, Choutet P. Infective endocarditis due to Streptococcus suis serotype 2. Eur J Clin Microbiol Infect Dis. 1996;15:765-6.

[202] Francois B, Gissot V, Ploy MC, Vignon P. Recurrent septic shock due to Streptococcus suis. J Clin Microbiol. 1998;36:2395.

[203] Durand F, Perino CL, Recule C, Brion JP, Kobish M, Guerber F, et al. Bacteriological diagnosis of Streptococcus suis meningitis. Eur J Clin Microbiol Infect Dis. 2001;20:519-21.

[204] Bungener W, Bialek R. Fatal Streptococcus suis septicemia in an abattoir worker. Eur J Clin Microbiol Infect Dis. 1989;8:306-8.

[205] Strangmann E, Froleke H, Kohse KP. Septic shock caused by Streptococcus suis: case report and investigation of a risk group. Int J Hyg Environ Health. 2002;205:385-92.

[206] Rosenkranz M, Elsner HA, Sturenburg HJ, Weiller C, Rother J, Sobottka I. Streptococcus suis meningitis and septicemia contracted from a wild boar in Germany. J Neurol. 2003;250:869-70.

[207] Heidt MC, Mohamed W, Hain T, Vogt PR, Chakraborty T, Domann E. Human infective endocarditis caused by Streptococcus suis serotype 2. J Clin Microbiol. 2005;43:4898-901.

[208] Mazokopakis EE, Kofteridis DP, Papadakis JA, Gikas AH, Samonis GJ. First case report of Streptococcus suis septicaemia and meningitis from Greece. Eur J Neurol. 2005;12:487-9.

[209] Voutsadakis IA. Streptococcus suis endocarditis and colon carcinoma: a case report. Clin Colorectal Cancer. 2006;6:226-8.

[210] Chatzopoulou M, Voulgaridou I, Papalas D, Vasiliou P, Tsiakalou M. Third Case of Streptococcus suis Infection in Greece. Case Rep Infect Dis. 2015;2015:505834.

[211] Baddeley PG. Streptococcus suis infection. Occup Med (Lond). 1995;45:222.

[212] Perseghin P, Bezzi G, Troupioti P, Gallina M. Streptococcus suis meningitis in an Italian blood donor. Lancet. 1995;346:1305-6.

[213] Camporese A, Tizianel G, Bruschetta G, Cruciatti B, Pomes A. Human meningitis caused by Streptococcus suis: the first case report from north-eastern Italy. Infez Med. 2007;15:111-4.

[214] Manzin A, Palmieri C, Serra C, Saddi B, Princivalli MS, Loi G, et al. Streptococcus suis meningitis without history of animal contact, Italy. Emerg Infect Dis. 2008;14:1946-8.

[215] Mancini F, Adamo F, Creti R, Monaco M, Alfarone G, Pantosti A, et al. A fatal case of streptococcal toxic shock syndrome caused by Streptococcus suis carrying tet (40) and tet (O/W/32/O), Italy. J Infect Chemother. 2016;22:774-6.

[216] van de Beek D, Spanjaard L, de Gans J. Streptococcus suis meningitis in the Netherlands. J Infect. 2008;57:158-61.

[217] Zalas-Wiecek P, Michalska A, Grabczewska E, Olczak A, Pawlowska M, Gospodarek E. Human meningitis caused by Streptococcus suis. J Med Microbiol. 2013;62:483-5.

[218] Bojarska A, Molska E, Janas K, Skoczynska A, Stefaniuk E, Hryniewicz W, et al. Streptococcus suis in invasive human infections in Poland: clonality and determinants of virulence and antimicrobial resistance. Eur J Clin Microbiol Infect Dis. 2016;35:917-25.

[219] Taipa R, Lopes V, Magalhaes M. Streptococcus suis meningitis: first case report from Portugal. J Infect. 2008;56:482-3.

[220] Sena Esteves S, Carvalho de Almeida J, Abrunhosa J, Almeida ESC, Arshad Q. Pig's ear: Streptococcus suis Meningitis and its associated inner ear implications. IDCases. 2017;10:55-7.

[221] Tarradas C, Borge C, Arenas A, Maldonado A, Astorga R, Miranda A, et al. Suilysin production by Streptococcus suis strains isolated from diseased and healthy carrier pigs in Spain. Vet Rec. 2001;148:183-4.

[222] Hidalgo A, Ropero F, Palacios R, Garcia V, Santos J. Meningitis due to Streptococcus suis with no contact with pigs or porcine products. J Infect. 2007;55:478.

[223] Vela AI, Aspiroz C, Fortuno B, Tirado G, Sierra J, Martinez R, et al. Meningitis caused by an unusual genotype (ST3) of Streptococcus suis. Infection. 2013;41:701-3.

[224] Gomez-Zorrilla S, Ardanuy C, Lora-Tamayo J, Camara J, Garcia-Somoza D, Pena C, et al. Streptococcus suis infection and malignancy in man, Spain. Emerg Infect Dis. 2014;20:1067-8.

[225] Gustavsson C, Rasmussen M. Septic arthritis caused by Streptococcus suis serotype 5 in pig farmer. Emerg Infect Dis. 2014;20:489-90.

[226] McLendon BF, Bron AJ, Mitchell CJ. Streptococcus suis type II (group R) as a cause of endophthalmitis. Br J Ophthalmol. 1978;62:729-31.

[227] Joynson DH. Infections of man and group R streptococci. Br J Clin Pract. 1980;34:147-9, 54.

[228] Twort CH. Group R streptococcal meningitis (Streptococcus suis type II): a new industrial disease? Br Med J (Clin Res Ed). 1981;282:523-4.

[229] Clements MR, Hamilton DV, Clifton-Hadley FA, O'Reilly JF. Streptococcus suis type II infection. A new industrial disease? Practitioner. 1982;226:323-5.

[230] McNeil NI, Gordon T. Meningitis caused by Streptococcus suis type II. Postgrad Med J. 1986;62:743-4.

[231] Hay PE, Cunniffe JG, Kramer G, France AJ, Gray JA, Watt B. Two cases of Streptococcus suis meningitis. Br J Ind Med. 1989;46:352-3.

[232] Maher D. Streptococcus suis septicaemia presenting as severe acute gastro-enteritis. J Infect. 1991;22:303-4.

[233] Meecham JS, Worth RC. Persistent diplopia following streptococcus suis type 2 meningitis. J R Soc Med. 1992;85:579-80.

[234] Watkins EJ, Brooksby P, Schweiger MS, Enright SM. Septicaemia in a pig-farm worker. Lancet. 2001;357:38.

[235] Rao SS, Mariathas A, Teare L. Meningitis in a butcher. Emerg Med J. 2008;25:607-8.

[236] Normile D. Infectious diseases - WHO probes deadliness of China's pig-borne disease. Science. 2005;309:1308-9.

[237] Ye C, Zhu X, Jing H, Du H, Segura M, Zheng H, et al. Streptococcus suis sequence type 7 outbreak, Sichuan, China. Emerg Infect Dis. 2006;12:1203-8.

[238] Yu H, Jing H, Chen Z, Zheng H, Zhu X, Wang H, et al. Human Streptococcus suis outbreak, Sichuan, China. Emerg Infect Dis. 2006;12:914-20.

[239] Chau PY, Huang CY, Kay R. Streptococcus suis meningitis. An important underdiagnosed disease in Hong Kong. Med J Aust. 1983;1:414-6, 7.

[240] Woo J. Streptococcus suis meningitis in man in Hong Kong. Trans R Soc Trop Med Hyg. 1986;80:848-9.

[241] Ip M, Fung KS, Chi F, Cheuk ES, Chau SS, Wong BW, et al. Streptococcus suis in Hong Kong. Diagn Microbiol Infect Dis. 2007;57:15-20.

[242] Ma E, Chung PH, So T, Wong L, Choi KM, Cheung DT, et al. Streptococcus suis infection in Hong Kong: an emerging infectious disease? Epidemiol Infect. 2008;136:1691-7.

[243] Taniyama D, Sakurai M, Sakai T, Kikuchi T, Takahashi T. Human case of bacteremia due to Streptococcus suis serotype 5 in Japan: The first report and literature review. IDCases. 2016;6:36-8.

[244] Choi SM, Cho BH, Choi KH, Nam TS, Kim JT, Park MS, et al. Meningitis caused by Streptococcus suis: case report and review of the literature. J Clin Neurol. 2012;8:79-82.

[245] Raberahona M, Rasoanandrasana S, Rahajamanana VL, Ranaivo-Rabetokotany F, Andriananja V, Rakotomalala FA, et al. Novel Streptococcus suis Sequence Type 834 among Humans, Madagascar. Emerg Infect Dis. 2018;24:391-2.

[246] Rajahram GS, Hameed AA, Menon J, William T, Tambyah PA, Yeo TW. Case report: two human Streptococcus suis infections in Borneo, Sabah, Malaysia. BMC Infect Dis. 2017;17:188.

[247] Tambyah PA, Kumarasinghe G, Chan HL, Lee KO. Streptococcus suis infection complicated by purpura fulminans and rhabdomyolysis: case report and review. Clin Infect Dis. 1997;24:710-2.

[248] Chan YC, Wilder-Smith A, Ong BK, Kumarasinghe G, Wilder-Smith E. Adult community acquired bacterial meningitis in a Singaporean teaching hospital. A seven-year overview (1993-2000). Singapore Med J. 2002;43:632-6.

[249] Tan JH, Yeh BI, Seet CS. Deafness due to haemorrhagic labyrinthitis and a review of relapses in Streptococcus suis meningitis. Singapore Med J. 2010;51:e30-3.

[250] Yen MY, Liu YC, Wang JH, Chen YS, Wang YH, Cheng DL. Streptococcus suis meningitis complicated with permanent perceptive deafness: report of a case. J Formos Med Assoc. 1994;93:349-51.

[251] Tsai HY, Liao CH, Liu CY, Huang YT, Teng LJ, Hsueh PR. Streptococcus suis infection in Taiwan, 2000-2011. Diagn Microbiol Infect Dis. 2012;74:75-7.

[252] Vilaichone RK, Mahachai V, Nunthapisud P. Streptococcus suis peritonitis: case report. J Med Assoc Thai. 2000;83:1274-7.

[253] Fongcom A, Pruksakorn S, Mongkol R, Tharavichitkul P, Yoonim N. Streptococcus suis infection in northern Thailand. J Med Assoc Thai. 2001;84:1502-8.

