Identification of a putative IS element in *Streptococcus equi* subspecies *equi*

Åsa Karlström

Institutionen för mikrobiologi
Sveriges lantbruksuniversitet

Uppsala

Examensarbete 2001:12

ISSN 1101-8151
ISRN SLU-MIKRO-EX—01/02--SE

Identification of a putative IS element in *Streptococcus equi* subspecies *equi*

Åsa Karlström

Examensarbete för bioteknologiprogrammet
utfört vid Institutionen för mikrobiologi, SLU

Handledare: **Bengt Guss**
Karin Jacobsson
Lars Frykberg

Institutionen för mikrobiologi
Sveriges lantbruksuniversitet

Uppsala

Examensarbete 2001:12

ISSN 1101-8151
ISRN SLU-MIKRO-EX—01/02--SE

Abstract

The gram-positive, catalase negative bacteria *Streptococcus equi* subspecies *equi* is the causative agent for a disease in horses called strangles. The disease is spread worldwide and apart from the suffering of horses, it causes severe economical losses. The classic description of strangles includes swelling and infection of the throat region, although the symptoms can vary from case to case. The bacteria enter the horse via the mouth or nose and are then transported to the lymphoid glands.

To investigate the infection process the bacterial virulence factors have to be identified and analysed. When studying the bacteria/ host interaction various methods are used, one method called phage display has previously been successfully used to identify bacterial proteins interacting with host components.

In this work different studies have been performed. A shotgun phage display library of *S. equi* subsp. *equi* strain 62 was affinity selected against both horse serum and horse blood respectively. The aim was to investigate if any new potential virulence factors involved in the adhesion process between the bacteria and the horse could be discovered.

The clones found in two types of experiments had inserts homologous to previously described proteins. The panning against horse serum resulted in phagemid clones which harboured inserts which were homologous to the genes encoding the fibronecting-binding protein FNZ (Lindmark *et al*, 1996) and the ?$_2$-macroglobulin, albumin and IgG-bindning protein ZAG (Jonsson *et al*, 1995). Where as in the panning against erythrocytes the inserts of the phagemid clones were homologous to an IS element 861 found in *Streptococcus agalactiae* (Rubens *et al*, 1989) as well as an IS element associated protein found in *S. pyogenes* (Feretti *et al*, 2001).

CONTENTS

INTRODUCTION

Bacteria are present in our everyday life and can be divided into helpful and harmful. The bacterium *Streptococcus equi* is a species responsible for a disease in horses called strangles. For the horse industry the possibility of getting an improved treatment and diagnostics for the disease Strangles is an important achievement, both in order to reduce vast economical losses but also to reduce the suffering in horses. Strangles is a disease that can result, not only, in decreased health and stamina but also, in the worst-case scenario death of the horse. When studying the bacteria/ host interaction various methods are used. One method called phage display has previously successfully been used to identify bacterial proteins interacting with host components.

Strangles

Strangles is a commonly used term to describe the disease that is caused by the bacterium *S. equi* subsp. e*qui* that affects horses. The disease is spread worldwide and the classic description of strangles includes swelling and infection of the throat region, although the symptoms can vary from case to case. In uncomplicated cases the disease will be localised to upper respiratory tract together with fever, coughing, pus discharging from the nostrils and abscesses forming in the lymph glands. In more severe cases when the bacteria are found in the lower tract pneumonia as well as heart muscle inflammation can be further complications. Although strangles is not primarily a mortal disease it can cause chronic illness and sometimes death (Timoney, 1988).

The most common causes of bacterial infections in horses are due to opportunistic bacteria. This is, however, not the case with strangles that is caused by subsp. *equi*. The other subspecies of the genus *S. equi* is subsp. *zooepidemicus*. This subspecies is one of these opportunistic bacteria that will cause similar symptoms, which sometimes makes it difficult to determine whether or not there is an outbreak of strangles.

Subsp. *equi* enters the horse via the mouth or nose and the bacteria are transported to the lymphoid glands through the mucosal layer. In the glands the bacteria form long chains of cocci. The incubation time for the disease varies from 3 to 14 days. The normal temperature of a horse range between 37.5?C and 38.5?C but during the disease the temperature may increase to above 39?C. During the incubation time the infected horse is highly contagious which is one of the reasons why stables, where an infected horse is found, is put in quarantine for at least 30 days. With the increased temperature comes forming of abscesses in lymph glands, a slight cough and difficulties in swallowing follows. As the disease progresses the abscesses can grow hard and painful and the pressure on the retorpharyngeal lymph nodes can cause respiratory difficulties as the name of the disease indicates. In severe cases the bacteria move into the lower respiratory tract and may cause pneumonia. In horses with the most severe form of strangles, any part of the body or viscera may be affected, including the heart (Timoney 1993).

The disease is most commonly transmitted through infective secretions. Once the disease establishes itself on a farm or ranch, it often becomes a persistent as well as a recurrent problem even though there may be long periods when no resident horse has typical signs of the disease. Once the organism establishes itself in a susceptible population, the percentage of

animals affected is generally quite high and often approaches 100%. However, the mortality rate in most uncomplicated cases is low. The disease is often treated with penicillin even though in several cases that will not be necessary as long as the horse has an appetite and not too high fever.

