STRANGLES
The Molecular Identification and Epidemiology of *Streptococcus equi* subsp. *equi*

A thesis

submitted in fulfillment

of the requirements for the degree

of

Master of Science in Biological Sciences

at

The University of Waikato

by

Olivia Patty

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Abstract

A conventional PCR diagnostic test was used to confirm the microbiological isolation of *Streptococcus equi* subsp. *equi* (*S. equi*), the causative agent of strangles. This test was based on the amplification of the *seeI* gene, which is species-specific for *S. equi*. Furthermore, a multiplex PCR was developed using species-specific primers to identify the presence of *S. equi* and two other streptococci known to complicate the diagnosis of strangles in horses, *Streptococcus equi* subsp. *zooepidemicus* (*S. zooepidemicus*) and *Streptococcus dysgalactiae* subsp. *equisimilis* (*S. equisimilis*). A total of 18 clinical isolates of *S. equi* plus the Pinnacle® IN vaccine isolate, two isolates of *S. zooepidemicus* and four isolates of *S. equisimilis* were obtained via culture and used in the development of the multiplex diagnostic tool. Two multiplex tests were trialled, a conventional multiplex PCR and a real-time multiplex PCR.

Both the conventional and real-time multiplex PCR approaches were able to distinguish between the three streptococci and accurately identify all isolates. However, further testing on 26 field specimens revealed that the real-time multiplex PCR had lower specificity, sensitivity and diagnostic accuracy as compared to the conventional multiplex PCR. This was theorised to be the result of the PEG/KOH solution used in the DNA extraction, possibly interfering with the intercalating dye in the real-time reaction. Based on these preliminary results, the conventional multiplex PCR diagnostic test developed here is recommended for further trials to determine its robustness. The test could provide a practical alternative to culture assays for the detection of *S. equi* in clinical specimens.

The 19 *S. equi* isolates obtained, including the vaccine, were further subjected to epidemiological studies. Sequencing of the variable N-terminal region of the antiphagocytic M-protein SeM to determine SeM allele subtypes and a Sau-PCR

amplification method were performed. Two novel strains of *S. equi* were found within NZ based on the variable region of *seM*, SeM alleles 99 and 100. SeM allele 100 had a higher prevalence over allele 99 as it was isolated in 6 out of 9 outbreaks and was found to occur on both the North and South Islands of New Zealand. SeM allele 99 was only found to occur on the North Island. In addition, the Pinnacle® IN vaccine strain, SeM 2 was isolated from lymph node abcesses of two horses. It was unclear as to whether this 'vaccine breakdown' was just a severe adverse reaction to the vaccine or if the vaccine reverted to a more virulent type. The Sau-PCR method, involving digestion of genomic DNA and subsequent amplification, was able to differentiate between the field isolates of *S. equi* and the vaccine strain but was unable to further differentiate between the field isolates. Sau-PCR was therefore determined not as valuable for *S. equi* epidemiological studies over SeM allele subtyping. Sequencing of the SeM gene may allow differentiation or linkage of strangles outbreaks within New Zealand as well as identify vaccine related isolates.

Acknowledgements

Firstly, a huge thank you goes to my supervisor Dr Ray Cursons. Your guidance, advice and encouragement were invaluable. Thank you for enabling this opportunity for me, I am truly grateful.

Thank you to the New Zealand Equine Research Foundation for supporting and funding this project. I am grateful to all the New Zealand veterinarians who provided the specimens making this project possible. Thank you to Rebecca Sutorius for taking horse photos for me to use in my work.

To all my work colleagues, thank you for your continued encouragement throughout this undertaking. A special thanks to Barry for your help with photographing plates and for sparking the initiative leading me to move over to 'the dark side' (as Ray would say). To my office mates, thank you for being there to field questions and providing the much needed office humour.

My dear friends and family, thank you for being interested or at least pretending to be ☺. Who would have thought that I could have fitted a thesis in and around my life; it would not have happened without your support. To Marshall and Azra thank you for putting up with the weekends and evenings that were taken up with this thesis. Bruce, you never waivered in your support and confidence in me; this work would not have been achieved without you.

This thesis is dedicated to my parents, Roy and Sue. Thank you for your support and belief in my abilities throughout all my schooling.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Standard International Unit abbreviations were used throughout this thesis.

BHI	Brain heart infusion
BOX PCR	Box element PCR
bp	Base pairs
BSA	Bovine serum albumin
C3	Complement component 3
CI	Confidence interval
CWA	Cell wall anchor
ddH2O	Double distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double-stranded DNA
EDTA	Ethylenediamine tetra acetic acid
ERIC	Enterobacterial repetitive intergenic consensus
Fb	Fibrinogen
FN	False negative
FP	False positive
GITC	Guanidine thiocyanate
GPNT	Gentamicin, polymyxin B, naladixic acid and thallium acetate
IgG	Immunoglobulin G
MQ-H_2O	Milli-Q (Millipore) filter-purified water (at least 15 megaohms resistance)
NZ	New Zealand
OD	Optical density
PCR	Polymerase chain reaction

PEG/KOH	Polyethylene glycol/potassium hydroxide
PFGE	Pulsed field gel electrophoresis
RAPD-PCR	Randomly amplified polymorphic DNA
rep-PCR	Repetitive element PCR
rcf	Relative centrifugal force
SB	Standardbred
SB buffer	Sodium Borate buffer
SD	Standard deviation
S. equi	*Streptococcus equi* subsp. *equi*
S. equisimilis	*Streptococcus dysgalactiae* subsp. *equisimilis*
SNP	Single nucleotide polymorphism(s)
S. *pneumoniae*	*Streptococcus pneumoniae*
S. pyogenes	*Streptococcus pyogenes*
SS	Signal sequence
ST	Sequence type
S. zooepidemicus	*Streptococcus equi* subsp. *zooepidemicus*
$\mathbf{T^A}$	Annealing temperature
TAE buffer	Tris (hydroxymethyl) amino-methane, acetic acid, EDTA buffer
TE	Tris (hydroxymethyl) amino-methane, EDTA
TB	Thoroughbred
THY	Todd Hewitt plus yeast extract
$\mathbf{T_m}$	Melting temperature
TN	True negative
TP	True positive
v/v	Volume per volume
WB	Warmblood
w/v	Weight per volume

List of Abbreviations for Manufacturers

All items were purchased through New Zealand or Australian Distributers (except GenScript products, which were purchased directly from the USA supplier).

Ajax	Ajax Chemicals, Ltd., Australia
AppliChem	AppliChem GmbH, Germany
APS	Asia Pacific Specialty Chemicals Ltd, Australia
Axygen	Axygen Scientific, Inc., USA
Barnstead	Barnstead International, USA
BIOER	Bioer Technology CO., LTD, China
Bio-Rad	Bio-Rad, NZ
BD	BD Diagnostics, USA
Boehringer	Boehringer Mannheim GmgH, Germany
Corbett	Corbett Research, Australia
Difco	Difco Laboratories Incorporated, USA
Eppendorf	Eppendorf, Germany
Fermentas	Fermentas GmbH, Switzerland
Fort Dodge	Fort Dodge, USA
GenScript	GenScript, USA
IDT	Integrated DNA Technologies, USA
Life Technologies	Life Technologies, NZ
Nanodrop	Nanodrop Technologies, USA
NEB	New England BioLabs, USA
Owl	Owl Separation Systems, USA
Pfizer	Pfizer, NZ
Riedel-de Haën	Riedel-de Haën, Germany
Roche	Roche Diagnostics GmbH, Germany

Scharlau	Scharlau, Spain
Sigma	Sigma-Aldrich Corporation, USA
Solis BioDyne	Solis BioDyne, Estonia
USB	USB Corporation, USA

1. Strangles Overview

1.1. Introduction

Strangles is a highly contagious upper respiratory tract infection of equids caused by the bacterium *Streptococcus equi* subsp. *equi* (*S. equi*). The first description of the disease goes back to 1251 (Timoney 1993; Sweeney, Timoney et al. 2005) and has been referenced to in Europe since the 13th century (Taylor and Wilson 2006). It is now a worldwide problem as one of the most diagnosed infectious diseases in horses, posing major welfare and economic impacts. In the UK during 2008 there were over 700 outbreaks of strangles, some outbreaks involving over 200 horses and having significant associated costs (Waller, Paillot et al. 2011). *S. equi* has been proven to be one of the important pathogens of equine upper respiratory tract infections. In a 24 month surveillance study in the United States during 2008-2010, equids with acute infectious upper respiratory tract disease and/or acute onset of neurological disease were screened for the presence of equine influenza virus (EIV), equine herpes virus-1 and -4 (EHV-1, EHV-4) and *S. equi* (Pusterla, Kass et al. 2011). Out of 201 positive cases for one of the four pathogens, the highest detection rate was for EHV-4 (82 cases), followed by EIV (60 cases), *S equi* subspecies *equi* (49 cases) and EHV-1 (23 cases). Further, *S. equi* was found to occur simultaneously in some cases with EIV, EHV-1, or EHV-4 and in a triple infection with EVH-1 and -4.

Currently, strangles is not a notifiable disease in New Zealand and the economic impact is not documented. A verbal account inferred that an outbreak on a New Zealand breeding property early in 2011 cost a minimum of $100,000 in feed cost, lost clients, and increased staffing. A cross-sectional survey of biosecurity

practices on thoroughbred stud farms in New Zealand by Rogers and Cogger (2010) during the 2006/2007 breeding seasons raised concerns over the lack of good biosecurity practices leaving farms open to endemic diseases such as strangles. The fear of a strangles outbreak is very real within the New Zealand horse industry.

1.2. *Streptococcus equi* subsp. *equi*

The *Equidae* host-adapted parasite, *S. equi* is the causative agent of strangles. It is a Gram-positive, β-haemolytic *Streptococcus* belonging to lancefield group C. Typically it is highly encapsulated, forming large mucoid colonies with a wide zone of haemolysis on blood agar. *S. equi* is separated from other group C streptococci by an inability to ferment lactose, sorbitol and trehalose (Grant, Efstratiou et al. 1993; Efstratiou, Colman et al. 1994; Holden, Heather et al. 2009). Its genome consists of a circular chromosome of approximately 2.25 million bps encoding over 2000 predicted coding sequences (Holden, Heather et al. 2009).

1.2.1. *S. equi* Virulence Factors

The following description of *S. equi* virulence factors is based on reviews written by Harrington et al. (2002) and Timoney (2004). The three categories for virulence stated by Harrington et al. (2002) will be discussed: adherence, immune evasion and nutrient acquisition.

The ability of adhering to host tissue increases an organism's chance of survival and success. Fibronectin-binding proteins mediate attachment of host cells; a fibronectin-binding protein in *Streptococcus equi* subsp. *zooepidemicus* (*S. zooepidemicus*), the archeal type of *S. equi*, was also found in *S. equi*. However

the *S. equi* version is lacking the C-terminal anchor and may not be functional. Fibrinogen (Fb), an abundant plasma protein, is also a site for adherence. *Streptococcus pyogenes* (*S. pyogens)* has an M protein that is involved in Fb binding. Similarly, *S. equi* has two M-like proteins; SeM is unique to *S. equi* and SzPSe, a homologue to a *S. zooepidemicus* M-like protein. Both show strong binding to equine Fb. Also, the hyaluronic acid capsule of *S. zooepidemicus* is significant in the role of adherence to its host and it is suggested that the mucoid hyaluronic acid capsule of *S. equi* could have a similar role.

Pathogenic bacteria need to be able to evade the immune challenges of their host. Again, the mucoid capsule has been implicated in this role. The capsule reduces the chance of recognition and subsequently phagocytosis. M-like proteins are involved not only in binding Fb but also have antiphagocytic activity. Of the two M-like proteins in *S. equi*, SeM plays a more dominate role in resisting phagocytosis due to its binding of both Fb and immunoglobulin G (IgG) which interfere with immune related proteins binding to the bacteria, decreasing the ability of the hosts immune system to recognise it for phagocytosis. Interestingly, a deletion in the SeM gene was associated with reduced virulence and increased persistence in hosts. *S. equi* also releases a toxin, which damages host immune cells making them unable to phagocytose the bacteria. Another method of survival against the host defences is by encoding superoxide dismutase that is involved in converting toxic host reactive oxygen intermediates into less toxic hydrogen peroxide and water. The one found in *S. equi* is *sodA* and has been suggested to have a similar role. Further, it has four known pyrogenic mitogens (superantigens). They bind to Major histocompatibility complex molecules and T-cells simultaneously resulting in non-specific T-cell stimulation, proliferation, and proinflammatory cytokine release causing the acute phase of strangles.

Finally, pathogenic bacteria need to acquire nutrients from their host to survive. *S. equi* produces two cell-associated acid phosphatase, which are thought to participate in nutrient acquisition. It also produces degradative enzymes used in the utilisation of peptide and carbohydrate substrates by degrading protein substrates such as Fb, gelatin and casein. Streptolysin S is an oligopeptide

responsible for β-haemolysis; the lysing of red blood cells makes essential nutrients such as iron available to the bacteria.

1.3. Other Streptococci Horse Pathogens

Streptococci are Gram-positive, catalase-negative cocci that grow in pairs or chains and are non-motile or spore forming, which vary in their pathogenic nature. Besides *S. equi*, there are three other pathogenic streptococci in equids, the β-haemolytic *S. zooepidemicus* and *Streptococcus dysgalactiae* subsp. *equisimilis* (*S. equisimilis*) as well as the α-haemolytic *Streptococcus pneumoniae* (S. *pneumoniae*) (Timoney 2004).

S. zooepidemicus is a mucosal commensal of a wide range of animals, including humans. This zoonotic pathogen is opportunistic and causes disease when its host is under situations of stress such as virus infection, heat stress or tissue injury (Timoney 2004). In horses, *S. zooepidemicus* has been associated with respiratory disease, foal pneumonia, endometritis and abortion (Chanter, Collin et al. 1997; Laus, Preziuso et al. 2007; Proietti, Bietta et al. 2011). It has been isolated from equine joints, lymph nodes, nasal cavities, lungs and uteruses (Timoney 1993; Proietti, Bietta et al. 2011). The bacterium *S. equisimilis* normally inhabits skin and mucous membranes of humans and animals. Infection in horses is thought to be infrequent, opportunistic and is poorly understood (Timoney 2004). It has been isolated in horses from aborted placentas, uteruses, lymph nodes and nasopharyngeal swabs including ones from strangle-like disease (Timoney 2004; Laus, Preziuso et al. 2007; Preziuso, Laus et al. 2010; Proietti, Bietta et al. 2011). The bacterium *S. pneumoniae* is an important human pathogen associated with several diseases including pneumoniae. It has been linked to lower airway disease in horses including pneumonia in a neonatal foal (Timoney 2004).

S. equi, *S. zooepidemicus*, and *S. equisimilis* are all Lancefield group C streptococci; yet, *S. equisimilis* is genetically distinct from the other two species

(Timoney 2004). *S. zooepidemicus* is thought to be the archetype of *S. equi* based on multilocus enzyme electrophoresis (Jorm, Love et al. 1994) and the composition of 16S-23S RNA gene intergenic spacer sequences (Chanter, Collin et al. 1997). *S. equi* and *S. zooepidemicus* share greater than 98 % DNA homology. Through comparative genomics, Holden et al. (2009) believe that *S. equi* has passed through a genetic bottleneck evolving to the *Equidea* host-restricted pathogen from an ancestral *S. zooepidemicus* strain. They found evidence of functional loss due to mutation and deletion, and acquisition of new genes via mobile genetic elements. They proposed that the key speciation event was the acquisition of a novel system involved in the uptake of iron, hypothesizing that a more efficient uptake of iron would enhance the ability of *S. equi* to generate lymph node abscession critical to establishing carriers and therefore the worldwide success of *S. equi*. Strangles has been referred to as 'strep throat of horses', which makes another finding in their study very interesting; *S. equi* and *S. zooepidemicus* have about 80 % DNA homology with *S. pyogenes,* a human pathogen that causes several diseases including 'strep throat'. They demonstrated that these three organisms share a common phage pool allowing for cross-species evolution and influence their pathogenicity.

1.4. Strangles

1.4.1. Clinical Signs & Conditions

The first clinical signs appear 3-14 days after exposure with a fever reaching 103-106°F (approximately 39 – 41°C) (Sweeney, Timoney et al. 2005; Taylor and Wilson 2006). Within 24hrs of the fever onset, a nasal discharge will develop which becomes mucopurulent and the horse can become anorexic, depressed and some may have a moist cough (Taylor and Wilson 2006). The submandibular lymph nodes usually abscess becoming enlarged, firm and painful; the retropharyngeal lymph nodes may also be affected (Taylor and Wilson 2006). The abscessed lymph nodes will mature and rupture 7-10 days after disease onset that then leads to recovery of the horse within 1-2 weeks (Taylor and Wilson 2006). The abscesses can become large and painful enough to cause difficultly in breathing and swallowing resulting in the horse standing with an outstretched neck to ease the pressure (Taylor and Wilson 2006). Figure 1.1 shows a New Zealand horse displaying clinical strangles.

The majority of horses will fully recover within 4-6 weeks (Waller and Jolley 2007). The severity can vary greatly depending on their level of immunity, with younger horses generally experiencing more severe clinical signs for a longer duration and older horses experiencing a milder form of the disease over a shorter duration (Sweeney, Timoney et al. 2005; Taylor and Wilson 2006). Seventy five percent of horses will develop immunity after recovery, which can last up to 5 years; the other 25 % can become susceptible to a second attack within several months (Timoney 2004; Sweeney, Timoney et al. 2005). This acquired resistance appears to be mediated by mucosal and systemic antibodies to SeM and other antigens unique to *S. equi* (Timoney 1993; Sweeney, Timoney et al. 2005). Suckling foals benefit from the milk of recovering mares and are usually resistant until weaned (Timoney 2004; Taylor and Wilson 2006).

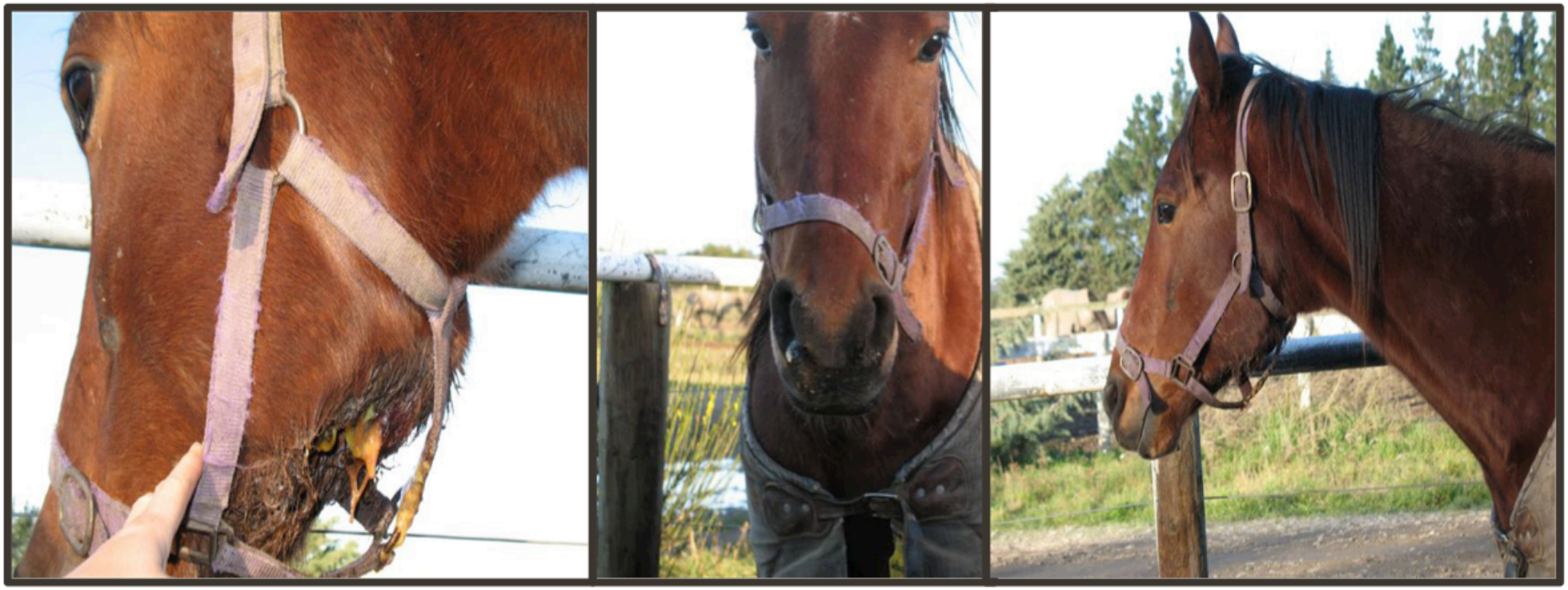


Figure 1.1: Photos supplied by New Zealand veterinarian Rebecca Sutorius showing a horse displaying clinical strangles: discharging abscesses of the submandibular lymph nodes, purulent nasal discharge, and swollen lymph nodes.

Complications occur in approximately 20 % of strangles cases and of those, 40 % may die or be euthanized (Sweeney, Timoney et al. 2005; Taylor and Wilson 2006). The review by Sweeney et al. (2005) has listed the complications with the most important being associated with metastatic spread of infection and immune-mediated complications. In metastatic spread of infection, *S. equi* can spread via the circulatory system, lymphatic channels, or by close association. Metastatic abscessation is referred to as 'bastard strangles'; common sites being the lung, mesentery, liver, spleen, kidneys and brain. Symptoms usually consists of weight loss, colic, persistent or intermittent fever, depression and loss of appetite (Taylor and Wilson 2006). Empyema of the guttural pouch is the most common complication involving the spread of infection with symptoms of chronic mucopurulent nasal discharge and difficulty breathing (Taylor and Wilson 2006). The guttural pouch is a pair of air-filled chambers off the Eustachian tube (Figure 1.2). Purpura hemorrhagica is an immune-mediated complication characterized by limb and ventral endema and petechial or ecchymotic hemorrhagica (Sweeney, Timoney et al. 2005; Taylor and Wilson 2006). It is usually associated with depression, reluctance to move, fever, tachycardia, colic or diarrhoea (Taylor and Wilson 2006). Myopathies, also thought to be immune-mediated, have been associated with strangles infection (Sweeney, Timoney et al. 2005). Pneumonia is an important secondary infection to strangles and is the most common complication resulting in death (Sweeney, Timoney et al. 2005).

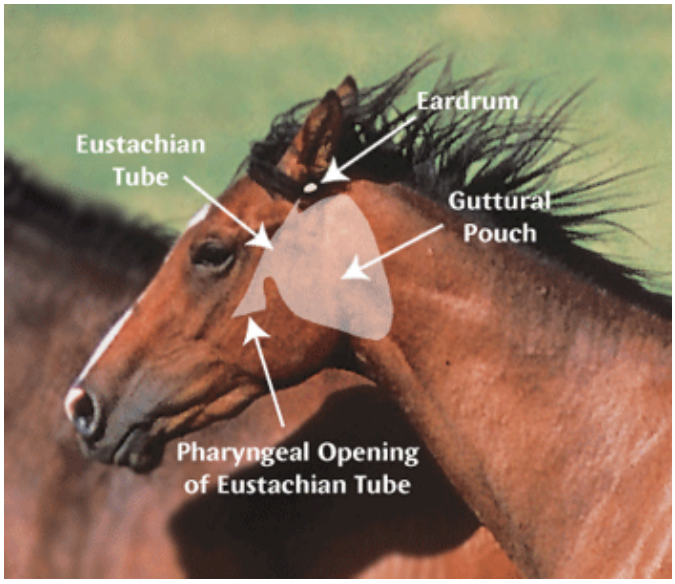


Figure 1.2: Guttural pouch location. Image used with permission from MSD Animal Health (http://www.equine-strangles.co.uk/Transmission.asp)

1.4.2. Pathogenesis

S. equi enters via the mouth or nose by ingestion or inhalation where it then attaches to tonsil cells and adjacent lymphoid nodules (Timoney 2004; Taylor and Wilson 2006). Within hours of initial infection, it is translocated to the submandibular and retropharyngeal lymph nodes that drain the pharyngeal/tonsillar region (Timoney 2004; Taylor and Wilson 2006), where accumulation causing abscesses. Disposal of the bacterium occurs via lysis of mature abscesses and evacuation of the contents through the skin or nasal passage (Timoney 2004). Metastasis of the infection can occur spreading throughout the body called 'bastard strangles'. The guttural pouch may become infected during the early stages and harbour infection with intermittent shedding after clinical recovery (Timoney 2004; Sweeney, Timoney et al. 2005).

1.4.3. Transmission & Environmental Persistence

Morbidity of strangles infections can be as high as 85-100 % in susceptible populations influenced by factors such as overcrowding, mixing from different areas, and stress (Taylor and Wilson 2006). An outbreak can last up to 4-6 months in a highly susceptible population under poor management procedures such as isolation (Taylor and Wilson 2006). Infected purulent discharge is transmitted either directly by head to head contact or indirectly by formites such as feeders, water buckets, tack, flies and people (Sweeney, Timoney et al. 2005; Taylor and Wilson 2006). Nasal shedding occurs 2-3 days after the onset of fever and persists anywhere from 2-6 weeks after clinical recovery with the exception of chronic carriers (Timoney 2004; Sweeney, Timoney et al. 2005; Taylor and Wilson 2006).

Jorm (1992) conducted a laboratory study on the persistence of *S. equi* on surfaces. It was able to survive up to 63 days on wood at 2 °C and 48 days at 20 °C. However, this study was done on sterile surfaces in a controlled, stable environment not reflecting the changing natural environment such as temperature, humidity, rainfall and sunlight. A second study by Weese et al. (2009) was conducted in an outdoor environment incorporating two different types of inoculums; one in phosphate buffered saline and the other in mucus from the upper respiratory tract of horses. The conditions of this study better reflected the natural environment including competing microflora from the environment and horse normal flora. Survival of *S. equi* in this study was very short, <1-3 days. There was a significant effect of sunlight on the persistence of *S. equi* ($P = 0.002$) indicating the susceptibility of the organism to ultraviolet rays and desiccation. There was no effect of rain or surface type, but bacterium's survival was significantly longer in mucus inoculum than saline ($P = 0.024$) when >24 hrs was used as the cutoff; mucus provides nutrition and protection from the physical environment. The authors warn that these results are only applicable to the conditions of the study and that survivability is likely to be greater in shady areas, or within soil or grass where there is little or no sunlight and lower temperatures.

It can be implied from these two studies, that barns and stables pose a higher risk in the transmission of *S. equi* than pastures especially in the summer due to ultraviolet exposure.

1.4.4. Chronic Carriers

Chronic carriage, a subclinical infection, is believed to be due to the harbouring of *S. equi* in the guttural pouch (Figure 1.2) and greatly influences the survival and spread of this disease. As described by Waller and Jolley (2007), incomplete drainage of the bacterium from the guttural pouches (and/or sinuses) after the rupture of abscesses can lead to chronic carriers. This seems to occur by the drying and hardening of the bacterium leading to the formation of discrete bodies called chondroids (Figure 1.3). These can remain in the guttural pouch for several years with intermittent shedding of the bacterium. One study found prolonged shedding up to 39 months after the resolution of clinical signs (Newton, Wood et al. 1997). It is estimated that about 10 % of infected horses become carriers (Sweeney, Timoney et al. 2005; Taylor and Wilson 2006; Waller and Jolley 2007). Another study indicated that infection persisted in 9-44 % of horses in their study after clinical signs had disappeared (Newton, Verheyen et al. 2000). Carriers are the major contributors to outbreaks in naïve horses and are a threat to the control and eradication of Strangles. Waller (2011) suitably states: "These persistent infections in 'carrier' horses are likely to have been critical to the global success of *S. equi* because intermittent shedding from the guttural pouches of otherwise normal horses contributes to the interepizootic maintenance of the disease" (p. 8).

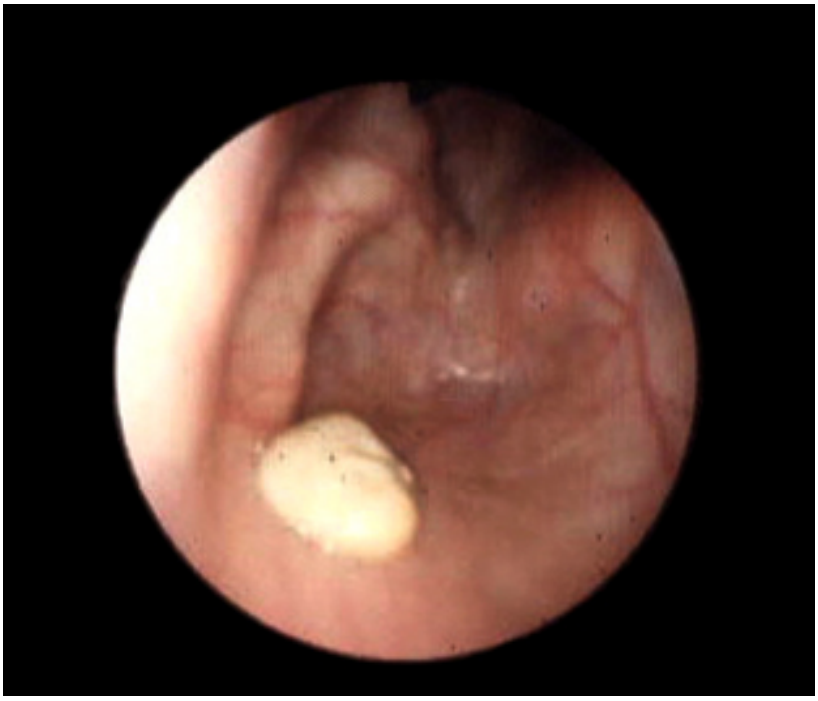


Figure 1.3: A chondroid identified by guttural pouch endoscopy (Waller and Jolley 2007). Image used with permission from copyright holder Elsevier.

1.4.5. Treatment

Generally, the stage and severity of the disease will determine the type of treatment required with the majority of cases requiring no treatment and letting the disease run its course. The use of antimicrobials in the early clinical stage of infection can reduce recovery time, reduce lymph node abscession and prevent complications (Sweeney, Timoney et al. 2005; Taylor and Wilson 2006). However, once lymph nodes are abscessed, antimicrobials may slow the progression of the disease by delaying the maturation of the abscesses. Treatment at this stage should be directed towards speeding abscess maturation (Sweeney, Timoney et al. 2005; Taylor and Wilson 2006). In severely affected horses, antimicrobials are recommended (Taylor and Wilson 2006). Carriers should be treated with topical and systemic antimicrobials and in the presence of guttural pouch empyema, pouches should be lavaged daily with isotonic saline or

polyionic fluids; the addition of 20 % acetylcysteine may help fragment and liquefy chondroids (Sweeney, Timoney et al. 2005; Taylor and Wilson 2006). Complications from strangles should receive symptomatic therapy as discussed by Sweeney et al. (2005).