[254] Vilaichone RK, Vilaichone W, Nunthapisud P, Wilde H. Streptococcus suis infection in Thailand. J Med Assoc Thai. 2002;85 Suppl 1:S109-17.

[255] Donsakul K, Dejthevaporn C, Witoonpanich R. Streptococcus suis infection: clinical features and diagnostic pitfalls. Southeast Asian J Trop Med Public Health. 2003;34:154-8.

[256] Wangkaew S, Chaiwarith R, Tharavichitkul P, Supparatpinyo K. Streptococcus suis infection: a series of 41 cases from Chiang Mai University Hospital. J Infect. 2006;52:455-60.

[257] Rusmeechan S, Sribusara P. Streptococcus suis meningitis: the newest serious infectious disease. J Med Assoc Thai. 2008;91:654-8.

[258] Kerdsin A, Dejsirilert S, Sawanpanyalert P, Boonnark A, Noithachang W, Sriyakum D, et al. Sepsis and spontaneous bacterial peritonitis in Thailand. Lancet. 2011;378:960.

[259] Kerdsin A, Dejsirilert S, Puangpatra P, Sripakdee S, Chumla K, Boonkerd N, et al. Genotypic profile of Streptococcus suis serotype 2 and clinical features of infection in humans, Thailand. Emerg Infect Dis. 2011;17:835-42.

[260] Pachirat O, Taksinachanekit S, Mootsikapun P, Kerdsin A. Human Streptococcus suis Endocarditis: Echocardiographic Features and Clinical Outcome. Clin Med Insights Cardiol. 2012;6:119-23.

[261] Takeuchi D, Kerdsin A, Pienpringam A, Loetthong P, Samerchea S, Luangsuk P, et al. Population-based study of Streptococcus suis infection in humans in Phayao Province in northern Thailand. PLoS One. 2012;7:e31265.

[262] Tien le HT, Sugiyama N, Duangsonk K, Tharavichitkul P, Osawa R. Phenotypic and PCR-based identification of bacterial strains isolated from patients with suspected streptococcus suis infection in northern Thailand. Jpn J Infect Dis. 2012;65:171-4.

[263] Praphasiri P, Owusu JT, Thammathitiwat S, Ditsungnoen D, Boonmongkon P, Sangwichian O, et al. Streptococcus suis infection in hospitalized patients, Nakhon Phanom Province, Thailand. Emerg Infect Dis. 2015;21:345-8.

[264] Takeuchi D, Kerdsin A, Akeda Y, Chiranairadul P, Loetthong P, Tanburawong N, et al. Impact of a Food Safety Campaign on Streptococcus suis Infection in Humans in Thailand. Am J Trop Med Hyg. 2017;96:1370-7.

[265] Nghia HD, Tu le TP, Wolbers M, Thai CQ, Hoang NV, Nga TV, et al. Risk factors of Streptococcus suis infection in Vietnam. A case-control study. PLoS One. 2011;6:e17604.

[266] Hoa NT, Chieu TT, Nghia HD, Mai NT, Anh PH, Wolbers M, et al. The antimicrobial resistance patterns and associated determinants in Streptococcus suis isolated from humans in southern Vietnam, 1997-2008. BMC Infect Dis. 2011;11:6.

[267] Chen FL, Hsueh PR, Ou TY, Hsieh TC, Lee WS. A cluster of Streptococcus suis meningitis in a family who traveled to Taiwan from Southern Vietnam. J Microbiol Immunol Infect. 2016;49:468-9.

[268] Callejo R, Zheng H, Du P, Prieto M, Xu J, Zielinski G, et al. Streptococcus suis serotype 2 strains isolated in Argentina (South America) are different from those recovered in North America and present a higher risk for humans. JMM Case Rep. 2016;3:e005066.

[269] Kennedy KJ, Jadeer AA, Ong CW, Senanayake SN, Collignon PJ. Two cases of Streptococcus suis endocarditis in Australian piggery workers. Med J Aust. 2008;189:413.

[270] Tramontana AR, Graham M, Sinickas V, Bak N. An Australian case of Streptococcus suis toxic shock syndrome associated with occupational exposure to animal carcasses. Med J Aust. 2008;188:538-9.

[271] Soares TC, Gottschalk M, Lacouture S, Megid J, Ribolla PE, Pantoja JC, et al. Streptococcus suis in employees and the environment of swine slaughterhouses in Sao Paulo, Brazil: Occurrence, risk factors, serotype distribution, and antimicrobial susceptibility. Can J Vet Res. 2015;79:279-84.

[272] Trottier S, Higgins R, Brochu G, Gottschalk M. A case of human endocarditis due to Streptococcus suis in North America. Rev Infect Dis. 1991;13:1251-2.

[273] Michaud S, Duperval R, Higgins R. Streptococcus suis meningitis: First case reported in Quebec. Can J Infect Dis. 1996;7:329-31.

[274] Haleis A, Alfa M, Gottschalk M, Bernard K, Ronald A, Manickam K. Meningitis caused by Streptococcus suis serotype 14, North America. Emerg Infect Dis. 2009;15:350-2.

[275] Koch E, Fuentes G, Carvajal R, Palma R, Aguirre V, Cruz C, et al. [Streptococcus suis meningitis in pig farmers: report of first two cases in Chile]. Rev Chilena Infectol. 2013;30:557-61.

[276] Prince-David M, Salou M, Marois-Crehan C, Assogba K, Plainvert C, Balogou KA, et al. Human meningitis due to Streptococcus suis in Lome, Togo: a case report. BMC Infect Dis. 2016;16:651.

[277] Willenburg KS, Sentochnik DE, Zadoks RN. Human Streptococcus suis meningitis in the United States. N Engl J Med. 2006;354:1325.

[278] Lee GT, Chiu CY, Haller BL, Denn PM, Hall CS, Gerberding JL. Streptococcus suis meningitis, United States. Emerg Infect Dis. 2008;14:183-5.

[279] Gomez E, Kennedy CC, Gottschalk M, Cunningham SA, Patel R, Virk A. Streptococcus suis-related prosthetic joint infection and streptococcal toxic shock-like syndrome in a pig farmer in the United States. J Clin Microbiol. 2014;52:2254-8.

[280] Cheung PY, Lo KL, Cheung TT, Yeung WH, Leung PH, Kam KM. Streptococcus suis in retail markets: how prevalent is it in raw pork? Int J Food Microbiol. 2008;127:316-20.

[281] Gottschalk M, Xu J, Lecours M, Grenier D, Fittipaldi N, Segura M. *Streptococcus suis* Infections in Humans: What is the prognosis for Western countries? (Part I). Clinical Microbiology Newsletter. 2010;32:89-96.

[282] Nakayama T, Takeuchi D, Matsumura T, Akeda Y, Fujinaga Y, Oishi K. Alcohol consumption promotes the intestinal translocation of Streptococcus suis infections. Microb Pathog. 2013;65:14-20.

[283] Gallagher F. Streptococcus infection and splenectomy. Lancet. 2001;357:1129-30.

[284] Vecht U, Wisselink HJ, Stockhofe-Zurwieden N, Smith HE. Characterization of virulence of the Streptococcus suis serotype 2 reference strain Henrichsen S 735 in newborn gnotobiotic pigs. Vet Microbiol. 1996;51:125-36.

[285] Hu X, Zhu F, Wang H, Chen S, Wang G, Sun J, et al. [Studies on human streptococcal infectious syndrome caused by infected pigs]. Zhonghua Yu Fang Yi Xue Za Zhi. 2000;34:150-2.

[286] Zhu J, Tang JQ, Guo HB. [Identification of the pathogens responsible for toxic shock syndrome in the severe acute outbreak]. Chin J Contag. 2001;19:17-9.

[287] Tang JQ, Bai W, Zhu J, Guo HB. [Epidemiological and pathogenic study on the outbreak of toxic shock syndrome and meningocephalitis caused by swine streptococcus]. Chin J Med Colle PLA. 2004;19:59-63.

[288] Salasia SI, Lammler C, Devriese LA. Serotypes and putative virulence markers of Streptococcus suis isolates from cats and dogs. Res Vet Sci. 1994;57:259-61.

[289] Devriese LA, Cruz Colque JI, De Herdt P, Haesebrouck F. Identification and composition of the tonsillar and anal enterococcal and streptococcal flora of dogs and cats. J Appl Bacteriol. 1992;73:421-5.

[290] Devriese LA, Haesebrouck F. Streptococcus suis infections in horses and cats. Vet Rec. 1992;130:380.

[291] Hojo K, Tamura A, Mizoguchi C, Kato D, Ohshima T, Maeda N. Predominant bacteria recovered from a periodontitis site in a hamster model raised by silk-ligature with Porphyromonas gingivalis infection. Biosci Biotechnol Biochem. 2008;72:1348-51.

[292] Devriese LA, Sustronck B, Maenhout T, Haesebrouck F. Streptococcus suis meningitis in a horse. Vet Rec. 1990;127:68.

[293] Baums CG, Verkuhlen GJ, Rehm T, Silva LM, Beyerbach M, Pohlmeyer K, et al. Prevalence of Streptococcus suis genotypes in wild boars of Northwestern Germany. Appl Environ Microbiol. 2007;73:711-7.

[294] Risco D, Fernandez-Llario P, Cuesta JM, Garcia-Jimenez WL, Goncalves P, Martinez R, et al. Fatal Case of Streptococcus Suis Infection in a Young Wild Boar (Sus Scrofa) from Southwestern Spain. J Zoo Wildl Med. 2015;46:370-3.

[295] Sanchez del Rey V, Fernandez-Garayzabal JF, Briones V, Iriso A, Dominguez L, Gottschalk M, et al. Genetic analysis of Streptococcus suis isolates from wild rabbits. Vet Microbiol. 2013;165:483-6.

[296] Devriese LA, Haesebrouck F, de Herdt P, Dom P, Ducatelle R, Desmidt M, et al. Streptococcus suis infections in birds. Avian Pathol. 1994;23:721-4.

[297] Grebe T, Bergenthal D, Fahr AM, Scheja HW. [Meningitis caused by Streptococcus suis type 2 in an adult]. Dtsch Med Wochenschr. 1997;122:1244-7.

[298] Halaby T, Hoitsma E, Hupperts R, Spanjaard L, Luirink M, Jacobs J. Streptococcus suis meningitis, a poacher's risk. Eur J Clin Microbiol Infect Dis. 2000;19:943-5.