The genus *Streptococcus*

Streptococcus is described in Bergey's manual (1986) as spherical or sometimes ovoid shaped cells, gram-positive cocci, catalase negative bacteria that usually are facultatively anaerobic. They grow in pairs or chains (Figure 1) and temperature optimum is usually 37?C even though minimum and maximum temperature vary within the genus. Not all species form capsules, but it is not uncommon. Streptococci are responsible for a large number of important diseases in man and in animals. They are found on mucose membranes of the mouth, respiratory and genitourinary tracts as well as on the skin. The pyogenic group are regarded as pathogenic or opportunistic pathogens of mammals (Hardie & Whiley 1995). The most studied member of the pyogenic group is *Streptococcus pyogenes*, which belongs to the Lancefield serological group A streptococci, which is the causative agent of various diseases in humans, ranging from mild infections in the skin area to more severe diseases as scarlet fever.

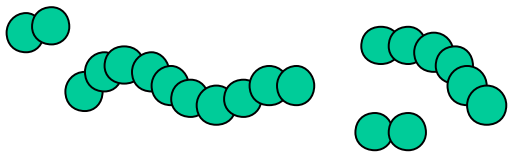


Fig 1. Streptococci grow in chains or in pair.

Streptococcus equi

Streptococcus equi is ?-hemolytic, belongs to the Lancefield group C streptococci and compromise two subspecies, subsp. *equi* and subsp. *zooepidemicus* (Lancefield 1933, Hardie & Whiley 1995). The cells are 0.6-1.0 ?m in diameter and capsules are demonstrated in some strains either when young cultures are examined or when serum is added to the growth medium.

Subsp. *zooepidemicus* is regarded as an opportunistic pathogen that belongs to the normal bacterial flora in horses and will normally not cause disease in a healthy horse. The organism is found in the respiratory region, the uterus and wounds where it occasionally may give rise to an infection. This subsp. has also been isolated from humans as well as from a number of mammals including rat, pig, cat, cattle and mouse (Ruoff, 1992).

Subsp. *equi* is an obligate pathogen and considered to be a parasite in the upper respiratory tract in horses (equimilius). Since donkeys and mules belong to the same taxonomic group, Equidae, they also can be infected. The bacteria cause strangles, a disease spread worldwide which give rise to a pathogenic colonisation in the upper respiratory tract and will affect the animal in various ways (SVAs hästbok, 1998).

Streptococci and host infection

Streptococci and their pathogenesis vary between the different species and subspecies. Most of the streptococci belonging to the *pyogenic* group however are considered pathogens. Pathogenic streptococci express a number of different compounds in order to establish infections of their host. Of these some are secreted into the growth medium whereas others are associated to the cell wall. The compounds expressed include various products that are considered important for the bacteria such as a capsule and cell surface anchored proteins. The capsule, made by hyaluronic acid, may not be present in all pathogenic species but is present in some (Ruoff, 1992). Many of the cell surface anchored proteins in gram positive bacteria, like streptococci, display a common structural organisation and a certain kind of motifs that bind the protein in the cell wall (Navarre & Schneewind, 1999). The motifs are; a signal sequence in the N-terminal which directs the secretion through the cell membrane, a hydrophobic domain close to the C-terminal and a LPXTG motif preceding the hydrophobic C-terminal domain. All these motifs are essential in the protein/cell wall linkage process. Example of host proteins streptococcal cells can bind to the surface are IgG, ?$_2$-macroglobulin, albumin, fibronectin and fibrinogen. In addition the bacteria can secret proteins that have other activities, for example haemolysins and streptolysins (Navarre & Schneewind, 1999). The relative importance of the various potential virulence factors between species and site probably varies.

Identification of potential virulence factors

In order to identify potential virulence factors the molecules involved in adhesion and infection have to be both identified and analysed. The processes of investigating such factors include gene cloning and different strategies to express the genes in another host, usually *E. coli*. Phage display is a technique developed for expressing proteins in fusion with the phages' own coat proteins (Smith, 1985). The *E. coli* filamentous phage M13 has a capsid consisting of approximately 2700 copies of the major coat protein, pVIII, and a fewer number of the proteins pIII, pVI, pVII and pIX. (Figure 2). The phages used are filamentous phages that will infect a bacterium and their progeny will be secreted without killing the bacteria. If foreign DNA is inserted in fusion with a gene encoding a phage coat protein, the foreign genes will be expressed on the surface of the phage. The phage display technique was enhanced by Bass *et al* in 1990 who instead of a phage used phagmid-vectors. A phagemid is a plasmid with a phage origin of replication and the filamentous phage genes encoding phage proteins, usually pIII or pVIII. If foreign DNA is inserted with an open reading frame in fusion with a phage protein a fusionprotein could be expressed. The fused protein encoded by the phagemid can by co-infection of a helper phage be displayed on the surface of a phagemid particle. This is more valuable since the copy number of the expressed fusion protein will decrease, thus allow studies of weak and strong binding phages. The use of a pVIII phagemid-vector will result in an expression of a variant