1.5. Vaccines

Currently, there are two vaccines for strangles available in New Zealand. One is an inactivated bacterin, Equivac® S (Pfizer), used intramuscularly. It is also available with a tetanus vaccine, Equivac® 2 in 1. The other is an avirulent, non-encapsulated, attenuated mutant strain of *S. equi*, Pinnacle® IN (Fort Dodge), used as an intranasal vaccine. Pinnacle® IN is known to have adverse side affects and has not been licensed for sale in Europe due to safety concerns (Guss, Flock et al. 2009). Its small dry colony phenotype can revert to a mucoid phenotype, which cannot be distinguished from the wild type strains (Walker and Timoney 2002). Borst (2011) evaluated Pinnacle IN® in ponies. Their results indicated that this vaccine is not safe for use in <1-year-old naïve ponies, causing substantial morbidity and some deaths. However, they mentioned that the modified-live vaccine could be safely and successfully administered to adult ponies with low background antibody titers against *S. equi* resulting in increased antibody titers. Progress towards safe and efficient vaccines is still being made (Walker and Timoney 2002; Flock, Karlström et al. 2006; Waller and Jolley 2007; Guss, Flock et al. 2009).

2. Development of Diagnostic Multiplex PCR

2.1. Introduction

The high morbidity rate and major welfare and economic impacts of strangles require a diagnostic for this disease that is reliable and rapid. In New Zealand, diagnosis of strangles is currently based on culture (serological and biochemical identification). This typically takes ≥2 days and was considered the 'gold standard' for confirmation of clinical diagnoses. However, the presence of other β-haemolytic streptococci may make it difficult to detect *S. equi*, such as *S. zooepidemicus* and *S. equisimilis*. They can complicate a strangles diagnoses that is based on colony morphology and fermentation properties. Phenotypically, *S. equi* forms mucoid colonies, yet; although colonies from *S. zooepidemicus* are typically nonmucoid, they can sometimes take on a mucoid formation appearing more like *S. equi* (Sweeney, Timoney et al. 2005). The mucoid appearance of *S. equi* cannot always be relied upon as a less virulent isolate produces a matte appearance (Taylor and Wilson 2006). Moreover, atypical *S. equi* isolates can occur and can be confused with *S. equisimilis* or *S. zooepidemicu*s on the basis of sugar fermentation (Grant, Efstratiou et al. 1993).

Further to this complication, *S. equisimilis* has been identified as the causative agent in cases of strangles-like diseases in the absence of *S. equi* (Timoney 2004; Laus, Preziuso et al. 2007). And *S. zooepidemicus* is a mucosal commensal, opportunistic bacterium that can result in a serious pathogen of the respiratory

tract (Timoney 2004). Therefore, *S. equi*, *S. zooepidemicus* and *S. equisimilis* can all be important pathogens of the upper respiratory tract.

With the need for a rapid and more reliable method, diagnosis has moved to molecular techniques. Several studies have now been involving PCR (polymerase chain reaction) as a diagnostic tool for upper respiratory tract infections of horses (Alber, El-Sayed et al. 2004; Anzai, Hobo et al. 2006; Laus, Preziuso et al. 2007; Jannatabadi, Mohammadi et al. 2008; Preziuso, Laus et al. 2010; Pusterla, Kass et al. 2011). A study by Laus et al. (2007) investigating chronic upper respiratory diseases in horses found that PCR was more sensitive in detecting *S. equisimilis* and *S. zooepidemicus* then bacteriological results. PCR is also about three times more sensitive than culture in detecting *S. equi* which is suggested to be due to initial low numbers of the organism, slow growth and/or overgrowth of contaminants (Timoney and Artiushin 1997). Båverud et al. (2007) developed a real-time PCR to detect and differentiate *S. equi* and *S. zooepidemicus* based on the Alber et al. (2004) multiplex PCR. This method was also more sensitive then cultivation for detecting the two *Streptococcus* species. Pusterla (2011) also used real-time PCR to identify the pathogen of upper respiratory tract infections, being either viral pathogens or *S. equi*.

Asymptomatic carriers of the *S. equi* infection further complicate matters. One study found that the detection of *S. equi* via culture had a sensitivity of 45 % from nasopharyngeal swabs and 88 % from guttural pouch lavages (Newton, Wood et al. 1997). The detection sensitivity from the culture of the swab increased to 85 % for three consecutive swabs; the variance is a result of intermittent shedding of the organism. They later showed that PCR in combination with culture of nasopharyngeal swabs increases the detection of asymptomatic carriers of *S. equi* (Newton, Verheyen et al. 2000). Smith (2006) describes a method for the elimination of asymptomatic carriers without culture and using only PCR for detecting *S. equi*. He indicated that even though PCR alone may result in unnecessary treatment of horses due to detection of dead bacteria, the cost of combined culture and PCR tests exceeded that of treatment administered by farm personnel.

2.1.1. Hypotheses

The objective of this study was to develop a PCR based multiplex that can rapidly and reliably detect *S. equi* and distinguish it from the two other pathogens which have been know to make diagnoses difficult, *S. zooepidemicus* and *S. equisimilis* as well as have the potential to detect asymptomatic carriers. In order to achieve this two hypothesis were proposed.

Hypothesis 1:
S. equi could be isolated from swabs and differentiated from *S. zooepidemicus* and *S. equisimilis* using molecular and biochemical methods.

Hypothesis 2:
Using specimen swabs, a conventional multiplex PCR and a real-time multiplex PCR could be developed to detect and distinguish between *S. equi*, *S. zooepidemicus*, and *S. equisimilis*.

2.2. Methods

2.2.1. Specimens

Collaborating veterinarians submitted specimens taken from areas known to harbour the bacterium: nasal passages, abscesses and/or guttural pouches. These consisted of nasal swabs, guttural pouch washes and abscess aspirates or swabs. Specimens were taken from symptomatic horses, horses in contact with symptomatic horses and from horses requiring *S. equi* screening for carriage or prior to overseas shipping. Some culture positive specimens from Gribbles Veterinary Pathology facilities within New Zealand were also submitted. The specimens were received during the running of this project between October 2010 and February 2012. As a reference control the *S. equi* live culture vaccine, Pinnacle® IN (Fort Dodge), was received in October 2010 and prepared as to manufacturer's recommendations.

2.2.2. Specimen Processing and DNA Isolation

Throughout the project, various DNA isolation methods were trailed to ascertain the most efficient method. These ranged from extracting directly from the specimen with a rapid high alkaline boil; or a lysis, chloroform, isopropanol precipitation method; to DNA purification from an overnight culture of the provided specimen sample. Reported here is the final method deemed most efficient based on ease and time required, and was used on the final specimens received.

Swabs were initially immersed into 300 μl of sterile saline (0.85 % *w/v*) in a 1.7 ml microfuge tube (Axygen) and mixed by vortex. To ensure all the specimen was removed from the swab, a pair of flame-sterilized forceps was used to ring out and

place the swab into a 0.6 ml microfuge tube (Axygen) with a hole pierced in the bottom (using a 19 gauge hot needle) and a 3 mm sterile glass bead (Ajax) covering the hole. This tube was then placed into another 1.7 ml microfuge tube and centrifuged for 2 mins at 16,000 rcf. The collected fluid was then combined with the original saline mixture and vortexed to ensure thorough mixing of the sample. Ten to 50 µl was streaked onto blood agar (10 % sheep blood, Life Technologies and Columbia Blood Agar, Difco) and incubated at 37 °C overnight. The remaining sample was then centrifuged at 16,000 rcf. After removal of the supernatant, the pellet was resuspended in 200 µl PEG/KOH (Chomczynski and Rymaszewski 2006) (Appendix A.I), vortexed and heated at 95 °C for 15 mins in a thermomixer (Eppendorf) to allow for cell lysis. The lysate was finally centrifuged for 1 min at 16,000 rcf to pellet the cellular debris; 1 µl of the supernatant was used for PCR analysis.

Abscess aspirates and guttural pouch washes were treated similarly. The vial of obtained fluid was vortexed, 1.5 ml extracted and centrifuged at 16,000 rcf for 1 min. The resulting pellet was resuspended in 200 µl of PEG/KOH, heated at 95 °C for 15 mins to lyse the bacterial cells. The mixture was then centrifuged for 1 min at 16,000 rcf to pellet the cellular debris and 1 µl of the supernatant was used for PCR analysis. The remaining specimen not used for DNA extraction was used to culture on blood agar plates.

Due to the difficulty of isolating *S. equi* from specimen samples on blood agar plates because of normal flora contaminates, an alternative media was later used; GPNT selective media (see section 2.2.5 and Appendix A.I). Specimens were either streaked out on selective blood agar or initially grown overnight in selective broth.

2.2.3. Uniplex PCR analysis of Specimens

Primers used in this study are listed in appendix (A.II). All specimens were tested for the presence of *S. equi* using the species-specific *seM* and/or *seeI* primer pairs previously published (Alber, El-Sayed et al. 2004; Kelly, Bugg et al. 2006). A portion of the specimens was also tested for the presence of *S. zooepidemicus* using the *srtz* primer pair (present study) and for *S. equisimilis* using the previously published *eqsim* primer pair (Preziuso, Laus et al. 2010). The *srtz* primer was designed using Primer 3 (Rozen and Skaletsky 2000) based on a putative sortase gene *srtC2* (SZO 1827). This is specific to *S. zooepidemicus* and may have a role in host cell interactions as described by Holden et al. (2009). All primers were synthesized by IDT and reconstituted in TE buffer to 200 µM according to the concentration given on the information sheet received with the primer. Reconstituted primers were stored at -20°C until required.

PCR reactions were carried out on either a LifePro Thermal Cycler (BIOER) or a DNA Engine PTC-200 (Bio-Rad). Each reaction consisted of 1 x Buffer B1 (Solis BioDyne), 2 mM $MgCl_2$ (Solis BioDyne), 0.25 mM of each dNTP (deoxyribonucleotide triphosphates; dATP, dCTP, dGTP, dTTP; Solis BioDyne), 0.5 U Hot FIREPol® DNA Polymerase (Solis BioDyne) and 0.2 µM of the species-specific primer pair. The amplification protocol for the uniplex PCR consisted of an initial activation/denaturation at 95 °C for 15 mins followed by 40 cycles of denaturing at 95 °C for 20 secs, annealing at 53 °C for 20 secs and elongating at 68 °C for 30 secs. A final elongation was done at 68 °C for 5 mins. Ten µl of the PCR products were electrophoresed in a 1.0 % (*w/v*) agarose gel in SB buffer (Appendix A.I) and 1.5 µg/ml of ethidium bromide (USB) using an Owl gel electrophoresis system.

2.2.4. Acquisition of *Streptococcus* Species

The bacteria were attempted to be cultured and isolated, for use in developing a diagnostic multiplex PCR, from any of the *S. equi*, *S. zooepidemicus* and *S. equisimilis* PCR positive samples. This was achieved by screening any β-haemolytic colonies from the original sample blood agar plates using two methods: PCR followed by sequencing and biochemical analysis.

2.2.4.1. PCR and sequencing for species confirmation

Selected β-haemolytic colonies were grown overnight in BHI (Brain Heart Infusion, Difco) media at 37 °C. Bacterial cells were collected by centrifuging 1.5 ml of the overnight culture at 16,000 rcf for 1 min. The supernatant was removed and the cells resuspended in 200 µl PEG/KOH. Then the cellular suspension was incubated at 95 °C for 15 mins and centrifuged at 16,000 rcf for 1 min to pellet cellular debris. One µl of the supernatant was used in a uniplex PCR using a species-specific primer: *seM*, *seeI*, *strz* or *eqsim* (Appendix A.I). The amplification protocol consisted of an initial activation/denaturation at 95 °C for 15 mins followed by 40 cycles of denaturing at 95 °C for 20 secs, annealing at 53 °C for 20 secs and elongating at 68 °C for 30 secs. A final elongation was done at 68 °C for 5 mins. Following amplification, the PCR products were purified for sequencing by adding 1 U of rAPid Alkaline Phosphatase (Roche) and 10 U exonuclease1 (Fermentas) to each 20 µl reaction, incubating at 37 °C for 30 mins followed by 85 °C for 15 mins. Sequencing of both strands with the primers used in the initial PCR amplification was carried out by the Waikato DNA Sequencing Facility (University of Waikato, Hamilton, New Zealand) using BigDye fluorescent terminators (Applied Biosystems Big Dye v 3.1.).

Sequence results of each isolate were analyzed by aligning and manually editing the forward and reverse sequences using Geneious v5.4 (Drummond, Ashton et al. 2011). The resulting high quality double stranded sequence data was entered into the NCBI website in FASTA format to facilitate a BLAST search

(http://blast.ncbi.nlm.nih.gov/Blast.cgi) which allowed confirmation that the primers were amplifying the correct region. Using species-specific primers, sequencing analysis also allowed for species confirmation using the microbial DNA database at NCBI.

2.2.4.2. Biochemical confirmation of *Streptococcus* species

Any isolate confirmed by PCR as *S. equi*, *S. zooepidemicus* or *S. equisimilis* were also confirmed through testing their fermentation abilities on Columbia blood agar supplemented containing pH indicators with 0.25 % (*w/v*) sorbitol or 0.025 % (*w/v*) trehalose (Appendix A.I). Isolates were streaked onto both a sorbital plate and a trehalose plate. The plates were incubated at 37 °C overnight and fermentation results recorded. On the purple sorbital plate a colour change to yellow indicated fermentation. The pale blue trehalose plate results in a colour change to dark blue if fermentation occurred.

2.2.5. Streptococci selective media

Due to difficulties in isolating *S. equi* from clinical samples on 10 % blood agar plates, a streptococci selective media was trialled both as broth and agar plates. Confirmed isolates of *S. equi*, *S. zooepidemicus*, and *S. equisimilis* were used for this trial to ensure that all three species of interest would grow in the selective media. The Pinnacle® IN vaccine strain was also used in this trial.

Two types of broth were trialled: BHI (Difco) and THY (Todd Hewitt Broth (Difco), with an addition of 1 % yeast extract). Both were supplemented with GPNT (2.5 μg/ml, gentamicin; 10 μg/ml, polymyxin B; 10 μg/ml, naladixic acid; and 1 mg/ml, thallium acetate) (Appendix A.I). A single colony of a representative of each species; *S. equi*, *S. zooepidemicus*, *S. equisimilis*, and the Pinnacle® In vaccine strain, were grown in BHI/GPNT and THY/GPNT at 37 °C.

The optical density at 600 nm (OD_{600}) was measured at 24 hrs and 48 hrs (SmartSpec™ 3000, Bio-Rad). The standard t-test was used to compare which media the isolates grew better in based on the OD_{600} measurements, with a $P < 0.05$ considered a significant difference in growth.

Blood agar plates were made with 10 % sheep blood and Columbia blood agar (Difco), supplemented as in the broth; GPNT/blood agar (Appendix A.I). All isolates of *S. equi*, *S. zooepidemicus*, and *S. equisimilis* obtained from the clinical samples, including the Pinnacle® IN vaccine strain; were streaked out onto GPNT/blood agar plates. They were grown overnight at 37 °C and any growth was recorded as a positive result.

2.2.6. Conventional Multiplex PCR Development

Using the Pinnacle® IN vaccine strain as a control and other confirmed isolates of *S. equi*, *S. zooepidemicus* or *S. equisimilis* species, a multiplex diagnostic PCR was developed. For consistency and comparison abilities, a known concentration of DNA was required. To achieve this, DNA from these isolates were extracted and purified using a later mentioned method, the GITC/column method (see sections 3.2.2 and 3.2.3 for a description) and between 40 and 50 ng of DNA was used for each species as the PCR template. To ensure that all products of the multiplex could be amplified in the same reaction, DNA from all three species was used as the template DNA in the multiplex reaction. To ensure that the multiplex was able to identify each isolate correctly, DNA from each isolate was used individually in a separate multiplex reaction.

The primer pairs used were *seeI*, *srtz*, *eqsim* and 16S, as reported in Appendix A.II. *S. equi* was identified by amplification with the *seeI* primers, producing a 520 bp band. *S. zooepidemicus* was identified by amplification with the *srtz* primers, producing a 217 bp product. *S. equisimilis* was identified by amplification with the *eqsim* primers, producing a 279 bp band. The 16S rDNA

primer pair was included as an internal control in order to determine if any PCR inhibitors were present resulting in a false negative as it is a common sequence in all bacteria. All reactions should amplify the 410 bp 16S rDNA product. The mix of these primers makes this multiplex a tetraplex, amplifying up to four products.

For optimization of the multiplex, the modified version of Taguchi was used as described in Cobb and Clarkson (1994). This method arranges the components, which are thought to affect the process; into an orthogonal array with each factor occurring at one of three predetermined levels (Table 2.1). The eight components chosen, which would have an effect on the yield of each multiplex product, were the annealing temperature (T^A) and the final concentrations of Taq polymerase, $MgCl_2$, dNTPs, and the four primer pairs. The three probable levels chosen for each component are as follows: T^A at 52 °C, 55 °C and 60 °C; Taq polymerase at 0.5 U, 0.75 U and 1 U; $MgCl_2$ at 2 mM, 3 mM and 4 mM; dNTPs at 0.2 mM, 0.25 mM and 0.3 mM; primer pairs at 0.15 μM, 0.2 μM and 0.3 μM (Table 2.5). Reactions were carried out in a final volume of 20 μl, made up in 1 x Buffer B1 (Solis BioDyne) and MQ-H_2O (Barnstead). The Taq polymerase (Hot FIREPol® DNA Polymerase), $MgCl_2$, and dNTPs were products of Solis BioDyne. All reactions were carried out in a LifePro Thermal Cycler (BIOER).

Table 2.1 PCR orthogonal array for 8 variable factors at 3 levels as proposed by Cobb and Clarkson (1994).

Reaction	Components for Optimisation (at 3 probable levels A, B, C)							
	1	2	3	4	5	6	7	8
1	A	A	A	A	A	A	A	A
2	A	A	B	B	B	B	B	B
3	A	A	C	C	C	C	C	C
4	A	B	A	A	B	B	C	C
5	A	B	B	B	C	C	A	A
6	A	B	C	C	A	A	B	B
7	B	C	A	B	A	C	B	C
8	B	C	B	C	B	A	C	A
9	B	C	C	A	C	B	A	B
10	B	A	A	C	C	B	B	A
11	B	A	B	A	A	C	C	B
12	B	A	C	B	B	A	A	C
13	C	B	A	B	C	A	C	B
14	C	B	B	C	A	B	A	C
15	C	B	C	A	B	C	B	A
16	C	C	A	C	B	C	A	B
17	C	C	B	A	C	A	B	C
18	C	C	C	B	A	B	C	A

The PCR parameters consisted of an initial activation/denaturation at 95 °C for 15 mins followed by 40 cycles of denaturation at 95 °C for 15 secs, annealing at 52, 55 or 60 °C for 15 secs, and extending at 72 °C for 20 secs; ending with a final extension at 72 °C for 5 mins. A touchdown protocol was later employed to reduce the potential of nonspecific amplification (Don, Cox et al. 1991), starting with a T^A 10 °C higher then the optimal/final T^A and then decreasing by 1 °C every cycle until the optimal T^A was reached. This was followed by an additional 35 cycles of denaturing, annealing and extending at the parameters stated above; keeping the T^A constant at the determined optimal temperature.

Hot FIREPol® DNA Polymerase (Solis BioDyne) and Hot FIREPol® Blend Master Mix (Solis BioDyne) were also compared using the final optimized components. The blend master mix comes pre-made so the final concentration of

$MgCl_2$ is at 1.5, 2, 2.5 or 3mM per reaction. The 3 mM $MgCl_2$ Hot FIREPol® Blend Master Mix was used in this study. The blend master mix contains a proof reading enzyme and bovine serum albumin (BSA) to help increase the specificity of the reaction as well as a gel loading dye to decrease sample-handling time.

The DNA (1 μl) from all specimens submitted was also used to test the multiplex during its stages of development. Due to the uncertainty of the DNA stability in PEG/KOH only the last 26 specimens submitted were used to test the specificity, sensitivity and diagnostic accuracy of the final optimized multiplex PCR.

All PCR results were analysed by electrophoreses of 10 μl in a 2 % (*w/v*) agarose gel in TAE buffer (Appendix A.I) and 1.5 μg/ml of ethidium bromide using an Owl gel electrophoresis system.

2.2.7. Real-Time Multiplex and Melt Curve Analysis

A multiplex real-time PCR was trialled using melt curve analysis with components based on the optimized results from the conventional multiplex PCR.

Each reaction consisted of 1 x Buffer B1 (Solis BioDyne), 4 mM $MgCl_2$ (Solis BioDyne), and 0.25 mM of each dNTP (Solis BioDyne), 1 U Hot FIREPol® DNA Polymerase (Solis BioDyne), and 0.02 mM SYTO® 82 (Life Technologies), an orange fluorescent nucleic acid stain. The same DNA templates were used as in the multiplex diagnostic PCR, 40 - 50 ng of DNA from a representative of each of the three *Streptococcus* species of interest. The reactions were performed using the Rotor-Gene 6000 (Corbett) and melt curves analysis was done using the supplied software (Rotor-Gene 6000 Series Software 1.7).

The reaction parameters were the same as used in the multiplex PCR with additional acquiring steps to measure fluorescence. The initial

activation/denaturation was 95 °C for 15 mins. Touchdown thermocycling was employed with 10 cycles of denaturing at 95 °C for 15 secs, annealing at 65 °C for 15 secs (-1 °C /cycle) and an extension at 72 °C for 20 secs. Another 35 cycles were done at the optimal annealing temperature consisting on 95 °C for 15 secs, 55 °C for 15 secs, 72 °C for 20 secs and an acquiring step at 80 °C for 15 secs. The acquiring step detected fluorescent at 555 nm (excitation at 470 nm, cycling on yellow).

To generate melt curves, the PCR products were heated from 75 °C to 99 °C in 0.25 °C increments every 5 secs. This caused a dissociation of the amplified dsDNA into single strands. Depending on the length and sequence of the PCR product and therefore the strength of the bonds between the strands, each specific PCR product will dissociate at different temperatures. This is referred to as the melting temperature (T_m). As the products 'melted' the decrease of fluorescence was measured due to the release of SYTO® 82, which as an intercalating fluorescent dye, was incorporated into the dsDNA. The Rotor-Gene 6000 Series Software 1.7 formulated this measure of fluorescent decrease into a melt curve, plotting it in a graph as the negative derivative of fluorescence (- dF) with reference to temperature as described by Ririe et al. (1997). The peak of the melt curve represented the T_m for the product. The absence or presence of the appropriate melt curves was used to determine the presence of *S. equi*, *S. zooepidemicus* and *S. equisimilis*.

The Real-time PCR products were checked by electrophoreses of 10 µl in a 2 % (*w/v*) agarose gel in TAE buffer (Appendix A.I) and 1.5 µg/ml of ethidium bromide using an Owl gel electrophoresis system. This was to confirm that the correct products were amplified and that all amplified products were being detected by melt curve analysis.

Using the same 26 specimens used to test the multiplex PCR; the specificity, sensitivity and accuracy of the real-time multiplex PCR was also tested by comparing its results to that of the uniplex PCR.

2.2.8. Multiplex Specificity, Sensitivity and Diagnostic Accuracy Analysis

Multiplex analysis was done by comparing the multiplex results (using all four primer pairs in one reaction) to those of the uniplex PCR (using each primer pair in separate reactions) based on the assumption that the uniplex PCR would be more reliable (refer to section 2.2.3, Uniplex PCR analysis of Specimens, for details on the uniplex reactions). A multiplex result was considered a true positive (TP) if a positive result for both multiplex and uniplex PCR's of the specimen was achieved; this is conversely the same for a true negative (TN) result, both reactions resulting in a negative. A multiplex result was considered a false positive (FP) if a positive result was given in the multiplex and a negative uniplex result for the same specimen. A multiplex result was considered a false negative (FN) if a negative result was given in the multiplex and a positive uniplex result for the same specimen. The specificity, sensitivity and diagnostic accuracy of the multiplex was calculated using the TP, TN, FP and FN results based on the following equations:

Specificity = TN/(TN+FP)
Sensitivity = TP/(TP+FN)
Diagnostic Accuracy = (TP+TN)/(TP+FP+TN+FN)

2.3. Results

2.3.1. Presence of *S. equi* in Specimens

A total of 201 specimens were submitted between October 2010 and February 2012. All specimens were PCR tested for *S. equi* resulting in 26 confirmed cases of strangles. All 26 specimens were from horses bearing clinical signs of strangles except two, where this information was not available. The specimen details are presented in Table 2.2. Upon culture, specimens 136 and 142 had atypical dry colony isolate morphologies and after 2-3 passages of dry isolates, typical mucoid colonies started to appear. Both of these specimens came from horses that became symptomatic after receiving the Pinnacle® IN vaccination. Through typing, the dry and mucoid isolates from both specimens were confirmed as resulting from the Pinnacle® IN vaccine (see chapter 3). Nineteen of the 26 *S. equi* specimens were successfully cultured and isolated (Table 2.2).

Table 2.2: Specimen and isolate details, PCR positive for *S. equi*.

Specimen #*	Date	Isolate ID**	Specimen Type	Location	Sex	Age	Breed†	Symptoms
001	19/10/10	OP1	Nasal swab	Matamata	F	3 yrs	TB	Submandibular swelling and nasal discharge, initial pyrexia
008	15/12/10	OP3	Nasal swab	Cambridge	F	15 yrs	TB	Purulent bilateral nasal discharge, swollen submandibular glands + abscesses
029	24/12/10	OP2	Nasal swab	Cambridge	F	9 yrs	TB	Purulent bilateral nasal discharge, swollen submandibular glands + abscesses
041	17/01/11	OP5	Nasal swab	Cambridge	F	~15 yrs	Andalusian	Purulent bilateral nasal discharge
043	19/01/11	OP4	Nasal swab	Cambridge	F	~10 yrs	WB	Purulent bilateral nasal discharge
103	10/03/11	OP6	Abscess swab	Matamata	M	4 mths	TB	Abscessed submandibular lymph node
131	10/06/11	OP7	Abscess swab	Palmerston North	F	3 yrs	TB	Clinical, abscessed submandibular lymph nodes
132	10/06/11	OP8	Abscess swab	Palmerston North	F	4 yrs	TB	Clinical, abscessed submandibular lymph nodes
134	4/07/11	OP10	Abscess swab	Rangiora	M	3 yrs	SB	Abscessed submandibular lymph node
135	17/05/11	OP9	Culture from Gribbles	Rangiora	F	1 yr	SB	Abscess below mandible, 39.4 temp but bright
138	16/06/11	OP12	Nasal and abscess swab	Mosgiel	M	3 yrs	SB	Clinical strangles
141	22/07/11	OP11	Abscess swab	Matamata	F	2 yrs	TB	Abscess surrounding TMJ joint, 40 temp
143	29/07/11	OP13	Culture from Gribbles	Rangiora	F	18 yrs 11mths	SB	Discharging submandibular lymph nodes and respiratory signs
174	18/08/11	OP14	Abscess aspirate	Bulls	M	1 yr		Abscess
136 dry	27/06/11	OP18	Culture from Gribbles	Christchurch	F	1 yr	SB	Swollen retropharyngeal abscess, horse received first Pinnacle Strangles vaccination ~10days prior to sample taken. Initially only dry colony morphology - after 2 passages mucoid colonies started to appear
136 mucoid		OP17						

Continued next page….

Specimen #*	Date	Isolate ID**	Specimen Type	Location	Sex	Age	Breed†	Symptoms
142 dry	27/07/11	OP16	Abscess swab	Otago	F	10 mths	TB	Weaned 5months prior, 03/06/11 1st Pinnacle vaccine; 17/06/11, 2nd Pinnacle vaccine. Blood, puss, swelling, loss of hair of both mandibular lymph nodes, bright, no fever. Initially only dry colony morphology - after 3 passages mucoid colonies started to appear
142 mucoid		OP15						
Pinnacle® IN S. equi Vaccine Dry	13/10/10	OP19	Lot#: 139192C, 15/12/11	Fort Dodge, USA				Dry colony morphology
019	20/12/10	N‡	Nasal swab	Cambridge	F	10 yrs	WB	Purulent bilateral nasal discharge, swollen submandibular glands + abscesses
030	24/12/10	N	Nasal swab	Cambridge	F	5 yrs	WB	Purulent bilateral nasal discharge, swollen submandibular glands + abscesses
105	14/03/11	N	Nasal swab	Whangarei				
113	15/04/11	N	Nasal swab	Karaka				Purpura hemorrhagica, limb edema, pyrexia and petechiations of mucus membranes
119	29/04/11	N	Nasal swab	Cambridge				Clinical strangles
120	29/04/11	N	Nasal swab	Cambridge				Clinical strangles
133	27/06/11	N	Swab	Karaka				
137	8/07/11	N	Abscess swab	Rangiora	M	4 yrs	SB	Abscessed submandibular lymph node
139	16/06/11	N	Nasal swab	Mosgiel	F	3 yrs	SB	Clinical strangles
140	16/06/11	N	Nasal swab	Mosgiel	F	3 yrs	SB	Clinical strangles

*Specimen #: All isolates had the typical mucoid colony morphology unless specified as "dry". **Isolate ID: An identification assigned to any *S. equi* that was successfully cultured and isolated. ‡N indicates that *S. equi* was unable to be cultured and isolated.
†Breed: TB, Thoroughbred; SB, Strandardbred; WB, Warmblood.

2.3.2. Presence of Multiple *Streptococcus* Species in Specimens

From the 201 specimens received, 73 that resulted in a positive PCR for one of the streptococci of interest were also PCR tested for all three of the *Streptococcus* species (Table 2.3). Of these, 60 were positive for *S. equisimilis* (82.19 %), 23 were positive for *S. zooepidemicus* (31.51 %) and seven were positive for *S. equi* (9.59 %). For use in the development of the multiplex diagnostic PCR, four of the *S. equisimilis* and two of the *S. zooepidemicus* were able to be culture isolated and DNA extracted. *S. equi* was found together with *S. equisimilis* in four of the 73 samples (5.48 %) but was not found to occur with *S. zooepidemicus*. However, *S. zooepidemicus* and *S. equisimilis* were present together in 13 specimens giving an occurrence of 17.81 %. No specimen contained all three species (Table 2.3).

Table 2.3: Occurrence of *S. equi*, *S. zooepidemicus*, and *S. equisimilis* within positive PCR specimens.

	Combination of Strep Species in Specimens*	**Percentage**
S. equi	3	4.11 %
S. zooepidemicus	10	13.70 %
S. equisimilis	43	58.90 %
S. equi/zooepidemicus	0	0 %
S. equi/equisimilis	4	5.48 %
S. zooepidemicus/equisimilis	13	17.81 %
S. equi/zooepidemicus/equisimilis	0	0 %
Total Specimens	73	
	Total Occurrence of each Strep Species across the 73 specimens*	
S. equi	7	9.59 %
S. zooepidemicus	23	31.51 %
S. equisimilis	60	82.19 %

*Tallied only from specimens PCR tested for all three *Streptococcus* species and resulting in a positive for one or more of the three *Streptococcus* species tested.

2.3.3. Isolate Species Confirmation

The 19 isolates of *S. equi* were confirmed through PCR and sequencing of the *seeI* and *seM* gene. The two isolates of *S. zooepidemicus* and four isolates of *S. equisimilis* were also confirmed through PCR and sequencing of the sortase gene (*srtz* primers) and the streptokinase precursor gene (*eqsim* primers) respectively. The sequences and BLAST results can be found in Appendix (A.V). Further, all isolates obtained identified as their species with their fermentation abilities (Table 2.4, Figure 2.1). The 19 isolates of *S. equi*, including the vaccine isolate, were unable to ferment the sugar sorbital or trehalose. *S. zooepidemicus* isolates fermented sorbital but not trehalose, whereas the isolates of *S. equisimilis* were able to ferment trehalose but not sorbital.