[299] Si Y, Yuan F, Chang H, Liu X, Li H, Cai K, et al. Contribution of glutamine synthetase to the virulence of Streptococcus suis serotype 2. Vet Microbiol. 2009;139:80-8.

[300] Baums CG, Kaim U, Fulde M, Ramachandran G, Goethe R, Valentin-Weigand P. Identification of a novel virulence determinant with serum opacification activity in Streptococcus suis. Infect Immun. 2006;74:6154-62.

[301] Zhang H, Fan H, Lu C. Identification of a novel virulence-related gene in Streptococcus suis type 2 strains. Curr Microbiol. 2010;61:494-9.

[302] Segura M, Fittipaldi N, Calzas C, Gottschalk M. Critical Streptococcus suis Virulence Factors: Are They All Really Critical? Trends Microbiol. 2017;25:585-99.

[303] Segura M, Zheng H, de Greeff A, Gao GF, Grenier D, Jiang Y, et al. Latest developments on Streptococcus suis: an emerging zoonotic pathogen: part 1. Future Microbiol. 2014;9:441-4.

[304] Segura M, Zheng H, Greeff A, Gao GF, Grenier D, Jiang Y, et al. Latest developments on Streptococcus suis: an emerging zoonotic pathogen: part 2. Future Microbiol. 2014;9:587-91.

[305] Charland N, Harel J, Kobisch M, Lacasse S, Gottschalk M. Streptococcus suis serotype 2 mutants deficient in capsular expression. Microbiology. 1998;144 (Pt 2):325-32.

[306] Gottschalk M, Higgins R, Quessy S. Dilemma of the virulence of Strptococcus suis strains. J Clin Microbiol. 1999;37:4202-3.

[307] Vecht U, Stockhofe-Zurwieden N, Tetenburg BJ, Wisselink HJ, Smith HE. Virulence of Streptococcus suis type 2 for mice and pigs appeared host-specific. Vet Microbiol. 1997;58:53-60.

[308] Vecht U, Stockhofe-Zurwieden N, Tetenburg BJ, Wisselink HJ, Smith HE. Murine and pig models of Streptococcus suis type 2 infections are incompatible. Adv Exp Med Biol. 1997;418:827-9.

[309] Allen AG, Bolitho S, Lindsay H, Khan S, Bryant C, Norton P, et al. Generation and characterization of a defined mutant of Streptococcus suis lacking suilysin. Infect Immun. 2001;69:2732-5.

[310] Gottschalk M, Higgins R, Jacques M, Dubreuil D. Production and characterization of two Streptococcus suis capsular type 2 mutants. Vet Microbiol. 1992;30:59-71.

[311] Tan C, Fu S, Liu M, Jin M, Liu J, Bei W, et al. Cloning, expression and characterization of a cell wall surface protein, 6-phosphogluconate-dehydrogenase, of Streptococcus suis serotype 2. Vet Microbiol. 2008;130:363-70.

[312] Fu S, Yuan F, Zhang M, Tan C, Chen H, Bei W. Cloning, expression and characterization of a cell wall surface protein, 6-phosphogluconate dehydrogenase, of Haemophilus parasuis. Res Vet Sci. 2012;93:57-62.

[313] Aranda J, Garrido ME, Fittipaldi N, Cortes P, Llagostera M, Gottschalk M, et al. The cation-uptake regulators AdcR and Fur are necessary for full virulence of Streptococcus suis. Vet Microbiol. 2010;144:246-9.

[314] Tikkanen K, Haataja S, Finne J. The galactosyl-(alpha 1-4)-galactose-binding adhesin of Streptococcus suis: occurrence in strains of different hemagglutination activities and induction of opsonic antibodies. Infect Immun. 1996;64:3659-65.

[315] Winterhoff N, Goethe R, Gruening P, Rohde M, Kalisz H, Smith HE, et al. Identification and characterization of two temperature-induced surface-associated proteins of Streptococcus suis with high homologies to members of the arginine deiminase system of Streptococcus pyogenes. Journal of Bacteriology. 2002;184:6768-76.

[316] Gruening P, Fulde M, Valentin-Weigand P, Goethe R. Structure, regulation, and putative function of the arginine deiminase system of Streptococcus suis. J Bacteriol. 2006;188:361-9.

[317] Ferrando ML, Fuentes S, de Greeff A, Smith H, Wells JM. ApuA, a multifunctional alpha-glucan-degrading enzyme of Streptococcus suis, mediates adhesion to porcine epithelium and mucus. Microbiology. 2010;156:2818-28.

[318] Willenborg J, Fulde M, de Greeff A, Rohde M, Smith HE, Valentin-Weigand P, et al. Role of glucose and CcpA in capsule expression and virulence of Streptococcus suis. Microbiology. 2011;157:1823-33.

[319] Wilson TL, Jeffers J, Rapp-Gabrielson VJ, Martin S, Klein LK, Lowery DE, et al. A novel signature-tagged mutagenesis system for Streptococcus suis serotype 2. Vet Microbiol. 2007;122:135-45.

[320] Li J, Tan C, Zhou Y, Fu S, Hu L, Hu J, et al. The two-component regulatory system CiaRH contributes to the virulence of Streptococcus suis 2. Vet Microbiol. 2011;148:99-104.

[321] Pan X, Ge J, Li M, Wu B, Wang C, Wang J, et al. The orphan response regulator CovR: a globally negative modulator of virulence in Streptococcus suis serotype 2. J Bacteriol. 2009;191:2601-12.

[322] Chabot-Roy G, Willson P, Segura M, Lacouture S, Gottschalk M. Phagocytosis and killing of Streptococcus suis by porcine neutrophils. Microb Pathog. 2006;41:21-32.

[323] Jobin MC, Martinez G, Motard J, Gottschalk M, Grenier D. Cloning, purification, and enzymatic properties of dipeptidyl peptidase IV from the swine pathogen Streptococcus suis. J Bacteriol. 2005;187:795-9.

[324] Ge J, Feng Y, Ji H, Zhang H, Zheng F, Wang C, et al. Inactivation of dipeptidyl peptidase IV attenuates the virulence of Streptococcus suis serotype 2 that causes streptococcal toxic shock syndrome. Curr Microbiol. 2009;59:248-55.

[325] Fittipaldi N, Sekizaki T, Takamatsu D, Harel J, Dominguez-Punaro Mde L, Von Aulock S, et al. D-alanylation of lipoteichoic acid contributes to the virulence of Streptococcus suis. Infect Immun. 2008;76:3587-94.

[326] Niven DF, Ekins A. Iron content of Streptococcus suis and evidence for a dpr homologue. Can J Microbiol. 2001;47:412-6.

[327] Haataja S, Penttinen A, Pulliainen AT, Tikkanen K, Finne J, Papageorgiou AC. Expression, purification and crystallization of Dpr, a ferritin-like protein from the Gram-positive meningitis-associated bacterium Streptococcus suis. Acta Crystallogr D Biol Crystallogr. 2002;58:1851-3.

[328] Esgleas M, Li Y, Hancock MA, Harel J, Dubreuil JD, Gottschalk M. Isolation and characterization of alpha-enolase, a novel fibronectin-binding protein from Streptococcus suis. Microbiology. 2008;154:2668-79.

[329] Smith HE, Reek FH, Vecht U, Gielkens AL, Smits MA. Repeats in an extracellular protein of weakly pathogenic strains of Streptococcus suis type 2 are absent in pathogenic strains. Infect Immun. 1993;61:3318-26.

[330] Smith HE, Vecht U, Wisselink HJ, Stockhofe-Zurwieden N, Biermann Y, Smits MA. Mutants of Streptococcus suis types 1 and 2 impaired in expression of muramidase-released protein and extracellular protein induce disease in newborn germfree pigs. Infect Immun. 1996;64:4409-12.

[331] Vanier G, Fittipaldi N, Slater JD, de la Cruz Dominguez-Punaro M, Rycroft AN, Segura M, et al. New putative virulence factors of Streptococcus suis involved in invasion of porcine brain microvascular endothelial cells. Microb Pathog. 2009;46:13-20.

[332] de Greeff A, Buys H, Verhaar R, Dijkstra J, van Alphen L, Smith HE. Contribution of fibronectin-binding protein to pathogenesis of Streptococcus suis serotype 2. Infect Immun. 2002;70:1319-25.

[333] Aranda J, Cortes P, Garrido ME, Fittipaldi N, Llagostera M, Gottschalk M, et al. Contribution of the FeoB transporter to Streptococcus suis virulence. Int Microbiol. 2009;12:137-43.

[334] Brassard J, Gottschalk M, Quessy S. Cloning and purification of the Streptococcus suis serotype 2 glyceraldehyde-3-phosphate dehydrogenase and its involvement as an adhesin. Vet Microbiol. 2004;102:87-94.

[335] Wang K, Lu C. Adhesion activity of glyceraldehyde-3-phosphate dehydrogenase in a Chinese Streptococcus suis type 2 strain. Berl Munch Tierarztl Wochenschr. 2007;120:207-9.

[336] Allen AG, Lindsay H, Seilly D, Bolitho S, Peters SE, Maskell DJ. Identification and characterisation of hyaluronate lyase from Streptococcus suis. Microb Pathog. 2004;36:327-35.

[337] Zhang A, Mu X, Chen B, Liu C, Han L, Chen H, et al. Identification and characterization of IgA1 protease from Streptococcus suis. Vet Microbiol. 2010;140:171-5.

[338] Zhang A, Mu X, Chen B, Han L, Chen H, Jin M. IgA1 protease contributes to the virulence of Streptococcus suis. Vet Microbiol. 2011;148:436-9.

[339] Schreur PJW, Rebel JMJ, Smits MA, van Putten JPM, Smith HE. Lgt Processing Is an Essential Step in Streptococcus suis Lipoprotein Mediated Innate Immune Activation. Plos One. 2011;6.

[340] Aranda J, Teixido L, Fittipaldi N, Cortes P, Llagostera M, Gottschalk M, et al. Inactivation of the gene encoding zinc-binding lipoprotein 103 impairs the infectivity of Streptococcus suis. Can J Vet Res. 2012;76:72-6.

[341] De Greeff A, Hamilton A, Sutcliffe IC, Buys H, Van Alphen L, Smith HE. Lipoprotein signal peptidase of Streptococcus suis serotype 2. Microbiology. 2003;149:1399-407.

[342] Han XG, Lu CP. Detection of autoinducer-2 and analysis of the profile of luxS and pfs transcription in Streptococcus suis serotype 2. Curr Microbiol. 2009;58:146-52.