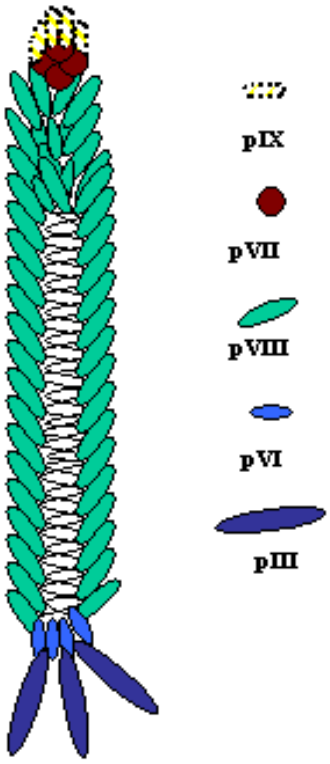


Fig. 2 Coat proteins on phage

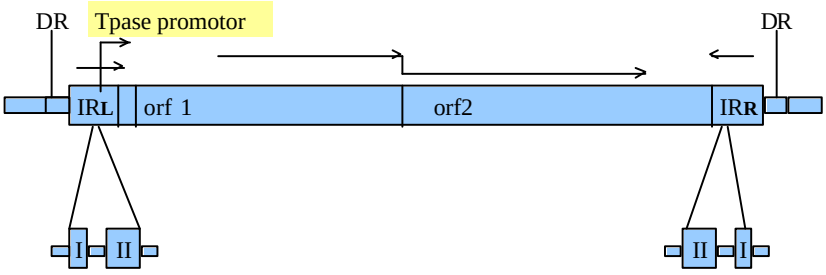
numbers of copies of the protein in question per phages compared to the pIII phage system. Most commonly used for phage display are the proteins pIII and pVIII (Jacobsson 1996, Jacobsson & Frykberg 2001).

Transposable elements

Transposable elements, or transposons, are mobile parts of DNA that are able to change their location within genomes. They are present in chromosomal DNA as well as in plasmids. They have no apparent function apart from their own replication but when inserted into a genome they usually cause differences in expression of the gene affected (Basic genetics, 1995).

Bacterial insertion sequences (IS) contain only the elements required for transposition, usually they only code for the necessary transposase that enables the transposition itself. They are the simplest transposons found in both prokaryotes and eukaryotes. The number of IS elements in prokaryotes reported is steadily increasing, in 1998 more than 500 had been identified (Mahillon & Chandler 1998). The vast numbers of elements as well as the different positions and species they are associated with makes the nomenclature a problem. As of today, two different systems are in use, one presented in 1978 and later refined (Lederberg 1981) numbers the IS elements i.e. IS2. The other system provides a small piece of information on the source of the element, i.e. ISSp1 for the first described element in *Streptococcus pyogenes*. Some of the closely related elements with few differences in nucleotide sequence level are classified as isoforms of a parent element, thus named differently than stated above (Mahillon & Chandler 1998).

The IS elements have a general characteristic structure even though the elements are different; short terminal inverted repeats and an open reading frame (or sometimes several) between these sequences (Figure 3). The direct repeats are a consequence of the insertion of the



DR – Direct repeats
I – Terminal base pairs, recognition for Tpase-mediated cleavage
II – Base pairs necessary for site-specific recognition and binding by the Tpase
IRL – left inverted repeat (10-40 bp)
IRR – right inverted repeat (10-40 bp)
Orf 1 and 2 – open reading frame 1 and 2 (~500 bp and ~800 bp respectively)

Fig 3. Insertion sequence element
(Mahillon & Chandler, 1998)

element in the DNA sequence of the host genome. The short terminal inverted repeats (IR) are generally between 10 and 40 bp and present on each side of the element and are more closely similar than actually identical. The open reading frame consists of either one or two open reading frames and these encode the transposase and usually consumes the entire length of the

element. The shift between the two reading frames are due to a translational frame shifting, where a sequence of "slippery" codons makes it possible to change reading frame under certain circumstances (Farabaugh, 1996).

Aim of the project

The idea was to use phage display in order to identify novel interactions between subsp. *equi* and the host. This was planned to be done in two parallel set of experiments,

- in a first set of experiments panning of a genomic phage display library of subsp. *equi* strain 62 against horse serum
- in another set of experiments a similar panning against horse RBC (erythrocytes)

The results could not only increase the knowledge of adhesion and infection by streptococci but also in the long run be of help in order to produce vaccines and diagnostics.

MATERIALS AND METHODS

Panning against serum from horse

Microtiter wells (Nunc, Copenhagen, Denmark) were coated with 190 ?l coating buffer (0.5 M $NaHCO_3$) and 10 ?l serum from horse for 2 h in room temperature (RT). The microtiter wells were then washed 15 times with phosphate-buffered saline supplemented with Tween 20 at concentration 0.05% (PBST). The wells were then blocked with PBST and 8 ?l of casein ?200 ?g/?l? for 30 min in RT and then washed thoroughly with PBST. All the incubations were performed on a shaker (IKA-Schüttler MTS4) in RT. 200 ?l of the phagemid library of *S. equi* subsp. *equi* strain 62 (Lindmark *et al.*, 2001) was then added to each well. Prior to addition the samples had been blocked for 15 min in RT with two different mixtures. Sample S1 had been blocked with 4 ?l human IgG ?10 mg/ml?, 10 ?l human Fg ?20 mg/ml?, 5.9 ?l human serum albumin ?3,5 mg/ml?, and 2 ?l bovine casein ?10mg/ml? whereas sample S2 had been blocked with 8 ?l casein ?10mg/ml?. After incubation for 2 h in the microtiter wells were washed thoroughly with PBST, then incubated with PBST for 5 minutes, before the washing continued. After the extensive washing the wells were eluted with 200 ?l elution buffer ?50mM citrate, 140 mM NaCl, pH 2.0] and then neutralised with 75 ?l Tris ?pH 8?. *E. coli* TG1 cells (20 ?l) from an over night culture in 100 ?l Luria-Bertani medium (LB) were infected for 30 min in RT with different conc. of the eluted phagemid particles, i.e. with 1 ?l, 5 ? and 50 ?l phagemid particles respectively. The infected cells were then spread on plates containing LB medium supplemented with 1.5 % agar and ampicillin [50?l/ml] (LAA-plates) and incubated at 37? C over night. Next day the colonies were counted and ~100 picked for E-tag screening. After picking the colonies the remaining colonies were pooled together in 6 ml of LB medium in a sterile glass tube where 20 ?l of the helper phage R408 [10^{12} pfu/ml] was added and incubated for 20 minutes. LB soft agar (0.5 %) was then added and the solution was spread on LAA plates for incubation over night in 37? C. After incubation the phagemid particles was extracted and the panning procedure was repeated.