Table 2.4: Biochemical fermentation and GPNT blood agar selective plate results.

Species	**Fermentation**** Sorbital	Trehalose	**GPNT Blood Agar Growth*****
*S. equi** (n = 19)	-	-	+
S. zooepidemicus (n = 2)	+	-	+
S. equisimilis (n = 4)	-	+	+

*The 19 *S. equi* isolates include the Pinnacle® IN Vaccine isolate.
** (+) = a positive fermentation result; (-) = a negative fermentation result.
*** (+) = growth of bacterial species.

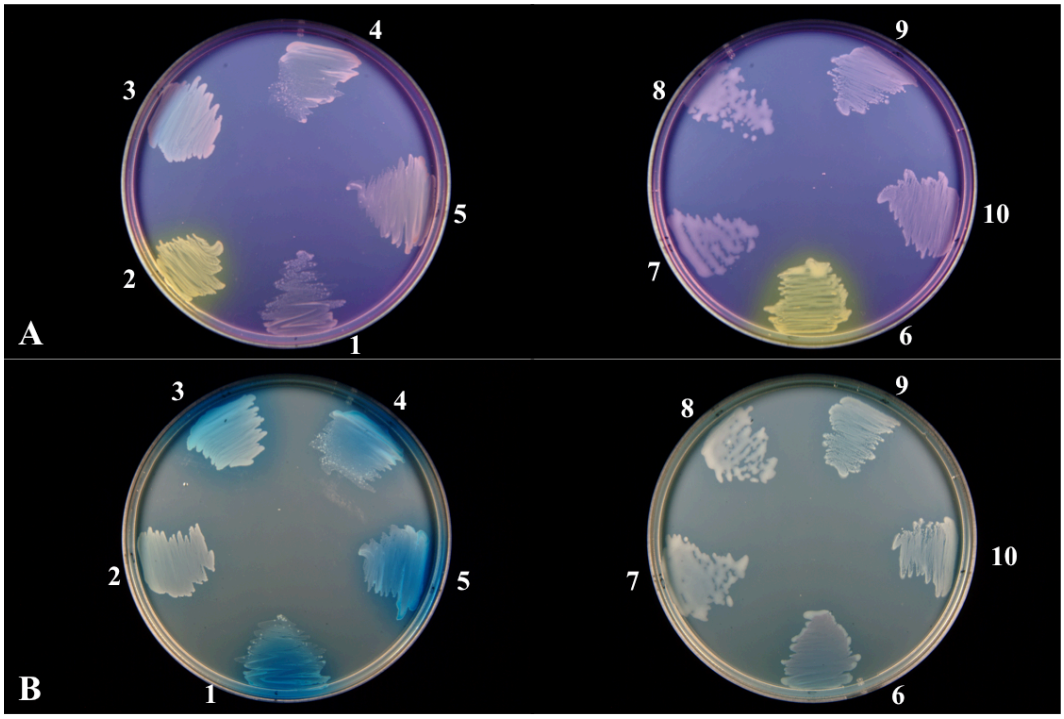


Figure 2.1: An example of the biochemical plate results. Purple plates are supplemented with sorbitol (A); positive sorbitol fermentation results in a change of color to yellow. Pale blue plates are supplemented with trehalose (B); positive trehalose fermentation results in bright blue coloring. 2 and 6, *S. zooepidemicus*; 1, 3- 5, *S. equisimilis*; 7-10, *S. equi*.

2.3.4. Selective Media

All representative isolates of *S. equi*, *S. zooepidemicus*, *S. equisimilis* and of the Pinnacle® IN Vaccine were able to grow in both selective broths trialled (Figure 2.2,Figure 2.3). There was no significant difference between BHI/GPNT and THY/GPNT for each isolate after 24 hrs (t-test, $P > 0.05$). *S.* equisimilis and the vaccine isolates had slow initial growth however, after 48 hrs both had ample growth. *S. equisimilis* grew better in the BHI/GPNT after 48 hrs (t-test, $P = 0.00423$); there was no significant difference between the two selective broths for *S. equi*, *S. zooepidemicus*, or the vaccine (t-test, $P > 0.05$).

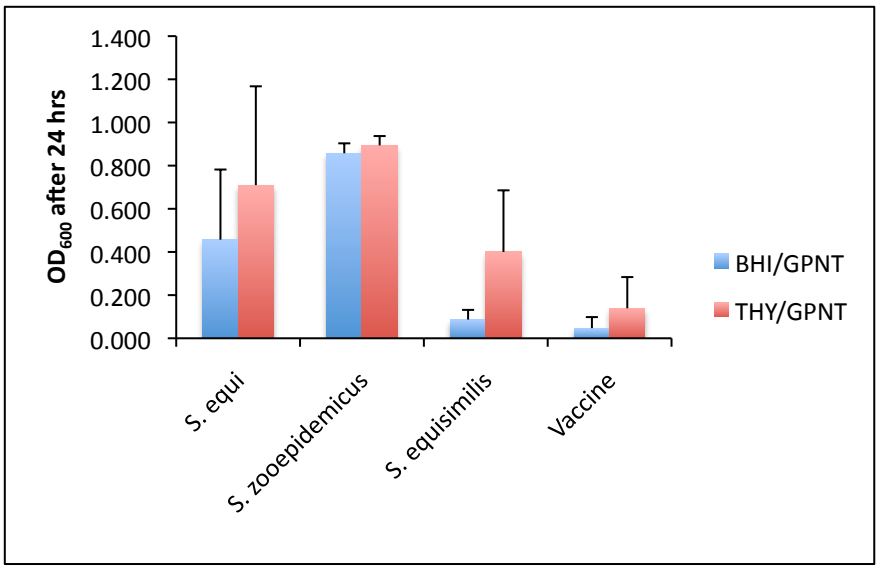


Figure 2.2: Growth of streptococci in selective media after 24 hrs. Graph represents the mean optical density at 600 nm of *S. equi, S. zooepidemicus, S. equisimilis* and Pinnacle® IN vaccine after 24 hrs in BHI/GPNT (Brain Heart Infusion broth supplemented with gentamicin, polymyxin B, nalidixic acid and thallium acetate), represented in blue and THY (Todd Hewitt broth with yeast extract)/GPNT, represented in red. Error bars show the 95 % CI of triplicates.

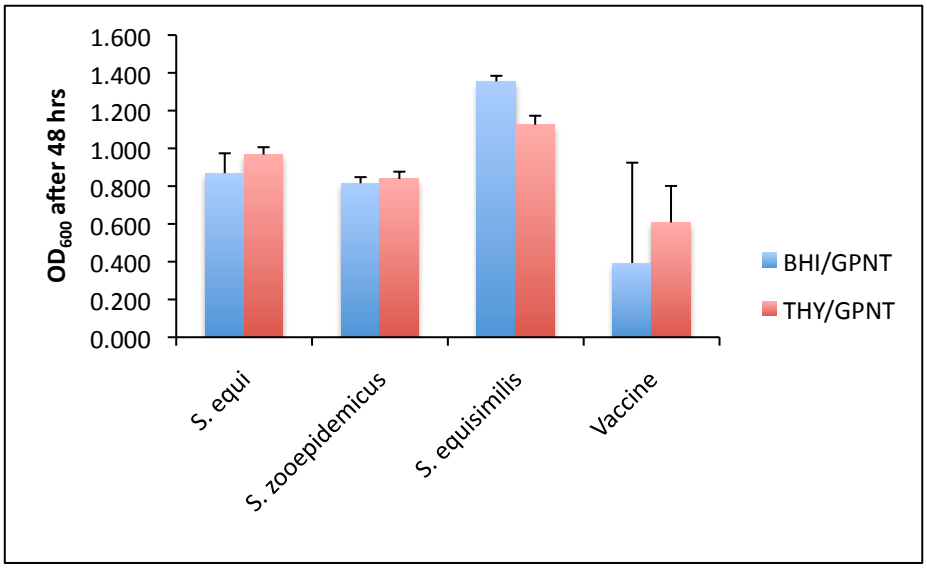


Figure 2.3: Growth of streptococci in selective media after 48 hrs. Graph represents the mean optical density at 600 nm of *S. equi*, *S. zooepidemicus*, *S. equisimilis* and Pinnacle® IN vaccine after 48 hrs in BHI/GPNT (Brain Heart Infusion broth supplemented with gentamicin, polymyxin B, nalidixic acid and thallium acetate), represented in blue and THY (Todd Hewitt broth with yeast extract)/GPNT, represented in red. Error bars show the 95 % CI of triplicates.

All isolates had substantial growth after 24 hrs on the selective blood agar plates, GPNT/blood agar (Table 2.4). Growth was observed after 24hrs for all 19 isolates of *S. equi* (including the Pinnacle® IN Vaccine isolate), the two isolates of *S. zooepidemicus*, and the four isolates of *S. equisimilis*. Figure 2.4 shows an example of 50 µl of a nasal swab specimen, processed using the method mentioned prior (see 2.2.2 Specimen Processing and DNA Isolation), grown on a plain 10 % blood agar plate verses a GPNT/blood agar plate after 24 hrs. The small β-haemolytic colonies present on the GPNT/blood agar plate are not present on the blood agar and were possible streptococci. However, these colonies on this example plate did not identify as any of our three streptococci of interest *S. equi*, *S. zooepidemicus* or *S. equisimilis* under PCR tests.

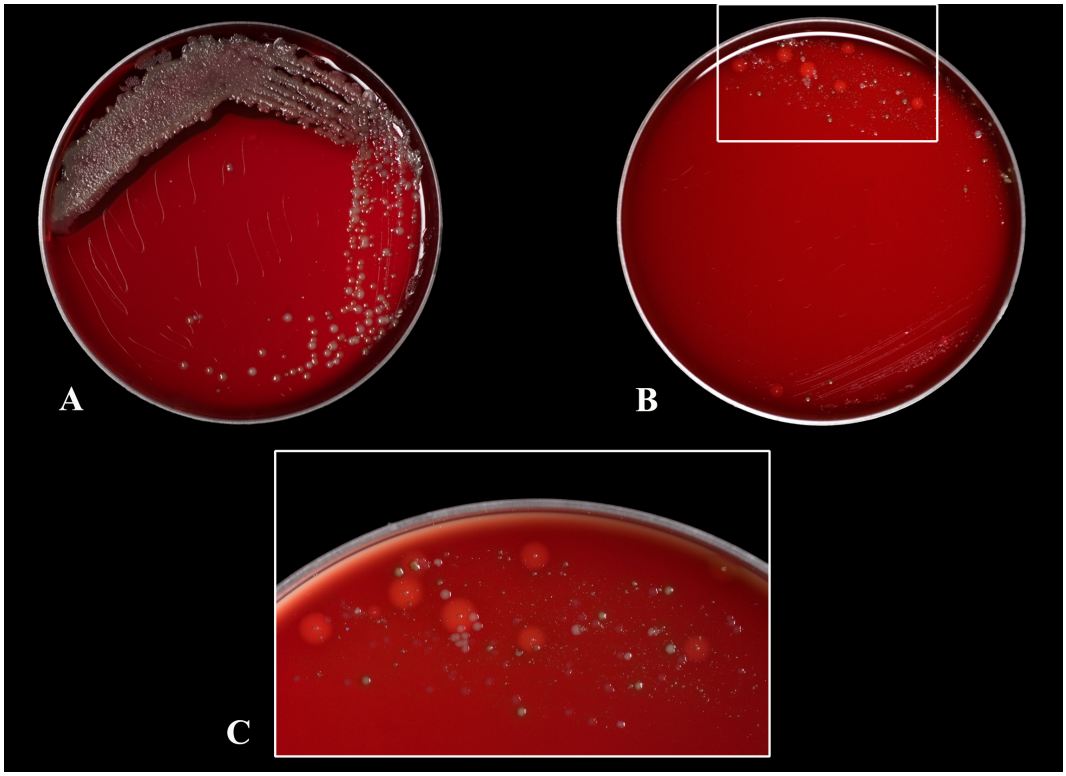


Figure 2.4: Comparison of blood agar and selective GPNT/blood agar plates. 50 µl of a specimen was streaked out on a blood agar plate (A) and on a GPNT/blood agar plate (B). The squared area in B is enlarged in C to show the β-haemolytic colonies not identified on the blood agar plate (A) due to overgrowth of other contaminates.

2.3.5. Conventional Multiplex PCR Optimization Results

2.3.5.1. Orthogonal Array Optimization

Eighteen multiplex reactions were run using the orthogonal array as displayed in Table 2.5. Reaction nine resulted in the best simultaneous amplification of all four products as indicated by the banding pattern (Figure 2.5). The levels for the components in reaction nine were as follows: T^A at 55 °C, Taq at 1 U, $MgCl_2$ at 4 mM, dNTPs at 0.2 mM, *seeI* primer pair at 0.3 μM; *srtz* primer pair at 0.2 μM, *eqsim* primer pair at 0.15 μM and the 16S rDNAprimer pair at 0.2 μM.

The annealing temperature appeared optimal at 55 °C. At 52 °C, the *seeI* product did not amplify (lanes 1-6, Figure 2.5). Lower amplification of the two larger products, *seeI* and 16S; occurred at 60 °C (lanes 14, 15, 17, and 18) when compared to 55 °C (lane 9, Figure 2.5). The optimal Taq concentration was > 0.5 U. Amplification of the largest product, *seeI*, only resulted when the Taq concentration was either 0.75 U (lanes 14 and 15) or 1 U (lanes 9, 17, and 18, Figure 2.5). The $MgCl_2$ appeared to have the greatest affect; all reactions containing 2 mM $MgCl_2$ only produced the smaller products, *srtz* and/or *eqsim* (lanes 1, 4, 7, 10, and 16; Figure 2.5). Reaction 13 resulted in no products likely due to human error, however as it only contained 2 mM $MgCl_2$ the assumption was made that it to would only amplify the smaller products (lane 13, Figure 2.5). The concentration of dNTPs appeared to have little effect as 0.2, 0.25 and 0.3 mM resulted in amplification of all four products (lanes 9, 14, 15, 17, and 18; Figure 2.5).

Each of the four primer pairs had different optimal conditions for amplification across the 18 reactions with the larger products more restrictive in the levels of each component (Figure 2.5). The *seeI* product was produced at all three levels trialled; 0.15, 0.2 and 0.3 μM, when the T^A was 55 or 60 °C, when Taq was > 0.5 U and when $MgCl_2$ was at 3 or 4 mM (lanes 6, 8, 9,14, 15, 17, and 18; Figure 2.5). The *srtz* product amplified in all conditions. Nearly all reactions amplified

the *eqsim* product but to differing extents. Generally, 16S rDNA amplified only when $MgCl_2$ was at 3 or 4 mM; with the higher Mg^{2+} it was able to amplify in all three levels of each of the other components (lanes 2, 3, 6, 9, 12, 14, 15, 17 and 18; Figure 2.5).

Table 2.5: The orthogonal array used for optimization of the multiplex PCR

Reaction	Multiplex Components for optimization							
	T^{A*} °C	Taq U	$MgCl_2$ mM	dNTPs mM	*seeI* µM	*srtz* µM	*eqsim* µM	16S µM
1	52	0.5	2	0.2	0.15	0.15	0.15	0.15
2	52	0.5	3	0.25	0.2	0.2	0.2	0.2
3	52	0.5	4	0.3	0.3	0.3	0.3	0.3
4	52	0.75	2	0.2	0.2	0.2	0.3	0.3
5	52	0.75	3	0.25	0.3	0.3	0.15	0.15
6	52	0.75	4	0.3	0.15	0.15	0.2	0.2
7	55	1	2	0.25	0.15	0.3	0.2	0.3
8	55	1	3	0.3	0.2	0.15	0.3	0.15
9	**55**	**1**	**4**	**0.2**	**0.3**	**0.2**	**0.15**	**0.2**
10	55	0.5	2	0.3	0.3	0.2	0.2	0.15
11	55	0.5	3	0.2	0.15	0.3	0.3	0.2
12	55	0.5	4	0.25	0.2	0.15	0.15	0.3
13	60	0.75	2	0.25	0.3	0.15	0.3	0.2
14	60	0.75	3	0.3	0.15	0.2	0.15	0.3
15	60	0.75	4	0.2	0.2	0.3	0.2	0.15
16	60	1	2	0.3	0.2	0.3	0.15	0.2
17	60	1	3	0.2	0.3	0.15	0.2	0.3
18	60	1	4	0.25	0.15	0.2	0.3	0.15

The different shaded cells in each column represent the three levels chosen for each factor. The 18 rows represent 18 different reactions. The reaction resulting in the best amplification of all four products is highlighted with a bold boarder, reaction 9.

*T^A = Annealing Temperature

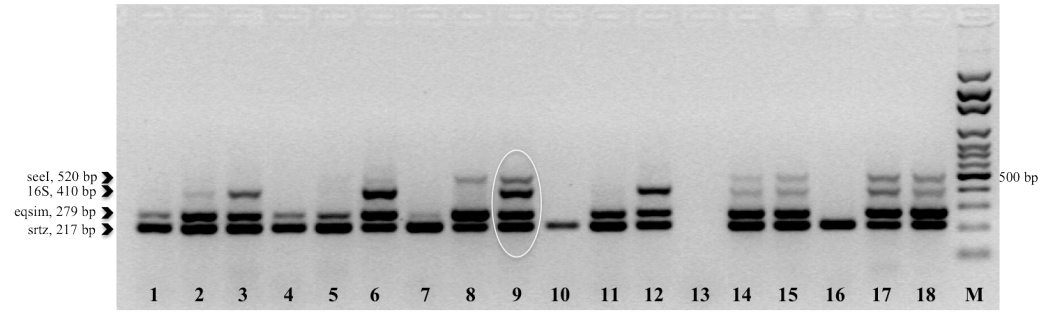


Figure 2.5: Orthogonal array gel results. Lanes 1-18 display the band pattern for the 18 reactions of the orthogonal array. Lane M = 100 bp ladder (Solis Biodyne). Reaction 9 resulted in the best amplification of all four products (circled).

2.3.5.2. Further Multiplex Component Optimization

In order to improve the multiplex reaction obtained in reaction nine of the orthogonal array (Table 2.5), an increase in the amplification of the *seeI* product (presence of *S. equi* indicator) was needed. As there appeared to be preferential amplification of the smallest product, *srtz*, decreasing this primer pair concentration was trialled in reactions containing either 3 or 4 mM $MgCl_2$ (Table 2.6). Decreasing the $MgCl_2$ resulted in a decrease in the amplification of the larger products (Figure 2.6). A decrease in the concentration of the primer pair producing the smallest product, *srtz*, increased the amplification of the largest product, *seeI* (Figure 2.6).

Table 2.6: Trial of different component levels to further optimize the multiplex PCR.

Reaction	Multiplex Components							
	T^{A*} °C	Taq U	$MgCl_2$ mM	dNTPs mM	*seeI* µM	*srtz* µM	*eqsim* µM	16S rDNA µM
1	55	1	4	0.2	0.3	0.15	0.15	0.2
2	55	1	3	0.2	0.3	0.15	0.15	0.2
3**	55	1	4	0.2	0.3	0.2	0.15	0.2

Shaded areas highlight the variances in each reaction.
* T^A =Annealing Temperature
** Reaction 3 = optimal reaction 9 in Table 2.5, included for comparison.

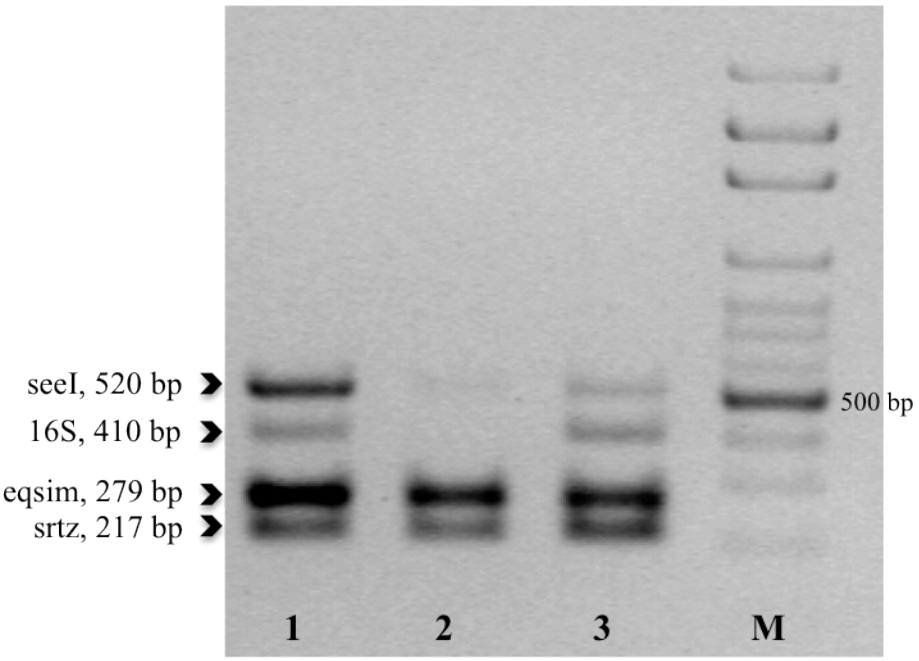


Figure 2.6: Results of further optimization of the multiplex PCR. Lane 1 shows the result of 0.15 µM *srtz* primer pair and 4 mM $MgCl_2$, Lane 2 shows the result of 0.15 µM *srtz* and 3 mM $MgCl_2$; both compared to lane 9, 0.2 µM *srtz* and 4 mM $MgCl_2$. Lane M = 100 bp ladder (Solis Biodyne).

2.3.5.3. Hot FIREPol® DNA Polymerase versus Hot FIREPol® Blend Master Mix

Overall there was very little difference between the Hot FIREPol® Blend Master Mix and the reactions containing only the Hot FIREPol® DNA Polymerase (Figure 2.7). The 16S rDNA and *seeI* product intensity was greater in the reaction containing only the Hot FIREPol® DNA Polymerase.

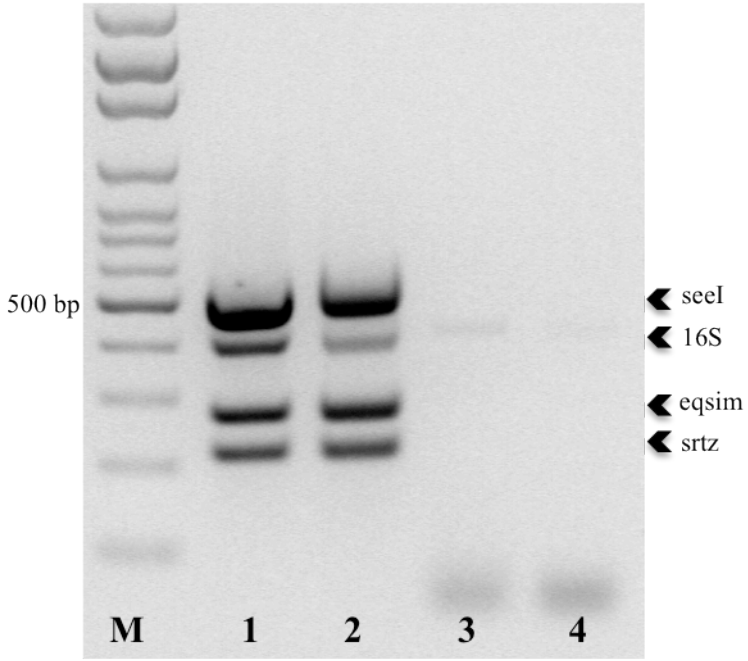


Figure 2.7: Comparison of Hot FIREPol® DNA Polymerase (lane 1) to Hot FIREPol® Blend Master Mix (lane 2). Lane M = 100 bp ladder (Solis Biodyne). Lane 3 = no template control of the Hot FIREPol® DNA Polymerase and Lane 4 is the no template control of the Hot FIREPol® Blend Master Mix.

2.3.6. Final Optimized Diagnostic Multiplex

The final reaction mixture contained 1 x Buffer B1 (Solis BioDyne), 1 U Hot FIREPol® DNA Polymerase (Solis BioDyne), 4 mM $MgCl_2$ (Solis BioDyne), 0.2 mM dNTPs (Solis BioDyne), 0.3 μM of *seeI* primer pair, 0.15 μM of *srtz* primer

pair, 0.15 μM of *eqsim* primer pair and 0.2 μM of 16S primer pair. A touchdown PCR protocol was used to reduce the potential of nonspecific amplification. The final PCR parameters consisted of an initial activation/denaturation at 95 °C for 15 mins. This was followed by touchdown thermocycling with 10 cycles of denaturing at 95 °C for 15 secs, annealing at 65 °C for 15 secs (-1 °C /cycle) and extending at 72 °C for 20 secs. Then an additional 35 cycles at consisting of 95 °C for 15 secs, 55 °C for 15 secs, and 72 °C for 20 secs. Figure 2.8 shows the expected banding result of a positive diagnosis for each streptococci of interest. The 16S product was produced in all reactions. The presence of *S. equi* was confirmed by the amplification of the *seeI* product, 520 bp; *S. zooepidemicus* was confirmed by the amplification of the *srtz* product, 217 bp; and the presence of *S. equisimilis* was confirmed by the amplification of the *eqsim* product, 279 bp.

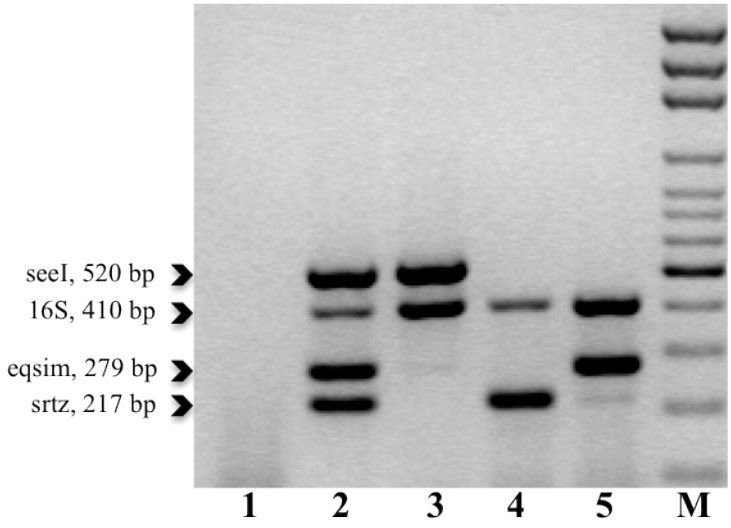


Figure 2.8: Expected banding pattern for the diagnostic multiplex of streptococci species. Lane 1 is the no-template negative control, lane 2 is the positive control representing all 3 streptococci species, lane 3 represents a positive result for *S. equi*, lane 2 represents a positive result for *S. zooepidemicus*, lane 5 represents a positive result of *S. equisimilis* and lane M is a 100 bp ladder (Solis Biodyne).

2.3.7. Conventional Multiplex PCR Results

The final optimized diagnostic multiplex was trialled on all obtained isolates. The multiplex correctly identified all 19 isolates of *S. equi* (including the vaccine isolate), the two isolates of *S. zooepidemicus* and the four *S. equisimilis* isolates (Table 2.7). All isolates were positive for the 16S rDNA product. The 19 isolates of *S. equi* produced the 520 bp *seeI* product and the 410 bp 16S rDNA product. *S. zooepidemicus* isolates produced the 217 bp *srtz* product and the 410 bp 16S rDNA product. All the *S. equisimilis* isolates produced the 279 bp *eqsim* product and the 410 bp 16S rDNA product.

Table 2.7: Diagnostic multiplex results on identifying *S. equi*, *S. zooepidemicus*, *and S. equisimilis*.

	Primer Pairs (Species Identified)			
Isolates	*seeI* (*S. equi*)	*srtz* (*S. zooepidemicus*)	*eqsim* (*S. equisimilis*)	16S rDNA (common to all bacteria)
S. equi ($n = 19$)	+	-	-	+
S. zooepidemicus ($n = 2$)	-	+	-	+
S. equisimilis ($n = 4$)	-	-	+	+

(+) indicates a positive PCR result with the primer pair and a (-) indicates a negative result.

The specificity, sensitivity and diagnostic accuracy was deduced by testing the multiplex PCR with 26 submitted specimens and comparing the results with those of the uniplex of each primer-pair individually for all the 26 specimens (Table 2.8). The multiplex had 100 % specificity for the three streptococci species. The 16S rDNA primer will amplify all bacterial species, making the specificity not applicable for this product as there will be no TN's. The sensitivity for detecting the presence of bacteria (16S rDNA) was 100 %. The sensitivity for detecting the presence of *S. zooepidemicus* was 100 % and was 87.5 % for *S. equisimilis*. The multiplex was unable to detect *S. equisimilis* in one specimen out of the 8 that were detected in the uniplex PCR. The sensitivity for detecting *S. equi* was unable to be deduced; as *S. equi* was not detected in any of the 26 specimens

either in the multiplex or uniplex PCR reactions. The Diagnostic Accuracy for correctly diagnosing *S. equi*, *S. zooepidemicus* and bacterial presence (16S rDNA) was 100 %. This was reduced to 96.2 % with diagnosing *S. equisimilis* due to the one false negative result as compared to the uniplex PCR.

Table 2.8: Specificity, Sensitivity and Diagnostic Accuracy of the multiplex PCR using the Uniplex PCR results as a reference standard*

	Uniplex/ Multiplex	Primer Pairs (Species Identified)			
		seeI (*S. equi*)	*srtz* (*S. zooepidemicus*)	*eqsim* (*S. equisimilis*)	16S rDNA (common to all bacteria)
TP	+/+	0	1	7	32
FP	+/-	0	0	0	0
TN	-/-	26	25	18	0
FN	-/+	0	0	1	0
Specificity		100 %	100 %	100 %	NA
Sensitivity		NA**	100 %	87.5 %	100 %
Diagnostic Accuracy		100 %	100 %	96.2 %	100 %

TP = true positive; FP = false positive; TN = true negative; FN = false negative. Specificity = TN/(TN+FP); Sensitivity = TP/(TP+FN); Diagnostic Accuracy = (TP+TN)/(TP+FP+TN+FN). **NA = Not applicable

2.3.8. Real-Time Multiplex PCR Results

In a real-time multiplex where an equal mix of *S. equi*, *S. zooepidemicus*, and *S. equisimilis* DNA (40-50 ng) was used, the melt curve analysis was able to clearly identify all three species. Each species had different T_m's resulting in separate peaks on the graphed melt curves; the presence of *S. equi*, *S. equisimilis*, and *S. zooepidemicus* identified by the peak of *seeI*, *eqsim* and *srtz* respectively (Figure 2.9). The 16S rDNA peak identifies the presence of bacteria and in the melt curve analysis was barely detectable as indicated by the height of the melt curve (Figure 2.9). The results were confirmed by gel eletrophoresis, all four products were present with the 16S rDNA product weak comparatively (Figure 2.10).