[343] Wang Y, Zhang W, Wu Z, Zhu X, Lu C. Functional analysis of luxS in Streptococcus suis reveals a key role in biofilm formation and virulence. Vet Microbiol. 2011;152:151-60.

[344] Wang Y, Yi L, Zhang Z, Fan H, Cheng X, Lu C. Overexpression of luxS cannot increase autoinducer-2 production, only affect the growth and biofilm formation in Streptococcus suis. ScientificWorldJournal. 2013;2013:924276.

[345] Smith HE, Vecht U, Gielkens AL, Smits MA. Cloning and nucleotide sequence of the gene encoding the 136-kilodalton surface protein (muramidase-released protein) of Streptococcus suis type 2. Infect Immun. 1992;60:2361-7.

[346] Feng Y, Cao M, Shi J, Zhang H, Hu D, Zhu J, et al. Attenuation of Streptococcus suis virulence by the alteration of bacterial surface architecture. Sci Rep. 2012;2:710.

[347] Takamatsu D, Osaki M, Tharavichitkul P, Takai S, Sekizaki T. Allelic variation and prevalence of serum opacity factor among the Streptococcus suis population. J Med Microbiol. 2008;57:488-94.

[348] Fittipaldi N, Gottschalk M, Vanier G, Daigle F, Harel J. Use of selective capture of transcribed sequences to identify genes preferentially expressed by Streptococcus suis upon interaction with porcine brain microvascular endothelial cells. Appl Environ Microbiol. 2007;73:4359-64.

[349] Fittipaldi N, Sekizaki T, Takamatsu D, de la Cruz Dominguez-Punaro M, Harel J, Bui NK, et al. Significant contribution of the pgdA gene to the virulence of Streptococcus suis. Mol Microbiol. 2008;70:1120-35.

[350] Jobin MC, Fortin J, Willson PJ, Gottschalk M, Grenier D. Acquisition of plasmin activity and induction of arachidonic acid release by Streptococcus suis in contact with human brain microvascular endothelial cells. FEMS Microbiol Lett. 2005;252:105-11.

[351] de Greeff A, Buys H, van Alphen L, Smith HE. Response regulator important in pathogenesis of Streptococcus suis serotype 2. Microb Pathog. 2002;33:185-92.

[352] Wu T, Chang H, Tan C, Bei W, Chen H. The orphan response regulator RevSC21 controls the attachment of Streptococcus suis serotype-2 to human laryngeal epithelial cells and the expression of virulence genes. FEMS Microbiol Lett. 2009;292:170-81.

[353] Zheng F, Ji H, Cao M, Wang C, Feng Y, Li M, et al. Contribution of the Rgg transcription regulator to metabolism and virulence of Streptococcus suis serotype 2. Infect Immun. 2011;79:1319-28.

[354] Li M, Wang C, Feng Y, Pan X, Cheng G, Wang J, et al. SalK/SalR, a two-component signal transduction system, is essential for full virulence of highly invasive Streptococcus suis serotype 2. PLoS One. 2008;3:e2080.

[355] Osaki M, Takamatsu D, Shimoji Y, Sekizaki T. Characterization of Streptococcus suis genes encoding proteins homologous to sortase of gram-positive bacteria. J Bacteriol. 2002;184:971-82.

[356] Dramsi S, Trieu-Cuot P, Bierne H. Sorting sortases: a nomenclature proposal for the various sortases of Gram-positive bacteria. Res Microbiol. 2005;156:289-97.

[357] Vanier G, Sekizaki T, Dominguez-Punaro MC, Esgleas M, Osaki M, Takamatsu D, et al. Disruption of srtA gene in Streptococcus suis results in decreased interactions with endothelial cells and extracellular matrix proteins. Vet Microbiol. 2008;127:417-24.

[358] Wang C, Li M, Feng Y, Zheng F, Dong Y, Pan X, et al. The involvement of sortase A in high virulence of STSS-causing Streptococcus suis serotype 2. Arch Microbiol. 2009;191:23-33.

[359] Takamatsu D, Nishino H, Ishiji T, Ishii J, Osaki M, Fittipaldi N, et al. Genetic organization and preferential distribution of putative pilus gene clusters in Streptococcus suis. Vet Microbiol. 2009;138:132-9.

[360] Fittipaldi N, Takamatsu D, de la Cruz Dominguez-Punaro M, Lecours MP, Montpetit D, Osaki M, et al. Mutations in the gene encoding the ancillary pilin subunit of the Streptococcus suis srtF cluster result in pili formed by the major subunit only. PLoS One. 2010;5:e8426.

[361] Okura M, Osaki M, Fittipaldi N, Gottschalk M, Sekizaki T, Takamatsu D. The Minor Pilin Subunit Sgp2 Is Necessary for Assembly of the Pilus Encoded by the srtG Cluster of Streptococcus suis. Journal of Bacteriology. 2011;193:822-31.

[362] Fontaine MC, Perez-Casal J, Willson PJ. Investigation of a novel DNase of Streptococcus suis serotype 2. Infect Immun. 2004;72:774-81.

[363] Slater JD, Allen AG, May JP, Bolitho S, Lindsay H, Maskell DJ. Mutagenesis of Streptococcus equi and Streptococcus suis by transposon Tn917. Vet Microbiol. 2003;93:197-206.

[364] Bonifait L, Vaillancourt K, Gottschalk M, Frenette M, Grenier D. Purification and characterization of the subtilisin-like protease of Streptococcus suis that contributes to its virulence. Vet Microbiol. 2011;148:333-40.

[365] Bonifait L, Grenier D. The SspA subtilisin-like protease of Streptococcus suis triggers a pro-inflammatory response in macrophages through a non-proteolytic mechanism. BMC Microbiol. 2011;11:47.

[366] Shao Z, Pan X, Li X, Liu W, Han M, Wang C, et al. HtpS, a novel immunogenic cell surface-exposed protein of Streptococcus suis, confers protection in mice. FEMS Microbiol Lett. 2011;314:174-82.

[367] Jacobs AA, Loeffen PL, van den Berg AJ, Storm PK. Identification, purification, and characterization of a thiol-activated hemolysin (suilysin) of Streptococcus suis. Infect Immun. 1994;62:1742-8.

[368] Gottschalk MG, Lacouture S, Dubreuil JD. Characterization of Streptococcus-Suis Capsular Type-2 Hemolysin. Microbiology-Uk. 1995;141:189-95.

[369] Jacobs AAC, Vandenberg AJG, Baars JC, Nielsen B, Johannsen LW. Production of Suilysin, the Thiol-Activated Hemolysin of Streptococcus-Suis, by Field Isolates from Diseased Pigs. Veterinary Record. 1995;137:295-6.

[370] Lun S, Perez-Casal J, Connor W, Willson PJ. Role of suilysin in pathogenesis of Streptococcus suis capsular serotype 2. Microb Pathog. 2003;34:27-37.

[371] Langford P, Williams AE, Kroll JS. Superoxide dismutases of pathogenic and non-pathogenic Streptococcus suis type 2 isolates. FEMS Microbiol Lett. 1991;61:347-50.

[372] Li Y, Gottschalk M, Esgleas M, Lacouture S, Dubreuil JD, Willson P, et al. Immunization with recombinant Sao protein confers protection against Streptococcus suis infection. Clin Vaccine Immunol. 2007;14:937-43.

[373] Zhang A, Xie C, Chen H, Jin M. Identification of immunogenic cell wall-associated proteins of Streptococcus suis serotype 2. Proteomics. 2008;8:3506-15.

[374] Hsueh KJ, Lee JW, Hou SM, Chen HS, Chang TC, Chu CY. Evaluation on a Streptococcus suis Vaccine Using Recombinant Sao-L Protein Manufactured by Bioreactors as the Antigen in Pigs. Transboundary and Emerging Diseases. 2014;61:E35-E43.

[375] Roy D, Fittipaldi N, Dumesnil A, Lacouture S, Gottschalk M. The protective protein Sao (surface antigen one) is not a critical virulence factor for Streptococcus suis serotype 2. Microb Pathog. 2014;67-68:31-5.

[376] Schreur PJW, Rebel JMJ, Smits MA, van Putten JPM, Smith HE. TroA of Streptococcus suis Is Required for Manganese Acquisition and Full Virulence. Journal of Bacteriology. 2011;193:5073-80.

[377] Zheng BW, Zhang QM, Gao J, Han HM, Li M, Zhang JR, et al. Insight into the Interaction of Metal Ions with TroA from Streptococcus suis. Plos One. 2011;6.

[378] Li P, Liu J, Zhu L, Qi C, Bei W, Cai X, et al. VirA: a virulence-related gene of Streptococcus suis serotype 2. Microb Pathog. 2010;49:305-10.

[379] Feng Y, Li M, Zhang H, Zheng B, Han H, Wang C, et al. Functional definition and global regulation of Zur, a zinc uptake regulator in a Streptococcus suis serotype 2 strain causing streptococcal toxic shock syndrome. J Bacteriol. 2008;190:7567-78.

[380] Moxon ER, Kroll JS. The role of bacterial polysaccharide capsules as virulence factors. Curr Top Microbiol Immunol. 1990;150:65-85.

[381] Jacques M, Gottschalk M, Foiry B, Higgins R. Ultrastructural study of surface components of Streptococcus suis. J Bacteriol. 1990;172:2833-8.

[382] Graveline R, Segura M, Radzioch D, Gottschalk M. TLR2-dependent recognition of Streptococcus suis is modulated by the presence of capsular polysaccharide which modifies macrophage responsiveness. Int Immunol. 2007;19:375-89.

[383] Segura M, Vanier G, Al-Numani D, Lacouture S, Olivier M, Gottschalk M. Proinflammatory cytokine and chemokine modulation by Streptococcus suis in a whole-blood culture system. FEMS Immunol Med Microbiol. 2006;47:92-106.

[384] Quessy S, Dubreuil JD, Jacques M, Malouin F, Higgins R. Increase of capsular material thickness following in vivo growth of virulent Streptococcus suis serotype 2 strains. FEMS Microbiol Lett. 1994;115:19-26.

[385] Wibawan IW, Lammler C. Relation between encapsulation and various properties of Streptococcus suis. Zentralbl Veterinarmed B. 1994;41:453-9.

[386] Charland N, Kobisch M, Martineau-Doize B, Jacques M, Gottschalk M. Role of capsular sialic acid in virulence and resistance to phagocytosis of Streptococcus suis capsular type 2. FEMS Immunol Med Microbiol. 1996;14:195-203.