Panning experiments using washed erythrocytes from horse

To obtain the erythrocytes (RBCs) horse blood was prepared by an extensive wash with cold phosphate-buffered saline (PBS) and centrifuged in order to remove as much of the additional components in the blood as possible. The washing procedure was as follows. Five ml blood and five ml PBS was mixed and centrifuged, the supernatant removed and the pellet resuspended in PBS. One ml of the blood was then mixed with 1 ml of PBS in a 2 ml Eppendorf tube. The solution was centrifuged and the supernatant removed once more. The procedure was repeated four times until the supernatant was clear. The final pellet was diluted with PBS and 20 ?l casein [200 ?g/?l? and the phage library (200 ?l) of *S. equi* sp. *equi* strain 62 was added. Approximately 1.8 ml of PBS was added and the Eppendorf tube was placed in +2?C (IKA-Schüttler MTS4) for 3 h. After incubation the RBC were washed with cold PBS and after the last withdrawal of supernatant, 100 ?l PBS was added in order to make the solution easier to handle.
E. coli TG1 cells (20 ?l) from an over night culture in 100 ?l LB medium were infected for 30 min in RT with different conc. of the washed red blood cells with attached phagemid particles, i.e. with 1 ?l, 5 ? and 50 ?l blood respectively. The cells were then spread on LAA-plates and incubated at 37? C over night. Next day the colonies were counted and ~100 picked for E-tag screening. After picking the colonies the remaining colonies were pooled together in 6 ml of LB medium in a sterile glass tube where 20 ?l of the helper phage R408 [10^{12} pfu/ml] was added and incubated for 20 minutes. Five ml of LB soft agar (0.5 %) was then added and the solution was spread on LAA plates for incubation over night in 37? C. After incubation the phagemid particles was extracted and the panning procedure was repeated.

Enhanced experiment

After the initial washing of the blood the washed blood was divided into two tubes; A and B. Five hundred ?l blood and 500?l of cold PBS was added in the Eppendorf tubes respectively. In A 200 ?l phagemid library of *S. equi* sp. *equi* strain 62 and 20 ?l of casein [200 ?g/?l? was added. In B 220 ?l pre-treated phagemid library of *S. equi* sp. *equi* strain 62 and 20 ?l casein [200 ?g/?l? and 20 ?l salmon sperm DNA were added. The pre-treated phagemid library was 200 ?l phagemid library with added 20 ?l salmon sperm DNA [10 mg/ml]and incubated for 15 min before mixing with the horse RBCs. Both the tubes were incubated at +2?C for 4 h. After incubation the RBCs were washed with cold PBS ten times. After four washings PBS was added and incubated at 2?C (IKA-Schüttler MTS4) for 5 min, then the washings continued. After the last withdrawal of supernatant 100 ?l of PBS was added in order to make the solution easier to handle.

E. coli TG1 cells (20 ?l) from an over night culture in 100 ?l LB medium were infected for 30 min in RT with different conc. of the washed red blood cells with attached phagemid particles, i.e. with 1 ?l, 5 ? and 50 ?l blood respectively. The infected cells were then spread on six LAA-plates incubated at 37? C over night. After the incubation the colonies were counted and picked for E-tag screening and the remaining colonies were pooled together, within each sample A and B, in 5 ml of LB medium in a sterile glass tube where 20 ?l of the helper phage R408 [10^{12} pfu/ml] was added and incubated for 20 minutes. Five ml LB soft agar (0.5%) was added and the solution was spread on LAA plates for incubation over night in 37? C. After the incubation the phagemid particles was extracted and the panning procedure was repeated once, though the volumes of the eluate used for infection of *E. coli* TG1 cells used ranged from 10 ?l to $1x10^{-5}$?l with the interval of 10 (Table 2).

E-tag screening

Colonies were picked from each of the LAA-plates onto new LAA-plates using a grid. The colonies were let to grow for a few hours, and then a nitro-cellulose (NC) filter was used in order to make replicates. The colonies on the filter were left on a LAA-plates and incubated at 37?C over night. The next day the filters were put in chloroform vapour for 10 min before they were thoroughly washed with PBST. The filters were then transferred into an empty petri dish together with 10 ml PBST and 10 ?l casein [10mg/ml]. After blocking 10 ?l a mouse antiE-tag antibody was added. After the filters were washed with PBST and a secondary HRP-labelled goat anti mouse IgG A4416, whole molecules was added and incubated for two h. The filters were washed with PBST four times and then a last time only with PBS before the filters were moved to the development bath containing 10 ml PBS, 2 ml 4-chloro-1-naphtole and 10 ?l H_2O_2. To stop the developing process the filters were washed in water and the E-tag positive clones (visible as dark spots) were counted (Figure 4).