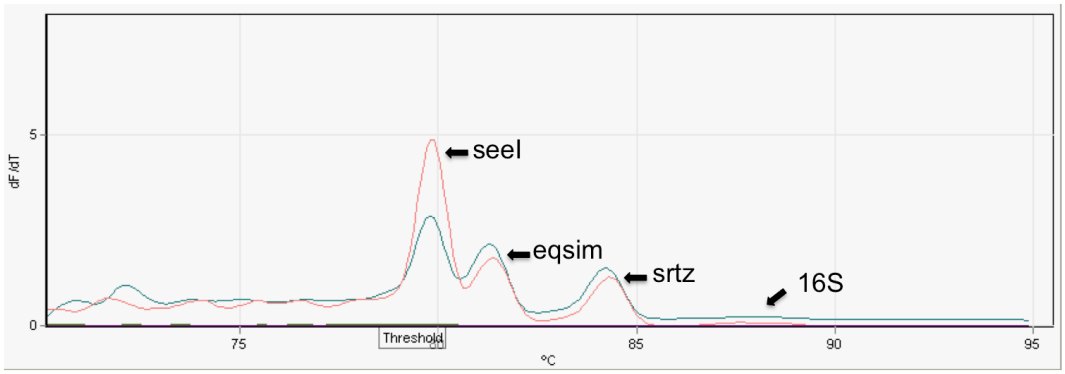


Figure 2.9: The melt curve graph showing the results of two real-time multiplex reactions, each with a mix of the three streptococci species of interest. They are graphed as the decrease of fluorescence with respect to time (dF/dT) in relation to temperature (°C) and each peak represents the products T_m (melt temperature) of the primer pairs as indicated, *seeI*, *eqsim*, *srtz* and 16S rDNA.

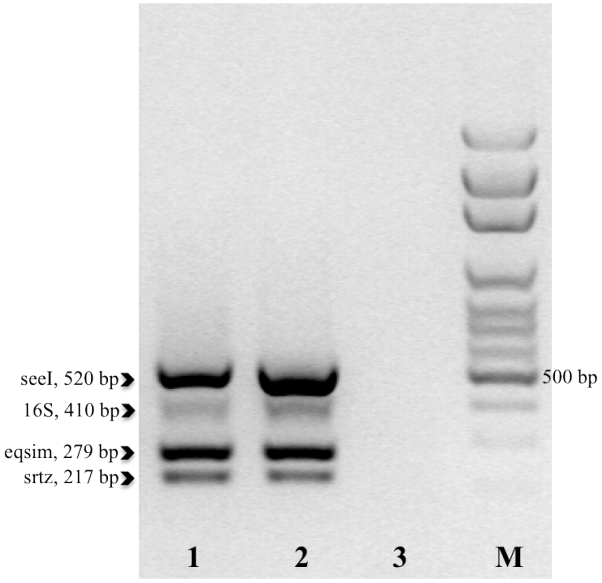


Figure 2.10: Real-time multiplex gel results. Lane 1 and 2 are the banding results of two real-time multiplex reactions with a mix of *S. equi*, *S. zooepidemicus* and *S. equisimilis* as templates per reaction. Lane 3 is the no template control reaction and lane M is a 100 bp ladder (Solis Biodyne).

Through melt curve analysis, the real-time multiplex was able to accurately identify the 19 isolates of *S. equi*, the two isolates of *S. zooepidemicus* and the four isolates of *S. equisimilis* and distinguish the three species from each other (Figure 2.11 and Table 2.9). The melt curve of the *seeI* product peaks at a T_m of 80.05 (standard deviation (SD) = 0.06) and identifies the presence of *S. equi*. The melt curve of the *eqsim* product peaks at a T_m of 81.60 (SD = 0.08) and identifies the presence of *S. equisimilis*. As for the *srtz* product, the melt curve peaks at a T_m of 84.53 (SD = 0.04) and identifies the presence of *S. zooepidemicus*. The 16S product melt curve was more variable as seen in the SD of each species mean melt curve T_m (Table 2.9) and ranging from a T_m of 89.97 – 89.39 between the three species (Figure 2.11 and Table 2.9). The amount of product produced by the 16S rDNA primer pair was extremely low as seen earlier in the real-time multiplexes with mixed DNA templates.

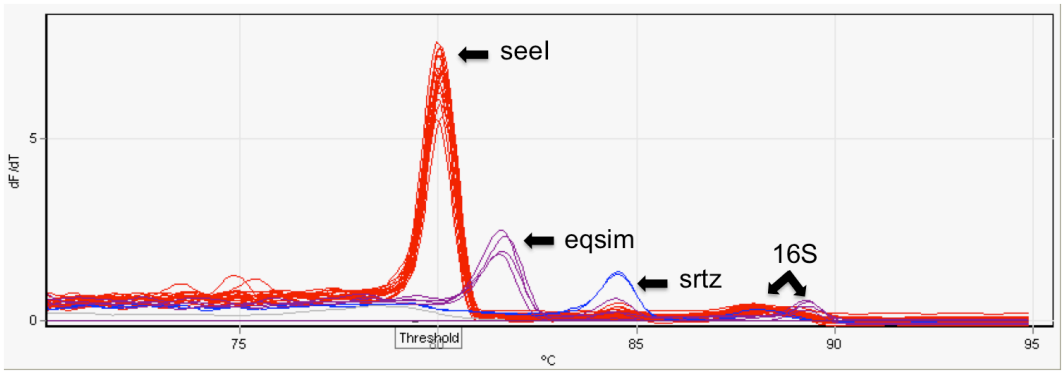


Figure 2.11: Melt curves of *S. equi*, *S. zooepidemicus* and *S. equisimilis* isolates graphed as the decrease of fluorescence with respect to time (dF/dT) in relation to temperature (°C). The products from each of the four primer pairs in the multiplex are indicated; the 19 *S. equi* isolates giving peaks representative of *seeI* and 16S rDNA products, the four *S. equisimilis* isolates giving peaks representative of *eqsim* and 16S rDNA products and the two *S. zooepidemicus* isolates giving peaks representative of *srtz* and 16S rDNA products.

Table 2.9: The mean melt temperatures (T_m) of *S. equi*, *S. zooepidemicus* and *S. equisimilis* isolates.

	Mean T_m (SD)**			
Isolates*	*seeI*	*srtz*	*eqsim*	16S rDNA
S. equi (*n* = 19)	80.05 (0.06)	-	-	87.97 (0.20)
S. zooepidemicus (*n* = 2)	-	84.53 (0.04)	-	88.16 (0.12)
S. equisimilis (*n* = 4)	-	-	81.60 (0.08)	89.39 (0.24)

**n* = number of isolates. **SD = the standard deviation around the mean T_m

The specificity, sensitivity and diagnostic accuracy of the real-time multiplex was deduced in the same way as the convention multiplex PCR; by testing the real-time multiplex with 26 specimens and comparing the results with that of the uniplex results of each primer pair individually for all 26 specimens (Table 2.8). The multiplex had 88 % specificity for *S. equi* due to three FP results. 100 % specificity was found for *S. zooepidemicus* and *S. equisimilis*. The 16S primer-pair should amplify all bacterial species, making the specificity not applicable. However, the sensitivity and diagnostic accuracy for detecting the presence of bacteria (16S rDNA) was 0 %, as it did not detect any 16S rDNA products in any of the reactions. The sensitivity for detecting the presence of *S. zooepidemicus* was 100 % and was 0 % for *S. equisimilis*. The real-time multiplex was unable to detect *S. equisimilis* in any of the eight specimens that the uniplex PCR detected. The sensitivity for detecting *S. equi* was unable to be deduced; as there were no TP's or FN's. The Diagnostic Accuracy for correctly diagnosing *S. equi* was 88 % and for *S. zooepidemicus* was 100 %. This was reduced to 69.2 % with diagnosing *S. equisimilis* due to the 8 FN results as compared to the uniplex PCR. The diagnostic accuracy for correctly identifying the presence of bacteria (16S rDNA) was 81 % owing to five FN.

To confirm the results, these real-time reactions were run on an agarose gel to verify the absence and presence of the products produced (Table 2.10). Zero *seeI*

products (*S. equi*) were present on the gel, unlike the three detected in the melt curve analysis whose products were actually the size of the *eqsim* products (*S. equisimilis*) on the gel. This gave a total of five reactions resulting in *eqsim* products on the gel compared to zero with melt curve analysis. Both the gel and melt curve analysis results were the same for accurately detecting the *srtz* (*S. zooepidemicus*) product. Twenty out of the 26 specimen reactions produced the 16S rDNA product, whereas the melt curve analysis was unable to detect any.

Table 2.10: Specificity, Sensitivity and Diagnostic Accuracy of the real-time multiplex melt curve analysis verses the corresponding agarose gel results using the Uniplex PCR results as a reference standard*

	Uniplex/ Real-time Multiplex	**Primer Pairs (Species Identified)****			
		seeI (*S. equi*)	*srtz* (*S. zooepidemicus*)	*eqsim* (*S. equisimilis*)	16S rDNA (common to all bacteria)
TP	+/+	0/0	1/1	0/5	0/20
FP	+/-	3/0	0/0	0/0	0/0
TN	-/-	23/26	25/25	18/18	0/0
FN	-/+	0/0	0/0	8/3	26/6
Specificity		88 % /100 %	100 % /100 %	100 % /100 %	NA
Sensitivity		NA	100 % /100 %	0 % /62.5%	0 % /77 %
Diagnostic Accuracy		88 % /100 %	100 % /100 %	69.2 % /88.5%	0 % /77 %

*TP = true positive; FP = false positive; TN = true negative; FN = false negative. Specificity = TN/(TN+FP); Sensitivity = TP/(TP+FN); Diagnostic Accuracy = (TP+TN)/(TP+FP+TN+FN). NA = Not applicable. **Displayed as the melt curve analysis / agarose gel results.

2.4. Discussion

2.4.1. Identification and Isolation of Streptococci Species

Throughout this study, 19 isolates of *S. equi*, two isolates of *S. zooepidemicus* and four isolates of *S. equisimilis* were obtained and used towards developing and testing a molecular diagnostic method for the detection of *S. equi* and discriminate it from the other two streptococci. These isolates were genetically characterized by amplification and sequencing of species-specific genes. Initially, the isolation of these species proved difficult due to overgrowth of other bacteria and fungi so a selective media was needed. The GPNT selective media greatly reduced these contaminates and made identification and isolation of the three streptococci of interest easier. All of the streptococci isolates in this study were able to grow on and in the GPNT selective media. Other studies have used selective media containing different supplements: colistin sulphate and oxolinic acid supplements (Newton, Verheyen et al. 2000; Båverud, Johansson et al. 2007), *Streptococcus* selective supplement from Oxoid, Italy (Preziuso, Laus et al. 2010), colistin sulphate, nalidixic acid (Proietti, Bietta et al. 2011) and amikacin supplements (Lanka, Borst et al. 2010). A comparative study to determine which mix of supplements best select for *S. equi*, *S. zooepidemicus*, and *S. equisimilis* was beyond the scope of this project but could be an avenue of future investigation.

Biochemical attributes and colony morphology of *Streptococcus* species have been shown to be unreliable for species identification due to atypical strains. Atypical isolates of *S. equi*, *S. zooepidemicus* and *S. equisimilis* occur on the basis of sugar fermentation. A failure to ferment lactose, ribose, sorbitol and trehalose is typical for *S. equi*; *S. equisimilis* typically ferments ribose and trehalose and some may ferment lactose; typical *S. zooepidemicus* ferments lactose, ribose, and sorbitol (Grant, Efstratiou et al. 1993). Some atypical isolates of *S. equi* were shown to ferment trehalose alone or lactose and trehalose which could be mistaken for *S. equisimilis* (Grant, Efstratiou et al. 1993; Efstratiou, Colman et al.

1994). One study found an atypical strain of *S. equisimilis* to ferment sorbitol in a delayed reaction and one of *S. zooepidemicus* to ferment trehalose (Efstratiou, Colman et al. 1994). Also atypical strains of *S. zooepidemicus,* which did not ferment ribose have been described (Kuwamoto, Anzai et al. 2001). These results suggest that these three streptococci cannot be definitively identified by only a sugar fermentation test. In this present study, no atypical isolates of the three streptococci species were found based on their fermentation abilities of the two sugars sorbitol and trehalose. All of the *S. equi* isolates were unable to ferment either sugar. Even the Pinnacle® IN vaccine strain identified as *S. equi* by not fermenting sorbitol or trehalose. The *S. zooepidemicus* isolates fermented sorbitol and not trehalose and the *S. equisimilis* isolates fermented trehalose and not sorbitol.

Phenotypically, *S. equi* forms mucoid colonies distinct from *S. zooepidemicus*. Although colonies from *S. zooepidemicus* are typically nonmucoid, they can sometimes take on a mucoid formation appearing more like *S. equi* (Sweeney, Timoney et al. 2005). The mucoid appearance of *S. equi* cannot always be relied upon as a less virulent isolate produces a matte appearance (Taylor and Wilson 2006), which is the appearance that the Pinnacle® IN vaccine strain takes on. In this study small, matte atypical colonies were isolated from two horses displaying clinical strangles shortly after vaccinations with the Pinnacle® IN vaccine. These isolates were later typed to the same strain as the vaccine (refer to chapter 3). The questions remain as to whether these 'vaccine breakdown' isolates are just the cause of a bad adverse reaction or whether the vaccine strain has reverted back to a virulent form. Upon sequential passages of these isolates, mucoid colony formations like that of the wild type strains started to appear, which has been previously noted for the Pinnacle® IN vaccine strain (Walker and Timoney 2002). It would be interesting to genome sequence these isolates to determine if reversions have taken place. This vaccine was derived following chemical mutagenesis, inducing random mutations throughout the genome that could be prone to back mutations and thus the possibility of reversion back to full

virulence. Due to safety concerns, it was not been licensed for sale in Europe (Waller and Jolley 2007).

2.4.2. Presence of *Streptococcus* Species in Specimens

Out of the 73 specimens resulting in a positive PCR for one of the three species of interest: *S. equi*, *S. zooepidemicus*, and *S. equisimilis*; it was *S. equisimilis* that occurred most frequent at 82.19 %. *S. equi*, the causative agent of strangles only occurred in 9.59 % of the positive specimens and *S. zooepidemicus* occurred in 31.51 %. Nearly 18 % of these specimens contained both *S. zooepidemicus* and *S. equisimilis* concurrently. A limitation to this study was the unknown infection status of the majority of the horses from which the specimens were obtained. Due to this, correlations between the presence of the bacterium and the clinical state of the horses were unable to be made. Future studies could be done to ascertain if any new *S. zooepidemicus* and *S. equisimilis* positive results obtained are from horses suffering with upper respiratory tract infections and to what degree. These results could help determine the significance of these bacteria in causing respiratory disease within the New Zealand horse population. *S. equisimilis* has been identified as the causative agent in cases of strangles-like diseases in the absence of *S. equi* (Timoney 2004; Laus, Preziuso et al. 2007). *S. zooepidemicus* is a mucosal commensal, opportunistic bacteria which can result in a serious pathogen of the respiratory tract (Timoney 2004). It also has been isolated from horses with strangles-like symptoms ((Timoney 2004; Laus, Preziuso et al. 2007).

Out of the specimens tested for all three streptococci, all of the seven resulting in a positive result for *S. equi* were from horses displaying clinical strangles. Of these, four were also positive for *S. equisimilis*. It is likely that the symptoms were caused by *S. equi* however a pathogenic role for *S. equisimilis* in the disease cannot be ruled out. Or is *S. equisimilis* only acting as an opportunistic secondary invader? The diagnostic multiplex developed here can assist in determining the

role and significance of *S. zooepidemicus* and *S. equisimilis* in respiratory diseases within the New Zealand horse population.

A study investigating β-haemolytic *Streptococcus* species from horses found that *S. zooepidemicus* was the most frequent isolate at 72 % followed by *S. equisimilis* at 21.3 %; both isolated from several organs, placenta and fetal tissues, and from the respiratory tract, nasal passages, guttural pouches and lymph node abscesses (Erol, Locke et al. 2012). This is in contrast to the current study where *S. equisimilis* was most frequently detected from the respiratory tract of New Zealand horses. It must be noted that no other organs or tissues were involved in the current study. Further, the Erol study also found that *S. equi* had a much smaller incident rate at only 5.8 % with a narrower range of tissue tropism then *S. zooepidemicus* and *S. equisimilis*; it was recovered mainly from lymph node abscesses, the upper and lower respiratory tract, nasal samples and the guttural pouches. Interestingly, only a few *S. equi* isolates were recovered from foetal tissue and placentas. Of the 391 cases in the Erol study where β-haemolytic streptococci were isolated, more than one *Streptococcus* species was recovered in 106 cases giving indication to the co-existence of these bacteria (Erol, Locke et al. 2012). This can also be reflected in the present study where 23.29 % of samples tested for all three streptococci had more than one of the species present.

There were no confirmed carrier cases identified in this study. All of the specimens containing *S. equi* were presenting clinical symptoms with the exception of two where the information was not available. In order to confirm a *S. equi* carrier the recommendation for PCR as an adjunctive test along with culture of nasopharyngeal swabs has been suggested (Newton, Verheyen et al. 2000). In that study, horses were considered clear of infection following three consecutive weekly samples from which *S. equi* was not isolated and the last of which was also negative by PCR. They further recommend endoscopic investigations of the guttural pouch in all asymptomatic horses with positive PCR results.

2.4.3. Diagnostic Multiplex PCR

The identification of *S. equi* traditionally relies upon phenotypic characteristics: morphology, serological and biochemical properties. However, significant time is required for this type of culture based diagnosis and atypical strains can lead to a misdiagnosis. This has led several researchers to develop rapid, more specific and sensitive molecular diagnostic methods, both conventional PCR and real-time PCR methods. However, to the researcher's knowledge, no test is currently available to differentiate *S. equi*, *S. zooepidemicus* and *S. equisimilis* in a single reaction. Multiplex PCR assays have been shown to be a valuable technique for the identification of viruses, bacteria, fungi, and/or parasites (Elnifro, Ashshi et al. 2000). A multiplex able to distinguish these three streptococci of interest will save time and cost when compared to doing the reactions separately. Alber et al. (2004) designed a duplex PCR to differentiate *S. equi* and *S. zooepidemicus* based on the *sodA* gene specific to both and the *seeI* gene specific to *S. equi*. Later a real-time duplex PCR was developed using these two primers to differentiate *S. equi* and *S. zooepidemicus* (Båverud, Johansson et al. 2007). The real-time duplex PCR diagnostic abilities were compared to diagnostics via conventional cultivation, they found the real-time PCR to be more sensitive as two additional field isolates of *S. equi* and four of *S. zooepidemicus* were detected.

The PCR mediated diagnostic multiplex introduced here was shown to successfully identify and differentiate *S. equi*, *S. zooepidemicus* and *S. equisimilis*. Four gene targets were used for this multiplex, making it a tetraplex PCR. These included the *seeI* primer pair specific to the *S. equi* superantigenic toxin SeeI (Alber, El-Sayed et al. 2004), the *srtz* primer pair specific to a *S. zooepidemicus* putative sortase gene (present study), the *eqsim* primer pair specific to *S. equisimilis* streptokinase precursor gene (Preziuso, Laus et al. 2010) and a primer pair amplifying a universal 16S rDNA region common to all bacteria (Boye, Høgdall et al. 1999). Inclusion of the 16S rDNA primers was used as an internal control; failure to amplify this product would indicate that an inhibitor to the PCR reaction might be present resulting in a possible false negative and a misdiagnosis.

2.4.3.1. Conventional Multiplex PCR

Optimization of the conventional multiplex PCR was achieved using the modified Taguchi's method as described in Cobb and Clarkson (1994). This allowed for eight components to be varied and tested together in just 18 reactions allowing for a prediction of a combination that will lead to an optimal reaction. If each of the eight components were to be tested independently at the three variable levels it would take 3^8 separate reactions. Similarly, researchers developing a tetraplex PCR to detect and differentiate *Mycobacterium tuberculosis* complex and nontuberculous *Mycobacteria* used the modified Taguchi method to successfully optimize their tetraplex PCR (Anilkumar, Madhavilatha et al. 2012). With the components varied in the current diagnostic multiplex, the concentration of $MgCl_2$ had the greatest effect with only the reactions containing high concentrations of Mg^{2+} resulting in the amplification of the two larger products and thus resulting in amplification of all four products.

As Taq DNA polymerase requires free Mg^{2+}, an increase of $MgCl_2$ in the reaction mix should generally have a positive effect on the PCR yield as described here and in Henegariu et al. (1997). However, excess Mg^{2+} can increase amplification of nonspecific products and can also stabilize the DNA double strand preventing complete denaturation and therefore prevent amplification (Markoulatos, Siafakas et al. 2002). Although, inadequate Mg^{2+} can reduce the amount of product and if primers or template DNA contain chelators, an increase in Mg^{2+} may be needed (Markoulatos, Siafakas et al. 2002). Henegariu et al. (1997) showed that an increase in $MgCl_2$ from 1.8-10.8 mM made the amplification of multiple products more specific with the unspecific bands disappearing, however in PCRs with up to 20 mM $MgCl_2$ the reactions were inhibited; products became barely visible.

Primer concentrations also played an important part in optimizing the PCR. All of the 18 reactions amplified the smallest length product; *srtz* (*S. zooepidemicus*) and nearly all contained the second smallest, *eqsim* (*S. equisimilis*). There seemed to be preferential amplification of the smaller length products. Therefore, a lower concentration of primer (0.15 μM) was used to amplify the smaller products that

had amplification bias, to a higher concentration of the larger products (0.2 and 0.3 μM). Although the reason for this bias was not determined, preferential amplification of PCR products over others is a known phenomenon as discussed in Markoulatos (2002) and in Elnifro et al. (2000). It is suggested that it is either due to random fluctuations in the interactions of the PCR reagents or due to properties of the target, the target's flanking sequence or the entire target genome such as GC content, secondary structures, and gene copy number.

The final optimized conventional multiplex PCR was able to accurately detect and differentiate all isolates of *S. equi*, *S. zooepidemicus* and *S. equisimilis* tested in this study. High specificity, sensitivity and diagnostic accuracy was determined for the conventional multiplex PCR of all three species as concluded by the test results of 26 submitted specimens compared to the uniplex PCR results as the reference standard. Specificity is the proportion of actual negatives correctly identified and sensitivity is the proportion of actual positives correctly identified. Diagnostic accuracy measures the degree of veracity of the test, the proportion of true results either positive or negative. The failure of the conventional multiplex PCR to detect the presence of *S. equisimilis* in one of the specimens reduced the sensitivity and diagnostic accuracy below 100 %. However, as the specimens DNA were stored for several months prior to testing with the final optimized multiplex (the stability of DNA in PEG/KOH was unknown) degradation of the DNA may have occurred and may have influenced this result. Also, as *S. equi* was not detected in any of these 26 specimens, the respective sensitivity of the multiplex could not be determined. These results should be taken as preliminary with more trials on field specimens needed to accurately test its robustness. Also, further tests could also be run to test the multiplex sensitivity such as determining the lower limit of detection based on the concentration of DNA and/or the number of colony forming units. This could be done by spiking processed field specimens from non-infected healthy horses with the streptococci of interest at various concentrations of DNA or colony forming units.

2.4.3.2. Real-Time Multiplex PCR and Melt Curve Analysis

In contrast to conventional PCR, real-time PCR has the advantage of no post-PCR handling for identification of amplicons reducing the potential for sample contamination and speeding up the diagnosis process (Monis and Giglio 2006). Melt curve analysis is analogous to amplicon detection by gel electrophoresis; the difference being that it is via closed tube verification of amplification using an intercalating dye and measuring the decrease in fluorescence of the dissociation of the amplicon fragment from double stranded DNA to single stranded. This is then converted and plotted as the negative derivative of fluorescence with respect to temperature (Ririe, Rasmussen et al. 1997; Monis and Giglio 2006). As the dissociation of an amplicon and resulting melt curve is dependent on the GC content, length and sequence; amplicons can be distinguish by their melt curves (Ririe, Rasmussen et al. 1997; Monis and Giglio 2006).

The conventional PCR multiplex developed here was trialled as a real-time PCR with the prospect of increasing sensitivity and decreasing sample handling time. It performed equally as well by accurately identifying all *S. equi*, *S. zooepidemicus*, and *S. equisimilis* isolates in this study using purified DNA as templates. However, it was unable to accurately amplify and detect all products when trialled on the 26 specimens PEG/KOH extracted DNA, using the uniplex conventional PCR as a standard. The melt curve analysis gave three false positives by identifying the presence of three *S. equi* that were later confirmed as *S. equisimilis* when the reaction was run on an agarose gel. The melt curve analysis also failed to detect any of the eight specimens containing *S. equisimilis* or any 16S rDNA amplicons. The gel results of the real-time PCR multiplex were slightly better due to detecting more of the amplicons and accurately differentiating *S. equi* from *S. equisimilis*. These results indicated that there was interference with the fluorescence; as all the bands produced on the agarose gel were not detected with melt curve analysis. This was theorized as possibly being due to the processing method of the specimens. The specimens were in a high alkaline PEG/KOH solution (pH of 13.5), from which 1 μl was used as the template for the reaction, this solution might have had an inhibitory affect on the

intercalating dye. Further, as the melt curve analysis for three *S. equisimilis* specimens were more indicative of *S. equi* the PEG/KOH may have also had an affect on the temperature required for disassociation of the double stranded DNA. A 10-fold dilution (as used in this study) of the PEG/KOH solution in standard PCR buffer results in a decrease of pH to 8.63, well within the effective range for PCR and has been shown to support DNA amplification in a conventional multiplex PCR (Chomczynski and Rymaszewski 2006). The only difference of the reaction components in the conventional PCR and the real-time PCR was the SYTO® 82 intercalating dye. Therefore, the decreased specificity, sensitivity and diagnostic accuracy was thought to be the cause of the relationship between the PEG/KOH template DNA mix and the SYTO® 82 intercalating dye. To further develop the real-time multiplex PCR, an investigation into the actual cause of the inhibition and into techniques to reduce these effects is needed. One possibility to improve the real-time PCR diagnostic ability is further purification of the field specimen's DNA.

2.4.4. Diagnostic Multiplex PCR Conclusions

A review by Sweeney et al. (2005) has listed PCR methods as useful for the detection of asymptomatic carriers, to establish *S. equi* infection status prior to transport and commingling, as well as to determine the success of elimination of *S. equi* from the guttural pouch. However, as PCR will detect dead organisms it is liable to yield false positive results and therefore culture of the bacteria is needed to confirm infection status. The multiplex PCR developed here along with culture has the potential to enhance the detection of *S. equi* and the diagnosis of upper respiratory tract infections in the New Zealand horse population.

As the real-time multiplex PCR did not perform as well as the conventional multiplex PCR on submitted specimens with respect to specificity, sensitivity and diagnostic accuracy; the multiplex via conventional PCR is recommended for future diagnostics.

The use of this multiplex diagnostic tool has potential of use beyond analysing upper respiratory tract infections. Another application can be in determining the identity of β-haemolytic streptococci associated with endometritis. One study using PCR based methods, found that over 90 % of β-haemolytic isolates from the uterus of mares suffering endometritis identified as *S. zooepidemicus*, the other 5 % identified as *S. equi* and *S. equisimilis* (Proietti, Bietta et al. 2011). If screening for just *S. equi* is required, the uniplex PCR with the species-specific primers can be used. It must be noted that there will be an increased possibility of being unable to detect false negatives due to the absence of the 16S rDNA primer internal control.

3. Typing of New Zealand *S. equi* Strains

3.1. Introduction

In epidemiological studies, researchers have trialled various methods to discriminate between *S. equi* isolates, many with minimal success. A multilocus enzyme electrophoresis study found that 69 out of 70 *S. equi* isolates fell into a single electrophoretic type (Jorm, Love et al. 1994). Multilocus sequence typing, using five genes was unable to find any variation in 70 clinical isolates of *S. equi* from different geographical locations (Kelly, Bugg et al. 2006). Immunoblotting, restriction length polymorphism analysis and southern blotting also failed to detect any differences among 21 isolates (Galán and Timoney 1988). The 16S-23S RNA gene intergenic spacers were identical in *S. equi* isolates with no size or sequence variation within the spacer region resulting in the same subtype (Chanter, Collin et al. 1997; Hassan, Khan et al. 2003). One can deduct from these results that *S. equi* is a highly homogeneous, colonial species.

Regardless of the failures, there has been some success in typing *S. equi*. Pulsed-field gel electrophoresis (PFGE) was able to type 20 isolates into seven distinct groups (Takai, Anzai et al. 2000). Another study was done comparing several molecular methods: ribotyping, PFGE, enterobacterial repetitive intergenic consensus (ERIC) PCR, repetitive extragenic palindrone PCR (rep-PCR), BOX element (BOX) PCR and DNA sequencing of the regulatory region of *S. equi's szp* gene which encodes the M-like protein SzP (Lanka, Borst et al. 2010). The researchers found that only PFGE and the *szp* sequencing could distinguish the modified live vaccine strain from wildtype strains. PFGE was further able to

distinguish wildtype strains from each other. In contrast to this study, researchers were able to differentiate between *S. equi* isolates, obtained from various worldwide geographical locations, using rep-PCR as well as differentiate vaccine strains from wildtype strains (Al-Ghamdi, Kapur et al. 2000; Al-Ghamdi, Amonsen et al. 2008). Still, these successful methods are not ideal. PFGE is highly labour intensive, the reproducibility of rep-PCR tends to be unreliable and *szp* sequencing can only distinguish the vaccine strain from wildtype strains.

Sequencing of the *seM* gene is currently the method of choice for typing *S. equi*. This method involves sequencing of the hypervariable region found following the N-terminal signal sequence (Figure 3.1) and identifying single nucleotide polymorphisms (SNPs), thereby discriminating SeM alleles resulting in different isolate subtypes. A database of SeM alleles has been established, (http://pubmlst.org/szooepidemicus/seM); currently there are 114 recorded alleles from around the world (accessed January 21, 2012). This method of subtyping has been shown to be more sensitive then PFGE (Lindahl, Soderlund et al. 2011). It also has been recommended and used successfully in discriminating outbreaks of strangles (Anzai, Kuwamoto et al. 2005; Kelly, Bugg et al. 2006; Ivens, Matthews et al. 2011; Lindahl, Soderlund et al. 2011; Parkinson, Robin et al. 2011).

The *seM* gene (1,605 bp) produces a novel cell wall associated M-like protein, unique to *S. equi*, which has been a proposed link for the increased virulence of *S. equi* over its archetype *S. zooepidemicus* (Timoney, Artiushin et al. 1997). This M-like protein is important to the survival of *S. equi* due to its antiphagocytic activity; it binds to host fibrinogen (Fb) and immunoglobulin G (IgG) blocking access of C3b (Timoney 2004) (Figure 3.1). C3b is a component of the complement component 3 (C3) protein of the host immune system, which marks microbials for phagocytosis. A study found that a deletion in *seM* was associated with persistence in hosts; 24 % of *S. equi* isolates from outwardly healthy horses contained this deleted region of *seM* resulting in a truncated protein (Chanter, Talbot et al. 2000) (Figure 2.2). The deletion occurs within the Fb and IgG binging region (Figure 3.1) and was found to be associated with decreased

resistance to phagocytosis (Chanter, Talbot et al. 2000). A recent retrospective study describes an isolate containing a single nucleotide insertion in the *seM* gene giving rise to a premature stop codon and therefore a putative truncated SeM protein (Parkinson, Robin et al. 2011). This isolate was obtained from the guttural pouch of an asymptomatic carrier.