[387] Wessels MR, Rubens CE, Benedi VJ, Kasper DL. Definition of a bacterial virulence factor: sialylation of the group B streptococcal capsule. Proc Natl Acad Sci U S A. 1989;86:8983-7.

[388] Billington SJ, Jost BH, Songer JG. Thiol-activated cytolysins: structure, function and role in pathogenesis. FEMS Microbiol Lett. 2000;182:197-205.

[389] Palmer M. The family of thiol-activated, cholesterol-binding cytolysins. Toxicon. 2001;39:1681-9.

[390] Xu L, Huang B, Du H, Zhang XC, Xu J, Li X, et al. Crystal structure of cytotoxin protein suilysin from Streptococcus suis. Protein Cell. 2010;1:96-105.

[391] Feder I, Chengappa MM, Fenwick B, Rider M, Staats J. Partial Characterization of Streptococcus-Suis Type-2 Hemolysin. Journal of Clinical Microbiology. 1994;32:1256-60.

[392] Berthelot-Herault F, Morvan H, Keribin AM, Gottschalk M, Kobisch M. Production of muraminidase-released protein (MRP), extracellular factor (EF) and suilysin by field isolates of Streptococcus suis capsular types 2, 1/2, 9, 7 and 3 isolated from swine in France. Vet Res. 2000;31:473-9.

[393] Fittipaldi N, Fuller TE, Teel JF, Wilson TL, Wolfram TJ, Lowery DE, et al. Serotype distribution and production of muramidase-released protein, extracellular factor and suilysin by field strains of Streptococcus suis isolated in the United States. Vet Microbiol. 2009;139:310-7.

[394] Tenenbaum T, Asmat T, Seitz M, Schroten H, Schwerk C. Biological activities of suilysin: role in Streptococcus suis pathogenesis. Future Microbiol. 2016;11:941-54.

[395] Takeuchi D, Akeda Y, Nakayama T, Kerdsin A, Sano Y, Kanda T, et al. The contribution of suilysin to the pathogenesis of Streptococcus suis meningitis. J Infect Dis. 2014;209:1509-19.

[396] He Z, Pian Y, Ren Z, Bi L, Yuan Y, Zheng Y, et al. Increased production of suilysin contributes to invasive infection of the Streptococcus suis strain 05ZYH33. Mol Med Rep. 2014;10:2819-26.

[397] Charland N, Nizet V, Rubens CE, Kim KS, Lacouture S, Gottschalk M. Streptococcus suis serotype 2 interactions with human brain microvascular endothelial cells. Infect Immun. 2000;68:637-43.

[398] Vanier G, Segura M, Lecours MP, Grenier D, Gottschalk M. Porcine brain microvascular endothelial cell-derived interleukin-8 is first induced and then degraded by Streptococcus suis. Microb Pathog. 2009;46:135-43.

[399] Benga L, Fulde M, Neis C, Goethe R, Valentin-Weigand P. Polysaccharide capsule and suilysin contribute to extracellular survival of Streptococcus suis co-cultivated with primary porcine phagocytes. Vet Microbiol. 2008;132:211-9.

[400] Silva LM, Baums CG, Rehm T, Wisselink HJ, Goethe R, Valentin-Weigand P. Virulence-associated gene profiling of Streptococcus suis isolates by PCR. Vet Microbiol. 2006;115:117-27.

[401] Baums CG, Kock C, Beineke A, Bennecke K, Goethe R, Schroder C, et al. Streptococcus suis bacterin and subunit vaccine immunogenicities and protective efficacies against serotypes 2 and 9. Clin Vaccine Immunol. 2009;16:200-8.

[402] Ellen RP, Gibbons RJ. M protein-associated adherence of Streptococcus pyogenes to epithelial surfaces: prerequisite for virulence. Infect Immun. 1972;5:826-30.

[403] Cunningham MW. Pathogenesis of group A streptococcal infections. Clin Microbiol Rev. 2000;13:470-511.

[404] Signas C, Raucci G, Jonsson K, Lindgren PE, Anantharamaiah GM, Hook M, et al. Nucleotide sequence of the gene for a fibronectin-binding protein from Staphylococcus aureus: use of this peptide sequence in the synthesis of biologically active peptides. Proc Natl Acad Sci U S A. 1989;86:699-703.

[405] Pian Y, Wang P, Liu P, Zheng Y, Zhu L, Wang H, et al. Proteomics identification of novel fibrinogen-binding proteins of Streptococcus suis contributing to antiphagocytosis. Front Cell Infect Microbiol. 2015;5:19.

[406] Wang J, Kong D, Zhang S, Jiang H, Zheng Y, Zang Y, et al. Interaction of fibrinogen and muramidase-released protein promotes the development of Streptococcus suis meningitis. Front Microbiol. 2015;6:1001.

[407] Towers RJ, Fagan PK, Talay SR, Currie BJ, Sriprakash KS, Walker MJ, et al. Evolution of sfbI encoding streptococcal fibronectin-binding protein I: horizontal genetic transfer and gene mosaic structure. J Clin Microbiol. 2003;41:5398-406.

[408] Gottschalk M, Lebrun A, Wisselink H, Dubreuil JD, Smith H, Vecht U. Production of virulence-related proteins by Canadian strains of Streptococcus suis capsular type 2. Can J Vet Res. 1998;62:75-9.

[409] Wisselink HJ, Smith HE, Stockhofe-Zurwieden N, Peperkamp K, Vecht U. Distribution of capsular types and production of muramidase-released protein (MRP) and extracellular factor

(EF) of Streptococcus suis strains isolated from diseased pigs in seven European countries. Vet Microbiol. 2000;74:237-48.

[410] Zhang W, Lu CP. Immunoproteomics of extracellular proteins of Chinese virulent strains of Streptococcus suis type 2. Proteomics. 2007;7:4468-76.

[411] Esgleas M, Dominguez-Punaro Mde L, Li Y, Harel J, Dubreuil JD, Gottschalk M. Immunization with SsEno fails to protect mice against challenge with Streptococcus suis serotype 2. FEMS Microbiol Lett. 2009;294:82-8.

[412] King SJ, Allen AG, Maskell DJ, Dowson CG, Whatmore AM. Distribution, genetic diversity, and variable expression of the gene encoding hyaluronate lyase within the Streptococcus suis population. J Bacteriol. 2004;186:4740-7.

[413] Kock C, Beineke A, Seitz M, Ganter M, Waldmann KH, Valentin-Weigand P, et al. Intranasal immunization with a live Streptococcus suis isogenic ofs mutant elicited suilysin-neutralization titers but failed to induce opsonizing antibodies and protection. Vet Immunol Immunopathol. 2009;132:135-45.

[414] Vollmer W. Structural variation in the glycan strands of bacterial peptidoglycan. FEMS Microbiol Rev. 2008;32:287-306.

[415] Vollmer W, Blanot D, de Pedro MA. Peptidoglycan structure and architecture. FEMS Microbiol Rev. 2008;32:149-67.

[416] Li Y, Martinez G, Gottschalk M, Lacouture S, Willson P, Dubreuil JD, et al. Identification of a surface protein of Streptococcus suis and evaluation of its immunogenic and protective capacity in pigs. Infect Immun. 2006;74:305-12.

[417] Feng Y, Zheng F, Pan X, Sun W, Wang C, Dong Y, et al. Existence and characterization of allelic variants of Sao, a newly identified surface protein from Streptococcus suis. FEMS Microbiol Lett. 2007;275:80-8.

[418] Osaki M, Takamatsu D, Shimoji Y, Sekizaki T. Allelic variation in srtAs of Streptococcus suis strains. FEMS Microbiol Lett. 2003;219:195-201.

[419] Smith HE, Buijs H, de Vries RR, Wisselink HJ, Stockhofe-Zurwieden N, Smits MA. Environmentally regulated genes of Streptococcus suis: identification by the use of iron-restricted conditions in vitro and by experimental infection of piglets. Microbiology. 2001;147:271-80.

[420] Esgleas M, Lacouture S, Gottschalk M. Streptococcus suis serotype 2 binding to extracellular matrix proteins. FEMS Microbiol Lett. 2005;244:33-40.

[421] Tan C, Liu M, Jin M, Liu J, Chen Y, Wu T, et al. The key virulence-associated genes of Streptococcus suis type 2 are upregulated and differentially expressed in vivo. FEMS Microbiol Lett. 2008;278:108-14.

[422] Graham ML, Prescott MJ. The multifactorial role of the 3Rs in shifting the harm-benefit analysis in animal models of disease. Eur J Pharmacol. 2015;759:19-29.

[423] Velikova N, Kavanagh K, Wells JM. Evaluation of Galleria mellonella larvae for studying the virulence of Streptococcus suis. BMC Microbiol. 2016;16:291.

[424] Zaccaria E, Cao R, Wells JM, van Baarlen P. A Zebrafish Larval Model to Assess Virulence of Porcine Streptococcus suis Strains. PLoS One. 2016;11:e0151623.

[425] Auger JP, Chuzeville S, Roy D, Mathieu-Denoncourt A, Xu J, Grenier D, et al. The bias of experimental design, including strain background, in the determination of critical Streptococcus suis serotype 2 virulence factors. PLoS One. 2017;12:e0181920.

[426] Clifton-Hadley FA. Studies of Streptococcus suis type 2 infection in pigs. Vet Res Commun. 1984;8:217-27.

[427] Andresen LO, Tegtmeier C. Passive immunization of pigs against experimental infection with Streptococcus suis serotype 2. Veterinary Microbiology. 2001;81:331-44.

[428] Liu P, Pian Y, Li X, Liu R, Xie W, Zhang C, et al. Streptococcus suis adenosine synthase functions as an effector in evasion of PMN-mediated innate immunit. J Infect Dis. 2014;210:35-45.

[429] Yu J, Pian Y, Ge J, Guo J, Zheng Y, Jiang H, et al. Functional and Structural Characterization of the Antiphagocytic Properties of a Novel Transglutaminase from Streptococcus suis. J Biol Chem. 2015;290:19081-92.

[430] Huang YY, Haines DM, Harding JCS. Snatch-farrowed, porcine-colostrum-deprived (SF-pCD) pigs as a model for swine infectious disease research. Canadian Journal of Veterinary Research-Revue Canadienne De Recherche Veterinaire. 2013;77:81-8.

[431] Clifton-Hadley FA. The epidemiology, diagnosis, treatment and control of Streptococcus suis type 2 infection. Proceedings of the American Association of Swine Practioners. 1986:471–91.