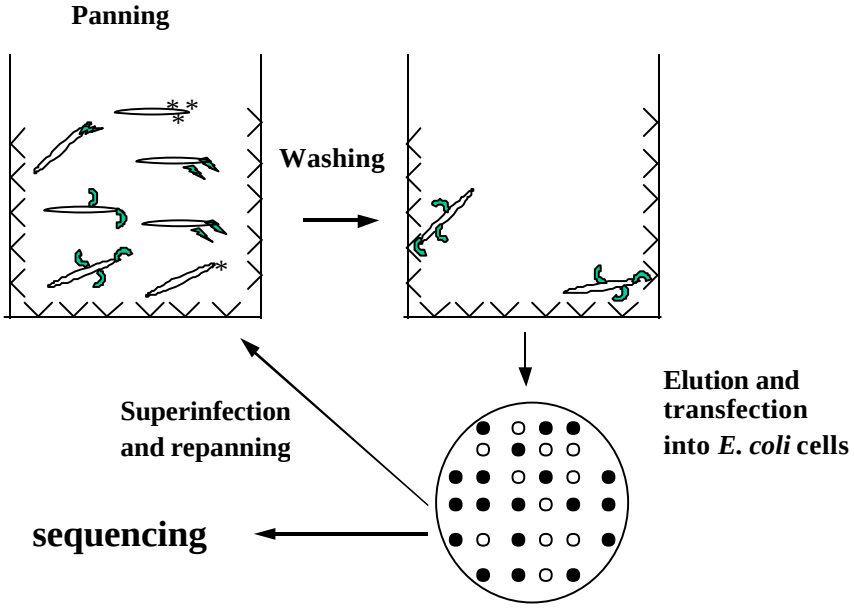


Fig 4. Dark spots represent E-tag positive clones on the filters.

Sequencing and homology studies

Phagemides from E-tag positive clones were prepared and their inserts were sequenced using the primer Sesk or E-tag [20 pmol/ml] and the DYEnamic^{TM} ET Terminator cycle sequencing premix kit (Amersham Life Science Inc. Cleveland, OH, USA) was used together with a PCR-cycle program with preheated lid (95?C for 30 s, 50?C for 15 s, 72?C for 60 s repeated 30 times) was run. The PCR-product was ethanol precipitated and after the centrifugation, washing and drying the pellet was resolved in loading dye and subjected to gel electrophoresis. The obtained sequences were analysed using the computer programs Chromas and PCGENE. In Chromas the sequences were read and interpreted for the final analysis in PCGENE were the DNA sequences were transcribed into amino acids. The inserts of the different clones were aligned with the complete genome published at The Sanger

Centre (http://www.sanger.ac.uk) and program BLAST was used in order to find homologies to other reported sequences.

RESULTS

Panning against horse serum

The phagemid library of subsp. *equi* strain 62 was panned against horse serum. The obtained phage stock from the first panning was screened for the presence of E-tag positive clones and then panned a second time against horse serum (Table 1, S2). The experiment was also performed in a separate test in which soluble eukaryotic proteins (human IgG, human Fg, human serum albumin and bovine casein) were included in the panning process to reduce the clones finally obtained expressing IgG and albumin binding proteins, S1.

Table 1. Cfu/ml and E-tag positives on nitro-cellulose filters with panning against horse serum

	Cfu/ml	**E-tag positive**	**Cfu/ml**	**E-tag positive**
	Serum S1	**Serum S1**	**Serum S2**	**Serum S2**
First panning	4×10^5	-	5×10^5	-
Second panning	6×10^5	~1%	1.4×10^6	10%

S1: soluble eukaryotic proteins added S2: no soluble eukaryotic proteins added

The results from the panning against serum resulted in E-tag positives clones when analysed were found to harbour inserts homologous to the genes encoding the previously characterised subsp. *zooepidemicus* proteins ZAG (Jonsson *et al*, 1995) and FNZ (Lindmark *et al*. 1996, Lindmark 1999), as well as the recently described protein FNE (Lindmark *et al*., 2001). The S1, the inhibited samples, had no E-tag positive clones in the first panning and only two in the second. The S2, the uninhibited samples, the first panning resulted in 0 % E-tag positive clones and the second panning resulted in an increase number of positives. The clones in both samples were sequenced and they also encoded the above mentioned proteins even though it was different clones.

Panning against erythrocytes

To identify potential components interacting with RBC two independent panning experiments were performed. The phagemid library of subsp. *equi* strain 62 was panned against horse

Table 2. Cfu/ml and E-tag positives on nitro-cellulose filters with panning against erythrocytes

	First experiment		**Second experiment**			
	Cfu/ml	**E-tag positive**	**Cfu/ml**	**E-tag positive**	**Cfu/ml**	**E-tag positive**
	RBC first	**RBC first**	**A**	**A**	**B**	**B**
First panning	3.2×10^5	17%	4.6×10^5	15%	2.4×10^5	33%
Second panning	2.6×10^9	85-90%	2.4×10^7	53%	3×10^7	44%

RBC first: the first experiment A: untreated B: treated with salmon sperm DNA

erythrocytes. The horse blood had been washed with PBS in order to reduce the numbers of serum proteins. The experiment was repeated and modified by adding salmon sperm DNA in order to confirm the previous results.