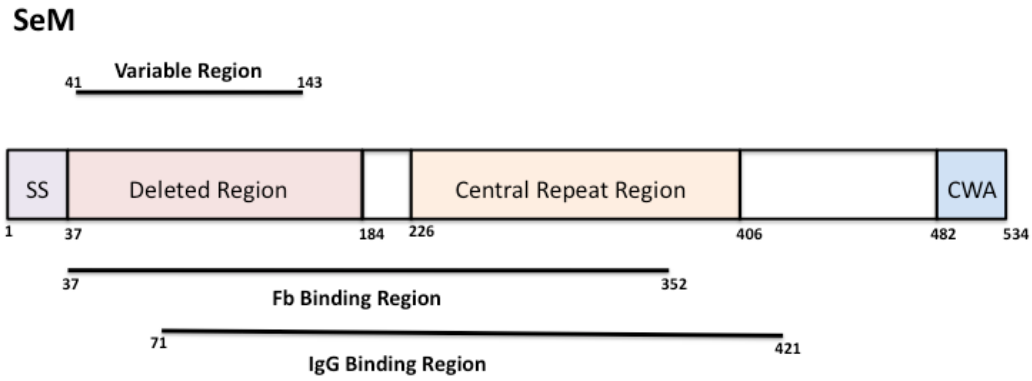


Figure 3.1: A schematic representation of the *S. equi* SeM protein, identifying the variable N-terminal sub-typing region. Numbers represent amino acids. The signal sequence (SS), deleted region in some carriers, central repeat region, cell wall anchor (CWA) region, as well as the Fb and IgG binding regions are all indicated. Adapted from (Kelly, Bugg et al. 2006). Image used with permission from copyright holder American Society for Microbiology and corresponding author.

3.1.1. Hypotheses

The objective of this study was to use the New Zealand *S. equi* isolates obtained in the previous molecular diagnostic development to determine if more than one strain exists in New Zealand. Two epidemiological methods were employed, the SeM allele sequencing method and a Sau-PCR amplification method, which previously has not been trialled on *S. equi* isolates. Sau-PCR is based on the digestion of genomic DNA with the restriction endonuclease Sau3AI and subsequent amplification using primers with a core sequence based on the Sau3AI recognition site (Corich, Mattiazzi et al. 2005).

Hypothesis 1:
New Zealand may have different *S. equi* strains in circulation based on SeM allele subtyping and that these strains may differ from those already recorded in the SeM allele database.

Hypothesis 2:
The Sau-PCR amplification method could provide an alternative subtyping method for *S. equi* or be used in conjunction with SeM allele subtyping to further differentiate strains.

3.2. Methods

3.2.1. Samples

All 19 isolates of *S. equi* were used in the typing trials (Table 2.2), inclusive of the Pinnacle® IN vaccine isolate.

3.2.2. DNA Isolation

In order to normalize across samples, DNA was extracted and purified using a GITC/column method and an equal concentration of purified DNA from each isolate was used in the typing trials. A single colony was grown overnight in 10 ml of either BHI or THY (Appendix A.I) at 37 °C. The entire 10 ml was then centrifuged at 1,200 rcf for 10 mins. All of the supernatant was removed; the bacterial pellet was resuspended in 100 µl TE and transferred to a 1.7 ml microfuge tube (Axygen). Then 20 µl of lysozyme (100 mg/ml) (Boehringer) was added and the suspension incubated at 37 °C for 30-45 mins in a thermomixer (Eppendorf). Once the incubation was done, 350 µl of SDS lysis solution (Appendix A.I) and 20 µl of Proteinase K (60 µg/µl) (Roche) was added and the suspension further incubated at 65 °C in a thermomixer for 1-2 hrs to complete cell lysis. To the lysate, 350 µl of 5 M LiCl was added and mixed before the addition of 750 µl of chloroform and vortex for 20 secs to form an emulsion. The tubes were placed on a rotating wheel for 15 mins and centrifuged at 16,000 rcf for 10 mins, after which the top aqueous phase of the biphasic solution containing the DNA was transferred to a new 1.7 ml microfuge tube.

To precipitate the DNA, an equal volume of isopropanol was added, mixed by inversion and then incubated at -20 °C for 15 mins. The tubes were then spun at 16,000 rcf for 15 mins at 4 °C (5415R Bench Top Centrifuge, Eppendorf) to pellet

the DNA. All the alcohol was then removed and the DNA resuspended in 50 µl TE. Once dissolved, four volumes of 5M GITC solution (Appendix A.I) was added and mixed. This mixture was placed in a column (Genscript spin columns, QuickClean PCR Purification Kit) and centrifuged and wash as per manufacture's directions. After the wash, 50 µl of heated (~50 °C) TE was added to the column and left to sit for 1 min to elute the DNA. The DNA was recovered by spinning for 2 mins at 16,000 rcf. All DNA samples were stored at 4 °C until subsequent analysis.

For comparison, variations of the GITC/column method were trialled: A, the method in its entirety (GITC/column); B, the method using the GenScript binding solution in place of GITC (non-GITC/column); and C, stopping at the alcohol precipitation and then resuspending in TE (non-GITC/non-column).

3.2.3. DNA Sample Quantity and Quality

The quantity of nucleic acids was calculated by measuring absorbance at a wavelength of 260 nm using a Nanodrop ND-1000 spectrophotometer (Nanodrop) within 2 µl of each sample, reflecting the total nucleic acid concentration present in ng/µl. The ratio of sample absorbance at 260 and 280 nm was used to assess the purity of the DNA with an optimal 260/280 ratio range of 1.8 – 2.0. To confirm the Nanodrop results, 5 µl of each sample was electrophoresed in a 1.0 % (*w/v*) agarose gel in SB buffer (Appendix A.I) and 1.5 µg/ml of ethidium bromide (USB) using an Owl gel electrophoresis system.

3.2.4. PCR Components

The PCR master mix used in the PCR based typing methods was the same unless otherwise indicated. It was prepared with Solis Biodyne components; 10 x Buffer B1, 25 mM $MgCl_2$, and 20 mM dNTP mix made up in MQ-H2O (Barnstead). Final concentrations are shown in Table 3.1. The reaction mix was aliquoted into small volumes and stored at -18 °C until required at which time the Hot FIREPol® DNA Polymerase (Solis BioDyne) was added. Primers, primer concentration and template DNA concentration depended on typing method used and are noted in the following sections.

Table 3.1: Components of typing PCR master mix

Typing PCR Reaction Master Mix	
Components	**Final Concentration**
10 x Buffer B1	1 x
$MgCl_2$ (25 mM)	2 mM
dNTP mix (20 mM each)	0.25 mM
Hot FIREPol® DNA Polymerase (5 U/μl)	0.5 U/20 μl reaction
Primers and Templates	Variable

3.2.5. *seM* Sequence Typing

3.2.5.1. *seM* PCR protocol and sequencing

The primers used to amplify 541 bps of the 5' variable region of the *seM* gene were *seM* F and *seM* R (Kelly, Bugg et al. 2006) (Appendix A.I). The primers were used at a concentration of 0.2 μM with 0.5 U Hot FIREPol® DNA Polymerase in a 20 μl reaction made up in the mastermix (Table 3.1). Between 50 and 100 ng/μl of template DNA was used for each reaction. The amplification protocol consisted of an initial activation/denaturation at 95 °C for 15 mins followed by 40 cycles of denaturing at 95 °C for 20 secs, annealing at 53 °C for

20 secs and elongating at 68 °C for 30 secs. A final elongation was done at 68 °C for 5 mins. The PCR products were cleaned up by adding 1 U of rAPid Alkaline Phosphatase (Roche) and 10 U exonuclease1 (Fermentas) to each 20 µl reaction and incubating at 37 °C for 30 mins followed by 85 °C for 15 mins. Sequencing of both forward and reverse strands with the *seM* primers used in the initial PCR amplification was carried out by the Waikato DNA Sequencing Facility (University of Waikato, Hamilton, New Zealand) using BigDye fluorescent terminators (Applied Biosystems Big Dye v 3.1.).

3.2.5.2. *seM* sequencing and phylogenic analysis

Forward and reverse sequences for each isolate were aligned and manually edited using Geneious v5.4 (Drummond, Ashton et al. 2011). The resulting high quality double stranded sequence data was compared to those of the SeM alleles in the PubMLST SeM database (http://pubmlst.org/szooepidemicus/seM). New alleles were identified by one or more SNP, assigned allele numbers and uploaded to the database. A chi-square test (χ^2-test) was used to determine if the SeM alleles found in New Zealand were more likely to occur on one island over the other. This was based on an expected equal distribution and a $P < 0.05$ considered a significant diversion from the expected equal distribution.

Using all the nucleotide *seM* sequences in the database, a maximum likelihood tree was constructed using MEGA v5.05 (Tamura, Peterson et al. 2011) and the bestfit model (HKY+I+G) selected by jModeltest 0.1.1 (Posada 2008) which uses the Phyml alogrithm (Guindon and Gascuel 2003). The robustness of the isolate grouping was tested by use of bootstrap resampling with 1,024 replicates.

3.2.6. Sau-PCR Typing

3.2.6.1. Sau-PCR DNA Prep

Purified DNA was digested with the restriction endonuclease MboI (NEB) an isoschizomer of Sau2AI. The digestion mix consisted of 200 ng DNA, 10 U MboI, and 1 x NEBuffer 4 made up in MQ-H_2O (Barnstead). It was digested at 37 °C overnight and heat inactivated at 65 °C for 20 mins.

3.2.6.2. Sau-PCR Protocol

The PCR reaction was performed in 20 µl of master mix (3.2.4) with 0.5 U Hot FIREPol® DNA Polymerase (Solis Biodyne), and 0.6 µM of one of the Sau-PCR primers (Appendix A.I). For the template, 2 µl of MboI digested DNA was added.

A graphical representation of the amplification protocol used is summarized in (Figure 3.2); some modifications were made as listed below.

The initial 'filling in'step, fills in the Sau3AI-cut ends starting with an activation/denaturation at 95 °C for 15 mins followed by 24 °C for 5 secs, ramping at 0.1 °C/sec to 60 °C and then holding at 60 °C for 30 secs. Next was the 'low-stringency cycle', which allows for the GC tail formation on both ends of each Sau3AI fragment. This phase was cycled through twice starting with denaturation at 95 °C for 60 secs, annealing at 50 °C for 15 secs, ramping at 0.1 °C/sec to 25 °C, ramping at 0.1 °C/sec to 50 °C and then holding at 50 °C for 30 secs. The 'high-stringency phase' follows, amplifying the Sau3AI genomic regions flanked by the GC tails. This phase was cycled through 40 times with denaturation at 95 °C for 15 secs, annealing at 48 °C for 60 secs and extension at 65 °C for 2 mins. The final extension step was at 65 °C for 5 mins.

Once amplification was done, 10 µl of the product was electrophoresed in a 2.0 % (*w/v*) agarose gel in TAE buffer (Appendix A.I) and 1.5 µg/ml of ethidium bromide (USB) using an Owl gel electrophoresis system.

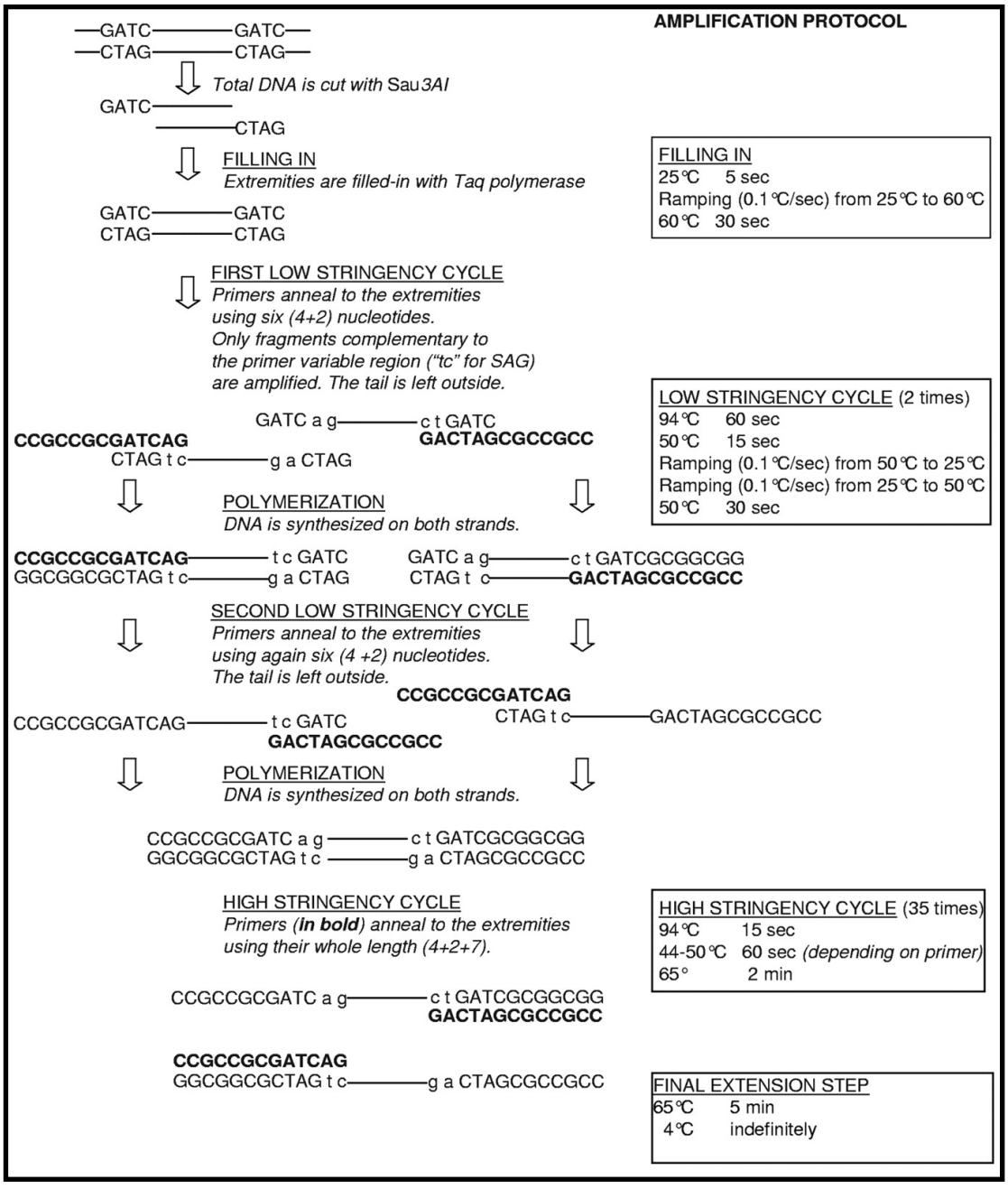


Figure 3.2: Graphical representation of the Sau-PCR using the SAG primer (bold). Small letters indicate nucleotides that are complementary to the variable part of the primers and that therefore differ depending on the primer used (Corich, Mattiazzi et al. 2005). Image used with permission from copyright holder American Society for Microbiology.

3.3. Results

3.3.1. DNA Isolation and Purification

The GITC-column method gave high quality DNA (Figure 2.2). Although the resulting concentration of the nucleic acids is lower then the other methods trialled (Table 3.2), the amount of high quality DNA is much greater (Figure 3.3). The non-GITC/column and non-GITC/non-column methods resulted in lower quality of DNA with RNA and degraded DNA (smears and lower bands in figure Figure 3.3) making up most of the nucleic acid concentrations in the nanodrop results (Table 3.2). DNA for each *S. equi* isolate was extracted using the GITC/column method resulting in similar quality and purity as shown in these results and used for further analysis for the typing methods.

Table 3.2: Nanodrop results comparing the DNA purification methods: A, GITC/column; B, non-GITC/column; and C, non-GITC/non-column (see methods sections 3.2.2 and 3.2.3).

DNA Purification Trial Nanodrop Results		
Method	Nucleic Acid Concentration ng/μl	260/280 ratio
A	49.43	1.74
B	192.51	2.00
C_1	832.78	1.96
C_2	2412.8	1.98

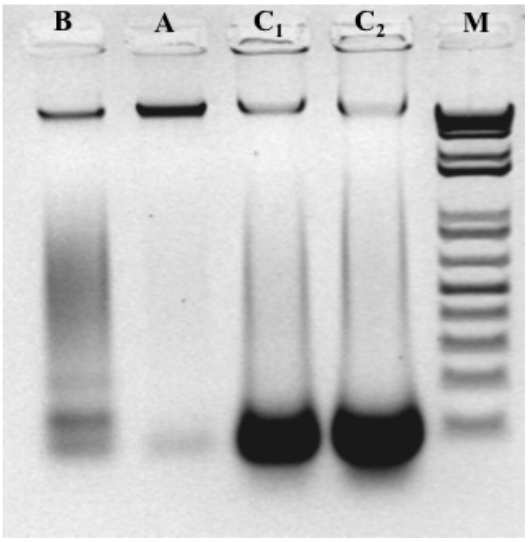


Figure 3.3: Gel results comparing purification methods: A, GITC/column; B, non-GITC/column; and C_1 and C_2, non-GITC/non-column (see methods sections 3.2.2 and 3.2.3).

3.3.2. New Zealand SeM alleles

PCR of the *seM* genes of all 18 clinical isolates of *S. equi* and the Pinnacle® IN vaccine strain generated the expected product size of 541 bp. Comparison of the resulting *seM* sequences with the SeM alleles sequence data, on the SeM database, identified two new alleles novel to New Zealand numbered 99 and 100 (Table 3.3, Figure 3.4). Further, the dry and mucoid colony isolates of clinical specimens from the horses, which became symptomatic after receiving the Pinnacle® IN vaccine (Table 2.2), typed as *seM* 2; identical to the vaccine (Table 3.3). The novel strains were added to the SeM database. The nucleotide and peptide sequences of all 114 currently identified SeM alleles are shown in Appendix (A.IV). Allele 99 was found to only occur in the north island of New Zealand whereas, allele 100 occurred in both the north and south islands (Table 3.4). With half of the isolates, containing SeM allele 100, occurring on the North Island and the other half on the South; there was no significant difference (χ^2-test, P = 1) in its occurrences between the two islands based on an expected equal distribution.

SeM allele 99 was significantly more likely to occur on the North Island then the South (χ^2-test, P = 0.014) as all six isolates containing the SeM allele 99 were only found on the North Island (Table 3.4).

Table 3.3: SeM allele types from clinical isolates

SeM allele	Isolate	Source	Island	Sample Date**	Outbreak[†]
99	OP1	Matamata	North	19/10/10	A
99	OP2	Cambridge	North	24/12/10	B
99	OP3	Cambridge	North	15/12/10	B
99	OP4	Cambridge	North	19/01/11	B
99	OP5	Cambridge	North	17/01/11	B
99	OP14	Bulls	North	18/08/11	I
100	OP6	Matamata	North	10/03/11	C
100	OP7	Palmerston North	North	10/06/11	D
100	OP8	Palmerston North	North	10/06/11	D
100	OP9	Rangiora	South	17/05/11	E
100	OP10	Rangiora	South	4/07/11	E
100	OP11	Matamata	North	22/07/11	F
100	OP12	Mossgiel	South	16/06/11	G
100	OP13	Rangiora	South	29/07/11	H[¥]
2	OP15, OP16	Otago	South	27/07/11	NA
2	OP17, OP18	Christchurch	South	27/06/11	NA
2	OP19	Pinnacle® IN vaccine	NA*	14/10/10	NA

*NA: not applicable; **day/month/year; [†]Outbreaks were considered different based on sample date and location. [¥]Outbreak H was considered separate from outbreak E as they were from separate properties.

Table 3.4: Occurrences of novel New Zealand SeM allele isolates and outbreaks in the North and South Islands.

Novel NZ SeM Allele	Number of *S. equi* Isolates		
	North Island	South Island	Total
99	6	0	6
100	4	4	8
Total	10	4	14
	Number of *S. equi* Outbreaks		
	North Island	South Island	Total
99	3	0	3
100	3	3	6
Total	6	3	9

Isolates OP2-OP5 were obtained from the same outbreak in Cambridge. They all typed as SeM allele 99 and make up four of the six SeM 99 isolates obtained during the study (Table 3.3). Isolates OP7 and OP8 were isolated from the same outbreak in Palmerston North and isolates OP9 and OP10 were from the same outbreak in Rangiora; *S. equi* subtype, SeM allele 100 was the cause of these two outbreaks (Table 3.3). The rest of the isolates, based on the location and the date the samples were taken on, were considered to be separate outbreaks. This resulted in a total of nine outbreaks from which isolates were obtained (this excludes the vaccine related isolates, SeM allele 2). Taking the nine outbreaks into account, with all the isolates from the same outbreak counting as one, then the significance of only finding allele 99 on the North island becomes non-significant due to a low sample number (χ^2-test, P = 0.083). Allele 100 remains at equal distributions between the islands. Matamata is the only location from which both New Zealand *S. equi* subtypes 99 and 100 were isolated, however they were not isolated during the same outbreaks.

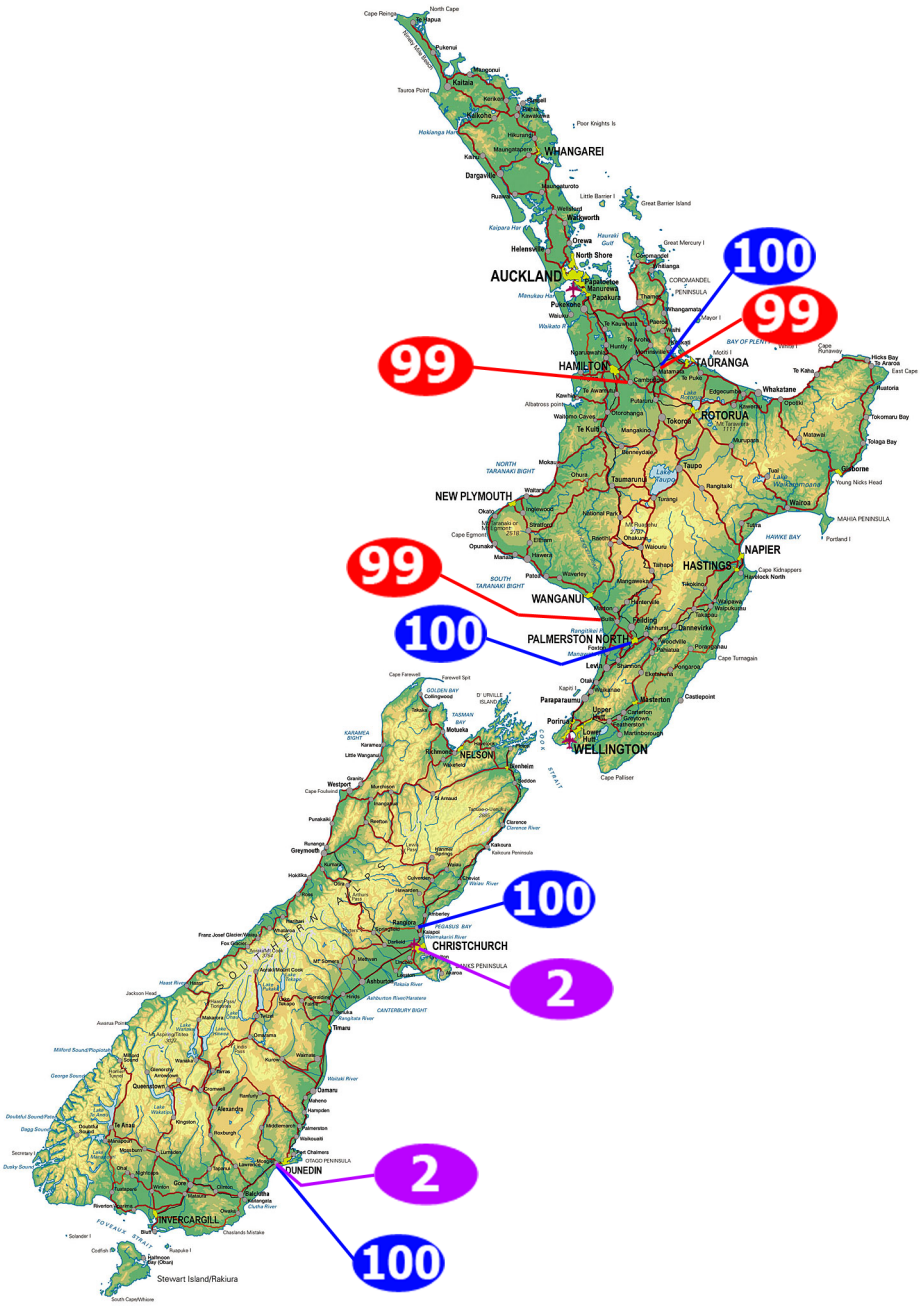


Figure 3.4: New Zealand locations of SeM alleles; *seM* 2, 99 and 100. Map adapted from Land Information New Zealand (LINZ) (www.linz.govt.nz/).

There are three SNPs that differentiate these two SeM alleles from each other (Figure 3.5). Only one of these is unique amongst all 114 SeM alleles and is found in allele 100 (Figure 3.5 and Appendix A.III). All three SNPs are nonsynonymous substitutions resulting in an altered amino acid sequence (Figure 3.5). The vaccine allele 2 is shown for comparison.

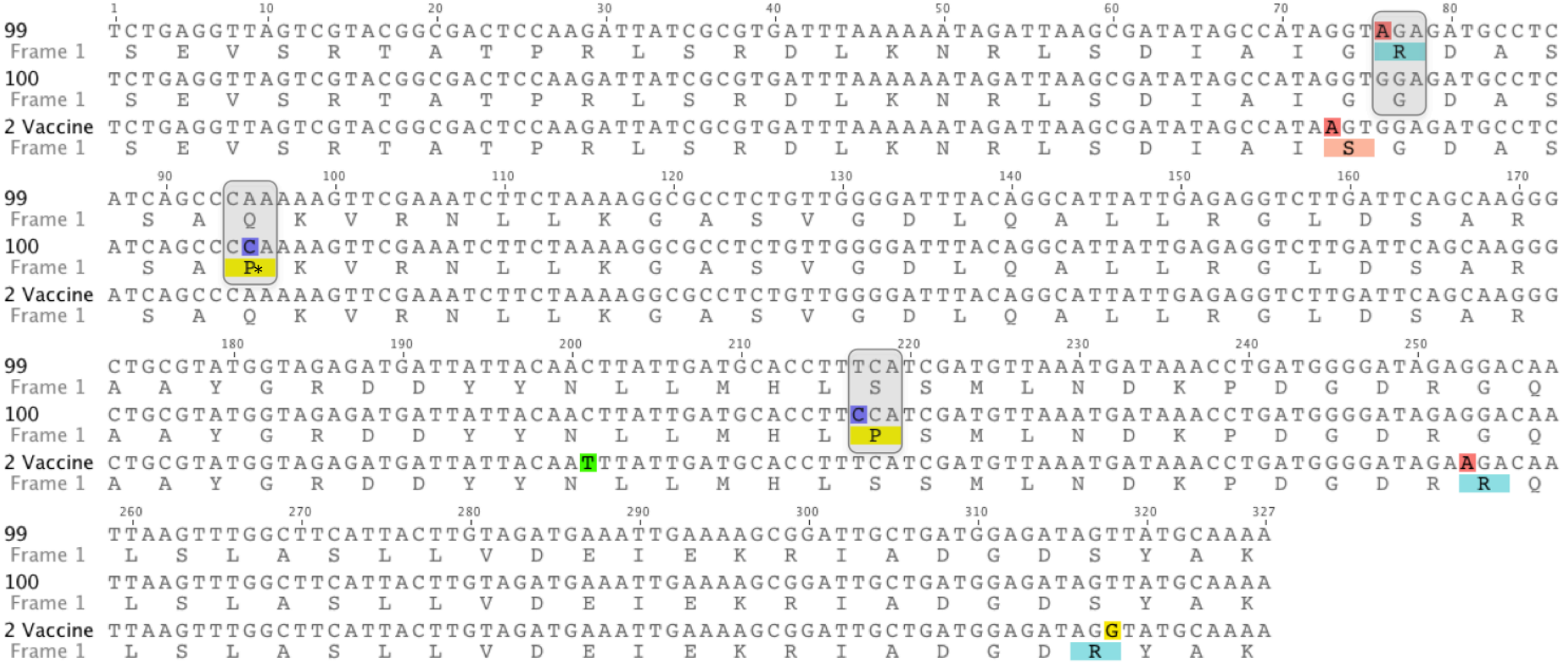


Figure 3.5: Alignment of New Zealand SeM allele subtypes 99, 100 and 2 (Pinnacle® IN vaccine). For each allele subtype the top row shows the DNA sequence and the bottom row is the resulting amino acid sequence (Frame 1). Differences in the codes are highlighted and the three SNP's resulting in different amino acids between allele 99 and 100 are shaded grey. The starred (*) SNP and resulting amino acid sequence in *seM* 100 is the only unique SNP of the three identified here among all 114 SeM alleles (supplement 114 alignment). The vaccine is shown for comparison.

Two clinical swabs from different horses resulted in *S. equi* isolates with dry and mucoid colony morphologies each (OP15, OP16 and OP17, OP18 respectively, see Table 2.2). The *seM* sequence of both the dry and mucoid isolate morphologies from each of these horses identified as SeM allele 2, which is identical to the Pinnacle® IN vaccine allele.

3.3.3. *seM* 99 and 100 Sequence and Phylogenic Analysis

All 114 SeM allele sequences available in the database are shown in Appendix (A.IV) and the alignment of these sequences is shown in Appendix (A.III). Based on nucleotide sequence alignments, New Zealand novel SeM alleles 99 and 100 shared the highest sequence identity with *seM* 89, with 99.7 % and 98.8 % respectively. Although poorly supported by bootstrap analysis, < 50 % of replicate trees, the maximum likelihood tree based on the nucleotide sequences maintained this relationship (Figure 3.6). SeM allele 89 was isolated in Pakistan (2010) and in Germany (1991) (Appendix A.VI). This phylogenic tree also shows that the New Zealand strains are related to *seM* 62 and 86; *seM* 99 shares 99.4 % and 93.3 % sequence identity with alleles 62 and 86 respectively, *seM* 100 shares 98.5 % sequence identity with both alleles. SeM allele 62 was isolated in the UK (2008) and *seM* 86 was isolated in Sweden (1990) (Appendix A.VI).