[432] Mwaniki CG, Robertson ID, Trott DJ, Atyeo RF, Lee BJ, Hampson DJ. Clonal analysis and virulence of Australian isolates of Streptococcus suis type 2. Epidemiol Infect. 1994;113:321-34.

[433] Vahle JL, Haynes JS, Andrews JJ. Experimental Reproduction of Haemophilus-Parasuis Infection in Swine - Clinical, Bacteriological, and Morphologic Findings. Journal of Veterinary Diagnostic Investigation. 1995;7:476-80.

[434] Rapp-Gabrielson VJ, Kocur GJ, Clark JT, Muir SK. Haemophilus parasuis: Immunity in swine after vaccination. Veterinary Medicine. 1997;92:83-90.

[435] Vahle JL, Haynes JS, Andrews JJ. Interaction of Haemophilus parasuis with nasal and tracheal mucosa following intranasal inoculation of cesarean derived colostrum deprived (CDCD) swine. Can J Vet Res. 1997;61:200-6.

[436] Brockmeier SL, Loving CL, Mullins MA, Register KB, Nicholson TL, Wiseman BS, et al. Virulence, transmission, and heterologous protection of four isolates of Haemophilus parasuis. Clin Vaccine Immunol. 2013;20:1466-72.

[437] Brockmeier SL, Register KB, Kuehn JS, Nicholson TL, Loving CL, Bayles DO, et al. Virulence and draft genome sequence overview of multiple strains of the swine pathogen Haemophilus parasuis. PLoS One. 2014;9:e103787.

[438] Struve R. A source of CDCD swine for research. American Association of Swine Veterinarians. 1999;117.

[439] Charland N, Jacques M, Lacouture S, Gottschalk M. Characterization and protective activity of a monoclonal antibody against a capsular epitope shared by Streptococcus suis serotypes 1, 2 and 1/2. Microbiology-Uk. 1997;143:3607-14.

[440] Williams AE, Blakemore WF, Alexander TJL. A Murine Model of Streptococcus-Suis Type-2 Meningitis in the Pig. Research in Veterinary Science. 1988;45:394-9.

[441] Gottschalk M. Streptococcosis. In: Zimmerman J, Karriker L, Ramirez A, Schwartz K, Stevenson G, editors. Diseases of Swine (Tenth Edition). USA: Wiley Publishers; 2012. p. 841-55.

[442] Talbot WS, Trevarrow B, Halpern ME, Melby AE, Farr G, Postlethwait JH, et al. A homeobox gene essential for zebrafish notochord development. Nature. 1995;378:150-7.

[443] Gritsman K, Zhang JJ, Cheng S, Heckscher E, Talbot WS, Schier AF. The EGF-CFC protein one-eyed pinhead is essential for nodal signaling. Cell. 1999;97:121-32.

[444] Ober EA, Verkade H, Field HA, Stainier DY. Mesodermal Wnt2b signalling positively regulates liver specification. Nature. 2006;442:688-91.

[445] Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, et al. The zebrafish reference genome sequence and its relationship to the human genome (vol 496, pg 498, 2013). Nature. 2014;505:248-.

[446] Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. Genetics. 2011;188:799-808.

[447] van der Sar AM, Musters RJ, van Eeden FJ, Appelmelk BJ, Vandenbroucke-Grauls CM, Bitter W. Zebrafish embryos as a model host for the real time analysis of Salmonella typhimurium infections. Cell Microbiol. 2003;5:601-11.

[448] van der Sar AM, Appelmelk BJ, Vandenbroucke-Grauls CM, Bitter W. A star with stripes: zebrafish as an infection model. Trends Microbiol. 2004;12:451-7.

[449] Prajsnar TK, Cunliffe VT, Foster SJ, Renshaw SA. A novel vertebrate model of Staphylococcus aureus infection reveals phagocyte-dependent resistance of zebrafish to non-host specialized pathogens. Cell Microbiol. 2008;10:2312-25.

[450] Neely MN, Pfeifer JD, Caparon M. Streptococcus-Zebrafish Model of Bacterial Pathogenesis. Infection and Immunity. 2002;70:3904-14.

[451] Tang F, Zhang W, Lu C. Lysogenic Streptococcus suis isolate SS2-4 containing prophage SMP showed increased mortality in zebra fish compared to the wild-type isolate. PLoS One. 2013;8:e54227.

[452] Wu Z, Zhang W, Lu C. Comparative proteome analysis of secreted proteins of Streptococcus suis serotype 9 isolates from diseased and healthy pigs. Microb Pathog. 2008;45:159-66.

[453] Wu Z, Zhang W, Lu Y, Lu C. Transcriptome profiling of zebrafish infected with Streptococcus suis. Microb Pathog. 2010;48:178-87.

[454] Herbomel P, Thisse B, Thisse C. Ontogeny and behaviour of early macrophages in the zebrafish embryo. Development. 1999;126:3735-45.

[455] Trede NS, Langenau DM, Traver D, Look AT, Zon LI. The use of zebrafish to understand immunity. Immunity. 2004;20:367-79.

[456] Cui C, Benard EL, Kanwal Z, Stockhammer OW, van der Vaart M, Zakrzewska A, et al. Infectious disease modeling and innate immune function in zebrafish embryos. Methods Cell Biol. 2011;105:273-308.

[457] Ellis JD, Graham JR, Mortensen A. Standard methods for wax moth research. Journal of Apicultural Research. 2013;52.

[458] Kavanagh K, Reeves EP. Exploiting the potential of insects for in vivo pathogenicity testing of microbial pathogens. FEMS Microbiol Rev. 2004;28:101-12.

[459] Desbois AP, Coote PJ. Wax moth larva (Galleria mellonella): an in vivo model for assessing the efficacy of antistaphylococcal agents. J Antimicrob Chemother. 2011;66:1785-90.

[460] Evans BA, Rozen DE. A Streptococcus pneumoniae infection model in larvae of the wax moth Galleria mellonella. Eur J Clin Microbiol Infect Dis. 2012;31:2653-60.

[461] Harding CR, Schroeder GN, Collins JW, Frankel G. Use of Galleria mellonella as a Model Organism to Study Legionella pneumophila Infection. Jove-Journal of Visualized Experiments. 2013.

[462] Loh JM, Adenwalla N, Wiles S, Proft T. Galleria mellonella larvae as an infection model for group A streptococcus. Virulence. 2013;4:419-28.

[463] Buckley AA, Faustoferri RC, Quivey RG, Jr. beta-Phosphoglucomutase contributes to aciduricity in Streptococcus mutans. Microbiology. 2014;160:818-27.

[464] Ba X, Harrison EM, Lovering AL, Gleadall N, Zadoks R, Parkhill J, et al. Old Drugs To Treat Resistant Bugs: Methicillin-Resistant Staphylococcus aureus Isolates with mecC Are Susceptible

to a Combination of Penicillin and Clavulanic Acid. Antimicrob Agents Chemother. 2015;59:7396-404.

[465] Smith TL. External morphology of the larva, pupa, and adult of the wax moth. Journal of the Kansas Entomological Society. 1965;38:287- 310.

[466] Kang AR, Lee ML, Lee MY, Kim HK, Yoon MY, Choi YS. Biological Control of Wax Moth, Galleria mellonella L. (Lepidoptera: Pyralidae) by Bacillus thuringiensis. Journal of Apiculture. 2015;30:275-80.

[467] Nielsen RA, Brister CD. Greater Wax Moth (Lepidoptera, Pyralidae) - Behavior of Larvae. Annals of the Entomological Society of America. 1979;72:811-5.

[468] Krams I, Kecko S, Kangassalo K, Moore FR, Jankevics E, Inashkina I, et al. Effects of food quality on trade-offs among growth, immunity and survival in the greater wax moth Galleria mellonella. Insect Science. 2015;22:431-9.

[469] Banville N, Browne N, Kavanagh K. Effect of nutrient deprivation on the susceptibility of Galleria mellonella larvae to infection. Virulence. 2012;3:497-503.

[470] Konkel ME, Tilly K. Temperature-regulated expression of bacterial virulence genes. Microbes Infect. 2000;2:157-66.

[471] Smoot LM, Smoot JC, Graham MR, Somerville GA, Sturdevant DE, Migliaccio CA, et al. Global differential gene expression in response to growth temperature alteration in group A Streptococcus. Proc Natl Acad Sci U S A. 2001;98:10416-21.

[472] Cotter G, Doyle S, Kavanagh K. Development of an insect model for the in vivo pathogenicity testing of yeasts. FEMS Immunol Med Microbiol. 2000;27:163-9.

[473] Ramarao N, Nielsen-Leroux C, Lereclus D. The Insect Galleria mellonella as a Powerful Infection Model to Investigate Bacterial Pathogenesis. Jove-Journal of Visualized Experiments. 2012.

[474] Fedhila S, Buisson C, Dussurget O, Serror P, Glomski IJ, Liehl P, et al. Comparative analysis of the virulence of invertebrate and mammalian pathogenic bacteria in the oral insect infection model Galleria mellonella. J Invertebr Pathol. 2010;103:24-9.

[475] Tsai CJ, Loh JM, Proft T. Galleria mellonella infection models for the study of bacterial diseases and for antimicrobial drug testing. Virulence. 2016;7:214-29.

[476] Chen C, Tang J, Dong W, Wang C, Feng Y, Wang J, et al. A glimpse of streptococcal toxic shock syndrome from comparative genomics of S. suis 2 Chinese isolates. PLoS One. 2007;2:e315.

[477] Holden MT, Hauser H, Sanders M, Ngo TH, Cherevach I, Cronin A, et al. Rapid evolution of virulence and drug resistance in the emerging zoonotic pathogen Streptococcus suis. PLoS One. 2009;4:e6072.

[478] Hu P, Yang M, Zhang AD, Wu JY, Chen B, Hua YF, et al. Complete Genome Sequence of Streptococcus suis Serotype 3 Strain ST3. Journal of Bacteriology. 2011;193:3428-9.

[479] Hu P, Yang M, Zhang AD, Wu JY, Chen B, Hua YF, et al. Complete Genome Sequence of Streptococcus suis Serotype 14 Strain JS14. Journal of Bacteriology. 2011;193:2375-6.

[480] Hu P, Yang M, Zhang AD, Wu JY, Chen B, Hua YF, et al. Comparative Genomics Study of Multi-Drug-Resistance Mechanisms in the Antibiotic-Resistant Streptococcus suis R61 Strain. Plos One. 2011;6.