The different experiments resulted in clones harbouring inserts with similar DNA sequence encoding a part of an IS like element homologous to IS861 (Rubens et al., 1989) (Figure 5 and 6). In the first experiment the number of E-tag positive clones in the first panning increased dramatically compared to the second panning. The second experiment with two different samples, A and B, showed little difference compared in E-tag positive clones both between first and second panning as well as between the samples themselves (Table 2). The DNA sequence comparisons of the inserts of the respective clones show that the inserts have similarities after the putative IRL sequence but in the four clones that contained a sequence upstream this element the clones were not similar (Figure 6). They also did not show homologies to other sequences when a BLAST search was conducted (Figure 7). The clones were, with some exceptions, sequenced both from the 5' end (using primer Sesk) and 3' end (using primer E-tag) since the most interesting open reading frame is found in fusion with the E-tag sequence. The putative IS element was similar in all clones with some smaller base pair

S. equi **GTA**TGAACTGCCCCC*AAAAGTTAGACAT
IS861 **ACA**TGAACTGCACCCCAAAAGTTAGACAA

Variations found within the clones:
first three bases: **TTG ATA GAG AAC CAC** no of Cs: **CCCCC CCCCCC**

Fig. 6 Comparison of IRL-sequence between IS861 from *S. agalaticae* and the sequence found in subsp. *equi* strain 62 (underlined bases are subject to variation between individual clones examined)

varieties.

a) match ~70% (positives 83%)
```
S. equi MIRLIDRYGLAIVQKGKNSYYSPELKQEIIDKVLIDGQSQKQTSLDYALPNSSMLSRWIAQYKKNGYTILEKRRGRPPKMGRQPKKTLEQMTE
Cons.   MI+L+DRYG+ IV+KG+N YY PELKQE+IDKVLI G SQ   SLDYAL N S+L+ W++Q+KKNGYTI+EK RGRP KMGR+ KKT E+MTE
IS861   MIKLMDRYGVEIVEKGRNEYYPPELKQEMIDKVLIHGCSQLSVSLDYALSNCSILTNWLSQFKKNGYTIVEKPRGRPSKMGRKRKKTWEEMTE
```

b) match ~87% (positives 91%)
```
S. equi  MIRLIDRYGLAIVQKGKNSYYSPELKQEIIDKVLIDGQSQKQTSLDYALPNSSMLSRWIAQYKKNGYTILEKRRGRPPKMGRQPKKTLEQMTE
Cons     MIRLIDRYG+ IVQK KN YYSPELKQEII+KVLIDGQSQKQTSLDYALP SSMLSRWIAQYKKNGYTILEK RGRP KMGR+ KK LE+MTE
IS861ass MIRLIDRYGVTIVQKCKNHYYSPELKQEIINKVLIDGQSQKQTSLDYALPTSSMLSRWIAQYKKNGYTILEKPRGRPSKMGRKRKKNLEEMTE
```

Fig 7. a) Comparison between hypothetical protein 1, *S. agalactiae* IS861 (Rubens et al. 1989) and the sequence found in subsp. *equi* strain 62
b) Comparison between conserved hypothetical protein, IS861 (Ferretti et al. 2001) associated in *S. pyogenes* and the sequence found in subsp. *equi* strain 62

DISCUSSION

Panning against horse serum

The bacteria can produce extracellular cell surface bound proteins interacting with host compounds like extracellular matrix proteins or soluble proteins of the host. In addition the bacteria can produce toxins affecting the cell membrane of host cells.
S. equi might have the possibility to interact with host cells in various ways. Furthermore it can not be ruled out that the bacteria possesses the ability to produce proteins which bind to carbohydrate structures on host cells which opens new possibilities for the bacteria to interact with the host.

The panning against serum resulted in several clones harbouring inserts with homologies to the previous reported genes *fnz* and *zag*. Interestingly one clone harbouring an insert homologous to a *fnz* like gene was found. This insert was also identified by another student in our research group (Malin Florby, data not shown) who continued the characterisation of this putative protein.

Panning against erythrocytes

The panning against RBC resulted in an enrichment of E-tag positives clones. Surprisingly the isolated and characterised inserts from a number of E-tag positives clones revealed that the insert had similarities to a previously reported IS-element of *S. agalactiae* (Rubens *et al*, 1989).

The panning process with blood itself must be considered as an experimental process under development and evaluation and therefore the results may be questionable. One possible explanation for the IS like element to show up would be that the RBCs are negatively charged and the expressed orf1 on the outside of the phagemids are positively charged, thus creating an electrostatic binding in the laboratory environment. The second time the panning was proceeded salmon sperm DNA was added in order to see if that made an effect on the bindings. That was, however, not the case (Table 2). It is hard to say if these results would turn up in an *in vivo* investigation since the IS element is an intracellular component. The full understanding and detectable use of these elements are still to be investigated in the future.

IS-like elements in *Streptococcus dysgalactiae* has been described by Vasi et. al. 2000 and have similarities to the structure described in this work. The IS-element found here turned out to have large similarities to the IS element 861 in *S. agalactiae* (Rubens et al, 1989) and also to the recent described IS associated protein 861 in *S. pyogenes* (Feretti *et al*, 2001). These results indicate that IS elements are present in many of the genomes of various streptococci and that they might play an important role in the pyogenic group of *Streptococcus*.