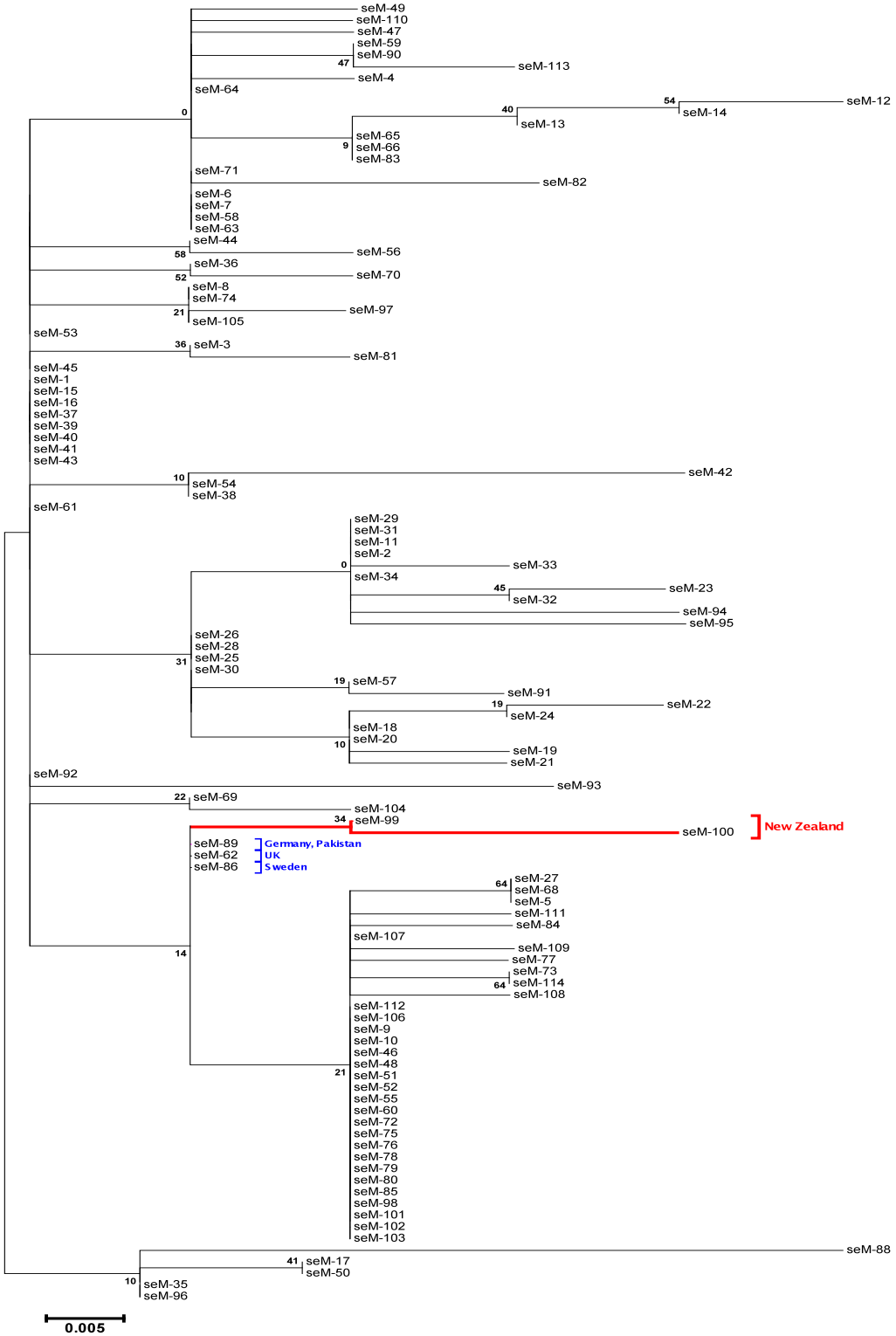


Figure 3.6: Maximum likelihood tree generated on the nucleotide sequences of the SeM allele types of *S. equi*. The scale bar represents the number of substitutions per site. Bootstrap values of 1,024 replicates trees are shown next to branches as the percentage of trees supporting the associated taxa cluster. Colouring represents New Zealand alleles (red) and the alleles highly similar (blue).

3.3.4. Sau-PCR Typing Results

Out of the seven Sau-PCR primers trialled, the primer SAUT was determined to be the best one able to clearly differentiate between the Pinnacle® IN vaccine isolate and the clinical *S. equi* isolates (Figure 3.7). This confirmed the use of only SAUT in further Sau-PCR typing. One of the *S. equi* isolates used was OP3, which was later typed as SeM allele 99 and the other was OP7, SeM allele 100 (Table 3.3); however, none of the seven Sau-PCR primers were able to differentiate these two isolates (Figure 3.7, Table 3.5).

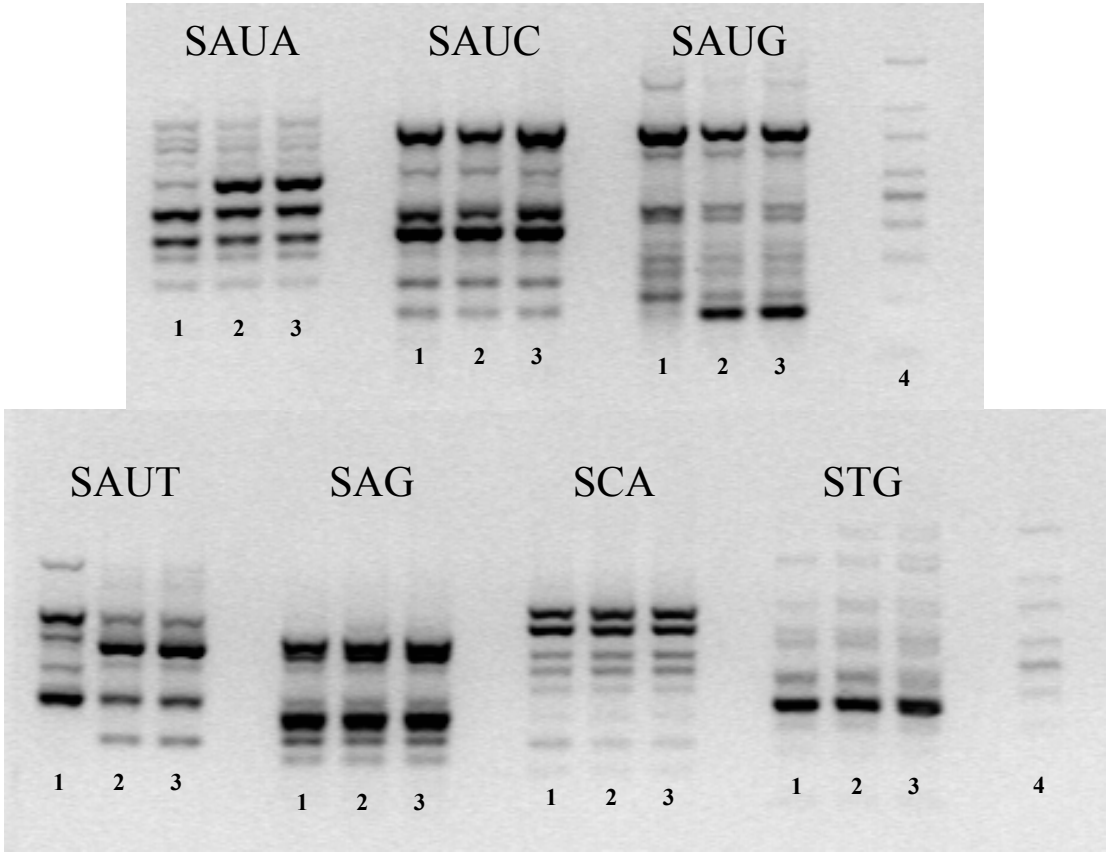


Figure 3.7: Sau-PCR primer trial. Seven Sau primers (indicated in capital letters above the lanes) were trialled with 3 *S. equi* isolates: (1) Pinnacle® IN Vaccine; (2) OP3, SeM allele 99; and (3) OP7, SeM allele 100 (refer to Table 2.2).

Using the SAUT primer, all *S. equi* isolates were tested using Sau-PCR. This method clearly identified two strains, labelled as Sau1 and Sau2 (Figure 3.8). Isolates OP15-OP18 (both the dry and mucoid colony morphologies, refer to Table 2.2 for isolate details) resulted in strain Sau2, the same strain as the Pinnacle® IN vaccine. All the other isolates typed as Sau1 (Table 3.5, Figure 3.8).

Table 3.5: Comparison of isolates SeM allele and Sau-PCR subtypes

Isolate	SeM allele Subtype	Sau Subtype
OP1	99	Sau1
OP2	99	Sau1
OP3	99	Sau1
OP4	99	Sau1
OP5	99	Sau1
OP14	99	Sau1
OP6	100	Sau1
OP7	100	Sau1
OP8	100	Sau1
OP9	100	Sau1
OP10	100	Sau1
OP11	100	Sau1
OP12	100	Sau1
OP13	100	Sau1
OP15, OP16	2	Sau2
OP17, OP18	2	Sau2
OP19 (Pinnacle® IN vaccine)	2	Sau2

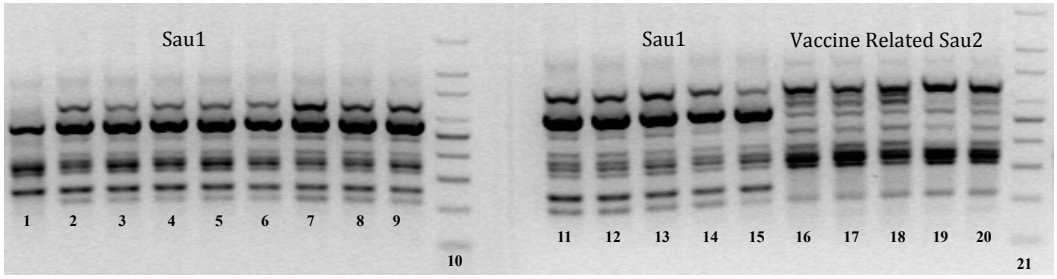


Figure 3.8: Sau-PCR results using SAUT primer. All *S. equi* isolates (Table 2.2) were digested with MboI and amplified with the SAUT primer. Sau-PCR strain, Sau1 is in lanes 1-9 and 11-15 from *S. equi* isolates: OP6, OP1, OP2, OP3, OP4, OP5, OP7, OP8, OP9, OP10, OP11, OP12, OP13 and OP14 respectively. Sau-PCR strain Sau2 is in lanes 16-20 from *S. equi* isolates: OP15, OP16, OP17, OP18 and OP19 respectively. Lane 20 is a Pinnacle® IN vaccine isolate. A 100bp ladder (Solis BioDyne) was loaded into Lanes 10 and 21.

3.4. Discussion

3.4.1. New Zealand Novel SeM Alleles

The variable N-terminal region of the antiphagocytic M-protein SeM has been suggested and used as a tool for epidemiological studies of *S.* equi (Anzai, Kuwamoto et al. 2005; Kelly, Bugg et al. 2006; Ivens, Matthews et al. 2011; Parkinson, Robin et al. 2011). In the present study, a total of 19 *S. equi* isolates were obtained totalling three SeM alleles. Six identified as the novel New Zealand strain *seM* 99 consisting of three outbreaks on the North Island. Eight identified as the novel New Zealand strain *seM* 100 consisting of three outbreaks on the North Island and three outbreaks on the South Island. The remaining isolates were vaccine related, identifying as the strain *seM* 2.

It appears that *seM* 100 is more prevalent in New Zealand causing six out of the nine outbreaks in this study and appearing on both islands; although, due to the low numbers more isolates would be needed to confirm the dominance of *seM* 100. Interestingly, a recent retrospective study (Parkinson, Robin et al. 2011) found that there were significant changes in allele prevalence between 2007, 2008 and 2010 in the UK with an increasing prevalence in *seM* 9 related alleles. This corresponded with a decreasing prevalence in *seM* 6 related alleles. It was suggested that this increased dominance of one allele group over others could be related to an increase in fitness. They theorize that the SeM 9 allele may have an enhanced ability to survive and persist in the guttural pouch, providing greater opportunity to cause recurrent outbreaks or benefit from enhanced transmission, such as a lower infective dose being required to cause infection. As there was no vaccine against strangles marketed in the UK when these *S. equi* strains were isolated, the positive selective pressure on dominant *seM* 9 was not related to enhanced fitness against a vaccine (Parkinson, Robin et al. 2011).

The variable *seM* region was first identified by Anzai et al. (2005) and Kelly et al. (2006). Both studies found a much higher proportion of non-synonymous variations than synonymous. Data from other studies (Waller and Jolley 2007; Ivens, Matthews et al. 2011; Lindahl, Soderlund et al. 2011) further support this with relative frequencies of non-synonymous and synonymous substitutions (d_N/d_s) all >1 and some much greater, giving indication that this portion of the *seM* gene is under strong diversifying selective pressure. The d_N/d_s ratio is used as a measure of selective pressure at the protein level, with a ratio >1 indicative of positive selective pressure. This pressure was presumably exerted by the host environment and lead to a dominance of those alleles with greater fitness.

The increase in fitness of one *S. equi* strain over another may not be entirely due to *seM* variants; other gene variants may also have a role in fitness. One such gene is the thymidylate kinase housekeeping gene (*tdk*). A multilocus sequence typing scheme involving seven housekeeping genes including *tdk* found that 23 of their 24 *S. equi* isolates were of the same sequence type (ST) 179 and the remaining isolate was of ST-151 (Webb, Jolley et al. 2008). ST-179 and ST-151 are differentiated only by a variation in *tdk* (*tdk* allele 45 and *tdk* allele 23 respectively) as designated in the *S. zooepidemicus* MLST database (http://pubmlst.org/szooepidemicus/). The other six housekeeping genes were of the same allele type in all 24 *S. equi* isolates (Webb, Jolley et al. 2008). Parkinson et al. (2011) indicate that the dominant SeM 9 strains in the UK could be further differentiated on the basis of a SNP in *tdk* and any increase in fitness may not be due to SeM variations alone. It would be interesting to determine the *tdk* alleles of the New Zealand strains, *seM* 99 and 100 and determine if there are any correlations with the apparent increased fitness of *seM* 100.

In the present study, *seM* 99 and 100 were found to be closely related to *seM* 89, which suggest that the New Zealand alleles originated and diverged from SeM allele 89 originally isolated in Germany in 1991 and later in Pakistan in 2010. Out of the two New Zealand alleles, *seM* allele 100 was found to be more divergent from allele 89. It also contained a unique SNP resulting in a different amino acid, as compared to all 114 SeM alleles analysed in this study. These

findings give suggestive evidence that it may have an increased fitness within the New Zealand horse population and may be indicative of why *seM* 100 is more prevalent. It would be interesting to ascertain if allele 100 has increased virulence abilities and if any isolates from carriers are allele 100 or are more closely related to allele 100 than allele 99. As there are currently two vaccines used in New Zealand, Equivac® S and Pinnacle® IN, one or both could be driving the selective pressure. Allele 100 may have a greater fitness against the vaccines. Future test could be focused around the effectiveness of the vaccines on these two New Zealand alleles.

It must be noted that the Maximum likelihood phylogenetic tree produced based on the HKY+I+G model is poorly supported by bootstrap analysis. This indicates that the allele clusters generated are not robust and therefore relatedness between alleles extracted from this data is only suggestive and should be interpreted with care. This result is most likely due to the low numbers of SNPs between alleles. Regardless, SeM alleles 18, 19, 20 and 21 were all isolated in Japan and are clustered together within this phylogenetic tree; New Zealand alleles clustered together and the Brazil alleles (40 and 41) remained together (Figure 3.6 and Appendix A.VI). This gives evidence that related alleles are geographically linked and has also been shown in other studies using other phylogenetic tree methods (Anzai, Kuwamoto et al. 2005; Ivens, Matthews et al. 2011; Parkinson, Robin et al. 2011).

3.4.2. Affects of SeM Variation

SeM is important to the survival of *S. equi* due to its antiphagocytic activity; it binds to host Fb and IgG blocking access of opsonin molecules thus affecting recognition by the host immune system (Timoney 2004). The effects of the SeM variation among *S. equi* isolates were attempted to be ascertained by Timoney et al. (2010). Their results suggested that the SeM variation did not significantly affect Fb binding or susceptibility of *S. equi* to an opsonic equine serum; in other

words, it did not affect antibody mediated phagocytosis and killing. They also found that the N-terminal variation has significance in mucosal IgA and systemic T cell responses.

Further the SeM variable region was found to play a role in the establishment of asymptomatic carriers. Chanter et al. (2000) found that a deletion in the *seM* gene within this variable region was associated with the development of carriers; 24 % of *S. equi* isolates from outwardly healthy horses contained this deleted region of *seM* resulting in a truncated protein. The deletion occurred within the Fb and IgG binging region and was found to be associated with decreased resistance to phagocytosis (Chanter, Talbot et al. 2000). Parkinson et al. (2011) describes an isolate, obtained from the guttural pouch of an asymptomatic carrier, containing a single nucleotide insertion in the *seM* gene giving rise to a premature stop codon and therefore a putative truncated SeM protein. In the present study, no carriers were detected as all confirmed *S. equi* cases were from symptomatic horses. Further, none of the isolates obtained in this study carried deletions or insertions in the SeM variable region.

3.4.3. Sau-PCR Typing

A technique based on the digestion of genomic DNA with the restriction endonuclease Sau3AI and subsequent amplification was proposed for the genetic fingerprinting of virtually any cultivable microorganism (Corich, Mattiazzi et al. 2005). A Sau-PCR analysis of *Burkholderia cepacia* strains was compared to PFGE (Yan, Shi et al. 2008). It was found to be a less complex method that provided bacterial genotyping results within a single working day, whereas PFGE required at least four days to produce data of almost identical clinical significance. In a *S. equi* study, PFGE was able to distinguish wildtype strains of *S. equi* from each other (Lanka, Borst et al. 2010). Yan et al. (2008) reported that the Sau-PCR assay was highly reproducible and sensitive enough to discriminate the clonal diversity of *B. cepacia* in tracing the source and the routes of a hospital outbreak.

However, Sau-PCR has also been shown to be not as sensitive as random-amplified polymorphic DNA (RAPD)-PCR for comparing isolates of *Lactococcus garvieae* (Foschino, Nucera et al. 2008). There have been no epidemiological studies of *S. equi* using Sau-PCR and its usefulness was trialled in this present study.

Out of the 19 *S. equi* isolates obtained in this study, Sau-PCR was able to group them into two types, Sau1 and Sau2. Sau1 contained both *seM* 99 and 100, with no discriminating between them. Sau2 contained all the vaccine related isolates. Sau-PCR was only able to differentiate between wildtype and vaccine strains and was not able to further subtype *seM* 99 and 100. The possibility of coupling different Sau primers may result in increasing the ability of Sau-PCR to further subtype the wildtype strains, however, this was not explored in this study.

3.4.4. *S. equi* Typing Conclusions

This data demonstrates that the SeM allele sequencing is superior in regards to sensitivity then the Sau-PCR method. Sau-PCR was able to only discriminate between wildtype strains and the vaccine strain. The benefit of a sequencing method over the Sau-PCR method based on comparisons of patterns of electrophoretic bands is that across countries and laboratories, sequencing is universally applicable, reliable and repeatable. There are too many variables, which could change the outcome of banding patterns thereby decreasing the ability of cross-laboratory comparisons. The sequencing of the *seM* variable region identified three strains of *S. equi* within New Zealand, *seM* 99 and 100 being novel wildtype strains and *seM* 2 being the Pinnacle® IN vaccine strain. SeM allele 100 is more dominate then SeM allele 99 and appears on both the North and South Islands of New Zealand.

4. Summary and Recommendations

4.1. Multiplex PCR for Diagnostic purposes

In the present study, a conventional multiplex PCR was developed to identify the presence of *S. equi* (the causative agent of Strangles) and two other streptococci known to cause disease in horses, *S. zooepidemicus* and *S. equisimilis*. The method described involves DNA extraction from swabs or washes for direct PCR with results available within 3-4 hrs upon receiving specimens. As the PCR will detect dead bacteria, the results must be further confirmed with an overnight culture on streptococci selective media such as the GPNT media used in this study. These results are preliminary and it is recommended that further clinical specimens be trialled to test its robustness. Further tests will also be required to ascertain the multiplex's ability on detecting asymptomatic carriers as none were identified in this study. The molecular test developed here could provide a practical alternative to culture assays for the detection of *S. equi* in clinical specimens.

To increase detection ability in clinical specimens, it is recommended that nasopharyngeal swabs or nasal washes be used over nasal swabs. Pus from abscesses can also be used. For suspected carriers, guttural pouch lavages or three consecutive weekly nasopharyngeal swabs should be used to detect any harboured *S. equi* in conjunction with culture (Newton, Verheyen et al. 2000; Taylor and Wilson 2006). To confirm carriage, any positive result should be followed up by endoscopic examination of the guttural pouch for evidence of abnormalities such as empyema or chondroids (Newton, Verheyen et al. 2000). Earlier studies have

shown that *S. equi* was isolated from only the left or right guttural pouch and/or nasopharynx (Newton, Wood et al. 1997); therefore, it is recommended that samples be taken from both left and right sides of the head.

This multiplex PCR will be valuable to the New Zealand horse industry as it decreases the amount of time for a diagnosis and allows differentiation of the causative agent of upper respiratory tract infections, which all three streptococci have been known to be involved in. Further, with the increasing need for PCR negative results for *S. equi* to ship to the overseas market, the uniplex PCR to identify *S. equi* will be of value.

Recently, a simple and portable assay for detection of *S. equi* was developed based on amplification of a *S. equi* specific sequence using a thermophilic helicase-dependent reaction followed by visual detection of the amplicon in a disposable lateral flow cassette (Artiushin, Tong et al. 2011). This assay was shown to be more sensitive than culture but less sensitive than nested-PCR. Thermophilic helicase-dependant amplification uses a heat-stable helicase to unwind complementary DNA strands instead of heat, allowing isothermal DNA amplification without the need for thermocycling which can be done at one temperature (Vincen, Xu et al. 2004). Also, with the use of a disposable cassette for amplicon detection there is no need for gel electrophoresis. Although this new assay requires some equipment and training of personnel, it provides an alternative to culture or PCR assays for detection of *S. equi* in clinical samples and could offer a simple portable *S. equi* diagnostic tool to be used in the field and at the point-of-care. It maybe worth further investigation with regards to costs and ability for use as a diagnostic tool within New Zealand. However, unlike the multiplex PCR developed in the present study, it does not differentiate between *S. equi*, *S. zooepidemicus*, and *S. equisimilis*.

4.2. Typing of New Zealand *S. equi* Strains

Two novel strains of *S. equi* were found within New Zealand based on the variable region of *seM*, SeM alleles 99 and 100. Currently, these strains have only been found within New Zealand. One strain was found on both islands, allele 100, and the other only found on the North island, allele 99. Further to this study, the Pinnacle® IN vaccine strain, SeM 2 was isolated from lymph node abcesses of two horses. It is unclear as to whether this 'vaccine breakdown' is just a severe adverse reaction to the vaccine or if the vaccine reverted to a more virulent type. Sequencing of the SeM gene may allow differentiation or linkage of strangles outbreaks within New Zealand as well as identify vaccine related isolates.

The dominance of SeM allele 100 needs to be further ascessed with more isolates. Further, it is unclear as to if this perceived dominance is due to an increased fitness with regards to the host environment, such as an increased ability to persist in the guttural pouch or an increased fitness with regards to the strangles vaccines used in New Zealand. It will be interesting to determine the affectiveness of these two vaccines, Equivac® S and Pinnacle® IN, against the two novel New Zealand strains and whether a combination of the vaccines will provide better effectivness against these strains. Combining the two available vaccines might also prevent the severe adverse reactions, such as the ones found in this study.

Appendices

A.I. Media and Buffers

A.I.1. Selective Media

BHI/GPNT:

Make BHI (Bacto™ Brain Heart Infusion, Difco) broth to manufacturer's directions adding 1 mg/ml thallium acetate (Riedel-de Haën).
Sterilize by autoclaving and cool to ≤ 50°C.
Add:
2.5 μg/ml gentamicin
10 μg/ml polymyxin B
10 μg/ml nalidixic acid

THY/GPNT:

Make TH (Todd Hewitt Broth, Difco) broth to manufacturer's directions adding 1 % (*w/v*) yeast extract (BD) and 1 mg/ml thallium acetate (Riedel-de Haën).
Sterilize by autoclaving and cool to ≤ 50 °C.
Add:
2.5 μg/ml gentamicin
10 μg/ml polymyxin B
10 μg/ml nalidixic acid

GPNT/blood agar:

Make Columbia Blood Agar Base (Difco) to manufacturer's directions adding 1 mg/ml thallium acetate (Riedel-de Haën) and ensuring room is left for the addition of 10 % blood (*v/v*).
Sterilize by autoclaving and cool to ≤ 50 °C.
Add:
10 % (*v/v*) sheep blood (Life Technologies)
2.5 μg/ml gentamicin
10 μg/ml polymyxin B
10 μg/ml nalidixic acid
Gently mix and pour into petri dishes (approximately 20 ml each).

A.I.2. Fermentation Plates

Sorbital:

8.8 g Columbia Blood Agar Base (Difco)
2 mls 1 M Tris (Tris (hydroxymethyl) amino-methane, APS) solution, pH 7.4
1.75 ml Bromocresal Purple (2 % *w/v* stock solution)
0.5 g Sorbitol (MW: 182.2 g/mol, Sigma)
Made up to 200 ml with distilled water
Sterilize by autoclaving and pour into petri dishes (approximately 20 ml each).

Trehalose:

8.8 g Columbia Blood Agar Base (Difco)
2 mls 1 M Tris (Tris (hydroxymethyl) amino-methane, APS) solution, pH 7.4
0.02 g Aniline Blue
0.05 g Trehalose-dihydrate (MW 378.34 g/mol, AppliChem)
Made up to 200 mls with distilled water
Sterilize by autoclaving and pour into petri dishes (approximately 20 ml each).

A.I.3. Buffers and Solutions

5M GITC Solution

295.4 g Guanidine thiocyanate (Sigma)
2.5 g Sakanosyl (aka: N-Lauroylsarcosine sodium salt, Sigma)
3.9 g tri-Sodium citrate (BDH)
3.6 ml 2-mercaptoethanol (Scharlau)
Made up to 500 ml with MQ-H_2O
pH to 7.0 with ~ 0.5 ml 1N NaOH

PEG/KOH

27 ml Polyethylene glycol (PEG) 200 (Sigma)
465 µl 2 M KOH (potassium hydroxide) (Riedel-de Haën)
Made up to 50 ml with MQ-H_2O
pH 13.3-13.5

SB Running Buffer

10 x stock solution
56 g boric acid (MW 61.83 g/mol, Ajax)
10 g NaOH (sodium hydroxide, MW 40.0 g/mol, Ajax)
Made up to 2 L in ddH_2O
pH should be at 8.5
Dilute to 1 x in ddH_2O for working solution
Working solution final concentrations:
45 mM boric acid
12.5 mM NaOH

SDS Lysis Solution

5 mls 1M Tris (APS) solution, pH 9
5 mls 0.5 M EDTA (Ethylenediamine tetra acetic acid , Scharlau) solution, pH 8
5 mls 10 % SDS (Roche) solution
1 ml 5 M NaCl solution
Made up to 50 ml with MQ-H_2O

TAE Running Buffer

50 x stock solution
242 g Tris (MW 121.14 g/mol, APS)
100 ml 0.5 M EDTA (Ethylenediamine tetra acetic acid , Scharlau) solution, pH 8
57.1 ml acetic acid (MW 60.05 g/mol, Ajax)
Made up to 1 L in ddH_2O
Diluted to 1 x in ddH_2O for working solution
Working solution final concentrations:
40 mM Tris
1 mM EDTA
20 mM acetic acid

A.II. Primer Sequences

Primer Name	Sequence (5'-3')	Gene/Genomic Region	Species	Product Size	Reference	Method
seeI F seeI R	GAA GGT CCG CCA TTT TCA GGT AGT TTG GCA TAC TCT CTC TGT CAC CAT GTC CTG	Superantigenic toxin	*S. equi*	520 bp	Alber et al. (2004)	Uniplex Multiplex Sequencing
seM F *seM* R	CAG AAA ACT AAG TGC CGG TG ATT CGG TAA GAG CTT GAC GC	Novel M-like protein	*S. equi*	541 bp	Kelly et al. (2006) (ASW73 (Fwd) and ASW74 (Rev))	Uniplex Multiplex SeM Typing Sequencing
eqsim F *eqsim* R	TCA AAT CGG TTG GCA CAG AC CGT CCT TAG CAT AGA AGG ATT GG	Streptokinase precursor gene	*S. equisimilis*	279 bp	Preziuso et al. (2010)	Uniplex Multiplex Sequencing
srtz F *srtz* R	GCT GAC TGA TCA AGC CAC AA AAC TGA TCG TGT TCC CAA GC	Putative sortase	*S. zooepidemicus*	217 bp	Present Study	Uniplex Multiplex Sequencing
16S F 16S R	GGA TTA GAT ACC CBB GTA GTC C GAC GTC RTC CNC DCC TTC CTC	16S rDNA	Universal to bacteria	410 bp	(Boye et al. (1999) (DA71 and DA72)	Uniplex Multiplex
SAUA SAUC SAUG SAUT SAG SCA STG	CCG CCG CGA TCA CCG CCG CGA TCC CCG CCG CGA TCG CCG CCG CGA TCT CCG CCG CGA TCAG CCG CCG CGA TCCA CCG CCG CGA TCTG	Primers based on Sau3AI recognition site GATC with a G+C tail		Variable	Corich et al. (2005)	Sau-PCR

A.III. Alignment of all 114 SeM alleles in the SeM database

(http://pubmlst.org/szooepidemicus/seM, accessed January 12, 2012)

The alignment of all 114 SeM alleles is supplied as an image file on the provided compact disk. It was generated using Geneious v5.4 (Drummond, Ashton et al. 2011). Nucleotide and amino acid differences are highlighted with color. The dot (•) indicates a match to the consensus sequence.