[481] Zhang A, Yang M, Hu P, Wu J, Chen B, Hua Y, et al. Comparative genomic analysis of Streptococcus suis reveals significant genomic diversity among different serotypes. BMC Genomics. 2011;12:523.

[482] Wang K, Yao H, Lu C, Chen J. Complete Genome Sequence of Streptococcus suis Serotype 16 Strain TL13. Genome Announc. 2013;1.

[483] Ye CY, Zheng H, Zhang J, Jing HQ, Wang L, Xiong YW, et al. Clinical, Experimental, and Genomic Differences between Intermediately Pathogenic, Highly Pathogenic, and Epidemic Streptococcus suis. Journal of Infectious Diseases. 2009;199:97-107.

[484] Beres SB, Musser JM. Contribution of exogenous genetic elements to the group A Streptococcus metagenome. PLoS One. 2007;2:e800.

[485] Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A. 2005;102:13950-5.

[486] Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, Powell E, et al. Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. J Bacteriol. 2007;189:8186-95.

[487] Croucher NJ, Walker D, Romero P, Lennard N, Paterson GK, Bason NC, et al. Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone Streptococcus pneumoniaeSpain23F ST81. J Bacteriol. 2009;191:1480-9.

[488] Jacob AE, Hobbs SJ. Conjugal transfer of plasmid-borne multiple antibiotic resistance in Streptococcus faecalis var. zymogenes. J Bacteriol. 1974;117:360-72.

[489] Stuart JG, Zimmerer EJ, Maddux RL. Conjugation of antibiotic resistance in Streptococcus suis. Vet Microbiol. 1992;30:213-22.

[490] Zaccaria E, van Baarlen P, de Greeff A, Morrison DA, Smith H, Wells JM. Control of competence for DNA transformation in streptococcus suis by genetically transferable pherotypes. PLoS One. 2014;9:e99394.

[491] Zaccaria E, Wels M, van Baarlen P, Wells JM. Temporal Regulation of the Transformasome and Competence Development in Streptococcus suis. Front Microbiol. 2016;7:1922.

[492] Zaccaria E, Wells JM, van Baarlen P. Metabolic Context of the Competence-Induced Checkpoint for Cell Replication in Streptococcus suis. PLoS One. 2016;11:e0153571.

[493] Athey TB, Teatero S, Takamatsu D, Wasserscheid J, Dewar K, Gottschalk M, et al. Population Structure and Antimicrobial Resistance Profiles of Streptococcus suis Serotype 2 Sequence Type 25 Strains. PLoS One. 2016;11:e0150908.

[494] Takamatsu D, Osaki M, Sekizaki T. Sequence analysis of a small cryptic plasmid isolated from Streptococcus suis serotype 2. Curr Microbiol. 2000;40:61-6.

[495] Harel J, Martinez G, Nassar A, Dezfulian H, Labrie SJ, Brousseau R, et al. Identification of an inducible bacteriophage in a virulent strain of Streptococcus suis serotype 2. Infect Immun. 2003;71:6104-8.

[496] Tang F, Bossers A, Harders F, Lu C, Smith H. Comparative genomic analysis of twelve Streptococcus suis (pro)phages. Genomics. 2013;101:336-44.

[497] Tang F, Bossers A, Harders F, Lu C, Smith H. Complete Genome Sequence of the Streptococcus suis Temperate Bacteriophage varphiNJ2. Genome Announc. 2013;1.

[498] Harper DR, Anderson J, Enright MC. Phage therapy: delivering on the promise. Ther Deliv. 2011;2:935-47.

[499] Ryan EM, Gorman SP, Donnelly RF, Gilmore BF. Recent advances in bacteriophage therapy: how delivery routes, formulation, concentration and timing influence the success of phage therapy. J Pharm Pharmacol. 2011;63:1253-64.

[500] Brussow H, Hendrix RW. Phage genomics: small is beautiful. Cell. 2002;108:13-6.

[501] Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc Natl Acad Sci U S A. 1999;96:2192-7.

[502] Wagner PL, Waldor MK. Bacteriophage control of bacterial virulence. Infect Immun. 2002;70:3985-93.

[503] Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brussow H. Phage as agents of lateral gene transfer. Curr Opin Microbiol. 2003;6:417-24.

[504] Alisky J, Iczkowski K, Rapoport A, Troitsky N. Bacteriophages show promise as antimicrobial agents. J Infect. 1998;36:5-15.

[505] Ma YL, Lu CP. Isolation and identification of a bacteriophage capable of infecting Streptococcus suis type 2 strains. Vet Microbiol. 2008;132:340-7.

[506] Maurelli AT, Fernandez RE, Bloch CA, Rode CK, Fasano A. "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of Shigella spp. and enteroinvasive Escherichia coli. Proc Natl Acad Sci U S A. 1998;95:3943-8.

[507] Maurelli AT. Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens. FEMS Microbiol Lett. 2007;267:1-8.

[508] Bliven KA, Maurelli AT. Antivirulence genes: insights into pathogen evolution through gene loss. Infect Immun. 2012;80:4061-70.

[509] Merhej V, Georgiades K, Raoult D. Postgenomic analysis of bacterial pathogens repertoire reveals genome reduction rather than virulence factors. Brief Funct Genomics. 2013;12:291-304.

[510] Clifton-Hadley FA. Studies of *Streptococcus suis* type 2 infection in pigs: University of Cambridge. 1981.

[511] Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821-9.

[512] Li L, Stoeckert CJ, Jr., Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13:2178-89.

[513] Chen F, Mackey AJ, Stoeckert CJ, Jr., Roos DS. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 2006;34:D363-8.

[514] Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, et al. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Curr Protoc Bioinformatics. 2011;Chapter 6:Unit 6 12 1-9.

[515] R Core Team. R: A language and environment for statistical computing. 2013.

[516] Team RDC. R: A language and environment for statistical computing.: R Foundation for Statistical Computing; 2008.

[517] Jombart T, Pontier D, Dufour AB. Genetic markers in the playground of multivariate analysis. Heredity (Edinb). 2009;102:330-41.

[518] Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. Bmc Genetics. 2010;11.

[519] Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24:1403-5.

[520] Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011;27:3070-1.

[521] Heinze G, Ploner M, Dunkler D, Southworth H. logistf: Firth's bias reduced logistic regression. 2013.

[522] Pedhazur E. Multiple Regression in Behavioral Research. 2nd ed. ed. New York: Holt, Rinehart & Winston; 1982.

[523] Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics. 2005;21:3940-1.

[524] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.

[525] Kosinski AS. A weighted generalized score statistic for comparison of predictive values of diagnostic tests. Stat Med. 2013;32:964-77.

[526] Stock C, Hielscher T. DTComPair: comparison of binary diagnostic tests in a paired study design. 2014.

[527] Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. New York: Springer; 2000.

[528] Therneau TM. A Package for Survival Analysis in S. 2015.

[529] Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag 2009.

[530] Wisselink HJ, Reek FH, Vecht U, Stockhofe-Zurwieden N, Smits MA, Smith HE. Detection of virulent strains of Streptococcus suis type 2 and highly virulent strains of Streptococcus suis type 1 in tonsillar specimens of pigs by PCR. Vet Microbiol. 1999;67:143-57.

[531] Todd EW, Hewitt LF. A NEW CULTURE MEDIUM FOR THE PRODUCTION OF ANTIGENIC STREPTOCOCCAL HAEMOLYSIN. J Path Bact. 1932;35:973-4.

[532] Bonifait L, Grignon L, Grenier D. Fibrinogen induces biofilm formation by Streptococcus suis and enhances its antibiotic resistance. Appl Environ Microbiol. 2008;74:4969-72.

[533] Grenier D, Grignon L, Gottschalk M. Characterisation of biofilm formation by a Streptococcus suis meningitis isolate. Vet J. 2009;179:292-5.

[534] Wang Y, Zhang W, Wu Z, Lu C. Reduced virulence is an important characteristic of biofilm infection of Streptococcus suis. FEMS Microbiol Lett. 2011;316:36-43.

[535] Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 2009;55:611-22.

[536] Grant AJ, Restif O, McKinley TJ, Sheppard M, Maskell DJ, Mastroeni P. Modelling within-host spatiotemporal dynamics of invasive bacterial disease. PLoS Biol. 2008;6:e74.

[537] Rutledge RG, Cote C. Mathematics of quantitative kinetic PCR and the application of standard curves. Nucleic Acids Res. 2003;31:e93.

[538] Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing. 2000;10:325-37.

[539] Sturtz S, Ligges U, Gelman A. R2WinBUGS: A package for running WinBUGS from R. Journal of Statistical Software. 2005;12:1-16.

[540] Horton RM, Cai ZL, Ho SN, Pease LR. Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. Biotechniques. 1990;8:528-35.

[541] Athey TB, Auger JP, Teatero S, Dumesnil A, Takamatsu D, Wasserscheid J, et al. Complex Population Structure and Virulence Differences among Serotype 2 Streptococcus suis Strains Belonging to Sequence Type 28. PLoS One. 2015;10:e0137760.

[542] Falush D, Bowden R. Genome-wide association mapping in bacteria? Trends Microbiol. 2006;14:353-5.

[543] Read TD, Massey RC. Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies: a new direction for bacteriology. Genome Med. 2014;6:109.

[544] Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, et al. The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. J Bacteriol. 2008;190:6881-93.

[545] Grad YH, Godfrey P, Cerquiera GC, Mariani-Kurkdjian P, Gouali M, Bingen E, et al. Comparative genomics of recent Shiga toxin-producing Escherichia coli O104:H4: short-term evolution of an emerging pathogen. MBio. 2013;4:e00452-12.

[546] McNally A, Cheng L, Harris SR, Corander J. The evolutionary path to extraintestinal pathogenic, drug-resistant Escherichia coli is marked by drastic reduction in detectable recombination within the core genome. Genome Biol Evol. 2013;5:699-710.

[547] Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science. 2011;331:430-4.

[548] Sheppard SK, Didelot X, Meric G, Torralbo A, Jolley KA, Kelly DJ, et al. Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in Campylobacter. Proc Natl Acad Sci U S A. 2013;110:11923-7.

[549] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559-75.

[550] Laabei M, Recker M, Rudkin JK, Aldeljawi M, Gulay Z, Sloan TJ, et al. Predicting the virulence of MRSA from its genome sequence. Genome Res. 2014;24:839-49.