The IS associated protein 861 consists of two open reading frames, orf1 and orf2, as well as inverted repeats IRL and IRR. The panning results in this work only shows an expression of orf1 and none of a potential orf2 or IRR which would be present in an IS element. The whole hypothetical element was put together through shotgun clone walking made possible by the database at Sanger (www.sanger.ac.uk) where almost the complete genome is available. Since the sequence found have variations compared to the sequence found in strain 62, the IRL and IRR are not exactly alike. This may be a result due to the fact that when the different shot gun clones were assembled the resulting genome sequence could be a consensus sequence

between several different IS element situated in various positions in the genome. It is also possible that the IS elements differ between strains within the subspecies.

The smaller heterogeneities between the sequences found in the panning may be a result due to the sequencing or might indicate that several copies of the IS element are present in the genome. It is not a strange assumption to make that there would be several copies of the IS element in the genome, nor that the different copies would display small variations.

Concluding remarks

The possible use of IS elements found in subsp. *equi* are not obvious but it could be a future possible way of investigate epidemiological transfer paths as well as investigate how these elements affect the genome and the virulence of the bacteria i.e. between different strains. The most interesting part would be to see in what number these elements are present in various strains of subsp. *equi* and subsp. *zooepidemicus* and other streptococcal species. It would also be interesting to try and see if these variations affect the virulence of different strains in subsp. *equi*. In order to investigate this a Southern blot analysis would be a good method to use and the different bands that show up would tell in how many various positions the element can be found.

Finally, these elements may not be important in the vaccine process or even in the epidemiological studies but studies of these elements may contribute to the overall picture of what the genome of subsp. *equi* looks like and the structure of IS elements present in *S. equi*.

References

Bass S., Greene, R & Wells, J.A., 1990, Hormone phage; an enrichment method for variant proteins with altered binding properties, Proteins 8:309-314

Brock, T., 1997, Biology of Microorganisms 8:th edition, (eds: Madigan, M., Martinko, J., Parker, J.), Prentice-Hall Inc, Simon & Schuster, New York, USA

Farabaugh, P.J., 1996, Programmed translational frameshifting, Microbiology Review vol 60(1):103-34

Ferretti JJ, McShan WM, Ajdic D, Savic DJ, Savic G, Lyon K, Primeaux C, Sezate S, Suvorov AN, Kenton S, Lai HS, Lin SP, Qian Y, Jia HG, Najar FZ, Ren Q, Zhu H, Song L, White J, Yuan X, Clifton SW, Roe BA, McLaughlin R., 2001, Complete genome sequence of an M1 strain of Streptococcus pyogenes, Proc Natl Acad Sci 98(8):4658-63

Hardie, J.M & Whiley, R.A, 1995, The genus *Streptococcus*, In "The Lactic Acid Bacteria" (eds. Wood, B.J.B & Holzapfel, W.H.) vol 2, Blackie Academic & Professional, Glasgow, UK

Jacobsson, K. & Frykberg, L. 2001, Shotgun Phage Display Cloning, Combinatory Chemistry & High Throughput Screening. 4:135-143

Jacobsson, K., 1997, Shotgun Phage Display, Doctoral thesis, Uppsala Sweden

Jonsson, H., Lindmark, H & Guss, B. 1995, A Protein G-Related Cell Surface Protein in *Streptococcus zooepidemicus*, Infect. Immun. 63(8):2968-2975

Kleckner, N., 1990, Regulation of transposition in bacteria, Annual Rev. Cellbiology vol 6:297-327

Lederberg E. M., 1981, Plasmid reference center registry of transposon (Tn) allocations through July 1981, Gene 16(1-3):59-61

Lindmark, H., Jacobsson, K., Frykberg, L. & Guss, B., 1996, Fibronectin-Binding protein of *Streptococcus equi* subsp. *zooepidemicus*, Infect. Immun. 64(10):3993-3999

Lindmark, Hans, 1999, Doctoral thesis Characterization of adhesive Extracellular proteins from Streptococcus equi, Uppsala, Sweden

Lindmark, Nilsson and Guss, 2001, Comparison of the fibronectin-binding protein FNE from *Streptococcus equi* subspecies *equi* with FNZ from *S. equi* subspecies *zooepidemicus* reveals a major and conserved difference. Infect. Immun May 69(5):3159-63.

Mahillon, J. & Chandler, M., 1998, Insertion sequences, Microbiology and Molecular Biology Reviews 62(3):725-774

Murrey et al, 1986, Bergey's manual of systematic bacteriology vol 2, Williams and Wilkins, Baltimore, USA, pp 1043-1053

Navarre, W,W. & Schneewind, O. 1999, Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope, Microbiol. Mol. Biol. Rev. 63:174-229

Rubens, C.E., Heggen, L.M, Kuypers, J.M., 1989, IS861, a group B Streptococcal Insertion sequence related to IS150 and IS3 of Escherichia coli, J. of Bact. 171(10):5531-5535

Ruoff, K.L, 1992, The prokaryote, (Balows et al eds.) Springer-Verlag New York inc., New York, USA pp 1450-1461

Singer, M. & Berg, P., 1991, Genes and genomes, University science books, Blackwell scientific Publications, Oxford, UK

Smith, G.P., 1985, Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface, Science, 228:1315-1317