A.IV. *seM* sequences from the SeM database (http://pubmlst.org/szooepidemicus/seM)

(Accessed January 12, 2012)

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
1	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCGATGGGGATAGAAGACAATTAAGTTTG GCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	1	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
2	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	2	SEVSRTATPRLSRDLKNRLSDIAISGDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
3	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGGTGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	3	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLVHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
4	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAGGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	4	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYYNLLMRLSSMLNDKPDGDRGQLSL ASLLVDEIEKRIADGDSYAK
5	TCTGAGGTTAGACGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCTTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	5	SEVRRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMRLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
6	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	6	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYYNLLMRLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
7	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	7	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
8	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATCCAGCAAGGGCTGCGTATGGTAGAGATGATTATTAC AATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTT TGGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	8	SEVSRTATPRLSRDLKNRLSDIAISRDASSA QKVRNLLKGASVGDLQALLRGLDPARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
9	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	9	SEVSRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
10	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	9	SEVSRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
11	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAGATA GCCATAAGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	11	SEVSRTATPRLSRDLKNRLSEIAISGDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
12	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAGGTAGATTAAGCGAAATA GCCATAAGTAGAGATGTCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAATTT GTCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	12	SEVSRTATPRLSRDLKGRLSEIAISRDVSSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYYNLLMRLSSMLNDKPDGDRRQLNL SSLLVDEIEKRIADGDSYAK
13	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAGTAGATTAAGCGAAATA GCCATAAGTAGAGATGTCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTCCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAATTT GTCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	13	SEVSRTATPRLSRDLKSRLSEIAISRDVSSAQ KVRNLLKGASVGDLQALLRGFDSARAAYG RDDYYNLLMHLPSMLNDKPDGDRRQLNLS SLLVDEIEKRIADGDSYAK
14	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAGGTAGATTAAGCGAAATA GCCATAAGTAGAGATGTCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTCCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAATTT GTCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	14	SEVSRTATPRLSRDLKGRLSEIAISRDVSSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYYNLLMHLPSMLNDKPDGDRRQLN LSSLLVDEIEKRIADGDSYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
15	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAGATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCAGCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	15	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMQLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
16	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAGATA GGCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	16	SEVSRTATPRLSRDLKNRLSEIGISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
17	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	17	SEVSRTATPRLSRDLKNRLSEIAISGDASSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYYNLLMRLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
18	CCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	18	PEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMRLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
19	CCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCAGGGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	19	PEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSGMLNDKPDGDRRQLS LASLLVDEIEKRIADGDRYAK
20	CCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	20	PEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
21	CCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGGTGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	21	PEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLVHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
22	CCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGAAGCACCTTCCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	22	PEVSRTATPRLSRDLKNRLSEIAISGDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLKHLPSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
23	CCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATCAC AATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTT TGGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	23	PEVSRTATPRLSRDLKNRLSEIAISGDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYHNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
24	CCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	24	PEVSRTATPRLSRDLKNRLSEIAISGDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
25	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	25	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
26	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATTGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	26	SEVSRTATPRLSRDLKNRLSDIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMRLSLMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
27	TCTGAGGTTAGACGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	27	SEVSRTATPRLSRDLKNRLSDIAISRDTSSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
28	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	28	SEVSRTATPRLSRDLKNRLSDIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
29	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATCGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	29	SEVSRTATPRLSRDLKNRLSDIAISGDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLSMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
30	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATCGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	30	SEVSRTATPRLSRDLKNRLSDIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLSMRLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
31	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTTTTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	31	SEVSRTATPRLSRDLKNRLSDIAISGDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDFYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
32	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTTTCACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	32	SEVSRTATPRLSRDLKNRLSDIAISGDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDFHNLLMRLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
33	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTCCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	33	SEVSRTATPRLSRDLKNRLSDIAISGDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLPSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
34	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTGGAGGTGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	34	SEVSRTATPRLSRDLKNRLSDIAISGGASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
35	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	35	SEVSRTATPRLSRDLKNRLSDIAISGDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
36	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTAGAGATGCCTCATCAGCCTTAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	36	SEVSRTATPRLSRDLKNRLSDIAISRDASSA LKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
37	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAGTAGATTAAACGAGATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	37	SEVSRTATPRLSRDLKSRLNEIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
38	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAACGATATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAAAGATGATTATTACA ATTTATTGATGCACCTTCCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	38	SEVSRTATPRLSRDLKNRLNDIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GKDDYYNLLMHLPSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
39	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAACGATATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAAAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	39	SEVSRTATPRLSRDLKNRLNDIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GKDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
40	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAGATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGTCAATTAAATTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	40	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRSQLNL ASLLVDEIEKRIADGDSYAK
41	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTTAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTCAAATGATAAACCTGATGGGGATAGAAGACAATTAAATTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	41	SEVSRTATPRLSRDFKNRLSEIAISRDASSAQ KVRNLLKGASVGDLQALLRGLDSARAAYG RDDYYNLLMHLSSMSNDKPDGDRRQLNLA SLLVDEIEKRIADGDSYAK
42	TCTGAGGTTAGTCGTACGGCGACTCCAAAATTATCGCGTAATTTAAGAAATAGATTCAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATAAC AATTTATTGATGCGCCTTCCACCGATGTTAAATGATAAACCTGATGGGGATAGAAGTCAATTAAGTT TGGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	42	SEVSRTATPKLSRNLRNRFSEIAISRDASSAQ KVRNLLKGASVGDLQALLRGLDSARAAYG RDDYNNLLMRLPPMLNDKPDGDRSQLSLA SLLVDEIEKRIADGDSYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
43	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAGATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	1	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
44	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAGATA GGCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTATAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	44	SEVSRTATPRLSRDLKNRLSEIGISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLIDEIEKRIADGDSYAK
45	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAGATA GGCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATTGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	45	SEVSRTATPRLSRDLKNRLSEIGISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMRLSLMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
46	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAGATA GCCATAGGTAGAGGTGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	46	SEVSRTATPRLSRDLKNRLSEIAIGRGASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
47	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGACTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	10	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGFDSARTAY GRDDYYNLLMRLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
48	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	43	SEVSRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMRLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
49	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATCGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	49	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGFDSARAAY RRDDYYNLLMRLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
50	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGTTTAAGCGAAATA GCCATAAGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	50	SEVSRTATPRLSRDLKNSLSEIAISGDASSAQ KVRNLLKGASVGDLQALLRGFDSARAAYG RDDYYNLLMRLSSMLNDKPDGDRRQLSLA SLLVDEIEKRIADGDSYAK
51	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAGGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	47	SEVSRTATPRLSRDLKNRLSEIAIGRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
52	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACTTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	48	SEVSRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKLDGDRRQLSL TSLLVDEIEKRIADGDSYAK
53	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAGATA GGCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATTGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	51	SEVSRTATPRLSRDLKNRLSEIGISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSLMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
54	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAGATA GGCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTCCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	52	SEVSRTATPRLSRDLKNRLSEIGISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLPSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
55	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	53	SEVSRTATPRLSRDLKNRLSEIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
56	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAGATA GGCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATTGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTATAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	54	SEVSRTATPRLSRDLKNRLSEIGISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY CRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLIDEIEKRIADGDSYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
57	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTCCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	55	SEVSRTATPRLSRDLKNRLSDIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLPSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
58	TCTGAGGTTAGTCGTACGGTGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAACGAAATA GCCATAAGTAGAGATGTCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	56	SEVSRTVTPRLSRDLKNRLNEIAISRDVSSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
59	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATACCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	57	SEVSRTATPRLSRDLKNRLSEIAISRDTSSAQ KVRNLLKGASVGDLQALLRGFDSARAAYG RDDYYNLLMRLSSMLNDKPDGDRRQLSLA SLLVDEIEKRIADGDSYAK
60	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGAGGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	58	SEVSRTATPRLSRDLKNRLSDIAIDREASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
61	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAATTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	59	SEVSRTATPRLSRDLKNRLSDIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMRLSSMLNDKPDGDRRQLNL ASLLVDEIEKRIADGDSYAK
62	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	60	SEVSRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
63	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGTCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	61	SEVSRTATPRLSRDLKNRLSEIAISRDVSSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
64	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTTATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	62	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYYNLLMHLLSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
65	TCTGAGGTTAGTCGTACGGTGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAACGAAATA GCCATAAGTAGAGATGTCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTCCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	63	SEVSRTVTPRLSRDLKNRLNEIAISRDVSSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYYNLLMHLPSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
66	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTCCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	64	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYYNLLMRLPSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
67	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAATTAGA	65	SEVSRTATPRLSRDLKIR
68	TCTGAGGTTAGACGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	5	SEVRRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMRLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
69	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	66	SEVSRTATPRLSRDLKNRLSDIAINRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
70	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCTAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	67	SEVSRTATPRLSRDLKNRLSEIAISRDASSA KVRNLLKGASVGDLQALLRGFDSARAAYG RDDYYNLLMRLSSMLNDKPDGDRRQLSLA SLLVDEIEKRIADGDSYAK
71	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGTAAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	68	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYYNLLMRLSSMLNVKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
72	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAGATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	53	SEVSRTATPRLSRDLKNRLSEIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
73	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCAATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	69	SEVSRTATPRLSRDLKNRLSNIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
74	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGGTATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATCCAGCAAGGGCTGCGTATGGTAGAGATGATTATTAC AATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTT TGGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	70	SEVSRTATPRLSRDLKNRLSGIAISRDASSA QKVRNLLKGASVGDLQALLRGLDPARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
75	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGAAGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	71	SEVSRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLKRLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
76	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGTTATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	72	SEVSRTATPRLSRDLKNRLSVIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
77	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTCCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	73	SEVSRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLPSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
78	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGTTATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACATTTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	74	SEVSRTATPRLSRDLKNRLSVIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRHLSL TSLLVDEIEKRIADGDSYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
79	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTTAAAAATAGATTAAGCGAAATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	75	SEVSRTATPRLSRDFKNRLSEIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
80	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGTTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	76	SEVSRTATPRLSRDLKNRLSDIAIVRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMRLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
81	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAGATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATCCAGCAAGGGCTGCGTATGGTAGAGATGATTATTAC AATTTATTGGTGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTT TGGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	77	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGLDPARAAY GRDDYYNLLVHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
82	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATAAC AATTTATTGATGAGCCTTTCATCGATGTCAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTT TGGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	78	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYNNLLMSLSSMSNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
83	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTCCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	79	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYYNLLMHLPSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
84	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGACGAAGCAAGGGCTGCGTATGGTAGAGATGATTATTAC AACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTT TGACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	80	SEVSRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDEARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
85	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGTTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAATTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	81	SEVSRTATPRLSRDLKNRLSDIAIVRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLN LTSLLVDEIEKRIADGDSYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
86	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGAAAGGGCTGCGTATGGTAGAGATGATTATTAC AACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTT TGGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	83	SEVSRTATPRLSRDLKNRLSDIAIGRDASSA QKVRNLLKGASVGDLQALLRGLDSERAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
87	AGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAGATAGCCATAAGT AGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGGGATTTACAG GCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTAGTACAATTTATTGA TGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTTGGCTTCATT ACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAACT	82	SRTATPRLSRDLKNRLSEIAISRDASSAQKV RNLLKGASVGDLQALLRGLDSARAAYGRD D
88	TCTGAGGTTAGTCGTACGGCGACTCCAACGTTATCGCGTGATTTTAAAAATAGATTAAACGAGATA GCCATAACTGGAGATCATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTT AGGGATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTAT TACAATTTATTGATACAGCTTTCATCGATGCCAAATGATAAACCTGATGGGGATAGATCACAATTAA ATTTGAGTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	84	SEVSRTATPTLSRDFKNRLNEIAITGDHASS AQKVRNLLKGASVRDLQALLRGLDSARAA YGRDDYYNLLIQLSSMPNDKPDGDRSQLN LSSLLVDEIEKRIADGDSYAK
89	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	85	SEVSRTATPRLSRDLKNRLSDIAIGRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
90	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GACATAAGTAGAGATACCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTCGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	86	SEVSRTATPRLSRDLKNRLSEIDISRDTSSAQ KVRNLLKGASVGDLQALLRGFDSARAAYG RDDYYNLLMRLSSMLNDKPDGDRRQLSLA SLLVDEIEKRIADGDSYAK
91	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATCAC AATTTATCGATGCACCTTCCATCGATGTCAAATGATAAACCTGATGGGGATAGAAGACAATTAAGT TTGGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	87	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYHNLSMHLPSMSNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
92	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAACGATATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAAAGATGATTATTACA ATTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	88	SEVSRTATPRLSRDLKNRLNDIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GKDDYYNLLMRLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
93	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTTTTGAGGGCCTTTTTTTCCATGTTAAAGGATAAACCCGCTGGGGATTTCAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	89	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNFLRAFFSMLKDKPAGDFRQLSL ASLLVDEIEKRIADGDSYAK
94	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAGAAATAGATTAAGCGTTATA GCCATAAGTGGAGATGCCCAAAAAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	90	SEVSRTATPRLSRDLRNRLSVIAISGDAQKA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDRYAK
95	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTTATCACTGTTAAATGATAAACCTCATGGGGAGAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGGTATGCAAAA	91	SEVSRTATPRLSRDLKNRLSDIAISGDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLLSLLNDKPHGERRQLSL ASLLVDEIEKRIADGDRYAK
96	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAACGATATA GCCATAAGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAAAGATGATTATTACA ATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	92	SEVSRTATPRLSRDLKNRLNDIAISGDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GKDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
97	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAAGTAGAGATACCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATCCAGCAAGGGCTGCGTATGGTAGAGATGATTATTAC AATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTT TGGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	93	SEVSRTATPRLSRDLKNRLSDIAISRDTSSA QKVRNLLKGASVGDLQALLRGLDPARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
99	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAGGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	95	SEVSRTATPRLSRDLKNRLSDIAIGRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRGQLSL ASLLVDEIEKRIADGDSYAK
100	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGGTGGAGATGCCTCATCAGCCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTCCATCGATGTTAAATGATAAACCTGATGGGGATAGAGGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	96	SEVSRTATPRLSRDLKNRLSDIAIGGDASSA PKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLPSMLNDKPDGDRGQLSL ASLLVDEIEKRIADGDSYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
101	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTCAAATGATAAACTTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	97	SEVSRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMSNDKLDGDRRQLSL TSLLVDEIEKRIADGDSYAK
102	TCTGAGGTTAGTCGTACGGCGACTCCGAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	9	SEVSRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
103	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAGAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	43	SEVSRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMRLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
104	TCTGAGGTTAGTCGTACGGCGACTCCCAGATTATCGCGTGATTTAAAAAATACATTAAGCGATATA GCCATAAATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCCTTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAC	98	SEVSRTATPRLSRDLKNTLSDIAINRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAN
105	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGTTATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATCCAGCAAGGGCTGCGTATGGTAGAGATGATTATTAC AATTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTT TGGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	99	SEVSRTATPRLSRDLKNRLSVIAISRDASSA QKVRNLLKGASVGDLQALLRGLDPARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
106	TCTGAGGTTAGTCGTACGGCGACTCCAATATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	100	SEVSRTATPILSRDLKNRLSDIAIDRDASSAQ KVRNLLKGASVGDLQALLRGLDSARAAYG RDDYYNLLMHLSSMLNDKPDGDRRQLSLT SLLVDEIEKRIADGDSYAK
107	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTTTCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	101	SEVSRTATPRFSRDLKNRLSDIAIGRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK

Continues next page…..

Allele id	Nucleotide Sequence	Peptide id	Peptide Sequence
108	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGGTGCACCTTTCATCGATGTTAATTGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	102	SEVSRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLVHLSSMLIDKPDGDRRQLSLT SLLVDEIEKRIADGDSYAK
109	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATCCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	103	SEVSRTATPRLSRDLKNRLSDIAIDRDPSSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
110	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTATTGATGCACCTTTCACCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	104	SEVSRTATPRLSRDLKNRLSEIAISRDASSA QKVRNLLKGASVGDLQALLRGFDSARAAY GRDDYYNLLMHLSPMLNDKPDGDRRQLSL ASLLVDEIEKRIADGDSYAK
111	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGAAGGGAGATAGTTATGCAAAA	105	SEVSRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIAEGRLCK
112	GATGTTGAGTCTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	9	SEVSRTATPRLSRDLKNRLSDIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK
113	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGAAATA GCCATAAGTAGAGATACCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTTTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ATTTACGGATGCGCCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	106	SEVSRTATPRLSRDLKNRLSEIAISRDTSSAQ KVRNLLKGASVGDLQALLRGFDSARAAYG RDDYYNLRMRLSSMLNDKPDGDRRQLSLA SLLVDEIEKRIADGDSYAK
114	TCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCAGTATA GCCATAGATAGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAAAAGGCGCCTCTGTTGGG GATTTACAGGCATTATTGAGAGGTCTTGATTCAGCAAGGGCTGCGTATGGTAGAGATGATTATTACA ACTTATTGATGCACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTT GACTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAA	107	SEVSRTATPRLSRDLKNRLSSIAIDRDASSA QKVRNLLKGASVGDLQALLRGLDSARAAY GRDDYYNLLMHLSSMLNDKPDGDRRQLSL TSLLVDEIEKRIADGDSYAK

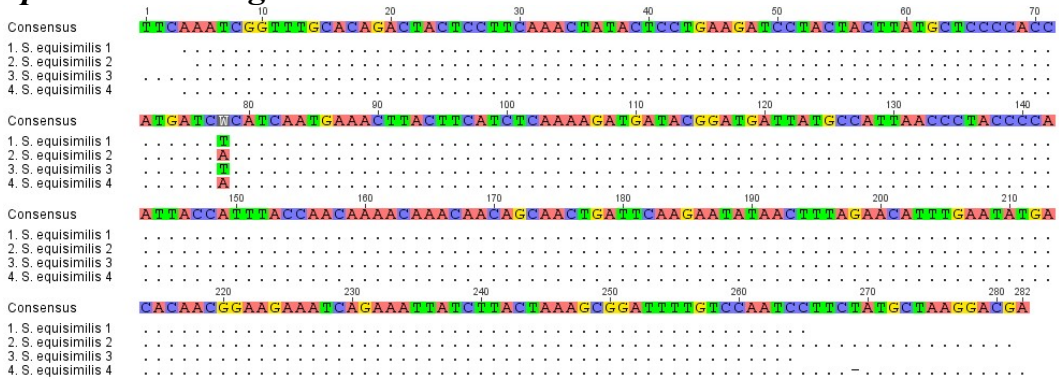
A.V. Isolate alignment and BLAST results from sequence data

A.V.1. *eqsim* Amplicons of the Four *S. equisimilis* Isolates in the Present Study

>*eqsim*_consensus_sequence:

AATCGGTTTGCACAGACTACTCCTTCAAACTATACTCCTGAAGATCCT
ACTACTTATGCTCCCCACCATGATCWCATCAATGAAACTTACTTCATC
TCAAAAGATGATACGGATGATTATGCCATTAACCCTACCCCAATTACC
ATTTACCAACAAAACAAACAACAGCAACTGATTCAAGAATATAACTTT
AGAACATTTGAATATGACACAACGGAAGAAATCAGAAATTATCTTACT
AAAGCGGATTTTGTCCAATC

***eqsim* isolate alignment:**



Parameters from BLAST Search of *eqsim* consensus sequence:

BLASTN 2.2.26+ (Mar 24, 2012)
Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14.
RID: PZSNM42J01N
Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,environmental samples or phase 0, 1 or 2 HTGS sequences)
15,916,285 sequences; 40,589,149,171 total letters
Query= *eqsim*_consensus_sequence
Length=260

eqsim BLAST results:

Sequences producing significant alignments:

Sequences	Score (Bits)	E Value
gb|AF104301.1|AF104301 *Streptococcus equisimilis* strain 87-542-W streptokinase precursor	472	7e-130

Alignments of *eqsim* Query Sequence with Streptococci Sequences:

Alignment of *eqsim* query sequence with: gb|AF104301.1|AF104301 *Streptococcus equisimilis* strain 87-542-W streptokinase precursor,(skc) gene, partial cds

Length=1248
Score = 472 bits (255), Expect = 7e-130
Identities = 258/260 (99%), Gaps = 0/260 (0%)
Strand=Plus/Plus

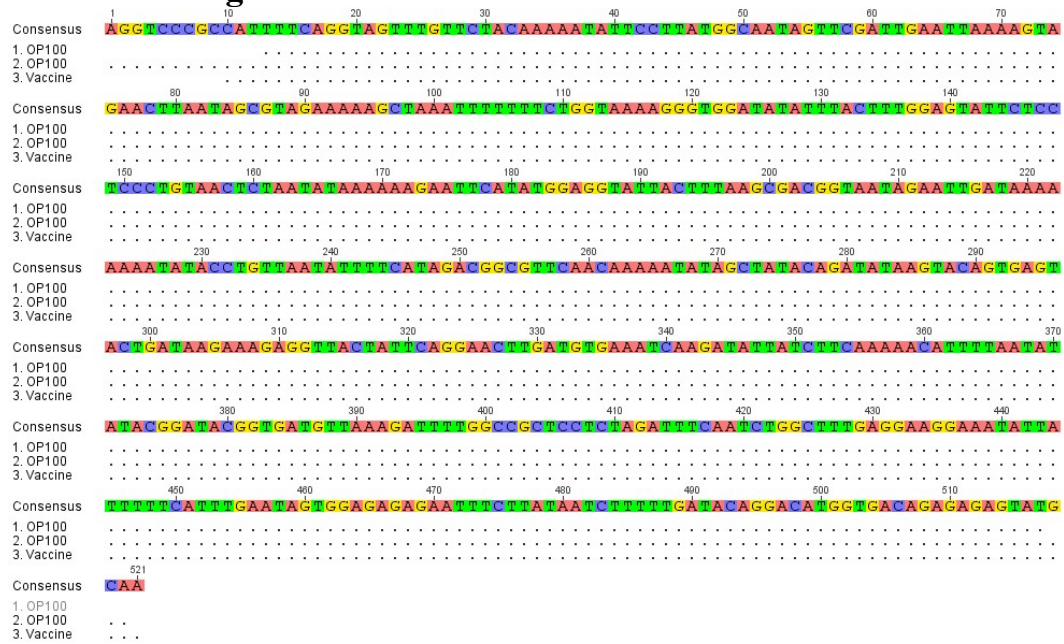
```
Query 1      AATCGGTTTGCACAGACTACTCCTTCAAACTATACTCCTGAAGATCCTACTACTTATGCT  60
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  892   AATCGGTTGGCACAGACTACTCCTTCAAACTATACTCCTGAAGATCCTACTACTTATGCT  951

Query  61    CCCCACCATGATCWCATCAATGAAACTTACTTCATCTCAAAAGATGATACGGATGATTAT  120
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  952   CCCCACCATGATCACATCAATGAAACTTACTTCATCTCAAAAGATGATACGGATGATTAT  1011

Query  121   GCCATTAACCCTACCCCAATTACCATTTACCAACAAAACAAACAACAGCAACTGATTCAA  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1012  GCCATTAACCCTACCCCAATTACCATTTACCAACAAAACAAACAACAGCAACTGATTCAA  1071

Query  181   GAATATAACTTTAGAACATTTGAATATGACACAACGGAAGAAATCAGAAATTATCTTACT  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1072  GAATATAACTTTAGAACATTTGAATATGACACAACGGAAGAAATCAGAAATTATCTTACT  1131

Query  241   AAAGCGGATTTTGTCCAATC  260
             ||||||||||||||||||||
Sbjct  1132  AAAGCGGATTTTGTCCAATC  1151
```

A.V.2. *seeI* Amplicons from Isolates of *seM* 99, 100 and 2, Representing All *S. equi* Isolates in the Present Study

>*seeI*__consensus_sequence:

TTTCAGGTAGTTTGTTCTACAAAAATATTCCTTATGGCAATAGTTCGAT
TGAATTAAAAGTAGAACTTAATAGCGTAGAAAAAGCTAAATTTTTTTC
TGGTAAAAGGGTGGATATATTTACTTTGGAGTATTCTCCTCCCTGTAAC
TCTAATATAAAAAAGAATTCATATGGAGGTATTACTTTAAGCGACGGT
AATAGAATTGATAAAAAAAATATACCTGTTAATATTTTCATAGACGGC
GTTCAACAAAAATATAGCTATACAGATATAAGTACAGTGAGTACTGAT
AAGAAAGAGGTTACTATTCAGGAACTTGATGTGAAATCAAGATATTAT
CTTCAAAAACATTTTAATATATACGGATACGGTGATGTTAAAGATTTT
GGCCGCTCCTCTAGATTTCAATCTGGCTTTGAGGAAGGAAATATTATTT
TTCATTTGAATAGTGGAGAGAGAATTTCTTATAATCTTTTTGATACAGG
ACATGGTGACAGAGAGAGTAT

seeI isolate alignment:



Parameters from BLAST search of *seeI* consensus sequence:

BLASTN 2.2.26+ (Mar 24, 2012)
Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and
Webb Miller (2000), "A greedy algorithm for aligning DNA
sequences", J Comput Biol 2000; 7(1-2):203-14.
RID: PZTCBEZZ01N
Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS,
GSS,environmental samples or phase 0, 1 or 2 HTGS sequences)
15,916,285 sequences; 40,589,149,171 total letters
Query= seeI__consensus_sequence
Length=505

seeI BLAST results:

Sequences producing significant alignments:

Sequences	Score (Bits)	E Value
emb|FM204883.1| *Streptococcus equi* subsp. *equi* 4047, complete genome	933	0.0
gb|AF186180.1| *Streptococcus equi* 49.7 kDa protein, 25.7 kDa protein, and SeeH (seeH) genes, complete cds	933	0.0
gb|CP000259.1| *Streptococcus pyogenes* MGAS9429, complete genome	917	0.0
gb|AF438524.1|AF438524 *Streptococcus pyogenes* exotoxin I (speI) gene, complete cds	917	0.0
emb|AM295007.1| *Streptococcus pyogenes* Manfredo complete genome	915	0.0
gb|AE004092.1| *Streptococcus pyogenes* M1 GAS, complete genome	911	0.0
gb|GQ923932.1| *Streptococcus dysgalactiae* subsp. *equisimilis* strain B81 exotoxin I (speI) gene, partial cds	346	1e-91

Alignments of *seeI* Query Sequence with Streptococci Sequences:

Alignment of *seeI* query sequence with: emb|FM204883.1| *Streptococcus equi* subsp. *equi* 4047, complete genome
Length=2253793

Features in this part of subject sequence:
exotoxin I precursor

Score = 933 bits (505), Expect = 0.0
Identities = 505/505 (100%), Gaps = 0/505 (0%)
Strand=Plus/Minus

```
Query  1        TTTCAGGTAGTTTGTTCTACAAAAATATTCCTTATGGCAATAGTTCGATTGAATTAAAAG  60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2066442  TTTCAGGTAGTTTGTTCTACAAAAATATTCCTTATGGCAATAGTTCGATTGAATTAAAAG
2066383

Query  61       TAGAACTTAATAGCGTAGAAAAAGCTAAAttttttttCTGGTAAAAGGGTGGATATATTTA  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2066382  TAGAACTTAATAGCGTAGAAAAAGCTAAATTTTTTTTCTGGTAAAAGGGTGGATATATTTA
2066323

Query  121      CTTTGGAGTATTCTCCTCCCTGTAACTCTAATATAAAAAAGAATTCATATGGAGGTATTA  180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2066322  CTTTGGAGTATTCTCCTCCCTGTAACTCTAATATAAAAAAGAATTCATATGGAGGTATTA
2066263

Query  181      CTTTAAGCGACGGTAATAGAATTGATaaaaaaaaTATACCTGTTAATATTTTCATAGACG  240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2066262  CTTTAAGCGACGGTAATAGAATTGATAAAAAAAAATATACCTGTTAATATTTTCATAGACG
2066203

Query  241      GCGTTCAACAAAAATATAGCTATACAGATATAAGTACAGTGAGTACTGATAAGAAAGAGG  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2066202  GCGTTCAACAAAAATATAGCTATACAGATATAAGTACAGTGAGTACTGATAAGAAAGAGG
2066143

Query  301      TTACTATTCAGGAACTTGATGTGAAATCAAGATATTATCTTCAAAAACATTTTAATATAT  360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2066142  TTACTATTCAGGAACTTGATGTGAAATCAAGATATTATCTTCAAAAACATTTTAATATAT
2066083

Query  361      ACGGATACGGTGATGTTAAAGATTTTGGCCGCTCCTCTAGATTTCAATCTGGCTTTGAGG  420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2066082  ACGGATACGGTGATGTTAAAGATTTTGGCCGCTCCTCTAGATTTCAATCTGGCTTTGAGG
2066023

Query  421      AAGGAAATATTATTTTTCATTTGAATAGTGGAGAGAGAATTTCTTATAATCTTTTTGATA  480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2066022  AAGGAAATATTATTTTTCATTTGAATAGTGGAGAGAGAATTTCTTATAATCTTTTTGATA
2065963

Query  481      CAGGACATGGTGACAGAGAGAGTAT  505
                |||||||||||||||||||||||||
Sbjct  2065962  CAGGACATGGTGACAGAGAGAGTAT  2065938
```

Alignment of *seeI* query sequence with: gb|CP000259.1| *Streptococcus pyogenes* MGAS9429, complete genome
Length=1836467

Features in this part of subject sequence:
enterotoxin

Score = 917 bits (496), Expect = 0.0
Identities = 502/505 (99%), Gaps = 0/505 (0%)
Strand=Plus/Plus

```
Query  1       TTTCAGGTAGTTTGTTCTACAAAAATATTCCTTATGGCAATAGTTCGATTGAATTAAAAG  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  817073  TTTCAGGTAGTTTGTTCTACAAAAATATTCCTTATGGCAATAGTTCGATTGAATTAAAAG 817132

Query  61      TAGAACTTAATAGCGTAGAAAAAGCTAAAttttttttCTGGTAAAAGGGTGGATATATTTA  120
               |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct  817133  TAGAACTTAATAGCGTAGAAAAAGCTAATTTTTTTTTCTGGTAAAAGGGTGGATATATTTA 817192

Query  121     CTTTGGAGTATTCTCCTCCCTGTAACTCTAATATAAAAAAGAATTCATATGGAGGTATTA  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  817193  CTTTGGAGTATTCTCCTCCCTGTAACTCTAATATAAAAAAGAATTCATATGGAGGTATTA 817252

Query  181     CTTTAAGCGACGGTAATAGAATTGATaaaaaaaaTATACCTGTTAATATTTTCATAGACG  240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  817253  CTTTAAGCGACGGTAATAGAATTGATAAAAAAAAATATACCTGTTAATATTTTCATAGACG 817312

Query  241     GCGTTCAACAAAAATATAGCTATACAGATATAAGTACAGTGAGTACTGATAAGAAAGAGG  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  817313  GCGTTCAACAAAAATATAGCTATACAGATATAAGTACAGTGAGTACTGATAAGAAAGAGG 817372

Query  301     TTACTATTCAGGAACTTGATGTGAAATCAAGATATTATCTTCAAAAACATTTTAATATAT  360
               |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  817373  TTACGATTCAGGAACTTGATGTGAAATCAAGATATTATCTTCAAAAACATTTTAATATAT 817432

Query  361     ACGGATACGGTGATGTTAAAGATTTTGGCCGCTCCTCTAGATTTCAATCTGGCTTTGAGG  420
               |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  817433  ACGGATTCGGTGATGTTAAAGATTTTGGCCGCTCCTCTAGATTTCAATCTGGCTTTGAGG 817492

Query  421     AAGGAAATATTATTTTTCATTTGAATAGTGGAGAGAGAATTTCTTATAATCTTTTTGATA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  817493  AAGGAAATATTATTTTTCATTTGAATAGTGGAGAGAGAATTTCTTATAATCTTTTTGATA 817552

Query  481     CAGGACATGGTGACAGAGAGAGTAT  505
               |||||||||||||||||||||||||
Sbjct  817553  CAGGACATGGTGACAGAGAGAGTAT  817577
```

Alignment of *seeI* query sequence with: gb|GQ923932.1| *Streptococcus dysgalactiae* subsp. *equisimilis* strain B81 exotoxin I (speI) gene, partial cds
Length=215

Score = 346 bits (187), Expect = 1e-91
Identities = 189/190 (99%), Gaps = 0/190 (0%)
Strand=Plus/Plus