[551] Van Calsteren MR, Gagnon F, Lacouture S, Fittipaldi N, Gottschalk M. Structure determination of Streptococcus suis serotype 2 capsular polysaccharide. Biochem Cell Biol. 2010;88:513-25.

[552] Willemse N, Schultsz C. Distribution of Type I Restriction-Modification Systems in Streptococcus suis: An Outlook. Pathogens. 2016;5.

[553] Manso AS, Chai MH, Atack JM, Furi L, De Ste Croix M, Haigh R, et al. A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. Nat Commun. 2014;5:5055.

[554] Kouki A, Haataja S, Loimaranta V, Pulliainen AT, Nilsson UJ, Finne J. Identification of a novel streptococcal adhesin P (SadP) protein recognizing galactosyl-alpha1-4-galactose-containing glycoconjugates: convergent evolution of bacterial pathogens to binding of the same host receptor. J Biol Chem. 2011;286:38854-64.

[555] Ferrando ML, Willemse N, Zaccaria E, Pannekoek Y, van der Ende A, Schultsz C. Streptococcal Adhesin P (SadP) contributes to Streptococcus suis adhesion to the human intestinal epithelium. PLoS One. 2017;12:e0175639.

[556] Segura M, Vadeboncoeur N, Gottschalk M. CD14-dependent and -independent cytokine and chemokine production by human THP-1 monocytes stimulated by Streptococcus suis capsular type 2. Clin Exp Immunol. 2002;127:243-54.

[557] Winkler C, Krumsiek J, Lempainen J, Achenbach P, Grallert H, Giannopoulou E, et al. A strategy for combining minor genetic susceptibility genes to improve prediction of disease in type 1 diabetes. Genes Immun. 2012;13:549-55.

[558] Winkler C, Krumsiek J, Buettner F, Angermuller C, Giannopoulou EZ, Theis FJ, et al. Feature ranking of type 1 diabetes susceptibility genes improves prediction of type 1 diabetes. Diabetologia. 2014;57:2521-9.

[559] Loenen WA, Dryden DT, Raleigh EA, Wilson GG, Murray NE. Highlights of the DNA cutters: a short history of the restriction enzymes. Nucleic Acids Res. 2014;42:3-19.

[560] Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev. 2013;77:53-72.

[561] Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, et al. The Epigenomic Landscape of Prokaryotes. PLoS Genet. 2016;12:e1005854.

[562] Messier S, Lacouture S, Gottschalk M, Groupe de Recherche sur les Maladies Infectieuses d, Porc, Centre de Recherche en Infectiologie P. Distribution of Streptococcus suis capsular types from 2001 to 2007. Can Vet J. 2008;49:461-2.

[563] Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, et al. Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. Nat Genet. 2003;35:32-40.

[564] Clifford RJ, Milillo M, Prestwood J, Quintero R, Zurawski DV, Kwak YI, et al. Detection of bacterial 16S rRNA and identification of four clinically important bacteria by real-time PCR. PLoS One. 2012;7:e48558.

[565] Baele M, Chiers K, Devriese LA, Smith HE, Wisselink HJ, Vaneechoutte M, et al. The gram-positive tonsillar and nasal flora of piglets before and after weaning. J Appl Microbiol. 2001;91:997-1003.

[566] MacInnes JI, Gottschalk M, Lone AG, Metcalf DS, Ojha S, Rosendal T, et al. Prevalence of Actinobacillus pleuropneumoniae, Actinobacillus suis, Haemophilus parasuis, Pasteurella multocida, and Streptococcus suis in representative Ontario swine herds. Can J Vet Res. 2008;72:242-8.

[567] Chengappa MM, Pace LW, Williams JA, Herren CH, Ascher SE. Efficacy of tiamulin against experimentally induced Streptococcus suis type-2 infection in swine. J Am Vet Med Assoc. 1990;197:1467-70.

[568] Sihvonen L, Kurl DN, Salmela P. Infection with Streptococcus suis serotypes 1 and 2 in the same diseased pig. Acta Vet Scand. 1986;27:626-8.

[569] Brisebois LM, Charlebois R, Higgins R, Nadeau M. Prevalence of Streptococcus suis in four to eight week old clinically healthy piglets. Can J Vet Res. 1990;54:174-7.

[570] Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. Cold Spring Harbor Symposia on Quantitative Biology. 1986;LI:263-73.

[571] Lazcka O, Del Campo FJ, Munoz FX. Pathogen detection: a perspective of traditional methods and biosensors. Biosens Bioelectron. 2007;22:1205-17.

[572] Glazunova OO, Raoult D, Roux V. Partial recN gene sequencing: a new tool for identification and phylogeny within the genus Streptococcus. Int J Syst Evol Microbiol. 2010;60:2140-8.

[573] Hatrongjit R, Akeda Y, Hamada S, Gottschalk M, Kerdsin A. Multiplex PCR for identification of six clinically relevant streptococci. J Med Microbiol. 2017.

[574] Davies PR, Ossowicz CJ. Evaluation of methods used for detecting Streptococcus suis type 2 in tonsils, and investigation of the carrier state in pigs. Res Vet Sci. 1991;50:190-4.

[575] Ochert AS, Boulter AW, Birnbaum W, Johnson NW, Teo CG. Inhibitory effect of salivary fluids on PCR: potency and removal. PCR Methods Appl. 1994;3:365-8.

[576] Numminen E, Chewapreecha C, Turner C, Goldblatt D, Nosten F, Bentley SD, et al. Climate induces seasonality in pneumococcal transmission. Sci Rep. 2015;5:11344.

[577] Kennaway CK, Taylor JE, Song CF, Potrzebowski W, Nicholson W, White JH, et al. Structure and operation of the DNA-translocating type I DNA restriction enzymes. Genes Dev. 2012;26:92-104.

[578] Loenen WA, Dryden DT, Raleigh EA, Wilson GG. Type I restriction enzymes and their relatives. Nucleic Acids Res. 2014;42:20-44.

[579] Murray NE. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiol Mol Biol Rev. 2000;64:412-34.

[580] Kneale GG. A symmetrical model for the domain structure of type I DNA methyltransferases. J Mol Biol. 1994;243:1-5.

[581] Sekizaki T, Osaki M, Takamatsu D, Shimoji Y. Distribution of the SsuDAT1I restriction-modification system among different serotypes of Streptococcus suis. J Bacteriol. 2001;183:5436-40.

[582] Sekizaki T, Otani Y, Osaki M, Takamatsu D, Shimoji Y. Evidence for horizontal transfer of SsuDAT1I restriction-modification genes to the Streptococcus suis genome. J Bacteriol. 2001;183:500-11.

[583] Sekizaki T, Takamatsu D, Osaki M, Shimoji Y. Different foreign genes incidentally integrated into the same locus of the Streptococcus suis genome. J Bacteriol. 2005;187:872-83.

[584] Xu B, Zhang P, Li W, Liu R, Tang J, Fan H. hsdS, Belonging to the Type I Restriction-Modification System, Contributes to the Streptococcus suis Serotype 2 Survival Ability in Phagocytes. Front Microbiol. 2017;8:1524.

[585] Johnston C, Martin B, Fichant G, Polard P, Claverys JP. Bacterial transformation: distribution, shared mechanisms and divergent control. Nat Rev Microbiol. 2014;12:181-96.

[586] Wang Y, Yi L, Zhang Z, Fan H, Cheng X, Lu C. Biofilm formation, host-cell adherence, and virulence genes regulation of Streptococcus suis in response to autoinducer-2 signaling. Curr Microbiol. 2014;68:575-80.

[587] Wang Y, Yi L, Wang S, Fan H, Ding C, Mao X, et al. Crystal Structure and Identification of Two Key Amino Acids Involved in AI-2 Production and Biofilm Formation in Streptococcus suis LuxS. PLoS One. 2015;10:e0138826.

[588] Wang Y, Yi L, Wu Z, Shao J, Liu G, Fan H, et al. Comparative proteomic analysis of Streptococcus suis biofilms and planktonic cells that identified biofilm infection-related immunogenic proteins. PLoS One. 2012;7:e33371.

[589] Bello-Orti B, Deslandes V, Tremblay YD, Labrie J, Howell KJ, Tucker AW, et al. Biofilm formation by virulent and non-virulent strains of Haemophilus parasuis. Vet Res. 2014;45:104.

[590] Donlan RM, Costerton JW. Biofilms: survival mechanisms of clinically relevant microorganisms. Clin Microbiol Rev. 2002;15:167-93.

[591] Lunetta KL, Hayward LB, Segal J, Van Eerdewegh P. Screening large-scale association study data: exploiting interactions using random forests. BMC Genet. 2004;5:32.

[592] Melles DC, Gorkink RF, Boelens HA, Snijders SV, Peeters JK, Moorhouse MJ, et al. Natural population dynamics and expansion of pathogenic clones of Staphylococcus aureus. J Clin Invest. 2004;114:1732-40.

[593] Meijerink M, van Hemert S, Taverne N, Wels M, de Vos P, Bron PA, et al. Identification of genetic loci in Lactobacillus plantarum that modulate the immune response of dendritic cells using comparative genome hybridization. PLoS One. 2010;5:e10632.

[594] Bayjanov JR, Starrenburg MJ, van der Sijde MR, Siezen RJ, van Hijum SA. Genotype-phenotype matching analysis of 38 Lactococcus lactis strains using random forest methods. BMC Microbiol. 2013;13:68.

[595] Touw WG, Bayjanov JR, Overmars L, Backus L, Boekhorst J, Wels M, et al. Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle? Brief Bioinform. 2013;14:315-26.

[596] Galina L, Vecht U, Wisselink HJ, Pijoan C. Prevalence of various phenotypes of Streptococcus suis isolated from swine in the U.S.A. based on the presence of muraminidase-released protein and extracellular factor. Can J Vet Res. 1996;60:72-4.

[597] Schultsz C, Jansen E, Keijzers W, Rothkamp A, Duim B, Wagenaar JA, et al. Differences in the population structure of invasive Streptococcus suis strains isolated from pigs and from humans in The Netherlands. PLoS One. 2012;7:e33854.

[598] Willemse N, Howell KJ, Weinert LA, Heuvelink A, Pannekoek Y, Wagenaar JA, et al. An emerging zoonotic clone in the Netherlands provides clues to virulence and zoonotic potential of Streptococcus suis. Sci Rep. 2016;6:28984.

[599] Fatehullah A, Tan SH, Barker N. Organoids as an in vitro model of human development and disease. Nat Cell Biol. 2016;18:246-54.