SVAs hästbok, 1998, Sveriges Veterinärmedicinska Anstalt, Uppsala, Sweden

Timoney, J.F, 1988, Protecting against 'Strangles': A contemporary view, Equine Vet. J, vol 20(6):392-396

Timoney, J.F., 1993, Strangles, Vet Clinical Journal North America Equine Pract., Vol 9:2 pp 356-374

Vasi, J, 2000, Doctoral thesis Characterization of two potential virulence factors in *Streptococcus dysgalactiae*, Uppsala, Sweden

Weaver, R. F. & Hendrick, P. W., 1995, Basic genetics 2nd ed, Wm C Brown Communications, Duberque, USA

Acknowledgement

I would like to thank my supervisor Bengt Guss for all help and encouragement as well as my co-supervisors Karin Jacobsson and Lars Frykberg. I also want to thank the PhD students in the adhesion group and the other PhD students, scientists, researchers and other staff at the Department of Microbiology. Also thanks to Susanna Sternberger and Per Jonsson for lending me photos of horses infected with *S. equi* and sick in Strangles for my oral presentation.

Appendix 1

Nucleotide sequence found in *S. equi* subsp. *equi*. Underlined sequence was identified by sequencing various inserts of phagemid clones panned against erythrocytes whereas the remaining sequence was obtained by computer analysis using the sequence database of *S. equi* (http://www.sanger.ac.uk/Projects/S_equi/).

Appendix 1

IR$_L$ Orf1

GTATGAACTGCCCCCAAAAGTTAGACATAAAATCTAACAATTGGGGGGCTATTTTT**ATG**A 60

M N C P Q K L D I K S N N W G A I F M

CATTGAGTTATGAAGACAAGGTTCAAATCTATGAGCTACGGCACATTGGAAAGTCCATTA 120

T L S Y E D K V Q I Y E L R H I G K S I

AATGCTTATCAGAAAAGTTTAGTATTGCAGAATCTGACCTCAAATACATGATTCGCCTGA 180

K C L S E K F S I A E S D L K Y M I R L

TTGACAGGTATGGGTTAGCCATTGTCCAAAAAGGTAAGAATAGTTATTATTCTCCAGAAC 240

I D R Y G L A I V Q K G K N S Y Y S P E

TGAAGCAAGAGATAATAGATAAAGTTCTGATTGATGGTCAATCTCAAAAACAGACGTCCT 300

L K Q E I I D K V L I D G Q S Q K Q T S

TAGACTATGCTTTACCAAATTCTAGTATGCTTTCAAGGTGGATAGCGCAATACAAGAAAA 360

L D Y A L P N S S M L S R W I A Q Y K K

ACGGCTACACTATTCTTGAGAAAAGAAGAGGGAGGCCACCAAAGATGGGACGTCAACCAA 420

N G Y T I L E K R R G R P P K M G R Q P

AGAAGACTTTAGAACAAATGACAGAGTTGGAGCGACTCCAAAAAGAATTAGACTACCTTA 480

K K T L E Q M T E L E R L Q K E L D Y L

GAGCGGAGAATGCTGTGCT**AAAAAA**GCTGAGAGAATACCGGTTGAGGGACGAAGCAAAGC 540

S G E C C A K K A

L K K L R E Y R L R D E A K

Orf2

TCAAAGAGCAACAGAAATCATCCAAGAATTAATCGGTCAATTNTCTCTAGCAACTTTGC**T** 600

TGAAATCCTTGATTTATCGCGGTCAACCTATTATTATCAAGTCAAGCAACTAGCTCAAGA 660

AGATAAGGACATGGACTTAAAGGAGCTCATTCANRGCATCTATGATGAACATCATGGNCA 720

NTTATGGCTATCGTCGCATTCATCTGGAACTAAGAAATCGTGGTTTTATCGTCAATCACA 780

AAAAAGTACAACGTTTGATGACTGTCATGGGCTTAAAAGCTCGTATCCGTCGTAAGCGCA 840

AGTATTCTTCTTACAAAGGTGAGGTTGGCAAAAAGGCTGATAATCTGATTAAACGTCAGT 900

TTGAAGGTTCTAAGCCCTACGAGAAGTGCTATACCGATGTGACGGAATTTACCTTACCTG 960

AGGGGAAACTCTATCTATCGCCTGTTCTTGACGGCTATAACAGTGAGATTATTGATTTCA 1020

CCCTGTCTCGATCGCCTGACTTGAAGCAAGTACAAACCATGCTTGAGAAGGCTTTTCCAG 1080

CGGATTCGTACAATGGAACGATTCTCCACAGCGATCAAGGCTGGCAATATCAACATCAGT 1140

CTTATCATCACTTTTTGGAGACTAAAGGCATTCGTCCATCCATGTCTCGCAAGGGAAATA 1200

GTCCAGATAATGGGATGATGGAGTCCTTCTTTGGTATTCTCAAATCTGAGATGTTTTACG 1260

GCCTTGAGACAACTTATCAATCCCTTAATGAGCTTGAACAAGCTATTACAGATTACATTT 1320

TTTACTACAACAACAAACGCATTAAAGCAAAGCTAAAAGGACTTAGCCCTGTGCAATACA 1380

IR$_R$

GAACTAAATCCTTTCACTAAT**ATGTCGCGTCCAACTTTTTGCGGTCAG**TACACAGGATAG 1440

ACAG 1444