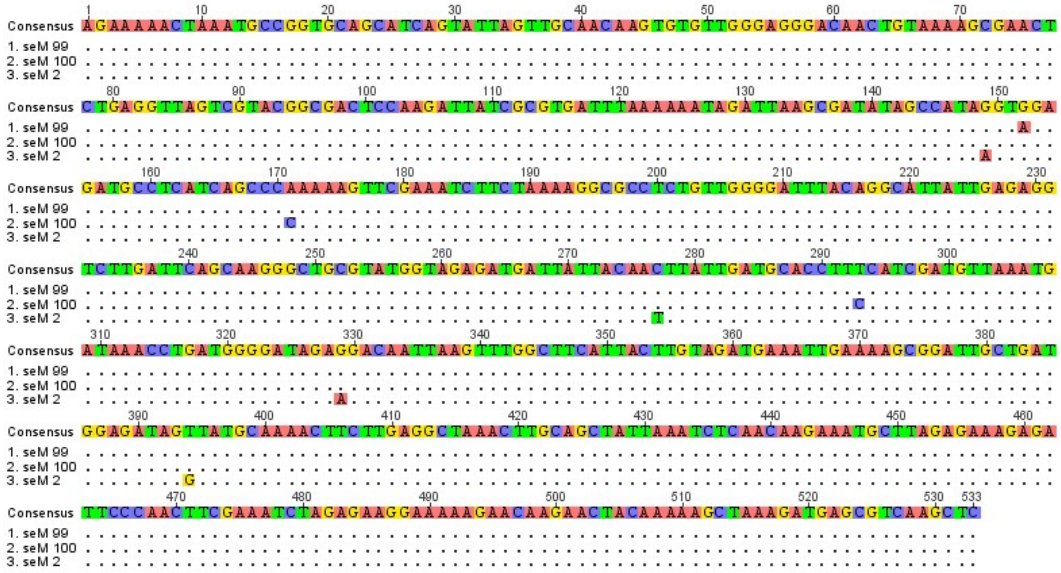
```
Query  1    TTTCAGGTAGTTTGTTCTACAAAAATATTCCTTATGGCAATAGTTCGATTGAATTAAAAG  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  26   TTTCAGGTAGTTTGTTCTACAAAAATATTCCTTATGGCAATAGTTCGATTGAATTAAAAG  85

Query  61   TAGAACTTAATAGCGTAGAAAAAGCTAAAttttttttCTGGTAAAAGGGTGGATATATTTA  120
            |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct  86   TAGAACTTAATAGCGTAGAAAAAGCTAATTTTTTTTTCTGGTAAAAGGGTGGATATATTTA  145

Query  121  CTTTGGAGTATTCTCCTCCCTGTAACTCTAATATAAAAAAGAATTCATATGGAGGTATTA  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  146  CTTTGGAGTATTCTCCTCCCTGTAACTCTAATATAAAAAAGAATTCATATGGAGGTATTA  205

Query  181  CTTTAAGCGA  190
            ||||||||||
Sbjct  206  CTTTAAGCGA  215
```

A.V.3. *seM* Amplicons from Isolates of *seM* 99, 100 and 2, Representing All *S. equi* Isolates in the Present Study

>*seM*_consensus_sequence:

AGAAAAACTAAATGCCGGTGCAGCATCAGTATTAGTTGCAACAAGTGT
GTTGGGAGGGACAACTGTAAAAGCGAACTCTGAGGTTAGTCGTACGG
CGACTCCAAGATTATCGCGTGATTTAAAAAATAGATTAAGCGATATAG
CCATAGGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAATCTTCTAA
AAGGCGCCTCTGTTGGGGATTTACAGGCATTATTGAGAGGTCTTGATT
CAGCAAGGGCTGCGTATGGTAGAGATGATTATTACAACTTATTGATGC
ACCTTTCATCGATGTTAAATGATAAACCTGATGGGGATAGAGGACAAT
TAAGTTTGGCTTCATTACTTGTAGATGAAATTGAAAAGCGGATTGCTG
ATGGAGATAGTTATGCAAAACTTCTTGAGGCTAAACTTGCAGCTATTA
AATCTCAACAAGAAATGCTTAGAGAAAGAGATTCCCAACTTCGAAATC
TAGAGAAGGAAAAAGAACAAGAACTACAAAAAGCTAAAGATGAGCG
TCAAGCTC

***seM* isolate alignment:**



Parameters from BLAST search of *seM* consensus sequence:

BLASTN 2.2.26+ (Mar 24, 2012)
Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14.
RID: T26DBZ5C013
Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,environmental samples or phase 0, 1 or 2 HTGS sequences)
15,971,215 sequences; 40,892,569,651 total letters

Query= seM_consensus_sequence
Length=533

seM BLAST results:

Sequences producing significant alignments:

Sequences	Score (Bits)	E Value
gb|AF012927.1|AF012927 *Streptococcus equi* fibrinogen-binding protein (fbp) gene, complete cds	946	0.0
emb|FM204883.1| *Streptococcus equi* subsp. *equi* 4047, complete genome	941	0.0
gb|U73162.1|SEU73162 *Streptococcus equi* M-protein (seM) gene, complete cds	929	0.0
emb|AJ249868.1| *Streptococcus equi* partial M gene for M-protein, isolate 972112	928	0.0
emb|AJ249871.1| *Streptococcus equi* partial M gene for M-protein, isolate 983257	246	1e-61
emb|AJ249869.1| *Streptococcus equi* partial M gene for M-protein, isolate 973945	202	2e-48
emb|AJ249874.1| *Streptococcus equi* partial M gene for M-protein, isolate 935844	134	9e-28
emb|AJ249873.1| *Streptococcus equi* partial M gene for M-protein, isolate 93758	117	9e-23
emb|AJ249870.1| *Streptococcus equi* partial M gene for M-protein, isolate 975102	117	9e-23
emb|AJ249872.1| *Streptococcus equi* partial M gene for M-protein, isolate 980719	104	7e-19

Alignments of *seM* Query Sequence with Streptococci Sequences:

Alignment of *seM* query sequence with: gb|AF012927.1|AF012927
Streptococcus equi fibrinogen-binding protein (fbp) gene, complete
cds
Length=2075

Score = 946 bits (512), Expect = 0.0
Identities = 524/530 (99%), Gaps = 0/530 (0%)
Strand=Plus/Plus

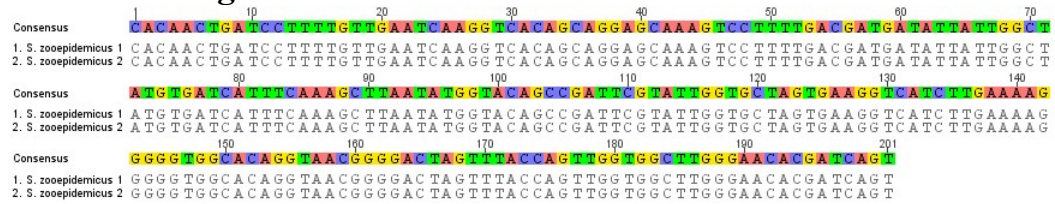
```
Query  4    AAAACTAAATGCCGGTGCAGCATCAGTATTAGTTGCAACAAGTGTGTTGGGAGGGACAAC  63
            ||||||||                                                    |||
Sbjct  362  AAAACTAAGTGCCGGTGCAGCATCAGTATTAGTTGCAACAAGTGTGTTGGGAGGGACAAC  421

Query  64   TGTAAAAGCGAACTCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAA  123
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  422  TGTAAAAGCGAACTCTGAGGTTAGTCGTACGGCGACTCCAAGATTATCGCGTGATTTAAA  481

Query  124  AAATAGATTAAGCGATATAGCCATAGGTGGAGATGCCTCATCAGCCCAAAAAGTTCGAAA  183
            ||||||||||||||| ||||||||| || ||||||||||||||||||||||||||||||
Sbjct  482  AAATAGATTAAGCGAAATAGCCATAAGTAGAGATGCCTCATCAGCCCAAAAAGTTCGAAA  541

Query  184  TCTTCTAAAAGGCGCCTCTGTTGGGGATTTACAGGCATTATTGAGAGGTCTTGATTCAGC  243
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  542  TCTTCTAAAAGGCGCCTCTGTTGGGGATTTACAGGCATTATTGAGAGGTCTTGATTCAGC  601

Query  244  AAGGGCTGCGTATGGTAGAGATGATTATTACAACTTATTGATGCACCTTTCATCGATGTT  303
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Sbjct  602  AAGGGCTGCGTATGGTAGAGATGATTATTACAATTTATTGATGCACCTTTCATCGATGTT  661

Query  304  AAATGATAAACCTGATGGGGATAGAGGACAATTAAGTTTGGCTTCATTACTTGTAGATGA  363
            |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct  662  AAATGATAAACCTGATGGGGATAGAAGACAATTAAGTTTGGCTTCATTACTTGTAGATGA  721

Query  364  AATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAACTTCTTGAGGCTAAACTTGC  423
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  722  AATTGAAAAGCGGATTGCTGATGGAGATAGTTATGCAAAACTTCTTGAGGCTAAACTTGC  781

Query  424  AGCTATTAAATCTCAACAAGAAATGCTTAGAGAAAGAGATTCCCAACTTCGAAATCTAGA  483
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  782  AGCTATTAAATCTCAACAAGAAATGCTTAGAGAAAGAGATTCCCAACTTCGAAATCTAGA  841

Query  484  GAAGGAAAAAGAACAAGAACTACAAAAAGCTAAAGATGAGCGTCAAGCTC  533
            |||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  842  GAAGGAAAAAGAACAAGAACTACAAAAAGCTAAAGATGAGCGTCAAGCTC  891
```

A.V.4. srtz Amplicon Consensus Sequence of the Two *S. zooepidemicus* Isolates in the Present Study:

>*srtz*_consensus_sequence:

CACAACTGATCCTTTTGTTGAATCAAGGTCACAGCAGGAGCAAAGTCC
TTTTGACGATGATATTATTGGCTATGTGATCATTTCAAAGCTTAATATG
GTACAGCCGATTCGTATTGGTGCTAGTGAAGGTCATCTTGAAAAGGGG
GTGGCACAGGTAACGGGGACTAGTTTACCAGTTGGTGGCTTGGGAACA
CGATCAGT

srtz isolate alignment:



Parameters from BLAST Search of *srtz* consensus sequence:

BLASTN 2.2.26+ (Mar 24, 2012)
Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14.
RID: PZTZ224A01S
Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,environmental samples or phase 0, 1 or 2 HTGS sequences)
15,916,285 sequences; 40,589,149,171 total letters
Query= srtz_consensus_sequence
Length=201

srtz BLAST results:

Sequences producing significant alignments:

Sequences	Score (Bits)	E Value
emb|FM204884.1 *Streptococcus equi* subsp. *zooepidemicus* H70, complete genome	372	6e-100

Alignment of *srtz* Query Sequence with Streptococci Sequences:

Alignment of *srtz* query sequence with: emb|FM204884.1| *Streptococcus equi* subsp. *zooepidemicus* H70, complete genome
Length=2149868

Features in this part of subject sequence:
sortase SrtC2

Score = 372 bits (201), Expect = 6e-100
Identities = 201/201 (100%), Gaps = 0/201 (0%)
Strand=Plus/Minus

```
Query  1        CACAACTGATCCTTTTGTTGAATCAAGGTCACAGCAGGAGCAAAGTCCTTTTGACGATGA  60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2026845  CACAACTGATCCTTTTGTTGAATCAAGGTCACAGCAGGAGCAAAGTCCTTTTGACGATGA
2026786

Query  61       TATTATTGGCTATGTGATCATTTCAAAGCTTAATATGGTACAGCCGATTCGTATTGGTGC  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2026785  TATTATTGGCTATGTGATCATTTCAAAGCTTAATATGGTACAGCCGATTCGTATTGGTGC
2026726

Query  121      TAGTGAAGGTCATCTTGAAAAGGGGGTGGCACAGGTAACGGGGACTAGTTTACCAGTTGG  180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2026725  TAGTGAAGGTCATCTTGAAAAGGGGGTGGCACAGGTAACGGGGACTAGTTTACCAGTTGG
2026666

Query  181      TGGCTTGGGAACACGATCAGT  201
                |||||||||||||||||||
Sbjct  2026665  TGGCTTGGGAACACGATCAGT  2026645
```

A.VI. SeM Allele Isolate Locations

The isolate locations of the SeM alleles were sourced from the SeM database (http://pubmlst.org/szooepidemicus/seM, accessed January 12, 2012) and various papers.

seM allele	Location	Reference
1	Netherlands, UK, Ireland, Sweden, Scotland	SeM database, Anzai et al. (1999), Kelly et al. (2006), Ijaz et al. (2011), Ivens et al. (2011), Lindahl et al. (2011)
2	USA, Canada, NZ, Japan	SeM database, Anzai et al. (1999), Kelly et al. (2006), Ijaz et al. (2011)
3	UK	SeM database, Kelly et al. (2006), Ivens et al. (2011)
4	UK	SeM database, Kelly et al. (2006)
5	UK	SeM database, Kelly et al. (2006)
6	UK, Sweden	SeM database, Lindahl et al. (2011), Kelly et al. (2006), Ivens et al. (2011)
7	UK	SeM database, Kelly et al. (2006), Ivens et al. (2011)
8	UK	SeM database, Kelly et al. (2006), Ivens et al. (2011)
9	UK, Sweden, USA	SeM database, Kelly et al. (2006), Lindahl, et al. (2011), Ivens et al. (2011), Anzai, et al. (1999)
10	UK	SeM database, Kelly et al. (2006)
11	UK	SeM database, Kelly et al. (2006)
12	UK	SeM database, Kelly et al. (2006)
13	UK	SeM database, Kelly et al. (2006)
14	UK	SeM database, Kelly et al. (2006)
15	Australia	SeM database, Kelly et al. (2006), Ijaz et al. (2011), Anzai et al. (1999)
16	UK	SeM database
17	No Data	
18	Japan	Anzai et al. (1999)
19	Japan	Anzai et al. (1999)
20	Japan	Anzai et al. (1999)
21	Japan	Anzai et al. (1999)
22	USA	Ijaz et al. (2011), Anzai et al. (1999)
23	Ireland	Ijaz et al. (2011), Anzai et al. (1999)
24	USA	Anzai et al. (1999)
25	USA	Anzai et al. (1999)
26	USA	Ijaz et al. (2011), Anzai et al. (1999)
27	USA	Ijaz et al. (2011), Anzai et al. (1999)
28	UK, USA	SeM database, Ivens et al. (2011), Ijaz et al. (2011), Anzai et al. (1999)
29	USA	Anzai et al. (1999)
30	USA	Anzai et al. (1999)
31	USA	Ijaz et al. (2011), Anzai et al. (1999)
32	USA	Ijaz et al. (2011), Anzai et al. (1999)
33	USA	Anzai et al. (1999)
34	USA	Anzai et al. (1999)
35	Ireland	Anzai et al. (1999)
36	Sweden	Ijaz et al. (2011), Anzai et al. (1999)
37	USA	Anzai et al. (1999)
38	USA	Anzai et al. (1999)
39	Sweden, USA	Lindahl et al. (2011), Ijaz et al. (2011), Anzai et al. (1999)
40	Brazil	(Anzai, Timoney et al. 1999)
41	Brazil	(Anzai, Timoney et al. 1999)
42	USA	Ijaz et al. (2011), Anzai et al. (1999)
43	Sweden	SeM database, Lindahl et al. (2011), Anzai et al. (1999)
44	UK	SeM database
45	UK	SeM database
46	UK	SeM database, Ivens et al. (2011)
47	UK	SeM database
48	UK	SeM database, Ivens et al. (2011)
49	No Data	
50	UK	SeM database
51	UK	SeM database
52	UK	SeM database

Continues next page….

seM allele	Location	Reference
53	UK	SeM database
54	UK	SeM database
55	UK	SeM database
56	UK	SeM database
57	UK	Ivens et al. (2011)
58	UK	SeM database, Ivens et al. (2011)
59	UK	SeM database, Ivens et al. (2011)
60	UK	Ivens et al. (2011)
61	UK	Ivens et al. (2011)
62	UK	SeM database, Ivens et al. (2011)
63	UK	SeM database, Ivens et al. (2011)
64	UK	Ivens et al. (2011)
65	UK	Ivens et al. (2011)
66	UK	Ivens et al. (2011)
67	UK	SeM database
68	UK	SeM database
69	UK	SeM database
70	UK	SeM database
71	Sweden	SeM database, Lindahl et al. (2011)
72	UK, Sweden	SeM database, Lindahl et al. (2011)
73	UK	SeM database
74	UK	SeM database
75	UK, Pakistan	SeM database, Ijaz et al. (2011)
76	UK, Sweden, Germany	SeM database, Lindahl et al. (2011), Ijaz et al. (2011)
77	Sweden, Denmark	SeM database, Lindahl et al. (2011), Ijaz et al. (2011)
78	Sweden, USA	SeM database, Lindahl et al. (2011), Ijaz et al. (2011)
79	UK, USA	SeM database, Ijaz et al. (2011)
80	UK, Scotland	SeM database, Ijaz et al. (2011)
81	UK	SeM database
82	UK	SeM database
83	UK	SeM database
84	UK	SeM database
85	UK	SeM database
86	Sweden	SeM database, Lindahl et al. (2011)
87	UK, Sweden	SeM database, Lindahl et al. (2011)
88	No Data	
89	Pakistan, Germany	SeM database
90	Denmark	SeM database
91	USA	SeM database
92	USA	SeM database
93	Scotland	SeM database
94	USA	SeM database
95	USA	SeM database
96	USA	SeM database
97	No Data	
98	No Data	
99	NZ	SeM database, Present Study
100	NZ	SeM database, Present Study
101	No Data	
102	No Data	
103	No Data	
104	No Data	
105	No Data	
106	No Data	
107	No Data	
108	No Data	
109	No Data	
110	No Data	
111	No Data	
112	No Data	
113	No Data	
114	No Data	

Supplementary Information

All supplementary information is included on the compact disk supplied.

S.I License Agreements for the Use of Figures Sourced in Jounal Publications

S.I.1. Licence agreement for Figure 1.3: guttural pouch chondroid

S.I.2. Licence agreement for Figure 3.1: SeM schematic

S.I.3. Licence agreement for Figure 3.2: Sau-PCR graphical representation

References

Al-Ghamdi, G. M., A. Amonsen, et al. (2008). "Use of Repetitive Sequence-Based Polymerase Chain Reaction to Charaterize *Streptococcus equi* Subspecies *Equi* Isolates Cultured from Horses Experiencing Post-Vaccinal Reation." Journal of Animal and Veterinary Advances **7**(8): 892-896.

Al-Ghamdi, G. M., V. Kapur, et al. (2000). "Use of repetitive sequence-based polymerase chain reaction for molecular epidemiologic analysis of *Streptococcus equi* subspecies *equi*." American Journal of Veterinary Research **61**(6): 699-705.

Alber, J., A. El-Sayed, et al. (2004). "Multiplex Polymerase Chain Reaction for Identification and Differentiation of *Streptococcus equi* subsp. *zooepidemicus* and *Streptococcus equi* subsp. *equi*." Journal of Veterinary Medicine Series B **51**(10): 455-458.

Anilkumar, A. K., G. K. Madhavilatha, et al. (2012). "Standardization and evaluation of a tetraplex polymerase chain reaction to detect and differentiate *Mycobacterium tuberculosis* complex and nontuberculous *Mycobacteria*-a retrospective study on pulmonary TB patients." Diagnostic Microbiology and Infectious Disease **72**: 239-247.

Anzai, T., S. Hobo, et al. (2006). "Development of a Diagnostic Shuttle PCR Test Targeting the *Streptococcus equi* SeM Gene." Journal of Equine Science **17**(4): 101-104.

Anzai, T., Y. Kuwamoto, et al. (2005). "Variation in the N-terminal region of an M-like protein of *Streptococcus equi* subsp. *equi* and evaluation of its protential as a tool in epidemiologic studies." American Journal of Veterinary Research **66**: 2167-2171.

Anzai, T., J. F. Timoney, et al. (1999). "In vivo pathogenicity and resistance to phagocytosis of *Streptococcus equi* strains with different levels of capsule expression." Veterinary Microbiology **67**: 277-286.

Artiushin, S. C., Y. Tong, et al. (2011). "Thermophilic helicase-dependant DNA amplification using the IsoAmpTM SE experimental kit fro rapis detectio of *Streptococcus equi* subspecies *equi* in clinical samples." Journal of Veterinary Diagnostic Investigation **23**(5): 909-914.

Båverud, V., S. K. Johansson, et al. (2007). "Real-time PCR for Detection and Differentiation of *Streptococcus equi* subsp. *equi* and *Streptococcus equi* subsp. *zooepidemicus*." Veterinary Microbiology **124**: 219-229.

Borst, L. B., S. K. Patterson, et al. (2011). "Evaluation of a commercially available modified-live *Streptococcus equi* subsp *equi* vaccine in ponies." American Journal of Veterinary Research **72**(8): 1130-1138.

Boye, K., E. Høgdall, et al. (1999). "Identification of bacteria using two degenerate 16S rDNA sequencing primers." Microbiological Research **154**: 23-26.

Chanter, N., N. Collin, et al. (1997). "Characterization of the Lancefield Group C *Streptococcus* 16S-23S RNA Gene Intergenic Spacer and its Potential for Identification and Sub-Specific Typing." Epidemiology and Infection **118**: 125-135.

Chanter, N., N. C. Talbot, et al. (2000). "*Streptococcus equi* with truncated M-proteins isolated from outwardly healthy horses." Microbiology **146**(6): 1361-1369.

Chomczynski, P. and M. Rymaszewski (2006). "Alkaline Polyethylene Glycol-Based Method for Direct PCR from Bacteria, Eukaryotic Tissue Samples, and Whole Blood." BioTechniques **40**: 454-458.

Cobb, B. D. and J. M. Clarkson (1994). "A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods." Nucleic Acids Research **22**(18): 3801-3805.

Corich, V., A. Mattiazzi, et al. (2005). "Sau-PCR, a Novel Amplification Technique for Genetic Fingerprinting of Microorganisms." Applied and Environmental Microbiology **71**(10): 6401-6406.

Don, R. H., P. T. Cox, et al. (1991). "'Touchdown' PCR to circumvent spurious priming during gene amplification." Nucleic Acids Research **19**(14): 4008.

Drummond, A. J., B. Ashton, et al. (2011). "Geneious v5.4." Available from http://www.geneious.com/.

Efstratiou, A., G. Colman, et al. (1994). "Biochemical Differences Among Human and Animal Streptococci of Lancefield Group C or Group G." Journal of Medical Microbiology **41**: 145-148.

Elnifro, E. M., A. M. Ashshi, et al. (2000). "Multiplex PCR: Optimization and Application in Diagnostic Virology." Clinical Microbiology Reviews **13**(4): 559-570.

Erol, E., S. J. Locke, et al. (2012). "Beta-hemolytic *Streptococcus* spp. from horses: a retrospective study (2000-2010)." Journal of Veterinary Diagnostic Investigation **24**(1): 142-147.

Flock, M., Å. Karlström, et al. (2006). "Protective Effect of Vaccination with Recombinant Proteins from *Streptococcus equi* subspecies *equi* in a Strangles Model in the Mouse." Vaccine **24**: 4144-4151.

Foschino, R., D. Nucera, et al. (2008). "Comparison of *Lactococcus garvieae* strains isolated in northern Italy from dairy products and fishes through molecular typing." Journal of Applied Microbiology **105**(3): 652-662.

Galán, J. E. and J. F. Timoney (1988). "Immunologic and genetic comparison of *Streptococcus equi* isolates from the United States and Europe." Journal of Clinical Microbiology **26**(6): 1142-1146.

Grant, S. T., A. Efstratiou, et al. (1993). "Laboratory diagnosis of strangles and the isolation of atypical *Streptococcus equi*." The Veterinary Record **33**: 215-216.

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

Guss, B., M. Flock, et al. (2009). "Getting to Grips with Strangles: An Effective Multi-Component Recombinant Vaccine for the Protection of Horses from *Streptococcus equi* Infection." PLoS Pathog **5**(9): e1000584.

Harrington, D. J., I. C. Sutcliffe, et al. (2002). "The Molecular Basis of *Streptococcus equi* Infection and Disease." Microbes and Infection **4**: 501-510.

Hassan, A. A., I. U. Khan, et al. (2003). "Inter- and Intraspecies Variations of the 16S-23S rDNA Intergenic Spacer Region of Various Streptococcal Species." Systematic and Applied Microbiology **26**: 97-103.

Henegariu, O., N. A. Heerema, et al. (1997). "Multiplex PCR: Critical Parameters and Step-by-Step Protocol." BioTechniques **23**(2): 504-511.

Holden, M. T. G., Z. Heather, et al. (2009). "Genomic Evidence for the Evolution of *Streptococcus equi*: Host Restriction, Increased Virulence, and Genetic Exchange with Human Pathogens." PLoS Pathog **5**(3): e1000346.

Ijaz, M., S. Velineni, et al. (2011). "Selective pressure for allelic diversity in SeM of *Streptococcus equi* does not affect immunoreactive proteins SzPSe of Se18.9." Infection, Genetics and Evolution.

Ivens, P. A. S., D. Matthews, et al. (2011). "Molecular characterisation of 'strangles' outbreaks in the UK: The use of M-protein typing of *Streptococcus equi* ssp. *equi*." Equine Veterinary Journal **43**(3): 359-364.

Jannatabadi, A. A., G. R. Mohammadi, et al. (2008). "Molecular Identification of *Streptococcus equi* subsp. *equi* and *Streptococcus equi* subsp. *zooepidemicus* in Nasal Swabs Samples from Horses Suffering Respiratory Infections in Iran." Pakistan Journal of Biological Sciences **11**(3): 468-471.

Jorm, L. R. (1992). "Laboratory studies on the survival of *Streptococcus equi* subspecies *equi* on surfaces." Equine Infectious Diseases VI: Proceedings of the Sixth International Conference, Newmarket (United Kingdom): R and W Publications (Newmarket) Ltd **7-11 July 1991**: 39-43.

Jorm, L. R., D. N. Love, et al. (1994). "Genetic structure of populations of beta-haemolytic Lancefield group C streptococci from horses and their association with disease." Research in Veterinary Science **57**: 292-299.

Kelly, C., M. Bugg, et al. (2006). "Sequence Variation of the SeM Gene of Streptococcus equi Allows Discrimination of the Source of Strangles Outbreaks." Journal of Clinical Microbiology **44**(2): 480-486.

Kuwamoto, Y., T. Anzai, et al. (2001). "Microplate Sugar-Fermentation Assay Distinguishes *Streptococcus equi* from Other Streptococci of Lancefield's Group C." Journal of Equine Science **12**(4): 47-49.

Lanka, S., L. B. Borst, et al. (2010). "A Multiphasic Typing Approach to Subtype *Streptococcus equi* subspecies *equi*." Journal of Veterinary Diagnostic Investigation **22**: 928-936.

Laus, F., S. Preziuso, et al. (2007). "Clinical and Epidemiological Investigation of Chronic Upper Respiratory Diseases Caused by Beta-Haemolytic Streptococci in Horses." Comparative Immunology, Microbiology & Infectious Diseases **30**: 247-260.

Lindahl, S., R. Soderlund, et al. (2011). "Tracing outbreaks of *Streptococcus equi* infection (strangles) in horses using sequence variation in the seM gene and pulsed-field gel electrophoresis." Veterinary Microbiology **In Press, Corrected Proof**.

Markoulatos, P., N. Siafakas, et al. (2002). "Multiplex Polymerase Chain Reaction: A Practical Approach." Journal of Clinical Laboratory Analysis **16**: 47-51.

Monis, P. T. and S. Giglio (2006). "Nucleic acid amplification-based techniques for pathogen detection and identification." Infection, Genetics and Evolution **6**: 2-12.

Newton, J. R., K. Verheyen, et al. (2000). "Control of Strangles Outbreaks by Isolation of Guttural Pouch Carriers Identified Using PCR and Culture of *Streptococcus equi*." Equine Veterinary Journal **32**(6): 515-526.

Newton, J. R., J. L. N. Wood, et al. (1997). "Naturally occuring persistent and asymptomatic infection of the guttural pouches of horses with *Streptococcus equi*." The Veterinary Record **140**: 84-90.

Parkinson, N. J., C. Robin, et al. (2011). "Molecular epidemiology of strangles outbreaks in the UK during 2010." Veterinary Record **168**(25): 666.

Posada, D. (2008). "jModelTest: Phylogenetic Model Averaging." Molecular Biology and Evolution **25**: 1253-1256.

Preziuso, S., F. Laus, et al. (2010). "Detection of *Streptococcus dysgalactiae* subsp. *equisimilis* in equine nasopharyngeal swabs by PCR." Journal of Veterinary Science **11**(1): 67-72.

Proietti, P. C., A. Bietta, et al. (2011). "Isolation and charaterizaton of β-haemoytic-Streptococci from endometritis in mares." Veterinary Microbiology **152**: 126-130.

Pusterla, N., P. H. Kass, et al. (2011). "Surveillance programme for important equine infectious respiratory pathogens in the USA." Veterinay Record **169**(1): 12.

Ririe, K. M., R. P. Rasmussen, et al. (1997). "Product differentiation by analysis of DNA melting curves during the polymerase chain reaction." Analytical Biochemistry **245**(2): 154-160.

Rogers, C. W. and N. Cogger (2010). "A cross-sectional survey of biosecurity practices on Thoroughbred stud farms in New Zealand." New Zealand Veterinary Journal **58**(2): 64-68.

Rozen, S. and H. Skaletsky (2000). Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S., Misener, S., (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Totowa, NJ, Humana Press.

Smith, P. A. (2006). "How to eliminate strangles infections caused by silent carriers." AAEP Proceedings **52**: 101-103.

Sweeney, C. R., J. F. Timoney, et al. (2005). "*Streptococcus equi* infections in horses: guidelines for treatment, control, and prevention of strangles." Journal of Veterinary International Medicine **19**: 123-134.

Takai, S., T. Anzai, et al. (2000). "Detection of DNA restriction fragment polymorphisms in *Streptococcus equi*." The Veterinary Record **146**: 159-161.

Tamura, K., D. Peterson, et al. (2011). "MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods." Molecular Biology and Evolution **28**: 2731-2739.

Taylor, S. D. and D. Wilson (2006). "*Streptococcus equi* subsp. *equi* (Strangles) Infection." Clinical Techniques in Equine Practice **5**: 211-217.

Timoney, J. F. (1993). "Strangles." The Veterinary clinics of North America: Equine Practice **9**(2): 365-374.

Timoney, J. F. (2004). "The Pathogenic Equine Streptococci." Veterinary Research **35**: 397-409.

Timoney, J. F. and S. C. Artiushin (1997). "Detection of *Streptococcus equi* in equine nasal swabs and washes by DNA amplification." The Veterinary Record **141**: 446-447.

Timoney, J. F., S. C. Artiushin, et al. (1997). "Comparison of the Sequences and Functions of *Streptococcus equi* M-Like Proteins SeM and SzPSe." Infection and Immunity **65**(9): 3600-3605.

Timoney, J. F., R. DeNegri, et al. (2010). "Affects of N-terminal variation in the SeM protein of *Streptococcus equi* on antibody and fibrinogen binding." Vaccine **28**(6): 1522-1527.

Vincen, M., Y. Xu, et al. (2004). "Helicase-dependent isothermal DNA amplification." EMBO reports **5**: 795-800.

Walker, J. A. and J. F. Timoney (2002). "Construction of a Stable Non-Mucoid Deletion Mutant of the *Streptococcus equi* Pinnacle Vaccine Strain." Veterinary Microbiology **89**: 311-321.

Waller, A. S. and K. A. Jolley (2007). "Getting a Grip on Strangles: Recent Progress Towards Improved Diagnostics and Vaccines." The Veterinary Journal **173**: 492-501.

Waller, A. S., R. Paillot, et al. (2011). "*Streptococcus equi*: a pthogen restricted to one host." Journal of Medical Microbiology.

Webb, K., K. A. Jolley, et al. (2008). "Development of an unambiguous and discriminatory multilocus sequence typing scheme for the *Streptococcus zooepidemicus* group." Microbiology **154**: 3016-3024.

Weese, J. S., C. Jarlot, et al. (2009). "Survival of *Streptococcus equi* on surfaces in an Outdoor Environment." Canadian Veterinary Journal **50**: 968-970.

Yan, H., L. Shi, et al. (2008). "Usefulness of Sau-PCR for molecular epidemiology of nosocomial outbreaks to *Burkholderia cepacia* which occured in a local hospital in Guangzhou, China." Microbiology and Immunology **52**(5): 283-286.